MSc. Thesis

# Mapping Tourism in Aruba: Leveraging Geosocial Data for Sustainable Development

| | |
|---|---|
| Title | Mapping Tourism in Aruba: Leveraging Geosocial Data for Strategic Decision-Making and Sustainable Development |
| Name | Margreet Havermans<br>6384137<br>m.c.l.havermans@students.uu.nl |
| Commissioned by | University of Twente/ITC Enschede, TU Delft, Utrecht University and Wageningen University<br>Geographical Sciences (crohocode 60732)<br>Geographical Information Management and Applications (GIMA) |
| Supervisor | prof.dr. I (Ioannis) Athanasiadis<br>Wageningen University & Research |
| Responsible Professor | prof.dr.ir. L (Lammert) Kooistra<br>Wageningen University & Research |
| Thesis written on behalf of | dr. PJFM (Peter) Verweij, DLO researcher<br>C (Charlotte) van Haren, DLO researcher<br>Wageningen Environmental Research |
| Date and place | February 23, 2024 Utrecht, The Netherlands |

# Abstract

This research maps the spatial distribution of several types of visitors in Aruba by using geosocial data from Flickr image content. Computer processing power combined with the analytical ability of the researcher the method is formed. With these methods geovisualizations and charts are made to examine the geosocial data, discover patterns, and make conclusions based on the results. 19.189 images were extracted from Flickr and stored in an SQLite database. They were categorized into six tourist categories which resulted in a spatial distribution map for each category where eventually 18.300 images were accounted for. So the division in land of origin of the photographer and their time of visit is taken into account. The results also show which land use type and protected nature or marine areas many tourists go. A critical assessment is executed whereby the findings of the research are presented to several tourism agencies in Aruba to see if the findings correspond with their perception of reality, and the numbers of this research are compared to official statistics. The outcomes showed that the results were useful for these organizations to comprehend the movement of tourists and strive to regulate it. It considers both environmental sustainability (preservation of nature) and the locations of visited attractions, as well as devising strategies to manage, improve, and sustain them. is study contributes to a deeper understanding of tourist behavior patterns and offers insights into nature-inclusive tourism management in Aruba.


**Keywords**: Aruba, Flickr, Geosocial data, Spatiotemporal distribution, Tourism.

# Acknowledgment

For this thesis, I would like to express my gratitude to my supervisor, Ioannis, who has always been helpful and kind. When we first met, I was somewhat uncertain about my programming abilities. Before this thesis, I had never used R and had only employed Python and SQL during the GIMA courses, albeit at a basic level. Although I felt compelled to excel in programming, Ioannis reassured me by emphasizing that the thesis is a testament to academic capabilities and independent research. While programming is one way of conducting research, there are multiple approaches. Moreover, he did not want me to be a stressed-out student, as a thesis is not meant for that purpose. This assurance accompanied me throughout the thesis, motivating me to explore the programming aspect. If it did not align with my expectations, I knew I had the option to design another research project.

Surprisingly, the learning curve for programming was faster than I anticipated. The Flickr API proved to be clear, and resolving errors through online searches resulted in quick solutions. Scraping geosocial data from Flickr also proceeded faster than initially envisaged. The most challenging aspect for me was categorizing the images. Initially exploring various options, I found it difficult to decide what was feasible for me without causing excessive stress. Since this research is part of the TRUPIAL program, almost the same method used in the article by Slijkerman et al. (2020) was chosen. This decision was influenced by the desire to compare results for other research easily. Peter Verweij from WenR, for which the thesis was written, agreed to this approach.

I found out that emulating a method is maybe even harder than making your method and scrips. Although adjustments were made to fit the method to Aruba, most parts of the methods are the same. I eventually decided that my code was better for the Flickr API as I didn't understand the given script very well. For the ‘PhotoCategorizer’, it also took a long time before everything worked the way I wanted it to and everything was set up correctly. I had to work with programs I had never dealt with before. This was a huge challenge, but I am all the more proud of myself and this is why I succeeded.

The method employed in Slijkerman et al. (2020) utilizes a 'PhotoCategorizer' designed and created by Jan Tjalling van der Wal. I extend my thanks to Jan for sharing this tool with me and explaining its functionality. Although the Python script was somewhat outdated and required adjustments, it eventually worked and a new version was made. I appreciated the fact that I could always reach out to Jan via email for any clarifications. Following the categorization of images, I had a meeting with Peter Verweij to discuss which results were most useful for him, and the geosocial data was shared. I am happy to share that Peter Verweij has already mentioned and used my results in another research.

The additional surveys I sent to tourism agencies for feedback made it clear that conducting surveys is time-consuming. Using geosocial data has been a game-changer in making this research possible. I managed to get responses from seven people, but it took at least five emails each to get a reaction. Especially the fact that I was not able to visit Aruba made it more challenging. Everyone's so busy, it's a real challenge trying to reach out to all the tourists in Aruba.

# Table of contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| AIA | Automatic Image Annotation |
| API | Application Programming Interface |
| CSV | Comma-separated values |
| CBIR | Content Based Image Retrieval |
| DMO | Destination Marketing Organisation |
| GDP | Gross Domestic Product |
| GIS | Geographic Information Systems |
| HTML | Hyper Text Markup Language |
| JSON | Javascript Object Notation |
| UGC | User Generated Content |
| URL | Uniform Resource Locator |
| POI | Point of Interest |
| PUD | Photo User Days |
| REST | REpresentational State Transfer |
| SIDS | Small Island Development State |
| SMA | Social Media Analytics |
| SQL | Structured Query Language |
| TDI | Tourism Destination Image |
| TRUPIAL | social-ecological TRansformation for bottom-UP Integrated Approach in caribbean Landscapes |
| VGI | Volunteered Geographic Information |
| WCDI | Wageningen Centre for Development Innovation |
| WECR | Wageningen Economic Research |
| WENR | Wageningen Environmental Research |
| WMR | Wageningen Marine Research |
| WPR | Wageningen Plant Research |
| XML | Extensible Markup Language |

# 1 Introduction

Aruba is a Caribbean Island that attracts a lot of tourists each year. Tourists come to Aruba for the white sandy beaches, crystal clear waters, and hospitable culture. According to the world travel and tourism council report 2022, Travel & Tourism Relative Contribution to the GDP of Aruba was 67.9% in 2019 and 59.6% in 2021. The report states that in 2022 the sector would grow even faster and was expected to return to the level of 2019 (WTTC, 2022). In 2019 two million tourists visited the relatively small island, which has 106.000 inhabitants. Besides Aruba has one of the highest densities of tourism and population in the Caribbean (Cole & Razak, 2004). Examining tourist preferences and patterns of interest on the Caribbean Island of Aruba could serve as a valuable stride toward enhancing the management of tourism. While tourism has indisputably catalyzed economic expansion on the island, it has concurrently exerted considerable stress upon the island's ecosystems (ATA, 2019).

In this research, an approach will be adopted leveraging geosocial data for the analysis of tourist image-worthy places and tracing the contours of their specific preferences. The primary objective herein is to contribute to the enhancement of strategic decision-making in the realm of tourism. Finding out which categories of tourists visit Aruba where and how to extract the geosocial data from the internet including the metadata is an interesting source of information. Metadata includes information about when, where, and what tourists visit Aruba (Lusiana et al., 2020). Employing advanced techniques on vast repositories of geosocial data, with the overarching aspiration to foster greater sustainability within the tourism industry (Wood et al., 2013).

Nowadays people use social media to express their feelings and show their daily lives, and thus their perceptions can be interpreted (Marine-Roig & Clavé, 2015). The advantage of this is that not only the people with a loud voice will be incorporated in the research but the ordinary people that use social media to share will be as well. Therefore, *user generated content* (hereafter, UGC) is the new data source that researchers use to find people’s *tourism destination image* (hereafter, TDI) (Deng et al., 2018). Especially images posted on social media are extremely valuable to measure people’s TDI. They could contain geographical information such as the coordinates of the image taken, have extra information about the photographer (username, origin, etc.) and they have additional metadata such as a description or a hashtag that is added (Deng, et al. 2019). The type of UGC used in this research is voluntarily posted on social media and made publicly available. It is *volunteered geographic information* (hereafter, VGI) which is geotagged geosocial data. To map the tourism of Aruba these geotagged social media are extracted from the internet (Yan et al., 2022).

## 1.1 Research problem and objectives

### *1.1.1 Research problem*

Tourist visits typically exhibit a higher degree of spatial concentration compared to visits made by residents. They are focused on *Points of Interest* (hereafter, POI) or landmarks, which usually also have good accessibility (Muñoz et al., 2019). In comparison to nature areas, there is a division in how tourists and residents perceive the wilderness (Ghermandi et al., 2020). Even if they visit the same locations for example a city, the activities they do in the city are different between locals and tourists (Muñoz et al., 2019; Séraphin et al., 2020). Therefore, it is valuable to know where the tourists travel to. Secondly, the amount and type of tourists that visit vulnerable nature areas, can be harmful. To map what type of tourists visit which destinations is therefore valuable information for planners and destination marketing (Fyall & Garrod, 2020).

The relationship between geosocial data and tourism has been studied previously. In the scientific literature, some research has been conducted to map nature-based tourism (Da Mota & Pickering, 2020; Jailani et al., 2021; Mirzaalian & Halpenny, 2021; Schep et al., 2016; Slijkerman et al., 2020). However, in Aruba, this has never been done before and therefore this research contributes to the scientific literature. With the pressure of tourism on Aruba’s nature and time-consuming methods such as surveys and interviews, the use of geosocial data is interesting to apply (Callegaro & Yang, 2018; Hassink et al., 2015). Geosocial data can handle bigger sample sizes, frequency is higher, and level of detail is better than visitor surveys. Research has been conducted into forecasting tourism demand through search queries and machine learning (Kort, 2017). However, what type of tourists visit Aruba and which environments they visit has not been discovered yet. Therefore, this research will provide insight into the perceptions of tourists which in turn can be used by destination marketing organizations and policy makers to act upon these findings. The use of social media by tourists who visit Aruba will be analyzed to explore their preferences and interests.

### *1.1.2 Research objective*

This research is part of a research program named TRUPIAL (*social-ecological TRansformation for bottom-UP Integrated Approach in caribbean*) from *WENR, WCDI, WECR, WPR, and WMR* which focused first on nature-inclusive planning for Bonaire 2050 with a positive future for people and nature. After this successful research program, the aim is to transfer the method applied for Bonaire to Aruba. This research is therefore part of this TRUPIAL research for Aruba and focuses on the tourism effect on nature and people.

This research tries to map what tourists are most interested in, the places they go to, the duration of their stay, and their origin. The objective of this research is to explore both the potential as well as the constraints associated with collecting geosocial data to estimate the geographical patterns of tourist activity and to facilitate the advancement of sustainable tourism initiatives, this is done by comparing geosocial data with nature and marine protected areas.

The research objective is to investigate methods to map diverse types of tourists in Aruba with geosocial data. The result of this research explores the different options on how to map tourism, what are the methods and technologies available? How to collect and analyze geosocial data? What are the limitations and challenges? And eventually, how can these findings be used in practice? First, the main research question will be formulated. Second, the concrete sub-questions are displayed and finally, the research limitations and scope are explained.

**Main research question:** *How can geosocial data be utilized to map the spatial distribution of tourism in Aruba and contribute to strategic decision-making in the realm of tourism management?*

**Sub-Questions:**

Sub-question 1: How can social media analysis contribute to understanding tourist preferences?
Sub-question 2: How can geosocial data be analyzed to create meaningful insights for mapping tourism interest?
Sub-question 3: What collecting methods and technologies are available for collecting and analyzing geosocial images in the context of tourism mapping?
Sub-question 4: What are the implications of using platforms such as Instagram, TripAdvisor, Pinterest, and Flickr for gathering geosocial data in the context of tourism mapping in Aruba?
Sub-question 5: What are the various categories of tourists visiting Aruba and how do they distribute across different environments?
Sub-question 6: What insights can be drawn from the spatiotemporal analysis of geosocial data regarding the impact of tourism on the environment and tourist characteristics in Aruba?

In conclusion, the findings of this study offer a deeper understanding of tourism interest in Aruba, linking tourist types with various environmental sites. The research explores the implications of these findings for Aruba's tourism industry and government, emphasizing the potential for data-driven decision-making. Secondly, as mentioned above, this research is also carried out to see if the method used for Bonaire is transferrable to Aruba.

## 1.2 Research scope

Before answering these questions, it is important to also state the research limitations by specifying the scope of the project. The main goal is to categorize the tourists who visit Aruba and map tourists with the help of social media. Social media is still very broad and needs some more specification. In this research images from a bounding box around Aruba between 2012 and 2022 are extracted from Flickr together with their metadata. A bounding box is a rectangular area set around a geographic area, such as the island of Aruba, and is defined by a set of coordinates that encompass the minimum and maximum latitude and longitude values for that area. It encloses the desired geographic area (Bessagnet et al., 2022).

There are several limitations within the research which are not incorporated in this research. Firstly, this study does not cover deep learning and machine learning techniques. Secondly, it does not compare different methods to gather geosocial data from several platforms. Thirdly, due to new privacy and ethical regulations, there is limited access to several platforms and metadata. It caused some social media platforms to become inoperable.

## 1.3 Reading guide

The next chapter gives an overview of the related work on executing tourism research with geosocial data. As a result, the first and second sub-questions of how geosocial data can be analyzed and the current collecting methods and technologies for analyzing geosocial data are explained. The third and fourth and the first part of the fifth sub-question is answered in Chapter 3 where also the method used within this research is explained. The geosocial data extraction, preparation, and classification are outlined in this chapter. The results are presented in Chapter 4, which gives an answer to the second part of the fifth and addresses partly the sixth sub-question. First, the overall dataset is explained followed by a mapping of the spatial distribution of several types of tourists. The origin of the photographers and their spatial distribution are looked at. Also, the duration of the visit is applied and mapped to see the difference between a cruise, stay over tourists, and local. Eventually, the findings from the interviews with the tourism agencies are elaborated on, which also addresses sub-question 6. In the discussion and conclusions, the limitations and findings of this research are presented. Thereafter, recommendations for further research are provided. Finally, some final words are written.

# 2 Related work

First, this chapter will explain the role of social media in tourism and especially the effect of images. Secondly, the rise of social media in scientific research is addressed. Discussing why geosocial data is chosen, its significance, and how it is applied. Finally, several methods for analyzing geosocial data such as spatiotemporal analysis, content analysis, and network analysis are discussed. This chapter answers the first and second sub-questions: '*How can social media analysis contribute to understanding tourist preferences?*' And '*How can geosocial data be analyzed to create meaningful insights for mapping tourism interest?*'

The role of social media in tourism

There is an overwhelming amount of data generated by digital technologies and new information sources. This vast and diverse data is known as big data (Che et al., 2013; Hausmann et al., 2018). Social media plays a crucial role in preserving and sharing individual experiences through images and comments (Latorre-Martínez et al., 2014). These platforms facilitate sharing with both close contacts and strangers. Together it forms a dynamic network for sharing information about tourist experiences and destinations. Users can interact by commenting, sharing, rating, tagging, and posting their content (Latorre-Martínez et al., 2014).

Tourism relies heavily on images, both in marketing to attract potential tourists and also increasingly on the images tourists take on their journeys themselves. These images represent the fundamental characteristics of the tourist destination and can be seen as '*The Hermeneutic Circle of destination representation*' (Jenkins, 2003; Xiaoyang, 2019). As can be seen in Figure 2-1, these images are disseminated by public and private entities to attract tourists and become iconic landmarks that tourists aim to experience (Lojo et al., 2020). The integration of digital photography in mobile devices allows tourists to not only capture visual memories but also share them on social media. This process solidifies the social representation of a specific destination, shaping the collective imagination. Digital photography and social media serve as crucial tools in shaping and disseminating a destination's image, and revolutionizing tourist marketing (Lin et al., 2021).

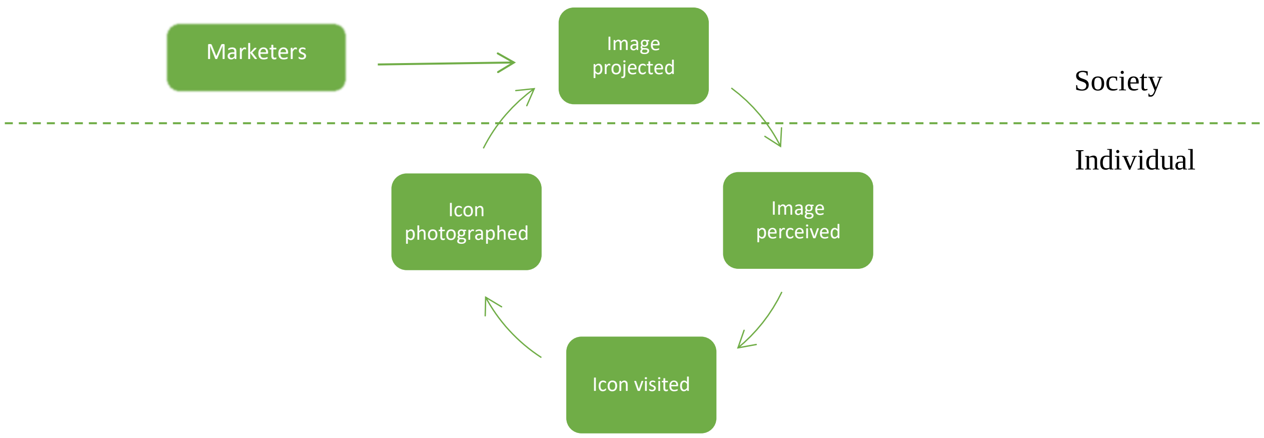



*Figure 2-1: The Hermeneutic Circle of Destination Representation (Jenkins, 2003)*

Starting from the top, the mass media collectively projects images of a particular destination. These images are perceived by individuals, potentially sparking their desire to travel to that destination. Once at the location, tourists are inclined to visit the prominent attractions or iconic spots they had seen in those projected images. They often document their experiences using cameras and later share these personal photographs with friends and family, partly as evidence of their visit. These images can be seen as another form of image projection, effectively initiating a cycle where they influence the perceptions of the destination held by others.

Furthermore, tourism advertisers and marketers, intending to disseminate appealing imagery, also play a role in this process of image projection (Calcagni et al., 2019; Jenkins, 2003). Moreover, the impact of social media imagery on tourists' expectations, experiences, and overall satisfaction with destinations cannot be understated (Miller et al., 2019).

## 2.1 The rise of geosocial data in scientific research

### *2.1.1 Why use geosocial data?*

Traditional surveys are limited in space and time. Surveys heavily rely on individuals' recollections whereby issues could arise such as accuracy, memory recall biases, and nonresponse biases. Conversely, among tourists, the information extracted from geosocial data tends to be perceived as credible and independent. *Social Media Analytics* (hereafter, SMA) has more options for in-depth analysis related to space, time, and various subgroups. Online image posts on social media platforms empower researchers to capture a wide range of communicable attributes of TDI, reflecting both tangible and intangible aspects of the destination. They can be utilized to trace changes in the destination image over different periods (Callegaro & Yang, 2018). However, the downfall of SMA is the accuracy and the structure of the data is cluttered. Here also the validity is a concern, only a handful of tourists share their perceptions on social media and a specific platform. Data availability is becoming more and more of a concern within SMA, due to privacy regulations getting access to the data gets increasingly harder. Data integrity is a big concern within SMA research (Owuor & Hockmair, 2020). And eventually, the interpretation of the images meaning by the researcher plays a big role in SMA research (Stepchankova et al., 2013).

### *2.1.2 The significance of social media in scientific research*

The use of social media in tourism promotion has garnered significant attention from the global scientific community researching emerging technologies in tourism (Marine-roig & Clavé, 2015). The rise of image-centric platforms such as Flickr, Instagram, Pinterest, Facebook, and Panoramio highlights the significance of digital photography as a means of expression and communication. The shared images offer credible proof of journeys and vacations, enabling researchers to develop innovative observation and analysis methods to comprehend tourism dynamics (Giglio et al., 2019). This extends not only to friends and family but also to geographically distant strangers (Li et al., 2023). Lo et al. (2011) revealed that 76.1% of travelers share their travel images on platforms such as Flickr, Instagram, or similar social media. Another noteworthy aspect is that tourists tend to place greater trust in images and opinions from fellow travelers in comparison to those offered by official companies and destinations (Lo et al., 2011). This opens a new avenue for research where social media can serve as crucial sources of knowledge for tourism companies and institutions (Latrrorre-Martínez, 2014).

Using social media in scientific research can be a valuable new source of information, which has its advantages and disadvantages, but due to its widespread nature can be of great added value in scientific research, especially for collecting research data, finding respondents, and testing hypotheses. Numerous review studies and meta-analyses offer insights into the analysis of data derived from various social media platforms and the diverse applications of social media apps in different contexts and settings (Ouwuor & Hockmair, 2020).

### *2.1.3 How is social media used in scientific research?*

The use of geosocial data for evaluating tourism is still a relatively new area of research. It first emerged in 2012, coinciding with the launch of some of the most popular social media platforms at the time such as Instagram and Facebook (Da Mota & Pickering, 2020).

Using geosocial data to examine tourism has several methods such as spatial analysis which focuses on POI, hotspots, and clusters but also on where people post these images, photo user days (PUD), which is a metric that determines how many different users submit at least one image in a specific location on a given day. Research is also looking at and evaluating CES (Kim et al., 2021; Zhang et al., 2022; Runge et al., 2020; Hale, 2018), sentiment analysis, and descriptive analysis, which focuses on text to discover the perceptions and values of the tourists give more insight into the feelings (Becken et al., 2017; Manoharan & Sathesh, 2020; Xiang et al., 2017).

Some papers compare their findings with surveys or expand them with interviews and visitor counts (Lin et al., 2021; Sultan et al., 2020; Zhang et al., 2021) while others compare different areas. In addition, several social media platforms can be compared with one another, such as Flick, Panoramio, Facebook, and Instagram (Da Mota & Pickering, 2020; Van Zanten et al., 2016). Some researchers focus on factors that possibly influence a tourist such as social-demographic characteristics such as age, gender, and income, and estimating the economic values with travel costs methods (Rashidi et al., 2017).

In addition, also environmental factors such as natural parks/areas, natural disasters, and the use of social media are taken into account to monitor the visitors or calamities afterward (Ding et al., 2021; Heikinheimo et al., 2017; Kim et al., 2021; Memon et al., 2015; Yan et al., 2017). Researchers also look at people's travel patterns or human mobility of tourists such as tracking someone's posts on Instagram. (Chua et al., 2016; Kovács et al., 2021; Majid et al., 2013). Network mapping analysis, whereby the focus lays on followers/following you can see what groups are formed when looking at the COVID-19 pandemic. A network of connections can yield valuable insights into patterns of communication and the spread of information (Park et al., 2016).

## 2.2 Different methods for analyzing geosocial data

For analyzing geosocial data there are several methods, the three overarching categories that are most used are Spatiotemporal analysis, content analysis, and network analysis. Below these three categories are explained, which all have their sub-categories.

### *2.2.1 Spatio-temporal analysis*

Various data types can be extracted from the internet such as texts, images, and route data, which all contain geographical information and a time stamp. An analysis of this data can lead to insight into spatial and temporal patterns (Ghermandi & Sinclair, 2019; Hamstead et al., 2018).

Spatial analyses focus on where people post their images, the specific trajectories of one person, or for example the origin of the tourists (Barros et al., 2020; Hasnat & Hasan, 2018; Maeda et al., 2018). Temporal analysis is done by looking at seasonal, yearly, or monthly variations or patterns. Also, predicting the number of tourists is valuable for tourism agencies and policymakers who are interested in these numbers (Heikinheimo et al., 2017; Reif & Schmücker, 2020; Tenkanen et al., 2017). However, spatial and temporal analysis are mostly combined and therefore this category is called spatio-temporal analysis. The focus here lies on the timestamps and the geographical details of the geosocial data (Yan et al., 2017).

Broadly, there are two key categories of spatio-temporal analysis approaches that prove valuable in the realm of geosocial data analytics: location-based and person-based analyses (Toivonen et al., 2019). With the arrival of smartphones, people were no longer bound to their computers. Sending location data into the world became more mobile and thus created a new traveler, who could share their experiences whenever and wherever they desired. This has led to more information about where and when people visit a place (Gretzel, 2018; Jimenez, 2012).

#### *2.2.1.1 Person-based approach*

The person-based approach focuses on the individual routes of a tourist in a national park, whereby movement trajectories span but on the other hand can focus on the origin of all social media users. More methods for human mobility analysis are used for detailed examination of visited sites and movement paths over a longer period. These methods include sequence alignment to for example generate exploration-based route plans in a small-sized city. But another example is space-time prisms showing the potential range of movement within an individual's world when accounting for various constraints on movement for example could give insights into where local events are detected (El Ali et al., 2013; Krumm & Horvitz, 2015; Seo & Cho, 2021). Looking at an individual trajectory can provide valuable insight into someone's preference or behavior. The posted images and texts represent people's interest and their activities during their visit (Sun et al., 2015). These insights can lead to classifying the visitors into different categories, for example, backpackers, ecotourists, thrill-seekers, cultural aficionados, diving tourists, cruise visitors, etc (Gehrmandi et al., 2020).

#### *2.2.1.2 Location-based approach*

The location-based approach gives the spatial characteristics of the social media posts and users. The frequency of tourists can be monitored but also predicting tourist counts and their trajectories. It has different outcomes such as mapping the location of the image with points, creating density maps, hot spot detecting, and grid-based maps (Hirota et al., 2014; Lee & Tsou, 2018; Wibowo et al., 2019). Shared images offer access to up-to-the-minute data, enabling researchers to create temporal patterns in specific places such as beaches urban green parks, or protected areas space utilization. Moreover, social media posts are captured and shared year-round, facilitating longitudinal analysis (Cui et al., 2021). Monitoring these changes can be quite challenging when using traditional data collection methods (Toivonen et al., 2019).

### *2.2.2 Content analysis*

Social media content analysis encompasses a range of qualitative and quantitative techniques designed to systematically examine the content posted by users on social media platforms. The content of the images itself, or tags can be used to provide more detail about the perspective of the producer (Pickering et al., 2020). Evaluating the contents of images or tags offers valuable insights into how individuals perceive and assign significance to locations, encompassing tourism and cultural aspects (Oteros-Rozas et al., 2017).

Identifying the content of the images and categorizing them is labor intensive but a nice way to map distinct categories, for example into diverse types of tourists such as wildlife, coastal, etc. (Slijkerman et al., 2020). Content analysis has gained much more attention with the rise of artificial intelligence. With deep learning techniques classifying the images is possible, this is done by detecting certain objects in the image and labeling these objects, for example, certain landscapes, animals but also emotions of persons in the image (Egarter Vigl et al., 2021; He et al., 2017).

#### *2.2.2.1 Visual and Textual Content Analysis*

It is possible to focus on the visual content and the textual content of social media (Shin et al., 2020). The visual content contains images but could also contain videos. This can be done manually by the researcher itself, however, analysis of such material, especially videos, can be time-consuming. The prevailing method of analyzing the content of the image involves manually scrutinizing the content of individual images to categorize them into groups, such as nature, wildlife, and heritage (Figueredo et al., 2017).

This categorization is determined by identifying the presence or absence of distinct elements in the images, such as natural scenery, historical landmarks, or an animal (Goldspiel et al., 2023; Tieskens et al., 2018). However, this categorization can be biased, and therefore some authors chose not to focus on these categorization options. Other researchers try to categorize the images on the intention of the photographer (Kaiser et al., 2020). Computer vision methodologies can speed up this progress and recognize certain species, emotions, and objects but with these identifications, it can classify the content (Araujo et al., 2020; Kumar et al., 2020; Zhang et al., 2020).

Focusing on the textual content covers the tags, titles, captions, hashtags, and other texts posted on social media. Keywords can be used to search for images with a specific topic, but also for categorizing the images, it can be useful to look at the tags or titles. It can also be used to select only the relevant images that will be used for the research (Mirzaalian & Halpenny, 2019; Obembe et al., 2021). This is where machine learning algorithms can help categorize the data (Lee et al., 2019; Menk et al., 2019). With Natural Language Processing (NLP) methods, the text can also be categorized in emotions, locations, organizations, etc. (Kim et al., 2022). NLP uses word embedding which is a word representation that computers use to understand language better. They learn from lots of other texts and help machines grasp the meaning and relationships between words in a way that makes more sense to them (Tejaswini et al., 2022). However, Language is personal to everybody, and therefore sometimes what is written on the internet can be interpreted differently (Khurana et al., 2023).

Most studies combine both visual and text content analysis, this is since they both enhance each other. When someone posts an image on social media a description, title, hashtags, location, or hashtags are given (Jailani et al., 2021). For example, an image of food in a restaurant with the caption "Yummy" and the name of the restaurant shows what the food is, how the person experienced it, and where exactly it was. This gives thus more comprehensive information about the social media user (Jindal & Aron, 2021).

However, both visual and textual content analysis also have their downsides. The availability of big volumes of data is necessary. In addition, the image can be misinterpreted by the computer vision methodology due to limited training datasets. Finally, when categorized by the researcher manually the research has a certain amount of biasedness (Morstatter & Liu, 2017).

### *2.2.3 Network analysis*

Focuses on the social network of relationships between users of social media. The nodes of this social network of social media are the users, images, or topics and they are linked by views, likes, comments, favorites, and followers of shares which are types of interactions (Vassey et al., 2023). Social network analysis is usually topic-oriented or topology-oriented. Topic-oriented social network analysis can provide valuable insights for various purposes, for example gaining a deeper understanding of the connections between users and sellers in the context of e-cigarette brands, and the discourse of anti-vaccine. A topology-oriented social network analysis focuses on for example the reach of an influencer or link clustering methods. A network of connections can yield valuable insights into patterns of communication and the spread of information (Gunaratne et al., 2019; Park et al., 2016).

*Table 2-1: Overview of the relevant literature*

| Author | Title | Research area | Geosocial data | Methodology | Findings |
|---|---|---|---|---|---|
| Figueredo et al. (2017) | Using Social Media Photos to Identify Tourism Preferences in Smart Tourism Destination | Natal, northeast of Brazil | Find a trip Platform with three thousand images | Convolutional Neural Network and fuzzy logic to classify tourists into five categories | 'A set of photos were analyzed and the results reached over 82% on accuracy metric for all classifiers'. |
| Garcia-palomares et al. (2015) | Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS | Eight major European cities | Images uploaded to Panoramio between 2005 and 2014 in the eight major European cities | Data aggregated by hexagons, for density maps and descriptive statistics, calculating standard distance to the mean center, Spatial autocorrelation was analyzed to identify the location and extent of spatial clusters of images. | 'The results indicated the concentration and dispersion of images in each city and their main hot spots and revealed marked differences between tourists' and residents' images since the former showed higher spatial concentrations' |
| Ghermandi et al. (2020) | Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico | Coastal region Mexico | 8.245 Flickr images between 2004 and 2017 | combine analysis of social media images and high-resolution land cover mapping to identify different cultural services and their association with specific ecosystem and land cover types | 'Locals are 2.2–2.5 times more likely than international visitors to be associated with aesthetic appreciation and birdwatching. Locals upload more images of coastal lagoons, mangroves, beaches, and sea' |
| Hale, (2018) | Mapping Potential Environmental Impacts from Tourists Using Data from Social Media: A Case Study in the Westfjords of Iceland | Westfjords region of Iceland | 10,172 Flickr images between 2014 and 2016 | Measures ecological sensitivity combined with geosocial data to assess the vulnerability of the locations frequented by foreign tourists | 'Tourists cluster primarily around six hotspots that represented some of the major known tourist destinations of the region. Although tourists generally frequented areas with lower ecological sensitivity and rarely went far beyond the main roads, one of the hotspots was in an area of higher ecological sensitivity. Further, tourists also appeared to have higher-intensity stays when they entered areas of higher ecological sensitivity |
| Jailani et al. (2021) | A Machine Learning Approach to Study Tourist Interests and Predict Tourism Demand on Bonaire Island from Social Media Data | Bonaire island | From 2003 to 2019, 13,706 geotagged Flickr data points assigned keywords and CBS (Centraal Bureau voor de Statistiek) data | TF-IDF (Term FrequencyInverse Document Frequency), DBSCAN (Density-Based Spatial Clustering of Noise Applications) | 'Tourism demand was forecasted using both Flickr and CBS data. The result considered a 'fine' result in this research was where the most relevant keywords and least relevant keywords show contradicting activities, which then define the unique activities in the area. The contrast between both sets for keywords reinforces the finding'. |
| Kim et al. (2021) | Coastal Tourism Spatial Planning at the Regional Unit: Identifying Coastal Tourism Hotspots Based on Social Media Data | South Korean regions | Data from Flickr and Twitter with 30" spatial resolution | Getis-Ord Gi used to derive the hotspots | 'Coastal Tourism Spatial Planning at the Regional Unit: Identifying Coastal Tourism Hotspots Based on Geosocial Data' |
| Lee & Tsou, (2018) | Mapping Spatiotemporal Tourist Behaviors and Hotspots Through Location-Based Photo-Sharing Service (Flickr) Data | Grand Canyon | 38,127 Flickr images between 2014 and 2015 | kernel density estimate (KDE) mapping, Exif (Exchangeable image file format) data, and dynamic time warping (DTW) methods. Different spatiotemporal movement patterns of tourists and popular points of interest (POIs) in the Grand Canyon area are identified and visualized in GIS maps | 'Winter tourists in the Grand Canyon explore fewer POIs compared to summer tourists based on their Flickr data. Tourists using high-end cameras are more active and explore more POIs than tourists using smartphone images. Weekend tourists are more likely to stay around the lodge area compared to weekday tourists who have visited more remote areas in the park, such as the north of Pima Point. These tourist activities and spatiotemporal patterns can be used for the improvement of national park facility management, regional tourism, and local transportation plans. |
| Pickering et al. (2020) | Using social media images and text to examine how tourists view and value the highest mountain in Australia | Highest mountain Australia | 498 Flickr images and text associated | Content analysis of images and semantic analysis of text | 'In Summer more visitors compared to the winter. In winter more recreation and in summer more biodiversity' |

| | | | | | |
|---|---|---|---|---|---|
| Runge et al. (2020) | Quantifying tourism booms and the increasing footprint in the Arctic with social media data | The Arctic | 800,000 Flickr images between 2006 and 2016 | created square spatial grids (rasters) at 10km and 100km resolutions. We then calculated the photo-unit-days in each grid cell for summer and winter aggregating data for each season | 'The footprint of summer tourism quadrupled and winter tourism increased by over 600% between 2006 and 2016, although large areas of the Arctic remain untouched by tourism'. |
| Schep et al. (2016) | Nature on Bonaire and social media behavior<br>Assessing the spatial distribution of tourism and<br>recreation on Bonaire through social media | Bonaire island | The total number of Panoramio unique users per kilometer uploads was 1478 images (between 2005 and 2016). Flickr photos from 2004-2016. 5,847 images from Instagram per km from 2014-2016. | Used also the classification of tourists based on the environment and thus the content of the image. Also compared the PUD with natural areas in Bonaire. | 'By far the largest share of images on social media are taken along the western side of the island. most tourists on Bonaire focus on marine-based activities such as diving, snorkeling, and surfing. More remarkable is that a very high percentage of images contained a terrestrial landscape rather than a coastal landscape. The results offer valuable input to support sustainable tourism development in Bonaire and have great potential to guide the management of human impacts on the natural environment of the island' |
| Slijkerman et al. (2020) | Tracking digital footprints in Bonaire's landscapes | Bonaire island | 13026 Flickr images within a bounding box around Bonaire | self-built Python application "PhotoCategoriser" to categorize the images. Metadata used of images. Resolution: 0.301 square km | 'Provides various figures and maps presenting the spatial distribution of PUD as a proxy for tourist distribution. Temporal aspects in PUDs reflect the annual dynamics in tourist numbers. PUDs are a proxy of tourist distribution, but not a strong indicator for trends in absolute numbers and intensity. |
| Yan et al. (2017) | Monitoring and Assessing Post-Disaster Tourism Recovery Using Geotagged Social Media Data | Philippines | 71.329 geotagged Flicker Images (from 1 April 2016 – 6 July 2016) | viewshed-based data quality enhancement, a space-time bin-based quantitative photo analysis, and a crowdsourcing-based qualitative photo analysis | 'Discovered spatiotemporal knowledge about the post-disaster tourism recovery, including the recovery statuses and trends mined from tourist images and images visually showing unfixed damage from both tourist and non-tourist images' |
| Zhang et al. (2020) | Mapping destination images and behavioral patterns from user generated images: a computer vision approach | Hong Kong | 58.392 Flickr images | artificial intelligence computer vision technologies to identify the differences in the perceived destination image and behavioral patterns between residents and tourists from user-generated images | 'The perceptual differences between the two groups lay on seven types of perceptions'. |

# 3 Methodology

In this chapter, the method used to carry out this research is explained. First, the study area of this research is briefly described. Secondly, the research steps will be explained. The first step is the investigation of the possible usage of geosocial data for this research is described. Which eventually will answer the third and fourth sub-questions '*What collecting methods and technologies are available for collecting and analyzing geosocial images in the context of tourism mapping?'* and *'What are the implications of using platforms such as Instagram, TripAdvisor, and Flickr for gathering geosocial data in the context of tourism mapping in Aruba?*. Sub-question 5 will only be addressed in this chapter namely, '*What are the various categories of tourists visiting Aruba?'*. After choosing the right social media platform the procedure to carry out the methodology with several steps is explained: data extraction, data preparation and classification, analysis of the data, and eventually the critical assessment of the data is explained.

## 3.1 Study area

Together with Bonaire, Curacao, Sint Maarten, Saba, and St. Eustatius, Aruba forms the Dutch Caribbean and is a member of the Kingdom of the Netherlands. It has around 106.000 inhabitants only 180 square kilometres and has a coastline of 69 km. Aruba is a Small Island Developing State (hereafter, SIDS), which means that the economy is dependent on foreign trade, and they have an open economy (Taylor, 2018). Figure 3-1 shows the island of Aruba, as can be seen, there are long sandbars along the southern coast of the island and the eastern side of the island is mostly covered by Arikok National Park.


*Figure 3-1: Aruba physiographic map and location (Schmutz et al., 2017)*

SIDS typically have few natural resources and are vulnerable to catastrophic weather conditions. They are thought to have difficult economic conditions in large part because of their vulnerable environment, location, and resource shortage (Petzold & Magnan, 2019). They are challenged by a lack of internal markets, institutional capacity, economies of scale, innovation, and competitiveness due to their small populations (Zhuawu et al., 2021). Their geographical location also adds to high import prices, including inputs for manufacturing, agriculture, and high transportation expenses. SIDS also have open economies which increases their vulnerability even more (Dagher, 2019).

### *3.1.1 High dependency on tourism*

Aruba and Bonaire were merely Curaçao's dependents throughout the first three centuries of Dutch colonization. The history of Aruba's economic development started in 1924. The oil refinery operated by the Lago Oil & Transport Company served as the primary economic driver for the island of Aruba. When the oil refinery closed in 1985, it led to a lot of unemployment (Croes & Vanegas, 2003). A new economic development started, and the focus shifted towards tourism. This led to a lot of jobs due to an increase in restaurants and hotels. Tourism has since then been most dominant in Aruba. However, having tourism as your primary source of income adds to the SIDS vulnerability (Taylor, 2018). Until now this dependency on tourism has hurt Aruba's economy three times, the first shock was in 2001 after 9/11, in 2008 was the second one with the financial crisis and the third was in 2019 during the coronavirus pandemic (Goede, 2021).

Aruba has undergone extensive development to accommodate the growing demand for tourism, including both overnight guests and cruise passengers. Growing international competitiveness, a small carrying capacity (as a microstate), and general sustainable growth are obstacles to Aruba's tourism business (Dipietro & Peterson, 2017). As can be seen in Figure 3-2 the numbers are decreasing from 2019-2023, this can be explained by the corona-pandemic. The numbers for 2023 are expected to be on the level of 2019.

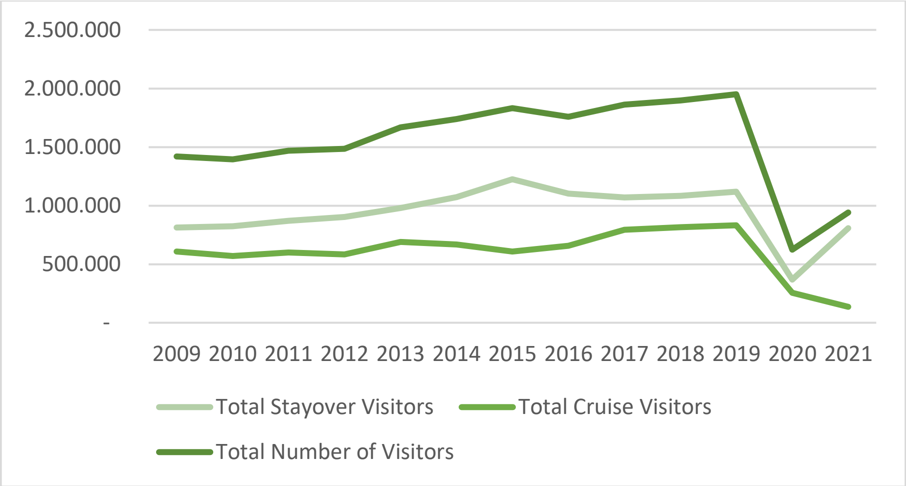



*Figure 3-2: Total number of tourists in Aruba 2009-2021 (CBS, 2021)*

## 3.2 Research Steps

The rest of the methodology chapter focuses on the research steps. First, a schematic overview of the steps is displayed (Figure 3-3), which shows the steps that need to be taken in this research. In total five steps will lead to realizing the research objectives as stated in Chapter 1. The first step focuses on the wide variety of possibilities to extract geosocial data from the internet. Important here is that the location of the images can be extracted. After a decision is made for a platform, data extraction will commence using APIs (Application Programming Interface) and collecting the relevant metadata. Thereafter follows the data preparation and classification. In this step, the data is structured in a way that the tourism of Aruba can be mapped per tourist category and graphs can be created. This is done by leveraging the programming languages Python and R. In addition, a 'PhotoCategorizer' will be used to group the images into several types of tourists, which uses a SQLite database. When the images are categorized, a c-square grid is created to display the frequency of each type of tourist in Aruba. This will lead to analyzing the data, meaning explaining the graphs and maps. Thereafter a critical assessment of results is conducted using surveys sent to tourist agencies in Aruba. Eventually, conclusions can be drawn, and meaningful knowledge is created.

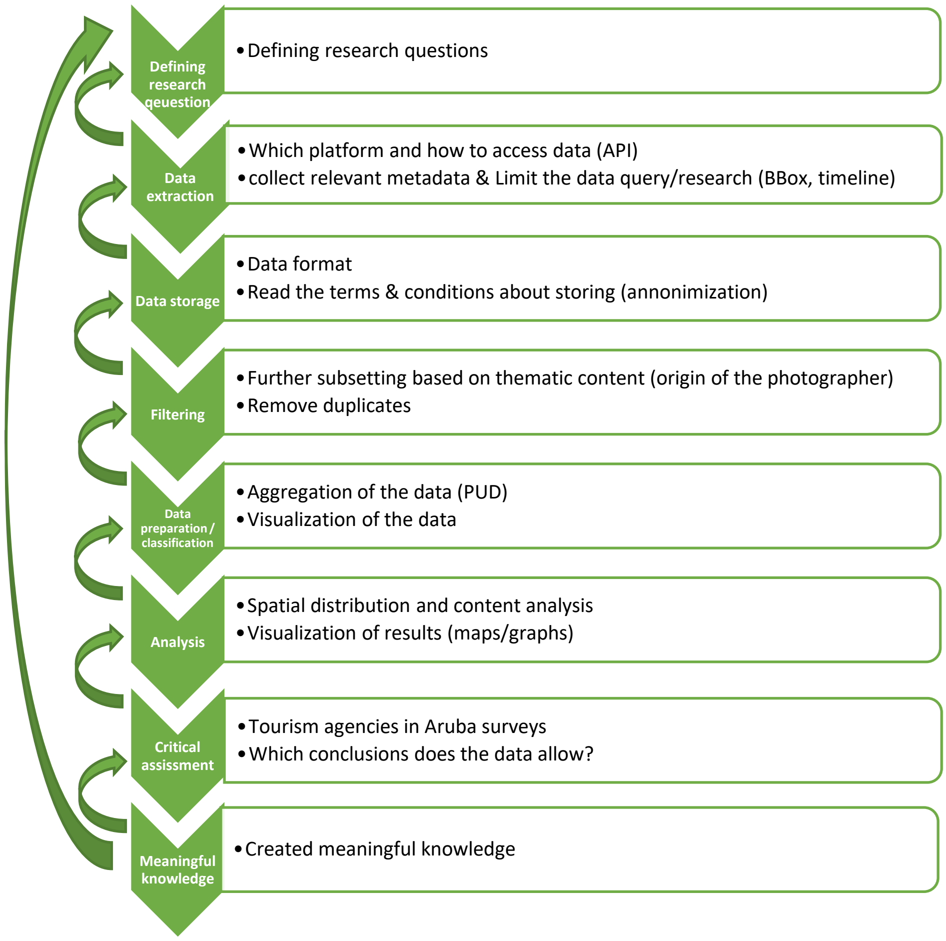



*Figure 3-3: Schematic view of the research steps (Toivonen et al., 2019)*

To derive valuable insights from geosocial data an iterative procedure is needed. According to Toivonen et al. (2019) within this procedure, several stages should be checked (Figure 3-3). The process is as follows, first data extraction, then data storage, filtering the data, data preparation, and classification, which then leads to the actual analysis. When the analysis is done, a critical assessment takes place where the questions arise: can I compare the results with 'Ground Truth', and what conclusions can be drawn from the data? These steps will be followed iteratively, refining and adjusting at each stage to enhance the insights gleaned from the data.

With the help of this process enormous amounts of data, which exhibit considerable size and complexity are transformed to create meaningful knowledge. In the context of data quality, the substantial volume of geosocial data can effectively counteract challenges related to sample size limitations and sampling bias (Lin et al., 2021). Transforming the data to only extract the pertinent data for further analysis is therefore crucial, whereby the data is filtered and cleaned to make sure the representative data will be taken into the analysis (Dash et al., 2019). Being aware of the state of the art and the challenges are displayed below.

## 3.3 Data extraction

According to Vu et al. (2019), privacy statements have been more tightly defined for social media platforms, therefore the first step is to investigate what the respective terms of service are, followed by the ethical considerations and the collection methods. Then it is critical to determine which platforms still provide the metadata required for this research and take other new developments into account, such as API access.

### *3.3.1 Social media Terms of Service and Ethical use*

According to Da Mota & Pickering (2020), who looked at current research trends that use geosocial data for tourism, only 25% of the papers evaluated made some form of reference to ethical concerns, with the majority providing only brief mentions. These include mentioning that access was limited due to privacy, mentioning the importance of privacy, or that there were privacy policy rules they needed to follow.

Privacy of users on social media is gaining much more attention, this has led to several restrictions with the automated gathering of data from these platforms. The restrictions are described in the Terms of Service (hereafter, TOS). TOS are agreed upon between the user and the social medium upon which the user is active (Hidalgo, 2021). Usually, the TOS is accepted by the user before the platform can be (actively) used and interacted with. Depending on the platform, the TOS can contain provisions in which the user also agrees with the terms of using their data for external agencies (Stasi. 2019). The possibilities of web scraping and API access with the platform and the re-use of content uploaded on social media are described in the TOS of the platform (Marapengopi, 2017).

Social media research should always be done with caution because every study of social media will involve human-produced content, content that might contain sensitive or personally identifiable information. There are large volumes of data used within social media research, thus this data source has a large volume and a substantive reach, which in turn makes it harder to control the ethics of research using SMA. With the emergence of computational network analysis, text mining, and deep learning techniques, the nature of social media has also evolved. Platforms seem to be moving away from text-only and shifting towards multimedia (Bruns, 2019).

Social media users are mostly not aware that their data on the internet is used in research. When you want to use personal data as compared to other sorts of data gathering, informed consent is needed. However, for social media informed consent is not regulated. There are several ways in which the researcher can let the user know that their data is used, for example emailing all users or contacting them via the platform. When using information from enormous amounts of users this can be quite hard and challenging. The fact is that using detailed information about users is something to be aware of and the consequences should be considered (Di Minin et al., 2021).

### *3.3.2 Collection of geosocial data*

There are several ways of extracting data from social media, the two main ways are manual collection, and automated collection (Dewi & Chandra, 2019). Manual collecting can be done by making screenshots of images, collecting videos, or Twitter posts on public accounts. Without violating the TOS, you are free to manually collect information regarding content on social media. The same kinds of guidelines apply while observing individuals in the actual world. This way of collecting data is not against the TOS. However, when collecting data from private accounts it is good to read the TOS from that social media platform (Chaudhary et al., 2019).

Automated collection is also possible, the most used method here is API, although web scraping is also used a lot. Web scraping is a technique that uses software to extract, store, and analyze information from websites. It can be done with automated processes using a bot, script, or web crawler in which mostly an ‘invisible’ link is used on a social media platform. Web scraping is usually used when API access is not possible. Therefore, in most cases, web scraping is against the TOS and not supported by the social media platform (Dewi & Chandra, 2019).

#### 3.3.2.1 Application Programming Interface

Because manual collection will take a lot of time and web scraping mostly is against the TOS, API is the method applied in this research. Access to a sizeable portion of geosocial data on various platforms is regulated through the API, which is implemented by platforms to define the guidelines that all third-party entities must adhere to when accessing the data (Bruns, 2019). Through the API data can be downloaded, in the same way as the platform handles the data, providing access to a huge amount of detailed information about the users who post certain information publicly (Viñán-Ludeña & Campos, 2022). API documentation offers in-depth information about the functionality of these APIs and provides details on how to retrieve data (Robinson et al., 2016).

APIs enable third-party entities to access for instance friendship connections in a controlled and programmatic manner, allowing them to interact with the recorded connections between registered users. Access to publicly available web data may sometimes require login credentials, which are used to enforce download limits and oversee data access. To achieve this, social media platforms have implemented authentication procedures using keys and secrets. While rate limits are essential to prevent data abuse and enhance security. API rate limiting can be achieved through methods such as request queues, API throttling, and the application of rate limit algorithms (Ali & Zafar, 2022).

Each platform has its API and the limitation to reach certain detailed information varies per platform. Also, the format in which the data can be extracted differs per social media platform (Tenkanen et al., 2017). Some platforms offer APIs-specific access for "academic research" that extends beyond the typical call restrictions (Acker & Kreisberg, 2020).

### 3.3.3 Social media platforms

Each social media platform has its type of users, tools, techniques, relevance, or authorization to decide if the platform is suited for the research. The options displayed in Table 3-1 below mainly focus on images because even though the options for text or hashtags were also possible, including them would make it increasingly complex for tourism to look at tags due to the language barriers and preferences for certain choice of words. Research can miss a lot of information when choosing one language whereas the tourists would post something in another language (Chen et al., 2022). In comparison, the TRUPIAL research that has been conducted on Bonaire was also carried out using images.

In Table 3-1, an overview can be found that describes what the best platform was to use for this research. Combined research on the visual content found in tourist-generated social media images primarily concentrated on platforms such as Flickr and Instagram, with additional scrutiny directed towards Pinterest and TripAdvisor. (McMullen, 2024). The availability of their API is explained, per platform, the amount and type of users are explained and eventually, the relevance of the platform itself is added. This gives an insight into why people use this platform.

*Table 3-1: Options and details on why to use various geosocial data sources*

| Source | Availability | Users | Relevance |
|---|---|---|---|
| Instagram | Instagram has tight restrictions on its API. Also, their TOS mentions a ban on scraping collecting legal data from Instagram automatically is extremely hard. With the "graph API", it is possible to search for hashtags, however, the results will only be from business accounts and not personal accounts (Igartua et al., 2020). | 1.5 billion users currently, People between 25 and 34 are 33 percent of the users and people between 18 and 24 are 29.8 percent of | Instagram is widely used (as can be seen in the number of users), and young people use Instagram. However, people show their most beautiful experiences on the platform which only |

| | | | |
|---|---|---|---|
| | | the users (Güven et al., 2023). | shows their beauty (Manovich, 2019). |
| TripAdvisor | TripAdvisor’s API costs money but this also not gives full access. It gives five images and reviews per location max. Also for academic research, it is not allowed to access (Lu et al., 2016). | TripAdvisor has reached 1 billion users (Batabyal et al., 2023). | Platforms are used by frequent travelers to express their opinions due to the review function of the platform (Karaceper & Ramadanoglu, (2023). |
| Flickr | Flickr’s API is used most often in research and also provides a lot of different methods and options to gather several metadata (Wang et al. 2023). | 122 million users from 63 countries (Pereira, 2023) | The platform is used the most in scientific research. People are less picky in what they post due to the personal photo archive Flickr has. A lot of nature images are posted on Flickr (Wong et al., 2017). |
| Pinterest | Pinterest does provide an API for developers to access. However, access to the API may not be entirely free, and there may be restrictions or usage limits depending on the specific features and data you need. Pinterest's API allows developers to retrieve information such as user data, pins, and boards. However, it does not allow for downloading images and semantic information (Axelsson et al., 2020) | 463 million active users. In total, 60% of Pinterest's user base falls within the age range of 18–34, with females comprising 80% (Kemp, 2023). | Pinterest is a way of gathering inspiration and ideas. Users thus document their travel aspirations via Pinterest. This exemplifies the emotional effort users invest in curating their collections of wanderlust (Gretzel, 2021). |

Instagram has sharpened its rules for using the API and only gathering hashtags for business accounts will not map the tourism in Aruba. The fact that it is not allowed to use the Tripadvisor API for scientific research has led to the choice to not use this platform (Ricardo et al., 2022). The Pinterest API is easily accessible, however, the content of the images is not always their images, and mainly people's dreams or ideas to go visit that are posted on their pins or boards (Gretzel, 2021). This can impact the results since people pin a certain attraction of Aruba, which is not visited by that person. Downloading images is not allowed according to the API and therefore this platform is left out (Axelsson et al., 2020). Flickr has been used a lot recently for several studies, in addition, or maybe by default, it is one of the most used social media platforms for tourism research see Figure 3-4 (Da Mota & Pickering, 2020). One-third of Flickr users geotag their images, voluntarily choosing to share the location where the image was taken. Which is great for mapping tourism. Flickr also allows for building personal photo archives and sharing images with friends. This archiving is handy for researchers because it gives a more comprehensive record of the destinations visited by the users (Wong et al., 2017).

Slijkerman et al. (2020) carried out a similar study for Bonaire as part of the aforementioned TRUPIAL program, in which Flickr was also used. This research aims to transfer the method applied for Bonaire to Aruba. That is another reason the choice of Flickr is the most compelling one.

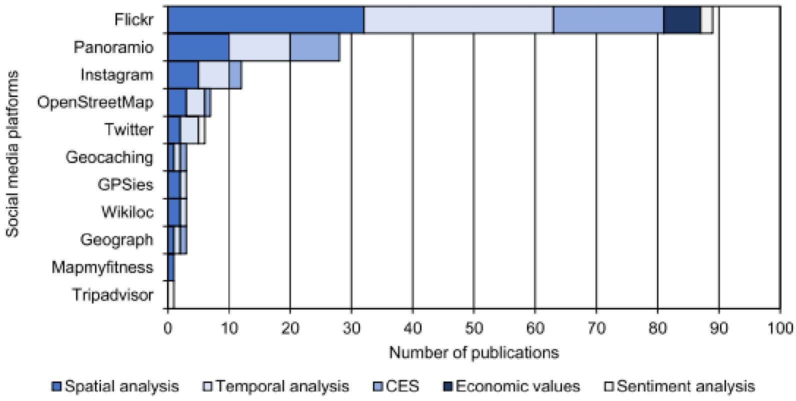



*Figure 3-4: Types of data analysis for different social media platforms (Da Mota & Pickering, 2020)*

Besides, Flickr is the most used platform to research TDI, this is due to its academic-friendly API (Lim, 2015). As can be seen in Figure 3-4, where research by Da Mota and Pickering (2020) compared several methods for mapping tourism uncovering that most methods used Flickr. In comparison to for instance, Instagram and TripAdvisor the amount of users on Flickr is not that high; however, the APIs of Instagram and Pinterest are not academic friendly and therefore also not used that often in research (Deng et al., 2019). For example, Instagram only allows images to be extracted when pointing to a specific location, but the location of an image is not given, and owner consent is needed (Wong et al., 2017). So, for this study in which bounding-box scope is determined, collecting data from this platform is not feasible. Since the API does not allow a bounding box methodology.

While the number of users on Flickr is decreasing there were still 25 million images uploaded in 2022 and the platform has 122 million registered users of which 60 million are active. An active user is someone who uses Flickr once a month (Broz, 2022). Research findings have highlighted Flickr's distinctive character, as it outshines other platforms by hosting a wealth of nature-centric images. The platform users have the preferences of tourists, including their fascination with landscapes and biodiversity found in parks and natural environments (Pickering et al., 2020). This is also a good reason why choosing Flickr for this study because the classification of the images is focused on the environment within the image.

### *3.3.4 Relevant metadata & limit the data query/research*

Social media has several options and valuable information that can be extracted (Figure 3-5). It is therefore good to decide what information is needed for the research and if that information can be extracted. You can get information about the user, the content, or comments and likes by other users (Poorthuis, et al., 2016).


*Figure 3-5: Social media metadata attributes*

### *3.3.4.1 Relevant metadata needed*

Flickr has a detailed explanation of exactly how and what you can cite using the Flickr API. In the beginning, the focus was laid on what information was needed for this study. Table 3-2 shows the metadata needed to extract the information such as the geotag of the image, the place of residence of the user, but also the URL of the image to display the image while categorizing.

*Table 3-2: Metadata extracted from Flickr*

| Metadata attribute | Description |
|---|---|
| **Id** | A unique identifier of the image |
| **Owner** | A unique identifier for the user that posted the image |
| **Secret** | A Unique key to get additional metadata of an image |
| **Server** | The server number where the image is stored |
| **Dateuploaded** | The date on which the image was uploaded on Flickr |
| **Description** | Description of the image |
| **Views** | The number of views the image has |
| **Comments** | The number of comments the image has |
| **Tags** | Set of labels that describe the content of the image |
| **Title** | Title of the image |
| **Ispublic** | Public images |
| **url_o, url_m, url_c** | URL to the image and profile of the photographer |
| **Latitude** | Latitude coordinate of image location (EPSG:4326) |
| **Longitude** | Longitude coordinate of image location (EPSG:4326) |
| **Accuracy** | Accuracy of the image location (World level is 1, Country is ~3, Region is ~6, City is ~11, Street is ~16) |
| **Context** | Geo context is a numeric value representing the image geotag beyond latitude and longitude (0 not defined, 1 indoors, 2 outdoors) |
| **Locality** | What Place of residence Flickr user has filled in |
| **Woied** | Where On Earth Identifier which is a unique 32-bit reference identifier |
| **Neighbourhood** | What neighborhood Flickr user has filled in |
| **Media** | Media type (videos or images) |

### *3.3.5 Limit the data query/research*

A GET-request was used to find the metadata, and thus images uploaded, within a bounding box around Aruba. The box ranges from -70.28 and -69.64 longitude and 12.17 and 12.81 latitude. The years that the data was collected ranged between 2012 and 2022. This period was chosen because entire years (1 January to 30 December) were only included and this was the most recent entire year when starting this research. A large period of ten years was possible due to the usage of geosocial data.

### *3.3.6 Flickr geosocial data extraction process*

In Figure 3-6, a flowchart can be seen that describes the process that was followed to extract and store the data.

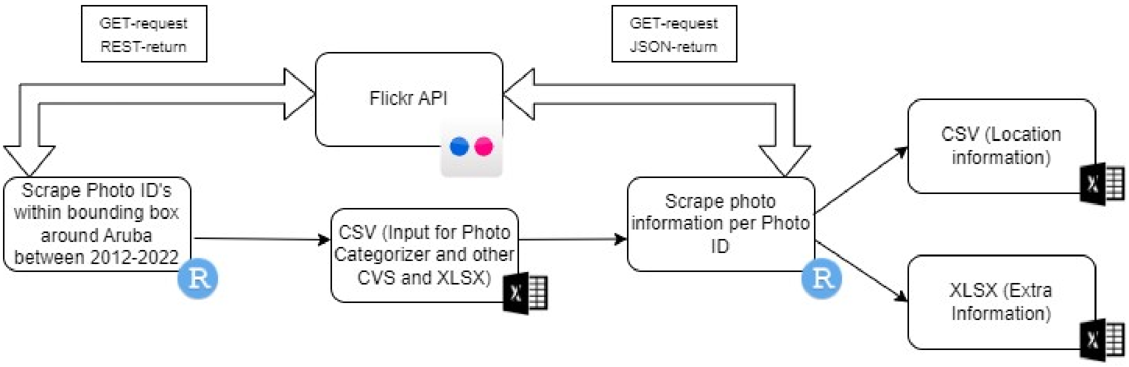



*Figure 3-6: Flowchart of data extraction process*

To utilize the Flickr API, an email and a Flickr account were created to obtain access to Flickr and the API keys. A key and secret were requested to extract images from Flickr. The R script, also used for the Slijkerman et al. (2020) article made by Pepijn de Vries, was used as inspiration. The script itself was outdated and the packages were no longer useful, but it did give the researcher more insight into how to use R studio which information was necessary to collect and in which format the data should be gathered. The Flickr API Flickr.photos.search was used to gather the images. Flickr itself also has one limitation, being that the amount of information per request is limited and therefore the data that will be extracted per year. After finding out which metadata was needed and taking the data query/research limits into account, in total 19.189 images were found within the bounding box around Aruba between 2012-2022.

A second R script was made to collect the other extra relevant information about the location, user, and content itself. The question that emerged was how the exact location of the images could be retrieved. The Flickr API Flickr.photos.geo.getLocation was used, however it could only be collected per photo ID. Therefore, the R script included a loop to insert 19.189 photo IDs in the Get-URL request. This resulted in a separate CSV file with the locations of each photo ID, which later could be correlated. Another Flickr API namely flickr.photos.getInfo was also only usable per photo ID. Therefore it was used to extract other extra information. This last code resulted in the extra information of the images such as origin being collected in the xlsx-format.

## 3.4 Data storage

As seen in Figure 3-6, the data resulted in a CSV file and an xlsx file. These were stored on a local computer which only the researcher could access. The data was also stored in an SQLite 3 database, which was also stored on the local computer. When categorizing the images, a small image was stored in a folder on the local computer. In case of a repetition of the research (to check for, for instance, biasedness) this can be done by requesting access to the researcher.

## 3.5 Data preparation and classification

### 3.5.1 *Categorizing Flickr Aruba Images*

To make the categories for the tourists in Aruba current literature was explored. When looking at the literature the type of tourists that visit Aruba versus Bonaire were quite similar. They both offer the same branding strategies (Straathof, 2019). Therefore, the categories used by Schep et al. (2016) and Slijkerman et al. (2020) for Bonaire will also be used in this research.

The difference between tourism in Aruba and Bonaire is that more tourists that visit Aruba go there to party and in Bonaire more people go diving or visiting nature. Also, Aruba is known for the big all-inclusive resorts that lie at the beach. The absolute number of tourists that visit Aruba is higher compared to Bonaire (MacDonald, 2020). Therefore a category called party/all-inclusive was added compared to Slijkerman et al. (2020). Peter Verweij was also requested by Peter Verweij, for whom this research is conducted. The categories that are used are:

1. Coastal: images of the coastline (from the mainland)
2. Party/all-inclusive: all-inclusive hotels and people partying
3. Seascapes: scenic images of the seascape (above water) and water sports (taken from the sea)
4. Landscape: images of terrestrial natural landscapes (excluding terrestrial wildlife)
5. Wildlife: images of terrestrial wildlife and birds and marine wildlife
6. Other: mainly indoor and urban images

Each image has been assigned to one of these six categories and eventually, for each of these six categories, a unique map was created. These categories were chosen because Aruba and Bonaire both possess a common ancestry that includes the impact of African slaves, European Spaniards, Amerindians from continental South America, and the Dutch. They are both focused on tourism and they compete with each other in the cruise tourism industry (Schmutz et al., 2017). In addition, both islands are relatively small, with the focus on sun, sea, and sand as the main selling point for tourism (Ridderstraat, 2022).

### 3.5.2 *Different categorizing techniques*

To categorize the images there are several methods to do so. As explained in paragraph 2.3.2 the visual content analysis, in the case of images, could be analyzed as well. This could be done manually by the researcher or by using computer vision methodologies (Zhang et al., 2020). Initially, an examination was conducted to assess the potential categorization of images through computer vision techniques.

One way of using computer vision techniques is content-based image retrieval (hereafter, CBIR), here the images are categorized and rely on low-level content features such as color, pattern, and shape (Latif et al., 2019). However, this technique categorizes the images in a way that mainly does not suit the researcher's wishes. There is a notable disparity between the low-level content features and the semantic concepts employed by humans for image interpretation. Moreover, it proves impractical for general users, given the necessity for users to furnish query images (Zhang et al., 2012). As this research already has predefined categories this would not be the best method to use.

There is also an Automatic Image Annotation (Hereafter, AIA) method. AIA methods concentrate on creating models and algorithms to label images by their semantic content. They can also evaluate the similarity between a he amount of images and labelled with semantic content, with high efficiency and decreased subjectivity compared to the manual and CBIR methods. These approaches forecast pertinent labels for untagged images through a weak-supervision method or fully automated processes (Cheng et al., 2018). Here the images could be categorized into the categories that are also used by Slijkerman et al (2020) However, these would need to be predefined and that is the challenge with applying AIA. For example the challenge of variations in capturing the same object from different angles, distances, or under varying lighting conditions. Two real-world objects with the same name may exhibit different visual characteristics, such as shape and color. Describing image content automatically using labels is subjective (Bahrololoum et al., 2017).

In conclusion, there are several techniques to categorize the images, manual, CBIR, and AIA. For this research, the manual method was chosen. Even though this is a time-consuming method, the researcher is more aware of what images are classified and gets to see all the data. Also, the automated image annotation approaches are expensive (Sager et al., 2021). Manual annotations of images could be done in two ways: tagging and browsing (Figure 3-7). With tagging one image gets a set of 'tags' based on the image itself. Browsing the images means that you look at the images and based on that a group of images gets a 'tag' that is pre-defined (Fruchard et al., 2023). In this research, the manual Browsing method was chosen due to the fact the types of tourists can easily be grouped manually by the researcher. In the next section, more explanation about the pre-defined tourist categories is explained.


*Figure 3-7: Manually categorizing images, Tagging versus Browsing (Fruchard et al., 2023)*

### 3.5.3 Convert the data to categorize the images

For the data classification and preparation, a second flowchart was made to show the steps taken (Figure 3-8).

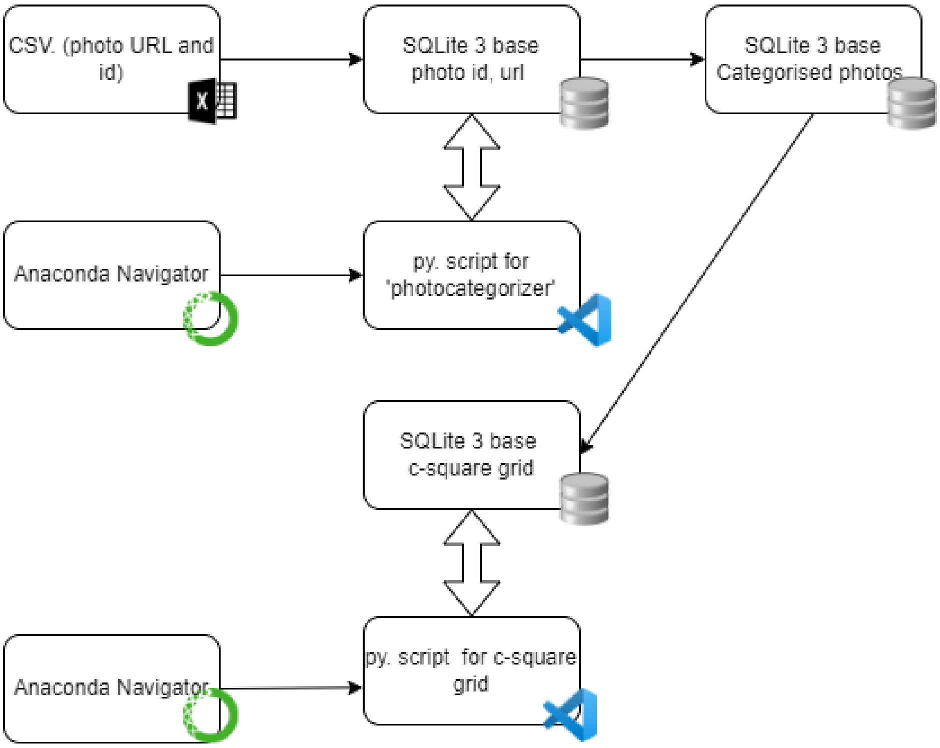
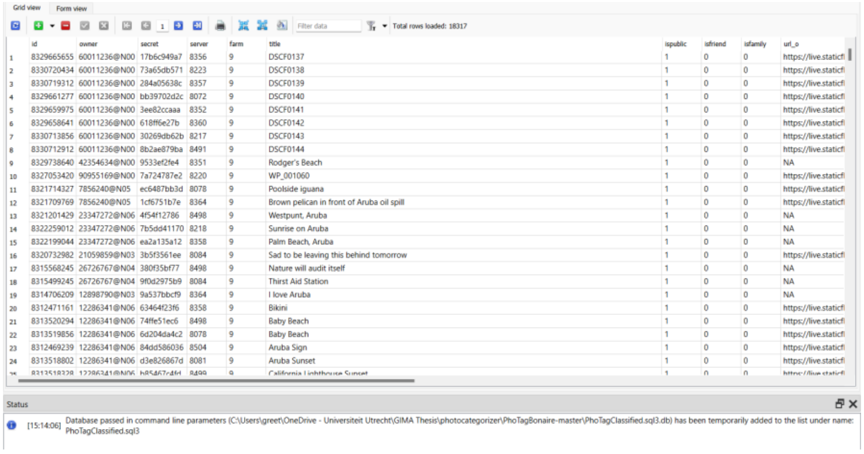



*Figure 3-8: Flowchart of data Preparation*

The data was converted in R to a CSV file where all the information was structured and could be interpreted by the next useful product, which was also used in the Slijkerman et al. (2020) research namely, the 'PhotoCategorizer' made by Jan Tjalling van der Wal.

However, before using the ‘PhotoCategorizer’, there were several things that had to be done. First, an SQLite 3 database was constructed where the CSV file was integrated. When uploading the CSV file an error occurred that several photo IDs were duplicated. Therefore the duplicated images, were identified based on their photo ID and subsequently removed from the dataset, eventually 18.300 images were categorized and were taken into account in the analysis. This SQLite 3 database was used to read the images’ URLs and show the image in the ‘ PhotoCategorizer’ (Figure 3-9)



Grid view Form view

Total rows loaded: 18317

| | id | owner | secret | server | farm | title | ispublic | isfriend | isfamily | url_o |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 8329665655 | 60011236@N00 | 17b6c949a7 | 8356 | 9 | DSCF0137 | 1 | 0 | 0 | https://live.staticfl |
| 2 | 8330720434 | 60011236@N00 | 73a65db571 | 8223 | 9 | DSCF0138 | 1 | 0 | 0 | https://live.staticfl |
| 3 | 8330719312 | 60011236@N00 | 284a05638c | 8357 | 9 | DSCF0139 | 1 | 0 | 0 | https://live.staticfl |
| 4 | 8329661277 | 60011236@N00 | bb39702d2c | 8072 | 9 | DSCF0140 | 1 | 0 | 0 | https://live.staticfl |
| 5 | 8329659975 | 60011236@N00 | 3ee82ccaaa | 8352 | 9 | DSCF0141 | 1 | 0 | 0 | https://live.staticfl |
| 6 | 8329658641 | 60011236@N00 | 618ff6e27b | 8360 | 9 | DSCF0142 | 1 | 0 | 0 | https://live.staticfl |
| 7 | 8330713856 | 60011236@N00 | 30269db62b | 8217 | 9 | DSCF0143 | 1 | 0 | 0 | https://live.staticfl |
| 8 | 8330712912 | 60011236@N00 | 8b2ae879ba | 8491 | 9 | DSCF0144 | 1 | 0 | 0 | https://live.staticfl |
| 9 | 8329738640 | 42354634@N00 | 9533ef2fe4 | 8351 | 9 | Rodger's Beach | 1 | 0 | 0 | NA |
| 10 | 8327053420 | 90955169@N00 | 7a724787e2 | 8220 | 9 | WP_001060 | 1 | 0 | 0 | https://live.staticfl |
| 11 | 8321714327 | 7856240@N05 | ec6487bb3d | 8078 | 9 | Poolside iguana | 1 | 0 | 0 | https://live.staticfl |
| 12 | 8321709769 | 7856240@N05 | 1cf6751b7e | 8364 | 9 | Brown pelican in front of Aruba oil spill | 1 | 0 | 0 | https://live.staticfl |
| 13 | 8321201429 | 23347272@N06 | 4f54f12786 | 8498 | 9 | Westpunt, Aruba | 1 | 0 | 0 | NA |
| 14 | 8322259012 | 23347272@N06 | 7b5dd41170 | 8218 | 9 | Sunrise on Aruba | 1 | 0 | 0 | NA |
| 15 | 8322199044 | 23347272@N06 | ea2a135a12 | 8358 | 9 | Palm Beach, Aruba | 1 | 0 | 0 | NA |
| 16 | 8320732982 | 21059859@N03 | 3b5f3561ee | 8084 | 9 | Sad to be leaving this behind tomorrow | 1 | 0 | 0 | https://live.staticfl |
| 17 | 8315568245 | 26726767@N04 | 380f35bf77 | 8498 | 9 | Nature will audit itself | 1 | 0 | 0 | NA |
| 18 | 8315499245 | 26726767@N04 | 9f0d2975b9 | 8084 | 9 | Thirst Aid Station | 1 | 0 | 0 | NA |
| 19 | 8314706209 | 12898790@N03 | 9a537bbcf9 | 8364 | 9 | I love Aruba | 1 | 0 | 0 | NA |
| 20 | 8312471161 | 12286341@N06 | 63464f23f6 | 8358 | 9 | Bikini | 1 | 0 | 0 | https://live.staticfl |
| 21 | 8313520294 | 12286341@N06 | 74ffe51ec6 | 8498 | 9 | Baby Beach | 1 | 0 | 0 | https://live.staticfl |
| 22 | 8313519856 | 12286341@N06 | 6d204da4c2 | 8078 | 9 | Baby Beach | 1 | 0 | 0 | https://live.staticfl |
| 23 | 8312469239 | 12286341@N06 | 84dd586036 | 8504 | 9 | Aruba Sign | 1 | 0 | 0 | https://live.staticfl |
| 24 | 8313518802 | 12286341@N06 | d3e826867d | 8081 | 9 | Aruba Sunset | 1 | 0 | 0 | https://live.staticfl |

Status

[15:14:06] Database passed in command line parameters (C:\Users\greet\OneDrive - Universiteit Utrecht\GIMA Thesis\photocategorizer\PhoTagBonaire-master\PhoTagClassified.sql3.db) has been temporarily added to the list under name: PhoTagClassified.sql3

*Figure 3-9: SQLite 3 base for 'PhotoCategorizer'*

With the ‘PhotoCategorizer’ the images posted on Flickr were displayed as a copy accompanied with several types of tourist categories used in the research on the side. The ‘PhotoCategorizer’ was used to assign images to six categories of tourists. It used the SQLite3-database to store the categorized images. The researcher had to choose to which category the image belonged (Figure 3-10). In the research of Slijkerman et al. (2020), the researchers stated that it took about two working days to categorize all the images. However, after speaking to Jan Tjalling van der Wal, he explained that the categorization was carried out by two people at the same time which ensured that it took only two days. Important to mention is, that this research was carried out by one researcher and therefore took around 8 full days of categorizing. Using the ‘PhotoCategorizer’ had the advantage that the image's qualitative aspect and thus the content of the image itself were reviewed by the researcher. Because of this, every image was inspected, and no images were included that did not fit into this research.

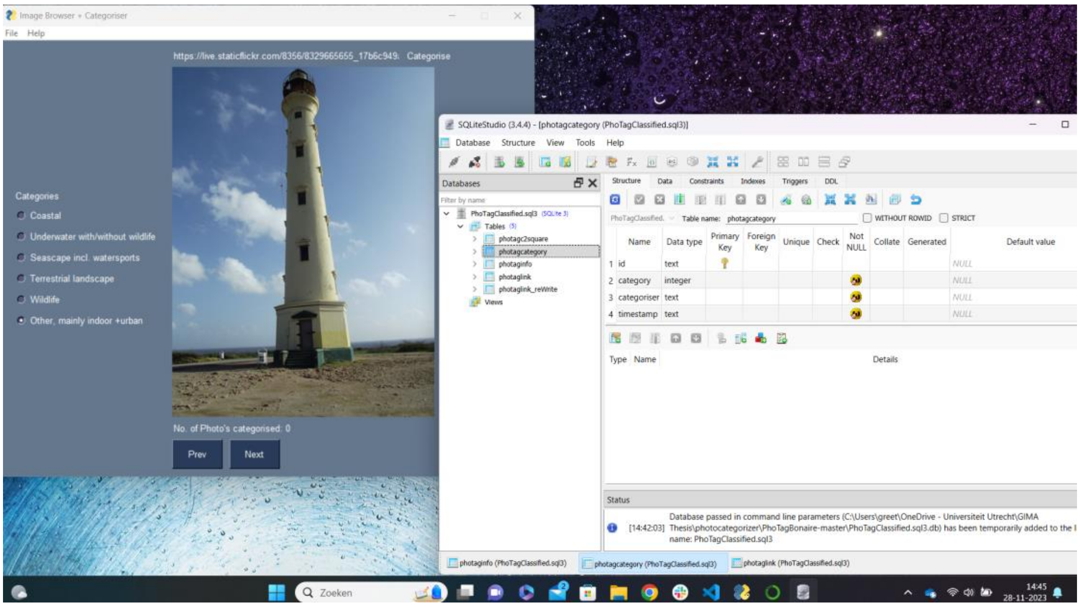



*Figure 3-10: The ‘PhotoCategorizer’*

Below the phase of categorizing the images is displayed. As you can see you press the category and then press next (Figure 3-11). A second image will appear and then you decide which category matches the image.  In the Python output, you can see if everything went well and you see under which category the image was stored.

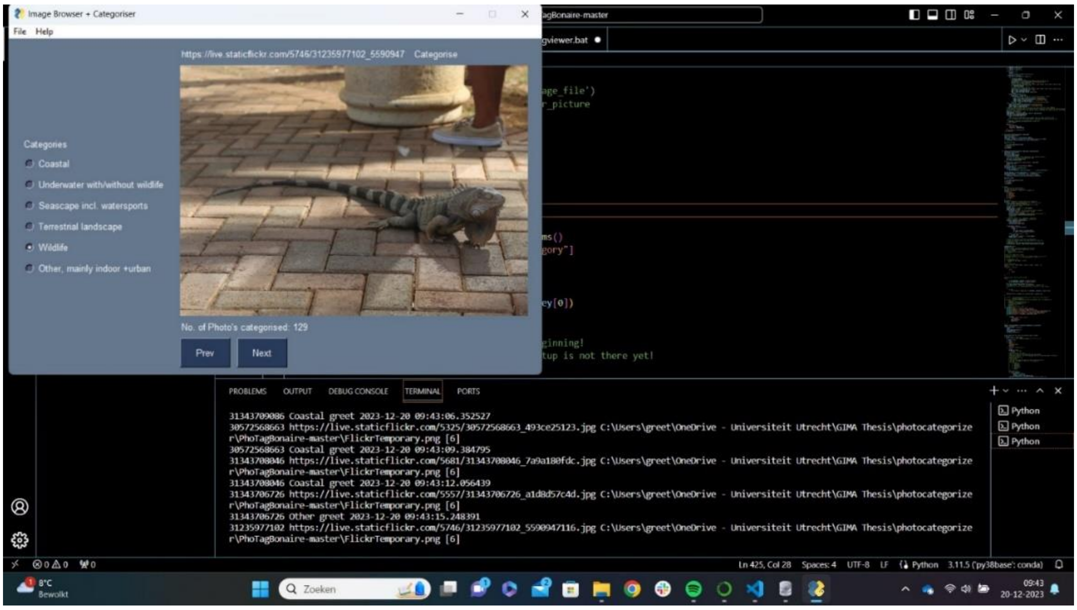



*Figure 3-11: Categorizing the images*

Figure 3-12 shows the output in the database, which shows the photo ID, the category, the person who categorized the images, and the time. With that time it was possible to see how long categorizing took and in total, this took about 30,5 hours more or less (Table 3-3). This is a time-intensive way of categorizing the images, however, the researcher knows what the content of the images is. However, compared to conducting surveys or interviews with 18.300 tourists that visited Aruba it has saved time.

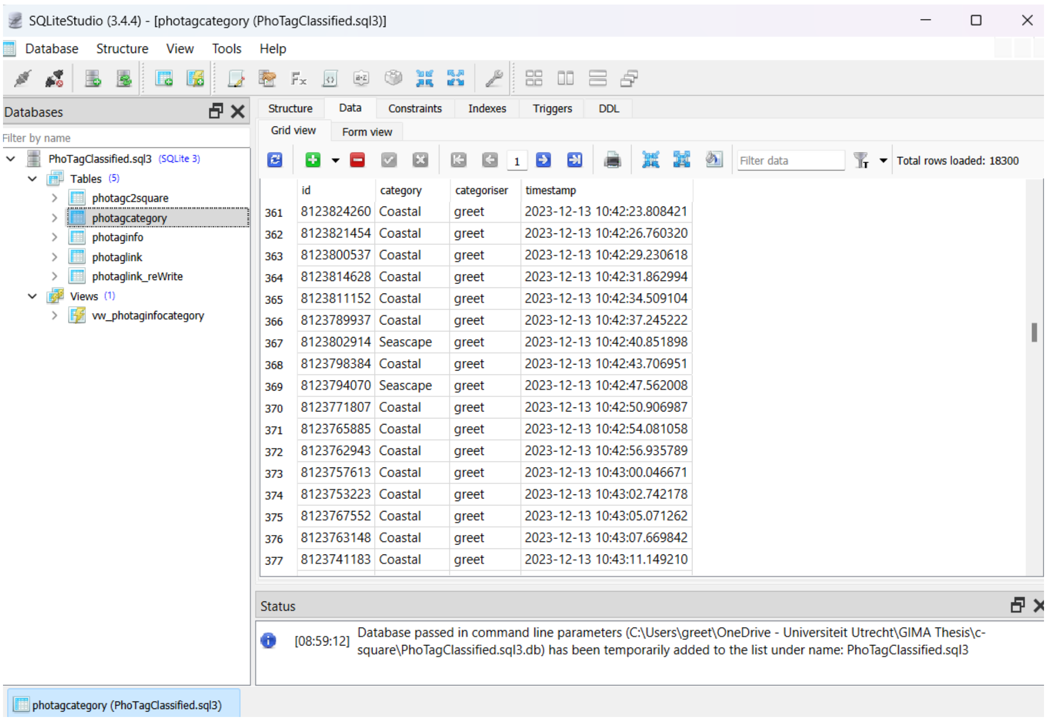



*Figure 3-12: Outcome of the categorized images*

*Table 3-3: Time spent categorizing the images*

| Date and time | Hours spent on categorizing | Number of images categorized |
|---|---|---|
| **2023-12-13** 09:58 – 11:37 | 1,5 Hours ± | 1.274 |
| **2023-12-15** 10:52 – 11:53 / 15:59 – 16:50 | 2 Hours ± | 1.197 |
| **2023-12-16** 11:02 – 13:02 | 2 Hours ± | 1.376 |
| **2023-12-18** 12:01 – 18:24 | 6,5 Hours ± | 1.601 |
| **2023-12-19** 09:43 – 12:07 | 2,5 Hour ± | 1.462 |
| **2023-12-20** 08:17 – 10:39 / 22:11 – 22:58 | 3 Hours ± | 2.370 |
| **2023-12-22** 09:30 – 12:39 | 3 Hours ± | 1.908 |
| **2023-12-24** 18:20 – 22:00 | 2 Hours ± | 2.154 |
| **2023-12-26** 13:24-13:45 / 21:29 – 23:16 | 2,5 Hours ± | 1.998 |
| **2024-01-03** 09:00 –12:37 / 14:33 – 15:35 / 18:47 – 19:51 | 5,5 Hours ± | 2.960 |
| **Total** | 30,5 Hours ± | 18.300 |

### *3.5.4 Converting Flickr images on the map*

First, a reliable spatial metric for image intensity was used. The total number of days that at least one image was taken in a location, in this case, the grid cells of Aruba, is referred to as the photo user days (PUD) (Spalding, 2017). There are several ways in which the results can be displayed on the map. Below is an explanation of which spatial indexing system is used and why.

To display geosocial data there are several ways in which this could be done. There are several ways in which you can index spatial data about metadata standards (Directory Interchange Format and Federal Geographic Data Committee). Firstly, you can use location keywords, such as the Netherlands. Secondly, the bounding polygon approach. In which you define a polygon that encloses the dataset's borders. Thirdly, a bounding box or Minimum Bounding Rectangle (MBR) can be used, which is also used to scrape the data from Flickr (Lin & Chiang, 2018; Loncomilla et al., 2022; Rees, 2003).

Overplotting is a big problem when having a large amount of points on the map. Overplotting can be prevented by creating a 'heat map', with a kernel density or kriging to create a better visualization of the data (Engel et al., 2021). However, using this method presupposes that the underlying spatial processes are continuous because they interpolate the values in between real data points. This warning is more troublesome when applied to social phenomena, such as tourism geosocial data, than it is to many natural phenomena, such as temperature data. This is because there can be stark differences between demographics (Poorthuis & Zook, 2015)

The PUD measurement shows the geosocial data with a lot of points and therefore aggregating these points to larger areas or polygons to represent the intensity of the data and calculations creates meaningful insights (Gugulica & Burghardt, 2023; Wang et al., 2023). A common way for this is using regular units with the shape of rectangles, c-squares, or hexagons for example (Gröbe & Burghardt, 2020). This is better compared to using administrative boundaries because every unit has the same size and thus the same chance of receiving points within the unit and does not stand out more visually than others in means of size (Poorthuis & Zook, 2015)

All proposed grids thus far exhibit identified problems or concerns warranting additional investigation. For this research c-squares grids were chosen because the grid is easy to implement and is already used by Slijkerman et al (2020) which makes comparison with Bonaire in the future possible.

To explore the images on the map of Aruba a concise spatial query and representation system (hereafter, C-squares grid) was used to identify where the images were taken, which resulted in an easy mapping per category of all images. A pleasant and noteworthy feature of C-squares is that it is a hierarchical grid, making it quite feasible (even without GIS) to merge smaller units into a larger one (Slijkerman et al., 2020).

C-squares is a system of geocodes and thus a global grid that can be used for spatial indexing of data. This type of spatial indexing was designed in 2001 by Tony Rees of CSIRO Marine and Atmospheric Research. The world can be geographically classified using latitude and longitude coordinates. However, c-squares an easily human readable the underlying coordinate domain is divided into four segments (NE, SE, NW, SW). According to Mahdavi- Amiri et al., 2015 ‘Each cell then receives an index of the form iyxx; where i is 1, 3, 5, or 7 if the cell is, respectively, located in the NE, SE, SW, or NW segment; y is the first digit of the cell's latitude; and xx are the first two digits of the cell's longitude’. An example can be seen below (Figure 3-13).

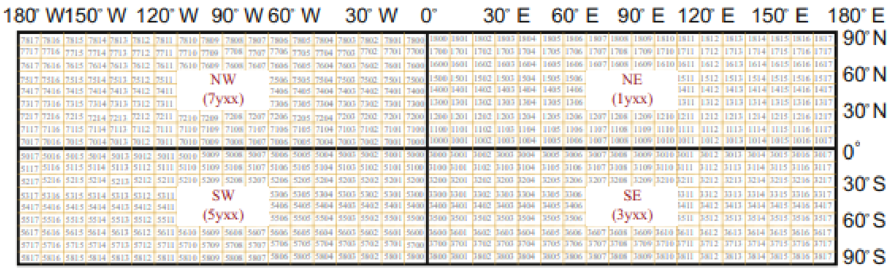



*Figure 3-13: The Concise Spatial Query and Representation System indexing example (Mahdavi-Amiri et al., 2015)*

C-squares are available in degrees from 10, 5, 1, and 0,5 are available for download, via https://www.cmar.csiro.au/csquares/ if you want smaller sizes these have to be aggregated. These are not available for download since the files are unwieldy (too large) when offering global coverage. The approach taken to generate polygon grids (aka fishnets or vector grids) assumes that a UTM (WGS84) zone is taken as the base to determine the sides (for NL UTM zone == 31, western limit = 0, eastern limit = 6 ). This is helpful when calculating the column area_km2. First, it was determined which UTM zone was best to use for Aruba, which resulted in UTM zone 19. The C-squares used is 0.005 degrees, which at the latitude of Aruba means that it has an area of 0.3 km2. And that immediately relates to the choice to specify a UTM zone when generating. Then you know which conversion is valid to obtain exact results. Figure 3-14 shows the c-squares grid in UTM zone 19. The degree of 0,005 was chosen since Aruba is a small island. To make spatial distinctions, keeping the analysis small (lots of detail) soon becomes necessary. C-squares are easy to aggregate, first, the idea was to start with as fine a resolution as we thought possible. Eventually, this was a good size. But if it was pushed too far (overdone it), it would have been easy to scale it up one or two sizes, which is also why c-squares were chosen.


*Figure 3-14: C-square grid UTM zone 19*

Figure 3-15 displays the grid over Aruba, which shows that the cells are a nice resolution to map the results. They are small but not too small so that the results can be seen easily.

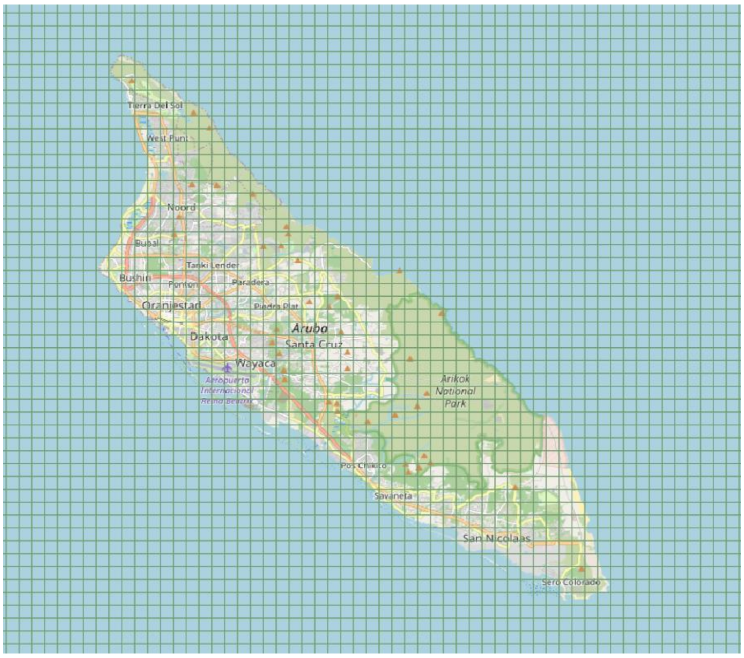



*Figure 3-15: C-squares grid in UTM zone 19, 0.005 cells*

## 3.6 Analysis

The analysis itself and the outcomes of this research will be displayed in the results section. For the analysis of this research, the main outcome was the combination of a spatial distribution analysis and a visual content analysis. The location-based analysis was used for the grid of each type of tourist, where the images were posted most frequently. For the type of tourists, a visual content analysis was used for categorizing the images.

Maps of the several types of tourists have been analyzed, which showed where certain types of tourists frequently stay. The measure used to display the intensity of the images is PUD (photo user days). PUD calculates the number of individual users that upload at least one image on a unique day, in a particular location (Wood et al., 2013). This resulted in a map of PUD intensity for all images and also a map per tourist category. These maps show how certain types of tourists are spread over Aruba and the darker the color, the higher the amount of PUD in that cell and thus the area on Aruba. The decision for PUD as a measuring unit was because otherwise, one photographer who made 50 images of one phenomenon accounted for 50. This could lead to a high amount of images at a certain place however, it only was one photographer in one moment.

A person-based analysis was done focussing on the origin of the users. There is a division between different origins: North America, Latin America, Europe, other and local. The problem here is that the privacy has been sharpened and you can only see the origin of the photographers when they made this publicly open to every one of their Flickr profiles. That is why only a limited amount can be categorized and a lot of people fall into the category ‘Other’. Under Latin America other Caribbean islands are taken into account except for Aruba., this division was also used in Slijkerman et al (2020). However, first was checked if this division was also applicable for Aruba. The CBS 2021 also looked at the number of stay over tourists by place of residence. The countries included can be seen in Table 3-4. As you can see most of these countries also were included in the division of Slijkerman et al (2020) and therefore the decision was made to also use the division.

*Table 3-4: Number of tourists in Aruba 2021 (CBS, 2021)*

| Total Stayover tourists | 806.534 |
|---|---|
| United States | 677.349 |
| Venezuela | 1.192 |
| Netherlands | 37.532 |
| Colombia | 21.053 |
| Brazil | 4.420 |
| Canada | 12.879 |
| Argentina | 2.030 |
| Germany | 3.194 |
| Rest of Europe | 11.062 |
| South-Central America | 11.180 |
| United Kingdom | 1.908 |
| Rest of the World | 22.735 |

Also, the number of days people visit Aruba is analyzed. A difference between the three photographers' time of visit is taken into account: local, Stay over, and Cruise tourists. Local tourists are people who stay longer than 30 days, and cruise tourists are included if they only stay half a day. According to Ludmer (2011) cruise ships dock between 6-12 hours. When including getting on and off the ship tourists have only 8 hours to see the island. The stay-over tourists are the people who stay more than half a day but no more than 30 days. This was chosen because tourists who visit by cruise mainly stay no longer than a whole day. Slijkerman et al (2020) also used these three divisions.

Finally, the PUD is also compared to a land-use type map and two maps of protected areas in Aruba (marine areas and nature areas), which can be seen in Appendix D. This is done to see if the tourists frequently visit the natural protected areas of Aruba and thus if the impact there is big or not. How many images are taken in these areas and what is the intensity of PUD in this area? For the marine area map the coastal and seascape tourists PUD intensity will be mapped. For the protected natural area the 'Landscape' and 'Wildlife' PUD intensity is mapped. Also, the amount of images taken in these areas is shown in a chart.

## 3.7 Critical assessment

Critical assessment of the data and outcomes has been realized by sharing the outcomes of the research with tourist agencies in Aruba and conducting surveys with them. The following agencies have been contacted:

- Aruba Tourism Authority (ATA)
- Aruba Hotel and Tourism Association (AHATA)
- Renaissance

They have a comprehensive view of the current situation and could therefore provide insight into whether the maps and their hotspots showed a realistic image of touristic preferences. Two surveys have been sent to the participants, the first one (Appendix D) has been sent before even the results have been mapped for this research. In the first survey respondents were first asked to provide some basic information about the tourism expert themselves. Second, the experts have been asked to indicate on a map of Aruba where they believe certain types of tourists visit the island, based on their expertise. These outcomes have been compared with the outcome of this research. The second survey included the outcomes of this research and asked more about whether the outcomes represented the actual situation and what agencies could do with this freshly obtained insight.

However, contacting these organizations proved difficult. Many did not respond or were busy. this was discussed with Peter Verweij (the submitter of this study). Peter provided additional contact information that was also approached. This had led to eventually seven respondents in total.

In addition to looking at the answers provided by the tourism experts in Aruba, the data has also been compared to other statistics. For example, the origin of the photographers has been compared to see if the sample used for this research was in line with the monitored statistics from for example CBS (Central Bureau of Statistics). The critical assessment will be explained in Chapter 5.

# 4 Results

In this chapter, results are given and the second half of sub-question 5 and 6 partly addressed: '*How do tourists visiting Aruba distribute across different environments?* 'and '*What insights can be drawn from the spatio-temporal analysis of geosocial data regarding the impact of tourism on the environment and visitor characteristics in Aruba?*'. First, a first glimpse of the image locations is given. Secondly, a closer look into the PUD is given which first focuses on all tourists, and after that divided into categories. Thirdly, the origin of the photographer will be looked at. Hereafter the time spent by the photographer in Aruba, the photographer's time of visit, will be displayed. Finally, the PUD will be compared with a land-use map of Aruba and with a protected marine and natural areas map.

## 4.1 Locations of Images in Aruba

A preliminary analysis of the data is presented in Figure 4-1 some images are places far from the coast but remain within the bounding box of Aruba, these could be taken from a ship or airplane. Later on in this research, these images will be looked at carefully. Visible is already that the coastlines are in demand, but also Oranjestad, San Nicolas, the national park, and the Spanish lagoon. At the moment it is hard to see the division of images due to the overlapping points, this is called overplotting. Therefore the points are aggregated to a c-squares grid.

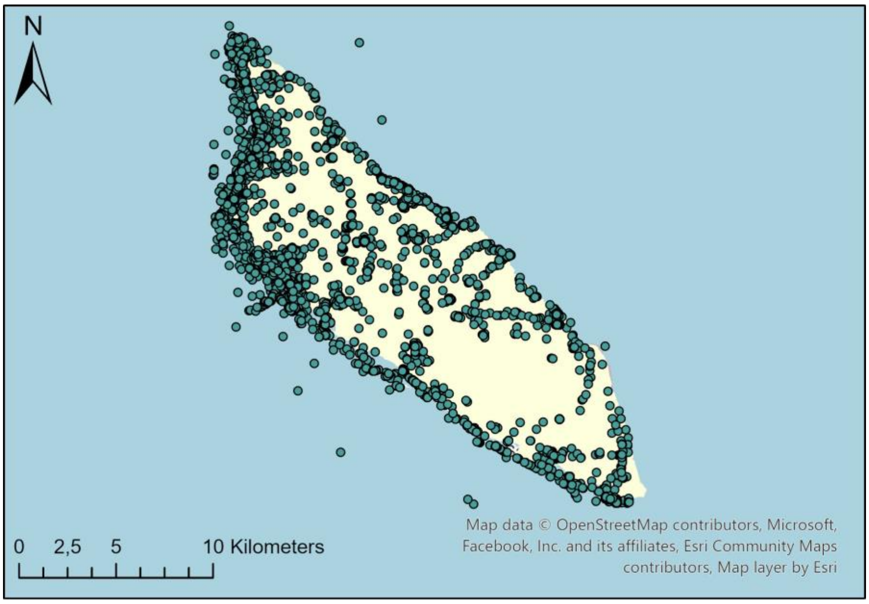



*Figure 4-1: Locations of the Flickr images extracted*

## 4.2 Photo User Days intensity

Below first the overall data will be shown, the PUD of all images and the relative division of categories. Thereafter the categories themselves and the spatial division will be looked at closely.

### *4.2.1 Phot User Day Tourists Aruba*

Looking at Figure 4-2 it is visible that tourism in Aruba is spread over the whole island except for the eastern part of Arikok National Park. However, the center of this park is frequently visited. The coastlines and Oranjestad also have a high PUD. Figure 4-3 displays the division of PUD per category. Most images fall into the 'Coastal' category however, the second biggest category is 'Other". While categorizing the images it was clear that people post also images on Flickr that are not easy to categorize into environmental categories: such as airplanes, cars, and urban images (houses, shops, museums), I <3 Aruba signs and selfies.

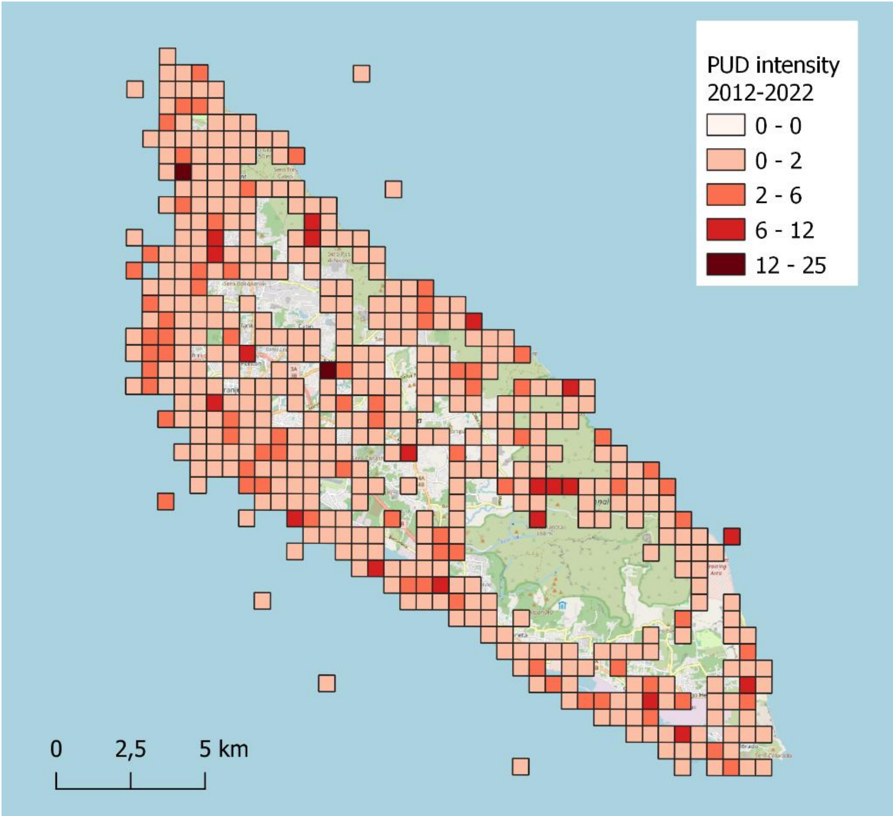



*Figure 4-2: Photo User Days intensity Aruba 2012-2022*

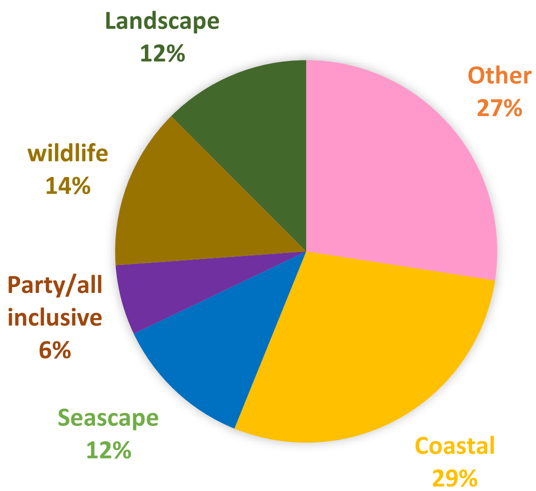



*Figure 4-3: Photo User Days division of categories in relative numbers*

### *4.2.2 Photo User Days per type of tourist*

Below the maps are shown per tourist category (Figure 4-4 to Figure 4-9). Looking at the ‘Coastal’ maps almost all images are taken at the coast. Most images are on the east coast, but also north-south, and above Oranjestad are popular destinations (Figure 4-4). The ‘Party/all-inclusive’ are also located in the urban areas, and at the coast where the casino is located (figure 4-5). The ‘Seascape’ category is mainly on the west coast compared to the coastal and some images are more into the sea (figure 4-6). For the ‘Landscape’ category most are located in the Arikok national park (Figure 4-7). ‘Wildlife’ is mainly on the east coast located, where the mangrove areas and reef islands are (Figure 4-8). While categorizing the most wildlife seen were iguanas after that it was Flamingo and Fish in the sea and birds. The ‘Other’ category is mainly located in the urban area of Aruba (Figure 4-9).

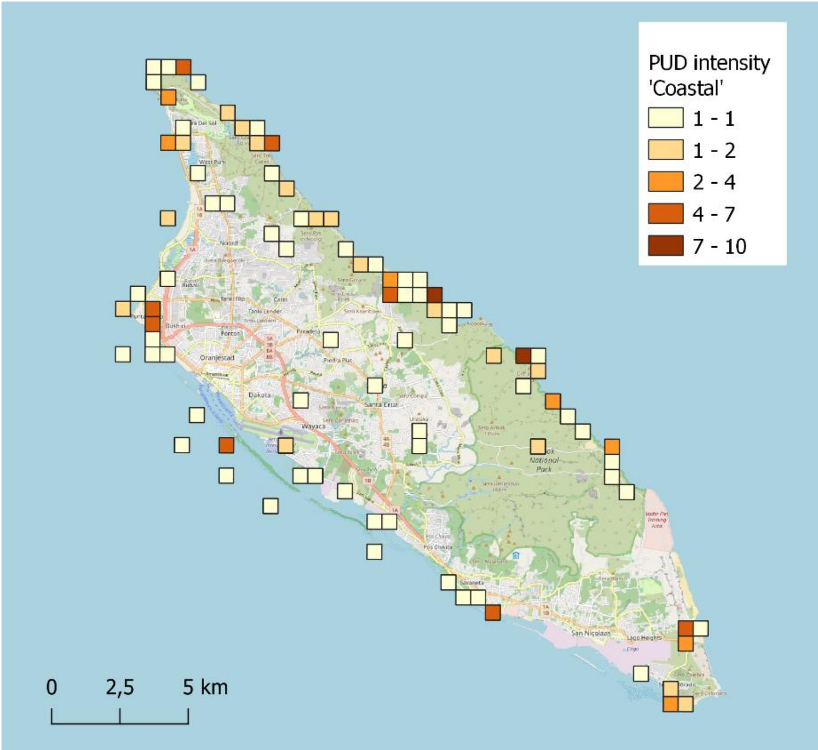



*Figure 4-4: Spatial distribution of Photo User Days 'Coastal' tourist*

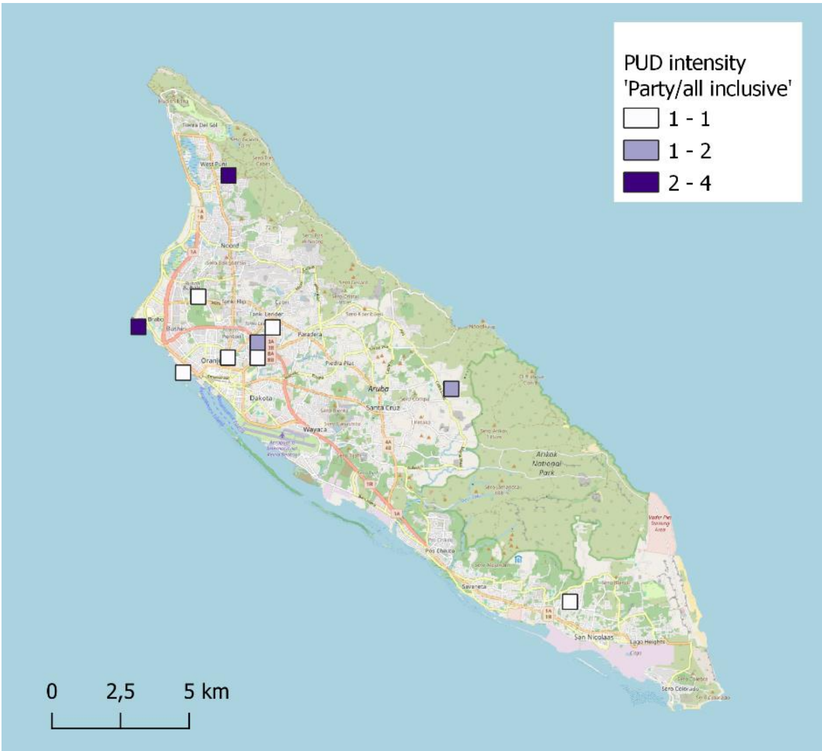



*Figure 4-5: Spatial distribution of Photo User Days 'Party/all-inclusive' tourist*

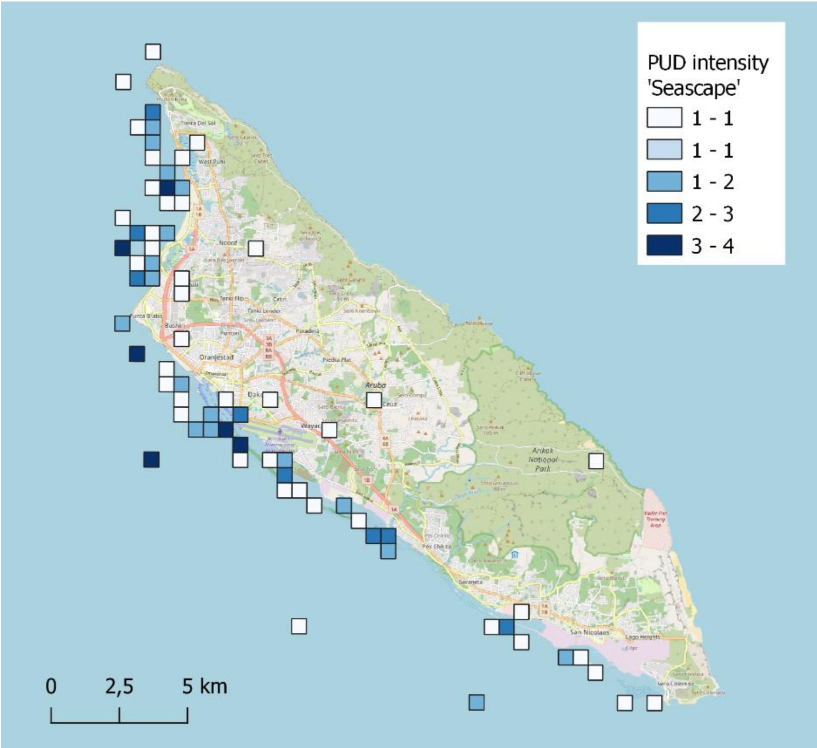



*Figure 4-6: Spatial distribution of Photo User Days 'Seascape' tourist*

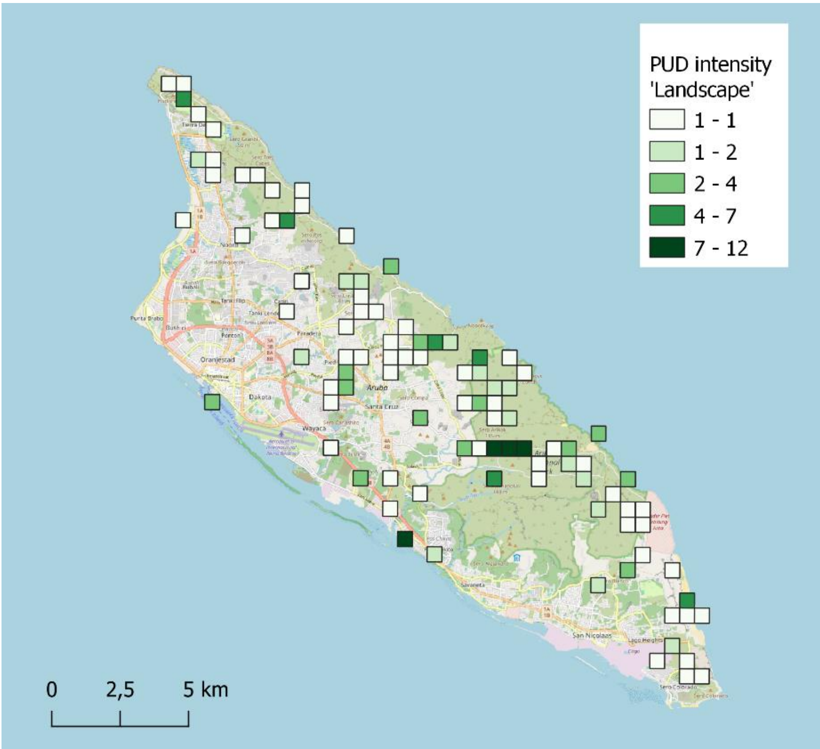



*Figure 4-7: Spatial distribution of Photo User Days 'Landscape' tourist*

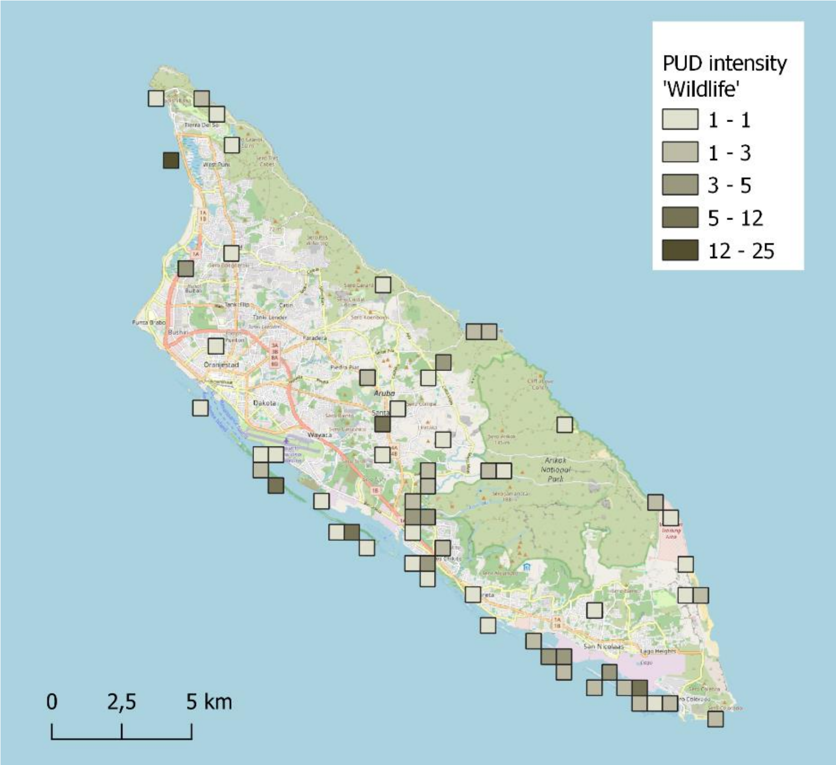



*Figure 4-8: Spatial distribution of Photo User Days 'Wildlife' tourist*

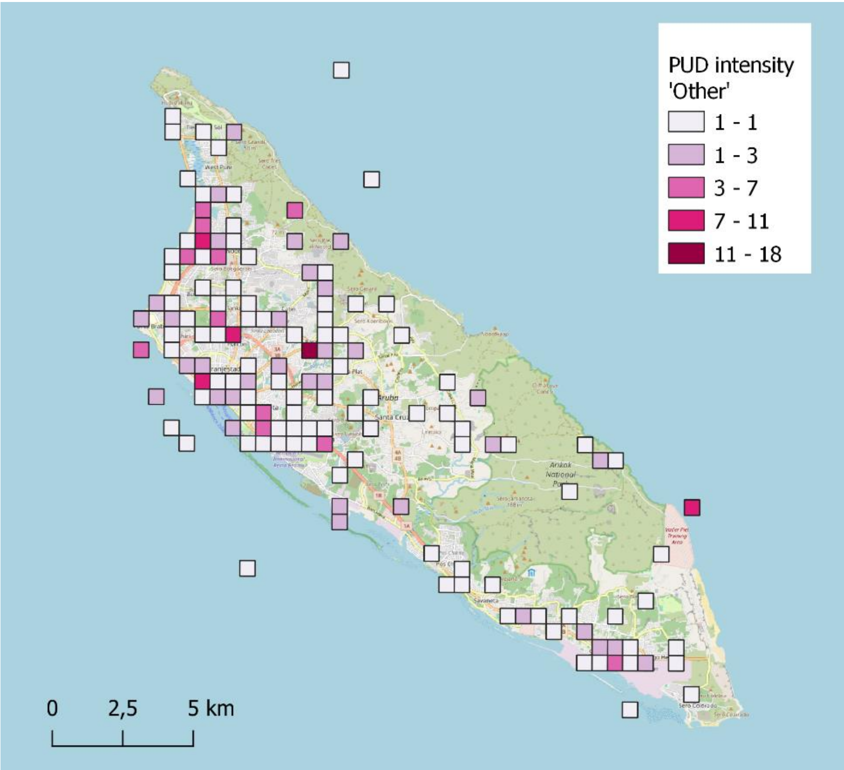



*Figure 4-9: Spatial distribution of Photo User Days 'Other' tourist*

### 4.2.3 Valid/invalid position

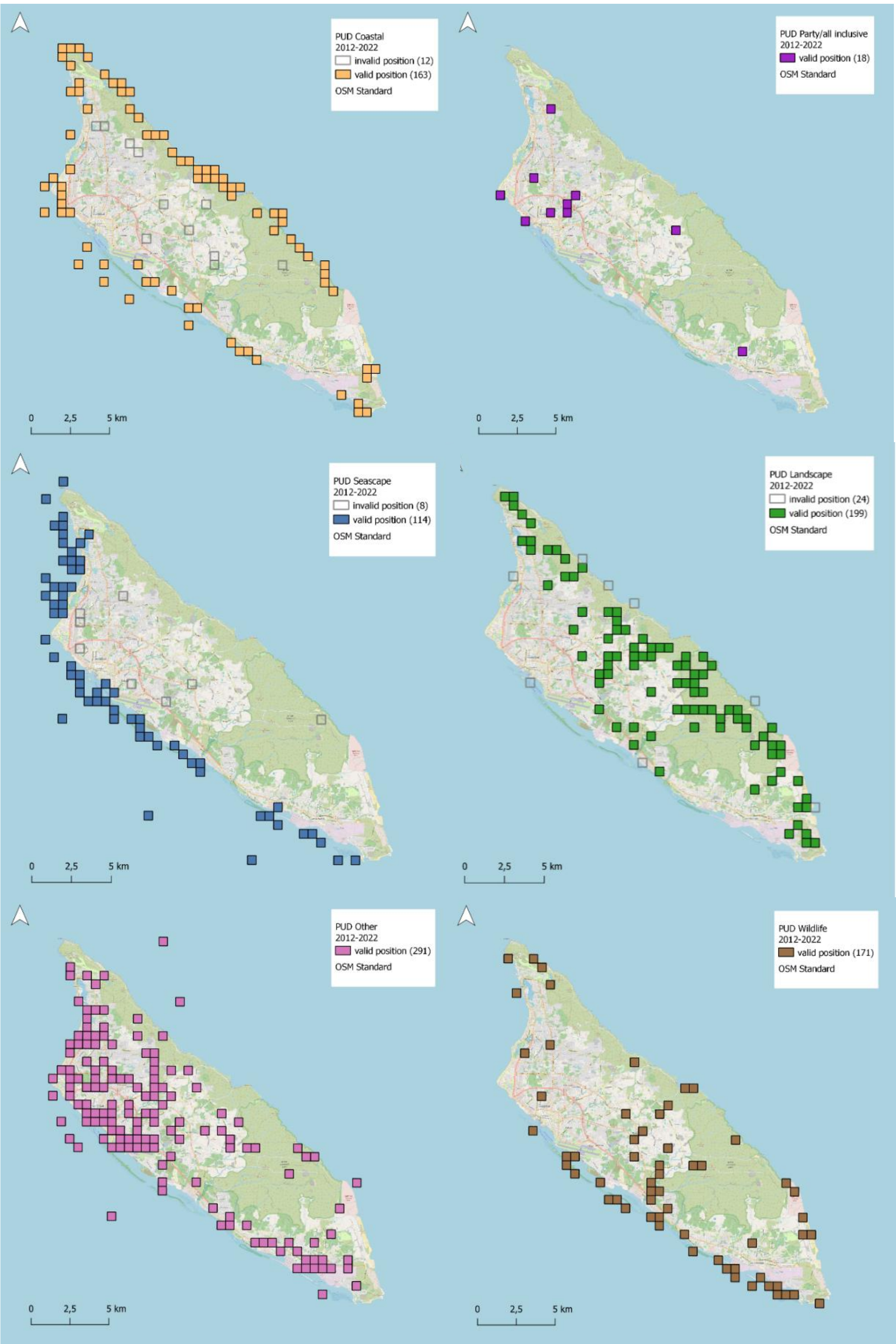



*Figure 4-10: Photo User Days invalid/valid position (Coastal, Party/all-inclusive, Seascape, Landscape, Wildlife, Other)*

For the 'Coastal', 'Seascape' and 'Landscape' category a few positions of the PUD were found invalid, which are grey outlined (Figure 4-10). For the 'Coastal' map these are locations in where the position of the PUD cell was too far from the coast. For the 'Seascape' map it is a similar story as to the coastal map. For the 'Landscape' map this are the PUD cell position was close to the coast. The numbers behind the valid/invalid position show how many PUD per category shows how many were invalid.

## 4.3 Origin of the Photographer

From Flickr, the place of origin filled in by several users of Flickr has been extracted. The results can be seen in Figures 4-11 to 4-16. From all the images gathered from Flickr, only 386 people have their locations listed publicly. The 'Europe' origin photographer mainly visits the urban areas but also the northwest coast is popular. (Figure 4-11). A total of 8 photographers came from Latin America who are all spread over the island (Figure 4-12). The 'local' origin of the photographers is based on .. photographers, with one concentration of three which lays in the middle of Aruba (Figure 4-13). In Figure 4-16 'North-America' is the second biggest origin of photographer category, and mainly visits the coast but also Arikok national park (Figure 4-14). The photographers from which the place of origin was not given, were placed under the 'Other' (Figure 4-15).

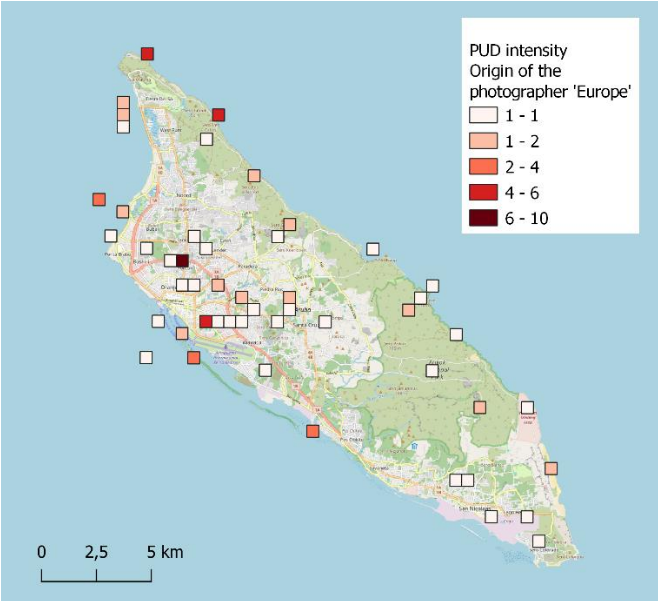



*Figure 4-11: Spatial distribution and intensity of Photo User Days divided by the origin of the photographer 'Europe'*

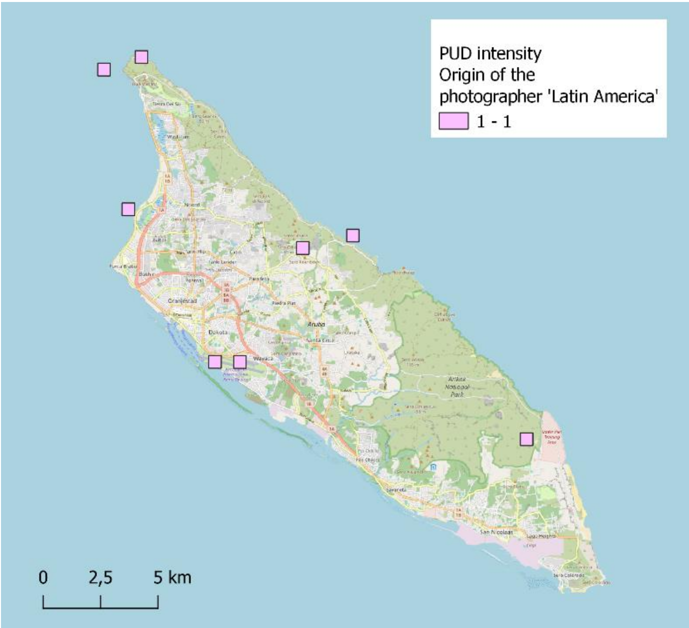



*Figure 4-12: Spatial distribution and intensity of Photo User Days divided by the origin of the photographer 'Latin America'*

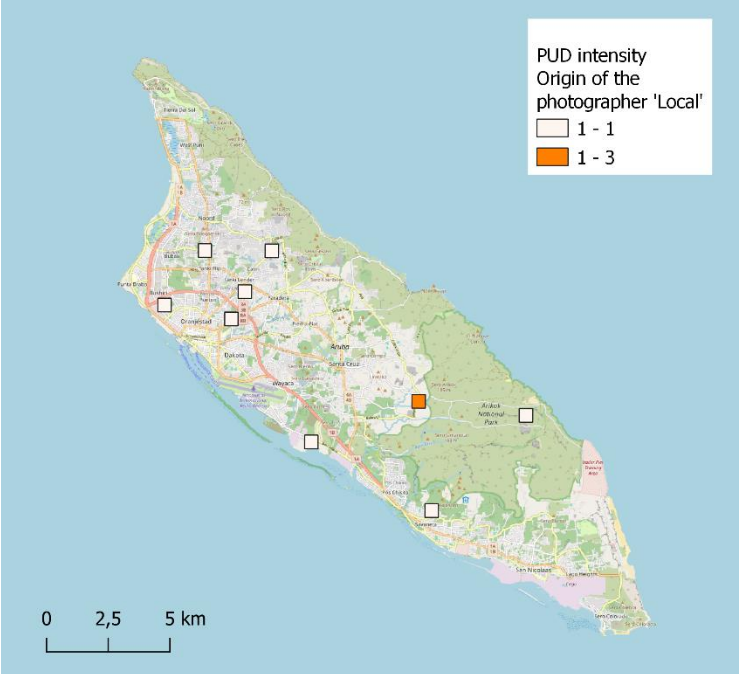



*Figure 4-13: Spatial distribution and intensity of Photo User Days divided by the origin of the photographer 'Local'*

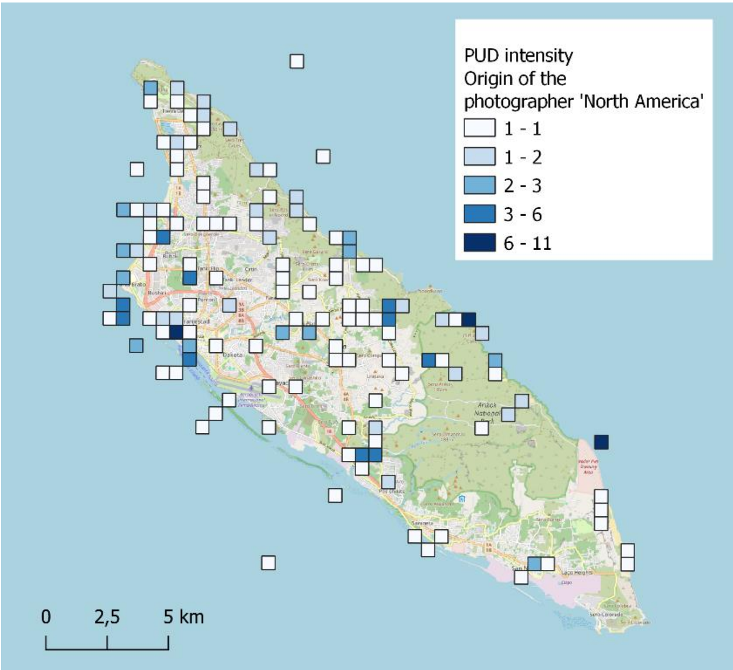



*Figure 4-14: Spatial distribution and intensity of Photo User Days divided by the origin of the photographer 'North America'*

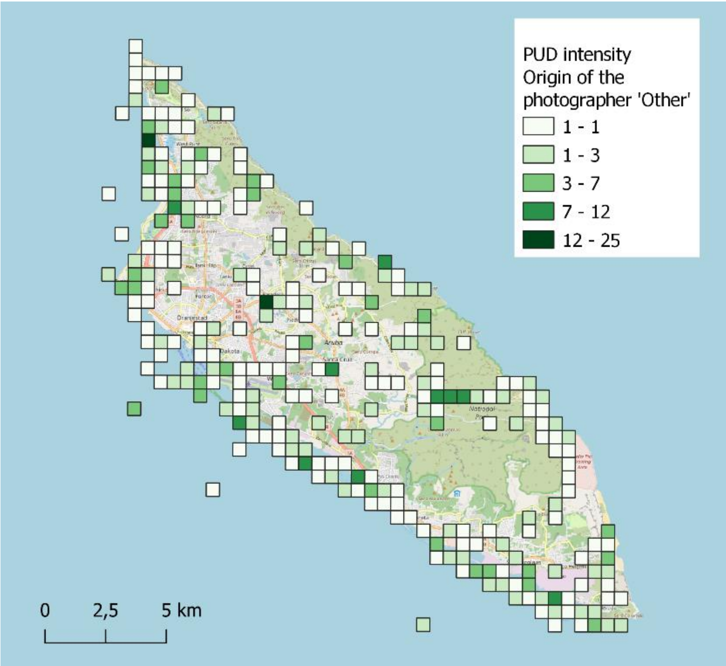



*Figure 4-15: Spatial distribution and intensity of Photo User Days divided by the origin of the photographer 'Other'*

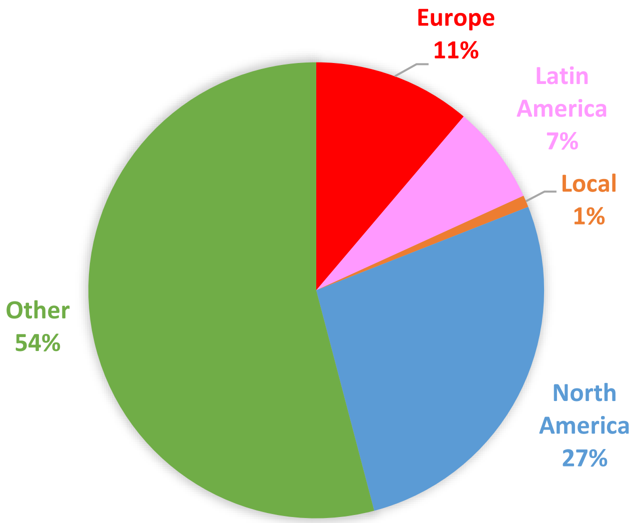



*Figure 4-16: Relative numbers of the division of the origin of the photographer*

## 4.4 Photographer time of visit

The spatio-temporal analysis of this research looks at the photographer's time of visit to the island. The time spent in this research is the time between uploading the first and last image on Flickr. Half of the photographers fall into the ‘Cruise’ category (Figure 4-20). Looking at the ‘Stay over’ PUD map there is a variety in where this photographer goes, Oranjestad itself, but also the lighthouse and California sand dunes are popular. The spread is also quite evenly distributed over the island (Figure 4-17). For the ‘Cruise’ photographer the harbor front of Oranjestad also has some PUD. Most PUDs however are seen at the Arikok natural park (Figure 4-18). The ‘Local’ photographer's time of visit is mainly in the urban area located (4-19).

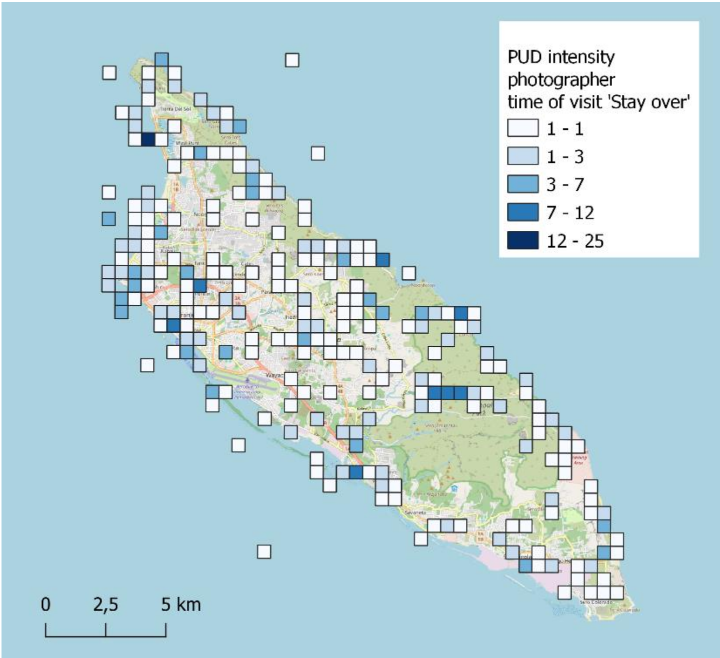



*Figure 4-17: Spatial distribution and intensity of Photo User Days by photographer time of visit 'Stay over'*

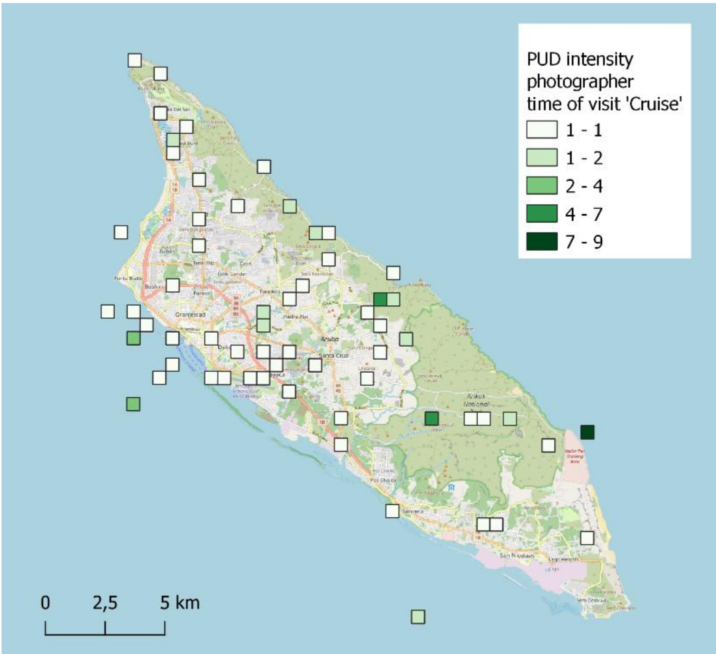



*Figure 4-18: Spatial distribution and intensity of Photo User Days by photographer time of visit 'Cruise'*

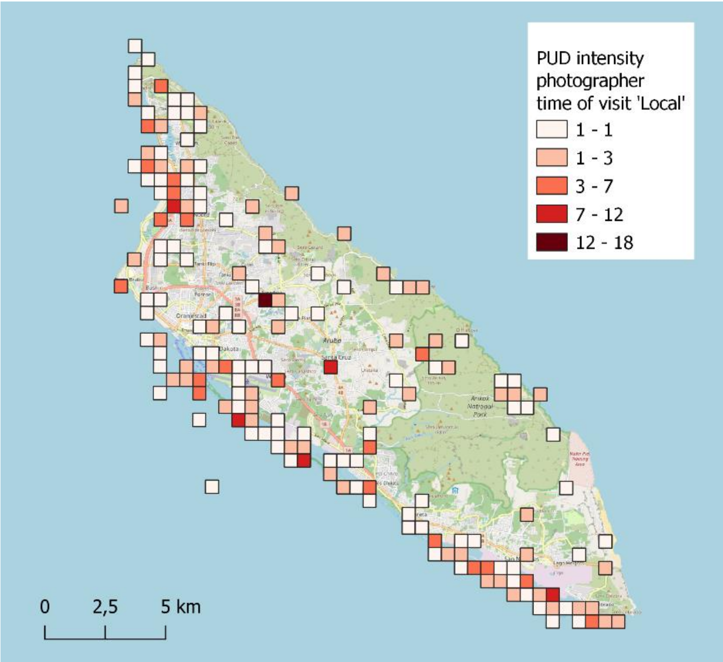



*Figure 4-19: Spatial distribution and intensity of Photo User Days by photographer time of visit 'Local'*

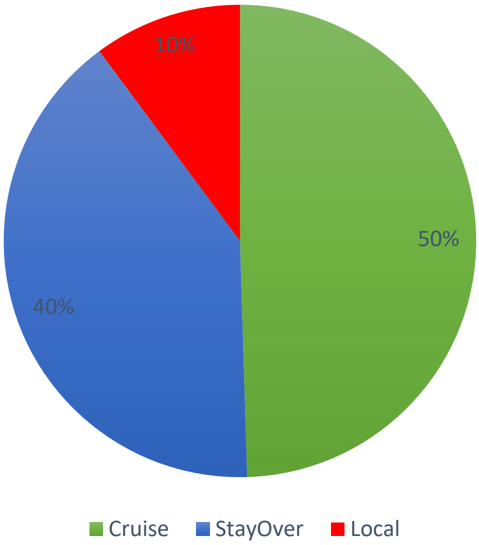



*Figure 4-20: Relative number of distribution by photographer time of visit*

## 4.5 Tourists versus land-use types and protected areas

### *4.5.1 Land-use Types Photo User Days and Images*

Figure 4-21 shows how the PUD distribution intensity is compared to a land-use type map of Aruba (Appendix D). Figure 4-22 shows the number of images per land use type, most images were taken in the nature and landscape and beach land-use types. The Oranjestad harbourfront is thereafter a popular location where a lot of images were taken. A total of 6660 images were taken in these three types of land use and therefore almost 40% of the images were captured there.

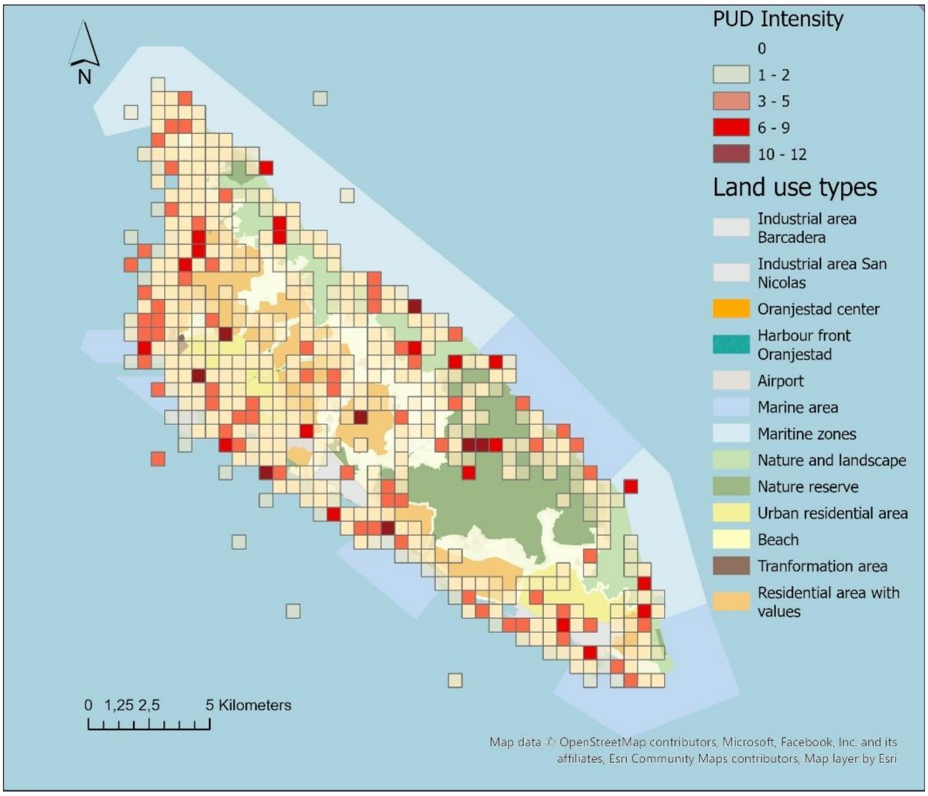



*Figure 4-21: Photo User Days intensity on land-use types Aruba*

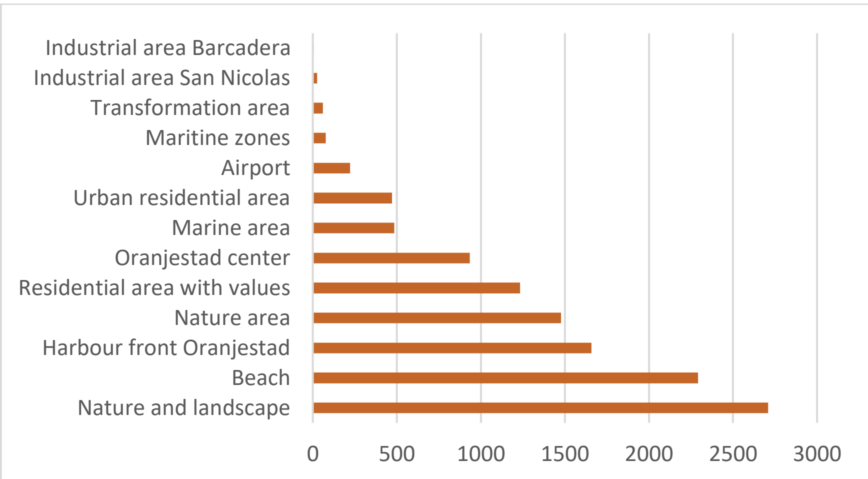



*Figure 4-22: Number of images per land-use type Aruba (total of 11648 images)*

### 4.5.2 Protected marine areas Photo User Days and images

As can be seen in the maps the 'Coastal' and 'Seascape' high PUD are not necessarily within the protected marine areas (Figure 4-23 and 4-24). Especially the northwest of Aruba is mainly popular with these tourists and here there are no protected marine areas. The most frequently visited protected marine area is Marine area Sero Colorado which lies in the South-east of Aruba and 268 images were captured there, see Figure 4-25.

mi

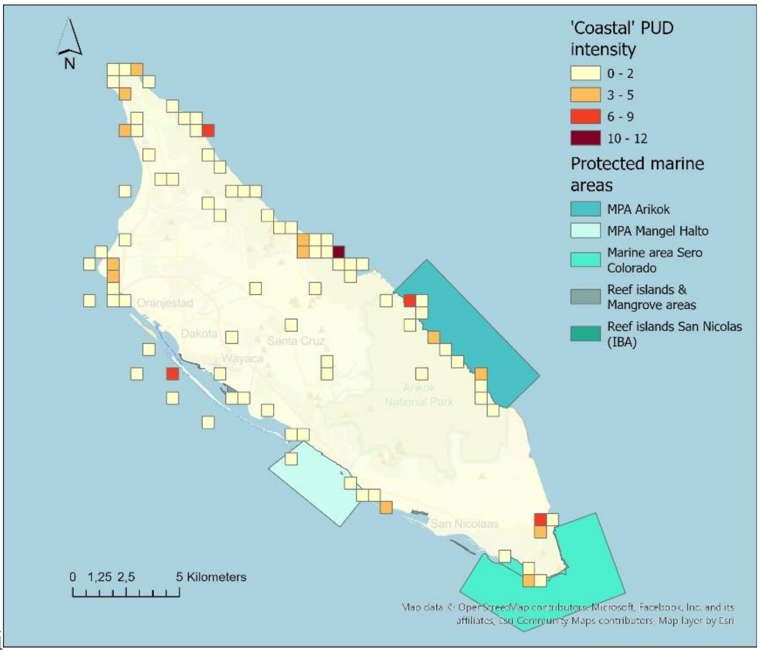



*Figure 4-23: Photo User Days Intensity of 'Coastal' tourists in protected marine areas Aruba*

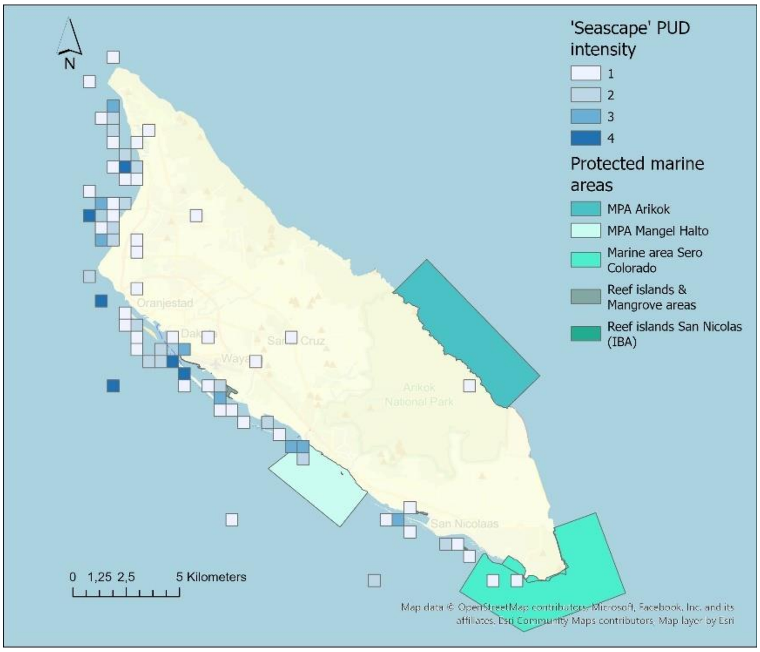



*Figure 4-24: Photo User Days intensity of 'Seascape' tourists in protected marine areas Aruba*

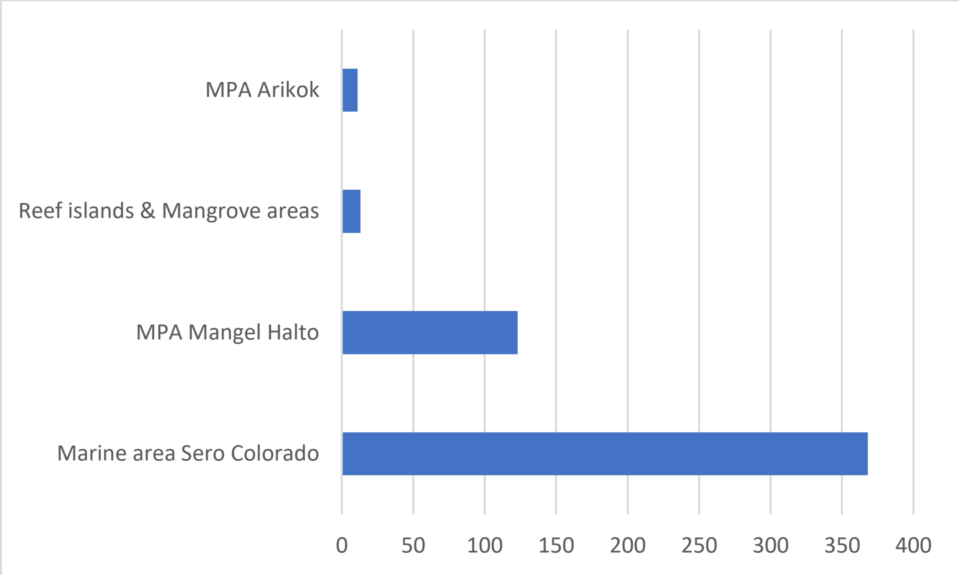



*Figure 4-25: Number of images per Protected marine area on Aruba*

### *4.5.3 Protected nature areas Photo User Days and images*

Looking at the division of 'Landscape' and 'Wildlife' tourists PUD the Arikok Natural Park is popular (670 images were taken there, see Figure 4-28). At the eastern coastline, a lot of wildlife is seen, at the moment it is not well visible but when looking at the protected nature area map in Appendix D you see a lot of islands on the East coast which are protected natural areas. The 'Wildlife' images are thus also in these protected Reef islands and Mangrove areas (Figure 4-26).

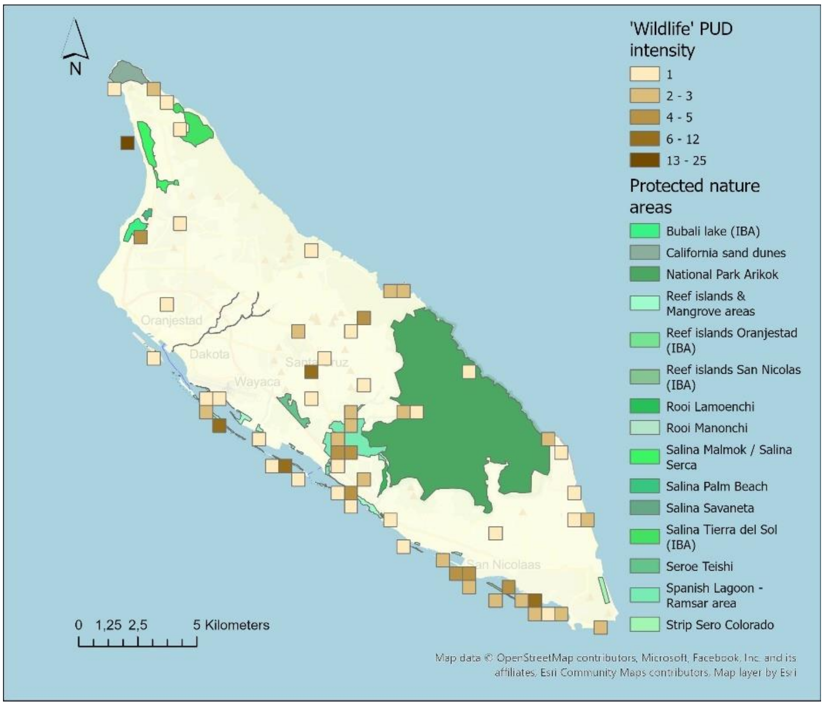



*Figure 4-26: Photo User Days intensity of 'Wildlife' tourists on protected nature areas Aruba*

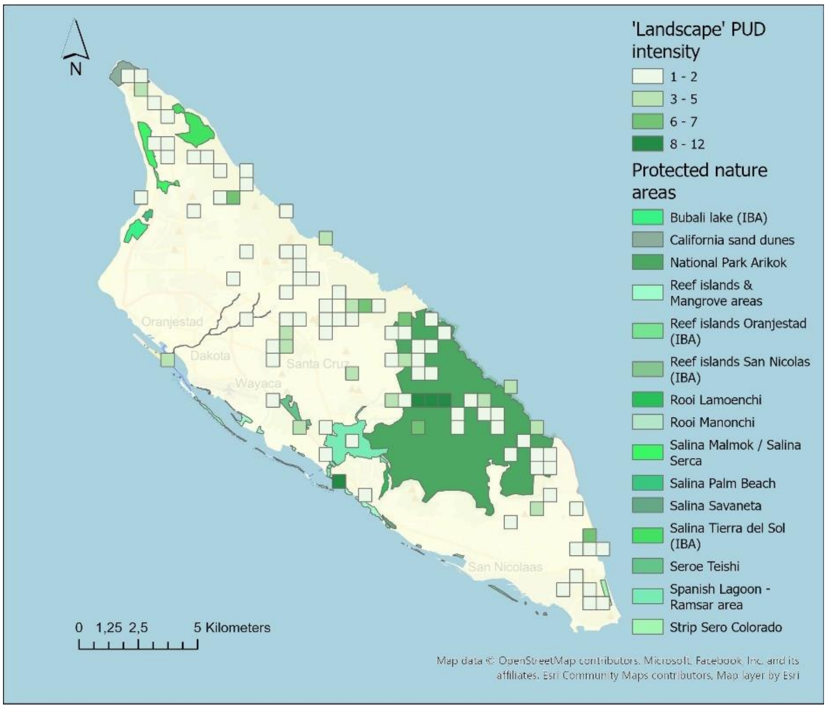



*Figure 4-27: Photo User Days intensity of 'Landscape' tourists on protected nature areas Aruba*

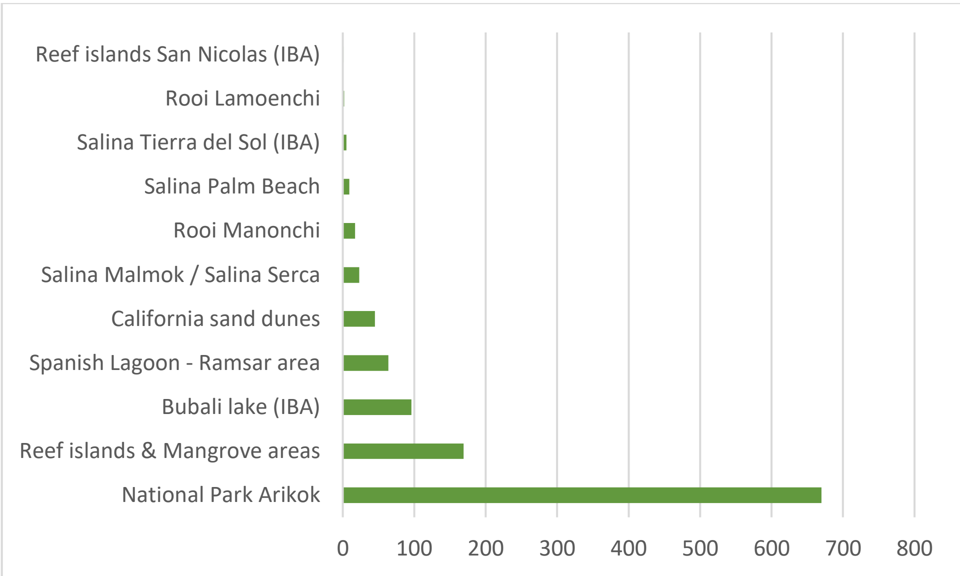



*Figure 4-28: Number of images per Protected Nature areas in Aruba*

# 5 Critical Assessment

In this chapter sub-question number 6 is answered: '*What insights can be drawn from the spatio-temporal analysis of geosocial data regarding the impact of tourism on the environment and tourist characteristics in Aruba?*' To get answers to this question, surveys were circulated to several Tourism Agencies. Appendix G shows the responses to the surveys of tourism agents participating in this study. Seven individuals took part in the study, with one person failing to complete the second survey. Their knowledge of the island of Aruba was on average 4,33 out of 5. One respondent added a six to the scalebar. Some live in the northern part of Aruba, some in Oranjestad. Most of the respondents work at Aruba Tourism Authority, Budget Marine Aruba, and Renaissance Wind Creek Aruba Resort. Below first the categorized maps are compared to the maps drawn by the respondents. Thereafter the results of the second survey are displayed.

## 5.1 Results compared to the first survey

### *5.1.1 Coastal*

For the 'Coastal' tourists they highlighted where the tourists would mainly go. Most respondents mentioned: palm beach and Eagle Beach in the northwest, where the high and low-rise hotel area is. However Respondent 4 did mention this too, it also said that more towards the north side, the waters are a bit rougher and it is unsafe to swim in most parts. In the southeast Baby Beach and Rodgers Beach were mentioned quite often, with good sightseeing spots and tourist activities. Mangel and Halto beaches on the west side of the island, where a lot of inshore waters are. These areas were also highlighted in the 'Coastal' map but did not have a high-intensity PUD, except for the South-East.

Looking at the 'Coastal' map result in this research the east side has a higher intensity of PUD, only 50% of the respondents have penciled in this coast. Such as the natural pool for off-road tours. Especially respondent 5 highlighted the whole East Coast and mentioned: tourists exploring the North Coast area on their own (jeeps, UTV/ATV, bus, hiking, horseback, mountain biking, etc.). Also, the Arikok National Park coastal area is highlighted.

### *5.1.2 Party/all-inclusive*

For the 'Party' map most have drawn Palm Beach and downtown areas. Palm Beach has high-rise hotels and a lot of bars and restaurants. But also low-rise and downtown hotels are mentioned quite often. One respondent also mentioned the strip that contains many nightlife activities. The low and high-rise hotels also have a high PUD in the 'Party map' as can be seen in the results. However, most of the respondents have drawn these types of tourists to the coast, but in the 'Party' map there are also places further away from the coast that are not drawn by the respondents.

### *5.1.3 Seascape*

Compared to the 'Coastal' tourists the east side coast of Aruba is frequently mapped here by some respondents. However, others did draw the west coast. Again Palm Beach is mentioned where a lot of water sports activities can be done. Also diving near the reef islands, and wreck diving are mentioned too. The north is highlighted, for sightseeing. Two respondents highlighted the whole coast around the island. Because every common beach includes optional sports activities, such as kites and windsurfing. For views, there are no beaches that tourists do not visit except near the landfill.

Comparing these findings with the 'Seascape' map it is clear that the north has a lot of options for water sports and the wreck diving on the west coast is also visible. However, the south and east coasts are not highlighted in the PUD intensity. Yet, this coastline is drawn by most tourism agencies.

### *5.1.4 Landscape*

The Arikok National Park is here the most common place that came forward in the drawings of the respondents, which also can be seen in the 'Landscape' map But also ATV/UTV tours, and horseback riding activities are mainly on the East coast of the island, which also can be seen as the east part of the island also have high PUD intensity there. Ayo Rock Formation and Casibari rock formation are also mentioned, also the northernmost tip of the island is highlighted, where the Sasaeawichi dunes, terra cora are located, and the wetlands. Here again, the 'Landscape' map highlights this with PUD intensity.

### *5.1.5 Wildlife*

Again the Arikok National Park is highlighted, but also the north and west coast, with bird watching in the wetlands which also are visible in the PUD intensity of the 'Wildlife' map. The respondents mention: Lizards, snakes, crabs, hatching turtles, monarch butterflies, mountain-climbing goats, donkeys, and the neon-green prikichi. Bubali place to see the migratory birds, and the butterfly farm is the place up in the north with the highest PUD and this is also mentioned a lot. While categorizing these butterflies and donkeys also were seen often while categorizing.

### *5.1.6 Other*

The 'Other' map is a hard category to compare since anything not included in the other categories falls within this category. However, every respondent did highlight specific parts of Aruba, mainly due to the fact the urban also fell within this category. For example, Oranjestad was highlighted often because you can go shopping or have dinner, also the north with Palm Beach was mentioned. But San Nicolas and the art walk in the south also is mentioned by three respondents. Comparing these results to the PUD map the results are quite similar however more land inward there is also a high intensity of PUD which is not highlighted by all respondents, only a few.

## 5.2 Results of the second survey

### *5.2.1 Outcomes resemble the real-world situation?*

On a scale from 1 to 5, every respondent had to say how well the study outcomes resemble the real-world situation. A lot of people graded 4 and some 3, which eventually led to a mean of 3,7. Which Is a sufficient grade however, there is room for improvement. Therefore the question with comments and suggestions was added to see where these improvements could have been.

One respondent mentioned that the results should be looked at periodically to see if the spatial patterns shift over time. Which should be looked at in follow-up research indeed. Two respondents were not sure if Flickr was the right medium or if other social media accounts such as Instagram or Facebook would affect the results. This is indeed something to take into account, however looking at the data accessibility and the financial options, Flickr was the only suited social media platform to use. The method options and explanations were not shared with the respondents. One respondent commented about the lack of trash disposal throughout the island and facilities that make the tourist experience more natural in inclusive (to avoid damage to ecosystems).

### *5.2.2 Outcomes useful for the organization?*

The second and third questions were about how useful these outcomes are for the organization. One question again had a scale from 1 to 5, respondent seven graded a 2, however, the rest of the respondents graded 4 or 3 with a mean of 3,4.

The other question was about what purpose the study outcomes could be used for. These answers varied. For example, one respondent mentioned that it could be used to identify areas that may require community enhancements, wildlife conservation, and voluntary clean-up efforts to reduce the impact on nature. This was also one of many reasons why this research has been done. But also to identify the interest of past tourists and provide customized suggestions for future tourists to appeal to and add value to their stay in Aruba. To gain more insights into tourism activity, inform marketing strategies, guide infrastructure development, plan tourism initiatives, prioritize conservation efforts, enhance tourist experiences, and foster collaboration with stakeholders. One respondent summarized these answers: 'to understand tourist flow and attempt to manage such, both from a sustainability standpoint (nature preservation) and where visited attractions are, and how to manage/enhance/maintain these'.

## 5.3 Results compared to other statistics

### *5.3.1 Origin of the Photographer*

In section 3.6 Analysis of the number of tourist arrivals in Aruba in 2021 is shown (Table 3-4). A total of 806.534 tourists visited Aruba in 2021. Comparing these numbers with the origin of the photographer's statistics can give useful insights into the Flickr data. Therefore the countries represented in Table 3-4 were divided into the categories used for the origin of the photographer without the 'Local' category. Also to compare a new pie chart was made for the geosocial Flickr data, which excluded the 'Local' category.

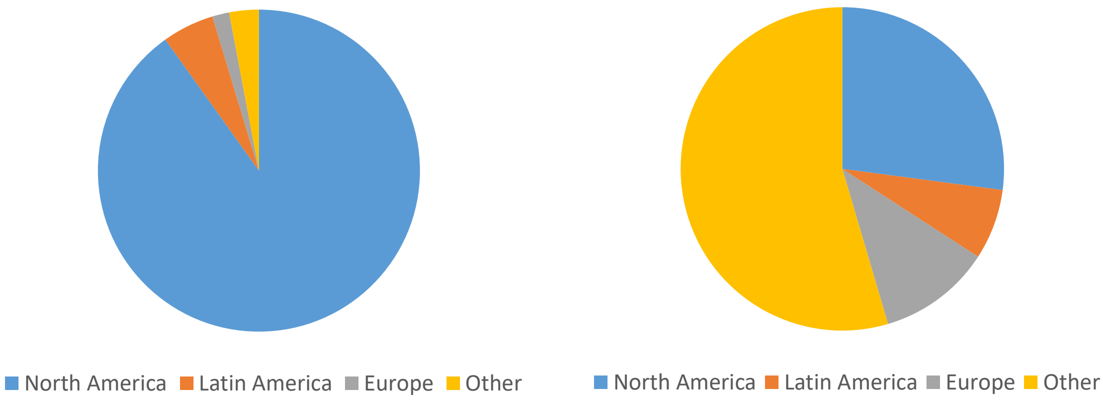



*Figure 5-1: Division Visitors Aruba 2021 (left) versus origin of photographer division 2012-2022 (right)*

Figure 5-1 shows that the representation of the origin of the photographer between 2012-2022 is not comparable with the Division of visitors from CBS. The amount of tourists from North America is more than 80% of the total. Whereas, for the origin of the photographer division it only covers a bit over 25%. Only 830 photographers shared their location, from which seven photographers were 'Locals'. Therefore only 823 photographers were accounted for in the origin of the photographer division. Even though the Flickr division represents ten years, 829 photographers compared to the total amount of tourists in 2021 represents only 0,10%.

### *5.3.2 Photographer time of visit*

The total number of Stay-over tourists has always been higher compared to the number of Cruise tourists (Figure 3-2). In the last couple of years, the percentage of cruise tourist visitors relative to stay over tourists has grown (Table 5-1).

*Table 5-1: Tourists that have been to Aruba by cruise, in percentages (CBS, 2021)*

| Year | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Yes | 8,6% | 9,2% | 12,2% | 12,7% | 11,8% | 11,6% | 12,3% | 13,0% | 22,8% | 25,5% | 20,7% |
| No | 91,4% | 90,8% | 87,8% | 87,3% | 88,2% | 88,4% | 87,7% | 87,0% | 77,2% | 74,5% | 79,3% |

Comparing these numbers to the results of this research, more photographers fall into the Cruise category (50% ±) see Figure 4-20. Therefore the results of this research do not resemble the real world situation.

# 6 Discussion, conclusion, and recommendations

## 6.1 Discussion

### *6.1.1 Geosocial Flickr Data Characteristics*

It is common knowledge that social media is universally used by a lot of people. However, even here there was a bias in what type of people use social media. The most active people who use social media are young, wealthy, educated, English-speaking people from the First World. This was also the case for the users of Flickr (Da Mota & Pickering, 2020). Using only Flickr in research to map tourism for a certain destination therefore leads to only capturing a specific group of tourists which does not necessarily represent the whole range of tourists visiting Aruba. Also, the frequency and number of uploaded images per user influenced the dataset and outcomes of the analysis. For example, one person who posted their images every day compared to one user who posted one image results in an unequal division of images per user, and could therefore influence the results regarding for instance popularity of a particular destination (Mangachena et al., 2023).

The content of the images from Flickr could vary a lot. As mentioned by Wong et al (2017) a lot of nature images are posted on Flickr. While categorizing this was true for more than half of the images. However, the other images were of for example food, a poster/invitation for a party, or an art piece in a museum. While gathering the data is easy, it is hard to fully understand what the content is of it. Manually categorizing the images therefore did give insights into how many images were not useful for this research and thus directly fell under the ‘Other’ category (Figure 4-9). Also while categorizing 889 images had no URL working, which then led to leaving these images out of the research.

### *6.1.2 Image annotation method*

As explained in Section 3.5.2 there are several techniques to annotate the images, manually or with computer vision methodologies (Zhang et al., 2020). The ‘Photocategorizer’ used in this research did contribute to understanding the data well. However, it was a time-consuming method, and later on, several other methods and tools were advised. For example automatic image annotation tools could be used to categorize the maps as well. The options of these were explored but did not cost money, or a training dataset was needed. Later on (when categorizing was done) it was clear that Wageningen University possessed the V7 tool. Looking back this would be a good option for categorizing the images. Most tools thought of beforehand would be too detailed to categorize the images by annotating a lot of features in the image. Later was found out that this could also be done with the categories used for this research. A sidenote of using these tools is that eventually, they will categorize the images automatically by learning from manual categorizing at the beginning. That means that at the beginning the researcher still has to put in a lot of work, and at the end still has to check if everything was well handled by the tool.

Looking at the invalid/valid maps of the categories (Figure 4-10) there can still be made some improvements. Some pictures of certain categories are placed in spots where the likelihood of for example seascape is rare. 4% of the total amount of PUD had an invalid position, 44 of 1100. To understand if this is due to the locations of the photos themselves or due to the categorizing method it would be good to add another categorizing method and compare the outcomes.

### *6.1.3 Temporal and spatial aspects of the data*

The origin of the photographer was compared with CBS data, which shows that the division of origin of the photographer does not comply with the division of CBS (Figure 5-1). Also, the number of tourists is compared with the number of Photographers. However, the origin of the photographer, due to the updated privacy policy, was harder to obtain. Therefore only half of the origin of the photographer could be located, 823 in total without locals. Besides, as can be seen in the ‘Other’ map (Figure 4-9), a lot of images were taken in the sea, this is because a lot of cruise and airplane tourists made images from the island from above or in the Cruise ship. This was sometimes troublesome while categorizing because an image from an airplane of a coast was indeed coastal however was made in the plane. Which then resulted in placing this image in the 'Other' category.

### *6.1.4 Tourist Classification Method*

How many tourists visit Aruba and what type of tourists were they? This research used a classification method used by Schep et al. (2016) and by Slijkerman et al. (2020) for a study in Bonaire. The same classifications were used since the same tourists visit Bonaire and Aruba. However, the underwater category was excluded and party/all-inclusive was added (Figure 4-5). The underwater wildlife were included in the 'Wildlife' category. Images of people diving without a fish were included in the 'Seascape' category. This was done because this thesis was written on behalf of Wageningen and Environmental Research which appointed that in Aruba more people go to parties and stay at resorts instead of diving compared to Bonaire. While categorizing there were not that many images taken while partying, but there were images of the all-inclusive resorts. However, still, the number of images compared to the other categories is only 6% of the total amount (Figure 4-3).

### *6.1.5 C-squares grid*

There were several ways in which the geosocial data could have been displayed on the map. This spatial resolution was thus more accurate than the square kilometer according to Slijkerman et al (2016). However, other ways of displaying the geosocial data could also have been investigated, see section 3.5.4. For example, hexagonal grids are better for visual appearance and division. Most studies would rather use circles however, these do not connect well (Chen et al., 2016). As explained in section 3.5.4. the c-squares grid is easy to aggregate and therefore this grid was chosen. It could have been explored if other visualizations would have led to other or better visualizations.

### *6.1.6 Ethical and privacy*

Ethical considerations were important to consider when using UGC from social media. This research aimed to collect the information that was needed and anonymize the information. Controversies about how social media is used in certain research have led to more implications and restrictions regarding the use and access of geosocial data. Also, public opinion about the use and access has led to several protection laws which also have implications for further research opportunities (Da Mota & Pickering, 2020). The uniqueness of UGC published by the user was something that also had to be considered. Some were put together in the same dataset of Aruba, but this image was due to human error made somewhere else. Therefore, with the pre-processing of the geosocial data, the researcher eliminated those incorrect geotagged data (Wong et al., 2017).

This research was carried out by one person and therefore the collected images from social media were also categorized by one person. This led to decreasing the total number of images due to time limitations but it also resulted in a biased view since only one person's perception is considered. Also, technical challenges because the researcher was working with APIs and a programming language for the first time needed to be considered.

However, with these increasing numbers, privacy has also been a bigger concern. Within this research, the metadata of the Flickr images has also been extracted. The metadata contained personal information about the users of Flickr. Even if it was the user's own choice to publish the images it was not necessarily agreed upon that their images could be used for academic purposes. With the metadata of the images, the researcher could identify for example the origin of the photographer and more information. In this research, this has been considered with care and therefore the geosocial data was anonymized. The metadata was gathered by creating pre-built tools, or by writing a program to request the API (Gerber & Lynch, 2017).

## 6.2 Conclusion

This research has contributed to the enhancement of strategic decision-making in the realm of tourism. Finding out which categories of tourists visit Aruba where and how to extract geosocial data from the internet including the metadata is an interesting source of information. Metadata includes information about when, where, and what tourists visit Aruba.

As mentioned, the research was part of the TRUPIAL program and focused on the tourism effect on nature and people. The usage of geosocial data contributed to mapping the tourists and finding out the tourist's preferences and interests. These are categorized based on the environments the tourists visited. Therefore, this research provided insights into the perceptions of tourists which in turn can be used by destination marketing organizations and policy makers to act upon these findings. For Aruba the use of geosocial data to map tourism was not done yet before, therefore this also contributed to the scientific literature.

The objective of this research was to explore both the potential as well as the constraints associated with collecting geosocial data to estimate the geographical patterns of tourist activity, to facilitate the advancement of sustainable tourism initiatives.

The result of this research explored the different options on how to map tourism, what are the methods and technologies available? How to collect and analyze geosocial data? What are the limitations and challenges? And eventually, how can these findings be used in practice?

There are several ways of analyzing geosocial data: spatiotemporal analysis, content analysis, and network analysis. For this research a visual content analysis was used to categorize the images and the spatial-temporal analysis was used to analyse and map the geosocial data.

There are also three collecting methods and technologies available for analyzing geosocial data in the context of tourism mapping. Within this research, the focus is put on methods for scraping images from social media. You can collect the images manually, by making screenshots. In addition, there are two automated collecting methods frequently used: web scraping and Application Programming Interface (API). Web scraping is used a lot but could be against the Terms of Service of the social media platform. In this research, the API method is chosen because this is a way of gathering geosocial data while adhering to the Terms of Service.

The relevant key sources of photographical geosocial data relevant to tourism interest in Aruba that were taken into account in this research are Instagram, TripAdvisor, Pinterest, and Flickr. Looking at the platform's relevance, users, and availability. Instagram had the most amount of users which also was a relatively young age group. The downside of Instagram was the limited accessibility to the geosocial data, there is no API available for research. Trip-advisor also has a lot of users, but the users mainly give reviews about the restaurants or places visited. For this research, the focus is put more on the natural environment in which the tourist is located. Also here the API has limitations with accessing the geosocial data due to a maximum number of reviews and places. This is also the case for Pinterest, and the content of the images is also about their travel wanderlust. Flickr does have a research-friendly API and the content of the images on Flickr (mostly nature-oriented) is therefore the best platform for this research.

Using social media in scientific research can be a valuable new source of information, which has its advantages and disadvantages, but due to its widespread nature can be of great added value in scientific research, especially for collecting research data, finding respondents, and testing hypotheses. Conversely, among tourists, the information extracted from geosocial data tends to be perceived as credible and independent. SMA has more options for in-depth analysis related to space, time, and various subgroups.

The results show a diverse spread of the tourist categories in Aruba. Coastal and Seascape however quite different in the spatial distribution pattern and temporal pattern. The Coastal map is more represented at the east coast, compared to the Seascape which is more at the west coast. Wildlife tourists mainly go to the beachside, for landscape tourists, the Arikok National Park and Rock Formation are mainly the most visited landscape. The Party tourists mainly go a lot to the urban area of Aruba, however, this category was not well represented in this study to do a really good analysis of it. This also comes forward when comparing the maps with the maps drawn by the tourism agencies.

The research outcomes have contributed to Aruba's tourism industry in different ways according to tourism agencies. One side on the environmental impact of tourism with wildlife conservation and clean-up efforts. But also with gaining more insights into the tourism activity, informing marketing strategies, and enhancing tourist experiences. With Aruba having tourism as the primary source of income, mapping tourism is crucial.

In conclusion, the findings of this study have offered a deeper understanding of using geosocial data to map the spatial distribution of tourism in Aruba. With an envisaged manner of choosing the right geosocial data and platform. Eventually had led to Mapping where several types of tourists' environmental sides. Secondly, as mentioned above, this research is also carried out to see if the method used for Bonaire is transferrable to Aruba. This is a transferrable method, as the article of Slijkerman et al (2020) has played a big role in this research. A side note here is that this research contributes to a small part of the TRUPIAL program, and therefore other methods and research should be looked at as well.

## 6.3 Recommendations

The significant potential of geosocial data for knowledge discovery comes forward with the several methods outlined and the literature mentioned in this research. Engaging with the geosocial data and analyzing the outcomes sparked numerous conceptual insights. Consequently, this section offers several suggestions for future research endeavors.

The geosocial data retrieved from Flickr could be enlarged by adding geosocial data from other platforms. Also, the results could be compared with each other. Section 5.3 showed that the representation of the tourists of Aruba is not representative of the tourists that visit Aruba. Therefore to add more value maybe when adding other geosocial data a better representation can be met. Comparing several techniques to gather geosocial data provided more comparative strength to the research. Combining surveys with geosocial data has led to fewer biases by showing the weaknesses and strengths of the methods used (Lin et al., 2021). A side note here is that geosocial data is not freely available and asks for time, skills, and maybe even money to collect. However, some respondents did mention this, and therefore future research could be looking into the options.

The Negative impacts of tourism are overcrowding, noise pollution, and littering (Goliath-Ludic & Yekela, 2020). While conducting the research the possibility of linking the geosocial data with these sorts of data was also thought of. For example, the OpenLitterMap was looked at. With this tool, people can make an image of litter and upload it. The map shows the places where all the people have taken these images, and thus where litter is found. However, in Aruba, zero images were taken. Maybe in the future governments or other nature conservation organizations can promote the OpenLitterMap and later compare the results of this study to the OpenLitterMap. This is also mentioned by one of the respondents who stated that there is a lack of trash disposal around the island. By adding such an initiative this problem can come more to the attention and gain public support for the problem of littering.

Maps could be added per period to see if there is a spatial difference in tourists. For policy, this would add value to maybe per season looking closer at certain places, and making sure that the environment is prepared for the tourists to come ahead. This is also mentioned by one of the respondents in this research.

For example, in the summer the Arikok National Park is more frequently visited compared to other hotspots. You could also look at the geosocial data more individually by adding a network analysis of where tourists go in the park. The policy or nature conservatism can keep more watch on specific areas of the park, such as patrolling or walking around to protect the park.

This research, conducted by Wageningen Environmental Research as part of the TRUPIAL program, presents findings already integrated into reports by the study's commissioner, Peter Verweij. A recommendation is to compare these results with those of Slijkerman et al. (2020) to assess differences relative to other studies on tourism in Aruba and Bonaire. As indicated by input from tourism agencies, the study's outcomes are relevant for environmental monitoring and tourism management. Collaboration between Aruba and Bonaire could leverage shared challenges to mutual benefit.

## 6.4 Final words

Geosocial data can provide valuable insights into the spatial division of tourists, however, some limitations should be mentioned. Tourism relies heavily on images, both in marketing to attract potential tourists and also increasingly on the images tourists take on their journeys themselves. However, this only shows the beautiful side of their vacations. Besides, the age group of Flickr only represents a small part of all tourists that visit Aruba. The results demonstrate that geosocial data can offer significant insight into the spatiotemporal patterns of tourists, even in the face of geosocial data flaws. Yet, conducting surveys or interviews would be more time-consuming. Furthermore, the tourism agencies believe that the findings are a useful supplement to conventional tourist research. With the geosocial data now gathered, endless analysis options can be utilized.

# Literature


Acker, A., & Kreisberg, A. (2020). Social media data archives in an API-driven world. *Archival Science*, *20*, 105-123.

Ali, S. A., & Zafar, M. W. (2022). API GATEWAY ARCHITECTURE EXPLAINED. *INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY*, *6*(4), 54-98.

Araujo, T., Lock, I., & van de Velde, B. (2020). Automated visual content analysis (AVCA) in communication research: A protocol for large scale image classification with pre-trained computer vision models. *Communication Methods and Measures*, *14*(4), 239-265.

Axelsson, B., Holmer, D., Ahrenberg, L., & Jönsson, A. (2020). Studying Emerging New Contexts for Museum Digitisations on Pinterest. In *CLARIN Annual Conference* (pp. 24-36).

Bahrololoum, A., & Nezamabadi-pour, H. (2017). A multi-expert based framework for automatic image annotation. *Pattern Recognition*, *61*, 169-184.

Barros, C., Moya-Gómez, B., & Gutiérrez, J. (2020). Using geotagged photographs and GPS tracks from social networks to analyse visitor behaviour in national parks. *Current Issues in Tourism*, *23*(10), 1291-1310.

Batabyal, D., Rawal, Y. S., Chatterjee, P., & Goswami, S. (2023). Temple service quality model for future pilgrimages in Indian sub-continent through user-generated contents (UGC) in TripAdvisor. *International Journal of Tourism Anthropology*, *9*(2), 113-135.

Becken, S., Stantic, B., Chen, J., Alaei, A. R., & Connolly, R. M. (2017). Monitoring the environment and human sentiment on the Great Barrier Reef: Assessing the potential of collective sensing. *Journal of environmental management*, *203*, 87-97.

Broz, M. (2022). Flickr Statistics, User Count, & Facts0.. (September 2023). Retrieved from

Bruns, A. (2019). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, *22*(11), 1544-1566.

Calcagni, F., Amorim Maia, A. T., Connolly, J. J. T., & Langemeyer, J. (2019). Digital co-construction of relational values: Understanding the role of social media for sustainability. *Sustainability Science*, *14*(5), 1309-1321.

Callegaro, M., & Yang, Y. (2018). The role of surveys in the era of "big data". *The Palgrave handbook of survey research*, 175-192.

CBS (2021, 5 December). Tourism In Aruba 2009-2021. Retrieved from https://cbs.aw/wp/index.php/2021/12/15/tourism-in-aruba-2018/

Chaudhary, P., D'Aronco, S., Moy de Vitry, M., Leitão, J. P., & Wegner, J. D. (2019). Flood-water level estimation from social media images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *4*(2/W5), 5-12.

Che, D., Safran, M., & Peng, Z. (2013). From big data to big data mining: challenges, issues, and opportunities. In *Database Systems for Advanced Applications: 18th International Conference, DASFAA 2013, International Workshops: BDMA, SNSM, SeCoP, Wuhan, China, April 22-25, 2013. Proceedings 18* (pp. 1-15). Springer Berlin Heidelberg.

Chen, J., Becken, S., & Stantic, B. (2022). Assessing destination satisfaction by social media: An innovative approach using Importance-Performance Analysis. *Annals of Tourism Research*, *93*, 103371.

Chen, S., Chen, S., Wang, Z., Liang, J., Yuan, X., Cao, N., & Wu, Y. (2016, October). D-Map: Visual analysis of ego-centric information diffusion patterns in social media. In *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 41-50). IEEE.

Chua, A., Servillo, L., Marcheggiani, E., & Moere, A. V. (2016). Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management*, *57*, 295-310.

Cole, S., & Razak, V. (2004). A framework for sustainable tourism in Aruba. *Prepared for The National Tourism Council of Aruba and The Minister of Tourism and Transportation, December*.

Croes, R. R. & Vanegas, M. (2003). Growth, development and tourism in a small economy: Evidence from Aruba. *International journal of tourism research*, *5*(5), 315-330.

Cui, N., Malleson, N., Houlden, V., & Comber, A. (2021). Using VGI and social media data to understand urban green space: a narrative literature review. *ISPRS International Journal of Geo-Information*, *10*(7), 425.

Da Mota, V. T., & Pickering, C. (2020). Using social media to assess nature-based tourism: Current research and future trends. *Journal of Outdoor Recreation and Tourism*, *30*, 100295.

Dagher, R. (2019). Policy space under a constraining combination–open economies, austerity and small island states. *Third World Quarterly*, *40*(6), 1040-1063.

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of big data*, *6*(1), 1-25.

Deng, N., Liu, J., Dai, Y., & Li, H. (2019). Different cultures, different photos: A comparison of Shanghai's pictorial destination image between East and West. *Tourism Management Perspectives*, *30*, 182-192.

Deng, N., Zhong, L., & Li, H. (2018). Perception of travel destination image based on user generated photograph metadata: The case of Beijing. Tourism Tribune, 33(1), 53–62.

Dewi, L. C., & Chandra, A. (2019). Social media web scraping using social media developers API and regex. *Procedia Computer Science*, *157*, 444-449.

Di Minin, E., Fink, C., Hausmann, A., Kremer, J., & Kulkarni, R. (2021). How to address data privacy concerns when using social media data in conservation science. *Conservation Biology*, *35*(2), 437-446.

Ding, K., Yang, M., & Luo, S. (2021). Mountain Landscape Preferences of Millennials Based on Social Media Data: A Case Study on Western Sichuan. *Land*, *10*(11), 1246.

DiPietro, R. B., & Peterson, R. (2017). Exploring cruise experiences, satisfaction, and loyalty: The case of Aruba as a small-island tourism economy. *International Journal of Hospitality & Tourism Administration*, *18*(1), 41-60.

Egarter Vigl, L., Marsoner, T., Giombini, V., Pecher, C., Simion, H., Stemle, E., ... & Depellegrin, D. (2021). Harnessing artificial intelligence technology and social media data to support Cultural Ecosystem Service assessments. *People and Nature*, *3*(3), 673-685.

El Ali, A., Van Sas, S. N., & Nack, F. (2013, February). Photographer paths: sequence alignment of geotagged photos for exploration-based route planning. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 985-994).

Engel, C., Müller, E. F., & Weber, G. (2021, June). Tactile Heatmaps: A Novel Visualisation Technique for Data Analysis with Tactile Charts. In *The 14th PErvasive Technologies Related to Assistive Environments Conference* (pp. 16-25).

Fiesler, C., Beard, N., & Keegan, B. C. (2020, May). No robots, spiders, or scrapers: Legal and ethical regulation of data collection methods in social media terms of service. In *Proceedings of the international AAAI conference on web and social media* (Vol. 14, pp. 187-196).

Figueredo, M., Cacho, N., Thome, A., Cacho, A., Lopes, F., & Araujo, M. (2017, December). Using social media photos to identify tourism preferences in smart tourism destination. In *2017 IEEE international conference on big data (big data)* (pp. 4068-4073). IEEE.

Fruchard, B., Malacria, S., Casiez, G., & Huot, S. (2023, April). User Preference and Performance using Tagging and Browsing for Image Labeling. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-13).

Fyall, A., & Garrod, B. (2020). Destination management: a perspective article. *Tourism Review*, *75*(1), 165-169.

García-Palomares, J. C., Gutiérrez, J., & Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography, 63*, 408-417.

Ghermandi, A., & Sinclair, M. (2019). Passive crowdsourcing of social media in environmental research: A systematic map. *Global environmental change*, *55*, 36-47.

Ghermandi, A., Camacho-Valdez, V., & Trejo-Espinosa, H. (2020). Social media-based analysis of cultural ecosystem services and heritage tourism in a coastal region of Mexico. *Tourism Management*, *77*, 104002.

Goede, M. (2021). COVID-19 and Change. *Journal of Applied Business and Economics, 23*(3).

Goldspiel, H., Barr, B., Badding, J., & Kuehn, D. (2023). Snapshots of Nature-Based Recreation Across Rural Landscapes: Insights from Geotagged Photographs in the Northeastern United States. *Environmental Management*, *71*(2), 234-248.

Goliath-Ludic, K., & Yekela, S. (2020). Resident's perception of the environmental impact of tourism: A case study of the bawa community in Butterworth, South Africa. *Geo Journal of Tourism and Geosites*, *33*, 1527-1531.

Gretzel, U. (2018). Tourism and social media. *The Sage handbook of tourism management*, *2*, 415-432.

Gretzel, U. (2021). Dreaming about travel: A Pinterest netnography. In *Information and Communication Technologies in Tourism 2021: Proceedings of the ENTER 2021 eTourism Conference, January 19–22, 2021* (pp. 256-268). Springer International Publishing.

Gröbe, M., & Burghardt, D. (2020). Micro diagrams: visualization of categorical point data from location-based social media. *Cartography and Geographic Information Science*, *47*(4), 305-320.

Gunaratne, K., Coomes, E. A., & Haghbayan, H. (2019). Temporal trends in anti-vaccine discourse on Twitter. *Vaccine*, *37*(35), 4867-4871.

Güven, E. Y., Boyacı, A., Sarıtemur, F. N., Türk, Z., Sütçü, G., & Turna, Ö. C. (2023). ICON: Instagram Profile Classification Using Image and Natural Language Processing Methods. *IEEE Transactions on Computational Social Systems*.

Hale, B. W. (2018). Mapping potential environmental impacts from tourists using data from social media: A case study in the Westfjords of Iceland. *Environmental management*, *62*(3), 446-457.

Hamstead, Z. A., Fisher, D., Ilieva, R. T., Wood, S. A., McPhearson, T., & Kremer, P. (2018). Geolocated social media as a rapid indicator of park visitation and equitable park access. *Computers, Environment and Urban Systems*, *72*, 38-50.

Hasnat, M. M., & Hasan, S. (2018). Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. *Transportation Research Part C: Emerging Technologies*, *96*, 38-54.

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters*, *11*(1), e12343.

Hassink, W. Kort, R. E. de. & Ridderstraat, J. 2015. De economische consequenties van de verdwijning van Natalee Holloway. Me JUdice economen in debat

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

Heikinheimo, V., Di Minin, E., Tenkanen, H., Hausmann, A., Erkkonen, J., & Toivonen, T. (2017). User-generated geographic information for visitor monitoring in a national park: A comparison of social media data and visitor survey. *ISPRS International Journal of Geo-Information*, *6*(3), 85.

Hidalgo, N. S. C. (2021). Terms of Service in Social Media, Artificial Intelligence and the Third-Party Doctrine. *Rev. Juridica U. Inter. PR*, *56*, 661.

Hirota, M., Shirai, M., Ishikawa, H., & Yokoyama, S. (2014, June). Detecting relations of hotspots using geo-tagged photographs in social media sites. In *Proceedings of Workshop on Managing and Mining Enriched Geo-Spatial Data* (pp. 1-6).

Igartua, M. A., Mendoza, F. A., Redondo, R. P. D., Vicente, M. I. M., Forné, J., Campo, C., ... & Rico-Novella, F. J. (2020). INRISCO: INcident monitoRing in smart communities. *IEEE access*, *8*, 72435-72460.

Jailani, Z. F., Verweij, P., Van Der Wal, J. T., & Van Lammeren, R. (2021, October). A Machine Learning Approach to Study Tourist Interests and Predict Tourism Demand on Bonaire Island from Social Media Data

Jenkins, O. (2003). Photography and travel brochures: The circle of representation. *Tourism geographies*, *5*(3), 305-328.

Jimenez, F. S. (2012). The 2.0 tourist as recipient of touristic promotion: linguistic strategies and importance of its study. *PASOS: Revista de Turismo y Patrimonio Cultural*, *10*(4), 143-153.

Jindal, K., & Aron, R. (2021). A novel visual-textual sentiment analysis framework for social media data. *Cognitive Computation*, *13*, 1433-1450.

Kaiser, C., Ahuvia, A., Rauschnabel, P. A., & Wimble, M. (2020). Social media monitoring: What can marketers learn from Facebook brand photos?. *Journal of business research*, *117*, 707-717.

Karaceper, E. I. & Ramadanoglu, E. (2023). Evaluation of Foreign Visitors' Turkish Cuisine Experiences on Tripadvisor, an Online Platform: The Case of İstanbul-Beyoğlu. *Journal of Tourism & Gastronomy Studies*, *11*(3), 2223-2237.

Kemp, S. (2023). *Pinterest Users, Stats, Data, Trends, and More — DataReportal – Global Digital Insights*. DataReportal – Global Digital Insights. Retrieved from, https://datareportal.com/essential-pinterest-stats

Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia tools and applications*, *82*(3), 3713-3744.

Kim, G. S., Chun, J., Kim, Y., & Kim, C. K. (2021). Coastal tourism spatial planning at the regional unit: Identifying coastal tourism hotspots based on social media data. *ISPRS International Journal of Geo-Information*, *10*(3), 167.

Kim, N. H., Kim, J. M., Park, D. M., Ji, S. R., & Kim, J. W. (2022). Analysis of depression in social media texts through the Patient Health Questionnaire-9 and natural language processing. *Digital Health*, *8*, 20552076221114204.

Kort, R. E. de. (2017). Forecasting tourism demand through search queries and machine learning. *IFC Bulletins chapters*, *44*.

Kovács, Z., Vida, G., Elekes, Á., & Kovalcsik, T. (2021). Combining social media and mobile positioning data in the analysis of tourist flows: A case study from Szeged, Hungary. *Sustainability*, *13*(5), 2926.

Krumm, J., & Horvitz, E. (2015, November). Eyewitness: Identifying local events via space-time signals in twitter feeds. In *Proceedings of the 23rd sigspatial international conference on advances in geographic information systems* (pp. 1-10).

Kumar, P., Ofli, F., Imran, M., & Castillo, C. (2020). Detection of disaster-affected cultural heritage sites from social media images using deep learning techniques. *Journal on Computing and Cultural Heritage (JOCCH)*, *13*(3), 1-31.

Latorre-Martínez, M. P., Iñíguez-Berrozpe, T., & Plumed-Lasarte, M. (2014). Image-focused social media for a market analysis of tourism consumption. *International Journal of Technology Management*, *64*(1), 17-30.

Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., ... & Khalil, T. (2019). Content-based image retrieval and feature extraction: a comprehensive review. *Mathematical problems in engineering*, *2019*.

Lee, H., Seo, B., Koellner, T., & Lautenbach, S. (2019). Mapping cultural ecosystem services 2.0–Potential and shortcomings from unlabeled crowd sourced images. *Ecological Indicators*, *96*, 505-515.

Lee, J. Y., & Tsou, M. H. (2018). Mapping spatiotemporal tourist behaviors and hotspots through location-based photo-sharing service (Flickr) data. In *Progress in Location Based Services 2018 14* (pp. 315-334). Springer International Publishing.

Lin, H., & Chiang, Y. Y. (2018). SRC: automatic extraction of phrase-level map labels from historical maps. *SIGSPATIAL Special*, *9*(3), 14-15.

Li, H., Zhang, L., & Hsu, C. H. (2023). Research on user-generated photos in tourism and hospitality: A systematic review and way forward. *Tourism Management*, *96*, 104714.

Lim, K. H. (2015, May). Recommending tours and places-of-interest based on user interests from geo-tagged photos. In *Proceedings of the 2015 ACM SIGMOD on PhD Symposium* (pp. 33-38).

Lin, M. S., Liang, Y., Xue, J. X., Pan, B., & Schroeder, A. (2021). Destination image through social media analytics and survey method. *International Journal of Contemporary Hospitality Management*, *33*(6), 2219-2238.

Lo, S.I., McKercher, B., Lo, A., Cheung, C. and Law, R. (2011) 'Tourism and online photography', Tourism Management, *32*(4), 725–731.

Lojo, A., Li, M., & Xu, H. (2020). Online tourism destination image: Components, information sources, and incongruence. *Journal of Travel & Tourism Marketing*, *37*(4), 495-509.

Loncomilla, P., Samtani, P., & Ruiz-del-Solar, J. (2022). Detecting rocks in challenging mining environments using convolutional neural networks and ellipses as an alternative to bounding boxes. *Expert Systems with Applications*, *194*, 116537.

Lu, C., Laublet, P., & Stankovic, M. (2016). Travel attractions recommendation with knowledge graphs. In *Knowledge Engineering and Knowledge Management: 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings 20* (pp. 416-431). Springer International Publishing.

Ludmer, L. H. (2011). *Cruising the Southern and Western Caribbean: A Guide to the Ships and the Ports of Call*. Hunter Publishing, Inc.

Lusiana, F., & Beng, J. T. (2020, July). Grouping of Tourism Objects Using Geotagged Photo with Hierarchical Clustering Method in Bantul and Sleman. In *IOP Conference Series: Materials Science and Engineering* (Vol. 852, No. 1, p. 012166). IOP Publishing.

MacDonald, S. B. (2020). The Dutch Caribbean. *Handbook of Caribbean Economies*, 442-456.

Maeda, T. N., Yoshida, M., Toriumi, F., & Ohashi, H. (2018). Extraction of tourist destinations and comparative analysis of preferences between foreign tourists and domestic tourists on the basis of geotagged social media data. *ISPRS International Journal of Geo-Information*, *7*(3), 99.

Mahdavi-Amiri, A., Alderson, T., & Samavati, F. (2015). A survey of digital earth. *Computers & Graphics*, *53*, 95-117.

Majid, A., Chen, L., Chen, G., Mirza, H. T., Hussain, I., & Woodward, J. (2013). A context-aware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science*, *27*(4), 662-684.

Mangachena, J. R., Geerts, S., & Pickering, C. M. (2023). Spatial and temporal patterns in wildlife tourism encounters and how people feel about them based on social media data from South Africa. *Journal of Outdoor Recreation and Tourism*, 100642.

Manoharan, D. S., & Sathesh, A. (2020). Geospatial and social media analytics for emotion analysis of theme park visitors using text mining and gis. *Journal of Information Technology and Digital World*, *2*(2), 100-107.

Manovich, L. (2019). The aesthetic society: Instagram as a life form. *Data Publics, Routledge, forthcoming*, *2*, 19.

Marapengopi, S. V. (2017). *Reusability of volunteered geographic information supported by Semantic Web technologies: a case study for environmental applications* (Master's thesis).

Marine-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, *4*(3), 162-172.

McMullen, M. (2024). Let's see the whole picture: a multi-platform visual analysis of historic Mission San José on Instagram, Pinterest, Flickr, and TripAdvisor. *Journal of Heritage Tourism*, 1-15.

Memon, I., Chen, L., Majid, A., Lv, M., Hussain, I., & Chen, G. (2015). Travel recommendation using geo-tagged photos in social media for tourist. *Wireless Personal Communications*, *80*, 1347-1362.

Menk, A., Sebastia, L., & Ferreira, R. (2019). Recommendation systems for tourism based on social networks: A survey. *arXiv preprint arXiv:1903.12099*.

Miller, Z. D., Taff, B. D., Newman, P., & Lawhon, B. (2019). A Proposed Research Agenda on Social Media's Role in Visitor Use and Experience in Parks and Protected Areas. *Journal of Park and Recreation Administration*, *37*(3), 134-143.

Mirzaalian, F., & Halpenny, E. (2019). Social media analytics in hospitality and tourism: A systematic literature review and future trends. *Journal of Hospitality and Tourism Technology*, *10*(4), 764-790.

Mirzaalian, F., & Halpenny, E. (2021). Exploring destination loyalty: Application of social media analytics in a nature-based tourism setting. *Journal of Destination Marketing & Management*, *20*, 100598.

Morstatter, F., & Liu, H. (2017). Discovering, assessing, and mitigating data bias in social media. *Online Social Networks and Media*, *1*, 1-13.

Muñoz, L., Hausner, V., Brown, G., Runge, C., & Fauchald, P. (2019). Identifying spatial overlap in the values of locals, domestic-and international tourists to protected areas. *Tourism Management*, *71*, 259-271.

Obembe, D., Kolade, O., Obembe, F., Owoseni, A., & Mafimisebi, O. (2021). Covid-19 and the tourism industry: An early stage sentiment analysis of the impact of social media and stakeholder communication. *International Journal of Information Management Data Insights*, *1*(2), 100040.

Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C., & Plieninger, T. (2018). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. *Ecological Indicators*, *94*, 74-86.

Owuor, I., & Hochmair, H. H. (2020). An overview of social media apps and their potential role in geospatial research. *ISPRS International Journal of Geo-Information*, *9*(9), 526.

Park, S. B., Ok, C. M., & Chae, B. K. (2016). Using Twitter data for cruise tourism marketing and research. *Journal of Travel & Tourism Marketing*, *33*(6), 885-898.

Pereira, M. D. C. D. (2023). *Using Flickr to identify and connect tourism Points of Interest: The case of Lisbon, Porto and Faro* (Doctoral dissertation).

Petzold, J., & Magnan, A. K. (2019). Climate change: thinking small islands beyond Small Island Developing States (SIDS). *Climatic Change*, *152*(1), 145-165.

Poorthuis, A., & Zook, M. (2015). Small stories in big data: Gaining insights from large spatial point pattern datasets. *Cityscape*, *17*(1), 151-160.

Rashidi, T. H., Abbasi, A., Maghrebi, M., Hasan, S., & Waller, T. S. (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, *75*, 197-211.

Rees, T. (2003). C-squares, a new spatial indexing system and its applicability to the description of oceanographic datasets. *Oceanography*, *16*(1), 11-19.

Reif, J., & Schmücker, D. (2020). Exploring new ways of visitor tracking using big data sources: Opportunities and limits of passive mobile data for tourism. *Journal of Destination Marketing & Management*, *18*, 100481.

Ridderstaat, J. (2022). Tourism development and nation-building: The case of Aruba. *Cultural heritage and the nation: Aruba, Bonaire and Curaçao in Caribbean perspective*.

Robinson, R., Rittenbruch, M., Foth, M., Filonik, D., & Viller, S. (2016). Street computing: Towards an integrated open data application programming interface (API) for cities. In *Street Computing* (pp. 1-23). Routledge.

Runge, C. A., Daigle, R. M., & Hausner, V. H. (2020). Quantifying tourism booms and the increasing footprint in the Arctic with social media data. *PloS one*, *15*(1), e0227189.

Sager, C., Janiesch, C., & Zschech, P. (2021). A survey of image labelling for computer vision applications. *Journal of Business Analytics*, *4*(2), 91-110.

Schep et al (2016). Nature on Bonaire and social media behaviour. Assessing the spatial distribution of tourism and recreation on Bonaire through social media

Seo, Y. D., & Cho, Y. S. (2021). Point of interest recommendations based on the anchoring effect in location-based social network services. *Expert Systems with Applications*, *164*, 114018.

Séraphin, H., Christodoulidi, F., & Gladkikh, T. (2020). Case Study 2: Broadly Engaging with Interaction Between Visitors and Locals—Towards Understanding Tourismphobia and Anti-tourism Movements. *Overtourism: Causes, Implications and Solutions*, 207-227.

Shin, D., He, S., Lee, G. M., Whinston, A. B., Cetintas, S., & Lee, K. C. (2020). *Enhancing social media analysis with visual data analytics: A deep learning approach* (pp. 1459-1492). SSRN.

Schmutz, P. P., Potter, A. E., & Arnold Modlin, E. (2017). Aruba, Bonaire, and Curaçao. *Landscapes and landforms of the Lesser Antilles*, 293-317.

Slijkerman, D., van der Wal, J. T., de Vries, P., & Verweij, P. (2020). *Tracking digital footprints in Bonaire's landscapes: spatial distribution and characterization of tourists on Bonaire using social media* (No. C052/20). Wageningen Marine Research.

Spalding, M., Burke, L., Wood, S. A., Ashpole, J., Hutchison, J., & Zu Ermgassen, P. (2017). Mapping the global value and distribution of coral reef tourism. *Marine Policy*, *82*, 104-113.

Stepchenkova, S., & Zhan, F. (2013). Visual destination images of Peru: Comparative content analysis of DMO and user-generated photography. *Tourism management*, *36*, 590-601.

Straathof, M., van Stipriaan Luïscius, A., & Dibbits, H. (2019). Constructing a Dutch gaze: imaging (post) colonial Aruba and Bonaire.

Stasi, M. L. (2019). Social media platforms and content exposure: How to restore users' control. *Competition and Regulation in Network Industries*, *20*(1), 86-110.

Suau Jiménez, F. (2012). The tourist 2.0 as a receiver of the tourist promotion: linguistic strategies and importance of his study/Steps. *Journal of tourism and cultural heritage*, (10), 4.

Sultan, M. T., Sharmin, F., Badulescu, A., Stiubea, E., & Xue, K. (2020). Travelers' responsible environmental behavior towards sustainable coastal tourism: An empirical investigation on social media user-generated content. *Sustainability*, *13*(1), 56.

Sun, Y., Fan, H., Bakillah, M., & Zipf, A. (2015). Road-based travel recommendation using geo-tagged images. *Computers, Environment and Urban Systems*, *53*, 110-122.

Taylor, D. (2018). Tourism and development in the Senian context: does it help or hurt SIDS? The case of Aruba. *Journal of Tourism and Hospitality Management*, *6*(3), 100-108.

Tejaswini, V., Babu, K. S., & Sahoo, B. (2022). Depression Detection from Social Media Text Analysis using Natural Language Processing Techniques and Hybrid Deep Learning Model. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., & Toivonen, T. (2017). Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific reports*, *7*(1), 17615.

Tieskens, K. F., Van Zanten, B. T., Schulp, C. J., & Verburg, P. H. (2018). Aesthetic appreciation of the cultural landscape through social media: An analysis of revealed preference in the Dutch river landscape. *Landscape and urban planning*, *177*, 128-137.

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., ... & Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*, *233*, 298-315.

Van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., & Verburg, P. H. (2016). Continental-scale quantification of landscape values using social media data. *Proceedings of the National Academy of Sciences*, *113*(46), 12974-12979.

Vassey, J., Valente, T., Barker, J., Stanton, C., Li, D., Laestadius, L., ... & Unger, J. B. (2023). E-cigarette brands and social media influencers on Instagram: a social network analysis. *Tobacco Control*, *32*(e2), e184-e191.

Viñán-Ludeña, M. S., & de Campos, L. M. (2022). Discovering a tourism destination with social media data: BERT-based sentiment analysis. *Journal of Hospitality and Tourism Technology*, *13*(5), 907-921.

Vu, H. Q., Law, R., & Li, G. (2019). Breach of traveller privacy in location-based social media. *Current Issues in Tourism*, *22*(15), 1825-1840.

Wang, L., Ai, T., Burghardt, D., Shen, Y., & Yang, M. (2023). A hexagon-based method for polygon generalization using morphological operators. *International Journal of Geographical Information Science*, *37*(1), 88-117.

Wang, Y., Chen, J., Huang, Q., Xia, X., & Jiang, B. (2023). Deep learning-based open API recommendation for Mashup development. *Science China Information Sciences*, *66*(7), 1-18.

Wibowo, T. W., Rofi'i, A., & Sulistyaningrum, N. A. (2019, November). Density analysis of Flickr data as a proxy to reveal the intensity of tourism activity in Borobudur. In *Sixth Geoinformation Science Symposium* (Vol. 11311, pp. 122-129). SPIE.

WTTC. (2022). Travel and Tourism Economic Impact 2022, World Travel & Tourism Council: London.

Wong, E., Law, R., & Li, G. (2017). Reviewing geotagging research in tourism. In *Information and Communication Technologies in Tourism 2017: Proceedings of the International Conference in Rome, Italy, January 24-26, 2017* (pp. 43-58). Springer International Publishing.

Wood, S. A., Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature- based tourism and recreation. *Scientific reports, 3*(1), 2976.

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51-65.

Xiaoyang, W. (2019). Analysis on the Communication Path of Image Symbols of Tourist Destinations in the New Media Era. In *AISALLC 2019: 1st Asia international Symposium on Arts, literature, Language and culture* (pp. 27-31). Guangdong University of Finance & Economics Guangzhou, Guangdong, China.

Yan, Y., Eckle, M., Kuo, C. L., Herfort, B., Fan, H., & Zipf, A. (2017). Monitoring and assessing post-disaster tourism recovery using geotagged social media data. *ISPRS International Journal of Geo-Information*, *6*(5), 144.

Zhang, D., Islam, M. M., & Lu, G. (2012). A review on automatic image annotation techniques. *Pattern Recognition*, *45*(1), 346-362.

Zhang, H., Huang, R., Zhang, Y., & Buhalis, D. (2022). Cultural ecosystem services evaluation using geolocated social media data: A review. *Tourism Geographies, 24*(4-5), 646-668.

Zhang, H., van Berkel, D., Howe, P. D., Miller, Z. D., & Smith, J. W. (2021). Using social media to measure and map visitation to public lands in Utah. *Applied Geography*, *128*, 102389.

Zhang, K., Chen, Y., & Lin, Z. (2020). Mapping destination images and behavioral patterns from user-generated photos: A computer vision approach. *Asia Pacific Journal of Tourism Research*, *25*(11), 1199-1214.

Zhuawu, M. C., Trade, E. A. M., & Secretariat, C. (2021). Identifying alternative economic development strategies for diverse SIDS.

# Appendices

## Appendix A Flickr extraction code for images and location

```
install.packages("FlickrAPI")

library(FlickrAPI)


# Set the API key for a single session

setFlickrAPIKey(api_key = "")


getPhotoSearch <- function(bbox, min_upload_date, max_upload_date) {

  page <- 1

  result <- NULL

  ## loop all pages

  repeat({

    search_url <- sprintf(

"https://www.flickr.com/services/rest/?method=flickr.photos.search&format=rest&api_key=%s&bbox=%s&min_upload_date=%s&max_upload_date=%s&page=%s&extras=%s",

      key,

      paste(bbox, collapse = ","),

      as.numeric(as.POSIXct(as.Date(min_upload_date))),

      as.numeric(as.POSIXct(as.Date(max_upload_date))),

      page,

      "url_o,url_m,url_c"

    )

    search_result <- curl::curl_fetch_memory(search_url)

    page_result   <- XML::xmlToList(rawToChar(search_result$content))[[1]]

    npages        <- as.numeric(page_result$.attrs["pages"])

    page_result   <- page_result[names(page_result) != ".attrs"]

    page_result   <- as.data.frame(do.call(dplyr::bind_rows, page_result))

    rownames(page_result) <- NULL

    result <- dplyr::bind_rows(result, page_result)

    if (page == npages) break
```

```
    page <- page + 1
  })
  result
}

result <- lapply(2012:2022, function(year) {
  getPhotoSearch(bbox = c(-70.0665, 12.4111, -69.8737, 12.6325),
                 sprintf("%s-01-01", year),
                 sprintf("%s-01-01", year + 1))
})

result <- do.call(dplyr::bind_rows, result)
write.csv(result, "C:/Users/greet/OneDrive - Universiteit Utrecht/GIMA
Thesis/rfoto's/ARUBAFOTOS.csv", row.names = F)

install.packages("readr")
library (readr)
install.packages("dplyr")
library (dplyr)
library(jsonlite)
install.packages("httr")
library(httr)
install.packages("xml2")

# Read the csv-file to get the photo IDs
get_locationofphoto <- read.csv("result")
my_photo_ids <- get_locationofphoto$id

# Create an empty list to store the data frames
results_list <- list()

# Define the batch size
batch_size <- 100
```

```
# Calculate the number of batches needed
num_batches <- ceiling(length(my_photo_ids) / batch_size)

# Create an empty data frame to accumulate the results
results_df <- data.frame()

# Iterate over batches
for (batch in 1:num_batches) {
  # Define the range of rows for the current batch
  start_row <- (batch - 1) * batch_size + 1
  end_row <- min(batch * batch_size, length(my_photo_ids))

  # Iterate over rows in the current batch
  for (i in start_row:end_row) {
    photo_id <- strsplit(my_photo_ids[[i]], ',')[[1]][1]
    print(photo_id)
    search_url <- sprintf(

"https://www.flickr.com/services/rest/?method=flickr.photos.geo.getLocation&api_key=%s&photo_id
=%s&format=json&nojsoncallback=1",
      key,
      photo_id
    )
    response <- GET(search_url)

    # Extract the necessary information from the response and add it to a data frame
    response_data <- content(response, "parsed")

    if (!is.null(response_data$photo) && !is.null(response_data$photo$location)) {
      location <- data.frame(response_data$photo$location)
      results_list[[i]] <- location
    }
  }
```

```
  # Combine data frames from the current batch into a single data frame
  batch_df <- bind_rows(results_list)

  # Add the batch data to the results data frame
  results_df <- bind_rows(results_df, batch_df)

  # Clear the results list for the next batch
  results_list <- list()
}

# Write the combined data frame to a single csv-file
write.csv(results_df, "C:/Users/greet/OneDrive - Universiteit Utrecht/GIMA Thesis/rfoto's/ARUBAFOTOSLOCATIE.csv", row.names = FALSE)
```

# Appendix B Flickr extraction code extra photo(grapher) information

```
install.packages("FlickrAPI")

library(FlickrAPI)

library(readr)

# Set the API key for a single session

setFlickrAPIKey(api_key = "")

install = TRUE


secret <- ""


getPhotoSearch <- function(bbox, min_upload_date, max_upload_date) {

  page <- 1

  result <- NULL

  ## loop all pages

  repeat({

    search_url <- sprintf(

"https://www.flickr.com/services/rest/?method=flickr.photos.search&format=rest&api_key=%s&bbox
=%s&min_upload_date=%s&max_upload_date=%s&page=%s&extras=%s",

      key,

      paste(bbox, collapse = ","),

      as.numeric(as.POSIXct(as.Date(min_upload_date))),

      as.numeric(as.POSIXct(as.Date(max_upload_date))),

      page,

      "url_o,url_m,url_c"

    )

    search_result <- curl::curl_fetch_memory(search_url)

    page_result   <- XML::xmlToList(rawToChar(search_result$content))[[1]]

    npages        <- as.numeric(page_result$.attrs["pages"])

    page_result   <- page_result[names(page_result) != ".attrs"]

    page_result   <- as.data.frame(do.call(dplyr::bind_rows, page_result))

    rownames(page_result) <- NULL

    result <- dplyr::bind_rows(result, page_result)
```

```
    if (page == npages) break
    page <- page + 1
  })
  result
}

result <- lapply(2012:2022, function(year) {
  getPhotoSearch(bbox = c(-70.0665, 12.4111, -69.8737, 12.6325),
                 sprintf("%s-01-01", year),
                 sprintf("%s-01-01", year + 1))
})

result <- do.call(dplyr::bind_rows, result)
write.csv(result, "C:/Users/greet/OneDrive - Universiteit Utrecht/GIMA
Thesis/rfoto's/ARUBAFOTOS2.csv", row.names = F)

install.packages("readr")
library (readr)
install.packages("dplyr")
library (dplyr)
library(jsonlite)
install.packages("httr")
library(httr)
install.packages("xml2")

# Read the CSV file to get the photo IDs
get_locationofphoto <- read.csv("C:/thesis GIMA/rfoto's/arubaphotosnieuw3.csv", stringsAsFactors =
FALSE)

# Convert 'id' column to character
my_photo_ids <- as.character(get_locationofphoto$id)

# Define the batch size
```

```
batch_size <- 100

# Calculate the number of batches needed
num_batches <- ceiling(length(my_photo_ids) / batch_size)

# Create an empty data frame to accumulate the results
results_df <- data.frame()

# Iterate over batches
for (batch in 1:num_batches) {
  # Define the range of rows for the current batch
  start_row <- (batch - 1) * batch_size + 1
  end_row <- min(batch * batch_size, length(my_photo_ids))

  batch_result <- NULL
  # Iterate over rows in the current batch
  for (i in start_row:end_row) {
    photo_id <- strsplit(my_photo_ids[i], ',')[[1]][1]
    print(photo_id)
    search_url <- sprintf(

"https://www.flickr.com/services/rest/?method=flickr.photos.getInfo&api_key=8fa26d3c2e0ad46e767
ffabd238a8bf0&photo_id=%s&secret=%s&format=json&nojsoncallback=1",
      photo_id,
      secret
    )
    response <- GET(search_url)

    # Extract the necessary information from the response and add it to a data frame
    response_data <- content(response, "parsed")
    batch_result <- bind_rows(
      batch_result,
      as_tibble(lapply(response_data$photo, list))
```

```
    )

    # if (!is.null(response_data$photo) && !is.null(response_data$photo$location)) {
    #   location <- data.frame(response_data$photo$location)
    #   location$photo_id <- photo_id  # Add the 'photo_id' column
    #   results_list[[i]] <- location
    # }
  }

  #warnings()

  # Combine data frames from the current batch into a single data frame
  # batch_df <- bind_rows(results_list)

  # Add the batch data to the results data frame
  # results_df <- bind_rows(results_df, batch_df)
  results_df <- bind_rows(results_df, batch_result)

  # Clear the results list for the next batch
  results_list <- list()
}

results_df <- results_df %>%
  mutate(across(everything(), ~unlist(lapply(., paste0, collapse = "|"))))



writexl::write_xlsx(results_df, "C:/thesis GIMA/rfoto's/arubaphotosinfo_combined1710.xlsx")
```

## Appendix C 'PhotoCategorizer' code

```
import PySimpleGUI as sg
import os
import datetime
import requests

"""
    Simple Image Browser based on PySimpleGUI
    Adjusting to PhotoTagBonaire..
"""
# DONE: copy paste relevant code from SimpeImageViewer_v0.1.3
# a.o. Flickr and sqlite imports
# a.o. switch to sql3 database (instead of a geopackage.)
#
# 20231101: JTvdW, improving the current code to
# detect when all photos have been categorised and then
# switch mode from 'Categorise' to 'Review'
# Addition triggered by re-use of the code by Greetje Havermans (UU)

# In pct_mode = 'Review' the code works as expected with the following quirks.
# Active pct_mode = 'Review' is visible in the window. Above right upper
corner of photo.
# 1 The first photo always comes up as category 'Other', check and correct if
necessary!
# 2 The moving to before the first photo is not possible, however it may
require
# several clicks on Next or rolling the scroll wheel to get moving again.
# 3 UNTESTED what would happen when moving past the last photo.
# Expectation is similar to what happens at zero. Last photo is shown,
# returning may require several clicks on Prev or rolling the scroll wheel a
bit more.

from Flickr_administrative import Flick_APIkey as Fk
import sqlite3
from sqlite3 import Error
from PIL import Image
import PIL


def PTC_Init_Category_Table():
    # connection named conn already inclused in global code.
    cursor_PTCict = conn.cursor()
    #
    sql_string = """
                Create table if not exists
                photagcategory (
                id text PRIMARY KEY,
```

```
                    category integer NOT NULL,
                    categoriser text NOT NULL,
                    timestamp text NOT NULL
                    );
                 """
    # SQl-statement includes a check for existence of the table so this should
avoid
    # creating a fresh table at each start while ensuring that there is a
table to accept the categories
    try:
        cursor_PTCict.execute(sql_string)
        conn.commit()
    except Error as e:
        print(e)
    #
    # And next create a UNIQUE index if it does not exist.
    sql_string = """
                    CREATE UNIQUE INDEX idx_uniq_photagcategory ON
photagcategory (id);
                 """
    try:
        cursor_PTCict.execute(sql_string)
        conn.commit()
    except Error as e:
        print(e)


def Previous_Flickr_Picture(before_timestamp):
    """
    Function: Retrieves data from the database for the photagcategory-records
that precedes the timestamp
    passed to the function, and based on the ID continues to select that
record from photaglink
    and finally returns one dictionary much like Get_FLickr_Picture
    URL_dict = {'id': row['id'], 'image_URL': image_URL, 'image_file':
image_TMP_png}
    - extended with the already assigned category and
    - the picture_timestep
    offering the GUI the required data  to be able
    - show that picture
    - know it's timestamp (so that there is an option to go further back
    - show the currently assigned category.
    """
    cursor_PTC = conn.cursor()
    previousID = 0
    sqlStr_previous_photagcategory = (
        "select id, category, timestamp from photagcategory "
        "where timestamp <'" + before_timestamp + "'"
        "order by timestamp DESC limit 1;"
```

```
    )
    print(sqlStr_previous_photagcategory)
    cursor_PTC.execute(sqlStr_previous_photagcategory)
    fieldList = []
    for (
        field
    ) in (
        cursor_PTC.description
    ):  # list of tuples, so add [0] to get the text from the initial element
        fieldList.append(field[0])
    for values in cursor_PTC.fetchmany(size=1):
        # fetchmany(size = 1) works, fetchone() results in an error about a
'missing schema'.
        row = dict(zip(fieldList, values))
        print(row)
    previousID = row["id"]
    timestamp_current = row["timestamp"]
    category_current = row["category"]
    sqlStr_previous_photaglink = (
        "Select id, url_m, url_c, url_o from photaglink where id = '"
        + str(previousID)
        + "';"
    )
    print(sqlStr_previous_photaglink)
    cursor_PTC.execute(sqlStr_previous_photaglink)
    fieldList = []
    for (
        field
    ) in (
        cursor_PTC.description
    ):  # list of tuples, so add [0] to get the text from the initial element
        fieldList.append(field[0])
    for values in cursor_PTC.fetchmany(size=1):
        # fetchmany(size = 1) works, fetchone() results in an error about a
'missing schema'.
        row = dict(zip(fieldList, values))
        # print('url_m=', row['url_m'], 'url_c=', row['url_c'], 'url_o=',
row['url_o'])
        # image_URL = row['photo.urls.url.text']
        if row["url_m"] != "NA":
            image_URL = row["url_m"]
        elif row["url_c"] != "NA":
            image_URL = row["url_c"]
        elif row["url_o"] != "NA":
            image_URL = row["url_o"]
        else:
            image_URL = None
            continue
            # No URL available proceed to the next line!
```

```
        r = requests.get(image_URL, stream=True)
        r.raise_for_status()
        r.raw.decode_content = True
        # image_TMP_jpg = r'D:\GIS_data\PhotoTagBonaire\FlickrTemporary.jpg'
        image_TMP_jpg = os.path.join(work_dir, "FlickrTemporary.jpg")
        # format of file follows extension '.jpg'
        # Needs adjusting, PySimpleGUI stays closer to what tkInter can do, 
which is png not jpg!
        with open(image_TMP_jpg, "wb") as f:
            f.write(r.content)
        # format of file follows extension '.jpg'
        # Needs adjusting, PySimpleGUI stays closer to what tkInter can do, 
which is png not jpg!
        image_TMP_png = image_TMP_jpg[:-4] + ".png"
        image = Image.open(image_TMP_jpg)
        image = image.save(image_TMP_png, format="PNG")
        # Added code below to keep a copy of the picture for future reference. 
JTvdW, 20200226
        # Code added to safe-guard against Flickr going off-line, leaving us 
with no options to review the assigned categories
        # or allow other scientist to review our data and options. Intending 
not to have to use this, but purely a safe-guard.
        ARC_dir = "Flickr_archive"
        image_ARC_png = row["id"] + ".png"
        image_ARC_path = os.path.join(work_dir, ARC_dir, image_ARC_png)
        image_ARC = Image.open(image_TMP_png)
        if os.path.exists(image_ARC_path):
            pass
            # As the used file name based on ID is unique, there is no need to 
rewrite a file
            # if it already exists. Thus avoiding OSerror raised when 
overwriting an existing file.
        else:
            image_ARC = image_ARC.save(image_ARC_path, format="PNG")
        # JTvdW, 20200226 \
        r.close()

        URL_dict = {
            "id": previousID,
            "image_URL": image_URL,
            "image_file": image_TMP_png,
            "category": category_current,
            "time_stamp": timestamp_current,
        }
        # print(URL_dict, type(URL_dict))
    return URL_dict


def Retrieve_Flickr_Picture(id2retrieve, offset_j):
```

```
    """
    Function: Retrieves data from the database and returns one dictionary with
the
    required data for the GUI, to be able show the it and record the category.

    The picture is retrieved as a link/pointer to the 'locally' store picture
    """
    URL_dict = {}
    cursor_PTC = conn.cursor()
    # Ensuring that offset_j does not go negative!
    offset_j = max(0, offset_j)
    # preparing the sql-statements for an initial retrieval
    # a NEXT-picture or
    # a PREVious-picture.
    sql_string_init = "select id, category from photagcategory order by id
limit 1"
    # sql_string_next = f"select id, category from photagcategory where id >
'{str(id2retrieve)}' order by id limit 1"
    # sql_string_prev = f"select id, category from photagcategory where id <
'{str(id2retrieve)}' order by id limit 1"
    # The above statements would not proceed reliable to a next or previous
photo
    # Trying with a _move-statement that uses offset j (int) to achieve this.
    # No need for a _next/_previous variation provided j is updated in
advance.
    sql_string_move = f"select id, category from photagcategory order by id
limit 1 offset {str(offset_j)};"
    sql_string_first = (
        "select min(id) as id, category from photagcategory order by id limit
1"
    )
    sql_string_last = (
        "select max(id) as id, category from photagcategory order by id limit
1"
    )
    # Now figure out which to use
    if id2retrieve == 0 or id2retrieve is None:
        str_string2execute = sql_string_init
    # elif movement == "next" or movement is None or movement == 0:
    #     str_string2execute = sql_string_next
    else:
        # str_string2execute = sql_string_prev
        str_string2execute = sql_string_move
    cursor_PTC.execute(str_string2execute)
    # if cursor_PTC.lastrowid == 0:
    #     # No results available try something else: i.e. first or last
    #     if movement == "next" or movement is None or movement == 0:
    #         str_string2execute = sql_string_first
    #     else:
```

```
    #         str_string2execute = sql_string_last
    #     cursor_PTC.execute(str_string2execute)
    # Check for moving beyond first or last

    fieldList = []
    for (
        field
    ) in (
        cursor_PTC.description
    ):  # description is a list of tuples, so add [0] to get the text from the
initial element
        fieldList.append(field[0])
    cursor_PTC.execute(str_string2execute)
    for values in cursor_PTC.fetchmany(size=1):
        row = dict(zip(fieldList, values))
        # print(row, movement)
        print(f"{row} ', offset_j='{offset_j}")
    ARC_dir = "Flickr_archive"
    image_ARC_png = row["id"] + ".png"
    image_ARC_path = os.path.join(work_dir, ARC_dir, image_ARC_png)
    URL_dict = {
        "id": row["id"],
        "image_URL": image_ARC_png,
        "image_file": image_ARC_path,
        "category": row["category"],
    }
    return URL_dict


def Get_Flickr_Picture():
    """
    Function: Gets data from the database and returns one dictionary with the
    required data for the GUI, to be able show the it and record the category.

    While at it, it uses request to fetch the picture, stores it locally
(.jpg)
    Then converts that (PIL used) to .png (since PySimpleGui does not do jpg).
    """
    URL_dict = {}
    cursor_PTC = conn.cursor()
    cursor_PTC.execute(
        "Select id, url_m, url_c, url_o from photaglink where id not in "
        + "(select id from photagcategory "
        + ") limit 1;"
    )
    # Retrieving url's for different sizes, smallest first
    # m = small, 240 pixels on longest side
    # c = medium, 800 on longest side
    # o = original. NB seems to return the web-page rather than the image!
```

```
    fieldList = []
    for (
        field
    ) in (
        cursor_PTC.description
    ):  # list of tuples, so add [0] to get the text from the initial element
        fieldList.append(field[0])
    for values in cursor_PTC.fetchmany(size=1):
        # fetchmany(size = 1) works, fetchone() results in an error about a
'missing schema'.
        row = dict(zip(fieldList, values))
        # print("url_m=", row["url_m"], "url_c=", row["url_c"], "url_o=",
row["url_o"])
        # image_URL = row['photo.urls.url.text']
        if row["url_m"] != "NA":
            image_URL = row["url_m"]
        elif row["url_c"] != "NA":
            image_URL = row["url_c"]
        elif row["url_o"] != "NA":
            image_URL = row["url_o"]
        else:
            image_URL = None
            continue
            # No URL available proceed to the next line!
        if image_URL is not None:
            r = requests.get(image_URL, stream=True)
            try:
                r.raise_for_status()
                r.raw.decode_content = True
                # image_TMP_jpg =
r'D:\GIS_data\PhotoTagBonaire\FlickrTemporary.jpg'
                image_TMP_jpg = os.path.join(work_dir, "FlickrTemporary.jpg")
                # format of file follows extension '.jpg'
                # Needs adjusting, PySimpleGUI stays closer to what tkInter
can do, which is png not jpg!
                with open(image_TMP_jpg, "wb") as f:
                    f.write(r.content)
                # format of file follows extension '.jpg'
                # Needs adjusting, PySimpleGUI stays closer to what tkInter
can do, which is png not jpg!
                image_TMP_png = image_TMP_jpg[:-4] + ".png"
                image = Image.open(image_TMP_jpg)
                image = image.save(image_TMP_png, format="PNG")
                r.close()
            except requests.HTTPError as e:
                status_code = r.status_code
                if status_code == 410:  # Exception: File not found! Picture
removed!
                    # Do the same thing as when no valid URL is encountered.
```

```
                    image = Image.open(os.path.join(work_dir,
"URL==None.png"))
                    image_TMP_png = os.path.join(work_dir,
"FlickrTemporary.png")
                    image = image.save(image_TMP_png)
                # add elif: clauses for other exceptions with status_code if
needed.
        else:  # image_URL == None
            image = Image.open(os.path.join(work_dir, "URL==None.png"))
            image_TMP_png = os.path.join(work_dir, "FlickrTemporary.png")
            image = image.save(image_TMP_png)
        # Added code below to keep a copy of the picture for future reference.
JTvdW, 20200226
        # Code added to safe-guard against Flickr going off-line, leaving us
with no options to review the assigned categories
        # or allow other scientist to review our data and options. Intending
not to have to use this, but purely a safe-guard.
        ARC_dir = "Flickr_archive"
        image_ARC_png = row["id"] + ".png"
        image_ARC_path = os.path.join(work_dir, ARC_dir, image_ARC_png)
        image_ARC = Image.open(image_TMP_png)
        if os.path.exists(image_ARC_path):
            pass
            # As the used file name based on ID is unique, there is no need to
rewrite a file
            # if it already exists. Thus avoiding OSerror raised when
overwriting an existing file.
        else:
            image_ARC = image_ARC.save(image_ARC_path, format="PNG")
        # JTvdW, 20200226

        URL_dict = {
            "id": row["id"],
            "image_URL": image_URL,
            "image_file": image_TMP_png,
        }
        # print(URL_dict, type(URL_dict))
    return URL_dict


def insert_record_table(record_dict, table_name):
    cur_insert = conn.cursor()
    table = table_name
    field_list, value_list = zip(*record_dict.items())
    sql_string = "INSERT into {table} {field_list} VALUES
{value_list}".format(
        table=table, field_list=field_list, value_list=value_list
    )
    # print(sql_string)
```

```
    try:
        cur_insert.execute(sql_string)
    except Error as e:
        print(e)
    conn.commit()


def update_record_table(record_dict, table_name, unique_key_name):
    cur_update = conn.cursor()
    table = table_name
    key_field = unique_key_name
    if record_dict[key_field]:
        # Checks for existence of the provided key in the supplied dictionary.
        # N.B. Not for existence in the destination table!
        key_value = record_dict[key_field]
        del record_dict[key_field]
    else:
        print("Error: unique_key for update missing.", key_field)
        print(record_dict)
        exit()
    field_list, value_list = zip(*record_dict.items())
    sql_set_substring = ""
    for nfield in range(0, len(field_list)):
        sql_set_substring = (
            sql_set_substring + field_list[nfield] + " = '" +
value_list[nfield] + "', "
        )
    sql_set_substring = sql_set_substring[:-2]  # removing the trailing ", "
    sql_where_substring = "" + key_field + " = '" + key_value + "'"
    sql_string = "Update {table} SET {set} WHERE {where};".format(
        table=table, set=sql_set_substring, where=sql_where_substring
    )
    print(sql_string)
    try:
        cur_update.execute(sql_string)
    except Error as e:
        print(e)
    conn.commit()


def main():
    # cat_code and cat_list moved here
    # so they exist in time for their use
    cat_list = ["category" + str(i) for i in range(1, 7)]
    cat_code = {
        1: "Coastal",
        2: "party",
        3: "Seascape",
        4: "Landscape",
```

```python
        5: "Wildlife",
        6: "Other",
    }
    i, j = 0, 0
    ptc_mode = "Categorise"  # default mode set to 'Categorise'
    if i == 0:  # Using i==0 to retrieve an initial picture just once.
        flickr_picture = Get_Flickr_Picture()
        # For the initial picture set the cateogory to 6 / Other.
        # Assume cat_key =6 : 'Other'
        cat_key = [6]
        category_to_set = "category" + str(cat_key)
        image_category = cat_key
        if flickr_picture == {}:  # no flickr_picture found to categorise
            ptc_mode = "Review"  # swith to mode = 'Review'
            # and Retrieve_Flickr_Picture
            print(f"pct_mode = {ptc_mode}, time to Retrieve_Flickr_Picture.")
            flickr_picture = Retrieve_Flickr_Picture(0, 0)
            # initialise j
            j = 0
        # Tuple with content ('id', 'image_URL', 'image_file')
        # image_id, image_URL, image_TMP_png = flickr_picture
        image_id, image_URL, image_TMP_png = (
            flickr_picture["id"],
            flickr_picture["image_URL"],
            flickr_picture["image_file"],
        )
        if ptc_mode == "Review":
            if "category" in flickr_picture:
                cat_key = [
                    key
                    for (key, value) in cat_code.items()
                    if value == flickr_picture["category"]
                ]
            else:  # Assume cat_key =6 : 'Other'
                cat_key = [6]
            category_to_set = "category" + str(cat_key[0])
            image_category = cat_key

        # Ensure that Category is set once at the beginning!
        # Cannot be done here already, the window-setup is not there yet!
        # image_URL = 'dummy'
        # image_TMP_png = r'D:\GIS_data\PhotoTagBonaire\FlickrTemporary.png'
        print(image_id, image_URL, image_TMP_png, image_category, ptc_mode)

    timestep_back = None
    movement = "next"
    # # Get the folder containing the images from the user
    # folder = sg.popup_get_folder('Image folder to open')
    # if folder is None:
```

```
    #     sg.popup_cancel('Cancelling')
    #     return
    #
    # # get list of PNG files in folder
    # png_files = [folder + '\\' + f for f in os.listdir(folder) if '.png' in
f]
    # filenames_only = [f for f in os.listdir(folder) if '.png' in f]
    #
    # if len(png_files) == 0:
    #     sg.popup('No PNG images in folder')
    #     return

    # define menu layout
    # menu = [['File', ['Open Folder', 'Exit']], ['Help', ['About', ]]]
    menu = [
        ["File", ["Exit"]],
        [
            "Help",
            [
                "About",
            ],
        ],
    ]

    # define layout, show and read the window
    col = [
        [
            sg.Text(image_URL, size=(40, 1), key="filename"),
            sg.Text(ptc_mode, size=(20, 1), key="ptc_mode"),
        ],
        [sg.Image(filename=image_TMP_png, key="image")],
        # [sg.Button('Prev', size=(8, 2)), sg.Button('Next', size=(8, 2)) # ,
        # sg.Text('File 1 of {}'.format(len(png_files)), size=(15, 1),
key='filenum')
        # ]
        [
            sg.Text(
                "No. of Photo's categorised: {}".format(0), size=(60, 1),
key="filenum"
            )
        ],
        [sg.Button("Prev", size=(8, 2)), sg.Button("Next", size=(8, 2))],
    ]

    # col_files = [[sg.Listbox(values=filenames_only, size=(60, 30),
key='listbox')],
    #              [sg.Button('Read')]]
    # col_files = [[sg.Text('RadioButtons go here',size=(60,30), key =
'rb_categories')]]
```

```
    #     ('Coastal', 1),
    #     ('Party', 2),
    #     ('Seascape incl. watersports', 3),
    #     ('landscape', 4),
    #     ('Wildlife', 5),
    #     ('Other, mainly indoor +urban', 6)
    col_categories = [
        [sg.Text("Categories", key="rb_category_label")],
        [sg.Radio("Coastal", "rb_category", key="category1")],
        [sg.Radio("Party", "rb_category", key="category2")],
        [sg.Radio("Seascape incl. watersports", "rb_category",
key="category3")],
        [sg.Radio("landscape", "rb_category", key="category4")],
        [sg.Radio("Wildlife", "rb_category", key="category5")],
        [
            sg.Radio(
                "Other, mainly indoor +urban",
                "rb_category",
                key="category6",
                default=True,
            )
        ],
    ]
    layout = [[sg.Menu(menu)], [sg.Col(col_categories), sg.Col(col)]]
    window = sg.Window(
        "Image Browser + Categoriser",
        layout,
        return_keyboard_events=True,
        location=(0, 0),
        use_default_focus=False,
    )

    # loop reading the user input and displaying image, filename
    while True:
        event, values = window.read()
        # --------------------- Button & Keyboard ---------------------
        if event is None:
            break
        elif event in (
            "Next",
            "MouseWheel:Down",
            "Down:40",
            "Next:34",
        ):  #  and i < len(png_files)-1:
            i += 1
            movement = "next"
            if ptc_mode == "Review":
                j += 1
            # Next here read() the window
```

```
        # Extract the value for the radiobutton that was selected
        # Pass onto a function, with the accompanying data from either the
window or Cursur_PTC
        # to store the data
        # DONE: proces the result
        # print(values)
        # cat_list & cat_code have moved up. They are required earlier!
        # cat_list = ["category" + str(i) for i in range(1, 7)]
        # cat_code = {
        #     1: "Coastal",
        #     2: "Party",
        #     3: "Seascape",
        #     4: "Landscape",
        #     5: "Wildlife",
        #     6: "Other",
        # }
        # print(cat_list)
        for cat in cat_list:
            if values[cat] == True:
                cat_int = int(cat[-1])
                category_chosen = cat_code[cat_int]
        timestamp = datetime.datetime.now().strftime("%Y-%m-%d
%H:%M:%S.%f")
        categoriser = os.getenv("username")
        print(image_id, category_chosen, categoriser, timestamp)
        category_dict = {
            "id": image_id,
            "category": category_chosen,
            "categoriser": categoriser,
            "timestamp": timestamp,
        }
        if timestep_back is None:
            # Direction of movement == Next
            movement = "next"
            # if ptc_mode == "Review":
            # j += 1   # not here again, otherwise not achieving step-
size=1
            if ptc_mode == "Categorise":
                insert_record_table(category_dict, "photagcategory")
            else:  # Review, so update the existing record
                update_record_table(category_dict, "photagcategory", "id")
        else:
            # Direction of movement is or was Previous
            movement = "prev"
            if ptc_mode == "Review":
                j -= 1
            update_record_table(category_dict, "photagcategory", "id")
```

```
            # sg.popup_auto_close(('IMAGE id: ', image_id, 'category chosen:
', category_chosen))
            #                      '\n at: ', datetime.now(), '\t by: ',
'user'))

            # and continue
            # Once the data about the previously categorised photagraph has
been entered,
            # Get_Flickr_Picture should notice return with a new/next
photograph to view.
            if ptc_mode == "Categorise":
                flickr_picture = Get_Flickr_Picture()
                if i == 0:
                    # And ensure that the Category is set and showing
                    # update wind with current categoy
                    category_to_set = "category" + str(cat_key[0])
                    image_category = cat_key
            else:  # Review active
                flickr_picture = Retrieve_Flickr_Picture(image_id, j)
            # Dictionary with content ('id', 'image_URL', 'image_file')
            # image_id, image_URL, image_TMP_png = flickr_picture
            image_id = flickr_picture["id"]
            image_URL = flickr_picture["image_URL"]
            image_TMP_png = flickr_picture["image_file"]
            # image_id = flickr_picture
            # image_URL = 'dummy'
            # image_TMP_png =
r'D:\GIS_data\PhotoTagBonaire\FlickrTemporary.png'

            # update window with new image
            window["image"].update(filename=image_TMP_png)
            # update window with filename
            window["filename"].update(image_URL)
            # update page display
            # window['filenum'].update('File {} of {}'.format(i+1,
len(png_files)))
            window["filenum"].update("No. of Photo's categorised:
{}".format(i))
            window["ptc_mode"].update(ptc_mode)
            timestep_back = None
            movement = "next"
            # if ptc_mode == "Review":
            #     j += 1  # not here again, otherwise not achieving step-
size=1
            if ptc_mode == "Review":
                if "category" in flickr_picture:
                    cat_key = [
                        key
                        for (key, value) in cat_code.items()
```

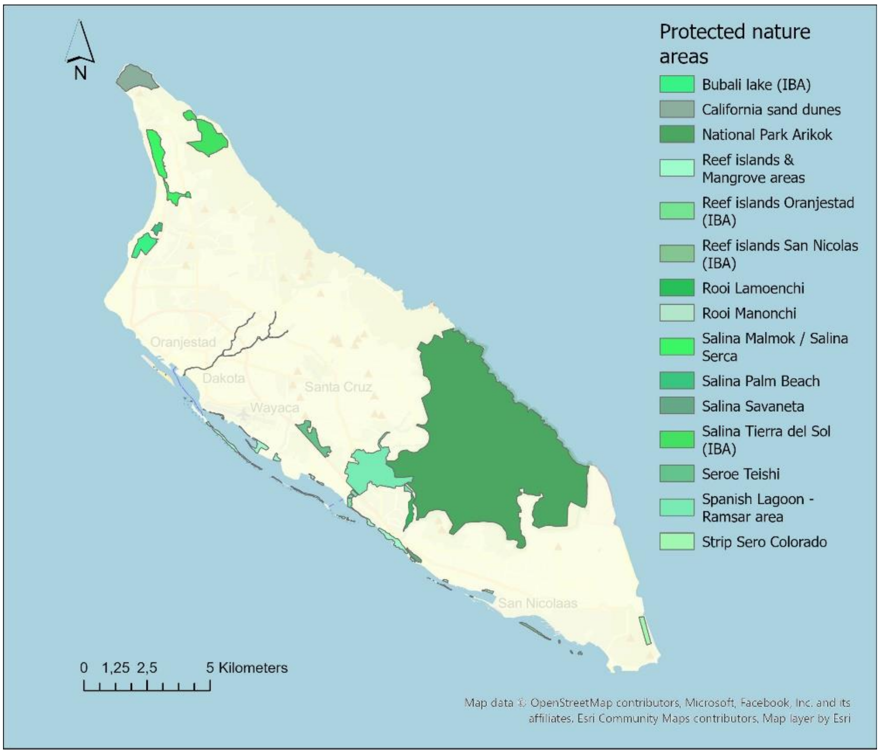
```
                        if value == flickr_picture["category"]
                    ]
                else:  # Assume cat_key =6 : 'Other'
                    cat_key = [6]
                category_to_set = "category" + str(cat_key[0])
                image_category = cat_key
            # update wind with current categoy
            window.Element(category_to_set).Update(value=True)
            print(image_id, image_URL, image_TMP_png, image_category)
        elif event in ("Prev", "MouseWheel:Up", "Up:38", "Prior:33") and i >
0:
            i += 1
            for cat in cat_list:
                if values[cat] == True:
                    cat_int = int(cat[-1])
                    category_chosen = cat_code[cat_int]
            timestamp = datetime.datetime.now().strftime("%Y-%m-%d
%H:%M:%S.%f")
            categoriser = os.getenv("username")
            print(image_id, category_chosen, categoriser, timestamp)
            category_dict = {
                "id": image_id,
                "category": category_chosen,
                "categoriser": categoriser,
                "timestamp": timestamp,
            }
            if timestep_back is None:
                # first move backwards!
                timestep_back = datetime.datetime.now().strftime("%Y-%m-%d
%H:%M:%S.%f")
                # earlier than now should do nicely.
                # Make sure to replace with returned value when and if it
becomes available later on.
                if ptc_mode == "Categorise":
                    insert_record_table(category_dict, "photagcategory")
                else:  # 'Review'
                    update_record_table(category_dict, "photagcategory", "id")
                movement = "prev"
                if ptc_mode == "Review":
                    j -= 1
            else:  # the GUI has moved back and before moving again the
current state must be recorded
                # i.e. Update the record.
                # Start this by collecting the data from the GUI.
                update_record_table(category_dict, "photagcategory", "id")
                movement = "prev"
                if ptc_mode == "Review":
                    j -= 1  # not here again, otherwise not achieving step-
size=1
```

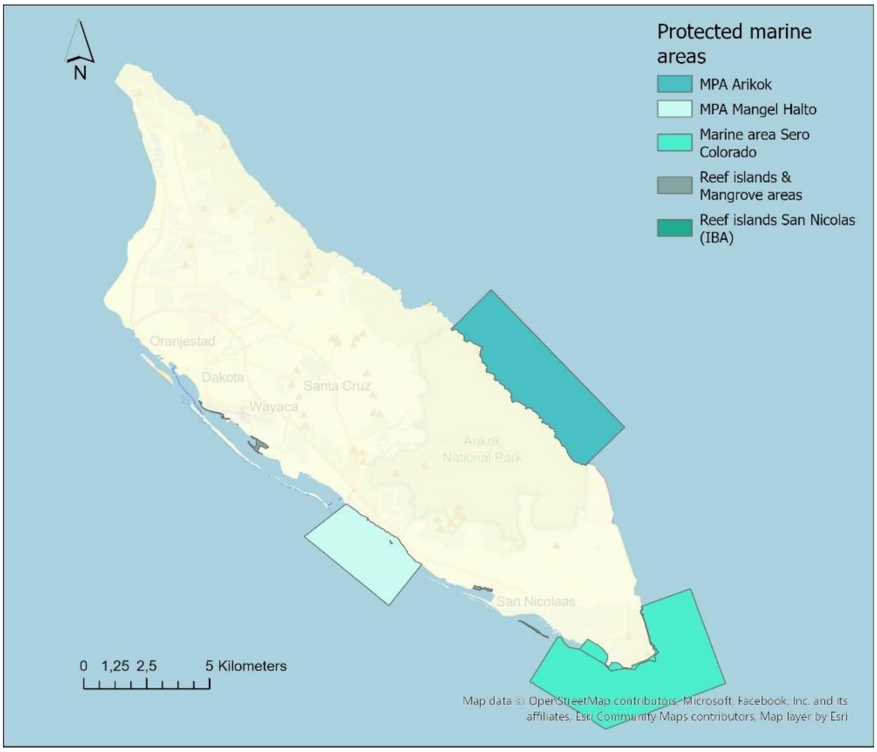
```
            if ptc_mode == "Categorise":
                movement = "prev"
                # if ptc_mode == "Review":
                #     j -= 1  # not here again, otherwise not achieving step-
size=1
                flickr_picture = Previous_Flickr_Picture(timestep_back)
                # URL_dict = {'id': previousID, 'image_URL': image_URL,
'image_file': image_TMP_png,
                #                      'category': category_current
,time_stamp': timestamp_current}
                image_id = flickr_picture["id"]
                image_URL = flickr_picture["image_URL"]
                image_TMP_png = flickr_picture["image_file"]
                image_category = flickr_picture["category"]
                print(image_id, image_URL, image_TMP_png, image_category)
                timestep_back = datetime.datetime.now().strftime("%Y-%m-%d
%H:%M:%S.%f")
                # Decode the retrieved Category to the appropriate integer
                # cat_code = {1: 'Coastal', 2: 'Party', 3: 'Seascape',
                #             4: 'Landscape', 5: 'Wildlife', 6: 'Other'}
                # and use that to set the current selection.
                cat_key = [
                    key for (key, value) in cat_code.items() if value ==
image_category
                ]
                category_to_set = "category" + str(cat_key[0])
                print(cat_key, category_to_set)
                # update wind with current categoy
                window.Element(category_to_set).Update(value=True)
                # update window with new image
                window["image"].update(filename=image_TMP_png)
                # update window with filename
                window["filename"].update(image_URL)
                # update page display
                # window['filenum'].update('File {} of {}'.format(i+1,
len(png_files)))
                window["filenum"].update("No. of Photo's categorised:
{}".format(i))
            else:  # ptc_mode = 'Review'
                movement = "prev"
                # if ptc_mode == "Review":
                #     j -= 1  # not here again, otherwise not achieving step-
size=1
                flickr_picture = Retrieve_Flickr_Picture(id, j)
                image_id = flickr_picture["id"]
                image_URL = flickr_picture["image_URL"]
                image_TMP_png = flickr_picture["image_file"]
                image_category = flickr_picture["category"]
```

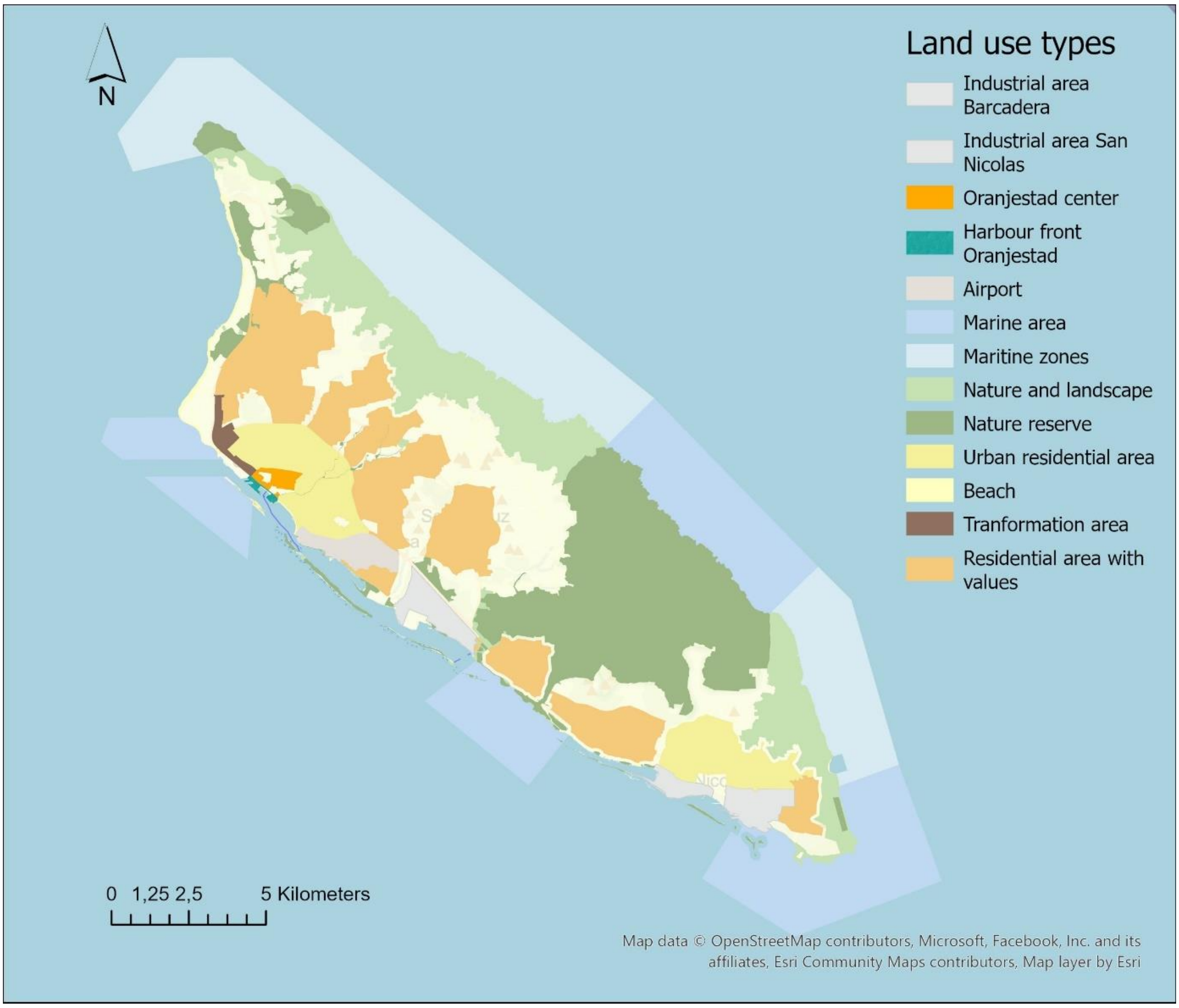
```
                cat_key = [
                    key for (key, value) in cat_code.items() if value ==
image_category
                ]
                category_to_set = "category" + str(cat_key[0])
                print(cat_key, category_to_set)
                # update wind with current categoy
                window.Element(category_to_set).Update(value=True)
                # update window with new image
                window["image"].update(filename=image_TMP_png)
                # update window with filename
                window["filename"].update(image_URL)
                # update page display
                # window['filenum'].update('File {} of {}'.format(i+1,
len(png_files)))
                window["filenum"].update("No. of Photo's categorised:
{}".format(i))
                window["ptc_mode"].update(ptc_mode)

            if ptc_mode == "Review":
                if "category" in flickr_picture:
                    cat_key = [
                        key
                        for (key, value) in cat_code.items()
                        if value == flickr_picture["category"]
                    ]
                else:  # Assume cat_key =6 : 'Other'
                    cat_key = [6]
                category_to_set = "category" + str(cat_key[0])
                image_category = cat_key
            # update wind with current categoy
            window.Element(category_to_set).Update(value=True)
            print(image_id, image_URL, image_TMP_png, image_category)
        elif event == "Exit":
            break

        # if event == 'Read':
        #     filename = folder + '/' + values['listbox'][0]
        # else:
        #     filename = png_files[i]

        # ----------------- Menu choices -----------------
        # if event == 'Open Folder':
        #     newfolder = sg.popup_get_folder('New folder', no_window=True)
        #     if newfolder is None:
        #         continue
        #
        #     folder = newfolder
        #     png_files = [folder + '/' +
```

```
        #                  f for f in os.listdir(folder) if '.png' in f]
        #     filenames_only = [f for f in os.listdir(folder) if '.png' in f]
        #
        #     window['listbox'].update(values=filenames_only)
        #     window.refresh()
        #
        #     i = 0
        # elif event == 'About':
        if event == "About":
            sg.popup(
                "Demo PNG Viewer Program",
                "Please give PySimpleGUI a try!",
                "Modified to Categorise Flickr-images."
                "Wageningen Marine Research, J.T. v/d Wal",
            )
    window.close()


if __name__ == "__main__":


    api_key = Fk.FRapi_dict["Key"]
    api_secret = Fk.FRapi_dict["Secret"]
    movement = None
    production = False
    if production == True:
        work_dir = r"W:\IMARES\Data\GIS_applicaties\PhotoTagBonaire"
        db_name = "PhoTagClassified.sql3.db"
    else:
        work_dir = "D:\\GIS_data\\PhotoTagBonaire\\"
        # db_name = "PhoTagClassified_dev.sql3.db"
        db_name = "PhoTagClassified.sql3.db"
    # f'Working with database= {work_dir}{db_name}'
    print(
        "Working with database= {work_dir}{db_name}".format(
            work_dir=work_dir, db_name=db_name
        )
    )
    conn = sqlite3.connect(os.path.join(work_dir, db_name))
    # And moving even more stuff.
    # Creates Category-table (if not exist)
    PTC_Init_Category_Table()

    main()
```

# Appendix D	Land-use and protected areas maps




Land use types
Industrial area Barcadera
Industrial area San Nicolas
Oranjestad center
Harbour front Oranjestad
Airport
Marine area
Maritine zones
Nature and landscape
Nature reserve
Urban residential area
Beach
Tranformation area
Residential area with values
N
0 1,25 2,5 5 Kilometers
Map data © OpenStreetMap contributors, Microsoft, Facebook, Inc. and its affiliates, Esri Community Maps contributors, Map layer by Esri

## Appendix E Survey Tourism Agencies Part 1

Thank you for participating in this research. This research answers the question: 'Mapping the spatial distribution of tourism in Aruba through social media'. In this research, images from the social media platform Flickr are extracted to map tourism in Aruba, distinguishing between different types of tourists. Ultimately, using this method, I only catch part of the tourists who visit Aruba, namely those who post an image on Flickr. That is why the help of your tourism expertise is needed to have a critical assessment of the outcomes.

First, we start with some basic questions.

What is your gender? ………………………………………………

How old are you? ………………………………………… ……...

In which city do you live? ………………………………………………

In what part of Aruba do you live? ………………………………………………

For how long have you lived in Aruba? ………………………………………………

For what organization do you work? ………………………………………………

What is your profession?

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

How well do you know the Aruba 1=not, 5=very well

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

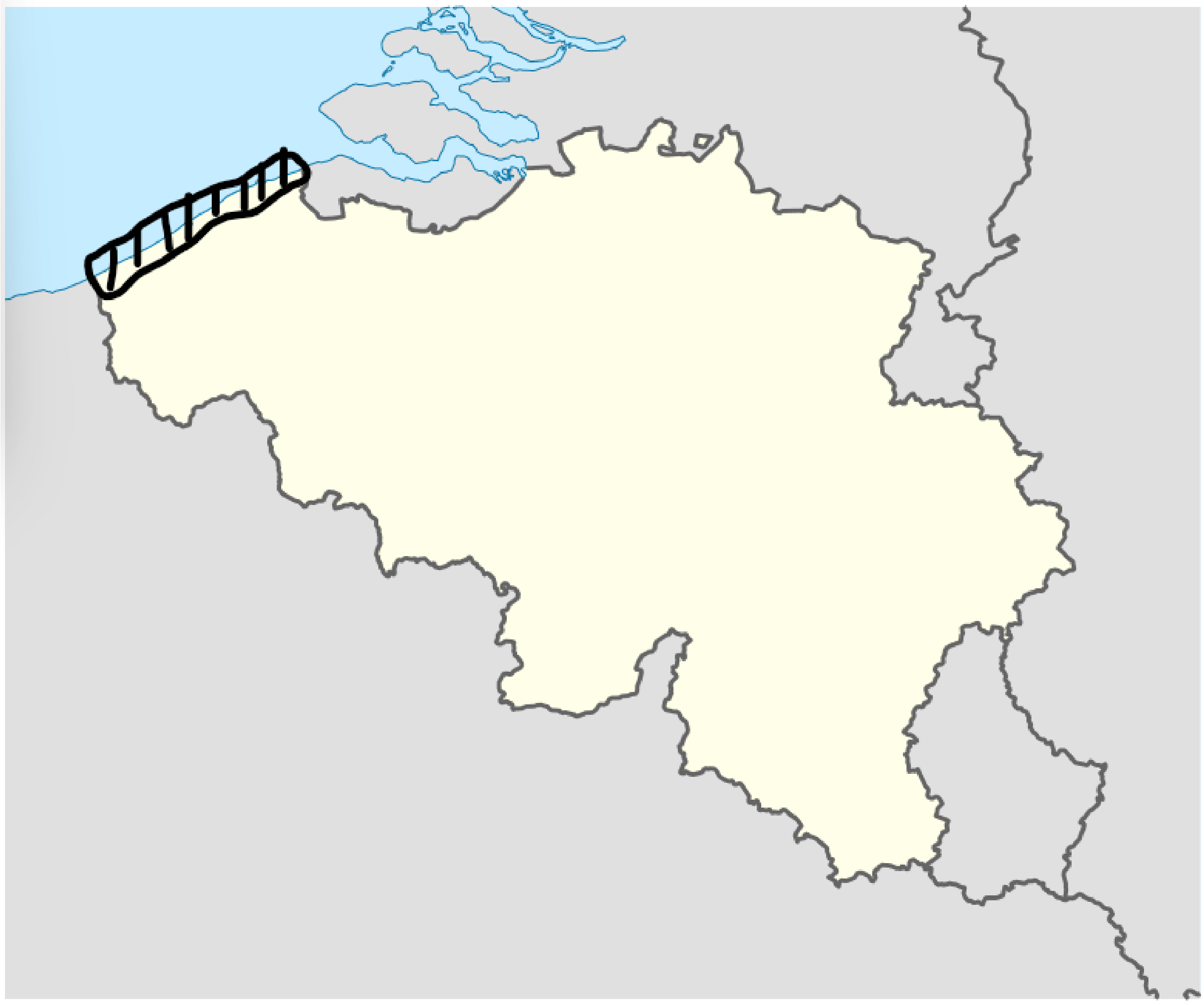
Below several maps will be displayed in which you have to draw the areas which type of tourist would mainly go to. The classifications of the types of tourists are:

1. Coastal: coastline from the mainland
2. Party/all-inclusive: all-inclusive hotels and people partying
3. Seascapes: scenic seascape views (above water) and water sports (from the sea)
4. landscape: natural landscapes (excluding terrestrial wildlife)
5. Wildlife: images of terrestrial wildlife and birds and marine wildlife
6. Other: mainly indoor and urban images

First, an example is given on how to exactly tackle these maps, after that you can fill in per type of tourists where you think the tourists would be present

EXAMPLE:

Coastal tourists in France:



Explain why here:

This is quite a simplified example but here then explain why and where you chose these region(s), name for example some famous beaches or other place names that you think tourists would go there for.

Coastal tourists go mainly to:

Explain why here:

…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………
…………………………………………………………………………………………………

Party/All-inclusive tourists mainly go to:

Explain why here:
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………
………………………………………………………………………………………………

Seascape tourists mainly go to:

Explain why here:

Landscape tourists mainly go to:

Explain why here:

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

………………………………………………………………………………………………………

Wildlife tourists mainly go to:

Explain why here:
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

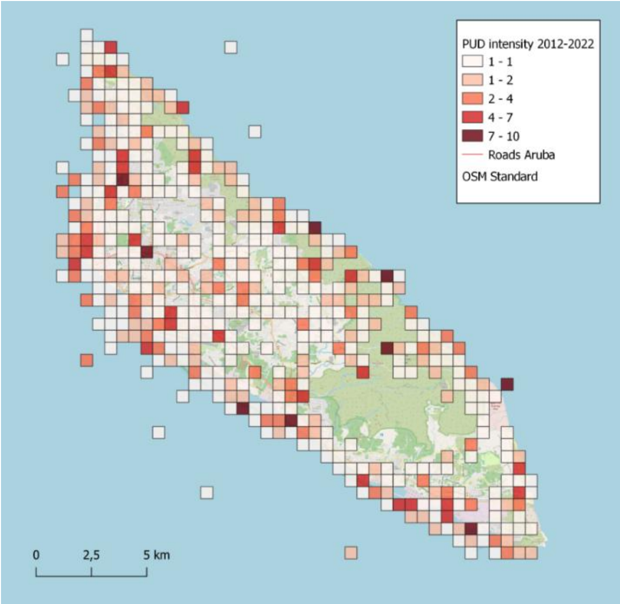
Other types of tourists would mainly go to:

Explain why here:
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………
……………………………………………………………………………………………………

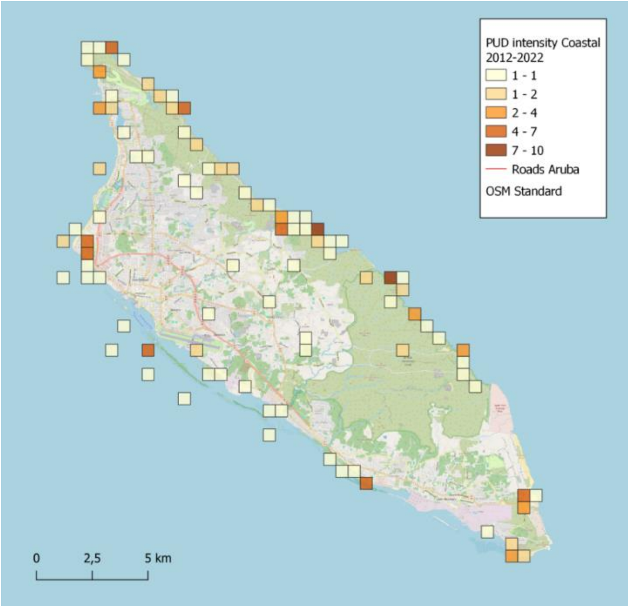
## Appendix F Survey Tourism Agencies Part 2

Thank you again for participating in this research, the results of the social media outcomes are displayed. The tourists are displayed in the intensity of PUD (photo user days) this is a measure that calculates the number of individual users that upload at least one image on a unique day, in a particular location (Wood et al., 2013). After seeing the results below there are some questions that would be helpful if they would be answered.

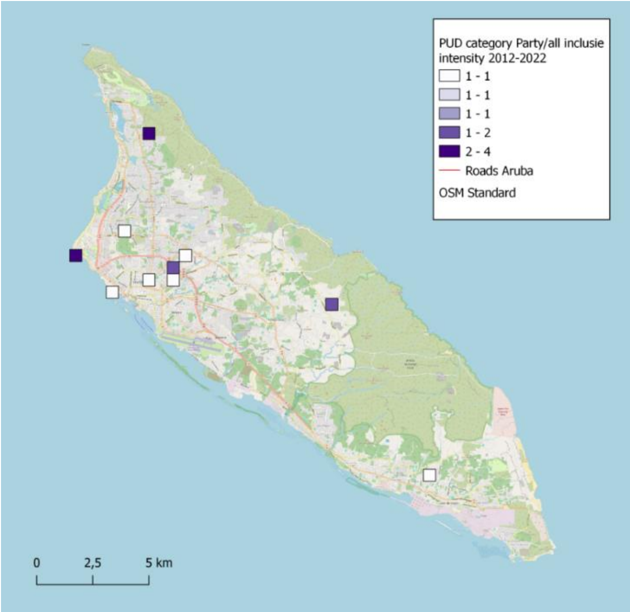
All tourists in Aruba:



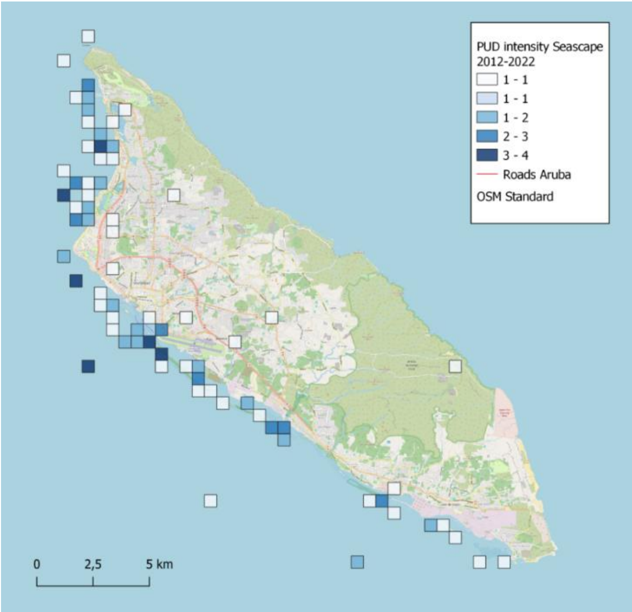

Coastal tourists:



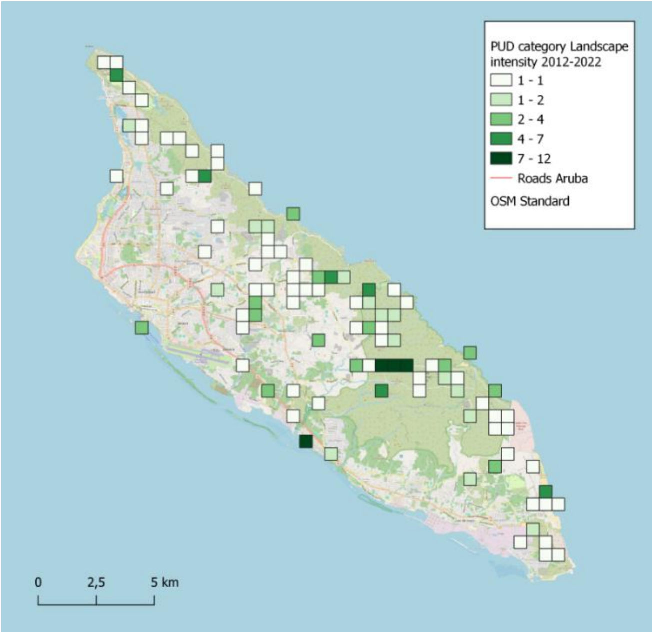
Party/all-inclusive tourists:



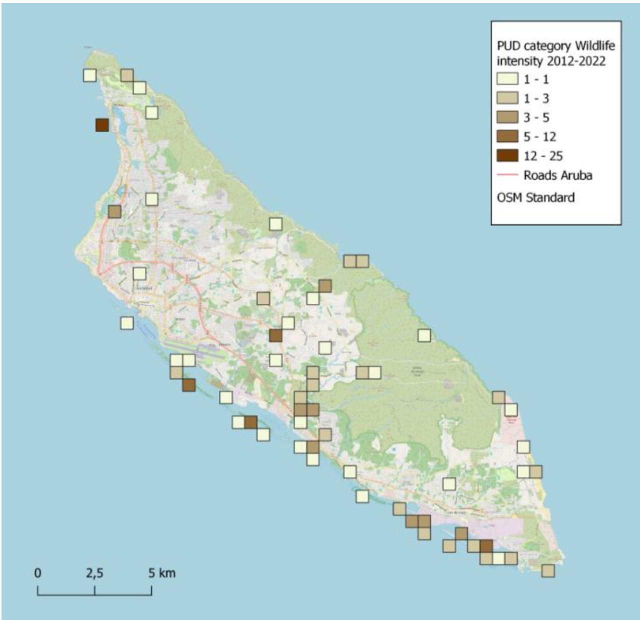

Seascape tourists:



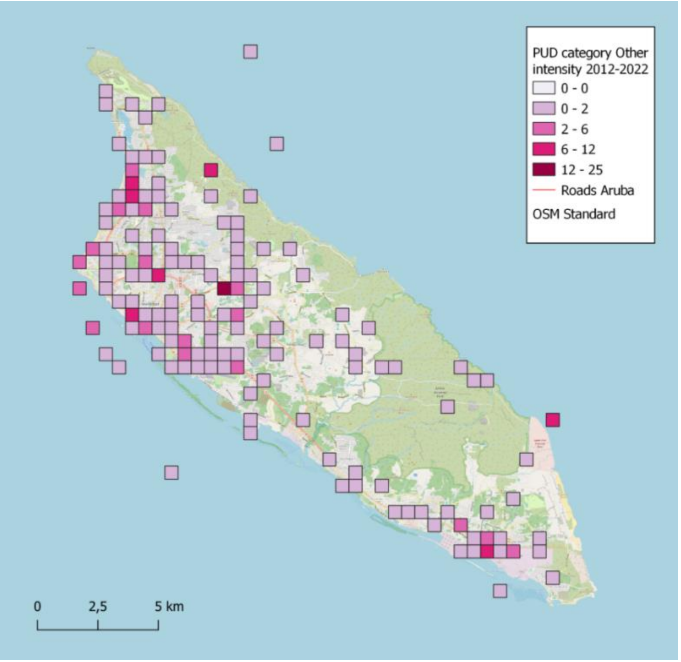
Natural landscape tourists




Wildlife tourists:



Other tourists




How well do the study outcomes resemble the real-world situation? 1=no, 5=very well

1 2 3 4 5

☐ ☐ ☐ ☐ ☐

Are the study outcomes useful for you or for your organization? 1=no, 5=very useful

1 2 3 4 5

☐ ☐ ☐ ☐ ☐

For what purposes could you use the study outcomes?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

Do you have any other comments or suggestions?

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

……………………………………………………………………………………………………

………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………

## Appendix G Comments and remarks of tourism experts

Below the respondent's answers to the two surveys can be found. Only the answers are included in the Appendix.

## Respondent 1

First we start with some basic questions.

| | |
|---|---|
| What is your gender? | Female |
| How old are you? | 38. |
| In which city do you live? | Noord |
| In what part of Aruba do you live? | Sero Pela |
| For how long have you lived in Aruba? | 5 years |
| For what organization do you work? | Renaissance Wind Creek Aruba Resort |

What is your profession?

Director of Operations

How well do you know the Aruba 1=not, 5=very well

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| ☐ | ☐ | ☒ | ☐ | ☐ |

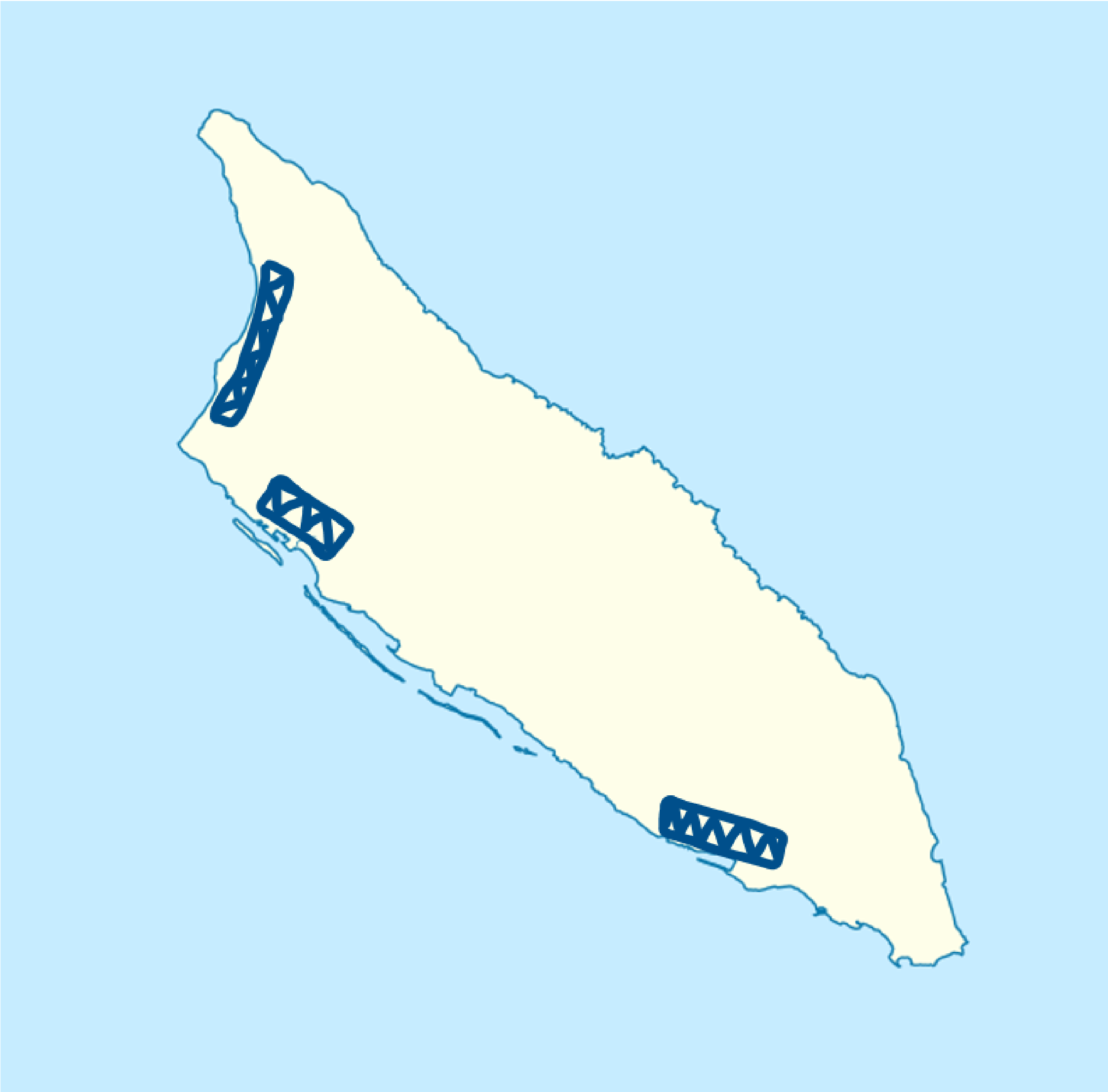
Coastal tourists go mainly to:

Explain why here:

*Aruba's beaches are great for Coastal Tourtists, as they are easy to access, many have facilities and waters are diverse calm to wavy for kite and wind surfers. Below lists the popular coastal areas:*

*Beaches such as Arashi, Boca Catalina, Malmok, Eagle, Palm, Druf, Mangel Halto, Surf Side, Rodgers, Baby and Boca Grandi*

Resort visitors/party tourists mainly go to:



Explain why here:

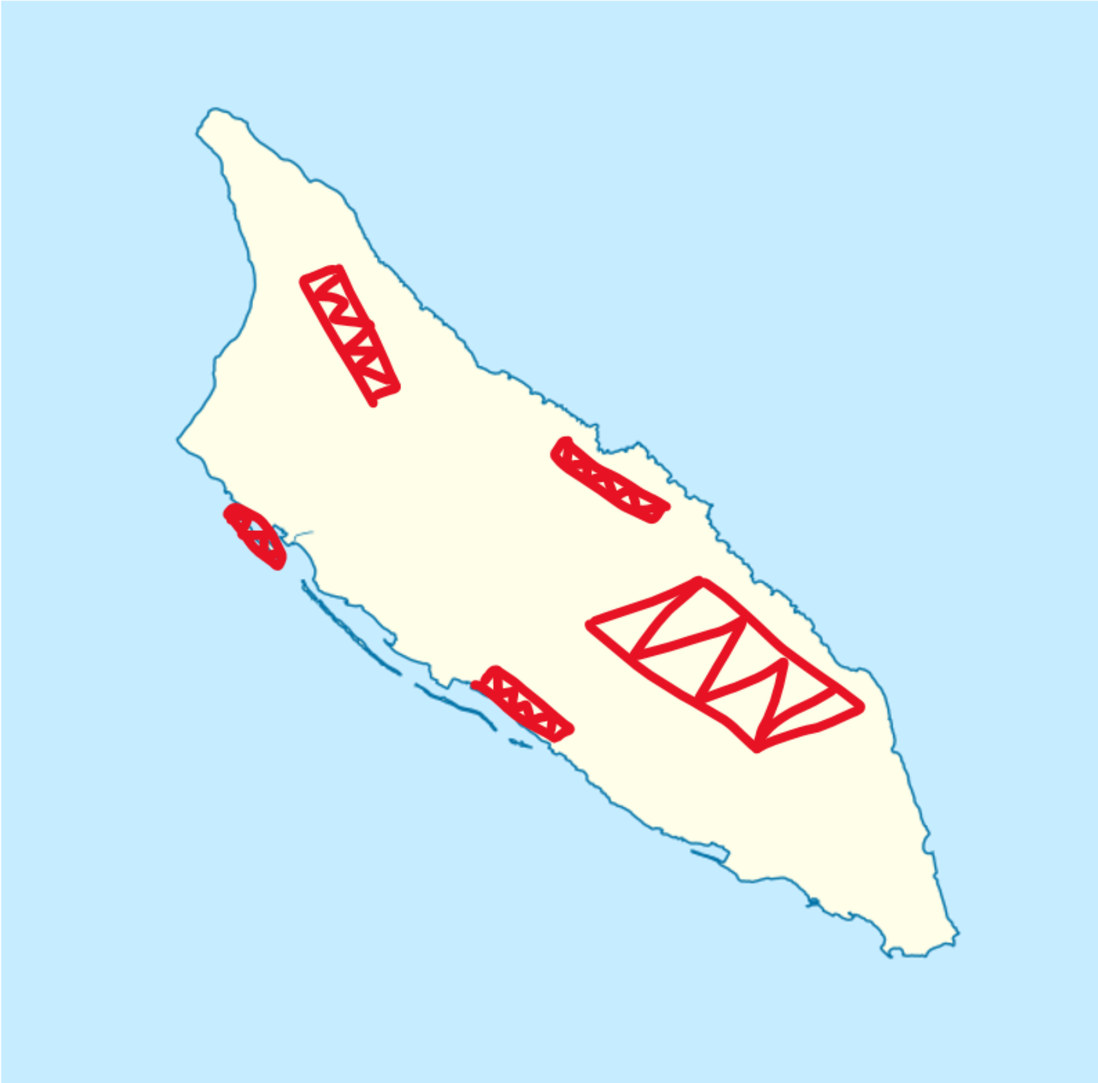
*Resort Visitors/party tourists will stay close to the Palm Beach and Downtown areas, they may also visit the San Nicolaas area. Palm Beach especially has a high concentration of restaurants and bars, it is easily accessible, safe, and open until late hours at night. The downtown area is a main attraction during the day, however restaurants and bars attract Resort Visitors as well. These visitors tend to explore the area of San Nicolaas for a cultural and unique experience.*

Seascape tourists mainly go to:

Explain why here:

*Boca Catalina, Arashi Beach, West Punt Coasts, Druif Beach, Anndicuri Beach, Natural Bridge, Dunes are the beaches where Seascape Tourists will frequent the most, besides the visual pleasure tourists may also appreciate the wave diversity and accessibility. Most wind surfing/kite surfing schools have facilities near these areas.*

Terrestrial landscapes tourists mainly go to:



Explain why here:

*Many of the terrestrial landscape that is not part of the national park are used for hiking, atv/utv tours horseback riding and multiple tourist activities.*

Wildlife tourists mainly go to:

Explain why here:

*These landscapes are visited to view and experience the local wildlife such as lizards, snakes, crabs, hatching turtles, monarch butterflies, mountain-climbing goats, donkeys and of course the neon-green Prikichi*

Other, mainly urban, types of tourists would mainly go to:

Explain why here:

*Tourists mainly frequent the Palm beach area and down town area for urban activities.*

How well do the study outcomes resemble the real world situation? 1=no, 5=very well

1 2 3 4 5

☐ ☐ ☐ ■ ☐

Are the study outcomes useful for you or for your organization? 1=no, 5=very useful

1 2 3 4 5

☐ ☐ ■ ☐ ☐

For what purposes could you use the study outcomes?

To identify the interests of past visitors and provide customized suggestions for future visitors to appeal to and add value to their stay in Aruba. To identify areas that may require community enhancements, wildlife conservation and voluntary clean up efforts to reduce the impact to nature.

Do you have any other comments or suggestions?

There is a lack of trash disposal throughout the island and facilities that make the tourist experience more nature inclusive (educational signage, photo taking facilities to avoid damage to eco systems etc.)

## Respondent 2

First we start with some basic questions.

What is your gender? MALE…………………………………………

How old are you? 58………………………………… ……...

In which city do you live? ORANJESTAD (SAN BARBOLA)

In what part of Aruba do you live? CENTER ………………………………

For how long have you lived in Aruba? 25 YRS………………………………………

For what organization do you work? ARUBA TOURISM AUTHORITY

What is your profession?

DESTINATION SERVICE MANAGER

As Destination Services Manager, responsibilities include:

- Development of multi-annual commercial business plan for 5 business units, namely a. product development, b. destination development, c. trade relations, d. cruise, and e. events & sponsorships
- Setup and maintenance of budgets
- Responsible for developing and maintaining product/ destination portfolio
- Responsible for business results
- Structuring/ shaping business unit (commercially and organizationally)
- Negotiations stakeholders (both locally and international)
- Assist in determination of strategic course ATA/ tourism

How well do you know the Aruba 1=not, 5=very well

1 2 3 4 5

☐ ☐ ☐ ✔ ☐

Coastal tourists go mainly to:


Explain why here:

Aruba is most famous for its beaches, especially EAGLE BEACH (award winning beach); Palm Beach is the strip where most large hotels are situated, visitors will automatically make a lot of use of those beaches. The more remote areas are Baby Beach, Rodgers Beach and Mangel Halto; visitors will have to specifically travel to those to enjoy.

Resort visitors/party tourists mainly go to:


Explain why here:

The bulk of hotel activity is mainly centralized in the 'Noord' area (Palm Beach), and adjacent are all of the locations where visitors could go out to party.

Seascape tourists mainly go to:


Explain why here:

Visitors interested in water(sports) activities can resort to diving, snorkeling, sailing on different vessels, motor boating and more. This happens in the most frequented areas of Palm Beach mainly, diving will take place at other locations based on offerings (wreck diving).

Terrestrial landscapes tourists mainly go to:


Explain why here:

Visitors have the option to go on different tours, or rent vehicles, go hiking, mountain biking on different locations all over the island, depending on their interest.

Wildlife tourists mainly go to:


Explain why here: identical to terrestrial landscapes tourists

Other, mainly urban, types of tourists would mainly go to:


Explain why here:

Options for shopping, tours through different urban areas, art & crafts (murals, e.g.) especially in San Nicolas.

How well do the study outcomes resemble the real world situation? 1=no, 5=very well

1 2 3 4 5

☐ ☐ ☐ ■ ☐

Are the study outcomes useful for you or for your organization? 1=no, 5=very useful

1 2 3 4 5

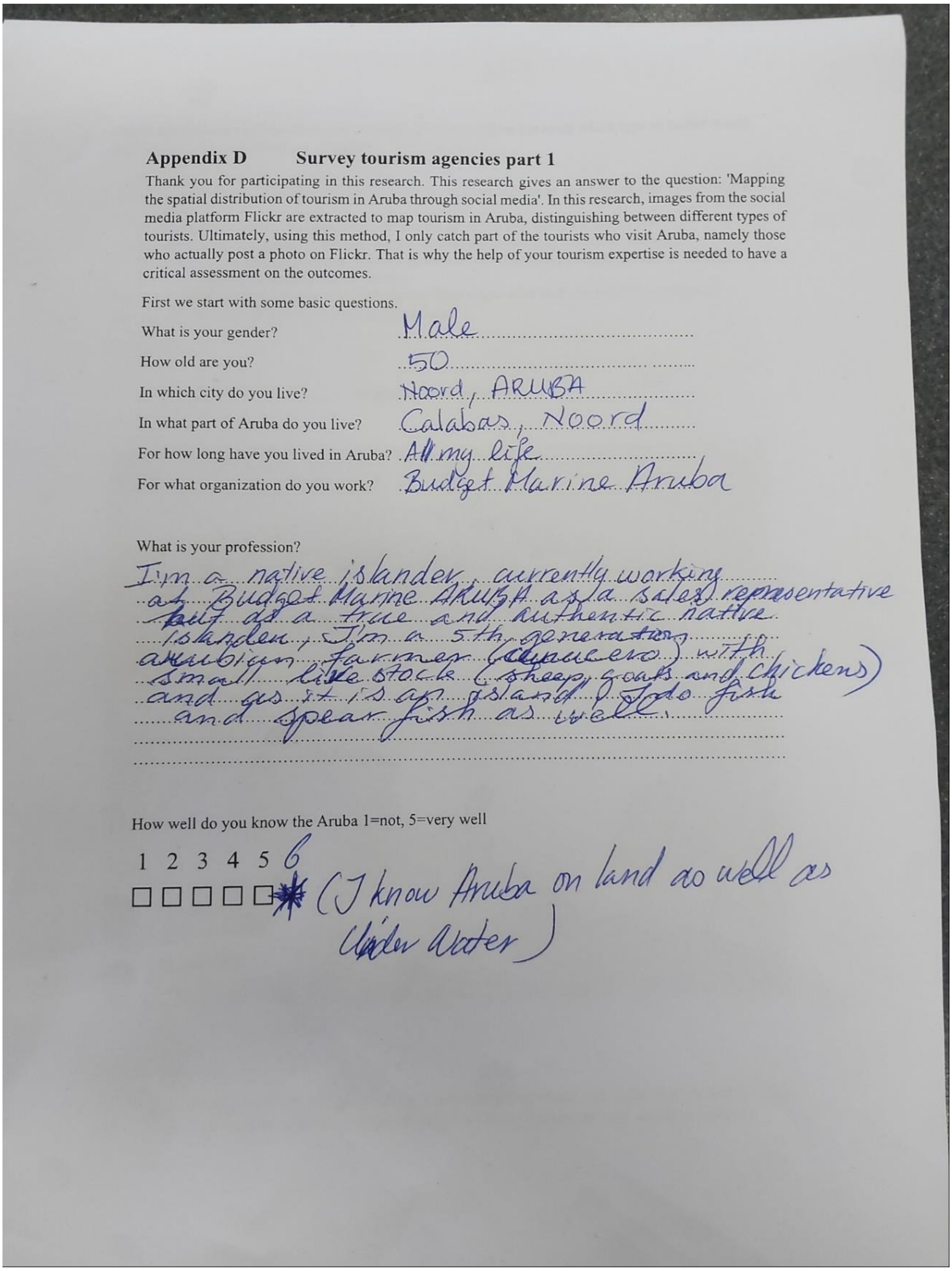
☐ ☐ ☐ ■ ☐

For what purposes could you use the study outcomes?

To understand visitor flow and attempt to manage such, both from a sustainability standpoint (nature preservation) and where most visited attractions are, and how to manage/enhance/maintain these.

Do you have any other comments or suggestions?

## Respondent 3



**Appendix D Survey tourism agencies part 1**

Thank you for participating in this research. This research gives an answer to the question: 'Mapping the spatial distribution of tourism in Aruba through social media'. In this research, images from the social media platform Flickr are extracted to map tourism in Aruba, distinguishing between different types of tourists. Ultimately, using this method, I only catch part of the tourists who visit Aruba, namely those who actually post a photo on Flickr. That is why the help of your tourism expertise is needed to have a critical assessment on the outcomes.

First we start with some basic questions.

What is your gender? Male

How old are you? 50

In which city do you live? Noord, ARUBA

In what part of Aruba do you live? Calabas, Noord

For how long have you lived in Aruba? All my life

For what organization do you work? Budget Marine Aruba

What is your profession?

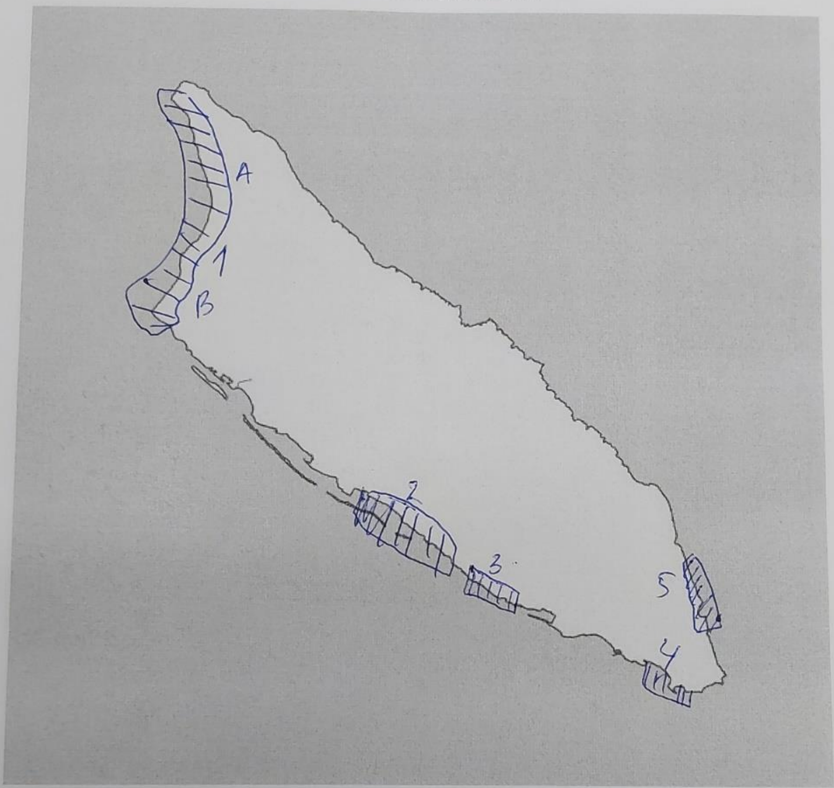
I'm a native islander, currently working at Budget Marine ARUBA as a sales representative but as a true and authentic native islander, I'm a 5th generation aruban farmer (cunucero) with small live stock (sheep, goats and chickens) and as it is an island I do fish and spear fish as well.

How well do you know the Aruba 1=not, 5=very well

1 2 3 4 5 6

☐ ☐ ☐ ☐ ☐ * (I know Aruba on land as well as under water)

Coastal tourists go mainly to:




Explain why here:

mega activities

(1) High Rise (A) and ~~[illegible]~~ Low Rise Hotel AREA

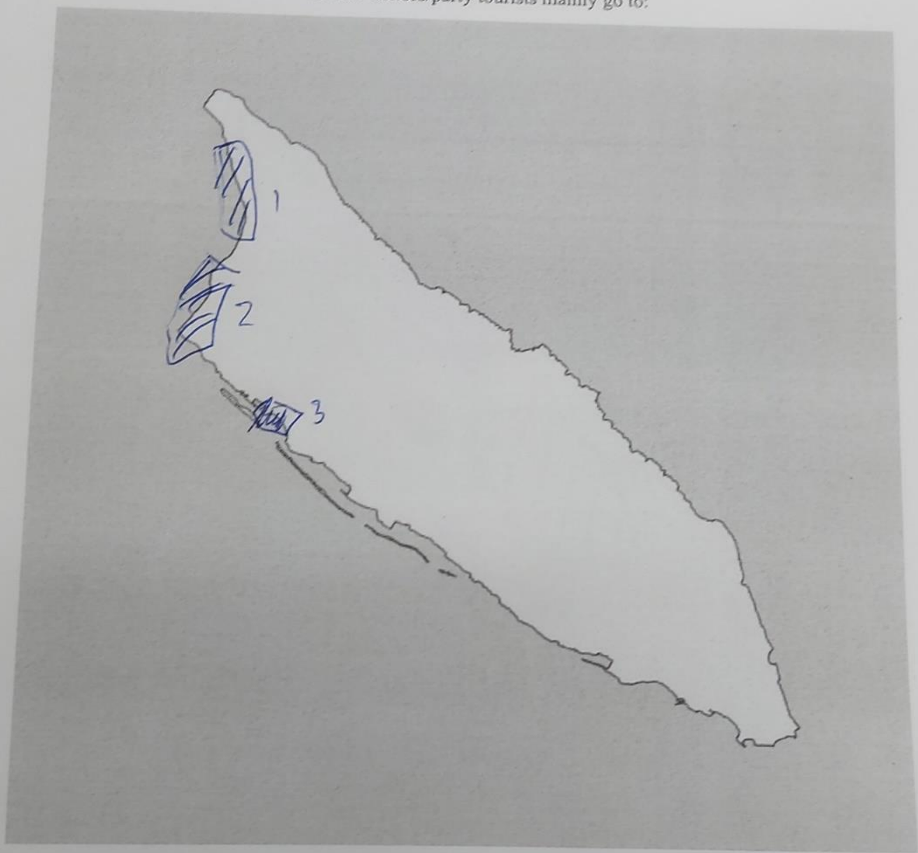
(2) South Side ~~Island~~ Attraction as well activities - watersports

(3) Another type of beach and inshore waters with loads of activities

(4) Major Tourist activities and beach area

(5) Busy sight seeing as well as kite Surfing grounds.

Resort visitors/party tourists mainly go to:



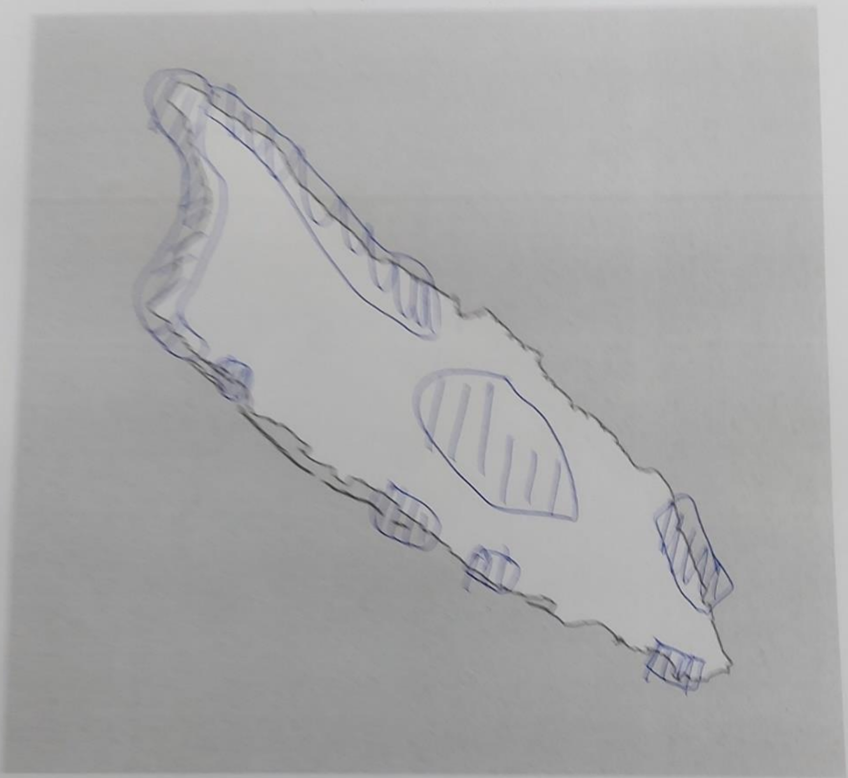

Explain why here:

(1) High Rise Hotels

(2) Low Rise Hotels

(3) Downtown Hotel

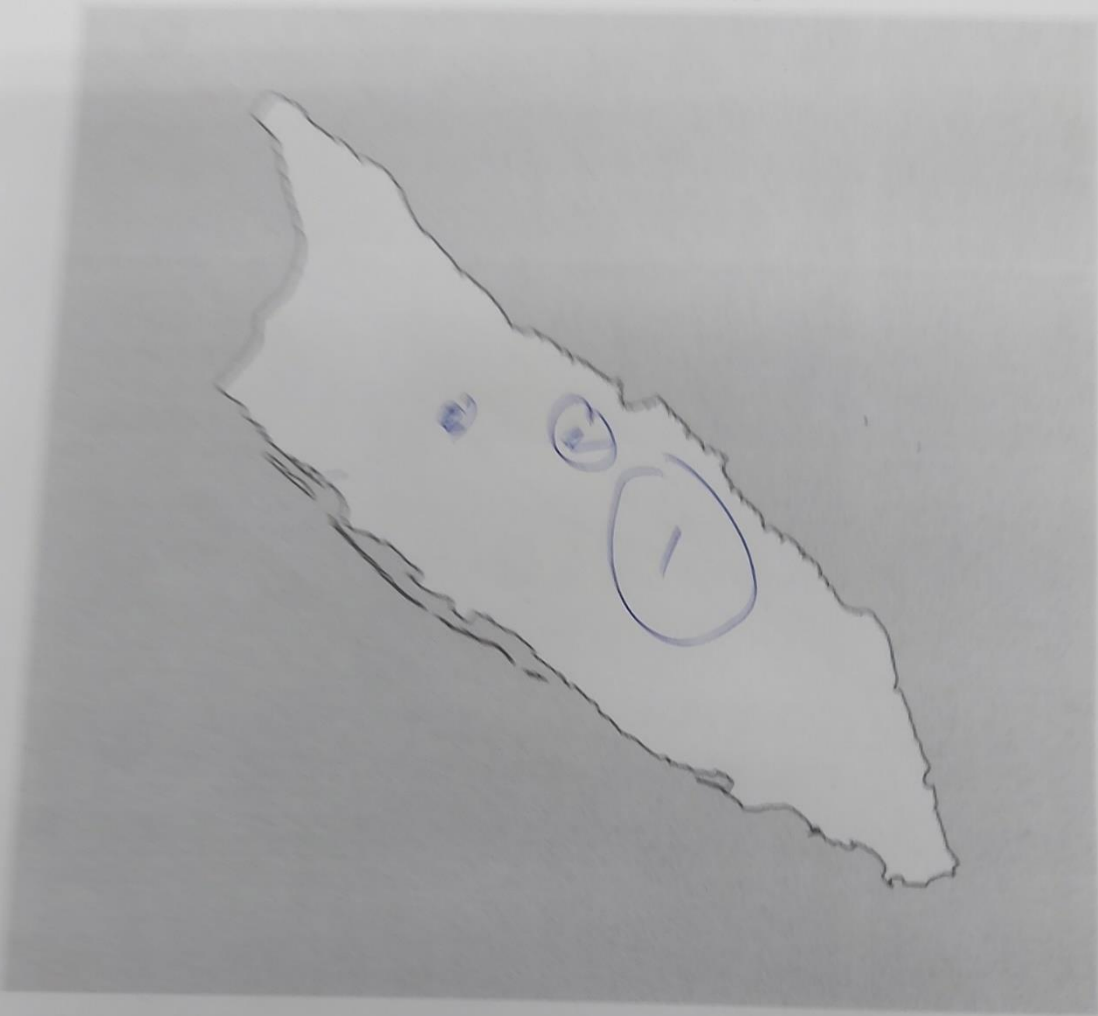
Seascape tourists mainly go to:



Explain why here:

From Hotel area which area beaches to North Shore sightseeing and touring boats to as well as the National Park in the middle

Terrestrial landscapes tourists mainly go to:



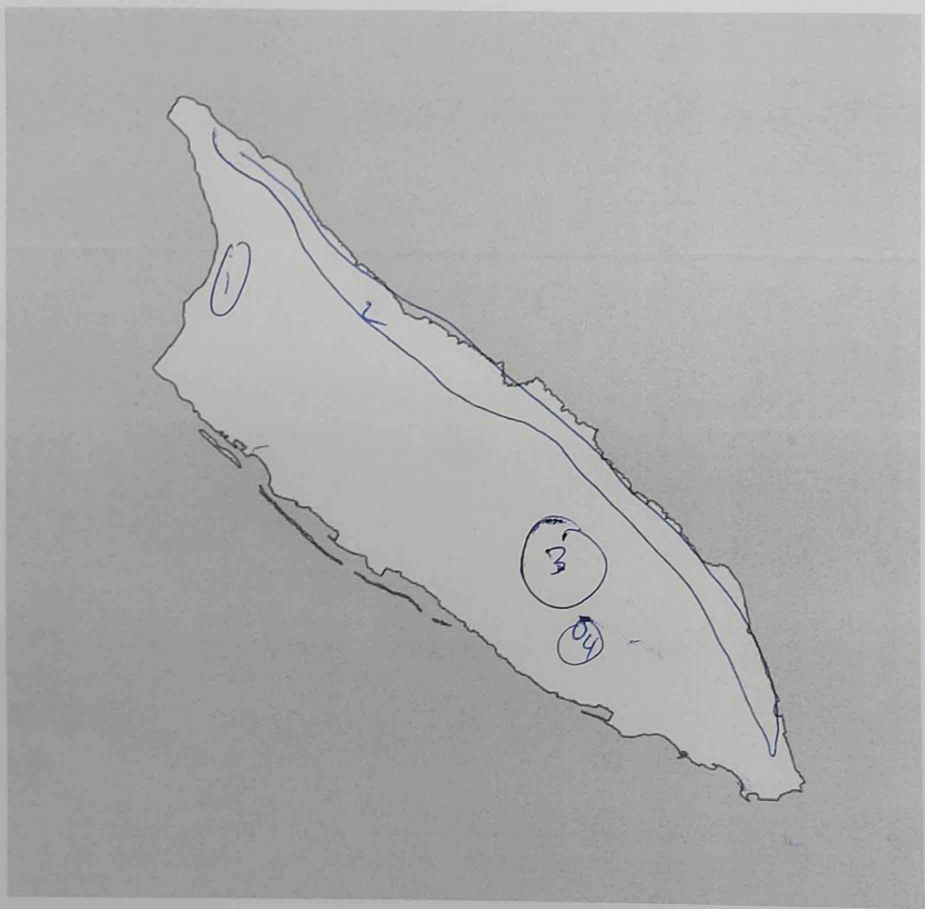
Explain why here:

1 National Park

2 Ayo Rock Formation

3 Casibari Rock Formations

Wildlife tourists mainly go to:




Explain why here:

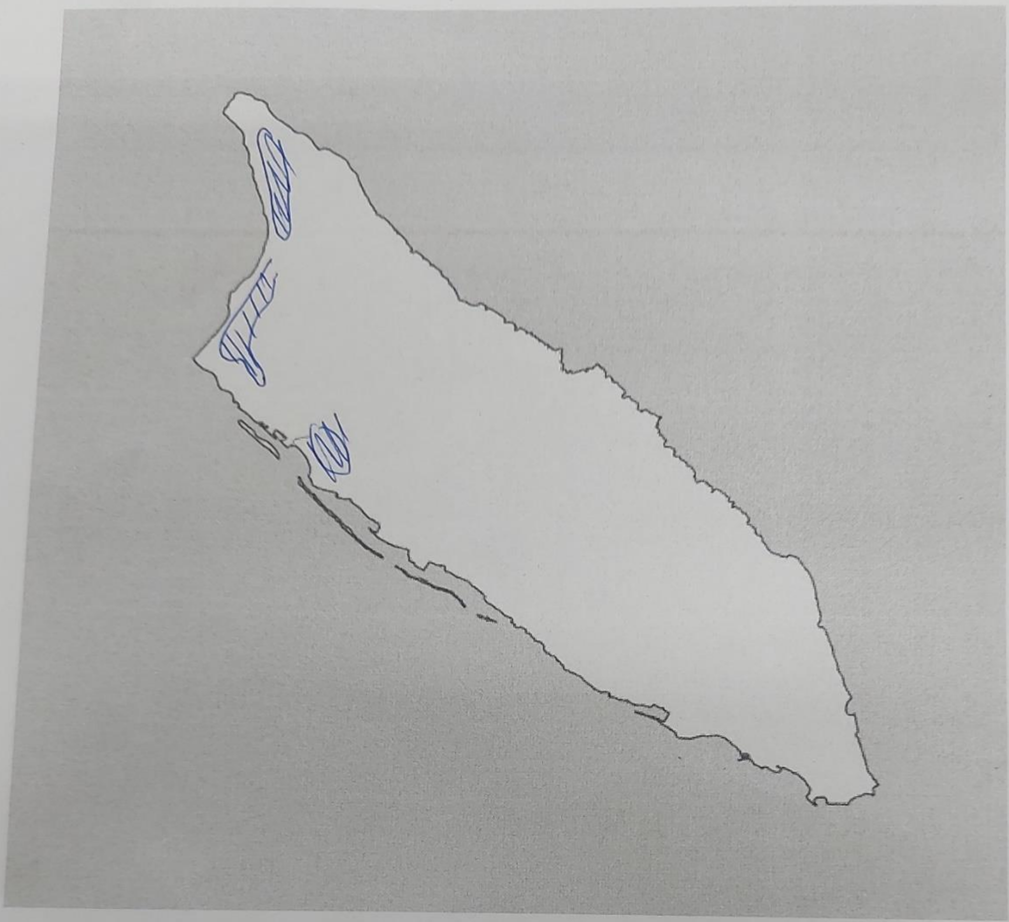
(1) Bubali Plas → Migratory Bird refuge

(2) North Shore landscape

(3) National Park

(4) Donkey Sanctuary

Other, mainly urban, types of tourists would mainly go to:



Explain why here:

* Hotel zone with the many shopping
and comfort F&B establishments
as well as down town

## Respondent 4

First we start with some basic questions.

What is your gender? …………………Female……………

How old are you? ……………………25………………

In which city do you live? ………………………Oranjestad…………………

In what part of Aruba do you live? ………………………Sabana Blanco……………

For how long have you lived in Aruba? ……………………23 years……………………

For what organization do you work? ……I am an MSc student at WUR………

What is your profession?

Environmental sciences student.

How well do you know the Aruba 1=not, 5=very well

1 2 3 4 5

☐ ☐ ☐ ☒ ☐

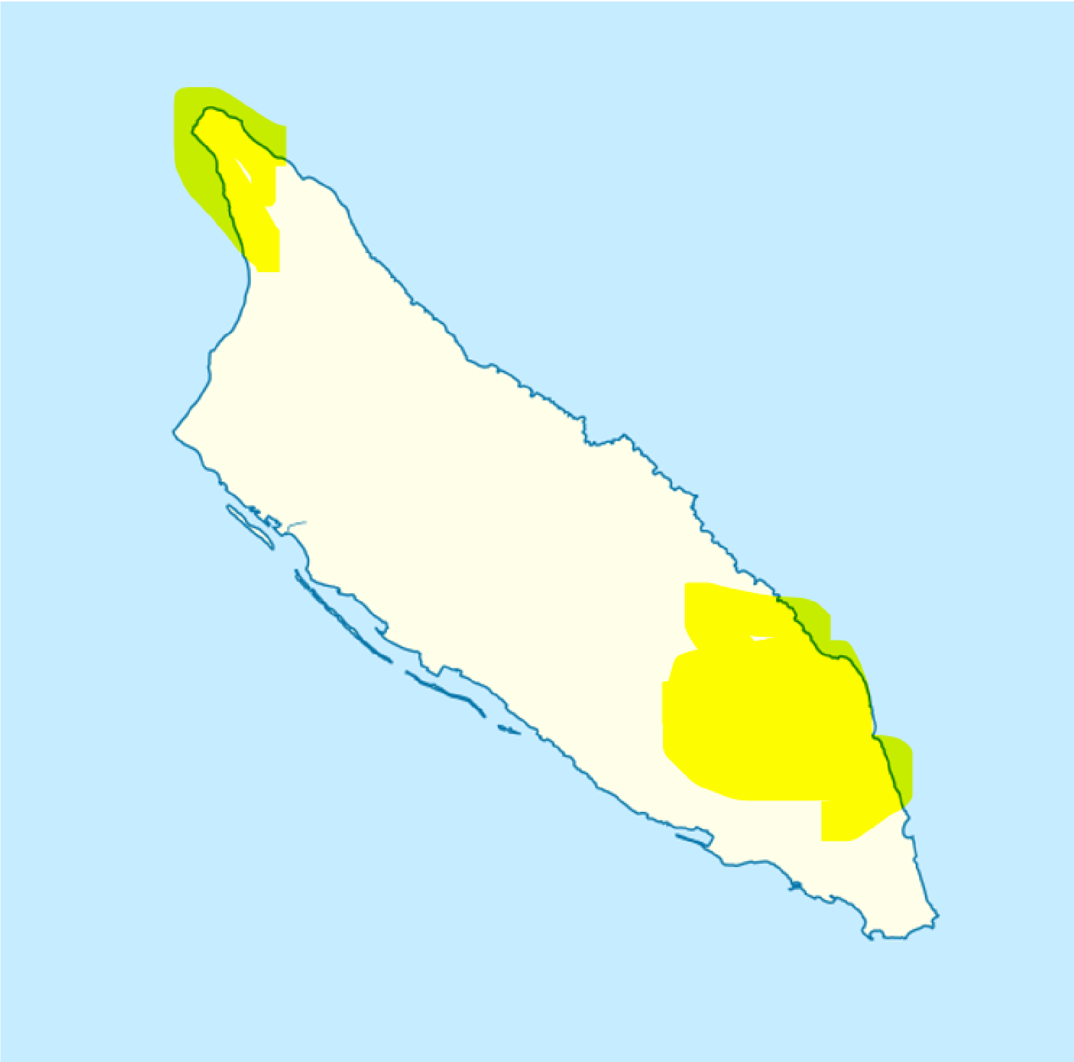
Coastal tourists go mainly to:

Explain why here: Examples of Eagle Beach, Boca Catalina, Renaissance Islands' beaches, Baby Beach in San Nicolas, the Palm Island, etc, (except near the landfill). Also, the Natural Pool is famous for off-road tours. However, toward the North side, the waters are a bit rougher, and it is unsafe to swim in most parts.

Seascape tourists mainly go to:

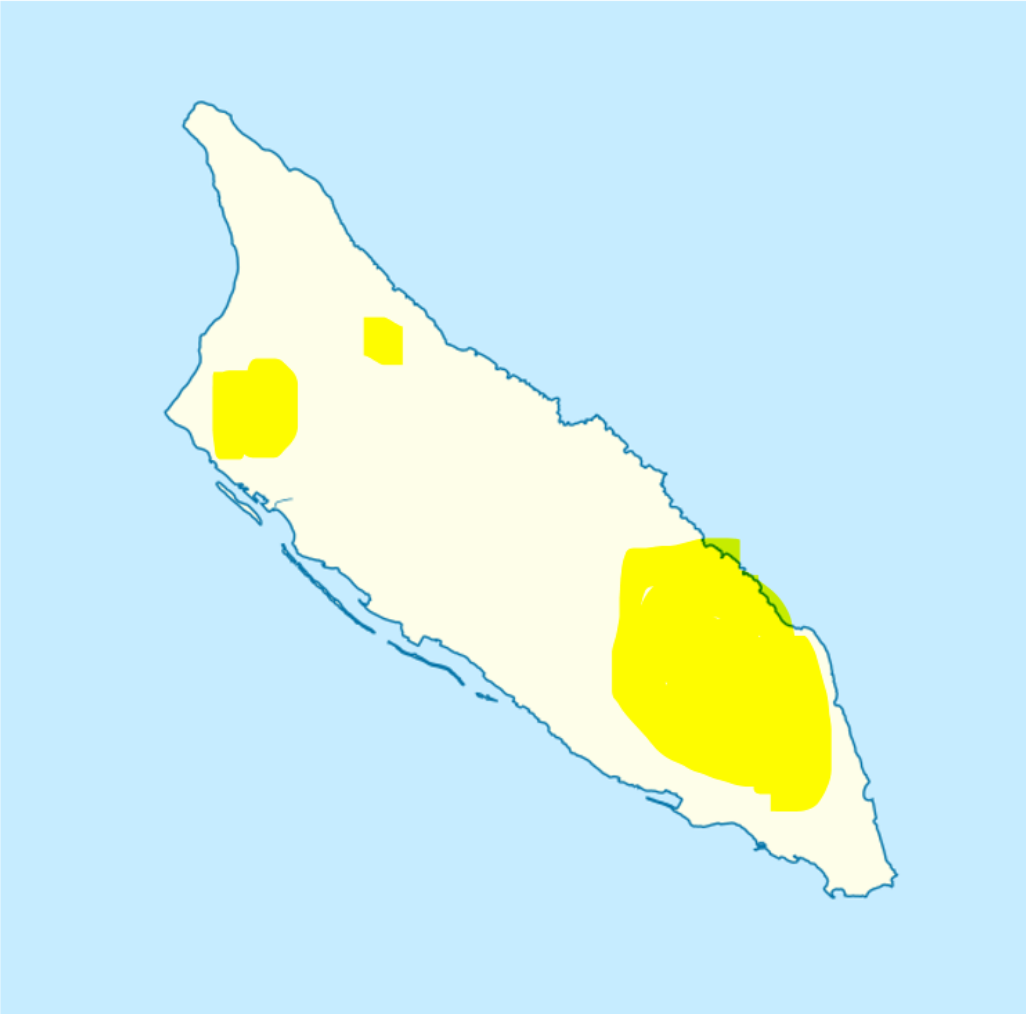
Explain why here: Every common beach includes optional sports activities, and beaches such as Fisherman's Hut and Grapefield are known to have kite- and windsurfing. For views, however, there are no beaches that tourists do not visit, except near the landfill (I would assume).

Terrestrial landscapes tourists mainly go to:



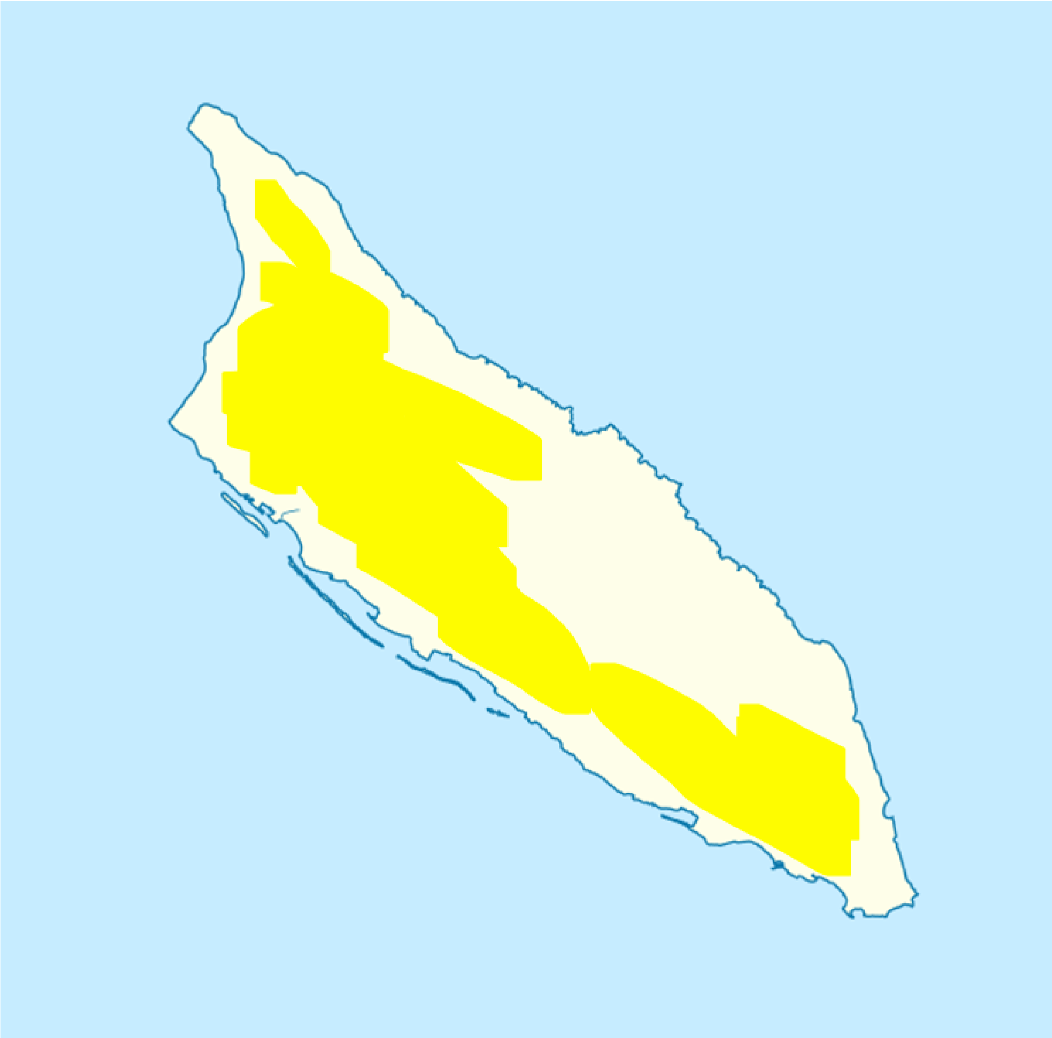
Explain why here: Arikok National Park, Bubali Wetlands, Boca Prins Sand Dunes, and Casibari Rock Formations offer terrestrial landscapes.

Wildlife tourists mainly go to:



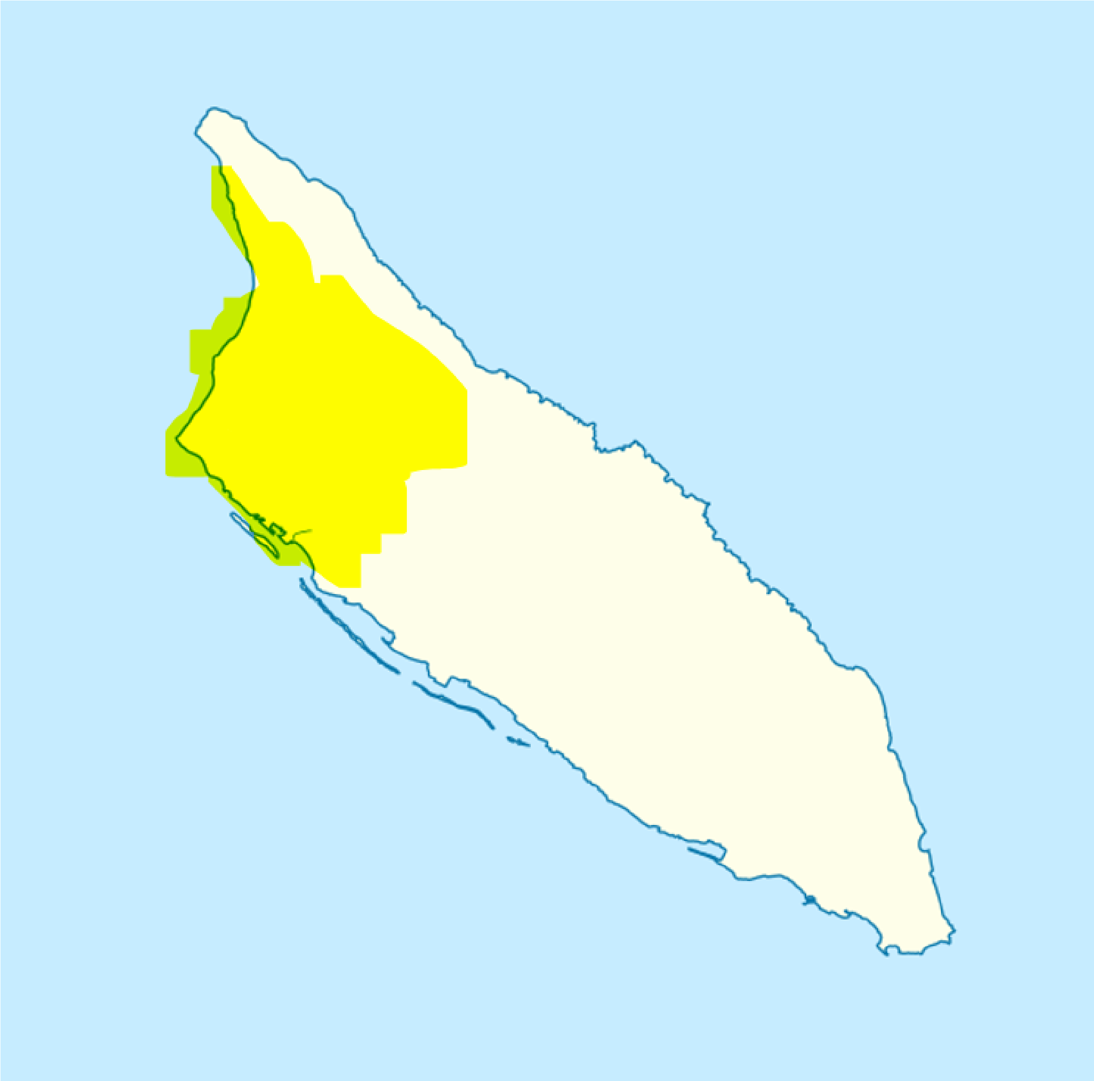
Explain why here: Arikok National Park, Bubali Wetlands, Phillip's Animal Garden, and the Butterfly Farm are common places to encounter wildlife.

Other, mainly urban, types of tourists would mainly go to:



Explain why here: Other than the low- and high-rise hotels, there are Airbnbs scattered around the island in addition to smaller-scale hotels and rentable apartment complexes.

Resort visitors/party tourists mainly go to:



Explain why here: Main locations of low- and high-rise hotels, and a strip that contains many nightlife activities.

How well do the study outcomes resemble the real world situation? 1=no, 5=very well

1 2 3 4 5

☐ ☐ ☒ ☐ ☐

Are the study outcomes useful for you or for your organization? 1=no, 5=very useful

1 2 3 4 5

☐ ☐ ☐ ☒ ☐

For what purposes could you use the study outcomes?

It may be handy for those who would like to research a similar topic in Aruba or any organization or persons who would like to gain more insight into tourism activity.

Do you have any other comments or suggestions?

I don't think so, but best of luck with your research! I wondered whether Instagram or Facebook would affect the results compared to Flickr nowadays, though.

## Respondent 5

First we start with some basic questions.

What is your gender? Female

How old are you? 35 ……...

In which city do you live? Noord……………………………………

In what part of Aruba do you live? Noord…

For how long have you lived in Aruba? Born and raised in Aruba, was abroad for university then returned and have been working for 8 years in Aruba

For what organization do you work? ATA…………………………

What is your profession?

Sr. Product Specialist at the ATA
………………………………………………………………………………………………………

How well do you know Aruba 1=not, 5=very well

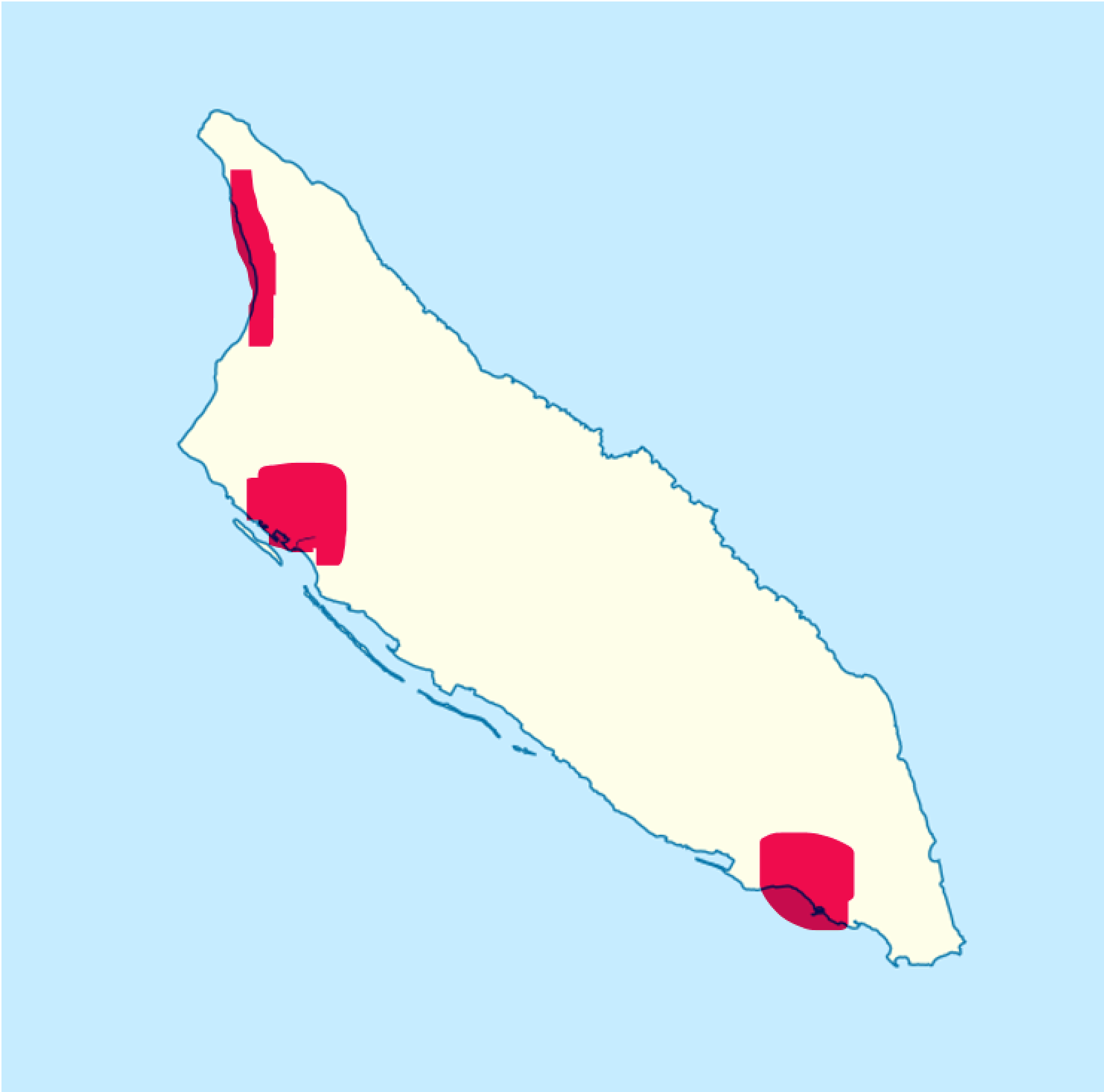
1 2 3 4 5

☐ ☐ ☐ ☐ x

Coastal tourists go mainly to:

Explain why here: Yellow highlights: Hotel strip and popular beaches and or reef islands (renaissance, palm island) . Orange highlight: tours or visitors exploring the north coast area on their own (Jeeps, UTV/ATV, bus, hiking, horseback, mountainbiking, etc.). Also the National Park Arikok's coastal area is highly visited.

Resort visitors/party tourists mainly go to:

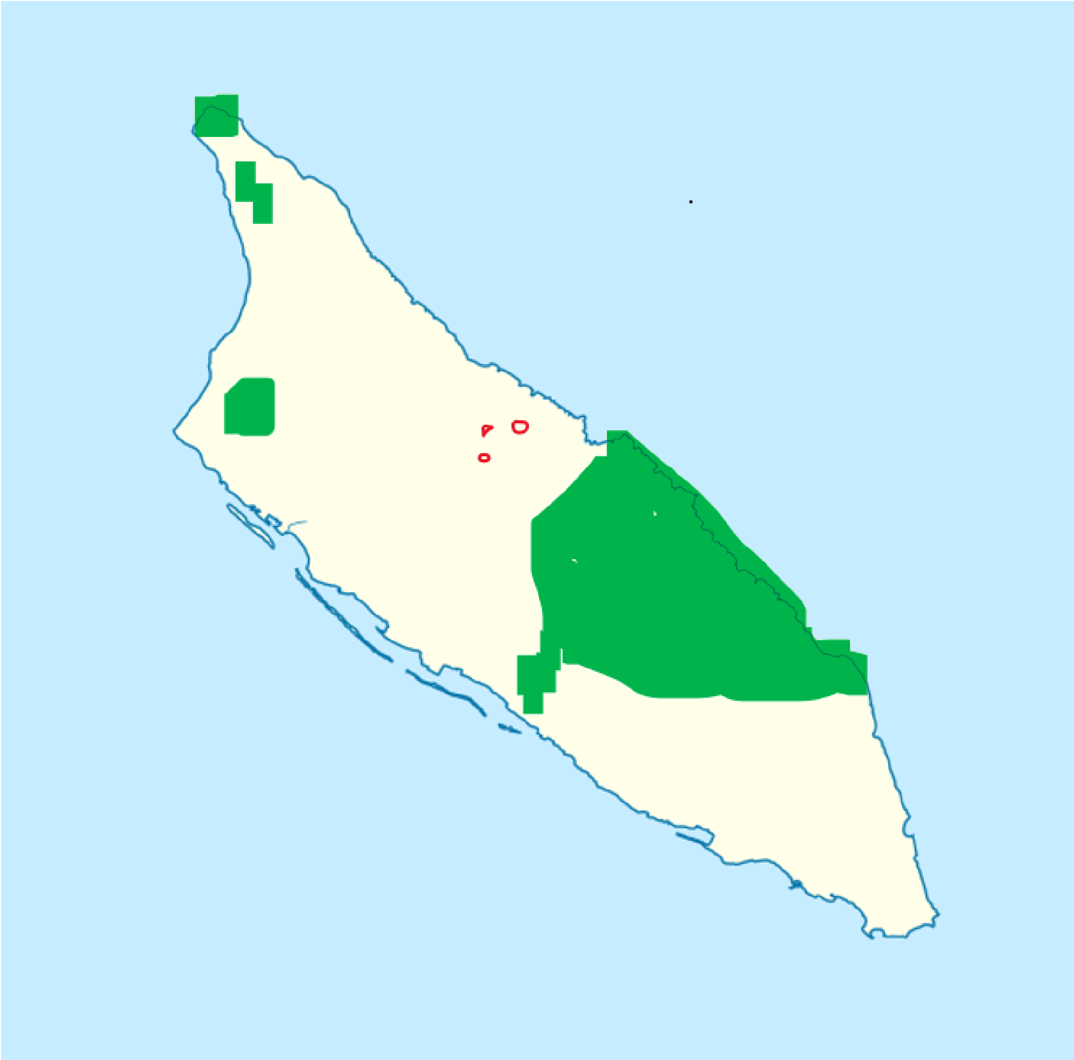


Explain why here: Same areas as the coastal tourists (as the resort tourist would also go to all the popular beaches and partake in island tours etc). Additionally they would party in the palm beach area along the High-rise hotels strip, and visit the Oranjestad and San Nicolas city centers

Seascape tourists mainly go to:

Explain why here: Same highlights as the coastal tourists. Additionally -: The watersports are mainly located along the hotel strip (pink highlights) The scenic seascape views are appreciated at every beach, cliff, or coast line. Some more popular than others.

Terrestrial landscapes tourists mainly go to:



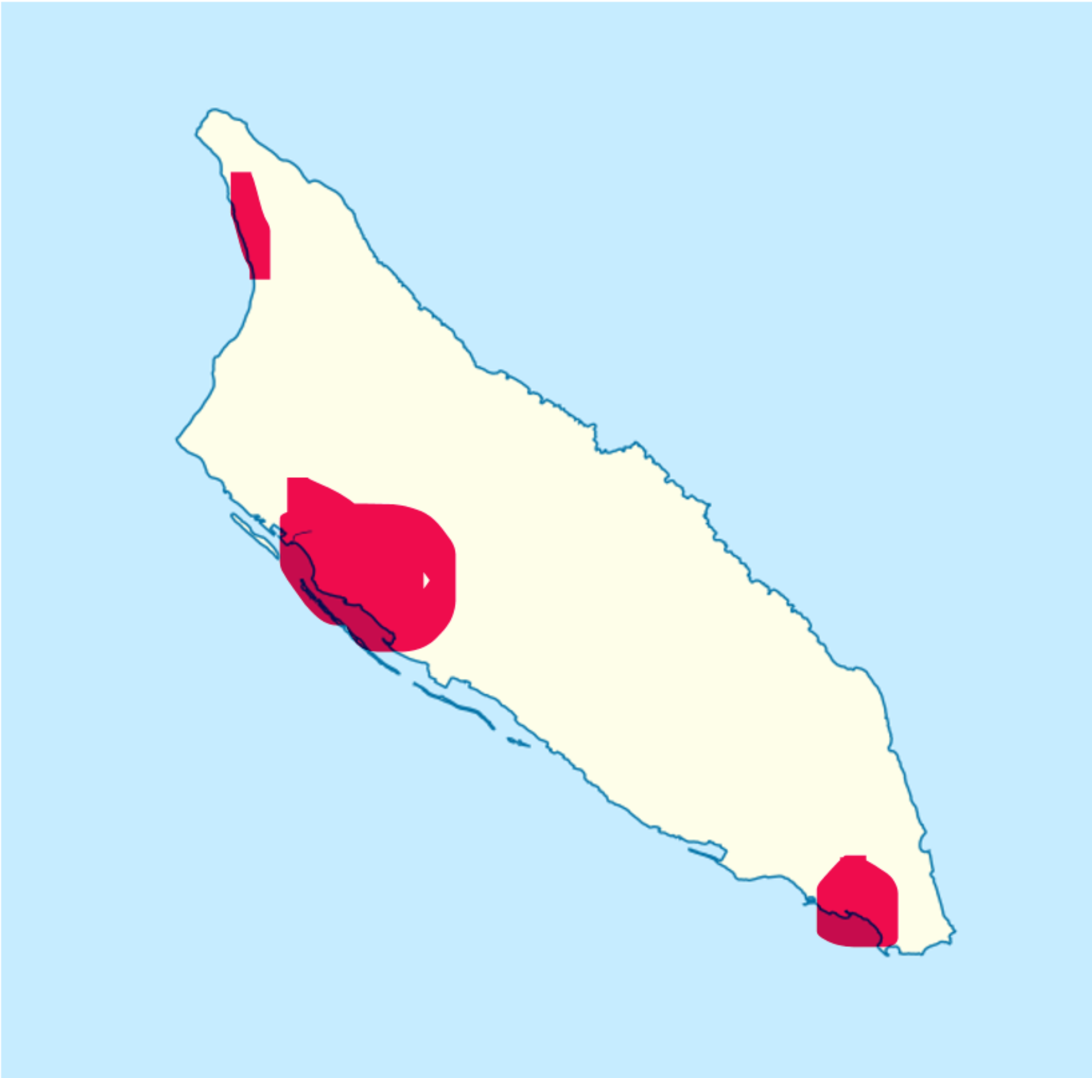
Explain why here: Arikok National Park, Ayo rock formation, Casibari rock formation, Hooiberg, bubaliplas wetlands, spaanslagoen area and wetlands, the wetlands in Noord and terra cora, the sasarawichi dunes.

Wildlife tourists mainly go to:

Explain why here:

All the green highlights of the previous question: bird watching in all the wetlands, birds and other terrestrial wildlife mainly at the Arikok National Park. Not sure is applicable: there is an animal sanctuary “phillips animal garden” in Noord

Other, mainly urban, types of tourists would mainly go to:



Explain why here: The urban type of tourist would mainly visit Oranjestad and San Nicolas in addition to the palm beach area. As it is a small island most “urban tourists” that choose Aruba would still visit most of the touristic highlights and beaches as well

How well do the study outcomes resemble the real world situation? 1=no, 5=very well

1 2 3 4 5

☐ ☐ ☐ x ☐

Are the study outcomes useful for you or for your organization? 1=no, 5=very useful

1 2 3 4 5

☐ ☐ ☐ x ☐

For what purposes could you use the study outcomes?

By understanding the preferences of different types of tourists, ATA can work with local businesses to develop new tourism products and experiences that cater to these specific interests. For instance, if there's a significant demand among coastal tourists, ATA could encourage the development of more beachfront accommodations or water-based activities.

..........................................................................................................................

..........................................................................................................................

Do you have any other comments or suggestions?

..........................................................................................................................

..........................................................................................................................

..........................................................................................................................

..........................................................................................................................

..........................................................................................................................

..........................................................................................................................

..........................................................................................................................

..........................................................................................................................

# Respondent 6

## Appendix D Survey tourism agencies part 1

Thank you for participating in this research. This research gives an answer to the question: 'Mapping the spatial distribution of tourism in Aruba through social media'. In this research, images from the social media platform Flickr are extracted to map tourism in Aruba, distinguishing between different types of tourists. Ultimately, using this method, I only catch part of the tourists who visit Aruba, namely those who actually post a photo on Flickr. That is why the help of your tourism expertise is needed to have a critical assessment on the outcomes.

First we start with some basic questions.

What is your gender? Female

How old are you? 38

In which city do you live? Oranjestad

In what part of Aruba do you live? Paradera

For how long have you lived in Aruba? Almost 9 years

For what organization do you work? ATA

What is your profession?

Intern marketing director

How well do you know the Aruba 1=not, 5=very well

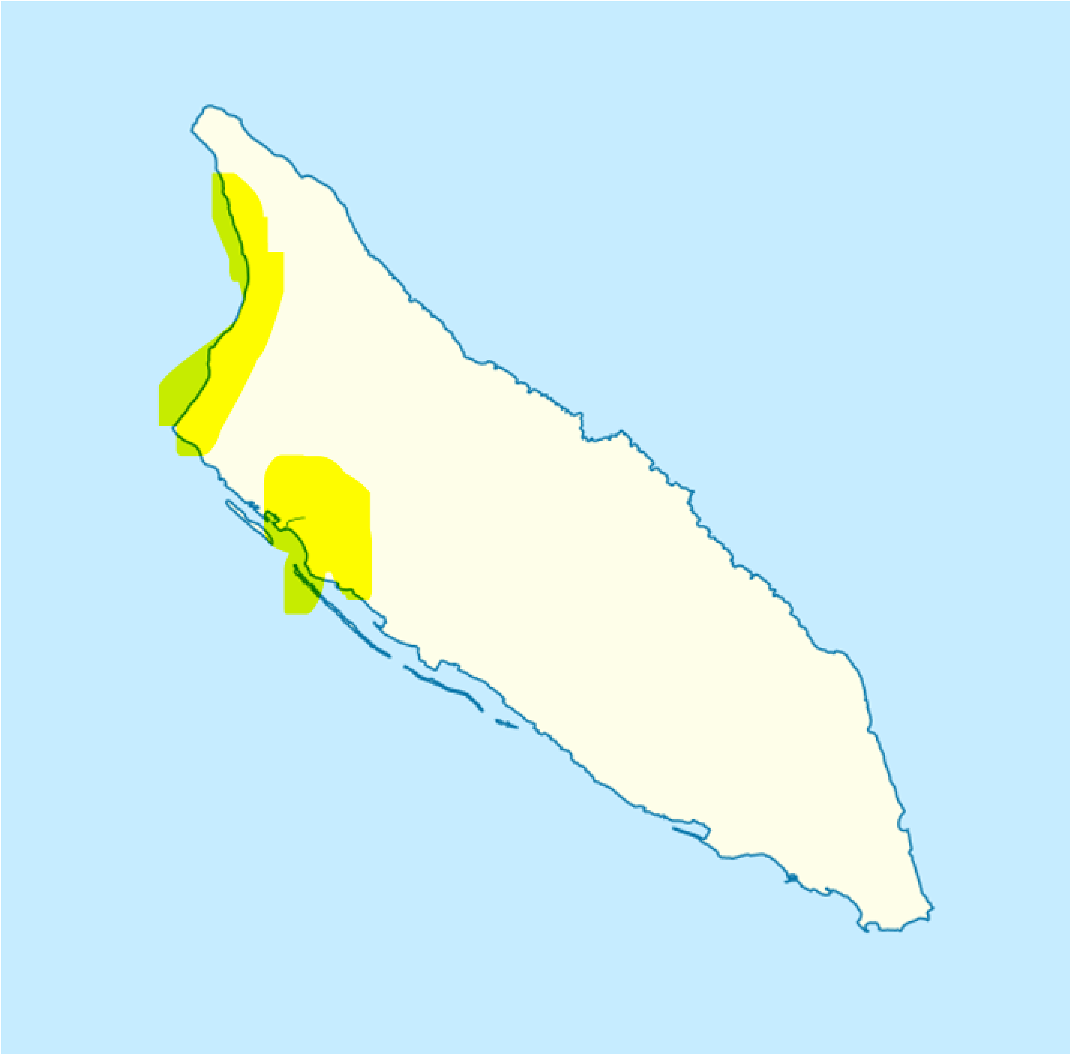
| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| ☐ | ☐ | ☐ | ☐ | ☐ |

Coastal tourists go mainly to:

Explain why here:

Eagle Beach, Palm Beach, Baby Beach, and Arashi Beach for sunbathing, swimming, snorkeling, and water sports activities. They may also explore natural attractions like Arikok National Park, which features rugged coastline and picturesque vistas, or take part in various water-based excursions such as snorkeling tours, sailing trips, or sunset cruises

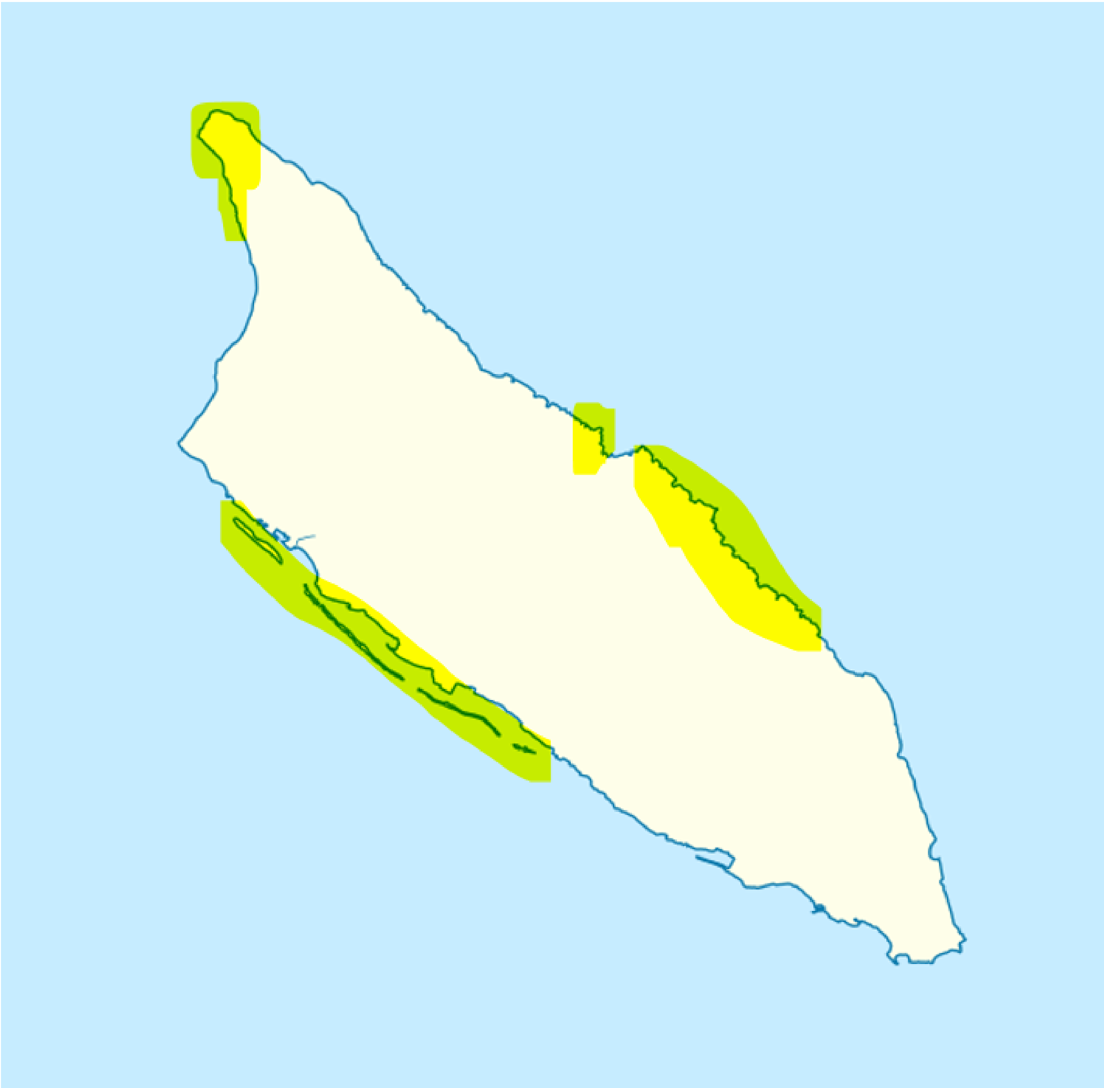
Resort visitors/party tourists mainly go to:



Explain why here:

The main hub for nightlife in Aruba is Oranjestad, where visitors can find numerous bars, clubs, and casinos. Specifically, areas such as Palm Beach and the high-rise hotel district are popular among tourists seeking lively nightlife experiences, with a plethora of beach bars, nightclubs, and entertainment venues. Nearby Paseo Herencia and Palm Beach Plaza offer shopping, dining, and entertainment options for those looking to extend their evenings after enjoying the beach and resort activities during the day.

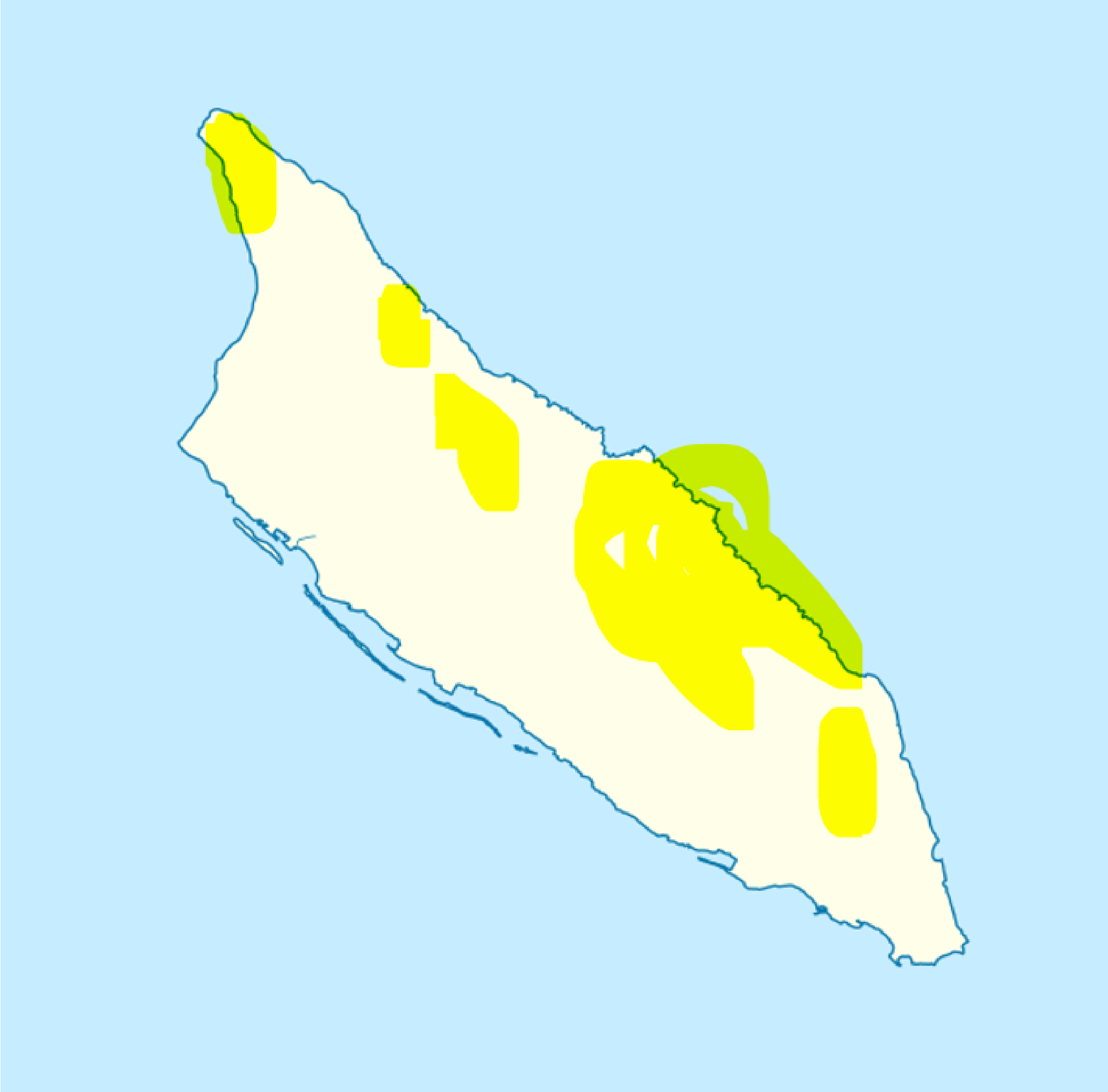
Seascape tourists mainly go to:



Explain why here:

Natural Bridges, California Lighthouse, Arikok National Park, Boca Catalina, Malmok Beach, De Palm Island, and enjoy catamaran and sailing tours along the coastline for scenic views and water activities

Terrestrial landscapes tourists mainly go to:



Explain why here:

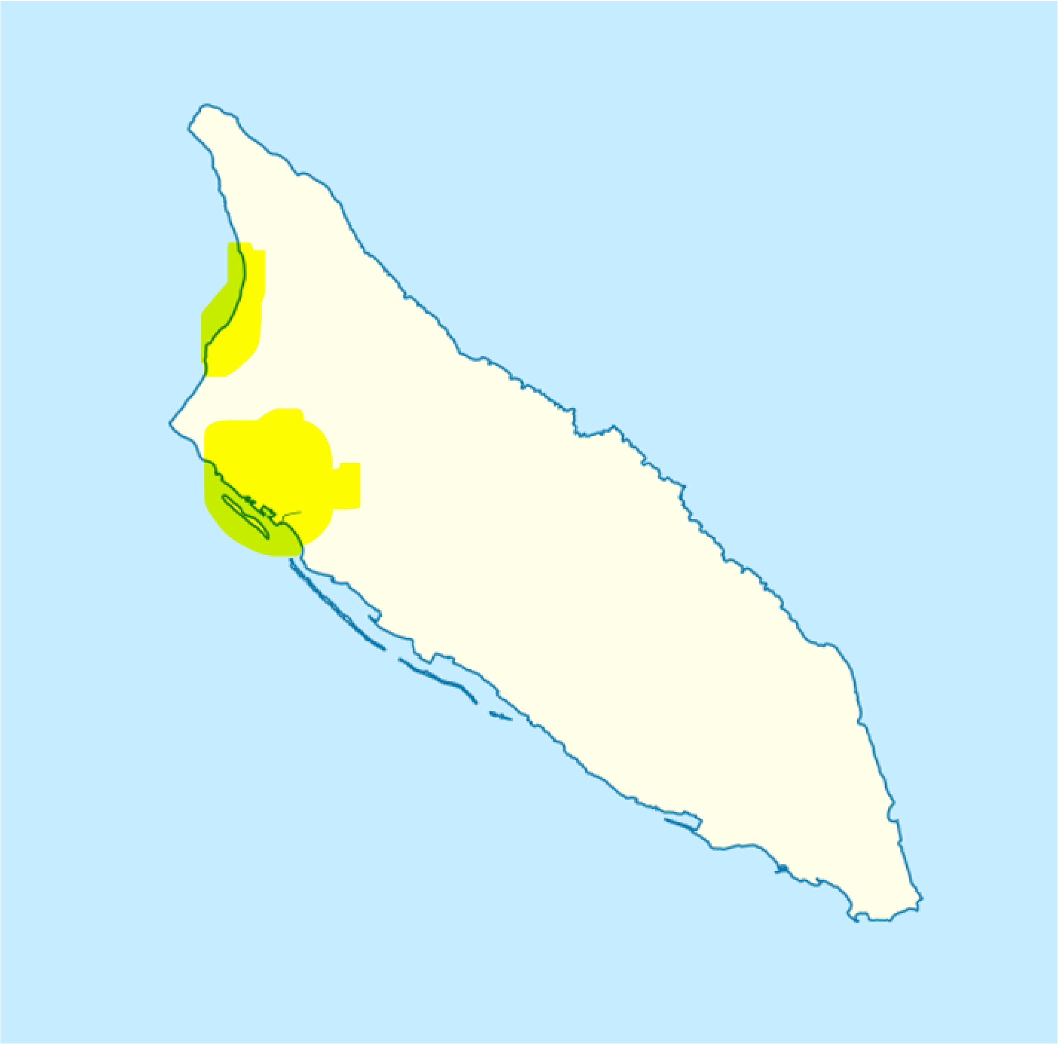
Arikok National Park, which offers hiking trails, caves, and unique desert landscapes. They may also visit landmarks like the California Lighthouse, rock formation and the Alto Vista Chapel. Or taking off-road tours to discover Aruba's rugged interior terrain and hidden gems.

Wildlife tourists mainly go to:

Explain why here:

Visit Arikok National Park, which is home to diverse flora and fauna, including native species like the Aruban burrowing owl, Aruban whiptail lizard, and Aruban rattlesnake. They may also explore the park's coastal areas to observe marine life such as sea turtles, iguanas, and various bird species. Also the Additionally, wildlife enthusiasts may participate in guided tours focused on birdwatching, nature walks, and educational programs aimed at conservation efforts on the island

Other, mainly urban, types of tourists would mainly go to:



Explain why here:

The capital city of Oranjestad, where they can explore its colourful Dutch colonial architecture, bustling markets, and vibrant street art scene. Attractions such as Fort Zoutman, home to the Historical Museum of Aruba, and the Parliament building draw urban tourists interested in the island's history and culture. Shopping districts like Main Street and Renaissance Mall offer a variety of shops, restaurants, and entertainment options. Additionally, visitors can experience Aruba's urban nightlife by exploring the city's bars, clubs, and casinos.

How well do the study outcomes resemble the real world situation? 1=no, 5=very well

1 2 3 4 5

☐ ☐ ☐ ☐ ☐

Are the study outcomes useful for you or for your organization? 1=no, 5=very useful

1 2 3 4 5

☐ ☐ ☐ ☐ ☐

For what purposes could you use the study outcomes?

to inform marketing strategies, guide infrastructure development, plan tourism initiatives, prioritize conservation efforts, enhance visitor experiences, and foster collaboration with stakeholders.

Do you have any other comments or suggestions?

Not sure if Flickr is the best medium to map tourism

## Respondent 7

First we start with some basic questions.

What is your gender? Female

How old are you? 54

In which city do you live? Mdko

In what part of Aruba do you live? Noord

For how long have you lived in Aruba? 46 jaar

For what organization do you work? Vertegenwoordiging van Nederland in Aruba

What is your profession?

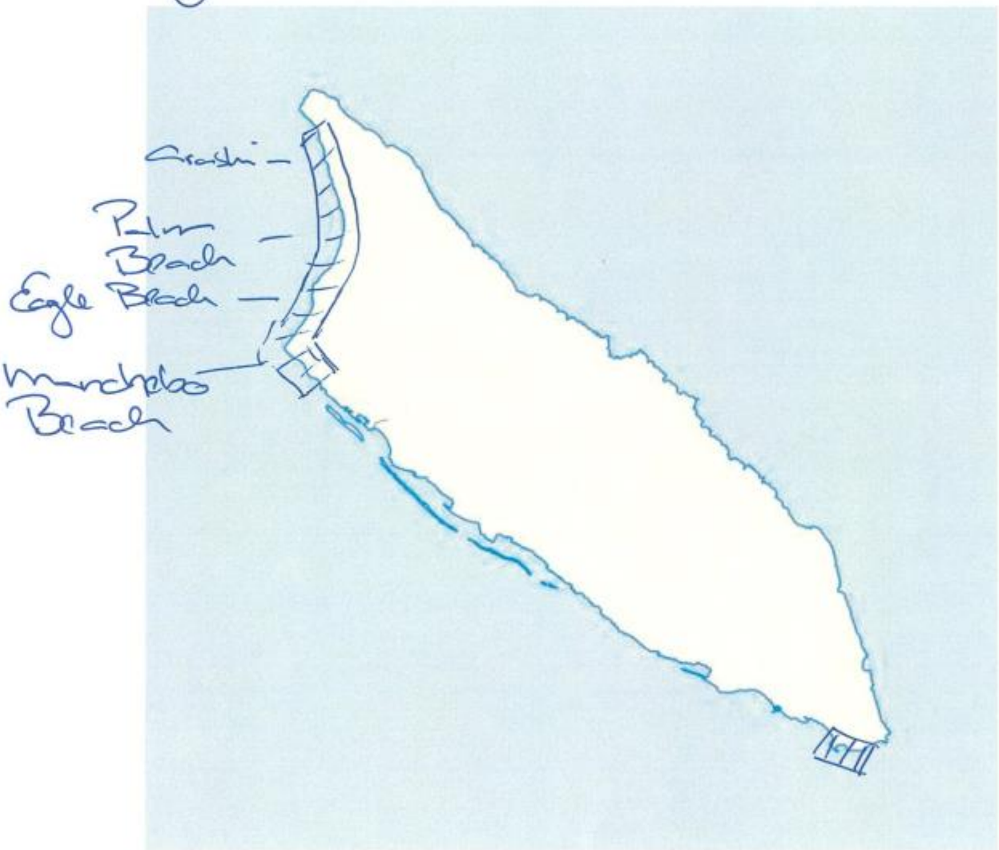
policy support officer

How well do you know the Aruba 1=not, 5=very well

1 2 3 4 5

☐ ☐ ☐ ☐ ☒

Coastal tourists go mainly to:




Explain why here:

- Grashi beach
- Palm Beach
- Eagle Beach
- Manchebo Beach
- Baby Beach

the most popular beaches

2

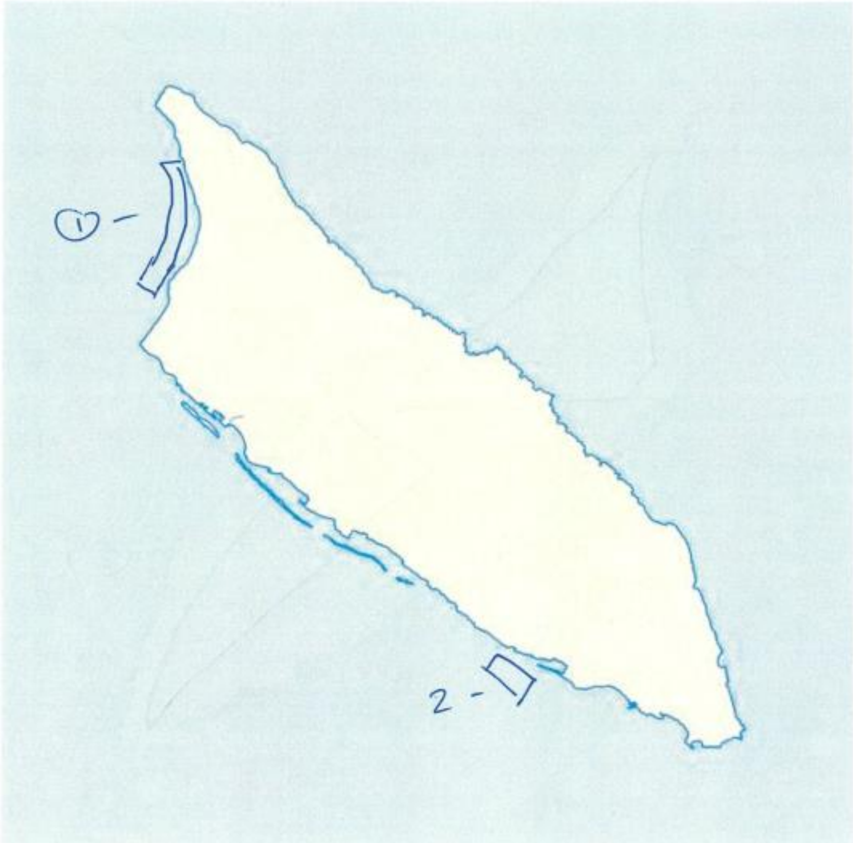
Resort visitors/party tourists mainly go to:

Explain why here:

Close to the beaches in Noord and Oranjestad.

③

Seascape tourists mainly go to:



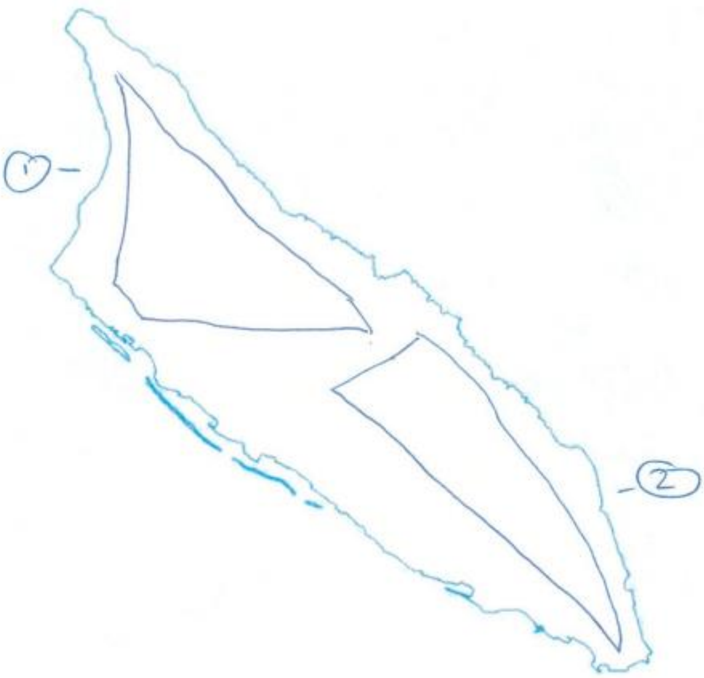

Explain why here:

① Watersports

② Diving

4

Terrestrial landscapes tourists mainly go to:



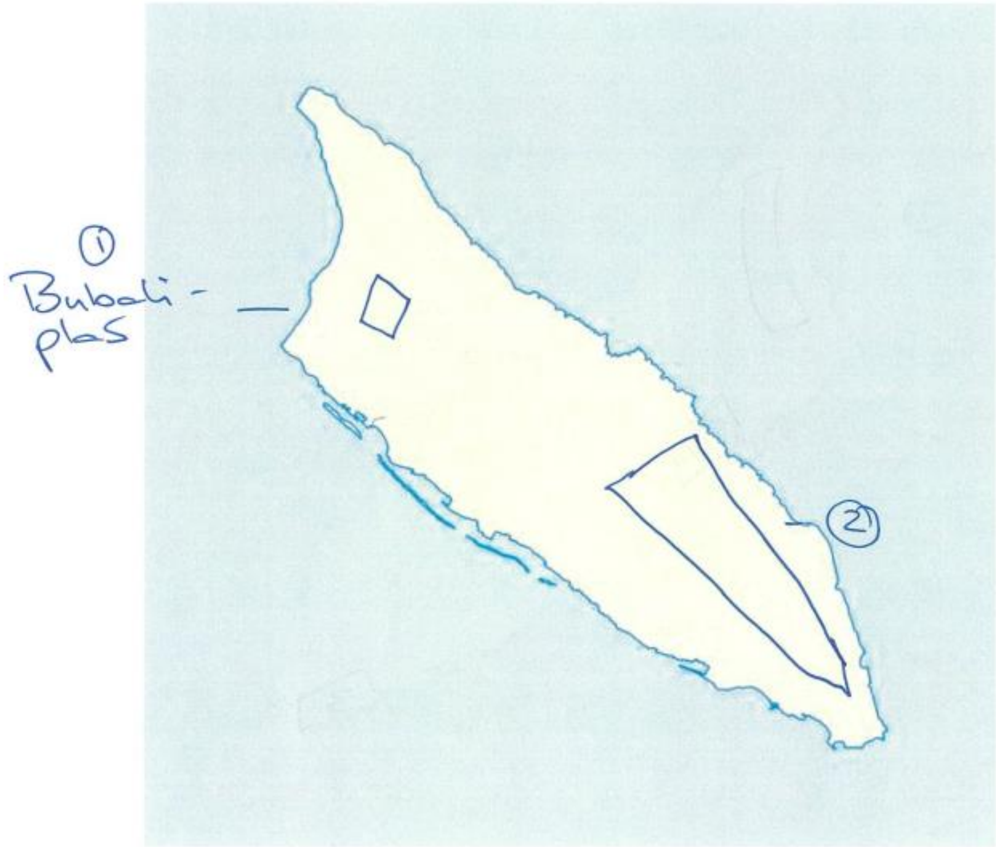

Explain why here:

1 Off the road terrain vehicles

2 Parke Nacional Arikok

5

Wildlife tourists mainly go to:



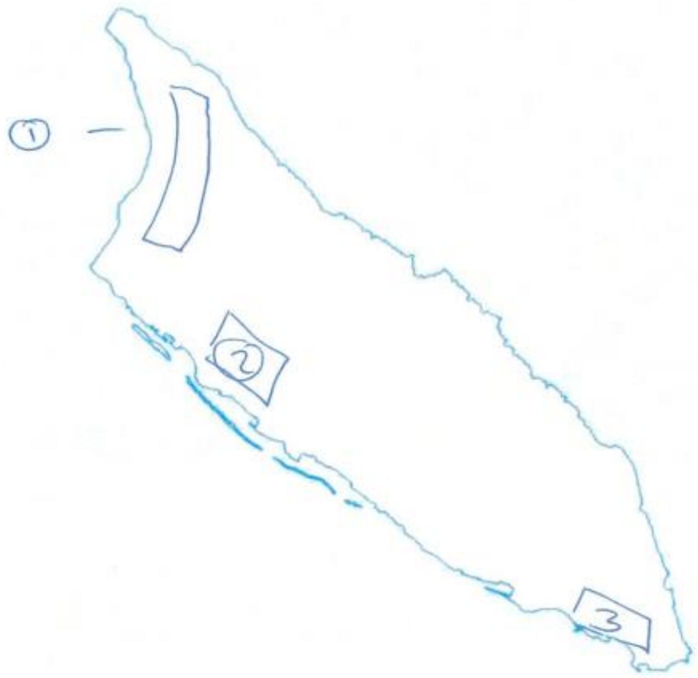

Explain why here:

① Bubali plas - birds habitat

② Parke Nacional Arikok -

6

Other, mainly urban, types of tourists would mainly go to:




Explain why here:

1- Noord - near hotels, beaches, restaurants
2- Oranjestad, down town - cruise port
3- San Nicolas - art walk

How well do the study outcomes resemble the real world situation? 1=no, 5=very well

1 2 3 4 5 *

☐ ☐ ☐ ☒ ☐

Are the study outcomes useful for you or for your organization? 1=no, 5=very useful

1 2 3 4 5

☐ ☒ ☐ ☐ ☐

For what purposes could you use the study outcomes?

None, but it is good to know the outcome. I think it would be important to do this periodically to see if it shifts.

* Het is opvallend dat de coastal tourists meer gaan a.d. ruwe zijde van Aruba en minder a.d. stranden waar ze zwemmen.

Do you have any other comments or suggestions?

Succes!