Creating insights in Tourism with Flickr Photography

Visualizing and Analysing Spatial and Temporal patterns in Venice

MASTER THESIS: ROBERT BEKKER

MSC ENVIRONMENTAL AND INFRASTRUCTURE PLANNING

rijksuniversiteit groningen

faculteit ruimtelijke wetenschappen

Colophon

Abstract

With the increase in urban tourism around the world planners stand before a new challenge. To keep cities sustainable and decrease the effects of urban strains such as mass-tourism. To achieve this, planners and decisionmakers must implement environmentally friendly urban developments. These decisions, however, require more insights to implement more informed decision making. This includes the whereabouts of tourists within the urban environment but also the identification of popular point of interests, throughout multiple seasons. This thesis, therefore, focusses on spatial patterns and temporal distributions of tourists within a case-study of Venice. By analysing and visualizing geotagged photographs from social media site Flickr maps and graphs are created, which show urban activity patterns over multiple time-intervals. Additionally, are pathways constructed that track routes throughout a city over one or multiple days. All this is done by exploring and describing a method, which uses big data that could ultimately improve the way we plan today.

Table of Contents

List of figures

List of tables

I.   Background

Venice is worldwide known as a tourist destination thanks to its rich history as well as her beauty experienced in architecture and artwork. Once established as a major financial and maritime power, becoming one of the world's most important centres of commerce. Venice has become a place of architecture and culture influenced through many differing periods. With the many landmarks and sightseeing opportunities, the city beholds, Venice has become an increasingly popular destination among tourists. Even so much, that the city is struggling to cope with the influx of travellers. This struggle, as Edwards (2017) describes, stems from the cheap flights, huge cruise ships, as well as the cities Instagram appeal that attract so many tourists. Here, the main perpetrators are the Doge's Palace, Canal Grande, and the Basilica di San Marco, which are on top of the list of many traveling guides and agencies. This leads to an incredible number of visitors of up to 28 million tourists per year who flood the city of Venice (Giuffrida, 2017).

And this has not gone without consequences, as the population is dwindling as the surge in number of bed & breakfasts among other tourists facilities appear, making it difficult for residents to find a home as housing prices rise (Edwards, 2017; Giuffrida, 2017). This growing tension and frustration with the visitors ultimately led to 2.000 Venetians marching against the tourism industry. This stemmed from the main concerns the residents of Venice must cope with such as: the eroded quality of life, the damaging of the environment, and the drive away of residents from the historic centre (Giuffrida, 2017). In addition, has the Mayor of Venice, Luigi Brugnaro, been campaigning for more regulation in the city to regulate the influx of tourists (Persio, 2018). Leading to the city launching the #EnjoyRespectVenezia campaign to increase the public awareness with infographics to help inform about the crowded areas, with limited success (Persio, 2018). However, despite the hatred of many residents towards the tourists, there are locals that understand that the city thrives on tourism, which calls for further insight and regulation on how to improve the city despite the influx of visitors.

This thesis will therefore explore tourists' patterns in both space and time, by using new ways of data collection methods. Here, the importance of both space- and time- coordinates when studying the problems of people, in this case tourists, is that the location in space cannot effectively be separated from the flow of time (Hägerstrand, 1970). Therefore, it is important to take both spatial patterns and temporal distributions into consideration with the goal of increasing knowledge and thus ultimately leading to a more informed decision-making process to help solve urban strains. Additionally, can the results encourage principles of sustainability for planners and policymakers in urban places (Gössling, 2017). Thus, a focus on environmentally friendly urban development could contribute to a proper balance between the increase in tourism and the quality of life of the residents of Venice.

II.   Research Problem

A few years ago, Ashworth and Page (2011), noted how little attention was being paid to how tourists use the urban environment, as there was little insight on the whereabouts of tourists in a city. However, to cope with the problems that arise from mass tourism, such as in Venice, it is important for planners and decisionmakers to have detailed information about the whereabouts of tourists in a city (Edwards et al., 2010; Girardin et al., 2007). Traditionally, were locations of people studied at specific moments in time by expansive and time-consuming fieldwork methods (Woods et al., 2013). And recent years have shown an increase in research from a multitude of scholars as it comes to analysing the behaviour of tourists in cities. However, detailed insights for many cities coping with mass-tourism is still little developed. Therefore, defining the spatial and temporal patterns of tourists' activities through social media applications, is important to better understand tourist motivations, the usage of urban space and the socio-economic impacts within urban destinations (Encalada et al., 2017; Kádár & Gede, 2013).

Nowadays, an unprecedented wealth of digital geographic information is in reach to planners and decisionmakers to help support in designs, spatial analysis, and decision-making (Massa & Campagna, 2014). This is provided in the form of Social Media, where every day, around the world, a vast amount of content is generated, which is expected to grow in the years to come (Croitoru et al., 2013; Gössling, 2017; Massa & Campagna, 2014; Xiang & Gretzel, 2010). Here, as argued by Campagna (2014), can complex problem-solving and decision-making in spatial planning be supported by an abundance of digital data. These, geospatial analytics, and computing functions can help planners and decisionmakers to achieve more informed and, sustainable decision-making as it comes to improving and adjusting the urban environment (Campagna, 2014). And as noted by Kádár and Gede (2013), can visualizing the geographical positions of photographs taken by tourist measure the tourist activity in urban spaces by defining spatial patterns, in space, as well as temporal distributions, in time. To ultimately, improve the planners decision-making as friction between locals and residents run high and proper solutions are in dire need.

Here, social media can help increase insights of tourists within the urban environment, however, it is not entirely clear in what way this is achievable. This leads to the **primary research question**: How can Social Media, such as Flickr, contribute in providing spatial and temporal patterns of tourism in Venice for improved urban planning?

Additionally, **secondary research questions** help answer the primary research question, and are formulated as follows:

- o   What are the advantages and disadvantages in creating spatial patterns and temporal distributions from social media networks in general?
- o   What techniques are recently studied in scientific literature to analyse and visualize spatial patterns and temporal distributions?
- o   Which steps need to be undertaken to identify and analyse spatial patterns and temporal distributions of tourists using Flickr photography?
- o   Which insights can be gained by spatial and temporal data from Flickr for planners and decisionmakers in Venice for refined urban planning?

III.   Structure of the Thesis

It remains unclear how cities can handle this new influx of urban tourists. Therefore, the aim of this study is to provide greater insights into urban tourism created from Flickr photography. By gathering user-generated geographical data research is conducted and an attempt is made in creating a better understanding of the whereabouts, in both space and time, of tourists within the urban environment. This is important to understand the tourist motivations, urban space usage and socio-economic impacts (Kádár & Gede, 2013). To ultimately, better inform planners in creating sustainable solutions in the urban environment.

The first chapter, the Theoretical Framework, will cover the generalities as well as the first two secondary research questions about the advantages and disadvantages in creating spatial and temporal patterns of social media network data, and it will cover some of the current methods available used by other researchers.

The second chapter, about the methodology, will include the third secondary research question. Here, an explanation of the multiple steps is undertaken to identify and ultimately analyse the extracted data. In addition, is a research framework included that shows graphically the multitude of steps taken.

The third chapter, results, will include the final secondary research question. Here, results created from social media metadata will be displayed. Additionally, will insights be discussed created from spatial patterns and temporal distributions.

At last, will the primary research question be discussed in the final chapter of this thesis. Here, the insights created from the secondary questions will help provide the answer and insights needed to conclude as well as discuss the main question to this thesis. In addition, will future recommendations be considered.

IV.　Theoretical Framework

With the remarkable success of online social media sites in recent years, a shift has occurred in the way people connect and share information (Croitoru et al., 2013). The internet, as Xiang and Gretzel (2010) describe, has fundamentally reshaped the way tourists plan for and consume travel. One of the reasons for this trend are the new ways in which information is distributed via the internet. Here, opportunities for both tourists as well as planners and scholars arise to become better informed and thus more knowledgeable. Additionally, as Encalada et al. (2017) notes, is the increase in online information access a great opportunity to better understand tourist behaviour within the urban environment. This results from a combination of tourists being avid for online content as well as tourists who share a great deal of information on social medias. Tourists have become increasingly abundant, especially in cities. It is thus important, as Encalada et al. (2017) emphasizes, to deal with this urban pressure, that will raise a variety of problems and important challenges, which can question the economic, social, and environmental sustainability of cities all over the world. However, as Gössling (2017) discusses have there only been a limited number of authors that discuss this interrelationship of information technologies, tourism, and possible sustainable solutions.

This chapter starts off with exploring the benefits and insights social media metadata gives to help improve urban planning. Additionally, will the rise of social media be explored as well as emerging tourism trends, and the relationship that exists with one another. It will further investigate the positives as well as negatives that hide behind social media applications. This is important to understand used techniques from related works, which could be beneficial to comprehend, create, analyse, and visualize spatial and temporal patterns. Lastly, does this chapter end with a display of the conceptual model. By visualizing the relationships of concepts mentioned within this chapter an understanding is created how they are influenced by one another.

A.　Benefits for Planning

A greater understanding of the places tourist visit, the time they spend and the services they utilize can help provide valuable information for many engaged in the management or study of tourism (Edwards et al., 2010). Additionally, as explained by Kádár and Gede (2013), can the distribution of tourist in urban environments be of relevant information for scholars within hospitality research, tourism studies and urban studies. With the global increase in urban tourism, planners stand before a challenge to keep cities liveable as well as sustainable. With the right tools and techniques, it is possible to create insights, which can help planners and decisionmakers to have an increased understanding to make more informed decisions. Planning includes, among other things, future-oriented activities, in which goal orientation, the selection of resources and possibilities, area development, the local regional and international scale level, and the individual behavior of actors within a spatial context are core themes (Woltjer et al., 2011). This, as Woltjer et al. (2011) discussed, is important as society cannot do without a well-considered spatial policy. With the rise of new social and technological developments such as social media, there are new opportunities that can help create alternative insights for planners to analyse tourist behaviour within the urban environment. These insights could in turn lead to more appropriate decision-making.

One of the insights as shown by Girardin et al. (2007), is the increased understanding in touristic pulses throughout a city. In addition, do Edwards et al. (2010) and Kisilevich et al. (2010) uncover human behaviours, by understanding tourists' spatial behaviour as well as the identification of points of interests within the urban environment by interpreting attractive areas. Moreover, was Sagl et al. (2012), able to create an understanding of how collective social activities can shape urban systems.

Lastly, did Massa and Campagna (2014) conclude in their research, that insights into opinions and perceptions can help improve the dialogue between planners, decisionmakers and citizens for improved feedback. However, it remains difficult to give clear-cut actions that insights such as these can give to help improve planning. With these insights it is, however, possible to increase the knowledge and information that is put into planning to ultimately improve the urban environment with environmentally friendly development. Especially, since the data available from social media applications is oftentimes enormous. In turn does big data lend itself well for analyzations and visualizations, which could help planners increase their understanding from another perspective.

B.   The Rise of Social Media

In recent years has social media attracted tourism researchers by generating interests in understanding the role of social media as well as transforming travel experiences of tourists (Xiang & Gretzel, 2010). With the rise of social media there is now lots of data on the internet that contains some form of geographical content and it is for the first time that we can observe human activities on a greater scale, which was impossible without the use of computers (Croitoru et al., 2013). It has since been discovered that social media photos have, in general, the capability to improve tourism-related research. Here, Sun et al. (2013) describe, the usefulness to investigate spatio-temporal human activities, which opens new possibilities for further location and event-based analysis. These social media sites contain billions of photos, which are publicly available with different kinds of useful metadata such as: image size, tags, titles and especially spatio-temporal information like, where and when a photo was taken (Kisilevich et al., 2010). Here, a multitude of researchers have shown multiple takes on doing research based on social media metadata. This includes collecting geographical spatial information about the whereabouts of the tourists themselves. All due to the introduction of location-based services in social media applications of smartphones and other devices that has enabled people to share their activity related choices in their virtual social networks. Providing an unprecedented amount of user-generated data on human movements and activities (Hasan et al., 2013).

This phenomenon of the deployment of location-based devices has led to a massive increase in the volume of records of where people have been and when they were there. And as Girardin et al. (2008) describe, are the analysis of spatio-temporal data of great value to urban planners and local authorities to study human behaviour. This opens new doorways for planners and policymakers to improve their city in a possible sustainable way (Ashworth & Page, 2010). As this data has the potential to impact many other areas including travel demand modelling, computing, urban planning, security, and health monitoring (Hasan et al., 2013). Therefore, a tremendous opportunity exists to develop fundamental tools to analyse large-scale spatial and temporal data that allows one to understand social and behavioural characteristics of the users of location-based services, which was not possible without the existence of social media applications and its ability to share data (Hasan et al., 2013).

C.   Tourism Trends

As Ashworth and Page (2011) describe, is urban tourism a world-wide form of tourism. However, despite its significance, has urban tourism only been imprecisely defined and vaguely demarcated with little development of a systematic structure of understanding (Ashworth & Page, 2011). Importantly, to plan, regulate and facilitate effectively, governments require an understanding of the functioning and needs of the many stakeholders within the urban environment (Edwards et al., 2008). Thus, the study of urban tourism and associated focus on urban tourists' behaviour is a growing area of interest as practitioners, researchers and policymakers seek to understand the phenomenon of tourism within

the urban environment (Edwards et al., 2008). This correlates with the study of Environmental Planning, where the analysis of urban and regional strategies and government approaches for highly dynamic and complex situations is combined with tools to help create more environmentally friendly urban and regional developments. Ultimately, planning is there to promote common interests such as health, safety, and prosperity, to monitor market failures, promote efficient developments and safeguard the rights of residents (Woltjer et al., 2011). Additionally, as Gössling (2017) discusses, exist a consensus that information technologies have far-reaching consequences in the ways in which tourism is applied and consumed. And as Xiang and Gretzel (2010) note, can tourism-related information be a stepping-stone for the development of better information systems in tourism.

However, to better understand what spatio-temporal techniques can do, it is important to first understand the phenomenon of urban tourism. Here, Ashworth and Page (2011) describe cities, as characterised by density and diversity, whether of functions, facilities, built forms, cultures or people that form the urban way of life. These urban areas are primarily signified by a busy, interactive built environment purposely built to meet the needs of many differing stakeholders (Edwards et al., 2008). A city often offers a wide variety of reasons why tourists visit cities. Some of these main purposes to visit a city, as described by Edwards et al. (2008) are: leisure, business, shopping, conferences, and visiting friends and family. This in combination with the wide variety of people, places, mixes of culture, values, expectations, and experiences result in a city with an exciting landscape for the urban tourist to explore (Edwards et al., 2008). However, while tourists make an intensive use of many of the urban facilities and services, little of the city has been created specifically for the use of tourists (Ashworth & Page, 2011). Additionally, as Ashworth and Page (2011) note, are cities that accommodate most tourists oftentimes large multifunctional entities, where tourists often are barely visible in both economical and physical aspects. Despite this, does tourism increasingly have an influencing role in shaping spatial and economic characteristics of the urban environment. To improve the long-term sustainability of touristic places it is important to make research of urban tourism destinations (Edwards et al., 2008). Here, Venice is a city that is the example of a city being heavily affected by mass-tourism. As the city is being toppled by the masses of tourists that flood the city, which leads to urban strains such as leaving little space for the residents and their desired quality of life (Edwards, 2017).

D.   Social Media's Value

With a wide variety of social media applications, including many differing ways of collecting metadata from social media, it is important to understand the benefits and negatives as well as the values that different social media networks can give. It is, therefore, this section that shall include the first sub-question of: *What are the advantages and disadvantages in creating spatial patterns and temporal distributions from social media networks in general?*

Nowadays, social media networks have become increasingly popular and the contents that are being uploaded on these social media platforms are increasing significantly (Xiang & Gretzel, 2010; Batty, 2013; Croitoru et al., 2013; Gössling, 2017). This wealth of digital data is available to be collected, analysed, understood, and used to help solve complex problems and support informed decision-making (Campagna, 2014; Massa & Campagna, 2014; Gössling, 2017). This signifies a shift that has occurred in recent years of digital data production. With Flickr users alone uploading more than 3000 images per minute, which are remarkable volumes of user generated data (Croitoru et al., 2013). Additionally, as Massa and Campagna (2014) discuss, are the traditional spatial analysis methodologies and techniques as surveys, interviews, and observations, no longer adequate to manage and take advantage of this knowledge potential. And as Batty (2013) notes, it is this kind of big data that can benefit planners and decisionmakers to help understand the world around us. Here, Flickr is an ideal

platform to use in the research of this thesis, as tourists are quickly associated with taking pictures and Flickr is predominantly a photo sharing platform from which the photos contain spatio-temporal data of when and where the photos were taken. Most importantly does Girardin et al. (2007), consider uploading, tagging, and disclosing the location of a photo an act of a physical presence in time, which makes tourist research based on their uploaded photos an interesting technique to gain insights.

There are many advantages as it comes to social media networks and the metadata it offers. More and more studies, especially within human behaviours, are increasingly reliant on the wealth of digital data produced by human social activity (Sobolevsky et al., 2015). Moreover, as Kádár and Gede (2013) note, is defining the spatial patterns of tourists' activities a valid agenda, as it can help to gain insights into clearer definitions of tourists motivations, urban space usage and socio-economic impacts. However, to achieve this, it is necessary to analyse and convert digital data into understandable information. Predominantly, as Blok et al. (1999) describes, is analysation the finding of patterns and analysing their characteristics, comparing patterns to discover anomalies and relationships, but also seeking trends in pattern developments. This ultimately allows bridging the gap between independent spatial and social network analyses and supports the discovery of additional knowledge (Croitoru et al., 2013; Vassakis et al., 2019).

Having location-based information thus, offers multiple advantages over traditional methods, as it allows the precise and continuous tracking throughout space and time, which leads to valuable information in analysing human activity patterns (Kwan & Lee, 2004; Edwards et al., 2010; Sun et al., 2013). This makes it possible to derive underlying information on the social structure of the user community of different social network platforms (Croitoru et al., 2013). Additionally, users of social media networks can easily access information and be the producers and broadcasters of personal geographically referenced contents on location-based social media networks (Massa & Campagna, 2014). Here, as Massa and Campagna (2014) describe, are users able to include geographic information into their own generated contents, driving geography into daily routines. As such, social media networks could be considered as affordable and potentially boundless sources for information about daily life, events, but also opinions, feelings and needs related to geographical locations and facts (Massa & Campagna, 2014). This leads to data that can be regarded as a set of trajectories of multiple users or as independent spatio-temporal events, which can then be used for analysis of attractive areas, places of interests, landmarks or user travel preferences (Kisilevich et al., 2010). Moreover, can these insights lead to recommendations and customizations of services in tourism, improvements in local administration, advertising, the rescheduling of monuments opening times or even reallocating of existing service infrastructures (Girardin et al., 2008; Kisilevich et al., 2010). Ultimately, does this allow domain specialists as planners and decisionmakers to explore and analyse the temporal correspondence in connections using the visualizations made from social media networks. These insights are easier to investigate in forms of graphic visualizations and maps than to compare the complex patterns through numbers (Blok et al., 1999; Clifford et al., 2010).

One of the major challenges is the translation of the enormous amounts of data into a comprehensible form (Vassakis et al., 2019). Here, a risk as described by Batty (2013), stems from the fact that big data often needs to be augmented with other independent data sets to create a good estimation. However, to create an estimation there are chances for errors as there is human handling and bias involved. This is partly, due to the field in which big data operates, being not yet developed well enough (Vassakis et al., 2019). Additionally, as it comes to the analysation of photos made by tourists is the representation of data skewed towards those people that upload their photos on social media networks with spatio-temporal data (Sun et al., 2013).

Another major concern that exists is the privacy and ethical issues related to collecting data without the individual's consent. Here, as Girardin et al. (2008) describe, lies a risk of identifying people and organizations when the time comes that spatially rich data precision improves. Here, Clifford et al. (2012) and Batty (2013) mention, that the ethics for internet research are still developing and are subjected to new laws and regulations regarding questions of privacy and confidentiality as time continues. And this discussion of ethical implications for research is conducted using the internet is still ongoing. As Tiidenberg (2020) describes, are some scholars in agreement that individuals are responsible for maintaining their own privacy as it comes to the internet, however, not all agree. This unclarity stems from the fact that the internet and its ethics are embedded in deep structures of software, politics and habits, which despite increasing regulations, creates an environment where the research ethics lie with the personal choices of the researchers (Tiidenberg, 2020). However, the data which is ultimately used relies upon being publicly shared by those who use the social media networks. Despite social media user data being subject to controversy, as it comes to the ethics of using an individual their data, does the possibility offer many options for analysations and insights. To summarize the many advantages and disadvantages an overview is included in the figure below.
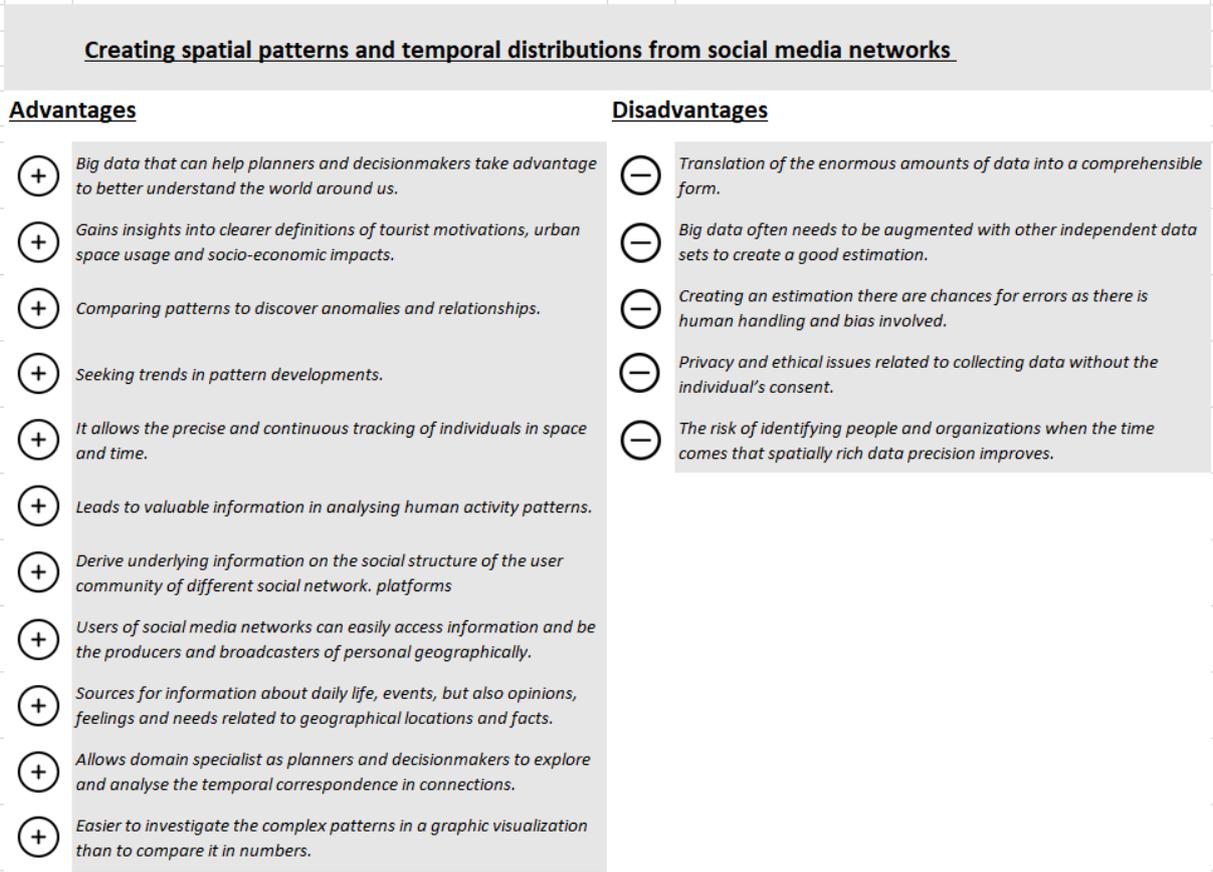
Creating spatial patterns and temporal distributions from social media networks

Advantages

(+) Big data that can help planners and decisionmakers take advantage to better understand the world around us.

(+) Gains insights into clearer definitions of tourist motivations, urban space usage and socio-economic impacts.

(+) Comparing patterns to discover anomalies and relationships.

(+) Seeking trends in pattern developments.

(+) It allows the precise and continuous tracking of individuals in space and time.

(+) Leads to valuable information in analysing human activity patterns.

(+) Derive underlying information on the social structure of the user community of different social network. platforms

(+) Users of social media networks can easily access information and be the producers and broadcasters of personal geographically.

(+) Sources for information about daily life, events, but also opinions, feelings and needs related to geographical locations and facts.

(+) Allows domain specialist as planners and decisionmakers to explore and analyse the temporal correspondence in connections.

(+) Easier to investigate the complex patterns in a graphic visualization than to compare it in numbers.

Disadvantages

(−) Translation of the enormous amounts of data into a comprehensible form.

(−) Big data often needs to be augmented with other independent data sets to create a good estimation.

(−) Creating an estimation there are chances for errors as there is human handling and bias involved.

(−) Privacy and ethical issues related to collecting data without the individual's consent.

(−) The risk of identifying people and organizations when the time comes that spatially rich data precision improves.

Figure 1: The advantages and disadvantages in using Social Media to create spatio-temporal patterns.

E. Techniques for Analysis

Understanding the advantages and disadvantages in creating spatial patterns and temporal distributions of social media networks, it is additionally, important to understand the multitude of approaches one can take to analyse harvested metadata. Within this section related works will be explored to gain insights in used techniques, which could help improve and form a basis to help with the analysis and visualization within this research. Therefore, the second research sub-question will be

discussed: *What techniques are recently studied in scientific literature to analyse and visualize spatial patterns and temporal distributions?*

The development of various general techniques for analysis of movement data and human activities have been proposed in the past years. Before social media networks became increasingly popular, pioneers within this research field, focussed on data available from survey-data. Here, Kwan and Lee (2004) presented different techniques for 3d geographical visualizations of space and time patterns of people's travel experience and mobility. By studying differences between genders and ethnicities, they revealed a complex interaction that showed a difference in space-time activity patterns. Other researchers started to focus on the increasingly available geographic locations of cell phones and other hand-held devices. Here, Ratti et al. (2006), started with exploring the urban environment by geographically mapping cell phone data at different times of the day, which resulted in a graphic representation of the intensity of urban activities through space and time.

However, one of the first to study the behaviour of tourists by identifying spatial and temporal features from Flickr photos were Girardin et al. (2007). By using visual analytics, they explored temporal distributions of tourists in Florence. These temporal granularities were studied by analysing days of the week, months of the year and days of the year, to better understand the touristic pulse within the urban environment. One of the techniques Girardin et al. (2007) used, was classifying photographers as tourists if all their photos were taken in a certain period within the study area. Additionally, were Edwards et al. (2010), interested in increasing the understanding of tourists' spatial behaviour in urban environments. Their research focussed on exploring tourist spatial movement, by only focussing on the trails taken, which were collected from GPS data. Here, one of the first possibilities of path creations throughout a city were explored while making use of digital data.

Other researchers then started to use alternative approaches that involved Flickr, which was used as a basis to uncover human behaviours from photographical data. Such as Kisilevich et al. (2010), who analysed attractive places, points of interest and compared behavioural patterns of different communities based on geographical imagery. By using heatmaps they interpreted attractive areas, which showed the density of photos in a certain area and reasoned that this was an alternative method of ranking points of interest in the world, based on the count of photographs in an area. In addition, were the identified point of interests compared with a tourist site where places are ranked based on user feedback [*tripadvisor*.com]. This ultimately, is a form of data triangulation to compare the outcome of results and their validity. However, also attention was brought to other means of interpretation of the metadata of Flickr photos. Hollenstein and Purves (2010) tested the location accuracy of user generated textual tags in London by identifying city cores. Here, the conclusion was that Flickr's accuracy value is a useful method for filtering out images that are placed at a wrong location. In other words, the accuracy of Flickr photos was found to be considerately accurate. Also focussing on the spatial accuracy, were Wood et al. (2013), which used Flickr to approximate visitation rates, instead of using traditional surveys. Additionally, did they verify that the entered location information given by Flickr users is generally correct. Within their study they showed how geographical data can be a useful resource for studying the nationalities of visitors.

With techniques and possibilities emerging from social media platforms, other interesting ideas came about. Sagl et al. (2012), showed examples of spatio-temporal patterns of collective human dynamics. However, interestingly, activity and mobility hotspots were created with seasonality taken into consideration. This, to provide additional insights into how collective social activities shape urban systems. Also interested in the human activities throughout space and time are Sun et al. (2013), who explored the patterns of tourists' accommodations in a city. Here, Kernel Density estimations were applied to produce estimates of densities, which resulted in a smooth surface of tourist

accommodations. This was a useful technique to help identify possible location clusters. Where tourists actually go, is however, explored by Kádár and Gede (2013), who visualized and analysed the spatial distributions of geographically tagged photos and made a differentiation between tourist attractions and local visitors, while exploring the differences. Interestingly, did Massa and Campagna (2014) conduct a textual analysis, which provided opinions and perceptions of users in different neighbourhoods. Their results showed how social media data can help support design, analysis and decision making in planning by stimulating dialogue with citizens and integrating this information with professional knowledge. Lastly, Sobolevsky et al. (2015) leveraged big data to study aspects of human activity for qualifying city attractiveness for foreign visitors. Here, seasonal patterns were revealed through temporal variations.

Knowing the many variations of social media data analysations and visualizations available will be useful within this research to help explore social media metadata and its possibilities, summarized in the table below. To explore the spatial patterns and temporal distributions of tourists in Venice, the Flickr photo sharing website will be used to gain insights into the whereabouts of tourist in the urban environment. Here, a combination of methods and steps of the above-mentioned authors will be used to explore outcomes. This thesis will follow their footsteps and make use of techniques and methods of visualization to draw conclusions and gain insights of the case-study in question, Venice.

Year	Techniques	Authors	Insights
2004	Activity Diary (GIS).	Kwan and Lee	Human activities and their movements.
2006	Geographic locations of cell phones.	Ratti et al.	Graphic representation of urban activities through space and time.
2007	Visual analytics (Flickr).	Girardin et al.	Temporal distributions of tourists.
2010	Exploring tourist spatial movements (GPS).	Edwards et al.	Understanding of tourists' spatial behaviour in urban environments.
2010	Heatmaps and use of external tourist ranking websites.	Kisilevich et al.	Identifying points of interests and use of data triangulation.
2012	Social sensor data to study the dynamics within the urban environment.	Sagl et al.	Spatio-temporal patterns of collective human dynamics.
2013	Visitation rates and visitations over time.	Wood et al.	Substituting surveys for Social Media data and studying the nationality of visitors.
2013	Creating Kernel Density estimations from Volunteered Geographic Information	Sun et al.	Recognizing patterns of tourists' accommodations.
2013	Visualized and analysed the spatial distributions	Kádár and Gede	Differentiation between tourist attractions and local visitors of points of interest.
2014	Conduction of textual analysis	Massa and Campagna	Opinions and perceptions of users from different neighbourhoods.
2015	Big Data used to quantify city attractiveness	Sobolevsky et al.	Seasonal patterns were revealed through temporal variations.

Table 1: Summarization of techniques and insights to explore social media metadata.

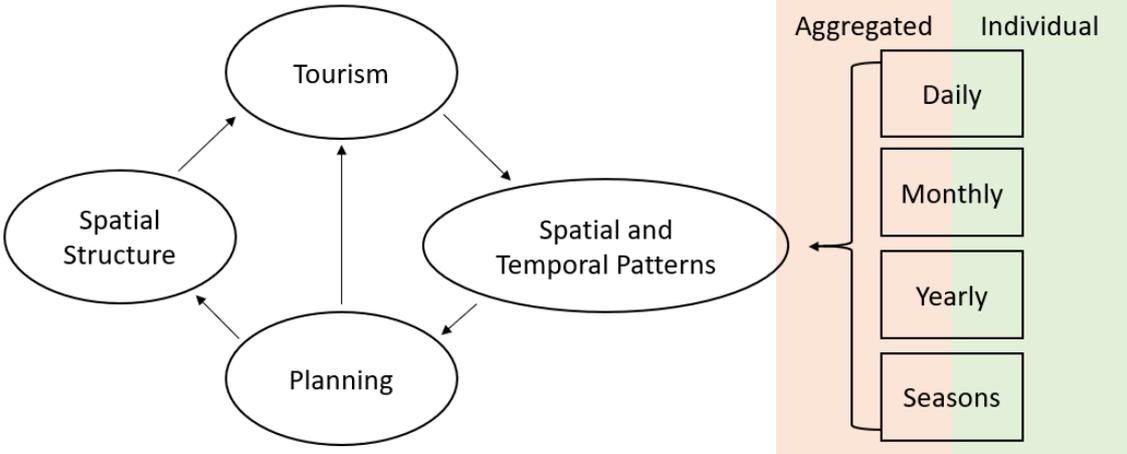## F.   The Conceptual Model

At last, does the conceptual model show the theoretical concepts in relation with one another. Affected by global trends as the increase in urban tourism, especially in major cities worldwide, but also, the increase in an online presences via social media applications that give new opportunities to analyse behaviours of tourists by creating spatial and temporal patterns in many new ways. These patterns are ultimately important to help solve urban strains caused by over-tourism as new insights can be created. As shown within the conceptual model can tourism be affected in two ways. Either by planning influencing the spatial structure, such as cities, roads, pathways, or the allocation of placement of buildings or attractions, to help alleviate urban strains. Or directly influencing the tourist, by limiting access, increasing pricing, improving information distribution or additional measures. However, without proper insights and little knowledge it is difficult to take appropriate actions and decisions. To gain a better insight into a cities spatial and temporal patterns can be used. This based on pictures taken by tourists and uploaded on social media applications. These patterns can be aggregated or be based on an individual level and can show daily, monthly, yearly, or even seasonal patterns as well as pathways, all constructed from the metadata of social media photos. These patterns and the information it provides could in turn help create insights to improve or change the urban environment and thus the behaviour of the tourists indirectly or by limiting tourist behaviour directly.

V.   Research Approach

Within this chapter the steps taken to gather and visualize metadata from Flickr will be explained. As a basis for this explorative research the multi-steps as described by Croitoru et al. (2013) will be used to collect photo information, such as when and where the images were made. Additionally, the steps surrounding this goal of processing and storing the retrieved data will be covered. The process of gathering data is important to eventually create spatial and temporal patterns, which in turn could help improve the urban environment or change tourist behaviour as shown within the conceptual model. These steps will be explained through answering the third research sub-question of: *which steps need to be undertaken to identify and analyse spatial patterns and temporal distributions of tourists using Flickr photography?*
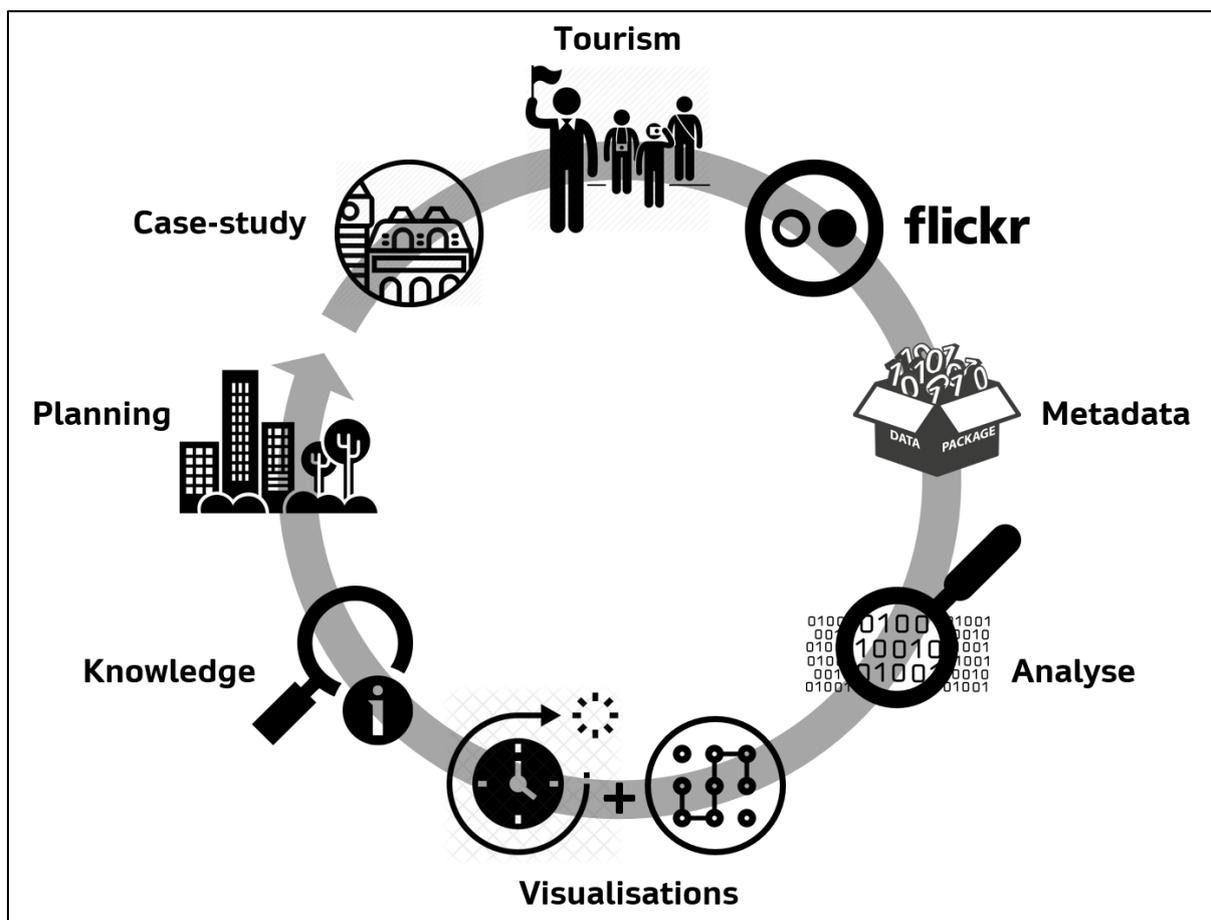
A.   Research Framework

Figure 3: Research Framework.

Depicted in *figure 3* is the research framework used for this thesis. It shows in general steps the cycle that has been undertaken to conduct research, which combined with the conceptual model can show a better understanding of the underlying relationships. Within the research framework it is shown how urban strains, such as tourism, can help create insights. These insights can than lead to an improved form of planning, specifically for the case-study in question, as the phenomenon of mass-tourism may occur in many more places, most of those places have their own ways of planning and each place is unique as it comes to the specific issues tourism causes upon their urban environment. Thus, increasing the understanding of motivations and activities of tourists in a specific city is the goal when conducting

this type of research, as it is this step of knowledge, that ultimately leads to planning decisions. This as shown, is done by collecting metadata of pictures that are uploaded on social media networks, specifically Flickr, as it is research friendly and supports researchers by providing Application Programming Interfaces (API), which are needed to gather data. By gathering the metadata of social media photo's, a rich amount of geographic information is unveiled, which is necessary to create visualization of both spatial patterns and temporal distributions. With the insight of tourist patterns that emerge from the visualizations, a better understanding of the urban environment can be created, to improve the understanding of planners and policymakers and their decisions, which influence the spatial structure of a city and can help create a more sustainable environment to alleviate the urban strains.

B. Methodologies

The combination of metadata collected from uploaded photos online in addition to the results being visualized in informational maps and graphs relate to research conducted by mixed methods. The reasoning behind this use of methods is the fact that quantitative data is used to create and visualize maps and images extracted from raw social media data. However, to interpret these results qualitatively methods are used to give meaning to these results and their usefulness. In addition, is desk research conducted by researching a self-created database of public social media data. This was done based on earlier conducted literature research about social media in general and the use of methods that have been used by other researchers to create, visualize, and analyse spatial patterns and temporal distributions from mostly social media data. To explore the available possibilities and understand what benefits the results can give, a case-study was conducted of predominantly a single-case: Venice, which based on earlier discussed theory forms a great example of how complex a city can be affected by mass-tourism. As in addition to Venice, additional insights are tried to be replicated in a lesser touristic city, namely Groningen. With a research approach that could be applied to multiple cities, it is important to understand the context of the urban environment is different all around the world. The focus, however, remains on understanding the touristic pulse of a city through analysing and visualizing social media photos their metadata. Ultimately, is the case-study beneficial to explore the possibilities of the techniques used and their validity.

C. Flickr Photography

The following section will explain the research strategy and the subsequent steps taken to recreate the conducted research and the tools used to ultimately increase environmentally friendly urban development. It all starts with Flickr, which provides a service that lets you retrieve data with your computer. This is done via API, which is a service where you send a command and in return you receive the requested data. Importantly, to operate these services you need to request access with a so-called key, which can be requested via the Flickr website. To get locations and dates from the Flickr API there is a routine named [flickr.photos.search] that makes it possible to search their database with a multitude of requests. Before the search for photos can take place a choice for data collection instruments has to be made. As there are many options for different databases as well as languages available to communicate with the internet and the database. Within this research the used programmer language is PHP and the chosen database to store the retrieved data is MySQL, which are used within the webserver environment XAMPP to communicate with one another. The choice for these instruments was predominantly chosen in terms of difficulty to work with and this combination seemed less complicated, in the end does this come down to personal preference.

Similarly, to the first step of Croitoru et al. (2013), the search for photos can begin by identifying the area of our-case study, Venice. As the goal is to better unders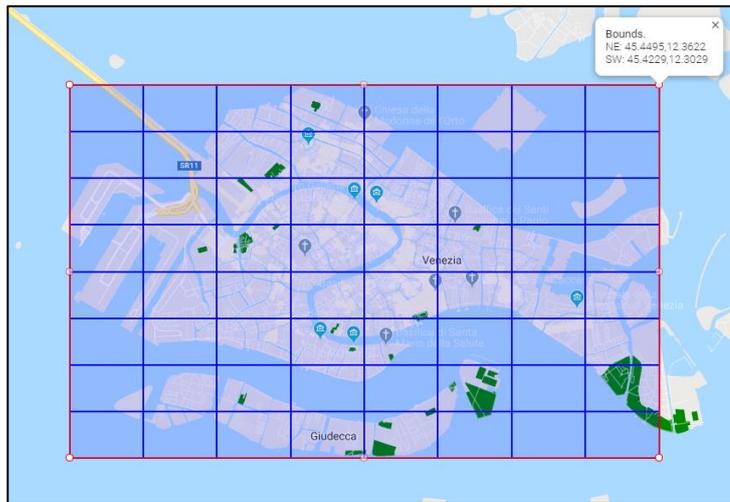tand the touristic pulse through new emerging methods, as research through social media data. It is important to focus the area on the touristic isle, as to avoid an abundance of data outside the case-study. This is done via a bounding box in combination with specified time frames, to specify the area as well as the time-period the photos have been taken in. Here, the bounding box consists of North-East and South-West longitude and latitude coordinates. To retrieve these coordinates a Google-Maps API routine was used, which shows the coordinates of a drawn box, covering the case-study area. Additionally, a start and end date need to be specified, within this case a period between 01/01/2014 and 01/01/2020 have been chosen. The reasoning for a multitude of years is to explore differences throughout the years, as well as differences between seasons and to avoid having too little or too much available data.

Figure 4: The Bounds of Venice in a raster format.

However, to process this area and not stress the Flickr servers, the area is divided into smaller boxes. This also makes it easier to keep track of already searched areas. After the area is scanned you are left with several identified photos that have been made within the pre-selected area. However, with these identified photos, the actual metadata of when and where a photo was taken has not yet been requested.

D.   Collecting Metadata

The second step becomes available after the identification of available photos within the case-study area has taken place and it is this step that the actual harvesting of photo data is done. This is done via the [flickr.photos.getinfo] API routine, which makes it possible to request metadata of earlier identified photos. This process can take up to several days, depending on the size of the bounding box as well as the specified time period. However, once completed the photo information is stored in a private database. Here, information such as: titles, descriptions, times, locations, and owners are stored for later use. Another element that is harvested is the Where On Earth Identification description, which can be used to later pin-point a location on a map with use of a Geographical Information System (GIS).

E.   Processing

With the data retrieved and stored in your personal database, it is now important as the third step to process the data by cleaning the errors and inconsistencies it may have. Here, it is often inaccurate locations that are the main problem. To solve this issue, Hollenstein and Purves (2010) and Wood et al. (2013) showed how the accuracy level of a photo location on Flickr can help filtering out images that are at a wrong location. This is done by selecting photos that are categorized with accuracy level 16 on Flickr, which is the highest accuracy available on street level. Other remaining inconsistencies

are often explored via GIS, where leftover errors can be edited out. These are often photos tagged within the case-study area but displayed in another location.

F.   Criteria

With the general data being cleaned of inconsistencies, the fourth step would be to distinguish the photos made by either tourists or locals, as for this research it is necessary to explore the whereabouts of tourists within the urban environment. However, there are not only tourists who make use of the social media platform. Here, additional information by Flickr is offered concerning the nationality of a photo owner, which makes it possible to filter users based on their nationality linked to their profile. To include domestic tourists, who reside outside of Venice, a selection was made for locals that had their place of residence set to Venice. This makes it possible to distinguish locals, tourists, and unknown photo owners. Within this research, patterns of tourists are mostly analysed, and a comparison is made to the behaviour the residents show, which leaves the unidentified photos unused. This leaves a set of locals and another of tourists within the case-study area.

G.   Conversion

The fifth step is to make the database now stored within the database useable within a GIS environment. This is possible by converting the data into shapefiles, which is done via exporting the database into Geographical JavaScript Object Notation feature files (GeoJSON). Here, the longitude and latitude information are converted into a point feature with additional data attached to each feature, such as times, dates, and owners for analysation within GIS. Via [mapshaper.org] it is then possible to convert these features into a dataset of shapefiles, which are then able to be used in GIS.

H.   Visualization

What is left is to create spatial patterns. ArcGIS offers many tools to calculate and visualize maps. However, for most calculations within a GIS environment it is important for the data to be projected with coordinates, in this case an Italian format, specifically: RDN2008. This gives the Kernel Density tool and others the ability to calculate distances in meters. Additionally, as shown in *table techniques* is it possible to visualize these patterns in a multitude of ways. One such way would be by taken the seasonality into consideration over a span of multiple years. This way changes can be analysed throughout different seasons as well as different years.

I.   Ethical Considerations

Additionally, are ethical considerations taken during the collection of data. As noted by Girardin et al. (2008), is the topic of privacy a sensitive subject. Within this research, user data of Flickr is extracted in the forms of metadata that stems from photos uploaded by the users themselves. The photos used during this process are publicly shared. Moreover, were usernames displayed as owner numbers and thus stored anonymously. The difference between private photos and public photos is unknown as it comes to the representation of the tourist group. In addition, is the audience that uses Flickr not representative for the entire tourist community, as there are older age-groups that use these social media platforms less than current generations.

VI.   Results

With an understanding of the research approach to gather social media metadata it is now possible to apply earlier mentioned techniques of data exploration and visualization from other authors. By creating spatial and temporal patterns of tourists in Venice insights are created to help improve the urban understanding as well as tourists' behaviour to improve urban planning. It is this section of explorative research that will test some of the earlier discussed techniques in theory, which results are discussed with help of the fourth secondary research question: *which insights can be gained by spatial and temporal data from Flickr for planners and decisionmakers in Venice for refined urban planning?*

A.   Case-study: Venice

Within this case-study, metadata of Flickr photos has been collected. Currently, there are *174.441* different photos recognized and stored in the database as records. These photos were made by *8194* different individuals sharing their photos publicly online. After processing the data, a selection was made based on nationality and/or place of residence, if provided. This is important as explained within the research approach to identify photo owners as either tourist or local. This led to a numerous amount of tourist photos as can be seen in table 2, below. Here, a notable trend is the decline in number of tourist photos taken in more recent years, which are most likely a result of a social media platform in decline.

Year	Tourist #photos	Local #photos
2014	17.022	843
2015	17.236	439
2016	12.538	599
2017	12.730	610
2018	10.586	425
2019	7.271	564
Total	77.383	3.480

Table 2: Amount of identified tourist and local photos.

B.   Points of Interest

The next figure shows an aggregate of all data points, which represent the location in space and time of the tourists in Venice. The image is visualized as described by Sun et al. (2013), using Kernel density estimations. By using the program ArcGIS, the results show a smooth surface, which makes it possible to identify popular photo locations as well as points of interests within the city of Venice. In figure *5* does the colours range from blue, with a low density of taken photographs, to red, with a high density of photographs taken. The result is a surface map that shows in detail the popular points of interests that are photographed and uploaded within the collected timeframe. To get familiar with the case-area are the red surface areas analysed and identified in table 3. This identification is important to later analyse the earlier identified techniques of different authors as shown in table 1. This involves the analysation of differences between tourist and local points of interests (Girardin et al. 2007; Kádár & Gede, 2013), or to study the differences over different time-periods (Girardin et al. 2007; Sagl et al. 2012), and also makes it possible to take seasonality into consideration (Sagl et al. 2012; Sobolevsky et al. 2015). Additionally, does some familiarity with the case-study area make it easier to spot changes throughout years or seasons. To validate these results a comparison is made with a popular tourist website [*tripadvisor.com*] (Kisilevich et al. 2010), which ranks locations or activities based on user feedback scores. This ultimately shows how the data in form of photos represent actual touristic places in terms of popularity. This understanding can also explain why some points of interests are more popular during summer or winter time, as some outdoor activities are not available all year round.
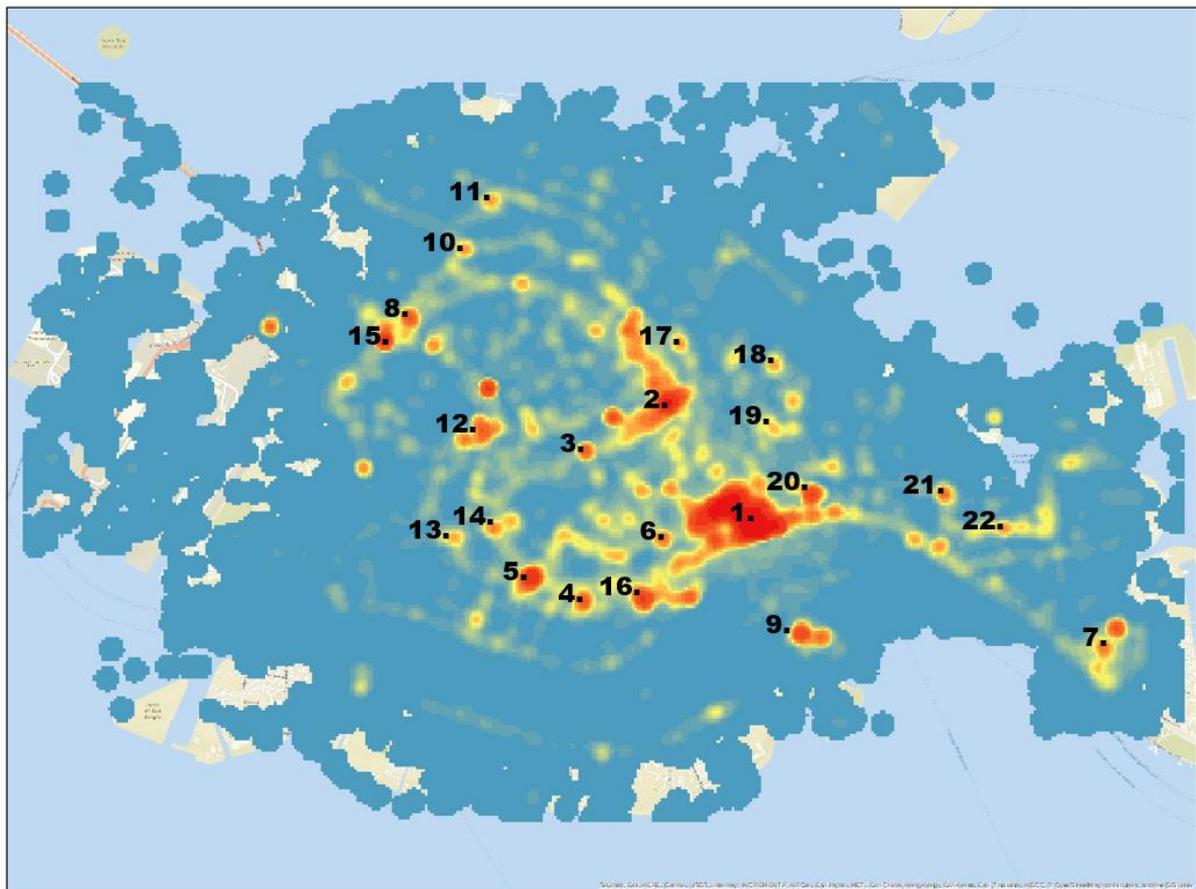
Figure 5: Kernel density estimations of tourist photos from 2014-2019.

	Point of Interest	Description
1.	Piazza San Marco, Doge Palace, Giardini Reali di Venezia and the Basilica di San Marco	The principal public square of Venice, with multiple popular destinations to visit.
2.	Ponte di Rialto.	The oldest bridge spanning the Canal Grande.
3.	Canal Grande.	The channel of Venice.
4.	Collezione Peggy Guggenheim.	A museum for modern art.
5.	Ponte dell'Accademia	One of four bridges that cross the Canal Grande.
6.	Ponte S. Moisè.	Small bridge, located near souvenir shops and a Gondola terminal.
7.	Giardini Della Biennale.	Relaxing gardens with art exhibitions and national pavilions, built by famous architects.
8.	Ponte degli Scalzi.	Another bridge that spans the Canal Grande.
9.	San Giorgio Maggiore.	A 16th century church, with a great viewpoint over the Piazza San Marco.
10.	Ponte delle Guglie.	Bridge spanning the Cannaregio Canal.
11.	Campo del Ghetto.	Neighbourhood with Jewish history.
12.	Basilica di Santa Maria Gloriosa dei Frari.	A 14th century church.
13.	Ca' Rezzonico.	A palace connected to the Canal Grande.
14.	Palazzo Grassi.	Another palace opposite the Ca' Rezzonico.
15.	Ferrovia.	One of many ferry-terminals.
16.	Basilica di Santa Maria della Salute.	Another church, next to the Canal Grande.
17.	Campo Santi Apostoli	A touristic square with small shops and street merchants.
18.	Basilica dei Santi Giovanni e Paolo	Another 14th century church.
19.	Parrocchia di Santa Maria Formosa	A 15th century church.
20.	Chiesa di San Zaccaria	Another 15th century church.
21.	Arsenale di Venezia	A complex of former shipyards and armories.
22.	Arsenale della Biennale di Venezia	Used to be one of the largest production centres in the pre-industrial era, concerning mostly ships.

Table 3: Identification of Point of Interests in Venice.

It comes as no surprise that Venice, with a great historic past and many sightseeing opportunities, thus has many famous locations highlighted on the above created map. The resulting image in figure 5, shows multiple areas with a higher density in photos taken. Of these locations there are 22 points of interest identified in table 3. The location densities were analysed using ArcGIS and Google Maps, to zoom into different locations and explore their surroundings. Here, some location densities are unknown and thus not numbered as it was unclear what was photographed on these locations without checking the actual photographs.

The resulting points of interest do show similarities as it comes to the rankings on tourist sites such as TripAdvisor, which stems from the fact that most pictures taken are oftentimes of tourist attractions like museums, historic buildings, and sightseeing locations in general. With one of the main locations highlighted in number one, is the main tourist area, which consist of multiple points of interest. This area includes the Piazza San Marco, the Doge Palace, the Royal Gardens as well as the Basilica di San Marco and many more intriguing buildings in close vicinity. Secondly, is the highlight around the area of the Rialto bridge, which crosses the Grand Canal and is according to the number of pictures taken a very popular destination. In addition, there are numerous photos taken within the vicinity, as there are a lot of touristic venues close-by as well that people consider noteworthy enough to capture on paper. Additionally, many museums, bridges, churches, and touristic neighbourhoods show up on the map created out of pictures. As mentioned, do most of these results match the top touristic attractions as shown on touristic sites, such as TripAdvisor. However, one place of note, would be the Teatro la Fenice. A theatre, which is highly ranked on TripAdvisor, yet not visible in figure 5. This might be because it is a location mainly visited for indoor activities and therefore it might not be permitted to take pictures inside. All in all, does the metadata of Flickr photos in combination with a Kernel density display give insight to the most attractive tourist places in Venice.

C.   Tourists versus Locals

For planners in Venice, could this data of a presence in space be of useful information. As this information shows the most attractive tourist locations, as well as places that are perhaps unfamiliar. With spatial planning being the systematic preparation of policy-making and executive actions (Woltjer et al., 2011). These insights could help increase conscious interventions through physical measures and regulations. And as Ashworth and Page (2011) mentioned, is it important to create this understanding of where tourists are in a city to help improve the spatial structure of a city. Here, the differences between tourist attractions and the interests of residents could help create this understanding. This is done by assessing the metadata of Flickr photography using the principles discussed by Girardin et al. (2007), where photos of tourists and residents were analysed separately. By looking at the nationality and/or place of residency of users, which have this information displayed publicly, the following figures were created of differences in points of interests using kernel density estimations.
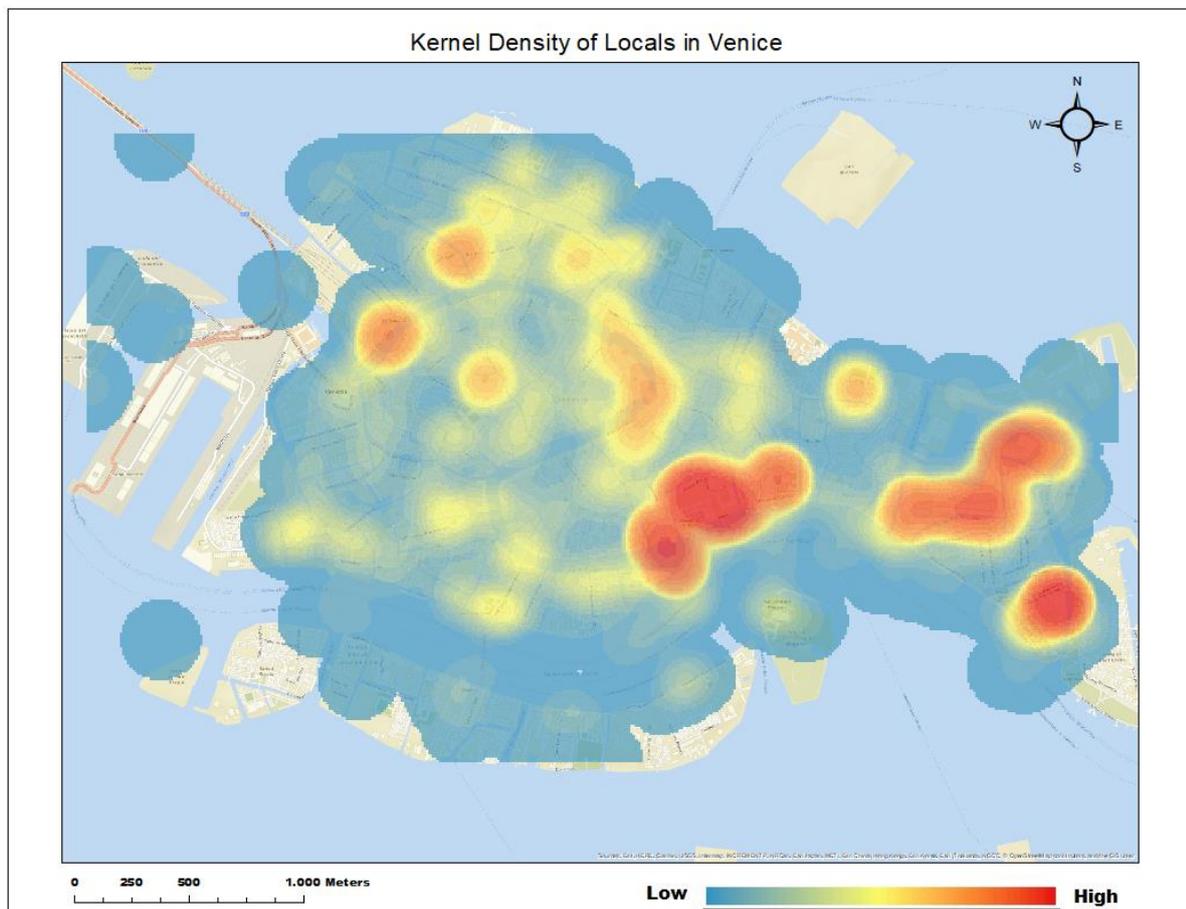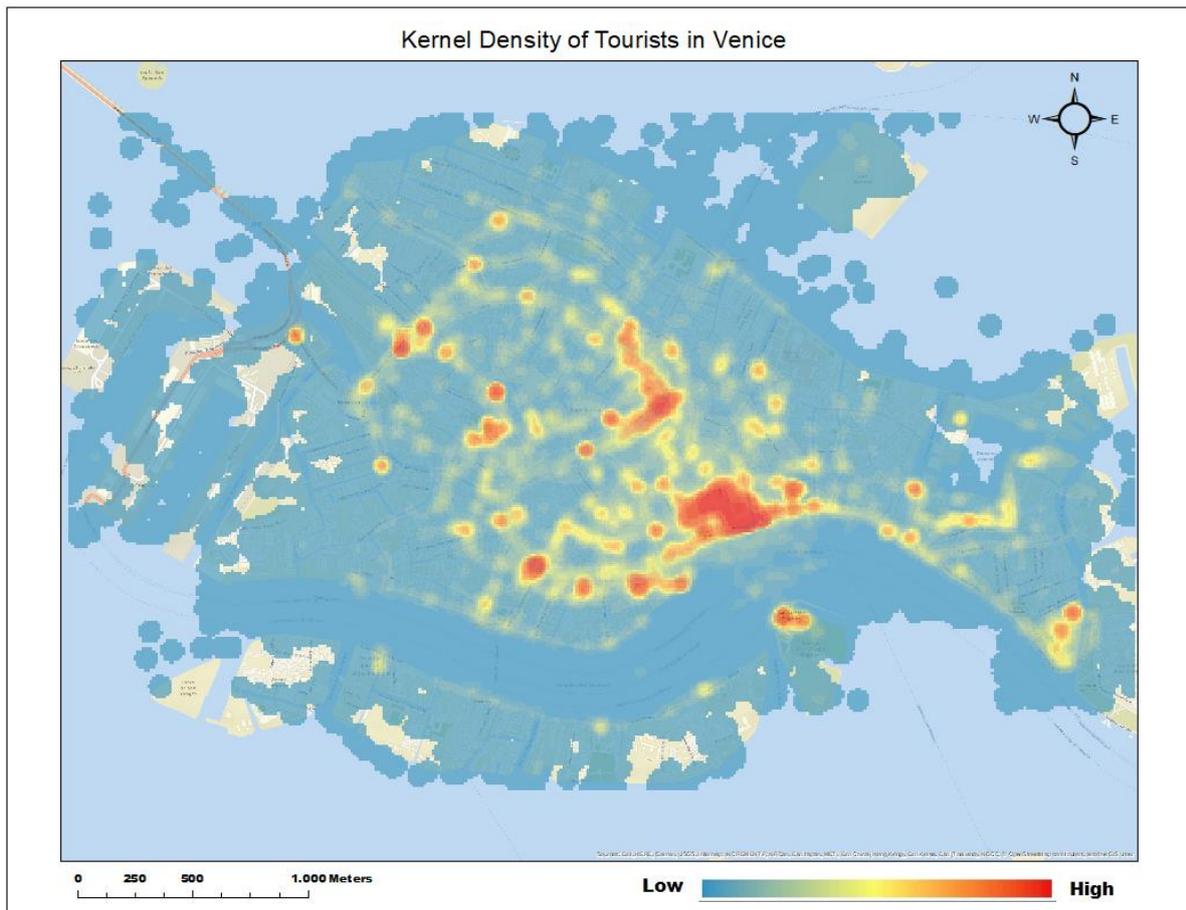
Kernel Density of Tourists in Venice

Low | High

0   250   500   1.000 Meters

Kernel Density of Locals in Venice

Low | High

0   250   500   1.000 Meters

Figure 6 & 7: Displaying the Kernel density of Tourists (upper picture) and Locals (lower picture).

Important to note are the differences in accuracy that can be created with big data and a more limited data set, as it comes to pictures taken by residents. The result is a map of locals that is less detailed, however, there are still some remarkable insights to gain. In both figures the main area highlighted is the Piazza San Marco (*#1 in table 4*). This in combination with the Rialto bridge, which has a lesser density of pictures taken for locals than it has for tourists. However, the area is one of the few locations that is identifiable from the local density map. Of note is the higher density areas of both the Arsenals of former shipyards as well as the gardens with art exhibitions. The reason for discrepancy might be the fact that these attractions are a little off-centre for the average tourist, who might have limited visiting times and mainly wants to visit the main points of interest. In addition, are the changes of expositions for the art gallery a reason for locals to make a return visit, which could be a reason for the distinction. Another highlighted area, would be the Cannaregio, including the Guglie bridge, which is similar compared to the tourists map, however, a bit less noticeable as there is a big discrepancy in available data between both groups.
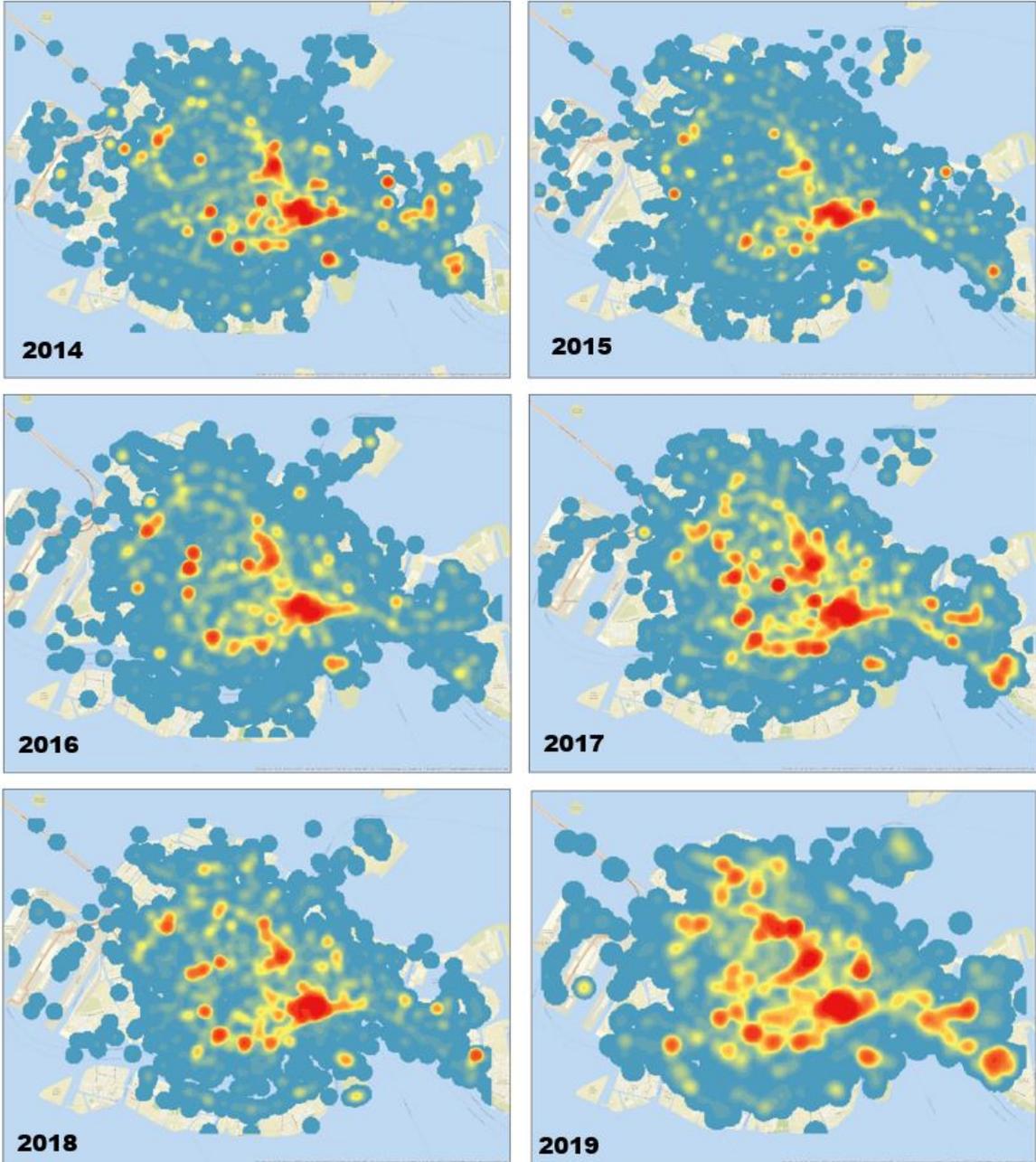
D.   Annual Display

Figure 8: Kernel density maps of each collected year from 2014-2019.

It is also possible to study differences over different time-periods, as can be seen in figure 5 above. Here, changes over time are noticeable with occurring and disappearing red spots over the years, which means a change in photographed objects or locations. Whereas the main visiting locations such as: The Piazza San Marco and the Rialto bridge are visible continually throughout the years. Other spots brighten up irregular. This might be caused as a result of different expositions, new openings, a change in popularity or because of the dependency on user uploaded pictures, that can cause changes to the yearly patterns. While the maps created throughout the years 2014 to 2019 mostly show similarities in points of interests, there are some notable changes. One of them is the increase in activity east of the Rialto bridge since 2016. While there is no clear-cut answer for this phenomenon, there are some possibilities how this came to be. Either due to new attractions, better informed tourists or due to the regulations attempted by the Venetian government to regulate tourists. Additionally, in year 2016 there is less exposure of the Bienalla gardens, which reappears in the following years. This could be due to an unpopular exhibition of art or just due to the bias people have while taking pictures.

The available information could help improve planning decisions, marketing activities, visitor flows and transport policies (Edwards et al. 2013). These regulations can take place in a multitude of ways. As the insights gained from social media can help target problematic areas for planners and decision makers as well as help monitor and evaluate changes over time after certain regulations have been put into place. Here, a uniform solution is difficult to describe, as the intricacies of each city differ and require solutions on a local level made by those knowledgeable of the urban environment.

E.   Seasonality

Looking throughout time by having multiple years at display there are many additional options for visualization. One of them as shown by Sagl et al. (2012) and Sobolevsky et al. (2015) is looking at seasonality. Here, a 3-month period is taken for each season. With winter from December to February, spring from March to May, summer from June to August and lastly, autumn from September to November. This results in the figure 9 below.
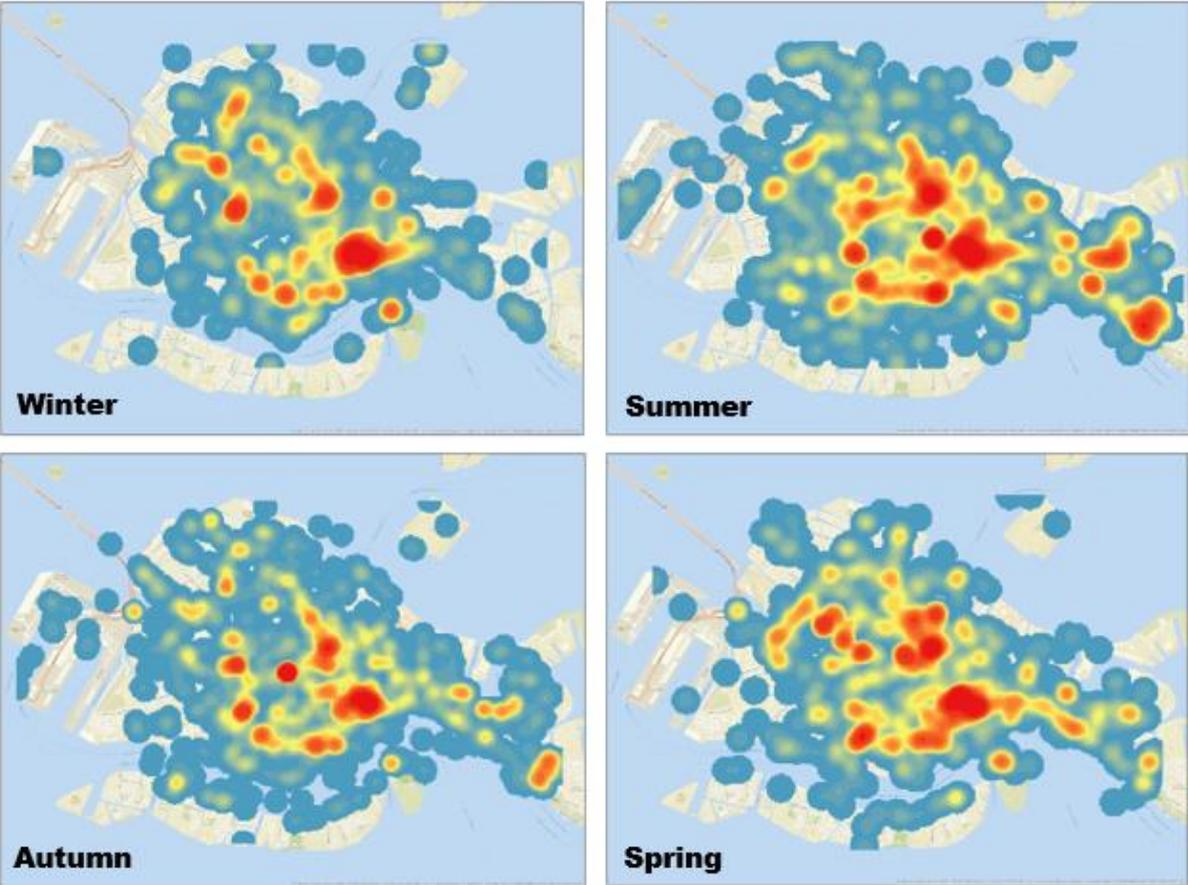

Figure 9: Seasonal display of tourists their spatial patterns.

Although relatively similar at first glance, there are some notable differences between seasons. However, it does show how the summer and spring periods contain more highlights. This should come as no surprise as these seasons have favourable weather conditions, which oftentimes defines the tourist season. Another difference would be the popularity of the arsenals, the former shipping yards to the east, during the summer period. This is the opposite of the other seasons, where barely any pictures are taken and uploaded online. The same goes for the art exhibition gardens, which attract many more tourists during the summer period. However, despite these seasonal activities, the main attractions such as: The Piazza San Marco and the Rialto bridge remain some of the most visited locations throughout the whole year. Here, the intensity of photos taken on those locations is lessened during the off seasons.

F.   Temporal Distributions

Having explored some of the options to visualize and analyse spatial patterns available. It is this section that will discuss the temporal distributions that have been extracted from the metadata of Flickr photos. To create an understanding of the whereabouts of tourists in Venice it is additionally possible to explore the data throughout different time-periods, this time however, with a focus on when the pictures were taken. These tourist pulses can be created with activity charts that show when a picture was taken (Girardin et al., 2007). The figure 10 below shows how the activity is displayed with a distinction between tourists and locals in relative numbers. Here, charts of days of the week, months of the year, days of the year as well as hours of the day are aggregated of the collected time-period.
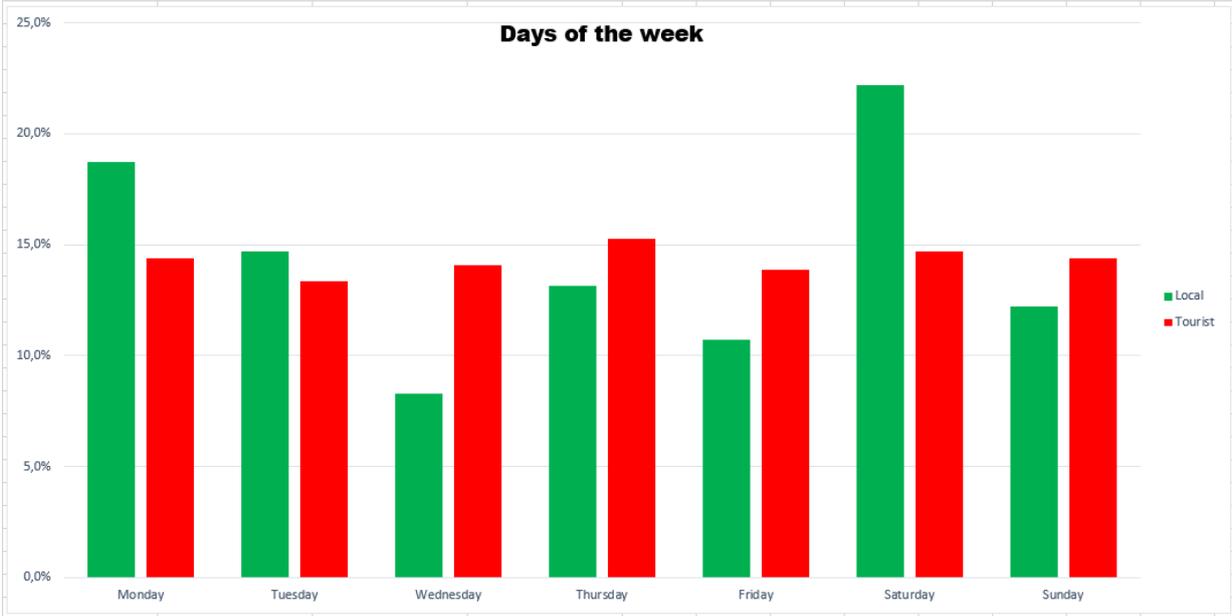

Figure 10: Temporal distributions of both tourists (red) and locals (green) for each day of the week.

As the overall number of photographs taken from tourist dwarf those of locals, percentages are used to identify changes within both tourist and local photographs. What is noticeable in this chart is that locals take relatively more pictures on a Saturday, which could be their day off, than during the week. For tourists, the number of uploaded pictures is rather equal throughout the week. This means the tourist presence in Venice is rather constant throughout the days of the week.

Additionally, the months of the year could be analysed, as can be seen in figure 11. Here, clear differences are noticeable for both locals and tourists. With locals taking most of their pictures during February and March, which coincides with the festivities that are present during those periods. What

is noticeable, is that during this period also more tourists are attracted towards the city, taking the most pictures during the month of February. Moreover, it is noticeable that during the summer period there is a relative peak, whereas the months of January and December, during winter are the quietest months as it comes to uploading photos on Flickr.
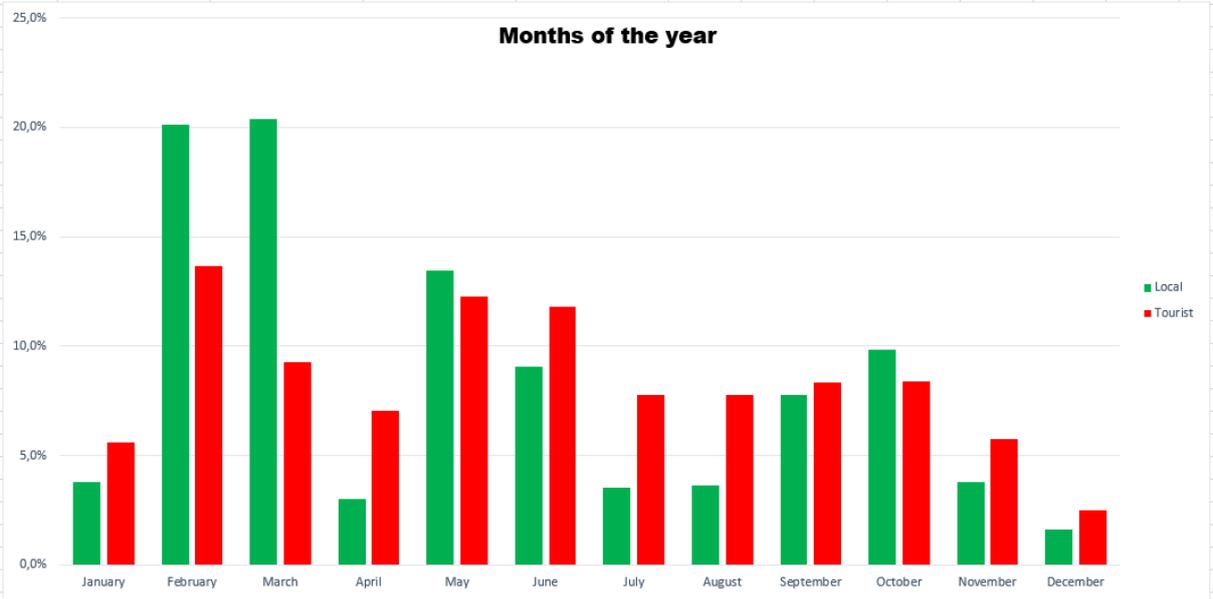

Figure 11: Temporal distributions of both tourists (red) and locals (green) for each month of the year.

With days of the year it is possible to correlate the spikes to festivities or national holidays, as they oftentimes occur on the same day each year. Even here, aggregate data is used, but to further explore individual years in more depth it is also possible to consider looking at the analyses from each year, to better pin-point certain events. However, with the following chart, clear spikes are noticeable. The Venetian carnival is clearly noticeable during the month of February. This period also draws in a lot of tourists, as it is one of the periods when tourists upload the most pictures on Flickr. Also, noticeable, is the Vogalonga, a local festivity where venetians row through the canals of the city in late May or early June. It could also be possible to reverse search these tables and look at events that happened on the calendar and analyse if it is noticeable in the given data.
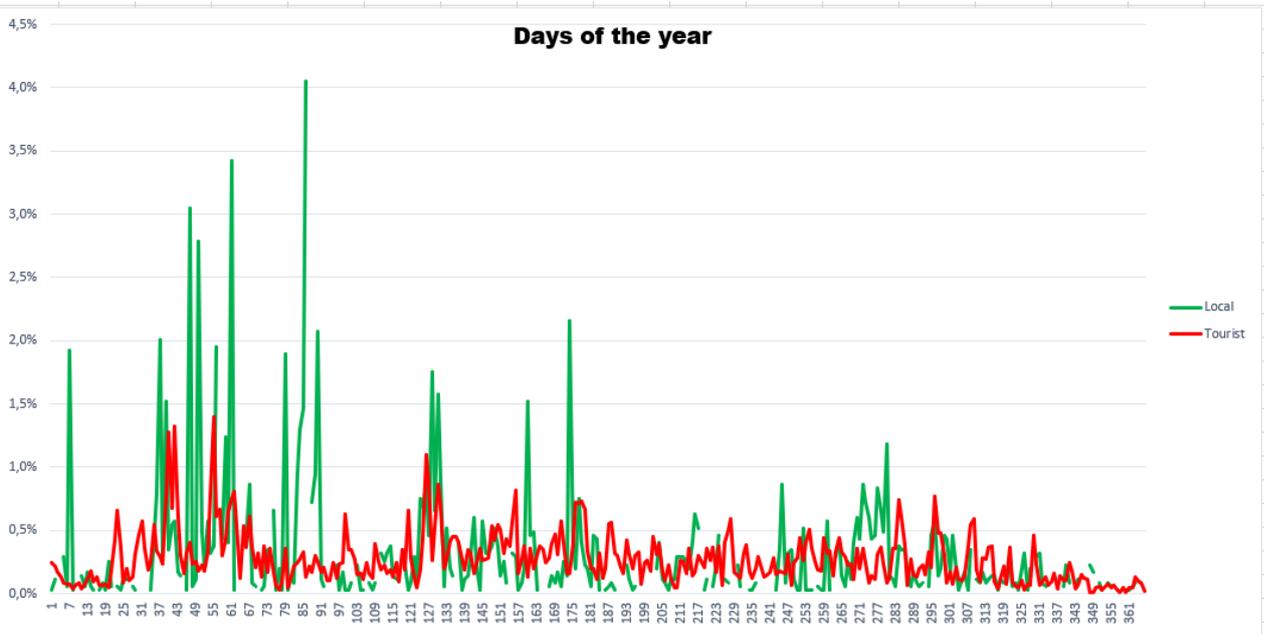

Figure 12: Temporal distributions of both tourists (red) and locals (green) for each day of the year.

The last addition is to look at activities during a given day. What is noticeable in figure 13 below, is that most photos are taken during the afternoon, which is understandable as with daylight the Venetian city and her landmarks show their beauty. Additionally, during night-hours it are the tourists who are continuing to take pictures in comparison to the locals. As well as in the early mornings when tourists start off their day earlier to probably enjoy some of the serene moments the city has to offer and before the city is overwhelmed by tourists.
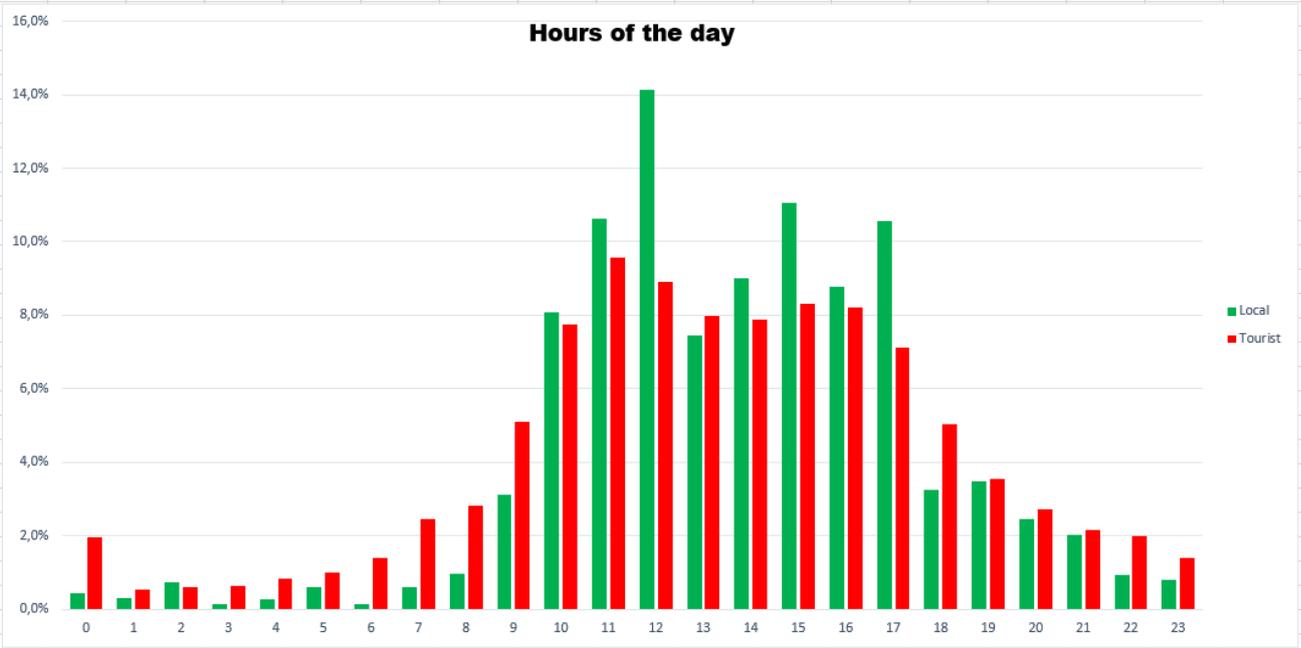

Figure 13: Temporal distributions of both tourists (red) and locals (green) for each hour of the day.

G.   Pathways

With a greater insight of the whereabouts of when and where tourists spend their time in a city. It is additionally, a possibility to create and analyse pathways of tourists in Venice. To construct pathways, it is important that the metadata includes when and where a photo is taken as well as the availability of multiple photos made during a day. Ideally this method could be implemented on all available data. However, to avoid an unreadable cluster, a decision was made to track individuals anonymously. This was done to filter out many unusable picture owners and to highlight the possibility of this method to

create insights how people move throughout the urban environment. This was done via ArcGIS by chronologically drawing lines. The results are visible in the following figure 14, where the photographer followed a path starting at the Rialto bridge (north). This was continued by moving along the alleyways next to the Grand Canal, where eventually the photographer most likely got on a ferry. The next photograph was made in vicinity of the Basilica di Santa Maria della Salute, from which the path continues across the Ponte dell' Accademia and ended near the Piazza San Marco.

Figure 14: Construction of an individual's pathway.

Additionally, it is possible to analyse individuals over multiple days. The following figure 15, shows three days of photographs made throughout the city of Venice. Here, the first day is colorized blue, the second day red and the third day as black. Now despite Venice being a city very much known for their canals and the use of ferries and gondolas, which makes certain routes very difficult to understand as the pathways across water are not very linear. There are clear pathways to analyse. On the first day, the Rialto bridge was visited and places in its vicinity. On the second day a tour throughout the city was taken, visiting some of the more southern regions by boat, and crossing throughout most of the eastern part of the city. On the third day a clear path is noticeable along the Grand Canal, as well as visitations to the Piazza of San Marco and more western parts of Venice. With this method the potential of analysing visitor flows becomes possible, especially when aggregated. Aggregating the pathways would make it possible for the most used pathways to become visible, despite possibly not being the quickest or most economic route. This in turn, could help benefit planners and policymakers to help regulate the city, as there is more insight into paths taken, which could help in making targeted solutions for the city.

Figure 15: An individual's pathways spread over multiple days.

All in all, does the data offer insights into human behaviour within the urban environment, with a special focus on tourists. To ultimately give clear examples as to what this information might be used for is rather difficult to say. However, the insights gained like the places people visit, the differences in behaviour of tourists and the residents and the paths they take can help towards improving planning principles of creating a better urban environment with concise planning interventions and actions. Here, measures can be focused on those areas that are highlighted as crowded, which shows through greater amounts of photo densities. By spreading information and perhaps a bit of nudging can tourists be steered away or limited to overcrowded streets or locations. Moreover, is it possible with the

insights from social media to consider specific time-periods of the day or even during specific events in a month, to prevent visitors being in the same place at the same times. Additionally, do the result lend themselves well for comparisons over time. By evaluating interventions made within the urban environment changes over time could be seen through new social media metadata. To show an example of this possibility an additional case study was conducted about the city of Groningen. Where the opening of the Groningen Forum is analysed via Flickr photography.

H.   Case-study: Groningen

While the inclusion of Groningen may seem out of place. Curiosity remained about using the same techniques on a smaller city, which is not much affected by tourist flows. For this reason, I chose my own hometown Groningen, a city in the north of the Netherlands with over 200.000 inhabitants. It is included to highlight the possibilities of monitoring change in spatial patterns after the opening of a new attraction. In this case, the Groningen Forum, which was opened on 29-11-2019. This cultural institution offers activities as movies, lectures, study space, exhibitions, and a great rooftop view. With over a million visitors in the first three months the Forum had a great start, despite Groningen being a relatively small city, especially as it comes to tourism. By applying the same method as described in the research approach, metadata of almost 2.000 pictures were extracted over an almost year long period as shown in table 4.

Opening Groningen Forum	#photos
Before	1177
After	758

Table 4: Number of photos before and after the opening of the Groningen Forum.

With a time period of 5 months before and after the opening of the Groningen Forum, a shift in photo density has taken place, as can be seen in the figures below. The deficit in uploaded photos in the period after the opening might be caused due to upload lag, where pictures have yet to be shared online. Most interestingly, does the Groningen Forum appear as a new location marked with an arrow, shifting away from the Grote Markt (Grand Market), located across the street. The more consistent density spot in the middle of the map, located around a restaurant as well as a catering business remains a dense location.

Additionally, has the Groninger Museum become less visible after the opening of the Groningen Forum, which could have been caused due to the newness of the Forum as well as old expositions. Due to the opening of the Forum being close around the Christmas period, it also shows a winter event. Namely, the WinterWelVaart, which shows along the west side of the inner city of Groningen along the canal. Despite having a relatively small amount of available data, spatial patterns were uncovered that show changes within the urban environment after the opening of a new attraction. These are insights that can be replicated also for other interventions or regulations to help improve the urban environment and monitor changes.
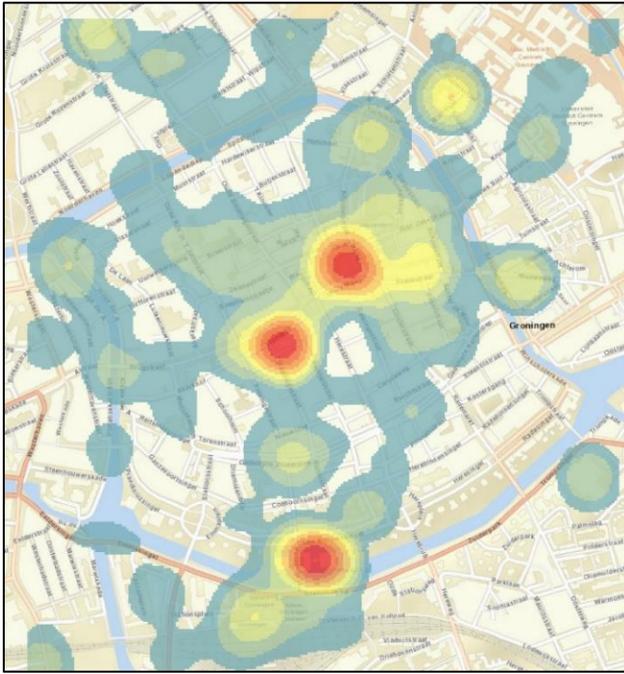
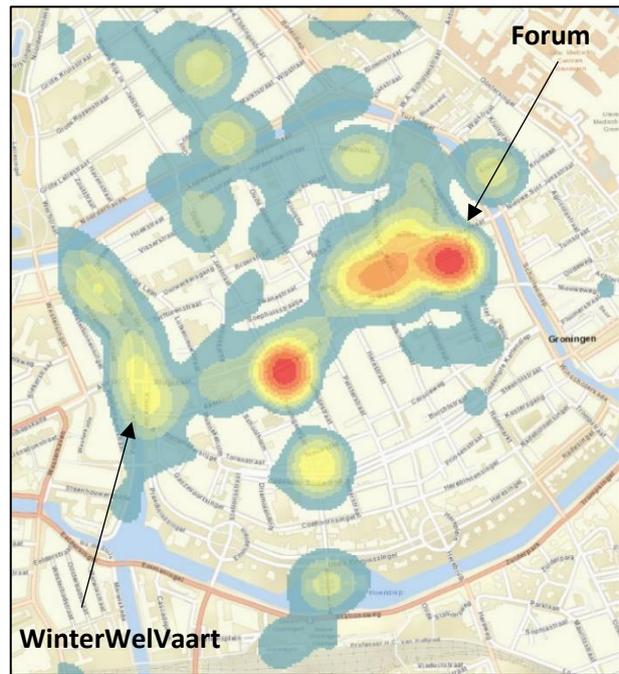*Figure 16: Before the opening of Groningen Forum.*

Figure 17: After the opening of Groningen Forum.

VII.   Conclusion & Discussion

At last, it is this section that concludes this study of analysing and visualizing spatial and temporal patterns of tourism in Venice using Flickr as a social media source. The aim of this study has been to provide greater insights into urban tourism. This has been attempted with the help of multiple sub-questions of which the answers provided greater insights into different topics. Here, the advantages and disadvantages of social media as a source of data have been discussed. As well as the methods for visualization that have been explored to make sense of the rich amount of spatial data that is available online. Additionally, is the used method explained how harvested metadata of Flickr photos is collected and transformed into understandable images and figures of spatial behaviours. Moreover, are insights of these spatial patterns and temporal distributions analysed and discussed to better understand the touristic pulse within the city of Venice. This even includes a small inclusion of the city of Groningen to highlight possible insights the used technique can give when changes occur to the spatial structure. What remains is the main research question that needs answering, which is the question of: *How can Social Media, such as Flickr, contribute in providing spatial and temporal patterns of tourism in Venice for improved urban planning?*

It was shown that Flickr could be a suitable source for harvesting spatial rich data, as it contains the whereabouts and timings of when and where a photograph was taken, which can be transformed into spatial patterns and temporal distributions using a program like ArcGIS. One of the main issues for planners and policymakers as discussed by Ashworth and Page (2011), is the lack of insights of the whereabouts of tourists within a city, which makes it difficult to accurately improve the urban environment. These insights are achieved by showing the touristic pulse throughout the city of Venice and partly Groningen. Here, insights such as kernel density estimations, points of interest, comparison in behaviour between tourists and residents, seasonal changes as well as differences in temporal activities. All in all, does it show behaviours and usage of the urban environment in space and time. Additionally, was a comparison made over multiple years, that shows changes over time, which could be especially useful to show changes before and after made interventions to indicate success rates or the need of adjustments while intervening within the urban environment.

In addition, making density maps according to the seasons creates insight into the behaviour of the tourists. Changes of point of interests are noticeable as the seasons change, which could help in creating spatial strategies. As ultimately, the purpose of planning is to promote common interests such as health, safety, and prosperity, which is done via efficient and environmentally friendly developments with physical spatial measures (Woltjer et al., 2011). The created results can help give insights to improve the systematic preparation of policy-making and executive actions. Additionally, temporal distributions are made in the form of graphs that show tourist activity of days of the week, months of the year, days of the year as well as hours of the day. This made it possible to identify events and festivities as well as understand the popularity of different locations throughout the seasons. Lastly, a change in points of interests with the opening of the Groningen Forum is evaluated. It was shown how a change in locations can also change the representation of Flickr photography and thus the change in spatial behaviour. All this was done with the help of 77.383 images made from publicly shared tourist photos.

Additionally, were tourist flows constructed that show the path of individuals throughout the city. Expanding on this method and aggregating the pathways could help to give insight into the tourist flows of a city. This understanding in general, could help alleviate urban strains. However, in the case of Venice, were pathways ultimately a little les appropriate, due to the many waterways available. This made it more difficult to track consistent routes and is also reason for highlighting two individual cases instead of an aggregated group of pathways. Another issue social media has in general, is the fact that

most photographs have a period of upload lag, which means pictures are often uploaded a few days, weeks sometimes even months later than the moment they were taken. This limits the possibility of this method to help improve tourist activity on a daily basis and thus regulate a city. However, this does not limit new technologies and other ways of social media to do provide contributions with more real-time solutions. In addition, sites like Flickr are sensitive to change and the information may no longer be available.

To answer the main research question, it is shown that spatial patterns and temporal distributions can be created from social media metadata and there are opportunities for improved urban planning. To summarize some of the improvements would be the additional insights that are created of the whereabouts of tourists within the urban environment. This is necessary for planners to intervene within the urban environment and this method showed some alternative perspectives that could help improve the knowledgeability of planners analysing a city. Additionally, are there opportunities to indicate success rates of taken interventions as well as possibilities to adjust these interventions by comparing before and after situations within the urban environment. By seeing the effects of taken measures there could be changes in the behaviour of tourists that show within available social media metadata. These effects could then be evaluated and considered either positive or negative upon which a decision could be taken to take further actions. The created results can thus offer insights to improve the systematic preparation of policy-making and executive actions, which are key components of spatial planning. Additionally, is it possible to identify events and festivities as well as understand the popularity of different locations throughout the seasons. And most importantly, is it possible to create tourists flows within a city, which could be of great importance to help alleviate urban strains of over-tourism and help consider measures to control the over crowdedness and better understand their location.

However, despite the insights created from spatial and temporal patterns. There are no clear outcomes as to how to improve the urban environment considering urban strains as over-tourism. What the spatial and temporal patterns do offer are insights for planners, which could help enlighten their vision and improve the decision-making process. It is, however, important to understand the outcomes created are metadata of online photographs. And this social media metadata, does not come without its own flaws. As the hotspots portrayed are not necessarily hotspots within the real-life setting. They portray objects or attractions of locations which are heavily photographed, and this could be a subtle difference with the actual situation in a city. Therefore, basing urban planning upon this method alone should not be a consideration. Though, there are opportunities to enhance current insights available as is shown through the construction of seasonal differences, time-intervals as well as the opportunity to construct pathways. Additionally, does it remain uncertain in what capacity the density of photographs display actual tourist hotspots. Despite, Girardin et al. (2007), seeing photographs as a presence in space and time and the densities often correlate with major tourist attractions. It does not display situations that are crowded but not necessarily appealing to photograph, which is something to be taken into consideration for planners. However, it remains a suitable technique to enhance current insights created from surveys, observations or interviews and create a more all-rounded point of view.

All in all, this study has found both interesting spatial patterns and temporal distributions of tourists in the city of Venice. With literature research that showed interesting tools and techniques from related works a great many ways of visualizations have been shown within this study. In the end, a selection was made with what was possible and limited to the available software. However, this study is not the end, as other scholars have shown and created many more interesting ideas and techniques that could

help create insights from social media metadata. This thesis has explored some of the possibilities of big data use and the planning benefits it can give, which shows the possibilities are there.

A.   Reflection

While this study showed spatial patterns and temporal distributions of tourist patterns within the urban environment. It remains difficult to answer what planning actions can be taken with the created insights. This is something I have struggled with throughout the duration of this study. Within this thesis I have thus tried creating an understanding how the results can lead to greater insights, which are valuable for planning. As planning's function is to help create environmentally friendly developments with physical interventions. Since this is done via conscious planning decisions, a better understanding of the urban environment is needed.

In hindsight I would have liked to have created more diverse results. The ArcGIS program has helped me a lot, but it also has its limitations, especially when it comes to my knowledge of the program. Recording 3D visualisations to better show density patterns would have been a good shot. This is something I felt was important, as most images that are shown are meant to tell most of the story by themselves. However, it was ultimately not something I managed to do. Overall, I am glad about the diversity of outcomes that are created. The results show multiple spatial perspectives combined with different time intervals, which does show a broad picture and might be a catalysator for others to improve upon with their own skills and creativity.

B.   Future Recommendations

Within this project a lot of options were explored as it comes to the visualization and analysation of data. It is, however, something which could be improved upon by showing three dimensional visualizations. Additionally, by undertaking interviews with experts in the fields, which could be planners that make spatial interventions, to better understand the benefits of the created result. This would be important to get a better understanding as to what this data and its interpretation may be used for as well as examples on how it could improve the urban environment, despite the insights it clearly gives. Moreover, is it a possibility to increase the number of photos used, by increasing the time-period that is subject to study. This could further improve the detail and quality of the results.

Acknowledgements

VIII. References

Ashworth, G., & Page, S. J. (2011). Urban tourism research: Recent progress and current paradoxes. Tourism management, 32(1), 1-15.

Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, *3*(3), 274-279.

Blok, C., Köbben, B., Cheng, T., & Kuterema, A. A. (1999). Visualization of relationships between spatial patterns in time by cartographic animation. Cartography and Geographic Information Science, 26(2), 139-151.

Campagna, M. (2014, June). The geographic turn in Social Media: opportunities for spatial planning and Geodesign. In International Conference on Computational Science and Its Applications (pp. 598-610). Springer, Cham.

Clifford, N., Cope, M., Gillespie, T., & French, S. (Eds.). (2010). Key methods in geography. Sage.

Croitoru, A., Crooks, A., Radzikowski, J., & Stefanidis, A. (2013). Geosocial gauge: a system prototype for knowledge discovery from social media. International Journal of Geographical Information Science, 27(12), 2483-2508.

Edwards, C. (2017). 'Tourism is killing Venice, but it's also the only key to survival'. [online] The Local. Available at: https://www.thelocal.it/20170718/mass-tourism-killing-crowded-venice-survival-authentic-travel-local [Accessed on 8 Jul. 2018]

Edwards, D., Dickson, T., Griffin, T., & Hayllar, B. (2010). Tracking the urban visitor: Methods for examining tourists' spatial behaviour and visual representations. Cultural tourism research methods, 104-114.

Edwards, D., Griffin, T., & Hayllar, B. (2008). Urban tourism research: developing an agenda. Annals of Tourism Research, 35(4), 1032-1052.

Encalada, L., Boavida-Portugal, I., Cardoso Ferreira, C., & Rocha, J. (2017). Identifying tourist places of interest based on digital imprints: Towards a sustainable smart city. Sustainability, 9(12), 2317.

Girardin, F., Dal Fiore, F., Blat, J., & Ratti, C. (2007). Understanding of tourist dynamics from explicitly disclosed location information. In Symposium on LBS and Telecartography (Vol. 58).

Girardin, F., Fiore, F. D., Ratti, C., & Blat, J. (2008). Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. Journal of Location Based Services, 2(1), 41-56.

Giuffrida, A. (2017). 'Imagine living with this crap': tempers in Venice boil over in tourist high season. [online] The Guardian. Available at: https://www.theguardian.com/world/2017/jul/23/venice-tempers-boil-over-tourist-high-season [Accessed 8 Jul. 2019]

Gössling, S. (2017). Tourism, information technologies and sustainability: an exploratory review. Journal of Sustainable Tourism, 25(7), 1024-1041.

Hasan, S., Zhan, X., & Ukkusuri, S. V. (2013, August). Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. In Proceedings of the 2nd ACM SIGKDD international workshop on urban computing (p. 6). ACM.

Hao, J., Zhu, J., & Zhong, R. (2015). The rise of big data on urban studies and planning practices in China: Review and open research issues. Journal of Urban Management, 4(2), 92-124.

Hollenstein, L., & Purves, R. (2010). Exploring place through user-generated content: Using Flickr tags to describe city cores. Journal of Spatial Information Science, 2010(1), 21-48.

Kádár, B., & Gede, M. (2013). Where do tourists go? Visualizing and analysing the spatial distribution of geotagged photography. Cartographica: The International Journal for Geographic Information and Geovisualization, 48(2), 78-88.

Kisilevich, S., Krstajic, M., Keim, D., Andrienko, N., & Andrienko, G. (2010, July). Event-based analysis of people's activities and behavior using Flickr and Panoramio geotagged photo collections. In 2010 14th International Conference Information Visualisation (pp. 289-296). IEEE.

Kwan, M. P., & Lee, J. (2004). Geovisualization of human activity patterns using 3D GIS: a time-geographic approach. Spatially integrated social science, 27, 721-744.

Massa, P., & Campagna, M. (2014). Social Media Geographic Information: recent findings and opportunities for smart spatial planning. Tema. Journal of Land Use, Mobility and Environment.

Persio, S. L. (2018). Traveling to Italy? These cities are fighting against mass tourism. [online] Newsweek. Available at: https://www.newsweek.com/traveling-italy-these-cities-are-turning-against-mass-tourism-909632 [Accessed on 8 Jul. 2019]

Ratti, C., Frenchman, D., Pulselli, R. M., & Williams, S. (2006). Mobile landscapes: using location data from cell phones for urban analysis. Environment and Planning B: Planning and Design, 33(5), 727-748.

Sagl, G., Resch, B., Hawelka, B., & Beinat, E. (2012, July). From social sensor data to collective human behaviour patterns: Analysing and visualising spatio-temporal dynamics in urban environments. In Proceedings of the GI-Forum (pp. 54-63). Berlin: Herbert Wichmann Verlag.

Sobolevsky, S., Bojic, I., Belyi, A., Sitko, I., Hawelka, B., Arias, J. M., & Ratti, C. (2015, June). Scaling of city attractiveness for foreign visitors through big data of human economical and social media activity. In 2015 IEEE International Congress on Big Data (pp. 600-607). IEEE.

Sun, Y., Fan, H., Helbich, M., & Zipf, A. (2013). Analyzing human activities through volunteered geographic information: Using Flickr to analyze spatial and temporal pattern of tourist accommodation. In Progress in location-based services (pp. 57-69). Springer, Berlin, Heidelberg.

Tiidenberg, K. (2020). Research ethics, vulnerability, and trust on the internet. Second International Handbook of Internet Research, 569-583.

Woltjer, J., van Dijk, T., & Voogd, H. (2011). Facetten van de Planologie. Kluwer Academic Publishers.

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, 2976.

Vassakis, K., Petrakis, E., Kopanakis, I., Makridis, J., & Mastorakis, G. (2019). Location-Based Social Network Data for Tourism Destinations. In Big Data and Innovation in Tourism, Travel, and Hospitality (pp. 105-114). Springer, Singapore.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. Tourism management, 31(2), 179-188.