



TECHNISCHE
UNIVERSITÄT
WIEN

MASTERARBEIT

**Extracting and Mapping Spatial Patterns of Interest
from Social Media**

Ausgeführt am Department für Geodäsie und Geoinformation der
Technischen Universität Wien

unter der Anleitung von
Univ.Prof.Mag.rer.nat. Dr.rer.nat. Georg Gartner, TU Wien
Und
MSc. Francisco Porras-Bernardez, TU Wien

Durch
Mohamed Ibrahim
Treustraße 92, Wien

Wien, 12 December 2019

Unterschrift (Student)



TECHNISCHE
UNIVERSITÄT
WIEN

MASTER'S THESIS

**Extracting and Mapping Spatial Patterns of Interest
from Social Media**

Conducted at the Department of Geodesy and Geoinformation
Technical University Vienna

Under the supervision of
Univ.Prof. Mag.rer.nat. Dr.rer.nat. Georg Gartner, TU Wien
and
MSc. Francisco Porrás-Bernardez, TU Wien

Durch
Mohamed Ibrahim
Treustraße 92, Vienna

Vienna, 12 December 2019

Signature (Student)

Statement of Authorship

Herewith I declare that I am the sole author of the submitted master's thesis entitled:

“Extracting and Mapping Spatial Patterns of Interest from
Social Media”

I have fully referenced the ideas and work of others, whether published or unpublished. Literal or analogous citations are clearly marked as such.

Vienna, 12/12/2019

Ibrahim, Mohamed

Acknowledgement

Praise be to God, by whose grace, good deeds and works are accomplished. Thank God for giving me the strength to accomplish this research and for the countless graces especially the good people in my life who are my family and friends.

The deepest thanks goes to my family (My father, mother and sister Nesma). You are the ones who really care and you are the ones who really matter. Without you, I wouldn't have accomplished anything. My deepest gratitude for you, wishing you all the happiness, health and peace of mind. I wish you were all here or I were there! Such time I spent with you, was priceless. I wish it could be back again!

To my all my Egyptian friends especially, Mohamed Mostafa, Mohamed Khamis, Mohamed Gomaa, Islam Barakat, Islam Hefzy, Amena Magdy, Nour, Khaled and my newest friend Ahmed Makky. You really cannot be replaced! I'm so grateful to have such good and supportive friends like you. I miss you all. Best wishes to you guys!

I would like to thank my supervisors Prof. Georg Gartner and MSc. Francisco Porras-Bernardez for their patience and concern in addition to the supervision work of the thesis and Dr.-Ing. Christian Murphy for reviewing the thesis.

And a very very very special thanks to Toot, thanks a lot for the company, support and the good time during last year. For a year full of ups and downs you were by my side. I'm truly grateful. Thanks for sharing such a great time of cooking, travelling and doing random things together. I wish you health, strength, happiness, peace of mind and satisfaction!

Abstract

Spatial patterns of interest are those spatial extents of interest to people. These patterns can be identified in two forms. The first form is areas of interest (AOIs) which refers to those areas that attract peoples' interest within the city borders. The spatial extent of these areas is not defined because AOIs are subjective to the behavior of different groups of people. The second form of visualizing spatial patterns of interest is footprints which refers to spatial spread of human activity.

Studying and understanding spatial patterns of interest, would undoubtedly help us to understand the spatial behavior of different groups of people. Thus, the extracted information about their patterns can be very useful in a lot of applications such as urban planning, location based services, national security, tourism, transportation and much more. Geosocial media GSM as stores different types of data such as tags, comments and photos that correlates to a space. Thus, GSM as a data source has proven reliability to provide data for spatial pattern analysis.

This thesis introduces a work scheme to extract and analyze spatial patterns of interest of different groups of tourists in Vienna. The proposed work paradigm can be used to identify spatial patterns of interest in similar cases. Vienna has been chosen as the study area in this research. Vienna has special attributes that make the city a standard area for touristic pattern analysis such as history, culture, location, event hosting, political situation and climate.

A dataset of Flickr geotagged photo data has been used to analyze the interest patterns in Vienna. Flickr is one of the oldest photo sharing platforms, thus Flickr data is very useful for performing a spatio-temporal analysis to compare the density evolution of tourists' activities during a wide time period than what other GSM data would provide.

A density based clustering algorithm HDBSCAN has been used to identify AOIs from Flickr data with help of concave hull enveloping algorithm. Different tourist groups' footprints have been identified using Kernel density estimation KDE. Both the followed methodology have proven reliability to identify interest patterns of tourists in Vienna. Using two different methodologies in this case study, has opened a door to criticize both methodologies during the research.

Abbreviations

| | |
|------------|------------------------------------|
| GPS | Global Positioning System |
| VGI | Volunteered geographic information |
| UGC | User Generated Content |
| GSM | Geosocial media |
| AOI | Area of Interest |
| POI | Point of Interest |
| KDE | Kernel Density Estimation |
| UTM | Universal Transverse Mercator |
| MRD | Mutual Reachability Distance |
| MST | Minimum Spanning Tree |
| TIN | Triangulated Irregular Network |

List of figures

| | |
|--|----|
| Figure 1: A diagram shows the research objectives, research questions the relational sequence of the research..... | 3 |
| Figure 2: A map shows potential heritage AOIs within and near the city center of Vienna. | 10 |
| Figure 3: A map shows potential diplomatic AOIs in Vienna..... | 13 |
| Figure 4 A map shows potential AOIs where events take place in Vienna..... | 15 |
| Figure 5Areas of interest in Vienna..... | 17 |
| Figure 6: A work flow to prepare the data for analysis..... | 20 |
| Figure 7: A pie of pie chart showing a brief summary of the raw data..... | 21 |
| Figure 8: Stack bar chart shows (a) the number of photos taken by different nationalities in Vienna; (b) the photos number drop ratio of every nationality than the preceding nationality in list..... | 22 |
| Figure 9: A pie of pie chart showing a brief summary the data after aggregation..... | 23 |
| Figure 10: Stack bar chart shows (a) the number of visits by different nationalities in Vienna; (b) the visits number drop ratio of every nationality than the preceding nationality in list..... | 24 |
| Figure 11: Flickr number of public photo uploads from January 2004 to December 2017..... | 25 |
| Figure 12: A line chart shows the total number of visits in Vienna on yearly basis. The demonstrated data is extracted from the Flickr dataset, which covers a 17 year time-span from 2002 till 2018. | 26 |
| Figure 13: A stacked bar chart shows the visits rate of different nationalities to all POIs within Vienna on yearly basis. | 27 |
| Figure 14: A stacked bar chart shows the visits rate of different nationalities to all POIs within Vienna on seasonal basis..... | 28 |
| Figure 15: A stacked bar chart shows the visits rate of different nationalities to all POIs within Vienna on monthly basis..... | 30 |
| Figure 16: A framework to visualize AOIs using a GSM dataset..... | 31 |
| Figure 17: A dot map of tourists' visits within Vienna..... | 32 |
| Figure 18: comparing k-means, DBSCAN and HDBSCAN methods using Scikit-learn toy data-set... | 34 |
| Figure 19: HDBSCAN example. | 35 |
| Figure 20: Sample data of 100 points on a Cartesian plane..... | 37 |
| Figure 21: MRD calculation example..... | 38 |
| Figure 22: The minimum spanning tree (MST) of the sample data..... | 39 |
| Figure 23: A dendrogram of the sample data that sorts all the edges of MST based on length. | 39 |

| | |
|--|----|
| Figure 24: condensed cluster tree of the sample data. | 40 |
| Figure 25: Visualization of the sample data after clustering..... | 41 |
| Figure 26: demonstrates the effect on the generated hull with different l values: | 42 |
| Figure 27: demonstrates the effect of different h values on heat map smoothness: | 45 |
| Figure 28: AOIs of all tourists in Vienna with minPts equals to .05% of this group number of points. | 50 |
| Figure 29: AOIs of all tourists in Vienna with minPts equals to .25% of this group number of points. | 50 |
| Figure 30: AOIs of all tourists in Vienna with minPts equals to .5% of this group number of points. .. | 51 |
| Figure 31: AOIs of Russian tourists in Vienna with minPts equals to 1% of this group number of points. | 51 |
| Figure 32: AOIs of Russian tourists in Vienna with minPts equals to 0.1% of this group number of points | 52 |
| Figure 33: A line chart shows the average fitness values of different λP | 54 |
| Figure 34: AOIs of all tourists in Vienna | 54 |
| Figure 35: Footprints of all tourists in Vienna | 56 |
| Figure 36: Footprints of Italian tourists in Vienna..... | 57 |
| Figure 37: Example for subtraction operation in Map Algebra | 58 |
| Figure 38: The footprints difference between The American tourists and all foreign tourists in Vienna | 58 |
| Figure 39: AOIs of foreign tourists in Vienna..... | 60 |
| Figure 40: AOIs of locals in Vienna..... | 60 |
| Figure 41: A bar chart showing the 10 most visited AOIs in Vienna and visits rates of different tourists groups | 61 |
| Figure 42: The footprints difference between the locals and all foreign tourists in Vienna. | 61 |
| Figure 43: AOIs of German tourists in Vienna | 63 |
| Figure 44: The footprints difference between the German tourists and all foreign tourists in Vienna... .. | 63 |
| Figure 45: AOIs of British tourists in Vienna | 64 |
| Figure 46: The difference in footprints between the British tourists and all foreign tourists in Vienna. | 64 |
| Figure 47: AOIs of Spanish Tourists in Vienna..... | 65 |
| Figure 48: The difference in footprints between the British tourists and all foreign tourists in Vienna. | 65 |
| Figure 49: A bar chart shows 10 of the most visited AOIs in Vienna and visits rates on annual basis... .. | 66 |
| Figure 50: The difference in tourist footprints between 2008 and 2013 in Vienna | 67 |
| Figure 51: The difference in tourist footprints between 2013 and 2008 in Vienna | 67 |
| Figure 52: A bar chart showing 10 of the most visited AOIs in Vienna and visits rates on seasonal basis. | 68 |

| | |
|---|----|
| Figure 53: AOIs of tourists in Vienna during summer..... | 69 |
| Figure 54: AOIs of tourists in Vienna during winter..... | 69 |
| Figure 55: The difference in tourists footprints between winter and summer in Vienna..... | 70 |

List of tables

| | |
|---|----|
| Table 1: Thesis structure | 6 |
| Table 2: List of some summer events in Vienna | 14 |
| Table 3: a list of the AOIs where the Christmas markets are being held in Vienna yearly | 14 |
| Table 4: Areas of interest in Vienna. | 18 |
| Table 5: A brief summary of the data in Vienna..... | 21 |
| Table 6: A brief summary of the data after aggregation in comparison with the raw data in Vienna..... | 23 |
| Table 7: minPts of different normalized values for different data groups | 48 |
| Table 8: Number of clusters of different data groups using different normalized values of minPts. | 49 |
| Table 9: Proposed minPts threshold levels. | 53 |
| Table 10: h_{opt} for different classes. | 55 |
| Table 11: Comparison between KDE and HDBSCAN as different methods to visualize spatial patterns of interests | 71 |

Table of Contents

| | |
|--|-----------|
| 1 Introduction | 1 |
| 1.1 Motivation and problem statement..... | 1 |
| 1.2 Research identification..... | 2 |
| 1.2.1 Research objectives..... | 2 |
| 1.2.2 Research questions..... | 2 |
| 1.2.3 Innovation aimed at..... | 4 |
| 1.3 Related work..... | 4 |
| 1.3.1 AOIs (areas of interest)..... | 4 |
| 1.3.2 Footprints..... | 6 |
| 1.3.3 Thesis structure..... | 6 |
| | |
| 2 Study Area | 8 |
| 2.1 History and culture..... | 8 |
| 2.2 Location and accessibility..... | 11 |
| 2.3 Political situation..... | 12 |
| 2.4 Events..... | 13 |
| 2.5 Climate..... | 16 |
| | |
| 3 Data | 19 |
| 3.1 Raw data..... | 20 |
| 3.2 Aggregated data..... | 23 |
| 3.3 Data analysis..... | 25 |
| 3.3.1 Temporal analysis on annual basis..... | 25 |
| 3.3.2 Temporal analysis on seasonal basis..... | 28 |
| 3.3.3 Temporal analysis on monthly basis..... | 29 |

| | |
|--|-----------|
| 4 Methodology | 31 |
| 4.1 Visualizing areas of interest | 31 |
| 4.1.1 Visualizing the data..... | 31 |
| 4.1.2 Generating clusters | 33 |
| 4.1.2.1 Density-based spatial clustering of applications with noise (DBSCAN) | 35 |
| 4.1.2.2 Hierarchical Density-based spatial clustering of applications with noise (HDBSCAN) | |
| | 36 |
| 4.1.2.2.1 Calculating the mutual reachability distances (MRDs) between the sample points. | |
| | 37 |
| 4.1.2.2.2 Structure the minimum spanning tree (MST) of the data | 38 |
| 4.1.2.2.3 Structure the cluster hierarchy | 39 |
| 4.1.2.2.4 Build the condensed cluster tree | 40 |
| 4.1.2.2.5 Extracting the clusters | 40 |
| 4.1.3 Bounding the clusters with polygons..... | 41 |
| 4.1.3.1 Convex hull..... | 41 |
| 4.1.3.2 Concave hull | 42 |
| 4.2 Visualizing footprints | 44 |
| | |
| 5 Implementation | 47 |
| 5.1 Visualizing AOIs of different groups of tourists in Vienna | 47 |
| 5.1.1 Identifying a suitable minimum number of points <i>minPts</i> | 47 |
| 5.1.2 Finding a suitable normalized length parameter λP | 53 |
| 5.2 Visualizing footprints of different tourists' nationalities in Vienna | 55 |
| 5.2.1 Defining a suitable search radius (bandwidth) <i>h</i> | 55 |
| 5.3 Technologies..... | 57 |
| 5.3.1 Map Algebra | 57 |

| | |
|---|-----------|
| 6 Outputs..... | 59 |
| 6.1 Spatial differences of interest patterns | 59 |
| 6.1.1 Locals and foreigners..... | 59 |
| 6.1.2 Different tourists' nationalities. | 62 |
| 6.2 Spatio-temporal differences of interest patterns..... | 66 |
| 6.2.1 Annual basis..... | 66 |
| 6.2.2 Seasonal basis | 68 |
| | |
| 7 Conclusion..... | 71 |
| 7.1 Possible future work..... | 72 |
| | |
| 8 References | 73 |

1 Introduction

1.1 Motivation and problem statement

In the last 20 years, the involvement of Internet users in the production of geographic information has increased. This is due to the emergence of Web 2.0 technologies, advancement of GPS enabled devices (e.g., smart phones) and advancement of broad band communication links (Ariffin, Solemon, Anwar, Din, & Azmi, 2014). As a result Volunteered Geographic Information (VGI) has gained growing interest (Mooney, 2015) among researchers and corporations. VGI is crowdsourced geographic information provided by variant participants, of different levels of education and knowledge (Haklay, Antoniou, Basiouka, Soden, & Mooney, 2014). VGI is a special case of user-generated content (UGC) that can be provided in different forms (e.g. restaurant reviews, travel logs, or geo-tagged photos). VGI data have explicit or implicit embedded spatial location components which can be analyzed and visualized in a Geographic Information System (GIS) (Parker, 2014).

Social media platforms such as Flickr, Instagram, Twitter etc. contain a huge amount of UGC data posted by users (Cao et al., 2010). Due to the ease of VGI data retrieval from social media platforms using their embedded APIs at low costs and by virtue of the rich characteristic of this kind of data to reflect human behavior (Hu et al., 2015), VGI data became an essential data source for both commercial analysis and research purposes (Oleksiak, 2014). Geospatial media (GSM) such as Flickr (Kennedy, Naaman, Ahern, Nair, & Rattenbury, 2007), Instagram and Foursquare (Silva, Vaz de Melo, Almeida, Salles, & Loureiro, 2013) have proven reliability to reflect users' interest of certain spaces or in other words areas of interest (AOIs).

Areas of interest (AOIs) are frequently visited urban areas. These areas may bound pivotal landmarks that attract people's attention. Thinking about AOIs definition, it can be inferred that AOIs are subjective. It is not a necessity that official boundaries, such as of a neighborhood correlate with AOIs boundaries. Adding to that, AOIs can be different from a group of people to another since peoples' interests are different due to their nationality, age, gender etc. As research has previously demonstrated, AOIs can be derived from clustering the points (Geotagged social media content) (Hu et al., 2015). Furthermore, VGI data can be used to spatially and temporally distinguish footprints of people's' interest (Koylu, Zhao, & Shao, 2019). Since the term footprint has been defined as “an area subject to human activity” (Antarctic

Environmental Officers Network AEON, 2005) and further described as “the geographical spread of outputs” (Antarctic and Southern Ocean Coalition ASOC, 2011) an aggregation method such as Kernel Density Estimation (KDE) can be used to generate temporal and spatial footprints (Grothe & Schaab, 2009).

1.2 Research identification

1.2.1 Research objectives

The general objective of the research is to extract urban areas of interest (AOIs) and footprints from geosocial media GSM and visualize them using suitable aggregation methods. As mentioned in the previous section, areas of interest can be different according to the difference in nationality of a certain group (class) of users. Considering this fact, this will help us to spatially analyze and compare the spatial distribution of densities between the classes and the general spatial density of the data. In addition, performing a temporal analysis to compare the evolution of densities during time on annual/seasonal/monthly basis, will definitely help to study and understand the evolution of the spatial interest pattern for a certain group of people.

An aggregation method is required to visually model and analyze different spatial interest patterns of GSM posts. The points can be aggregated using a density estimation method such as Kernel Density Estimation (KDE) to generate footprints (Grothe & Schaab, 2009) or a suitable clustering algorithm such as K-Means, DBSCAN or HDBSCAN to identify AOIs (Hu et al., 2015). Using two methods to visualize and analyze the data would give us the chance to criticize the capabilities of the used methods in spatial patterns analysis context. Especially in comparing between different spatial patterns extracted from classes with different sample sizes.

1.2.2 Research questions

1. Are the methods used in this research suitable to visualize spatial patterns of interest and perform spatial pattern analysis?
2. Is it possible to identify different spatial patterns of interest of different nationalities in the same space depending on social media data?

3. Is it possible to detect a change in spatial patterns of interest of the same nationality in the same space in monthly, seasonal and annual basis?
4. Is it possible to compare between different patterns of interest using data with big difference in sample size?

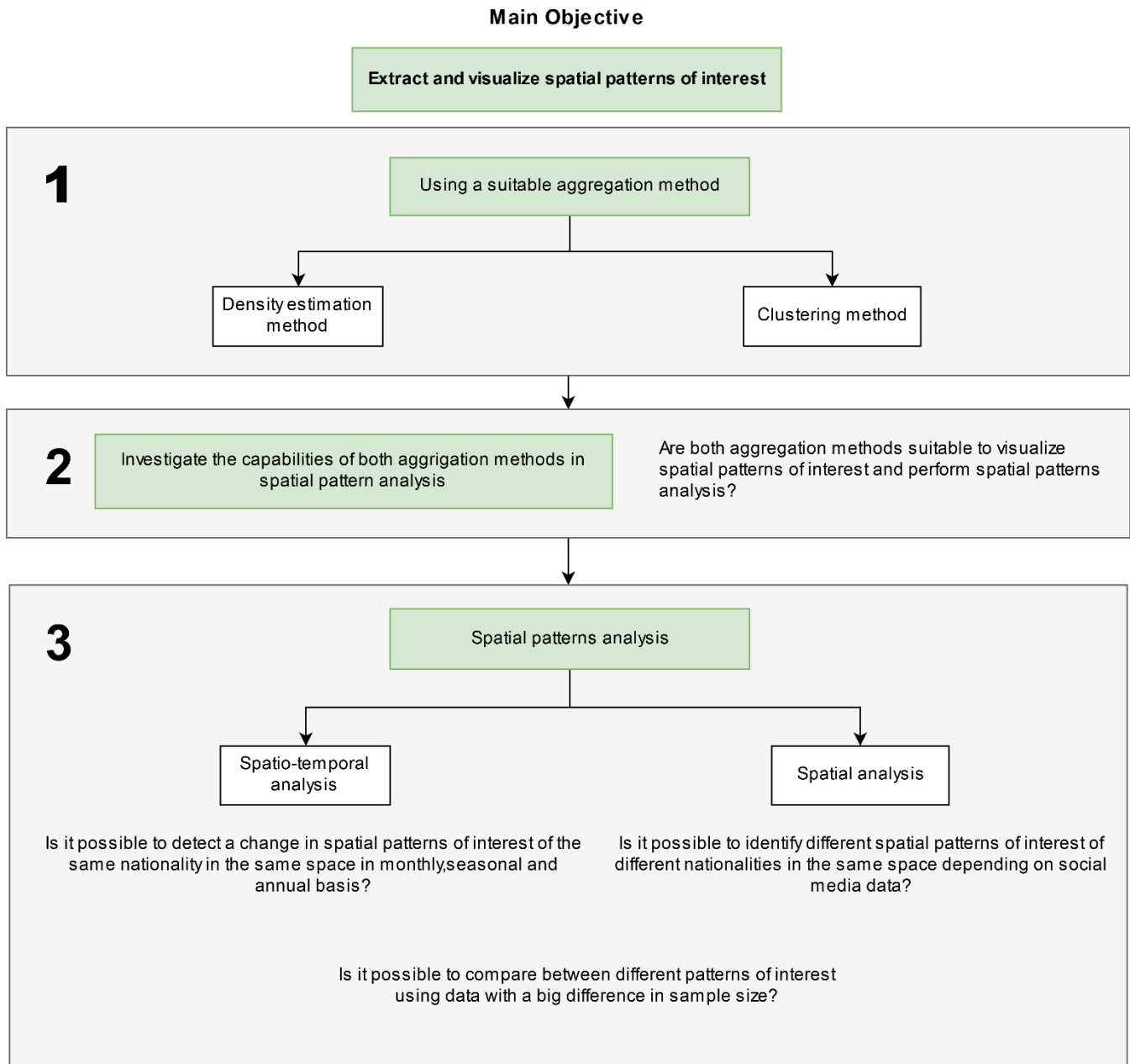


Figure 1: A diagram shows the research objectives, research questions the relational sequence of the research

1.2.3 Innovation aimed at

The key strength of this research is using two different aggregation methods to visualize the spatial patterns of interest for GSM users. This gives us the opportunity to go further with the research to criticize the suitability of both methods in spatial patterns analysis.

In addition, using an existing dataset classified based on the users' nationalities would open the door for performing a spatial patterns analysis. That will help to detect the change in these patterns based on the difference of nationality. The classified dataset used in this research was provided by (Verstockt et al., 2019)

1.3 Related work

In order to answer the research questions, it is a must to examine previous related research works. One of the most inspiring researches for this thesis is the EURECA project by (Verstockt et al., 2019) One of the main focuses of this research was identifying AOIs and footprints of tourists in different cities in Europe. Flickr data was extracted and processed to identify the origin of every Flickr user who has taken photos in the region of the extracted dataset. The used method to identify was proposed previously by (Linna Li & Michael F. Goodchild, 2012) They also used the coordinates of every photo to visualize it as points on map. Thus, they managed to visualize POIs, AOIs and footprints in Vienna and Ghent of different tourists' nationalities and in different time periods. It must be noted again that the data was courteously provided by them.

In addition, some of the mentioned terms in the introduction section need to be defined in details to avoid any ambiguity. One of the main purposes of this section is to show the difference between two important definitions in the research which are AOIs and Footprints.

1.3.1 AOIs (areas of interest)

The research of Hu *et al* is very inspiring for this thesis. They introduced a comprehensive definition for AOI. AOIs were described as:

1. Those areas that attracts peoples' interest within the city borders. For instance, those areas that bound:

- Recreation areas, defined as areas or facilities designed for either physical or psychological leisure activities (e.g., swimming pools, parks, amusement parks, sports arenas, theaters, museums) (USLegal, n.d.; Yukic, 1970).
- Landmarks, defined as remarkable spatial objects with special attributes that make them unique from the other surrounding spatial objects (e.g., buildings, memorials, statues) (Lamit, 2004).
- Shopping Centers

AOIs is a more precise term than urbanized areas. People's activities and interests is the main factor to determine AOIs inside a city. However, urbanized areas are not necessarily to be a center of people's interest.

2. Subjective to the behavior of different groups of people.

In such research, surveys could also provide insight into people's behavior in spaces. However, this would be highly consumptive of money, time and labor. In addition to surveys as a traditional data source, remote sensing is well-known as a good method to reflect physical phenomena (e.g., land use) in urban studies. But, remote sensing does not have the capabilities to track human behavior.

GSM as an alternative data source has proven reliability to provide data for spatial pattern analysis. It is more successful than surveys or remote sensing due to the high availability, low cost and the capability of GSM to track human interaction within spaces (Hu et al., 2015).

3. Can be a useful source of information for decision making.

Extraction and studying the patterns of AOIs in a city would help in decision making in many fields (e.g., urban planning, tourism, emergencies). For instance, by detecting the temporal change of AOIs, many research questions regarding future planning and prediction can be answered (Hu et al., 2015).

4. Can be identified using clustering algorithms.

The research of Hu et al introduced a solid approach using Flickr's geotagged photos for six different cities to extract and model urban areas of interest (AOIs). By virtue of HDBSCAN as a clustering algorithm, AOIs can be extracted and visualized. The followed approach of this research can also be applied to detect AOIs of other different cities using different GSM (Hu et al., 2015).

1.3.2 Footprints

The second term that needs to be discussed is footprints. As mentioned in the introduction, footprints are those areas determined by the human activity within a space (Antarctic Environmental Officers Network AEON, 2005). At this point, the definition of footprints is similar to AOIs. What makes footprints special, is that it is described as the spatial spread of human activity (Antarctic and Southern Ocean Coalition ASOC, 2011).

A suitable cartographic representation such as heat maps (intensity maps) is needed to visualize and analyze the spread of phenomena such as human behavior. By investigating the characteristics of heat maps, it can be defined as “a method for showing the geographic clustering of a phenomenon” as described by Dempsey (Dempsey, 2014). In addition, according to Yeap and Uy it can be described as “geospatial data on a map using different colors or different hues of colors to represent areas with different concentrations of points, showing overall shape and concentration trends” (Yeap & Uy, 2014).

According to Koylu *et al*, the authors used Kernel Density Estimation (KDE) as an aggregation method to estimate footprints of geotagged photos from GSM. The followed methodology has proven reliability to generate heat maps (Koylu et al., 2019).

1.3.3 Thesis structure

As shown in the table below, the master thesis consists of five chapters:

Table 1: Thesis structure

| Chapter | Content |
|---------|----------------|
| 1 | Introduction |
| 2 | Study area |
| 3 | Data |
| 4 | Methodology |
| 5 | Implementation |
| 6 | Outputs |
| 7 | Conclusions |

In the introductory chapter it has been presented to you the motivation, objectives and research questions. In the related work section, some important definitions for the research have been reviewed. The role of

the second chapter is to introduce the study area of the case study of the thesis. In the following two chapters will be discussing about the data and methodology on how to answer the research questions. In the fifth chapter, the stated methodology has been applied and redefined. Chapter six, explores in details the final implementation outputs. And finally, the conclusions and future work.

2 Study Area

One of the key strengths of the current research is to define a workflow for visualization of tourists' patterns of interests to be used to perform spatial pattern analysis in different cities. In order to accomplish this, a city with specific characteristics has to be defined as a study area, which will be defined in this chapter. Vienna has been chosen to be the study area, because of the familiarity of the author with the city, which is necessary to verify the results of the research (Hu et al., 2015). And due to the ability to verify the results in comparison with other similar researches conducted in the same study area such as (Verstockt et al., 2019) However, choosing Vienna as the study is not only because of the author familiarity of the city or the ability to verify the results, but also because of many other attributes that make the city a suitable area for touristic pattern analysis. In this section, the following attributes will be discussed:

1. History and culture
2. Location and Accessibility
3. Political situation
4. Events
5. Climate

One of the roles of this chapter is to increase the reader's familiarity level of Vienna for better understanding of the study results in further chapters. Also, it must be noted that all of the maps in this chapter were created using a series of pre-processing experiments to extract AOIs of the data provided by (Verstockt et al., 2019) The grey circles in this chapter's maps represent potential AOIs which were repeatedly identified in different steps in the pre-processing experiments to visualize the data. The diameter of each circle is defined to enclose all the POIs within each identified AOI.

2.1 History and culture

Cultural tourism is now experiencing a significant continuous growth, more than any other tourism type and more than the overall tourism rate worldwide. It is estimated by 2020 that cultural tourism will increase to 60% of the total international tourism which is estimated to be 1.6 billion arrivals worldwide (*Affiliate Members Global Report, Volume 12 – Cultural Routes and Itineraries*, 2015; Shopova, Grigorova, & Timareva, 2017). According to (International Council on Monuments and Sites ICOMOS, 1997) cultural tourism can be defined as 'That activity which enables people to experience the different ways of life of other people, thereby gaining at first hand an understanding of their customs, traditions, the physical environment, the intellectual ideas and those places of architectural,

historic, archaeological or other cultural significance. Cultural tourism differs from recreational tourism in that it seeks to gain an understanding or appreciation of the nature of the place being visited.' (Csapo, 2012). Transforming a cultural place of heritage to a touristic site, will definitely reflect the historical identity of the local inhabitants. According to the previous definition of cultural tourism, these transformed sites would fascinate and attracts those tourists whom are concerned to experience the traditional lifestyle of the inhabitants (Shopova et al., 2017).

Vienna has two of the most important world heritage sites in Europe; the Palace and Gardens of Schönbrunn and the Historic Centre of Vienna. Both of these sites has been added to the UNESCO's World Cultural Heritage list due to their recognized value (Vienna Tourist Board, n.d.-d). The palace and gardens of Schönbrunn was built over different stages. In 1569, the land was purchased by the Austrian royal House of Habsburg. However, the palace was built in 1696 for the wife of Emperor Ferdinand II, Eleonore von Gonzaga. After 1743, the palace and the gardens were redesigned and completed under Maria Theresa.

Only 45 rooms of the total 1.441 rooms in the palace can be visited. However, the gardens are open all year to be visited by public for free. The Gardens contains a lot of sculptures, flowers and the Gloriette structure. In addition, there are other places within that can be visited for a fee such as:

1. Imperial Carriage Museum
2. Crown Prince Garden
3. Orangery Garden
4. Maze & Labyrinth
5. Zoo
6. Palm House
7. Desert Experience House

The palace and gardens of Schönbrunn is considered as the most visited site in all Austria (Vienna Tourist Board, n.d.-a).

On the other hand, the historic Centre of Vienna has not been added to that list out of the blue. It is a very rich area of diverse and harmonious architecture (UNESCO World Heritage Centre, n.d.). The area is a prominent witness on the convergence of values change during the last thousand years. It remarkably contains a considerable number of buildings, gardens and monuments from middle ages; the Baroque period, and the Gründerzeit. These historical periods were used to be pivotal for the political and cultural development in Europe. In addition, Vienna is well known as the musical capital of Europe since the 16th

century (Wehdorn, Biermayer, Kreppenhofer, Scheuchel, & Zunke, 2009). The historic center of Vienna is surrounded by the 5.3 kilometer long Ringstraße. The street took 50 years to be completed. The construction work was started pursuant to the famous statement of Kaiser Franz Joseph, "It is my will ..." in December 1857. The best architects at that time carried the responsibility to make the 57 meter wide street decorated with remarkable buildings of the Historicism style, incorporating influence from Renaissance, Baroque and Gothic styles (Vienna Tourist Board, n.d.-e, n.d.-c).

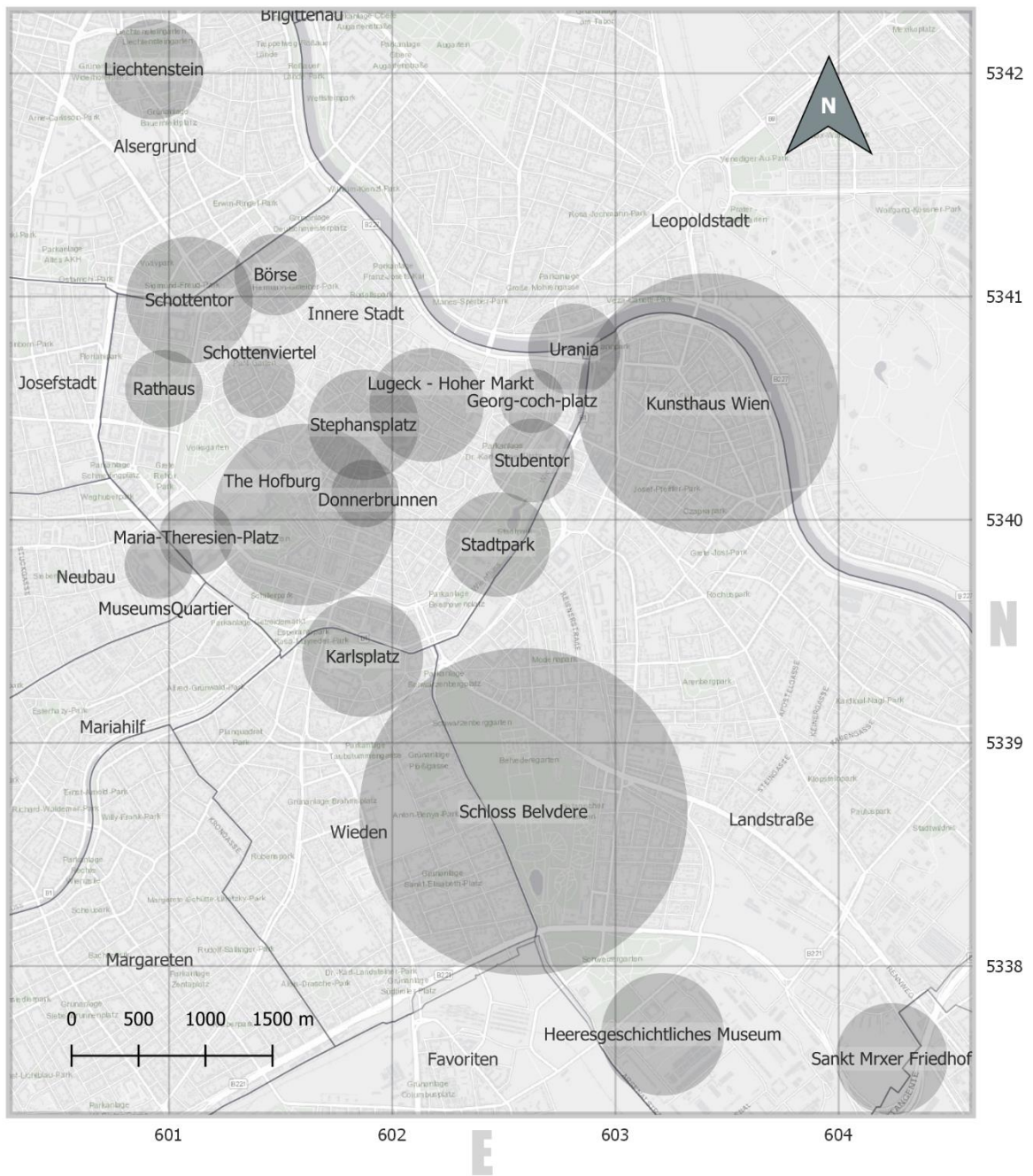


Figure 2: A map shows potential heritage AOIs within and near the city center of Vienna.

2.2 Location and accessibility

Accessibility to a touristic sites within a city is a very important touristic factor (Więckowski et al., 2014). It points out to a high quality of life and considered to be an important element for the social integration (Cristureanu & Bobirca, 2007), which in turn, definitely helps tourists to reach their touristic sites easily (Więckowski et al., 2014) and would positively affects the reputation of a city as a touristic destination. In addition, improved transportation system to a city has the most powerful leverage on location decisions for those who choose their touristic destinations based on the ease of accessibility (Cristureanu & Bobirca, 2007).

According to Buehler *et al*, the continuous renovation and enhancement of the public transport and the application of parking systems made the Viennese public transportation one of the most sustainable in the world (Buehler, Pucher, & Altshuler, 2017). In a time span of 23 years (from 1993 to 2016), the car trips has been reduced from 40% to 27% within the city. In the meantime, the public transport share has been increased from 29% to 39%. Five underground, 29 tram and 127 bus lines forming all together the well planned and implemented Viennese public transportation network (Vienna City Administration, n.d.). Visitors of Vienna can benefit of the transportation network with very cheap prices. In addition, a 1346 km of bicycle paths and lanes have been paved for public usage which collaborated to rise the share of bicycle use in the city by 3% since 1993 (Brauner, 2017). In 2003, the city of Vienna has initiated the Citybike project as the city public bike rental system (Citybike Wien, n.d.). Tourists can also benefit of using the Viennese Citybike with very low price per day.

Vienna international airport is one of Europe's leading airports. It is the main hub that connects east and west Europe (Flughafen Wien AG, 2017). The infrastructure of the airport allows it to be a regional train hub for central Europe (Antalovsky & Löw, 2018). In a 2-hour flying radius, Vienna can be reached from any capital in Europe through the city airport. It provides daily trips from/to all major cities across the globe (Austrian Economic Chamber WKO, 2018). In addition, the airport also provides a professional service for travelers. For three times in a row, it won "Best Airport Staff Europe" award for the time period of 2015 to 2017 (Flughafen Wien AG, 2017). Such reputation of an airport will be in the favor of the city tourism.

2.3 Political situation

The political stability within a country and political relations with other political entities indicates the safety level within and it is highly important for tourists to determine their destination. The political situation of a country is being reflected through media such as newspapers, TV streams, and social media networks, etc. Media is considered to be the major influencer of a country's safety reputation. Media can be selective, as it can emphasize specific interpretation of an event or a certain image of a place or a group of people. Having good relations with other political entities would help the country to be stereotyped in their media in a good way. In addition, unstable political situations such as wars, coups and protests, etc. is definitely affecting the decision of tourists for visiting a country (Ryu, 2005).

Tourism is very sensitive to the economic situation of a country (Kordic, Živković, Stanković, & Gajic, 2015). There is a positive correlation between the economic situation and local security. On the other hand, low GDP of a country usually indicates inequality and lack of security among other factors. The economic situation of Austria is pretty stable, as it is ranked fourth highest country in Europe by GDP per capita (Austrian Economic Chamber WKO, 2018). In turn, this affirmably affects the safety level of the country (Northrup & Klaer, 2014). According to Numbeo, Vienna is considered one of the safest cities in Europe with a crime rate from very low to low and a very high safety index ("Crime in Vienna," n.d.).

Austria in general and Vienna in specific are considered as a vital diplomatic center in Europe. Vienna is a home for over 40 international organizations such as (Austrian Economic Chamber WKO, 2018): (Figure.3)

- The Organization for Security and Cooperation in Europe (OSCE)
- The Organization of the Petroleum Exporting Countries (OPEC)
- United Nations (UN) as one of four headquarters around the world

Austria also is an influential member of the European Union (EU) since 1995.

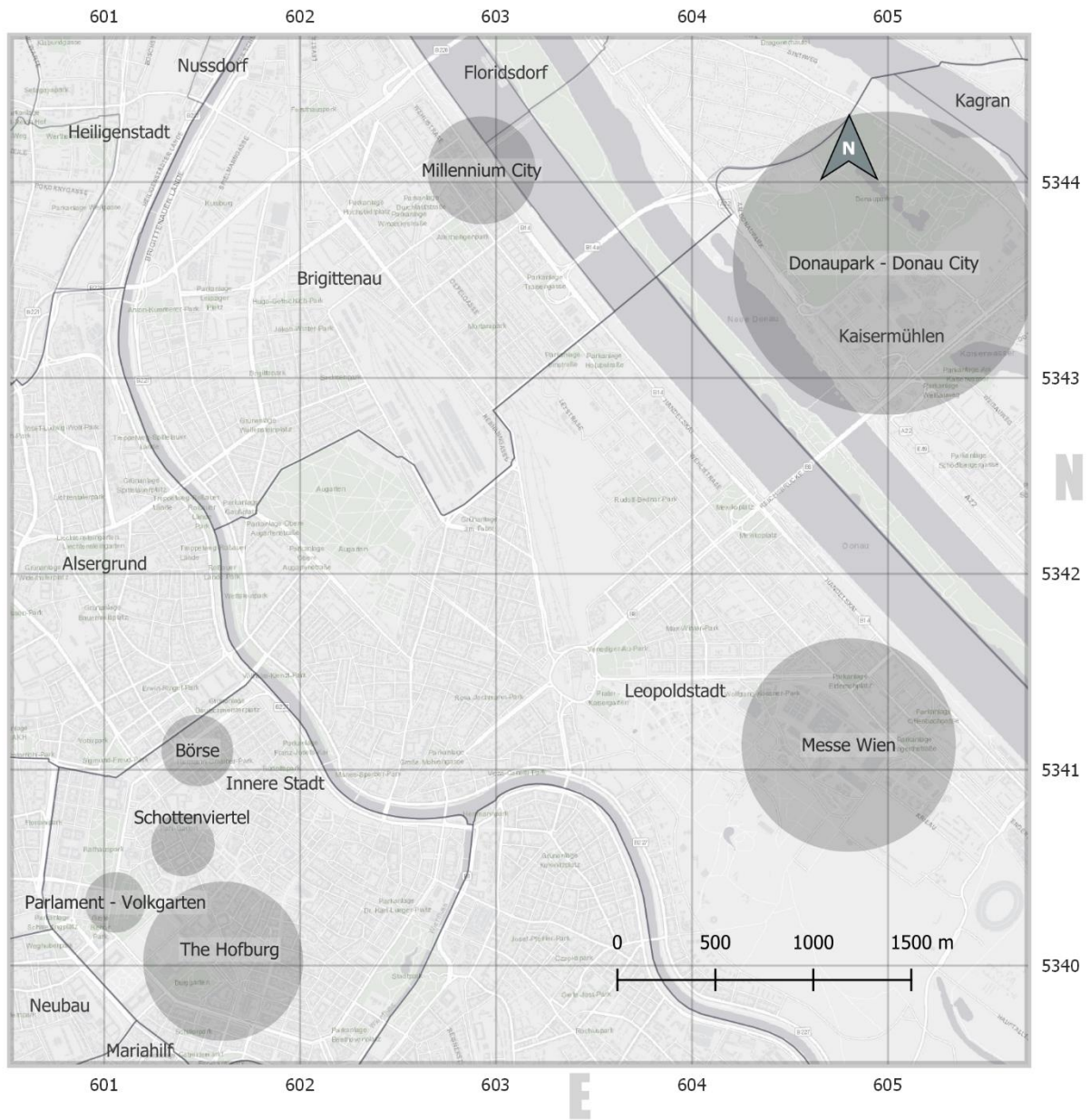


Figure 3: A map shows potential diplomatic AOIs in Vienna

2.4 Events

According to Skoultzos and Tsartas (Skoultzos & Tsartas, 2009), events is one of the most important factors for tourism growth. It contributes in:

1. Promoting the touristic destination
2. Increasing the popularity of the region as a touristic destination

3. Stimulating the accommodation business in the destination
4. Motivating the government of the destination to improve and maintain the infrastructure such as increasing the accessibility.

Luckily, the city of Vienna organizes a considerable number of events yearly on seasonal bases which in return contribute to the tourism growth in the city. During spring, a lot of events are being held in Vienna yearly such as: 1) The Vienna City Marathon on April, 2) Vienna blue spring music festival takes place during the Ester markets in April yearly, 3) There is a considerable number of concerts being held in Schönbrunn Palace through the year. The most famous of them is the Summer Night classical concert in the gardens of Schönbrunn Palace by the end of May. During summer, the city of Vienna organizes a lot of open air events to make good use of the good weather during summer (Table. 2). Winter is considered to be the off-season for tourism in Vienna and in Europe in general due to the cold climate during this time of the year. However, from the middle of November till Christmas, different Christmas markets (Table. 3) are being held in Vienna in different places which contributes to tourism growth during winter.

Table 2: List of some summer events in Vienna (Vienna Tourist Board, n.d.-b)

| Event | Date | Place |
|---|-------------------|-----------------|
| Frameout open-air cinema | Mid July - August | MuseumsQuartier |
| Rathaus Film Festival | July - August | Rathausplatz |
| Beach volleyball yearly championship (Vienna major) | July - August | Neue Danube |
| Popfest music festival | July | Karlsplatz |
| The Radio Wien Lichterfest | July | Alte Danube |
| Africa festival | Jul - August | Neue Danube |

Table 3: a list of the AOIs where the Christmas markets are being held in Vienna yearly (Vienna Tourist Board, n.d.-b)

| | | | |
|-------------------|-----------------------|------------------|-----------------|
| Rathaus | Maria-Theresien-Platz | Belvedere Palace | Prater |
| Hof 1 | Schottenviertel | Am Hof | The Hofburg |
| Schönbrunn Palace | Karlsplatz | Stephansplatz | MuseumsQuartier |

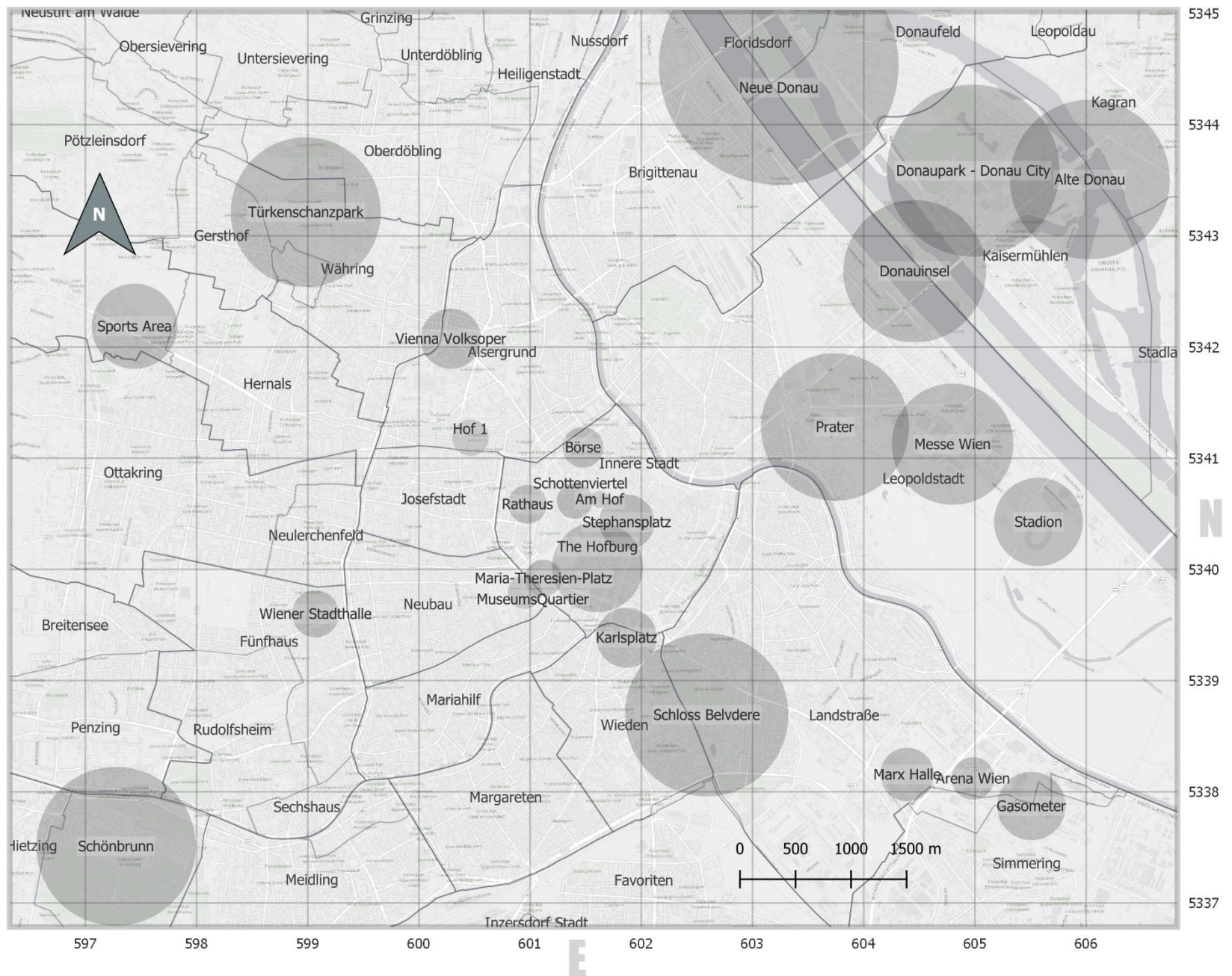


Figure 4 A map shows potential AOIs where events take place in Vienna

2.5 Climate

Climate is not only an important natural resource for tourism in any region, but it also can be the main reason for tourists' attraction (De Freitas & Higham, 2005). A lot of tourism types completely depend on certain condition of weather (European Market for Climate Services EU-MACS, n.d.) such as sport tourism or certain weather dependent events, etc. That makes tourism in some regions sensitive to weather and general climate change, which makes the economy in these regions venerable. This has motivated many institutes and governments to assess the future of climate change and its impact on tourism and economy along with forming strategies for risk management (Prettenthaler, Kortschak, & Ortmann, 2016) which in turn reflects on the tourism as well.

According to Köppen's climate classification, the climate in Vienna is described as an oceanic climate. It has a warm summer with moderate precipitation in July and August, which makes summer perfect season for touristic and recreational activities around the Donauinsel and the Donaukanal. In addition, winter in Vienna is not very snowy in comparison of other European cities, which makes Vienna a perfect European destination during this time of the year as well.

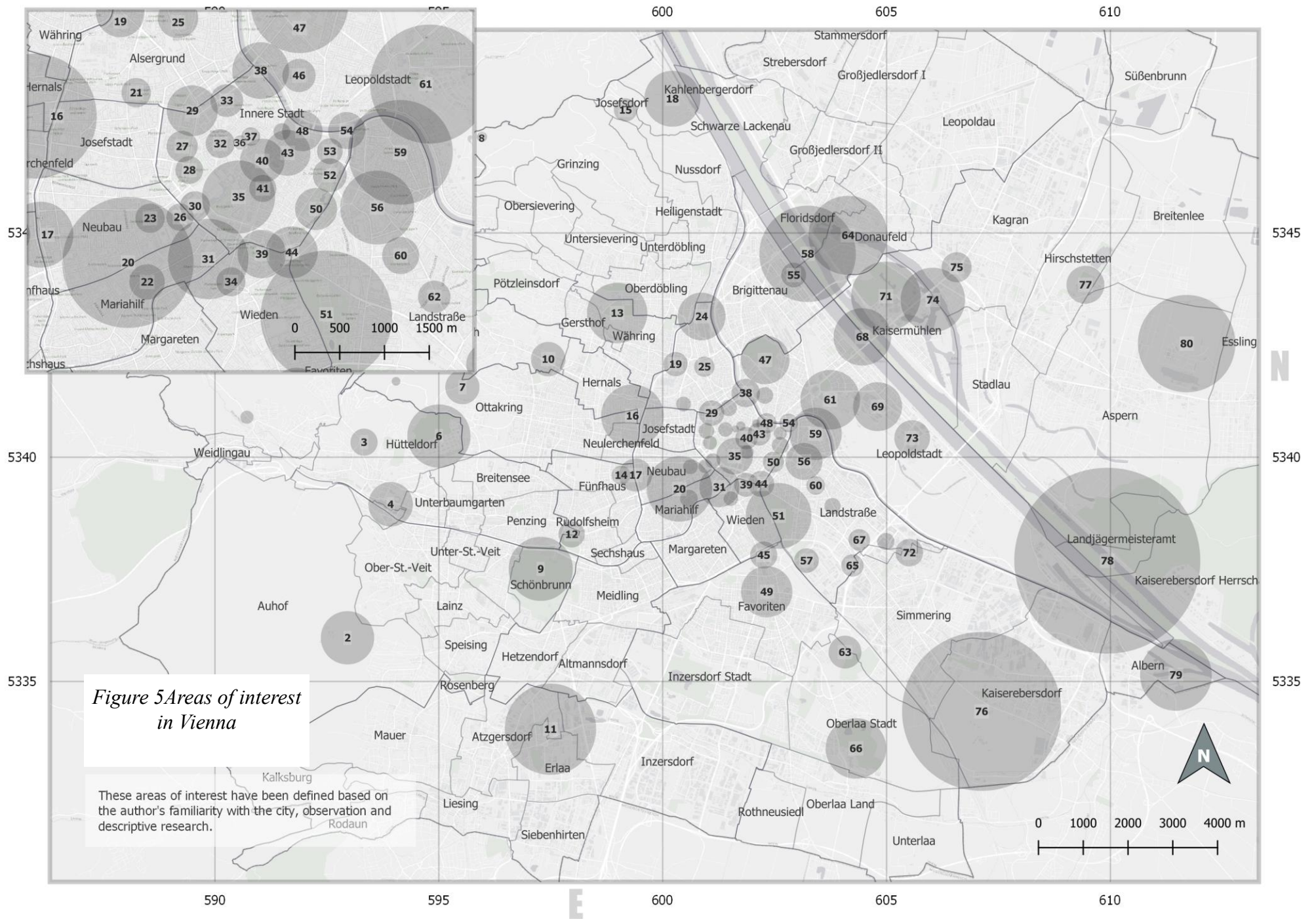


Figure 5 Areas of interest in Vienna

These areas of interest have been defined based on the author's familiarity with the city, observation and descriptive research.

Table 4: Areas of interest in Vienna.

| | | | | | | | | | |
|----|----------------------------------|----|-----------------------------------|----|-------------------------|----|---------------------------------|----|-------------------------------|
| 1 | Laudon Palace | 17 | Hauptbücherei am Gürtel | 33 | Börse | 49 | Favoritenstraße | 65 | Sankt Mrxer Friedhof |
| 2 | Hermesvilla | 18 | Leopoldsberg | 34 | Schleifmühlgasse | 50 | Stadtpark | 66 | Kurpark Oberlaa |
| 3 | Ernst Fuchs-Museum | 19 | Vienna Volksoper | 35 | The Hofburg | 51 | Schloss Belvedere | 67 | Marx Halle |
| 4 | Allianz Stadion | 20 | Mariahilfersreasse | 36 | Am Hof | 52 | Stubentor | 68 | Donauinsel |
| 5 | Jubiläumswarte | 21 | Hof 1 | 37 | Judenplatz | 53 | Georg-coch-platz | 69 | Messe Wien |
| 6 | Steinhof | 22 | Haus des Meeres Aqua Terra Zoo | 38 | Schottenring | 54 | Urania | 70 | Arena Wien |
| 7 | Wilhelminenberg | 23 | Spittelberg | 39 | Karlsplatz | 55 | Millennium City | 71 | Donaupark - Donau City |
| 8 | Hermannskogel | 24 | Spittelau incinerator | 40 | Stephansplatz | 56 | Landstraßer Hauptstraße | 72 | Gasometer |
| 9 | Schönbrunn | 25 | Liechtenstein | 41 | Donnerbrunnen | 57 | Heeresgeschichtliches Museum | 73 | Stadion |
| 10 | Sports Area | 26 | MuseumsQuartier | 42 | Ruprechtsplatz | 58 | Neue Donau | 74 | Alte Donau |
| 11 | Alterlaa | 27 | Rathaus | 43 | Lugeck - Hoher Markt | 59 | Kunsthau Wien | 75 | Donauzentrum |
| 12 | Technisches Museum | 28 | Parlament - Volkgarten | 44 | Schwarzenbergplatz | 60 | Arenbergpark | 76 | Zentralfriedhof |
| 13 | Türkenschanzpark | 29 | Schottentor | 45 | Wien Hauptbahnhof | 61 | Prater | 77 | Blumengärten Hirschstetten |
| 14 | Wiener Stadthalle | 30 | Maria-Theresien- Platz | 46 | Karmelitermarkt | 62 | Victor-Braun-Platz | 78 | Lobau |
| 15 | Kahlenberg | 31 | Naschmarkt | 47 | Augarten | 63 | Bohemian Prater, Tivoli | 79 | Alberner Hafen |
| 16 | Yppenplatz - Josefstädter Str | 32 | Schottenviertel | 48 | Swedenplatz | 64 | Alte Donau 2 | 80 | Seestadt |

3 Data

In this chapter, an overview on the used dataset will be presented. In addition, I will discuss the reasons for preferring media sharing platforms, such as Flickr, over other types of GSM, such as microblogging platforms (e.g. Twitter), as the data reference for the research.

Social platforms such as Flickr, Instagram, Twitter etc., contain a huge number of UGC posted data (Cao et al., 2010). However, a minority of the posted content of the social media platforms is geotagged (Flatow, Naaman, Xie, Volkovich, & Kanza, 2014). As previous researches have demonstrated, Twitter users outside the USA tweet around 5 to 15 million geotagged tweets per day, giving rise to a percentage of 1 to 3% of all tweets. Other social media platforms provide higher percentage of geotagged content (Oleksiak, 2014) such as Instagram, which 25% of its content is geotagged (Flatow et al., 2014). Given the fact that geotagged tweets are not necessarily related to the location where they are posted from, and the low percentage of geotagged tweets, there would be some doubts about Twitter as a reliable VGI data source for spatial analysis (Hu et al., 2015). On the other hand, Flickr (Kennedy et al., 2007), Instagram and Foursquare (Silva et al., 2013) have proven reliability to reflect users' interest of the space or in other words areas of interest (AOI). According to these facts, a dataset of Flickr data will be used for the following research.

The choice of Flickr as the source of data for the research over other media sharing platforms such as Instagram is due to the release date of both platforms. Flickr as a media sharing platform is six and a half years older than Instagram. Instagram was released on October, 2010 (Salomon, 2013) and Flickr was released on February 10, 2004 (Mao, 2015). The time difference between the two release dates makes Flickr more useful for performing a temporal analysis to compare the density evolution of tourists' activities during a wider time period than what Instagram data would provide.

Before working to extract AOIs or footprints, Flickr data has been acquired and processed by (Verstockt et al., 2019). In this regard, a working flow (Figure.6.) has been designed by them.

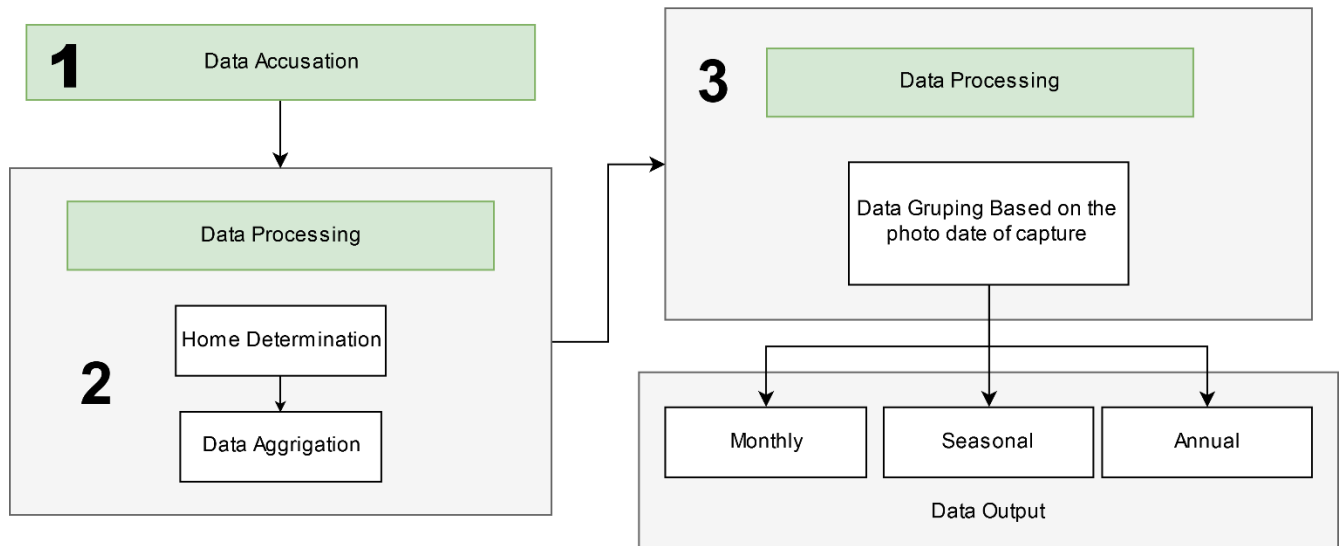


Figure 6: A work flow to prepare the data for analysis (Verstockt et al., 2019)

3.1 Raw data

(Verstockt et al., 2019) have used two Flickr APIs to get all the geo-tagged photos taken in the continent of Europe including 479 126 records were taken in Vienna from 2002 till the end of 2018. All the records of Vienna were stored by (Verstockt et al., 2019) in a geospatial postgresQL dataset. The photos were taken by 13 187 number of users. Every User will be considered as a visitor stayed in Vienna for a specific time span. The record also carries a time stamp tells when the photo has been taken.

As mentioned before, AOIs can be different from a group of people to another since people interests are different due to their nationality, age, gender, etc. In order to identify AOIs of different groups of people, the nationality of the users needs to be determined. However, the user home location was not available for more than 60% of the users. In order to determine the home location of the contributors, (Verstockt et al., 2019) developed a home classification algorithm in order to determine the country of origin of each user. The algorithm was benchmarked, reaching a precision of 81% in the classification task. The results for Vienna were as follows: The number of photos taken by Austrians (the locals) is 248.893. 44 868 photos, their owners nationalities were not identified (Table. 5)

Table 5: A brief summary of the data in Vienna

| | Number of photos |
|-----------------------------------|-------------------------|
| All records | 479 126 |
| Austrians (the locals) | 248 893 |
| Identified nationalities | 185 365 |
| Unidentified nationalities | 44 868 |

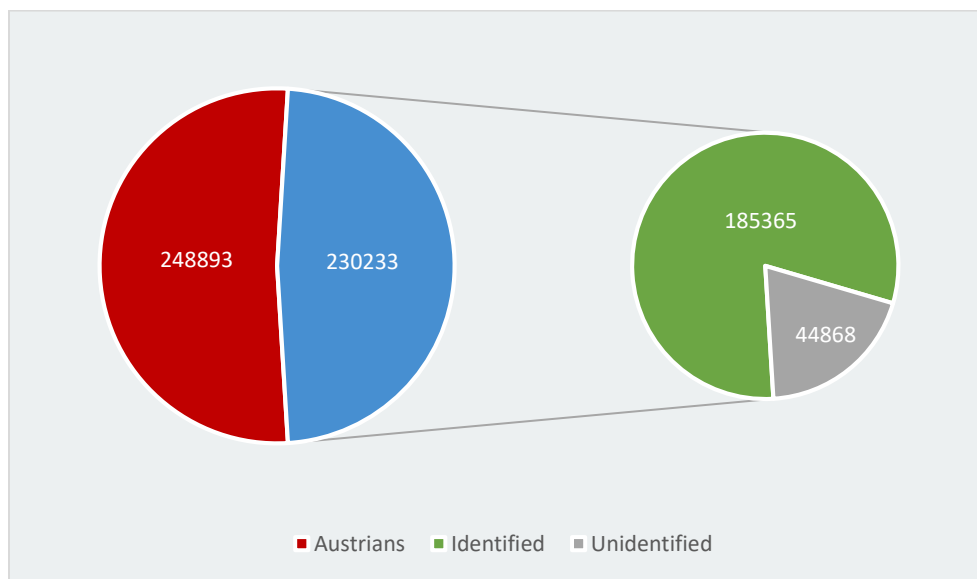


Figure 7: A pie of pie chart showing a brief summary of the raw data

In the following figure, it has been observed that:

- The numbers of photos taken by American tourists is 87.65% less than the photos taken by the locals.
- The numbers of photos taken by German tourists is 24.99% less than the photos taken by American tourists.
- All the numbers of points have dropped down by 15% and more have been highlighted with red in the previous stacked bar chart.

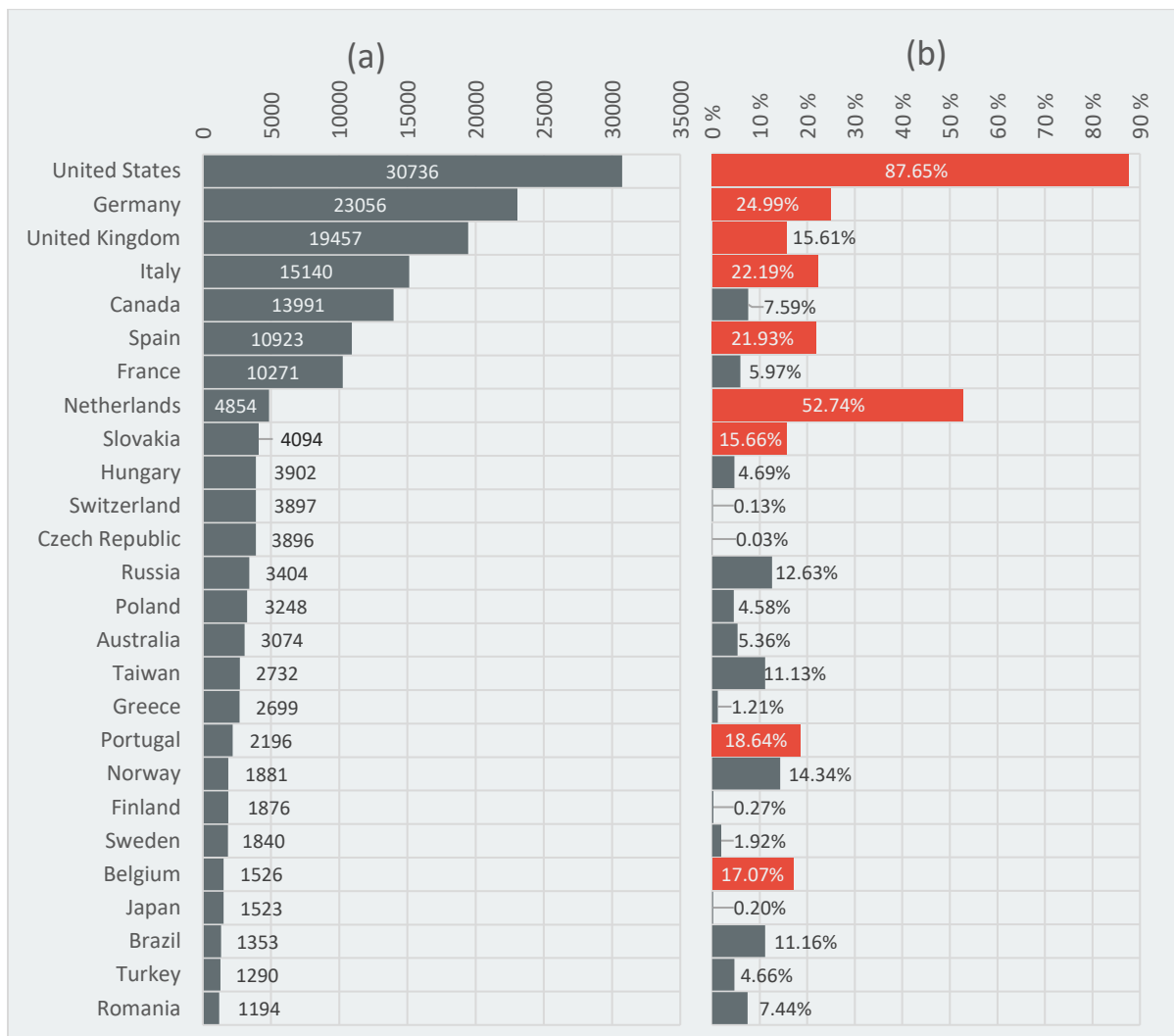


Figure 8: Stack bar chart shows (a) the number of photos taken by different nationalities in Vienna; (b) the photos number drop ratio of every nationality than the preceding nationality in list.

In order to calculate the drop ratio of every nationality than the preceding nationality in list, the following equation has been used:

$$\frac{\text{Pointscountoftheprecedingcountry} - \text{pointscountofthecountry}}{\text{Pointscountoftheprecedingcountry}} * 100\%$$

3.2 Aggregated data

Intuitively, it is highly expected that a tourist may take more than one photo of every POI visited. It is also possible that tourists may share the taken photos or some of them on their favorite social media platforms. Therefore, in the Flickr dataset it may be found that multiple photos have been taken in the same coordinates and the same date by the same user. Whereas, only one record of every user per POI is enough to be counted as one visit per POI. Thus, all the other records are considered as redundant data to be removed (Verstockt et al., 2019).

In Vienna, it was found that 287 817 number of photos are redundant and removed. The output database now contains 191 309 records representing the number of visits for every POI in Vienna.

Table 6: A brief summary of the data after aggregation in comparison with the raw data in Vienna

| | Number of photos Before aggregation | Number of visits After aggregation |
|-----------------------------------|--|---------------------------------------|
| All records | 479 126 | 191 309 |
| Austrians (the locals) | 248 893 | 81 626 |
| Identified nationalities | 185 365 | 86 633 |
| Unidentified nationalities | 44 868 | 23050 |

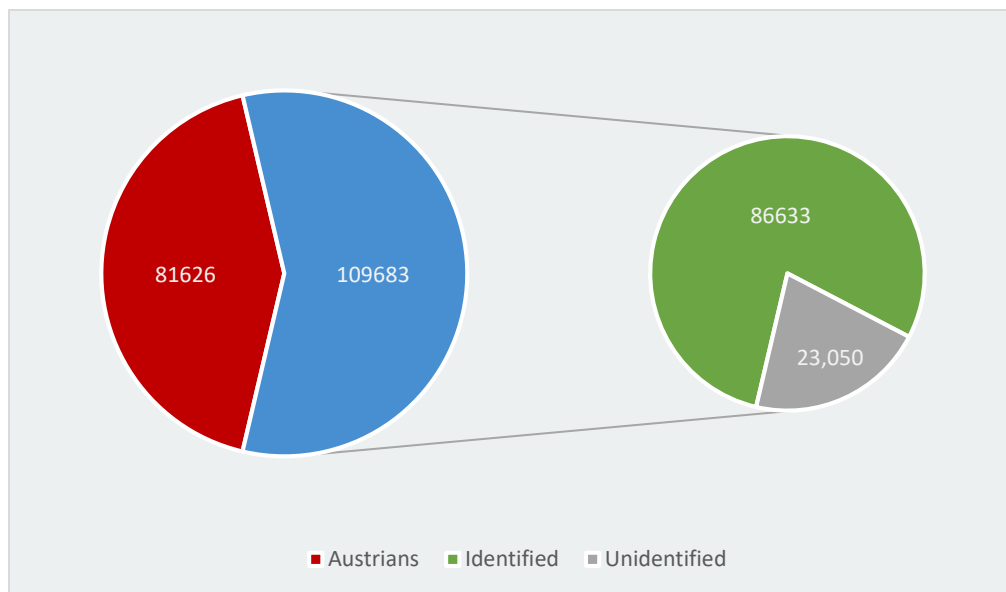


Figure 9: A pie of pie chart showing a brief summary the data after aggregation

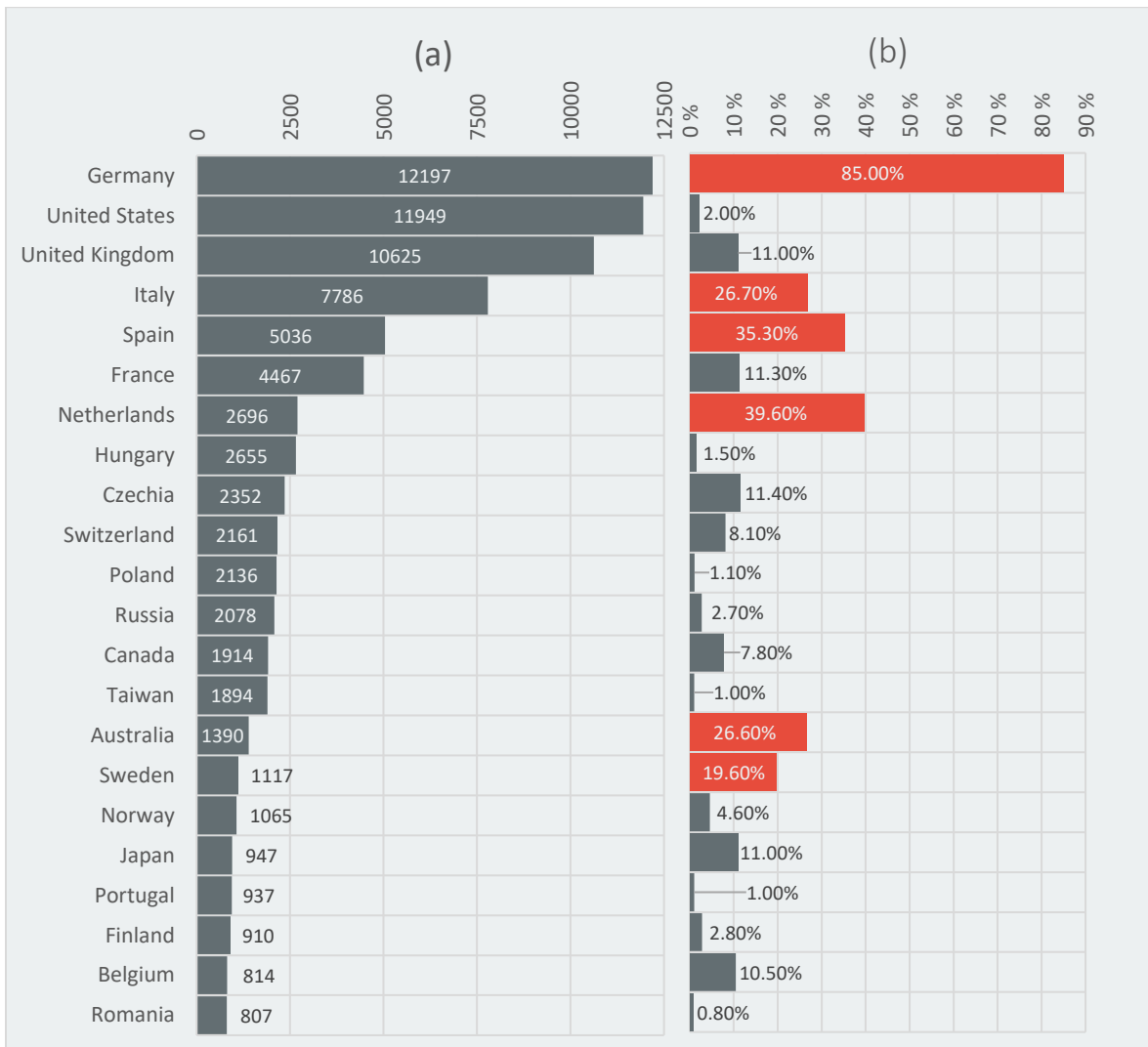


Figure 10: Stack bar chart shows (a) the number of visits by different nationalities in Vienna; (b) the visits number drop ratio of every nationality than the preceding nationality in list.

It has been found that:

- German tourists' number of visits is 85% less than the locals' number of visits.
- American tourists' number of visits is 2% less than the German tourists' number of visits.

It is very clear that Dutch tourists' visits is 39.60% less than French tourists' visits in Vienna. This drop ratio is the highest in the previous figure after the drop ratio of German tourists' visits. In addition, the sample size of Dutch tourists' visits is relatively small. I wouldn't be able to derive much information by analyzing a relatively small sample (Gerokostopoulos, Guo, & Pohl, 2015) in comparison to the Austrians sample size. Considering this fact, the required sample size to identify areas of interest should be as big as possible. In this regard, identifying AOIs in this research will be limited to the highest 7 countries in visits number.

3.3 Data analysis

Performing a temporal analysis to compare the evolution of visits numbers during time on annual/seasonal/monthly basis, will definitely help to study and understand the evolution of the spatial interest pattern for a certain group of people.

3.3.1 Temporal analysis on annual basis

As mentioned before in this chapter, Flickr was released on February 10, 2004. The number of Flickr users has increased since then. The number of photo uploads has reached its peak in the period between 2013 till 2015. The number of photo uploads started to decline starting from May 2015 since Google Photos was released (Michel, 2013, n.d.) Now Flickr is acquired by SmugMug since April 2018. Since the acquiring, the number of photo uploads started to decline even more due to the new policies over the free accounts.

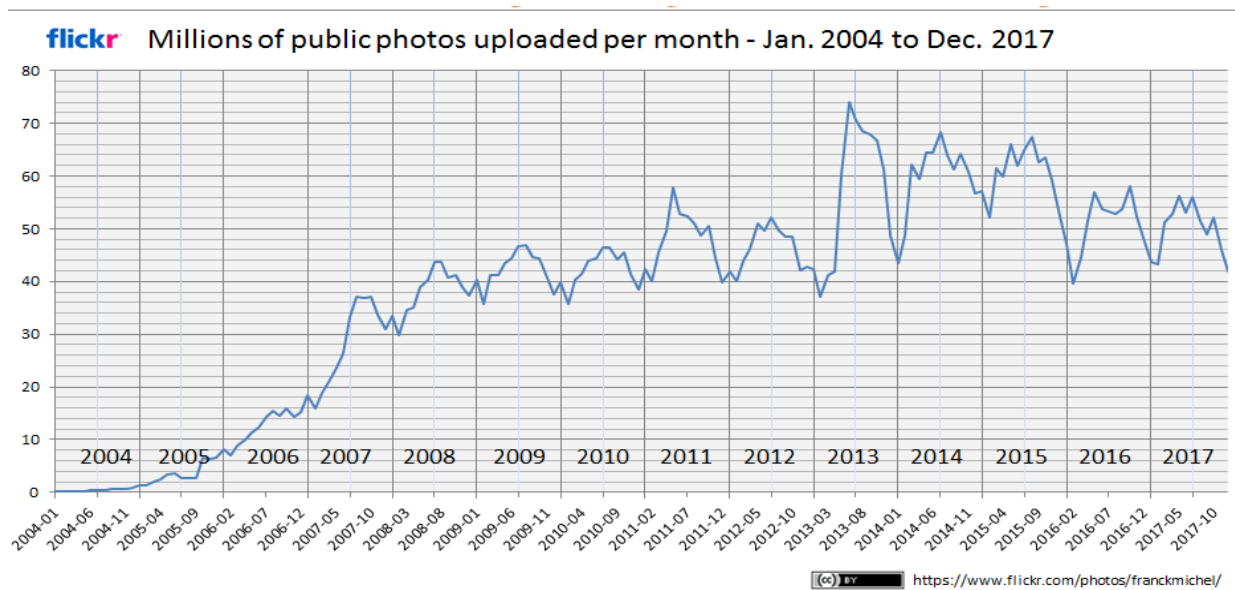


Figure 11: Flickr number of public photo uploads from January 2004 to December 2017 (Michel, 2013, n.d.)

The data of the previous figure is collected by the author via Flickr API (Michel, n.d.)

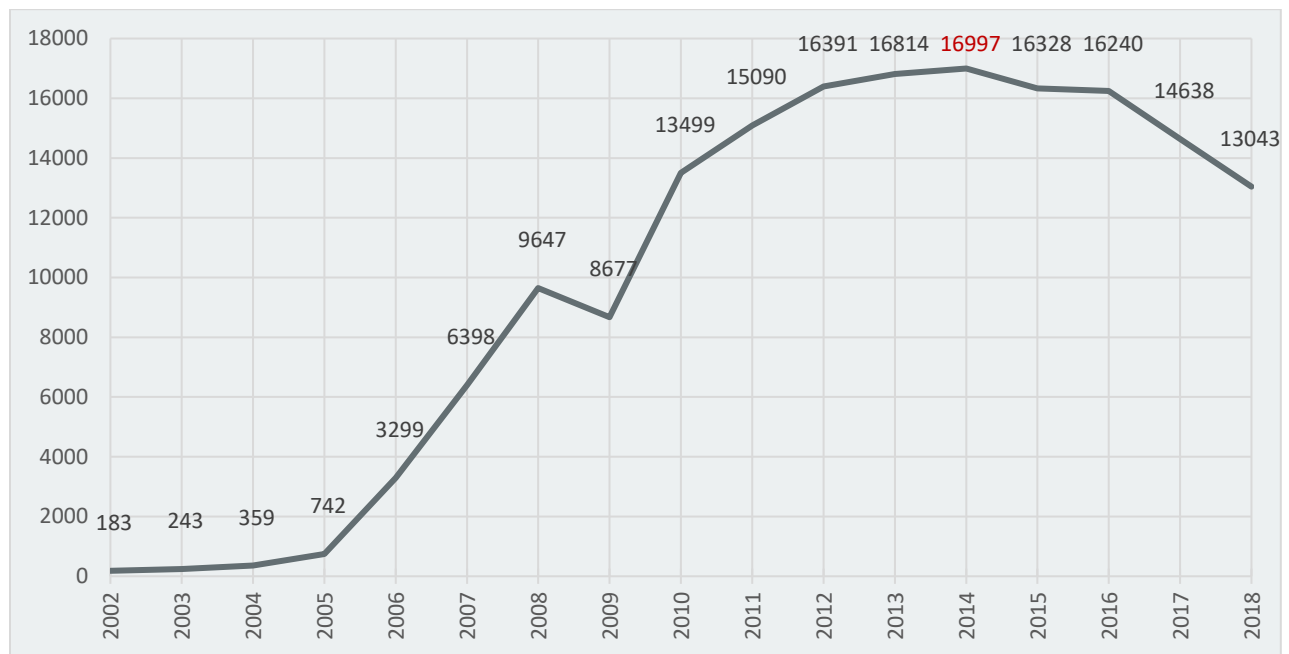


Figure 12: A line chart shows the total number of visits in Vienna on yearly basis. The demonstrated data is extracted from the Flickr dataset, which covers a 17 year time-span from 2002 till 2018.

By analyzing both graphs, it was found out that the trend of the total number of visits in Vienna is similar to the total photo upload yearly trend. Thus, the extracted number of visits doesn't represent an actual number. However, it acts like an indicator that helps us to compare the visit rate between different groups of people. By analyzing Figure 13, the visits rates of different nationalities in the same year can be compared, for example:

- The rate of the Taiwanese tourists' visits in 2012 is higher than the average rate during the same year for the entire dataset.
- On the other hand, the visits rate of the Taiwanese people in 2014 is much less than the average rate during the same year for the entire dataset.
- The year of 2014 has witnessed an increase in the visits rate of Hungarian tourists more than the average rate during the same year for the entire dataset.

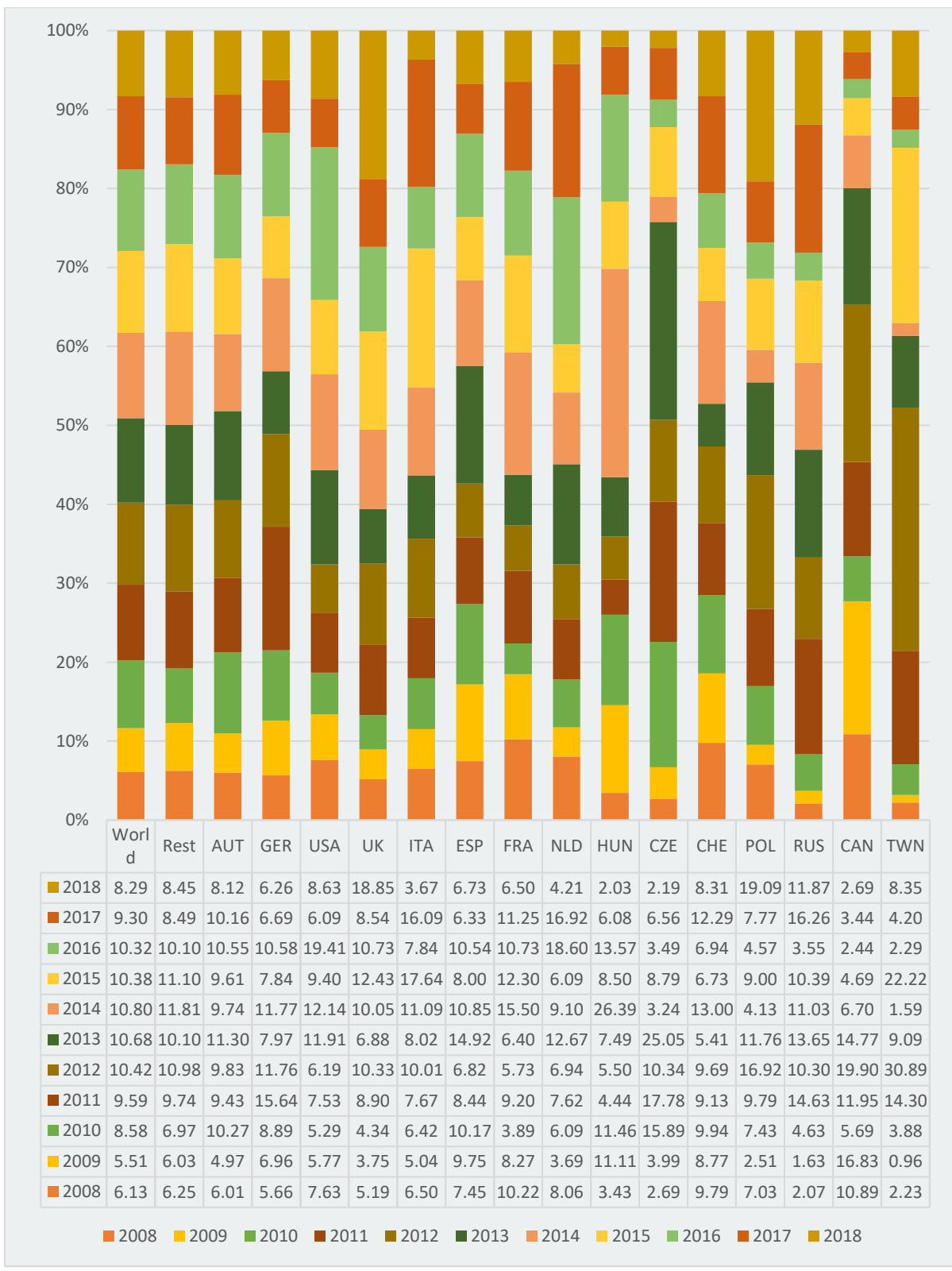


Figure 13: A stacked bar chart shows the visits rate of different nationalities to all POIs within Vienna on yearly basis.

3.3.2 Temporal analysis on seasonal basis

As it can be derived from the figure below, the busiest tourism season in Vienna is summer with average visit rate of 30.86%. Meanwhile, it can be clearly derived that winter is the off-season for touristic activity in Vienna with average visit rate of 18.37%. With a visit rate of 31.68%, it can be inferred that the Czech tourists visited Vienna during winter more frequent than the average visit rate during the same season for the entire dataset. And this is the highest visit rate of Czech tourists throughout the year. The Taiwanese tourists has the highest visit rate during spring than any other nationality as shown in Figure 14.

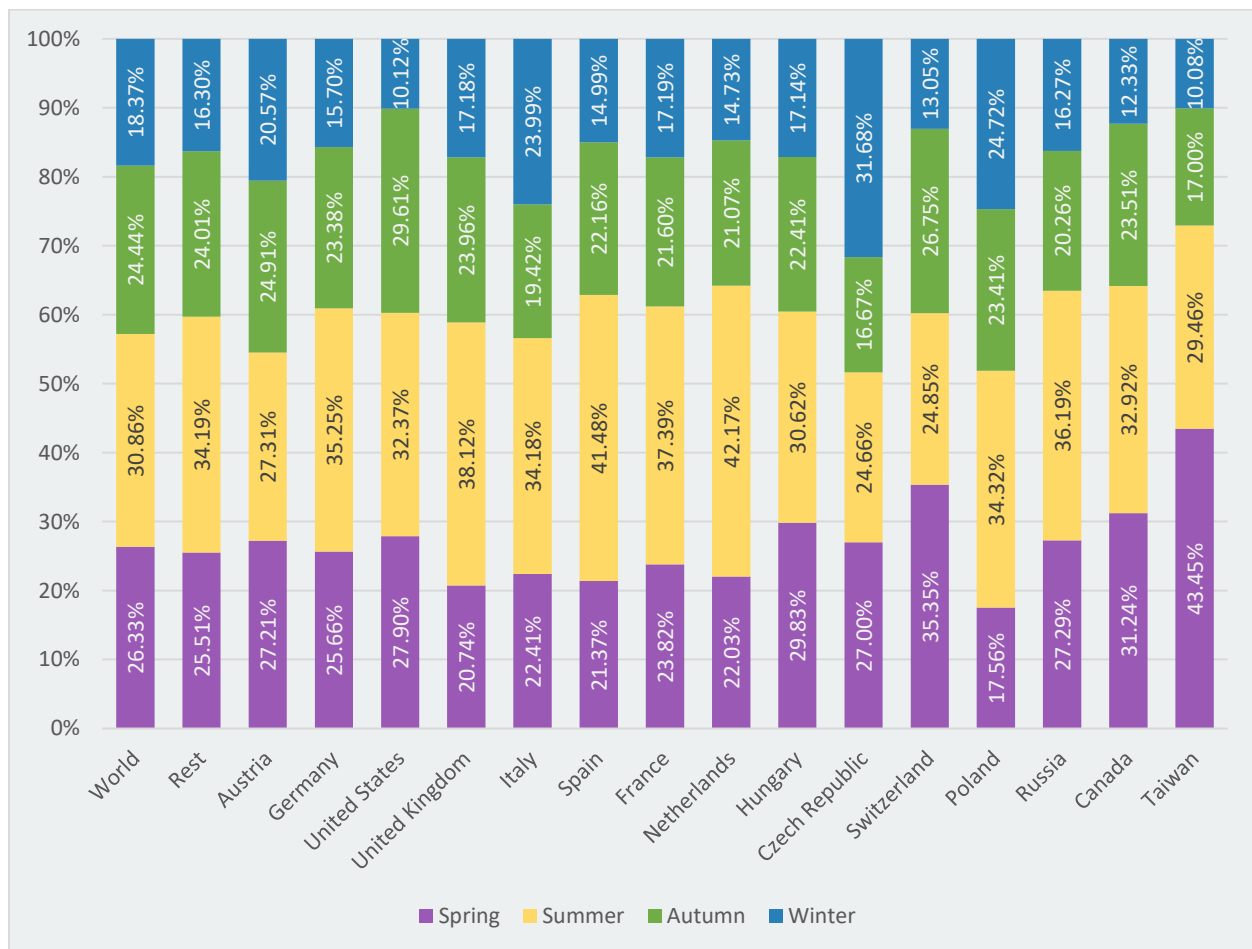


Figure 14: A stacked bar chart shows the visits rate of different nationalities to all POIs within Vienna on seasonal basis.

3.3.3 Temporal analysis on monthly basis

By analyzing Figure 15, the visits rates of different nationalities in the same month can be compared, for example:

- The rate of the Taiwanese tourists' visits in May is 13.9% higher than the average rate during the same Month for the entire dataset.
- On the other hand, the visits rate of the Taiwanese people in November is much less than the average rate during the same year for the entire dataset.
- In June, the visits rate of Russian tourists is 9.22% higher than the average rate during the same month for the entire dataset.

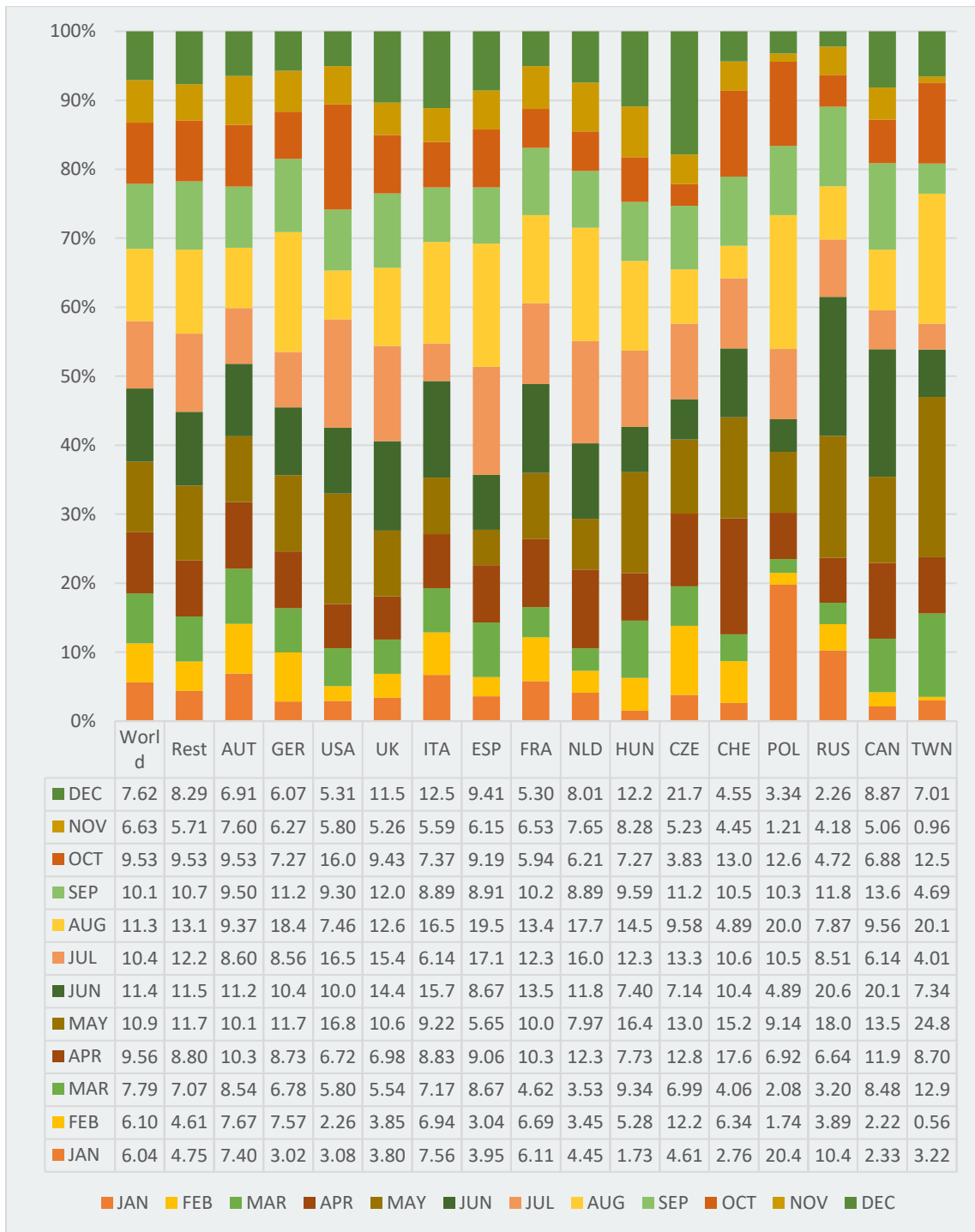


Figure 15: A stacked bar chart shows the visits rate of different nationalities to all POIs within Vienna on monthly basis.

4 Methodology

4.1 Visualizing areas of interest

Areas of interest are those areas that may interest a group of people but not necessarily to attract everyone's attention. In addition, the interest in those areas could be temporal (seasonal, annual, or even event-dependent). With a view to visualize AOIs in Vienna, the following framework has been designed. In the previous chapter (the data section), a designed framework by (Verstockt et al., 2019) for data extraction and processing has been discussed. It was also mentioned in the referred section that each record in the extracted data represent a geo-tagged Flickr photo taken in Vienna. After extracting and processing the data by (Verstockt et al., 2019) the Flickr geotagged records can be visualized with a suitable spatial reference system.

A clustering method will be used in order to spatially cluster the processed data based on the density of the points. Each generated cluster should represent an AOI on the map. In the last step, concave polygons will be generated to embrace each cluster points within. In the following sections, detailed explanations of every step will be discussed and demonstrated with examples.

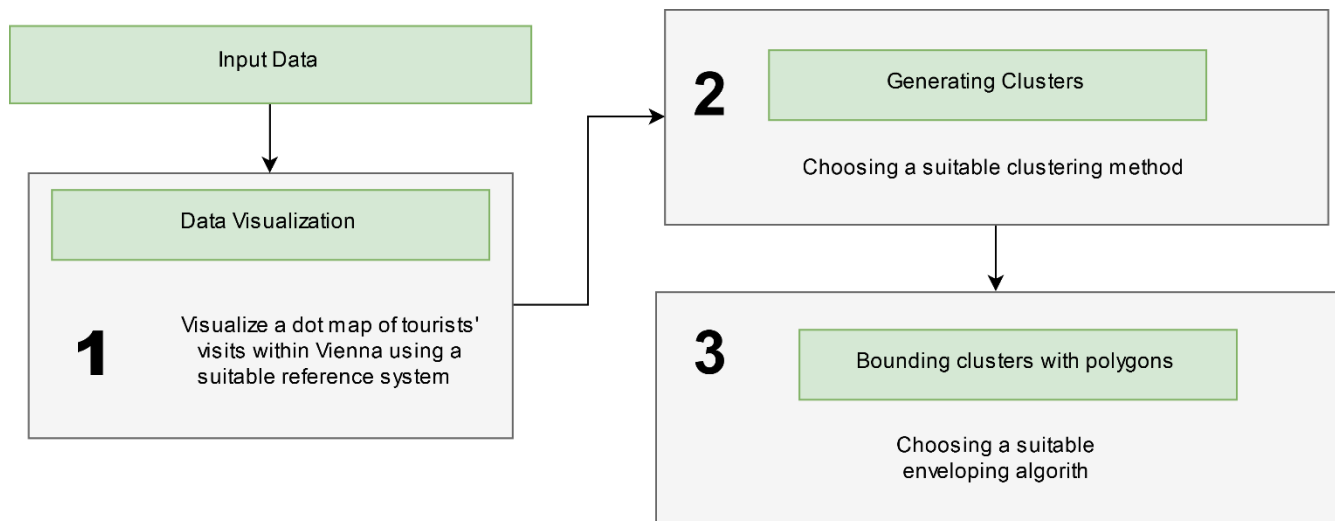


Figure 16: A framework to visualize AOIs using a GSM dataset

In the previous figure, the input data can be any suitable GSM data from any GSM platform.

4.1.1 Visualizing the data

Every record (Flickr geotagged photo) contains the coordinates of where the picture was taken. These coordinates can be used to visualize the Flickr geotagged photos as points on the map. The Universal Transverse Mercator (UTM) as a projected coordinate system and the WGS 84 datum surface will be

used as the spatial reference system throughout the research. According to the UTM system, Vienna is entirely located in UTM zone 33N.

The following figure shows a visualization of the entire records of the used Flickr data-set. As can be clearly seen from the map, the concentration of photo taking in Vienna is located in:

1. The first district
2. Schloss Belvedere
3. Prater
4. Neue Donau and Donau City
5. Schönbrunn palace and gardens

However, it is not clear the density degree of these areas. In addition, other areas that might be important but not as much as these areas are not very much identified. Also, due to the wide area of the first district and the diversity of POIs within, it is hard to consider the first district as a single AOI. As discussed in the previous chapter, 21 different AOIs within the first district can be arbitrarily defined according to the Author's familiarity with the city by reading the map. However, the visualized data presents the first district as one, wide and dense AOI.

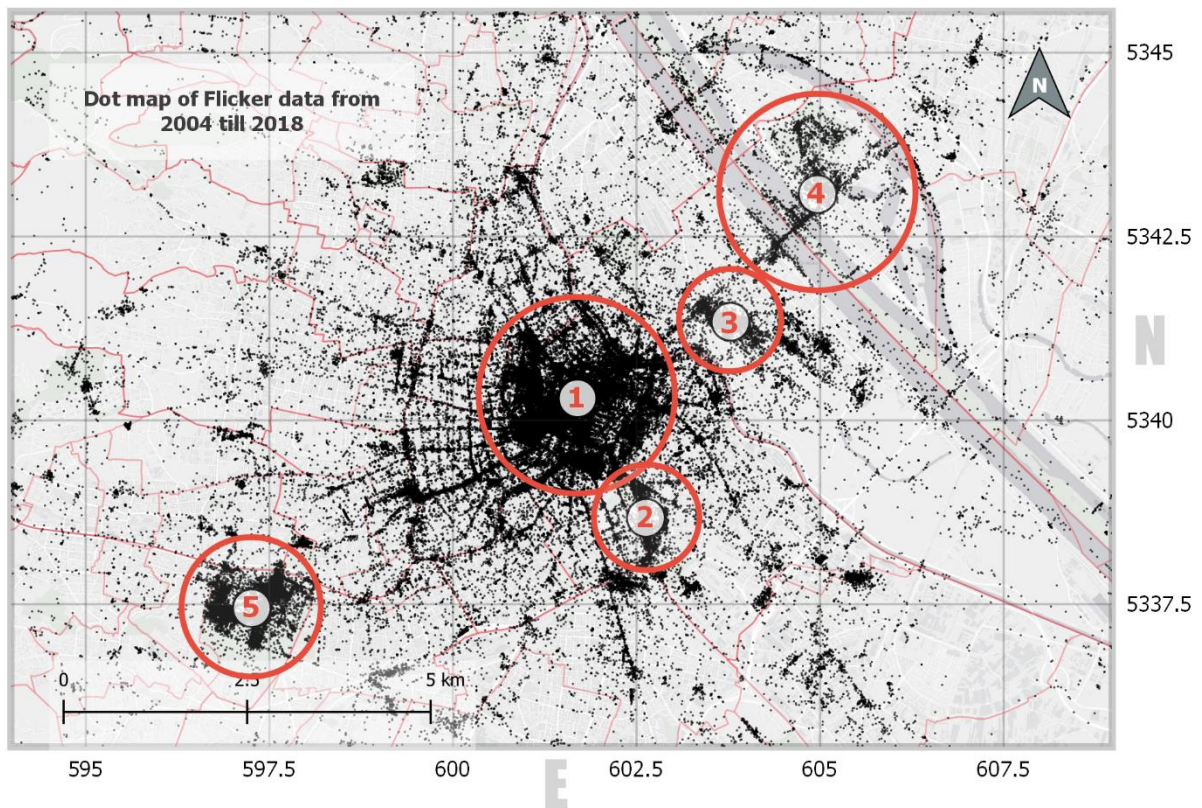


Figure 17: A dot map of tourists' visits within Vienna.

The method used in the previous visualization example is (dot mapping). It is one of the traditional methods to visualize quantitative data. The main usage of this type of visualization is to show whether the phenomenon under study is uniformly distributed throughout the study area or not (Arnold, Jenny, & White, 2017). On the other hand, identifying the similarity of densities can be delusional if the comparable areas are similar in density and different in space. In addition, it is hard to identify the boundaries of different densities.

4.1.2 Generating clusters

Due to the inability of the previous demonstrated visualization method to clearly visualize AOIs, other visualization methods have to be considered. Clustering in general is an unsupervised machine learning branch that aims to detect hidden patterns in an unreferenced data (Agostinelli, 2017) Density-based clustering methods aim to detect the patterns of similar densities in space and segment them from low density regions (noise data, outliers) (51).

According to the research of Hu *et al*, density-based clustering methods have proven ability to identify the spots of high concentration of photo taking (AOIs). The choice of density-based algorithms was based on the advantages that this type provides in comparison to other types of clustering methods such as K-means. These advantages can be listed as follow (Hu et al., 2015):

1. In density-based clustering, it is not required to identify the number of clusters beforehand. As mentioned before, every generated cluster will be considered as an AOI. Identifying the number of AOIs in a space in general or in specific time (season, month or year) beforehand is very complicated. On the other hand, in K-Means method which is considered to be a partitioning clustering method, the number of clusters has to be identified beforehand (Hu et al., 2015).
2. Arbitrarily shaped clusters can be identified effectively using density-based clustering algorithms in spherical and non-spherical data as it is clearly shown in the first, second and third row of Figure 18. It is also clearly noticeable in the same figure in the first and second row that k-means method fails to identify arbitrary shapes from non-spherical data. K-means is only effective to identify spherical-like shapes as demonstrated in the third row of Figure 18. Whereas, AOIs can be found in a wide diversity of shapes. For example, popular streets, beaches, canal sides, etc. can be identified in an elongated form. Also, AOIs can be rectangular, convex, concave, etc. according to the attraction type of the AOI. This endorses the usage of density-based clustering methods in AOIs detection (Hu et al., 2015).

3. K-means is sensitive to outliers and noise. Whereas density-based clustering is robust against data noise which can be identified as those scattered points that form less density than the typical density required to form a cluster (51, 52). Considering the fact that noise data is very common in GSM data, the usage of density-based clustering is inevitable (Hu et al., 2015).

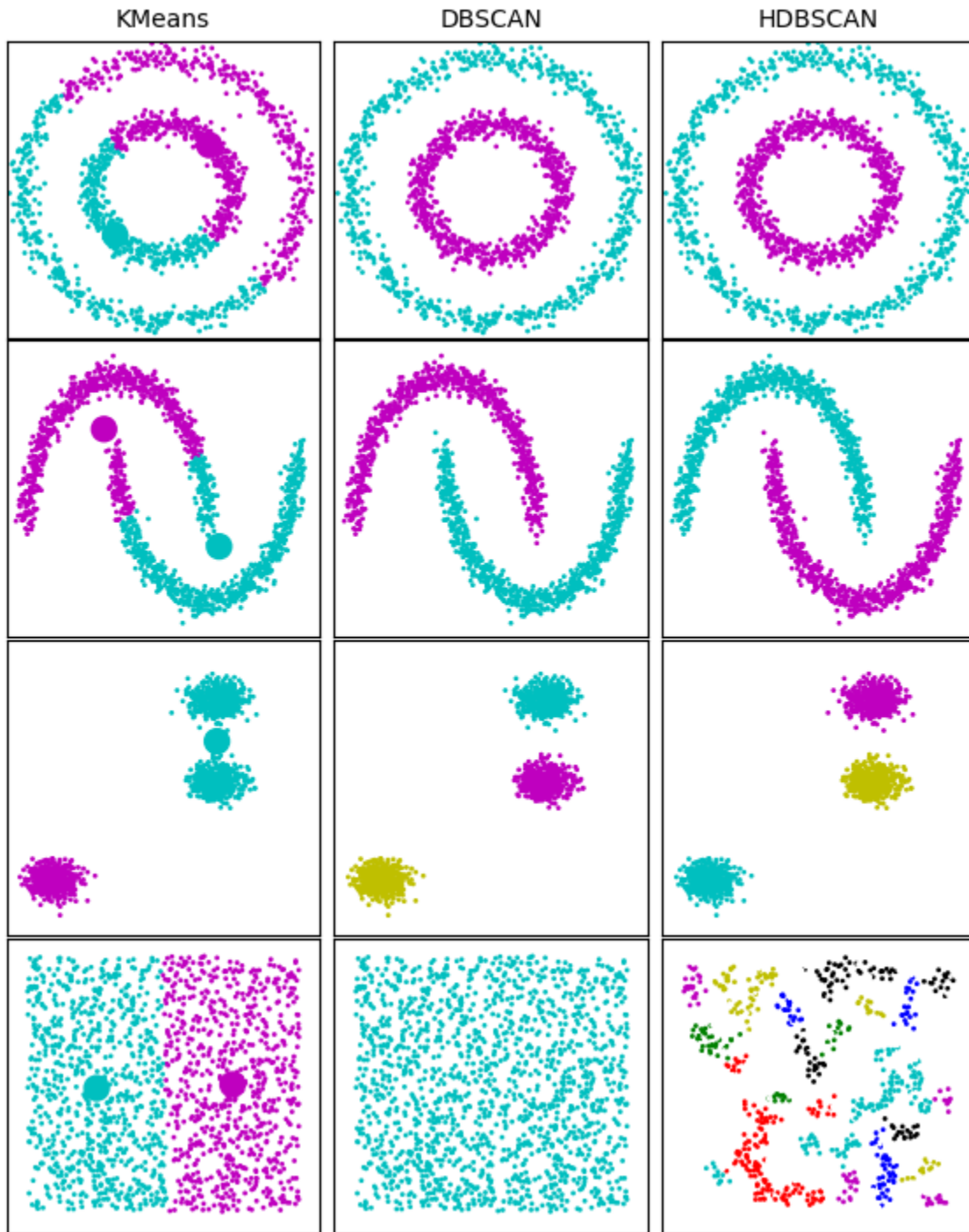


Figure 18: comparing *k*-means, DBSCAN and HDBSCAN methods using Scikit-learn toy data-set.

The author has edited following python script “Comparing different clustering algorithms on toy datasets” (Pedregosa et al., 2011) to generate the previous figure.’

There are different algorithms of type density based clustering:

1. Density-based spatial clustering of applications with noise (DBSCAN)
2. Hierarchical Density-based spatial clustering of applications with noise (HDBSCAN)

4.1.2.1 Density-based spatial clustering of applications with noise (DBSCAN)

The algorithm was first introduced by Ester *et al* in knowledge discovery in databases (KDD conference in 1996 (Ester, Kriegel, & Xu, 1996). The algorithm has been implemented in different data-science tool-kits such as scikit-learn, R and many others (Schubert, Sander, Ester, Kriegel, & Xu, 2017). It follows non-hierarchical (*i.e.* flat) clustering paradigm. The working mechanism behind this algorithm relies on identifying the minimum number of points *minPts* within an identified neighborhood radius *Eps* (Ester et al., 1996). For example, if the identified *minPts* is 4, then the number of points within a specified neighborhood *Eps* needs to be at least 4 in order to be considered as core points for a cluster. However, if the number of points within the *Eps* is less than 4 but the radius already embraces at least one core point, then all the other points within the radius are considered part of the cluster but not core points. In this example, it can be concluded that any cluster has to include at least 4 points. But it can include higher number of points than 4.

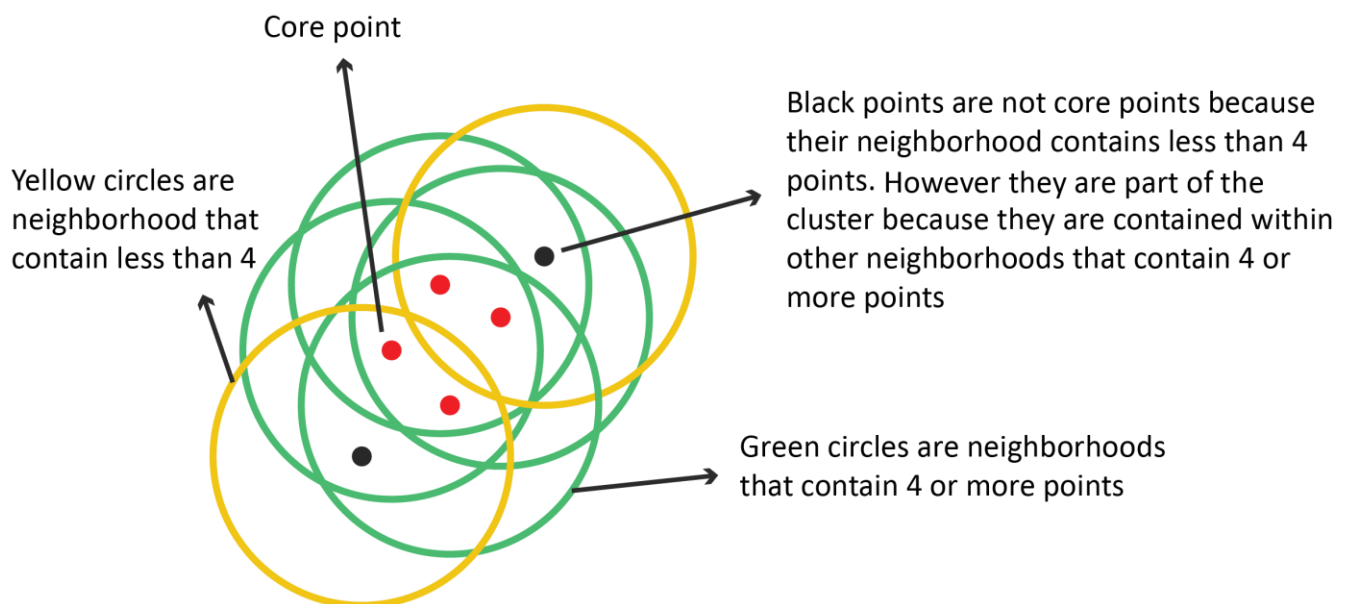


Figure 19: HDBSCAN example.

4.1.2.2 Hierarchical Density-based spatial clustering of applications with noise (HDBSCAN)

HDBSCAN algorithm has outperformed DBSCAN in AOIs detection. As demonstrated in Figure 18 rows 1, 2 and 3, DBSCAN has proven capability to detect clusters of different shapes. But as shown in row 4, HDBSCAN has proven supremacy to identify clusters of different densities in scattered data points (Yadav, 2015) On the other hand, DBSCAN has treated the same data as one cluster failing to identify the dense areas of the scattered data points. This example explains one of the drawbacks of DBSCAN algorithm. For these type of limitations and drawbacks of the density-based clustering algorithms, it became necessary to modify these algorithms for better results. In 2013, Campello *et al* have proposed a theoretical and practical method to enhance the efficiency of DBSCAN by converting it into a hierarchical clustering algorithm (Campello, Moulavi, & Sander, 2013).

HDBSCAN was implemented by (McInnes, Healy, & Astels, 2017) as a scikit-learn compatible project. However, understanding how the software works would be beneficial to identify suitable values for the parameters. The working mechanism behind this method, can be broken into 5 steps as implemented by (McInnes et al., 2017)

1. Calculating the mutual reachability distances between the sample points.
2. Structuring the minimum spanning tree of the data.
3. Structuring the cluster hierarchy.
4. Building the condensed cluster tree.
5. Extracting the cluster

The sample data used for the following example is randomly generated. The sample size is 100 points. The sample size meant to be small to make to the example simple to analyze and understand. The *minPts* in this example is 5.

The author has edited following python script (McInnes et al., 2017) to generate all the following figures in this section.

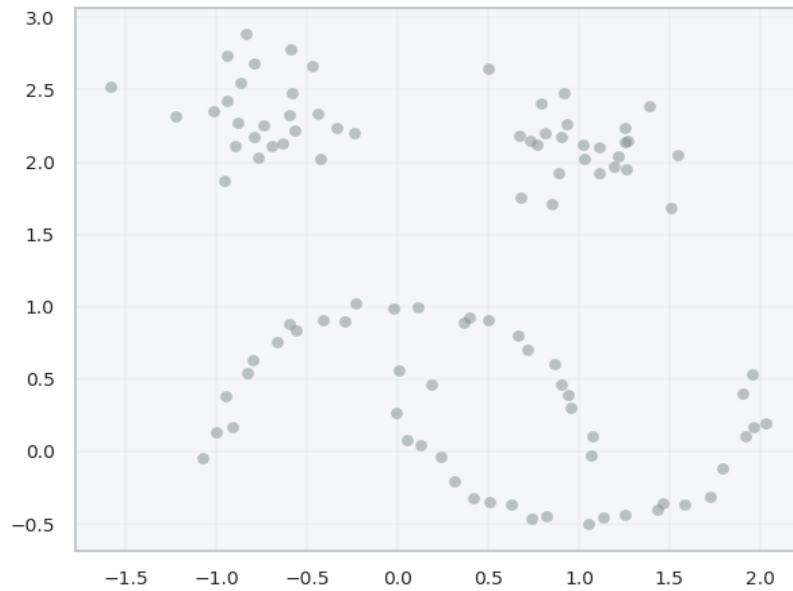


Figure 20: Sample data of 100 points on a Cartesian plane

4.1.2.2.1 Calculating the mutual reachability distances (MRDs) between the sample points.

The mutual reachability distance (MRD) is the distance in space between two points that equals or greater than the longest core distance of these points (Moulavi, Jaskowiak, Campello, Zimek, & Sander, 2014). The core distance of a point is the radius of a circle having the point as the circle's center as the circle contains other n number of points and the circumference of a circle touches the outermost point of them (McInnes et al., 2017). The value of n is identified by *minPts* which is 5 in this example.

In the following figure:

- 3 circles have been created over the points p_1 , p_2 , and p_3 as their circle centers. Every circle contains 5 points other than its center.
- The core distance of the point p_1 is the distance between p_1 and the outermost point of its n neighbors n_5 .
- The MRD between p_1 and p_2 is the distance between both points.
- The MRD between p_1 and p_3 is the distance between p_1 and p_1' . As the MRD cannot be shorter than the longest core distance of both points.

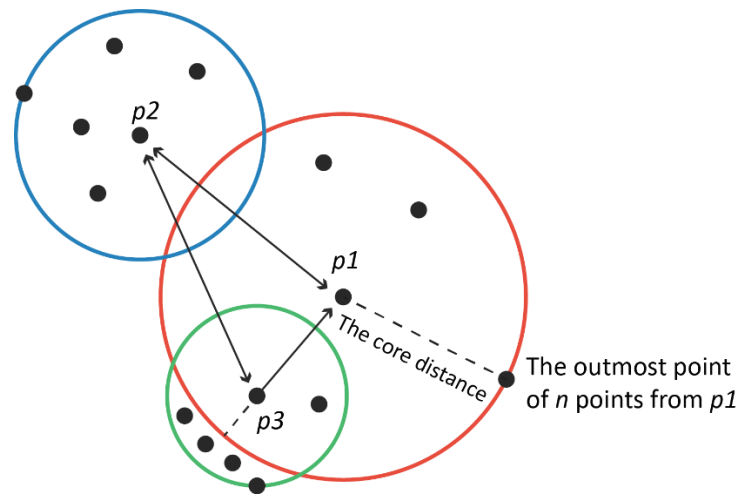


Figure 21: MRD calculation example.

Identifying all MRDs between the points gives us a weighted graph of the sample points as the graph's vertices and MRDs of the sample points as the weights of the graph's edges.

4.1.2.2.2 Structure the minimum spanning tree (MST) of the data

The next step after building the weighted graph of the mutual reach ability distance, is to find the minimum spanning tree (MST) for the weighted graph. MST is defined as follows (Akbar & Shahid-ul-Islam, 2018):

1. MTS is a tree of connected weighted edges.
2. The cost of weight is the minimum among all the other possible spanning trees.
3. It has to connect all the vertices (nodes) in the weighted graph.
4. Any node cannot loop back to itself (i.e. acyclic).

A huge number of spanning trees can be found in any weighted graph. Thus, considering an algorithm in order to find the MST is a must. It is fortunate that there are a few algorithms designed for this reason such as Kruskal's algorithm, Prim's algorithm and Boruvka's algorithm. In the following figure, Prim's algorithm has been used for finding the MST of the generated weighted graph in the previous step.

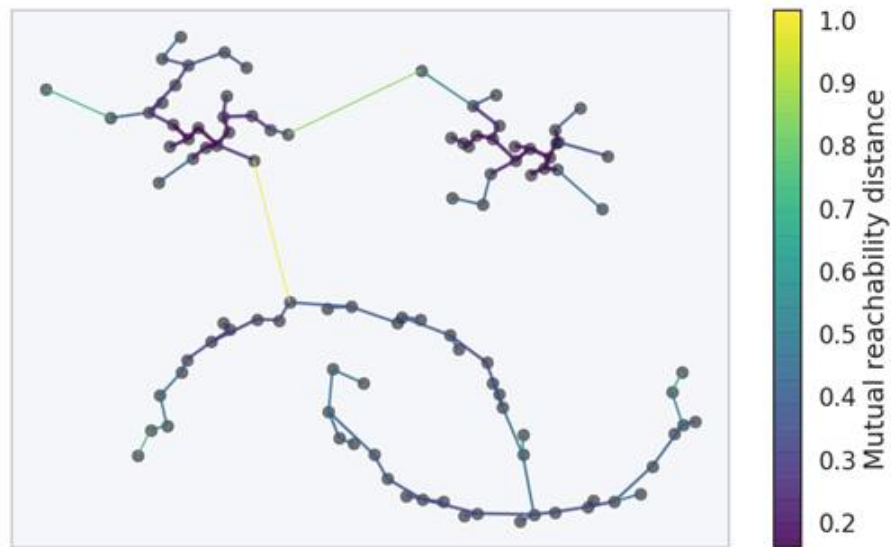


Figure 22: The minimum spanning tree (MST) of the sample data.

4.1.2.2.3 Structure the cluster hierarchy

In the previous example, clusters of different densities can be identified only by looking. That is because the simplicity of the data used in the example. However, it is very difficult to identify clusters in MST (Yu et al., 2015) in more complex data-sets.

MST is the spine of the data weighted graph and it contains implicit information about the clusters. These information can be uncovered using an agglomerative hierarchical clustering approach (Yu et al., 2015). The approach aims at creating a dendrogram of the MST by sorting all the edges of MST based on length from the smallest to the longest. Then iterating through them to create gradually merged cluster for every edge of the MST.

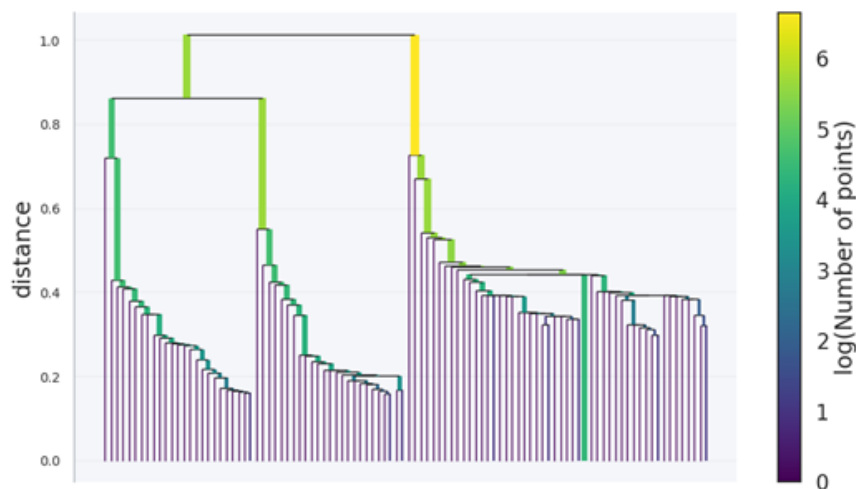


Figure 23: A dendrogram of the sample data that sorts all the edges of MST based on length.

4.1.2.2.4 Build the condensed cluster tree

To extract flat clusters, horizontal line that cuts the previous dendrogram needs to be drawn. The final clusters are those which the horizontal line cuts through. However, HDBSCAN is a hierarchical clustering algorithm. In order to perform a hierarchical clustering, a condensed cluster tree based on the previous dendrogram needs to be generated. A condensed tree can be generated by iterating through the previous dendrogram starting from the longest distance to the lowest. While iterating, the number of points of every cluster should be compared with the previously identified *minPts* at every split in the dendrogram. If the cluster size is less than the *minPts*, then this branch will fall out of the cluster and the other sibling branch will keep being part of the cluster.

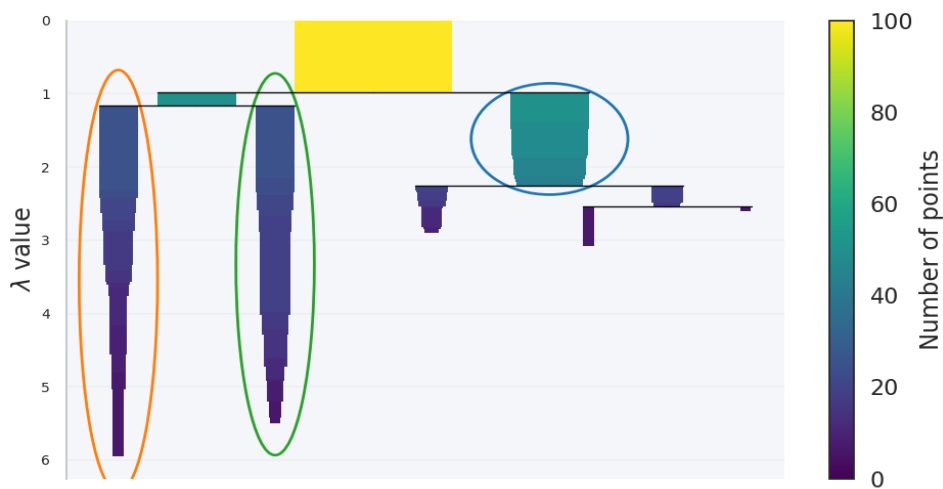


Figure 24: condensed cluster tree of the sample data.

λ is the life span of the clusters. The clusters which have longer life span are what the algorithm selects as the final clusters.

4.1.2.2.5 Extracting the clusters

In the previous step the data has been clustered into 3 clusters of points. Any point that doesn't belong to the 3 clusters will be considered as noise.

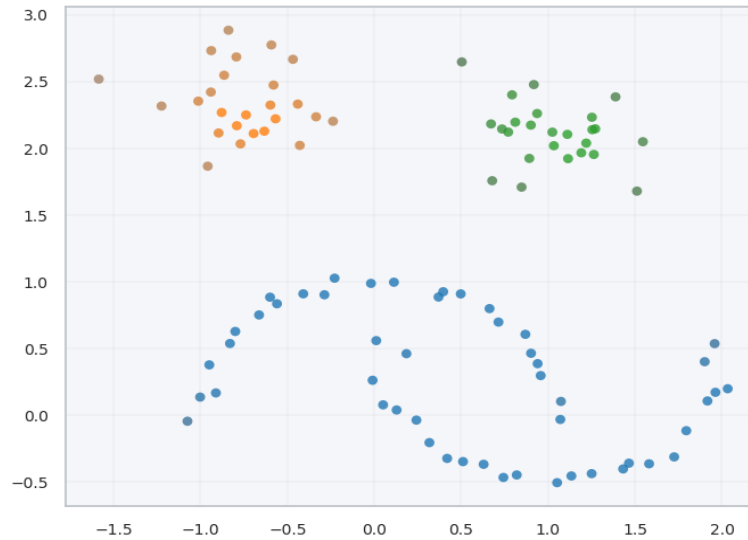


Figure 25: Visualization of the sample data after clustering

In HDBSCAN the only required parameter to identify is *minPts*. On the other hand, it is necessary to identify *minPts* and *Eps* in order to cluster data with DBSCAN.

4.1.3 Bounding the clusters with polygons

Representing areas with points on maps may not be optimum visualization approach in some cases. It could be a sufficient approach if the points were forming a block of high density dots that appear to be an area on the map. In addition, this is also dependent on the used dot symbol for representing each point and the map scale. For example, if the point symbol is small enough in contrast to the map scale, it would be very hard for the visualized points to represent an area in low densities. Thus, converting the point clusters into polygons became necessary.

4.1.3.1 Convex hull

Convex-hull algorithm is the most popular for enveloping points in a flat space. Many algorithms has been proposed for convex-hull computation such as (Graham 1972, Jarvis 1973, Preparata 1977, Eddy 1977). In-spite of the popularity of this algorithm, in some cases it has some drawbacks. In some arbitrary shaped group of points, the algorithm may not be able to compute a tight boundary around the points (Moreira & Santos, n.d.). For example, the computed polygons may contain a large empty spaced areas in some concave shaped group of points (Akdag, Eick, & Chen, 2014).

4.1.3.2 Concave hull

In 2008, Duckham *et al.* have developed an algorithm that computes possible non-convex (i.e. concave) shaped polygons for a group of points. The computed polygons should have the following characteristics (Duckham, Kulik, Worboys, & Galton, 2008):

1. The polygon should be simple. In other words, the computed polygon should have as less notches as possible.
2. It should embrace all the concerned points within.
3. The computed polygon area should be equal or less than the possible convex hull area of the same points.

The steps for concave hull computation of a finite set of points can be summarized as follows:

1. Generate a triangulated irregular network (TIN) of a set of discrete points using Delaunay triangulation method.
2. Remove all the exterior edges (*i.e.* edges which are part of only one triangle of the network) that are longer than a pre-defined length parameter l .
3. Repeat step 2 until all the edges in the TIN are shorter than l .
4. Generate the resulted shape.

The parameter l can equal to any positive number. If l equals to or less than the shortest exterior edge \min_p , then all the exterior edges of the TIN will be removed and the complicity of the generated hull will be the highest. In other words, the number of edges in the generated hull will be at its maximum same as for the number of notches which will lead to unsatisfactory shapes. If l is longer than the longest exterior edge \max_p , then none of the exterior edges will be removed and the generated hull will be the same as the possible convex hull for this set of points (Duckham *et al.*, 2008).

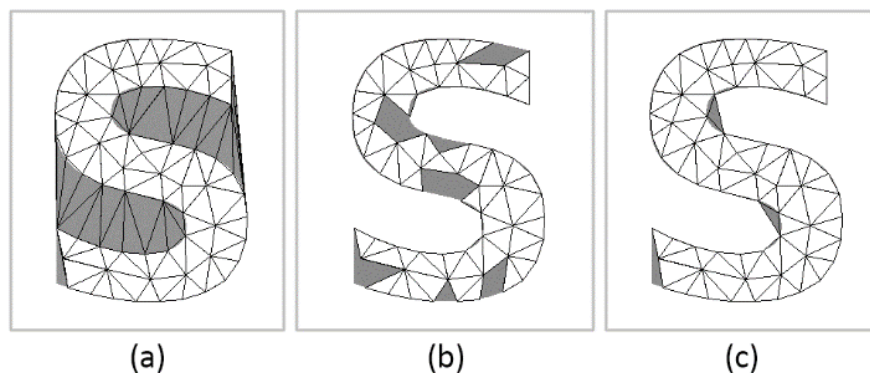


Figure 26: demonstrates the effect on the generated hull with different l values:

(a) small l value; (b) high l value; (c) optimum l value

The figure was generated using *Non-convex hulls software* (Duckham et al., 2008)

According to the stated facts in the previous paragraph, identifying the longest and shortest exterior edge is very important. Identifying them can help us to define normalized length parameter λ_P (Duckham et al., 2008).

$$\lambda_P \in [0,1]$$

$$\begin{aligned} \text{if } l < \min_P \\ \lambda_P = 0 \end{aligned}$$

$$\begin{aligned} \text{if } l \geq \max_P \\ \lambda_P = 1 \end{aligned}$$

The following fitness function is aiming to identify the optimum λ_P for a smooth computed shapes that minimizes the value of the function (Akdag et al., 2014).

$$\Phi(P, D) = \text{Emptiness}(P, D) + C * \text{Complexity}(P)$$

The computed polygons should be smooth. In other words, the polygon P should have low emptiness (i.e. empty areas) with respect to the concerned set of points D and a low complexity (Akdag et al., 2014). The shape complexity can be identified based on (Brinkhoff, Kriegel, Schneider, & Braun, 1995).

1. The number of notches.

The number of notches has a positive correlation with the vertices number of a polygon. This correlation can be described as follows:

$$\text{notches}_P \leq \text{vertices}_P - 3$$

Thus the normalized number of notches can be calculated as follows:

$$\text{notches}_{normP} = \frac{\text{notches}_P}{\text{vertices} - 3}$$

2. The convexity (i.e. deviation) of the computed concave hull from the convex hull.

Convex hull is the simplest representation of any polygon or a group of discrete points. It has an equal or bigger area than the potential concave hulls of the same point group. On the other hand, concave hulls usually have more complex shapes and a smaller areas. The deviation between both shapes areas is called the convexity of a polygon. It can be calculated as follows:

$$ConvP = \frac{Convexhullarea - Concavehullarea}{Convexhullarea}$$

3. The concave hull perimeter length.

The perimeter length of the concave hull is usually shorter than the perimeter length of the convex hull of the same point group. The deviation between both perimeter lengths can be calculated as follows:

$$\frac{Concavehullperimeter - Convexhullperimeter}{Concavehullperimeter}$$

If the result of the equation is 0, then the perimeter of the concave hull equals the perimeter of the convex hull. If the result of the equation is 1, then the deviation between both perimeters is in its highest.

4.2 Visualizing footprints

As mentioned in the introduction, footprints are those areas determined by the human activity within a space (11). At this point, the definition of footprints is similar to AOIs. What makes footprints special, is that it is described as the spatial spread of human activity (12). A suitable cartographic representation such as heat maps (intensity maps) is needed to visualize and analyze the spread of phenomena such as human behavior. According to Chainey *et al*, Kernel Density Estimation (KDE) gives more accurate results than the other common mapping algorithms for intensity maps creation (Chainey, 2013). Thus, KDE will be used throughout this research as an aggregation method to estimate footprints of geotagged photos from GSM. The followed methodology has proven reliability to generate heat maps. As it is described as “one of the most popular methods for analyzing the first order properties of a point event distribution.” According to Bailey *et al*, Silverman *et al* and Xie *et al* (Bailey & Gatrell, 1995; Silverman, 1986; Xie & Yan, 2008).

The working mechanism behind KDE calculation is very simple (Esri, 2011; Silverman, 1986):

- The method considers every pixel in the map has a parallel area on reality. Thus, this area has to be identified by the user before running the KDE algorithm. For instance, if the map units are in meters, the pixel size can be 10 meters X 10 meters or 5 meters X 5 meters. However, identifying cell size f for different applications has no specific guidelines (71).
- After running the algorithm, every pixel (kernel) will carry a value that represents the density of point features within a circular neighborhood and the concerned kernel is the center of it. This process is called mean smoothing. However, the search radius (bandwidth) h of the neighborhood has to be defined previously.

$$f(x, y) = \frac{1}{nh^2} \sum_{i=1}^n k\left(\frac{d_i}{h}\right)$$

$f(x,y)$ is the kernel calculated intensity value at x,y coordinate.

n is the number of points.

h is the bandwidth.

d_i is the distance between point i and the pixel

k is the kernel function. There are a few popular k functions to use such as Quartic, Triangular, Uniform, Triweight, Epanechnikov and more. The output of these functions usually very similar (Chainey, 2013).

As it can be derived from the previous equation, if large bandwidth h has been used, a lot of points will fall within its boundary. In return, the output heat map will be more generalized and smoother (Mikelbank, 2001).

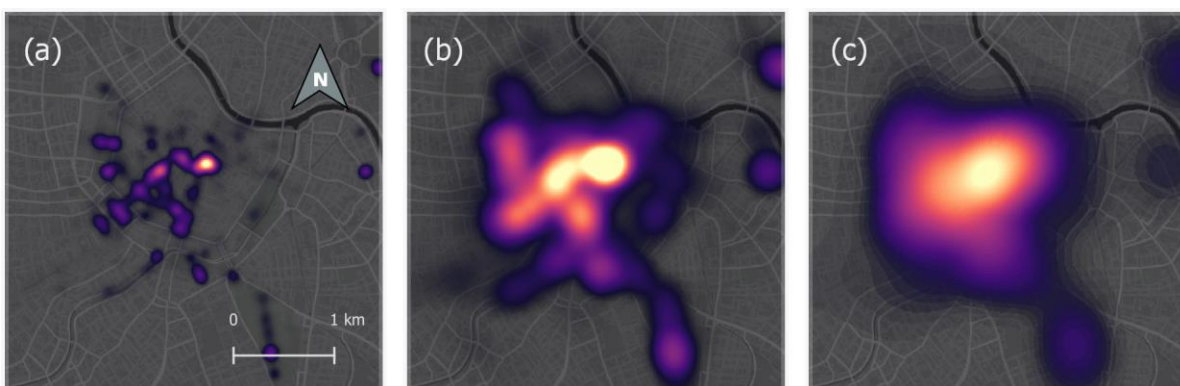


Figure 27: demonstrates the effect of different h values on heat map smoothness:

(a) $h = 100$; (b) $h = 300$; (c) $h = 600$

In order to keep the balance between smoothness and edginess, a suitable h value should be identified in different studies. The following equation can be used for that purpose.

$$h_{opt} = \left[\frac{2}{3n} \right]^{1/4} \sigma$$

h_{opt} is the optimum value of h

n is the number of points.

σ is the standard distance of the points sample

5 Implementation

In the previous chapter, work methodologies have been explained. However, some parameters values are not yet identified due to the dependency of these parameters on the case study. This chapter, will put forth some guidelines to identify the optimum values of these parameters in this case study. Therefore, these guidelines can be followed to identify the parameters values in similar cases.

5.1 Visualizing AOIs of different groups of tourists in Vienna

5.1.1 Identifying a suitable minimum number of points *minPts*

In the previous chapter *minPts* was defined as the minimum number of points required to assemble a cluster. Thus, *minPts* has a negative correlation with the produced number of clusters. For example, an experimental pre-process has been conducted on the entire dataset which consists of 191 309 points. A value of 96 was assigned to *minPts* while identifying clusters of this set of points using HDBSCAN, and the number of the output clusters was 229. On the other hand, when *minPts* value was raised to 1913, the output cluster count reduced to 9.

As mentioned in chapter 3, the points dataset is classified by different nationalities. As Figure 9 demonstrates, there is a big difference ratio between different classes. For example, the sample size of the visits by all nationalities is 191 309. On the other hand, the sample size of visits by Russian tourists is 2 078. This is considered a large difference ratio. Thus, fixing an absolute *minpts* value to identify AOIs for two groups with a large difference ratio may not be appropriate. For example, experiments assigning a fixed value of 1 000 for *minPts* while computing the clusters for two different groups of points with big difference ratio in points count gave the following results:

When the points count was 191 309, the number of clusters was 20

When the points count was 2 078, the number of clusters was 1. The DBSCAN algorithm identified the whole group of points as one cluster.

Thus, a normalized value was decided to be a better approach, and the schema of normalized values was assigned to *minPts*. For example:

- If *minPts* equals .5% of the visits number of all tourists, then *minPts* equals 957 points.
- If *minPts* equals .5% of the visits number of Russian tourists, then *minPts* equals 1 point.

Table 7: *minPts* of different normalized values for different data groups

| CTRY | World | Rest | AUT | GER | USA | UK | ITA | ESP | FRA | NLD | HUN | CZE | CHE | POL | RUS | CAN | TWN |
|------|--------|--------|-------|-------|-------|-------|------|------|------|------|------|------|------|------|------|------|------|
| | 191309 | 109683 | 81626 | 12197 | 11949 | 10625 | 7786 | 5036 | 4467 | 2696 | 2655 | 2352 | 2161 | 2136 | 2078 | 1914 | 1894 |
| .05% | 96 | 55 | 41 | 6 | 6 | 5 | 4 | 3 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| .1% | 191 | 110 | 82 | 12 | 12 | 11 | 8 | 5 | 4 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 |
| .15% | 287 | 165 | 122 | 18 | 18 | 16 | 12 | 8 | 7 | 4 | 4 | 4 | 3 | 3 | 3 | 3 | 3 |
| .2% | 383 | 219 | 163 | 24 | 24 | 21 | 16 | 10 | 9 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | 4 |
| .25% | 478 | 274 | 204 | 30 | 30 | 27 | 19 | 13 | 11 | 7 | 7 | 6 | 5 | 5 | 5 | 5 | 5 |
| .3% | 574 | 329 | 245 | 37 | 36 | 32 | 23 | 15 | 13 | 8 | 8 | 7 | 6 | 6 | 6 | 6 | 6 |
| .35% | 670 | 384 | 286 | 43 | 42 | 37 | 27 | 18 | 16 | 9 | 9 | 8 | 8 | 7 | 7 | 7 | 7 |
| .4% | 765 | 439 | 327 | 49 | 48 | 43 | 31 | 20 | 18 | 11 | 11 | 9 | 9 | 9 | 8 | 8 | 8 |
| .45% | 861 | 494 | 367 | 55 | 54 | 48 | 35 | 23 | 20 | 12 | 12 | 11 | 10 | 10 | 9 | 9 | 9 |
| .5% | 957 | 548 | 408 | 61 | 60 | 53 | 39 | 25 | 22 | 13 | 13 | 12 | 11 | 11 | 10 | 10 | 9 |
| .55% | 1052 | 603 | 449 | 67 | 66 | 58 | 43 | 28 | 25 | 15 | 15 | 13 | 12 | 12 | 11 | 11 | 10 |
| .6% | 1148 | 658 | 490 | 73 | 72 | 64 | 47 | 30 | 27 | 16 | 16 | 14 | 13 | 13 | 12 | 11 | 11 |
| .65% | 1244 | 713 | 531 | 79 | 78 | 69 | 51 | 33 | 29 | 18 | 17 | 15 | 14 | 14 | 14 | 12 | 12 |
| .7% | 1339 | 768 | 571 | 85 | 84 | 74 | 55 | 35 | 31 | 19 | 19 | 16 | 15 | 15 | 15 | 13 | 13 |
| .75% | 1435 | 823 | 612 | 91 | 90 | 80 | 58 | 38 | 34 | 20 | 20 | 18 | 16 | 16 | 16 | 14 | 14 |
| .8% | 1530 | 877 | 653 | 98 | 96 | 85 | 62 | 40 | 36 | 22 | 21 | 19 | 17 | 17 | 17 | 15 | 15 |
| .85% | 1626 | 932 | 694 | 104 | 102 | 90 | 66 | 43 | 38 | 23 | 23 | 20 | 18 | 18 | 18 | 16 | 16 |
| .9% | 1722 | 987 | 735 | 110 | 108 | 96 | 70 | 45 | 40 | 24 | 24 | 21 | 19 | 19 | 19 | 17 | 17 |
| .95% | 1817 | 1042 | 775 | 116 | 114 | 101 | 74 | 48 | 42 | 26 | 25 | 22 | 21 | 20 | 20 | 18 | 18 |
| 1% | 1913 | 1097 | 816 | 122 | 119 | 106 | 78 | 50 | 45 | 27 | 27 | 24 | 22 | 21 | 21 | 19 | 19 |

The approach was to iterate through each group of points 20 times. Each time, a different normalized value was assigned to *minPts* to identify the number of clusters in every iteration. The results were as follows:

Table 8: Number of clusters of different data groups using different normalized values of *minPts*.

| CTRY | World | Rest | AUT | GER | USA | UK | ITA | ESP | FRA | NLD | HUN | CZE | CHE | POL | RUS | CAN | TWN |
|------|--------|--------|-------|-------|-------|-------|------|------|------|------|------|------|------|------|------|------|------|
| | 191309 | 109683 | 81626 | 12197 | 11949 | 10625 | 7786 | 5036 | 4467 | 2696 | 2655 | 2352 | 2161 | 2136 | 2078 | 1914 | 1894 |
| .05% | 230 | 192 | 264 | 326 | 339 | 366 | 379 | 403 | 633 | | | | | | | | |
| .1% | 117 | 109 | 151 | 143 | 145 | 132 | 144 | 163 | 242 | 209 | 212 | 346 | 321 | 290 | 305 | 288 | 284 |
| .15% | 73 | 63 | 94 | 92 | 84 | 89 | 90 | 86 | 117 | 142 | 147 | 141 | 161 | 185 | 163 | 158 | 175 |
| .2% | 57 | 48 | 66 | 69 | 62 | 68 | 58 | 71 | 76 | 111 | 119 | 102 | 118 | 122 | 113 | 100 | 113 |
| .25% | 43 | 46 | 55 | 53 | 46 | 53 | 51 | 48 | 65 | 69 | 87 | 83 | 91 | 93 | 81 | 78 | 88 |
| .3% | 37 | 39 | 48 | 40 | 44 | 51 | 38 | 45 | 56 | 57 | 73 | 72 | 67 | 70 | 62 | 63 | 63 |
| .35% | 38 | 32 | 40 | 34 | 41 | 38 | 35 | 36 | 44 | 53 | 65 | 63 | 45 | 59 | 55 | 58 | 46 |
| .4% | 30 | 30 | 36 | 34 | 34 | 36 | 36 | 31 | 38 | 49 | 52 | 56 | 42 | 45 | 44 | 47 | 35 |
| .45% | 25 | 27 | 29 | 31 | 29 | 28 | 31 | 28 | 34 | 42 | 50 | 48 | 39 | 42 | 34 | 41 | 34 |
| .5% | 21 | 24 | 26 | 25 | 28 | 28 | 28 | 25 | 29 | 38 | 44 | 45 | 33 | 34 | 32 | 35 | 34 |
| .55% | 21 | 23 | 18 | 22 | 25 | 28 | 24 | 22 | 27 | 33 | 39 | 39 | 29 | 32 | 28 | 29 | 30 |
| .6% | 19 | 23 | 16 | 21 | 24 | 24 | 21 | 22 | 27 | 31 | 36 | 39 | 24 | 28 | 28 | 29 | 29 |
| .65% | 19 | 20 | 20 | 20 | 23 | 23 | 21 | 21 | 25 | 30 | 33 | 38 | 21 | 26 | 23 | 28 | 26 |
| .7% | 17 | 22 | 15 | 9 | 21 | 22 | 20 | 20 | 23 | 29 | 28 | 35 | 20 | 23 | 23 | 26 | 26 |
| .75% | 13 | 18 | 14 | 19 | 19 | 19 | 19 | 17 | 17 | 28 | 26 | 26 | 19 | 22 | 21 | 20 | 25 |
| .8% | 14 | 16 | 12 | 17 | 18 | 19 | 19 | 17 | 19 | 26 | 26 | 24 | 19 | 21 | 21 | 18 | 22 |
| .85% | 14 | 17 | 12 | 15 | 18 | 19 | 18 | 17 | 16 | 22 | 23 | 23 | 18 | 18 | 21 | 16 | 21 |
| .9% | 14 | 16 | 10 | 7 | 17 | 18 | 14 | 17 | 16 | 21 | 23 | 23 | 18 | 19 | 21 | 18 | 21 |
| .95% | 13 | 15 | 10 | 8 | 15 | 16 | 14 | 16 | 16 | 17 | 21 | 16 | 17 | 19 | 20 | 16 | 23 |
| 1% | 10 | 13 | 9 | 8 | 15 | 14 | 14 | 16 | 15 | 17 | 21 | 9 | 7 | 18 | 18 | 16 | 20 |

To choose a suitable normalized value of *minPts*, clusters were generated for all the possible values shown in Table 8. This yielded the following:

1. By giving a small normalized value (0.05%, 0.1% etc.) for *minPts*, HDBSCAN algorithm is able to identify AOIs in the outskirts of the city. The reason is that the tourists are more concentrated in AOIs within or near the city center. This in turn makes the photo density lower in the outskirts of the city. Thus these low density clusters can be identified by giving a small normalized value to *minPts*.
2. When *minPts* value is small, some of the clusters within or near the city center don't represent AOIs, however they represent POIs. For example (See Figure 28), Some AOIs should be identified as one unified AOI such as Karlsplatz, Schottentor and Donnerbrunnen etc. However, the AOIs were divided into many small clusters representing POIs within a presumed AOI.

- When increasing the *minPts* value bit by bit, the low density clusters (AOIs) in the outskirts of the city start to disappear. And the small clusters within or near the city center where dense areas are start to merge forming bigger AOIs.

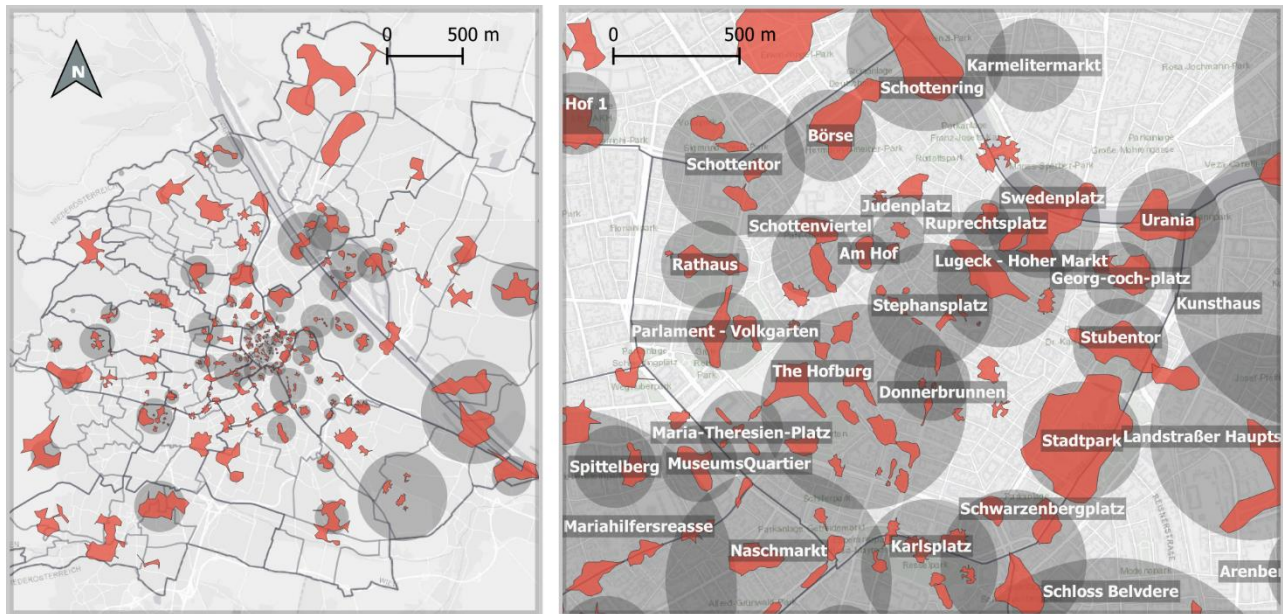


Figure 28: AOIs of all tourists in Vienna with *minPts* equals to .05% of this group number of points.

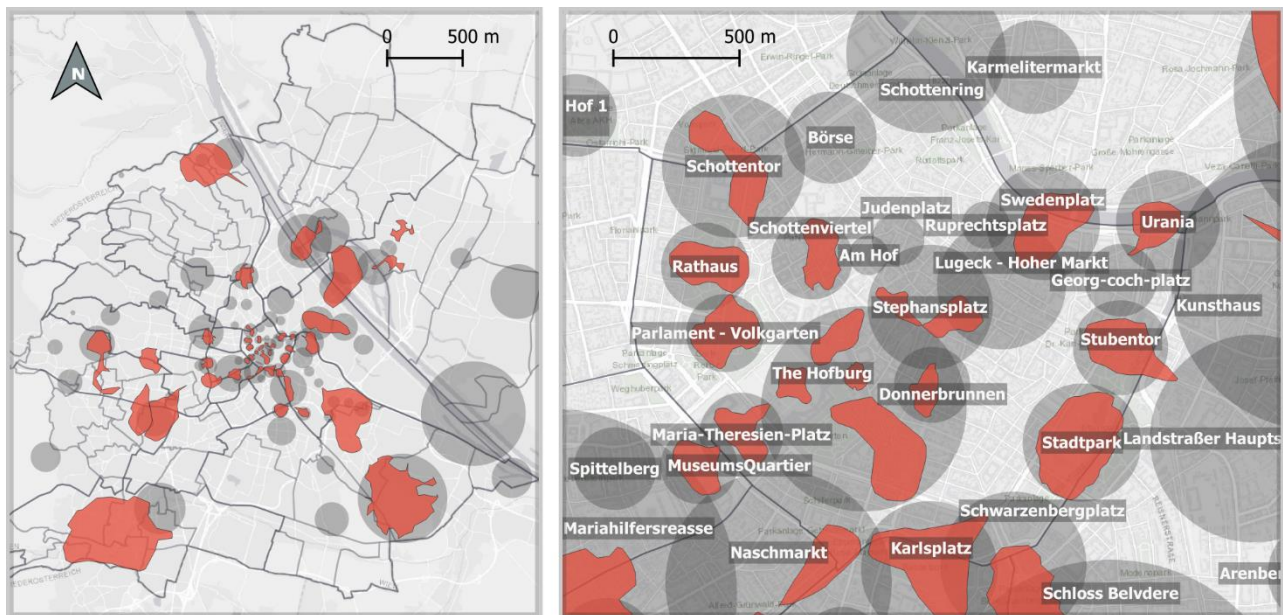


Figure 29: AOIs of all tourists in Vienna with *minPts* equals to .25% of this group number of points.

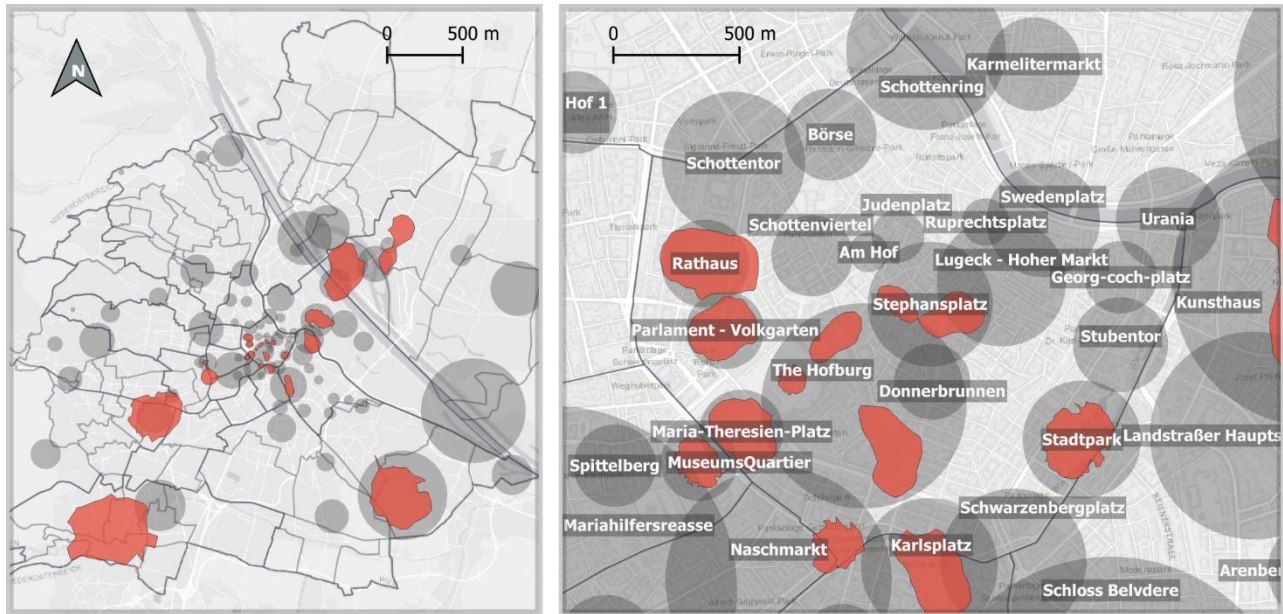


Figure 30: AOIs of all tourists in Vienna with $minPts$ equals to .5% of this group number of points.

4. In classes with small sample sizes, the normalized value of $minPts$ needs to be high enough to detect AOIs that represent reality (See Figure 31). It needs to be higher than what is should be to detect realistic AOIs in larger sample sized groups (See Figure 29).

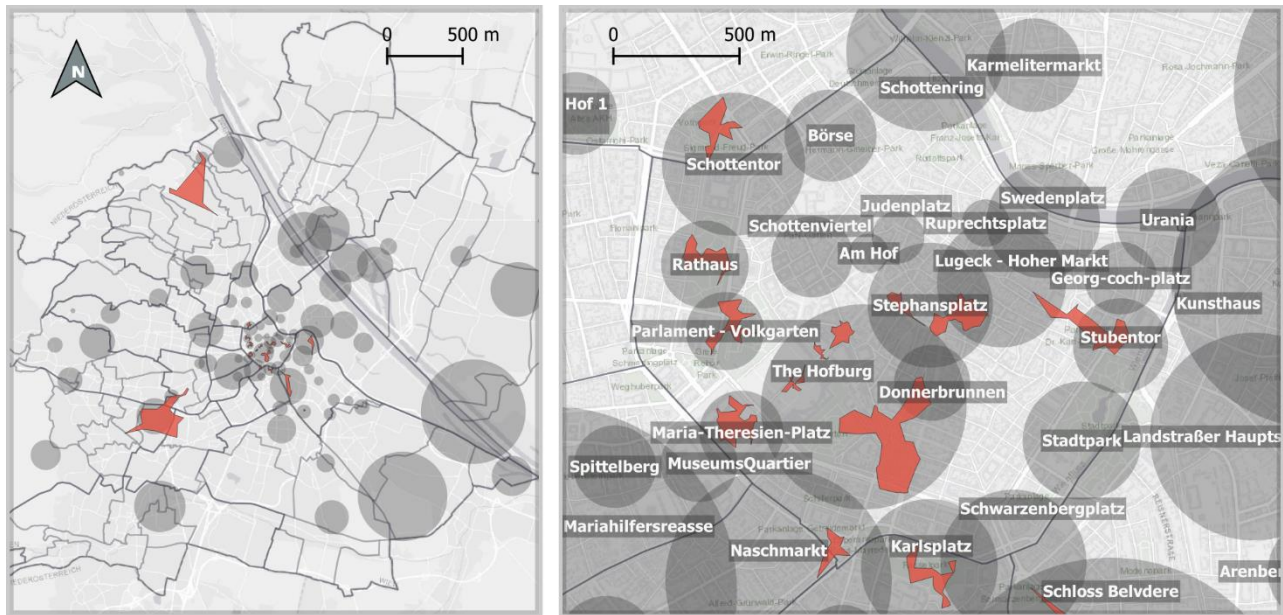


Figure 31: AOIs of Russian tourists in Vienna with $minPts$ equals to 1% of this group number of points.

5. In a lot of cases AOIs in the outskirts of the city may not be detected in classes with small sample size when the used minimum visits rate to extract AOIs is high in those classes (0.75%, 1% etc.) Figure 31.
6. Even by giving a small normalized value (0.05%, 0.1% etc.) for *minPts*, HDBSCAN algorithm is not able to identify AOIs in the outskirts of the city in data classes with small sample size (See Figure 32 in comparison with Figure 28). In addition, the algorithm was not able to identify AOIs by giving a small normalized of 0.05% for *minPts* in data classes with very small sample size (See Table 8).

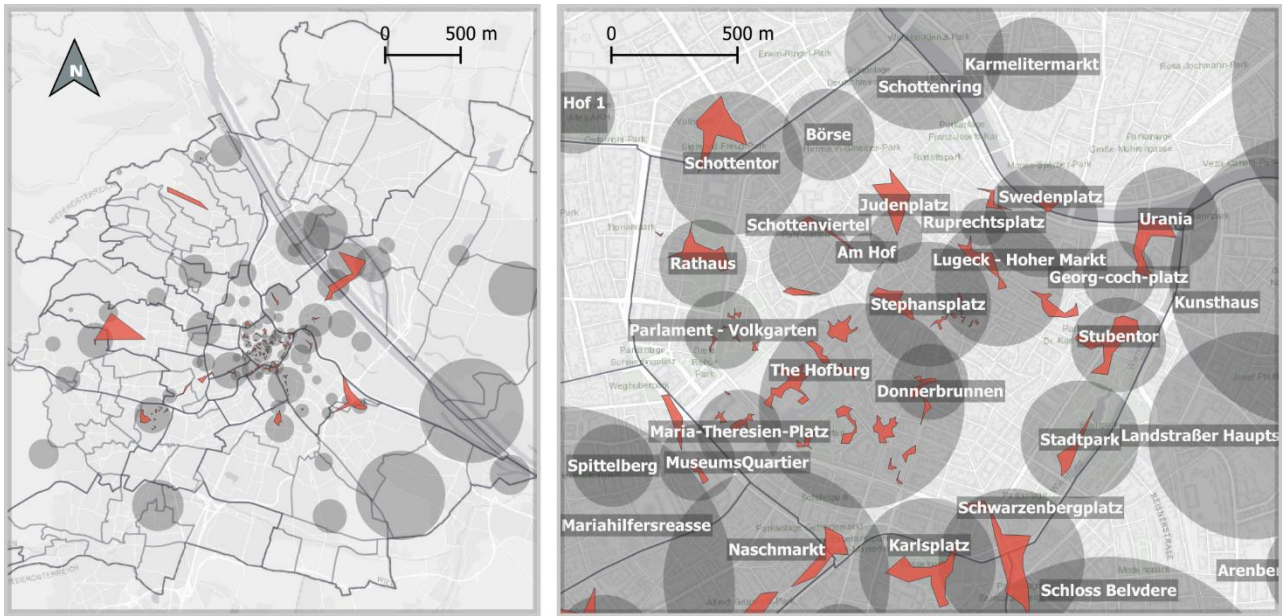


Figure 32: AOIs of Russian tourists in Vienna with *minPts* equals to 0.1% of this group number of points

7. The AOIs shapes of data classes with big sample size (See Figures 28, 29 and 30) are smoother than the AOIs shapes of data classes with small sample size (See Figures 31, 32) even though using the same λ_P in all figures.

Considering all these conclusions, the best way to identify a suitable *minPts* value is to treat each class of points individually based on its sample size. For classes with large sample sizes such as the total visits of all tourists, AOIs are best identified when the assigned value of *minPts* is .25% of the total sample size (191 309). However, it is better to assign a value of 1% of the total sample size to *minPts* if the class sample size is small, e.g. the group of total visits of Russian tourists. Accordingly, by examining all the generated clusters of all the possibilities in Table 8, *minPts* threshold levels have been defined based on the total points count of every group.

Table 9: Proposed *minPts* threshold levels.

| Points count | 200 000 – 75 000 | 75 000 – 30 000 | 30 000 – 10 000 | 10 000 – 4 000 | 4 000 – 1 500 |
|----------------------|------------------|-----------------|-----------------|----------------|---------------|
| <i>minPts</i> = | | | | | |
| % of points count | .25% | .4% | .5% | .75% | 1% |

5.1.2 Finding a suitable normalized length parameter λP

As mentioned in Chapter 4, representing AOIs with points on maps may not be an optimum visualization approach in this case study. Thus, converting the point clusters into polygons became necessary. In the same chapter, a Concave hull algorithm was proposed to compute bounding polygons for the generated point clusters. The computed polygons should have low complexity and low emptiness (i.e. smooth polygons). In order to compute smooth polygons, an optimum normalized length parameter λP should be identified. The following function was proposed for that reason.

$$\Phi(P, D) = Emptiness(P, D) + C * Complexity(P)$$

This method was implemented in a tool called Balanced Concave Hull by Hu *et al.* The tool iterates 100 times through every point cluster and during every iteration it assumes a different λP value from 1 to 100 for concave hull computation. Then it applies this equation on every computed polygon in every iteration to calculate the average fitness of the λP value during this iteration. The value of λP with the lowest average fitness is the optimum λP (Hu et al., 2015). A test of the tool was made on 16 groups of the case study dataset and the result was as follows in Figure 32. The lowest (optimum) value of λP is 28 where the average fitness value is 0.392233368. The tool used this λP to compute the optimum concave hull for every point cluster.

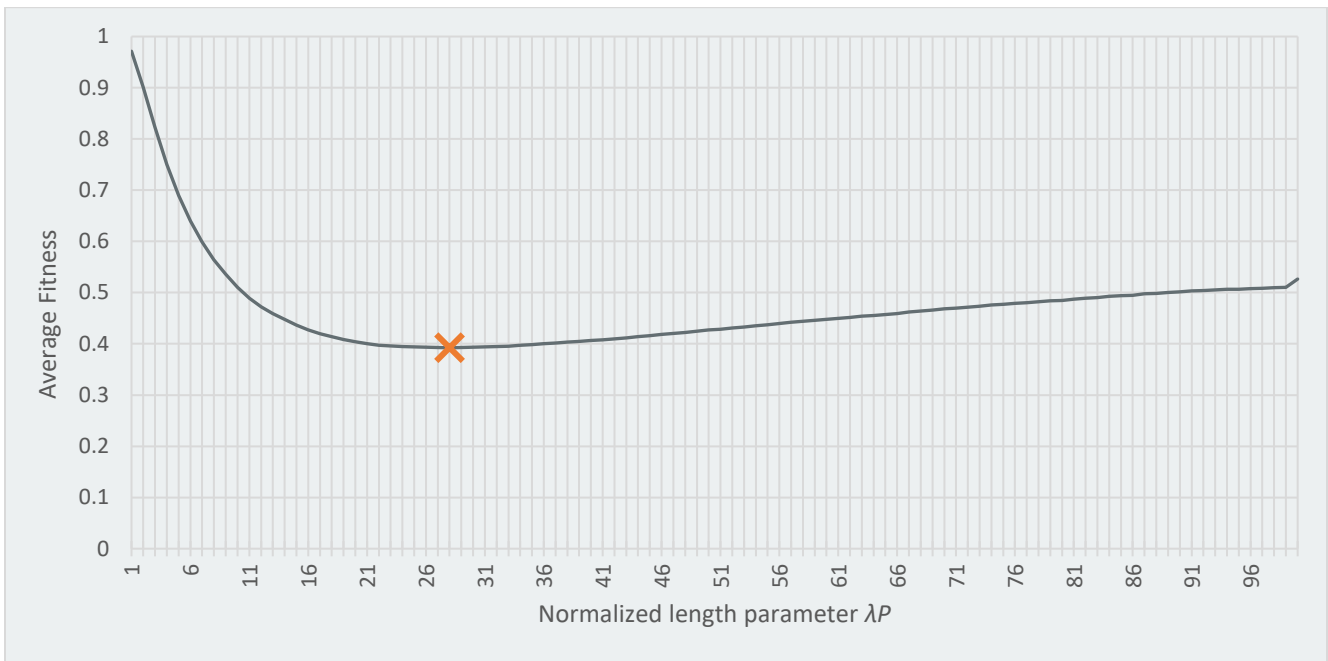


Figure 33: A line chart shows the average fitness values of different λP .



Figure 34: AOIs of all tourists in Vienna

The previous figure shows the AOIs of all tourists in Vienna with $minPts$ equals to 0.25% of this group number of points. This means that AOIs with a visit rate of more than 0.25% of all tourists will be visible

on the map. The most visited places in Vienna are Schönbrunn, The Hofburg and Stephansplatz, all with visit rates more than 3%. Other places such as the Spittelau incinerator and Heeresgeschichtliches Museum have visit rates of less than 0.5%.

5.2 Visualizing footprints of different tourists' nationalities in Vienna

5.2.1 Defining a suitable search radius (bandwidth) h

As mentioned in the previous chapter, if large bandwidth h has been used, a lot of points will fall within its boundary. In return, the output heat map will be more generalized and smoother (Mikelbank, 2001). In order to keep a balance between smoothness and edginess, a suitable h value should be identified. The following equation can be used for that purpose.

$$h_{opt} = \left[\frac{2}{3n} \right]^{1/4} \sigma$$

Table 10: h_{opt} for different classes.

| Country | No. Points | Standard Distance by m | h_{opt} by m |
|-------------------------------------|------------|------------------------|----------------|
| All tourists | 191309 | 3510.032769 | 151.6542221 |
| Foreign tourists | 86962 | 2711.963808 | 142.7015515 |
| Locals | 81626 | 4336.941731 | 231.8481547 |
| German tourists | 12197 | 3064.009508 | 263.4535396 |
| American tourists | 11949 | 2469.358374 | 213.4167087 |
| British tourists | 10625 | 2436.634164 | 216.8629093 |
| Italian tourists | 7786 | 2465.20543 | 237.1382994 |
| Spanish tourists | 5036 | 2202.643944 | 236.2653611 |
| French tourists | 4467 | 2503.197347 | 276.6739605 |
| All tourists 2008 | 10801 | 3453.268496 | 306.0846428 |
| All tourists 2013 | 19464 | 3530.531579 | 270.0904706 |
| All tourists 2018 | 14412 | 3672.30157 | 302.8547993 |
| All tourists Winter | 34470 | 3469.13889 | 230.058642 |
| All tourists Summer | 60528 | 3422.73923 | 197.1795708 |
| All tourists Spring | 50099 | 3584.585398 | 216.5004744 |
| All tourists Autumn | 46212 | 3567.797343 | 219.8814903 |
| Average h_{opt} | | | 232.0415498 |

The value of h_{opt} of every class of points is different. However, that will give misleading results when comparing between different footprints. The solution for that is to calculate the h_{opt} average of the whole groups that I'm working on.

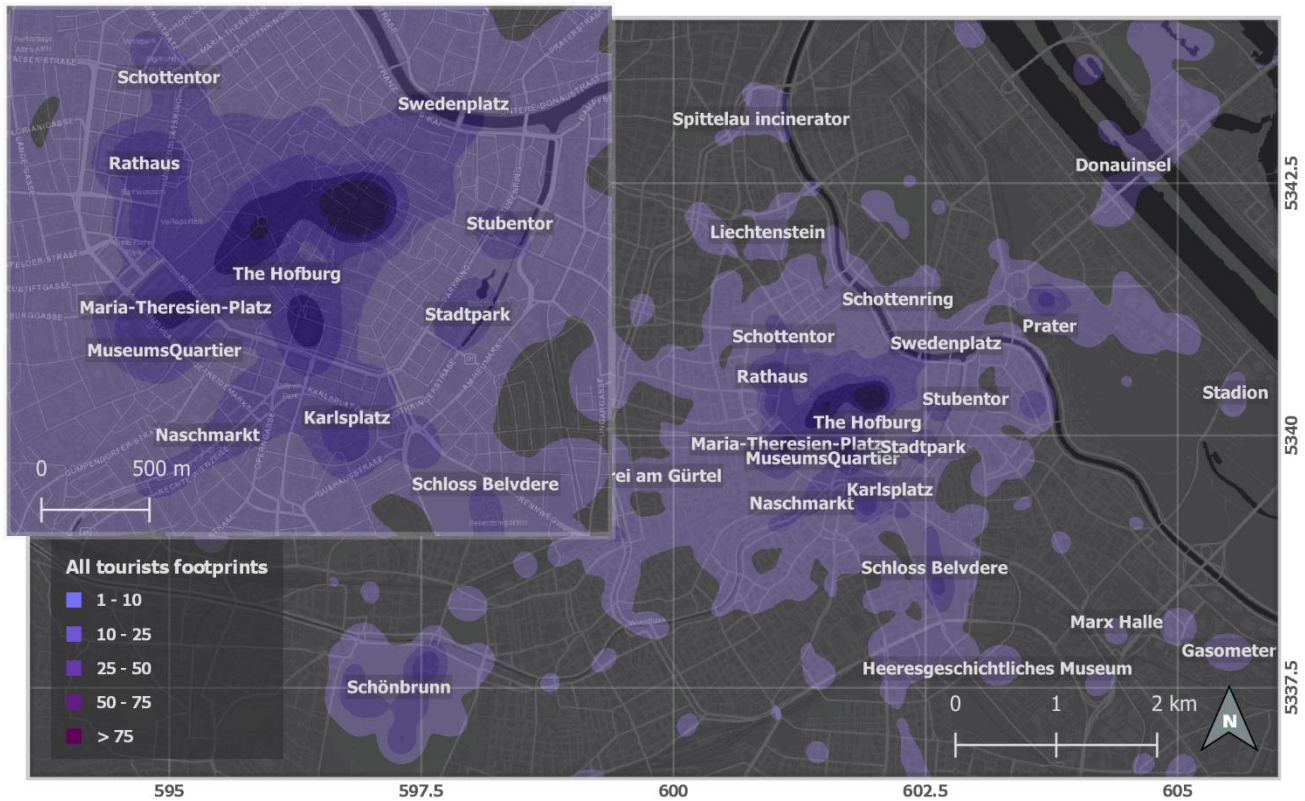


Figure 35: Footprints of all tourists in Vienna

The value of h_{opt} of every class of points is different. However, that will give misleading results when comparing between different footprints. The solution for that is to calculate the h_{opt} average of the whole groups that I'm working on.

Even with applying the same value of h_{opt} to other classes with smaller sample size like the class of the Italian tourists' visits in Vienna, the results are still reliable since the footprints of this class are still visible even in the outskirts of the city (See Figure 36).



Figure 36: Footprints of Italian tourists in Vienna

5.3 Technologies

All the technologies used in this research are open source tools. QGIS is the geographic information system application used for viewing, editing, and analysis of the geospatial data of the case study. The application provides a built in Python support called PyQGIS for data processing. That opened the door for the research to benefit from other Python libraries for processing the data. For Example:

- scikit-learn: is an open source machine learning Python library. It was an essential tool for the data analysis in the research.
- HDBSCAN: is a scikit-learn compatible project. It was used to identify the clusters.

5.3.1 Map Algebra

Map Algebra is a powerful tool for raster data analysis. It iterates over all the cells in a raster and performs operations and functions over the cells. For instance, two raster images can be subtracted from each other using the subtraction arithmetic operation.

| | | |
|---|---|---|
| 5 | 5 | 8 |
| 1 | 5 | 9 |
| 3 | 7 | 5 |

 $-$

| | | |
|---|---|---|
| 8 | 2 | 6 |
| 4 | 0 | 2 |
| 8 | 4 | 6 |

 $=$

| | | |
|---|----|----|
| 3 | -3 | -2 |
| 3 | -5 | -7 |
| 5 | -3 | 1 |

Figure 37: Example for subtraction operation in Map Algebra

The tool is implemented in QGIS and widely used in a lot of GIS applications and studies. The tool was very useful in this study for comparing the footprints of two different groups of data. For instance, the subtraction operation was used to subtract the footprints of the American tourists from the footprints of all foreign tourists. It was found out that, Americans are more interested in Ruprechtsplatz, Heeresgeschichtliches Museum and Victor-Braun-Platz than the other foreign tourists.



Figure 38: The footprints difference between The American tourists and all foreign tourists in Vienna

6 Outputs

6.1 Spatial differences of interest patterns

6.1.1 Locals and foreigners

According to the results, the most visited place by both foreigners and locals in Vienna is Schönbrunn. However, the percentage of Austrians who are interested visiting Schönbrunn is 3.7% out of the total number of Austrian tourists. Meanwhile, 9.6% of foreigners are interested in visiting Schönbrunn (see Figures 39, 40 and 41). In addition, it can be derived from Figure 42 that Austrians are more interested in the Schönbrunn zoo than the gardens and the palace.

In general, the locals' interest in the touristic places is less than the foreigners. However, the locals' footprints are widely spread over a very big area of Vienna (See Figure 42), in contrast to the foreigners who are more concentrated in specific AOIs. The reason for that is that other people from different nationalities who take and share photos of places in Vienna are highly likely to be tourists. They came from different countries and target specific AOIs to visit in a short time span. On the other hand, an Austrian may or may not be a tourist. They could simply be Wieners who spotted something interesting on their way to work and decided to share it with their friends on Flickr. This kind of data is considered to be noise in our case study because a lot of this data might be out of the context of tourism.

KDE as an aggregation method is not able to detect and get rid of this noise data. On the other hand, HDBSCAN managed to successfully filter out the noise data. The best clue for that is the visit density of locals which is extracted by KDE seems to be higher in most of the areas in Figure 42 than the visit density by tourists. For example the interest of Austrians to visit Rathaus is higher than the interest of foreigners in Figure 40. Whereas, it is the opposite if the Rathaus AOIs was compared by HDBSCAN in figures 39 and 40.

However, the HDBSCAN results might still be biased to some degree in terms of interest rate. For example, the standard deviation of Austrians interest AOIs rates is 0.86%. On the other hand it is 3% for foreign tourists. That means the Austrians have neutral interest in almost all the 55 AOIs which they are interested in. Whereas, foreign tourists have contrast in their interests. The reason for that might be due to the noise data in the Austrians data-set which may affect the extracted interest rate by HDBSCAN. That indicates that there might be a correlation between data noise of this kind and the spatial density of the points.



Figure 39: AOIs of foreign tourists in Vienna



Figure 40: AOIs of locals in Vienna

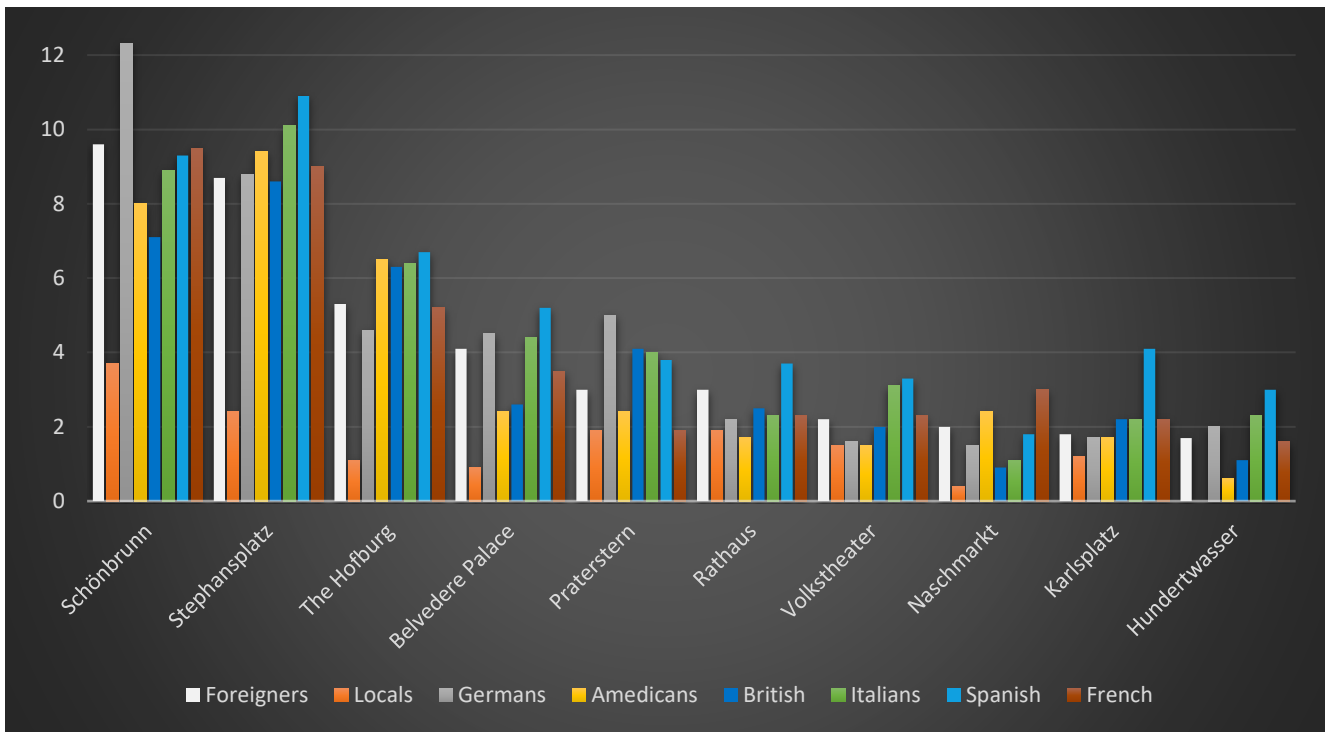


Figure 41: A bar chart showing the 10 most visited AOIs in Vienna and visits rates of different tourists groups

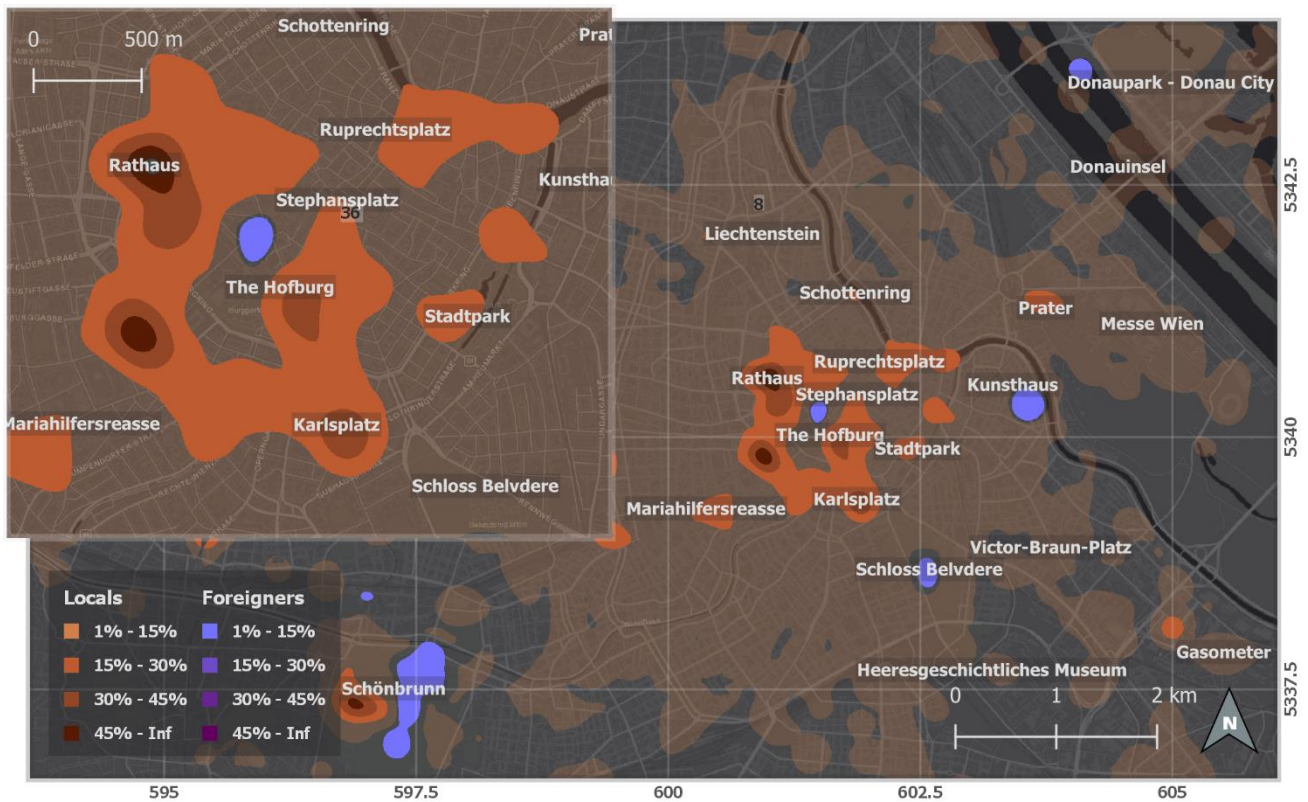


Figure 42: The footprints difference between the locals and all foreign tourists in Vienna.

6.1.2 Different tourists' nationalities.

With a 12.3% visit rate to the Schönbrunn, the German tourists have the highest interest in this AOI. And similarly, with 2% higher than the visit rate of foreign tourists, the German tourists have the highest visit rate to Prater (see Figure 41 and 43). That also can be detected in the footprints difference between the German tourists and all foreign tourists in Vienna (see Figure 44). The same figure also shows that the visit intensity of Hauptbahnhof and Schottenring is from 1% to 5% higher than the visit intensity of all foreign tourists combined.

The interest pattern of British tourist is neutral. However, they are the most interested in Maria-Theresien-Platz with a visit rate of more than 3% of the total British tourists. On the other hand, they are less interested in the Graben, a street leading to St. Stephen's Cathedral (see Figure 45). Also they are less interested in visiting the Albertina with visit rate of 1.3%, compared to the German tourists who have interest rate of 4.1%.

In Figure 46, it is very clear that the visit intensity of British tourists to Maria-Theresien-Platz is very high with an intensity rate 15% higher than the general intensity rate of foreign tourists of the same place. They also have a higher visit intensity than foreign tourists in Landstraße. In addition, they are less interested in Kunsthaus (Hundertwasserhaus) with an intensity rate from 5% to 10% less than foreign tourists (see Figure 46).

Spanish tourists have very interesting pattern of interest. They have the highest visit rate in 7 of the top 10 AOIs in Vienna which are Stephansplatz, The Hofburg, Belvedere Palace, Rathaus, Volkstheater, Karlsplatz and Hundertwasserhaus. All of these places have visit rate 3.3% to 10.9% of the total Spanish tourists in Vienna (see Figure 41 and 47). However they have a high contrast in their visits rate in different AOIs with standard deviation of 3.1.

The difference in visit intensity is also very notable between the Spanish tourists' footprints and the foreign tourists' footprints. The highest difference can be spotted in both Karlsplatz and Kunsthaus (Hundertwasserhaus) with an intensity difference higher than 15% in favor of the Spanish tourists. Some other places with high intensity differences can be spotted as well such as Volkstheater, Rathaus, Belvedere Palace, Albertina and Stadtpark (see Figure 48).



Figure 43: AOIs of German tourists in Vienna

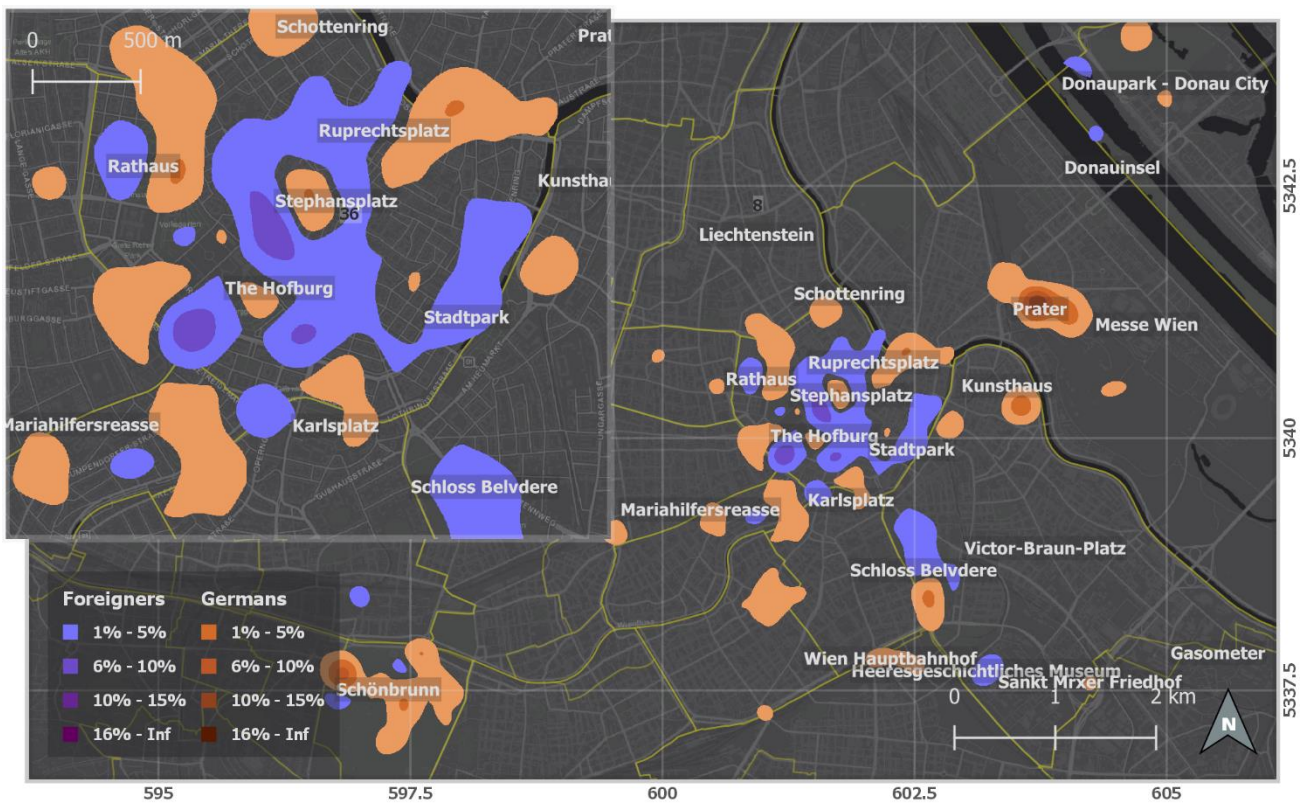


Figure 44: The footprints difference between the German tourists and all foreign tourists in Vienna



Figure 45: AOIs of British tourists in Vienna

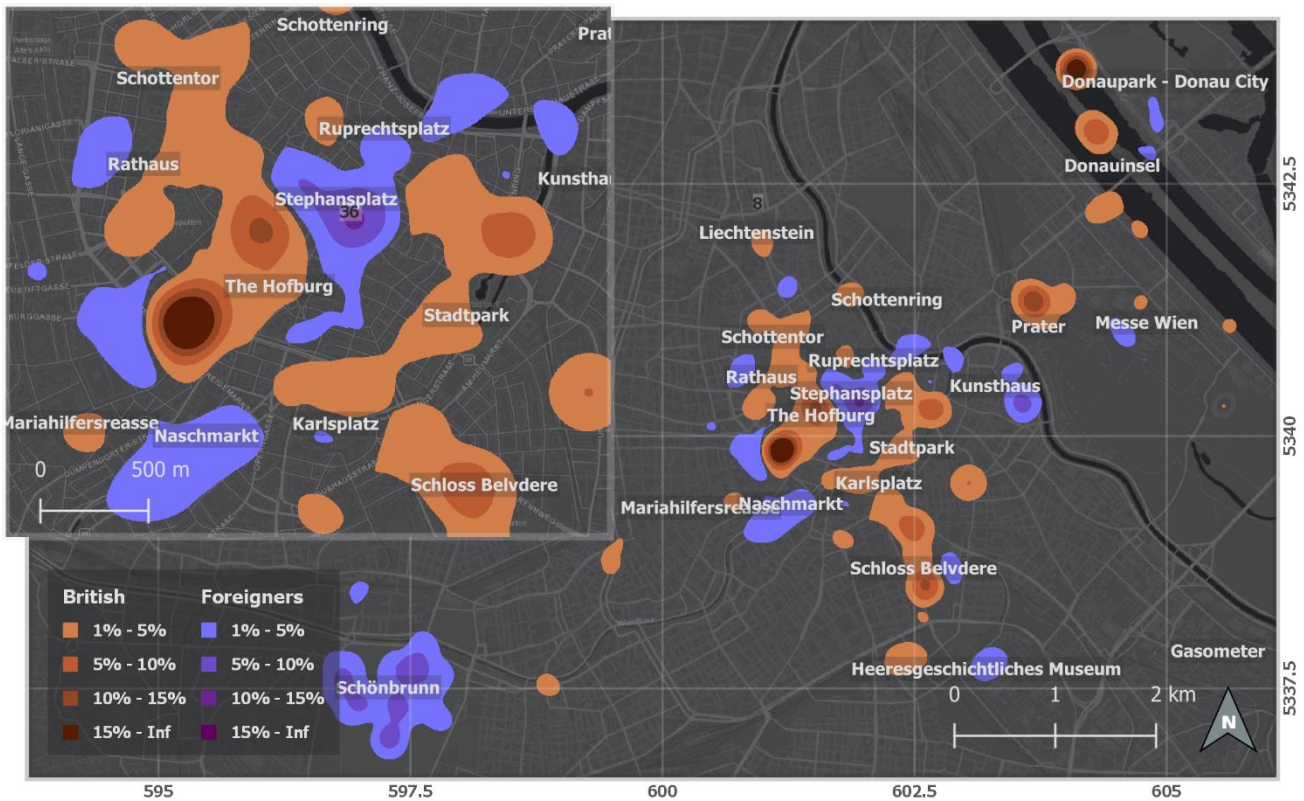


Figure 46: The difference in footprints between the British tourists and all foreign tourists in Vienna

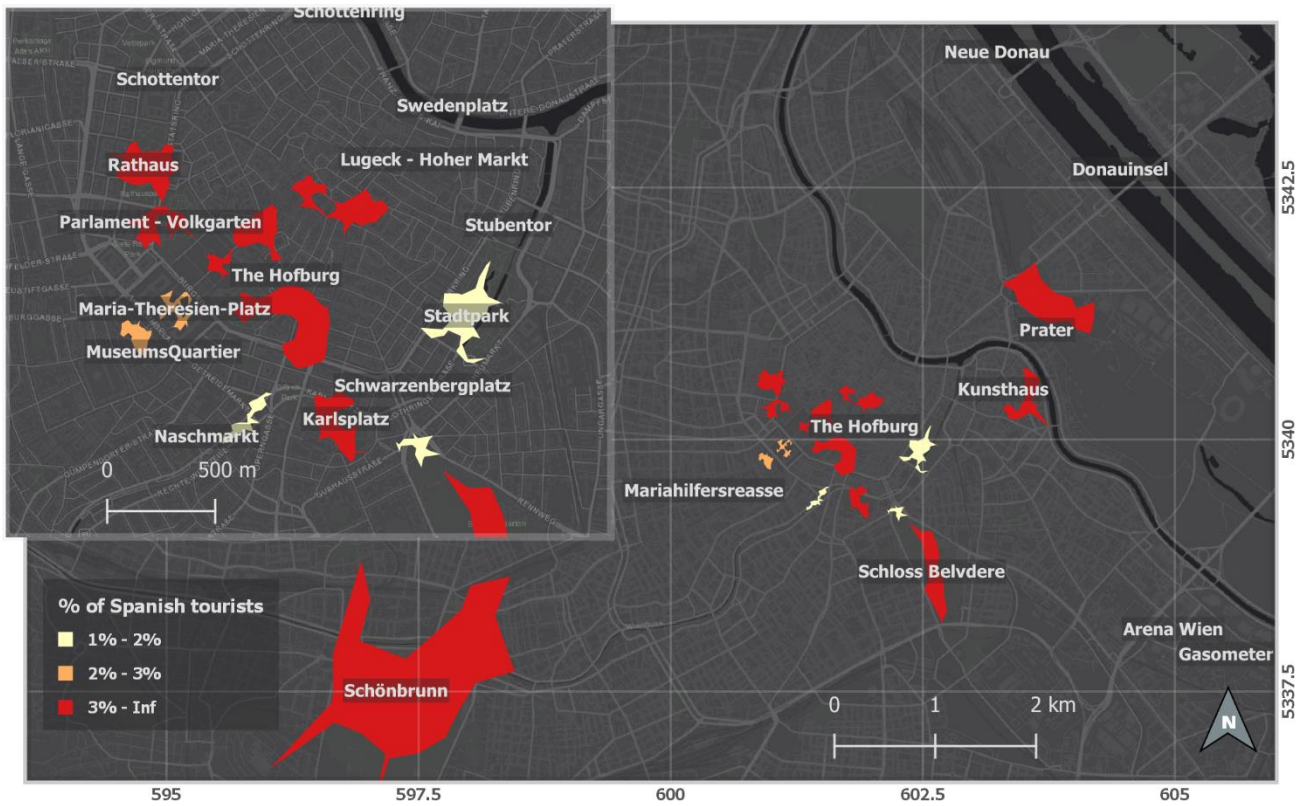


Figure 47: AOIs of Spanish Tourists in Vienna

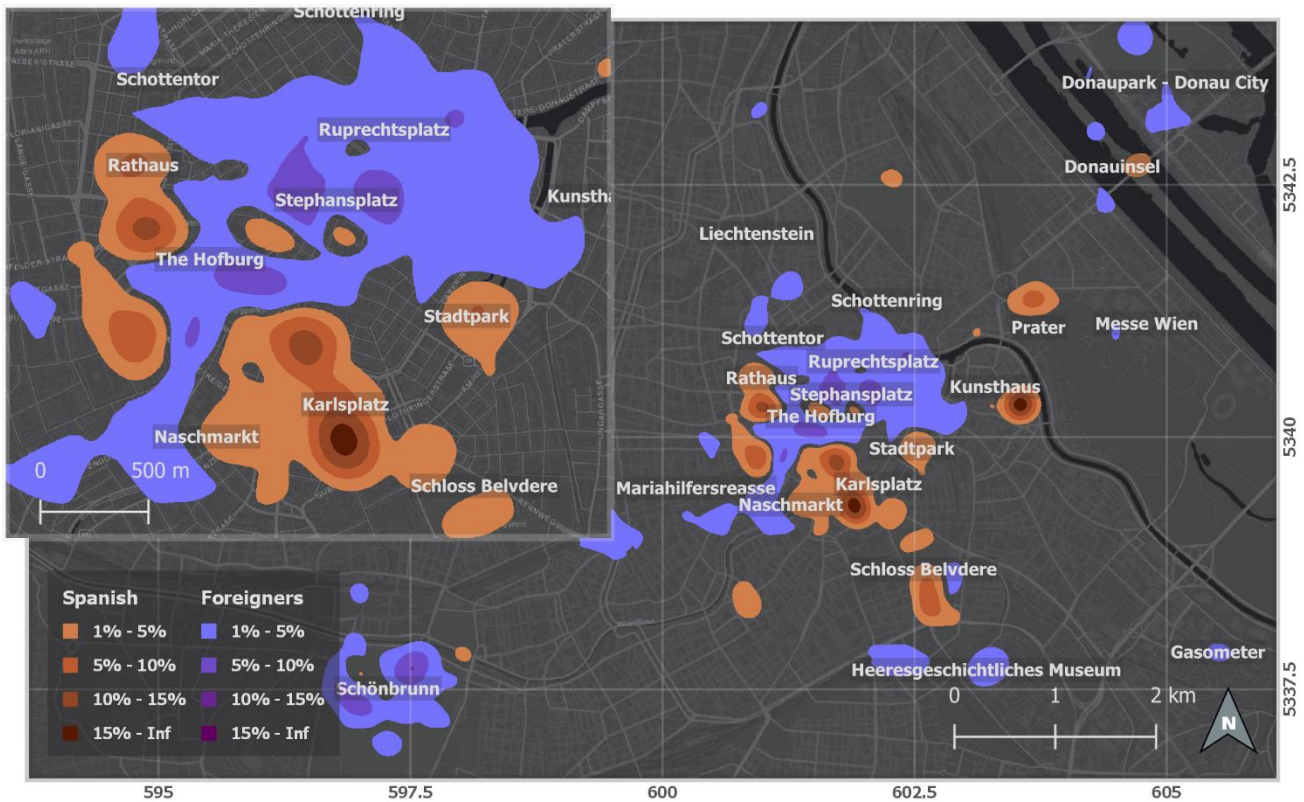


Figure 48: The difference in footprints between the British tourists and all foreign tourists in Vienna

6.2 Spatio-temporal differences of interest patterns

6.2.1 Annual basis

Spatio-temporal analysis to detect the difference of interest patterns on annual basis has been conducted on a time period of 10 years from 2008 to 2018. The analysis has yielded the following:

- The visit rates of some AOIs has increased over the years such as Albertina, Maria-Theresien-Platz and Donauinsel (see figure 49). However, the increase in visit rate for the Donauinsel was only in 2018. Meanwhile it was constant from 2008 to 2013. The reason for that could be the hosting of Vienna Major in Donauinsel, 2018 (see Table 2).
- The visit rates of some AOIs has declined over the years such as Rathaus, The Hofburg and Stephansplatz (see figure 49).
- The Schönbrunn and Belvedere Palace gained more visits in 2013 than usual.
- In 2018, it is very notable that visit intensity of the Parliament has dropped since 2013. The reason for that is likely the long-term construction work on the Parliament building.

Comparing between different AOI classes might be misleading visualization methodology if the comparison includes more than 2 classes to compare. On the other hand, heat maps provide suitable representations of these differences in one map. The reason is that the footprints of different classes can be visualized in one map. However, AOIs of different classes have to be visualized in different maps.

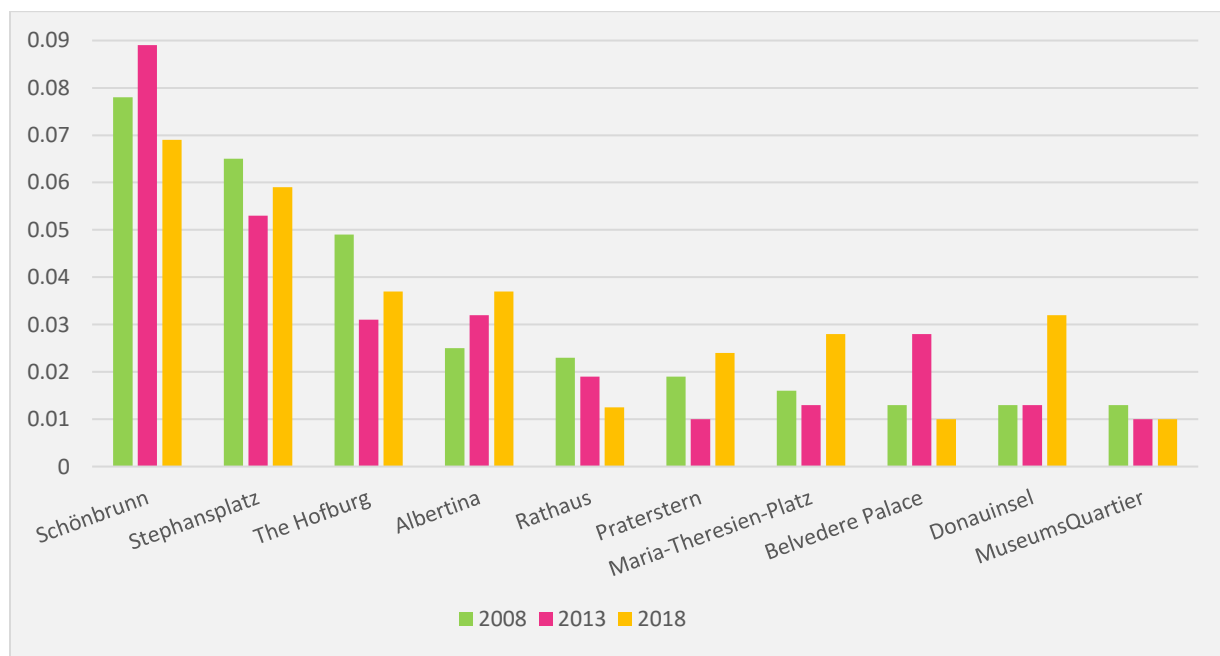


Figure 49: A bar chart shows 10 of the most visited AOIs in Vienna and visits rates on annual basis.



Figure 50: The difference in tourist footprints between 2008 and 2013 in Vienna



Figure 51: The difference in tourist footprints between 2013 and 2018 in Vienna

6.2.2 Seasonal basis

Spatio-temporal analysis to detect the difference of interest patterns on seasonal basis has been conducted. The analysis has yielded the following:

- The visit rates of open air AOIs notably increase during the summer such as Schönbrunn, Stephansplatz and Belvedere. However in winter, the visits rates increase in the Rathaus, possibly for the Christmas market, and Maria-Theresien-Platz where two famous museums (Museum of Natural History and Kunsthistorisches Museum) are located and another large Christmas market (see the following figures).
- The visit rate to Prater is notably increases during autumn. Possibly for hosting some festivals like the Halloween.
- During spring, the visit rates are neutral in all AOIs. However, it is very low in the Albertina.

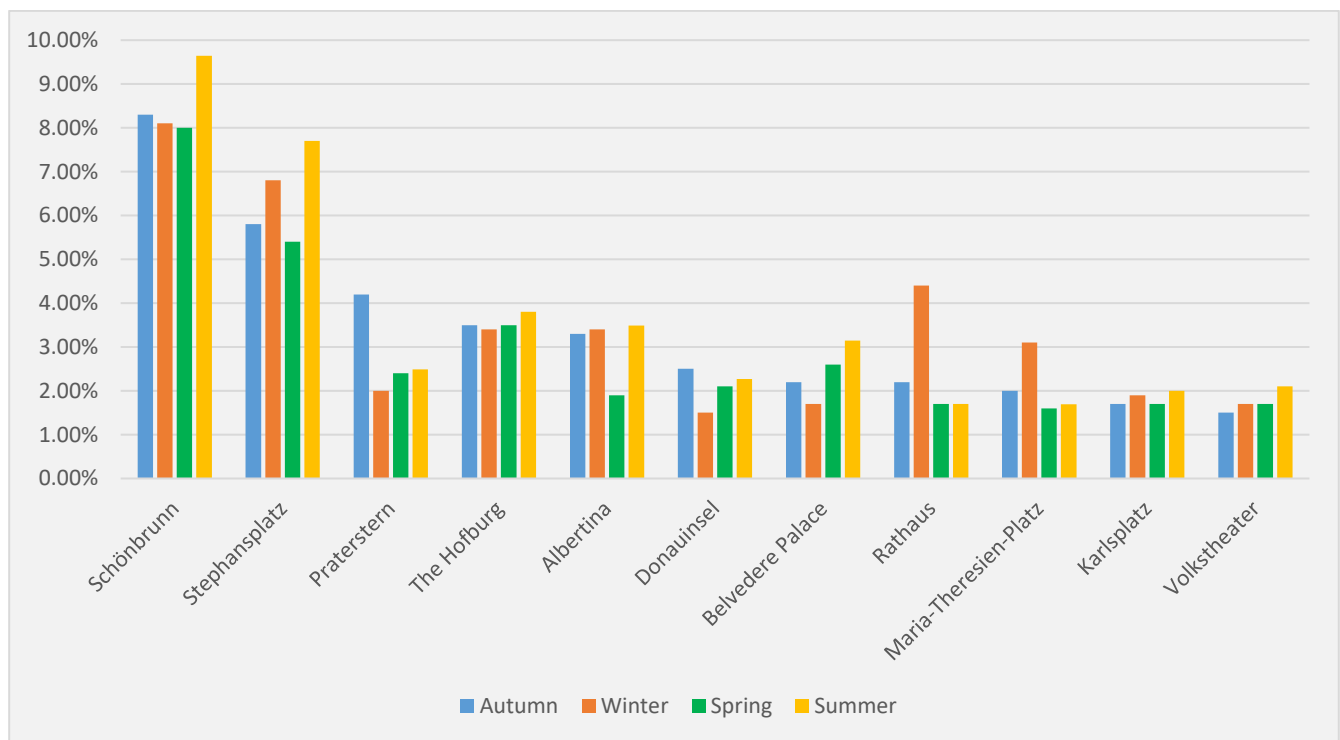


Figure 52: A bar chart showing 10 of the most visited AOIs in Vienna and visits rates on seasonal basis.



Figure 53: AOIs of tourists in Vienna during summer



Figure 54: AOIs of tourists in Vienna during winter

Comparing the footprints of the tourists in winter and summer in Vienna has yielded the following:

The contrast between the visit intensities of winter and summer is very high. A high contrast can be spotted very easily in Schönbrunn, Volksgarten and Schloss Belvedere in favor of summer with 15% intensity value higher than winter. On the other hand, the Maria-Theresien-Platz and the Rathaus witness a remarkable increase in the visits intensity during winter.

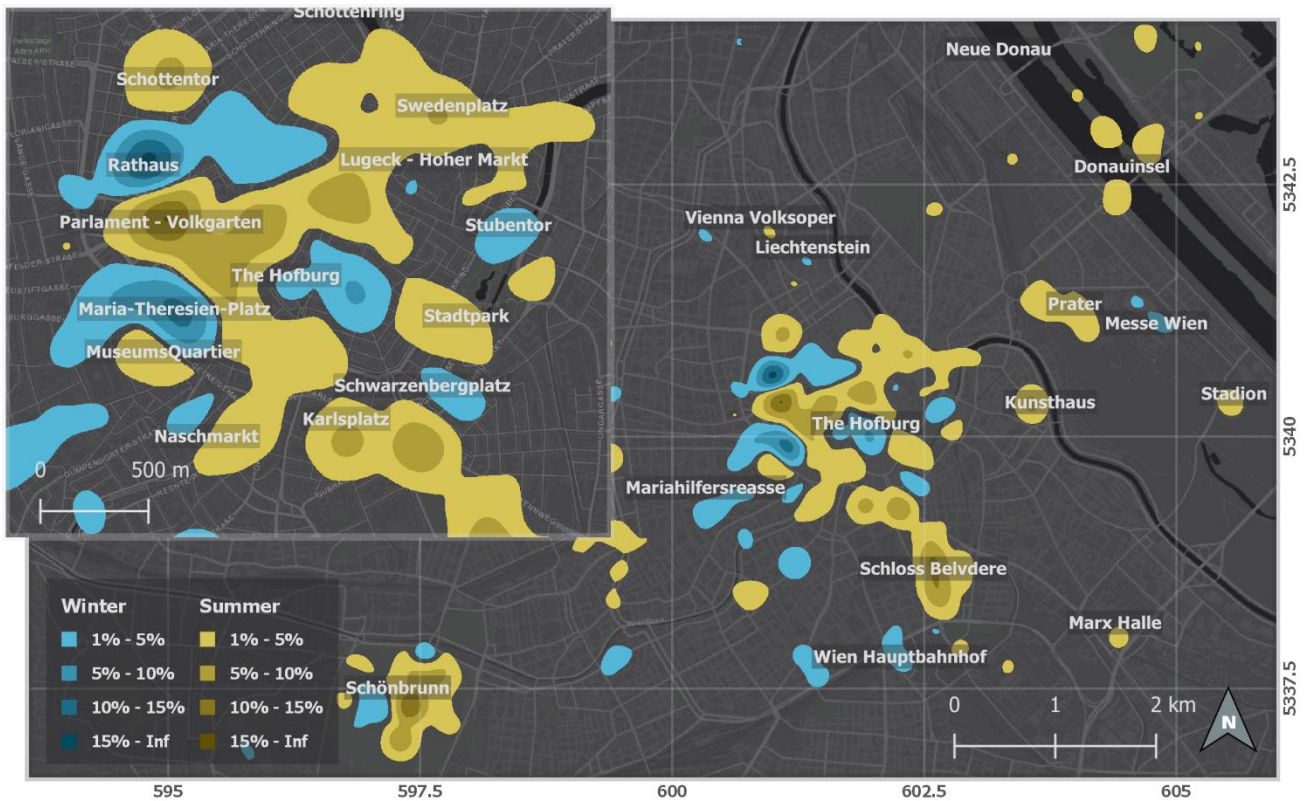


Figure 55: The difference in tourists' footprints between winter and summer in Vienna

7 Conclusion

In this research, spatial patterns of interests have been extracted and visualized in Vienna in context of tourism from geosocial media GSM. Different patterns of different groups of tourists have been detected in the same space. In addition, the methodologies in this research were capable to help in understanding the evolution of the spatial interest patterns by detecting different interest patterns in the same space in seasonal and annual basis. KDE as an aggregation method is not able to detect and get rid of noise data. On the other hand, HDBSCAN managed to successfully filter out the noise data. Representing interest patterns as footprints is a more suitable representation methodology if the comparison includes more than 2 classes to compare.

By giving a small normalized value (0.05%, 0.1% etc.) for *minPts*, HDBSCAN algorithm is able to identify AOIs in the outskirts of the city. However, data with big sample size has to be used in that case. On the other hand, AOIs in the outskirts of the city might not be detected using HDBSCAN in data classes with small sample size even by giving a small normalized value for *minPts*. KDE as an aggregation method was able to detect footprints in the outskirts of the city even in classes with small sample size, this is due to the ability of KDE to consider all the sample points in computing the footprints.

Table 11: Comparison between KDE and HDBSCAN as different methods to visualize spatial patterns of interests

| | HDBSCAN | KDE |
|--|--|---|
| Data noise | Able to filter out the noise data and exclude it from the final clusters. | Not able to detect and get rid of noise data. |
| Comparing results | Every class has to be visualized in separate map. | Comparing the results of more than 2 classes in one map is possible |
| Visualizing interest patterns using small sample size | AOIs in the outskirts of the city might not be detected. The AOIs polygons are smoother in bigger sample sizes. | Was able to detect footprints in the outskirts of the city. |

Overall, the interest patterns of this research were very accurate in comparison with the reality. This is due to the integration between both methodologies and the suitability to use georeferenced Flickr data in context of tourism. However, Flickr data needs to go through an aggregation process first (Verstockt et al., 2019) in order to prepare the data for AOIs visit rate computation.

7.1 Possible future work

GSM data gives the opportunity for researchers to go deeper in spatial and temporal analysis of people's interests. The associated text of the georeferenced posts in GSM (title, descriptions, and tags) can be a very good data source for sentiment analysis techniques (Huang, Gartner, & Turdean, 2013). The use of Natural Language Processing (NLP) is essential to perform sentiment analysis to derive the valence of an area (Hobel, Fogliaroni, & Frank, 2016). Since valence is described as the pleasure or the umbrage of people's experiences, the valence of an area can be classified as positive or negative (Hauthal, 2015). As valence of an area is subjective from a group of people to another and from person to another, this valence can be analyzed and associated to the AOIs or footprints modelled from social media.

Determination of nationality, age, gender of the users would indeed help to visualize different classes of AOIs and footprints. In return, this will help to spatially analyze and compare the spatial distribution of densities between different tourists to understand their interest patterns based on their ages and genders. Thus, developing user profiling determination methods for data grouping would be a good idea.

Moreover, visualizing data from different geosocial media platforms provides the space to compare between these GSM. This helps to find out the level of suitability of these sources for spatial analysis in general and to recognize spatial patterns of interests in specific.

8 References

- Affiliate Members Global Report, Volume 12 – Cultural Routes and Itineraries.* (2015).
<https://doi.org/10.18111/9789284417704>
- Agostinelli, M. (2017). *Density-Based Clustering Heuristics for the Traveling Salesman Problem.* University of Rhode Island.
- Akbar, S. I., & Shahid-ul-Islam. (2018). Minimum spanning tree algorithms and techniques. *IJCIRAS, 1(8)*, 2581–5334.
- Akdag, F., Eick, C. F., & Chen, G. (2014). Creating polygon models for spatial clusters. *Foundations of Intelligent Systems, 8502.*
- Antalovsky, E., & Löw, J. (2018). *Why Vienna Gets High Marks.* Retrieved from https://www.urbaninnovation.at/tools/uploads/citytransformed_WhyViennagetshighmarks.pdf
- Antarctic and Southern Ocean Coalition ASOC. (2011). *Evolution of Footprint: Spatial and Temporal Dimensions of Human Activities.* Presented at the XXXIV ATCM Buenos Aires, Argentina.
- Antarctic Environmental Officers Network AEON. (2005). *Practical guidelines for developing and designing environmental monitoring programmes in antarctica.*
- Ariffin, I., Solemon, B., Anwar, R. M., Din, M. M., & Azmi, N. N. (2014). Exploring the potentials of volunteered geographic Information as a source for spatial data acquisition. *IOP Conference Series: Earth and Environmental Science, 20*, 012041.
- Arnold, N. D., Jenny, B., & White, D. (2017). Automation and evaluation of graduated dot maps. *International Journal of Geographical Information Science, 31(12)*, 2524–2542.
- Austrian Economic Chamber WKO. (2018, September). *This is Austria-facts & figures.* Austrian Federal Economic Chamber - Economic Policy Department.
- Bailey, T. C., & Gatrell, A. C. (1995). Interactive spatial data analysis. *Longman Scientific & Technical; J. Wiley, Harlow Essex, England.*
- Brauner, R. (2017). *Vienna in Figures 2017.* Vienna City Administration.

- Brinkhoff, T., Kriegel, H.-P., Schneider, R., & Braun, A. (1995). *Measuring the complexity of polygonal objects*.
- Buehler, R., Pucher, J., & Altshuler, A. (2017). Vienna's path to sustainable transport. *International Journal of Sustainable Transportation*, 11(4), 257–271.
- Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Advances in Knowledge Discovery and Data Mining*, 7819.
- Cao, L., Luo, J., Gallagher, A., Jin, X., Han, J., & Huang, T. S. (2010). *A worldwide tourism recommendation system based on geotagged web photos*. Presented at the IEEE International Conference on Acoustics, Speech and Signal Processing.
- Chainey, S. (2013). Examining the influence of cell size and bandwidth size on kernel density estimation crime hotspot maps for predicting spatial patterns of crime. *BSGLg*, 60, 7–19.
- Citybike Wien. (n.d.). 15 Jahre Citybike Wien.
- Crime in Vienna. (n.d.). Retrieved December 8, 2019, from <https://www.numbeo.com/crime/in/Vienna>
- Cristureanu, C., & Bobirca, A. (2007). Airports driving economic and tourism development. *Romanian Economic Journal*, 10, 31–44.
- Csapo, J. (2012). The Role and Importance of Cultural Tourism in Modern Tourism Industry. In M. Kasimoglu & H. Aydin (Eds.), *Strategies for Tourism Industry*. <https://doi.org/10.5772/38693>
- De Freitas, C., & Higham, J. (2005). *The climate–tourism relationship and its relevance to climate change impact assessment*.
- Dempsey, C. (2014). What is the difference between a heat map and a hot spot map?
- Duckham, M., Kulik, L., Worboys, M., & Galton, A. (2008). Efficient generation of simple polygons for characterizing the shape of a set of points in the plane. *Pattern Recognition*, 41(10), 3224–3236.
- Esri. (2011). How kernel density works.
- Ester, M., Kriegel, H.-P., & Xu, X. (1996). *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. 6.

- European Market for Climate Services EU-MACS. (n.d.). How does climate impact tourism? Retrieved from <http://eu-macs.eu/outputs/tourism-faq/1-how-does-climate-impact-tourism>
- Flatow, D., Naaman, M., Xie, K., Volkovich, Y., & Kanza, Y. (2014). On the accuracy of hyper-local geotagging of social media content. *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*.
- Flughafen Wien AG. (2017). *Sustainability report 2017*. Retrieved from [/va/uploads/data-uploads/Konzern/Investor Relations/Nachhaltigkeitsbericht/NHB 2018_EN.pdf](#)
- Gerokostopoulos, A., Guo, H., & Pohl, E. (2015). Determining the right sample size for your test: Theory and application. *2015 Annual RELIABILITY and MAINTAINABILITY Symposium*.
- Grothe, C., & Schaab, J. (2009). Automated footprint generation from geotags with kernel density estimation and support vector machines. *Spatial Cognition & Computation*, 9(3), 195–211.
- Hauthal, E. (2015). *Detection, Modelling and Visualisation of Georeferenced Emotions from User-Generated Content*. Technische Universität Dresden.
- Hobel, H., Fogliaroni, P., & Frank, A. U. (2016). Deriving the Geographic Footprint of Cognitive Regions. In S. M. Sarjakoski T. (Ed.), *Geospatial Data in a Changing World. Lecture Notes in Geoinformation and Cartography*. Springer, Cham.
- Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., & Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, 54, 240–254.
- Huang, H., Gartner, G., & Turdean, T. (2013). Social media data as a source for studying people's perception and knowledge of environments. *Mitteilungen Der Österreichischen Geographischen Gesellschaft*, 155, 291–302.
- International Council on Monuments and Sites ICOMOS. (1997). *Charter for Cultural Tourism*.
- Kennedy, L., Naaman, M., Ahern, S., Nair, R., & Rattenbury, T. (2007). *How flickr helps us make sense of the world: Context and content in community-contributed media collections*. Presented at the MM '07 Proceedings of the 15th ACM international conference on Multimedia.

- Kordic, N., Živković, R., Stanković, J., & Gajic, J. (2015). *Safety and security as factors of tourist destination competitiveness*.
- Koylu, C., Zhao, C., & Shao, W. (2019). Deep neural networks and kernel density estimation for detecting human activity patterns from geo-tagged images: A case study of birdwatching on flickr. *ISPRS International Journal of Geo-Information*, 1(45).
- Lamit, H. (2004). Redefining landmarks. *Jurnal Alam Bina*, 6(1).
- Linna Li, & Michael F. Goodchild. (2012). *Constructing places from spatial footprints*. Presented at the Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, Redondo Beach, California.
- Mao, T. (2015). Mining One Hundred Million Creative Commons Flickr Images Dataset to Flickr Tourist Index. *International Journal of Future Computer and Communication*, 4(2), 104–107.
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, 2(11), 205.
- Michel, F. (2013). *Flickr explorer: What makes explored photos so special?*, Technical report.
- Michel, F. (n.d.). *Flickr-statistics* [Java]. Retrieved from <https://github.com/frmichel/flickr-statistics> (Original work published 2011)
- Mikelbank, B. A. (2001). Quantitative geography: Perspectives on spatial data analysis, by A. S. Fotheringham, C. Brunsdon, and M. Charlton. *Geographical Analysis*, 33(4), 370–372.
- Mooney, P. (2015). *Is volunteered geographic information big data?* Presented at the Proceedings of the 23rd GIS Resarch UK Conference (GISRUK 2015), Leeds, United Kingdom.
- Moreira, A., & Santos, M. Y. (n.d.). *Concave hull: A k-nearest neighbours approach for the computation of the region occupied by a set of points*. Presented at the GRAPP 2007 - International Conference on Computer Graphics Theory and Applications.
- Moulavi, D., Jaskowiak, P. A., Campello, R. J. G. B., Zimek, A., & Sander, J. (2014). *Density-based clustering validation*. Presented at the Proceedings of the 14th SIAM International Conference on Data Mining (SDM).

- Northrup, B., & Klaer, J. (2014). Effects of GDP on violent crime. *Georgia Institute of Technology*. Retrieved from <https://smartech.gatech.edu/handle/1853/51649>
- Oleksiak, P. (2014). Analysing and processing of geotagged social media. *Information Systems in Management*, 3(4), 250–260.
- Parker, C. J. (2014). *The fundamentals of human factors design for volunteered geographic information. 1 ed.* : Springer International Publishing.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Prettenthaler, F., Kortschak, D., & Ortman, P. (2016). The use of weather driven demand analysis in recreation site management. *Journal of Tourism & Hospitality*, 4, 188.
- Ryu, S. J. (2005). *Political Instability And Its Effects On Tourism*.
- Salomon, D. (2013). Moving on from Facebook: Using Instagram to connect with undergraduates and engage in teaching and learning. *College & Research Libraries News; Vol 74, No 8 (2013): September* DO - 10.5860/Crln.74.8.8991.
- Schubert, E., Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN. *ACM Trans. Database Syst.*, 42(3), 21.
- Shopova, I., Grigorova, Z., & Timareva, S. (2017). *The impact of culture and history on tourism: Case study of trakia tourist region*.
- Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Salles, J., & Loureiro, A. A. F. (2013). *A comparison of Foursquare and Instagram to the study of city dynamics and urban social behavior*. Presented at the UrbComp '13 Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing, Chicago, Illinois.
- Silverman, B. W. (1986). Density estimation for statistics and data analysis. *Monographs on Statistics and Applied Probability*, London: Chapman and Hall.

- Skoultzos, S., & Tsartas, P. (2009). Event tourism: Statements and questions about its impacts on rural areas. *Tourismos*, 4, 293–310.
- UNESCO World Heritage Centre. (n.d.). Historic Centre of Vienna. Retrieved December 8, 2019, from UNESCO World Heritage Centre website: <https://whc.unesco.org/en/list/1033/>
- USLegal. (n.d.). Recreational Areas and Facilities Law and Legal Definition. Retrieved December 8, 2019, from <https://definitions.uslegal.com/r/recreational-areas-and-facilities/>
- Verstockt, S., Milleville, K., Ali, D., Porras-Bernardez, F., Gartner, G., & Van de Weghe, N. (2019). *EURECA - European Region Enrichment in City Archives and collections*. Presented at the Proceedings 14th ICA Conference Digital Approaches to Cartographic Heritage.
- Vienna City Administration. (n.d.). Vienna Public Transport—Tickets, timetables and plans. Retrieved December 8, 2019, from <https://www.wien.gv.at/english/transportation-urbanplanning/public-transport/>
- Vienna Tourist Board. (n.d.-a). Schönbrunn Palace. Retrieved December 8, 2019, from VIENNA – Now. Forever website: <https://www.wien.info/en/sightseeing/sights/imperial/schoenbrunn-palace>
- Vienna Tourist Board. (n.d.-b). The online travel guide for Vienna. Retrieved December 8, 2019, from VIENNA – Now. Forever website: <https://www.wien.info/en>
- Vienna Tourist Board. (n.d.-c). The origin of the Ringstrasse. Retrieved December 8, 2019, from VIENNA – Now. Forever website: <https://www.wien.info/en/sightseeing/ringstrasse/construction-of-ringstrasse>
- Vienna Tourist Board. (n.d.-d). UNESCO’S World Cultural Heritage. Retrieved December 8, 2019, from VIENNA – Now. Forever website: <https://www.wien.info/en/sightseeing/sights/cultural-heritage>
- Vienna Tourist Board. (n.d.-e). Vienna’s Ringstrasse. Retrieved December 8, 2019, from VIENNA – Now. Forever website: <https://www.wien.info/en/sightseeing/architecture-design/ringstrasse>
- Wehdorn, M., Biermayer, U., Kreppenhof, A., Scheuchel, P., & Zunke, R. (Eds.), Szabó, S. (Trans.). (2009). *The Historic Centre of Vienna—World Cultural Heritage and Vibrant Hub*. Retrieved from <https://www.wien.gv.at/stadtentwicklung/studien/pdf/b008028.pdf>

- Więckowski, M., Michniak, D., Bednarek-Szczepanska, M., Chrenka, B., Ira, V., Tomasz, K., ...
Wiśniewski, R. (2014). Road accessibility to tourist destinations of the Polish-Slovak borderland:
2010-2030 prediction and planning. *Geographia Polonica*, 87, 5–26.
- Xie, Z., & Yan, J. (2008). Kernel Density Estimation of traffic accidents in a network space. *Computers,
Environment and Urban Systems*, 32(5), 396–406.
- Yadav, J. (2015). *Fuzzer test log analysis using machine learning (Framework to analyze logs and
provide feedback to guide the fuzzer)*. Stockholm, Sweden.
- Yeap, E., & Uy, I. (2014). Marker Clustering and Heatmaps: New features in the Google Maps Android
API Utility Library.
- Yu, M., Hillebrand, A., Tewarie, P., Meier, J., van Dijk, B., Van Mieghem, P., & Stam, C. J. (2015).
Hierarchical clustering in minimum spanning trees. *Chaos*, 25(2).
- Yukic, T. S. (1970). *Fundamentals of recreation, 2nd edition*. Harpers & Row.