SUSTAINABLE TOURISM IN AMSTERDAM

Final Report

FairBnB & Science Shop

Monitoring Impact and Assessing Data Driven Solutions
July 2019 | Group 12

Intan Ika Apriani, Melanie Arp, Joey Hodde, Joep van Leeuwen, William Tjiong
Coach: Dr. H. Bartholomeus

Executive Summary

Almost four out of ten tourists that travelled to the Netherlands in 2017 visited Amsterdam. This number is expected to keep increasing in the future. The growth of tourism in popular destinations is beneficial for the economy and regional development. However, the problems with tourism arise when the benefits of increased tourism overshadow the costs. An increase of negative attitude towards tourist has been observed in popular destinations such as Amsterdam. The aim of this study is exploring data sources that serve as indicators for the impact of tourism on the attitude of inhabitants towards tourists and liveability in neighbourhoods. Attention is given on exploring independent variables that can provide insights on tourist whereabouts, activities and how tourism changes the scene of the city. For each dataset a guideline is provided that explains how to retrieve the data and analyse it if applicable.

Relevant variables to explain the whereabouts of tourism include AirBnB and hotel bed densities and the infiltration of AirBnBs. Additionally, data sources such as housing prices and shop establishments indicate to what extent tourists contribute to gentrification in the city. Besides, social media data contain useful information in explaining tourist activities in the city. To assess the attractiveness of touristic places, TripAdvisor reviews were also explored. There is a relationship between number of reviews and the number of visitors. Other potential data sources, that were not extensively explored but can add value to the study include: Amsterdam Pass, GVB OV-chip card data, KVK data, neighbourhood surveys and Visa card data.

I Amsterdam City Card data provide information concerning where tourist visits frequently. GVB OV-chip card data may explain how tourist move in the city using public transport. An annual neighbourhood survey is done by the municipality of Amsterdam. The survey data provides an indicator about the liveability of the city based on residents' opinions. A more elaborate research on the shops can be done via the use of data from the chamber of commerce (KVK) of the Netherlands. Lastly, there is data available on where and on what people spend their money due to the availability of APIs on the website of VISA. With these data sources and potential data sources an impact assessment of tourism can be done.

The key datasets that we found in combination with the potential datasets are recommended for further use due to their ability to provide insights on tourism impacts in neighbourhoods. There are concerns when it comes to the validity of the datasets as there are ethical considerations and data policy issues. A factor of importance is the ability to distinguish patterns in data relating solemnly to tourists and their behaviour. Recommendations reflect the need for this phenomenon but with the emphasis on keeping in line with data mining policies and ethical aspects. Furthermore, extensive data guidelines and workflows make it possible to reproduce many of the analyses with data from other regions.

Table of contents

1. INTRODUCTION

Tourism is a major driver of both the global economy as well as regional development. According to the World Travel and Tourism Council (WTTC), the travel and tourism industry contributed up to 8.8 trillion dollars and provided more than 300 million jobs to the world economy in 2018 (World Travel and Tourism Council, 2019). The tourism industry has shown a growth rate of 3.9% compared to 3.7% global GDP in 2018 (Statista, 2019), which indicates that the industry is growing faster than the world economy. Tourism is also an important sector for the economy in the Netherlands. According to 'Het Nederlandse Bureau voor Toerisme en Congressen' (NBTC), about 19.1 million of international tourists visited the Netherlands in 2018 which is an increase of 7% compared to the previous year. Travelers contributed about 12.7 million euro to the economy of the Netherlands in 2018. 43% of the tourists in 2018 originate from Germany and Belgium, which indicates that the Netherlands is increasingly popular for neighbouring countries as well (Nederlandse Bureau voor Toerisme en Congressen, 2019). Moreover, the number of tourists in the Netherlands is expected to reach 29 million by 2030 (Solanki, 2018). Tourism has multiple impacts at popular vacation destinations. Impacts of tourism become more pronounced when the quality of residents' lives, values, norms and identities are subject to change (Zhuang et al., 2019). However, there are positive and negative impacts of tourism.

According to a study done by the Transport and Tourism Committee (TRAN) of the European Parliament, 60 popular destinations in Europe are subject to the negative impact of increased tourism. Six of the touristic places are situated in the Netherlands and include Giethoorn, Maastricht, Scheveningen, Amsterdam, Zaanse Schans and Kinderdijk (Peeters et al., 2018). 37% of the tourists that visited the Netherlands in 2017 went to Amsterdam, and the amount of tourists that are visiting the Netherlands keeps on growing (Centraal Bureau Statistiek, 2018). In European cities, a trend has come up where there is a negative attitude against tourism and tourists (Martin et al. 2018).

Sometimes the term overtourism is used to describe the problems in an area that are a consequence of large amounts of tourists. Overtourism has been named as a big problem in Amsterdam (Dickinson 2018, de Vries 2018). The term was first coined in 2012 in a Twitter hashtag (Goodwin, 2017) and can be defined as a situation where host, locals or travelers, feel that the quality of life and experience in the area has declined due to the mass flow of visitors. Today, low travel costs and affordable accommodations are stimulating citizens to travel more to various destinations and sometimes for multiple short haul flights each year (Alexis, 2017). Several different types of impacts can be generated by the tourist activity in an area. Almeida et al. (2016) were able to distinguish three major negative effects of tourism: ecological, sociocultural and economic effects . In an ecological context, the tourism increases pollution, noise and garbage. These factors deteriorate the natural environment. In a sociocultural context, the tourism generates loss of festivals and traditions, increases the use of drugs and alcohol, produces more congestion, accidents and parking problems and causes more crime. From an economic perspective, tourism has a negative effect on the price of housing, the cost of living and it generates instability in the employment. As a result people are pushed out of the city centre and the centre is taken over by tourists (Milikowski, 2016). This results in the city centre having only shops and buildings that are interesting for tourists and the people that live in Amsterdam get pushed out of the centre.

On the other hand, more tax and revenue generated from tourism can be used to develop new regional infrastructure, improved public services and more variety in recreational activities. These have positive impacts on the living conditions and quality of life of local residents. Moreover, travelers share stories about their visits, which can help in promoting cultural awareness and preserving tradition and cultural heritage as well (Goodman, 2019). Besides, it promotes the feeling of nationality to locals (Zhuang et al., 2019). Tourism can also strengthen communities by improving earnings prospects through job training and business developments in small areas. Depending on the destination and length of the trip, tourism can actually help the environment. National parks, wilderness areas, and conserved areas in developing countries derive income from tourists to manage and preserve plant and wildlife habitats (Castley, 2011). This made evident that tourism is beneficial for the economy and socio-economic development in countries.

The people in Amsterdam experience most of the indirect and direct impacts of tourism in the area. Whereas some people might experience economic benefits due to tourism others might find it a nuisance that these tourists reside in their neighbourhood. To be able to better understand the local differences in sentiment we will look into the similarities and dissimilarities between the different neighbourhoods in Amsterdam. These (dis)similarities might indicate why people that live in a neighbourhood show certain sentiments. In this research we will not be able to focus on the actual sentiments that people have in the city, instead we will focus on collecting data that might explain the sentiments. Therefore our main goal of the research will be:

“Exploring data sources that serve as indicators for the impact of tourism on the attitude of inhabitants towards tourists and liveability in neighbourhoods”

2. METHODOLOGY

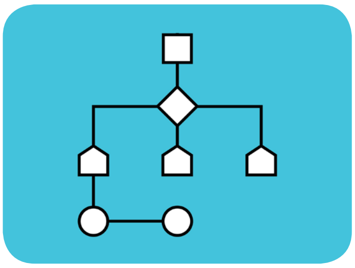
Potential data sources must give insights on the impacts of tourism in a city. One important factor that impacts tourists' dispersal throughout a city are the accommodation locations. Secondly, the movements of tourists through a city and the attractions they visit give insights on tourist behaviour. Ideally, this provides options to change their behaviour and relieve the pressure of tourism in a city. Finally, it is of great interest to get variables that define the street scene on which different actors have their distinctive perspectives. The information gathered in this report falls in one of these categories. For any dataset we harvest or recommend it is invaluable to be able to make a distinction between temporary residents and long-term residents.

To assess spatial patterns in the different neighbourhoods of Amsterdam we have focussed on finding datasets that already have a neighbourhood attribute or which have a geospatial component and thus gives us the ability to perform spatial operations.

Besides the spatial component, there is a great benefit in finding data with an additional temporal aspect. The development of neighbourhoods through time shows where tourism is making changes most rapidly. The scope of this study is limited to the aforementioned neighbourhoods division which can be found through https://maps.amsterdam.nl/gebiedsindeling/.

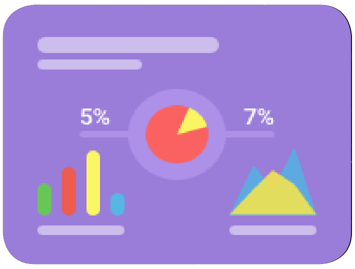
For each dataset there is a guideline in the form of a level 1 product which delineates where and how to retrieve the data. This enables the repeatability of data retrieval. Additionally we provide information pertaining to the data quality and privacy issues. If data sources were not available during this project but they are highly relevant for the goal of the project we include them as potentials. When data retrieval was possible, we have either produced visualizations or calculations pertaining to the validity with a subset of the data (level 2) or we provide a proof of concept integrated in an R-Shiny dashboard (level 3).

level 1



A product of this level contains a clear and structured elaboration about the data source and the ways it can be operationalized in order to provide valuable insights.

level 2



A level 2 product contains certain visualization elements in the form of graphs, static or dynamic maps. Possibly only covering a part of the dataset or a part of Amsterdam. It also includes an assessment of the data quality and usability.

level 3

A level 3 product is a proof of concept. This entails a working dynamic visualization, integrated in our R Shiny dashboard on our server.

3. RESULTS

This chapter provides an overview of potential data sources that serve as indicators and metrics for the impact of tourism on the attitude of inhabitants towards travelers in Amsterdam. Investigated indicators include: AirBnB and hotel beds intensities, shopping establishments, geotagged Flickr photos and Twitter posts, Tripadvisor reviews and housing prices. Other data sources that can support the analyses but which were not retrieved during this project include OV-chipcard check-ins, the Amsterdam Pass, KVK data and Visa card transactions.

3.1 CONCEPT DASHBOARD

To provide an idea about what a dashboard could look like, we have made a demonstration of a dashboard. For this demonstration we used subsets of the data that we were able to retrieve. The dashboard serves to visualize the different datasets in an easily understandable way. It gives the viewer an immediate insight on what is going on in the different neighbourhoods of the city. It allows you to select the different data sources and choose which part of the data you want to visualize and a small description about the dataset can be seen on the right of the screen. The dashboard can be found using this link: https://tinyurl.com/tourismdashboard. Alternatively, you can access the dashboard on mobile phone by scanning the QR code below. It should be noted that the dashboard was designed for desktop use, so mobile users might lose some functionality.

Scan me

3.2 TOURIST WHEREABOUTS / ACCOMMODATIONS

3.2.1 Airbnb and hotel beds

AirBnB was founded in 2008 as an online platform that provides a marketplace for hosts to accommodate guests with short-term lodge. Initially, the idea of AirBnB is to stimulate house owners to rent spare rooms for a small fee. Hence it provides travelers or tourists with an affordable alternative to expensive hotel rooms, especially for short-stay accommodations. Besides, house owners can earn extra income which yields a win-win situation (Milou, 2018). Today there are more than six million AirBnB listings worldwide spread over 100,000 cities and 191 countries and regions (AirBnB, n.d.). The rapid growth of AirBnB disrupts the hospitality industry by providing a way to find low-cost accommodations for low-value travelers without owning the properties (Christensen, Raynor, & McDonald, 2015). This is different than traditional hotel visits in which investment are considerably large for hotel owners to attract customers.

The abundance of accommodations also has an impact on communities and the housing market in Amsterdam. The number of vacation rentals in Amsterdam has tripled between 2015-2019 (AT5, 2019) mainly due to the explosive number of AirBnB listings, many of which are illegally rented all year round. This leads to a displacement in the housing market because less affordable houses become available for local people. Increasing numbers of tourist accommodations leads to increasing pressure in the city and changing sentiment towards tourism. Hence, data regarding number of beds can be used as an indicator in explaining the capacity of the city. Tourist intensity, calculated as the number of beds per inhabitant, is a common index used to assess the relative importance of tourism (Gutiérrez et al., 2017; Silva, et al., 2018). In order to normalize these numbers for the different neighbourhoods in the city of Amsterdam, we have calculated the intensity as follows:

$$Normalized\ Tourism\ Intensity = \frac{AirBnB\ beds}{AirBnB\ beds + inhabitants}$$

Furthermore, in order to properly assess the effects of AirBnB, it is useful to get an idea about the extent to which AirBnB is bringing holiday rentals to neighbourhoods that are, by means of zoning plans, not intended for tourism. In Amsterdam, especially the city centre contains many hotels and thus the infrastructure to support many temporary residents. However, since AirBnB rentals can occur without the need for any permit, the city of Amsterdam has no control over where these rentals take place. Peeters et al., (2018, p. 44) propose using the distance between an AirBnB address and the nearest commercial accommodation as a proxy for determining the extent that neighbourhoods that are not traditionally destined for tourism or accommodation business are infiltrated.

L3/Airbnb

Data provider	Inside AirBnB
Description	Inside AirBnB is an independent, non-commercial, open-source tool that provides listing data from the original AirBnB site for many popular destinations around the world including Amsterdam. The tool and site is provided by Murray Cox.
Data Access	Data can be retrieved through a URL link. Listing data of all cities are published on the Inside AirBnB site http://insideairbnb.com/get-the-data.html. An example of a link address to retrieve the listing data of a city from a certain month can be found at http://data.insideairbnb.com/the-netherlands/north-holland/amsterdam/2019-05-06/data/listings.csv.gz.
Time component	Historical listing data can be retrieved back to 2015 for each month for many cities around the world.
Data content	The listing data contains 106 attributes including: number of listings, type of accommodation, rental price, availability, number of reviews, number of beds, location and many others.
Data quality	According to the Inside AirBnB site (n.d.): Accuracy of the geolocation of each listing can vary from 0-150 meter. Listing in the same building appears scattered around the actual address to anonymize hosts from each other. Published listing data is just a snapshot of the listings on AirBnB site. Some reviews may be attributed as spam Some hosts may not update their calendar frequently or keep them highly available while they do live in the house or apartment. The accuracy of the number of listings may deviate by 10% from the true number (Slee, n.d.)
Privacy	Inside AirBnB (n.d.), does not retrieve private information of the host. However, usernames, photos, listings and review information, which are displayed on the original AirBnB site, are retrieved. The use of the data for analysis and visualization are available under Creative Commons CC0 1.0 Universal (CC0 1.0) “Public Domain Dedication”.
Reproducible in other regions	Inside AirBnB publishes listing data for many popular destinations around the world including: Barcelona, Venice, Antwerp, Athens, Singapore and many more.

Data analysis

Analyzing Inside AirBnB data

Figure 1 depicts a brief workflow on how Inside AirBnB listing data can be processed and analyzed to derive tourist intensity (tourist beds per inhabitant). Other potential derivatives include rental prices and the number of illegal listings. As of January 1st, 2019, the municipality of Amsterdam limited AirBnB listing to 30 days per calendar year to reduce tourist flow (Municipality of Amsterdam, n.d.). Listing data can be used to assess hosts who violate the rules. The data are then aggregated per neighbourhood (476 in total). The final product is a map containing tourist intensity, average listing price and number of illegal listings per neighbourhood.

Analyzing Hotel bed intensity

Hotel bed intensity is analyzed in a similar way as the Airbnb bed intensity. Hotel listing data from the Amsterdam Data Portal is aggregated with the population per neighbourhood data.

L3/Hotel Data

Data provider	Amsterdam City Data
Description	The city of Amsterdam has an online database containing relevant information about the city and its inhabitants. This is openly available and updated regularly
Data Access	The data can be collected as a XLS from the website as well as a CSV file. As mentioned, the data is open source and can be found through https://api.data.amsterdam.nl/dcatd/datasets/-Ja51qSD_owMzg/purls/3.
Time component	The data is available for one year only.
Data content	Data contains information about all the hotels. Per hotel the name, address, number of stars, number of rooms and the number of beds are available.
Data quality	The data is from 2014. This means that not everything might be up-to-date. Due to the hotel stop of the city of Amsterdam we do not expect there to be major changes.
Privacy	Amsterdam City Data provides open data under Creative Commons Attribution and Creative Commons CCZero licenses.
Reproducible in other regions	This source covers only the city of Amsterdam. It might be available for other cities depending on the data structure at hand but that requires extra searches.

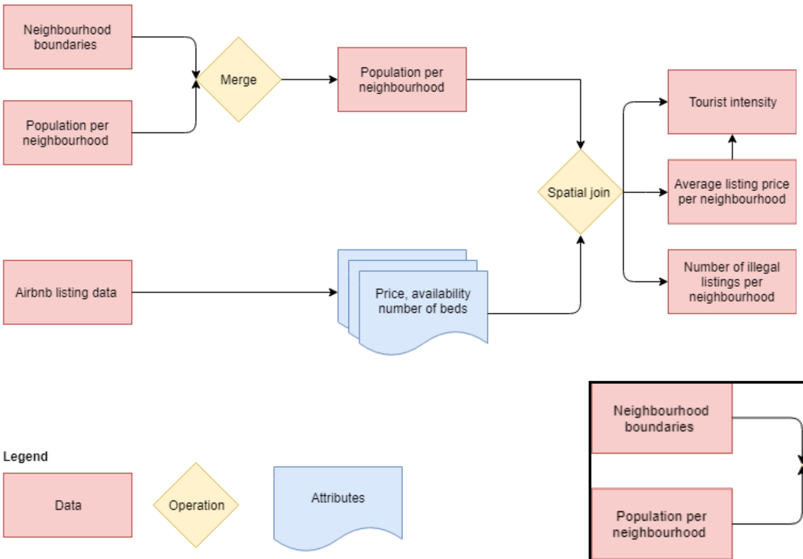

Neighbourhood boundaries
Population per neighbourhood
Merge
Population per neighbourhood
Spatial join
Tourist intensity
Average listing price per neighbourhood
Number of illegal listings per neighbourhood
Airbnb listing data
Price, availability number of beds
Legend
Data
Operation
Attributes

Figure 1 Processing Inside AirBnB data

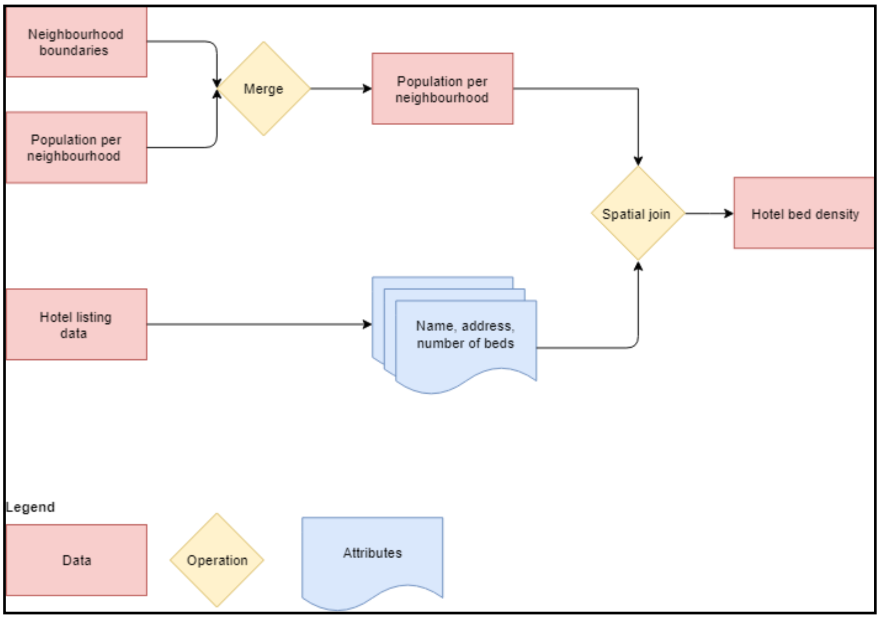


Figure 2 Hotel bed data processing workflow

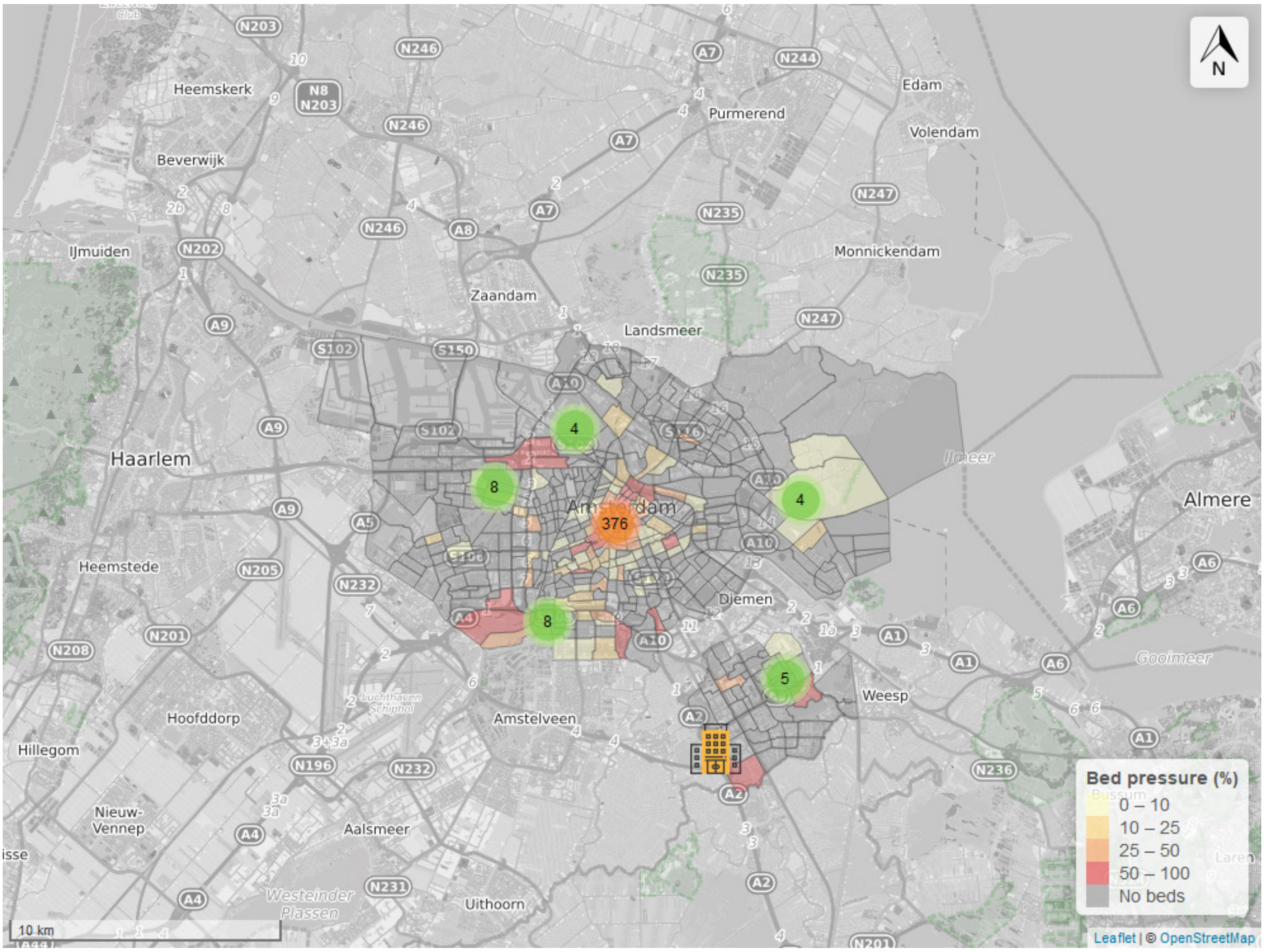


Figure 3 Hotel bed intensity map

Analyzing infiltration

In order to assess the amount of infiltration taking place we have calculated the average distance to the nearest hotel for every known AirBnB address. These differences have been aggregated on a neighbourhood level by calculating the mean distance per neighbourhood. In our 2D visualization, these distances are shown. In our 3D visualization, the aforementioned AirBnB tourist intensity has been added. The Z-value corresponds with this variable. Figure 4 depicts the flowchart corresponding with this methodology.

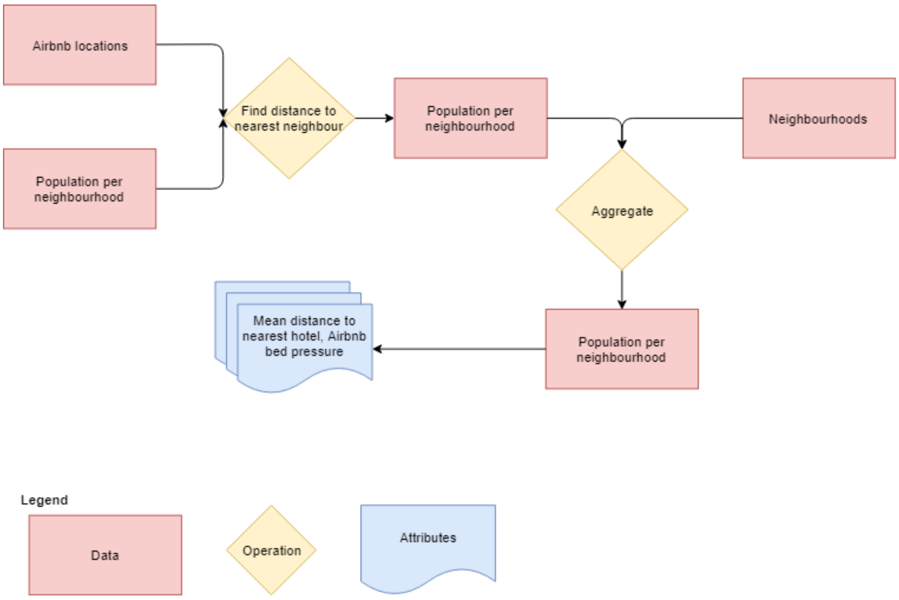

Airbnb locations
Population per neighbourhood
Find distance to nearest neighbour
Population per neighbourhood
Neighbourhoods
Aggregate
Population per neighbourhood
Mean distance to nearest hotel, Airbnb bed pressure
Legend
Data
Operation
Attributes

Figure 4: AirBnB infiltration workflow

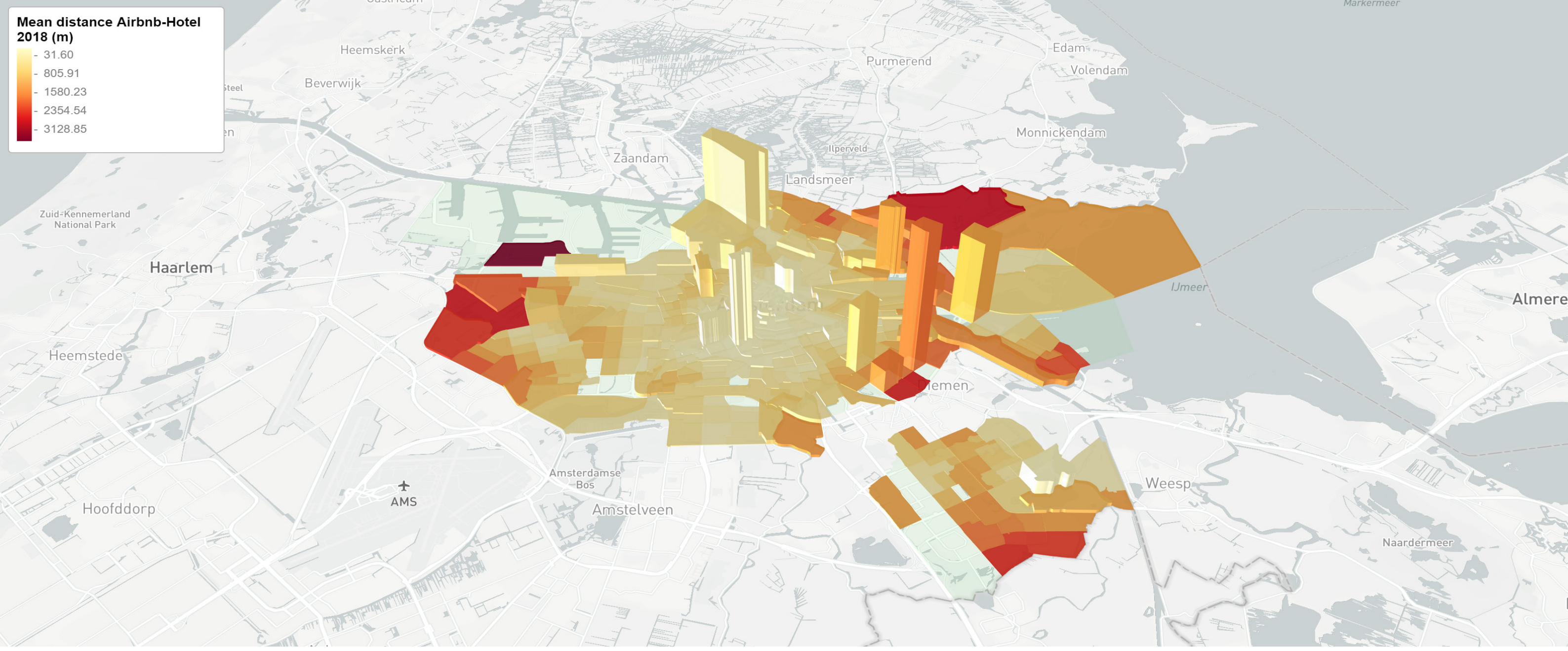


Figure 5 3D map of Airbnb infiltration

3.3.1 Geotagged Tweets

The geotagged tweets are a textual representation of people's behaviour at a given time and place. The data harvested from geolocated tweets can be used in a number of interesting ways. To distinguish between tourists and locals tweeting, the hashtags on which you harvest can be chosen such that there is a large certainty only tourists use them. By looking at where in Amsterdam the tourists are tweeting you can derive tourist densities throughout the city as well as hot-spots of tourists. Tweets also offer a potential to investigate spatio-temporal patterns of tourist movement and behaviour (Di Minin et al., 2015). The time component plays a big role and could even be a real-time product, however we have chosen to show only a demo with a subset of the data harvested over a short period.

L2/Twitter data

Data provider	Twitter API
Description	Twitter is a popular social media platform where users can share 'tweets' with the world. Tweets may contain text, photos and location information. Additionally, users can retrieve tweets about a certain topic using hashtags. The Twitter API provides a way for developers to collect tweets,
Data Access	To access the data connecting to the Twitter API is needed. The following steps need to be taken in order to use the twitter API: getting twitter API Key (https://apps.twitter.com/), connecting to twitter API, downloading data, reading and cleaning data
Time component	Tweets are created year-round and could be harvested in near real-time. In this project a subset of tweets was harvested in the week of the 15th of June till the 25th of June 2019.
Data content	The raw data contains 85 attributes, but we filter those attributes and keep the important attributes such as identity number, tweet time created, tweet text, twitter username, and location (latitude, longitude). Note: we will not make these attributes available to conform with the Twitter policies but we only use them to validate the use of certain hashtags
Data quality	There will be a bias in the amount of tourists and non-tourist because of certain people do not activate their geo-location in their twitter application for privacy matter. To get a real time tweet it takes a long time to continuously retrieve the twitter data from the API service. The hashtag #amsterdam has given a lot of tweets from the server like the most widely used hashtag but several local people also use this hashtag, so this hashtag not really main representative to distinguish tourists and non-tourists.
Privacy	Developers have to make an agreement with Twitter to be granted access to use the API. The full terms and conditions covering terms of use and privacy policy of the API can be found at https://developer.twitter.com/en/developer-terms/agreement.html
Reproducible in other regions	The Twitter API can be used in other regions in the world. There can be differences in the use of hashtags, namely concerning the hashtag language.

Data analysis

The tweet data is analyzed based on common hashtags used by tourists to show tourist density in Amsterdam. We filtered the tweets based on a location filter with a radius of 12 km from the city centre of Amsterdam to cover all neighbourhoods. Additionally, sixteen commonly used hashtags by tourist are used as search parameters: #amsterdam, #canals , #damsquare, #holiday, #holland, #netherlands, #photography, #redlightdistrict, #rijksmuseum, #summer, #travel, #travellling, #travelphotography, #visitamsterdam ,#weekend and #iamsterdam. For the demo, we retrieved over 1600 tweets from the 15th of June till the 25th of June 2019 for the aforementioned hashtags, however only 800 tweets contained a geo-location tag.

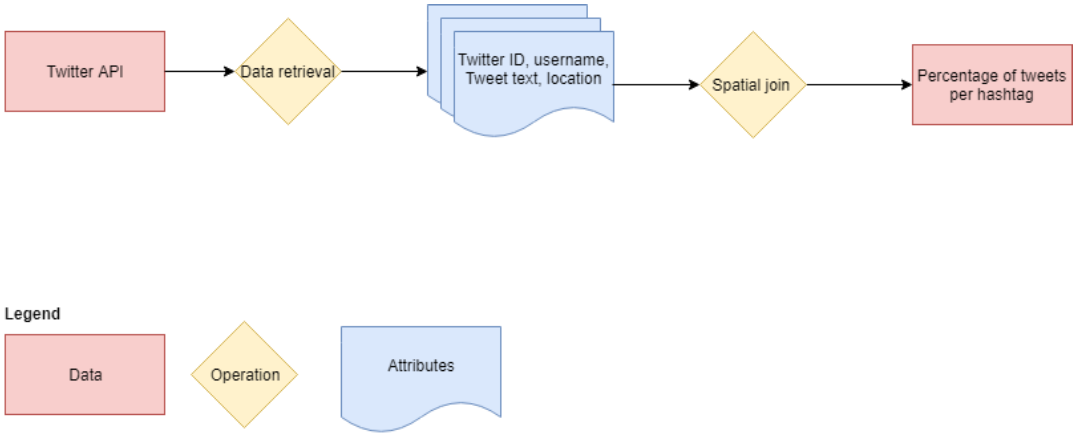

Twitter API
Data retrieval
Twitter ID, username, Tweet text, location
Spatial join
Percentage of tweets per hashtag
Legend
Data
Operation
Attributes

Figure 6 Workflow for harvesting Twitter data

3.3.2 Geotagged Flickr photos

A project done by Eric Fischer (2010) analyzed spatial patterns between tourists and locals by using geotagged Flickr Photos. The analysis was done for 130 cities including Amsterdam. The number of pictures taken in an area is an indicator for the number of people visiting a place. To discern tourists from local people, Flickr data can be labeled according to certains treats of these groups. For example, tourists tend to photograph touristic places, but locals prefer to avoid these places. Moreover, tourists might take several pictures in shorter periods and distances since they only stay at locations near their accommodation. Locals do not take pictures as frequently as tourists. Local people visit and take pictures of places where not many tourists may visit. This method is useful for investigating the distribution of residents and tourists across cities (Bliss, 2015). Flickr data also contains information about the users country of origin This information can be used to distinguish geotagged photos more accurately and to assess which people from which country visited the city.

Data analysis

Geotagged Flickr photos can be analyzed to create a map that shows the spatial distribution between international , locals and domestic travelers across the city. Domestic tourists refer to people living in the Netherlands but outside the city. The schema below shows how Flickr data can be analyzed and retrieved using the hashtag “amsterdam” and a location filter. The attribute user location is used to distinguish between tourists and locals. Additionally, a similar method as proposed by Eric Fischer was used to label users as tourist or non-tourist that do not provide user location information. For those users we used a rough estimation of the photo frequency:

- More than 100 photos taken in a month: international tourists
- Between 10 to 100 photos taken in a month: domestic tourists
- Less than 10 photos taken in a month: locals

L2/Flickr

Data provider Flickr

Description Flickr is a popular photo sharing platform with over 100 million users. Photographers and amateurs can upload up to one terabyte of pictures to the platform. The people that take the pictures are still the owner. These pictures are searchable on the Flickr site through tags. Flickr also provide an API in which developers can retrieve photo and user information based on tags.

Data Access To use the API and access the data, developers need to create a Flickr account and requests an API key through the Flick App Garden site: https://www.flickr.com/services/. This site also provide a documentation of how to properly use the API service. There are two types of API that Flickr provide: non-commercial and commercial keys. Non-commercial keys can be requested freely by anyone. Access to commercial keys requires a review of the project detail by Flickr. Requests made with the Flickr API are limited to 3600 queries per hour (a request per second). Besides an API key, a secret key is given for encoding and decoding requested information.

Time component 1.63 million public photos were uploaded daily on average in 2017 (Michel, n.d.). Depending on the tags and activity of the users, daily or most recent photos can be retrieved from the Flickr API.

Data content The Flickr API provide an extensive amount of information including:
user profile, user location, number of photos, date of photos taken, type of camera, photos in a certain location and many others. Additionally, users can edit their own photos and comments through the API. A list of Flickr API methods can be found at: https://www.flickr.com/services/api/

Data quality Several limitation of retrieving user information include the location accuracy of the photos. This is highly related to the accuracy of the smartphone or camera device. Current technology in smartphone GPS can achieve 5-10 meters accuracy. Furthermore, users can adapt their profile and add information such as country of origin. However, common semantic of country names are missing. For example, some users use 'The Netherlands', 'Nederland' or 'Holland' as country of origin and some only use city of residence.This requires proper data treatment prior to analysis.

Privacy Pictures are owned by the users. Developers need to comply with the terms and conditions which users attached to their photos. Additionally, Flickr provides a terms of use for their API to commercial and non-commercial users. The full terms of use can be found at :https://www.flickr.com/help/terms/api.

Reproducible in other regions A location filter can be used to retrieve geotagged photos from many places around the world.

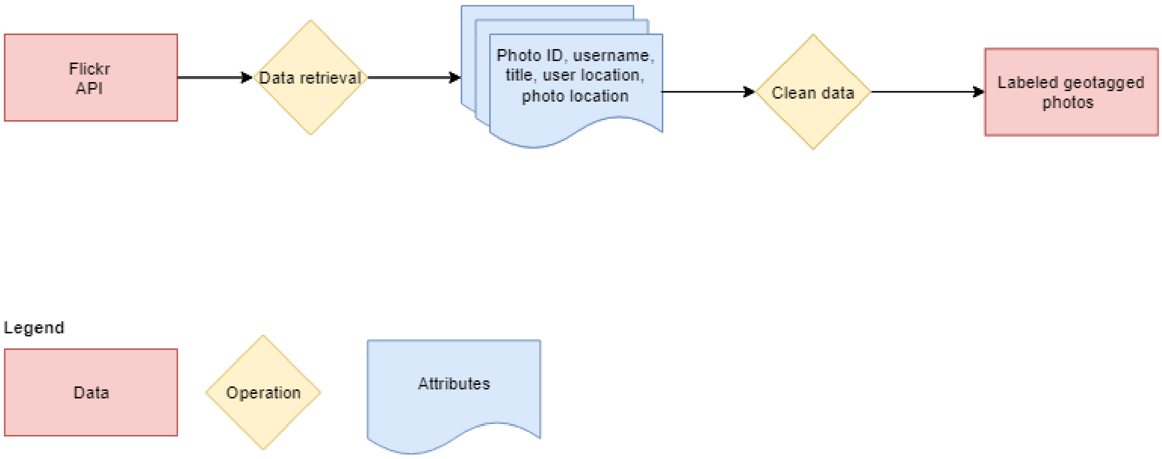


Figure 7 Processing Flickr API data

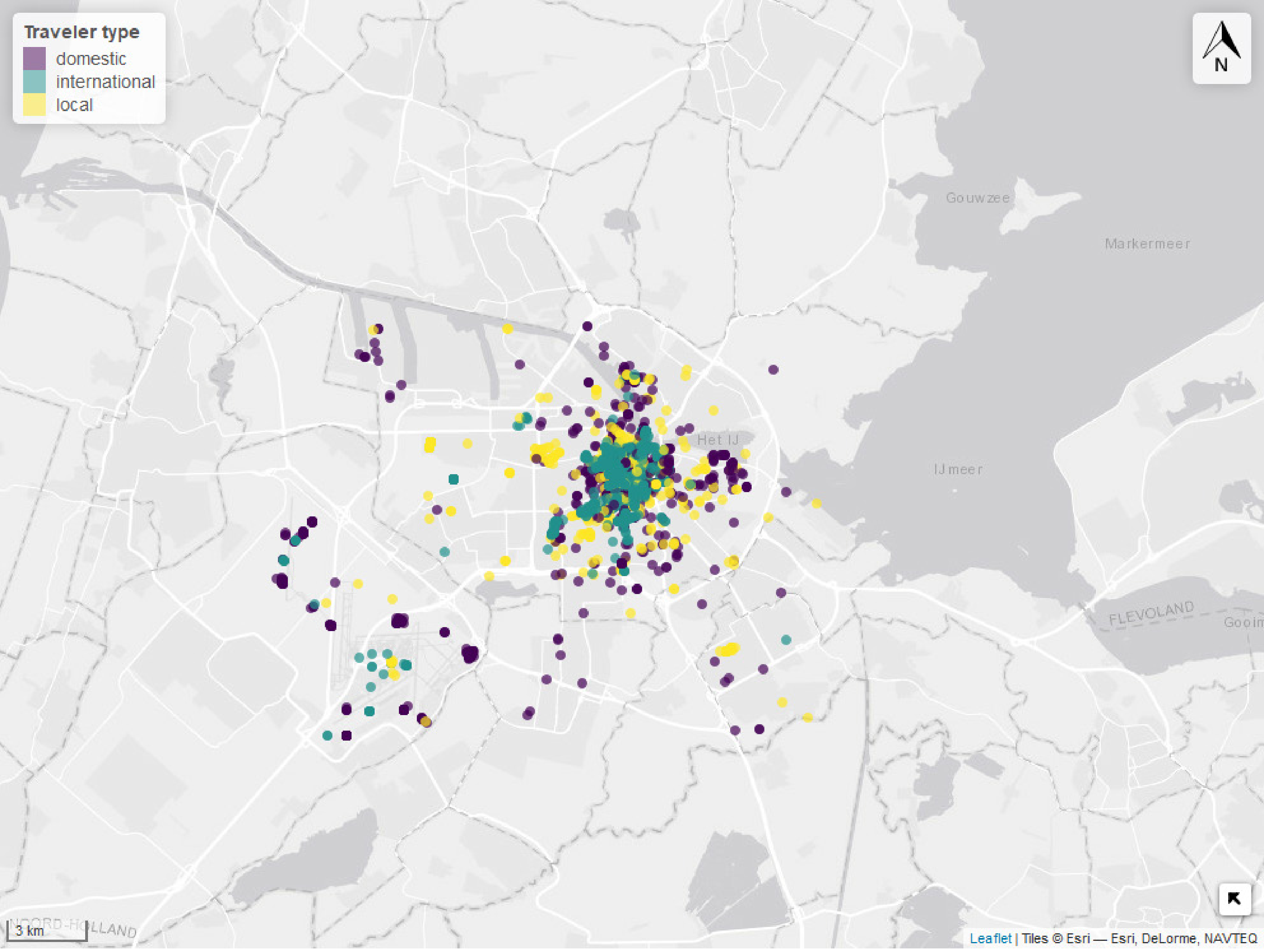


Figure 8 Geo-referenced Flickr posts from locals and tourists

3.3.3 TripAdvisor Reviews

In order to get an overview of what attractions tourists are visiting and in what numbers they are doing so, we have investigated the role a travel review website like TripAdvisor can play. We have assessed if the amount of reviews written by tourists in any way corresponds with the amount of visitors reported by the establishments themselves.

This work requires data acquisition from Twitter as the social media that we use. Twitter has provided access to their API, through which developers are able to interact with its. So we can retrieve tweet available on twitter using search function. Twitter search function enables us to find tweets based on keyword and location.

Data analysis

Figure 9 shows how review densities can be calculated. For all venues in Amsterdam, the timestamp of reviews can be used to tally the total amount of reviews for a specific year, month or week. This data can be aggregated on a neighbourhood level as a proxy for touristic pressure.

Validation

In order to assess the measurement validity of this method we have compared the amount of reviews with reported visitor numbers. The results have been plotted in figure XX and show a significant relationship with a p-value < 0.05. However, it should be noted that, since manual counting of reviews is a very time-intensive procedure, the sample size is 10 and the sample consists out relatively popular venues. Even though these results look promising, more research is needed if, in accordance with TripAdvisor, this data can be unlocked on a larger scale.

L2/TripAdvisor

Data provider Tripadvisor.com

Description By looking at the amount of reviews written within a certain time frame we can make assumptions on the amount of visitors of that specific attraction.

Data Access The pages of venues are in the public domain, which enables manual counting of reviews. In order to scale up this method, TripAdvisor must be contacted in order to attain this data in a structural and automated method.

Time component In order to be able to compare the reviews with the amount of reported visitors we have looked at the year 2017, since most venues have not reported visitor numbers of more recent years on their website.

Data content The only thing we have looked at is the count of reviews in the year 2017. The content of these reviews has not been used in any way, shape or form.

Data quality The reviews are stored in an orderly fashion and easily accessible. However, there are several biases that need to be taken into account. Firstly, the type of venue can have an effect on the fraction of visitors that will actually leave a review after having visited. For our preliminary exploration of the data, we have only looked at musea, so this inter-type bias could not be quantified. Secondly, within the museum-sector, different musea can have different target audiences. Some target audiences (e.g. younger audiences) might be more heavily inclined to leave a review at a website like TripAdvisor ad hence skew the image. Since the counting of reviews was done manually, we did not have enough time to gather a large enough sample in order to assess this inter-venue bias.

Privacy Since the only thing we are doing is counting the amount of reviews, no privacy-sensitive information is being used or processed.

Reproducible in other regions Since TripAdvisor is the leading social review site in the world and operates on a global scale, this data source i excellently suited to be used in other regions.

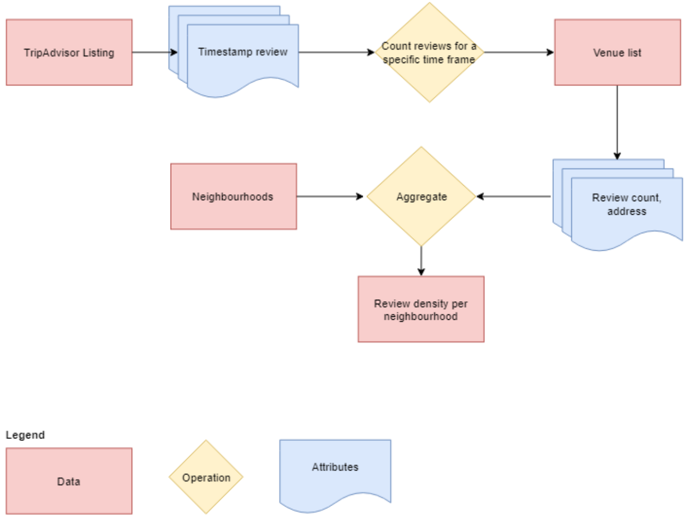


Figure 9 Workflow for analysing TripAdvisor Data

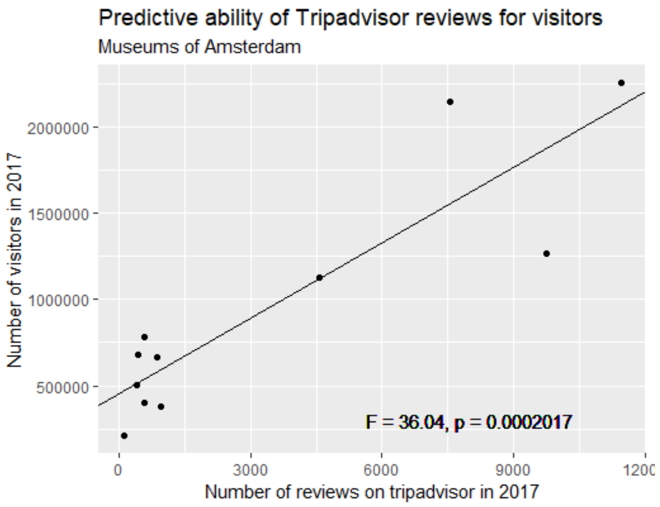


Figure 10 Statistical test research for measurement validity

3.4 CITY SCENE

3.4.1 House pricing

There is a shared sentiment among locals that the tourism industry contributes to the increase of property values (Almeida et al. 2016). This might be due to the fact that buildings in the centre of a city that has many tourists also has to house these tourists. The city of Amsterdam has reduced the amount of days people can have guests through AirBnB from 60 to 30 days in a year (Niemantsverdriet, 2018) and is taking more measures to overcome the problem. In addition to this, in combination with the AirBnB density data it should be possible to see if the house prices rise in areas where the density of AirBnB's are the highest.

Data-analysis

To calculate the average house prices per neighbourhood, the (geo)location of the neighbourhoods and the value data is needed. The data will be averaged for each of the different neighbourhoods which will result in an average value per neighbourhood.

Figure 11 provides the flowchart concerning the data collection and processing. The final product will be a map containing the average values for different years which makes it comparable with other years. This means that the trend for each neighbourhood is visualized.

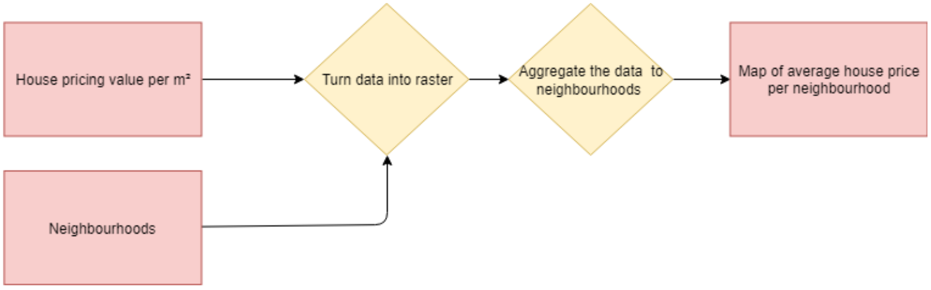
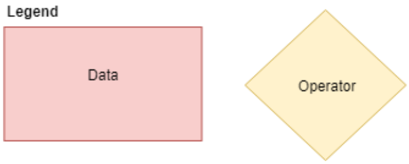

House pricing value per m²
Turn data into raster
Aggregate the data to neighbourhoods
Map of average house price per neighbourhood
Neighbourhoods

Figure 11 Workflow for analysing housing prices data



L2/Housing Prices

Data provider	Amsterdam City Data
Description	The city of Amsterdam has an online database containing relevant information about the city and its inhabitants. This is openly available and updated regularly
Data Access	The data can be collected as a shapefile from the website as well as an HTML. As mentioned, the data is usable and open to everybody. It can be downloaded on: https://maps.amsterdam.nl/open_geodata/
Time component	There is data for each year. So differences over the years can be compared quite easily.
Data content	When processed, the data shows the average price per municipality. This means that the differences between neighbourhoods can be distinguished and together with data from other data sources it can provide additional insights on where tourist pressure is (too) high.
Data quality	There have not been any big problems with the data quality. But due to the fact that there is only data about the house prices it can be hard to validate the results. From this dataset only it doesn't become clear if tourism is a factor that drives up housing prices.
Privacy	Amsterdam City Data provide open data under Creative Commons Attribution and Creative Commons CCZero licenses.
Reproducible in other regions	This source covers only the city of Amsterdam. It might be available for other cities depending on the data structure at hand but that requires extra searches.

L2/Shops

Data provider	Google Places API
Description	The Google Places API offers the possibility to retrieve up-to-date, location-based information. Places can be points of interest, geographic locations or establishments. For this project we focussed on the abundance and spatial patterns of establishments to get an indication for the street scene in cities.
Data Access	Many programming platforms have functionality to retrieve data from the Google API. In this project we have used R (version 3.5.1) which has the package 'googleway' (version 2.7.1) (Kahle & Wickham, n.d.) that can call on the Google API functionality. The data is commercial in nature and requires to create a developers account. However, Google grants every new user an amount of free credits worth 267 euros. The conversions differ per API, but for the places API 1000 requests cost 5 cents.
Time component	There is no opportunity to retrieve historical data on the places to be requested through the API.
Data content	A Google places request renders a number of attributes such as the address, name, photo and ratings. You can perform a text search which was used in this project to render souvenir shops. There is also an option to do a type search with which you can get only restaurants in a certain area. The supported types can be found at https://developers.google.com/places/supported_types.
Data quality	A request to the places API returns a maximum of 60 unique outputs. When wanting to retrieve data for a large area it is necessary to request data at multiple locations using the 'location' and 'radius' argument. There is however a high chance of duplicates amongst the requests which would have to be cleaned as well as it takes up credits. Another limitation is the 1000 requests per 24 hours limit that Google Places API has set. To cover large areas you might be restricted by this measure and it could take multiple days to retrieve the complete desired dataset.
Privacy	There is no data in the Google Places outputs which can be linked to individuals. All data is public on the Google Maps website. For the general terms of use of Google APIs please refer to https://developers.google.com/terms/.
Reproducible in other regions	The retrieval methods from the Google Places API are completely reproducible in other regions of the world as Google Maps has global coverage. There might be variability in the accuracy and amount of establishments that register themselves around the globe which should be taken into consideration when requesting information from the API.

3.4.2 Shop establishments

The value of a given area or neighbourhood can be derived from a range of indicators which make up the scene of that area. The scene can explain the development of neighbourhoods (Silver et al., 2007). An example of these indicators are the type and spread of establishments over the neighbourhoods. The Google Places API provides up-to-date establishment locations which draws from the Google Maps platform. Though the API also allows to get points of interest and a range of place types, we focus on the shopping possibilities and placement in the city.

Data Analysis

To retrieve a complete set of souvenir shops as well as total amount of shops, a regular sampling grid was created. At every sampling spot you can retrieve 60 shop locations so the outputs will have to be joined into one shop locations entity. To ensure all objects are unique a duplicate check is executed after which duplicates are removed. The amount of souvenir shops and total amount of shops per neighbourhood are counted. The final product then shows the percentage of shops that are souvenir shops (Figure 12).

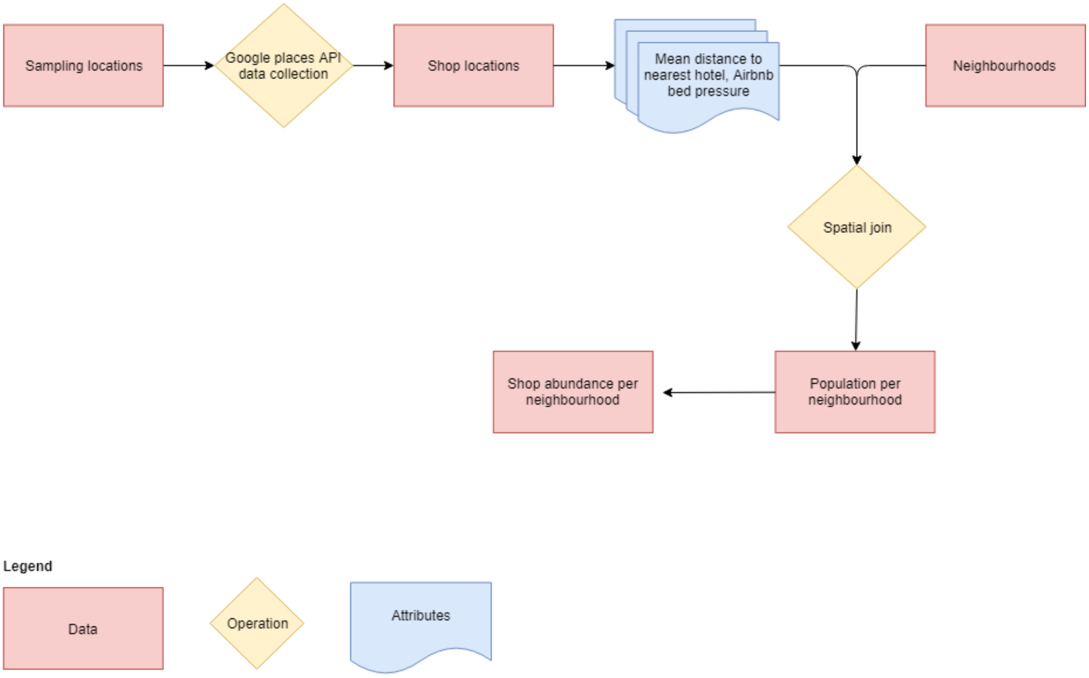

Sampling locations
Google places API data collection
Shop locations
Mean distance to nearest hotel, Airbnb bed pressure
Neighbourhoods
Spatial join
Population per neighbourhood
Shop abundance per neighbourhood
Legend
Data
Operation
Attributes

Figure 12 Workflow for retrieving shop abundances in an area with the Google Places API

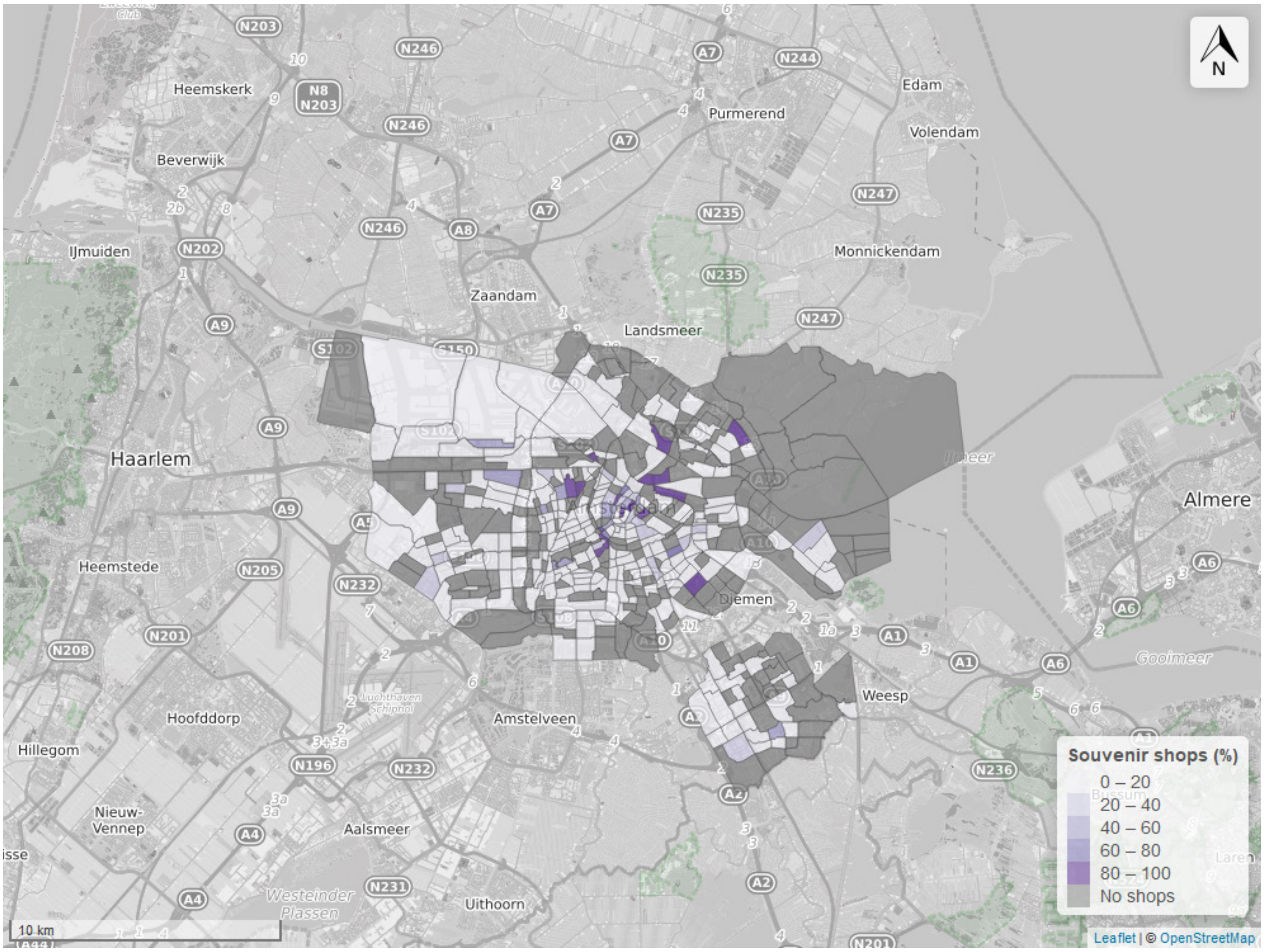


Figure 13 Souvenir shop spread

3.4.3 KVK Data

One of the impacts of increased tourism described by Peeters et al. (2018) are changing street scenes in which local entrepreneurs disappear and get replaced by chain-stores that are geared towards servicing tourists rather than local inhabitants. In order to get a grasp on these changing street scenes, it is imperative to collect information pertaining to historic registrations of shops on specific locations. Google Maps Places API does not provide historical data. The Dutch Chamber of Commerce (KVK) however, has a database with historical registrations, from which data can be requested against payment. Historical registrations for specific addresses can be classified into touristic and non-touristic categories, enabling the possibility of creating a time series analysis.

The data has to be requested through a sales representative of the KVK. This can be done based on an address register or SBI-code corresponding to the type of business.

Data analysis

Figure 14 depicts the workflow corresponding with this analysis. First, the zoning plan can be used to assess which addresses should be checked. Historical data associated with these addresses can be requested at the Chamber of Commerce. This data can be classified to filter out the touristic shops and be compiled into a time series.

L1/KVK

Data provider	Kamer van Koophandel (Chamber of Commerce)
Description	By requesting historical registrations of shops a time series can be constructed with information about changing street scenes.
Data Access	A request has to be made at the Chamber of Commerce specifying which addresses and SBI-codes can be used. The data has to be requested against payment The rate posted in the offer we have received is a start-up fee of €250,00, a fee of €0,07 per address and shipping costs of €11,00.
Time component	Historical registrations going back to 1992 can be requested.
Data content	90 different rubrics pertaining to anything from starting date to correspondence address can be requested. For this methodology, mainly starting date, end date, data about core business activities, the name of the company and address are relevant.
Data quality	The database is actively maintained by the Chamber of Commerce, sold commercially and accompanied by extensive metadata descriptions. Therefore, the quality of the business information can be assumed to be highly accurate.
Privacy	Even though the data can be purchased and used, personal ethical considerations lead us to believe that data concerning bankruptcies of small entrepreneurs can be regarded as privacy-sensitive. In order to be able to properly classify a shop, some information about the name of the shop is needed. Special care should be given to ensure that before and after aggregation on street level data any information that comprimises the privacy of enterpreneurs is anonymized or deleted.
Reproducible in other regions	Whether or not this type of analysis can be performed in regions outside of the Netherlands completely depends on how well data of this nature is being maintained by the local Chamber of Commerce.

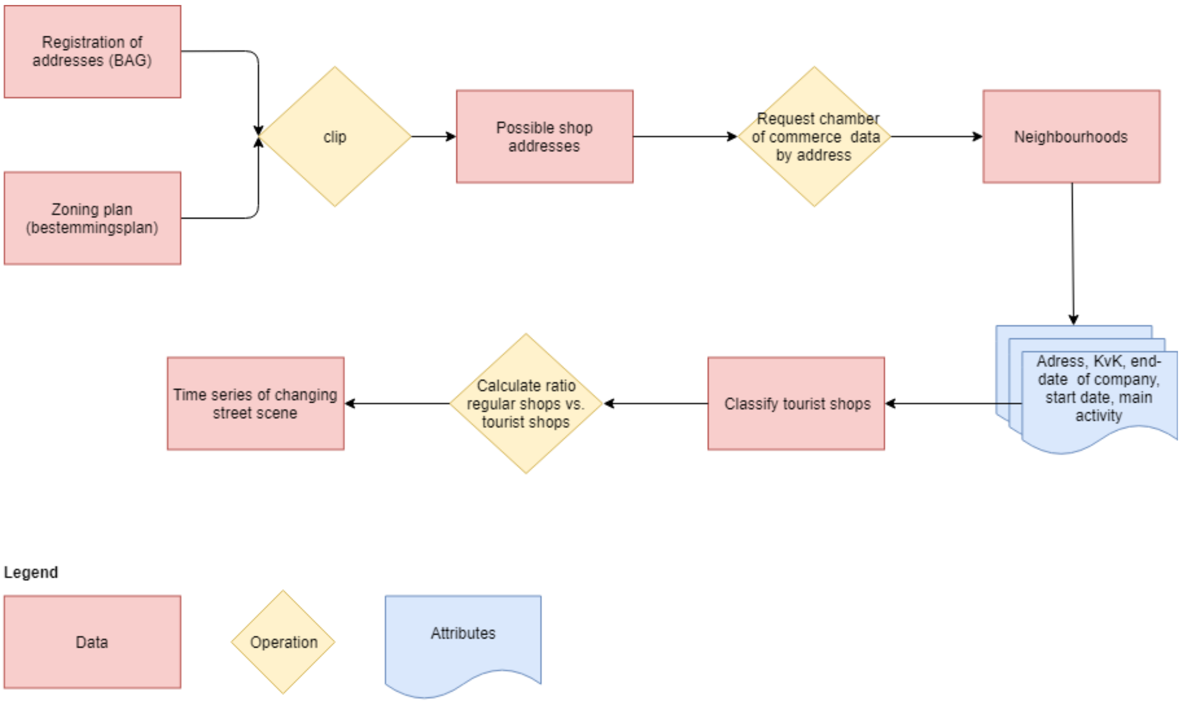

Registration of addresses (BAG)
Zoning plan (bestemmingsplan)
clip
Possible shop addresses
Request chamber of commerce data by address
Neighbourhoods
Adress, KvK, end-date of company, start date, main activity
Classify tourist shops
Calculate ratio regular shops vs. tourist shops
Time series of changing street scene
Legend
Data
Operation
Attributes

Figure 14 Workflow for analysing the Chamber of Commerce data

3.5 OTHER POTENTIAL DATA SOURCES AND FRAMEWORKS

3.5.1 I Amsterdam City Card

The I Amsterdam City Card provides tourists with the ability to use the public transportation and visit attractions for free or at a discount. Hence, the data that is produced by this card, provides valuable information about the movement of itineraries of tourists. Furthermore, since the provider of this card has the ability to market or provide a discount for certain attractions, the platform can also be used for influencing tourist behaviour. At the time of writing requests for access to this data from both our team and the commissioner have been denied, but since Amsterdam&Partners, the provider of the card, is operating in the public-private sphere, data requests may still be possible in coordination with the municipality of Amsterdam.

3.5.2 GVB

A different way to access information pertaining to how tourists move through the city would be to look at data from the public transportation agency of the city of Amsterdam, the GVB. In order to ensure that movement patterns of tourisms are distilled from all transport data we can distinguish local inhabitants from tourists by means of the tickets they use. There are several tickets intended specifically for tourists (IAmsterdam, n.d.), including, but not limited to:

- I Amsterdam City Card
- Amsterdam & Region Travel Ticket
- Amsterdam Travel Ticket
- Old Holland Tour

Furthermore, the GVB provides day passes that are not specifically meant, but frequently used by tourists. The GVB was contacted, but denied our request due to not having time at the moment. Again, our suspicion is that data requests in coordination with the municipality of Amsterdam will prove fruitful.

3.5.3 Neighbourhood surveys

The research, information and statistics department of the municipality of Amsterdam has conducted a research where they held a survey about the different sentiments that people have in the neighbourhoods. The results of the survey show in which neighbourhood people have a certain attitude towards the statements that are presented. In the survey there are also

statements that show an opinion towards tourist related issues, like where people do their groceries and why they do not use other places. The only problem is that the report is in Dutch. We tried getting the data that they collected in an excel file so we could provide it. In the end the excel file was not handed over to us and thus we were not able to make it part of our products. This data is highly relevant because it provides information about the dependent variable to which we tried to supply the independent variables. In combination with the independent variables, the data can say something about the usability of the spatial differences we observed. The data is from 2017, as of now we do not know if the survey is an annual survey or if it was held only once. The survey can be found here: https://amsterdamcity.nl/wp-content/uploads/2017/11/Ondernemers-versie-def_rapport_Stadsdeel_Centrum.docx.pdf

3.5.4 Visa

Visa provides an API which is able to retrieve anonymised and aggregated data from credit card transactions. Credit card data contains a lot information about the transaction and its user. In this way you can create a profile of the different transactions for each local merchant. This way it should be possible to see where tourists actually spend their money in the city. The Visa API provide only credit card data. In the Netherlands credit cards are not a payment method that are very popular (Credit Card, 2017) . Therefore the data for Amsterdam is mostly about tourists that visit Amsterdam and people that moved to Amsterdam from other countries. It gives an extent of where these people move and buy things in the city. From money withdrawals to buying groceries. The API and its description as well as the readme can be found at:
https://developer.visa.com/capabilities/visanet-data-services.

3.5.5 Amsterdam Open data

On the website of Amsterdam open data (data.amsterdam.nl) a lot of relevant information can be found. We provide an overview of what can be found on the Amsterdam Open Data platform and why it is relevant in Appendix B.

4. CLUSTERING ANALYSIS

The mentioned data sources and potentials can showcase quantitative differences between neighbourhoods. On their own they reveal tourism-related processes and if they are integrated they might unveil relationships and a multi-facet outlook on tourism in the city. Taking it one step further, there is a possibility to group neighbourhoods based on the multiple data sources and link certain characteristics to each group of neighbourhoods. An approach as such is a clustering method which groups entities based on similarities and dissimilarities. Especially when we enter the field of data mining it is useful to explore patterns in high-volume datasets through clustering (Yu, 1977; Han et al., 2001).

In this project, a clustering analysis has been performed on a few variables to demonstrate the capabilities and usefulness of this method. Neighbourhoods were clustered in three groups based on the variables population density, hotel bed pressure and airBnB bed pressure. To understand how the clusters were derived, a Principal Component Analysis (PCA) gives insight into which variables lead to a distinction between the neighbourhoods. In this example, there are a number of neighbourhoods which are typified as having distinct hotel bed pressures when comparing them to the other neighbourhoods (Figure 15). These are mainly a part of city district 'Centrum' and thus the cluster that contains neighbourhoods of this district is characterized by unusual hotel bed pressure. Similarly, the neighbourhoods from the 'Zuid' and 'West' districts are differentiable from the others by airBnB bed pressures. This coincides with a cluster that typifies neighbourhoods adjacent to the central neighbourhoods by distinctive airBnB bed pressures. Other neighbourhoods show a lot of overlap in the PCA plot and are thus not distinctive when it comes to these two variables. Figure 16 shows the distribution of these grouped neighbourhoods on a map.

This method of clustering could be used on or including a different subset of the variables from chapter 3 to expand the analysis. The groups then show characterizations related to tourist whereabouts, behaviour or sentiment towards tourism in the neighbourhoods, depending on the purpose. Eventually, this specification of neighbourhoods and their connection to tourism can support the detection of neighbourhoods that are open to an increase in (sustainable) tourism and where this poses more of a challenge.

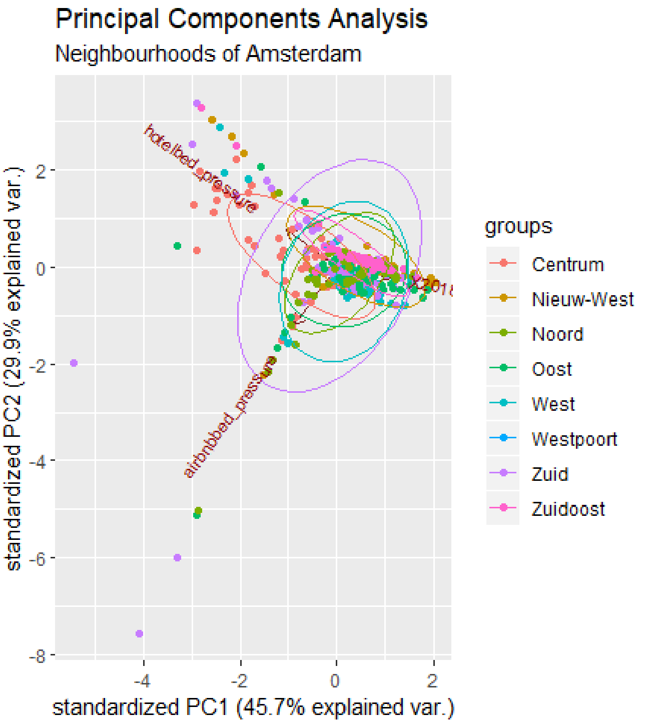


Figure 15 Principal component analysis using hotel beds and airBnB pressures per neighbourhood in Amsterdam

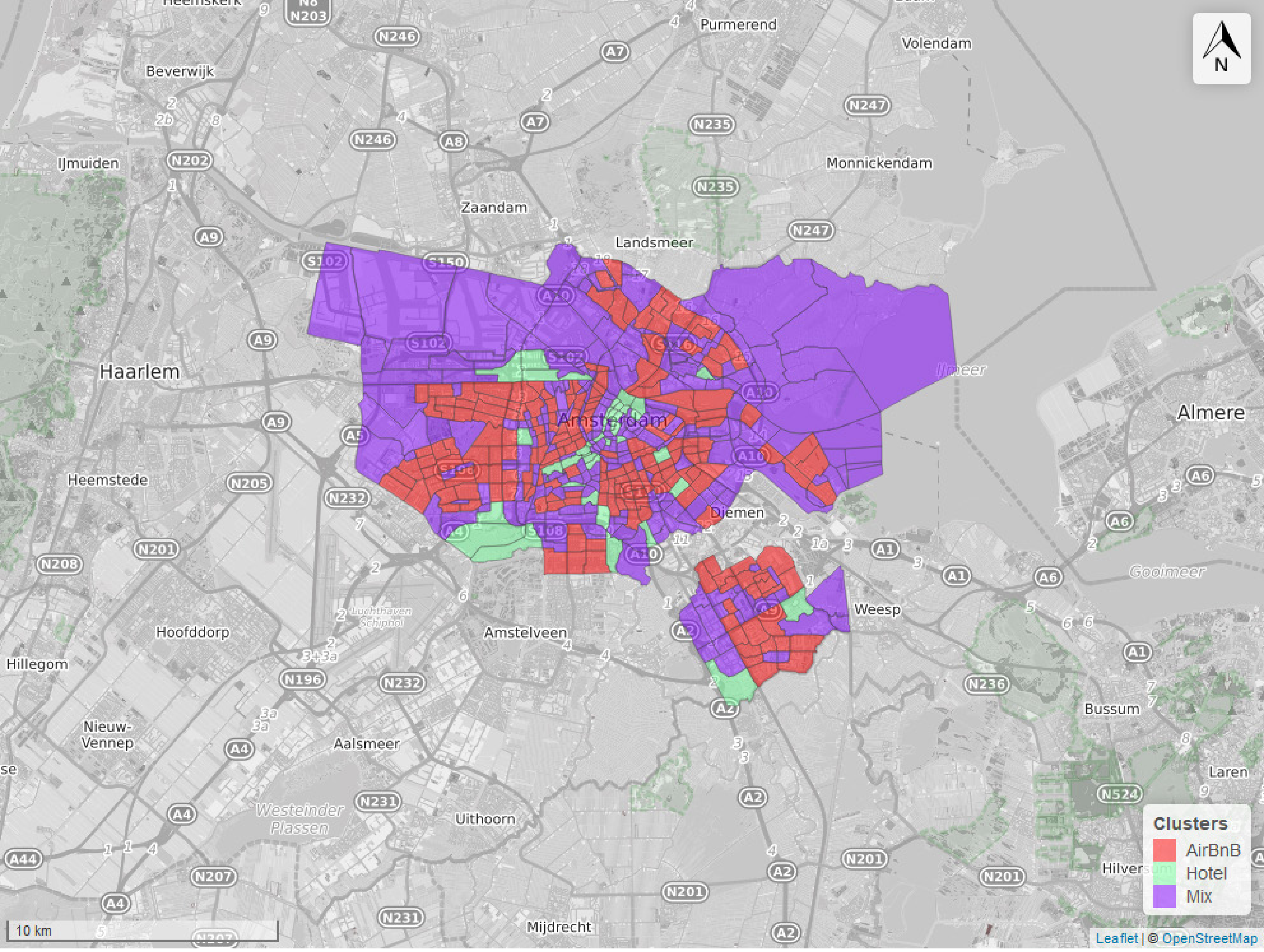


Figure 16 Map Clustering of neighbourhoods with indication regarding what makes the clusters separable

5. CONCLUSIONS AND RECOMMENDATIONS

The aim of this study is to investigate how data can be used to assess the dissimilarities in sentiment towards tourism in the city. Various online data sources and frameworks that provide insights on the impact of tourism in Amsterdam are explored and evaluated. Potential data sources are assessed based on its relevance in explaining tourist dynamics such as where travelers mostly stay or visit across the city. More importantly, the retrieved dataset should function as an explanatory variable which could explain the differences in behaviour between tourists and local residents. Furthermore, time and reproducibility are also essential data components which provide insight on how tourism change over time and the usability of the data sources in other popular cities. Key indicators and datasets that could explain the dynamics of tourism and discrepancy between locals and tourists include: AirBnB and hotel beds intensities, geotagged social media data, house pricing and shop establishments.

Whereabouts

To gain an inside in the whereabouts of tourists in the city we were able to collect several datasets. These include the bed intensities for AirBnB and hotels and we were able to calculate an average distance from an AirBnB to a hotel. These datasets all tell something about where people stay and sleep during their visit in the city. When a certain neighbourhood is filled with accomodations where tourists can sleep, the residents of Amsterdam most probably feel the direct impact of tourists on a daily basis. The visualizations we made show clearly where the pressure (number of beds per inhabitant) of both the hotels and the AirBnBs is the highest. We can depict that for the city centre the amount of beds in hotels is limited whereas the amount of beds of AirBnBs is higher.

In addition to this, we calculated the average distance from an AirBnB to a hotel. This result is called the infiltration of AirBnB in the city. The infiltration shows that in some neighbourhoods where no hotels are allowed to be built and the existing hotels are far away still AirBnBs can be found. This means that the measures the municipality tries to take to avoid having accommodations in certain areas are not working to at least some extent. Due to the timeline we can also see the changes that are happening overtime.

Activities

Evidently, the impacts of tourism are not only visible in the places where the tourists sleep. It is also relevant to look at what they do during the day. For this we were able to find data from social media. This data shows where people are and what they do. We did this for both Twitter and Flickr data. The Twitter data contains geolocated tweets that people send out. We were able to find these by looking at certain hashtags. A difficulty with this is that it is hard to make a distinction between what the tourist tweets are and what the local residents tweets are. This was more clear for the Flickr data. Because this contains user location information, like the country that the user is registered in. This provides a distinction between the residents of Amsterdam and the Netherlands and people that visit from other countries.

There is a lot of potential in finding out where people are moving through the city. The Twitter and Flickr data both have a risk of being biased because the chance that for example a photograph was taken at a tourist hotspot are high. Therefore to find out where people are actually moving within the city other data can be very useful.

Although we were not able to provide the actual data the potential for payment data and public transport data is big. The public transport data can be used to see where people are travelling in the city. But they also include information about where people use the public transport for the first time in the morning and the last time in the evening. Using the GVB and Iamsterdam card data therefore can tell a lot about both the movement patterns and the accommodation locations. The payment data that can be retrieved using the VISA API will provide information on where people spend money. They will show the moving patterns of tourists in the city and show where people actually spend their money. This might be an influence on how the people in that neighbourhood think about tourists and tourism in general and therefore can be combined with information on possible dependent variable.

City scene

Due to the increasing amount of tourists that visit the city every year the city itself is subdued to change as well. We were able to retrieve datasets that give an insight on how the city has changed. The average housing prices per neighbourhood over the years shows that every year the house selling value per m^2 increases for some neighbourhoods. In addition to this, it can also be seen that more and more neighbourhoods are subdued to high house prices, especially in the city centre. The problem with this is that it cannot be linked to tourism directly. Although it might give an indication we can never be sure.

What's also important for the city scene and the residents of Amsterdam are the shops that can be found in the city. If there are more shops for tourists in a neighbourhood than there are shops for local residents it can be a nuisance. We were able to retrieve the shops in Amsterdam and make a distinction between souvenir shops and other shops. This shows to what extent the street view is influenced by souvenir shops for the neighbourhoods. This would be even more relevant if this included historic data about what shops there were in the past. We were not able to retrieve this but have found a way to do it. Using the data of the chamber of commerce of the Netherlands, it should be possible to get the historic data. In this way it is possible to make an assessment about the changes over the years. This will provide a holistic overview on how tourist presence in the city streets.

Data Terms of Use

Terms of use and privacy policy may differ across data sources and frameworks.
For example, Inside AirBnB (CC0 1.0) and all the data in the Amsterdam Data Portal (CC0 1.0 and CC BY 4.0) are open data and licensed under Creative Common License. Other data providers such as Google Places API, Flickr API and Twitter API provide commercial and non-commercial use services. The scope of this study is limited in exploring the non-commercial service of the APIs. Users and developers are recommended to refer to the terms of use and privacy policy for each data provider prior building their own application. Extra care should be taken into account when handling sensitive social media data. In this study, username information, tweet content and text were anonymized to comply with Twitter API Restrictions on Use of Licensed Materials section (Twitter, 2018). The final data product retrieved from the Twitter API only contains Twitter ID, time, location information and other non-personal information. A Twitter ID can be traced to an account. That's why we have ensured that nowhere in the dashboard it is visible for users. The final data product retrieved from the Flickr API also only contains non-personal information including photo title, photo ID and country of origin. This is done to comply with Flickr API Licensed Uses and Restrictions section (Flickr, 2018).

6. DISCUSSION

Validation

Before conducting the research, literature research was performed. From this we were able to derive certain indicators for (over)tourism in cities. Important factors are for example, hotel bed density and other accommodation densities. In our research we used the found indicators to highlight spatial patterns concerning tourism-related variables in Amsterdam. In addition to these indicators that can be found in literature, we looked for (spatial) data that has an obvious connection to tourism such as the TripAdvisor platform. Were it is possible we provide validation to why it is relevant to use a dataset, for example by the manual inventory of number of reviews for a small set of musea.

All the datasets reveal similarities and dissimilarities between the neighbourhoods when it comes to quantitative variables, but there is a lack of a dependent variable in our study case. Eventually, we would like to test the relationship between the indicator variables mentioned in this report and qualitative data on sentiment of the local actors towards tourism. This will unveil which neighbourhood characteristics link to either positive, neutral or negative sentiments towards tourism.

The processes that the recommended data sources refer to do not cover any short-term events. When looking at the time components most of the data sources are published annually (e.g. Amsterdam Data Portal), once in a month (e.g. Inside Airbnb) or real-time (i.e. Twitter). Therefore we do not foresee issues concerning the ecological validity since the data capture daily activities of the residents and locals in Amsterdam (Bryman, 2004).

Reproducibility and repeatability

In order for this research to be reproducible we added flowcharts and data guidelines for each of the datasets found. These guidelines provide an overview on where the data was found and how it can be retrieved. The flowcharts show the work that is needed to visualize the raw data. In this way, the products that were provided can be reproduced with minimum effort. The scripts that we used to retrieve and visualize the data are also made available on GitHub (Appendix A). They can be used to repeat the analysis conducted in this report, they can be slightly modified to include a larger subset of the data source if applicable or they can be run using the same data sources but in other regions. It is important to state that for APIs where keys are required the user needs to be willing to create their own accounts.

Something of added value for FairBnB is the reproducibility of the recommended data sources for other cities. This is mainly because of the fact that FairBnB is not only interested in the situation in the city of Amsterdam but also in other cities around the world. Data from globally operating platforms like TripAdvisor, Twitter, Flickr and Airbnb can be used place-independent, but the data provided by the municipal bodies might vary. The municipality of Amsterdam has a good-quality and elaborate, open database containing much data on the city itself and its residents. Research would have to be invested to find similar datasets on city statistics in different regions. In terms of external validity, being able to use the data or the workflows elsewhere (Bryman, 2004), there are many options with the recommended data sources to do so. It should also be clear from the descriptions of the explored municipality data what the extent is of a necessary dataset to implement it in one of the workflows presented in this report. This is under the condition that time and effort is spent on finding equivalent city data in other regions.

Other data sources that we were not able to retrieve for Amsterdam might be available for other cities. Therefore we advice to not only look to the datasets we used in this research but also to look for additional data sources that might tell something about the tourism-related differences within a city.

Limitations

A challenge in this research is the fact that it is difficult to trace whether the information is solemnly related to tourists. It is possible for example that people from the Netherlands used the different hashtags that were mainly used by the tourists as well for their tweets. In this way, a tweet can be misidentified as a tourist location. With some of our potential data sources we have the same problem. This issue concerns the internal validity of the research. Distinguishing between the different actors is highly important but challenging at times. Moreover, the extent to which each dataset can directly explains the impact of tourism is debatable. However, the majority of the explored datasets are satisfactory.

Where some of the data sources are applicable in the United States they are not (yet) usable in big parts of Europe. This mainly concerns the APIs of different banking or credit card companies. Where this market is still very much under development in, in this case, the Netherlands. American companies already use it with data collected in the United States. We expect that in a few years, these APIs will also be usable in Europe.

Another limitation we stumbled upon during our work is that for almost every dataset, the city is divided into different zones. To overcome this, we used the official neighbourhoods dataset that is provided by the municipality and tried to fit our data to each of these neighbourhoods. We did this to be able to compare all the different results efficiently.

Ethics and data policy

Although the types of ethical conflicts that come with data mining and looking for online data sources are not new, the way in which these ethical conflicts occur are. The people whose privacy or individuality is scathed are most of the time not aware of the fact, and unable to give their consensus about sharing their personal data (van Wel and Royakkers, 2004). A solution to this problem might be the grouping of individuals with (roughly) the same characteristics.

In this project, we read and stuck to the data policies and terms of use documents of all the companies that are involved in the research. This means that all of the data was legally obtained. Some personal information came with the retrieval from social media APIs such as Twitter and Flickr. However, information such as usernames, tweet content and any other sensitive data that can be directly linked to the person were removed during analyses or anonymized as far as needed. The main features we retrieved from the APIs are location data and ID numbers.

As already mentioned, every company or website has a different policy regarding their data. During the whole research we took careful measures not to violate any of these. Therefore we were limited in what we were able to show. Moreover this also limits the data sources and framework we can retrieve and search online.

ACKNOWLEDGEMENTS

The research process and creation of outputs was accompanied by a lot of guidance and feedback from different parties involved in the project. Firstly we want to thank our commissioners, Sito Veracruz and Garán Hobbelink of FairBnB and Margriet Goris of the WUR Science Shop. Thank you for the guidance and recommendations despite your busy schedules during the project. During the Skype meetings it became more clear what was expected and this helped to improve the overall quality of the products. In addition to this, we feel that the collaboration with both parties went smoothly and we enjoyed working with you. Secondly we want to thank Harm Bartholomeus, the coach of our team, for the personal guidance during the project. During meetings with Harm we were able to get the group on the same level and knew what we could expect from each other.

REFERENCES

Almeida, F.; Peláez, M.A.; Balbuena, A.; Cortés, R. Residents' perceptions of tourism development in Benalmadena (Spain). Tour. Manag. 2016, 54, 259–274. (crossref)

AT5. (2019, January 10). 60 dagen verhuren via AirBnB nog steeds mogelijk ondanks nieuwe regels. Retrieved June 17, 2019, from AT5: https://www.at5.nl/artikelen/190501/60-dagen-verhuren-via-airbnb-nog-steeds-mogelijk-ondanks-nieuwe-regels

Baltussen, L.B, Oomen, J, Brinkerink, M, Zeinstra, M, & Timmermans, N. (2013). Open Culture Data: Opening GLAM Data Bottom-up. In N Proctor & R Cherry (Eds.), Museums and the Web 2013. Museums and the Web.

Bliss, L. (2015, February 23). Where Do Locals Go in Major Cities? Check Out This Interactive World Map. Retrieved June 24, 2019, from CityLab : https://www.citylab.com/transportation/2015/02/where-do-locals-go-in-major-cities-check-out-this-interactive-world-map/385768/

Bryman, A. (2004). Social research methods. Oxford: Univ. Press.

Credit Card. (2017, March 14). Nederland ondanks voordelen nog niet weg van creditcard. Retrieved July 2, 2019, from Credit Card: https://www.creditcard.nl/nieuws/nederland-ondanks-voordelen-nog-niet-weg-van-creditcard

Castley, G. (2011, July 12). Can tourism really have conservation benefits? Retrieved July 2, 2019, from The Conversation: http://theconversation.com/can-tourism-really-have-conservation-benefits-1337

Centraal Bureau Statistiek. (2018, April 4). Grootste groei toerisme in ruim tien jaar. Retrieved May 24, 2019, from CBS: https://www.cbs.nl/nl-nl/nieuws/2018/14/grootste-groei-toerisme-in-ruim-tien-jaar

Chen, J. S. (2003). Market segmentation by tourists' sentiments. Annals of tourism research, 30(1), 178-193.

Christensen, C. M., Raynor, M., & McDonald, R. (2015). What Is Disruptive Innovation? Retrieved june 17, 2019, from Pedrotrillo: http://pedrotrillo.com/wp-content/uploads/2016/01/Whatisdisruptiveinnovation.pdf

Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. Frontiers in Environmental Science, 3, 63.

Dickinson, G.(2018, October 10) Mapped: How tourism is taking over Amsterdam. Retrieved from The Telegraph: https://www.telegraph.co.uk/travel/news/amsterdam-overtourism-travelbird/

D. Kahle and H. Wickham. ggmap: Spatial Visualization with ggplot2. The R Journal, 5(1), 144-161. URL http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

Dutch news. (2018, January 10). Amsterdam slashes Airbnb rental period from 60 to 30 days. Retrieved July 2, 2019, from Dutch News: https://www.dutchnews.nl/news/2018/01/amsterdam-slashes-airbnb-rental-period-from-60-to-30-days/

Flickr. (2018, May 9). Flickr APIs Terms of Use. Retrieved July 1, 2019, from Flickr API: https://www.flickr.com/help/terms/api

Gemeente Amsterdam City Data. Amsterdam City Data. Retrieved June 28, 2019, from Data Amsterdam: https://data.amsterdam.nl/

Goodman, P. (2019, March 27). The Advantages and Disadvantages of Tourism. Retrieved May 24, 2019, from Soapboxie: https://soapboxie.com/economy/Advantages-and-disadvantages-of-tourism

Goodwin, H. (2017). The Challenge of Overtourism. Retrieved May 21, 2019, from Responsible Tourism: https://haroldgoodwin.info/pubs/RTP'WP4Overtourism01'2017.pdf

Gutiérrez, J., García-Palomares, J. C., Romanillos, G., & Salas-Olmedo, M. H. (2017). The eruption of AirBnB in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. Tourism Management, 62, 278-291. doi:10.1016/j.tourman.2017.05.003

Han, J., Kamber, M., & Tung, A. K. (2001). Spatial clustering methods in data mining. Geographic data mining and knowledge discovery, 188-217.

Ismagilova, G., Safiullin, L., & Gafurov, I. (2015). Using historical heritage as a factor in tourism development. Procedia-social and Behavioral sciences, 188, 157-162.

IAmsterdam. (n.d.). Public transport in Amsterdam. Retrieved July 1, 2019, from IAmsterdam: https://www.iamsterdam.com/en/plan-your-trip/getting-around/public-transport

Inside AirBnB. (n.d.). About Inside AirBnB. Retrieved June 18, 2019, from Inside AirBnB: http://insideairbnb.com/about.html

Martín, J., Martínez, J. and Fernández, J. (2018). An Analysis of the Factors behind the Citizen's Attitude of Rejection towards Tourism in a Context of Overtourism and Economic Dependence on This Activity. Sustainability, 10(8), p.2851.

Michel, F. (n.d.). How many public photos are uploaded to Flickr every day, month, year? Retrieved June 28, 2019, from Flickr: https://www.flickr.com/photos/franckmichel/6855169886

Municipality of Amsterdam. (n.d.). Particuliere vakantieverhuur. Retrieved June 23, 2019, from Gemeente Amsterdam: https://www.amsterdam.nl/wonen-leefomgeving/wonen/particuliere/

Milkowski, F. (2016, April 13). Een luchtbedje met ontbijt. Retrieved May 24, 2019, from De Groene Amsterdammer: https://www.groene.nl/artikel/een-luchtbedje-met-ontbijt

Milikowski, F. (2016, July 27). Amsterdam als koelkastmagneetje. Retrieved May 24, 2019, from De Groene Amsterdammer: https://www.groene.nl/artikel/amsterdam-als-koelkastmagneetje

Milou. (2018, February 2019). Alles wat je moet weten over Airbnb: zo werkt het. Retrieved June 17, 2019, from Explorista: https://explorista.nl/airbnb-tips/

NBTC Holland Marketing. (2018, February) Tourism in Perspective. from NBTC: https://www.nbtc.nl

Nederlandse Bureau voor Toerisme en Congressen. (2019, January 29). Minder groei internationale toeristen in Nederland in 2018. Retrieved from NBTC: https://www.nbtc.nl/nl/homepage/artikel/minder-groei-internationale-toeristen-in-nederland-in-2018.htm

Niemantsverdriet, T. (27-09-2018). Airbnb laat zich niet zomaar aan banden leggen door Amsterdam. NRC. Available at: https://www.nrc.nl/nieuws/2018/09/27/airbnb-laat-zich-niet-zomaar-aan-banden-leggen-door-amsterdam-a1817406.

Oklobdžija, S. (2015). The role of events in tourism development. Bizinfo (Blace), 6(2), 83-97.

Panfiluk, Eugenia. "Impact of a Tourist Event or a Regional Range on the Development of Tourism." Procedia-Social and Behavioral Sciences 213 (2015): 1020-1027.

Peeters, P., Gössling, S., Klijs, J., Milano, C., Novelli, M., Dijkmans, C., ... & Mitas, O. (2018). Overtourism: impact and possible policy responses, Research for TRAN Committee. European Parliament, Policy Department for Structural and Cohesion Policies, Brussels.

Peeters, P., Gössling, S., Klijs, J., Milano, C., Novelli, M., Dijkmans, C., Eijgelaar, E., Hartman, S., Heslinga, J., Isaac, R., Mitas, O., Moretti, S., Nawijn, J., Papp, B. and Postma, A. (2018). Research for TRAN Committee -Overtourism: impact and possible policy responses (pp. 1–260). Brussels.

Silver, D., Clark, T. N., & Rothfield, L. (2007). A theory of scenes. University of Chicago, http://tnc. research. googlepages. com/atheoryofscenes.

Slee, T. (n.d.). AirBnB Data Collection: Methodology and Accuracy. Retrieved June 18, 2019, from Tom Slee: http://tomslee.net/airbnb-data-collection-methodology-and-accuracy

Solanki, M. (2018, October 12). Tourism numbers in the Netherlands set to explode to 29 million by 2030. Retrieved May 24, 2019, from I Am Expat: www.iamexpat.nl/expat-info/dutch-expat-news/tourism-numbers-netherlands-set-explode-29-million-2030

Statista. (2019). Growth of the global gross domestic product (GDP) from 2012 to 2022 (compared to the previous year). Retrieved May 24, 2019, from Statista:
https://www.statista.com/statistics/273951/growth-of-the-global-gross-domestic-product-gdp/

Twitter. (2018, May 25). Developer Agreement and Policy. Retrieved July 1, 2019, from Developer Twitter: https://developer.twitter.com/en/developer-terms/agreement-and-policy.html

van Wel, L. and Royakkers, L. (2004). Ethical issues in web data mining. Ethics and Information Technology, 6(2), pp.129-140.

World Travel and Tourism Council. (2019, February 27). Travel & Tourism continues strong growth above global GDP. Retrieved May 24, 2019, from World Travel and Tourism Council: https://www.wttc.org/about/media-centre/press-releases/press-releases/2019/travel-tourism-continues-strong-growth-above-global-gdp/

Yu, C. H. (1977). Exploratory data analysis. Methods, 2, 131-160.

Zhou, D., Yanagida, J., Chakravorty, U. and Leung, P. (1997). Estimating economic impacts from tourism. Annals of Tourism Research, 24(1), pp.76-89.

Zhuang, X., Yao, Y., & Li, J. (. (2019). Socio Cultural Impacts of Tourism on Residents of. Sustainability, 1-18. doi:10.3390/su11030840

Zillinger, M. (2007). Tourist routes: A time-geographical approach on German car-tourists in Sweden. Tourism Geographies, 9(1), 64-83.

A. Repository

The GitHub repository contains all the datasets as they are fed into the script to create the dashboard. It also holds all scripts that were used to create and manipulate the data as it is presented in the concept dashboard. Please refer to the readme.md for the structure of the repository. You can find the repository through:
https://github.com/futureplant/tourismdashboard/

Scan me

B. Relevant data list of Amsterdam open data

Data	Description
'Werk en inkomen (wijken)' - (jobs and incomes per neighbourhood)	Shows data on the different jobs and incomes that people have. Also includes data about possible allowances people receive. This is relevant because there might be a difference in how people with allowances have a different view on tourism. Also it shows how many people are directly working in the tourism sector. Also because people might get economic benefits in areas where more tourists are. Literature research: It is stated that tourism has a positive effect on the economics of cities and countries as well as the employment rates. (Zhou et al., 1997)
'Cultuur en monumenten' - (culture and monuments)	Shows the touristic places in the city. This can explain tourist behaviour and point towards possible hotspots of tourists.
'Op- en afstapplaatsen vaartuigen & ligplaatsen passagiersvaart' (Places to hop on and off boats and the anchorage place of passenger boats)	Places where the canal cruises pass start and stop are big hotspots where tourists gather are places where the canal cruises pass start and stop. Therefore it is relevant to look at these places.
'Bevolking - stand van de bevolking' (Population, civil status)	This dataset contains information about the people of Amsterdam, their ethnicity etc. It is usable because it might explain the sentiment a neighbourhood has towards tourists. Literature research: A person relates to the people that are from the same country and that speak their native language. This means that people from neighbourhoods with more different ethnicities might have different sentiments towards them. (Almeida et al., 2016)
'Economie en haven - Toerisme' (economics and harbour, tourism)	This dataset holds additional information on the hotels and places where tourists can sleep.
'Toerisme (Metropoolregio Amsterdam)'	This dataset holds additional information on the hotels and places where tourists can sleep and the duration of their stay.

‘Tellingen touringcars 2016-2018’ (countings touringcars)	The datasets holds information on where there have been sightings of touringcars and other transports used by tourists. The locations where they have been counted are mostly near big hotels. Would be very relevant if other parts of the city were also included. Literature research: The transportation used by tourists can be defined as a travel pattern that tourists relate to, independent from the tourist sites that are visited (Zillinger, 2007)
‘Opleidingsniveau’ (level of education)	Gives information of the level of education of different locations in the city. This is relevant because there can be differences in sentiment based on the level of education. Literature research: (Chen, 2003)
Events and festivals in Amsterdam	This dataset contains information on regular events and festivals in Amsterdam and the surrounding area, which is an attraction for tourists to visit this city. This data is relevant for temporal analysis the tourist existence based on when is the event held. Literature research: As the number of tourists worldwide is rising, so do their expectations and needs for specific experiences. There lays the role of events and their significance in modern tourism. Events and festivals have no direct bearing on the development of tourism in the region, but they contribute to raising the satisfaction connected with staying in the region (Panfiluk, 2015).
Cultural-historical values	This dataset gives a brief overview on traces, objects, patterns, and structures that form a part of living environment and gives an impression of historical development. This relevant information to start building sustainable tourism in Amsterdam while maintaining the cultural historical. Literature research: No doubt that objects of historical and cultural heritage, being an important asset of the cities, make a profit and significantly influence the economic development (Ismagilova, 2015)
Open culture data	This is a digital representation of collectables, items and/or knowledge and information from cultural institutions and initiatives Literature research: There is an urge for cultural institutions to open up control of their data, this can be an opportunity to show how cultural

	material can contribute to innovation and how it can be a driver of new developments. Everyone can consult, spread and use Open Culture Data. Open culture data makes a clear distinction between content and metadata (Baltussen, 2015)