Geospatial analysis and representation for data science – Antwerp Airbnb analysis

Fattorel Marta

1. The city

This analysis focuses on the city of Antwerp, which is located in the north of Belgium. The municipality is divided into nine districts (Antwerpen, Berchem, Berendrecht-Zandvliet-Lillo, Borgerhout, Deurne, Ekeren, Hoboken, Merksem and Wilrijk) and each of them is comprised of neighbourhoods (63 in total).

2. The data

The data used during the analysis comes from different sources. Data about the neighbourhoods of Antwerp, its population density and its museums are geojson files retrieved from the city data portal[1]. Additional data about the population come from an *.xlsx* file downloaded from the database of the city[2]. "Inside Airbnb"[3] offers data about all the Airbnb of Antwerp and, in particular, I used the *listings.csv* file. From export.hotosm.org I created and extracted the *.pbf* bounding box of Antwerp[4] in order to retrieve data about the tourist activities and other services of the city from Openstreetmap. Finally, for the analysis of the spatial autocorrelation of price I created a *.shp* file with data about the neighbourhoods of the city and its Airbnb price coming from the previous analyses.

3. Statistical information on neighbourhoods

I firstly analysed the gender composition of the population of Antwerp by comparing with a stacked barplot the percentage of men and women by neighbourhood. Overall, women and men are quite balanced, even though the number of neighbourhoods that count more women is higher than the one representing neighbourhoods with more men.

Secondly, I plotted a clorophlet map showing the average age by neighbourhood. The younger one counts a mean age of 28.7, while the older registers 48.6 as average age. Concerning the city of Antwerp, the mean age results to be 39.4.

Thirdly, I visualized the population density by neighbourhood with a web clorophlet map. The 3 most densely populated neighbourhoods are: Amandus – Atheneum, Stuivenberg and Borgerhout Intra Muros Zuid.

4. Neighbourhoods with the highest Airbnb prices

After a preliminary visualization of the spatial distribution of Airbnb with a heatmap, I applied a spatial join operation in order to identify the neighbourhoods in which each Airbnb is located. Then, I grouped the obtained dataframe by neighbourhood and I aggregated the Airbnb prices by computing the mean. Surprisingly, the "most expensive" neighbourhoods for Airbnb average price (Bezali –

[1] City data portal: https://portaal-stadantwerpen.opendata.arcgis.com/
Neighbourhoods data: https://hub.arcgis.com/datasets/aae52c04520e4f08bb187c4e68d163d1_97
Population density: https://opendata.arcgis.com/datasets/95a92f58b1ee40eb9736837b19c05579_867
[2] Statistics on the population: https://stadincijfers.antwerpen.be/databank
[3] Airbnb data: http://insideairbnb.com/get-the-data.html
[4] Bounding box of Antwerp: https://export.hotosm.org/en/v3/exports/b0f5f5b5-935b-45d6-a890-22d521e4c9de

Polder, Deurne – Zuidoost, Hoboken – Zuidoost) do not belong to the city center, to the main shopping street or are close to them. To get more insights on that result, I counted the number of Airbnb for each neighbourhood and I retrieved the most expensive Airbnb per neighbourhood. I found out that the "most expensive" neighbourhoods found before have few Airbnb with a price above the average. Instead, the most popular neighbourhoods (the ones that belong to the Antwerpen district) have lots of Airbnb: some of them are cheap, others are really expensive (more than 2000 euros). Finally, I plotted all these analyses on web maps.

5. Neighbourhoods with the greatest number of tourist activities

As described before, with the *.pbf* file and the geometry of the entire city, I was able to extract the tourist activities of Antwerp (artworks, galleries, information points). I combined them with the museums dataset provided by the city data portal and, in order to group and count the total number of tourist activities by neighbourhood, I performed another spatial join operation to detect the neighbourhood in which each activity is located. The top three entities for number of activities resulted to be the historic centre (Antwerpen - Historisch centrum), the university neighbourhood (Antwerpen – Universiteitsbuurt) and the main shopping street (Antwerpen - Theaterbuurt & Meir). This time I used a clorophlet map to visualize the results.

6. Three closest Airbnb to one of the city's museum

I decided to consider the Museum aan de Stroom (MAS), located in the Antwerpen - Eilandje neighbourhood. I used *Nominatim* as a geocoding tool to extract the latitude and longitude coordinates of the MAS museum. Then, I performed a street network analysis with the package osmnx in order to spot the shortest walking paths from each Airbnb to the museum. The three closest Airbnb resulted to be 51 and 114 meters away from MAS and are all located in its same neighbourhood (Antwerpen – Eilandje). I visualized the routes on a folium map.

7. Airbnb with more services around (300 m area) considering only the three closest to MAS

In order to retrieve the supermarkets, pharmacies and restaurants of the city, I applied the same procedure done before with the tourist activities. To compute the 800 m area around each Airbnb (I extended the area to 800 m because the 300 m threshold resulted to be too low), I had to change the coordinate reference system (from WGS84 to UTM32N) of the 3 Airbnb points as well as the dataset with all the services. This is because the UTM32N crs allows us to compute a buffer around our Airbnb in meters. Finally, I filtered the services inside the 800 m buffer for each of the three Airbnb and I found out that the closest to the MAS museum has a higher number of services nearby. However, I also realized that Openstreetmap is not really precise insofar it does not retrieve all the services of the areas.

8. Spatial autocorrelation of price

After loading the shapefile in R, the first step to test spatial autocorrelation is to build a row standardized spatial weights matrix. For consistency and robustness of the results, it is a good practise to build different ones, according to different definitions of neighbourhood relationships amongst the spatial units. In particular, I used 1 knn criterion, critical cut-off neighbourhood criterion with different cut-off levels (4, 8, 12 km) and a matrix with weights that are inverse functions of the distance among centroids. Then, I performed a global analysis of spatial autocorrelation of the average Airbnb price by computing the Moran's I index according to the different weights matrices created before and according to different assumptions on the distribution of the data (normalization, randomization, bootstrap). All p-values were coherent in showing no evidence of global spatial autocorrelation of price. I moved on by looking for local spatial patterns of autocorrelation through the Moran scatterplot and the local Moran's I index. Both methods confirmed the absence of spatial autocorrelation of price even locally. As motivated in the R file, if instead of considering the average Airbnb price, I had considered the number of Airbnb by neighbourhood, probably the results would have been different.