Elena Zordan
Student number 211663
M. Sc. Data Science, University of Trento

Geospatial analysis and representation for Data Science

BRISTOL AIRBNB ANALYSIS REPORT

Motivation

Bristol is the most populous city in South West England, with a population that reaches more than 463 thousand people. The reasons behind this city choice are both personal and practical: I personally choose this city because I was interested in it since high school, when we studied England cities in detail. I was fascinated by Bristol, and I am planning to visit it as soon as the overall situation will be restored. On the other hand, I decided to work on Bristol because the open data of the city were available and clear, especially with respect to other cities; moreover, there was no discrepancy between airbnb data and open data obtained from the city data portal.

Data

Data were gathered from two several sources. First of all, Bristol airbnb data were downloaded from "Inside Airbnb" (data under *CC0 license*); in particular:

- "listing.csv", a Csv containing all the information about the airbnbs of Bristol, including latitude and longitude for each point;
- "neighbourhood.geojson", a geojson file containing the geometries of neighbourhoods of Bristol

Second, all the data regarding insights about neighbourhood of the city (like the distribution of the population) were collected from the Bristol city data portal:

- "Wards", a folder which contains files needed to R session analysis such as shapefile (.shp), a DBase database file (.dbf) and projection file (.prj)
- "estim_pop_by_ward", a folder with the shapefile of the population by ward
- "ethnicity_by_ward", a folder with the shapefile of main ethnies by ward

- “pop_sex_by_ward.csv”, a csv of the male and female population, divided by age groups, for each neighborhood

Third, point of interests and services of the city from OpenStreetMap services, data are available under Open data Commons Open Database Licence (ODbL) by the OpenStreetMap Foundation

- Bristol.osm.pbf

Last but not least, the github repo contains two additional folders that were created throughout the analysis. These are needed in the last point of the analysis, and are executed in R. In particular the are:

- “newgdf”
- “geo_data_regression”

Analysis

The analysis is computed using *Python 3.8.1* and *R 3.6.1* on the jupyter notebook. Using the library rp2 it was possible to put all the analysis within the same jupyter notebook. All the libraries were installed by the use of a virtual environment.

The analysis required is divided into five main points:

1. Retrieve data on neighbourhoods of the city
2. (2.1) View statistical information about Bristol neighbourhoods and
   (2.2) identify which are the neighbourhoods with the highest prices in Airbnb
3. Identify the districts with the greatest number of tourist activities
4. Find the three nearest airbnb to the Red Lodge museum on walking distance
5. (5.1) Of the three hosts, pick the one that has the greatest number of services (supermarkets, pharmacies, restaurants) in an area of 300m
   (5.2) Analyze and test spatial autocorrelation of price (in R)

Moreover, for each point visualise data on maps (web mapping or not).

Overall, the analysis follows the order of the five main points, but the beginning where all the libraries needed are listed and the “Data preparation for R session”, which creates the data required by R libraries (point 5).

Code and results are commented throughout the jupyter notebook.

Among the skills acquired from the course, the one under the name of “raster concepts” is the only unexploited one. This is because I did not manage to find clean and free raster data

of the city nor a way to use raster data concepts in a useful and coherent way for the analysis. However, the simplest form of raster data consists of a matrix of cells -pixels- organized into a grid, and each cell contains information. Because of its structure, a raster is considered a continuous data and some examples are the digital elevation models (DEM), solar power cells and orthophotos. In particular, a DEM can be a DSM (Digital Surface Model) or a DTM (Digital Terrain Model).
Moreover, while the structure of raster data is simple, it is useful for a wide range of applications such as basemaps, surface maps and thematic maps.

Conclusion

Overall, the results are quite consistent with the expectations: neighbourhoods with higher population are the central and the most spatially extended (in terms of territorial area) ones (respectively: Central, Avonmouth & Lawrence Weston, Westbury-on-Trym & Henleaze and Lawrence Hill, which cover almost 18% of total population).

With respect to airbnb prices, the analysis finds out that the district with higher airbnb prices on average is Brislington East, followed by Hotwells & Harbourside and Clifton.
In light of this, the analysis computed in R shows that the autocorrelation of the price is very low and significant only within small distances: this means that each airbnb may have an influence on the nearest ones, but limited.
Moreover, the analysis of the residuals displays that, using price as dependent variable and population (by district) as predictor, the value of the observed Moran I is almost zero, and the p-value associated with this “global Moran's I for regression residuals” analysis, is greater than 0.05, leading to accepting the null hypothesis (no autocorrelation).

As regards tourism, the neighborhood with highest number of “tourist activities” is the “Central” one (n=243), followed by the nearby Hotwells & Harbourside on west (n=107) and Lawrence Hill on the east side (n=70).

Finally, some considerations about weakness of the analysis are needed, especially regarding the point 5.1 of the exercise (services within an area of 300m): I did not manage to find a way to plot all the points on a map without being confusing, hence I limited myself to the dataframe. Another limit can be found with respect to the autocorrelation analysis: maybe another dataframe containing lesser points could have achieved better results.

Resources

- http://insideairbnb.com/get-the-data.html
- https://opendata.bristol.gov.uk/pages/homepage/