# Geospatial analysis and representation for data science 2020/21

## Dall'Asen Nicola (ID 211662) - M.Sc. Data Science

**nicola.dallasen@studenti.unitn.it**

## Software and data used

### Software

The analysis was conducted using `Python` 3.8.6 with the following libraries list and versions: `geopandas` 0.8.1 `folium` 0.11.0 `contextily` 1.0.1 `geopy` 2.1.0 `matplotlib` 3.3.3 `seaborn` 0.11.1 `pyrosm` 0.6.0 `osmnx` 1.0.0 `Shapely` 1.7.1, `pyproj` 3.0.0.post1 and `rpy2` 3.4.1

For the autocorrelation and regression part the software used was `R` 4.0.3 with `rgdal` 1.5-19 and `spdep` 1.1-5

### Data

Several data sources were brought together to conduct the analysis, in particular:

- AirBnB data and neighborhoods GeoJSON from [InsideAirBnB](#) under [CC0 license](#)
- Population data for the city of Florence from the [Open Data City Portal](#) under [CC BY 3.0](#)
- Point of interest and services data from [OpenStreetMap](#) downloaded through [BBBike](#), OSM data are available under [Open Data Commons Open Database License](#) (ODbL) by the [OpenStreetMap Foundation](#)

Some snippets of the code were taken and adapted from the notebooks and files of Maurizio Napolitano and Diego Giuliani for the [course](#) under the [MIT license](#).

## Analysis for the city of Florence, Tuscany

Florence is one of the most tourists-attractive cities of Italy, given its artistic and historic relevance and the city center provides tourists with many attractions and activities. The city is composed by 5 districts, *Centro Storico* (the city center), *Campo di Marte*, *Rifredi*, *Isolotto Legnaia* and *Gavinana Galluzzo*. Even though one would expect the center to be expensive, we are going to see this is not the case.

The attractiveness is reflected by the number of hosts on AirBnB (more than 11,000), which is higher than other Italian and abroad cities (e.g. 8,000 in Venice and Naples, 6,500 in Edinburgh).

The aim of this study was to analyze and correlate AirBnB data of the city of Florence with data coming from the City Data Portal on the population and data coming from Open Street Map on number of services offered to tourists.

The code follows this structure:

- Retrieving data about AirBnB and districts
- Analysis of districts population data
- Identification of high AirBnB price areas
- Retrieval of tourist activities data in relation to the districts

- Street network analysis of AirBnB in relation to one of the city museum
- Analysis of the availability of services around the AirBnBs
- Study of the autocorrelation and regression of the prices based on demographic data

As expected from this city, most of the AirBnBs are located in the city center, where we can find almost 75% of them, tourists want to be as close as possible to the main attractions and this number confirms it.

Unexpectedly, the most expensive district is not the center, but *Gavinana Galluzzo*, the district situated in the South-East, but the reason is simple. This neighborhood is outside the city, and has a lot of green areas around small villages, and the AirBnB structures situated here are luxury *villas* and their price can reach 8,000 euros per night.

Apart from the price consideration, Florence's attractive sites are almost entirely contained in the city center, while the rest of the neighborhoods have almost no points of interest, out of the 500 points marked as tourism on OpenStreetMap, 410 are located in the city center, while *Gavinana Galluzzo* holds 16 of these activities, suggesting a different type of customers for that zone.

Taking Palazzo Vecchio as a reference museum we have seen that the nearest AirBnBs are extremely close, so much that OpenStreetMap calculates a 0 meters walk-distance to reach it, meaning they get mapped to the same point on the street graph. Similar results have been noticed around the Uffizi gallery indicating that many structures try to be more appealing to tourists by being close to the main museum and most famous city points.

The city center is also rich in services, and our study showed that from the AirBnBs located around Palazzo Vecchio you can reach in an area of 300 meters almost 100 different services, including pharmacies, supermarkets and restaurants.

Lastly we studied the autocorrelation and regression of price in the various districts but we found that there is almost none correlation and that therefore neighboring districts do not affect the price of AirBnBs, and little statistical significance has been found when trying to correlate the price to other demographical data such as population, number of structures and services, the price is not determined by these factors.

## Conclusions

The study confirmed that Florence is one of the most attractive cities in Italy for tourists, and this is clear by the number of structures and attractions in the city center, the most famous part of the city known worldwide. A surprising phenomenon emerged from the study, that the most expensive area of Florence is not the center, as one would expect, but a district to the South-East, *Gavinana Galluzzo*, a district rich in green areas and luxury villas.

The final part of the study showed that the mean price presents almost no autocorrelation, so every district is independent of the others when determining the AirBnB price and also independent from other factors as population and number of POIs, as shown by the case of *Gavinana*, where there is little population, a low number of POIs but a higher mean price.

This last part has some clear limitations due to the low number of districts in the city, and therefore providing low statistical significance.

## Considerations on the availability of data

It is appreciable to notice more and more Open Data available from local cities and from government institutions and, in particular for the Florence case, we appreciated the use of open format such as CSV (although some closed source formats are available, e.g. XLS) and the use of permissive licenses, in this case CC BY. On the other hand, a noticeable drawback is the

interaction with the City Data Portal and the explanation of the data, which is left to the final user based only on the name of the dataset and the manual inspection of the latter.

A final note has to refer to raster data, which in the end have not been used in these analysis because of the interactions with the *GEOscopio Fototeca* of Tuscany, which does not provide an accessible way to download raster images; there are no clear instructions on how to download and interact with the downloaded data.

The service is provided also via WMS, but the images provided through this service cover the whole Tuscany, which was not suitable for our analysis.