2 Data

2.1 Data sources

In order to develop our analysis, we will be leveraging location data via the free access API provided by *Foursquare*. This will enable us to cluster Venice's neighbourhoods - Venice's main islands and the areas historically called "Sestieri" - on the basis of the typologies of venues users have reviewed over the years. Such an homogenous source of data will be integrated with a more diverse series of datasets, which will be functional to the geocoding of the neighbourhood themselves, as well as to providing useful insights into features such as population density and tourism demand.

Considering the peculiarity of the analysis, some datasets were manually assembled into tabular form, while other were web-scraped or directly downloaded by open access platforms such as *Inside Airbnb*.

The sourced data set - in order of appearance within the analysis - are the ones that follow:

- User-defined dataset listing Venice's neighbourhoods and respective geographical coordinates;
- User-defined dataset organising Venice's neighbourhoods on the basis of their population density;
- Dataset directly downloaded from the platform *Inside Aribnb*, which catalogues all *Airbnb* offerings in the city, providing useful insights such as price per night and location information;
- Data fetched via the *Foursquare* API, containing information about the categories of the food services offered in each Venetian neighbourhood;
- Publicly available web-scraped data on rental cost for commercial actives in each Venetian neighbourhood;

Finally, it is worth to mention that, when closely examined, this collection of data is flawed in some respect. *Foursquare* data itself, for instance, is sometimes too granular to be throughly functional, as it seldom results in a series of noisy, redundant entries - e.g. "Veneto Restaurant", "Italian Restaurant" ,"Mediterranean", and "Local Restaurant" all labelled under different categories.

Nevertheless, it cannot underestimated how the aforementioned collection allows non-enterprise users to build an accurate and commercial viable analysis as the one that follows.