

# Data

In order to develop our analysis, we will be leveraging location data via a free access API provided by the world famous service *Foursquare*. This will enable us to **cluster Venice's neighbourhoods** - and those areas historically called "Sestieri" - into **4 groups**, on the basis of the typologies of venues *Foursquare* users have reviewed over the years. Of course, 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 needed to be manually assembled into tabular form, while other were web-scraped or directly downloaded by open access platforms such as *Inside Airbnb*.

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 specific to be thoroughly functional, as, being *Foursquare* users allowed to create their own venue labels, it often results in a series of noisy, redundant entries - e.g. "Veneto Restaurant", "Italian Restaurant", "Mediterranean Restaurant", and "Local Restaurant" are 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.