Data

Description

Foursquare

Its API will be used to obtain data related to the first criterion in the selection of a neighbourhood to open a restaurant. It will be fed the coordinates of each neighbourhood, returning a json file. From it the relevant information will be extracted, presenting mainly venues, their categories and their coordinates.

ArcGIS

Their API will also be used, feeding it with the name of a location in order to obtain its geographical coordinates. In this study in particular, it will be used mainly to get the coordinates of Buenos Aires and all its neighbourhoods.

Buenos Aires – Datasets

Additionally to the already mentioned sources datasets provided by the government of the City of Buenos Aires.

Neighbourhoods

https://data.buenosaires.gob.ar/dataset/barrios/archivo/juqdkmgo-191-resource (in Spanish).

In this dataset different fields can be found inside the .csv file, but the site does not give any explanation about them. Upon inspection of these fields, it is possible to gain some understanding and to infer the following:

- WKT: this is a string field and contains polygons in wkt format, which is widely used and, in this dataset, contains the shape of each neighbourhood.
- Barrio: this is another string field, containing, in this case, the name of the neighbourhood.
- Comuna: this is an integer field, containing the number of the commune to which the neighbourhood belongs.
- Perimetro: this is a float field, containing the perimeter of the neighbourhood, measured in meters.
- Area: this is a float field, containing the area of the neighbourhood, measured in square meters.

Population

https://data.buenosaires.gob.ar/dataset/estructura-demografica/archivo/c44be985-8d7f-4aa4-972e-a7f8f0b796dc (in Spanish)

The explanation given about the fields in this csv file is quite straight-forward: “BARRIO” is a string containing the name of the neighbourhood and “POBLACION” is an integer with the

population of said neighbourhood. Unfortunately, this is a dataset with information from 2010, but it's the best made available by the City's government.

Bus Stops

https://data.buenosaires.gob.ar/dataset/colectivos-gtfs (in Spanish).

This dataset includes a number of files compressed in a .zip file. The one of interest for this study is "stops.txt", which internally is actually a .csv file. The site gives no explanation about the dataset, but the fields "stop_lat", "stop_lon" (both floats) will prove to be the latitude and longitude coordinates of each bus stop. It must be noticed that the listed stops belong not only to the City of Buenos Aires, but also to Buenos Aires Province. This will have to be considered when this set is cleaned and prepared.

Train Stations

https://data.buenosaires.gob.ar/dataset/estaciones-ferrocarril/archivo/juqdkmgo-1021-resource (in Spanish).

In this case, the site explains the dataset, presented in .csv format. For this study, the fields "long", "lat" (both floats) and "barrio" (string) contain the relevant information: longitude, latitude and neighbourhood, correspondingly. As in the previous case, the stations belong to both the City and the Province. However, those belonging to the Province have no listed neighbourhood, which simplifies the cleaning and preparation.

Subway Stations

https://data.buenosaires.gob.ar/dataset/subte-estaciones/archivo/juqdkmgo-1992-resource (in Spanish).

Once again, the site does not explain this dataset, which is comprised of a .csv file. From the name of its fields and their inspection it can be inferred that "long" and "lat" (both floats), are the geographical coordinates of each subway station. This is the relevant information to extract for this study. All subway stations are located in the City, as it is the only place in Argentine with such a transportation system.

Tram Stations

https://data.buenosaires.gob.ar/dataset/premetro/archivo/juqdkmgo-1651-resource (in Spanish).

No explanation can be found in the site about this dataset. Analogously to previous cases, upon inspection of the .csv file, the fields "long" and "lat" (both floats) can be recognized and the geographical coordinates of the tram stations. These are the only relevant fields for this study. Also, all tram stations are found in the City only.

Usage

The starting point will be the Neighbourhood dataset, because after some preparation of this set, the names of the neighbourhoods will be fed to the ArcGIS API in order to obtain the coordinates of each of them. These coordinates will be appended to this dataset.

The datasets Bus Stops, Train Stations, Subway Stations and Tram Stations will be cleaned and prepared with the aim of obtaining only the geographical coordinates of each stop/station reported in them. Later, using these datasets and the field WKT of the Neighbourhood ones, each stop/station will be assigned to a neighbourhood. Consequently, all stops/stations outside of the limits of the City of Buenos Aires will be dropped. Finally, each of these datasets related to the public transportation system will be turned into dataframes counting how many stops/stations there are in each neighbourhood.

Finally, all the dataframes will be combined in a single dataframe in which each row will represent a neighbourhood containing all the prepared information described up to this point and allowing the analysis which ultimately will help answer the question stated in the initial parts of this study.