

Using Data Science to identify the optimal location for a new Italian restaurant in Sydney

Data

To solve this problem, we will need the following data:

- List of neighborhoods in Melbourne. This defines the scope of the project, as confined to the city of Melbourne, Australia
- Latitude and Longitude co-ordinates of these neighborhoods. We will need these to plot the map and also to get the venue data
- Venue data, related to restaurants, particularly Italian restaurants. We will use this data to perform clustering and other analysis on the neighborhoods

Sources and extraction of Data

- We will use the data from the Australian government datasets located at Data.com.au which is a geojson file containing the administrative boundaries of Sydney. We can extract the district names from there as well as any map boundaries if required.
(there is no example of data i can provide as it's just a geojson file from which the names of the districts can be extracted with a single line of code. It does not require any cleaning or wrangling)
- Getting latitude and longitude data of these neighborhoods via geocoder package. This is straight forward and doesn't require any further explanation. You pass a line through the library with a string and it returns the latitude/longitude
- Using foursquare API to get details of venues in these neighborhood. We will be extracting all the closest venues to a geographical location and categorize them into restaurants etc and extract the Italian restaurants in them.

Data cleaning : All the data is extracted cleanly from above sources, so there is no need to wrangle/clean/process any data for this problem