# Battle of Neighbourhoods

## Introduction/Background

Mrs A has decided to open a new Italian restaurant in London, having moved from Italy but she has not yet decided on the location. Mrs A wants to find the neighbourhood with successful existing restaurants to ensure there is high enough footfall, however she does not want there to be too much direct competition from other nearby Italian restaurants.

This data science problem will be addressed with Mrs A as the target audience to help her decide where the best location for her new Italian restaurant will be.

## Data

There were multiple factors to consider in the approach to this problem. These included the number of similar restaurants in the immediate area (I have used a 500 metre radius), the success of other food businesses and the affluence of the area in general.

A recent report found that "full-service restaurant revenues are directly correlated with the available level of customers' disposable income". (*Parsa, H. & Kreeger, Jeff & van der Rest, Jean-Pierre & Xie, Karen & Lamb, Jackson. (2019). Why Restaurants Fail? Part V: Role of Economic Factors, Risk, Density, Location, Cuisine, Health Code Violations and GIS Factors. International Journal of Hospitality & Tourism Administration. 10.1080/15256480.2019.1598908.*).

This should be a contributing factor in deciding upon the most suitable neighbourhood for a new Italian restaurant, therefore, I brought in an external dataset showing the regional gross disposable household income (GDHI) in London from the gov website (<https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/datasets/regionalgrossdisposablehouseholdincomebylocalauthoritiesbynuts1region>) to identify the borough with the highest GDHI.

Other data sources used in the analysis include:

* a list of postcodes and borough names obtained from <https://www.doogal.co.uk/london_postcodes.php>
* Geopy library (used to obtain coordinates)
* Foursquare API calls made to obtain venue details

## Methodology

The first dataset I called into the analysis was a list of postcodes in each borough in London (obtained from <https://www.doogal.co.uk/london_postcodes.php>). This dataset did include coordinates for each postcode but as there were thousands of postcodes it made more sense to obtain one set of coordinates for each ward using the geopy library.

Next, I imported the dataset showing regional gross disposable household income (GDHI) in London. As this dataset only covered the period up to the end of 2018 I decided to take an average of the last 5 years’ figures for each region and take the highest of these to narrow the search down to one region. This turned out to be Kensington and Chelsea.

The next step was to use the geopy library to obtain the coordinates for each ward in the region of Kensington and Chelsea. I plotted the coordinates using Folium to check that they were correct and what the density of wards was.

To obtain details of all restaurants in the borough I made an API call to Foursquare. As the aim was to identify a suitable area where there are no Italian restaurants within a 500 metre radius but where there are other highly rated restaurants in the vicinity, I listed all the venue types in the borough and then filtered the dataframe to just those containing the string ‘restaurant’ in the venue category column.

I then used one hot encoding to convert the categorical data into numerical, allowing for further statistical testing. To see how common each type of restaurant is in each neighbourhood, I took the mean of each restaurant type for each neighbourhood and then listed the top 10 most common.

As a result, I managed to narrow the possible locations down to two neighbourhoods, Dalgarno and Holland, neither of which have Italian restaurants as a common venue.

Each of the two neighbourhoods have two restaurants so to choose between them I obtained the average rating for each of the restaurants from Foursquare using the ‘venues’ endpoint.

## Results

One restaurant in Dalgarno has an average rating of 7.7 but the second has had no ratings, whereas the two restaurants in Holland have ratings of 7.9 and 7.5. This suggests that Holland would be the preferable area to open the restaurant in.

## Conclusion

As a result of this analysis I would recommend that Mrs A plan to open her new Italian restaurant in the neighbourhood of Holland as the area has a high gross disposable household income, there is no competition from other Italian restaurants and the general rating for other restaurants is high.