COURSERA CAPSTONE PROJECT

Predicting Best Location to open an Asian Restaurant in London

By: Vaishali

Description :

To find the perfect location to help make your operation a true success. It's important to remember that a restaurant's location is as important to its success as great food and service.

Data acquisition and cleaning:

All the required information is available at LONDON DATASTORE, which is a free and open data-sharing portal where anyone can access data relating to the city. The data is available in XLS and CSV format, which we can download and can use as-is for solving our problem.

The link for the LONDON DATASTORE, – https://data.london.gov.uk/

Machine Learning Algorithm to predict the best location to open an Asian Restaurant

Machine Learning Algorithm Simple Linear Regression is used to predict the data for Rated Value for the year 2018 for the Newham borough.

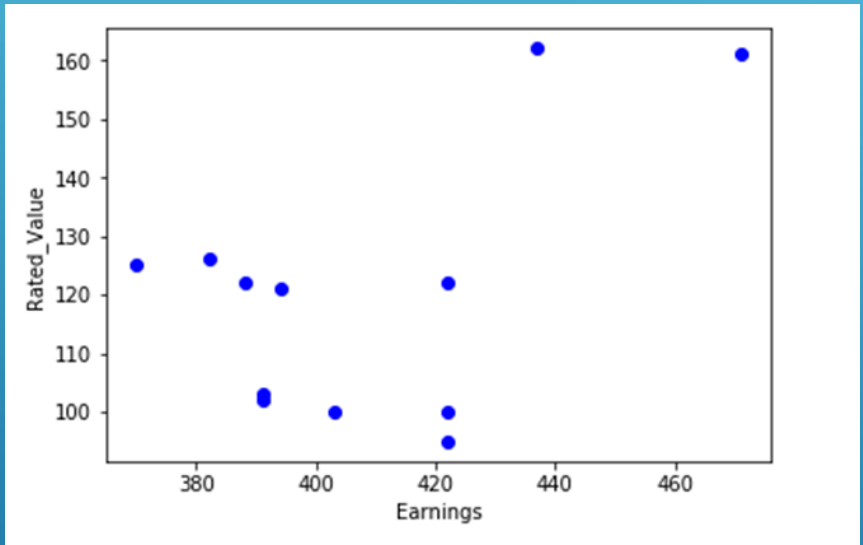


Figure 1 Earnings vs. Rated Value

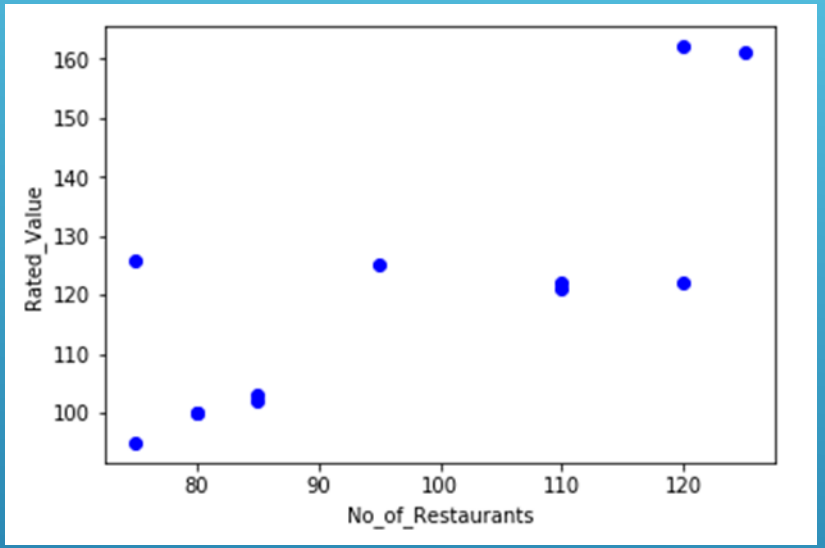
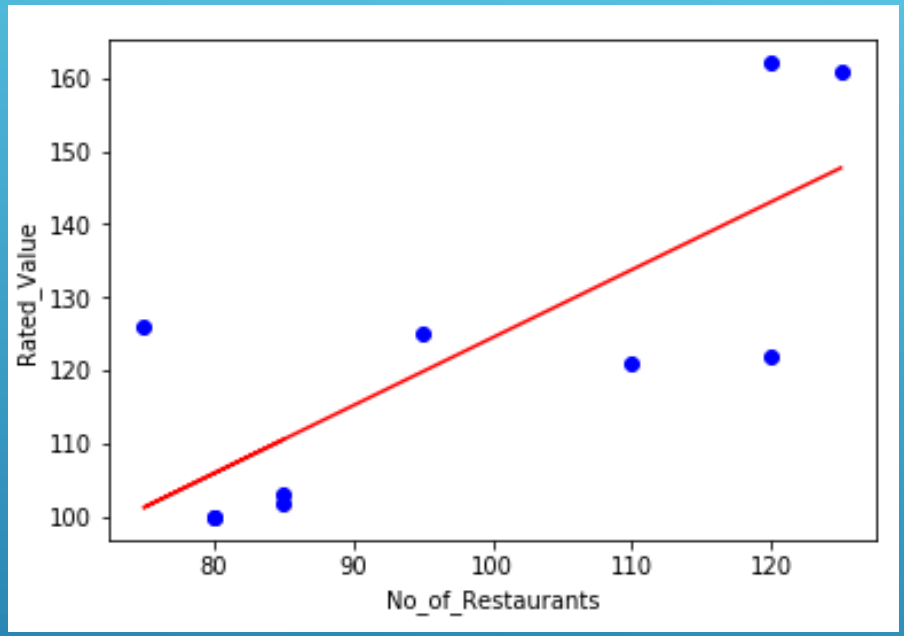


Figure 2 No of Restaurants vs. Rated Value

The train data is used to train the model and the test data is used to evaluate the model.



Conclusion: By seeing the above Scatter Plot with a best possible fit line, the Rated Value in Newham borough for the year 2018 can be predicted somewhere between 160 to 165

Segmenting and Clustering:

After the prediction of rated value per sq. meter of a retail space is completed and when we are convinced that a particular borough will be the preferred location for the restaurant, we have to get the necessary data of that borough.

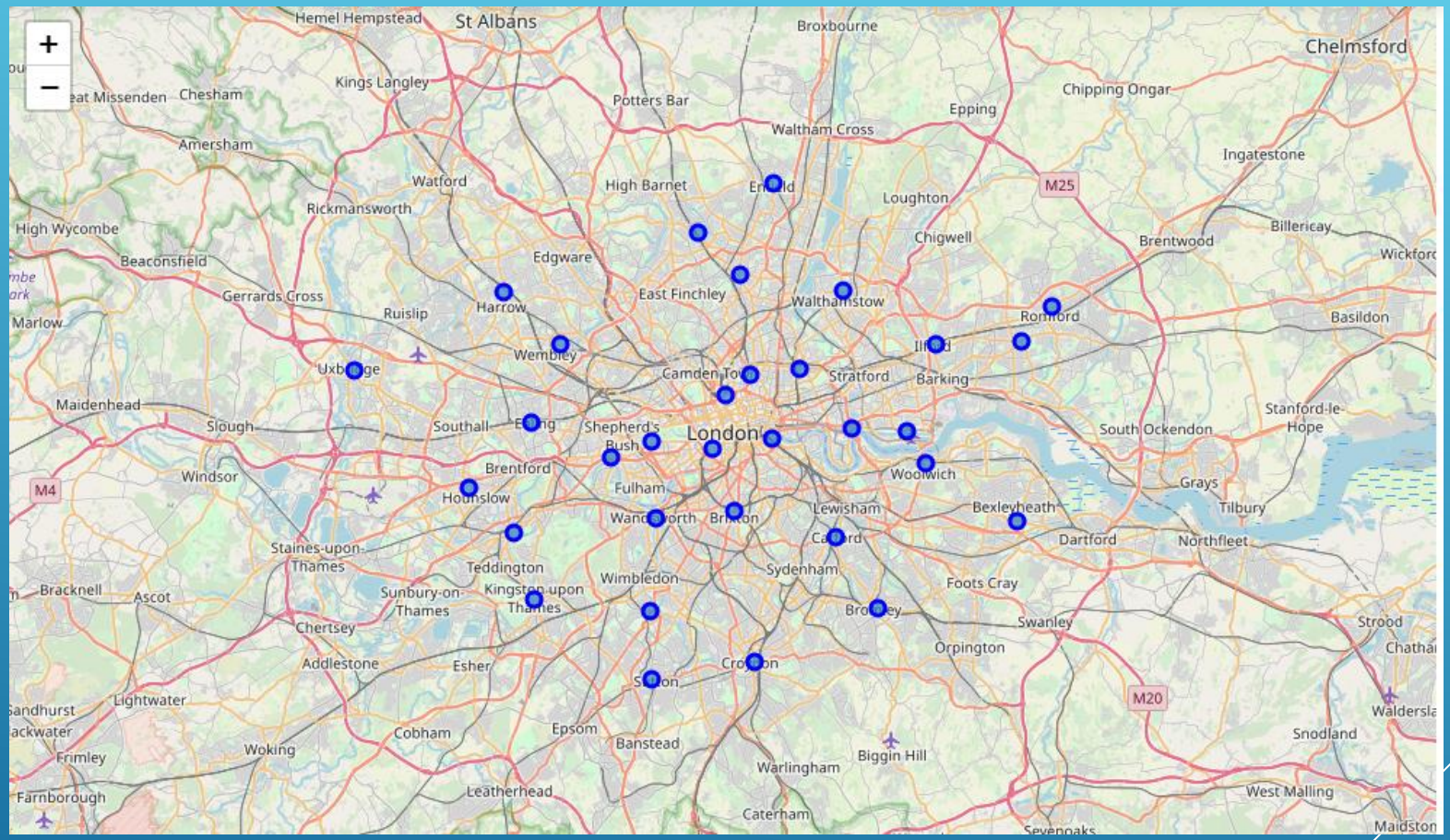
First we need to get the geo-coordinates of the borough and the geo-coordinates of the neighborhoods of the borough from the web. I have used the Wikipedia pages to get this data.

The link for List of Boroughs:
https://en.wikipedia.org/wiki/List_of_London_boroughs

The link for List of areas of London:
https://en.wikipedia.org/wiki/List_of_areas_of_London

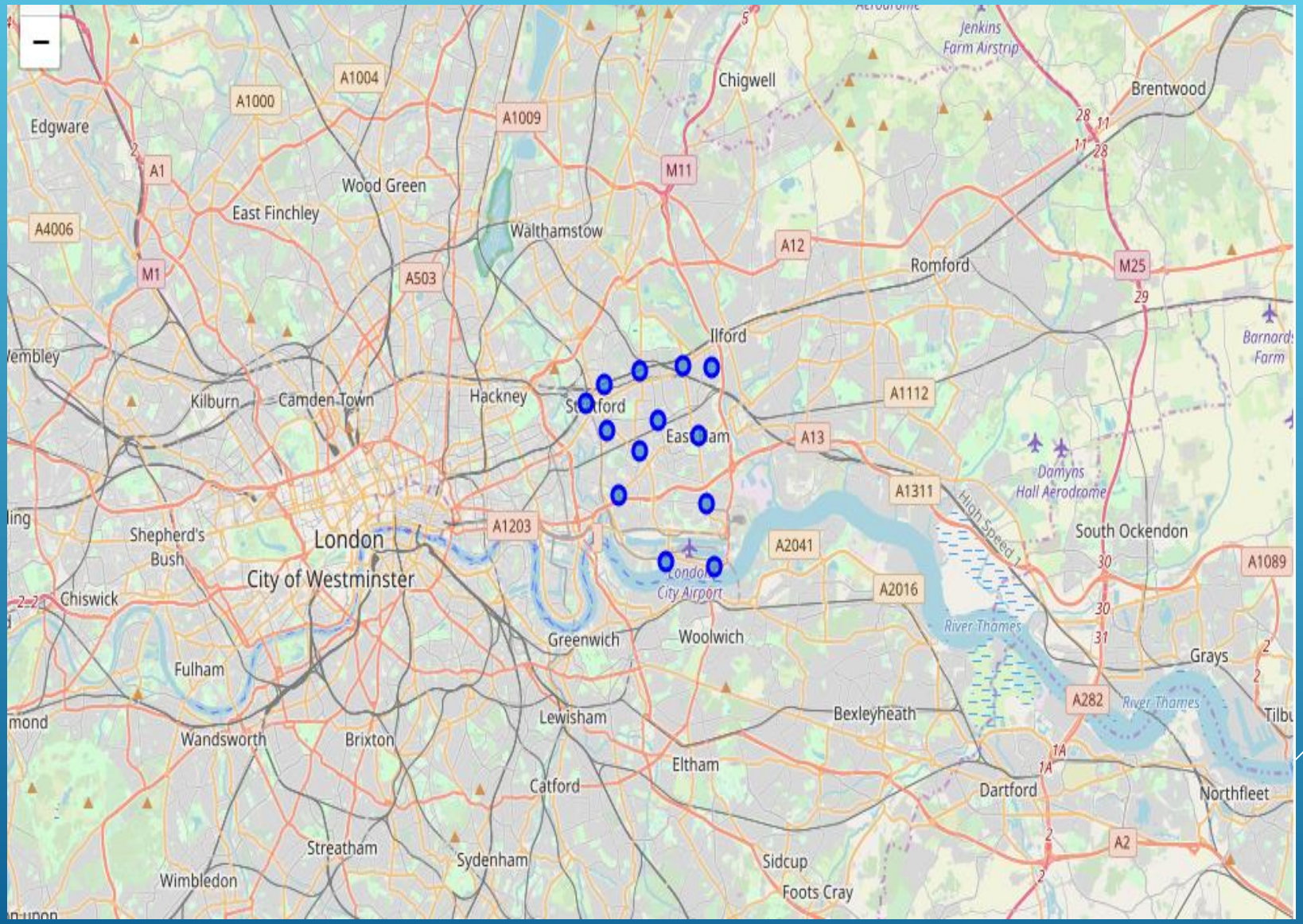
To read data from these URLs, I have used the requests, urllib and Beautiful Soup libraries of python.

Geopy module and Nominatim library is used to convert a given address into the latitude and longitude values.

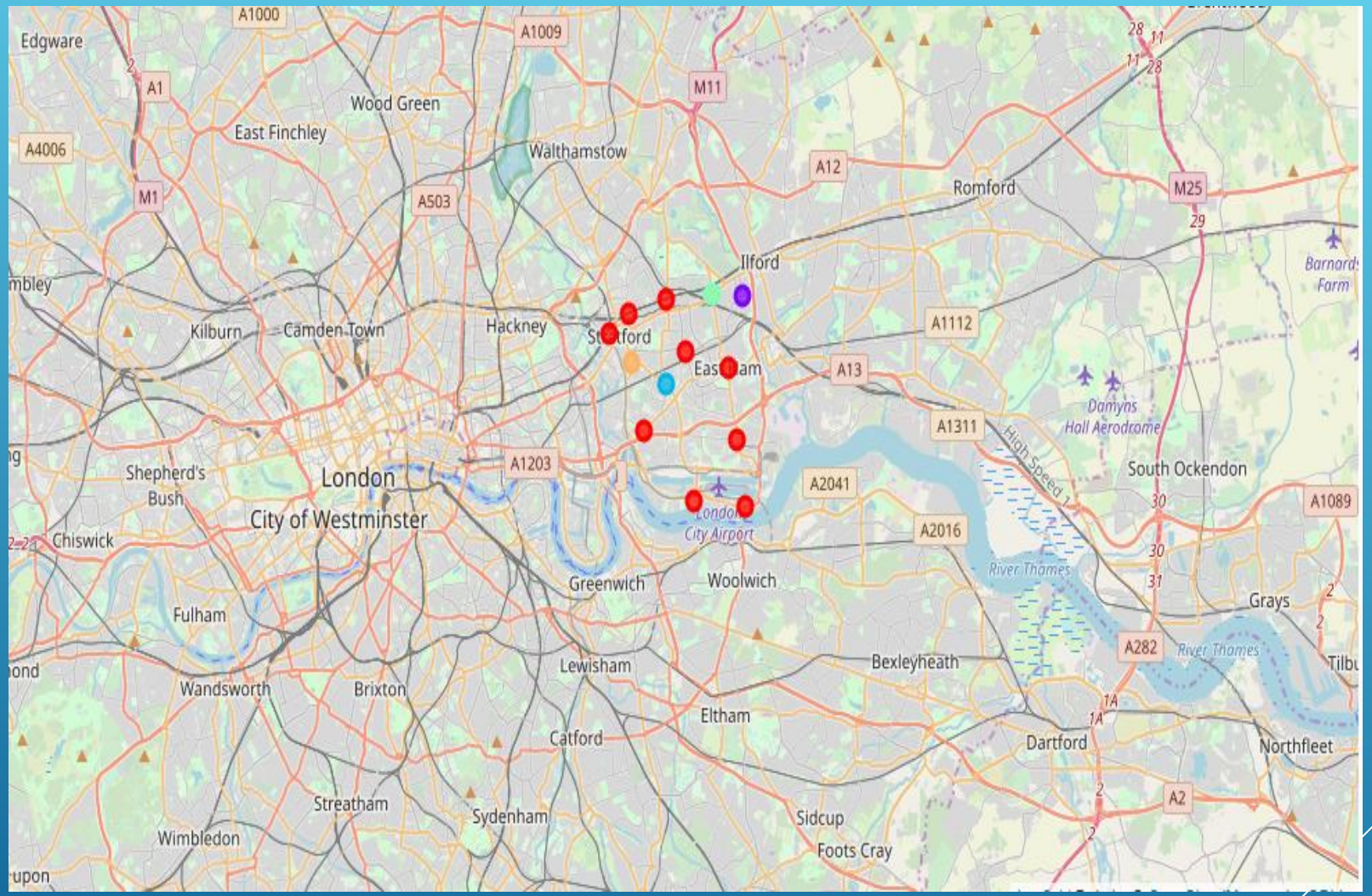


Map of London with Boroughs Superimposed on Top

Map of Newham as it has the largest population of Asians in London



Newham Neighborhood Clusters



Result:

We found that Newham borough is having highest Asian population and from the rated value data set, the prices in this borough are less compared to other boroughs. These 2 factors influenced more on my decision of choosing Newham borough as the preferred location for our restaurant. The Newham borough has 146 existing restaurants and taking this as an independent variable (X variable) I have predicted the rated value per sq. meter (dependant variable) between 160 to 165 using the Linear Regression model.

In the Segmenting and Clustering section, the neighbourhoods of Newham borough are explored, and the top 10 venues of each neighbourhood are listed.

Discussion:

My observation after doing this analysis is the model we used could have given better results, if we had huge data to train and test our model. In spite of that this model gives us a better insight for our problem and also help us to gain better results. From the clustering results our problem finds a better solution of identifying the best location for the Asian restaurant. We could explore all the neighbourhoods of the borough and could list the most common venues based on their frequency of occurrence.

Conclusion:

There is always room for improvement and hence the above solution I have provided can also be improved and the machine learning models can be trained and tested for best results depending upon the data we have.