**Coursera Capstone Project -**

**The Battle of Neighborhoods**

**Shankar Eaga**

**Dt 20 Oct 2020**

**Table of contents**

1. **The introduction : Business problem**

The project report is prepared as a part of IBM Data Science certification course offered by Coursera. The scope of the project to use the Foursquare API and Machine Learning algorithm to analyze the city of London, UK with respect to establishing a business by an Entrepreneur. The city of London is the capitol of United Kingdom, which is a 5th largest economy in the world (as per businessinsider.com), and London is very attractive for business, tourist and is financial capital of the United Kingdom. In the final project, I am going to focus on identifying 5 most famous things that Londoner’s do as neighborhood vide and highlight the potential business opportunity. Also this would help tourists to find easily where to shop, eat and sight seeing while enjoying their holidays in UK. This project has been chosen mainly keeping business owners in mind, provide sufficient evidence about business opportunities.

1. **Data collection and preparation** 
   1. Getting the information about City of London

I have used Google, Python pandas and Beautifulsoup library to extract data and grouping into data frames. The city of London has 54 neighborhood boroughs. I am going to use Foursquare API for extracting the data from the web.

Using Python library – Geopy, neighborhood longitude and latitude have been generated , and has been added to the data frame (created above), see below for neighborhood and longitude/latitude data frames. Refer Table 1 and 2 for Neighborhood and Neighborhood with longitude and latitude respectively.

* 1. Foursquare API to get venue locations.

Foursquare API has been used to get large amount of data about city of London in the json file format. Refer table 3 for most popular locations for each neighborhood within a radius of 1000 m.

* 1. I would like to use one of the simple and famous machine learning algorithm – KMeans clustering - (unsupervised machine learning algorithm) segregating the based on similar characteristics.

Table 1: Neighborhood data frame

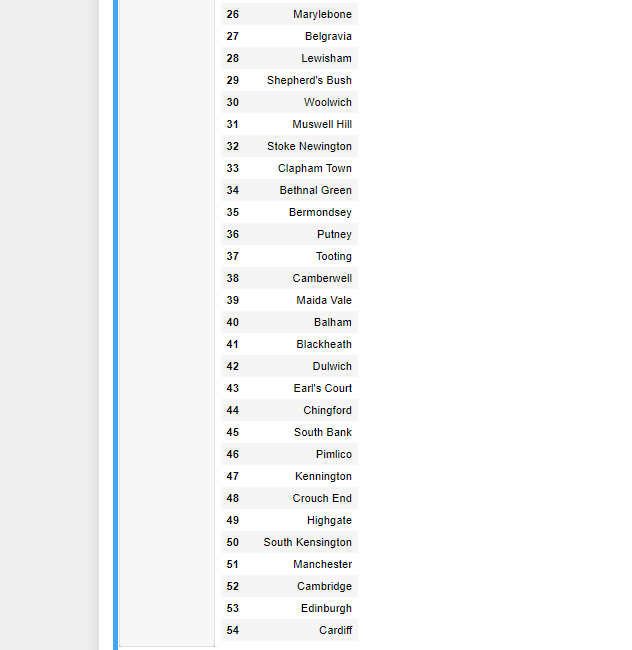
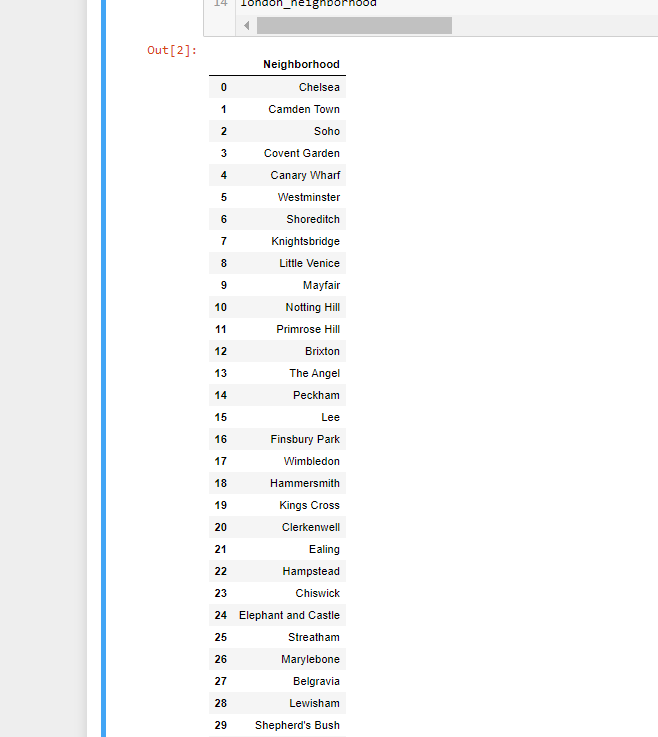
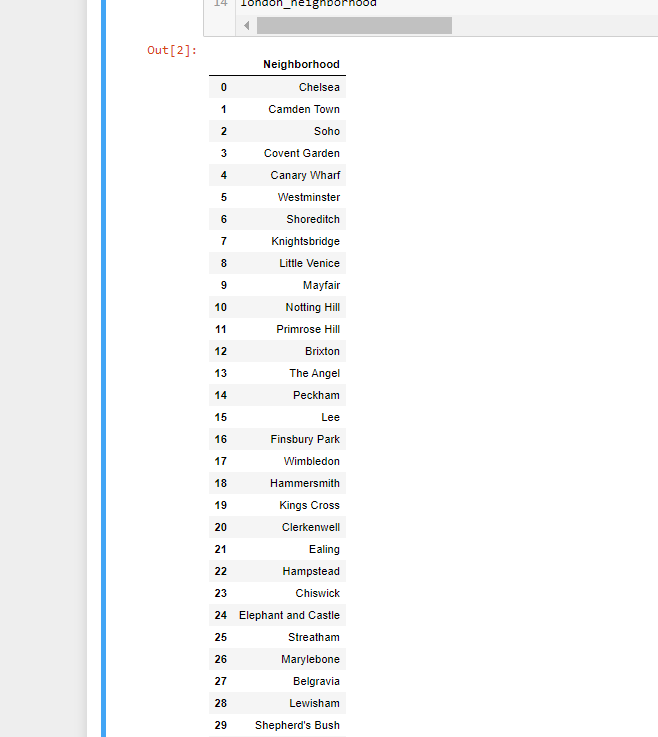
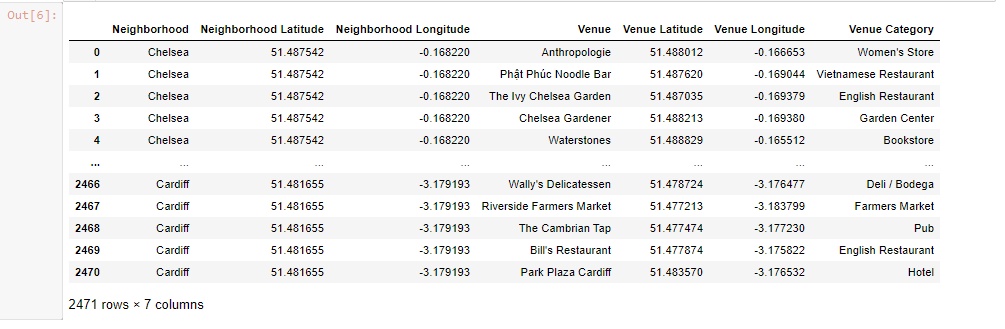


Table 2: Neighborhood with longitude and latitude data frame

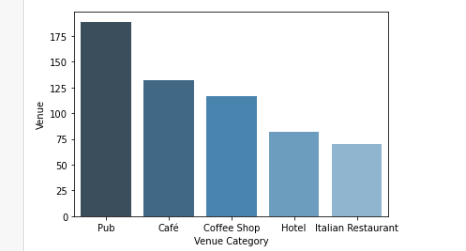


Table 3 – Formatted data from json file using Foursquare API

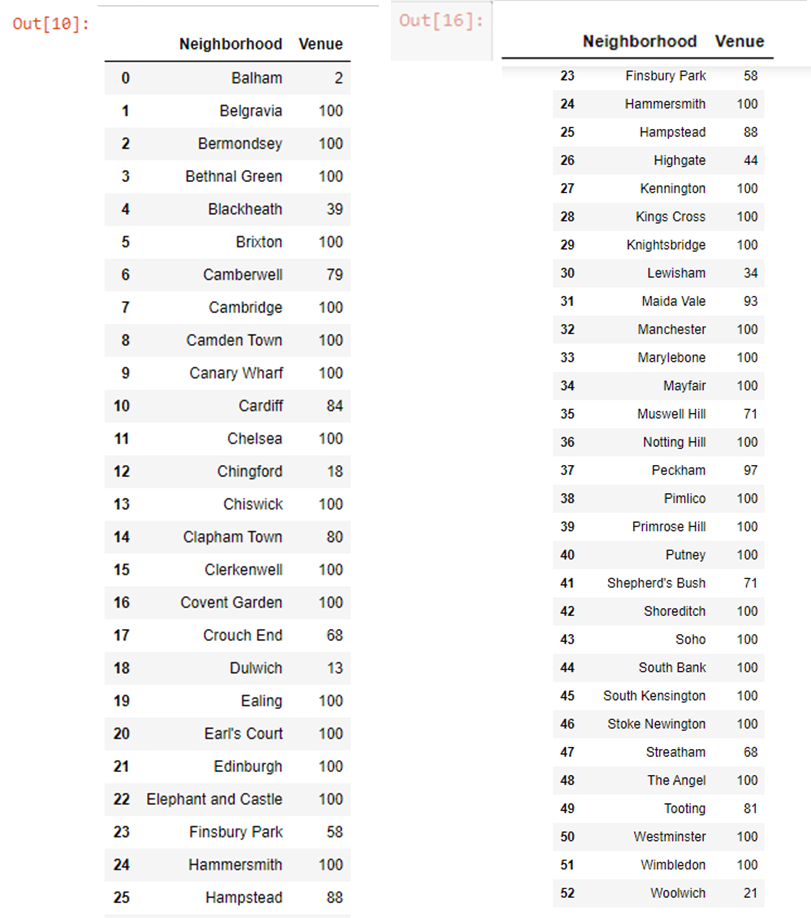


1. **Methodology** 
   1. Visualization and Data Exploration

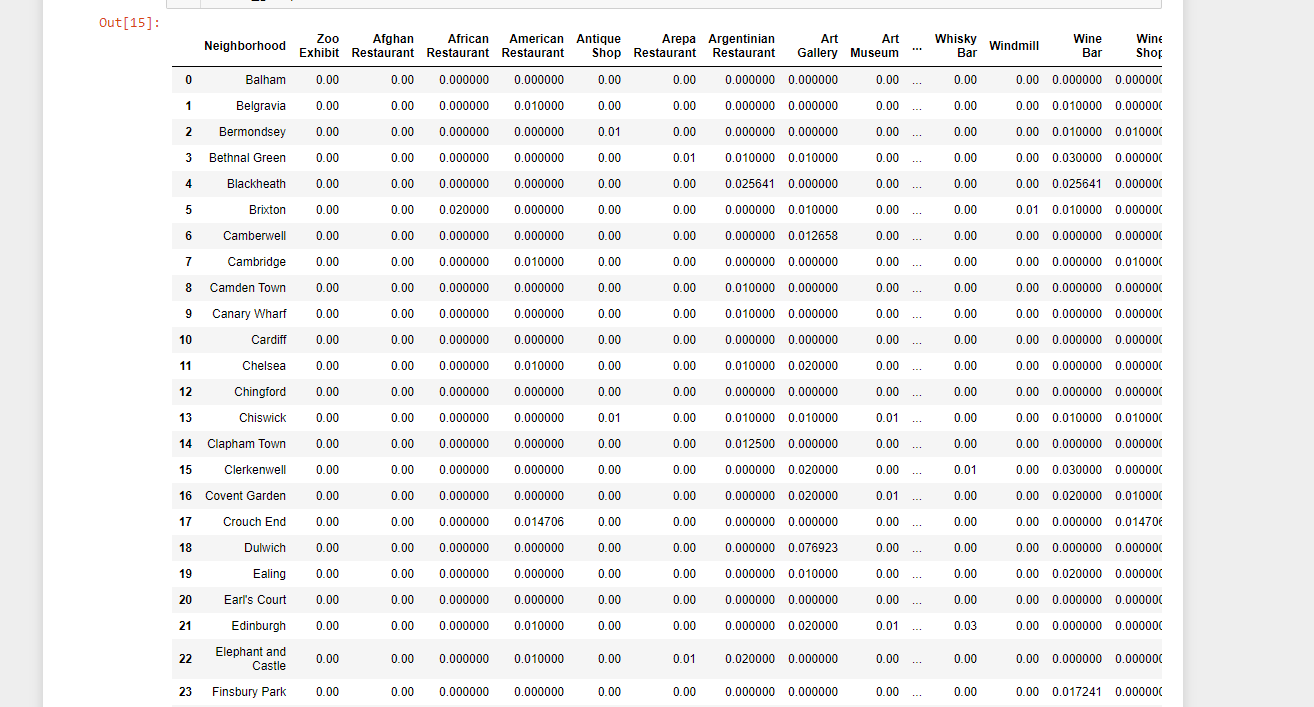
Further to collecting the data and created data frames, a quick analysis has been done to understand the data on top 5 venues in each neighborhood. See below chart for top 5 venues. It can be seen from below Bar chart that Pub is the popular venues in most of the neighborhood areas.

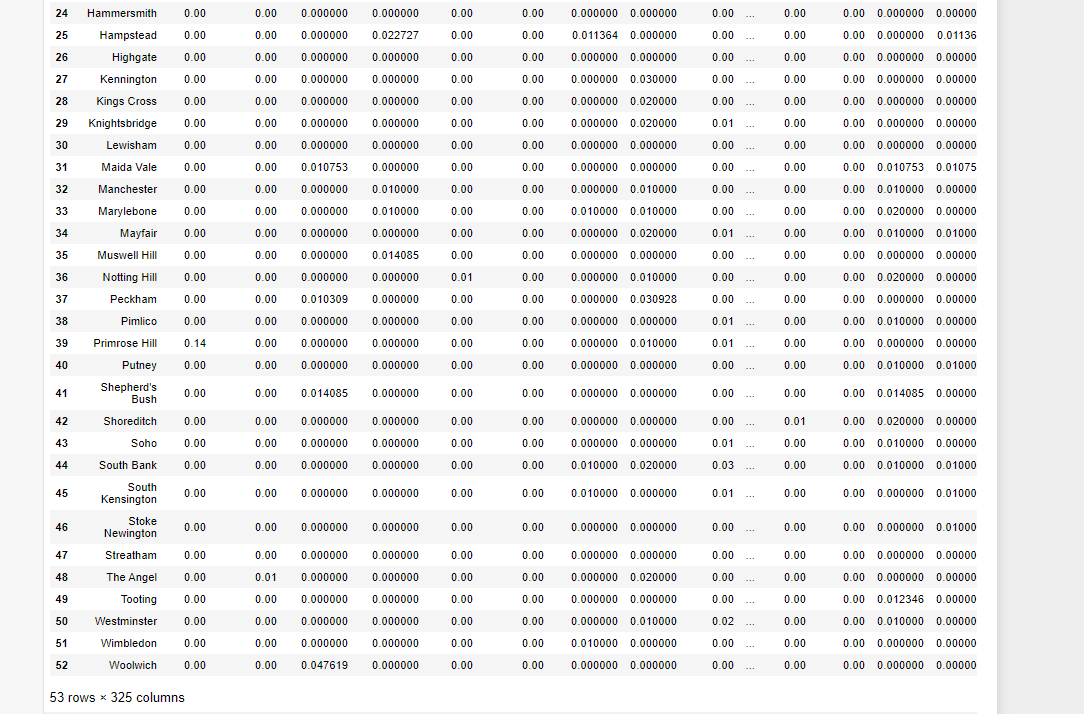


There are 325 unique venue categories in total and top 5 are Pub, Café Caffee shop, Hotel and Italian restaurant. A data frame has been created for venue , grouped by neighborhood. It can be seen that there are plenty of neighborhoods more that 100 venues (put max 100 as limit), shows that there is significant contribution to UK economy.



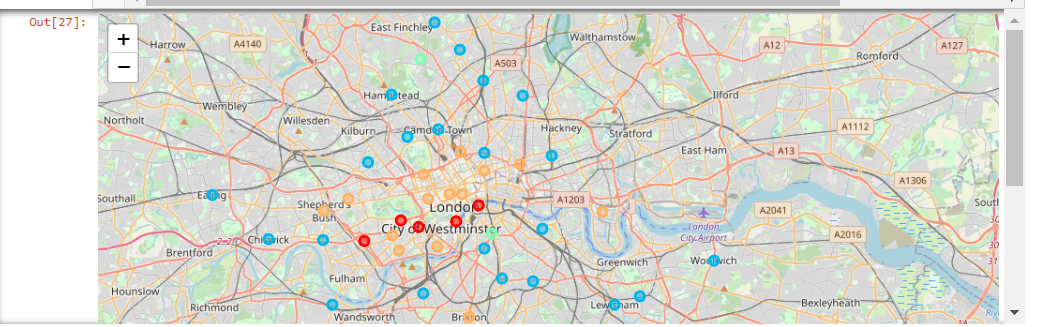
Data has been normalized used one hot encode normalization technique to reduce the impact of bias in analysis.



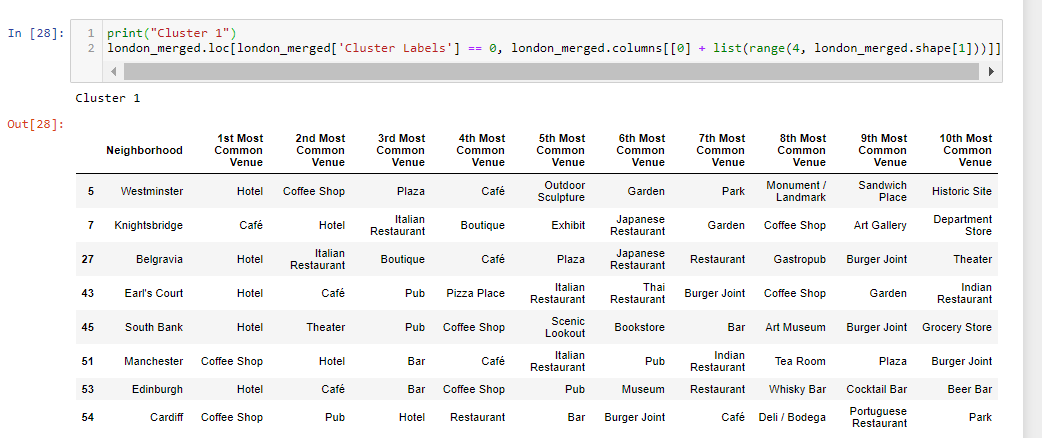


1. **Analysis – KMean clustering**

Machine learning algorithm – KMean clustering is used to cluster all neighborhood areas based on venue category and is divided into 5 clusters. By using Folium library – London map has been created with clustered neighborhood and following tables show details of the each cluster (1 to 5)



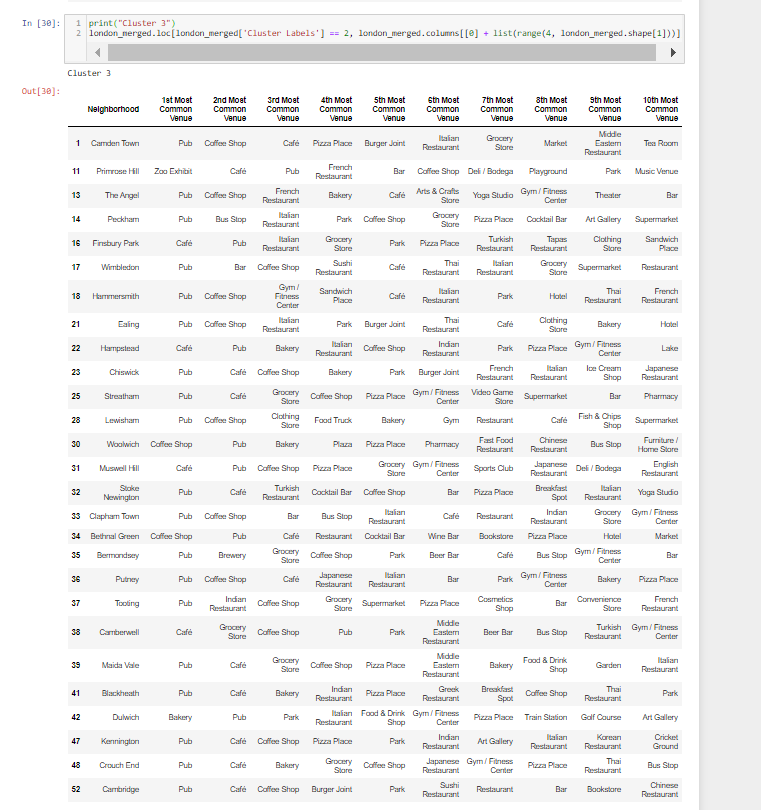
Cluster 1



Cluster 2



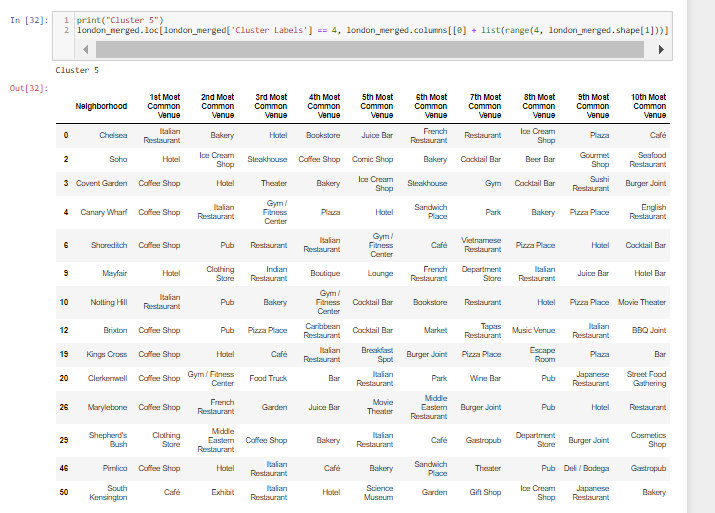
Cluster 3



Cluster 4



Cluster 5



1. **Results and Discussion**

With above K-mean clustering, we can observe that cluster 3, 4 and 5 are group of neighborhood commonly known for pub , café and coffee shop. Cluster 1 is mainly for hotel and unexpectedly cluster 2 is having unique venue – watch-shop.

By using the data collected from Foursquare, I have completed a small project to understand most popular venues in the City of London neighborhood areas. This project report could be helpful to entrepreneurs , to know what is most sought in the popular cosmopolitan city of the world, and could be used as guide to tourist to find location for pubs, to eat out and explore city.

1. **Conclusion**

There is a potential opportunities to further explore the data to know more insight of the data from business perspective like start a restaurant, hotel businesses and travel companies etc.