

Capstone Project

“The Battle of Neighborhoods”

By Iaroslav Mokroguz

03/16/2020

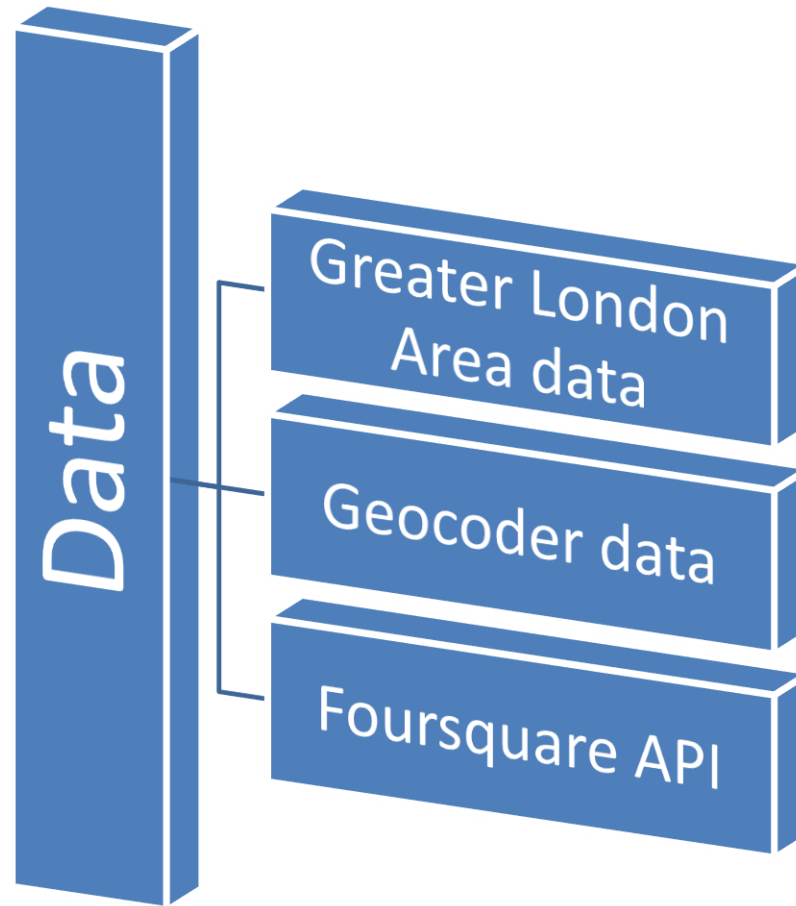
Description of the Problem

- Open new business in London is quite challenge
- Another way to increase chance to survive in competition
- We can use Foursquare API and ethnic distribution data to find our market share

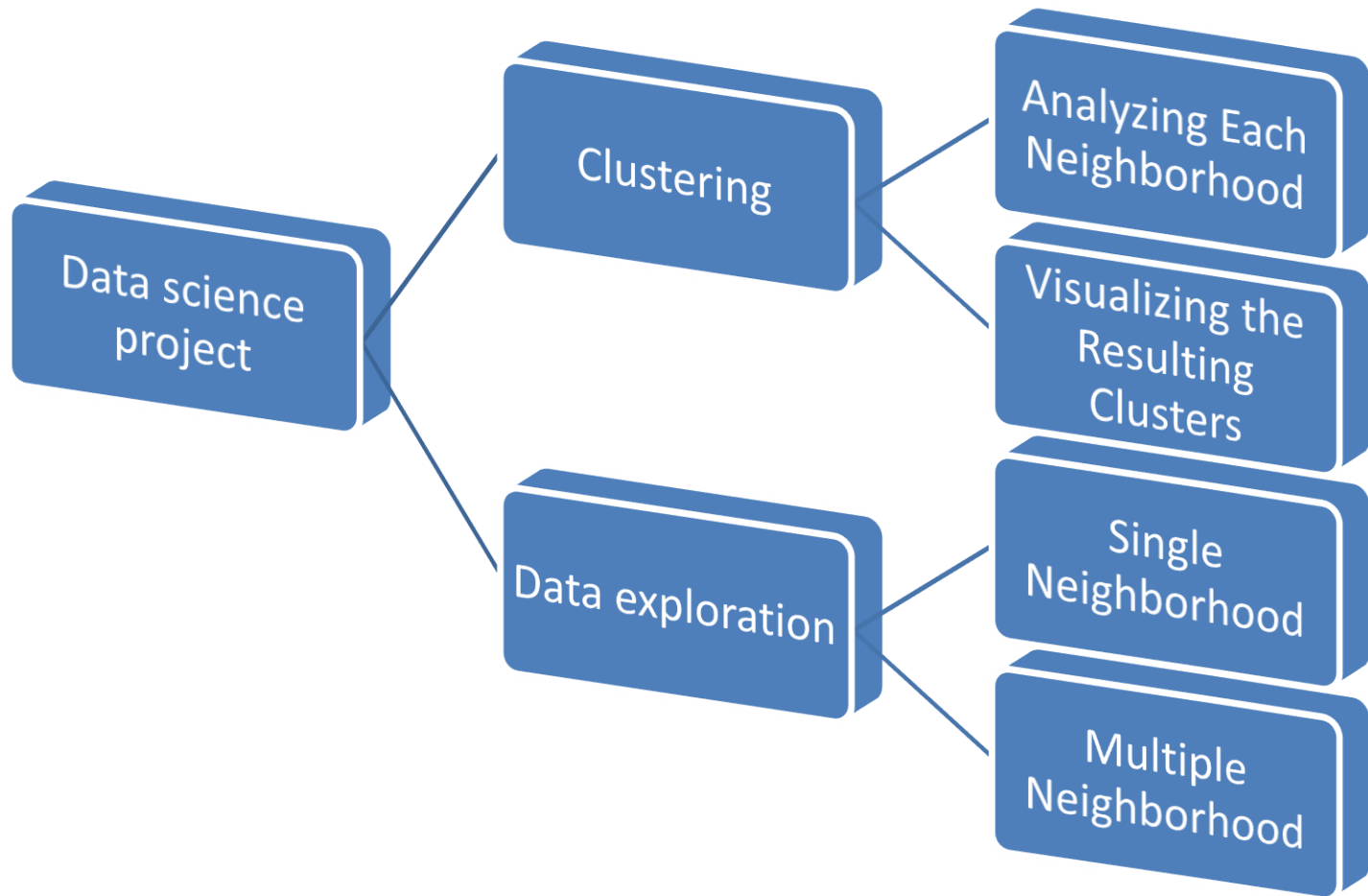
Discussion of the Background and Stakeholders

- "Ukudla meals" restaurant
- African cuisine in Europe
- Targets in London

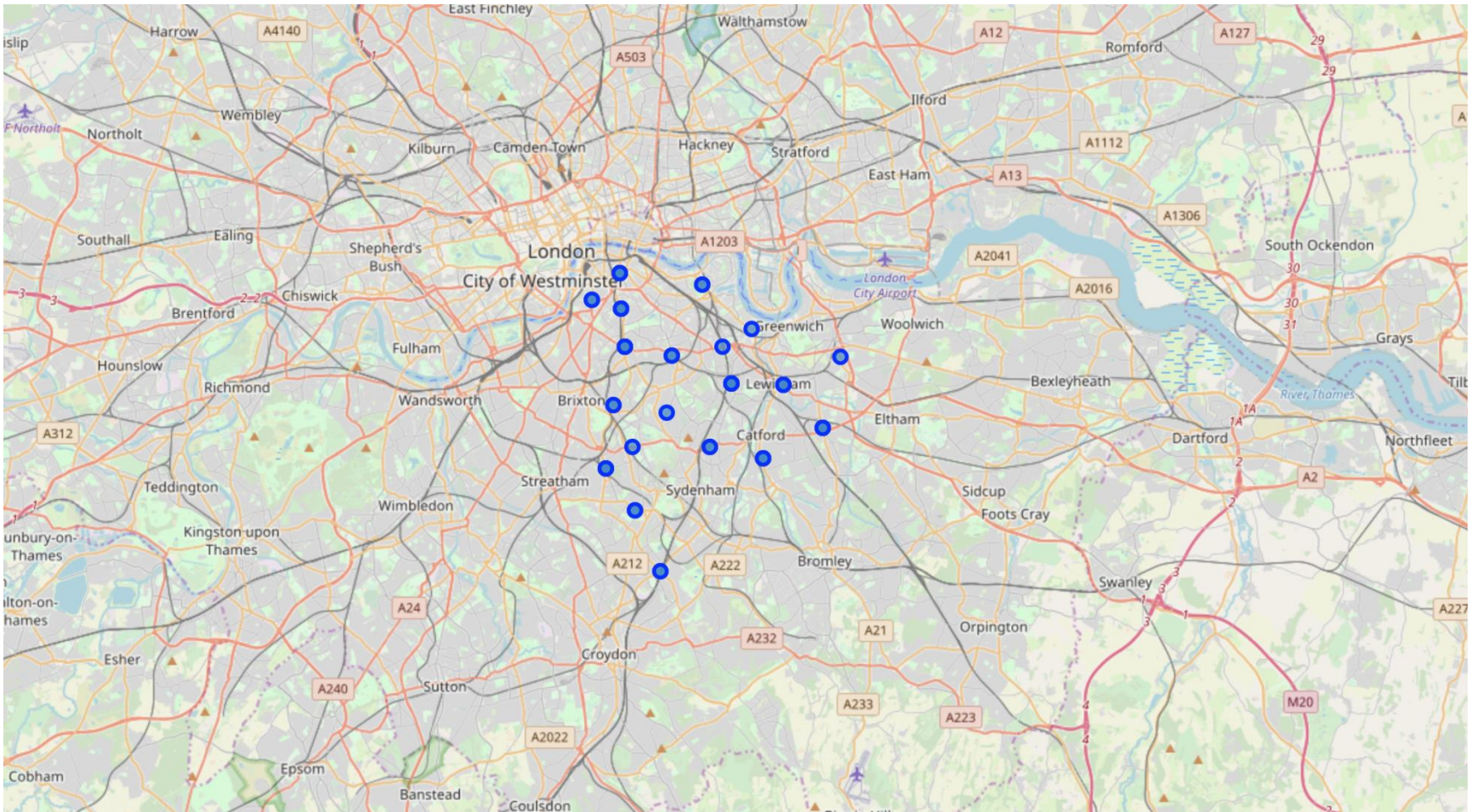
Data acquisition and cleaning



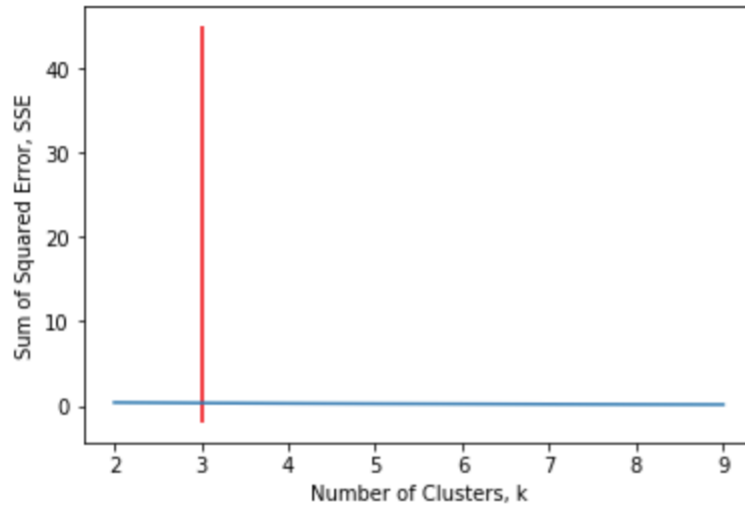
Methodology



Map Visualization



Elbow Method



Depending on the number of iteration (in this case, 500 iterations were used), the number of cluster, k is 3.

Silhouette Coefficient

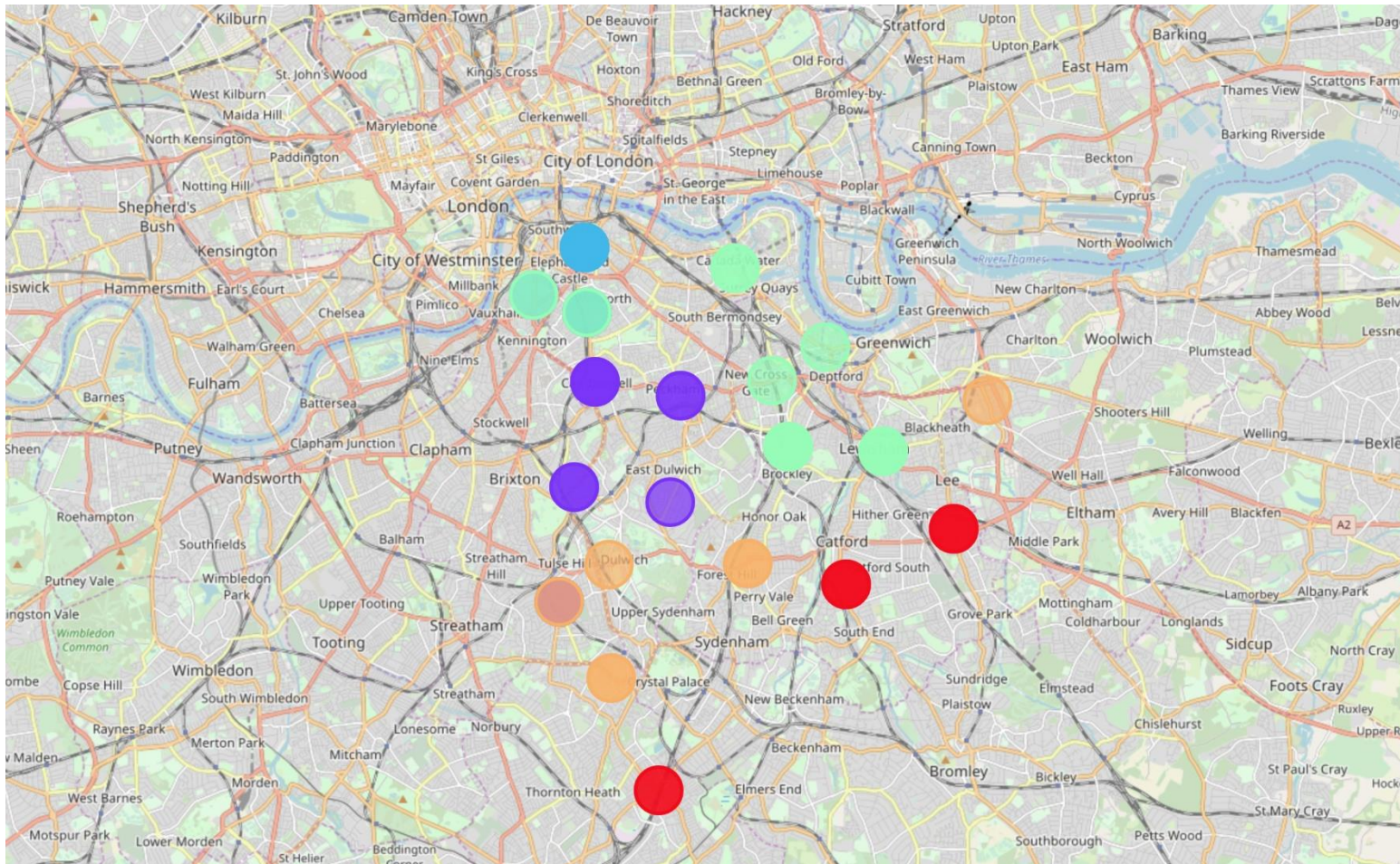
```
In [109]: from sklearn.metrics import silhouette_score
from sklearn.cluster import KMeans

for n_cluster2 in range(2, 10):
    kmeans2 = KMeans(n_clusters = n_cluster2, random_state = 0).fit(se_grouped_clustering)
    label2 = kmeans2.labels_
    sil_coeff = silhouette_score(se_grouped_clustering, label2, metric = 'euclidean')
    print("Where n_clusters = {}, the Silhouette Coefficient is {}".format(n_cluster2, sil_coeff))
```

```
Where n_clusters = 2, the Silhouette Coefficient is 0.6081085930798573
Where n_clusters = 3, the Silhouette Coefficient is 0.6157992881218252
Where n_clusters = 4, the Silhouette Coefficient is 0.6365055758470731
Where n_clusters = 5, the Silhouette Coefficient is 0.674324874789089
Where n_clusters = 6, the Silhouette Coefficient is 0.769792007465332
Where n_clusters = 7, the Silhouette Coefficient is 0.8095362649872676
Where n_clusters = 8, the Silhouette Coefficient is 0.8666861116629072
Where n_clusters = 9, the Silhouette Coefficient is 0.9337555386129051
```

From the result, the high the `n_clusters` the better the silhouette coefficient. For this project, a cluster value of 5 will be used.

Visualizing the Resulting Clusters



Results

- Pubs, Cafe, Coffee Shops are popular in the South East London
- As for restaurants, the Italian Restaurants are very popular in the South East London area (especially in Southwark and Lambeth areas)
- With the Lewisham area being the most condensed area of Africans in the South East Area, it is surprising to see how in the top 10 venues, you can barely see restaurants in the top 5 venues
- Although, the Clusters have variations, a very visible presence is the predominance of pubs

Discussion

- It is very important to note that Clusters 2 and 3 are the most viable clusters to create a brand African Restaurant. Their proximity to other amenities and accessibility to station are paramount. These 2 clusters do not have top restaurants that could rival their standards if they are created

Conclusion

- In conclusion, this project would have had better results if there were more data in terms of crime data within the area, traffic access and allowance of more venues exploration with the Foursquare (limited venues for free calls).
- Also, getting the ratings and feedbacks of the current restaurants within the clusters would have helped in providing more insight into the best location