THE LONDON COFFEE SHOP PROJECT

CAPSTONE PROJECT - THE BATTLE OF

THE NEIGHBORHOODS (WEEK 2)

APPLIED DATA SCIENCE CAPSTONE BY IBM/COURSERA

BUSINESS PROBLEM

- A young entrepreneur has decided to open a coffee shop in London.
- Location of the shop has not been selected yet
- We will need to explore the venues in London's boroughs to understand where the competition is located

DATA

- The data required for this project can be collected from multiple sources.
- The project uses foursquare API .
- Our project used following libraries pandas, requests, bs4, plotly, folium, geocoder, and numpy.
- Scrapping is used to get content from Wikipedia.
- Columns is table is refined through iterative process

METHODOLOGY

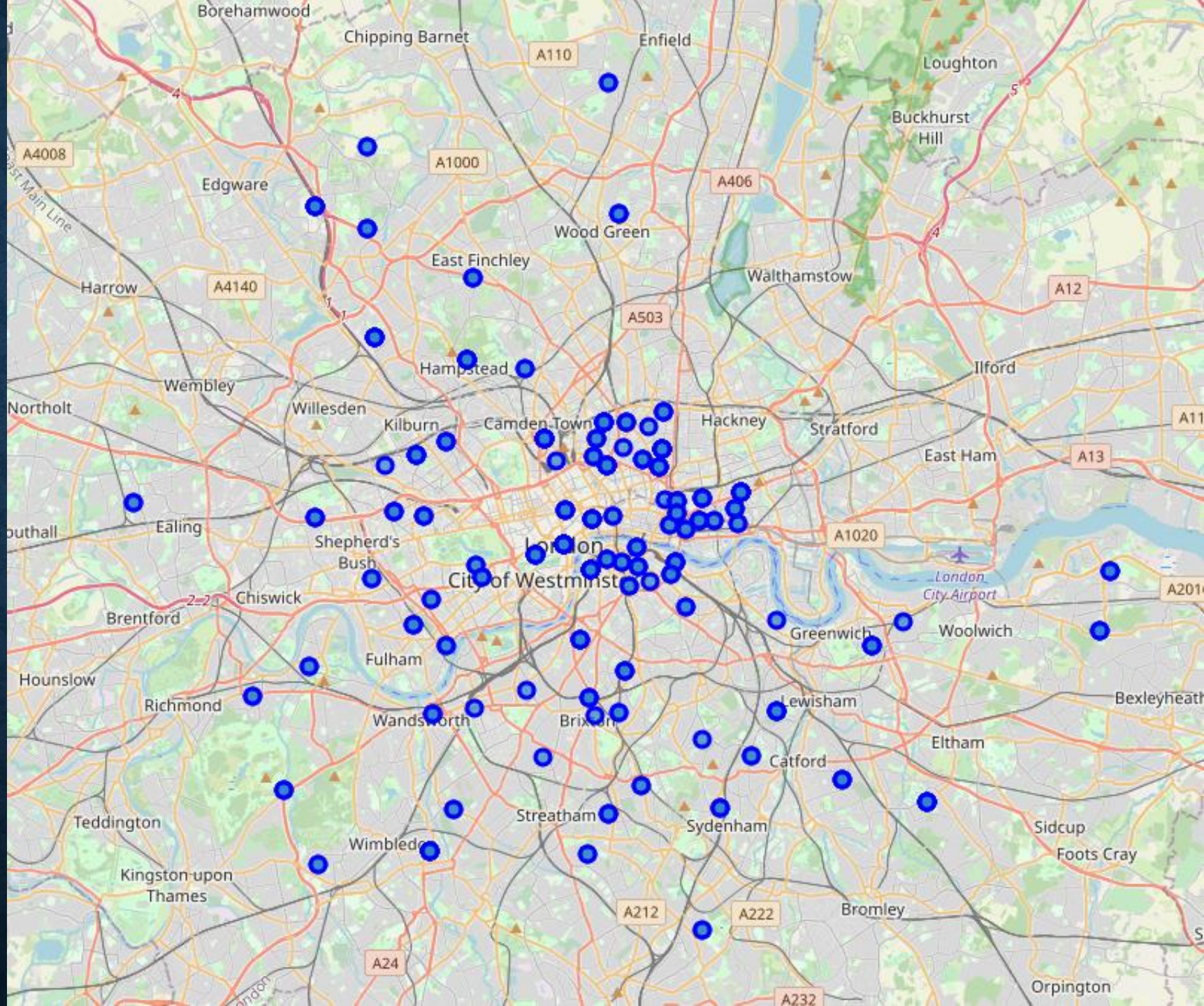
Our methodology consist of following steps

- Data Cleaning.
- Conversion of multiple records in a single row to multiple rows.
- Use of subet.
- Getting coordinates using geocoder and ArcGIS.
- Plot location on map.
- Explore nearby venues
- Apply k mean algorithm.
- Find best cluster.
- Plot best cluster.

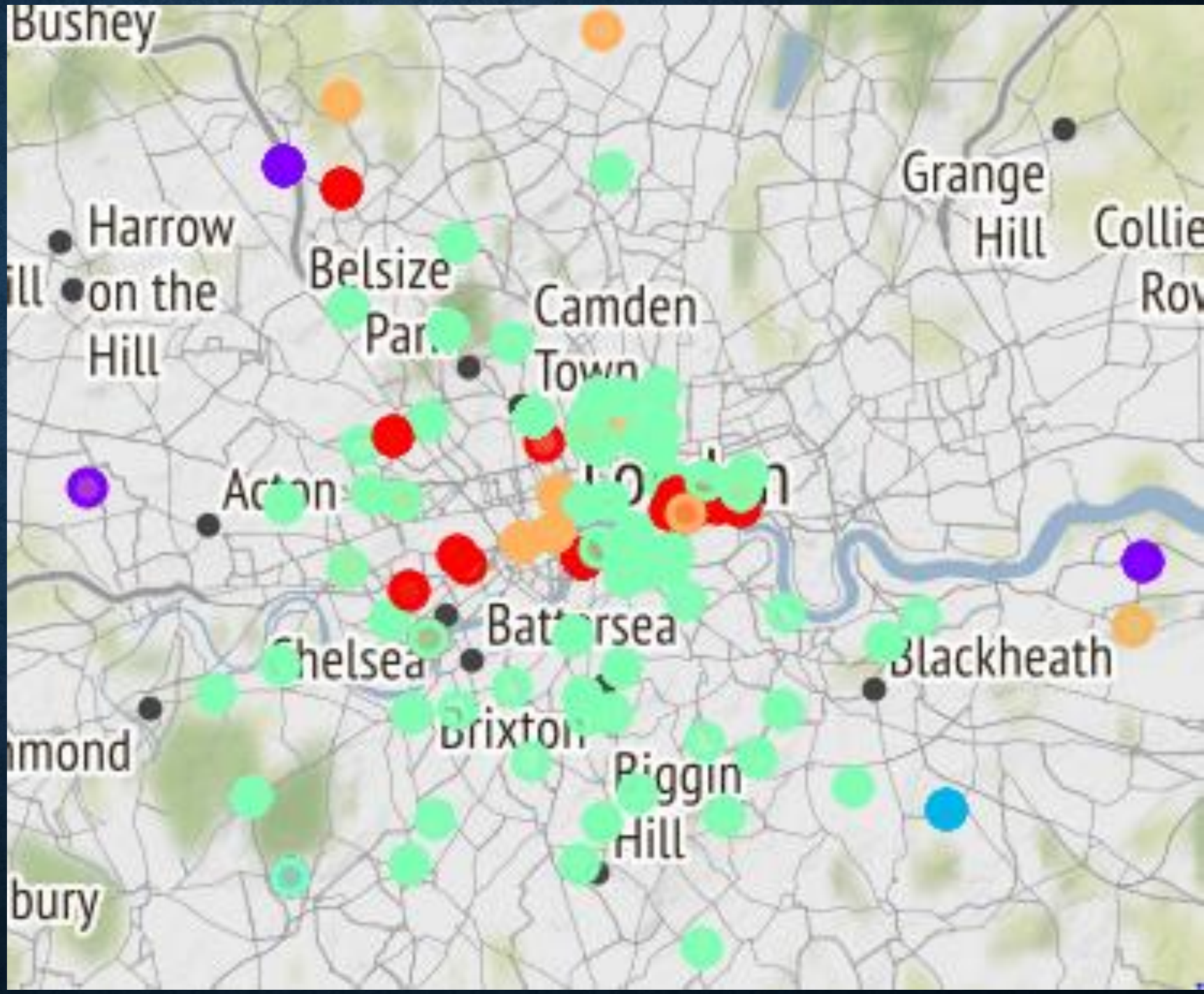
ANALYSIS

- This analysis will gives us area that is most suitable for this business.
- Use of python language
- Importing libraries
- Use of unsupervised learning.
- Cluster 0 is chosen.

MAP OF LONDON NEIGHBORHOODS



MAP OF NEIGHBORHOODS CLUSTERS BASED ON COMMERCIAL ACTIVITY

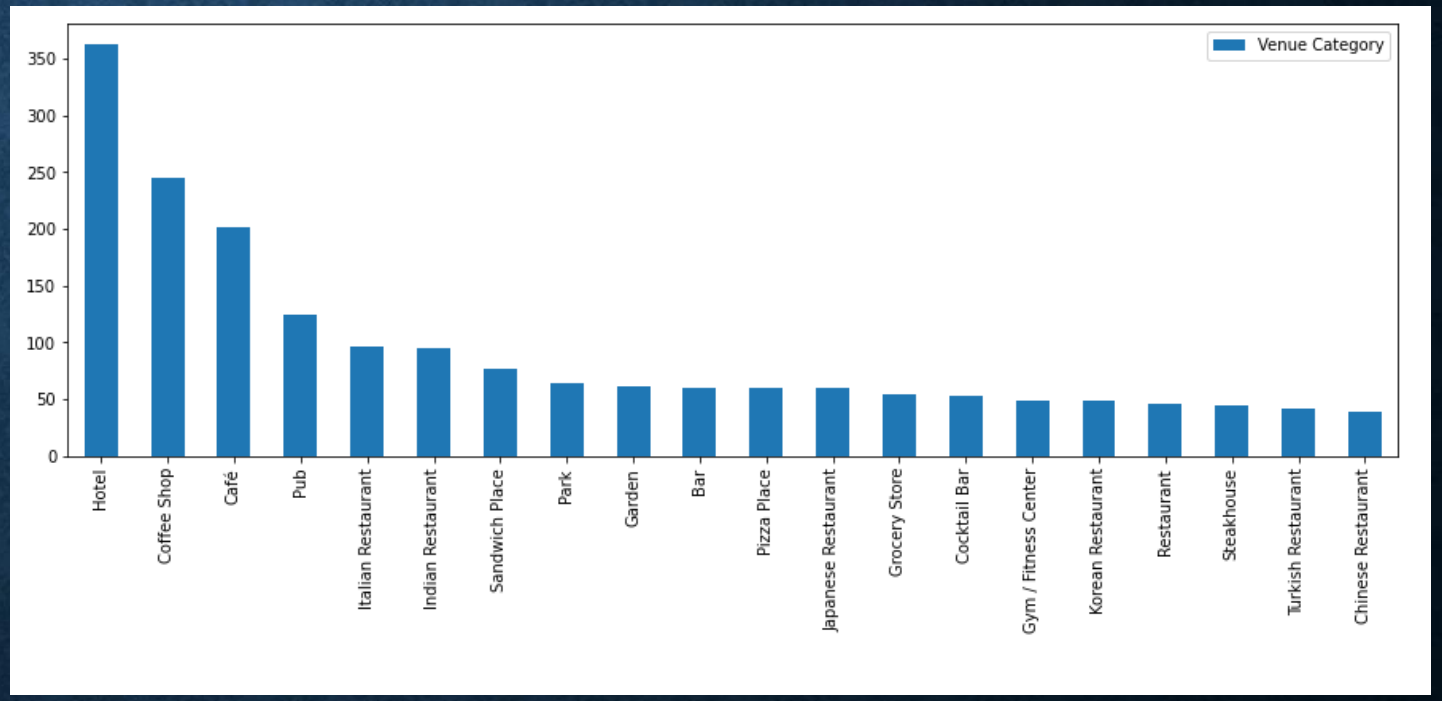


CLUSTER 0

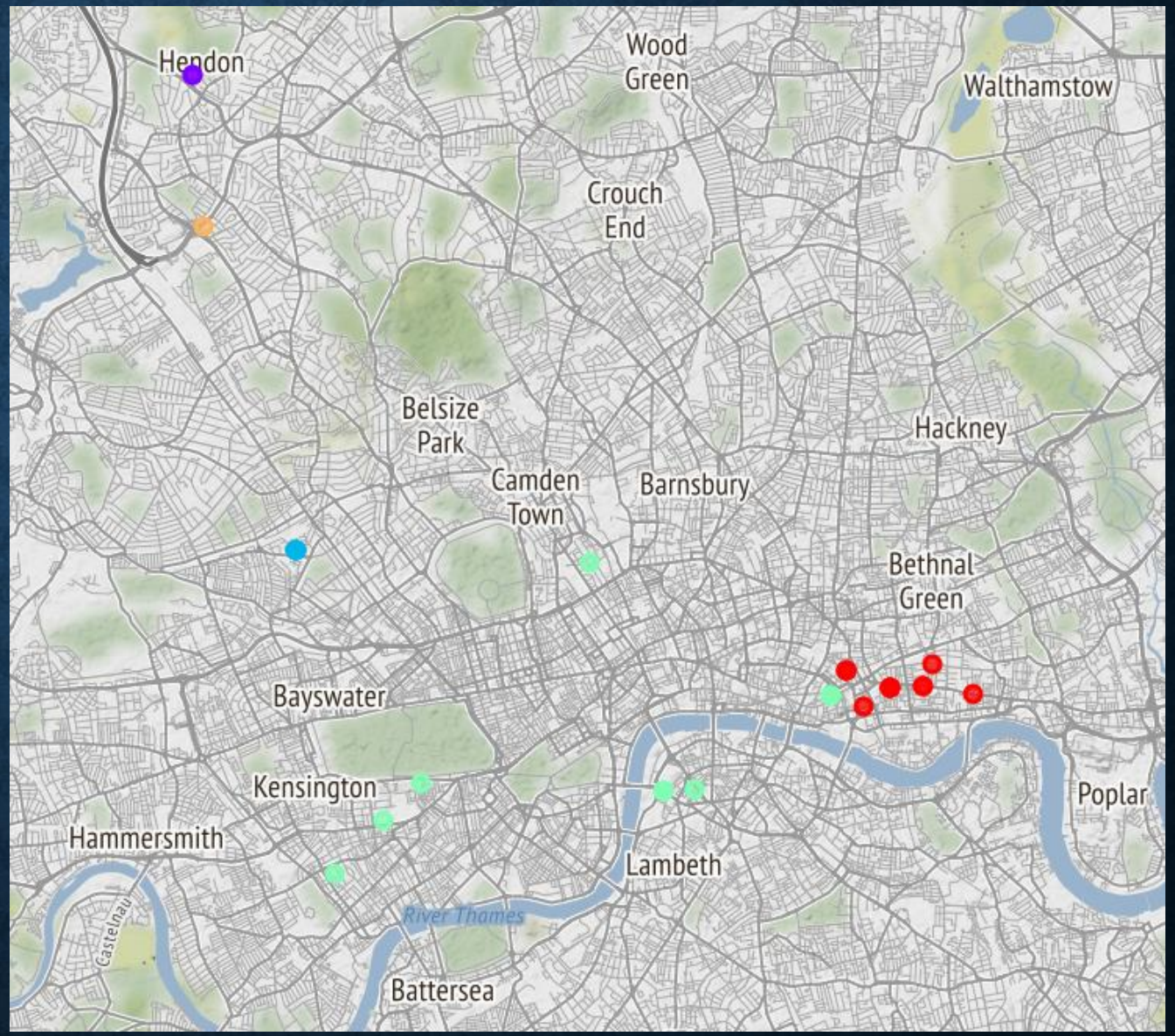
- Low competition
- High number of supporting venues

The total number of venues is: 3304

	Venue Category
Hotel	363
Coffee Shop	245
Café	201
Pub	124
Italian Restaurant	96
Indian Restaurant	95
Sandwich Place	76
Park	64
Garden	61
Bar	60
Pizza Place	60
Japanese Restaurant	59
Grocery Store	54
Cocktail Bar	52
Gym / Fitness Center	49
Korean Restaurant	49
Restaurant	46
Steakhouse	44
Turkish Restaurant	41
Chinese Restaurant	38



SECOND ITERATION OF CLUSTERING FOR CLUSTER 0



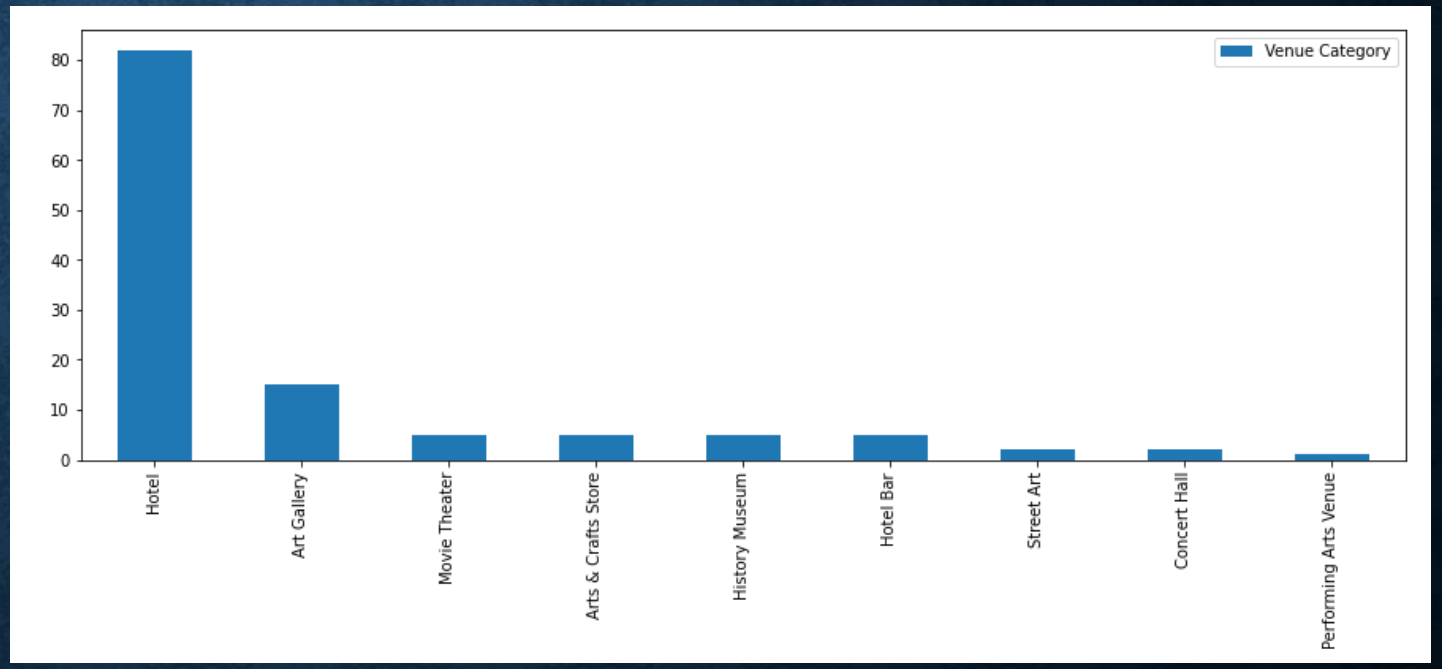
EXPLORING SUPPORTING VENUES FOR CLUSTER 0

Cluster 0.0

- 122 supporting venues

The total number of venues is: 122

	Venue Category
Hotel	82
Art Gallery	15
Movie Theater	5
Arts & Crafts Store	5
History Museum	5
Hotel Bar	5
Street Art	2
Concert Hall	2
Performing Arts Venue	1



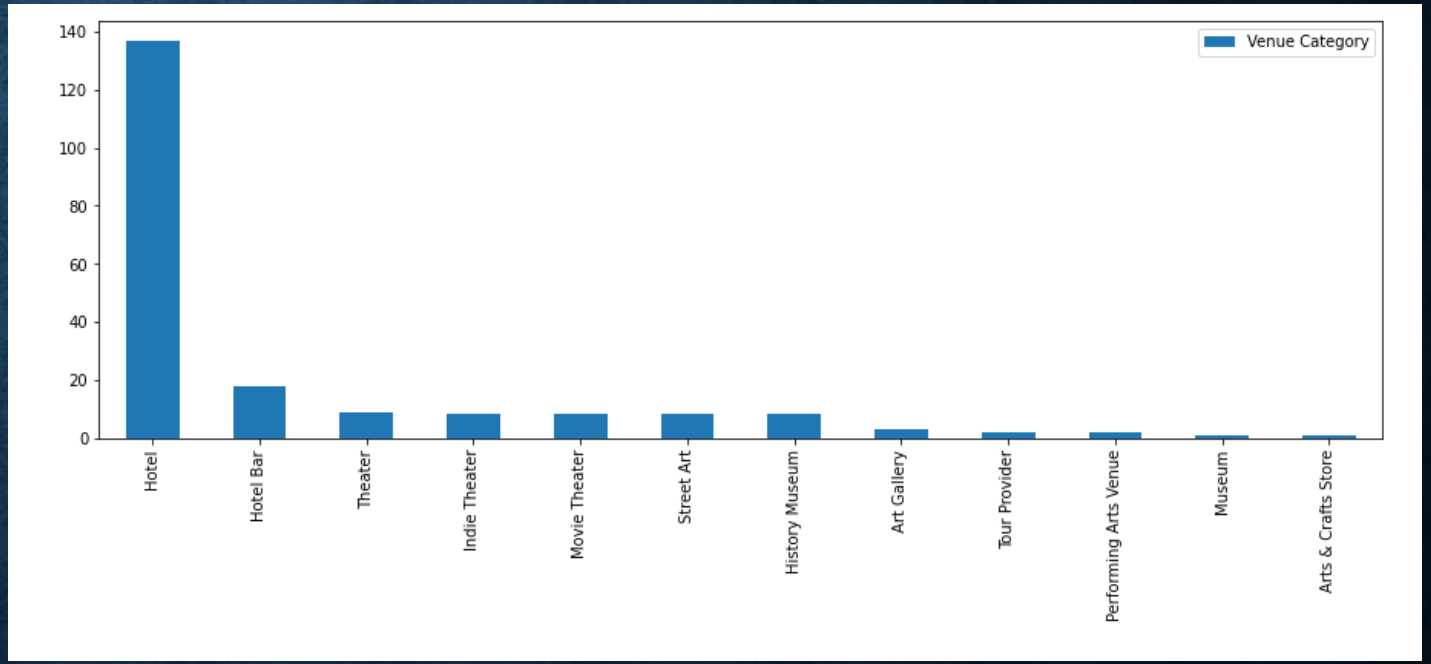
EXPLORING SUPPORTING VENUES FOR CLUSTER 0

Cluster 0.3

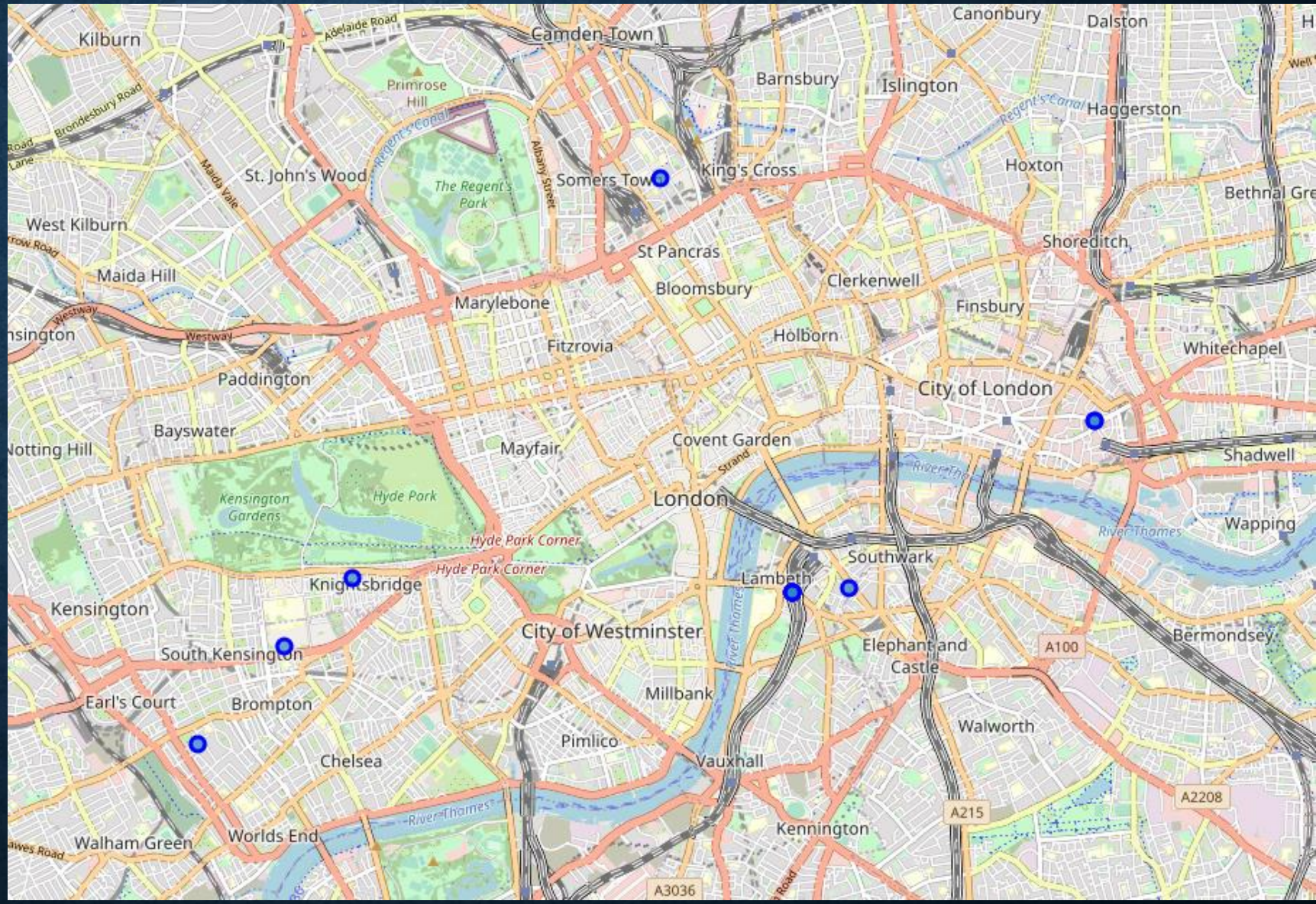
- 205 supporting venues
- It is the optimal cluster

The total number of venues is: 205

	Venue Category
Hotel	137
Hotel Bar	18
Theater	9
Indie Theater	8
Movie Theater	8
Street Art	8
History Museum	8
Art Gallery	3
Tour Provider	2
Performing Arts Venue	2
Museum	1
Arts & Crafts Store	1



OPTIMAL CLUSTERS FOR NEW COFFEE SHOP



RESULTS AND DISCUSSION

- Cluster 0 and 3 performed better
- Cluster 0 is best after one more iteration.
- Other important factors for business success would be customer satisfaction, customer loyalty, customer retention, market share, and the firm's ability to charge a premium price
- Atmosphere is added as one of the quality factors.
- Another important factor for customer satisfaction is quality on food being served

CONCLUSION

- Implementation of k mean algorithm gives many clusters.
- Cluster 0 performed better.
- Café should be open in places that are located in cluster 0.