IBM - CAPSTONE PROJECT

1. Introduction

Amsterdam is the largest city in the Netherlands with multifarious venues. Therefore, to start a food and beverage service is not an easy task.

A job owner who is living another country wants to open a food and beverage service (like bar, cafe, restaurant) in Amsterdam. Since Amsterdam has lots of place to start a business, the job owner wants to earn money as much as s/he can. There are lots of criteria for this purpose like, population density of a location, number of cafes, restaurants, bars etc.

In this project, I will show you the analysis of the problem, and try to find the best place to open a food and beverage service in the Amsterdam with an unsupervised learning technique K-means.

2. Data Description

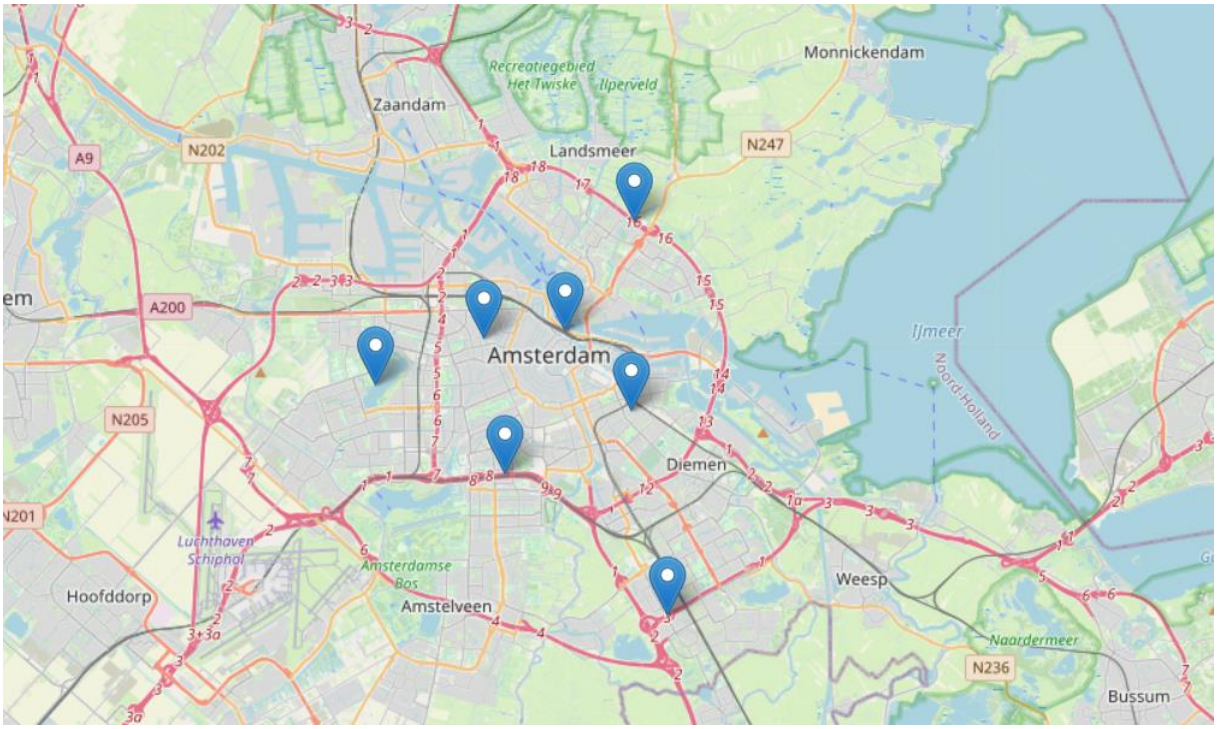
- Related Amsterdam data was taken from https://en.wikipedia.org/wiki/Boroughs_of_Amsterdam
- Coordinates of Amsterdam Boroughs were obtained from Geocoder
- Venues of Amsterdam were obtained from Foursquare

	Borough	Population	Population density	Neighbourhoods
0	Centrum (Centre)	86422	13,748/km²	Binnenstad, Grachtengordel, Haarlemmerbuurt, J...
1	Noord (North)	94766	2,269/km²	Banne Buiksloot, Buiksloot, Buikslotermeer, Fl...
2	Nieuw-West(New West)	151677	4,478/km²	Geuzenveld, Nieuw Sloten, Oostoever, Osdorp, O...
3	Oost (East)	135767	7,635/km²	IJburg, Indische Buurt, Eastern Docklands, Oud...
4	West	143842	15,252/km²	Frederik Hendrikbuurt, Houthaven, Spaarndammer...
5	Westpoort(West Gateway)	192	10/km²	Westpoort
6	Zuid (South)	144432	9,349/km²	Apollobuurt, Buitenveldert, Hoofddorppleinbuur...
7	Zuidoost(Southeast)	87854	4,391/km²	Bijlmermeer, Venserpolder, Gaasperdam, Driemond

3. Methodology and Analysis

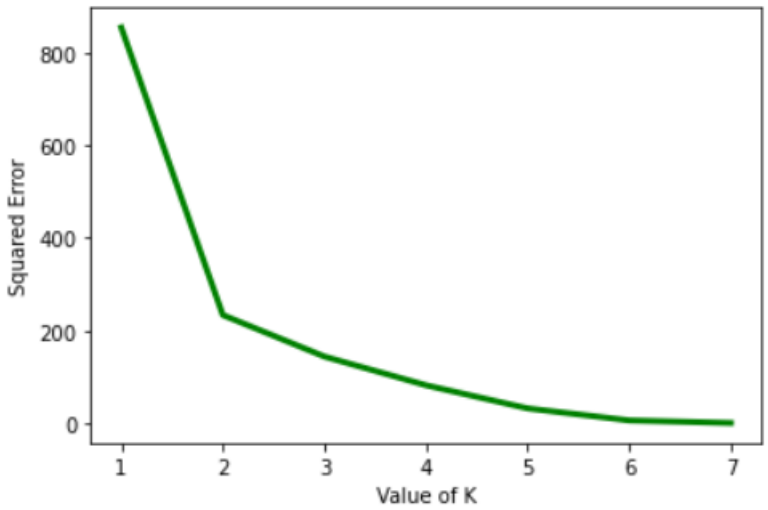
After the data preparation the Boroughs of Amsterdam looks like this.

Note: Maximum 100 venues are collected for each district because of the limitation for collecting the venue data.

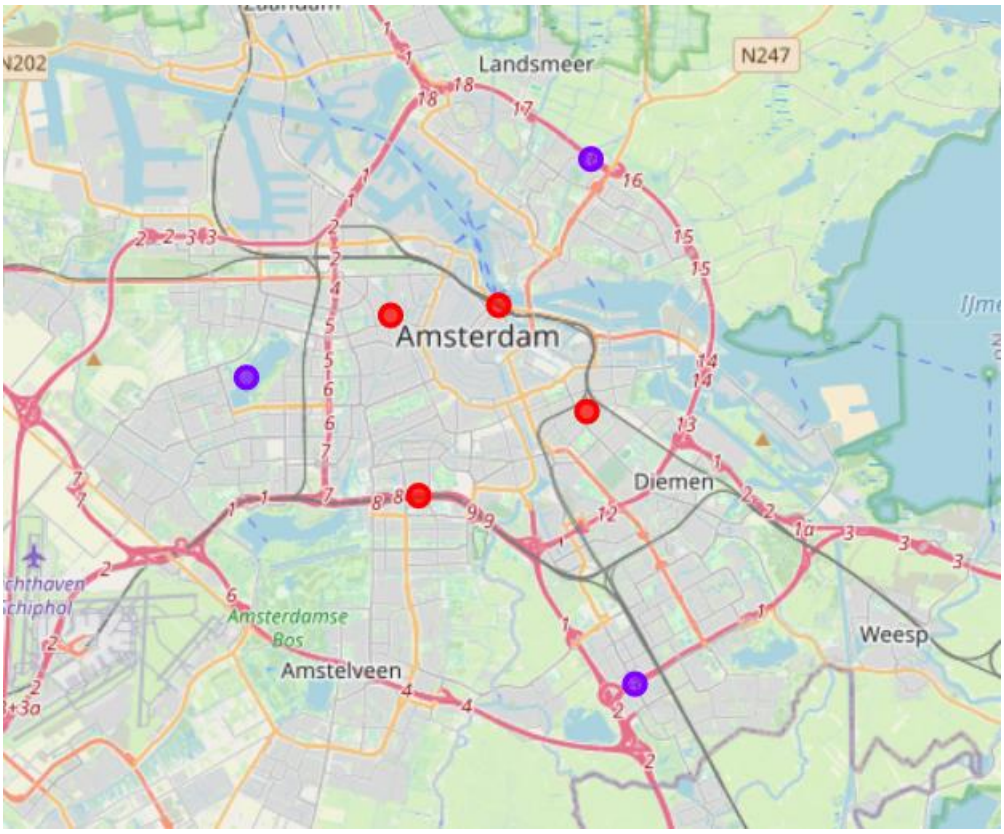


For the machine learning part, unsupervised technique K-means were used.

To obtain the best K, Elbow method was used.



Based on the 2 clusters, the result as follows:



	Cluster Labels	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	0	Amsterdam-Centraal	Hotel	Coffee Shop	Restaurant	Bar	Marijuana Dispensary	Pub	Café	French Restaurant	Chocolate Shop	Church
3	0	Amsterdam-Oost	Bar	Café	Coffee Shop	Italian Restaurant	French Restaurant	Ice Cream Shop	Market	Restaurant	Bakery	Cocktail Bar
4	0	Amsterdam-West	Bar	Coffee Shop	Italian Restaurant	Restaurant	Café	Pizza Place	Grocery Store	Park	Vegetarian / Vegan Restaurant	Gym / Fitness Center
5	0	Amsterdam-Zuid	Coffee Shop	Italian Restaurant	Restaurant	Bakery	Supermarket	Hotel	Department Store	Breakfast Spot	Plaza	Salad Place

Cluster 2

```
venues_sorted[(venues_sorted['Cluster Labels'] == 1)]
```

	Cluster Labels	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	1	Amsterdam-Nieuw-West	Gym	Trail	Soccer Field	Bus Stop	Clothing Store	Supermarket	Drugstore	Theater	Shopping Mall	Mediterranean Restaurant
2	1	Amsterdam-Noord	Bus Stop	Golf Course	Monument / Landmark	Gym / Fitness Center	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Court	Food & Drink Shop	Fast Food Restaurant
6	1	Amsterdam-Zuidoost	Restaurant	Furniture / Home Store	Hotel	Coffee Shop	Fast Food Restaurant	Café	Bagel Shop	Hotel Bar	Food & Drink Shop	Food Court

4. Conclusion

To conclude, two clusters were defined. Cluster 1 (Cluster Label = 0), which includes Centraal, Oost, West, and Zuid, has mostly Bars, Cafes and Restaurants. And Cluster 2 (Cluster Label = 1), which includes Nieuw-West, Noord, Zuidoost, mostly has Stores, Bus Stops, Gyms etc.

Since density rates in descending order as follows: West, Centraal, Zuid, Oost, we can say that:

- If s/he want to run a bar; s/he should prefer West, Centraal, Oost respectively,
- If s/he want to run a café; s/he should prefer Oost, West,
- If s/he want to run a restaurant; s/he should prefer West, Zuid, Oost, respectively.