

Capstone Project — The Battle of Neighborhoods in Amsterdam: Restaurants

Introduction/Business problem

Amsterdam, the city author works, attracting millions of tourists every year because of its stunning architecture, rich culture, history and nightlife. While enjoying this beautiful city, eating a delicious meal ideally meets tourists' expectations. However, finding the right place to eat can be a challenge. Different tourists prefer different dishes due to various reasons like culture, faith, moral, etc.

Thus, the goal of this exercise is to give a simple recommendation to tourists in Amsterdam: in which neighborhood of the city fit their needs in terms of food venues. For instance, Where to eat Mediterranean food, where to find Turkish food, where to get fast food?

Data Requirements

List of neighbourhoods in Amsterdam

Data source:

https://en.wikipedia.org/wiki/Boroughs_of_Amsterdam

<https://opencagedata.com>

Description: I will Scrap 69 Amsterdam neighbourhoods from the Wikipedia page and get the coordinates of those neighbourhoods using the OpenCage Geocoding API.

Restaurants in each neighborhood of Amsterdam

Data source: Foursquare APIs:

Description : Foursquare is a US tech company from New York focusing on location data. Their technology and data powers apps such as Apple's Maps, Uber, Twitter . By using this API we will get all the venues in each neighborhood. We can filter these venues to get only restaurants.

This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

Methodology

Data Preparation

Firstly, I scraped the neighbourhoods in Amsterdam from the wikipedia page on https://en.wikipedia.org/wiki/Boroughs_of_Amsterdam in order to create a data-frame by using pandas. As a result following data-frame is created:

	Borough	Area	Population	Population density	Location (in green)	Neighbourhoods
0	Centrum (Centre)	8.04 km ²	86422	13,748/km ²	NaN	Binnenstad, Grachtengordel, Haarlemmerbuurt, J...
1	Noord (North)	49.01 km ²	94766	2,269/km ²	NaN	Banne Buiksloot, Buiksloot, Buisklootmeer, FI...
2	Nieuw-West(New West)	32.38 km ²	151677	4,478/km ²	NaN	Geuzenveld, Nieuw Sloten, Oostoever, Osdorp, O...
3	Oost (East)	30.56 km ²	135767	7,635/km ²	NaN	IJburg, Indische Buurt, Eastern Docklands, Oud...
4	West	9.89 km ²	143842	15,252/km ²	NaN	Frederik Hendrikbuurt, Houthaven, Spaarndammer...
5	Westpoort(West Gateway)	10 km ²	192	10/km ²	NaN	Westpoort
6	Zuid (South)	17.41 km ²	144432	9,349/km ²	NaN	Apollobuurt, Buitenveldert, Hoofddorpplainbuur...
7	Zuidoost(Southeast)	22.08 km ²	87854	4,391/km ²	NaN	Bijlmermeer, Venserpolder, Gaasperdam, Driemond

After that I need to get the data-frame which contains a list of every neighbourhood. For that purpose, I cleaned and transformed above data-frame to as following data-frame:

	Neighbourhoods
0	Binnenstad
1	Grachtengordel
2	Haarlemmerbuurt
3	Jodenbuurt
4	Jordaan
5	Kadijken
6	Lastage
7	Oosterdokseiland

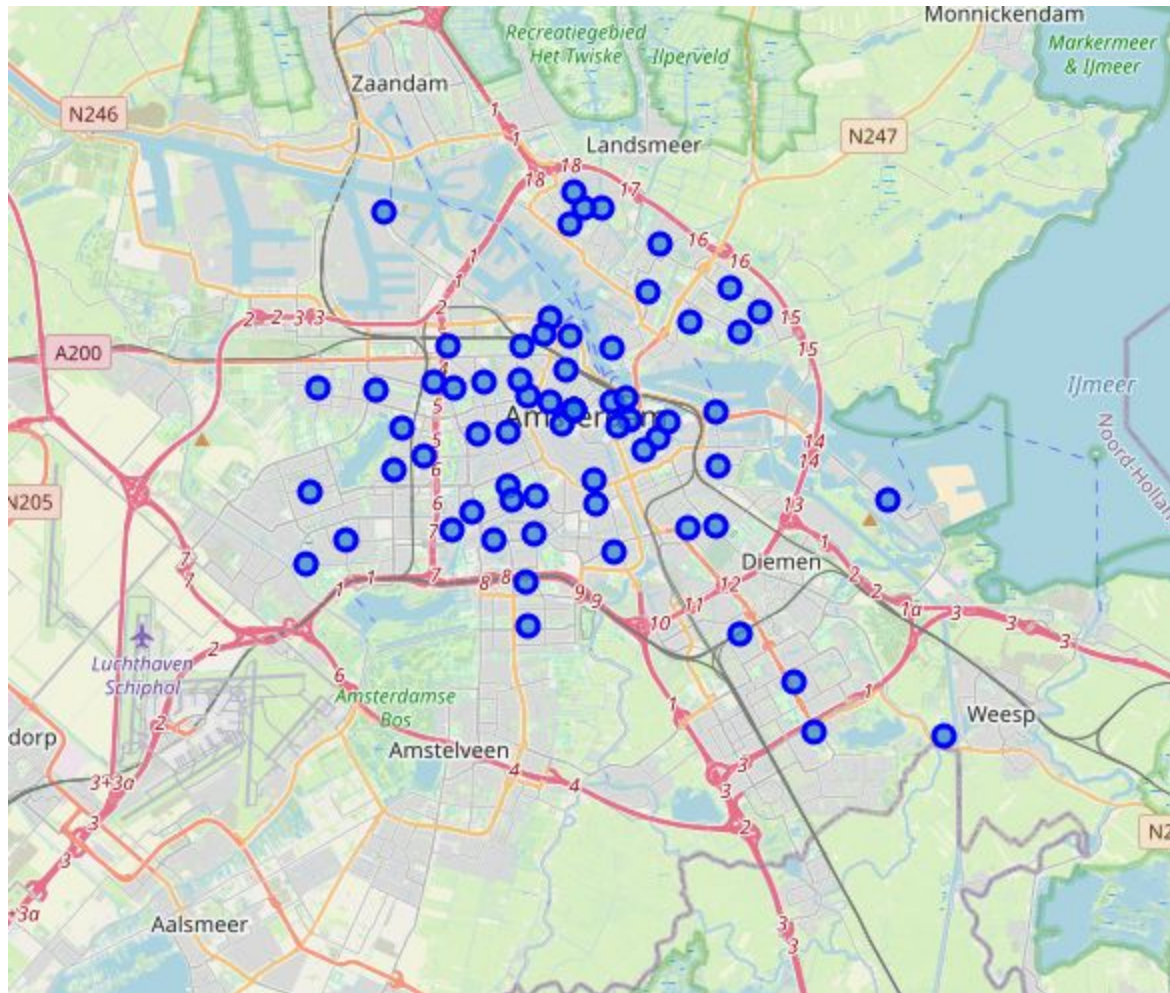
Getting coordinates of neighbourhoods

I tried google geocoding API and Geopy for the coordinates of neighbourhoods, however some of the neighbourhoods' coordinates were not correctly obtained. After doing some research I found that the OpenCage Geocoding API on <https://opencagedata.com> did a good job except one neighbourhood. The result as follow:

:

	Neighbourhoods	lat	lon
0	Binnenstad	52.093622	5.119690
1	Grachtengordel	52.370837	4.885478
2	Haarlemmerbuurt	52.382441	4.887193
3	Jodenbuurt	52.374030	4.889690
4	Jordaan	52.375416	4.881096
5	Kadijken	52.368155	4.918586
6	Lastage	52.375459	4.902402
7	Oosterdokseiland	52.376250	4.907307
8	Oostelijke Eilanden	52.371553	4.921840
9	Plantage	52.365705	4.913661
10	Rapenburg	52.371681	4.909008

Thereafter, I used the python **folium** library to visualize geographic details of Amsterdam and 69 neighborhoods. As I mentioned earlier one neighborhood namely, Binnenstad's coordinates, was not correctly obtained hence, it was placed in another city. It gives as below result:



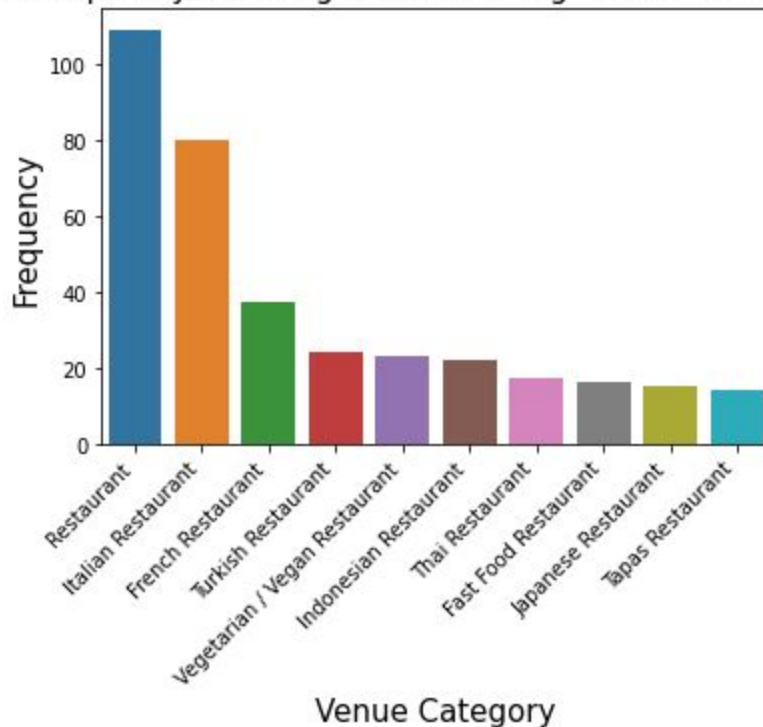
Using Foursquare Location Data

Now it is Foursquare's turn, I retrieved the foursquare data for all venues within a radius of 500 meters from each center of each neighbourhood. I got 2562 venues all over Amsterdam city from them 545 are restaurants and from these restaurants come from 56 unique restaurant categories.

After that 10 most frequently visiting restaurants were found and visualized by bar chart as following:

	Venue_Category	Frequency
0	Restaurant	109
1	Italian Restaurant	80
2	French Restaurant	37
3	Turkish Restaurant	24
4	Vegetarian / Vegan Restaurant	23
5	Indonesian Restaurant	22
6	Thai Restaurant	17
7	Fast Food Restaurant	16
8	Japanese Restaurant	15
9	Tapas Restaurant	14

10 Most Frequently Occuring Venues in Neighbourhoods of Amsterdam



As we can see from the barchart Restaurant (it may include diverse restaurants) has the highest frequency and Tapas restaurant has the lowest. It gives pretty well overview about popularity of restaurants in Amsterdam and obviously mediteraian foods are highly appreciated Turkish and asian foods come second and third places.

To find clusters of restaurant types in different neighbourhoods following steps needs to be taken:

- Create a one hot encoding data-frame by pandas for the venue categories:


```
]: amsterdam_onehot_grouped=amsterdam_onehot.groupby(["Neighbourhoods"]).sum().reset_index()
print(amsterdam_onehot_grouped.shape)
amsterdam_onehot_grouped.head()
```

(60, 56)

	Neighbourhoods	Afghan Restaurant	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Belgian Restaurant	Cantonese Restaurant	Caribbean Restaurant	Caucasian Restaurant	Chinese Restaurant	French Restaurant
0	Bos en Lommer	0	0	0	0	0	0	0	0	0	0	0	0
1	Buikslotermeer	0	0	0	0	0	0	0	0	0	0	0	0
2	Buitenveldert	0	0	0	0	1	0	0	0	0	0	0	0
3	De Baarsjes	0	0	0	0	0	0	0	0	0	0	1	0
4	De Pijp	0	0	0	0	0	0	0	0	0	0	1	1

- Use pandas groupby on neighbourhoods column and calculate the mean of the frequency of occurrence of each venue category:

```
amsterdam_grouped = amsterdam_onehot.groupby('Neighbourhoods').mean().reset_index()
amsterdam_grouped.head()
```

	Neighbourhoods	Afghan Restaurant	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Belgian Restaurant	Cantonese Restaurant	Caribbean Restaurant	Caucasian Restaurant	Chinese Restaurant	French Restaurant
0	Bos en Lommer	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
1	Buikslotermeer	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
2	Buitenveldert	0.0	0.0	0.0	0.0	0.083333	0.0	0.0	0.0	0.0	0.0	0.000000	0.0
3	De Baarsjes	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.090909	0.0
4	De Pijp	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.028571	0.5

- Each neighborhood along with the top 5 most common venues output as follows:

```
---- Bos en Lommer----
      venue  freq
0      Restaurant  0.15
1  Vegetarian / Vegan Restaurant  0.15
2      Turkish Restaurant  0.15
3      Fast Food Restaurant  0.15
4      Thai Restaurant  0.08
```

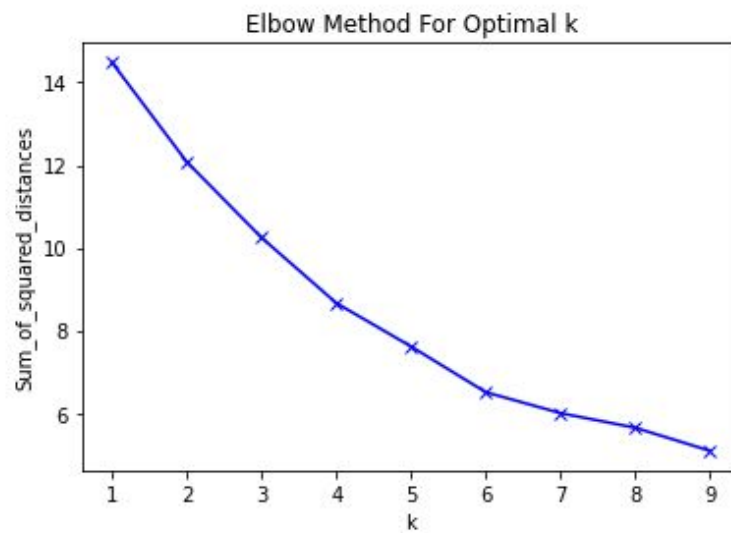
```
---- Buikslotermeer----
      venue  freq
0  Seafood Restaurant  0.33
1  Japanese Restaurant  0.33
2      Restaurant  0.33
3  Afghan Restaurant  0.00
4  Ramen Restaurant  0.00
```

```
---- Buitenveldert----
      venue  freq
0      Sushi Restaurant  0.17
1      Restaurant  0.17
2  Vietnamese Restaurant  0.08
3  Middle Eastern Restaurant  0.08
4      Persian Restaurant  0.08
```

Finally with all that information, a k-means clustering algorithm from scikit-learn package which is categorized as an unsupervised machine learning algorithm was applied. In this stage I need to define the value of the k. For that regard, I tried two different elbow methods and picked the best one which is 6.

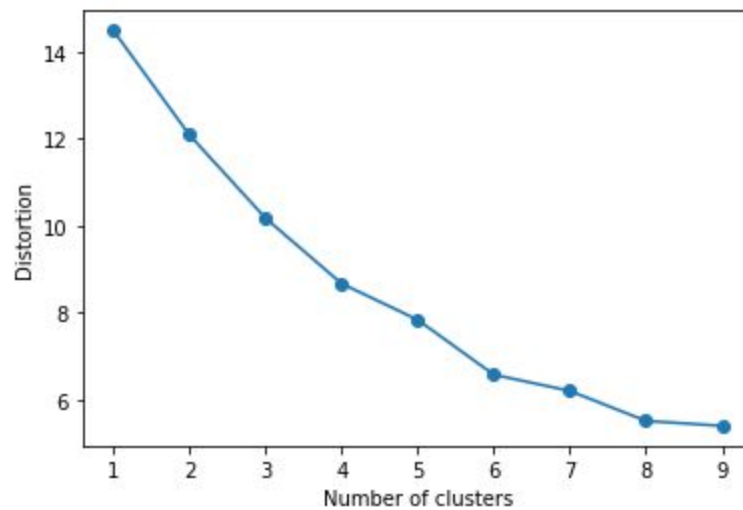
```
In [38]: Sum_of_squared_distances = []
K = range(1,10)
for k in K:
    km = KMeans(n_clusters=k)
    km = km.fit(amsterdam_grouped_clustering)
    Sum_of_squared_distances.append(km.inertia_)
```

```
In [39]: plt.plot(K, Sum_of_squared_distances, 'bx-')
plt.xlabel('k')
plt.ylabel('Sum_of_squared_distances')
plt.title('Elbow Method For Optimal k')
plt.show()
```



```
In [68]: # calculate distortion for a range of number of cluster
distortions = []
for i in range(1, 10):
    km = KMeans(
        n_clusters=i, init='random',
        n_init=10, max_iter=300,
        tol=1e-04, random_state=0
    )
    km.fit(amsterdam_grouped_clustering)
    distortions.append(km.inertia_)

# plot
plt.plot(range(1, 10), distortions, marker='o')
plt.xlabel('Number of clusters')
plt.ylabel('Distortion')
plt.show()
```



Result

The result of clustering as below:

Cluster 1

```
In [61]: amsterdam_merged.loc[amsterdam_merged['Cluster Labels'] == 0, amsterdam_merged.columns[[0] + list(range(6, amsterdam_merged.shape[1]))]]
```

	Neighbourhoods	Venue Category	Cluster Labels	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
1	Binnenstad	Turkish Restaurant	0	Restaurant	Italian Restaurant	French Restaurant	Seafood Restaurant	Greek Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Mexican Restaurar
2	Binnenstad	Vietnamese Restaurant	0	Restaurant	Italian Restaurant	French Restaurant	Seafood Restaurant	Greek Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Mexican Restaurar
3	Binnenstad	Italian Restaurant	0	Restaurant	Italian Restaurant	French Restaurant	Seafood Restaurant	Greek Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Mexican Restaurar
4	Binnenstad	American Restaurant	0	Restaurant	Italian Restaurant	French Restaurant	Seafood Restaurant	Greek Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Mexican Restaurar
5	Binnenstad	Indonesian Restaurant	0	Restaurant	Italian Restaurant	French Restaurant	Seafood Restaurant	Greek Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Mexican Restaurar
6	Binnenstad	Restaurant	0	Restaurant	Italian Restaurant	French Restaurant	Seafood Restaurant	Greek Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Mexican Restaurar
7	Binnenstad	Restaurant	0	Restaurant	Italian Restaurant	French Restaurant	Seafood Restaurant	Greek Restaurant	Indonesian Restaurant	Vietnamese Restaurant	Mexican Restaurar

Cluster 2

```
: amsterdam_merged.loc[amsterdam_merged['Cluster Labels'] == 1, amsterdam_merged.columns[[0] + list(range(6, amsterdam_merged.shape[1]))]]
```

	Neighbourhoods	Venue Category	Cluster Labels	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
81	Kadijken	Italian Restaurant	1	Restaurant	Italian Restaurant	Middle Eastern Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant	Fast Food Restaurant
82	Kadijken	Restaurant	1	Restaurant	Italian Restaurant	Middle Eastern Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant	Fast Food Restaurant
83	Kadijken	Middle Eastern Restaurant	1	Restaurant	Italian Restaurant	Middle Eastern Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant	Fast Food Restaurant
84	Kadijken	Restaurant	1	Restaurant	Italian Restaurant	Middle Eastern Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant	Fast Food Restaurant
171	Kadoelen	Restaurant	1	Restaurant	Vietnamese Restaurant	Dutch Restaurant	Indian Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant	Fast Food Restaurant

Cluster 3

```
: amsterdam_merged.loc[amsterdam_merged['Cluster Labels'] == 2, amsterdam_merged.columns[[0] + list(range(6, amsterdam_merged.shape[1]))]]
```

	Neighbourhoods	Venue Category	Cluster Labels	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
172	Nieuwendam	Indonesian Restaurant	2	Indonesian Restaurant	Dutch Restaurant	Indian Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant	Fast Food Restaurant	Falafel Restaurant
187	Tuindorp Nieuwendam	Indonesian Restaurant	2	Indonesian Restaurant	Dutch Restaurant	Indian Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant	Fast Food Restaurant	Falafel Restaurant

4

Cluster 4

```
amsterdam_merged.loc[amsterdam_merged['Cluster Labels'] == 3, amsterdam_merged.columns[[0] + list(range(6, amsterdam_merged.shape[1]))]]
```

	Neighbourhoods	Venue Category	Cluster Labels	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
115	Oostelijke Eilanden	Mediterranean Restaurant	3	Italian Restaurant	Mediterranean Restaurant	Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
116	Oostelijke Eilanden	Italian Restaurant	3	Italian Restaurant	Mediterranean Restaurant	Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
117	Oostelijke Eilanden	Restaurant	3	Italian Restaurant	Mediterranean Restaurant	Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
211	IJburg	Asian Restaurant	3	Italian Restaurant	Asian Restaurant	Greek Restaurant	Dutch Restaurant	Indian Restaurant	Indian Chinese Restaurant	Halal Restaurant	French Restaurant	Fondue Restaurant
212	IJburg	Italian Restaurant	3	Italian Restaurant	Asian Restaurant	Greek Restaurant	Dutch Restaurant	Indian Restaurant	Indian Chinese Restaurant	Halal Restaurant	French Restaurant	Fondue Restaurant

Cluster 5

```
amsterdam_merged.loc[amsterdam_merged['Cluster Labels'] == 4, amsterdam_merged.columns[[0] + list(range(6, amsterdam_merged.shape[1]))]]
```

	Neighbourhoods	Venue Category	Cluster Labels	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
190	Nieuw Sloten	Chinese Restaurant	4	Chinese Restaurant	Vietnamese Restaurant	Dutch Restaurant	Indian Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant	Fast Food Restaurant
234	Eastern Docklands	Spanish Restaurant	4	Chinese Restaurant	Spanish Restaurant	Restaurant	Vietnamese Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
235	Eastern Docklands	Restaurant	4	Chinese Restaurant	Spanish Restaurant	Restaurant	Vietnamese Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
236	Eastern Docklands	Chinese Restaurant	4	Chinese Restaurant	Spanish Restaurant	Restaurant	Vietnamese Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
237	Eastern Docklands	Chinese Restaurant	4	Chinese Restaurant	Spanish Restaurant	Restaurant	Vietnamese Restaurant	Dutch Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant

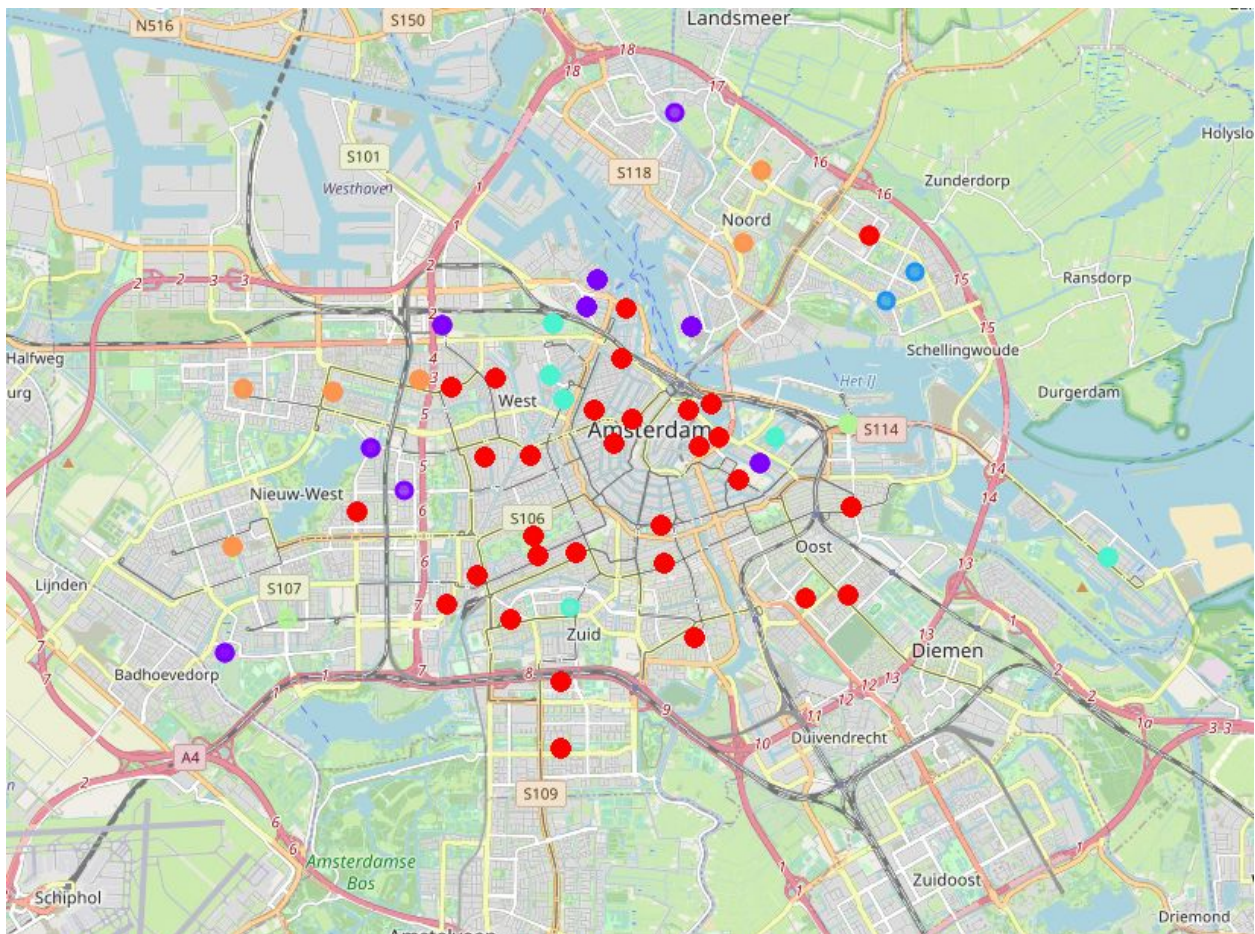
Cluster 6

```
amsterdam_merged.loc[amsterdam_merged['Cluster Labels'] == 5, amsterdam_merged.columns[[0] + list(range(6, amsterdam_merged.shape[1]))]]
```

	Neighbourhoods	Venue Category	Cluster Labels	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
164	Banne Buiksloot	Turkish Restaurant	5	Turkish Restaurant	Restaurant	Vietnamese Restaurant	Doner Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
165	Banne Buiksloot	Restaurant	5	Turkish Restaurant	Restaurant	Vietnamese Restaurant	Doner Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
169	Floradorp	Turkish Restaurant	5	Turkish Restaurant	Vietnamese Restaurant	Dutch Restaurant	Indian Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
170	Floradorp	Turkish Restaurant	5	Turkish Restaurant	Vietnamese Restaurant	Dutch Restaurant	Indian Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
188	Geuzenveld	Turkish Restaurant	5	Italian Restaurant	Turkish Restaurant	Dutch Restaurant	Indian Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant
189	Geuzenveld	Italian Restaurant	5	Italian Restaurant	Turkish Restaurant	Dutch Restaurant	Indian Restaurant	Indian Chinese Restaurant	Halal Restaurant	Greek Restaurant	French Restaurant	Fondue Restaurant

From the tables above one can see the neighborhoods and their most common venues as well as corresponding clusters.

As beneath the neighborhoods marked with a cluster specific map is represented by using Folium library:



Discussion

Let's summarize our findings based on final result:

- Cluster 1 is the biggest cluster among the other 6 clusters. When we examine the first 5 most common venues one can see Italian restaurants come first place then coming French, Seafood , Greek restaurants and if further looking you can find diverse food types of restaurants. For me as a working person in Amsterdam that makes sense. This cluster covers almost all the best known neighborhoods for restaurants like Jordaan, Kinkerbuurt, Grachtengordel, De pijp, Bos Lommer, De Baarsjes, etc.
- Cluster 2 is the second big cluster and it can be described as mixed with Italian and Turkish food clusters if we exclude general restaurant items. Additionally, this cluster might provide best options if someone prefers Dutch, Indian-Chinese, Vegetarian and Halal food restaurants.
- Cluster 3 the smallest cluster and it obviously can be described as Indonesian and Dutch food cluster.
- Cluster 4 can be described as Mediterranean food cluster. However, you can also find Dutch, Asian and halal (Turkish) food restaurants.
- Cluster 5 can be described as Chinese and Spanish food cluster.

- Cluster 6 obviously can be described as Turkish or Halal food cluster and I totally agree with that. But you can also find Asian and Fast food restaurants as well.

Lastly, this result is purely based on the most common venues obtained from Foursquare data. Hence, it gives travelers only an overview of Restaurants distribution by categories in the 69 neighborhoods in Amsterdam and combination of other factors like distance of the venues from closest stations, range of prices of restaurants and ratings of restaurants and so on could be the best way to choose the right restaurant.

Conclusion/Acknowledgement

This is the fruit of my online IBM Data Science Professional Certificate course on coursera. The goal of this project is to give a simple recommendation for tourists in Amsterdam. I applied all kinds of data skills that I learned from the course, like data cleaning, data visualization and machine learning tools and so on. It is amazing to see that real-life problems like this can be solved by using data science tools. Besides the knowledge which I've gained from the course, I am inspired by some similar pieces of works and I highly appreciate them for contribution to my work. I am also so grateful for all kinds of online open resources and insights.

Below I listed the main sources that I referenced:

<https://www.coursera.org/professional-certificates/ibm-data-science>

<https://www.linkedin.com/pulse/applied-data-science-capstone-project-restaurant-wagner-mba/?articleId=6670274875946622976>

<https://medium.com/@radialee/capstone-project-the-battle-of-neighborhoods-in-tokyo-restaurants-45a503e65ff>

<https://towardsdatascience.com/k-means-clustering-with-scikit-learn-6b47a369a83c>

<https://www.scikit-yb.org/en/latest/api/cluster/elbow.html#quick-method>