IBM CAPSTONE PROJECT MMK

The Battle of Neighborhoods (Rotterdam and The Hague)

Table of Contents

Introduction and Business Problem

Background

The Battle of Neighborhoods in this project I am going to compare the neighborhoods of the two cities (Rotterdam and The Hague) in South Holland province of the Netherlands. By the end of the project I am going to determine how similar or dissimilar they are. Rotterdam is a major port city and it is now known for bold and modern architecture. Rotterdam city was almost completely reconstructed after WWII. On the North Sea coast of the western Netherlands, the Hague city is located. The Hague city is home to the International Criminal Court (ICC) and the U.N.'s International Court of Justice. For people that are new to the Netherlands particularly to the South Holland province, both cities offer many interesting things to do such as visiting the beautiful buildings architecture in Rotterdam city and the beach in the Hague. At the same time, it would be fun and exciting to visit and eat in the restaurants which are customer's recommended and preferred choice. The locals are very outgoing and take every opportunity to taste different cuisines and food which is offered by the diverse and multinational community in the Netherlands. Similarly, the international students are trying their best to taste various food and find familiar taste of food offered by their country cooks. Importantly, **the main target audience of this project are the travelers** who have limited time to visit amazing places and will be provided a list of restaurants, marked on the map, in clusters based on the customers' (users) venue likes from Foursquare.

Business Problem

In this project, I am going to compare Rotterdam and the Hague city that how similar or dissimilar they are. In addition, I will provide two separate lists of the most liked restaurants marked on the map based on the users who have like the venues (restaurants) on Foursquare. One might think that travelers could google and/or use different websites to find restaurants which have the option of ratings and reviews. Similar to the machine learning and data science skills that google and other applications use to offer such choices, here I am going to replicate and offer that service with the skills that I have learned joining IBM Data Science course. **Thus, the business**

problem that will be tackled in this project is to assist the visitors/travelers (to the Netherlands) and anyone interested to find customer's preferred choice restaurants. For two main reasons it is important to solve this business problem. First, it is time saving and takes away the unnecessary anxiety of missing the opportunity to eat in the best restaurants in a single visit to the Netherlands. Second, for the businesses (restaurants) this project provides in clusters, categorization of restaurants, based the customer's experience, which could be used to improve their services.

Data

From Wikipedia, list of cities, towns and villages in South Holland are available. I am going to use BeautifulSoup, one of the available libraries in Python, to extract (Name, Municipality, Coordinates) data. Subsequently, after cleaning and preparing the data frame for the two cities through Foursquare API the following data on venues (restaurants) in two cities Rotterdam and the Hague of the Netherlands are compiled and stored for further analysis:

- Venue Name
- Venue ID
- Venue Location (lat and lng)
- Venue Category
- Venue User Likes

Data Acquisition Approach

Latitude and longitude coordinates for South Holland province focusing on Rotterdam and the Hague cities.

Through Foursquare API to get a list of all venues in Rotterdam and the Hague cities consist of venue name, venue id, location, category, and likes.

Methodology

Following similar process that, Christopher Jameson, has carried out that venue users "likes are a proxy for quality." The number of "likes" for each of the restaurants are counted and in stored in column (satisfactory category) categorizing the restaurant in four categories (Poor, Fair, Good, and Excellent). This will be done using k-means clustering algorithm grouping restaurants into 4 clusters. Finally, I will provide two separate lists of the most liked restaurants marked on the map, based on the users who have like the venues (restaurants) on Foursquare, to be able to choose among the restaurants that customers (users) have liked the venues (restaurants). Depending of the type of membership on Foursquare the regular and premium calls limits calls such as venue tips and photos.

Data requirement

Neighbourhood (name) and boroughs (municipality) in the table. In order to segment the neighbourhoods and explore them, I used the dataset that contains the municipality and the name as well as the latitude and longitude coordinates of each neighbourhood.

This dataset exists for free on the web. Feel free to try to find this dataset on your own: https://en.wikipedia.org/wiki/List_of_cities,_towns_and_villages_in_South_Holland

Data is available in a table format under: Name, Municipality, Coordinates, Notes columns. These columns are renamed as Neighborhood, Borough, Coordinates, Notes. Coordinates column is consist of several codes separated by / with N and E, also special symbols are included as well. Therefore, in multi steps latitude and longitude are extracted.

Geo data from Wikipedia

On this page geographical data is available for all the cities in South Holland. In this project columns are renamed to fit the consistency in the written code and for ease of tracking the outputs. Neighborhood (Name), Borough (Municipality), latitude and longitude (Coordinates). The focus of this project is two neighborhoods (Rotterdam city and Den Haag “The Hague” city).

Folium

South Holland map

Using folium I have created South Holland map and all the neighborhoods are marked on the map.

Exploratory data analysis

List of nearby neighborhood in Rotterdam with categories

```
nearby_venues.head()
```

Out[81]:

	name	id	categories	lat	lng
0	Sugo	54e35ace498e94443cfd14e5	Pizza Place	51.921030	4.477747
1	THOMS	5676cb0e498e19ce9f0f106f	Restaurant	51.921856	4.481555
2	Urban Residences Rotterdam	508aaff7e4b01ceeb29e0971	Residential Building (Apartment / Condo)	51.920041	4.477572
3	Stadsbrouwerij Thoms	5b23df7895d986002c24dda9	Beer Garden	51.922837	4.480854
4	The James	5ae23ca865211f0039a4f537	Hotel	51.921074	4.477826

List of nearby neighborhood in Den Haag with categories

```
d_nearby_venues.head()
```

Out[82]:

	name	id	categories	lat	lng
0	Baladi Manouche	555b50bd498e2843efb7c8c7	Bakery	52.078805	4.303832
1	Kua	52f529ae11d203e7b26d4453	Mexican Restaurant	52.078638	4.304469
2	Grapes&Olives	4cb8d4150180721e0db69061	Wine Bar	52.079949	4.303191
3	Korzo	4b42f16ef964a5201cdb25e3	Theater	52.078690	4.305699
4	Albert Heijn XL	4b10ff34f964a5204c7723e3	Supermarket	52.079887	4.295971

Let's check how many venues were returned for each neighborhood in Rotterdam

```
In [88]: rotterdam_venues.groupby('Neighborhood').count()
```

Out[88]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Groeneweg	4	4	4	4	4	4
Heijplaat	13	13	13	13	13	13
Hoek van Holland	24	24	24	24	24	24
Hoogvliet	28	28	28	28	28	28
Kandelaar	4	4	4	4	4	4
Kralingse Veer	10	10	10	10	10	10
Oud Verlaat	9	9	9	9	9	9
Pernis	4	4	4	4	4	4
Rotterdam	100	100	100	100	100	100
Rozenburg	17	17	17	17	17	17
Terbregge	9	9	9	9	9	9
Zweth	6	6	6	6	6	6

There are 100 unique categories from all the returned venues in Rotterdam.

Let's check how many venues were returned for each neighborhood in Den Haag

```
In [89]: denhaag_venues.groupby('Neighborhood').count()
```

Out[89]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Den Haag (The Hague)	100	100	100	100	100	100
Haagoord	26	26	26	26	26	26

There are 80 unique categories from all the returned venues in Den Haag.

The top 5 venues for each neighborhood in Rotterdam:

```
neighborhoods_venues_sorted.head()
```

Out[97]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Groeneweg	Beach	Flower Shop	Metro Station	Yoga Studio	Diner
1	Heijplaat	Harbor / Marina	Bus Stop	Boat or Ferry	Supermarket	Food Truck
2	Hoek van Holland	Boat or Ferry	Supermarket	Restaurant	Drugstore	Bus Stop
3	Hoogvliet	Supermarket	Gym / Fitness Center	Metro Station	Shopping Mall	Furniture / Home Store
4	Kandelaar	Historic Site	Restaurant	French Restaurant	Monument / Landmark	Fast Food Restaurant

The top 5 venues for each neighborhood in Den Haag:

```
d_neighborhoods_venues_sorted.head()
```

Out[98]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Den Haag (The Hague)	Restaurant	Coffee Shop	Bakery	Bar	Café
1	Haagoord	Restaurant	Supermarket	Soccer Field	Furniture / Home Store	Shopping Mall

Cluster Neighborhoods

Using k-means to cluster the neighborhood into clusters based in Rotterdam.

```
rotterdam_merged.head() # check the last columns!
```

Out[103]:

	Neighborhood	Borough	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Groeneweg	Rotterdam	51.98472	4.59722	0	Beach	Flower Shop	Metro Station	Yoga Studio	Diner
1	Heijplaat	Rotterdam	51.89306	4.42083	1	Harbor / Marina	Bus Stop	Boat or Ferry	Supermarket	Food Truck
2	Hoek van Holland	Rotterdam	51.97778	4.13333	1	Boat or Ferry	Supermarket	Restaurant	Drugstore	Bus Stop
3	Hoogvliet	Rotterdam	51.86250	4.36250	1	Supermarket	Gym / Fitness Center	Metro Station	Shopping Mall	Furniture / Home Store
4	Kandelaar	Rotterdam	51.95694	4.39861	2	Historic Site	Restaurant	French Restaurant	Monument / Landmark	Fast Food Restaurant

Using k-means to cluster the neighborhood into clusters based in Den Haag.

```
denhaag_merged.head() # check the last columns!
```

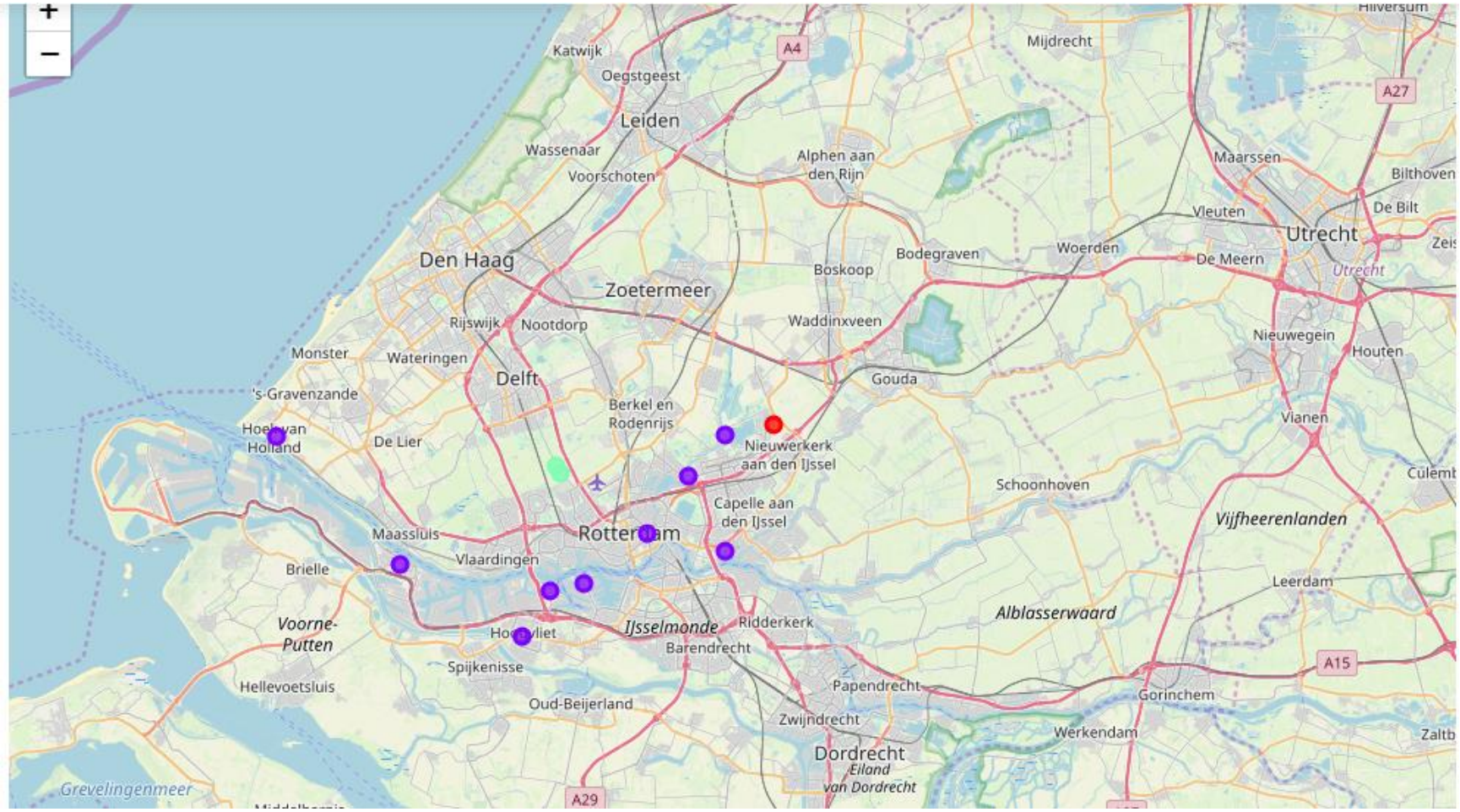
Out[104]:

	Neighborhood	Borough	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Den Haag (The Hague)	Den Haag (The Hague)	52.07639	4.29861	0	Restaurant	Coffee Shop	Bakery	Bar	Café
1	Haagoord	Den Haag (The Hague)	52.05972	4.40417	1	Restaurant	Supermarket	Soccer Field	Furniture / Home Store	Shopping Mall

Visualizing on the map

Finally, let's visualize the resulting clusters

Rotterdam clusters on map



Rotterdam City clusters

```
In [171]: rotterdam_merged.sort_values(by=['Cluster Labels', 'Neighborhood'])
```

Out[171]:

	Neighborhood	Borough	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Groeneweg	Rotterdam	51.98472	4.59722	0	Beach	Flower Shop	Metro Station	Yoga Studio	Diner
1	Heijplaat	Rotterdam	51.89306	4.42083	1	Harbor / Marina	Bus Stop	Boat or Ferry	Supermarket	Food Truck
2	Hoek van Holland	Rotterdam	51.97778	4.13333	1	Boat or Ferry	Supermarket	Restaurant	Drugstore	Bus Stop
3	Hoogvliet	Rotterdam	51.86250	4.36250	1	Supermarket	Gym / Fitness Center	Metro Station	Shopping Mall	Furniture / Home Store
5	Kralingse Veer	Rotterdam	51.91250	4.55278	1	Restaurant	Cosmetics Shop	IT Services	Office	Bar
6	Oud Verlaat	Rotterdam	51.97917	4.55278	1	Pet Store	Diner	Eastern European Restaurant	Bar	Beach
7	Pernis	Rotterdam	51.88889	4.38889	1	Harbor / Marina	Supermarket	Metro Station	Bus Station	Yoga Studio
8	Rotterdam	Rotterdam	51.92222	4.47917	1	Italian Restaurant	Bar	Coffee Shop	Hotel	Vegetarian / Vegan Restaurant
9	Rozenburg	Rotterdam	51.90417	4.24861	1	Supermarket	Gym	Soccer Field	Pizza Place	Drugstore
10	Terbregge	Rotterdam	51.95556	4.51806	1	Harbor / Marina	Tennis Court	Diner	Dog Run	Auto Dealership
4	Kandelaar	Rotterdam	51.95694	4.39861	2	Historic Site	Restaurant	French Restaurant	Monument / Landmark	Fast Food Restaurant
11	Zweth	Rotterdam	51.96111	4.39444	2	Farm	Fishing Spot	Restaurant	French Restaurant	Café

The 1st cluster (cluster label 0) is comprised of various venues which are: Beach, Flower shop, Metro Station, Yoga Studio, Diner. The 2nd cluster (cluster label 1) is very diverse which includes: Restaurants, Supermarket, Offices, Gym, and many more. The 3rd cluster locates Historic sites, Farm, Landmark and Restaurants.

The 2nd cluster is the recommended area in Rotterdam for visitors/travelers.

Den Haag cluster on map

Examine Clusters

Now, we can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, we can then assign a name to each cluster. First Rotterdam city clusters following that Den Haag (The Hague) clusters.

Den Haag city clusters

```
In [172]: denhaag_merged.sort_values(by=['Cluster Labels', 'Neighborhood'])
```

Out[172]:

	Neighborhood	Borough	latitude	longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Den Haag (The Hague)	Den Haag (The Hague)	52.07639	4.29861	0	Restaurant	Coffee Shop	Bakery	Bar	Café
1	Haagoord	Den Haag (The Hague)	52.05972	4.40417	1	Restaurant	Supermarket	Soccer Field	Furniture / Home Store	Shopping Mall

The 1st cluster (cluster label 0) is comprised of various venues which are: Coffee shop, Bakery, Bar and Restaurant. The 2nd cluster (cluster label 1) includes: Restaurants, Supermarket, Soccer Field, Furniture Store and Shopping mall.

The 1st cluster is the recommended area in Den Haag for visitors/travelers.

Section 2 – Satisfactory categorization (likes)

Unique list of venues and through Foursquare number of likes for each of the venue (Restaurant) is retrieved and stored in a new column.

For restaurants in Rotterdam city

```
venueRlist.head()
```

Out[132]:

	name	id	categories	lat	lng	total likes
0	THOMS	5676cb0e498e19ce9f0f106f	Restaurant	51.921856	4.481555	132
1	Roots Rotterdam	50cb82bce4b0262ed982ed88	Restaurant	51.923667	4.477332	30
2	Bertmans	5adb0538018cbb0f94b63923	Vegetarian / Vegan Restaurant	51.920812	4.474312	41
3	Little V	4b5c9d93f964a520173a29e3	Vietnamese Restaurant	51.921807	4.484580	291
4	Backyard	5afadf1adec1d6002cdec318	Vegetarian / Vegan Restaurant	51.919667	4.482279	13

For restaurants in Den Haag city

```
d_venueRlist.head()
```

Out[138]:

	name	id	categories	lat	lng	total likes
0	Kua	52f529ae11d203e7b26d4453	Mexican Restaurant	52.078638	4.304469	64
1	Juni	4b409643f964a52029b925e3	Restaurant	52.076651	4.306936	36
2	Hagedis	4b254ce1f964a520986f24e3	Vegetarian / Vegan Restaurant	52.078356	4.292688	16
3	Giuliano's	4b1ff42cf964a520462b24e3	Italian Restaurant	52.076784	4.307886	87
4	Bøg	54cbbea7498ede96b0832ffa	Scandinavian Restaurant	52.079956	4.303943	18

Finding the statistical analysis of the data frame.

```
#Rotterdam
venueRlist['total likes'].describe()
```

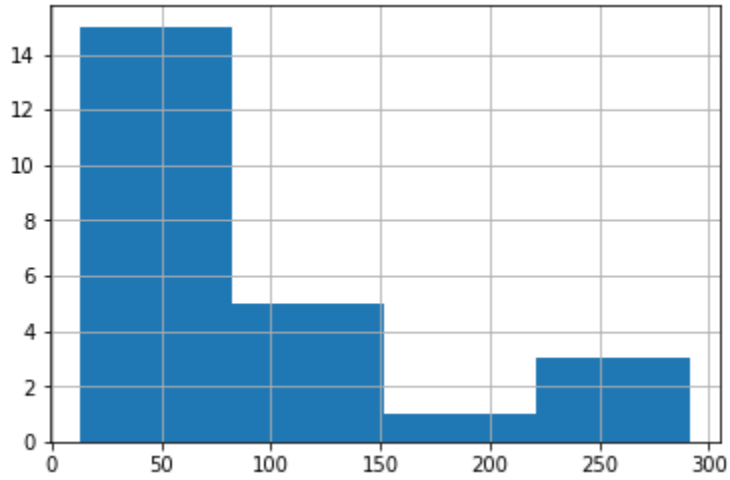
```
count     24.000000
mean      83.708333
std       80.233615
min       13.000000
25%       30.750000
50%       46.500000
75%       99.000000
max      291.000000
Name: total likes, dtype: float64
```

```
#Den Haag
d_venueRlist['total likes'].describe()
```

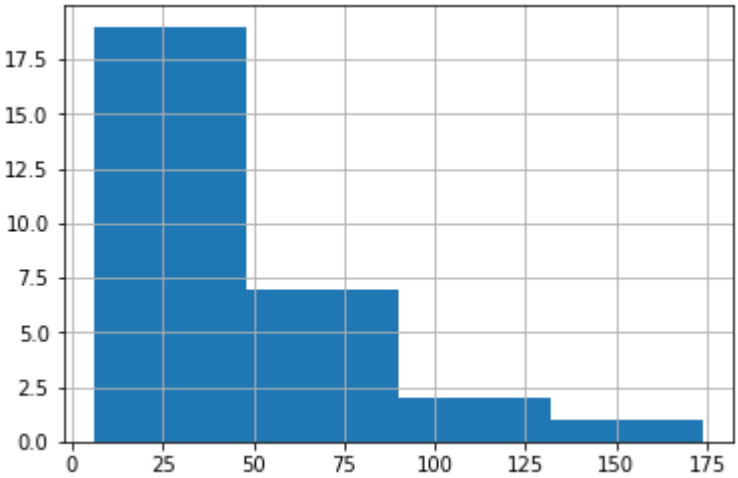
```
count     29.000000
mean      41.931034
std       37.367797
min        6.000000
25%       18.000000
50%       31.000000
75%       54.000000
max      174.000000
Name: total likes, dtype: float64
```

Histogram is plotted for the total number of likes based on the number of total likes category.

Rotterdam city restaurants likes



Den Haag city restaurants likes



Cluster Neighborhoods

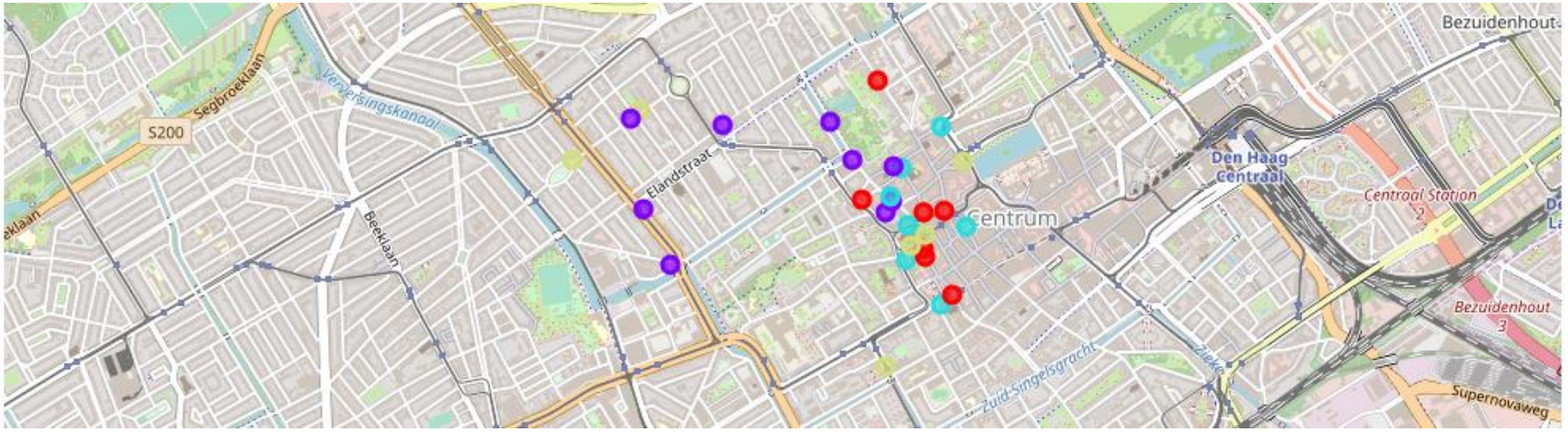
Running k-means to cluster the neighborhood into clusters

Rotterdam city restaurant in clusters with satisfactory category (likes)

	name	id	categories	lat	lng	total likes	satisfactory_category	cluster labels
0	THOMS	5676cb0e498e19ce9f0f106f	Restaurant	51.921856	4.481555	132	Excellent	2
1	Roots Rotterdam	50cb82bce4b0262ed982ed88	Restaurant	51.923667	4.477332	30	Poor	3
2	Bertmans	5adb0538018cbb0f94b63923	Vegetarian / Vegan Restaurant	51.920812	4.474312	41	Fair	0
3	Little V	4b5c9d93f964a520173a29e3	Vietnamese Restaurant	51.921807	4.484580	291	Excellent	2
4	Backyard	5afadf1adec1d6002cdec318	Vegetarian / Vegan Restaurant	51.919667	4.482279	13	Poor	3

Den Haag city restaurant in clusters with satisfactory category (likes)

	name	id	categories	lat	lng	total likes	satisfactory_category	cluster labels
0	Kua	52f529ae11d203e7b26d4453	Mexican Restaurant	52.078638	4.304469	64	Excellent	0
1	Juni	4b409643f964a52029b925e3	Restaurant	52.076651	4.306936	36	Good	2
2	Hagedis	4b254ce1f964a520986f24e3	Vegetarian / Vegan Restaurant	52.078356	4.292688	16	Poor	1
3	Giuliano's	4b1ff42cf964a520462b24e3	Italian Restaurant	52.076784	4.307886	87	Excellent	0
4	Bøg	54cbbea7498ede96b0832ffa	Scandinavian Restaurant	52.079956	4.303943	18	Poor	1



Results

Examining Clusters

Now, examining each cluster and determining the discriminated restaurants categories that distinguish each cluster. Based on the defining categories, a name to each cluster is assigned.

Rotterdam city restaurants with satisfactory category of excellent (most liked by the users through Foursquare API)

	name	id	categories	lat	lng	total likes	satisfactory_category	cluster labels
0	THOMS	5676cb0e498e19ce9f0f106f	Restaurant	51.921856	4.481555	132	Excellent	2
3	Little V	4b5c9d93f964a520173a29e3	Vietnamese Restaurant	51.921807	4.484580	291	Excellent	2
16	Happy Italy	4dd69beae4cd37c89396c024	Italian Restaurant	51.922947	4.485218	251	Excellent	2
18	Spirit	4b4e0d42f964a520b5df26e3	Vegetarian / Vegan Restaurant	51.924282	4.488177	189	Excellent	2
21	Vapiano	510bb50be4b0ef517b7cd1df	Italian Restaurant	51.922971	4.471810	240	Excellent	2
23	Trattoria Gusto	4b0c4fb2f964a520f73a23e3	Italian Restaurant	51.916827	4.480223	102	Excellent	2

Den Haag city restaurants with satisfactory category of excellent (most liked by the users through Foursquare API)

	name	id	categories	lat	lng	total likes	satisfactory_category	cluster labels
0	Kua	52f529ae11d203e7b26d4453	Mexican Restaurant	52.078638	4.304469	64	Excellent	0
3	Giuliano's	4b1ff42cf964a520462b24e3	Italian Restaurant	52.076784	4.307886	87	Excellent	0
11	Bar & Restaurant Milú	53540f30498e27fc50d7534a	Restaurant	52.078222	4.307831	110	Excellent	0
12	SET Restaurant	4e592daa2271886714efcfc0	Japanese Restaurant	52.077037	4.307828	91	Excellent	0
15	Irodion	54f1ab4b498e6903bf159b9e	Greek Restaurant	52.075522	4.309413	77	Excellent	0
23	Het Heden	4b63282cf964a52019672ae3	Restaurant	52.082638	4.305330	63	Excellent	0
28	't Goude Hooft	4b8d6472f964a52008f932e3	Restaurant	52.078294	4.308920	174	Excellent	0

Discussion

Using Foursquare and retrieving 100 venues and in particular focusing on category in which the word restaurant was stated, has helped to have consistency in identification of a specific category in this project. Although, I was able to prepare a list of restaurants marked on the map of both cities in clusters. However, in determining the satisfactory category, which was through the Foursquare venue users likes, the statistic shows that the number of likes for the restaurants in Rotterdam city had a higher range than the likes for the restaurants in The Hague. For a restaurant in Rotterdam in order to be categorized "Excellent" at least 99 likes were needed, whereas, for restaurants in The Hague only more than 54 likes were required. Future project could take such variations in their analysis.

Conclusion

Rotterdam city

The 1st cluster (cluster label 0) is comprised of various venues which are: Beach, Flower shop, Metro Station, Yoga Studio, Diner. The 2nd cluster (cluster label 1) is very diverse which includes: Restaurants, Supermarket, Offices, Gym, and many more. The 3rd cluster locates Historic sites, Farm, Landmark and Restaurants.

The 2nd cluster is the recommended area in Rotterdam for visitors/travelers.

Recommended restaurants are as follows

- Little V
- Happy Italy
- Vapiano
- Spirit
- THOMS
- Trattoria Gusto

The Hague city (Den Haag)

The 1st cluster (cluster label 0) is comprised of various venues which are: Coffee shop, Bakery, Bar and Restaurant. The 2nd cluster (cluster label 1) includes: Restaurants, Supermarket, Soccer Field, Furniture Store and Shopping mall.

The 1st cluster is the recommended area in Den Haag for visitors/travelers.

Recommended restaurants are as follows

- 't Goude Hooft
- Bar & Restaurant Milú
- SET Restaurant
- Giuliano's
- Irodion
- Kua
- Het Heden

Thank you for reviewing this document.

References

Christopher Jameson: https://medium.com/@chriswjameson/ibm-data-science-professional-certificate-capstone-project-blog-df3044cffe24

Foursquare: https://developer.foursquare.com/docs/api/endpoints

Google search Rotterdam: https://www.google.com/search?q=rotterdam&oq=ro&aqs=chrome.0.69i59j69i60l3j69i57j69i59.1193j0j4&sourceid=chrome&ie=UTF-8

Google search the Hague: https://www.google.com/search?ei=AMUDXbLDDLDjkgXz5LkI&q=the+hague&oq=the+&gs_l=psy-ab.3.0.0i67l6j0l4.75086.77585..78796...0.0..0.179.1104.8j4.....0....1..gws-wiz.....0..0i71.chRLr4Ni8GA

Wikipedia: https://en.wikipedia.org/wiki/List_of_cities,_towns_and_villages_in_South_Holland