

Capstone Project - The Battle of Neighborhoods

Retrieving insights on Medditerean/Moroccan restaurants establishments in Rotterdam, The Netherlands.



Author: Rachid de Graaff

Date: 21 - 04 - 2020

1.Introduction

This report attempts to present a dedicated exploration of the city of Rotterdam in The Netherlands. In addition, the exploration goes in pair with retrieving information of Medditerean/Moroccan restaurants. As such, with the retrieved insights one can determine the business opportunities in potential areas of the city. At least it gives a good representation where to explore the excellent cuisine of Medditerean/Moroccan food.

In order to grasp the information adequately of the restaurants in Rotterdam, Geoplots will be the main visualization applied. In particular, the retrieved data on restaurants will be plotted on maps and an attempt to determine clusters (k-means clustering) upon areas is shown. From there, one can identify areas which are more dense with restaurants with respect to other areas. Thus, potential locations for a new restaurant can be determined accordingly.

The target audience of the project are those who are interested in establishing/moving a Medditerean/Moroccan restaurant in an area in Rotterdam, or simply want to know where to enjoy a great meal.

2.Data acquisition and cleaning

2.1 Data Sources

The data acquired for the assignment is a combination of data from three sources. The first data source is related to acquiring the exact geographical city center of Rotterdam. For the sake of scope, the city center location is the Coolsingel. From that point, the centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Bing Maps API reverse geocoding.

The second data source is regarded to the number of Meditterean/Moroccan restaurants and exact location in every candidate area in Rotterdam. These locations will be obtained using Foursquare API. Finally, to mark the areas of each sector in Rotterdam, a neighborhood border geojson dataset is used provided by github:blackmad¹.

2.2 Data cleaning and preparation

Data cleaning and preparation is done for each source accordingly. First, the data retrieved from the Bing Map API for the candidate areas are constructed in a dataframe with the following columns:

¹ <https://github.com/blackmad/neighborhoods/blob/master/README.md>

- **Address:** The address of the center of the candidate areas
- **Latitude:** geographical coordinate for geoplotting purposes
- **Longitude:** geographical coordinate for geoplotting purposes
- **X:** X-coordinate for a two-dimensional Cartesian system
- **Y:** Y-coordinate for a two-dimensional Cartesian system
- **Distance from center:** The distance to the Coolsingel in meters.

	Address	Latitude	Longitude	X	Y	Distance from center
0	3085 Rotterdam	51.869029	4.486977	-222846.796994	5.798896e+06	5983.101203
1	Kapelburg 296, 3085 HV Rotterdam	51.869803	4.495545	-222246.796994	5.798896e+06	5983.101203
2	3084 Rotterdam	51.870908	4.455905	-224946.796994	5.799416e+06	5960.494946
3	Kruiningenstraat 127, 3086 KS Rotterdam	51.871684	4.464472	-224346.796994	5.799416e+06	5745.215401
4	Tiengemetensingel 34A, 3086 NR Rotterdam	51.872460	4.473040	-223746.796994	5.799416e+06	5586.367335
5	Biezeveld 11, 3085 RD Rotterdam	51.873235	4.481609	-223146.796994	5.799416e+06	5488.852339
6	Larenkamp 32A, 3085 GC Rotterdam	51.874009	4.490178	-222546.796994	5.799416e+06	5455.960044
7	Diepenhorst 32A, 3085 WL Rotterdam	51.874783	4.498747	-221946.796994	5.799416e+06	5488.852339
8	Vaanweg, 3076 Rotterdam	51.875556	4.507317	-221346.796994	5.799416e+06	5586.367335
9	Elsschotstraat 12, 3076 EA Rotterdam	51.876329	4.515887	-220746.796994	5.799416e+06	5745.215401

Table 1: Dataframe with the candidate areas

The second dataset is constructed on data retrieved from the Foursquare API with information of Mediterranean/Moroccan restaurants in Rotterdam. The dataset is constructed in a dataframe with the following columns:

- **Address:** The address of the restaurant
- **ResName:** The name of the restaurant
- **Latitude:** geographical coordinate for geoplotting purposes
- **Longitude:** geographical coordinate for geoplotting purposes
- **X:** X-coordinate for a two-dimensional Cartesian system
- **Y:** Y-coordinate for a two-dimensional Cartesian system
- **Distance to Rdam center:** The distance of the restaurant to the city center.

	Address	ResName	Latitude	Longitude	X	Y	Distance to Rdam center
0	NaN	Tekbir Döner	51.931725	4.482236	-222157.284492	5.805888e+06	1088.107007
1	Nieuwe Binnenweg 152	Ömür	51.914523	4.465676	-223569.996219	5.804148e+06	1253.050568
2	Nieuwe Binnenweg 65 A	Ten To Ten	51.916431	4.469877	-223251.322890	5.804318e+06	896.323174
3	Museumpark	La Cantine & Haute Friture	51.913204	4.472272	-223139.548859	5.803936e+06	1107.550807
4	Stationsplein 1	Cafe Restaurant 't Centrum	51.920952	4.407998	-227415.665964	5.805440e+06	4901.949821
5	NaN	Bar Pulpo	51.905760	4.488081	-222177.110878	5.802953e+06	1953.586778
6	Rotterdamse Rijnweg 128	beyaz saray	51.933617	4.427559	-225870.013237	5.806646e+06	3767.184802
7	Burgemeester baumannlaan 164	FP Markt	51.935057	4.428039	-225813.679450	5.806801e+06	3793.839687
8	Kruisplein 153	Ayla	51.921200	4.471302	-223076.462280	5.804832e+06	531.188297
9	NaN	Grillroom Hadarom	51.898307	4.511255	-220709.736332	5.801896e+06	3496.773972

Table 2: Dataframe with the restaurants

Finally, the borders of areas in Rotterdam are retrieved in a geojson file. Figure 1 shows how the data looks like and figure 2 is a geoplot presentation of the data.

Figure 1: borders of areas in Rotterdam as geojson file



3.1 Exploratory Data Analysis

3.1.1 Areas of interest in Rotterdam

As the fact that the Coolsingel is taken as the city center point, the next task is to determine the areas of interest to investigate further for our case. As such, a grid of area candidates which is equally spaced around the city center within approx. 6 kilometers from the Coolsingel. The areas are subsequently plotted as circular areas with a radius of 350 meters, which implies the centers to be 600 meters apart. To accurately calculate the distance, it is necessary to create the grid of locations in a Cartesian 2D coordinate system

which allows to calculate distances in meters as opposed to latitude/longitude degrees. From there, those coordinates will be projected back to latitude/longitude degrees shown on the folium map in figure 3. A total of 361 candidate areas are defined with the requirements.

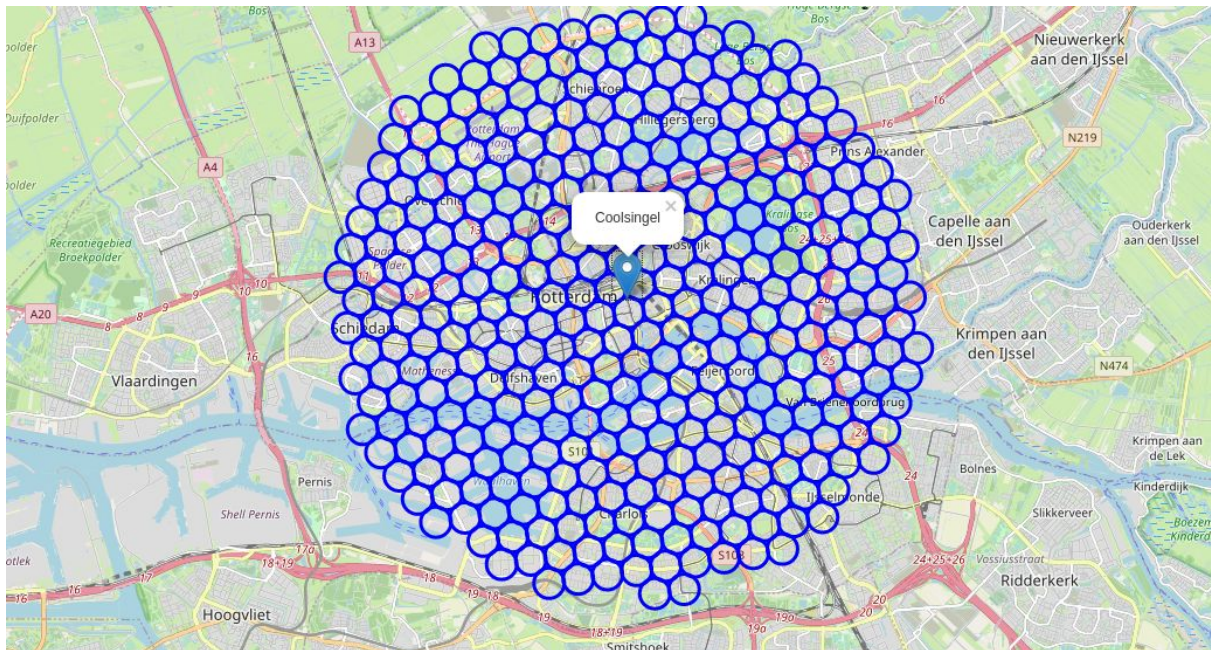


Figure 3: Candidate areas plotted as circular areas around the city center Coolingsingel

3.1.2 Medditerean/Moroccan Restaurants in Rotterdam

After defining the candidate areas, the next task is to retrieve and investigate the distribution of the location of the Medditerean/Moroccan restaurants. Important here is to determine which candidate areas include restaurants around the city center. Figure 4 shows the distribution of the restaurants in the candidate areas around the city center.

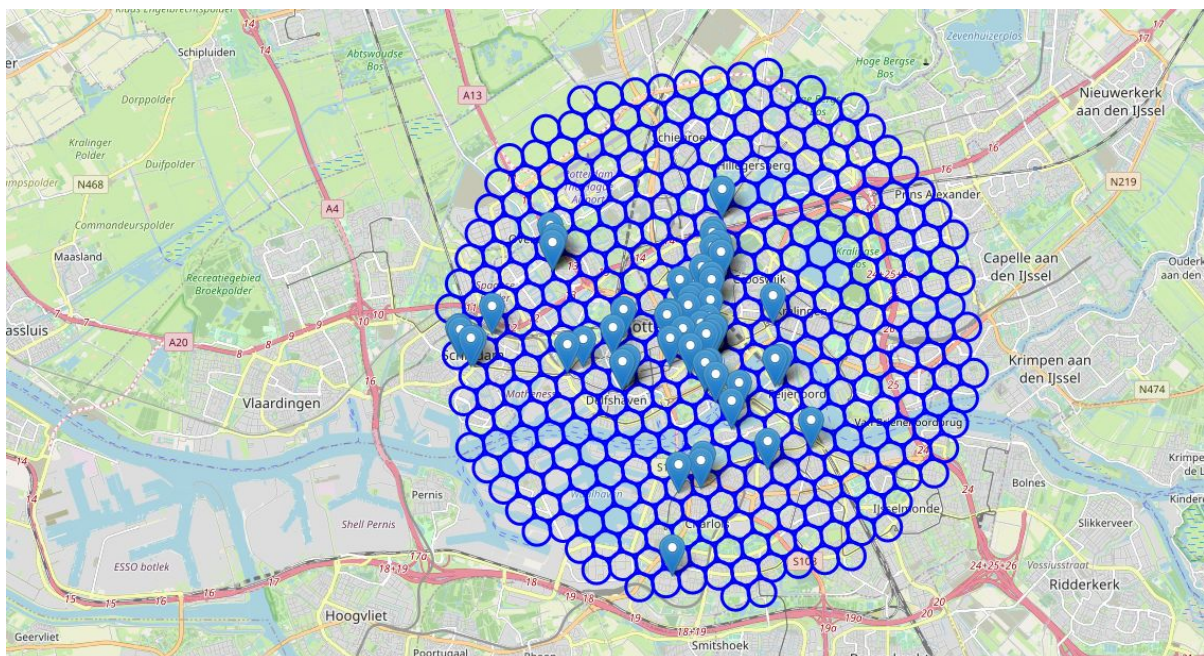


Figure 4: Medditerean/Moroccan restaurants plotted in the candidate areas

With the information retrieved above it concludes the exploratory data analysis. In the next chapter, the data is used to identify clusters using k-means clustering.

4. Modeling

4.1 Identify clusters

In the previous section, the restaurants retrieved were plotted on the map within the candidate areas. The next step is to identify the restaurant density across the candidate areas of Rotterdam. In order to do so accordingly, a heatmap is created to optically identify promising candidate areas close to the city center with low number of Medditerean/Moroccan restaurants. Figure 5 presents the heatmap of the location of restaurants considered in the candidate areas. Note that the higher the density of the number of restaurants within a certain candidate area, the warmer the color.

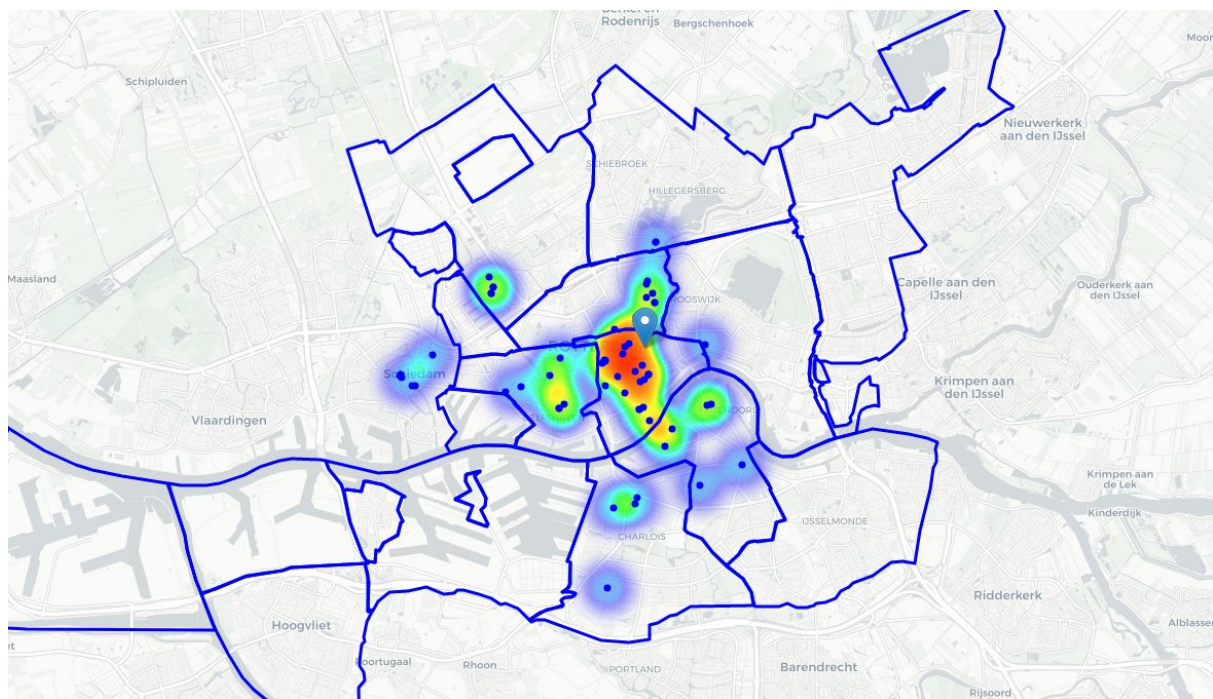


Figure 5: density of restaurants in candidate areas based on their location

As one can see in figure 6, the heatmap suggests the existence of clusters around the city center of Rotterdam. By applying the k-means clustering algorithm, an attempt is made to determine the existence of clusters. I have chosen to take $k=15$ as the number of possible clusters to take into consideration. After determining the clusters, a plot on the heat map is made to optically identify the clusters. Indeed, there are a significant number of clusters around the city center with groups of restaurants. Indeed, one should consider the area of clusters with a minimum and maximum number of corresponding restaurants to start one.

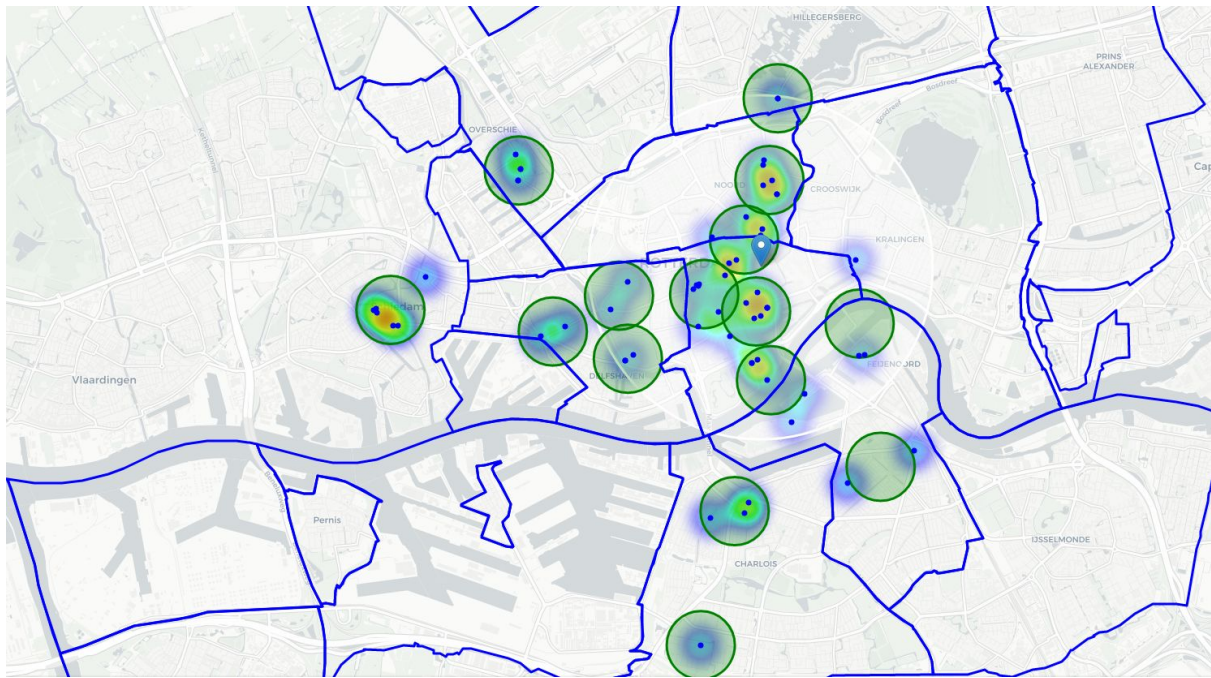


Figure 6: 15 Identified clusters using k-means clustering and plotted on the heatmap.

5. Conclusion

The purpose of this project was to identify Mediterranean/Moroccan restaurants in candidate areas from the city center of Rotterdam and identify areas with low density of restaurants in order to aid stakeholders in narrowing down the search for optimal location for a new business venture. The overall analysis shows that although there is a great number of restaurants in Rotterdam, there are interesting areas of low restaurant density fairly close to the city center as seen in figure 6. The highest concentration of restaurants is clearly in areas close by the city center. This is explainable due to the fact that the city center is the place masses gather. The remaining interesting clusters can be explained by a combination of popularity among tourists, closeness to the city center and strong socio-economic dynamics. Nonetheless, the clusters show a lower restaurant density. From there, these areas are recommended for potential business ventures and should therefore be considered as a starting point for more detailed analysis.

6. Future work

Indeed, one should consider the area of clusters with a minimum number of corresponding restaurants and high potentials within the candidate area. In order to come to a final decision on an optimal restaurant location, further research is essential. Based on specific characteristics of neighborhoods and locations in every recommended area, investors have to take into account additional factors like attractiveness of each location (proximity to park or water), levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc. Thus, further analysis is recommended per designated candidate area with the appropriate level of restaurant density.