Rotterdam – battle of neighborhoods

Coursera Capstone project

Anastasia Khomenko

November 6, 2019

Contents

Introduction

Background

Rotterdam is a very lively and bustling city that attracts students, expats, and tourists. Alongside Porto, Rotterdam was European Capital of Culture in 2001. The city has its own orchestra, the Rotterdam Philharmonic, with its well-regarded young music director Yannick Nézet-Séguin; a large congress and concert building called De Doelen; several theaters (including the new Luxor) and movie theatres; and the Rotterdam Ahoy complex in the south of the city, which is used for pop concerts, exhibitions, tennis tournaments, and other activities[1]. It is, therefore, no wonder that the city is always full of live and people. Therefore, the environment for entertainment industry is quite competitive, but at the same time all the hype creates a lot of opportunities for development.

Problem

This data analytics project will concern choosing the best location in Rotterdam to set up a cafe. The Neighborhood should ideally be trendy and visited by people, but there should not be too many cafes already so that the competition is not too high. Therefore, the question is: Which neighborhood is the best to set up a restaurant given that Rotterdam is a competitive environment for cafes?

Interest

Obviously, the future owner of the café would like to know the outcome of the analysis. Additionally, entrepreneurs in the entertainment industry can use the materials of the analysis to determine 1) level of competition in any given area; 2) estimate the amount of people in any given area, taking into account the assumption that where there are more restaurants, there are more people.

Data

Data sources

2 key datasets were used in the analysis. The first dataset was taken from the Rotterdam City website[2]. This data concerned the information on the main city neighborhoods. Please, find an example of how the dataset looked like in table 1.

Table 1 – Dataset 1 columns.

Index	DistrName	DistrNum	DistrCode	Latitude	Longitude

The second dataset used was generated with Foursquare API. Taking neighborhoods centres as starting location, Foursquare worked through all the locations in 2.5km range and returned them into a dataset. The dataset looked as presented in table 2.

Table 2 – Dataset 2 columns.

id	Venue	Category	Latitude	Longitude	address	city	Distance	Postcode

[1] https://en.wikipedia.org/wiki/Rotterdam#Culture
[2] ***https://www.rotterdam.nl***

Data cleaning

Data preparation and cleaning was one of the most important steps and it required several actions to be taken. This section will summarize the main steps taken in order to prepare the datasets for analysis.

First of all, the two datasets needed to be combined into one and venues needed to be allocated to different neighborhoods based on their locations. For this, a function was created that checked for neighborhood centers within radius of 2.5km and allocated venues to one or another neighborhood. Rotterdam neighborhoods have a radius of 3km, on average, therefore, the overlaps were not likely to happen[3].

From the initial dataset, only a few features of interest were selected. These included:

- Neighborhood Name;
- Neighborhood Latitude;
- Neighborhood Longitude;
- Venue;
- Venue Category;
- Venue Longitude;
- Venue Latitude;

The rest of the features, such as address or location id were not required as we are not interested in learning specificities about particular locations. Venue names, location data and category, were preserved as they provide more information on the location and help to get an overview and "flavor" of restaurant density in Rotterdam.

Second, Foursquare API generated data on all the venues in Rotterdam. However, the interest for this analysis is only places where you can have a drink and eat a dessert. Therefore, the next step was to sort through the locations and ensure that venue categories such as metro stations or historical site are removed.

Duplicates was not an issue because, as mentioned previously, Rotterdam neighborhoods have a 3km radius, on average, and the search was for 2.5km radius from neighborhood center. A quick check for duplicates confirmed that there were none.

Finally, I checked for missing data to ensure to remove it. Mostly, I was interested in missing location data and venue category data. However, no issues were present.

The data was cleaned and prepared for the analysis.

Methodology

In this project we will direct our efforts on detecting areas of Rotterdam that have low restaurant & cafe density, particularly those with low number of places where you can get a drink & dessert. We will limit our analysis to the 26 city districts.

[3] https://en.wikipedia.org/wiki/Rotterdam

In first step we have collected the required data: location and type (category) of every restaurant in the 26 districts. We have also identified which venues are of interest for us.

Second step in our analysis will be calculation and exploration of 'restaurant density' across different areas of Rotterdam - we will use heatmaps to identify a few promising areas close to center with low number of restaurants in general and focus our attention on those areas.

In third and final step we will generate the most promising areas, within 2.5km range from neighborhood center, and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will show locations with low restaurant & cafe density, and we want locations that are closer to the city center. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Please note. There is no aim to identify the locations that do not have any cafes or restaurants nearby. The aim is to bring visibility of cafe-dense areas and suggest several locations within the busiest and most centered neighborhoods, even if they fall into "heated" areas.

Data exploration

Understanding the data

The final dataset obtained from data cleaning consisted of 249 venues. Additionally, Rotterdam was established to be quite dense with cafes: on average, there were determined to be ≈ 10 venues of interest per neighborhood. At the same time, undoubtedly, some neighborhoods had more venues than the others. Table 3 below shows 5 most “venue-intense” locations.

Table 3 – Neighborhoods with the biggest amount of restaurants.

Neighborhood	**Num of Rest**
Delfshaven	18
Noord	17
Kralingen-Crooswijk	17
Hillegersberg-Schiebroek	17
Hoek van Holland	16

It is interesting that Kralingen-Crooswijk (the student hub of the city) and Stadscentrum are either low on the list or not even present. However, this is good for us since we are interested in setting up a café as close to the city center as possible, therefore, the fact that these 2 neighborhoods are less café-intense is good.

Furthermore, we can definitely conclude that Delfshaven is a poor candidate as it has a lot of venues already. Hoek van Holland also looks attractive as the number of venues is not too high. However, this neighborhood is quite far from the city center.[4]

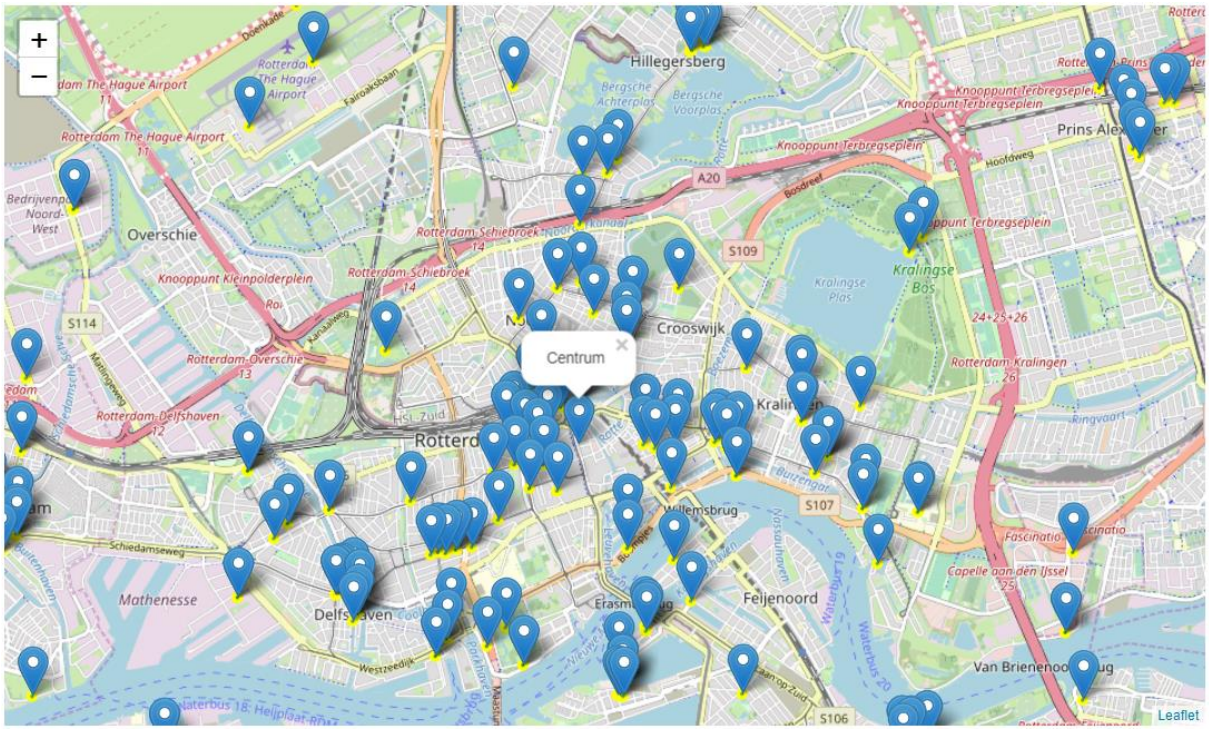
Data visualization

To further understand the data, it was put on a folium map, together with venue categories as labels, so that stakeholders could check what kind of venues are located at any given location.

Visualization clearly shows where most restaurants are situated and how they are split in between the neighborhoods. Indeed, the high amount of restaurants in Delfshaven is immediately apparent. Additionally, low density in Kralingen-Crooswijk and Stadscentrum becomes visible too.

Finally, all the different locations and neighborhoods get visualized against the city center. Therefore, it becomes apparent that Kralingen as well as Stadscentrum are the best candidates for further analysis because: they do not have too many restaurants already. Additionally, they are close to the city center. These were one of the main criteria identified by stakeholders which the two neighborhoods satisfy.

Picture 1. Venues visualized on Rotterdam map.



The marker for the city center was pre-chosen to increase image readability.

[4] https://www.rotterdam.nl

Analysis

Assessment of neighborhoods

After the two most promising neighborhoods were defined, it was necessary to explore some reviews online.

Rotterdam Stadscentrum is famous for sights like: Market Hall, Euromast, Beurstraverse (Koopgoot) with the Beurs-World Trade Center, Lijnbaan, Coolsingel with the city hall and Hofplein, Erasmusbrug, Willemsbrug, Various stations of the Rotterdam Metro, Grote of Sint-Laurenskerk, Library Rotterdam, Cube houses.

Reviews on Booking.com describe Centrum as: "This neighborhood is a great choice for travelers looking for museums, shops, and food." Furthermore, it is believed to be: "Today de Centrum is renowned for its innovative architecture which includes several modern masterpieces such as the Cube House complex and Rotterdam's state-of-the-art Markthal."

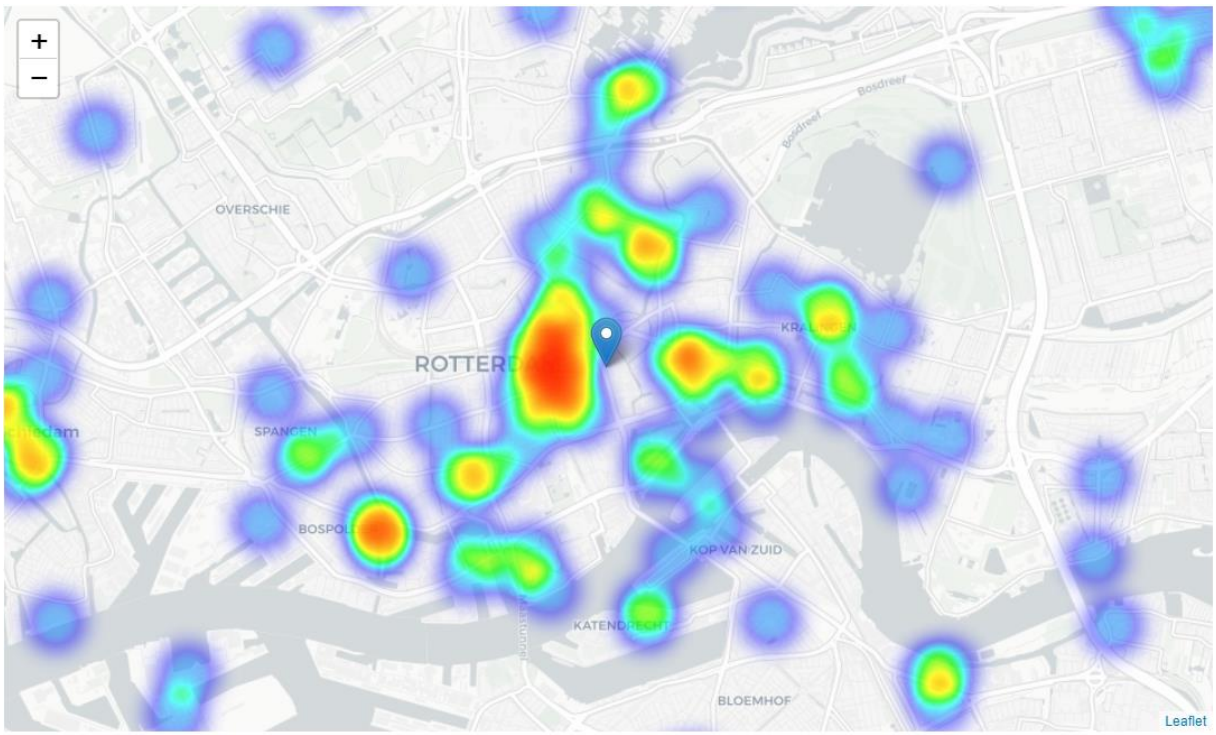
As for Kralingen, this neighborhood is more popular among the student population. Therefore, it is always full of young and vivid souls.

CityRotterdam describes the neighborhood as: "Kralingen is a green and attractive neighborhood, traditionally one of the richer areas of Rotterdam. Well-known places are the recreation area Kralingse Bos, the student pubs around Oostplein and the Erasmus University." And Agoda.com states that: "Kralingen-Crooswijk supplies the perfect mix of tranquility and entertainment. There are also several impressive landmarks to visit."

Therefore, both of these neighborhoods are full of live. They could be great candidates for a future cafe.

Rotterdam heatmap creation

Picture 2. Rotterdam heatmap.



To further understand what dense areas are, the heatmap was generated using folium. The number of restaurants in a given location was the basis on which the heatmap was color coded. For easiness of understanding, the marker in the middle shows the city center.

The heatmap can serve as a starting point of assessment of where to locate a restaurant for entrepreneurs. Additionally, it brings visibility of the busiest areas of Rotterdam, therefore, entrepreneurs who are interested to sell other services than dining, could try to stay within the busiest area to get more traction. Thus, the heatmap can solve multiple purposes and can be the beginning of another analysis in the future.

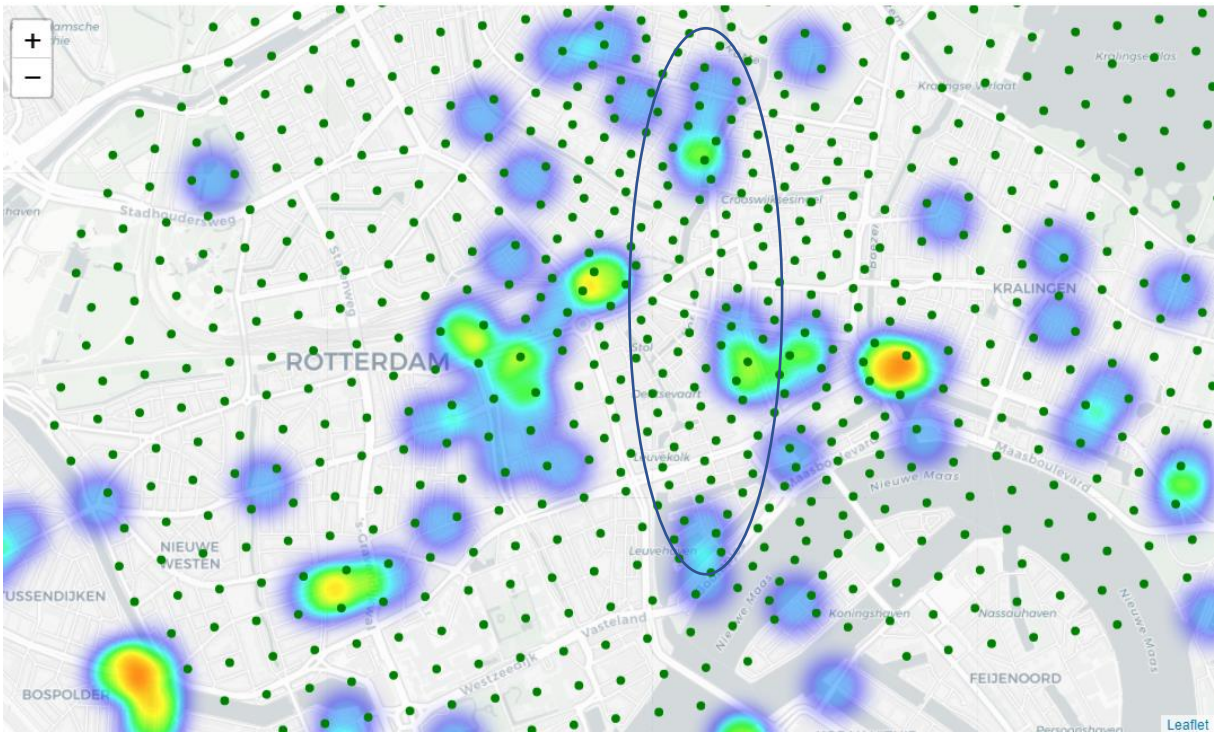
Generation of location candidates

Having decided on the neighborhood, we decided to generate candidate locations for a future café. Random location points were created with a step of 200 meters between each other. The location points were generated within 2.5km range from the centers of the 2 chosen neighborhoods, Kralingen and Stadscentrum.

The location candidates were stored in a dataset. Then, they were visualized with help of folium maps. Picture 3 depicts the result.

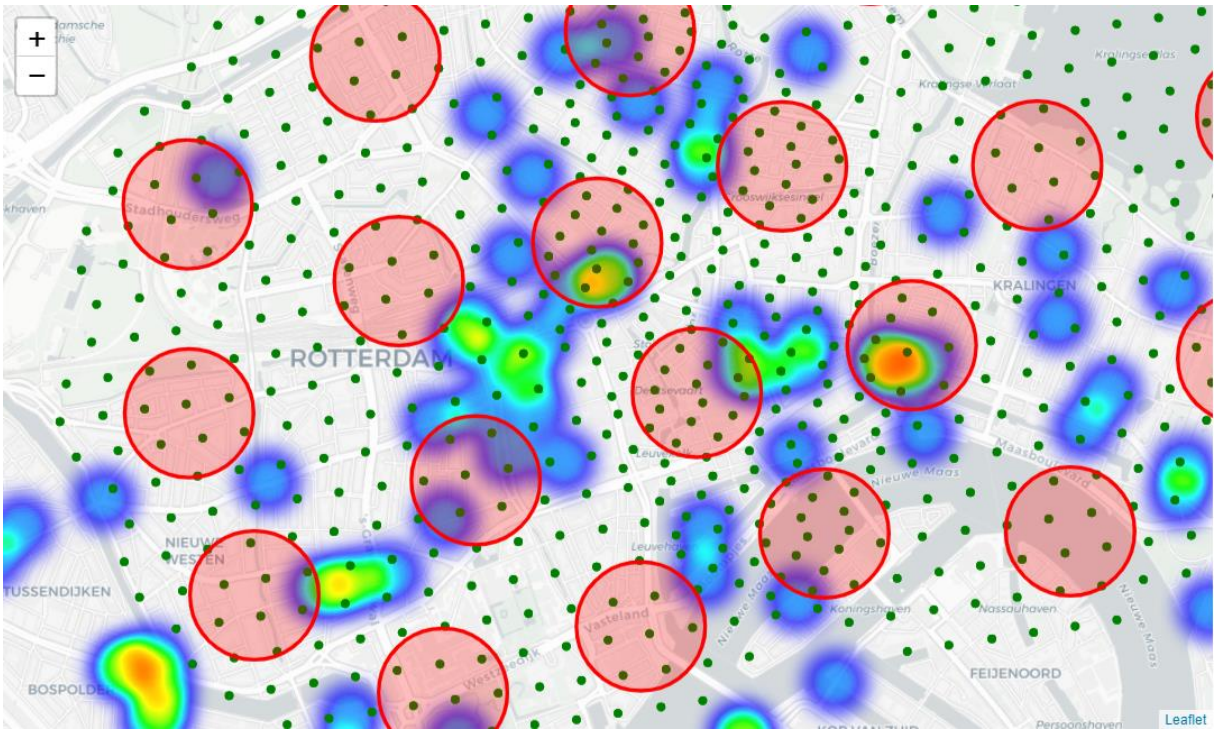
The map generated showed an area where the amount of dots was more dense. This is the most promising area because it does not have a lot of venues already. Additionally, it is located in between two “good” neighborhoods. Finally, it is close to the city center.

Picture 3 – The generated location candidates



Now, we can explore further and narrow down the amount of location candidates. To do this, we will use K Means and cluster the dots.

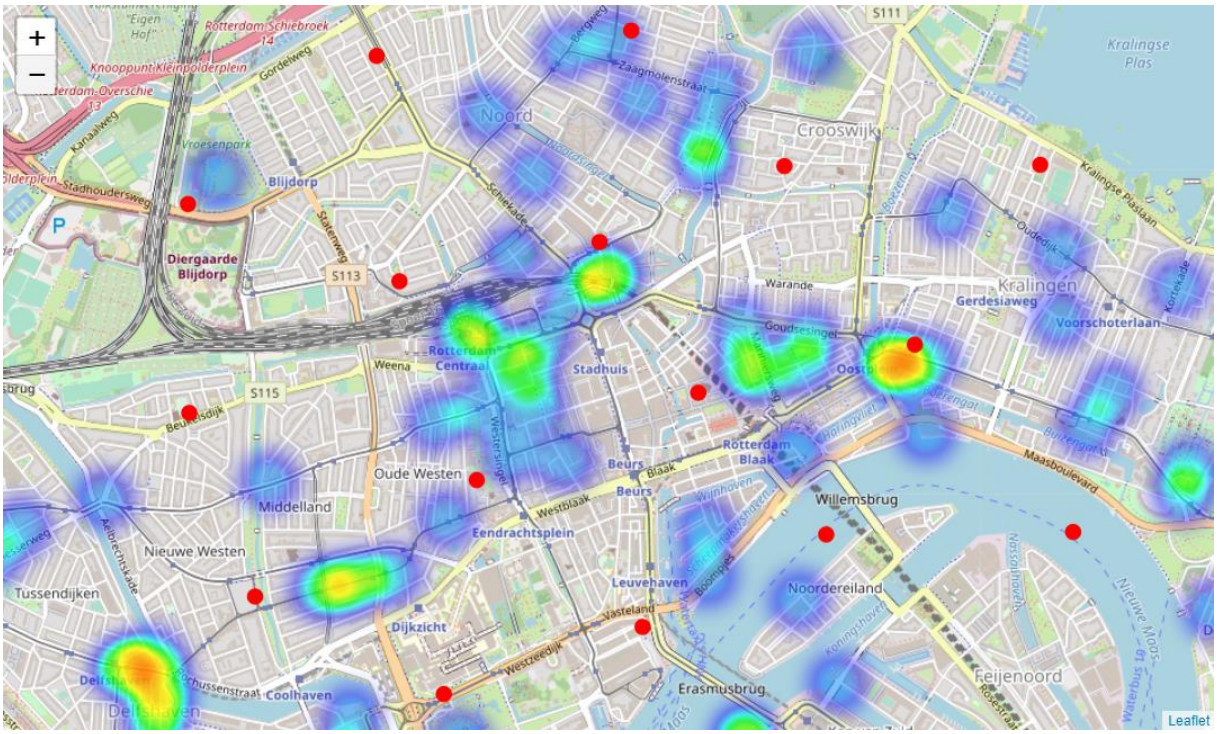
Picture 4 – Clusters generated via K-Means



The clusters were generated in a way that every cluster consists of the biggest amount of dots. Clusters are visualized on picture 4.

Finally, to be more specific, candidate locations were generated. Cluster centers were determined to be the future candidates locations.

Picture 5 – Clusters centers – suggested locations



Therefore, the final picture is the result of the analysis:

1. It shows a heatmap of Rotterdam venues that entrepreneurs can use for different purposes;
2. It shows possible locations for the café that are situated close to the city center, within 2.5km range from neighborhood center.

Results

According to the main aims of the project, the following results were achieved:

1. Heatmap of Rotterdam café-dense areas was generated to bring visibility of area with a lot of cafes/restaurants. The heatmap can be used by entrepreneurs to plan future locations for their businesses.
2. Out of the empty pockets available, the ones closest to the city center were chosen as proximity to the Centrum was one of the requirements by the stakeholder.
3. The reviews of two neighborhoods with not too many cafes, namely Centrum and Kralingen, were checked. All of them indicated that these 2 neighborhoods are lovely and popular.
4. Locations situated within 2.5km from neighborhood center were generated. At one place, they were more dense, indicating that the area is better for a future café.
5. The locations were clustered and cluster centers were defined to be the suggested locations for future cafes.

Discussion

Our analysis shows that although there is a great number of restaurants and cafes in Rotterdam (~600 in our initial area of interest which was 2500 meter from each neighborhood center), there are pockets of low restaurant density fairly close to city center. Highest concentration of restaurants was detected west from the city center and in Neighborhoods such as Noord, Kralingen, Delfshaven, Schiebroek, Hoek van Holland. Therefore, we decided to focus on the more eastern part of Rotterdam, specifically east Stadscentrum and Kralingen-Crooswijk neighborhoods. Both of these neighborhoods are reported to be high in popularity, either because of tourists & expats or students. Therefore, there was a lot of potential identified for these areas.

After directing our attention to this narrower area of interest we first created a dense grid of location candidates (spaced 200m apart). Since the stakeholder wanted to get an approximation of any location he can get, the locations were not filtered further. However, the heatmap was left for visibility to ensure that the stakeholder can check whether any of the given locations is in the heated area or not.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates.

Result of all this is 20 zones containing largest number of potential new cafe locations. This, of course, does not imply that those zones are actually optimal locations for a new restaurant! Purpose of this analysis was to only highlight what locations are full with

restaurants & cafes and then generate suggestions as to where new location candidates can sit. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no nearby competition but also other factors taken into account and all other relevant conditions met.

Limitations and future directions

There are several recommendations that can be made as a result of the analysis:

1. More limitations and factors need to be taken into account for further analysis. Currently, we only checked for the number of restaurants in 2.5km range from neighborhood center. However, further analysis could explore, for instance, what are the trendy areas at any given moment and bring this data to the map.
2. Based on the analysis and initial constraints, 20 recommended location points were generated as a starting point. The project, though, does not separate the suggested areas into 'good' and 'bad' based on the amount of cafes nearby as it was not the aim of the project. However, one could further explore which of these locations are situated far away from other venues.
3. A heatmap of Rotterdam with the current amount of restaurants was visualized. This can help with future explorations or it can serve as a starting point for further analysis of other neighborhoods.

Conclusion

In this study, I analyzed which places are the best to set up a cafe in the city of Rotterdam. I identified popularity of the location as well as the amount of the cafes in the Neighborhood as the main determinants of what Neighborhood is the best. I built classification models to understand which Neighborhood is the best. These models can be very useful in helping entrepreneurs to determine the best location for a cafe in the future. As one of the benefits, these models can drive growth for cafes, increase profits, and ensure cafe popularity and success.