



IBM Data Science Professional Certificate

Submission date: 07-17-2019

The Battle of Neighborhoods

Capstone project

Pascal Hakkers

Alex Akison

## Abstract

This report is on the project of finding a possible location for a new McDonald's venue in the city of Amsterdam. For this we look into different dataset and use those to cluster Amsterdam into several segments where a McDonald's venue could make sense based on population density and whether there is already a restaurant in the area or not. After clustering the population, several area's showed up without a McDonalds. Those area's where up for a closer look. Possible housing locations are considered for those areas if the location have a shop destination and a minimal floor area of 600 m2. In total 401 addresses come up meeting the set requirements.

## Content

Abstract .....	2
Content.....	3
Introduction.....	4
Background .....	4
Problem .....	4
Interest.....	4
Data .....	4
Methodology .....	5
Results.....	6
Discussion .....	9
Conclusion .....	9
Future directions .....	9
References.....	10

## Introduction

### Background

McDonald's is worldwide maybe the most well-known fast food franchise formula. In the Netherlands it has about 250 restaurants, weekly they serve together over 3 million guests. Restaurants are owned by a franchisee or by the McDonald's itself. The company is still looking to expand their businesses and therefore looking into possible new venues constantly. New restaurants have a need for a minimal floorspace and a minimal number of inhabitants in the area.

### Problem

For this project we are exploring the possibility if and where McDonald's could open a new restaurant in the city of Amsterdam, the Netherlands, with regards to the specified requirements. As floor surface a minimal of 500 square meters is necessary and 35000 inhabitants/potential customers per venue. McDonald's already has 17 active venues in the Amsterdam area so it will be a real challenge to find a new spot.

### Interest

This project and it's result are interesting for the McDonald's corporation as well as a potential franchisee looking to start a new venue in Amsterdam.

## Data

To analyse this problem we need to know how many people are living in Amsterdam, how they are distributed and whether or not if they have access to an already existing restaurant in the near vicinity. In the area's with enough potential customers and without a McDonald's venue we are going to look for potential housing locations. To solve this problem we need these data:

1. Data on how many people are living in Amsterdam, where they live and how they are distributed over the residential areas;

*The city of Amsterdam provides a set of data with all addresses in the city [1], categorized by type [e.g. house, store]. Every record in this dataset - relevant to this project - has a) an address and b) latitude and longitude coordinates. This data, however, does not contain the needed demographics of people there are living on an address. For this we use the average persons per household number 1.8, as stated by oozo.nl [2]. The combination of address, location of that address and the average of 1.8 persons per household provides us with what is needed.*

2. Current distribution of existing McDonald's restaurants;

*To gather the existing restaurants we utilize the Foursquare API [3], specifically the search endpoint. This endpoint returns a list of venues corresponding to the performed search query in JSON format. Each search result contains - at least - a) an address and b) latitude and longitude coordinates.*

3. An overview of the requirements for a new McDonald's venue;

*McDonald's has a few minimal requirements for their new venues, which can be found on their website [4].*

4. A list of addresses that possibly could reside a new McDonald's restaurant.

*The dataset from the city of Amsterdam as mentioned in [1] also contains addresses of all the business addresses and the floorspace of every address. This can be used to find housing options that meet the location requirements.*

### Methodology

In this project we will look into the possible housing opportunities to open a new McDonald's venue in the city of Amsterdam. We will be looking at area's populated by a minimum of 35,000 people which do not already have a McDonald's restaurant.

Firstly we divide Amsterdam into clusters of 35,000 inhabitants. For this a dataset from the municipality of Amsterdam is available [1] with all the buildings in the city and their latitude and longitude location data. To obtain clusters from this data we used the DBSCAN and K-means clustering algorithms and feed them with the coordinates of the buildings. DBSCAN was used on several different test sets and gave some promising results. When the test set was increasing in numbers the cluster grew larger and larger. On the full data set it would end up showing one big cluster with several outliers. This lead us to conclude that DBSCAN might not be the best possible algorithm for this job, therefore K-Means was used to form clusters. K-Means was fed with a number of clusters calculated by:

$$\text{No. of clusters} = \text{no. people in Amsterdam} / \text{no. people per McDonald's}$$

With the number of clusters specified K-Means is able to generate clusters based on location data. After obtaining the clusters we calculate the borders of the clusters to generate polygons, which can be displayed on a map representing the area's of Amsterdam for further analysis. After clustering the addresses we separate the clusters into two, the ones who have a restaurant present and those who don't.

Finally we look into the building which have a store indication and meet the requirement of a McDonald's restaurant. Any building that is in a cluster that already has a McDonald's present will be discarded, leading to an overview of possible buildings to be utilized as a new restaurant venue by McDonald's.

### Results

In this dataset are 523,487 addresses present, after cleaning, and according to the website of the municipality [2] the average persons per household in Amsterdam is 1.8. That leads to a total of inhabitants of 942,276 (at the time of the dataset is created) Based on the buildings dataset we form clusters of approximately 19,444 addresses ( $35000/1.8$ )

Our calculations for number of clusters leads to 27 different clusters:

$$\text{No. of clusters} = 942,276 / 35,000$$

When we plug this in into the K-Means algorithm along with the coordinates of the Amsterdam addresses we get the distribution as seen in figure 1.

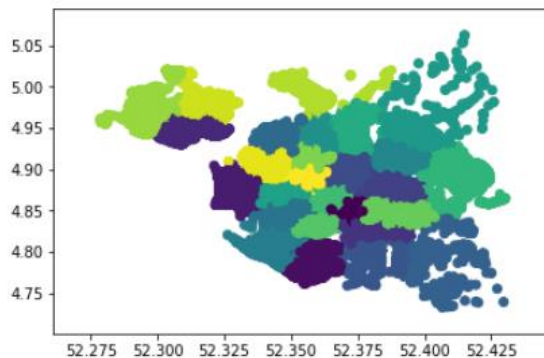


Figure 1: Cluster distribution of Amsterdam addresses by K-Means.

Next, for each cluster the borders are calculated leading to polygons which can be represented on a map. See figure 2 for a sample bordered cluster.

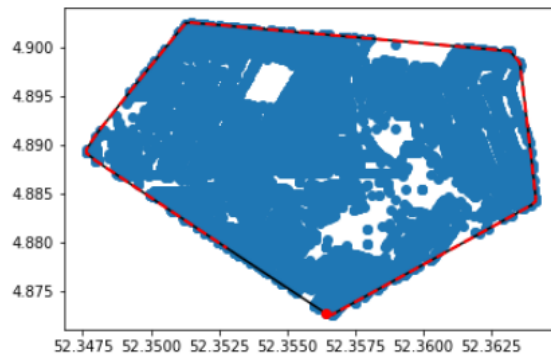


Figure 2: A sample of one of the calculated clusters, including the border.

For every cluster we can extract the polygon coordinates and plot them on a map of Amsterdam, see figure 3 for the overview of clusters plotted over the city of Amsterdam.

After plotting the clusters we add the current McDonald's venues – obtained by utilizing the Foursquare API, search endpoint – to the map and colour the clusters red or green representing the presence of a restaurant already in that area or not, see figure 4 for the result.

Next we show all the results of the possible buildings – venues with a store indication and a minimal floor area of 600 m<sup>2</sup> – on the map, see figure 5. Finally all the results are displayed as addresses and saved in a file to be sent to the McDonald's corporation.

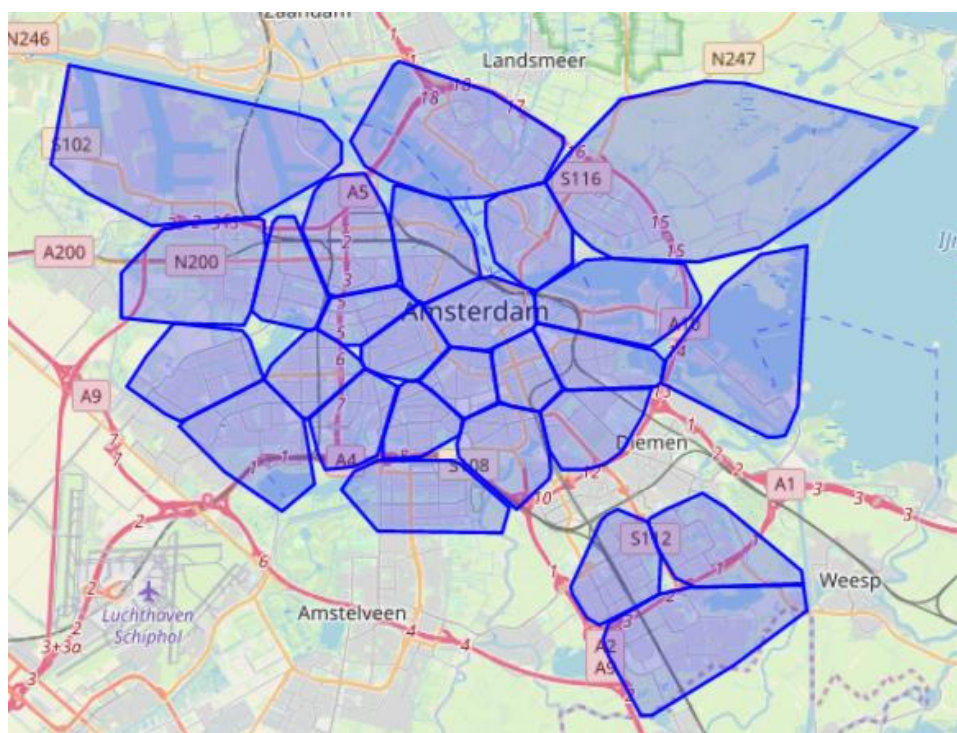


Figure 3: Clusters of 35,000 people plotted on Amsterdam.



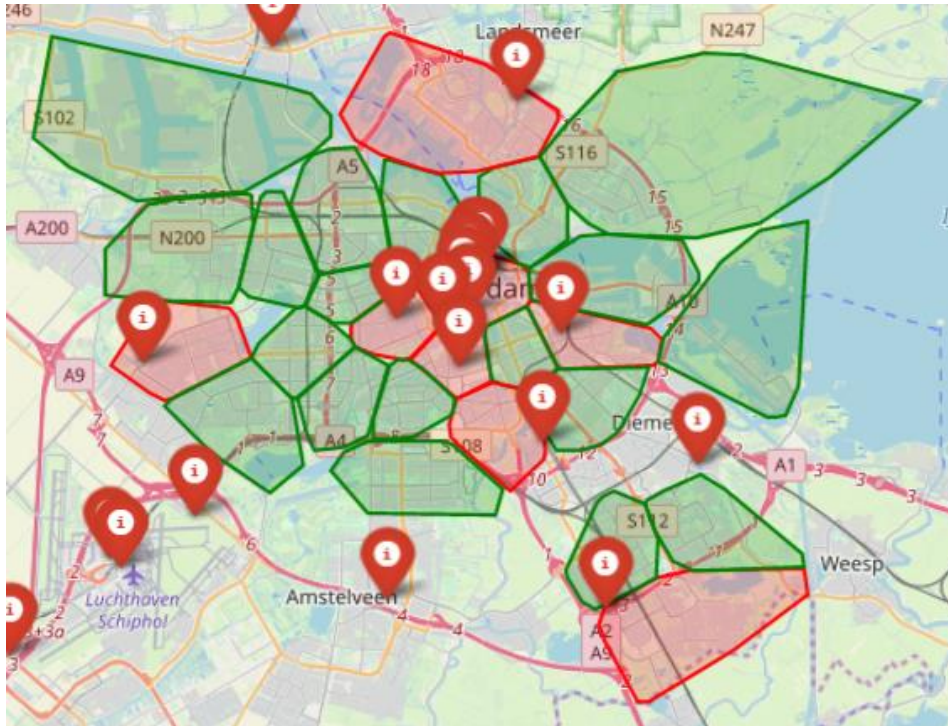


Figure 4: Clusters divided by McDonald's presence.

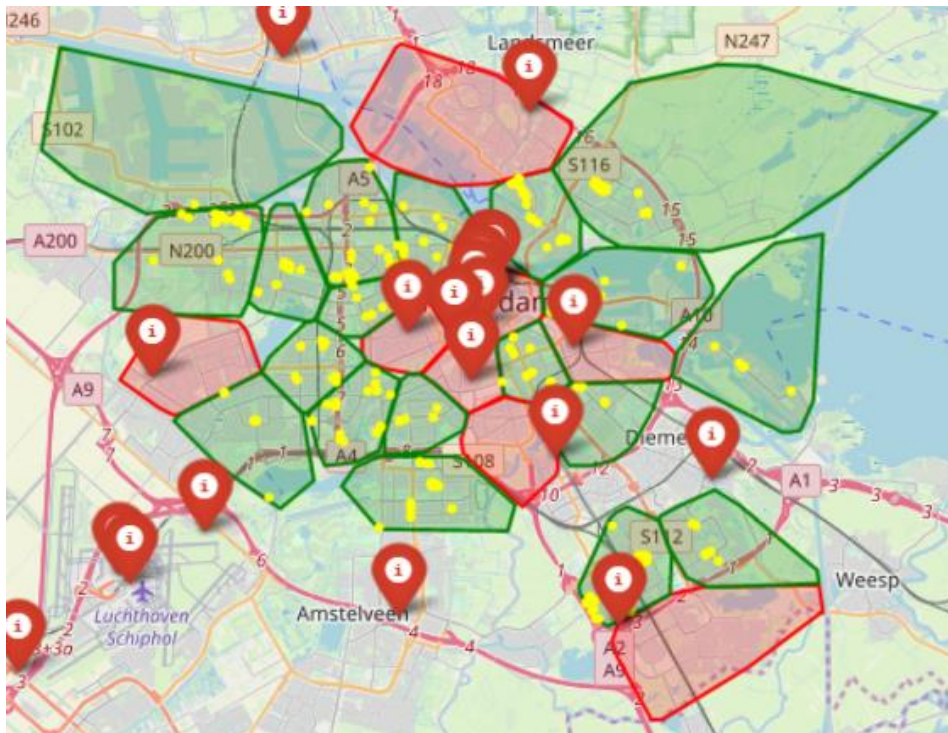


Figure 5: Clustered addresses, segregated areas and possible building options.



### **Discussion**

When we started this project we did not expect to get a result like this, maybe one or two area's of possible venues where to be expected. But this analysis showed that only 8 of the 27 clustered area's have an McDonald's restaurant at this moment. That leaves 19 densely populated area's for further exploration of the possibilities. The McDonald's corporation and their franchisees seem to have lost of possibilities for expansion in Amsterdam. Looking at the map of the end result, as displayed in figure 5, one could argue that there are already restaurant close to the border of a possible area and therefore not to go venture in these area. But on a closer look of the results three distinct groups of cluster show around the city centre. Combining these clusters can potentially be more interesting than the single clusters on their own.

### **Conclusion**

Amsterdam is a densely populated city and a high McDonald's density. Looking at the distribution of both McDonald's restaurants and inhabitants of Amsterdam we can conclude that there is still room for expansion in the majority of population clusters. 19 of 27 clusters show potential for new restaurants and looking at the bigger picture, when combining potential clusters, one could see three big area's of Amsterdam without a McDonald's restaurant at this time.

### **Future directions**

With this project over 250 housing possibilities came up in the city of Amsterdam. We looked at areas which did not already had an existing McDonald's restaurant within the borders of these areas. Possible extension to this project could consider location of existing restaurants and prevent results that are within a certain radius of an already existing McDonald's. Result can be further narrowed down by looking more into location details such as reachability, whether or not it the building is at ground floor level, next or close to main roads, more dinning venues in the vicinity and if the building already has a food permit from the municipality.

Also a different clustering method could result in more precise clustering, one could look into the DBSCAN method with an implementation of a haversine function that might calculate distances between buildings more precise and therefore conclude to -slightly-different clusters.

## References

- [1] <https://api.data.amsterdam.nl/dataselectie/bag/export/>
- [2] <https://www.oozo.nl/cijfers/amsterdam>
- [3] <https://api.foursquare.com/v2/venues/search>
- [4] <https://www.mcdonalds.nl/over-mcdonalds/nieuwe-restaurantlocaties>