

Applied Data science Capstone : Battle of Neighborhoods

« Best neighborhoods in Amsterdam »

1. Introduction : Business Problem

Starting a new business need more research to come up with the best idea in the best place. The aim of this project is to find for stakeholders **which is the most promising business in which the perfect neighborhoods** to start, especially in a city with high density like **Amsterdam**.

There are lots of business projects like **Restaurant, Grocery, Clothes Store,...** scattered between neighborhood in **Amsterdam**. It would be interesting if we have some groups with close similarities to define which is the **most common business in area**. Also, it is important to compare with **number of population in each neighborhood** to have more idea about suitable business.

Therefore, we will use data science to generate a few propositions of neighborhoods based on this criteria, in order to find suitable business kind in best location for stakeholders.

2. Data section :

To answer the main question in the previous section, we should define :

- Most common business (Restaurants, Groceries, Stores, Coffees,...) in each neighborhood
- number of total population in neighborhood

First, we need to download the list of all neighborhoods in Amsterdam (Wijk) with demographic informations to extract number of habitants, from website <https://claircitydata.cbs.nl/>

Then, we will get all venues located in the corresponding neighborhoods using **Foursquare API**, in order to classify the most common business in each neighborhood.

3. Methodology section :

In this project, our aim to define a few propositions of neighborhoods candidates with their profiles and match them with their demographic information such as number of habitants.

First, We suggest to use the explore function to get the most common venue categories in each neighborhood, and then use this feature to group the neighborhoods into clusters.

Then, we will use the k -means clustering algorithm to complete this analysis, and to have similar neighborhoods based on their profiles. Finally, we get our clusters sorted by number of population in each neighborhood and then the global idea about the suitable business in suitable neighborhood.

4. Results section :

We get 5 clusters of neighborhoods (*Figure Map below*), based on their most common venues around, and the corresponding number of habitants. Now, let's explore each cluster, to get most populated neighborhood candidates :

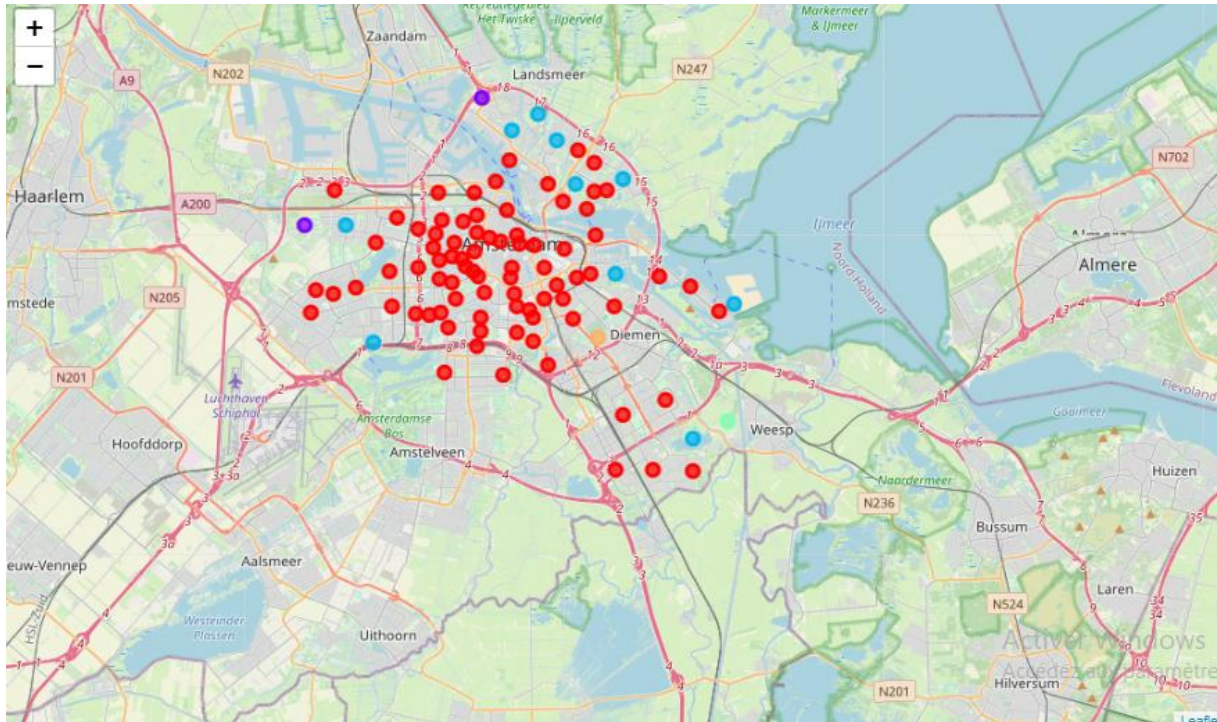
1st Cluster : Bijlmer Oost (E,G,K), Bijlmer Centrum (D,F,H), and Jordaan. This is Cluster has the most populated neighborhoods, and it's more likely to find clothes stores, Restaurants/Food places, Supermarkets, and transport stations.

2nd Cluster : Oostzanerwerf. There are more public places like Soccer field, Event space and Exhibits

3rd Cluster : Geuzenveld, Banne Buiksloot. This cluster has an average populated neighborhoods, with most common venues like transport stations, supermarkets and restaurants

4th Cluster : Driemond. Is the least populated neighborhood, with public places like park and zoo.

5th Cluster : Betondorp. Few population, and we can find places like Bars and Bakery.



5. Discussion section :

According to profiles that we get from results section, we can get some recommendations to start with a suitable business in each neighborhood. First, if we take the 1st Cluster, we can observe that there are the most populated and animated neighborhoods, but there aren't some vital places like **Gym clubs** or **Fitness centers**. In the same way, the 3rd Cluster has populated neighborhoods, but there aren't stores like **Clothes stores** or **Book stores**. In other hands, the 2nd cluster may hasn't much populations, but has more public places like Park and Soccer fields, and then much more flows of people, and it may be interesting to get **groceries** or special stores like **Sport stores** or **Ethic stores**.

6. Conclusion section :

We can say that our analysis of neighborhoods and their profile could help stakeholders to take decision about what kind of business to start with, and which neighborhoods is the most suitable. Note that our approach to solve this project take the number of population in neighborhoods out of clustering stage, and it serve to compare in the final stage between neighborhoods. The reason behind this approach is to let stakeholders to decide the suitable business according to number of population. Therefore, in this project we could also add other variables based on demographic informations that influence the choice of business nature, like gender or age.