Applied Data Science Capstone: The Battle of Neighborhoods

Use case: « 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 promissing business in which the perfect neighborhoods** to start, especially in a city with high density like **Amsterdam**.

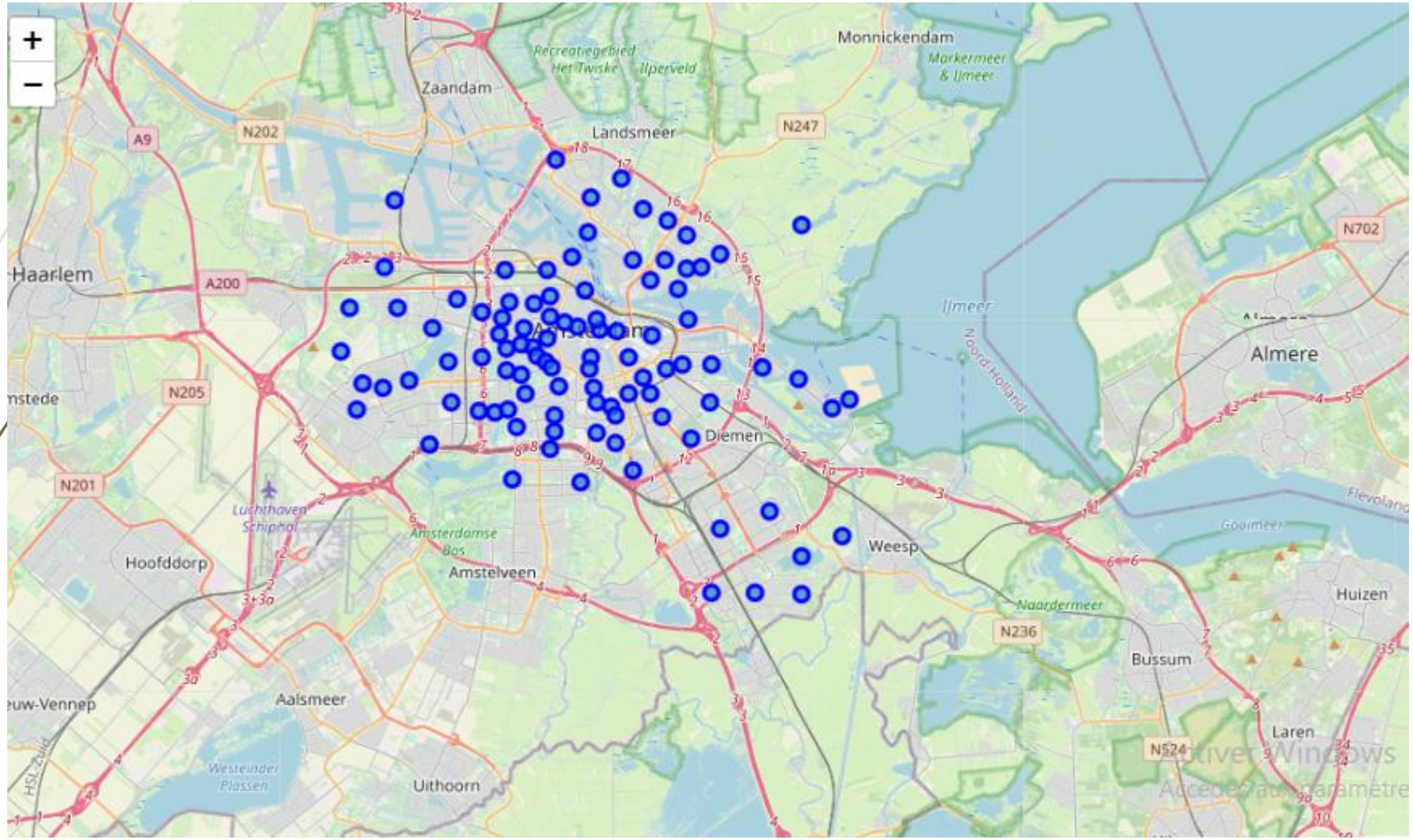
- There are lots of business projects like **Restaurant, Grocery, Clothes Store,…** scattered between neighborhood in **Amsterdam**. It would be interesting if we have some groups with close similarities to define which is the **most common business in area**. Also, it is important to compare with **number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

- 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 using **Foursquare API,** in order to classify the most common business in each neighborhood

2. Data: Map of neighborhoods in Amsterdam-Wijk



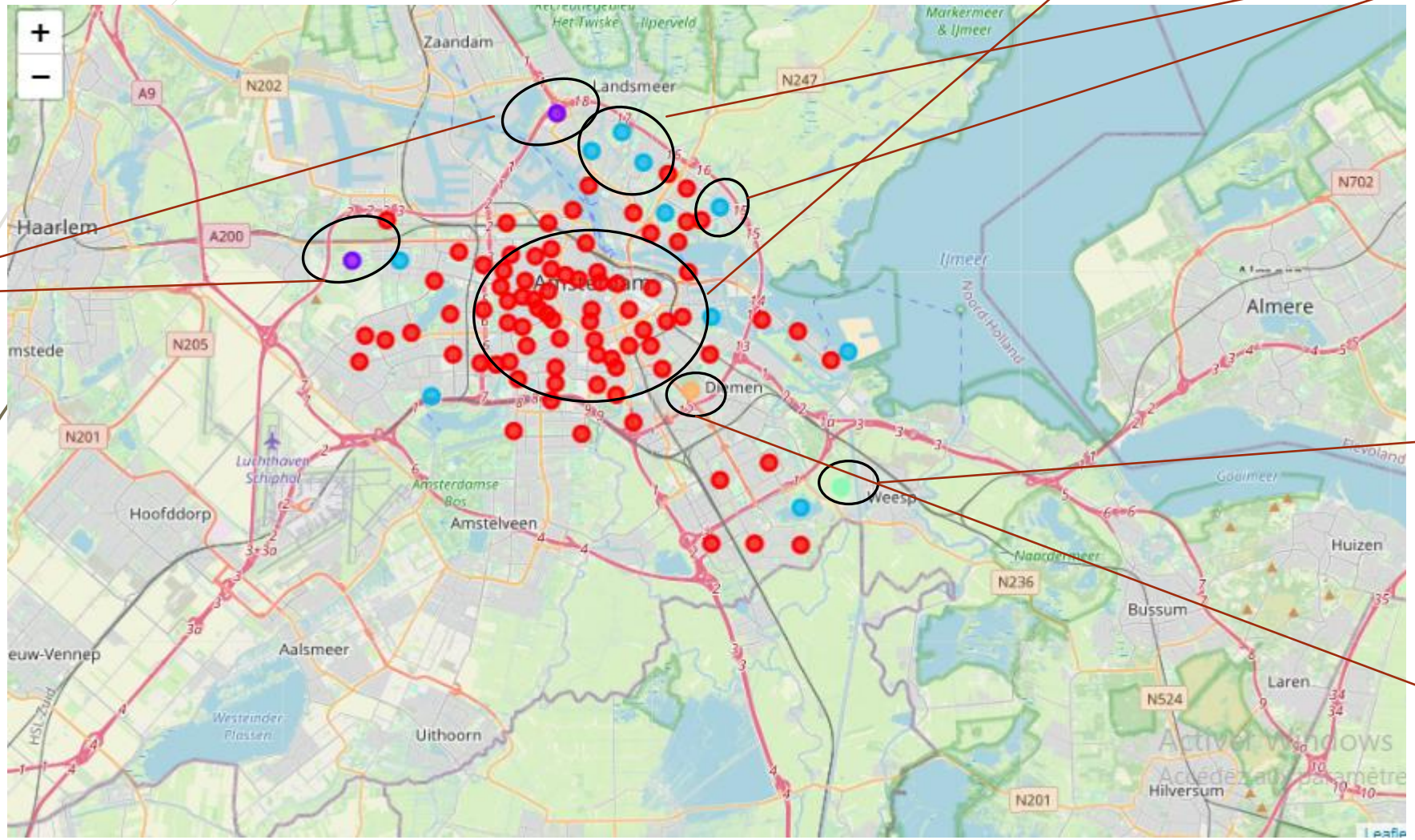
3. Methodology

- 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 profils. 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

	Neighborhood candidates	Most Common venues	Cluster Nº
Most populated	Bijlmer Oost	Restaurants/Food places	1st
	Bijlmer Centrum	Supermarkets	
	Jordaan	Transport station	
Average populated	Geuzenveld	Transport station	3rd
	Banne Buiksloot	Supermarkets	
Few populated	Oostzanerwerf	Soccer field, Event places, Exhebits	2nd
	Driemond	Park, Zoo	4th
	Betondorp	Bar, Bekerey	5th

4. Results: Map of clusters



1st Cluster

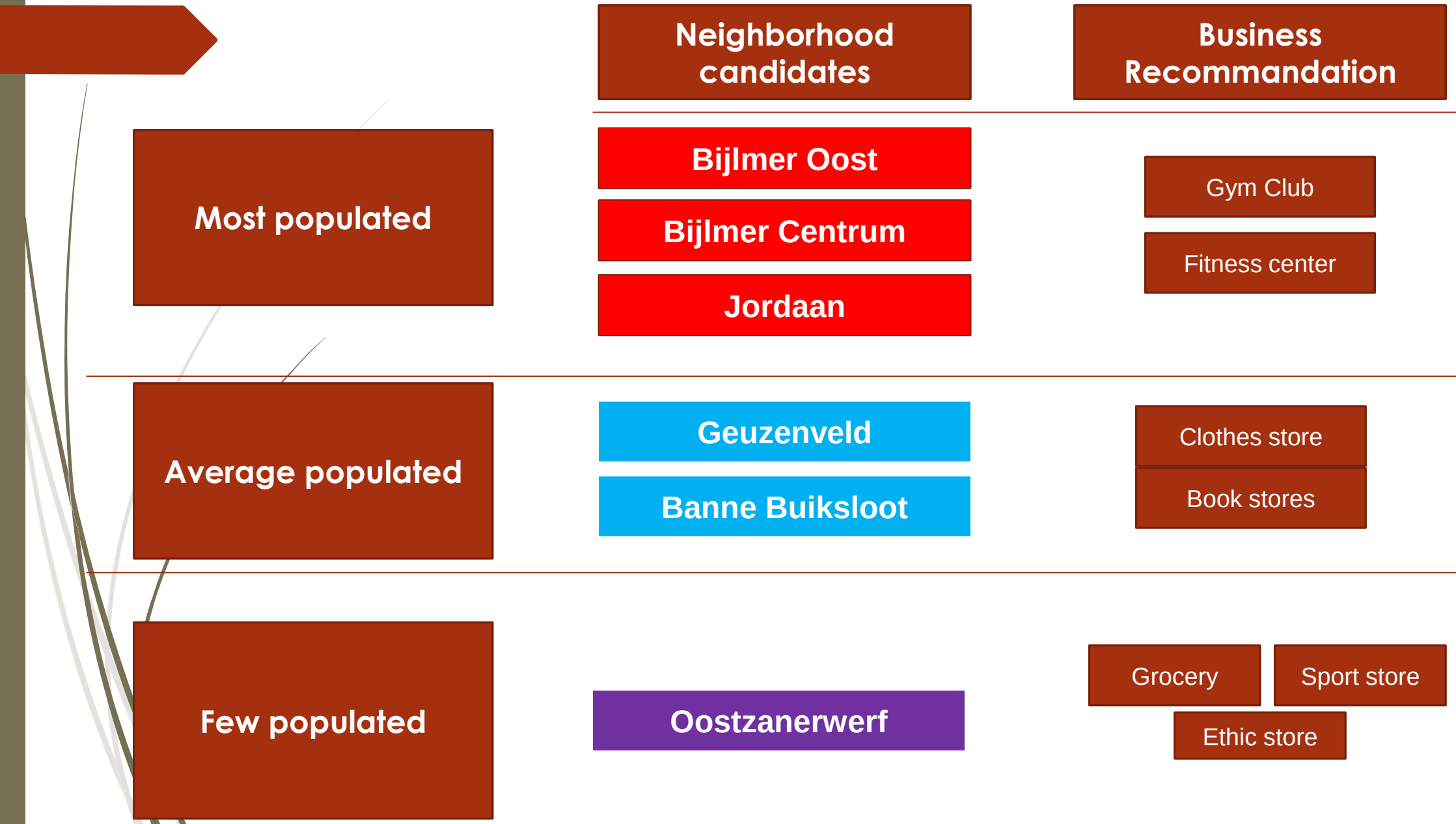
3rd Cluster

2nd Cluster

4th Cluster

5th Cluster

5. Discussion: Propositions



6. Conclusion:

- We can say that our analysis of neighborhoods and their profil 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 aproach 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.