Battle Of Neighborhoods Amsterdam

PRANJAL BIYANI

Introduction

- Amsterdam is the capital city of the Netherlands, and is also one of the most popular tourist destinations in Europe.
- The project attempts to provides the travelers to Amsterdam with a comprehensive comparison of different neighborhoods of the city to help them when deciding the location of their stay.
- We classify the neighborhoods into clusters of similar characteristic to give a better idea about the type of neighborhoods one is looking to stay.

Data Acquisition

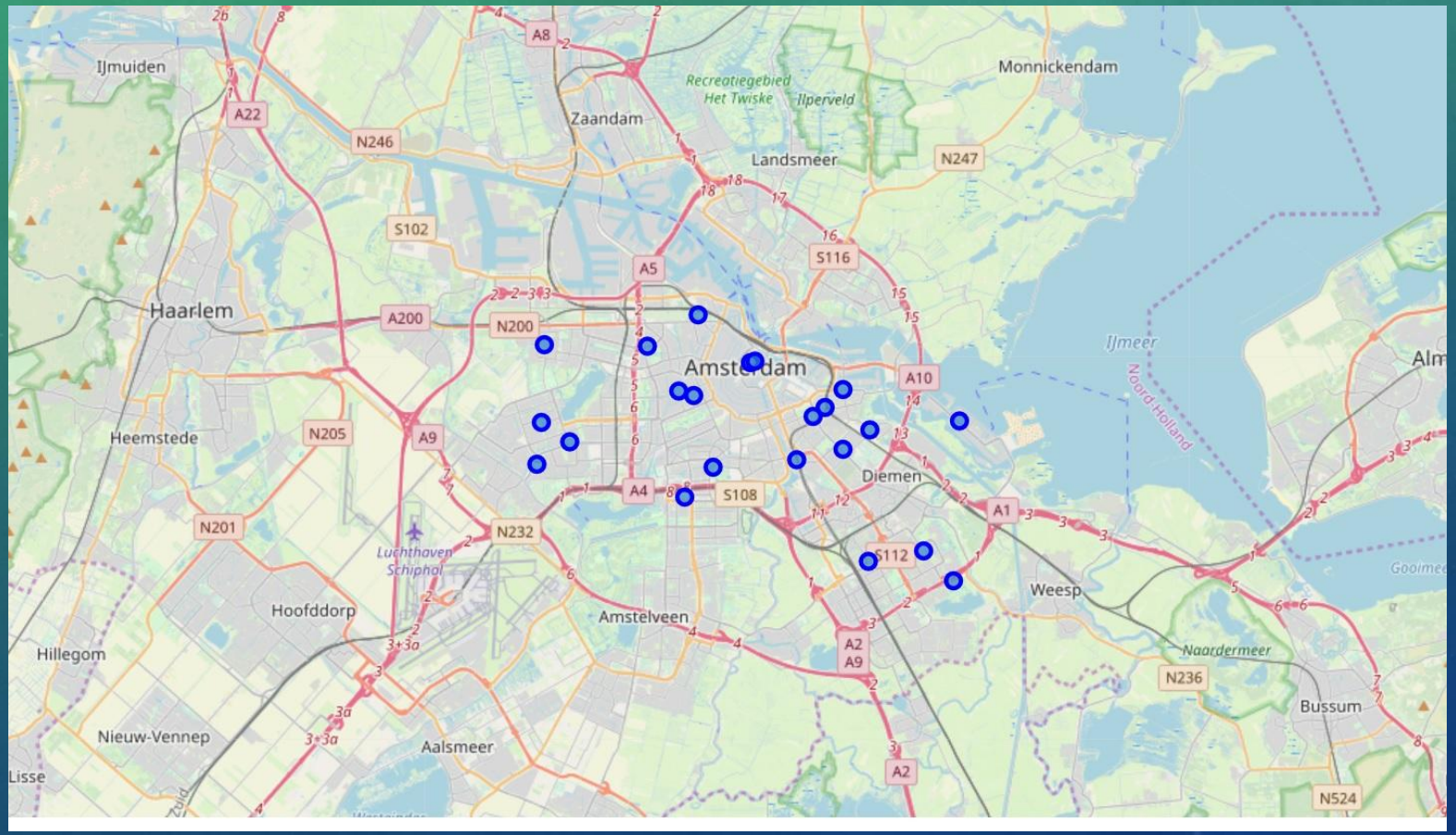
- Amsterdam Airbnb listing data and neighborhood list from Kaggle website
- Geospatial coordinates of Amsterdam neighborhood using Geopy package of Python.
- Use Foursquare API to return venues of interests by passing the geospatial coordinates.

Data Cleaning And Preparation

- Removed any inactive listing that was last reviewed before year 2018
- One listing where price value is zero is removed as bad data and another one with extreme outlier prize of 5,000 is also removed.
- After cleaning, 3 relevant features from the main Airbnb dataset were derived.
- This is combined with venue features to form a final dataset with total of 8 features.

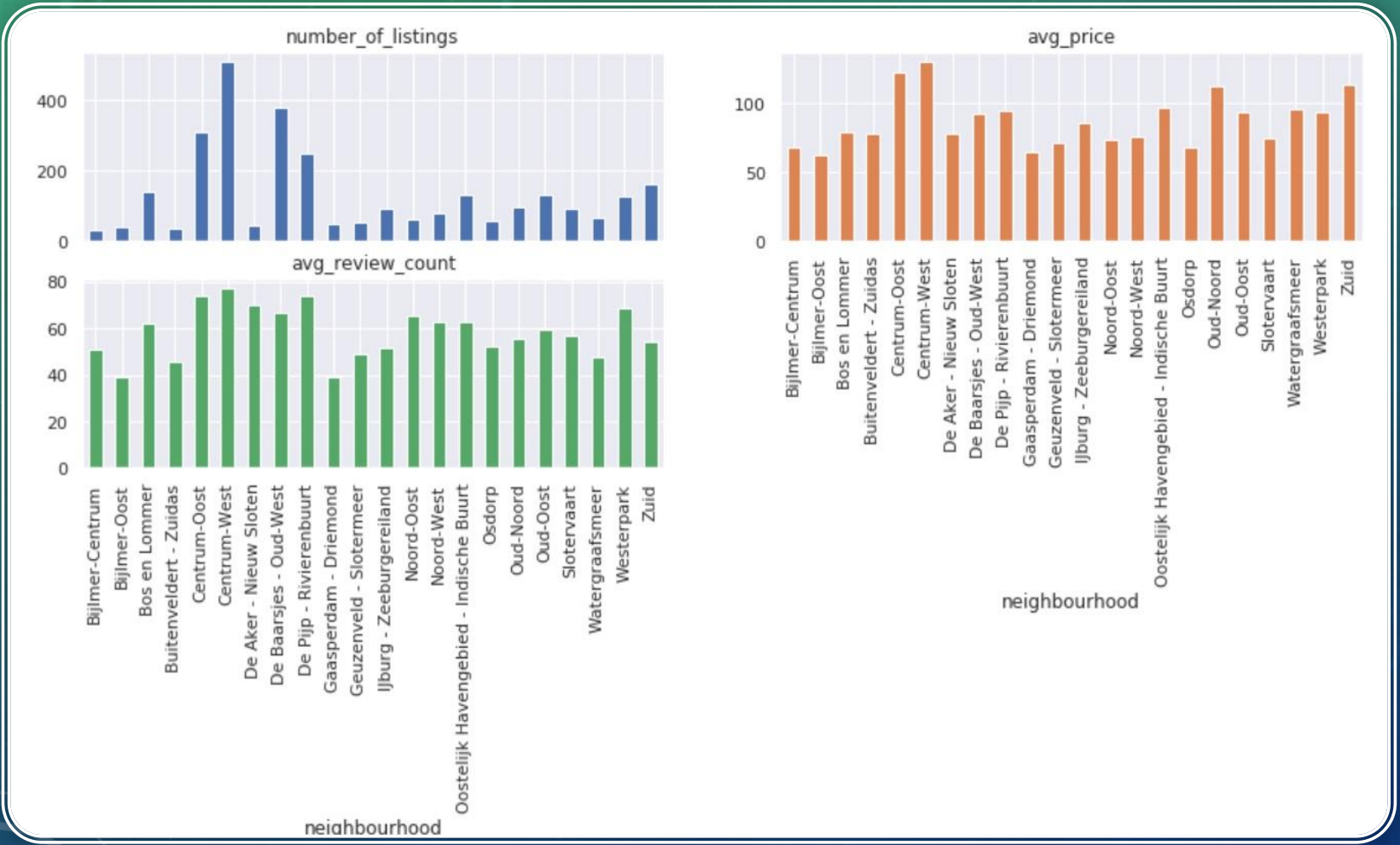
Exploratory Data Analysis

Neighborhood Map



- Most of the neighborhoods are located centrally
- Two distinct outer spatial clusters having three neighborhoods each.
- Two neighborhoods are further far from the center and also from other neighborhoods.

Exploratory Data Analysis



- Central neighborhoods score high for each category and these should form a cluster.
- Neighborhood “Centrum-West” appears to be most popular with maximum number of listings.
- Average prices are lower and are similar among the neighborhoods which are not central.

K-means Clustering

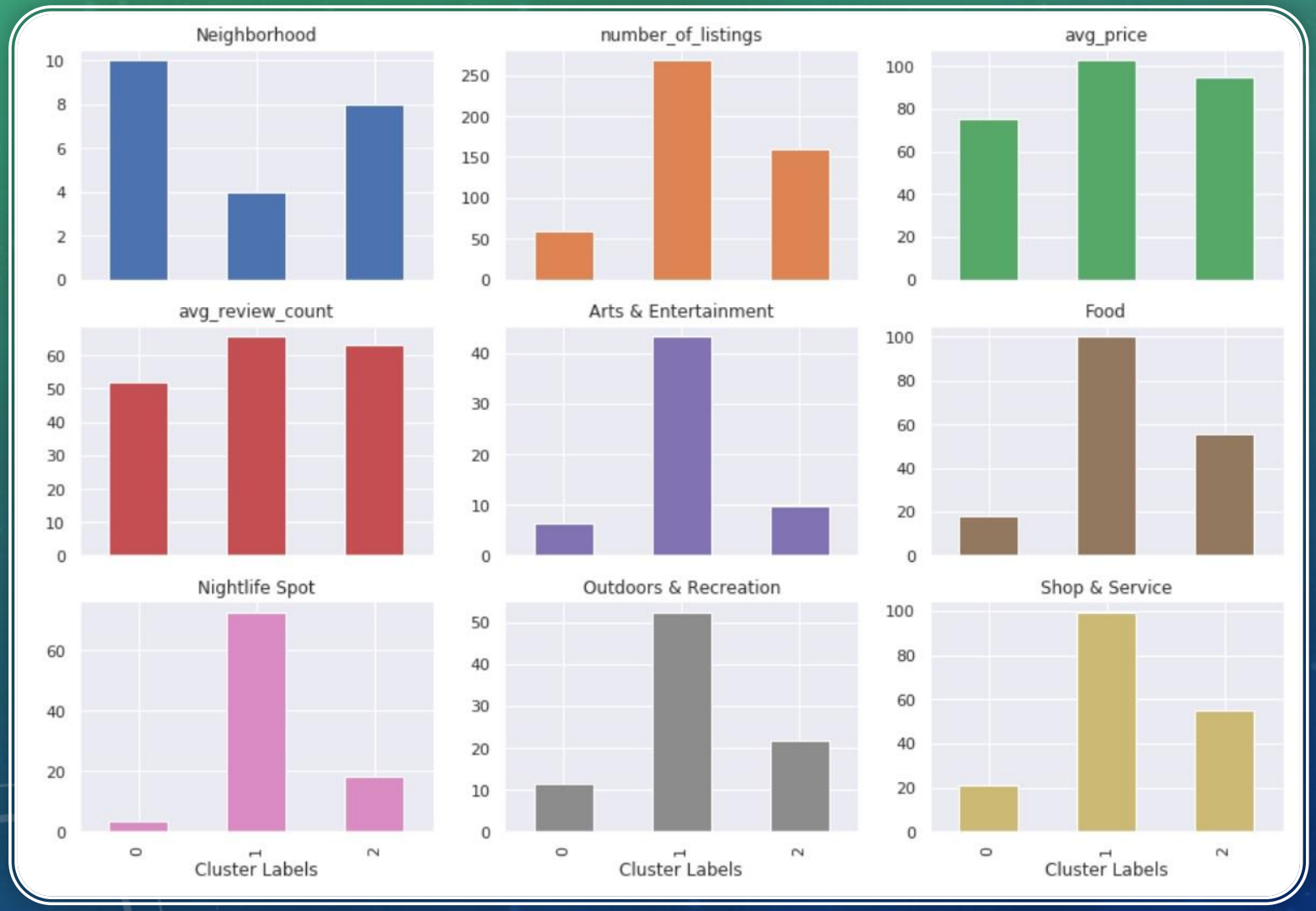
- We used the unsupervised classification algorithm K-means algorithm to classify the neighborhoods.
- K-Means algorithm is one of the most common cluster method of unsupervised learning.
- We choose value of K to be equal to 3.
- The features are normalized using a Min-Max normalization algorithm

Results

Cluster 1	Cluster 2	Cluster 3
Bijlmer-Oost	Centrum-West	Bijlmer-Centrum
Buitenveldert - Zuidas	De Baarsjes - Oud-West	Bos en Lommer
De Aker - Nieuw Sloten	Noord-West	Centrum-Oost
Gaasperdam - Driemond	Oud-Noord	De Pijp - Rivierenbuurt
Geuzenveld - Slotermeer		Oostelijk Havengebied - Indische Buurt
IJburg - Zeeburgereiland		Oud-Oost
Noord-Oost		Westerpark
Osdorp		Zuid
Slotervaart		
Watergraafsmeer		

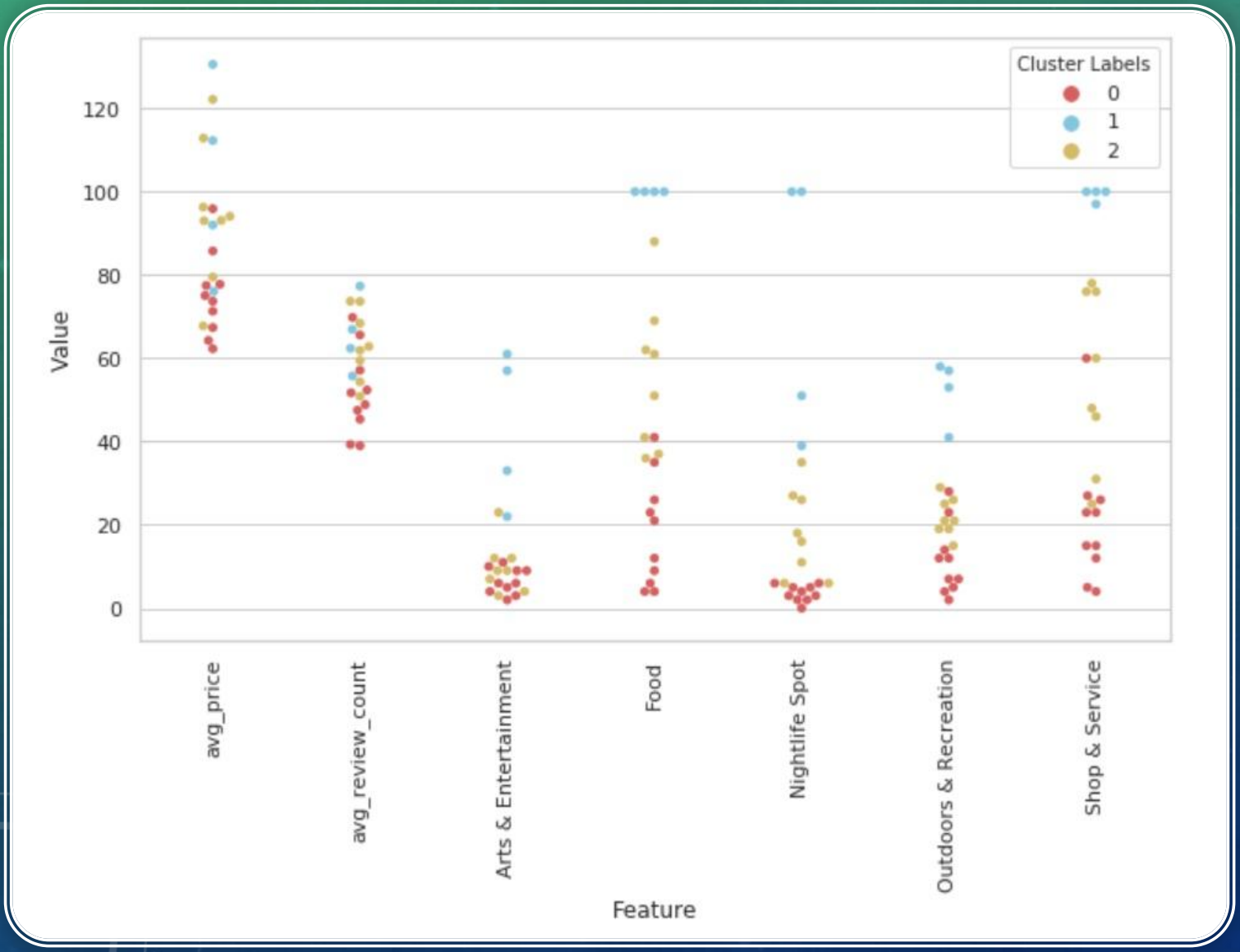
- Cluster 1 contains four central neighborhoods
- Number of neighborhoods in each cluster varies

Cluster Feature Analysis



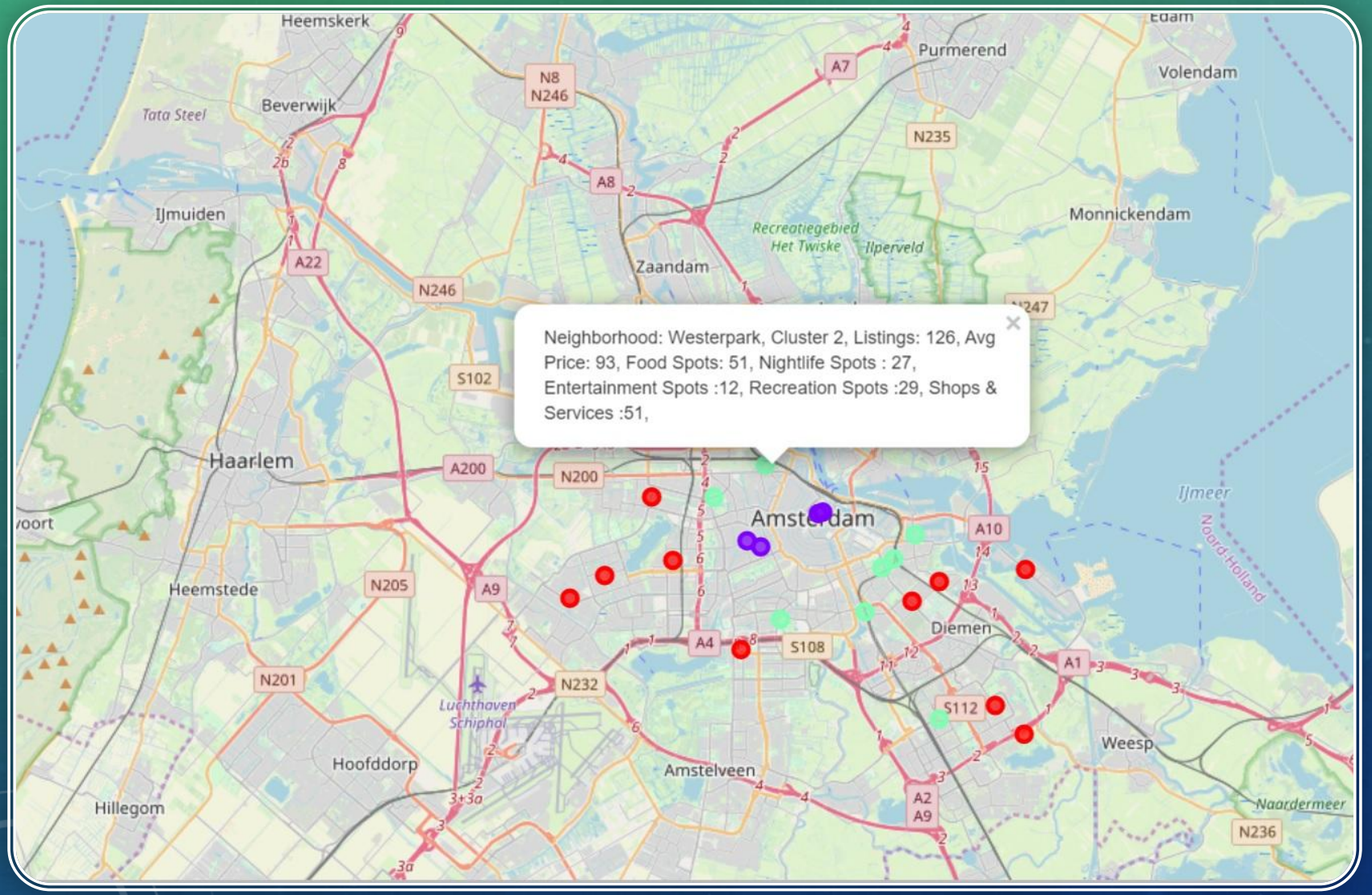
- Cluster 1 is the main central area and has the highest concentration of all types of venues but is also costlier on average.
- Cluster 2 has more(8), neighborhoods and is cheaper than cluster 0 on average
- It also matches or exceeds cluster 0 in total number of listings and venues

Neighborhood Features In Each Cluster



- Visualize how the feature values vary among neighborhoods within each cluster
- Revealed that even with in a cluster, most of the values varies significantly

Neighborhood Clusters On The Map

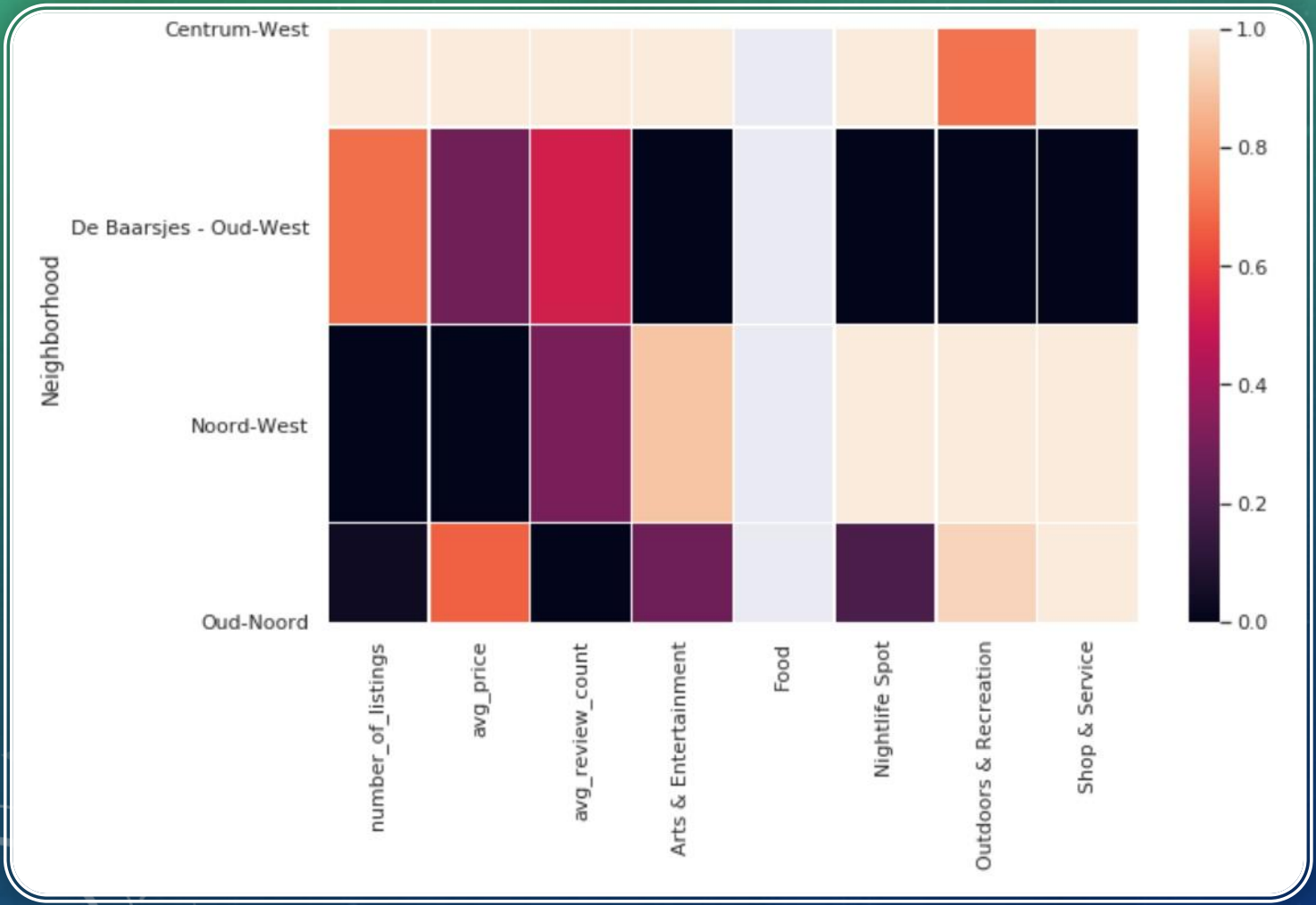


- Clusters are color coded
- Reveals strong correlation between actual geo-spatial clusters and logical clusters formed by the algorithm.
- A short info snippet about the neighborhood also pops up.

Heatmap - Cluster 0

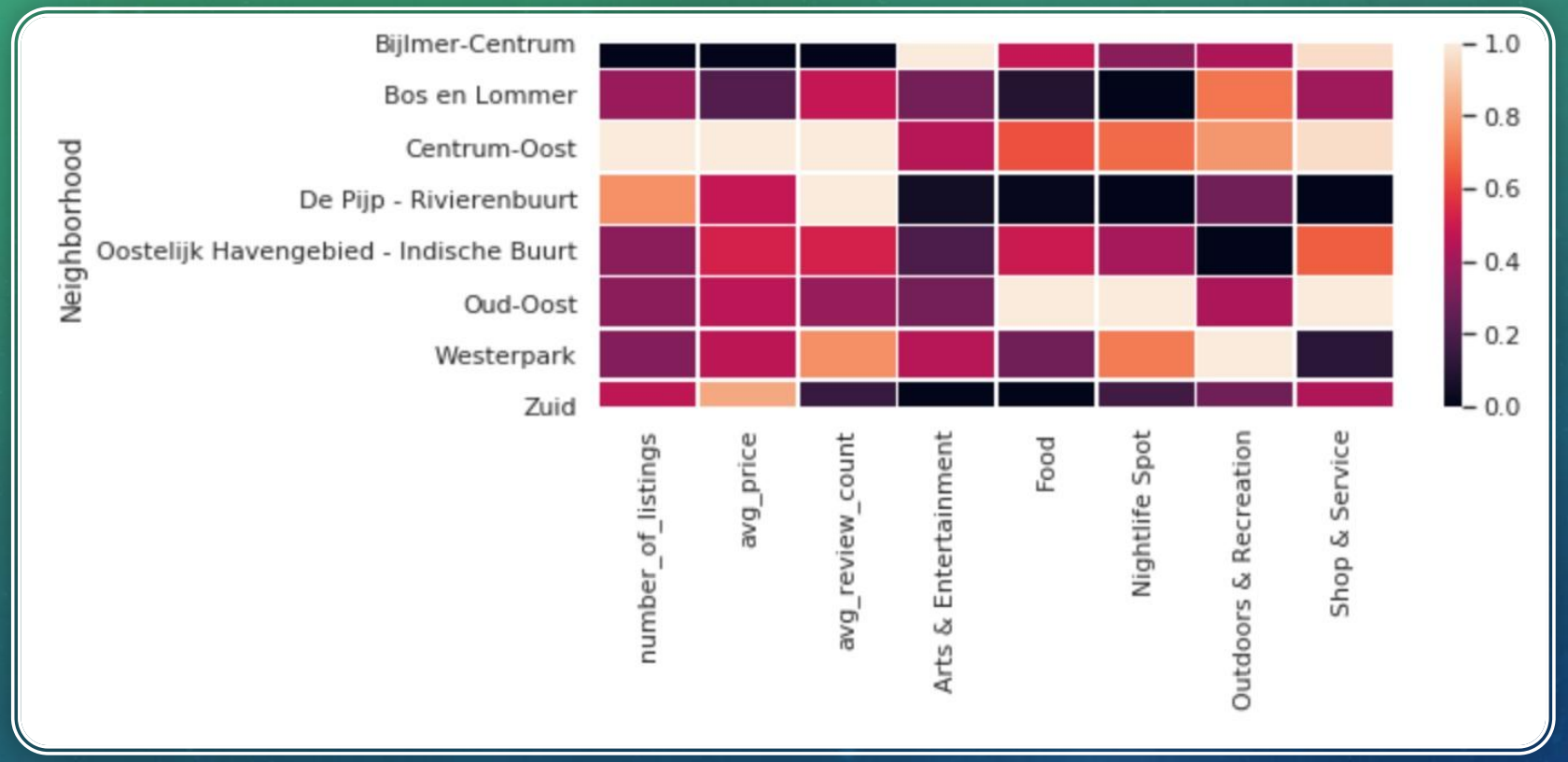
- This cluster has the most number of neighborhoods.
- Very diverse cluster with all types of neighborhoods.

Heatmap - Cluster 1



- Centrum- West stands out in most categories but is also costlier.
- Noord-West fares the best with low price and most venues around.

Heatmap -Cluster 2



- This cluster has second most number of neighborhoods and is also very diverse.
- Centrum-Oost is the outright winner in this cluster

Conclusion And Future Recommendations

- We used K-Means clustering algorithm to classify the neighborhoods
- We got three distinct clusters of different size.
- We characterized the clusters and analyzed the clusters individually
- Future recommendation include
  - Include full homes/apartments listings for analysis as this can bring in more listings located non-centrally
  - Obtain dataset with more recent listings and this can present a different case for each neighborhood