Capstone Project - Queens

Introduction: Business Problem

Introduction: Business Problem

A fastfood chain is interested in expanding to Queens. They want to open a few new restaurants and would like to find out the best neighborhoods to do so.

In order to answer that question, we first need to analyse the different neighborhoods in Queens, New York. We would like to understand the most common venues in each neighborhood are. Knowing this information we can group the neighborhoods which have similar characteristics. This helps to analyse each group to try to get insights or see patterns to help us to draw our conclusions and make good suggestions to the stakeholders.

Data

Data

The data we need must contain the following information:

- Names of the neighborhoods in Queens, New York
- Their latitude longitude
- Names of the 10 most common venues for each neighborhood

This dataset is accessible for free on the internet. It is acquired from https://geo.nyu.edu/catalog/nyu_2451_34572

The latitude and longitude values of the neighborhoods determined using the Geopy Library.

Finally, the most common venues of each city are obtained using the Foursquare API.

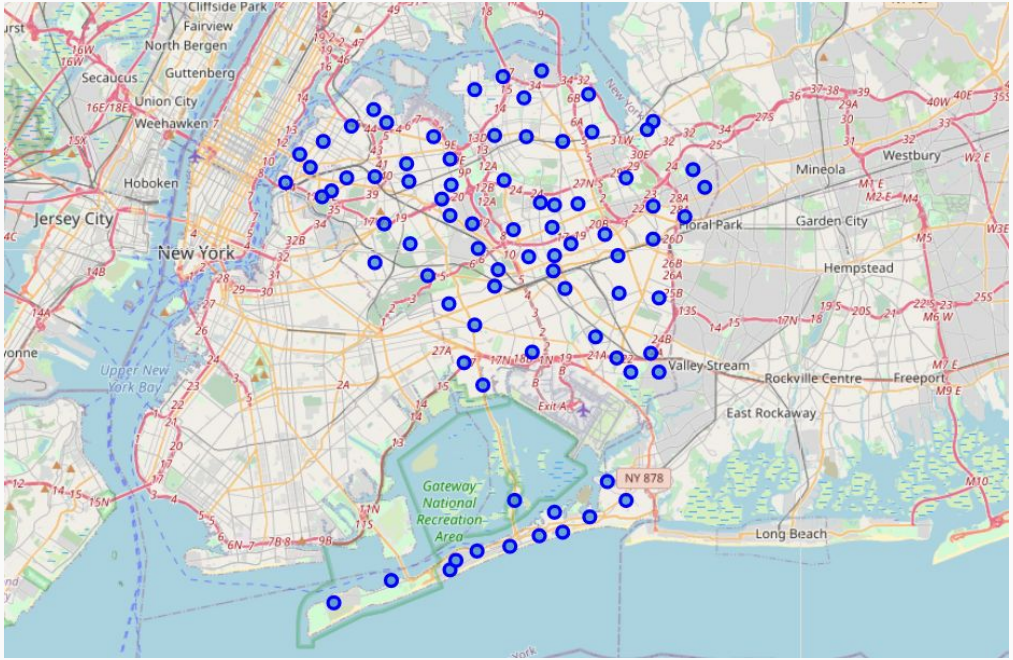
Dataframe

	Borough	Neighborhood	Latitude	Longitude
0	Queens	Astoria	40.768509	-73.915654
1	Queens	Woodside	40.746349	-73.901842
2	Queens	Jackson Heights	40.751981	-73.882821
3	Queens	Elmhurst	40.744049	-73.881656
4	Queens	Howard Beach	40.654225	-73.838138

Methodology

Methodology

We use the folium library to create the geographical map of Queens, in order to select Queens, York City as the central location in our map we need the coordinates. The geographical coordinates of Queens are 40.7498243, -73.7976337.

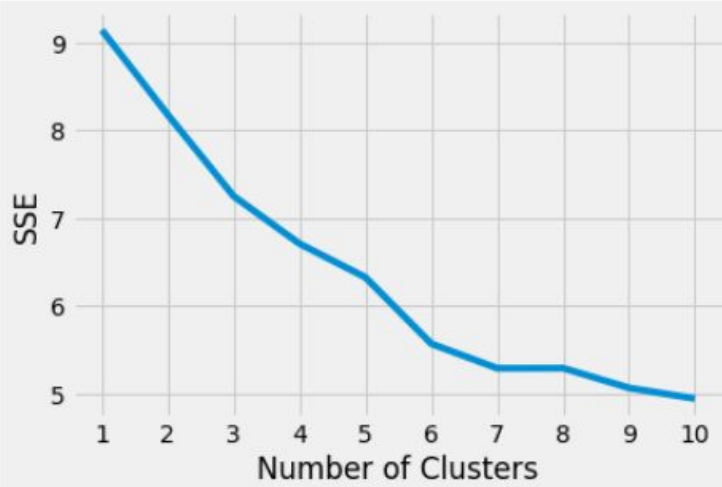


We create a new dataframe displaying the top 10 venues for each neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Arverne	Surf Spot	Playground	Metro Station	Sandwich Place	Pizza Place	Café	Thai Restaurant	Restaurant	Coffee Shop	Board Shop
1	Astoria	Hookah Bar	Middle Eastern Restaurant	Bar	Seafood Restaurant	Indian Restaurant	Greek Restaurant	Mediterranean Restaurant	Café	Bakery	Food Truck
2	Astoria Heights	Plaza	Italian Restaurant	Burger Joint	Laundromat	Bakery	Supermarket	Bowling Alley	Chinese Restaurant	Food	Playground
3	Auburndale	Italian Restaurant	Mobile Phone Shop	Train	Sushi Restaurant	Noodle House	Fast Food Restaurant	Bar	Gymnastics Gym	Mattress Store	Toy / Game Store
4	Bay Terrace	Clothing Store	Shoe Store	Women's Store	Mobile Phone Shop	Kids Store	Donut Shop	American Restaurant	Cosmetics Shop	Gift Shop	Coffee Shop

This enables us to cluster the neighborhoods according to their similarities in venues using the *k*-means method.

First we get the best number of clusters based on the elbow method.

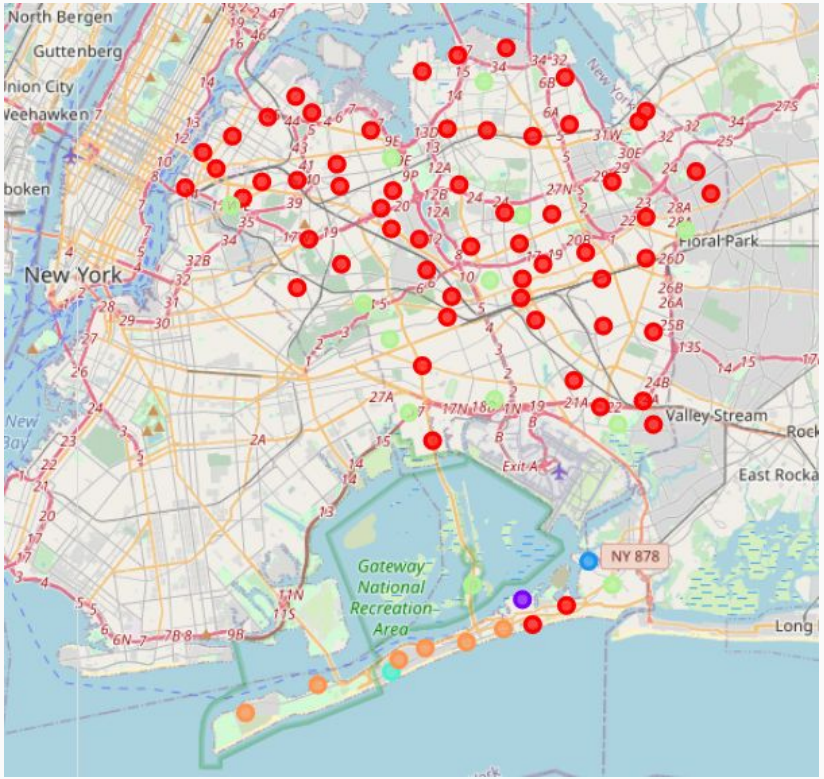


The optimal group number is determined by the convexity of the curve, here we find k=6.

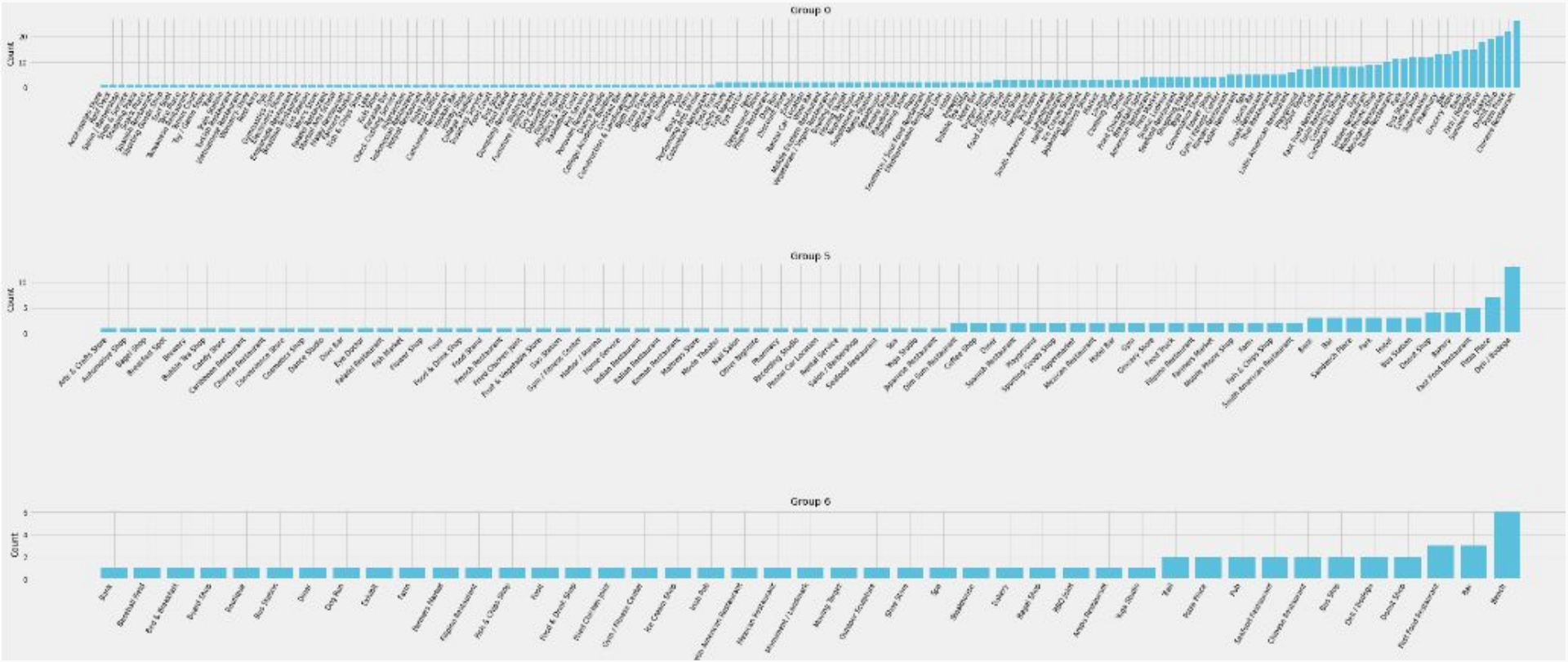
Results and Discussion

Results and Discussion

Finally, we can visualize the resulting clusters, where the neighborhoods with the same color belong to the same group.



To get an overview we will make bar charts for the clusters 1,5,6. We exclude the other clusters from the further analysis, since they only contain one neighborhood.



We can observe that in cluster 5 and 6 fast food restaurants seem to be very popular, they appear with a high frequency among the most 10 common venues. Since the cities in the same group are similar, it could be a good choice to open a fast food restaurant in the neighborhoods of these two clusters, which do not already show fast food restaurants among the 10 most common venues.

Conclusion

Conclusion

In this project we have analysed the neighborhoods in Queens and their most common venues. We have clustered the neighborhoods and found out which ones are similar. We found that in two clusters, fast food restaurants are especially popular.

After analysing the data we have the following recommendation: It would be a good choice to **open new fast food restaurants in those neighborhoods that belong to cluster 5 and 6, but don't have a fast food restaurant among their 10 most common venues** yet.