THE BATTLE OF THE NEIGHBORHOODS

Structure

- The business problem
- The data
- Methodology
  - No. of neighborhoods
  - Venues in the neighborhoods
  - Clustering
    - Neighborhoods per cluster
    - Cluster map
   - Heat map
- Results

Business problem

- Restaurant owner trying to open a new venues
- Currently established in other areas of New York City
- Goal is to open a new venue in Brooklyn
- Find a new neighborhood or potential neighborhoods with many potential costumers and growth potential

The data

- **New York City Data**
  - Inlcudes all five boroughs with its neighborhoods, their latitude and longitude values
  - This data builds the basis for further analysis
- Foursquare API to get information about all the venues in the different neighborhoods to determine the interest of the people living in the respective neighborhoods and the degree of competition
- Some important libraries like pandas, geopy, folium and more

Methodology

▶ Importing the data and creating a dataframe to analyse the boroughs of New York City
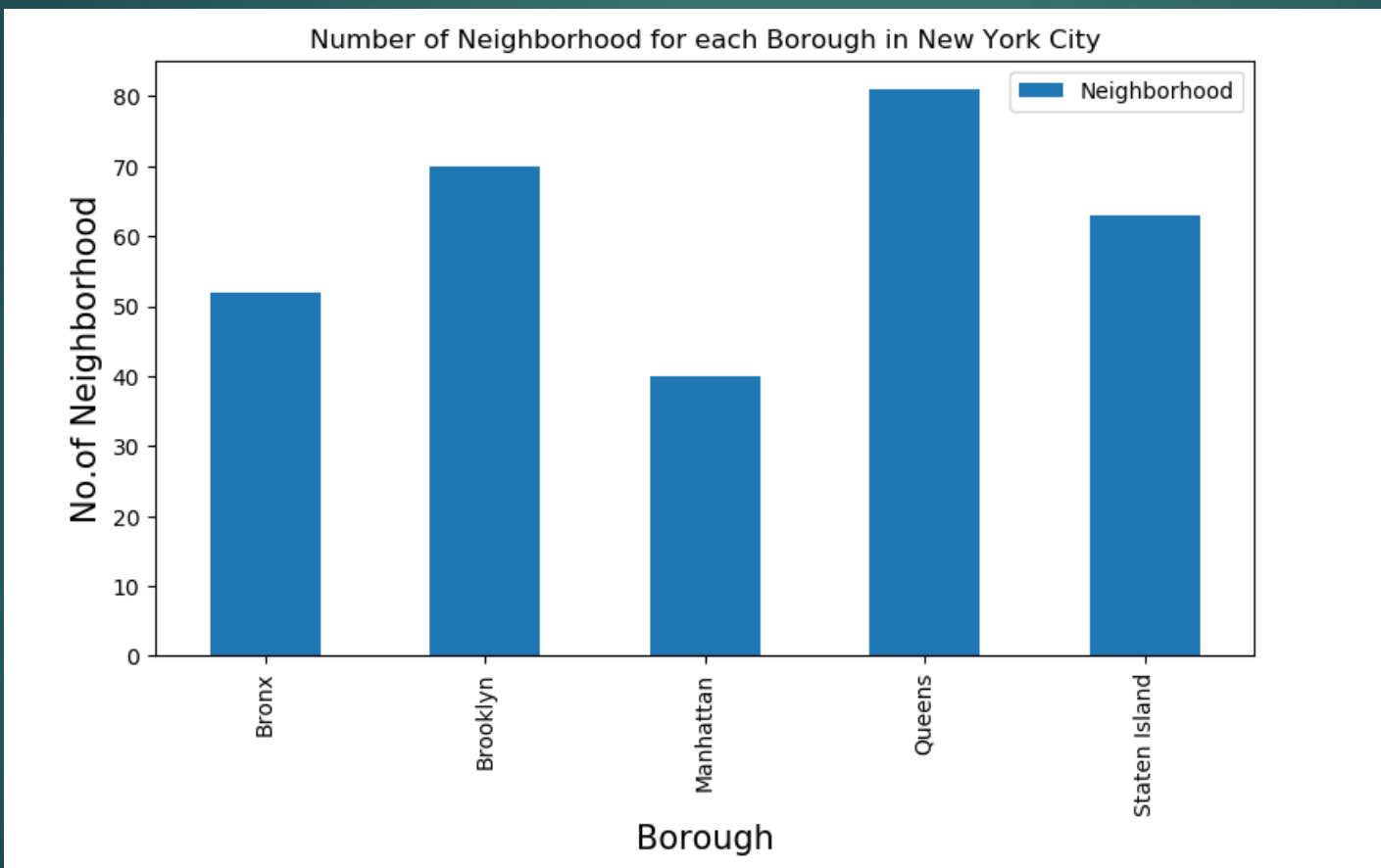
- Queens is the biggest borough with the most neighborhoods
- Brooklyn is the second biggest borough, showing that it is home to many peoply and a good potenital area to open a new venue

Venues in the neighborhoods

After calling for the venues for the individual neighborhhoods i captured them in a dataframe.
We can see that Mill Island only has two venues, whereas Brooklyn Heights, Caroll Gardens, Cobble Hill, Downtown, Greenpoint, North Side and South Side all have 100 venues (which is the cap for the individual neighborhood).(complete list can be found on github)

```
In [28]: brooklyn_venues.groupby('Neighborhood').count()
```

Out[28]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bath Beach	47	47	47	47	47	47
Bay Ridge	84	84	84	84	84	84
Bedford Stuyvesant	28	28	28	28	28	28
Bensonhurst	33	33	33	33	33	33
Bergen Beach	6	6	6	6	6	6
Boerum Hill	90	90	90	90	90	90
Borough Park	20	20	20	20	20	20
Brighton Beach	45	45	45	45	45	45
Broadway Junction	12	12	12	12	12	12

Clustering

Creating five clusters to further differentiate the neighborhoods which should help to identify interesting areas
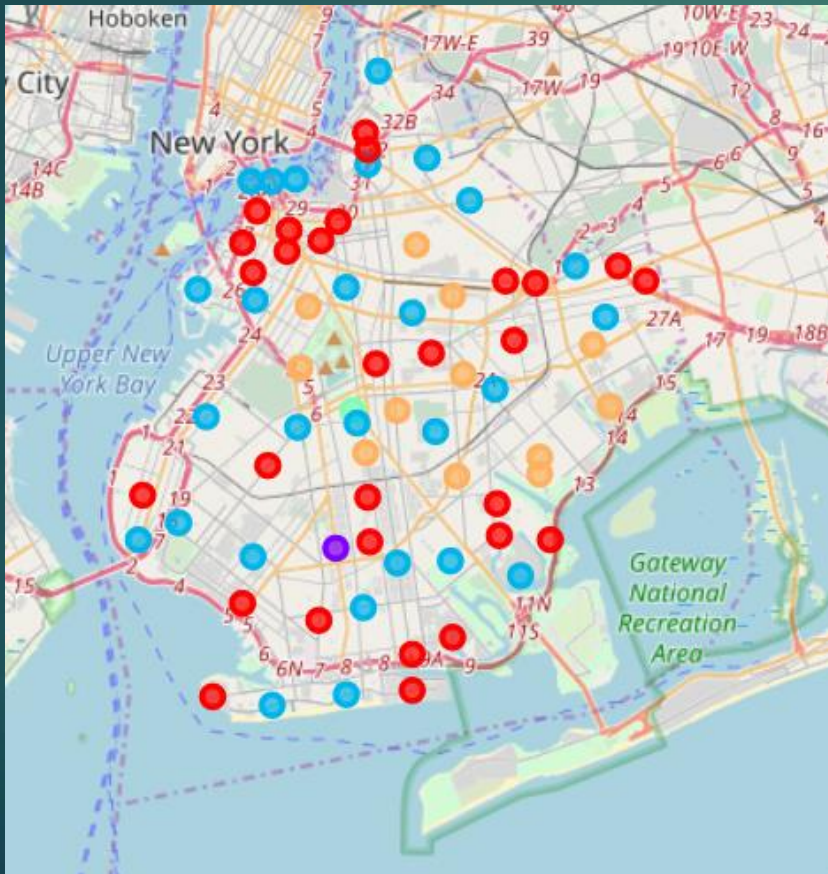
- brooklyn_cluster_0 = red
- brooklyn_cluster_1 = purple
- brooklyn_cluster_2 = blue
- brooklyn_cluster_3 = green
- brooklyn_cluster_4 = orange

No of neighborhhods for each cluster

```
In [43]: brooklyn_cluster_0['Neighborhood'].count()
Out[43]: 29

In [44]: brooklyn_cluster_1['Neighborhood'].count()
Out[44]: 1

In [45]: brooklyn_cluster_2['Neighborhood'].count()
Out[45]: 27

In [46]: brooklyn_cluster_3['Neighborhood'].count()
Out[46]: 1

In [47]: brooklyn_cluster_4['Neighborhood'].count()
Out[47]: 12
```
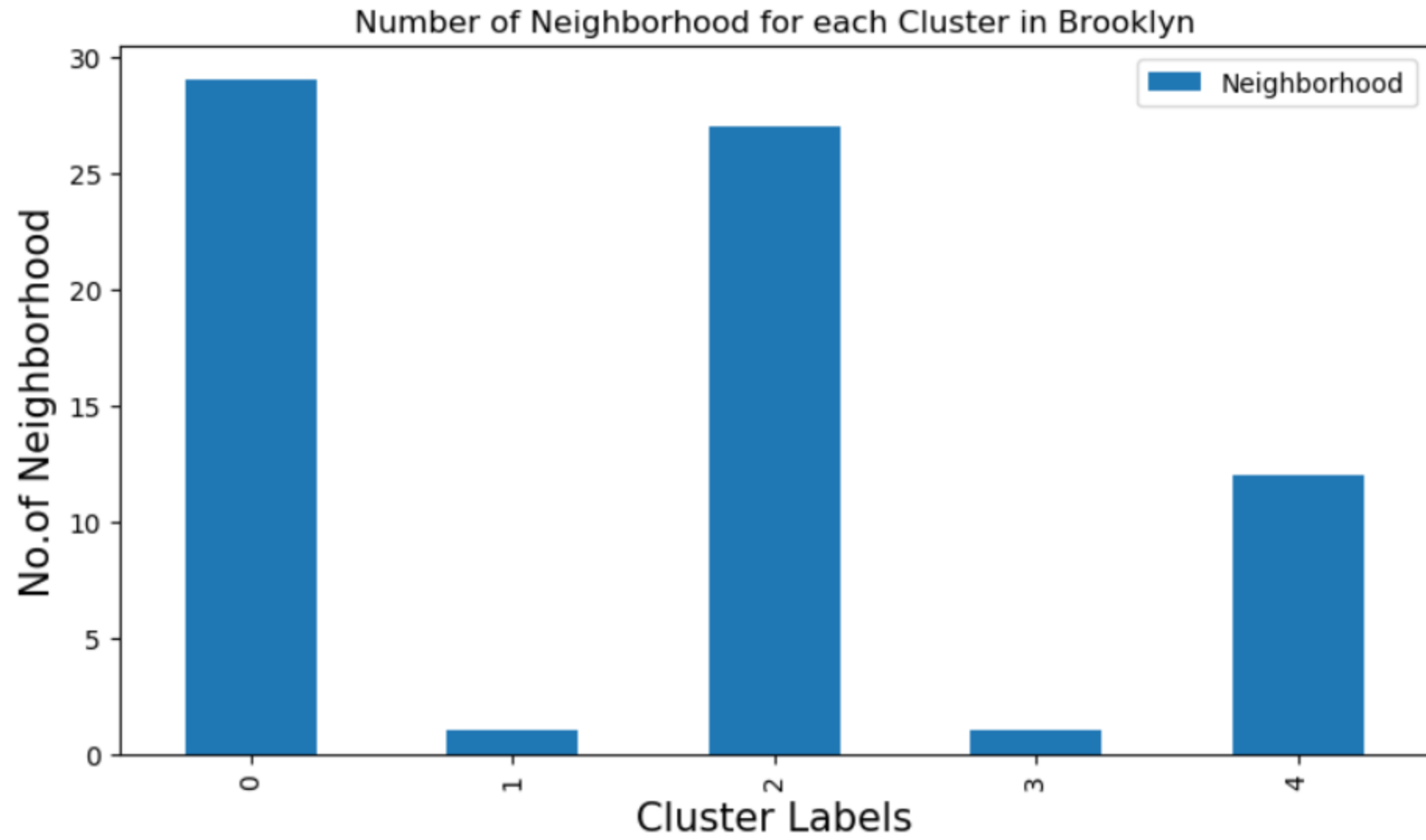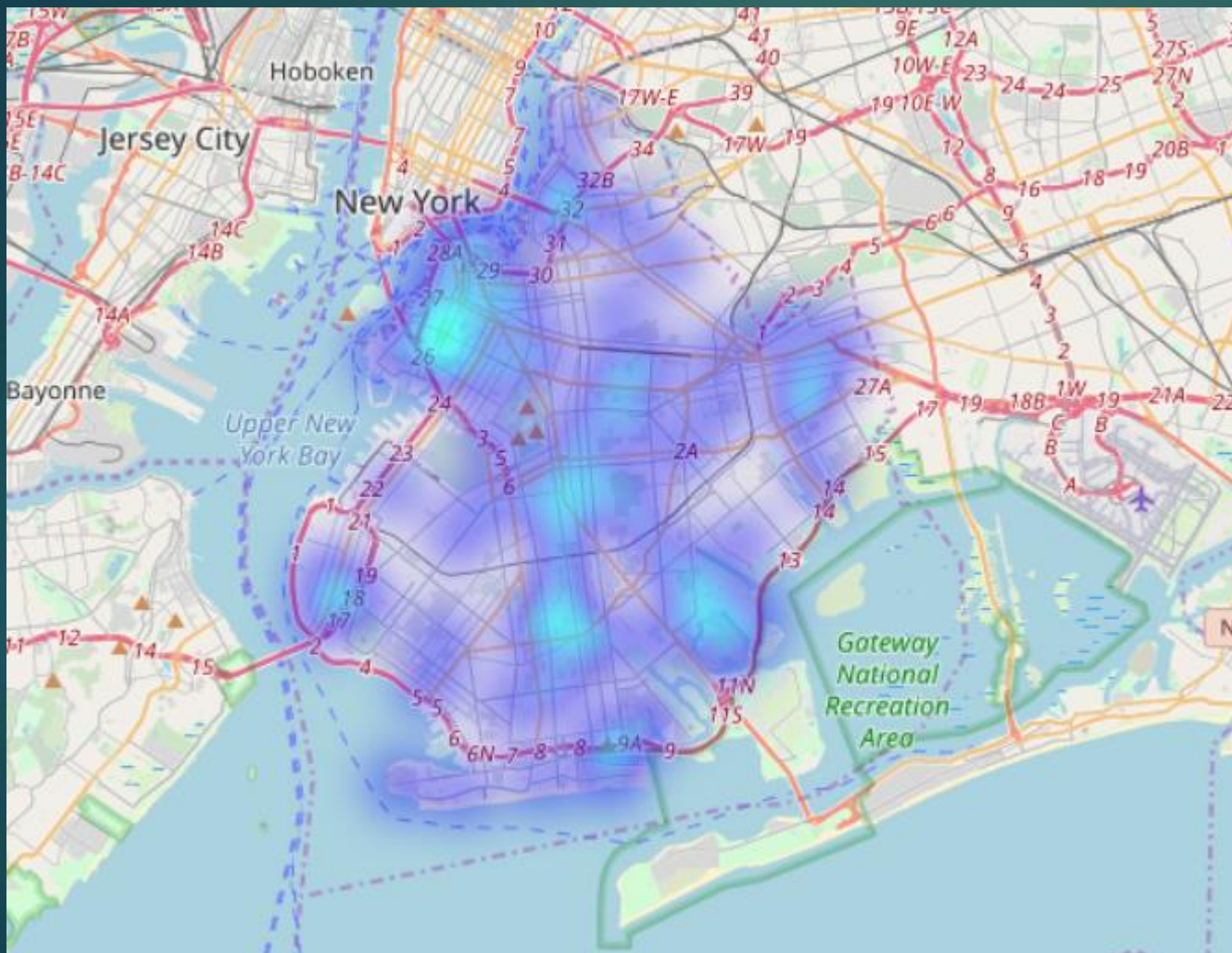
Number of Neighborhood for each Cluster in Brooklyn

Clustering Heat map

The heat map shows that especially the center and the north western region are filled with venues.

Results

- From the numbers above we can draw some conclusions regarding the most frequent venues in each cluster:

- brooklyn_cluster_0 is mainly populated by pizza places, italian restaurants, coffee shops and chicken joints. Because of that we can conclude that the people there love to go out to eat. From the two most frequent venues we can assume that the people who visit the venues from this cluster like to eat italian food.

- brooklyn_cluster_1 includes only one store which is a liquor store. Because of this, the second cluster does not seem to be a good fit.

- brooklyn_cluster_2 contains mostly bars, deli / bodegas and pizza places. This cluster covers venues for more outgoing people who like to drink. But delis / bodegas are not a good representation for venues where people would like to stay for longer and enjoy some food. Therefor the third cluster might not be the best fit.Peolpe their like to go out, but restaurants are not their highest preference.

- brooklyn_cluster_3 like cluster 2 this cluster only includes one venues, a carribean restaurant. This might resemble a good fit, but there can be better options out there.

- brooklyn_cluster_4 is dominated by carribean restaurants and asian restaurants. This cluster seems to be a good fit for a restaurant.

Results

- The used methods have shown that brooklyn_cluster_4 is characterised by a lot of restaurants and bears the potential to be a good place to open a new restaurant. Other clusters tend to be more populated with bars and places like supermarkets which are not a good fit to open a new restaurant. Cluster 5 consists of 12 neighborhoods. On the map they are marked as orange dots, that are pretty close together and not spread out like other clusters. This indicated that these neighborhoods are pretty similar. They tend to have similar people living there which is a really important aspect if you want to attract as many people as possible to your restaurant. It is much more convenient for the people to have a venue in their close proximity! Compared with the heat map we can see that cluster 5 (brooklyn_cluster_4) is not directly among the most heated areas. Tis is also an important aspect. Since the goal is to open a new venue,it is important to have an overview of the existing situation. If the current market is already flooded with venues (in this particular case with restaurants) it becomes way more difficult to establish a new venue. The market competition is fierce and to attract customers in a new environment is nearly impossible.