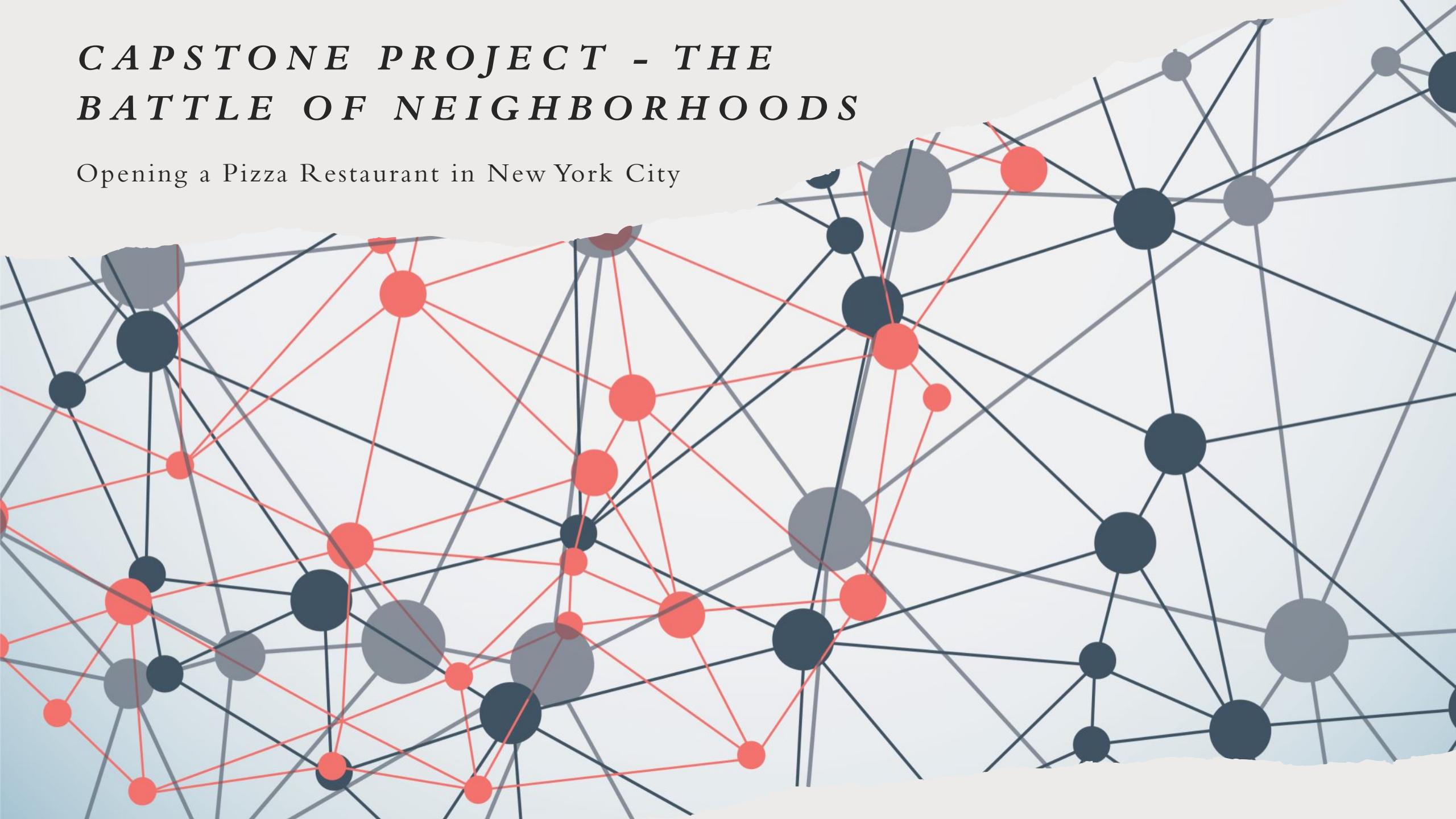


CAPSTONE PROJECT - THE BATTLE OF NEIGHBORHOODS

Opening a Pizza Restaurant in New York City



INTRODUCTION / BUSINESS PROBLEM

- In this project, I will find the best location to open a new pizza restaurant in New York. New York City is a very populated city as there are currently 8.4 million people currently living there. It is estimated that there are more than 1250 pizza places already in New York City. This means that opening a new pizza restaurant will be highly competitive. People who would be interested in this project include:
 -
 - 1. Businessmen looking to open a new pizza restaurant in New York City.
 - 2. People who want to invest in a successful pizza restaurant in New York City.

DATA



For this project we will need:

1. New York City data that contains Borough, Neighborhoods and their latitudes and longitudes.

Source for this data: https://cocl.us/new_york_dataset

We will use this data to analyze the different neighborhoods in New York City.

2. Pizza restaurants in New York City.

Source for this data: Foursquare API

We will use this data to find all of the pizza restaurants currently in New York City.

METHODOLOGY



OBTAI N DATA FOR BOROUGHS, NEIGHBORHOODS, LATITUDE AND LONGITUDE IN NEW YORK CITY

```
!wget -q -O 'newyork_data.json' https://cocl.us/new_york_dataset

with open('newyork_data.json') as json_data:
    newyork_data = json.load(json_data)
ny_data = newyork_data['features']

# define the dataframe columns
column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']

# instantiate the dataframe
newyork = pd.DataFrame(columns=column_names)
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

CLUSTERING



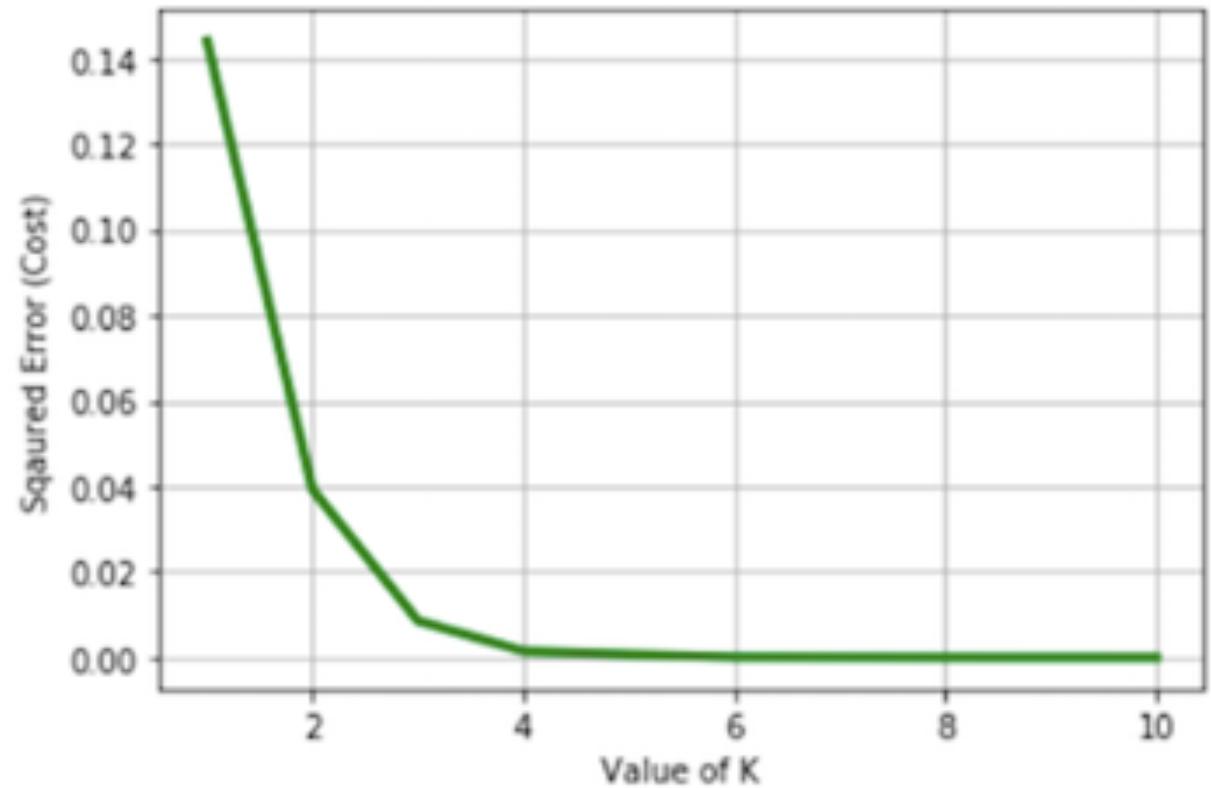
FINDING THE BEST VALUE OF K

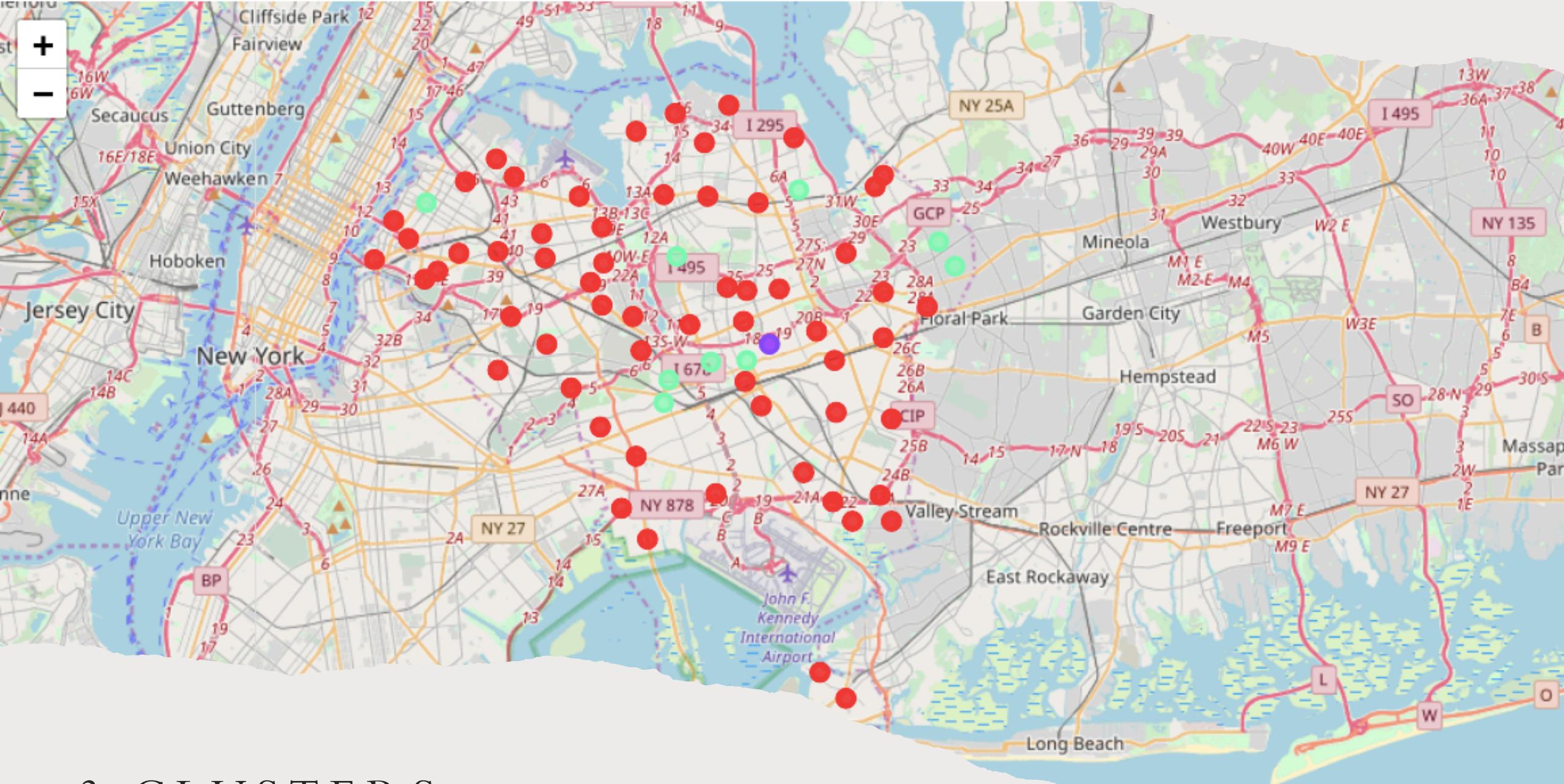
```
cost =[]
for i in range(1, 11):
    KM = KMeans(n_clusters = i, max_iter = 500)
    KM.fit(queens_grouped_clustering)

    # calculates squared error
    # for the clustered points
    cost.append(KM.inertia_)

# plot the cost against K values
plt.plot(range(1, 11), cost, color ='g', linewidth ='3')
plt.xlabel("Value of K")
plt.ylabel("Sqaured Error (Cost)")
plt.grid()
plt.show() # clear the plot
```

THE BEST
VALUE FOR K
IS 3



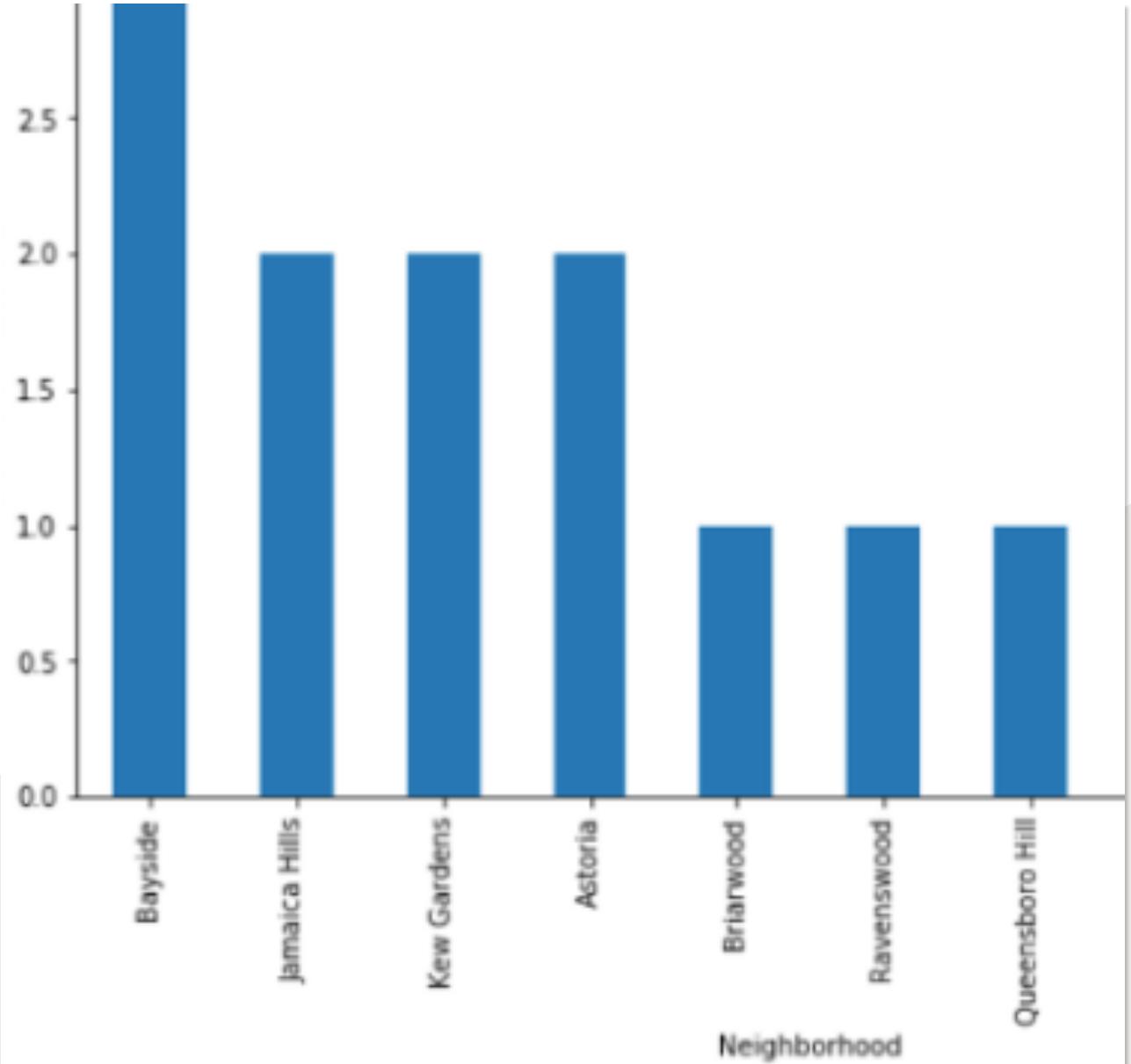


3 CLUSTERS

HOW MANY PIZZA RESTAURANTS ARE IN EACH NEIGHBORHOOD?

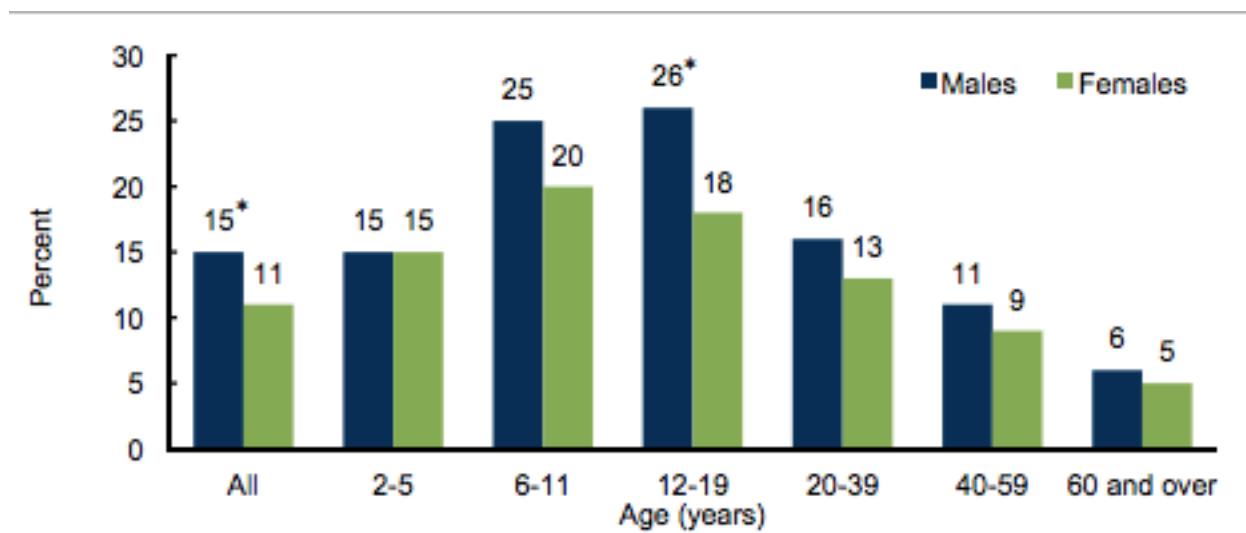
```
graph = pd.DataFrame(to_onehot.groupby('Neighborhood')['Pizza Restaurant'].sum())
graph = graph.sort_values(by='Pizza Restaurant', ascending=False)
graph.iloc[:10].plot(kind='bar', figsize=(10,6))
plt.xlabel('Neighborhood')
plt.ylabel('Number of Pizza Restaurants')
plt.show
```

BAYSIDE HAS THE
MOST
RESTAURANTS
WITH 3. THREE
NEIGHBORHOODS
ARE TIED FOR THE
LEAST WITH 1.



RESULT

- The neighborhood of Bayside has the most pizza restaurants.
- “About 1 in 8 Americans consumed pizza on any given day. More than 1 in 4 males, ages 6-19 years, consumed pizza on a day.” (www.ars.usda.gov)



DISCUSSION



I would recommend not opening the restaurant in Bayside as there is a lot of competition.



I would recommend opening the restaurant in a neighborhood that contains a lot of kids aged 6-11.



BATTERY PARK CITY

- About 36% of households there have one or more kids in Battery Park City. This is the highest percentage among all New York City neighborhoods. (www.addressreport.com)
- As stated earlier, this is the demographic that eats the most pizza. For this reason, I would recommend Battery Park City as the neighborhood to open the pizza restaurant.

CONCLUSION

In conclusion, this project is an example of how data science can solve real world problems. Data Science tools such as Python and Foursquare API can be used to make informed decisions. The tools and techniques used in this course can make a significant positive impact on our decisions if we know how to use them.