

# Coursera Applied Data Science Capstone Project

Eoghan Chelmiah

[eoghan.chelmiah.research@gmail.co](mailto:eoghan.chelmiah.research@gmail.com)  
[m](mailto:eoghan.chelmiah.research@gmail.com)

# Introduction

- population of 8,336,817 distributed over 784 square kilometres
- cultural, financial, and media capital of the world
- Brooklyn, Queens, Manhattan, the Bronx, and Staten Island
- **Business Problem**
- This objective of this project is to identify the best locations in New York City to open a gym of fitness studio facility.
- Location data will be analysed using a combination of machine learning and data science techniques, the optimal location for a property developer to build or establish a new gym facility will be determined.

## Target Audience

property developers and leisure centre managers who plan on opening a new gym or fitness studio in New York City.

Determining the optimal location for the gym will help in obtaining valuable investors for such a project.

# Data

- list of neighbourhoods in New York City
- coordinated details of each neighbourhood
- venues currently present in the surrounding area of New York City

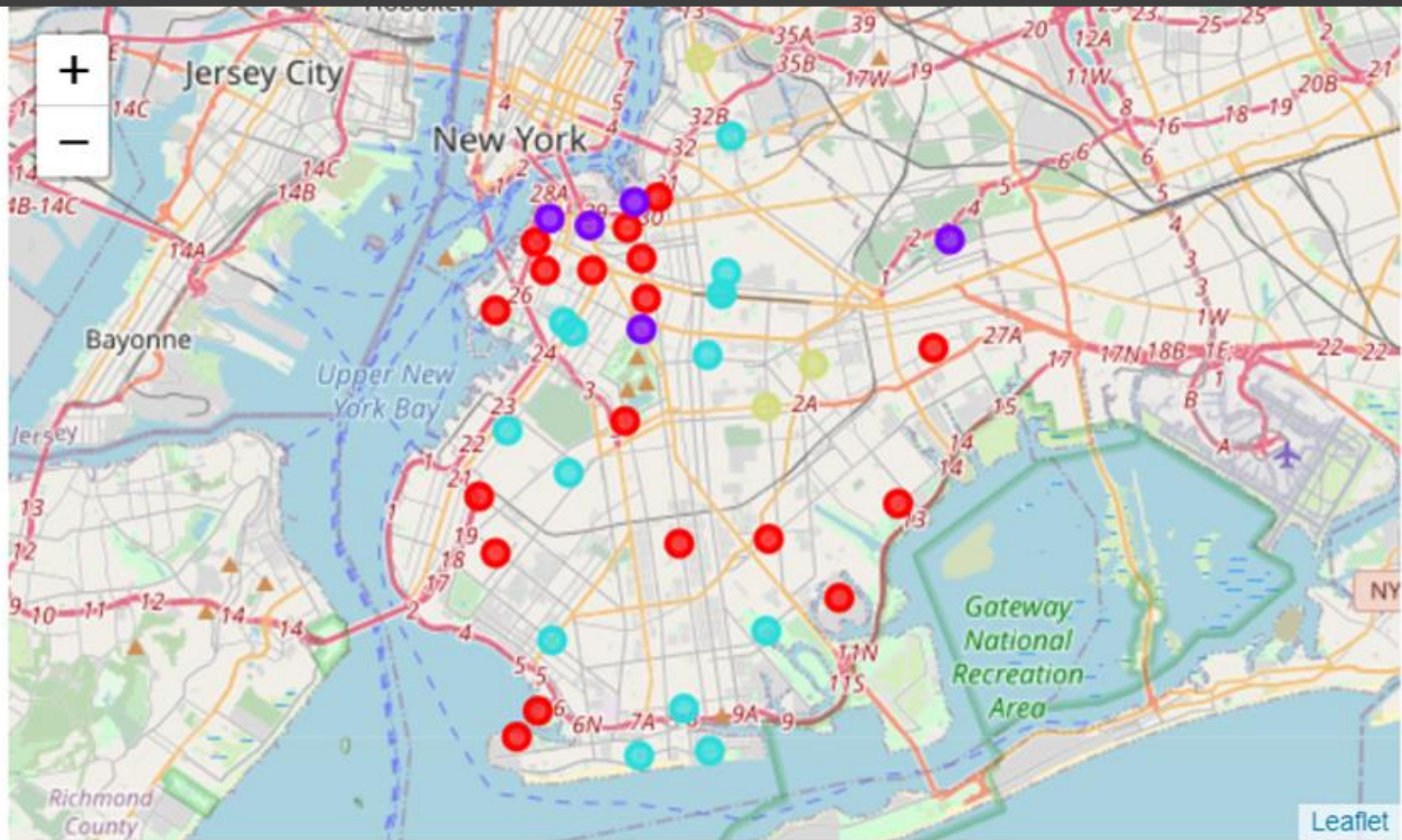
Out[7]:

	Neighborhood	Latitude	Longitude
0	► Bay Ridge, Brooklyn (1 C, 18 P)	40.633910	-74.014910
1	► Bedford–Stuyvesant, Brooklyn (1 C, 46 P)	40.683390	-73.942470
2	► Bensonhurst, Brooklyn (1 C, 28 P)	40.601860	-73.993900
3	► Boerum Hill (1 C, 15 P)	40.683720	-73.981790
4	► Borough Park, Brooklyn (1 C, 18 P)	40.638820	-73.989120
5	► Brighton Beach (1 C, 15 P)	40.576378	-73.968187
6	► Brooklyn Heights (3 C, 40 P)	40.695350	-73.994050
7	► Brooklyn Navy Yard (12 P)	40.678785	-73.944084
8	► Brownsville, Brooklyn (1 C, 12 P)	40.662900	-73.917290

# Methodology

- Data mining techniques to extract this list of locations from the webpage
- The location names and coordinates were then used to create a map visualization of New York City, using the Folium Python package.
- Foursquare is used to analyse the venue data for all the listed neighbourhoods.
- Filter using the K-means algorithm to identify the relative concentrations of gyms in each area

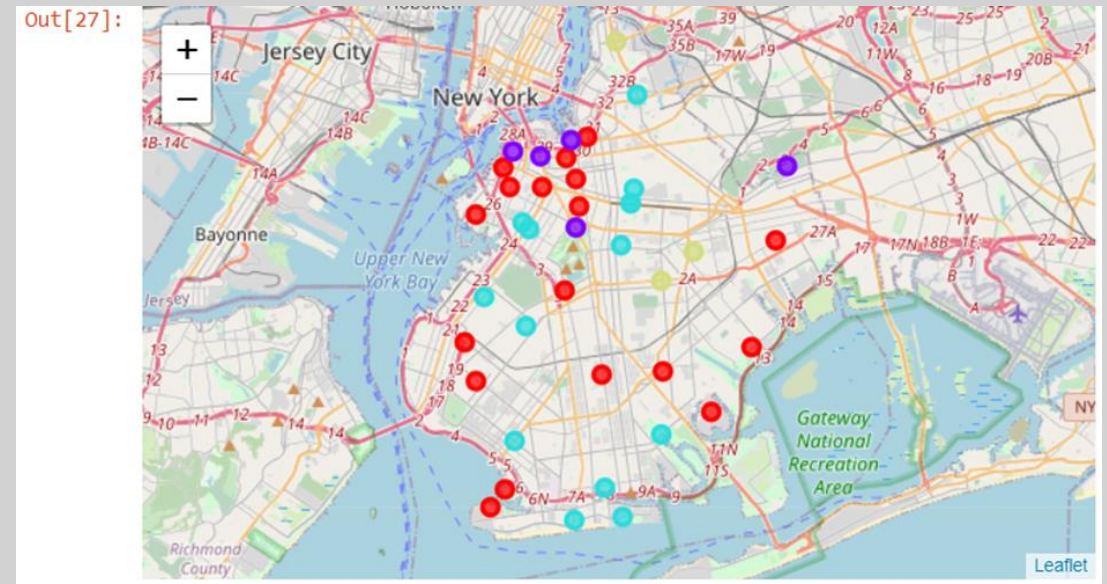
Out[27]:





# Results

- Clusters 0 and 1 had little or no gyms
- Cluster 2 had a high concentration of gyms
- Cluster 3 had a moderate concentration of gyms



# Conclusion

- Optimal location to build or establish a new gym or fitness suite would be the central neighbourhoods of Brooklyn where there are little to no competition with other fitness institutes.
- Through data mining and machine learning methods in the Python computer programming language, we were able to segment Brooklyn into clusters of neighbourhoods and analyse these 4 clusters to determine which could foster the most optimal conditions for a new gym.





Thank you  
for your  
attention

