Capstone Project - The Battle of Neighborhoods

By Arun Sridharan

Introduction

- SkoolestHub decides to setup its retail operations in New York City (NYC)
- Need the repository of high schools operating in NYC
- To setup the optimal number of stores in NYC that cater to the needs of all the schools in this area
- A type of greenfield analysis problem

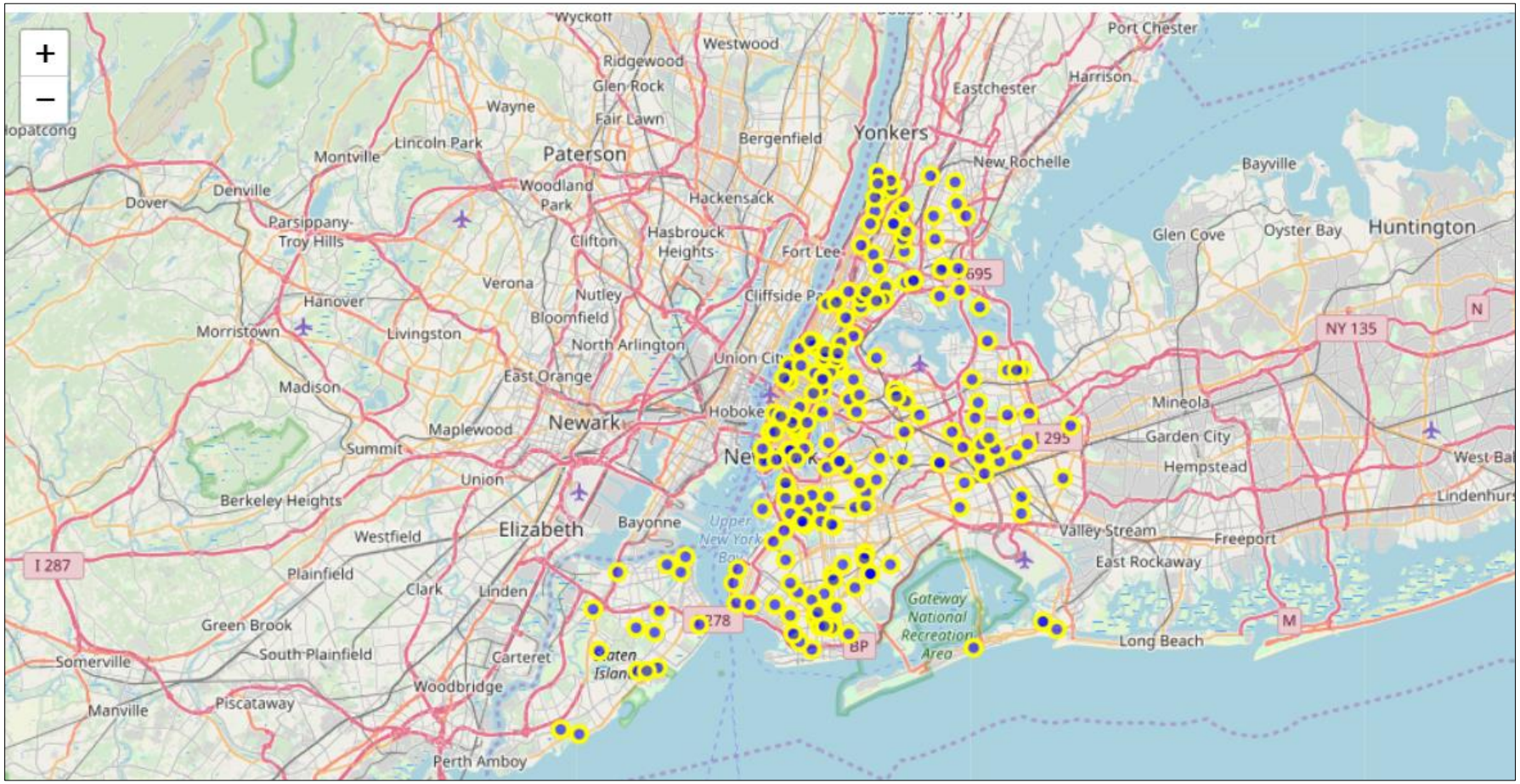
Data

- List of High Schools in New York City from Wikipedia
- Latitude and Longitude information of all schools using Foursquare API

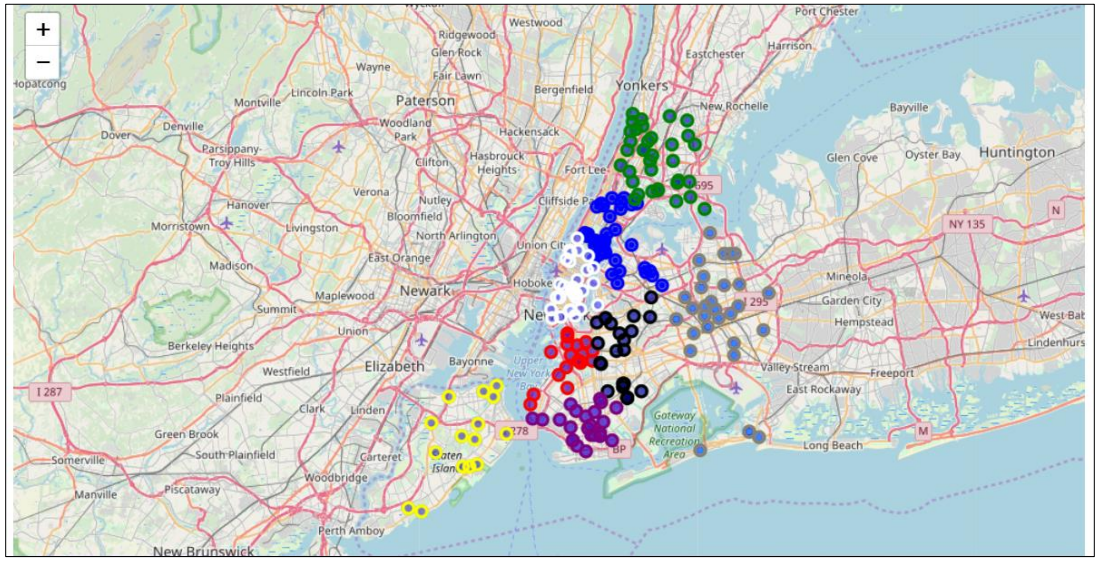
Methodology

- Data import and processing
  - Using pandas to import an excel file
- Get latitude and longitude information
  - geopy to convert address to geocodes using Foursquare API
- Data cleaning
  - Ignore schools where geocode values are NaN
- Visualization of schools on a map
  - Using folium map feature
- Segmentation of locations using clustering
  - Uses k-means clustering

Visualization of schools on a map



Results



- Based on k-means clustering technique, the optimal number of stores to be setup in NYC are is 8.
- Each 8 stores will cater to the needs of all the schools in the specific cluster.
- Map displays the 8 clusters.