

|  |
| --- |
| Capstone Project REPORT 2019  2018 |
|  |
| April 3  Coursera – Applied Data Science Course  Authored by: A Coursera Student |

# My Friendly Neighborhood Project:

# Safe School/Living Districts in New York City.

|  |
| --- |
| Introduction This project attempts to identify popular and safe areas in large cities like the City of New York based on the theory that popular places are more likely to be safe and convenient for living. Business Problem Large cities of the size of New York attracts people from all over the world. There are people who move in permanently as immigrants, some move for work, some to reunite with family and for various other reasons. For all these people, especially those with families, with kids of school going age, it is vital to identify safe neighborhood. This project attempts to solve this problem by identifying high crime rate areas which serves as an indicator for newcomers to watch out for before making decisions regarding renting or buying real estate. It will help them avoid places with relatively higher crime rates. Target Audience This project aims to benefit new comers to New York City. The data visualization will help them take an informed decision about the place to look for apartments and settle down in a safe neighborhood. |
| Data Listed are the data sources for help with this project:   1. List of boroughs and neighborhoods in the state of NYC. Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood. Here is the link to the dataset: <https://geo.nyu.edu/catalog/nyu_2451_34572> 2. GeoJSON data for New York: <https://cocl.us/new_york_dataset> 3. API calls to get location data from the following source: [https://Foursquare.com](https://foursquare.com/)   Crime Records for NYC were pulled out for year 2014 from maps.nyc.gov website. Methodology Used Listed are the steps majorly used to collect the information, interpret data and arrive at conclusion.   1. As a start, the New York City data related to its neighborhood and Burroughs was fetched using the available New York Json file. 2. The file was inspected to understand its structure to be able to traverse through it and fetch meaningful data.   Screen Shot 1: Typical structure of data fetched for a neighborhood.     1. All the NYC Burroughs and neighborhoods along with their coordinates were listed. The major places were plotted on a map.   Screen Shot 2: Sample Borough/ neighborhood data.     1. Foursquare API was invoked to gather information about venues for a certain neighborhood picked for example (- Manhattan). All the venues were examined, and information was fetched about categories of each of the venues – whether restaurants, movie theaters, public parks, tourist attraction, public schools, hospitals and so on. 2. *K-Means* *- Clustering* technique was used for clustering the neighborhoods and finally depicted such popular areas in a map. 3. A relation was being inferenced from the fact that neighborhoods with popular places like highly rated restaurants, utilities like groceries shops, public facilities were considered safe and convenient neighborhoods and are likely to attract new settlers.  Conclusion Hence safety of a neighborhood is a direct result of what categories of venues are available at a certain area. The popular the area is in terms of certain categories of venues like public utilities, hospitals, groceries, nice restaurants the better and safer the neighborhood for people to come in and settle especially for families with school going kids. |