Battle of The Neighborhoods: New York Edition

By: Hassan Sarhan

Date: 18/06/2019

# Introduction

New York City is the definition of a global city. It is the financial center of the United States and the main immigrant landing point for many immigrants to the United States. No city is represented in media as much as New York is. People across the states and all over the world envy the people that live there. However, living in New York has its downsides. The biggest downside for most people is that it is expensive. According to Investopedia.com the cost of living in Manhattan was 138.6% of the U.S. average in 2018 while Brooklyn was a little bit lower at 82% of the U.S average. The second most important thing for people who want to move to New York City is the crime rate. This is mostly because crime in New York has been overly exaggerated in the media, New York is a relatively safe city to live in. However, people moving to New York would still like to be sure that they will be safe when they arrive. The third most thing that is on people’s minds is that they want to be close to the trendiest bars, cafes, and venues. So, what I am suggesting is that there is this one perfect neighborhood that has it all. A neighborhood that is close to trendy venues, has a low crime rate, and is relatively affordable to live in. That mythical neighborhood is real, and this report will be about how I found it. We will also create a rudimentary ranking system and we will introduce a multiple linear regression model to predict the ranking of a neighborhood based on three different features.

# Data

In this report I will be working with five main data sets representing New York City’s neighborhoods, venue details, crime rate, average price by neighborhood, and neighborhood ranking. This will involve scraping websites, cleaning the data, feature selection, and merging the data

## New York City Neighborhood Names

The New York City Neighborhood Names data set comes from this NYU [website](https://geo.nyu.edu/catalog/nyu_2451_34572). It has many features for each neighborhood, but we are only interested in four of them: Borough name, Neighborhood name, Latitude, and Longitude

## Foursquare Venue Data

The data about the location of each venue comes from the Foursquare [Places API](http://www.foursquare.com/). Each venue comes with a bucket filled with features. In this report we’ll only extract the name, latitude, and longitude.

## New York City Criminal Activity Data

This data set comes from the City of New York [website](https://data.cityofnewyork.us/Public-Safety/Crime-Map-/5jvd-shfj). It details all crimes reported by the NYPD in 2016. It contains the details of the crime and the location the crime occurred; however, it does not list the name of the neighborhood, making it difficult to merge with the other data sets. We will address this issue later

## New York City Rent Cost by Neighborhood

This dataset cannot be downloaded from a singular website. It had to be scraped from multiple pages. I will explain the process later in the report, but what you need to know now is that this dataset contains various neighborhoods around New York City and the average rent cost in each one.

## New York City Neighborhood Ranking

This dataset comes from niche.com and it ranks each neighborhood in New York City based on a wide number of features, including proximity to schools, safety of the neighborhood, and job opportunities. This dataset also cannot be obtained through normal means from a particular website. It had to be scraped from multiple pages. I will explain the process later in the report.