NEIGHBOURHOOD SAFETY IN NYC

JULY 28

IBM DATA SCIENCE CAPSTONE PROJECT
Authored by: Ifeoluwa Anjorin

INTRODUCTION

When you choose to live in New York, NY, you are choosing a lot more than a place to call home. You are choosing a lifestyle, defined by a city teeming with possibilities. Encompassing 8.4 million residents spread throughout five boroughs, New York's intersecting cultures and communities have established the city as a global destination for arts and culture, commerce and cuisine, and everything in between.

Hence, if anyone is considering moving to a new area of the city, it does not matter whether they are making the switch from Manhattan or the Midwest — people are likely to be following the same basic checklist. After the standard affordability factor, people also ask; where are the safest areas in NYC? Does this potential new neighborhood have relatively easy access to the subway? Access to good restaurants and bars? Do I feel comfortable walking there alone at night?

Cost, convenience, neighborhood amenities and security are only some of the factors that renters and buyers consider in selecting a home, and it is not likely that any one factor would be decisive. While safety statistics are probably not the only factor a prospective renter or buyer should consider in choosing a home, safety data is one type of information that may help to evaluate housing choices.

It is important to be proactive about one's safety, especially when it comes to choosing a place to live. After all, the safety of your neighborhood affects everything from one's happiness and sense of security to one's home resale value.

PROBLEM STATEMENT

The objective of this capstone project is to analyze and profile neighborhoods across all boroughs in New York City based on safety. Using data science methodologies and machine learning techniques like clustering, this project aims to provide solutions to answer the question: If a homebuyer or renter were looking for the safest place to stay in NYC, where would you recommend?

DATA

To solve the problem, we will need the following data:

- Neighborhoods across New York. This defines the scope of the project that is confined to the city of New York and its five boroughs.
- Latitude and longitude coordinates of those neighborhoods. This is required in other to plot the map and give venue data
- Crime data related to arrests for some specific offences and venue data related to police departments. We will use this data to perform clustering on the Neighborhoods.

DATA SOURCES

The Crime Data is a List of every arrest in NYC in 2017. This data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning before being posted on the NYPD website. Each record represents an arrest effected in NYC by the NYPD and includes information about the type of crime, the location and time of enforcement. The data can be found here :

https://data.cityofnewyork.us/Public-Safety/NYPD-Arrests-Data-Historic-/8h9b-rp9u

In addition, Foursquare API will be used to get the venue data for those neighborhoods. Foursquare has one of the largest database for millions of locations and is used by thousands of developers. Foursquare provides many categories of venue data but for this project, the interest lies in Police department category.