Table of Contents

Introduction

Background

New York City (NYC) is one of the most important financial centers in the world, it's the most populous city in the United States (U.S.) with an estimated population of 8,398,748 in the year 2018 and covers 784 km^2.

NYC cosmists of 5 boroughs and its core is at Manhattan where all the world's major commercial, financial and cultural centers are located.

Problem

This study will rank neighborhoods from each borough according to its total crime rates and tag each neighborhood with a safety score. Using the safety score as a feature together with top 10 visited categories of venue (using Foursquare API) from each neighborhood to find clusters using K-Means Clustering.

By Clustering the neighborhoods, we should be able to find out some relationship between safeness and type of places people more likely to visit.

Target Audience

NYPD and new residents moving to NYC would benefit from the results from this study.

NYPD could use the results to better understand which area has higher crime rate and what type of places people are more likely to visit in those areas.

New residents moving to NYC will be able to know not only the safeness of the neighborhoods but also what kind of places they could visit in different neighborhoods. On top of that, it could also be a deciding factor for them to choose the area they wish to stay.

Dataset

Data Source

The New York City Crimes dataset can be obtained from Kaggle. The dataset contains all criminal offences recorded by the NYPD from 2014 to 2015.

Data Wrangling

Crimes dataset consists of 24 columns, 6 of them were chosen for this project:

- **CMPLNT_NUM**: The unique ID of the criminal offence.
- **OFNS_DESC**: Offence Descriptions.
- **LAW_CAT_CD**: Offence Categories.
- **BORO_NM**: Borough, where crime was committed.
- **Latitude**: Latitude of the crime scene.
- **Longitude**: Longitude of the crime scene.