Coursera Capstone Project
The Battle of Neighborhoods

Section 1: Introduction

1.1 Background

In this world, crimes are an inseparable part of our lives. Every day we hear about them. Being cautious and improve safety is not a simple instruction anymore. We need to use modern technology and data science techniques to more wisely act against this problem. There are so many records and documentation in the police department that have been gathered during the years, which can be used as a valuable source of data for the data analytics tasks. Applying analytical task to these data bring us valuable information that can be used to increase the safety of our society and lower the crime rate. The average American moves about eleven times in their lifetime. We should always do proper research when planning our next move in life. Safety is a top concern when moving to a new area. If you don't feel safe in your own home, you're not going to be able to enjoy living there.

1.2 Problem

In this project we analyze the New York Crime dataset, which is one of the richest open source data in this area, to get a better understanding about the security status of this city. The crime rates in each borough may have changed over time. The project aims to select the safest areas in each borough in New York based on the lowest crime rates, explore the neighborhoods of that borough to find the 10 most common venues in each neighborhood and finally cluster the neighborhoods using k-mean clustering.

1.3 Interest

People who are considering relocating to New York or the travelers planning to visit this city will be interested to identify the safest borough in NYC and explore its neighborhoods and common venues around each neighborhood.

Section 2. Data Description

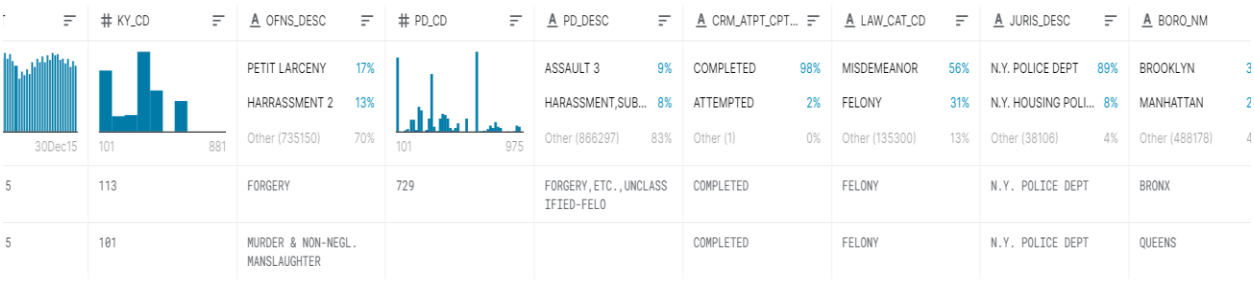
Based on definition of our problem, factors that will influence our decisions are:

- finding the safest borough based on crime statistics
- finding the 10 most common venues
- choosing the right neighborhood within the borough

We will be using the geographical coordinates of New York City to plot neighborhoods in a borough that is safe and in the city's vicinity, and finally cluster our neighborhoods and present our findings.

Following data sources will be needed to extract/generate the required information:

- New York City crimes reported : https://www.kaggle.com/adamschroeder/crimes-new-york-city?select=NYPD_Complaint_Data_Historic.csv
  This dataset consists of the crimes reported in New York city by NYPD. We can get the crimes committed in each borough of NYC and then compare them with each other to find the safest of them.



	KY_CD	OFNS_DESC	PD_CD	PD_DESC	CRM_ATPT_CPT...	LAW_CAT_CD	JURIS_DESC	BORO_NM
	30Dec15 / 101 / 881	PETIT LARCENY 17% HARRASSMENT 2 13% Other (735150) 70%	101 / 975	ASSAULT 3 9% HARASSMENT,SUB... 8% Other (866297) 83%	COMPLETED 98% ATTEMPTED 2% Other (1) 0%	MISDEMEANOR 56% FELONY 31% Other (135300) 13%	N.Y. POLICE DEPT 89% N.Y. HOUSING POLI... 8% Other (38106) 4%	BROOKLYN MANHATTAN Other (488178)
5	113	FORGERY	729	FORGERY,ETC.,UNCLASSIFIED-FELO	COMPLETED	FELONY	N.Y. POLICE DEPT	BRONX
5	101	MURDER & NON-NEGL. MANSLAUGHTER			COMPLETED	FELONY	N.Y. POLICE DEPT	QUEENS

- Population of NYC by Borough: https://www.kaggle.com/adamschroeder/crimes-new-york-city?select=Population_by_Borough_NYC.csv
  Population dataset can be used to analyze the proportion of total crimes to total population in that borough so that the crime rate is calculated without any bias or flaws.

Age Group Population group, usually classifies y age ranges, here the data it's general	Borough	1950 1950 Census population	1950 - Boro shar...	1960 1960 Census population	1960 - Boro shar...
1 unique value	6 unique values	6 total values	6 unique values	6 total values	6 unique values
Total Population	NYC Total	7,891,957	100%	7,781,984	100%
Total Population	Bronx	1,451,277	18.39%	1,424,815	18.31%
Total Population	Brooklyn	2,738,175	34.7%	2,627,319	33.76%
Total Population	Manhattan	1,960,101	24.84%	1,698,281	21.82%
Total Population	Queens	1,550,849	19.65%	1,809,578	23.25%
Total Population	Staten Island	191,555	2.43%	221,991	2.85%

- Scraping additional information of the different Boroughs in NYC from a Wikipedia page.: Boroughs of New York City

Borough	*County*	*Estimate (2019)*[3]	*billions (US$)*[4]	*per capita (US$)*	*square miles*	*square km*	*persons / sq. mi*	*persons / km²*
The Bronx	Bronx	1,418,207	42.695	30,100	42.10	109.04	33,867	13,006
Brooklyn	Kings	2,559,903	91.559	35,800	70.82	183.42	36,147	13,957
Manhattan	New York	1,628,706	600.244	368,500	22.83	59.13	71,341	27,544
Queens	Queens	2,253,858	93.310	41,400	108.53	281.09	20,767	8,018
Staten Island	Richmond	476,143	14.514	30,500	58.37	151.18	8,157	3,150
City of New York		**8,336,817**	**842.343**	**101,000**	**302.64**	**783.83**	**27,547**	**10,636**
State of New York		19,453,561	1,731.910	89,000	47,214	122,284	412	159

- Creating a new consolidated dataset of the Neighborhoods, boroughs, and the most common venues and the respective Neighborhood along with co-ordinates.: This data will be fetched using Four Square API to explore the neighborhood venues and to apply machine learning algorithm to cluster the neighborhoods and present the findings by plotting it on maps using Folium.