CORRELATIONS BETWEEN DATASETS USING DYNAMIC TIME WARPING

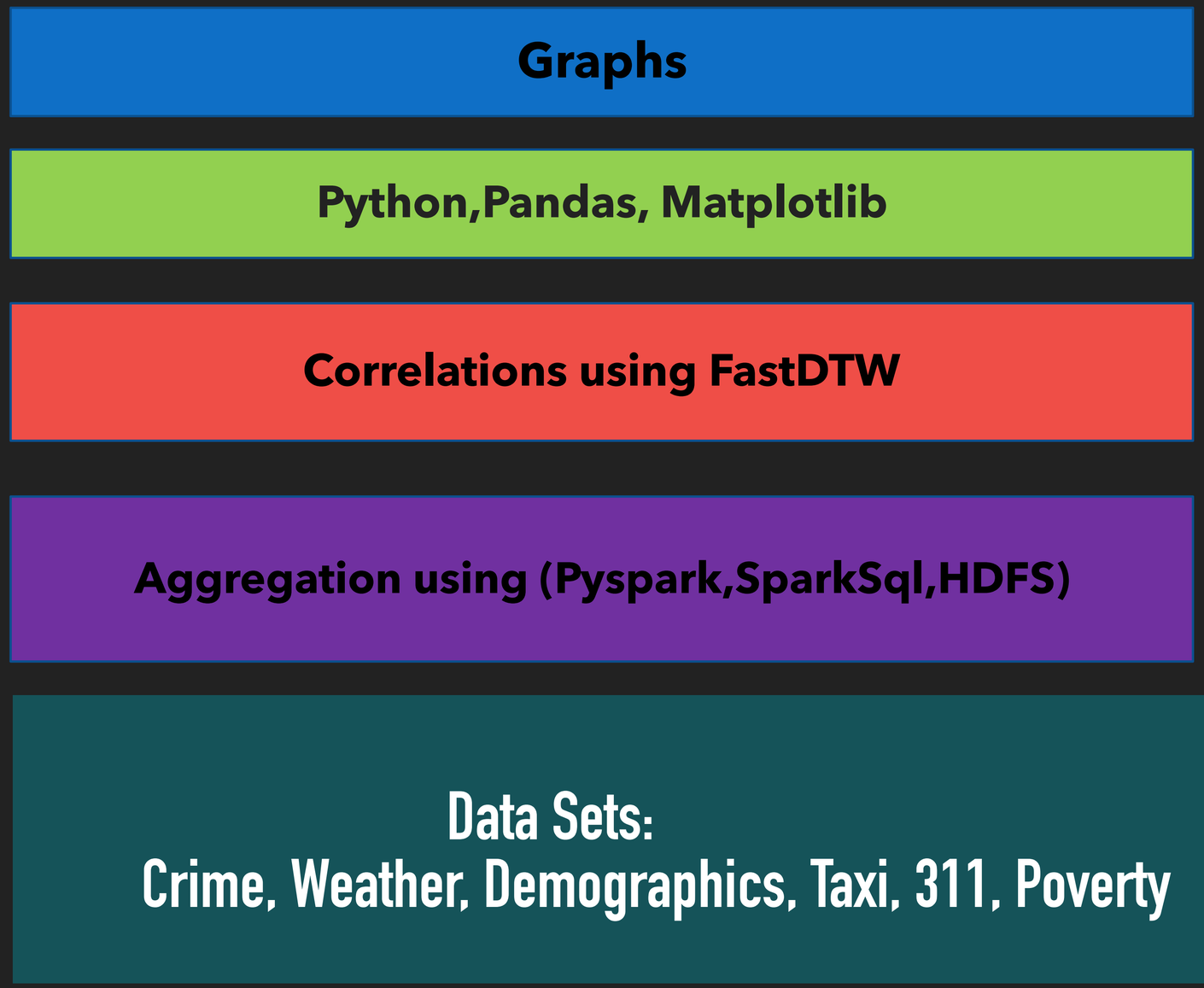
PREPARED BY:
KHUSHNASEEB ALI
HASSAN FAROOQUI
JEET KAMDAR

PROJECT GOAL

- To discover and understand correlations between various datasets related to New York using Dynamic Time Warping

- To find interesting and meaningful relationships between datasets - (Weather, Demographics, Crime, Poverty, Taxi and 311 Data)

ARCHITECTURE



DATA GATHERING

- Crime Data (NYC Open Data)
- Weather Data (NYU BOX)
- Demographics (NYC City Data Portal)
- Income Data (NYC City Data Portal)
- Taxi Data (NYC Open Data)
- 311 Data (NYC Open Data)
- Poverty Data (NYC Gov Data Portal)

DATA CLEANING

- Fixing Column Names (Same column different representation)
- Null Values
- Date formatting – different date formats in different datasets
- Renaming columns
- Slicing two datasets into same temporal and spatial resolution

CORRELATION ANALYSIS APPROACH

- Approach:
  Align and aggregate the two datasets into various sequences according to different temporal resolutions - yearly, monthly and daily.
- Calculate the correlation using DTW:

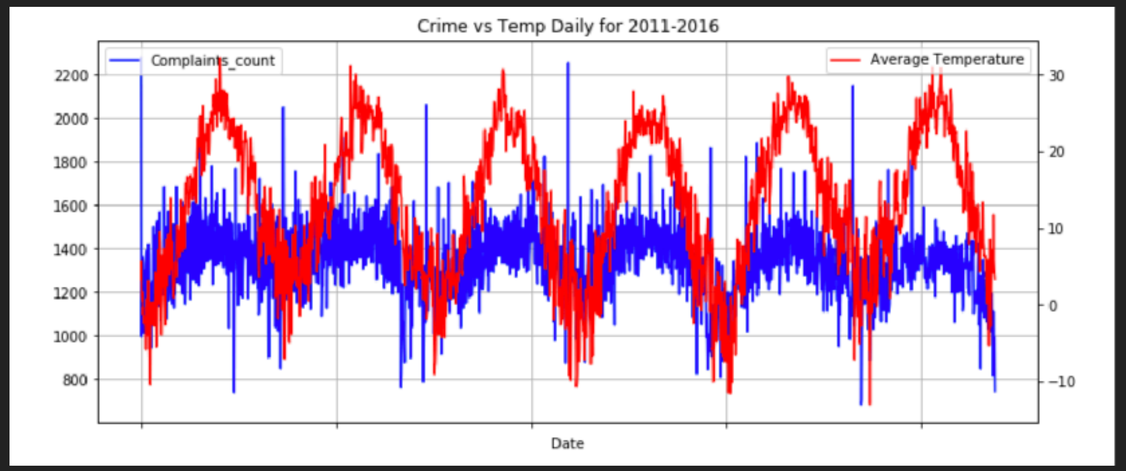
$$\beta_{DTW}(X,Y) = 1 - \frac{\mathrm{DTW}(X,Y)}{\mathrm{DTW}(X,0) + \mathrm{DTW}(0,Y)}$$

- If β is near 0 = > different/ not correlated sets
- If β is near 1 => identical/correlated sets

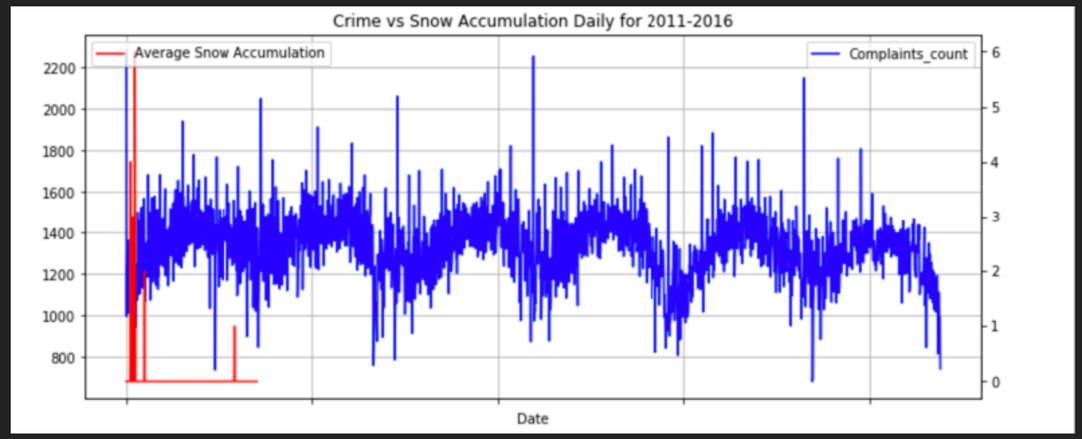
CRIME AND WEATHER

CRIME AND WEATHER CORRELATION ANALYSIS 2011–2016

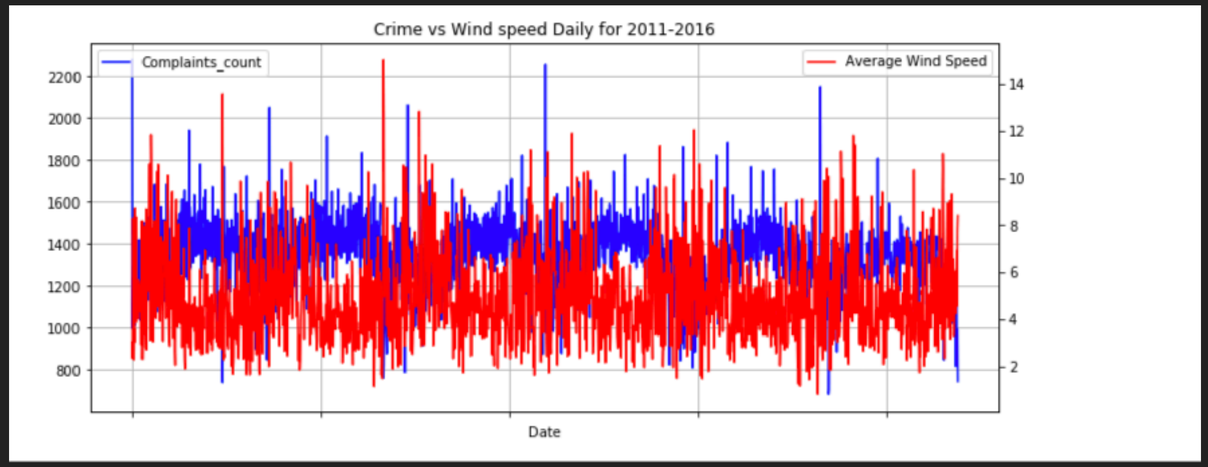
Temperature



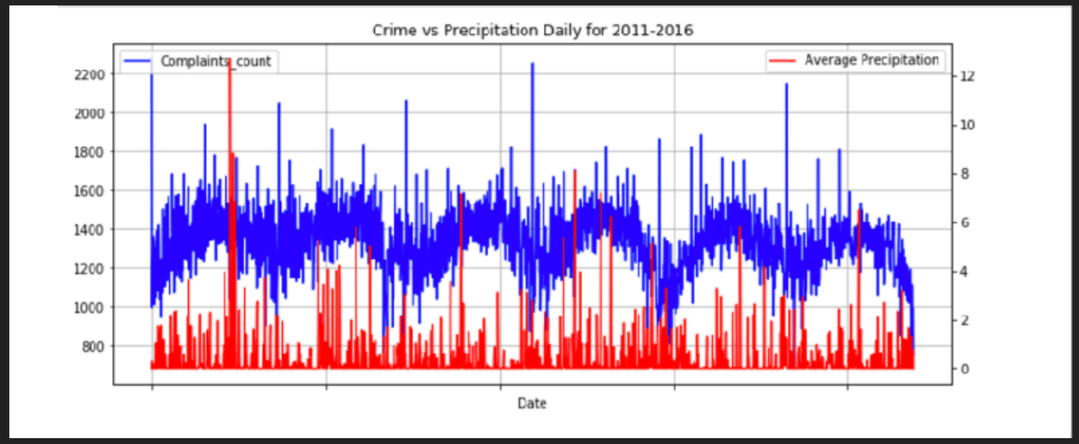
Snow



Wind Speed



Precipitation

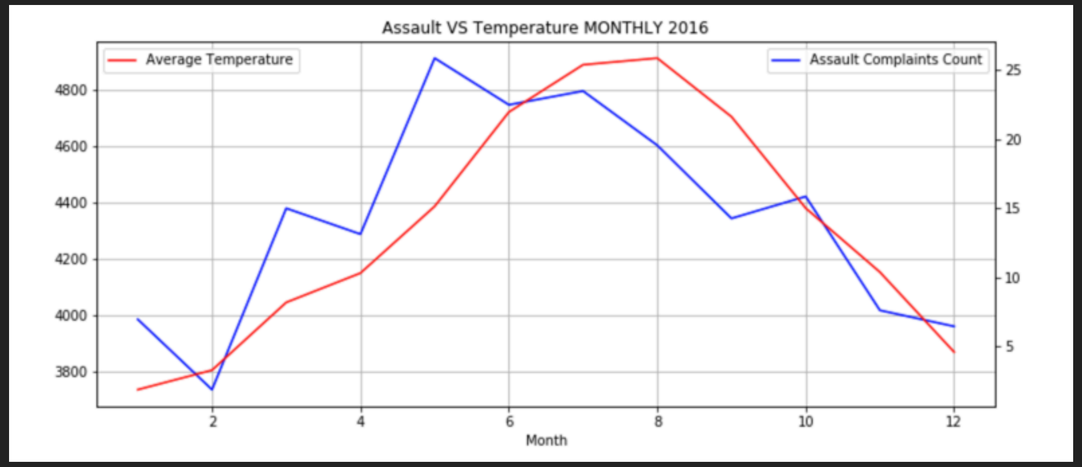


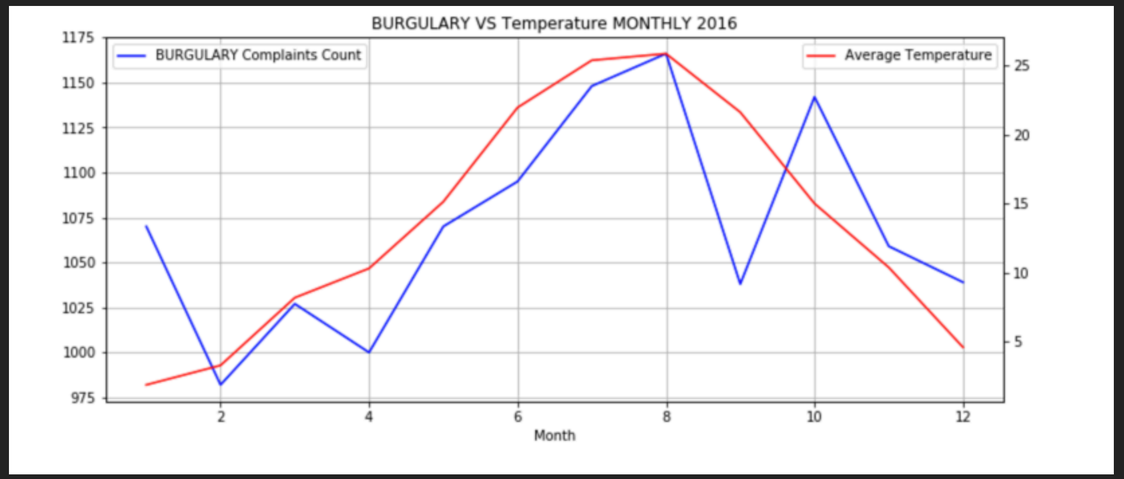
CORRELATION VALUES

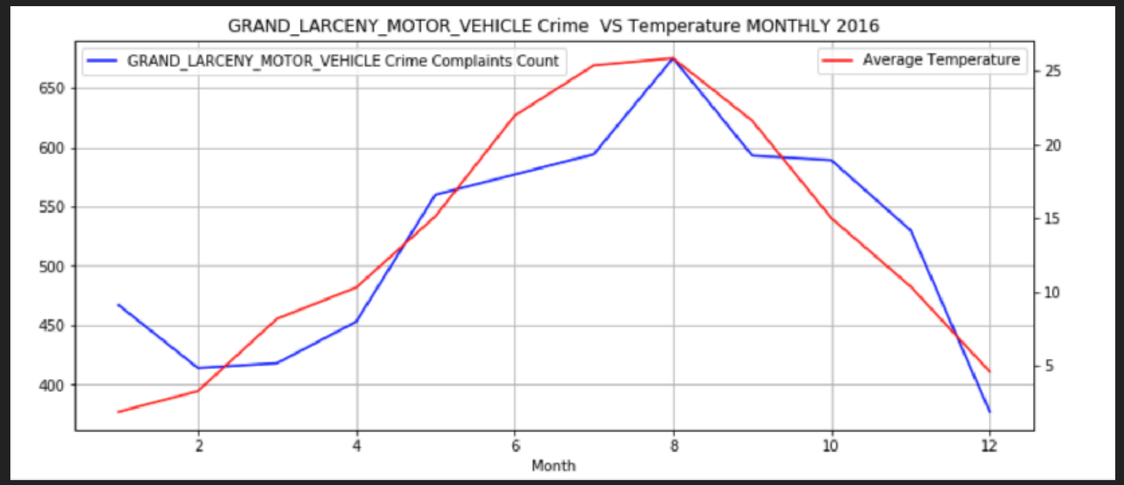
Crime and Weather From 2011-2016 Daily.

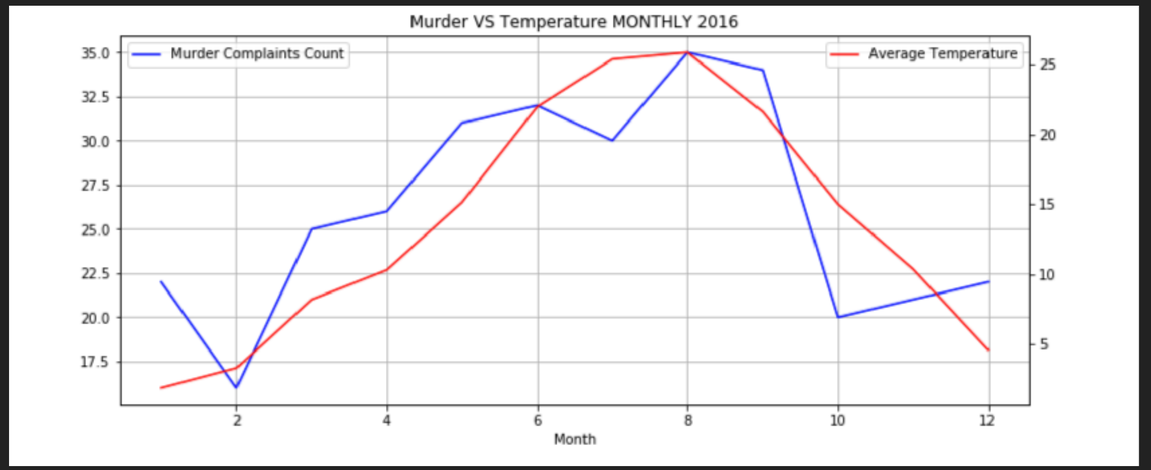
Weather Parameters	Total Complaints Registered
Avg. Temperature	0.62
Avg. Wind Speed	0.47
Avg. Snow Accumulation	0.14
Avg. Precipitation	0.39

CRIME CATEGORIES AND WEATHER CORRELATION 2016









CORRELATION VALUES – MONTHLY 2016

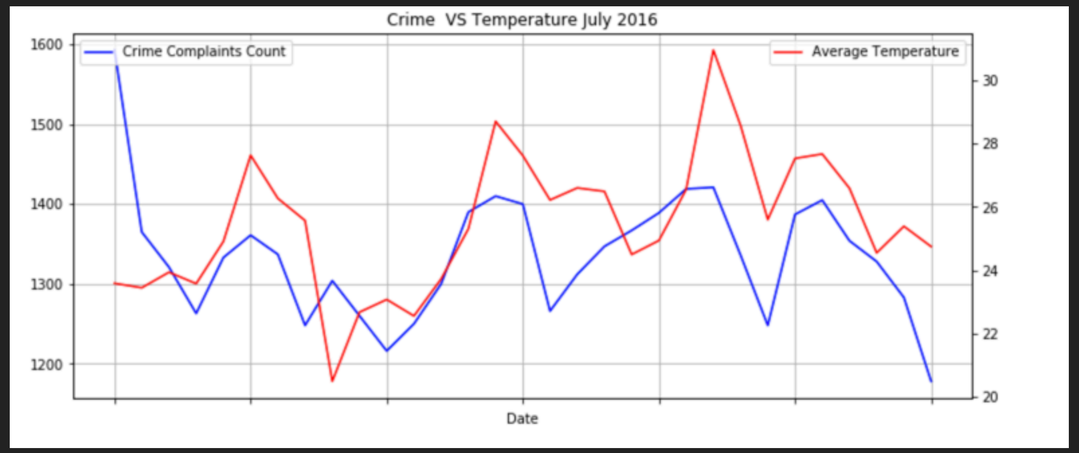
Crime Categories and Weather - 2016 Monthly

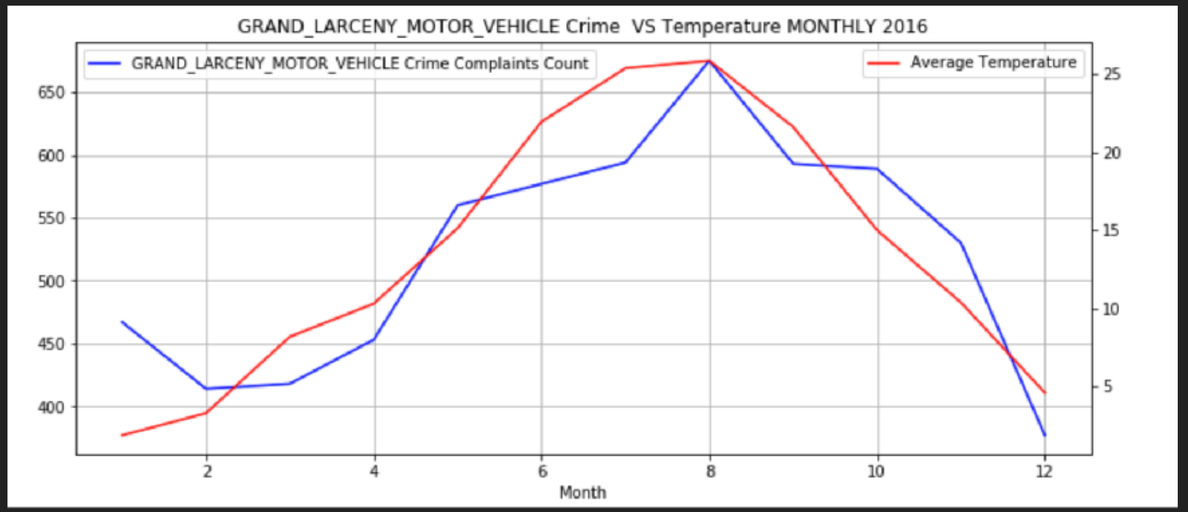
Crime Category	Average Temperature
Assault	0.62
Murder	0.75
Burglary	0.67
Vehicle and Traffic Crimes	0.64
Grand Larceny and Motor theft	0.79

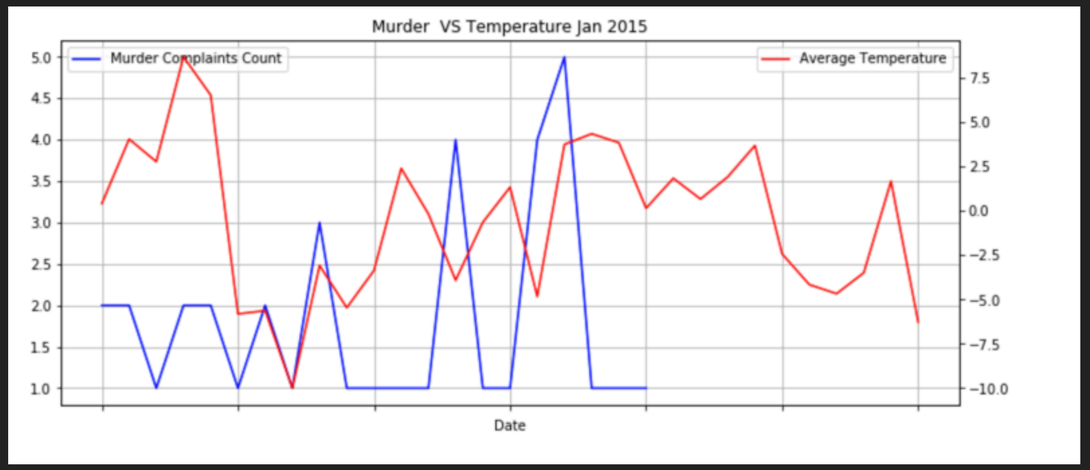
INTERESTING OBSERVATIONS

- **Assault and Murder complaints count are highest in summer** as compared to winter. Warmer temperatures bring out more aggression leading to more assaults. Also, there are more people out and about, which means there are more potential victims outside.
- **Highest number of burglaries in Summer**. During the summer, many people start to leave their windows open to try to cool off their houses. Additionally, they tend to leave the house more frequently as they attend different summer activities. Both of these activities give burglars a perfect opportunity to enter the home
- **Vehicle and traffic laws crime are maximum in summer** especially in June as higher temperatures bring aggression and thus road rage resulting in more violations

Focussed Analysis on Specific Crimes in Specific Months,





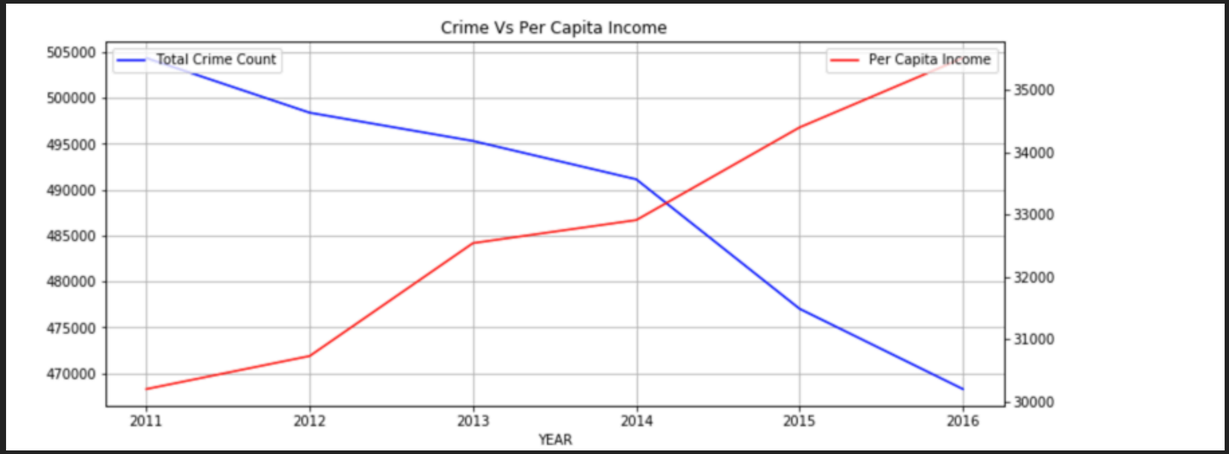


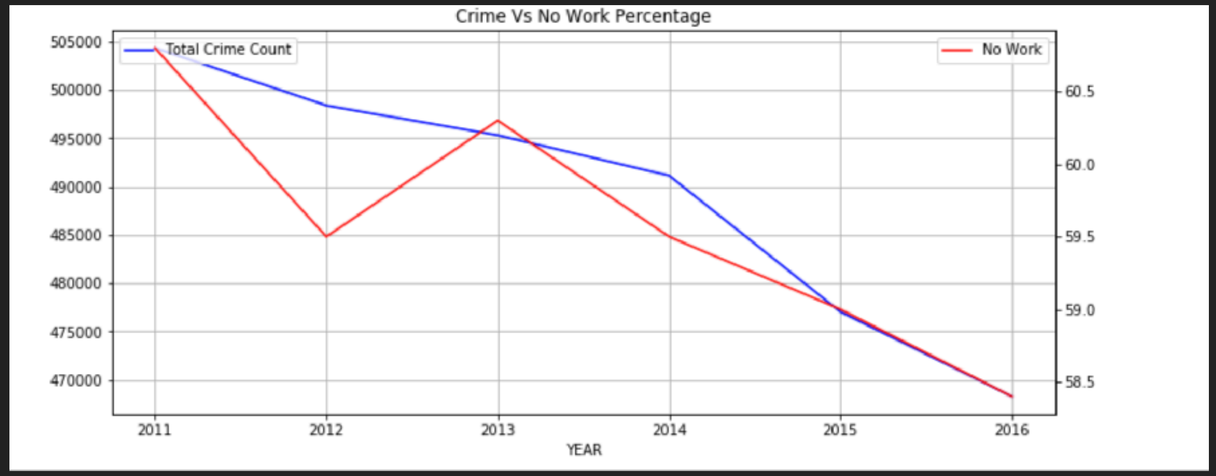
INTERESTING OBSERVATIONS

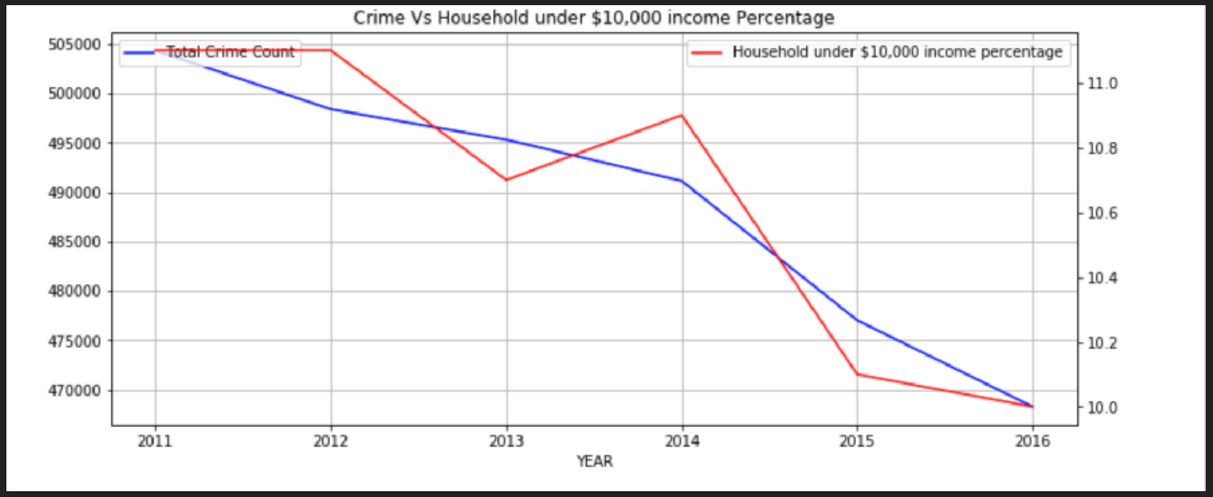
- Additionally, many **crimes continue to decrease as the temperatures get colder**. Zero cases of Homicides and very few murders in Jan 2015, went the temperature was extremely low
- One exception: car theft. In fact, **car theft incidences actually jump when temperatures takes negative values.**
- While crime seems to spike when it's warm outside, **if temperatures rise above 32 degrees, crime starts to decline.** People seek shelter inside and no longer spend as much time outside.

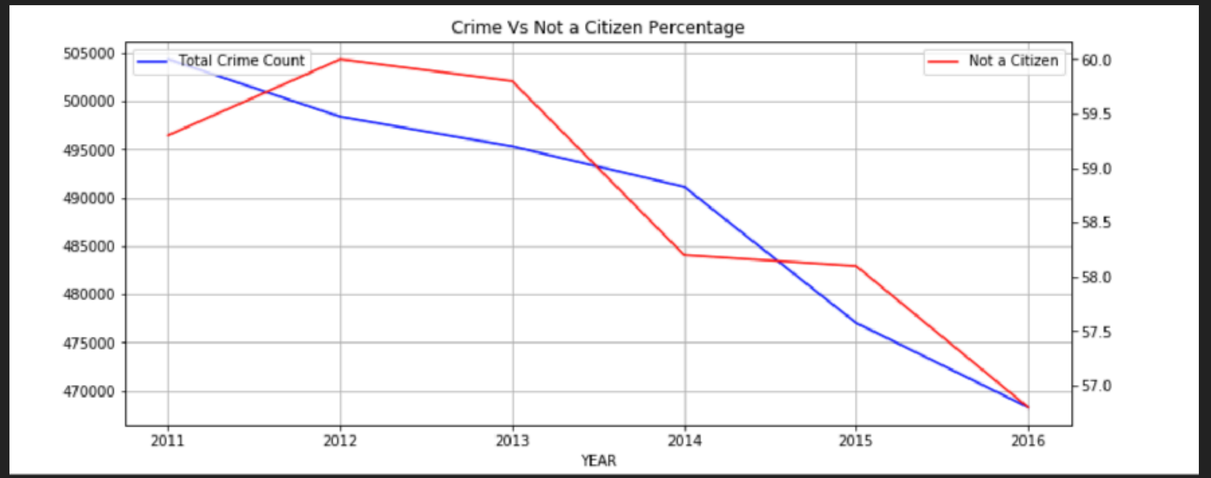
DEMOGRAPHICS AND CRIME

CRIME AND DEMOGRAPHICS CORRELATION ANALYSIS 2011–2016









CORRELATION VALUES

Crime and Demographics - 2011-2016

Demographic Factors	Correlation with Crime
Per CAPITA Income	0.02
No work Percentage	0.78
Not a Citizen Percentage	0.76
Full Time Work Percentage	0.65
Bachelor's Degree or higher Percentage	0.76
Household under $10,000 income percentage	0.84

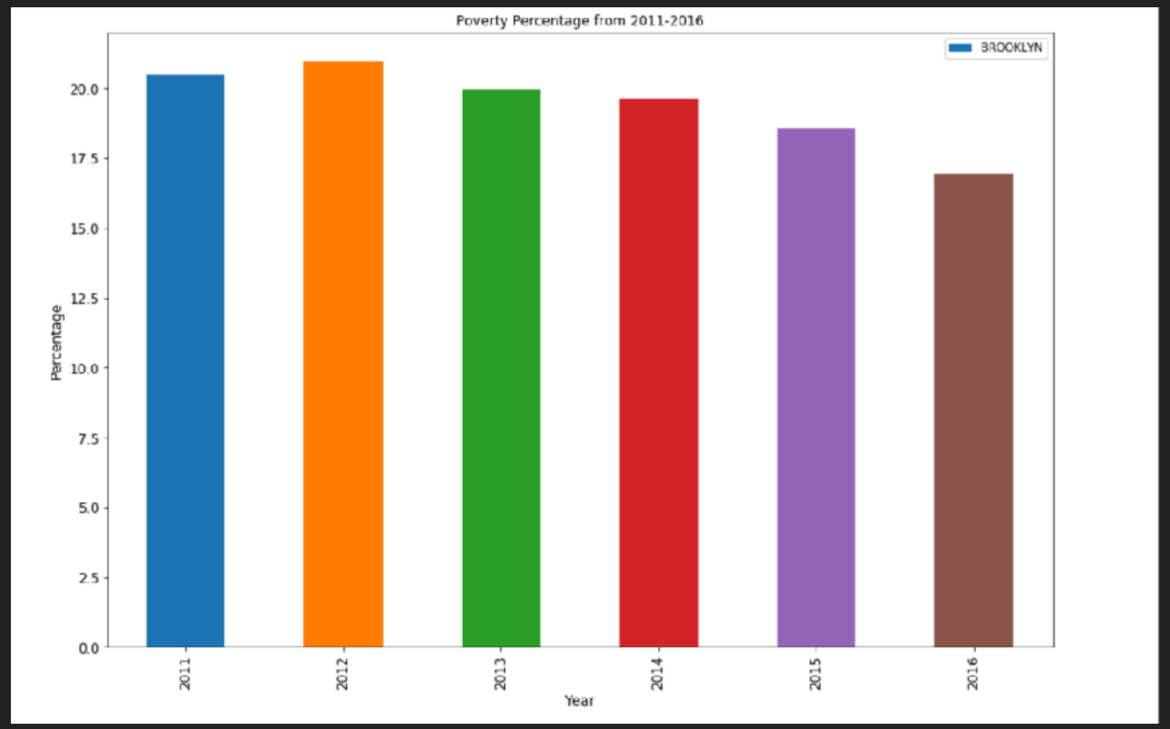
INTERESTING OBSERVATIONS

- Crime decreases as Per Capita Income Increases.
- Crime rates decline as the percentage of people not working also decreases
- Crime decrease with Decrease in percentage of households with an earning under $10,000
- Crime rates tend to decline as the percentage of people with no work falls.
- Crime rates tend to decline as the percentage of people with a full time work increases.

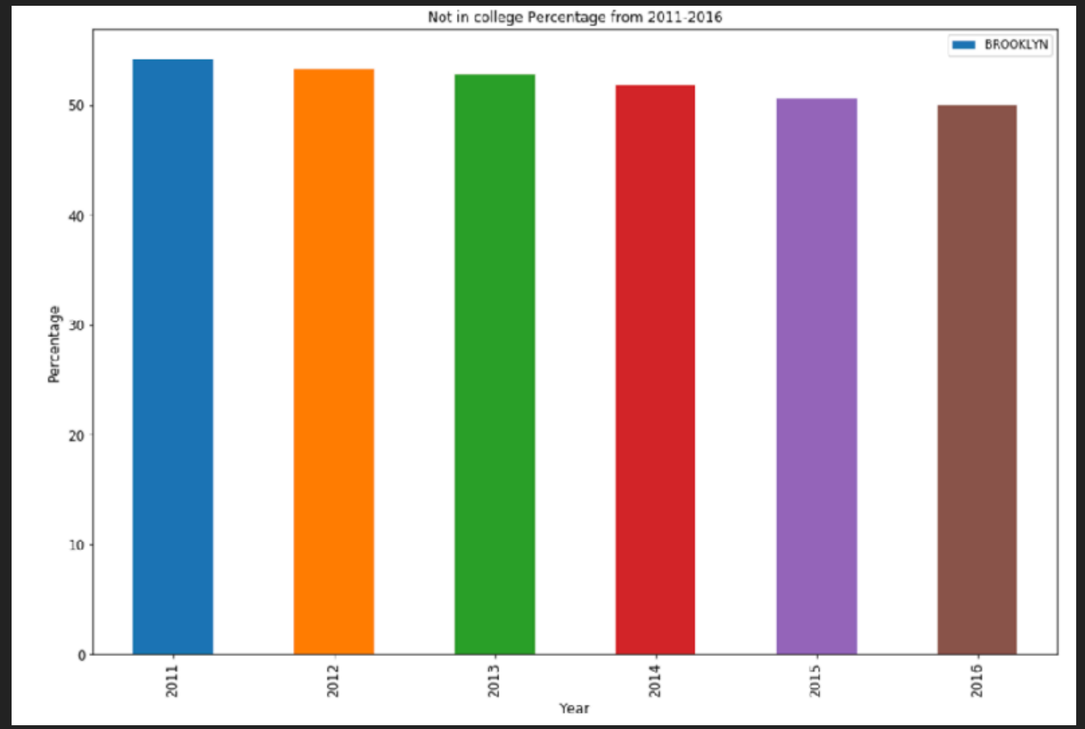
POVERTY AND CRIME

CRIME AND POVERTY CORRELATION ANALYSIS

BROOLKLYN, 2011-2016



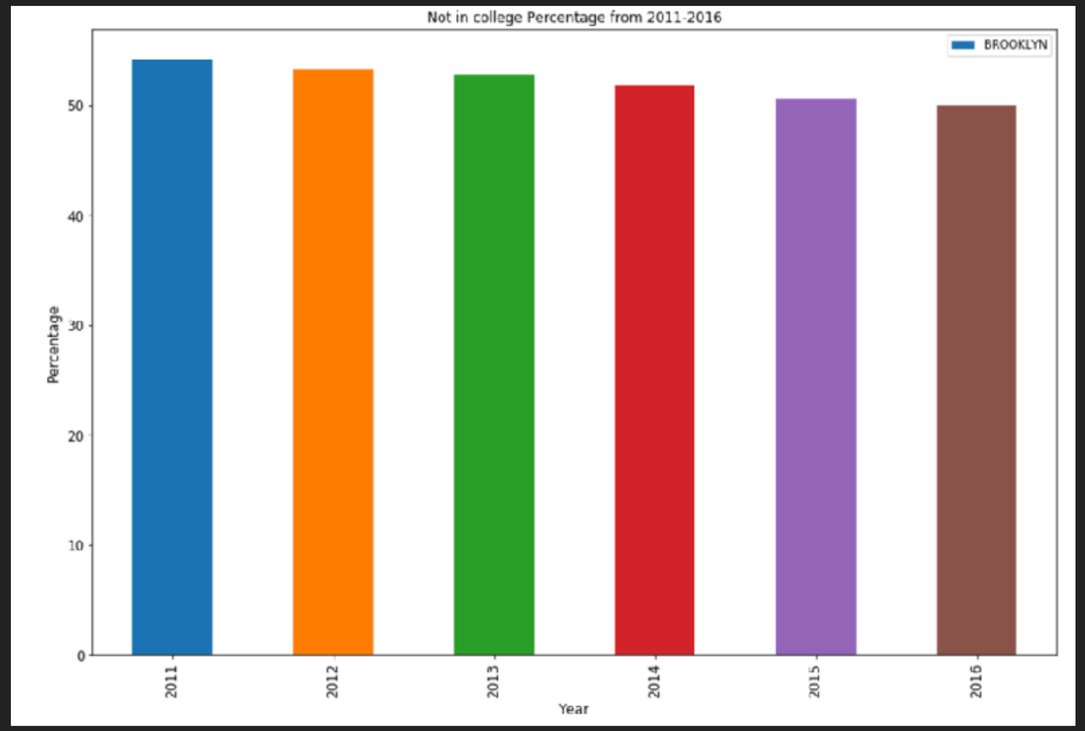
Percentage of Poor



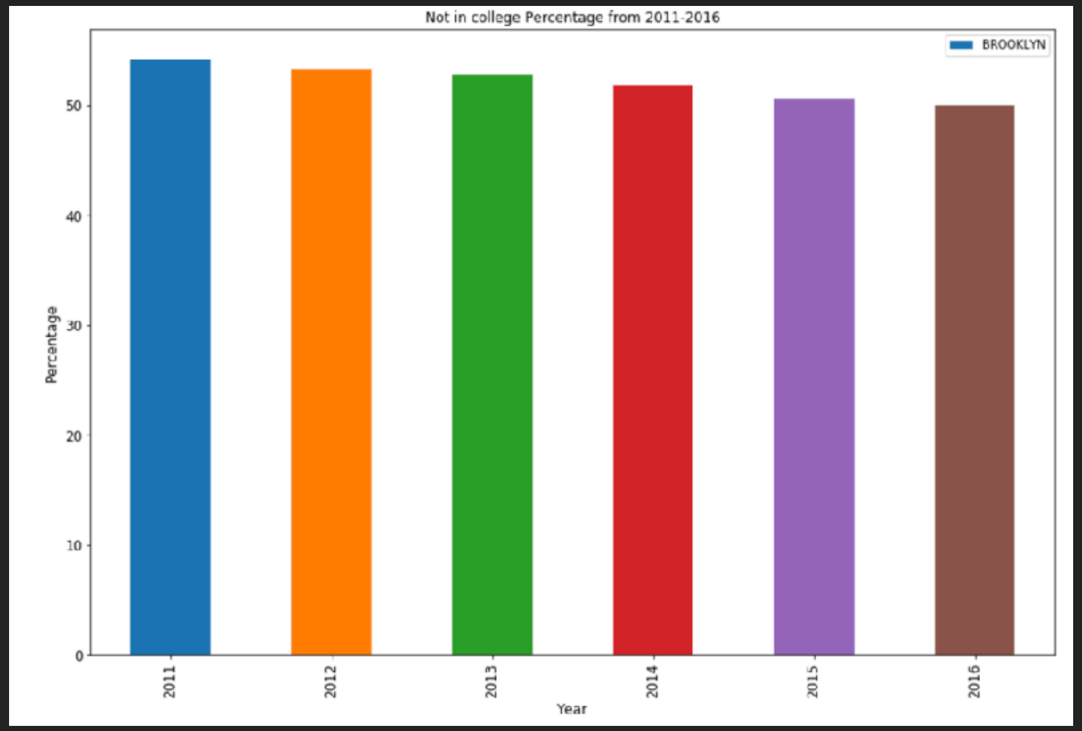
Number of Murder complaints

CRIME AND DID NOT GO TO COLLEGE CORRELATION ANALYSIS

BROOLKLYN, 2011-2016



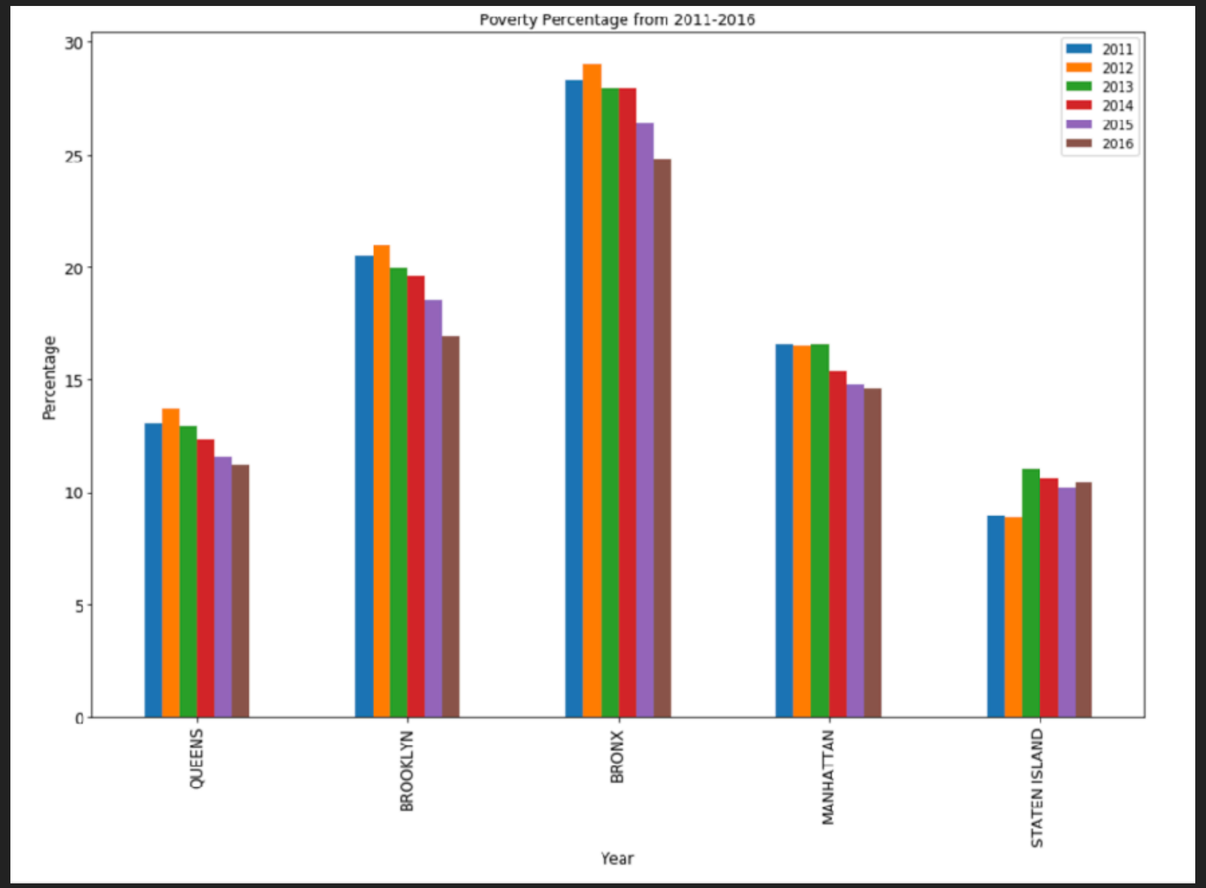
Percentage of People who did not go to College

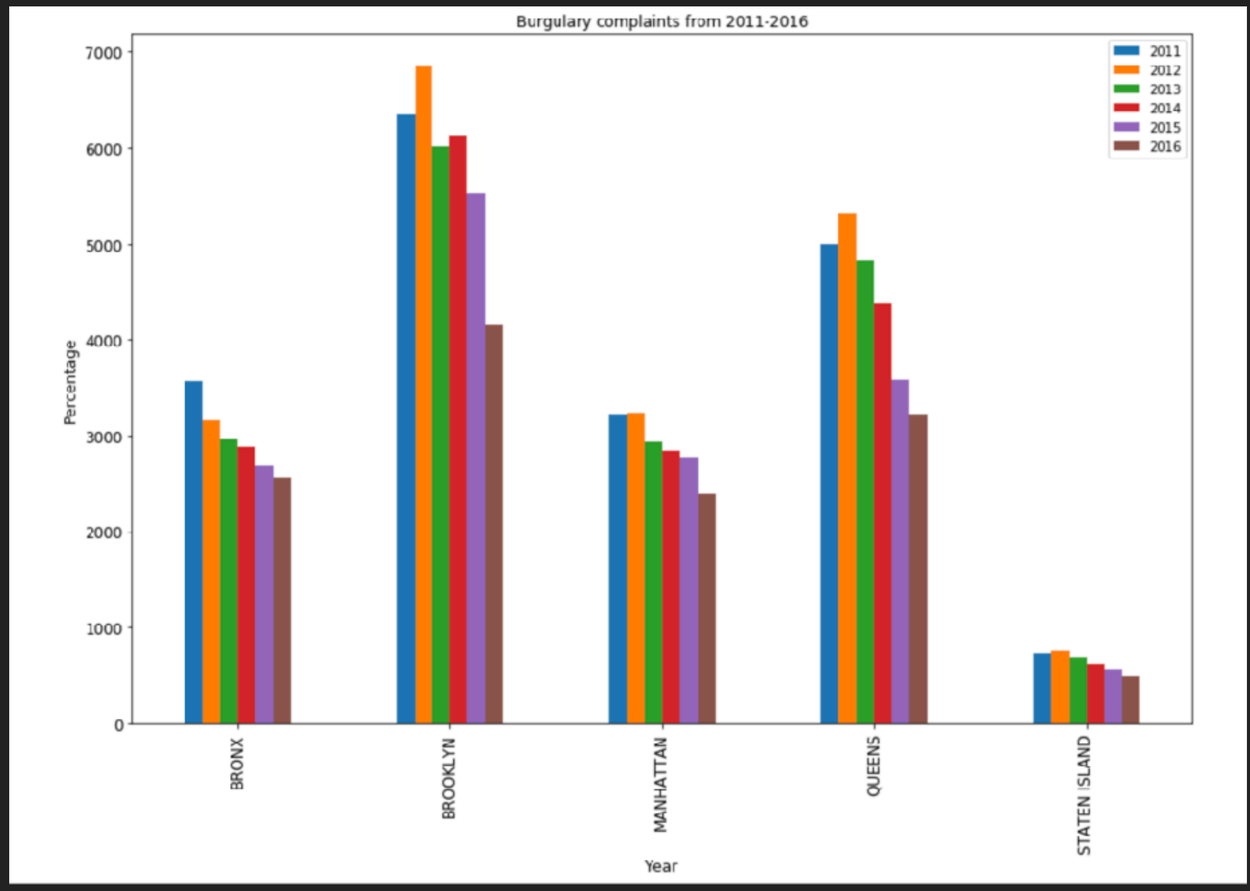


Number of Murder complaints

POVERTY AND BURGULARY CORRELATION ANALYSIS

2011-2016





CORRELATION VALUES

2011-2016

Crime Categories and Poverty - 2011-2016

Borough	Crime Categories		
	Assault	Burglary	Murder
Brooklyn	0.74	0.89	0.54
Bronx	0.24	0.64	0.56
Manhattan	0.30	0.77	0.76
Staten Island	0.37	0.22	0.63
Queens	0.79	0.94	0.78

CORRELATION VALUES

2011-2016

Crime Categories and Percentage of Population that did not go to College - 2011-2016

Borough	Crime Categories		
	Assault	Burglary	Murder
Brooklyn	0.73	0.70	0.69
Bronx	0.13	0.78	0.84
Manhattan	0.24	0.63	0.85
Staten Island	0.59	0.68	0.243
Queens	0.44	0.76	0.90

INTERESTING OBSERVATIONS

- Crime decreases with decrease in percentage of people officially regarded as poor.
- Crime decrease with Decrease in percentage of people who did not go to college.
- Burglaries increase with an increase in poverty

FUTURE WORK

- Investigate more correlations between more datasets.
- Range based correlations.
- Investigating correlations between sequences of different sizes.


ANY
QUESTIONS
?