

NYC Crime Analysis

Team Members

Joby Joy

Pravar Singh

Harshit Srivastava

Tools and Libraries:

- NYU HPC
- Pyspark
- Jupyter Notebook/Python
- GeoPandas
- Scikit Learn
- Statsmodels
- Matplotlib
- Numpy
- Pandas
- GitHub
- Microsoft Excel

Data Cleaning

Approach:

- Research on the data attributes and tabulated the expected results against each columns.
- Developed appropriate routines to validate against their respective column. Tabulated the number of invalid entries.
- Scrutinized the number of invalid and null entries and built logic to generate cleaned dataset.
- The cleaned data set was used in subsequent data analysis.

Data Cleaning Observations

Data Field	Valid	Invalid	Null
CMPLNT_NUM	5580036	0	0
CMPLNT_FR_DT	5111061	468319	656
CMPLNT_FR_TM	5579084	903	49
CMPLNT_TO_DT	3713616	393633	1472787
CMPLNT_TO_TM	4109777	1376	1468883
RPT_DT	5101229	478806	1
KY_CD	5580035		1
OFNS_DESC	5561144		18892
PD_CD	5575126		4910
PD_DESC	5575127		4909
CRM_ATPT_CPTD_CD	5580028		8
LAW_CAT_CD	5580035		1
JURIS_DESC	5580036		
BORO_NM	5579572		464
ADDR_PCT_CD	5579645	1	390
LOC_OF_OCCUR_DESC	4356430		1223606
PREM_TYP_DESC	5544838		35198
PARKS_NM	12539		5567497
HADEVELOPT	277818		5302218
X_COORD_CD	5384167		195869
Y_COORD_CD	5384167		195869
Latitude	5384167		195869
Longitude	5384167		195869

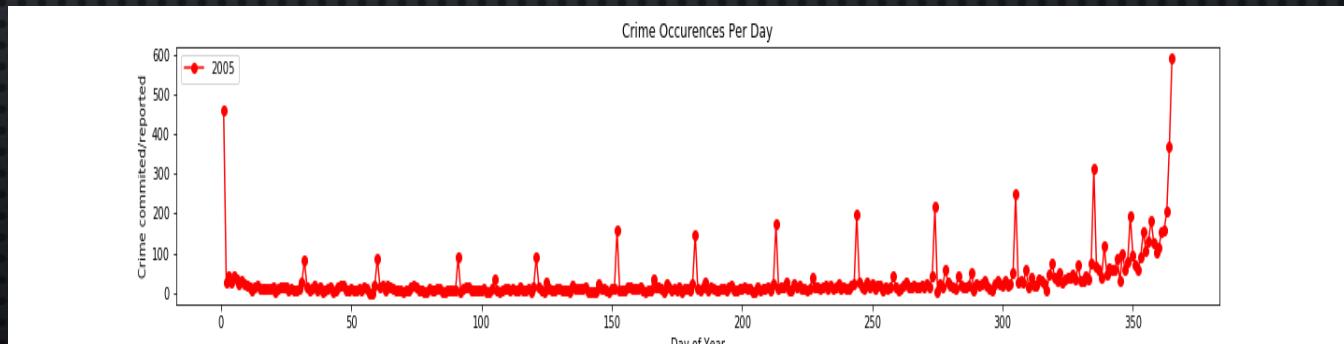
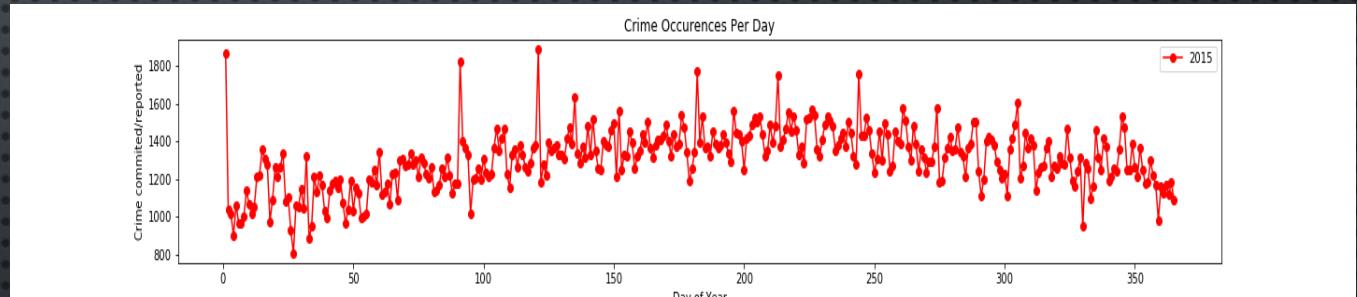
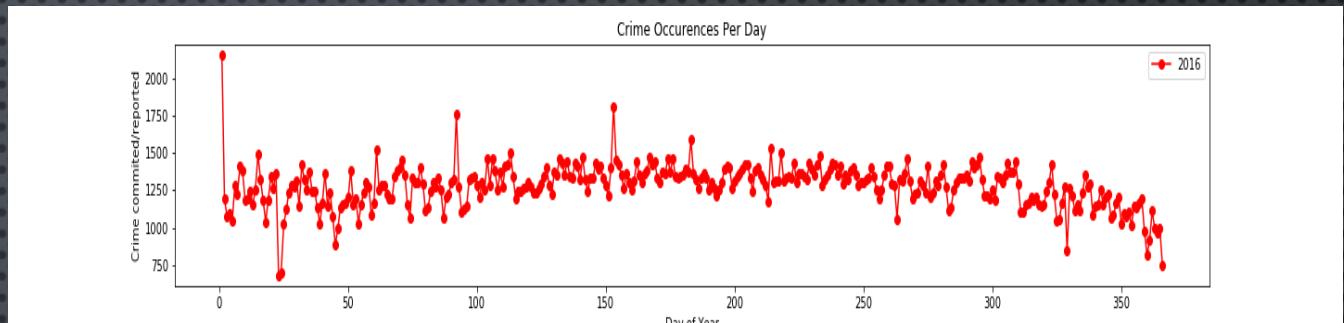
Data Analysis

Use Cases:

- Total Crime Occurrences across each year in New York City.
- Crime Occurrences across each year per Borough.
- Crime distribution percentage across each Borough every year.
- Crime Occurrences by the Nature of Crime.
- Crime Occurrences by precinct in each Borough.
- Average Incident count over the year.
- Correlation of Crime Occurrences with temperature.
- Time Series Forecast of Crime Occurrences over the year.
- Correlation of the Crime Patterns by Real Estate Demands in Brooklyn.

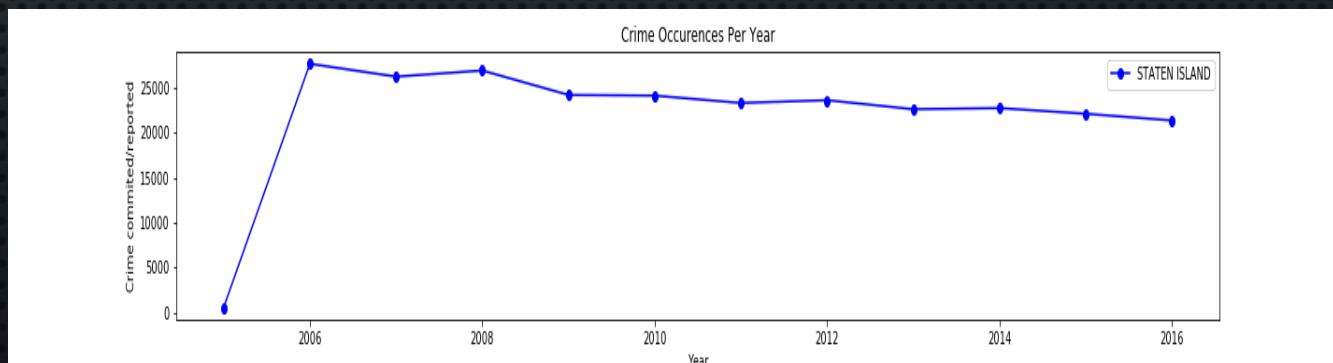
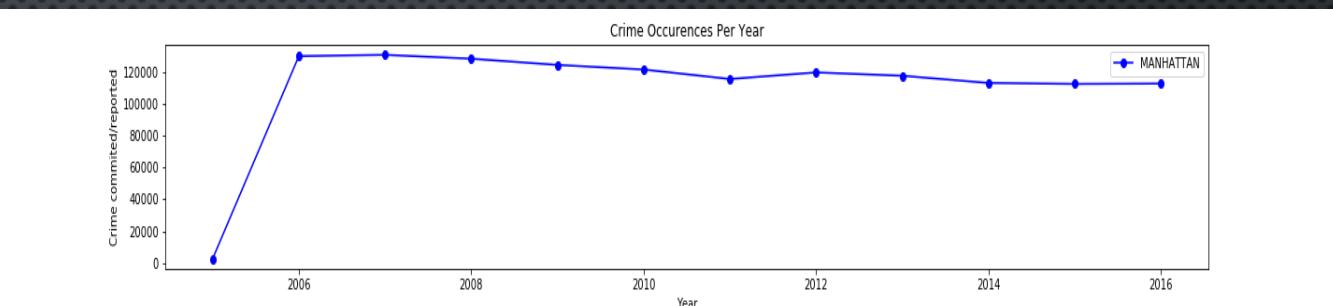
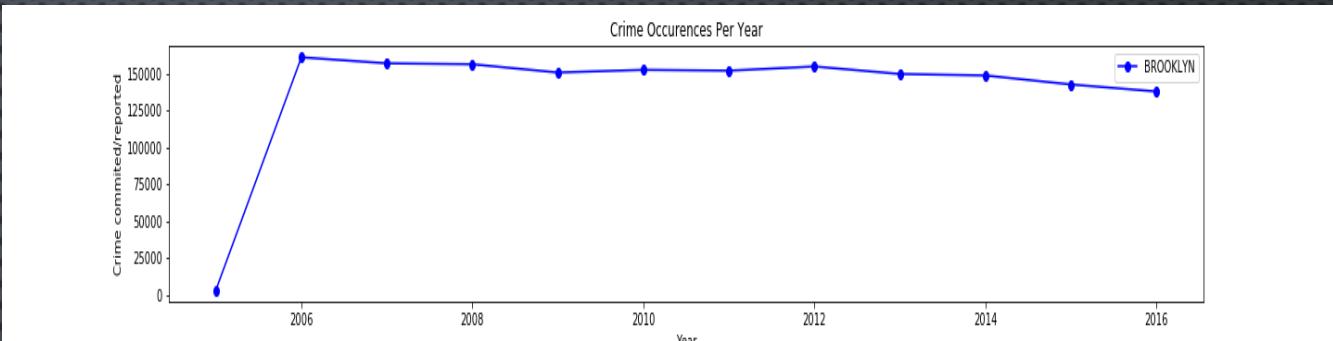
Total Crime Occurrences across each year in New York City.

- Crime Occurrence per day between 2005-2016.
- Crime distribution throughout the year was observed to be consistent between 2006-2016.
- Crime distribution steadily in the year 2005.

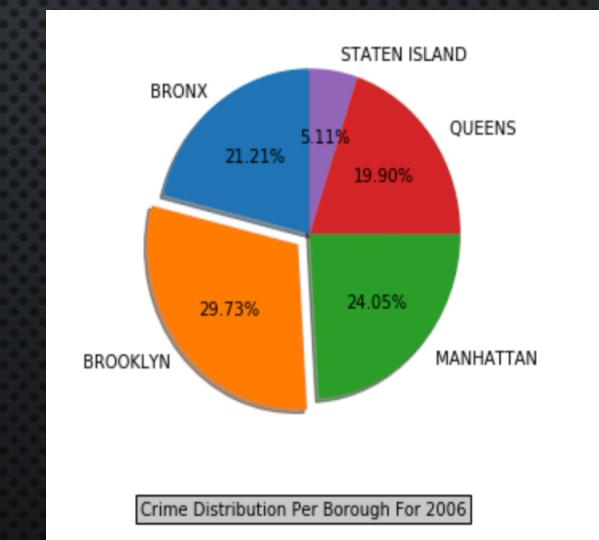
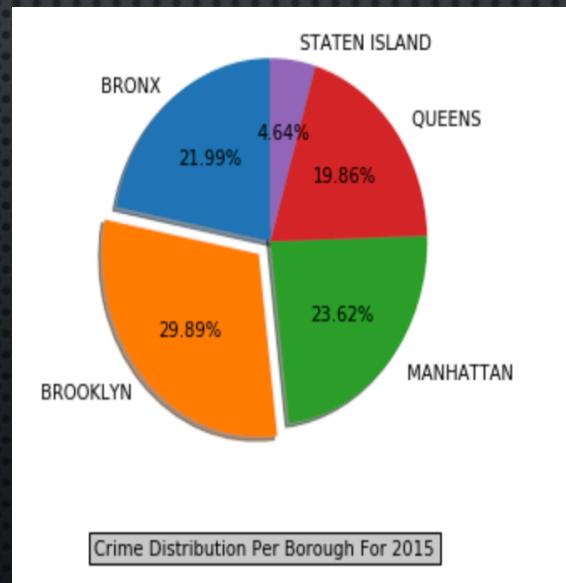
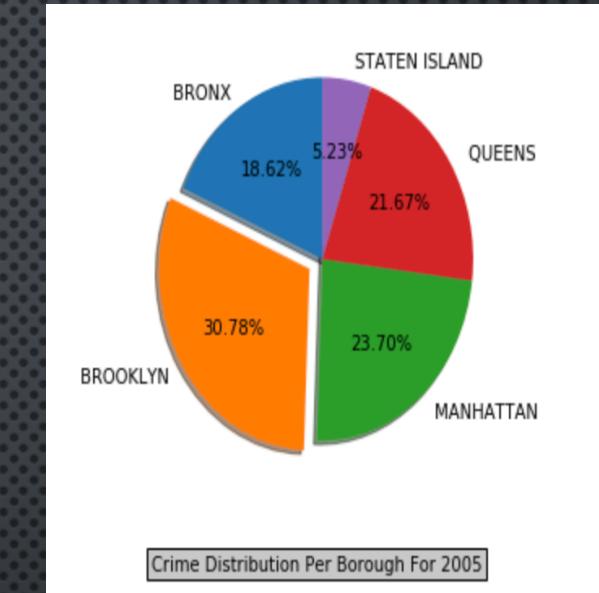
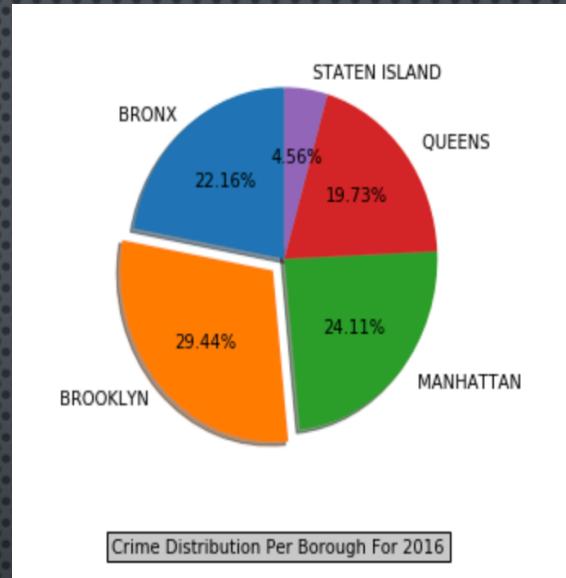


Crime Occurrences across each year per Borough

- Crime Distribution Per Borough between 2005 to 2016.
- Borough show a similar trend relating to the trend observed in the yearly crime count.
- Maximum crime count observed in Brooklyn.
- Staten Island has the lowest crime record.

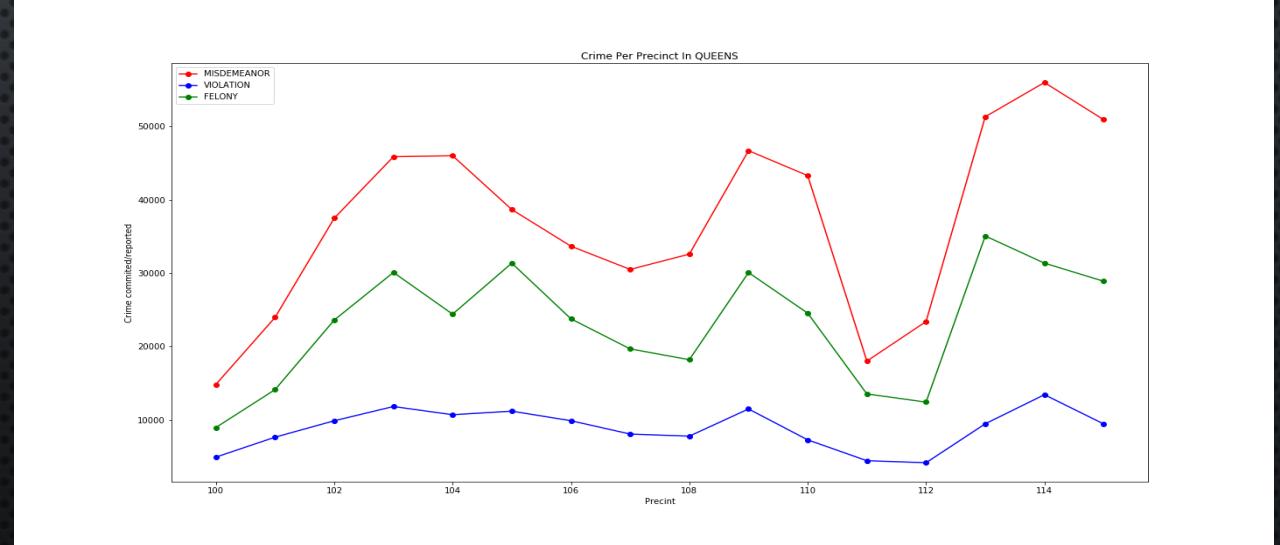
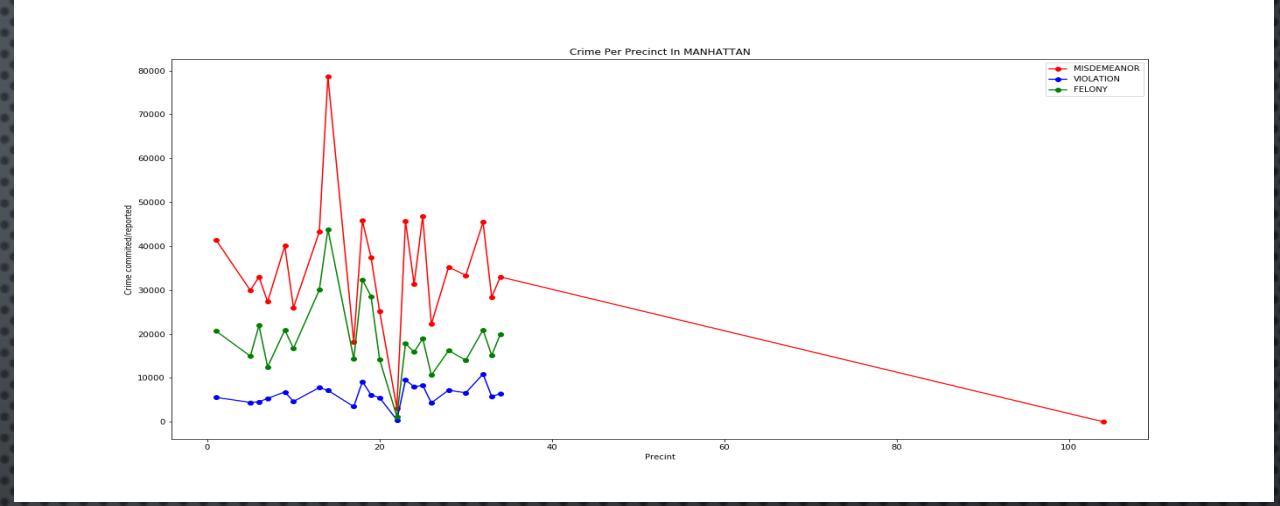


Crime distribution percentage across Borough by year



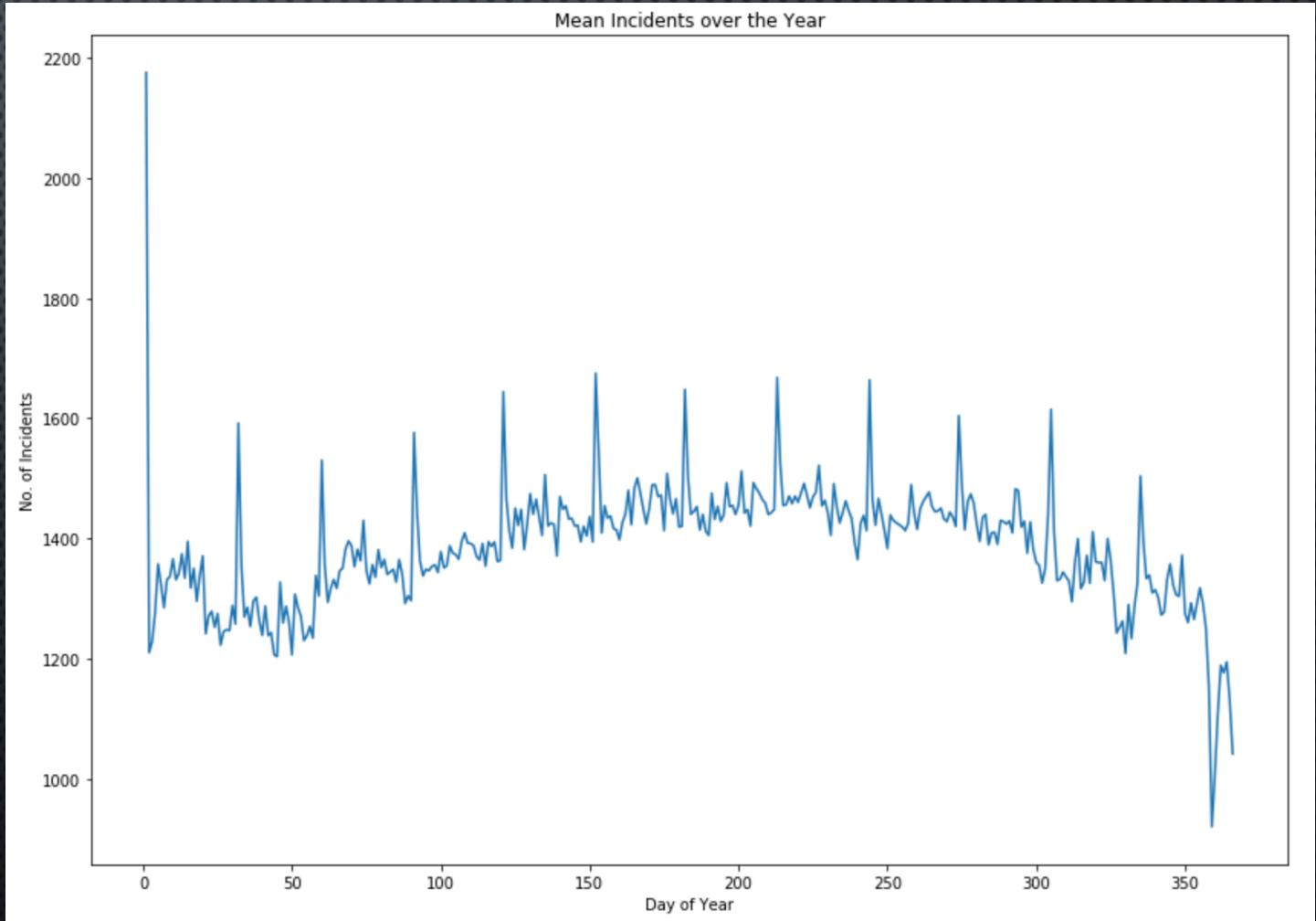
Crime Occurrences by Precinct and Nature in each Borough.

- Illustrates the crime distribution per Precinct segregated by the Nature of crime.
- Uneven Distribution in the Crime load being observed in Manhattan, Brooklyn.
- High Peaks observed in Precinct 15 -18 at Manhattan.
- Even Distribution observed in Queens.



Average Incidents over the year

- Crime Occurrence per day between 2005-2016.
- High number of Incidents on January 1st.
- Christmas time is safest in New York.



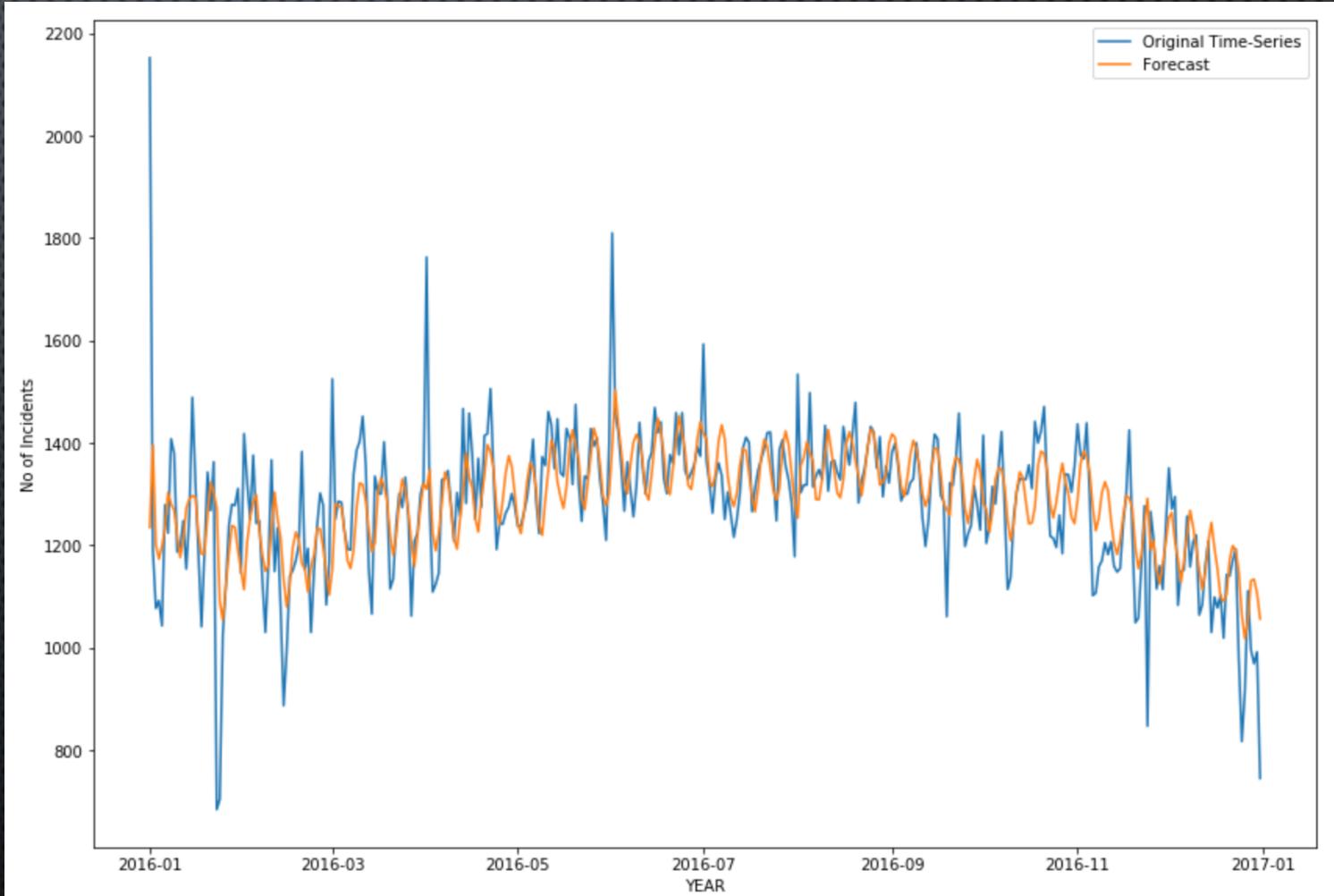
Crime Occurrences by Weather across Borough

- Weather dataset for 2016
- SpearmanR test for correlation between Crime Count and Temperature
- P-Value < 0.05 meaning Null Hypothesis can be rejected

Sample Size:	366
Correlation Coefficient:	0.608202764393
P-Value:	2.1195302026e-38

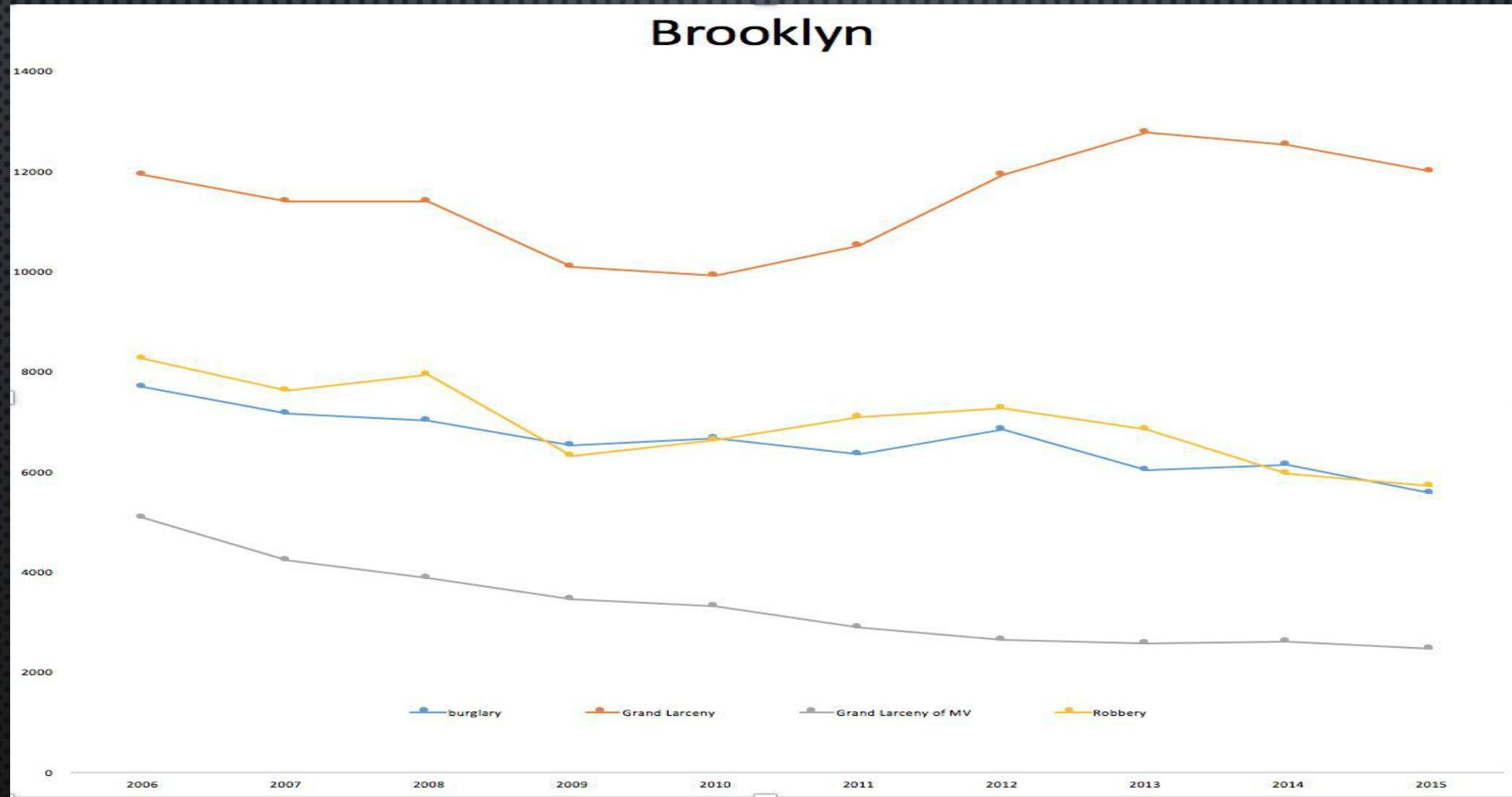
Time Series Forecast of Crime Occurrences

- ARMA model for prediction of Crime Occurrences for 2016
- AutoRegressive (AR) – 5
Moving Average (MA) - 3
- Mean Absolute Error (MAE):
69.1824291797
- Mean Forecast Error (MFE):
0.459864503948



Bonus Part: Exploring Correlation Between Crime Patterns and Real-Estate prices in Brooklyn

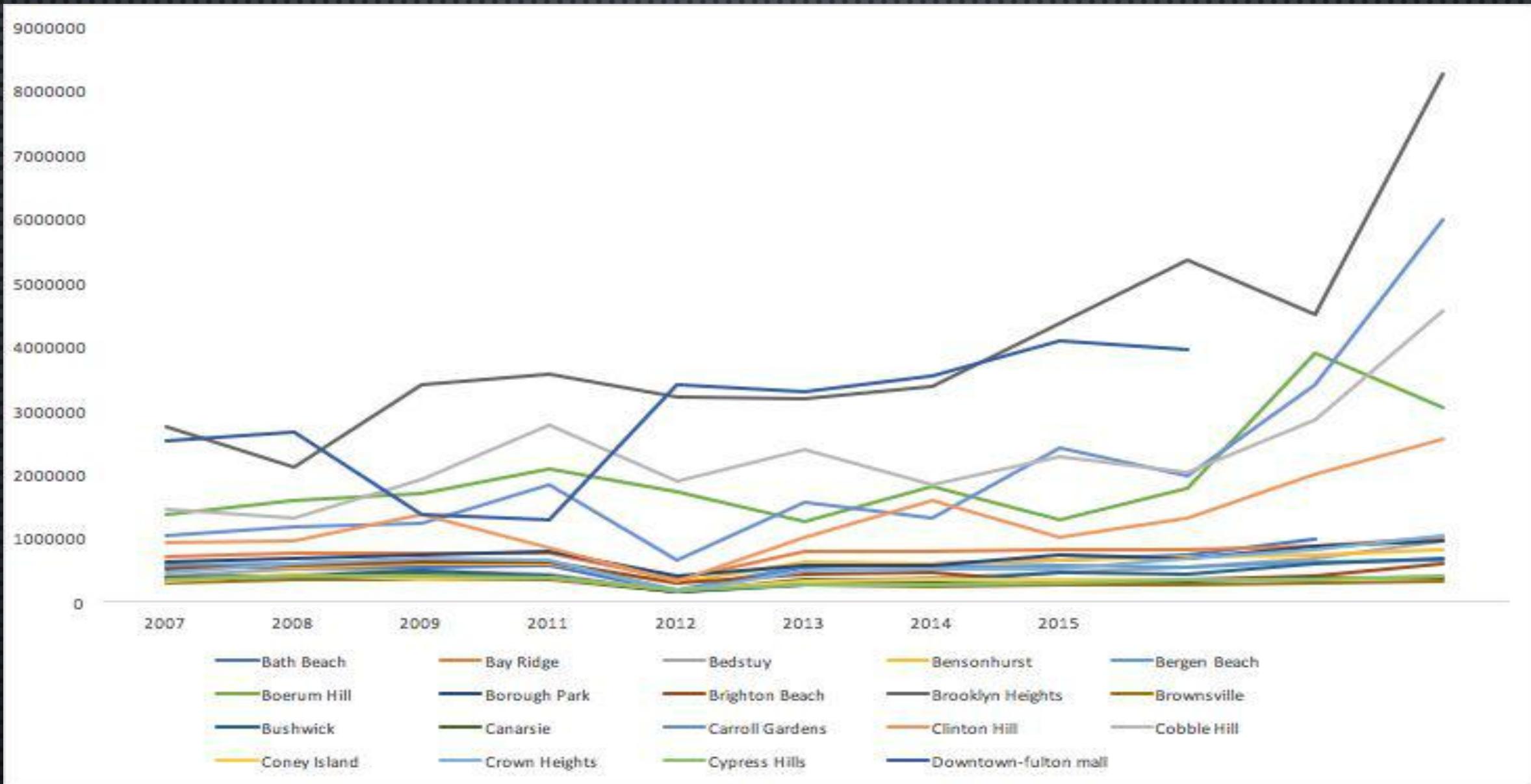
1. Number of Felonies in Brooklyn by Category



2: Real Estate Avg Sales Prices in Brooklyn per category



3: Real Estate Prices in Brooklyn by Neighborhood



Bedford-Stuyvesant - Average Price of all Real-Estate Sales and Count of Burglary, Robbery, Grand Larceny and Grand Larceny of Motor Vehicles

Bedford-Stuyvesant

10 Year Increase in Sale Cost: **61%**

10 Year Decrease in BRGL: **15%**

Year	Avg Sale Cost	Count of BRGL
2006	\$605,831.34	2,053
2007	\$599,073.63	2,133
2008	\$536,218.40	1,957
2009	\$414,480.32	1,781
2010	\$373,402.03	1,915
2011	\$419,655.64	1,978
2012	\$481,524.01	2,086
2013	\$626,258.19	1,812
2014	\$804,190.59	1,898
2015	\$978,138.33	1,750

Thank You