Data Bootcamp Project

Analysis of NYC Crime Data

Our project is an analysis of data from the New York Police Department Complaint history, which is a data set that details all complaints filed with the NYPD from 2006 to 2017. Although New York City has a very high number of crimes occurring per year, the “homicide number is the lowest in the modern era of NYPD Compstat record keeping, which began in 1994”. The spreadsheet organizes the data into a set of several different categories, from the type of crime to the borough a crime was committed in and what the demographics of the suspect and victim were.

We chose to look for trends in these statistics to get a better understand of the causation of these crimes. We compared the crime data to common associated factors, such as crime prevalence per borough in relation to age of the suspect, race of the suspect, as well as the distribution of crimes in relation to what borough the crimes were committed in. The analysis of this data will provide a better understanding of the distribution of crime in New York City over the last decade. This data can be further utilized by the NYPD themselves to help create tactics to combat these crimes in the coming years, for example, for our first analysis, which goes into detail about the crimes committed on Christmas Day, the NYPD can look at our data and chose to make strategic decisions such as increasing NYPD staff in heavily populated shopping areas to help catch and combat shoplifting crimes.

The data can also provide a profile of what the demographics of the average suspect are like, so the officers on patrol have a general idea of what type of criminal they should be looking for for each type of crime. For example, there may be a peak in white males under 44 who are shoplifters as opposed to hispanic women in their 60’s. This data provides the perfect profile for these criminals which will hopefully be utilized as a tool to watch for crime and not a bias which harms those who fall under the same profile but are not guilty of a crime. By looking at external data on crime and as well as borough statistics, we have been able to further search for trends in the data that may correspond with the outside statistics.

We expect that in all cases, the more violent crime will be committed in the outer boroughs as opposed to Manhattan. We expect the most number of crimes to occur in Manhattan due to the foot traffic, but it is expected that they will be mostly petty crimes. We have not made a prediction on the average profile of the suspects per each borough, but can strongly determine it will most likely be males in the youngest age group of 25-44.

Description of the Data

We used the NYPD Complaint Data Historic dataset to analyze the crimes. This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department (NYPD) from 2006 to the end of last year (2017).

The data has 6 million rows and 35 columns. Each row details the crime reported in order of date and time. Columns include the name of the borough in which the incident occurred, exact date of occurrence for the reported event, three digit offense classification code, suspect and victim information, etc.

The data can be found **here:**
https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i/data

Sorting of the Data

For each of our findings, our data was cleaned and sorted in different ways that worked best for the type of data we were using. Our data was cleaned of any incomplete reports, such as reports that had NaN's or unknowns in columns which that information was pertinent to our findings. The omission of uncomplete reports may lead to a skew in the data from unknown factors, but due to the sheer number of data rows we have, we believe if there is any skew from the missing data removal, it will be negligible.

We sorted the data through each finding primarily based off borough, as different boroughs inhabit different types of people from different socioeconomic and racial backgrounds. Sorting the data by borough helps provide a better understanding of what crimes are most frequent in each area, which will display trends in the data in relation to the statistics that each specific borough holds.

Christmas Crime Suspect Demographic by Borough

https://github.com/ys07/Data_Bootcamp/blob/master/Christmas%20Crime_FINAL.ipynb

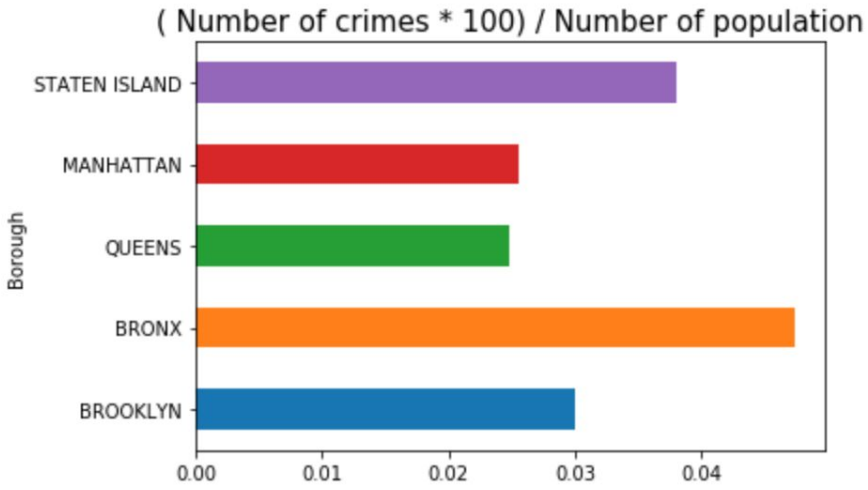
For this analysis, we chose to look at the data and create profiles of suspects who committed crimes on Christmas Day since 2006. Christmas is an incredibly high traffic time in New York City, and crime is prevalent throughout the boroughs. We first sorted through the data set and pinpointed all of the crimes occurring on 12/25/200X and 12/25/201X. After retrieving the Christmas specific crime data, we further shortened the data by selected the columns that were needed: borough name, suspect age, and suspect race.

The data was then organized by borough, with our data showing that the number of crimes occured during Christmas Day in Brooklyn was the highest, followed by the Bronx, Queens, Manhattan, with Staten Island having the lowest. The boroughs with the highest and the lowest reported crimes was in line with the overall dataset of reported crimes over the 10 year period, but other boroughs varied in order from the overall dataset. According to NYPD's statistics from 2006 to 2017, the order from most to least reported crimes by borough was Brooklyn - Manhattan - Queens - Bronx - Staten Island.

Based on the data, when compared to the Queens and the Bronx, Manhattan has a lower crime rate on Christmas than typically reported over the period under analysis. A possible explanation for this are the additional foot patrol officers in Manhattan to protect the increase in tourists in Manhattan on Christmas. According to "Broken Windows", 1982, by James Wilson, the presence of additional foot-patrol officers elevate the level of public order. Wilson explains that disorder and crime are usually inextricably linked. In his thesis, Wilson notes that social psychologists and police officers tend to agree that if a window in a building is broken and is left unrepaired, the remaining windows will soon be broken creating an area in disorder. When applied to criminal activity, an area in disorder is more vulnerable to criminal invasion. Therefore, an area with increased patrol officers will be less susceptible to disorder and, consequently, less criminal activity.

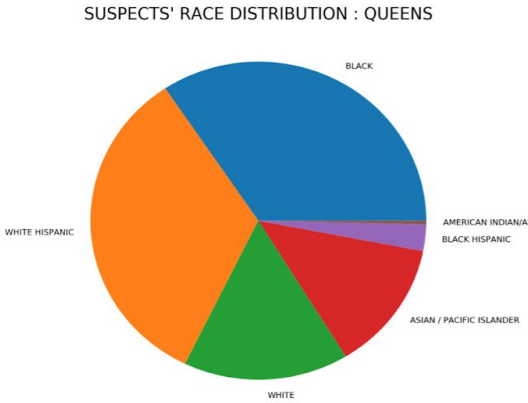
To determine whether the criminal activity was directly related to the population changes of each borough, we considered the population differences between the boroughs.

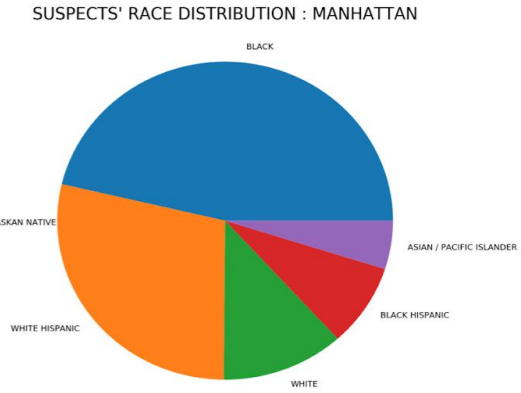
We used New York City Boroughs Population data from **US CENSUS Bureau**. Since the percentage of population increase in each borough was similar, we used 2010 data in the analysis. With this data we created a new column which shows (number of crimes * 100) / population per borough.

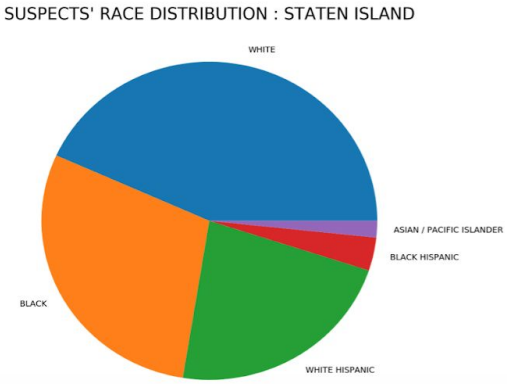


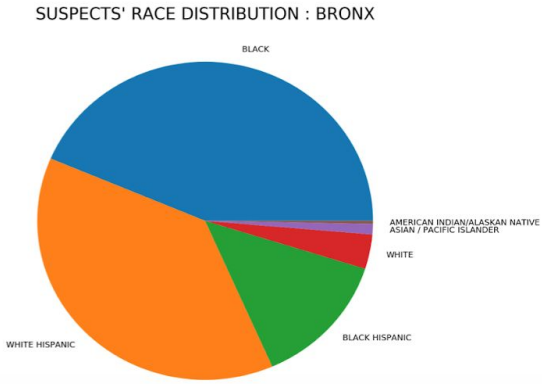
Considering each borough's population, on Christmas Day, the Bronx has the highest crime rate and Queens has the lowest rate. Staten Island has a high crime rate due to the comparatively less number of police and a bigger area to patrol.

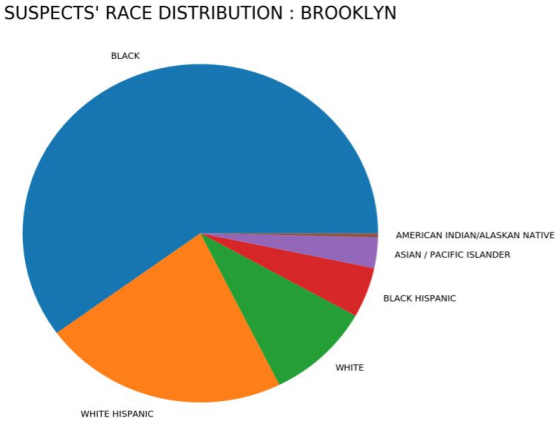
We then went on to analyze the suspect's race distribution by borough. The borough was set as the index and a series of pie charts were made to display the race prevalence between the five different boroughs.





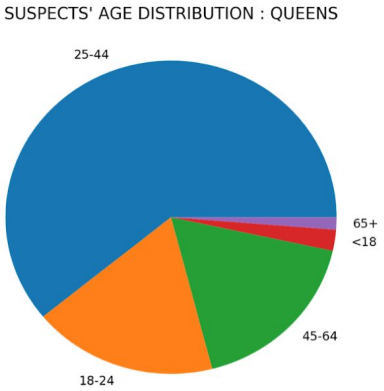


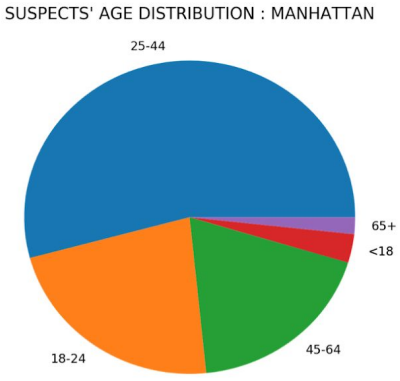


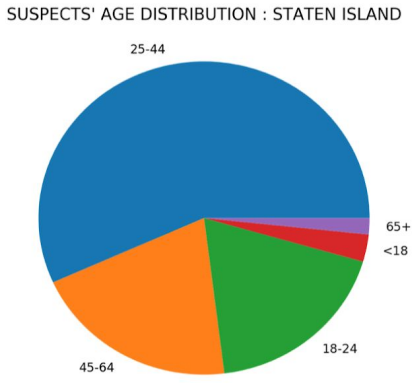


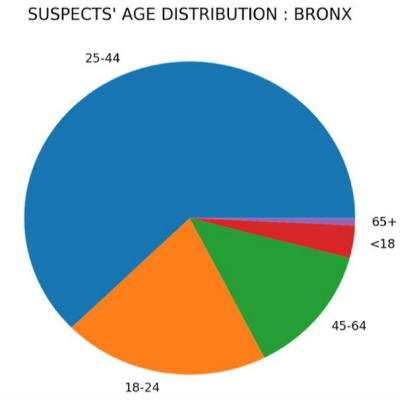
The graphs above show varying results. Aside from Staten Island, in all other boroughs Blacks had the highest crime rate with White Hispanics following in second. In Staten Island, Whites had the highest crime rate most likely because Staten Island's racial distribution is concentrated on the White racial grouping. According to Stats And The City, the two largest racial groups in Staten Island are White (72.8%) and Black (10.6%), Queens are White (39.7%) and Black (19.1%), and Brooklyn are White (42.8%) and Black (34.3%).

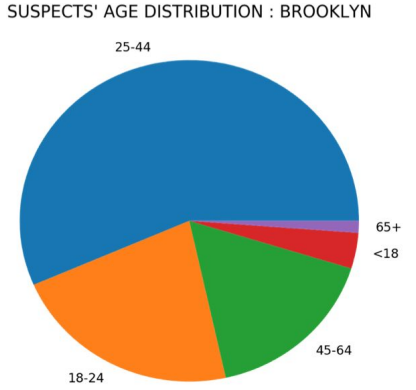
Our final analysis goes on to plot the suspect's age distribution by borough.











In all boroughs, the age group which committed the most amount of criminal activity was the 25-44 year old bracket. Aside from Staten Island, the age group with the second highest crime rate was the 18-24 year old grouping. In Staten Island, the age group which committed the second highest crime rate was surprisingly the 45-64 year old bracket.

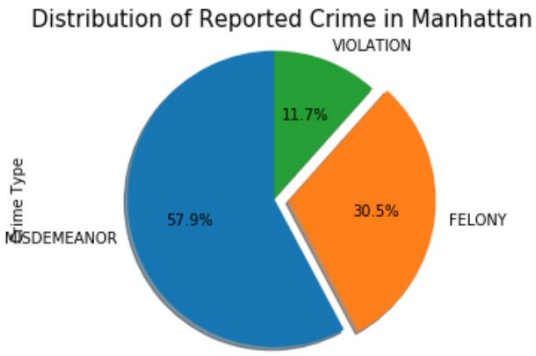
Felony Crimes Reported per Borough and Unemployment

https://github.com/nuomin1/Data-Bootcamp/blob/master/Project%20Final.ipynb

Continuing the use of the NYPD Complaint Data Historic, we analyzed the distribution of the classifications of crime reported in each borough and its correlation to unemployment. The data of analysis is specific to 2017, of which there are complete and revised data for crimes reported and unemployment.

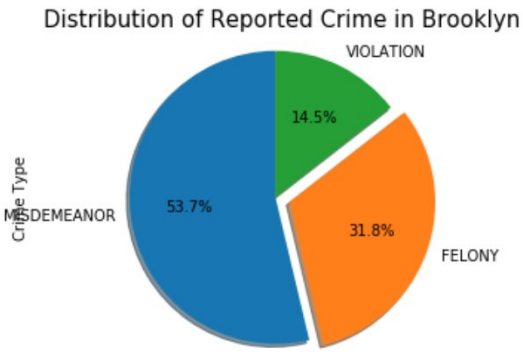
For formatting the data, we selected the column that details the type of crime in relation to the rows of crime reports listed. To ensure that the crime classification is complete for all police reports, we will check for unknown values and remove them. Once the unknown values were sorted out, we proceeded to classify the crimes based on employment in the borough.

Manhattan



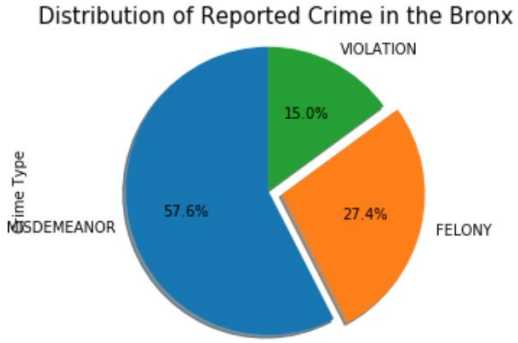
In 2017, Manhattan had approximately 111,765 crimes reported, 30.5% were felony charges, but a greater proportion of misdemeanors.

Brooklyn



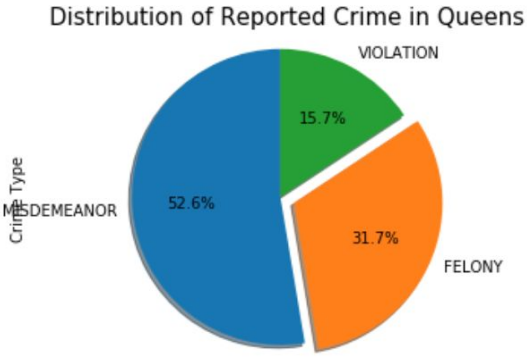
In 2017, Brooklyn had approximately 134,683 crimes reported, 31.8% were felony charges, but a greater proportion of misdemeanors. This is only slightly higher than Manhattan; Brooklyn has less misdemeanors and more violation crimes as well.

Bronx



In 2017, the Bronx had approximately 101,301 crimes reported, 27.4% were felony charges, lowest of the boroughs examined so far. Furthermore, the Bronx has a small percentage of violation crimes reported.

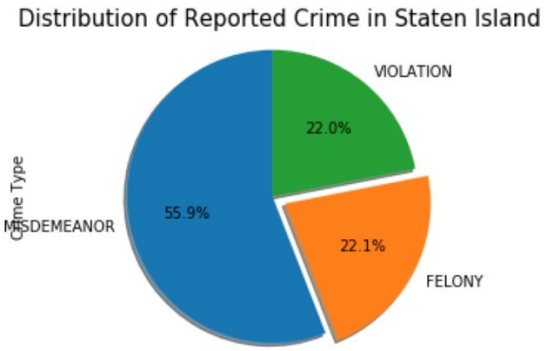
Queens



In 2017, Queens had approximately 89,361 crimes reported, 21.7% were felony charges, similar to Brooklyn. Queens also has the lowest misdemeanor crimes reported of the boroughs so far.

Staten Island

In 2017, Staten Island had 20,841 crimes reported, 22.1% of which were felony charges; lowest number of crimes and % of felony charges across all boroughs. This was the highest amount violation crimes reported.



To continue our analysis on the correlation of unemployment and felony crimes reported, we used data from the Labor Statistics for New York City. We related the unemployment rate to the felony crime rate via a statistics chart.

	Felony Crime Reported (%)	Unemployment Rate
Borough		
Kings County	31.8	4.6
Bronx County	27.4	6.2
New York County	30.5	4.0
Queens County	31.7	4.0
Richmond County	22.1	4.6

From the comparison of the tables, we see that while Brooklyn (Kings County) had a higher % of felony crime reported in 2017 than the Bronx, it still had a lower unemployment rate. Staten Island, alternatively, had the lowest % of felony crimes reported in 2017 yet still had a higher unemployment rate compared to Queens and Manhattan. For further insight, we will examine data from the 2013 American Community Survey on Median Household Income ($) throughout the boroughs in relation to the two factors.

We also related the median household income to the unemployment rate and felony crime rate.

	Median Household Income ($)	Unemployment Rate
Borough		
Kings County	46085	4.6
Bronx County	34388	6.2
New York County	69659	4.0
Queens County	54373	4.0
Richmond County	72569,	4.6

An observation that we have found including median household income into our analysis is that despite having a lower % of felony crimes reported, the Bronx has the highest unemployment rate and the lowest median household income. One plausible explanation is that the residents do not cooperate and follow-up with the police due to distrust, as previous persons who have committed felonies cannot find good work, as seen by median household income and the high unemployment rate. Meanwhile, Staten Island's low % of felony crimes reported may be attributed to a combination of its smaller population, and the high median household income, in which most residents are working high-earning jobs in NYC. Furthermore, higher % of crime in New York and Queens can be attributed to the large population, which is supported by the fact that the unemployment rates are low while median household income are still high although more felony crimes are being reported.

Violent Felony Offense vs. Property Crime Felony Offense by Borough

https://github.com/nuomin1/Data-Bootcamp/blob/master/analysis%203.ipynb

We used the data from the second analysis, as well as the 2017 New York County Index Crime Counts and Rates to help implement our analysis on the breaking of felony crimes into violent and property crimes.

	Boroughs	Violent Crimes	property crimes
0	Bronx	12,534	22,973
1	Brooklyn	14,593	35,173
2	Queens	8,671	38,475
3	New York	9,184	23,676
4	Staten Island	1,451	4,518

We can see from the table that in felony cases, there are significantly more property crimes than violent crimes for all of the counties. Merging the tables will allow us to compare the nature of felony charges with the number of reported felony crimes.

	Felony Crimes Reported	Violent Crimes	property crimes
Boroughs			
Brooklyn	42814	14,593	35,173
Bronx	43814	12,534	22,973
Queens	28320	8,671	38,475
Staten Island	4607	1,451	4,518

The numbers from the NYPD Complaint Data is more or less consistent with the index crimes count. One reason for the discrepancy is that multiple crimes could have occurred on a police complaint. However, there is a significant difference in the number of property crimes and violent crimes in which both are considered index crimes, with possible long-term sentencing demonstrates the carceral state of NYC.

Interpretation of Results

The outcome of the data analysis for each experiment run provided useful information on the crime statistics in NYC. In our findings of the Christmas Crime Suspect Demographic by Borough, it was evident that for four out the five boroughs Blacks and White-Hispanics the age of 25-44 seemed to be the most frequent suspects seen throughout the boroughs, with slight deviations for Staten Island. This maybe due to the demographic makeup of Staten Island, where 75% of people are White Non-Hispanic (World Population Review), therefore the there is a larger prevalence of white suspects in Staten Island. Surprisingly, the distribution of suspects by age and race remained fairly consistent, regardless of borough. The causation of this may be due to the fact that people from other boroughs commit crimes in boroughs they do not reside in, which led to a fairly even distribution.

With our data from the Felony Crimes Reported per Borough and Unemployment, the data shows the distribution of type of crime reported in each borough is also fairly similar to each other, with Staten Island and the Bronx having the highest number of misdemeanors. In relation to unemployment rate, we did not find a strong correlation between our unemployment statistics and felony crimes and our household median income statistics and felony crime rates. In our analysis of the statistics provided by the 2017 New York County Index Crime Counts and Rates, we also found that the numbers from the NYPD Complaint Data is more or less consistent from the index crimes count.

Final Statement

In conclusion, it is clear that our analysis provides a more segmented look into crime in New York City. By utilizing the categorical data to determine trends in the crime reports over the past decade, we were able to find steady and strong correlations in crime throughout all five boroughs. Although outside factors play a factor on the accuracy of this data, overall our data is consistent with outside reports created on crime statistics in New York City.

Bibliography

- Penn Wharton. “Criminal Records and Unemployment: The Impact on the Economy.” *Penn Wharton Public Policy Initiative*, Penn Wharton Public Policy Initiative, 20 Aug. 2017,
- publicpolicy.wharton.upenn.edu/live/news/2071-criminal-records-and-unemployment-the-impact-on.
- Stats and the City“Racial Breakdown by Borough Stats and the City.” *Stats and The City*, Crain's New York Business,
- mycrains.crainsnewyork.com/stats-and-the-city/2011/demographics/racial-breakdown-by-borough.
- Wilson, James Q., and George L. Kelling. 1982. "Broken Windows.” The Atlantic Monthly, March.

Data Sets

- https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i/data
- https://www.labor.ny.gov/stats/nyc/Revised%202017%20Borough%20Labor%20Force.xls
- http://www.criminaljustice.ny.gov/crimnet/ojsa/indexcrimes/2017-county-index-rates.pdf
- https://projects.newsday.com/databases/long-island/new-york-city-crime-rate/

FINAL PROJECT GITHUB *https://github.com/mariayannaarundel/DataBootcamp*