NYPD Shooting Incident Data

Name: Pranjal Anand Singh

Introduction

For this project, we will be analyzing the "NYPD Shooting Incident Data (Historic)" dataset available on the NYC Open Data portal *(https://data.cityofnewyork.us/Public-Safety/NYPD-Shooting-Incident-Data-Historic-/833y-fsy8)*. This dataset contains information on reported shooting incidents in New York City from 2006 to 2021. It is a structured dataset with a defined schema and data types, consisting of rows and columns.

The dataset includes information on various variables such as the location, date, time, victim and perpetrator demographics, and other relevant information. With over 34,000 rows of data, this dataset offers valuable insights into the trends and patterns of gun violence in the city.

In real life, this dataset can be used by law enforcement agencies, policymakers, and researchers to better understand the frequency, location, and circumstances surrounding shooting incidents in New York City. By analyzing this data, policymakers can make informed decisions on how to reduce gun violence in the city.

The dataset can be improved by providing more detailed information on the context and circumstances surrounding each shooting incident. This would enable researchers to identify the root causes of gun violence and develop targeted interventions to prevent it. Additionally, the dataset could be updated in real-time, allowing law enforcement agencies to respond quickly to incidents and improve public safety.

Preprocessing

Before we can perform analysis, we applied some preprocessing steps to prepare the data for visualization and analysis. Here are some of them:

- <u>Dropping unnecessary columns</u>: Some columns in the original data frame were not needed for the analysis, so they were dropped using the drop method.
- <u>Handling missing data:</u> In some cases, missing data were dropped using the dropna method, or filled using the fillna method. For example, in the code for generating the heat map, rows with missing latitude and longitude values were dropped using dropna(subset=['Latitude', 'Longitude']).
- <u>Converting data types:</u> In some cases, columns were converted to a different data type. For example, the 'OCCUR_TIME' column was converted to a datetime object using the pd.to_datetime method.
- <u>Aggregation and grouping:</u> Some analysis required grouping and aggregation of data using the groupby method. For example, the code for generating the bar chart of shooting incidents by hour of the day grouped the data by the 'HOUR' column and counted the number of incidents in each group using the size method.Overall, the preprocessing steps taken above were aimed at preparing the data for visualization and

analysis by ensuring that it is clean, complete, and in a format that is suitable for the intended analysis.
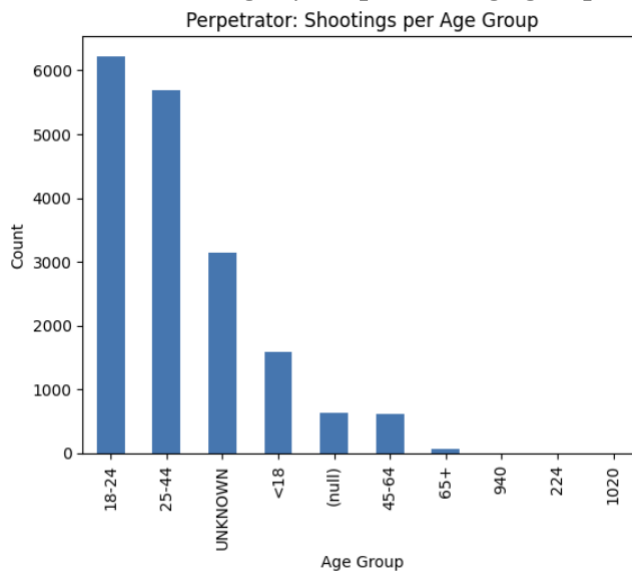
Analysis

The various analyses performed are:

- Analysis of the number of shootings by borough: This analysis uses the 'BORO' field of the data. The result is a bar chart that shows the number of shootings by borough.
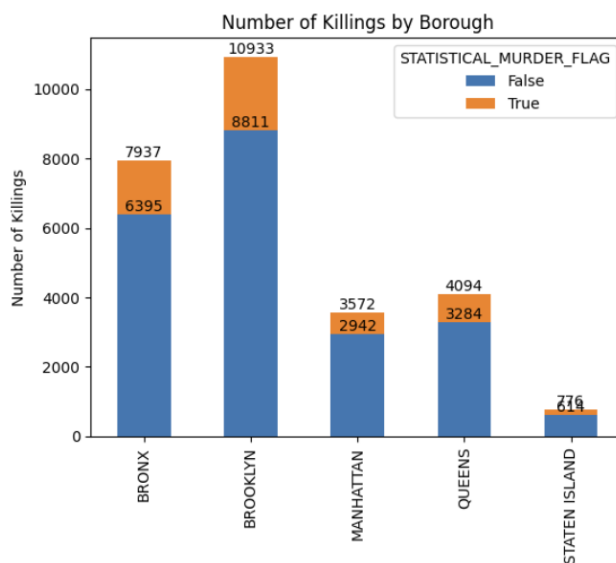

Shootings per Borough
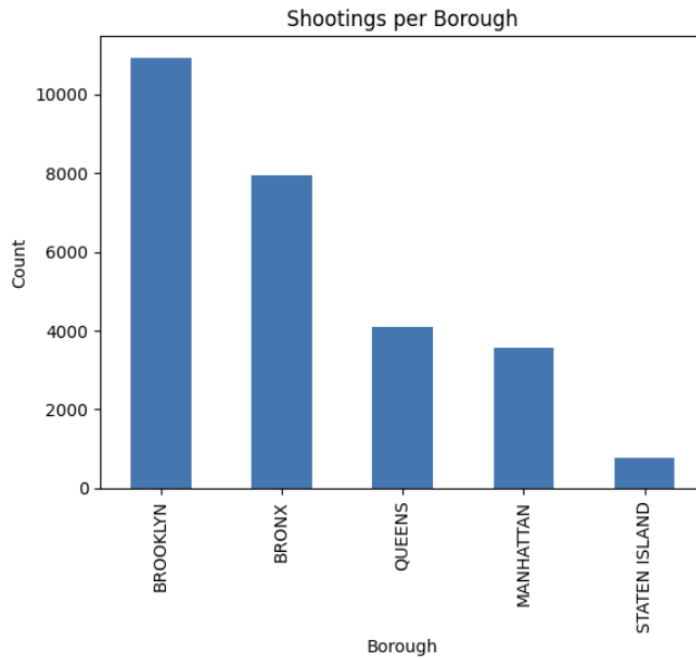
- Analysis of the distribution of shooting incidents by Perpetrator age group: This analysis uses the PERP_AGE_GROUP' field of the data. The result is a bar chart that shows the

number of shootings by Perpetrator age group.

- Analysis of the distribution of shooting incidents by Borough and whether murder was committed: This analysis uses the 'BORO' and 'STATISTICAL_MURDER_FLAG' field of the data. The result is a bar chart that shows the breakdown of shootings by Borough and whether there was a murder.


- Analysis of the distribution of shooting incidents by Perpetrator and VictimeGender: This analysis uses the 'VIC_SEX' and 'PERP_SEX' field of the data. The result is a stacked bar chart that shows the distribution of shooting incidents Victim and Perpetrator Gender.
The chart shows that the majority of victims were male, with male perpetrators being more common than female perpetrators.
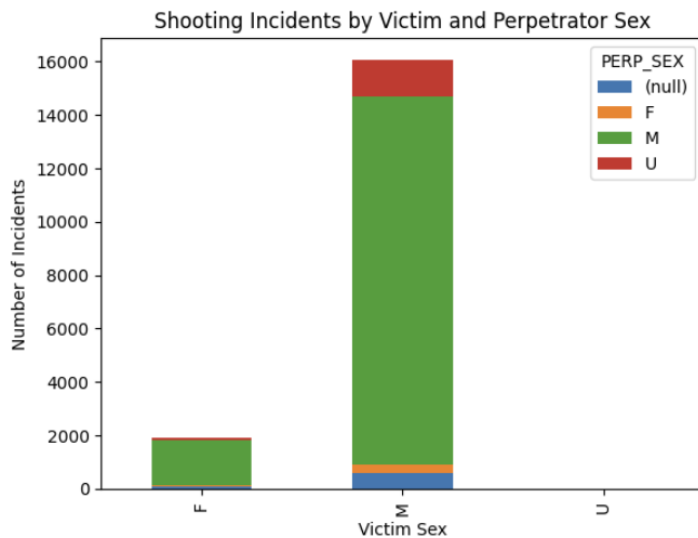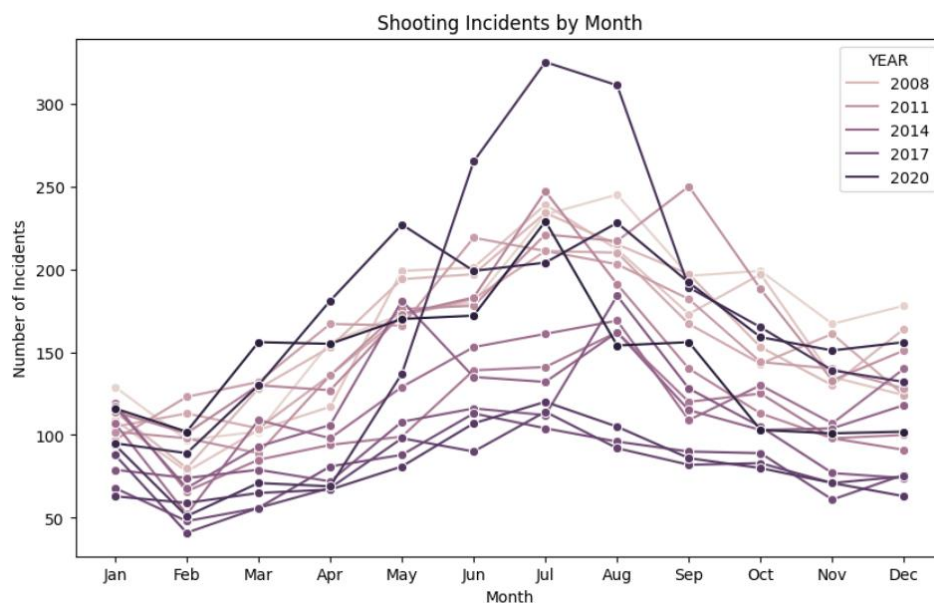
- Analysis of the distribution of shooting incidents by victim race: This analysis uses the 'VIC_RACE' field of the data. The result is a pie chart that shows the distribution of shooting incidents by victim race.
  The chart shows that the majority of victims were Black or Hispanic, with a smaller percentage of victims being White or Asian.
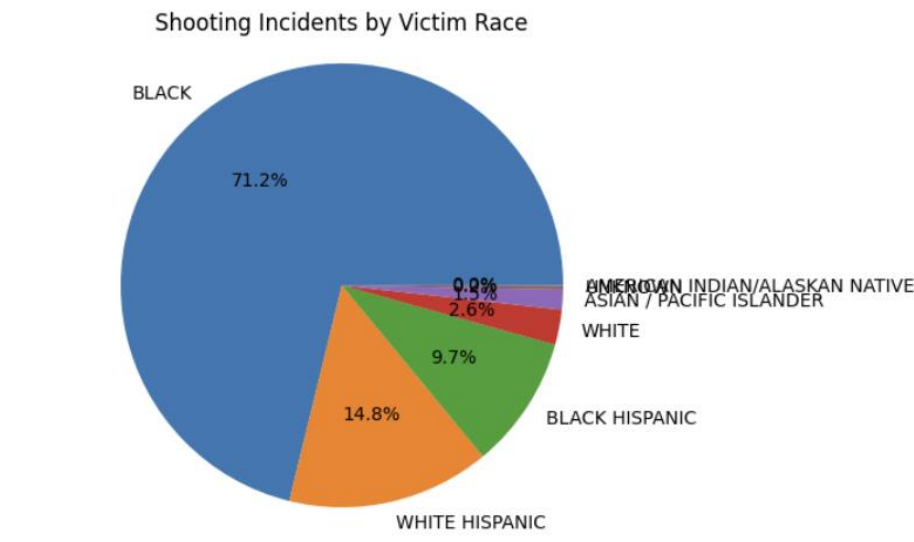
- Analysis of the shooting incidents by hour of the day: This analysis uses the 'OCCUR_TIME' field of the data. The 'OCCUR_TIME' field is converted to datetime format, and the 'HOUR' field is extracted from it. The result is a bar chart that shows the number of shooting incidents by hour of the day.
  The chart indicates that the number of incidents is highest between 8pm and early morning hours.

Shooting Incidents by Hour of the Day
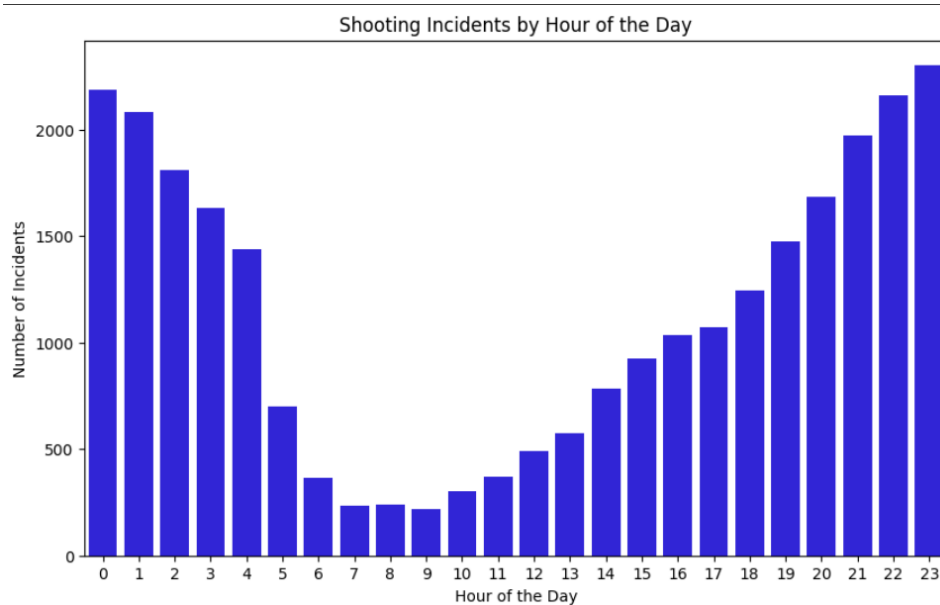
- Analysis of the shooting incidents by month: This analysis uses the 'OCCUR_DATE' field of the data. The 'OCCUR_DATE' field is converted to datetime format, and the 'YEAR' and 'MONTH' fields are extracted from it. The result is a line plot that shows the number of shooting incidents by month and year.
  The chart shows that shooting incidents tend to increase in the summer months and decrease in the winter months.
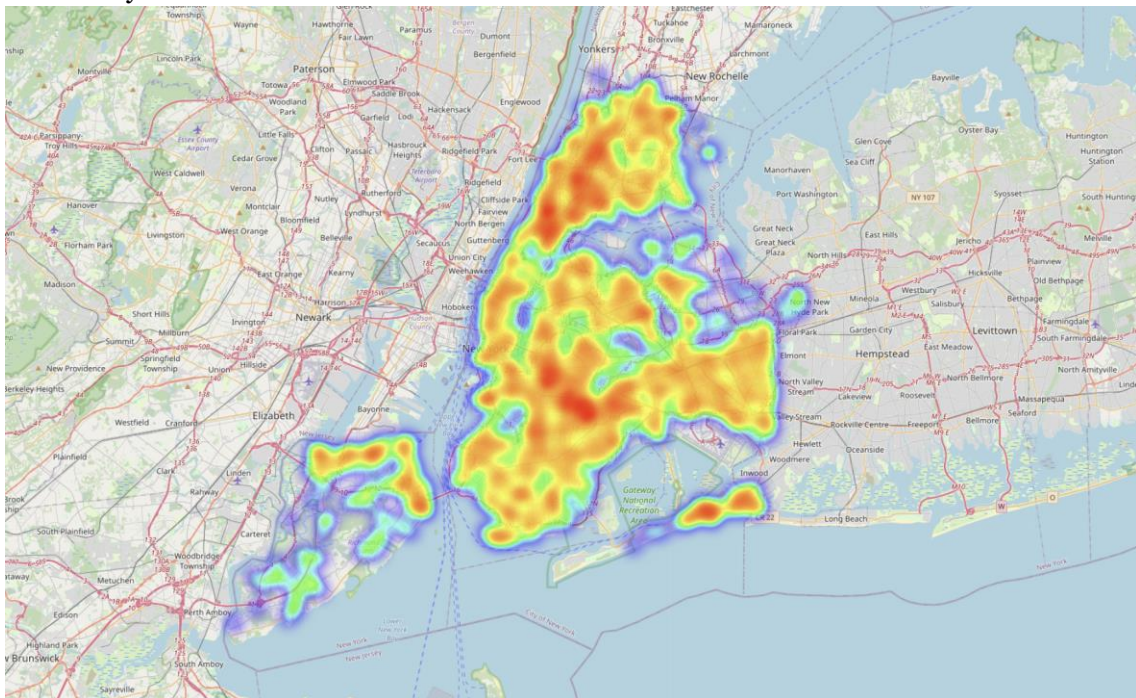

Shooting Incidents by Month

- Analysis of the locations of shooting incidents: This analysis uses the 'BORO' and 'LOC_CLASSFCTN_DESC' fields of the data. The result is a heatmap that shows the locations of shooting incidents by borough and location classification.

The heatmap shows that the highest number of shooting incidents occurred in public housing and on the street, with most incidents occurring in Brooklyn and the Bronx.
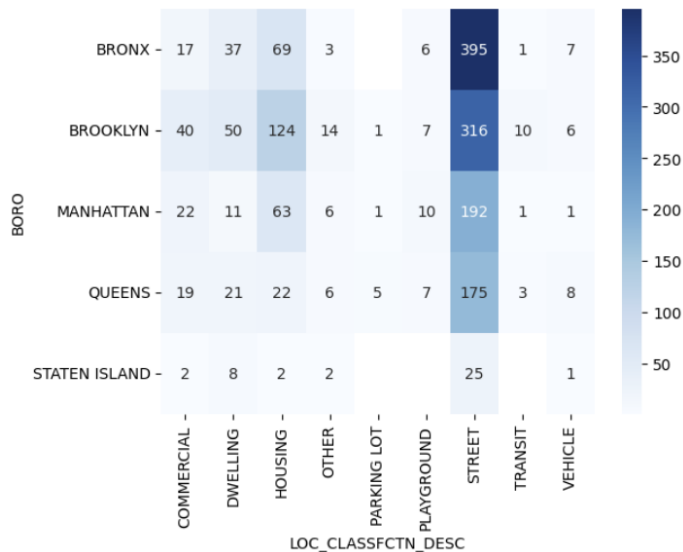
- The final visualization is a heatmap showing the density of shooting incidents across the city. The heatmap shows that shooting incidents are concentrated in certain areas, such as Brooklyn and the Bronx.

Overall Description

The program is a data analysis program that is focused on examining shooting incidents in New York City. The program reads the data from a CSV file and performs various preprocessing steps such as converting data types and filtering out incomplete data. It then proceeds to perform

various analyses such as examining the distribution of shootings by borough, race and sex of the victim and perpetrator, hourly and monthly trends of shooting incidents, and the locations of shooting incidents using maps to reveal important trends and patterns in the data that can be used to inform critical decision-making by policy makers.

Conclusions:

From the above visualizations, we can infer the below:

- The data shows that gun violence is a significant issue in New York City, and there are various trends that policy makers can observe to develop effective strategies to address this issue.
- The high number of shootings in Brooklyn, particularly on the streets, could lead to the implementation of community policing programs that target high-risk areas.
- The high number of young perpetrators and victims of gun violence highlights the need for early intervention programs that focus on youth outreach and education.
- Moreover, the disproportionate number of Black victims highlights the need for targeted initiatives that address the root causes of violence in the community.
- The majority of shooting incidents occur in Brooklyn and in public places like the streets, sidewalks, and parks.
- The incidents tend to occur mostly in the evening and early morning hours.
- The heatmap shows the concentration of shooting incidents in certain areas of New York City, especially in Brooklyn and the Bronx.
- The bar chart showing shooting incidents by victim and perpetrator sex highlights the fact that the majority of victims are male, and the majority of perpetrators are also male. This could suggest the need for targeted intervention programs aimed at reducing violent behavior among young men.
- Similarly, the pie chart showing shooting incidents by victim race suggests that Black and Hispanic individuals are disproportionately impacted by gun violence. This information can be used by policy makers to identify areas where additional resources may be needed to address this issue.
- The heatmap of shooting incidents by location can also provide valuable insights into where gun violence is most prevalent, which can help inform decisions around resource allocation and policing strategies.

Overall, the Python program provides a powerful tool for analyzing and visualizing complex data related to gun violence. The insights gained from these visualizations can be used to make more informed policy decisions, to make meaningful progress in addressing this critical issue.