

# **NYC Complaint Data Analysis**

**Presented by Yiqun Liu, Shuhua Yu, Xuchao  
Gao, Yi Li, Zhuoqing Xu, Yuxuan Liu**

# Research Question

**How to analyze ten years of criminal details and predict the severity of crime in New York based on multiple factors( time, area, etc.) and optimize the allocation of police resources based on the data prediction results.**

1. the most common crime months, the most common crime time of day, and the top 10 most common crime types during the year 2017-2023
2. How to predict the severity of the crime level
3. how to optimize the allocation of police resources and give the more effective crime prevention strategy.



# Data Selection

We collect the data from the NYPD official website which contains the crime data from 2009 to 2024.

01.

Considering that the huge data set may contain false positives, duplicates, omissions, and other error information, the final sample selection will involve data processing methods such as cleaning, filtering to eliminate inconsistencies, missing values, or outliers to get a sample with complete and accurate information. In addition, the sample will be segmented to train the prediction model.

02.

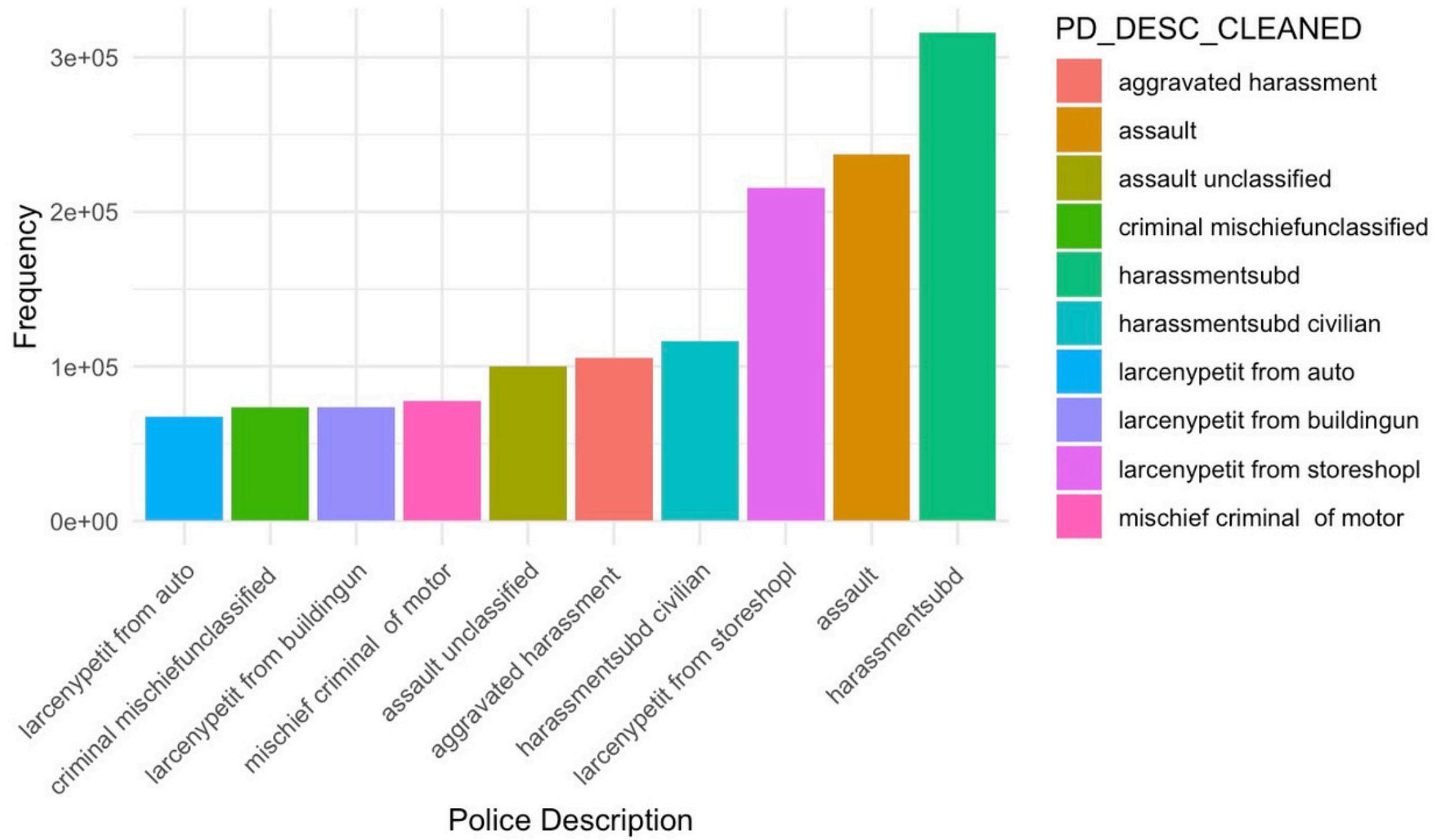
Given that this dataset was collected over a period of 15 years and the size of the data, we believe that factors such as false crime reports can be ignored and that this dataset can be used as a population representative of all crime incidents in the New York area.

03.

The unprocessed original data set contains error information (such as duplicate data, missing data, etc.), so it is necessary to use basic data set cleaning and filtering methods to eliminate useless data, reduce the error of the original data set, and obtain samples. Then, we will segment the sample (i.e. the processed data set) again.

# KEY FINDINGS FROM TIME SERIES ANALYSIS

Top 10 Most Common Police Descriptions



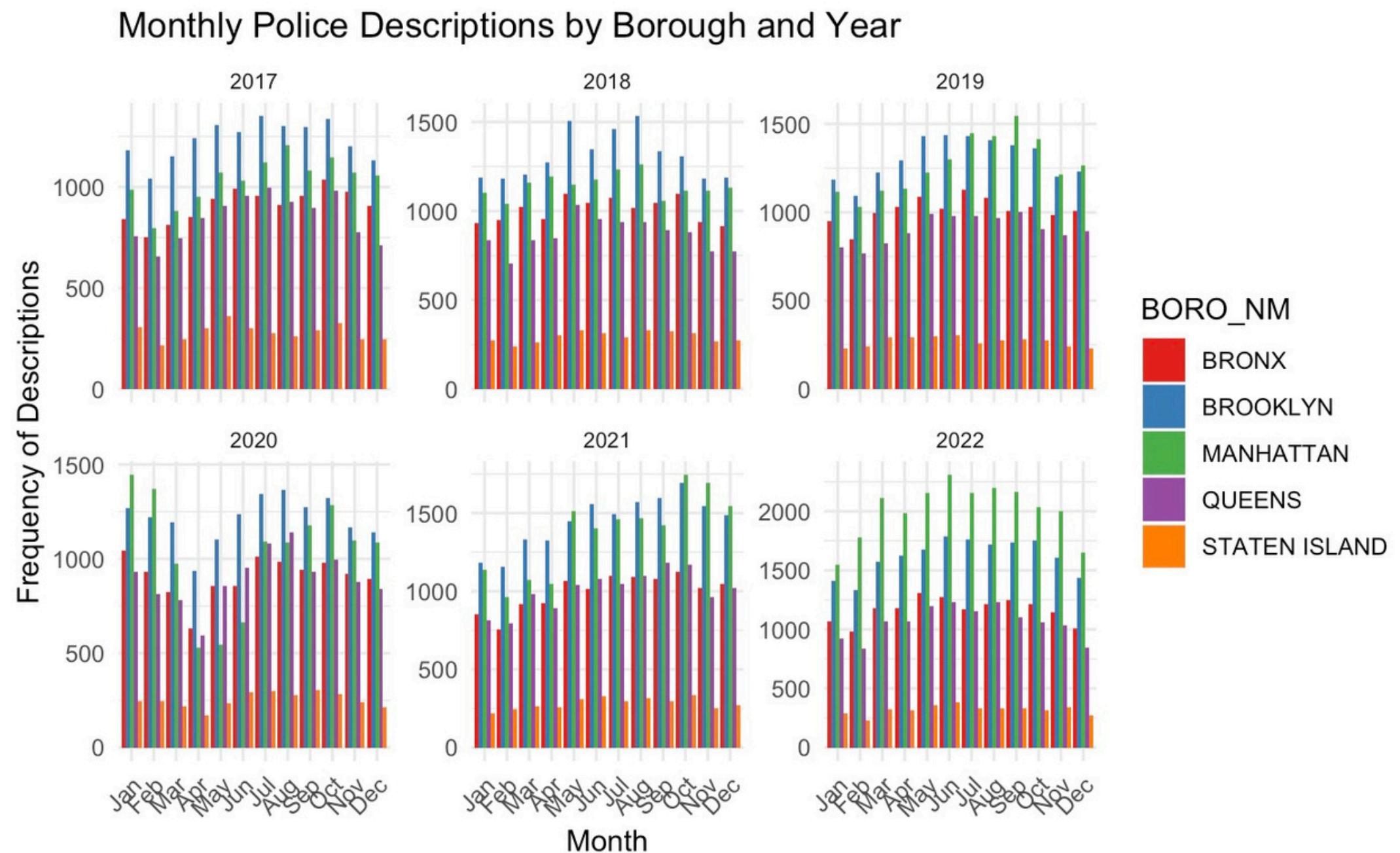
**This bar chart depicts the top 10 incident types by frequency, offering insights into common police reports.**

- Aggravated Harassment: Most frequent incident.
- Assault Categories: High occurrence, including classified and unclassified.
- Larceny Incidents: Notable reports of theft from autos, buildings, and stores.
- Criminal Mischief: Specifically related to motor incidents.

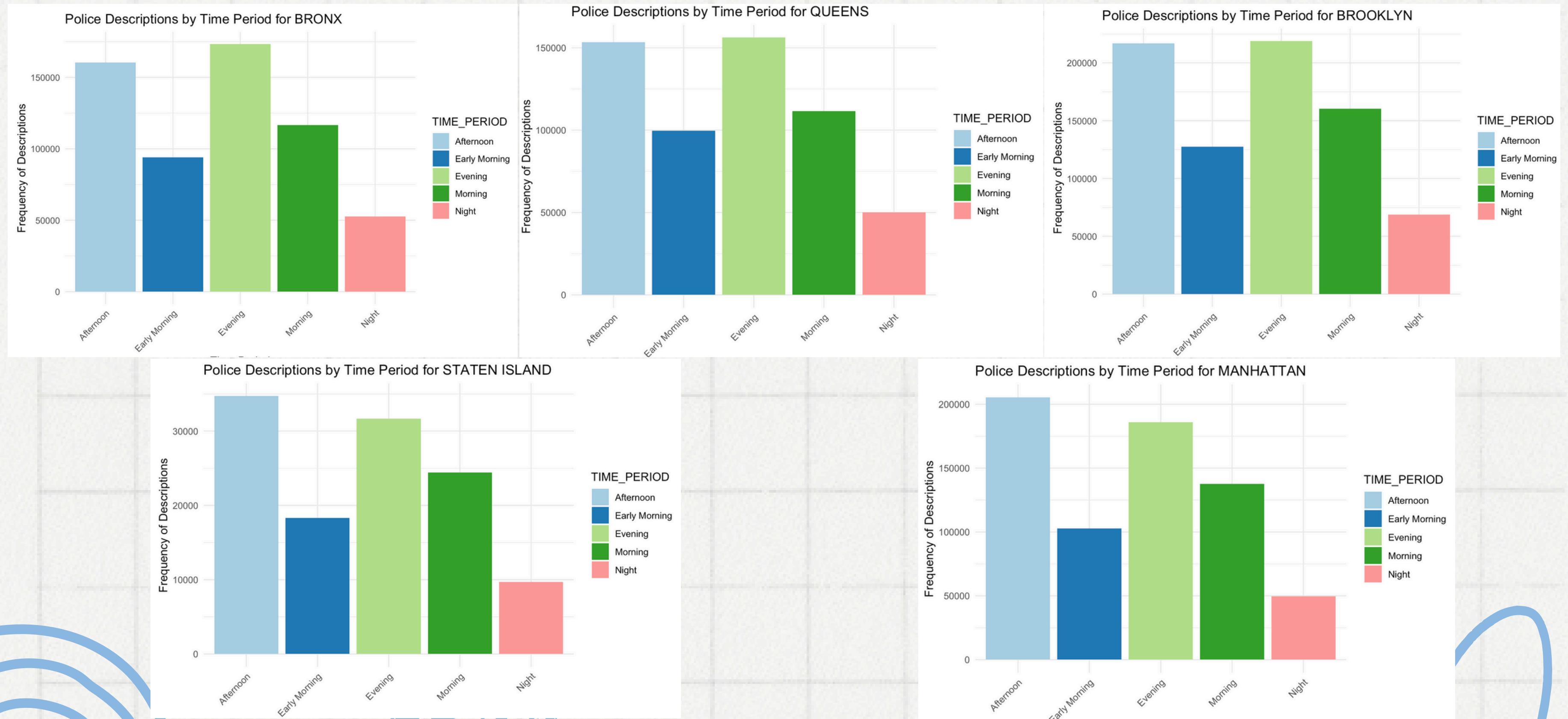
# CONTINUED...

## Key Observations:

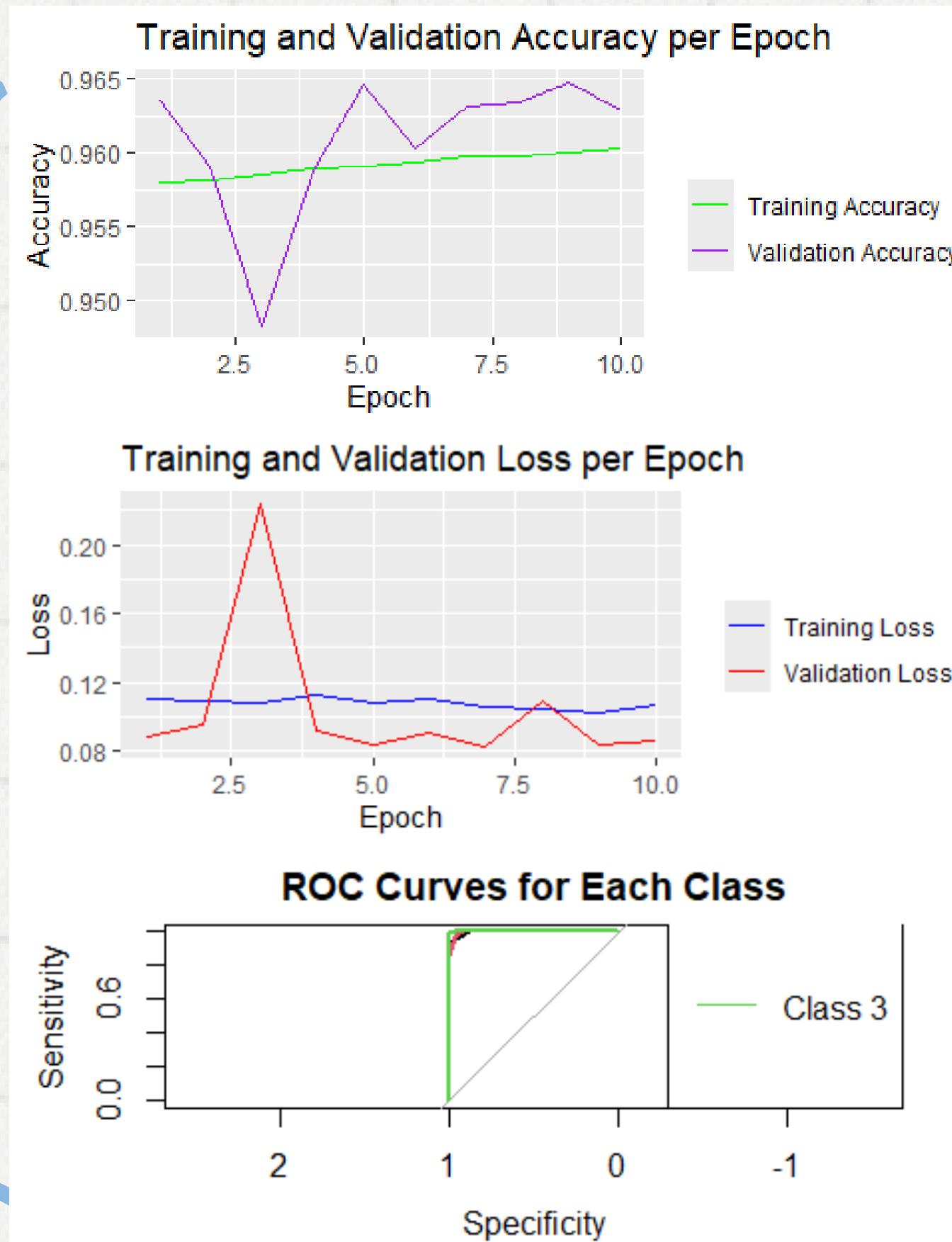
- **Bronx:** Notable peaks during mid-year and year-end across multiple years.
- **Brooklyn:** Sustained high frequency with occasional spikes, possibly indicative of larger population/events.
- **Manhattan:** Shows variability; some years have pronounced peaks, which may correlate with specific events.
- **Queens:** Exhibits mid-year peaks consistently, with a notable increase in the latter years.
- **Staten Island:** Lower frequency overall, but with clear spikes that may reflect localized incidents.



# FREQUENCY OF INCIDENT ACCROSS DIFFERENT TIMES OF THE DAY IN 5 BOROUGHS

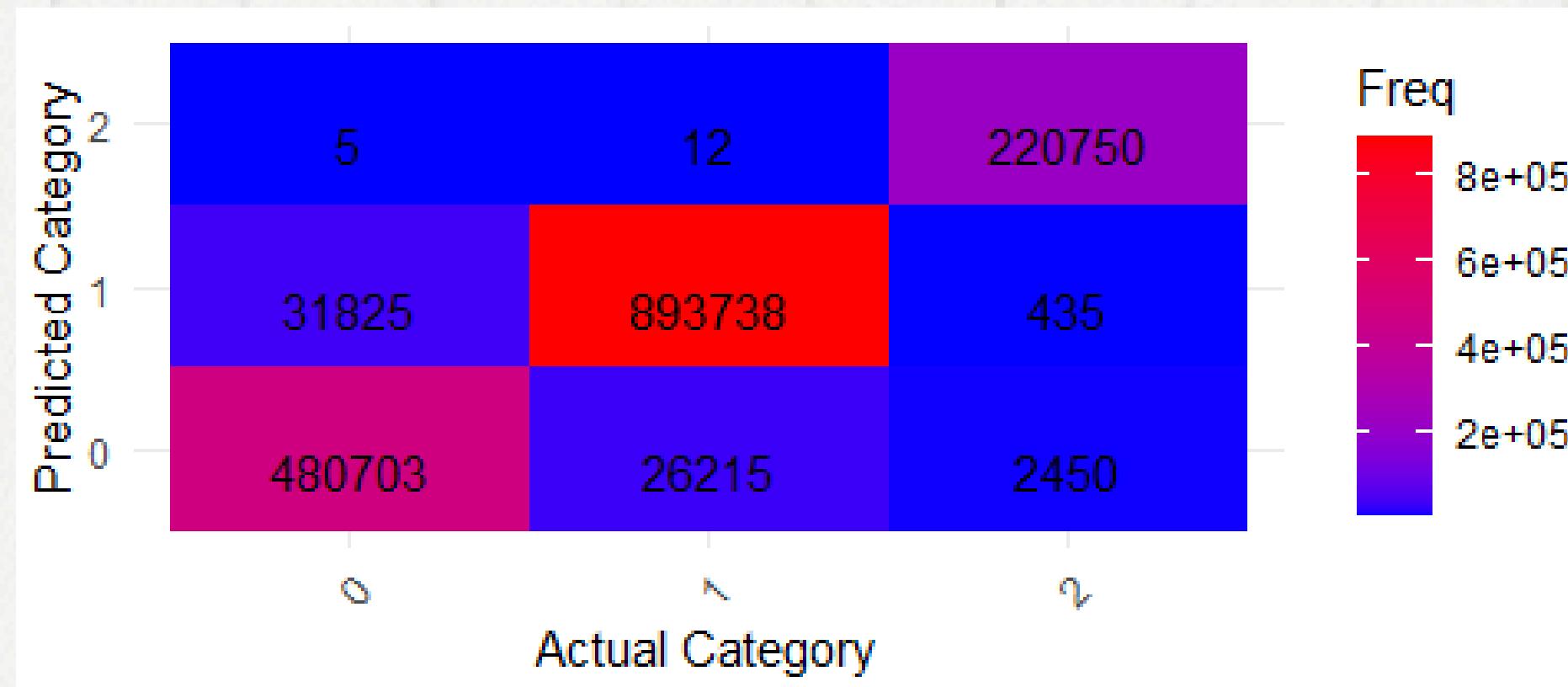


# PREDICTED THE OFFENSE



This is the result of Accuracy of the Neural network we build, we can see that the training accuracy is stable ,mean the Neural network has a very good converges behaviors, and the Validation line mean that the Lowest accuracy of prediction based on the test data is 95%

# PREDICTED THE OFFENSE



This confusion matrix says that most our predicted is right, but our neural network have a bigger chance wrong when it predict is result of 0 (Felony)and 1(Misdemeanor).

# **96%**

## **Success cases– Prediction Evaluation**

We clean the data and define Neural network by ourselves, the accuracy of predict on test dataset is 96%



# Improvements and Key Insights

**Feature Engineering:** The expansion of temporal features and the inclusion of binary indicators for each month and period of time provide a more detailed temporal context, which is crucial for capturing crime patterns. Reduce the complexity of the model, the matrix cost also lower than before.

**Data filtering:** By focusing on the most recent data (post-2016), the new model is able to better align with current crime trends, potentially improving its usefulness in contemporary forecasting tasks.

**Model complexity:** The transition from basic neural networks in V1 to more complex multinomial logistic regression in later versions allows for a more nuanced understanding of the effects of various predictors. V4 becomes a neural network with a hidden level of 64x10, and combined two activations : relu and softmax.

# Strategic Plan

We are divided into two parts, one is the analysis of crime details based on times series method, and the other is the prediction of crime severity based on prediction. As for the Prediction result, we suggest that NYPD may consider adjusting the allocation of police resources according to the prediction result as appropriate. However, because the prediction model needs to be perfected, it is suggested to make a reasonable judgment based on the 10-year police situation analysis given by us, rather than relying entirely on the prediction.

**Combined with the crime severity forecast, focus on the areas of Manhattan and Brooklyn throughout the year, increase the number of police officers, patrol stations, or better police supplies in these areas during the two high crime time period(afternoon and evening).**



## Crime type

According to the above analysis, we can learn that the top three major crime types with the highest frequency are harassment, assault and larceny/petit from shop, which are less harmful to the police and the public. We recommend that the NYPD equip most officers on duty with regular police protection, and for less frequent criminal cases or shootings, for special operations, officers involved in operations with high-grade protection or firearms.

## The peak crime hours of the day

In all areas, the frequency of crime is significantly higher in the afternoon and evening than in the other two time periods, and it is recommended that more officers be on patrol or on duty during these two times and be ready to respond at any time. It can be reduced in other time periods.

## Areas

the areas with high incidence of criminal incidents are Manhattan and Brooklyn, so it is suggested to arrange major police forces in these two areas and strengthen prevention by increasing the number and duty stations.

## Monthly police description

In the absence of significant high crime months throughout the year, it is suggested that NYPD does not need to make special police arrangements according to the month, but only needs to be fully prepared during the holidays and focus on the high crime periods every day.



**Thank you  
for  
listening.**