Analysis of NYPD Complaint Data Using Machine Learning Techniques

ADTA 5340

Professor: Tony Fantasia

Team
Brainy Bots

Tejamanikanta Gudla

Sai Preethi Jiddigam

Harish Chinnakadiri

Indusree Terala

Sai Swetha Puli

TABLE OF CONTENTS

Introduction

New York Police Department use the Complaint records for seeking insights to increase patroling to where the crime rates are gradually high based on time frames, Boroughs and Location axes of the map. New York being the most visited city in America by tourists, more than 95% of the resident's commute for work daily. Being the liveliest place, the crime rate of New York city should be taken into consideration to oversee the public safety.

How AI and ML Impacts

Many of the Departments PredPol type of tools to predict crime suspecting locations. Facial Recognition sensors in surveillances helps in two ways. Many algorithms used to find patterns in the crime based on the offensive description and the crime identification.

Business Understanding

1. Operational Efficiency:
   - Patrol Management
   - Emergency Response

2. Public Safety and Trust:
   - Enhanced Public Confidence
   - Community Relations

Problems to Solve

This data covers the crime recorded complaints data in New York region. The problem can be defined as how to use this data to better predict crime locations based on complaints.

Dataset

- NYPD Crime complaint records

Data Explanation

In this Dataset we have Unique complaint number and details of incident recorded data and time. These categorizes into crime codes by KY_CD and PD_CD with OFNS_DESC and PD_DESC and the status of the complaint with law category and location latitude and longitudes. It had 361740 records and 24 columns in the raw dataset.

Data Preparation

This steps follows the removal of columns with count above the - 50% missing values as well as all rows containing missing values in any cell to maintain data accuracy. Moreover, non-numeric data columns like the created date and time for each complaint listing were also removed due to redundancy.

Exploratory Data Analysis

Bar chart

Complaints based on Borough Names

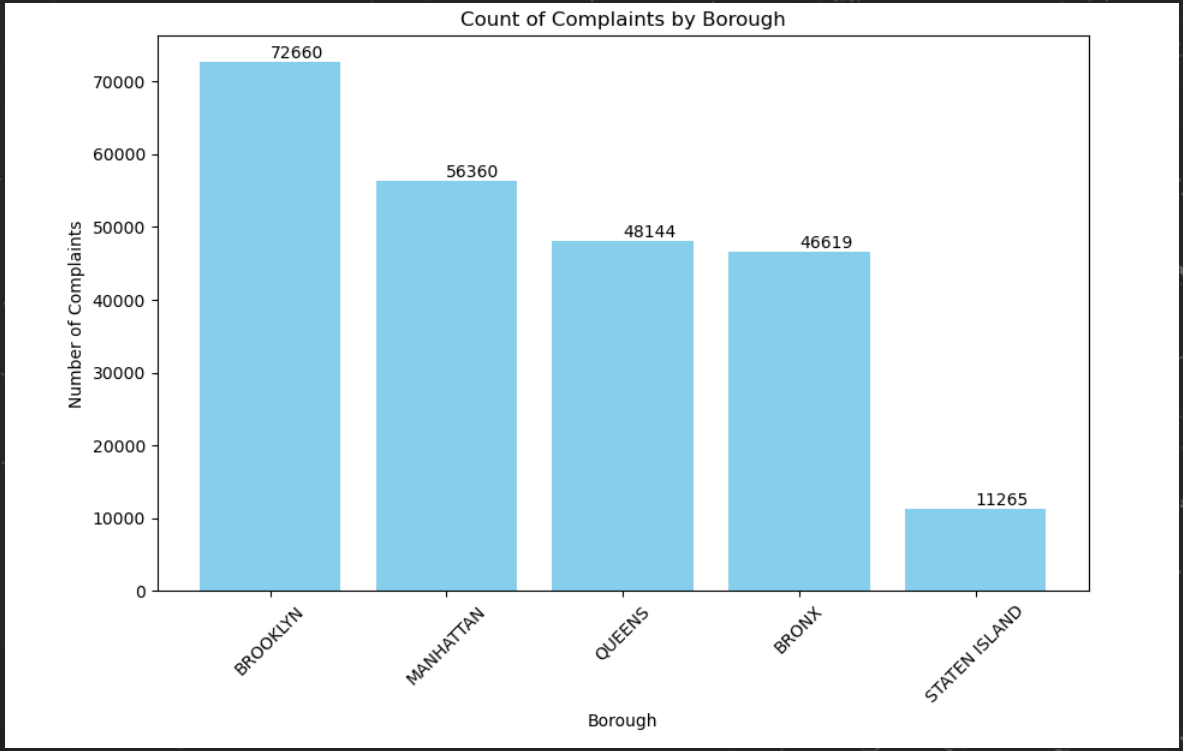
Pie Chart

Offensive description

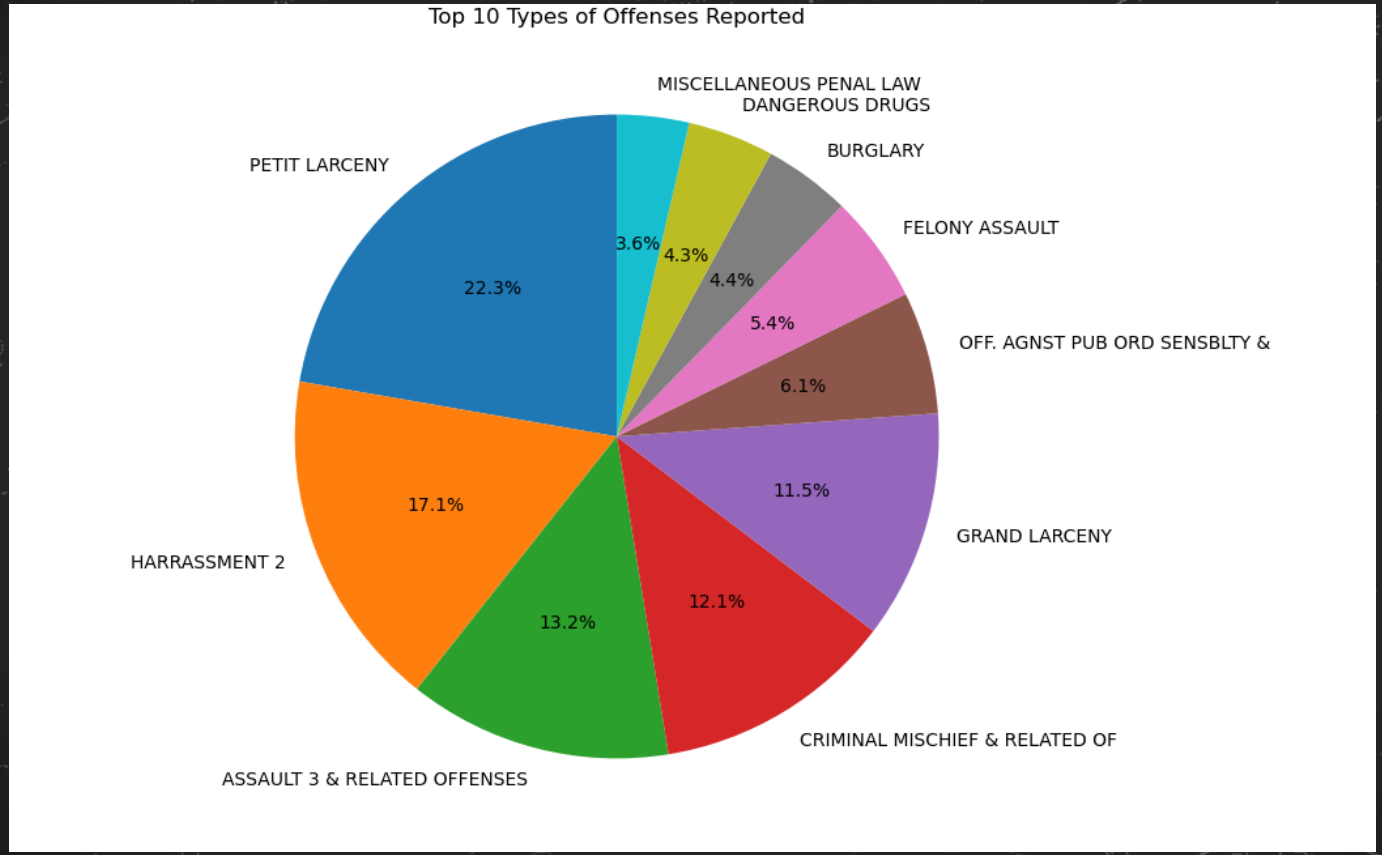
Stacked bar chart

Outcome of the law Category

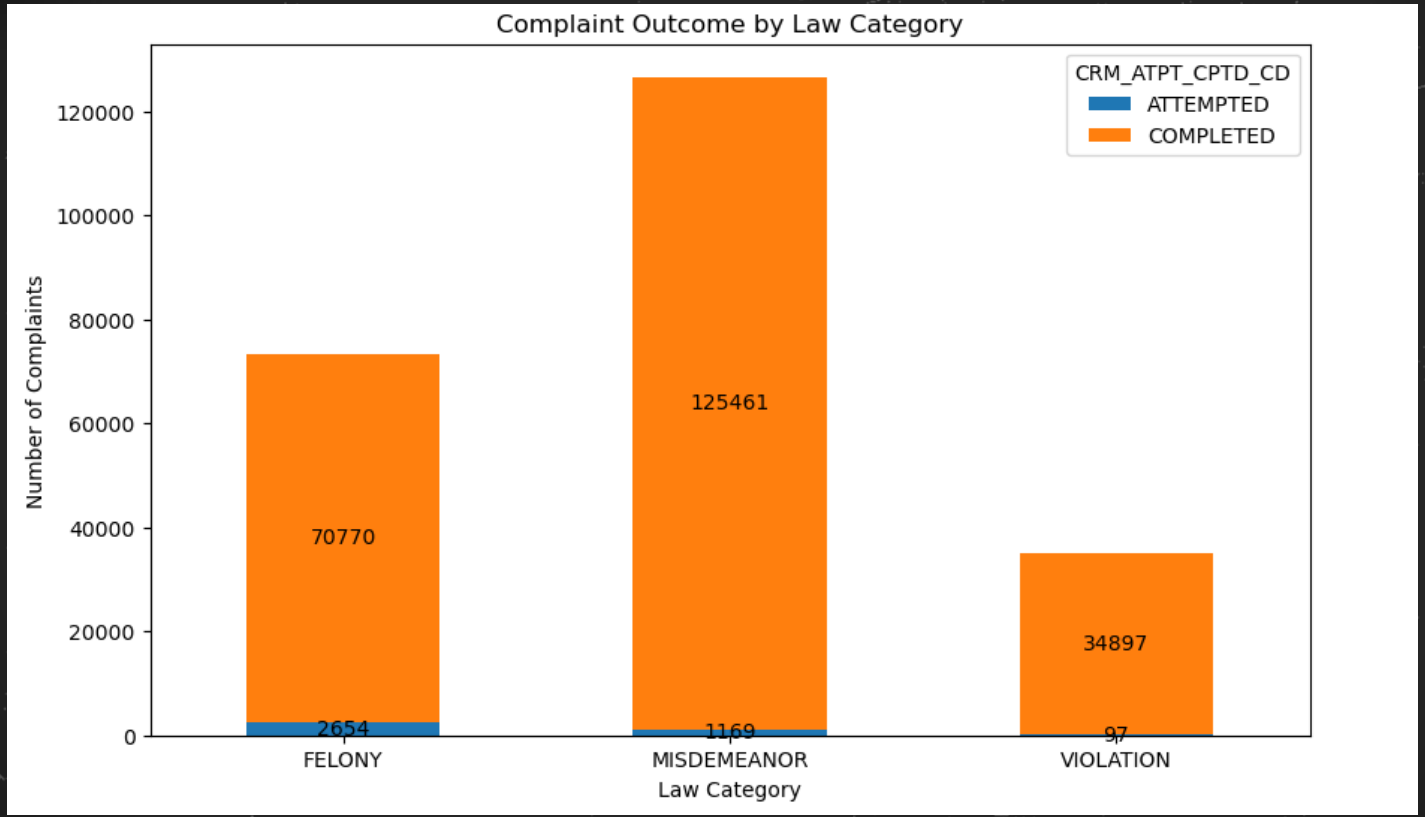
Plots


Count of Complaints by Borough
Number of Complaints
0
10000
20000
30000
40000
50000
60000
70000
72660
56360
48144
46619
11265
BROOKLYN
MANHATTAN
QUEENS
BRONX
STATEN ISLAND
Borough

Plots


Top 10 Types of Offenses Reported
MISCELLANEOUS PENAL LAW
DANGEROUS DRUGS
BURGLARY
FELONY ASSAULT
OFF. AGNST PUB ORD SENSBLTY &
GRAND LARCENY
CRIMINAL MISCHIEF & RELATED OF
ASSAULT 3 & RELATED OFFENSES
HARRASSMENT 2
PETIT LARCENY
3.6%
4.3%
4.4%
5.4%
6.1%
11.5%
12.1%
13.2%
17.1%
22.3%

Plots


Complaint Outcome by Law Category
CRM_ATPT_CPTD_CD
ATTEMPTED
COMPLETED
Number of Complaints
120000
100000
80000
60000
40000
20000
0
70770
2654
125461
1169
34897
97
FELONY
MISDEMEANOR
VIOLATION
Law Category

Data Encoding & validation,

In our data we have many categorical columns. So, we have used label encoding to manage the data columns. We have used K fold cross validation in the code for better results on training and testing data splits.

Algorithms Used

Logistic regression

Decision tree classification

Random forest classiification

Random forest regression

Algorithm Comparison

After using all the models, we are finalizing the model by the score produced by each model.

Model Prediction

We are calculating the scores for each model taken in the process

Future Suggestions

References

- *NYC Crime Data - dataset by data-society.* (2024, April 15). data.world. https://data.world/data-society/nyc-crime-data

- *NYPD Announces the Publication of its 2023 Strategic Plan.* (2023, January 25). The Official Website of the City of New York. https://www.nyc.gov/site/nypd/news/p00072/nypd-the-publication-its-2023-strategic-plan

Thank You