

A decorative graphic on the left side of the slide. It consists of a blue parallelogram and a light green parallelogram, both tilted at an angle. The blue shape is in the foreground, and the green shape is partially behind it. They are set against a dark blue background with faint, lighter blue diagonal stripes.

NYPD Crime Analysis



Team

Domenick Avanzi

Mary Kate Shea

Jason Yoo

Zac Zydonic



Overview

Topic

Reasoning

Data Source Description

Questions to be Answered

Data Exploration

Analysis Phase

Technologies, Languages, Tools, and Algorithms



Topic

The main objective of this analysis is to explore NYPD Complaint data provided from NYC Open Data using a the Random Forest classification logistic regression model.

Specifically, the analysis will examine crime descriptions, times, locations, etc. to create a binary classification model that will predict whether or not the type of crime committed is likely to be a violent crime. Furthermore, this will be supplemented by Tableau heat maps to show high crime areas per crime type. Time permitting and given the expanse of the data available, we can see if crime rates change over months/years.



Reasoning

This analysis is inspired by increased focused on policing in the United States over the past year. The goal is to explore and create unbiased observations on crime trends using the developed classification model.

This classification model, if successful, would infer/support where and when violent crimes & complaints may take place. Note the output of the model is to successfully deem a crime/complaint is violent or not. The model will support existing data to show pockets of violent data based on time and location trends.



Data Source Description

This current data source comes from the city of New York Open Data website. The data is provided by the NYPD and contains all valid felony, misdemeanor, and violation crimes reported.

- The dataset used contains data from 2020 to 2021 (324K rows of data). It can be found here <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243>
- There is an alternative dataset which contains data from 2006 to 2021 (7.38M rows of data) . It can be found here: <https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i>



Questions to be Answered

This analysis is intending to classify whether or not a crime is violent (output) based on a certain area, and/or at a certain time (input).



Data Exploration

To predict whether or not a crime is violent depending on different factors associated with the registered complaint and to analyze the factors in the dataset and eventually leading to the crime bucket.



Analysis Phase

Using regression model to forecast binary - “violent” or “nonviolent” - crime in NYC.

- Benefits:
 - Provides correlation and reasonable outputs among many variables.
- Limitations:
 - It is assumed that the cause and effect relationship between the variables remains unchanged. The assumption may not hold an accurate estimation of the values based on the model and it may lead to misleading results.



Technologies, Languages, Tools

- Technologies
 - Random Forest Model
- Languages
 - Python, PostgreSQL
- Tools
 - Jupyter Notebook, VS Code
- Imports
 - **import** pandas **as** pd
 - **from** sklearn.model_selection **import** train_test_split
 - **from** sklearn.preprocessing **import** StandardScaler, OneHotEncoder
 - **import** tensorflow **as** tf
 - **from** sklearn.metrics **import** balanced_accuracy_score
 - **from** sklearn.metrics **import** confusion_matrix
 - **from** imblearn.metrics **import** classification_report_imbalanced
 - **from** sklearn.model_selection **import** train_test_split
 - **from** imblearn.ensemble **import** BalancedRandomForestClassifier
 - **from** sqlalchemy **import** create_engine
 - **from** config **import** db_password
 -



Algorithms & Data Structures

- Algorithms
 - Import Dependencies
 - Import and Read Data
 - Drop Unnecessary Columns and Rows with Null Values
 - Explore the Shape of the Data
 - See Value Counts for Offense Descriptions
 - Group Violent Crimes into List
 - Add Violent Crime column in Dataframe
 - Add 1 if violent or keep 0 for non-violent crime
 - Clean Age Range
 - Export to CSV
- Data Structures
 - Dataframes, lists



Tableau Dashboard

Link to the Dashboard is [here](#):

The dashboard allows users to filter through the feature columns of the dataset.

Users can filter by the following:

- Violent vs Non-Violent Crime
- Borough of occurrence
- Month of complaint/crime
- Offense description
- Victim sex, race and age group

Tableau Dashboard

