

Crime Pattern Analysis and Predictive Modelling for Urban Safety

Group members:

- Ashutosh Singh -0043
- Saris Singh -0025
- Ashutosh Patil-0014

Column Name	Type	Description
DR_NO	int64	Division of Records Number – Unique ID for the crime report.
Date Rptd	object (date)	Date Reported – When the crime was reported to authorities.
DATE OCC	object (date)	Date Occurred – When the crime actually happened.
TIME OCC	int64	Time of Occurrence – 4-digit military time (e.g., 2130).
AREA	int64	Area code for police reporting divisions.
AREA NAME	object	Name of the corresponding police area (e.g., "Wilshire").
Rpt Dist No	int64	Reporting District Number – More specific location code.
LOCATION	object	Street address of the incident.
Cross Street	object	Nearest cross street (may be missing).
LAT	float64	Latitude – Geographic coordinate of the crime.
LON	float64	Longitude – Geographic coordinate of the crime.
Part 1-2	int64	FBI classification: 1 = Part I (e.g., homicide), 2 = Part II (e.g., drugs)
Crm Cd	int64	Primary numeric code for the offense.
Crm Cd Desc	object	Text description of the crime (e.g., "BURGLARY").
Crm Cd 1 to 4	float64	Secondary/tertiary codes if multiple crimes occurred.
Mocodes	object	Modus operandi codes – how the crime was committed.
Weapon Used	float64	Numeric weapon code (if any).
Weapon Desc	object	Description of weapon used (e.g., "KNIFE", "HANDGUN").
Vict Age	int64	Age of the victim.
Vict Sex	object	Victim's gender (M = Male, F = Female, X = Unknown/Other).
Vict Descent	object	Ethnicity/race of victim (e.g., B = Black, H = Hispanic, W = White, etc.)
Premis Cd	float64	Premise code (e.g., house, store).
Premis Desc	object	Textual description of the premise (e.g., "STREET", "STORE").
Status	object	Abbreviated case status (e.g., AA = Adult Arrest, IC = Investigation Continued).
Status Desc	object	Full description of the status (e.g., "Adult Arrest").

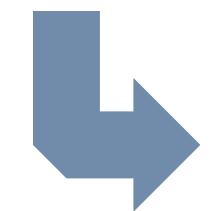
DATASET

LINK:

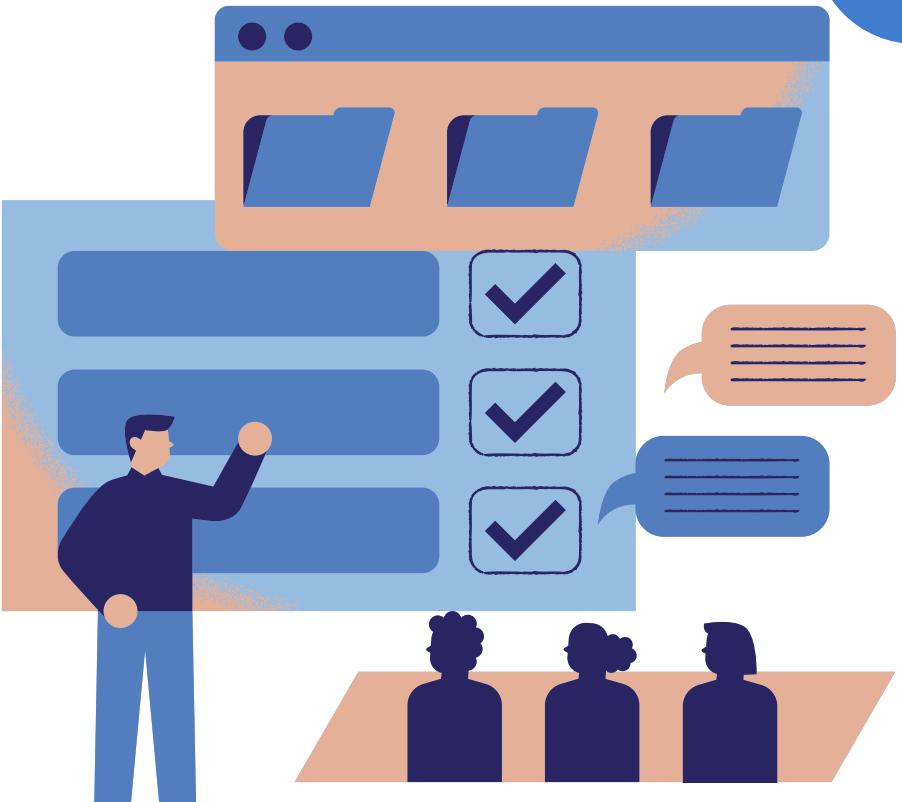
<https://catalog.data.gov/dataset/crime-data-from-2020-to-present>

DATA PREPROCESSING

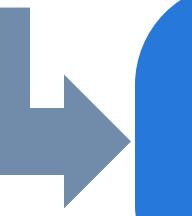
DATA
REPRESENTATION



IDENTIFYING NULL/GARBAGE
VALUE IN PERCENTAGE



CLEANING BY DROPPING
COLUMNS WITH $>60\%$

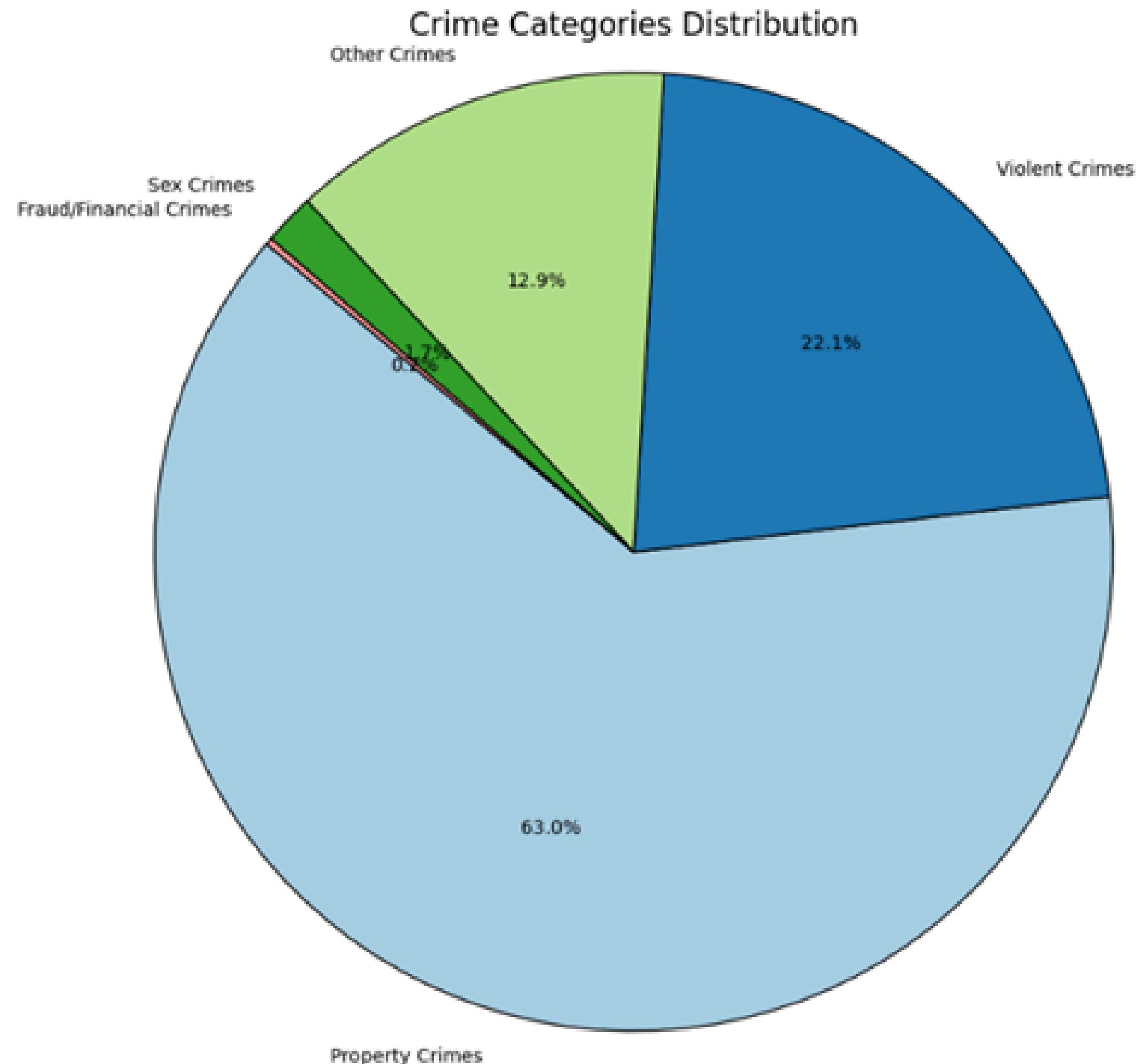
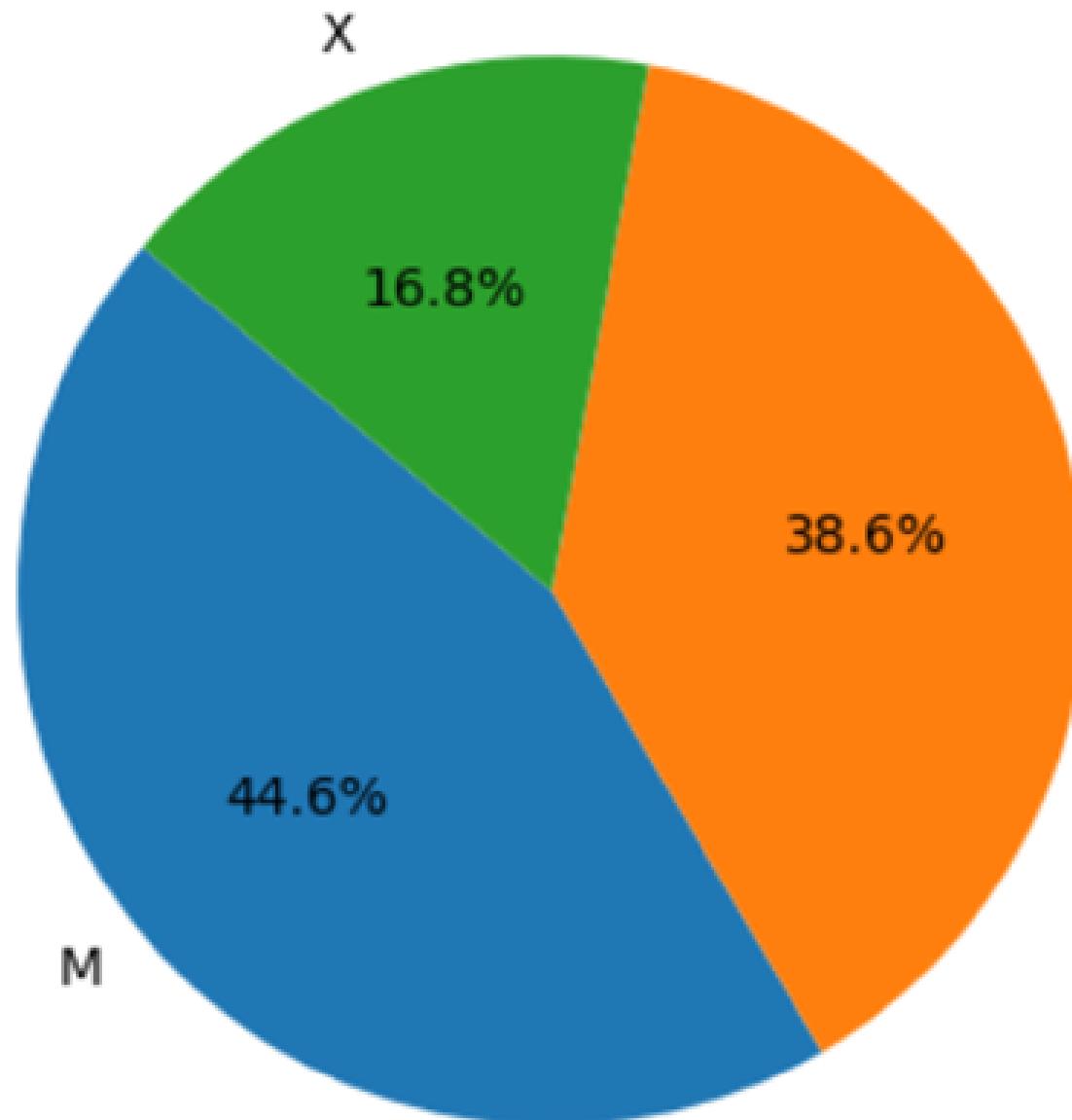


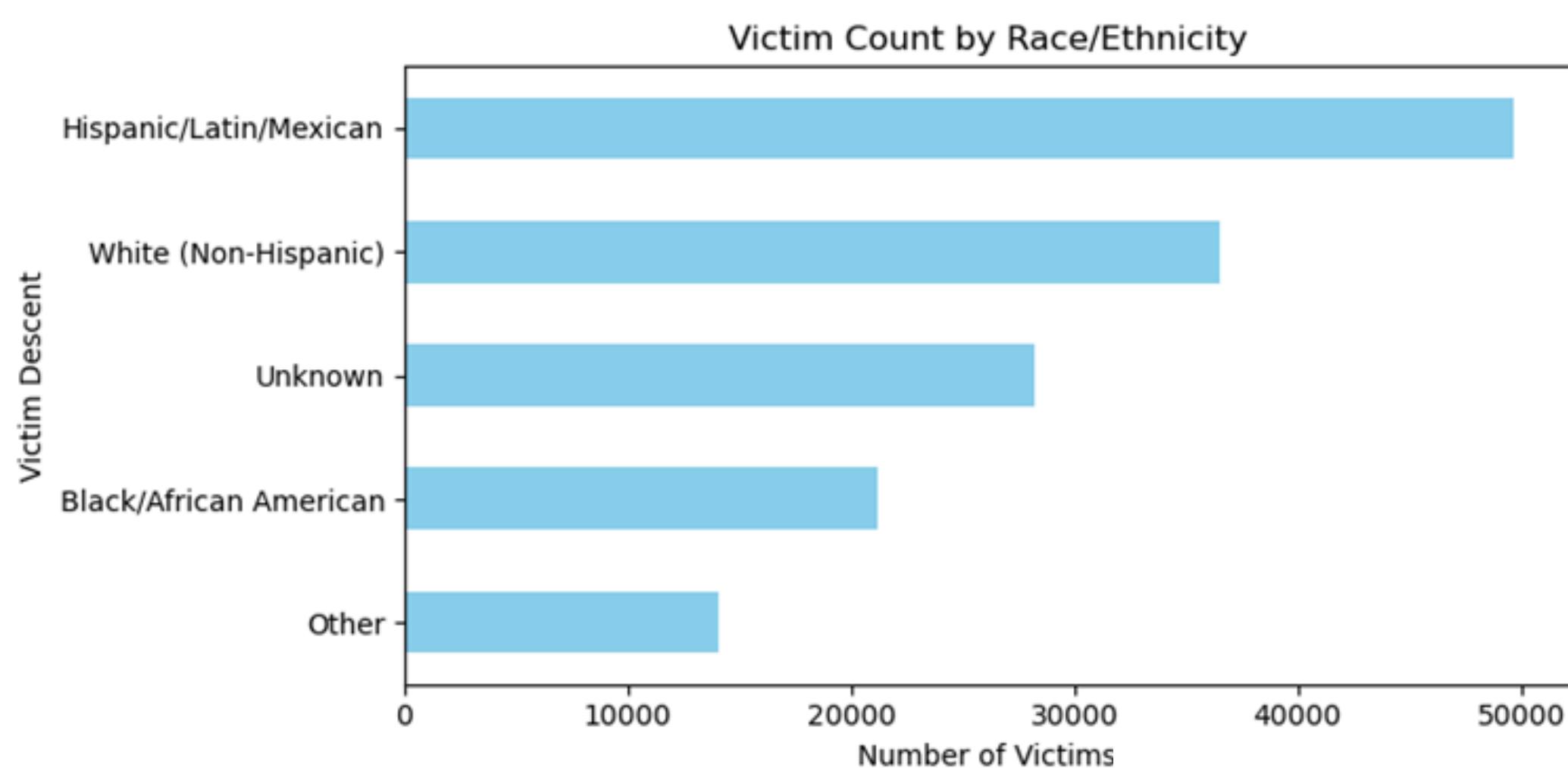
REMOVING ROWS WITH
NULL/GARBAGE VALUE



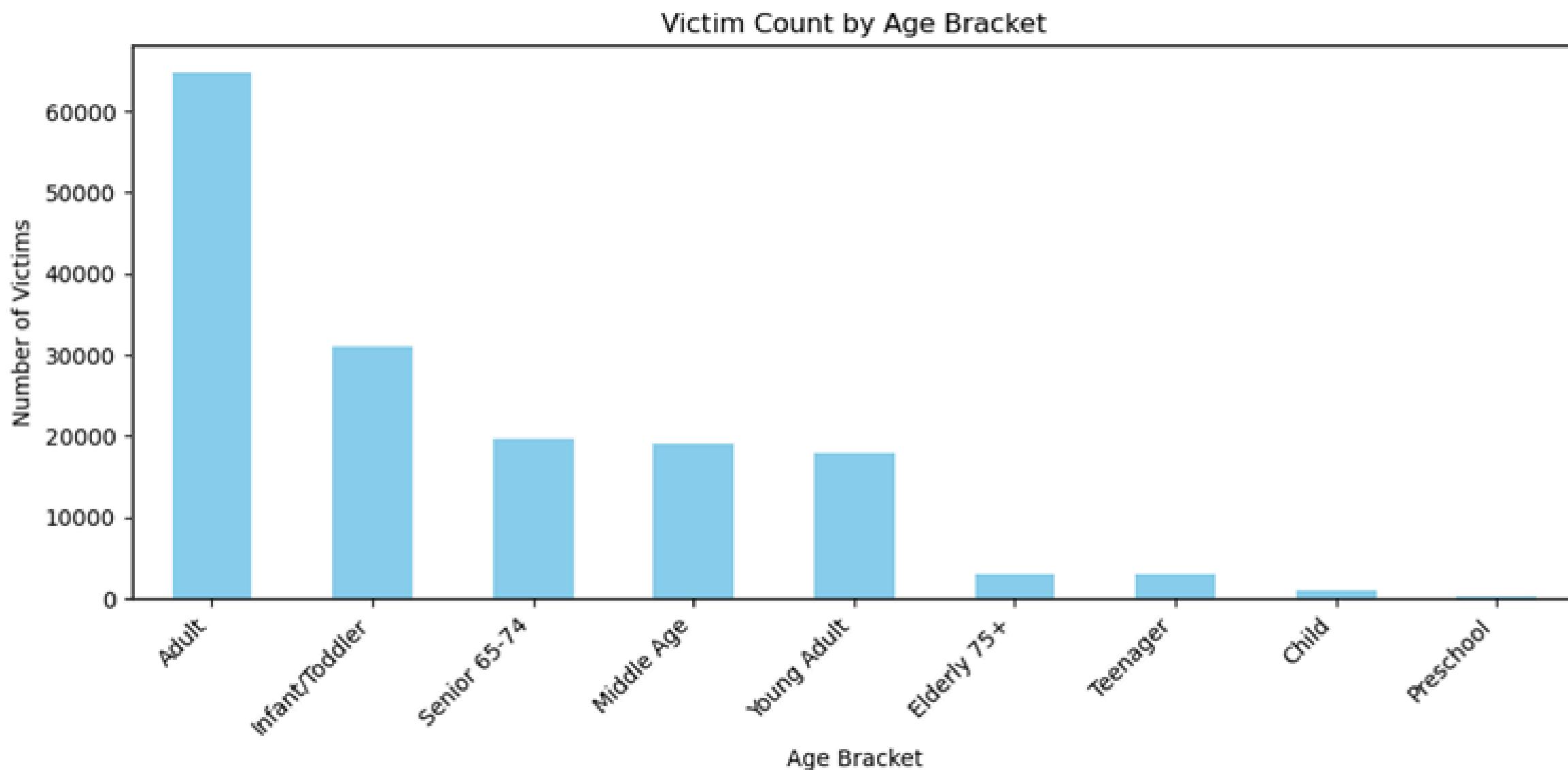
Exploratory Data Analysis (EDA)

Gender Distribution of Victims





- This bar plot shows the relationship in data between victims in different race based on total no of victims for each type of race.
- Findings are such that Mexicans/ Latins is the major race against whom the crime rate is quite very high (approx. – 50,000 cases)in similar fashion as per the data the victims count is quite very high for race non-Hispanic -> Black -> Other



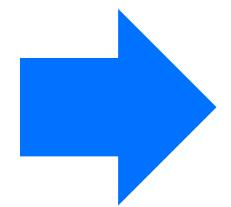
- Bar graph shows relationship between victims of different age groups classified and bracketed in different groups and number of incidents
- The major concerning age group as per the data is of Adults(60,000 +) cases followed by infant / toddler then senior citizens and so on...

- From the below bar chart we can deduce information about number of crimes committed in specific areas .
- The high crime rate is been observed in city Central followed by South west and Pacific at same rate later N-Hollywood with crime rate of 14k, 10k, 9k, 8.5k respectively



Model Preparation

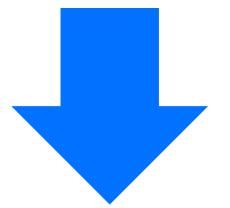
Converting Input data from categorical to numerical data



Splitting data into training and testing datasets



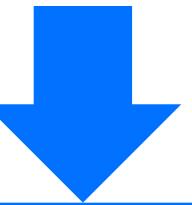
model fitting (Random forest)



Checking Accuracy

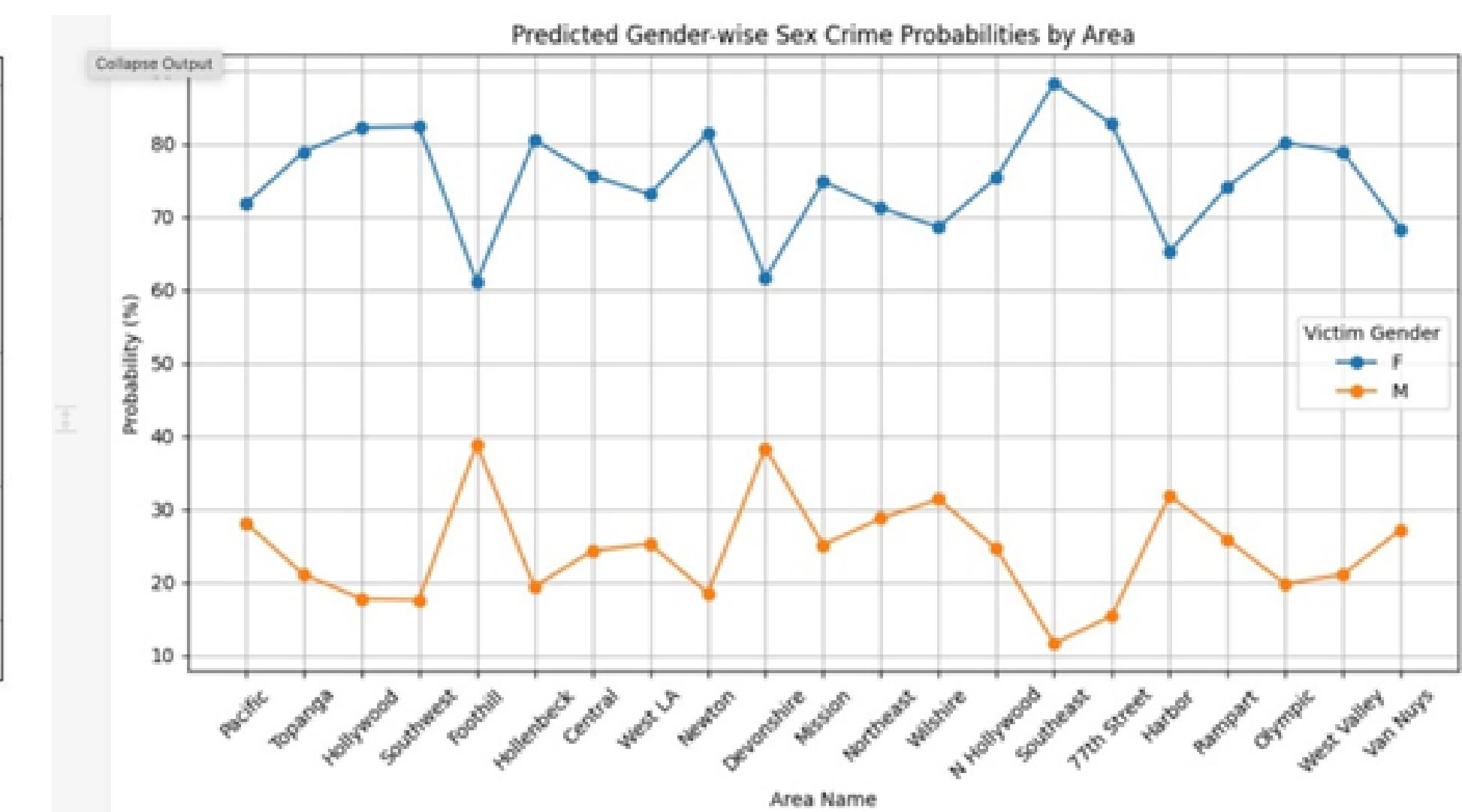
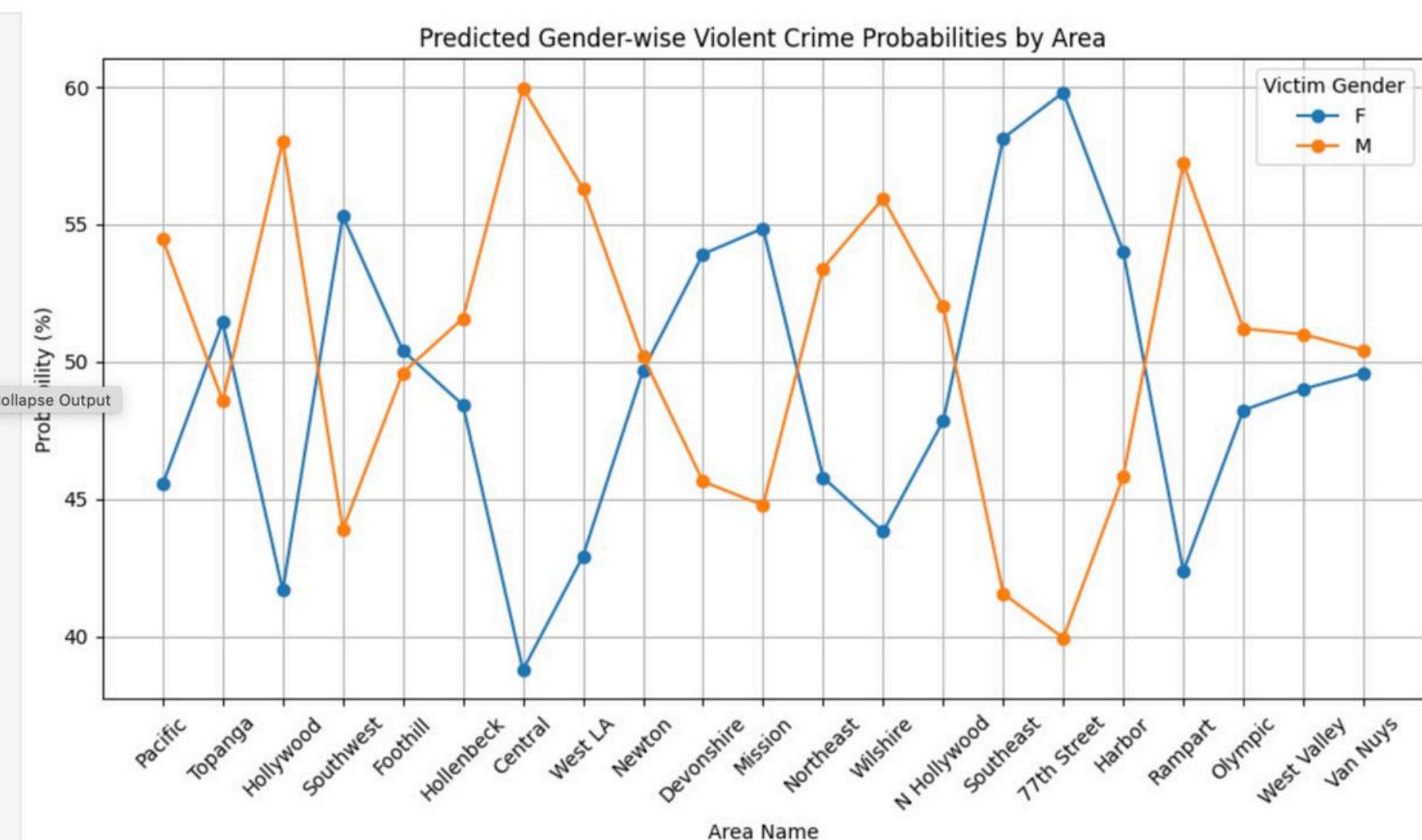


Probability prediction



Prediction

- The line plot illustrates predicted occurrences of violent crime and sex crime across different areas, categorized by gender.
- Predictions indicate higher crime rates against females in 77th Street for violent crimes and Southeast for sex crimes, while male victimization is notably higher in the Central region for violent crimes and Foothill for sex crimes.



Future Enhancement

- The model is being enhanced to analyze and predict crime hotspot areas using features such as crime type, time range, age group, and gender.
- The objective is to enable strategic deployment of law enforcement teams in high-risk locations where the probability of crime occurrence is significantly higher.





ANY
QUESTIONS?

THANKYOU!!



The screenshot shows a Jupyter Notebook interface running on a local host. The title bar indicates the notebook is titled "us_crime_data_final" and was last checked at "yesterday". The menu bar includes File, Edit, View, Run, Kernel, Settings, and Help. A toolbar below the menu bar contains icons for various operations like saving, running cells, and kernel selection. The status bar at the bottom right shows "Trusted" and "Python 3 (ipykernel)".

The notebook content consists of several code cells:

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from imblearn.over_sampling import SMOTE
import numpy as np
import xgboost as xgb
```

Importing Dataset

```
[2]: df = pd.read_csv('/Users/ashutoshpatil/Downloads/Crime_Data_from_2020_to_Present.csv').tail(200000)
df.head()
```

Data Pre-processing

```
[3]: df.info()
[4]: # checking null percentage of each column
round((df.isnull().sum() / len(df)) * 100,2)
[5]: # dropping columns whose null % are > 60
df = df.drop(columns = ['Weapon Used Cd','Weapon Desc','Crm Cd 2','Crm Cd 2','Crm Cd 3','Cross Street','Crm Cd 4'])
df.info()
[6]: # convert date,time columns to datetime
df['Date Rptd'] = pd.to_datetime(df['Date Rptd'],format = 'mixed',errors = 'coerce')
df['DATE OCC'] = df['DATE OCC'].str.split(' ').str[0]
df['DATE OCC'] = pd.to_datetime(df['DATE OCC'],format = 'mixed',errors='coerce')
[7]: # normalize values in same format
df['TIME OCC'] = df['TIME OCC'].astype(str).str.zfill(4)
# extracting features from date for better ml predictions
df['DayOfWeek'] = df['DATE OCC'].dt.day_name()
df['Month'] = df['DATE OCC'].dt.month
```

[**VIDEO LINK:**](#)