**PROJECT REPORT**

Project title

**Exploring Crime Analysis with LAPD Leveraging Machine Learning for Public Safety**

**Industrial Project Based Learning**

**Capstone Project**

By

**TEAM 7**

**Team Members:**

CHINTA SAI RANGANATH 22R15A6703

SIDRAMYNA RATHNAKAR 22R15A6707

MARIPALLI FARIDUDDIN 21R11A6730

LUNAVATH THIRUPATHI 22R15A6705

NANDIGAMA LIKITH CHOWDARY 21R11A6739

KOTA AMITH 21R11A6724

**Geethanjali College of Engineering and Technology**

**Cheeryal, Keesara, 500085, Telangana**

**APRIL 2024**

**ABSTRACT**

Crime remains a persistent and complex challenge affecting communities worldwide, necessitating effective strategies from law enforcement agencies and policymakers. In this we study a dataset provided by the Los Angeles Police Department (LAPD), covering the years 2020 to 2024, we aim to unveil insights into crime dynamics, victim demographics, spatial-temporal trends, and determinants of crime severity.

Additionally, we leverage advanced machine learning methodologies to develop predictive models. These models not only forecast crime types and severity but also endeavor to predict victim gender, offering invaluable tools for proactive intervention and resource allocation... To this end, we embark on the development of a user-friendly web interface utilizing the Flask framework. This interface serves as a conduit for users to interact with our insights seamlessly, empowering them to make informed decisions and implement targeted strategies to combat crime effectively.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **TOPICS** | **PAGE NUMBERS** |
| i | Abstract | i |
| 1 | Introduction | 1 |
| 2 | Literature Survey | 2 |
| 3 | Problem Statement | 2 |
| 4 | Objective | 3 |
| 5 | Methodology | 4 |
| 5.1 | Data Collection | 4 |
| 5.1.1 | Brief description of the dataset | 5 |
| 5.1.2 | Datatype of each feature | 5 |
| 5.2 | Importing dataset and required packages | 6 |
| 5.3 | Data Cleaning | 7 |
| 5.4 | Handling Missing values | 8 |
| 5.5 | Outlier detection | 9 |
| 5.6 | Outlier Handling | 12 |
| 5.7 | EXPLORATORY DATA ANALYSIS | 14 |
| 5.7.1 | Data Preprocessing | 14 |
| 5.7.2 | Feature Engineering | 14 |
| 5.7.3 | Visualisations | 18 |
| 5.8 | Research Problems | 25 |
| 5.8.1 | Crime Severity Prediction | 25 |
| 5.8.2 | Predicting Victim Gender | 27 |
| 5.8.3 | Crime Trend Analysis and Resource Allocation Optimization | 29 |
| 5.8.4 | Predicting Crime Type | 36 |
| 5.8.5 | High-Crime Area Identification | 38 |
| 6 | RESULT | 39 |
| 6.1 | Crime Severity Prediction | 39 |
| 6.2 | Predicting Victim Gender | 41 |
| 6.3 | Predicting Crime Type | 43 |
| 7 | CONCLUSION | 46 |
| 8 | FUTURESCOPE | 46 |
| 9 | REFERNECS | 47 |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Fig No** | **Figure Name** | **Page No** |
| 2.1 | Los Angeles crime spots view | 2 |
| 5.2 | Columns with Null values | 6 |
| 5.3.1 | Visualization of Null values | 7 |
| 5.3.2 | Visualization of Non Null values | 9 |
| 5.5.1 | Visualizing outliers using box plot for Vict Age | 10 |
| 5.5.2 | Visualizing outliers using box plot for LAT | 10 |
| 5.5.3 | Visualizing outliers using box plot for LON | 11 |
| 5.5.4 | Visualizing outliers using box plot for Vict sex and descent | 11 |
| 5.6.1 | Visualization of outlier handling for Vict Age | 12 |
| 5.6.2 | Visualization of outlier handling for LAT | 12 |
| 5.6.3 | Visualization of outlier handling for LON | 13 |
| 5.6.4 | Visualization of outlier handling for Vict sex and Vict descent | 13 |
| 5.7.1 | bar graphs for Area Nmae | 18 |
| 5.7.2 | bar graphs for vict sex | 18 |
| 5.7.3 | pie chart for Vict sex | 19 |
| 5.7.4 | bar graph for top 10 vict descent | 19 |
| 5.7.5 | bar graph for bottom 10 vict descent | 20 |
| 5.7.6 | pie chart for status | 20 |
| 5.7.7 | Histogram for Vict age | 21 |
| 5.7.8 | bar graph of distribution of incidents by year | 21 |
| 5.7.9 | pie chart for report lag | 22 |
| 5.7.10 | Map view of Los Angeles(source=google) | 22 |
| 5.7.11 | Scatter plot Map view of Los Angeles | 23 |
| 5.7.12 | Area wise divided scatter plot map view | 23 |
| 5.7.13 | bar plot for holidays | 24 |
| 5.7.14 | Severity plots | 24 |
| 5.8.1 | Metrices for Crime Severity Prediction | 25 |
| 5.8.2 | Web screen 1 for Crime Severity Prediction | 26 |
| 5.8.3 | web screen 2 for Crime Severity Prediction | 26 |
| 5.8.4 | Metrices for Predicting Victim Gender | 27 |
| 5.8.5 | Web screen 1 for Predicting Victim Gender | 28 |
| 5.8.6 | Web screen 2 for Predicting Victim Gender | 28 |
| 5.8.7 | Trend of incident by year | 29 |
| 5.8.8 | Trend of incident by Month | 29 |
| 5.8.9 | Trend of incident by weeks | 30 |
| 5.8.10 | Trend of incident by days | 30 |
| 5.8.11 | Trend of incident by day of week | 31 |
| 5.8.12 | Trend of incident by season | 21 |
| 5.8.13 | Distribution of incidents by season | 32 |
| 5.8.14 | Trend of incidents by hour | 32 |
| 5.8.15 | Distribution of incidents by hour | 33 |
| 5.8.16 | Trend of incidents over all years | 33 |
| 5.8.17 | daily crime counts over time | 34 |
| 5.8.18 | Area crime counts | 34 |
| 5.8.19 | Average crime severity over time | 35 |
| 5.8.20 | Web screen 1 for Predicting Crime Type | 36 |
| 5.8.21 | Web screen 2 for Predicting Crime Type | 37 |
| 5.8.22 | Web screen 3 for Predicting Crime Type | 37 |
| 5.8.23 | Scatter plot of coordinates with weapon categories | 38 |
| 6.1.1 | Running Flask application on localhost | 39 |
| 6.1.2 | Web address of Flask Application | 39 |
| 6.1.3 | Input to the web 1 | 39 |
| 6.1.4 | 4 Input to the web 2 | 40 |
| 6.1.5 | Output 1 for Crime Severity Prediction | 40 |
| 6.1.6 | Output 2 for Crime Severity Prediction | 41 |
| 6.2.1 | Web address of Flask Application | 41 |
| 6.2.2 | Input to the web 1 | 41 |
| 6.2.3 | Input to the web 2 | 42 |
| 6.2.4 | Output for Gender Prediction | 42 |
| 6.3.1 | Web address of Flask Application | 43 |
| 6.3.2 | Input to the web 1 | 43 |
| 6.3.3 | Input to the web 2 | 43 |
| 6.3.4 | Input to the web 3 | 44 |
| 6.3.5 | Output 1 for crime type prediction | 44 |
| 6.3.6 | Output 2 for crime type prediction | 45 |
| 6.3.7 | Output 3 for crime type prediction | 45 |

1. **Introduction**

Crime is a multifaceted and pervasive challenge that transcends geographical boundaries, impacting communities worldwide. Its ramifications extend beyond mere law enforcement, encompassing societal well-being, economic stability, and public safety. In urban landscapes like Los Angeles, the complexities of crime are particularly pronounced, necessitating nuanced approaches and evidence-based strategies for effective intervention.

The Los Angeles Police Department (LAPD) stands at the forefront of combating crime in the city, employing rigorous investigative techniques and proactive measures to maintain law and order. Central to their efforts is the collection and analysis of comprehensive data on reported crimes, which serves as a cornerstone for understanding the intricate dynamics of criminal activity within the city.

In this study, we embark on a journey to unravel the complexities of crime in Los Angeles by leveraging the rich dataset provided by the LAPD. Spanning the years 2020 to 2024, this dataset offers a wealth of information on reported crimes, including detailed records of crime occurrences, victim demographics, and spatial-temporal trends.

Our overarching objective is to gain a deeper understanding of crime dynamics in Los Angeles, with the ultimate goal of informing targeted interventions to address crime and enhance public safety. By delving into the intricacies of the dataset through exploratory data analysis (EDA), we aim to uncover hidden patterns, trends, and correlations that shed light on the underlying drivers of criminal behavior.

1. **Literature Survey**

A review of existing literature in the field of criminology reveals a wealth of research focused on understanding various aspects of crime, including its causes, patterns, and consequences. Scholars have employed diverse methodologies, ranging from statistical analyses to qualitative investigations, to explore the complex dynamics of criminal behavior. Key themes in this body of literature include the role of socio-economic factors, environmental influences, and demographic characteristics in shaping patterns of crime. Additionally, advancements in technology and data analytics have opened up new avenues for studying crime, such as spatial analysis and predictive modeling. Our study builds upon this foundation of knowledge to provide a nuanced understanding of crime patterns and victim demographics in Los Angeles.

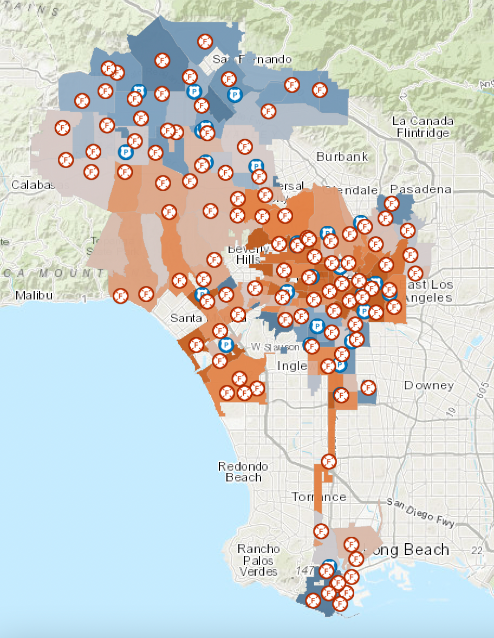


Fig:2.1 Los Angeles crime spots view

1. **Problem Statement**

The prevalence of crime in Los Angeles presents a significant challenge for law enforcement agencies and policymakers, who must allocate resources and devise effective strategies for crime prevention. To address this challenge, comprehensive analysis of crime data is essential to identify patterns, trends, and factors influencing crime occurrence and severity.

Additionally, understanding victim demographics is crucial for developing targeted interventions to support affected communities and enhance public safety.

Within this context, several research problems emerge. Firstly, there is a need to develop predictive models to anticipate the type of crime, such as theft or assault, based on factors like location and day of the week.

Secondly, exploring whether factors like the weapon used or victim demographics can predict the severity of a crime is imperative.

Furthermore, leveraging geo-spatial data, such as latitude and longitude, to identify areas with a higher likelihood of crime occurrence is essential for proactive intervention.

Additionally, analyzing historical data to uncover seasonal or temporal trends in crime patterns can provide valuable insights for strategic planning.

Also, a model to generate or predict the gender based on the independent variables which is related to it.

Lastly, exploring how predictive models can aid in optimizing resource allocation, such as patrols, based on predicted crime hotspots, is crucial for enhancing the efficiency of law enforcement efforts.

Therefore, addressing these research problems through data-driven analysis and modeling is paramount for improving public safety and crime prevention strategies in Los Angeles.

1. **Objective**

* To uncover insights into crime patterns, victim demographics, spatial-temporal trends, and factors influencing crime severity.
* Analyze crime trends over time: Identify seasonal variations (e.g., property crimes might increase during summer) or daily/weekly patterns.
* Explore spatial patterns: Investigate high-crime areas and identify potential factors (e.g., demographics, socio-economic indicators).
* Understand crime type relationships: Analyze how different crime types might be linked or occur together.

1. **Methodology**

**5.1 Data Collection**

* The dataset has 28 features and 925721 columns observation.

**5.1.1 Brief description of the dataset**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **Feature Name** | **Description** |
| 1 | DR\_NO | Unique Department Report Number assigned to each reported crime in Los Angeles. |
| 2 | Date Rptd | Date when the crime was reported to the police department. |
| 3 | DATE OCC | Actual date when the crime occurred. |
| 4 | TIME OCC | Time of the day when the crime occurred, in 24-hour format. |
| 5 | AREA | Numerical identifier for the community police station within Los Angeles (1-21). |
| 6 | AREA NAME | Name corresponding to the numerical area division. |
| 7 | Rprt Dist. No | Code indicating a sub-area within a geographic area, often prefixed by the area code. |
| 8 | Part code | Indicates the severity or type of crime. |
| 9 | Crm Cd | Numeric code representing the specific crime committed. |
| 10 | Crm Cd Desc | Description of the crime corresponding to its code. |
| 11 | Mocodes | Modus Operandi code providing additional crime details. |
| 12 | Vict Age | Age of the victim at the time of the crime. |
| 13 | Vict Sex | Gender of the victim (F: Female, M: Male, X: Unknown). |
| 14 | Vict Descent | Code representing the ethnic descent of the victim. |
| 15 | Premise Cd | Code indicating the type of structure where the crime occurred |
| 16 | Premise Desc | Description corresponding to the premise code. |
| 17 | Weapon Used Cd | Code representing the type of weapon used in the crime. |
| 18 | Weapon Desc | Description of the weapon used. |
| 19 | Status | Code indicating the status of the case. |
| 20 | Status Desc | Description of the case status. |
| 21 | Crm Cd 1 | Additional crime status codes associated with the case. |
| 22 | Crm Cd 2 | Additional crime status codes associated with the case. |
| 23 | Crm Cd 3 | Additional crime status codes associated with the case. |
| 24 | Crm Cd 4 | Additional crime status codes associated with the case. |
| 25 | LOCATION | Specific location where the crime occurred. |
| 26 | Cross Street | Cross street reference from the crime location |
| 27 | LAT | Latitude coordinate of the crime location. |
| 28 | LON | Longitude coordinate of the crime location |

**5.1.2 Datatype of each feature**

These are the feature names and their datatype.

|  |  |  |
| --- | --- | --- |
| **S.NO** | **Feature Name** | **DataType** |
| 1 | DR\_NO | Numerical variable |
| 2 | Date Rptd | Categorical variable |
| 3 | DATE OCC | Categorical variable |
| 4 | TIME OCC | Numerical variable |
| 5 | AREA | Numerical variable |
| 6 | AREA NAME | Categorical variable |
| 7 | Rprt Dist. No | Numerical variable |
| 8 | Part code | Numerical variable |
| 9 | Crm Cd | Numerical variable |
| 10 | Crm Cd Desc | Categorical variable |
| 11 | Mocodes | Categorical variable |
| 12 | Vict Age | Numerical variable |
| 13 | Vict Sex | Categorical variable |
| 14 | Vict Descent | Categorical variable |
| 15 | Premise Cd | Numerical variable |
| 16 | Premise Desc | Categorical variable |
| 17 | Weapon Used Cd | Numerical variable |
| 18 | Weapon Desc | Categorical variable |
| 19 | Status | Categorical variable |
| 20 | Status Desc | Categorical variable |
| 21 | Crm Cd 1 | Numerical variable |
| 22 | Crm Cd 2 | Numerical variable |
| 23 | Crm Cd 3 | Numerical variable |
| 24 | Crm Cd 4 | Numerical variable |
| 25 | LOCATION | Categorical variable |
| 26 | Cross Street | Categorical variable |
| 27 | LAT | Numerical variable |
| 28 | LON | Numerical variable |

* 1. **Importing dataset and required packages**

**Pandas, NumPy, Matplotlib, Seaborn**, and **Scikit-learn** are the libraries which are used for this dataset. The dataset is loaded using Panda’s library, NumPy library is used to handle numerical computing, Seaborn and Matplotlib is used to visualize the data distributions, trends, and relationships.

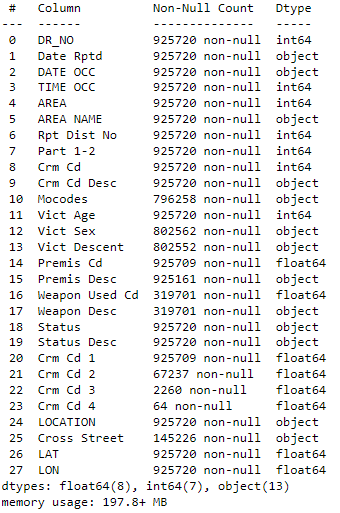


Fig 5.2 Columns with Null values

The above figure specifies that there are columns with null values.

Mocodes, Vict Sex, Vict Descent, Premis cd, Premis Desc, Weapon Used Cd, Weapon Desc, Crm cd 1, Crm cd 2,Crm cd 3, Crm cd 4, Cross Street features have null values. So, from this we can specify that there are Null values to handle the data.

**5.3 Data Cleaning**

* **Detecting Missing Values**

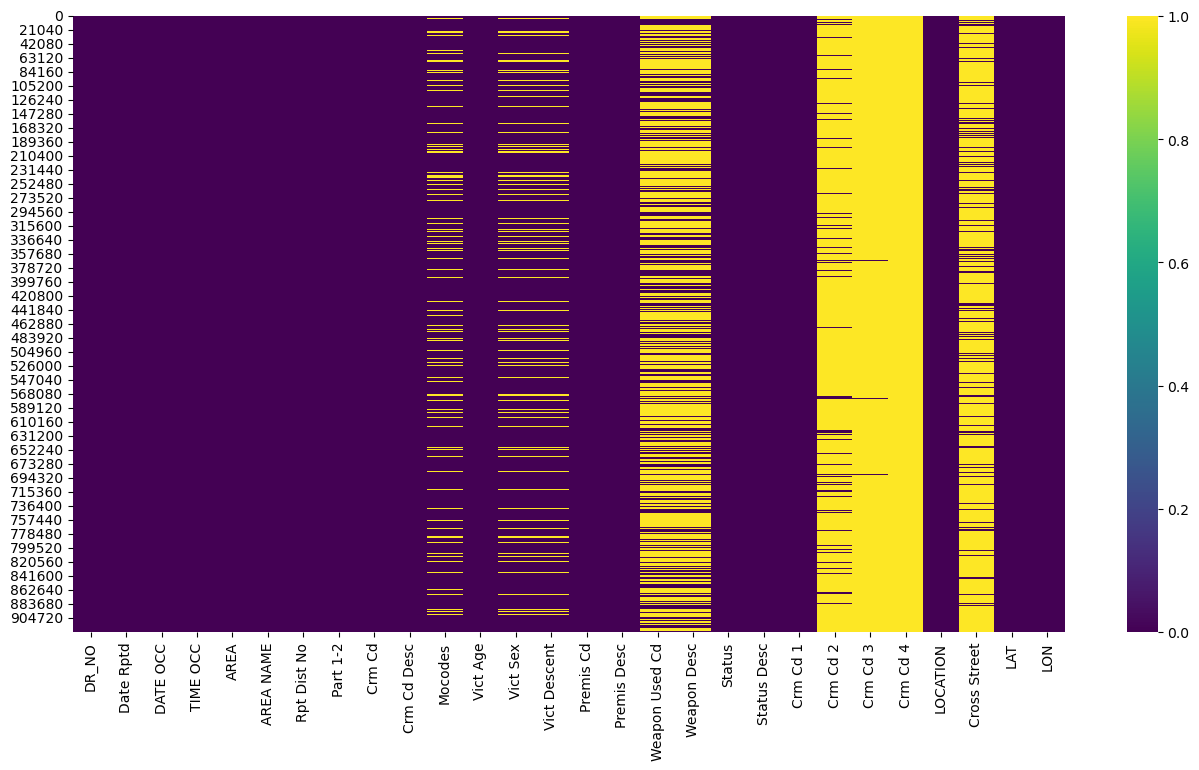


Fig 5.3.1 Visualization of Null values

The above figure shows that there are some missing values present in the dataset. The missing values are handled through diverse techniques such as the backfill method and zero insertion.

* 1. **Handling Missing values**

1. **Vict Age Feature**

Here we found out that the age column was recorded as 0, which indicates the potential missing data rather than an actual age value. So, we removed the values with 0 and replace it with null afterwards we used back fill to fill in the null values with the last observed age value.

1. **Vict Sex Feature**

Here for the missing values, it is replaced with x which specifies that sex of victim is unknown.

1. **Vict Descent**

Here for the missing values, it is replaced with x which specifies that descent of victim is unknown.

1. **Crm Cd 2**

For this feature the null values are replaced with zero (0) because it is referred as a additional status code so we kept zero meaning no additional crime code.

1. **Mocodes**

Here, the code first removes any extra whitespace from the 'Mocodes' value and then splits it into a list. Each value in the list is then formatted as a time string in the HH:MM format. If any null values are encountered during this process, they are replaced with the corresponding value from the 'TIME OCC' column.

1. **Cross street**

The values here are concatenated to the feature Location and the column was dropped.

1. The features like **Crm cd 1 ,Premis cd, Premis Desc** have every low null values so we the row which have null for these columns.
2. **Crm Cd 3, Crm Cd 4** are dropped from the dataset.
3. The **Weapon Used cd** and **Weapon desc** features have null values so we replaced with the values like WEAPON NOT MENTIONED for weapon desc and 0 for Weapon used cd.

After handling the null values with this technique, we found there is no null values in the dataset.

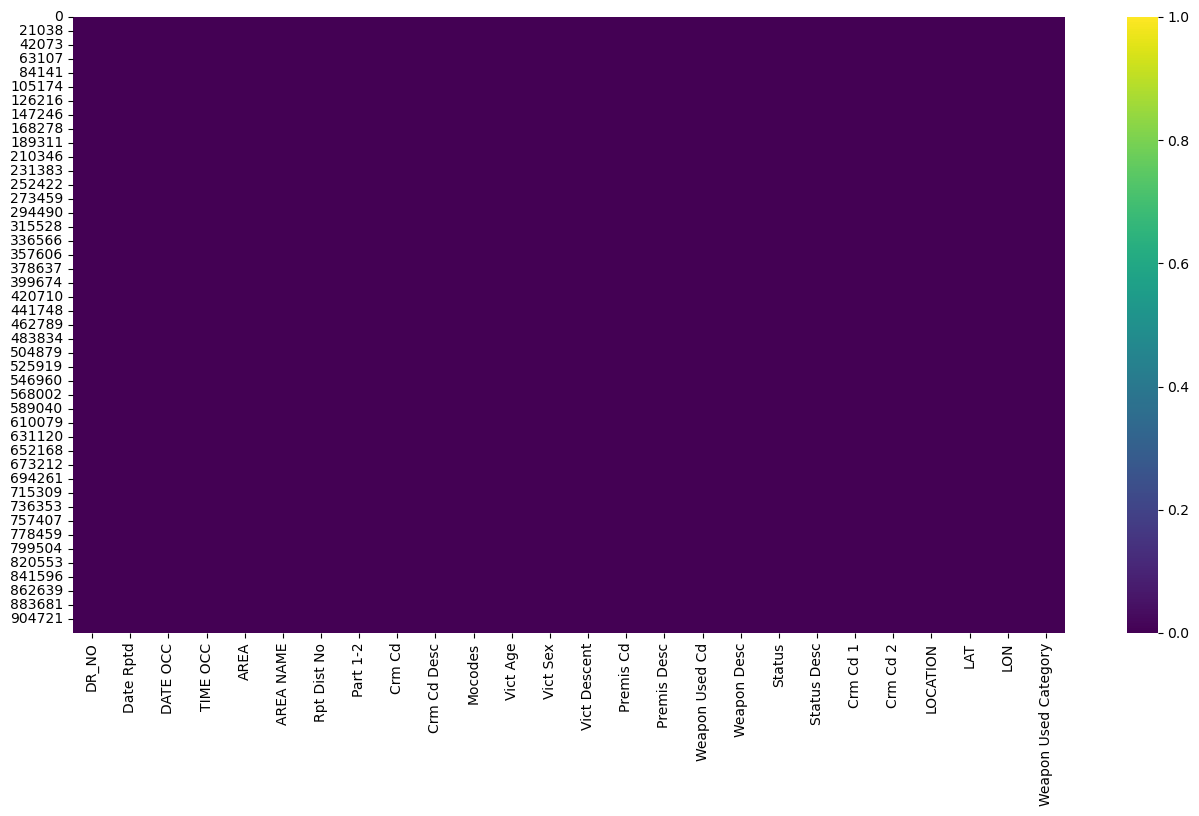


Fig 5.3.2 Visualization of Non Null values

The above figure shows that there are no null values or missing values in any of the feature.

**5.5 Outlier detection**

In the dataset there are some features which have outliers. These outliers are identified by a commonly used technique called Interquartile Range (IQR) Method. This IQR method identifies outliers based on the interquartile range, which is the difference between the third quartile (Q3) and the first quartile (Q1). Outliers are defined as data points that fall below

Q1 – 1.5 \* IQR or above Q3 + 1.5 \* IQR.

The outliers are represented visually using box plots which are shown below:

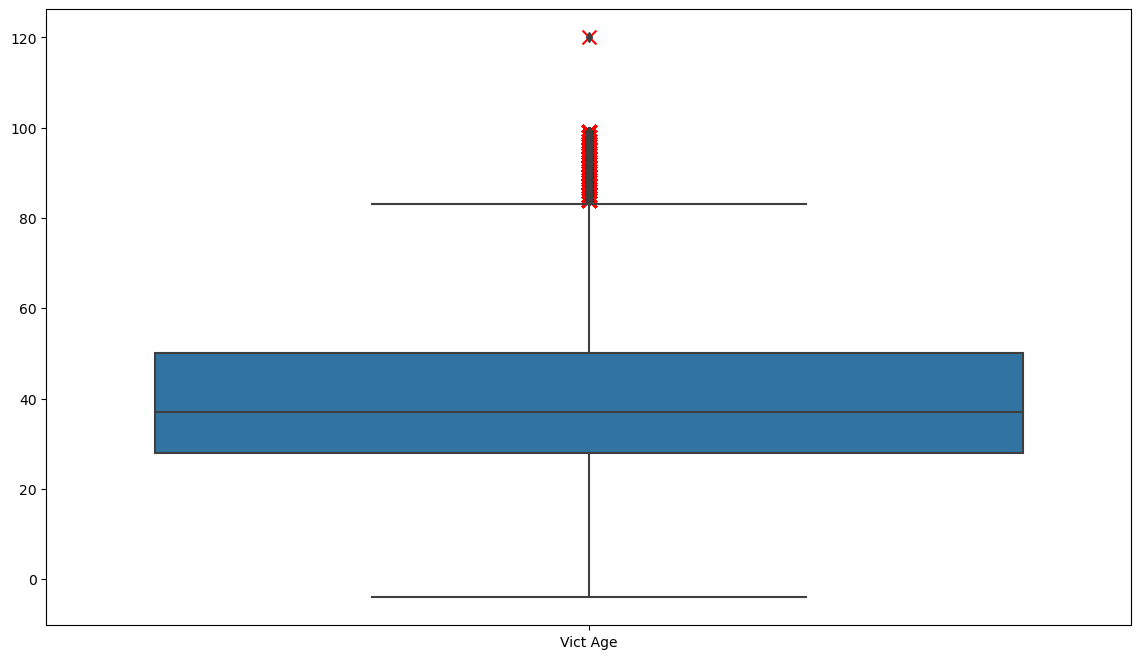
****

Fig 5.5.1 Visualizing outliers using box plot for Vict Age

The Vict Age feature have values above 120 and below 0 which are identified as outliers.



Fig 5.5.2 Visualizing outliers using box plot for LAT

The LAT feature has values 0 which are identified as outliers.

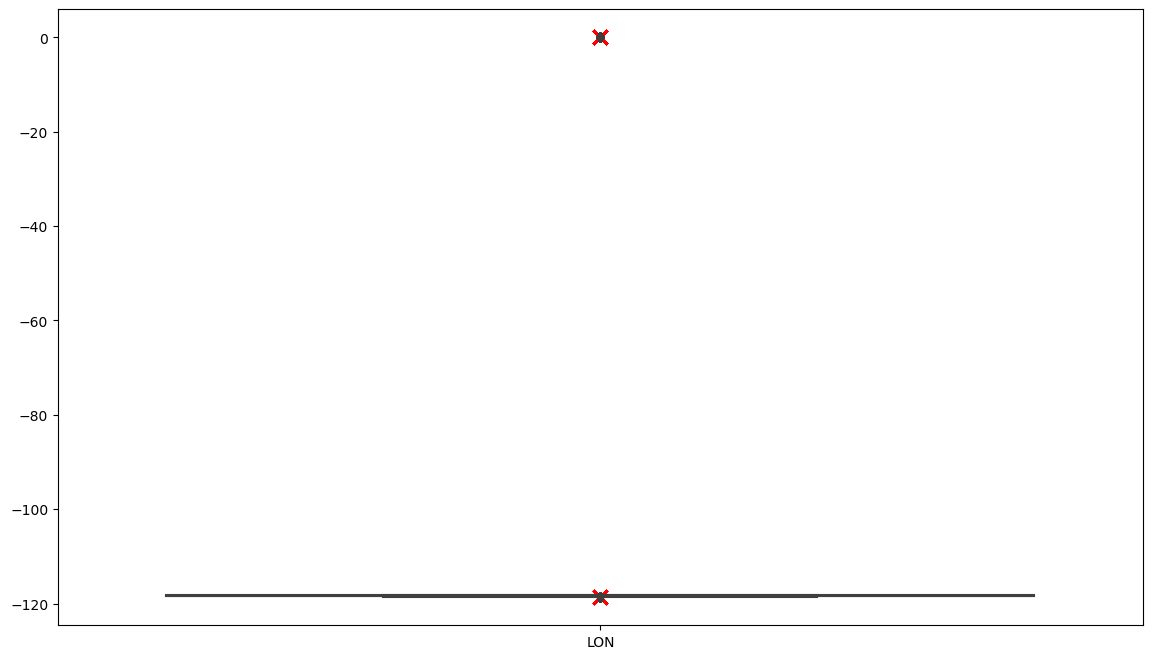
****

Fig 5.5.3 Visualizing outliers using box plot for LON

The LON feature has values 0 which are identified as outliers.

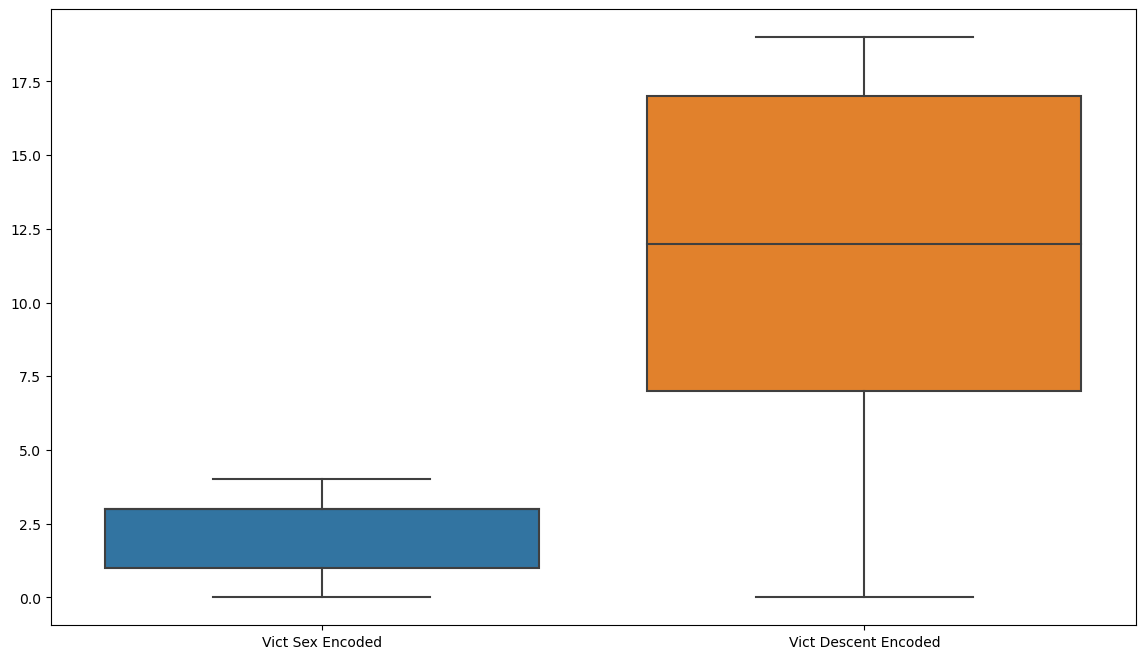
****

Fig 5.5.4 Visualizing outliers using box plot for Vict sex and descent

The Vict Sex and Vict Descent feature have values like ‘- ‘and some unidentified values which are identified as outliers.

**5.6 Outlier Handling**

1. **Vict Age**

Here the identified outliers with respect to its row have been dropped from the dataset.

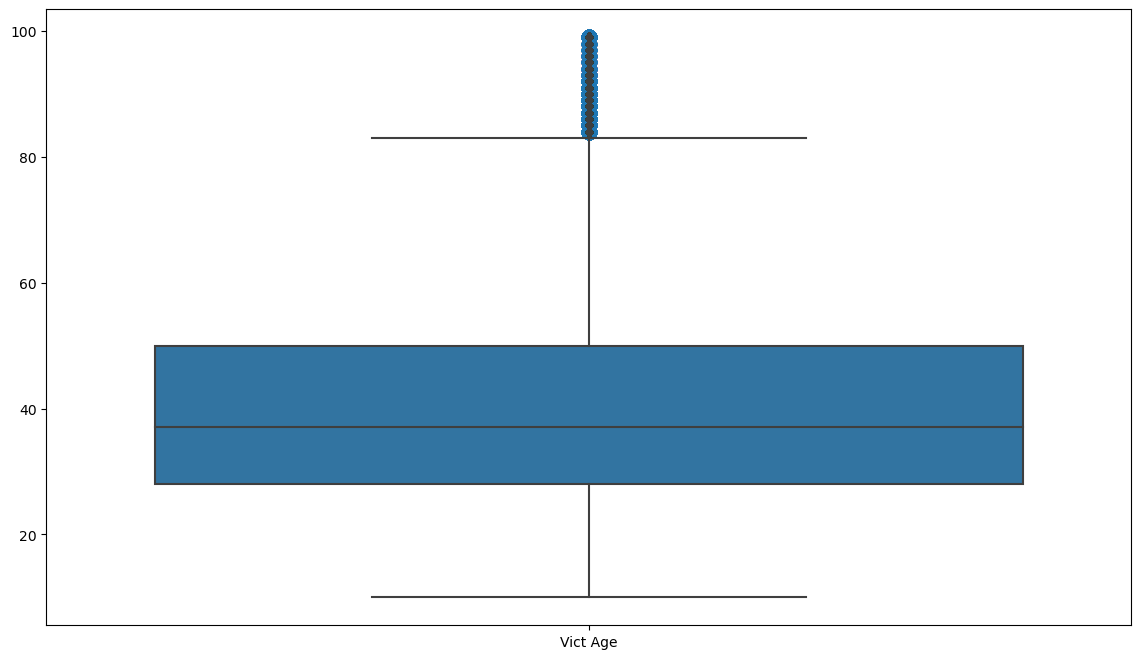


Fig 5.6.1 Visualization of outlier handling for Vict Age

1. **LAT**

Here the identified outliers which have value 0 with respect to its row have been dropped from the dataset.

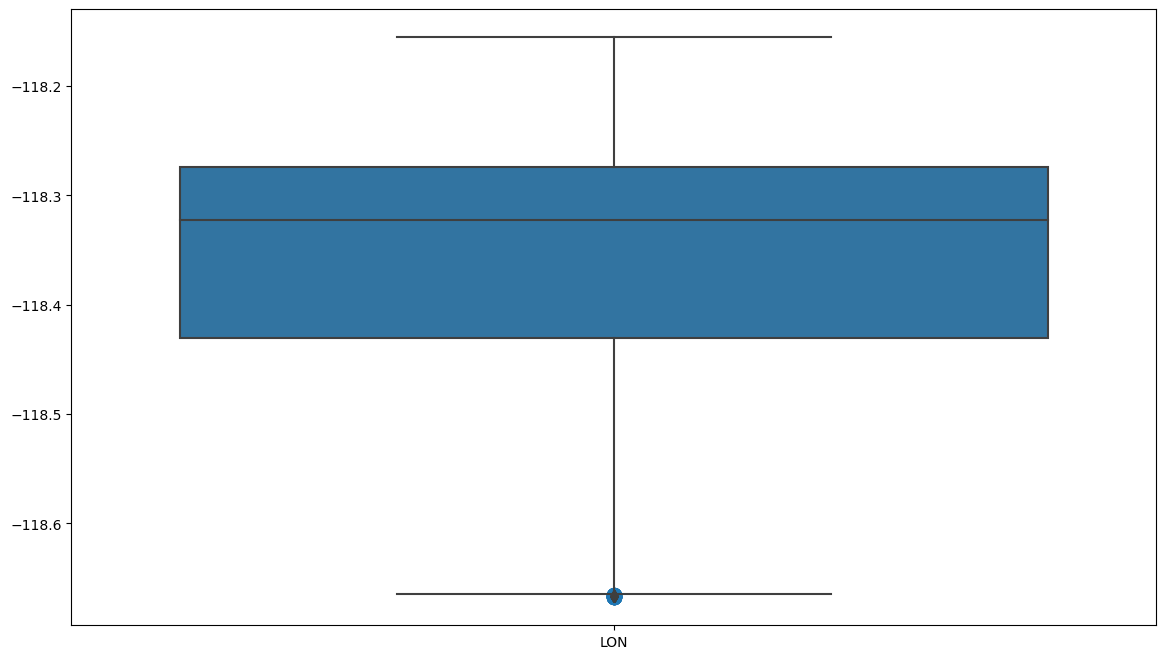


Fig 5.6.2 Visualization of outlier handling for LAT

1. **LON**

Here the identified outliers which have value 0 with respect to its row have been dropped from the dataset.

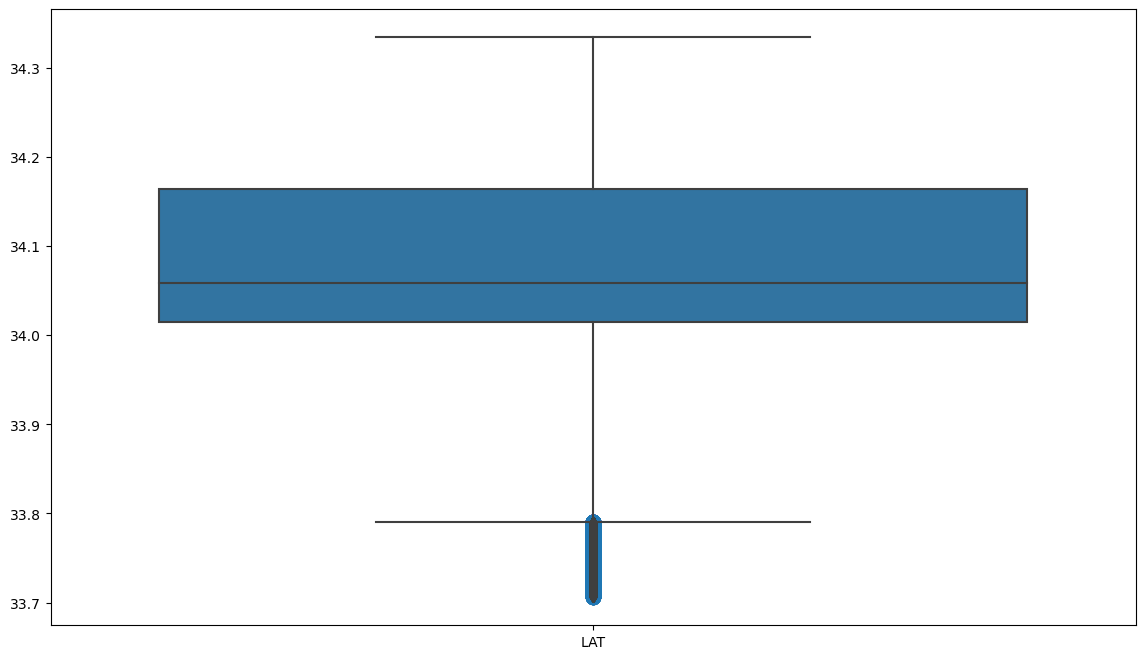


Fig 5.6.3 Visualization of outlier handling for LON

1. **Vict Sex and Vict Descent**

Here the identified outliers which have value which is not considerable is dropped from the dataset.

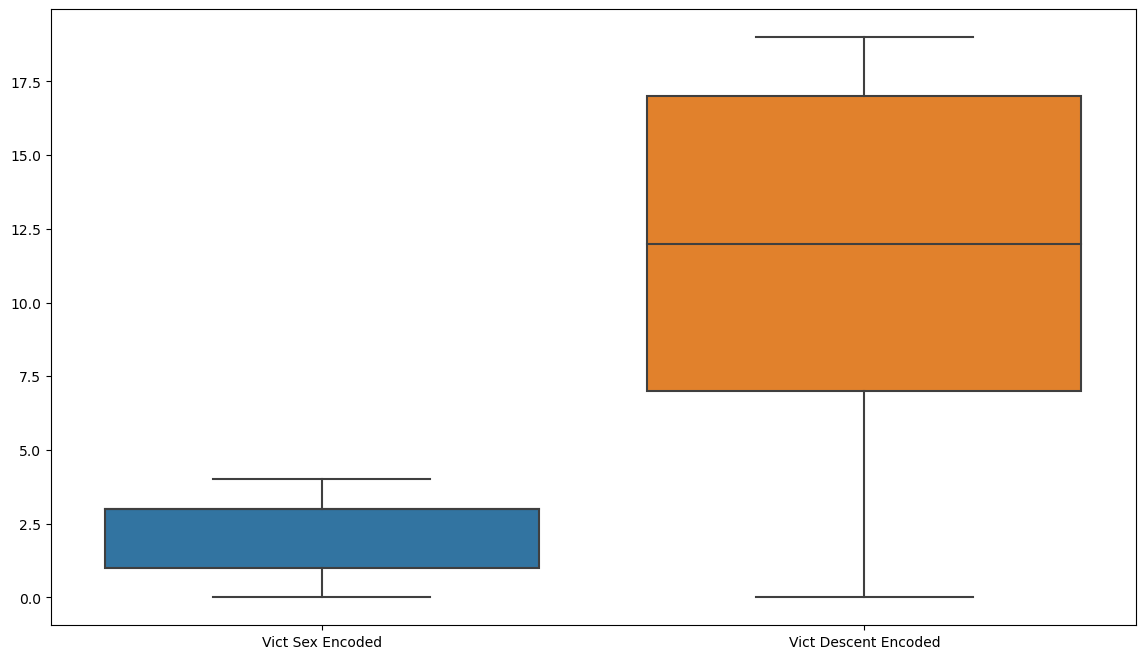
****

Fig 5.6.4 Visualization of outlier handling for Vict sex and Vict descent

**5.7 EXPLORATORY DATA ANALYSIS**

**5.7.1 Data Preprocessing**

* **Label Encoding:**

Label Encoding is a technique used in data preprocessing. Label Encoding is a technique used to convert categorical features into numerical values. It assigns a unique integer to each category in the feature.

* **Converting the time feature to datetime datatype**

Date Rptd, DATE OCC are in object datatype which are convert to datetime.

**5.7.2 Feature Engineering**

* **Date Based Features**

DATE OCC Year, DATE OCC Month, DATE OCC Week, DATE OCC Day, DATE OCC Day of Week are the new features which are derived from the **DATE OCC** feature.

1. DATE OCC Year: This feature represents the year component extracted from the DATE OCC feature. It allows the model to capture any yearly trends in the data.

2. DATE OCC Month: This feature represents the month component extracted from the DATE OCC feature. It enables the model to capture any monthly patterns in the data.

3. DATE OCC Week: This feature represents the week number within the year extracted from the DATE OCC feature. It helps in capturing weekly patterns in the data.

4. DATE OCC Day: This feature represents the day of the month extracted from the DATE OCC feature. It can capture any variations or patterns specific to certain days of the month.

5. DATE OCC Day of Week: This feature represents the day of the week (e.g., Monday, Tuesday) extracted from the DATE OCC feature.

* **Time Based Features**

Hour, Time of Day are the new features which are derived from the **TIME OCC** feature.

1. Hour: This feature represents the hour component extracted from the TIME OCC feature.

2. Time of Day: This feature categorizes the time of day into segments such as morning, afternoon, evening, or night.

* **Season Based Feature**

Season feature categorizes the time of years or months into segments such as summer, winter, fall and rainy.

* **Holiday Feature**

BY using the DATE OCC Day of Week, we derived is crime happened on holiday or not.

* **Extra Features**

1. **Weapon used category**

To simplify the analysis of the "Weapon Used" feature, which contains numerous unique values, we categorized them into nine distinct categories based on their characteristics:

1. Guns: This category includes all weapons related to firearms, such as pistols, rifles, shotguns, etc.
2. Physical Attack: Weapons used in hand-to-hand combat or physical assault fall under this category. Examples include punches, kicks, slaps, and other forms of physical violence.
3. Blade Weapons: This category encompasses weapons with blades, such as knives, swords, machetes, and other sharp-edged instruments.
4. Firearms/Explosives: Weapons that involve the use of explosives or firearms to cause harm are categorized here. This includes bombs, grenades, and other explosive devices.

5. Sharp Objects: Weapons like blades, scissors, or any sharp-edged objects that can cause injury fall into this category.

6. Verbal Threat: Instances where the perpetrator uses verbal threats or intimidation without physical weapons are categorized as verbal threats.

7. Chemical/Poison: This category includes incidents involving the use of chemical substances or poisons to harm individuals.

8. Miscellaneous/Mixed: Incidents where the weapon used does not fit into any of the above categories or involves a combination of different types of weapons are classified as miscellaneous or mixed.

1. Throwable Objects: Weapons that are intended to be thrown, such as rocks, stones, or projectiles, fall into this category.

By categorizing the weapons into these nine distinct groups, we can simplify the analysis and gain insights into the types of weapons most commonly used in incidents. This classification enables a clearer understanding of the nature of the crimes and can aid in developing strategies for crime prevention and law enforcement.

1. **Reporting Lag**

This is derived based on the Date of occurred and Date of reported to determine whether crimes are reported promptly after they occur represents the time elapsed between the occurrence of the crime and its reporting.

These features which are encoded, derived feature and datatype convert features helps to build different models and predict the outcomes.

1. **Crime type**

In our dataset, we encountered 139 unique descriptions for crimes (CRM CD DESC) where finding or predicting the specific crime type is particularly challenging. To address this issue, we categorized these descriptions into 14 different crime types:

1. Theft and Property Crime: Involves the unlawful taking of someone else's property without their permission.
2. Burglary: The act of unlawfully entering a building or structure with the intent to commit a theft, property damage, or other felony.
3. Battery: Refers to the intentional and unlawful use of force or violence against another person.
4. Sexual Offenses: Crimes of a sexual nature, including sexual assault, rape, and other forms of non-consensual sexual activity.
5. Assault: Involves the threat or attempt to inflict physical harm on another person, causing them to fear imminent bodily harm.
6. Domestic Violence/Child Abuse: Acts of violence or abuse that occur within familial or intimate relationships, including violence against children.
7. Threats: Expressions of intent to cause harm or to engage in unlawful activities that instill fear in others.
8. Miscellaneous Crimes: Other criminal activities that do not fit neatly into specific categories, such as public intoxication or disorderly conduct.
9. Robbery: The unlawful taking of property from a person or place through the use of force or intimidation.
10. Fraud: Deceptive practices or schemes intended to deceive others for financial gain, often involving false representations or omissions of information.
11. Legal Order Violations: Breaches of court orders or legal injunctions, including violations of restraining orders or probation terms.
12. Vandalism: The wilful destruction or defacement of property belonging to others.
13. Weapon Offenses: Crimes involving the unlawful possession, use, or trafficking of weapons such as firearms, knives, or other dangerous instruments.
14. Property Damage: Actions resulting in the destruction or impairment of property, excluding intentional acts such as vandalism.

At last, we left with 916994 observations and 44 features which includes encoded, derived or changed datatype.

**5.7.3 Visualisations**

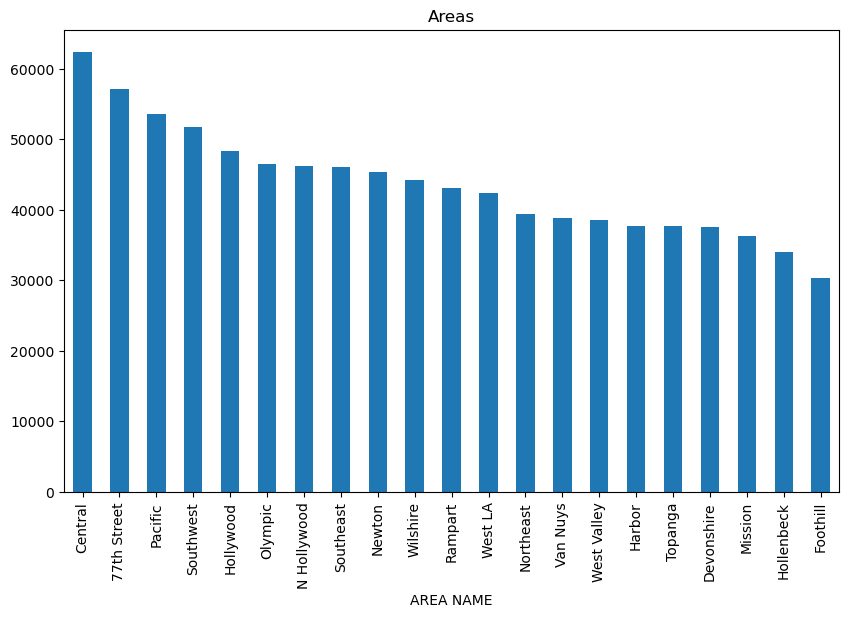


Fig 5.7.1 bar graphs for Area Nmae

The above figure shows the number of crimes happened according to the area wise.

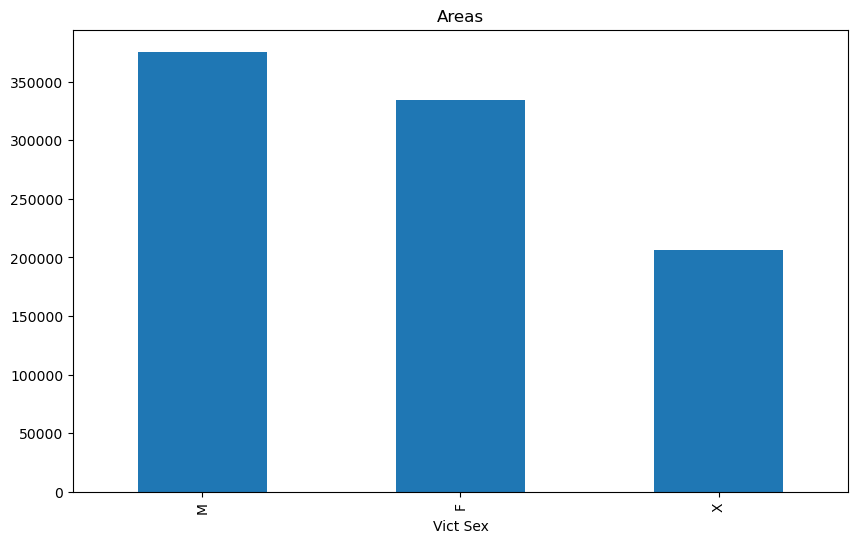


Fig 5.7.2 bar graphs for vict sex

The above figure shows the number of crimes happened according to the gender wise category where m is male, f is female, and x is unknown.

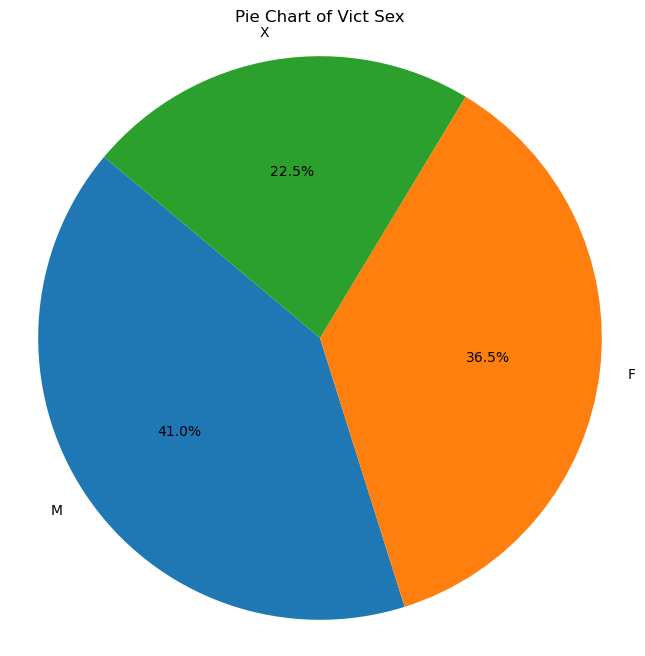


Fig 5.7.3 pie chart for Vict sex

The above figure specifies the gender-wise ratio percentage of crime happened to each gender.

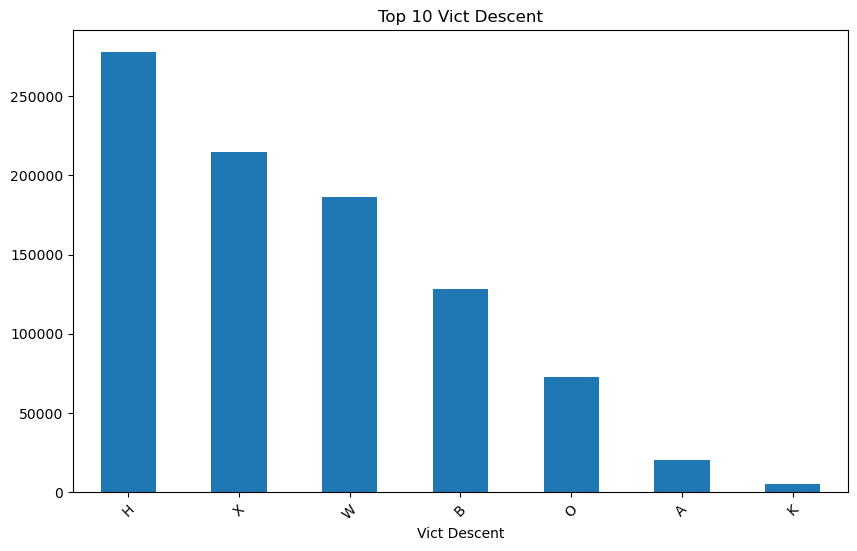


Fig 5.7.4 bar graph for top 10 vict descent

The above figure shows the top 10 Victims based on descent who was mostly affected.

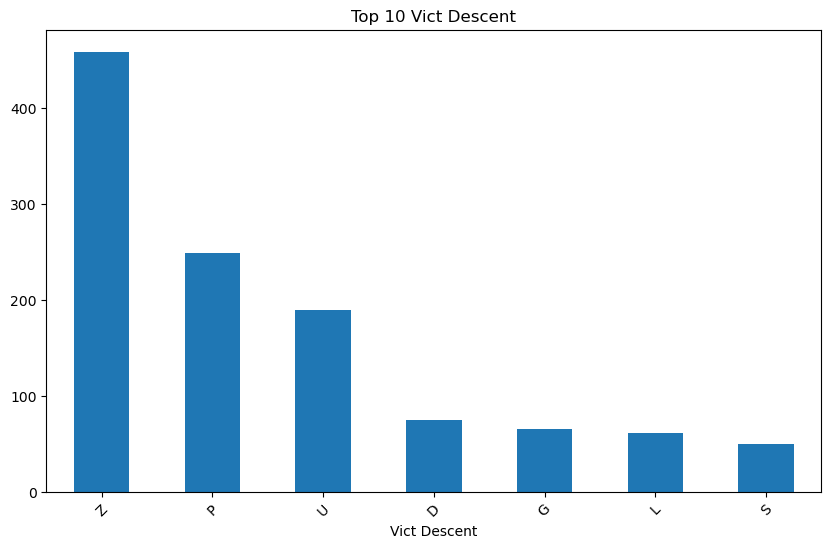


Fig 5.7.5 bar graph for bottom 10 vict descent

The above figure shows the top 10 Victims based on descent who was not affected badly.

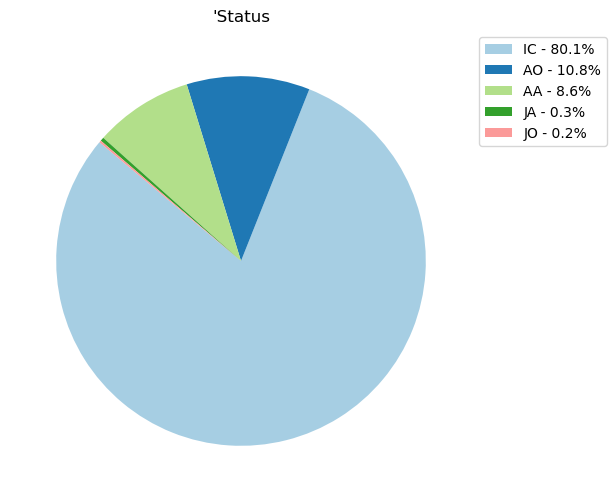


Fig 5.7.6 pie chart for status

The above pie shows the status of the crime happened from 2020 to 2024 where it shows that have 80 % of crimes are in Ic(Invest Cont) meaning Investigation is continuing till now from the date it happened.

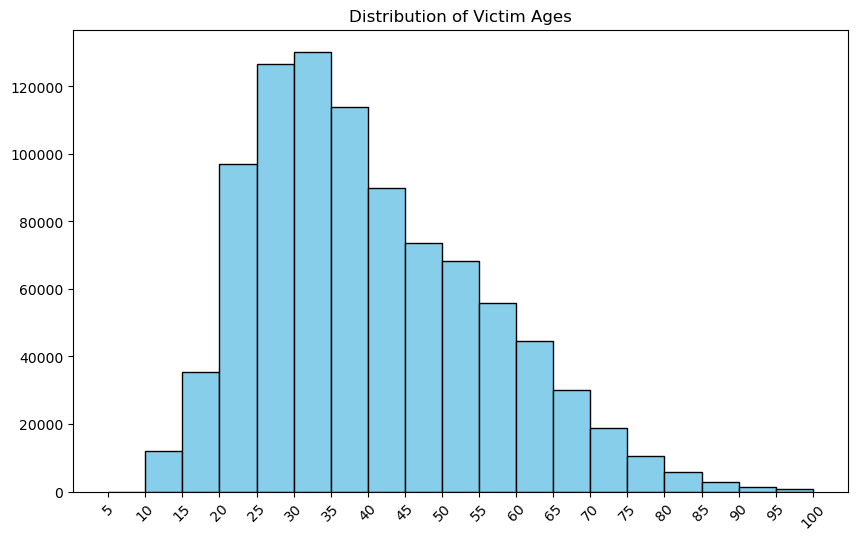


Fig 5.7.7 Histogram for Vict age

The above plots show the how different age people where effect with number of crimes.

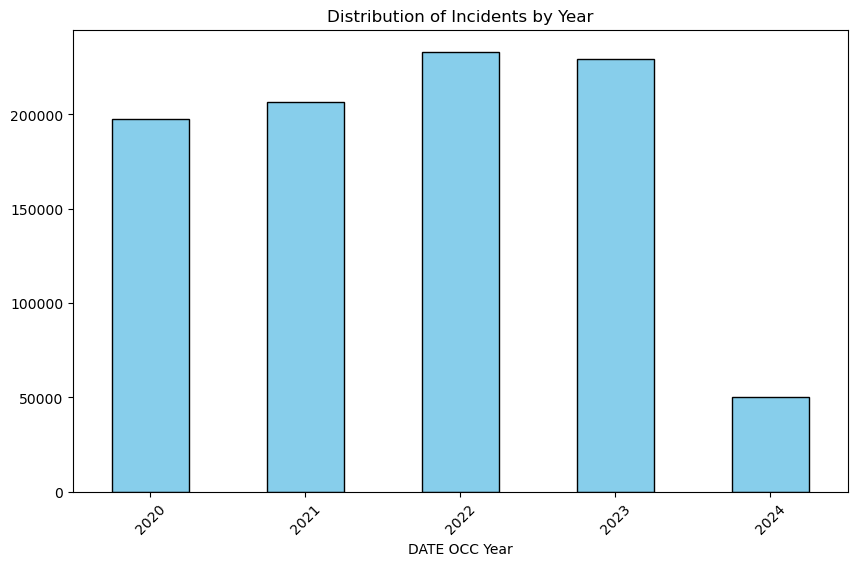


Fig 5.7.8 bar graph of distribution of incidents by year

The above image shows how crime has increased over the years passed.

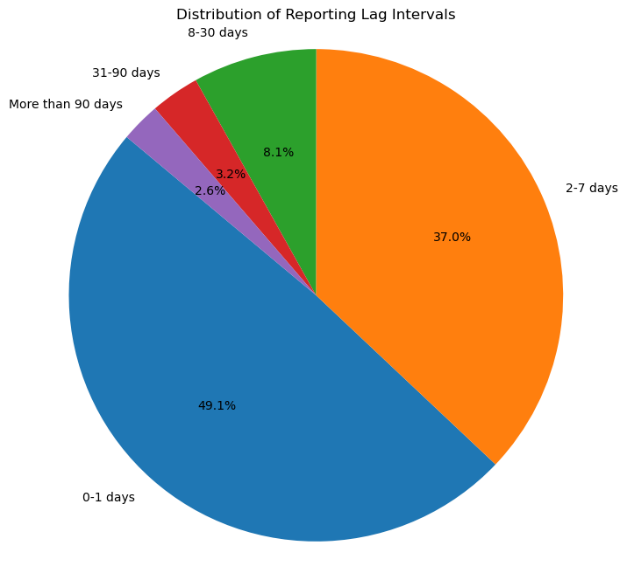


Fig 5.7.9 pie chart for report lag

The above pie shows the Lag taken to report to the station after the incident happened. There is immediate reporting in 0-1 days after the incident is taken by 49.1% people in the city. As well as there is also 37% in between 2-7 days.

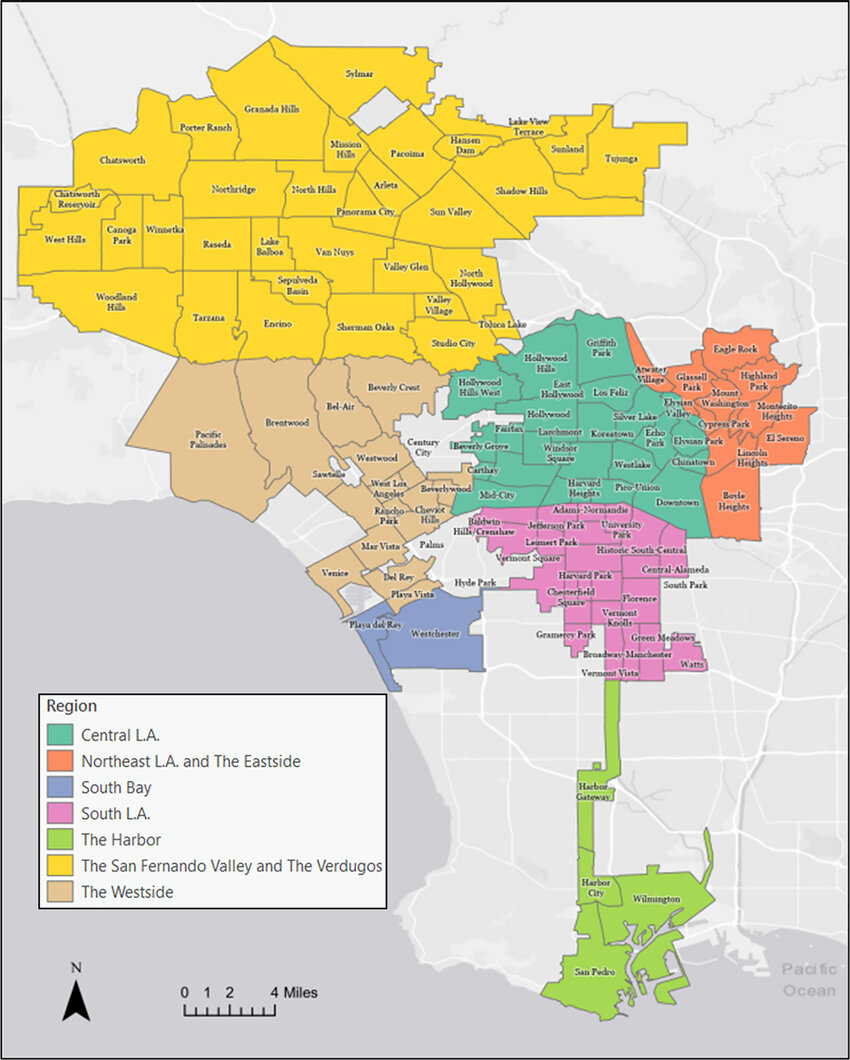


Fig 5.7.10 Map view of Los Angeles(source=google)

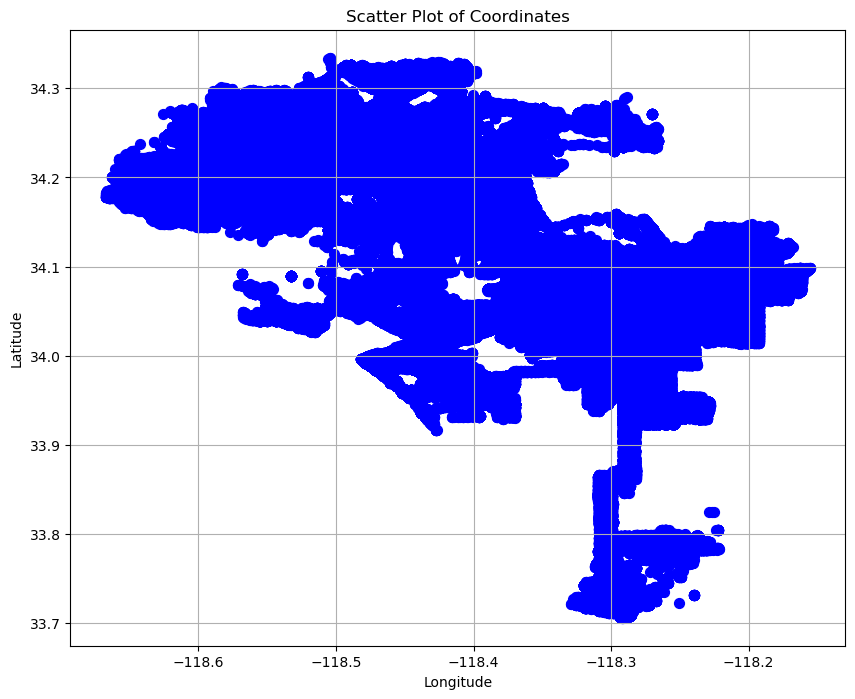


Fig 5.7.11 Scatter plot Map view of Los Angeles

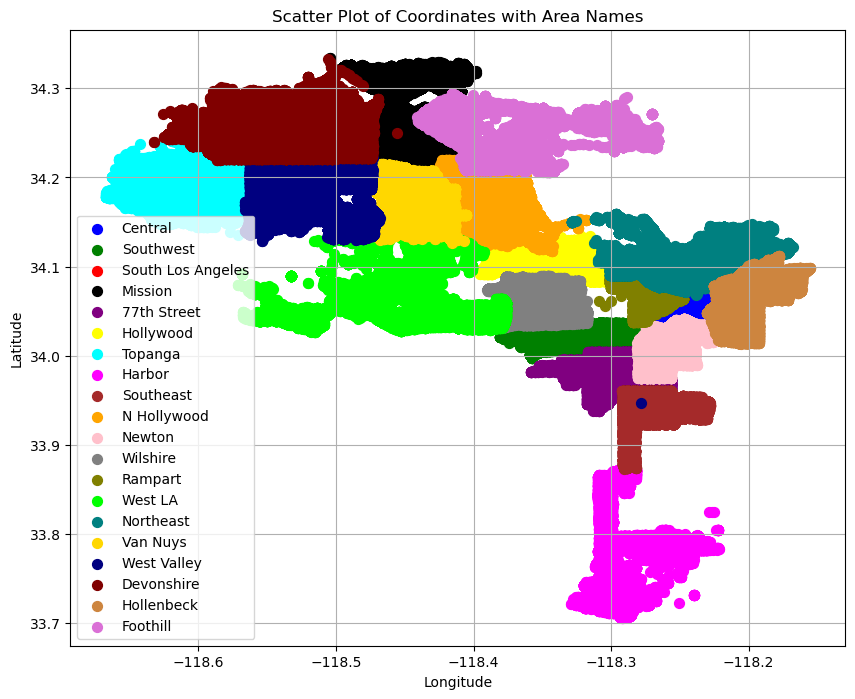
The above image shows the real time city map and the below ones shows the area where crime has happened. This is shown visually through scatter plots. So, we can identify that the almost all areas represent the widespread impact of crime.

Fig 5.7.12 Area wise divided scatter plot map view

The above image shows the areas where crime occurred over 4 years. It is showing different colours for different areas.



Fig 5.7.13 bar plot for holidays

The figures specifies that the non-holidays are most impact to spread crime rather than on holidays.

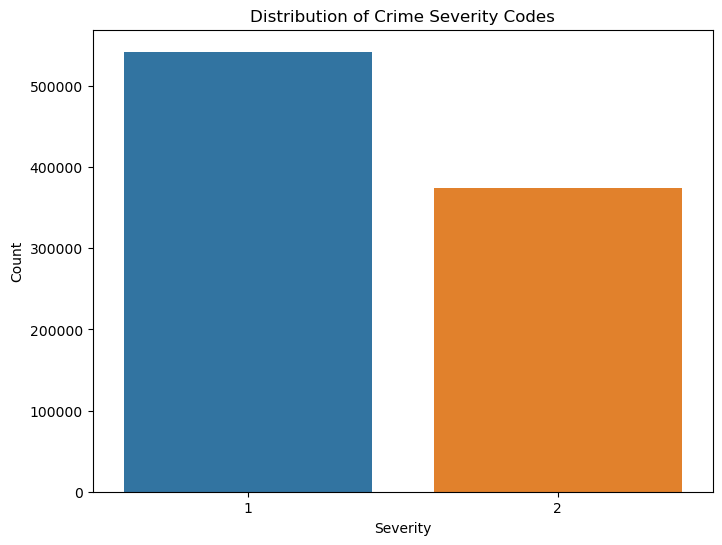


Fig 5.7.14 Severity plots

The figure shows that over 500000 cases are not such great crime type and less than that have high severity.

**5.8 Research Problems**

**5.8.1 Crime Severity Prediction**

We have a feature Part 1-2 which specifies the crime severity basically. We have encoded the numerical to categorially to train using classification.

For 1 it is named as Misdemeanor and for 0 it is named as Felony.

"Misdemeanor" typically refers to a less serious type of criminal offense compared to a felony. In legal terms, misdemeanors are offenses that are punishable by fines, community service, probation, or a jail sentence of less than one year.

"Felony" refers to a more serious category of criminal offense compared to a misdemeanor. Felonies are typically crimes that carry heavier penalties, including imprisonment for more than one year, fines, or even capital punishment in some jurisdictions.

We have trained classification and regression models but we got best metrics for classification.

In classification we have trained a random forest classifier with independent variables of 'Weapon Used Category, Vict Age, Vict Sex, AREA, Vict Descent, Season, DATE OCC Day of Week and dependent variable is Severity that is Part 1-2 feature.

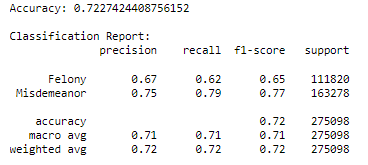


Fig 5.8.1 Metrices for Crime Severity Prediction

As the model is trained, the trained model is saved into a single pickle file, enabling their reuse for future predictions without the need for retraining and we have developed a web application to find or predict based on given values.

* **Web application for Crime Severity Prediction**

The Web application is developed using HTML (Hyper Test Markup Language) and CSS (Cascading Style sheet). The web application was developed to predict Severity of crime based on the inputs given. It allows users to input various parameters to find how much it has affected.

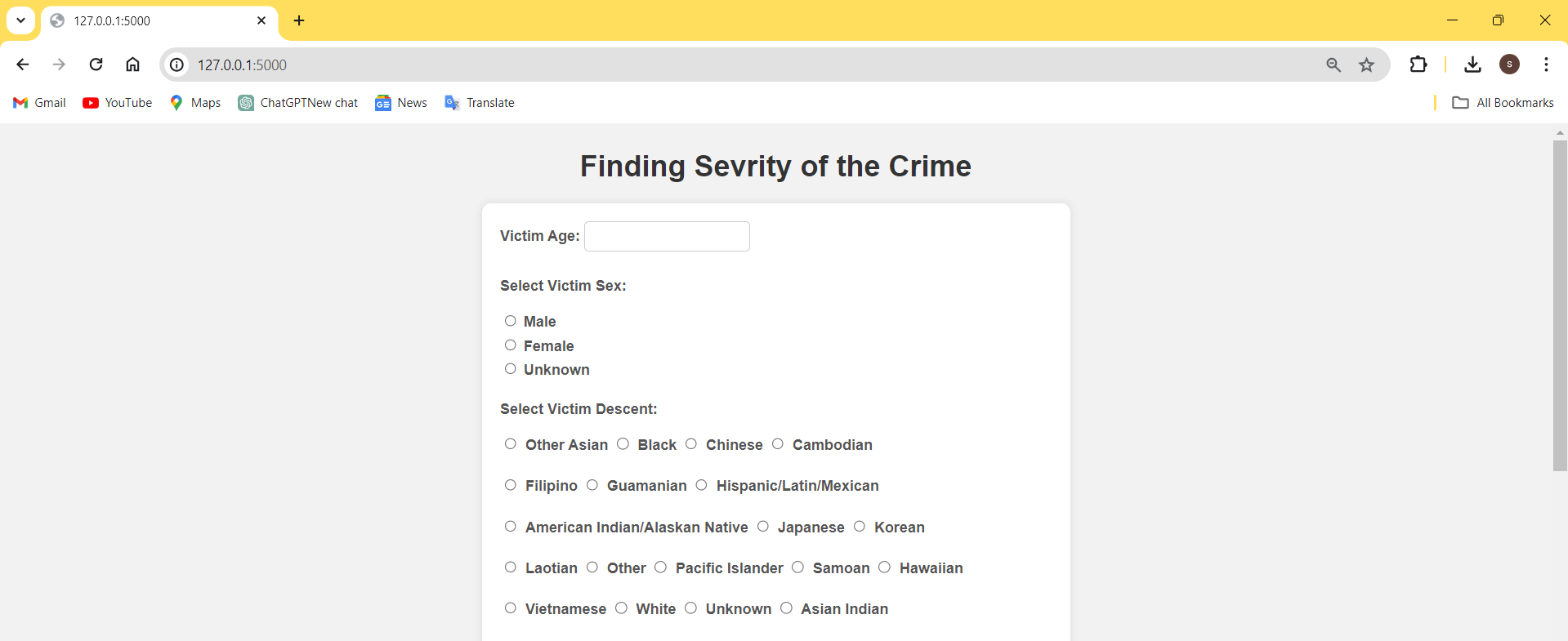


Fig 5.8.2 Web screen 1 for Crime Severity Prediction

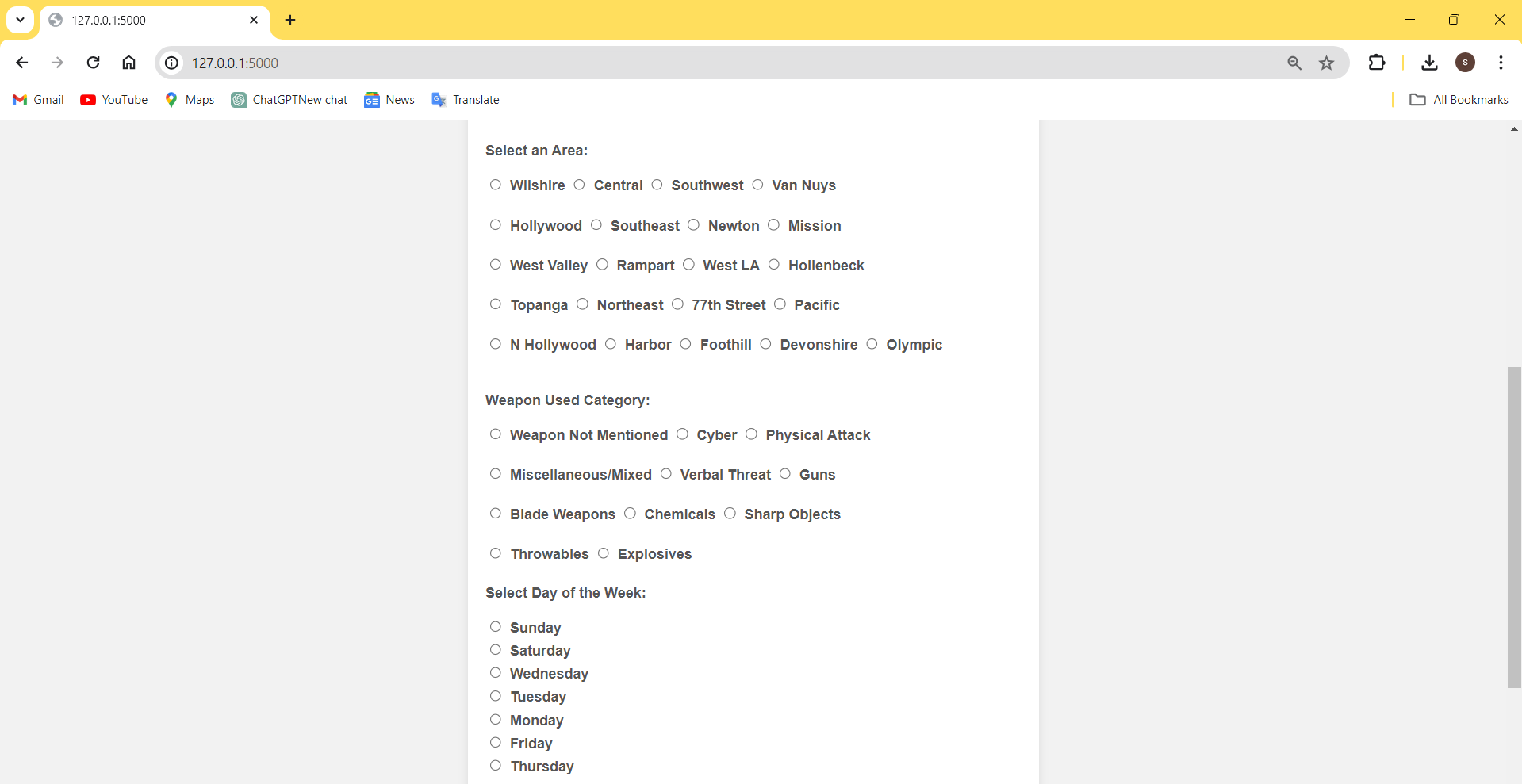
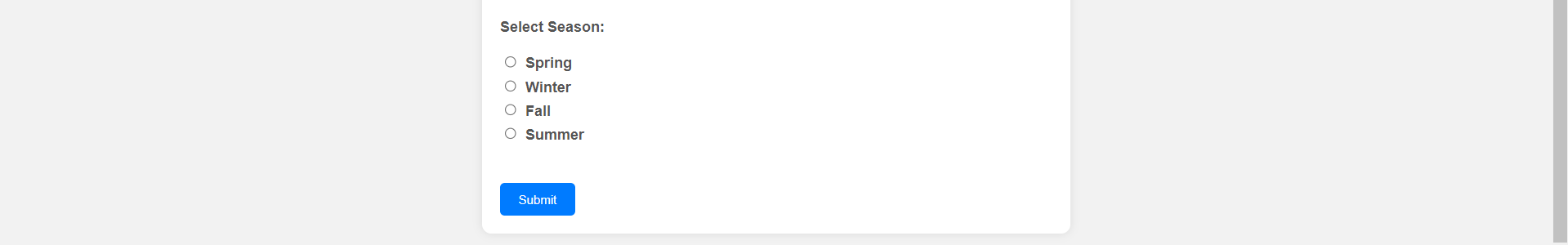


Fig 5.8.3 web screen 2 for Crime Severity Prediction

Based on the above input parameters the model gets trained and predict the result.

* + 1. **Predicting Victim Gender**

Here we are going to predict the gender based on model trained on historical data. Here the model is trained on random forest classifier on independent variables such as AREA, Vict Descent, Hour, Time of Day, Part 1-2, On Holidays, Vict Age, Season, DATE OCC Day of Week and dependent variable Vict Gender.

Here the model will predict either Male or Female gender which may affect based on the trained input parameters.

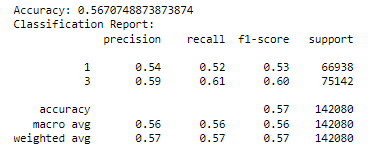


Fig 5.8.4 Metrices for Predicting Victim Gender

Here the accuracy is low because there is no gender which can identified as unknown so we excluded unknown values while training which effect the accuracy. But the f1 score shows that the model will not skew on one side and predict only one value rather it predicts both based on inputs.

As the model is trained, the trained model is saved into a single pickle file, enabling their reuse for future predictions without the need for retraining and we have developed a web application to find or predict based on given values.

* **Web application for Predicting Victim Gender**

The Web application is developed using HTML (Hyper Test Markup Language) and CSS (Cascading Style sheet). The web application was developed to predict Victim Gender based on the inputs given. It allows users to input various parameters to find how much it has affected.

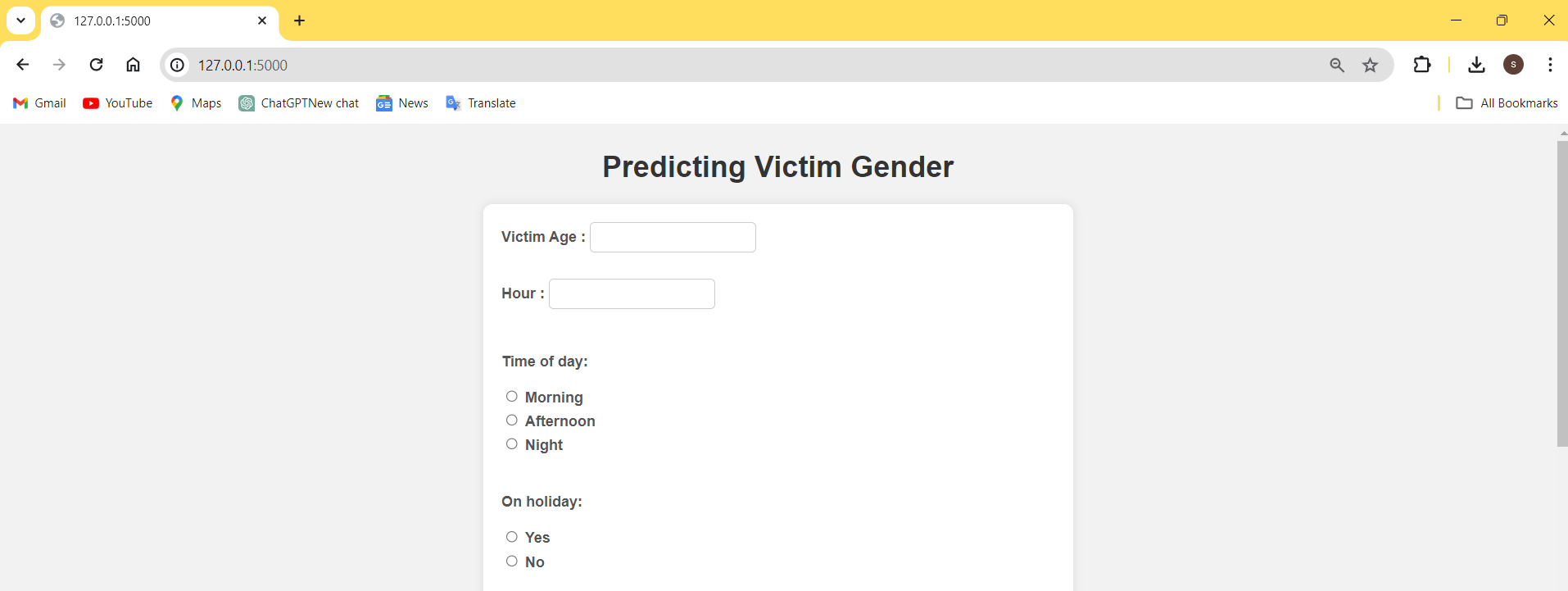


Fig 5.8.5 Web screen 1 for Predicting Victim Gender

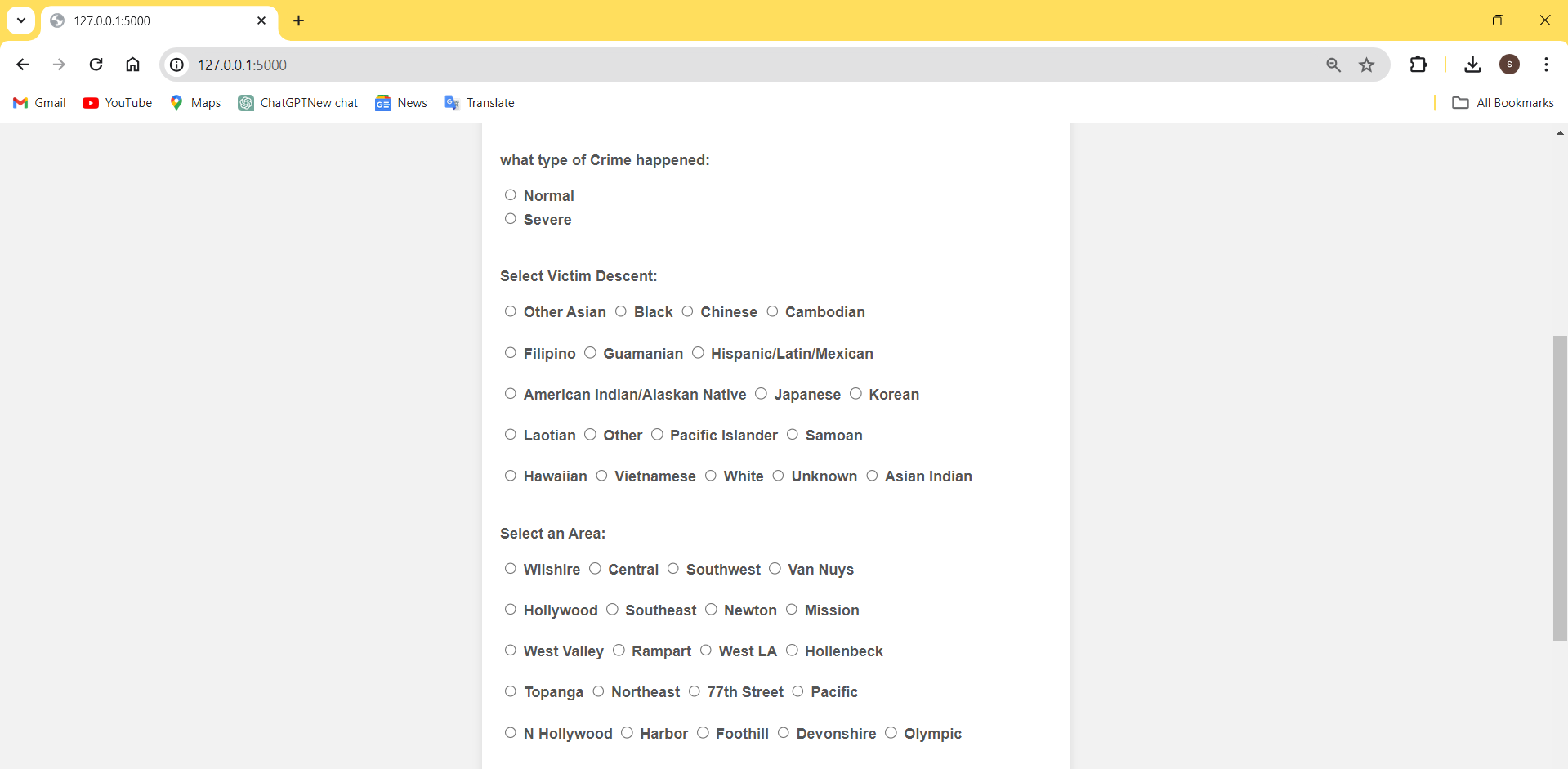
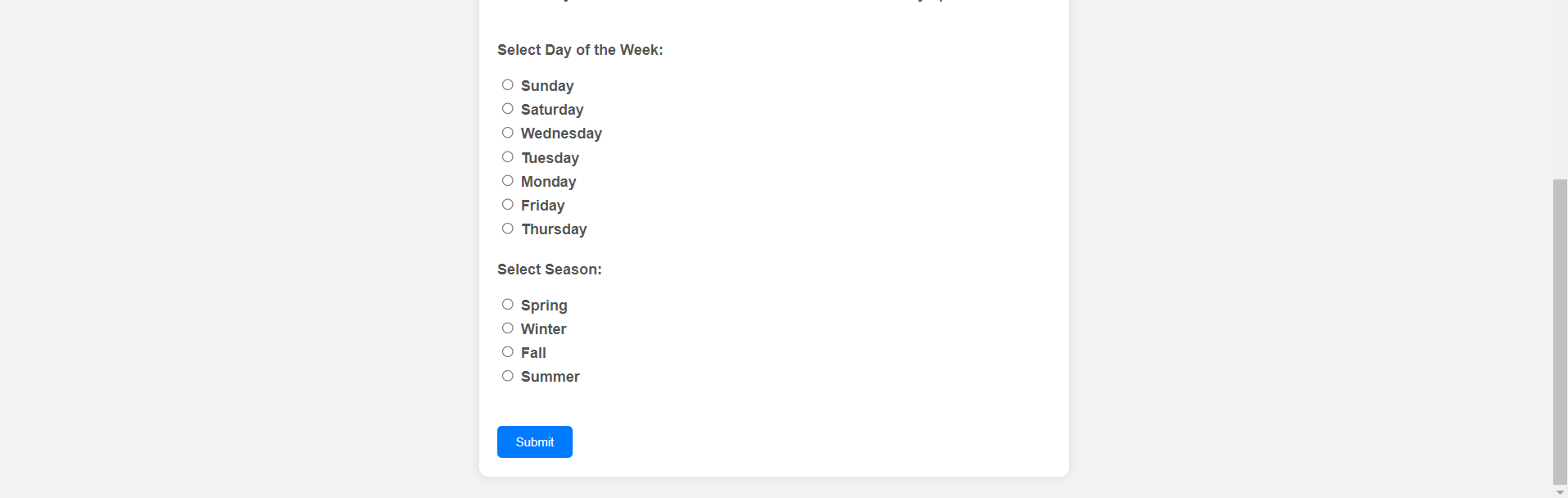


Fig 5.8.6 Web screen 2 for Predicting Victim Gender

Based on the above input parameters the model gets trained and predict the result.

* + 1. **Crime Trend Analysis and Resource Allocation Optimization**

As we derived new features from Time and date-based features. We have plotted some graphs and trends to see any patterns occurs.

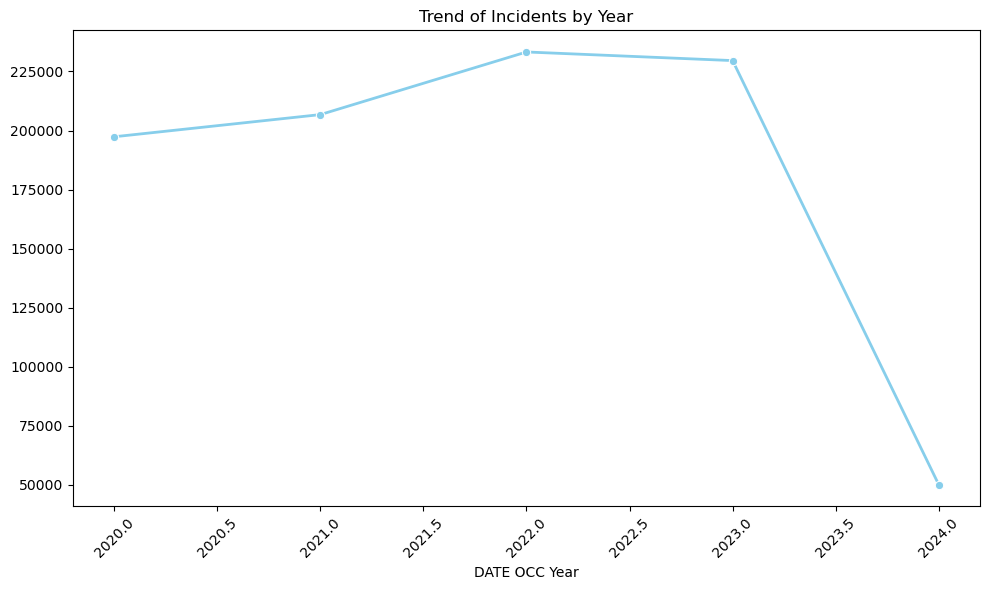


Fig 5.8.7 Trend of incident by year

The above figure shows how the crime rate increases from 2020 to 2023. The sudden fall is due to the dataset contains only 3 months of 2024 year data.

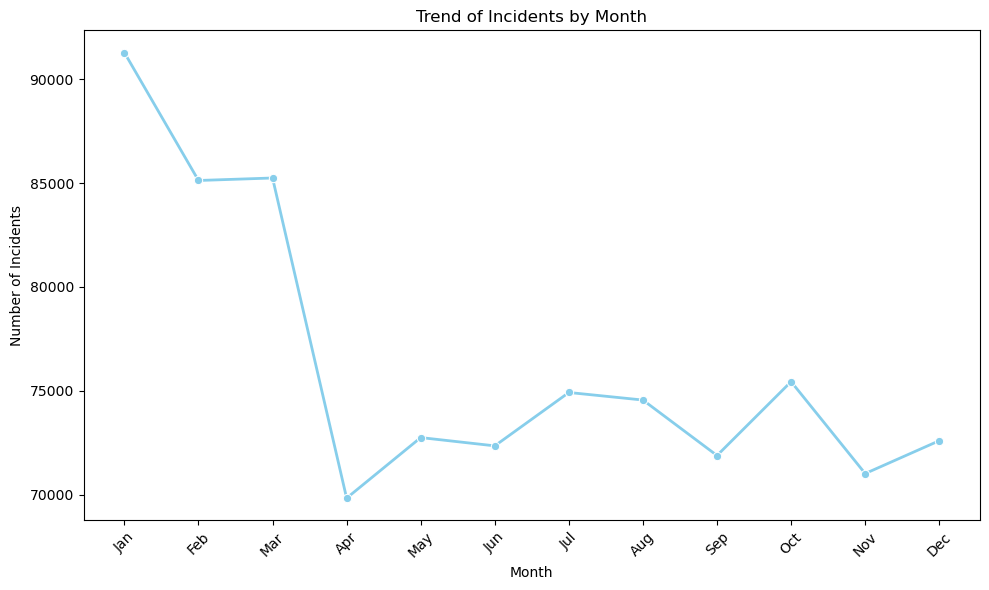


Fig 5.8.8 Trend of incident by Month

The above figure shows how the crime rate between January to December months. It is reducing at the time of Christmas and increasing at starting of year.

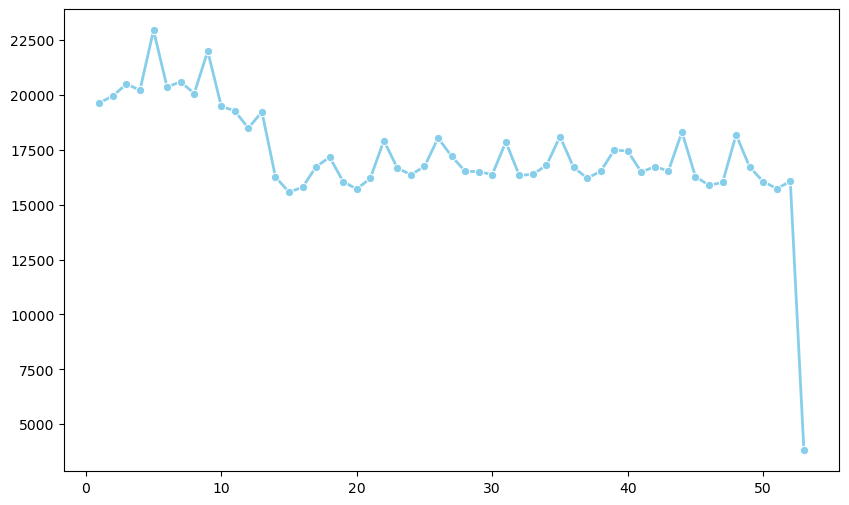


Fig 5.8.9 Trend of incident by weeks

The above figure shows how the cases for all 52 weeks in a year.

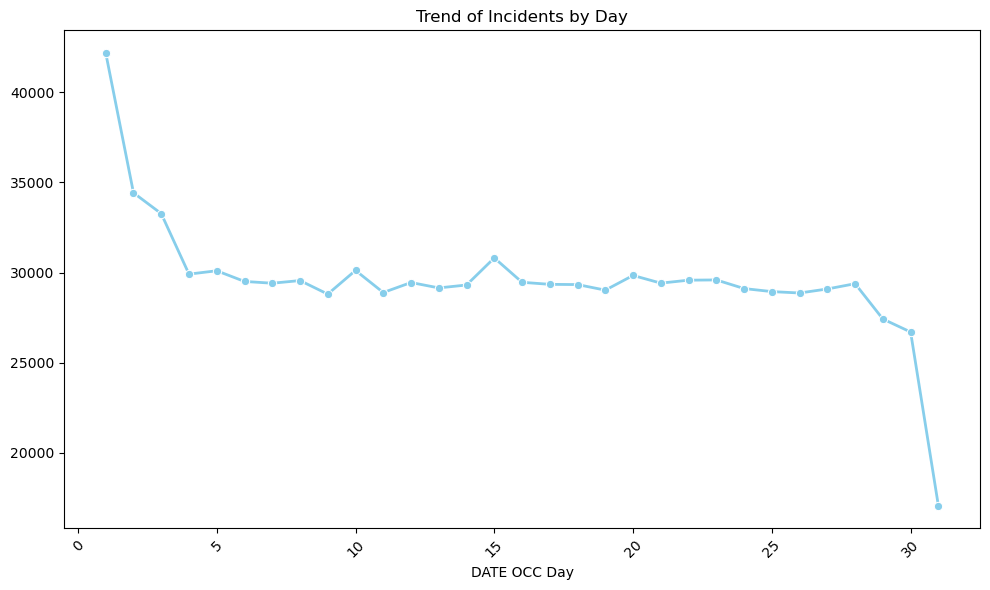


Fig 5.8.10 Trend of incident by days

The above figure shows how the cases for all 30 to 31 days in a month.

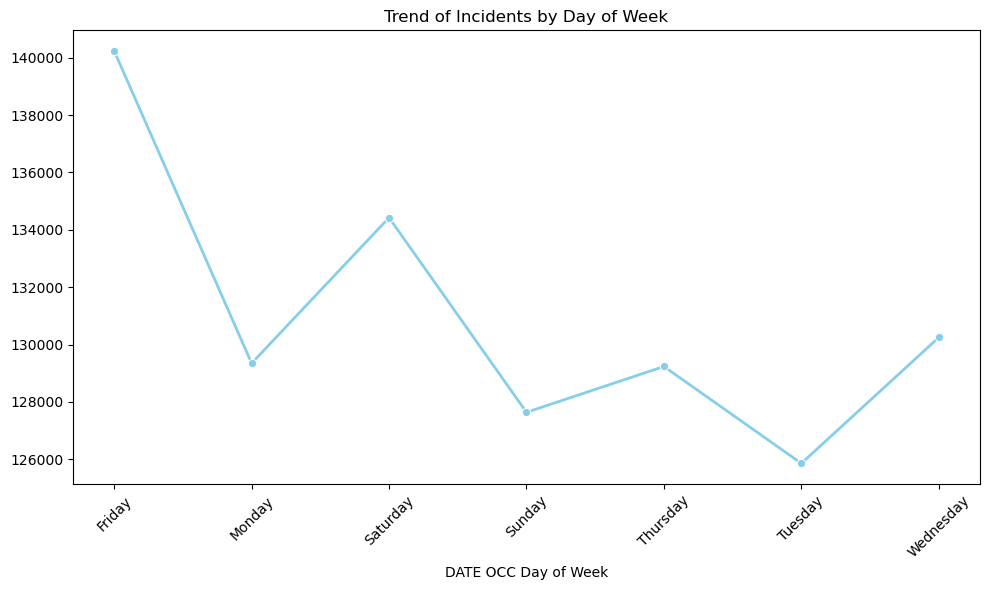


Fig 5.8.11 Trend of incident by day of week

The above figure shows how the cases trend for all 7 days in a week.

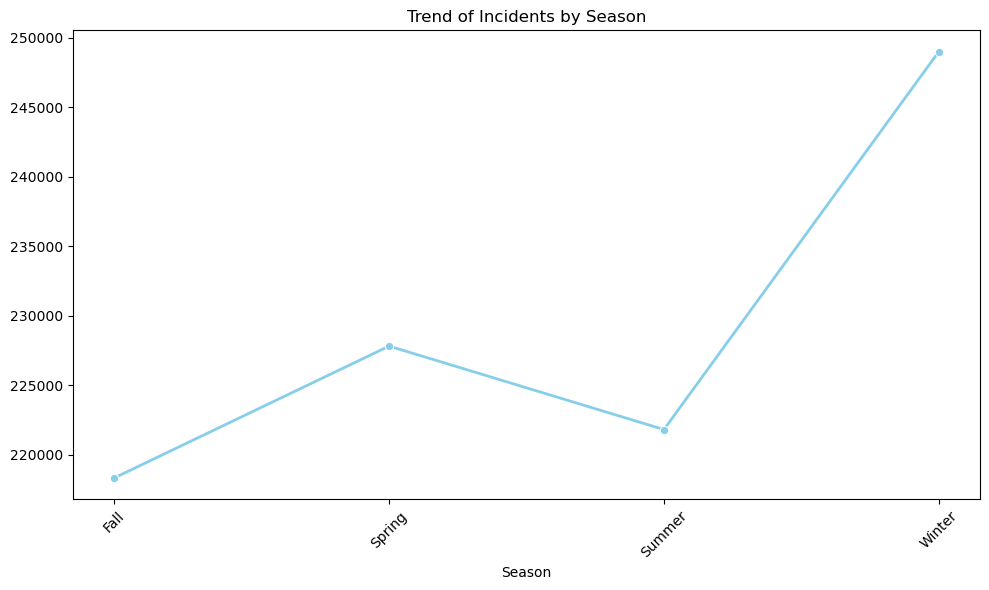


Fig 5.8.12 Trend of incident by season

The above figure shows the increase in crime rate at winter season than other seasons.

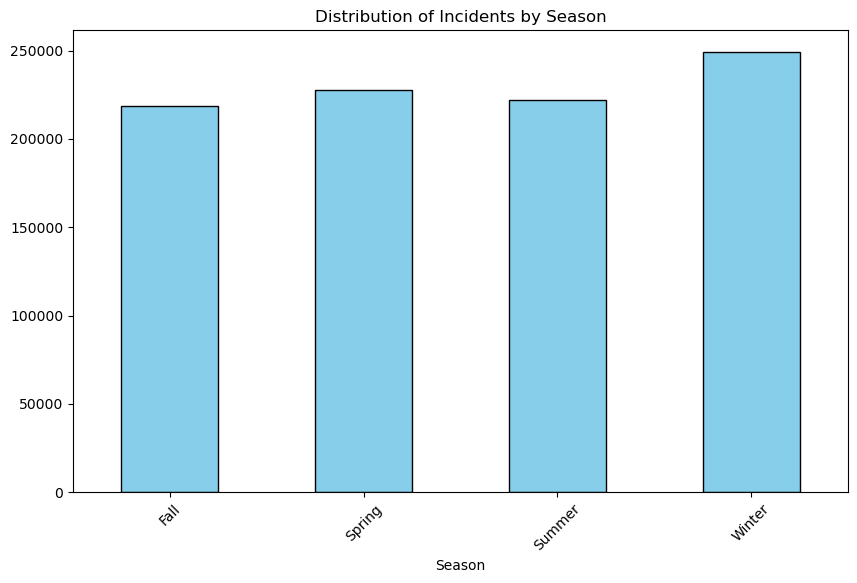
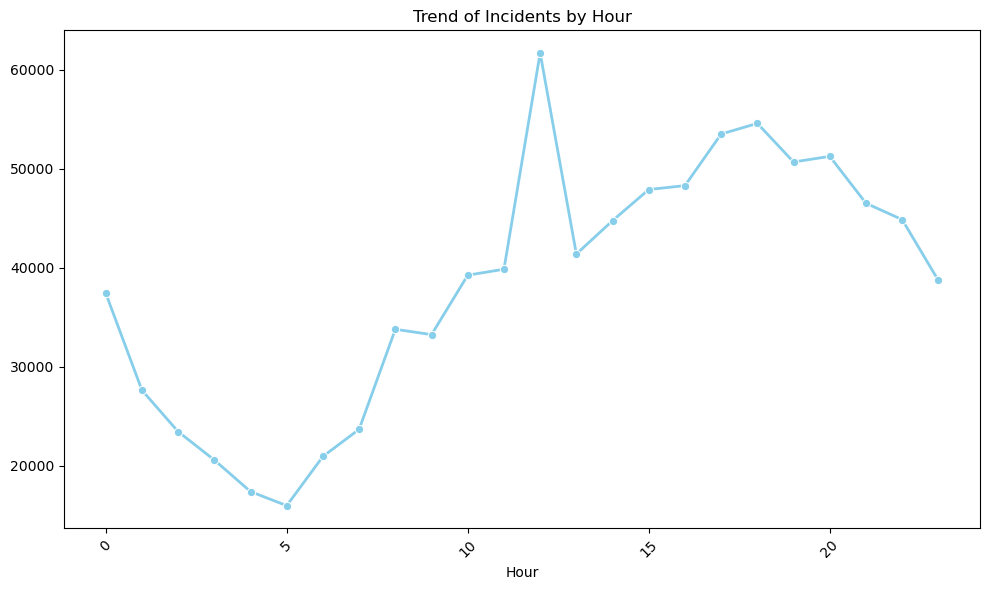


Fig 5.8.13 Distribution of incidents by season

Similarly, same shown through bar plots.

Fig 5.8.14 Trend of incidents by hour

The hour at 12 is a peak hour where most of the crime is happening.

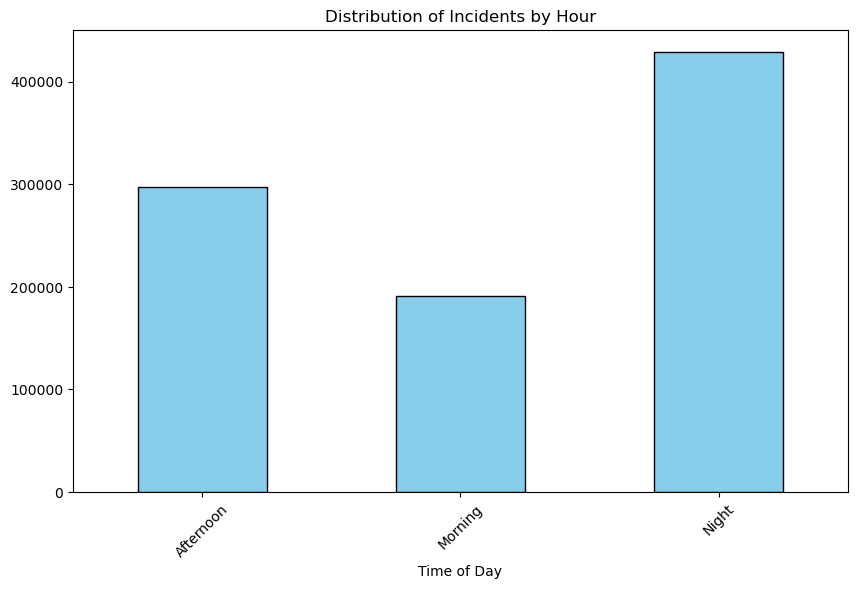


Fig 5.8.15 Distribution of incidents by hour

**The night time the crime happening is more rather than other times.**

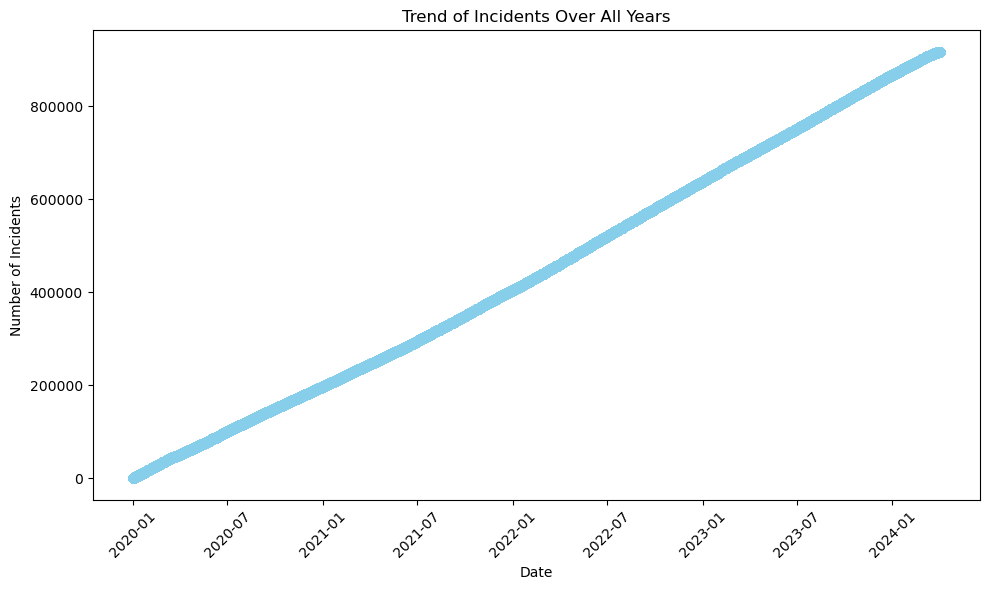


Fig 5.8.16 Trend of incidents over all years

There is significantly increasing in crime rather than decreasing as time changes further.

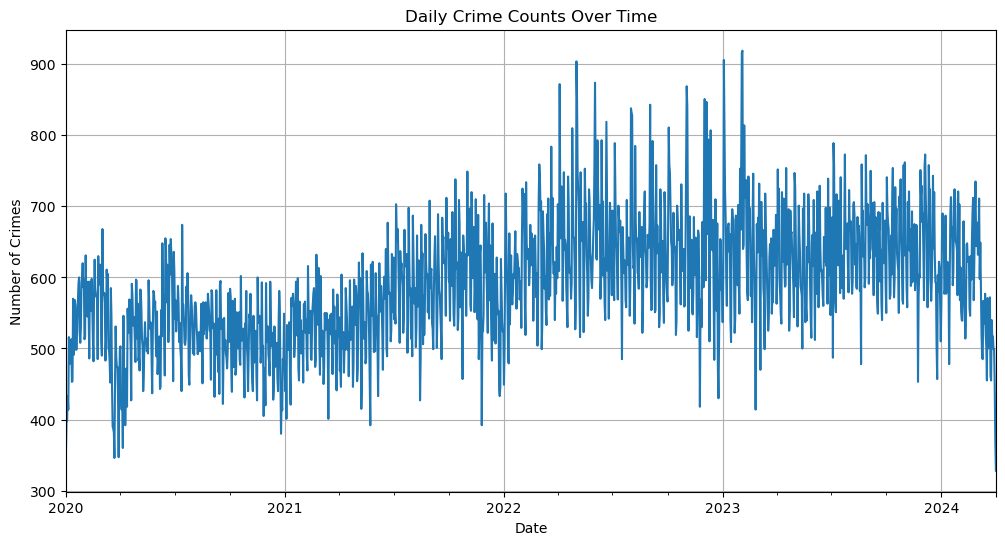


Fig 5.8.17 daily crime counts over time

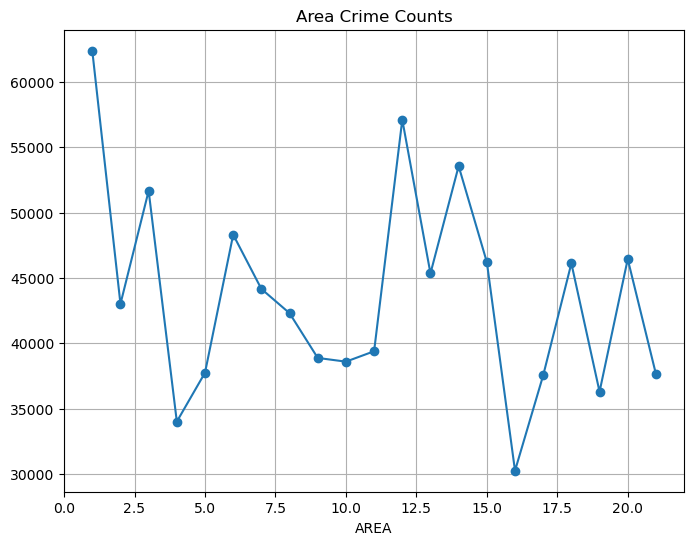
**The daily count of all cases from 2020 to 2024**

Fig 5.8.18 Area crime counts

The crime rates accordingly to area wise.

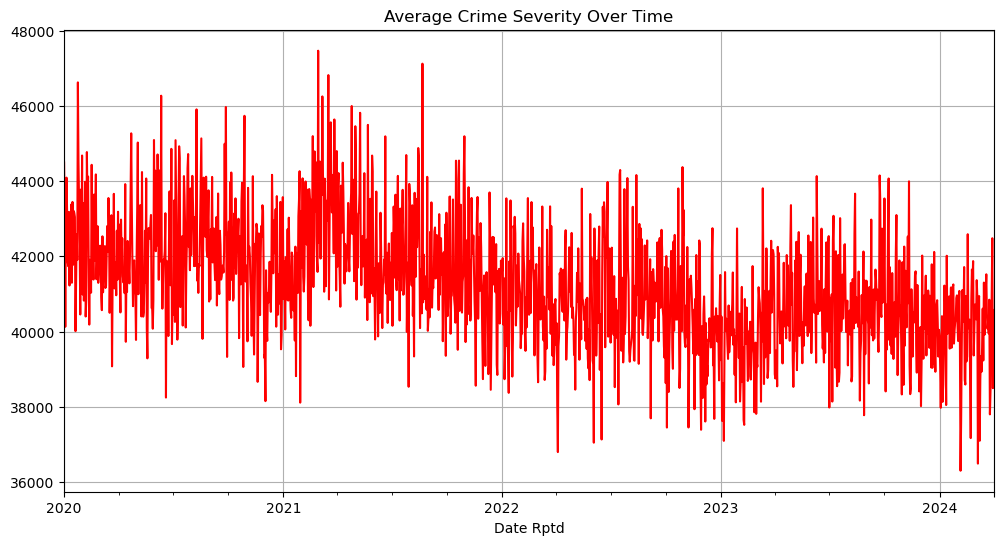


Fig 5.8.19 Average crime severity over time

Based on the trends analysed from various figures and graphs we can say that Overall Crime Rate has shown an increasing trend from 2020 to 2023, with a sudden fall in 2024 due to limited data availability for that year. Crime rates fluctuate throughout the year, typically decreasing during the Christmas season and increasing at the beginning of the year. Crime rates vary by day of the week, with some days experiencing higher crime rates than others. Additionally, crime rates tend to be higher during the night hours. Crime rates are observed to increase during the winter season compared to other seasons. There is a peak in crime rates around the 12th hour of the day, suggesting a peak in criminal activity during that time. Despite fluctuations, there is an overall increasing trend in crime rates over time.

Analysis of these trends suggests that resource allocation strategies should consider peak times, seasonal variations, and long-term trends to effectively target areas with higher crime rates and mitigate criminal activities. Crime rates vary across different geographical areas, indicating the need for tailored approaches to resource allocation and law enforcement strategies based on the specific characteristics of each area.

In summary, the analysis of crime trends provides valuable insights for optimizing resource allocation and developing targeted interventions to reduce crime rates effectively. By understanding the patterns and factors influencing crime, law enforcement agencies can allocate resources more efficiently.

* + 1. **Predicting Crime Type**

Here we are going to predict the crime type based on model trained on data. Here the model is trained on random forest classifier on independent variables such as AREA, Vict Descent, Hour, Time of Day, Part 1-2, On Holidays, Vict Age, Season, DATE OCC Day of Week, Weapon Used Cd, DATE OCC Week and dependent variable is crime type.

Here the model will predict crime type based on the trained input parameters.

Once the model is trained on historical data with known crime types, it can make predictions on new or unseen data by inputting values for the independent variables. The model then applies the learned patterns from the training data to classify the crime type of the new observations.

The following shows the development of web application after model is trained.

* **Web application for Predicting Crime Type**

The Web application is developed using HTML (Hyper Test Markup Language) and CSS (Cascading Style sheet). The web application was developed to predict crime type based on the inputs given. It allows users to input various parameters to find how much it has affected.

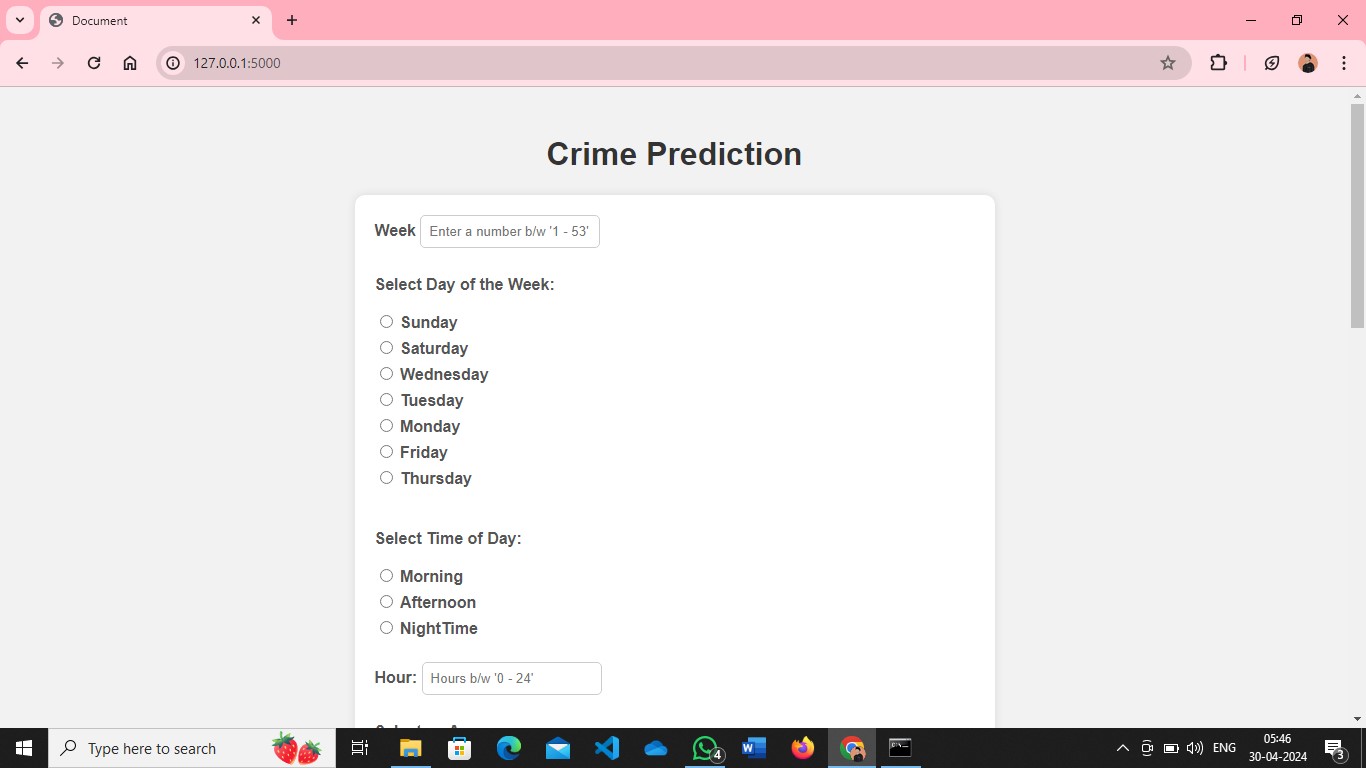


Fig 5.8.20 Web screen 1 for Predicting Crime Type

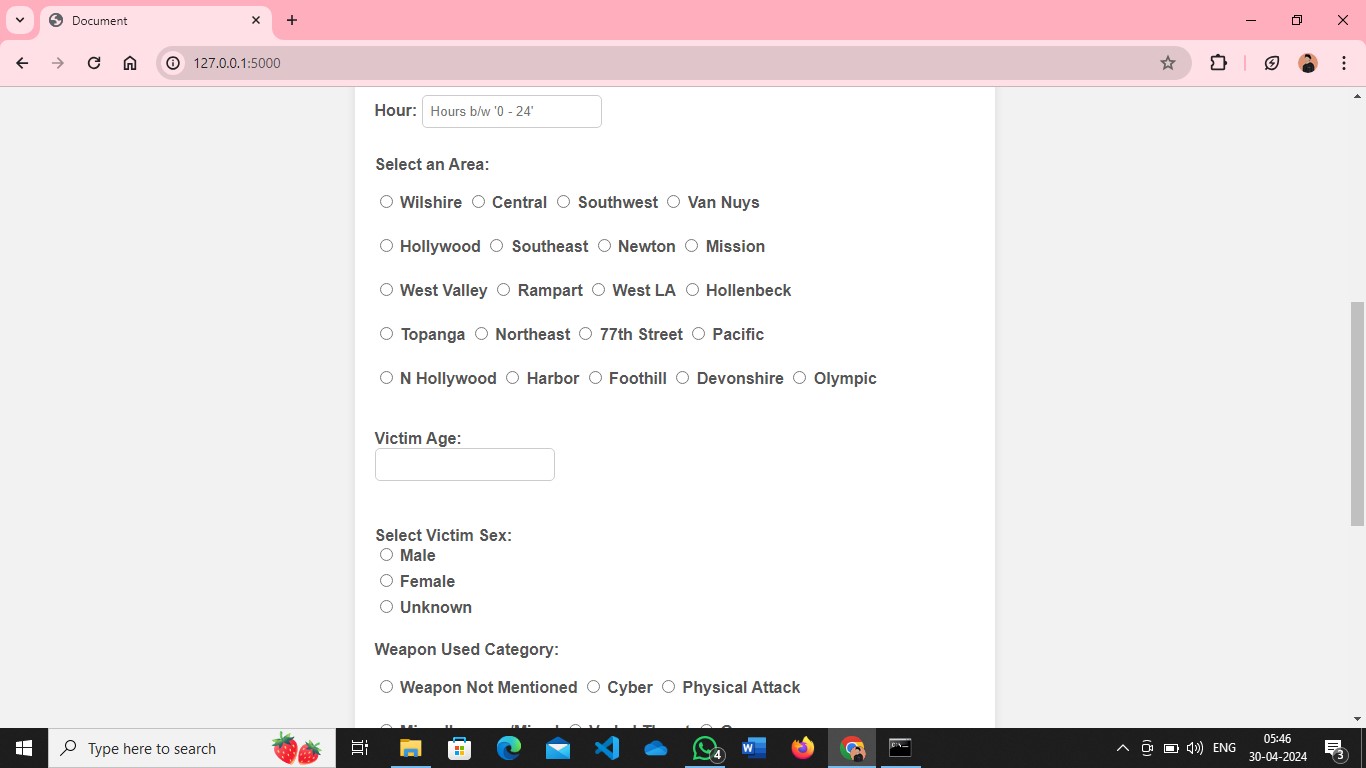


Fig 5.8.21 Web screen 2 for Predicting Crime Type

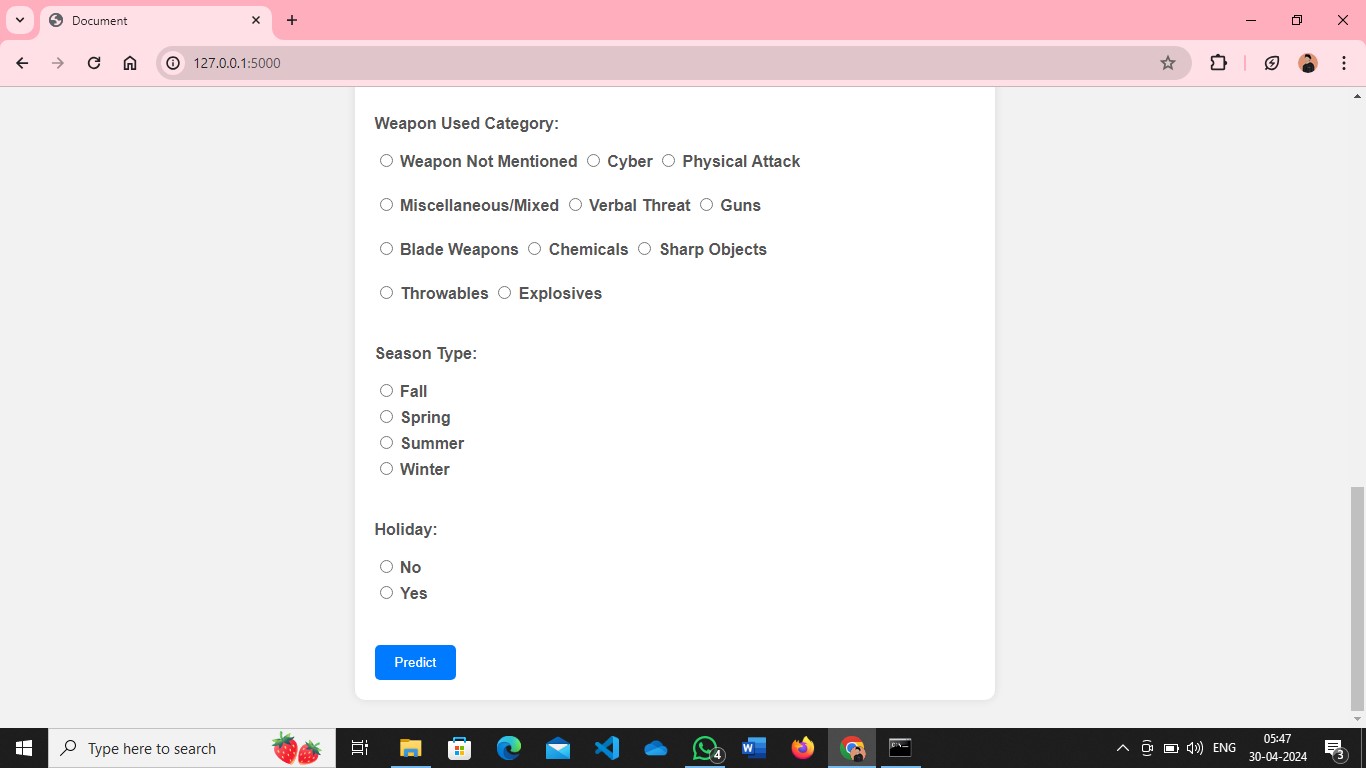


Fig 5.8.22 Web screen 3 for Predicting Crime Type

* + 1. **High-Crime Area Identification**

The high crime area identification is done to find which is mostly spread crime area across the city.

Here the LAT, LON features are used so we can identify the high crime area. For identifying the high crime, we used weapon used category feature which we derived from the weapon used and plot based on LAT and LON provided in the dataset.

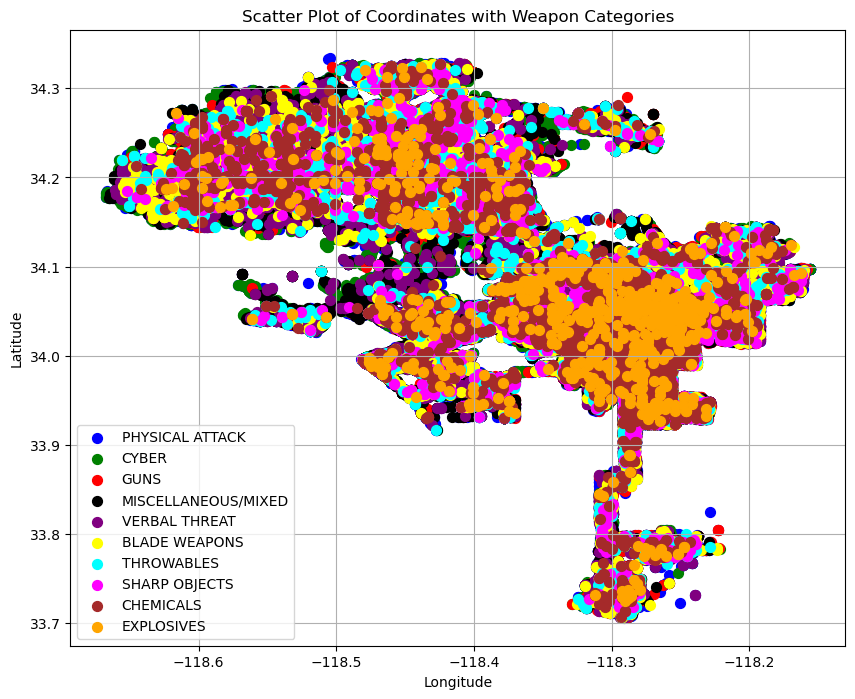


Fig 5.8.23 Scatter plot of coordinates with weapon categories

Based on weapon used category we identify the high crime areas we can see that the orange plots where resembles explosives is mostly used weapons and causing that area as high crime area.

**6. RESULT**

After taking inputs through web application, we created a new Python file to define Flask application and we defined routes for our application. Here one route is used to render the form for users to input their data and another route to handle form submission and generate the result.

When running a Flask application, the framework automatically assigns a port number for the application to listen on. By default, Flask uses port 5000. This means that when the Flask application is started, it will be accessible through a web browser using the appropriate port number. Accessing the application in a web browser would then be done by navigating to “http://localhost:5000”. The port number is crucial as it enables communication between the Flask application and the web browser, allowing users to interact with the web application seamlessly.

* 1. **Crime Severity Prediction**



Fig 6.1.1 Running Flask application on localhost

The URL "http://127.0.0.1:5000" you would be accessing a Flask web application running on your local machine through a web browser. Upon accessing this URL, web browser would display the home page in Flask application, allowing to interact with the web application's features and functionalities.



Fig 6.1.2 Web address of Flask Application

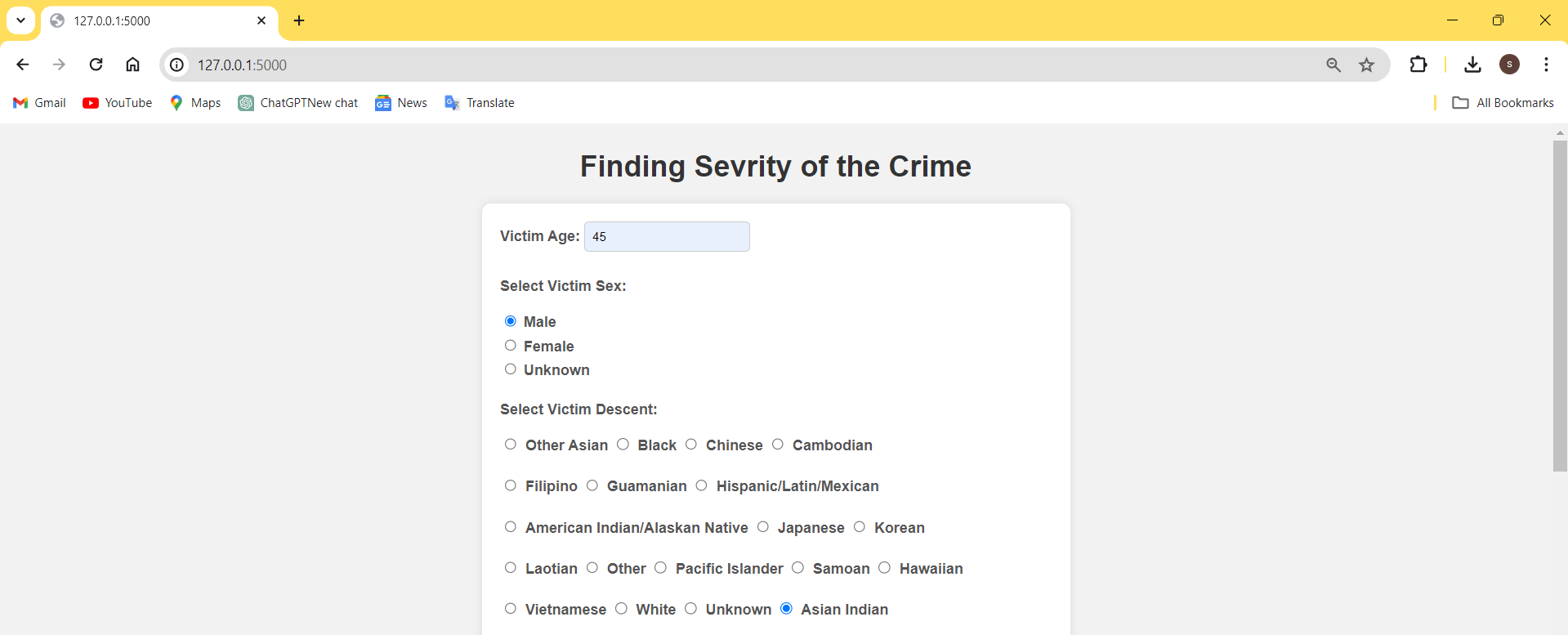


Fig 6.1.3 Input to the web 1

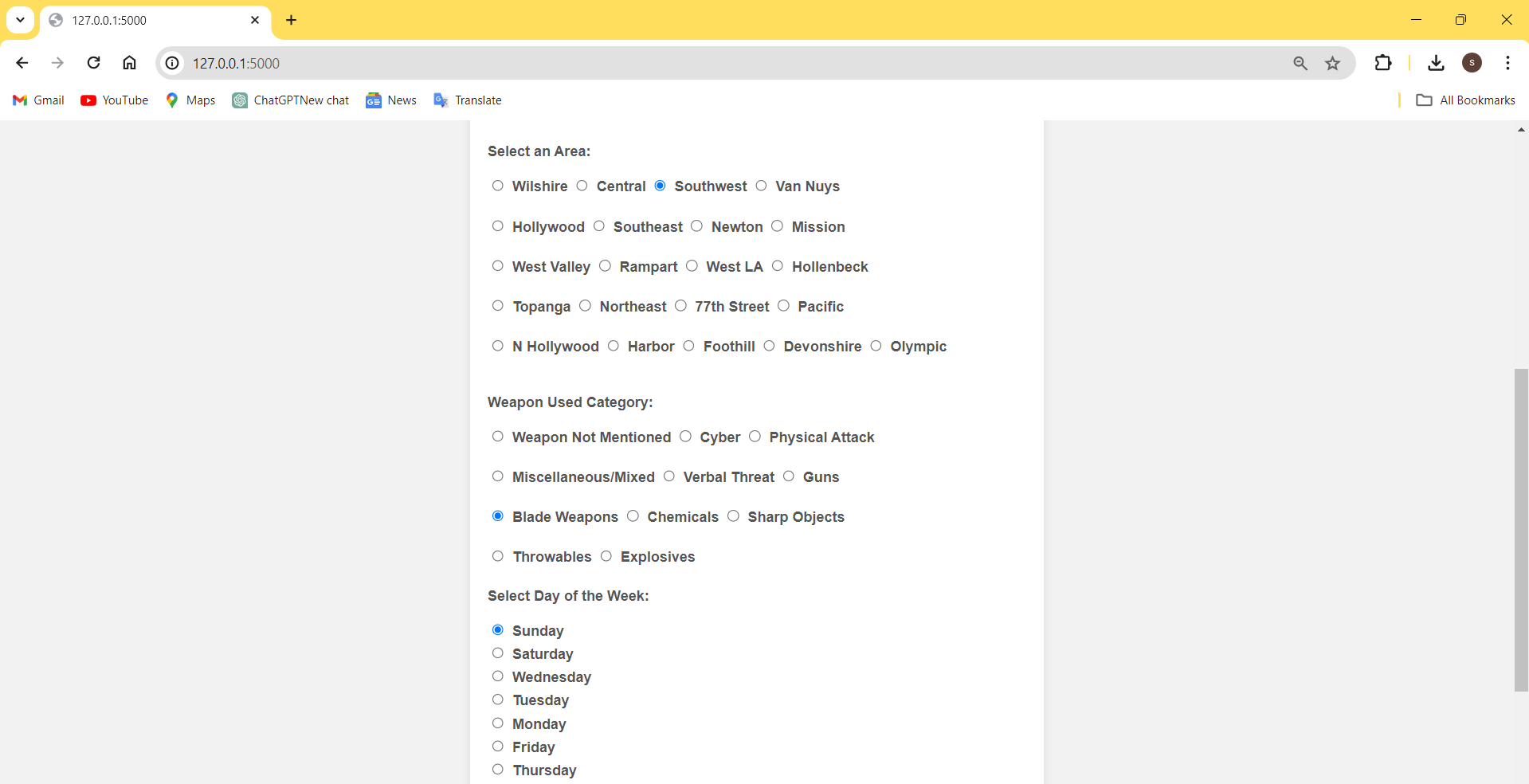
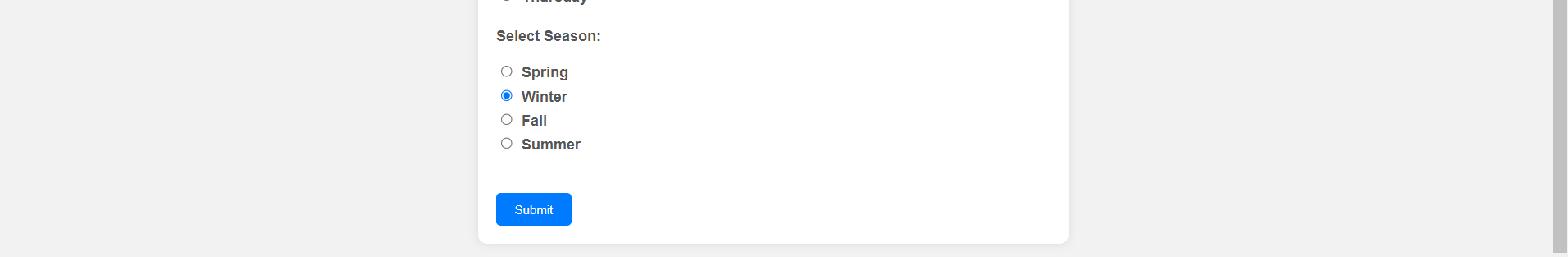


Fig 6.1.4 Input to the web 2

After giving inputs when we click on submit button the predicted value will be displayed on result page.

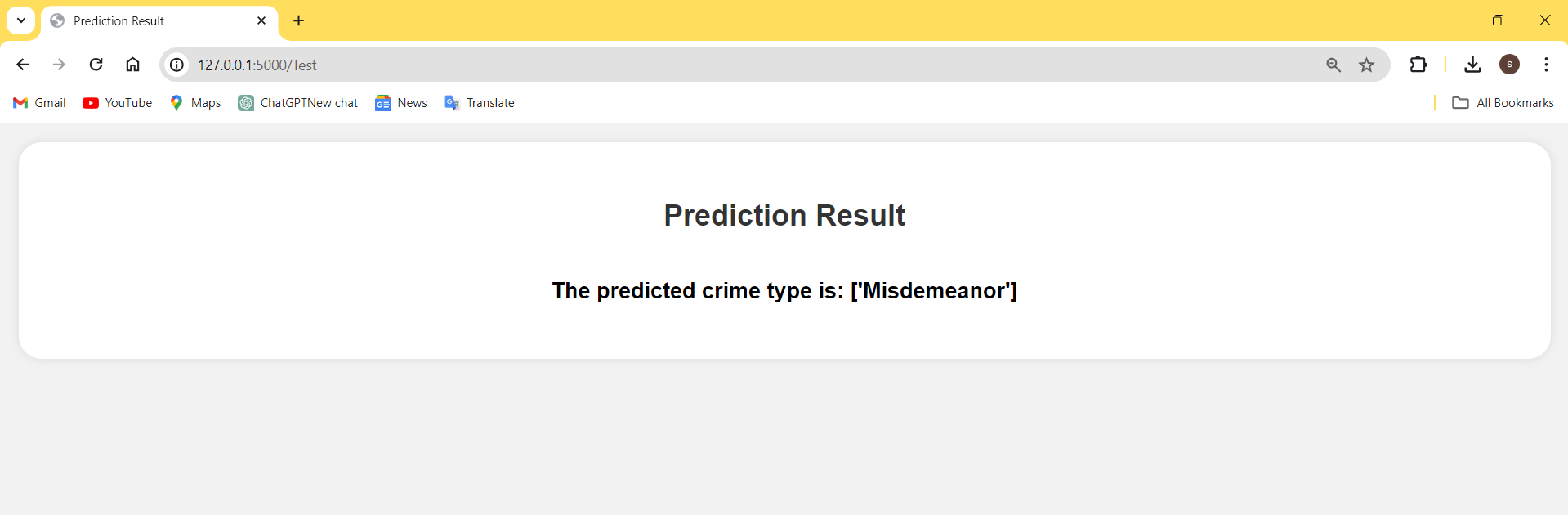


Fig 6.1.5 Output 1 for Crime Severity Prediction

The predicted result says that it has predicted Misdemeanours severity typically refers to a less serious type of criminal offense.

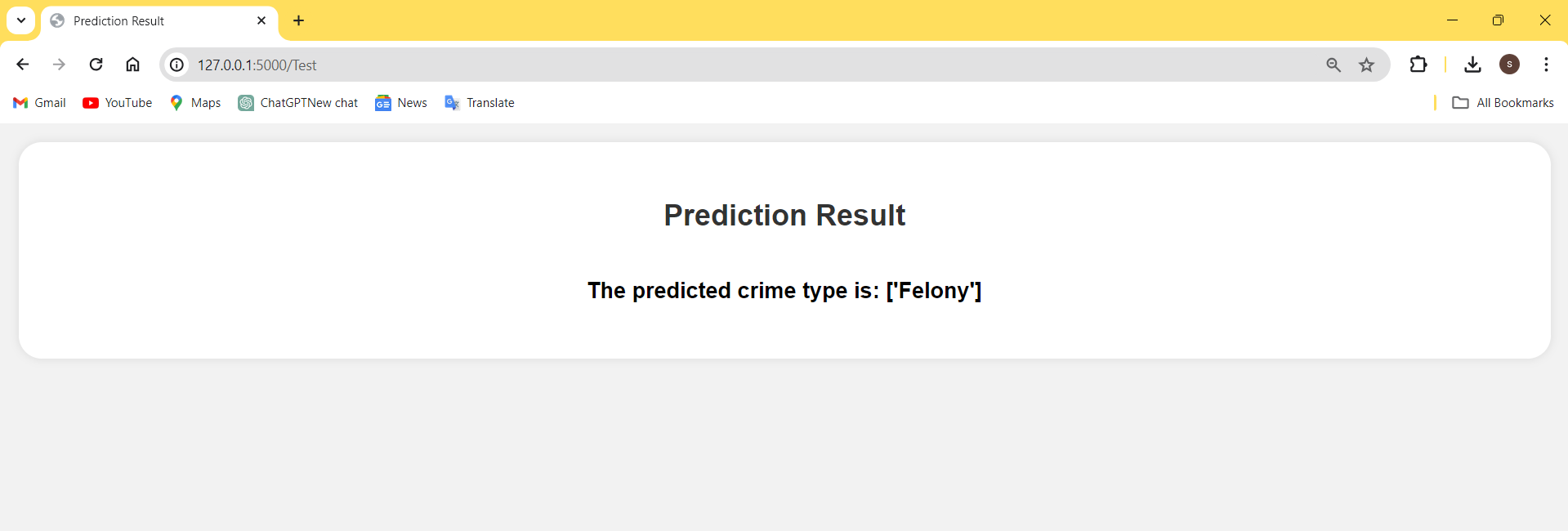


Fig 6.1.6 Output 2 for Crime Severity Prediction

The input parameters have been changed and predicted result says that it has predicted Felony severity typically refers to a serious type of criminal offense

* 1. **Predicting Victim Gender**



Fig 6.2.1 Web address of Flask Application

We will access the main page using the above port number.

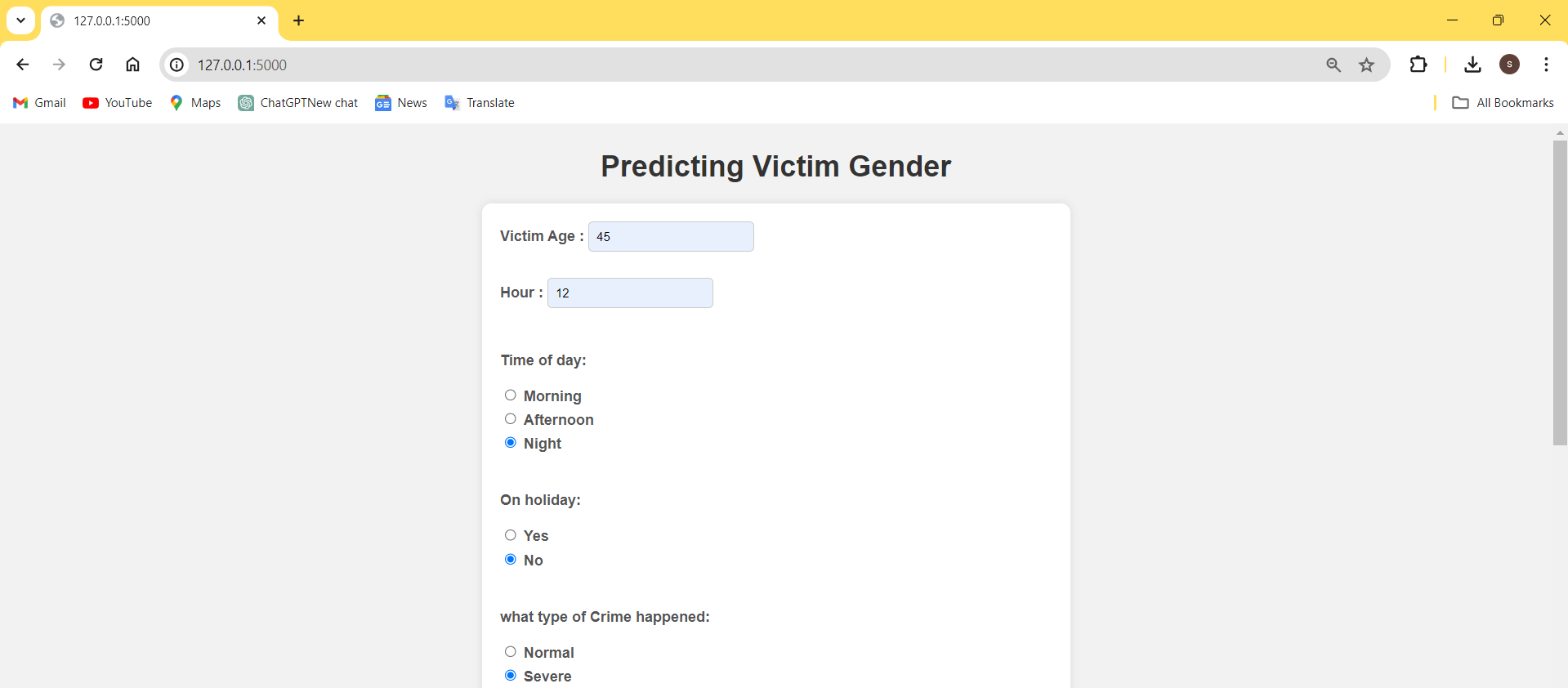


Fig 6.2.2 Input to the web 1

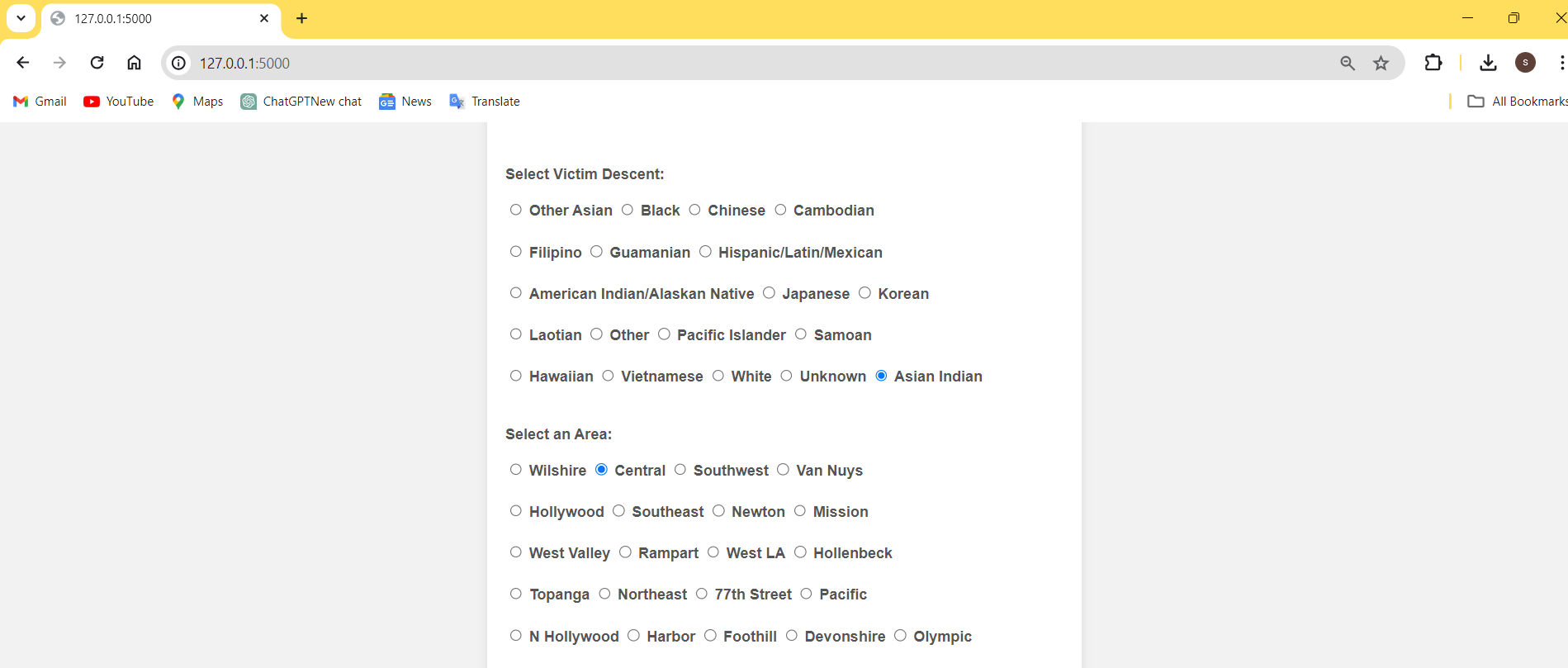


Fig 6.2.3 Input to the web 2

After giving inputs when we click on submit button the predicted value will be displayed on result page.

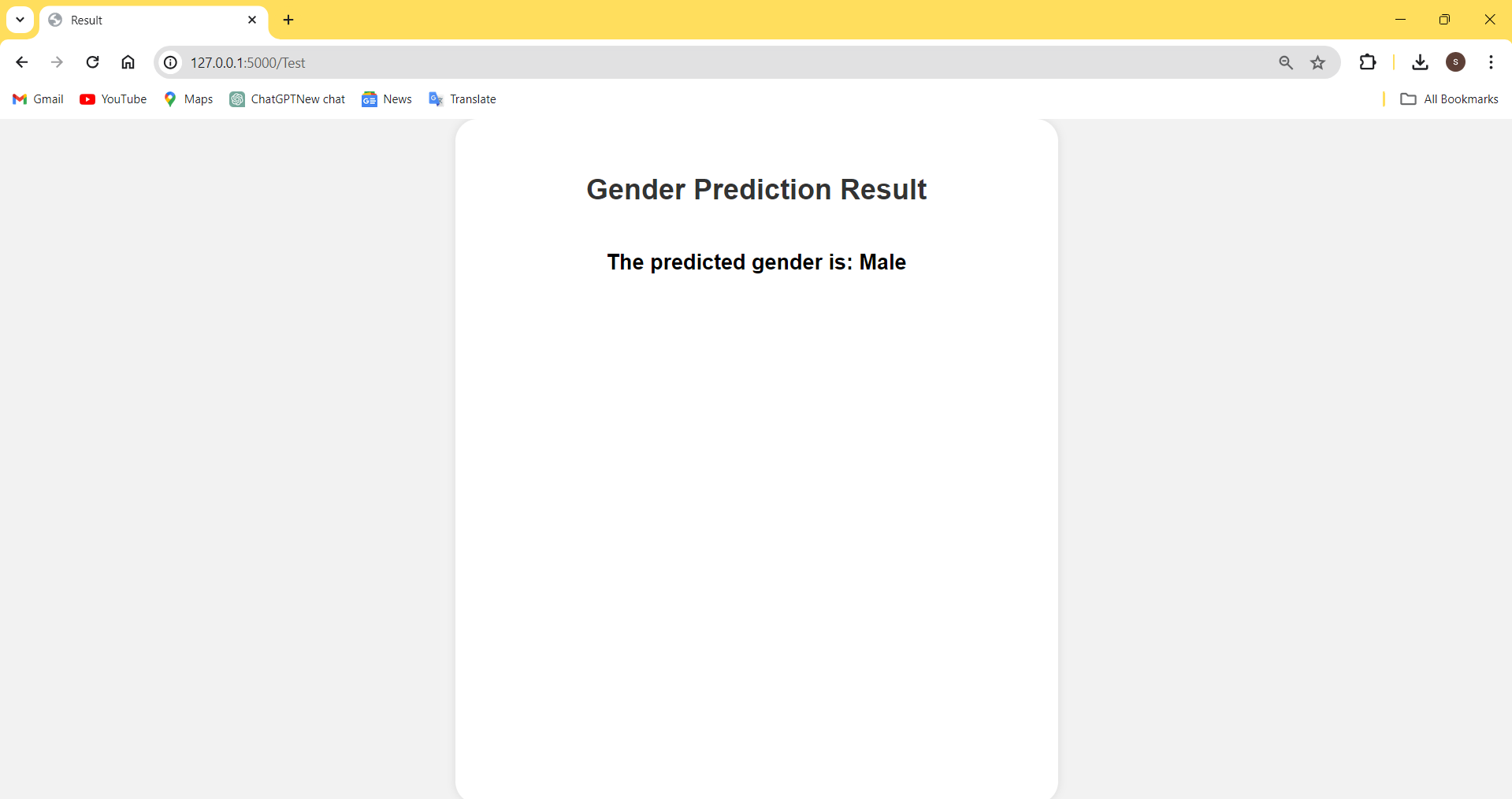


Fig 6.2.4 Output for Gender Prediction

The predicted result specifies that gender type is male based on input parameters and trained model.

* 1. **Predicting Crime Type**

We will access the main page using the below port number.



Fig 6.3.1 Web address of Flask Application

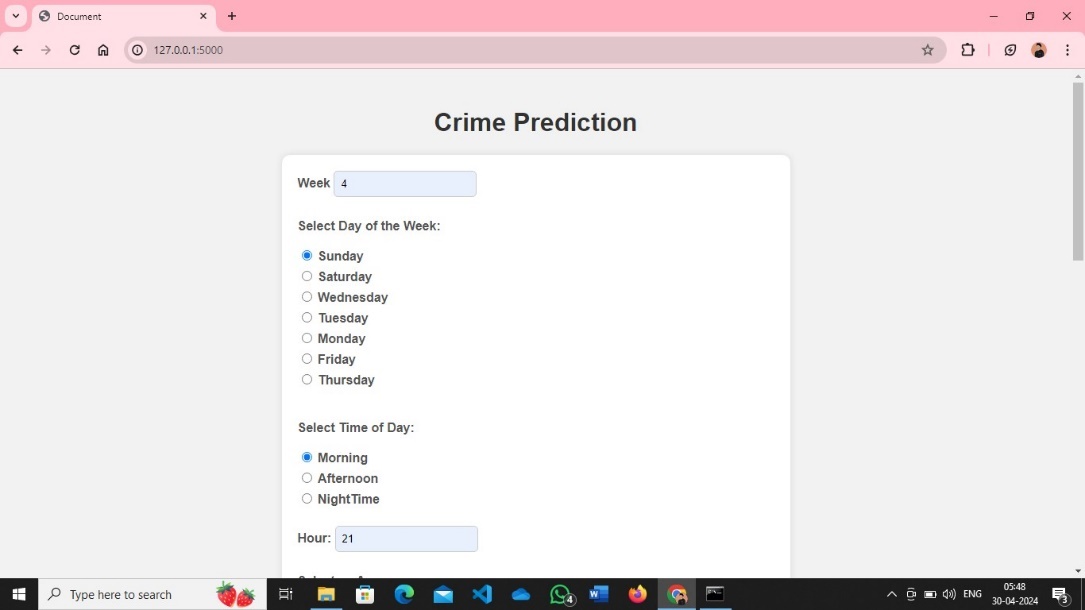


Fig 6.3.2 Input to the web 1

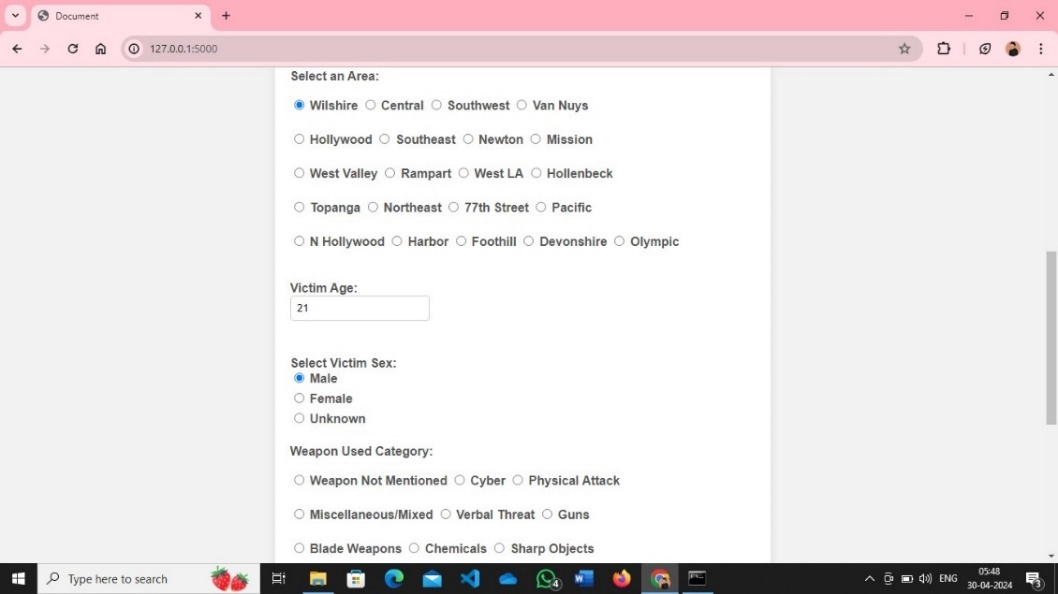


Fig 6.3.3 Input to the web 2

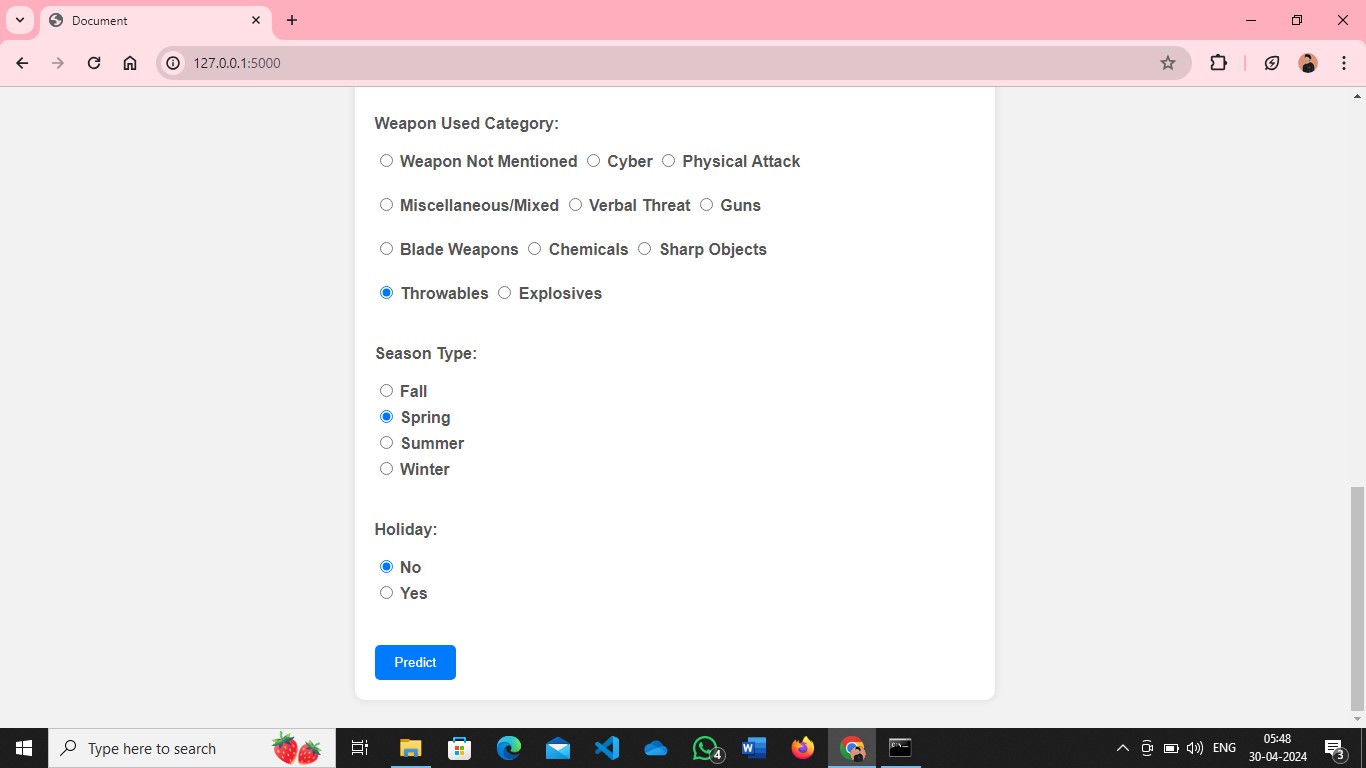


Fig 6.3.4 Input to the web 3

After giving inputs when we click on submit button the predicted value will be displayed on result page.

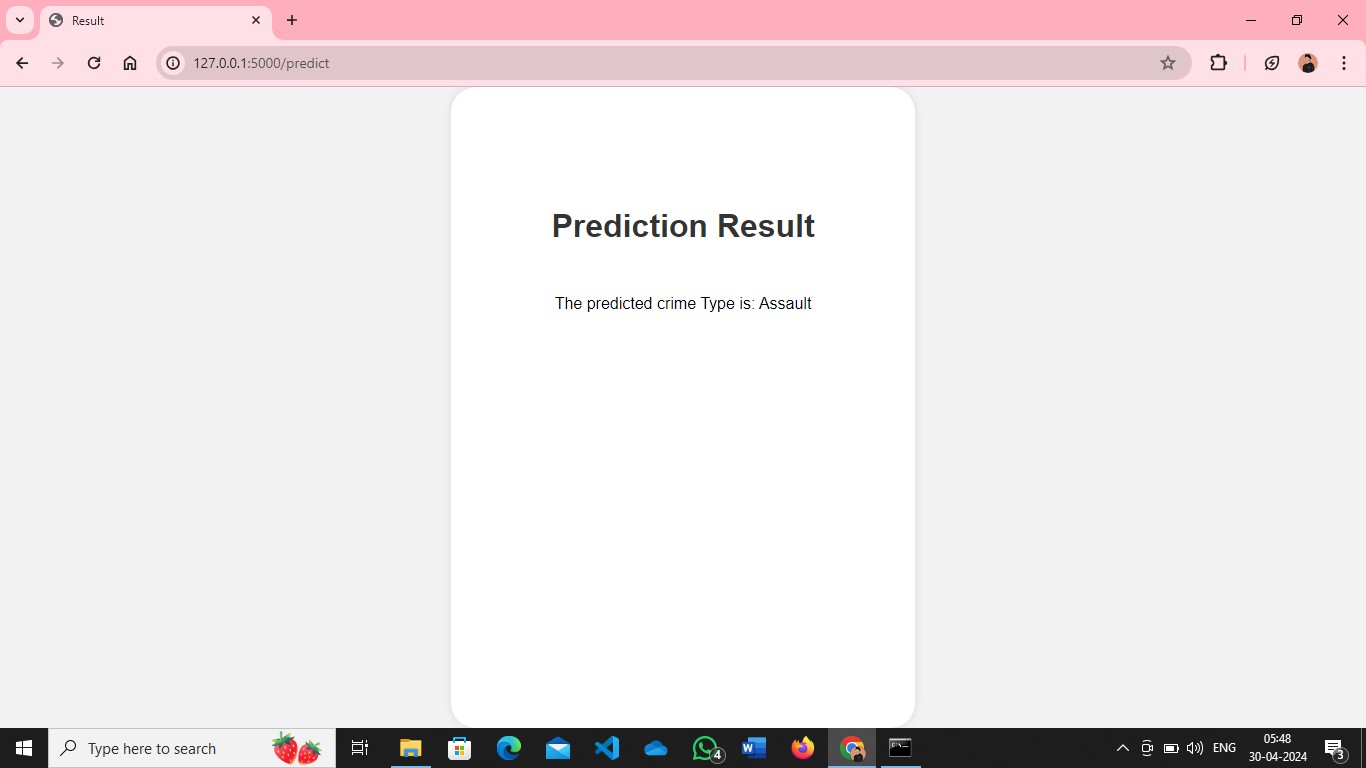


Fig 6.3.5 Output 1 for crime type prediction

The predicted result is assault based on given input parameters and trained model.

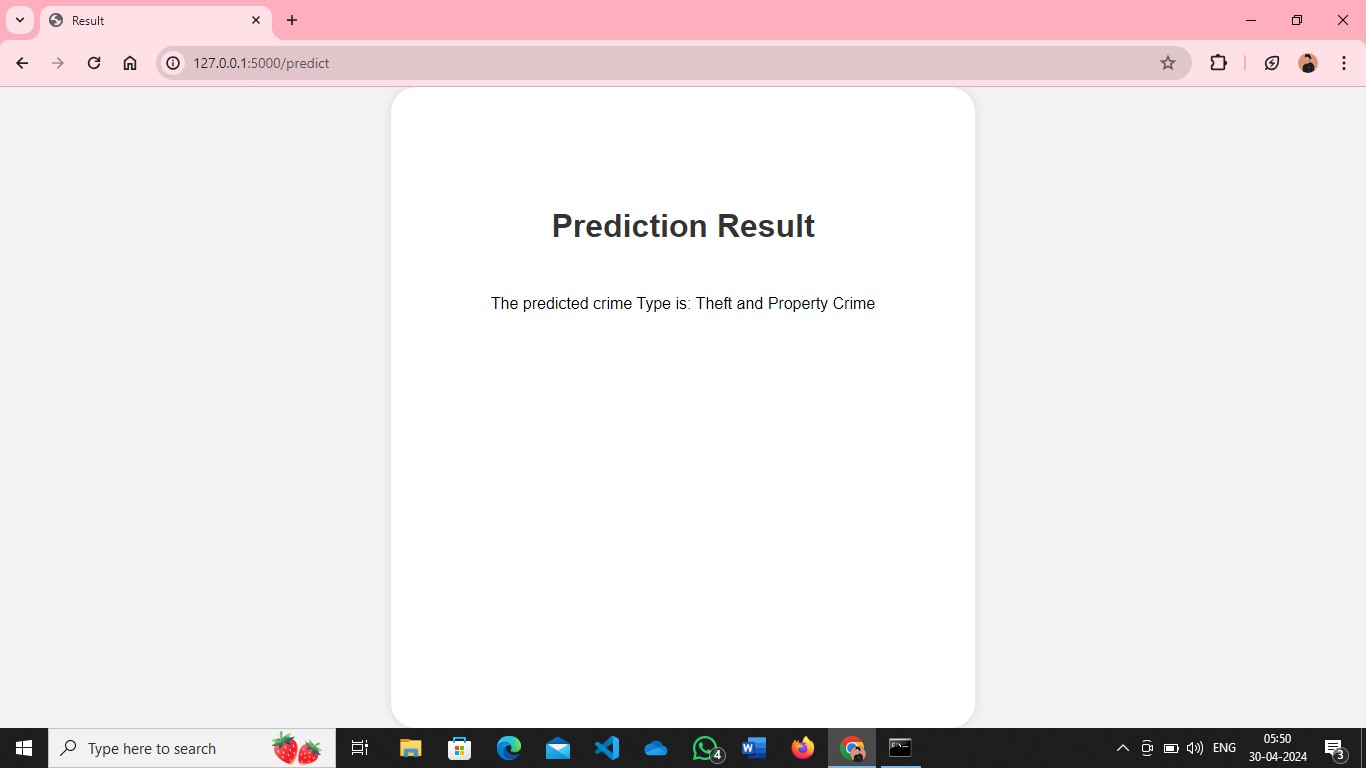


Fig 6.3.6 Output 2 for crime type prediction

The predicted result is theft and property crime based on given input parameters and trained model.

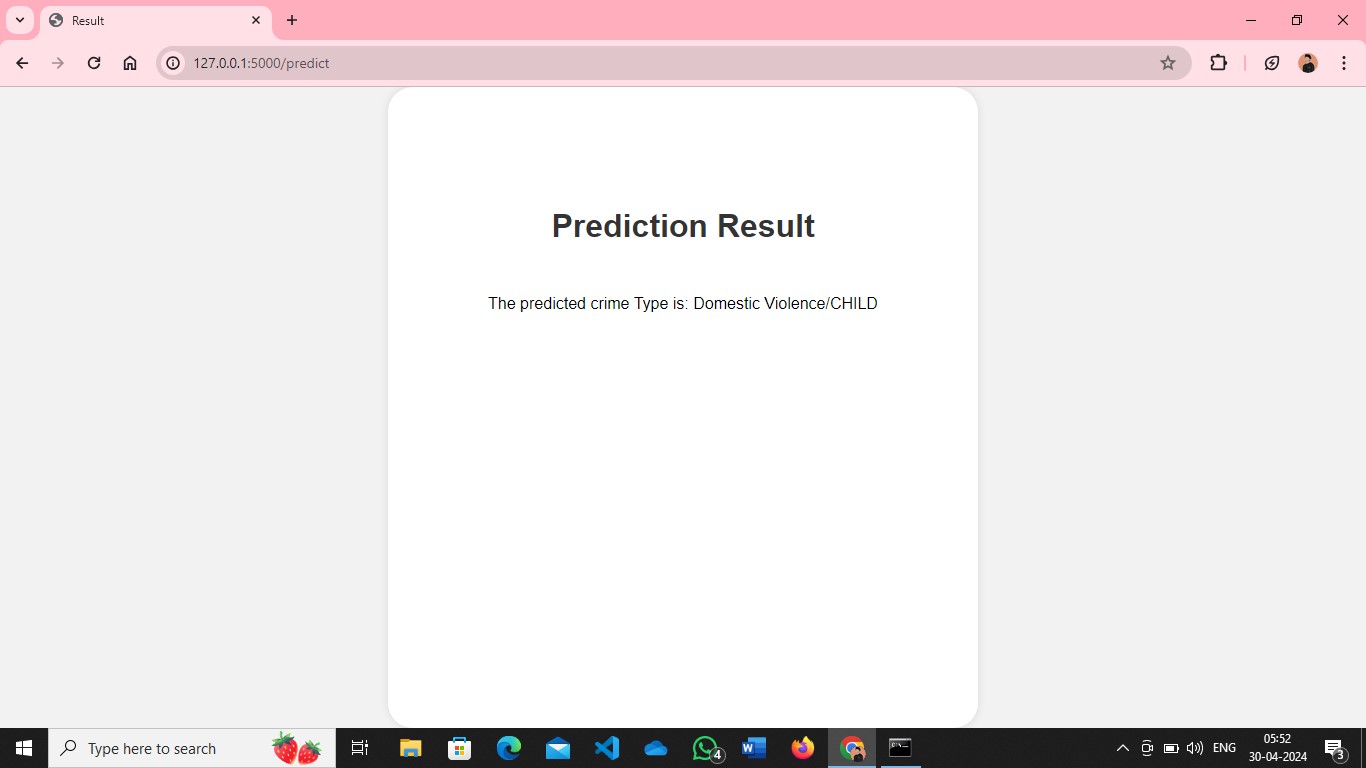


Fig 6.3.7 Output 3 for crime type prediction

The predicted result is domestic violence based on given input parameters and trained model.

1. **Conclusion**

In conclusion, our study used a comprehensive approach to analyse crime data and create predictive models to address various research goals. We meticulously cleaned and pre-processed the data to ensure its accuracy, handling missing values and outliers, and transforming variables for analysis. Through Exploratory Data Analysis (EDA), we uncovered valuable insights into crime trends and patterns, which guided the development of predictive models.

We tackled several research problems, including predicting crime severity, victim gender, crime type, and identifying high-crime areas. We employed machine learning techniques, particularly random forest classifiers, to make predictions based on factors like location, time, demographics, and weapon involvement.

Moreover, we developed user-friendly web applications to allow stakeholders to access our models and obtain real-time predictions. These tools offer valuable insights for crime prevention and intervention strategies. Overall, our models demonstrated promising accuracy in predicting crime-related outcomes. The insights gained from our analysis can help law enforcement agencies and policymakers optimize resource allocation and improve public safety in urban areas.

1. **Future Scope**

In future crime data projects, it's crucial to understand the dataset's structure and conduct thorough exploratory data analysis (EDA) to uncover insights. Handle missing values appropriately and encode categorical features for machine learning compatibility. Feature engineering, especially extracting temporal and spatial patterns, enhances model performance. Select suitable algorithms, split data for training/testing, and fine-tune models for robustness. Develop a user-friendly web interface for accessing predictive models. Document all steps and findings comprehensively for transparency and reproducibility, providing valuable insights for decision-makers.

1. **REFERNECS**

* <https://trysakai.longsight.com/portal/site/520b8591-6130-4776-bf04-504e18e885f1/tool/e59fe5bb-2464-4b5c-9679-4d9af5857ede?panel=Main>
* [www.youtube.com](http://www.youtube.com)
* [www.google.com](http://www.google.com)
* <https://data.lacounty.gov>