**Exploring Crime Analysis with LAPD Leveraging Machine Learning**

**for Public Safety**

**CAPSTONE PROJECT REPORT**

By

Team 03

A Harshith 21R11A0504

D Nissi 21R11A0511

G Harshit 21R11A0519

G Pranav 21R11A0515

P Rohith 21R11A0537



**Department of Computer Science and Engineering**

**Accredited by NBA**

**Geethanjali College of Engineering and Technology**

**(UGC Autonomous)**

(Affiliated to J.N.T.U.H, Approved by AICTE, New Delhi)

Cheeryal (V), Keesara (M), Medchal.Dist.-501 301

March 2024

**ABSTRACT**

This capstone project delves into LAPD crime data from 2020 to 2024, aiming to reveal patterns and trends in crime occurrence across Los Angeles. It investigates temporal and spatial aspects, relationships between crime types, victim demographics, severity assessments, case statuses, and location specifics. The objective is to uncover seasonal fluctuations, identify high-crime areas, and understand common methods used by criminals. By profiling demographics and assessing severity, the project seeks to address inequalities and enhance crime prevention tactics. Geospatial analysis is utilized to map crime hotspots, aiding targeted interventions and resource allocation. Ultimately, this study aims to contribute to public safety improvement and the creation of a safer environment in Los Angeles.

Top of Form

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S No** | **Title** | **Page No** |
|  | **ABSTRACT** | 2 |
| **1** | **INTRODUCTION** | 4 |
| **2** | **PROBLEM STATEMENT** | 5 |
| **3** | **OBJECTIVES** | 5 |
| **4** | **Literature Survey** | 6 |
| **5** | **METHODOLOGY** | 7 |
|  | 5.1 Data Source | 7 |
|  | 5.2 Exploratory Data Analysis | 8 |
| **6** | **Final Implementation** | 19 |
| **7** | **Conclusion** | 21 |
| **8** | **Future Scope** | 21 |

1. **INTRODUCTION**

Crime presents a persistent challenge that extends beyond borders, impacting communities globally and presenting substantial hurdles for law enforcement and policymakers alike. Los Angeles, California, is no stranger to this reality, contending with intricate crime dynamics that necessitate a comprehensive grasp and strategic interventions. Addressing this need, the Los Angeles Police Department (LAPD) diligently documents reported crimes, compiling a comprehensive dataset spanning multiple years. This dataset serves as a crucial asset for scrutinizing crime trends, victim characteristics, and spatiotemporal patterns, guiding focused interventions and resource distribution strategies to bolster public safety.

The aim of this capstone project is to perform an exhaustive analysis of LAPD crime data spanning 2020 to 2024, emphasizing the discovery of practical insights to aid law enforcement agencies and policymakers in their endeavors to address crime and enhance community welfare. Utilizing the extensive dataset at our disposal, our goal is to pinpoint significant trends, associations, and determinants impacting crime rates in Los Angeles.

The analysis will be guided by multiple dimensions, encompassing temporal scrutiny to unveil seasonal fluctuations and daily/weekly crime patterns, spatial examination to pinpoint high-crime zones and potential influencers, and exploration of crime type relationships to grasp the interconnections among various criminal activities. Moreover, demographic profiling of victims, severity evaluations, case status assessments, and utilization of geospatial visualization techniques will be employed to offer a comprehensive insight into crime dynamics in Los Angeles. Through this comprehensive approach, we aim to facilitate evidence-based decision-making processes aimed at bolstering crime prevention strategies and cultivating a safer, more secure environment for all Los Angeles residents.

**2. PROBLEM STATEMENT**

Crime represents a significant menace to the safety and welfare of Los Angeles communities, profoundly impacting residents' daily lives. Despite diligent efforts from law enforcement, comprehending the intricate dynamics of crime remains challenging. It is imperative to conduct a thorough analysis of LAPD crime data to discern patterns, trends, and underlying factors influencing crime occurrence. Through this analysis, we can enhance resource allocation and devise tailored interventions to prevent crime and bolster public safety. This project endeavors to unveil the root causes driving crime in Los Angeles, furnishing actionable insights for law enforcement and policymakers to effectively address these challenges. By embracing data-driven methodologies, we aim to empower communities and foster a more secure environment for all residents.

**3. OBJECTIVES**

1. Examine LAPD crime data spanning 2020 to 2024 to reveal temporal patterns and trends in crime occurrence.

2. Delve into the spatial distributions of crime across Los Angeles's 21 community police stations, pinpointing high-crime areas and potential contributing factors.

3. Explore connections between different crime types and methods of operation to comprehend patterns in criminal behavior.

4. Profile crime victims demographically to uncover discrepancies in victimization rates concerning age, gender, and ethnicity.

# **4.**Literature survey****

1. "Temporal Analysis of Crime Patterns: A Review of Literature" - This survey explores existing research on temporal patterns in crime occurrence, focusing on seasonal variations and daily/weekly trends, to provide insights into how time influences criminal activity.

2. "Spatial Analysis of Crime: A Comprehensive Review" - This literature review examines studies on spatial distributions of crime, including hotspot identification and analysis of geographical factors influencing crime occurrence, to understand the spatial dynamics of criminal behavior.

3. "Understanding Crime Type Relationships: A Survey of Literature" - This survey synthesizes research on the relationships between different types of crimes and modus operandi, exploring patterns in criminal behavior and the interconnectedness of various crime categories.

4. "Demographic Profiling of Crime Victims: A Review of Research" - This literature review investigates studies on demographic profiling of crime victims, including age, gender, and ethnicity, to identify disparities and trends in victimization rates and inform equitable crime prevention strategies.

5. "Geospatial Visualization Techniques in Crime Analysis: A Comprehensive Review" - This survey examines existing literature on geospatial visualization techniques for crime analysis, including mapping crime hotspots and clusters, to facilitate resource allocation and targeted interventions by law enforcement agencies and policymakers.

**5. METHODOLOGY**

**5.1 Data Source**

* **Description of Data Set**

This data is extracted from a US Government website Data.gov.

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

The daily crime occurrence in Los Angeles state is recorded and updated on the website starting from 2020 till the latest.

Columns/ Variables - 28

Rows/ Observations – 925721

**Columns/Variables:**

1. DR\_NO: TheDepartment Report Number is represented as DR\_NO which is generated for every crime that is reported to the Los Angeles department.

2. Date Rptd:The date when the crime was reported to the police department is recorded as Date Rptd and the date lies between 2020 to 2024.

3. DATE OCC:The actual date when the crime occurred is recorded under this column.

4. TIME OCC:The time the crime occurred is represented in a 24-hour format.

5. AREA: The Los Angeles Police Department has 21 community police stations within the geographical location sequencing from 1-21.

6. AREA NAME:The 21 divisions also have an area name in Los Angeles

7.Rprt Dist. No: Code that represents a sub-area within a Geographic Area this is usually prefixed by area code.

8. Part code: Indicate whether the crime is serious or less offensive.

9. Crm Cd: Indicates the crime committed.

10. Crm Cd Desc: Defines the Crime Code provided.

11. Mocodes: Modus Operandi code provides additional details about the crime.

12. Vict Age: Indicates the age of the victim.

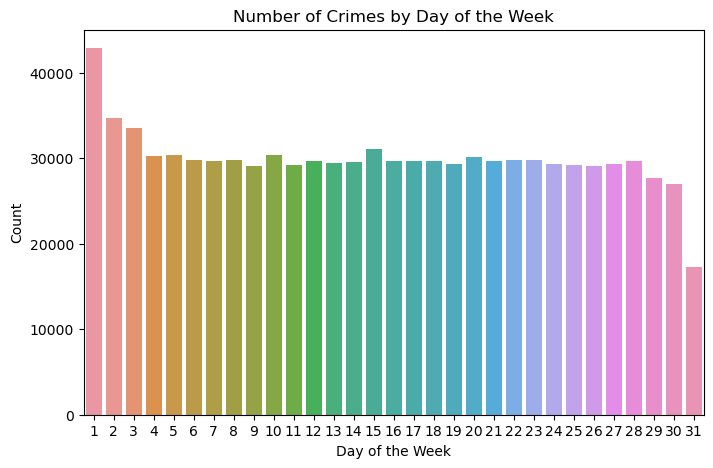
13. Vict Sex: F: Female M: Male X: Unknown

14. Vict Descent: Descent Code: A - OtherAsian B - Black C - Chinese D - Cambodian F - Filipino G - Guamanian H - Hispanic/Latin/Mexican I - American Indian/Alaskan Native J - Japanese K - Korean L - Laotian O - Other P - Pacific Islander S - Samoan U - Hawaiian V - Vietnamese W - White X - Unknown Z - Asian Indian

15. Premise Cd: Type of structure the crime took place in (vehicle, building, parking lot, etc

**5.2. Exploratory Data Analysis**

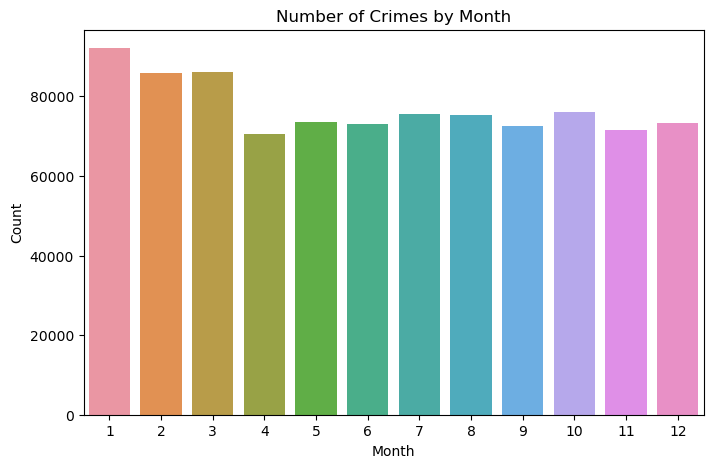
1. **Monthly Crime Trend: A surge is observed on the first day, succeeded by a gradual decline in criminal activity throughout the month.**



**Observations:**

* The bar graph illustrates a notable spike in crime rates exceeding 40,000 incidents on the 1st day of each month, implying increased criminal activity at the start of the month
* Over time, there is a consistent decline in the frequency of crimes as the month progresses, indicating a downward trend in criminal activity throughout the month.
* Despite minor fluctuations, there are no consistent patterns indicating specific days with significantly higher crime rates, implying variability in daily crime occurrences without discernible trends.

1. Number of Crimes By Month



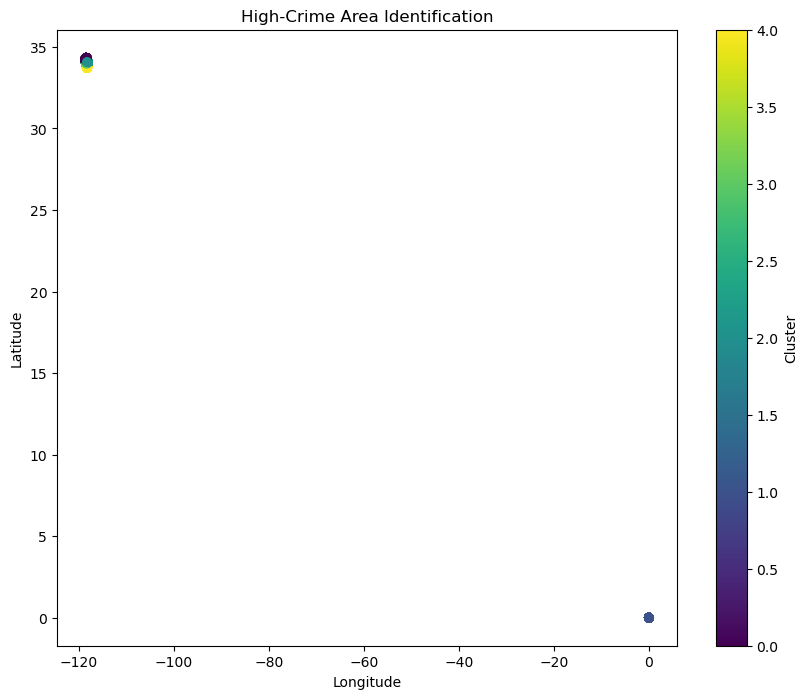
**Observations:**

• Data analysis reveals that January (Month 1) demonstrates the highest crime rate, surpassing 80,000 incidents, indicating elevated criminal activity at the onset of the year.

• Crime counts remain relatively steady around 60,000 from May (Month 5) to September (Month 9), indicating a period of consistent crime rates during the summer months.

• There is a slight uptick in crime counts in October (Month 10) and November (Month 11) followed by a subsequent decline in December (Month 12), suggesting seasonal fluctuations in criminal activity towards the end of the year

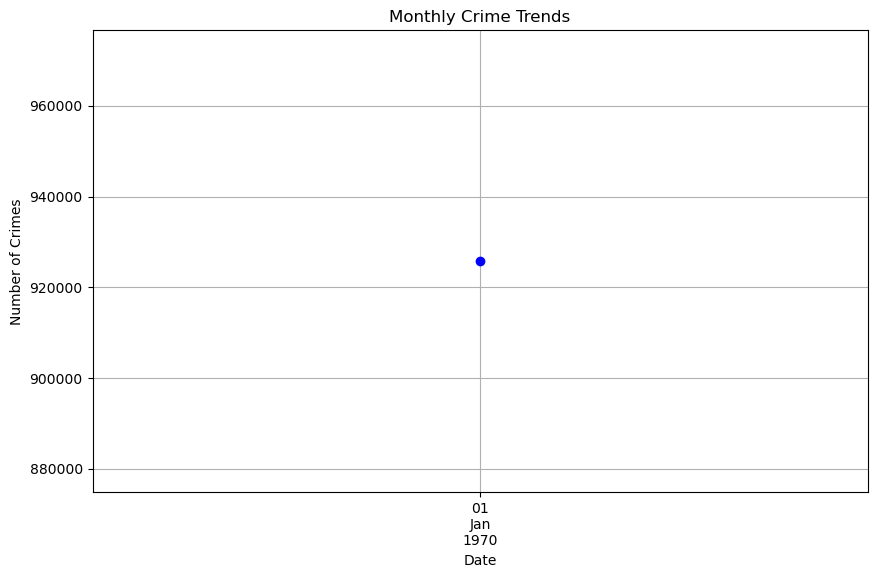
1. **Graph showing crime intensity based on geographical coordinates**



**Observations:**

* The chart showcases longitude and latitude coordinates, with the x-axis representing longitude spanning from roughly -120 to 0, and the y-axis representing latitude ranging from 0 to 35.
* Two distinct data points are noticeable: one positioned close to the upper left corner at around (-120, 35), and another at the origin (0, 0).
* The color bar on the right illustrates clustering values ranging from 0 to 4. The data point at (-120, 35) exhibits a higher cluster value, denoted by its color, compared to the data point at (0, 0).

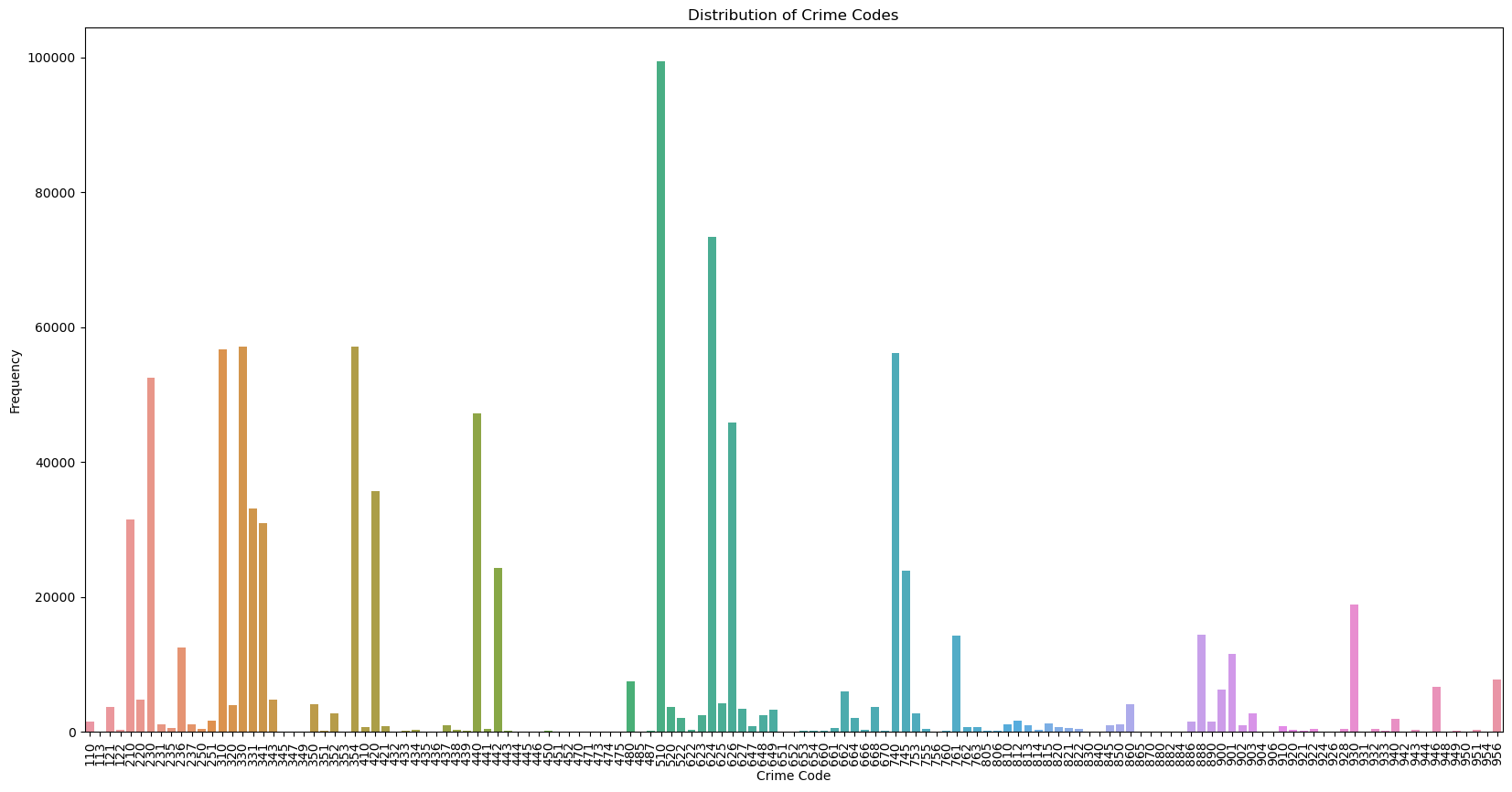
1. **Monthly Crime Trends**



**Observations:**

* Within the "Monthly Crime Trends" graph, a lone data point is illustrated for January 1, 1970, depicting roughly 920,000 reported crimes.
* The absence of multiple data points poses a challenge in identifying trends or conducting comparative analyses regarding crime occurrences over time.

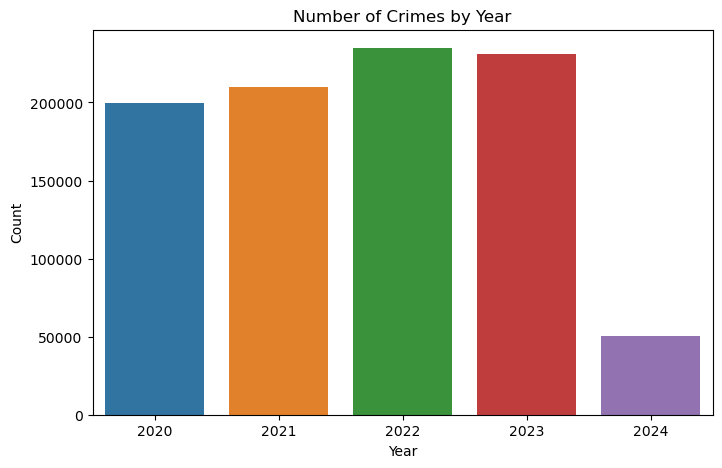
**5. Understanding Crime Patterns: Frequencies of Different Crime Codes in Los Angeles**



**Observations:**

* Some crime codes demonstrate markedly higher frequencies, signaling urgent areas of focus for law enforcement and policymakers
* .Examining temporal patterns can unveil fluctuations in crime rates over time, spotlighting periods of heightened or diminished criminal activity related to specific crime codes.
* Through pinpointing crime hotspots using geographical data, law enforcement can strategically allocate resources and enact focused interventions to combat prevalent criminal activities in susceptible areas

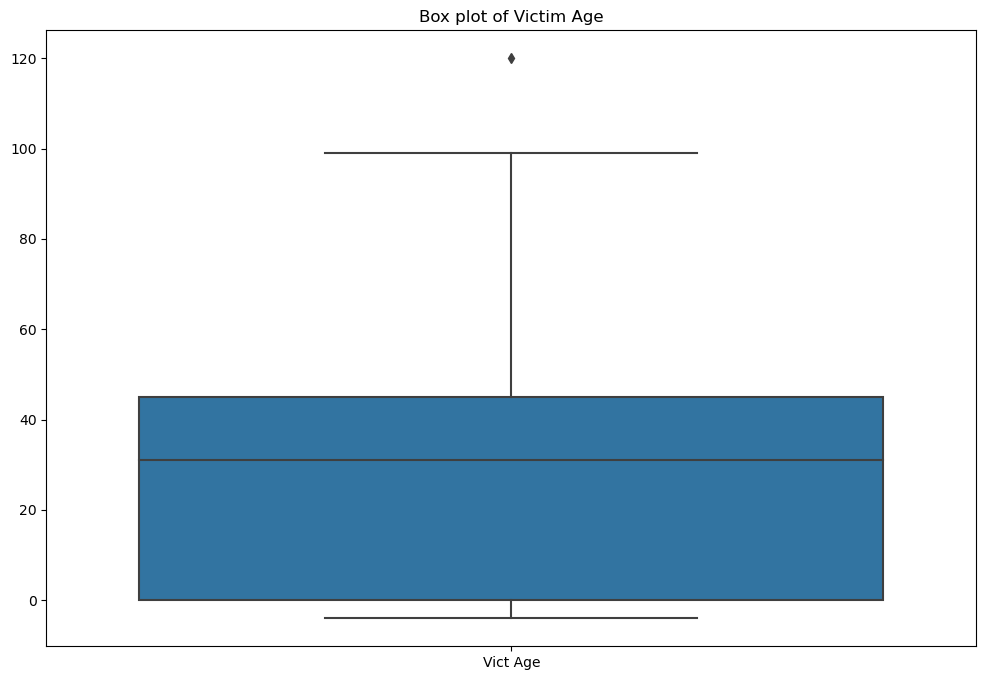
1. Number of Crimes By Year



**Observations:**

* Consistent Crime Rates (2020-2023), Throughout the period from 2020 to 2023, the incidence of crimes maintains a steady pattern. Each year records approximately 200,000 reported incidents.
* Drastic Reduction in 2024, Conversely, in 2024, there is a notable decline in crime rates. The tally experiences a dramatic plunge to below 50,000 incidents

**7. Box plot illustrating the distribution of victim ages**



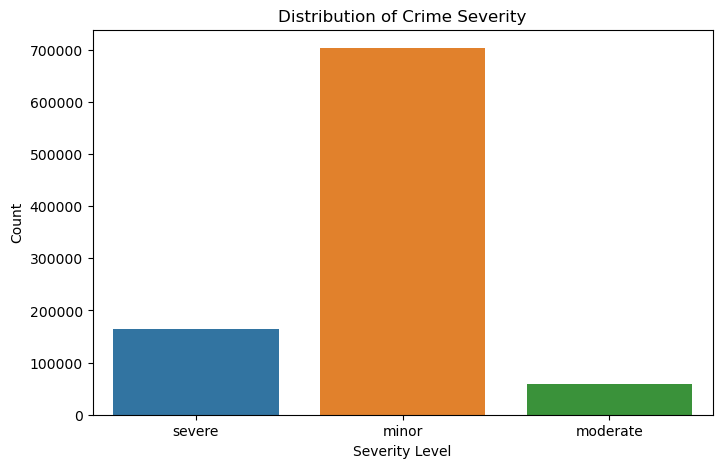
**Observations:**

• The median age of victims stands at approximately 40 years, indicating a relatively even distribution of ages among victims.

• The middle 50% of victim ages range from around 20 to 60 years, representing the typical age range of victims.

• An outlier exceeding 100 years underscores an exceptional case, underscoring the importance of accounting for extreme instances in victim age analysis

8. Severe crimes are outnumbered by minor offenses



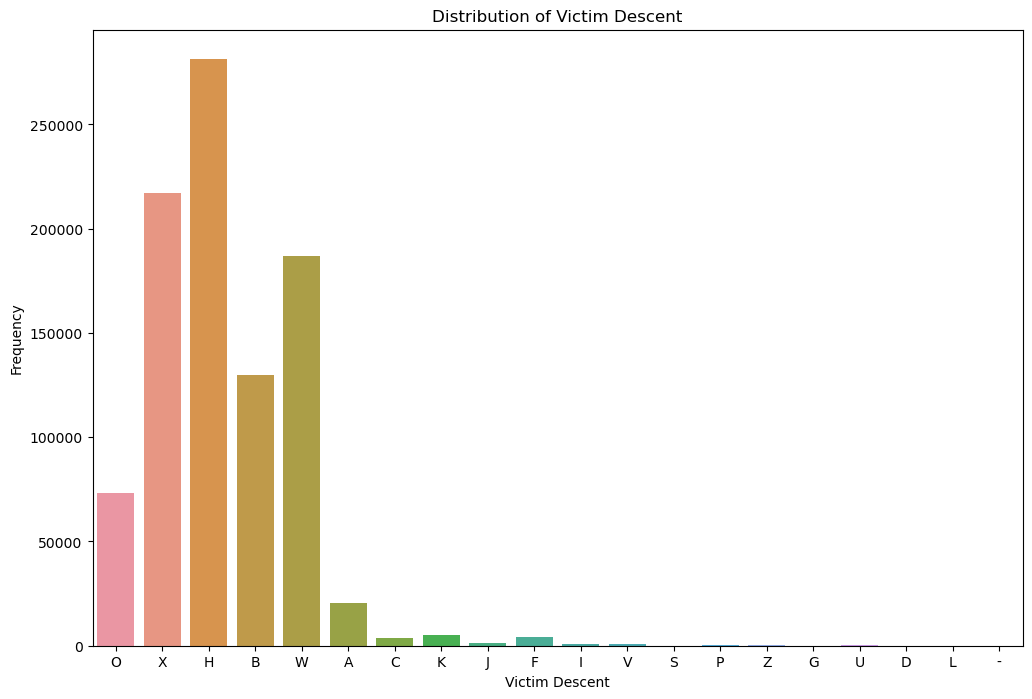
**Observations:**

• The bar graph illustrates the distribution of crime severity within the LAPD crime dataset

. • Roughly 100,000 incidents are classified as severe crimes, indicating a significant occurrence of high-severity offenses.

• More than 600,000 incidents are categorized as minor crimes, underscoring the prevalence of lower-severity offenses within the dataset.

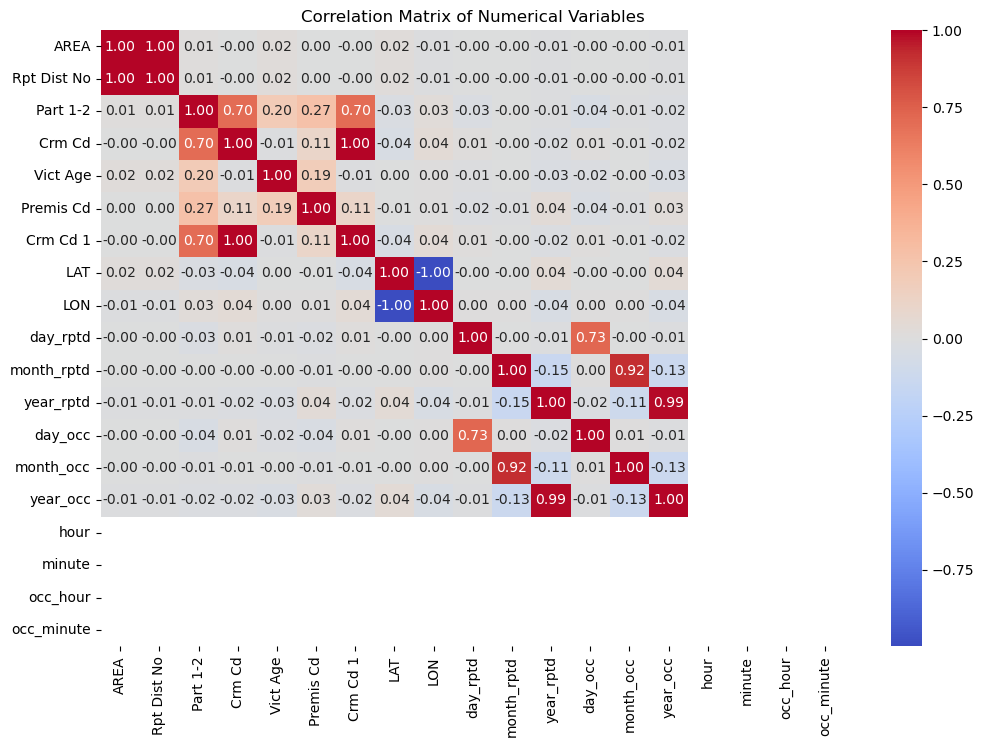
**9. The bar graph highlights victim descent frequencies**



**Observations:**

* The 'H' victim descent category prevails in the dataset, exceeding 250,000 occurrences, indicating its significant presence.
* Following closely, the 'O' category demonstrates a substantial frequency, just shy of 200,000 occurrences, emphasizing its importance
* .Categories 'B' and 'W' show similar frequencies, both hovering around 150,000 occurrences, while 'A' registers slightly below 50,000 occurrences. Conversely, other descent categories (C, K, J, F, V, S, P, Z, G, U, D, L) exhibit notably lower frequencies, suggesting their comparative rarity among victim demographics.

10. Strong positive correlations between reporting year/month and occurrence day/reporting day are evident.



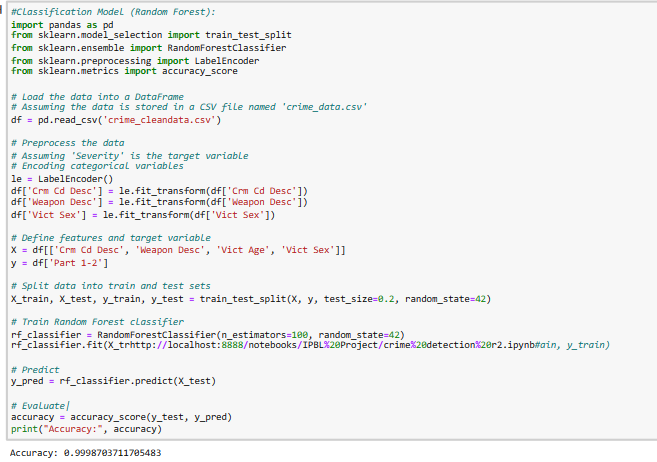
**Observations:**

• "year\_rptd" and "month\_rptd" display a robust positive correlation of 0.92, while "day\_occ" and "day\_rptd" demonstrate a correlation of 0.73, indicating close connections between reporting year/month and occurrence day/reporting day.

• The majority of other correlations are weak, with values nearing zero. Variables such as "LAT," "LON," "hour," and "minute" exhibit insignificant correlations with other numerical variables

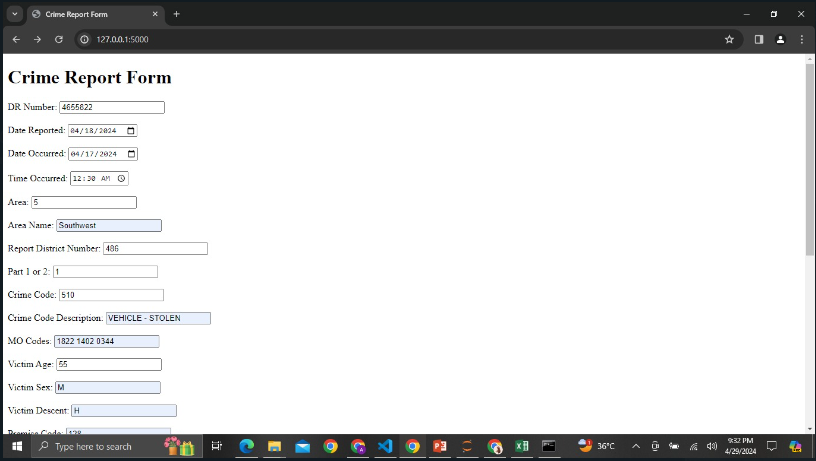
• The correlation matrix underscores notable relationships between specific variables but primarily reveals weak correlations, implying limited linear associations among the numerical variables in the LAPD crime data.

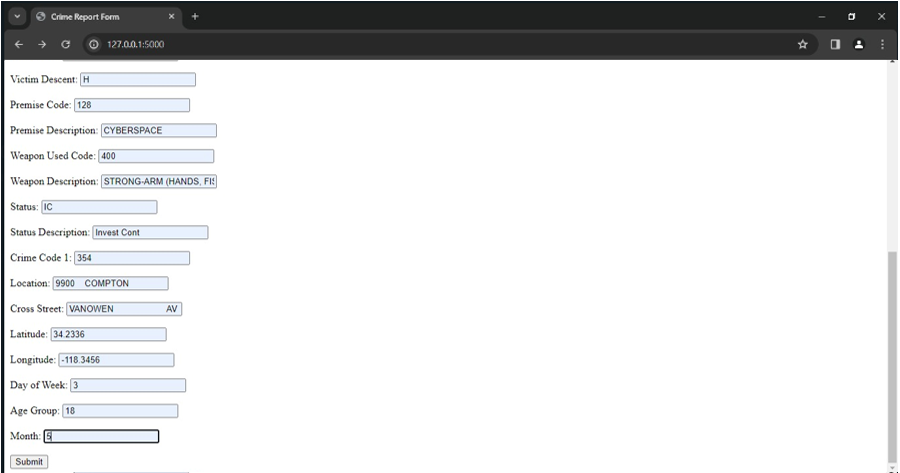
1. **Final Implementation**
   1. **Random Forest**

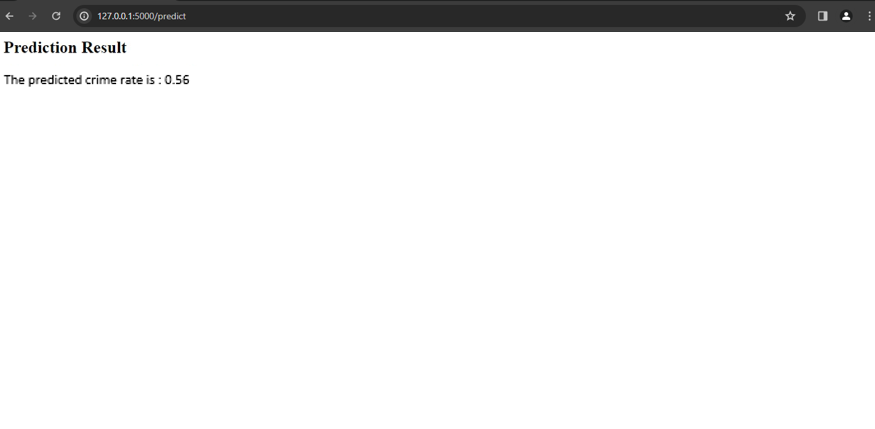


Accuracy: 0.9998703711705483

Observation**:** The identified features ('Crm Cd Desc', 'Weapon Desc', 'Vict Age', 'Vict Sex') exhibit robust predictive capability in discerning between 'Part 1' and 'Part 2' crimes. The model's elevated accuracy implies that these chosen features encapsulate valuable insights for categorizing crimes into the specified groups. Leveraging the Random Forest algorithm proves advantageous for this classification task, as it effectively manages a plethora of features while maintaining performance integrity without overfitting. The model's high accuracy suggests its potential utility for law enforcement agencies and policymakers in both predicting and preempting various types of crimes. Overall, the Random Forest classification model demonstrates commendable efficacy in precisely categorizing crimes based on the designated features.







**7. Conclusion**

In summary, our examination of LAPD crime data has yielded valuable insights into the dynamics of crime in Los Angeles. We've discerned seasonal and daily patterns, charted high-crime zones, and delved into the interplay between various crime types. Demographic analysis has shed light on disparities in victimization rates, while geospatial visualization has facilitated resource allocation efforts. These discoveries underscore the necessity for tailored interventions to tackle crime and enhance public safety in Los Angeles. Ongoing research and cooperation are essential for crafting effective strategies to confront these multifaceted challenges..

**8. Future Scope**

Looking ahead, the project can delve into advanced predictive modeling techniques and incorporate socio-economic data to achieve a more comprehensive grasp of crime dynamics. Initiating community engagement programs, implementing real-time data analysis systems, and assessing intervention strategies can amplify the effectiveness of crime prevention endeavors. Longitudinal studies and cross-city comparative analyses offer promising avenues for gaining insights into the evolving nature of crime patterns and devising tailored interventions. Embracing technological advancements such as artificial intelligence and IoT sensors holds the potential to further refine crime analysis capabilities, enabling more targeted and proactive law enforcement approaches. These future directions have the potential to bolster ongoing endeavors aimed at enhancing public safety and curbing crime rates in Los Angeles