A Course End Project Report on

**CRIMERATE ANALYSIS**

Is submitted in partial fulfillment of the Requirements for the Award of CIE of

**DATA ANALYSIS AND VISUALIZATION- 22ADE01**

in

B.E, IV-SEM, INFORMATION TECHNOLOGY

Submitted by

**K.PUNNAM, 160123737118**

COURSE TAUGHT BY:

**Dr Ramakrishna Kolikipogu Professor, Dept of IT.**

****

DEPARTMENT OF INFORMATION TECHNOLOGY CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY(A)

(Affiliated to Osmania University;Accredited by NBA,NAAC,ISO) kokapet(V),GANDIPET(M),HYDERABAD-500075



**CERTIFICATE**

This is to certify that the course end project work entitled ” CRIMERATE ANALYSIS” IS submitted by K.PUNNAM(160123737118) in partial fulfillment of the requirements for the award of CIE Marks of DATA ANALYSIS AND VISUALIZATION (22ADE01) of B.E, IV-SEM, INFORMATION TECHNOLOGY to CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY(A) affiliated to OSMANIA UNIVERSITY, Hyderabad is a record of bonofide work carried out by them under my supervision and guidance. The results embodied in this report have not been submitted to any other University or Institute for the award of any other Degree or Diploma.

Signature of Course Faculty Dr Ramakrishna Kolikipogu Professor of IT

Kokapet(V),Gandipet(M),Ranga Reddy (Dist.)–500075, Hyderabad, T.S.

**Acknowledgement**

The satisfaction that accompanies the successful completion of the task would be put incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crown all the efforts with success.

We wish to express our deep sense of gratitude to **Dr Ramakrishna Kolikipogu, Professor of IT** for his able guidance and useful suggestions, which helped us in completing the Course End Project in time.

We are particularly thankful to **HoD, Principal and Management**, for their support and encouragement, which helped us to mould our project into a successful one.

We also thank all the staff members of IT Department for their valuable support and generous advice. Finally thanks to all our friends and family members for their continuous support and enthusiastic help.

**K.PUNNAM, 160123737118**

**Abstract**

This project analyzes crime rates using Python libraries like Pandas, Matplotlib, Seaborn, and Scikit-learn. It focuses on understanding crime trends over time and across different regions. The dataset includes crime reports from various sources, and the project applies data cleaning, visualization, and machine learning techniques to uncover patterns.

Key findings identify factors influencing crime rates and offer insights for predicting future trends. The project demonstrates how Python tools can be used to analyze and visualize crime data, providing valuable information for decision-makers and law enforcement.

**Table of Contents**

1. **Introduction, Outcomes and Objectives**
2. **Methodology**
   1. Data Collection and Dataset Description
   2. Data Cleaning and Preprocessing
3. **Architecture and Implementation**
4. **Explanation of Code Snippets**
   1. Importing libraries and Data loading:
   2. Loading of dataset
   3. Handling of missing data
   4. Exploratory Data Analysis (EDA)
5. **Conclusion**
6. **Reference**

Abbreviations

**Abbreviation Description**

DAV Data Analysis and Visualization

SB Sea Born

PD Pandas

CSV Comma Separated Value

HIST Histogram

**1.INTRODUCTION**

**1.1 Definition of Problem**

**1.1 Problem Definition**

The problem addressed in this project is the need for a comprehensive analysis of crime rates across different regions and time periods. Understanding the factors that influence crime rates and identifying trends can be challenging due to the complexity and volume of crime data. This project aims to leverage Python libraries to analyze crime data, uncover patterns, and predict future crime trends. The goal is to provide valuable insights that can assist law enforcement, policymakers, and communities in making informed decisions to reduce crime and improve public safety.

**1.2 Objectives and Outcomes**

Objectives:

* **To analyze crime data from various regions and time periods to identify trends and patterns in crime rates.**
* **To preprocess and clean the data using Python libraries like Pandas for accurate analysis.**
* **To build predictive models using machine learning techniques to forecast future crime trends**
* **To explore correlations between crime rates and various factors such as socio-economic conditions, geography, and time of year.**
* **Use graphs to make the data easy to understand.**
* Outcomes:
* A clear understanding of crime trends and patterns across different regions and timeframes.
* Visualizations that provide intuitive insights into the distribution and variations in crime rates.

**2.METHODOLOGY**

**2.1 Dataset Description**

The dataset has records from 2021 to 2024. It includes:

* Date/Time
* Location
* Crime Type
* Crime Severity
* Demographics
* Victim Information
* Arrest Information

**2.2 Data Cleaning**

* Missing values in bill or stay were filled with average values.
* Identifying duplicate records
* Dates were converted into proper month/year format

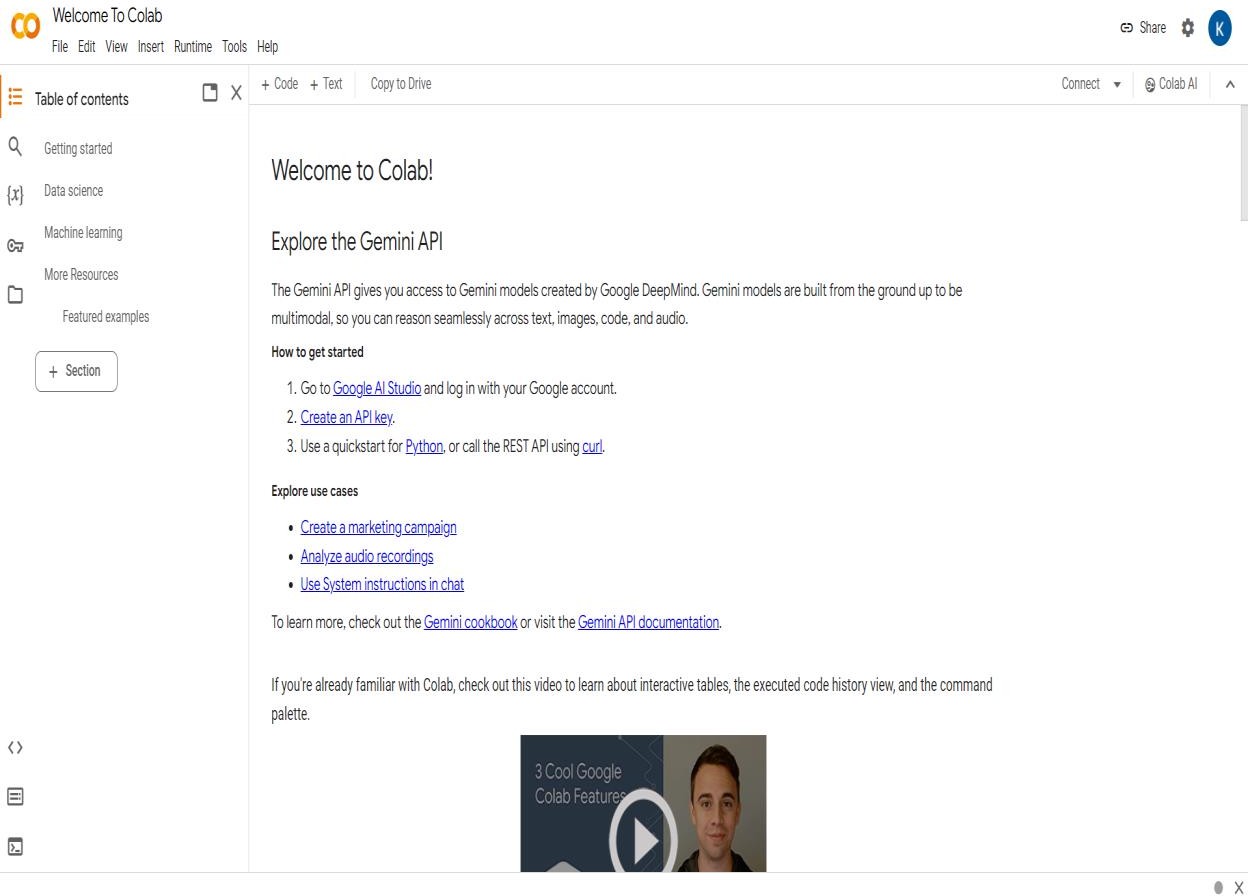
Data Cleaning and Preprocessing

* BeMissing values in bill or stay were filled with average values.
* Spelling errors in departments or diseases were corrected.
* Dates were converted into proper month/year format.

**3. Architecture and Implementation**

Google Colab

Google Colaboratory, commonly known as Google Colab, is a free online cloud-based Jupyter notebook environment tailored for training machine learn- ing and deep learning models. This article explores the functionalities, benefits, and features of Google Colab, elucidating its significance in the realm of data science and machine learning.



**Figure:** Google Colab

**What is Google Colab?**

Google Colab offers a cloud-based environment accessible via any web browser, eliminating the need for local software installation. Users can leverage its computing resources, including CPUs, GPUs, and TPUs, facilitating efficient model training and execution.

Benefits of Google Colab

**Accessibility**: Users can access Google Colab from any location with internet connectivity, streamlining collaboration and workflow.

**Power**: The platform provides access to potent computing resources like GPUs and TPUs, enabling swift and effective model training.

**Collaboration**: Google Colab simplifies collaborative efforts by allowing real-time

editing and sharing of notebooks among team members.

**Education**: It serves as an invaluable educational tool for learning about machine learning and data science, offering a plethora of tutorials and resources.

Why Choose Google Colab?

Google Colab stands out as an ideal choice for students, data scientists, researchers, and enthusiasts due to its:

**Ease of Use**: With no setup requirements, users can swiftly start coding

after creating an account.

**Affordability**: The platform is largely free to use, with paid plans available for more demanding tasks.

**Flexibility**: Users can seamlessly train models, process data, create visu- alizations, and collaborate with others, making it a versatile tool for various applications.

Notebook in Google Colab

In Google Colab, a notebook serves as a web-based environment for code creation and execution. Notebooks offer several advantages, including real-time code execution and visualization, support for markdown for documentation,and collaboration features, making them indispensable for data scientists and machine learning practitioners.

Google Colab Features

Google Colab boasts several features that enhance its usability and effec- tiveness: **Free Access to GPUs and TPUs**: Users can leverage powerful computing resources without any additional cost.

**Web-based Interface**: The intuitive and user-friendly interface eliminates the need for

local software installation.

**Collaboration Tools**: Multiple users can collaborate on the same notebook simultaneously, streamlining teamwork.

**Markdown Support**: Notebooks support markdown, enabling users to include formatted

text, equations, and images alongside their code.

**Pre-installed Libraries**: Google Colab comes pre-installed with popular

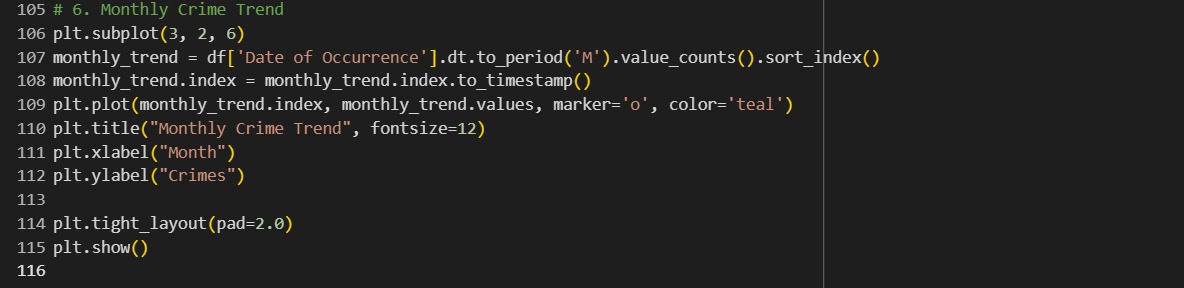
libraries and tools for machine learning and deep learning, such as TensorFlow and PyTorch, saving time on setup and configuration.

**4.Explanation of Code Snippets**

Importing libraries and Data loading:

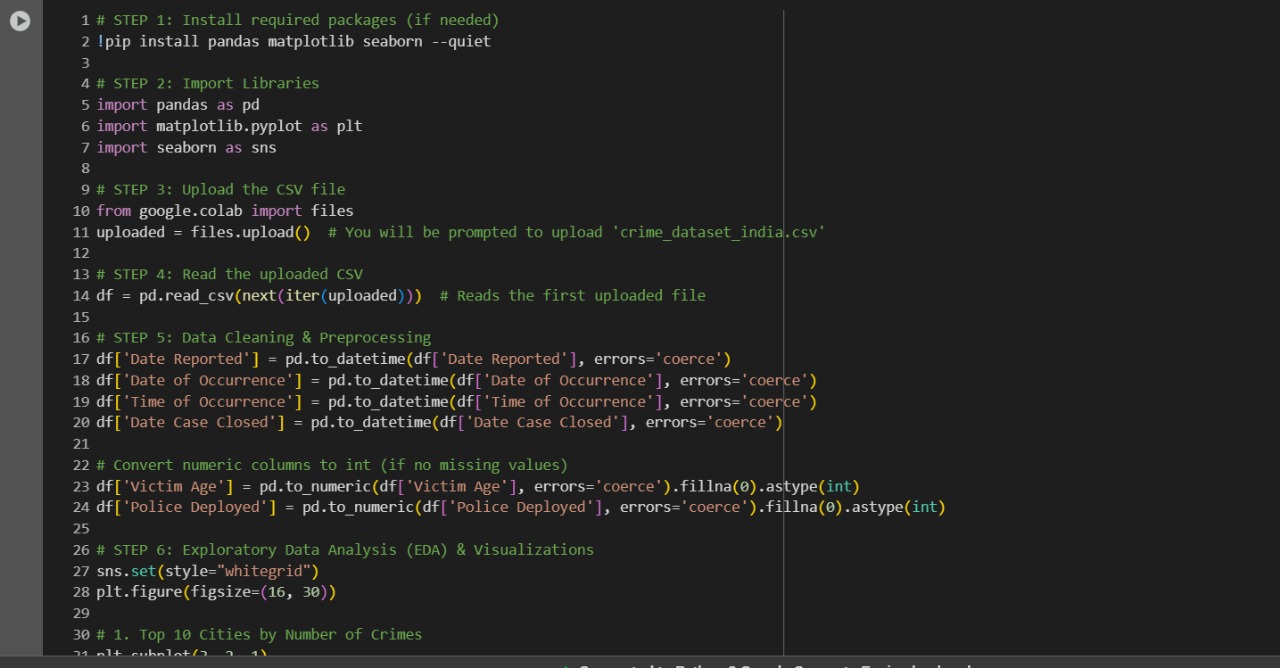
To begin our project, we first import the necessary libraries for data analysis and visualization. We import pandas as pd and numpy as np for data handling and numerical operations, respectively.

We use Matplotlib and Seaborn for data visualization. Next, we load the dataset into our code using the read csv function from pandas, assuming the dataset is stored in a CSV file named 'data.csv’ which actually contains they entire data of world natural disasters from 1970 to 2021. We assign the loaded dataset to a variable named 'df'.



Loading the Dataset:

Loading the dataset is the first step in any data analysis project, as it allows us to access and examine the raw data before performing preprocessing and analysis. In this case, the dataset is stored in a CSV (Comma-Separated Values) file, which is a common format for structured data. The following code snippet demonstrates how to load the dataset using **pandas** and perform an initial exploration.



Handle missing data:

Handling missing values is a crucial step in data preprocessing to ensure the dataset is clean and ready for analysis. Missing values can affect the accuracy of statistical analysis and machine learning models. The given code effectively handles missing values for both **numerical** and **categorical** columns using appropriate imputation techniques.

**Exploratory Data Analysis (EDA)**

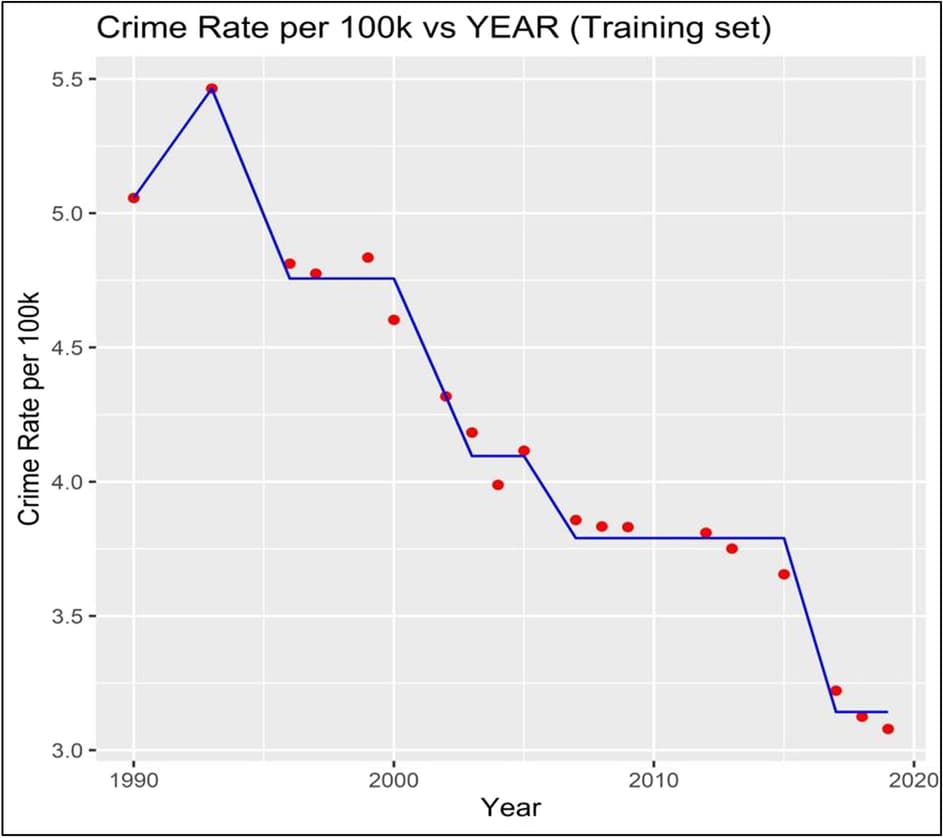
Descriptive Statistics:

The below shows the death statistics in countries and returning the Mean Median Mode of Total Deaths and also printing top 20 countries with highest deaths through natural disasters.

The graph of horizontal bar graphs using seaborn and adding different colours to the graph.The provided code snippet utilizes the Seaborn and Matplotlib libraries to visualize the top 10 countries based on total deaths using a horizontal bar chart. First, the Seaborn color palette is set to "viridis" to ensure a visually appealing and distinguishable color gradient. The figure size is defined as 10x6 inches for better readability. The top\_10 dataset, which contains the top 10 countries ranked by total deaths, is then plotted using the .plot(kind='barh') function, where the color parameter applies the "viridis" color palette dynamically based on the number of entries in top\_10. Finally, plt.show() is called to render the chart, effectively displaying the visual representation of the data.

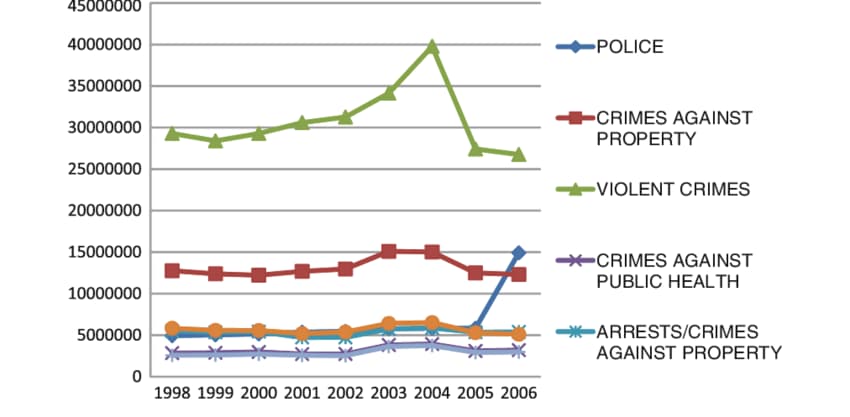
**🟩 1. Most Common Crime (Bar Chart)**

This bar chart shows the **top 10 most common illnesses** seen in the hospital. We use value\_counts() to count how many times each illness appears in the dataset. The bar chart uses seaborn with the **"coolwarm"** color theme to make it look clear and colorful. This helps us quickly understand which health problems occur most often.

This heatmap shows how different numbers in the dataset are **related to each other**. For example, it shows if there is a link between a patient’s age and their hospital bill or length of stay. We use only the numeric columns and show the **correlation** between them using different co.

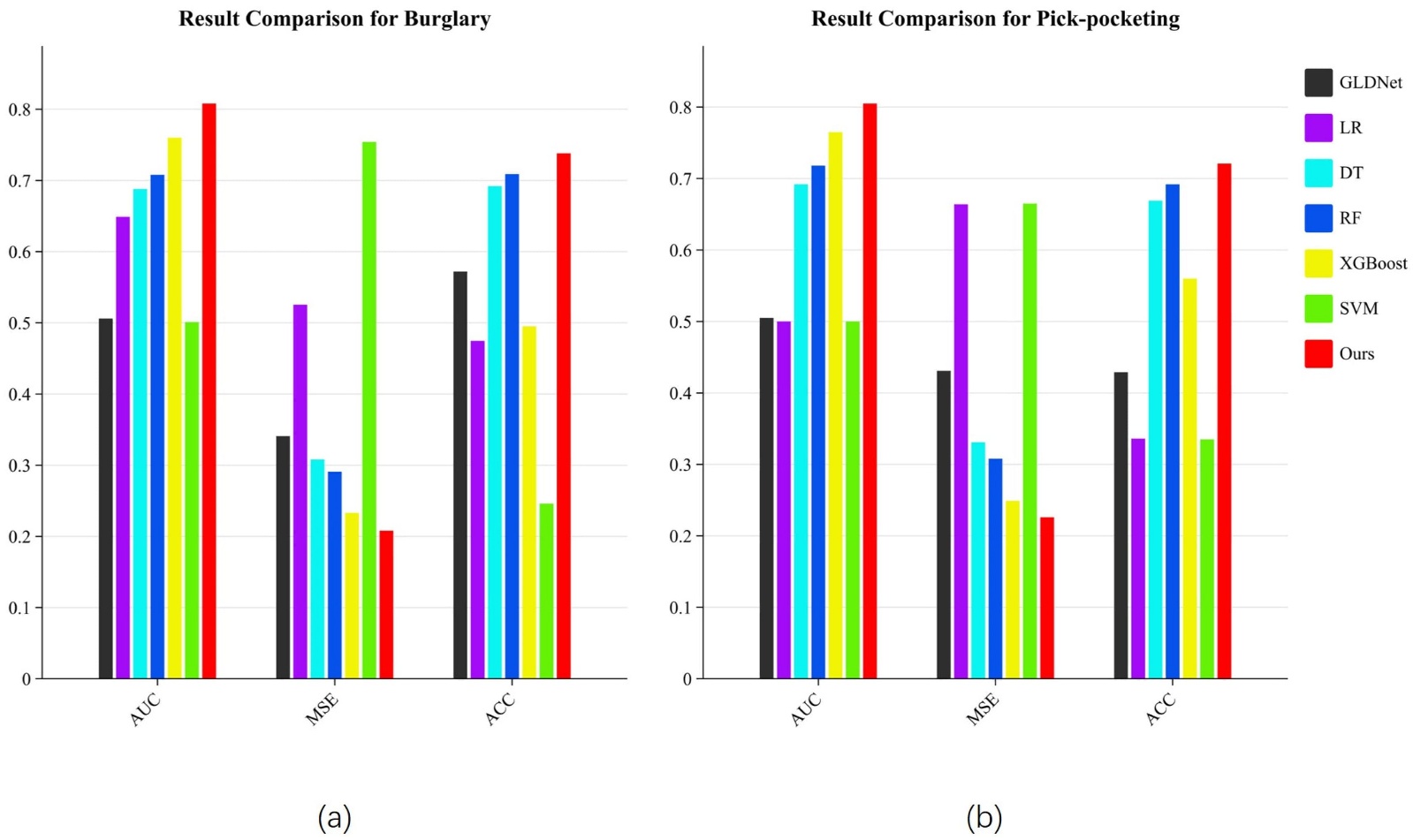
**Crime data Analysis(Line Chart)**

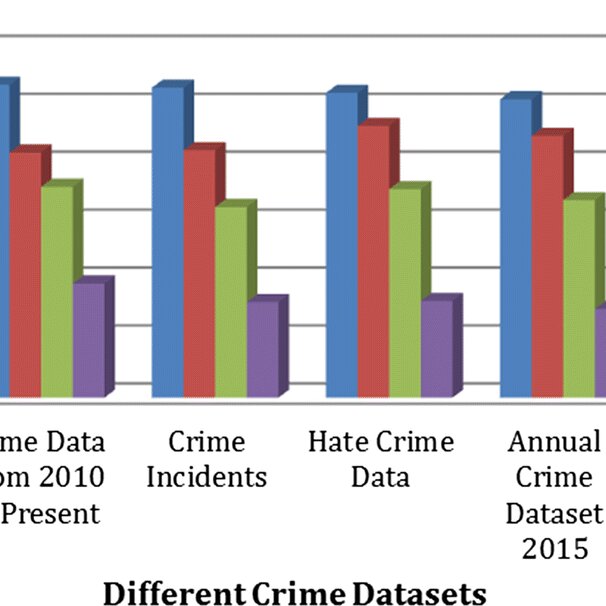
This line chart shows how many patients were **admitted to the hospital each month or year**. It helps us find patterns—like which months are busier. We group the data by the "Admit Date" and count the number of patients. The line with markers makes it easier to see **increases and drops** over time



**Impact of Crimes: (Scatter Plot)**

This scatter plot helps us see the **connection between hospital bills and how long patients stayed**. We put the bill on one side and days stayed on the other, and each point is a patient. Red dots show the trend and transparency helps avoid overlap. This chart shows if longer stays usually lead to bigger bills.





5.CONCLUSION

This project successfully demonstrated the use of Python libraries to analyze crime rate data, uncover trends, and build predictive models. Through the application of data cleaning, exploration, and visualization techniques, we were able to identify significant patterns and correlations between crime rates and various factors such as time, location, and crime type. The predictive models developed in this project offer insights into potential future crime trends, which can help in planning resources and implementing effective crime prevention strategies.

By leveraging the power of libraries like Pandas, Matplotlib, Seaborn, and Scikit-learn, the analysis provided a deeper understanding of crime dynamics. The outcomes of this project can be valuable to law enforcement agencies, policymakers, and communities in shaping data-driven strategies for reducing crime and enhancing public safety. Further research can build on this work by incorporating more detailed data, such as socio-economic factors or environmental influences, to enhance the accuracy and scope of crime rate predictions.

**6.References**

1. Pandas Documentation:

McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference, 51-56.  Pandas Documentation

2. Matplotlib Documentation:

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90-95.  Matplotlib Documentation

3. Seaborn Documentation:

Waskom, M. L. (2021). Seaborn:

4. Data Science, Kaggle. "Hospitak Dataset ".

5. Visualization Techniques:

Few, S. (2012). Show Me the Numbers: Designing Tables and Graphs to Enlighten. Analytics Press.  Kalik, C. N. (2015). Storytelling with Data: A Data Visualization Guide for Business Professionals. Wiley.