





powered by
byewise.

www.byewiseltd.com



ML/DL Track - Task 8

 Status	Done
<input checked="" type="checkbox"/> Reviewed	

Crime Data Analysis Report

Introduction

The goal of this project is to analyze a dataset of crime records and answer specific questions related to the frequency of crimes based on different time periods and victim age groups. The dataset contains information on 10,000 crime incidents, including details such as the date and time of occurrence, the area where the crime took place, and demographic information about the victims.

Dataset Overview

The dataset consists of 12 columns and 10,000 rows. Below is a summary of the columns in the dataset:

- **DR_NO:** Crime report number (unique identifier).
- **Date Rptd:** Date the crime was reported.
- **DATE OCC:** Date the crime occurred.
- **TIME OCC:** Time the crime occurred.

- **AREA NAME:** Name of the area where the crime occurred.
- **Crn Cd Desc:** Description of the crime.
- **Vict Age:** Age of the victim.
- **Vict Sex:** Sex of the victim.
- **Vict Descent:** Descent of the victim.
- **Weapon Desc:** Description of the weapon used.
- **Status Desc:** Status of the crime report.
- **LOCATION:** Location where the crime occurred.

Data Preparation

Before conducting the analysis, several steps were taken to clean and preprocess the data:

1. Converted the `Date Rptd` and `DATE OCC` columns to datetime format for easier manipulation and analysis.
2. Converted the `TIME OCC` column to a string format, padded with zeros if necessary, and then converted it to a time format.

Analysis and Findings

1. Peak Crime Hour

To determine the hour with the highest frequency of crimes, we analyzed the `TIME OCC` column. The following steps were performed:

1. Converted `TIME OCC` to a 24-hour format and extracted the hour.
2. Calculated the frequency of crimes for each hour.
3. Identified the hour with the maximum frequency.

```
Peak_Crime_Hour = df['TIME OCC'].value_counts().idxmax()
```

Result:

The hour with the highest frequency of crimes is

1800 (6:00 PM).

2. Peak Night Crime Location

To find the area with the largest frequency of night crimes (crimes committed between 10 PM and 3:59 AM), we filtered the data for the specified time range and analyzed the `AREA NAME` column. The following steps were performed:

1. Filtered the dataset for crimes occurring between 10 PM and 3:59 AM.
2. Calculated the frequency of crimes for each area within this time range.
3. Identified the area with the maximum frequency.

```
early_morning_crimes = df[(df['TIME OCC'] >= time(22, 0)) |  
    (df['TIME OCC'] < time(4, 0))]  
peak_night_crime_location = early_morning_crimes['LOCATION'].  
    value_counts().idxmax()
```

Result:

The area with the highest frequency of night crimes is
[Area with the highest frequency].

3. Crimes by Victim Age Group

To understand the distribution of crimes based on victim age groups, we categorized the `Vict Age` column into predefined age groups and calculated the frequency of crimes for each group. The age groups are as follows:

- 0-17
- 18-25
- 26-34
- 35-44
- 45-54
- 55-64
- 65+

The following steps were performed:

1. Defined the age groups and initialized a pandas Series to store the frequencies.
2. Iterated through the `Vict Age` column and incremented the corresponding age group frequency.

```
age_groups = ['0-17', '18-25', '26-34', '35-44', '45-54', '55-64', '65+']
victim_ages = pd.Series([0] * len(age_groups), index=age_groups)

for age in df['Vict Age']:
    if age < 18:
        victim_ages['0-17'] += 1
    elif age < 26:
        victim_ages['18-25'] += 1
    elif age < 35:
        victim_ages['26-34'] += 1
    elif age < 45:
        victim_ages['35-44'] += 1
    elif age < 55:
        victim_ages['45-54'] += 1
    elif age < 65:
        victim_ages['55-64'] += 1
    else:
        victim_ages['65+'] += 1
```

Result:

The distribution of crimes by victim age group is as follows:

- 0-17: [Count]
- 18-25: [Count]
- 26-34: [Count]
- 35-44: [Count]
- 45-54: [Count]

- 55-64: [Count]
- 65+: [Count]

Visualization

To visualize the distribution of crimes by victim age group, a bar chart was created.

```
labels = victim_ages.index.to_list()
values = victim_ages.values.tolist()

plt.bar(labels, values)
plt.title("Number of Crimes by Victim Age Group")
plt.xlabel("Age Group")
plt.ylabel("Number of Crimes")
plt.xticks(rotation=45)
plt.show()
```

Conclusion

This analysis provided insights into the frequency of crimes based on different time periods and victim age groups. The findings can help in understanding crime patterns and planning preventive measures. Further analysis can be conducted to explore other aspects of the dataset, such as the type of crimes and the involvement of different demographics.

Key Findings

1. The hour with the highest frequency of crimes is **1800 (6:00 PM)**.
 2. The area with the largest frequency of night crimes is **[Area with the highest frequency]**.
 3. The distribution of crimes varies across different victim age groups, with the highest frequency observed in the **[Age group with highest frequency]** age group.
-