DATA MINING PROJECT – SPRING 2022

(LAPD CRIME DATA ANALYSIS)

Introduction

This dataset reflects incidents of crime in the City of Los Angeles dating back to 2010. This data is transcribed from original crime reports that are typed on paper and therefore there may be some inaccuracies within the data. Some location fields with missing data are noted as (0°, 0°). Address fields are only provided to the nearest hundred block in order to maintain privacy.

LINK: https://drive.google.com/drive/folders/1ZIL-yCEkXWMWka-nyShq96lIJiswPUTK?usp=sharing

Deliverables

There are going to be a total of three deliverables for this project. The details of each deliverable is mentioned below.

Deliverable 1: Pre-processing & Exploratory Data Analysis (EDA) [40%]:

This deliverable is primarily focused on extracting as getting your hands dirty with the dataset. This will consists of data cleaning / pre-processing, initial data exploration, visualizations etc. This could include but not limited to the following points:

- Change types of columns as per requirements
- Find and handle missing values or incomplete rows
- Correlation between attributes
- Most frequently occurring incident
- District wise analysis based on number of incidents
- Using maps to visualize incidents

These are just some suggestions of tasks you can do. You need to figure out what other useful information can you figure out from this data. You need to elaborate your technique along with reasoning for each task you perform in your report.

Deliverable 2: Clustering and Frequent Pattern Mining [40%]:

- Perform cluster analyses on the data, primarily on location and offense type. Other clustering could be done based on time of day etc.
- Find frequent patterns in the dataset and explain your findings Your report for this deliverable must justify your choice of clustering algorithm.

 NOTE: You are free to implement your own algorithm or use any library to perform selected tasks.

Deliverable 3: Final Report [20%]:

 In this deliverable, you will be incorporating feedback from the previous deliverables and presenting your findings. You need to explain what information you have extracted from the data and suggest ways to help the police department overcome the rising crime. In your report, also mention other information, if any, that could be extracted from the data and how you plan to do that.