CRIME IN LOS ANGELES

2020 - PRESENT

TEAM MEMBERS: LEDY DE LA ROSA

DESCRIPTION

My project will analyze a dataset of crime data from Los Angeles, California, spanning from the year 2000 to the present. The dataset consists of 28 unique attributes such as the type of crime committed, area codes where it occurred , victim's descent, the timestamp of the crime, the time the crime was reported, and additional details.

I believe we can answer multiple questions with this information, but I've decided to specifically focus on the victims of these crimes. We often hear reports on the news or social media about crime rates, but rarely do we hear about who is more likely to be affected. I hope to answer questions such as:

- Which race is at risk for certain crimes?
- What age group tends to be a victim of a specific crime?
- Which area in Los Angeles has the most crime?
- The locations with higher murder rates, sexual assault, etc.
- On average, how long does it take to report a crime after it has been committed?

PRIOR WORK

There is a vast amount of previous work we can find on crime data from different parts of the world. Here are a few I gathered that aline perfectly with my project's goal:

1. 2020 - 2023 Los angeles crime data analysis with a focus on identifying patterns in criminal behavior.

   https://medium.com/@rpangarego/case-study-lapd-crime-data-from-2020-to-oct-2023-analysis-using-python-1d5d6dbf58f8

2. 2020 - 2023 LA data analysis with a focus on where/when crimes are happening and who are most vulnerable to such crimes.

   https://medium.com/@ptuan5/exploratory-data-analysis-los-angeles-crime-2020-2023-5adab44973c9

DATASETS

Dataset to use:

Crime_Data_from_2020_to_Present.csv

Where found?

Kaggle and provided by LA local police.

URL:

https://www.kaggle.com/datasets/middlehigh/los-angeles-crime-data-from-2000/data?select=Crime_Data_from_2020_to_Present.csv

Whether it you have it downloaded (on who's machine) ?

Yes, it has been downloaded on Ledy's Laptop.

PROPOSED WORK

Data cleaning: Fill in missing values and detect any outliers.

Dimensionality reduction: Remove attributes that do not add value to our research such as “Division of Records Number”

Data Mining: Extract insights from our dataset to answer our questions.

Data Visualization: Develop visual models that effectively convey our findings.

TOOLS

The tools I plan on using are the following:

1. Python
    a. Pandas
    b. Numby
2. Jupyter Notebook
3. Power BI

EVALUATION

I intend to evaluate my results by using cross-validation. This process involves comparing my results with other data available that resembles the one I used and check for consistency. Additionally, I plan on searching for errors by thoroughly cleaning the data and utilizing visualization tools which facilitate a quick and effective way of inspecting the data.