SCHOOL OF DATA by Springboard

Capstone Project Proposal: Crime Prediction for Areas within Los Angeles City

1. Problem Identification

Problem formulation statement

How can develop a predictive crime model for specific areas within Los Angeles Country within the next 12 years to help policy makers to enhance security, ensure the safety of locals, and assist the community in making informed decisions about the crime trend in certain areas.

Context

Los Angeles is the second-most populous city in the United States and has a high crime rate. According to the FBI's Uniform Crime Reports, Los Angeles has a crime rate of 2,759 per 100,000 people, which is higher than the national average of 2,580 per 100,000 people.By providing a data-driven crime prediction model , policy makers and certain communities can be empowered to implement targeted security measures and make informed decisions.

Criteria for Success

The success of this project will be measured by the accuracy of the predictive model and the usefulness of the model to the community.

Scope of Solution Space

The development of a machine learning model that integrates demographics, and historical crime reports to predict the likelihood of specific crimes occurring in specific areas within Los Angeles. The model will provide forecasts for different types of crimes, such as property crimes, violent crimes, and thefts.

Constraints

The project is constrained by the availability and quality of data.

Stakeholders

- The policymakers to understand crime trend.
- The community to be aware of crime patterns.

Data Sources

The data sources for this project will include the demographics, and historical crime reports in Los Angeles. The datasets will be obtained from the following resources:

- **Crime Data from 2010 to Present**(https://catalog.data.gov/dataset/crime-data-from-2020-to-present , https://data.lacity.org/Public-Safety/Crime-Data-from-2010-to-2019/63jg-8b9z/explore**)**

2. Project Approach

A machine learning model that predicts the likelihood of the most frequent crimes occurring in certain areas will be built. The model will be developed by combining past 12 years historical crime reports by area and victim characteristics.

Problem to Solve

The problem I want to solve is the uncertainty faced by new business owners in Los Angeles when choosing the location for their establishments due to a lack of real-time crime data.

Data Acquisition
I will acquire the necessary data from the provided sources:

- Crime data from 2012 to the present will provide historical crime reports.Crime Datasource includes Date Rptd,DATE OCC,TIME OCC,AREA,AREA NAME,Rpt Dist No,Part 1-2,Crm Cd,Crm Cd Desc,Mocodes,Vict Age,Vict Sex,Vict Descent,Premis Cd,Premis Desc,Weapon Used Cd,Weapon Desc,Status,Status DescLOCATION,Cross Street,LAT,and LON).

Solution Overview

My approach involves:

1. Data Waggling: Cleaning and integrating demographics, and historical crime reports, Crime Location,General Crime type,Occurred date will be used as primary data. And, matching crime location with current location to find out the where is the most common place that current businesses are operating.

2. Exploratory Data Analysis: Statistically exploring the dataset to understand its characteristics, patterns, and potential issues, and creating relevant features that capture the characteristics of crime areas.

3. Preprocessing & Training: Building a machine learning model (e.g., a predictive classifier or regressor) that utilizes historical crime data and area features to predict future crime rates.

4. Modeling: Assessing the model's performance using appropriate metrics such as MAE,MSE and RMSE.

Project Deliverable

A GitHub repo containing:

1. Notebooks & output models
    i. Data Wiggling notebook and cleaned datasets
    ii. EDA and Feature engineering work
    iii. Preprocessing & Training
    iv. Modeling and model output
2. Documentation
    i. A slide deck
    ii. A project report

In summary, this project aims to develop a data-driven solution that empowers governments or locals in Los Angeles to make informed decisions about the safety of their chosen locations by predicting crime rates in specific areas. This tool will not only enhance security but also contribute to the overall economic development of the city.