Austin Crime Data

Group 2 Final Project

Presentation Agenda

The presentation outlines the project, including the following:

- Selected topic
- Reason topic was selected
- Description of the source of data
- Questions the team hopes to answer with the data
- Description of the data exploration phase of the project
- Description of the analysis phase of the project
- Segment 3 - Technologies, languages, tools, and algorithms used throughout project

Topic: Austin Crime Data

Why did we pick this?

- Austin, TX is relative to our team
- Team-wide genuine interest in crime as a subject
- Curious how crime data can be leveraged with ML

Data Source

data.austintexas.gov
the official City of Austin open data portal

Timeframe of Crime	Criteria of Data	Dataset	Source and Ownership
2017 - 2022	This dataset contains a record of incidents where the Austin Police Department responded to calls for police service **where a report was written**.	• 200,000+ rows of data • 27 Columns • 3 tables	**Data Provided by:** Austin Police Department **Dataset Owner** APD PIO

Predictions we hope to answer with the data

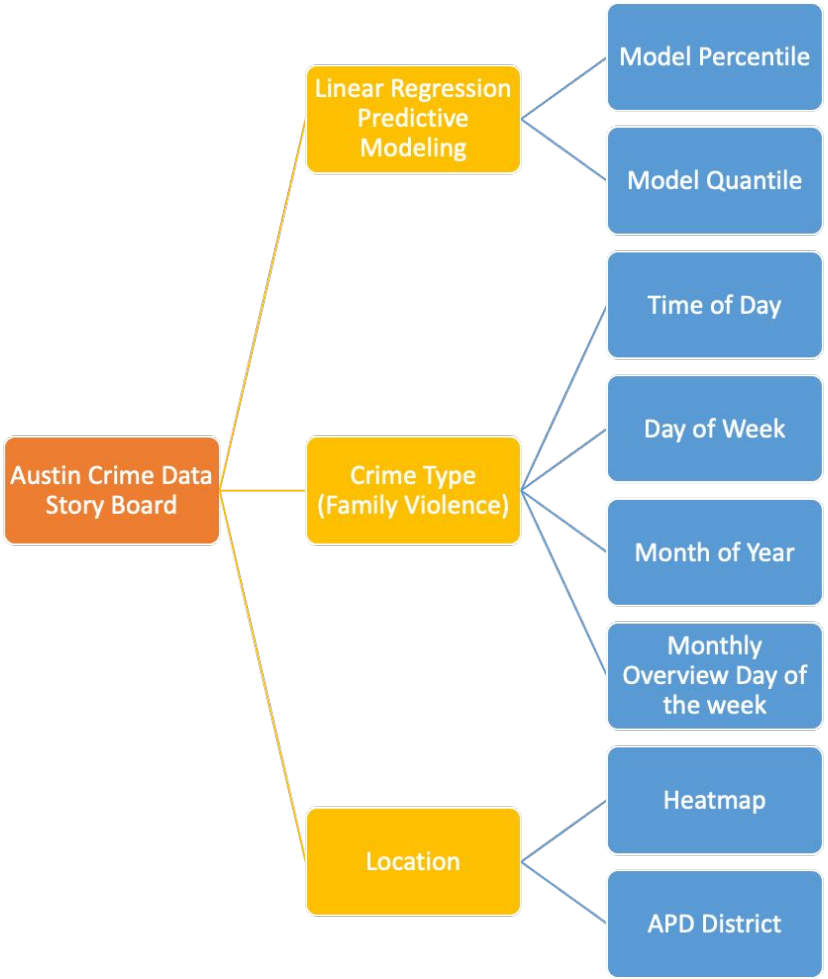
We hope to use machine learning models to accurately predict if a crime would be classified as Family Violence or not.

The rest of our analysis around the crime data will help us answer questions about crime that is most likely to occur in relation to:

- Logistic Regression Prediction Modeling
- Crime type
- Time of day crime occurs
- Day of the week
- Month of the year
- Location

Dashboard Blueprint

- Dashboard & Storyboard will be created with the aid of Tableau Public
- Interactive elements include:
  - Logistic Regression Predictive Modeling
  - Crime based
  - Location based



Database

Using Pandas, several columns were dropped to remove data that we didn't feel was applicable to our analysis. The dataset was then split into two tables in preparation of fulfilling the join requirement. The ERD was created in quickdatabasediagrams.com. Finally, in order to fulfil the requirement for joining two tables, the dataset that was previously split into two tables was joined back together using SQL Query.

Machine Learning Model

For our analysis we are using three different machine learning models. Each model's performance will be evaluated and compared to the others.

Preprocessing Steps:

- Drop rows with null values
- Drop unnecessary/redundant columns
- Encode categorical variables
- Split the dataset
- Scale the subsets

Feature Engineering and Selection:

- Split into features and the target array "Family Violence_Y"

Splitting the Data:

- Split into training and testing sets

Machine Learning Models:

- Logistic Regression
- Random Forest Classifier
- Neural Network

SEGMENT 3
DRAFT WORK

Technologies, languages, tools, and algorithms

Database

Machine Learning

01	Technologies	• Github • Google Colab • Jupyter Notebook • pgAdmin • Terminal / CMD prompt
02	Languages and Tools	• Python • SQLAlchemy • Scikit-learn libraries • TensorFlow
03	Algorithms	• Logistic Regression • Random Forest Classifier • Neural Networks

Github, Communication and Collaboration

01	Technologies	• Github • Slack • Google Docs & Slides • Jupyter Notebook • Terminal / CMD prompt
02	Languages	100% Jupyter Notebook
03	Algorithms	Step by step procedures: ○ Commits ○ Peer Reviews ○ Branch Merges