**Outline**

**First Segment - Decide on your overall project:**

Sketch It Out:

* Selected topic

For topic of the final project, we selected Burglary and Robbery crimes in the state of Colorado define crime indicators, patterns and use Machine Learning Models to predict when, and where a crime will take place based on the indicators.

* Reason the topic was selected

Crimes are intricate and interesting, and can be both predictable and unpredictable. Crimes have existed for a long time and will remain parts of society. Therefore, I am interested in predicting when and where crime could take place.

* Select your question -- Questions the team hopes to answer with the data

1. Can we predict the location, date and type of a crimes based on key indicators: historical crime data, incident year and number of crimes?
2. Can we predict types of crimes w with key indicators: historical crime data, incident year and number of crimes?
3. What crimes are increasing or decreasing, at what rate and over what timeframes with a given regional?
4. Can we confirm the findings against published statistics?

* Description of the source of data

The data was sourced from Colorado Crime Data Explorer site. The Burglary and Robbery crimes data were available as a download in comma separated values format from the year 2016 to 2019. Since the data is being sourced from the state of Colorado, the dataset will be small. Datasets from the rest of the states need to be incorporated to improve the outcomes of this project.

The following are the column names and their data types from NIBRS\_incident and NIBRS\_offense datasets:

NIBRS\_incident dataset:

DATA\_YEAR INT

AGENCY\_ID INT

INCIDENT\_ID INT,

NIBRS\_MONTH\_ID INT

CARGO\_THEFT\_FLAG VARCHAR

SUBMISSION\_DATE DATE

INCIDENT\_DATE INT

INCIDENT\_Month VARCHAR

INCIDENT\_DAY INT

REPORT\_DATE\_FLAG VARCHAR,

INCIDENT\_HOUR INT

DATA\_HOME VARCHAR

ORIG\_FORMAT VARCHAR

DID INT

NIBRS\_offense dataset:

DATA\_YEAR INT

OFFENSE\_ID INT

INCIDENT\_ID INT

OFFENSE\_TYPE\_ID VARCHAR

ATTEMPT\_COMPLETE\_FLAG VARCHAR

LOCATION\_ID VARCHAR

METHOD\_ENTRY\_CODE VARCHAR

Links:

NIBRS: https://crime-data-explorer.app.cloud.gov/downloads-and-docs

* Description of the data exploration phase of the project

In the initial data analysis, we explored the structured dataset to uncover initial patterns, characteristics, creating a broad picture of important trends and major points to study in greater detail.

Data Types Breakdown:

Nominal Data examples

NIBRS\_Offense\_16\_19.csv:

Variables:DATA\_YEAR,OFFENSE\_ID,INCIDENT\_ID,OFFENSE\_TYPE\_ID, ATTEMPT\_COMPLETE\_FLAG,LOCATION\_ID,METHOD\_ENTRY\_CODE

nibrs\_incidents\_17\_19.csv

Variables: DATA\_YEAR, AGENCY\_ID, INCIDENT\_ID

Numerical Data:

NIBRS\_Offense\_16\_19.csv:

Variables:

nibrs\_incidents\_17\_19.csv

Variables:

* Description of the analysis phase of the project

Which model did you choose and why?

Logistic Regression model is a classification algorithm so it is best applied to categorical data to predict discrete outcomes. This model’s algorithm will attempt to learn patterns from the data. If the model is successful, it will help predict if crime will take place? The datasets in this project contain categorical features that will be included in the machine learning algorithms. The categorical features will be converted to numerical data to use in the machine learning process. Also, the features needed with the Logistic Regression are pulled from two datasets which we will pull into a dataframe which are coded in the DataAnalysis\_phase jupyter notebook.

For the new dataframe we will be pulling the following features and targeting INCIDENT\_Month:

NIBRS\_incident dataset:

DATA\_YEAR INT

INCIDENT\_ID INT

INCIDENT\_DATE INT

INCIDENT\_Month VARCHAR

INCIDENT\_DAY INT

INCIDENT\_HOUR INT

NIBRS\_Offense dataset:

OFFENSE\_ID INT

OFFENSE\_TYPE\_ID VARCHAR

LOCATION\_ID VARCHAR

The following features will be encoded using python libraries:

INCIDENT\_Month VARCHAR

OFFENSE\_TYPE\_ID VARCHAR

LOCATION\_ID VARCHAR

Logistic Regression

Pearson Correlation

Random forest Classifier Result

Decision Tree

SVM

How are you training your model?

Logistic model analyzes the

What is the model's accuracy?

How does this model work?

* Result of analysis
* Recommendation for future analysis
* Anything the team would have done differently

**Deliverables will be posted via GitHub repository containing the following items:**

* Deliverables for the presentation
* GitHub
* Machine Learning Model

Pearson Correlation

Logistic Regression

Random forest Classifier Result

Decision Tree

SVM

* Database sections
* Technologies use

1. Use Python or Excel to clean, prepare, and explore the data as well as to complete the initial analysis.
2. Use database integration (Postgres, MongoDB, or SQLite) to store your cleaned data.
3. Use machine learning to enhance your topic.
4. Use Python libraries, JavaScript libraries (such as Data-Driven Documents, or D3, and Plotly), and Tableau can be used to create visuals to help tell your data story.
5. Use Tableau or JavaScript to build a dashboard to present your results.
6. Use Tools to prepare and deliver the final presentation to walk your class through your project, step by step.

**Second Segment:**

Build the Pieces: Train your model and build out the database you'll use for your final presentation.

**Deliverables will be posted via GitHub repository containing the following items:**

* Deliverables for the presentation

Outine the dashboard and create the presentation slide deck.

* GitHub
* Machine Learning Model
* Database sections

The data for the nibrs\_incident and nibrs\_incident tables have been successfully ingested via sqlalchemy library. The script is incorporated in the DataExploration\_phase.ipynb.

* Technologies use

Use Tableau to create both dashboard and presentation slide deck.

**Third Segment:**

Plug It In: Connect your final database to your model, continue to train your model, and create your dashboard and presentation.

**Deliverables will be posted via GitHub repository containing the following items:**

Deliverables for the presentation

GitHub,

Machine Learning Model, and

Database sections

Technologies use

**Fourth Segment:**

Put It All Together: Put the final touches on your model, database, and dashboard. Lastly, create and deliver your final presentation to your class.

**Deliverables will be posted via GitHub repository containing the following items:**

Deliverables for the presentation

GitHub,

Machine Learning Model, and

Database sections

Technologies use