Project Proposal
INFO 7390- Advance Data Science

Chicago Crime Rate

Presented By:

Birwa Galia (001824469)

Milony Mehta (001869360)

Shantanu Deosthale (001851612)

Overview:

1. Abstract:

Problem:

- Crime in Chicago has been tracked by the Chicago Police Department's Bureau of Records since the beginning of the 20th century.
- The city's overall crime rate, especially the violent crime rate, is higher than the US average.
- Chicago was responsible for nearly half of 2016's increase in homicides in the US, though the city's crime rates remain near historic lows.
- The reasons for the higher numbers in Chicago remain unclear.
- More than half of the US's average Crime comes from Chicago State.
- Chicago is ranked one in homicide rate as compared to other metropolitan cities such as Los Angeles and New York. Homicide remains more than 50% of the crime in Chicago.

Result:

- With the help of historical data, patterns, and variations there is possibility to interpret the reasons leading to increasing crime in Chicago.
- The police department will come to know which area needs more patrolling.
- With the analysis performed on the dataset, we can conclude which area is the safest by taking arrest rate into consideration.

2. Data:

- **Chicago Crime Data:** https://www.kaggle.com/currie32/crimes-in-chicago/data

Dataset:

- The Dataset showcases reported incidents of crime that have been occurred in the city of Chicago from 2001 to 2016.
- For this research project, we are using data provided by the open source Chicago Police Department's system.
- The entire dataset has 6.49M rows. Due to the size of this dataset, we decided to focus on data from 2008 to 2016.

3. Objective:

- The inspiration behind this project is to help Chicago Police Department to improvise and derive suitable measures to reduce the crime.
- The following measures are projected:
  1. Is arrest rate equivalent to the crime rate?
  2. Analyze Number of Crime by Month, Day and Date of the Year
  3. Deriving the most common locations of the crime which will be beneficial to the city based on the arrest made.
  4. Analyzing different types of crimes in Chicago
     4.1. How big is the increase in homicides?
  5. Forecasting the crime in Chicago for the year 2018.
- This dataset is used to correlate the types of the crimes occurred, number of criminals arrested and predicting the probability of crime occurrences at a given date and location.

4. Use Cases:

1. Demographic generated by the arrest rates:

By taking that into consideration, we can increase the security in that area thus contributing to reduce crimes in Chicago

2. Analyzing crime rates every year:

Providing us with way to predict which time is safe for the people to travel in and around the city.

5. Process Outline:

1. Data Preprocessing: Data Wrangling, handling missing values
2. Exploratory Data Analysis
3. Study of Supervised approaches and select the best model for prediction
4. Design of a pipeline and system to implement this approach and discussion on the system's capabilities
5. Deploy the Model on AWS
6. Build a web application to demonstrate the prediction.

6. Deployment Details:

1. Language: Python.
2. Pipeline: Sklearn / Airflow
3. Container: Docker
4. Cloud Tools/Platforms: AWS (Amazon Web Services) EC2