

# 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:

- 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.

#### 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.

#### Result:

- With the help of historical data, patterns, and fluctuations, a clear picture can be framed, about the reasons leading to increasing crime in Chicago.

During the Presidential campaign, Chicago crime number and its analysis are widely used.

### 2. Data

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

We will work with Chicago Crime data for the years 2001-2016

### 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. Predicting the relationship between weather, crime rate and arrest rate.
  4. Deriving the most common locations of the crime will be beneficial to the city.
  5. Analyzing different types of crimes in Chicago
    - 5.1. How big is the increase in homicides?
  6. Predicting the outcome if an arrest would occur or not.
  7. 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 crime 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 as per time of the year:

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

**4. Process Outline:**

## 1. Data Preprocessing:

Data Wrangling, handling missing values

## 2. Exploratory Data Analysis

Feature Engineering

## 3. Study of Supervised approaches and select the best model for prediction

Feature Selection

## 4. Design of a pipeline and system to implement this approach

## 5. Deploy the Model on Azure/AWS or Google Cloud Computing Platform

## 6. Build a web application to demonstrate the prediction.

**Deployment Details:**

## 1. Language: Python

## 2. Pipeline: Airflow

## 3. Container: Docker

## 4. Cloud Tools/Platforms: Microsoft Azure Machine Learning Studio, AWS (Amazon Web Services) EC2

## 5. Other Considerations: Google Cloud Platform