Project Proposal
Chicago Crime Data Analysis

Introduction and background of the Data:

In this world, crimes are an inseparable part of our lives. Every day we hear about them and some of us are even involved in at least one of them during our life. Being cautious and improve safety is not a simple instruction anymore. We need to use modern technology and data science techniques to more wisely act against this problem. There are so many records and documentation in the police department that have been gathered during the years, which can be used as a valuable source of data for the data analytics tasks. Applying analytical task to these data bring us valuable information that can be used to increase the safety of our society and lower the crime rate.

Objectives Of Study(Aim) :

The main idea behind this project is to create a user story of the crimes Dataset, which involved geographical analysis, Crime Data analysis and the use of machine learning models on the Chicago crime dataset. Analyzing and examining of crimes gives an understanding of crime regions and can used to take the precautions to reduce the crime rates.By identifying the patterns will allow us to tackle problems.My approach involves the prediction crimes and visualization of patterns.Using of past data could help us to correlate factors which might help to understand the possibility of happening of the particular crime.

About Data:

This dataset reflects reported incidents of crime (with the exception of murders where data exists for each victim) that occurred in the City of Chicago from 2001 to present, minus the most recent seven days. Data is extracted from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system.

- Dataset consists of 22 columns
- Consists of 6.3 millions of rows
- Size of the data set 1.4 GB

Data sources:

https://www.kaggle.com/chicago/chicago-crime

https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data

Research Design and Methodology:

- **Data download:** This may include the use of BigQuery API in python to download the dataset. This data set consists of data upto data.
- **Data cleaning/Wrangling:** understanding the data structure ,checking the missing values and invalid records

- **Exploratory Data analysis:**Finding the crime pattern and identification of correlation factors that could serve better training features
- **Feature Engineering:** This process to our data could helps us increasing the accuracy of the models.which creates the huge difference between a good model and bad model
- **Machine Learning models :** Building of prediction modes for the measure of accuracy and well as root mean square errors across the modes to find the best model among them
- Draw of Recommendations from the user story

Research Questions:

1. How has the number of various crimes changed over time in Chicago?
2. How have the number of arrests corresponding to the crimes changed over time in Chicago?
3. Are there any trends in the crimes being committed?
4. Which crimes are most frequently committed?
5. Which locations are these frequent crimes being committed to?
6. Are there certain high crime locations for certain crimes (etc Sexual offense)?
7. How has the number of certain crimes (etc homicide) changed over the years in Chicago?

Reference:

- https://blog.dataiku.com/predicting-london-crime-rates-using-machine-learning
- https://www.bjs.gov/content/pub/pdf/fca-eaibcd.pdf
- https://www.researchgate.net/publication/280722606_Crime_Analysis_and_Prediction_Using_Data_Mining
- https://towardsdatascience.com/indian-crime-data-analysis-85d3afdc0ceb

Conclusion:

Finally with this project i would like to tell a story from the day,identifying the most common crime and helping them to prevent.By applying the models we can accurately predict the upcoming event from the past data.

Deliverables:

Python code,presentation slide of my results.All the files will be uploaded to Github Folder here

Bibliography:

- Cva, K. (2017). https://www.medwinpublishers.com/JOBD/JOBD16000139.pdf. Journal of Orthopedics & Bone Disorders, 1(7). doi: 10.23880/jobd-16000139
- Sathyadevan, S., S, D. M., & S., S. G. (2014). Crime analysis and prediction using data mining. 2014 First International Conference on Networks & Soft Computing (ICNSC2014). doi: 10.1109/cnsc.2014.6906719
- Dandir, S. (2019, November 16). Indian Crime Data Analysis. Retrieved from https://towardsdatascience.com/indian-crime-data-analysis-85d3afdc0ceb
- Gakrelidz, N. (n.d.). Predicting London Crime Rates Using Machine Learning. Retrieved from https://blog.dataiku.com/predicting-london-crime-rates-using-machine-learning