CHICAGO CRIME DATASET ANALYSIS

INTRODUCTION –

The project report outlines how predictive crime analysis can help assist Chicago Police Department to prevent criminal activities in the city and ergo reduce crime rate. The Police Department of City of Chicago strives to improve their services to reduce the crime rate and our goal is to provide resourceful insights which in turn lead to reduction in crime rate. The data opted is from 2016 to 2020.

The main aims of the Project are:

- Common Crimes in Chicago-
  1. A graph depicting the most occurring offences in Chicago from 2016 to 2020.

- Arrests and the City of Chicago-
  1. What is the probability of an arrest?
  2. Distribution of Arrests across the month
  3. How does arrest vary according to the type of crime?

- Crime vs Time-
  1. How does crime vary across the year?
  2. Which hours are the most unsafe?
  3. Is your house safe from a burglary during the day?
  4. 24hour pattern of a crime.

- Crime vs Locations
  1. Which district is the safest to live in? Which district is "Sin-district"?
  2. Visualize most occurring crimes per district
  3. Concentration of Crime in the City

- Predicting
  1. Visualizing the crime rate graph for 2021 and rest of the 2020 year
  2. Visualizing the crime rate day-vise for 2021 and rest of the 2020 year

The features included are –

1. ID: Unique identifier for the record.
2. Case Number: The Chicago Police Department RD Number (Records Division Number), which is unique to the incident.
3. Date: Date when the incident occurred. this is sometimes a best estimate.
4. Block: The partially redacted address where the incident occurred, placing it on the same block as the actual address.
5. IUCR: The Illinois Uniform Crime Reporting code. This is directly linked to the Primary Type and Description.
6. Primary Type: The primary description of the IUCR code.
7. Description: The secondary description of the IUCR code, a subcategory of the primary description.
8. Location Description: Description of the location where the incident occurred.
9. Arrest: Indicates whether an arrest was made.

10. Domestic: Indicates whether the incident was domestic-related as defined by the Illinois Domestic Violence Act.
11. Beat: Indicates the beat where the incident occurred. A beat is the smallest police geographic area – each beat has a dedicated police beat car. Three to five beats make up a police sector, and three sectors make up a police district. The Chicago Police Department has 22 police districts.
12. District: Indicates the police district where the incident occurred.
13. Ward: The ward (City Council district) where the incident occurred.
14. Community Area: Indicates the community area where the incident occurred. Chicago has 77 community areas.
15. FBI Code: Indicates the crime classification as outlined in the FBI's National Incident-Based Reporting System (NIBRS).
16. X Coordinate: The x coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.
17. Y Coordinate: The y coordinate of the location where the incident occurred in State Plane Illinois East NAD 1983 projection. This location is shifted from the actual location for partial redaction but falls on the same block.
18. Year: Year the incident occurred.
19. Updated On: Date and time the record was last updated.
20. Latitude: The latitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
21. Longitude: The longitude of the location where the incident occurred. This location is shifted from the actual location for partial redaction but falls on the same block.
22. Location: The location where the incident occurred in a format that allows for creation of maps and other geographic operations on this data portal. This location is shifted from the actual location for partial redaction but falls on the same block.

WORKING –

1. STUDYING THE DATA:
   This includes:
   i. Checking the number of records and features that are 1206828 and 22 respectively.
   ii. Checking the top rows.
   iii. Checking the description of data.
   iv. Checking all data stored in different columns.

2. PREPARING THE DATA SET FOR ANALYSIS:
   i. Checking data set for missing values.
   ii. Dropping columns with missing values.
3. ANALYSIS:
   i. Visualizing all the objectives using matplotlib module and formatting the dataset according to our needs.

4. PREDICTION:
    i. Making the predictions from 2020 to 2021 using fbprophet module and formatting the dataset according to our needs.

REQUIREMENTS –

1. Chicago crime dataset
2. Jupyter Notebook (anaconda)
3. Modules used-
    numpy
    matplotlib
    pandas
    seaborn
    datetime
    fbprophet

CONCLUSION –

1. The prediction and analysis of the data could be useful for the Police Department, when deciding which areas to allocate more resources (this also depends on the crime type, which we have covered in our analysis).
2. If they want to increase the number of arrests for a particular crime type, then where (area) should they focus their efforts?
3. Which type of crime is more prone to happen at particular location (sidewalks, street, ATM, ..).
4. We plot the concentration of a crime in the city. Analysis of this will help the Police Department of City of Chicago to allocate more resources to red zones with high criminal activities.

FUTURE WORK –

1. Predicting the hotspots for crimes in future based on current data.
2. Extending the current predictions for more years.