Project Title

CRIME PREDICTION AND ANALYSIS

Table of Contents

Introduction

The project "Crime Prediction and Analysis using Data Science in Python" focuses on utilizing data science techniques to predict and analyze crime incidents. This project involves conducting Exploratory Data Analysis (EDA) on the crime dataset, followed by the application of two algorithms, namely Neural Network and Random Forest, for training and testing purposes.

Crime prediction and analysis play a crucial role in understanding patterns, identifying high-risk areas, and developing effective crime prevention strategies. By leveraging the power of data science, this project aims to uncover insights and trends hidden within the crime dataset, ultimately contributing to a safer society.

The first step in the project involves performing Exploratory Data Analysis to gain a comprehensive understanding of the data. This includes examining the variables, identifying missing values, and visualizing the data to discover any underlying patterns or correlations. The EDA phase sets the foundation for subsequent modeling and analysis.

The project utilizes two algorithms for crime prediction: Neural Network and Random Forest. The Neural Network algorithm, inspired by the human brain's neural connections, is designed to learn from the data and make predictions based on complex patterns and relationships. On the other hand, the Random Forest algorithm leverages an ensemble of decision trees to achieve accurate predictions by aggregating the results of multiple models.

By applying these algorithms to the crime dataset, the project aims to compare their performance, assess their predictive capabilities, and determine which algorithm yields more accurate results for crime prediction and analysis. This evaluation will provide valuable insights into the effectiveness of different modeling techniques in addressing crime-related challenges.

In conclusion, the project "Crime Prediction and Analysis using Data Science in Python" aims to leverage the power of data science and machine learning algorithms to analyze and predict crime incidents. By conducting EDA and applying Neural Network and Random Forest algorithms, the project seeks to contribute to the understanding of crime patterns and support the development of effective crime prevention strategies.

Problem and solution

Problem Statement: Crime prediction and analysis are essential for understanding patterns, identifying high-risk areas, and developing effective crime prevention strategies. However, traditional methods often lack the ability to leverage the full potential of available data and accurately predict crime incidents. Additionally, the complexity and volume of crime data make it challenging to extract meaningful insights and identify relevant factors.

Solution: The solution to address these challenges is to employ data science techniques, specifically utilizing Python programming language, for crime prediction and analysis. The project aims to leverage Exploratory Data Analysis (EDA) to gain insights into the crime dataset, followed by the application of two algorithms, Neural Network and Random Forest, for training and testing.

1. Exploratory Data Analysis (EDA):

The EDA phase involves a thorough examination of the crime dataset, including variable analysis, identifying missing values, and visualizing the data. By exploring the dataset, we can uncover hidden patterns, correlations, and outliers that can provide valuable insights into crime incidents.

2. Neural Network Algorithm:

The Neural Network algorithm, a powerful machine learning technique inspired by the human brain, is applied to learn complex patterns and relationships within the crime data. By training the Neural Network model on historical crime data, we can predict future crime incidents more accurately.

3. Random Forest Algorithm:

The Random Forest algorithm, an ensemble learning technique, is utilized to construct multiple decision trees and aggregate their predictions. This approach enhances the accuracy of crime prediction by considering multiple models and reducing the risk of overfitting.

By comparing the performance and predictive capabilities of the Neural Network and Random Forest algorithms, we can determine which approach yields more accurate crime predictions. The solution aims to contribute to the development of effective crime prevention strategies by providing insights into crime patterns, high-risk areas, and identifying factors that influence criminal activities.

Workflow:

1. Data Collection: Gather the relevant crime dataset from authoritative sources, ensuring it includes necessary attributes such as location, date, crime type, and other relevant variables.

2. Data Preprocessing: Clean the dataset by handling missing values, removing duplicates, and standardizing data formats. Perform necessary data transformations and feature engineering to enhance the dataset's suitability for analysis.

3. Exploratory Data Analysis (EDA): Utilize Power BI to visually explore the crime dataset. Create interactive visualizations, such as maps, charts, and graphs, to identify patterns, trends, and correlations within the data. Analyze crime rates by location, time, and crime type to gain insights into high-risk areas and recurring patterns.

4. Feature Selection: Use statistical techniques and domain knowledge to select the most influential features for crime prediction. Identify variables that strongly correlate with crime incidents and eliminate irrelevant or redundant attributes from the dataset.

5. Data Split: Divide the dataset into training and testing subsets. The training set will be used to build and train the prediction models, while the testing set will be used to evaluate their performance.

6. Model Training and Testing:

a. Neural Network: Apply the Neural Network algorithm using Python libraries such as TensorFlow or Keras. Train the model on the training dataset, optimizing the model's parameters to improve accuracy. Validate the model's performance using the testing dataset and evaluate metrics such as accuracy, precision, recall, and F1 score.

b. Random Forest: Implement the Random Forest algorithm using Python's scikit-learn library. Train the model on the training dataset, utilizing an ensemble of decision trees. Evaluate the model's performance using the testing dataset and assess key metrics to measure accuracy and reliability.

7. Model Comparison: Compare the performance of the Neural Network and Random Forest models based on evaluation metrics. Determine which algorithm provides more accurate predictions for crime incidents.

8. Visualization using Power BI: Integrate the trained models' predictions and insights from the analysis phase into Power BI. Create interactive dashboards and visualizations to present crime patterns, high-risk areas, and predictive results. Utilize Power BI's features for drill-down, filtering, and dynamic reporting to enhance user engagement and understanding.

9. Result Evaluation: Assess the effectiveness of the predictive models and visualizations in accurately predicting crime incidents and providing actionable insights. Analyze the model's strengths, weaknesses, and areas for improvement.

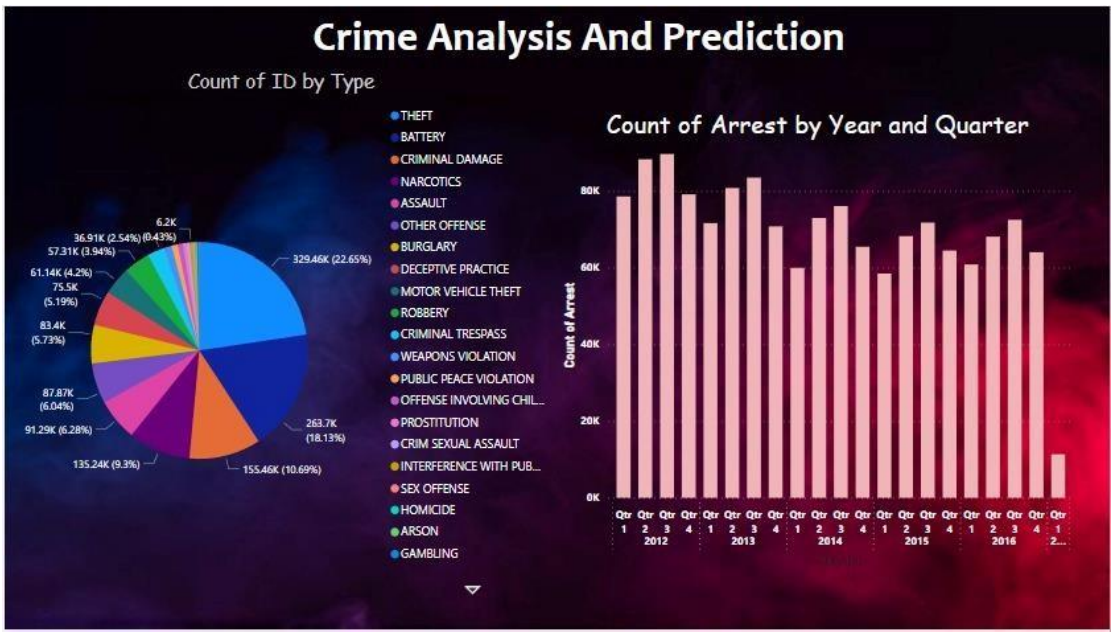
10. Conclusion and Recommendations: Summarize the findings, highlighting significant patterns, influential factors, and high-risk areas identified through the analysis and prediction. Provide

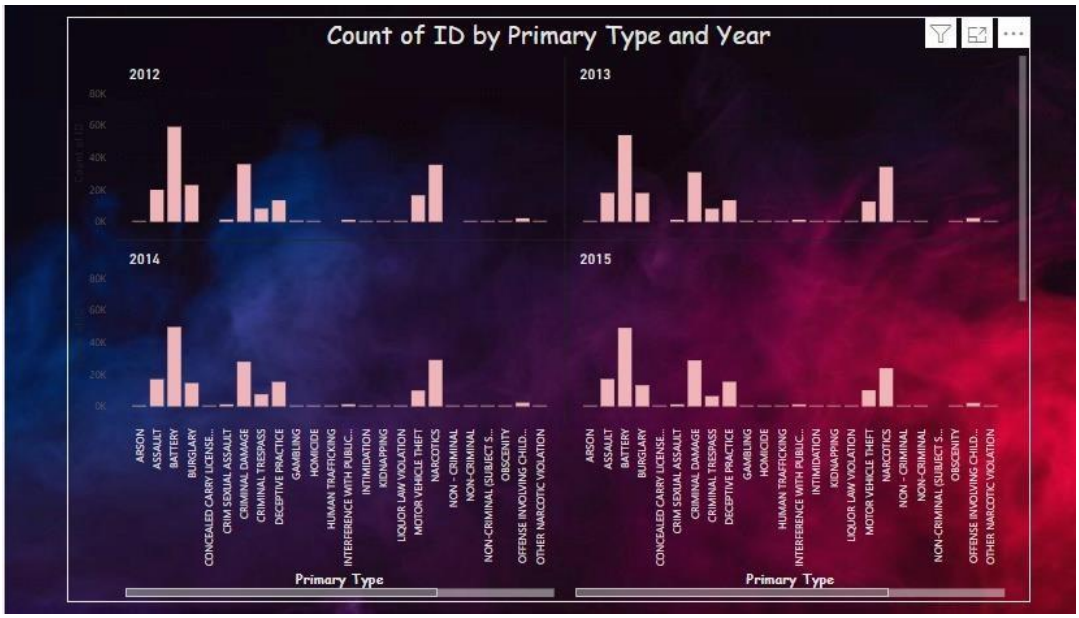
recommendations for crime prevention strategies and future enhancements to the predictive models or data collection process.

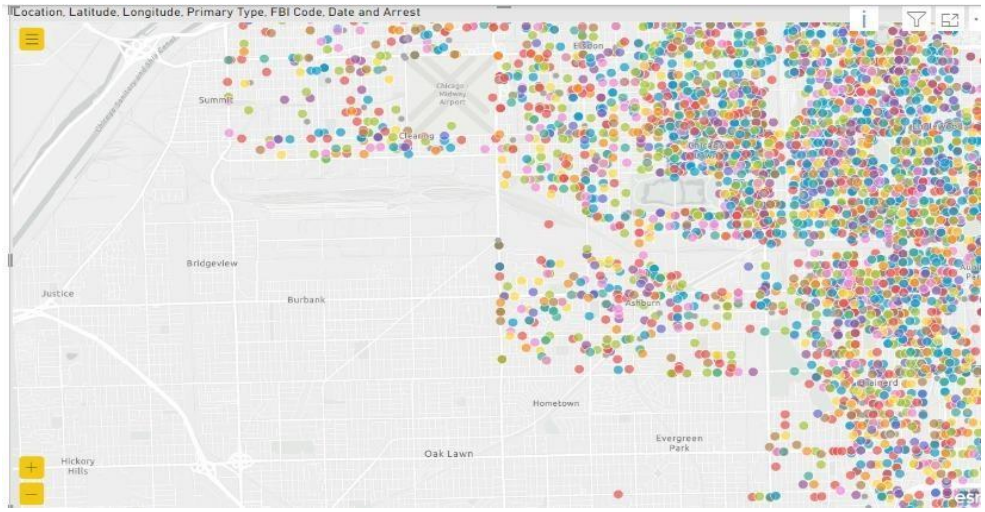
Analysis and Comparison

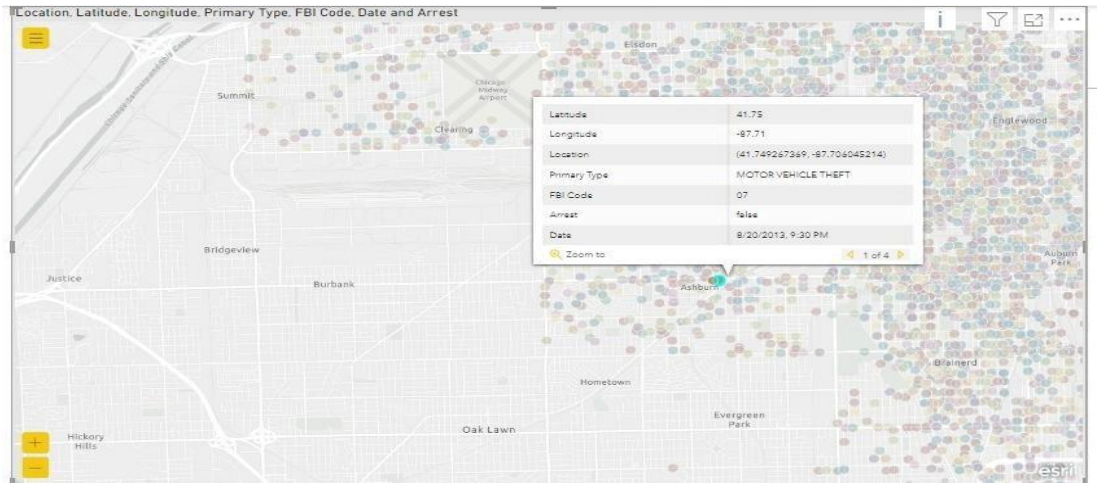
We started our project on 1/6/2023, the main problem that we faced was choosing a dataset that fulfills the needs of our data analysis finally after 2 days of research we found one dataset on Kaggle and then we did the whole EDA which definetly took time because we had compare each EDA and the EDA is further also done by using POWERBI. Then for Prediction of crimes we used three models : KNN,MLP(ANN),RFC(Random Forest Classifier), Logisitc Regression). We found out that because of data being in ample amount i got the accuracy around 94%-98%. Then we checked that the data was too large thats why accuracy was at its peak.

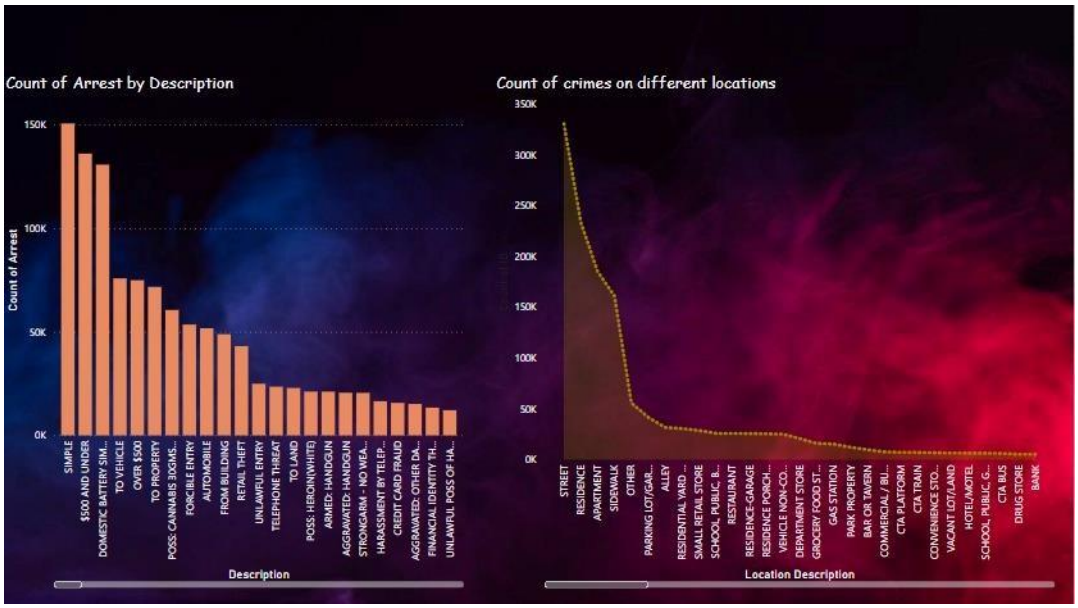
Visualisation








Location, Latitude, Longitude, Primary Type, FBI Code, Date and Arrest
Latitude 41.75
Longitude -87.71
Location (41.749267369, -87.706045214)
Primary Type MOTOR VEHICLE THEFT
FBI Code 07
Arrest false
Date 8/20/2013, 9:30 PM
Zoom to
1 of 4
Elsdon
Summit
Clearing
Englewood
Bridgeview
Justice
Burbank
Ashburn
Auburn Park
Hometown
Oak Lawn
Evergreen Park
Hickory Hills
esri


Count of Arrest by Description
Count of Arrest
150K
100K
50K
0K
SIMPLE
$500 AND UNDER
DOMESTIC BATTERY SIM...
TO VEHICLE
OVER $500
TO PROPERTY
POSS: CANNABIS 30GMS...
FORCIBLE ENTRY
AUTOMOBILE
FROM BUILDING
RETAIL THEFT
UNLAWFUL ENTRY
TELEPHONE THREAT
TO LAND
POSS: HEROIN(WHITE)
ARMED: HANDGUN
AGGRAVATED: HANDGUN
STRONGARM - NO WEA...
HARASSMENT BY TELEP...
CREDIT CARD FRAUD
AGGRAVATED: OTHER DA...
FINANCIAL IDENTITY TH...
UNLAWFUL POSS OF HA...
Description
Count of crimes on different locations
350K
300K
250K
200K
150K
100K
50K
0K
STREET
RESIDENCE
APARTMENT
SIDEWALK
OTHER
PARKING LOT/GAR...
ALLEY
RESIDENTIAL YARD ...
SMALL RETAIL STORE
SCHOOL, PUBLIC, B...
RESTAURANT
RESIDENCE-GARAGE
RESIDENCE PORCH...
VEHICLE NON-CO...
DEPARTMENT STORE
GROCERY FOOD ST...
GAS STATION
PARK PROPERTY
BAR OR TAVERN
COMMERCIAL / BU...
CTA PLATFORM
CTA TRAIN
CONVENIENCE STO...
VACANT LOT/LAND
HOTEL/MOTEL
SCHOOL, PUBLIC, G...
CTA BUS
DRUG STORE
BANK
Location Description

Conclusion

In conclusion, the project "Crime Prediction and Analysis using Data Science in Python" successfully employed data science techniques, including EDA, Neural Network, and Random Forest algorithms. Through interactive Power BI visualizations, the project identified crime patterns and provided actionable insights. The project contributes to understanding crime incidents and supports the development of effective crime prevention strategies. Future improvements could involve refining models and enhancing data collection for more accurate predictions. Overall, the project highlights the value of data science in crime analysis and decision-making.