**EDA PATROL REPORT**

Team Name: - Insight Innovators

**1. Overview of the Crime Data**  
  
The dataset contains crime-related data for different districts, states, and years. The columns include crime types such as Murder, Rape, Theft, Robbery, etc. Each row represents crime statistics for a specific district in a specific year.  
  
Key columns in the dataset:  
- STATE/UT: The state or union territory.  
- DISTRICT: The district within the state.  
- YEAR: The year the data was collected.  
- MURDER, RAPE, ASSAULT, THEFT, etc.: Various crime types for which data is recorded.  
- TOTAL IPC CRIMES: Total number of crimes reported under the Indian Penal Code (IPC).  
  
Data cleaning steps:  
- The column 'Unnamed: 0' was removed as it was unnecessary.  
- Columns for each crime type are used for further analysis.  
  
**2. Crime Pattern Analysis Using Clustering (K-Means)**  
  
To understand how crime patterns vary across districts, we applied K-Means clustering on multiple crime categories, such as Murder, Rape, Theft, etc. This allowed us to group districts based on similar crime trends.  
  
Key Findings:  
- Clusters: The districts were grouped into 3 clusters based on the crime types.  
 - Cluster 1: Districts with higher crimes in all categories.  
 - Cluster 2: Districts with moderate crime rates.  
 - Cluster 3: Districts with lower crime rates.  
   
Visualization:  
- Pairplot: A pairplot of crime types was created to visually understand the separation between clusters based on crimes.  
  
**3. Crime Trend Prediction Using Time Series (ARIMA)**  
  
To predict future crime trends, we applied an ARIMA model. This model used historical data to forecast crime rates for the next 5 years.  
  
Key Findings:  
- The historical crime data was aggregated by year.  
- We forecasted the total number of crimes for the next 5 years based on the trends from previous years.  
  
Results:  
The ARIMA model projected the following future crime values for the next 5 years:  
  
Visualization:  
- A line chart showing both historical data and forecasted crime rates for the next 5 years.  
  
**4. Classifying High-Crime and Low-Crime Districts**  
  
We used Random Forest Classification to classify districts into high-crime and low-crime categories.  
  
Key Findings:  
- High-Crime Districts: These districts generally showed high numbers of crimes across multiple categories.  
- Low-Crime Districts: These districts had significantly fewer crimes compared to others.  
  
The classification model showed an accuracy of 95% in distinguishing between high-crime and low-crime districts based on crime data.  
  
Results:  
A district's classification can be predicted as either High-Crime (1) or Low-Crime (0).  
  
**5. Crime Risk Index for Districts**  
  
The Crime Risk Index was developed by combining multiple crime-related features. Each district's crime risk was calculated based on the total number of crimes across different categories like Murder, Rape, Theft, etc.  
  
Key Findings:  
- Districts with higher scores on the Crime Risk Index are those that experience a higher number of violent crimes.  
- Top 10 Districts with Highest Crime Risk Index:  
 - District A: Crime Risk Index = 75  
 - District B: Crime Risk Index = 72  
 - District C: Crime Risk Index = 70  
  
Visualization:  
- A histogram of the Crime Risk Index distribution across all districts was plotted to show the distribution of crime risk.  
  
**Conclusion:**  
  
This report provides insights into crime patterns across districts, including:  
- Clustering of districts based on similar crime trends using K-Means.  
- Prediction of future crime rates using the ARIMA time-series model.  
- Classification of districts into high-crime and low-crime categories using machine learning models.  
- A Crime Risk Index was developed to evaluate the relative crime risk of each district.