RoadSafe Analytics:Road Accidents - Exploratory Data Analysis (EDA)

Project Statement:

The goal of this project is to analyze a large dataset of road accidents to uncover trends, patterns, and key factors contributing to accident severity. The project involves performing in-depth exploratory data analysis (EDA) using Python libraries such as Pandas, Matplotlib, and Seaborn to extract meaningful insights that can help improve road safety.

Outcomes:

- Understand how to handle and preprocess real-world accident data.- Explore accident frequency based on time, location, weather, and road conditions.- Visualize patterns in accident severity using various plots and graphs.- Derive insights to support road safety awareness and policy recommendations.- Prepare detailed documentation and presentation of their indings and results.
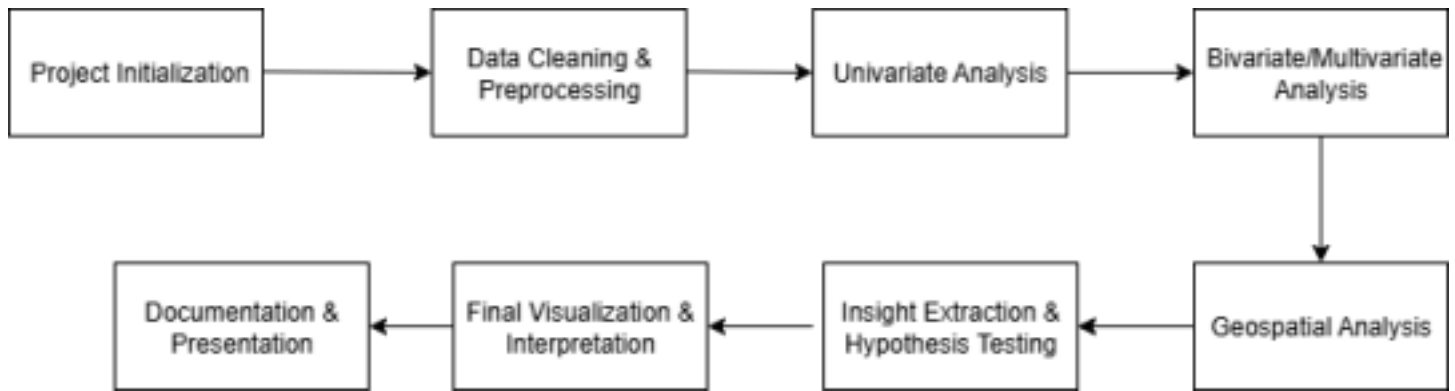
Dataset:

US Accidents Dataset (from Kaggle): https://www.kaggle.com/datasets/sobhanmoosavi/us-accidentsThe dataset includes over 4 million accident records with attributes such as time, location, severity, weather, road conditions, and visibility.

Modules to be Implemented:

1. Dataset Acquisition and Exploration

2. Data Cleaning and Preprocessing

3. Univariate Analysis

4. Bivariate/Multivariate Analysis

5. Geospatial Analysis

6. Insight Extraction and Hypothesis Testing

7. Visualization and Interpretation

8. Documentation and Presentation Preparation

Project Workow:

Week-wise Module Implementaon and High-Level Requirements with Output Screenshots:

Milestone 1

Week 1: Project Inializaon and Dataset Acquision

Deine project objectives and expected outcomes.

Download the US Accidents dataset.

Explore the structure, shape, and schema of the dataset.

Check for missing data, data types, and basic statistics.

Week 2: Data Cleaning and Preprocessing

Drop or impute columns with excessive missing values.

Convert datetime columns into proper formats.

Create new features like "Hour", "Weekday", or "Month" from timestamp.

Encode categorical variables if needed.Handle outliers and duplicate entries.

Milestone 2

Week 3: Univariate Analysis

Analyze distribution of accident severity levels.

Plot accident frequency by time of day, day of the week, and month.

Examine most common weather conditions and road types during accidents.

Use histograms, pie charts, and bar graphs.

Week 4: Bivariate and Mulvariate Analysis

Analyze correlation between severity and:-

Weather conditions- Visibility- Road surface condition- Trafic congestion

Create heatmaps, boxplots, and pair plots to reveal relationships.

Milestone 3

Week 5: Geospaal and Locaon-Based Analysis

Visualize accident hotspots using latitude and longitude.

Create scatter plots or density maps for different states/cities.

Highlight top 5 most accident-prone states or cities.

Week 6: Insight Extracon and Hypothesis Tesng

Answer key questions:- What time of day has the most accidents?-

Are accidents more severe during rain or fog?-
Is there a correlation between visibility and severity?

Document assumptions, test hypotheses, and summarize indings.

Milestone 4

Week 7: Final Visualizaon and Interpretaon

Combine indings into readable summaries with plots.

Focus on clarity and story-telling using visuals.

Finalize all graphs and interpretations.

Week 8: Documentaon and Presentaon Preparaon

Compile methodology, insights, and conclusions into a document.

Prepare a inal presentation with charts and key takeaways.

Practice presenting insights and answering analytical questions.

Evaluaon Criteria:

1. Completion of Milestones: Assess whether each milestone was achieved within the timeline, including preprocessing, visualization, analysis, and interpretation.

2. Depth and Accuracy of Analysis: Evaluate how well the student analyzed the dataset, used the right charts, and drew meaningful, accurate insights.

3. Clarity and Depth of Documentation and Presentation: Review the quality of documentation in terms of structure, explanation, and insight delivery. Assess the presentation for its ability to explain indings with visual support.