



DEPARTMENT OF CONSTRUCTION ENGINEERING AND SRIJAN '25
PRESENTS

DATA DRIFT

Date of Event : 20th April, 2025

Problem Statement

Team member :

01 (min.) to 02 (max.) Students from various educational institutions and departments of any year of study at the Undergraduate level can work together to form a team.

Introduction :

Welcome to the Data Analysis Event on Road Accidents

We are excited to host this data analysis event, where we will delve into the factors contributing to road accidents. By focusing on critical elements such as road conditions, road type, surface type, lane type, weather conditions, and casualties, our aim is to uncover insights that will help improve safety measures and reduce road risks. This event brings together experts and enthusiasts to explore how these factors influence accidents and to identify trends that can inform future safety strategies.

The dataset provided covers road accident casualties in Karnataka from 2016 to 2023, and includes key variables like road type, surface type, weather conditions, and other significant factors. As a data analyst or graduate engineer in the traffic department of your city's municipal corporation, you are tasked with analyzing this dataset to uncover actionable insights that can drive improvements in road safety.

Through collaborative efforts and the use of advanced tools, we hope to identify effective measures to prevent accidents and enhance public safety. Join us in this effort to analyze the data, explore trends, and work towards creating a safer environment for all road users.

Task :-

Data cleaning : It is a critical step in transforming raw datasets into organized and relevant information. In the provided dataset, several unrefined entities need to be addressed to ensure the quality and accuracy of analysis:

1. **Duplicate Entries:** Instances of identical records appearing multiple times, which can lead to redundancy and distort the integrity of the analysis.
2. **Blank Entries:** Records with missing essential information, hindering a thorough and reliable data analysis process.
3. **Spelling Errors:** Incorrect spellings within the dataset can introduce confusion, leading to misinterpretation and potential errors in subsequent analyses.
4. **Null Entries:** Records with null values in critical fields, which can limit the scope and accuracy of the analysis if not properly handled.

Data Processing :

1. **Grouped** similar row categories to identify and analyze broader data trends.
2. Introduce a new feature, "**Month**", derived from the date column, to gain insights into the monthly distribution and trends of accident-related casualties.
3. Create a feature, "**Day of Week**", extracted from the date column, to analyze the weekly distribution and trends of accident occurrences.

Primary KPIs :-

1. **Total Number of Accidents by District (Top 5)**

- Calculate and visualize the total number of accidents reported in each district. Highlight the **top 5** districts with the highest accident counts.

2. Accident Severity Distribution

- Analyze and present the distribution and percentage share of casualties by severity level (Fatal, Serious, Slight).

3. Accident Patterns by Time of Day and Day of Week

- Identify accident trends across different times of the day and days of the week. Additionally, compare accident frequencies between weekdays and weekends.

4. Road Type Contribution to Accidents

- Determine which types of roads are associated with the highest number of accidents.

5. Casualty Distribution by Weather Conditions

- Analyze the breakdown of total casualties under different weather conditions.

6. Accidents by Lane Type

- Assess the number of accidents based on lane configurations.
(Note: Participants are encouraged to address the missing data creatively.)

7. Accidents by Road Character and Severity

- Visualize how different road geometries (e.g., straight, curve, slope) impact accident severity levels.

8. Accidents by Surface Type

- Analyze which types of road surfaces are more prone to accidents.

9. Top Causes of Accidents

- Identify the most common causes of accidents and evaluate their impact on the overall accident count and severity.



Secondary KPI (Deeper Insights) :-

1. Monthly Casualty Trend Analysis (2021–2023)

- Analyze trends in total casualties across months from 2021 to 2023 to uncover seasonal patterns and consistently high-risk periods.

2. Casualty Analysis across Surface Condition by Surface Type and Weather Condition

- Analyze the impact of surface conditions on casualties by combining surface type and weather condition.

3. Vehicle Involvement by Time of Day and Road Type

- Show how the number of vehicles involved in accidents varies across different times of the day and types of roads.

4. Casualty Analysis at Junction Controls by Accident severity

- Analyze how different types of junction controls (e.g., controlled, uncontrolled) correlate with accident severity outcomes.

Create a Dashboard:

1. Develop an intuitive dashboard interface that efficiently consolidates and displays key metrics and data points, providing clear insights at a glance.
2. Utilize visualization tools to present complex data in a visually engaging and easily interpretable format, enhancing user experience and data accessibility.

Create a Presentation (ppt) with Insights:

1. Develop a well-structured presentation that effectively synthesizes data-driven insights and analysis, ensuring clarity and coherence.
2. Emphasize key trends, patterns, and correlations identified through in-depth analysis of the data, providing actionable insights for decision-making.

Finally, you are encouraged to explore and analyze the data further to uncover additional insights beyond the scope of the tasks outlined above. The primary objective of these new findings should be to support the city's traffic planning board in revising existing policies, with the goal of preventing future accidents and enhancing road safety.

Time :

A total time of **2 hours and 30 minutes** will be provided for analysing the data, creating the dashboard and preparing a PPT.

Judging criteria :

1. Accurate data cleaning is essential for reliable analysis.
2. Explore and visualize the KPIs for actionable insights.
3. Find additional insights beyond the given metrics for extra credit.
4. Respond thoughtfully to judges' questions for extra points.
5. Completing the task early may earn bonus points.

