DEPARTMENT OF CONSTRUCTION ENGINEERING AND SRIJAN '24
PRESENTS

DATA DRIFT

Date of Event : 23rd March, 2024

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 our data analysis event focusing on accidents, where we'll explore factors like speed, weather, and casualties. Here, experts and enthusiasts gather to study how these factors affect accidents. By analyzing data closely, we hope to find connections and trends that can help make safety measures better and reduce risks. Through teamwork and using advanced tools, we'll work together to find practical ways to prevent accidents and keep communities safer. Join us as we dig into the data to find ways to make our world safer for everyone.

The dataset provided is about road accident casualties in UK in the year 2021 and 2022. The set consists of accidents reported to respect police authorities which contains various important factors such as speed,weather condition for every accident reported. Imagine you are a data analyst/graduate engineer in the traffic department of your city's municipal corporation. You are given this dataset by your supervisor and your task is as follows:

Task :

Data Cleaning : Data cleansing is a crucial step in refining raw datasets into organised and pertinent entries. Within the provided dataset, the following unrefined entities needs to be identified:

1. Duplicate entries: Instances where identical records appear more than once, which can introduce redundancy and inaccuracies in analysis.
2. Blank entries: Records lacking essential information, which can impede comprehensive data analysis and interpretation.
3. Spelling mistakes in some data: Incorrect spellings in the data can cause confusion and mistakes in understanding, which can then lead to errors in later analyses.

Data Processing :

1.Grouped similar row categories to get a more general data trend
2.Add a new feature "month" by extracting month from the date column to get a better idea of monthly trend of accident casualties
3.Add a new feature "year" by extracting year from the date column to get a better idea of yearly trend of accident casualties

Primary KPI :

1. Analysis of total casualties and their percentage distribution based on accident severity.
2. Examination of casualties categorized by vehicle type.
3. Identification of monthly casualty trends throughout the years 2021-2022.
4. Assessment of casualties relative to road types.
5. Investigation into casualties in correlation with the number of vehicles involved in accidents.
6. Exploration of the relationship between casualties and geographical factors such as urban and rural areas.
7. Analysis of casualties concerning temporal factors such as day and night occurrences.
8. Examination of casualties in relation to road surface conditions.

9. Evaluation of casualties during weekdays versus weekends, segmented by urban and rural settings.

Secondary KPI :

1. Analysis of casualties segmented by time intervals (3 hours) and day of the week.
2. Examination of casualties at junction-controlled areas in rural and urban settings, specifically for incidents involving high-speed vehicles.
3. Comparative study of casualties between rural and urban areas, focusing on fatal accidents during high-speed incidents and considering the lighting conditions of the road at night.
4. Investigation into casualties based on weather conditions and speed variables, aiming to discern correlations and patterns.

(**NOTE** : Merge similar types of data for better analysis (e.g., In Vehicle type Motorcycles over 125cc and up to 500cc, Motorcycles over 500cc, and Motorcycles 50cc and under can be categorized under a single category: motorcycles).

Create a Dashboard:

1. Develop an intuitive dashboard interface to consolidate and display key metrics and data points effectively.
2. Utilize visualization tools to present complex information in a visually appealing and accessible format.

Create a Presentation (ppt) with Insights:

1. Construct a structured presentation that synthesizes data-driven insights and analysis.
2. Highlight significant trends, patterns, and correlations discovered through thorough examination of the data.

Finally, you can also extract/analyze the data and find some new insights which are not given in the above tasks. The fundamental intention behind the new findings should be such that it helps the city's traffic planning board to modify the existing policies, to prevent more accidents in the future.

Time :

A total time of 2 hours and 15 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. Start with at least four primary Key Performance Indicators (KPIs).
3. Explore and measure as many KPIs as possible.
4. Find additional insights beyond the given metrics for extra credit.
5. Respond thoughtfully to judges' questions for extra points.
6. Completing the task early may earn bonus points.
7.Use Excel exclusively for creating your dashboard and calculations, ensuring consistency and accuracy.