Batch: B.Tech (CSE) – V Semester

SQL Project: Road Accident Data Analysis using MySQL

Problem Statement:

Imagine you are a Junior Data Analyst at the National Road Safety Bureau (NRSB). The bureau has given your team a cleaned dataset — accident_data_india.xlsx — containing detailed records of road accidents across India. Your analyst team must use SQL (MySQL) to load the data, perform exploratory analysis, identify high-risk patterns, and prepare recommendations for policy makers.

Dataset Columns

Find the Dataset link here: https://tinyurl.com/sutd3wtc

State Name	City Name
Year	Month
Day of Week	Time of Day
Accident Severity	Number of Vehicles Involved
Vehicle Type Involved	Number of Casualties
Number of Fatalities	Weather Conditions
Road Type	Road Condition
Lighting Conditions	Traffic Control Presence
Speed Limit (km/h)	Driver Age
Driver Gender	Driver License Status
Alcohol Involvement	Accident Location Details

Objective

Using the dataset above, write SQL queries to find patterns of risk, time and location trends, vehicle and driver characteristics related to severity and fatalities. Produce a short briefing document (and presentation) for the NRSB summarizing key findings and actionable recommendations.

Project Tasks

Setup & Data Load

1. Create the database `nrsb_accidents` and table `accidents` with appropriate data types.
2. Load the data from the Excel (or CSV) file and verify with SELECT COUNT(*) and SELECT * LIMIT 10.

Data Quality & Cleaning

3. Identify columns with NULL or missing values.
4. Standardize the Month column and normalize Vehicle Type Involved values.

Exploratory Analysis

5. List total accidents per state (top 10).
6. Show accidents per city for a chosen state.
7. Show accident counts per year, weekday, and time of day.

Severity & Outcomes

8. Compute total fatalities and casualties nationwide.
9. Average fatalities per severity type.
10. Top 10 fatal accidents with details.
11. Calculate fatality rate per accident.

Environmental & Road Factors

12. Identify which weather conditions are linked with the highest fatalities.
13. Compare road type, condition, and lighting for their impact.

Driver & Vehicle Factors

14. Accidents split by gender and age groups.
15. Compare alcohol involvement, license status, and fatalities.

Combined Risk Scenarios

16. Find top combinations of weather and road type by total fatalities.
17. Compare accidents with/without traffic control.

Reporting & Presentation

18. Create a state summary table showing accidents, fatalities, and fatality rate.
19. Write policy recommendations and present findings in a short report or slides.

Deliverables

- SQL script file with all queries and cleaning steps.
- Short report (2 pages) summarizing top 7 insights with SQL results.
- 3–5 slide presentation highlighting insights and recommendations.

Your focus should be on clear, reproducible SQL steps.