Transportation Database SQL Analysis Project

1. Project Overview

This project demonstrates how SQL can be used to analyze transportation and traffic-related data. The dataset represents synthetic transportation records that include vehicle counts, traffic density, accident information, alerts, and road-level details.

The main goal of the project is to:

- Explore transportation records
- Analyze traffic volume and density
- Identify accident patterns
- Understand alert generation
- Find roads with high traffic and safety risks

The project uses a MySQL-style database named **transportation** and a table called **transportation123**.

2. Database Creation

CREATE DATABASE transportation;

This command creates a database named **transportation**, which stores all traffic and transportation-related data for analysis.

3. Dataset Description (Column Explanation)

Below is a description of each column typically found in the transportation123 dataset:

Column Name	Description
Record_ID	Unique identifier for each transportation record (primary key).
Date	The date when the traffic data was recorded.
Time	Time of day when the observation was made.
Road_ID	Unique identifier for each road or route.
Location	Name or description of the road location.
Vehicle_Type	Type of vehicle observed (e.g., Car, Bus, Truck, Bike).

Column Name	Description
Vehicle_Count	Total number of vehicles recorded during that observation.
Traffic_Density	Traffic condition level (Low, Medium, High).
Average_Speed	Average speed of vehicles on the road (km/h or mph).
Accident_Occurred	Indicates whether an accident occurred (Yes / No).
Accident_Severity	Severity level of accident (Low, Medium, High) – applicable only if accident occurred.
Weather_Condition	Weather at the time of observation (Clear, Rainy, Foggy, etc.).
Alert_Generated	Indicates whether a traffic alert was generated (Yes / No).

4. SQL Queries Explanation

Query 1: Display First 10 Transportation Records

```
SELECT * FROM transportation123 LIMIT 10;
```

Purpose:

- Displays a preview of the dataset
- Helps understand the structure and sample values

Query 2: Total Number of Trips Recorded

```
SELECT COUNT(*) AS total_trips FROM transportation123;
```

Purpose:

- Counts the total number of records (trips)
- Gives an idea of dataset size

Query 3: Average Number of Vehicles Recorded

```
SELECT AVG(Vehicle_Count) AS avg_vehicle_count
FROM transportation123;
```

Purpose:

- Calculates average traffic volume per record
- Useful for traffic planning and capacity analysis

Query 4: Records by Vehicle Type

```
SELECT Vehicle_Type, COUNT(*) AS total_records
FROM transportation123
GROUP BY Vehicle_Type;
```

Purpose:

- Shows distribution of different vehicle types
- Helps analyze dominant vehicle categories

Query 5: Records by Traffic Density Level

```
SELECT Traffic_Density, COUNT(*) AS count
FROM transportation123
GROUP BY Traffic_Density;
```

Purpose:

- Identifies how often roads experience low, medium, or high traffic
- Useful for congestion analysis

Query 6: Accident Occurrence Count

```
SELECT Accident_Occurred, COUNT(*) AS total
FROM transportation123
GROUP BY Accident_Occurred;
```

Purpose:

- Compares number of records with and without accidents
- Measures accident frequency

Query 7: Accident Count by Severity

```
SELECT Accident_Severity, COUNT(*) AS severity_count
FROM transportation123
WHERE Accident_Occurred = 'Yes'
GROUP BY Accident_Severity;
```

Purpose:

- Analyzes seriousness of accidents
- Helps identify high-risk accident levels

Query 8: Alert Generation Count

```
SELECT Alert_Generated, COUNT(*) AS alert_count
FROM transportation123
GROUP BY Alert_Generated;
```

Purpose:

- Shows how often traffic alerts are triggered
- Useful for monitoring system responsiveness

Query 9: Roads with Highest Traffic Volume

```
SELECT Road_ID, SUM(Vehicle_Count) AS total_vehicles
FROM transportation123
GROUP BY Road_ID
ORDER BY total_vehicles DESC
LIMIT 5;
```

Purpose:

- Identifies top 5 busiest roads
- Helps prioritize infrastructure improvements

Query 10: Roads with High Traffic and Accidents

```
SELECT DISTINCT a.Road_ID
```

```
FROM transportation123 a

JOIN transportation123 b

ON a.Road_ID = b.Road_ID

WHERE a.Traffic_Density = 'High'

AND b.Accident_Occurred = 'Yes';
```

Purpose:

- Finds roads with both congestion and accident history
- Useful for safety planning and traffic control

5. Conclusion

This SQL project demonstrates how transportation data can be analyzed using structured queries to gain insights into traffic patterns, accident risks, and alert generation. Such analysis is valuable for:

- Traffic management systems
- Smart city planning
- Road safety improvement
- Transportation policy decisions

6. Tools & Technologies

- **Database:** MySQL
- **Dataset:** Synthetic Transportation Dataset
- **Skills Used:** SQL Queries, Aggregation, Grouping, Joins