Accident Data Analysis Documentation

Project design

1.1 Objective

The primary objective of this project is to analyze traffic accident data to uncover patterns, identify key contributing factors, and understand the characteristics of drivers, vehicles, locations, and environmental conditions associated with accidents.
The dashboard aims to support data-driven decisions by providing insights into:

- Overall accident frequency
- Driver demographics (age, gender)
- Vehicle involvement
- Time and weather patterns
- Accident severity and hotspot locations

1.2 Project Scope

Geographical Scope

- All accident locations included in the dataset.
- Road types and areas shown through map visualizations and location-based charts.

Driver Scope

- Driver age distribution
- Gender breakdown
- Driver category (if available)

Vehicle Scope

- Vehicle types involved in accidents
- Accident distribution across vehicle categories

Environmental Scope

- Road types and other environmental factors affecting accidents

1.3 Stakeholders

The analysis is designed for use by:

- **Traffic & Road Safety Authorities**
  To identify accident hotspots and reduce accident rates.
- **Transportation & Logistics Companies**
  To improve driver training and reduce operational risks.
- **Government Agencies & City Planners**
  For infrastructure planning and road safety improvements.
- **Insurance Companies**
  To better understand driver risk profiles.
- **Data Analysts & Researchers**
  For further modeling and exploration.

1.4 Tools and Technologies

The following tools were used to complete the project:

- **Power BI** – Data modeling, DAX calculations, and dashboard visualization
- **Microsoft Excel** – Data cleaning and preprocessing
- **SQL Server** (if used) – Querying, filtering, and transforming data
- **Power Query** – ETL (Extract, Transform, Load) operations inside Power BI

1.5 Business Questions

The project is designed to answer several key analytical questions:

1. **What is the total number of accidents per year and how does it change over time?**
   → Trend lines and KPIs show annual and monthly accident patterns.
2. **What are the most common causes of accidents?**
   → Charts highlight the top accident causes and their frequency.
3. **Do accidents follow a specific pattern according to day, month, or time of day?**
   → Time-of-day and seasonal analyses reveal when accidents most frequently occur.
4. **What is the age distribution of drivers involved in accidents?**
   → KPIs and charts show average driver age and accident counts by age group.
5. **Which vehicle types are most often involved in accidents?**
   → Bar charts display vehicle involvement across different categories.
6. **Where do accidents occur most frequently (hotspot analysis)?**
   → The map view identifies locations with the highest accident concentration

2. Data Source and Description

This project uses a structured dataset that contains detailed information about road accidents. The data is collected from multiple tables, and each table contributes to understanding different dimensions of the accident records.

1. Drivers Table

- Contains personal information about drivers.
- Key fields include:
    - **Driver ID**
    - **Age**
    - **Gender**
    - **Experience Level** (if available)

2. Vehicles Table

- Provides details about the vehicles involved in accidents.
- Key fields include:
    - **Vehicle ID**
    - **Vehicle Type**
    - **Car Model**
    - **Plate Number**

3. Accidents Table

- Represents the core dataset containing accident events.
- Key fields include:
    - **Accident ID**
    - **Accident Date & Time**
    - **Location**
    - **Severity Level**
    - **Weather Conditions**
    - **Road Conditions**
    - **Driver ID**
    - **Vehicle ID**

Purpose of each table

Table	Purpose
Drivers	Analyze age, gender, and experience impact on accident frequency.
Vehicles	Compare accident rates by vehicle type and model.
Accidents	Core analysis: when, where, and why accidents happened.
Lookup Tables	Improve reporting accuracy and enable filtering by categories.

3. Data Model (Data Architecture Overview)

The data model represents how all tables in the accident dataset are connected. A well-structured model ensures accurate analysis, correct relationships, and reliable dashboard insights.

3.1. Tables Included in the Data Model

1. Accidents Table (Fact Table)

This is the central table in the data model.

- **Primary Key:** `Accident_ID`
- **Contains details about each accident**, such as:
    - Date & Time
    - Location
    - Weather condition
    - Road condition
    - Severity level
    - Linked Driver ID
    - Linked Vehicle ID
    - Linked Road ID

---

2. Drivers Table (Dimension Table)

Stores information related to drivers involved in accidents.

- **Primary Key:** `Driver_ID`
- **Fields include:**
  - Age
  - Gender
  - Experience

---

3. Vehicles Table (Dimension Table)

Contains detailed information about the vehicles involved.

- **Primary Key:** `Vehicle_ID`
- **Fields include:**
  - Vehicle type
  - Model
  - Car category

---

4. Roads Table (Dimension Table)

Includes details about the road where the accident occurred.

- **Primary Key:** `Road_ID`
- **Fields include:**
  - Road type
  - Speed limit
  - Surface condition

3.2. Relationships Between Tables

The relationships follow a **Star Schema**, where the **Accidents table is the central fact table**, and all other tables are connected as dimensions.

Relationship Summary

From Table	To Table	Relationship Type	Nature	Key Used
Drivers → Accidents	One-to-Many	One driver can be involved in many accidents	PK → FK	`Driver_ID`

From Table	To Table	Relationship Type	Nature	Key Used
Vehicles → Accidents	One-to-Many	One vehicle can appear in several accident records	PK → FK	`Vehicle_ID`
Roads → Accidents	One-to-Many	One road can have multiple accidents	PK → FK	`Road_ID`

Explanation of Relationship Types

1. One-to-Many (1 → ∞)

This is the main type used in the accident model.
Examples:

- **One driver can be associated with many accident records.**
- **One vehicle can appear in multiple accidents.**
- **One road can have multiple accidents.**

This ensures clean filtering and avoids duplicated results in visualizations.

3.3. Keys Used in the Model

Primary Keys (PK)

Each table has a unique identifier:

- `Accident_ID` → Accidents table
- `Driver_ID` → Drivers table
- `Vehicle_ID` → Vehicles table
- `Road_ID` → Roads table

Foreign Keys (FK)

Stored inside the Accidents table to connect the dimensions:

- `Driver_ID` → connects to Drivers
- `Vehicle_ID` → connects to Vehicles
- `Road_ID` → connects to Roads

3.4. Example Relationship Description

Example:
A driver with `Driver_ID = 102` appears in the Drivers table only once.
But in the Accidents table, this same driver ID may appear 5 times.
This means:

- The driver **is involved in 5 different accidents**.
- The relationship is **One Driver → Many Accidents**.

4. Visualizations (Dashboard Visual Components)

This section explains every visual used in the final accident analysis dashboard, why it was selected, and what insight it delivers.

4.1. KPI Cards

1. Total Accidents

- **Type:** KPI Card
- **Purpose:** Shows the overall number of accidents in the dataset.
- **Reason:** Provides an instant understanding of the scale of the problem.

2. Average Driver Age

- **Type:** KPI Card
- **Purpose:** Displays the average age of all drivers involved in the reported accidents.
- **Reason:** Helps identify whether certain age groups are more frequently involved in accidents.

3. Fatal Accidents %

- **Type:** KPI Card
- **Purpose:** Calculates the percentage of accidents resulting in fatal outcomes.
- **Reason:** Measures the severity of incidents and highlights critical safety risks.

4. Male vs Female Accidents (Gender Ratio)

- **Type:** KPI Card
- **Purpose:** Summarizes total accidents for each gender.
- **Reason:** Quickly shows whether one gender is significantly more involved in accidents.

4.2. Donut / Pie Charts

1. Gender Distribution of Drivers

- **Purpose:** Shows the percentage split between male and female drivers involved in accidents.
- **Reason:** Visualizes disproportionate involvement and supports demographic analysis.

2. Accident Severity Breakdown

- **Purpose:** Illustrates how accidents are divided across severity levels (Minor, Major, Fatal).
- **Reason:** Helps prioritize focus areas for safety improvements.

4.3. Bar Charts

1. Accidents by Vehicle Type

- **Purpose:** Displays the number of accidents categorized by vehicle type.
- **Reason:** Helps identify which types of vehicles (Cars, Trucks, Motorcycles…) are most frequently involved.

2. Accidents by Weather Condition

- **Purpose:** Shows how weather conditions (Clear, Rainy, Foggy, etc.) affect accident frequency.
- **Reason:** Weather is a key risk factor — this chart highlights environmental impact on accident rates.

3. Accidents by Road Condition

- **Purpose:** Compares road surface conditions (Dry, Wet, Icy…).
- **Reason:** Helps determine whether poor road infrastructure contributes to higher accident rates.

4.4. Tables

1. Detailed Accident Table

- **Purpose:** A searchable, sortable table listing full accident records.
- **Reason:** Gives stakeholders access to raw, detailed information for deeper investigation.

5. Report

- Enhance Data Monitoring

Implement continuous monitoring of accident records to ensure timely identification of trends related to drivers, injuries, and accident causes.

2. Improve Driver Awareness Programs

Develop targeted awareness programs focusing on the most affected driver groups to reduce accident frequency

3. Focus on High-Risk Accident Types

Allocate attention and resources toward the accident types that appear most frequently in the dataset to reduce repeated patterns.

4. Strengthen Vehicle Safety Measures

Encourage monitoring and enforcement of vehicle safety standards to reduce injuries in multi-vehicle accidents.

5. Utilize Weather and Time Patterns

Use insights about weather conditions and accident timings to optimize safety messaging and preventive actions during high-risk period.

- KPIs

5.Key Performance Indicators (KPIs)

1. Total Number of Accidents

Shows the overall accident count in the dataset

2. Number of Male drivers

3. % Male Drivers

Proportion of accidents involving male drivers.

 4. % Female Drivers

Proportion of accidents involving female drivers.

5.Number of Severe Accidents

Number of accidents marked as fatal.

7.% Severe accidents

8.AVG vehicle age

9.AVG driver age