Vehicle Accident Statistics Dashboard

Prepared by
Dhvanilkumar Patel

Contents

1.Introduction

1.1 Purpose

This document provides a comprehensive design for the Vehicle Accident Statistics Dashboard, aimed at analyzing road traffic accidents and identifying key factors contributing to their occurrence and severity. The dashboard provides actionable insights to enhance road safety, guide policymaking, and improve decision-making.

1.2 Scope

The project covers accident analysis based on **demographic factors, environmental conditions, vehicle types, and road characteristics**. It does not include real-time traffic monitoring or predictive accident modeling.

2. System Overview

2.1 System Architecture

The system architecture supports comprehensive analysis and visualization of vehicle accident data sourced from traffic records and weather databases, stored in a secured local environment . The ETL process uses Python for data cleaning and transformation, with Power BI Power Query for loading. Data modeling in Power BI structures flat tables into a star schema for optimized analysis. Relationships between accidents, vehicles, drivers, and environmental factors enable multi-dimensional insights. Power BI Desktop creates interactive dashboards, while Power BI Service facilitates sharing via its cloud platform. This streamlined process ensures efficient data-driven decision-making for road safety improvements.

Data Sourcing	https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents
Data Storage	Cleaned data is stored in structured formats compatible with Power BI.
ETL Process	Data is extracted, transformed, and loaded using Python scripts.
Data Modeling	Relationships and data models are established in Power BI
Visualization	Power BI dashboards visualize insights for decision-making.

2.2 System Components

The major components for the system include:

2.2.1. User interface

The user interface (UI) will be designed using **Power BI Desktop** to create interactive dashboards and reports that visualize vehicle accident statistics. It will feature intuitive navigation, interactive charts, filters, and slicers, allowing users to explore accident trends, identify high-risk factors, and analyze data based on various criteria such as vehicle type, weather conditions, and driver demographics.

2.2.2. Backend Processing

Backend processing consists of data preparation tasks such as cleaning, transforming, and modeling. **Python** will handle the initial data preprocessing steps, which include addressing missing values, correcting inconsistencies, and merging multiple data sources. Subsequent data refinement and organization will be managed through **Power BI's Power Query** tool, optimizing the data for effective visualization and analysis.

2.2.3 Database

Accident data will be securely stored in a local environment, enabling direct manipulation and analysis with local software tools. To maintain data integrity and ensure easy access, regular backups will be stored in cloud-based platforms like Google Drive. The organized datasets will then be integrated into Power BI for streamlined modeling and reporting.

3. Data Management

3.1 Data Sourcing

- **Traffic Authority Records:** Official reports on road accidents.
- **Weather Data:** Information on environmental conditions at the time of accidents.

3.2 ETL Process (Extract, Transform, Load)

1. **Extract**:
   Raw data is extracted from **kaggle.com** by downloading available datasets directly from the website. This step ensures that all relevant data pertaining to the chosen topic is captured in its raw format for further processing and analysis.

2. **Transform**:
   The extracted data often contains inconsistencies, missing values, or formats unsuitable for analysis. Using **Python**, the data is cleaned and standardized. This may involve:

   - Removing duplicates or null values
   - Formatting dates and numbers
   - Normalizing text (e.g., making sure names are consistently capitalized)
   - Creating calculated fields or aggregating data for better insights

3. **Load**:
   Once the data is transformed, it is imported into **Power BI**, a data visualization and business intelligence tool. In Power BI:
   - The data is modeled to establish relationships between tables
   - Dashboards and reports are created to visualize key metrics and insights
   - Data refresh schedules are set to ensure reports always reflect the latest information

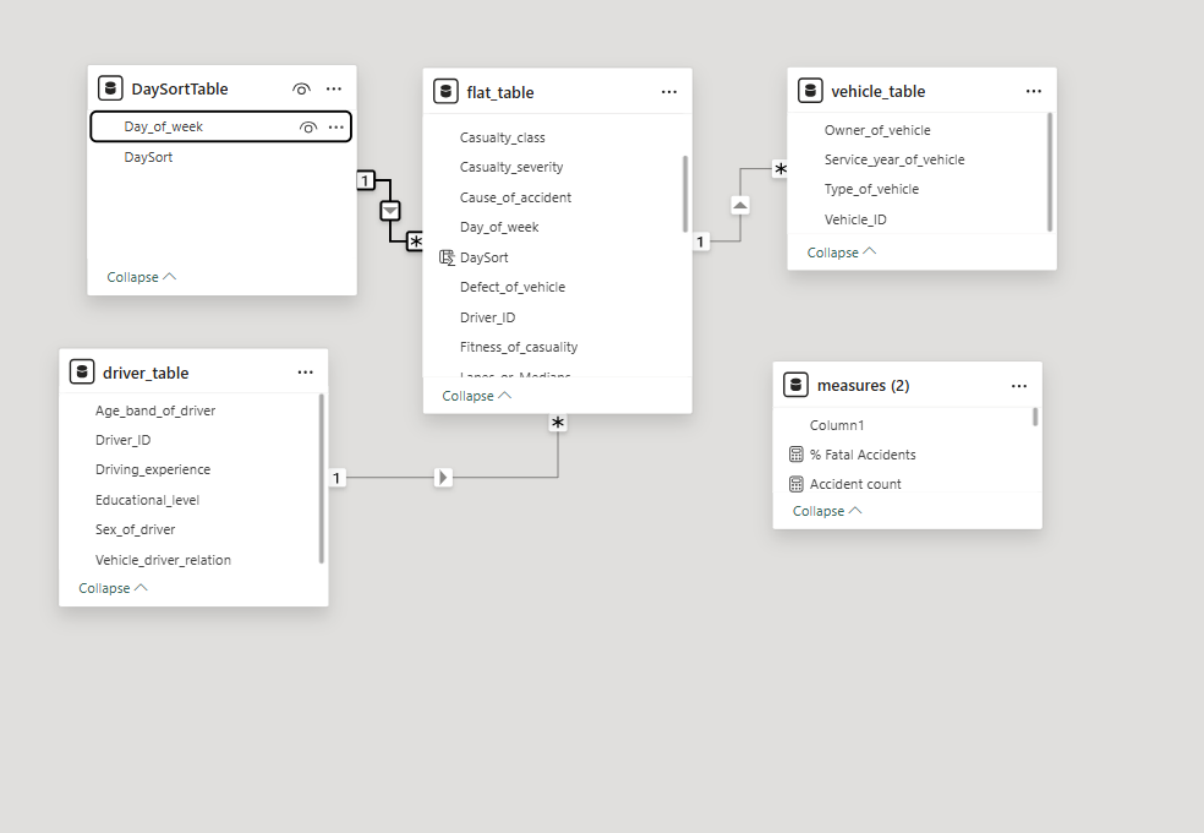
4. Data Modeling

4.1 Entity-Relationship Diagram

The original dataset was normalized into multiple tables and transformed from a flat structure into a **star schema**. This approach enhances analysis and visualization capabilities by organizing data into a **fact table** (**flat_table**, containing accident details) and **dimension tables** (**driver_table**, **vehicle_table**, **DaySortTable**).

4.2 Schema Design

- The **flat_table** holds core accident-related information such as casualty details, causes of accidents, and vehicle or driver IDs.
- The **driver_table** provides demographic and experience-related details of the drivers involved.
- The **vehicle_table** contains information about the vehicles, such as type, ownership, and service history.
- The **DaySortTable** ensures proper chronological sorting for time-based analysis.



5. Visualization & User Interface

5.1 Layout

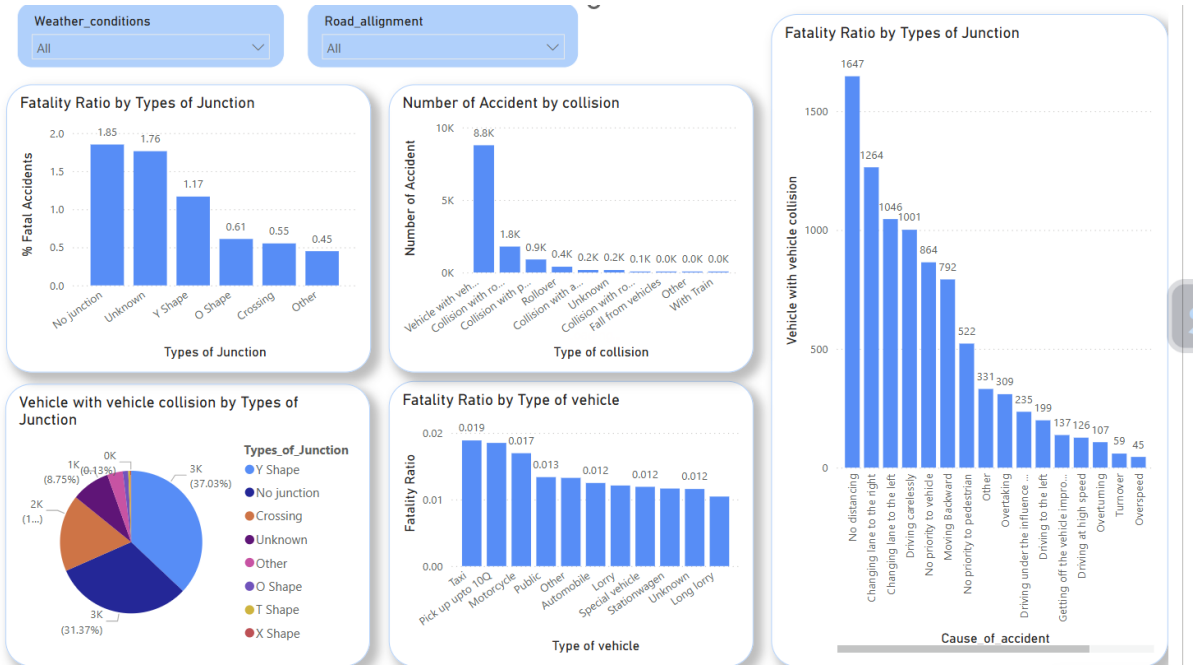
5.1.1 Vehicle Accident Analysis Dashboard

This dashboard provides a comprehensive analysis of vehicle accidents, highlighting key metrics and trends to support data-driven decision-making for improving road safety. The top section showcases the total number of accidents, the percentage of fatal accidents, and the proportion of severe accidents. Users can filter the data by **weather conditions** and **road alignment** to observe how these factors influence accident patterns.
Visualizations include:

- **Accident Distribution by Day of the Week:** Shows the frequency of accidents throughout the week, with **Thursday** and **Friday** experiencing the highest counts.
- **Accident Count by Hour:** Highlights peak accident times, with noticeable spikes during **rush hours** around 8 AM and between 5-7 PM.
- **Driver Demographics:** A pie chart representing the distribution of accidents by the **sex of drivers**, indicating a majority of male drivers involved.
- **Fatalities by Age Group:** Displays fatality counts across different age bands, showing the **18-30** age group as the most affected.

This interactive dashboard, designed using **Power BI**, allows stakeholders to explore accident data dynamically, enabling targeted interventions and policies to reduce accident rates and improve public safety.

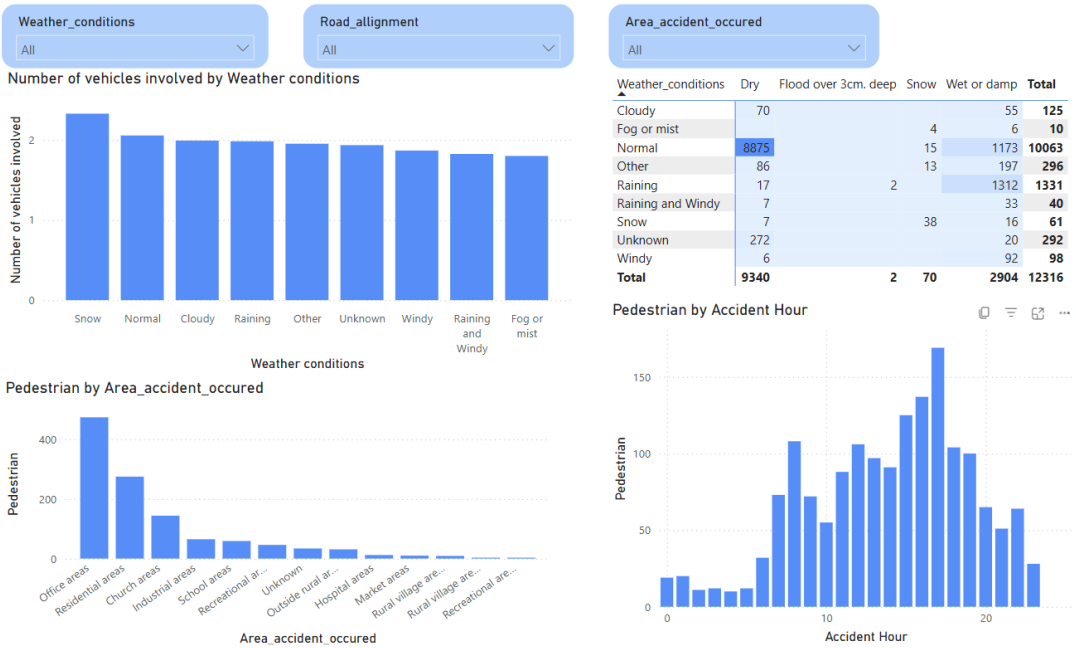
5.1.2 Accident Severity and Collision Analysis Dashboard



This dashboard provides an in-depth analysis of accident severity, collision types, and contributing factors to vehicle accidents. It highlights key metrics like fatality ratios across different **junction types**, **vehicle categories**, and **collision types**. The dashboard includes filters for **weather conditions** and **road alignment** to explore how environmental factors influence accident severity.
Visualizations include:

- **Fatality Ratio by Types of Junctions:** Shows that **No Junction** and **Y-Shape** junctions have the highest fatality ratios, indicating critical areas for safety interventions.
- **Number of Accidents by Collision Type:** Highlights that **vehicle-to-vehicle collisions** are the most common, followed by **collisions with road objects**.
- **Vehicle Collisions by Junction Type:** A pie chart detailing the distribution of collisions at various junction types, with **Y-Shape** and **No Junction** areas contributing to the majority.
- **Fatality Ratio by Type of Vehicle:** Indicates that **taxis** and **motorcycles** have the highest fatality ratios, pointing to the need for stricter regulations.
- **Causes of Accidents:** Lists leading causes such as **no distancing**, **changing lanes improperly**, and **careless driving**.

5.1.3 Weather Conditions and Pedestrian Accident Analysis Dashboard



Weather_conditions	Dry	Flood over 3cm. deep	Snow	Wet or damp	Total
Cloudy	70			55	**125**
Fog or mist			4	6	**10**
Normal	8875		15	1173	**10063**
Other	86		13	197	**296**
Raining	17	2		1312	**1331**
Raining and Windy	7			33	**40**
Snow	7		38	16	**61**
Unknown	272			20	**292**
Windy	6			92	**98**
Total	**9340**	**2**	**70**	**2904**	**12316**

This dashboard offers insights into the relationship between weather conditions, road environments, and pedestrian accidents. It enables users to explore how different **weather conditions** and **areas** impact the frequency and severity of vehicle and pedestrian accidents. Filters for **weather conditions**, **road alignment**, and **accident areas** allow for dynamic data exploration.
Key visualizations include:

- **Number of Vehicles Involved by Weather Conditions:** Highlights that **snowy** conditions lead to a higher number of vehicle-involved accidents, while **normal** conditions account for the majority of overall incidents.
- **Weather Conditions Breakdown:** A detailed table shows the total number of accidents under various weather scenarios, with **normal** and **wet/damp** conditions being the most frequent contributors.
- **Pedestrian Accidents by Area:** Reveals that **office areas** and **residential areas** experience the highest pedestrian accident counts, suggesting targeted safety interventions in these zones.
- **Pedestrian Accidents by Hour:** Demonstrates peak pedestrian accident times, with noticeable spikes during **morning rush hours** and **late afternoon**, aligning with typical work commute periods.

5.2 Features & Functionalities

- **Interactive Elements:** Filters for date, vehicle type, and weather conditions.
- **Filter & Search Options:** Users can filter data based on multiple criteria.
- **Navigation:** Simple navigation between different dashboard sections.

6. Technical Requirements

6.1 Software & Tools

- **Python:** Used for data cleaning, transformation, and initial analysis.
- **Power BI:** Used for data modeling and interactive visualization.

6.2 Hardware Requirements

- **Computational requirements:** Standard computing resources sufficient for processing large datasets.

7. Milestones & Timeline

Milestone	Completion Date
Data Collection	24-jan-2025
Data Cleaning & Transformation	25-jan-2025
Data Modeling	27-jan-2025
Visualization & Dashboard Creation	30-jan-2025
Final Review & Submission	4-Feb-2025