Road Accident Analysis

Submitted by: Lagisetty Bhupala Vignesh

Submitted for: CiSTUP – IISc Bangalore Internship

Date: 22-07-2025

Abstract:

This task includes the improvement of an interactive Power BI dashboard aimed toward analysing road accident data to derive significant insights and trends. The dataset used consists of a huge range of variables which include accident date, severity, climate conditions, road surface types, and time of occurrence. The aim is to recognize the factors contributing to road accidents and present the findings in a user-pleasant visible format.

The raw data was first cleaned and converted to deal with missing values, make certain consistency, and put together it for correct analysis. Key performance indicators (KPIs) have been decided on to monitor patterns which include coincidence frequency over time, accident severity distribution, influence of weather and lighting conditions, and high-risk junctions.

The ensuing dashboard allows dynamic exploration via filters and visuals, imparting stakeholders an effective tool to pick out important risk regions and timeframes. This visualization now no longer most effective complements the information of accident dynamics however additionally helps knowledgeable decision-making for traffic safety enhancements and urban planning.

Objective:

The aim of this assignment is to carry out a complete analysis of road accident data the usage of Power BI to find significant patterns, trends, and risk factors related to traffic incidents. The primary attention is to expand a user-friendly, interactive dashboard that visualizes key signs including coincidence frequency, severity, time-based patterns, environmental conditions, and geographic hotspots.

This analysis seeks to reply to important questions like:

- When do accidents occur most frequently (e.g., by hour, day, month)?
- What conditions—such as weather, road surface, and lighting—are linked to higher accident severity?
- Which types of junctions or road configurations are more prone to accidents?
- How do vehicle count, and casualties vary across different accident scenarios?

The dashboard serves as a decision-support tool for transportation authorities, urban infrastructure planners, and road safety officers. By visualizing the underlying reasons and outcomes of road accidents, the tool facilitates discover precedence regions for intervention and guides the improvement of regulations aimed toward improving road safety, decreasing

fatalities, and enhancing public awareness. Additionally, the project demonstrates technical proficiency in data transformation, data modelling, KPI development, and visualization design within Power BI—highlighting both analytical and storytelling capabilities.

Data Cleaning Steps:

To make sure the accuracy and reliability of the analysis, several records cleansing steps had been completed earlier than visualizing the data. These steps had been crucial to address missing or inconsistent values and put together the dataset for powerful interpretation in Power BI.

1. Import and Initial Inspection:
   The dataset was imported into Power BI, and an initial inspection was performed to recognize the structure and data varieties of every column.
2. Handling Missing Values:
   For data with missing values within the Time column, an upfill method was applied (We can use Downfill or Upfill). Instead of removing the rows or leaving them null, the subsequent method was used:
   - Upfill Method: The missing time values had been full of the most frequent (mode) time value discovered within the dataset. This approach assumes that accidents tend to cluster around certain hours (e.g., 08:00 or 17:00), so the usage of the most common time allows hold the integrity of time-based analysis.
   - Alternative Options Considered In a few scenarios, filling missing values with a neutral price like “00:00” or labelling them as “Unknown” should lie to trends. Upfilling with mode (or the usage of the median hour) avoids introducing synthetic skew.
   - Impact Filling in missing time data enabled the development of accurate visuals for hourly trends, time-of-day KPIs, and helped save you visual gaps in heatmaps or line graphs that depend on entire time information.
3. Data Type Corrections:
   Fields were converted to appropriate data types, such as numeric or date formats, to enable correct aggregation and sorting in visuals.

These data cleaning methods helped create an elegant dataset that helps accurate insights and an efficient analytical revel in in the dashboard.

Derived Columns:

To allow extra powerful analysis and visualization, several new columns have been derived from existing fields within the dataset. These extra fields have been created to higher understand time-based trends, enhance filtering, and simplify visual illustration throughout the dashboard. The key derived columns include:

1. Month: Extracted the month name from the accident date. This made it less complicated to analyse accident trends throughout different months.
2. Month Number: A numeric version of the month (from 1 to 12) was created to ensure that months were sorted chronologically in charts and tables.
3. Year: Separated the year from the accident date to allow year-wise trend analysis and comparison.

4. Quarter: Each accident was assigned to a quarter (Q1 to Q4) to identify seasonal trends in accident occurrence.
5. Week Number: The week of the year was derived from the accident date to explore short-term patterns and weekly fluctuations.
6. Hour: By extracting the hour from the time column, we could study the distribution of accidents across different times of the day.
7. Time of Day: Accidents were categorized into broader time segments such as Morning, Afternoon, Evening, and Night. This helped in identifying high-risk periods during the day.

These additional fields enhanced the ability to drill down into specific timeframes and supported more flexible and insightful dashboard visuals.

Key Insights:

The Power BI dashboard provides valuable insights into various factors influencing road accidents:

- Peak Hours and Time of Day:
  The analysis indicates that most accidents occur during early morning and evening hours, suggesting peak traffic periods. This is visualized using bar charts based on the "Hour" and "Time_of_Day" fields.
- Monthly and Quarterly Trends:
  A line chart and drill-down hierarchy help reveal trends in accidents over months and quarters. Seasonal patterns are observed, with certain months showing higher accident counts.
- Junction Risk Analysis:
  Charts based on "Junction_Control" and "Junction_Detail" highlight that uncontrolled junctions or those with complex layouts are more prone to accidents.
- Weather and Light Conditions:
  Most accidents occurred under clear weather but poor light conditions, especially during dark periods with or without streetlights. These factors are visualized using stacked column charts.
- Road Type and Speed Limits:
  Accidents are more frequent on single carriageways and roads with speed limits around 30–40 mph, as shown in relevant visualizations.
- Vehicle Type Involvement:
  Analysis of "Vehicle_Type" indicates that cars are involved in the majority of accidents, followed by motorcycles and goods vehicles.

KPI Used:

Key Performance Indicators (KPIs) play a crucial role in summarizing the overall state of road safety and helping users quickly assess critical accident statistics. In this dashboard, four major KPIs have been included at the top of the main page to provide a high-level overview:

1. Total Number of Accidents:
   This KPI reflects the total number of road accident cases recorded in the dataset. It provides a quantitative snapshot of how frequently accidents occur and helps assess the

general traffic safety conditions. This value dynamically adjusts based on filters like year, weather, or road type, allowing for focused analysis.

2. Total Casualties:
   This metric displays the cumulative number of people affected by the accidents—both injured and deceased. It gives insight into the human impact of road incidents and can help prioritize safety policies in areas with high casualty rates.
3. Total Vehicles Involved:
   This indicator tracks the total number of vehicles that were involved in the reported accidents. Monitoring this KPI helps understand the scale of collisions and whether incidents typically involve single or multiple vehicles.
4. High Severity Accidents:
   This KPI focuses on the number of accidents labeled with the highest severity level. It helps identify extremely dangerous situations and serves as a red flag for areas or conditions where such accidents occur frequently.
5. Interactivity with Filters and Slicers:
   Each KPI is designed to interact with slicers like Year, Quarter, Month, Junction Type, Light Conditions, Weather Conditions etc... When a user selects a filter, the KPIs instantly update, enabling comparative and exploratory data analysis.

These KPIs not only offer a summary of accident data but also guide users toward deeper insights by connecting with visual elements throughout the dashboard.

Dashboard Page Overview:

The Power BI dashboard is structured into two well-organized and interactive pages: Page 1 and Page 2. Each page is designed with a focus on usability, visual clarity, and insightful analytics, making it easier for users to understand patterns and derive conclusions from the road accident data.

- Page 1:

  This Page serves as a dashboard landing page, providing a snapshot of road accident trends and conditions through rich visuals and KPIs. It includes the following key components:

  1. KPI Cards:
     - No. of Accidents by Hour
     - No. of Accidents by Week
     - No. of Accidents by Month
     - No. of Accidents by Year

  2. Visual Analysis:
     - Column Chart: Accidents by Severity – categorizes incidents into Fatal, Serious, and Slight.
     - Line Chart: Accidents Over Time – highlights accident trends over different time periods (e.g., months), allowing seasonal patterns to be observed.
     - Bar Chart: Accidents by Day of Week – helps identify the riskiest days.

- Bar Chart: Top Junction Controls Contributing to Accidents – highlights which types of traffic control mechanisms at junctions are most associated with road accidents.

3. Interactive Slicers:
   - Slicers include Accident Date, Junction Control, Weather Conditions, Road Surface Conditions, Time of Day, Light Conditions, Day of Week, Month, Quarter, Year
   - Slicers enable dynamic data filtering across all visuals, enhancing analytical flexibility.

- Page 2:

This page dives deeper into specific aspects of the data, offering more focused and analytical views for advanced users.

1. Visual Analysis:
   - Line chart: Road Accident Trends: Year → Quarter → Month – Tracks accident counts over time using a date hierarchy (Year > Quarter > Month).
   - Donut chart: Accidents by Road Surface – Shows how the number of road accidents varies based on the condition or type of the road surface at the time of the accident.
   - 3D Map Visualization: A 3D map visual is used to geographically plot the accident locations using latitude and longitude – Shows accident counts at what time of day (e.g. Afternoon, Evening etc...)
   - Bar Chart: Weather Conditions – shows how environmental conditions correlate with accident frequency.
2. Synced Filters:
   - Slicers are synced with Page 1 and Page 2 ensuring consistency in filtering across pages.

- Navigation Bar:
  - To enhance user experience and enable seamless movement between the two report pages, a custom navigation bar has been implemented using Power BI buttons:
  - Navigation buttons are placed clearly on both Page 1 and Page 2 labelled as "➡" and "⬅"
  - Clicking these buttons allows users to switch between the summary and detailed analysis views without needing to scroll or search manually.
  - This navigation setup improves report usability, especially for non-technical users or stakeholders who want to move between pages effortlessly.

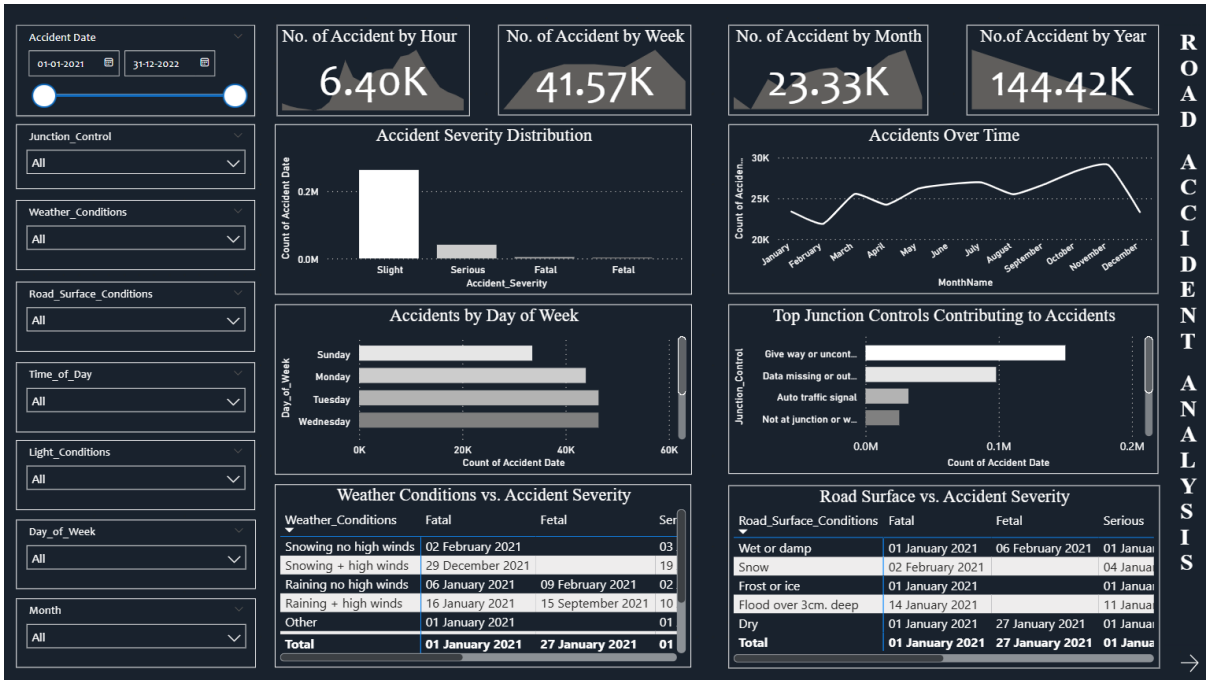

Fig 1: Page 1

Fig 2: Page 2

Conclusion:

The road accident analysis dashboard presents a clean and interactive view of the various elements contributing to road mishaps throughout regions. By inspecting data inclusive of accident severity, road surface conditions, lighting, weather, time, and location details, we had been able to find extensive tendencies and insights.

The dashboard discovered that maximum injuries came about on dry roads and under daylight conditions, indicating that human factors like distraction or speeding might also additionally

play a bigger role than environmental conditions. Peak hours and certain days also confirmed better accident frequencies, which could assist in scheduling traffic manage measures. Additionally, the insights approximately risky junction sorts and affected regions can manual city planners and traffic authorities in implementing protection interventions.

The use of interactive visuals, navigation controls, filters, and a 3-D map made the dashboard user-pleasant and insightful for stakeholders. Overall, this analysis highlights the significance of non-stop monitoring, data-driven decision making, and focused strategies to enhance road safety and decrease accident rates.