**MIS 6940, Strategic Safety Solutions: A Data-Driven Approach to Vehicle Design**

**Deepa Palariya**

Final Project Report

## ***Organization description, problem statement and approach used.***

The project aims to utilize crash investigation data from the [National Highway Traffic Safety Administration](https://www.nhtsa.gov/) (NHTSA) to enhance vehicle safety features. This data, sourced from the [Fatality Analysis Reporting System (FARS)](https://www.nhtsa.gov/file-downloads?p=nhtsa/downloads/FARS/), offers comprehensive insights into fatal traffic crashes nationwide, covering various factors influencing crash dynamics. With a focus on the Fatality Analysis Reporting System (FARS) dataset, which encompasses fatal traffic crash data from all 50 states, the District of Columbia, and Puerto Rico, the project will specifically concentrate on the latest year's data (2021) for the mid-western states due to the vastness of the dataset. The primary objective is to identify trends, patterns, and contributing factors in crash occurrences to guide the development, adjustment, or introduction of safety solutions for vehicles. Adherence to safety regulations mandated by organizations like the NHTSA is crucial to maintain public trust and avoid regulatory penalties. Furthermore, the project aims to gain a competitive edge in the automotive market by pioneering state-of-the-art safety features that resonate with safety-conscious consumers.

My approach involves leveraging crash investigation data to inform safety solutions aimed at enhancing vehicle safety standards. This includes collecting, cleaning, and validating data to create a robust data warehouse with master and transaction tables for final analysis. By utilizing this data-driven approach, we aim to develop innovative safety features that not only address existing safety concerns but also comply with regulatory requirements and provide a competitive edge in the market.

This model is dedicated to advancing vehicle safety standards, addressing existing safety concerns, and innovating safety technology to reduce the frequency and severity of traffic crashes. Compliance with safety regulations enforced by entities like the NHTSA is paramount, ensuring the organization maintains public trust and avoids penalties. Additionally, my goal is to gain a competitive advantage in the automotive market by developing cutting-edge safety features that attract safety-conscious consumers.

The primary challenge lies in enhancing vehicle safety standards to mitigate the risks associated with traffic crashes. Compliance with stringent safety regulations, such as those mandated by the NHTSA, is essential to uphold public trust and avoid regulatory penalties. Furthermore, there is a need to innovate safety technology to stay ahead in the competitive automotive market.

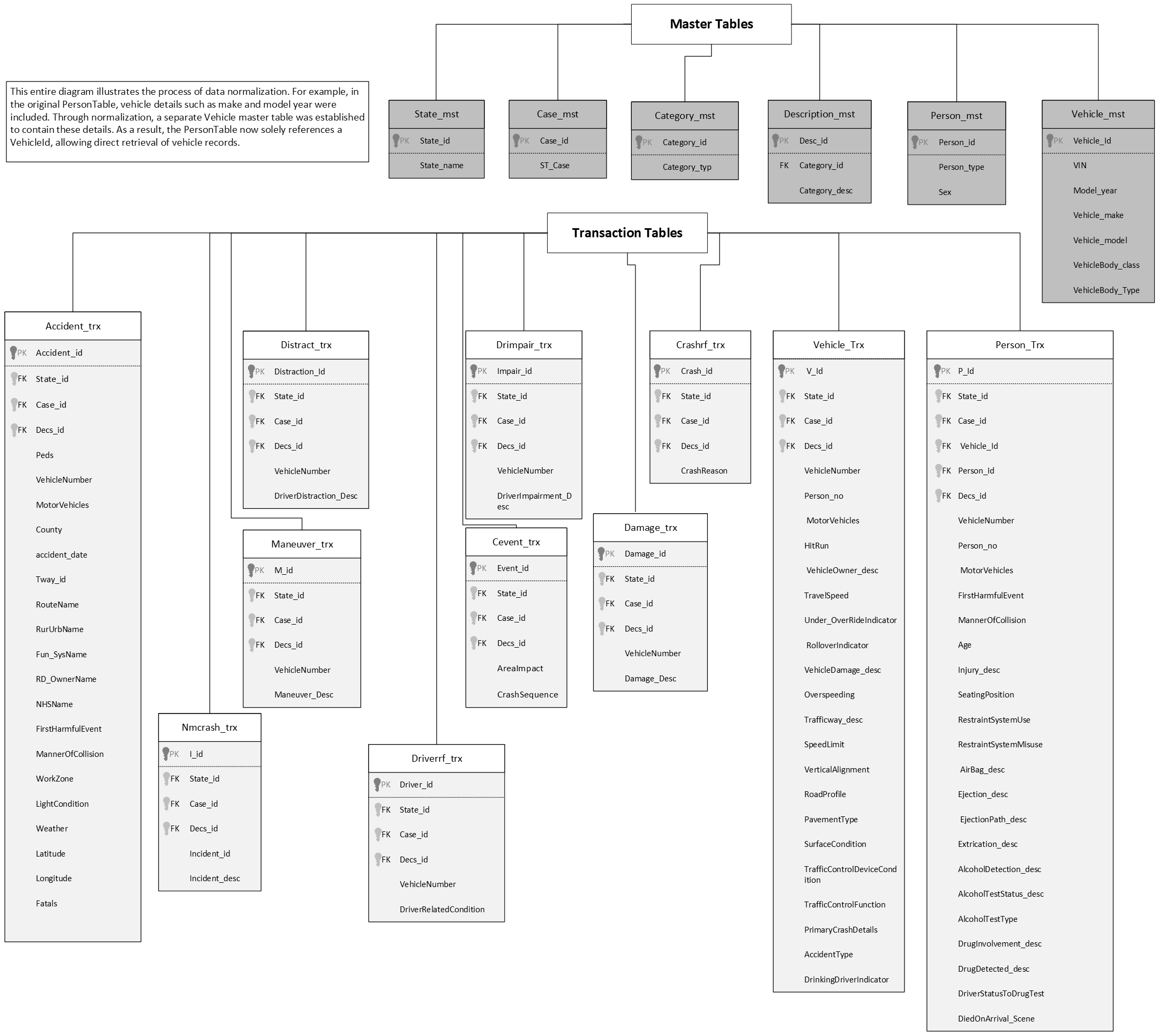
## ***Overview of Decision-Making Challenges and the Role of DW/BI Applications:***

1. Analyzing Crash Causes: The DW/BI application enables the identification of trends and patterns in crash occurrences, shedding light on primary causes such as human error, mechanical failure, or environmental factors.
2. Assessing Safety Measures: Through the DW/BI application, stakeholders can evaluate the effectiveness of existing safety features and protocols in vehicles, guiding decisions on enhancements or the introduction of new safety measures.
3. Prioritizing Safety Solutions: By examining crash severity and incident frequency, stakeholders can prioritize safety solutions to reduce severe crashes or address common contributing factors.
4. Monitoring Performance: The DW/BI application offers real-time monitoring of safety performance metrics, facilitating the tracking of implemented safety solutions' impact and allowing for timely adjustments.

Overall, the DW/BI application serves as a comprehensive tool for analyzing crash investigation data and informing strategic decisions aimed at enhancing vehicle safety standards, compliance with regulations, and gaining a competitive edge in the automotive industry.

## ***Sample Master Table and Transaction Table Structure:***

Figure 1 illustrates the fundamental layout of my database, comprising six master tables and 11 transaction tables. Each master table serves as a dimension table, while each transaction table functions as a Fact table, embodying a snowflake schema. These tables collectively facilitate the derivation of essential KPIs necessary for devising strategic safety solutions for vehicles.



**Figure 1 – Sample Table Structure**

## ***The specifications for the DW/BI solution and Schema and design of the DW/BI application.***

I utilized Excel, MSSQL, Access, and SSIS to manage data effectively. Although I didn't employ cubes, as I had already incorporated their functionality during the normalization of data and creation of master and transaction tables.

In Figure 2, each transaction table is linked to its corresponding master table, illustrating the interconnectedness of the data within the complex data warehouse. While detailing every relationship here is impractical due to the warehouse's complexity, it's evident that I effectively managed data using Excel, MSSQL, Access, and SSIS. Despite not utilizing cubes, their functionality was integrated during the data normalization and creation of master and transaction tables.

A diagram of a chart

Description automatically generated with medium confidence

**Figure 2 – DW Structure**

Encountering a dataset fraught with redundant and denormalized data prompted me to rectify these inefficiencies and potential integrity issues. I initiated a normalization process by crafting multiple master tables as necessary.

Figure 3 elucidates the normalization procedure using two master tables: "Category" and "Description." The former houses distinct categories like Route Name and First Harmful Event, while the latter stores corresponding descriptions for each category. Instead of duplicating data in transaction tables, we now reference the appropriate descriptions via unique description IDs.

For example, let's consider the "Accident" transaction table, which initially contained redundant data for Route Name and First Harmful Event. Through normalization, redundant values were eradicated, replaced with corresponding description IDs from the master tables.

**Normalization brings several benefits:**

1. Enhanced Data Efficiency: The elimination of redundant data reduces storage requirements and enhances query performance.
2. Improved Data Integrity: Centralizing category descriptions ensures consistency and accuracy throughout the dataset.
3. Streamlined Data Retrieval: Analysts can effortlessly link description IDs to category variables, simplifying data retrieval and analysis.

In summary, this normalization approach has optimized our dataset, bolstering efficiency, integrity, and usability. Furthermore, it obviates the need for a separate cube/DW creation.**A screenshot of a computer

Description automatically generated**

**Figure 3 – Demo**

Figure 4 illustrates, the schema for the DW/BI is of a snowflakes variety showing the list of tables and their relationships in the DW database. All Transaction tables act as fact tables followed by master tables acting as dimension tables. These were necessary to allow BI analysis and keep the system flexible for future expansions including keeping track of additional data such as category of accidents, new vehicle details, additional states, persons, crash reasons etcetera which is stored in unique master tables.

In the Description Master table, the Desc\_id serves as the primary key, acting as a foreign key in nearly all transaction tables. For example, in the Accident\_trx table, attributes like RouteName, LightCondition, and Weather each have a corresponding Desc\_id linked to the Description master table, ensuring a unique record connection.

This normalization process involved assigning integer values to each category across the entire database, simplifying its structure and eliminating redundancy. Furthermore, it facilitates streamlined query execution, enhancing database management efficiency.

**Figure 4 – DW Database Schema**

A computer screen shot of a computer

Description automatically generated

## ***e) Helping views and queries created to account for the KPIs (screen shots) demonstrating how my application is used to address the problems/KPIs identified.***

Now that the DW is set up and populated, my next step was to establish some views to calculate KPIs. I considered either using MSSQL to write queries or Power BI to write DAX functions. Eventually, I opted for MSSQL for both views and KPIs. Below are some examples of the views created, which served as the basis for querying and building KPIs for visualization purposes.

1. A screenshot of a computer

   Description automatically generatedCrash Severity Analysis
2. Vehicle and Person

**A screenshot of a computer

Description automatically generated**

1. Primary Crash Details and Severity

A screenshot of a computer

Description automatically generated

1. A screenshot of a computer

   Description automatically generatedCrash Characteristics
2. Injury/Damage

A screenshot of a computer

Description automatically generated

A white text on a white background

Description automatically generated

The below screenshots helped answer all the above-listed questions.

A close-up of a graph

Description automatically generated

A close-up of a graph

Description automatically generatedA screenshot of a graph

Description automatically generated

A close-up of a graph

Description automatically generatedA close-up of a chart

Description automatically generated

## ***f) Final Analysis/Findings:***

Based on findings in Power BI, below are the key observations:

1. The maximum crash count was noted in the age group ranging from 19 - 55, leading to fatal injuries, with Illinois recording the highest number of crashes.
2. The top 5 vehicle makes and models involved in crashes in the mid-west region were Chevy Silverado, Impala, Malibu, Equinox, Ford F-150, Fusion, Escape, Explorer, Dodge Ram, Toyota Camry, and Honda Civic.
3. Certain vehicles such as Chevy Silverado, Impala, Malibu, Ford F-150, and Explorer recorded fatal injuries more frequently than others.
4. Illinois, Missouri, Ohio, and Michigan recorded the highest crash counts, primarily dominated by males.
5. Maximum crashes occurred during clear weather with wet surface conditions.
6. Most crashes occurred when going straight, with the incident being failure to yield right of way.
7. When crashes were highest in states like Illinois and Missouri, the prevailing light conditions were dark - lighted and dark - not lighted.
8. Speeding crashes accounted for a significant portion of total crashes, with Iowa having the highest percentage of speeding crashes to total crashes.
9. Michigan had the highest percentage of impaired driving crashes, followed by Wisconsin, Iowa, and Kansas.
10. Distracted driving leading to crashes was highest in Minnesota, followed by Michigan and Nebraska.
11. The manner of collision was reported as not a collision with a motor vehicle in transport in most cases.
12. Michigan had the highest hit-and-run count.
13. State highways were the most common location for crashes, with Ohio recording the highest number, followed by Missouri, Illinois, and Michigan.
14. Fatal injuries were the most common type of injury reported, followed by no injury.
15. Shoulder and lap belts were used in 90% of crashes for restraint system use.
16. The front seat left side seating position was most involved in crashes.
17. Airbags were not deployed in most cases.
18. Extrication was not required in most crashes.

Below are the safety solutions proposed based on the findings:

1. **Targeted Educational Campaigns and Law Enforcement Initiatives:**

* Implement educational campaigns and law enforcement initiatives aimed at high-risk age groups to promote safer driving behaviors.
* Integrate driver monitoring systems in vehicles to track and alert drivers of risky behaviors like distracted driving, speeding, and fatigue in real-time.

## **Integration of Advanced Driver Assistance Systems (ADAS):**

* Enhance vehicle safety by integrating ADAS features such as lane departure warning, blind-spot detection, and automatic emergency braking.
* These systems assist drivers in avoiding collisions by enhancing situational awareness and providing active intervention when necessary.

## **State-Specific Road Safety Initiatives and V2V Communication Systems:**

* Implement state-specific road safety initiatives and integrate vehicle-to-vehicle (V2V) communication systems.
* These initiatives and systems enable vehicles to exchange critical information about speed, position, and direction with nearby vehicles, enhancing cooperation and helping drivers avoid collisions.

## **Integration of Weather-Awareness Systems:**

* Improve driver awareness and response to adverse weather conditions by integrating weather-awareness systems into vehicles.
* These systems provide real-time weather updates and alerts to drivers, enabling informed decision-making and safer driving practices.

## **Installation of Speed Limiters:**

* Enhance speed reduction efforts by installing speed limiters in vehicles.
* These devices restrict the vehicle's speed to comply with posted speed limits, reducing the risk of speeding-related crashes.

1. **Development of Alcohol Ignition Interlock Devices:**

* Prevent impaired driving incidents by developing alcohol ignition interlock devices.
* These devices require drivers to pass a breathalyzer test before starting their vehicle, preventing intoxicated individuals from driving, and mitigating the occurrence of alcohol-related crashes.

## **Integration of Driver Monitoring Systems with Distracted Driving Detection:**

* Combat distracted driving by integrating driver monitoring systems equipped with distracted driving detection capabilities.
* These systems use cameras and sensors to detect driver distraction and provide alerts or interventions to help them refocus on the road.

## **Equipping Vehicles with Onboard Cameras:**

* Enhance hit-and-run prevention measures by equipping vehicles with onboard cameras that record video footage of their surroundings.
* This footage could serve as valuable evidence in identifying perpetrators and holding them accountable for hit-and-run incidents.

## **Installation of Connected Vehicle Infrastructure:**

* Improve road safety infrastructure by installing connected vehicle infrastructure on highways and roads.
* This technology allows vehicles to communicate with traffic signals and signs, providing drivers with real-time traffic information and warnings about road hazards.

## **Integration of Active Seat Belt Systems:**

* Promote optimal restraint systems in vehicles through the integration of active seat belt systems.
* These systems automatically adjust seat belt tension based on driving conditions and crash severity, reducing the risk of injury to occupants during a crash.

## **Development of Advanced Occupant Sensing Systems:**

* Enhance airbag deployment effectiveness through the development of advanced occupant sensing systems.
* These systems use cameras and sensors to detect the presence, position, and size of occupants in the vehicle, optimizing airbag deployment to reduce the risk of injury.

## **Equipping Vehicles with Emergency Communication Systems:**

* Improve emergency response capabilities by equipping vehicles with emergency communication systems.
* These systems automatically alert emergency responders in the event of a crash, providing critical information about the crash location and severity for a faster and more effective response.

## ***g) Summary/Lessons learned/Limitations:***

Collecting and preparing data from various sources and storing it in a data warehouse can be incredibly valuable for analyzing crashes in the US and deriving necessary solutions. By leveraging decision support systems like Power BI to analyze this data, insights can be gained into various aspects of crashes, including severity, causative factors, and demographic trends. This enables informed decision-making and the development of targeted interventions to improve road safety.

By implementing these measures in a coordinated manner, road safety can be significantly enhanced, resulting in a reduction in the frequency and severity of crashes. Ultimately, this approach saves lives and reduces the societal and economic impact of road traffic accidents, fostering safer roadways for all.

However, it's important to note a limitation: the absence of detailed information on specific vehicle parts and safety features may hinder the ability to conduct in-depth vehicle analysis. Access to such data could provide valuable insights into the effectiveness of safety features and inform strategies for improving vehicle safety standards.

***h) Instructions to run the entire DW/BI application:***

The essential components required to run the entire Data Warehouse (DW) and Business Intelligence (BI) application include the Database Management System (DBMS) Microsoft SQL Server (MSSQL) and the BI tool Power BI. Attempting to execute scripts in alternative DBMS platforms like PostgreSQL or Oracle may encounter issues due to differences in syntax, as the tables are created using MSSQL queries.

To initiate the process, attach the provided .mdf file named Capstone to your MSSQL server, ensuring seamless operation of the DW. Subsequently, import the data directly into Power BI via SQL Database, as the dataset is extensive, containing approximately 20 million rows in one table. Once the data is loaded into Power BI, the pre-configured dashboards will be readily accessible.

**List of files inside the folder:**

1. **FARS2021NationalCSV\_RawData:** This folder contains all the Excel files sourced from the FARS data on the NHTSA website for the year 2021.
2. **All cleaned files:** Here, you'll find the cleaned data files derived from the FARS raw datafiles folder, each containing pertinent data.
3. **MasterTables:** This folder comprises all master tables utilized in normalized form for data insertion into the MSSQL server DW.
4. **TransactionTables:** Contains transaction tables in normalized form for data insertion into the MSSQL server DW.
5. **MSSQL .mdf DW\_Capstone:** This folder contains the complete MSSQL database, including the .mdf file and log file (Capstone.mdf and Capstone\_log), intended for attachment to any MSSQL server.
6. **PowerBIAnalysis:** Contains the Power BI dashboard files (.pbix and .pbit formats) along with a PDF version for convenience.
7. **Scripts:** Includes scripts for Create, Insert, Views, and KPIs queries, covering the entire database.
8. **VisioDiagrams:** Contains diagrams illustrating the table structure and demo utilized in this report.
9. **FinalReport:** This folder houses the final project report ("Deepa Palariya\_Final Project Report") in both Word document and PDF formats.

These resources collectively facilitate the setup and execution of the DW/BI application, providing comprehensive insights into crash investigation data sourced from the FARS dataset for informed decision-making and safety solution development.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated