Road Accident Analysis 2021 - 2022 Using Power BI

Ashley Vanessa Williams
LinkedIn: https://www.linkedin.com/in/ashleyvanessawilliams/
Email: ashleyvanessawilliams@gmail.com

Summary

1 INTRODUTION 3

2 REQUIREMENTS 4

2.1 DESCRIPTION 4

3 BI PROCESS PROPOSAL 5

4 ETL PROJECT 6

4.1 DESCRIPTION OF THE ETL PROJECT 6

5 DATA WAREHOUSE ELABORATION 7

5.1 DESCRIPTION 7

6 DASHBOARD 7

6.1 ELABORATION OF THE DASHBOARD 7

DASHBOARD SCREENS 7

7 CONCLUSION 8

1 Introduction

This is Business Intelligence (BI) project on Road Accident, aimed at providing valuable insights and analytics related to road safety and accident data. The goal of this project is to help stakeholders make informed decisions, identify areas for improvement, and take proactive measures to reduce road accidents and their impact on human lives and property.

2 Requirements

Road accidents pose a significant threat to public safety and can result in devastating consequences. Therefore, it is crucial to analyze and understand the underlying factors that contribute to these accidents and their severity. By leveraging business intelligence tools and data analytics, we aim to empower our clients with comprehensive visualizations and reports that shed light on key performance indicators (KPIs) related to road accidents.

2.1 Description

Data Collection and Integration: The project begins with the collection and integration of diverse road accident data from various sources, such as traffic authorities and law enforcement agencies. The data includes accident dates, locations, vehicle types, accident severity, casualties, and other relevant attributes.

Key Performance Indicators (KPIs) Definition: The project defines primary and secondary KPIs that will be the core focus of the data analysis. The primary KPIs include Total Casualties, Total Accidents, and Total Casualties by Accident Severity, with a focus on current year values and year-on-year (YoY) growth. The secondary KPIs involve Casualties with respect to vehicle type, Casualties by Road Type, Current Year Casualties by Area/Location & Day/Night, and Total Casualties and Total Accidents by Location.

Data Cleansing and Transformation: To ensure the accuracy and consistency of the analysis, the collected data undergoes a thorough cleansing and transformation process. This step involves handling missing values, standardizing data formats, and correcting any errors or inconsistencies.

Data Visualization and Reporting: Leveraging cutting-edge BI tools and techniques, the project generates interactive and visually appealing data visualizations and reports. These visualizations enable stakeholders to explore the data intuitively and gain valuable insights into accident trends, patterns, and correlations.

Current Year vs. Previous Year Comparison: One of the key highlights of the project is the comparison between the current year's accident data and the data from the previous year. This comparison allows stakeholders to identify trends and deviations in accident patterns, helping them assess the effectiveness of safety initiatives.

Accident Hotspot Identification: The project employs spatial analysis techniques to identify accident hotspots, areas with a high frequency of accidents and casualties. By pinpointing these hotspots, stakeholders can prioritize their efforts and allocate resources where they are most needed.

Recommendations and Actionable Insights: Based on the data analysis and visualizations, the project provides actionable insights and recommendations for enhancing road safety. These recommendations may include targeted safety campaigns, infrastructure improvements, and traffic management strategies.

3 Proposal of BI processes

In response to the following company´s requirements, aiming to better understand the data, a Data Warehouse was created. The data sources used in the project consist of Excel files, as previously mentioned in item 3. The Data Warehouse modeling was carried out in Power Architect (V1.0.9), an open source and free tool. For data transmission, transformation and loading, PDI (Pentaho Data Integration - v9.3.0.0) was used, an open-source software developed in Java. The data was then stored in Postgres (v15), an open-source relational database management system.

3 Extract, Transform and Load (ETL) Phase

3.1 Description

The ETL process is a critical component of the Road Accidents BI Project, responsible for extracting raw data from various sources, transforming it into a consistent and usable format, and loading it into a centralized data repository. This process ensures that the data used for analysis and reporting is accurate, reliable, and easily accessible for business intelligence purposes.

3.1.1 Extract:

The first stage of the ETL process involves extracting data from multiple sources that contain information related to road accidents. These sources may include traffic authorities, law enforcement agencies, insurance companies, and other relevant data providers. The data extraction process can be automated using scripts, APIs, or data integration tools to retrieve data from various formats, such as databases, spreadsheets, or flat files.

3.1.2 Data Sources:

1. Traffic authorities' accident databases containing details about accidents, casualties, accident severity, and location information.
2. Law enforcement reports and citations related to road accidents.
3. Insurance company records with information about vehicle types and accident claims.
4. Geographic Information System (GIS) data for road types and locations.

3.1.3 Transform:

In the transform stage, the extracted data undergoes a series of data cleansing, validation, and transformation operations to ensure its quality and consistency. The primary goal is to prepare the data for meaningful analysis and reporting by converting it into a standardized format.

Data Transformation Steps:

1. Data Cleansing: Identify and handle missing or erroneous data, ensuring data accuracy and integrity.
2. Data Standardization: Convert data into a consistent format, such as standardizing date formats or location codes.
3. Data Aggregation: Group and aggregate data to create meaningful KPIs, such as total casualties and accidents for each year and month.

3.1.4 Load:

In the load stage, the transformed and cleansed data is loaded into a centralized data repository, often referred to as a data warehouse. This repository acts as a single source of truth for the BI project, providing a unified view of road accident data for analysis and reporting.

4 Data Warehouse:

The data warehouse will store data from all sources in a structured and optimized manner. It will allow for quick and efficient data retrieval during analysis and reporting processes. It will be designed to support complex queries and aggregations required for generating BI insights.

5 Data Integration and Scheduling:

To ensure the data warehouse remains up-to-date with the latest information, a data integration process is established. This process can include regular scheduled data loads or real-time data feeds, depending on the data sources' update frequency.

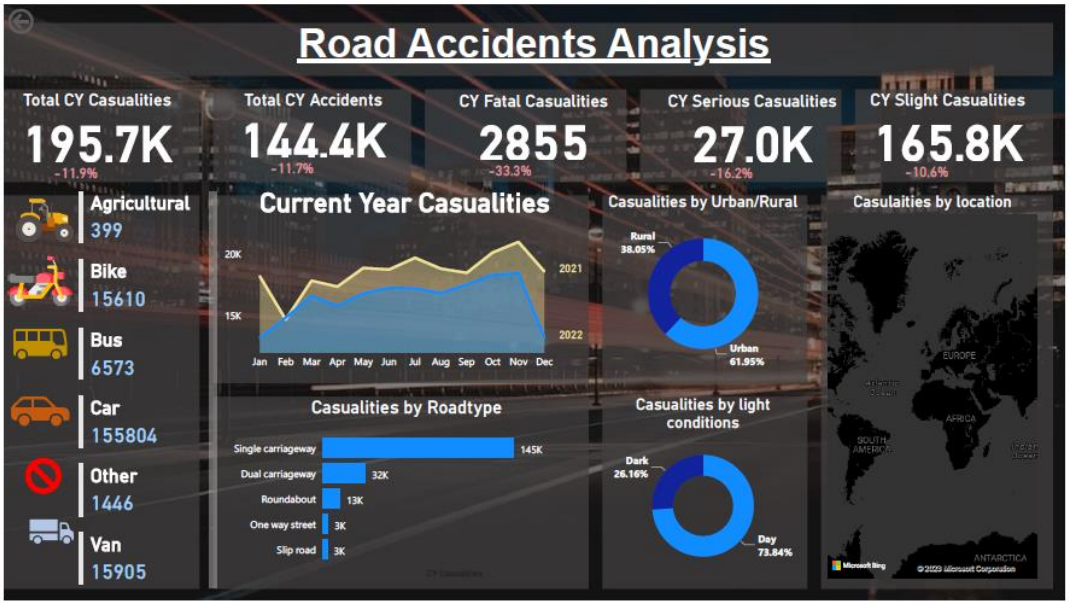
6 Dashboard

6.1 Description

O Dashboard was created on Power BI Desktop. It cam be found HERE

6.2 Dashboard Screen

Dashboard



7 Conclusion

The completion of the Road Accidents Business Intelligence (BI) Project marks a significant milestone in our pursuit of enhancing road safety and reducing the impact of road accidents. Throughout this endeavor, we have harnessed the power of data analytics and visualization to gain valuable insights into road accident trends, casualties, and contributing factors. The project's findings and actionable recommendations hold the potential to make our roads safer for all users and create a positive impact on public safety.

Through in-depth analysis of primary and secondary key performance indicators (KPIs), we have gained a comprehensive understanding of road accident patterns. The comparison of current year data with previous years has revealed trends, enabling stakeholders to gauge the effectiveness of safety initiatives. By examining accidents based on severity, vehicle types, road types, and locations, we have identified high-risk areas, accident hotspots, and specific risk factors contributing to accidents.