

BUSINESS INTELLIGENCE AND DATA WAREHOUSE TECHNOLOGIES FOR TRAFFIC ACCIDENT DATA ANALYSIS IN BOTSWANA

Monkgogi Mudongo, Edwin Thuma, Nkwebi Peace Motlogelwa,
Tebo Leburu-Dingalo and Pulafela Majoo

Department of Computer Science,
University of Botswana, Gaborone, Botswana

ABSTRACT

Road traffic accidents are a serious problem for the nation of Botswana. A large amount of money is used to compensate those who are affected by road accidents. Traffic accidents are one of the major causes of Deaths in Botswana. It is important for relevant organizations to have a reliable source of data for accurate evaluation of traffic accidents. Similarly, data on vehicle registration must be transformed and be readily available to assist managerial decision makers. In this article, we deploy a Business Intelligence (BI) and Data Warehouse (DW) solution in an attempt to assist the relevant departments in their road traffic accidents and vehicle registration evaluation. In Our evaluation of the traffic accidents our findings suggest that across accident severity, Damage Only accidents had the most interesting recent trend with a 11.93% decrease in the last 3 years on record. Count of Accident Severity for Damage Only accidents dropped from 13,491 to 11,881 between 2018 and 2020 whilst Minor accidents experienced the longest period of growth. Most accidents take place in rural locations and more accidents take place during the weekend. At 28,439, Sunday had the highest number of accidents and was 47.59% higher than Wednesday, which had the lowest count of accidents at 19,269. The results for vehicle registration reveal that the number of vehicle registration decreased for the last 3 years on record. The number of vehicles registered dropped from 65535 to 24457 during its steepest decline between 2019 and 2021.

KEYWORDS

Business Intelligence, Data Warehousing, ETL, Accident and Vehicle registration.

1. INTRODUCTION

Organizations generate and accumulate data on daily basis. These data become so huge with time and can be of no help if the right tools are not deployed to process it. It is imperative that data be converted to information then knowledge and ultimately to wisdom that decision makers can use to make quality decisions thus the need to manage information and knowledge in any organization. Due to constant change in business environment, organizations are forced or prompted to respond quickly to the changes in the environment [1]. It is imperative for organizations to have tools and technologies that will help them transform data into valuable information to assist in comprehensive managerial decision making. Traffic departments in Botswana are not exempt to the ever-increasing data. For example, the traffic department under the Botswana police captures data on accidents while the Department of Road Transport and Safety (DRTS) registers new vehicles and issue driver licenses daily. Road accidents are a serious

concern as a lot of lives have been lost to accidents. Furthermore, there has been a growing number of vehicles in the roads in the past years. It is against this background that decision makers must be provided with the right information, in the right format in order to help them make decisions that will assist in injury prevention and in vehicle control and management. This can be made possible by building a BI platform that will integrate data from different data sources and transform these data to quality information. This process will help identify the information needed to support decisions at different organizational levels more especially the strategic decision makers. Even though various departments deal with different aspects of traffic issues, the need for a consolidated data repository cannot be over emphasized. Currently data on traffic and accidents is stored in different and dispersed data sources. This poses a problem because there is no central data store for easy access and evaluation of accidents and vehicle registrations as well as other issues pertaining to traffic data. The capturing and recording of road traffic accidents is done by police department and they are the custodians of that data while vehicle registration is done by DRTS. Therefore, there is a need for a BI platform to integrate data from disparate system sources, which can then be prepared for analytical use. The other limitation is that the data is kept in transactional databases, which are limited in the sense that they store current transactions for ongoing business processes. On the other hand, a BI platform has a data warehouse which has the capability to store large quantities of historical and current data, which enables fast, complex queries across all the data, typically using Online Analytical Processing (OLAP) and Power BI technologies.

In this article, we develop a BI solution for traffic accidents and vehicle registration data analysis in order to avail information required for proper decision making. First, we develop a data warehouse for our BI solution. In developing this data warehouse, we carry out the following steps: we identify the data sources, then identify dimensions, design the DW star schema and we demonstrate the ETL process. This solution model is developed to address the following selected strategic areas:

- **Traffic Accidents Analysis**

This area seeks to investigate the traffic accidents in relation to location, causalities, driver details, vehicle details and road details. This investigation will unearth details pertaining to accident circumstances. The following are sample strategic questions that will be addressed later;

1. How many accidents have been registered across the years as classified by severity and year?
2. Who are involved in these accidents as classified by gender, class and age?
3. Where do most of the accidents happen?
4. When do these accidents happen as classified by time and day of the week?
5. What is the status of the roads where accidents happen?

- **Vehicle Registrations Analysis**

This strategic area provides an investigation into the vehicle registration to understand vehicle registration by model/make, year of registration and registration location. The following are some of the strategic questions which managerial decision makers may be interested in;

1. How many vehicles were registered in the past five years?
2. Where are most vehicles registered?
3. How many vehicles are currently active?
4. Which vehicle model is mostly registered?

The rest of this article is organized as follows: In Section II, we present Related Works, this is followed by a description of data and its sources in Section III. Section IV describes the data warehouse development methodology while section V is about identifying data warehouse dimensions. In section VI, we represent the data warehouse dimensions with a star schema, it is followed by business intelligence platform, the ETL process, data analysis and results and lastly conclusion and future work.

2. RELATED WORKS

To date, several studies have demonstrated that Business Intelligence solutions can be used in critical decision making such as in insurance as well as in traffic accident evaluation. Traffic accidents are a serious worldwide problem that requires robust action from decision makers. According to Hemalatha & Krishnaveni [2] a traffic collision occurs when a road vehicle collides with another vehicle, pedestrian, animal, or geographical or architectural obstacle resulting in injury, property damage, and death. Data warehouse is one of the important BI components that plays a vital role in integrating data sources. The concept of data warehousing arose in mid 1980s with the intention to support huge information analysis and management reporting [3]. Kimball [4] defines a Data Warehouse as a source of data in the enterprise that extracts, cleans, conforms, and delivers source data into a dimensional data store and then supports and implements querying and analysis for the purpose of decision making. Moreover, a Data Warehouse is a subject-oriented, integrated, non-volatile, and time-variant collection of data in support of management's decisions [5]. Data warehousing is critical in data consolidation where there exists dispersed and isolated data systems. For the remainder of this section, we present previous works that describe how other countries have consolidated traffic data from different departments through BI and data warehousing. Lastly, we summarize what has been found through literature in section C.

2.1. Business Intelligence

Companies accumulate and generate data through transactions as they interact with customers on daily basis. BI systems and tools play a vigorous role in bringing together data sources in order to extract the right information from data generated by companies. Without the right tools and technologies, it is difficult to make use of the data as it increases over the years. In modern business, increasing standards, automation and technologies have led to vast amounts of data becoming available [6].

BI is simply a combination of various components and tools that work together collectively to transform operational data into quality information useful for decision making. BI is the process of taking large amount of data, analysing that data and presenting a high-level set of reports of business action to enable management to make fundamental daily business decisions [7]. The objective of BI is to improve the timeliness and quality of information, and enable managers to be able to better understand the position of their firm as in comparison to competitors [8]. The authors further explain that BI explores several technological tools, producing reports and forecasts, in order to improve the efficiency of the decision making. BI components and tools include Data Warehouse (DW), Extract-Transform and Load (ETL), On-Line Analytical Processing (OLAP), Data Mining (DM), Text Mining, Web Mining, Data Visualization, Geographic Information Systems (GIS), and Web Portals. All these tools are combined to create BI platform to facilitate data analytics.

2.2. Application of BI and data warehousing in traffic accident evaluation

The data warehouse is an important component of BI as it consolidates data from various operational sources and make it readily available for data analysis. BI solutions have been used and applied in many areas. A BI solution was proposed in a study in Serbia on traffic accidents analysis and evaluation. In the proposed BI solution, the accidents data is extracted from operational databases to establish the possibility of using those databases as the sources of accident data required for quality analysis of road traffic safety [9]. Quality traffic accident data analysis are enabled in the proposed model by application of the OLAP (Online Analytical Processing) data warehouse concept, which represents the possibility of creating a multidimensional database with a large number of options. The data sources are traffic accident databases by Ministry of Interior (Mol), health authorities, insurance companies, road directorate, statistical office and results of researches. This Serbian model demonstrates how data is extracted from various sources and loaded into the data warehouse by integrating data from multiple sources to create a central storage of data.

Another BI solution was demonstrated in a study conducted in Victory on traffic accident analysis. This BI approach was applied within a SAS Enterprise-based Data Warehouse called Statistical Application for Population Health and Intelligence (SAPHaRI) [10]. The data collections held within SAPHaRI were available for analysis through a secure internet connection that requires password and security token access. This data warehouse includes some indicators that relate to road traffic fatal and non-fatal injuries. From this data warehouse, annual trend graphs, the annual numbers and incidence rate can all be easily accessed. The data sources for this solution were from road safety agencies and hospital's emergency departments [10].

The European Union (EU) also has a system for accident evaluation known as Community database on Accidents on the Roads in Europe (CARE). This system's main objective is to identify and quantify road safety problems, evaluate the efficiency of road safety measures, determine the relevance of community actions and facilitate the exchange of experience in this field. CARE is the European centralized data storage on road accidents which result in death or injury across the EU [11]. This data warehouse provides member states access to this central data store which is hosted by the European Commission at the Luxembourg data center. Through EU data warehouse decision makers are able to identify and quantify of road safety problems, evaluate the efficiency of road safety measures and also determine the relevance of community actions. The CARE DW pulls together non-confidential data from across the EU member states into one central database [11]. Each country is responsible for producing road safety statistics every year, which it then submits in the form of a report to the European Commission. The categories of information used to build statistics on road traffic accidents include: Person Class, Gender, Age group, Vehicle group, Area type, Motorways, Junctions, Collision type, Lighting conditions, Weather conditions and Day of the week.

Furthermore, The Accident Compensation Corporation (ACC) in New Zealand also applied the BI techniques in traffic accidents through their accident injury compensation scheme, which provides injury insurance for all citizens, residents and temporary visitors to New Zealand [13]. The primary source of crash data is the crash analysis system (CAS) database, which contains and summarizes police reported crashes. The database categorizes accident into fatal, injury and non-injury crash types [13]. Information about the nature of the injury, like what, when, where, how, by what, is also captured and stored in the database. The data is extracted from these sources in order to assist decision makers.

2.3. Summary

All the examples above acutely demonstrate that BI tools can be and have been successfully applied in the traffic, accident analysis. The literature indicates that indeed data warehouse plays a pivotal role in BI solutions required for traffic data analysis and evaluation. In Botswana, we are proposing a BI solution with integration of the accident data and vehicle registration data into a central repository for easy access and comprehensive data analysis. The data sources are traffic database system and vehicle registration system. The solution provides an integrated and total view of the enterprise data to facilitate easy decision making. This will enable data analysis through a set of powerful tools for calculations, analysis and report generation. The results of the analysis and evaluation will result in development of measures and accident prevention strategies.

3. DATA SOURCES

Traffic departments in Botswana have accumulated huge amount of data in the past years therefore drowning in data and yet starving in strategic information needed for good decision making. Most of data are kept in transactional information systems for different departments. The data sources for our platform are the files from the traffic accident database from Botswana police and vehicle registration system from DRTS. The traffic accident system has three important tables; Attendant circumstances, Causality details and the Vehicle details. All these tables have different attributes that describe circumstances of the accidents. The vehicle registration file was provided as an excel file with attributes that are captured when a new vehicle is registered. The accident data comprise data from 2012-2020 whereas data on vehicle registration is from 2015-2020.

4. DATA WAREHOUSE DEVELOPMENT METHODOLOGY

In this work, we develop a BI and Data Warehouse solution using the Kimball approach which uses dimensional modelling and bottom-up methodology [4]. Dimensional modelling is a technique for making databases simple to ensure that users can easily understand the data, as well as allow software to navigate and deliver results quickly and efficiently [14]. In building the dimensional model, the following phases are critical; identifying business process requirements, identifying the grain, identifying the dimensions, identifying the facts, verifying the model, physical design considerations and metadata management [15]. Dimensions are developed around identified strategic questions. These dimensions help in the interrogation of different aspects of the business in order to answer the desired strategic questions.

The Data Warehouse was implemented using SQL server platform which supports SQL Server Integration Services (SSIS), SQL Server Analysis Services (SSAS) and SQL Server Reporting Services (SSRS). Power BI was also deployed for interactive data visualization and reports. Microsoft visual studio was used to develop and build the required packages.

5. DATA WAREHOUSE DIMENSIONS

Based the data provided a number of dimensions were identified as follows;

5.1. Accident Dimension

This dimension was created to interrogate the accident data in order to answer the important queries pertaining accident circumstances.

5.2. Causality Dimension

This dimension was created to interrogate the details of those who die because of accidents. Such details include; gender, age, class, injury and other factors pertaining to victims.

5.3. Vehicle Dimension

This dimension is loaded with data about the details of the vehicles involved in the accident. Details such as vehicle make, year and vehicle ownership are captured in this dimension.

5.4. Road Dimension

The road dimension will provide the details of the road when the accident occurred. This dimension has attributes such as road curvature, road slope as well as slippery.

5.5. Driver Dimension

The driver dimension will help answer questions about the driver. These will include details such as driver age and driver gender.

5.6. Vehicle Registration Dimension

This dimension is created to respond to questions regarding all registered vehicles in Botswana. Details such as vehicle model, year of registration and others are captured in this dimension.

5.7. Time Dimension

The time dimension consists of all the dates in the data warehouse. It also consists of the levels of year, quarter, and month. The dimension hierarchies are the paths for drilling down or rolling up in our analysis [17].

6. DATA WAREHOUSE SCHEMA

The dimensional model can be represented with a star schema. Kimball and Ross [14] define a star schema as the generic representation of a dimensional model in a relational database in which a fact table with a composite key is joined to a number of dimensional tables, each with a single primary key.

Figure. shows the star schema for the identified dimensions. The schema also shows the attributes that are essential in answering the strategic questions.

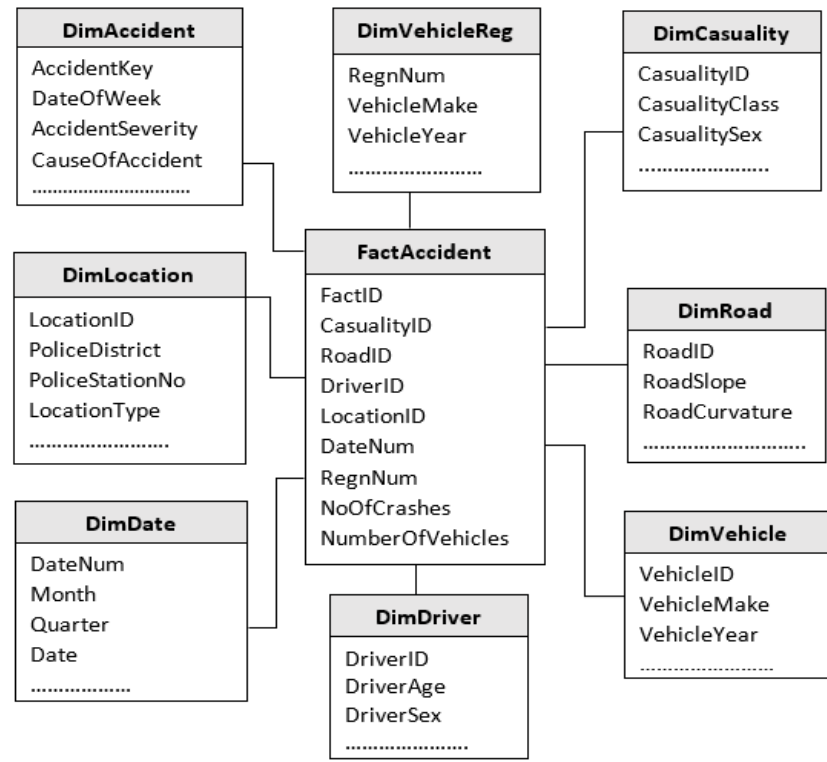


Figure. 1. DW star schema

7. BUSINESS INTELLIGENCE PLATFORM FOR ROAD TRAFFIC

Business intelligence plays a pivotal role in business analytics. Sharda et al. [1] describe BI as an umbrella term that combines architectures, tools, databases, analytics tools, applications and methodologies with the aim of extracting valuable information and knowledge from data. For our proposed solution, data is extracted from the operational systems from Botswana police database and DRTS system. The data goes through the Extraction, Transformation and Loading (ETL) process. ETL is a data integration function that involves extracting data from operational systems sources, transforming it to fit business needs, and ultimately loading it into a Data Warehouse [17]. Relevant data is extracted to address the strategic questions, followed by the transformation which involves the cleaning and correcting of errors and inconsistencies to ensure that only quality data is loaded in the Data Warehouse. Data loading refers to loading the data into the end target [18]. A comprehensive description of this ETL process is provided in Section VIII. Figure.2 represents the architectural blueprint of the proposed solution.

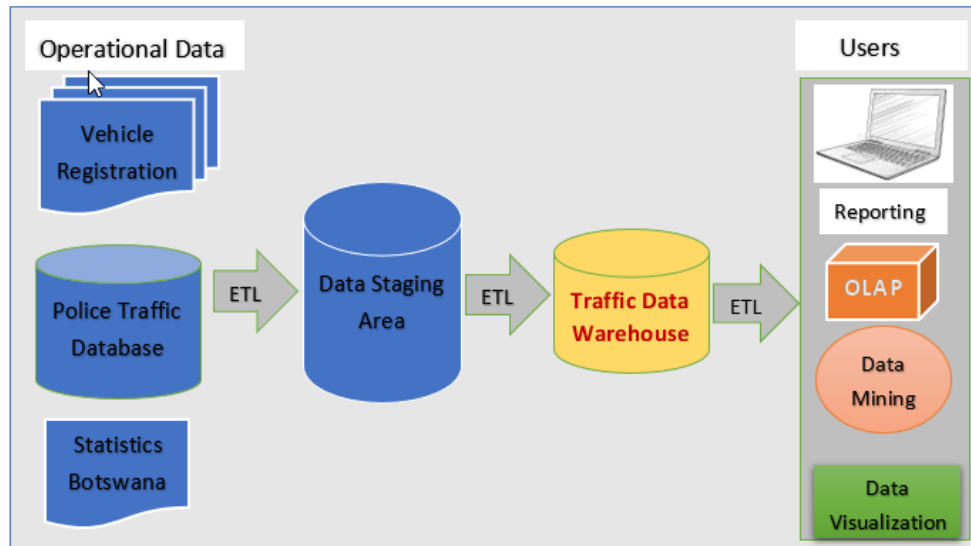


Figure. 2. Business Intelligence platform

8. THE ETL PROCESS

In this section, we demonstrate the implementation of the ETL process for the traffic data warehouse we are proposing. ETL is the process of extracting data from source systems and carrying it into the data warehouse [19]. This involves extraction, transformation and loading of relevant data into the target dimensions to provide strategic information. The ETL process was done through the SQL Server platform which supports data integration, analysis and reporting. SQL Server Integration Services (SSIS) contain a data-flow engine to transfer and transform data to and from various data sources through its operations such as Aggregate, Sort, Lookup, Merge, Merge Join, Union All, Data Conversion and Audit. It also has graphical tools and wizards for creating an extraction, transform, and loading. SQL Server Analysis Services (SSAS) enable users to accommodate multiple analytic needs within one solution and also creation of the cubes. SQL Server Reporting Services (SSRS) provide a platform that supports the authoring, management and delivery of interactive reports to the entire organization. Microsoft visual studio was used to develop and build the integration packages. Power BI was then used to generate reports to address strategic question for traffic accident managers. After the development and loading, the data was used to generate reports for data analysis. Data was extracted from different source files and stored in a central location through the use of SSIS for the purpose of answering strategic questions.

8.1. Implementation of ETL Process

The ETL was implemented using SQL Server Integration Services and the Visual studio which develops the packages required to build the solution. The data from the traffic accident database had three tables. Dimensions were identified around these three files. The Accident details files had data about the accident and road details. This file was targeting the accident dimension and the road dimension. The causality details table captured data on the gender, age and class of the accident victims. This file was used to populate the causality dimension. The vehicle data file had data about vehicles and driver details. This file was targeted at the vehicle and driver dimension. Vehicle Registration file contained details on the registration of new vehicles. The data were extracted from this file and loaded in the vehicle registration dimension.

8.2. Data Staging Area

The staging area which can also be referred to as landing area sit between the data sources and the data warehouse. It is the immediate data storage before data can be loaded in the data warehouse. The ETL process is used to extract data sets from the various sources, bring them to a data staging area, apply a sequence of processes to prepare the data for migration into the data warehouse, and actually load them [20]. Lans [20] further asserts that the data that is loaded in the staging areas undergo a lot of processing before it's in a form suitable for storage in a data warehouse. Incorrect values have to be transformed, missing data values have to been replaced and so on.

8.3. Data Transformation

During the extraction process data types were converted through the transformation process. Data types were changed to the ones that conforms to SSIS. The varchar data types are converted to String [DT_STR] and integer values are converted to four-byte signed integer [DT_I4]. The conversion is shown in figure 3.

| Output Alias | Data Type | Length |
|-------------------------|----------------------------------|--------|
| Copy of Accident Key | four-byte signed integer [DT_I4] | |
| Copy of NUMBER OF ... | string [DT_STR] | 250 |
| Copy of POLICE DISTR... | string [DT_STR] | 250 |
| Copy of POLICE STATI... | string [DT_STR] | 250 |
| Copy of DATE | date [DT_DATE] | |
| Copy of TIME | database time [DT_DBTIME] | |
| Copy of DAY OF WEEK | string [DT_STR] | 250 |
| Copy of ACCIDENT SE... | string [DT_STR] | 250 |
| Copy of TRAFFIC | string [DT_STR] | 250 |

Figure. 3. Data type conversion

For our solution data was extracted from the source files into the staging area as demonstrated in figure 4.

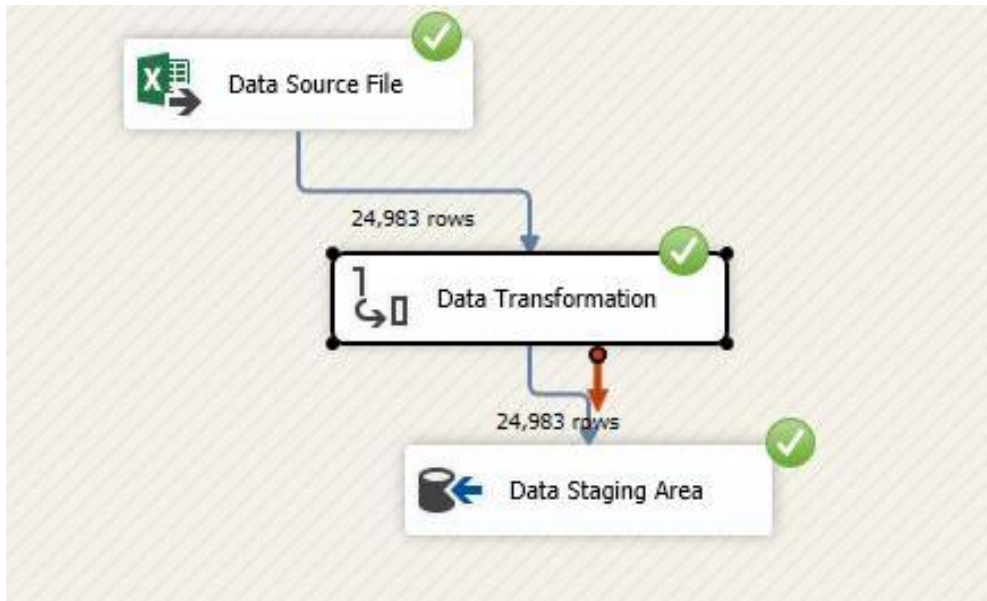


Figure. 4. ETL Process

When the data in the staging area has been transformed it can be copied to the dimensions in a warehouse. Data are copied from the staging area into the identified dimensions. Figure 5 demonstrates the loading of the Road dimension and the Location dimension.

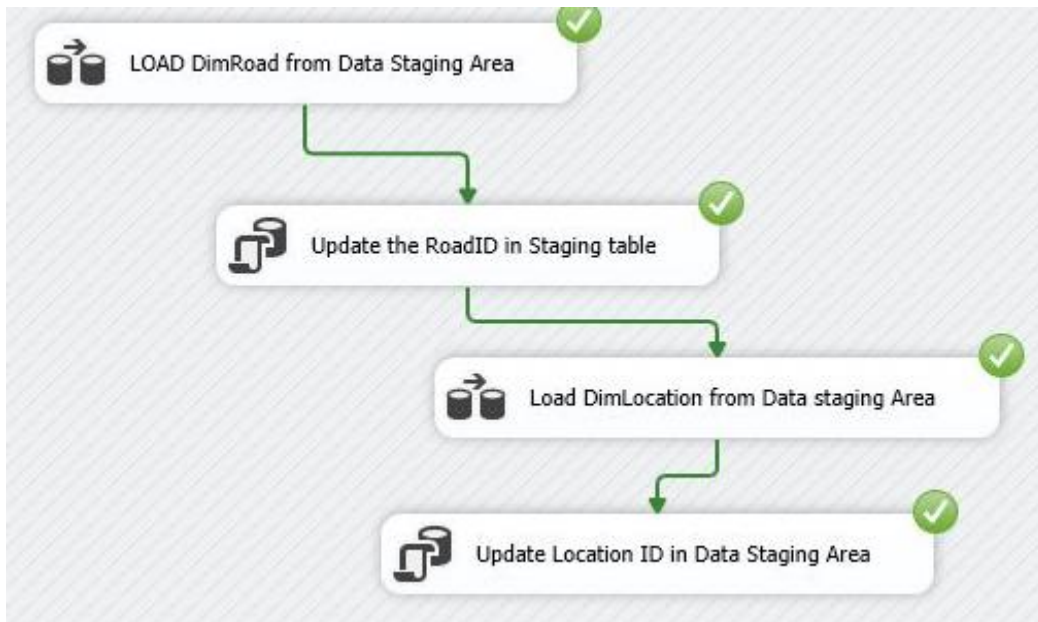


Figure. 5. Data Loading

After the data from these files are loaded into the data warehouse, decisions makers should be able to query the data warehouse in order to answer all questions pertaining to accidents and vehicle registration. For this research, the dimensions should provide answers to the aforementioned research questions as will be demonstrated in results section. Figure 6 shows the data in the road dimension after successful loading.

| | RoadID | AccidentKey | RoadCurvature | RoadSlope | JunctionType |
|----|--------|-------------|---------------|-----------|--------------|
| 1 | 1 | 53903 | 1 | 1 | 1 |
| 2 | 2 | 57052 | 1 | 1 | 1 |
| 3 | 3 | 57053 | 1 | 1 | 2 |
| 4 | 4 | 57054 | 1 | 1 | 1 |
| 5 | 5 | 57055 | 1 | 1 | 1 |
| 6 | 6 | 57056 | 1 | 1 | 1 |
| 7 | 7 | 57057 | 1 | 1 | 1 |
| 8 | 8 | 57058 | 1 | 1 | 1 |
| 9 | 9 | 57059 | 1 | 1 | 1 |
| 10 | 10 | 57060 | 1 | 1 | 3 |

Figure. 6. Road data dimension

9. DATA ANALYSIS AND RESULTS

In this Section, we deploy power BI and SSRS to build reports based on the strategic questions across the dimensions first introduced in Section I. The results demonstrate how information can become available for strategic decisions by addressing just a few questions from each dimension.

9.1. Traffic Accidents Results

9.1.1. Results from strategic questions

The results in this section are based on queries from different dimensions. A few sample questions that a manager may need answers to are addressed. One of the powerful aspects of a data warehouse is that users can change how they ask their questions from time to time. The following are results based on the stated strategic questions.

- **Strategic question 1:** How many accidents have been registered across the years as classified by severity and year?

The distribution of accidents by type reveals four types of accidents being fatal, serious injuries, minor injuries and damage only. Across accident severity, damage Only accidents had the most interesting recent trend with a 11.93% decrease in the last 3 years on record. The number for damage only dropped from 13,491 to 11,881 during its steepest decline between 2018 and 2020. Minor accidents experienced the longest period of growth in between 2012 and 2020, and serious accidents experienced the longest decline (-283) during the same period as demonstrated in table 1.

Table 1. distribution of accidents by type

| Year | Damage Only | Fatal | Minor | Serious | Total |
|--------------|---------------|-------------|--------------|-------------|---------------|
| 2012 | 13931 | 328 | 2467 | 795 | 17521 |
| 2013 | 13473 | 321 | 2486 | 778 | 17058 |
| 2014 | 13077 | 288 | 2536 | 732 | 16633 |
| 2015 | 13950 | 329 | 2607 | 768 | 17654 |
| 2016 | 14425 | 348 | 2861 | 739 | 18373 |
| 2017 | 13984 | 366 | 2726 | 710 | 17786 |
| 2018 | 13491 | 379 | 2835 | 636 | 17341 |
| 2019 | 14652 | 361 | 2909 | 701 | 18623 |
| 2020 | 11881 | 265 | 2417 | 512 | 15075 |
| Total | 122864 | 2985 | 23844 | 6371 | 156064 |

- **Strategic question 2:** How many accidents are recorded in each police district as classified by severity and location?

The distribution of accidents by location indicates that many accidents occur in the Broadhurst area where the district accounted for highest number of 27778 followed by G/West and Serowe. The least number of accidents was recorded in Tsabong which recorded 659 accidents. This information can be used to inform managers to deploy more resources there, such as road blocks, patrols and speed traps. Table 2 shows accidents by police district.

Table 2. Accidents by police district

| Police District | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | Total |
|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| Broadhurst | 4715 | 5000 | 4685 | 4612 | 4861 | 3905 | 27778 |
| Francistown | 792 | 827 | 756 | 657 | 685 | 564 | 4281 |
| G/West | 4746 | 4759 | 4627 | 4618 | 4941 | 4025 | 27716 |
| Gantsi | 217 | 263 | 235 | 223 | 207 | 168 | 1313 |
| Kasane | 119 | 195 | 237 | 217 | 200 | 172 | 1140 |
| Kutlwano | 1011 | 1007 | 926 | 879 | 1015 | 763 | 5601 |
| Letlhakane | 236 | 232 | 237 | 196 | 274 | 191 | 1366 |
| Lobatse | 268 | 263 | 218 | 184 | 260 | 215 | 1408 |
| Mahalapye | 598 | 735 | 771 | 796 | 823 | 560 | 4283 |
| Maun | 663 | 679 | 659 | 585 | 709 | 645 | 3940 |
| Mochudi | 435 | 483 | 485 | 543 | 624 | 425 | 2995 |
| Molepolole | 612 | 600 | 680 | 544 | 758 | 611 | 3805 |
| Sejelo | 609 | 838 | 804 | 764 | 867 | 669 | 4551 |
| Selibe Phikwe | 460 | 359 | 301 | 249 | 283 | 225 | 1877 |
| Serowe | 988 | 1026 | 1005 | 1088 | 1006 | 735 | 5848 |
| Tsabong | 123 | 91 | 123 | 100 | 123 | 99 | 659 |
| Grand Total | 16592 | 17357 | 16749 | 16255 | 17636 | 13972 | 98561 |

- **Strategic Question 3:** When do these accidents happen as classified by time and day of the week?

Most accidents were recorded over the weekend. At 28,439, Sunday had the highest number of recorded accidents and was 47.59% higher than Wednesday, which had the lowest count of accidents at 19,269. Sunday accounted for 18.22% of the accidents. Across all 7 Days of the week, count of number of accidents ranged from 19,269 to 28,439. Accident prevention measures must be intensified over the weekend to curb the ever-increasing number of accidents. Figure.7 shows the distribution of these accidents by day of the week from list to highest.

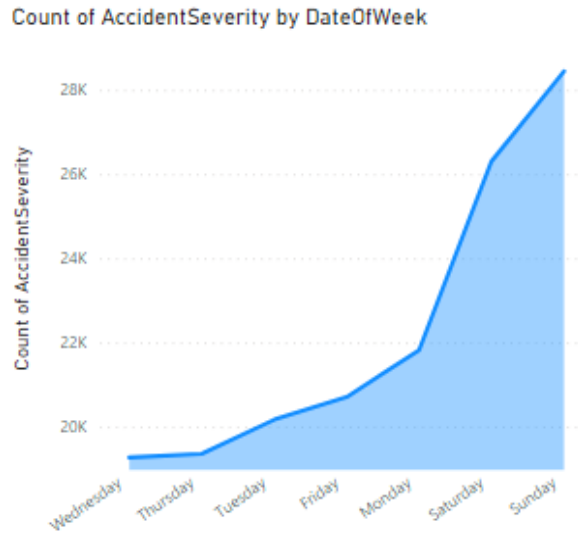


Figure.7. Accident by day of the week

- **Strategic question 4: Who are involved in these accidents as classified by age and accident severity?**

The distribution of casualties by age reveals that the 20-40 age group recorded the highest number at 33002 whilst there are less accidents for age group 81-100. Table 1 shows the distribution of accident by severity and age.

Table 3. accident by severity and age

| Age | Damage Only | Minor | Serious | Grand Total |
|--------------|----------------|--------------|--------------|--------------|
| 1-20 | 151329 | 6383 | 2273 | 10000 |
| 21-40 | 685058 | 20209 | 7667 | 33002 |
| 41-60 | 311685 | 6057 | 2485 | 10258 |
| 61-80 | 4421 | 1031 | 428 | 1884 |
| 81-100 | 50 | 85 | 38 | 173 |
| Total | 1188543 | 33765 | 12891 | 55317 |

9.1.2. Data Visualization through interactive Dashboards

A dashboard is a data visualisation tool that facilitates information delivery to the business and data warehouse users. Dashboards provide visual displays of important information that is

consolidated and arranged on a single screen so that information can be digested at single glance and easily drilled in and further explored [2]. Power BI allows for a creation of interactive dashboards where user can see many other factors or details just by way of hovering the mouse over the required fields. Dashboards can include graphs from different dimensions of the data warehouse as well as maps to show locations. Figure 8 shows a dashboard for our traffic data warehouse solution.

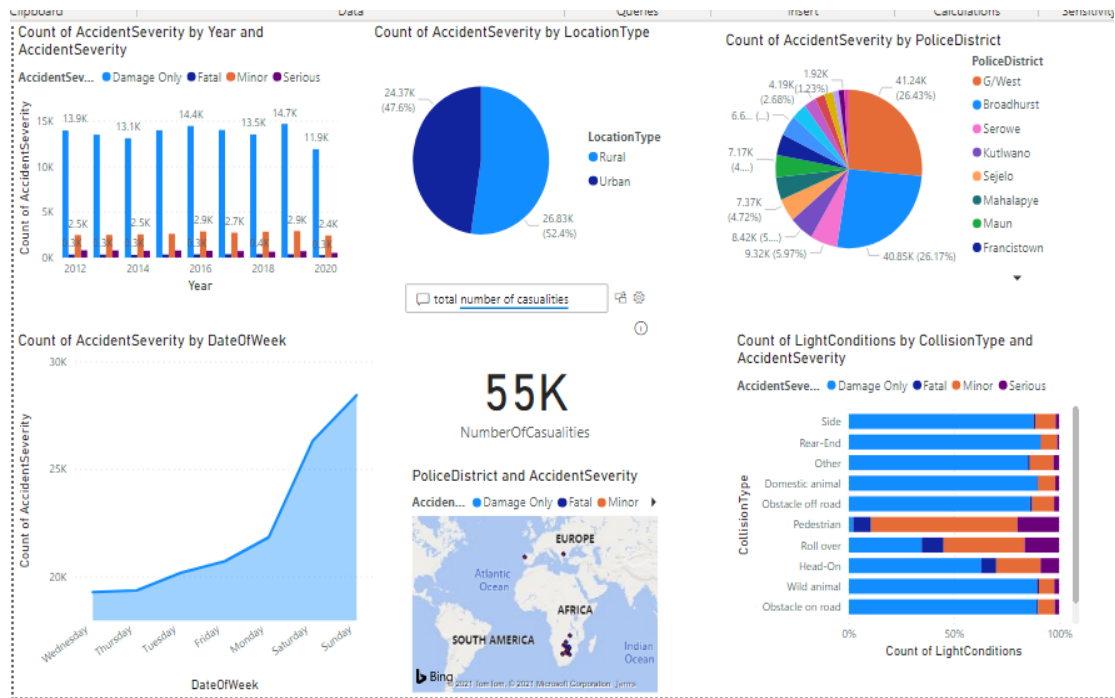


Figure. 8. Data visualization dashboard

9.2. Vehicle Registration Results

The vehicle registration dimension was used to answer sample questions pertaining to vehicle registration. Queries were based on the strategic questions below;

- **Strategic question 1:** How many vehicles were registered in the past five years?

The number of vehicle registration decreased for the last 3 years on record. Number of vehicles dropped from 65535 to 24457 during its steepest decline between 2019 and 2021. Vehicle registration experienced the longest period of growth (+9243) between 2015 and 2018. At 65,535, 2019 had the highest number of registered vehicles and was 57.00% higher than 2020, which had the lowest number at 41,741. The year 2019 accounted for 21% of the registered vehicles. Across all 6 Year, the numbers of registration ranged from 41,741 to 65,535.

- **Strategic question 2:** Where are most vehicles registered?

At 75,438, ROAD TRANSPORT & SAFETY (HQ) had the highest count of registered cars. Decision makers can deploy more resources at this center to prevent slow service delivery. Sowa registered the lowest number of 70 vehicles across a period of 2015 to 2021. ROAD TRANSPORT & SAFETY (HQ) accounted for 22.40% of registered vehicles. Across all 29 transport offices, the number of registered vehicles ranged from 70 to 75,438.

- **Strategic question 3:** How many vehicles are currently active?

An evaluation of the vehicle status indicates that of all the registered vehicles, Run (327,026) was higher than EXPORTED (9,666). Run refers to active vehicles. RUN accounted for 97.13%. The results also reveal that 142 vehicles have been scrapped. Figure 9 shows the vehicle status.

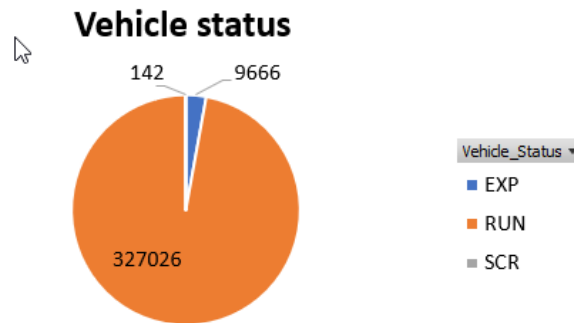


Figure. 9. Vehicle Status

- **Strategic question 4:** Which is the mostly registered vehicle model?

Of all the vehicle registered, a total count was highest for TOYOTA at 136,988, followed by HONDA, and MAZDA. FIT made up 9.14% as the least. We can safely conclude that Toyota is the mostly registered car model in Botswana.

10. CONCLUSION AND FUTURE WORK

The development of the BI platform for traffic accidents has proven to have an immense impact on information delivery for accident data analysis and vehicle registration as demonstrated by our results. Quality decision making is achieved when decision makers have the information they need at the right time and in the right quality. Through the data visualisation tools demonstrated above, the injury prevention departments can view the statistics on different categories and can query the DW to get answers to any questions they have. Easy accessibility to accident data will help decision makers to formulate accident prevention strategies and measures for traffic safety improvement.

The vehicle registration data mart also has shown the possibility of providing answers to questions pertaining to new vehicles registration. This information will inform managerial decision makers about the number of vehicles that are active in our roads and other finer details about vehicles.

In conclusion BI is a viable computerized support tool for decision making. The framework developed in this research is a solution for accidents and vehicle registration data evaluation that other organization in the same domain must consider. Data integration is critical in ensuring easy access to strategic information required to make decisions at strategic level.

In future other analysis techniques can be employed such as OLAP, drilling and data mining to further unearth interesting revelations from the data.

ACKNOWLEDGMENTS

Our sincere gratitude goes to the Botswana police and DRTS for providing us with relevant data to model our solution.

REFERENCES

- [1] R. Sharda, D. Delen and E. Turban, *Business Intelligence and Data Analytics: Systems for Decision Making*, Pearson, 2014.
- [2] M. Hemalatha and S. Krishnaveni, "A Perspective Analysis of Traffic Accident using Data Mining Techniques," *International Journal of Computer Applications*, 2011.
- [3] O. E. Sheta and A. N. Eldeen, "The technology of using a data warehouse to support decision making in Health care," *International journal of database management systems (IJDMMS)*, pp. 75-86, 2013.
- [4] R. Kimball, *Data warehouse Life cycle toolkit*, Wiley, 2013.
- [5] H. W. Inmon, *Building the Data Warehouse*, New York: John Wiley & Sons, Inc, 2002.
- [6] J. Ranjan, "Business Intelligence: Concepts, Components, Techniques and Benefits," *Journal of theoretical and Applied Information Technology*, 2009.
- [7] R. Stackowiak, J. Rayman and R. Greenwald, *Oracle Data Warehousing and Business Intelligence Solutions*, Wiley Publishing, 2006.
- [8] A. R. Khan and. K. S. Quadri, "Business Intelligence: An Integrated Approach," *Business Intelligence Journal*, 2012.
- [9] O. Pantelic, D. Pesic, M. Vujanic and D. B. Vujaklija, "Towards an analytical information system of traffic accidents in the function of traffic safety monitoring," *Scientific Research and Essays*, pp. 398-409, 2012.
- [10] R. Mitchell, M. Babanch, A. Williamson and R. Grzebieta, "Transport And Road Safety (TARS) Research submission to the Victoria serious injury Inquiry of the road safety committee," University of New South Wale, victoria, 2013.
- [11] IDABC, "Idabc," 2004. [Online]. Available: <http://ec.europa.eu/idabc/en/document/2281/5926.html>. [Accessed 5 April 2020].
- [12] European Commission, 2017. [Online]. Available: https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/care_flowchart_a0.jpg. [Accessed 26 January 2018].
- [13] Statistics New Zealand, "Injury Statistics Project Pilot: Output Report," Statistics New Zealand, Wellington, 2004.
- [14] R. Kimball and M. Ross, *The Data Warehouse toolkit: The Complete Guide to Dimensional Modelling*, Second ed., New York: John Wiley and Sons, Inc., 2002.
- [15] C. Ballard, D. M. Farrell, A. Gupta, C. Mazuela and S. Vohnik, *Dimensional Modelling: In a Business Intelligence Enviroment*, First ed., IBM Corp, 2006.
- [16] P. Ponniah, *Data warehouse fundamentals*, New york: John Wiley and Sons, INC, 2001.
- [17] V. Gour, S. Sarangdevot, G. S. Tanwar and A. Sharma, "Improve Performance of Extract, Transform and Load (ETL) in Data Warehouse," *International Journal on Computer Science and Engineering*, pp. 786-789, 2010.
- [18] F.S. Esmail Ali, "A Survey of Real-Time Data Warehouse and ETL," *International Scientific Journal of Management Information Systems*, vol. 9, no. 3, 2014.
- [19] P. Dhadha and N. Sharma, "Extract Transform Load Data with ETL Tools," *International Journal of Advanced Research in Computer Science*, vol. 7, pp. 152-160, 2016.
- [20] R. v. d. Lans, *Data Virtualization for Business Intelligence Systems*, Morgan Kaufmann, 2012.

AUTHORS

Mr Monkgogi Mudongo is a lecturer in the department of Computer Science in the University of Botswana. He obtained his MSc in Computer Information Systems from the university Of Botswana. His research interests are in the areas of Data Warehousing, Knowledge Management, Business Information Systems, Big data analytics, and Information Retrieval.



Dr Edwin Thuma has a broad background in Computing Science with specific expertise in Information Retrieval (the science of search engines) and Big Data Systems. In particular, his research has been focused primarily on the development of search engines tailored to support health professionals and laypeople when searching for health content on the web. Recently he has started working on search engines that are tailored to support legal professionals when searching for precedent cases or statutes that support the current case.



Mr Nkwebi P. Motlogelwa works as a Lecturer in the Department of Computer Science, University of Botswana. He is currently actively involved in two research areas: Information retrieval in the medical domain research, and Natural Language Processing specifically focusing on the Setswana Language. In the past he was engaged in a Microsoft funded research that explored how wireless and mobile technologies could improve public health in under-served communities



Mrs Tebo K. Leburu-Dingalo is a Lecturer in the Department of Computer Science, University of Botswana. Her areas of interest are Health Informatics, Knowledge Based Systems and High-Performance Computing. She is also interested in the use of Information and Communications Technology towards social and economic development. Tebo holds a Bachelor of Science (BS) in Computer Engineering from Florida Institute of Technology, and Master of Science in Information Systems from the University of Botswana.



Ms Pulafela A. Majoo is a Graduate Teaching Assistant in the Disability Support Services Unit in the University of Botswana. She is currently awaiting to be awarded with Msc Computer Information Systems after submission of her dissertation from University of Botswana. Her areas of interest are in information retrieval in the medical domain, data warehousing, software development frameworks and data analytics.

