

Group 8 BI Project- Second Draft

Name & ID - Muskan Muskan (C70303902)

Name & ID - Vatsal Rameshbhai Gohel (C70303897)

Name & ID - Sakina Hassani (C70306262)

Course - Business Intelligence (MSIT3820-02-F23)

➤ **What are some of the problems that the company has?**

The company is currently grappling with a surge in nationwide road accident complaints, which are not only causing chaos but also resulting in tragic loss of human lives. In response to this pressing issue, the company is diligently working on reducing these accidents. To achieve this, they have taken steps to collect data from MassDOT for the National Road Safety Intelligence Bureau, which will be used to gain insights into the recurring patterns of these road accidents.

One of the key challenges we face is the occasional difficulty in obtaining data from all regions, leading to a reduction in the accuracy of our predictive models. This lack of accuracy poses a significant hurdle in our mission to prevent road accidents and increases the burden on the National Road Safety Intelligence Bureau. Additionally, connectivity issues with transportation authorities, stemming from inadequate infrastructure, hinder our ability to collaborate effectively with medical and law enforcement agencies.

The task of influencing legislative changes to enhance road safety can be particularly daunting, given that these matters involve both state and federal laws. Implementing changes in these domains requires extensive processing and time, making it a complex and time-consuming endeavor.

In light of these challenges, our focus has shifted toward Business Intelligence (BI) as a means to enhance road safety. We aim to integrate the data we have collected into a mobile application. This app will utilize the longitude and latitude coordinates of accident locations to promptly dispatch paramedic services and towing assistance. The goal is to ensure the swift and efficient management of accident scenes without causing extended road closures.

It's important to note that the development of this app will take some time, as it entails a meticulous analysis of the collected data to identify patterns. Subsequently, the app will be built based on this

information. The ultimate aim is to provide valuable support to various departments and agencies involved in road safety, addressing this urgent issue effectively.

➤ **What are the company's overall goals?**

The company is dedicated to a primary mission: raising awareness about road safety with the ultimate goal of reducing the number of accidents. We aim to achieve this by harnessing the wealth of data derived from past accidents. Through comprehensive data analysis, we intend to identify recurring trends and patterns, subsequently using these insights to disseminate critical information about road safety.

In addition to our core objective, our company is committed to addressing road safety concerns in collaboration with transportation authorities, law enforcement agencies, and other key stakeholders. Our collective aim is to enhance safety measures and mitigate the tragic consequences of road accidents.

Looking ahead, our long-term vision is to leverage Business Intelligence (BI) as a powerful tool for advancing road safety. We plan to integrate our data into a user-friendly mobile application, capable of utilizing the precise longitude and latitude coordinates of accident locations. This app will serve as a crucial platform for dispatching paramedic services promptly and coordinating efficient towing operations. By doing so, we aim to minimize road closures and ensure the smooth functioning of affected areas, ultimately contributing to improved road safety.

➤ **How could Business Intelligence help the company to solve the problems?**

Our vision hinges entirely on the strategic utilization of business intelligence and data analysis. By meticulously gathering data on road accidents and applying advanced analytical techniques, we empower stakeholders to unearth invaluable insights. Through this process, they can then implement precise, targeted interventions aimed at enhancing road safety and mitigating the risks of accidents.

We firmly believe that harnessing business intelligence to scrutinize the data will unveil critical information about accident-prone scenarios. These insights will encompass factors such as road conditions, weather conditions, vehicle type, speed limits, lighting conditions, junction details, and

more. This approach enables us to proactively address and rectify the conditions and circumstances that contribute to accidents, ultimately making our roadways safer for all.

➤ **What software does the company utilize? (e.g. Excel, SQL, Python, Tableau, etc.)**

At the National Road Safety Intelligence Bureau, our approach to data management and analysis is a well-structured process. We harness a combination of powerful tools to effectively address our mission of enhancing road safety.

We initiate our data collection process using Excel, where we gather historical information related to road accidents and safety incidents. To maintain this valuable data, we employ SQL databases, ensuring that we retain only the information essential for our specific use case, which is road safety data.

Data analysis is a critical step in our workflow, and we turn to Python for this purpose. Python allows us to conduct both descriptive and predictive analyses, providing us with actionable insights into road safety trends and potential areas of concern.

To present this data in a comprehensible and visually engaging manner, we utilize Tableau. This tool empowers us to transform our tabular data into insightful visualizations and interactive dashboards. These visuals are designed with the aim of making complex information easily understandable for our business stakeholders and partners in various agencies, thereby fostering effective communication and collaboration in pursuit of our organization's shared goals.

In the future, we are poised to expand our capabilities by incorporating APIs and application development. These elements will play a pivotal role in the realization of our earlier mentioned app, which is a key component of our ongoing efforts to improve road safety.

➤ **What are the logical components of the star schema? Or, if they do not have a Business Intelligence system in place, what would the components be in order to analyze data that is useful in solving the problems?**

At present, we do not have a Business Intelligence system in place, but we have organized our data based on the dataset we have. As a result of this structured schema, we can efficiently analyze and

interpret insights from the data to address our road safety concerns. Here's how we can structure the logical components:

1. Fact Table:

- Fact table will contain quantitative data that we want to analyze. In this case, it could be a table named "Accident_Fact" with measures such as "Number_of_Casualties" and "Number_of_Vehicles."

2. Dimension Table:

- Create separate dimension tables for descriptive attributes:
- **Date Dimension:** This table can include attributes related to the accident date, such as "Accident_Date", "Day_of_Week" and "Time."
- **Location Dimension:** Store details about the location, including "Latitude", "Longitude", "Local_Authority_(District)" and "Urban_or_Rural_Area".
- **Weather Dimension:** Include attributes like "Weather_Conditions" and "Light_Conditions" related to the weather and lighting conditions during accidents.
- **Road Dimension:** This table can contain attributes such as "Road_Surface_Conditions", "Road_Type", and "Speed_limit."
- **Junction Dimension:** Include attributes like "Junction_Control" and "Junction_Detail" for information about road junctions.
- **Vehicle Dimension:** Store information about the involved vehicles, such as "Vehicle_Type."

3. Hierarchies:

- We can define hierarchies within dimension tables if needed. For example, in the Date Dimension, we create hierarchies like "Year > Month > Day."

4. Relationship:

- Establish relationships between the fact table and dimension tables using common keys. For instance, we link the "Date Dimension" to the "Accident_Fact" using the "Accident_Date" field.

5. Data Transformation and Cleaning:

- Ensure data quality by cleaning and preprocessing the data. This may involve handling missing values, correcting errors, and standardizing formats.

6. Data Analysis:

- Use data analysis tools like Excel, SQL, or Python to perform data analysis. We can calculate accident rates, identify trends, and derive insights.

7. Visualizations:

- Create charts and graphs to visualize our findings. Tools like Excel, Tableau, Python visualization libraries like matplotlib and seaborn can help us create visual reports.

8. Documentation:

- Document our data analysis process, findings, and insights for future reference and collaboration.

9. Communication:

- We can Share our findings and insights with relevant stakeholders, such as transportation authorities or law enforcement agencies, for collaborative efforts to improve road safety.

Through this simplified star schema structure, we can organize and analyze our data more effectively without using a full-fledged business intelligence system. A more advanced BI system may be used for better automation and scalability as the project progresses.

➤ What are the sources of data?

The source of our data is Kaggle.com, where we conducted in-depth research with the aim of contributing to the reduction of road accident casualties. Delving into various datasets available, we identified one comprehensive dataset that provided abundant information for initiating our research. This dataset equips us with valuable insights that can significantly benefit stakeholders such as the Ministry of Transport, Road Transport Department, Police Force, Emergency Services Department, Road Safety Corps, Transport Operators, Traffic Management Agencies, as well as the general public and media.

Notably, the dataset includes critical geospatial data like longitude and latitude, enabling us to precisely pinpoint the locations of accidents. This geographical information holds immense potential for enhancing our ability to strategize and implement targeted measures aimed at reducing the overall number of casualties. Our research is poised to provide actionable insights that can contribute meaningfully to the efforts of various stakeholders in the realm of road safety.

➤ **What is the frequency of the data?**

The term "frequency" typically refers to how often a particular value occurs in a dataset, and it is commonly used in the context of categorical variables. The categorical variable present in our dataset are as follows:

- Day_of_Week
- Junction_Control
- Junction_Detail
- Accident_Severity
- Light_Conditions
- Local_Authority_(District)
- Carriageway_Hazards
- Police_Force
- Road_Surface_Conditions
- Road_Type
- Time
- Urban_or_Rural_Area
- Weather_Conditions
- Vehicle_Type

We did the pivot analysis on these columns to find the frequency of each categorical variable to find any useful insight. We found few data cleaning needs like correcting spelling of same categorized so they are categorized as one like fetal and fatal in Accident severity.

Junction Control	Count of Junction_Control
Authorised person	460
Auto traffic sigl	93
Auto traffic signal	32256
Data missing or out of range	98056
Give way or uncontrolled	150045
Not at junction or within 20 metres	25378
Stop sign	1685

Week Day	Count of Day_of_Week
Sunday	33544
Monday	43918
Tuesday	46386
Wednesday	46381
Thursday	45649
Friday	50529
Saturday	41566

Junction Detail	Count of Junction_Detail
Crossroads	29948
Mini-roundabout	3346
More than 4 arms (not roundabout)	4148
Not at junction or within 20 metres	123094
Other junction	8315
Private drive or entrance	10875
Roundabout	27264
Slip road	4265
T or staggered junction	96718

Accident Severity	Count of Accident_Severity
Fatal	3904
Fetal	49
Serious	40740
Slight	263280

Light Conditions	Count of Light_Conditions
Darkness - lighting unknown	2924
Darkness - lights lit	60093

Darkness - lights unlit	1142
Darkness - no lighting	16528
Daylight	227286

Carriageway Hazards	Count of Carriageway_Hazards
Any animal in carriageway (except ridden horse)	1620
None	302546
Other object on road	2243
Pedestrian in carriageway - not injured	715
Previous accident	511
Vehicle load on road	335

Top 10 Police Forces	Count of Police_Force
Devon and Cornwall	8804
Greater Manchester	11954
Hampshire	9492
Kent	9995
Lancashire	9143
Metropolitan Police	46789
Sussex	8548
Thames Valley	11483
West Midlands	13509
West Yorkshire	12016

Top 10 Local Authority District	Count of Local_Authority_(District)
Barnet	2302
Birmingham	6165
Bradford	3006
Bristol, City of	2270
Cornwall	2606

Leeds	4140
Liverpool	2611
Manchester	3132
Sheffield	2750
Westminster	2811

Road Surface Conditions	Count of Road_Surface_Conditions
Dry	208967
Flood over 3cm. deep	374
Frost or ice	12078
Snow	4758
Wet or damp	81479

Road Type	Count of Road_Type
Dual carriageway	45467
One way street	6197
Roundabout	20929
Single carriageway	230612
Slip road	3234

Top 10 hours of accident	Count of Time
8 AM	22552
11 AM	16141
12 PM	18342
1 PM	18971
2 PM	19067
3 PM	24151
4 PM	24903

5 PM	26964
6 PM	21063
7 PM	15851

Weather Conditions	Count of Weather_Conditions
Fine + high winds	3148
Fine no high winds	244496
Fog or mist	1690
Other	8802
Raining + high winds	3526
Raining no high winds	34877
Snowing + high winds	538
Snowing no high winds	4839

Area Urban/Rural	Count of Urban_or_Rural_Area
Rural	109441
Urban	198532

Vehicle Type	Count of Vehicle_Type
Agricultural vehicle	749
Bus or coach (17 or more pass seats)	8686
Car	239794
Goods 7.5 tonnes mgw and over	6532
Goods over 3.5t. and under 7.5t	2502
Minibus (8 - 16 passenger seats)	821
Motorcycle 125cc and under	6852
Motorcycle 50cc and under	3703
Motorcycle over 125cc and up to 500cc	3285
Motorcycle over 500cc	11226
Other vehicle	2516

Pedal cycle	66
Ridden horse	3
Taxi/Private hire car	5543
Van / Goods 3.5 tonnes mgw or under	15695

Speed Limit	Count of Speed_limit
10	3
15	2
20	2899
30	200040
40	25650
50	10191
60	46826
70	22362

- **What else can you do to help the company achieve their goals? (E.g. Increase sales? Lower costs? Increase organizational efficiency? Provide better decision-making capacity?)**

A comprehensive data analysis is a key component of the company's primary mission of promoting road safety and reducing accidents. We work with transportation authorities, law enforcement agencies, and key stakeholders to improve safety measures and minimize the consequences of road accidents. To advance road safety, we intend to leverage Business Intelligence (BI) in the long run. Our plan is to integrate our data into a user-friendly mobile application, enabling paramedics and towing services to be dispatched efficiently according to precise accident location coordinates. By taking this strategic approach, road closures will be minimized and affected areas will function smoothly, thereby improving road safety.

In order to further support the company's goals, we propose several initiatives. Predictive analytics would allow us to identify hotspots and high-risk areas based on historical data, allowing us to allocate resources and take preventative measures in advance. In addition, public awareness campaigns, including social media, are conducted to educate the public about safe driving practices, traffic rules adherence, and distracted driving dangers. Thirdly, collaboration with smart

city initiatives in an effort to incorporate road safety data into broader urban planning efforts, which will result in safer and more efficient transportation.

A real-time incident reporting and response system could also be implemented using mobile apps. This would allow users to report accidents or road hazards immediately, so emergency services can respond more quickly. Furthermore, the developed mobile application should include features facilitating quick and accurate emergency responses, including integration with emergency services, the provision of real-time accident data, and effective communication among all parties involved.

The implementation of successful strategies from similar contexts would significantly improve road safety measures, as would benchmarking studies conducted to identify best practices. Additionally, it is crucial to establish a continuous data analysis process and feedback loop for road safety. The effectiveness of road safety initiatives will be maintained if new data are regularly analyzed, models are updated, and stakeholders are incorporated. Taking these initiatives will not only meet the company's mission, but will also enhance organizational efficiency, decision-making capacity, and overall effectiveness.

➤ **Who is/are the consumer(s) of the Business Intelligence insights?**

Road safety analysis provides Business Intelligence (BI) insights to a wide array of stakeholders who are both interested in understanding and improving road safety as a whole. Government entities such as the Ministry of Transport, which shapes transportation policies and regulations, are the primary consumers of BI insights in this context. In order to improve road safety across the nation, these insights become crucial. As well, agencies like the Road Transport Department use BI to implement targeted interventions and enhance overall road infrastructure based on data-driven insights to effectively manage and regulate road transport.

BI insights have a significant impact on law enforcement agencies, as represented by the Police Force. By identifying patterns in accidents, optimizing resource allocation, and refining enforcement strategies, these insights can reduce road accidents. In order to improve preparation and coordination, emergency service departments, such as fire departments and paramedics, benefit from BI for prompt and efficient responses to accidents. In addition to tailored campaigns,

educational programs, and enforcement initiatives, road safety organizations can use BI insights to focus on targeted campaigns, education, and enforcement.

Public transportation companies and logistics companies use BI insights to optimize routes, improve vehicle safety, and reduce accident risks. By leveraging BI, traffic management agencies can improve traffic flow, improve signage, and improve road safety. Through awareness campaigns, educational programs, and public service announcements, business intelligence insights can be spread to the general public, empowering individuals to make informed decisions and adopt safer driving practices. The media can also use BI insights to raise awareness about road safety, report on trends, and contribute to public discourse.

The insights from business intelligence serve as a powerful tool for fostering collaboration and decision-making among diverse stakeholders, which results in improved road safety. In order to enable evidence-based interventions, policy changes, and public awareness initiatives, actionable information is the overarching goal. These stakeholders work together in order to reduce road accidents and improve overall road safety in their communities.

➤ **What might you recommend the company do in order to improve their Business Intelligence systems?**

There are several key recommendations for improving the Business Intelligence (BI) systems of the company for road safety analysis. As a result of integrating advanced analytics and machine learning algorithms into predictive modeling, insights can be significantly more accurate, aiding in the development of proactive accident prevention measures. A real-time integration of data is crucial to ensuring that the BI system reflects the most recent information, especially when responding to road safety emergencies as soon as possible. To enhance the accuracy and reliability of BI insights, robust data quality processes and governance standards, such as regular data cleaning and validation, are essential.

By incorporating social media, traffic camera feeds, and other IoT devices into data sources, road safety factors can be seen from a broader perspective. For effective emergency response, it is crucial that the planned mobile application is continuously enhanced, incorporating user feedback and ensuring user friendliness. By integrating APIs with external data sources and systems, other

road safety agencies can collaborate. The BI system is effective when users receive training and assistance, along with documentation and training sessions.

The ability to notify stakeholders of critical insights or emerging trends through automated reporting and alert systems ensures timely awareness and response to changing road safety conditions. In order to accommodate future increases in data volume and user demand, scalability planning is essential. Furthermore, if data visualization techniques are aligned with best practices, insights will be presented in an intuitive and actionable manner. Collaborative BI platforms promote real-time collaboration and insight sharing between different stakeholders.

For sustained effectiveness, it is essential to monitor and optimize the BI system for efficiency, including addressing bottlenecks and optimizing query performance. Continuous improvement is driven by creating a feedback mechanism for users to provide input on usability and functionality. Security measures, such as encryption and access controls, and regular security audits, help protect sensitive road safety information. By maintaining a culture of continuous improvement within the BI team, the company can ensure it remains on top of its BI game. With the implementation of these recommendations, the company can optimize its business intelligence systems, ensuring they are effective, scalable, and aligned with stakeholders' evolving needs to improve road safety.