EXAMINER MARKING RUBRIC
(MARKS: 40)

This part only for examiners (Each ROW only needs to choose one):

(A) Presentation's Marking Rubric: / 5

Requirement /Marks	Exceptional (5) The group able to present it well and answer the entire questions correctly.	Acceptable (3-4) The group able to present it well and able to answer partially questions correctly.	Amateur (2) The group able to present it but not able to answer the entire questions correctly.	Unsatisfactory (1) The group unable to present well and unable to answer the entire questions correctly.
Presentation				

(B) Data Visualization's Marking Rubrics: / 20

Requirement/ Marks	Good (4-5)	Satisfactory (2-3)	Unsatisfactory (0-1)
Marks below are average for all questions asked			
1. Relevant question asked			
2. Suitable Data Visualization Techniques used			
3. Suitable selection of colours / fonts / shapes.			
4. Completeness / Depth of Analysis.			
		Total /20	

(C) Report's Marking Rubric: / 15

Requirement/ Marks	**Exceptional (8 - 10)** The report is well written and clearly explains what the program is accomplishing and how.	**Acceptable (5 - 7)** The report is well written but not clearly explains what the program is accomplishing and how.	**Amateur (3 - 4)** The report is not well written and not clearly explains what the program is accomplishing and how.	**Unsatisfactory (0 - 2)** The report is not well written, not clearly explains what the program is accomplishing and does not fulfill report's requirements.	
Report i. Project Introduction, Problem Statement, and Objective ii. Dataset Information					**/5**
iii. Findings and Conclusion iv. Lesson Learnt from Project.					**/10**
					/15

TABLE OF CONTENT

CONTENT		PAGE
1.0	Problem Statement	1
2.0	Dataset Information	1
2.1	Description of attributes	1
2.2	Data Dictionary	3
3.0	Objective	5
4.0	Questions or Hypotheses	6
4.1	Data Preprocessing / Data Transformation	7
4.2	Data Visualizations	7
Summary / Conclusion		17
Lesson Learnt from Project		17

1.0 Problem Statement

Road traffic crashes are the leading cause of death among citizens. As vehicle technology advanced and speeds increased, these accidents became more and more destructive. Every year the lives of approximately 1.35 million people are cut short as a result of the road traffic crash. Furthermore, between 20 and 50 million more people suffer nonfatal injuries, with many incurring a disability because of their injuries.

Accidents are relatively unpredictable. The widely known contributing factor deduce to road traffic crashes involved road and environmental factors. It is challenging to keep pace with the factors and effects of these accidents. In certain cases, incidents due to poor road environment management occur in many ways, such as road surface conditions, roadside conditions, speeding, and the conditions driving ambiance itself. Prior to this issue, this project was conducted to explore factors that lead to accident fatalities to avoid the increasing number of road accidents year by year.

2.0 Dataset Information

The dataset information will be described in the next subsection.

2.1 Description of attributes

Table 2.1 Attributes Description of the Dataset

No	Attribute	Description
1	ACCNUM	Accident Number
2	YEAR	Year of Accident
3	MONTH	Month of Accident
4	DAY	Day of Accident
5	HOUR	Hour of Accident
6	MINUTES	Minute of Accident
7	WEEKDAY	Weekday of Accident

8	LATITUDE	Latitude of the Accident
9	LONGITUDE	Longitude of the Accident
10	WARD_NAME	City Ward
11	WARD_ID	City Ward ID
12	HOOD_NAME	Neighbourhood Name
13	HOOD_ID	Neighbourhood ID
14	DIVISION	Police Division
15	DISTRICT	City District
16	STREET1	Street of Accident
17	STREET2	Street of Accident
18	OFFSET	Distance and direction of the accident
19	ROAD_CLASS	Road Classification
20	LOCCOORD	Location Coordinate
21	ACCLOC	Accident Location
22	TRAFFCTL	Traffic Control Type
23	VISIBILITY	Environment Condition
24	LIGHT	Light Condition
25	RDSFCOND	Road Surface Condition
26	ACCLASS	Classification of Accident
27	IMPACTYPE	Initial Impact Type
28	INVTYPE	Involvement Type
29	INVAGE	Age of Involved Party
30	INJURY	Severity of Injury
31	FATAL_NO	Sequential Number of Number of Fatal
32	INITDIR	Initial Direction of Travel
33	VEHTYPE	Type of Vehicle
34	MANOEUVER	Vehicle Manouever
35	DRIVACT	Apparent Driver Action
36	DRIVCOND	Driver Condition
37	PEDTYPE	Pedestrian Crash Type

38	PEDACT	Pedestrian Action
39	PEDCOND	Condition of Pedestrian
40	CYCLISTYPE	Cyclist Crash Type
41	CYCACT	Cyclict Action
42	CYCCOND	Cyclist Condition
43	PEDESTRIAN	Pedestrian Involved in Collision
44	CYCLIST	Cyclists Involved in Collision
45	AUTOMOBILE	Driver Involved in Collision
46	MOTORCYCLE	Motorcyclist Involved in Collision
47	TRUCK	Truck Driver Involved in Collision
48	TRSN_CITY_VEH	Transit or City Vehicle Involved in Collision
49	EMERG_VEH	Emergency Vehicle Involved in Collision
50	PASSENGER	Passenger Involved in Collision
51	SPEEDING	Speeding Related Collision
52	AG_DRIV	Aggressive and Distracted Driving Collision
53	REDLIGHT	Red Light Related Collision
54	ALCOHOL	Alcohol Related Collision
55	DISABILITY	Medical or Physical Disability Related Collision
56	FATAL	Fatal Injury in Collision

2.2 Data Dictionary

Table 2.2 Data Dictionary of the Dataset

No	**Attribute**	**Type**	**Size**
1	ACCNUM	Integer	4000
2	YEAR	Integer	11
3	MONTH	Integer	12
4	DAY	Integer	31
5	HOUR	Integer	24
6	MINUTES	Integer	60

7	WEEKDAY	Integer	7
8	LATITUDE	Double	3000
9	LONGITUDE	Double	3000
10	WARD_NAME	String	44
11	WARD_ID	Integer	44
12	HOOD_NAME	String	140
13	HOOD_ID	Integer	140
14	DIVISION	String	17
15	DISTRICT	String	6
16	STREET1	String	1000
17	STREET2	String	2000
18	OFFSET	String	283
19	ROAD_CLASS	String	8
20	LOCCOORD	String	5
21	ACCLOC	String	10
22	TRAFFCTL	String	11
23	VISIBILITY	String	9
24	LIGHT	String	10
25	RDSFCOND	String	10
26	ACCLASS	String	3
27	IMPACTYPE	String	10
28	INVTYPE	String	20
29	INVAGE	String	21
30	INJURY	String	5
31	FATAL_NO	Integer	78
32	INITDIR	String	4
33	VEHTYPE	String	27
34	MANOEUVER	String	16
35	DRIVACT	String	13
36	DRIVCOND	String	10

37	PEDTYPE	String	16
38	PEDACT	String	15
39	PEDCOND	String	10
40	CYCLISTYPE	String	22
41	CYCACT	String	11
42	CYCCOND	String	10
43	PEDESTRIAN	Boolean	2
44	CYCLIST	Boolean	2
45	AUTOMOBILE	Boolean	2
46	MOTORCYCLE	Boolean	2
47	TRUCK	Boolean	2
48	TRSN_CITY_VEH	Boolean	2
49	EMERG_VEH	Boolean	2
50	PASSENGER	Boolean	2
51	SPEEDING	Boolean	2
52	AG_DRIV	Boolean	2
53	REDLIGHT	Boolean	2
54	ALCOHOL	Boolean	2
55	DISABILITY	Boolean	2
56	FATAL	Boolean	2

3.0 Objective

Specifically, the project has focused on these three main objective:

i. To analyze the factor that causes road accidents from the year 2013 until 2017.
ii. To extract information from the dataset and make an analysis about the pattern of road accidents that result in killed or injured.
iii. To develop a comprehensive dashboard of the dataset and visualize the pattern in a graphical manner.

4.0 Questions or Hypotheses

4.1 Data Preprocessing / Data Transformation

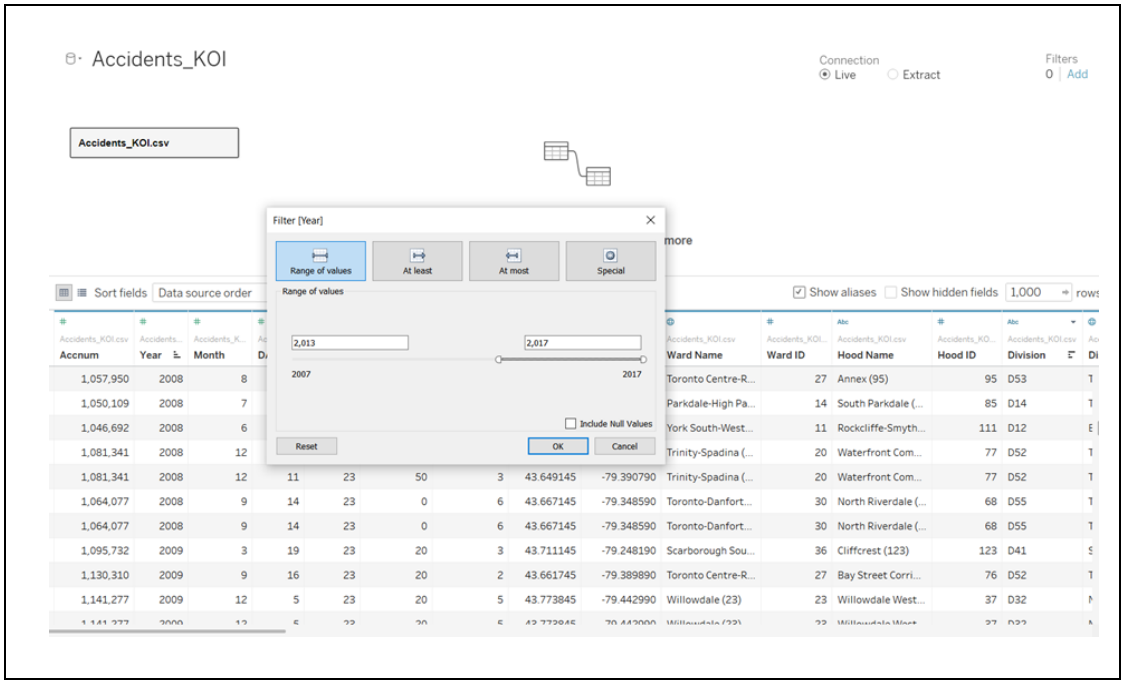

Figure 4.1 Filtering the Dataset

Figure 4.1 shows the filtered years, which is from 2013 until 2017. By removing the year 2008 until 2012, the data were pre-processed, and only the last five years of data were selected for analysis.

4.2 Data Visualizations

Question One: How does the number of accidents differ among different types of the vehicle each year?

Attributes used: Month, Year, Automobile, Cyclist, Emergency Veh, Motorcycle, Pedestrian, Truck

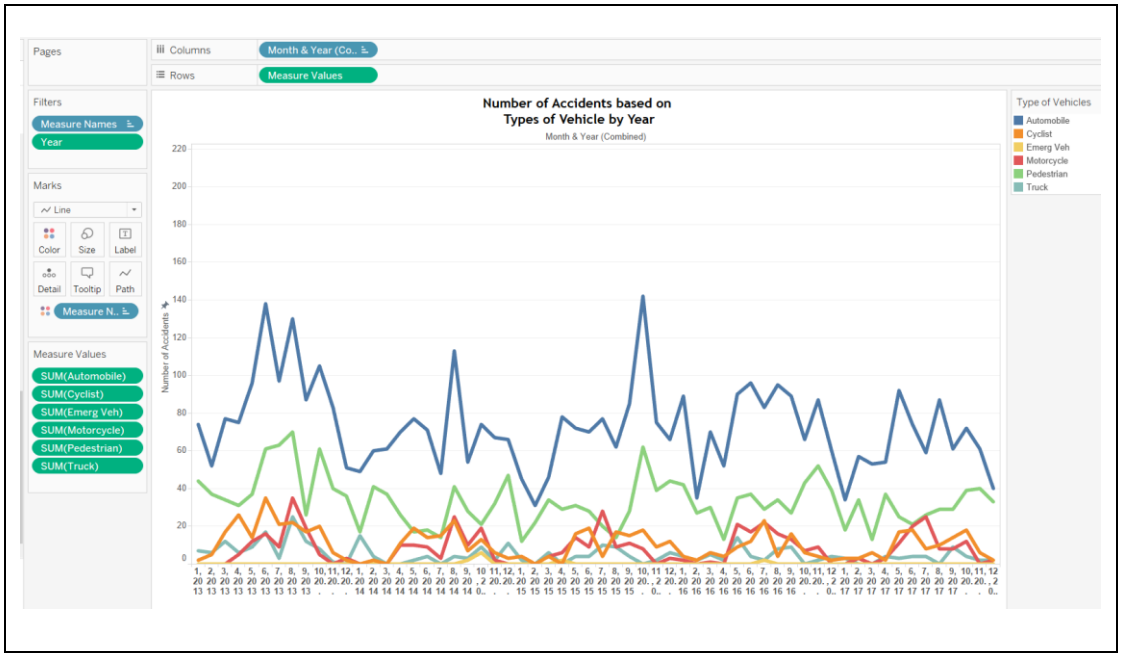

Figure 4.2 Number of Accidents based on Types of Vehicle by Year

Figure 4.2 shows a line chart of the number of accidents based on the type of vehicle from 2013 until 2017. Each of the colours represents a different type of vehicle. Blue for automobile, orange for bicycle, yellow for the emergency vehicle, red for motorcycle, green for pedestrian and cyan for truck.

The graph shows that throughout the years, the main cause for accidents are consistently automobiles. As seen in the Figure, automobiles continuously remain first in ranks and peaked in October 2015 with a total of 142 accidents. Second in ranks are pedestrian. Same with automobiles, it consistently remains in second, making it the second main cause of accidents. As for the other four types, the ranking between them keeps changing. Therefore, it can be concluded that automobiles are the main cause for the majority of accidents, followed closely by pedestrians.

Question Two: What is the percentage of different total cases of accidents?

Attributes used: AccNum, AccClass

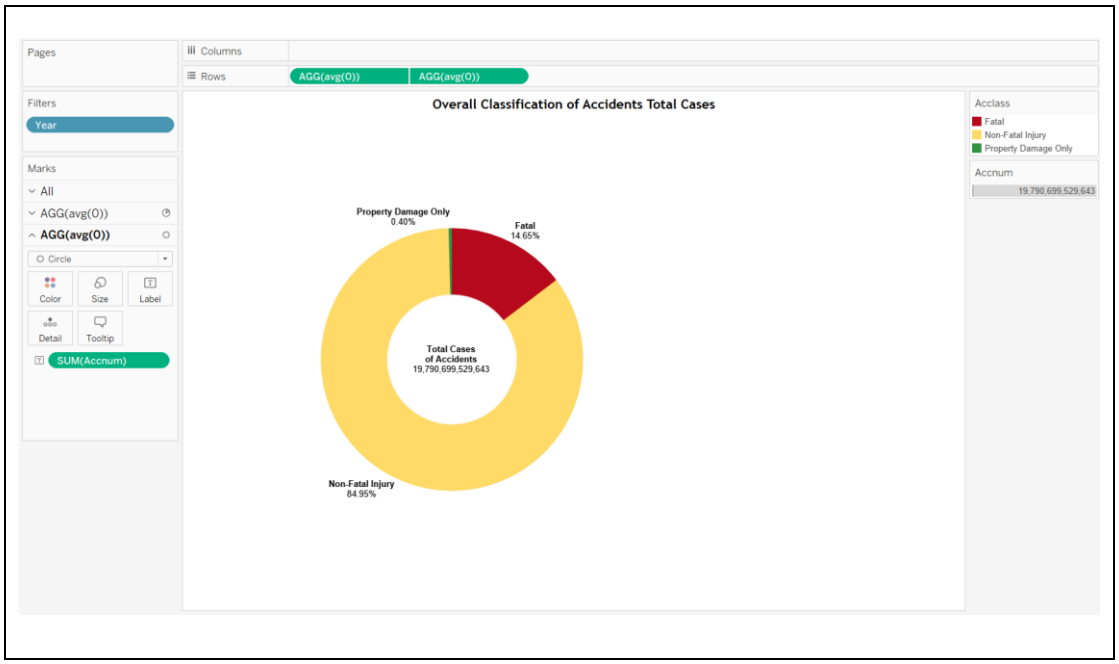

Figure 4.3 Overall Classification of Accidents Total Cases

Figure 4.3 shows a Donut chart of total accident cases on the different classifications of accidents from 2013 until 2017. The red colour represents a fatal accident, the yellow colour represents nonfatal injury, while property damage only is represented by green colour.

The chart shows that the nonfatal injury constituted a large proportion of total cases of accidents with 16, 820, 090, 521, 457 cases, indicating that most accidents occurred in districts of Toronto, Canada caused major and minor injury. Besides, fatal accidents accounted for 14.64% of overall cases with 2, 899, 357, 131, 154, followed by property damage with only 80, 026, 642, 080 cases. Thus, it can be inferred that fatal injury and property damage made up the minority of accident classification cases.

Question Three: Do environmental factors lead to road accidents? If so, which of these factors contributes most to the number of road accidents case?

Attributes: Visibility, Light, Road Surface Condition, Accident classification, Accident Number

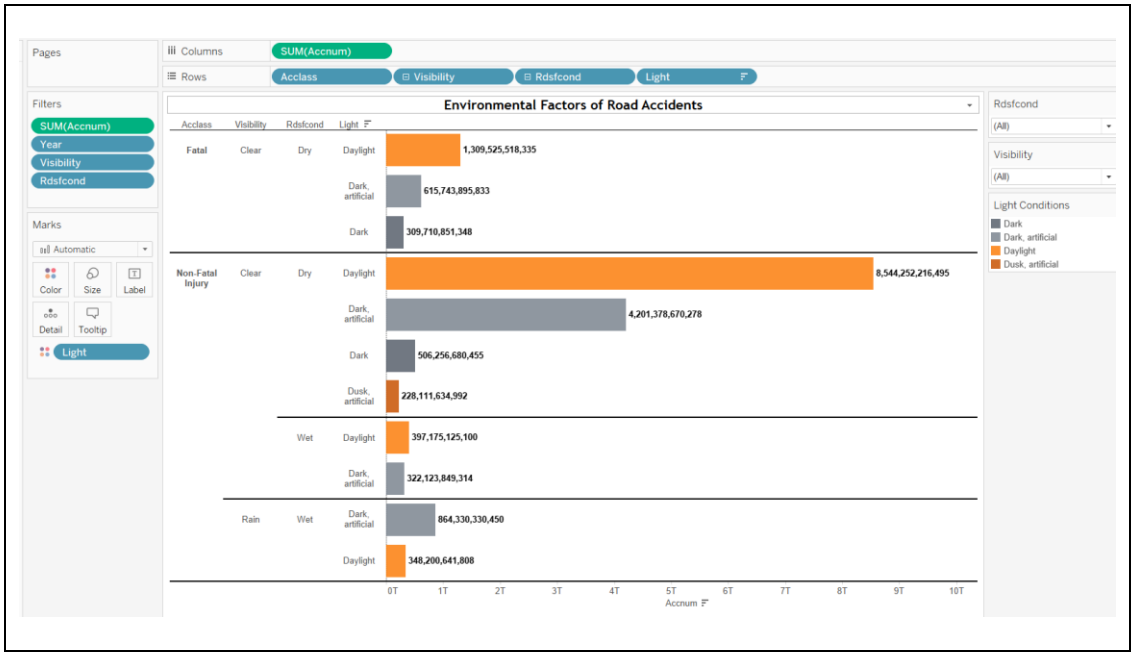

Figure 4.4 Environmental Factors of Road Accidents

Environmental factors causing accidents refer to factors that lead to accidents that are not attributable to human actions. Alternatively, the environment is partially responsible for certain characteristics of human actions, such as driving speed and safety measures being taken. Various environmental factors may cause accidents. Some common examples found on the dataset include visibility such as clear and rain, road surface condition either dry or wet, and light conditions such as dark, daylight, dark and artificial, or dusk and artificial. The bar graph in Figure 4.4 analyzed different environmental factors that causing fatal and nonfatal injury

From the bar graph, it can be observed that there are significantly more cases of nonfatal injury than fatal injury. For the classification of fatal accidents, most cases occur in clear visibility and dry surface conditions of the road. Under light conditions, most accidents occur in daylight rather than dark, with 1, 309, 525, 518, 335 cases. This

implies that most people misjudged their visual performance in daylight, and therefore, do not take precautionary measures.

Moving on to the nonfatal injury, it is noticeable that the highest cases occurred in clear visibility and dry road surface condition in daylight with 8, 544, 252, 216, 495 cases, followed by dark and artificial visibility with 4,201,376,670,278. The lowest cases under this environment are dusk, artificial with 228, 111, 634, and 992 cases. Visibility is particularly poor at dusk because natural light is reduced, and it is not dark enough for headlights to be effective. It can be inferred that most people must be very alert to the pedestrian heading to and from work, runners, or joggers that use the roads at these times for a training run. By examining the dataset, pedestrian and cyclist collisions with the automobile are the leading causes of accidents. Thus, it is essential for pedestrians and cyclists to walk on the side of the road heading towards oncoming vehicles and to increase their visibility to others through reflective and high visibility clothing. In comparison, not many cases are reported during rain visibility and wet road surface condition. The highest cases are 864, 330, 339, 450 cases in dark, artificial light conditions. Thus, it suggests that drivers must be extra careful driving in the rain because it is slippery and worsens if the light condition is dark.

The results indicate that road accidents, in general, depend on many independent variables. Analyses of environmental factors such as visibility, road surface, and light conditions proved to be attributable to the number of cases reported. Based on the strong correlation established between environmental factors and accident classification, increased light levels are believed to mitigate the frequency and severity of accidents directly. Besides, dry road surface conditions cause more overall cases than the wet road surface. This is a significant environmental factor that causes a large number of road accident cases. Although the road is dry, people should take extra precautionary steps. These results suggest a strong association between the total number of road accident cases and the environmental factors that caused them.

Sub-question: What driver action caused the majority of accidents?
Attributes used: Count of rows in accidents_KOI dataset, Drivact

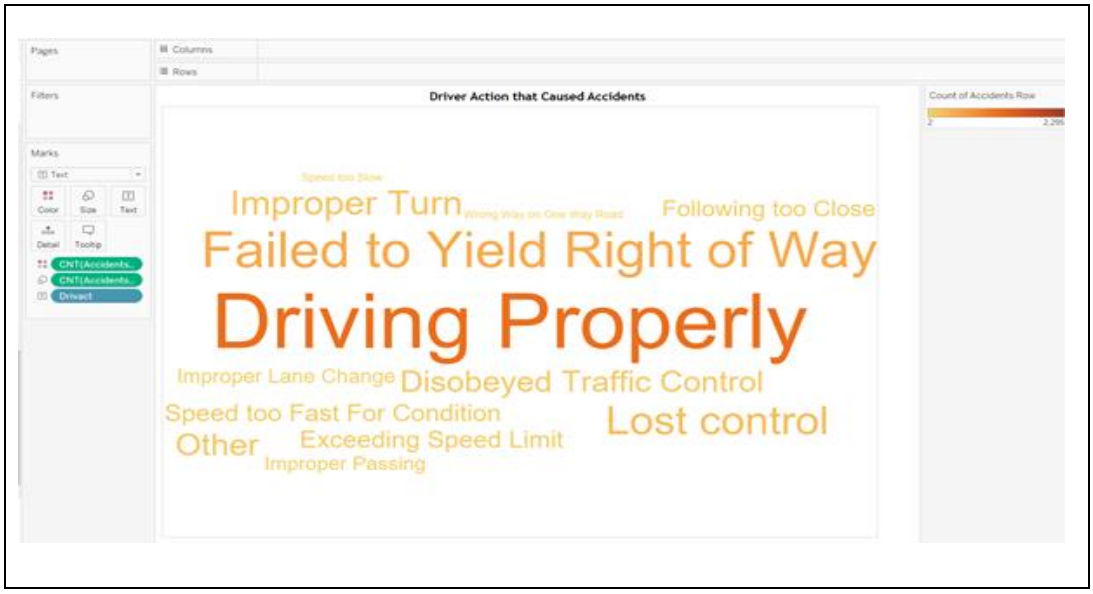

Figure 4.5 Word Cloud of Driver Action that Caused Accidents

The word cloud produced in Figure 4.5 represents words in 'Drivact' attribute text corpus row. From the word cloud, it can be observed that most driver action that causes an accident is 'Driving Properly' with 1,193 rows denoted by the strong colour contrast and largest font. A possible explanation for this is that even if the driver drives properly, other drivers may drive in an unsafe way and cause the risk of accidents to occur. Besides, 'Failed to yield right of way', 'improper turn', and 'lost control' are also notable in the word cloud generated. For such a reason, these common aggressive driving actions must be avoided at all costs. The least number of driver actions that caused accidents is 'speed too slow' and 'wrong way on one way road' with only 2 rows. Although these driver actions are less likely to cause accidents, dangerous driving situations may also arise from all these related behaviors.

Question four: Based on the data, what is the major type of injury caused by accident?
Attributes: Injury, No of accident

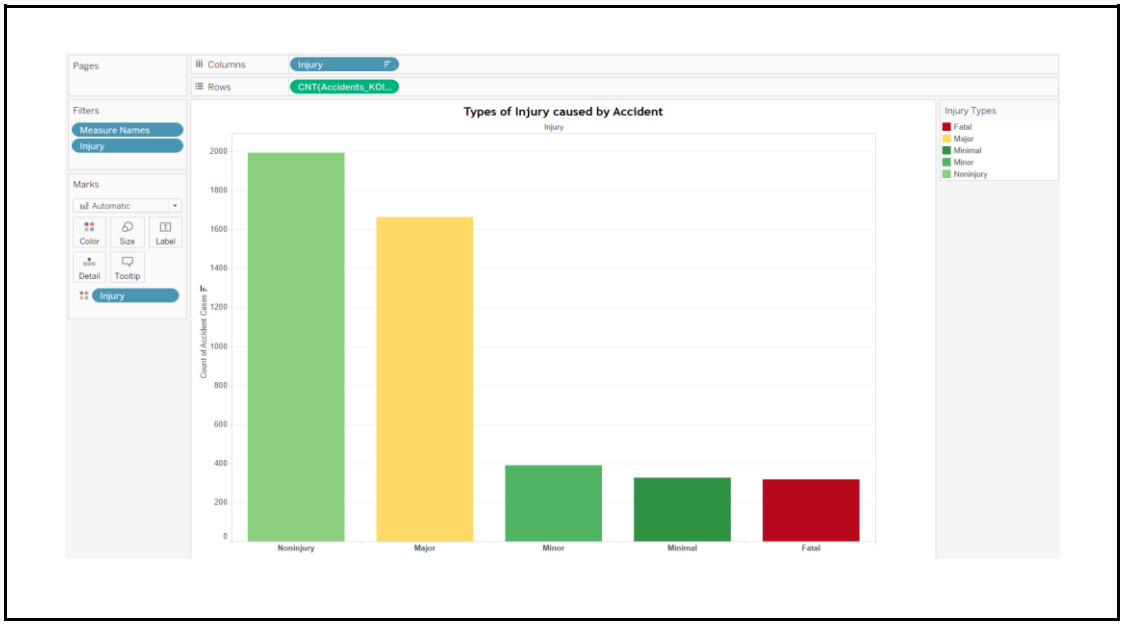

Figure 4.6 Major Type of Injury caused by accident

Injury is damage to the body caused by external force, which in this case it is caused by accidents. Accidental injuries can come from falling, hitting or throwing from the vehicle during the accidents. There are two classes for the accident, which is fatal, and nonfatal. In nonfatal accidents, there are several types of injury. There are major, minor, minimal and non-injury.

The graph shows in figure 4.6 are the relationship of number of accidents and the type of injury caused by the accident. From the graph, the highest count of types of injury caused by the accident are non-injury with 4576 of accidents. Followed by major injury with 4154 accidents. For minor injury, the number of accidents that caused the injury are 969 cases. Minimal injury have 793 cases that caused by the accident. And the lowest count of injury type that caused the accident is major type, with 593 cases only.

With that, we can conclude that most of the accidents from 2013 to 2017 did not cause any injury, as the higher type of injury is in the nonfatal accident class, which is non-injury.

Sub-question: Then, what is the highest mode of transportation that was involved in the accident?

Attributes used: Injury, Automobile, Cyclist, Emergency Vehicle, Motorcycle, Truck

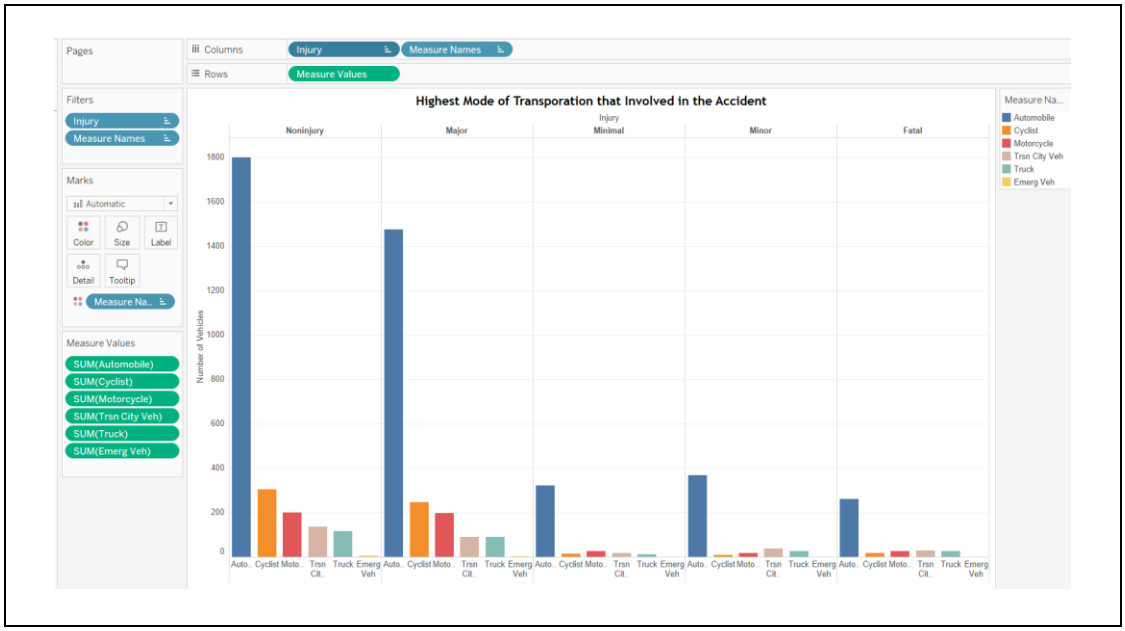

Figure 4.7 Highest Mode of Transportation involved in the Accident

The Figure above shows that the highest mode of transportation involved in the accident is the automobile, for each type of injury. Automobiles have 1799 cases in non-injury, 1474 cases in major, 369 cases in minor, 323 in minimal, and 261 cases in fatal.

For other modes of transportation have a different rank on other types of injury. For non-injury and major, they have the same rank of the type of transportation involved in accidents is the automobile, cyclist, motorcycle, transit or city vehicle, truck and emergency vehicle.

Therefore, the result indicates that automobile have the highest mode of transportation that involved in accidents, compared to other modes of transportation for accidents from 2013 to 2017.

Question Five: Which district in Canada has the highest and least number of accidents?

Attributes: Longitude, Latitude, District, Month, Year, Accnum

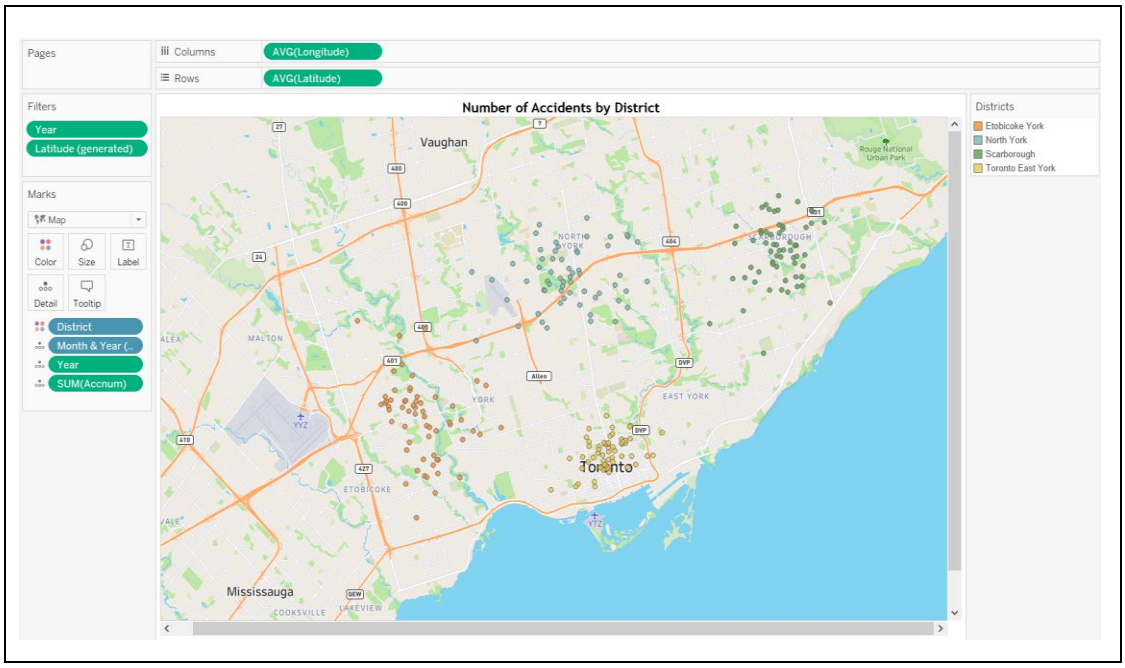

Figure 4.8 Number of Accidents by District

Figure 4.8 above shows the basic mapping of spatial data for the number of accidents by the district in Canada. The type of cartographic uses is a conventional plot where the points are plotted on a map using latitude and longitude. This map visualization can give readers a direct focus on the area or location they want to know. There are four main districts in this dataset, which are Etobicoke York (Orange), North York (Light Blue), Scarborough (Green), and Toronto East York (Yellow).

The highest number of accidents occurred in Etobicoke York district, which has the largest number of plots. The second highest is followed by North York and Scarborough district. In comparison, the least number of accidents in Canada was in Toronto East York, which has a smaller number of plots compared to the other districts.

Sub-questions: For each district, what type of vehicle contributes the most and the least in accidents?

Attributes used: District, Automobile, Cyclist, Emerg veh, Motorcycle, Pedestrian, Truck

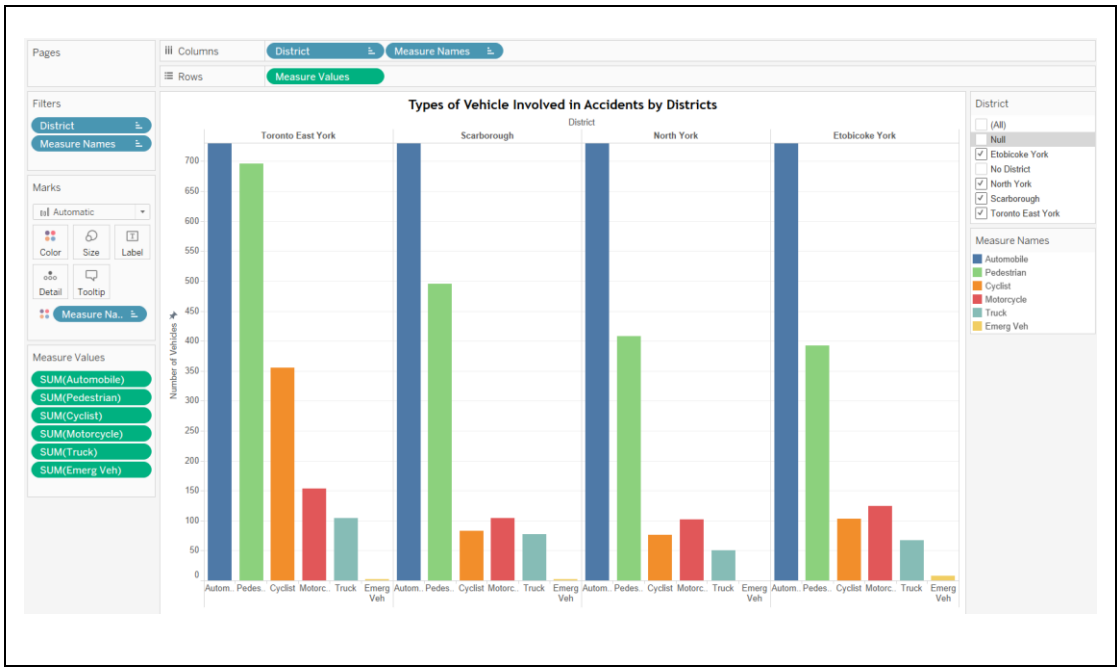

Figure 4.9 Types of Vehicles Contributing to Accidents for Each District

Figure 4.9 above shows a bar chart for the types of vehicles that contribute to accidents in each district (Toronto East York, Scarborough, North York, and Etobicoke York,Toronto East York). In the Toronto East York district, the most type of vehicle that contributes to accidents is an Automobile. Then followed by Pedestrian, Cyclist, Motorcycle, Truck, and the least contributing to the accidents is an Emergency Vehicles. For Scarborough district, the most type of vehicle that contributes to accidents is an Automobile. Then followed by Pedestrian, Motorcycle, Cyclist, Truck and the least contributing to the accidents is an Emergency Vehicles.

Moreover, for North York and Etobicoke York district, the most type of vehicle that contributes to accidents is an Automobile. Then followed by Pedestrian, Motorcycle,

Truck, Cyclist, and the least contributing to the accidents is an Emergency Vehicles.

It can be concluded that Automobiles are the most vehicles that contribute to accidents while Emergency vehicles are the fewest vehicles that contribute to accidents for each district in Canada. Pedestrians also were the second highest that lead to accidents. The total number of each vehicle that contributes to accidents in each district can be shown in Table 4.1 below.

Table 4.1 Total Number of Vehicles That Contributes to Accidents

District	Type of Vehicles (Descending Order)
Etobicoke York	i. Automobile (1, 039) ii. Pedestrian (392) iii. Cyclist (103) iv. Motorcycle (124) v. Truck (67) vi. Emergency Vehicles (8)
North York	i. Automobile (934) ii. Pedestrian (408) iii. Cyclist (76) iv. Motorcycle (102) v. Truck (50) vi. Emergency Vehicles (0)
Scarborough	i. Automobile (1, 092) ii. Pedestrian (495) iii. Cyclist (83) iv. Motorcycle (104) v. Truck (77) vi. Emergency Vehicles (2)
Toronto East York	i. Automobile (1, 307) ii. Pedestrian (696) iii. Cyclist (355) iv. Motorcycle (154) v. Truck (104) vi. Emergency Vehicles (2)

Summary / Conclusion

In conclusion, our main purpose for the study is to analyze the factors that lead to accident fatalities to avoid increasing road accidents year by year. By visualizing the data, we can see the related data pattern that shows the leading causes or hotspots for fatal or nonfatal accidents. As we can see, by using data visualization, it is easier to understand more about the datasets. Beyond that, all of our objectives we successfully achieved.

Lesson Learnt from Project

The lesson learned from this project is that data visualization is effective when the data visualizations are designed with the right data visualization. It is important to design the right visualizations to avoid any confusion and unnecessary complexity. The data visualization should be able to reflect and display important information. Moreover, a good data visualization can be determined by the understanding and proper interpretation by team members and viewers regarding the insights that can be analyzed from the visualization. There are many graphs and charts such as line graphs, pie charts, heat maps, and bubble charts that can be used to represent data visualization. Each graph and chart has its own purpose and is suitable for use with a particular type of data. For example, in this project, map visualization is used for geographical data where accident hotspots based on district need to be analyzed, and donut charts are used to visualize the entire data with classification. Hence, using the right visualization helps to analyze and comprehend the data easier and faster.