Analyzing Road Traffic Collisions in Canada

Group 2: Stanley Subrata, Jessie Ma, Amit Kaushik

Introduction:

Road traffic collisions continue to be a leading cause of injuries, fatalities, and economic losses in Canada, posing a significant threat to public safety and well-being. The primary goal of this study is to leverage the wealth of information present in the National Collision Database and perform a comprehensive examination of road traffic collisions in Canada in the year 2019. The report aims to address the following key challenges:

1. **Identify root causes and visible trends of the number of collisions in 2019:**
   Determine the distribution of major variables and problem areas contributing to road collisions. We will look at the number of collisions relating to time, driver behaviour, road conditions, environmental variables, and vehicle characteristics.

2. **Analyzing the relationship between the variables and the severity of the accident:**
   Determine how each variable can affect the severity of the collision. Additionally, prove that using a safety device can decrease the severity of the collisions.

3. **Addressing Data Quality and Completeness:** Ensuring the accuracy, consistency, and completeness of data in the NCDB to mitigate potential biases and uncertainties in the analysis. This involves identifying and addressing data gaps and limitations.

By conducting a rigorous and data-driven analysis of the National Collision Database, this project aims to contribute valuable insights that will empower policymakers, law enforcement agencies, and relevant stakeholders to develop evidence-based road safety initiatives. Ultimately, the project aspires to reduce the frequency and severity of road traffic collisions in Canada, safeguarding lives, preserving public health, and minimizing the economic burden associated with road accidents.

Data Preparation:

The National Collision Database (NCDB) in Canada serves as a comprehensive repository of collision-related data, capturing crucial information about the circumstances, causes, and outcomes of road accidents across the country. As the data is maintained by the Government of Canada, the quality of the data is acceptable for analysis once a moderate effort of cleanup is completed. Additionally, a data dictionary is provided by the NCDB to describe the variables contained in the raw data.

Unfortunately, different jurisdictions do not always provide a certain data element due to different reporting requirements. Additionally, some variables are recorded as “unknown”, as they can be impossible to determine i.e., cases that involve a hit and run. For the purpose of this report, data that do not contain the full list of variables (those with unknowns and non-reports) are excluded from the analysis.

Additional features were created to simplify some variables for the purposes of our analysis. The following features were engineered based on the data dictionary:

- **Datetime**: Time in the raw data is provided to the nearest hour, and data from multiple variables containing time information of was converted to time series to prepare for any analysis that involves a certain time range, i.e., how many recorded collisions in the month of January.

- **Collision Type**: The raw data provides a comprehensive typing of collisions, and the level of granularity is too high for the purpose of this analysis. A new feature is created by grouping the types of collisions into 4 separate buckets: One vehicle, two vehicles in the same direction, two vehicles in different directions, and two vehicles where one is stationary.

- **Age and Vehicle Age**: The ages of people involved in the collisions were binned into 3 categories = Teens, Adults, and Seniors. Vehicle ages were binned similarly as well.

Data including the data dictionary can be found at this link: https://open.canada.ca/data/en/dataset/1eb9eba7-71d1-4b30-9fb1-30cbdab7e63a

Analysis:

Time-related factors:

The analysis starts by looking at the distribution of accidents by day of the week in figure 1 below. In 2019, **Friday had more reported collisions compared to the other days of the week** (~17% higher than the average for all days). On the other hand, the number of reported collisions dropped by a significant margin on Sundays (~19% lower than the average for all days).

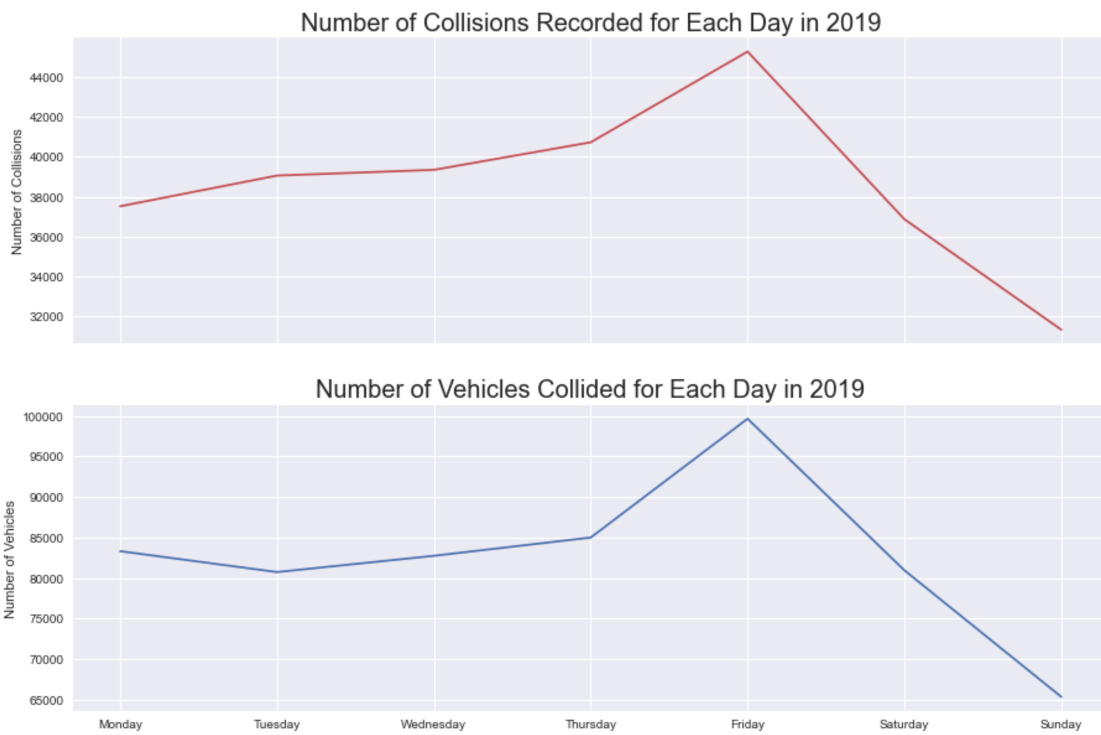

Figure 1

However, this does not always mean that a motor vehicle operator is more likely to record a collision on a Friday as the information on total trips taken is unknown. It is possible that there were more collisions recorded on Fridays because there were simply more trips taken on Fridays.

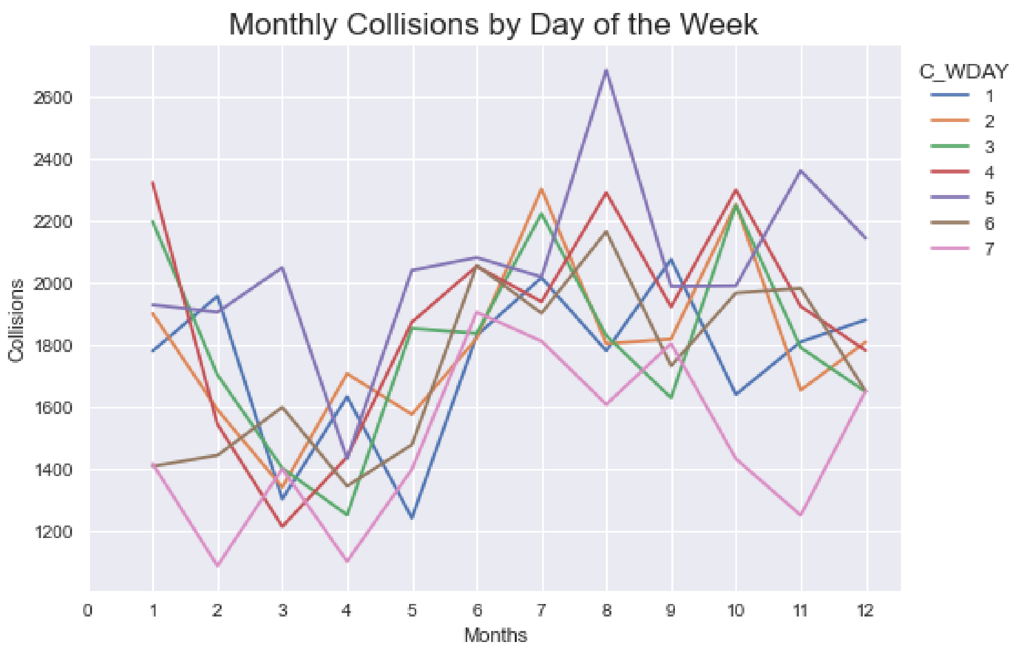

Figure 2

Additionally, by looking at Figure 2, there seems to be no significant differences in number of collisions reported for each day of the week in a specific month. Fridays in August have a slightly higher peak than most days in other months.

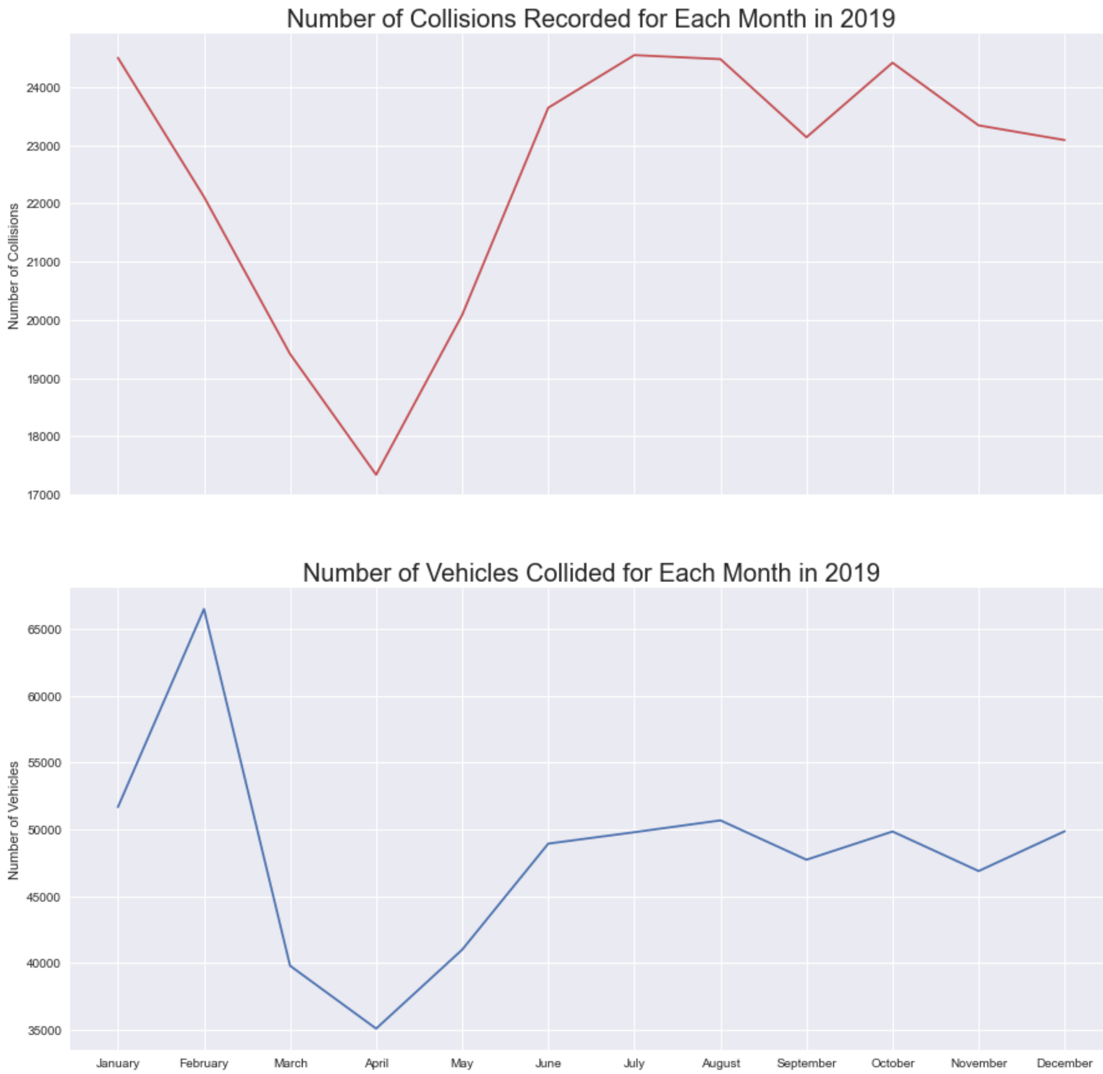

Figure 3

Figure 3 shows the number of reported collisions for each month in 2019. We observe that April has the least number of collisions recorded. Initially, we have hypothesized that the winter months would have more collisions recorded as the roads would be tougher to navigate as snowfall is relatively heavy in Canada. However, the winter months of December-March show no significant increase in number of collisions compared to the summer and fall months. **In fact, the number of collisions is generally quite level with the exception of March to May – which refer to the spring months**. The same conclusion can be extracted by examining the number of vehicles involved in a collision instead of the total number of incidents reported, where the spring months were observed to have a smaller number of vehicles involved in a collision compared to the other months.

One interesting observation is that during the month of February, there are significantly more vehicles involved in a collision compared to the other months, despite the fact that the number of collisions reported is on the lower end compared to other months i.e., January. **In fact, in February each collision on average involved around 3 vehicles compared to around 2 for all other months**.

<u>Weather Conditions:</u>

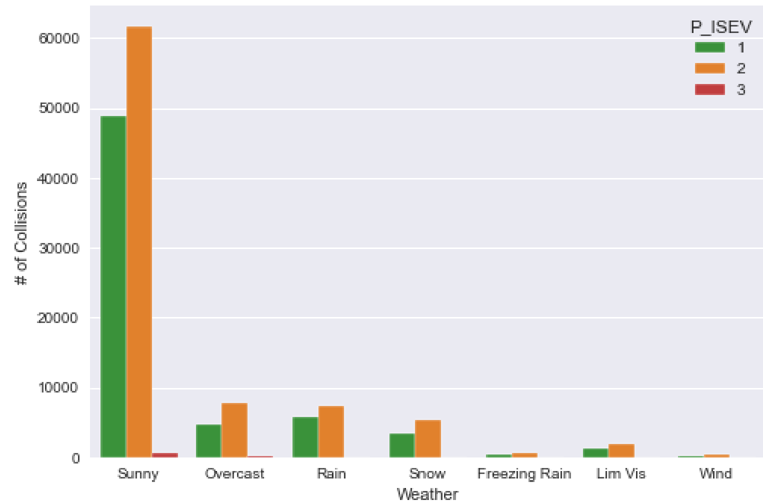

Figure 4

Figure 4 shows the distribution of reported collisions in different weather conditions. The total of collisions is further categorized by the severity of the collision, where the color green represents collisions resulting in no injury, orange represents collisions resulting in an injury, and red represents collisions resulting in a casualty.

The distribution of the number of collisions based on a specific weather condition shows that **most collisions happened during sunny weather, contrary to what most people would expect**. Around 73.9% of collisions happened on sunny days, followed by 8.7% on overcasts and 8.3% on rainy days.

It can be difficult to see in Figure 4 if a specific weather condition plays a role in determining how severe a collision would be. A new graph must be produced to analyze such correlation.

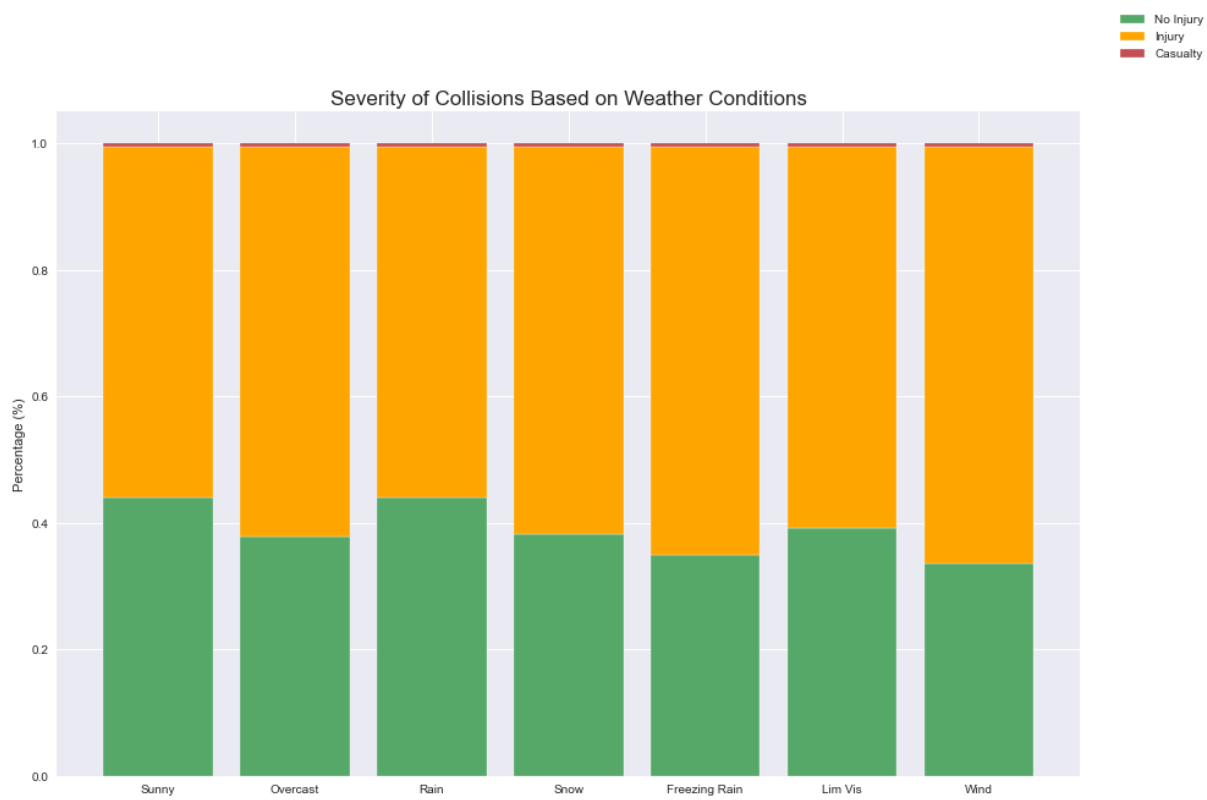

Figure 5

To further determine if a certain weather condition affects the severity of a collision, we have created another graph (figure 5). Figure 5 determines the proportion of the severity of the collisions for each weather condition, created by stacking the colored bars in Figure 4 and showing the proportion/percentage for each severity.

The first observation we can deduce from Figure 5 is that the red color is constant for all weather conditions, i.e., **a driver is as likely to be involved in a collision that results in a casualty during sunny weather as they would during freezing rain or a limited visibility condition**. Another observation is that collisions that happened during freezing rain or strong wind report a higher percentage of injuries compared to collisions that happened on a sunny or rainy day.

<u>Vehicle Age:</u>

The next few graphs follow the same methodology as the 2 graphs shown in the Weather Conditions section. Figure 6 and 7 below describes the distribution and severity of the collisions based on the age of the vehicle:

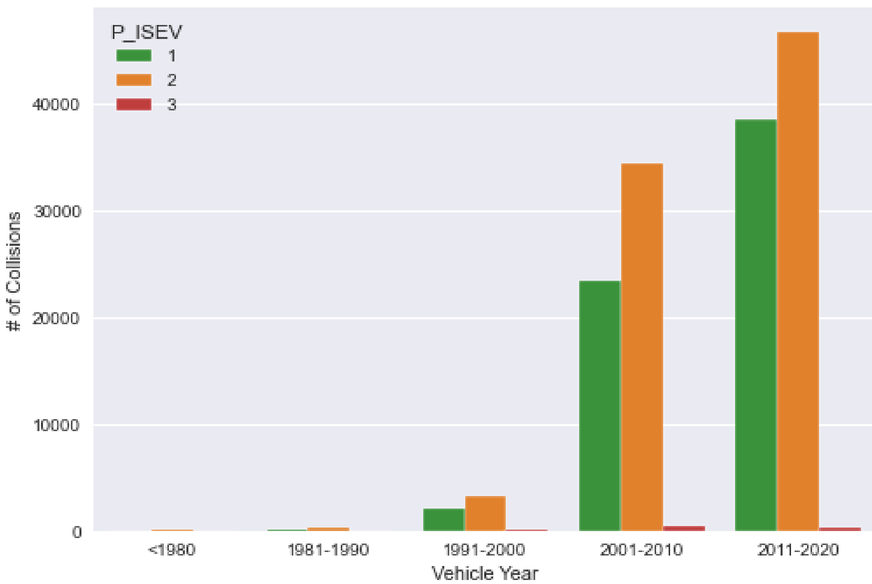

Figure 6

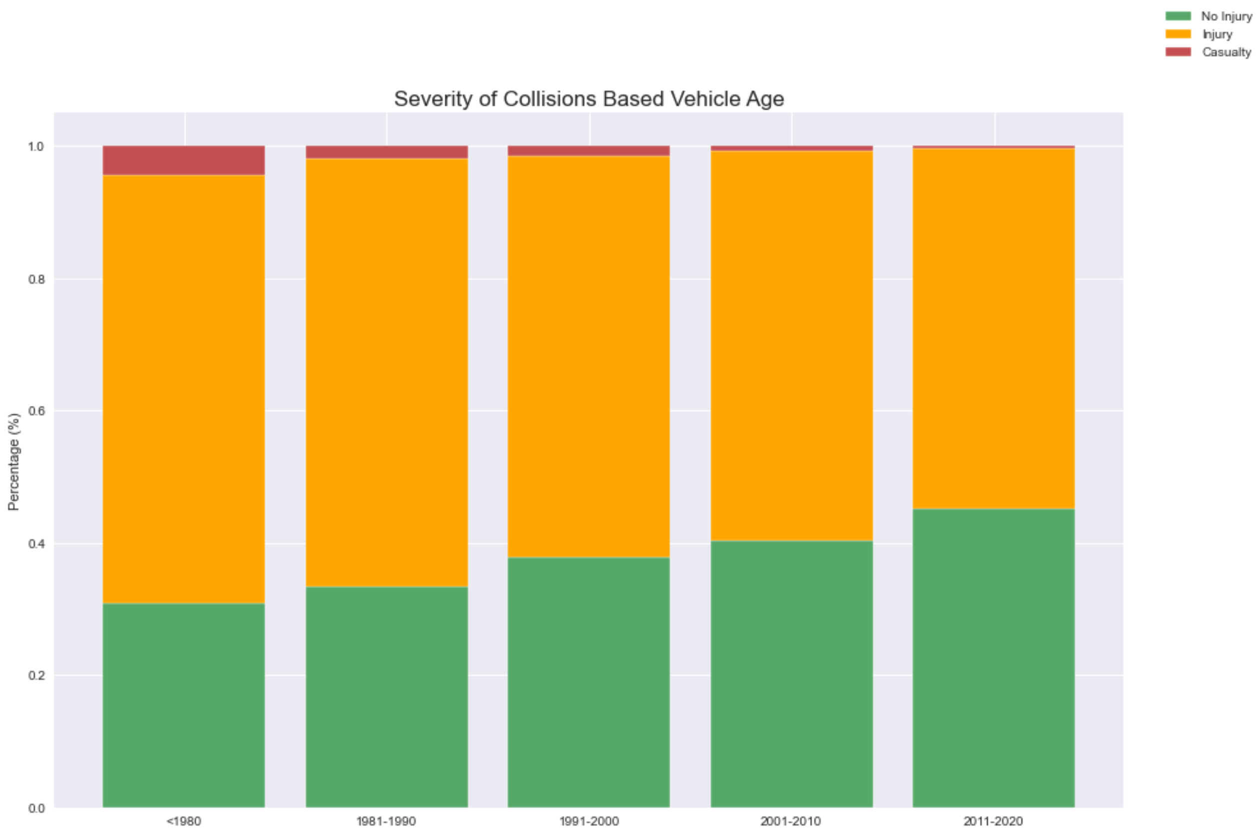

Figure 7

Figure 6 shows that the majority of the vehicles involved in a collision were made in 2001 and later. Figure 6 also shows a typical vehicle lifetime, which is around 20-30 years.

With vehicle age on the x axis, sorted from old to new, Figure 7 shows an increased safety for newer vehicles compared to the older models. **The proportion of recorded collisions with no injuries (green bars) increase with newer vehicles. On the other hand, the proportion of recorded collisions with casualties (red bars) decrease with newer vehicles**. For comparison, 3.3% of collisions for cars made between the year 1991-2000 resulted in a casualty compared to 0.7% of collisions for cars made between the year 2011-2020. This shows that newer models were most likely built with better materials to withstand collisions and features such as sensors to detect nearby objects.

<u>Collision Type:</u>

Figure 8 and 9 below describe the distribution and severity of the collision based on the type of collision.

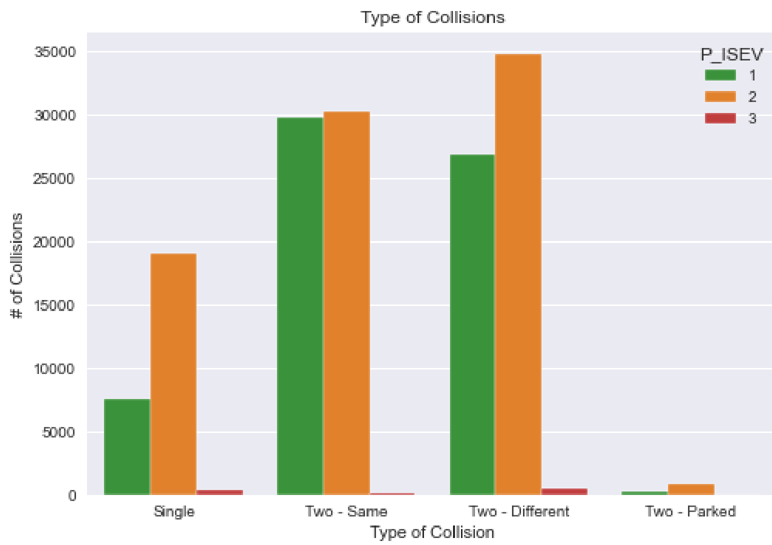

Figure 8

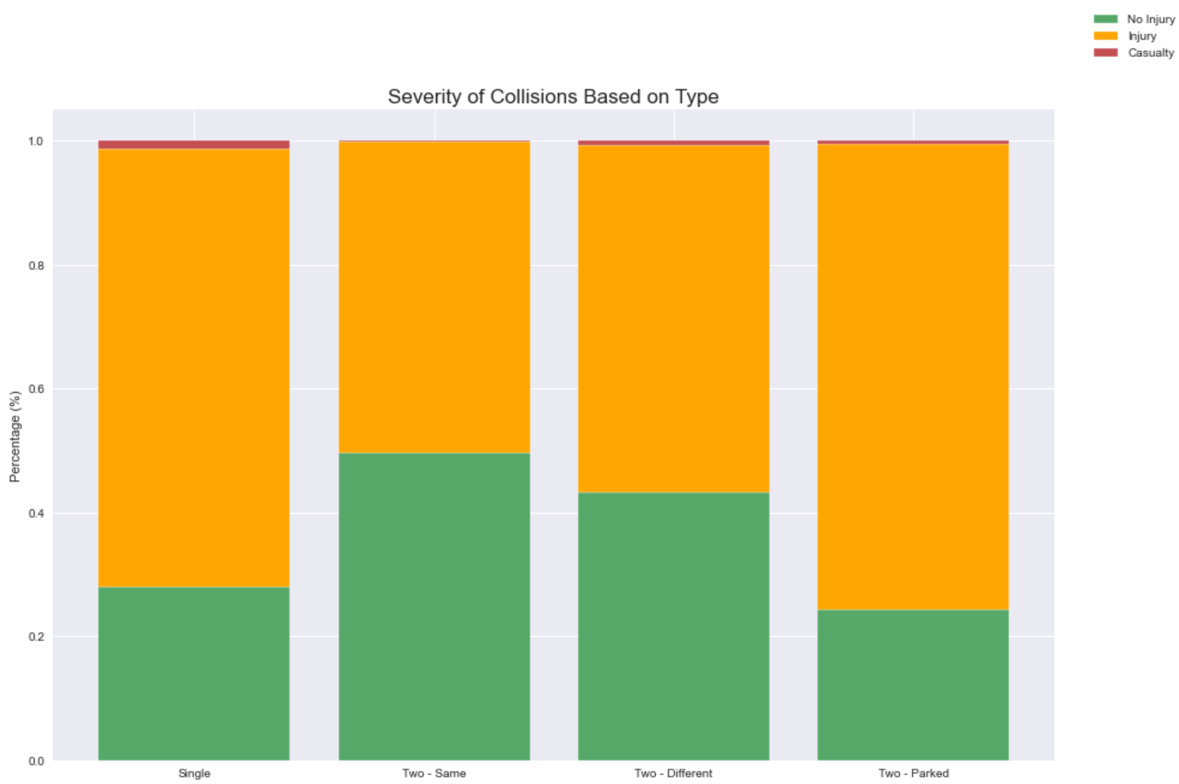

Figure 9

Figure 8 and 9 distributes the number of recorded collisions into 4 different types:

- Single: Involves one vehicle
- Two – Same: Involves two vehicles travelling in the same direction
- Two – Different: Involves two vehicles travelling in different directions
- Two – Parked: Involves two vehicles, with one vehicle stationary

Figure 8 shows that most collisions involved 2 different vehicles in motion (40% for same direction of travel, and 41% for different direction of travel), and that collisions involving a stationary vehicle is relatively rare (only 0.7%).

Figure 9 shows an interesting observation: **collisions involving one vehicle were more severe on average compared to collisions involving 2 different vehicles in motion**. This observation makes sense as Single type collisions include hitting a stationary object (e.g., a tree) or a moving object (e.g., an animal), and also include running off both right and left shoulders. Note that **there may be a bias towards a more disastrous situation for collisions of this type**, i.e., a vehicle that runs over a raccoon is less likely to be recorded in the database, whereas a vehicle that collides with a moose most likely would appear on the record.

Other minor observations from Figure 9:

- The average severity of collisions involving one vehicle and collisions involving a stationary vehicle is relatively similar
- The average severity of collisions involving two vehicles travelling on different directions is higher compared to collisions involving two vehicles travelling in the same direction, which makes sense as the impact would be greater
- Collisions involving two vehicles travelling in the same direction has almost no recorded casualties

<u>Sex:</u>

Figure 10.a and 10.b below compares the sex for all recorded collisions for each month in 2019. The graph shows that **more males are involved in collisions compared to females for all months**. It is also interesting to note that collisions involving females are more likely to result in an injury whereas in males the injury/no injury probabilities are an equal split.

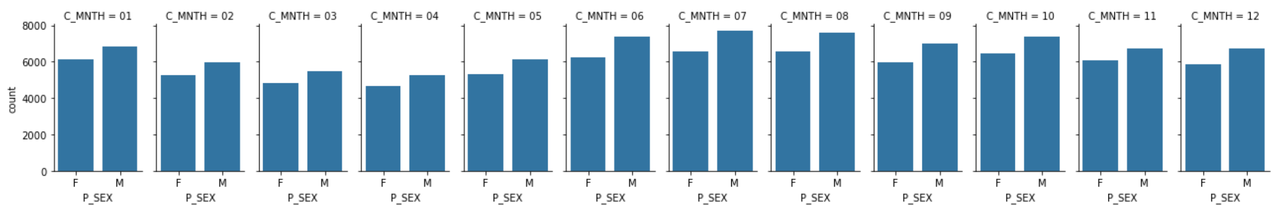

Figure 10.a

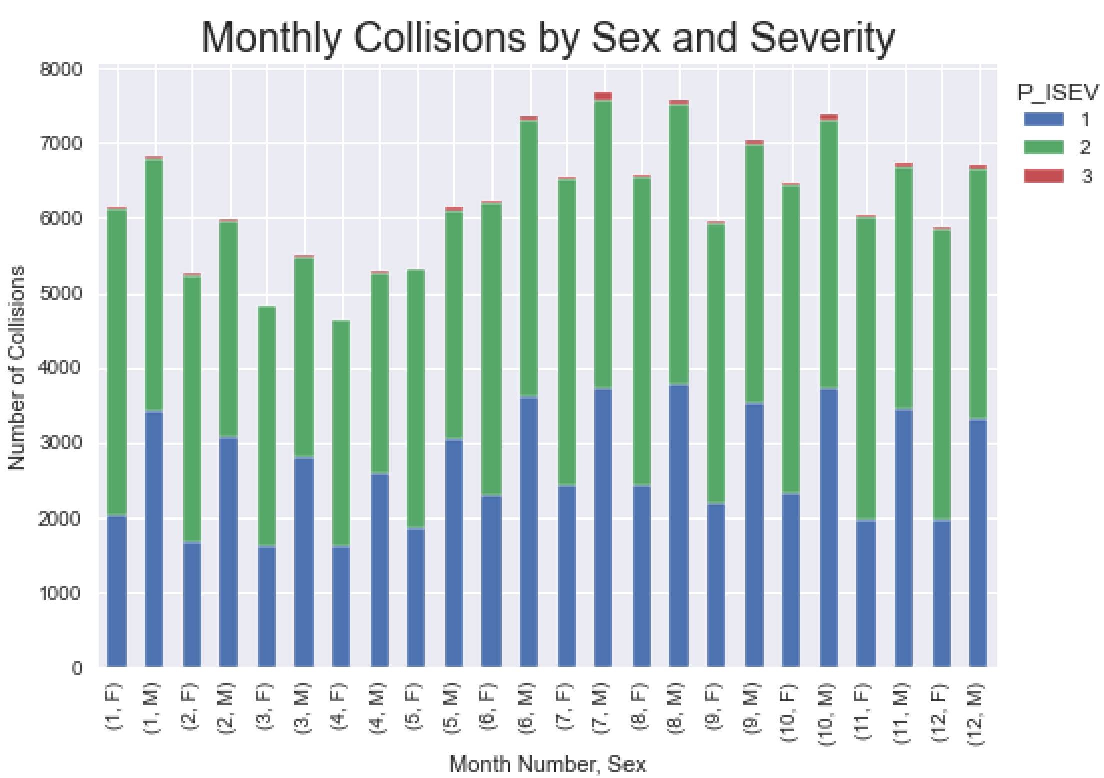

Figure 10.b

<u>Safety Device:</u>

One goal of this report is to prove that the usage of a safety device can decrease the severity of a collision. To show that this is true, we created a proportional stacked bar graph similar to Figure 7 and 9, with the safety variables below:

- None: no safety device used
- Safety: safety device used
- Helmet: helmet worn (for motorcyclists, snowmobilers, etc.)
- Other: other kinds of safety device used
- N/A: no safety device equipped (e.g., buses)

Figure 11 below shows the proportion of collision severity based on the safety device used:

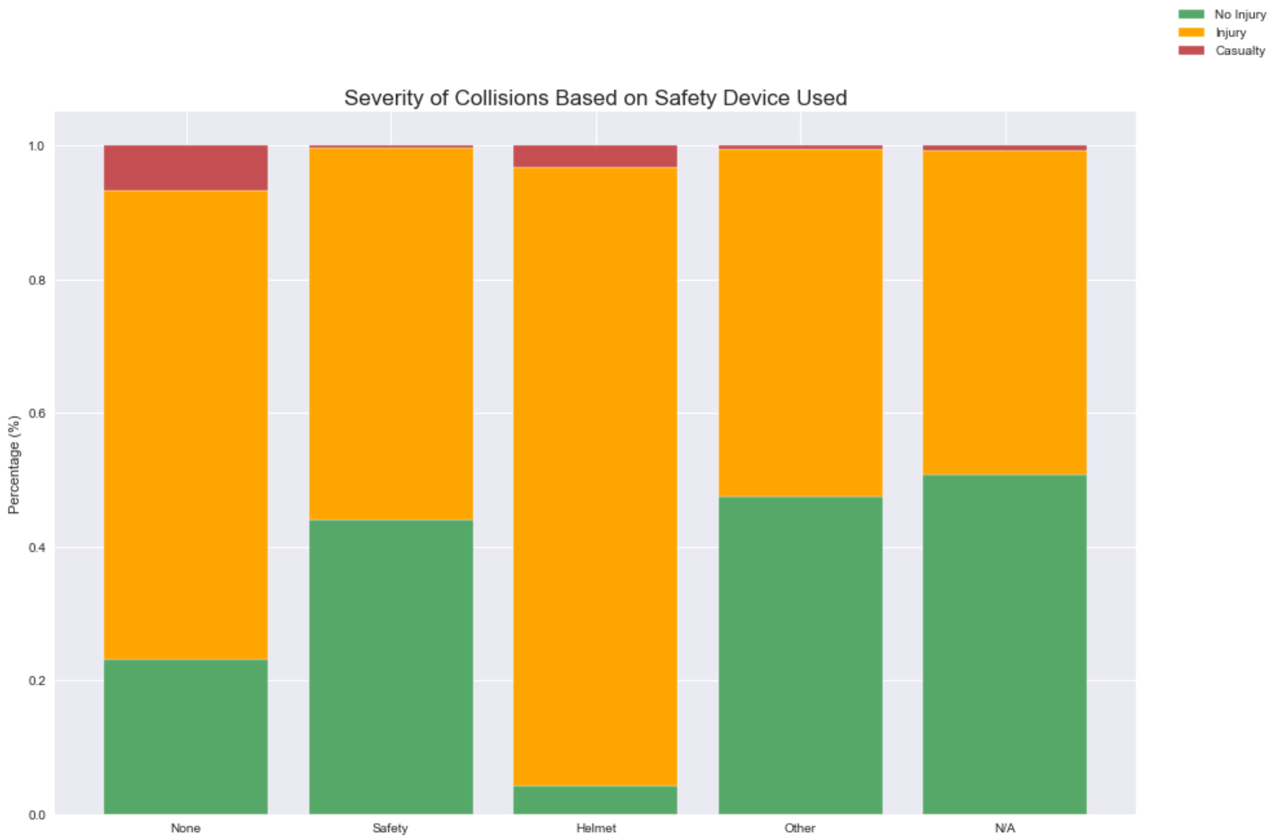

Figure 11

The proportion of collisions with no injuries (green bars) for those wearing a safety device is significantly higher compared to those without a safety device. Additionally, the usage of a safety device almost eliminates the proportion of collisions with casualty (red bars), i.e., **wearing a safety device whilst using your vehicle significantly reduces your chance of dying**. In fact, 6.8% of collisions without a safety device results in a casualty, compared to 0.4% for collisions with a safety device equipped.

Other interesting observations we can deduce from Figure 11:

- Vehicles requiring a helmet as a safety device recorded the highest proportion of collisions with injuries
- Vehicles that do not have a safety device equipped (e.g., buses) recorded the highest proportion of collisions with no injuries

To further analyze how a safety device can save a life, we take a look at Figure 12 below:

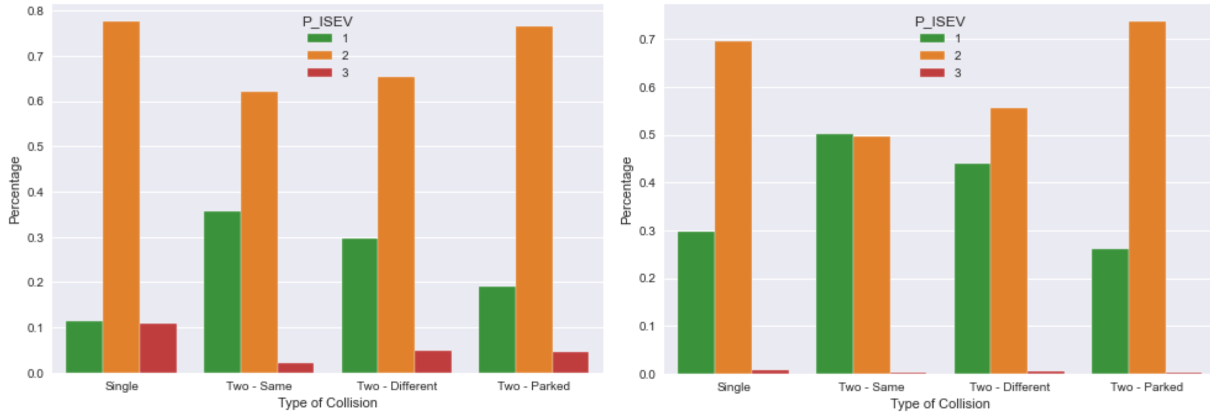

Figure 12

Figure 12 shows the distribution of collision severity based on the type of collision, with the left side showing the distribution for collisions without a safety device and the right side representing collisions with a safety device. Note that these bars represent the percentage of collisions and not the total recorded number. There are a few observations we can conclude from Figure 12:

- **The percentage of collisions with casualties (red bars) decrease to a non-existent level once a safety device is equipped**
- **The percentage of collisions with no injuries (green bars) increase by a significant margin once a safety device is equipped**. An interesting additional observation here is that collisions involving two vehicles travelling in the same direction with a safety device equipped recorded more non-injuries than injuries.
- The percentages of collisions with injury (orange bars) decrease across the board once a safety device is equipped

<u>Collision Location:</u>

The next few plots will be shown as bubble plots, where the size of the bubbles represents the total number of collisions and the axes represent 2 different variables relating to the accident.

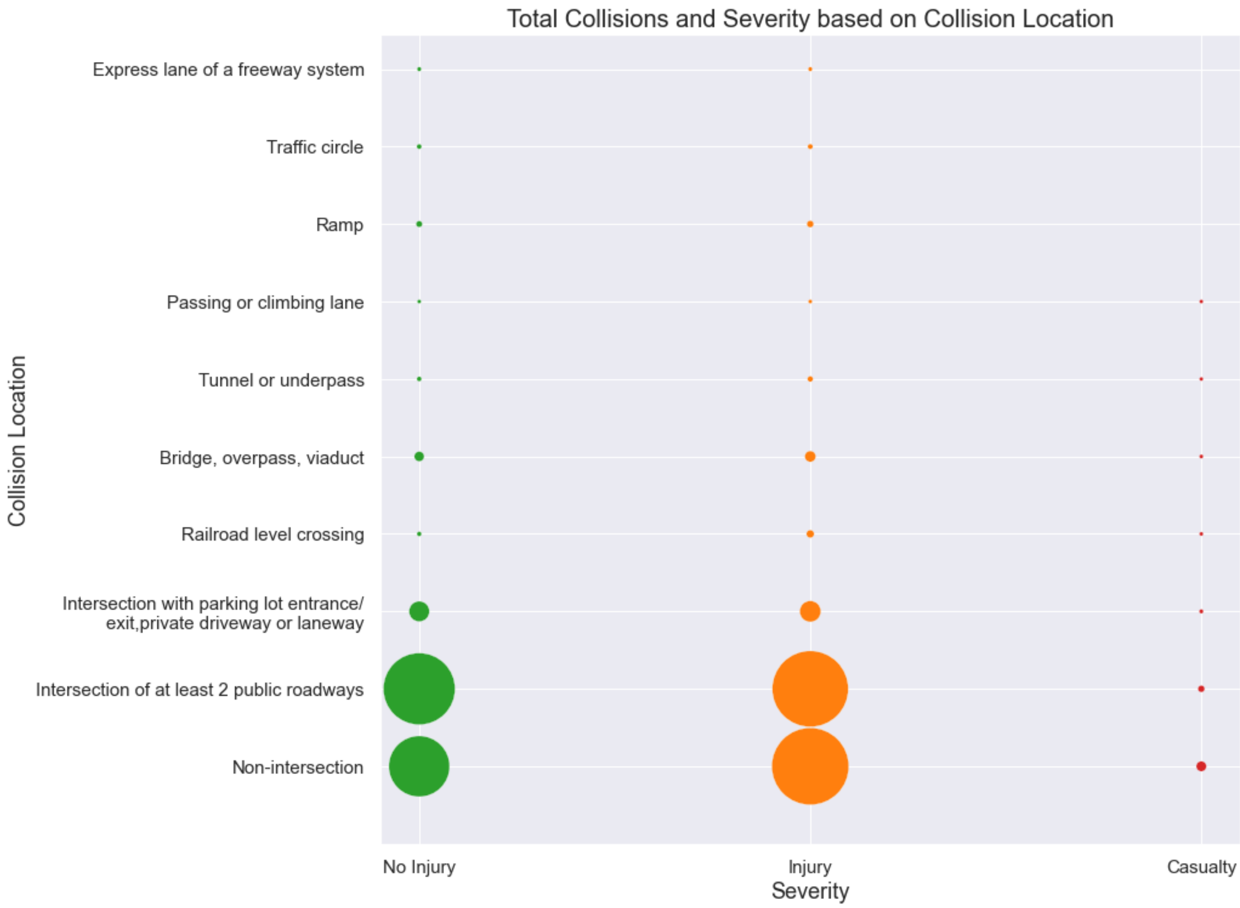

Figure 13

Figure 13 shows that **collisions happened in non-intersections tend to be more severe compared to collisions in an intersection**.

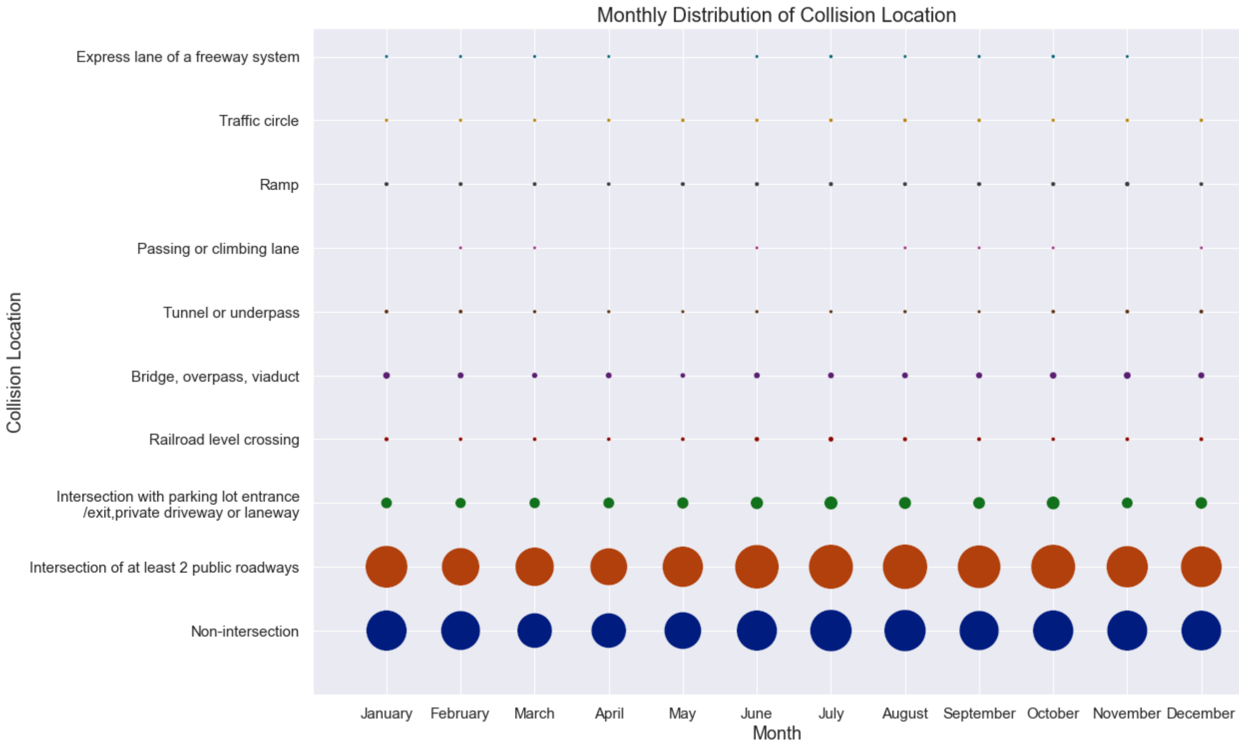

Figure 14a

Figure 14a shows the monthly distribution of collision location. There are no extra trends/patterns that stood out from this bubble plot other than most of the collisions happen at non-intersections and intersections of at least 2 public roadways. This bubble plot also shows the slight dip in the number of collisions during the Spring months which we saw in the earlier line graphs.

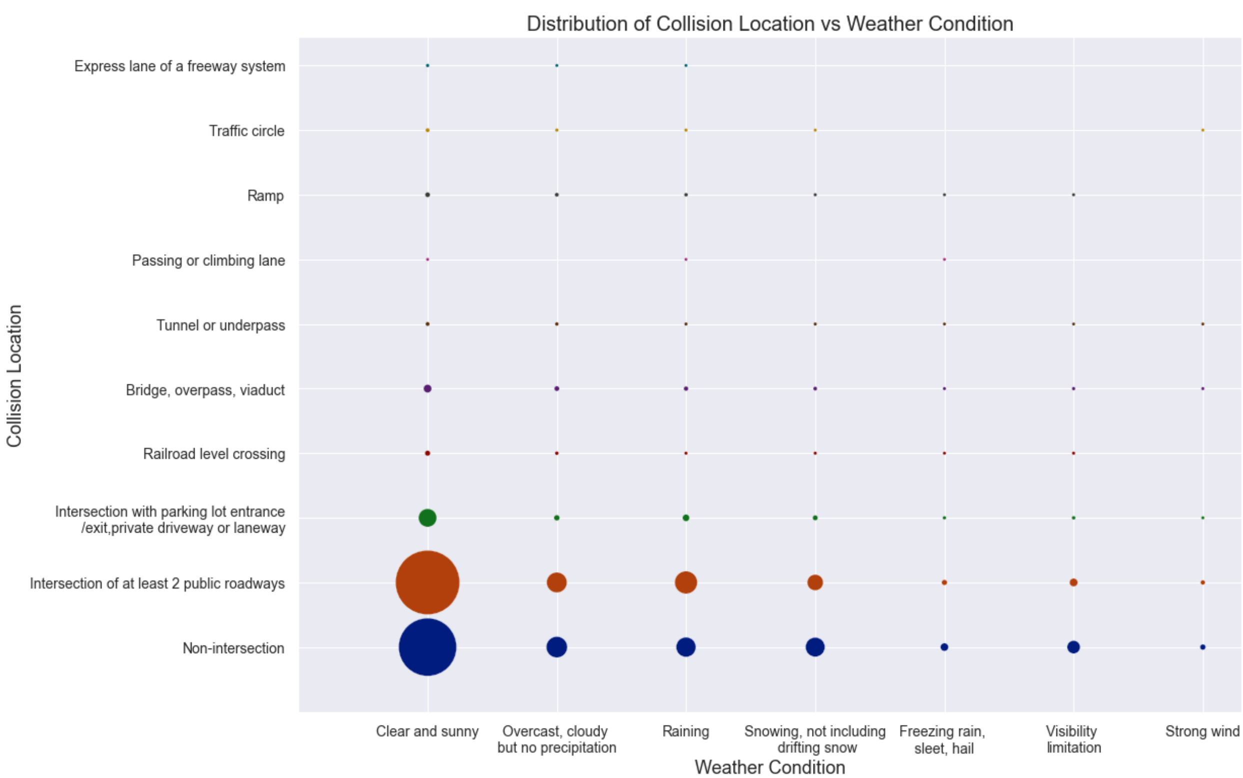

Figure 14b

Figure 14b further shows that most collisions happen on a sunny day, regardless of location.

<u>Surface Condition:</u>

Figure 15a below shows the distribution of collisions based on both road and weather conditions. This bubble plot reinforces our previous findings: that most collisions in 2019 happened on a sunny day and when the roads were dry.

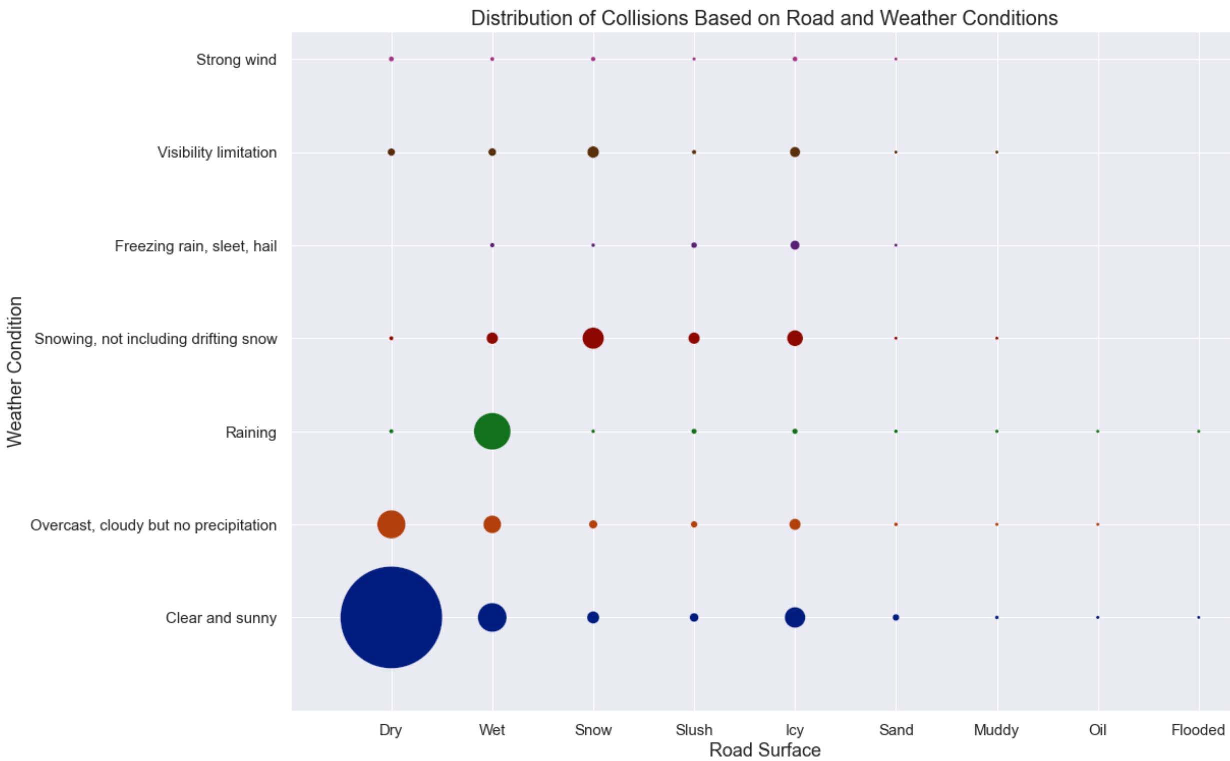

Figure 15a

Additionally, Figure 15b further shows that most collisions happened on a dry surface when the surface level is straight and flat.

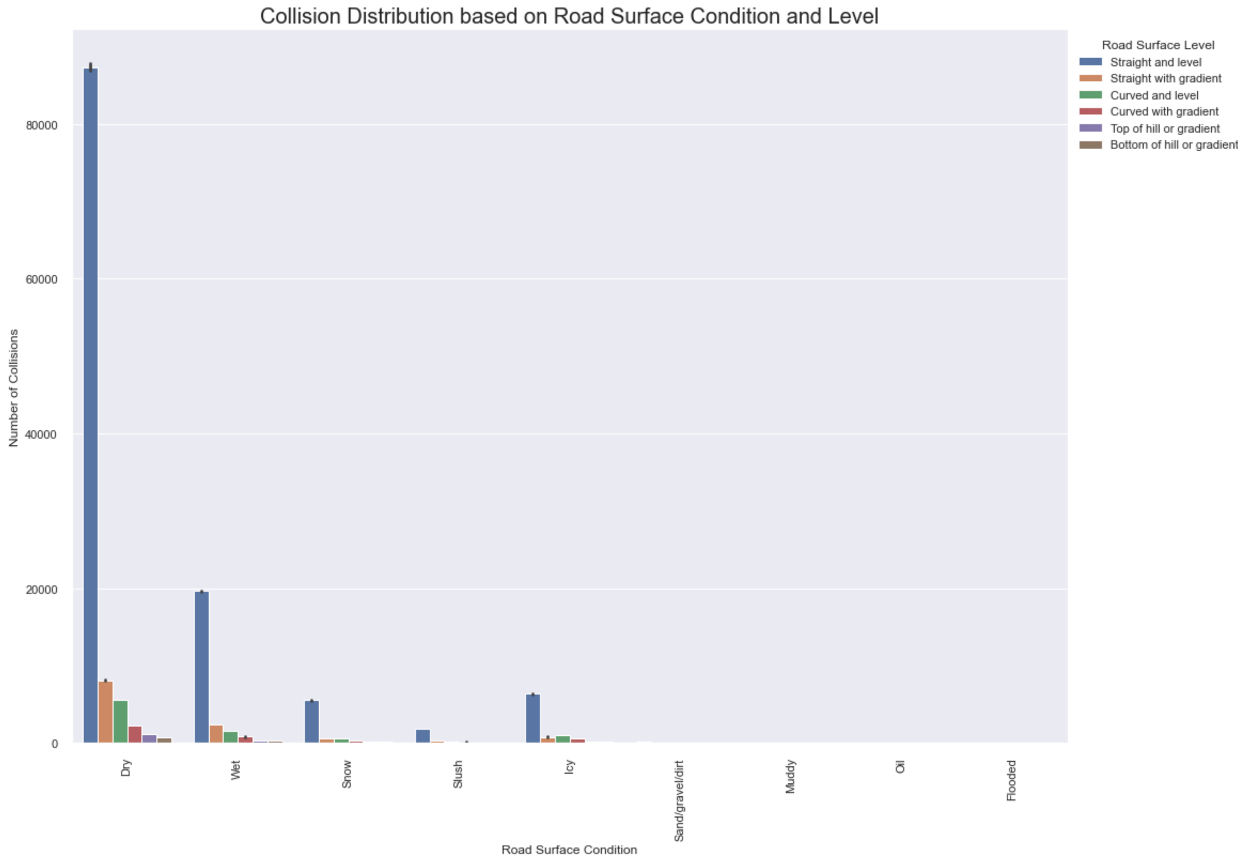

Figure 15b

<u>Traffic Sign:</u>

Figure 16 shows the distribution of collisions based on traffic sign and surface level. There are a few observations we can make from this graph:

1. Most collisions happened when there were no traffic signs present. **In fact, a significant number of collisions happened in a location where traffic signals were fully operational.**
2. Most collisions happened on a level surface
3. There is also a sizeable number of collisions recorded at stop signs

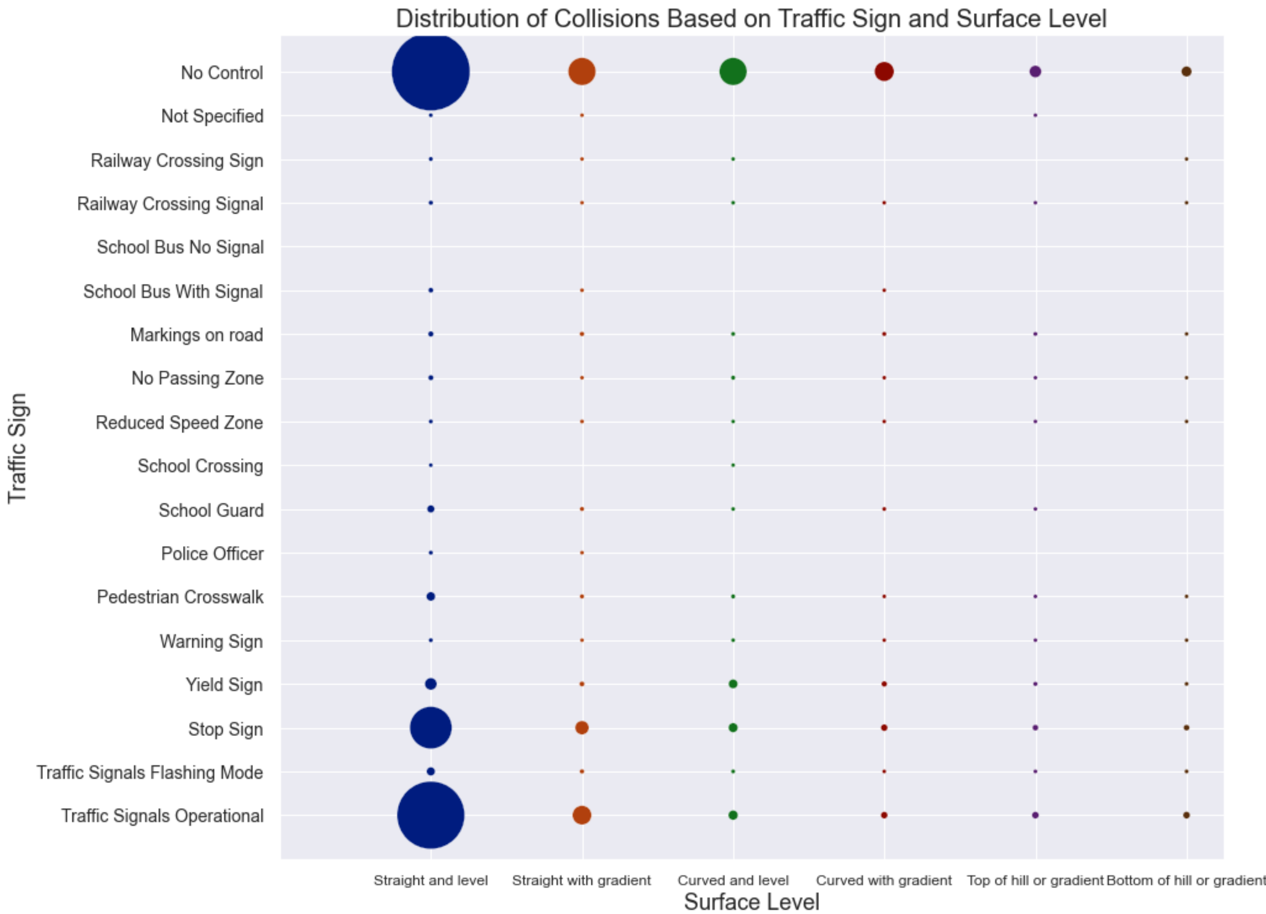

Figure 16

An interesting observation from Figure 17 is that a location without traffic controls results in the most collisions with injuries & casualties, which stresses the importance of traffic controls in road safety.

Figure 17

Conclusions:

This report provides valuable insights to the distribution and severity of collisions in 2019 that can be leveraged to increase road safety and empower decision makers to implement constructive public policies:

- Current initiatives to promote safety during the Winter months (i.e., snow clearing, driver education) are working well as collision severity for the Winter months is not significantly higher compared to the other months
- There is good evidence to implement laws subsidizing newer vehicles / apply penalties to older vehicles to decrease the overall number of casualties and increase passenger safety
- Road design must be such that it implements traffic signs where necessary and minimize the possibility of a collision involving a single vehicle
- There is evidence that driver behaviour is different between the sexes, which can potentially be addressed by driver education; this is evident based on insights derived from the data showing most collisions happen even on clear sunny days where the roads are dry and level
- Usage of safety devices is proven to be critical in reducing the severity of a collision
- Major improvements can be made in regards to traffic signals as collisions still happen even when signals are fully operational, the effectiveness of current traffic signals need to be addressed