Exploring Factors Associated with Road Crashes in New Zealand

Michael Saw

27 October 2025

Contents

1 Executive Summary

This report aims to investigate how lighting conditions and weather can influence the severity of road crashes in New Zealand between 2000 and 2024. Using the Crash Analysis System (CAS) data published from the Waka Kotahi NZ Transport Agency. In order to answer this question, binary logistic regression model was developed to estimate the probability of a crash being high severity based on factors in the dataset such as streetlight status, natural light, and weather conditions. The dataset has gone through a large amount of preparation and cleaning before running it through the selected model.

The analysis found that street light availability was the strongest selected predictor of crash severity. Crashes that occurred in areas with no street lights are highly likely to result in a serious or even fatal crash when compared to areas with functioning and present street lights. Similarly, crashes that occur during the night were about 45% more likely to be serious or fatal than those during daylight hours. Surprisingly, adverse weather conditions such as rain, fog or even snow was actually associated with lower crash severity, this may be due to driver behaviour adapting to hazardous conditions, resulting in fewer severe outcomes, it may also be due to a decrease in drivers on the road in harsher weather conditions.

In light of these findings, several recommendations may be proposed to the council:
1. Improve street lighting infrastructure
2. Implement public awareness campaigns
3. Enhance driver licensing and education programmes

Glossary of key terms
Variable : A characteristic or attribute that can take different values in a dataset (e.g., weather, light, streetLight).
Observation : A single recorded data point or row in your dataset — one crash event.
Binary logistic regression : A statistical model used to estimate the probability of one of two possible outcomes (e.g., High vs Low severity).
p-value : A number that shows whether an effect is statistically significant (usually < 0.05)
Predictor Variable (Independent Variable): A variable used to explain or predict the outcome variable. In this report, examples include weatherA, light, and streetLight.
Crash Severity : A measure describing the outcome of a crash, typically classified into categories such as Fatal, Serious, Minor, or Non-Injury.

2 Background

Road crashes remain as one of the most persistent safety issues in New Zealand. Every year, tens of thousands of crashes are reported to the New Zealand Police with around 300 of them being fatal deaths (of Transport, 2025). These crashes results public damage, injuries, fatalities and significant social and economics costs. There are many adding attributes that may influence a likelihood of a crash, an important factor that will be taken into consideration in this report is lighting and weather. Visibility and weather conditions are an important factor to pay attention to when driving.

The study aims to answer the question: **How lighting conditions and weather interact to influence the severity of road crashes in New Zealand between 2000 and 2024 ?** By analysing the presented dataset from the Waka Kotahi Crash Analysis System (CAS).(Agency, 2025). This study aims to identify whether crashes that occur in areas that may not have the best visibility or under adverse weather conditions will impact the result of a crash having fatal or serious casualties, and if the combination of these attributes amplifies crash severity.

The key aims of this analysis are to:
- Identify Patterns of Crash Severity
Examine how crash severity levels vary under different lighting and weather conditions across New Zealand to determine whether specific environmental settings are consistently linked with more severe outcomes.
- Investigate Interaction Effects Between Lighting and Weather
Assess whether poor lighting and adverse weather together will increase the likelihood of high-severity crashes beyond the effect of either factor alone.
- Generate Insights for Road-Safety Decision-Making:
Use the results of the analysis to develop and suggest practical recommendation for transport authorities and policymakers. These results can help prioritise improvements and additions in street lighting infrastructure and also help improve the quality of drivers on the road.

3 Data Description

The analysis in this research is based on the data from the Crash Analysis System dataset that is published by Waka Kotahi NZ Transport Agency. This dataset records all road crashes that have been reported to the New Zealand Police from 2000 to 2025, updating every month.(Agency, 2025).

Overall, the dataset contains a variety of variable types, including numerical count data (e.g., crashYear), categorical data (e.g., streetLight), and ordinal categorical data (e.g., crashSeverity), with a total of 72 variables and 885184 observations. This is considered a relatively large dataset that will be suitable for future analyse. However, only a subset of the data relevant to the research question is needed for study. The key variables selected for analysis is listed below.

3.1 Selected Variables in Detailed Analysis

- **CrashSeverityBinary (Target Variable)**
  This variable has been derived from the original crashSeverity variable. It is a categorical variable with two levels of High and Low. High, including Fatal + Serious, then low, including Minor + Non-injury. This transformation helps simplify interpretation from having 4 outcomes to 2, enabling the application of a binary logistic regression model.

- **streetLight - Variable**
  This is a categorical variable that describes the streetlight conditions at time of the crash location. It consists of 4 categories which are 'On' (Street lights are turned on), 'Off' (Street lights are turned off), 'None' (No Street lights in the area), and 'Unknown' (Unsure if street lights are in the area).

- **light - Variable**
  A categorical variable that states the light conditions at the time of the crash location. Consisting of 4 categories which are 'Bright Sun', 'Overcast', 'Twilight', and 'Dark'.

- **weatherA - Variable**
  A categorical variable that states the weather conditions at the time of the crash location. Consisting of 6 categories, which are 'Fine', 'Mist or Fog', 'Light rain', 'Heavy rain', 'Hail or Sleet', and 'Snow'.

3.2 Data Completeness

The provided dataset is generally well structured and reliable for analysis; however, it does contain a noticeable amount of missing data. For example, the variables 'crashRoadSideRoad' and 'intersection' are completely blank across all records. Additionally, some categorical variables, such as streetLight, weatherA and light, included 'Unknown' categories that may not provide meaningful information. These may need to be filtered out after further consideration. Missing or incomplete data in the dataset is important to look into, as it can introduce bias and lead to misleading outcomes if not handled appropriately.

4 Ethics, Privacy, and Security

4.1 Ethics

When reporting or publishing research findings based on the Crash Analysis System (CAS) dataset(Agency, 2025), it is essential to consider ethical considerations. Although the data contains no personal identifiers, the public reporting of crash findings can raise ethical concerns if misinterpreted. For example, if the results showed that there is higher crash severity in certain groups of vehicles or regions, it may lead to unfair generalisations and create stereotypical views on those certain groups.

Additionally, some aspects may have more observations recorded, such as the region. Auckland would naturally have higher crash counts due to its larger population and higher traffic volumes. If results were presented in raw counts, it may create a misleading impression that a certain region or group is inherently more dangerous. To avoid this bias, the analysis that will be done will use proportions and rates rather than raw counts when comparing crash severity across the factors being compared. By normalising the data in this way, it helps reduce the risk of reinforcing stereotypes or unfair interpretations.

Additionally, certain vehicle types, such as taxis, delivery vehicles and other work vehicles types are naturally exposed to greater driving hours and therefore a higher likelihood of being involved in crashes. If not considered, statistical results may lead to misinterpretation that these groups are more 'dangerous'. This could be mitigated in future analysis by calculating and measuring by proportions rates rather than raw counts, ensuring equal exposure.

While the Crash Analysis System dataset does not include iwi or ethnicity-specific variables, it is still important to recognise Māori data sovereignty principles.(Raraunga, 2024). These principles that are outlined by Te Mana Raraunga mentioned that Māori have inherit rights over data relating to Aotearoa and its people, and data should be managed in such way that uphold tikanga, kaitiakitanga, and the values of Te Tiriti o Waitangi. In this research, these principles are acknowledged through ensuring that results are interpreted responsibly and do not misrepresent or stigmatise communities or regions within New Zealand.(Raraunga, 2024).

4.2 Privacy

The Crash Analysis System (CAS) dataset is extensively and largely impersonal, with it continuing to be updated every month. Containing no personal identifiers such as name or license plate numbers. It does not include unique identifiers that could easily be linked with external datasets to identify specific individuals, and vehicle information in the data set has been limited to general vehicle types rather than registration details.Waka Kotahi NZ Transport Agency, 2025.

However, some risks of potential privacy breach have been identified, some variables in the dataset does contain variables that include street names and regional identifiers, which may pose a privacy risks if cross referenced with other publicly information that may be available on the web. For example, combining variables such as vehicle type, region and year may be able to allow a crash to be linked to a certain news article or incident on the web.

This research will not be eyed towards location and region, however if needed to minimise this risks, variables such as ”crashLocation1” and ”crashLocation2” may be excluded from the analysis, ensuring that while broader information such as region remain available for meaningful insights, individual crash events cannot be associated with identifiable locations. Given the dataset's large size and aggregated nature, the likelihood of backtracking and identifying should remain extremely low once these variables are excluded.

4.3 Security

Data security was an important consideration throughout this project. The dataset has been collected carefully and securely from an official trusted public source - Waka Kotahi — New Zealand Transport Agency, almost reducing the risk to zero of using working with unverified, corrupted or malicious files ensuring the integrity and authenticity of the data.

All files and scripts for this project have been stored locally on private protected devices. The raw dataset has been stored in a secure and private folder with access restricted to authorized users, helping to strengthen security and prevent unwanted modifications and changes to the dataset.

With further allocation of resources, a few recommendations has been raised to consider

1. Version control and access management
   By allocating further resources to a private github or repository would allow for secure version control, allowing transparency, traceability of code and documentation used for the analysis. Ensuring that changes to the data or data cleaning is recorded and having the ability to revert to previous versions of the dataset if need be.

2. Applying Encryption Backups

By allocating further resources to being able to maintaining regular encrypted backups of the dataset in a secure drive or cloud may help prevent data loss or such if there were an act of hardware failure or so.

3. Network and device security
   By allocating further resources to use a dedicated device solely for work without personal files or third party applications.This can help reduce the risks of malware and cyberattacks.In addition,ensuring the device is being connected on trusted, firewall protected networks to strengthen the overall security of the data.

5 Exploratory Data Analysis

5.1 Total Injury Crashes

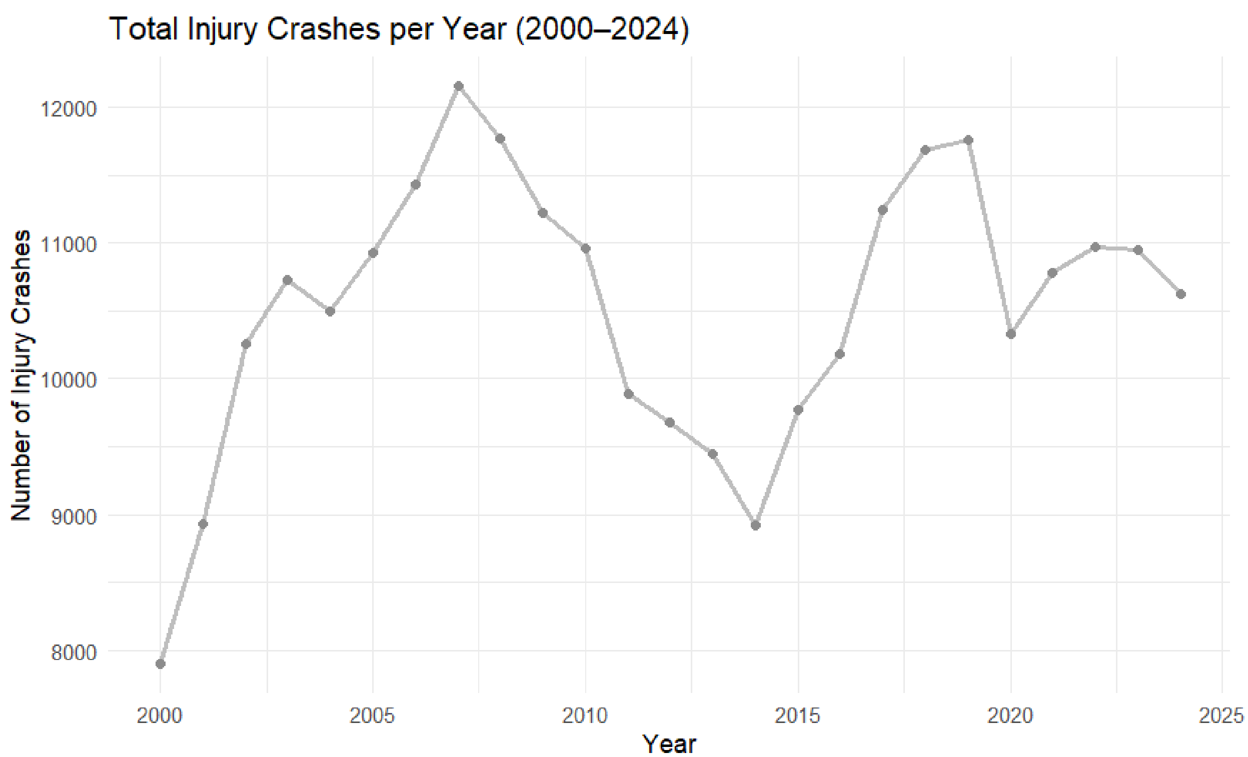

Figure 1: Total injury crashes per year in New Zealand (2000–2024)

In figure 1, we have a time series plot that shows how the total number of reported crashes that involves an injured person (fatal, serious, or minor) has changed between 2000 and 2024. On the X axis is the variable 'crashYear', which contains the year in which these crashes occurred. On the Y axis is the count of the variable 'crashSeverity' excluding recordings of crashes that did not have any injury.

From 2000 to around 2007 there has been a clear upward trend in crash counts peaking around 2007-2008. Followed by a steady decline until around 2014 almost being the lowest level in the period. From 2015 onward it seemed to steady rise again, reaching a second peak around 2019 till it declined again.

The fluctuations in injury crash counts from 2007-2014 may be a reflection of cars with better safety features being implemented into the market. However, from 2013, minor and serious injuries incline, until they drop in 2020. This makes sense as in 2020 the virus 'COVID-19' spread greatly across the entire world, causing lockdowns in many countries, which resulted in less vehicles on the road.

5.2 Crash Severity by Weather Conditions.

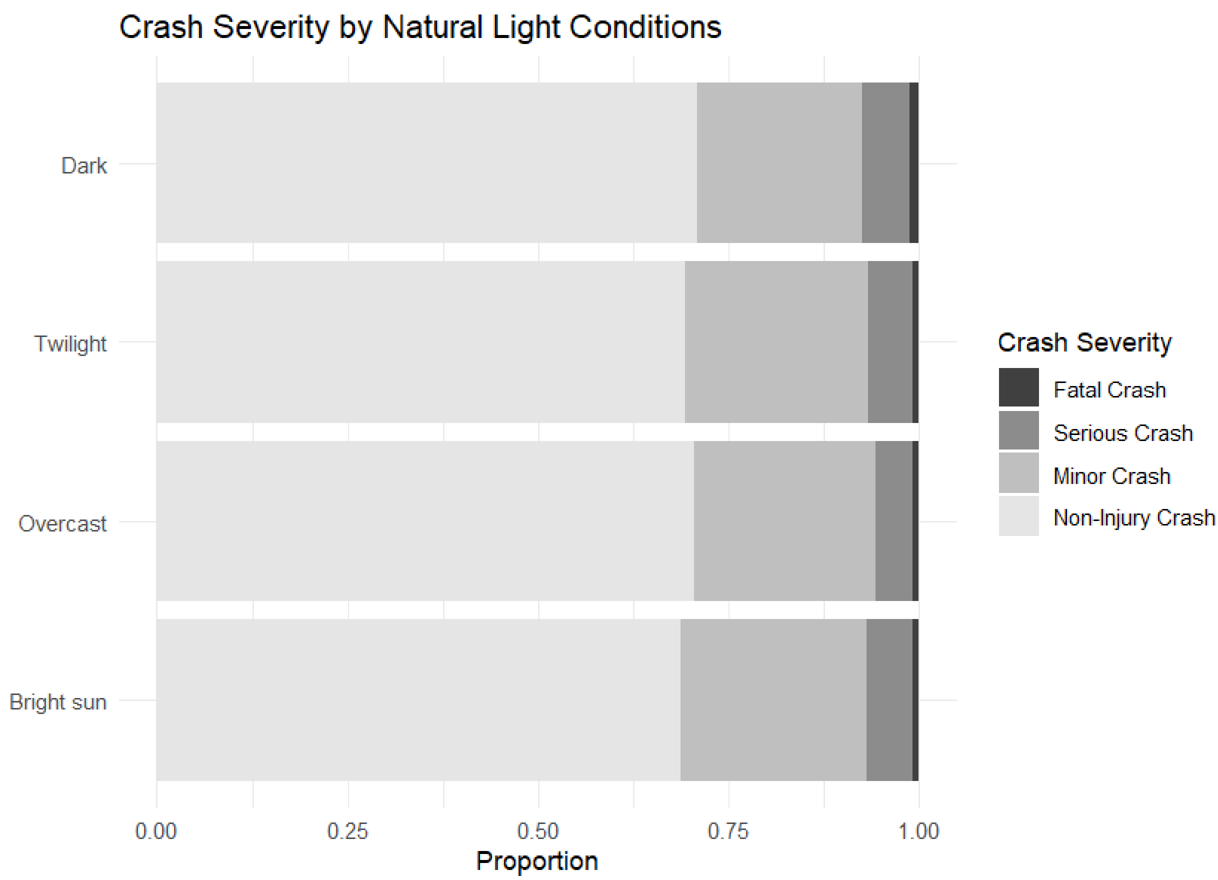

Figure 2: Crash Severity by Natural Light Conditions

Figure 2 presents a stacked bar plot displaying the proportions of crashes under different light conditions, these include - Bright Sun, Overcast, Twilight, and Dark. The X axis shows the proportion of 'crashSeverity', while the Y represents the variable 'Light', describing the natural lighting at the time of the crash. With the bars divided by crashSeverity, with categories ordered from Fatal Crash (darkest shade) to Non-Injury Crash (lightest shade).

When interpreting figure 2, it shows that the majority of crashes occur under bright or overcast daylight conditions, this makes sense as during the day there naturally will be more traffic exposure during the day. However, the proportion of serious and fatal crashes increase notably under low-light conditions, primarily during Twilight and Dark periods.

5.3 Urban vs Open Area crashes

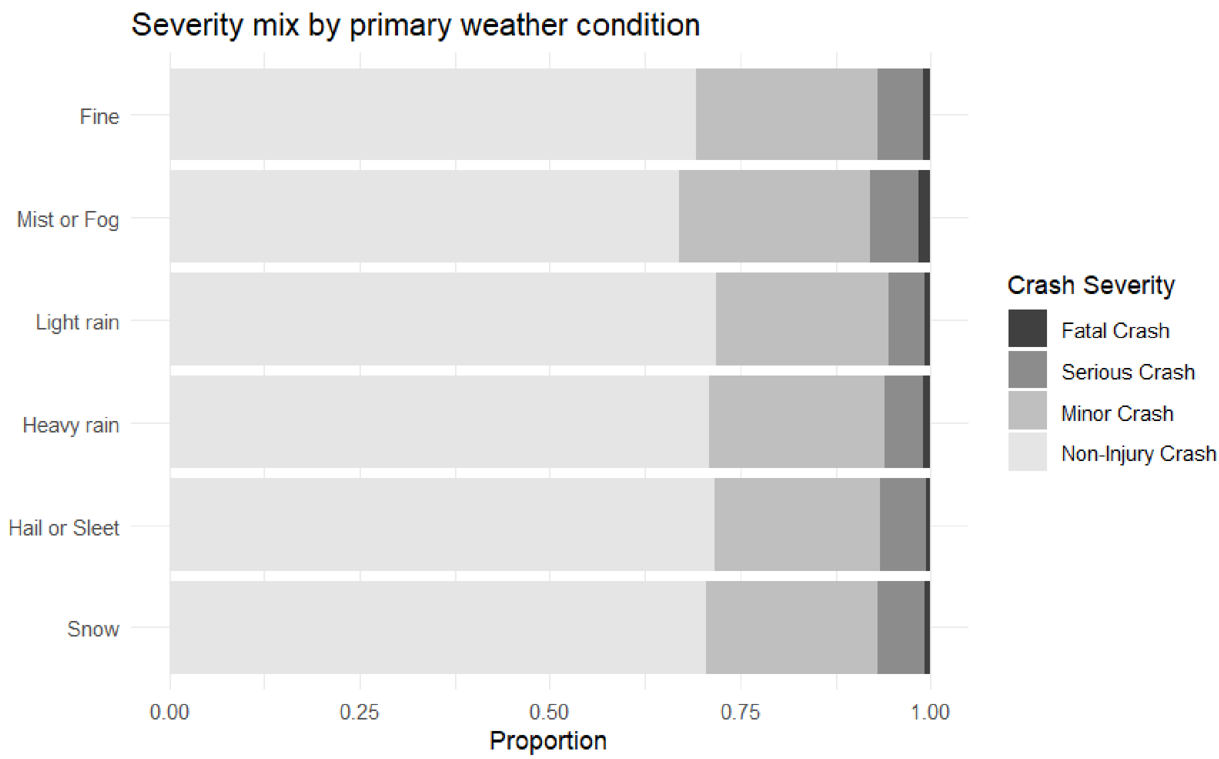

Figure 3: Crash Severity by Primary Weather Condition

Figure 3 contains the variable 'WeatherA', which contains the different weather conditions for each crash. The plot displays the severity of crashes in different weather conditions. The stacked bar chart displays the proportional distribution of crash severity levels within each category.

The distribution of crash severity seems to appear relatively consistent across the weather conditions, although serious crashes have a slightly higher proportion during snow, mist/fog, and hail/sleet, although it matches the proportion of serious crashes in fine conditions. Fatal crashes seem to have a slightly higher proportion in mist/fog and fine conditions. Although the plot presents proportions rather than raw counts, Fine weather occurs far more frequently overall. Even though its proportional severity is similar to other conditions the total number of crashes under fine conditions remain much higher due to greater exposure.

5.4 Weather and Light Crash Conditions

Street Light Condition	Low	High	Total	High (%)
On	181707	10681	192388	5.6
Off	236930	11181	248111	4.5
None	126129	16989	143118	11.9
Unknown	281194	20373	301567	6.8

Table 1: Crash Severity by Street Light Condition

Table 1 summarises the relationship between 'streetLight' conditions at the time of the crash and the level of crash severity. In this table, crashSeverity have been split into a binary group of 'High' and 'Low', with Fatal and Serious crashes in the 'High' category, and 'Non-injury and 'Minor' in the 'Low' category. The results show that crashes with no street lights in the area had the highest proportion of severe outcomes 11.9%. Followed by crashes with unknown of 6.8%, while areas with street lights on had the lowest proportion of severe crashes of 5.6%.The lower proportion of high-severity crashes in lit areas suggests that adequate street lighting may help mitigate the risk or severity of nighttime crashes by improving driver visibility.

6 Detailed Analysis

6.1 Data Manipulation

To investigate the research question "How lighting conditions and weather interact to influence the severity of road crashes in New Zealand between 2000 and 2024?".Before being able to do any sort of detailed analysis, first we must fully prepare our dataset to ensure that no problems occur in the future.

While the original Crash Analysis contained a generous amount of variables to work with, only a few variables were needed to answer the question. Selecting 4 variables 'CrashSeverityBinary', 'weatherA', 'light' and 'streetLight'. However, these 4 had to be manipulated first to fit our analysis. As stated earlier, the target variable originally was 'crashSeverity', containing 4 category outcomes but has been reclassified into two broader categories of High severity and Low severity. weatherA, light, and streetLight has all been simplified and recoded in order to make the analysis more interpretable and statistically stable. weatherA was collapsed into four broader categories which are 'Snow/Hail' (Snow, Hail or Sleet), 'Rain' (Light Rain, Heavy Rain), 'LowVisibility' (Mist or Fog), and 'Fine'. Whereas light has simplified into two categories which are 'Daylight' (Bright sun, Overcast), and 'Night' (Twilight, Dark).While streetlight stayed similar containing three category outcomes of 'On', 'Off' and 'None'. The dataset has also been trimmed to show records from 2000 to 2024, as it may lead to false findings to include 2025 as the year has yet to conclude.

To handle missing and unknown data, all entries that were recorded as unknown have been dropped from the dataset. After removing those entries and using a smaller subset of our original data, about 35.4% of total rows has been removed from removing the unknown values. This allowed the dataset to be finally completed and ready for analysis

6.2 Model

A binary logistic regression model has been selected, this approach was selected due to the manipulation of our selected Target variable.(for Digital Research Education, 2025). The model is specifically designed to estimate the probability of a high severity crash based on the predictor variables (weatherA, streetLight, light). The logistic regression models the log-odds of an event occurring rather than modeling the event directly, allowing the prediction to remain within the valid probability range of 0 to 1.

Variable	**Estimate**	**Std. Error**	**z value**	**Pr(> $\lvert z \rvert$)**
(Intercept)	-3.1309	0.0805	-38.905	<2e-16
streetLightOff	0.0804	0.0811	0.992	0.3212
streetLightNone	1.1436	0.0813	14.062	<2e-16
lightNight	0.3721	0.0810	4.594	4.35e-06
weatherALowVisibility	-0.1387	0.0379	-3.655	0.00026
weatherARain	-0.2949	0.0140	-21.059	<2e-16
weatherASnow/Hail	-0.4462	0.1273	-3.505	0.00046
streetLightOff:lightNight	-0.0560	0.0873	-0.642	0.5211
streetLightNone:lightNight	-0.3026	0.0827	-3.658	0.00025

Table 2: Logistic Regression Coefficient Estimates

Hypothesis test :

- Null Hypothesis (H0) :
  There is no significant relationship between lighting conditions, streetlight status, or weather conditions and the likelihood of a crash being high severity.

- Alternate Hypothesis (H1) :
  At least one of the predictors (lighting condition, streetlight status, or weather condition) has a significant effect on the probability of a high-severity crash.

After stating our Hypothesis, it can be observed that the Intercept has an estimate of -3.1309. It is the baseline log-rate of crash severity when the selected predictors are at their respective reference level which are when the street light is 'On', when light is 'Day', and weatherA is 'Fine'. The log odds is able to be referred in the appendices in table 4

streetLightOff:
The estimate of this variable is 0.0804 with a log odds of 1.08, this means that while holding all other variables constant, the odds of having a high severity crash is 8% more likely if there are street lights but it is off compared to our reference level. The p value is higher then the 5% significance level, therefore the Null hypothesis cannot be rejected suggesting that having street lights off does not have a significance effect on the probability of a high severity crash.

streetLightNone:
The estimate of this variable is 1.1436 with a log odds of 3.14, this means that while holding all other variables constant, the odds of having a high severity crash is 214% more likely if there is no street lights compared to our reference level. The p value for this variable is lower then the 5% significance level, therefore the Null hypothesis is rejected and that having no street lights has a significance effect on the probability of a high severity crash.

lightNight:
The estimate of this variable is 0.3721 with a log odds of 1.45, this means that while holding all other variables constant, the odds of having a high severity crash is 45% more likely during night time compared to daytime. The p value is lower then the 5% significance level, therefore the null hypothesis is rejected indicating that light conditions has a significance effect on the probability of a high severity crash.

weatherALowVisibility:
The estimate of this variable is -0.1387 with a log odds of 0.87, this means that while holding all other variables constant, the odds of having a high severity crash is 13% lower in low visibility conditions compared to base reference. The p value is lower than the significance level, therefore the null hypothesis is rejected indicating that low visibility conditions have a significance effect on the probability of a high severity crash.

weatherARain:
The estimate of this variable is -0.2949 with a log odds of 0.74, this means that while holding all other variables constant, the oods of having a high severity crash is 26% lower during any rainy conditions compared to base references. The p value is lower than the significance level, therefore the null hypothesis is rejected indicating that rainy conditions have a significance effect on the probability of a high severity crash.

weatherASnow/Hail:
The estimate of this variable is -0.4462 with a log odds of 0.64, this means that while holding all other variables constant, the odds of having a high severity crash is 36% lower during snowy or hail condition compared to the base reference. The p value is lower than the significance level, therefore the null hypothesis is rejected indicating that snow or hail weather has a significance effect on the probability of a high severity crash.

streetLightOff:lightNight:
The estimate of this variable is -0.0560 with a log odds of 0.95, this means that while holding all other variables constant, the odds of having a high severity crash is 5% lower when it is night time and the street lights are off the base reference. The p value is greater than the significance level, therefore the null hypothesis is accepted indicating that there is no significance effect on the probability of a high severity crash.

streetLightNone:lightNight:
The estimate of the variable is -0.3026 with a log odds of 0.74, this means that while holding all other variables constant, the odds of having a high severity crash is 26% lower when there are no street lights and when it is night time compared to the base reference. The p value is lower than the significance level, therefore the null hypothesis is rejected indicating that there is a significant effect on the probability of a high severity crash.

Referring to figure 4 in the appendices illustrate that the probability of a crash being severe increases as lighting conditions worsen. Predicting that crashes occuring in areas with no street lights is almost twice as likely compared to streets with working street lights.

In figure 5, the ROC illustrates how well the model distinguishes between High severity and Low severity crashes. The AUC displayed of 0.623, indicating that it is better then random, indicating that while lighting and weather conditions do have influence on crash severity, some other factors may also play a role.

6.3 Estimates Of Uncertainty

Variable	**Lower 95%**	**Upper 95%**
(Intercept)	0.0372	0.0510
streetLightOff	0.9279	1.2753
streetLightNone	2.6851	3.6945
lightNight	1.2423	1.7070
weatherALowVisibility	0.8076	0.9371
weatherARain	0.7244	0.7653
weatherASnow/Hail	0.4942	0.8146
streetLightOff:lightNight	0.7941	1.1184
streetLightNone:lightNight	0.6259	0.8659

Table 3: Exponentiated 95% Confidence Intervals (Odds Ratios) for Logistic Regression Model

In order to calculate uncertainty of our model predictors, we use a 95% confidence interval in order to measure our uncertainty. This table can be interpreted by first observing the lower 95% and upper 95% values of each variable and calculating the differences between them. If the differences is considered large, the confidence in the prediction would be less than if it is small.

For example :
weatherARain has a small interval width 0.7653 – 0.7244 = 0.0409). This narrow range shows a high degree of certainty, showing that the effects of crash severity is estimated in great confidence. weatherASnow/Hail has a much wider interval width of 0.8146 – 0.4942 = 0.320. Reflecting a great amount of uncertainty potentially due to fewer observations under snowy or hail conditions. Thus, predictors such as weatherARain is considered statistically stable and precise, whereas predictors with wider intervals such as weatherASnow/Hail carry a higher amount of uncertainty in their estimated impact on crash severity.

6.4 Potential Bias

While the selected logistic regression model most definitely did provide meaningful insights into how lighting and weather influences crash severity, some identifiers of several potential bias and limitations have been acknowledged when interpreting the results.
To start off, the data set is imbalanced, with a noticeably larger number of recorded observations of low severity crashes compared to high severity, while it is obvious that the chances of a high crash severity should be lower. The imbalance was quite large of around 38000 records of High crash severity compared to about 534000 of low crash severity. This imbalance may cause a model to favour the majority class and underestimate the probability of a much rarer high severity crash. Although the model is generally robust, imbalance still occurs and may effect results and probability of outcomes.

Secondly, it was stated earlier that around 35.4% of the observations has been removed due to unnecessary reports of unknown values. This cleaning was necessary but may also introduce to missing-data bias, missingness is unlikely to be random, since the missingness is unlikely to be random. For example, crashes occurring under complex or severe conditions may be less consistently reported, leading to under-representation of those circumstances compared to the large number of crashes reported in fine weather. In addition, the distribution of weather condition in the model is heavily uneven : fine weather accounts for about 77% of all crashes, whereas Snow/Hail conditions represent less than 0.2Ṫhis imbalance leads to exposure bias, where coefficients for rare weather categories are estimated from very few observations inflating uncertainty and widening the width in confidence intervals, while the model may suggest that snow or hail reduces crash severity, it may not be reflect to reality due to a small sample size to work with.

7 Conclusion

To reflect back to the main question of this project ""How lighting conditions and weather interact to influence the severity of road crashes in New Zealand between 2000 and 2024 ?". A binary logistic regression model was used with our desired predictors to answer this.
The results from the model showed that lighting conditions were a major contributing factor to crash severity. Crashes that occurred in areas with no present streetlights were much more likely to result in a fatal or serious crash in comparison to areas with streetlights present and functioning. Likewise, crashes that occur during the night is 45% more likely to be fatal or serious when compared to those in the day. Suggesting that light in general plays a big contributing factor in crash severity. However, the interaction between night time and street lights being off or absent resulted in a lower chance of being severe than expected, this may be drivers may be adapting to their environment and being more cautious or a lack of sample size.

In contrast, the model found that crashes in fine weather were more likely to be involved in a serious or fatal crash when compared to crashes occurring in rain, fog or snow. Although this may seem odd, it also make sense due to driver behaviour and also traffic exposure affected from weather rate. During adverse weather, drivers are typically more likely to adjust speed accordingly, reducing speed, increasing distance between others and being more cautious. adverse weather may also lead to less drivers on the road, resulting in lesser fatal/serious accidents, Whereas fine weather drivers may be more comfortable and make more mistakes. However, it has to be taken to account that the results of this may be due to data bias.

7.1 Potential Recommendations

Based on these findings, here are several recommendations that could be applied :

1. Improve street lighting infrastructure:
   Allocate extra resources to lighting infrastructure, covering and illuminating more streets in New Zealand with street lights.
2. Implement public awareness campaigns :
   Launch campaigns and awareness ads in New Zealand to remind drivers to drive safe and manage their speed and their attention level on the road.
3. Enhance driver licensing and education programmes:
   Strengthening licensing process to future drivers on the road, improve training in order to help future drivers be more prepared and qualified on the road.

7.2 Limitations

The findings of this study may not fully reflect real world driving conditions, it does not account for factors such as behaviour, traffic density and others which may influence the environment of a crash, additional, the dataset lacks in observations and reports of rarer weather conditions which may lead to data imbalance and bias.

References

Agency, W. K. N. T. (2025). *Crash analysis system (cas) data*. https://catalogue.data.govt.nz/dataset/crash-analysis-system-cas-data5

for Digital Research Education, U. I. (2025). *Logit regression — r data analysis examples* [Accessed 27 October 2025]. https://stats.oarc.ucla.edu/r/dae/logit-regression/

of Transport, M. (2025). *Crash and injury data — annual statistics* [Accessed 27 October 2025]. https://www.transport.govt.nz/statistics-and-insights/safety-annual-statistics/crash-and-injury-data/

Raraunga, K. (2024). *Māori data governance*. https://www.kahuiraraunga.io/maoridatagovernance

Waka Kotahi NZ Transport Agency. (2025). Crash analysis system (cas) data field descriptions [Accessed October 2025].

Appendices

Variable	Odds Ratio (exp(coef))
(Intercept)	0.0437
streetLightOff	1.0837
streetLightNone	3.1379
lightNight	1.4507
weatherALowVisibility	0.8705
weatherARain	0.7446
weatherASnow/Hail	0.6400
streetLightOff:lightNight	0.9455
streetLightNone:lightNight	0.7389

Table 4: Exponentiated Coefficients (Odds Ratios) from Logistic Regression Model

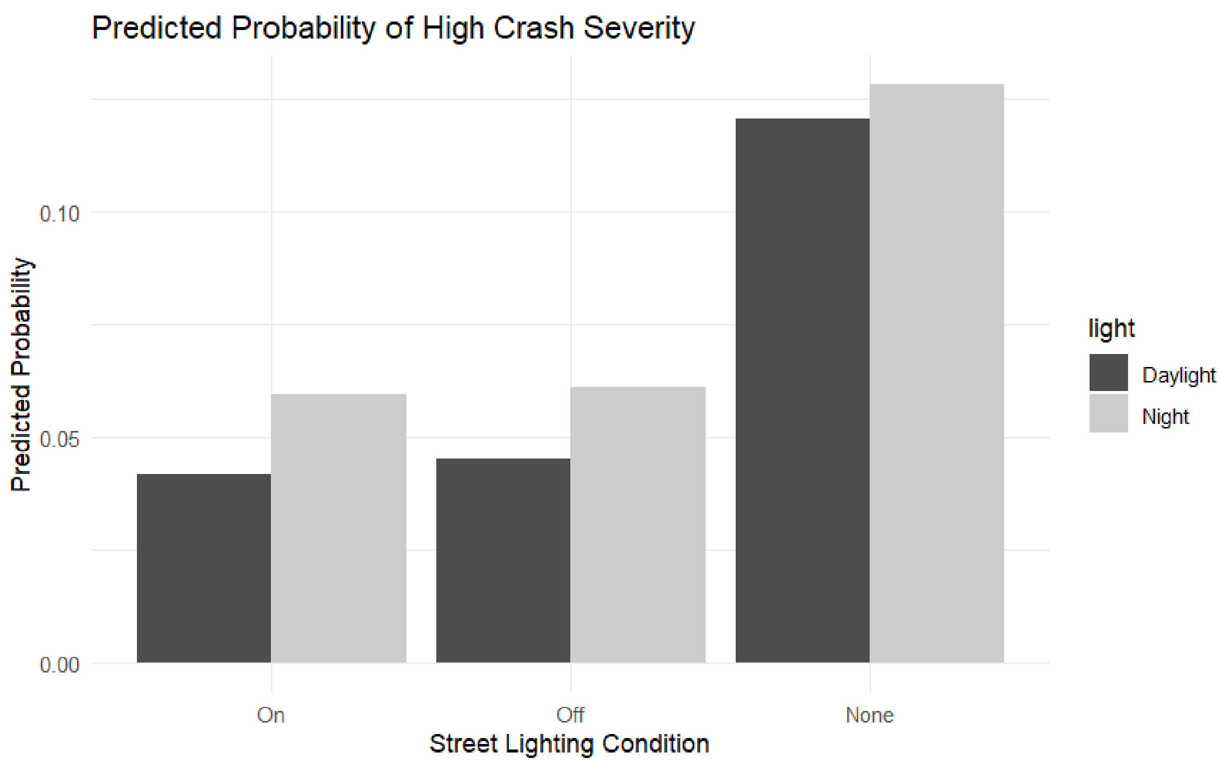

Figure 4: Crash Severity by Primary Weather Condition

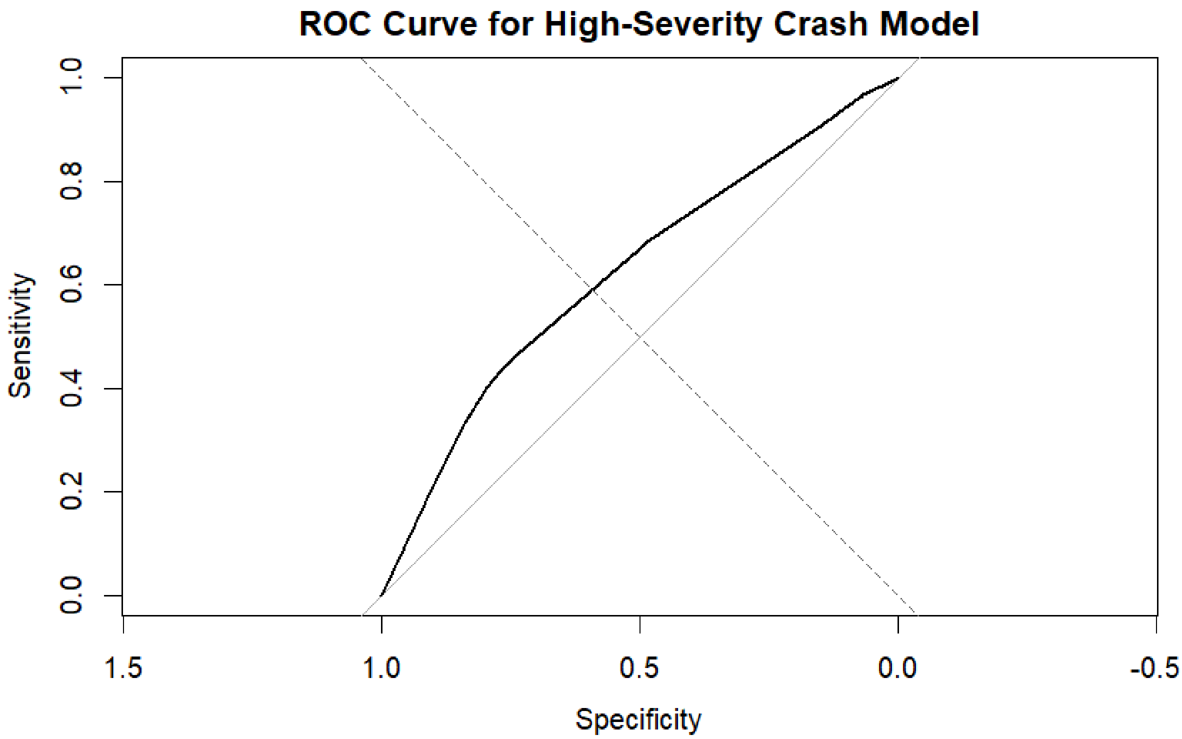

Figure 5: Area under the curve: 0.6229