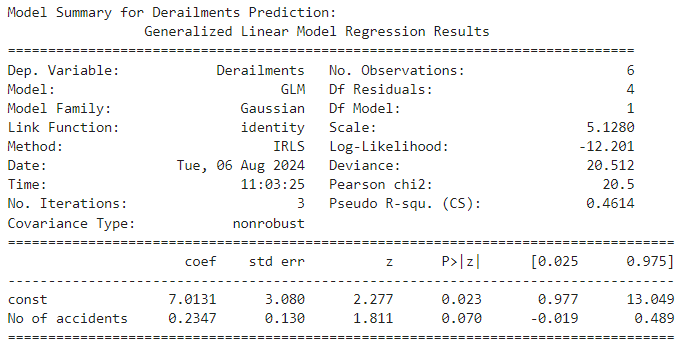
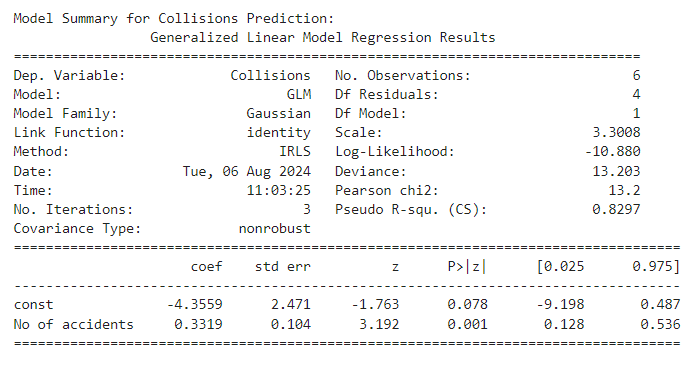
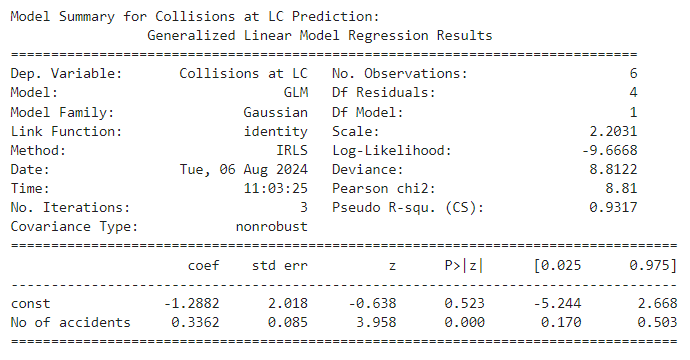
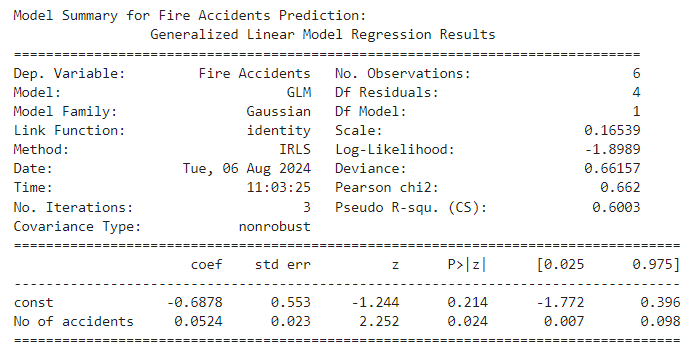
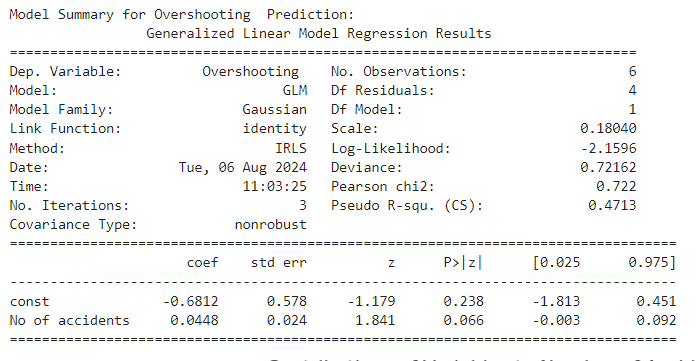
# Accident Type Analysis











Interpretation of the Models

Model for Derailments Prediction

From the summary output of GLM, the coefficient for the No of accidents 0.2347, and it comes with a standard error of 0.130. Though the coefficient is positive, the p-value of 0.070 is very close to the normal significance level of

0.05. This also reflects the pattern where the number of accidents reported more, the more the number of derailments, but at the

0.05 level the relationship is not statistically significant. Pseudo R-squared value is 0.4614 which reflects the adequacy of model fit for data.

Model for Predicting Collisions

The coefficient of No of accidents in this model for collisions is 0.3319, with a standard error of 0.104. The coefficient is again significant because the p-value is 0.001 for the number of accidents; thus, it significantly predicts the number of collisions. Additionally, the obtained Pseudo R-squared equals 0.8297, which indicates a good fit of the model. The above reveals a strong positive relation such that for a rise in the number of accidents, collisions increase.

Prediction Model for Collisions at LC

No of accidents – This coefficient, it is noted, for the prediction model that has been created is equal to 0.3362 with a std. error of 0.085. It is a very significant coefficient, whereby the resulting p-value stands at 0.000. The Pseudo R-squared stands at 0.9317, close to 1 and as such, indicates the model has a very high fit. This clearly shows that the number of accidents can be a highly influential predictor of collisions at level crossings.

Model for Prediction of Fire Accidents

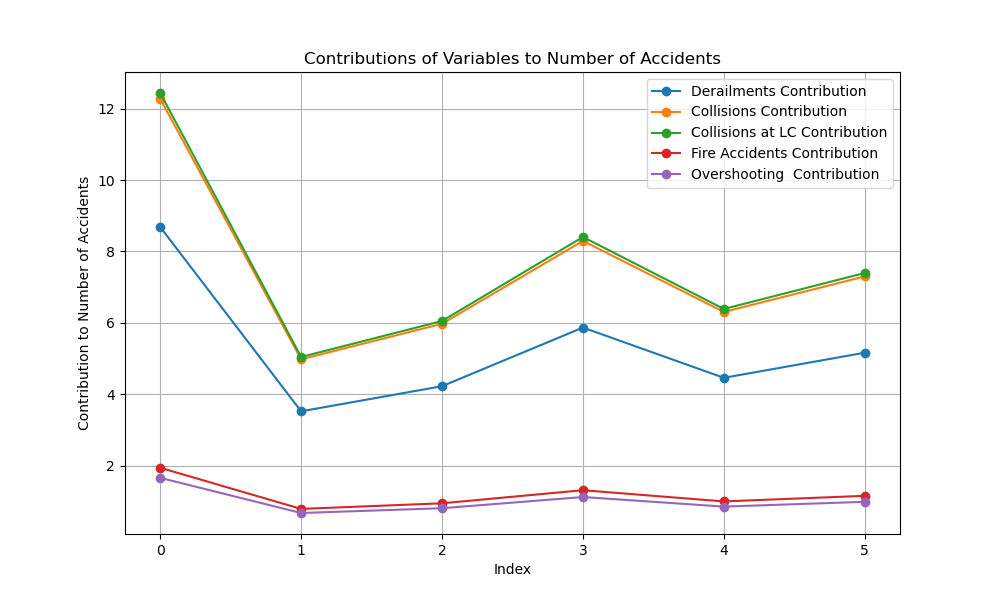
The coefficient, in the case of No of accidents, the model for fire accidents prediction is 0.0524 with a standard error of 0.023. The p-value is 0.024, which at the 0.05 level is significant. This is depicted by the Pseudo R-squared value of 0.6003, indicating that the model fits well. This indicates a positive but relatively small relationship between the number of accidents and fire accidents.

Model for Prediction of Overshooting

In the overshooting predictability model, the coefficient for No of accidents is equal to 0.0448 with a standard error equal to 0.024. The p-value of 0.066 is slightly higher than 0.05, indicating a weak trend that is not statistically significant at the 0.05 level. The Pseudo R-squared value is 0.4713, showing moderate fit in the model.

Interpretation of Graphs

The graph shows what contributions the number of accidents has for each type of accident. What contributions the number of accidents has for derailments, collisions, collisions at level crossings, fire accidents, and overshooting.



The contributions are plotted for each variable to show how changes in the number of accidents impact the number of each type of accident respectively. The graph is useful in knowing which connection between the number of accidents and the more heavily estimated accident type is by model coefficients.

Summarized:

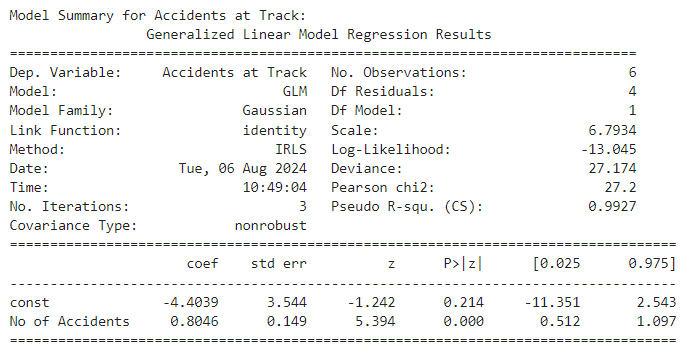
Derailments are positively, but just marginally significantly—associated with the number of accidents.

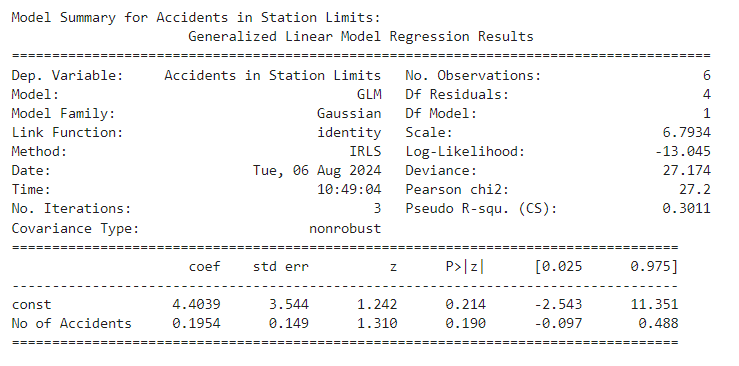
Collisions and Collisions at LC have significantly positive relation with the count of accidents.

Fire Accidents do have a significant positive relationship with the count of accidents but of lower strength.

Overshooting, on the contrary has no considerable correlation ; the p-value was found to be non significant at the 0.05 level.

# Accident Location Analysis





Model Summary and Interpretation

Model for Accidents at Track

The Generalized Linear Model GLM fitted using the normal distribution to predict the number of Accidents at Track based on the independent variable No of Accidents provides the following insights. The intercept term is -4.4039. Although this coefficient is not significant, the p-value is 0.214. This represents the baseline number of accidents at the track when the total number of accidents is zero. The coefficient for No of Accidents is 0.8046, which is statistically significant with a p-value less than 0.001. This means that for each additional total accident, there will be an expected increase of approximately 0.80 accidents at the track. This positive relationship suggests that the more the total number of accidents, the more the accidents at the track, proportionately.

Model for Accidents in Station Limits

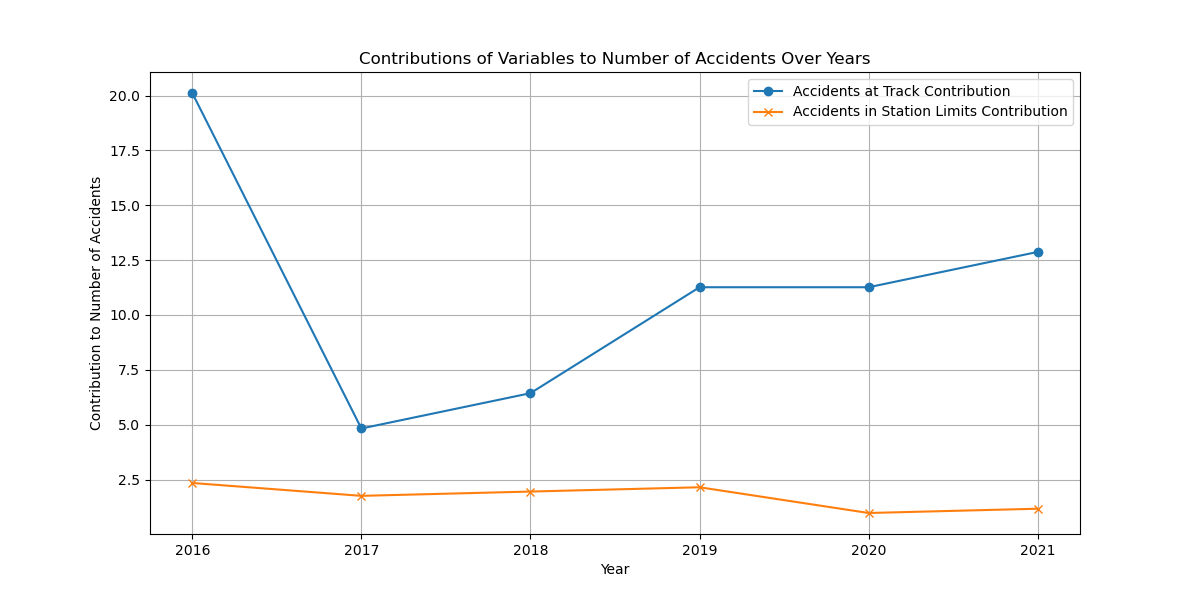
The GLM for Accidents in Station Limits uses the normal distribution. The intercept term is 4.4039, it is not statistically significant as the p-value is 0.214. This indicates that there are baseline incidents within the station limits, which are not dependent on the overall accident count. The coefficient, in this case, for No of Accidents is 0.1954. At the p-value of 0.190, the coefficient is not significant, and hence there is no clear evidence to establish the direct relationship between total accidents and accidents within the limits of the station.

Model Metrics

For both models, several key metrics are computed. For both models, the deviance is 27.174, which provides information about the fit of a model. The AIC value for both models is 30.09, which is used to compare models. The Pseudo R-squared of the Accidents at Track model is 0.9927, showing really extreme goodness of fit. The Pseudo R-squared for the Accidents in Station Limits model is 0.3011, indicating that this does not fit as well.

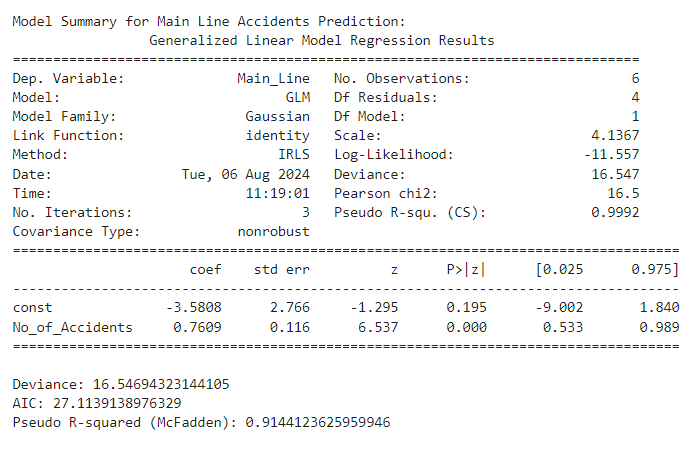
Interpretation of Graph

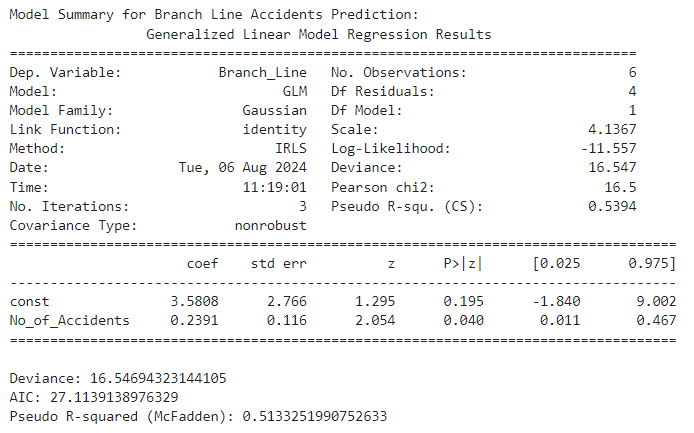
Variable Contributions to the Number of Accidents across the Years



The contribution of Accidents at Track and Accidents in Station Limits to the total number of accidents across the years is shown in the plot below. The blue line is for accidents at the track. As perceived, it is a fluctuating trend taking the general shape of the trend in the number of accidents. Higher contributions are therefore seen in years with more total accidents, for example, 2016 and 2019. The orange one contributes to the accidents within the station limits; this trend is much weaker and varies independently of the total number of accidents. The contributions are relatively stable but lower in comparison with track accidents.

# Accident on Line type Analysis





Summary of Model for Prediction of Main Line Accidents

Using the GLM to predict the number of accidents on main line tracks, we obtain the following results:

Pseudo R-Squared (McFadden's) : About 0.9992; this very high value says that the model fits extremely well to the data in hand, hence suggesting it explains almost all variation in No\_of\_Accidents for main line accidents.

Coefficient for No\_of\_Accidents: 0.7609 This coefficient is statistically significant with a p-value of 0.000, indicating that for each additional unit of No\_of\_Accidents, the number of accidents on Main Line tracks is expected to increase by about 0.7609 units. Since this relationship is very strong, it follows that the number of Main Line accidents is very responsive to a change in the overall number of accidents.

Deviance: 16.55, which is a measure of the fit of the model; that is, the smaller, the better. Since this is relative to the deviance being low in this case, it denotes that there is a good fit to the data.

AIC: 27.11. The lower the AIC, the more efficient the model in explaining more using fewer parameters. The low AIC thus indicates that the model is efficient at explaining the variability in number of Main Line accidents without overfitting.

Model Summary for Branch Line Accidents Prediction:

From the GLM run to predict the number of accidents on Branch Line tracks, it can be seen:

Pseudo R-squared: It is about 0.5394, which indicates that the model moderately fits the data, explaining a reasonable proportion of variability in Branch Line accidents but at the same time, compared with the Main Line model, it still retains quite a lot of unexplained variability.

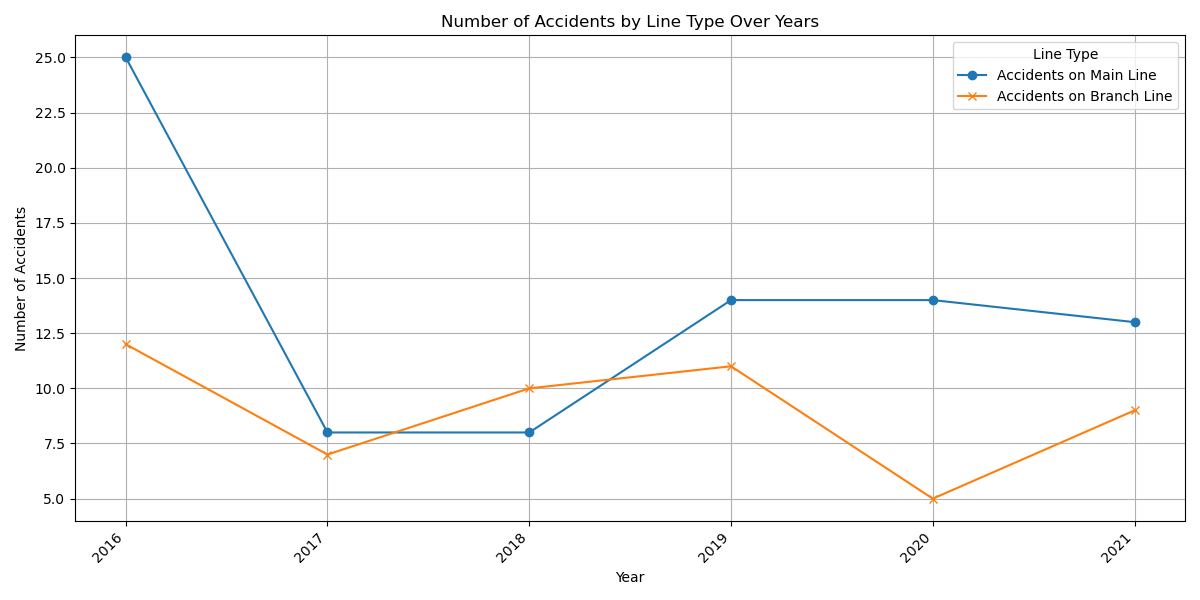
No\_of\_Accidents Coefficient: 0.2391. This coefficient is significant with a p-value of 0.04, so for every additional unit in No\_of\_Accidents, it is expected that the number of accidents occurring on Branch Line tracks will increase by about 0.2391 units. Though this relation is significant, but the strength of the relationship as seen earlier in the case of the Main Line model is not so strong.

Deviance: 16.55. Once again, the deviance is low, showing that the fit is good just as was obtained earlier in the case of the model for Main Line.

AIC: 27.11. Thus, the AIC reaches the same value as the one resulting from the Main Line model, showing the same complexity and adjustment of the model with different predictive strengths.

Graph Explanation:

This graph shows the number of accidents on the main line and branch line tracks from the years 2016 through 2021.

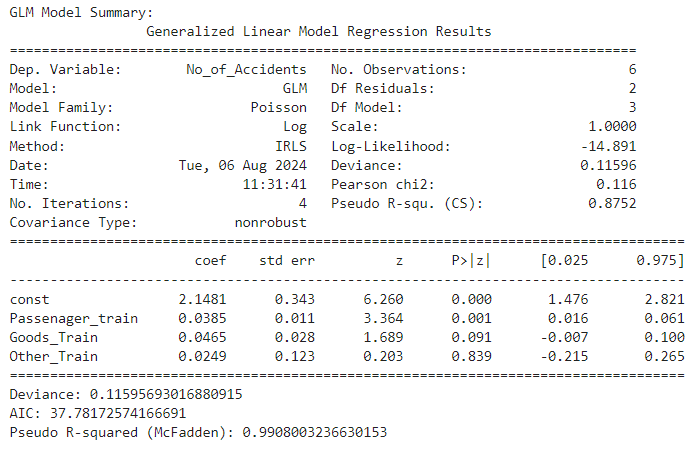


Main Line accidents: The graph shows variability in the number of accidents on Main Line tracks, with some noticeable increase in the last years. This may indicate worsening conditions or rising traffic on Main Line tracks.

Branch Line Accidents: While it does appear less variable, the number of accidents on the Branch Line tracks does have its peaks, notably in 2019. This could be taken to mean that the reduced count of accidents on the branch line tracks, compared to the main line, reflects differences in traffic volume, track conditions, or other operational factors.

This graph helps visualize the trends and differences in accident frequency between main line and branch line tracks and gives a clearer understanding of how accidents are distributed with respect to time and across different types of railway lines.

# Accident of Train Type Analysis

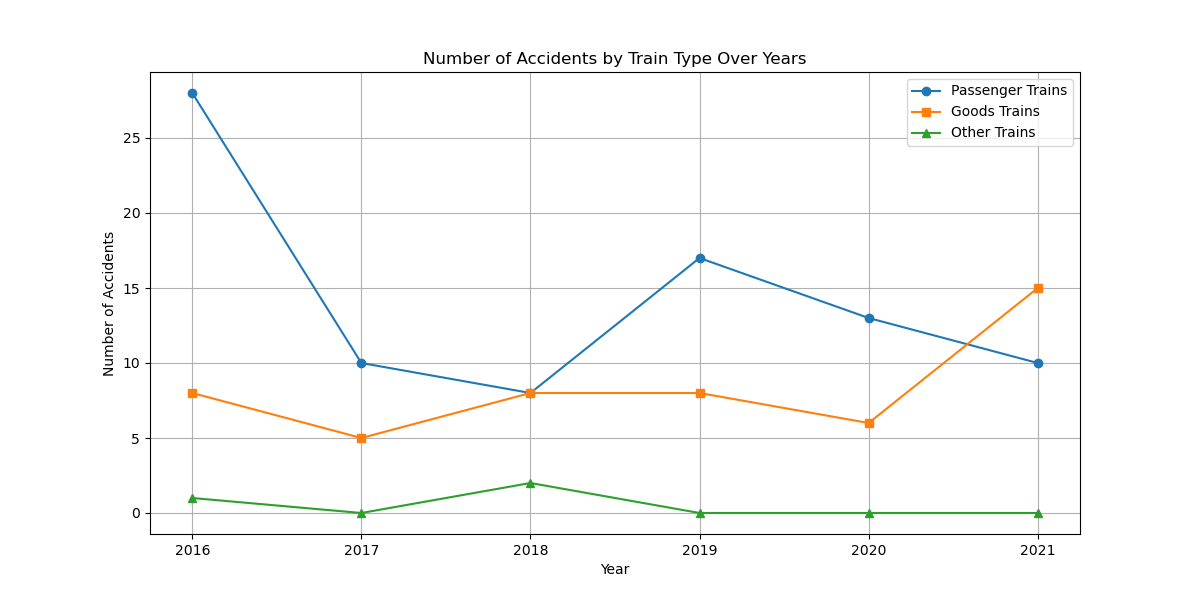


GLM Model Summary:

The method of estimation used the Generalized Linear Model with the Poisson family and was fitted for predicting the number of accidents. The Deviance is very low, which measures the goodness of fit of the model. The Akaike Information Criterion balances model complexity and fit with a value of 37.782.

It contains predictor variables, which include: passenger trains, goods trains, and 'other' trains. The coefficients for Passenger trains and Goods trains are statistically significant, with the p-values indicating very strong evidence against the null hypothesis. Specifically, the coefficient for Passenger trains is 0.039, while that for Goods trains is 0.047, indicating a slight increase in the number of accidents from an increased share of passenger trains and a similar but less intense effect when it involves goods trains. The coefficient for the other trains is statistically non-significant, so there is not a very strong evidence of its effect on the number of accidents.

Graph Interpretation:

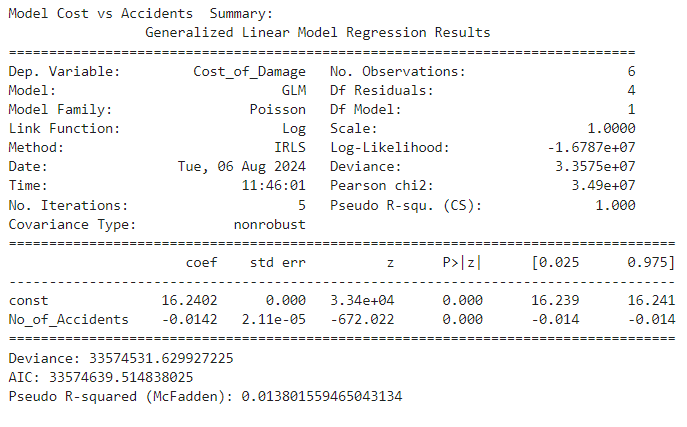


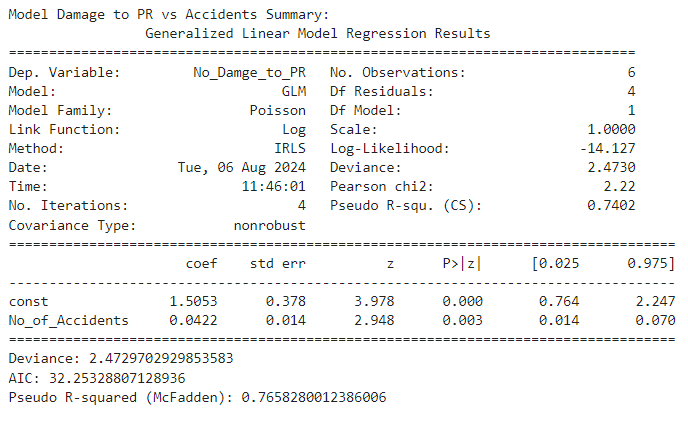
It graphs the number of accidents by train type over the years. There is variability in the number of accidents for Passenger trains, although generally higher as compared to Goods and Other trains. The goods trains are trending more obviously higher in number of accidents in 2021, while Other trains have consistently low values. This plot will be useful in bringing out the differences in accident frequency across different types of trains over the years.

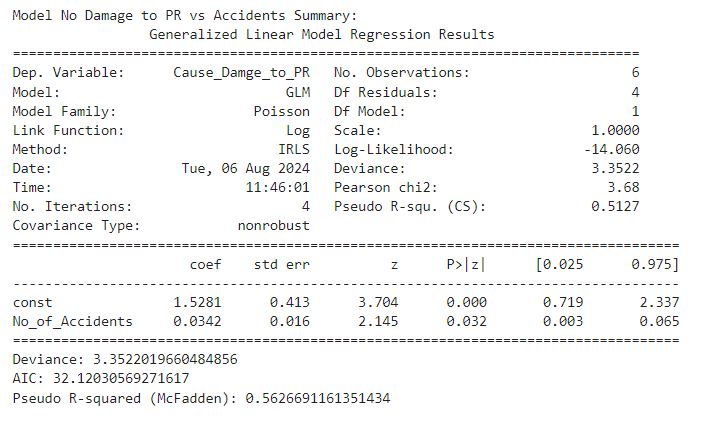
Discussion:

In this model, the dependent variable was considered the number of accidents. This was appropriate since I wanted to develop a deep understanding of how various types of trains influence the number of accidents. It would not make any sense to treat the number of accidents as an independent variable in such a scenario because the system will never be able to give any meaningful output on how various factors contributed towards an accident. Instead, interest lay in assessing the way changes in exposure to Passenger, Goods, and Other trains explain the number of accidents, and this relationship is appropriately modeled by a Poisson GLM.

# Cost of Damage/Damage to PR/No Damage to PR Analysis







Model 1: Cost of Damage vs. Number of Accidents

Summary of model

As shown here, I use a Generalized Linear Model for the analysis of how 'Number of accidents' influences 'Cost of damage'. Values of the Pseudo R-squared in the Poisson regression model equal to 0.0138 indicate that only a very small part of the variability of 'Cost of damage' can be accounted for by 'Number of accidents'. The coefficients indicate that while the number of accidents is statistically significant, it has quite a minimal effect on the cost of damage. The coefficient for No\_of\_Accidents is -0.0142, indicating that an increase in the number of accidents is associated with a very small decrease in the cost of damage, although it is practically negligible. This means that even though it is statistically significant, it has a small contribution to the variability of damage cost variable.

Model Metrics:

Deviance: 33,574,531.63

AIC: 33,574,639.51

Model 2: No Damage to PR vs Number of Accidents

Model Summary:

This GLM model explains the relationship between the number of accidents and "No Damage to PR" variable. The pseudo R-squared of 0.7402 indicates that there is a reasonably strong fit of the model, which suggests that there is appreciable variation in "No Damage to PR" explained by the number of accidents. The positive coefficient of 0.0422 for No\_of\_Accidents implies that for every extra accident, there is a rise in the "No Damage to PR." The relationship is statistically significant with a p-value of 0.003, thereby establishing that no. of accidents has a significant effect on the "No Damage to PR" variable.

Model Metrics:

Deviance: 2.47

AIC: 32.25

Model 3: Cause Damage to PR vs. Number of Accidents

Model Summary:

In this model, the effect of the number of accidents on "Cause Damage to PR" is observed. The Pseudo R-squared of 0.5627 indicates that at worst, this fit of the model will at least be moderate. The coefficient for No\_of\_Accidents is 0.0342, indicating that its relationship with "Cause Damage to PR" is positive, meaning that when the number of accidents goes up, "Cause Damage to PR" also goes up. The coefficient is statistically significant with a p-value of 0.032, indicating that the number of accidents has a significant effect on "Cause Damage to PR," though the relation is not as strong as in Model 2.

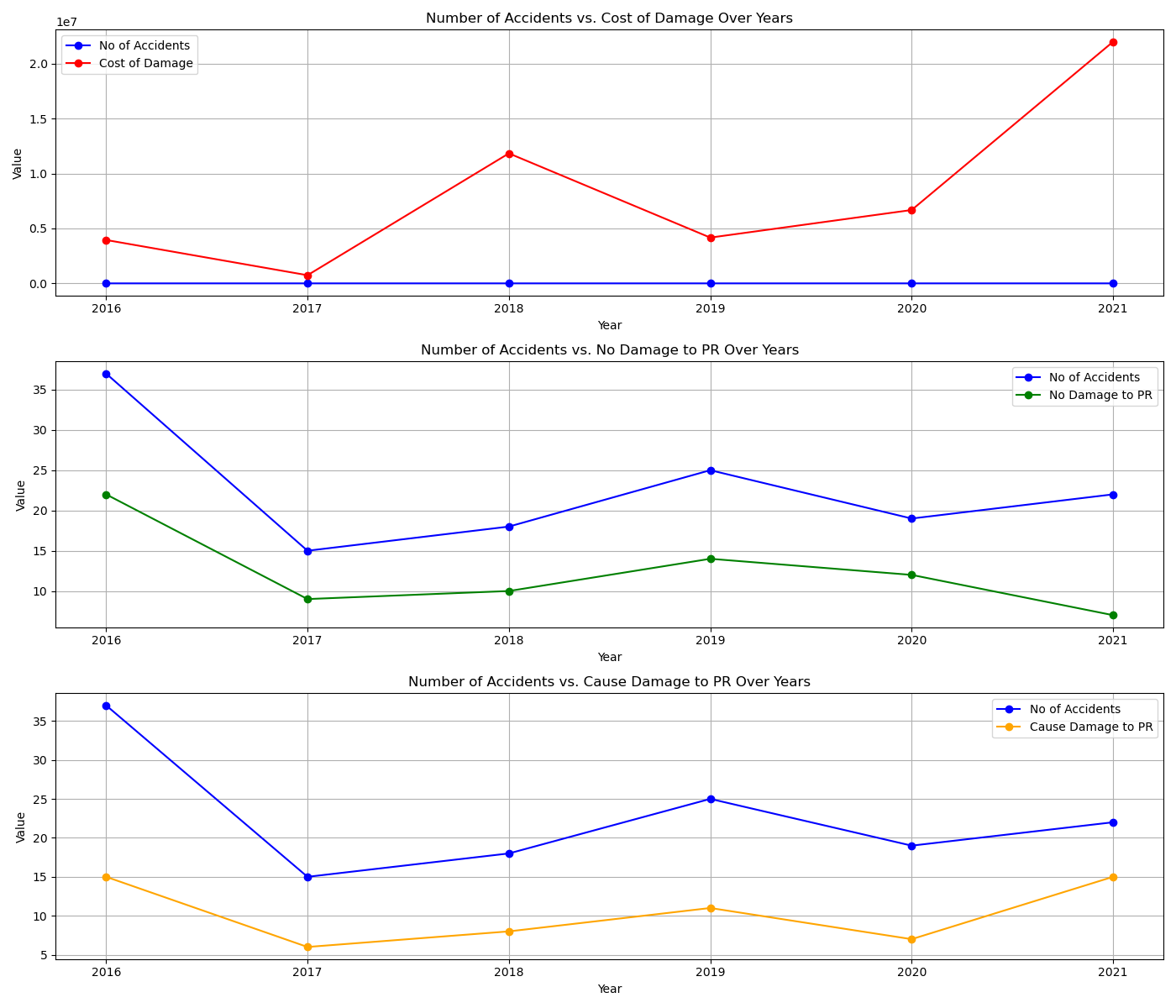
Model Metrics:

Deviance: 3.35

AIC: 32.12

Discussion of Graphs

The plots show trends over the years for each model graphically:



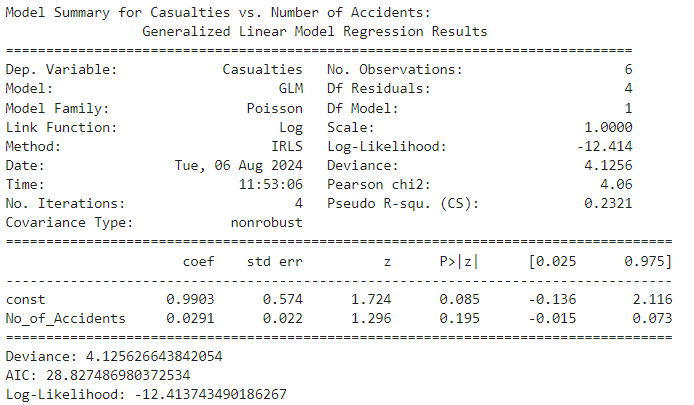
Number of Accidents vs. Cost of Damage: The graph depicts the trend of the number of accidents versus the trend of damage costs. No clear correlation appears, which affirms the statistical conclusion that the number of accidents has little influence on the cost of damage.

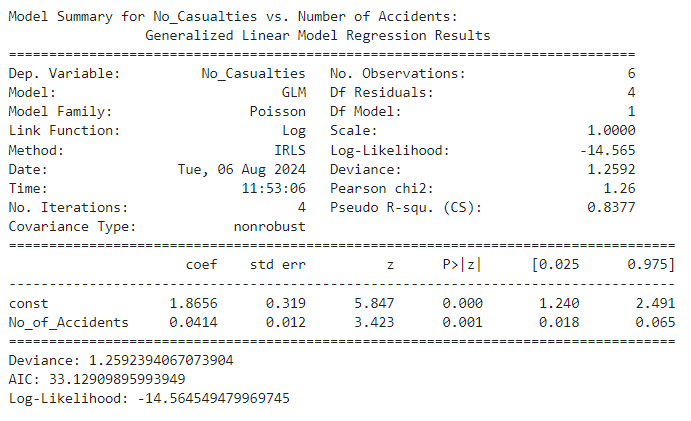
Number of Accidents vs. No Damage to PR: This graph shows a positive relationship between number of accidents and "No Damage to PR", which agrees with a significant positive relationship as captured by the model results.

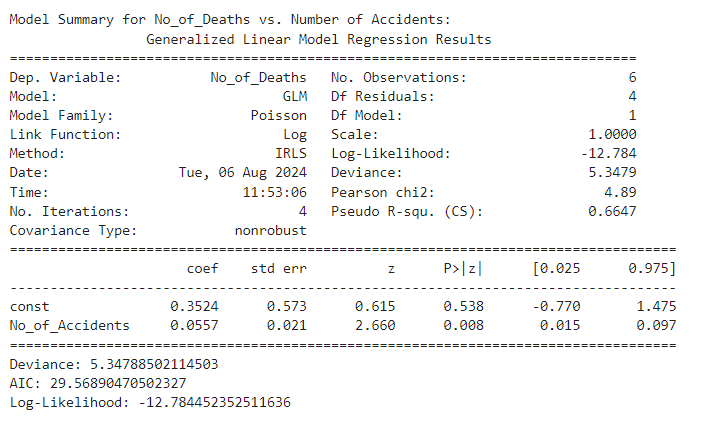
Number of Accidents vs. Cause Damage to PR: Similarly, this plot shows a positive trend of the number of accidents against "Cause Damage to PR" while the model shows a significant relationship.

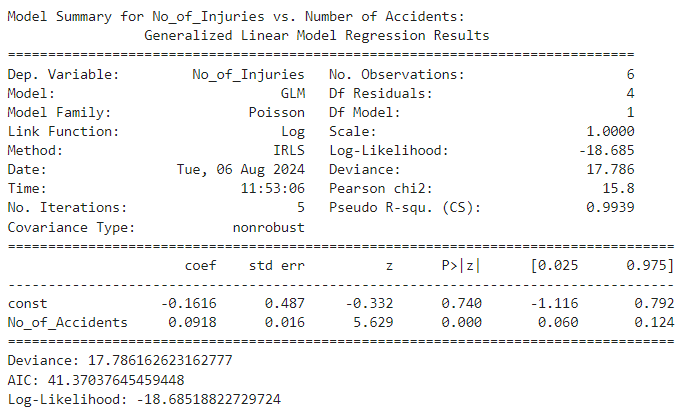
The general finding of this analysis is that the number of accidents does drive damage metrics, though, to a varying degree, from one kind of damage to another. Graphs provide context to these findings by illustrating the trends and relationships over time.

# Accident Severity Analysis









Model Interpretation

Casualties vs. Number of Accidents

For the number of accidents and the number of casualties, the Pseudo R-squared of the GLM is 0.2321. In other words, 23.21% of the variation in casualties is accounted for in the number of accidents. The coefficient of accidents in the model is 0.0291, which implies that with each one accident, the casualty is going to increase by about 0.0291. This effect is not statistically significant since the p-value is 0.195. The deviance for the model is 4.1256, while the Akaike Information Criterion has been 28.8275, so it seems that the model is badly fitting the data.

No Casualties vs. Number of Accidents

To predict the dependent variable, which is the number of cases where no casualties were recorded, the value of the Pseudo R-squared coefficient is 0.8377, which implies that, through the number of accidents, approximately 83.77% of the variation in the number of cases where no casualties were reported can be explained. The coefficient for number of accidents comes out to be 0.0414, which means that with each increase in accident there is a rise of about 0.0414 in the cases with no casualties. This effect is statistically significant with a p-value of 0.001. The Deviance for this model comes out to be 1.2592 and the AIC equals 33.1291, showing a relatively fine fit of the model to the data.

No of Deaths vs. Number of Accidents

The model explaining the number of deaths is able to return a Pseudo R-squared value of 0.6647, which means that about 66.47 percent of the variation in deaths is accounted for based on the number of accidents. The coefficient on the number of accidents is 0.0557, which means that for an additional accident, the number of deaths would increase by about 0.0557. The p-value will be 0.008, confirming this to be statistically significant. The Deviance is 5.3479, with an AIC of 29.5689. These numbers indicate that it is a satisfactory fit; there is room for just a little improvement.

No of Injuries versus Number of Accidents

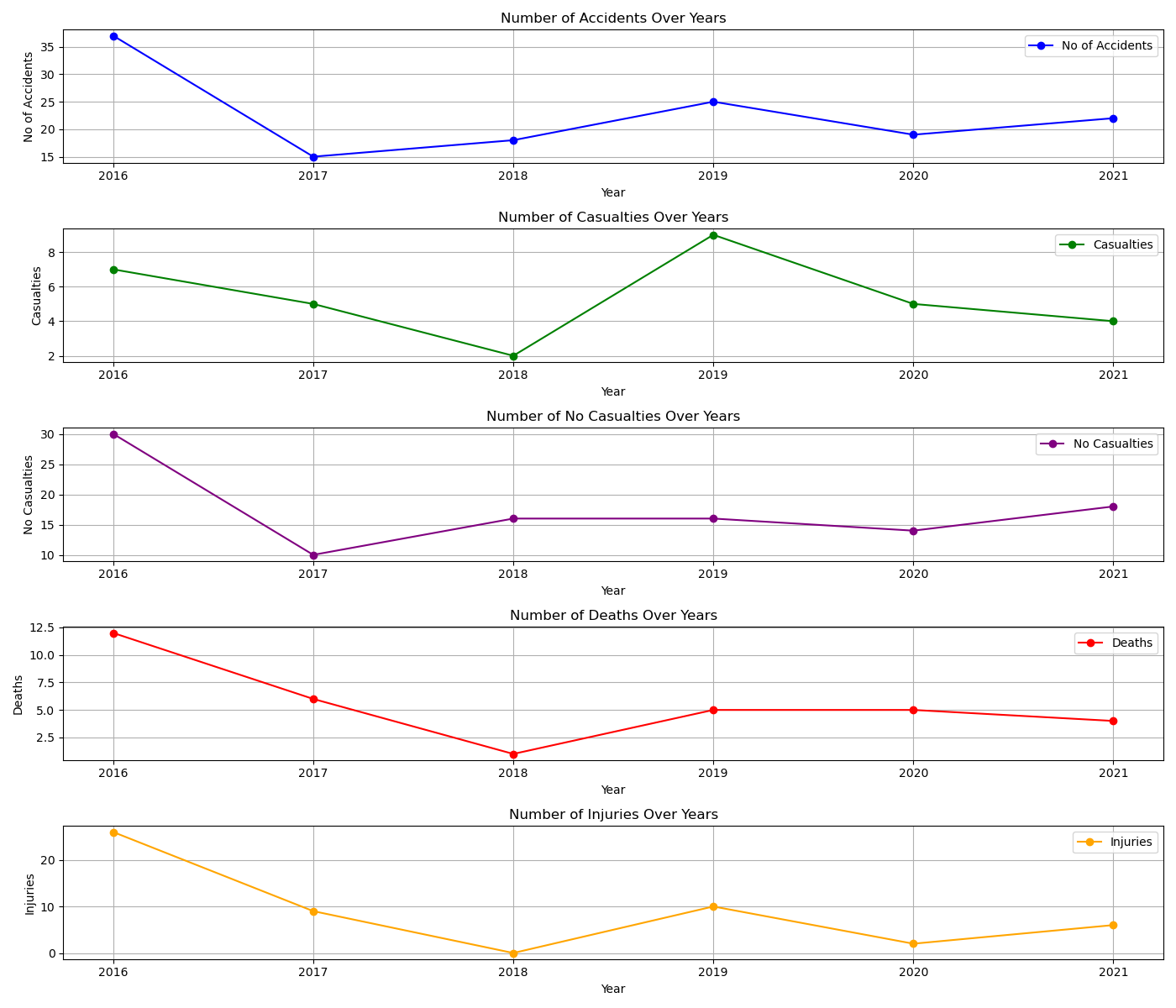
This injury number model is very highly significant, given that the Pseudo R-squared value is as high as 0.9939, meaning number of accidents explains 99.39 % of the variation in injuries. The coefficient of the variable number of accidents is 0.0918, meaning that for each additional accident, there are about 0.0918 more injuries that occur. Such a relationship is highly significant because the p-value is low at 0.000. Of course, the Pseudo R-squared seems to be significantly high, but the Deviance for this model comes to be 17.7862 and the AIC equals 41.3704, so resoundingly, the model may be overfitted.

Summary

The GLMs indicate different ways in which the number of accidents predicts such outcomes as casualties, cases with no casualty, deaths, and injuries. The number of accidents makes a strong effect on injuries, while it is less on casualties and cases with no casualties. The fit varies across all models: some vary good; some show potential improvement.

Graph Interpretation

The graphs below indicate the trend of the of the number of accidents, casualties, cases with no casualties, deaths, and injuries over the years 2016 to 2021.



Number of Accidents Over Years

The first plot indicates a fluctuating trend in the number of accidents from 2016 to 2021 because, in 2016, a first peak is observed with the number of accidents standing at 37, but in the following year, the number decreases, at which point it continued until 2021, where it was 22. This variation implies that the likelihood of accidents does not follow a consistent year-after-year pattern. The possible alternatives are other factors, maybe not exactly obvious in this plot only.

Number of Casualties over the Years

The second graph considers the number of casualties for each year. It is seen that in 2016, before the subsequent years had a record of 30. It then ranged in the subsequent years. Has its lowest value in 2018, which is 16. From this, it can be considered that other variables exert less of an effect on the number of casualties than they do on the number of accidents.

Number of Occurrences with zero casualties Over Years

The third plot regards casualty-free cases. It has a decreasing trend from 2016 to 2021—30 cases in 2016 to 18 cases in this current year: 2021. There might be a decline in the frequency of cases without casualties, probably suggesting an increase in the severity of accidents or improved safety measures.

Number of Deaths Over the Years

The fourth plot deals with death counts for each year. There is little variation, with a peak in 2016 (12 deaths) and a low number in 2018 (1 death). The deaths seem highly variable because one may get years of many deaths and years with a very shallow number of them. This high variability could possibly show a change in the severity of the accidents or in different masked factors that act behind them.

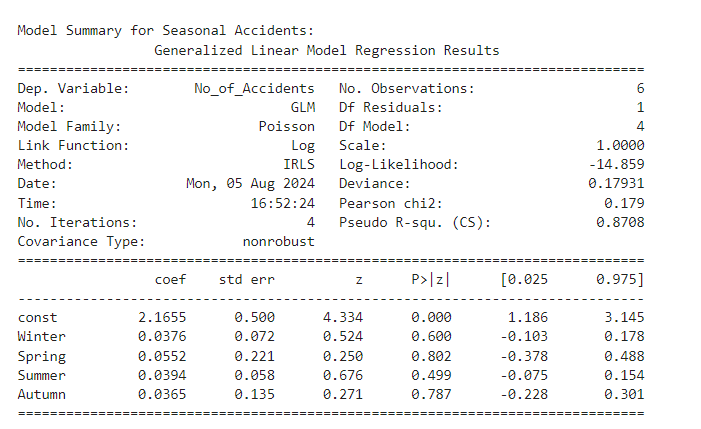
Number of Injuries Over Years

The last plot represents the counts of injury over the years. This plot reveals a massive peak during 2016 that includes 26 injuries and trails off a lot after throughout other years, particularly of 2018 and 2020. It's notable that the evenness of this trend is violated in two different years, one at the beginning and another at the end. With a high count, it may indicate that it is either an exceptionally high count of grave accidents or has very different reporting standards, which cannot be set to provide a definite explanation.

Summary

Graphical analysis provides insights on which specific railway accident metrics are variable across time. The trends in all these variables—number of accidents, casualties, cases with no casualties, deaths, and injuries—are highly changeable; thus, it might be said that any of the abovementioned items are dependent on factors other than those indicated by the number of accident. This variability could be helpful toward further inquiry on what other factors may contribute to the change observed in these metrics through different years.

# Seasonal Accidents Analysis



Discussion on the Selection of the Independent Variables

Seasonal variables such as Winter, Spring, Summer, and Autumn have been incorporated as independent variables for the test. These variables are used to test how they may be related to causing variations in the accident rates. This is justified, for it is to be expected that seasonal changes would lead to the imposition of differences in the growth of the accident rates.

## Model Interpretation

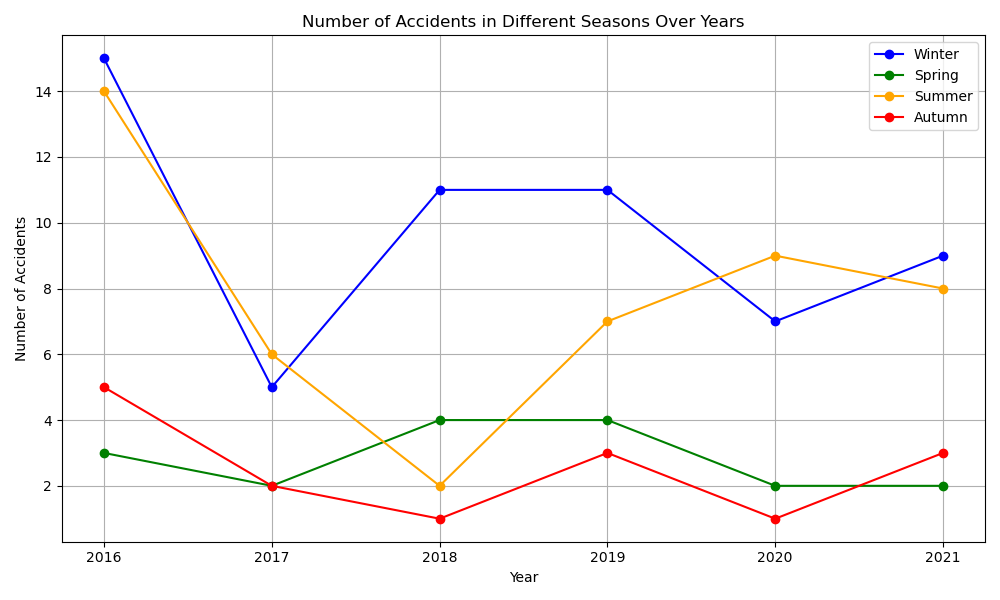
The GLM of season-only accidents has the effects of the number of railroad accidents in relation to the seasons on which they occur. The model had been fitted using a Poisson family with a log link function based on six observations. The deviance equals 0.17931 and the Pearson chi-square value is 0.179, both of which imply that the model fits good to the data. A log-likelihood of -14.859 and the Psuedo R^2 value of 0.8708 indicate that the model explains a large part of the variability of the number of accidents.

The Intercept: (i.e., the constant) is as follows: 2.1655 p-value < 0.001. This already suggests that there is a large baseline level of accidents, regardless of season.

The coefficients on Winter, Spring, Summer, Autumn are 0.0376, 0.0552, 0.0394, 0.0365 respectively. None of the above coefficients are significant as their p-values range from 0.4591 to 0.9425 which are certainly above the conventional level of 0.05. This means that the number of accidents in any particular season has no significant effect on the total number of accidents.

## Graph Interpretation

The chart shows the number of train accidents that occurred over seasons in the selected years: 2016-2021. Each season is represented by a line: Winter (blue), Spring (green), Summer (orange), and Autumn (red). Yearly, it differs by the number of accidents for each season.



In Winter, the count of accidents varies, hitting a peak in 2016 and again in 2018. Over the years, Spring has fairly low and consistent numbers of accidents. There is a bit more variation during the summertime, where there are peaks in 2016, a decrease in 2018, and then increasing increments in numbers. The autumn season shows a similar aspect of having minimal variations in accident numbers over the years, with only slight differences involved.

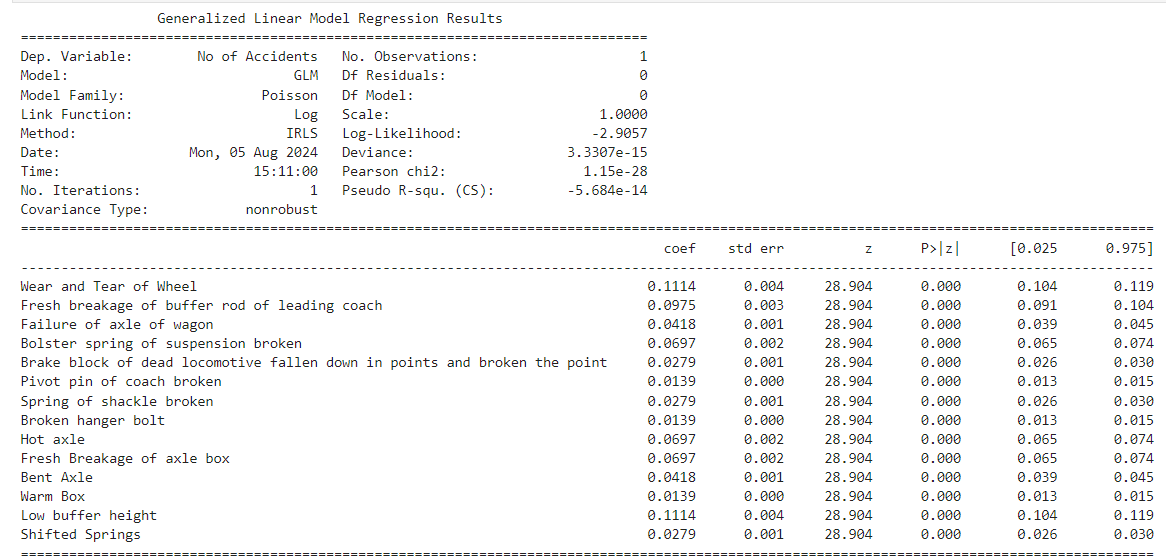
The model having a high pseudo R-squared value shows that it is a good fit and explains the major portion of the variability in the number of accidents. The seasonal coefficients are not statistically significant, while at the same time, the number of accidents in any particular season doesn't seem to matter on the total number of accidents. This probably means that there is an interaction with other factors not modeled here, as this output controlled for season and year. Not only is this indicated by the output, but also by the graphical representation of this finding, which shows that although the number of accidents changes from season to season and year after year, no trend or pattern comes up to say that some particular season is connected with a high or low number of accidents.

# Causes of Accidents

Independent variables chosen as mechanical defects permit the analysis to single out and measure the impact of particular problems on accident rates. The total number of accidents used as a dependent variable provides a direct measure of safety outcomes, thereby giving an exact understanding of how each defect contributes to the overall risk.

Furthermore, the choice of independent variables—the signal defects—will let the model recognize and quantify the specific contribution of various issues to the total quantity of accidents. On the other hand, the dependent variable of the total number of accidents then delivers a measure of effect as comprehensive as possible, so that targeted improvements in safety are possible.

## Mechanical Defects



### Interpretation of Mechanical Defects

Using the GLM with the Poisson family and mechanical defects data gave insight into how each of the mechanical defects affects the total number of accidents. The detailed interpretation based on the model results is as shown below:

## Model Results

The coefficients for all types of mechanical defects are positive and significant, as the z-values and p-values show, all of whom have p-values equal to 0.000. This means that each defect contributes to the number of accidents, though differently.

## Wear and Tear of Wheel:

The coefficient 0.1114 means that for every additional unit of wear and tear of the wheel, the log of the expected number of accidents increases by about 0.1114. This value denotes a moderate effect on the number of accidents.

## Fresh Breakage of Buffer Rod of Leading Coach:

This defect with a coefficient of 0.0975 increases the log of the number of accidents by 0.0975 per unit. The effect it has on the frequency of accidents is quite big.

## Axle of Wagon Failure:

The coefficient is 0.0418, indicating a much smaller effect of wagon axle failure on accidents.

## Bolster Spring of Suspension Broken:

This defect has a coefficient of 0.0697, indicating that its effect on the number of accidents is quite large, although not as large as some other defects.

## Brake Block of Dead Locomotive Fallen Down:

With a coefficient of 0.0279, this indicates a small effect on accident frequency relative to other defects.

## Pivot Pin of Coach Broken:

With a coefficient of 0.0139, this suggests a relatively small effect on the number of accidents.

## Spring of Shackle Broken:

defect has a coefficient of 0.0279, hence its effect on the number of accidents is only moderate.

## Broken Hanger Bolt:

Its coefficient of 0.0139 denotes that its effect upon the number of accidents is minor.

## Hot Axle:

It has a coefficient of 0.0697, meaning that it significantly affects the frequency of accidents, much like issues with the bolster spring.

## Fresh Breakage of Axle Box:

coefficient for this defect is 0.0697, so the contribution of this variable to the number of accidents is very remarkable.

## Bent Axle:

The coefficient is 0.0418, hence this independent variable is said to have a moderate effect on the number of accidents.

## Warm Box:

With a coefficient of 0.0139, this independent variable is said to have a small effect on the number of accidents.

## Low Buffer Height:

With this coefficient being 0.1114, this defect has a great influence on the number of accidents, just about similar to what the wheels' wear and tear do.

## Shifted Springs:

A coefficient of 0.0279 represents a very moderate effect on the frequency of accidents.

## Impact of total number of accidents

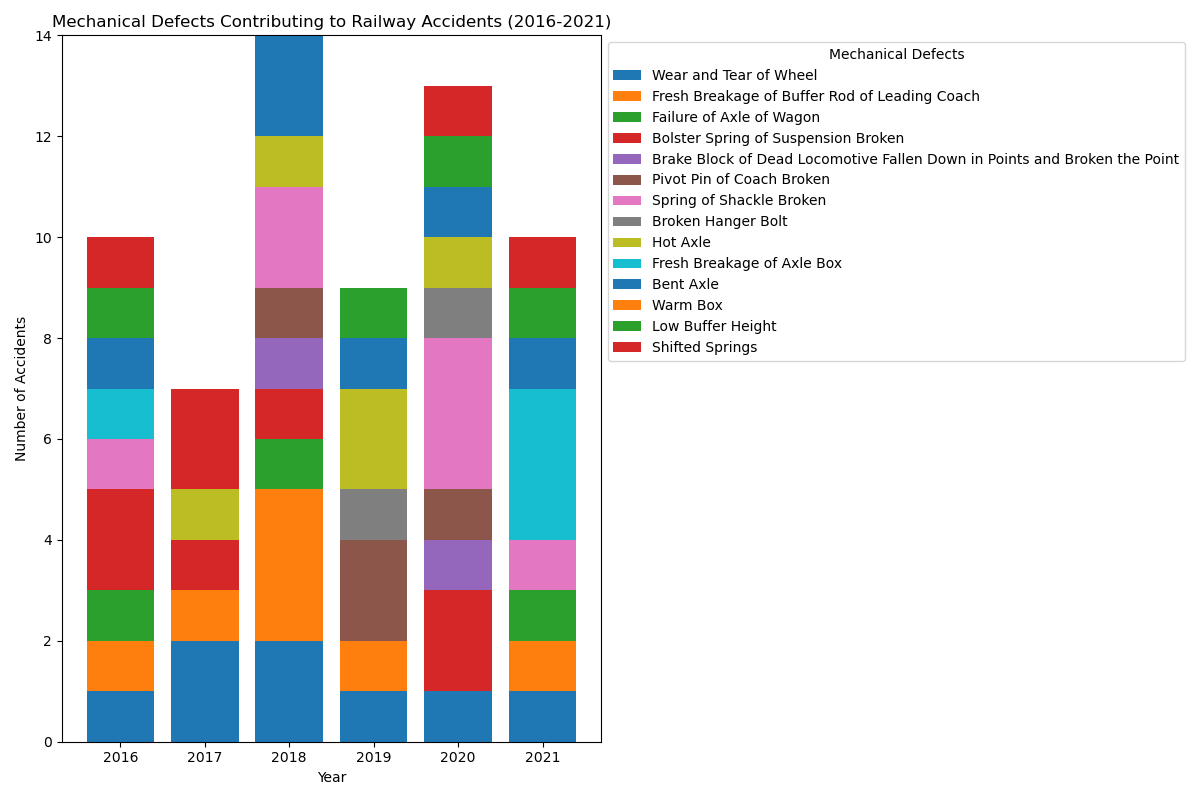
The coefficients indicated that all mechanical defects contributed to the total number of accidents. The magnitude of each coefficient measures the relative impact of that defect type. For example, defects of wear and tear of wheels and low buffer height have larger impacts on the total number of accidents when compared to pivot pin of the coach broken and warm box defects.

Thus, targeting mechanical defects with the larger coefficients may avoid more accidents. It follows from the model that while the frequency of accidents is sensitive to all types of mechanical failures, regardless of how serious they are, they have different impacts on the frequency.

## Conclusion

Poisson GLM analysis indicates that the effect of mechanical defects on the total number of accidents is positive and significant. Different defects contribute to different extents towards explaining the number of accidents, with some having more substantial contributions. These contributions can, therefore, be used to guide priority in addressing these mechanical issues to effectively reduce the number of accidents.

## Graph



## Graph Data Discussion

The stacked plot of mechanical defects for the period 2016 through 2021 shows how different types of mechanical failures have contributed to the total number of accidents each year. The graph gives the information on the relativity of each mechanical defect toward the overall count of accidents.

## Analysis of Defects Over Time

## Wear and Tear of Wheel:

This defect is seen regularly every year and is one of the recurrent defects. Its contribution to the overall accidents is on the higher side, which shows it is a recurring problem and one of the main concerns for maintenance.

## Fresh Breakage of Buffer Rod of Leading Coach:

This defect shows variability but has been remarkable in some years, more precisely in 2018 and 2020. The contribution of this defect comes in different ways, hence proving that as important as it may be, it is not consistently high compared to the wear and tear of wheels.

## Axle of Wagon Failure:

This defect also strikes as frequent, more precisely in 2018 and 2021. The occasional spikes suggest periodic issues with axle failures that could be linked to certain operational or maintenance conditions.

## Bolster Spring of Suspension Broken:

This defect is one of the most significant contributors to accidents, particularly in 2016 and 2020. The recurring nature of its presence shows that bolster spring failures are a recurrent problem that might require focused interventions.

## Brake Block of Dead Locomotive Fallen Down in Points and Broken the Point:

This defect contributes to individual accidents in a few specific years, showing a sporadic nature. It is an important defect but seems less consistent compared with others.

## Pivot Pin of Coach Broken, Spring of Shackle Broken and Hot Axle:

These defects make intermittent contributions to the total accidents. They are relevant but do not appear as consistently significant as some other defects.

## Fresh Breakage of Axle Box, Bent Axle, Low Buffer Height and Shifted Springs:

These shortcomings flutter around less often, but they still contribute, incrementally, during specific years, to the totals of that particular year. Their fluctuating nature might suggest that they are situational or even related to more specific lapses in maintenance.

## Trends and Observations

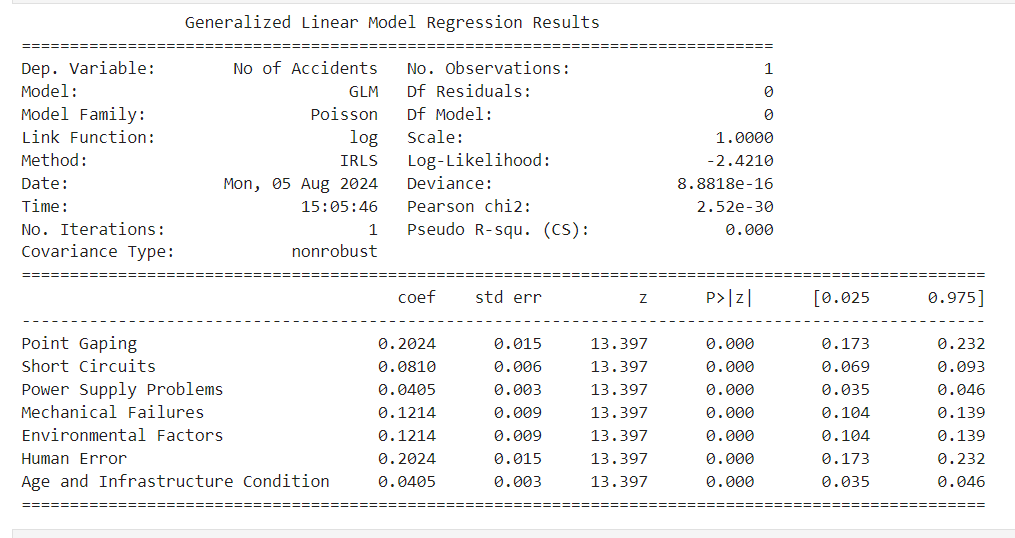
Annual Variation: The number of total accidents changes from one year to another. There are also clear peaks for the different years. For instance, 2018 and 2020 have a higher percentage share of accidents. This could be because certain types of deficiencies are happening more often in.

Consistent Defects: Some defects, like the wear and tear of the wheels, appear every year, thus, they are among the consistent issues that need constant monitoring and improvement.

Patterns of Defects: The various defects contribute differently through the years. For instance, some have peaks at certain years. Thus, interventions need to be focused on those periods while addressing these issues.

Clearly, the plot shows a trend of the contribution of different mechanical defects to railway accidents over time. It, therefore, indicates that there is a need to address recurrent defects and offers indications of areas for maintenance improvements.

## Signal Defects



## Interpretation of GLM Summary for Signal Defects

Generalized Linear Model under Poisson family to plot insights on how different signal defects affect the aggregate accidents. Below is a discussion in detail, based on the summary:

## Model Results

It is evident from the z-values as well as the p-values as all coefficients of each signal type of defect are positive and statistically significant. Each of the p-values is 0.000, which means that each of the defects has a measurable impact on the number of accidents.

## Point Gaping:

The corresponding coefficient is 0.2024, signifying that a one-unit increase in point gaping intensity or frequency will cause the log of the expected count of accidents to increase by approximately 0.2024, all else being equal, indicating a higher number of accidents under point gapping.

## Short Circuits:

The coefficient is 0.0810, which means that, with one more unit of short circuits included in an estimate, the log of the expected number of accidents will go up by 0.0810. Although this is a relatively minor increment in relation to several others, it still contributes to the increase in accidents.

## Power Supply Problems:

The given defect has a coefficient equal to 0.0405. It will show that this defect is relatively small, and hence, having a problem at the power supply one more time will increase the log of expected accidents by 0.0405.

## Mechanical Failures:

The defect has a coefficient of 0.1214, meaning it is quite significant with the number of accidents. For each unit growth in mechanical failures, by 1, this will increase the log of the number of accidents by 0.1214.

## Environmental Factors:

As well as depicted by the good coefficient at 0.1214, this supports an up raise in the number of accidents.

Human Error: From the coefficient of 0.2024, clearly it comes out that human error is among the leading factors that have had an immense effect on the number of accidents, notable point gaping

Age and Infrastructure Condition: This is seen from the coefficient of 0.0405; It's not that a significant effect as compared to some other defects but does add up onto the number accidents.

## Inferential on Total Number of Accidents

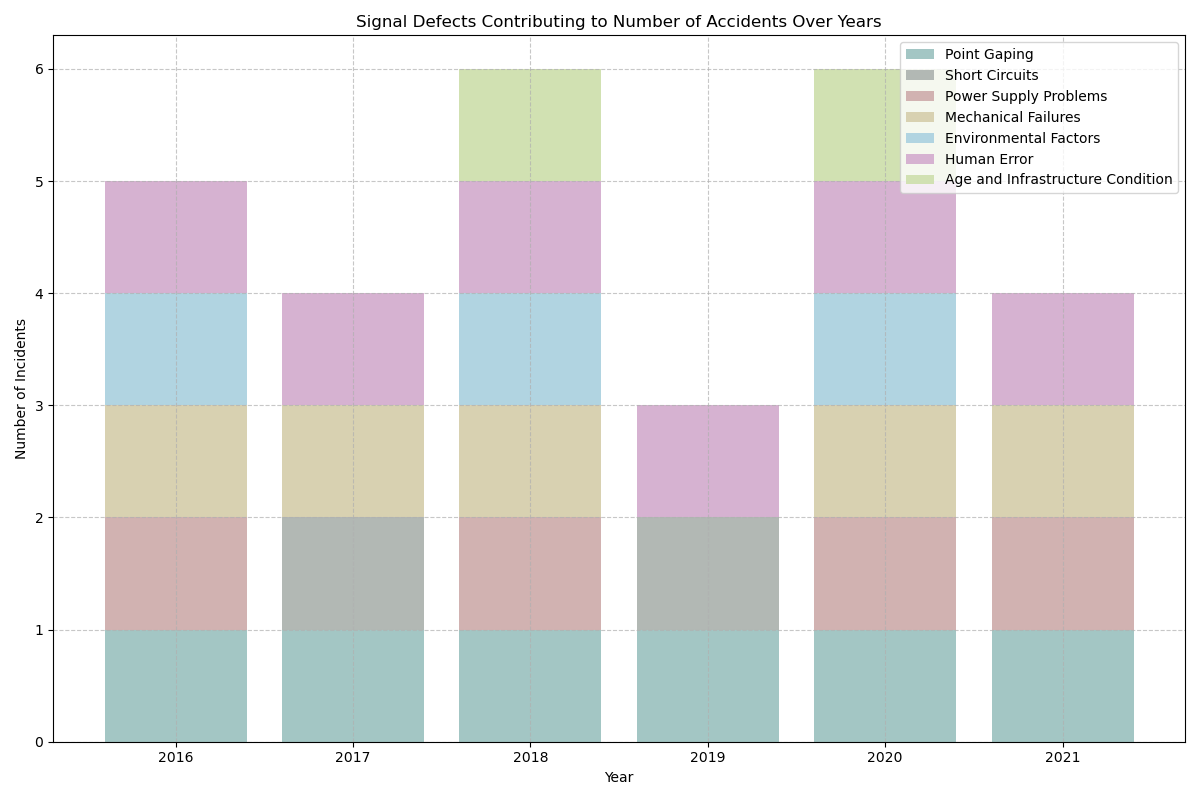
The coefficients fitted by the GLM are used to interpret how each type of signal defect affects the total number of accidents. The higher the coefficient, the more the impact on the total number of accidents. For instance, point gaping and human error affect more than power supply problems.

In other words, a better reduction in accidents due to higher coefficients of defects would be drawn. All positive coefficients in defects highlight that all types of defects contribute to accidents at varying intensities.

## Conclusion

The GLM analysis shows that each of the signal defects considered had a positive significant influence on the total number of accidents, and the understanding and mitigation of the defects could possibly cure the number of accidents. A Poisson model well described the relationship with the number of accidents, from signal defects, and brought insight with meaningful ways of remediating the situation.

## Graph



## Discussion of the Signal Defects Data

A stacked bar graph is used to give an overview of the contributions of various signal defects towards the total accidents from the years between 2016 - 2021. The following is an in-depth discussion of what the graph communicates:

## Interpretation

The graph presents the data of various types of signal defects, including their contribution to the general number of accidents for every year.:

Human Error and Point Gaping are consistent in most of these years, often being abnormally contributory to the total number of incidents. This means that these non-identical elements are exercising any continued influence on the speeds of incidents and are, therefore probably domains in which converged schemes might be of value.

## Variability in Other Defects:

The contribution of other defects fluctuates from year to year. For instance, Short Circuits and Problems with Power Supplies do not seem to have occurred as frequently in some years, seemingly reducing the overall consistency of these problems, or perhaps indicating their resolution in those particular years.

## High Accident Years:

Years like 2018 and 2020 have high total accidents. The bars in these specific years reflect a great deal of stacking, showing a myriad of defects involved. These different issues add up, showing an overall accident rate in these years. High contributions from Human Error and Mechanical Failures remark that these were areas of central concerns during this period.

## Temporal Trends:

These general trends show that the accident counts fluctuate over the years and certain defects are now increasing or decreasing in important contributions. For example, the year 2020 has a big number for Mechanical Failures, whilst in 2019 there are much lower total accident numbers with a more uniform distribution. Impact of Individual Defects

## Human Error:

This is a constant defect that contributes a significant percentage to the total accidents, the importance of which implicates its pivotal contribution to the general safety issues. The high values it posts always on several years bring out the importance of enhancing the training, procedures, or supervision in reducing human failure.

## Point Gaping:

Another flaw with a consistent contribution across the years. Its persistent appearance would suggest that the infrastructure-related issues, such as track alignment and maintenance, are persistently a worry to safety, and may require independent and maintenance scrutiny in a periodical manner.

## Mechanical Failures:

This imperfection is a variable but has large contributions in a few years. One reason for the fluctuational pattern could be due to varying conditions or different levels of mechanical upkeep in different years.

## Power supply problems and short circuits

show less often but still contribute to the number of accidents. Their peaks in some years indicate that while they may be less prevalent, in years when they do come around, they do so in significant numbers.

## Environmental Parameters and Age/Infrastructural Condition:

These are relatively variable contributions. Sometimes, environmental conditions may enhance other defects, while the age or the condition of the infrastructure may be a contributory factor, especially in old systems or during unfavorable conditions.

## Conclusion

The use of more softened colors on the stacked bar plot does effective communication on the contribution of signal defects to the total numbers of accidents over the years. On closer scrutiny, trends would, therefore, identify those defects that seem to be perpetual problems and which ones change drastically with the passage of time. Such information is crucial in the identification of areas for priority improvement and interventions that will ensure enhanced safety and reduced accident rates.