

General Insurance Project

Sivaram Ainkaran

Introduction

Aim

The aim of this project is to analyse the data of SuperCar Insurance's claim amounts made by claimants last year and provide them with a thorough summary and analysis of the state of the comprehensive car insurance market. The data will be used to provide SuperCar Insurance with details on their premium pricing and how it should be attacked in the following years.

Background

Comprehensive car insurance is the highest level of protection a claimant is able to put on their car and covers damage done to the claimant's vehicle, to other vehicles due to claimant's fault, damage done by storms, hail, fires, etc. and if the claimant's vehicle is stolen. The other forms of car insurance; Fire and Theft, Third Party Property and Compulsory Third Party insurance do not cover all of this and is not often recommended.

The Australian Car Insurance market consists of about \$21 billion which is approximately 30% of the total market share that the General Insurance market makes up (\$67 billion). Over the last 5 years the car insurance industry has been growing at a rate of 2.5% but is expected to decrease to around 1.9% due to increases in natural disasters greatly affecting the industry. Due to the time taken to introduce greater premium prices, current claim amounts have been devastating the industry and greatly reducing current profits, decreasing industry growth.

Analysis of Claims Data – Basic Excel

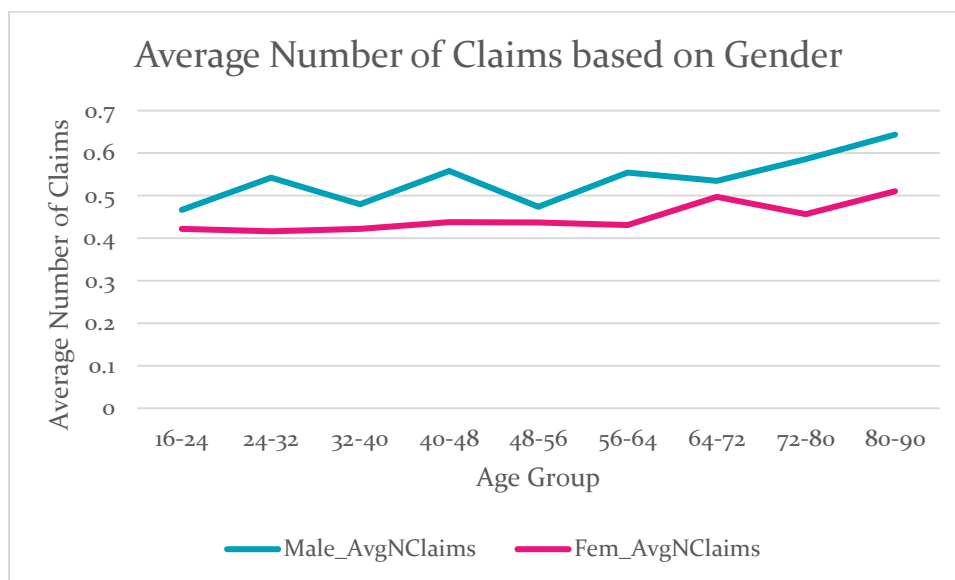
The claims data we are given can be used analysed to determine how different factors affect claim amounts and sizes and how those factors can be used to provide better premium pricing models. The main variables in the claims data from last year which can be easily translated into modelling price are Gender, Age, Marital Status, Postcode/Population, Number of Claims and Total Claim amounts. Using these factors we can begin to analyse the data to draw a framework, from which we can determine what type of model should be used to determine SuperCar Insurance’s pricing structure. This framework should consist of graphs and comparisons between the aforementioned variables to determine which factors impact claim amounts in what way.

We can look at how Gender, Marital Status, Age and Postcode/Population affects the number of claims and total amount of claims made by claimants throughout the year. This will help us create the initial framework required through simple excel analysis and graphing. To begin with, all the ages are split into 9 age groups of equal size (except the last one 80-90) which will make manipulating and graphing the data we are given much easier.

Gender and Age

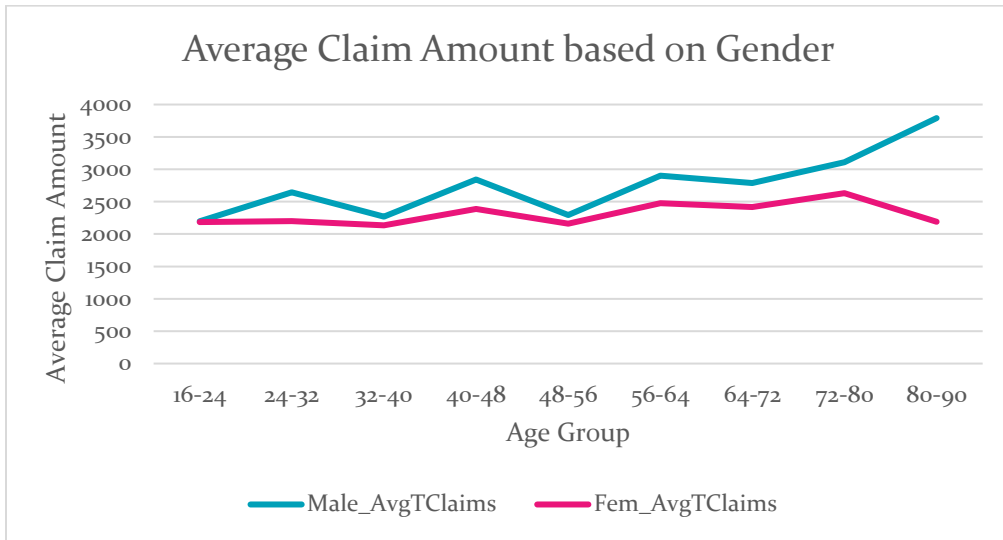
Number of Claims

The frequency of claims for each gender in each age group is divided by the number of members of that gender in that age group to determine an average number of claims per insurance holder for each gender in their given age group.



Total Amount of Claims

The Total Claim amounts for each gender in each age group is divided by the number of members of that gender in that age group to determine an Average Claim Amount per insurance holder for each gender in their given age group.

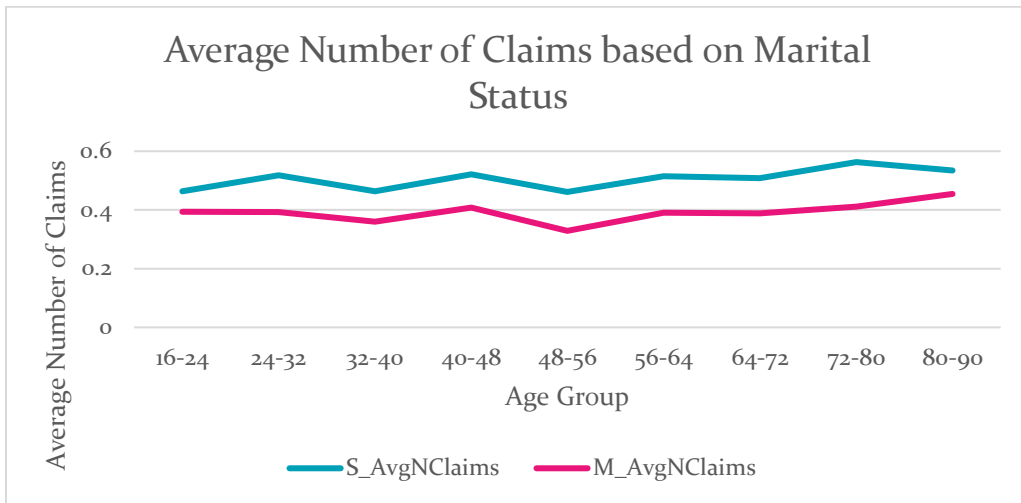


Both these graphs illustrate the fact that males tend to have higher insurance claim amounts and frequencies than females on average and that on average, the older a insurance holder is, the more claims they are likely to make.

Marital Status and Age

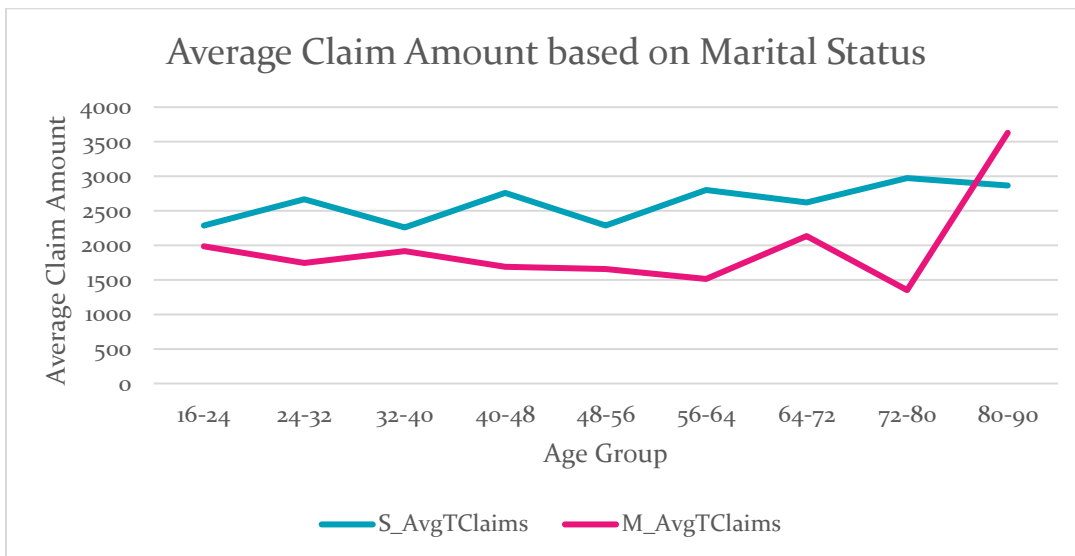
Number of Claims

The frequency of claims for each marital status in each age group is divided by the number of members of that marital status in that age group to determine an average number of claims per insurance holder for each marital status in their given age group.

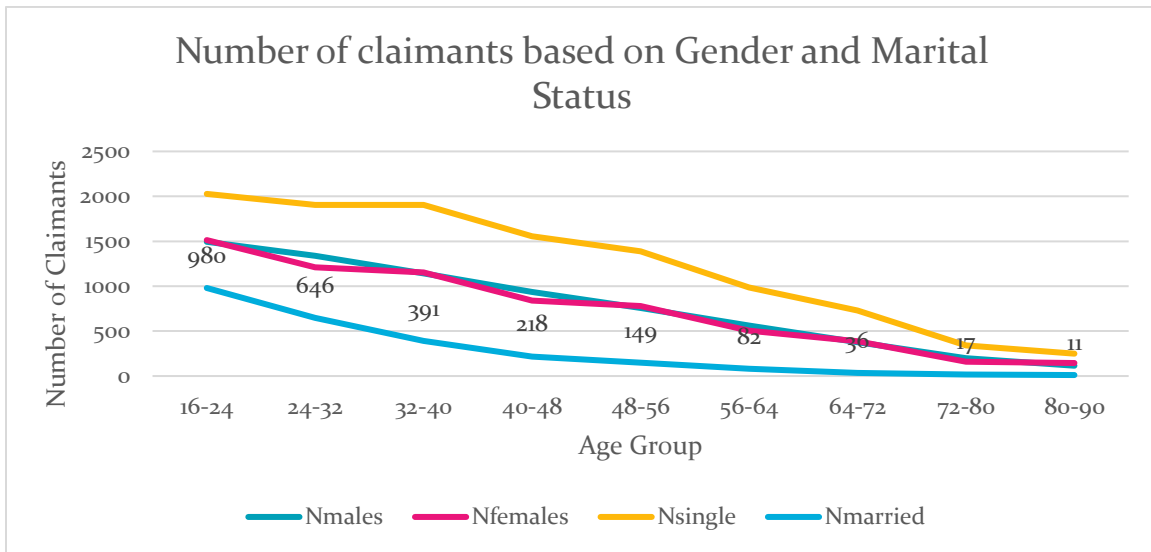


Total Amount of Claims

The Total Claim amounts for each marital status in each age group is divided by the number of members of that marital status in that age group to determine an Average Claim Amount per insurance holder for each marital status in their given age group.

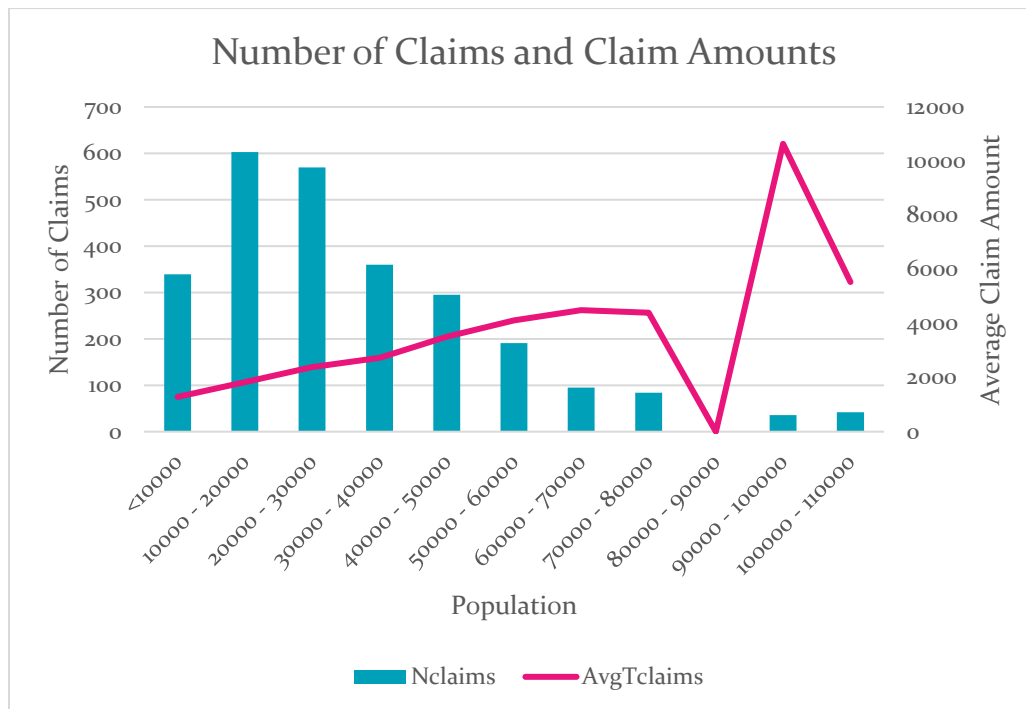


Both these graphs illustrate the fact that single people tend to have higher insurance claim amounts and frequencies than married people on average and that on average, the older a insurance holder is, the more claims they are likely to make. There is a spike in the average claim amounts of married people in the 80-90 age range which is likely just an outlier, being that the number of claimants within that range who are married is only 11 as shown on the graph below.



Population

The Total Claim amounts for each population range is divided by the number of members of that population range to determine an Average Claim Amount per insurance holder for each population range. The number of claims for each population is directly recorded from the given data from last year to provide the following graph.



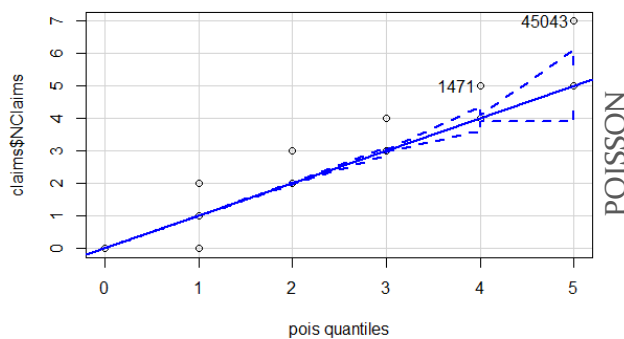
We can see that as population increases, the severity of the claim tends to increase steadily. There are some outliers though, being that there is no data recorded for the 80000-90000 population range and that population ranges higher than that have very few recorded data points. Using all this information about age, population, gender and marital status a suitable distribution to model this data can be determined.

Analysis of Claims Data – Monte Carlo

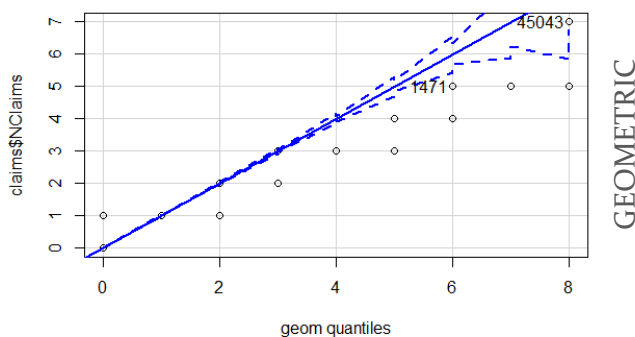
Using RStudio and by comparing multiple models within it using the data given and Monte Carlo Simulations, a model to be used for our pricing strategy can be determined based on how well it fits with our data.

Number of Claims

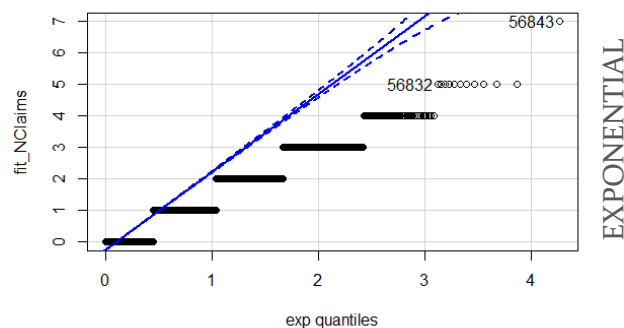
For the Number of Claims, given that it did have crucial values equal to 0, the main 3 distributions it could be modelled after are the exponential, poisson and geometric distributions. Using a Monte Carlo Simulation of the data we are given we can determine suitable values for each distribution. This simulation consisted of using the “fitdistr” function in R to determine a suitable rate, lambda and mean and standard deviation for the 3 distributions respectively and simulating 100,000 data points for each distribution. These data points can then be used in a qqplot to compare with the original data and determine which distribution is the best fit for our data.



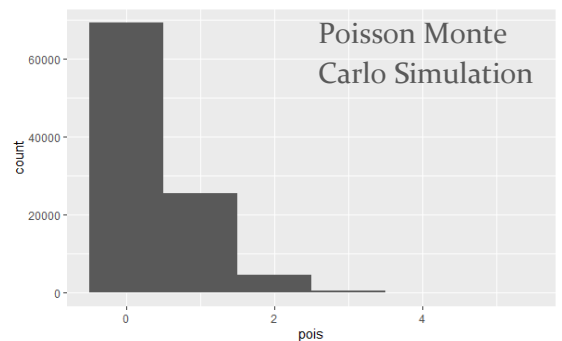
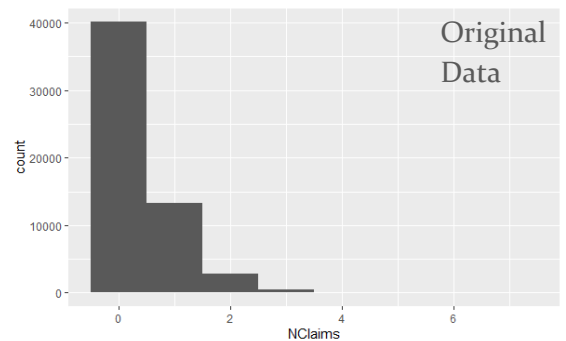
From these qqplots we are able to see that the poisson distribution with $\lambda = 0.3666006266$ provides the best fit to our data and should be used to model pricing for SuperCar Insurance.



GEOMETRIC



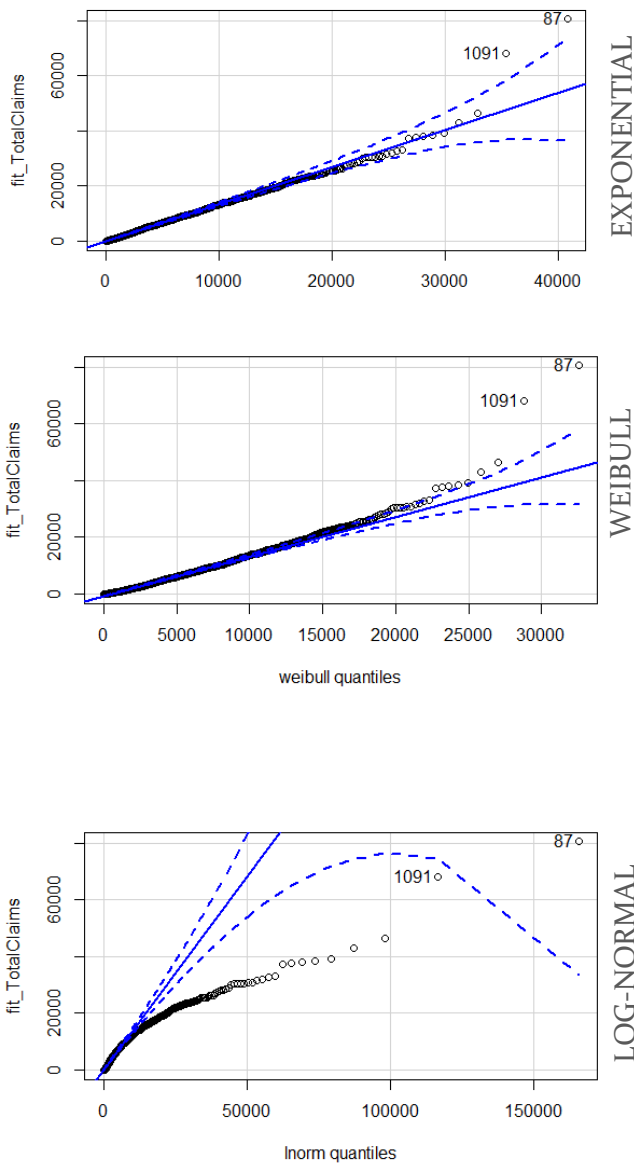
EXPONENTIAL



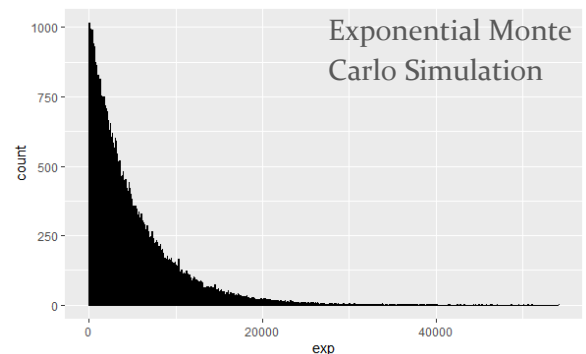
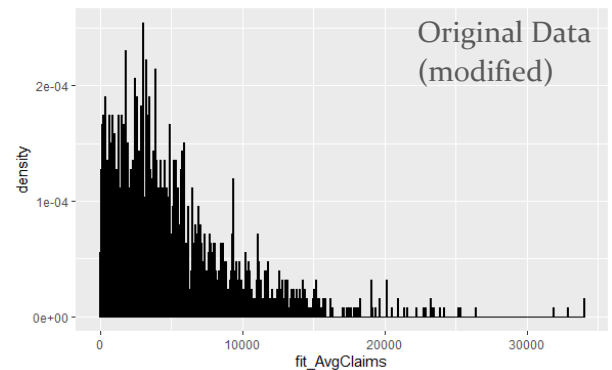
Total Claim Amount

For the Total Claim Amount, other than the minimum value of \$0.00, the next highest claim amount is \$300.00. There are 97 claimants with \$300.00 insurance claims which leads me to set a point of exceedance of \$300.00. Therefore we can focus on non-zero values and take away \$300.00 from them as this is the amount the claimant pays in excess for each insurance claim anyway, meaning all \$300.00 claims are of zero loss to SuperCar. Using these remaining, trimmed, non-zero values we can estimate exponential, Weibull and lognormal distribution models for this data.

Again, like for the Number of Claims, using a Monte Carlo Simulation of the data we are given we can determine suitable values for each distribution. These data points can then be used in a qqplot to compare with the original data and determine which distribution is the best fit for our data.



From these qqplots we are able to see that the exponential distribution with rate = 0.0002021486 provides the best fit to our data and should be used to model pricing for SuperCar Insurance.



Conclusion

SuperCar Insurance's claims data from last year is able to give us a inside view of the comprehensive car insurance industry in Australia and the different factors which can easily affect it. The way that a claimant's age, gender, marital status and even population of the city they live in affects the number and cost of claims made by them shows why premium pricing is specific for each person.

From the statistical analysis performed we found that males and single people were more likely to make an insurance claim than their counterparts. We also found that increasing age also tended towards a slightly upward sloping line of increased insurance claims, similar to that of increasing population. Although our data was skewed in terms of population due to there being less claimants from higher populations (none in the 80000-90000 range) it did show an increasing trend.

By using Monte Carlo analysis to determine suitable models to measure the number of claims made in a given year and the total amount of those claims in a given year we are also able to specify a model which can be used to forecast future trends and required premium prices. Through this analysis of SuperCar Insurance's claim data from last year overall it is able to be noted how thorough the insurance field must be to deliver insurance products to people.

References

Ibisworld.com. 2020. *Ibisworld - Industry Market Research, Reports, And Statistics*.
[online] Available at: <<https://www.ibisworld.com/au/industry/car-insurance/4122/>>