**INTRODUCTION**

The average insurance expenditure nationwide tends to rise from one year to another in the face of technological advancement, new automobiles, government policy, etc. it is therefore imperative to be proactive in measuring the probability of insurance claims and cost liable in the future and to ascertain policy premiums that will be suitable for both the policy maker and the policyholder.

As insurance policy is well known as a binging between the insurance provider and the insured to cover a percentage cost in a likely damage, accident, theft of property, therefore it is highly imperative to mitigate unforeseen cost by creating a suitable policy that will be guided by predictable clauses.

This project aimed to forecast the likelihood of claims, severity, frequencies, and anomalies by engaging the machine learning predictive model in creating a cost-effective insurance premium policy.

**Requirements:**

* Get an insight from 10,000 instances and 18 features of car insurance data to create predictive model to suit a new premium policy.
* We aim to reduce claims severity, frequency, and anomalies,
* Understand the probability of an insured making a claim.

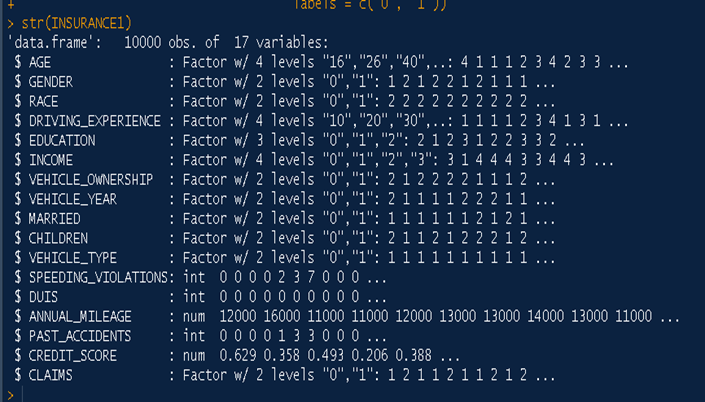
**Technique:**

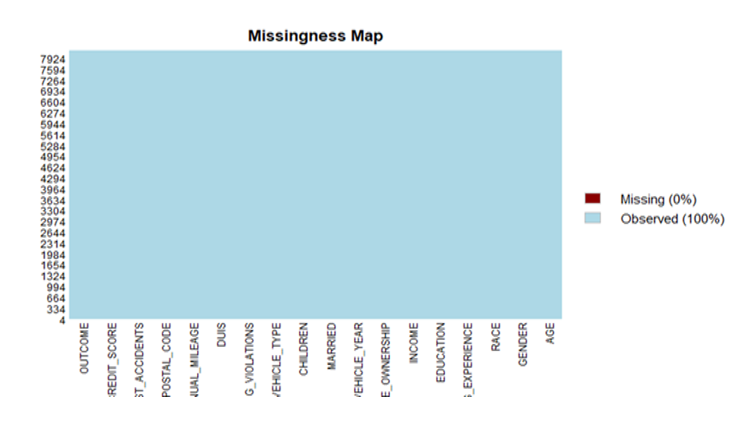
* Data Preparation and Cleaning
* Exploration
* Statistical Analysis Using R studio.
* Predictive Model Using R studio (Comparing Models using Excel)
* Conclusions

**DATA PREPARATION AND CLEANING**

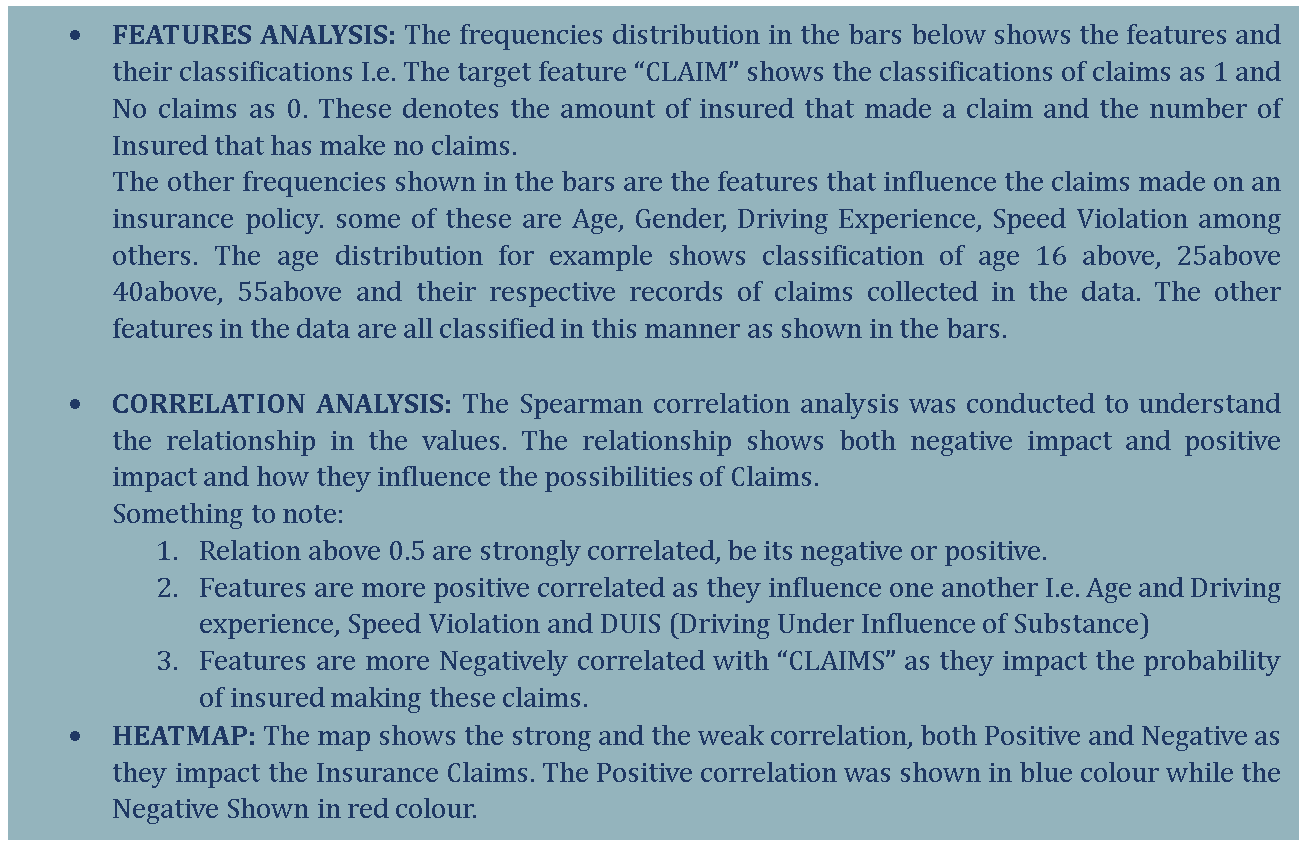
**A blue text on a white background

Description automatically generated**





**EXPLORATORY ANALYSIS**

****

A diagram of a number of gray squares

Description automatically generated

A screenshot of a graph

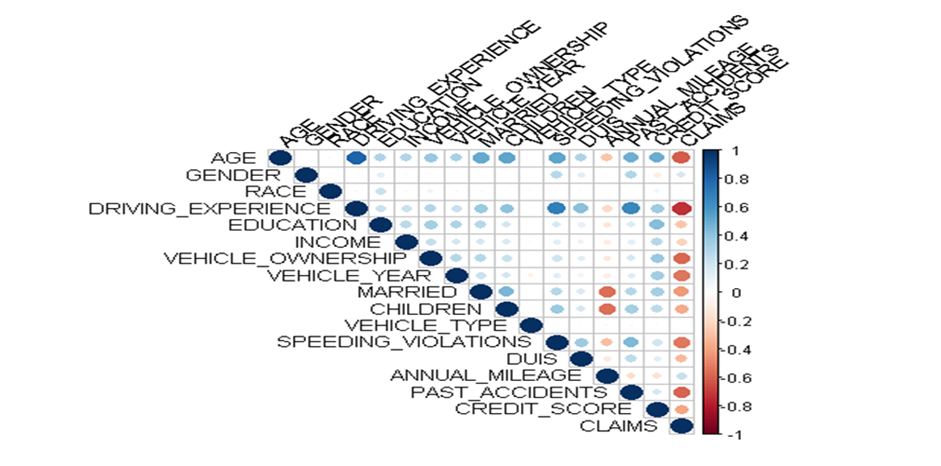
Description automatically generated

A screenshot of a computer screen

Description automatically generated

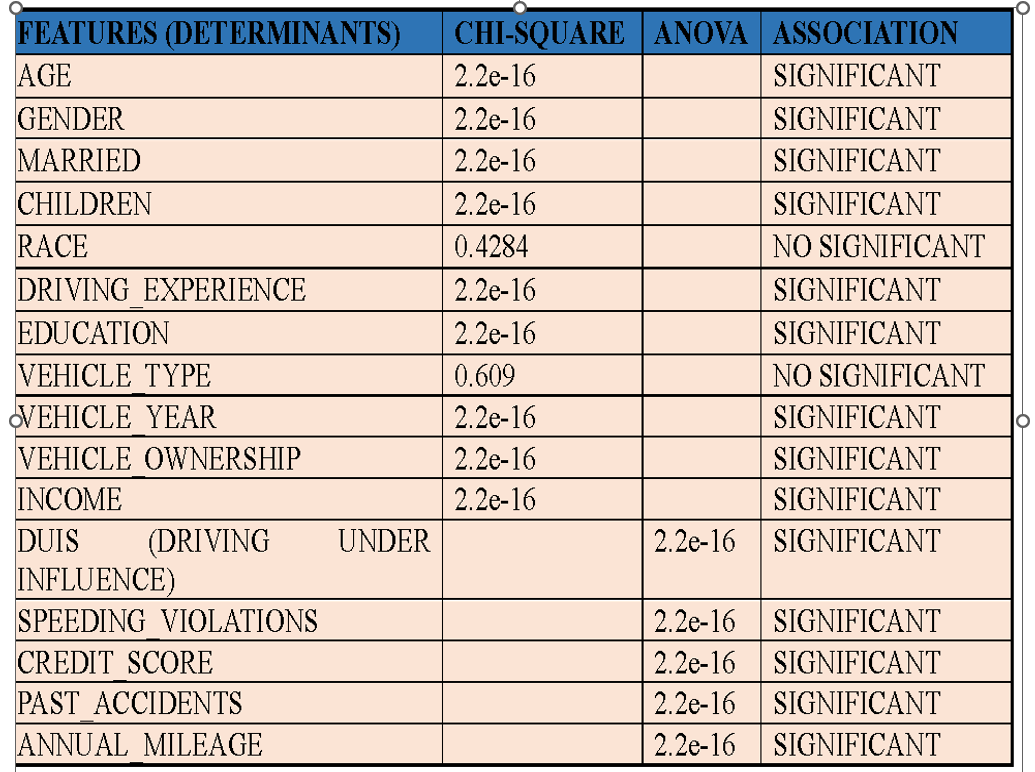
A screenshot of a graph

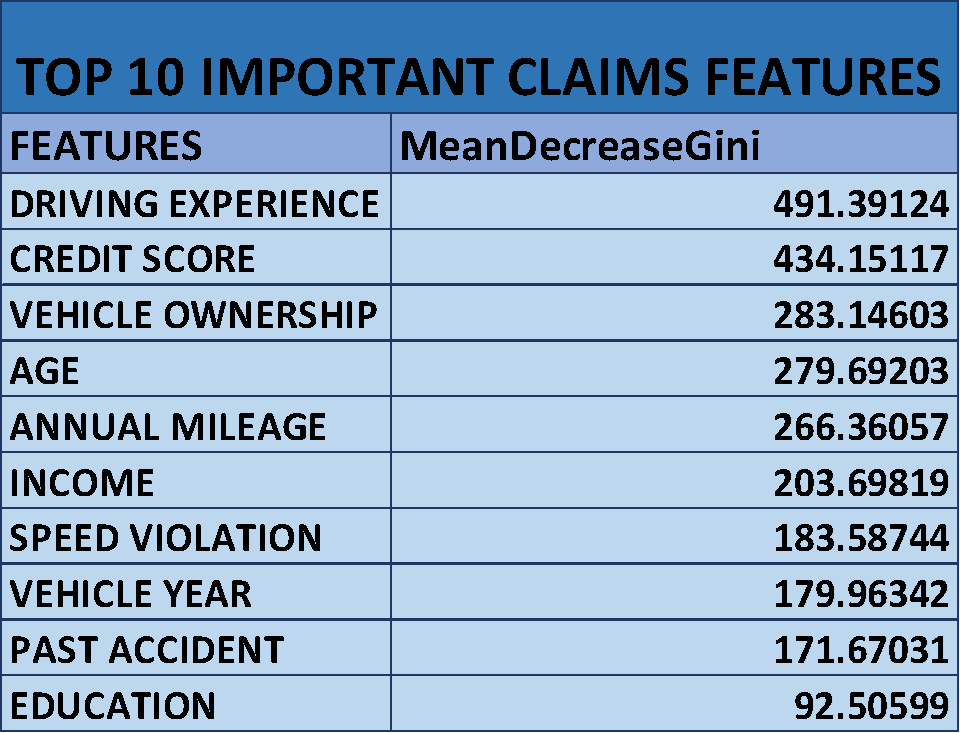
Description automatically generated

****

**STATISTICAL ANALYSIS**

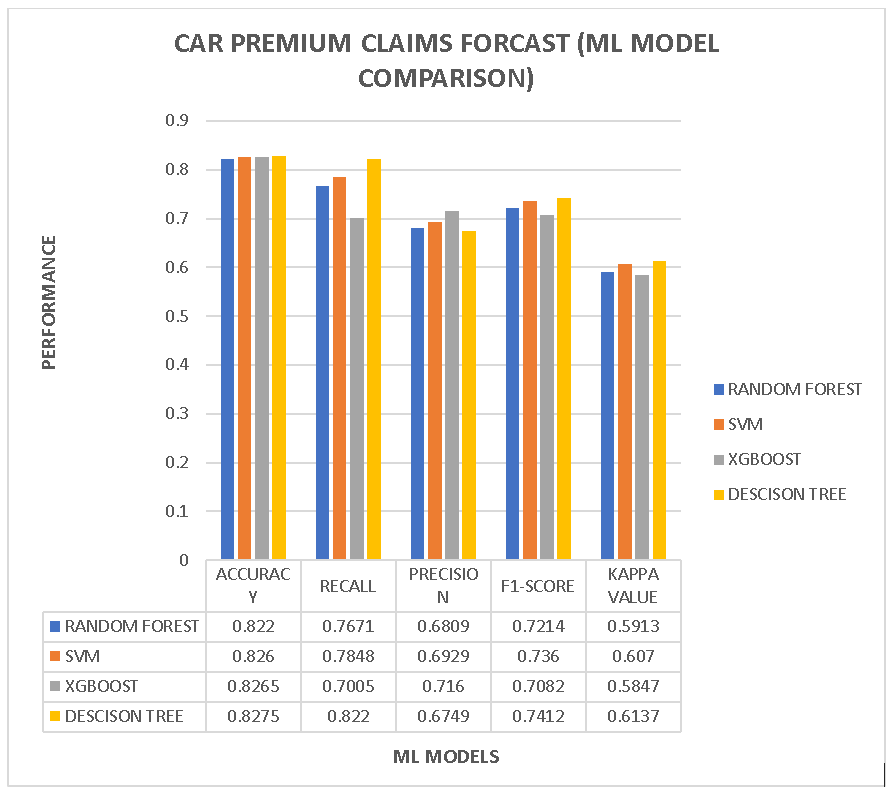
* **Hypothesis ANALYSIS:** To understand the significance of every feature that impacts the probabilities of claim, and hypothesis is conducted to analyse the important of each feature with Claims feature. The table of importance is shown below.
* **Features Significance:** Using the statical analysis, Top 10 features of are selected according to the impact the poses to car insurance claims. The table of Feature importance a is shown below.





**MACHINE LEARNING MODELS**

The project was conducted comparing four machine learning model to achieve consistency of prediction. The Random Forest, Support Vector Machine (SVM), Decision Tree and XGboost were compared to arrive at prediction of Claims. All the prediction resulted in similar prediction showing that a customer will make a claim on an insurance premium considering different feature listed in the statistical model, The chance of a customer making a claim is 83% on any premium considered. It is therefore imperative to consider all features related to individual customer history while considering the suitable premium in order to mitigate loss incurables.



**CONCLUSION**

Establishing a car insurance premium can be a challenging operation for an insurance company, it is therefore necessary to consider every historical option particular to every customer before a premium is established. The engagement of Machine learning and Artificial Intelligence is highly important in distinguishing probabilities, and it is highly important to the future of insurance in order to create a cost effect insurance strategy for every insurance company.