**Correlation Analysis of Customer Profile Contribution to Insurance Profitability**

**Problem Statement:**

The current study aims to conduct customer segment analysis to identify the contribution of different customer profiles to its profit margin.

**Sub Problems:**

Question 1:What are the different customer profiles?

Question 2: How to assess profit margin?

Question 3: How to evaluate a profile contribution to the company profit margin?

Question 4: What recommendation to give for each customer profile to maximize customer retainment and profit margin.

**Hypothesis:**

Hypothesis 1: Identify all elements that pertain to the characteristics of a customer.

Hypothesis 2: Identify several pertinent metrics to assess profit margin (disregards fraudulent claims).

Metric 1: Claim amount versus total premiums paid.

Metric 2: Frequency of claims versus customer profiles.

Metric 3: More fraudulent customer profiles.

Hypothesis 3: Metric 1 as Net Value against the customer profiles (density function) together with the average yearly premium paid for each of the customer profiles.

Hypothesis 4: Hypothesis 3 outcome interpretation

**Benefits of the study:**

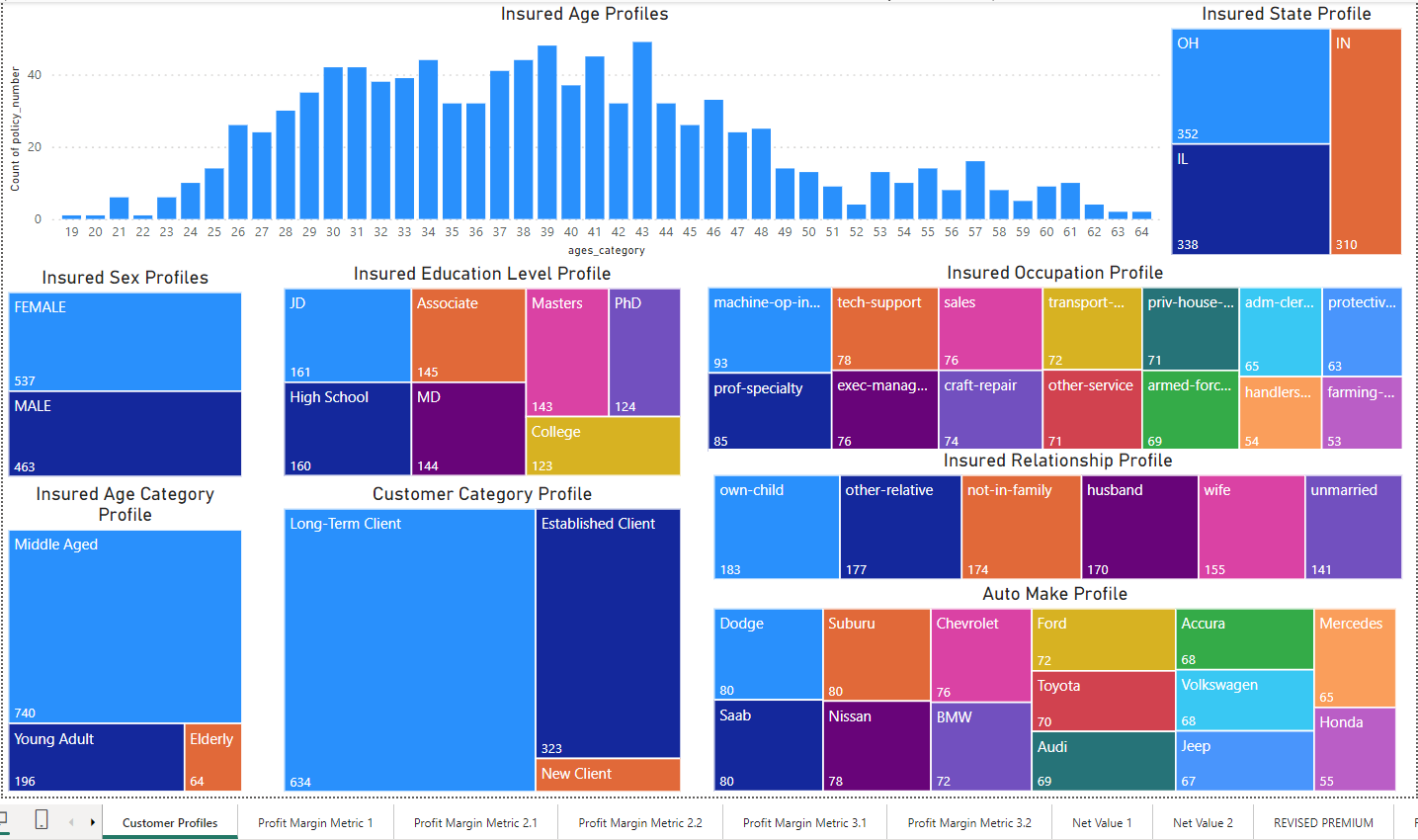
1. Assist the company to identify the most profitable customer profiles.
2. How to support all different customer groups and maximize retainment and profitability.
3. Identify more fraudulent profiles.
4. Assist the company to reevaluate the annual insurance premium for specific customer profiles.

**Outcome:**

* Hypothesis 1: Identify all elements that pertain to the characteristics of a customer.

Customer profiles were identified which included age, occupation, sex, insured state, education level, insured automobile make etc.

Here on can see for instance the age range of all customers from 19 – 64 years, with most customers in the mid-30s to mid-40s.



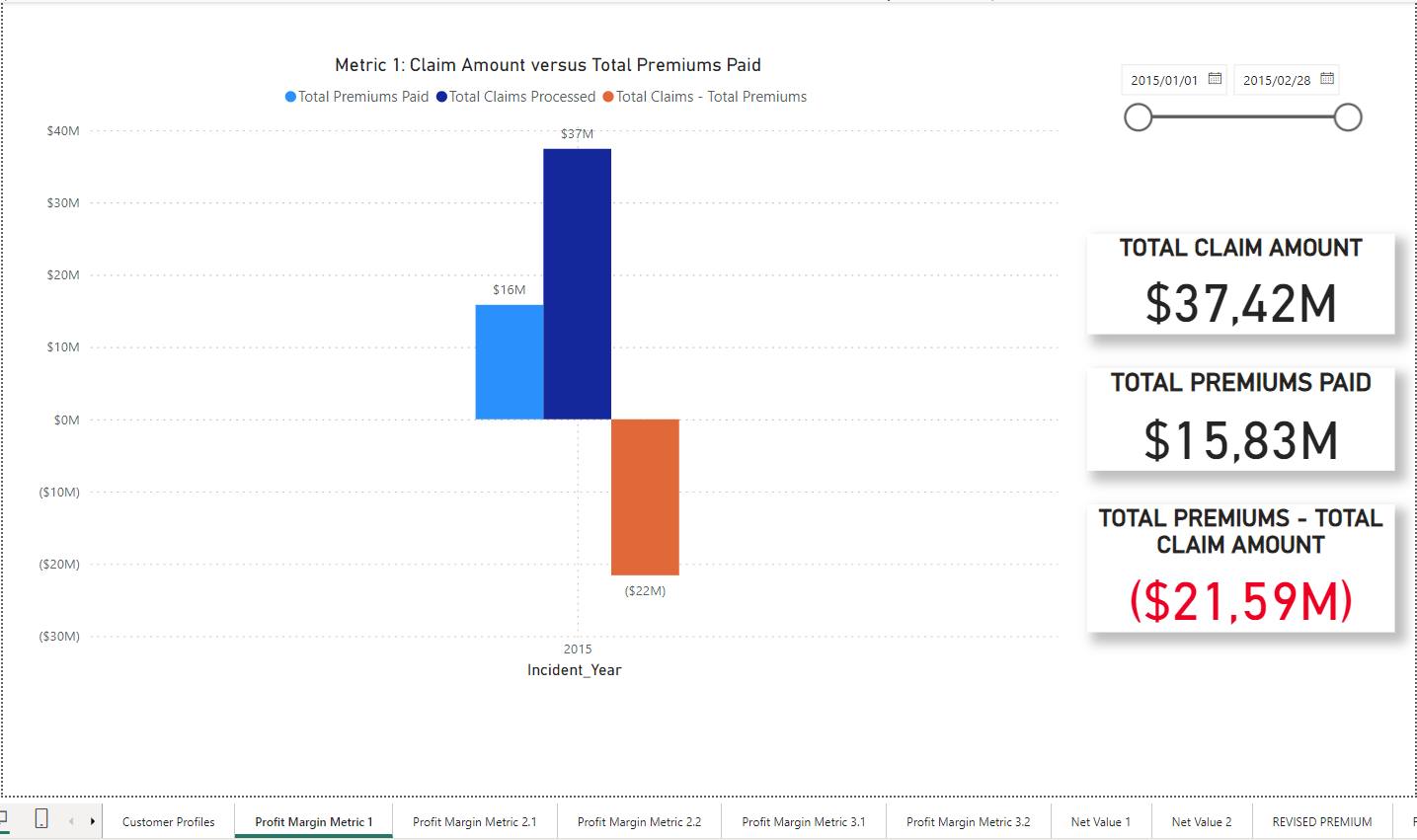
* Hypothesis 2: Identify several pertinent metrics to assess profit margin (disregards fraudulent claims).
  + Metric 1: Claim amount versus total premiums paid.

The total claim amount from 2015/01/01 to 2015/02/28 is shown as 37M.

The total premiums paid for the same duration is shown as 15.83M.

The total premiums – total claim amount is shown as -21.59M.

Only non-fraudulent claims were considered.



* + Metric 2: Frequency of claims versus customer profiles.

From the graph below one can see the frequency distribution for different profiles. The age distribution is skewed to the right which means that the younger age groups have a higher number of claims. Females claim more compared to their male counterparts. High School students claim more compared to the other education levels. Long term clients have much more claims compared to new clients and there are the most claims for a Jeep Wrangler compared to a BMW X6 with the least number of claims.

A screenshot of a graph

Description automatically generated

A screenshot of a graph

Description automatically generated

* + Metric 3: More fraudulent customer profiles.

From the graphs below one can see which customer profiles submit the most fraudulent claims. These profiles can be flagged, and claims investigated before any payout is processed. This will assist in increasing the company’s profit margin as less fraudulent claims will be paid out.

A screenshot of a graph

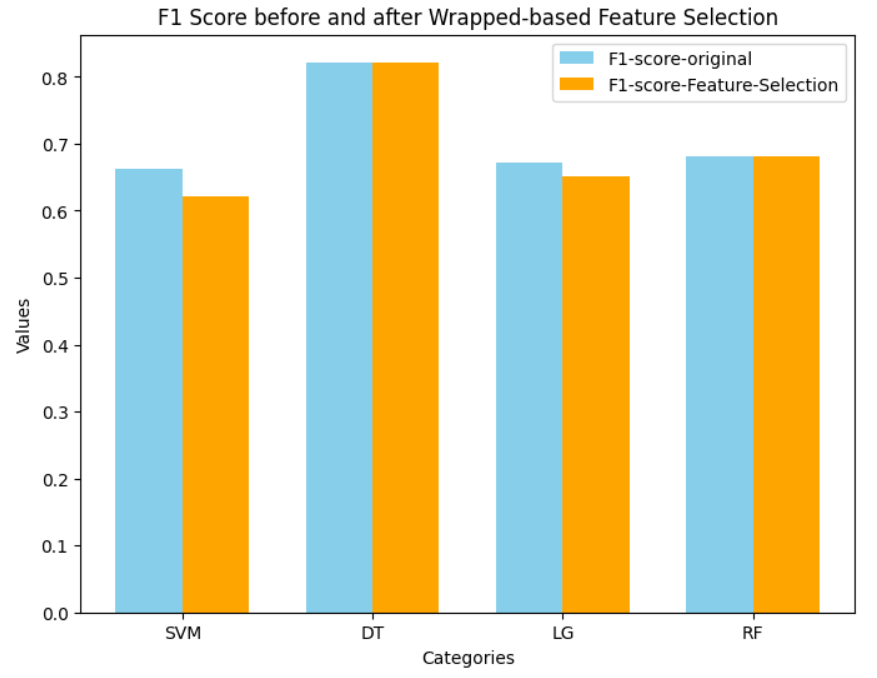
Description automatically generated

A screenshot of a graph

Description automatically generated

To further assist the company the labelled data provided was modelled to predict which customer profiles are more likely to submit a fraudulent claim. Since the target variable is categorical (yes/no), only classification models were used for supervised learning.

The Sequential Feature Selector wrapper-based method was used to see if the model’s performance could be increased. The figure below shows that the Decision Tree model had the best predictive power with a F1-score above 80% based on the test data.



A number of numbers in a row

Description automatically generated with medium confidence

A blue squares with white text

Description automatically generated

Based on the provided classification report for the Decision Tree (DT) model:

The model achieves relatively high precision for class 'N' (0.86), indicating that when it predicts an instance as 'N', it is correct about 86% of the time. However, the precision for class 'Y' (0.74) is slightly lower, suggesting that there might be some misclassification of instances predicted as 'Y'.

The recall for class 'N' is high (0.90), indicating that the model effectively captures a large proportion (90%) of actual instances of class 'N'. However, the recall for class 'Y' is relatively lower (0.66), indicating that the model misses a significant portion (34%) of actual instances of class 'Y'.

The F1-score for class 'N' is high (0.88), suggesting a good balance between precision and recall for this class. However, the F1-score for class 'Y' is lower (0.70), indicating a trade-off between precision and recall for this class.

The overall accuracy of the model is 0.82, which is relatively good. However, accuracy alone may not provide a complete picture, especially in imbalanced datasets.

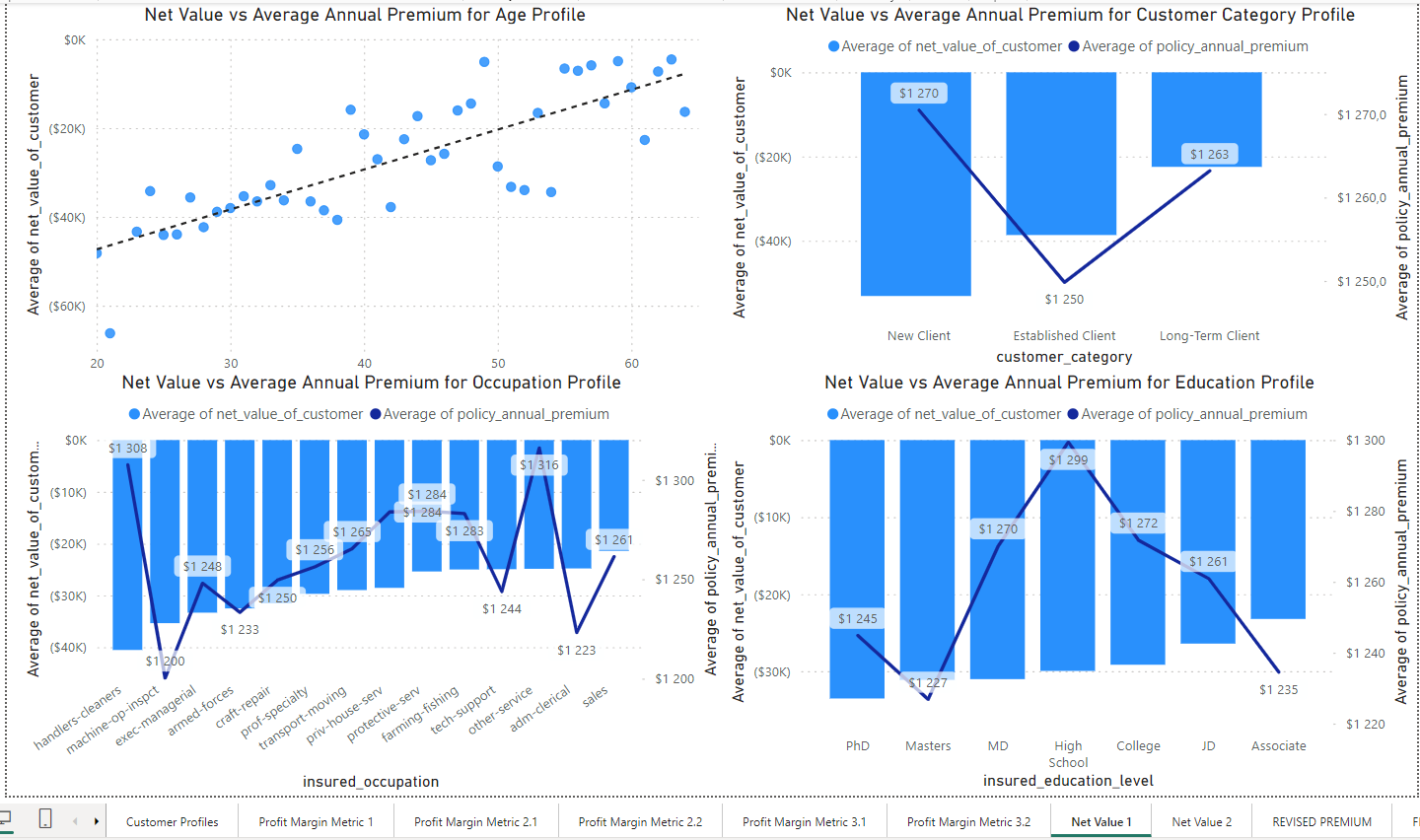
The macro-averaged precision, recall, and F1-score are slightly lower than the weighted averages, indicating some imbalance in class distribution.

Overall, the model performs reasonably well, with high precision and recall for class 'N' but lower recall for class 'Y'. Depending on the specific context and requirements, further optimization or exploration may be needed, particularly to improve the model's performance on predicting instances of class 'Y'.

* Hypothesis 2: Metric 1 as Net Value against the customer profiles (density function) together with the average yearly premium paid for each of the customer profiles.

The graph below shows that the net value is most negative for the lower age clients and increases linearly. This means that the lower age clients claim more than what was contributed through premiums. The average premium for new clients is $1270, whereas for long-term clients it is $1263. This equates to a small difference of $7. The premiums for new clients should be adjusted to minimize the loss encountered.

PhD and master’s students have the highest negative net value but have the lowest annual premium. Females have the highest negative net value, but have a lower annual premium compared to males.



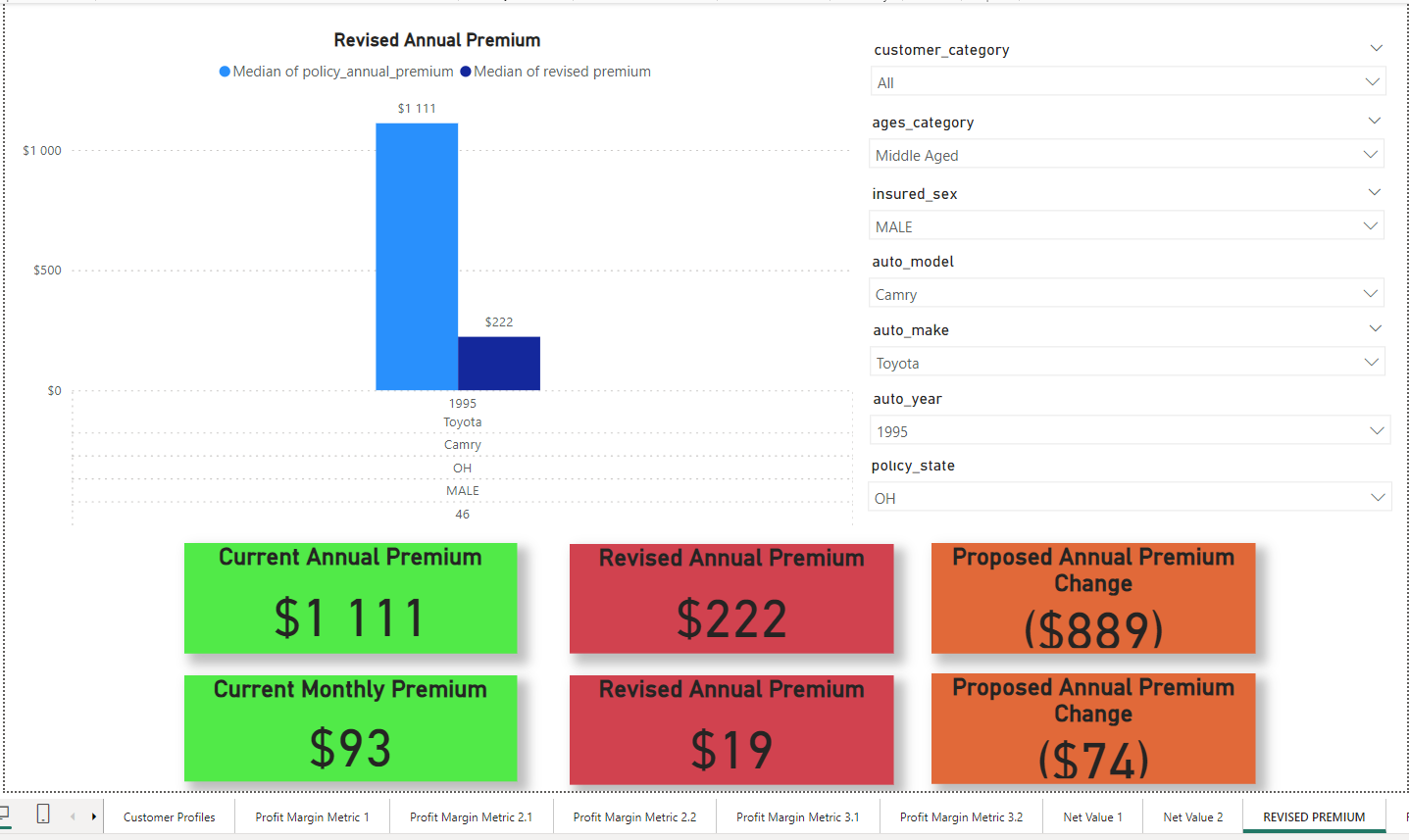
A screenshot of a graph

Description automatically generated

* Hypothesis 4: Hypothesis 3 outcome interpretation - What recommendation to give for each customer profile to maximize customer retainment and profit margin.

The revised annual premium tab can be used to influence annual premium adjustments for different customer profiles.

In the visual below it is shown that the annual premium for a middle-aged male driving a 1995 Toyota Camry in the state of OH, can actually be reduced with $889 per annum or $74 per month. This will assist with retainment of customers in this profile. We have seen that males have a lower negative net value and also middle-aged customers have a lower negative net value, thus one should expect a decrease in the annual premium paid.



In the following visual it is shown that the annual premium for a young female driving a 1997 Jeep Grand Cherokee in the state of IL, should be increased with $663 per annum or $53 per month. This might not assist with customer retainment, but with improving the profit margin of the company. Each case can be considered individually, and it might be beneficial to increase the excess paid rather than increasing the annual premium by too much of a margin. We have seen that females have a higher negative net value, and younger customers have a higher negative net value, thus one should expect an increase in the annual premium paid.

A screenshot of a computer

Description automatically generated

**Recommendations:**

1. The average customer net value of the age profile follows a linear trend line with the younger clients having higher negative net values. The inference made is that the younger clients are newer clients and have not paid sufficient premiums to overcome the claim amounts. The recommendation is that new clients’ (<25 months) excess payments should be higher or annual premiums be adjusted according to the ‘Revised Premium’ tab.
2. It is noted that High School students have the lowest average annual premium, although High School students have the highest number of claims. High School students also fall within the new client’s category as per point (1). It is recommended that the High School student client’s annual premium be adjusted with the associated risk.
3. In terms of client occupation, clients that are travelling more have the highest number of claims and negative net value, while office bound clients have less claims and a lower negative net value. The recommendation is that clients whose occupation requires travelling should have a revised annual premium which is associated with the risk of travelling. Clients who do not require travelling for work purposes can have a reduced annual premium, which will also retain more clients in this category.
4. The automobile make has the highest average negative net value for BMW, Dodge and Ford, while Saab, Volkswagen and Toyota have the lowest negative net value. Clients that insure more expensive cars should have a revised annual premium, while clients with lower valued cars can have a reduced annual premium, which will also retain more clients in this category.
5. The ‘Revised Premium’ tab can be utilized to assist the company on which customer profiles require readjustment on the annual premium paid. This will assist in increasing the companies profit margin, retain current customers and gain new customers through competitive market related premiums.
6. The Decision Tree model developed can be utilized to predict if a specific customer profile has submitted a possible fraudulent claim. These predictions can prevent the company to pay out possible fraudulent claims and therefore assist in increasing the company’s profit margin.