Abstract:

This project focuses on a dataset that describes auto claims filed by customers of an automobile insurance company located in the southwest and western regions of the United States.

For an insurance business to be profitable, their pricing projections must be precise and accurate. Given the costs of their clients' reimbursements for accident repairs, total loss car replacements, medical expenses, and legal fees, the auto insurance plans must be priced so that the insurance company earns a profit over the long term.

The goal of this project and research is to conduct an exploratory data analysis on the claims dataset and then prepare an executive summary of the key findings and suggestions for the insurance company's executive management to implement in the company to bring forth positive outcomes in the business.

This project used R language and RStudio to analyze and perform exploratory data analysis on the dataset and draw meaningful inferences from the dataset.

Dataset Information

The claims data being used for the purpose of this project consists of information of 6,249 auto claims that have been submitted by customers of a single insurance company. Every single row in this dataset represents the attributes of a single claim filed by a customer.
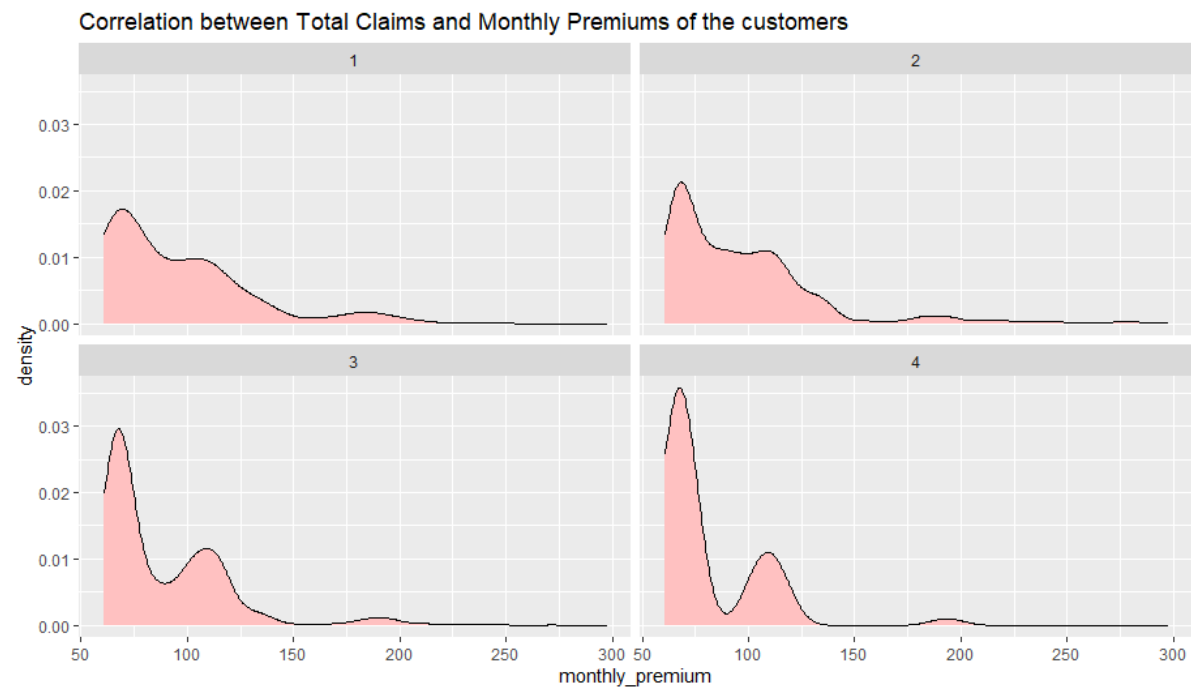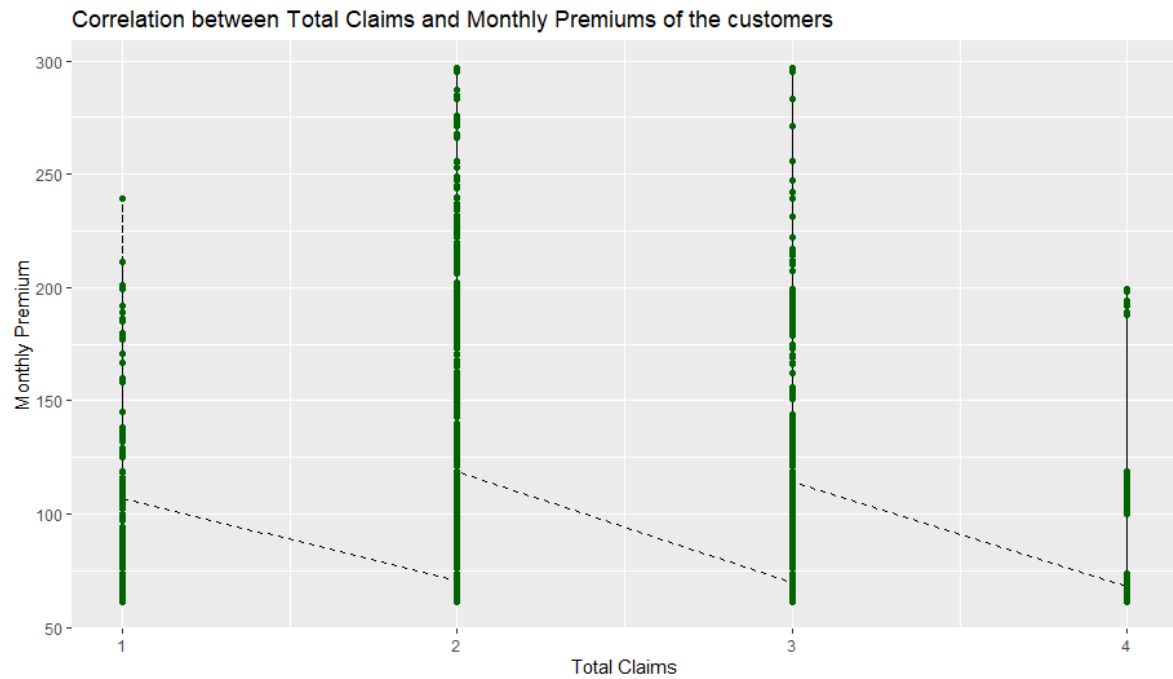
Question 1

Question: Is there a correlation between the total number of claims and the monthly premium, indicating that the premium could being increased after every claim?

Answer:

It is customary for an insurance company to increase the monthly premium that a customer has to pay after every claim that they make, in order to ensure that the customers do not make unnecessary and frequent claims and therefore do not cause a loss to the business. In this particular company however, this is not the case. The monthly premiums for all the customers, irrespective of their total number of claims is about the same. In fact, the premiums seem to be decreasing as the number of claims of the customers increases.

The same can be demonstrated in both the plots below.

Correlation between Total Claims and Monthly Premiums of the customers

Correlation between Total Claims and Monthly Premiums of the customers

The below summary table illustrates the same information that has been discussed above. The monthly premium keeps getting lower as the number of claims of the customers increases.

```
# A tibble: 4 x 6
  total_claims count min_premium avg_premium max_premium sd_premium
         <dbl> <int>       <dbl>       <dbl>       <dbl>      <dbl>
1            1   219          61        95.6         239       34.6
2            2  3709          61        97.3         297       37.3
3            3  1975          61        89.3         297       32.0
4            4   346          61        80.7         199       24.7
> |
```

Question 2

**Question:** Does a person being married makes them more likely to be on a Premium coverage?

Answer:

From the summaries obtained from this dataset using bar plots and mosaic plots, it is clear that majority of the customers who are taking up policies at this insurance company are married. This might be attributed to the higher stability in life that married customers have and their propensity to own an automobile.

It can be seen from the below graphs that across all the types of policies and coverages, it is the married customers that have the highest number of policies and thus play a significant role in the income generation for the company.

Effect of Marital Status on the type of coverage
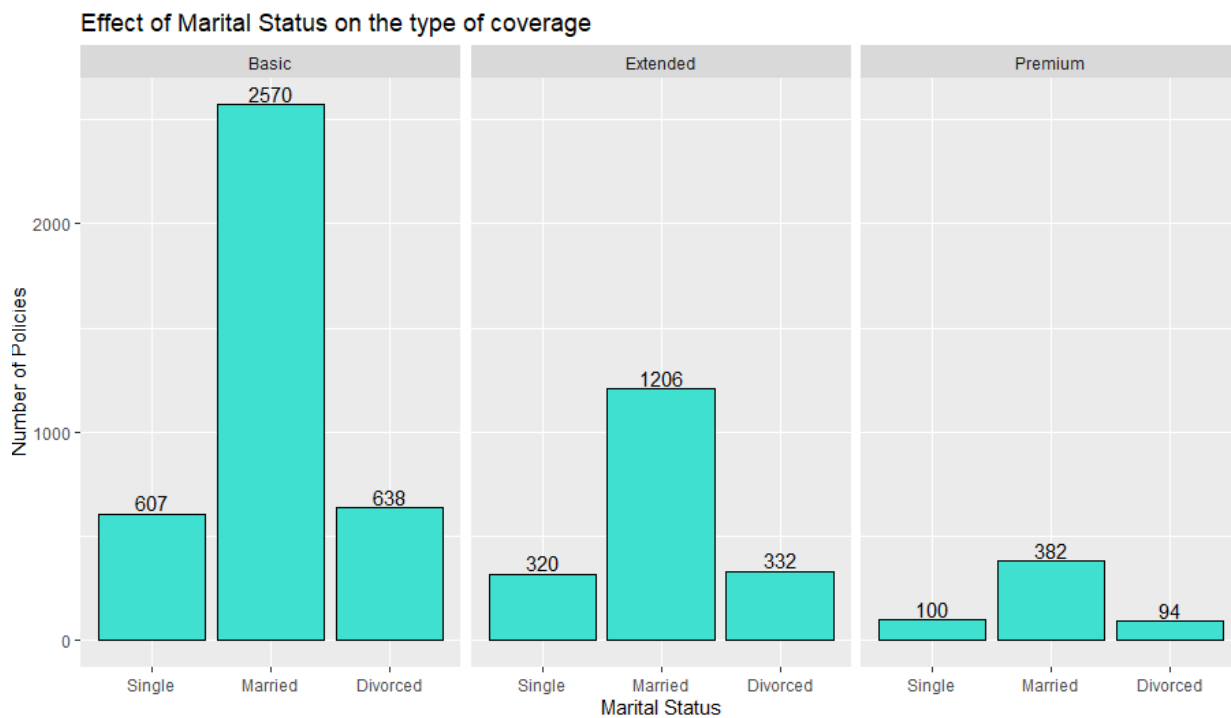
It is the same case for both the type of policy and the type of coverage, and as can be seen in both the graphs, married people take up most of the policies across all policy and category types.

Effect of Marital Status on the type of policy
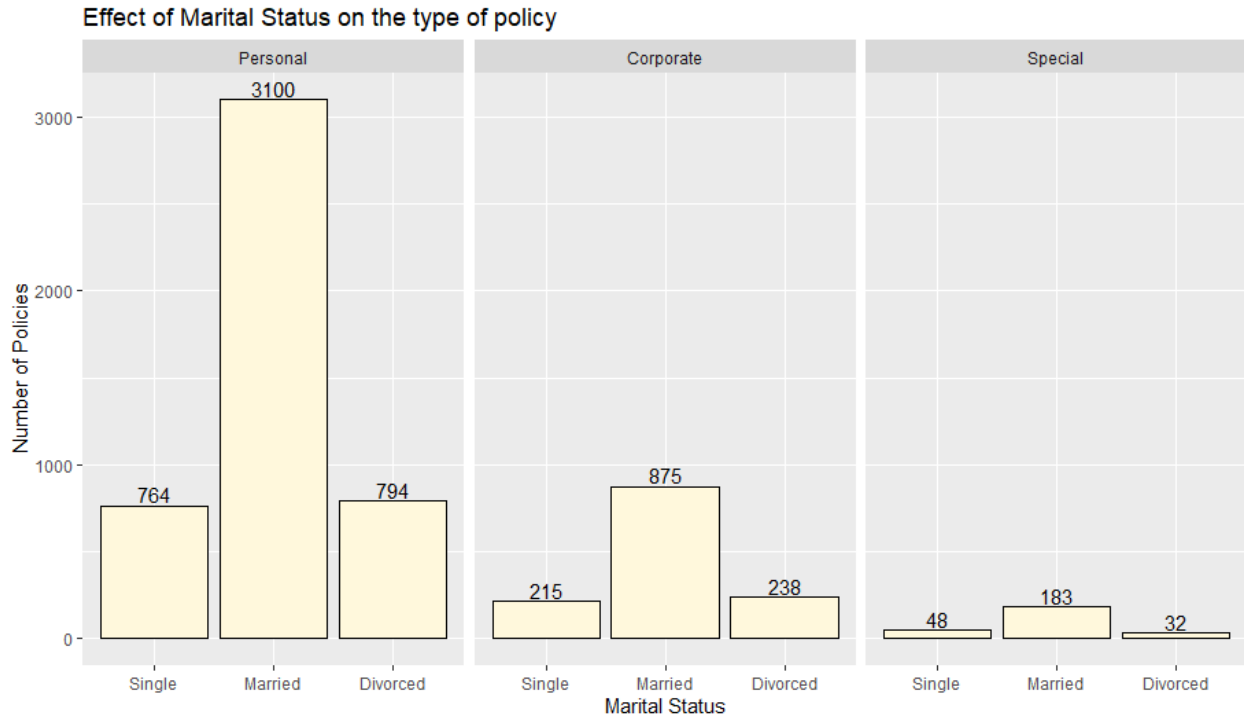
The same inference can be drawn from the below mosaic plot as well, where married customers take up most of the policies at the insurance company.


Effect of Marital Status on the type of coverage
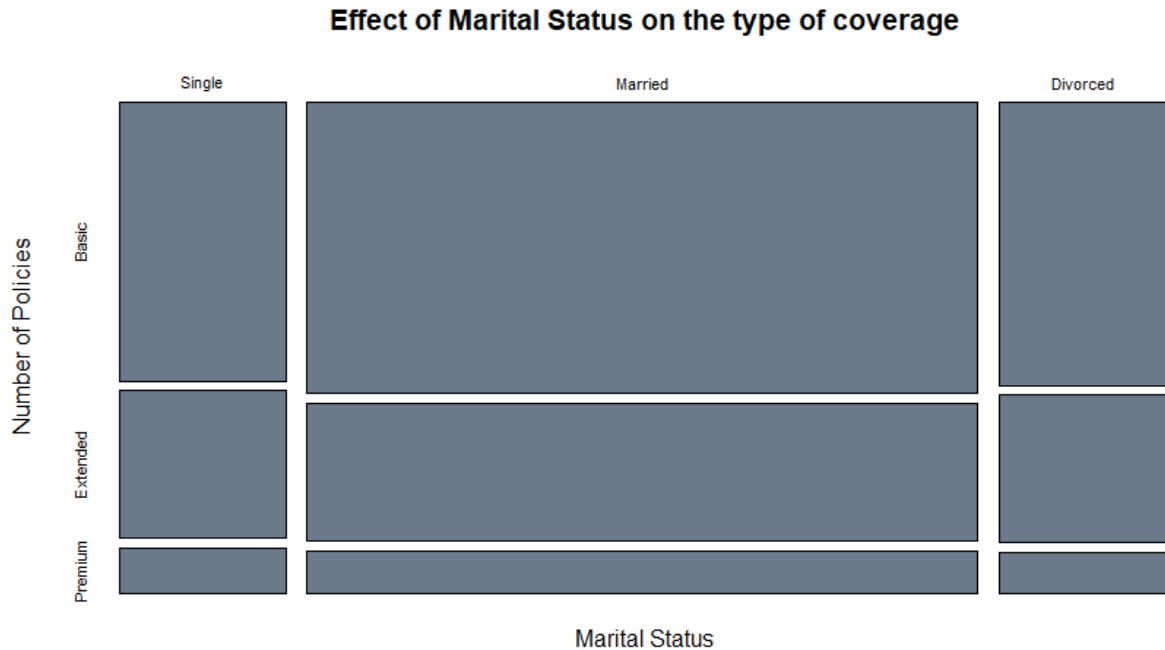
Question 3

Question: Is there a positive correlation between the type of vehicle that a person owns and their value to the company?

Answer:

A customer's lifetime value is calculated by taking all their defining attributes into consideration. As such, it is a measure of the value that the customer brings to the company and their overall profitability to the insurance company. Customers with higher lifetime values are preferred for the benefit of an insurance company.

With the case of this insurance company, it can be seen from the below graph that customers who own the top two vehicle classes have significantly higher lifetime values as compared to the customers who own other types of vehicles.

Relation between the Customer's vehicle class and their total lifetime value
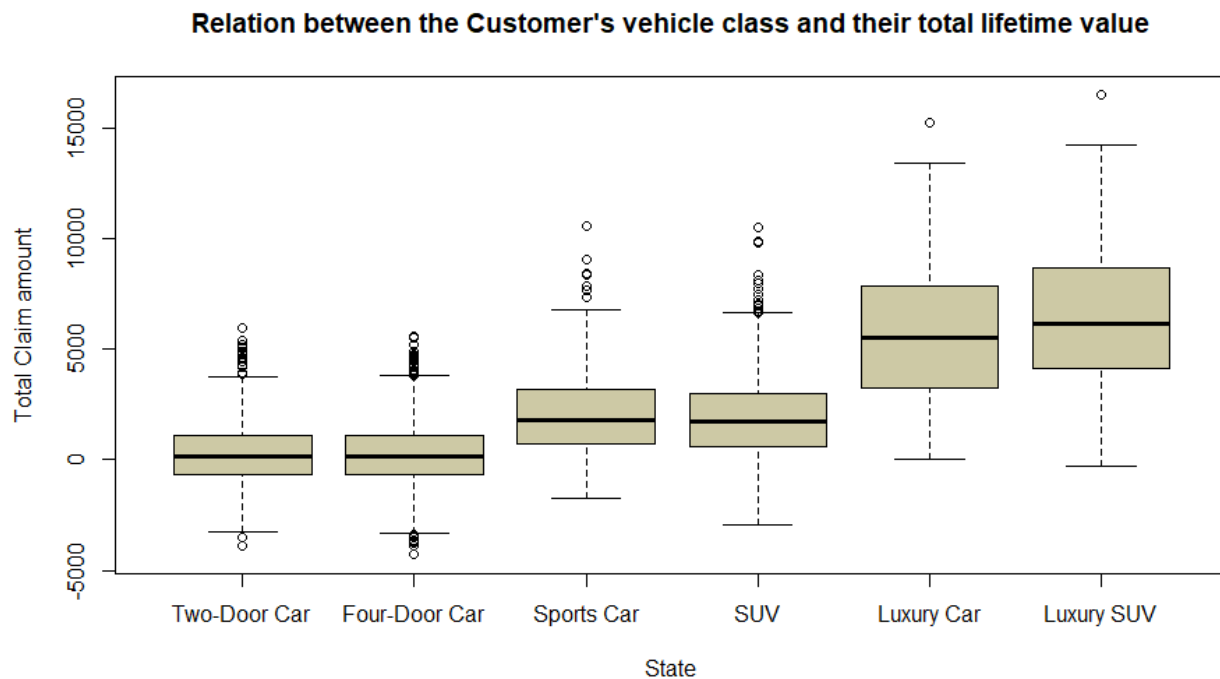

The same information can be demonstrated from the below summary table:

The type of vehicle that a customer owns bears a significant positive impact on their overall lifetime value to the company, judging from the significantly higher average lifetime values for the customers who have the top two vehicle types.

```
# A tibble: 6 x 6
  vehicle_class count min_lifetime_value avg_lifetime_value max_lifetime_value sd_lifetime_value
  <fct>         <int>              <dbl>              <dbl>              <dbl>             <dbl>
1 Two-Door Car   1292              -3890               269.               5941             1406.
2 Four-Door Car  3124              -4285               271.               5587             1371.
3 Sports Car      335              -1740              2159.              10588             1999.
4 SUV            1246              -2934              1861.              10525             1826.
5 Luxury Car      119                 60              5670.              15204             3106.
6 Luxury SUV      133               -287              6382.              16465             3297.
>
```

The below heatmaps explore the relation between the vehicle type of a customer and their lifetime value to the company, with respect to their type of policy and their type of coverage.

It can be seen for both the cases that the customers with the top two vehicle classes and the top type of policy and coverage have a higher lifetime value as compared to the remaining customers.

Relation between the vehicle type of a customer and their policy type

Relation between the vehicle type of a customer and their coverage type
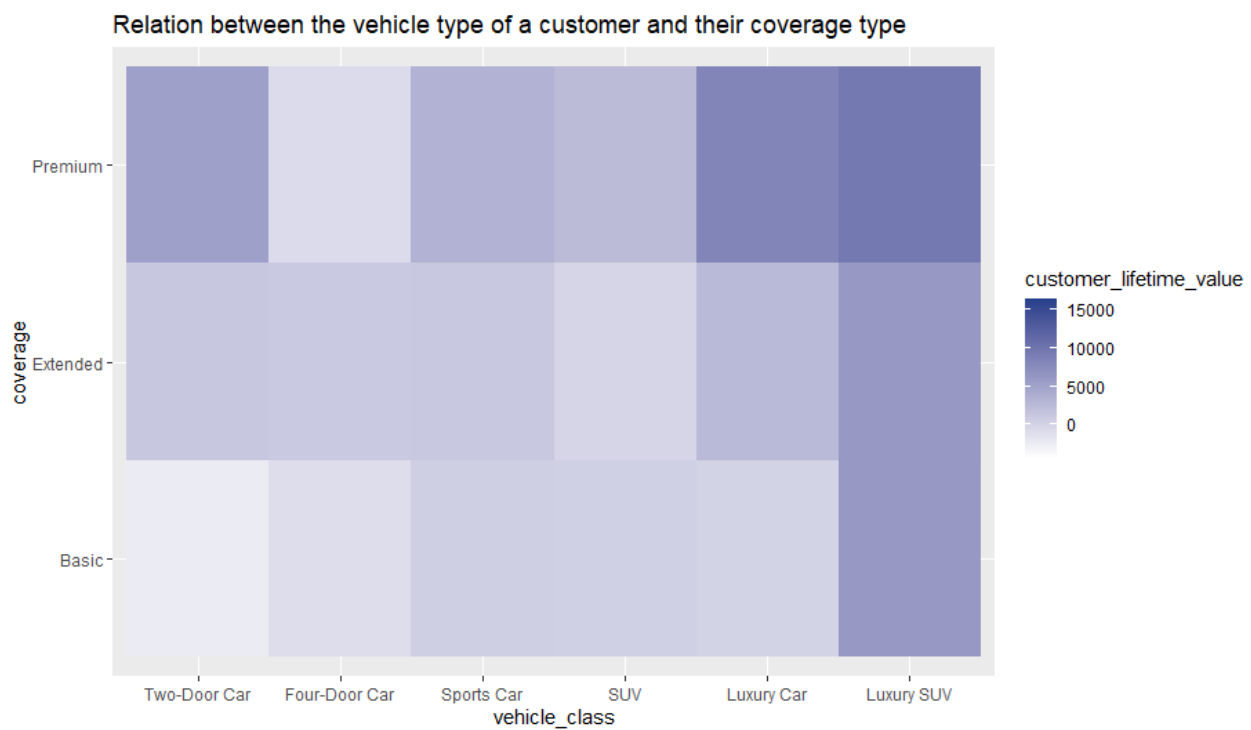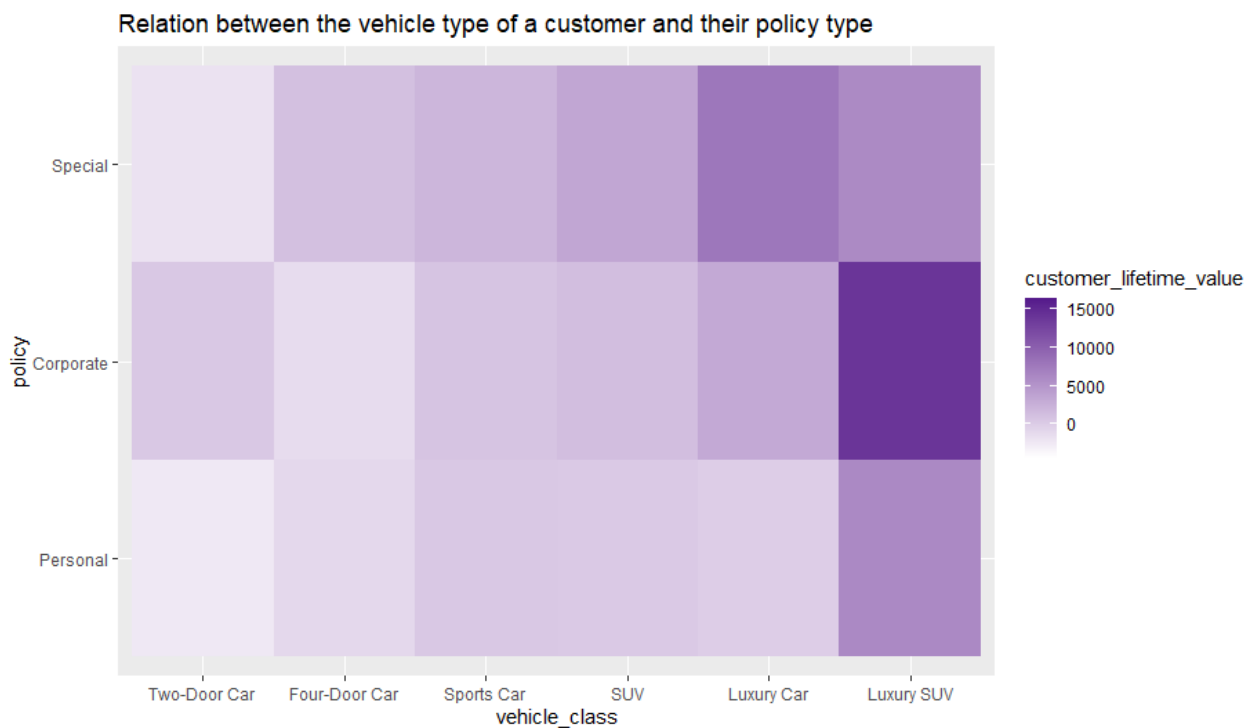
Question 4

Question: Does a person being employed have an impact on them claiming their policies less frequently, thus making them more profitable?

Answer:

It is generally assumed that a person who is employed is less likely to claim their policies too often and therefore would be preferable as customers to an insurance company. The same can be said about this company as well, inferring from the density plots and the heat map below.

It can be seen that the customers with their employment status as "Medical Leave" and "Disabled" have the highest claim amounts of all the other categories.
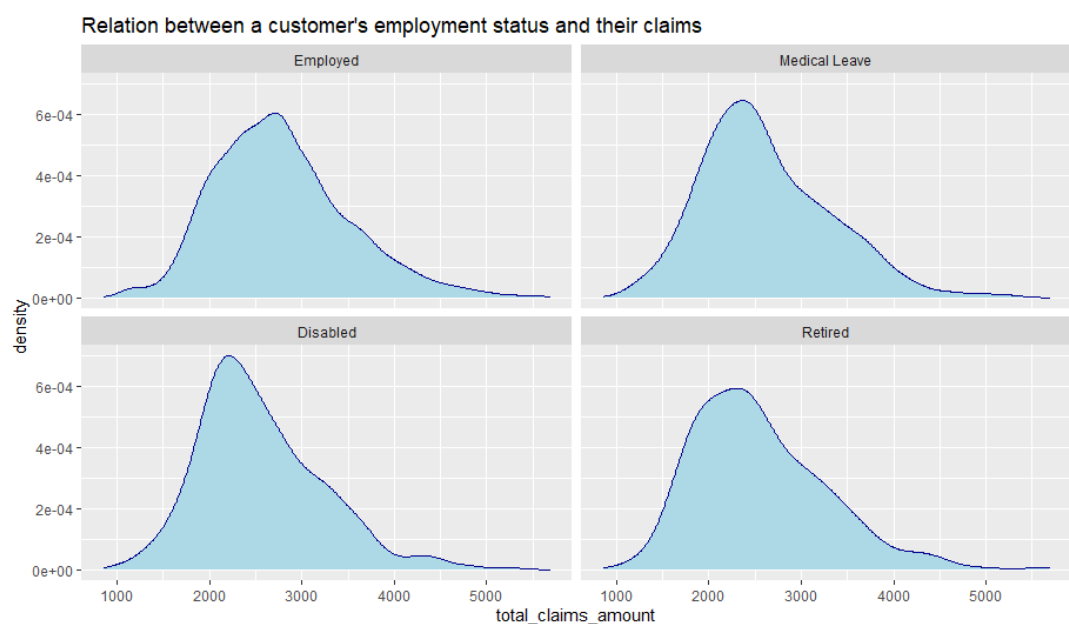
The same can be deduced from the below summary table as well.

It can be seen that on average, people who are disabled and on medical leave have the most number of claims.

```
# A tibble: 4 x 6
  employment_status count min_claims avg_claims max_claims sd_claims
  <fct>             <int>      <dbl>      <dbl>      <dbl>     <dbl>
1 Employed           5154          1       2.39          4     0.648
2 Medical Leave       421          1       2.42          4     0.641
3 Disabled            392          1       2.44          4     0.676
4 Retired             282          1       2.36          4     0.611
> |
```

Question 5

Question: Can the customers from a particular state be called more profitable to the company, based on their premiums and claims?

Answer:

Based on the geographical demographic, the customers of this company who are from the states of Nevada and Arizona have lower total claimed amounts as compared to customers from other states, thus being more beneficial for the company from a financial standpoint.

Relation between the Customer's state and their total claim amount
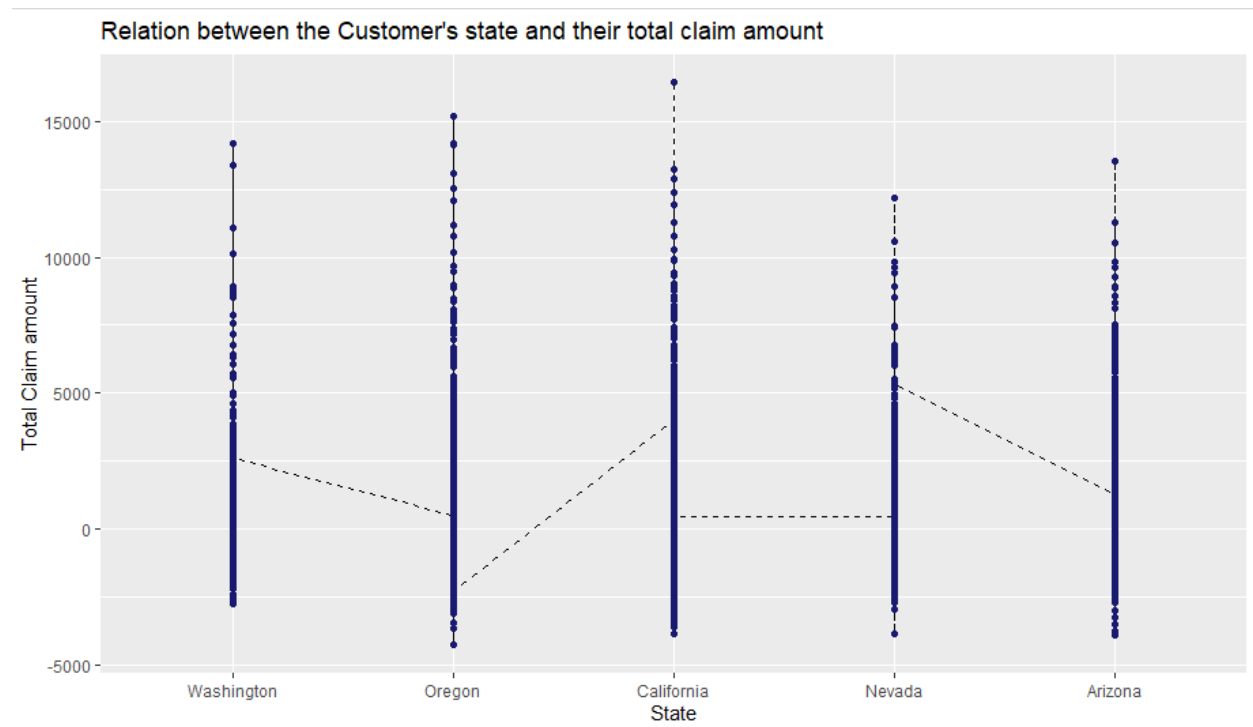
However, interestingly, the statistics of the monthly premiums being paid by all the customers, irrespective of their state, are fairly similar, as can be demonstrated from the below boxplots.

Relation between the Customer's state and their total monthly premium
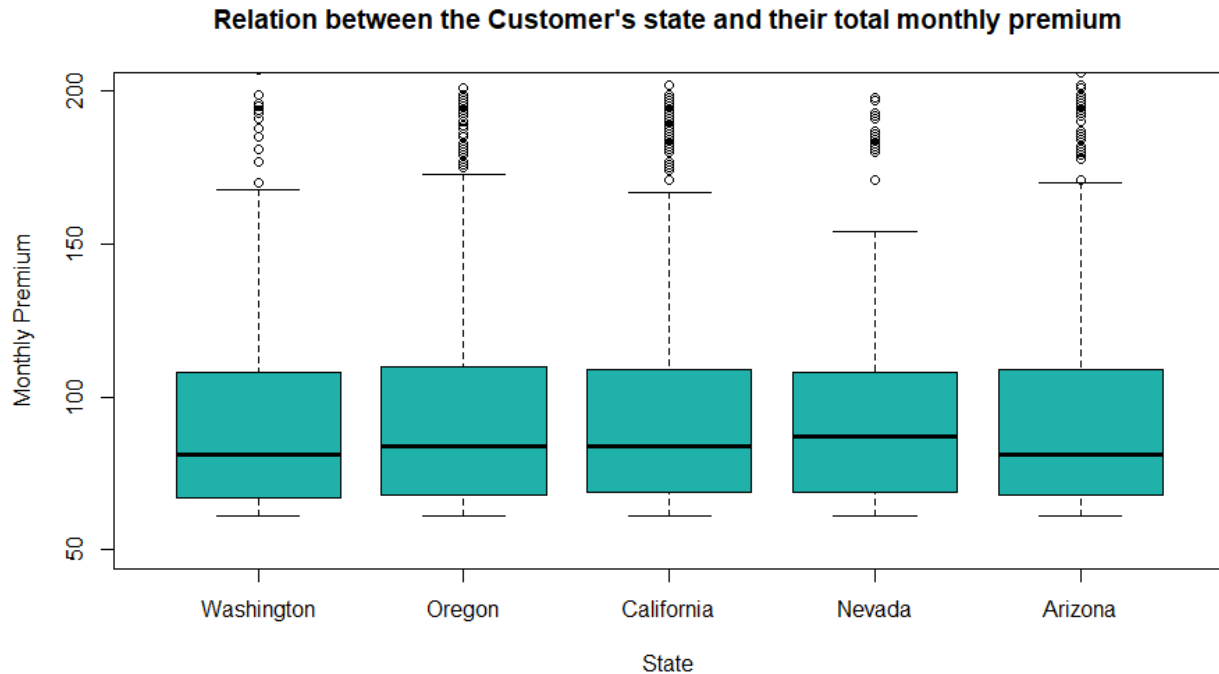
Thus for similar premiums, because the total claim amounts for the customers from the states of Nevada and Arizona are higher than that of other states, they have significantly higher lifetime values as compared to the customers from other states. The same can be demonstrated from the avg_lifetime_value column in the below summary table.

```
# A tibble: 5 x 6
  customer_state count min_lifetime_value avg_lifetime_value max_lifetime_value sd_lifetime_value
  <fct>          <int>              <dbl>              <dbl>              <dbl>             <dbl>
1 Washington       554              -2741               867.              14203             2058.
2 Oregon          1763              -4285               947.              15204             2091.
3 California       2150              -3890               883.              16465             2059.
4 Nevada           601              -3850               971.              12185             2086.
5 Arizona         1181              -3911               956.              13542             2024.
> |
```

Question 6

Question: Are certain coverage types more profitable, due to having a lower number of claims?

Answer:

From the below summary table and boxplots, it can be admitted that the policy type "Basic" has significantly higher number of customers that have claimed their policies more frequently as compared to the other two policy types.

```
# A tibble: 3 x 6
  coverage count min_claims avg_claims max_claims sd_claims
  <fct>    <int>      <dbl>      <dbl>      <dbl>     <dbl>
1 Basic     3815          1       2.54          4     0.704
2 Extended  1858          1       2.17          3     0.464
3 Premium    576          1       2.15          3     0.455
> |
```

Relation between the policy type and total number of claims

Question 7

Question: Do lower education levels of the customers imply that they are not as profitable, referring to their total claim amount and the total number of their claims?

Answer:

It can be seen that the education level of the customers has an impact on the total number of claims that they have taken from the insurance company. The customers with higher education levels ("Master", "Doctoral") have taken lower number of claims on average, as compared to the customers with lower education levels ("High School", "Associate", "Bachelor").

Relation between the highest education of a customer and their number of claims

The below heatmap explores the relation between the education level of a customer and their total number of claims with respect to the total amounts that they have claimed.

As a general pattern, it can be observed that the customers with higher education levels have claimed lower amounts, thus being more profitable to the company.

Relation between the highest education of a customer and their number of claims
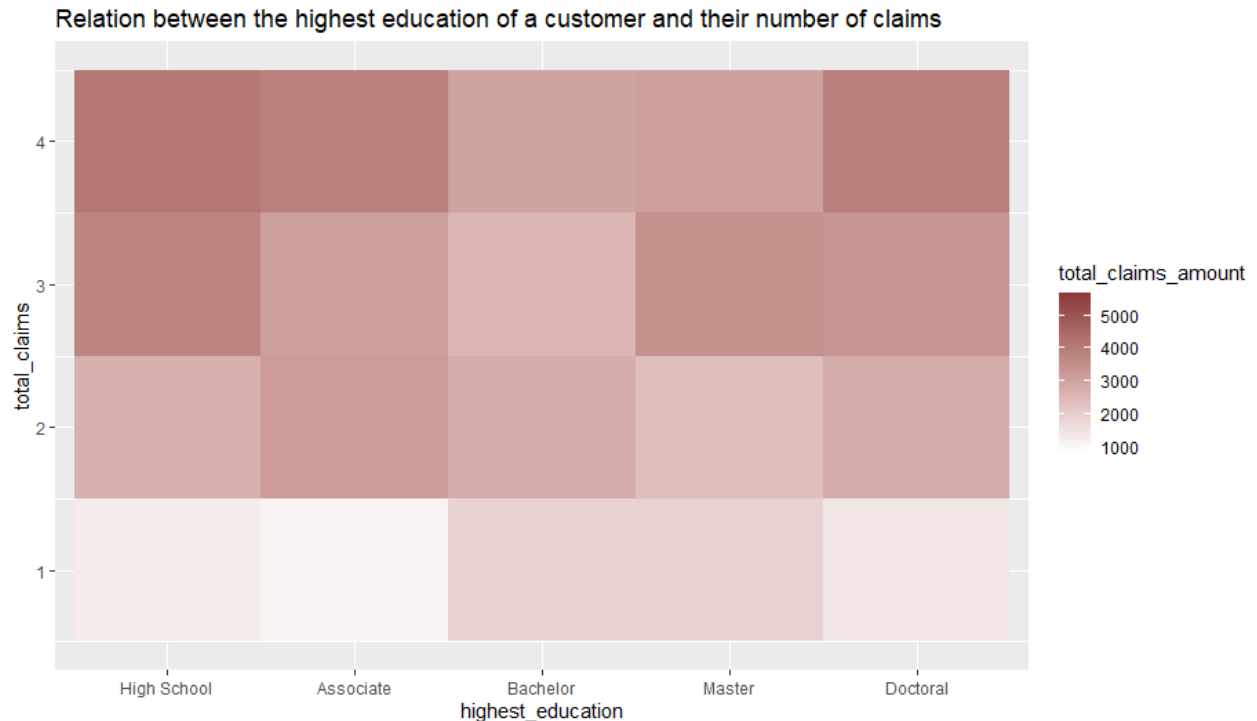
The same information can be conveyed through the below summary tables: that customers with higher education levels tend to claim their policies less frequently.

```
# A tibble: 5 x 6
  highest_education count min_claims avg_claims max_claims sd_claims
  <fct>             <int>      <dbl>      <dbl>      <dbl>     <dbl>
1 High School        1708          1       2.38          4     0.655
2 Associate          1843          1       2.41          4     0.657
3 Bachelor           1868          1       2.39          4     0.643
4 Master              578          1       2.36          4     0.613
5 Doctoral            252          1       2.40          4     0.633
> |
```

```
# A tibble: 5 x 6
  highest_education count min_claim_amount avg_claim_amount max_claim_amount sd_claim_amount
  <fct>             <int>            <dbl>            <dbl>            <dbl>           <dbl>
1 High School        1708             1022            2749.             5653            731.
2 Associate          1843              859            2772.             5656            737.
3 Bachelor           1868              989            2754.             5688            728.
4 Master              578              953            2622.             5522            673.
5 Doctoral            252             1092            2733.             5006            710.
> |
```
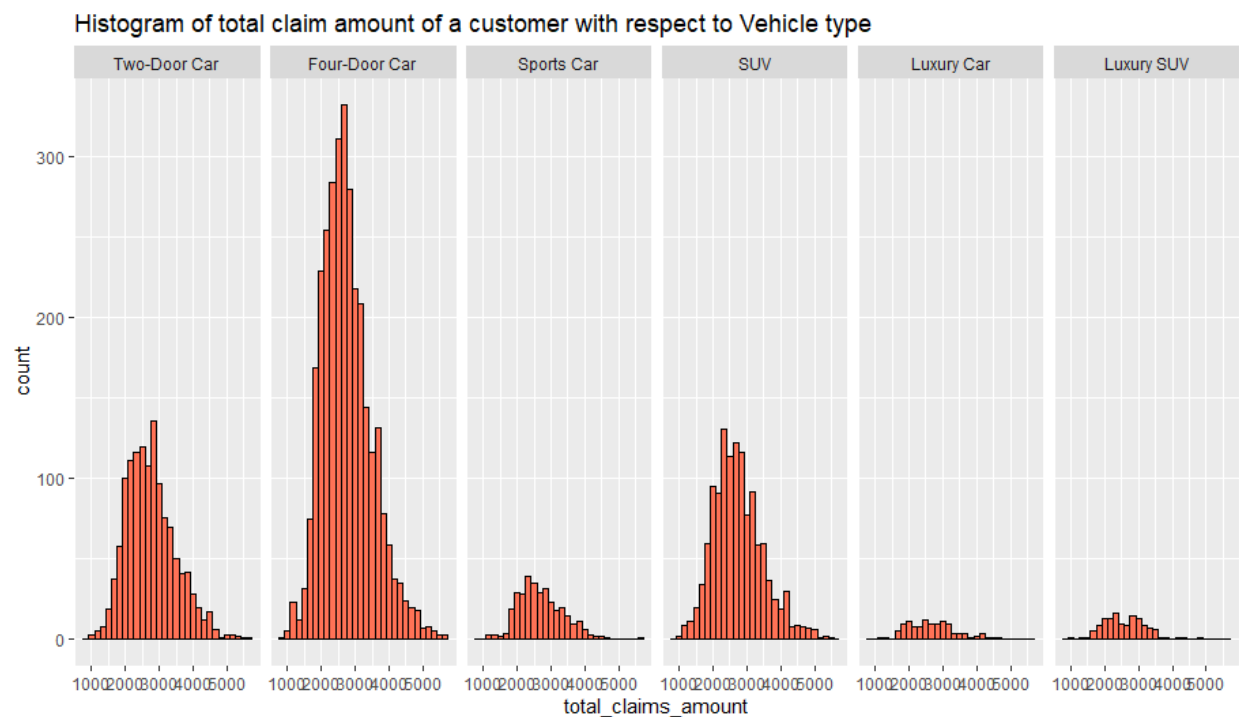
Question 8

Question: Is there a significant profit margin between the customers with different types of vehicle classes? Compare the top two vehicle classes with the rest to determine this.

Answer:

The below histograms and summary tables convey the relationships between the vehicle classes of the customers and their general value to the company.

The average claim amount of a customer goes down as the class of their vehicle increases.

```
# A tibble: 6 x 6
  vehicle_class count min_claim_amount avg_claim_amount max_claim_amount sd_claim_amount
  <fct>         <int>            <dbl>            <dbl>            <dbl>           <dbl>
1 Two-Door Car   1292              953            2757.             5606            729.
2 Four-Door Car  3124              859            2755.             5688            733.
3 Sports Car      335             1095            2702.             5593            674.
4 SUV            1246             1001            2740.             5553            734.
5 Luxury Car      119             1141            2692.             4659            716.
6 Luxury SUV      133             1066            2584.             4788            618.
> |
```

Histogram of total claim amount of a customer with respect to Vehicle type

The average monthly premium of a customer increases as the class of their vehicle increases.

```
# A tibble: 6 x 6
  vehicle_class count min_premium avg_premium max_premium sd_premium
  <fct>         <int>       <dbl>       <dbl>       <dbl>      <dbl>
1 Two-Door Car   1292          61        77.2         119       14.9
2 Four-Door Car  3124          61        77.5         119       14.7
3 Sports Car      335         100       123.          199       20.5
4 SUV            1246         100       120.          199       18.4
5 Luxury Car      119         180       213.          297       33.9
6 Luxury SUV      133         180       218.          287       29.9
> |
```

Histogram of monthly premium with respect to Vehicle type
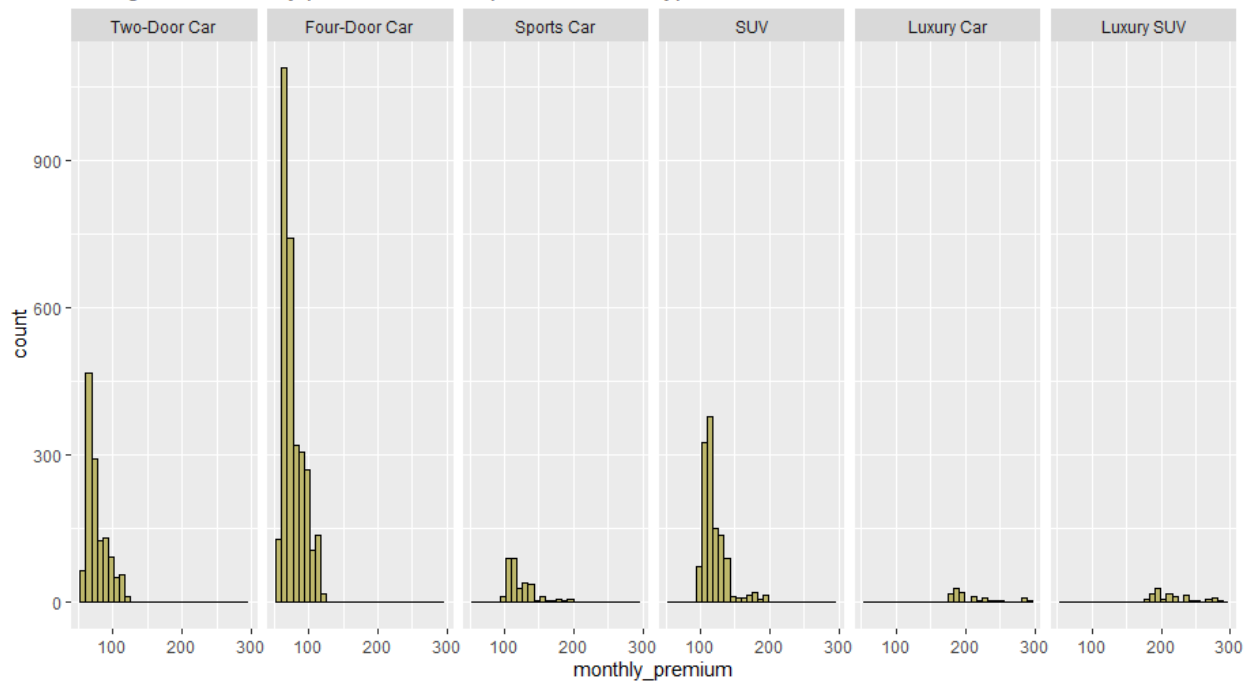
As the class of the vehicle increases, the customers tend to keep their policies alive for longer without making a claim, as can be seen in the below histograms. For the lower vehicle classes, there are a lot of policies that have been claimed frequently, as the number of months that the policy has been alive for is lower for a significantly higher number of policies for the lower vehicle types.

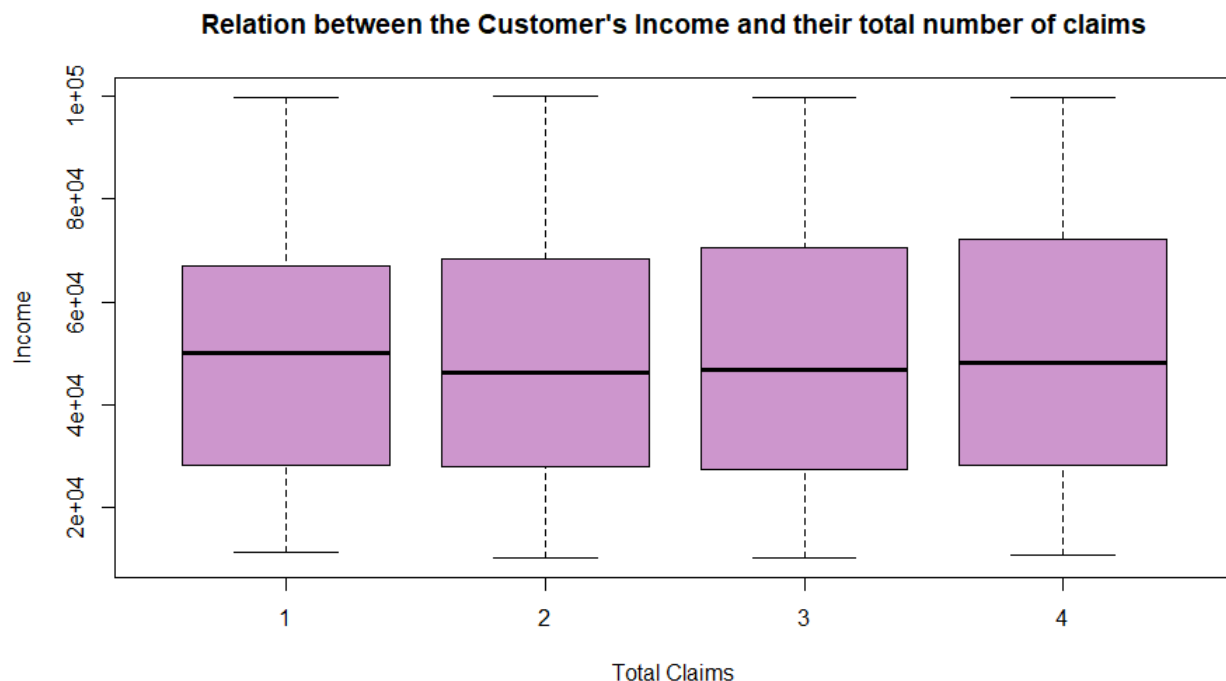**Histogram of the active time of policies with respect to Vehicle type**
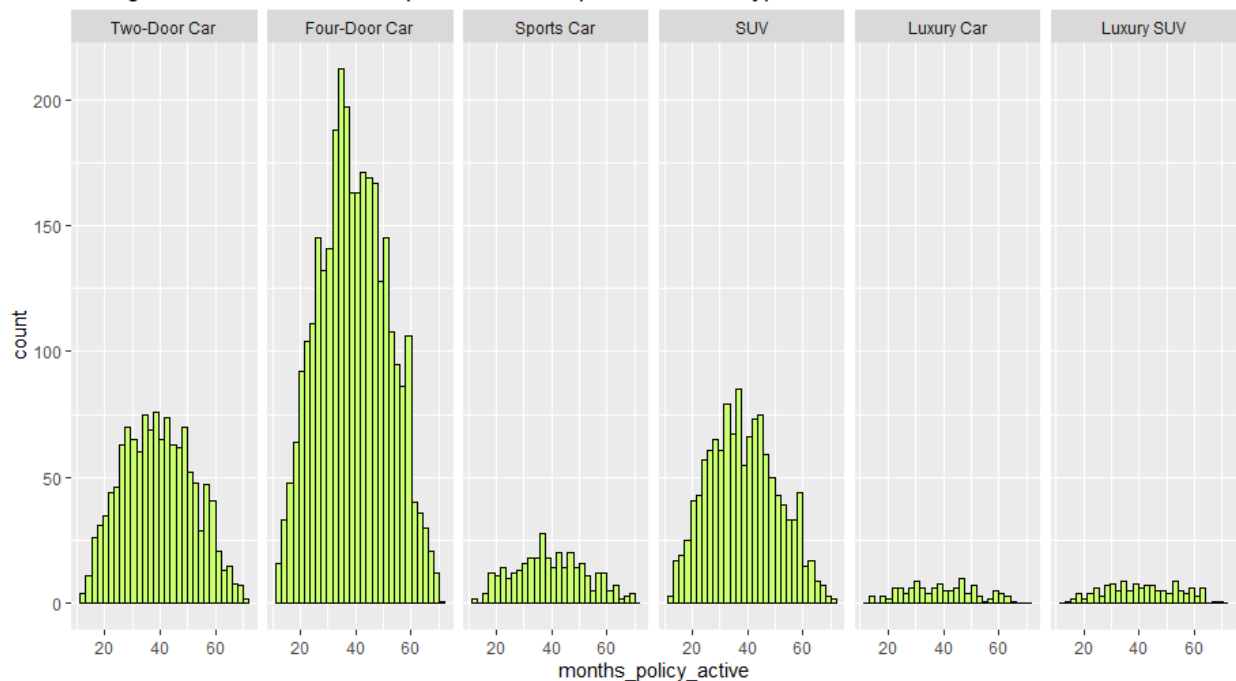
Question 9

Question: Do the customers with higher income keep their policies active for longer without less claims, thus making them long-term and preferable customers for the insurance company?

Answer:

It can be observed from the below summary table that the average income of the customers who have claimed their insurance twice or thrice is lower than that of the customers who have claimed only once. This indicates that the customers with higher income generally tend to claim their insurances lesser number of times. In this dataset though, interestingly, the average income of the customers who have claimed their insurances four times is higher than that of all the other customers.

Both the summary table and the boxplots listed below convey the same inference mentioned above.

```
# A tibble: 4 x 6
  total_claims count min_income avg_income max_income sd_income
         <dbl> <int>      <dbl>      <dbl>      <dbl>     <dbl>
1            1   219      11167     50127.      99816    23332.
2            2  3709      10074     49405.      99981    24139.
3            3  1975      10037     49648.      99874    24624.
4            4   346      10787     51086.      99845    25066.
> |
```

Relation between the Customer's Income and their total number of claims

Business Problem and goals of Analysis

The business problem with this insurance company is that they have been observing reduced profits over the last several years. The ultimate goal of this research and exploratory data analysis is to analyze the insurance policy data at this firm and observe trends and patterns related to the various attributes of a policy. These trends and patterns can be further studied to formulate

suggestions about changes that can be made in the day to day operations of the company to improve the financial stability and profitability of the company.

Highlights and Key Findings

Below are the key findings that have been observed in this exploratory data analysis:

1. The monthly premiums that have to be paid by the customers are not increasing with an increase in the total number of claims made by the customers. On the contrary, they are reducing.
2. Most of the policies at the insurance company are being taken by married customers, irrespective of the type of policy or the type of coverage.
3. There is a significant positive correlation between the vehicle class of a customer and their lifetime value to the company.
4. Customers who are employed tend to claim their policies less frequently, as compared to the remaining customers.
5. Customers from the states Nevada and Arizona have lower claim amounts for similar monthly premiums, as compared to customers from other states.
6. The 'Basic' policy types have been claimed significantly more frequently than that of the other policy types.
7. The customers with higher education levels ("Master", "Doctoral") have taken lower number of claims on average, as compared to the customers with lower education levels ("High School", "Associate", "Bachelor").
8. The average claim amount of a customer goes down as the class of their vehicle increases. As the class of the vehicle increases, the customers tend to keep their policies alive for longer without making a claim.
9. The average income of the customers who have claimed their insurance twice or thrice is lower than that of the customers who have claimed only once.

It is important to analyze these findings and draw conclusions and make necessary changes to the operations of the insurance company in order to improve the financial situation of the company and to ensure profits.

Recommendations

Below are the key recommendations and suggestions that can be given to the Insurance Company to make necessary changes to their operations to ensure that the company makes a profit in the long run:

1. The process of determining the monthly premiums that the customers have to pay needs to be changed completely in a way that a higher premium is set for customers who make frequent claims to their insurances. The premium for an insurance policy has to be increased with every claim that a customer makes from that policy. In the long run, this ensures that a customer does not make frequent claims to a policy which in turn results in more profitability to the company.

2. More emphasis should be put on customers whose marital status is 'Single' or 'Divorced'. The monthly premiums for these demographics of customers could be set at a slightly reduced price to encourage more of them to take up insurance policies with the company, thus resulting in increased revenues in the long run.
3. As the customers with high class vehicles are highly profitable to the company, it is important to attract as many of these types of customers as possible. Small incentives like a limited discount on the premium for a month could be provided to the customer for every policy taken on a high class vehicle, to boost the number of these types of policies being taken up.
4. As Nevada and Arizona are highly profitable for the company, advertisements and campaigns could be conducted in those two states to attract more consumers from these states to take up policies with the company.
5. Although there is no real reason as to why one type of coverage (Basic) is being preferred to the others (premium for all the coverage types is similar), there is a huge gap in the count of the number of basic coverage policies and the others. This has to be countered by appropriate measures – the benefits that the Extended and Premium coverages have over the basic coverage have to be clearly explained to the customer before they take up every new policy.
6. More preference should be given to the customers with higher education levels and higher incomes and the customers who are employed at the time of giving new policies, as they tend to be more profitable to the financial situation of the company.

Each of the recommendations listed above, if followed, would benefit the company in one or other way in their financial situation.