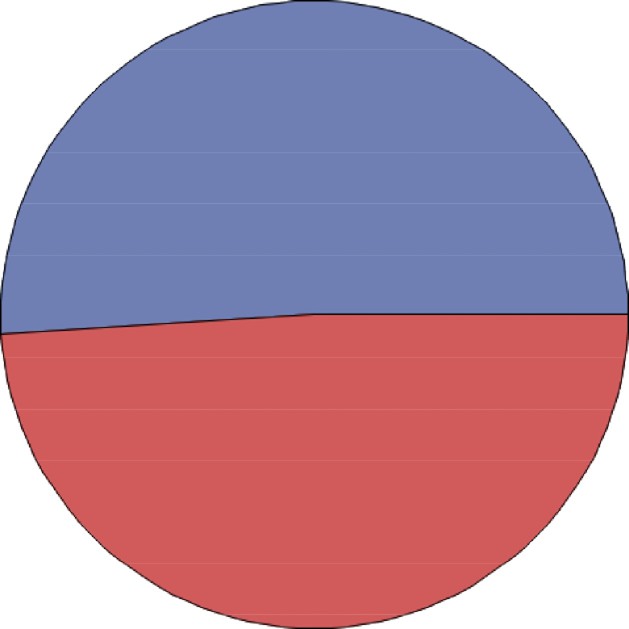
1.

# Frequency of Gender:

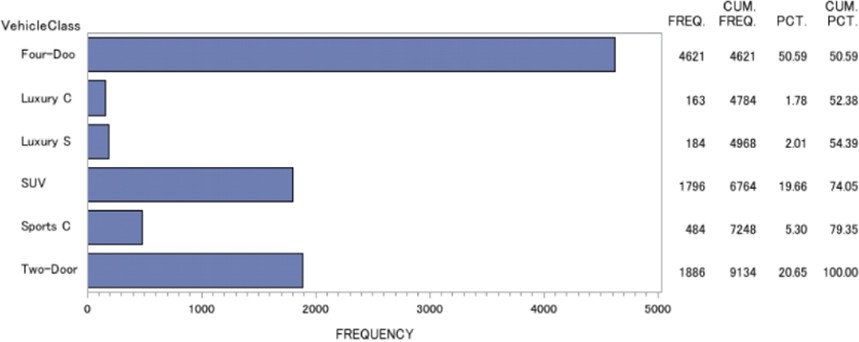
F 4658



M 4476

# Figure 1: Gender Distribution

From figure 1, the distribution of gender as per the data is as shown in the above chart, the number of females are more than males. The number of females exceeds by 182 i.e. by 2%.



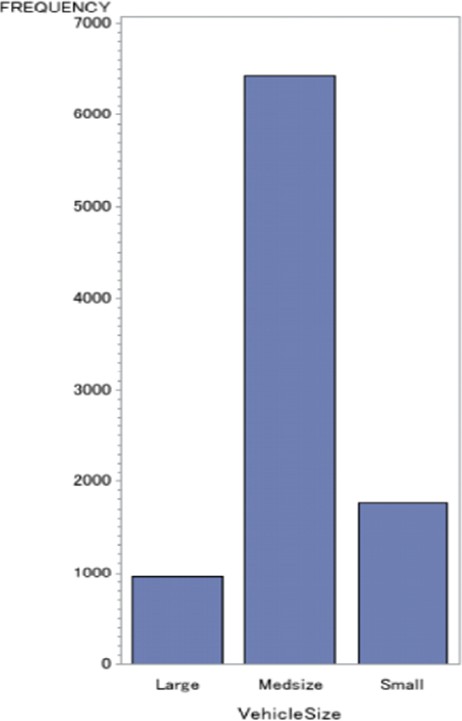
# Figure 2: Vehicle Class Distribution

The Four-Door vehicle class has the highest percentage (50.59 %) amongst all the vehicle classes, Two-Door vehicle class is next at (20.65 %) of total vehicle and it amounts to 1886 in numbers which is less than half of the highest. The Luxury Car amounts to the least number of vehicles (163 in number or 1.78%).

The luxury SUV amounts to (184 in number or 2.01%), Sports Cars amount to 5.3 % of total vehicles and SUV amounts to 19.66%.

From figure 2, we see that most of our customers are owners of either Four-Door or Two-Door Cars which amounts to (71.24%). The 3rd largest segment of vehicle class is SUV which is considerable in Number.

The above mentioned 3 classes of vehicle form a huge chunk (above 90%) and the rest 3 segments having just under 10%. This implies that a majority of vehicles insured are not in luxury segments which implies the owners of these cars are financially conservative in approach with only less than 10% of customers having high-disposable income and passion for the luxury class of vehicles.

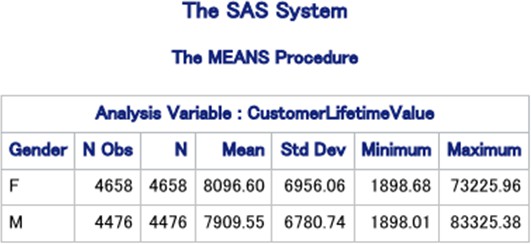


# Figure 3: Vehicle Size Distribution

By looking at the above bar graph we can understand that medium size vehicles dominate over the rest.

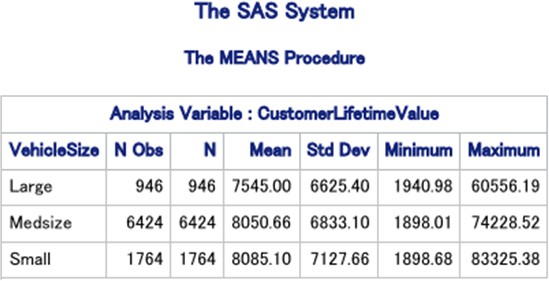
This once again reiterates the point made above that a bulk of our customers are financially conservative in approach as they own Mid-Size vehicles in majority. We also notice that the number of small car owners are more than large car owners.

2.



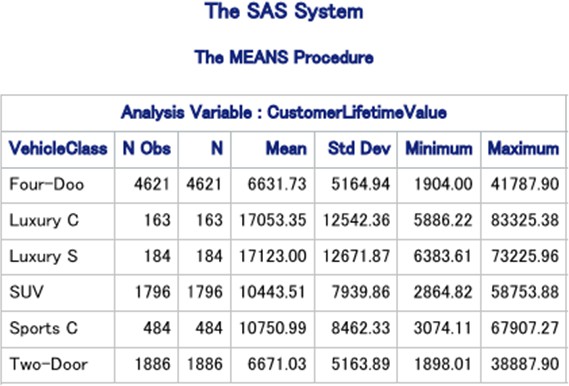
# Figure 4: Average Customer Lifetime Value grouped by gender

The mean of Customer Lifetime Value for females is more than that of males. The mean Customer Lifetime Value of females is **8096.6** and that of males is **7909.74.**



# Figure 5: Average Customer Lifetime Value grouped by vehicle size

The mean of Customer Lifetime Value for Small cars was the highest followed by Med-size and Large cars. The mean of Customer Lifetime Value for large sized vehicles is **7545**, medium sized vehicles is **8050.66** and small sized is **8085.1**



# Figure 6: Average Customer Lifetime Value grouped by vehicle type

By looking at the mean values of the different Vehicle Class we can say that Luxury SUV has the highest mean followed by Luxury car. The mean values are as follows: For Luxury SUV it is **17123**, Luxury Car it is **17053.35**, Sports Car it is **10750.99**, SUV it is **10443.51**, Two Door it is **6671.02** and Four Door it is **6631.73.**

The Customer Lifetime value for 4-Door and 2-Door cars are similar and on the lower side, However the Luxury cars and Luxury SUVs have the highest Customer Lifetime Values.

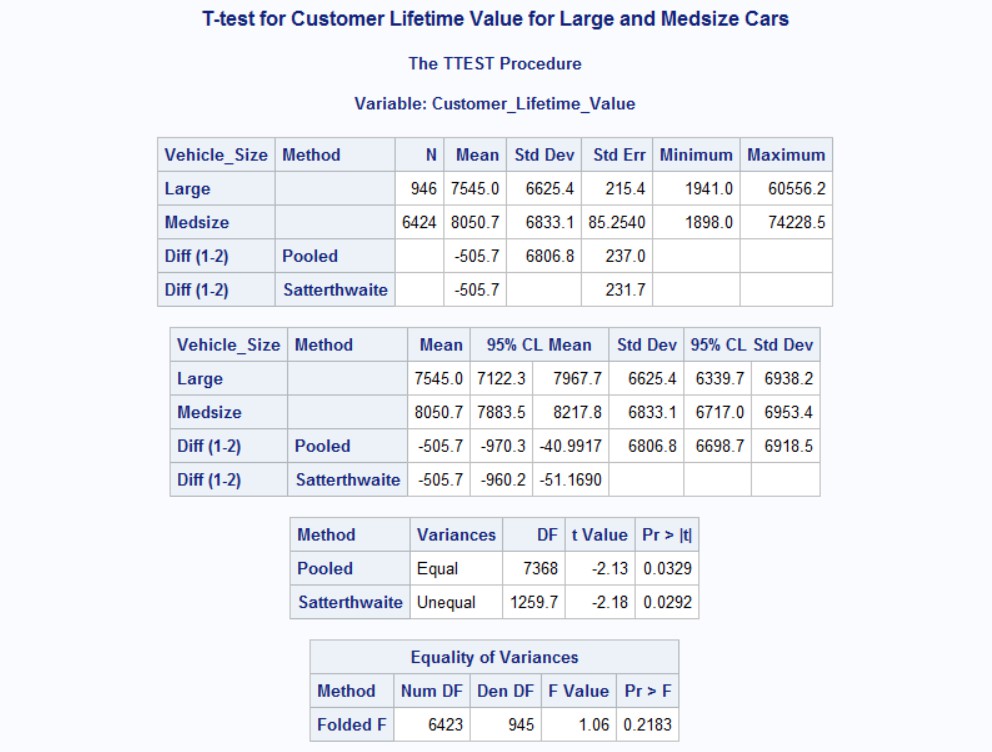
The Customer Lifetime value of SUV’s and Sports Cars are similar and, on the middle, as compared to the Luxury (Higher) and Car segments (Lower).

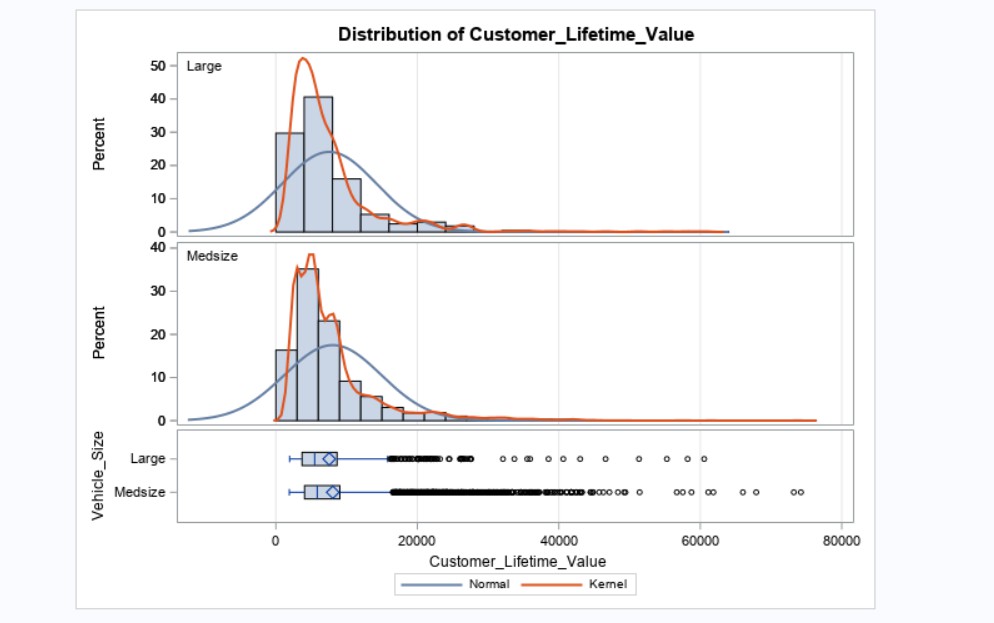
(NOTE: All these has been calculated assuming the distribution is Normal)

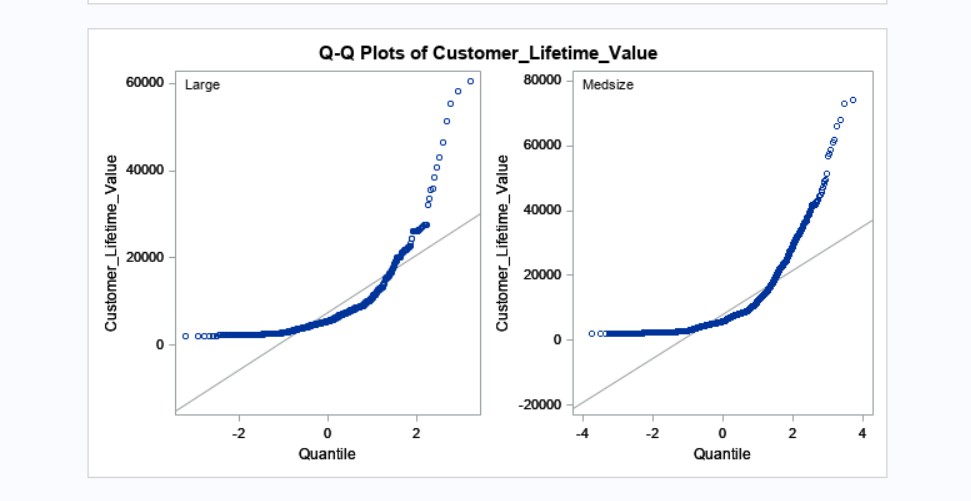
3)

To identify if large cars have higher lifetime value than medsize cars, we performed a t-test.

**Null Hypothesis** : Medium sized cars have less than or equal lifetime value as the Large sized cars.

**Alternative Hypothesis** : Medium cars have higher lifetime value than Large sized cars. The following are the results of the T-test performed:





# Figure 7: TTest results to identify if large cars have higher lifetime value that medium sized cars

From the above results of the t-test we can infer the following :

1. Looking at the Std Dev values of Large and Medsize vehicles we can assume that their variance is almost the same.So we consider the Pooled test values.
2. Pooled variance has a p-value of 0.0329 which is less than the level of significance(0.05), so we reject the null hypothesis.
3. This means that medium sized cars have a higher lifetime than large sized cars.

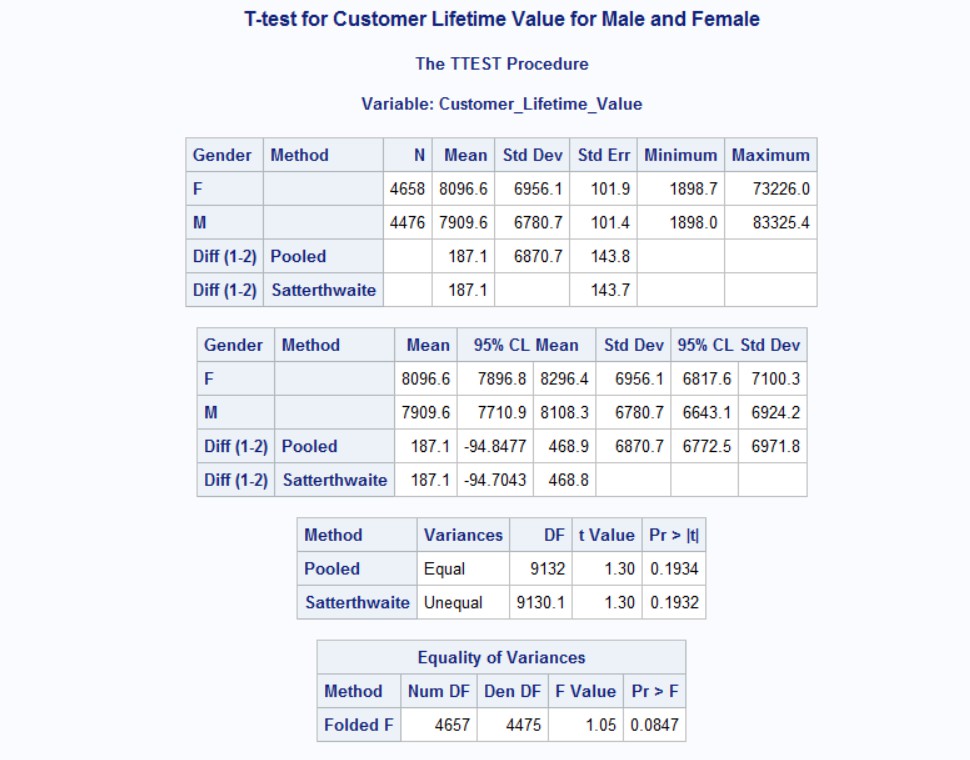
4)

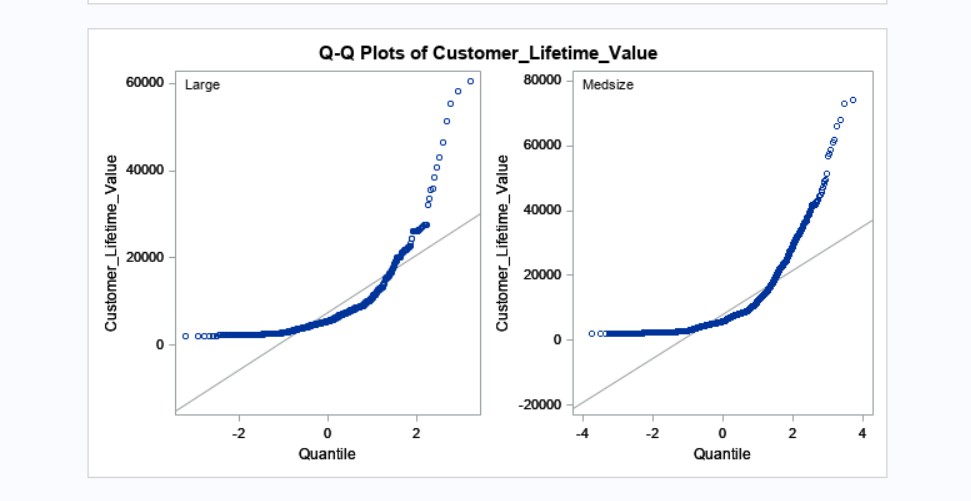
To identify if there is a significant difference between men and women in customer lifetime value we performed a t-test.

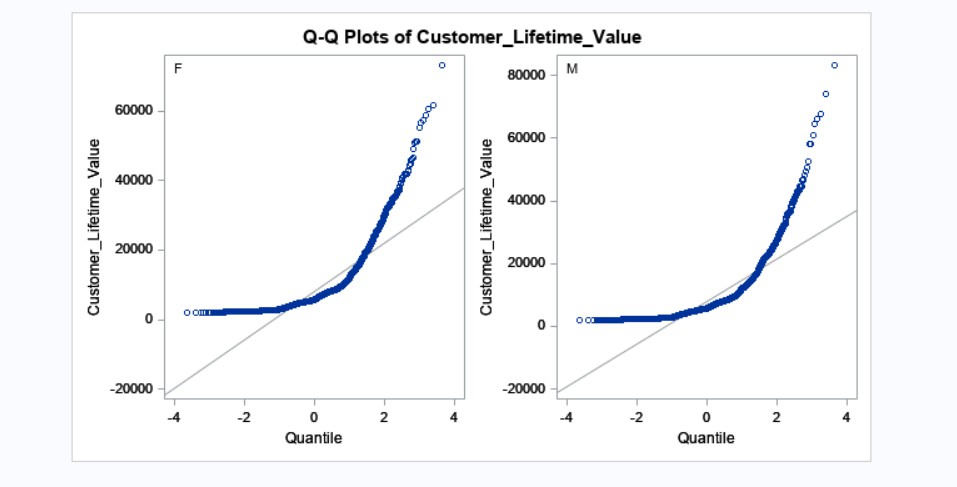
**Null Hypothesis** : There is no significant difference between men and women in customer lifetime value.

**Alternative Hypothesis** : There is a significant difference between men and women in customer lifetime value.

The following are the results of the T-test performed:





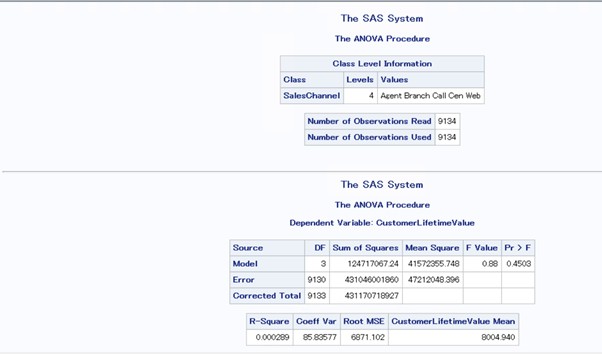


# Figure 8: TTest results to identify if there is significant difference between men and women in customer lifetime value

From the above results of the t-test we infer the following :

1. Looking at the Std Dev values of male and female we can assume that their variance is almost the same. So we consider the Pooled test values.
2. Pooled test has a p-value of 0.1934 which is not less than the level of significance(0.05), so we fail to reject the Null hypothesis. So there is not enough evidence to prove that there is a significant difference between the customer lifetime value of men and women.

5)

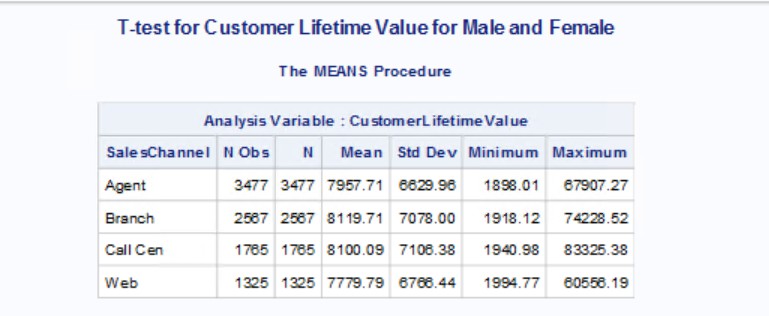


# Figure 9: ANOVA results to identify if there is significant difference in customer lifetime values across different sales

**Null Hypothesis** : There is no significant difference in customer lifetime value across different sales

**Alternative Hypothesis** : There is significant difference in customer lifetime value across different sales

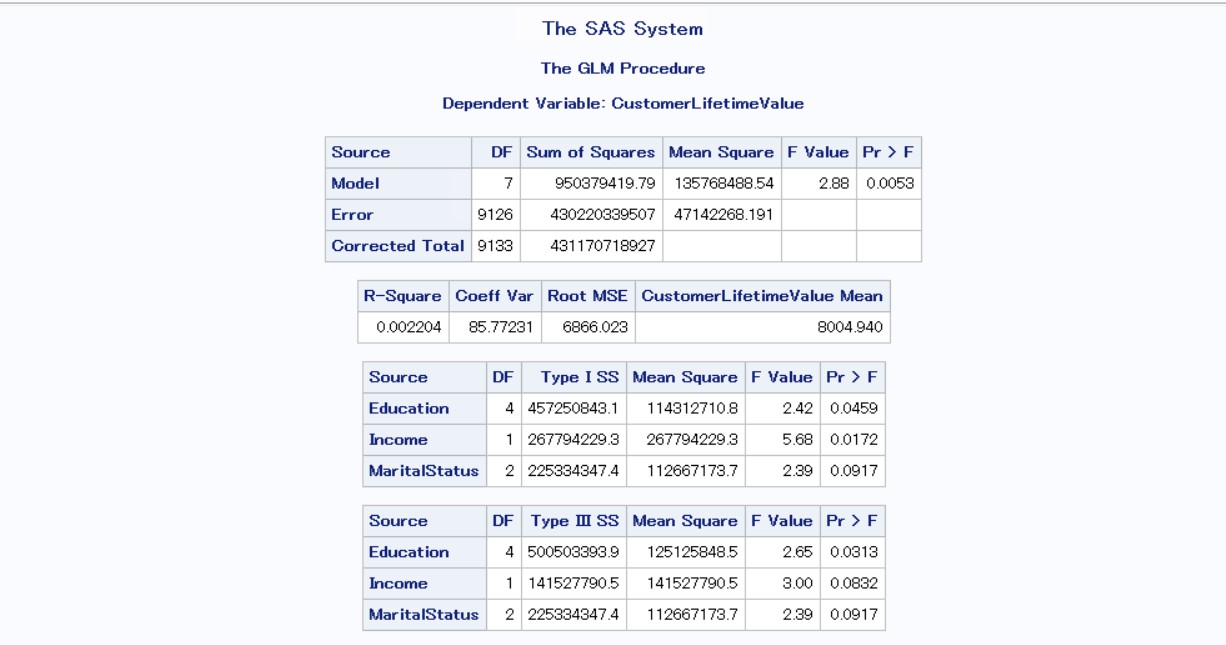
From the ANOVA test,the Pr>F value is 0.4503 which is not less than 0.05 and hence we can conclude that there is no significant difference in customer lifetime value across different sales channels at 95 % CI.



# Figure 10: Mean customer lifetime values for various sales channels

**The Branch sales channel has the highest customer Lifetime Value**. The mean is calculated to be 8119.71

6)

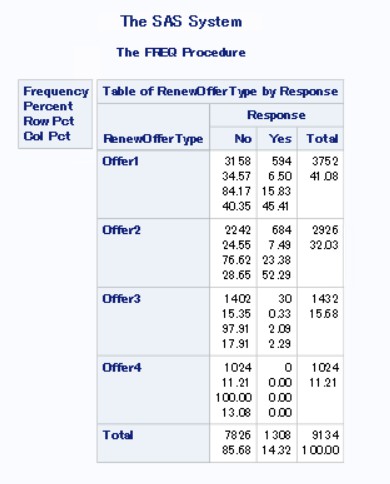
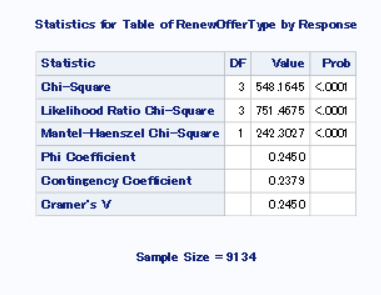


# Figure 11: Regression results for customer lifetime value for independent variables Education, Income and Marital Status

To test whether Education, Income and Marital Status affect Customer Lifetime Value: The concept of logistic regression has been used and the results snapshot is pasted above.

Education & Income affects CustomerLifetime Value whereas Marital Status does not in Type I. Only Education affects CustomerLifetime Value in Type III.

7)



# Figure 12: Chi-Square results

**H0: Null Hypothesis:** There is no relationship between response and renew offer type

**H1: Alternate Hypothesis:** There is relationship between response and renew offer type

From the chi-square results, we see that the p-value is <0.001 which is less than 0.05, which means we reject H0 at 5%, indicating that there is a relationship between response and renew\_offer\_type.

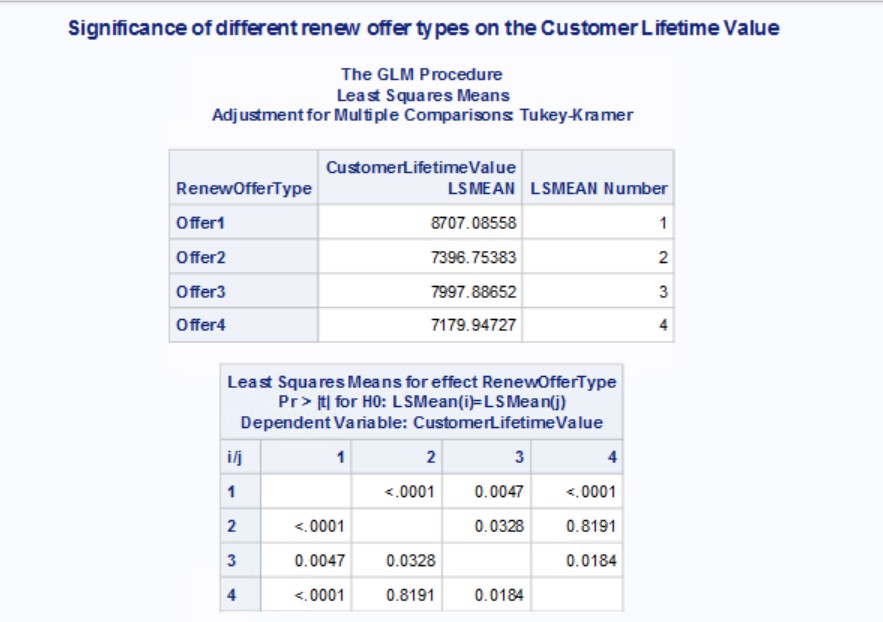
Now after checking the responses for the individual offer types, we see the below result: The response for renewal of Offer1 is mostly 'No' with 84.17% response rate.

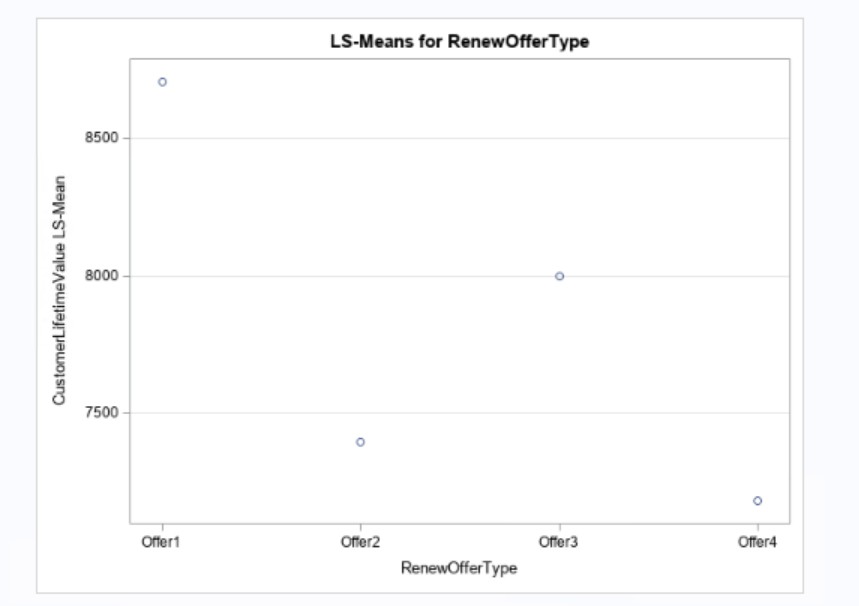
The response for renewal of Offer2 is mostly 'No' with 76.62% response rate. The response for renewal of Offer3 is mostly 'No' with 97.91% response rate. The response for renewal of Offer4 is mostly 'No' with 100% response rate. Among most of the renewal\_offer\_type, about 85.68% has 'No' response.

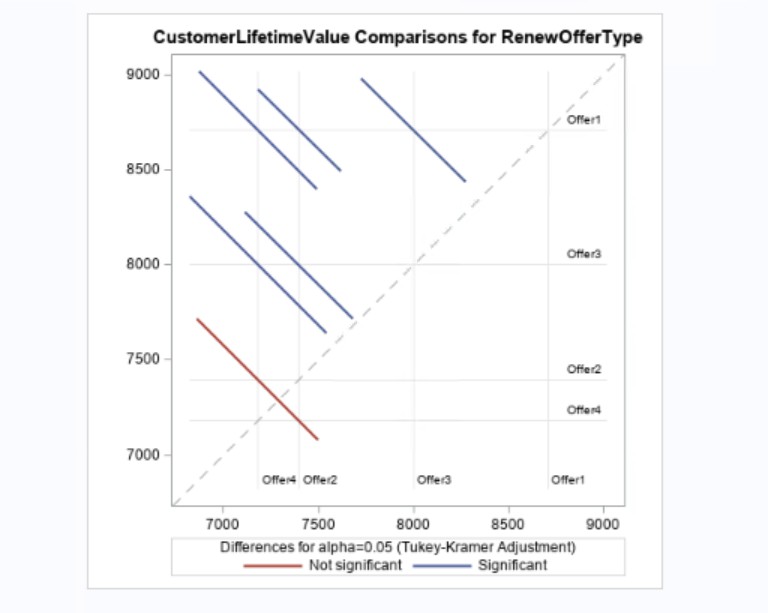
From the other table we can see that:

**Offer1** generates the highest overall response rate of about 41.08%. **Offer2** generates the highest response rate (Response = ‘yes’) of about 23.38%. **Offer4** generates the highest response rate (Response = ‘no’) of about 100.00%.

8)







# Figure 13: Tukey Kramer results

The null and alternative hypotheses are as follows:

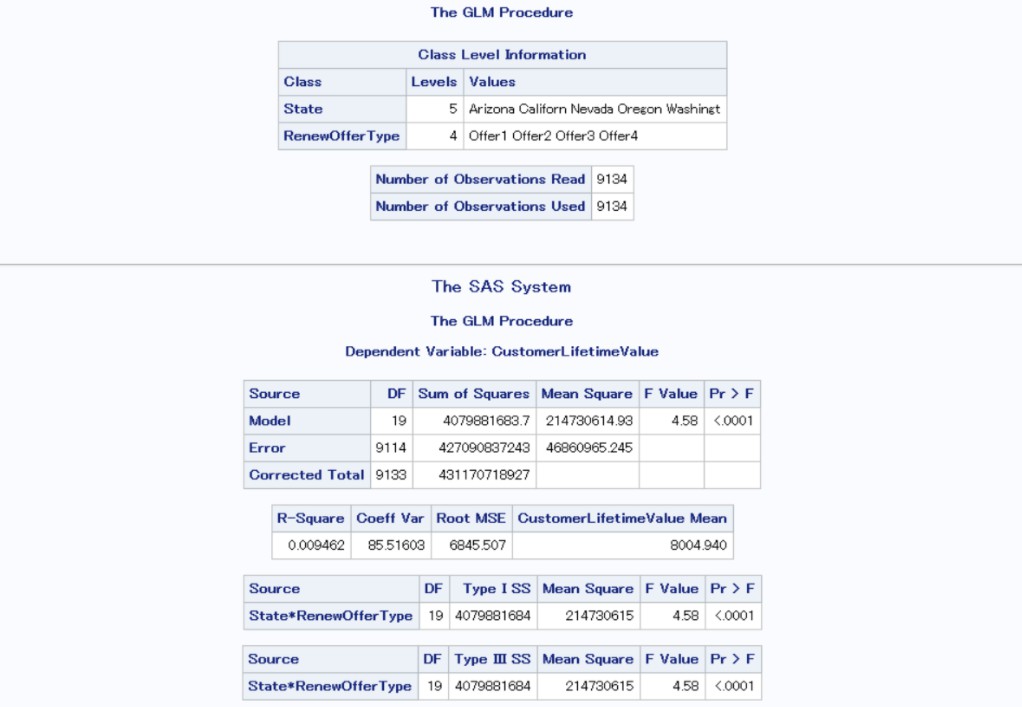
H0: The mean of Lifetime value is same across all renew\_offer\_type

H1: At least two means of Lifetime Value differ across all renew\_offer\_type

According to ANOVA analysis, we see that p-value is <0.01 which means that the results are statistically significant and different renew\_offer\_type have different lifetime values.

From the results, we can see that Offer1 has the highest Mean Customer Lifetime value of 8707.08. However, it should be noticed that the sample sizes are different for each offer type. So, we decided to perform the Tukey test. Among the different combinations, the maximum magnitude of difference is present across Offer 1. Hence, Offer1 is declared to be the best Offer Type.

9.



# Figure 14: Test results to test effectiveness of renew\_offer\_type is different across different states with respect to lifetime value

Null Hypothesis: Effectiveness of renew\_offer\_type is not different across different states with respect to lifetime value

Alternate Hypothesis: Effectiveness of renew\_offer\_type is different across different states with respect to lifetime value

Yes, effectiveness of renew\_offer\_type is different across different states with respect to lifetime value as Pr > F which is 0.0001 because at 95% interval.

10. a.

**Hypothesis 1**: There is a significant relationship between a customer's location and the monthly premium auto. This hypothesis is useful in understanding the relationship between location code and monthly premium auto which can help the firm determine if certain locations are associated with higher or lower premium costs

**Null Hypothesis**: There is no significant relationship between a customer’s location and monthly premium auto

**Alternate Hypothesis**: There is significant relationship between a customer’s location and monthly premium auto

**Hypothesis 2**: The total claim amount is different based on the customer's vehicle size. This hypothesis is is useful in understanding the relationship between vehicle size and total claim amount which can help the firm better understand the risk associated with different types of vehicles and help firm adjust pricing for different vehicle sizes

**Null Hypothesis**: The total claim amount is not different based on the customer's vehicle size

**Alternate Hypothesis**: The total claim amount is different based on the customer's vehicle size

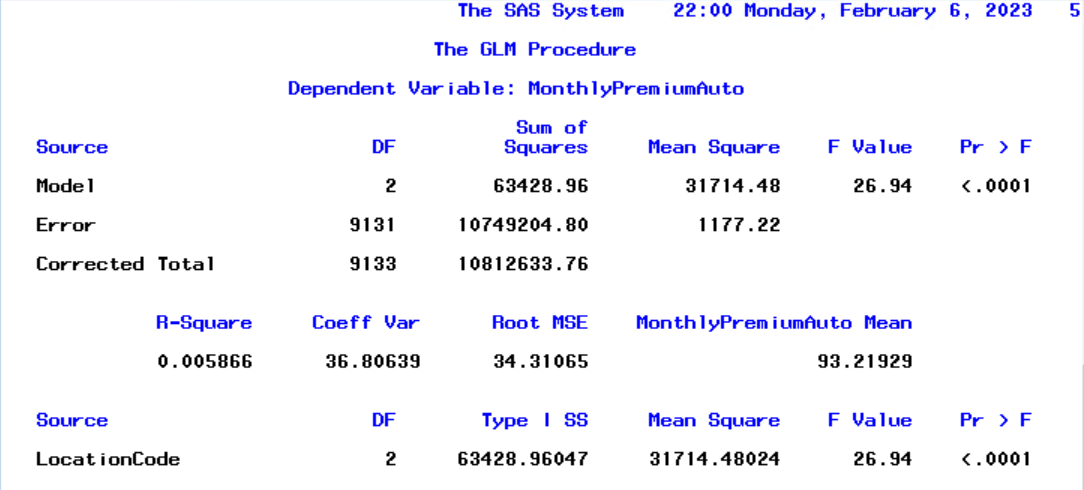
**Hypothesis 3**: There is a difference in Customer Lifetime Value (CLV) between employed and unemployed customers: This hypothesis can be useful to the insurance firm as it can help them understand the impact of employment status on CLV and target their marketing efforts towards the more profitable customer segments.

**Null Hypothesis**: There is no difference in Customer Lifetime Value (CLV) between employed and unemployed customer

**Alternate Hypothesis**:There is a difference in Customer Lifetime Value (CLV) between employed and unemployed customer

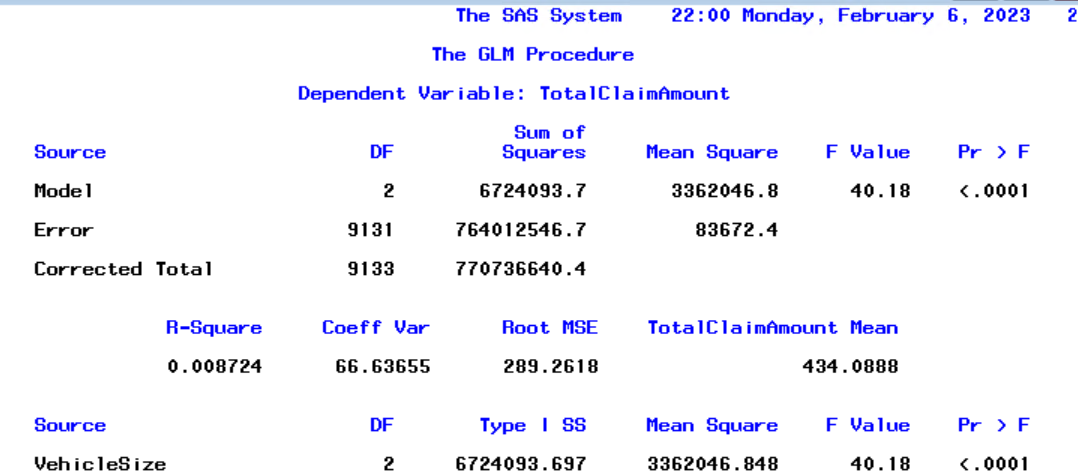
b.

**Anova Test for Hypothesis 1:**



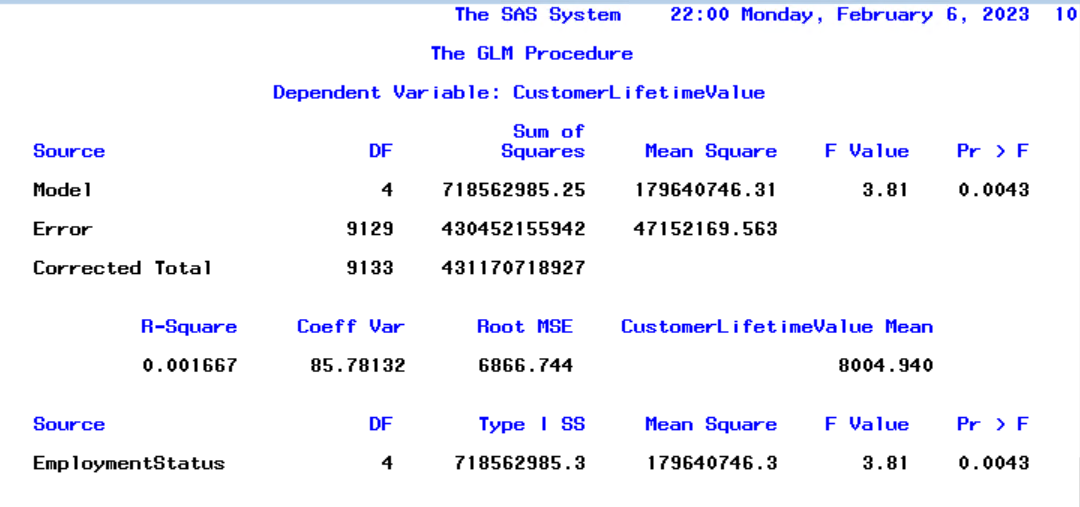
**Figure 15: ANOVA test results to test that is significant relationship between a customer’s location and monthly premium auto**

**Anova Test for Hypothesis 2:**



**Figure 16: ANOVA test results to test that total claim amount is different based on the customer's vehicle size**

**Anova Test for Hypothesis 3:**



**Figure 17: ANOVA test results to test that there is a difference in Customer Lifetime Value (CLV) between employed and unemployed customer**

C.

# Hypothesis 1 Report:

The GLM procedure is a statistical analysis that is used to assess the relationship between a dependent variable and one or more independent variables. In this case, the dependent variable is "MonthlyPremiumAuto" and the independent variable is "LocationCode."

The output shows the Model, Error, and Corrected Total, which represent the sum of squares and mean squares of the model and the error. The F Value and Pr > F are the test statistics that assess the significance of the independent variable in explaining the variation in the dependent variable.

The R-Square value of 0.005866 indicates that the independent variable explains about 0.6% of the variation in the dependent variable. The Coeff Var and Root MSE are measures of the variability of the dependent variable and the Root MSE is the square root of the mean of the residuals.

The Type I SS and Mean Square represent the sum of squares and mean squares for the independent variable "LocationCode." The F Value and Pr > F are the test statistics for the hypothesis test to determine if the independent variable has a significant effect on the dependent variable. In this case, the F Value of 26.94 and Pr > F of < .0001 indicate that the independent variable has a significant effect on the dependent variable.

Management can use this information to improve their operations by targeting their marketing and sales efforts to specific location codes that have a significant effect on the monthly premium auto.

# Hypothesis 2 Report:

The above output is from a General Linear Model (GLM) analysis that tests the relationship between Total Claim Amount and Vehicle Size. The output reports the results of the analysis in terms of the model, error, and corrected total.

The "Model" section shows that the model has 2 degrees of freedom (DF) and a sum of squares equal to 6724093.697. The "Mean Square" column shows the average contribution of each predictor in

the model to the total sum of squares. In this case, each predictor has a mean square of 3362046.848.

The "Source" section shows the individual contribution of each predictor to the model. In this case, there is only one predictor, Vehicle Size, with 2 degrees of freedom and a Type III Sum of Squares of 6724093.697.

The "F Value" column shows the F-statistic for each predictor and the "Pr > F" column shows the p-value. In this case, the F-statistic is 40.18 and the p-value is less than .0001, which indicates that there is a significant relationship between Total Claim Amount and Vehicle Size.

Based on the results of this analysis, it can be concluded that Vehicle Size has a significant impact on Total Claim Amount. Management can use this information to improve their operations by considering Vehicle Size when setting Total Claim Amount amounts and developing marketing strategies.

# Hypothesis 3 Report:

The GLM (Generalized Linear Model) procedure was used to analyze the relationship between the independent variable "Employment Status" and the dependent variable "Customer Lifetime Value". The results show that the "Employment Status" variable has a significant impact on the "Customer Lifetime Value" (p-value = 0.0043).

The R-square value of 0.001667 indicates that only 0.17% of the variation in "Customer Lifetime Value" can be explained by the "Employment Status" variable. The coefficient of variation (85.78%) suggests that the standard deviation of the dependent variable is 85.78% of its mean. The root mean square error (6866.744) is an estimate of the standard deviation of the error term.

Management can use this information to improve their operations by targeting specific segments of customers based on their employment status. For example, they can offer special promotions or discounts to unemployed customers to increase their lifetime value. They can also use this information to allocate resources more effectively, such as increasing marketing efforts in areas with a high concentration of unemployed customers.