1. **What is the distribution of gender, vehicle size, and vehicle class?**

   According to the output below are the interpretation:
   - The gender is approximately evenly distributed with 51% female customer with a frequency of 4658 and 49% male customer with the frequency of 4476 of entire population who have car insurance.
   - The car insurance data has approximately 70% of medium size vehicles, 10% of Large Vehicles and 19% of Small sized vehicles.
   - 50.59% of the vehicles in the dataset are four door cars, followed by 20.65% of two-door cars 19.66% of SUV. Luxury Cars, Luxury SUVs and Sports cars comprises of 1.78%, 2.01% and 5.30% respectively.

2. **What is the average customer lifetime value of each level of gender, vehicle size, and vehicle class?**

   Average Customer Lifetime Value of each level of gender:
   For female: Average Customer lifetime value is 8096.60
   For male: Average Customer lifetime value is 7909.55

   Average Customer Lifetime Value of each level of vehicle size:
   For Large Vehicle: Average Customer Lifetime Value is 7545.00
   For Medsize Vehicle: Average Customer Lifetime Value is 8050.66
   For Small Vehicle: Average Customer Lifetime Value is 8085.10

   Average Customer Lifetime Value of each level of vehicle class:
   For Four-Door Car: Average Customer Lifetime Value is 6631.73
   For Luxury Car: Average Customer Lifetime Value is 17053.35
   For Luxury SUV: Average Customer Lifetime Value is 17123.00
   For SUV: Average Customer Lifetime Value is 10443.51
   For Sports Car: Average Customer Lifetime Value is 10750.99
   For Two-Door Car: Average Customer Lifetime Value is 6671.03

3. **Do Large cars have a higher lifetime value than medsize cars? Do a ttest and report on your findings?**

   We have the following hypotheses:
   - H_0: Large sized cars have less than or equal lifetime value than med sized cars
   - H_1: Large sized cars have higher lifetime value than med sized cars

One tailed T-test Results:

F-test result: The P value of F test is 0.2183 > 0.05. This means that the difference between the variances for medium size cars and large size cars are insignificant. So, we will use equality of variances.

Results of one-tailed test for pooled variances: The P-value is 0.9835 > 0.05. Hence, we fail to reject the null hypothesis. This shows that large-sized cars have lower or equal customer lifetime value than mid-sized cars.

4. **Is there a significant difference between men and women in customer lifetime value?**

   The hypothesis will be:
   - H_0: There is no significant difference between men and women in customer lifetime value.

H$_1$: There is significant difference between men and women in customer lifetime value.

Two tailed T-test Results:

F-test result: The P value of F test is 0.0847 > 0.05. Hence, we fail to reject the null hypothesis. This means that the difference between the variances for male customers and female customers are insignificant. So, we will use equality of variances.

Results of two-tailed test for pooled variances: The P-value is 0.1934 > 0.05. Hence, we fail to reject the null hypothesis. There is not enough evidence to say that there is significant difference between men and women in customer lifetime value.

5. **Use ANOVA to test whether there is difference in customer lifetime value across different sales channels. Which sales channel generates the highest lifetime value?**

   The Hypothesis will be:
   H$_0$: There is no significant difference in average customer lifetime value across different sales channels.
   H$_1$: There is significant difference in average customer lifetime value with at-least one of the sales channels.

ANOVA Test Results:
The P-value is 0.4503 which is greater than 0.05. Hence, we fail to reject the null hypothesis here. Therefore, there is no significant difference in average customer lifetime value across all channels.

The highest average lifetime value is 8119.71 units generated by the branch sales channel.

6. **What demographic factors (education, income, marital_status) affect customer lifetime value?**

   The demographic factors observed in this dataset are:

- Income
- Education
- Marital_status
- State
- Location_Code
- Gender

   For Income, the result shows that the correlation of income and customer lifetime value is 0.02437 which is very less. So, it can be inferred that income has no effect on customer lifetime value.

   Now, we do univariate of Customer Lifetime Value to see the distribution of data.

   For education: The result shows that chi-square probability is 0.3577 which is greater than 0.05. So, education does not affect customer lifetime value.

   For marital status: The chi-square probability is 0.0290 which is less than 0.05. So, marital status affects customer lifetime value.

   For state: The chi-square probability is 0.2000 which is greater than 0.05. So, State does not affect customer lifetime value.

   For location: The chi-square probability is 0.8626 which is greater than 0.05. So, location does not affect customer lifetime value.

   For gender: The chi-square probability is 0.6947 which is greater than 0.05. So, gender does not affect customer lifetime value.

7. **Is there a relationship between renew_offer_type and response (use Chi-sq test)? Which offer type generates the highest response rate?**

Hypothesis is as follow:

H_0: There is no relationship between renew offer type and response.

H_1: There is relationship between renew offer type and response.

Since the P-value of chi-square is <0.0001. Hence, we reject the null hypothesis. So, there exists a relationship between renew offer type and response. Here, Offer 2 generates the highest response rate which 52.29%.

8. **Do different renew_offer_types have different lifetime values? Which offer type is the best?**

The Hypothesis will be:

H_0: Customer lifetime value is same for different renew offer types.

H_1: Customer Lifetime value is different for at least one renew offer types is different.

ANOVA test of the variable Renew Offer Type:

P value is <.0001. This means that we can reject the null hypothesis. So, it is inferred that different renew offer types have different customer lifetime value.

Now, we know that out of four offer types, at least one of them is different from the other. To find that we will check the mean of each offer types:

From mean we can see that top 3 mean customer lifetime belongs to offer 1, 2 and 3.

ANOVA to check if the variances of the 3 groups are different:

Frome ANOVA test we can see that, p-value is < 0.0001. So, by rejecting the null hypothesis, we can infer that atleast one of the offer is different from other.

To do so, we will carry out t-test for all combination of offer 1, 2 and 3.

T-test to check if Offer1 and Offer3 are different:

F-test: P-value < 0.0001. So, we will use the inequality of variance.

By looking at the inequality of variance, under the scatterwhite method we see that p-value is 0.0009. So, we reject the null hypothesis and infer that offer 1 and 3 are different from each other.

T-test to check if Offer1 and Offer2 are different:

F-test: The p-value is <0.0001. So we will select the inequality of variance.

Under inequality of variance, the p-value is <0.0001. we can infer that Offer1 is different than Offer2.

Conclusion: As all offers are significantly different from each other and by looking at the mean value, It can be concluded that Offer1 is the best Offer.

9. **Is the effectiveness of renew_offer_type different across different states with respect to lifetime value?**

Step 1: Making subset of the data by renew_offer_type into carins_offer1, carins_offer2, carins_offer3 and carins_offer4.

Step 2:

ANOVA on the subsets to check the effectiveness of offer across the states.

Hypothesis:

H_0: There is no significant difference between the effectiveness of renew offer types across different states with respect to lifetime values.

H_1: There is significant difference between the effectiveness of renew offer types across for at least one state with respect to lifetime value.

ANOVA Test Results:

For Offer1: P-value is 0.9797 which is greater than 0.05. So, we cannot reject the null hypothesis. Hence, we can infer that effectiveness of Offer1 is not different across different states with respect to lifetime values.

For Offer2: P-value is 0.8302 which is greater than 0.05. So, we cannot reject the null hypothesis. Hence, we can infer that effectiveness of Offer2 is not different across different states with respect to lifetime values.

For Offer3: P-value is 0.4106 which is greater than 0.05. So, we cannot reject the null hypothesis. Hence, we can infer that effectiveness of Offer3 is not different across different states with respect to lifetime values

For Offer4: P-value is 0.3096 which is greater than 0.05. So, we cannot reject the null hypothesis. Hence, we can infer that effectiveness of Offer4 is not different across different states with respect to lifetime values

10. **What other interesting insights that are useful to the company in terms of action can be obtained from the data?**
    a) **Write any three (3) hypotheses.**
    b) **Do appropriate statistical tests or analysis.**
    c) **Report what you found in each case and also write how management can use this information to improve their operations.**

Hypothesis 1: Relationship between marital_status and response type

H_0: There is no relationship between marital_status and response.
H_1: There is a relationship between marital_status and response.

Chi-Square Test Interpretation: The Probability of chi-square is <0.0001. So, we reject the null hypothesis. It can be inferred that there is a relationship between marital status and response type.

Suggestion: Since the married couples are giving more positive response, the management should give them extra discount on different coverage and provide them loyalty points which will make customers to stay with the same firm.

Hypothesis 2: Does total claims varies by the location? Which location code has the highest and lowest total claims?

H_0: The total claims does not vary by different locations.
H_1: There total claims differ for at least one location.

ANOVA Results:

The p-value is <.0001 which means we can reject the null hypothesis. Hence, it can be inferred that different locations have different Total Claim amount.

Step 2: Looking at the mean, total claim amount is maximum in suburban location and minimum in rural locations.

Step 3: T-test

1) **To check that rural and urban region are different:**
   The p-value is <0.0001, which means that the rural and urban groups are different.

2) **To check that rural and sub-urban region are different:**
   The p-value is <0.0001, which means that rural and sub-urban groups are also different.

Interpretation: The suburban location has the highest claim amount and the rural location has the least claim amount.

Suggestion: The management can focus more on the rural areas for customer as they will get the least claims from and will end up making more profit.

Hypothesis 3: Does customer lifetime value vary by level of education? Which level of education has highest customer lifetime value?

H_0: The customer lifetime values does not vary by different levels of education.
H_1: The customer lifetime value significantly varies for at least one or more level of education.

ANOVA Test Results:

The p-value = 0.046 which is less than 0.05. So, we reject the null hypothesis. Hence, it can be inferred that at least one of the groups have different lifetime value.

By looking at the means of lifetime value for each group we find that education level of High School or below has highest lifetime value and education level of Doctor has lowest customer lifetime value.

Step 2: T-Test

1) **To check that high school or lower and bachelor groups are different:**
   Since, p-value of F-test is 0.0006 which is less than 0.05, we reject the equality of variance and look at the p-value of unequal variances.
   The p-value is 0.0265 which is less than 0.05 which means that the high school or lower and bachelors education group are different.

2) **To check that high school or lower and doctor groups are different:**
   Since, p-value of F-test is 0.0009 which is less than 0.05, we reject the equality of variance and look at the p-value of unequal variances.
   The p-value is 0.0350 which is less than 0.05 which means that the high school or lower and doctor education group are different.

3) **To check that high school or lower and college groups are different:**
   Since, p-value of F-test is <0.0001, we reject the equality of variance and look at the p-value of unequal variances.
   The p-value is 0.198 which is less than 0.05 which means that the high school or lower and college education group are different.

4) **To check that high school or lower and Masters group are different:**
   Since, p-value of F-test is 0.1400 which is greater than 0.05, we fail to reject the equality of variance and look at the p-value of equal variances.

The p-value is 0.8581 which is greater than 0.05 which means that there is no difference the high school or lower and Masters education group. They both are same.

Interpretation: The high school or lower and master level of education groups possess the highest customer lifetime value.

Suggestions: The management should focus more on customers of these education groups in order to increase customer lifetime value which will lead to more profit for the firm.