

**Q1. What is the distribution of gender, vehicle size, and vehicle class?**

**A1.**

The distribution of **gender** from the frequency plot is close. Females take up 51% of the dataset and Males take up 49% of the dataset. The total number of rows with female instances is 4658 and males is 4476.

There are 3 **Vehicle Sizes** available in the dataset, Large, Medsize and Small vehicles. Medsize cars are spread the maximum in the dataset, covering 70.33% of the data whereas Large and Small sized cars cover 10.36% and 19.31% of the dataset

The Four-Door Car have the highest representation in the dataset, about 50.59% of the entire dataset.

**Q2. What is the average customer lifetime value of each level of gender, vehicle size, and vehicle class?**

**A1.**

**Highest Customer Lifetime Value:**

Compared to other size cars, luxury SUV has the highest average customer lifetime value for females, which is 28847.17 (Sort the code for highest value on top).

For males, the average customer lifetime value for small luxury cars is the highest at 24361.32 when other size cars are compared.

**Least Customer Lifetime Value:**

Two-door cars and four-door cars for both genders have the least customer lifetime value.

**Q3. Do Large cars have a higher lifetime value than medsize cars. Do a ttest and report on your findings?**

**A3.**

First, we need to perform an f-test to check for equality of variances.

**H<sub>0</sub>= Variances are equal**

**H<sub>1</sub>= Variances are unequal**

As the p-value is greater than 0.05 we conclude that we cannot reject the null hypothesis and that the variances are equal.

We perform a t-test to test the significant difference between customer lifetime value for medsize and large cars.

**H<sub>0</sub>= Average Customer Lifetime Value for Large cars is not higher than medsize cars**

**H<sub>1</sub>= Average Customer Lifetime Value for Large cars is higher than medsize cars**

The t-test output shows that the p-value is less than 0.05 so we reject the null hypothesis and conclude that the Average Customer Lifetime Value for Large cars is higher than the average customer lifetime value of medsize cars.

**Q4. Is there a significant difference between men and women in customer lifetime value?**

**A4.**

First, we need to perform an f-test to check for equality of variances.

**$H_0$ = Variances are equal**

**$H_1$ = Variances are unequal**

As the p-value is more than 0.05 we conclude that we cannot reject the null hypothesis and that the variances are equal.

We perform a t-test to test the significant difference between customer lifetime value for men and women.

**$H_0$ = Average Customer Lifetime Value for men and women is significantly same**

**$H_1$ = Average Customer Lifetime Value for men and women is not significantly same**

From the output of t-test we evaluate that the p-value is 0.19 which is higher than 0.05 and the t-value is 1.30, both higher than the threshold. Hence, we do not reject the null hypothesis at 5% level of significance, which means there is not enough evidence to prove that the Average Customer Lifetime Value for men and women is not significantly same.

**Q5. Use ANOVA to test whether there is difference in customer lifetime value across different sales channels. Which sales channel generates the highest lifetime value?**

**A5.**

First, we need to perform an f-test to check for equality of variances.

**$H_0$ = Variances are equal**

**$H_1$ = Variances are unequal**

As the p-value is more than 0.05 we conclude that we cannot reject the null hypothesis and that the variances are equal.

We now perform a t-test to test the significant difference in average customer lifetime across different sales channel

**$H_0$ = No difference in average customer lifetime across different sales channel**

**$H_1$ = Difference in average customer lifetime across with at least one of the sales channels**

ANOVA Test Results:

P-value = 0.4503 is greater than 0.05, thus we fail to reject the null hypothesis. We can conclude that there is no significant difference in the average customer lifetime across all channels.

The highest average lifetime value is generated by the Branch sales channel.

**Q6. What demographic factors (education, income, marital\_status) affect customer lifetime value?**

**A6.**

The Demographic factors given are education, income, marital status.

For Chi Square Test:

**H<sub>0</sub>= Customer lifetime value and Education are independent of each other**

**H<sub>1</sub>= Customer lifetime value and Education are dependent on each other**

The chi square probability for **education** is 0.1392 which is greater than 0.05 meaning education does not affect customer lifetime value, hence we fail to reject the null hypothesis.

**H<sub>0</sub>= Customer lifetime value and Income are independent of each other**

**H<sub>1</sub>= Customer lifetime value and Income are dependent on each other**

The correlation between income and customer lifetime value is 0.0242, signifying a very weak correlation. We can conclude that income has no effect on customer lifetime value.

**H<sub>0</sub>= Customer lifetime value and Marital Status are independent of each other**

**H<sub>1</sub>= Customer lifetime value and Marital Status are dependent on each other**

The chi square probability for **marital status** is 0.0319 which is less than 0.05 meaning marital status affects customer lifetime value.

**Q7. Is there a relationship between renew\_offer\_type and response (use Chi-sq test)? Which offer type generates the highest response rate?**

**A7.**

We can frame the hypothesis as:

**H<sub>0</sub>= No relationship between renew offer type and response**

**H<sub>1</sub>= There is a relationship between renew offer type and response**

As seen in the output that the p value of Chi square is <0.0001. Thus, we can reject the null hypothesis and conclude that there's a relationship between renew offer type and response. Here Offer 2 generates more response rate than offer 2 which is 7.49%.

**Q8 Do different renew\_offer\_types have different lifetime values? Which offer type is the best?**

**A8.**

We perform a t-test to test the significant difference in offer types across different customer lifetime

**H<sub>0</sub>= There is no significant difference in different renew offer types in different lifetime values**

**H<sub>1</sub>= Significant difference in different renew offer types in at least one of the lifetime values**

As seen in the output the probability of Chi Square is <0.0001, thus we can reject the null hypothesis and conclude that at least one of the renew offer types have different lifetime values.

We find that Offer 1 type is the best out of all

**Q9. Is the effectiveness of renew\_offer\_type different across different states with respect to lifetime value?**

**A9.**

We perform an ANOVA test, where:

**H<sub>0</sub>= There is no significant difference in Average Renew offer type according to lifetime value in different states**

**H<sub>1</sub>= There is a significant difference in Average Renew offer type according to lifetime value in different states**

As seen in the output, the chi square probability is much greater than 0.05 (0.9006) in all the cases, giving an insight that renew offer type does not differ in different states with respect to customer lifetime value at 5% significance level.

**Q10 What other interesting insights that are useful to the company in terms of action can be obtained from the data?**

- **Write any three (3) hypotheses.**
- **Do appropriate statistical tests or analysis.**
- **Report what you found in each case and also write how management can use this information to improve their operations.**

**Hypothesis 1:**

H<sub>0</sub>= Claim Amount is not significantly affecting Customer Lifetime Value

H<sub>1</sub>=Claim Amount is significantly affecting Customer Lifetime Value

After plotting an ANOVA for customer lifetime value and Claim Amount, we figure out that the p-value is less than 0.05 which means it is not significant in defining whether it matters.

**Hypothesis 2:**

H<sub>0</sub>= Gender does not affect Renew Offer Type significantly

H<sub>1</sub>= Gender affects Renew Offer Type significantly

For the range greater than 8999 customer lifetime value, we find that chi-square value is 0.1185 which is significantly greater than 0.05, hence we do not reject the null hypothesis and conclude that gender does not affect renew offer type.

**Hypothesis 3:**

$H_0$ = Location does not affect Renew Offer Type significantly

$H_1$ = Location affects Renew Offer Type significantly

When we look at the states and Renew Offer Type we find that the chi square value in all the ranges is greater 0.05, thus we do not reject the null hypothesis and conclude that location does not affect Renew Offer type

Managers can use these hypotheses to gain insights and work on these points by not focusing on unsubstantial variables.