Spring 2020 | BUAN6337.002

Predictive Analytics using SAS

Group2

Adam Butcher

Dong In Kim

Maria Phetteplace

Mahmoud Ali

Q1. The distribution of gender, vehicle size, and vehicle class

Q1-1. Gender

Gender	Frequency	Percent	Cumulative Frequency
F	4658	51%	4658
M	4476	49%	9134
Conclusion			
The distribution between male and female is almost an even split but the frequency of female car insurers is a bit higher than that of the male. (Female 51% vs. Male 49%)			

Q1-2. Vehicle Size

Vehicle Size	Frequency	Percent	Cumulative Frequency
Large	946	10.36	946
Medium	6424	70.33	7370
Small	1764	19.31	9134
Conclusion			
Medium-sized car accounts for 70.33%, Small car 19.31%, and Large the left. Medium sized cars have an overwhelming number compared to large and small sized cars.			

Q1-3. Vehicle Class

Vehicle Class	Frequency	Percent	Cumulative Frequency
Four-Door Car	4621	50.59	4621
Luxury Car	163	1.78	4784
Luxury SUV	184	2.01	4968
SUV	1796	19.66	6764
Sports Car	484	5.30	7248
Two-Door Car	1886	20.65	9134
Conclusion			
Four-door cars represent the greatest number of cars, and SUV and three-door cars have the next biggest number of cars.			

Q2. Average customer lifetime value of each level of gender, vehicle size, and vehicle class

Q2-1. Gender

Gender	Mean	Conclusion
Male	7909.55	From gender perspective both males and females almost have the same average lifetime value which is around 8000. The female's lifetime value is higher than male's lifetime value.
Female	8096.60	

Q2-2. Vehicle Size

Vehicle Size	Mean	Conclusion
Large	7545.00	Small vehicles has the highest average lifetime value and Large car has the lowest average lifetime value.
Medium	8050.66	
Small	8085.10	

Q2-3. Vehicle Class

Vehicle Class	Mean	Conclusion
Four-Door Car	6631.73	It was interesting that the average lifetime value of Luxury SUV is the highest out of all vehicle class, even higher than Sports Car. The lifetime value of luxury cars is the highest, that of SUV/Sports are the next, and Four-door/Two-Door cars have the lowest value.
Luxury Car	17053.35	
Luxury SUV	17123.00	
SUV	10443.51	
Sports Car	10750.99	
Two-Door Car	6671.03	

Q3. Vehicle size (Large vs Medium) in customer lifetime value [T-test]

H_0	Mean customer lifetime value of Large cars <= Medium cars	
H_A	Mean customer lifetime value of Large cars > Medium cars	
Result	Reject	Conclusion
t-value: -2.13 p-value: 0.0329	Reject the H_0	The average customer lifetime value of large cars is higher than medium size cars in the confidence level of 95%.

Q4. Gender (Male vs Female) in customer lifetime value [T-test]

H_0	There is no significant difference between men and women in customer lifetime value	
H_A	There is a difference between men and women in customer lifetime value	
Result	Reject	Conclusion
t-value: 1.30 p-value: 0.1934 (p-value of F-test: 0.0847 = equal variance)	Cannot Reject the H_0	There is not a significant difference between men and women in customer lifetime value.

Q5. The difference in customer lifetime value across different sales channels and the highest lifetime value of the sales channel [ANOVA]

H_0	There is no significant difference in customer lifetime value across different sales channels	
H_A	There is a difference in customer lifetime value across different sales channels	
Result	Reject	Conclusion
F-value: 0.88 p-value: 0.4503	Cannot Reject the H_0	There is no difference in customer lifetime value across different sales channels.

Q6. Correlation between Demographic factors (education, income, marital_status) and customer lifetime value

Q6-1. Education Level and Customer Lifetime Value [ANOVA]

H_0	Average customer lifetime value of people with different education level are equal.	
H_A	At least one group of people with different educational level has a different average customer lifetime value.	
Result	Reject	Conclusion
F-value: 2.42 p-value: 0.0460	Reject the H_0	At least one group of people with different educational level has a different average customer lifetime value, which means the

		educational level affect the lifetime value.

Q6-2. Income and Customer Lifetime Value [CORR]

Conclusion
Since 'Income' is the continuous variable, we use proc corr to see the correlation between Income and Customer Lifetime Value. Correlation is 0.02437, thus there is not a significant correlation with customer lifetime value. Income does not affect the lifetime value.

Q6-3. Marital Status and Customer Lifetime Value [ANOVA]

H_0	Average customer lifetime value of people with different marital stauts are equal.	
H_A	At least one group of people with different marital stauts has a different average customer lifetime value.	
Result	Reject	Conclusion
F-value: 3.32 p-value: 0.0363	Reject the H_0	At least one group of people with different marital status has a different average customer lifetime value, which means the marital status affect the lifetime value.

Q7. The relationship between renew_offer_type and response [Chi-test]

H_0	Renew Offer Type and Response are not related – independent	
H_A	Renew Offer Type and Response are related – dependent	
Result	Reject	Conclusion
Chi-square value: 548.1645 p-value: 0.0001	Reject the H_0	Renew Offer Type and Response are related. Thus, we can conclude that there is a relationship between Renew Offer Type and Response. Offer 2 has the highest response rate with 52%.

Q8. Different Renew Offer Type and the best Offer Type

Q8-1. Different Renew Offer in Customer Lifetime Value [ANOVA]

H_0	Average customer lifetime value of people with different renew offer type are equal.	
H_A	At least one group of people with different renew offer type has a different average customer lifetime value.	
Result	Reject	Conclusion
F-value: 25.83 p-value: 0.0001	Reject the H_0	At least one group of people with different renew offer type has a different average customer lifetime value, which means the different renew offer type have different customer lifetime value.

Q8-2. The Best Offer Type [T-test]

H_0	There is no significant difference between Offer X and Offer Y	
H_A	There is a difference between Offer X and Offer Y	
Result		Conclusion
Offer1 / Offer2 Offer1 / Offer3 Offer1 / Offer4	Offer1 > Offer2 Offer1 > Offer3 Offer1 > Offer4	Since offer1 is better than offer2, offer3, and offer4, Offer Type 1 is the best one.

Offer2 / Offer3 Offer2 / Offer4 Offer3 / Offer4	Offer2 < Offer3 Offer2 = Offer4 Offer3 > Offer4	

Q9. Different Renew Offer Type across different state in Customer Lifetime Value [ANOVA]

H_0	Average customer lifetime value of people with different renew offer type across different state are equal.	
H_A	At least one group of people with different renew offer type across different state has a different average customer lifetime value.	
Result	Reject	Conclusion
Renew Offer Type*State F-value: 0.71 p-value: 0.7388	Cannot reject the H_0	There is no interaction effect between renew offer type and state (Pr>F: 0.7388), which means Average customer lifetime value of people with different renew offer type across different state are equal. If no evidence for an interaction effect is found, then one can proceed to testing the main effects of the independent variables. (Can reject the Null Hypothesis of Renew Offer type, but cannot reject the Null Hypothesis of State)

Q10. Interesting insights & suggestions that can be obtained from the data

Q10-1. (Hypothesis 1) Does car class affect the profitability or loss of the user? What is the top car class that makes the customer more profitable ?

H_0	Car class does not affect the profitability of the user.	
H_A	Car class does affect the profitability of the user.	
Statistical Analysis	Filter sub data of only personal profiles →Add a loss/profit column →Run ANOVA test to check if car class has an effect on profitability	
Result		Conclusion
<table><tr><th>SNK Grouping</th><th>Mean</th><th>N</th><th>Vehicle_Class</th></tr><tr><td>A</td><td>9861.9</td><td>146</td><td>Luxury SUV</td></tr><tr><td>B</td><td>9370.1</td><td>116</td><td>Luxury Car</td></tr><tr><td>C</td><td>5526.1</td><td>355</td><td>Sports Car</td></tr><tr><td>C</td><td>5116.8</td><td>1329</td><td>SUV</td></tr><tr><td>D</td><td>3388.7</td><td>3433</td><td>Four-Door Car</td></tr><tr><td>D</td><td>3365.2</td><td>1409</td><td>Two-Door Car</td></tr></table>Means with the same letter are not significantly different. The most profitable one is Luxury SUV.		Sales team: Sales managers should have a KPI on selling for people who have luxury SUVs and corporate teams to focus on companies with larger Luxury SUV. Marketing side: we can cooperate with SUV car dealers to refer our company to their customers. Billboards location optimization: We can put billboards in front of Luxury SUV dealers locations.

Q10-2. (Hypothesis 2) Does car size affect the customer's total claims?

H_0	Having a large sized car versus a small sized car has no effect on the amount of a customer's total claims.
H_A	Having a large sized car versus a small sized car has an effect on the amount of a customer's total claims.

Statistical Analysis	Filter out only large and small size vehicles. → Do a t-test with the variable, 'total_claims_amount' and class 'vehicle_size'.	
Result	Reject	Conclusion
F-value: -5.55 p-value: 0.0001	Reject the H_0	Here is enough variance to reject the null hypothesis in favor of the alternate hypothesis. Our findings indicate that vehicle size does impact the total amount of claims for a customer. Because of this evidence, the company can look further into predicting total claims by what type of car a customer owns.

Q10-3. (Hypothesis 3) Do divorced people have a lower claim amount than married people?

H_0	Divorced People file have a lower claim amount than Married People	
H_A	Divorced People have a higher claim amount than Married People	
Statistical Analysis	Do a t-test with the variable 'Total_Claim_Amount' along with the classes 'Married' and 'Divorced'	
Result	Reject	Conclusion
F-value: 2.51 p-value: 0.0120	Reject the H_0	There is enough evidence to prove that divorced people have a higher claim amount than married people. The company could look into the marital status of people to see what their position in claim is and make decisions from there.