

Inferential Statistics Report

Overview

This Report presents the results of statistical methods applied on two cases :-
We want to check is there any relationship between

is_auto_renew and is_churn

is_cancel and is_churn

We will be using Chi-Squared test to check the association between these variables.

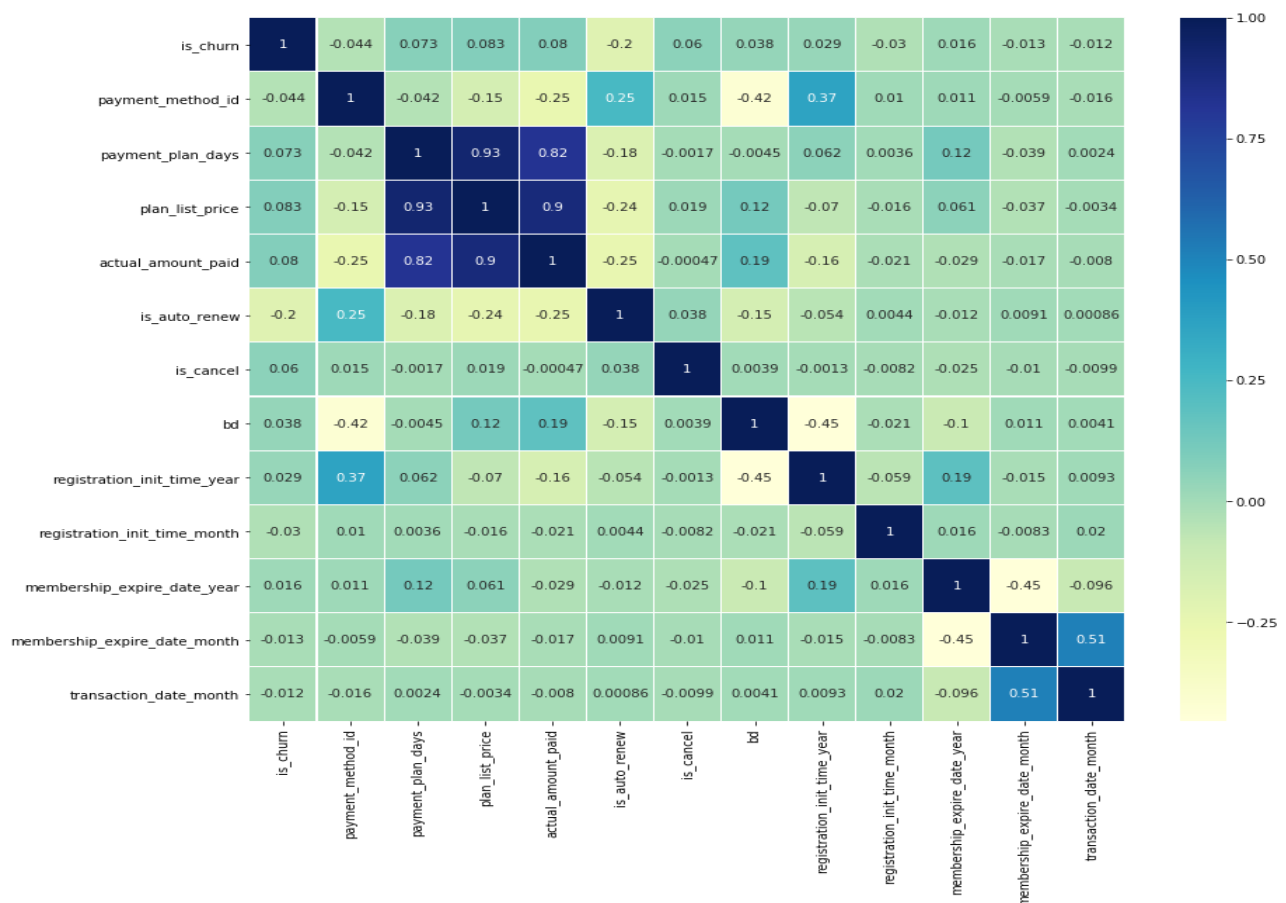
Set our significance level to be 0.05 .

Let's state our null hypothesis and the alternative hypothesis.

- H0: There is no statistically significant relationship between 'is_churn' and 'is_auto_renew'

- Ha: There is a statistically significant relationship between 'is_churn' and 'is_auto_renew'

Drawn Below graph will give us information about the correlation between the variables:-



We can see from above Graph, There is strong correlation between payment plan days, actual_amount paid, plan list price features. Also is_churn and is_auto_renewal are negatively correlated to each other.

Constructing the Contingency Table The next step is to format the data into a frequency count table. This is called a Contingency Table, we can accomplish this by using the `pd.crosstab()` function in pandas.

is_auto_renew	0	1	All
is_churn			
0	969187	14228390	15197577
1	220929	464642	685571
All	1190116	14693032	15883148

Chisq_test_statistic = 632301.26, P-value ~ 0, degree_of_freedom = 1

Conclusions:-

- With a p-value < 0.05 , we can reject the null hypothesis. There is definitely some sort of statistically significant relationship between 'is_churn' and the 'is_auto_renew' column.
- These two variables are not independent of each other.

Relationship between 'is_churn' and 'is_cancel' Variables

- The Null and Alternative Hypotheses:-

- Recall that we are interested in knowing if there is a relationship between 'is_churn' and 'is_cancel'. In order to do so, - we would have to use the Chi-squared test and we set our significance level to be 0.05 .But first, let's state our null hypothesis and the alternative hypothesis.

- H0: There is no statistically significant relationship between 'is_churn' and 'is_cancel'
- Ha: There is a statistically significant relationship between 'is_churn' and 'is_cancel'

Contingency Table

Is_Cancel	0	1	All
Is_Churn			
0	14958679	238898	15197577
1	648464	37107	685571
All	15607143	276005	15883148

Chisq_test_statistic = 56664.17, P-value ~ 0, degree_of_freedom = 1

Conclusions:-

- With a p-value < 0.05 , we can reject the null hypothesis. There is definitely some sort of statistically significant relationship between 'is_churn' and the 'is_cancel' column.
- These two variables are not independent of each other.