

Hypothesis: objective, method, result, discussion

Objective: Before we delve into the modeling part, we believe that conducting some hypothesis testing will help us discover insights from our data related to customer satisfaction, marketing opportunities, and functionality improvement. A plausible assumption regarding customer satisfaction is that if users with a higher number of support tickets per month tend to have a lower satisfaction score, it indicates a need for improvement in customer service. In terms of marketing opportunities, we can investigate whether there are distinct genre preferences between females and males. This insight could enable the company to send out gender-specific advertisements to users' registered email addresses. For functionality improvement, we aim to assess the effectiveness of parental control by testing the viewing hours and duration for people who use parental control compared to those who do not use it.

Method: Our chosen method is the ANOVA test, which allows us to determine if different groups are significantly different from each other. Under the large number of observations collected, our data points are fairly balanced in terms of the number of observations for different groups in each column. Due to this reason, we do not find a significant difference in genre preferences between females and males. On the other hand, when testing the relationship between the number of support tickets and satisfaction scores, the scatter plot shows a non-linear relationship with points randomly spread out. Initially, due to the large amount of data, our plot was overwhelming with points. In response to this, we employed the bootstrap method to obtain a random smaller sample size of 500 for graphing. After several attempts, we can confirm that there is no discernible relationship between support tickets and satisfaction scores. Customer satisfaction appears to be influenced by other factors outside our dataset.

Regarding our hypothesis about parental control, the viewing hours and durations do not differ significantly based on whether users employ parental control or not. This suggests that the parental control function does not significantly reduce total viewing hours and viewing duration. This insight might help us understand that factors other than parental control are influencing these aspects.

Under our efforts, we have fortunately obtained some meaningful findings in (WatchlistSize~SubscriptionType), (TotalCharges~ContentType), and (UserRating~Paymethod) using Kruskal-Wallis test, Dunn Test, and Mann-Whitney U test. The S-shape of the QQ plot and non-normal histogram of residuals indicate that our data is not normally distributed.

SubscriptionType can be ranked by basic, standard, and premium membership. WatchlistSize is a numerical column within the range of (0,24). ContentType is ranked as either one type (TV or Movies) or both (TV and Movies). TotalCharges is a numerical column within the range of (4.99,2378). Paymethod can be ranked by the speed of the

process: Mailed Check, Electronic Check, Bank Transfer, and Credit Card. UserRating is a numerical column within the range of (1,5).

Under the non-normal assumption, we attempted various transformations, including Box-Cox, log, square root, glms, and bootstrap. However, none of them proved effective for our data. Due to this, we opted for the Kruskal-Wallis test. Our columns satisfied independence from the correlation matrix. The large dataset supports a more precise approximation of testing results.

When we ran the Kruskal-Wallis test on WatchlistSize~SubscriptionType, we obtained a p-value of 0.07692. We discovered that Basic vs Premium and Basic vs Standard differ from each other with p-values of 0.04 and 0.06, respectively. Based on this, we decided to regroup Premium and Standard into Advanced Subscription.

Starting from this point, as we now only have two independent groups, Basic and Advanced Subscription, we utilized the Wilcoxon Rank Sum Test/Mann-Whitney U Test. The only assumptions for conducting a Mann-Whitney test are that the two groups must be independent and that the dependent variable is ordinal or numerical. However, to report the differences between groups as medians, the shape of the distributions of the dependent variable by group must be similar. After examining the frequency histogram, it is evident that the two groups have similar distributions. Therefore, we can use the Mann-Whitney U test to determine whether watchlist sizes are longer or shorter in basic membership versus advanced membership based on the use of mean ranks to describe the group differences.

Result: (WatchlistSize~SubscriptionType) have p-value as 0.02376 and negative values in difference of location. People with advanced subscriptions tend to use the app more frequently. When a customer with a basic but lengthy watchlist size appears, we can offer them a targeted upgrade discount. Once they upgrade to a Standard or Premium membership, the company can benefit from their potential longer-term commitment due to the higher-ranked subscription types and the enhanced benefits associated with the higher prices they pay.

(TotalCharges~ContentType) has p-value 0.04527 and a negative value in difference of location. Customers who watch both TV and movies tend to have higher charges. The company may consider sending promotions or advertisements for other types of content to customers who prefer a specific content type.

(UserRating~Paymethod) has 0.002939 and a negative value in difference of location. Specifically, individuals who paid with BankTransfer demonstrated higher satisfaction. Notably, we have already examined various factors, such as gender, watchlist preferences, subscription type, genre preferences, and other dataset columns, and found no significant differences among users based on their payment methods. Therefore, the noteworthy higher satisfaction observed among BankTransfer users warrants further investigation. If the cause is operational convenience, enhancing the functionality of other payment types to improve overall payment convenience may contribute to higher customer ratings.

Discussion for further improvement: In the process of hypothesis testing, our initial assumed insights on customer satisfaction, marketing opportunities, and functionality improvement have proven unsuccessful. However, this failure may result from the limitations of columns within our dataset, as there are likely more confounding variables outside the scope of our dataset. To analyze personalized preferences, we believe that incorporating personal user data, including information such as age, occupation, home address, or even having each row represent an individual viewing record instead of an account, could open up more possibilities for our hypothesis testing.