PROJECT PROPOSAL

Statistical Analysis on Customer Behavior on E-commerce Platform

Fall 2023: DATA 602- L01

Prashant Dhungana-30080130
Prateek Kaushik-30229287
Mariya Mathews-30192182
Nisha Pillai-30158934

Please note the data set used for this project is the same data set being used for the DATA-601 project.

Introduction:

The primary objective of our project is to conduct statistical analysis to understand attributes and factors that influence customer behavior on e-commerce platforms. The analysis will provide valuable and explicit information regarding customer behavior on e-commerce which is an important insight that businesses can use to improve their services and enhance customer satisfaction. Some of the variables we will be analyzing are Churn Rate, Hours spent on the app, Purchase Frequency, Satisfaction Score, Gender, Preferred Order Category.

DataSet:

The dataset used for this project was extracted from kaggle and it is important to note the data was collected directly through creativecommon.org as per the source. This dataset has 5360 unique values collected from 2019, that is 5360 unique customer interactions, and 20 columns. Each column provides crucial information that can be used for analysis. The type of visualization we will be using will be based on our guiding questions. To mention a few are Scatter-plots, Histograms, Box Plots and Normality plots will be used to visualize our analysis based on the requirements of our guiding questions.
The columns, descriptions and potential use are listed below:

Data	Variable	Description
Numerical	Customer ID	Unique customer ID
Categorical (Boolean)	Churn	Churn (0 - returning vs 1 - non-returning customers)
Numerical	Tenure	Tenure of customer in months
Categorical	Preferred Login Device	Preferred login device
Categorical	City Tier	City tier (1 - metropolitan, 2 - developing, 3 - town/village)
Numerical	Warehouse To Home	Distance from warehouse to customer
Categorical	Preferred Payment Mode	Preferred payment method
Categorical	Gender	Gender
Numerical	Hours Spent On App	Number of hours spent on mobile application or website
Numerical	Number Of Device Registered	Total number of devices registered for particular customer
Categorical	Preferred Order Category	Preferred order category of customer from the last month
Categorical	Satisfaction Score	Satisfaction score of customer on service
Categorical	Marital Status	Marital status of customer
Numerical	Number Of Addresses	Total number of addresses for customer
Categorical (Boolean)	Complain	If any complaints have been raised in the last month
Numerical	Order Amount Hike From last Year	Percentage increase of orders from last year
Numerical	Coupon Used	Total number of coupons used in the last month
Numerical	Order Count	Total number of orders placed in the last month
Numerical	Day Since Last Order	Days since last order made by customer
Numerical	Cashback Amount	Average cash back in the last month

Topics to Investigate:

Customer Demographic and Preferences

1. Does the hours spent on the app depend on the purchase frequency(Order count)?

Customer Behavior and Engagement:

2. Is there any correlation between Preferred Order category and gender? Does the preferred

order category have any correlation with marital status?

3. Does the coupon or cashback have an influence on customer order?

Customer Churn and Retention:

4. What is the correlation of satisfaction score and complaints with the churn rate?

Statistical Analysis:

1. **Linear Regression:** Using the number of hours spent by the users on the ecommerce website can lead to more Order Count (Frequence of orders). So using linear regression we will be able to identify whether there is a linear relationship between these 2 features or not. Similarly we can see trends between preferred order category and genders for example Male gender could be more inclined towards Electronic products rather than fashion products as compared to female gender.
2. **Multiple Regression:** We can see trends using two dependent features to determine the third feature such as count of coupons and cashback amount received by the customer can lead to more number of orders or not. To determine this effectively, we could be performing multiple regression statistical techniques.
3. **Logistic Regression:** Through logistic regression statistical technique, we could determine whether the customer will churn or not. Churning is basically when a customer is disappointed with the services or not happy with the ecommerce platform and never returns in future for buying anything. For example, we have a complaint feature and satisfaction score using these 2, we can determine whether the customer will churn (value = 1) or not (value = 0).
4. **Chi-Squared Test:** We will be checking for independence using the chi-square test by testing categorical variables like “Preferred Order” and “Gender” or “Preferred Order” and “Marital Status”. For example: our null hypothesis would be that there is no relation between the two variables suggesting independence and alternative hypothesis would be that there is a significant association between the two variables. We will be able to answer the following hypothesis by using Chi-Squared test

   These are some of the statistical analyses we will conduct for our analysis, and more analysis will be added if required to answer further questions.

References:

[1] Countants, “Why consumer behavior analysis is so relevant to the ecommerce business?,” Medium,Jan. 05, 2020. Available*:* *https://medium.datadriveninvestor.com/why-consumer-behavior-analysis-is-so-relevant-to-the-ecommerce-business-8f49c250ca9c. [Accessed: Oct. 03, 2023]*

[2] “Cc by-nc-sa 4. 0 deed | attribution-noncommercial-sharealike 4. 0 international | creative commons.” Available: *https://creativecommons.org/licenses/by-nc-sa/4.0/.* [Accessed: Oct. 03, 2023]

[3] “Ecommerce customer churn analysis and prediction.”Available: *https://www.kaggle.com/datasets/ankitverma2010/ecommerce-customer-churn-analysis-and-prediction.* [Accessed: Oct. 03, 2023]

[4] S.-C. Necula, “Exploring the Impact of Time Spent Reading Product Information on E-Commerce Websites: A Machine Learning Approach to Analyze Consumer Behavior,” Behavioral Sciences, vol. 13, no. 6, Jun. 2023, doi: 10.3390/bs13060439. Available: *https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10294865/.* [Accessed: Oct. 03, 2023]