Analyzing Customer Purchase Behavior

Insights into Influences on Satisfaction and Loyalty

Meghana Naramgaru  
Business Analytics- Management  
 Wentworth Institute of Technology  
 Boston, MA,USA  
 [naramgarum@wit.edu](mailto:naramgarum@wit.edu)

ABSTRACT

This project investigates customer purchase behavior by analyzing the provided dataset to identify key factors influencing satisfaction and loyalty. The study employs data visualization, statistical models, and predictive analytics to uncover trends and actionable insights for businesses.

KEYWORDS

Customer Behavior, Purchase Trends, Customer Satisfaction, Loyalty Analysis

1 Introduction

Understanding customer purchase behavior is essential for businesses aiming to optimize their marketing strategies and improve customer satisfaction. This project explores demographic trends, product preferences, and the impact of purchase frequency on loyalty. Leveraging analytics, this study seeks to answer critical business questions and identify actionable patterns for enhanced decision-making.

2 Data

2.1 Source of dataset

I downloaded the dataset from Kaggle, a reputable platform widely recognized for hosting credible datasets shared by organizations, researchers, and individuals. Kaggle ensures contributors provide sufficient metadata and maintain data quality, making it a reliable source for analytics projects. The dataset I used, titled 'Customer Purchase Data,' was last updated a year ago. According to the description on Kaggle, the dataset was generated using transaction logs from an e-commerce platform and includes customer demographics, product categories, and purchase details. While the exact methods of data generation were not specified, the dataset appears to reflect real-world scenarios based on its structure and attributes. I did not create the dataset; all credit for its creation goes to the original contributor, and it is publicly accessible on Kaggle at [[Consumer Behavior and Shopping Habits Dataset:](https://www.kaggle.com/datasets/zeesolver/consumer-behavior-and-shopping-habits-dataset)]

2.2 Characters of the datasets

The dataset is in CSV (Comma-Separated Values) format and contains [N rows] and 19 columns, including parameters such as Age (in years), Gender (Male/Female), Purchase Amount (in USD), Item Purchased (categorical), Product Category (categorical), and Payment Method (categorical). These columns provide information on customer demographics, purchasing behaviors, product preferences, payment and shipping options, and promotional activity. Missing values were handled by using forward fill for categorical data (such as Gender, Item Purchased, Payment Method) and mean imputation for numerical fields like Purchase Amount (USD). Additionally, units were standardized by ensuring the Purchase Amount was already in USD. Two new categories were created for analysis: (1) Age Group, which grouped customers into categories such as Teen, Young Adult, Adult, Middle Age, and Senior using defined age bins, and (2) Purchase Channel, which categorized purchases into Online and Offline. These transformations enabled more detailed insights into customer behavior and purchasing patterns, which were critical for the analysis.

3 Methodology

In this analysis, the method used is **Linear Regression**, a statistical technique used to model the relationship between a dependent variable (e.g., Satisfaction\_Score) and one or more independent variables (e.g., Purchase\_Amount). The primary assumption of this model is that the relationship between the dependent and independent variables is linear, and the residuals (errors) are normally distributed with constant variance (homoscedasticity). The advantages of Linear Regression include its simplicity, interpretability, and efficiency for datasets with linear relationships. However, it has disadvantages, such as being sensitive to outliers and performing poorly when the assumptions are violated. Linear Regression was chosen because it is straightforward to interpret the effect of variables like Purchase\_Amount on Satisfaction\_Score. The Python module used for this analysis is scikit-learn, specifically the LinearRegression function from sklearn.linear\_model. Optional adjustments made to improve results include standardizing numerical variables to handle scale differences and removing outliers to reduce their impact on the model. Hyperparameter tuning was not required for this method, but preprocessing steps like missing value imputation and feature scaling ensured better results.

3.1 Heading Level 2

3.2 Heading Level 2

…

Example format: The updated template, user manuals, samples, and required fonts, all are available at the URL <https://www.acm.org/publications/proceedings-template>. It contains said information for all three versions of MS Word (Windows and 2 versions of Mac). There are also separate links to the user guide, which can be referred to by the user. This URL also contains some useful video links, which describe how to add the template, structure the paper, and generate the layout, in different clips. **Display Formula with Number**

 (1)

**Continuation part of Paragraph Text** The user must style this paragraph in **ParaContinue** style, which follows immediately after the **DisplayFormula** (numbered equation). The **DisplayFormula** style is applied only in case of a numbered equation. A numbered equation always has a number to its right. Insert paragraph text here. **Display Formula without Number**



The **DisplayFormulaUnnum** style is applied only in case of an unnumbered equation. An unnumbered display equation never contains an equation number to its right, and this unique property distinguishes it from a numbered equation.



Figure 1: Figure Caption and Image above the caption [In draft mode, Image will not appear on the screen]

**Theorem/Proof/Lemma.** Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement. Insert text here for the enunciation or Math statement.

....Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract, Insert text here for the Quotation or Extract.

4 Results

In this part, you need to select a reasonable way to deliver the result of your topic. For example, equation or numerical results, or visualization of your result. You also need to provide a clear explanation of all results and how to understand the results. If there exist any unexpected results, please explain why or possible cause of this special result. You can use subsection 4.1, 4.2, … to separate your results.

4.1 Heading Level 2

Example format: In the below paragraph, it is explained how alt-txt value is placed in **MS Word 2010**. To add alternative text to a picture in Word 2010, follow these steps:

1. In a Word 2010 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. Select the **Alt Txt** option from the left-side panel options.
4. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

Below are steps to place alt-txt value in **MS Word 2013/2016**. To add alternative text to a picture in Word 2013/2016, follow these steps:

1. In a Word 2013/2016 document, insert a picture.
2. Right click on the inserted picture and select the **Format Picture** option.
3. In the settings at the right side of the window, click on the "Layout & Properties" icon (3rd option).
4. Expand **Alt Txt** option.
5. In the "Title:" and "Description:" text boxes, type the text you want to represent the picture, and then click "Close".

*1.1.1 Heading Level 3.* Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

*1.1.1.1 Heading Level 4.*Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here. Insert paragraph text here.

5 Discussion

Every method/project has its shortage or weakness. Please discuss the unsatisfied results in your project. And discuss the feasible suggestions of future work to revise/improve your result.

6 Conclusion

In this part, you should summarize your project. What important results did you find for your topic and what’s the effect of this result on the real-world?

ACKNOWLEDGMENTS

Insert paragraph text here. Insert paragraph text here.

REFERENCES

Use the following ACM Reference format for your citation

FirstName Surname, FirstName Surname and FirstName Surname. 2018. Insert Your Title Here: Insert Subtitle Here. In *Proceedings of ACM Woodstock conference (WOODSTOCK’18). ACM, New York, NY, USA, 2 pages.* https://doi.org/10.1145/1234567890

[1] Patricia S. Abril and Robert Plant, 2007. The patent holder's dilemma: Buy, sell, or troll? *Commun. ACM* 50, 1 (Jan, 2007), 36-44. DOI: <https://doi.org/>10.1145/1188913.1188915.

[2] Sten Andler. 1979. Predicate path expressions. In *Proceedings of the 6th. ACM SIGACT-SIGPLAN Symposium on Principles of Programming Languages (POPL '79)*. ACM Press, New York, NY, 226-236. DOI:https://doi.org/10.1145/567752.567774

[3] Ian Editor (Ed.). 2007. *The title of book one* (1st. ed.). The name of the series one, Vol. 9. University of Chicago Press, Chicago. DOI:https://doi.org/10.1007/3-540-09237-4.

[4] David Kosiur. 2001. *Understanding Policy-Based Networking* (2nd. ed.). Wiley, New York, NY..

Conference Name:ACM Woodstock conference

Conference Short Name:WOODSTOCK’18

Conference Location:El Paso, Texas USA

ISBN:978-1-4503-0000-0/18/06

Year:2018

Date:June

Copyright Year:2018

Copyright Statement:rightsretained

DOI:10.1145/1234567890

RRH: F. Surname et al.

Price:$15.00