Statistical and Predictive Analysis of Financial and Radical Progress and Performance of Nexus Apparel Store

Fouzan Asif[1]

10th December, 2023

- *Fouzan Asif. Author is with the National University of Computer and Emerging Sciences, Karachi, Pakistan 75300. E-mail: k238054@nu.edu.pk*

Abstract

This research delves into an in-depth analysis of a meticulously crafted dataset containing 300,000 samples and 14 distinct features, synthesized through a dataset_generator.py script. The study extensively explores and addresses over 20 crucial inquiries spanning data visualization, probability assessments (including conditional, total, independent, and dependent probabilities), Market Basket Analysis principles, and strategic evaluations of financial policies. Key objectives encompass identifying synergistic product combinations, isolating underperforming product categories impacting net profit margins, and extracting actionable insights vital for informed decision-making across marketing and financial landscapes.

Introduction

In contemporary business landscapes, the utilization of statistical and mathematical methods is imperative for comprehensive insights into complex datasets. This study focuses on leveraging such methodologies for a robust analysis of the Nexus Apparel Store dataset, a product of a custom dataset generation script. The dataset comprises 14 columns encompassing essential information such as Customer ID, Customer Name, Email, Order Date, Ordered Items, Item Type, Total Bill, Actual Cost, Tax, Net Profit, Mode of Shopping, Coupon Code, Product Prices, and Product Ratings.

The dataset generation process aimed to simulate real-world retail scenarios, capturing diverse customer behaviors, transactional details, and product-related metrics. Employing statistical and mathematical techniques ranging from exploratory data analysis to probability assessments, hypothesis testing,

and Market Basket Analysis, this study endeavors to extract meaningful insights. These methods enable the identification of latent patterns, uncovering correlations, and deriving predictive models essential for optimizing marketing strategies, discerning influential financial policies, and enhancing overall business performance.

Through a systematic analysis of the Nexus Apparel Store dataset, this research anticipates revealing critical associations between customer behaviors and financial performance metrics. The study's conclusions aim to offer actionable insights, enabling stakeholders to make informed decisions, optimize product offerings, and strategize more effectively for sustained growth and profitability.

Methodology

1. Dataset Description

The dataset utilized in this study comprises 17 features capturing various transactional and customer-related information within the Nexus Apparel Store. Below is a summary of the dataset information and descriptive statistics:

Table 1: Dataset Summary

Column	Non-Null Count	Data Type
Customer ID	300000	int64
Customer Name	300000	object
Email	300000	object
Order Date	300000	datetime64[ns]
Ordered Items	300000	object
Item Type	300000	object
Total Bill	300000	float64
Actual Cost	300000	float64
Tax	300000	float64
Net Profit	300000	float64
Mode of Shopping	285776	object
Coupon Code	285776	object
Product Prices	300000	object
Product Rating	300000	object
Cumulative Net Profit	300000	float64
YearMonth	300000	period[M]
Rating Mean	300000	float64

2. The Questions

Probability

1. Discover the Joint Probability Distribution among 3 Discrete Categorical variables. Find how does item type affect the Mode of Shopping Chosen

and the Coupon Code selected.

2. Find the probability of average Ratings exceeding 3.5 over the next 30 days.
3. What's the probability that a person will buy both Boots and Sandals in a single order?
4. What's the probability that a person will buy ONLY both Boots and Sandals in a single order?
5. Given that a person has purchased 'Traditional Dress' and 'Western Suit', what's the probability that he/she will purchase something else from close within the same order line?
6. Given that a person has purchased 'Traditional Dress' and 'Western Suit', what's the probability that he/she will purchase exactly 3 more items from clothes within the same order line?
7. Find the probability that a person will buy at least one item from Perfume, Clothes, Bags, and Shoes only in a single order.
8. Given that a person has purchased "Western Suit" and "Sandals", what's the probability that he/she will also buy a perfume and a designer bag?
9. Find probability distribution for ratings per maximum number of items bought within an order (to see if rating type is in any way affected by the behavior of customer and if the customer is giving away forged ratings).
10. What's the probability that every customer will shop again?
11. Find Probability of Choosing Mode of Shopping given the products.
12. Find the expected value for net profit for customers using coupon code 'DISCOUNT20'.
13. Which product has performed the least and should be removed?

Statistical Relational Findings

1. Find the average ratings each product has maintained monthly.
2. Find the expected time when the Net Cumulative Profit will double its yearly total.
3. Find the least contributing item type and its contribution.
4. Find the net profit contribution for each item type.
5. Find the least contributing and contributions of all the products to the net profit.

6. What is the effect of Mode of Shipping on Rating? Create a probability distribution to study the effect that what if rating is being disturbed due to Mode of Shopping.

Market Basket Analysis

1. Find the products that can be most purchased (in pairs).
2. As a shop owner, what should I know about couplings of my products and items bought together and which pairing should I choose for maximum benefit?

Predictive Analysis

1. Given Prior probabilities and likelihood, predict the coupon code.
2. Find probability distribution, expected value, and variance of ratings variable (to predict every next order's rating and improve it).

Statistical Testing

1. To test if there is a significant difference between the average total bill of Feb 2023 and Dec 2023.
2. To test if there is a significant difference between the average net profit of Bags and Clothes.
3. To test if there is a significant mean difference between Actual cost for FREESHIP and GET50OFF coupon codes.
4. To test if people with repeat orders have different average total bill than people with single orders.
5. To test if there is a significant difference in net profit for Discount20 and Get50Off coupon codes.
6. To test if there is a significant difference in average net profit among modes of shopping.

3. Data Visualization

Various visualization techniques were employed to explore and understand the dataset. Elements such as KDE plots, bar graphs for different columns, scatter plots, etc., were used to gain insights into the data's characteristics and relationships.

4. Statistical Methods

Key statistical methods utilized in this study include measures of central tendency, hypothesis testing (t-tail), conditional, total, independent, and dependent probability assessments, expected value calculations, and finding values using statistics. These methods were applied to aid the study's objectives.

- Relational Properties were used to find meaningful information and data patterns such as which value has outperformed or under-performed.
- To find expected values, variances, and probability distributions, the **probablistic methods** were used to exhibit likeliness of an event such as joint probability, conditional probability, total probability, Bayes rule, Exclusion principles etc.
- **Hypothesis testing** was used to test the inferential existence of sample's characteristics with population characteristics.
- **Market Basket Analysis** was used to find the most useful pairing of products and items (the ones that generate maximum profit and acquire most rating)

5. Predictive Analysis

The study involved the usage of Gaussian Naive Bayes and probability distributions for predictive analysis. Model accuracy, prior, likelihood, and posterior probabilities were evaluated to derive predictive insights.

Implementation and Testing

The Colab Notebook attached covers a good deal of clean work. A reader may not completely comprehend and a writer may not completely show his/her work. Thus, the colab notebook is preferred to be seen.

Conclusion

In this study, we delved into an extensive analysis of the Nexus Apparel Store dataset, encompassing a wide array of statistical, probabilistic, and predictive methodologies. The key findings and insights derived from this analysis are as follows:

- **Probability Insights:** We uncovered various probabilities associated with customer behavior, such as the likelihood of purchasing specific items together, the probability of repeat purchases, and the impact of different factors on shopping decisions.

- **Statistical Relationship Revelations:** Our analysis revealed significant statistical relationships, including the influence of product types on customer ratings and contributions of various product categories to net profits.

- **Market Basket Analysis Implications:** Insights from market basket analysis shed light on product pairings, providing valuable guidance for optimizing product offerings.

- **Predictive Analysis Results:** Leveraging predictive analysis techniques allowed us to make informed predictions regarding coupon codes and future ratings, enhancing our understanding of customer preferences and behaviors.

- **Statistical Testing Outcomes:** Various statistical tests were performed, elucidating differences in total bills, net profits, and customer behaviors across different periods and modes of shopping.

These findings collectively emphasize the significance of employing statistical and predictive analysis in understanding customer behaviors, optimizing product offerings, and maximizing profits within the Nexus Apparel Store.

Furthermore, the results obtained from this analysis serve as a foundation for formulating strategic decisions aimed at enhancing customer satisfaction, improving product recommendations, and optimizing business operations within the retail domain.

However, it's essential to acknowledge the limitations of this study, such as potential biases in the dataset, limited predictive accuracy, and assumptions made during the analysis. Future research and enhancements in data collection and analysis methodologies could further refine and augment the insights obtained in this study.

Overall, the multifaceted analysis presented in this study lays the groundwork for informed decision-making and strategic planning within the Nexus Apparel Store, contributing to its growth and success in the dynamic retail landscape.