Title: Customer Satisfaction Survey Analysis for Retailco

Final Project Report

ALY 6010: Probability Theory and Introductory Statistics

College of Professional Studies

Northeastern University

Fall B 2024

Instructor- Ganesh Subramanian, PhD.

Student:

Oluwatomisin V. Jegede

Date: December 14, 2024

Introduction

RetailCo, a fictional leading retail chain, aims to enhance its understanding of customer behavior to improve marketing strategies and loyalty programs. This analysis investigates key demographic and behavioral variables, evaluates group differences, and explores relationships between variables using inferential statistics and regression analysis.

Through hypothesis testing and regression modeling, this report identifies patterns, tests significant relationships, and provides actionable insights to improve customer engagement and drive revenue growth. Key questions were derived from an initial exploratory data analysis (EDA), which provided a foundation for deeper analysis. The findings are presented with visualizations and statistical summaries, and recommendations are tailored to align with RetailCo's strategic goals.

Key Analyses and Findings

1. Customer Demographics and Spending Behavior

Overview

The RetailCo dataset includes information on customer demographics, income levels, loyalty tiers, purchase frequencies, and spending scores. The data reveals the following:

- **Age Distribution**: Customers are predominantly middle-aged (mean: 34.5 years).

- **Annual Income**: Most customers fall within the $40,000–$60,000 range.

- **Spending Scores**: The average Spending_Score is close to the industry benchmark of 50, with a standard deviation of 14.3.

Insights

-Loyalty Tier Analysis: Higher loyalty tiers, such as "Platinum," tend to have higher spending scores, but income differences between tiers are negligible.

-Regional Trends: The South region hosts a larger proportion of high-spending customers compared to other regions, highlighting potential regional marketing opportunities.

---

2. Hypothesis Testing and Group Comparisons

a. Spending Score vs. Industry Benchmark

Question: Is RetailCo's average Spending_Score significantly greater than the industry standard of 50?
Test: One-sample t-test.

Dependent Variable: Spending_Score (Customer spending score).

Independent Variable: None (This is a one-sample test comparing the sample mean to a known industry standard of 50).

Null Hypothesis (H_0): The mean Spending_Score ≤ 50 (The average spending score is not significantly different from the industry standard of 50).

Alternative Hypothesis (H_a): The mean Spending_Score > 50 (The average spending score is significantly greater than the industry standard of 50).

Results: The T-value was -1.05 (P-value: 0.85).
Interpretation: The average Spending_Score does not differ significantly from the benchmark. RetailCo is competitive in this area but could improve by targeting low-spending customers through personalized campaigns.

---

b. Income Differences Across Loyalty Tiers

Question: Do customers in "Gold" and "Platinum" tiers differ in Annual_Income?

Test: Two-sample t-test.

Dependent Variable: Annual_Income (Customer income).

Independent Variable: Loyalty_Tier (Categorical variable with two levels: "Gold" and "Platinum").

Null Hypothesis (H_0): The mean Annual_Income is the same across the "Gold" and "Platinum" tiers.

Alternative Hypothesis (H_a): The mean Annual_Income is different across the "Gold" and "Platinum" tiers.

Results: The T-value was 1.56 (P-value: 0.12).
Interpretation: No significant income differences exist between these tiers, suggesting that loyalty progression is not income-dependent. RetailCo's loyalty program may benefit from offering more tier-specific rewards.

---

c. Regional Differences in High-Income Customers

Question: Are the proportions of high-income customers significantly different between the North and South regions?
Test: Z-test for proportions.

Dependent Variable: Proportion of customers in the "High" income bracket.

Independent Variable: Region (Categorical variable with two levels: "North" and "South").

Null Hypothesis (H_0): The proportion of "High" income customers is the same in both regions.

Alternative Hypothesis (H_a): The proportion of "High" income customers differs between the two regions.

Results: The Z-score was -0.37 (P-value: 0.71).
Interpretation: The proportion of high-income customers is consistent across regions, indicating uniform brand penetration. RetailCo can apply consistent marketing efforts across regions.

---

3. Regression Analysis and Relationships

a. Annual Income and Spending Score

Question: Is there a significant relationship between Annual_Income and Spending_Score?

Dependent Variable: Spending_Score (Customer spending score).

Independent Variable: Annual_Income (Customer annual income).

Null Hypothesis (H_0): There is no significant relationship between Annual_Income and Spending_Score.

Alternative Hypothesis (H_a): There is a significant relationship between Annual_Income and Spending_Score.

Results: Regression analysis revealed a slope of 0.00001 (P-value: 0.42).
Interpretation: No significant relationship exists. Spending behavior is likely driven by factors other than income, such as loyalty engagement or product preferences.

---

b. Purchase Frequency and Loyalty Tiers

Question: Does Purchase_Frequency vary significantly across loyalty tiers?

Dependent Variable: Purchase_Frequency (Number of purchases per year).

Independent Variable: Loyalty_Tier (Categorical variable with different loyalty levels).

Null Hypothesis (H_0): There is no difference in Purchase_Frequency across loyalty tiers.

Alternative Hypothesis (H_a): There is a difference in Purchase_Frequency across loyalty tiers.

Results: ANOVA yielded an F-statistic of 0.11 (P-value: 0.90).

Interpretation: No significant differences were found. Loyalty tiers do not seem to incentivize higher purchase frequencies, suggesting the need for more engaging tier benefits.

---

c. Age and Spending Score

Question: Is there a significant relationship between Age and Spending_Score?

Dependent Variable: Purchase_Frequency (Number of purchases per year).

Independent Variable: Annual_Income (Customer annual income).

Null Hypothesis (H_0): There is no significant relationship between Purchase_Frequency and Annual_Income.

Alternative Hypothesis (H_a): There is a significant relationship between Purchase_Frequency and Annual_Income.

Results: Regression analysis yielded a slope of -0.01 (P-value: 0.22).

Interpretation: No significant linear relationship was found. RetailCo should explore nonlinear patterns to capture potential age-based spending behaviors.

---

d. Purchase Frequency and Annual Income

Question: Is there a significant relationship between Purchase_Frequency and Annual_Income?

Dependent Variable: Purchase_Frequency (Number of purchases per year).

Independent Variable: Spending_Score (Customer spending score).

Null Hypothesis (H_0): There is no significant relationship between Purchase_Frequency and Spending_Score.

Alternative Hypothesis (H_a): There is a significant relationship between Purchase_Frequency and Spending_Score.

Results: Regression analysis yielded a slope of 25.45 (P-value: 0.31).

Interpretation: No significant relationship was observed. Purchase frequency is likely influenced by factors other than income.

Recommendations and Further Analysis

Based on the findings, RetailCo should consider the following strategies:

-Enhance Loyalty Programs: Redesign loyalty tiers to better differentiate rewards and incentivize frequent purchases. For example:

- Introduce time-sensitive promotions to encourage regular shopping.
- Offer exclusive rewards for high-spending or frequent customers.

-Leverage Regional Insights: The South region shows strong customer engagement. RetailCo could replicate successful strategies from this region across other locations.

-Focus on Personalization: Use behavioral data, such as purchase history and product preferences, to develop personalized campaigns that appeal to customers' unique needs.

-Investigate Nonlinear Relationships: Explore more complex patterns, such as quadratic models, to identify age-related spending trends or purchase behaviors.

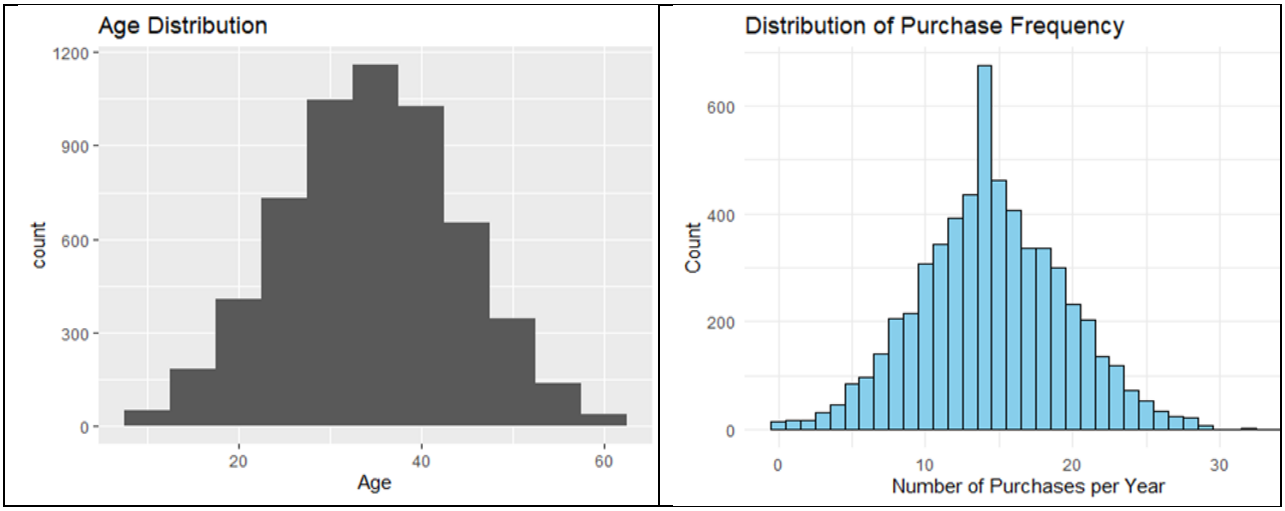
Further Analysis

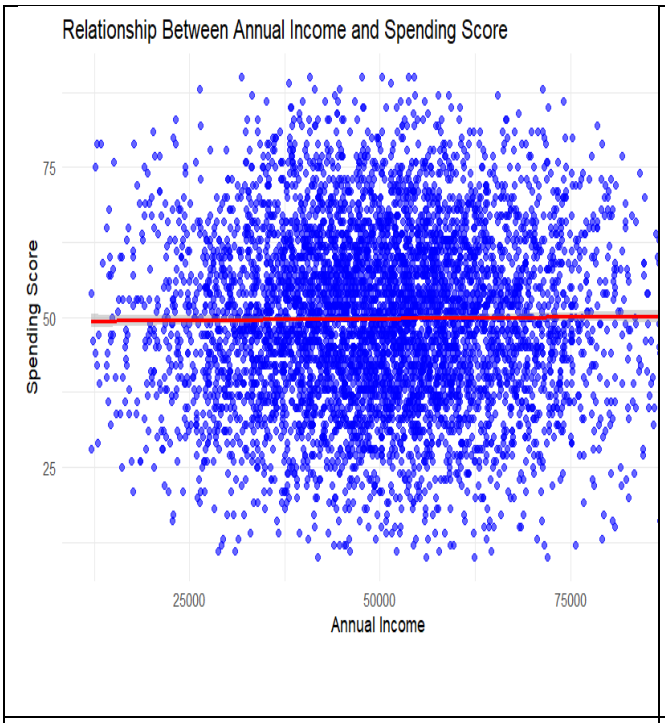
RetailCo could conduct cluster analysis or predictive modeling to segment customers more effectively. Additionally, qualitative data, such as customer feedback, may provide insights into improving satisfaction and loyalty.

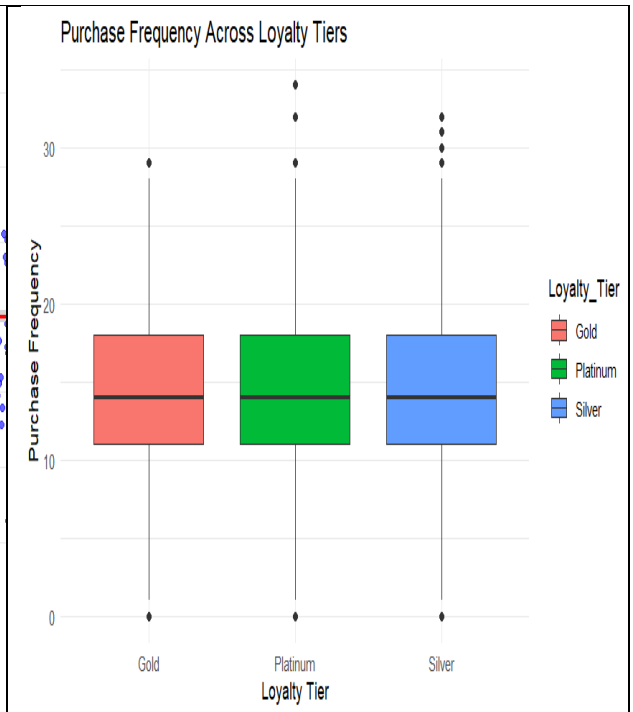
Conclusion

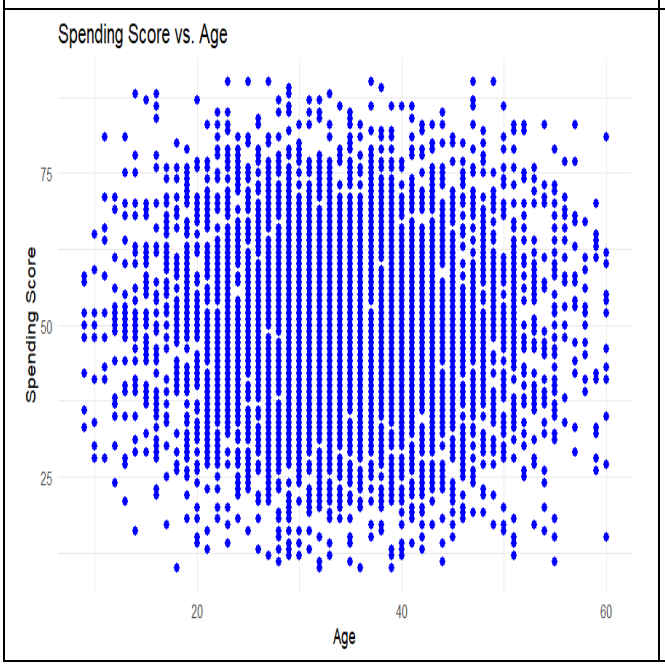
This analysis provides a detailed exploration of RetailCo's customer data, including demographic trends, group comparisons, and relationships between key variables. While many results showed no significant differences or relationships, the findings highlight opportunities for improvement in loyalty programs, targeted marketing, and regional campaigns. By implementing the recommended strategies, RetailCo can strengthen customer engagement and drive business growth.

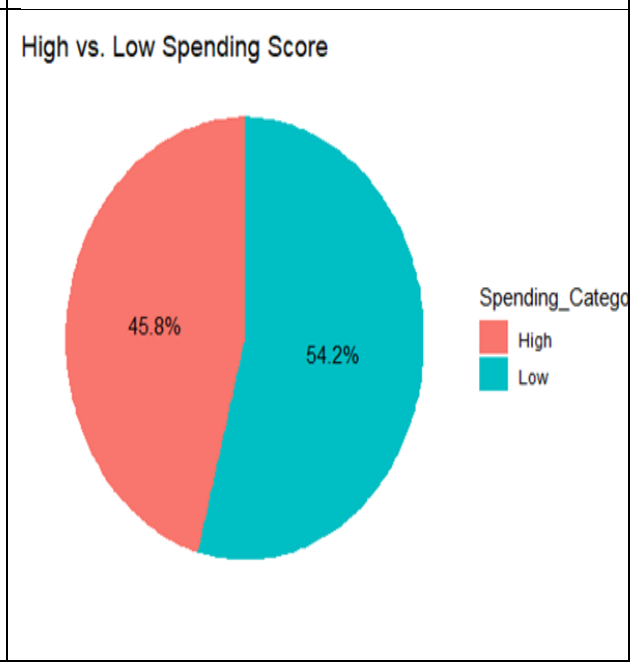
Appendix




Relationship Between Annual Income and Spending Score
Spending Score
75
50
25
25000
50000
75000
Annual Income


Purchase Frequency Across Loyalty Tiers
Purchase Frequency
30
20
10
0
Gold
Platinum
Silver
Loyalty Tier
Loyalty_Tier
Gold
Platinum
Silver


Spending Score vs. Age
Spending Score
75
50
25
20
40
60
Age


High vs. Low Spending Score
45.8%
54.2%
Spending_Catego
High
Low

References

ChatGPT. (2024). *Customer Satisfaction Dataset*.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage Publications.

R Core Team. (2023). *R: A language and environment for statistical computing* (Version 4.3.0) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/