Sarah Taghadosi
400222019
March 24, 2025

Customer Personality Analysis

Hypothesis Testing

Abstract

This report explores customer behavior using a personality dataset that includes demographic details, marketing responses, and spending patterns. The objective is to uncover insights about customer characteristics and their relationship with income, promotional effectiveness, and product preferences. We applied five non-parametric statistical tests, including the Kruskal-Wallis H Test, Wilcoxon Signed-Rank Test, Mann-Whitney U Test, Friedman Test, and McNemar Test. Each test addressed a specific research question related to marketing performance or demographic influence.The results reveal significant differences in income across education levels, changes in spending after campaigns, and behavioral variations based on household structure. These findings can support better targeting and personalization in future marketing strategies.

1. Introduction

This project analyzes customer data to understand how personality traits and demographic characteristics influence purchasing behavior. The dataset includes variables such as income, education level, number of children, and customer responses to various marketing campaigns.

The goal is to examine whether certain customer segments behave differently in terms of spending and purchasing decisions. We perform statistical hypothesis testing to validate these assumptions and generate insights that could help optimize marketing strategies.

2. Research Questions

1. The following research questions are addressed using non-parametric statistical methods:
2. 1. Do customers from different education levels have different income levels?
3. 2. Is there a significant difference in spending before and after a marketing campaign?
4. 3. Do customers with children spend differently than those without children?
5. 4. Is there a significant difference in customer satisfaction across multiple product categories?
6. 5. Did the number of purchases change significantly after a new marketing strategy?

3. Data

The dataset used in this project is a marketing campaign dataset containing customer-level information. It includes variables related to demographics (such as age, income, education, number of children), spending across product categories, and responses to past marketing campaigns.

Prior to analysis, we removed missing values from the Income column to ensure validity of the statistical tests. New variables were also created to support the hypotheses, including:

- `Total_Spending`: Sum of spending across six product categories
- `Has_Children`: Indicates whether the customer has children (Kidhome +Teenhome > 0)
- `Spending_Before` and `Spending_After`: Total purchases before and after the last campaign
- `Purchase_Before` and `Purchase_After`: Binary variables based on purchase response

The dataset was explored visually to better understand the distributions and guide the selection of appropriate tests.

4. Methods

To analyze the customer behavior in this project, we used a set of non-parametric statistical tests. These tests were selected based on the structure of the data and the nature of each research question:

- Kruskal-Wallis H Test : Used to compare the distribution of income across multiple independent education groups.
- Wilcoxon Signed-Rank Test : Applied to paired spending data before and after a marketing campaign to detect significant changes.
- Mann-Whitney U Test : Used to compare total spending between households with and without children.
- Friedman Test : Used to detect differences in spending across multiple product categories for the same individuals.
- McNemar Test : Applied to paired binary variables to assess changes in purchasing behavior before and after the final campaign.

All tests were conducted using Python's `scipy.stats` library, and their results were interpreted alongside visualizations such as boxplots and heatmaps.

5. Analysis and Results

5.1 Do customers from different education levels have different income levels?

To address whether customers with different education backgrounds earn significantly different income levels, we applied the Kruskal-Wallis H Test. This non-parametric test is suitable when comparing more than two independent groups where the income distribution is not assumed to be normal.

We grouped customers by education level and compared their income distributions.

- Test statistic: 142.17
- P-value: 9.69e-30

Interpretation:
Since the p-value is far below 0.05, we reject the null hypothesis. This means that at least one education group differs significantly from the others in terms of income.
This result implies that education level has a statistically significant impact on income, which is an important insight for customer segmentation and marketing strategy.

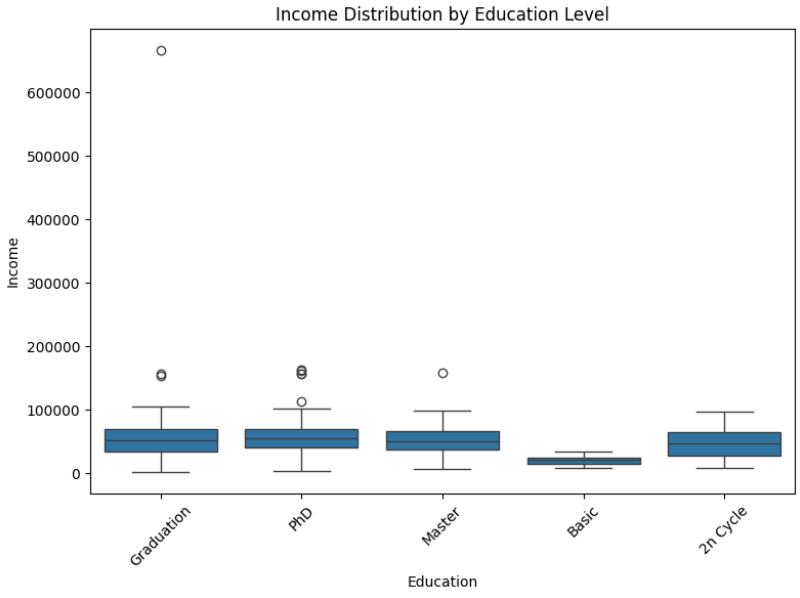


Figure 1. Income Distribution by Education Level

5.2 Is there a significant difference in spending before and after a marketing campaign?

To determine whether there is a significant difference in customer spending before and after the marketing campaign, we conducted the Wilcoxon Signed-Rank Test.
This non-parametric test is appropriate for comparing two related samples—in this case, the total spending of the same customers before and after the campaign.

- Test statistic: 0.000
- P-value: 2.44e-56

Interpretation:

Since the p-value is far below 0.05, we reject the null hypothesis. This indicates a statistically significant difference in customer spending before and after the campaign.
The boxplot visualization further confirms this result, showing a shift in the distribution of spending.

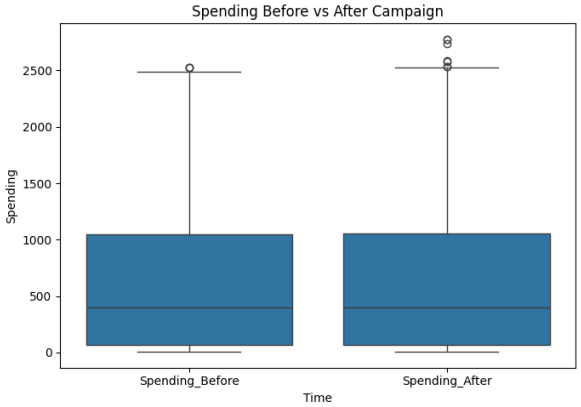


Figure 2. Spending Before and After Marketing Campaign

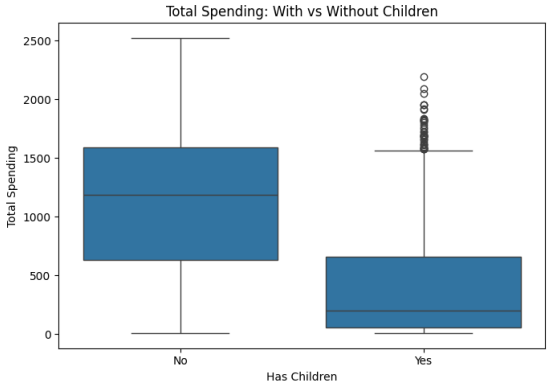
5.3 Do customers with children spend differently than those without?

To examine whether having children influences customer spending behavior, we categorized customers into two groups: those with children and those without. We calculated the total spending for each customer across all product categories and applied the Mann-Whitney U Test, a non-parametric test used to compare two independent groups.

- Test statistic: 203274.000
- P-value: 3.75e-106

Interpretation:
Given the p-value is significantly below 0.05, we reject the null hypothesis. This result indicates a statistically significant difference in total spending between customers with children and those without. Specifically, customers without children tend to spend more, which can be an important insight when tailoring marketing efforts or promotions based on family status.

Figure 3. Total Spending: With vs Without Children



5.4 Is there a significant difference in customer satisfaction across multiple product categories?

To evaluate whether customer satisfaction differs across various product categories, we used the Friedman Test. This non-parametric test is appropriate when analyzing repeated measures — in this case, spending amounts are used as a proxy for satisfaction across product categories.

We considered spending values for six product categories: Wines, Fruits, Meat, Fish, Sweets, and Gold Products. By focusing on customers who have made purchases in all categories, we ensured valid comparison within the same group.

- Test statistic: 5889.561
- P-value: 0.00000

Interpretation:
Since the p-value is significantly below 0.05, we reject the null hypothesis. This indicates that customers' satisfaction — inferred from their spending — varies significantly across product categories. Such variation provides valuable insights for inventory planning and product-specific marketing efforts.

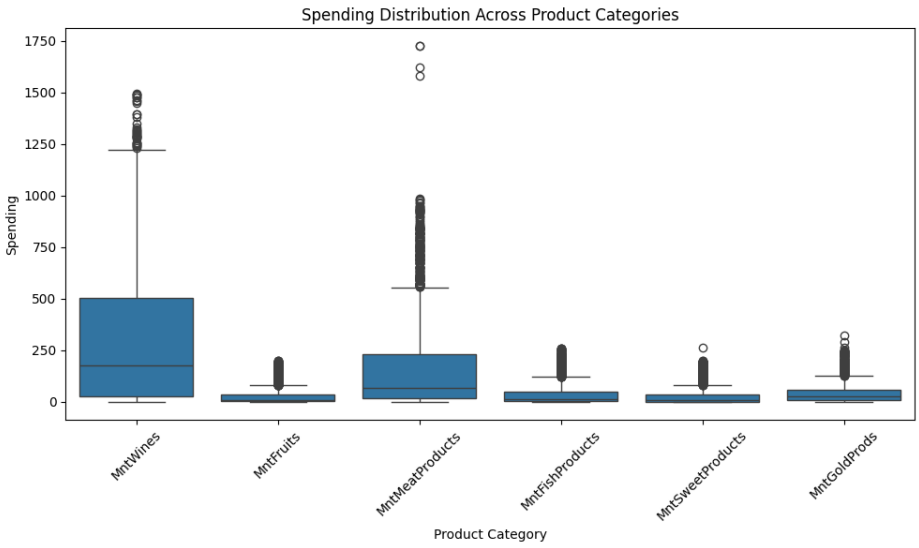


Figure 4. Spending Distribution Across Product Categories

5.5 Did the number of purchases change significantly after a new marketing strategy?

To evaluate whether the marketing campaign led to a significant change in purchasing behavior, we used the McNemar Test. This test is ideal for analyzing paired categorical data, especially for before-and-after designs.

We built a contingency table that compares the number of customers who made a purchase before and after the campaign:

	Purchase After: No	Purchase After: Yes
Purchase Befor: No	1611	146
Purchase Before: Yes	272	187

Table 1. Purchases Before and After Campaign

- Test statistic: 37.380
- P-value: 9.71949e-10

Interpretation:
Since the p-value is far below 0.05, we reject the null hypothesis. This indicates a statistically significant change in purchasing behavior after the campaign. The number of customers who began purchasing after the campaign (146) is substantially different from those who stopped purchasing (272), implying the marketing campaign had a measurable impact.

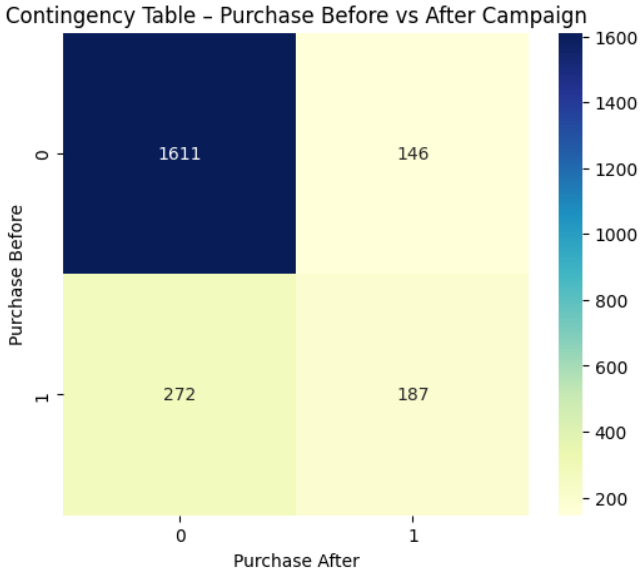


Figure 5. Contingency Table – Purchase Before vs After Campaign

6. Conclusion

This report analyzed customer behavior using statistical tests. We found that:

- Income differs significantly across education levels.
- Spending changed after the marketing campaign.
- Households without children spend more.
- Customers prefer certain product categories, especially wine and meat.
- The campaign had a measurable impact on purchasing behavior.

These insights support more targeted marketing and personalized strategies.