

MMAI 5040

Impact of Promo Code Usage on Purchase Amounts
Final Report

Pengbo Gao (220307831)

Yuhui Gu (220128716)

Yaosheng Liu (216173163)

Yisong Lu (216379927)

Lingrui Yu (220771929)

Jingze Zhang (220240248)

Tianchen Zhu (217755976)

Dec 3rd, 2023

Table of Contents

| | |
|---|----|
| Executive Summary | 1 |
| 1.0 Analytics Problems and Business Problems | 1 |
| 2.0 Descriptive Analysis | 1 |
| 2.1. Tabular and Graphical Analysis..... | 1 |
| 2.2. Segmentation and Control Analysis..... | 6 |
| 2.3. Summary of Descriptive Analysis | 9 |
| 3.0 Causality Analysis | 10 |
| 3.1. Propensity Score Matching and Ordinary Least Squares Regression..... | 10 |
| 3.2. Instrumental Variable Analysis..... | 12 |
| 4.0 Overall Conclusion | 14 |
| 4.1. Addressing the Analytics Problem..... | 14 |
| 4.2. Business Insights and Implementations | 15 |
| 4.3. Potential Implementation Issues | 15 |
| Reference | 16 |

List of Figures

| | |
|--|----|
| Figure 1 - Count of Promo Code Usage..... | 3 |
| Figure 2 - Promo Code Usage Across Locations..... | 3 |
| Figure 3 - Promo Code Usage Across Shipping Types | 4 |
| Figure 4 - Promo Code Usage Across Categories | 4 |
| Figure 5 - Promo Code Usage Across Items..... | 5 |
| Figure 6 - Promo Code Usage Across Season | 5 |
| Figure 7 - Top 10 Most Correlated Pairs | 6 |
| Figure 8 - Ordinary Least Squares (OLS) Regression Analysis of Purchase Amount | 11 |
| Figure 9 - Instrumental Variables Regression Summary | 13 |

List of Tables

| | |
|--|----|
| Table 1 - Statistical Information for the Dataset..... | 1 |
| Table 2 - Customer Purchase Behavior | 8 |
| Table 3 - Standardized Mean Differences for Customer Variables | 10 |

Executive Summary

This report investigates the causal relationship between the usage of promo codes and the purchase amounts in USD. By applying descriptive statistics and predictive modeling, we aim to provide insights into the effectiveness of promo codes as a sales strategy and their potential impact on customer purchasing behavior. This analysis will help determine the effectiveness of promo codes in potentially increasing the average order value.

1.0 Analytics Problems and Business Problems

The investigation will examine in depth the causal relationship between the use of promo codes and the corresponding impact on customer spending, focusing on the average purchase amount. The investigation aims to determine whether the strategic implementation of promo codes has played a significant role in catalyzing a change in customer spending patterns and has the potential to lead to an increase in purchases. By applying a suite of statistical methods to the analysis, the aim is to gain insights into how promo codes can be a lever for expanding sales and deepening customer engagement. This detailed review will guide how we deliver promo codes, illustrating how important they are to improving our marketing and making customers better shoppers.

2.0 Descriptive Analysis

2.1. Tabular and Graphical Analysis

Table 1 - Statistical Information for the Dataset

| | Age | Purchase_Amount_USD | Review_Rating | Previous_Purchases |
|-------|-------------|---------------------|---------------|--------------------|
| count | 3900.000000 | 3900.000000 | 3900.000000 | 3900.000000 |
| mean | 44.068462 | 59.764359 | 3.749949 | 25.351538 |
| std | 15.207589 | 23.685392 | 0.716223 | 14.447125 |
| min | 18.000000 | 20.000000 | 2.500000 | 1.000000 |
| 25% | 31.000000 | 39.000000 | 3.100000 | 13.000000 |
| 50% | 44.000000 | 60.000000 | 3.700000 | 25.000000 |
| 75% | 57.000000 | 81.000000 | 4.400000 | 38.000000 |
| max | 70.000000 | 100.000000 | 5.000000 | 50.000000 |

2.1.1. Age

The dataset includes 3,900 individuals with ages ranging from 18 to 70. The average age is approximately 44, with a standard deviation of 15.21. This indicates that there is a moderate spread of ages around the mean. Half of the individuals are between the ages of 31 and 57, indicating a middle-aged demographic.

2.1.2. Purchase Amount (USD)

The average purchase amount is approximately \$59.76, with a standard deviation of \$23.69, suggesting that while the average spending is around \$60, there is considerable variation in purchase amounts among customers. The range of purchase amounts is from \$20 to \$100. The interquartile range from 25th to 75th percentile is from \$39 to \$81, indicating that the middle 50% of the purchases fall within this range.

2.1.3. Review Rating

The product review ratings average 3.75 out of 5, with a standard deviation of 0.72. This suggests that customers generally leave positive reviews. The range of ratings is from 2.5 to 5, with a median very close to the mean, which implies a relatively even distribution of review ratings without extreme outliers.

2.1.4. Previous Purchases

The average number of previous purchases per customer is about 25, with a standard deviation of 14.45. This high variation shows that there is a wide range of customer loyalty and repeat purchase behavior. The minimum number of previous purchases recorded is 1, and the maximum is 50, indicating both newer and highly frequent customers are present in the dataset. The median value of 25 suggests that half of the customers have a purchase history of 25 or fewer transactions.

2.1.5. Promo Code Usage Frequency

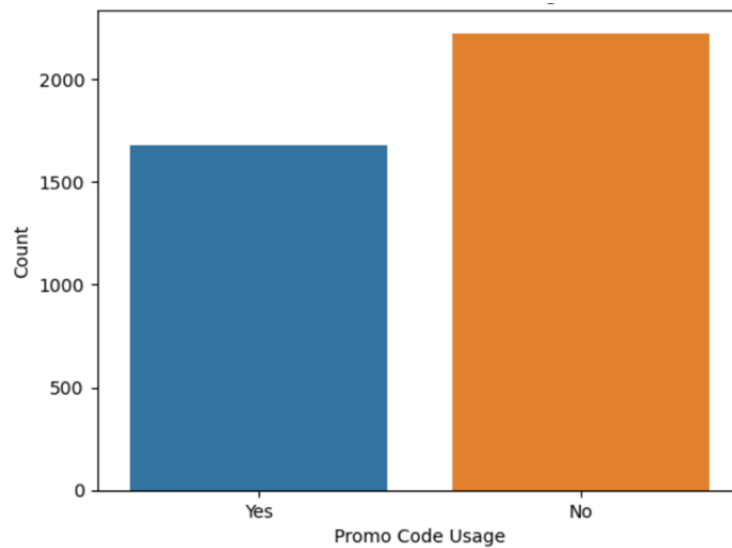


Figure 1 - Count of Promo Code Usage

In Figure 1, the bar chart indicates that most transactions do not involve the use of promo codes, as highlighted by the taller orange bar compared to the blue one.

2.1.6. Distribution Across Locations

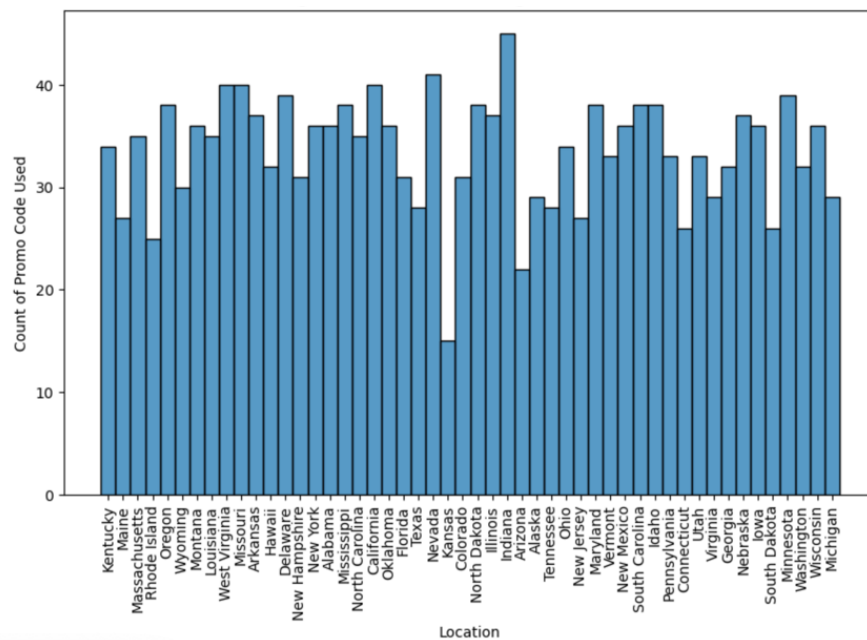


Figure 2 - Promo Code Usage Across Locations

This histogram (Figure 2) reveals a disparity in the usage of promo codes across locations, with Indiana having the highest count and Kansas the lowest. This suggests geographic differences in the adoption or effectiveness of promo codes.

2.1.7. Shipping Types

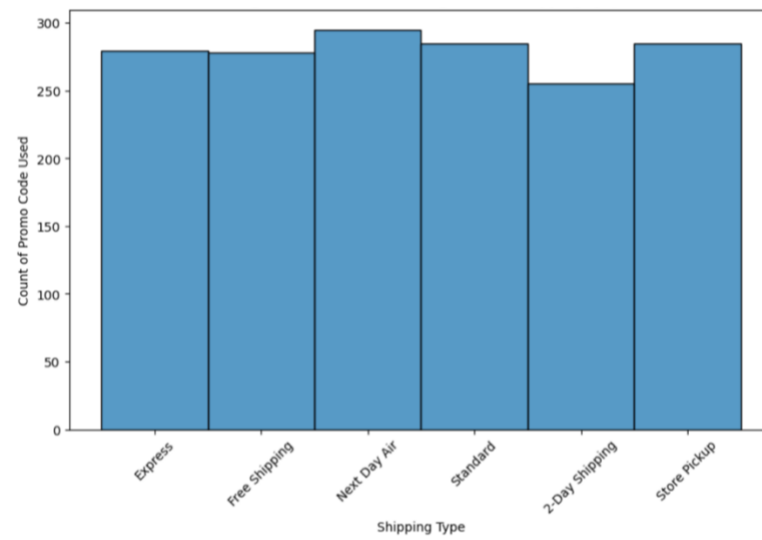


Figure 3 - Promo Code Usage Across Shipping Types

In Figure 3, the distribution of promo code usage among different shipping types appears uniform, as indicated by the nearly equal height of the bars. This consistency suggests that the choice of transportation type has little effect on the likelihood of using a promo code.

2.1.8. Different Categories

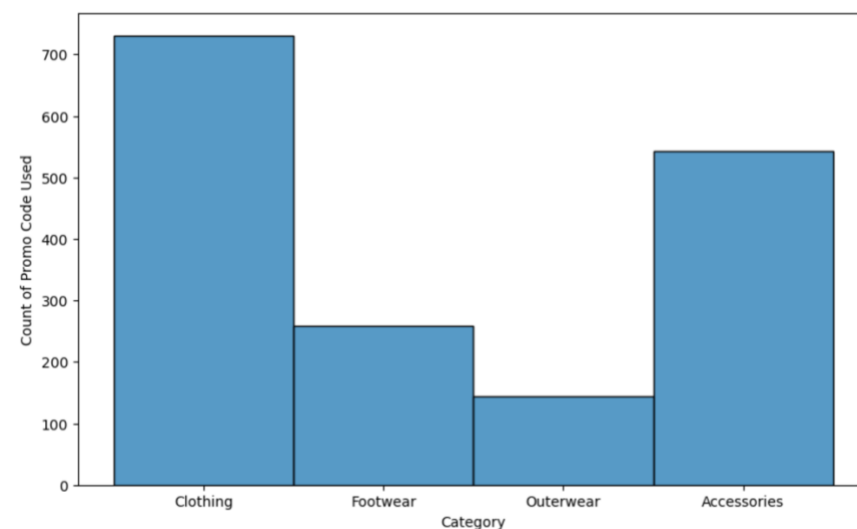


Figure 4 - Promo Code Usage Across Categories

From Figure 4, promo codes are used most frequently for clothing and least frequently for accessories. This pattern indicates that customers may be more inclined to use promo codes for higher-value items or items that are purchased more regularly.

2.1.9. Different Purchase Items

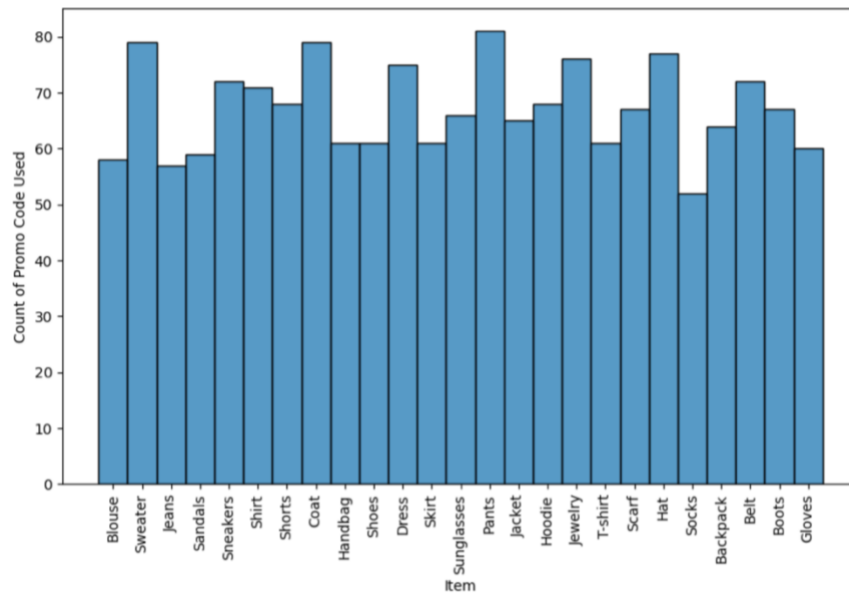


Figure 5 - Promo Code Usage Across Items

From Figure 5, while there is variation in promo code usage across different items, with pants being the highest and socks the lowest, no single item category shows an overwhelming advantage in promo code usage.

2.1.10. Seasonal Trends

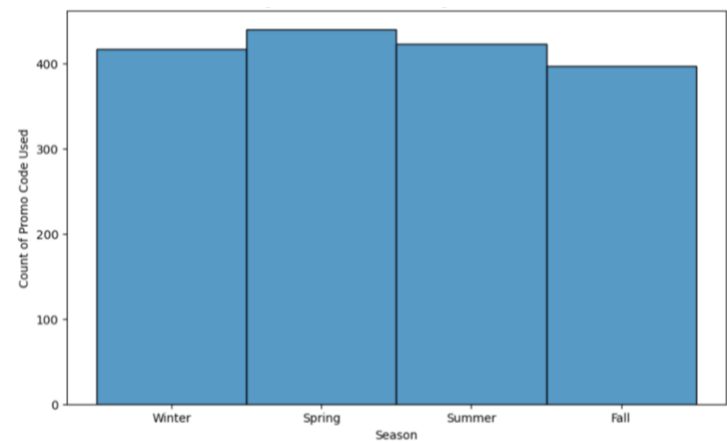


Figure 6 - Promo Code Usage Across Season

The usage of promo codes is relatively consistent throughout the year shown in Figure 6, with only slight variations observed between seasons. This suggests that promo codes are a steady part of the purchasing process regardless of the time of year.

2.1.11. Top 10 Most Correlated Pairs

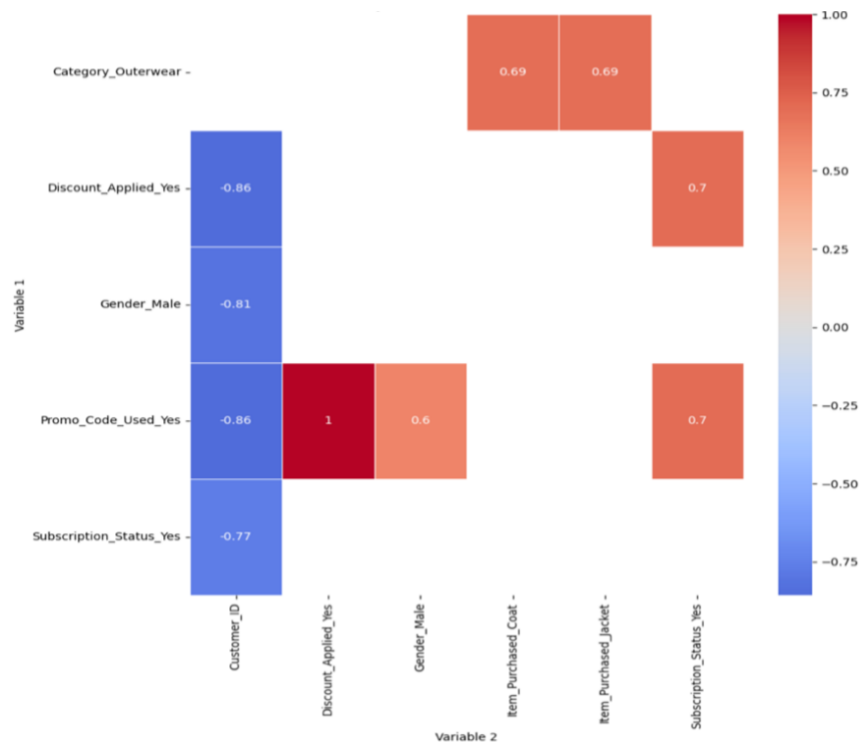


Figure 7 - Top 10 Most Correlated Pairs

This heatmap (Figure 7) shows the top correlated variable pairs. The color intensity represents the strength of the correlation, with red indicating a positive correlation and blue indicating a negative correlation. It seems that 'Discount_Applied_Yes' and 'Promo_Code_Used_Yes' are strongly negatively correlated, suggesting that when a discount is applied, a promo code is less likely to be used, or vice versa.

2.2. Segmentation and Control Analysis

2.2.1. Rationale for Choosing Gender and Location Segments

The application of Segmentation and Control Analysis, focusing on 'Gender' and 'Location', serves as a crucial preliminary step before conducting causality analysis in the project. This approach was chosen for the following reasons:

1. Data Understanding and Pattern Identification: Segmentation analysis provides an initial understanding of the data, uncovering patterns and behaviors across different customer groups. This foundational knowledge is vital for informed causal analysis.
2. Relevance of 'Gender' and 'Location':
 - Gender: Selected due to its influence on consumer behavior, as numerous studies indicate that shopping habits and responsiveness to marketing, including promo codes, can vary significantly between genders. This variation can provide critical insights into how promo codes are perceived and used by different gender groups.
 - Location: Chosen for its potential impact on shopping behavior and promo code usage. Geographic differences can reveal regional trends in spending, promotional response, and accessibility, which are essential for understanding the diverse customer base.
3. Enhancing Causal Inference Quality:
 - The segmentation analysis helps identify relevant covariates for the causality analysis, improving its accuracy by controlling for confounding variables.
 - Understanding the distribution of promo code usage across these segments assists in ensuring a balanced causality analysis.
4. Bias Mitigation: By analyzing different customer segments, potential selection biases that might influence the causal analysis are reduced. This is crucial for ensuring that the treatment (promo code usage) and control groups are comparable in the causal model.

2.2.2. Methodology

Data were segmented by gender and location to analyze the following variables:

1. Average Purchase Amount (USD): The mean spending per transaction within each segment.
2. Promo Code Usage: The proportion of transactions in each segment where a promo code was applied.
3. Review Ratings: The average customer satisfaction rating for purchases in each segment.
4. Previous Purchases: The mean number of prior purchases for customers in each segment.

2.2.3. Analysis Overview

Purchase Amounts by Segment: The data revealed variance in average spending by different gender-location segments, indicating distinct purchasing power or preferences across these groups.

Promo Code Usage Patterns: There was a range of promo code usage rates, with some segments exhibiting higher rates, potentially indicating different levels of price sensitivity or response to marketing promotions.

Customer Satisfaction: Review ratings across segments are generally positive, averaging around 3.5 to 4.0, suggesting a consistent level of product or service satisfaction among customers.

Customer Loyalty Indicators: The frequency of previous purchases varied by segment, hinting at varying degrees of customer loyalty or engagement.

2.2.4. Findings

Table 2 - Customer Purchase Behavior

| | Gender | Location | Purchase_Amount_USD | Promo_Code_Used | Review_Rating | Previous_Purchases |
|-----|--------|----------|---------------------|-----------------|---------------|--------------------|
| 0 | 0 | 0 | 62.148148 | 0.000000 | 3.600000 | 29.925926 |
| 1 | 0 | 1 | 72.083333 | 0.000000 | 3.712500 | 28.333333 |
| 2 | 0 | 2 | 66.320000 | 0.000000 | 3.648000 | 26.040000 |
| 3 | 0 | 3 | 69.640000 | 0.000000 | 3.588000 | 24.400000 |
| 4 | 0 | 4 | 59.862069 | 0.000000 | 3.706897 | 23.172414 |
| ... | ... | ... | ... | ... | ... | ... |
| 95 | 1 | 45 | 60.849057 | 0.547170 | 3.630189 | 22.603774 |
| 96 | 1 | 46 | 65.187500 | 0.666667 | 3.872917 | 24.958333 |
| 97 | 1 | 47 | 67.844828 | 0.689655 | 3.608621 | 22.879310 |
| 98 | 1 | 48 | 57.220000 | 0.720000 | 3.900000 | 22.780000 |
| 99 | 1 | 49 | 63.490196 | 0.588235 | 3.643137 | 29.019608 |

Spending Disparities: Certain gender-location segments spend more on average, which may guide targeted marketing and pricing strategies.

Promo Code Impact: Promo code usage varies significantly by segment, which could influence promotional campaign designs.

Satisfaction Correlation: While satisfaction is generally high, its relationship with spending and promo code usage warrants further examination.

Loyalty Insights: Previous purchase frequency provides a window into customer retention and loyalty, which can be leveraged for customer relationship management.

2.2.5. Limitations

The segmentation analysis sheds light on how different customer groups use promo codes and their spending habits. It aids in developing targeted marketing strategies. However, this analysis does not establish a direct cause-and-effect relationship between promo code use and purchase behavior, and further investigation may be necessary to understand this aspect fully. Future analyses should consider additional variables that may affect purchasing behavior, such as income, education, or specific product preferences.

2.2.6. Conclusion

The segmentation and control analysis provides valuable insights into how different customer segments interact with promo codes, their spending patterns, satisfaction levels, and loyalty. These insights are crucial for developing targeted marketing strategies, improving customer relationship management, and ultimately understanding the overall impact of promo codes on purchase amounts.

2.3. Summary of Descriptive Analysis

The descriptive analysis of a dataset of 3,900 individuals, aged 18 to 70 years with an average age of 44, provides a foundational understanding of promo code usage and its impact on customer behavior. With an average purchase amount of approximately \$59.76 and generally positive reviews averaging 3.75 out of 5, the data reveals varied customer loyalty, indicated by an average of 25 previous purchases per customer. Promo code usage, not widely prevalent in transactions, shows significant variability across locations, shipping types, product categories, and seasons. Additionally, a Segmentation and Control Analysis focusing on 'Gender' and 'Location' uncovers distinct spending patterns, promo code usage rates, satisfaction levels, and loyalty across different customer segments. This comprehensive analysis is instrumental for understanding the nuanced impacts of promo codes, shaping targeted marketing strategies, and informing customer relationship management, setting the stage for further statistical investigation into their causal effect on purchase amounts.

3.0 Causality Analysis

3.1. Propensity Score Matching and Ordinary Least Squares Regression

This method aimed at evaluating the effect of promo code usage on purchase amounts in USD.

The analysis was structured to test the hypothesis that promo codes influence customer spending.

3.1.1. Hypotheses

Null Hypothesis (H_0): The use of a promo code does not statistically affect the purchase amount in USD.

Alternative Hypothesis (H_1): The use of a promo code significantly affects the purchase amount in USD.

3.1.2. Methodology

A PSM analysis was conducted to create a balanced comparison between customers who used a promo code (treatment group) and those who did not (control group). Propensity scores were calculated using logistic regression, considering covariates such as 'Age', 'Review_Rating', 'Previous_Purchases', and 'Category'.

3.1.3. Findings

- Balance Check:

Table 3 - Standardized Mean Differences for Customer Variables

| Variables | Standardized Mean Differences (SMDs) |
|--------------------|--------------------------------------|
| Age | 0.0062093965620813735 |
| Review Rating | 0.01891176956064404 |
| Previous Purchases | 0.009384175251775343 |
| Category | -0.001868217259851074 |

The Standardized Mean Differences for all covariates post-matching were close to zero, indicating a successful matching process and suggesting that the treatment and control groups were well-balanced.

- Impact Analysis:

The Average Treatment effect on the Treated (ATT) was found to be -1.1491, suggesting that the use of a promo code is, on average, associated with a decrease in purchase amount by approximately \$1.15 among those who used the promo code. Also, in Figure 8, the coefficient for 'Promo_Code_Used' mirrored the ATT at -1.1491, reinforcing the observed negative association.

- Statistical Significance:

| OLS Regression Results | | | | | | |
|------------------------|---------------------|---------------------|-----------|-------|--------|--------|
| Dep. Variable: | Purchase_Amount_USD | R-squared: | 0.001 | | | |
| Model: | OLS | Adj. R-squared: | 0.000 | | | |
| Method: | Least Squares | F-statistic: | 1.940 | | | |
| Date: | Mon, 27 Nov 2023 | Prob (F-statistic): | 0.164 | | | |
| Time: | 18:42:11 | Log-Likelihood: | -15402. | | | |
| No. Observations: | 3354 | AIC: | 3.081e+04 | | | |
| Df Residuals: | 3352 | BIC: | 3.082e+04 | | | |
| Df Model: | 1 | | | | | |
| Covariance Type: | nonrobust | | | | | |
| | coef | std err | t | P> t | [0.025 | 0.975] |
| const | 60.4281 | 0.583 | 103.576 | 0.000 | 59.284 | 61.572 |
| Promo_Code_Used | -1.1491 | 0.825 | -1.393 | 0.164 | -2.767 | 0.469 |
| Omnibus: | 3840.315 | Durbin-Watson: | 1.960 | | | |
| Prob(Omnibus): | 0.000 | Jarque-Bera (JB): | 212.328 | | | |
| Skew: | -0.006 | Prob(JB): | 7.83e-47 | | | |
| Kurtosis: | 1.767 | Cond. No. | 2.62 | | | |

Figure 8 - Ordinary Least Squares (OLS) Regression Analysis of Purchase Amount

The p-value for 'Promo_Code_Used' was determined to be 0.164, which exceeds the conventional alpha level of 0.05, suggesting that the observed association was not statistically significant.

3.1.4. Model Efficiency

The R-squared value for the regression model was 0.001, indicating that the model explains only 0.1% of the variance in purchase amounts, signifying a very low explanatory power.

3.1.5. Conclusion

The analysis suggests that there is no statistically significant evidence to support the idea that promo code usage affects purchase amounts in USD. The null hypothesis cannot be rejected based on the data and statistical methods used. While the immediate impact on transaction value

appears not significant, it is important to consider other potential benefits of promo codes, such as customer acquisition, retention, and increased transaction frequency, which may contribute to long-term business value.

3.1.6. Limitations

1. The PSM and OLS methods used have limitations and assumptions that may affect the results.
2. The model's low R-squared value suggests that other factors not included in the model may be influencing purchase amounts.

3.1.7. Further Research

Future analyses could include a broader set of variables and explore different modeling techniques to better understand the dynamics between promo code usage and customer spending patterns.

3.2. Instrumental Variable Analysis

This method provides an analysis of the influence of promo codes on purchase amounts using an instrumental variable approach to address potential endogeneity in promo code usage. The analysis tests the hypothesis that promo codes affect customer spending, employing 'Gender' as an instrument to isolate the causal impact.

3.2.1. Hypotheses

Null Hypothesis (H_0): Promo code usage has no effect on the purchase amount in USD.

Alternative Hypothesis (H_1): Promo code usage has an effect on the purchase amount in USD.

3.2.2. Methodology

An IV analysis was conducted with 'Location' initially used as an instrument, but after checking its F-statistic, it was found to be a weak instrument. Then followed by 'Gender' based on its strong correlation with promo code usage, as suggested by the literature on gender differences in shopping behavior.

3.2.3. Findings

Instrumental variable 'Location'

- First stage F-statistic: The low F-statistic of 0.55 indicates that 'Location' is a weak instrument, poorly correlated with 'Promo_Code_Used'.

Instrumental Variable 'Gender'

- Relevance: Literature supports the hypothesis that 'Gender' influences the likelihood of using a promo code.
- First stage F-statistic: A high F-statistic of 2145 indicated that 'Gender' is a strong predictor of 'Promo_Code_Used'.
- Second stage F-statistic: The second stage F-statistic of 6.4459, while below the threshold of 10, reflects a weak model in explaining variations in 'Purchase_Amount_USD'.

Impact on Purchase Amount

| IV-2SLS Estimation Summary | | | | | | |
|-------------------------------------|---------------------|------------------|---------|---------|----------|----------|
| Dep. Variable: | Purchase_Amount_USD | R-squared: | 0.0017 | | | |
| Estimator: | IV-2SLS | Adj. R-squared: | 0.0004 | | | |
| No. Observations: | 3900 | F-statistic: | 6.4459 | | | |
| Date: | Tue, Nov 28 2023 | P-value (F-stat) | 0.2652 | | | |
| Time: | 13:28:13 | Distribution: | chi2(5) | | | |
| Cov. Estimator: | robust | | | | | |
| Parameter Estimates | | | | | | |
| | Parameter | Std. Err. | T-stat | P-value | Lower CI | Upper CI |
| const | 57.314 | 2.4817 | 23.094 | 0.0000 | 52.450 | 62.178 |
| Age | -0.0155 | 0.0251 | -0.6179 | 0.5366 | -0.0646 | 0.0336 |
| Review_Rating | 0.9990 | 0.5288 | 1.8891 | 0.0589 | -0.0375 | 2.0355 |
| Previous_Purchases | 0.0142 | 0.0264 | 0.5373 | 0.5911 | -0.0375 | 0.0658 |
| Category | -0.4661 | 0.4262 | -1.0936 | 0.2741 | -1.3015 | 0.3693 |
| Promo_Code_Used | -1.1759 | 1.2772 | -0.9207 | 0.3572 | -3.6791 | 1.3274 |
| Endogenous: Promo_Code_Used | | | | | | |
| Instruments: Gender | | | | | | |
| Robust Covariance (Heteroskedastic) | | | | | | |
| Debiased: False | | | | | | |

Figure 9 - Instrumental Variables Regression Summary

- Coefficient: The coefficient for 'Promo_Code_Used' is -1.1759, suggesting an average decrease in purchase amount by \$1.18 when a promo code is used.
- Statistical significance: With a p-value of 0.3572 for 'Promo_Code_Used', the results are not statistically significant, failing to provide strong evidence against the null hypothesis.

3.2.4. Conclusion

The IV analysis suggests that there is insufficient evidence to conclude a significant impact of promo codes on purchase amounts. Both instruments, 'Location' and 'Gender', provided different levels of strength in the first stage regression, but ultimately the impact of promo codes on purchase amounts was not statistically significant.

3.2.5. Limitations

1. The choice of instruments is crucial, and the instruments used in this analysis did not provide a significant link between promo code usage and purchase amounts.
2. The low explanatory power of the models suggests other variables may be at play that are not captured in the current analysis.

3.2.6. Further Research

Additional variables and alternative instruments could be explored to better understand the impact of promo codes.

4.0 Overall Conclusion

The comprehensive analysis encompassing Descriptive Analysis, Propensity Score Matching, Ordinary Least Squares Regression, Instrumental Variable Analysis, and Segmentation and Control Analysis provides a multi-dimensional understanding of the impact of promo code usage on purchase amounts.

4.1. Addressing the Analytics Problem

Despite the initial hypothesis that promo codes might encourage higher spending, the results from Propensity Score Matching, Ordinary Least Squares regression, and Instrumental Variable analysis consistently reveal that promo codes do not significantly affect customer purchasing behavior in terms of increased spending. This comprehensive analysis challenges the conventional marketing assumption about the effectiveness of promo codes, suggesting that their role might be more nuanced, potentially centered around customer engagement and loyalty rather than directly boosting sales. This finding is crucial for businesses to reconsider and strategize the use of promo codes in their marketing and sales initiatives.

4.2. Business Insights and Implementations

1. Promo Code Strategy

Since promo codes do not significantly boost purchase amounts, their role should be reconsidered. They could be more effectively used for customer acquisition, retention, and engagement rather than as a tool to increase immediate spending.

2. Targeted Marketing

Insights from the segmentation analysis can inform more nuanced marketing strategies. Businesses could tailor promotions and communication based on the distinct preferences and behaviors of different customer segments.

3. Customer Loyalty Programs

The variability in previous purchase patterns across segments suggests an opportunity for customized loyalty programs, targeting segments with higher loyalty potential.

4.3. Potential Implementation Issues

1. Data Limitations

The analyses are limited by the data available. Further research including additional variables (like income, specific product preferences) might provide deeper insights.

2. Causality Challenges

The methods employed, particularly PSM and IV, have limitations in establishing causality. Care should be taken in interpreting these results as definitive causal relationships.

3. Operational Integration

Implementing insights from this analysis into existing business processes may require significant changes in marketing strategies and customer relationship management systems.

4. Change Management

Shifting focus from promo codes as a sales-boosting tool to a customer engagement tool will require effective communication and change management within the organization.

In conclusion, while promo codes do not significantly impact average purchase amounts, they hold potential in other areas of customer engagement and marketing. Businesses should leverage these insights to refine their marketing strategies and customer relationship management practices, considering the operational and data challenges identified.

Reference

Banerjee, S. (2023, October 5). *Customer Shopping Trends Dataset*. Kaggle.

<https://www.kaggle.com/datasets/iamsouravbanerjee/customer-shopping-trends-dataset/data>