STAT 112- Introduction to Data Processing and Visualization Project

# How Promotion Codes Shape Customer Choices

# By

Cihangir Osman Dokucu

10 December 2023

**1. ABSTRACT:**

In the last 30 years customer’s online shopping behaviors have changed significantly fast and are still changing. Companies use various methods to bind possible customers. For example, they make subscription systems or adjust different discount and promotion codes. In this project, we will be analyzing a dataset that includes customers gender, ID, shipping type they use, payment method, frequency of purchases and purchase amounts, along with promotion code usage information. We aim to use this data to understand whether promotion code effect the customers positive or negative way. Also, relationship with age groups, genders, item types and seasonal changes. We use descriptive statistics and exploratory data analysis techniques to analyze data and from the analysis, the following were found:

* Customers who use promotion codes tend to purchase a greater quantity of products.
* Footwear is the category that has the highest amount of purchase spending among customers who use promotion codes.
* The age group that makes the most purchases among men using promotional codes is 55-64.
* The highest purchase numbers using a promotional code are made in Pennsylvania and lowest is Connecticut.
* Customers who use promotional codes spend the highest amount of money in the fall season.
* Among customers who used promotional codes, they spent the most money on t-shirts.

2. INTRODUCTION

This analysis mainly examined the promotion code usage in US states. The dataset has 14 different variables about customers shopping behavior:

* Category: Dataset contains 4 different clothing categories.
* Color: Dataset has 25 different color people bought.
* Customer ID: Customers’ personal ID.
* Discount Applied: Discount applied or not as “yes” and “no” in dataset.
* Frequency of Purchases: 7 different frequency type
* Gender: Gender of customers
* Item purchased: Dataset contains 25 different item type.
* Location: Location of the state in US
* Payment Method: Payment method customers’ use
* Promo Code Used: Promotion code used or not
* Season: Season in which the product was purchased
* Shipping Type: Shipping type of the product
* Size: Size of the products
* Subscription Status: Customer’s subscription status.

2.1 Research Questions:

*-* How do promotions (DiscountApplied and PromoCodeUsed variables) affect the total purchase amount and frequency of purchases among various consumer segments?

*-* What is the impact of using a promotional code on the product category?

*-* How do different age groups and genders react to promotions, and how does this affect their purchasing behavior?

*-* Are there regional variations in the utilization of promotions, and how do these disparities relate to overall purchase patterns and amounts?

*-* How does the impact of promotions vary seasonally, and are there specific seasons during which promotions have a more significant influence on consumer behavior?

*-* Do customers' purchasing frequency and promotional code usage have an impact on their purchase amounts?

*-* Does the promo code usage affect the customers’ decision about type of the product they buy?

3. DATA CLEANING

The dataset in this project demonstrates an impressive level of cleanliness and organization, which means that minimal intervention is needed during the data cleaning process. Consequently, the data cleaning process mainly consists of regular checks for outliers and ensuring data integrity. This efficient dataset not only speeds up the analysis but also showcases the meticulous attention to detail that was applied in creating the dataset, ensuring its reliability and accuracy for future stages of the project.

4. EXPLORATORY DATA ANALYSIS

Question 1

**Effectiveness of Promotions Code on Purchase Number:**

How do promotions affect the total purchase amount and frequency of purchases among various consumer segments?

To analyze this question, we should use box plot since we have one categorical data as promotion code used or not, and one numerical data as average previous purchase number.

In the graph, each dot represents US states. Thus, we can conclude that customers who use promotional codes purchase more items before that. We can suggest to companies give promotional codes to increase their sales numbers. However, they should be careful about the amount of promotional code because if they use too much, they may be negatively affected financially. Looking at the box plots, it is seen that graph of customers who do not use promotion codes might be little right skewed with the median around ≈ 25, and graph of customers who do not use promotion codes look like bell-shaped with the median average previous purchase of ≈ 26.

A graph with blue squares and dots

Description automatically generated

Question 2

**Promotion Code Usage Effect on Categories**

What is the impact of using a promotional code on the product category?

To analyze this question, we use lollipop chart since we have one categorical data as item category, and we have one numerical data as average purchase amount in USD. As we can see in the graph below, customers who apply promotion codes spend the highest amount of money on footwear products. Two categories in the middle follows footwear closely and outerwear followed far behind them. Thus, companies who produce outerwear should consider changing their prices or arrange the amount of their codes. Outerwear products usually have higher prices than other categories so there is a problem in the sales department. Thus, we suggest that companies who are involved in this research should make significant position changes in order to earn more money.

A graph with blue dots

Description automatically generated

Question 3

**Demographic Variation in Promotion Response**

How do different age groups and genders react to promotions, and how does this affect their purchasing behaviour?

We used stacked bar charts to analyze this research question. In the dataset, we only have age variable that contain all age groups, so we use calculated fields to visualize different age groups better by using tableau. Unfortunately, we do not have female number who use promotion codes, so we interpret without that number information. Surprisingly, the age group that uses the most promotional codes is 55-64 since it’s not expected that elder people use online shopping this amount. The largest age group that does not use a promotional code is 45-54. Leaving aside whether a promotional code was used or not, it is surprising that the second minimum shopping age group is 18. We suggest to companies used in this dataset that they should increase advertisements aimed at younger audiences because the future of companies comes from sales targeting the younger generations.

A graph of a bar graph

Description automatically generated with medium confidence

Question 4

**Geographical Disparities in Promotion Utilization:**

Are there regional differences in the use of promotions and how do these differences relate to overall purchasing patterns and quantities?

A graph of a number of people

Description automatically generated with medium confidenceAs we can see in the graph above the highest average spenders of people using promotional codes are in Pennsylvania. Moreover, 6 of the top 10 states are on the east coast. Therefore, Companies looking to increase sales in West Coast states should consider reorganizing their advertisements specifically for the West Coast market. Furthermore, the lowest average spenders of people using promotional codes are in Connecticut.

A screen shot of a computer

Description automatically generated

A graph of a number of people

Description automatically generated with medium confidence

As we can see in the graph above, the highest average spenders of people not using promotional codes are in Alaska. The average sales amount in Alaska is very high, but they can make more profit if they increase it with a simple edit in their promotional codes. We can observe that values in the middle part of the graph are very close and uniformly shaped. Furthermore, the lowest average spenders of people who not using promotional codes are in Delaware.

Question 5

**Seasonal Influence on Promotions:**

How does the impact of promotions vary seasonally, and are there specific seasons during which promotions have a more significant influence on consumer behavior?

A graph of blue and white bars

Description automatically generated

We use clustered bar chart in these variables since it suits very well. We can observe that average purchase amount is high in fall and winter which makes sense because items used in fall and winter cost more, so price is higher. For example, jackets and hoodies. Average purchase amount (apa) is around 62 in fall in both situation of promotion code. Apa is around 60 in winter in both circumstances too. However, spring and summer have some differences in both cases. To conclude, it looks like a promotion code is not very effective on average purchase amount. The only differences are in spring and summer, but the difference is still very low. We suggest to the companies that they should increase their items’ price’s in order to increase their profit.

Question 6

**Promo code usage effect on Type of Item Purchased:**

Does the promo code usage affect the customers’ decision about the type of the product they buy?

A screen shot of a chart

Description automatically generated

We used a highlight table to visualize this data. As depicted in the graph, individuals utilizing promotional codes have allocated the highest average spend toward T-shirt purchases and lowest spend toward dress. Surprisingly, in the situation promotion code used, lowest average money spends on jacket jewelry and jeans because they are usually more expensive than t-shirts and shorts. We can infer from this situation that the quality of the sold jewelry, jackets, and pants is below average. However, we cannot draw any conclusions about the profit generated by the companies, as we lack data on the number of sales. On the contrary, in cases where promotional codes are not used, the distribution of products and the expected price appears quite normal, like the highest one is dress and lowest one is sandals.

Question 7

**Purchase Frequency Impact on Purchase Amount:**

Do customers' purchasing frequency and promotional code usage have an impact on their purchase amounts?

In order to answer this question we used clustered bar chart since we have one categorical and numerical data. As we can see in the graph, in both scenarios, the annual average spending amount is the highest. Therefore, it can be inferred that companies would benefit from focusing on subscription systems that can keep their members engaged for as long as possible.

A blue and white bar chart

Description automatically generated

4. CONCLUSION

In conclusion, throughout our project on the interpretations and questionings on the tidied and cleaned data, we analyze and understand that how promotin codes effects customer shopping behaviour. We did a research about age groups, genders, geographical dispersion, season and various items. As a final result, we can see that promotion codes effect the companies profits in a positive way.

6. LINKS

Link to our dashboard:

<https://public.tableau.com/app/profile/cihangir.osman.dokucu/viz/CODassignment/Dashboard?publish=yes>