Marketing Analytics: MKT461-A Professor Harris

Consumer Behavior and Shopping Habits Analysis

Holly Masciarelli, Brooke Blanche, Kylie Donovan, Nicole North

Contents

[**Dataset** 2](#_Toc166187227)

[Introduction 2](#_Toc166187228)

[**Machine Learning (ML)** 3](#_Toc166187229)

[Models 3](#_Toc166187230)

[How ML Shapes Our Data 3](#_Toc166187231)

[Manipulating Our Data 4](#_Toc166187232)

[**Exploratory Data Analysis (EDA)** 5](#_Toc166187233)

[Dashboards 5](#_Toc166187234)

[Insights 6](#_Toc166187235)

[**Conclusion** 7](#_Toc166187236)

[**Appendix** 7](#_Toc166187237)

[Exhibit 1 7](#_Toc166187238)

[Exhibit 2 8](#_Toc166187239)

[Exhibit 3 8](#_Toc166187240)

[Exhibit 4 9](#_Toc166187241)

[Exhibit 5 10](#_Toc166187242)

[Exhibit 6 11](#_Toc166187243)

## **Dataset**

### Introduction

Throughout our project, we analyzed data on consumer behavior consisting of various categories including customer demographics, shopping choices, product characteristics, among others. The usage of Google Collab and Tableau provided us deeper insights into consumer trends, thus allowing us to facilitate future projections. We were faced with challenges which resulted in synthetic data leading to inconclusive graphs. Nonetheless, we can conclude that analyzing data through coding in Google Collab allows for comprehensive data processing, while creating the graphs in Tableau enhances visualization and presentation, offering a combination of insightful data exploration and decision-making.

When generating our machine learning models, we quickly discovered that our dataset was synthetic. This became apparent when attempting to extract insights about average customer spending, where both men and women coincidentally shared an average of $59.55. Similar patterns emerged across various visualizations, including age demographics, previous purchase behavior, types of clothing recently bought, and seasonal purchase trends. Due to the synthetic nature of our data, genuine insights remained elusive. Faced with this challenge, we adopted a creative strategy and decided to pivot. Leveraging Google Collab, we manipulated the dataset to showcase insights we deemed logical. We used Linear Regression to illustrate the dataset's synthetic nature and the difficulty in drawing meaningful conclusions. As depicted in Exhibit 1, the data points scattered aimlessly, reflecting the absence of a discernible trend. Subsequently, we used A/B testing for our second machine learning model, focusing on exploring the relationship between gender and purchase amounts. Here, we meticulously manipulated the dataset to align with our intuition that women typically spend more money on clothing. Exhibit 2 displays our manipulation efforts, where women have a noticeably higher average spending compared to men. As a team, we recognized the novelty of our approach to the final project. Despite working with synthetic data, we demonstrated our adeptness in machine learning by creating and manipulating models to convey insights.

## **Machine Learning (ML)**

### Models

Machine Learning (ML) models are ways to find patterns or make decisions from a set of data. These algorithms are the core component of machine learning and help to detect relationships between variables in a dataset. By analyzing lots of data, machine learning can uncover hidden patterns, trends, and insights that humans may overlook. They come in various forms to best fit a data type. From supervised learning models, which learn from labeled examples to unsupervised models that find patterns in unlabeled data, machine learning has changed the way that data is used and manipulated.

### How ML Shapes Our Data

The first ML model we used was linear regression. Linear regression is a supervised learning technique used to model the relationship between a dependent variable and one or more independent variables. It seems to find the best-fitting linear model that describes the relationship between variables. This shows how changes in the independent variables are associated with changes in the dependent variable. It is especially useful in understanding and predicting continuous outcomes like revenue or sales. In our dataset, we used linear regression to show the relationship between Age and Purchase Amount (USD) or Predicted Purchase Amount. Because our data is synthetic, we could not draw insights from the random clustering that was created with the linear regression code from Google Collab (See Exhibit 1).

Next, we analyzed our data using A/B testing. A/B testing involves comparing two or more versions of a product or two or more variables in the case of our dataset, to determine which one performs better in terms of a predefined metric. With our dataset, we were looking to determine whether Gender affects Purchase Amount (UDS) to get an estimate on consumer behavior and shopping habits. Ideally these conclusions would be used as a metric for customer engagement and functionality. Because our data is synthetic, however, there were few insights we could draw from the A/B testing model we conducted before manipulating the dataset (See Exhibit 2). By collecting and analyzing data on user behavior and outcomes, A/B testing allows organizations to make data-driven decisions about which variable is performing best.

### Manipulating Our Data

For A/B testing, we manipulated our dataset to extract meaningful insights. Initially, we exported the dataset to Excel and computed the average spending amounts for both genders. Surprisingly, we discovered a shared mean of $59.55, which prompted us to manually adjust the average spending per female to ensure a higher average. However, despite our efforts, it did not yield a significant deviation in the mean. This realization led us to reconsider our approach.

Subsequently, we opted for a more effective strategy by manipulating the dataset directly within Google Collab while scripting the A/B testing code. As shown in Exhibit 3, we deliberately set the average spending for males to $60 and females to $150. We believed this adjustment intuitively reflected real-world scenarios and introduced a substantial disparity, facilitating the extraction of actionable insights.

## **Exploratory Data Analysis (EDA)**

### Dashboards

One way we displayed our data was using Kaggle to create EDA graphs, or Exploratory Data Analysis. The purpose of this was to discover general patterns in our data. We created three dashboards to showcase different data insights. In particular we found gender to be interesting, so we focused two dashboards on gender data. In our first dashboard we looked at promo code use, categories of item purchased and finally season versus purchase amount. We found that more people did not use promo codes at 57.35% being no (Appendix 4). The item we found to be purchased the most was clothing and the least purchased was outerwear. Lastly, as far as season and purchase amount we found that Winter was the least popular season to make purchases. This could possibly mean that the store had better summer clothing as outerwear was not a popular item.

In our next dashboard focused on gender we looked at gender vs purchase amount, gender vs subscription status, and gender vs frequency of purchase. Something surprising in the data was that we found that males had a higher purchase amount at $157,890 vs females at $75,191. This was a surprise because the stereotype is that females shop more than males. Looking deeper this could mean that the store had better men’s clothing than women’s. Additionally, males also have higher subscription rate than females supporting that the brand may have better items for male. As far as frequency of purchase it was also shown that males purchase much more frequently than females which would make sense because they also had a higher purchase amount.

For our second gender dashboard, we looked at gender vs average previous purchases, gender vs review rating, and gender vs promo code used. For average previous purchase males had 25 while females had 24 meaning that most people despite gender had a similar number of previous purchases (Appendix 6). For gender and review rating, they were the same showing that gender did not affect the rating they gave. Lastly for gender and use of promo code, it showed that males used promo codes far more often than females. The data showed that males used promo codes almost two times the amount females did. This could have been due to more promo codes available to men or since men purchase more, they are more prone to using promo codes.

### Insights

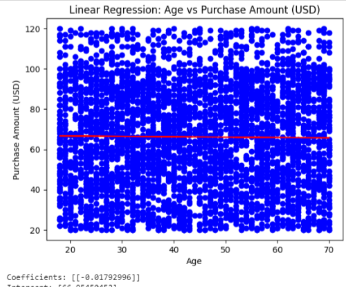
Dashboarding was valuable to visualize our data and gather insights from the visualizations. For instance, our exploration into gender-specific purchasing habits revealed that contrary to stereotypes, males exhibited a higher average purchase amount.   
Additionally, our manipulation of the dataset enabled us to simulate realistic scenarios, thus displaying consumer behaviors in different scenarios. Our exploratory data analysis revealed a correlation between promotional code usage and gender, offering additional insights for targeted marketing strategies. These findings showcase the importance of data analysis in uncovering hidden patterns.

## **Conclusion**

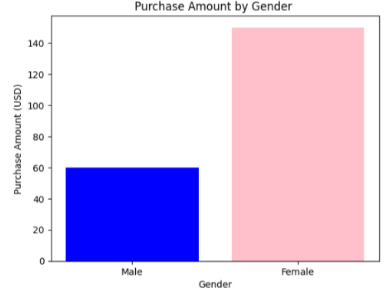
The integration of coding with learning techniques, A/B testing, and exploratory data analysis (EDA) allowed us to decipher trends within the data, thus enabling us to predict future consumer behaviors. Despite facing challenges with synthetic data, we showcased insights from linear regression and A/B testing. Manipulating the dataset allowed us to simulate realistic scenarios and extract meaningful conclusions. Gender-specific purchasing patterns were discovered from our various dashboards. Our project highlights the importance of combining coding tools and visualization techniques for comprehensive data exploration and decision-making in understanding consumer behavior.

## **Appendix**

### Exhibit 1



### Exhibit 2



### Exhibit 3

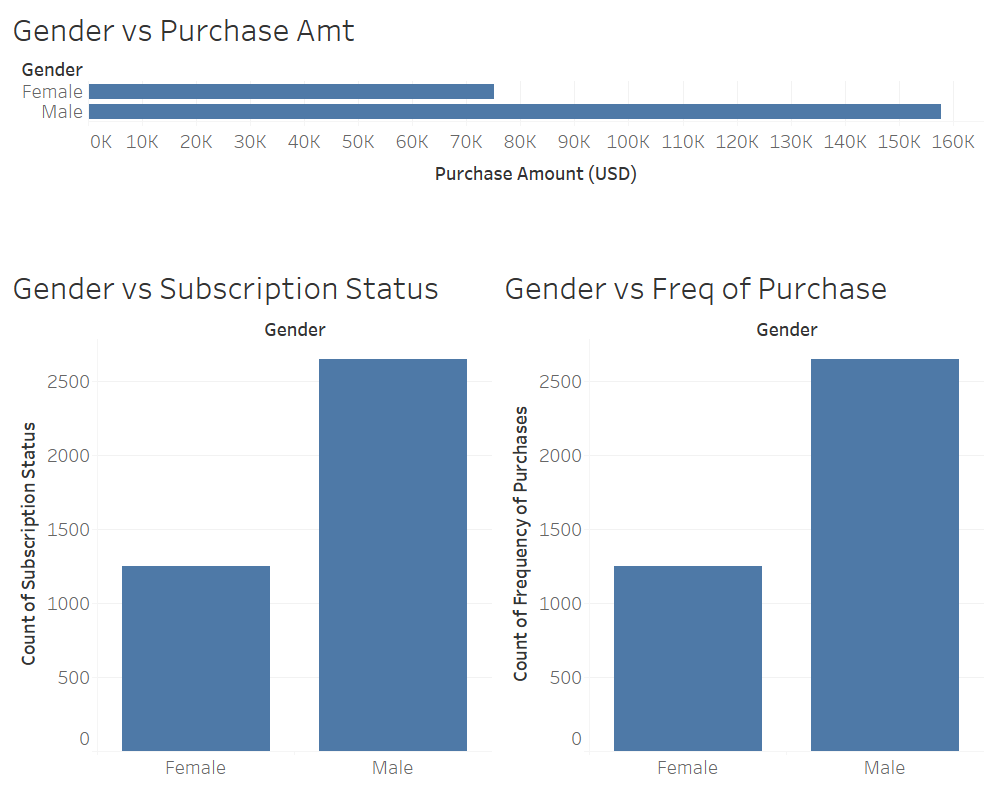
A computer code with black and red text

Description automatically generated

### Exhibit 4



### Exhibit 5



### Exhibit 6

