

E-COMMERCE BUSINESS ANALYTICS REPORT

Data Dictionary

Variable	Description
CustomerID	Unique identifier for each customer.
Age	Age of the customer, expressed in years.
Gender	Gender of the customer, expressed as either “Male” or “Female.”
Income	Annual nominal income of each customer, expressed in U.S. dollars (USD).
TotalSpent	Total amount spent by the customer, expressed in USD.
AvgOrderValue	Average value of each order placed by the customer, expressed in USD.
NumOrders	The total number of orders placed by the customer throughout their entire relationship.
CartAbandonmentRate	Proportion of total shopping carts that are abandoned by the customer, represented as a decimal value between 0 and 1.
TimeOnSite	The total duration of time, expressed in minutes, the customer has spent on the e-commerce website.
Clicks	The total number of clicks made by the customer on the e-commerce website.
ItemsViewed	The total number of items viewed by the customer on the e-commerce website.

Introduction

The e-commerce data set contains information on 500 unique customers. The data contains 11 attributes, which are composed of 1 identifier (CustomerID), 1 categorical (Gender), and 9 others. These quantitative measures contain demographic characteristics, such as the customer ages (Age) and their associated nominal income levels (Income). Purchasing features for these customers include their totals spent with the e-commerce platform (TotalSpent), their average order values (AvgOrderValues), the total number of orders they made (NumOrders), and their cart abandonment frequency (CartAbandonmentRate). Site behavior qualities of these same customers reflect their time spent on the website (TimeOnSite), the number of clicks they made (Clicks), and the number of items they viewed.

This report focuses mostly on the demographics characteristics of the customers in addition to their purchasing characteristics. Descriptive statistics and relationships between these variables are explored, revealing great insights into how income impacts key purchasing performance indicators. Age, income levels, gender, total spending, average spending, and order frequency per customer are explored.

Demographic Characteristics

The demographic characteristics analyzed within this report on e-commerce platform data are customer ages and income levels. The distribution of customer ages is shown, via a histogram, in Figure 1 below, which indicates somewhat of a “rough” bell curve shape. The central measure for customer age is around 40 years, where a large portion of the data is contained. Following are age ranges of between 45-50, 15-20, 30-35, and 55-60 years of age, in order of frequency.

Figure 2 below shows how the average age of customers in the data differ by gender. Female customers have a higher age compared to males; however, this difference is small and negligible. Although, it is also important to note that, for both genders, their average customer ages both center around 40 years of age, which is consistent with our original distribution of ages across both genders. The insight that can be gained here is that gender may not have an impact on customer ages. Nevertheless, the average age of a customer is around 40 years old, with other age groups described above being other potential customer bases for targeted marketing campaigns.

Male average incomes, as shown in Figure 3 below, are higher than women; however, this difference in mean income is negligible and doesn't suggest major differences in these income levels observed. For both genders, the average income level is above \$60,000, indicating many wealthier customers contained within the data. Figure 4 further shows the distribution of the customer incomes, illustrating no meaningful variations in the median income expressed and the spread of the data. Median income for both genders lie between \$60,000 to \$70,000. The insight here is that gender doesn't have an impact on income and that there are a majority of

customers that possess higher income, which can lead to higher potential spending. It would be optimal to provide offerings for both genders in these higher income areas.

Purchasing Behaviors

Figure 5 breaks down the distribution of totals spent per customer by gender. Male customers exert a higher median total spent with the e-commerce platform as well as a wider spread compared to women. Male customers have a median total amount spent with the e-commerce site that is closer to \$10,000 while female customers have spent between \$7,500 to \$10,000 overall. Women have many outliers above \$25,000 while men have fewer outliers above \$30,000. Removal of these extreme values would imply that women would have spent much lower overall as opposed to their male counterparts.

The distribution of average order values is influenced by extreme values, as indicated by its right-skewed shape, shown in Figure 6 below. Customers have demonstrated a high frequency of spending around \$500 per order, on average, indicating low spending behaviors per order. When broken down by gender, shown in Figure 7, women have a higher median as opposed to men in addition to greater spread. Women's outliers are mostly concentrated between \$4,500 and \$5,000 per order while men's outliers are spread across values between \$3,500 and \$5,000 per order. Higher spread and median values for women suggests that, although they have lower overall spending with the site, they spend more on average compared to male counterparts. The insight here is that women should be targeted with more expensive products and men should be targeted with more regular everyday items that encourage them to spend more overall.

Lastly, the distribution of the total number of orders placed overall with the e-commerce platform closely resembles a normal, bell-shaped curve. Yet, the highest frequency occurs for customers who have placed three orders throughout the duration of their relationship, while the central measure is four orders overall. In addition, the second highest frequent number of orders placed is five, followed by four. This tells the story that a majority of customers place between 3-5 orders throughout their relationship with the company, showing some degree of loyalty to the business. These customers should be targeted for promotional offers to increase the number of orders they place in the future.

Relationships

Figure 9 shows the correlation matrix for all quantitative variables analyzed within the e-commerce data. From this heatmap, there are only three potential relationships that can be derived from the plot (only the first two are important in this analysis), which includes variables such as the customer's income level, the total amount they have spent, as well as their average order values. These three quantitative variables narrate (a) how one economic demographic variable relates to two other indicators of purchasing behavior and (b) how those two purchasing variables coincide with one another. The following three potential relationships are listed below:

E-COMMERCE BUSINESS ANALYTICS REPORT

- (1) *The impact of customer income on their total amounts spent* (“Income” vs. “Total Spent”),
- (2) *the impact of customer income on their average order values* (“Income” vs. “AvgOrderValue”), and
- (3) *the impact of the customer’s total amount spent on their average order values* (“TotalSpent” vs. “AvgOrderValue”).

The first relationship that is meaningful is the one that is between the customer’s income level and the total that they have spent with the e-commerce company, demonstrated by Figure 10 below. The Pearson correlation coefficient is 0.68, according to Figure 9, which indicates a moderate-to-strong, positive, linear relationship between these two quantitative variables. Therefore, as the income of a customer increases, so does their total spending with the e-commerce platform. The business insight that can be derived from this relationship is that higher income customers are more likely to spend more overall. Therefore, targeting these groups with higher income levels can be meaningful in driving profits for the business.

The second relationship, shown in Figure 11, relates the customer’s income level with their average order value. Its Pearson correlation coefficient is much lower than the previous, at 0.48, also according to Figure 9. This tells the story that there exists a moderate, positive, linear relationship between these two variables. This indicates that there is, with some uncertainty, a factor of income in the customer’s average order value; although, this relationship is not perfect and there exists data that supports that low income customers have spent more on average as well as those in the middle and high income ranges. The business insight here is that average order values may not be most impacted by customer’s income and other factors may be necessary to explore.

Business Decision Making Recommendations

The first business decision making recommendation is to create premium offerings and exclusive bundles that target higher-income customers. The moderate-to-strong relationship between a customer’s income level and the total amount they spend suggests that as more people earn higher incomes, their propensity to spend increases. In turn, these income increases create elevated demand for their products, leading to higher pricing, revenues, and profits in the long run.

Next, increase marketing spending for women who possess higher average order values compared to men. These customers are more inclined to spend more per transaction as opposed to their male counterparts, on average. This contrasts the idea that they spend less overall, as they may have desires for higher priced, value-added products. Innovative products that have more elegant product features can drive in more revenues from this customer base as the demand rises when more women attain higher income levels.

The third suggestion is to utilize reward programs through repeat purchases, which allows the company to take great advantage of customer loyalty. The distribution of the total

number of orders placed with the e-commerce company shows that many customers make between 3-5 orders throughout their relationship. Incentivizing these customers via a rewards program, can encourage customers to make more frequent purchases, leading to increased revenue overall, as well as higher median and average order frequencies.

The next consideration the company should take is to differentiate product offerings by gender, as there are some purchasing behavioral differences observed. Women, as mentioned above, would respond strongly to higher-priced, value-added products. However, men have lower overall average spending patterns, but have higher total spending with the site. This can imply that men make more frequent purchases of everyday products rather than premium offerings. Creating unique bundles that are more affordable to male customers can generate steady revenue streams in addition to the higher streams by women.

Finally, the company should explore additional trends, attributes, and relationships that exist beyond this data. For instance, the company can gather data on the product categories, customer lifestyle preferences, and location characteristics. While the relationship between income and total spending is relatively strong compared to the impact of income on average purchasing behavior, there can also be additional relationships that are to be uncovered. An example would consist of understanding how location characteristics influence what categories of products are purchased and can give better insights into what women and men should be offered based on where they live. Uncovering these new relationships with additional attributes can also help control for random effects in the relationship between income and total spending, leading to precise estimations of future demand, which can help tailor offerings that will drive company revenues into the future.

Figures

Figure 1 – Histogram showing the distribution of customer ages.

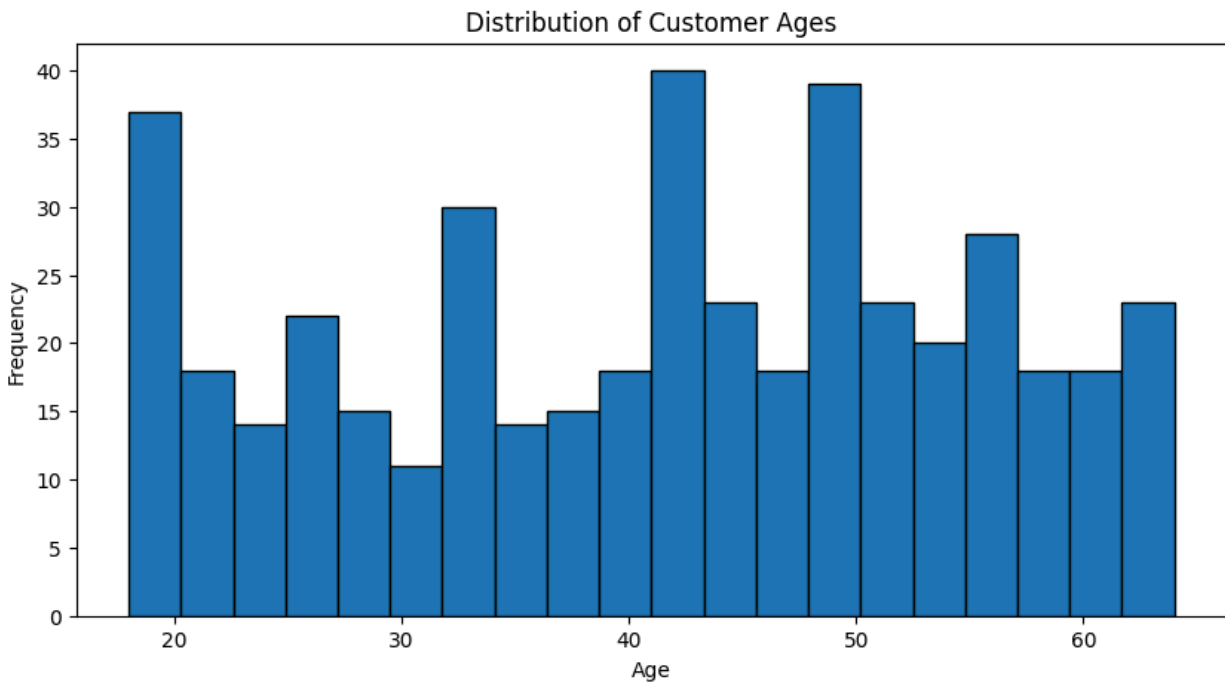


Figure 2 – Bar chart showing how average ages differ by gender.

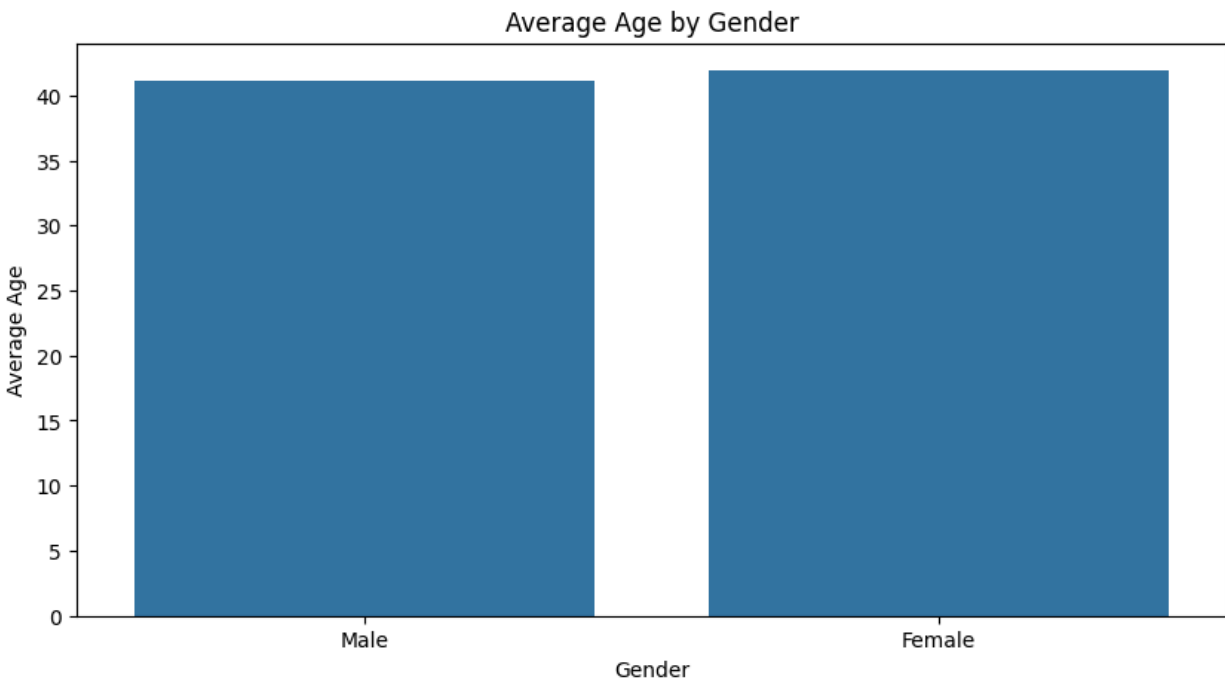


Figure 3 – Bar plot showing how average incomes differ by gender.

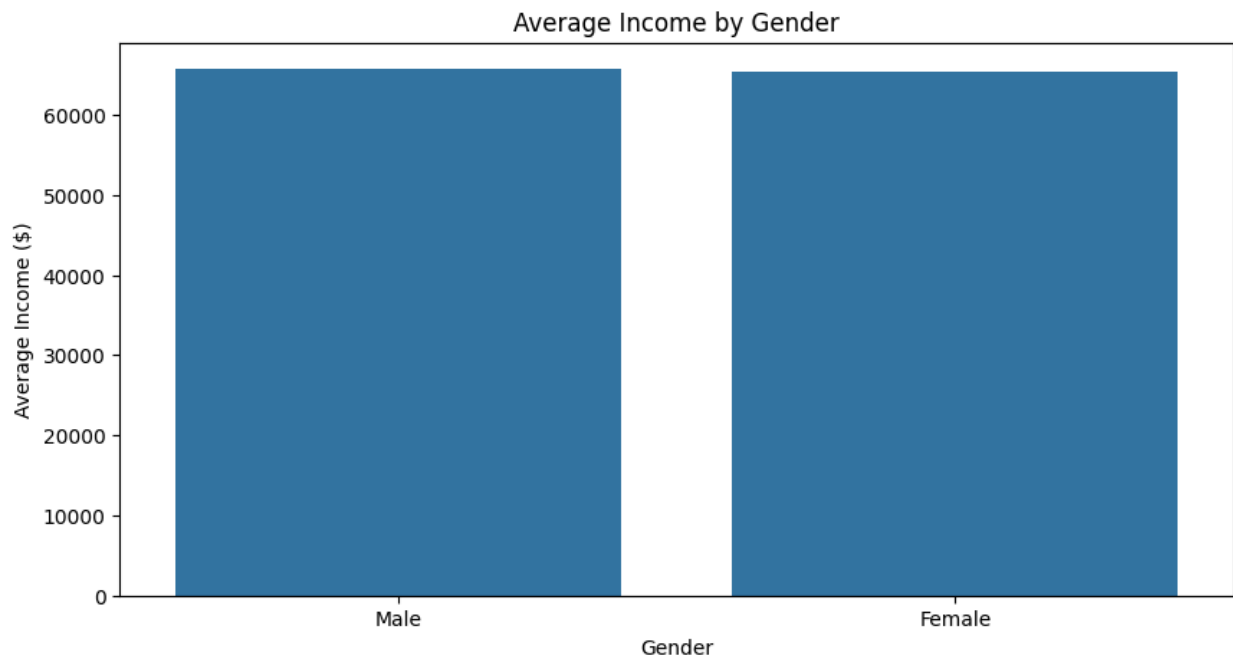


Figure 4 – Box plot showing how the distribution of incomes differ by gender.

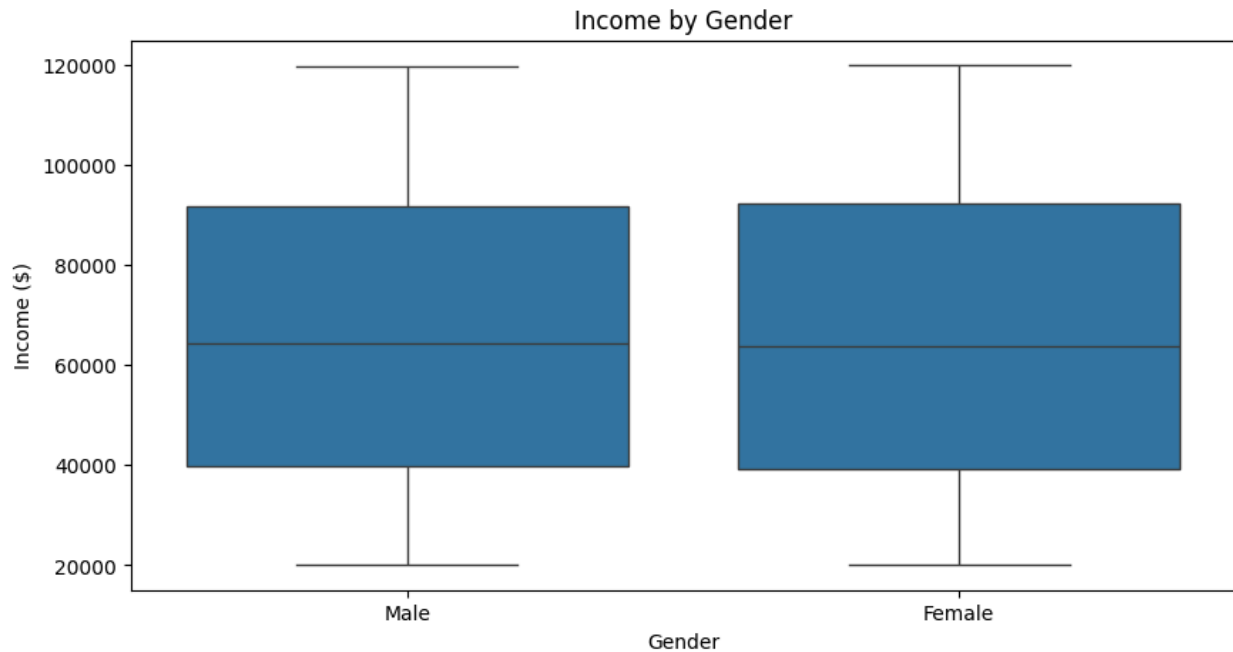


Figure 5 – *Box plots showing how the distribution of totals spent per customer differ by gender.*

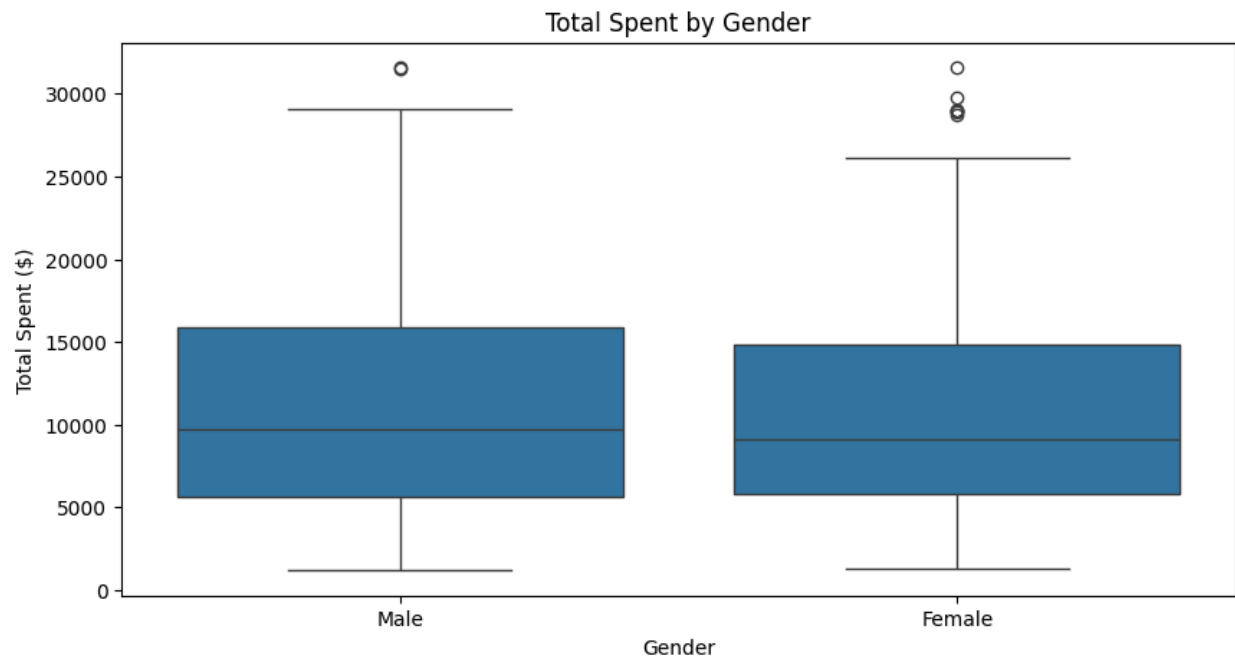


Figure 6 – *Histogram showing the distribution of customer average order values.*

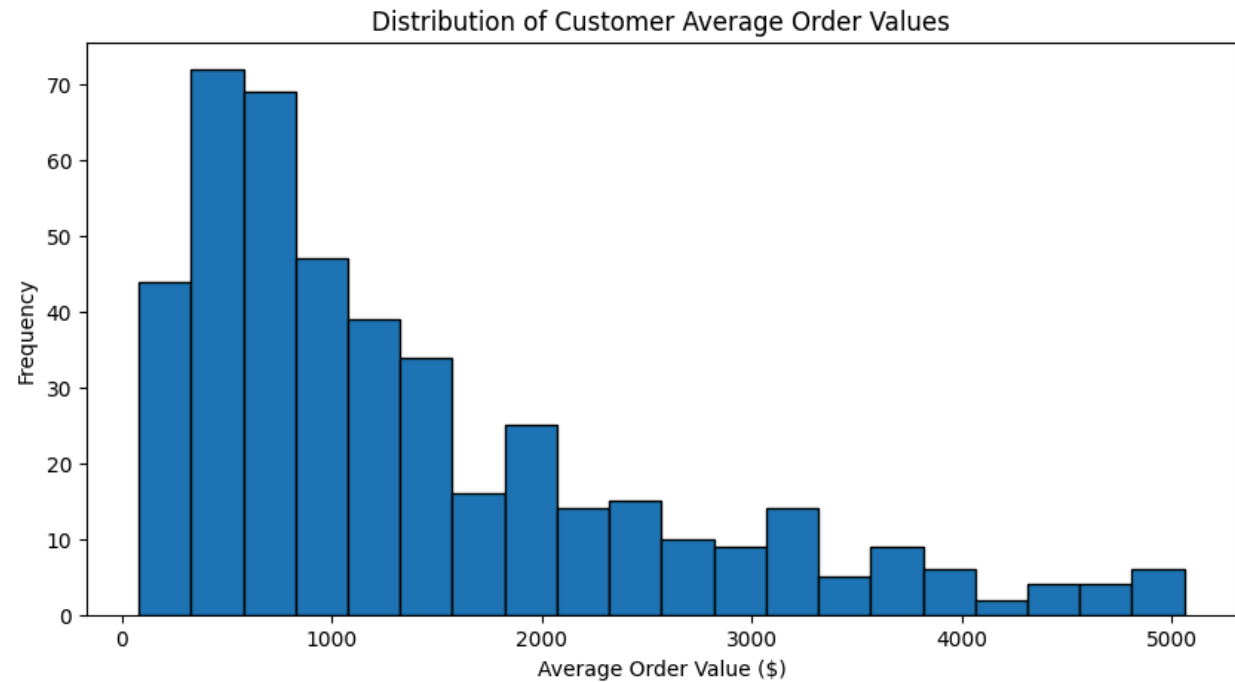


Figure 7 – *Box plots showing how the distributions of customer average order values differ by gender.*

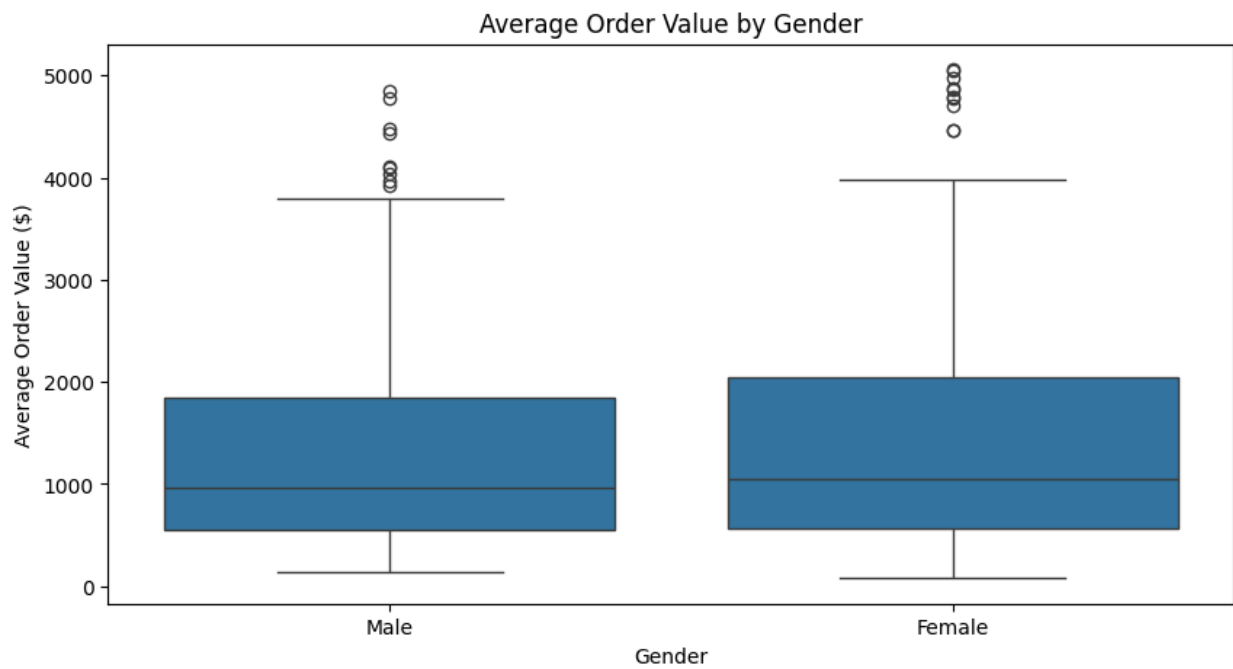


Figure 8 – *Histogram showing the distribution of the number of orders placed per customer.*



E-COMMERCE BUSINESS ANALYTICS REPORT

Figure 9 – Correlation matrix showing the relationships between quantitative variables in the e-commerce data set.

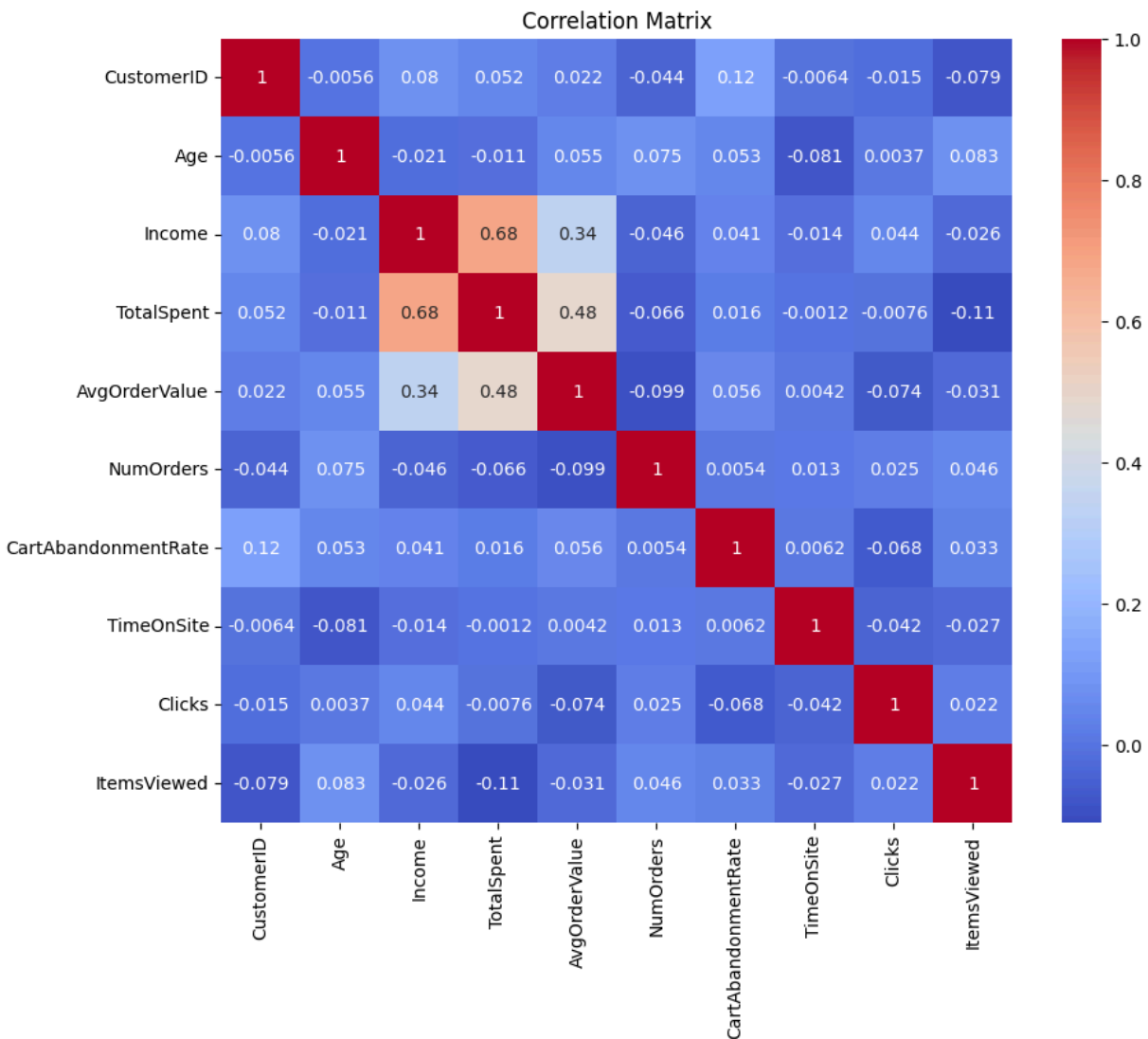


Figure 10 – *Scatter plot showing the positive relationship between customer incomes and their totals spent.*



Figure 11 – *Scatter plot showing the positive relationship between customer income and their associated average order values.*

