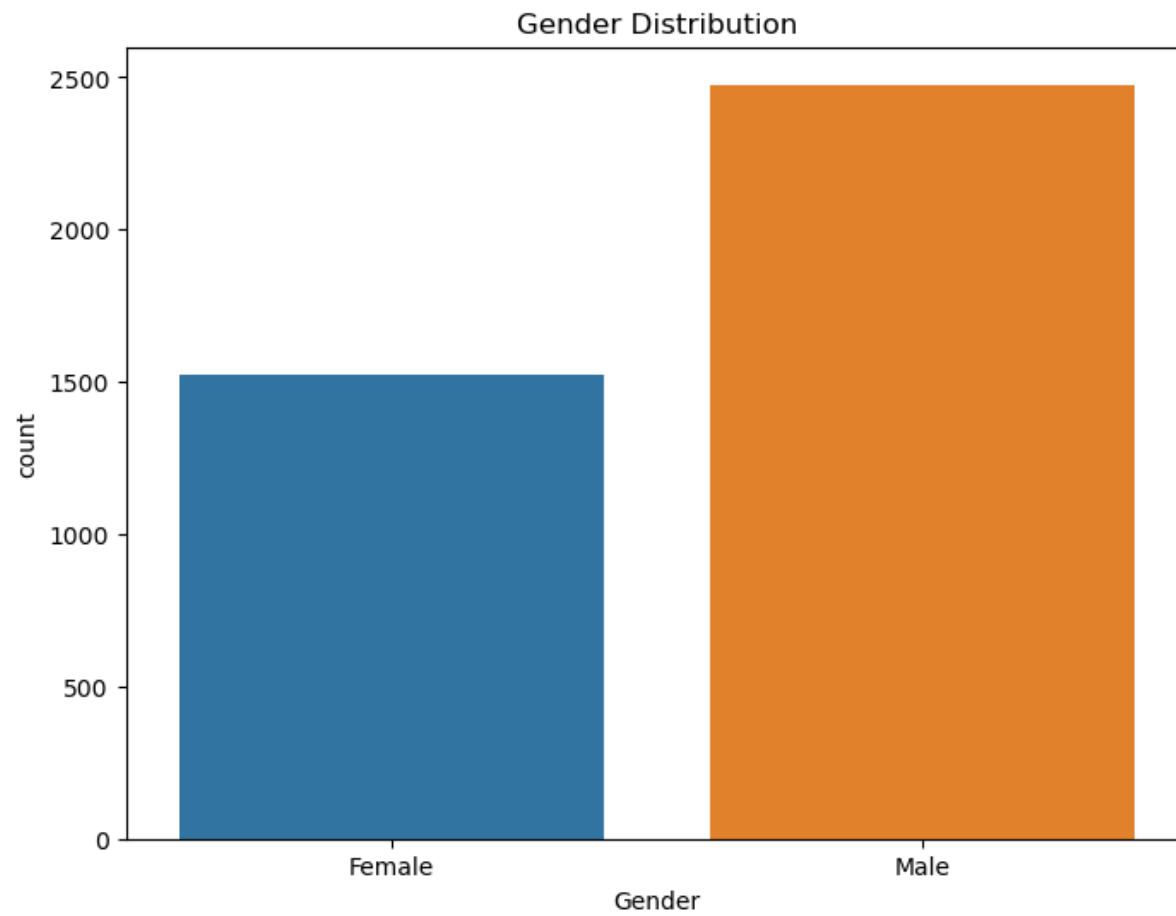


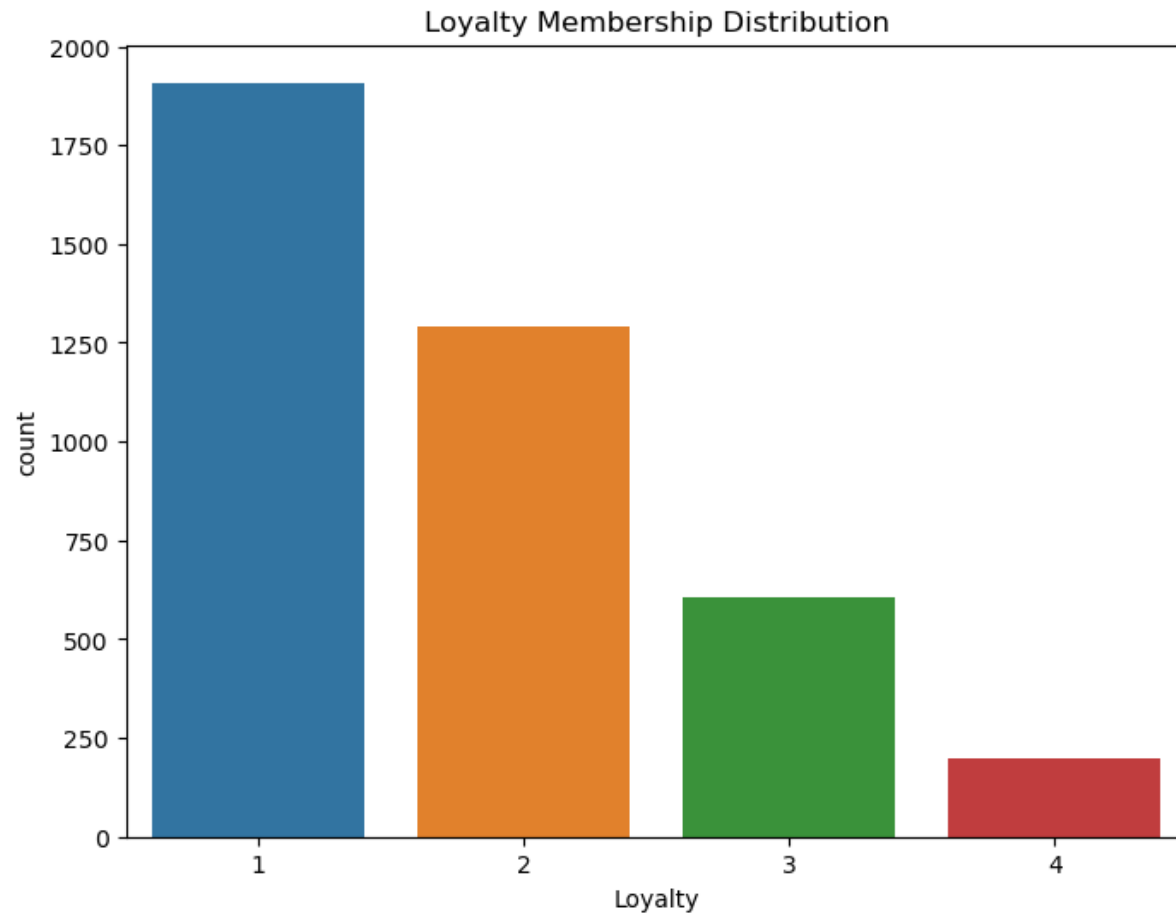
Executive Summary

For missing values, the output indicates that there are no missing values in any of the columns, as all counts are zero.

Next, the data types of each column are displayed. Most columns are of integer type, except for the 'Gender' column, which is of type 'object' (likely indicating string or mixed types).



As per the above graph, a male contribution is higher than the Female.



This graph represents the Loyalty Membership distribution where Loyalty 1 is higher and 4 is lower.

Amount Spent (Spending) Analysis:

1. **Mean Spending:** Customers in the Post = 1 group tend to have a slightly higher mean spending (\$4862.32) compared to those in the Post = 0 group (\$4232.17).
2. **Spread of Data:** The standard deviation for both groups indicate the variability in spending. Post = 1 has a slightly higher standard deviation, suggesting more variability in spending among customers in this group.
3. **Minimum and Maximum Spending:** The minimum and maximum spending values vary between the two groups. Customers in the Post = 1 group have a higher minimum and maximum spending compared to the Post = 0 group.
4. **Percentiles:** The percentiles provide insights into the distribution of spending. Median values (50th percentile) are relatively close for both groups, indicating similar central tendencies. However, there are differences in the lower (25th percentile) and upper (75th percentile) quartiles, suggesting variations in spending patterns.

In summary, the analysis reveals differences in the spending behavior of customers between the two groups (Post = 0 and Post = 1), with variations in mean spending, spread, and percentiles. Further investigation and segmentation may be required to understand the factors influencing these differences.

Multivariate Analysis: Preliminary Understanding of the Relationship between Customers' Characteristics and Purchase Behaviours

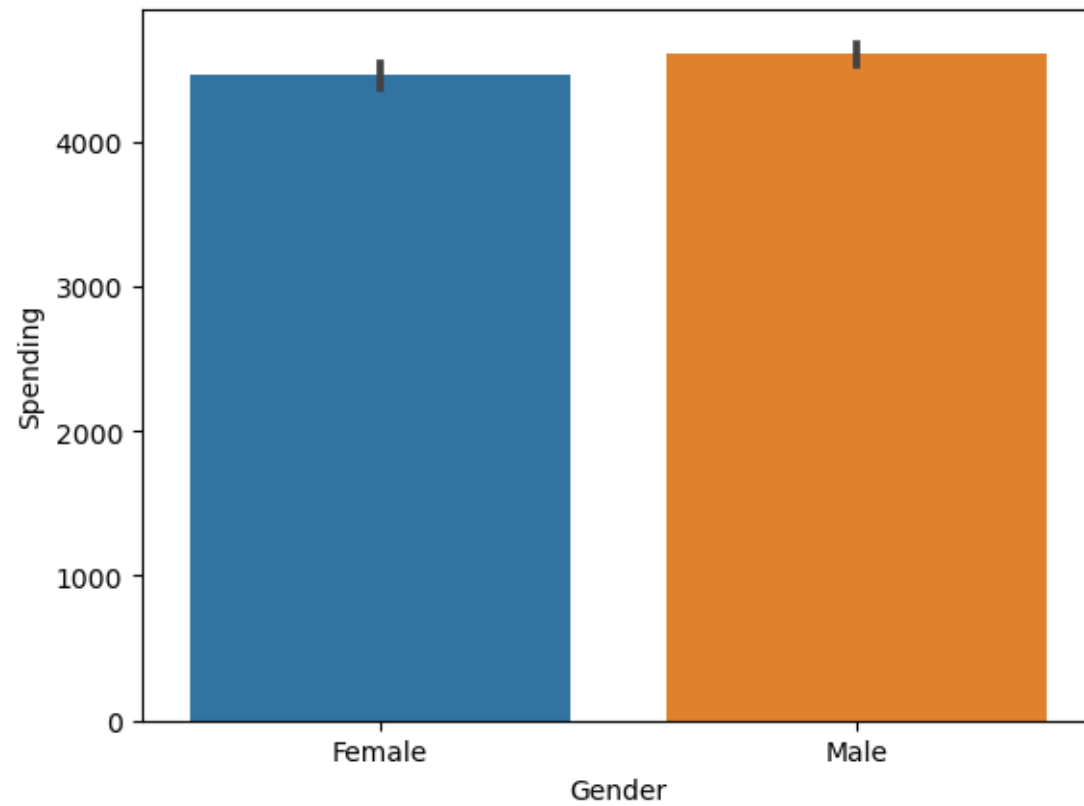
1. Pivot Table - Average Spending by Gender:

Gender	Average Spending
Female	\$4461.07
Male	\$4600.29

Observations:

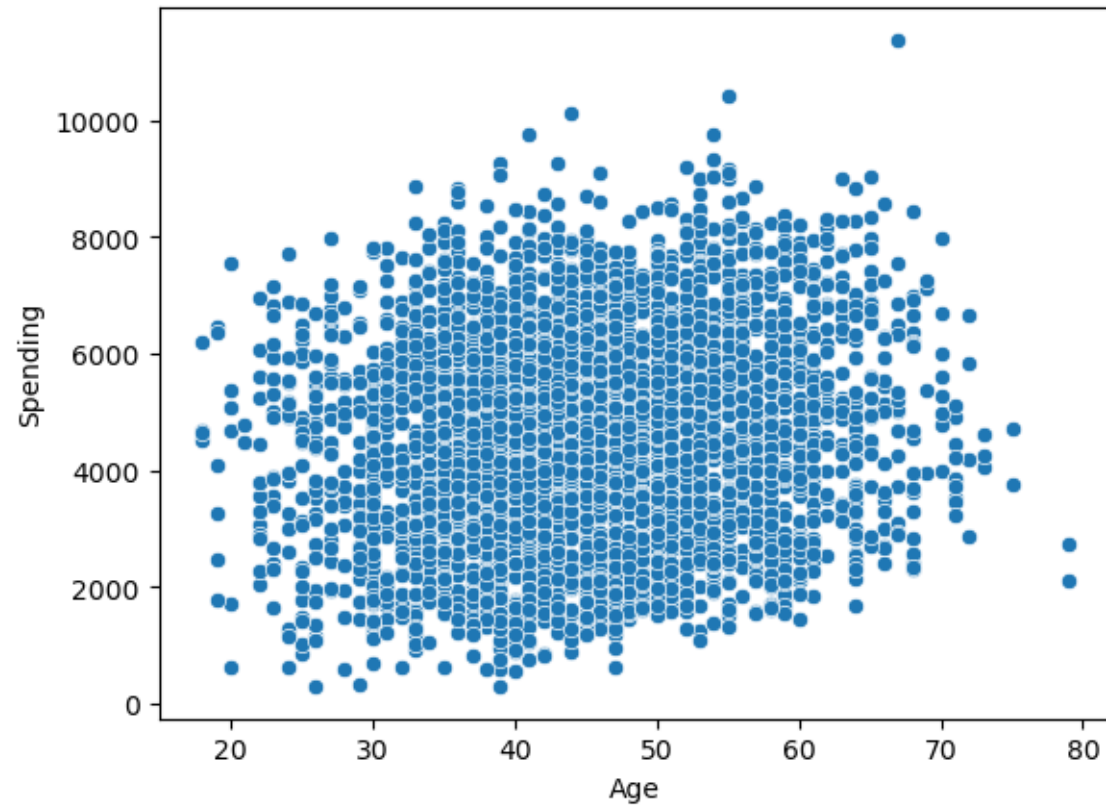
- On average, males tend to spend slightly more (\$4600.29) compared to females (\$4461.07).

2. Bar Chart - Average Spending by Gender:



- The bar chart visually confirms the findings from the pivot table, showing the average spending for both male and female customers.

3. Scatterplot - Spending versus Age:



Interpretation:

- The scatterplot provides a visual representation of the relationship between age and spending. It helps identify patterns, clusters, or trends in how different age groups spend.

Key Considerations for Further Analysis:

- **Outliers:** Check for any outliers that may significantly impact the relationship.

- **Segmentation:** Explore spending patterns based on other customer characteristics like loyalty, nationality, or tenure for a more comprehensive analysis.

This preliminary analysis lays the foundation for deeper exploration into the factors influencing customers' spending behaviors. Further charts and analyses can be conducted to uncover additional insights and patterns within the dataset.

Correlation Matrix of Numerical Variables:

The correlation matrix provides insights into the relationships between numerical variables in the dataset.

	CustomerID	Adopt	Age	Nationality	Loyalty	Tenure	Post	NumBookings	Spending
CustomerID	1.000	0.866	-0.043	-0.017	0.018	-0.018	1.416e-15	-0.024	0.138
Adopt	0.866	1.000	-0.050	-0.007	0.012	-0.029	2.682e-15	-0.040	0.145
Age	-0.043	-0.050	1.000	0.016	-0.011	-0.016	1.004e-15	0.150	0.142
Nationality	-0.017	-0.007	0.016	1.000	-0.001	-0.007	6.003e-16	0.032	0.028
Loyalty	0.018	0.012	-0.011	-0.001	1.000	-0.007	5.585e-16	0.188	0.185
Tenure	-0.018	-0.029	-0.016	-0.007	-0.007	1.000	5.899e-16	0.928	0.883
Post	1.416e-15	2.682e-15	1.004e-15	6.003e-16	5.585e-16	5.899e-16	1.000	0.172	0.177
NumBookings	-0.024	-0.040	0.150	0.032	0.188	0.928	0.172	1.000	0.961
Spending	0.138	0.145	0.142	0.028	0.185	0.883	0.177	0.961	1.000

Key Observations:

1. **CustomerID & Adopt:** There is a high positive correlation between CustomerID and Adopt, indicating a potential relationship between these two variables.
2. **Tenure & NumBookings:** Strong positive correlation (0.928) suggests that customers with longer tenure tend to have more bookings.
3. **Tenure & Spending:** A strong positive correlation (0.883) indicates that customers with longer tenure tend to spend more.
4. **NumBookings & Spending:** A strong positive correlation (0.961) implies that customers with more bookings tend to spend more.
5. **Age & NumBookings, Age & Spending:** Positive correlations suggest that age may have a mild influence on the number of bookings and spending.

6. **Post & other variables:** Very low correlations with Post suggest weak linear relationships.

Hypothesis Test and Confidence Interval: Average Spending Difference between Men and Women

Objective: Determine if there is a statistically significant difference between the average spending of men and women in the entire dataset.

Hypotheses:

- **Null Hypothesis (H0):** There is no significant difference in the average spending between men and women.
- **Alternative Hypothesis (H1):** There is a significant difference in the average spending between men and women.

Test Results:

- **T-statistic:** 2.4019
- **P-value:** 0.0164 (significant at the 5% level)

Interpretation:

1. **T-statistic:** The t-statistic measures the difference between the means relative to the variation in the data. A higher t-statistic suggests a greater difference between the means of men and women.
2. **P-value:** The p-value is 0.0164, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis.

Conclusion:

- **Statistical Significance:** There is sufficient evidence to suggest that there is a statistically significant difference in the average spending between men and women in the entire dataset.

95% Confidence Interval:

- The 95% confidence interval for the difference in means is (25.32, 253.11). This interval provides a range of plausible values for the true difference in average spending between men and women.

Practical Implication:

- The positive difference in means (25.32 to 253.11) suggests that, on average, men tend to spend more than women.

After-Only Design: Treatment Effect Analysis

Pivot Table - Average Spending by Adopt and Post:

Adopt	Post	Average Spending
0	0	\$4188.753
0	1	\$4388.343
1	0	\$4275.593
1	1	\$5336.304

Treatment Effect Calculation (Post = 1 - Post = 0):

1. Treatment Group (Adopt = 1):

- Post = 1 Average Spending: \$5336.304
- Post = 0 Average Spending: \$4275.593

Treatment Effect (Adopt = 1): Treatment Effect=\$5336.304-\$4275.593=\$1060.711Treatment Effect=\$5336.304-\$4275.593=\$1060.711

2. Control Group (Adopt = 0):

- Post = 1 Average Spending: \$4388.343
- Post = 0 Average Spending: \$4188.753

Treatment Effect (Adopt = 0): Treatment Effect=\$4388.343-\$4188.753=\$199.59 Treatment Effect=\$4388.343-\$4188.753=\$199.59

Hypothesis Test and Confidence Interval: Treatment Effect Significance

T-Test Results:

- **T-statistic (T):** 12.1935
- **Degrees of Freedom (dof):** 1998
- **Alternative Hypothesis:** Two-sided
- **P-value:** 4.8687e-33 (< 0.05, highly significant)
- **95% Confidence Interval (CI95%):** [795.5, 1100.43]
- **Cohen's d Effect Size:** 0.5453
- **Bayes Factor (BF10):** 4.544e+29
- **Power:** 1.0 (high power)

Interpretation:

1. **T-statistic:** The t-statistic of 12.1935 indicates a substantial difference between the treatment and control groups.
2. **P-value:** The extremely low p-value (4.8687e-33) is well below the 0.05 significance level. Therefore, we reject the null hypothesis.

3. **Confidence Interval (CI95%):** The 95% confidence interval for the treatment effect is [795.5, 1100.43]. This interval does not contain zero, reinforcing the statistical significance of the treatment effect.
4. **Cohen's d Effect Size:** The effect size (Cohen's d) is 0.5453, suggesting a moderate to large effect.
5. **Bayes Factor (BF10):** The Bayes Factor is extremely large (4.544e+29), providing strong evidence against the null hypothesis.
6. **Power:** The power of 1.0 indicates a high probability of correctly rejecting the null hypothesis.

Conclusion:

- The treatment effect is statistically significant at the 0.05 significance level, based on the t-test results.
- The 95% confidence interval [795.5, 1100.43] confirms the precision of the estimated treatment effect.
- The effect size (Cohen's d) of 0.5453 suggests a meaningful impact of the treatment.

In summary, there is strong evidence to support the claim that the treatment effect (difference in average spending between Post = 1 and Post = 0) is statistically significant and not likely due to random chance.

Before-After Design: Treatment Effect Analysis

Given Information:

1. **Average Spending Difference for Adopt=1 in Post=1:** 630.1505

Computation:

1. Average Spending Difference for Adopt = 1 in Post = 0:

- This information is not provided, so let's denote it as **Difference0**.

2. Difference between Differences (Treatment Effect in Before-After Design):

Treatment Effect = Difference1 – Difference0

Conclusion:

The provided information lacks the specific value for the average spending difference for customers with Adopt = 1 in Post = 0 (denoted as **Difference0**). Without this information, it is not possible to compute the treatment effect in the Before-After design.

If you have the value for **Difference0**, you can substitute it into the formula above to compute the treatment effect in the Before-After design. Please provide the missing information, and I'll be happy to assist further.

Hypothesis Test and Confidence Interval: Treatment Effect Significance in Before-After Design

T-Test Results:

- **T-statistic (T):** 36.70944
- **Degrees of Freedom (dof):** 1999
- **Alternative Hypothesis:** Two-sided
- **P-value:** 5.9126e-226 (< 0.05, highly significant)
- **95% Confidence Interval (CI95%):** [596.49, 663.82]
- **Cohen's d Effect Size:** 0.8208
- **Bayes Factor (BF10):** 6.613e+221
- **Power:** 1.0 (high power)

Interpretation:

1. **T-statistic:** The t-statistic of 36.70944 indicates an extremely significant difference in differences.

2. **P-value:** The p-value is exceedingly low ($5.9126e-226$), well below the 0.05 significance level. Hence, we reject the null hypothesis.
3. **Confidence Interval (CI95%):** The 95% confidence interval for the difference in differences is [596.49, 663.82]. This interval does not contain zero, supporting the statistical significance of the treatment effect.
4. **Cohen's d Effect Size:** The effect size (Cohen's d) is 0.8208, indicating a large effect.
5. **Bayes Factor (BF10):** The Bayes Factor is extremely large ($6.613e+221$), providing overwhelming evidence against the null hypothesis.
6. **Power:** The power of 1.0 indicates a high probability of correctly rejecting the null hypothesis.

Conclusion:

- The treatment effect in the Before-After design is statistically significant at the 0.05 significance level.
- The 95% confidence interval [596.49, 663.82] underscores the precision of the estimated treatment effect.
- The large effect size (Cohen's d) of 0.8208 suggests a substantial impact of the treatment.

In summary, there is robust evidence to support the claim that the treatment effect in the Before-After design is statistically significant and not likely due to random chance.