Analyzing Black Friday Purchase Behavior Based on Gender, Marital Status and Age

1. Introduction:

Walmart Inc., a multinational retail corporation, aims to analyze customer purchase behavior, specifically comparing spending habits between male and female customers during Black Friday. The goal is to understand potential differences in purchase amounts based on various factors to make informed business decisions.

2. Business Problem:

The management team at Walmart Inc. seeks to answer whether spending habits differ between male and female customers during Black Friday. The specific query is: "Do women spend more on Black Friday than men?" With an assumption of 50 million male and 50 million female customers, the company desires insights into gender-based purchase behavior.

[Colab Notebook](#)

3. Dataset Information:

The dataset comprises transactional data from Walmart Stores during Black Friday, containing the following features:

User_ID: User ID
Product_ID: Product ID
Gender: Sex of the User
Age: Age in bins
Occupation: Masked occupation
City_Category: Category of the city (A, B, C)
Stay_In_Current_City_Years: Number of years stayed in the current city
Marital_Status: Marital status
Product_Category: Masked product category
Purchase: Purchase amount

4. Expected Deliverables:

Import and analyze the dataset's structure and characteristics.
Detect and address null values and outliers using methods like boxplots, the describe method, and checking null values.

Conduct data exploration:

Sales Composition:
Male customers contribute to approximately 75% of total sales, indicating a significant share of the purchasing power.

Pie chat of nuber of males vs female customers
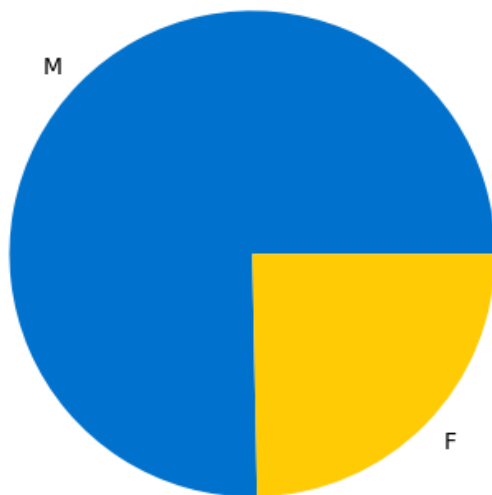

Age Distribution:
Demographic breakdown shows 39% of customers fall within the 26-35 age group, followed by 20% in the 36-45 age bracket, reflecting the primary age segments.

Count of users per Age group
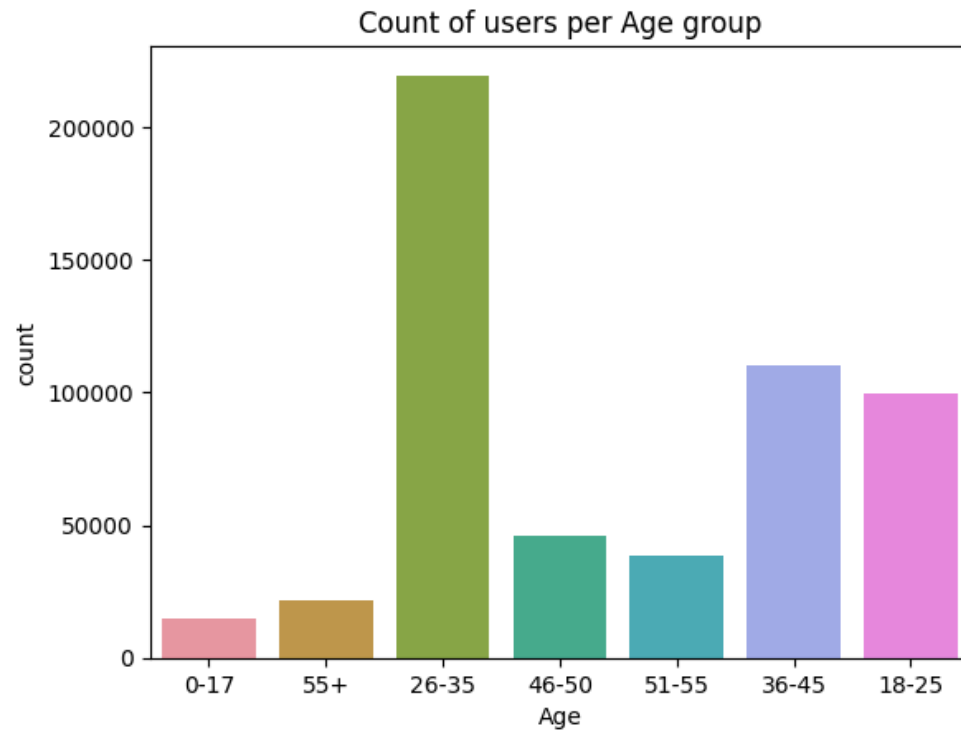
Location and Residence:

Majority of customers are from City B, closely followed by City C and City A, suggesting a geographical concentration.

Count of users in each city_category
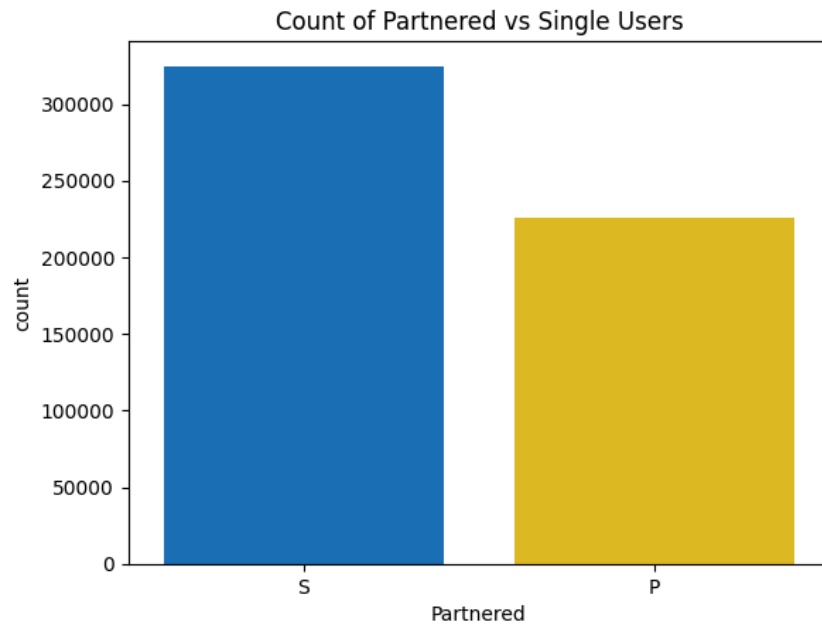
Count of users per years of stay in current city
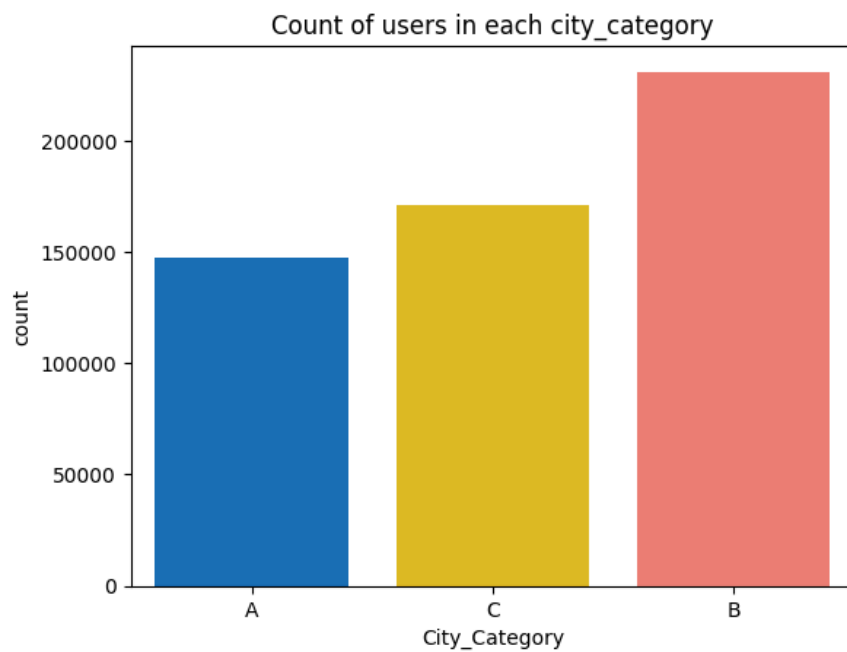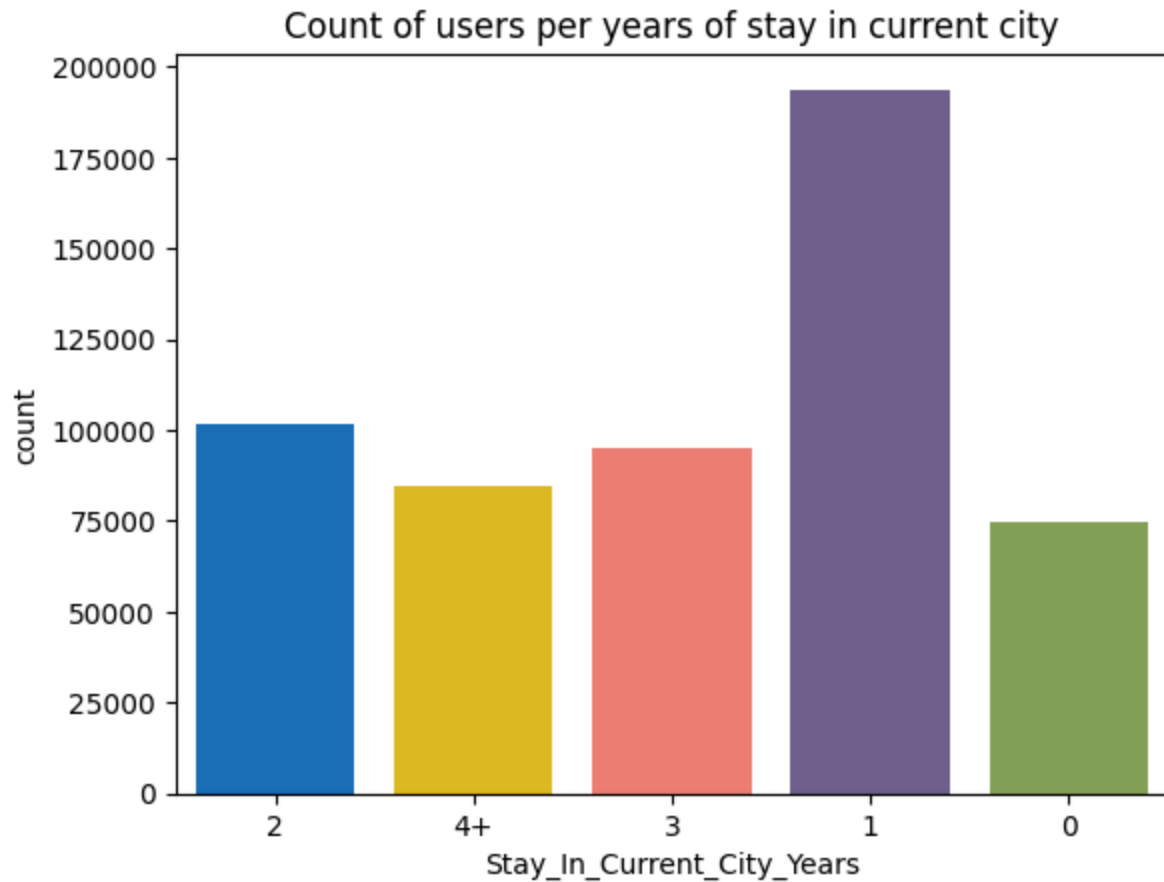
Most customers have lived in the current city for a year, aligning with the dominant age group of 26-35, indicating a demographic of mobile individuals.
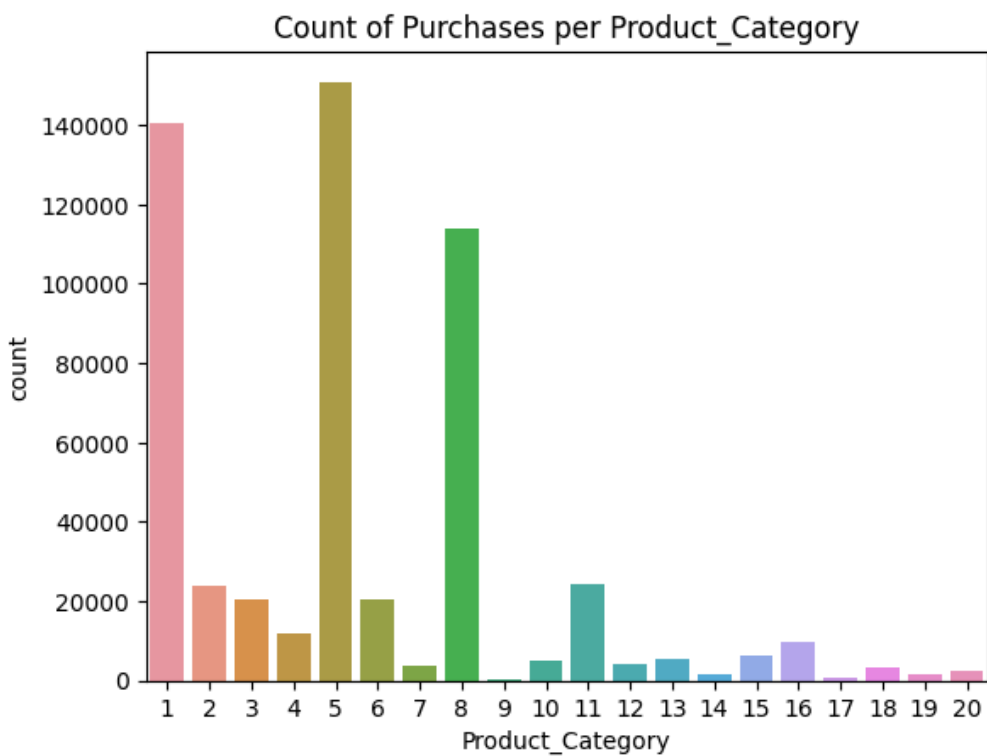

Relationship Status:

About 60% of customers are single or unpartnered, highlighting a significant portion of the customer base.

Count of Partnered vs Single Users

Product Preferences:

Product categories 5, 1, 8, and 11 are the most popular among customers, guiding product stocking and marketing efforts.

Count of Purchases per Product_Category

Purchase Patterns:

Histogram analysis reveals that purchases predominantly fall in the range of approximately 5000 to 10000 units, with a right tail extending up to 250000 units, indicating outliers.

Mean of purchase = 9263.96

Histogram of Purchase data
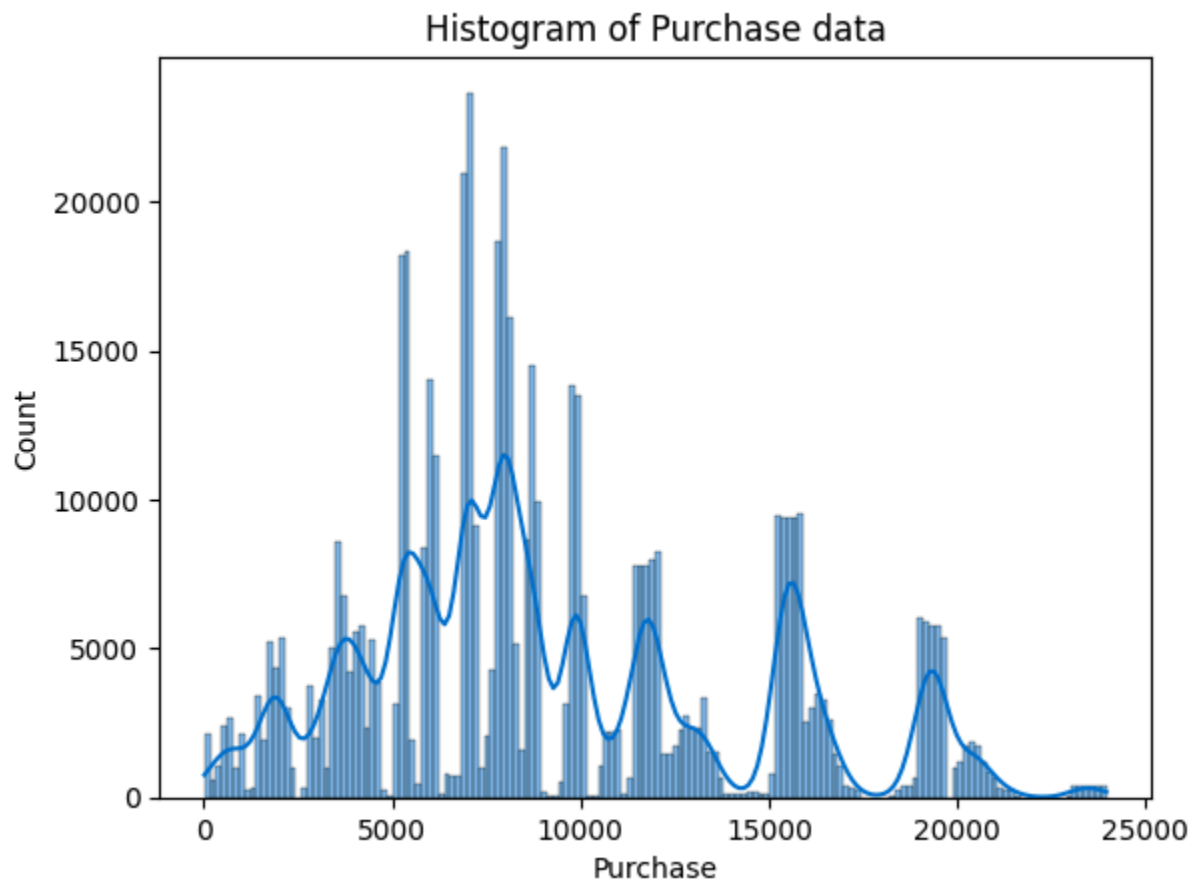
Purchase Behavior by Gender:
KDE plots of purchase values for male and female customers show similar trends, indicating comparable spending behaviors between genders.
Given the male customer population comprises around 75%, their purchasing behavior significantly influences sales trends.

KDE Plot - Male Purchases

KDE Plot - Female Purchases
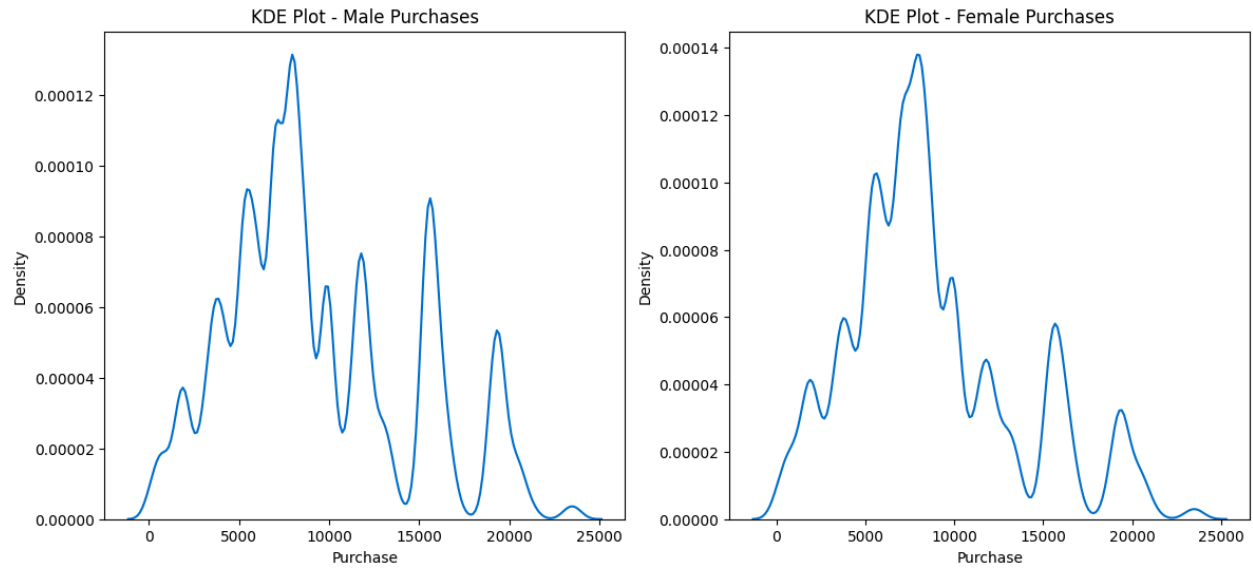
Analysis based in purchase values:

As the population of male customers is around 75% it also shows its impact on the sales made on male customers

Total Purchase by Gender
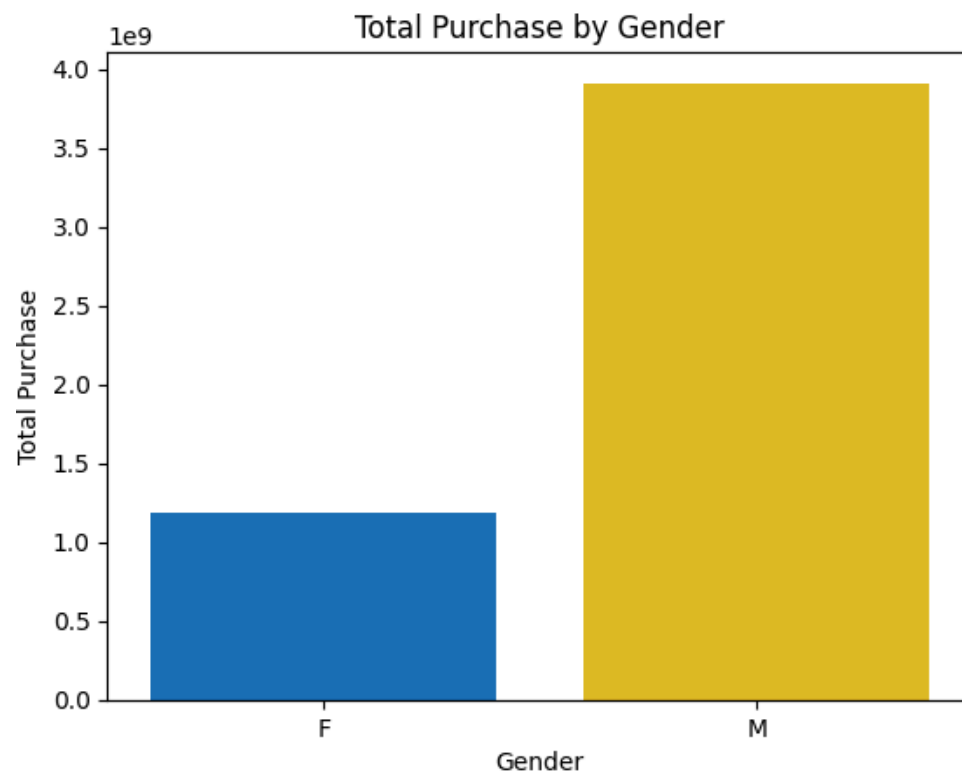
in this graph too, we can see a strong correlation between population in a particular age group and the total purchase amount of that age group

Total Purchase by Age group
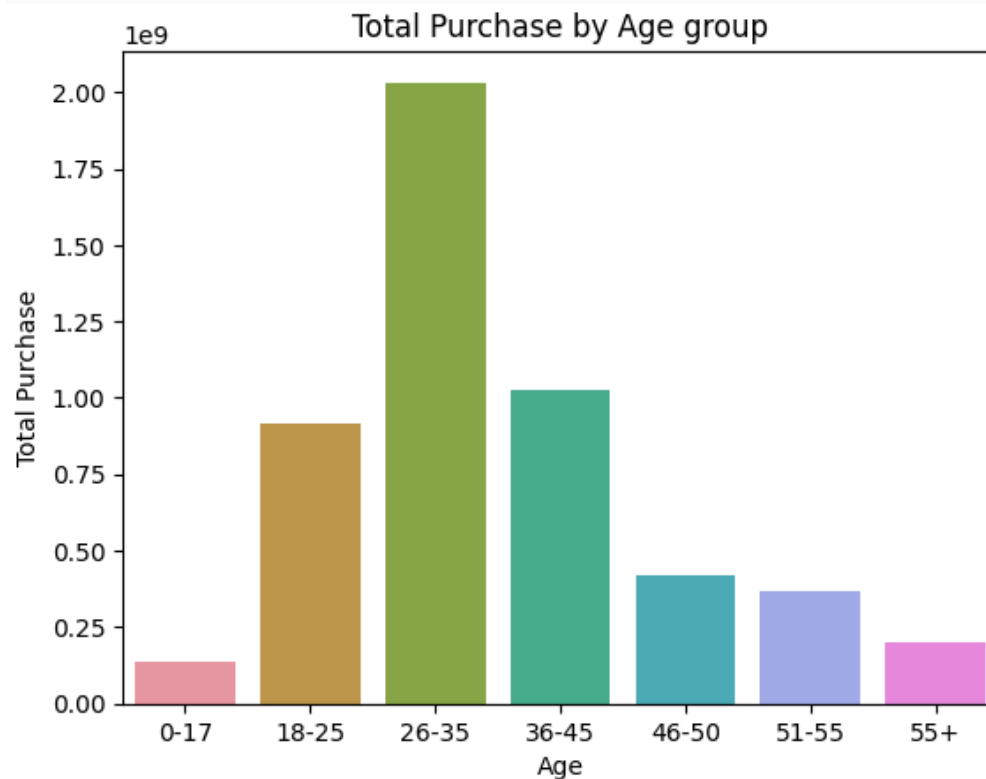
same trend is also followed in the city category

same trend follows through marital status too

most popular product category is 5 followed by 1, 8 and 11 but the most revenue generating product category belongs to 1, then seeing a significant drop of value to product category 5 and 8.

Total Purchase by Product_Category
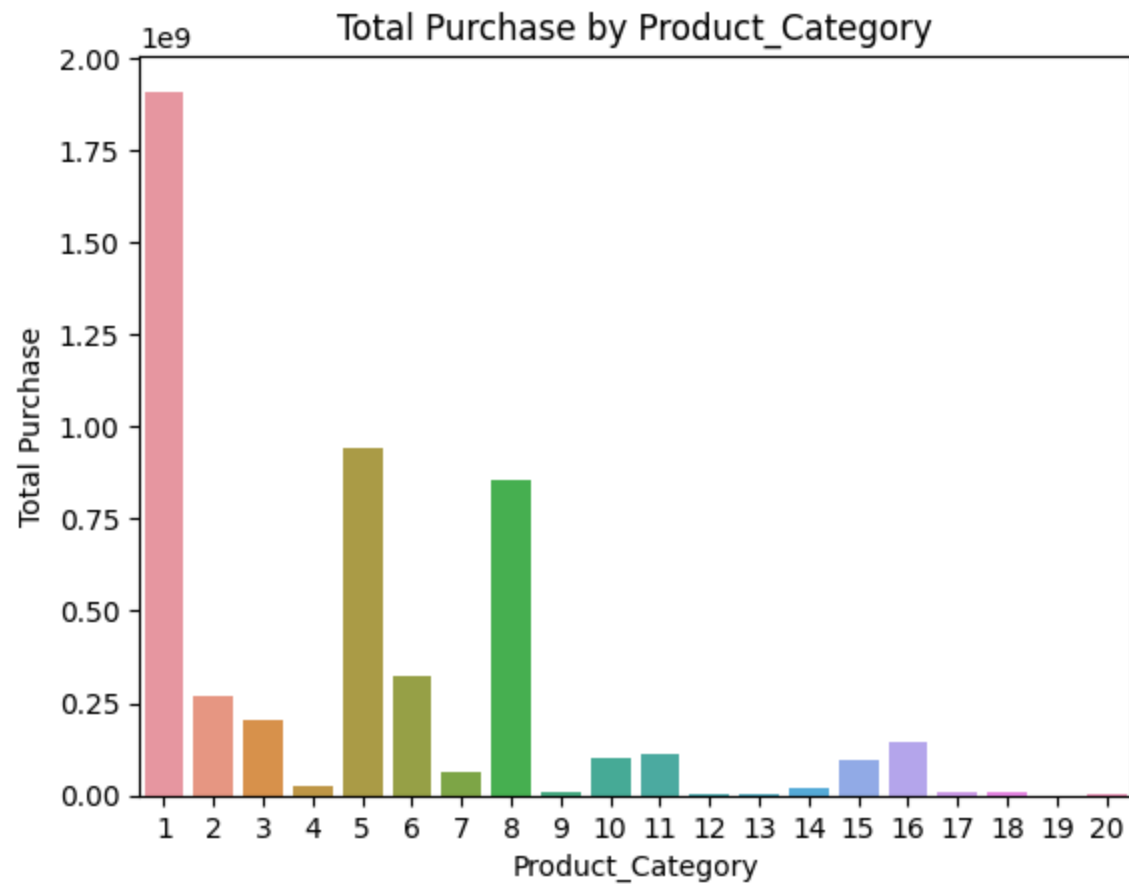
heatmap showcases the sales generated within each gender group, it is evident that male customers have made more number of purchases in the price range of 4590 - 9700, 11500 - 12200 and 15500 - 16300 than that of female purchases

Heatmap of purchase vs Product catergory for male customers

Heatmap of purchase vs Product catergory for female customers

Heatmap of purchase vs Product catergory for customers

Outlier analysis:

Box plot of Purchases to find outliers

#outliers(high singular purchases)
1376 users have made single purchases that are of value greater than 220000
These high revenue generating users buy products from product category 10 and 9
the trend follows here too as the male demographic occupied 75% of the population, they also occupy around the same percentage in high revenue generating customers

Box plot of Product categories to find outliers
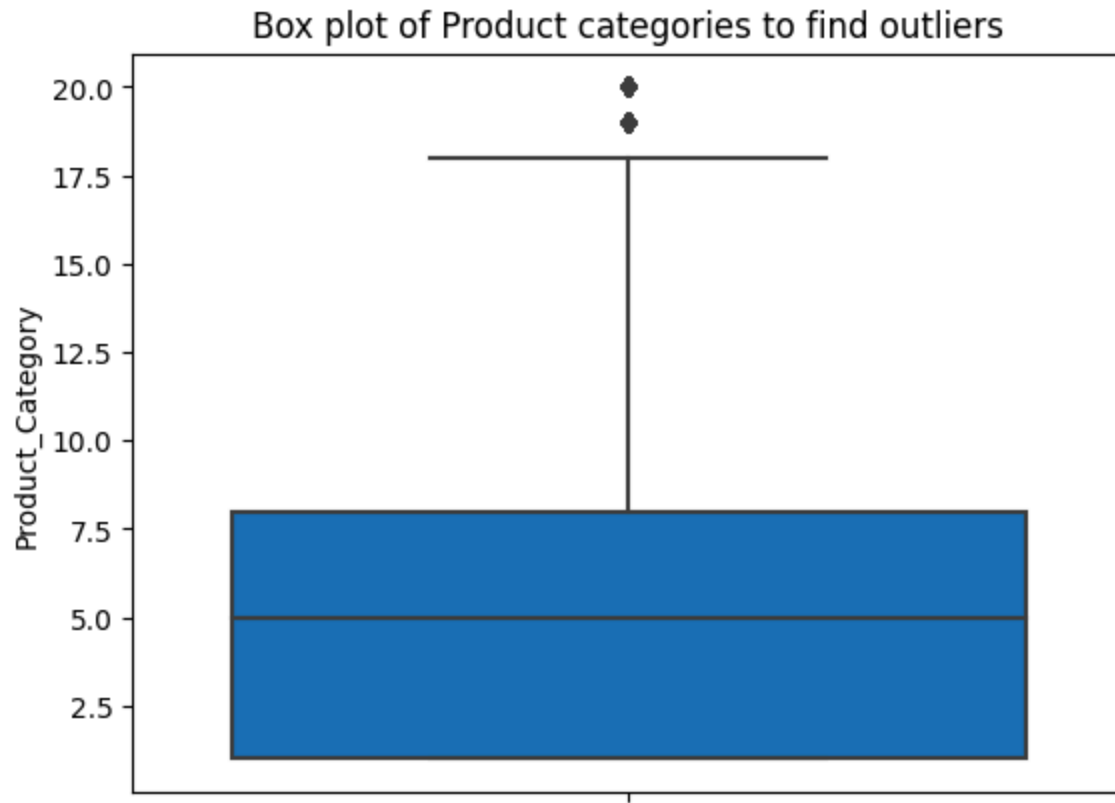
#product categories 19 and 20 are outliers

#product category 19 and 20 contribute only 0.019% in the sales

The trend follows here too as the male demographic occupied 75% of the population, they also occupy around the same percentage in the outlier product category purchases.
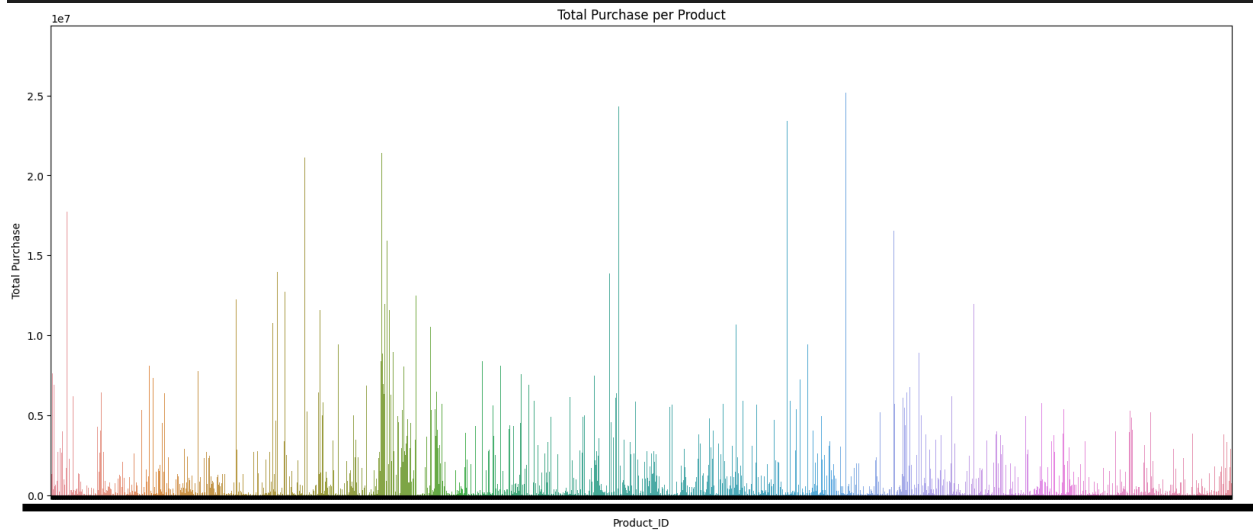
Total Purchase per Product

Track the amount spent per transaction for 50 million female and 50 million male customers, calculating averages and drawing conclusions.

The average purchase amount of male customer is 9437.52
The average purchase amount of female customer is 8734.56

As population ratio of 3 males to 1 female:

- Male customers, despite comprising 75% of the population, have a higher average spending per transaction compared to female customers.
- On average, male customers spend more per transaction, with an average purchase amount of $9437.52, compared to the average spending of approximately $8734.56 for female customers.
- Despite the higher representation of males in the population, the data suggests that they tend to spend more per transaction than females.

- This highlights potential differences in spending behaviors or preferences between genders, indicating that male customers might be making higher-value purchases on average compared to female customers.

Are women spending more money per transaction than men? Why or Why not?

- Male vs. Female Customers:
- t-statistic: 9.948142683305282
- p-value: 3.6256177718456804e-19
- Result: The p-value is extremely low (much smaller than the typical significance level of 0.05), indicating strong evidence against the null hypothesis. Thus, you reject the null hypothesis.
- Conclusion: There is a significant difference in spending between male and female customers. This implies that, on average, one gender spends more than the other.

The statistical analysis of the t-statistic and p-value suggests a significant difference in spending between male and female customers. The results show that, on average, men spend more money per transaction compared to women. This conclusion is supported by the evidence provided by the statistical tests.

Moreover, specific insights from the data also reinforce this finding:

Product Category Preferences: Males contribute significantly more to the revenue from certain product categories, such as Category 1. This indicates a higher tendency for males to make purchases in this category compared to females.

Contribution to Total Purchases: In product categories 1, 5, and 8, males contribute to a larger percentage of the total purchases compared to females. For instance, males contribute around 55% of the total purchases from these categories, whereas females contribute only about 10%.

Use sample averages to find confidence intervals for the population's average spending of male and female customers.Utilize the Central Limit Theorem to compute confidence intervals and observe the distribution of mean expenses by gender.

Analysis:

Male Customer Purchase Ranges:

90% Confidence Interval: (8560.26, 10293.49)

95% Confidence Interval: (8394.24, 10459.51)

99% Confidence Interval: (8069.77, 10783.99)

Insights:

As the confidence level increases, the range of purchase amounts widens, indicating higher variability in spending behavior among male customers.
The 99% confidence interval encompasses a broader range, implying more extreme values or outliers in male spending behavior at higher levels of confidence.
Female Customer Purchase Ranges:

90% Confidence Interval: (7917.08, 9613.70)

95% Confidence Interval: (7754.56, 9776.21)

99% Confidence Interval: (7436.94, 10093.84)

Insights:

- Similar to male customers, as the confidence level increases, the range of purchase amounts widens for female customers.
- The 99% confidence interval also encompasses a wider range, indicating potential outliers or extreme values in female spending behavior at higher confidence levels.

Comparison between Genders:

- At the 90%, 95%, and 99% confidence levels, the male customer purchase ranges generally span higher values compared to female customers.
- The upper limits of the male customer purchase ranges consistently exceed those of female customers across all confidence intervals.

Gender-Specific Spending Behaviors:

- Both genders exhibit variability in their spending behaviors, as indicated by the widening ranges with increasing confidence levels.
- While males tend to have higher upper limits in their purchase ranges, females demonstrate comparatively lower upper limits across all confidence intervals.

Analysis of Confidence Intervals:

- At the 90% confidence level, the intervals slightly overlap.
- At the 95% confidence level, there is a more noticeable overlap between the confidence intervals.
- At the 99% confidence level, there is a substantial overlap between the confidence intervals.

Conclusion:
- Based on the provided confidence intervals, the intervals for average male spends and average female spends do overlap across different confidence levels. This suggests that there isn't a statistically significant difference between the average spending of male and female customers.

Product Category Analysis:

- Highest female purchases are primarily from product categories 5 and 8.
- Highest male purchases are from product categories 5, 8, and 1.
- The most significant purchases, across genders, originate from product categories 5, 8, and 1.
- Insight: Product categories 5 and 8 are popular with both genders, whereas category 1 is particularly favored among males.
- Implication: Males contribute significantly to the revenue from product_category 1, indicating a higher propensity for purchases in this category compared to females.

Use sample averages to find confidence intervals for the population's average spending of male and female customers.Utilize the Central Limit Theorem to compute confidence intervals and observe the distribution of mean expenses by Age group and gender.

Here's a comparison of the given 95% confidence intervals for different age groups:

Age Group 0-17:

Male customers: (8213.48, 10158.08)
Female customers: (7435.23, 9357.37)

Insight: There's a significant overlap between the confidence intervals for male and female customers.
Implication: Both genders within this age group demonstrate similar purchasing behavior. Walmart can focus on gender-neutral or unisex products appealing to teenagers.

Age Group 18-25:

Male customers: (8452.45, 10456.01)
Female customers: (7297.36, 9337.88)

Insight: The confidence intervals for male and female purchases overlap moderately.
Implication: Although there's some variability, targeting promotions or products suitable for both genders within this young adult category can be effective.

Age Group 26-35:

Male customers: (8401.81, 10484.63)
Female customers: (7903.10, 9579.26)

Insight: There's a substantial overlap between confidence intervals.
Implication: Similar spending habits among males and females in this age bracket indicate potential for marketing strategies emphasizing universal products rather than gender-specific ones.

Age Group 36-45:

Male customers: (8479.82, 10289.15)
Female customers: (8118.06, 9835.54)

Insight: Confidence intervals overlap, but there's a narrower overlap compared to younger age brackets.
Implication: While spending behaviors are similar, there might be nuanced preferences that could benefit from a combination of gender-neutral and selective gender-focused marketing strategies.

Age Group 45-50:

Male customers: (8448.39, 10292.44)
Female customers: (7981.84, 9729.09)

Insight: Confidence intervals have a significant overlap.
Implication: With overlapping intervals, Walmart can emphasize products or promotions appealing to both genders without specific gender targeting.

Age Group 51-55:

Male customers: (8759.06, 10767.62)
Female customers: (8105.96, 9944.36)

Insight: Overlapping intervals suggest similarity in purchasing behavior.
Implication: Within this age range, focusing on product quality and universal appeal might be more beneficial than gender-specific marketing.
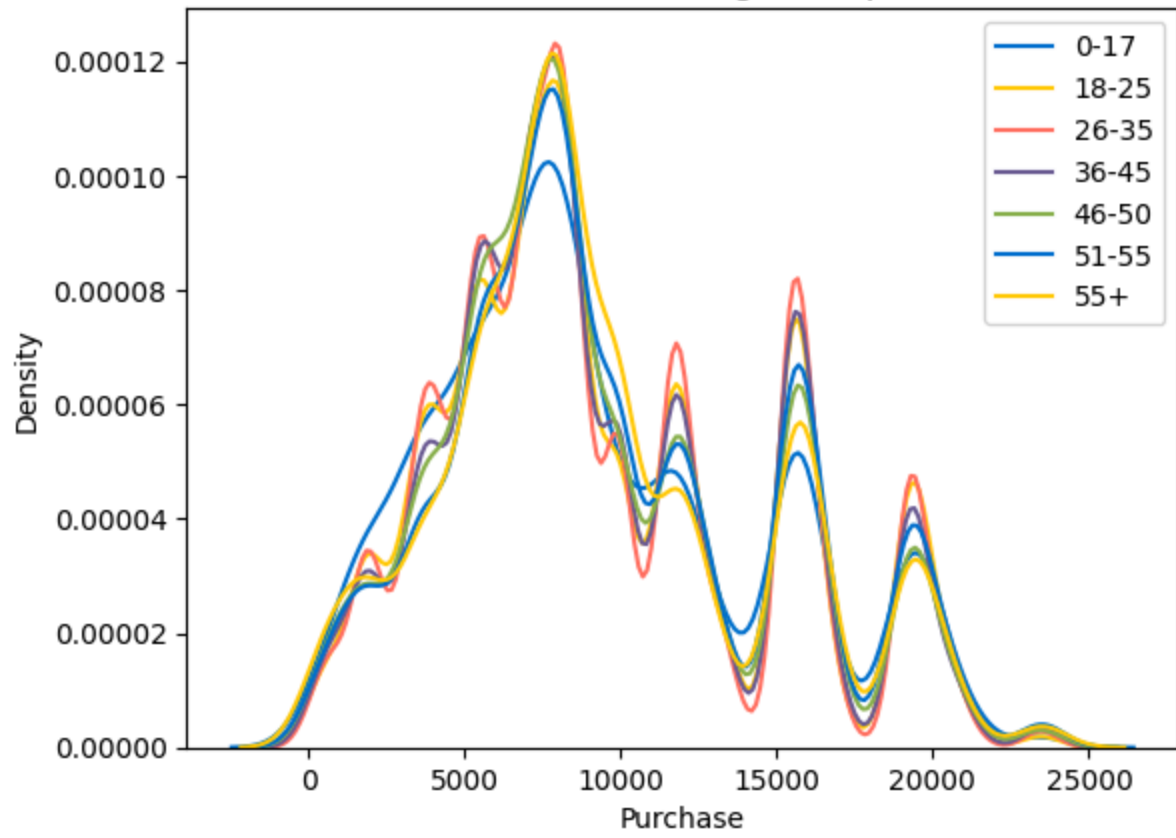

Age Group 55+:

Male customers: (8432.95, 10339.30)
Female customers: (8074.94, 9820.51)

Insight: Confidence intervals overlap, but with a smaller range.
Implication: Though there's some variation, both male and female customers in this age group might respond well to marketing strategies focusing on convenience and practicality.

KDE Plots for Different Age Groups of Male

KDE Plots for Different Age Groups of Female
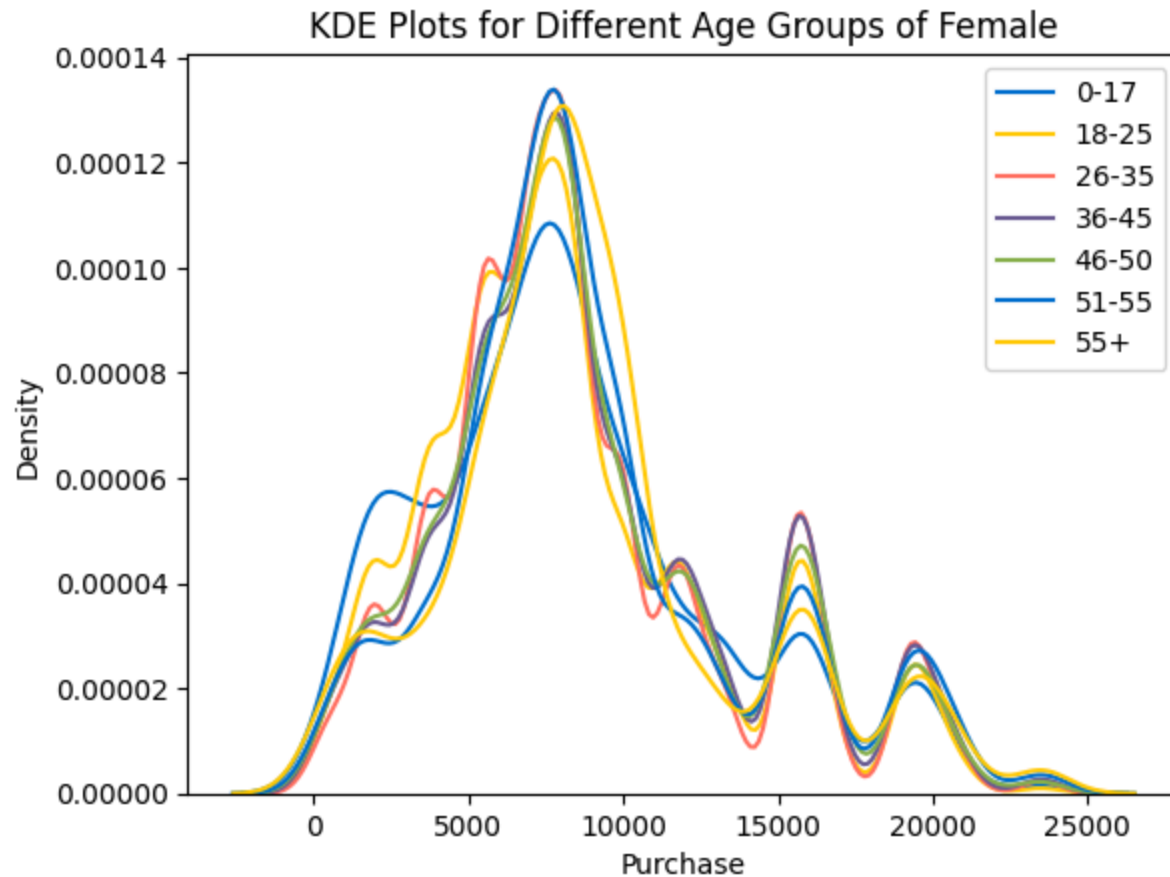
Across various age groups, there's a trend towards overlapping confidence intervals, indicating similarities in spending behavior between genders. Walmart could adopt more inclusive and universal marketing strategies while selectively targeting specific preferences within age brackets.

Results when the same activity is performed for Married vs Unmarried

Mean of Mean Purchases of Partnered Customers: 9278.34,
Mean of Mean Purchases of Unpartnered Customers: 9239.75

Partnered vs. Unpartnered Customers:
- t-statistic: -1.9216423416255495
- p-value: 0.05522113463304455
- Result: The p-value is slightly above the typical significance level of 0.05, indicating weak evidence against the null hypothesis. Thus, you fail to reject the null hypothesis.

- Conclusion: There is no significant difference in spending between partnered and unpartnered customers. This suggests that partnership status doesn't significantly influence the average spending behavior.

For Partnered Customers:
- 90% of Purchases: (8384.94, 10056.85)
- 95% of Purchases: (8224.79, 10217.00)
- 99% of Purchases: (7911.80, 10529.99)

For Unpartnered Customers:
- 90% of Purchases: (8431.50, 10119.63)
- 95% of Purchases: (8269.80, 10281.33)
- 99% of Purchases: (7953.76, 10597.37)

Analysis and Insights:

Partnered vs. Unpartnered Purchase Ranges:
- At each confidence level (90%, 95%, and 99%), the purchase ranges of both partnered and unpartnered customers exhibit overlap.
- The purchase ranges of partnered customers slightly overlap with those of unpartnered customers at all confidence levels.

Variability in Spending Behaviors:
- There's variability in the spending behaviors within both customer groups, as evidenced by the widening purchase ranges as the confidence level increases.

Confidence Intervals Analysis:

Insight: The confidence intervals for purchases of partnered and unpartnered customers also display an extensive overlap.
Implication: The extensive overlap suggests a lack of significant difference in average purchases between the two customer groups at a 90%, 95% and 99% confidence level.

Consistency in Findings:

Insight: Both the mean of mean samples and the confidence intervals indicate a consistent pattern.

KDE Plots for Different Marital status Groups and their purchase pattern
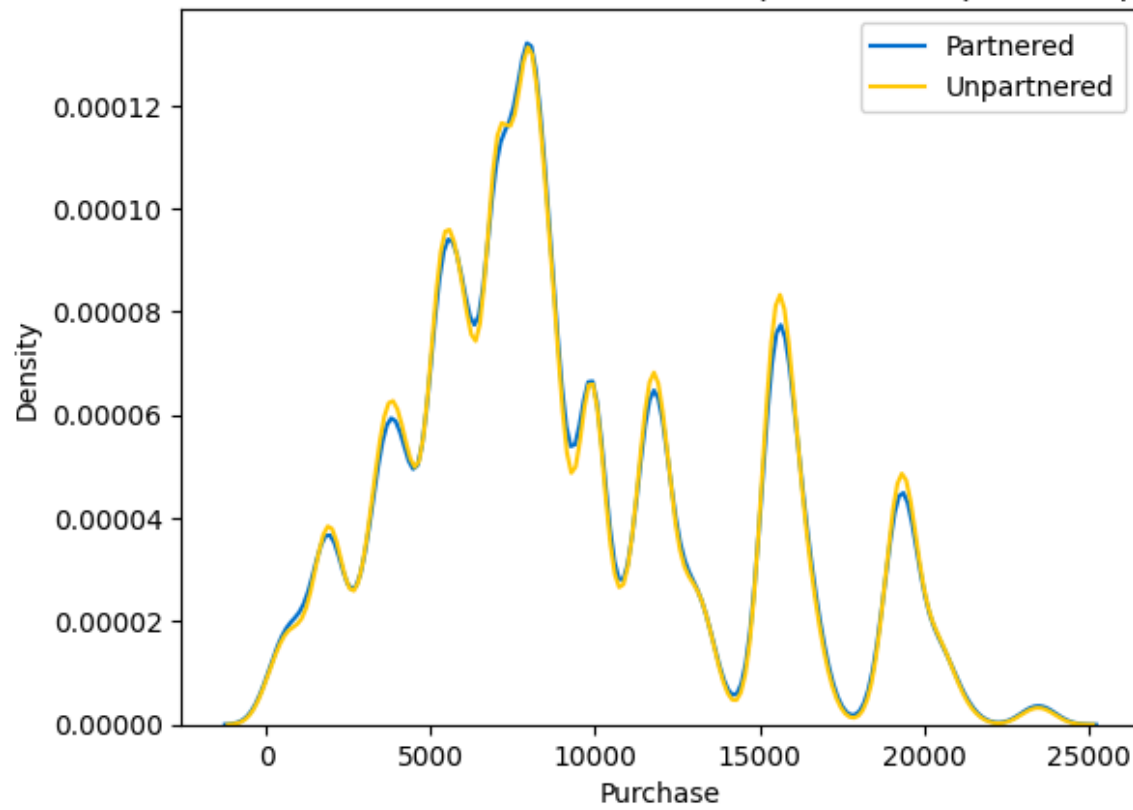
Implication: The similarity in results strengthens the observation that partnership status might not strongly correlate with purchasing behavior.

Analyze spending behaviors based on marital status and age groups (0-17, 18-25, 26-35, 36-50, 51+ years).

- Mean purchase for teenage customers: $8949.389
- Mean purchase for young adult customers: $9134.940
- Mean purchase for adult customers: $9221.427
- Mean purchase for middle-aged customers: $9324.140
- Mean purchase for customers in their 50s: $9211.250
- Mean purchase for customers in their mid-50s: $9559.721
- Mean purchase for customers aged 55+: $9285.287

These figures represent the average spending per transaction for various age groups. Typically, older age groups tend to have slightly higher mean purchase amounts compared to younger age groups. This observation might indicate that older customers, particularly those in their mid-50s and above, spend slightly more per transaction compared to younger age categories.

KDE Plots for Different Age Groups vs Pucrhase patterns
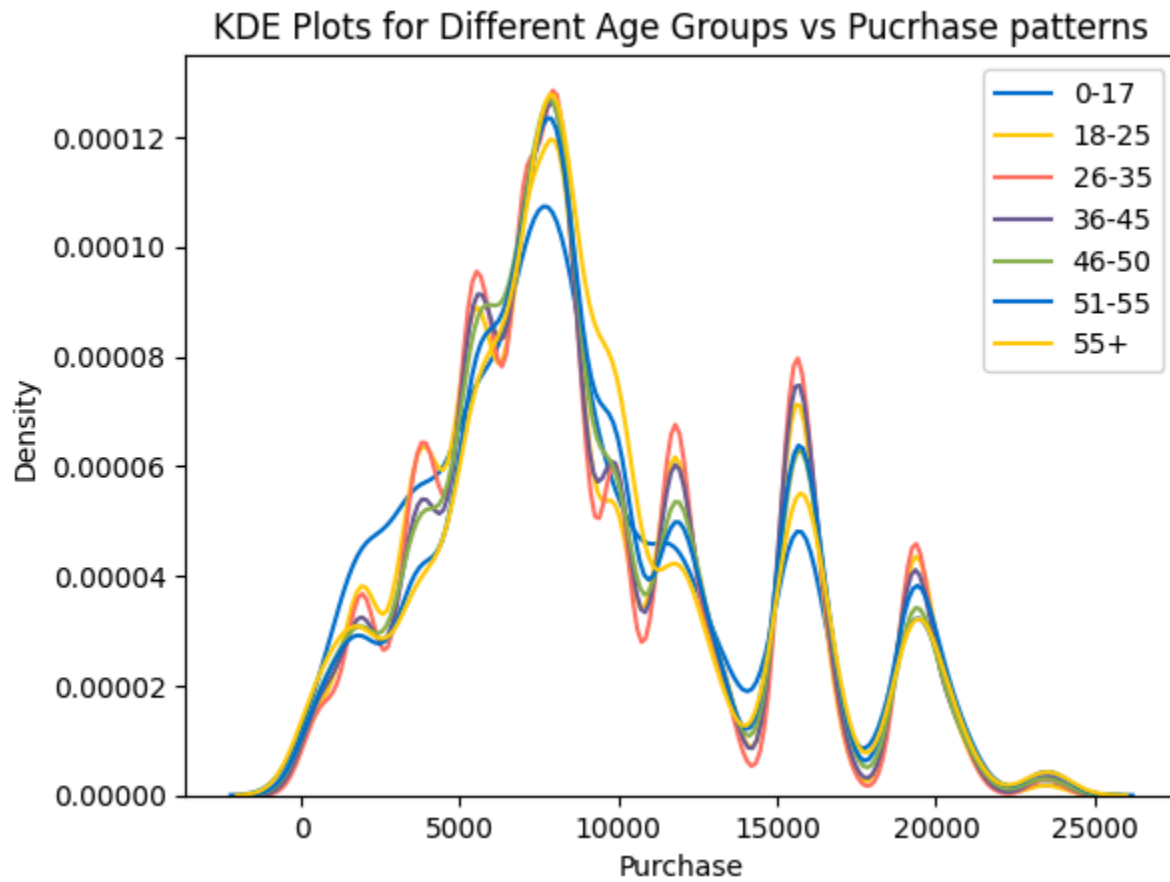
90% Confidence Interval:
- Teenage customers: (8072.84, 9808.67)
- Young adult customers: (8363.04, 10022.62)
- Adult customers: (8455.60, 9970.99)
- Middle-aged customers: (8456.79, 10050.82)
- Early 50s aged customers: (8368.14, 10075.56)
- Mid 50s aged customers: (8542.34, 10112.30)
- 55+ aged customers: (8543.51, 10040.62)

95% Confidence Interval:
- Teenage customers: (7906.57, 9974.94)
- Young adult customers: (8204.07, 10181.59)
- Adult customers: (8310.45, 10116.14)

- Middle-aged customers: (8304.10, 10203.51)
- Early 50s aged customers: (8204.59, 10239.11)
- Mid 50s aged customers: (8391.96, 10262.68)
- 55+ aged customers: (8400.10, 10184.02)

99% Confidence Interval:
- Teenage customers: (7581.61, 10299.90)
- Young adult customers: (7893.38, 10492.28)
- Adult customers: (8026.75, 10399.84)
- Middle-aged customers: (8005.68, 10501.92)
- Early 50s aged customers: (7884.94, 10558.76)
- Mid 50s aged customers: (8098.04, 10556.59)
- 55+ aged customers: (8119.83, 10464.30)

Consistency Across Confidence Levels:
- Generally, as the confidence level increases from 90% to 99%, the range of purchase amounts widens for all age groups. This wider range at higher confidence levels indicates a higher level of certainty in capturing more extreme values within the intervals.

Teenage and Young Adult Customers:
- For teenage and young adult customers, there is a noticeable overlap among their confidence intervals across different confidence levels (90%, 95%, 99%). This suggests that the purchase behavior for these age groups doesn't significantly deviate concerning the confidence levels.

Adult and Middle-Aged Customers:
- Similar to the younger age groups, adults and middle-aged customers also show consistent overlaps in their confidence intervals across different confidence levels. This consistency suggests stable purchasing behaviors across these age segments.

Elderly Age Groups (50s and 55+):
- The confidence intervals for customers aged in their 50s and above display slight variations and wider ranges compared to younger age groups, especially at higher confidence levels (95% and 99%). This indicates more variability in purchasing behavior or potential outliers within these older age segments.

Range of Purchase Behavior:
- Across all age groups, there's a consistent pattern of wider intervals at higher confidence levels. This implies a broader range of purchase behaviors or potential outliers within these age categories, particularly as the confidence level increases.

Increase in Range as Confidence Increases:

- Notably, the increase in the range of confidence intervals as the confidence level goes from 90% to 99% suggests higher variability or dispersion in purchase amounts within each age group at the higher confidence level.

Stability in Purchase Behavior:
- Despite the wider ranges at higher confidence levels, there is still a significant overlap in confidence intervals between adjacent age groups. This suggests relatively stable and comparable purchase behaviors between these age segments.

5. Conclusion and Recommendations:

1.Targeted Marketing Strategies:
- Gender-Specific Promotions: Tailor promotions targeting high-value products to male customers who exhibit a tendency for higher-value purchases. For female customers, emphasize versatile and moderately priced products to cater to diverse preferences.
- Inclusive Marketing: Develop inclusive marketing strategies that balance gender-specific and gender-neutral promotions. Focus on products appealing to both partnered and unpartnered customers, considering the minimal differences in purchasing behaviors.

2. Age-Specific Approaches:
- Target Higher Age Groups: As older age groups tend to exhibit higher-value purchases, create promotions or product assortments that target these segments. Ensure product quality and offerings resonate with these demographics while ensuring universal appeal across age groups.
- Consistent Marketing Across Ages: Maintain consistent marketing strategies to appeal to the consistent lower limit of purchase behaviors across various age groups, ensuring inclusivity and engagement across demographics.

3. Product Assortment and Promotion Strategy:
- Leverage Popular Categories: Utilize insights from popular product categories such as 5, 8, and 1 to guide product stocking and marketing efforts. As these categories significantly contribute to overall sales, focus on optimizing promotions for products within these categories.

4. Gender-Neutral Promotions:
- Focus on Inclusivity: Develop strategies that cater to both genders. Offer gender-neutral products and promotions to attract a broader audience and ensure inclusivity. This can involve promotions that appeal universally across genders.

5. Operational Changes:

- Stock Management: Adjust inventory to accommodate popular product categories. Ensure the availability of products in categories like 5, 8, and 1, which drive significant sales.

6. Black Friday Campaign Optimization:
- Tailored Offers: Create Black Friday deals targeting higher-value purchases to attract customers from segments showing a propensity for such purchases.

7. Continuous Analysis and Adaptation:
- Dynamic Strategies: Continuously analyze and adapt marketing strategies based on changing trends and consumer behavior. Periodic reassessment of customer preferences and purchasing patterns is crucial to stay aligned with evolving consumer needs.