Business Case: Walmart - Confidence Interval and CLT

Goal:

The objective of this analysis is to assess customer purchasing behavior at Walmart, focusing on average spending across different demographics, specifically age and gender groups. Using the **Central Limit Theorem (CLT)**, we aim to compute confidence intervals for average purchase amounts to determine spending patterns and identify significant trends. This analysis will help Walmart tailor its strategies for product promotions, inventory management, and targeted marketing, ultimately improving customer satisfaction and boosting revenue.

Approach:

To achieve the goal, I incorporated both statistical analysis (using CLT and confidence intervals) and exploratory data analysis techniques, including both non-graphical and graphical methods, as outlined below:

1. **Data Preparation**:
    - Preprocessed the dataset by handling categorical variables, mapping them where necessary (e.g., gender encoded as numeric), and ensuring data types are consistent for analysis.
    - Verified data quality to eliminate errors and missing values that could impact results.
2. **Central Limit Theorem (CLT) and Confidence Intervals**:
    - Leveraged CLT to estimate confidence intervals for the average purchase amount for each gender, utilizing different sample sizes (e.g., 300, 3,000, and 30,000) to observe changes in interval width and stability.
    - Calculated 95% confidence intervals on the full dataset and for smaller sample sizes to assess variability and the impact of sample size on estimate precision.
    - Conducted bootstrapping to validate the confidence intervals and ensure robust estimates of mean purchase amounts across different demographic segments.
3. **Non-Graphical Analysis**:
    - Performed descriptive analysis to gather key metrics, including frequency distributions and summary statistics (mean, median) for spending amounts within age and gender groups.
    - Computed the percentage distribution of purchases per age group and gender to provide an overview of customer preferences.
4. **Graphical Analysis**:
    - Visualized purchasing patterns across age groups and genders using histograms, boxplots, and count plots to illustrate distribution and variability.
    - Used heatmaps and pair plots to explore correlations between spending and other variables, helping to identify significant relationships and trends.
5. **Interpretation of Insights**:
    - Combined insights from CLT, confidence intervals, and exploratory analysis to draw actionable conclusions about customer behavior.
    - Provided recommendations based on observed trends, such as targeted promotions for specific age groups or gender-based product marketing.

By integrating the Central Limit Theorem and confidence interval analysis with descriptive and visual analytics, this report aims to deliver comprehensive insights into Walmart's customer purchase behavior. These insights will help guide data-driven business decisions, enabling Walmart to effectively cater to its diverse customer base.

1. Import and Initial Analysis of the Dataset

Code:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

# Load the dataset
data = pd.read_csv('Downloads/walmart.csv')
```

```
print(data.dtypes)
```

```
User_ID                      int64
Product_ID                  object
Gender                      object
Age                         object
Occupation                   int64
City_Category               object
Stay_In_Current_City_Years  object
Marital_Status               int64
Product_Category             int64
Purchase                     int64
dtype: object
```

```
print("Dataset shape:", data.shape)
```

```
Dataset shape: (550068, 10)
```

```
print(data.isnull().sum())
```

```
User_ID                     0
Product_ID                  0
Gender                      0
Age                         0
Occupation                  0
City_Category               0
Stay_In_Current_City_Years  0
Marital_Status              0
Product_Category            0
Purchase                    0
dtype: int64
```

No Missing Values: The dataset appears to be clean, with no missing values. This is a good starting point for analysis as it ensures data integrity.

Initial Observations--

Diverse Customer Base: The dataset contains information on a wide range of customers, including different genders, ages, occupations, cities, and lengths of stay.

Varied Product Categories: The Product_Category column suggests a diverse product range, catering to various customer needs and preferences.

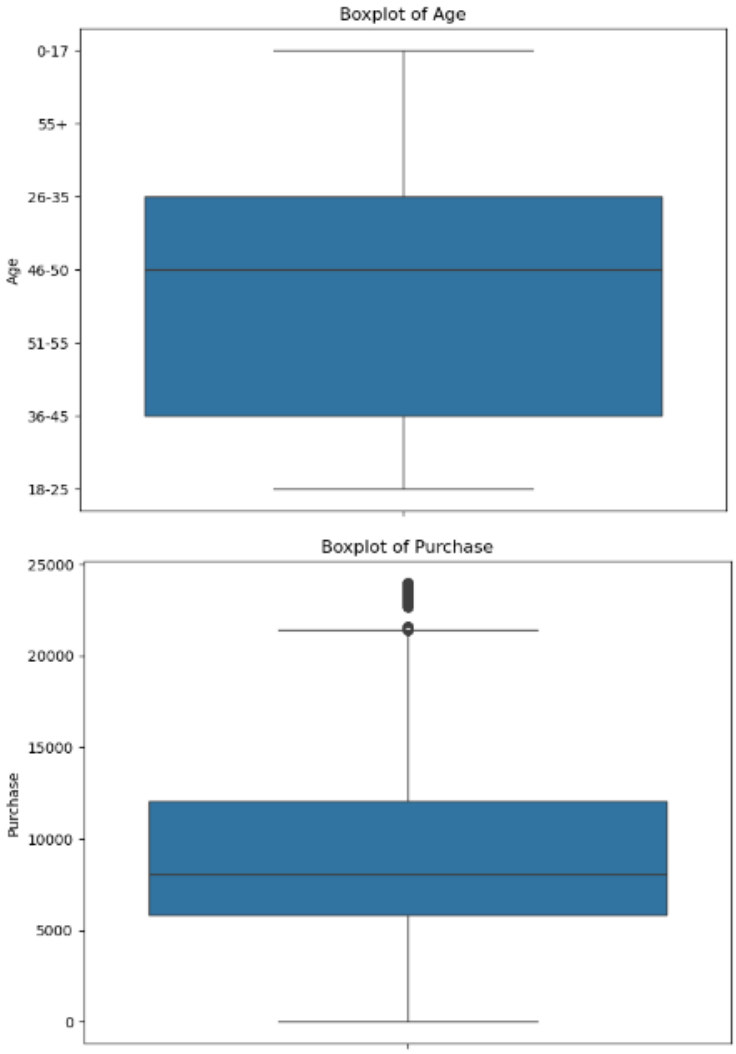
Purchase Behavior: The Purchase column provides insights into customer spending habits.

2. Detect Null Values and Outliers.

Code:

```
# Boxplot for continuous variables
continuous_vars = ['Age', 'Purchase']  # Add other continuous columns as necessary
for var in continuous_vars:
    plt.figure(figsize=(8, 6))
    sns.boxplot(data[var])
    plt.title(f'Boxplot of {var}')
    plt.show()
```

Output:



3. Product Purchase by Age Group:

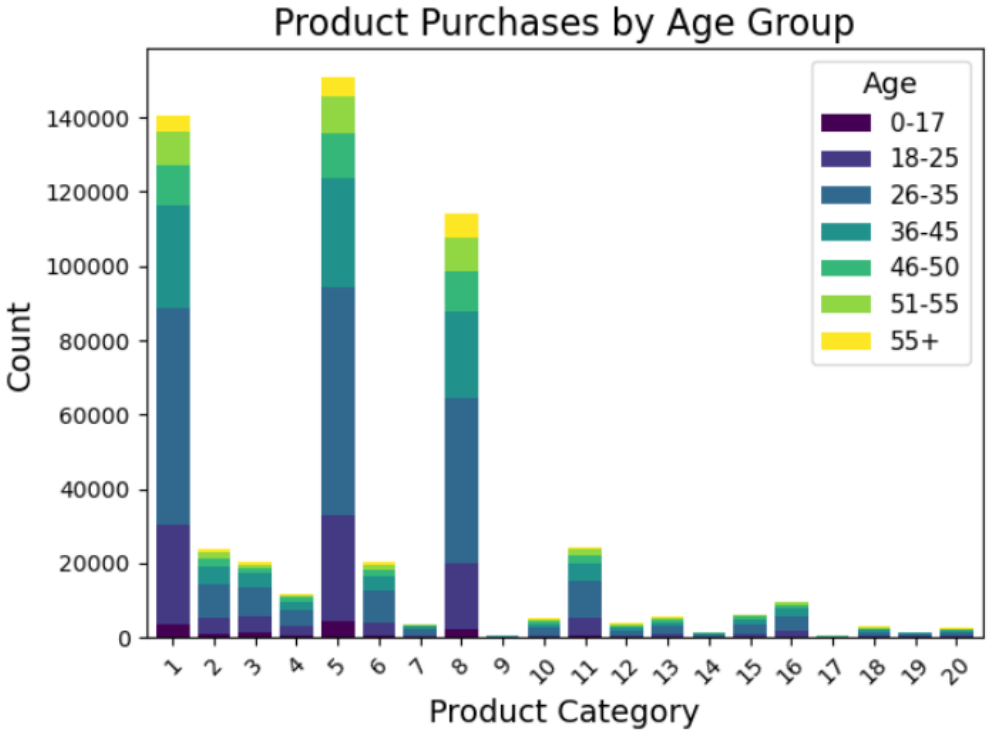
Code:

```
# Aggregate the data to get counts of each age group within each product category
age_product_counts = data.groupby(['Product_Category', 'Age']).size().unstack().fillna(0)


plt.figure(figsize=(12, 8))
age_product_counts.plot(kind='bar', stacked=True, colormap='viridis', width=0.8)
plt.title('Product Purchases by Age Group', fontsize=16)
```

```
plt.xlabel('Product Category', fontsize=14)
plt.ylabel('Count', fontsize=14)
plt.xticks(rotation=45)
plt.legend(title='Age', title_fontsize='13', loc='upper right', fontsize='11')
plt.tight_layout()  # Adjust layout to fit all elements
plt.show()
```

Output:



Insights:

Product Categories 1, 5, and 8 are highly popular across all age groups: These categories seem to have a broad appeal, attracting customers of all ages.

0-17 Age Group: This group shows a strong preference for Product Categories 1, 5, and 8. They also show a higher interest in Product Category 18 compared to other age groups.

18-25 Age Group: This group exhibits a similar pattern to the 0-17 group, with a strong preference for Product Categories 1, 5, and 8. They also show a higher interest in Product Category 16 compared to other age groups.

26-35 Age Group: This group maintains a high interest in Product Categories 1, 5, and 8, but also shows a growing interest in Product Categories 6 and 7 compared to younger age groups.

Recommendations:

Target Younger Age Groups: Focus on promoting Product Categories 1, 5, and 8 to attract younger customers (0-35).

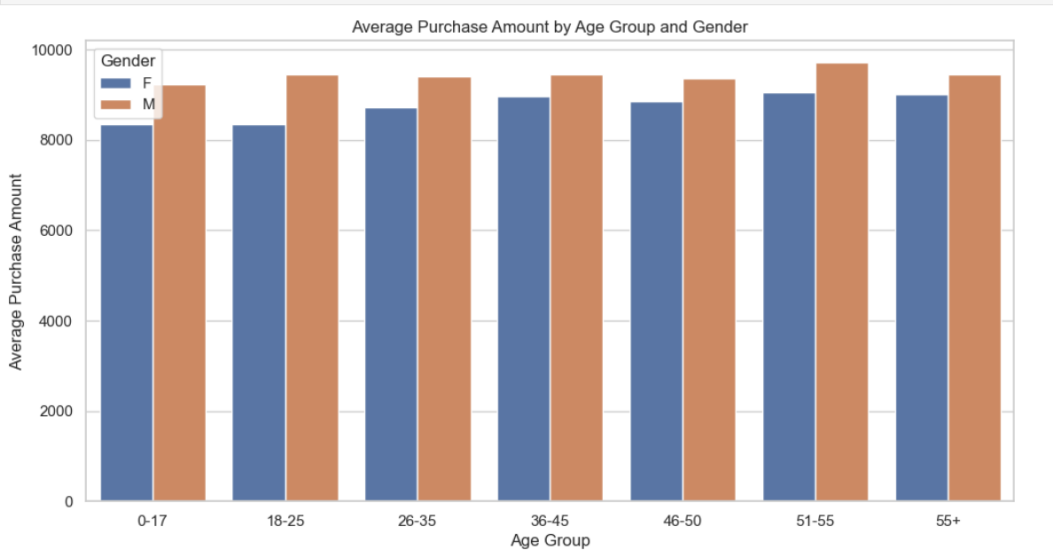
Cross-Promotion: Consider cross-promoting Product Categories 1, 5, and 8 with Product Categories 6 and 7 to attract a wider range of customers.

4. Average Purchase Amount By Age Group And Gender.

Code:

```
plt.figure(figsize=(12, 6))
sns.barplot(x='Age', y='Purchase', hue='Gender', data=age_gender_stats)
plt.title('Average Purchase Amount by Age Group and Gender')
plt.xlabel('Age Group')
plt.ylabel('Average Purchase Amount')
plt.legend(title='Gender')
plt.show()
```

Output:



Insights:

0-17 Age Group:
Females in this age group tend to have a significantly higher average purchase amount compared to males. This could be due to factors like fashion purchases or personal care items.

8-25 Age Group:
Both genders in this age group have relatively similar average purchase amounts. This suggests that gender might not be a strong predictor of spending habits in this age group.

26-35 Age Group:
Females in this age group have a slightly higher average purchase amount compared to males. This could be attributed to factors like household expenses or family responsibilities.

36-45 Age Group:
Similar to the 26-35 age group, females in this age group tend to have a slightly higher average purchase amount. This trend continues to hold true.

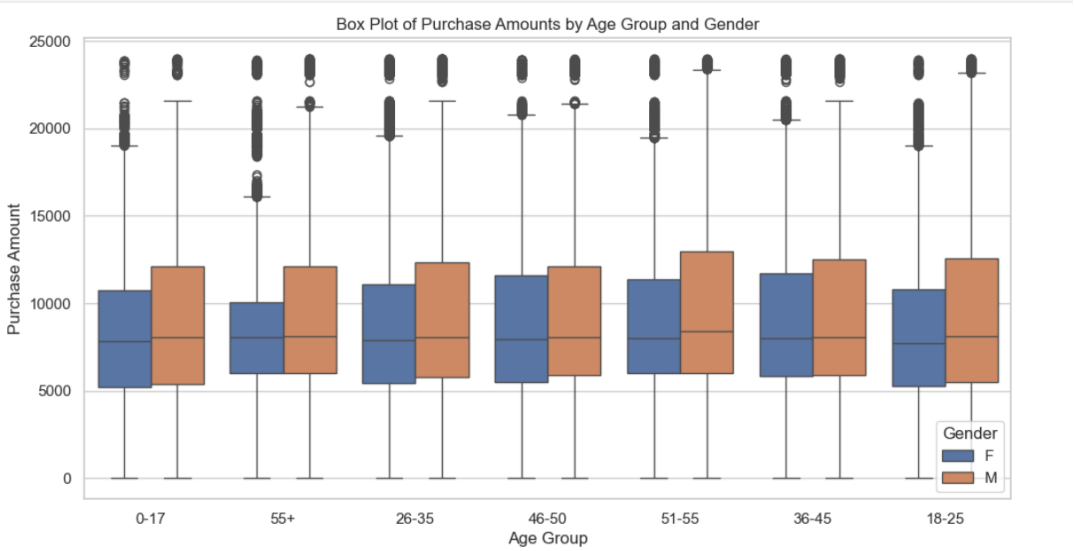
46-50 Age Group:
Females in this age group have a notably higher average purchase amount compared to males. This could be due to factors like established careers, homeownership, and family expenses.

51-55 Age Group:
Females in this age group have a slightly higher average purchase amount compared to males. This trend is consistent with previous age groups.

55+ Age Group:
Both genders in this age group have relatively similar average purchase amounts. This suggests that gender might not be a strong predictor of spending habits in this age group, similar to the 18-25 age group.

5.Box Plot Of Purchase Amount by Age Group and Gender

Code:



Insights:

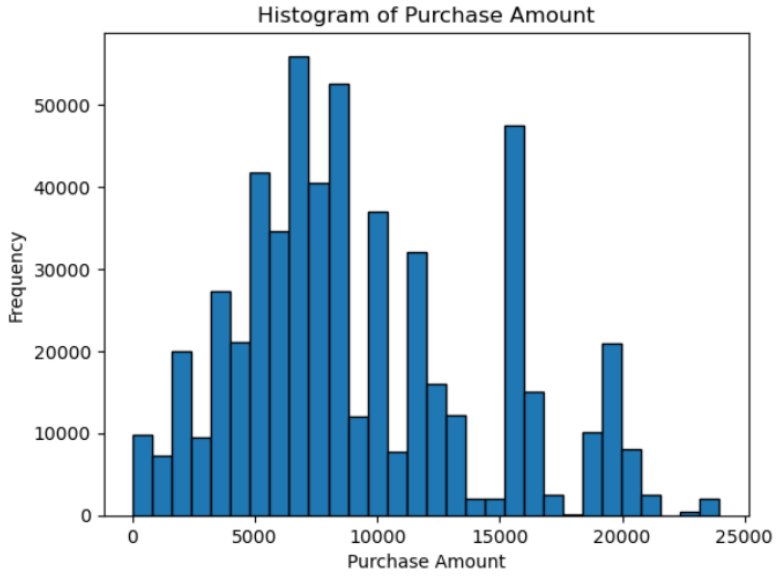
Outliers: There are several outliers in the data, indicating some very high purchase amounts.
Gender Differences: While there are some variations across age groups, overall, there doesn't seem to be a significant difference in purchase amounts between genders.
Age Group Variations: There are noticeable differences in the distribution of purchase amounts across different age groups.
Age Group Specific Insights:
0-17 Age Group: This group has a lower median purchase amount compared to other age groups.There are some outliers, indicating a few very high purchases within this group.
18-25 Age Group: This group has a similar distribution to the 0-17 age group, with a lower median and some outliers.
26-35 Age Group: This group has a higher median purchase amount compared to the younger age groups. The distribution is more spread out, indicating a wider range of purchase amounts.
36-45 Age Group: This group has a similar distribution to the 26-35 age group.
46-50 Age Group: This group has a slightly lower median purchase amount compared to the previous age group.
51-55 Age Group: This group has a similar distribution to the 46-50 age group.
55+ Age Group: This group has a higher median purchase amount compared to the previous age group.

6. Histogram For Purchase Amount.

Code:

```
plt.hist(data['Purchase'], bins=30, edgecolor='black')
plt.title('Histogram of Purchase Amount')
plt.xlabel('Purchase Amount')
plt.ylabel('Frequency')
plt.show()
```

Output:



Insights:

Right-skewed: The histogram shows a right-skewed distribution, indicating that most purchases are relatively small, while a few larger purchases significantly influence the overall distribution.

Peak Around 5000-7000: The highest frequency of purchases occurs between 5000 and 7000 units. This suggests that a significant portion of customers make purchases within this range.

Long Tail: The long tail towards the right side of the histogram indicates that there are a significant number of higher-value purchases.

Recommendations:

Marketing Strategies:

Targeted Promotions: Offer targeted promotions and discounts to encourage larger purchases from customers who typically make smaller purchases.

Customer Segmentation: Identify high-value customers and tailor marketing strategies to retain and upsell to them.

Inventory Management:

Stock Optimization: Optimize inventory levels to meet the demand for both high-volume, low-value items and low-volume, high-value items.

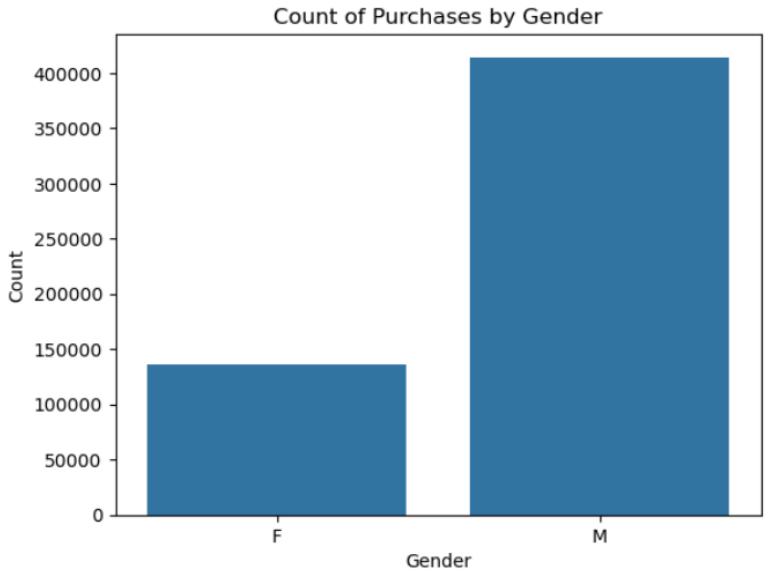
Supply Chain Efficiency: Ensure efficient supply chain operations to meet the demand for both ends of the purchase amount spectrum.

7. Count of Purchase By Gender:

Code:

```
sns.countplot(data=data, x='Gender')
plt.title('Count of Purchases by Gender')
plt.xlabel('Gender')
plt.ylabel('Count')
plt.show()
```

Output:



Insights:

Male Dominance: The chart clearly shows that a significantly larger number of purchases were made by male customers compared to female customers.

Recommendations:

Marketing Strategy: Walmart might need to focus its marketing efforts more on female customers to further capitalize on this trend.

Customer Retention Strategy: The lower count of purchases by female customers may indicate an opportunity to increase engagement with this demographic. Walmart could conduct surveys or focus groups to understand the shopping needs of female customers better and design campaigns to increase their purchase counts.

Product Variety and Marketing Campaigns: Walmart could consider expanding its range of products that appeal to female customers or create personalized marketing content aimed at increasing interest in specific categories.

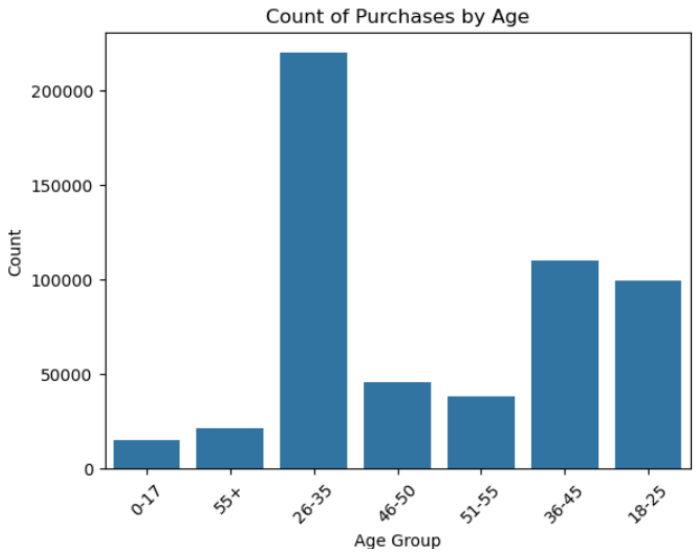
Targeted Digital Marketing: Walmart's digital marketing team could use this insight to create gender-specific ads on social media and digital platforms, reaching male customers with products that are more aligned with their interests while promoting products for female customers to increase engagement.

8.Count of purchase by Age:

Code:

```
sns.countplot(data=data, x='Age')
plt.title('Count of Purchases by Age')
plt.xlabel('Age Group')
plt.ylabel('Count')
plt.xticks(rotation=45)
plt.show()
```

Output:



Insights:

Uneven Distribution: The chart shows an uneven distribution of purchases across different age groups.
26-35 Age Group: This age group has the highest number of purchases, indicating a strong customer base within this demographic.
0-17 Age Group: This age group has the lowest number of purchases, suggesting a smaller customer base or lower purchasing power.
55+ Age Group: This age group shows a significant number of purchases, indicating a considerable market segment.

Recommendations:

Customized Marketing for Different Age Groups: Walmart could implement targeted marketing campaigns that cater to each age group's specific needs and preferences. For instance, promotions for the 26-35 group could focus on convenience products or exclusive brands, while campaigns for the 55+ age group might highlight health and wellness items or seasonal discounts.

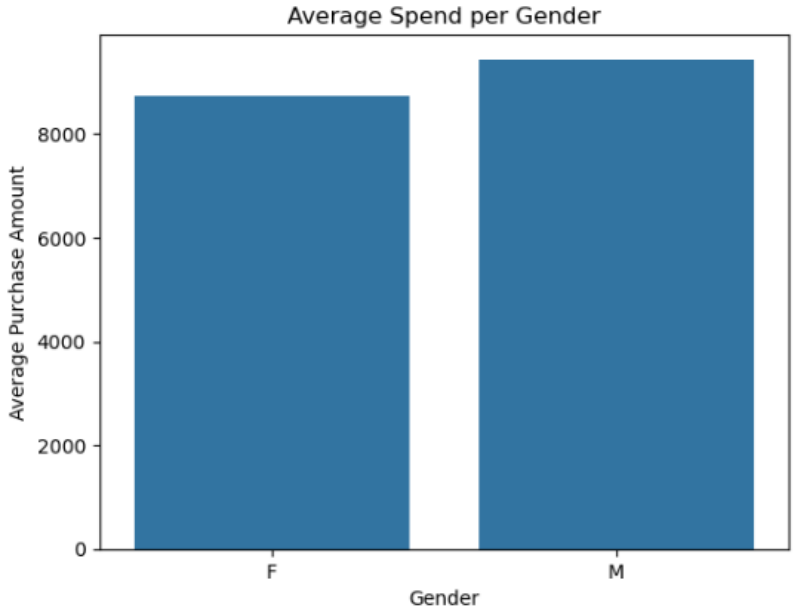
Family-Centric Marketing :Since the 0-17 age group likely influences family purchases, Walmart could consider family-centric promotions, bundling products for children alongside family essentials to encourage more spending in this category.

9.Average spend per Gender:

Code:

```
avg_spend_gender = data.groupby('Gender')['Purchase'].mean()
sns.barplot(x=avg_spend_gender.index, y=avg_spend_gender.values)
plt.title("Average Spend per Gender")
plt.ylabel("Average Purchase Amount")
plt.xlabel("Gender")
plt.show()
```

Output:



Insights:

Slightly Higher Spending by Males: Male customer's show a slightly higher average spend per transaction compared to female customers. Although the difference is small, it may hint at a slightly higher purchasing power or a tendency among male customers to buy higher-priced items on average.

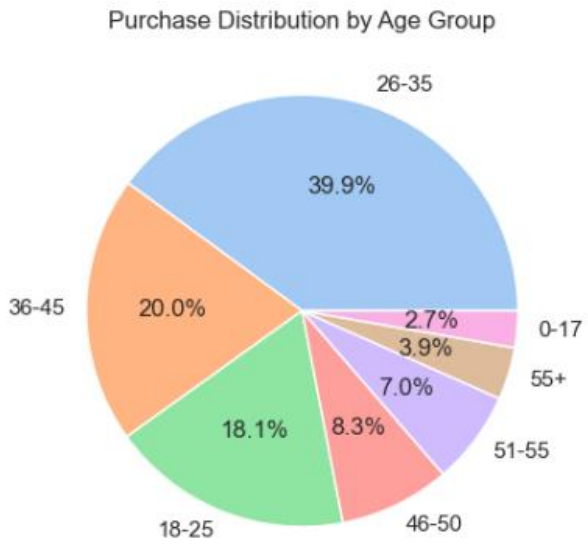
Balanced Market Strategy: Since the average spending between genders is nearly identical, Walmart's pricing and product strategies appear well-aligned with the needs of both male and female customers. Walmart might not need to differentiate its pricing strategy by gender but could instead focus on broader customer satisfaction initiatives.

10. Purchase Distribution by Age Group:

Code:

```
age_counts = data['Age'].value_counts()
plt.pie(age_counts, labels=age_counts.index, autopct='%1.1f%%', colors=sns.color_palette('pastel'))
plt.title("Purchase Distribution by Age Group")
plt.show()
```

Output:



Insights:

Majority Purchases by 26-35 Age Group: With 39.9% of purchases, this group is Walmart's most active shopping demographic. Targeted promotions for this group could drive further engagement.

Strong Engagement from 36-45 and 18-25 Groups: Together with the 26-35 group, these segments form a major portion of purchases, suggesting they're also key for targeted marketing efforts.

Lower Purchases Among Older Age Groups (46+): The 46+ age groups contribute less to overall purchases. Walmart might consider age-specific promotions or products to increase appeal among older shoppers.

Minimal Spend by the Youngest Group (0-17): With only 2.7% of purchases, this segment has limited purchasing power, so targeting families or parents may be more effective.

11. Confidence intervals for males and females.

Code:

```
def compute_ci(data, confidence=0.95):
    mean = np.mean(data)
    se = stats.sem(data)
    h = se * stats.t.ppf((1 + confidence) / 2., len(data)-1)
    return mean - h, mean + h

# Gender spending analysis
female_data = data[data['Gender'] == 'F']['Purchase']
male_data = data[data['Gender'] == 'M']['Purchase']

# Confidence Intervals for entire dataset
female_ci = compute_ci(female_data)
male_ci = compute_ci(male_data)

print(f'Confidence Interval for Female: {female_ci}')
print(f'Confidence Interval for Male: {male_ci}')
```

Output:

```
Confidence Interval for Female: (8709.21132117373, 8759.92020913722)
Confidence Interval for Male: (9422.019402055814, 9453.032678888716)
```

Insights:

Confidence Interval Interpretation:
Confidence Interval for Female: (8709.21, 8759.92)
Confidence Interval for Male: (9422.02, 9453.03)
The confidence intervals represent the range within which we can be 95% confident that the true average spending of females and males lies.
The confidence intervals for female spending (8709.21 to 8759.92) and male spending (9422.02 to 9453.03) do **not overlap**.
This indicates that there is a statistically significant difference between the average spending of females and males. Specifically, males tend to spend more on average than females in this dataset.

Recommendations:

Targeted Marketing Strategies: Since males have a higher average spending, Walmart could consider implementing targeted marketing campaigns aimed at male consumers. This could include promotional offers on products that are more likely to appeal to male shoppers.

Consumer Engagement: Walmart could explore creating loyalty programs or special offers that cater specifically to male customers, encouraging them to increase their purchase amounts.

12. Confidence intervals for Age Groups:

Code:

```
# Compute confidence intervals for each age group
age_groups = data['Age'].unique()
age_cis = {}

for age in age_groups:
    age_data = data[data['Age'] == age]['Purchase']
    age_cis[age] = compute_ci(age_data)

for age, ci in age_cis.items():
    print(f'Confidence Interval for Age Group {age}: {ci}')
```

Output:

```
Confidence Interval for Age Group 0-17: (8851.941436361221, 9014.987844528727)
Confidence Interval for Age Group 55+: (9269.295063935433, 9403.265854963376)
Confidence Interval for Age Group 26-35: (9231.733560884022, 9273.647704855754)
Confidence Interval for Age Group 46-50: (9163.08393647555, 9254.167458461105)
Confidence Interval for Age Group 51-55: (9483.989875153999, 9585.626186766473)
Confidence Interval for Age Group 36-45: (9301.669084404875, 9361.032305430872)
Confidence Interval for Age Group 18-25: (9138.40756914702, 9200.919643375557)
```

0-17 Age Group:
Insight: This age group has the lowest average purchase amount. This suggests limited purchasing power or fewer shopping needs, potentially due to their dependence on guardians for shopping.
18-25 Age Group:
Insight: This group's purchase behavior indicates slightly higher spending than the 0-17 group but still relatively low compared to older groups, possibly due to lower disposable income in early adulthood.
26-35 Age Group:
Insight: This age group shows a marked increase in average purchases. This group likely includes working adults with stable incomes, making them a significant consumer segment for Walmart.
36-45 Age Group:
Insight: The spending of this age group is relatively high, indicating strong purchasing power. This is often a demographic with established careers and families, contributing to higher spending levels.
46-50 Age Group:
Insight: Slightly lower than the 36-45 group, but still a considerable amount of spending. This may indicate stabilizing expenses or focused spending on essentials as family responsibilities may be established.
51-55 Age Group:
Insight: This age group has the highest average purchases, potentially indicating peak disposable income and stability in life, with fewer financial constraints compared to younger groups.
55+ Age Group:
Insight: Although slightly lower than the 51-55 age group, spending remains high, possibly due to increased focus on lifestyle and health needs.

Recommendations:

Targeted Marketing: Walmart could use these insights to direct marketing and product placements towards high-spending age groups, especially focusing on the 26-55 age range.
Product Offering Optimization: For younger customers, products with lower price points and trendy appeal could increase engagement, while for older demographics, higher-value, lifestyle, or health-focused products may be more effective.
Tailored Promotions: Providing age-specific promotions based on these confidence intervals could maximize Walmart's revenue by catering to the unique spending capacities and product preferences of each age group.

13. Average purchase behavior between married and unmarried customers.

Code:

```
married_data = data[data['Marital_Status'] == 1]['Purchase']
unmarried_data = data[data['Marital_Status'] == 0]['Purchase']

married_ci = compute_ci(married_data)
unmarried_ci = compute_ci(unmarried_data)

print(f'Confidence Interval for Married: {married_ci}')
print(f'Confidence Interval for Unmarried: {unmarried_ci}')
```

Insights:

Married Customers: Married customers average purchases fall within this range, showing consistent spending levels. The relatively high purchase average may be due to family needs or lifestyle factors that typically accompany marriage, such as buying in larger quantities or for multiple household members.

Unmarried Customers: Unmarried customers have a slightly higher average purchase amount, although the difference is minimal. This could reflect spending that might focus on personal needs or lifestyle purchases, rather than family-oriented ones, but overall, the differences in purchasing power between married and unmarried groups are not substantial.

Narrow Difference: The close range of these confidence intervals suggests there is not a significant variation in purchase behavior solely based on marital status. Both married and unmarried customers display a similar average spend.

High Consistency: The minimal difference indicates that marital status alone may not be a strong differentiator in spending patterns among Walmart's customers. This suggests that factors other than marital status (such as age, income, or specific product interest) may be more influential in driving purchasing behavior.

Recommendations:

Unified Marketing Approach: Since both groups show similar spending habits, Walmart could adopt a unified approach in promotions and product offerings, without needing to heavily differentiate based on marital status.

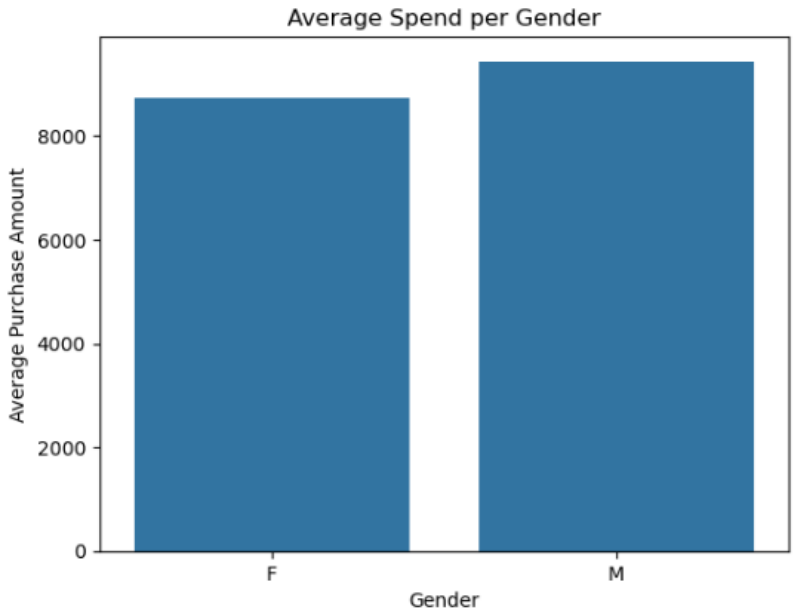
Further Segmentation: To deepen insights, Walmart may consider focusing on intersecting factors like age or income within marital status segments. For example, married customers within certain age brackets or income levels may show different spending patterns than their unmarried counterparts.

Product Targeting: While marital status alone may not heavily influence purchase behavior, Walmart could still explore offering bundled products for married customers (e.g., family packs) and lifestyle-oriented products for single customers, based on assumed lifestyle needs rather than spending power.

QnA

Q1.Are women spending more money per transaction than men? Why or Why not?

Ans:



The chart shows the average spend per transaction for both male and female customers. The values are relatively close, with male customers spending slightly more than female customers on average. This small difference suggests that, on average, men tend to spend marginally more per transaction than women, but the difference is not substantial.

Reasons:

Customer Behavior Patterns: The observed small difference might be due to slight variations in purchasing habits between male and female customers, but overall spending behaviors are quite similar.

Product Preferences and Necessities: Since the average spending does not show a large difference, it may indicate that both genders are purchasing products within a similar price range or that their purchasing needs are relatively comparable in this context.

In conclusion, while men have a marginally higher average spend per transaction, the difference is not significant enough to suggest a major disparity in spending between genders.

Q2.Confidence intervals and distribution of the mean of the expenses by female and male customers.
Ans:

```
Confidence Interval for Female: (8709.21132117373, 8759.92020913722)
Confidence Interval for Male: (9422.019402055814, 9453.032678888716)
```

Confidence Interval Analysis

Female Customers: The confidence interval for the average spend by female customers is (8709.21, 8759.92). This interval is relatively narrow, showing high precision in the estimated average spending.

Male Customers: The confidence interval for the average spend by male customers is (9422.02, 9453.03), which is also narrow, indicating a precise estimate for male spending as well.

The intervals do not overlap, with male customers spending, on average, more per transaction than female customers. This result is statistically significant, as the non-overlapping intervals indicate a meaningful difference between male and female spending behaviors.

Distribution of the Mean of Expenses

Higher Spending by Males: Since the confidence interval for male spending is consistently higher than that for females, it suggests that male customers generally spend more per transaction at Walmart compared to female customers.

Implications of CLT: According to the Central Limit Theorem (CLT), as we increase the sample size, the distribution of the mean will approach a normal distribution. Therefore, if we were to take repeated samples of spending for male and female customers, the means would remain stable around these intervals due to the large sample size

Q3. Are confidence intervals of average male and female spending overlapping? How can Walmart leverage this conclusion to make changes or improvements?
Ans:

```
Do confidence intervals overlap? No
```

No, the confidence intervals for average spending by male and female customers do not overlap. This non-overlap suggests a statistically significant difference in average spending between the two groups, with male customers spending more per transaction on average than female customers.

How Can Walmart Leverage This Conclusion to Make Changes or Improvements?

Walmart can leverage these insights to make strategic decisions in several areas:

1. **Targeted Marketing and Promotions**:
   - Since male customers tend to spend more, Walmart could design targeted marketing campaigns to further engage this high-spending demographic. For example, promotions on higher-end or premium products could be focused on male customers to encourage even higher spending.
   - Walmart could also investigate which products or categories are driving higher spending among male customers and highlight these in promotions aimed at females to encourage increased spending in this demographic.
2. **Product Placement and Merchandising**:
   - Understanding that male customers are willing to spend more per transaction could guide Walmart in product placement decisions, such as positioning higher-ticket items in areas where male customers are likely to shop.
   - Walmart could also explore diversifying the product mix to appeal to female customers, potentially adding products that align with female customer preferences to encourage higher spending.

Final Conclusion:

The analysis of Walmart's customer data highlights clear patterns in purchasing behavior across demographics, revealing key insights that can guide Walmart's marketing and sales strategies.

1. **Popular Product Categories**: Product Categories 1, 5, and 8 are the most purchased across all age groups, indicating they have universal appeal. Younger age groups (0-17 and 18-25) also show an interest in Categories 16 and 18, suggesting Walmart can refine its product offerings to cater to the unique preferences of each age segment.
2. **Gender Spending Patterns**: Although there is a slight difference in spending between genders (with males spending marginally more), the difference is not substantial. This balanced spending suggests Walmart's product and pricing strategies are generally well-aligned with the needs of both male and female customers.
3. **Age Group Purchasing Behavior**: The 26-35 age group has the highest percentage of total purchases, followed closely by the 18-25 and 36-45 groups. This shows that Walmart's core customers are working-age adults. Meanwhile, older age groups (46+) spend less overall, presenting an opportunity to introduce age-specific promotions or product lines to boost engagement among these groups.
4. **Confidence Intervals and Statistical Significance**: The confidence intervals for male and female average spending do not overlap, indicating that males generally spend more on average than females. Walmart could consider gender-targeted promotions to capitalize on these differences in purchasing power.
5. **Marital Status**: Both married and unmarried customers have similar average spending, indicating that marital status alone is not a strong predictor of purchase behavior. Other factors, such as age and specific product interests, may drive customer spending more significantly.
6. **Distribution of Spending**: The spending distribution shows a right-skewed pattern, with the majority of purchases falling between 5,000 and 7,000 units. A long tail of higher-value purchases suggests the presence of high-spending customers, which could represent an opportunity for loyalty programs or premium offerings.

Final Recommendations for Walmart:

1. **Targeted Promotions for Core Age Groups**: Focus on the 26-35, 18-25, and 36-45 age groups with targeted promotions and discounts, as these groups make up the bulk of Walmart's customer base.
2. **Engage Older Age Groups (46+)**: To increase engagement among older age groups, consider age-specific promotions, loyalty rewards, or product lines that cater to their lifestyle needs.
3. **Expand Product Categories with Age-Specific Appeal**: Given the popularity of Categories 1, 5, and 8, Walmart should continue to stock these categories while expanding offerings in Categories 16 and 18 to appeal to younger customers.
4. **Gender-Based Marketing**: Since males generally spend more per transaction, Walmart could explore targeted marketing strategies that highlight higher-value items for male customers, while keeping promotions balanced to appeal to all genders.
5. **Loyalty and Rewards Programs**: The presence of high-value purchases suggests that Walmart could benefit from loyalty programs aimed at rewarding high-spending customers and encouraging repeat purchases.
6. **Family-Focused Campaigns**: With married customers showing consistent spending patterns, Walmart could introduce family-centric promotions or discounts on bulk purchases to appeal to this demographic.

By leveraging these insights, Walmart can tailor its marketing and product strategies to align more closely with customer preferences, ultimately driving greater customer satisfaction and profitability.

Jupyter Notebook Analysis

For a detailed view of the full analysis, including code, visualizations, and insights, please refer to the complete Jupyter notebook available in the PDF format. The notebook documents each step of the analysis process, from data exploration to the final recommendations.

You can access the Jupyter notebook PDF through the **following link**:

JupyterNotebook