**[Jupyter Notebook Link](https://drive.google.com/file/d/1lPWIvkAYxUr4kyHGbKlK7aiz8CNX56jH/view)**

**Initial Insights**

* Shape is (550068,10).
* Columns are ['User\_ID', 'Product\_ID', 'Gender', 'Age','Occupation','City\_Category','Stay\_In\_Current\_City\_Years', 'Marital\_Status', 'Product\_Category','Purchase'].
* There are no nulls.
* There 3631 unique product ID’s.
* There are 7 age groups.
* Cities are categorised into A, B, C.
* Marital Status is Boolean value.
* There are 3631 products.
* There are 5891 user ID’s.
* There are 20 product categories.
* P00265242 is the most sold Product ID.
* Standard deviation is high thus Spread of purchase is more.
* There are 20 different types of Occupations.
* 73% users are between (18,45) age.
* Male count is significant with 75%.
* Male purchasing power is also significant with 76%.
* Nearly 78% of purchase is contributed by ages between (18,45)
* User ID with 1001680 has highest frequency of purchasing and 1004277 has highest total purchase.
* Occupations with value 0,4,7 has contributed nearly 37%.

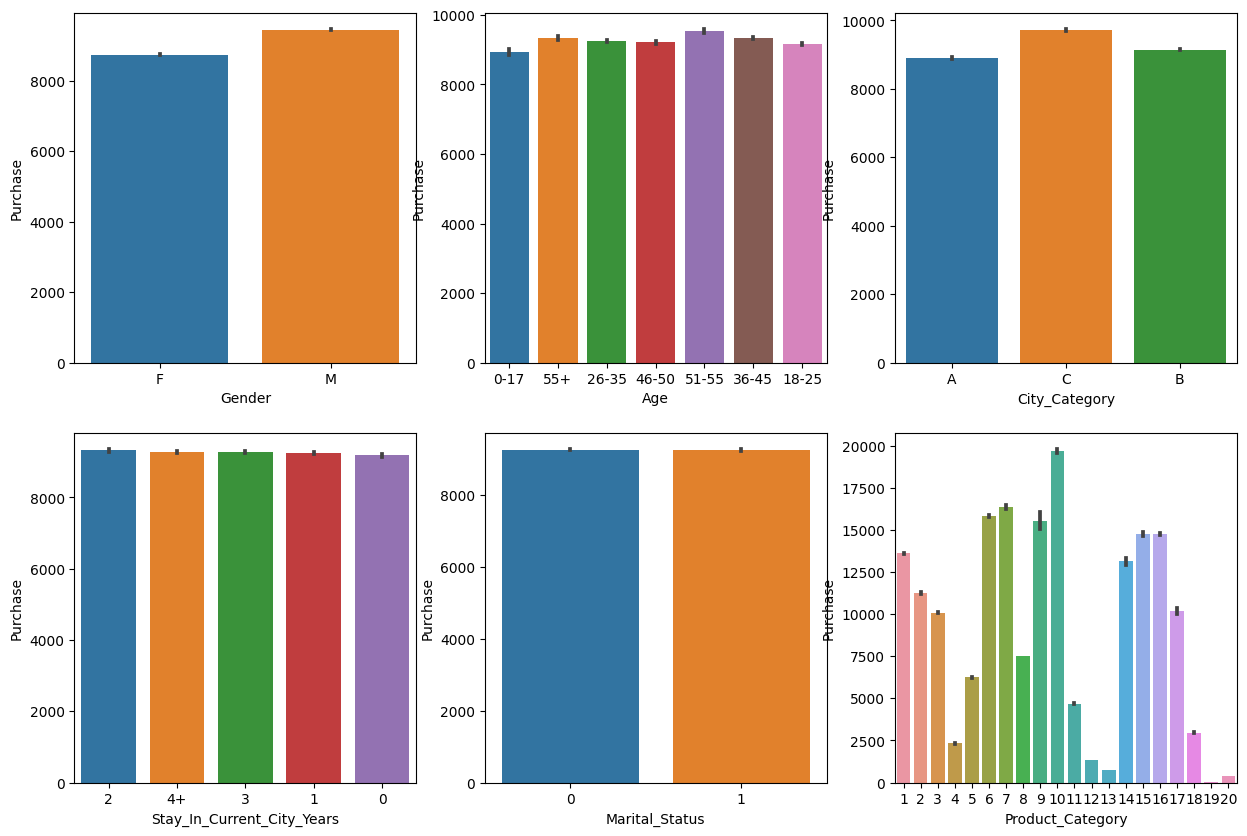
|  |  |  |
| --- | --- | --- |
| Mean | Median | Mean - Median |
| 9263.96 | 8047 | 1216.96 |

**Box Plots on Purchase Amount**

|  |  |
| --- | --- |
|  |  |
|  |  |

* There are significant outliers in all four Categories.

**Bar Charts to show Spread of Data**

****

**Heatmaps**

|  |  |
| --- | --- |
|  |  |

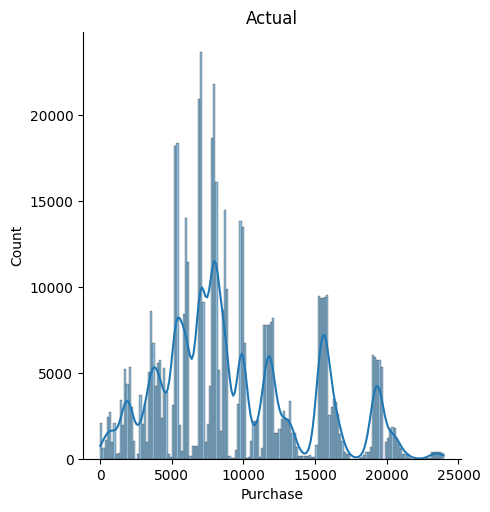
**NUMERICAL ANALYSIS**

Average

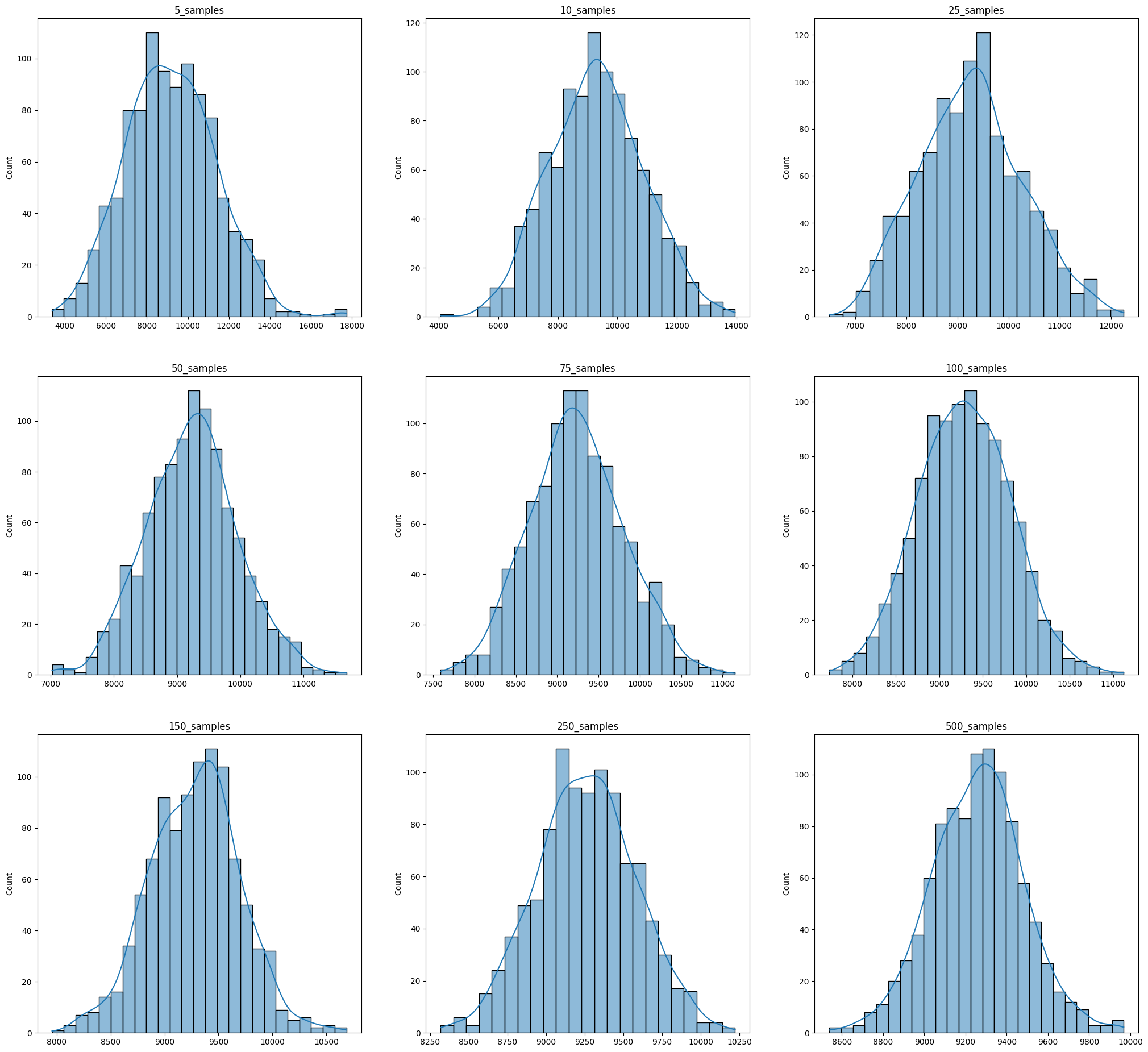
|  |  |  |
| --- | --- | --- |
| Column | Highest | Lowest |
| Gender | Male: 9437.53 | Female: 8734.57 |
| Age | (51–55): 9534.81 | (0-17): 8933.46 |
| City Type | 'C': 9719.92 | 'A': 8911.94 |
| Marital Status | Un-Married: 9265 | Married: 9261 |
| Product category | ‘10’: 19675 | ‘19’: 37.04 |
|  |  |  |

**Central Limit Theorem**

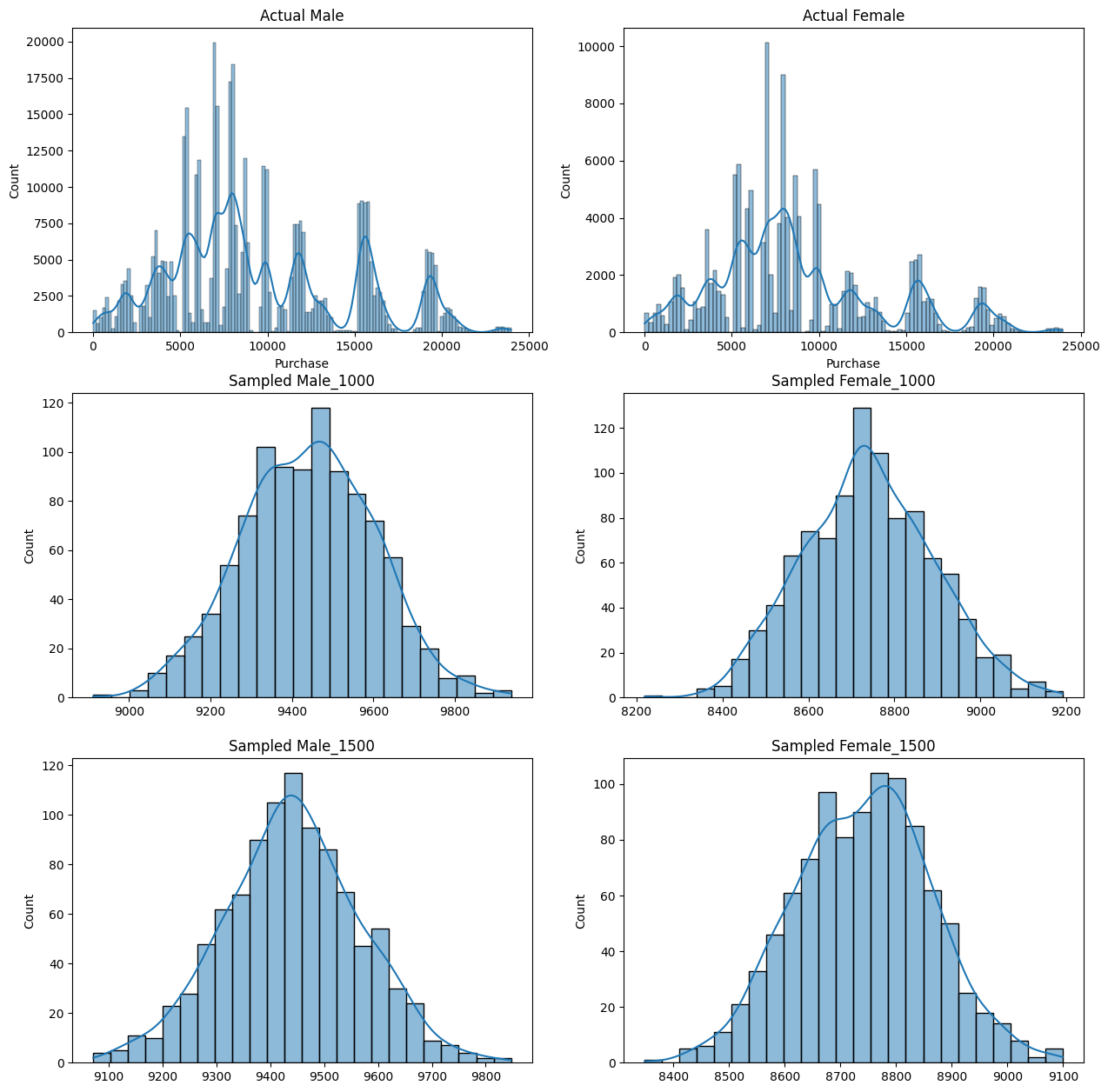
Purchase



Samples ran for 1000 times



**CLT of Male and Female**

* 1000 samples, 1500 samples
* 1000 iterations

**Confidence Interval**

**1000 Samples**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Male | Lower | Higher | | 90% | 9332.66 | 9539.03 | | 95% | 9303.41 | 9568.28 | | 99% | 9214.61 | 9657.10 | | |  |  |  | | --- | --- | --- | | Female | Lower | Higher | | 90% | 8634.24 | 8827.44 | | 95% | 8606.85 | 8854.82 | | 99% | 8523.72 | 8937.96 | |

**1500 Samples**

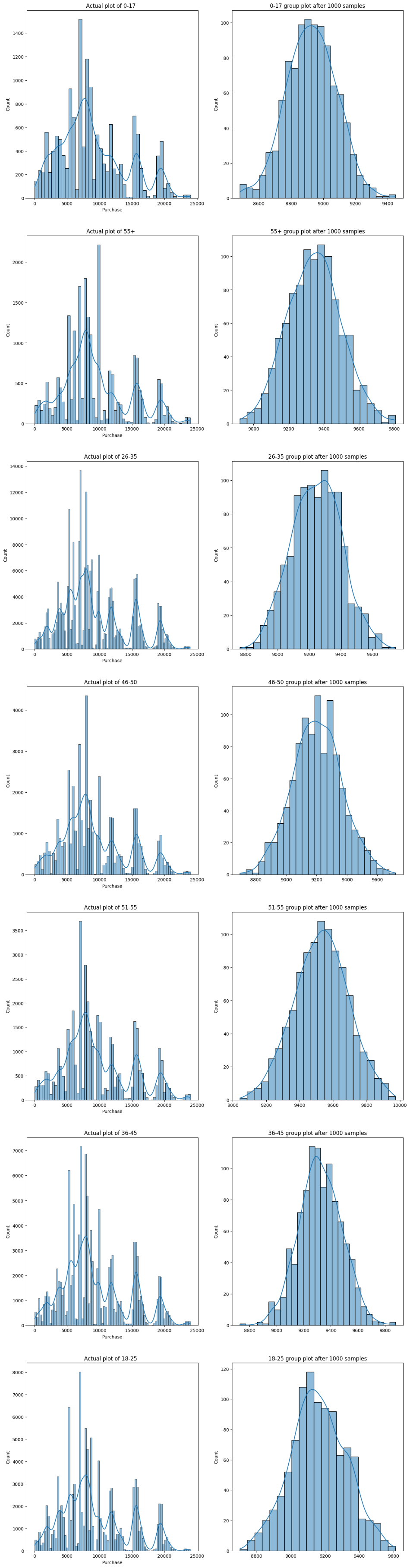
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Male | Lower | Higher | | 90% | 9352.22 | 9520.72 | | 95% | 9328.34 | 9544.60 | | 99% | 9255.83 | 9617.11 | | |  |  |  | | --- | --- | --- | | Female | Lower | Higher | | 90% | 8658.40 | 8816.15 | | 95% | 8636.04 | 8838.51 | | 99% | 8568.16 | 8906.39 | |

**1000 Samples**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Married | Lower | Higher | | 90% | 9151.02 | 9354.31 | | 95% | 9122.21 | 9383.16 | | 99% | 9034.72 | 9470.65 | | |  |  |  | | --- | --- | --- | | Un Married | Lower | Higher | | 90% | 9167.3 | 9371.04 | | 95% | 9138.42 | 9399.92 | | 99% | 9050.75 | 9487.59 | |

**1500 Samples**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  | | --- | --- | --- | | Married | Lower | Higher | | 90% | 9184.27 | 9350.28 | | 95% | 9160.74 | 9373.81 | | 99% | 9089.31 | 9445.24 | | |  |  |  | | --- | --- | --- | | Un Married | Lower | Higher | | 90% | 9185.8 | 9352.16 | | 95% | 9162.22 | 9375.74 | | 99% | 9090.64 | 9447.32 | |

**CLT ON AGE WITH 1000 Samples**

**Insights**

|  |  |
| --- | --- |
| 1 | Males have highest purchasing average |
| 2 | Product 10 has highest purchases and Product 1,5,8 have highest frequency in purchase. |
| 3 | The Width of bell curve of purchase is lowered with increased sample size. |
| 4 | Confidence interval is not overlapping in Males and Females |
| 5 | Confidence interval is overlapping in Married and Un Married. |
| 6 | Purchasing power of age 0-17 is less. |
| 7 | Purchasing power of age 51 – 55 is more |
| 8 | Un Married is spending more than married. |
| 9 | Actual Mean is similar to sampled mean for Gender, Marital-Status and Age groups. |
| 10 | People with 1 year stay contributed 35% of the Purchase. |
| 11 | 19, 20, 13 have lowest purchases. |
| 12 | There are only 1719 customers who appeared more than 100 times out of 5891 customers |

**Recommendations**

|  |  |
| --- | --- |
| **1** | Target marketing efforts towards males |
| **2** | Ensure products 1, 5, 8, and 10 are well-stocked, highlighted, and promoted to maximize sales potential |
| **3** | Continue to collect data to enhance the accuracy of your analysis and insights |
| **4** | Tailor marketing campaigns to each age group separately |
| **5** | Fine-tune marketing efforts to highlight specific benefits and preferences of both married and unmarried individuals |
| **6** | Design targeted promotions to attract customers in the 0-17 age range and offer age-appropriate discounts or product bundles |
| **7** | Capitalize on the higher purchasing power of the 51-55 age group through targeted promotions |
| **8** | Conduct further analysis to understand reasons behind this behavior and adjust marketing strategies accordingly |
| **9** | Monitor actual mean and sampled mean over time for potential shifts in customer behavior or preferences |
| **10** | Implement a customer loyalty program or exclusive offers to reward frequent customers and encourage repeat purchases |
| **11** | Investigate reasons behind low sales (pricing, quality, promotion) and consider targeted advertising, discounts, or product improvement |

**Answering:**

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

* No, men spend more money.
* Maybe men only purchase goods on behalf of family.
* Women are less in numbers.
* Maybe products are not attracting female customers.

**2.**Confidence intervals and distribution of the mean of the expenses by female and male customers?

1000 samples

|  |  |
| --- | --- |
| Male | Female |
| 90 % (9333.58, 9539.95)  95 % (9304.33, 9569.2)  99 % (9215.53, 9658.0) | 90 % (8633.83, 8827.03)  95 % (8606.45, 8854.41)  99 % (8523.31, 8937.55) |

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

* No, they are not overlapping at all at any sample rate.
* Need to focus more on female customers.

4. Results when the same activity is performed for Married vs Unmarried?

|  |  |
| --- | --- |
| Married | Un Married |
| for 90 % (9174.69, 9378.01)  for 95 % (9145.87, 9406.83)  for 99 % (9058.38, 9494.31) | for 90 % (9162.38, 9366.12)  for 95 % (9133.5, 9395.0)  for 99 % (9045.83, 9482.67) |

5. Results when the same activity is performed for Age? 1000 samples

|  |  |  |
| --- | --- | --- |
| Group | Lower | Higher |
| 0-17 | 8709.75 | 9153.86 |
| 55+ | 9118.08 | 9553.53 |
| 26-35 | 9031.28 | 9466.65 |
| 46-50 | 9002.31 | 9433.92 |
| 51-55 | 9309.93 | 9751.98 |
| 36-45 | 9114.13 | 9550.59 |
| 18-25 | 8943.17 | 9380.62 |