

BUSSINESS CASE: WALMART

Introduction

Walmart is an American multinational retail corporation that operates a chain of supercenters, discount departmental stores from the United States. Walmart has more than 100 million customers worldwide and is one of the key players in this section of business.

Business Problem

The Management team at Walmart Inc. wants to analyze the customer purchase behavior (especially, purchase amount) against the customers' gender and the various other factors to help the business make better decisions. They want to understand if the spending habits differ between male and female customers.

Approach

- Whole analysis is going to be Purchase amount centric.
- All-important attributes will be considered one by one and will see how average purchase amount will vary for that attribute.
- The analysis is being done on samples (i.e., out of 100 million customer bases nearly 550,068 have been provided) so sample average and population average will be calculated. If both are equal, then we will move further.
- All numerical parameters calculated will ultimately lead to Central Limit theorem and Confidence Interval.
- Every Insights and Recommendation will be backed by data and graph.

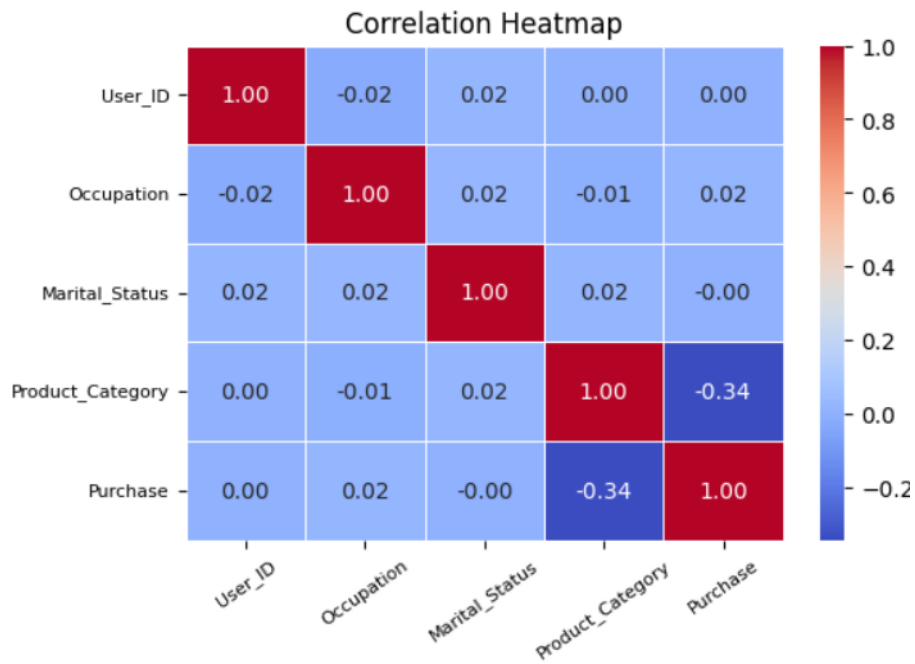
Non-Graphical Analysis of Sample Data:

- Shape of data= (550068 rows x 10 columns)
- Size of data = 5506800 in raw form
- Attributes of Data: User_Id, Product_Id, Gender, Age, Occupation, City_Category, Stay_In_Current_City_Years, Marital_Status, Product_category, Purchase.
- Any N/A or NaN value present: "Not found."
- Data Types initially present: 'Object', 'int64'
- Number of unique customers present: 5891
- Number of unique product present: 3631
- Top 5 Most purchased Product_Id: P00265242, P00025442, P00110742, P00112142 and P00057642 respectively.
- Number of Male and Female Customer in Sample: 4225 male and 1666 female.
- Number of Transactions made by each gender: 414259 by male and 135809 by female.
- Transactions done by each unmarried and married person: 324731 and 225337 respectively.
- Transactions made by each city: 147720 from City A, 171175 from City B and 231173 from city C.
- Maximum and Minimum transaction by age group: 219587 by age group (26-35) and 15102 by (0-17).
- Maximum and minimum purchase amount: 23961 and 12 respectively.

Final Insights:

- Above analysis and Insights below are derived from first look of the data.
- Shape of sample data denotes total number of transactions made which is 550068.
- Total 5891 unique customers are present of which 4225 are male and 1666 are female, they combined are responsible for 550068 transactions, more specifically 4225 males have done 414259 transactions and 1666 female have done 135809 transactions.
- Unmarried customers are doing more transactions.
- City C tops the chart in terms of transactions made with a staggering number of 231173.
- Age group (26-35) are more active customer base.

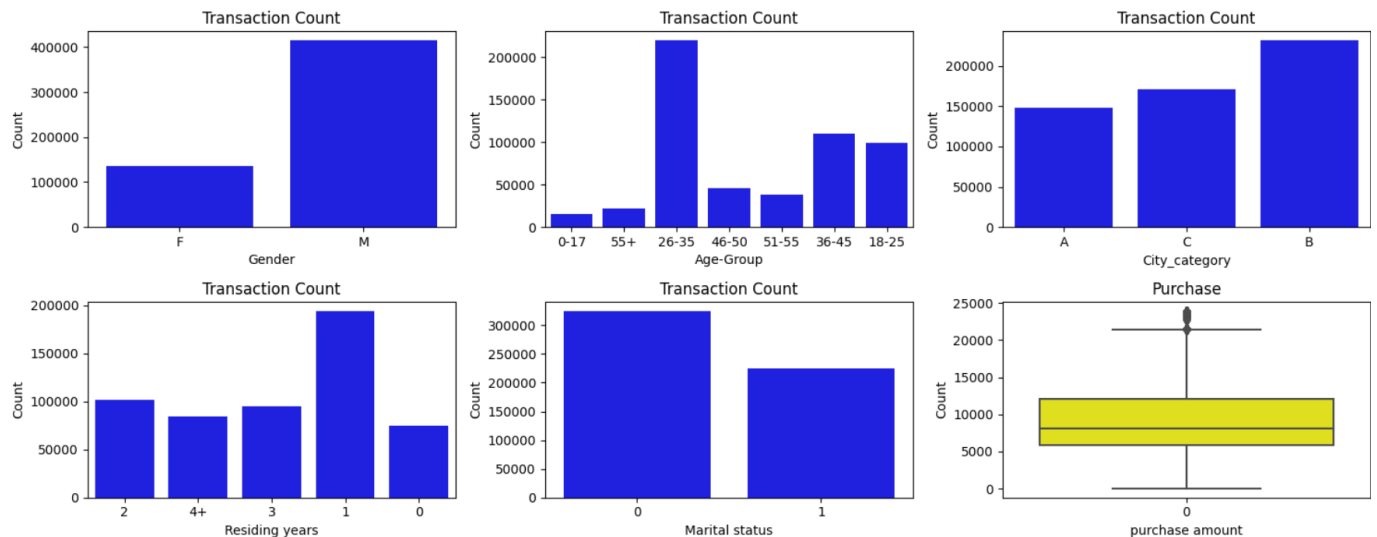
Correlation between different variables:



Final Insights

- There is no strong correlation between variables which could be considered.

Outlier detection:



Final Insights:

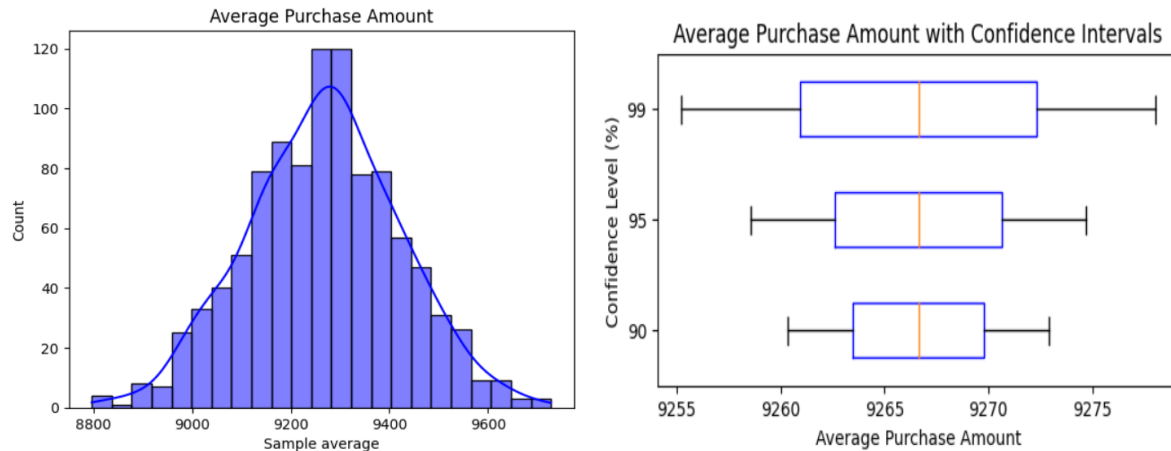
- For categorical data, in gender Male, in age group 26-35, In residing year 1 year is quite significant than the other ones.
- For numerical data, Purchase amount too have outliers present in their respective column.
- Outliers will not impact our further approach towards visualization as we are going to iterate randomly over data, and this nullify the effect of outlier.
- Best way to check if the effect of outlier has been nullified is to see if the graph after iteration will follow normal curve.

Before we Deep Dive:

There is extensive need for some calculations, assumption before moving ahead with Central limit theorem and Confidence Interval.

- Our Initial calculations and visualization will be for 90% confidence interval, 95% confidence interval, 99% confidence interval, but as we will move ahead, we will settle with 95% confidence interval.
- Z score for 90% is 1.2816, 1.6449 for 95% and 2.3263 for 99%.
- Blue and Yellow color will be used for visualization.
- Our approach will be like:
 - First calculate Population mean, then sample mean, see if they are close to each other, best is to check if sample mean graph follows normal distribution.
 - Sample mean will be calculated after 1000 iterations.
 - If above criteria are fulfilled, then calculate standard deviation and then move to calculation for confidence interval.
 - Insights and graphs will together follow our case study.

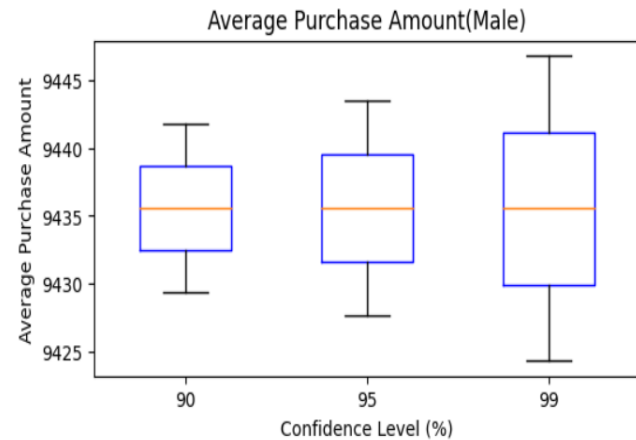
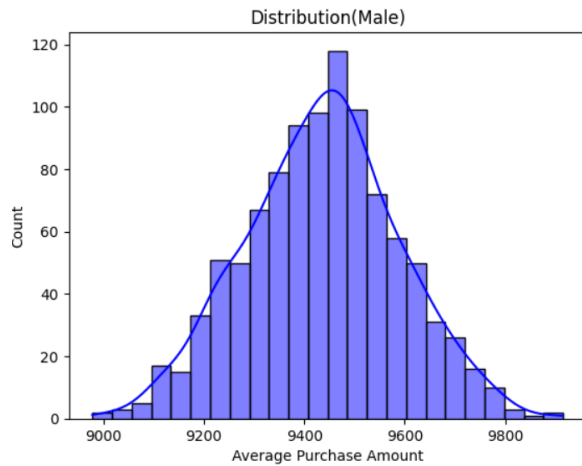
Purchase amount vs data



Final Insights:

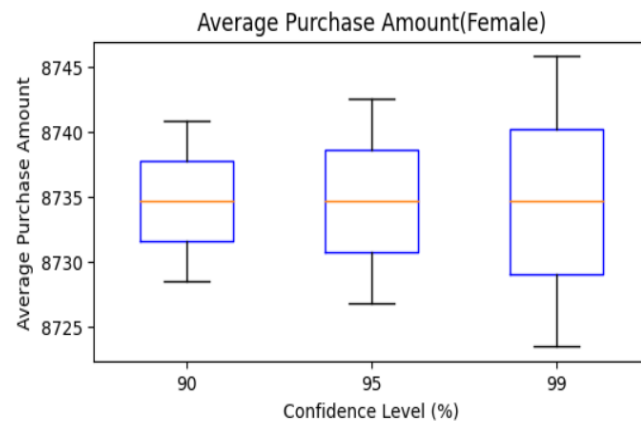
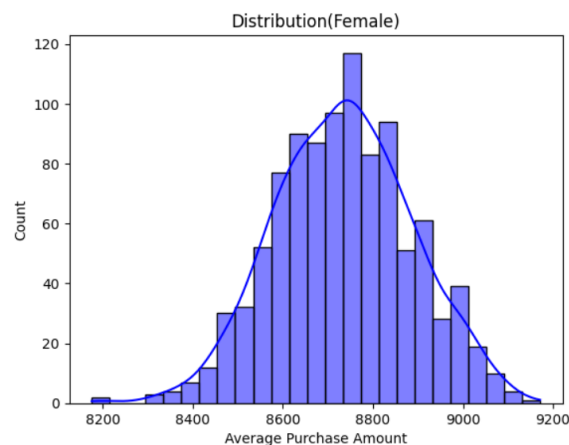
- Population average is 9263.968 and sample average is 9266.62, close ones.
- Sample average follows normal curve distribution.
- With 90% confidence, we can say average purchase amount is between 9260.35 and 9272.9.
- With 95% confidence, we can say average purchase amount is between 9258.35 and 9274.9.
- With 99% confidence, we can say average purchase amount is between 9255.35 and 9278.9.
- As the confidence interval increases, interval also increases, 95% is good number to comment.

Gender vs Purchase amount:



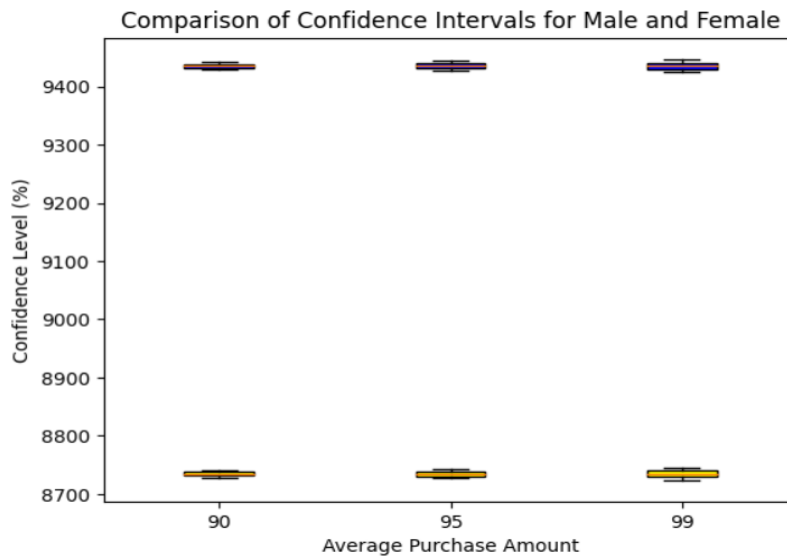
Final Insights(A):

- Male average purchase is 9437.524 and sample average is 9435.554, close ones.
- Sample average follows normal curve distribution.
- With 90% confidence, we can say average purchase amount is between 9429.37 and 9441.74.
- With 95% confidence, we can say average purchase amount is between 9427.62 and 9443.49.
- With 99% confidence, we can say average purchase amount is between 9424.33 and 9446.78.



Final Insights(B):

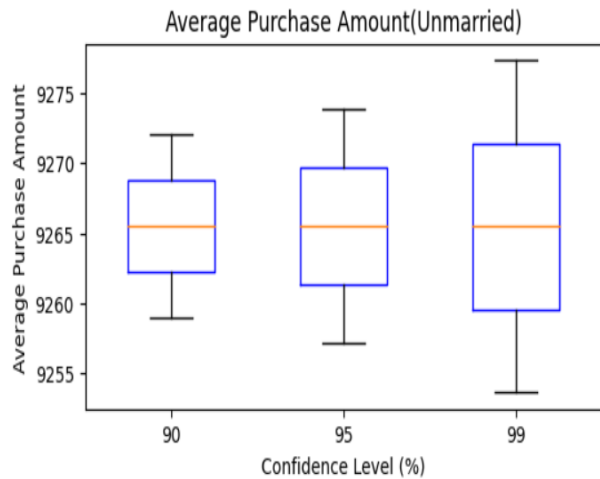
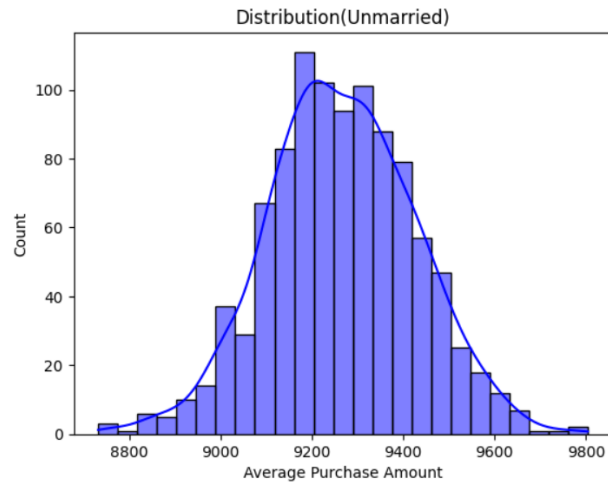
- Female average purchase is 8734.565 and sample average is 8734.636, nearly overlapping.
- Sample average follows normal curve distribution.
- With 90% confidence, we can say average purchase amount is between 8728.47 and 8740.8.
- With 95% confidence, we can say average purchase amount is between 8726.72 and 8742.55.
- With 99% confidence, we can say average purchase amount is between 8723.44 and 8745.83.



Final Insights(C):

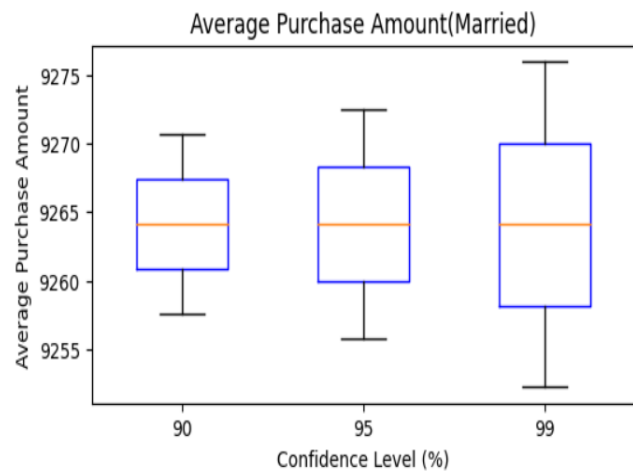
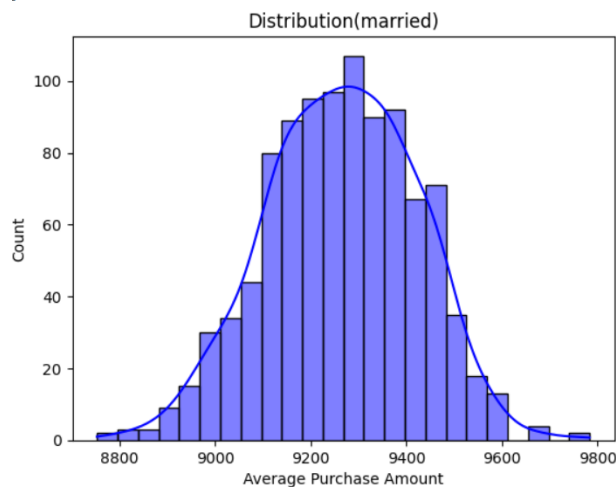
- Purchase power of male and female is significantly different. No overlap in the comparison plot confirms the statement.

Marital Status vs Purchase amount:



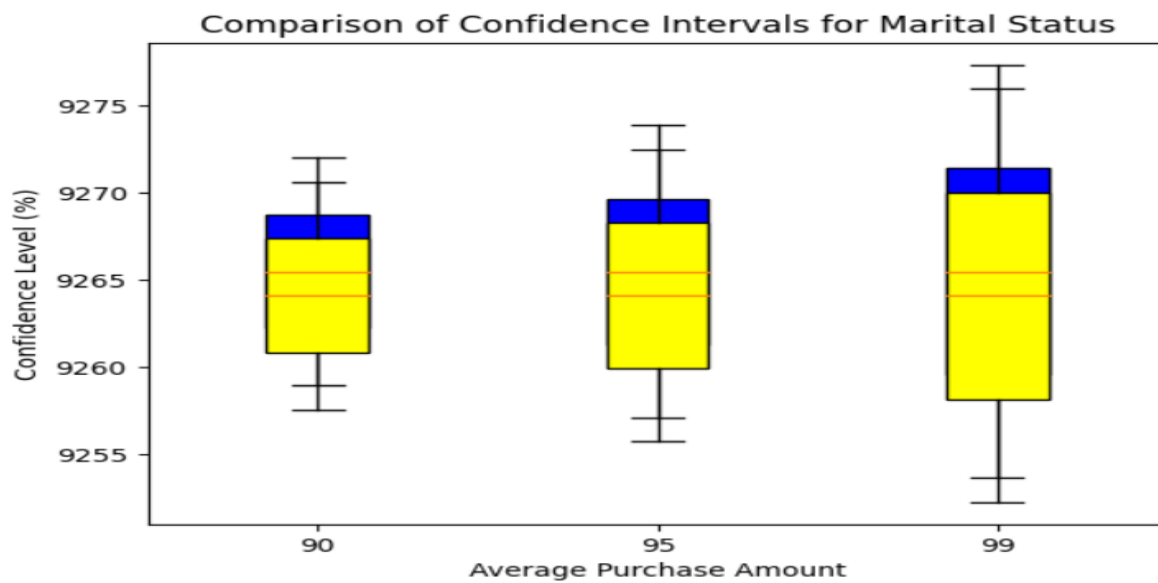
Final Insights(A):

- Unmarried customer average purchase is 9265.907 and sample average is 9565.488, nearly overlapping.
- Sample average follows normal curve distribution.
- With 90% confidence, we can say average purchase amount is between 9258.96 and 9272.01.
- With 95% confidence, we can say average purchase amount is between 9257.11 and 9273.87.
- With 99% confidence, we can say average purchase amount is between 9253.64 and 9277.4.



Final Insights(B):

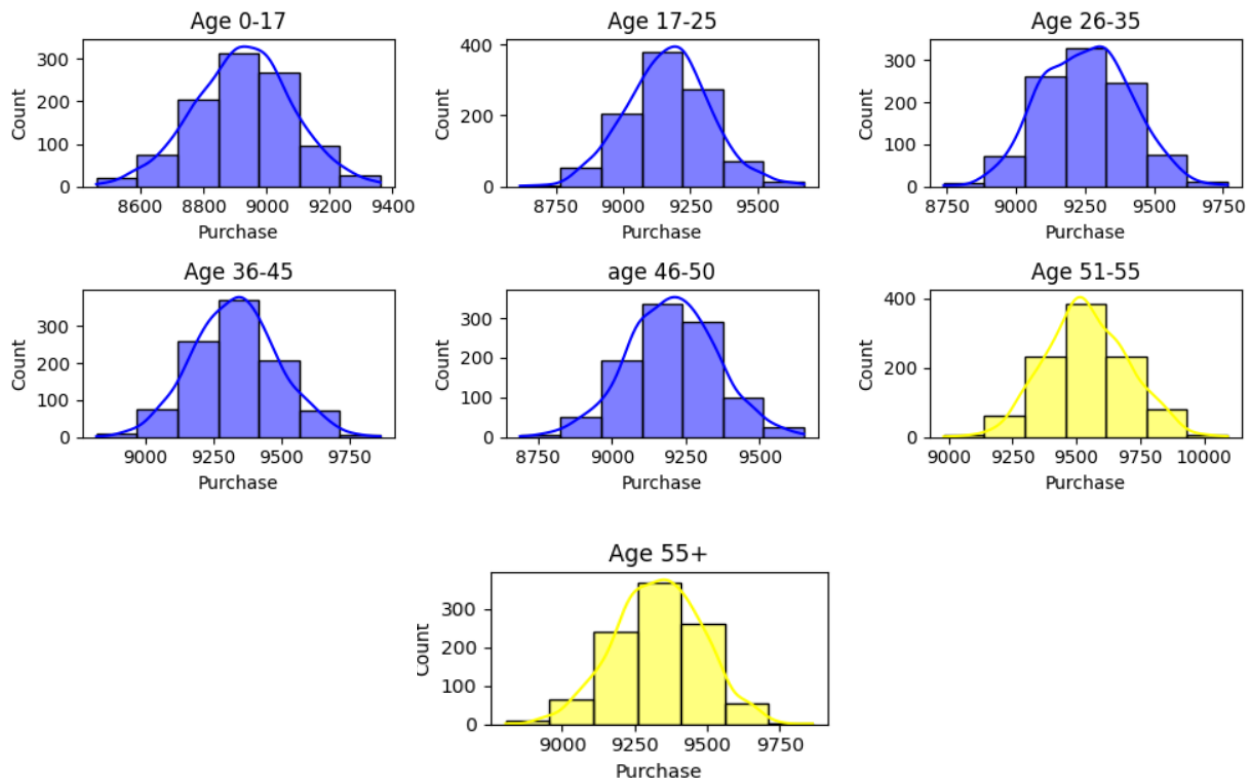
- Married customer average purchase is 9261.174 and sample average is 9264.109, quite close.
- Sample average follows normal curve distribution.
- With 90% confidence, we can say average purchase amount is between 9257.58 and 9270.64.
- With 95% confidence, we can say average purchase amount is between 9255.73 and 9272.49.
- With 99% confidence, we can say average purchase amount is between 9252.26 and 9275.96.



Final Insights(C):

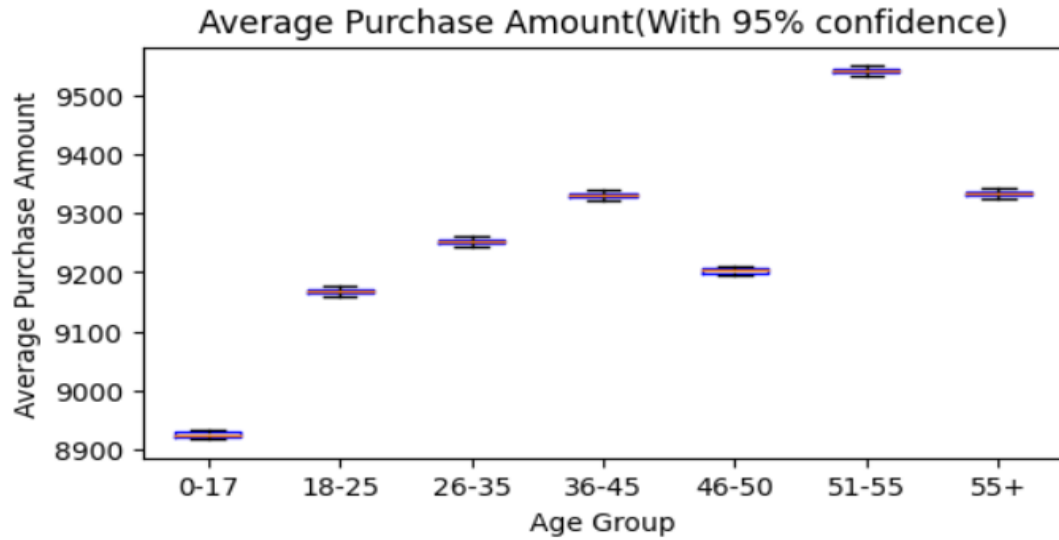
- Purchase capacity for married and unmarried customers are overlapping, so there is no direction when it comes to purchase.

Age Group vs Purchase amount:



Final Insights:

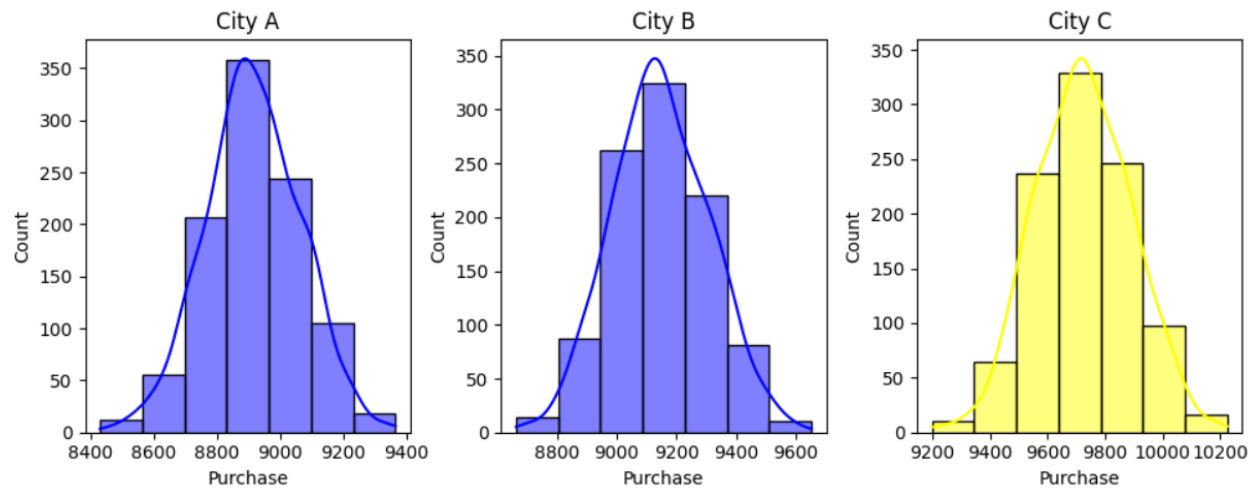
- Average purchase of customers aged (0-17) year is 8933.46 and sample average purchase is 8925.75, quite close.
- Average purchase of customers aged (18-25) year is 9169.66 and sample average purchase is 9167.86, quite close.
- Average purchase of customers aged (26-35) year is 9252.69 and sample average purchase is 9252.53, nearly overlapping.
- Average purchase of customers aged (36-45) year is 9331.35 and sample average purchase is 9330.54, quite close.
- Average purchase of customers aged (46-50) year is 9208.63 and sample average purchase is 9203.57, quite close.
- Average purchase of customers aged (51-55) year is 9534.81 and sample average purchase is 9542, close one.
- Average purchase of customers aged (55+) year is 9336.25 and sample average purchase is 9333.58, quite close.
- All sample average follows normal distribution.



Final Insights:

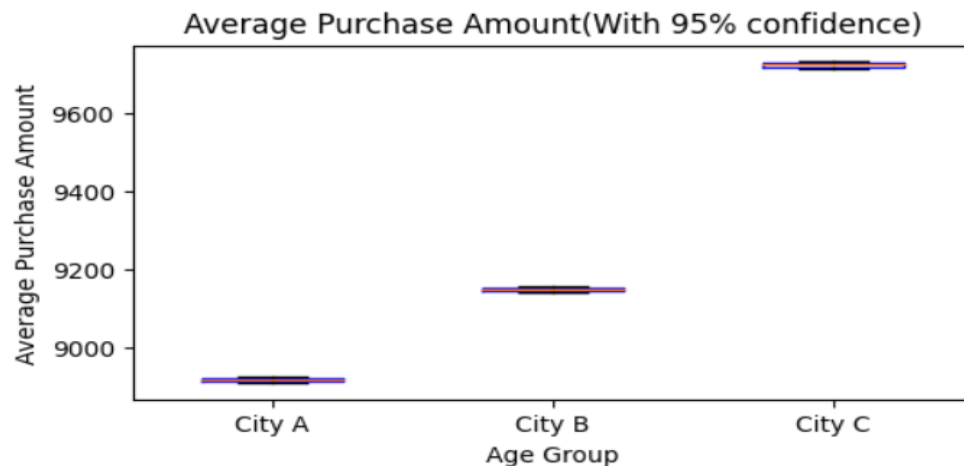
- With 95% confidence we can say, average purchase amount for customers aged 0-17 lies between 8917.67 And 8933.84.
- With 95% confidence we can say, average purchase amount for customers aged 18-25 lies between 9159.78 And 9175.95.
- With 95% confidence we can say, average purchase amount for customers aged 26-35 lies between 9244.45 And 9260.62.
- With 95% confidence we can say, average purchase amount for customers aged 36-45 lies between 9322.45 And 9338.62.
- With 95% confidence we can say, average purchase amount for customers aged 46-50 lies between 9195.48 And 9211.65.
- With 95% confidence we can say, average purchase amount for customers aged 51-55 lies between 9533.92 And 9550.08.
- With 95% confidence we can say, average purchase amount for customers aged 55+ lies between 9325.49 And 9341.66.

City vs Purchase Amount:



Final Insights(A):

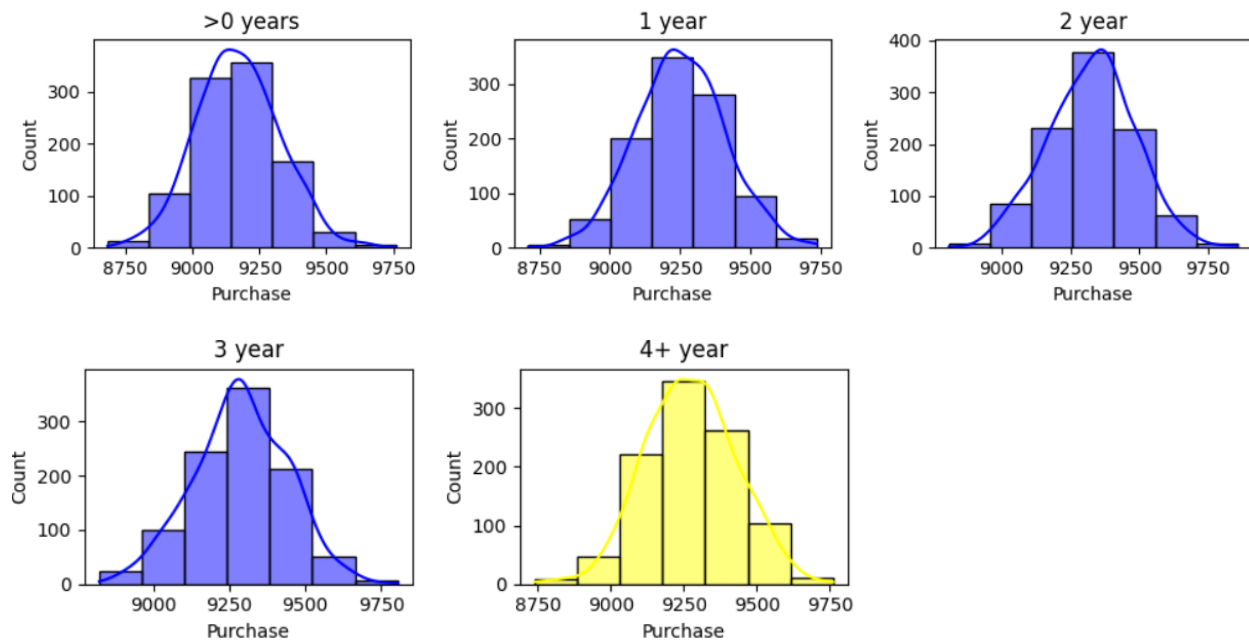
- Population average for 'City A' years is 8911.94 and Sample average for 'City A' is 8914.59.
- Population average for 'City B' years is 9151.3 and Sample average for 'City B' is 9147.21.
- Population average for 'City C' years is 9719.92 and Sample average for 'City C' years is 9724.77.



Final Insights(B):

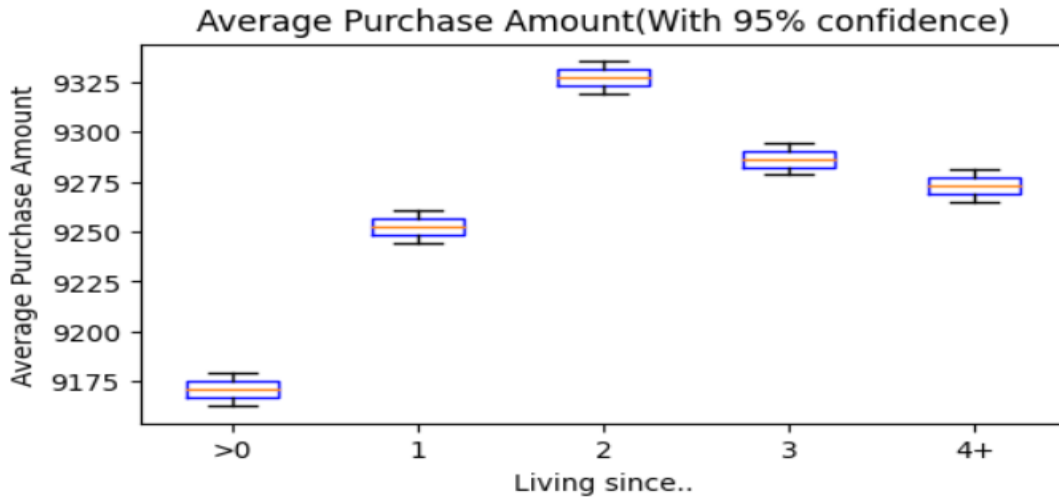
- With 95% confidence we can say, average purchase amount for City A customers lies between 8906.77 And 8922.41.
- With 95% confidence we can say, average purchase amount for City B customers lies between 9138.74 And 9155.68.
- With 95% confidence we can say, average purchase amount for City C customers lies between 9716.07 And 9733.47.

Total residing year vs Purchase amount:



Final Insights(A):

- Population average for '>0' years of residing is 9180.08 and Sample average for '>0' years of residing is 9170.97, quite close.
- Population average for '1' years of residing is 9250.15 and Sample average for '1' years of residing is 9252.16, quite close.
- Population average for '2' years of residing is 9320.43 and Sample average for '2' years of residing is 9327.06, quite close.
- Population average for '3' years of residing is 9286.9 and Sample average for '3' years of residing is 9286.54, nearly overlapping.
- Population average for '4+' years of residing is 9275.6 and Sample average for '4+' years of residing is 9273.02, quite close.
- Normal distribution is followed by every graph.



Final Insights(B):

- With 95% confidence we can say, average purchase amount for customers residing for >0 years lies between 9162.79 And 9179.14.
- With 95% confidence we can say, average purchase amount for customers residing for 1 year lies between 9243.92 And 9260.39.
- With 95% confidence we can say, average purchase amount for customers residing for 2 years lies between 9318.84 And 9335.28.
- With 95% confidence we can say, average purchase amount for customers residing for 3 years lies between 9278.42 And 9294.66.
- With 95% confidence we can say, average purchase amount for customers residing for 4+ years lies between 9264.86 And 9281.18.

Recommendations:

- There is some clear gap between male and female transactions and eventually there is gap in average amount spent. Women centric marketing campaign, with manageable discount on their favorite products can attract more transactions and can balance the gap.
- Customer aged 51-55 despite less transactions are spending more money, availability of their most consumed product can help in their retention, also 18-35 needs special attention as they are new earners and if properly persuaded, they will spend more, marketing campaign based on trending topic and availability of new stock based on it can increase revenue.
- There is dip in the amount spent by the consumer after 2 years of residence, maybe the customer is leaving the store or the people who are living in the area are not our customers. Feedback from customers and immediate action on it can help as it will help us to know if there is some problem from company end or is some taste change in the customer.
- City A and City B have very few average purchases in comparison to City C, Covering all the recommendations can help, as the average of a city will only increase if every section of customer is properly targeted.