WALMART - CONFIDENCE INTERVAL AND CLT

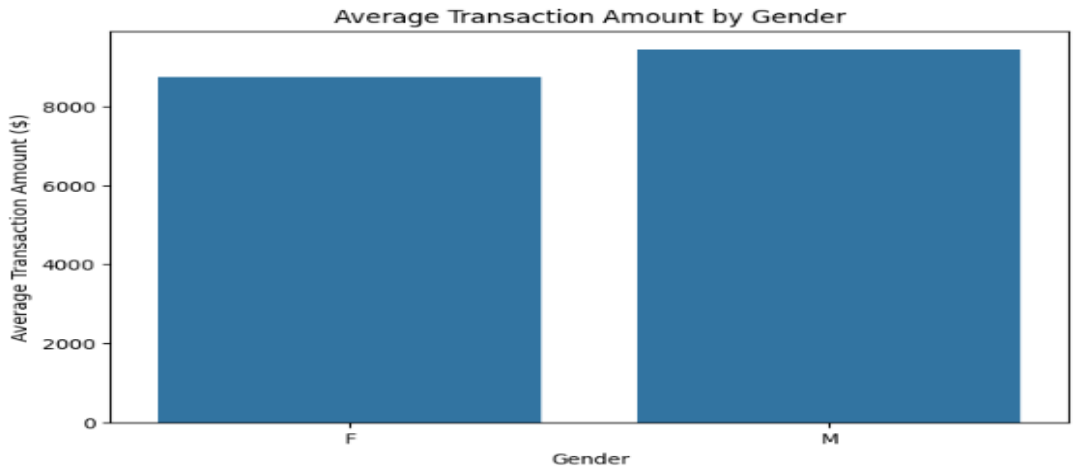
❖ Tracking the amount spent per transaction of all the 50 million female customers, and all the 50 million male customers, calculate the average, and conclude the results.

- We can use the groupby function to efficiently calculate the average transaction amount for both male and female customers.

```
avg_spent = df.groupby("Gender")['Purchase'].mean()
avg_spent
```

Gender	Purchase
F	8734.565765
M	9437.526040

- Using a bar plot, we can visually compare the average transaction amounts of male and female customers.



Insights:

The plot shows that male customers have a higher average transaction amount than female customers, suggesting greater spending per transaction.

❖ Inference after computing the average female and male expenses.

- The computed averages indicate that male customers tend to spend more per transaction than female customers. This insight can help businesses tailor marketing strategies, promotions, and product recommendations based on spending behaviour.

❖ Use the sample average to find out an interval within which the population average will lie. Using the sample of female customers you will calculate the interval within which the average spending of 50 million male and female customers may lie.

- Use sample mean and standard deviation to estimate population mean.
- Apply norm.interval for normal distribution.

- The confidence interval already applies to the population.

```
female_mean = []

for i in range(iteration):
  mean_val = df[df['Gender'] == 'F']['Purchase'].sample(sample_size).mean()
  female_mean.append(mean_val)
print(female_mean)
```

```
print("Female_sample_mean : ",pd.Series(female_mean).mean())
print("Female_sample_std : ",pd.Series(female_mean).std())
print("Female_sample_standard_err : ",pd.Series(female_mean).std() / np.sqrt(sample_size))
```

```
Female_sample_mean :  8760.858153333333
Female_sample_std :  285.8107106988157
Female sample standard err :  16.501289075923946
```

```
from scipy import stats
female_range = stats.t.interval(0.90, dof, loc = female_sample_mean,scale= male_standar_err)
female_range
```

```
(8732.763521773966, 8788.9527848927)
```

Insights:

The average spending per transaction for female customers is **$**8760.85, while for males it is **$** 9434.90. The 90% confidence intervals indicate that the true population average for 50 million customers likely falls within these ranges. This suggests spending trends between genders.

- The interval that you calculated is called Confidence Interval. The width of the interval is mostly decided by the business: Typically 90%, 95%, or 99%. Play around with the width parameter and report the observations.
  - For the confident interval of 90% is:

```
#CI = 90%
from scipy import stats
male_range = stats.t.interval(0.90, dof, loc = male_sample_mean,scale= male_standar_err)
print("Male_range :",male_range)

female_range = stats.t.interval(0.90, dof, loc = female_sample_mean,scale= male_standar_err)
print("female_range :",female_range)
```

```
Male_range : (9406.805441773968, 9462.9947048927)
female_range : (8732.763521773966, 8788.9527848927)
```

  - For the confident interval of 95% is:

```
#CI = 95%
from scipy import stats
male_range = stats.t.interval(0.95, dof, loc = male_sample_mean,scale= male_standar_err)
print("Male_range :",male_range)

female_range = stats.t.interval(0.95, dof, loc = female_sample_mean,scale= male_standar_err)
print("female_range :",female_range)
```

```
Male_range : (9401.41372827502, 9468.386418391648)
female_range : (8727.371808275018, 8794.344498391647)
```

  - For the confident interval of 99% is:

```
#CI = 99%
from scipy import stats
male_range = stats.t.interval(0.99, dof, loc = male_sample_mean,scale= male_standar_err)
print("Male_range :",male_range)

female_range = stats.t.interval(0.99, dof, loc = female_sample_mean,scale= male_standar_err)
print("female_range :",female_range)
```

```
Male_range : (9390.86074101642, 9478.939405650248)
female_range : (8716.818821016419, 8804.897485650246)
```

Insights:

Higher confidence levels (90% , 95% , 99%**)** give wider ranges. 90% CI is more precise but riskier, 95% CI balances both, and 99% CI is widest with more uncertainty. Males consistently spend more than females.

- Conclude the results and check if the confidence intervals of average male and female spends are overlapping or not overlapping. How can Walmart leverage this conclusion to make changes or improvements?
  - The male interval is entirely to the right of the female interval.
  - If True, there is no overlap, and male spending is significantly higher
  - The female interval is entirely to the right of the male interval.
  - If true, there is no overlap, and female spending is significantly higher.

```
print(Lower_m > Upper_f)
```

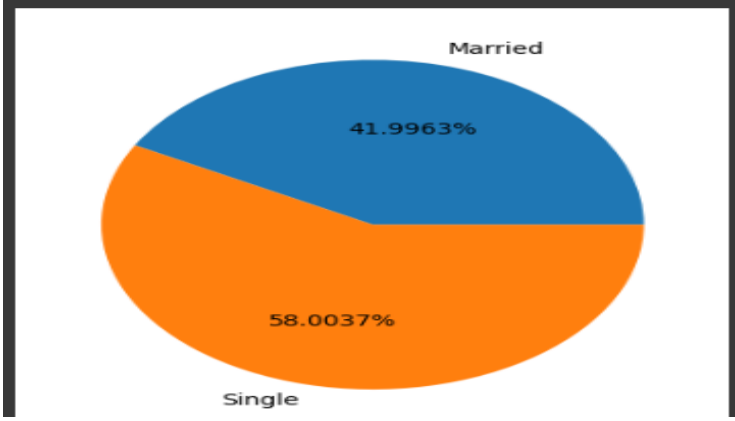
  - True
  - The Code shows there is no overlapping in the male and female Confidence interval for spend.

Insights:

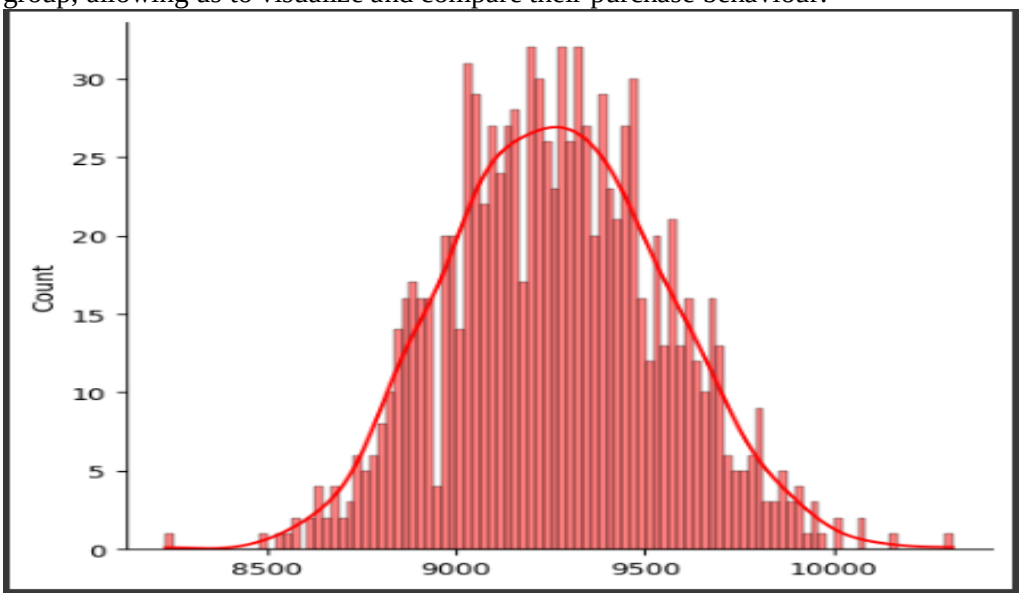
The non-overlapping confidence intervals show that males spend significantly more than females at Walmart. Walmart can leverage this by targeting male customers with tailored promotions and premium offerings to further boost sales.

- Perform the same activity for Married vs Unmarried and Age
  - For Age, you can try bins based on life stages: 0-17, 18-25, 26-35, 36-50, 51+ years.
    - Marital_status:
      - We can utilize the unique() function on the User_ID column to calculate the average purchase for distinct married and single customers.
      - Afterward, we can use a pie chart to visualize which marital status group contributes more to overall purchases.

```
plt.pie(unique_user, labels = unique_user.index, autopct = '%1.4f%%')
plt.show()
```



- We can use a for loop to generate 1,000 sample means for each marital status group. These sample means can then be used to plot a normal distribution chart for each group, allowing us to visualize and compare their purchase behaviour.



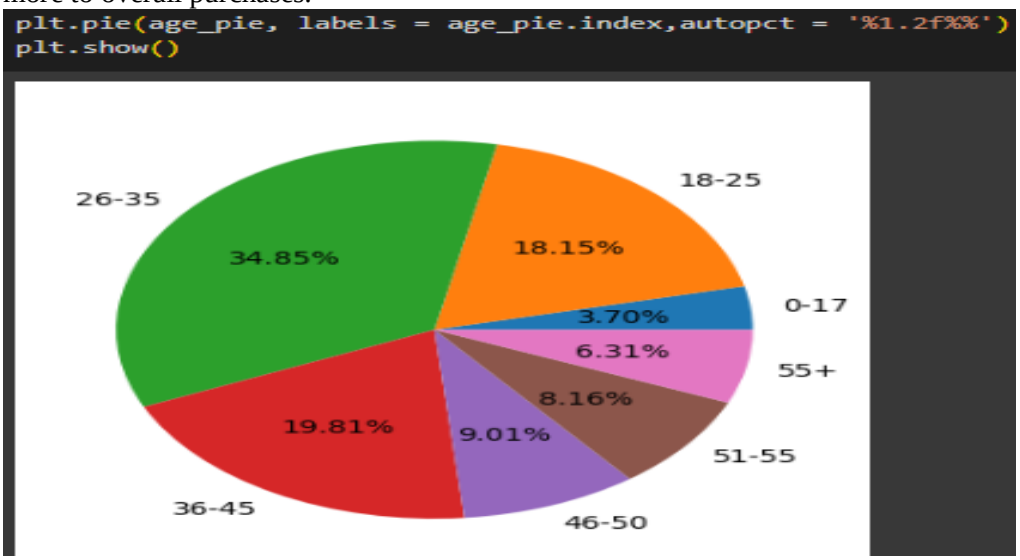
- After that, using confident interval of (0.90, 0.95, 0.99) we can get the range of mean where the population mean lies.

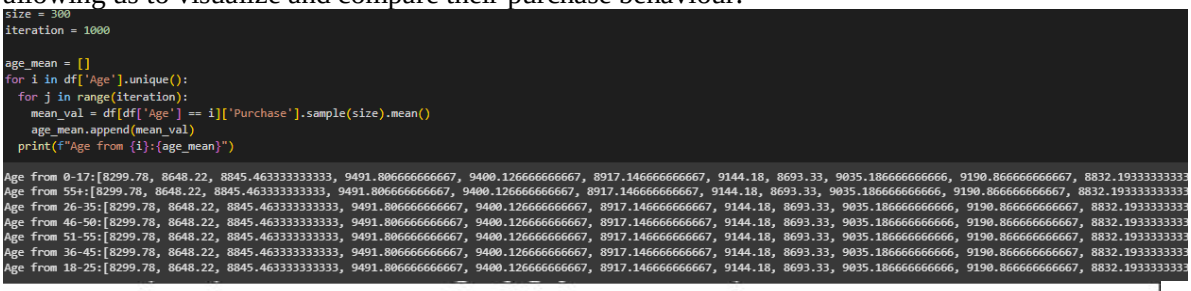
```
from scipy.stats import norm

Ci = [0.90, 0.95, 0.99]

for i in Ci:
  Single_interval = norm.interval(i, loc = Single_mean_val, scale = pd.Series(Single_mean).std())
  print(f"Single_range of {i}: {Singel_interval}")
```

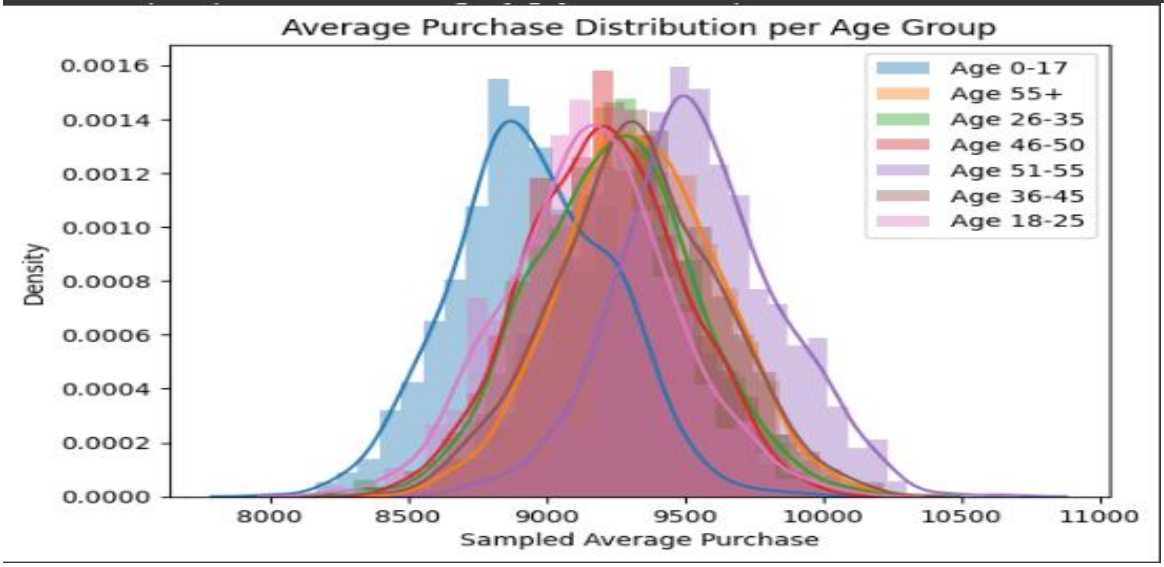
```
Single_range of 0.9: (8526.368812432765, 10008.809167567231)
Single_range of 0.95: (8526.368812432765, 10008.809167567231)
Single_range of 0.99: (8526.368812432765, 10008.809167567231)
```

Insights:

The confidence intervals indicate that the average purchase amounts for single and married customers overlap across all confidence levels (90%, 95%, and 99%). This suggests that there is no significant difference in purchase behaviour between single and married customers, as both groups fall within a similar spending range.

- Age:
  - We can utilize the unique() function on the User_ID column to calculate the average purchase for distinct Age group of customers.
  - Afterward, we can use a pie chart to visualize which Age status group contributes more to overall purchases.
  - 

  - We can use a for loop to generate 1,000 sample means for each Age group. These sample means can then be used to plot a normal distribution chart for each group, allowing us to visualize and compare their purchase behaviour.
  - 

  - 

  - After that, using confident interval of (0.90, 0.95, 0.99) we can get the range of mean where the population mean lies.

```
age of  0.9 : (8659.660913038522, 9652.326266961478)
age of  0.95 : (8564.576727578144, 9747.410452421856)
age of  0.99 : (8378.740080825617, 9933.247099174383)
--- Age from 0-17---
age of  0.9 : (8659.660913038522, 9652.326266961478)
age of  0.95 : (8564.576727578144, 9747.410452421856)
age of  0.99 : (8378.740080825617, 9933.247099174383)
--- Age from 55+---
age of  0.9 : (8659.660913038522, 9652.326266961478)
age of  0.95 : (8564.576727578144, 9747.410452421856)
age of  0.99 : (8378.740080825617, 9933.247099174383)
--- Age from 26-35---
age of  0.9 : (8659.660913038522, 9652.326266961478)
age of  0.95 : (8564.576727578144, 9747.410452421856)
age of  0.99 : (8378.740080825617, 9933.247099174383)
--- Age from 46-50---
age of  0.9 : (8659.660913038522, 9652.326266961478)
age of  0.95 : (8564.576727578144, 9747.410452421856)
age of  0.99 : (8378.740080825617, 9933.247099174383)
--- Age from 51-55---
age of  0.9 : (8659.660913038522, 9652.326266961478)
age of  0.95 : (8564.576727578144, 9747.410452421856)
age of  0.99 : (8378.740080825617, 9933.247099174383)
--- Age from 36-45---
age of  0.9 : (8659.660913038522, 9652.326266961478)
age of  0.95 : (8564.576727578144, 9747.410452421856)
age of  0.99 : (8378.740080825617, 9933.247099174383)
--- Age from 18-25---
```

Insights:

The confidence intervals across all age groups show a similar trend, with most intervals overlapping. However, the 26–35 and 36–50 age groups consistently show higher purchase ranges, suggesting they are the most significant spenders compared to other age segments.

❖ Give recommendations and action items to Walmart.

- **Focus on Male Shoppers**

Males spend more on average. Walmart should create special promotions and loyalty programs aimed at male customers.

- **General Promotions for Marital Status**

Since single and married customers have similar spending habits, Walmart can use the same promotions for both.

- **Target Age Groups 26–50**

Customers aged 26–50 spend more, so Walmart should focus marketing and product offerings on this group.

- **Adjust Inventory**

Stock more premium and bundled products that appeal to high-spending groups (males and customers aged 26–50).

- **Personalized Experience**

Use these insights to offer personalized discounts and product suggestions both online and in stores.