

About Walmart:

Walmart is an American multinational retail corporation that operates a chain of supercenters, discount departmental stores, and grocery stores from the United States. Walmart has more than 100 million customers worldwide.

Business Problem:

The Management team at Walmart Inc. wants to analyze the customer purchase behavior (specifically, purchase amount) against the customer's 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: Do women spend more than men? (Assume 50 million customers are male and 50 million are female).

Features and Descriptions:

User_ID:	User ID
Product_ID:	Product ID
Gender:	Sex of User
Age:	Age in bins
Occupation:	Occupation(Masked)
City_Category:	Category of the City (A,B,C)
StayInCurrentCityYears:	Number of years stay in current city
Marital_Status:	Marital Status
ProductCategory:	Product Category (Masked)
Purchase:	Purchase Amount

Insights:

- Product Category 1, 5, 8 are top 3 categories, in which product 5 is most preferred by each age group except the 55+ age group while product 8 is 2nd most preferred category by all age group except early to middle teenage i.e. 0-17 age group for this particular age group product 1 is 2nd most preferred category. All the product categories have extensive demand in 26-35 age groups. Product 9 is not that much demanding while 14 is still demanding by all age groups, product 17 is reducing for the early and middle teenage group i.e. 0-17
- Most of the purchases are made by marital status '0' for all age groups, marital status '1' are not showing purchase interest in early and middle teenage group i.e. 0-17
- Most of the users are males. Highest selling product category is 1 and 5 in males while in females product category 5 is highest selling followed by product category 8

Central Limit Theorem states that, if we take sample (of any size, here 300, 3000, 30000 will be taken below) from the population and plot the sample mean then will get the gaussian distribution. Below are the results of confidence intervals for different sample sizes based on gender:

- ❖ 95% confidence interval for mean expenses by 300 sample of male users is (9268, 9298) while female users is (9252, 9255)

- ❖ 95% confidence interval for mean expenses by 3000 sample of male users is (9236, 9236) while female users is (8691, 8731)
- ❖ 95% confidence interval for mean expenses by 30000 sample of males users is (8666, 8670) while female users is (8677, 8677)

With the help of these confidence intervals few things can be inferred:

- ❖ With 95% confidence i can say that male spends more than females
- ❖ As the sample size increases, width of confidence interval decreases and accuracy is more with large sample size
- ❖ As the sample size increases, the shape of the distribution of means is narrow

Below are the results of confidence intervals for different sample sizes based on marital status:

- ❖ 95% confidence interval for mean expenses by 300 sample of marital status 0 is (9085, 9119) while marital status 1 is (9098, 9102)
- ❖ 95% confidence interval for mean expenses by 3000 sample of marital status 0 is (9095, 9095) while marital status 1 is (9099, 9137)
- ❖ 95% confidence interval for mean expenses by 30000 sample of marital status 0 is (9109, 9113) while marital status 1 is (9112, 9112)

Nothing can be inferred anything about the purchase based upon marital status as the values are overlapping for all sample sizes, so it could be possible is both the marital status means are lying in same interval and both marital status showing the same purchase behaviour

Recommendations:

- we can increase the stock of product category 1, 5, 8 and 14 and ensure the stock expiry dates for product category 9 and 17 to make sure less wastage and less replacement of those products in walmart store
- Walmart can increase the sales by introducing the new products to the 26-35 age group people, as they are the most active customers for many product categories
- Walmart can focus on price and customer feedback to improve the sales for lowest selling product categories like 7 category in females and 9 in males
- Walmart can target male customers for the products that costs around 9240 and female customers for the products that costs around 8678, as per the population mean calculated for each gender
- Walmart can target marital status 0 customers for the products that costs around 9093 and marital status 1 customers for the products that costs around 9111, as per the population mean calculated for each marital status, but as per the analysis both marital status are showing equal behavior towards their purchase at walmart, walmart can introduce few surveys to know more about it.
- Walmart can gradually increase the sales of products with high cost for both genders and intriduce the new ad to promote more female customers

