

Dataset:

Our dataset consists of a sample size of 1000 randomly chosen datapoints (500 male, 500 female) from a total of 550,000 observations. Each observation corresponds to a transaction made in a retail store in India on Black Friday. The data was taken from Kaggle (<https://www.kaggle.com/code/ishanvardhan/black-friday-sales-analysis-and-prediction>).

The variables in the dataset that will be analyzed:

1. **User\_ID:** Unique identifier for each user.
  2. **Gender:** Categorical variable indicating the gender of the consumer (M or F).
  3. **Age:** Categorical variable indicating the age group of the consumer (0-17, 18-25 ...).
  4. **Marital\_Status :** Categorical variable indicating the marital status of the consumer (0 or 1).
  5. **Purchase\_Total (In Indian Rupees):** Continuous variable indicating the total purchase amount for each transaction.
- 2-4 are predictor variables, 5 is the response variable.

Analysis 1:

- Question: Is the average purchase amount in our sample significantly different from the overall average purchase amount in the population?
- Statistical Analysis: One-Sample Z-Test (stdev is known)

Analysis 2:

- Question: Is there a significant difference in mean spent between men and women?
- Objective: Explore gender-based variations in purchase behavior and determine if there's a statistically significant difference in spending between men and women.
- Statistical Analysis: Two-Sample T-Test

Analysis 3:

- Question: Are there significant differences in purchase amounts among different age groups?
- Objective: Understand purchasing habits across age groups and identify whether age has a significant impact on purchase amounts.
- Statistical Analysis: One-Way ANOVA

Analysis 4:

- Question: Is there a significant interaction effect between age groups and marital status on purchase amounts?
- Objective: Investigate the interaction between age and marital status to detect the combined effect on purchase amounts.
- Statistical Analysis: Two-Way ANOVA

Overall Objectives:

- Understand purchasing habits according to gender, age groups and marital status.
- Predict purchase spent for each customer given some preliminary information about them.