 # Assignment 5: Hypothesis Testing and Statistical Summary on SALES Dataset

#Objective

To perform descriptive statistics, visualize data, and conduct hypothesis testing on a business-related dataset to derive meaningful insights.

 # Dataset

Name: sales_data_sample Dataset

Source: Provided (custom/business dataset with 2823 records and 25 features)

Description: This dataset contains detailed sales order transactions including order quantity, pricing, revenue, product lines, customer regions, order status, and deal size. It is useful for analyzing sales performance and business trends.

 # Tools Used

- Python
  - Pandas
  - NumPy
  - SciPy
  - Seaborn
  - Matplotlib
- Jupyter Notebook

#Summary of Analyses

#1. Descriptive Statistics

- Calculated: Mean, Median, Mode, Std Dev, Min, Max, and Range for numerical columns like `SALES`, `QUANTITYORDERED`, and `PRICEEACH`.

2. Hypotheses

- **H_o (1):** The average sales (`SALES`) per order is ₹3500.
- **H_1 (1):** The average sales per order is not ₹3500.

- **H_o (2):** There is no significant difference in mean `SALES` across different `DEALSIZE` categories.

- **H_1 (2):** There is a significant difference in mean `SALES` across different `DEALSIZE` categories.

#3. Tests Conducted

- ✅ One-sample T-test on `SALES`

- ✅ ANOVA on `SALES` vs `DEALSIZE`

- ✅ Chi-square test for `STATUS` vs `PRODUCTLINE`

#4. Visualizations

- Histograms for `SALES`, `PRICEEACH`

- Boxplots comparing `SALES` across `DEALSIZE`

- Heatmap showing correlation among numeric fields

✅ Results

- The null hypothesis for the one-sample t-test was **rejected**, suggesting that the mean sales differ significantly from ₹3500.

- The ANOVA test showed a **significant difference in sales across different deal sizes**, confirming the alternative hypothesis.

- A moderate correlation was found between `QUANTITYORDERED` and `SALES` (`r ≈ 0.8`), indicating quantity drives revenue.