

## Task 11: A/B Testing – Hypothesis Testing in Python

---

### 1. Introduction

A/B testing is a statistical method used to compare two versions of a product, feature, or strategy to determine which performs better. In this task, an A/B test was conducted using Python to evaluate whether a **test group** performs better than a **control group** based on conversion metrics.

This task demonstrates the ability to:

- Design hypotheses
  - Perform statistical tests
  - Interpret p-values and confidence intervals
  - Make data-driven business decisions
- 

### 2. Objective of the Task

The main objective of this task is to:

- Perform an A/B test using an e-commerce dataset
  - Determine whether there is a statistically significant difference between control and test groups
  - Provide a final business recommendation based on the results
- 

### 3. Tools and Technologies Used

- **Platform:** Google Colab
  - **Programming Language:** Python
  - **Libraries Used:**
    - pandas – data manipulation
    - numpy – numerical operations
    - scipy – statistical testing
    - matplotlib – data visualization
- 

### 4. Dataset Description

#### Original Dataset

- **Dataset name:** E-commerce Product Dataset
- Contains information such as product sales and related attributes.

## Dataset Transformation

Since the original dataset was not designed for A/B testing, it was transformed to simulate an experiment:

- A **control/test group** column was created
- A **conversion metric** was defined based on sales performance

## Final A/B Dataset Columns

### Column Name Description

group      Indicates control or test group

converted      Binary conversion outcome (1 = converted, 0 = not converted)

---

## 5. Experimental Design

### Group Assignment

- Users/products were randomly assigned to:
  - **Control group** (existing version)
  - **Test group** (new version)

### Conversion Definition

- A conversion is marked as 1 if sales are above the median sales value
- Otherwise, conversion is marked as 0

This approach simulates real-world A/B testing scenarios.

---

## 6. Hypothesis Definition

- **Null Hypothesis ( $H_0$ ):**  
There is no significant difference between the control group and the test group.
- **Alternative Hypothesis ( $H_1$ ):**  
There is a significant difference between the control group and the test group.
- **Significance Level ( $\alpha$ ):**  
 $\alpha = 0.05$

---

## 7. Methodology

### Step 1: Data Loading

The dataset was loaded using pandas and inspected for correctness.

### Step 2: Data Preparation

- Random group assignment
- Creation of conversion metric
- Final A/B dataset exported as ab\_data.csv

### **Step 3: Group Segmentation**

Data was split into:

- Control group conversions
- Test group conversions

### **Step 4: Metric Calculation**

- Conversion rate for each group was calculated using mean values.

### **Step 5: Statistical Test**

- An **Independent Two-Sample t-test** was used.
- This test is suitable because:
  - The metric is numerical (0/1)
  - Two independent groups are compared

### **Step 6: Significance Evaluation**

- p-value was compared against  $\alpha = 0.05$
  - Decision made to reject or fail to reject the null hypothesis
- 

## **8. Statistical Analysis**

### **Conversion Rate Comparison**

- Control group conversion rate was calculated
- Test group conversion rate was calculated

### **p-value Interpretation**

- If  $p\text{-value} < 0.05 \rightarrow$  statistically significant difference
  - If  $p\text{-value} \geq 0.05 \rightarrow$  no statistically significant difference
- 

## **9. Confidence Interval**

A 95% confidence interval was calculated for the difference in conversion rates.

- The confidence interval helps estimate the range in which the true difference lies
  - If the interval does not include zero, the difference is considered statistically meaningful
-

## **10. Visualization**

A bar chart was created to visually compare conversion rates between:

- Control group
- Test group

This visualization helps stakeholders easily understand performance differences.

---

## **11. Results Summary**

- Control and test group conversion rates were compared
- Statistical testing was successfully performed
- The p-value and confidence interval guided the final decision

Results were summarized and saved in:

- ab\_test\_summary.csv
- 

## **12. Business Recommendation**

Based on the A/B testing results:

- If the test group shows statistically significant improvement, it is recommended to roll out the test version.
- If no significant difference is observed, the control version should be retained, and further testing may be conducted.

The final recommendation was documented in:

- final\_recommendation.txt
- 

## **13. Deliverables**

The following files were generated and uploaded to GitHub:

- task11\_abtest.ipynb – Complete Python analysis notebook
  - ab\_data.csv – A/B testing dataset
  - ab\_test\_summary.csv – Summary of conversion rates
  - final\_recommendation.txt – Business decision
  - README.md – Task overview and explanation
- 

## **14. Learning Outcome**

Through this task, the following skills were developed:

- Understanding of A/B testing concepts
  - Practical implementation of hypothesis testing
  - Statistical thinking and result interpretation
  - Translating analysis into business decisions
- 

### **15. Conclusion**

This task successfully demonstrates the use of A/B testing to make data-driven decisions. By applying statistical methods to real-world data, meaningful insights were derived, reinforcing the importance of experimentation and analytics in product and business strategy.



