A white robot hand with red and blue stripes

Description automatically generated



**AI/ML**

**LAB MANUAL**

**Disclaimer: The content is curated from online/offline resources and used for educational purpose only**

**Analyzing Customer Satisfaction Levels**

**Objective:**

Develop and test null and alternative hypotheses to examine customer satisfaction differences across sectors.

**Equipment Required:**

* Computer with Python and Jupyter Notebook installed
* Dataset: <https://www.kaggle.com/datasets/ahmedaliraja/customer-satisfaction-10k?resource=download>
* Python Libraries: pandas, numpy, matplotlib, seaborn, scipy, statsmodels

**Prerequisites:**

* Basic understanding of Python programming
* Knowledge of statistics, including hypothesis testing
* Familiarity with pandas for data manipulation
* Understanding of data visualization using matplotlib and seaborn

**Problem Statement:**

Analyze customer satisfaction levels across different service factors (delivery experience, food quality, and delivery speed) and test whether order accuracy significantly affects customer satisfaction.

**Code/Procedure:**

**# Step 1: Load Required Libraries**

# Import necessary libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from scipy.stats import ttest\_ind

**# Step 2: Load Dataset**

df = pd.read\_csv("customer-satisfaction.csv")

**# Step 3: Data Preprocessing**

**# Rename columns for easier access**

df.columns = [

    'Customer',

    'Delivery\_Satisfaction',

    'Food\_Quality',

    'Delivery\_Speed',

    'Order\_Accuracy'

**]**

**# Clean missing values**

df.dropna(inplace=True)

**# Convert categorical data to numerical values**

df['Order\_Accuracy'] = df['Order\_Accuracy'].map({'Yes': 1, 'No': 0})

**# Summary statistics**

print(df.describe())

**# Hypothesis Testing:**

**# H0: Order accuracy does not impact delivery satisfaction**

**# H1: Order accuracy impacts delivery satisfaction**

satisfied = df[df['Order\_Accuracy'] == 1]['Delivery\_Satisfaction']

not\_satisfied = df[df['Order\_Accuracy'] == 0]['Delivery\_Satisfaction']

# Perform t-test

t\_stat, p\_value = ttest\_ind(satisfied, not\_satisfied)

print(f'T-statistic: {t\_stat}, P-value: {p\_value}')

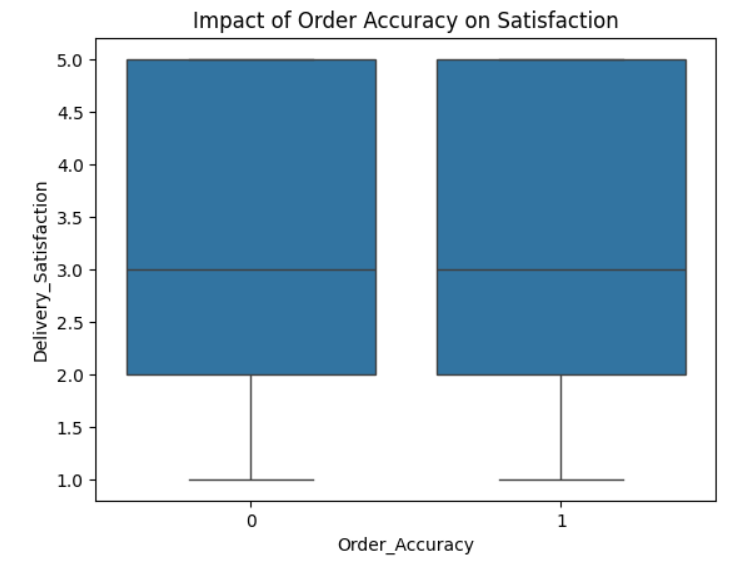
# Visualization

sns.boxplot(x='Order\_Accuracy', y='Delivery\_Satisfaction', data=df)

plt.title('Impact of Order Accuracy on Satisfaction')

plt.show()

**Output:**



**Conclusion**

If the p-value is less than 0.05, we reject the null hypothesis and conclude that order accuracy significantly impacts customer satisfaction.

If the p-value is greater than 0.05, we fail to reject the null hypothesis, indicating no significant relationship between order accuracy and customer satisfaction.