Assignment: Analysing Purchasing Patterns for Strategic Business Insights

Objective:

A retail company is looking to understand its customer's purchasing behavior and demographic trends. The company has provided two datasets: **purchase data** (containing information about orders) and **customer data** (containing customer demographic details). The goal is to derive actionable insights from these datasets.

You are tasked with analysing these datasets to answer key business questions and provide recommendations.

Problem Statement:

The company wants to address the following business questions:

1. Which products contribute the most to revenue, and how can discounts and shipping costs be optimised?
2. Are there any patterns in customer demographics (age, income, and country) that influence purchasing behavior?
3. Is there a significant difference in income or purchasing habits between genders?
4. Can the shipping process be optimized by analysing shipping costs and durations?
5. How strongly does customer income correlate with their purchasing behavior?

Dataset Details:

1. Purchase Data
2. Customer Data

Tasks:

Task 1: Revenue Analysis

1. Calculate the total revenue for each product by combining **price**, **discount**, **quantity**, and **tax**. Provide a list of the top 5 products that contribute the most to total revenue.
2. Visualise the contribution of these top 5 products using a bar chart.

Task 2: Customer Segmentation

1. Visualise the distribution of customers by age groups using a histogram or pie chart.
2. Identify the countries with the highest number of customers and visualize the distribution of customers across countries using a horizontal bar chart.

Task 3: Shipping Cost Optimisation

1. Calculate the average shipping cost for orders that were shipped within **3 days** vs. orders shipped after **3 days**.

Task 4: Gender-based Analysis

1. Compare the average income of male and female customers using a **t-test**.
    - Null Hypothesis (H0): There is no significant difference in average income between male and female customers.
    - Alternate Hypothesis (H1): There is a significant difference in average income between male and female customers.
2. Determine if there is a significant association between **product categories** (derived from product_name) and **customer gender** using a **Chi-Square test**.

Task 5: Correlation.

1. Calculate the correlation between **customer income** and the **total cost** of their purchases.

Task 6: Recommendations

1. Based on the insights derived from the analysis, provide actionable recommendations to the business on:
    - Optimizing product pricing, discounts, and shipping costs.
    - Targeting specific customer segments based on demographics.
    - Enhancing profitability by focusing on top-performing products.

Deliverables:

1. **Code Notebook**:
    - A Jupyter Notebook containing:
        - Python code for all tasks.
        - Visualizations with appropriate titles, labels, and legends.
        - Statistical test results with interpretations.