

Assignment 1: Customer Purchase Behavior Analysis Using Probability

Program: BTech CS-Data Science, IV sem

Batch: B5,6,7,8

Submission date: 24th Feb 2025

MM: 30 (10+10+10)

Submission platform: LMS

Given By: Dr. Sachi Choudhary

Objective:

This assignment will help you understand the application of **probability, random variables, and joint probability** in real-world scenarios using a **customer purchase dataset**. You will perform calculations **on paper, in Excel (mentioning formulas), and in Python**.

Dataset: customer_data.xlsx

Part 1: Solve on Paper

Perform the following calculations manually before implementing them in Excel and Python.

1. Basic Probability Calculation

- Calculate the probability of purchasing each **Product Category**.
- Example: Find **P(Electronics)** if 40 out of 200 transactions were for electronics.

2. Expected Value (Random Variable Concept)

- Compute the expected purchase amount ($E[X]$).
- Example: If purchase amounts are \$50, \$100, \$200 with probabilities 0.3, 0.5, 0.2, compute $E[X]$.

3. Probability Distribution of Spending Behavior

- Categorize spending into bins:
 - \$0-50, \$50-100, \$100-200, \$200+
- Find the probability of each range.

4. Joint Probability (Multiple Random Variables: Product & Payment Method)

- Compute $P(\text{Product} \cap \text{Payment Method})$

- Example: If **20 customers bought Clothing using Credit Card**, compute $P(\text{Clothing} \cap \text{Credit Card})$.

5. Conditional Probability

- Compute $P(\text{Payment Method} | \text{Product Category})$
- Example: Given that a customer bought Electronics, what is the probability they paid via Debit Card?

Part 2: Implement in Excel

Perform the same calculations in **Excel** using formulas. Submit your **Excel file** with formulas applied.

1. Basic Probability: Find the probability of a customer purchasing a particular product category.
2. Expected Purchase Amount
3. Probability Distribution of Spending: Create a probability distribution of spending ranges.
 1. Hint: categorize spending into bins, then calculate frequency and probability.
4. **Joint Probability (Product & Payment Method)**: Find the probability of purchasing a specific product AND using a particular payment method
 1. Hint: Use **Pivot Tables** to count occurrences and normalize values.
5. Conditional Probability: Given Purchase in Category, What's the Preferred Payment Method?
6. Use **Excel charts** to visualize probability distributions.
7. **Analyze trends** (e.g., which product category has the highest spending?).
8. **Time-based Trends**: Monthly and daily sales trends using Purchase_Date.
9. **Customer Behavior**: Spending patterns and frequency of purchases.

Hint for 8 and 9 : Extracting Month and Year from Purchase_Date for monthly trend analysis and create columns.

Group **Purchase_Amount** into bins (e.g., **Low, Medium, High**).

Part 3: Implement in Python (Submit .py or .ipynb file)

1. Basic Probability Calculation
2. Expected Purchase Amount
3. Joint Probability Calculation (Product & Payment Method)
4. Conditional Probability
5. Trend Analysis in Python
 - a. Monthly Sales Trends
 - b. Customer Retention Analysis
 - c. Spending Behavior Distribution
 - d. Outlier Detection in Spending
6. Trend Analysis & Plots
 - a. Monthly Sales Trend (Bar Chart) – Total purchase amount per month.
 - b. Category-Wise Revenue (Pie Chart) – Share of revenue by product category.
 - c. Payment Method Trends (Stacked Bar Chart) – Preferred payment methods over months.
 - d. Customer Purchase Frequency (Histogram) – Number of times customers make repeat purchases.
 - e. Spending Pattern (Box Plot) – Outliers and median purchase amounts.
 - f. Joint Probability Heatmap – Relationship between Product_Category and Payment_Method.
 - g. Write the observations from the trend in the comment section.

Submission

- **Paper Solutions:** Submit a scanned copy or PDF of handwritten calculations.
- **Excel Implementation:** Submit an Excel file (.xlsx) with formulas.
- **Python Implementation:** Submit a Python script (.py) or Jupyter Notebook (.ipynb).

Note: Every file should be named with your SAPId and have details (SAPId, Name, Batch and Program) at the top of each file. Put appropriate comments in the file wherever required.