

DAI_REPORT

Report on Data Analysis of Customer Purchases

- The dataset contains 1,050 records and 11 columns.
- It includes details about customer purchases, such as age, gender, purchase category, amount spent, frequency of purchases, and payment method.
- Some columns contain missing values.

Columns Overview:

| Column Name | Data Type | Description |
|-----------------------|-----------|---|
| Customer_ID | Integer | Unique identifier for customers |
| Age | Float | Age of the customer |
| Gender | Object | Customer's gender |
| Purchase_Category | Object | Category of purchased products |
| Purchase_Amount | Float | Total purchase amount (in currency) |
| Frequency_of_Purchase | Float | Number of times a customer has purchased |
| Purchase_Channel | Object | Mode of purchase (Online, In-Store, Mixed) |
| Brand_Loyalty | Float | Loyalty rating (scale of 1-5) |
| Product_Rating | Float | Rating is given to purchased products |
| Payment_Method | Object | Payment method used (Credit Card, PayPal, etc.) |
| Time_of_Purchase | Object | Date and time of purchase |

Data Cleaning and Preprocessing

Missing Values Summary

- **Age: 105 missing values (10%)**
- **Gender: 112 missing values (10.6%)**
- **Purchase Category: 110 missing values (10.5%)**
- **Purchase Amount: 104 missing values (9.9%)**
- **Frequency of Purchase: 103 missing values (9.8%)**
- **Purchase Channel: 100 missing values (9.5%)**
- **Brand Loyalty: 107 missing values (10.2%)**
- **Product Rating: 105 missing values (10%)**
- **Payment Method: 104 missing values (9.9%)**
- **Time of Purchase: 105 missing values (10%)**

Outliers and Duplicates Analysis

Duplicate Entries

- The dataset contains **47 duplicate** records, that were removed to ensure data integrity.

Outliers (Detected using IQR Method)

- **Age: 18 outliers detected** (values significantly below or above typical customer age).
- **Purchase Amount: 15 outliers detected** (extremely high or low purchase amounts).
- **Frequency of Purchase, Brand Loyalty, and Product Rating: No significant outliers detected.**

Exploratory Data Analysis (EDA)

Demographic Insights

Age Distribution:

- The average customer age is 35.1 years.
- Customers range in age from 18 to 100 years.
- The majority of customers fall between the 25-45 age group.

Gender Distribution:

- Male customers make up 54.7% of the dataset.

- Female customers account for 45.3%.
- Some entries contain missing or incorrect gender values.

Purchase Behavior

Most Popular Purchase Categories:

- The top three categories customers prefer are:
 - Sports & Outdoors
 - Food & Beverages
 - Home Appliances

Purchase Frequency:

- Customers make purchases between 2 to 12 times.
- The average purchase frequency is 6.98 transactions per customer.

Purchase Channels:

- In-Store Purchases: 34.2%
- Online Purchases: 29.1%
- Mixed Purchases (both online & in-store): 36.7%

Most Used Payment Methods:

- PayPal: 21.2%
- Credit Card: 19.5%
- Debit Card: 18.7%
- Other Payment Methods (UPI, cash, etc.): 40.6%

Purchase Amount Analysis

- The average purchase amount is ₹320.28.
- The lowest purchase amount is ₹50.71, while the highest is ₹4737.70.
- The standard deviation is ₹430.78, indicating high variation in spending patterns.

Brand Loyalty and Product Ratings

Brand Loyalty Scores:

- The average brand loyalty score is 2.99 (on a scale of 1-5).
- Loyal Customers (score ≥ 4): 32%
- Low Loyalty Customers (score ≤ 2): 28%

Product Ratings:

- The average product rating is 3.01 (on a scale of 1-5).

- High ratings (4-5 stars): 40%
- Low ratings (1-2 stars): 30%

Bivariate and Multivariate Analysis

Bivariate Analysis

Bivariate analysis examines relationships between two variables to uncover patterns, correlations, or trends.

1. Age vs. Purchase Amount

- Correlation Coefficient: 0.49 (Moderate Positive)
- Observation: Older customers tend to spend more per purchase.

2. Age vs. Frequency of Purchase

- Correlation Coefficient: -0.07 (Weak Negative)
- Observation: Younger customers make more frequent purchases than older ones, but the effect is minimal.

3. Purchase Amount vs. Brand Loyalty

- Correlation Coefficient: -0.04 (Weak Negative)
- Observation: Customers who spend more do not necessarily have higher brand loyalty.

4. Purchase Amount vs. Product Rating

- Correlation Coefficient: -0.006 (Almost No Correlation)
- Observation: Spending more on a product does not significantly impact ratings.

5. Purchase Channel vs. Payment Method

- **Insight:**
 - Online shoppers prefer PayPal and Credit Cards.
 - In-store shoppers use Debit Cards and Cash/UPI more frequently.

Multivariate Analysis

Multivariate analysis examines relationships between three or more variables to understand complex interactions.

1. Age, Purchase Amount, and Brand Loyalty

- Older customers tend to make higher-value purchases, but spending patterns do not significantly affect brand loyalty.

- Younger customers who make frequent, lower-value purchases tend to have higher loyalty scores.

2. Purchase Frequency, Payment Method, and Purchase Channel

- Customers who shop frequently online prefer digital payment methods like PayPal.
- Customers who shop in-store frequently tend to use cash, debit cards, or UPI payments.

3. Product Rating, Purchase Amount, and Category

- Low-priced items (under ₹500) receive higher ratings on average.
- Expensive products (₹3000+) have more mixed reviews, likely due to higher customer expectations.
- Home Appliances and Electronics categories show the highest rating variations.

4. Frequency of Purchase, Brand Loyalty, and Product Rating

- Frequent buyers tend to have higher brand loyalty and give better product ratings.
- Customers who shop infrequently tend to give lower ratings, possibly due to one-off negative experiences.