

Customer Insights & Spending Prediction

Author: Shubham Gadekar

Project Type: Data Analyst Portfolio Project

Tools Used: Python (Pandas, NumPy, Matplotlib, Scikit-learn)

1. Introduction

Understanding how customers behave and what influences their spending is important for any business. This project focuses on analysing customer transaction data to understand spending patterns and predict how much a customer is likely to spend in a transaction.

The goal was not only to analyze the data but also to extract insights that could help businesses improve marketing and customer retention strategies.

2. Problem Statement

The main purpose of this project is to:

- Study customer demographics and transaction behaviour
 - Identify patterns in customer spending
 - Understand which factors influence higher spending
 - Build a simple model to predict transaction-level spending
-

3. Dataset Overview

The dataset used in this project contains customer transaction records collected over a period of time.

- Total records: 2,500
- Total features: 11

The dataset includes customer details such as age, gender, location, segment, employment status, referral information, and the amount spent per transaction. The target variable for prediction is **Amount spent**.

4. Data Cleaning and Preparation

Before starting the analysis, the dataset was cleaned to ensure accuracy and consistency.

Missing values were handled using median and mode for numerical and categorical fields. In some cases, business assumptions were applied instead of removing data. The transaction date column was converted to a proper datetime format, duplicates were removed, and column names were standardized for easier analysis.

5. Feature Engineering

To make the analysis more meaningful, additional features were created. Customers were grouped into age ranges to simplify demographic analysis. Time-based features such as transaction month were extracted to study spending trends over time. Spending categories were also created to identify low, medium, and high spending transactions.

These features helped improve both analysis and model interpretation.

6. Exploratory Data Analysis

Exploratory data analysis was performed to understand customer spending behaviour.

The analysis showed that customers in the 26–45 age range generally spent more per transaction. Certain customer segments consistently contributed higher spending, while referral-based customers tended to spend more compared to non-referral customers. It was also observed that revenue was concentrated in a few states, and spending varied across different months, indicating seasonal patterns.

7. Spending Prediction Model

A linear regression model was built to predict the amount spent per transaction. Simple and interpretable features such as age, referral status, and transaction month were used as inputs.

The model was evaluated using R^2 score and Mean Absolute Error. While the model was not intended to provide perfect predictions, it offered reasonable accuracy and helped understand the relationship between customer attributes and spending behaviour.

8. Business Insights and Recommendations

Based on the analysis, a few practical recommendations can be made. Businesses can focus more on high-spending age groups and customer segments. Referral programs can be

strengthened, as referred customers showed higher spending. Marketing efforts can also be tailored for high-performing states and peak spending months.

9. Conclusion

This project demonstrates how customer transaction data can be used to gain insights into customer behaviour and predict spending using Python. The approach followed is simple, practical, and easy to explain, making it suitable for real-world business analysis.

10. Tools and Technologies

- Python
 - Pandas and NumPy
 - Matplotlib
 - Scikit-learn
 - Jupyter Notebook
-