

Week 17 - Graded Mini Project

Learning Outcome Addressed

- Understand the purpose and importance of Exploratory Data Analysis (EDA) in data science.
- Apply statistical summaries and visualizations to describe the key features of a dataset.
- Identify data patterns, correlations, and anomalies through graphical and quantitative methods.
- Formulate and test hypotheses based on EDA findings to inform further analysis.

Objective

A nationwide retail chain is exploring customer behavior and seasonal trends to improve its strategic decision-making. You have been hired as a data analyst to help the company uncover actionable insights from a sample of past transaction records. The leadership team is particularly interested in patterns related to purchases, customer segments, promotional impact, and seasonal performance.

Submission Instructions

Please document your response on the following pages.

Once you have completed the activity, save the file as a PDF and upload it. Be sure to name the file as **Module 17: Graded Mini Project_[Your last name]**.

Your submission will be considered complete when it meets these criteria:

- Includes all the key elements outlined in the activity instructions and the rubric.
- Adheres to the submission guidelines.
- Is submitted on time.

This is a required activity and counts towards programme completion.

Reflect on the task and respond to the following questions.

Data Description

You are provided with a CSV file titled 'Retail_Transactions_Dataset.csv'. Each row corresponds to a unique customer transaction. Below are the columns available in the dataset:

Column Name	Description
Transaction_ID	Unique identifier for each transaction
Date	Timestamp of the transaction
Customer_Name	Name of the customer
Product	List of items bought in the transaction
Total_Items	Number of different items purchased
Total_Cost	Total cost paid by the customer
Payment_Method	Mode of payment used
City	City in which the transaction took place
Store_Type	Type of store (e.g., Supermarket, Warehouse Club)
Discount_Applied	Whether a discount was used
Customer_Category	Segment the customer belongs to
Season	Season in which the transaction happened
Promotion	Type of promotion applied, if any

Project Objective

Perform a comprehensive analysis on this retail dataset to help answer key business questions. Your project should be organized into the following tasks:

Task 1: Data Preparation

- Read the CSV file.
- Parse and convert the Date column into appropriate format.
- Extract additional useful information like Year, Month, or DayOfWeek from the Date.
- Clean and preprocess the data if required.

Task 2: Basic Exploration

- How many total transactions are there?
- How many unique customers are in the dataset?
- What are the top 5 most common products sold across all transactions?
- Which cities have the highest number of transactions?



Task 3: Customer Behavior Analysis

- Which customer categories spend the most on average?
- Do certain customer categories prefer specific payment methods?
- What is the average number of items bought per transaction per store type?

Task 4: Promotion & Discount Impact

- What is the average cost of transactions where a discount was applied vs not applied?
- Compare the average number of items purchased for different promotion types.
- Which promotion type seems to be most effective in terms of increasing total cost?

Task 5: Seasonality Trends

- Which season has the highest total revenue?
- Are there seasonal preferences for certain store types or product categories?
- Create a plot showing average spending per season.

Task 6: Visualization Dashboard

- Bar plot of number of transactions per city
- Pie chart showing distribution of payment methods
- Line chart of monthly revenue trends (grouped by year if applicable)
- Heatmap or stacked bar showing revenue by season and customer category

Deliverables

Your project should include:

- Clean, commented Python code
- Output of all calculations and visualizations
- A short summary of key insights at the end