

# Advanced Python Project (First semester (2025/2026))

## Supermarket Sales Analysis (20 Marks) Deadline: 10/1/2026

Instructor Name: Dr. Ali Azwai and Dr. Ala Abuthawabeh

### Objective

Students will use **Pandas**, **NumPy**, **Seaborn**, and **Matplotlib** to perform deep data analysis on the **Supermarket Sales Dataset**.

They will clean, explore, visualize, and interpret real-world sales data.

### Supermarket Sales Dataset

**Dataset:** Supermarket Sales – Kaggle (Public)

**Link:** <https://www.kaggle.com/datasets/faresashraf1001/supermarket-sales>

### Description (Students)

In this assignment, you will work with the **Supermarket Sales Dataset**, which contains real transaction records for supermarkets in Myanmar. You will use Python (Pandas, NumPy, Matplotlib, Seaborn) to perform data exploration, cleaning, visualization, and statistical analysis.

Your task is to complete **all steps below**, answer the required questions, and submit both:

1. A **Jupyter Notebook** or A **Colad (file)** with code and graphs
2. A **PDF report** summarizing your findings

### Grading Breakdown (20 Marks)

Section	Task	Marks
1. Load & Inspect Data	Import dataset, show first rows, info, describe	2 marks
2. Data Cleaning	Handle missing values, adjust datatypes	2 marks
3. Exploratory Analysis	for Branch stats, customer types, gender, payment - Compute mean, median, max, and min (Basic Statistics) -Plot the distribution of each - Create a bar chart for frequency of each - Line plot of sales over time - Plot rating trends over the same period - Scatter plot of sales vs. rating. - Correlation heatmap of numerical variables. Box plot of gross income grouped by Product line	2 marks
4. Sales & Revenue Analysis	Revenue metrics, product line stats, correlations	2 marks
5. Visualization	6 professional visualizations	4 marks

Section	Task	Marks
<b>6. Advanced Questions</b>	Answer 5 analytical questions	<b>4 marks</b>

Q1: Which branch generates the highest revenue? Why might that be?

Q2: Do members spend more than normal customers?

Q3: Which payment method has the highest usage? Why?

Q4: Which product line has the highest average rating?

Q5: Is there a relationship between unit price and quantity purchased?

<b>7. Submit code on elearning</b>	send Github url	<b>1 marks</b>
<b>8. Discussion</b>	Run and presentation (record Video in English)	<b>3 marks</b>