

Project Specification (Lewis Furniture Store Project)

Module: Python for Data Science

Project Theme: Data Analytics for Lewis Furniture Store

Deliverables: Python programs, dataset(s) in .csv, and a professional report.

Section A

Question 1 – Data Creation (30 marks)

You are tasked with simulating sales data for **Lewis Furniture Store**. Generate a .csv dataset with at least **200 sales transactions**. Each row should represent one sales record. The dataset must include the following columns:

- Transaction ID
- Customer ID
- Customer Age Group (18–25, 26–40, 41–60, 60+)
- Product Category (Bedroom, Living Room, Dining, Appliances, Electronics, Décor)
- Product Name
- Unit Price (R)
- Quantity Purchased
- Total Sale Amount (calculated)
- Date of Purchase
- Payment Method (Cash, Credit Card, Store Account, EFT)

The dataset should reflect **realistic trends**, e.g., higher demand for certain products in certain months.

Question 2 – Data Cleaning & Visualization (40 marks)

- Clean your dataset: remove missing/zero values and duplicates.
- Create frequency tables for:
 - Product Category
 - Payment Method
 - Customer Age Group
- Use Matplotlib/Seaborn to create:
 - A bar chart showing sales per product category.
 - A pie chart showing payment method distribution.
 - A line chart showing monthly total sales trends.
 - A scatter plot comparing customer age vs. total amount spent.

Question 3 – Report & Analysis (30 marks)

Document your process and explain the insights you drew from your graphs, e.g.:

- Which product categories generate the most revenue?
- Which customer groups are most valuable?
- Which payment method is most commonly used?

Question 4 – Advanced Analysis (80 marks total)

4.1 Product & Revenue Analysis (15 marks)

- Find the **top 5 products by revenue** and the **bottom 3 products by revenue**.
- Display the results in both **dictionary format** (product: revenue) and a **horizontal bar chart**.
- Which product category contributed the most to **total store revenue**?

4.2 Customer Segmentation (15 marks)

- Calculate **average spend per customer age group**.
- Using a dictionary, map each age group to their **preferred payment method** (most frequently used).
- Which age group generates the **highest average transaction value**?

4.3 Time-Series & Seasonal Trends (15 marks)

- Group sales data by **month** and calculate:
 - Total sales per month
 - Average sales per transaction per month
- Use a **line graph** with dual axes (left = total sales, right = average sales).
- Identify the **three strongest sales months** and suggest why (seasonality, promotions, holidays, etc.).

4.4 Profit & Margin Analysis (15 marks)

Assume the following **cost multipliers** per category:

- Bedroom: 0.65
- Living Room: 0.60
- Dining: 0.55
- Appliances: 0.70
- Electronics: 0.75
- Décor: 0.50

- a. Create a new column Profit = Total Sale Amount * (1 - multiplier)
- b. Rank categories by **total profit** instead of revenue.
- c. Does the highest revenue category also have the highest profit? Why or why not?

4.5 Customer Loyalty & Repeat Purchases (10 marks)

- a. Count how many **unique customers** are in the dataset.
- b. Identify the **top 5 customers by total spend**.
- c. Calculate the percentage of customers who made **2 or more purchases**.
- d. What recommendation would you give Lewis Furniture to increase **customer loyalty**?

4.6 Predictive Insight Challenge (10 marks)

- a. Using NumPy arrays, extract all transactions for **Electronics** and calculate:
 - Mean
 - Median
 - Standard deviation of transaction amounts
- b. Based on this distribution, predict whether **Electronics sales are more volatile** than other categories.
- c. Suggest how Lewis Furniture could reduce volatility in sales (e.g., discounts, bundles, loyalty programs).