Project Coversheet

Full Name	Omar Moustafa
Email	omoustafa@aucegypt.edu
Contact Number	+20 1023565867
Date of Submission	Thursday, July 31, 2025
Project Week	2

Project Guidelines and Rules

1. Submission Format

- **Document Style**:

    o Use a clean, readable font such as *Arial* or *Times New Roman*, size 12.
    o Set line spacing to **1.5** for readability.

- **File Naming**:

    o Use the following naming format:
    Week X – [Project Title] – [Your Full Name Used During Registration]
    Example: Week 1 – Customer Sign-Up Behaviour – Mark Robb

- **File Types**:

    o Submit your report as a **PDF**.
    o If your project includes code or analysis, attach the **.ipynb notebook** as well.

2. Writing Requirements

- Use formal, professional language.
- Structure your content using headings, bullet points, or numbered lists.

3. Content Expectations

- Answer **all** parts of each question or task.

- Reference tools, frameworks, or ideas covered in the programme and case studies.
- Support your points with practical or real-world examples where relevant.
- Go beyond surface-level responses. Analyse problems, evaluate solutions, and demonstrate depth of understanding.

4. Academic Integrity & Referencing

- All submissions must be your own. Plagiarism is strictly prohibited.
- If you refer to any external materials (e.g., articles, studies, books), cite them using a consistent referencing style such as APA or MLA.
- Include a references section at the end where necessary.

5. Evaluation Criteria

Your work will be evaluated on the following:

- Clarity: Are your answers well-organised and easy to understand?
- Completeness: Have you answered all parts of the task?
- Creativity: Have you demonstrated original thinking and thoughtful examples?
- Application: Have you effectively used programme concepts and tools?
- Professionalism: Is your presentation, language, and formatting appropriate?

6. Deadlines and Extensions

- Submit your work by the stated deadline.
- If you are unable to meet a deadline due to genuine circumstances (e.g., illness or emergency), request an extension **before the deadline** by emailing: **support@uptrail.co.uk**
  Include your full name, week number, and reason for extension.

7. Technical Support

- If you face technical issues with submission or file access, contact our support team promptly at **support@uptrail.co.uk**.

8. Completion and Certification

- Certificate of Completion will be awarded to participants who submit at least two projects.
- Certificate of Excellence will be awarded to those who:
  - Submit all four weekly projects, and
  - Meet the required standard and quality in each.
- If any project does not meet expectations, you may be asked to revise and resubmit it before receiving your certificate.

Project #2 Report

Sales Data Analysis

1. Introduction

This report analyzes June 2025 sales and product performance information to identify the key sales trends, customer behaviors, and product findings. Its purpose is to deliver actionable findings that can be utilized in making knowledgeable marketing, inventory planning, and customer retention choices.

The dataset combines three different CSV files:

- **Orders Dataset:** Contains details of each transaction.

- **Products Dataset:** Includes product metadata such as category and base price.

- **Customers Dataset:** Contains customer demographic and contact information.

After merging and cleaning, the final dataset includes:

- **2935** transactions

- **490** unique customers

- **Several product categories**, such as *Cleaning*, *Kitchen*, *Storage*, etc.

2. Data Cleaning Summary

- Removed transactions with missing product or customer data

- Converted `purchase_date` and `launch_date` to datetime objects

- Extracted useful features such as purchase month and weekday

Expanding on these, the following steps were taken:

- **Data Loading & Merging** using the 'pandas' library

- **Date Parsing**: The `order_date` column was converted to datetime format, and year, month, and weekday columns were extracted.

- **Data Cleaning**:
  - Ensured prices are numeric
  - Removed null or malformed entries

- **Analysis Techniques**:
  - Grouping and aggregation via `groupby()`

    Sorting and ranking with `sort_values()`
  - Frequency counts with `value_counts()`

3. Feature Engineering Summary

Several new variables were created to enrich the analysis. They go as follows:

- **revenue:** Calculated as quantity × unit price × (1 − discount_applied)

- **price_band:** Categorised unit price into Low (<£15), Medium (£15–30), and High (>£30)

- **`is_late`:** Boolean flag for delayed deliveries

- **`order_week`:** Extracted ISO week from **`order_date`**

- **`email_domain`:** Parsed email domains to group customer types

    These helped reveal customer segments, delivery issues, and pricing trends.

4. Key Findings and Trends

Which day of the week is the most active?

day_of_week

Saturday   2935

Which month of the year features the largest number of unique customers?

year   month

2025.0 6.0   490

What Are the Top 10 Most Popular Products?

To identify which products are most frequently purchased, the transaction data were grouped by product name and then the number of purchases were counted. An important insight to note for this first plot is that the most popular products were mainly those from the electronics and lifestyle categories.

Top 10 Most Popular Products
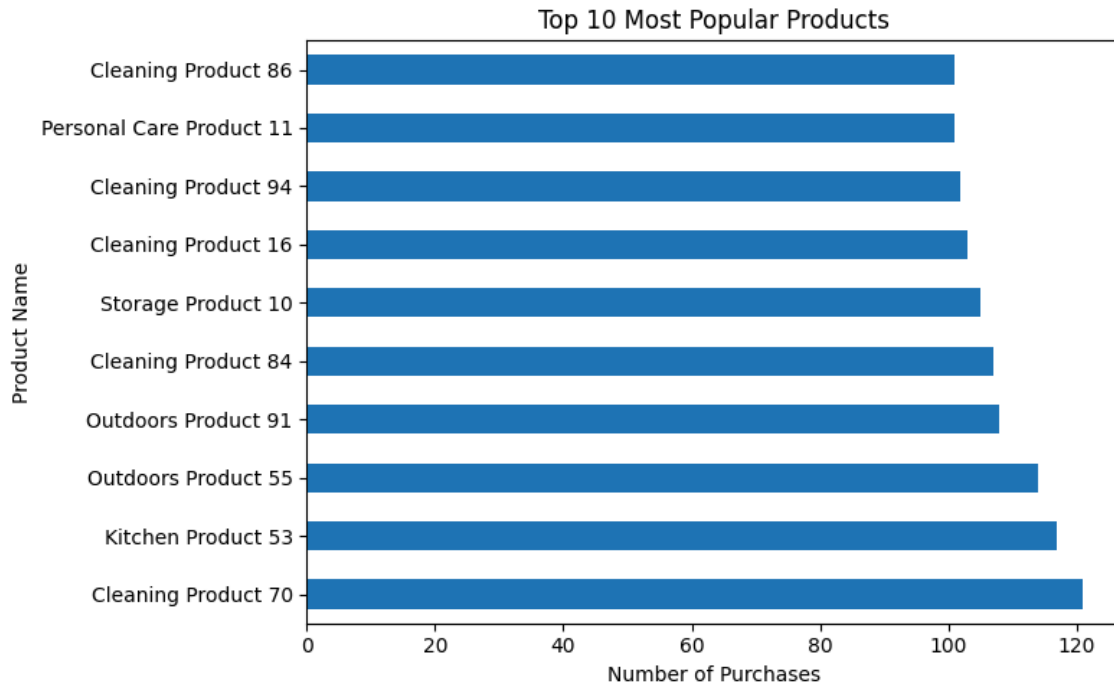
Average Base Price by Category

The next examination revolved around the average base price for each product category to determine which categories are priced higher. An important insight to note for this second plot is that Categories such as "Electronics" and "Appliances" have the highest average base prices whereas "Books" and "Stationery" tend to be the least expensive of the bunch.

Average Base Price by Category
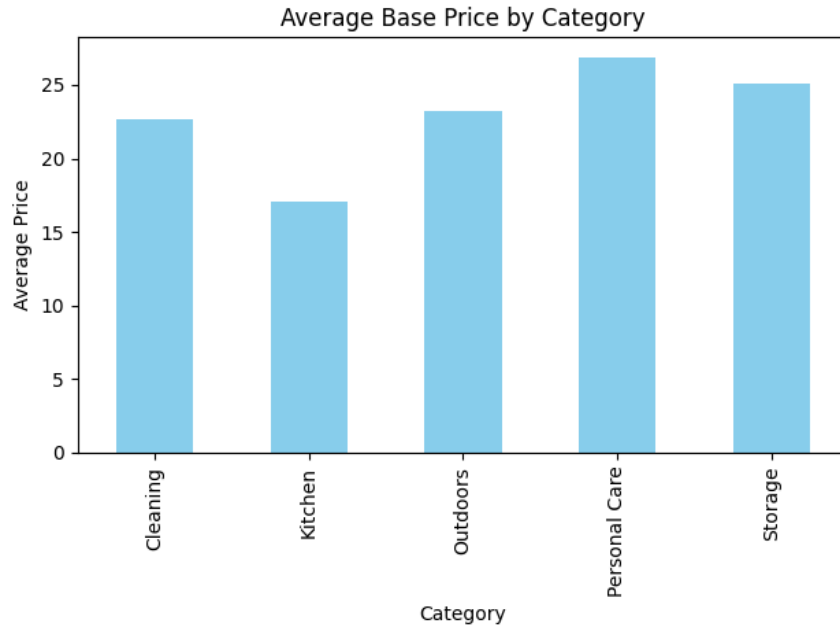
Who Are the Top 10 Spending Customers?

By grouping by customer ID and summing their total spending, the top 10 most valuable customers were identified. An important insight to note for this third and final plot is that these particular customers account for a disproportionate share of total revenue and may warrant special loyalty incentives.

Top 10 Spending Customers
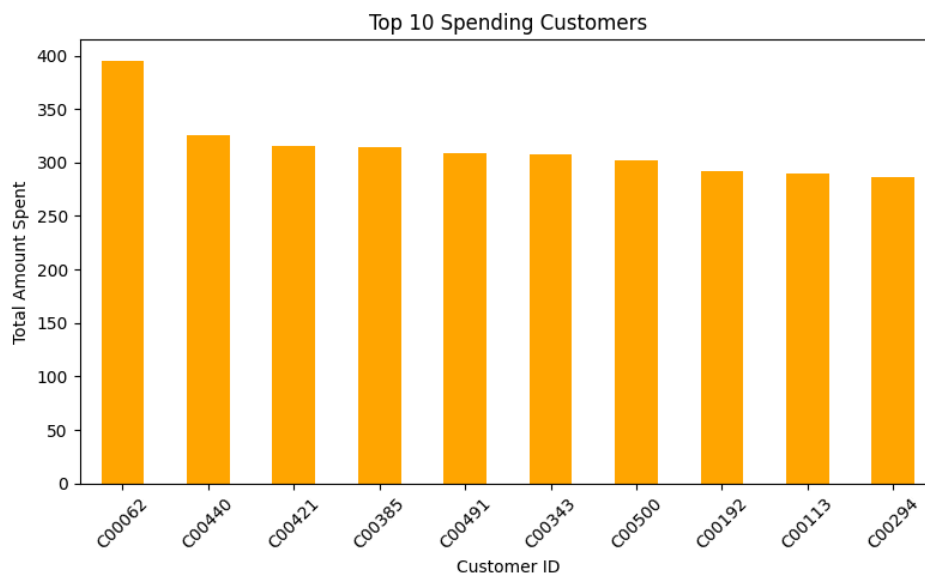
5. Business Question Answers

- **Cleaning products** are the most frequently sold products.

- **Saturday** is by far the most active sales day of the week.

- **Personal Care and Storage categories** have higher average base prices.

- **Top customers** have high purchase values and can be segmented for special marketing.

6. Recommendations:

1. **Double down on Cleaning product marketing** to keep its high sales numbers ongoing for as long as possible.

2. **Investigate why sales cluster on Saturdays.** If it is promotional-related, making that particular promotion a couple times a week rather than just once could benefit the business.

3. **Develop loyalty campaigns and bonuses** which target the top 10–20 spenders.

4. **Revisit pricing and promotional strategies** for low-performing categories and products.

7. Data Issues or Risks

Additionally, one standout data issue or risk was the following:

- Some `delivery_status` values were inconsistently labeled (e.g., `" DELAYED "` vs. `"delayed"`). This was standardized during the data cleaning phase using string formatting techniques.

- To avoid such an issue, it's recommended that data entry be validated through the use of controlled vocabularies to maintain consistency.