Project Coversheet

Full Name	Khalid Takar
Project Title (Example – Week1, Week2, Week3, Week 4)	Week 2

Instructions:

Students must download this cover sheet**,** use it as the first page of their project, and then save the entire document as a PDF before submission**.**

Project Guidelines and Rules

1. Formatting and Submission

- Format: Use a readable font (e.g., Arial/Times New Roman), size 12, 1.5 line spacing.
- Title: Include Week and Title (Example - Week 1: Travel Ease Case Study.)
- File Format: Submit as PDF or Word file
- Page Limit: 4–5 pages, including the title and references.

2. Answer Requirements

- Word Count: Each answer should be within 100–150 words; Maximum 800–1,200 words.
- Clarity: Write concise, structured answers with key points.
- Tone: Use formal, professional language.

3. Content Rules

- Answer all questions thoroughly, referencing case study concepts.

- Use examples where possible (e.g., risk assessment techniques).
- Break complex answers into bullet points or lists.

4. Plagiarism Policy

- Submit original work; no copy-pasting.
- Cite external material in a consistent format (e.g., APA, MLA).

5. Evaluation Criteria

- Understanding: Clear grasp of business analysis principles.
- Application: Effective use of concepts like cost-benefit analysis and Agile/Waterfall.
- Clarity: Logical, well-structured responses.
- Creativity: Innovative problem-solving and examples.
- Completeness: Answer all questions within the word limit.

6. Deadlines and Late Submissions

- Deadline: Submit on time; trainees who fail to submit the project will miss the “Certificate of Excellence”

7. Additional Resources

- Refer to lecture notes and recommended readings.
- Contact the instructor or peers for clarifications before the deadline.

YOU CAN START YOUR PROJECT FROM HERE

Week 2 Project: Sales and Customer Behaviour Insights

---

1. Introduction

This project analyses sales, product, and customer data for Green Cart Ltd with the aim of generating business insights related to revenue performance, customer behaviour, and operational efficiency. The analysis focuses on understanding sales trends over time, identifying high-performing product categories, assessing customer loyalty behaviour, and evaluating delivery performance. The findings are intended to support decision-making across marketing, operations, and customer experience teams.

The analysis uses three datasets: sales transactions, product information, and customer details. These datasets were cleaned, merged, and analysed using Python to produce summary tables and visualisations suitable for a non-technical audience.

---

2. Data Cleaning Summary

Several data quality issues were identified during initial inspection of the datasets. These included missing values, inconsistent text formatting, duplicate records, and incorrect data types.

Key cleaning steps included:

- Conversion of date columns such as order date, product launch date, and customer signup date into datetime format.
- Removal of duplicate records across sales, product, and customer datasets to prevent inflated counts.
- Standardisation of categorical fields such as region, loyalty tier, payment method, and product category.

- Correction of spelling inconsistencies, including variations such as "Brnze" and "Gld" in loyalty tiers, and misspellings in payment methods such as "Bank Tranfer".
- Handling of missing values by applying sensible defaults (e.g. "Unknown") or numeric imputation where appropriate.
- Validation of numeric fields to ensure values such as quantity, unit price, and discounts were non-negative.

These steps ensured the dataset was consistent, reliable, and suitable for aggregation and analysis.

---

3. Feature Engineering

To support deeper analysis, several new variables were created:

- **Revenue**, calculated using quantity, unit price, and applied discounts.
- **Order week**, derived from the order date to enable weekly trend analysis.
- **Price band**, categorising products into low, medium, and high price ranges.
- **Days to order**, measuring the time between product launch and purchase.
- **Email domain**, extracted from customer email addresses for potential segmentation.
- **Late delivery flag**, identifying delayed orders.
- **Signup month**, enabling analysis of customer behaviour over time.

These engineered features enabled more meaningful business insights across time, pricing, and customer segments.

---

4. Key Findings and Trends

Overall revenue showed relatively stable trends across weeks, with some variation by region. Certain regions consistently generated higher weekly revenue, indicating opportunities for region-specific marketing or operational focus.

Product category analysis revealed that a small number of categories accounted for a large proportion of total revenue, suggesting that Green Cart’s sales are driven by a core product set.

Customers in higher loyalty tiers generated significantly more orders and revenue than lower-tier customers, highlighting the value of customer retention programmes.

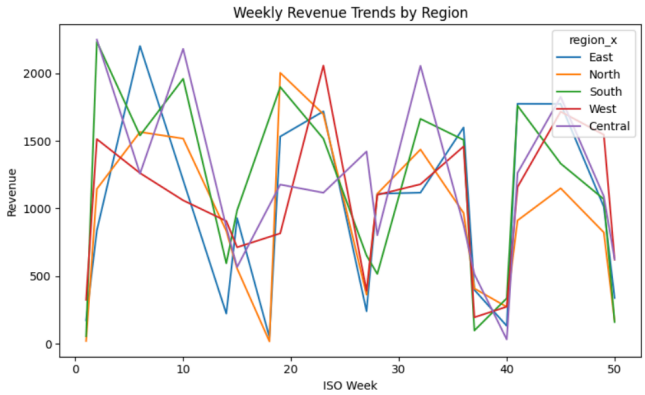
Delivery performance varied by price band, with higher-priced items showing a slightly higher likelihood of delayed delivery. This suggests potential operational challenges associated with premium products.

---

5. Business Question Answers

Which regions generate the highest revenue over time?

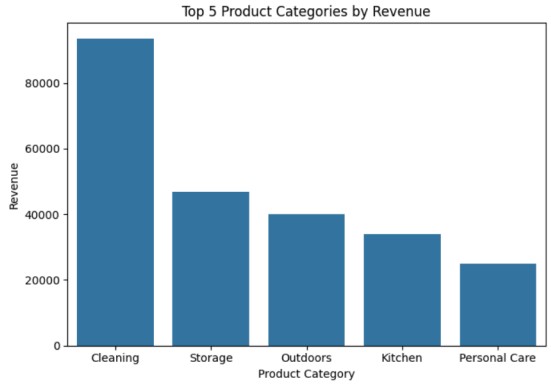
Weekly revenue analysis showed that certain regions consistently outperformed others across the year. This indicates that regional demand patterns are stable and can be leveraged for targeted campaigns and inventory planning.



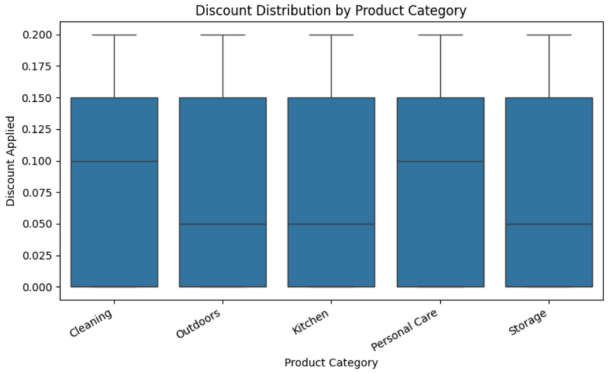
Which product categories contribute most to revenue?

A small number of product categories contributed the majority of total revenue. This suggests that Green Cart should prioritise these categories in promotional strategies and stock management.

How are discounts distributed across product categories?

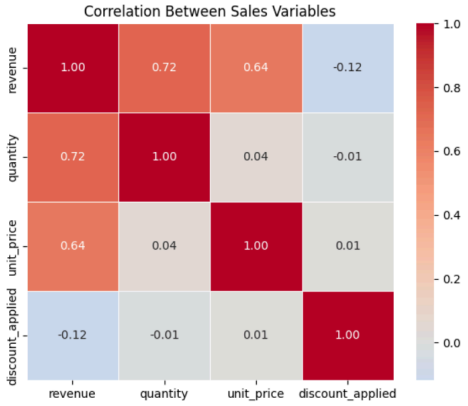


Discount levels varied significantly by product category. Some categories relied more heavily on discounting, which may impact profit margins.



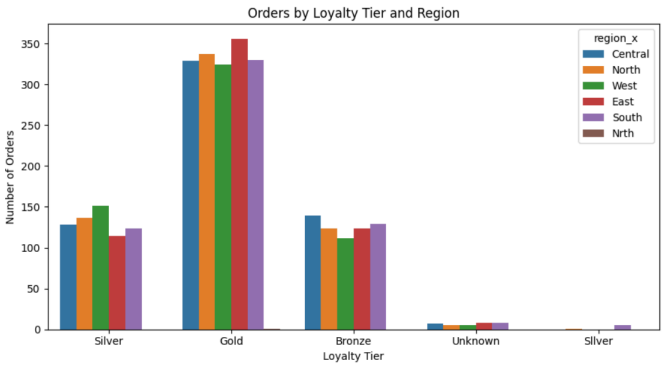
What factors are most strongly related to revenue?

Correlation analysis showed that revenue has a strong positive relationship with quantity and unit price, while discounts showed a weaker negative relationship. This indicates that sales volume and pricing are stronger drivers of revenue than discounting.



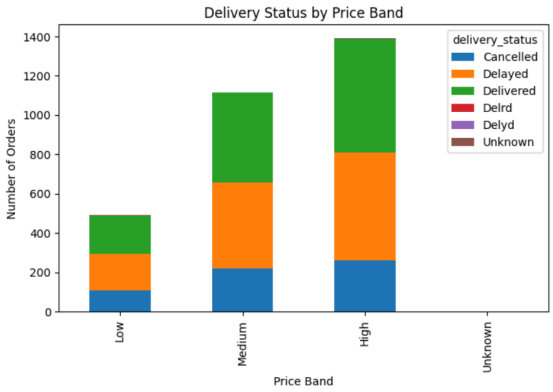
How does customer behaviour vary by loyalty tier and region?

Higher loyalty tiers consistently placed more orders across regions, demonstrating the importance of loyalty programmes. Regional differences were present but secondary to loyalty tier effects.



Does delivery performance vary by price band?

Delivery delays were more common in higher price bands, suggesting that premium products may require additional logistical attention.



6. Recommendations

Green Cart should focus marketing and inventory efforts on high-performing regions and product categories that drive the majority of revenue. Strengthening loyalty programmes could further increase revenue, as higher-tier customers demonstrate

significantly higher engagement. Operational improvements should be explored for higher-priced products to reduce delivery delays and improve customer satisfaction.

---

7. Data Issues and Risks

The main data quality risks identified were inconsistent categorical values and missing information in key fields such as region and payment method. If left unaddressed, these issues could lead to misleading summaries and incorrect business conclusions. Implementing stricter data validation rules and standardised input options at the data collection stage would significantly improve data reliability in future reporting.