

Project Coversheet

Full Name	Pranathi Sarma
Project Title (Example – Week1, Week2, Week3, Week 4)	Sales & Customer Behavior Insights – Green Cart Ltd.

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.

Table of Contents

Contents.....	3
Introduction.....	4
Business Task.....	4
Data Overview.....	4
Report Objectives.....	5
Data cleaning summary.....	5
Duplicate records.....	5
Missing Values.....	5
Standardising inconsistent values.....	6
Correcting Data Types.....	6
Feature Engineering Summary.....	7
Key Findings and Trends.....	8
Cleaning Products Drive Revenue Leadership.....	9
Gold Tier Customers Deliver Disproportionate Value.....	9
Ineffective Discount Strategy Identified.....	10
Business questions answers.....	10
1. Which product categories drive the most revenue, and in which regions?.....	10
2. Do discounts lead to more items sold?.....	11
3. Which loyalty tier generates the most value?.....	11
4. Are certain regions struggling with delivery delays?.....	12
5. Do customer signup patterns influence purchasing activity?.....	12
Recommendations.....	12
Data issues.....	14
Conclusion.....	14

Introduction

Business Task

Green Cart Ltd. is a growing UK-based e-commerce company specializing in eco-friendly household products. As the company prepares for its Q2 performance review, this analysis was commissioned to understand sales patterns and customer behaviour across different regions and product lines.

This report examines three key areas to support strategic decision-making:

- **Sales Performance:** Which products and regions drive the most revenue
- **Customer Behaviour:** How different customer segments interact with the business
- **Operational Efficiency:** How well the company meets delivery expectations

Data Overview

The analysis was conducted using three main data sources provided by the company:

1. **Sales Transactions** - 2,990 customer orders showing what products were purchased, when, and for how much
2. **Product Information** - Details for 30 eco-friendly products across 5 categories (Cleaning, Storage, Outdoors, Kitchen, and Personal Care)
3. **Customer Profiles** - Information for 494 customers including their signup dates, locations, and loyalty levels

After thorough data cleaning and validation, the final analysis includes:

- **1,177 complete customer orders** with all necessary information
- **Full year of sales data** from 2024-2025
- **Comprehensive coverage** across all UK regions and product categories

First few rows of each dataset:						
Sales Data:						
	order_id	customer_id	product_id	quantity	unit_price	order_date \
0	0966977	C00397	P0022	3	39.25	06-07-2025
1	0696648	C00236	P0023	5	18.92	06-07-2025
	delivery_status	payment_method	region	discount_applied		
0	Delivered	PayPal	Central	0.0		
1	DELAYED	credit card	North	0.0		
Product Info:						
	product_id	product_name	category	launch_date	base_price	\
0	P0001	Storage Product 39	Storage	11-03-25	15.88	
1	P0002	Cleaning Product 82	Cleaning	18-08-24	34.23	
	supplier_code					
0	S339					
1	S974					
Customer Info:						
	customer_id	email	signup_date	gender	region	\
0	C00001	shaneramirez@gmail.com	26-04-25	Male	Central	
1	C00002	jpeterson@bernard.com	11-08-24	Female	Central	
	loyalty_tier					
0	Silver					
1	gold					

This combined dataset provides a complete picture of customer purchasing behavior, product performance, and regional sales patterns, enabling reliable insights for business decision-making.

Report Objectives

This report aims to provide clear, actionable insights that help the business:

- Identify revenue growth opportunities
- Improve customer satisfaction and retention
- Optimize operational performance
- Make data-informed decisions for future planning

Data cleaning summary

All data underwent thorough cleaning to ensure accuracy and reliability. This process involved addressing common data quality issues that could affect the validity of our findings.

Duplicate records

- No duplicate records were found in any datasets, confirming each sales transaction, product entry, and customer profile was unique and required no removal actions.

Missing Values

- Critical identifiers such as order_id, customer_id, and product_id had 8 incomplete records that were removed to ensure every analysis point had complete information.

- For operational data like discount values, conservative assumptions were made by setting missing discounts to 0 to avoid overstating promotional impacts.
- Delivery status and payment method missing values were filled with 'Unknown' and 'Not Specified' respectively to maintain data completeness while clearly identifying records with missing information.
- Three customer records were removed due to missing customer IDs, as these could not be reliably analyzed.
- For four records with missing signup dates, the order date was used as a reference point to preserve these transactions in the analysis.
- Missing demographic information for gender and region were filled with 'Not Specified' and 'Unknown' to maintain dataset integrity while being transparent about data limitations.

Standardizing inconsistent values

Inconsistent labels were identified and standardized across all datasets:

- Delivery status was standardized from eight different variations including 'DELAYED', 'delrd', 'delyd', and 'Cancelled' into three clear statuses: 'Delivered', 'Delayed', and 'Canceled'.
- Loyalty tiers were consolidated from eight inconsistent formats such as 'GOLD', 'gold', 'gld', 'braz', and 'silver' into a clean three-tier system of 'Gold', 'Silver', and 'Bronze'.
- Customer demographics were standardized, with gender variations reduced from seven different formats to four consistent categories.
- Regional data was corrected to ensure consistent capitalization across all records.

Correcting Data Types

- Date fields including order_date, signup_date, and launch_date were converted to proper date format to enable accurate time-series analysis.
- Numerical values for prices and quantities were validated to ensure no negative values were present in the data.
- Categorical data was converted to efficient storage formats to optimize processing while maintaining data integrity.

Feature Engineering Summary

To enable deeper analysis of business performance, several new calculated fields were created from the existing data. These enhanced metrics provide additional insights beyond the basic transaction information.

- **Revenue Calculation** was developed by multiplying quantity by unit price and adjusting for any discounts applied. This provides a clear measure of the actual income generated from each transaction rather than just sales volume.
- **Order Week** was extracted from order dates to allow for weekly performance tracking and seasonal pattern analysis. This enables the business to understand sales rhythms throughout the year.
- **Price Band Categorization** was implemented by grouping products into three clear segments: Low-priced (under £15), Medium-priced (£15-30), and High-priced (over £30). This helps analyze how different pricing strategies affect customer purchasing behavior.
- **Days to Order** was calculated by measuring the time between product launch dates and when customers actually purchased those products. This metric helps understand how quickly new products gain market acceptance.
- **Email Domain Extraction** was performed by identifying the domain portion of customer email addresses. This allows for analysis of customer segments based on their email providers, which can indicate different demographic groups.
- **Delivery Performance Flag** was created to clearly identify late deliveries by marking any orders with 'Delayed' status. This enables focused analysis of delivery issues and their impact on customer satisfaction.

Key Findings and Trends

Critical Delivery Performance Issues Identified

The analysis reveals a significant operational challenge affecting customer satisfaction across all regions. Delivery delays impact 35-43% of all orders, with the East region experiencing the highest rate at 43% late deliveries. This consistent pattern across all price bands and product categories suggests systemic operational issues rather than isolated regional problems. The high incidence of delayed orders represents both a customer satisfaction risk and a potential barrier to business growth.



Figure 1: Heatmap highlighting delivery performance challenges across regions and price categories

Cleaning Products Drive Revenue Leadership

Eco-friendly cleaning products emerge as the company's strongest performing category, generating £37,482 in total revenue—significantly ahead of other product lines. This category demonstrates consistent strength across all UK regions, with Central and West regions showing particularly strong performance. The dominance of cleaning products suggests a clear market position that the company can leverage for future growth and category expansion.

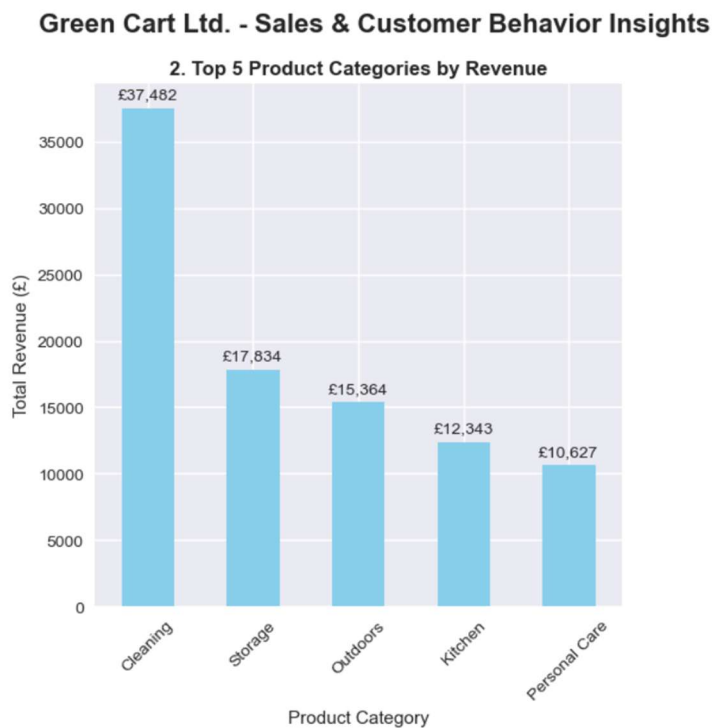


Figure 2: Bar chart displaying the highest-earning product categories with cleaning products leading

Gold Tier Customers Deliver Disproportionate Value

Customer loyalty program analysis reveals that Gold tier members, while representing a similar number of customers as other tiers, generate 44% of total company revenue. These high-value customers show an average order value of £81.96 and contribute £222.58 in revenue per customer, significantly higher than Bronze tier customers at £193.84. This finding highlights the critical importance of customer retention and loyalty program effectiveness for sustained business growth.

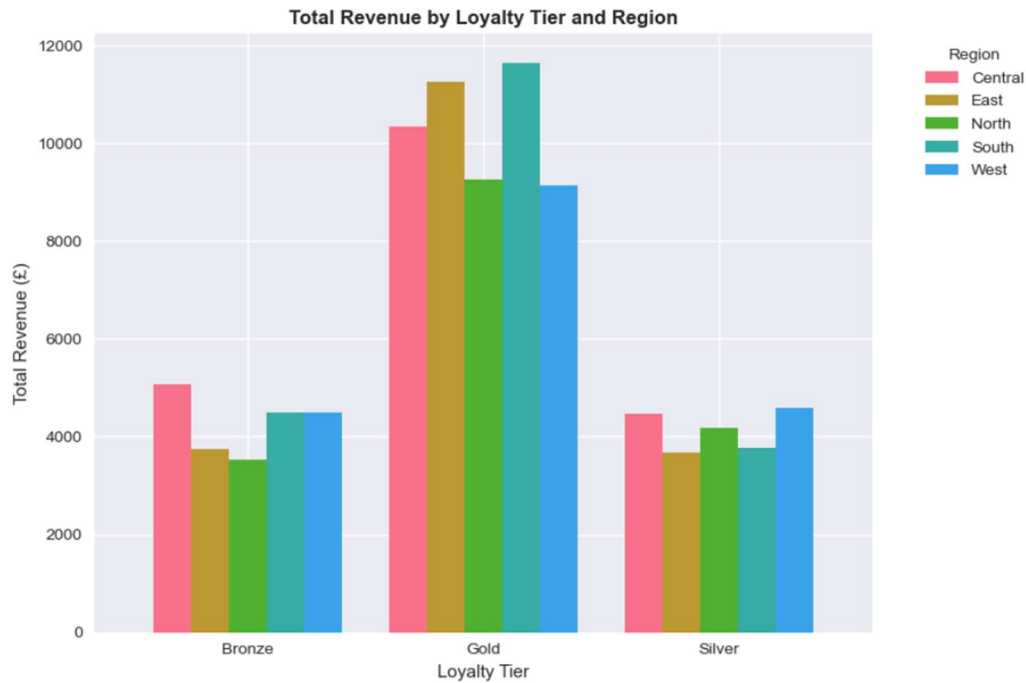


Figure 3: Grouped bar chart demonstrating Gold tier's revenue dominance across all regions

Ineffective Discount Strategy Identified

The correlation analysis between discounts and sales performance reveals that current discount practices show near-zero impact on driving higher order quantities (-0.001 correlation) and a slightly negative relationship with overall revenue (-0.085 correlation). This suggests that the company's discount strategy may be reducing profit margins without delivering the intended sales volume increases, indicating an opportunity for promotional strategy optimization.

Business questions answers

1. Which product categories drive the most revenue, and in which regions?

Cleaning products are the clear revenue leader across all UK regions, generating £37,482 in total revenue. This category demonstrates consistent strength throughout the company's operational areas, with particularly strong performance in the Central region (£7,952) and West region (£7,627).

The product category revenue ranking shows:

- **Cleaning:** £37,482 (32% of total revenue)
- **Storage:** £17,834
- **Outdoors:** £15,364
- **Kitchen:** £12,343
- **Personal Care:** £10,627

2. Do discounts lead to more items sold?

Discounts show no meaningful impact on driving higher sales volumes. The analysis reveals a near-zero correlation (-0.001) between discount levels and quantity sold. Furthermore, higher discounts actually correlate with slightly lower revenue (-0.085 correlation), suggesting that the current discount strategy may be reducing profit margins without achieving the intended sales growth.

Key findings:

- Average quantity remains consistent (2.94-3.08 items) across all discount ranges
- Higher discounts do not translate to larger basket sizes
- Revenue per order decreases as discount levels increase

3. Which loyalty tier generates the most value?

Gold tier customers deliver exceptional value, generating 44% of total company revenue (£51,637) despite representing a similar customer count to other tiers. These customers demonstrate higher spending patterns with an average order value of £81.96 and contribute £222.58 in lifetime value per customer—significantly outperforming Bronze tier customers at £193.84.

Loyalty Tier Performance Comparison:

- **Gold:** £51,637 total revenue | £222.58 per customer
- **Silver:** £20,689 total revenue | £204.85 per customer
- **Bronze:** £21,322 total revenue | £193.84 per customer

4. Are certain regions struggling with delivery delays?

All regions face significant delivery challenges, with late delivery rates ranging from 35% to 43%. The East region shows the poorest performance with 43% of orders delivered late, affecting 104 customer transactions. The South region, while performing best relatively, still experiences 35% late deliveries, indicating a company-wide operational issue requiring immediate attention.

Regional Delivery Performance:

- East: 43% late delivery rate (104 orders)
- Central: 41% late delivery rate (101 orders)
- North: 39% late delivery rate (90 orders)
- West: 37% late delivery rate (86 orders)
- South: 35% late delivery rate (79 orders)

5. Do customer signup patterns influence purchasing activity?

Customer signup timing significantly influences purchasing behavior. Analysis shows that 16% of customers make their first purchase within 7 days of signing up, representing a key conversion metric. Seasonal patterns are evident, with February, August, and October showing the strongest signup-to-purchase conversion rates and highest revenue per customer.

Notable patterns:

- April signups show highest revenue per customer (£237.33)
- February and August demonstrate strong conversion efficiency
- Consistent customer acquisition throughout the year with varying quality
- Quick purchasers (within 7 days) represent a valuable customer segment

Recommendations

1. Immediate Delivery Operations Overhaul

Address the critical 35-43% late delivery rates across all regions by implementing a comprehensive delivery improvement program. Begin with the East region, which shows the worst performance at 43% late deliveries, and establish regional performance benchmarks. Consider partnering with additional logistics providers in underperforming areas and implement real-time delivery tracking with proactive customer communication. The goal should be to reduce late deliveries by 50% within the next quarter, which would significantly improve customer satisfaction and reduce potential revenue loss from delivery-related issues.

2. Gold Customer Retention and Bronze Tier Upgrade Program

Leverage the exceptional value of Gold tier customers who generate 44% of total revenue by enhancing their loyalty benefits. Develop exclusive early access to new products, priority customer service, and personalized product recommendations. Simultaneously, create a targeted Bronze-to-Silver upgrade program offering strategic incentives for Bronze customers who increase their purchasing frequency or average order value. Given that Gold customers deliver £222.58 in lifetime value compared to £193.84 for Bronze customers, even small improvements in tier upgrades could generate substantial revenue growth.

3. Strategic Discount Optimization and Cleaning Category Expansion

Replace blanket discount strategies with targeted promotions based on the finding that current discounts show no correlation with increased sales volume. Instead, focus promotional efforts on slow-moving inventory and use data-driven personalization for high-value customers. Additionally, **capitalize on the dominant position of cleaning products** by expanding this category with new eco-friendly variants and complementary product bundles. Since cleaning products already generate £37,482 in revenue and lead in every region, strategic expansion could further solidify market leadership while reducing reliance on ineffective discounting.

These three focused initiatives address the most significant opportunities identified in the analysis while building on existing strengths, providing a clear roadmap for improved profitability and customer satisfaction in the upcoming quarter.

Data issues

The analysis identified inconsistent delivery status recording across systems, with eight different variations (such as 'DELAYED', 'delrd', 'delyd') requiring manual standardization. This inconsistency hampers accurate delivery performance tracking and operational decision-making. To resolve this, implementing standardized dropdown menus in all order management systems with automated validation checks is recommended. This improvement would ensure consistent data entry across regions, enabling reliable delivery performance monitoring and more effective customer service initiatives.

Conclusion

Green Cart Ltd. demonstrates strong potential with market-leading cleaning products and valuable Gold-tier customers. However, urgent action is needed to address critical delivery delays affecting 35-43% of orders and to optimize ineffective discount strategies. By focusing on these key areas while leveraging existing strengths, the company can significantly improve customer satisfaction and drive sustainable growth in the eco-friendly products market.