# Project Coversheet

| | |
|---|---|
| Full Name | Emmanuel Onyekachukwu Agada |
| Email | yeksagads@gmail.com |
| Contact Number | 07546779925 |
| Date of Submission | 28/07/2025 |
| Project Week | Week 2 |

# Green Cart Ltd. Annual Sales and Customer Insights.

**Introduction:**

The task assigned by the manager on the Data & Insights team is to examine sales and customer behaviour over the year across different regions and product lines. This information will be used to guide future marketing and operational strategies. Access to the sales, product, and customer datasets has been granted, containing essential details such as 'sales data', 'loyalty tier', and others.

The report will display weekly revenue trends by region, identify which categories generate the most revenue, and analyse customer behaviour regarding subscriptions and other relevant factors.

**Data Cleaning Summary:**

Customer Info:

- Changed the signup date column to a datetime dtype and added the sign-up month.
- Changed all the values of the gender column to lowercase and handled inconsistent values by changing them to simple 'm' and 'f' values.
- Changed all the values in the loyalty_tier to lowercase
- Forward fill interpolation to handle NAN values.

Sales Data:

- Filled in the missing values in the quantity column with the mode.
- Filled in the missing values in the unit price column with the mean.
- Filled in the missing values in the discount applied column with the most common value (0)

- Changed the values in both the delivery status and payment method columns to lowercase.
- Changed order date to datetime dtype.
- Interpolated the remaining NAN values to maintain consistency.
- Changed the quantity column to numeric and filled in the missing values with the mode.

Product Info:

- Changed launch date to datetime dytpe.
- Changed base price to numeric dytpe.

Merged df:

- Changed two column names to reflect the different regions. (order region and customer region)
- Standardised inconsistent labels in loyalty tier, delivery status, order region and payment method
- Encoded new order week column to make for better graphs.
- Forward fill interpolation for NAN values

**Feature Engineering Summary:**

- Revenue: This is the total amount generated by multiplying the quantity of the product by its unit price, while also applying the discount if available
- Order Week: This is the week the order was placed.
- Price band: This categorises the unit price of an item based on how much it costs.
- Email domain: The domain name of the customer's email account.
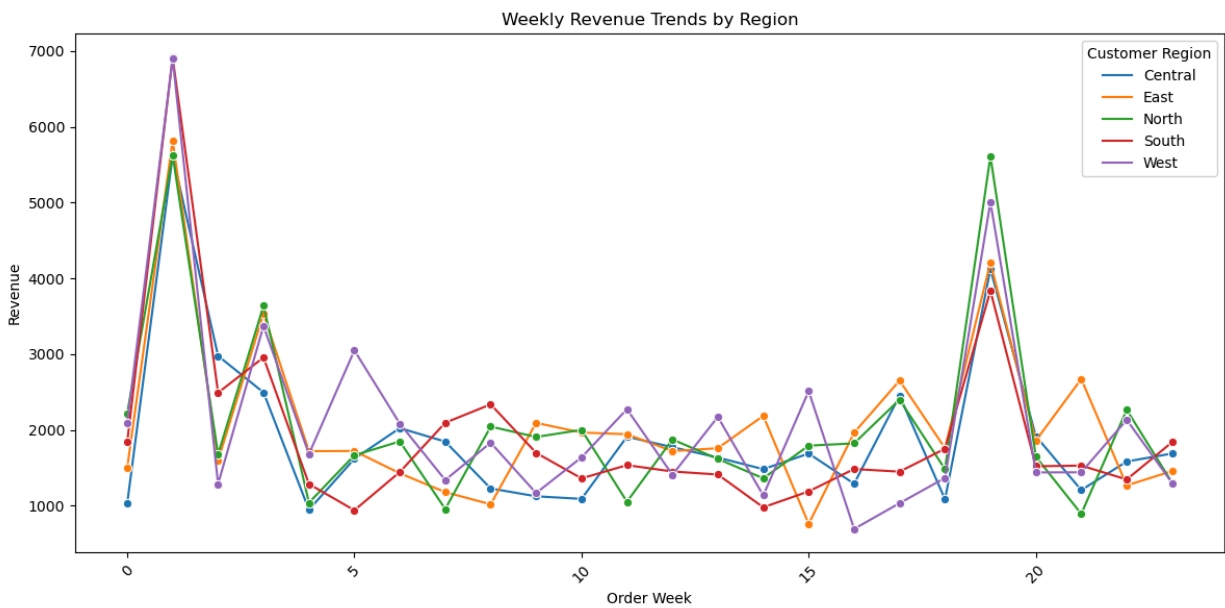- Is late: Shows True if the delivery status is noted down as delayed.

**Key Trends:**

Revenue peaked in Week 1, being driven by high-value orders being delivered from the Eastern region.
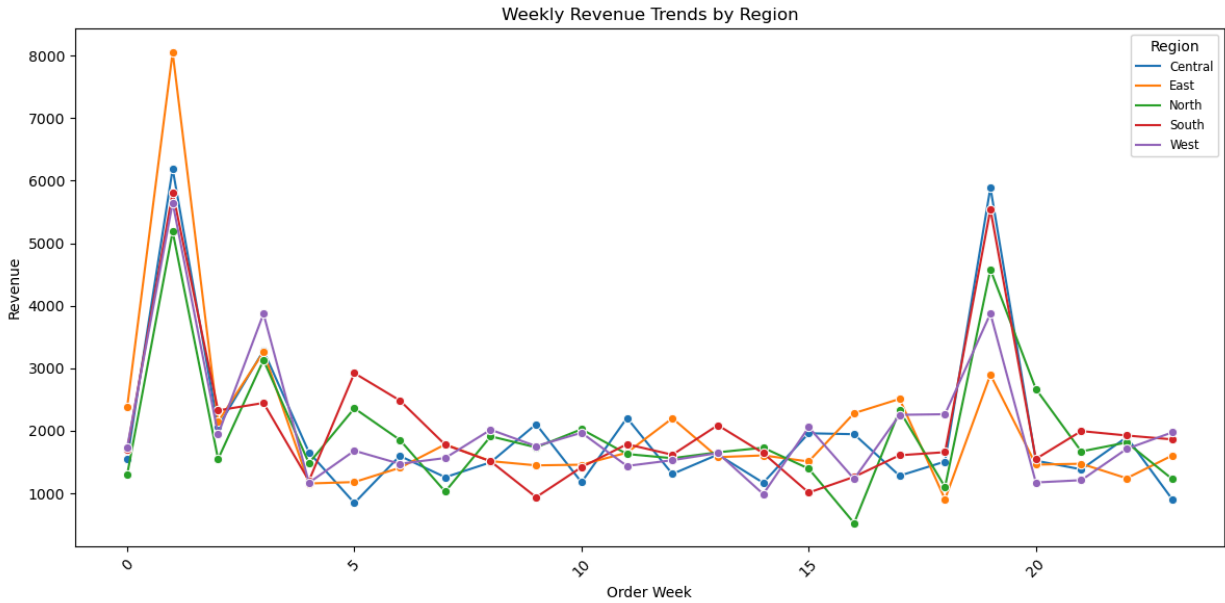
(Weekly Trends of Revenue by Order Region)

This refers to the location to which the order is being delivered, with the Western and Southern regions generating the most.



(Weekly Trends of Revenue by Customer Region.)

## Business Question Answers:

1. Cleaning products drive the most revenue, generating around £95000, with most of the orders coming from the Northern region.

| | category | revenue | quantity | discount_applied | customer_region |
|---|---|---|---|---|---|
| 0 | Cleaning | 94602.349000 | 3623.0 | 0.085791 | North |
| 4 | Storage | 47420.395000 | 1751.0 | 0.081357 | North |
| 2 | Outdoors | 40762.238321 | 1543.0 | 0.082490 | North |
| 1 | Kitchen | 34422.126500 | 1245.0 | 0.075184 | Central |
| 3 | Personal Care | 25159.302500 | 910.0 | 0.086842 | North |



Top 5 Product Categories by Revenue
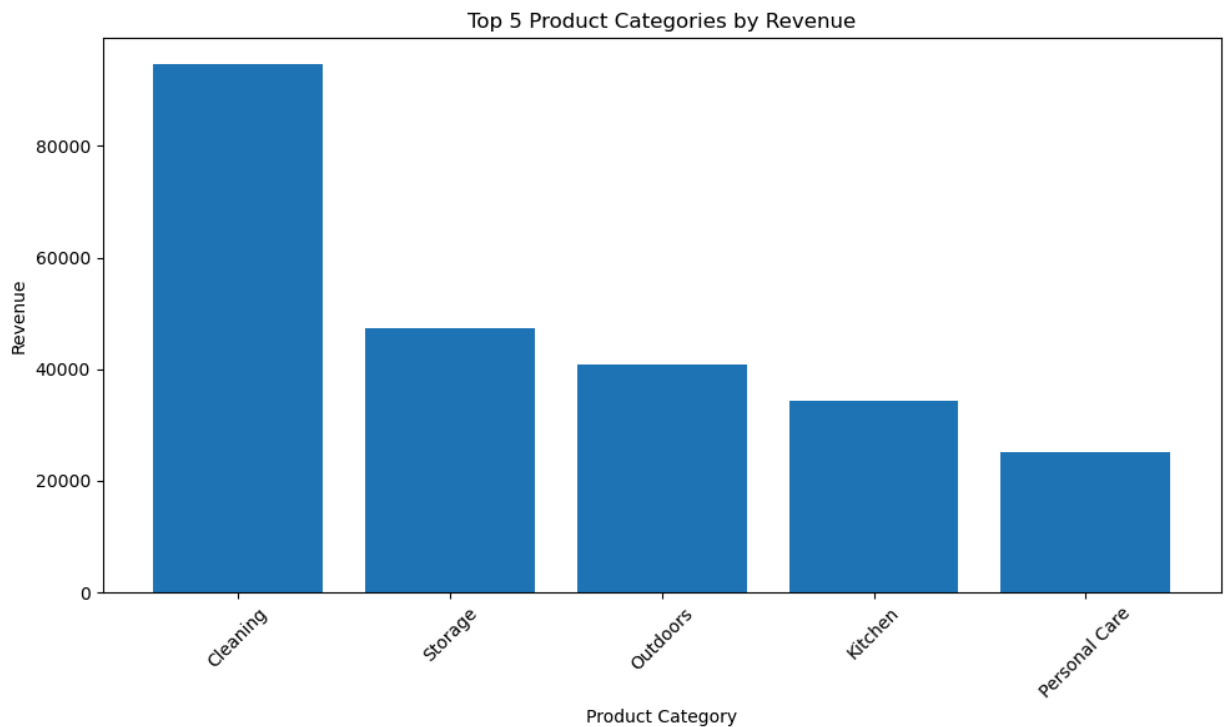
2. The graph below shows that discounts do lead to more products being sold. For cleaning and storage, a notable increase in the quantity is seen when the discount is applied.

Quantity vs Discount Applied Across Different Categories
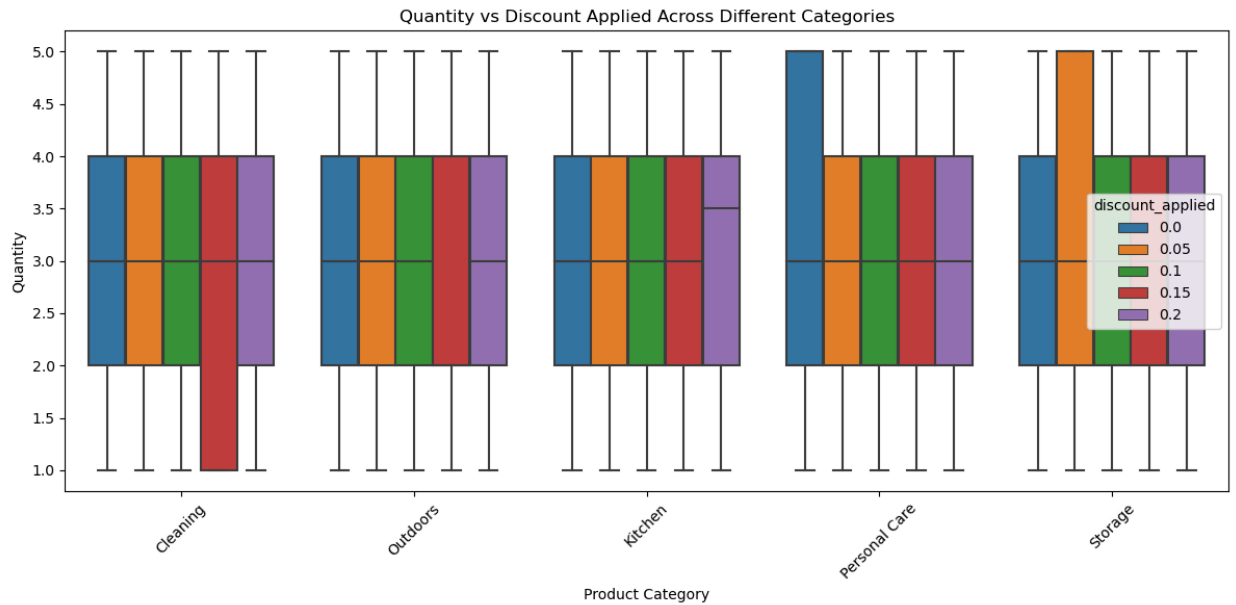
3. The gold tier generates the most value as it is the most preferred out of all the other options, with the central region having the most orders.


Count of Orders by Loyalty Tier with Customer Region
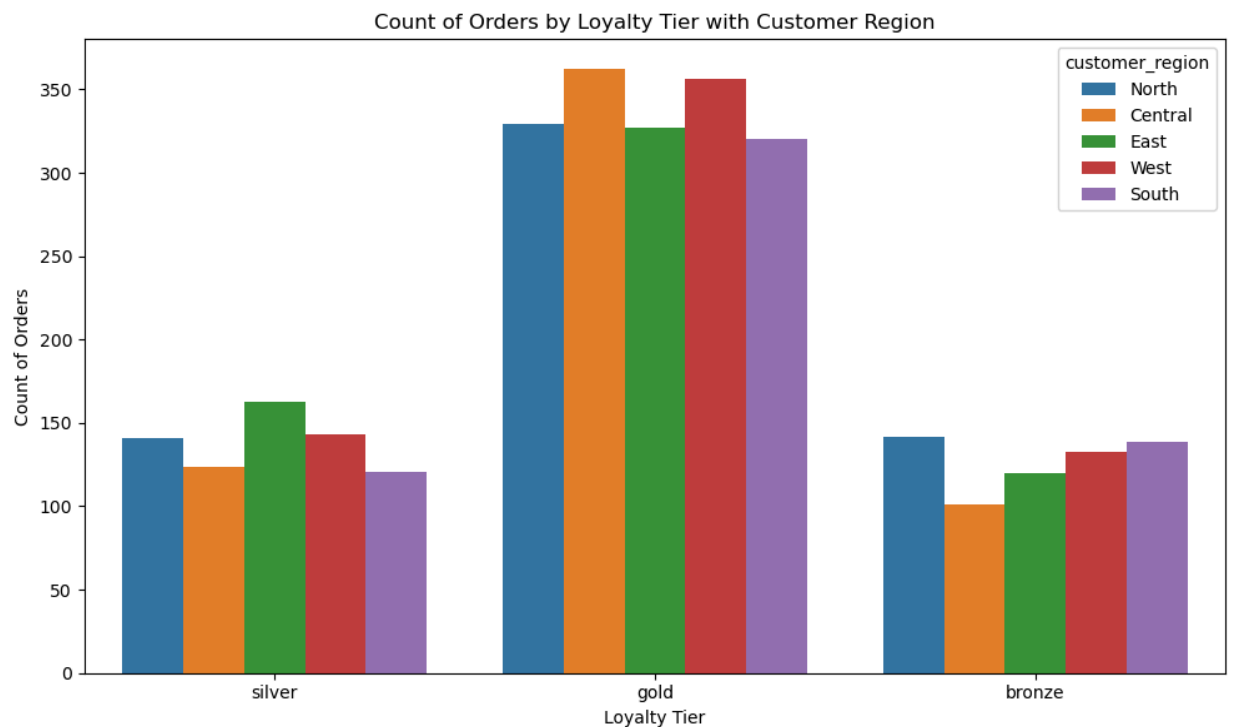
4. The data shows that all regions are struggling with delivery delays, with the Central, Southern and Western regions being affected the most.

| delivery_status | cancelled | delayed | delivered |
|---|---|---|---|
| customer_region | | | |
| Central | 109 | 245 | 233 |
| East | 115 | 212 | 283 |
| North | 111 | 237 | 264 |
| South | 125 | 235 | 220 |
| West | 133 | 258 | 241 |

5. Yes, as the number of signups correlates perfectly with the Christmas season, when customers will be more likely to make deliveries.

## Recommendations:

- The company should focus promotions on the 'Personal Care' category in the Northern region, especially during the winter season.
- Improvements should be made in the delivery of products to the Central, Southern and Western regions as they have higher delay rates than the others.

## Data Issues:

- Inconsistent labelling for regions, loyalty tiers, gender, delivery status and payment method
- Missing values for the majority of categorical columns.

## Solutions:

- Improve the data entry validation for customers, ensuring details are filled properly.