

YOU CAN START YOUR PROJECT FROM HERE

Introduction

This report is an analysis of recent sales and customer data across our regions and product lines. Three datasets were analysed together: sales data, customer info, and product info. Sales data consisted of 3000 entries, recording information including unit price, any discount applied, quantity and type of product ordered, the ordering date and region, the payment type, and whether delivery was successful or not. Customer info recorded 500 entries, detailing customer's emails, gender, region, loyalty tier, and signup date. Product info contains information regarding the 30 products we sell. Of particular note to us is which product category each product falls under. This report aims to showcase trends in revenue, customer behaviour, and delivery performance. It also aims to answer 5 specified business questions as requested, aiming to inform upcoming marketing and operational strategies.

Data Cleaning Summary

All 3 datasets were cleaned similarly. All texts were standardised to have only the first letter of each word capitalised. Missing text entries were replaced with "unknown" in most cases. Missing dates were filled using the previous entry's date. 'discount_applied' in the sales data had many missing values; these were assumed to be a discount of 0. Several spelling errors in the 'gender', 'region', 'loyalty_tier', and 'payment_method' columns were manually corrected. Duplicates were also removed, this only affected 1 entry in the sales data. A further 10 rows were dropped from sales data due to missing information which identified the entry (order_id, customer_id, product_id) and could not be inferred.

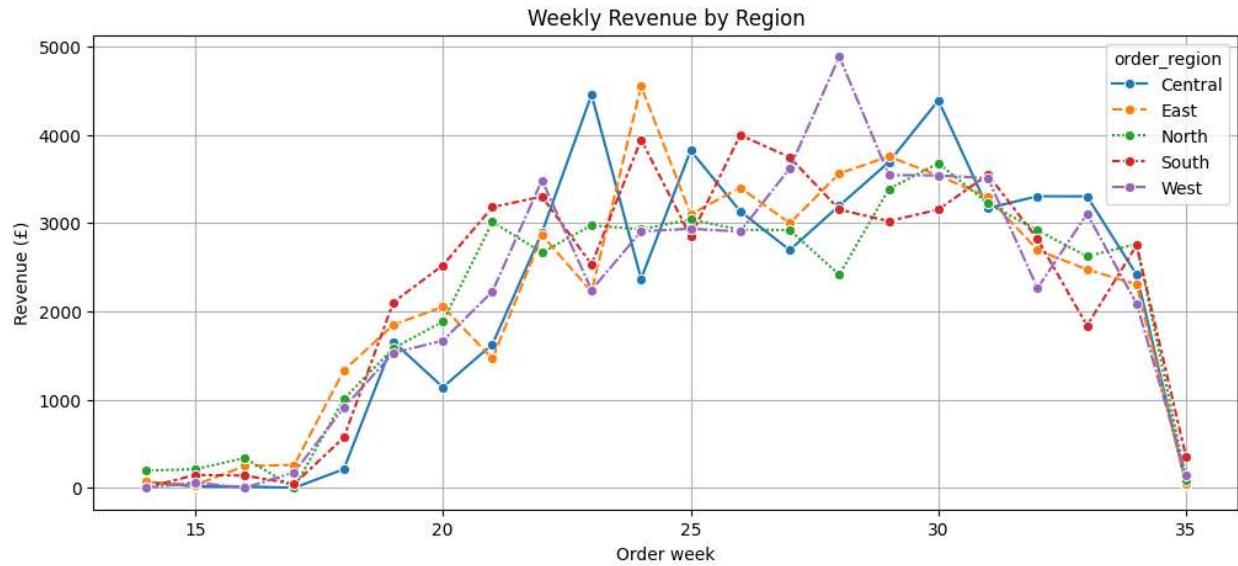
Feature Engineering Summary

To enable richer analysis, the 3 datasets were merged, and 7 new columns created. These were:

1. Revenue: calculated revenue of each transaction, taking discounts into account.
2. Order Week: the ISO week number each order was made in.
3. Price Band: Low, Medium or High, depending on the unit price (<£15 is Low, £15-30 is Medium, >£30 is High).
4. Days to Order: Number of days between product launch date and customer order date.
5. Email Domain: Email Domain of the ordering customer
6. Is Late: True or False depending on whether the delivery was delayed or not respectively.
7. Signup Month: Month and Year that the customer signed up in.

Key Findings and Trends

Here are some key findings and trends in addition to those answering a business question later.



From the figure above, we can see that weekly revenue in each region slowly increased for the first 6 weeks, then remained roughly in the range of £2500-4500, with a median weekly revenue of around £3000. The final week's numbers are likely incomplete as payments come in and orders are fulfilled, explaining why they are so low. This extends to the previous weeks' numbers to a smaller extent.

Table 1: Breakdown of Payment Methods by Loyalty Tier

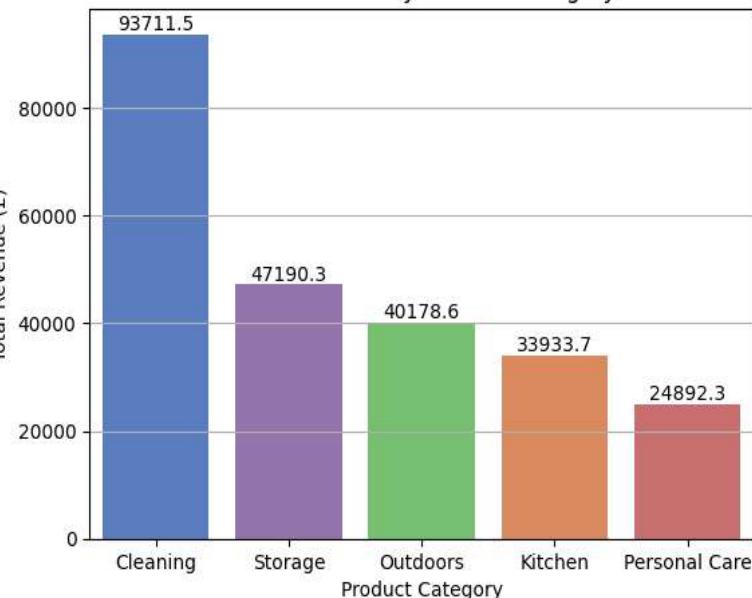
Loyalty Tier	Bank Transfer	Credit Card	Paypal	Unknown
Bronze	180	287	164	0
Gold	404	840	426	2
Silver	194	318	165	1
Unknown	0	6	3	0

The table above shows the number of times each payment type was used, separated by loyalty tier. We can see that across all tiers, credit cards are the most common payment type, accounting for about 50% of transactions in each tier. Between Bank Transfer and Paypal, Bronze and Silver tier customer seem to prefer bank transfers over Paypal, while Gold tier customers prefer Paypal, though the difference is small in each case.

Business Question Answers

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

Total Revenue by Product Category



From the figure on the left, we can see that **cleaning products generate the most total revenue**, followed by storage products and outdoor products. Personal care products generated the least revenue.

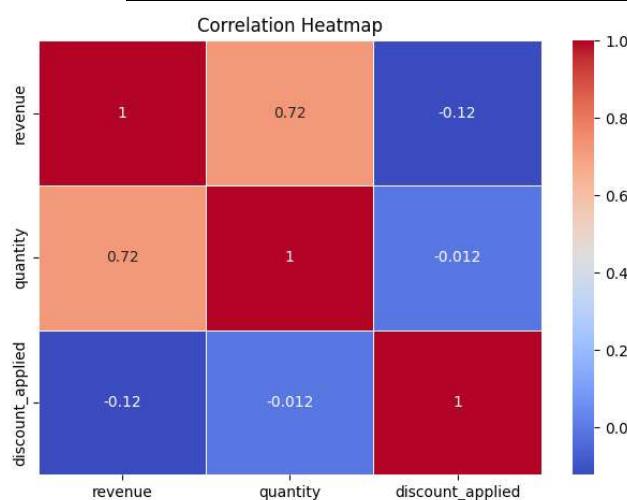
Table 2 below shows the revenue breakdown by region as well. We can see that the **overall trends and the regional trends are largely the same**. One other thing of note is that in the South, the revenue generated by storage products only barely beats out that by outdoor products. This is the only region with such a small difference between product categories.

Table 2: Revenue by Region and Product Category

Order Region

Product Category	Central	East	North	South	West
Cleaning	£18,018.84	£19,943.32	£18,766.10	£19,030.27	£17,952.93
Kitchen	£6,518.97	£6,695.84	£6,283.46	£7,761.40	£6,674.00
Outdoors	£7,931.85	£7,237.93	£7,806.33	£9,327.42	£7,875.10
Personal Care	£5,616.16	£4,616.20	£5,207.55	£4,239.03	£5,213.34
Storage	£9,511.04	£9,604.42	£8,715.05	£9,345.36	£10,014.45

2. Do discounts lead to more items sold?



No. The image on the left shows the correlation between revenue, quantity, and discount applied. If bigger discounts lead to more items sold, we would expect the correlation between quantity and discount applied to be positive and closer to 1 than 0. Here, it is -0.012, indicating that **larger discounts actually slightly lead to fewer items sold**. Note

that this analysis considers all items. Further analysis could reveal whether discounts affected individual products.

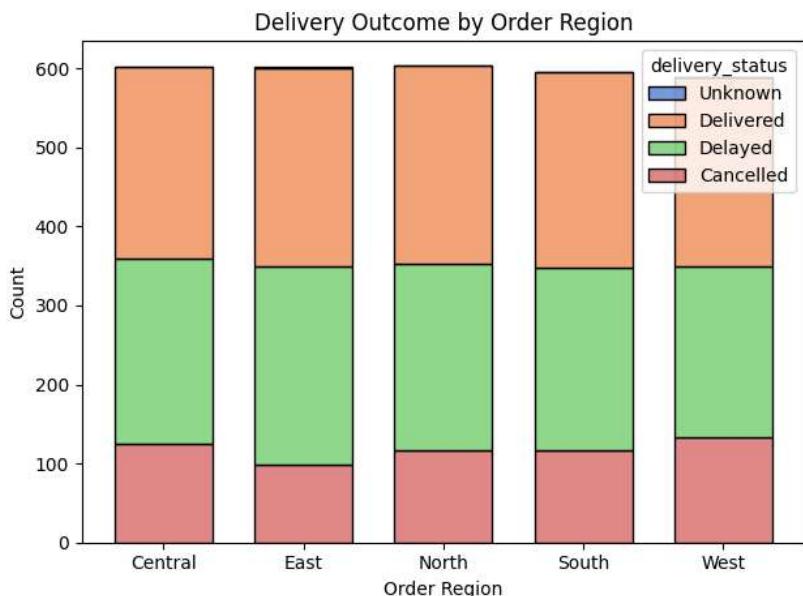
3. Which loyalty tier generates the most value?

Table 3: Revenue Breakdown by Customer Loyalty Tier

Loyalty Tier	Mean Revenue per order	Median Revenue per order	Total Revenue	No. of orders	No. of customers	Mean Orders per customer	Mean Revenue per customer
Bronze	£78.27	£63.70	£49,387.58	631	114	5.54	£433.22
Silver	£78.80	£65.12	£53,426.42	678	119	5.70	£448.96
Gold	£81.53	£69.77	£136,325.07	1672	265	6.31	£514.43
Unknown	£85.25	£73.30	£767.27	9	2	4.50	£383.64

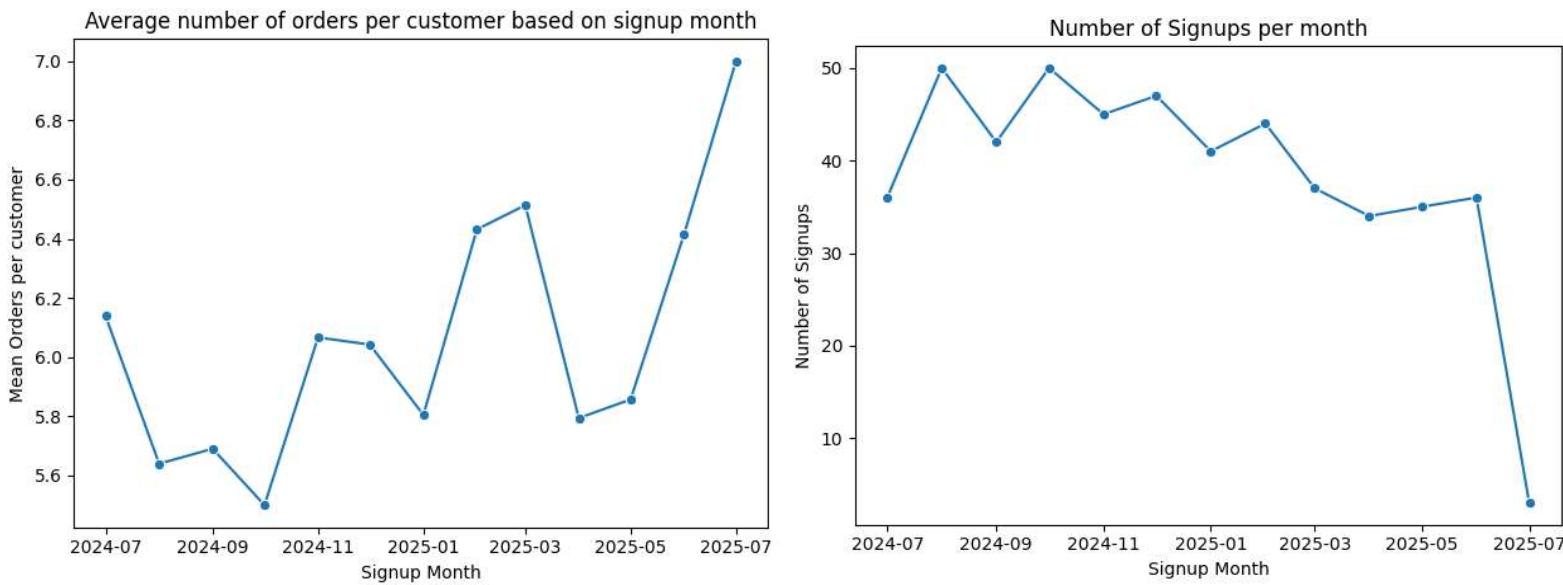
The table above shows revenue broken down by loyalty tier. We see that across all metrics, Gold tier customers achieve the highest numbers: the average Gold tier customer orders more often and spends more per order than the average Silver and Bronze tier customer. The data (not pictured here) also suggests that Gold tier customers order more per transaction on average (3.005 items per order) than Silver tier (3.004) and Bronze tier (2.977). This combines to give Gold tier customers the highest mean revenue per customer among the 3 loyalty tiers. Hence, it can be said that **Gold tier generates the most value**.

4. Are certain regions struggling with delivery delays?



The figure on the left summarises delivery outcomes per region. We can see that all regions experience around 40% successful delivery, 40% delayed delivery, and 20% cancelled delivery. Without information on how long the delays are, we are unable to say whether any region is experiencing worse delivery delays than others, but we can say that all regions experience delays around 40% of the time.

5. Do customer signup patterns influence purchasing activity?



The figure on the left shows the average number of orders per customer based on their signup month, while the figure on the right shows the number of signups each month. The final month is an outlier in the right figure, likely due to insufficient data since not enough time has passed.

We can see from the left that customers who signed up in Feb, Mar and Jun 2025 have the highest average number of orders per customer. We see on the right that these months had some of the lowest number of signups. Similarly, Aug-Oct 2024 had low numbers of orders per customer, and the similar period had fairly high signup rates.

It seems that **customer signup patterns could influence purchasing activity negatively**, with higher signup rates leading to decreased ordering per customer. Further statistical testing will be needed to confirm this.

Recommendations

Based on the above findings, we recommend the following:

1. Improve delivery reliability in all regions

Our analysis showed that across all regions, around 40% of orders were delayed. Improving this will help improve customer retention and experience.

2. Reconsider discounts

Our data showed that discounts did not lead to increased quantity of goods sold overall. Further analysis should be done product by product to see if this is true for all products,

or if this is the result of discounts increasing the quantity of one product sold but decreasing that of another product. We note that in the data, the maximum discount was 20%. Higher discounts could be considered if the demand for the product is truly price elastic.

3. Focus marketing on Gold tier customers

Gold tier customers already spend more and order more compared to other customers, so increased marketing to them is likely to increase their spending even more.

Data Issues/Risks

There were a number of spelling errors in the data, specifically in the categorical data (loyalty_tier, gender, region). This could be fixed during data entry by ensuring that the entered data meets one of the categories (e.g. for loyalty_tier, ensure that the entry is “Gold”, “Silver” or “Bronze”), and flagging it otherwise. If feasible, the method of entry could be changed from a freeform text field to a dropdown menu. This would eliminate all spelling errors from categorical data.