Project Coversheet

Full Name	Sai Kiran Dhandi
Email	Saikirandhandi25@gmail.com
Contact Number	+353 892402305
Date of Submission	20-07-2025
Project Week	Week - 2

Project Guidelines and Rules

1. Submission Format

- **Document Style**:

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

- **File Naming**:

  - 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**:

  - Submit your report as a **PDF**.
  - 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.

Introduction

The goal of this project was to analyze sales and customer behavior for a UK-based company selling eco-friendly products. I was provided with three datasets customers, products, and sales and tasked with cleaning and merging them, creating relevant features, and uncovering insights around revenue trends, customer acquisition, and regional performance. This report summarizes key patterns in the data, answers specific business questions, and highlights areas for improvement.

Data Cleaning Summary

Customers

- Filled missing customer_id values by inspecting and correcting them manually.
- Filled missing email values with 'unknown'.
- Replaced missing sign_up_date values with the most frequent date and converted the column to date_time.
- Standardized labels in gender, loyalty_tire, customer_region, filling nulls with each column's mode.
- Final cleaned dataset: 500 customers.

Products

- Cleaned dataset includes 30 products across 5 categories.
- The Cleaning category had the highest number of products.
- Converted launch_date to datetime.

	category	product_count
0	Cleaning	12
1	Kitchen	4
2	Outdoors	5
3	Personal Care	3
4	Storage	6

Sales

- Dropped records with missing order_id, product_id, or customer_id to maintain join integrity.
- Cleaned and standardized quantity, converted to float, filled nulls with median.
- Cleaned delivery_status, payment_method, and sale_region, filling missing values with mode.
- Filled discount_applied nulls with 0%.
- Removed duplicate orders and converted order_date to datetime.

Feature Engineering Summary

- **Revenue** : quantity × unit_price × (1 - discount_applied)
- **price_band** : Categorized unit price as Low if < 15, Medium if 15–30, High if > 30
- **days_to_order**: Number of days between product launch and order date.
- **email_domain**: Extracted domain from customer email.
- **is_late**: Boolean indicating if delivery status is marked as delayed.
- **same_region_customer**: Boolean flag for whether sale region matches customer region.

Summary

Weekly revenue trends by region

	sale_region	order_week	revenue
0	Central	27	47575.17
1	East	27	48153.20
2	North	27	46778.46
3	South	27	49732.13
4	West	27	47729.74

Product category performance by revenue

	category	sum	mean	min	max
0	Cleaning	93814.86	77.98	6.96	248.30
1	Kitchen	33933.66	84.41	7.76	249.15
2	Outdoors	40159.47	79.21	6.80	247.55
3	Personal Care	24892.25	82.70	7.50	248.10
4	Storage	47168.46	81.75	7.75	249.35

Product category performance by quantity

	category	sum	mean	min	max
0	Cleaning	3591.0	2.99	1.0	5.0
1	Kitchen	1226.0	3.05	1.0	5.0
2	Outdoors	1523.0	3.00	1.0	5.0
3	Personal Care	900.0	2.99	1.0	5.0
4	Storage	1736.0	3.01	1.0	5.0

Product category performance by discount applied

	category	mean	min	max
0	Cleaning	0.09	0.0	0.2
1	Kitchen	0.08	0.0	0.2
2	Outdoors	0.08	0.0	0.2
3	Personal Care	0.09	0.0	0.2
4	Storage	0.08	0.0	0.2

Top 5 weeks of customer acquisition

	signup_week	customer_count
5	2024-08-05/2024-08-11	135
13	2024-09-30/2024-10-06	97
31	2025-02-03/2025-02-09	95
33	2025-02-17/2025-02-23	89
17	2024-10-28/2024-11-03	88

Delivery performance by region and price band

price_band	High			Low			Medium		
delivery_status	Cancelled	Delayed	Delivered	Cancelled	Delayed	Delivered	Cancelled	Delayed	Delivered
customer_region									
Central	37	104	109	22	36	37	51	92	86
East	55	110	133	22	33	46	38	72	109
North	60	116	123	13	43	41	36	77	94
South	52	99	97	28	41	27	45	92	95
West	58	116	119	23	34	44	49	104	72

Preferred payment methods by loyalty_tier

payment_method	Bank Transfer	Credit Card	PayPal	credit card
loyalty_tier				
Bronze	180	154	164	133
Gold	406	427	429	419
Silver	195	153	165	165

Customers are from same sale region

	same_region	customer_count
0	False	2404
1	True	586

Total orders by each category

category	
Cleaning	1203
Storage	577
Outdoors	507
Kitchen	402
Personal Care	301

Category products purchased in each region

| category	Cleaning	Kitchen	Outdoors	Personal Care	Storage
customer_region					
Central	215	74	109	51	125
East	271	79	89	68	111
North	251	90	108	48	106
South	216	83	91	68	118
West	250	76	110	66	117

Top 5 products by Revenue

Top 5 Categories by Revenue

Quantity vs Discount Across Categories

Correlation Heatmap: Revenue, Discount, Quantity

Order Count by Loyalty Tier and Customer Region

Order Count by Loyalty Tier and Sale Region

Delivery Status by Price Band
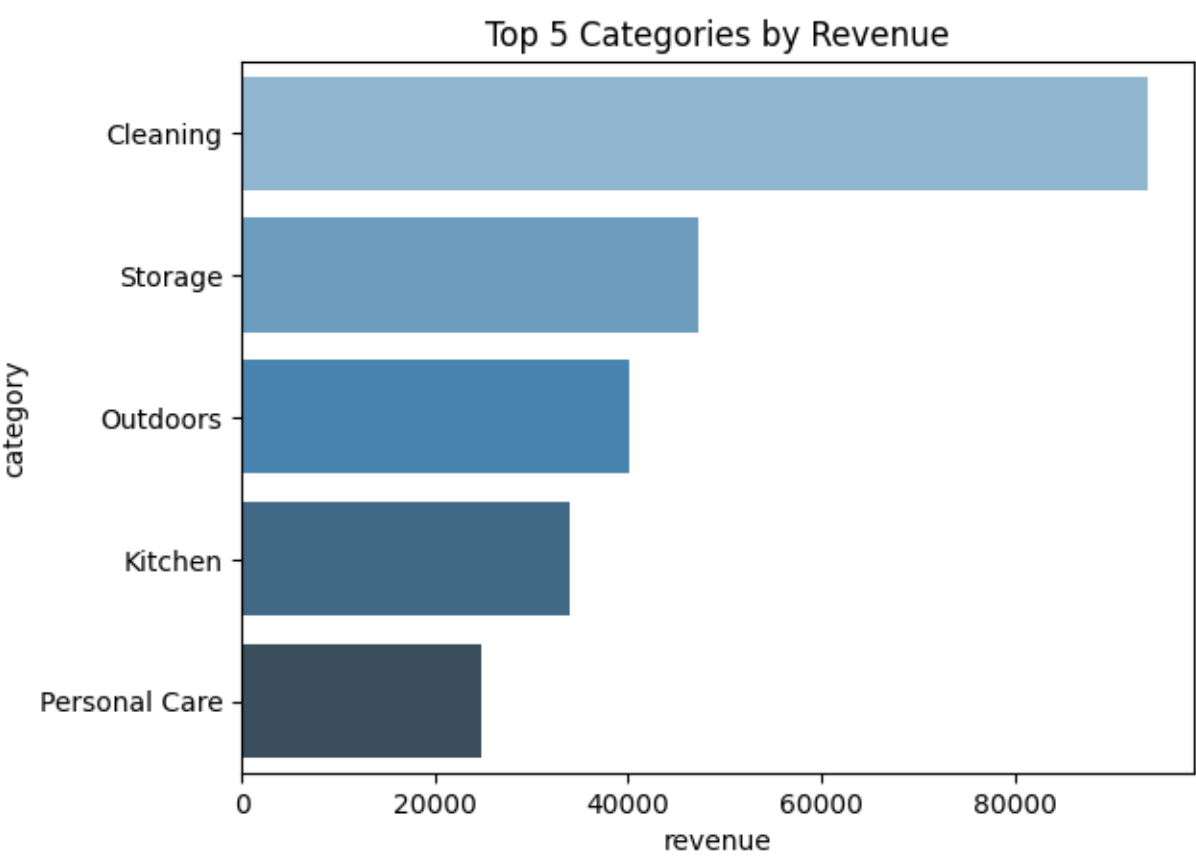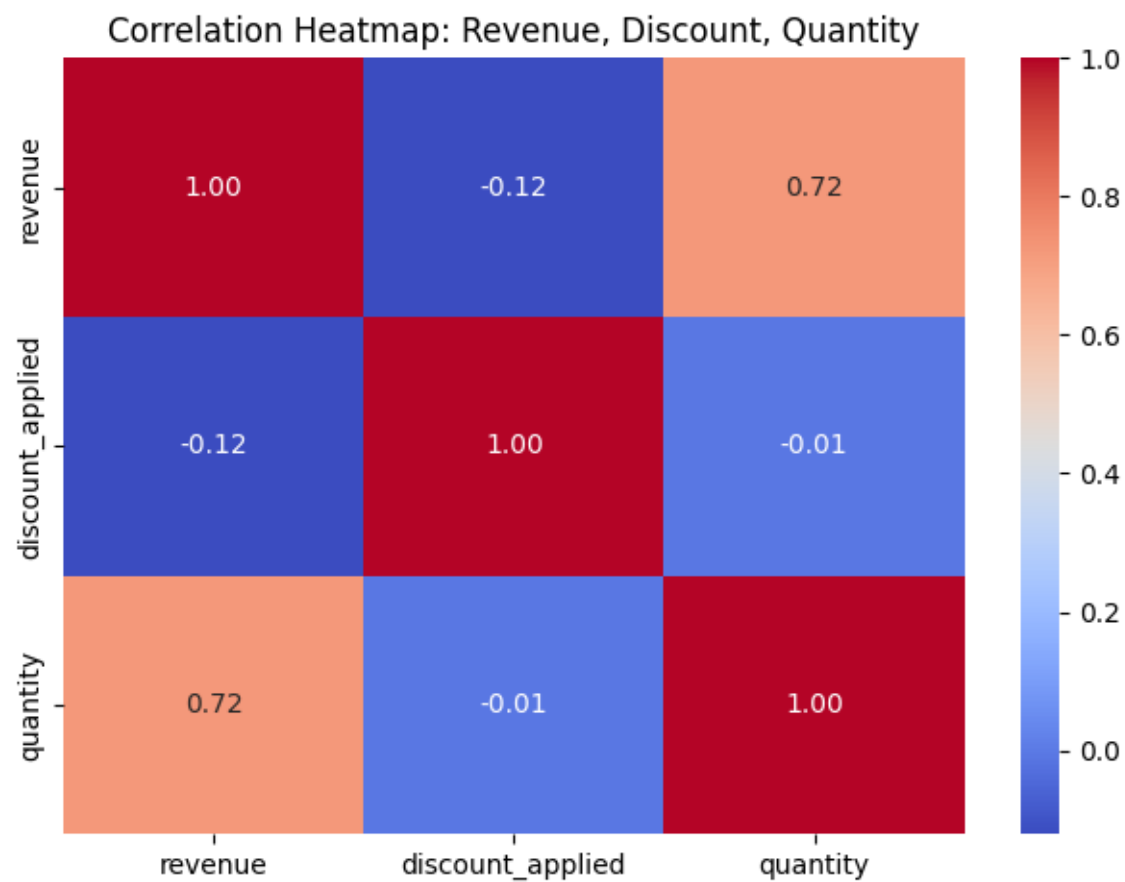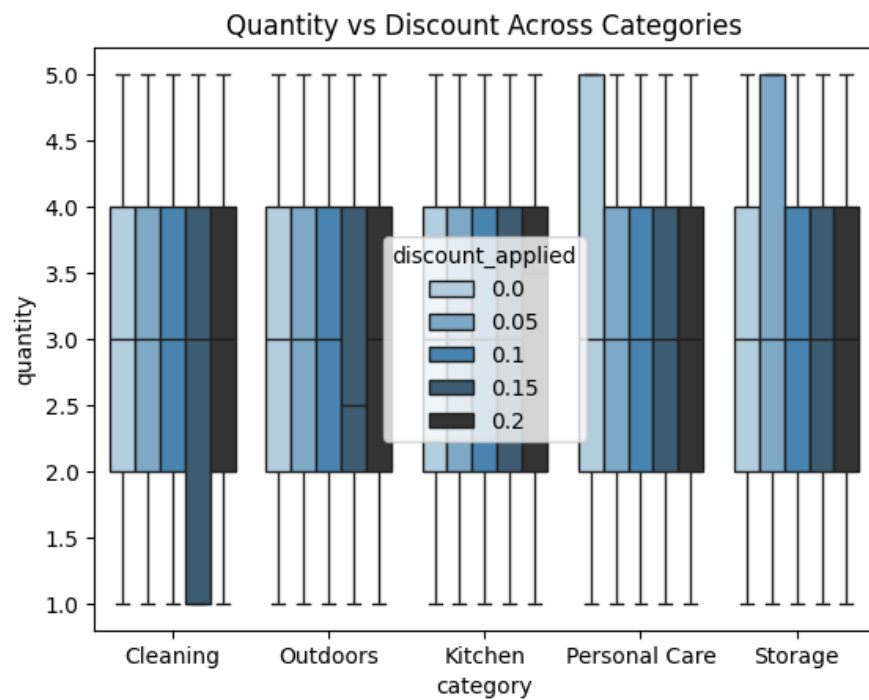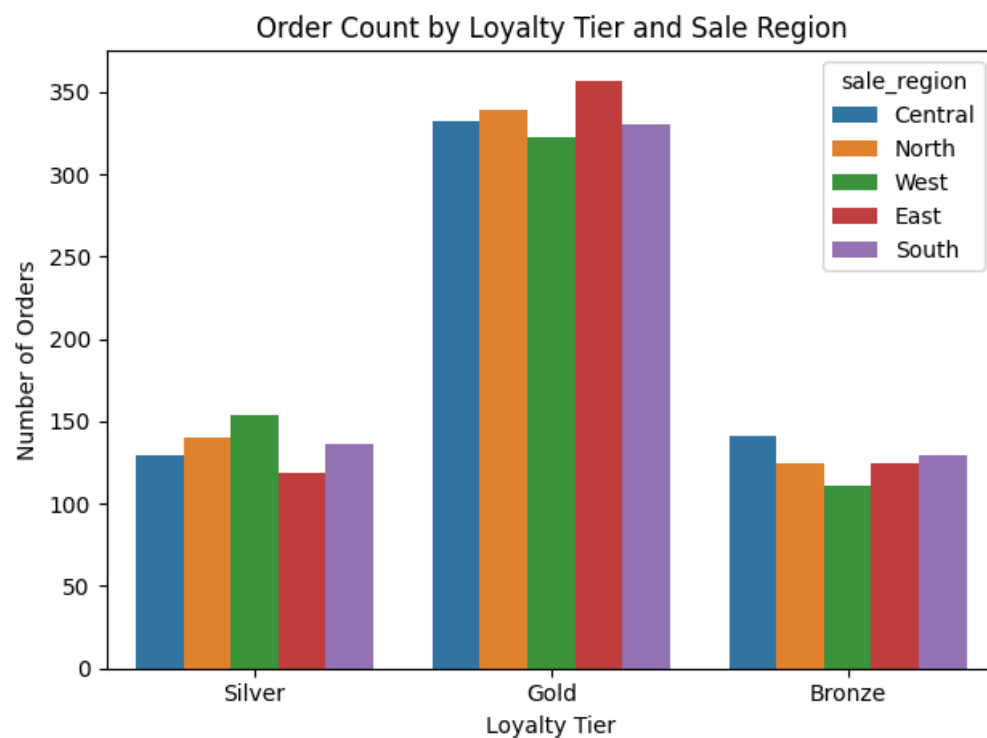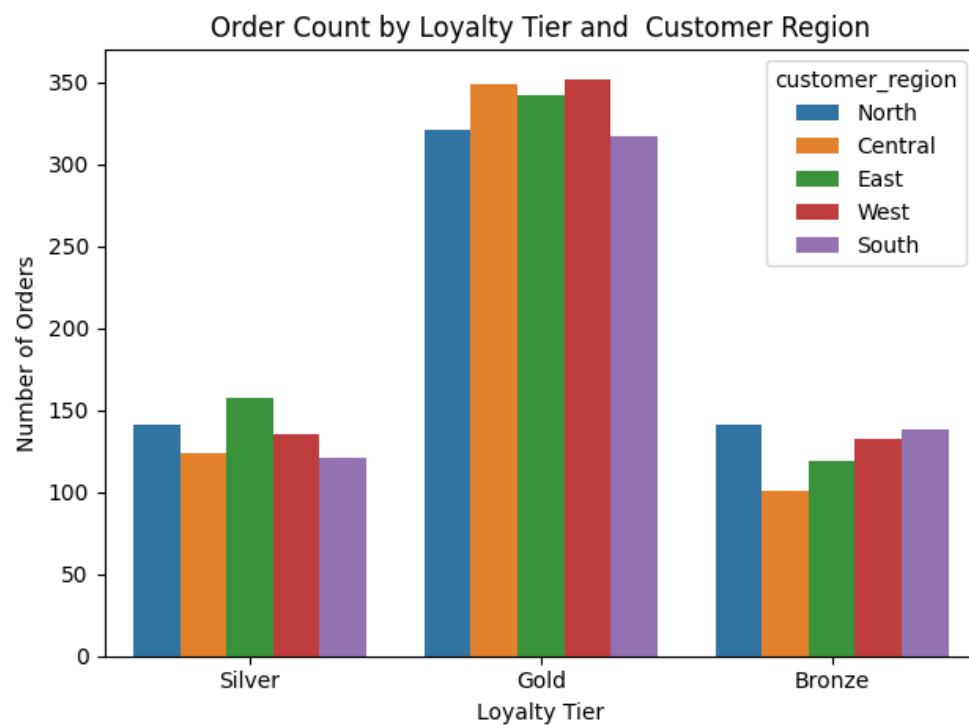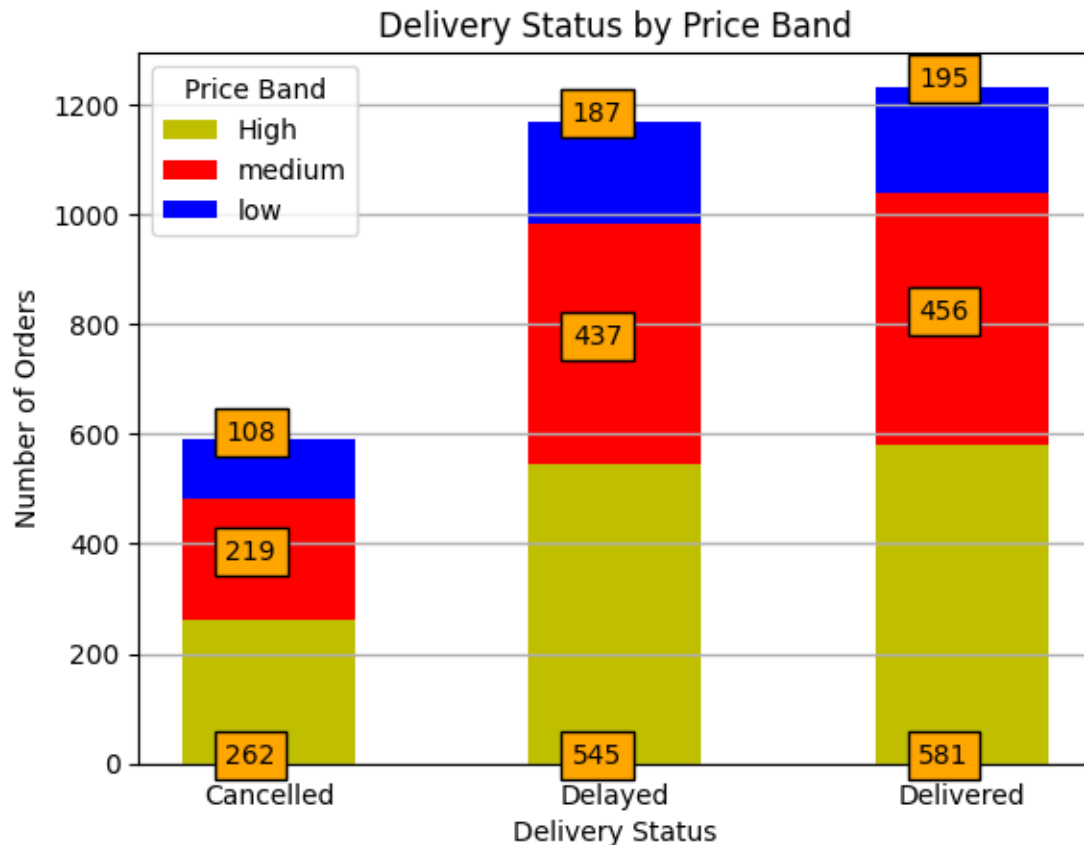
Business questions answered

- Cleaning category in east region has brought us most revenue.
- Discounts have not affected sale of any product.
- Gold loyalty tire has most valuable customer base.
- Customers who signed up in 2nd and 8th month has more purchases compared to other months.
- By looking at the numbers we can conclude that east region has delayed delivery issues compared to other regions.

Recommendations

- Improve delivery logistics in all regions especially in east region
- Personal care category products needs to be advertised more as they are very low in sales.

Data Issues or Risks

- Found inconsistent values and missing data in key columns like customer_id, quantity, and delivery_status.
- Enforce non-null constraints on order_id, customer_id, and product_id