Project Coversheet

May 23, 2025

Full Name	Jude Spellacy
Email	judespellacy@gmail.com
Contact Number	07891187181
Project Title	Week 1

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.

Project 1: Data Analysis for Business Insights

1 Data Cleaning

- There was a missing figure for the total sales for customer with ID C001 on 05/01/2024. This was filled by directly calculating with quantity and price as this was the trend observed in the other rows.
- Two customer IDs were repeated with the exact same purchases in both cases, it is possible these are repeat customers. Since this behaviour was unusual within the dataset it was instead assumed an error. This was atypical behaviour for the sample as usually electronics are bought individually, if left in C002 would be well outside of the range of all other electronics purchases.
- It was also assumed that in both cases the first recorded transaction was legitimate and the second a repeat. Since it is unlikely to be the other way around. The repeated sales on 10/02/2024 and the 05/04/2024.

Transaction_ID	Date	Customer_ID	Product	Category	Quantity	Price	Total_Amount	Payment_Method	Region
1001	2024-01-05	C001	Laptop	Electronics	1	800	nan	Credit Card	North
1002	2024-01-10	C002	Smartphone	Electronics	2	600	1200.0	Cash	South
1003	2024-01-12	C003	Headphones	Electronics	1	100	100.0	PayPal	West
1004	2024-02-05	C004	Tablet	Electronics	1	500	500.0	Debit Card	East
1005	2024-02-08	C005	Book	Books	3	20	60.0	Credit Card	North
1006	2024-02-10	C001	Laptop	Electronics	1	800	800.0	Credit Card	North
1007	2024-03-15	C006	Shoes	Clothing	2	50	100.0	Cash	South
1008	2024-03-18	C007	T-Shirt	Clothing	1	25	25.0	PayPal	West
1009	2024-03-20	C008	Smartwatch	Electronics	1	200	200.0	Debit Card	East
1010	2024-04-01	C009	Book	Books	2	20	40.0	Credit Card	North
1011	2024-04-05	C002	Smartphone	Electronics	2	600	1200.0	Cash	South
1012	2024-04-10	C010	Tablet	Electronics	1	500	500.0	Debit Card	East
1013	2024-05-01	C011	Shoes	Clothing	1	50	50.0	Cash	South
1014	2024-05-05	C012	Headphones	Electronics	1	100	100.0	PayPal	West
1015	2024-05-08	C013	Laptop	Electronics	1	800	800.0	Credit Card	North
1016	2024-05-10	C014	T-Shirt	Clothing	3	25	75.0	Cash	South
1017	2024-06-01	C015	Smartwatch	Electronics	1	200	200.0	Debit Card	East
1018	2024-06-05	C016	Book	Books	4	20	80.0	Credit Card	North
1019	2024-06-08	C017	Smartphone	Electronics	1	600	600.0	Cash	South
1020	2024-06-10	C018	Tablet	Electronics	1	500	500.0	Debit Card	East

Figure 1: Original dataset before cleaning with the marked cells removed in the cleaned dataset.

2 Exploratory Data Analysis and Data Visualisation

	Product	Category	Quantity	Price	Total Amount	Payment Method	Region
Mean	-	-	1.56	283.89	329.44	-	-
Median	-	-	1	150	150		-
Mode	Book, Tablet	Electronics	1	20, 500	100, 500	Credit, Debit, Cash	East, North, South
Range	-	-	3	780	1175	-	-
Standard Deviation	-	-	0.92	287.67	343.58	-	-

Indicated in this table is statistical values evaluated per transaction, for example the modal product being Tablet and Book is because they are sold in the most transactions but taking into account quantity per transaction Book becomes the modal product. It is implied that for the sample payment method was relatively consistent apart from PayPal and that sales are relatively even between all regions apart from West.

Category (Prices)	Electronics	Books	Clothing
Mean	445.45	20	37.5
Median	500	20	37.5
Mode	500	20	50/25
Range	700	0	25
Standard Deviation	258.32	0	14.43

Category (Quantity)	Electronics	Books	Clothing
Mean	1.09	3	1.75
Median	1	3	1.5
Mode	1	2	1
Range	1	2	2
Standard Deviation	0.30	1	0.96

Most electronics sell in single quantities per transaction, the modal quantity for Electronics and Clothing is 1. Books sell in a higher quantity per purchase.

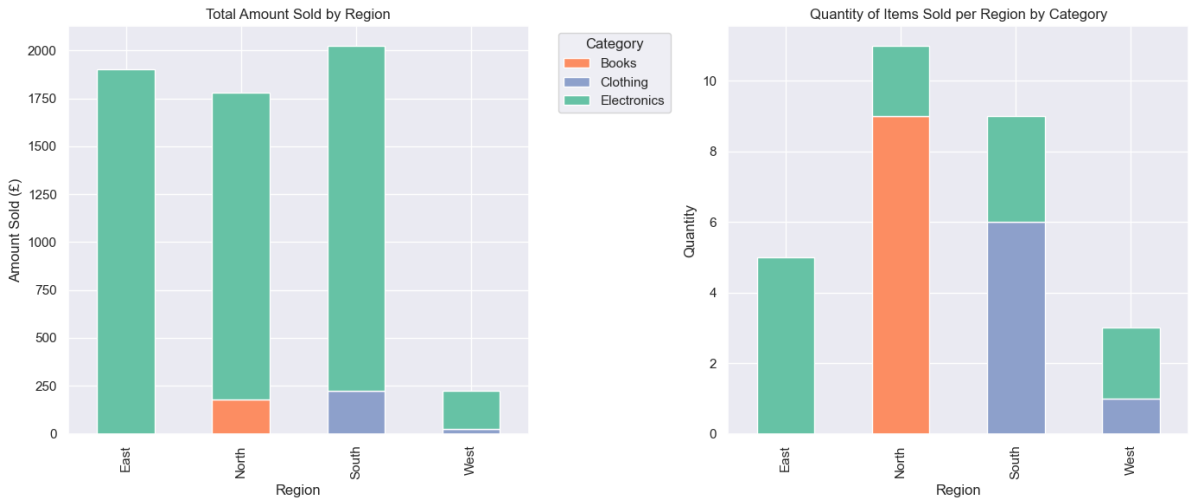

Figure 2: Distribution of summed sales per Region over all categories

In the figure for sales distribution over regions it can be seen that although total sales are similar between East and West, sales made in the East were worth more. It is also clear that the majority of

sale value comes from electronics especially in the North and South where book and clothing sales were 82% and 67% respectively.

From the sample of sales data over the period, there is indication of a normal distribution of sales over price for electronics, and over all categories a skewed distribution towards the least expensive products with the mean price of products sold being £211.79.

Figure 3: Distribution of prices in each purchase vs how many were sold in total by category and payment method

The distribution of total sales amount per month over the period could be indicating a skewed distribution towards January and June with the lowest sales value in March. The total quantity of sales in each month stays quite consistent between January and April with January having a much higher proportion of Electronics sales. May and June both have a higher total of sales and a higher proportion of those is in electronics, up to 33% and 43% from 25% in February and March.

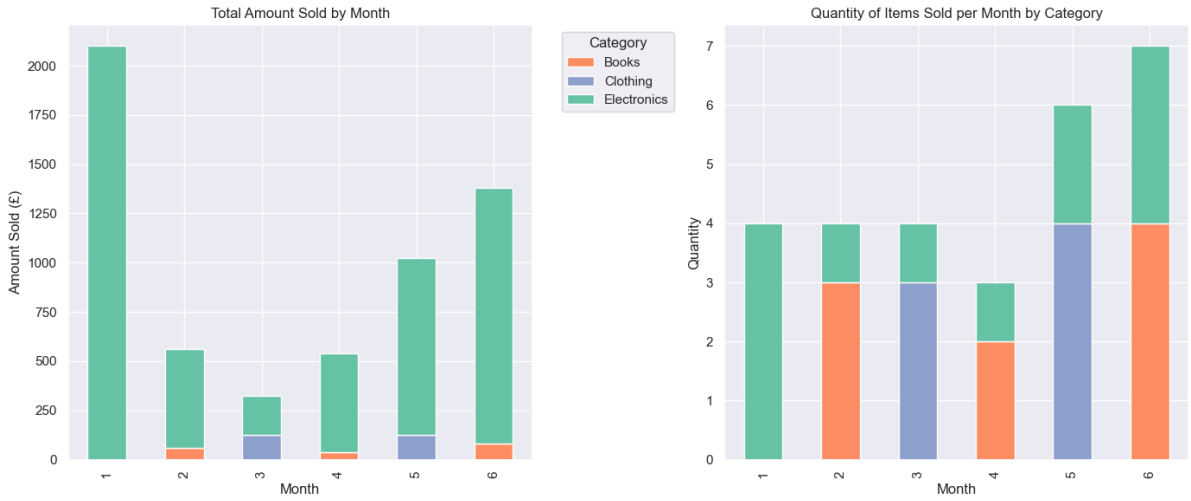

Figure 4: Distribution of summed sales per month over all categories

Within Categories it can be shown over the sample that Smartphones and Shoes make up a smaller share of sales 25% and 42.9% but represent as highest in total value 32.7% and 60%. Headphones and T-shirts return less value, this is not necessarily indicative of their profit margin.

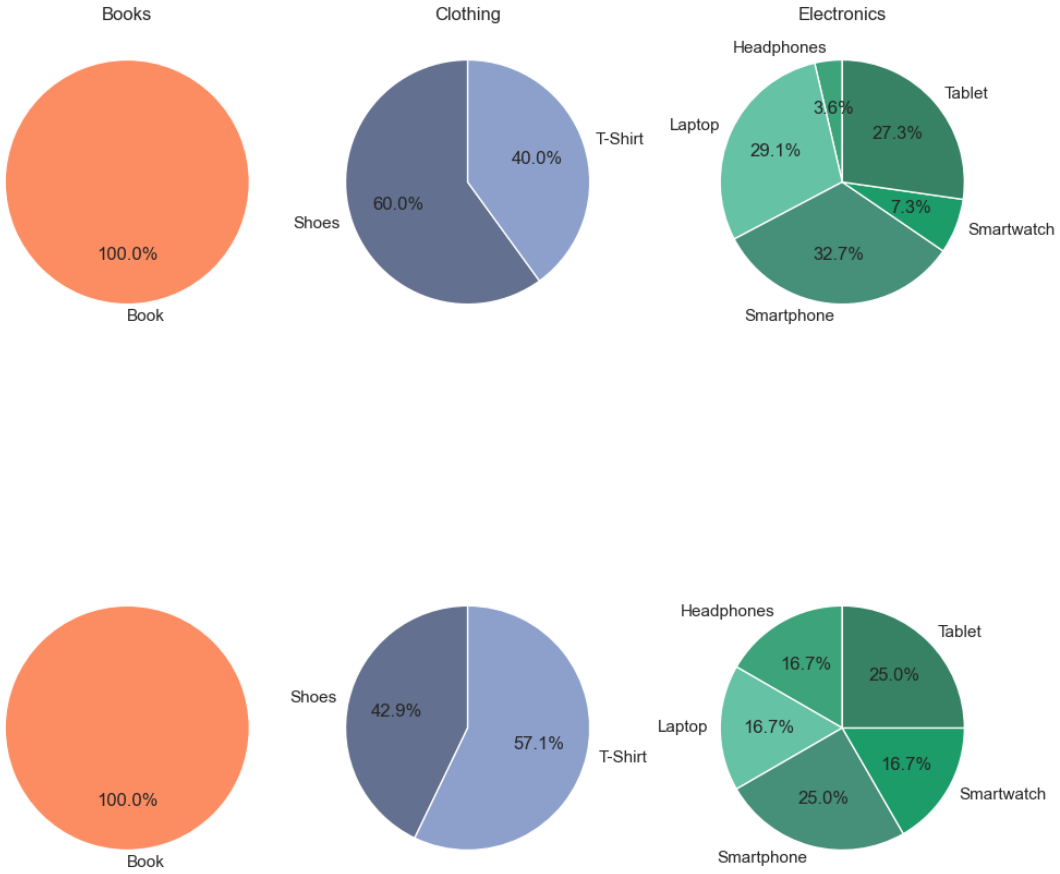

Figure 5: (1) Total amount in sales (£) and (2) Quantity of sales as a percentage of their category.

In summary, there is a relatively even purchase pattern between regions in electronics which generate the most in total sale value. Electronics outsell other products in January sell more in late spring and early summer than in early spring. Books sell at much higher quantities in individual sales and are only recorded sold in the North region. Clothing sales are recorded in the south and west over the sample but sell more frequently in the South than the West.

3 Data Insights Report

3.1 Key Findings

From the analysis for the sample provided:

- Electronics are the best selling item across all regions and months, both for quantity sold and value.
- Per region the best selling items vary, Electronics in the East and West, Books in the North, and Clothing in the South.
- By month Electronics are most popular in January, and Clothes and Books are alternately the most popular month to month from February to June.
- The majority of value from sales comes from Electronics for all months and regions.
- Most transactions were made with Cash, Credit, and Debit cards. With credit cards used in the sample exclusively on the most expensive and least expensive items.
- Total sales figures (value) were skewed towards Winter and early Summer, with the lowest performing month of March. March did not contain the least transactions, which was April.
- The distribution of purchases was skewed towards the cheapest items, books, clothing, headphones, and smartwatches. Only Electronics exceeded the mean transaction value of £329.44, with the average value of an item sold being £211.79.

3.2 Recommendations

January sales suggest that customers are more likely to purchase electronics in January this could indicate a post-Christmas sale, however there is not enough data to indicate confidently the specific reason. Pushing for more offers on Electronics in this month may capitalise on the increase, specifically it is recommended to focus marketing on electronics over winter.

Seasonally the lowest customer sales months for total spending value were in late winter and early spring, but purchase quantity suggests that these months are consistent in bringing in customers. It is recommended that offers on electronics that bundle with book sales could be useful as credit card purchases are recorded in this period on both books and electronics.

Regional marketing and specialising stock is recommended specifically electronics in the East and books in the North where these are already good sellers, offering a wider range could bring an increase in sales in these categories. Clothing in the South and West would benefit from a wider range, potentially offering luxury goods alongside existing products could make up for the disparity in value of sales.

Offering products like Kindles in the Northern region may work well based on purchase patterns in electronics and books. Depending on whether sales patterns persist up to September, a back to school or university offer on books and electronics could boost sales in the North.

Within categories it could be useful to market or bundle laptops with better selling items, such as headphones and smartwatches which represent a smaller share of value. This could potentially involve taking a reduced profit or even a small loss on these items in exchange for boosting sales quantities for laptops which generate a greater share of value.

In order to compensate for decreased value in sales for March, it may be beneficial to take advantage of footfall in the store as sales quantities are still high. Switching strategies in this period to market clothing or offer new products (e.g. snacks, accessories) at a lower cost could encourage more frequent sales.

3.3 Ethical Considerations

The data in use here is relatively standard for a company to keep track of, there is an implicit consent in a transaction that a receipt will be kept by the business. The ethical considerations come in when the data is used long-term by a business or third party to take action.

Keeping data long-term and using them for data analytics on payment methods, region, and products leaves purchases potentially traceable for third parties. If this information was communicated to customers clearly with for example signs on check out, then it could offset this problem.

It is also potentially a concern to capitalise on information such as credit card use on products, often credit card users are more vulnerable to overspending and utilising this information to get them to spend more is potentially ethically wrong.

Ensuring that data use is legal and ethical by using existing standards of practice is recommended.

3.4 Potential Strategies

In order to improve performance in the lowest performing months of February, March, and April, based upon the sample the following strategies are suggested:

Offer a wider range of products in books in the North and clothing in the South and West, marketing these products heavily after January sales begin to fall. By offering a more diverse range of products and potentially prices in these categories it may take advantage of the likely consistent footfall based on sales quantity between these months and January. Offering some clothes at a higher price point could close the gap in value between electronics which sell more seasonally.

Potentially focusing regions in specific categories could lower costs and allow a greater range of products in each store. Having the East focus on Electronics only may be advantageous if the sample is representative. Further to this focusing marketing and offers on laptops, smartphones, and tablets over the other electronics or bundling them with other popular lower value products may generate a higher income. This could then generate increased interest in electronics over the slower months, and pushing marketing on the middle to late spring could shorten the gap in sales between February and May.

Extending any offers or campaigns from January might lead to more sales in February in electronics, this is dependent on whether the trend seen in the data is actually indicating an offer rather than purely seasonal customer behaviour.