Project Coversheet

Full Name	Anagha Indulal Nair
Email	anaghaindulalgopal@gmail.com
Contact Number	+44 7407106517
Project Title (Example – Week1, Week2, Week3)	Week1: Data Analysis for Business Insights

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.

The dataset sales.xlsx contained 20 records of sales data.

1) Cleaned Dataset

- **Duplicates**: There were no duplicate records on preliminary examination. Upon checking, 0 duplicates were found as none of the Transaction IDs were duplicated.
- **Missing values handling**: Since the dataset given was manually examinable, there was only one missing value. Programming showed the same. One missing value in 'Total_Amount' for customer ID C001 for Transaction ID 1001. Total_Amount imputed as **Quantity × Price**.
- **Data types standardised**: Date parsed to datetime and reformatted DD/MM/YYYY. Quantity, Price, Total_Amount are converted to numeric.
- Categorical fields: Whitespaces were stripped and cases were standardized for categorical fields:- Product, Category, Payment Method and Region.
- **Outliers**: 0 outliers were found in Price and Total Amount.
- **Missing Dates**: There was an issue with missing dates for the month of march which was handled by manually examining and replacing the month as 3. Extracted month number by adding and extra column and ''Month_Extracted'' and changed data type to integer.

   The cleaned dataset was saved as "sales_data_cleaned". Now the dataset is clean and ready for EDA.

2) Exploratory Data Analysis – Summary

- **Understanding the dataset**: The file was loaded into a pandas dataframe and its structure was inspected. The dataset contains 11 fields:-
    - Date (DD/MM/YYYY), Transaction and Customer IDs
    - Categorical features:- Product, Category, Payment Method, Region,
    - Numerical features:- Quantity, Price and Total Amount.

- **Key Trends and Patterns:** We see that the information given to us was from January to June. It's a 6 months timeframe.
    - **Monthly Sales pattern:** Monthly sales peaked in January with the greatest revenue return of 2100, with its lowest dip in march to 325 and rebounded in April to 1740. (Figure 1)
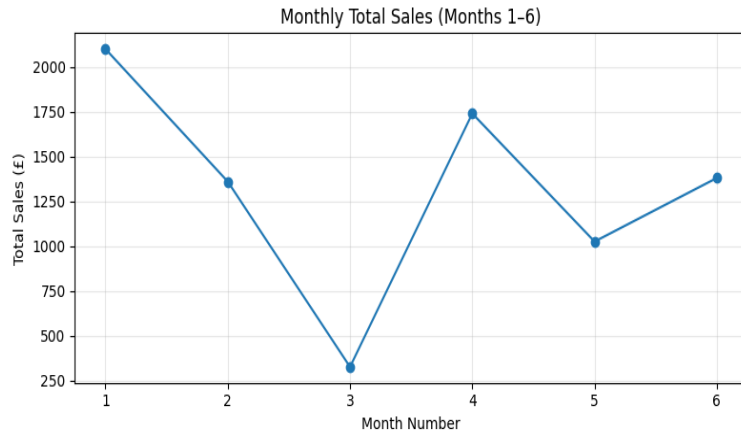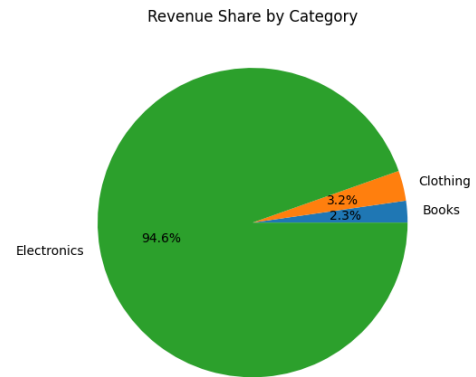
Figure 1

Figure 2

- **Product Popularity**:
    - ➢ Smartphones turned out to be the most sought after product (most purchases), with laptop coming in a close second.
    - ➢ Electronics seems to dominate the customer arena as they contribute to more than 90% of the revenue (Figure 2).
- **Payment Methods**: The histogram suggested that credit card and cash were the two popular mediums of payments.
- **Region**: From the heatmap (Figure 3) we can see that South has the most sales and west the least.
- **Summary Statistics of numerical columns**: The large standard deviations and wide Inter Quartile Ranges for both price and total amount underscore significant diversity in customer purchases—from low-cost, single-item buys to expensive, multi-unit electronics bundles. That mean, while most orders consist of one or two items, a minority of customers place significantly larger orders, which elevates the overall averages. Hence, they drive the revenue.
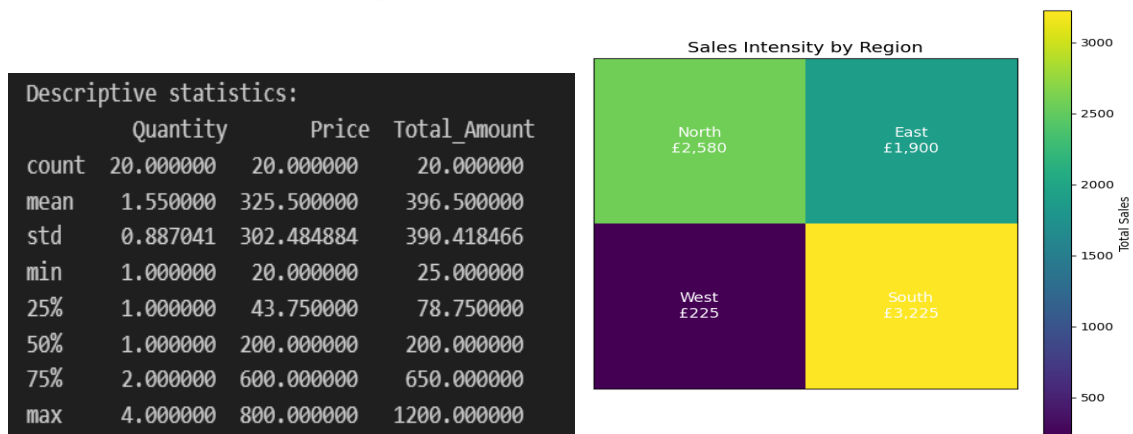
```
Descriptive statistics:
        Quantity      Price  Total_Amount
count  20.000000  20.000000    20.000000
mean    1.550000  325.500000   396.500000
std     0.887041  302.484884   390.418466
min     1.000000   20.000000    25.000000
25%     1.000000   43.750000    78.750000
50%     1.000000  200.000000   200.000000
75%     2.000000  600.000000   650.000000
max     4.000000  800.000000  1200.000000
```


Figure 3

- **Pearsons Correlation:**
  - Quantity vs Total Amount: The slight negative r(=-0.2) suggests that, in this dataset, orders with more units don't necessarily lead to higher total spend—and with p = 0.40 this relationship is not statistically significant.
  - Price vs Total Amount: A very strong, highly significant positive correlation. Higher-priced items almost always drive larger order values.

Summary – Exploratory Data Analysis

Our analysis of 6 months of sales data reveals clear seasonality and concentration in high-value segments: total monthly sales peaked in January (£2,100) and rebounded strongly in April (£1,750), but suffered a dip in March (£320). Electronics dominate revenue (more than 90%), led by Smartphones, Laptops, and Tablets, while Books and Clothing contribute 2.3% and 3.2 % respectively. Regionally, the South and North outperform the West and East, suggesting opportunities for targeted promotions. Finally, correlation analysis shows a very strong positive relationship between unit price and order value ($r = 0.89$, $p < 0.001$), whereas quantity per order has no significant effect ($r = -0.20$, $p = 0.40$), reflecting that high-ticket single-unit purchases drive revenue more than basket size.

3) Final Data Insights Report: January 2024–June 2024
KEY FINDINGS:-

1) **Seasonal Sales Patterns:**
   - The sales peaked in January(GBP 2100), most probably as a result of being new year and the took a hit and recovered in April(GBP 1750). This could be because of the start of the second quarter.
   - The numbers took a hit in March (GBP 320). This could be because of lack of promotional incentives or operational constraints, but we need more data points such as consumer reviews, import metrics, competitor sales, etc to reach a conclusion.
2) **Product Performance**:
   - The electronics dominate sales by taking up 94% of the total revenue earned. And out of that, smartphone, laptop and tablets are the top 3 products in demand.
   - Books and clothing contribute modestly but offer cross-sell potential.
3) **Regional Sales Patterns**:
   - Highest sales are recorded in the South region with a GBP total of 3225 and lowest in West with GBP 225.
   - This could be result of income difference between people of different regions, lack of awareness, reach of the company, past sales, presence of

competitors, etc. We do need more data points to reach a conclusion about this particular case.

4) **Correlations**: From analyzing correlations, it is clear that unit price strongly predicts order value (r = 0.89, p < 0.001)

5) Even though there are a few outliers here and there, the average purchase value is around GBP 400.

ACTIONABLE RECOMMENDATIONS:-

1) **Promotional Activities:**
   - Company could host an end-of-spring flash sale in July or August to recoup the loses of March.
   - In August, a back-to-school campaign could be hosted with games for children, to boost the sales of books and clothing.
   - Taking time to promote via social media (Instagram, TikTok, etc) could positively impact sales, by participating in trends, and giveaways and such.

2) **Inventory and Operations**: Pre-stock top electronics before January and April peaks; review mid-year supply chains to avoid July shortages.

3) **Targeted Marketing**:
   - To increase sales in the West, ShopEase could collaborate with an already existing retail company to spread awareness.
   - In the beginning, allocate 60 % of ad spend to smartphones and laptops, with region-specific creative, like more offers to least revenue generating regions as well as products.

4) **Loyalty Schemes**: Provide extended warranties or VIP support for premium purchasers to encourage loyalty

ETHICAL CONSIDERATIONS:-

1) **Data Privacy**: Anonymize all customer identifiers and comply with GDPR for storage and analysis of the data.

2) **Fair Targeting**: Ensure regional and demographic segmentation avoids any bias or exclusion.

3) **Transparency**: Communicate data usage and obtain consent from consumer wherever required to maintain trust.

4) Bonus Findings

1) **Low performing months:-** The lowest performing month is March (GBP 325), followed by May (GBP 1025). This is a stark difference. As suggested above, it is easily possible to boost sales.

2) **Strategies:-**
   - ➢ ShopEase could provide free shipping for orders above a certain threshold for those months, which motivates customers to make minimum purchases.
   - ➢ Give existing customers, a discount code to share, and reward both the referrer and referee with 10 % off—leverages word-of-mouth and expands reach.
   - ➢ A spring-flash-sale in march would also be a great idea, by selling not-so-sought-after products with a discount.

CONCLUSION

In summary, our analysis highlights strong seasonal peaks in January and April, a critical sales gap in March, and a heavy reliance on high-value electronics—especially smartphones and laptops—that drive the majority of revenue. Targeted off-peak promotions, bundled offers, and region-specific marketing can smooth seasonal changes and capitalize on under-served markets. By aligning inventory, campaigns, and incentives with these insights—and maintaining rigorous data quality checks—we can sustainably boost sales and customer engagement across the year.

REFERENCE:

1) Google
2) ChatGPT for grammatical errors.