Name	ASADUZZAMAN
Contact Number	07464104993
Project Title (Week2)	Data Cleaning, Analysis, and 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: TravelEase Case Study.)
   - **File Format:** Submit as PDF or Word file to contact@victoriasolutions.co.uk
   - **Page Limit:** 4–5 pages, including the title and references.

2. **Answer Requirements**
   - **Word Count:** Each answer should be 100–150 words; total 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 submit fail to submit the project will miss the "Certificate of Excellence"

7. **Additional Resources**
   - o Refer to lecture notes and recommended readings.
   - o Contact the instructor or peers for clarifications before the deadline.

START YOUR PROJECT FROM HERE:

Data Cleaning Results

1. Remove Duplicates

Remove Duplicates: No exact duplicates found in the dataset

The dataset had 8 rows, and no exact duplicates were found, so all rows were retained.

2. Fill Missing Values (IF/ISBLANK equivalent)

Fill Missing Values:

Email: 2 missing values filled with 'Unknown'

Phone: 2 missing values filled with 'Unknown'

Discount (%): 2 missing values filled with 0

3. Standardize Dates (TEXT formula equivalent)

Standardize Dates: All dates converted to YYYY-MM-DD format

12/31/2023 → 2023-12-31

01-05-2024 → 2024-01-05

2024/01/12 → 2024-01-12

02-15-2024 → 2024-02-15

FINAL Cleaned Dataset:

Order_ID	Customer_Name	Email	Phone	Product_Category	Order_Date	Revenue	Discount (%)
101	John Doe	john@email.com	9876543210	Electronics	2023-12-31	1200	10
102	Alice Smith	Unknown	9898989898	Clothing	2024-01-05	500	0
103	Bob Miller	bob@email.com	Unknown	Electronics	2024-01-12	3000	20
104	John Doe	john@email.com	9876543210	Electronics	2023-12-31	1200	10
105	David White	david@email.con	9123456789	Furniture	2024-02-15	2500	15
106	Emma Brown	emma@email.co	9234567890	Clothing	2024-03-08	700	5
107	Chris Green	Unknown	9345678901	Furniture	2024-04-10	1800	25
108	Alice Smith	alice@email.com	Unknown	Clothing	2024-03-08	500	0

Sales Trend Analysis Results
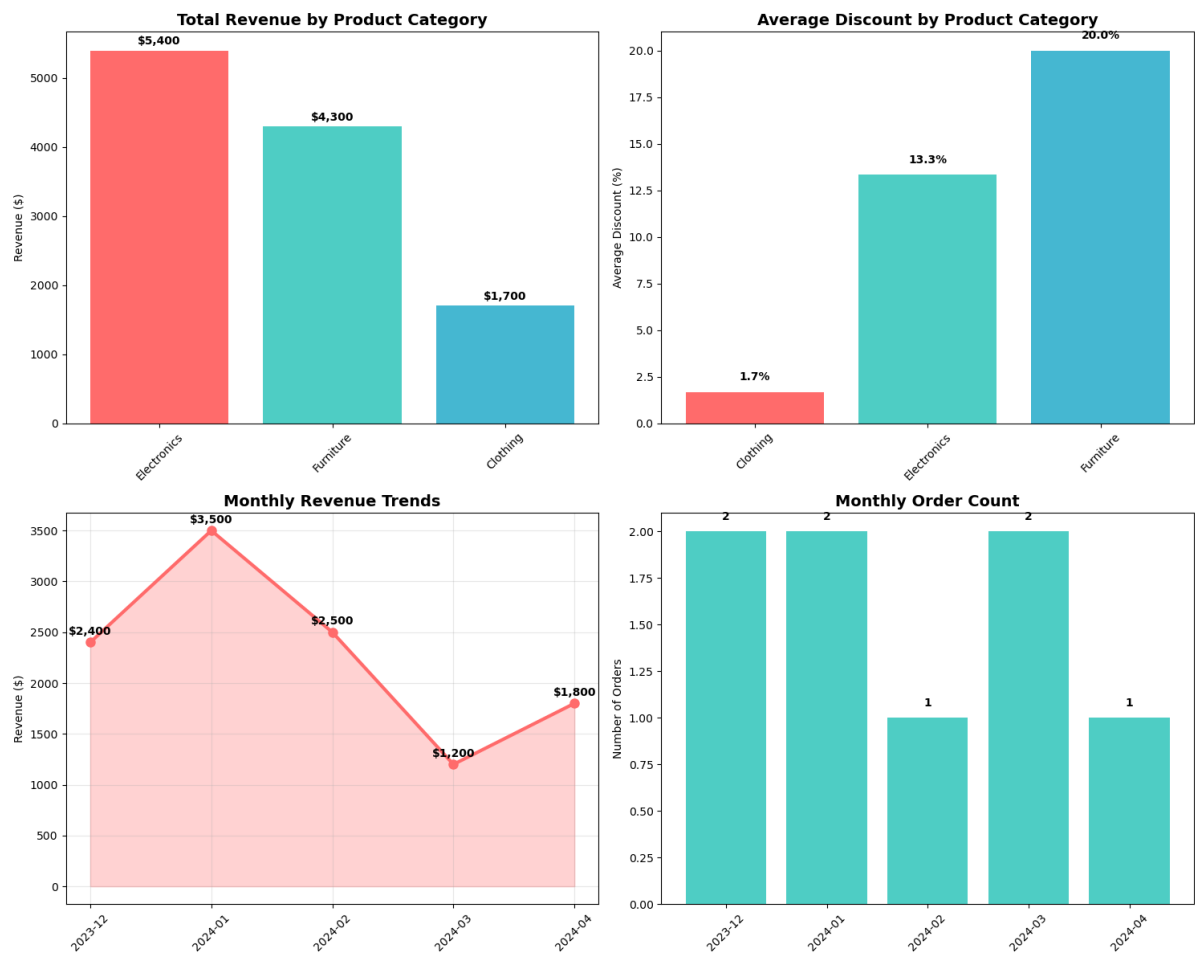
Sales Trend Analysis Results

Electronics dominates with 47.37% of total revenue ($5,400)

Furniture follows with 37.72% ($4,300) despite fewer orders

Clothing represents 14.91% ($1,700) with the lowest average order value

Discount Effectiveness Analysis

Furniture receives the highest average discounts (20%) but generates strong revenue

Electronics gets moderate discounts (13.33%) with excellent performance

Clothing has minimal discounts (1.67%) but also the lowest revenue per order
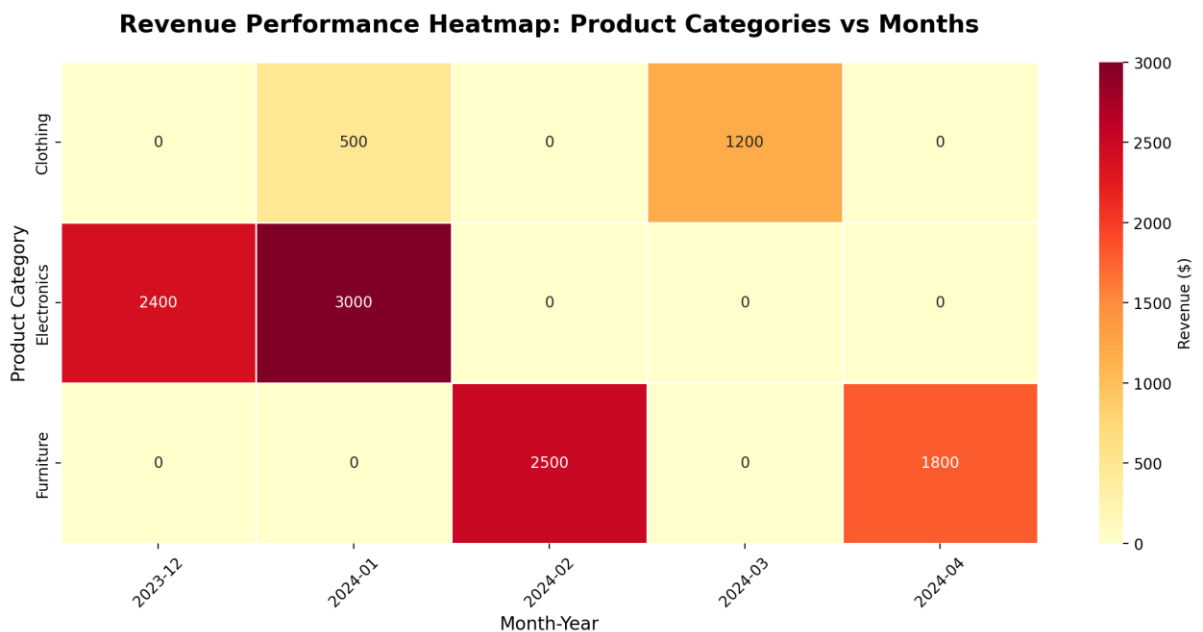
Overall average discount: 10.62 %

Monthly Sales Trends

Peak sales month: 2024-01 with $3500 revenue

Significant volatility with growth rates ranging from -52% to +50.83%

January 2024 showed strong 45.83% growth from December 2023

March 2024 experienced a notable 52% decline, followed by 50% recovery in April

**Revenue Performance Heatmap: Product Categories vs Months**

The heatmap shows revenue distribution across months and product categories. Electronics had strong performance in December 2023 and January 2024, while Furniture shows activity in February and April 2024, and Clothing appears in January and March 2024.

Relationship between Discount Percentage and Revenue
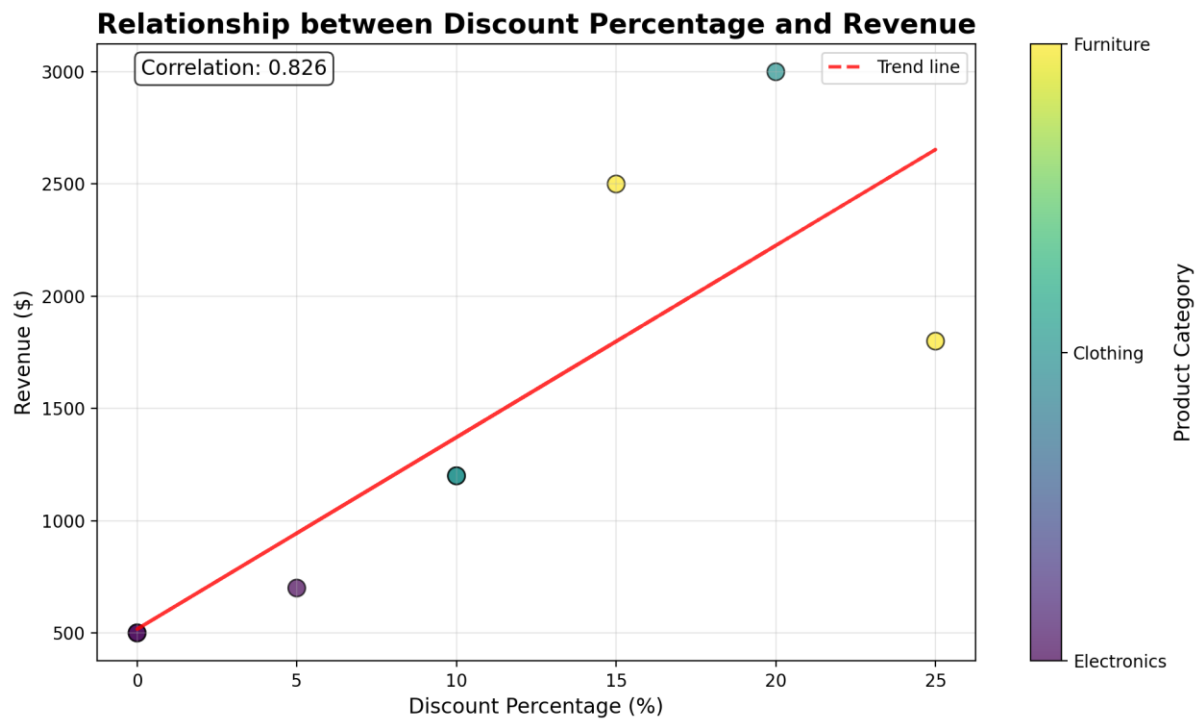
Correlation between discount and revenue: 0.826

The scatter plot reveals a strong positive correlation (0.826) between discount percentage and revenue, which is interesting - higher discounts are associated with higher revenue orders. This could indicate that discounts are being applied strategically to larger orders.
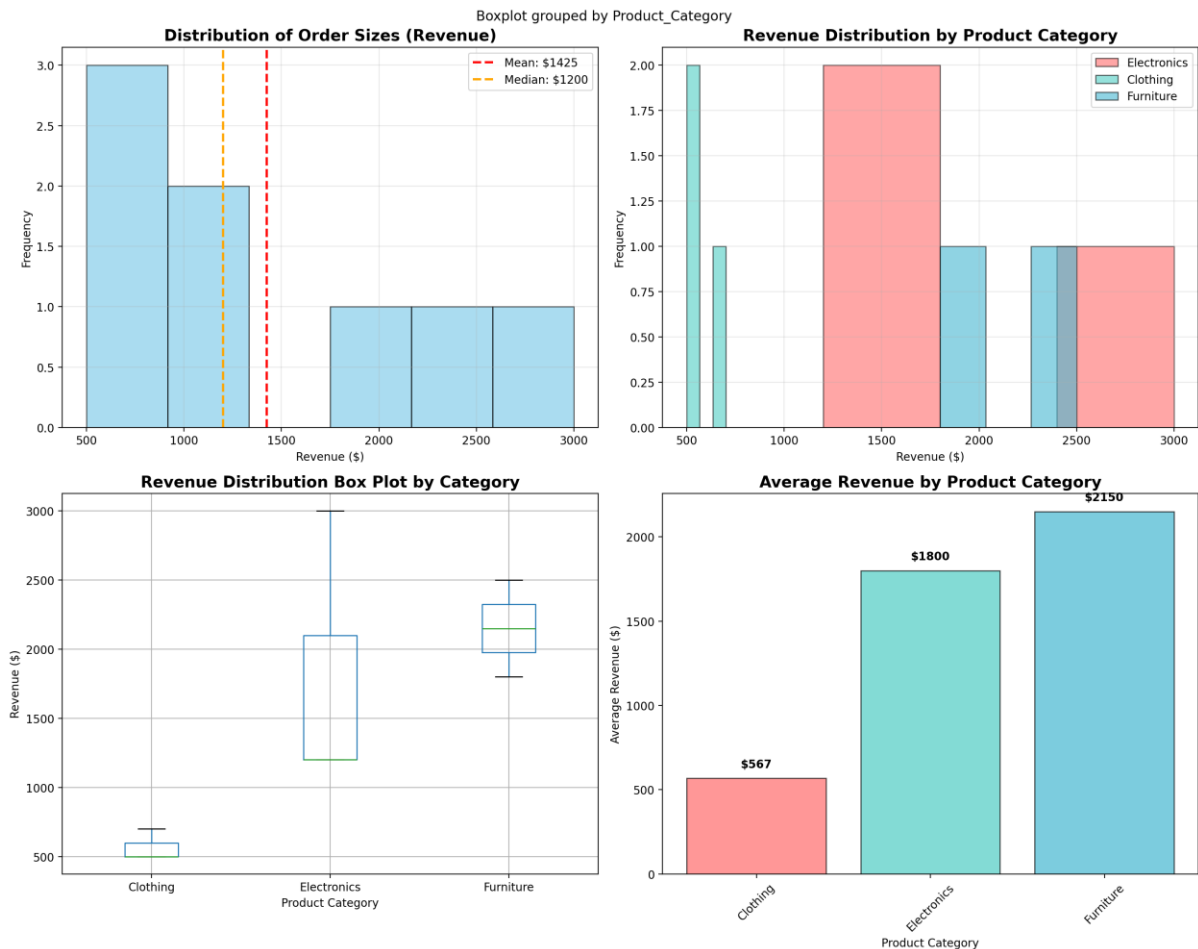
The comprehensive distribution analysis shows:

Mean: $1425.00

Median: $1200.00

Standard Deviation: $934.65

Range: $500 - $3000

Total Orders: 8

Total Revenue: $11400

Key Insights:

Electronics has the highest average revenue per order ($1,800) but also the highest variability

Furniture shows consistent high-value orders with less variation

Clothing has the lowest average revenue but represents steady business

The overall revenue distribution shows a wide range from $500 to $3,000, indicating diverse order sizes

The positive correlation between discounts and revenue suggests strategic discount application on larger orders

These visualizations provide a comprehensive view of your sales performance patterns, seasonal trends, and customer behavior across different product categories.