Name	Agata Gabara
Contact Number	+447904220617
Project Title (Example – Week1, Week2, Week3)	Week 2

VICTORIA SOLUTIONS
CONSULTING GROUP

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

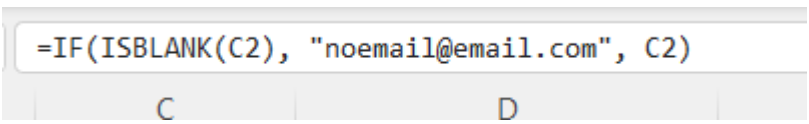
START YOUR PROJECT FROM HERE:

1.Cleaned dataset in Excel.

The dataset has been cleaned.

- I have removed duplicates; John Doe appears twice (Order_ID 101 and 104). One line has been removed.
- 6 values have been missing: 2 fields in Emil, 2 fields in Phone and 2 fields in Discount. The data has been filled by the usage appropriate function.
- Data format has been standardised.
- Data have been stored as floats and there have been changed for string.
- The empty fields in discount values column have value 0.

Email	Email (check if is blank)	F
john@email.com	john@email.com	
	noemail@email.com	
bob@email.com	bob@email.com	
john@email.com	john@email.com	
david@email.com	david@email.com	
emma@email.com	emma@email.com	
	noemail@email.com	
alice@email.com	alice@email.com	



E	G
Phone	**Phone (IF IS BLANK -> not available)**
9876543210	9876543210
9898989898	9898989898
	Not Available
9876543210	9876543210
9123456789	9123456789
9234567890	9234567890
9345678901	9345678901
	Not Available

```
=IF(ISBLANK(E2), "Not Available", E2)
```

I	J
Order_Date	**Order date**
12/31/2023	12/31/2023
01/05/2024	2024-05-01
12/01/2024	2024-01-12
12/31/2023	12/31/2023
02-15-2024	02-15-2024
08/03/2024	2024-03-08
04/10/2024	2024-10-04
03/08/2024	2024-08-03

```
=TEXT(I2,"yyyy-mm-dd")
```

L	M
Discount (%)	Discount
10	10
	0
20	20
10	10
15	15
5	5
25	25
	0

```
=IF(ISBLANK(L2),0,L2)
```

Cleaned dataset has 7 rows and 8 columns. The dataset contains numeric and non-numeric values.

Order_ID	Customer_Name	Email	Phone	Product_Category	Order date	Revenue	Discount
101	John Doe	john@email.com	9876543210	Electronics	12/31/2023	1200	10
102	Alice Smith	noemail@email.com	9898989898	Clothing	2024-05-01	500	0
103	Bob Miller	bob@email.com	Not Available	Electronics	2024-01-12	3000	20
104	John Doe	john@email.com	9876543210	Electronics	12/31/2023	1200	10
105	David White	david@email.com	9123456789	Furniture	02-15-2024	2500	15
106	Emma Brown	emma@email.com	9234567890	Clothing	2024-03-08	700	5
107	Chris Green	noemail@email.com	9345678901	Furniture	2024-10-04	1800	25
108	Alice Smith	alice@email.com	Not Available	Clothing	2024-08-03	500	0

2. SQL queries used for data cleaning and aggregation.

I have used SQL queries as below.

The code comes from MY SQL.

```
-- CREATE DATABASE ecommerce_data;

USE ecommerce_data;

drop table Sales_Data;

CREATE TABLE IF NOT EXISTS Sales_Data (
```

```
    Order_ID INT PRIMARY KEY,
    Customer_Name VARCHAR(100),
    Email VARCHAR(100),
    Phone VARCHAR(20),
    Product_Category VARCHAR(50),
    Order_Date DATE,
    Revenue DECIMAL(10,2),
    Discount_Percent DECIMAL(5,2)
);

INSERT INTO  `Sales_Data` (Order_ID, Customer_Name, Email, Phone, Product_Category,
Order_Date, Revenue, Discount_Percent)
VALUES (101, 'John Doe', 'john@email.com', '9876543210', 'Electronics', '2023-12-31', 1200.00,
10.00);

INSERT INTO  `Sales_Data` (Order_ID, Customer_Name, Email, Phone, Product_Category,
Order_Date, Revenue, Discount_Percent)
VALUES (102, 'Alice Smith', 'not_provided@email.com', '9898989898', 'Clothing', '2024-01-05',
500.00, 0.00);

INSERT INTO  `Sales_Data` (Order_ID, Customer_Name, Email, Phone, Product_Category,
Order_Date, Revenue, Discount_Percent)
VALUES (103, 'Bob Miller', 'bob@email.com', 'Not Available', 'Electronics', '2024-12-01', 3000.00,
20.00);

INSERT INTO  `Sales_Data` (Order_ID, Customer_Name, Email, Phone, Product_Category,
Order_Date, Revenue, Discount_Percent)
VALUES (105, 'David White', 'david@email.com', '9123456789', 'Furniture', '2024-02-15', 2500.00,
15.00);

INSERT INTO  `Sales_Data` (Order_ID, Customer_Name, Email, Phone, Product_Category,
Order_Date, Revenue, Discount_Percent)
VALUES (106, 'Emma Brown', 'emma@email.com', '9234567890', 'Clothing', '2024-08-03', 700.00,
5.00);

INSERT INTO  `Sales_Data` (Order_ID, Customer_Name, Email, Phone, Product_Category,
Order_Date, Revenue, Discount_Percent)
```

```
VALUES (107, 'Chris Green', 'not_provided@email.com', '9345678901', 'Furniture', '2024-04-10',
1800.00, 25.00);

INSERT INTO  `Sales_Data` (Order_ID, Customer_Name, Email, Phone, Product_Category,
Order_Date, Revenue, Discount_Percent)

VALUES (108, 'Alice Smith', 'alice@email.com', 'Not Available', 'Clothing', '2024-03-08', 500.00, 0.00);


ALTER TABLE Sales_Data ADD Discount DECIMAL(5,2);

select * from Sales_Data;

-- SET SQL_SAFE_UPDATES = 0;

UPDATE Sales_Data

SET Discount = Discount_Percent / 100.0

WHERE Order_ID IS NOT NULL;


ALTER TABLE Sales_Data DROP Discount_Percent


SELECT Product_Category, SUM(Revenue) AS Total_Revenue

FROM sales_data

GROUP BY Product_Category;


SELECT Product_Category, Discount_Percent
```

Screenshot from MySQL command line.

```
MySQL 8.0 Command Lin

owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show databases
    -> ;
+--------------------+
| Database           |
+--------------------+
| ecommerce_data     |
| information_schema |
| mysql              |
| performance_schema |
| retail_sales       |
| retaillab          |
| sys                |
+--------------------+
7 rows in set (0.01 sec)

mysql> use ecomerce_data
ERROR 1049 (42000): Unknown database 'ecomerce_data'
mysql> use ecommerce_data
Database changed
mysql> show tables;
Empty set (0.01 sec)

mysql> show tables;
+--------------------------+
| Tables_in_ecommerce_data |
+--------------------------+
| cleaned_sales_data       |
+--------------------------+
1 row in set (0.00 sec)

mysql> show tables;
+--------------------------+
| Tables_in_ecommerce_data |
+--------------------------+
| sales_data               |
+--------------------------+
1 row in set (0.00 sec)

mysql>
```

I received the result as below.

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Product_Category	Total_Revenue
Electronics	4200.00
Clothing	1700.00
Furniture	4300.00

Aggregated Query Result (Screenshot)

- Summed up revenue by Product_Category:
    - Electronics → 4200.00
    - Clothing → 1700.00
    - Furniture → 4300.00

Summary - relation to SQL queries.

Database Creation

- Created a new database ecommerce_data.
- Switched to using this database.

Table Setup

- Dropped Sales_Data table if it existed.
- Created a new Sales_Data table with fields for order details, including Revenue and Discount_Percent.

Data Insertion

- Inserted 7 records with sample orders across three categories:
    - Electronics
    - Clothing
    - Furniture

Table Modification

- Added a new column Discount.
- Populated Discount as a decimal fraction of Discount_Percent (Discount_Percent / 100).
- Dropped the original Discount_Percent column.

3.Three visualisations.

1. Heatmap

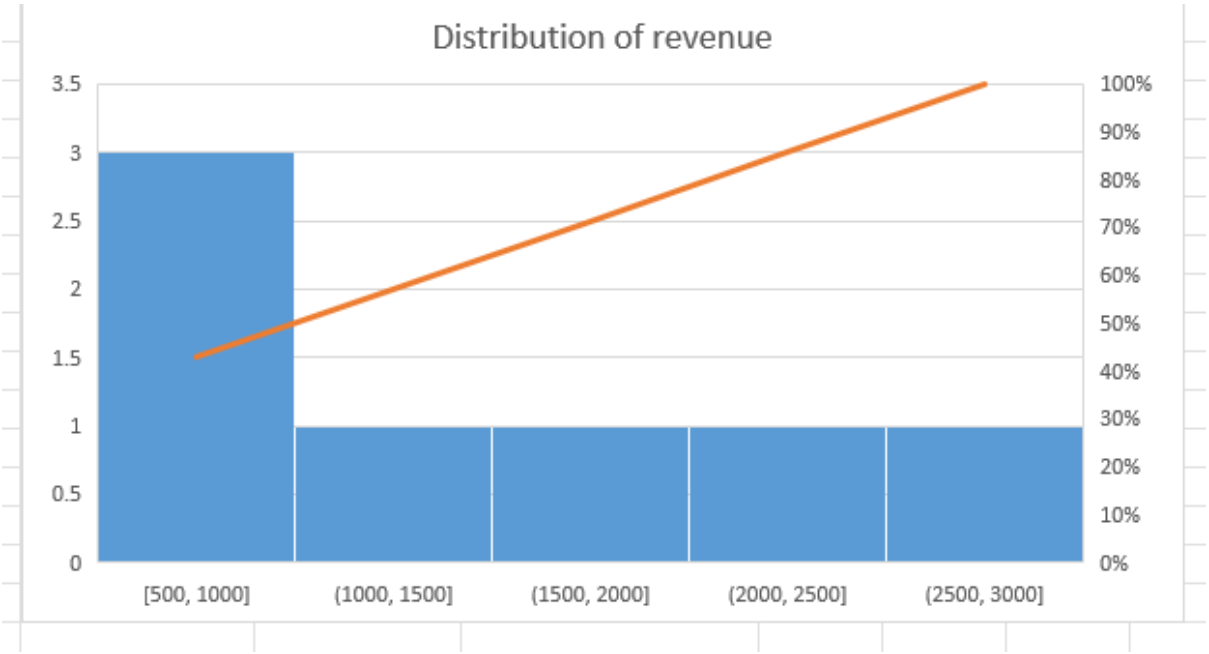
Sum of Revenue	Column Labels			
Row Labels	Clothing	Electronics	Furniture	Grand Total
Jan	500	0	0	500
Feb	0	0	2500	2500
Mar	500	0	0	500
Apr	0	0	1800	1800
Aug	700	0	0	700
Dec	0	4200	0	4200
Grand Total	**1700**	**4200**	**4300**	**10200**

Own collaboration based on MS Excel.

Summary.

- Furniture leads overall (4300), followed closely by Electronics (4200).
- Clothing lags (1700).
- Dec (Electronics) and Feb (Furniture) drive the largest monthly revenues.

2.Histogram

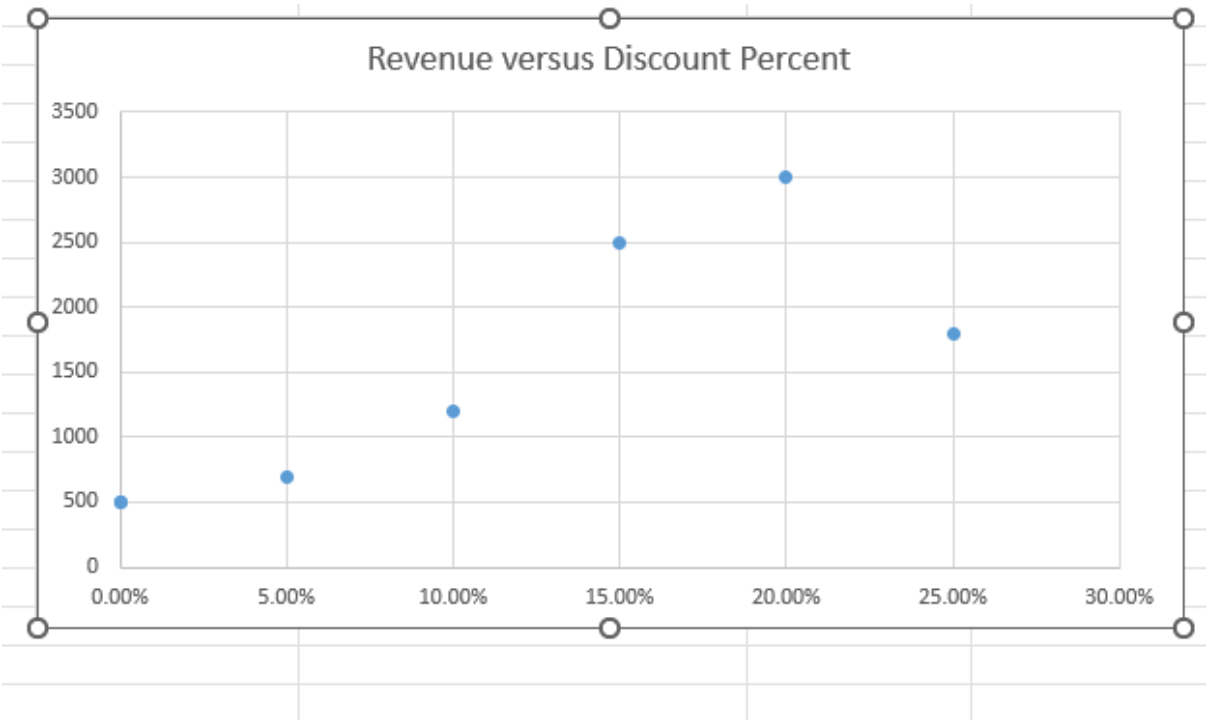


Own collaboration based on MS Excel.

Summary.

- Most common order size: Revenues fall in the [500–1000] range (3 orders).
- Larger orders are less frequent: Only 1 order each in the higher bins.
- Skewed distribution: Revenue is heavily concentrated in smaller orders (low values dominate frequency).
- Pareto implication: A small number of higher-value orders contribute significantly to overall revenue, even though they are rare.

3.Scatter.



Own collaboration based on MS Excel.

Summary.

- There seems to be a positive relationship between discount and revenue up to ~20%, suggesting moderate discounts encourage higher spending.
- Beyond 20%, revenue starts falling again, which might mean heavy discounts don't always guarantee higher sales value.
- The sweet spot for discounts in this dataset appears to be 15–20%, where revenue was highest.

4.Summary report detailing key findings and business recommendations.

Revenue Distribution by Category

Analysis revealed that Furniture (4300) generated the highest revenue, closely followed by Electronics (4200), while Clothing (1700) lagged behind. Seasonal trends showed that December (Electronics) and February (Furniture) were peak months, suggesting strong seasonal buying patterns.

Revenue Frequency

The histogram indicated that most transactions fall within the 500–1000 revenue range, highlighting a dominance of low-value purchases. High-value transactions, though rare, contributed disproportionately to overall revenue, reflecting a Pareto effect (80/20 rule).

Impact of Discounts

The scatter plot demonstrated a positive correlation between discounts and revenue up to 20%, with revenue peaking at this range. However, excessive discounts beyond 20% resulted in declining revenue, implying that overly aggressive discounting reduces profitability without boosting sales.

Business recommendations.

- Focus on profitable categories and seasons: Prioritize growth in Furniture and Electronics, while boosting Clothing sales through targeted campaigns; align promotions with peak months (December, February).
- Optimize discount strategy: Keep discounts in the 15–20% range to maximize revenue, avoid over-discounting, and use A/B testing to refine approaches by category.
- Enhance customer engagement: Segment customers to identify high-value buyers and introduce tiered loyalty programs to encourage both frequent small purchases and larger transactions.