Name	Mohamed abdelnasser hassan
Contact Number	+20 1060553978
Project Title (Example – Week1, Week2, Week3)	Week2: analysing sales data

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:

Part1: cleaning the data using excel:

Step1: removing the duplicates

Duplicate records were identified in the dataset, such as the repeated entry of "John Doe." The Remove Duplicates option from the Data tab was applied to eliminate these redundant rows. As a result, only unique customer records were retained.

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		9898989898	Clothing	1/5/2024	500	
103	Bob Miller	bob@email.com		Electronics	1/12/2024	3000	20
105	David White	david@email.com	9123456789	Furniture	2/15/2024	2500	15
106	Emma Brown	emma@email.com	9234567890	Clothing	3/8/2024	700	5
107	Chris Green		9345678901	Furniture	4/10/2024	1800	25
108	Alice Smith	alice@email.com		Clothing	3/8/2024	500	

Step 2: Handling Missing Values

Blank fields were observed in the *Email* and *Discount (%)* columns and also phone coulmn. To address this, the formula =IF(ISBLANK(cell), "Unknown", cell) was applied. In the case of discounts, missing values were replaced with **0** to avoid calculation errors. This process ensured that no empty cells remained and that the dataset maintained consistency.

| Order_ID	Customer_Name	Email	Email(resolved)	Phone	Phone(resolved)	Product_Category	Order_Date	Revenue	Discount (%)	Discount (%)(resolved)
101	John Doe	john@email.com	john@email.com	9876543210	9876543210	Electronics	12/31/2023	1200	10	10
102	Alice Smith		Unknown	9898989898	9898989898	Clothing	1/5/2024	500		0
103	Bob Miller	bob@email.com	bob@email.com		Unknown	Electronics	1/12/2024	3000	20	20
105	David White	david@email.com	david@email.com	9123456789	9123456789	Furniture	2/15/2024	2500	15	15
106	Emma Brown	emma@email.com	emma@email.com	9234567890	9234567890	Clothing	3/8/2024	700	5	5
107	Chris Green		Unknown	9345678901	9345678901	Furniture	4/10/2024	1800	25	25
108	Alice Smith	alice@email.com	alice@email.com		Unknown	Clothing	3/8/2024	500		0

Step 3: Standardizing Date Formats

Inconsistent date formats were identified, including variations such as *MM/DD/YYYY*, *DD-MM-YYYY*, and *YYYY/MM/DD*. The function =TEXT(cell, "yyyy-mm-dd") was applied to transform all dates into a uniform format. This standardization ensured that time-based analysis could be conducted accurately without format conflicts.

Order_ID	Customer_Name	Email(resolved)	Phone(resolved)	Product_Category	Order_Date	Revenue	Discount (%)(resolved)
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
105	David White	david@email.com	9123456789	Furniture	2024-02-15	2500	15
106	Emma Brown	emma@email.com	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

Part2: KPI Analysis with SQL(using Microsoft SQL Server)

A database named VictoriaProject was created, and the cleaned Excel file was imported using the Import Flat File option. This prepared the data for structured analysis. SQL queries were then applied to generate Key Performance Indicators (KPIs) on revenue, discounts, customer behavior, and sales trends.
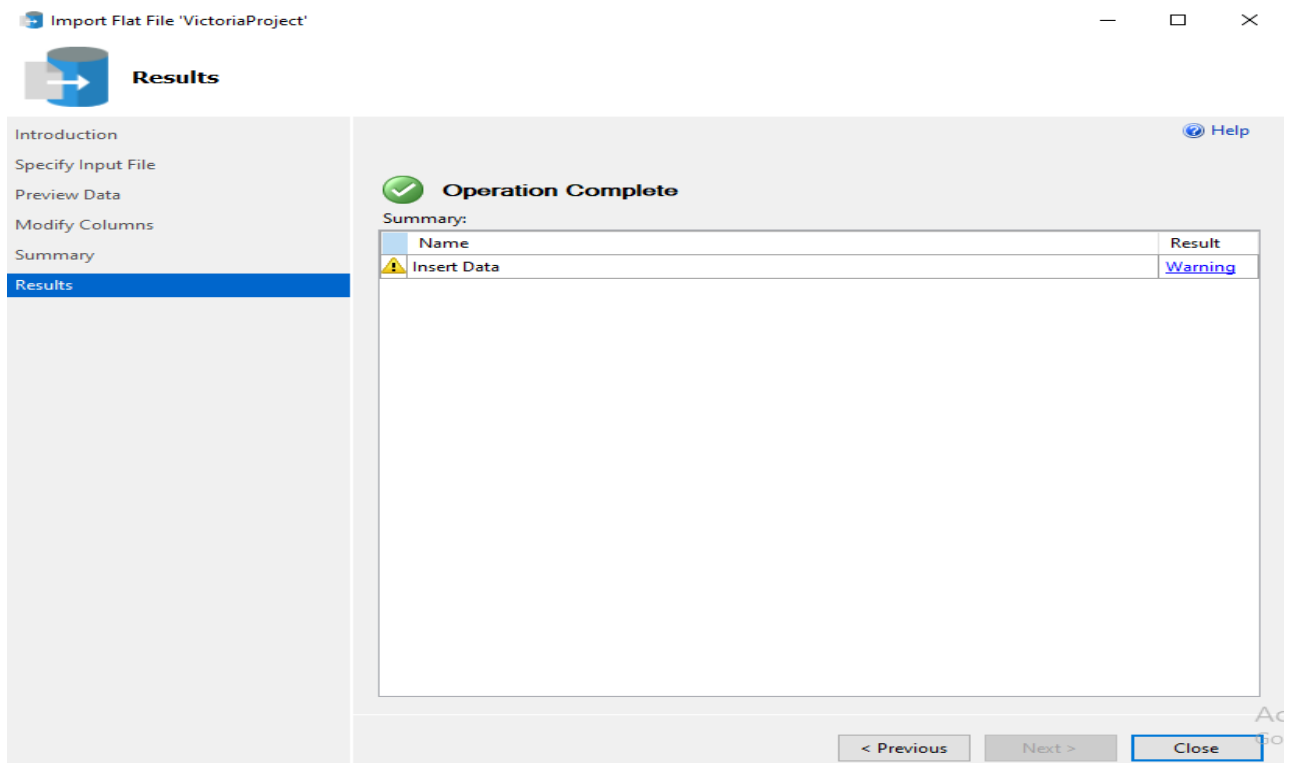
1-Total Revenue per Product Category:

Shows which product categories generate the most money as it Helps identify whether Electronics, Clothing, or Furniture drives the most revenue.


```
SELECT
    Product_Category,
    SUM(Revenue) AS Total_Revenue
FROM raw_sales_data
GROUP BY Product_Category
ORDER BY Total_Revenue DESC;
```

	Product_Category	Total_Revenue
1	Furniture	4300
2	Electronics	4200
3	Clothing	1700

2. Average Discount per Product Category:

Shows if certain categories rely more on discounts its Useful for evaluating discount effectiveness by category.


```
SELECT
    Product_Category,
    AVG([Discount]) AS Avg_Discount
FROM raw_sales_data
GROUP BY Product_Category;
```

	Product_Category	Avg_Discount
1	Clothing	1
2	Electronics	15
3	Furniture	20

3. Monthly Revenue Trend

Tracks revenue growth and seasonal sales it helps in Identifies peak months for sales (e.g., holidays, seasonal patterns).

```sql
SELECT
    FORMAT(Order_Date, 'yyyy-MM') AS Month,
    SUM(Revenue) AS Monthly_Revenue
FROM raw_sales_data
GROUP BY FORMAT(Order_Date, 'yyyy-MM')
ORDER BY Month;
```

	Month	Monthly_Revenue
1	2023-12	1200
2	2024-01	3500
3	2024-02	2500
4	2024-03	1200
5	2024-04	1800

4. High-Discount Customers

Identify customers who received above-average discounts. is important because it highlights segments where profit margins may be reduced, helping the business evaluate whether these discounts are driving loyalty or unnecessarily eroding revenue

```sql
SELECT
    Customer_Name,
    AVG([Discount]) AS Avg_Discount
FROM raw_sales_data
GROUP BY Customer_Name
HAVING AVG([Discount]) > (
    SELECT AVG([Discount]) FROM raw_sales_data
);
```

	Customer_Name	Avg_Discount
1	Bob Miller	20
2	Chris Green	25
3	David White	15

5- Orders Without Contact Information

Checks data quality by counting missing emails or phones this is important because Poor data quality can block marketing & customer follow-up.

```sql
SELECT
    SUM(CASE
            WHEN Email IS NULL OR LTRIM(RTRIM(Email)) IN ('', 'Unknown', 'unknown')
            THEN 1 ELSE 0
        END) AS Missing_Emails,
    SUM(CASE
            WHEN Phone IS NULL OR LTRIM(RTRIM(Phone)) IN ('', 'Unknown', 'unknown')
            THEN 1 ELSE 0
        END) AS Missing_Phones
FROM raw_sales_data;
```

Results | **Messages**

	Missing_Emails	Missing_Phones
1	2	2

Part3: Visualizations (using Power BI)

1- Heatmap (Seasonal Revenue Patterns)

Analyzing seasonal revenue patterns across categories helps the business forecast demand, optimize inventory, and align promotions with peak sales periods

Product_Category	January	February	March	April	December
Clothing	500		1200		
Electronics	3000				1200
Furniture		2500		1800	

2- Scatter Plot (Discount vs Revenue)

Examining the relationship between discounts and revenue helps determine whether higher discounts actually drive sales or unnecessarily reduce profitability

Impact of Discounts on Revenue

Customer_Name	Chris Green
Sum of Discount (%)	25
Sum of Revenue	1800

3- Histogram (Discount Distribution)

Understanding the distribution of discounts across orders highlights pricing strategies and reveals whether discounts are applied fairly or concentrated in specific ranges

Distribution of Discounts Across Orders

Part4: Analysis and Recommendations

this analysis provides a comprehensive overview of the sales data, integrating findings from both SQL queries and Power BI visualizations.

Key Insights

- **Discounting and Revenue:** The relationship between discounts and revenue is not linear. While some customers, like Bob Miller, generated a high revenue of $3,000 with a 20% discount, the customer with the highest discount, Chris Green (25%), generated a lower revenue of $1,800. The data also indicates that discounts are applied infrequently, with most orders receiving no discount.

- **Product Performance and Seasonal Trends:** Electronics and Furniture were identified as the top-performing categories. A clear seasonal trend was observed, with Electronics sales peaking in January, while Furniture sales were strongest in February and March. This pattern suggests that revenue can fluctuate significantly throughout the year based on product category performance.

- **Data Quality:** Gaps were identified in the data, specifically regarding customer contact information, with several missing email addresses and phone numbers.

Actionable Recommendations

Based on the insights gathered, the following recommendations are presented to improve sales strategies and data management.

- **A more strategic approach to discounting should be adopted.** It is recommended that the effectiveness of high-value discounts be re-evaluated. Focusing on customer segments that have demonstrated a higher return, such as those with behavior similar to Bob Miller, could lead to more profitable outcomes. Additionally, a test could be conducted to see if a higher discount for the **Clothing** category improves its low revenue performance.

- **Sales and marketing efforts should be aligned with seasonal trends.** By focusing promotional campaigns on specific categories during their peak months, such as **Electronics in January** and **Furniture in February and March**, revenue can be optimized.

- **A process to address data quality issues should be implemented.** The missing customer contact information represents a significant gap that could hinder targeted marketing and customer communication. A system for collecting and validating complete customer data is recommended to support future strategic initiatives.

LINK OF THE CODE ON GITHUB

References:

Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.

Tanimura, C. (2021). *SQL for data analysis*. O'Reilly Media.