King Abdulaziz University
The Applied College

Computer and Information Technology Department **Final**

Data Analysis Report

Developed py : Joury(Dif) AI-Mutairi

For professional inquiries, visit my LinkedIn profile:

linkedin.com/in/joury-al-mutairi-684573357

1. Introduction

This dataset contains sales transactions for various product categories across different regions.
It includes **279 records** and **11 variables**, namely:

- **Order ID**
- **Order Date**
- **Item Type**
- **Region**
- **Units Sold**

- **Unit Price**
- **Unit Cost**
- **Total Revenue**
- **Total Profit**
- **Customer Age**

The purpose of this project is to **clean the dataset**, conduct **descriptive statistical analysis**, assess **correlations and patterns**, and create **pivot table–based insights** to gain a comprehensive understanding of overall sales performance.

2. Data Cleaning

Several issues were identified in the raw dataset, including missing values, inconsistent formats, and inaccurate numerical entries. The following data-cleaning steps were performed:

• Handling Missing Values

A number of rows contained blank values in fields such as **Unit Price**, **Unit Cost**, **Total Revenue**, **Total Profit**, and **Customer Age**.

To maintain data integrity, rows with missing essential numerical values were removed.

• Correcting Incorrect Numeric Values

Some numeric fields used commas instead of decimal points (e.g., *99,99*).

These were corrected by replacing commas with periods and converting the values to proper numeric format.

A typographical error in **Total Revenue** (written as *1O00* using the letter "O") was also corrected.

- **Recalculating Financial Fields**

To ensure accuracy across financial metrics, key fields were recalculated:

- **Total Revenue** = Units Sold × Unit Price
- **Total Profit** = Total Revenue − (Units Sold × Unit Cost)
- **Ensuring Consistency**

All numeric columns were validated to ensure they contained only valid numerical entries. Text-based fields such as **Item Type** and **Region** were reviewed and standardized for formatting consistency.

3. Descriptive Statistics

A. Units Sold

Unit Sold	
Mean	258.4100719
Median	255.5
Mode	108
Range	488
Standard Deviation	142.2681311

Interpretation:

Units Sold varies widely (large range and SD), indicating high variability in demand across products.

B. Customer Age

Customer Age	
Mean	43.5
Median	43
Mode	64
Range	51
Standard Deviation	15.38688785

Interpretation:
The average customer age is around 43, with a moderate spread, suggesting a diverse customer base across age groups.

4. Correlation Analysis

1. Units Sold vs Total Revenue

Correlation Between Unit Sold & Total Revenue
0.615600714

Interpretation: A strong positive correlation; as Units Sold increases, Total Revenue increases significantly.

2. Unit Price vs Total Profit

Correlation Between Unit Price & Total Profit
0.536680573

Interpretation:
A moderate positive correlation; higher-priced items tend to generate more profit, but other factors (such as cost) also influence profitability.

3. Customer Age vs Units Sold

Correlation Between Customer Age & Units Sold
0.026529977

Interpretation:
No meaningful relationship; customer age does *not* significantly influence the number of units purchased.

5. Pivot Table Insights

1. Total Revenue by Item Type

Row Labels	Sum of Total Revenue
Beverages	2705392.09
Clothes	2024043.65
Cosmetics	2064812.64
Electronics	1892593.08
Snacks	1816135.53
Grand Total	**10502976.99**

Insight:
Beverages generate the highest revenue, while **Snacks** generate the lowest among the five categories.

2. Total Profit by Region

Row Labels	Sum of Total Profit
East	1004141.12
North	984926.94
South	814049.95
West	1219681.89
Grand Total	**4022799.9**

Insight:
The **West region** is the most profitable, while the **South region** has the lowest profit contribution.

3. Average Customer Age by Item Type

Row Labels	Average of Customer Age
Beverages	44.81428571
Clothes	41.43396226
Cosmetics	44.78723404
Electronics	44.92156863
Snacks	41.47368421
Grand Total	**43.5**

Insight:
Cosmetics, Electronics, and Beverages attract slightly older customers, while Clothes and Snacks tend to attract younger customers.

6. Conclusion

The dataset showcases diverse sales activity across various product types and geographic regions. Key insights include:

- **Beverages emerge as the most profitable product category.**

- **The West region records the highest overall profitability.**
- **Customer age shows no meaningful influence on purchase quantity**, as indicated by the near-zero correlation.
- **Recalculated financial fields** ensured accurate and reliable measurements of revenue and profit.
- **Descriptive statistics highlight substantial variability in units sold** and a **moderate range in customer ages**.

Overall, the cleaned and analyzed dataset offers a clearer understanding of sales performance and customer behavior, enabling more effective data-driven decision-making.