



## **Certified Data Analysts**

### **Assignment 1**

**Name: Nur Atiqah Binti Idris**

**Submission Date: 8<sup>th</sup> February 2025**

---

#### **Section 1: Data Cleaning and Preparation**

##### **Step 1: Import the Dataset**

- Open **Excel** and go to **Data** → **Get & Transform** → **Get Data** → **From File** → **From Workbook**.
- Select the dataset and load it into **Power Query**.

##### **Step 2: Identify Missing & Invalid Values**

- In Power Query, filter columns to find missing or invalid entries.
- Identify **null values** in critical fields like **Region, Sales, Profit, Quantity, and Discount**.

##### **Step 3: Data Transformation**

- **Replace missing values:**
  - For **Region** → Use "Unknown".
  - For **Sales & Profit** → Use the **mean or median** of the respective column.
  - For **Quantity** → Use the **mean or median** of the respective column.
  - For **Discount** → Use the **mean or median** of the respective column.
- **Remove invalid rows** that cannot be reasonably corrected.

##### **Step 4: Add a New Column for Profit Margin**

- Click **"Add Column"** → **"Custom Column"**.

- Use the formula:

$$[\text{Profit}] / [\text{Sales}] * 100$$

- If you get null values, check for missing or zero values in the **Sales** column (to avoid division by zero).

Order ID	Order Date	Ship Date	Region	Product Category	Product Name	Sales	Profit	Quantity	Customer Segment	Manager	Discount	Profit Margin (%)
1001	15/9/2024 0:00	28/12/2024 0:00	West	Technology	Control Smile	546.48	147.36	14	Small Business	Michael Brown	0.06	26.9653052
1002	13/8/2024 0:00	15/10/2024 0:00	East	Office Supplies	Wonder Figure	1002.27	380.1	4	Corporate	John Smith	0.05	37.9239127
1003	11/1/2025 0:00	29/1/2025 0:00	North	Technology	Recently Father	362.58	421.91	5	Home Office	Michael Brown	0.11	116.363285
1004	16/9/2024 0:00	20/12/2024 0:00	West	Technology	Check Charge	1426.55	72.51	5	Small Business	John Smith	0.05	5.08289229
1005	13/4/2024 0:00	10/1/2025 0:00	North	Technology	Company Now	234.62	358.74	12	Corporate	Sarah Johnson	0.08	152.902565
1006	30/6/2024 0:00	25/12/2024 0:00	South	Furniture	Ten Wind	510.05	313.58	9	Small Business	Sarah Johnson	0.06	61.4802470
1007	13/8/2024 0:00	7/11/2024 0:00	West	Technology	Bar Court	61.26	83.03	13	Small Business	John Smith	0.11	135.537055
1008	9/11/2024 0:00	11/11/2024 0:00	North	Office Supplies	Black Someone	305.04	429.5	3	Corporate	Emily Davis	0.18	140.801206
1009	1/5/2024 0:00	31/8/2024 0:00	East	Technology	Put However	1540.04	456.69	18	Consumer	John Smith	0.03	29.6544245
1010	9/7/2024 0:00	12/7/2024 0:00	North	Office Supplies	Southern Role	1112.7	108.4	5	Home Office	Michael Brown	0.18	9.74206884
1012	21/6/2024 0:00	5/10/2024 0:00	West	Office Supplies	Firm Represent	972.365	113.48	6	Consumer	Michael Brown	0.13	11.6705146
1013	22/7/2024 0:00	6/11/2024 0:00	East	Office Supplies	Drive Radio	1342.38	46.57	2	Consumer	Sarah Johnson	0.16	3.46921140
1014	5/2/2024 0:00	6/6/2024 0:00	West	Technology	If Cold	585.37	175.95	10	Corporate	John Smith	0.18	30.0579120
1016	24/9/2024 0:00	9/12/2024 0:00	East	Furniture	Recognize Keep	1446.74	408.83	20	Corporate	Sarah Johnson	0.05	28.2587057
1017	8/3/2024 0:00	4/5/2024 0:00	South	Technology	Key Lead	370.93	168.1	12	Home Office	John Smith	0.08	45.3185237
1018	2/1/2025 0:00	3/2/2025 0:00	East	Technology	Fall Yard	972.365	261.66	18	Home Office	Emily Davis	0.15	26.9096481
1019	25/5/2024 0:00	14/12/2024 0:00	East	Office Supplies	Stock Imagine	398.46	89.99	20	Consumer	Emily Davis	0.13	22.5844501
1020	17/11/2024 0:00	20/1/2025 0:00	Unknown	Technology	Trouble Media	1431.96	278.69	8	Corporate	Sarah Johnson	0.1	19.4621358
1021	3/1/2025 0:00	25/1/2025 0:00	North	Technology	Through Available	1954.15	48.62	6	Small Business	Michael Brown	0	2.48803827
1022	22/9/2024 0:00	22/12/2024 0:00	North	Furniture	Land Serve	1246.52	335.81	9	Small Business	John Smith	0.15	26.939800
1023	8/9/2024 0:00	17/11/2024 0:00	East	Office Supplies	Above Year	1004.36	270.26	16	Small Business	John Smith	0.01	26.9086781
1024	3/10/2024 0:00	27/1/2025 0:00	Unknown	Furniture	Quality Head	1268.05	394.17	20	Corporate	Emily Davis	0	31.0847364
1025	28/7/2024 0:00	15/10/2024 0:00	East	Technology	Than Discuss	1553.01	467.65	5	Consumer	Sarah Johnson	0.15	30.1124912
1026	9/4/2024 0:00	12/7/2024 0:00	West	Office Supplies	Drop Personal	1262.92	207.26	14	Home Office	John Smith	0.16	16.411174
1027	24/4/2024 0:00	19/7/2024 0:00	South	Technology	Establish Heavy	1906.82	132.28	8	Small Business	Michael Brown	0.1	6.93720435
1028	12/4/2024 0:00	30/8/2024 0:00	East	Office Supplies	Story Fear	663.08	410.76	9	Home Office	John Smith	0.01	61.9472763

## Section 2: Statistical Analysis

After cleaning the data, perform statistical analysis to derive insights.

### (a) Calculate Summary Statistics for Sales & Profit

Use **Excel Formulas** to compute:

- **Mean (Average)** → =AVERAGE(range)
- **Median** → =MEDIAN(range)
- **Mode** → =MODE.SNGL(range)
- **Variance** → =VAR.P(range)
- **Standard Deviation** → =STDEV.P(range)

	Sales	Profit
Mean	1004.2	250.861
Median	973.93	255.71
Mode	972.365	255.71
Variance	308713	20581.3
Standard deviation	555.62	143.462

## (b) Identify Top-Performing Regions & Customer Segments

1. Use a **PivotTable** to analyze total **Sales & Profit**:

- Select data → Insert → **PivotTable**
- Drag **Region** to Rows
- Drag **Sales & Profit** to Values
- Sort in descending order to find the **top regions & customer segments**.

Row Labels	Sum of Sales	Sum of Profit
<b>South</b>	<b>455879.405</b>	<b>109721.25</b>
Consumer	117036.81	27453.55
Corporate	115982.61	26476.25
Home Office	122344.94	28222.24
Small Business	100515.045	27569.25
<b>East</b>	<b>454669.26</b>	<b>113409.74</b>
Consumer	109761.98	27744.95
Corporate	108490.805	30361.26
Home Office	127862.945	30212.25
Small Business	108553.53	25091.34
<b>West</b>	<b>441560.84</b>	<b>109413.25</b>
Consumer	121126.155	31237.75
Corporate	113525.135	28312.25
Home Office	116658.35	28154.45
Small Business	90251.2	21708.85
<b>North</b>	<b>424526.875</b>	<b>107820.25</b>
Consumer	116366.995	29797.45
Corporate	97864.035	27200.15
Home Office	93232.73	20467.15
Small Business	117063.115	30355.55
<b>Unkown</b>	<b>96483.335</b>	<b>25016.15</b>
Consumer	27051.15	6650.45
Corporate	13923.28	3214.85
Home Office	25477.22	6409.45
Small Business	30031.685	8741.35
<b>Unknown</b>	<b>84060.145</b>	<b>23547.05</b>
Consumer	18166.435	5474.15
Corporate	25541.25	6478.65
Home Office	16397.67	5286.45
Small Business	23954.79	6307.75
<b>Grand Total</b>	<b>1957179.86</b>	<b>488927.65</b>

## (c) Detect Outliers Using the IQR Method

1. Compute **Q1 (25th percentile) & Q3 (75th percentile)** using:

=QUARTILE.INC(range, 1) → Q1

=QUARTILE.INC(range, 3) → Q3

2. Compute **Interquartile Range (IQR)**:  

$$= Q3 - Q1$$
3. Determine **Outlier Boundaries**:  

$$\text{Lower Bound} = Q1 - 1.5 * \text{IQR}$$

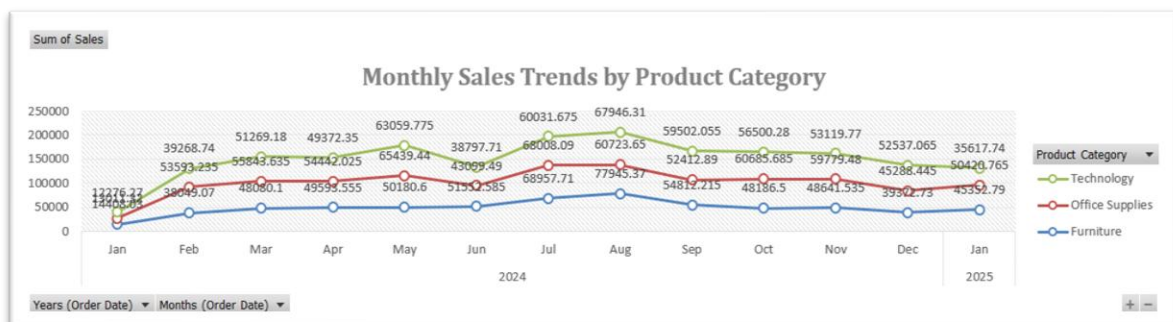
$$\text{Upper Bound} = Q3 + 1.5 * \text{IQR}$$
4. Identify outliers:  
 Use **Conditional Formatting** to highlight values **below Lower Bound** or **above Upper Bound**.

## Section 3: Data Visualization

### Step 1: Create Key Charts

#### 1. Monthly Sales Trend (PivotChart)

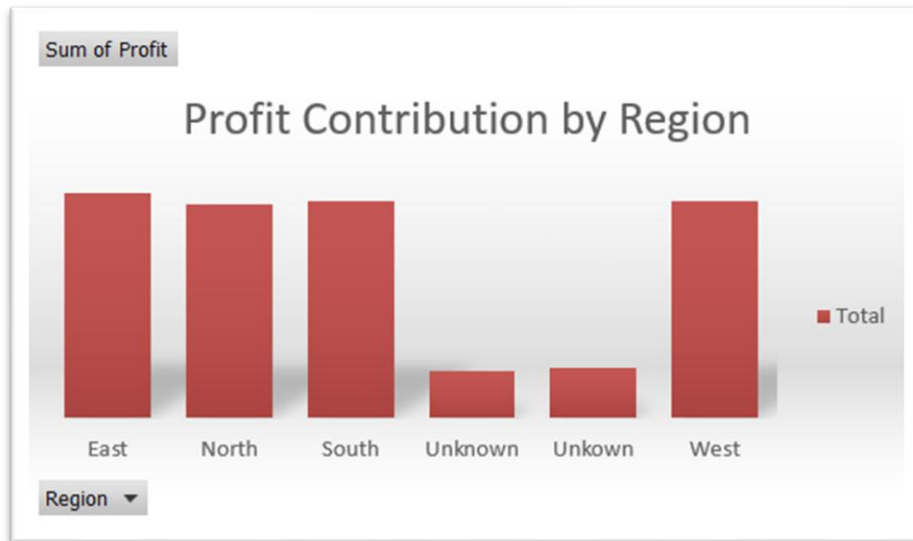
- Create a **PivotTable** with **Order Date (Grouped by Month)** and **Sales**.
- Insert a **Line Chart** to show monthly trends.



#### 2. Regional Profit Contribution (Bar Chart)

- Create a **PivotTable** with **Region** and **Total Profit**.

- Insert a **Bar Chart** to visualize contributions.



### 3. Sales vs. Profit Scatter Plot

- Select **Sales & Profit** columns → Insert → **Scatter Chart**.
- Click on the **data points** → Right-click → **Format Data Series** → Change marker colors.



### Step 2: Add Interactive Slicers

- Click on **PivotTable** → Go to **PivotTable Analyze** → Click **Insert Slicer**.

- Select **Region, Product Category, and Customer Segment** for dynamic filtering.

Region	Product Category	Customer Segment
East	Furniture	Consumer
North	Office Supplies	Corporate
South	Technology	Home Office
Unknown		Small Business
Unkown		
West		