**Exploratory Data Analysis**

**RDAMP-Sales-Analysis**

This project is the first of four projects that I have been working on as part of the Realcare Tech Mark LTD mentorship program. I have decided to utilize my knowledge of Excel and Power BI as tools to create a report that answers foundational business questions. This project gives an overview of key sales performance trends for ACE, a nationwide retail chain, between 2023 - 2025.

This report will serve as a baseline to identify regional performance gaps, customer behaviour patterns, and product category profitability, helping the executive team of ACE focus on high-impact areas in subsequent strategy sessions.

**Dataset Provided**

- Ace Superstore Retail Dataset.csv

- Columns: [Order ID], [Order Date], [Order Mode], [Customer ID], [City], [Postal Code], [Country], [Region], [Product ID], [Product Name], [Category], [Sub-Category], [Sales], [Cost Price], [Quantity], [Discount]

- Store Locations.xlsx

- Columns: [City], [Postal Code], [Country], [Region]

# Tools utilized for data analysis

- Microsoft Excel

- Microsoft Power BI

**Data Cleaning**

**Analysis using Excel**

- The datasets I am working with are Ace Superstore Retail and Store Locations.

- I ensured all columns within each dataset were of the correct datatype.

- Columns Country and Region within the Ace Superstore Retail dataset had missing information in several cells. I used XLOOKUP in excel to fill in the missing information. The columns with similar names within the Store Locations dataset had complete information that was able to fill in the missing information.

- I created two new columns because of the above step called Country-Adjusted and Region-Adjusted and removed the original columns Country and Region.

- I checked both datasets for duplicate rows. None were present.

**Analysis using Power BI**

- I imported the new datasets to Power BI to conduct further EDA and cleaning.

- I noticed that within my Adjusted Region column thaere were both "Yorkshire & the Humber" and "Yorkshire and the Humber". I used the Replace Value operation to replace the "&" with "and" so the datset had the correct Yorkshire and the Humber as a region.

- For the Category column, I needed to separate the information into two separate columns, Category and Segment. The information within the original Category column was separerated by "-". I was able to create the two columns using Custom Column operation using the M-codes:

if Text.Contains([Category], " - ")

then Text.BeforeDelimiter([Category], " - ")

else Text.BeforeDelimiter([Category], " ")

Named column Category1

&

if Text.Contains([Category], " - ")

then Text.AfterDelimiter([Category], " - ")

else Text.AfterDelimiter([Category], " ")

Named column Segment

- I removed error from the column named Category1.

- The segment column created had null values as a result of the initial Category column containing information without "-"

- Filled the null values in the Segment column by creating another custom column using the M-code:

if [Segment] = null or Text.Trim([Segment]) = "" then [Category] else [Segment]

Named new column Segment1

- Removed original Category column and the Segment column and renamed the Category1 column as Category and the Segment1 column as Segment.

- Discount column had null values so I treating these as no discount and fill the null values with 0s.

- I promoted the first row as header for the Store Locations dataset.

- Renamed tables: Ace Superstore RetailFact and Store LocationsDim

- Changed the Order Date datatype to Short Date format.

**Creation of Measure using Power BI**

- Created the following measures to aid in analysis prior to visualization:

1. Gross Profit per Unit =

AVERAGEX(

FILTER('Ace Superstore RetailFact', 'Ace Superstore RetailFact'[Sales] > 0 && 'Ace Superstore RetailFact'[Cost Price] > 0),

'Ace Superstore RetailFact'[Sales] - 'Ace Superstore RetailFact'[Cost Price]

)

2. Profit Margin (%) =

AVERAGEX(

FILTER('Ace Superstore RetailFact', 'Ace Superstore RetailFact'[Sales] > 0 && 'Ace Superstore RetailFact'[Cost Price] > 0),

DIVIDE('Ace Superstore RetailFact'[Sales] - 'Ace Superstore RetailFact'[Cost Price], 'Ace Superstore RetailFact'[Sales])

)

3. Profit per Unit = Sum('Ace Superstore RetailFact'[Sales]) - Sum('Ace Superstore RetailFact'[Cost Price])

4. Total Cost =

SUMX(

'Ace Superstore RetailFact',

IF('Ace Superstore RetailFact'[Cost Price] > 0, 'Ace Superstore RetailFact'[Cost Price] \* 'Ace Superstore RetailFact'[Quantity], 0)

)

5. Total Discount = AVERAGE('Ace Superstore RetailFact'[Discount])

6. Total Revenue =

SUMX(

'Ace Superstore RetailFact',

IF('Ace Superstore RetailFact'[Sales] > 0, 'Ace Superstore RetailFact'[Sales] \* 'Ace Superstore RetailFact'[Quantity], 0)

)

7. Total Units = SUM('Ace Superstore RetailFact'[Quantity])

- As the Sales and Cost Price columns contained negative values, I created the above measures to manage these without the analysis being affected.