**Retail Business Performance & Profitability Analysis**

# Project Objective

The primary objective of this project is to analyze the performance of a retail business in terms of sales, profit, and profitability across various dimensions such as product categories, customer segments, regions, and time periods. The goal is to derive actionable insights that support better business decision-making.

# Tools & Technologies Used

- Power BI: For interactive visual dashboards

- Python (Pandas): For initial data cleaning and preprocessing

- Excel: For raw data storage and exploration

# Dataset Overview

The dataset includes:  
- Sales Data: Order ID, Order Date, Ship Date, Quantity, Discount, Sales, Profit  
- Customer Info: Customer Name, Segment  
- Product Info: Product Name, Category, Sub-Category  
- Geographic Info: Region, State, City  
- Calculated Metrics: Profit Margin = Profit / Sales

# Dashboard Summary

## Page 1: Overview Dashboard

- KPIs: Total Sales, Total Profit, Average Profit Margin  
- Sales trend over time  
- Sales by region (bar/donut chart)  
-Profit by category and subcategory

## Page 2: Profitability Analysis Dashboard

- Profit margin by category and sub-category  
-Segment based analysis by using slicers   
- Relation between discount and profit by category by using Scatterplo

## Page 3: Top Performers & Loss Makers Dashboard

- Top 5 products by profit  
- Bottom 5 products by losses  
- Visuals sorted using Top N filters

# Key Findings

- Technology was the highest selling category but not always the most profitable.  
- Some products consistently generated negative profit (losses) despite sales volume.  
- Certain customer segments generated higher profit margins.  
- Regional analysis revealed variations in performance based on geography.

# Recommendations

- Promote and stock high-profit margin products more aggressively.  
- Investigate and possibly discontinue or reprice loss-making products.  
- Use customer segmentation insights to offer targeted marketing.  
- Consider regional pricing strategies based on regional performance trends.

# Conclusion

This analysis helps retail stakeholders identify where the business is succeeding and where improvements are needed. With the interactive Power BI dashboards, decision-makers can dynamically explore the data and apply filters to generate insights for specific periods, categories, or customer groups.

# Appendix

- Power BI Report File **(.[pbix](../OneDrive/Documents/Dashboards.pbix))**

- Processed dataset (.[csv](processedData_forPowerbi.csv))

**- Python code for data cleaning :**

import pandas as pd

df= pd.read\_csv("Superstore.csv")

#df.info()

# To check the null values are there or not

df.isnull().values.any()

# Here it returns false, so null values are not there in the data set

#Convert date columns to datetime

df['Order Date'] = pd.to\_datetime(df['Order Date'], errors='coerce')

df['Ship Date'] = pd.to\_datetime(df['Ship Date'], errors='coerce')

#Standardize column names

df.columns = df.columns.str.strip()

df.columns = df.columns.str.lower()

df.columns = df.columns.str.replace(' ', '\_')

df.dropna(inplace=True)

#Remove duplicate rows

df\_cleaned=df.drop\_duplicates()

#Remove outliers using IQR method

def remove\_outliers\_iqr(dataframe, column):

Q1 = dataframe[column].quantile(0.25)

Q3 = dataframe[column].quantile(0.75)

IQR = Q3 - Q1

lower\_bound = Q1 - 1.5 \* IQR

upper\_bound = Q3 + 1.5 \* IQR

return dataframe[(dataframe[column] >= lower\_bound) & (dataframe[column] <= upper\_bound)]

# Apply to selected numeric columns

for col in ['sales', 'profit', 'discount']:

df\_cleaned = remove\_outliers\_iqr(df, col)

df\_cleaned

#Save the cleaned data

#df.to\_excel("Cleaned\_Superstore.xlsx", index=False)

**After Cleaning, python code for processed Data:**

import pandas as pd

# Load cleaned dataset

df = pd.read\_csv('CleanedDataSet.csv')

#print(df.info())

df['Profit Margin'] = df['profit']/df['sales']

# 1. Profit Margin by Category and Sub-Category

category\_margin = df.groupby(['category', 'sub-category'])[['sales', 'profit']].sum().reset\_index()

category\_margin['Profit Margin']=category\_margin['profit']/ category\_margin['sales']

#2. Sales & Profit by Region

region\_margin=df.groupby('region')[['sales','profit']].sum().reset\_index()

#3. Discount vs Profitability correlation

dis\_corr=df[['discount','profit']].corr()

# Discount vs Profitability correlation by using scatterplot

import seaborn as sns

import matplotlib.pyplot as plt

sns.scatterplot(data=df, x='discount',y='profit')

plt.title("Discount vs Profit")

plt.show()

#Save the Final Data for power bi

df.to\_excel("processedData\_forPowerbi.xlsx",index=False)