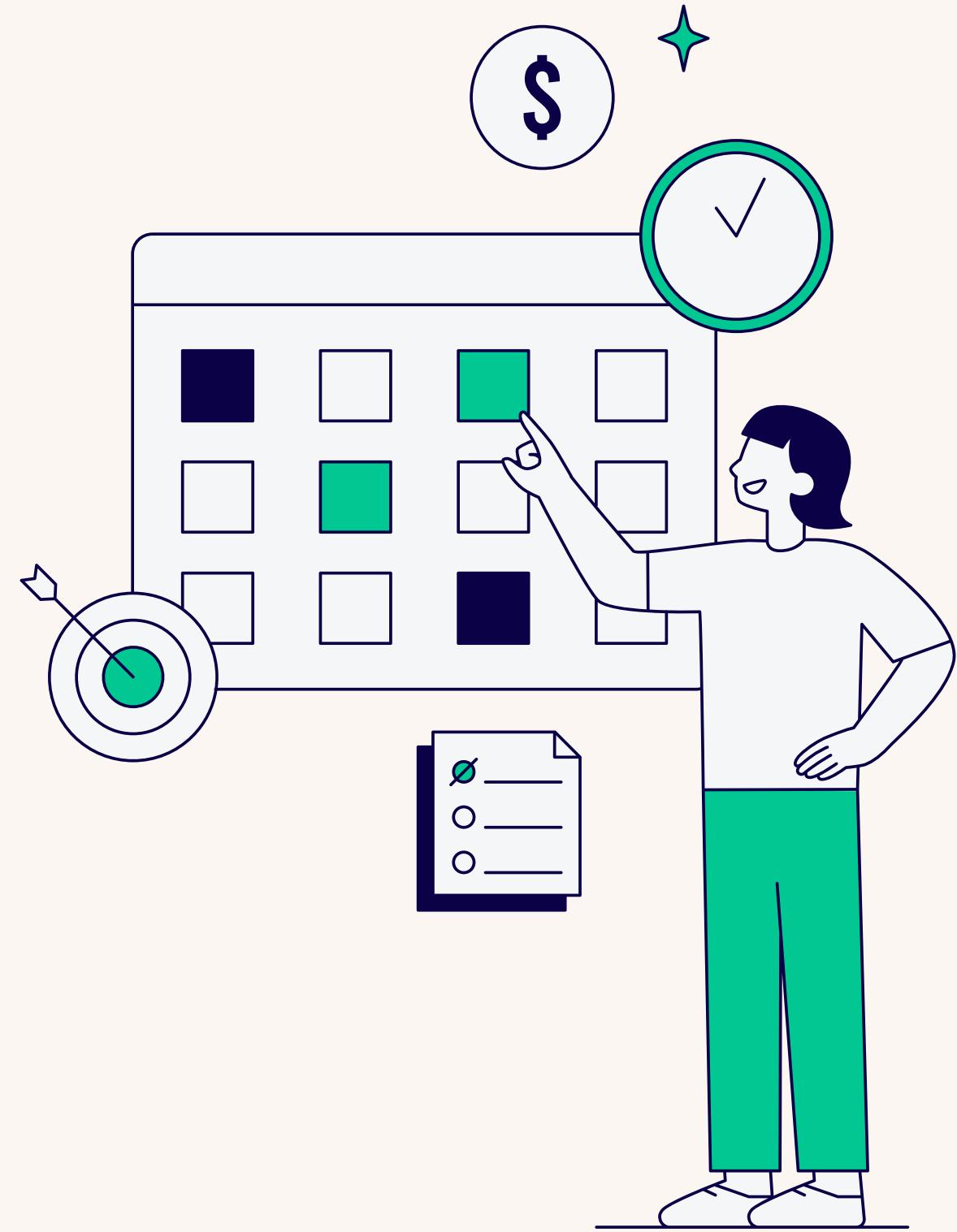


Presented by Soraia Tavares

# Retail Sales & Profit Analysis

## A Superstore Case Study

March 2025



# Project Brief

## Overview:

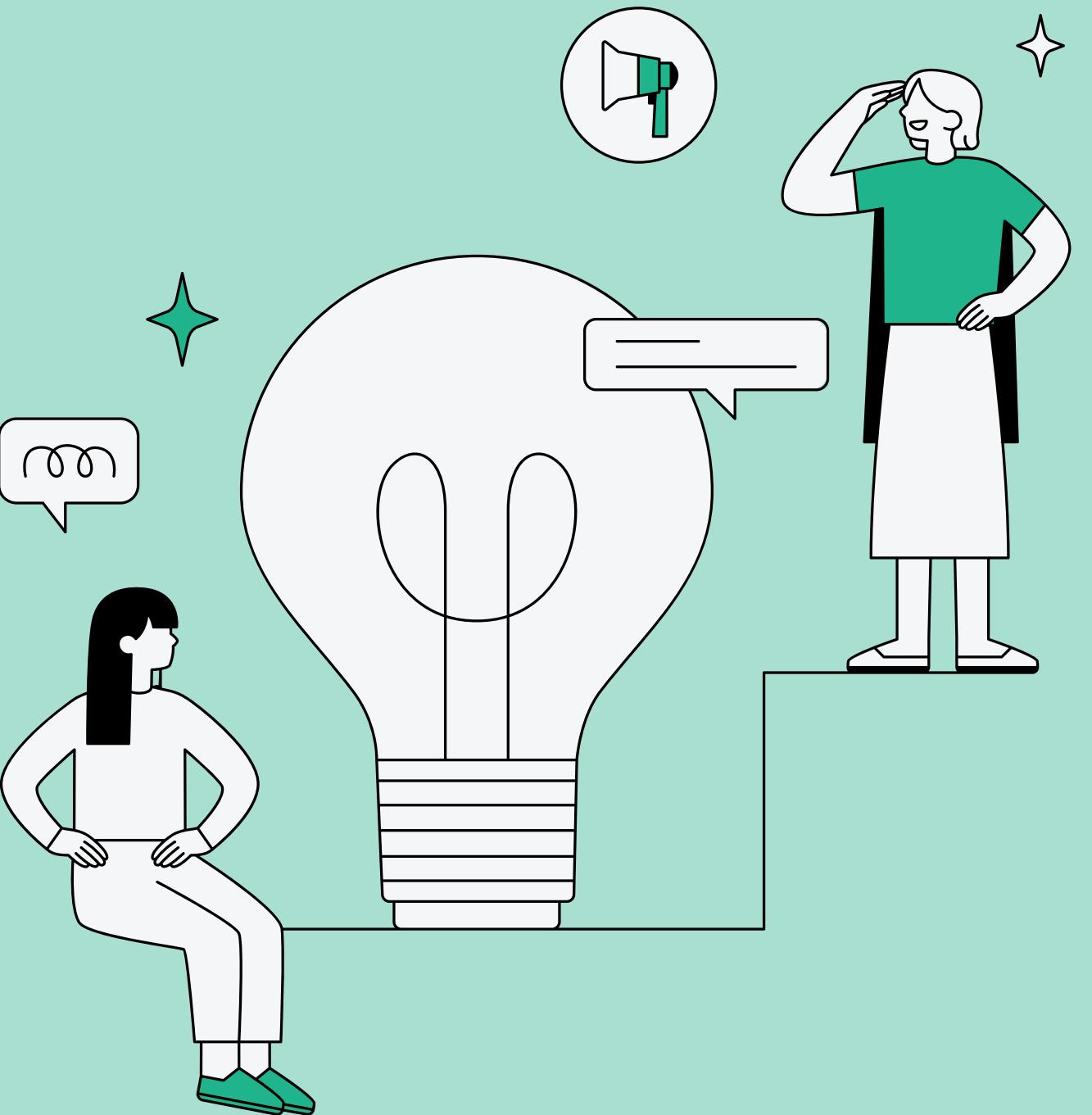
- Analysing Superstore sales data to uncover trends in customer behavior, profitability, and operational efficiency.

## Objectives:

- Identify key sales drivers
- Analyse profit trends
- Recommend strategies for growth
- 

## Key Questions:

- What factors influence sales and profit the most?
- Which customer segment generates the most revenue?
- How do discounts impact profitability?



# Dataset Overview

Source:

[Superstore Sales 2023 Dataset](#)

Description:

Contains transaction-level sales data, including customer details, product categories, sales amount, profit, and discounts applied.

Relevance:

Helps analyse sales performance and customer behaviour.



# Dataset Structure

## Key Variables:

- Order ID (Unique transaction identifier)
  - Customer Name & Segment
  - Product Category & Sub-Category
    - Sales, Profit, and Discount
    - Order Date & Ship Date

## Data Types:

- Categorical: Customer Segment, Product Category
  - Numerical: Sales, Profit, Discount
  - Date: Order Date, Ship Date

# Main challenges identified

01.

Missing Postal Codes  
→ Replaced blanks  
with "Unknown"

02.

Outliers in Sales & Profit  
→ Verified high-profit  
sales (copiers) and  
negative profit (high-  
discounted items)

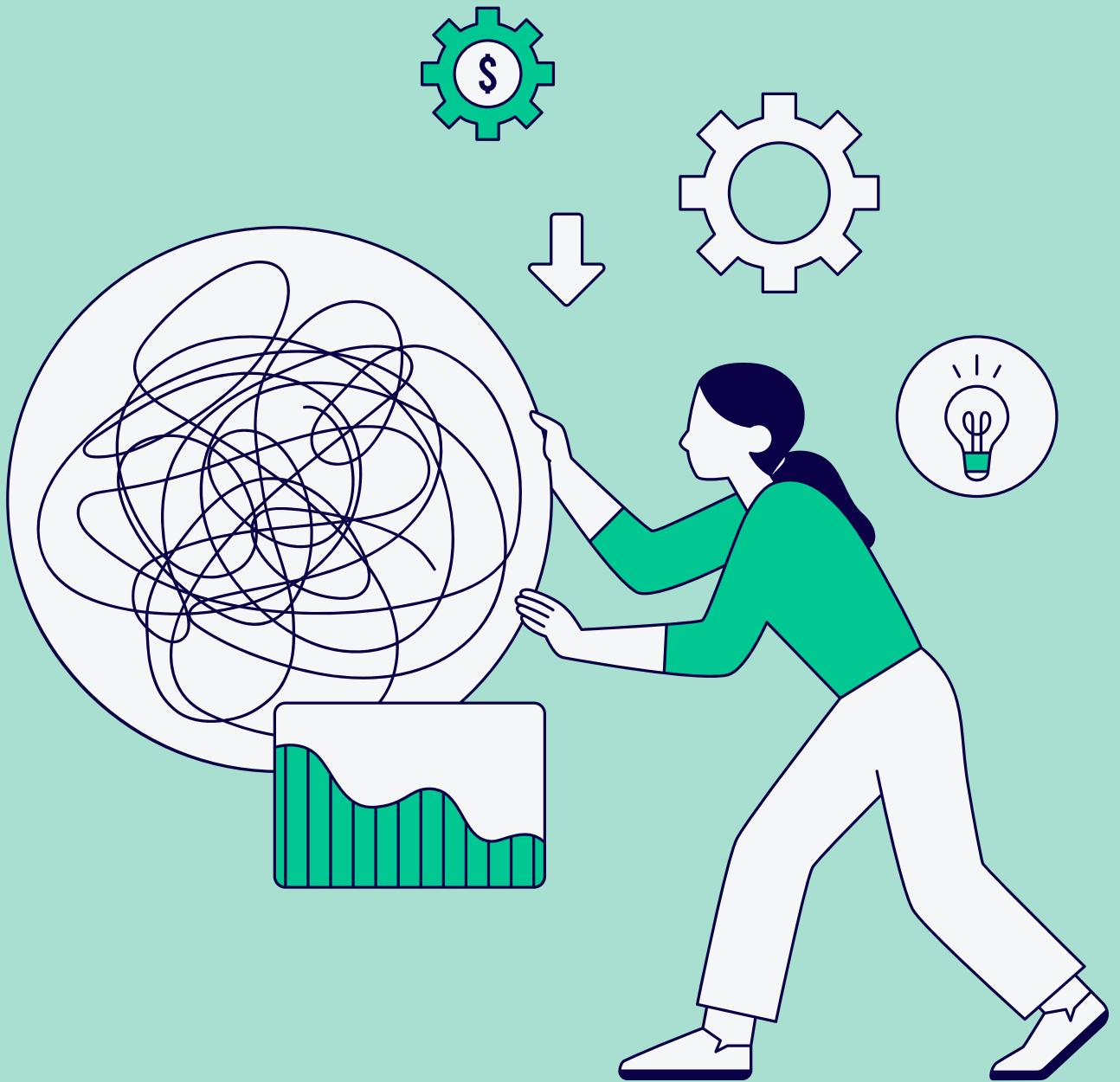
03.

Inconsistent Formatting  
→ Fixed date formats &  
standardised column  
names.

# Data Cleaning Process

## Steps Taken:

- Handled missing values → "Unknown" for missing Postal Codes.
- Standardised Formatting → Dates converted to DD/MM/YYYY, numbers formatted properly.
- Managed Outliers → Verified that negative profits resulted from high discounts and were legitimate.
- Order Date and Ship Date were initially swapped → corrected for accurate shipping analysis.
- RFM scoring was enabled by transforming order-level data into customer-level aggregations.



# Data Transformation

01.

## New Variables:

- **Profit Margin %** = Profit / Sales \* 100 (To analyse product profitability).
- **Loss Flag** = IF(Profit < 0, "Loss", "Profit") (To quickly identify loss-making transactions).
- **Extracted Order Month** = TEXT(Order Date, "MMMM") (To analyse sales trends over time).

02.

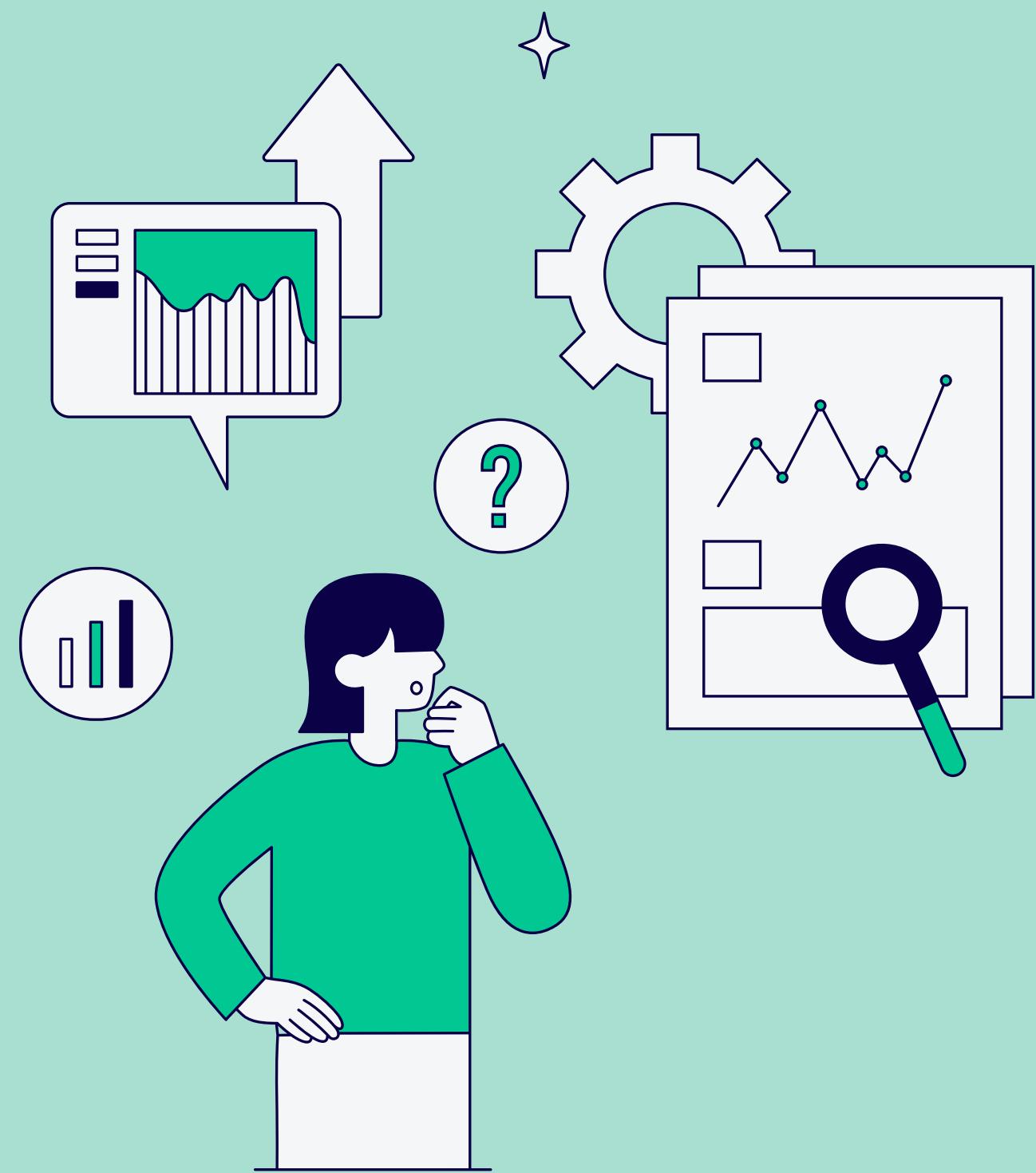
Encoded categorical data → Customer Segments, Product Categories used for analysis.

03.

- Examined **Discount Impact** by analysing how different discount levels affected profit
- Added a cut-off reference date for Recency scoring to ensure time-based consistency in RFM segmentation.

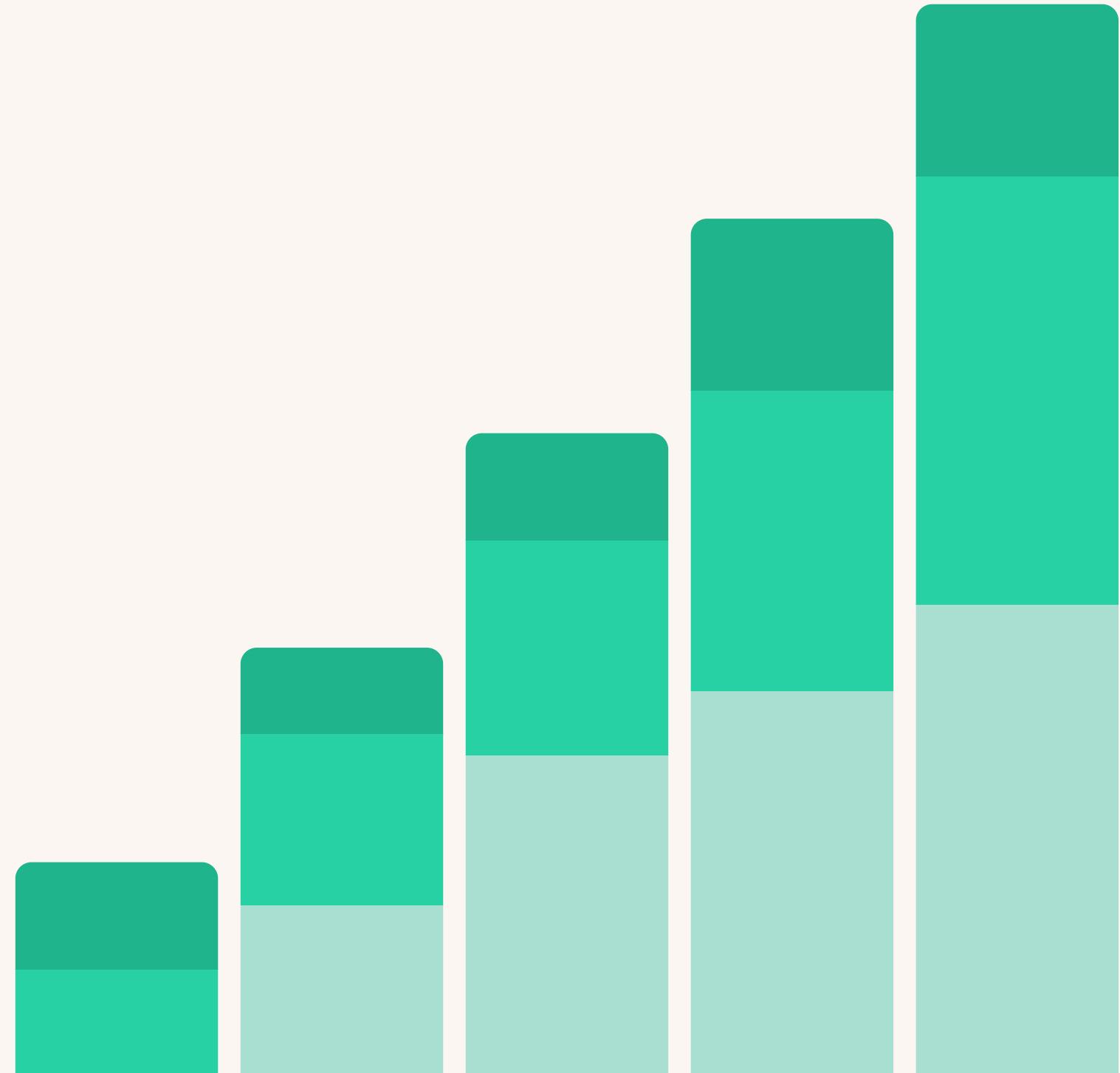
# Exploratory & Descriptive Analysis (EDA)

- **Top-performing category** → **Technology** had the highest profit margin.
- **Discount vs. Profit** → **Higher discounts led to negative profit margins** in some categories.
- **Customer Segment** → **Corporate** customers generated the highest revenue.



# Visual Analysis

To better understand **key trends and insights**, I have created interactive visualisations using Tableau and Power BI. These visuals provide a clear representation of sales performance, customer behaviour, profitability, and discount impact. The following pages will showcase key findings from my analysis.





# Foundational Insights from Initial Analysis



- 📈 **Sales performance** varies by state and product category, with **Technology** leading in **revenue**.
- 📦 **Office Supplies** and **Furniture** are widely purchased across customer segments, especially in the **East** and **West regions**.
- 📈 **Sales** have **increased steadily** over time, with **strong performance** in **Q4**.
- 💡 To explore **who** is driving these results, I applied **RFM segmentation** — analysing customer behaviour by **recency**, **frequency**, and **monetary** value.

# Customer Segmentation Using RFM

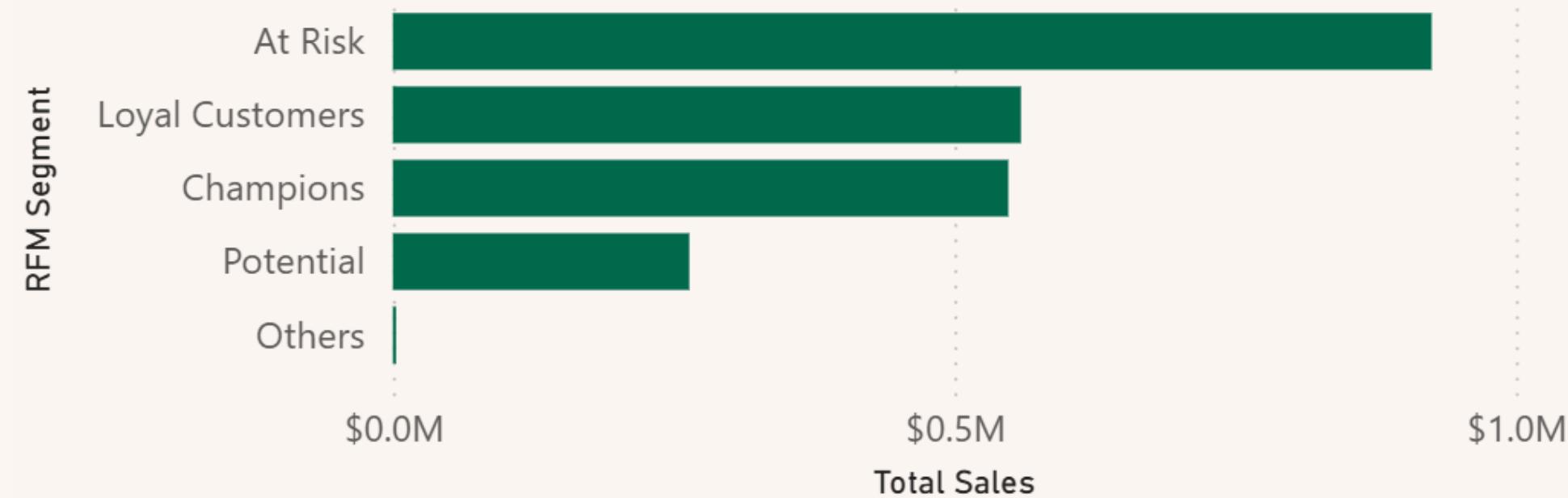
To better understand customer behaviour, I applied RFM segmentation – grouping customers based on:

- **Recency**: How recently they made a purchase
- **Frequency**: How often they made purchases
- **Monetary**: How much they spent

This allowed me to categorise customers into meaningful segments, such as:

- 🏆 Champions
- 💡 Loyal Customers
- ✅ Potential
- ⚠️ At Risk

Sales by RFM Segment



Note: Recency was calculated using a cut-off date of 28/02/2021, based on the latest order in the dataset.

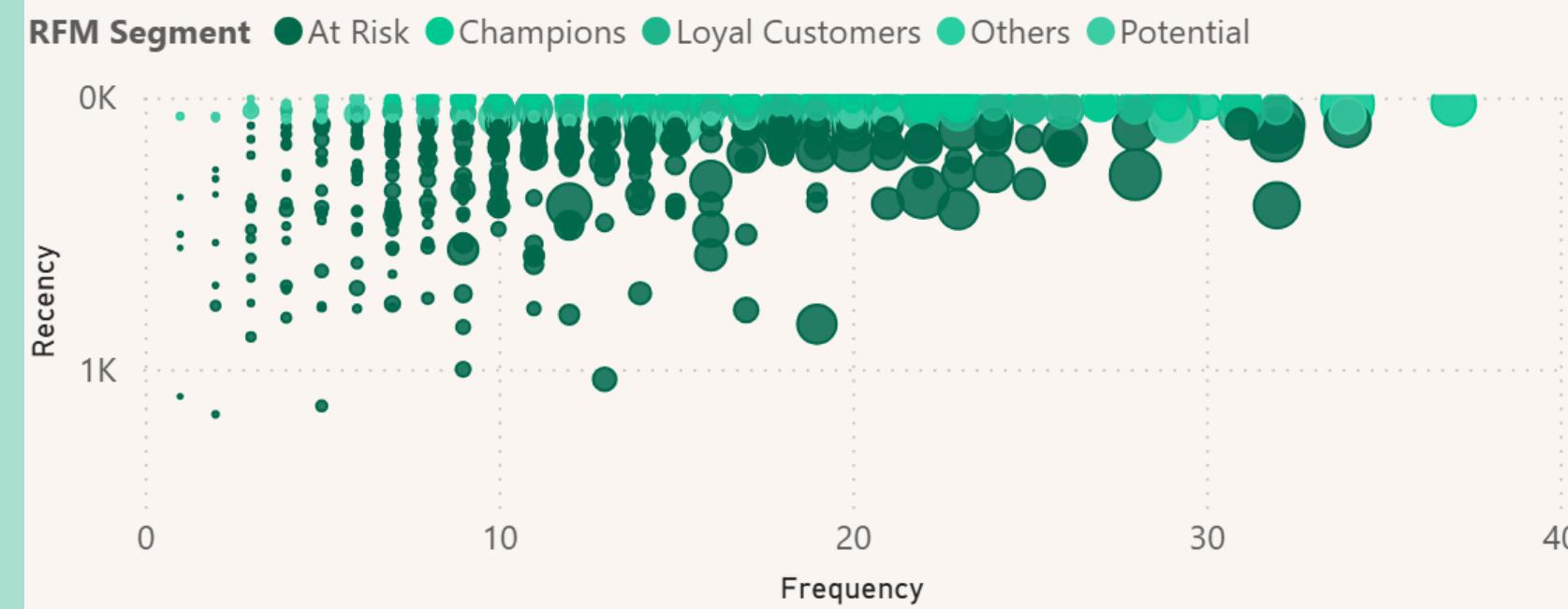
# Key Insights from RFM Segmentation

- 🏆 **Champions** represent **\$548K** in sales — they have the **highest average order value** and **purchase frequently**.
- ⚠️ **At Risk** customers contribute **40%** of total revenue but **haven't purchased** in a **long time**.
- 📈 Nearly **90%** of sales are from **recent customers (0–30 days)**, showing **strong short-term engagement**.
- 📦 **Loyal Customers** prefer categories like **Office Supplies** and **Technology**, making them ideal for **upselling**.

## Average Order Value by RFM Segment



## Customer Distribution by Recency & Frequency (RFM Segment)





# Customer Strategy Based on RFM Segmentation

## Segment-Specific Action Plan

Action	Target Segment	Rationale
🎁 Launch exclusive loyalty offers	🏆 Champions	Retain your highest-value, most engaged customers
✉️ Send win-back email campaigns	⚠️ At Risk	Recover revenue from customers who are fading
🧠 Personalise upselling emails	📦 Loyal Customers	Boost order value with product-focused recommendations
🛍 Promote tech bundles	🌱 Potential	Mirror Champions' behaviour to build loyalty
🔄 Reinforce retention strategy	📈 Recently Active	Keep engagement high through regular follow-ups

# Broader Business Recommendations

In addition to the customer-specific actions from the RFM analysis, here are broader strategic recommendations based on sales performance and category trends:

 **Focus marketing on high-performing regions (e.g., West, East)**

 **Consider bundling Office Supplies & Furniture for volume incentives**

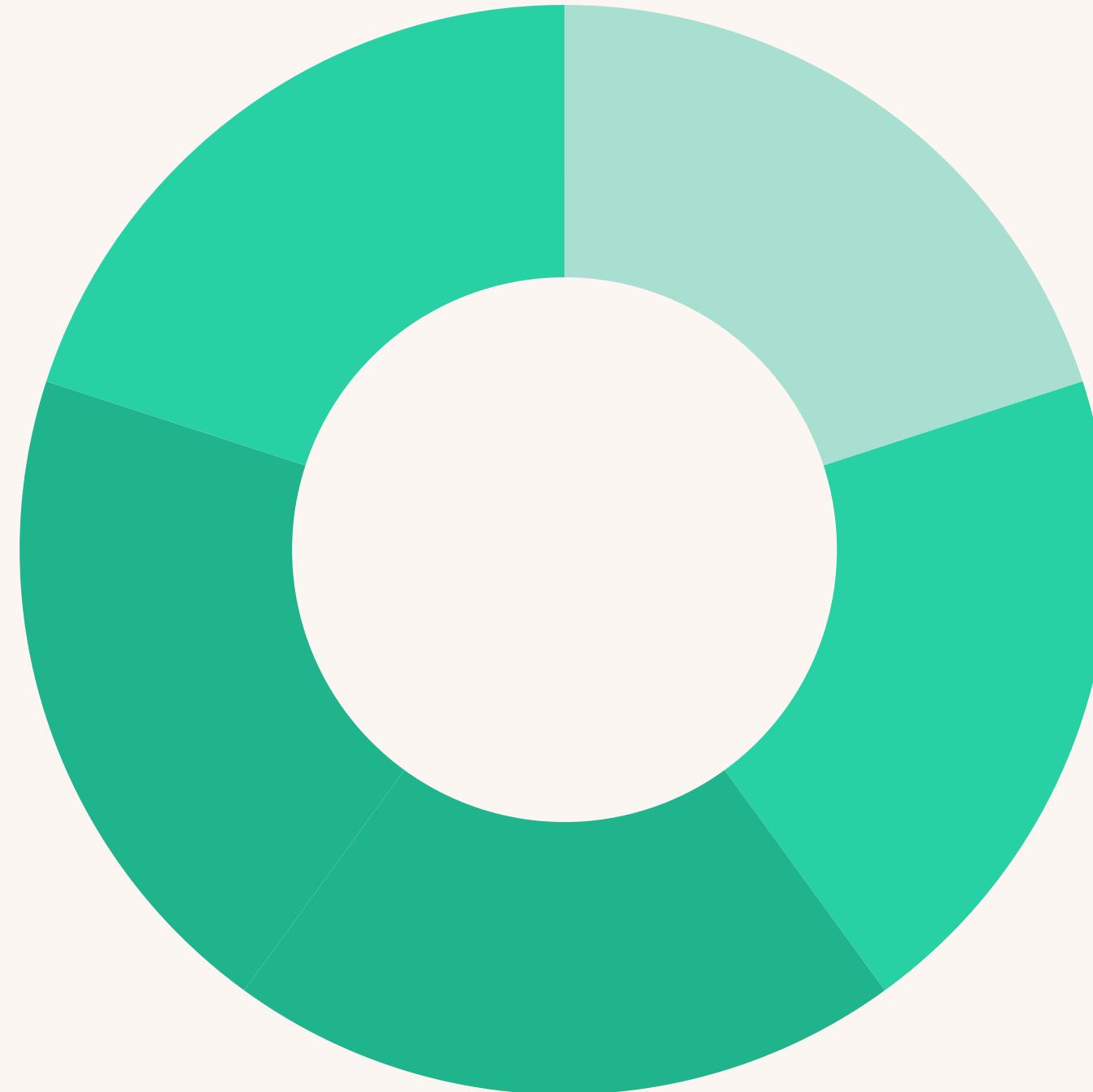


**Leverage Q4 seasonal peaks with targeted promotions**

 **Expand top-performing categories (e.g., Technology)**

# Next Steps & Strategic Opportunities

- Track customer movement across RFM segments
- Monitor changes in customer behavior over time to identify lifecycle shifts and retention opportunities.
- Combine RFM with product preferences
- Enhance segmentation by layering in product/category-level insights (e.g., Champions who prefer Technology).
- Run targeted campaigns by segment
- Test promotions, loyalty programs, and win-back strategies tailored to RFM groups and measure outcomes.
- Automate RFM updates in Power BI
- Set up dynamic refreshes to keep the dashboard current and usable for ongoing decision-making.
- Expand to additional customer features
- Include demographics, feedback, or channel preferences for deeper personalisation.



# Limitations of the Analysis

- The data only covers up to December 2022, which may not capture newer trends or seasonality shifts in 2023 and beyond.
- Some entries may contain inconsistencies or inaccuracies that could affect segmentation accuracy (e.g., outlier sales or missing dates).
- External factors (e.g., market conditions) not included
- Dataset may not represent future trends accurately
- Future iterations should incorporate automated data refreshes and deeper external context for more actionable insights.



Any questions? Let's Discuss!

# Thank you very much!

soraia.tavares@outlook.com

 [www.linkedin.com/in/soraiatavares](https://www.linkedin.com/in/soraiatavares)

 Access the full Power BI report online: [\[dashboard link\]](#)

