

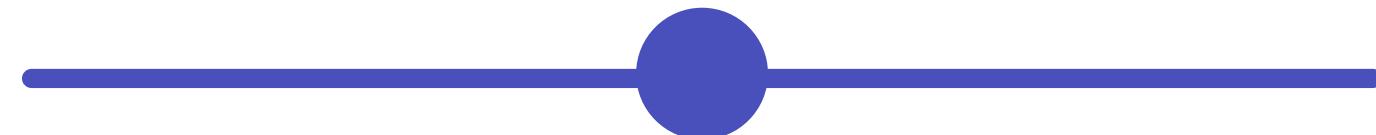
# Superstore Sales Dataset Analysis



# Our Project Team



**Nada Elsayed**



**Mariam Ehab**



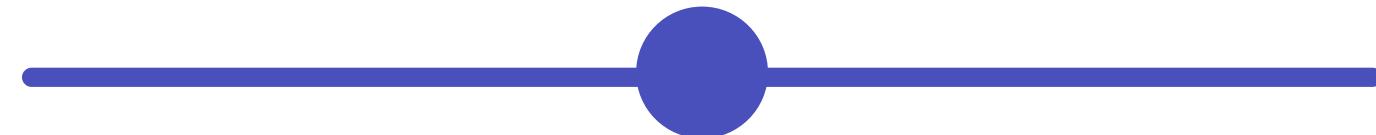
**Maha Ali**



**Esraa Saeed**



**Tabarak Mahmoud**



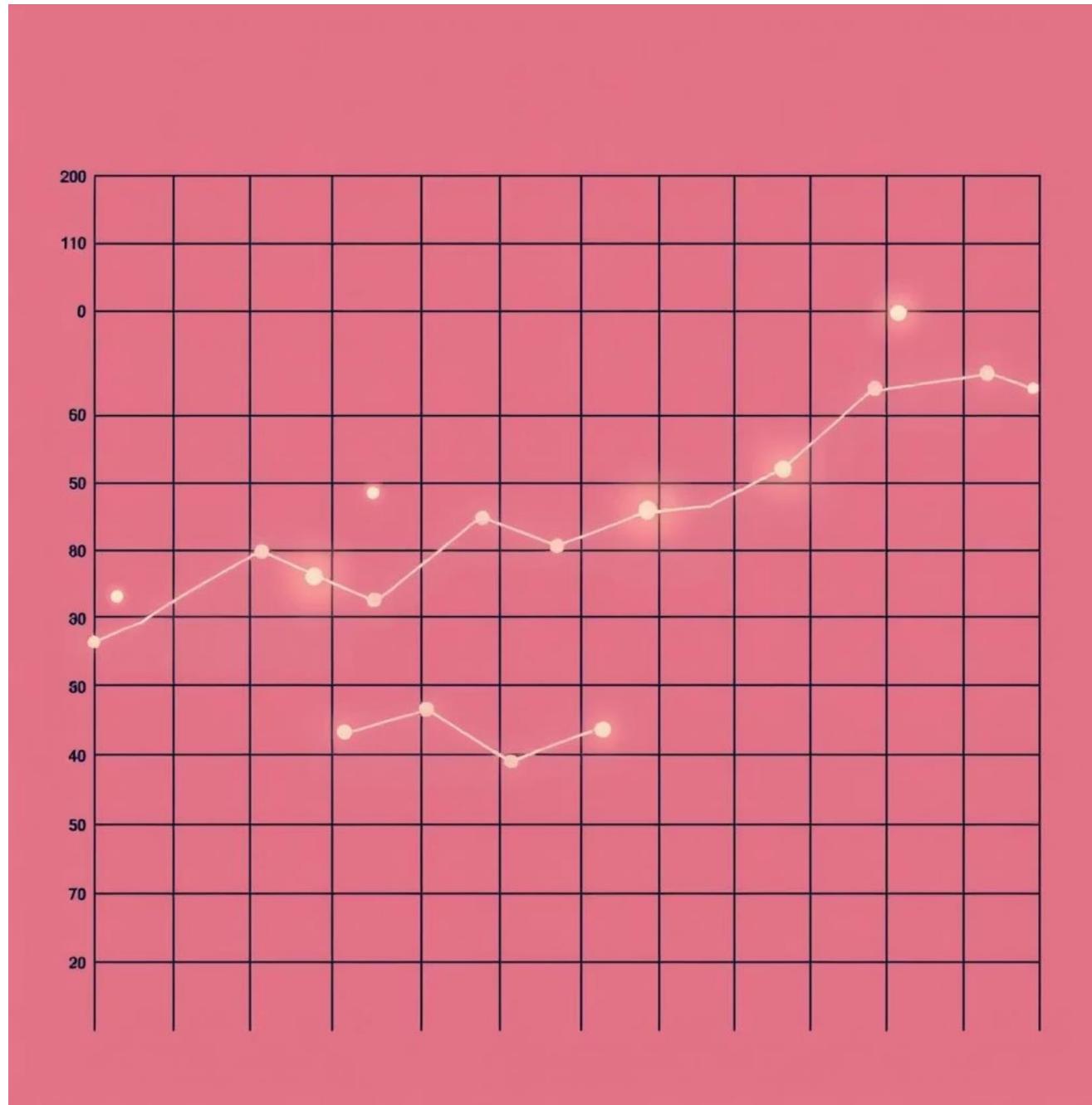
**Rana Mohamed**



## Project Goal: Unlocking Sales Potential

- Our primary objective is to meticulously analyse sales performance and profitability within the Superstore dataset. By delving deep into the available data, we aim to uncover crucial insights that will empower stakeholders to make more informed and strategic decisions, ultimately driving business growth.

# Dataset Overview



The Superstore Sales Dataset is a comprehensive collection of transactional data. It comprises 9,800 rows, each representing a unique sales record, and 18 distinct columns capturing various attributes of each transaction.

- **Shape:** 9,800 rows, 18 columns
- **Data Types:** Includes both numerical (e.g., Sales, Discount) and categorical (e.g., Region, Category, Product Name) data, offering a rich basis for analysis.
- **Completeness:** The dataset is remarkably clean, with no missing values or duplicate entries, ensuring data integrity for analysis.

# Identifying Key Limitations

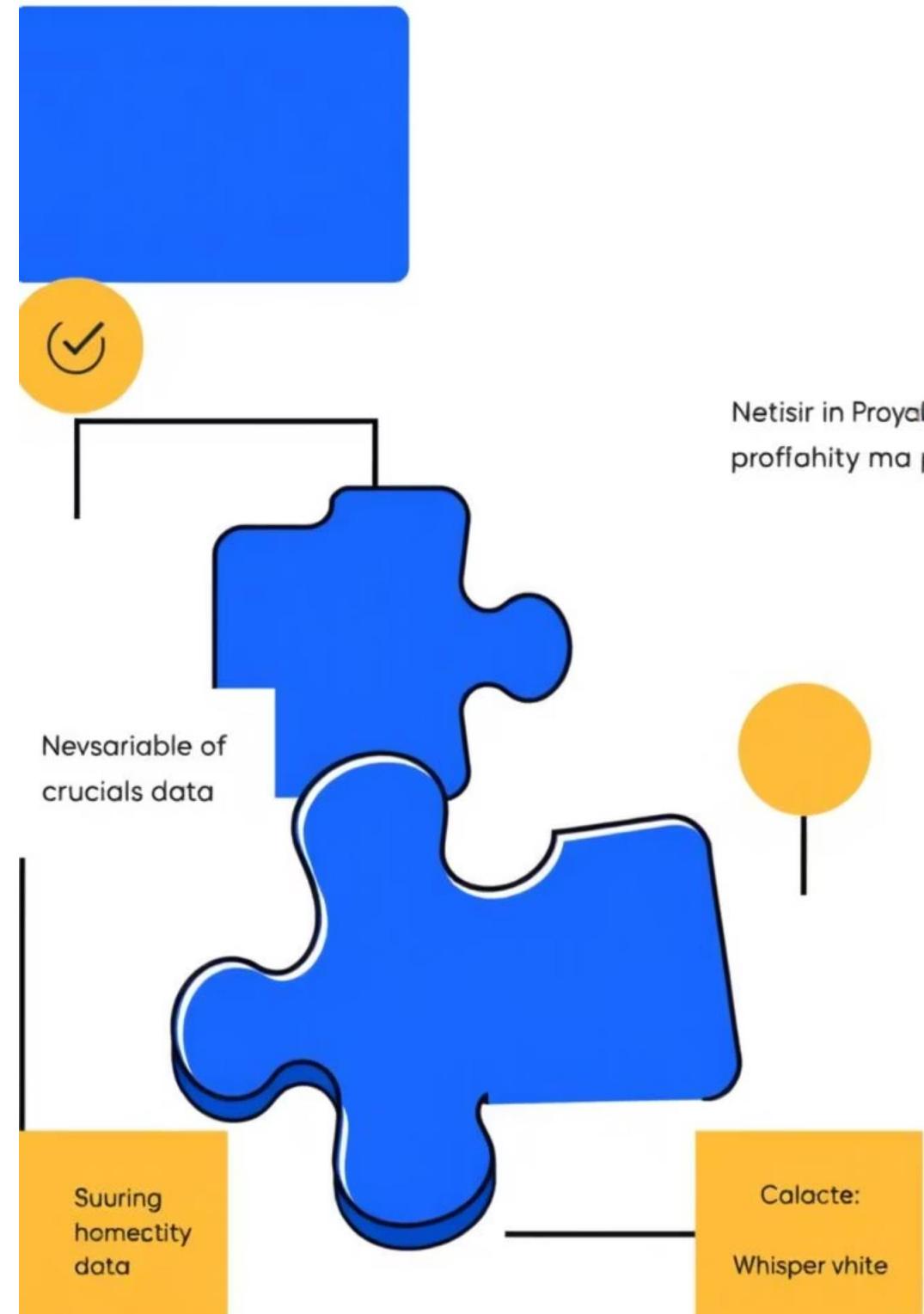
While the Superstore dataset is robust, a critical challenge emerged during our initial assessment. The absence of specific key columns significantly constrains our ability to conduct a truly comprehensive profitability analysis.

## Missing Columns for Profitability

Crucially, the dataset lacks direct columns for **Quantity**, **Cost**, and **Profit**. This limitation restricts our analysis primarily to sales revenue.

## Impact on Analysis Depth

Without these essential metrics, we cannot accurately calculate margins, evaluate product-level profitability, or understand the cost implications of sales strategies. This limits insights to surface-level sales performance.



# Data Transformation

To overcome the current dataset limitations and unlock deeper insights, we recommend adding new columns



# Preliminary Data Exploration

Despite the identified data limitations, our initial exploration has already yielded valuable insights into the superstore's sales dynamics, focusing on patterns discernible from available sales figures.

## 1 Top-Selling Products & Categories

We have identified the highest-grossing products and product categories, providing an initial understanding of customer preferences and market demand.

## 2 Regional Sales Performance

Analysis by region reveals geographical strengths and areas for potential improvement, indicating where sales strategies might need to be adapted.

## 3 Sales Trends Over Time

Examination of sales trends across different periods (e.g., monthly, quarterly) helps us understand seasonality and growth trajectories, informing future planning.



# Unlocking Insights: Superstore Sales Data Analysis

# Data Cleaning and Feature Engineering

We enhanced the dataset by creating new, valuable columns, providing deeper analytical capabilities and richer business understanding.



## Delivery Days

**Calculated as Ship Date - Order Date**, this metric is crucial for analysing shipping performance, identifying potential delays, and optimising logistics.



## Order Month

Derived from Order Date, this column helps us pinpoint seasonal sales trends, forecast demand, and plan inventory more effectively throughout the year.



## Order Year

Extracted from Order Date, this feature enables us to track year-on-year growth, compare performance, and identify long-term market shifts.

# Feature Engineering: Unveiling Financial Metrics

Our feature engineering also focused on creating critical financial metrics to support granular cost and profitability analysis.



## Cost per Unit

Determined by **Cost / Quantity**, this column provides a per-unit cost benchmark, essential for detailed cost analysis and margin calculations.



## Selling Price per Unit

Calculated as **Sales / Quantity**, this helps us understand our pricing strategies and their impact on revenue at an individual product level.

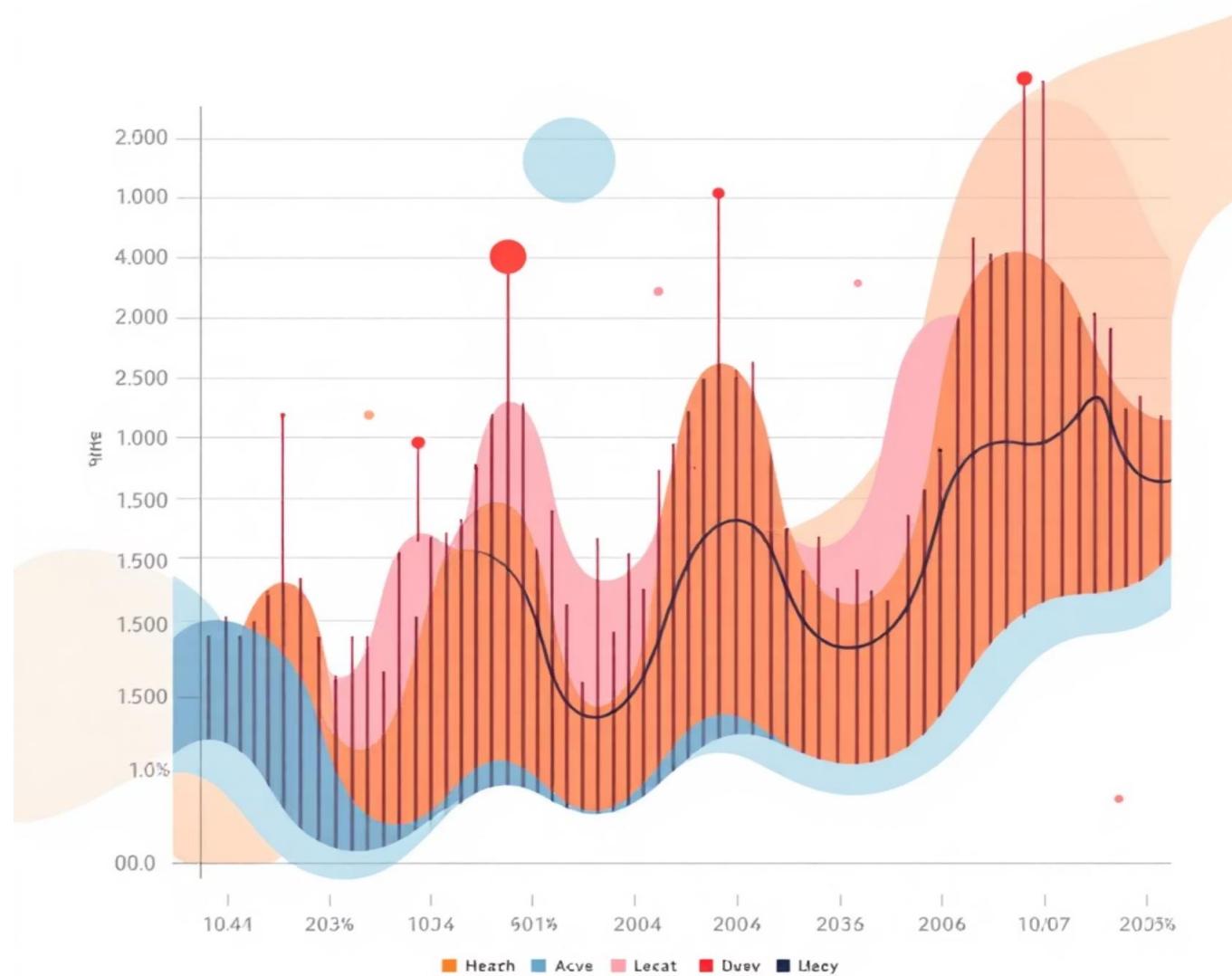


## Total Profit

The ultimate measure, **Sales - Cost**, representing the core profitability of each transaction, product, and customer segment.

# Addressing Outliers: A Realistic Approach

Our analysis identified approximately 1,145 outliers across Sales, Cost, and Profit data points.



**It is common to encounter outliers in real-world business datasets. These often represent legitimate, albeit extreme, transactions such as large corporate orders or significant promotional sales.**

**While statistical models might suggest their removal, we recommend retaining these outliers. They provide a complete picture of business activity and represent natural variations within the market.**

**Excluding them could lead to an incomplete or biased understanding of our operational landscape.**



# Data Modeling: Star Schema Design

---

- To optimise data retrieval and analysis, we implemented a Star Schema design, clearly distinguishing between fact and dimension tables.

# Data Model Components

## Fact Table

Row ID, Order ID, Sales, Quantity, Cost, Profit Per Unit, Delivery Days, and pricing data.

## Customer Dimension

Customer ID and Customer Name

## Product Dimension

Product ID, Category, Sub-Category, and Product Name for inventory insights.

## Date Dimension

Order Date, Order Month, Order Year, and Ship Date

## Location Dimension

Country, City, State, Postal Code, and Region for spatial analysis.

## Segment Dimension

segment

## Shipment Dimension

ship mode

# Supersales Dataset

## Scopes

General Measures

Geographical Distribution

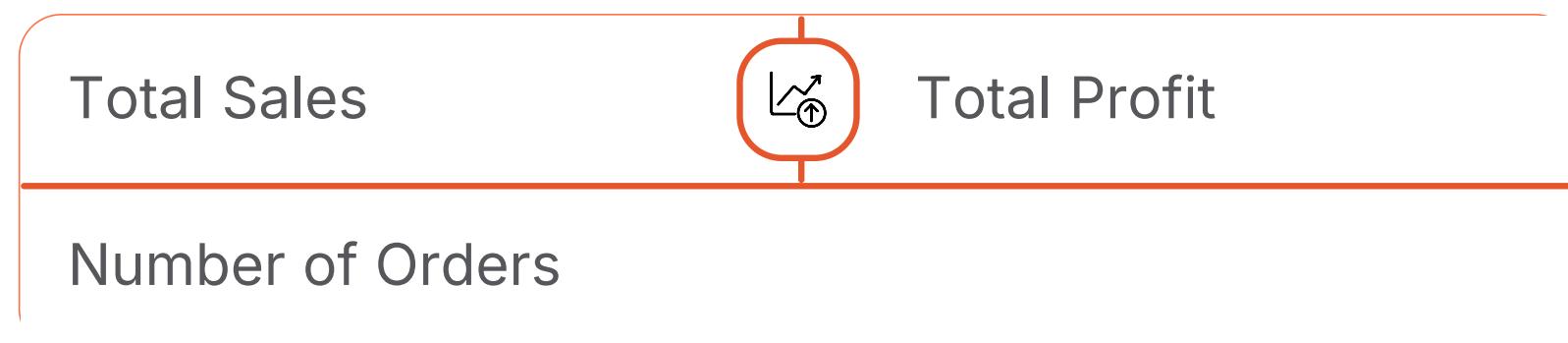
Category Analysis

Sales Performance

Shipments

# General Measures

Core financial metrics that provide a snapshot of overall business health and profitability



# Geographical Distribution

Understand customer reach and profitability across different markets and regions

## Number of Customers



Track regional customer acquisition and market penetration

## Average Profit per Customer



Measure customer value and profitability by geography





# Category and Product Analysis

Evaluate product performance, profitability, and cost structure across your portfolio

Total Quantity Sold

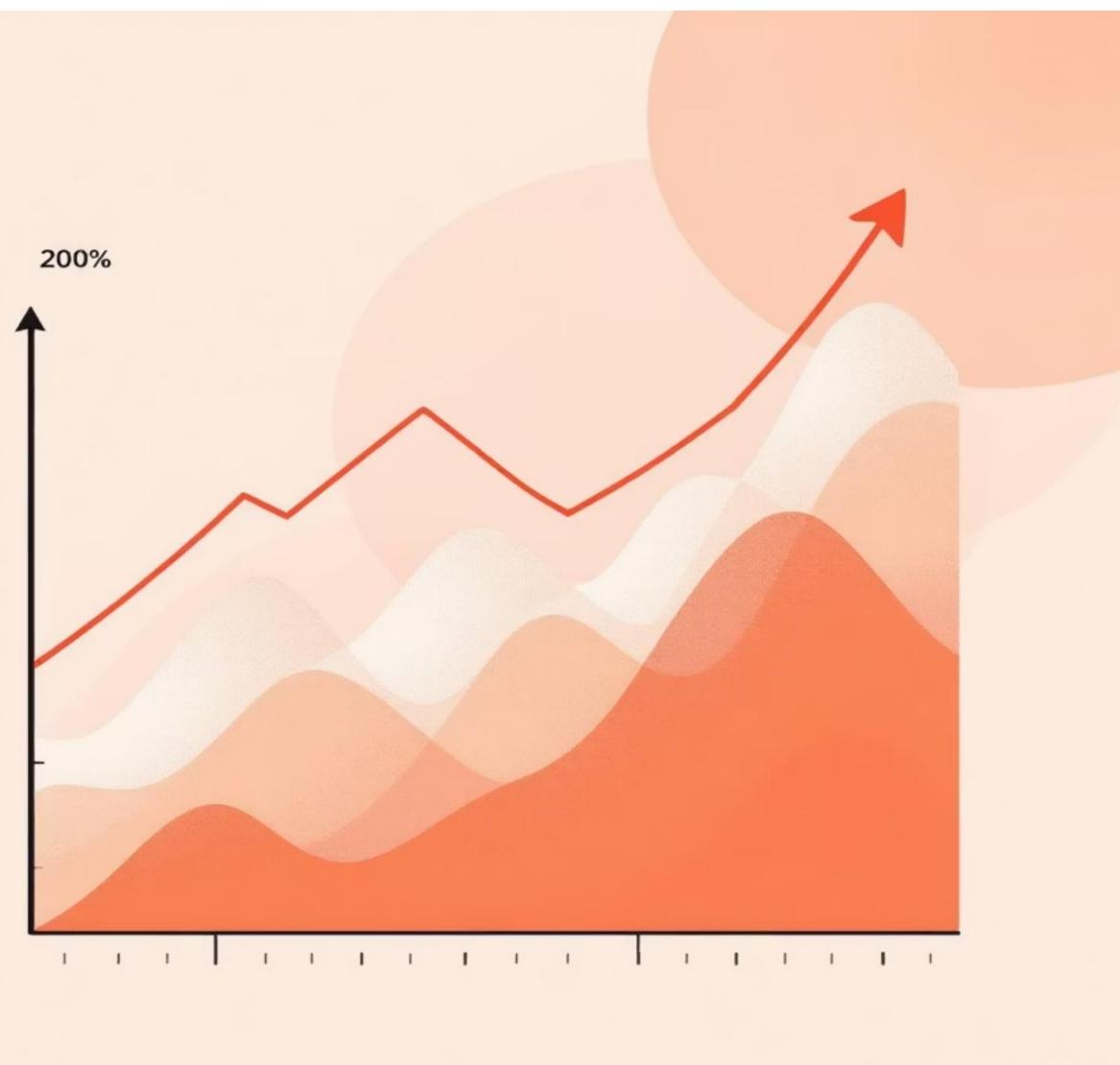
Average Profit Margin

Total Cost per Category



## Sales Analysis and Performance

Monitor growth trends and customer value to guide strategic decisions



### Sales Growth

Track period-over-period expansion and market momentum

### Average Order Value

Measure transaction size and revenue per order

# Shipments Performance

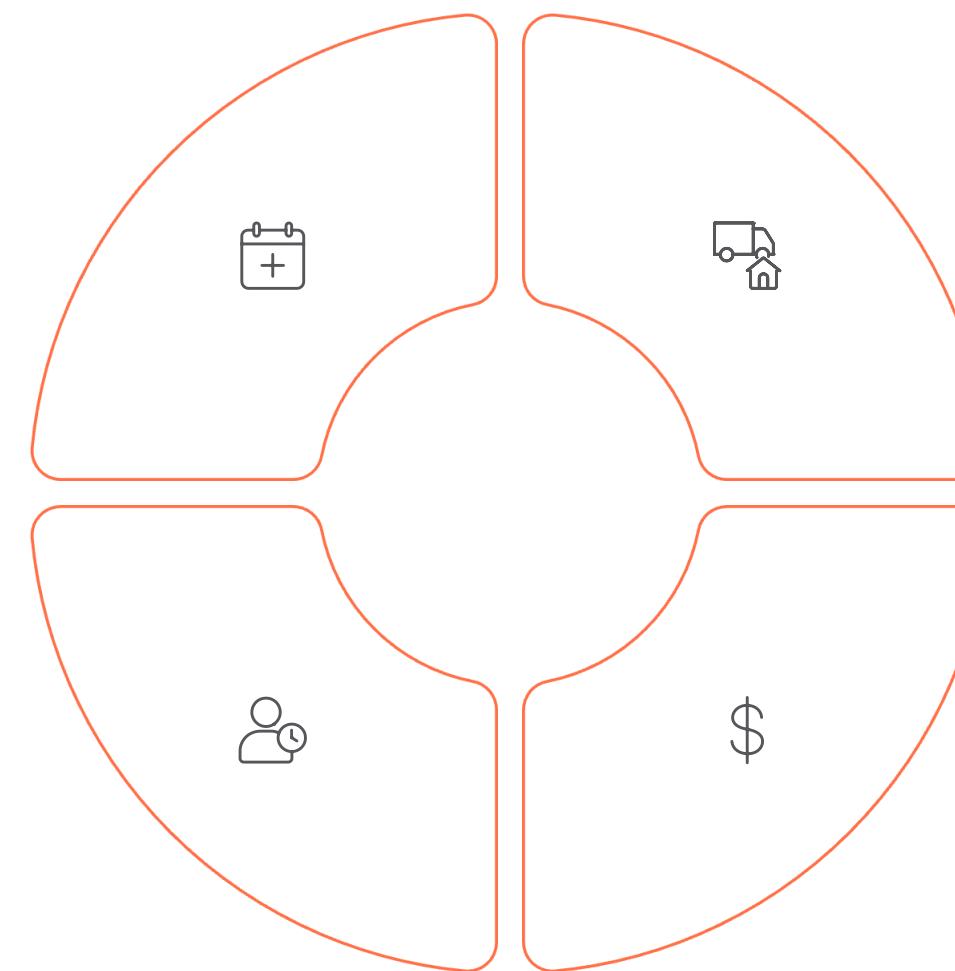
Optimise delivery operations and understand shipping efficiency across all modes

Average Delivery Days

Orders per Ship Mode

Delivery Days per Month

Average Sales per Mode

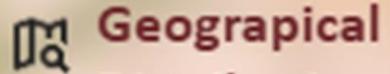


# Superstore Dataset Analysis

Welcome to the Superstore Analysis Project  
This project provides interactive insights across regions, categories, sales performance, and shipments through four main dashboards



Home



Geographical  
Distribution



Product Dashboard



Sales Performance



Shipments

## Team Members

Nada Elsayed

Rana Mohammed

Maha Ali

Mariam Ehab

Esraa Saeed

Tabark Mahmoud

Group Code : YAT525A SHR3 DAT1 S2 DEPI3

# Geographical Distribution Dashboard

Total Sales

\$2.74M

Total Profit

\$676K

Num of Orders

4916

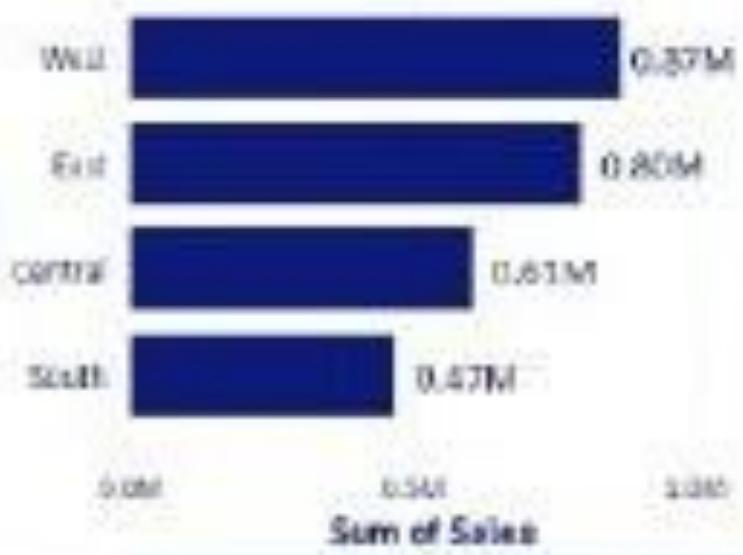
Avg Profit Per Customer

\$852

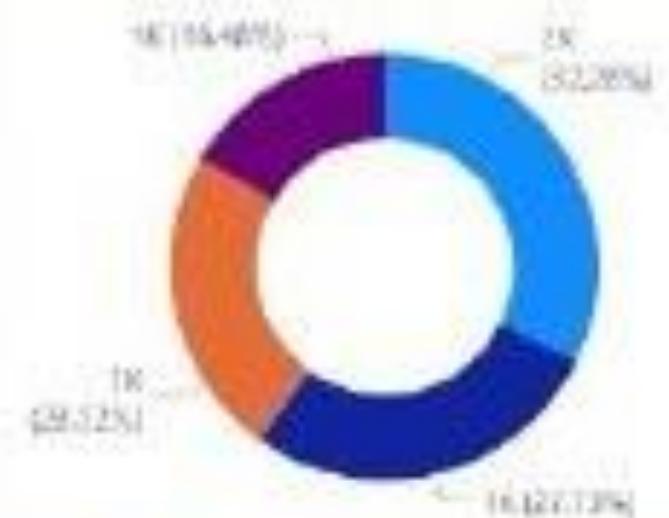
Num of Customers

793

Sum of Sales by Region



# Orders by Region



Region  
West  
East  
Central  
South



2015 2016 2017 2018 2019

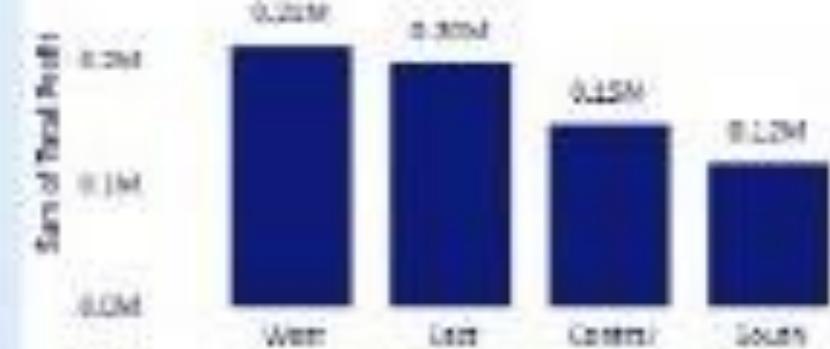
Central

East

South

West

Sum of Total Profit by Region



Sum of Total Profit by Year



# Product Dashboard



Total Sales by Year, Quarter and Month



Total Sales and Total Profit by Sub-Category



Total Profit by Category



2015

2016

2017

2018

2019

Avg profit margin

24.6%



Total Cost

\$1.6M



Total Revenue

\$2.7M



Avg sales per product

\$1.5K



Avg profit per product

\$372



worst product by sales

Eureka  
Disposable Bags...

Total Sales by Category



# Sales Performance

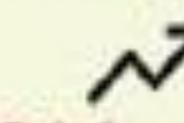
Total Rev.

\$725K



Total Profit

\$180K



Rev. Growth 28.10%

Avg Order Value

\$566



Growth 29.68%

2015

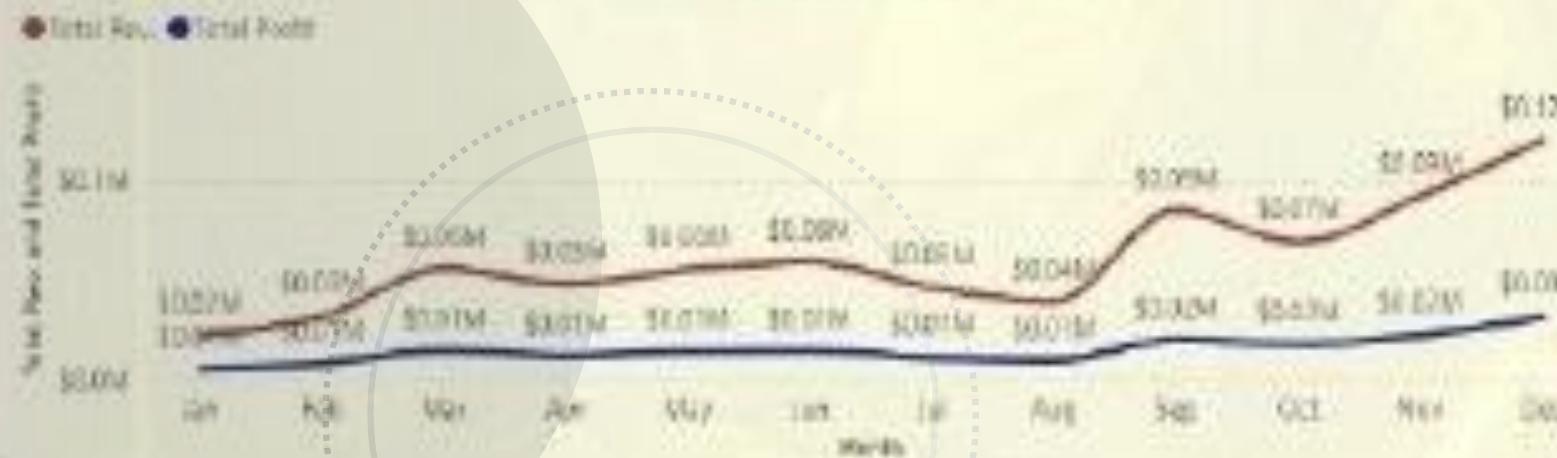
2016

2017

2018

2019

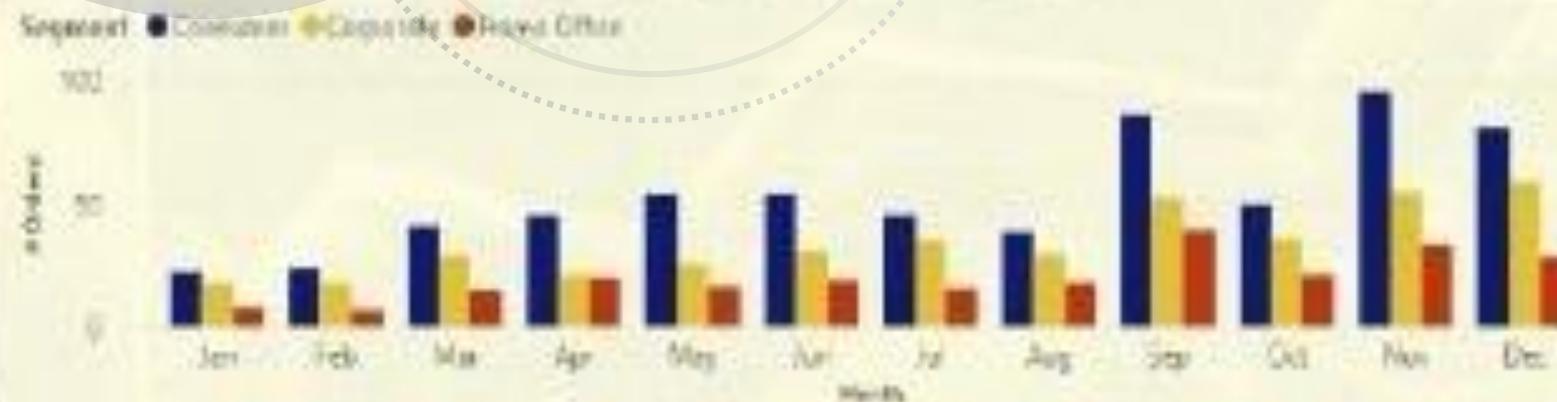
### Total Rev. and Total Profit by Month



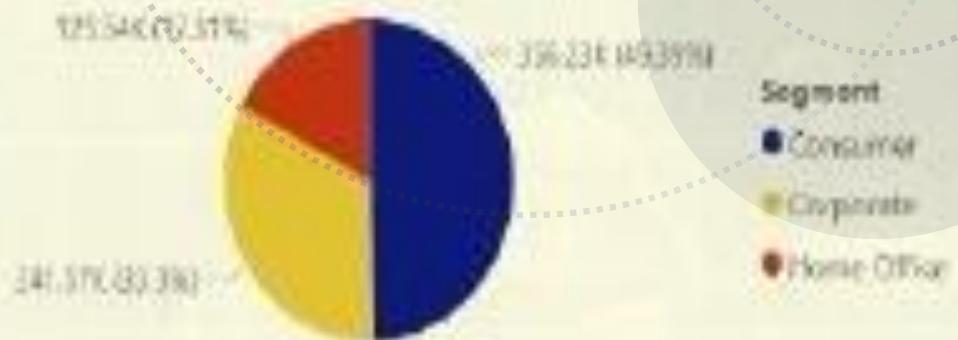
### # Orders and Total Profit by Month



### # Orders by Month and Segment



### Total Sales by Segment



# Shipment Dashboard

Total Shipment

9789

Orders Per Ship Mode

4916

Avg Sales Per Ship Mode

\$558

Avg Delivery Days

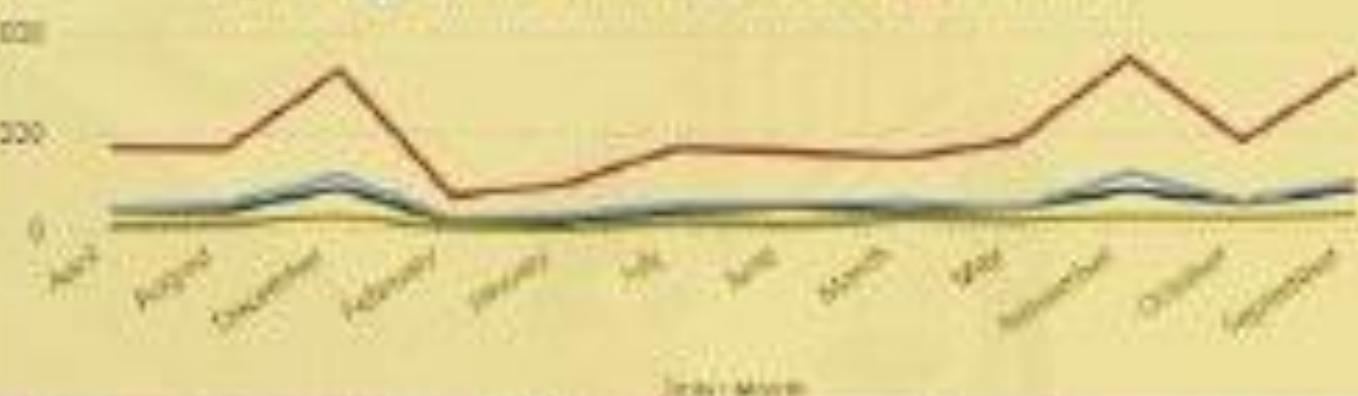
3.96

Year

Select all

## Shipping Trend Over Time

Ship Mode: ● First Class ● Same Day ● Second Class ● Standard Class



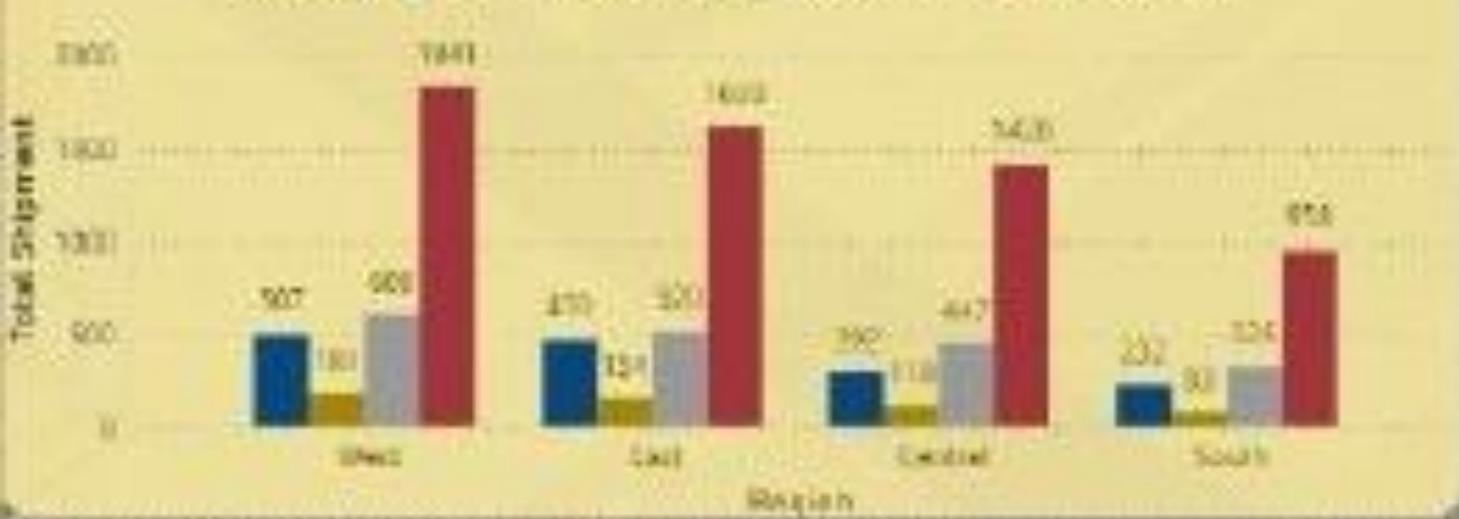
## Ship Mode Distribution

Ship Mode: ● Standard Class ● Second Class ● First Class ● Same Day



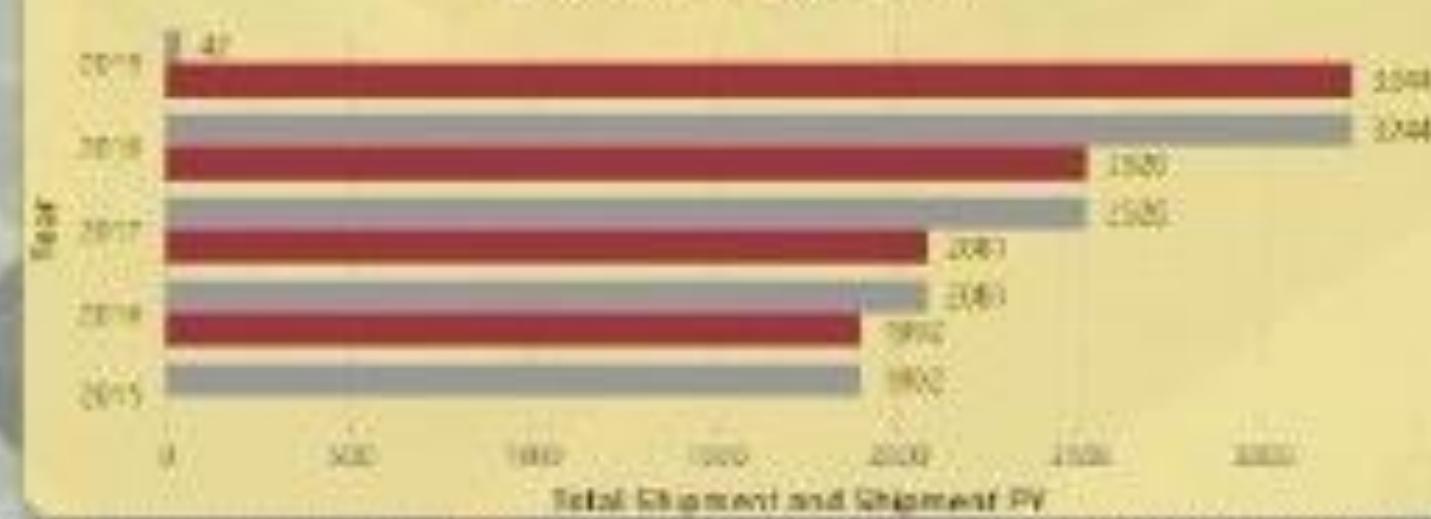
## Shipment Overview By Region & Mode

Ship Mode: ● First Class ● Same Day ● Second Class ● Standard Class



## Yearly Shipments By Ship Mode

● Total Shipment ● Shipment PV





# Thank You

---

We appreciate your attention  
and welcome any questions you  
may have regarding this  
analysis.

