Technical Report: Superstore Sales Trends & Business Drivers

1. Introduction

This technical report presents an analytical overview of **Superstore sales data from 2015 to 2018**.
The objective is to explore **sales trends**, **uncover customer behavior patterns**, and identify business drivers affecting performance.
The analysis focuses on discovering the **most profitable regions**, **top-selling product categories**, and underlying factors influencing sales such as **shipping modes**. The insights extracted will serve as a foundation for further business intelligence and strategic decision-making.

2. Objectives

The primary objectives of this analysis are:

1. To **analyze sales trends between 2015 and 2018** across different time periods.
2. To identify **top-performing product categories and sub-categories**.
3. To evaluate **customer purchasing behavior** and **segment performance**.
4. To determine **high-performing regions** and states based on revenue.

5. To investigate **key factors that influence sales performance** (e.g., discounts, shipping, segment).
6. To prepare the dataset for **dashboarding** and further business intelligence applications.

3. Data Description

The dataset was sourced from an **Excel file** and contains approximately **9,800 rows and 18 columns**. It covers key aspects of the Superstore's operations **from 2015 to 2018**, including **order information**, **sales figures**, **product details**, **customer data**, and **regional segmentation**.

Initially, the dataset was **not clean** and included missing values and formatting inconsistencies; These issues were resolved during the **data cleaning process to ensure data accuracy and enable meaningful analysis**.

4. Data Cleaning

The dataset was initially imported into **SSIS (SQL Server Integration Services)** for data preprocessing.
During the initial review, we identified duplicate records across the dataset, which were **removed to ensure data integrity**.

Next, we discovered **inconsistencies in Product IDs** , where identical or highly similar products had slightly different identifiers. These discrepancies were standardized using **SSIS** transformations to unify the product reference.

We also encountered **missing or incomplete data** in several fields, particularly the **Postal Code**.

To address this, we performed manual verification and **used external sources** to look up and fill in missing postal codes based on city and state information, **ensuring data completeness and consistency**.

5.Data Modeling

After cleaning, we structured the data using a **star schema model**. This involved separating the dataset into **several dimension tables**, such as Date, Product, Customer, and Region, with a central Fact table for sales transactions.
The Date dimension was generated and managed using **Power Query**

Finally, the entire data model was loaded into **Power BI**, where we built interactive visualizations and dashboards to support insightful and data-driven reporting.

6. Exploratory Data Analysis (EDA)

A comprehensive exploratory analysis was conducted using **PowerBI**, supported by **over 20 custom DAX measures**. The aim was to identify trends, patterns, and anomalies across various aspects of the Superstore dataset.

The analysis was structured across the following key areas:

1.Time-Based Analysis

Sales and profit trends were examined across years (2015–2018), months, and quarters

2.Product Performance

Product categories and subcategories were analyzed based on sales volume, sales contribution, and discount impact

3.Customer Behavior

Customers were segmented by type (consumer, business, home office) to assess purchase frequency, average order value, and overall contribution. Metrics such as number of customers

4.Regional Insights

Sales were **analyzed geographically** at the region and state level.
bar charts were used to highlight areas of high performance and underperformance.
A breakdown of shipping methods and their impact by region was also explored.

5.Shipping Impact

Additional analysis focused on the **effect of different shipping** modes on **sales performance and delivery efficiency**. The impact of shipping types across **regions and customer segments** was explored to better understand operational performance.

• Key Measures Used

Purpose	DAX Expression	Measure Name
Calculates Total Sales	`SUM('Fact Sales'[Sales])`	Total Sales
Calculates Units Sold	`COUNTROWS('Fact Sales')`	Total Units Sold

Calculates the Average Sales per Order	`DIVIDE(SUM('FactSales'[Sales]), DISTINCTCOUNT('FactSales'[Order_ID]),0)`	Average Order Value
Calculates the Average Shipping Time	`AVERAGE('Fact Sales'[Shipping time])`	Average Shipping Time
Calculates the Average Sales per State	`AVERAGEX(VALUES('Dim Location'[State]), CALCULATE(SUM('Fact Sales'[Sales])))`	Avg Sales per State
Calculates the Average Quantity per State	`CALCULATE(COUNTROWS('Fact Sales'), ALLEXCEPT('DimLocation','DimLocation'[State]))`	Avg Quantity per State
Calculates the Average Sales per Customer	`DIVIDE(DISTINCTCOUNT('Fact Sales'[Order_ID]), DISTINCTCOUNT('Fact Sales'[Customer_ID]))`	Average Orders Per Customer
Calculates number customer in 2018	`CALCULATE(DISTINCTCOUNT('DimCustomer'[Customer_ID]),YEAR('DimCustomer'[LastInteractionDate]) = 2018)`	CustomersIn2018
Measures the retention rate of 2018 customers	`DIVIDE([CustomersIn2018],[NumCustomers],0)`	Retention rate after 2018
Calculates Number of Orders	`DISTINCTCOUNT('Fact Sales'[order_ID])`	Num Orders
Calculates Number of Products	`DISTINCTCOUNT('Dim Product'[Product_ID])`	Num Products
Calculates Number of Customers	`DISTINCTCOUNT('Fact Sales'[Customer_ID])`	Num Customers
Calculate Number of States	`DISTINCTCOUNT('Dim Location'[State])`	Number of States
Calculate Number of Customers by Recency	`CALCULATE(DISTINCTCOUNT('DimCustomer'[Customer_ID]),ALLEXCEPT('Dim Customer','Dim Customer'[Recency Bucket]))`	Number of Customers by Recency
Calculate Frequency	`COUNT( 'Fact Sales'[Order_ID])`	Frequency
Calculate Monetary	`CALCULATE(SUM('Fact Sales'[Sales]), ALLEXCEPT('Fact Sales', 'Fact Sales'[Customer_ID]))`	Monetary

Measures how recently a customer interacted, in months, before Jan 2019	`DATEDIFF(MAX('Dim Customer'[LastInteractionDate]),DATE(2019, 1, 1), MONTH)`	Recency (month)
Retrieves the most recent interaction date for each customer	`MAX('DimCustomer'[LastInteractionDate])`	LastInteractionDate

• Calculated Columns Overview

Purpose	DAX Expression	Measure Name
Calculates Shipping time	`DATEDIFF('FactSales'[Order_Date],'FactSales'[Ship_Date], DAY)`	Shipping time
Calculates LastInteractionDate	`CALCULATE(MAX('Fact Sales'[Order_Date]), ALLEXCEPT('Dim Customer','DimCustomer'[Customer_ID]))`	LastInteractionDate
Calculates Customer Recency	`DATEDIFF('Dim Customer'[LastInteractionDate], DATE(2019, 1, 1), DAY)`	Customer Recency

- Calculates **`Recency Bucket`**

`Recency Bucket=`

```
SWITCH(
    TRUE(),
    'Dim Customer'[Customer Recency] <= 60, "0-60 Days",     'Dim Customer'[Customer Recency]<= 120, "61-120 Days",     'Dim Customer'[Customer Recency] <= 180, "121-180 Days",        'Dim Customer'[Customer Recency] <= 365, "181-365 Days",     "More than 365 Days")
```

- Calculates SpecialSalesDay

`Special_Sales_Day =`

```
VAR YearOfOrder = YEAR('Dim Date'[Date])
```

```
VAR Thanksgiving = SWITCH(YearOfOrder,2015, DATE(2015,11,26),2016,
DATE(2016,11,24),2017, DATE(2017,11,23),2018, DATE(2018,11,22))
VAR BlackFriday = Thanksgiving + 1
VAR CyberMonday = BlackFriday + 3 VAR EasterSunday =
SWITCH(YearOfOrder,2015, DATE(2015,4,5),2016,
DATE(2016,3,27),
2017, DATE(2017,4,16),2018, DATE(2018,4,1))
VAR NewYearsEve = DATE(YearOfOrder,12,31) VAR
ValentinesDay = DATE(YearOfOrder,2,14)
VAR MothersDay =
DATE(YearOfOrder,5,IF(YearOfOrder=2015,10,IF(YearOfOrder=2016,8,IF(YearOfOrder=
2017,14,13))))
VAR FathersDay =
DATE(YearOfOrder,6,IF(YearOfOrder=2015,21,IF(YearOfOrder=2016,19,IF(YearOfOrder
=2017,18,17))))
VAR BackToSchool = DATE(YearOfOrder,8,1)  VAR
Halloween = DATE(YearOfOrder,10,31)
VAR ChristmasEve = DATE(YearOfOrder,12,24) VAR
ChristmasDay = DATE(YearOfOrder,12,25)
RETURN
SWITCH(TRUE(),'Dim Date'[Date]  = NewYearsEve, "New Year's Eve",'Dim
Date'[Date] ValentinesDay, "Valentine's Day",'Dim Date'[Date] = EasterSunday,
"Easter Sunday",'Dim Date'[Date] = MothersDay, "Mother's Day",'Dim Date'[Date]
= FathersDay, "Father's Day",'Dim Date'[Date] = BackToSchool, "Back to
School",'Dim Date'[Date] = Halloween, "Halloween",'Dim Date'[Date] =
Thanksgiving, "Thanksgiving",'Dim Date'[Date] = BlackFriday, "Black
Friday",'Dim Date'[Date] = CyberMonday, "Cyber Monday",'DimDate'[Date] =
ChristmasEve, "Christmas Eve",'Dim Date'[Date] = ChristmasDay, "Christmas
Day","Regular Day")
```

7. Tools & Technologies Used

The following tools and technologies were used throughout the data analysis project :

Microsoft Excel : Used as the initial data source. The raw dataset was structured in Excel format (.xlsx) before being imported into the ETL process.

SQL Server Integration Services (SSIS) : Utilized for the ETL process, including importing the Excel file, cleaning the data, removing duplicates, and standardizing fields such as Product IDs.

Power Query : Applied for data transformation, generating Date dimension tables, filtering rows, and shaping the data for analysis within Power BI.

Power BI Desktop : Used for data modeling, building relationships using a star schema (Fact and Dimension tables), and creating interactive visualizations and dashboards.

DAX (Data Analysis Expressions) : Used to create calculated columns and measures for key metrics such as Recency, Frequency, Monetary (RFM), Retention Rate, and Average Order Value.

Git & GitHub : Version control tools used to manage project files, track changes, and document progress throughout the development of the report.