



# Super Store Sales Analysis Project

**Prepared for**  
**Dr. Yasser A. Rahman**

**By:**

Mariam Ragab Khedr Abdelaleam  
Bernadette Sameh Saman Yousef  
Mohamed Abdelwahab Mahmoud  
Yasser Elsayed Mohamed Seleem  
Yara Farid Mohamed Fathy Shahin  
Maged Rizk Eskander

# Table of Contents

- Executive Summary.
- Introduction.
- Project Objectives.
- Dataset Description.
- Methodology.
  - Data Cleaning (Excel).
  - Data Querying (SQL).
  - Visualization (Power BI).
- Business Questions and Analysis.
- Key Findings.
- Conclusion and Recommendations.
- Appendix.

# Executive Summary

This report provides a comprehensive analysis of SuperStore's sales performance, customer behavior, and operational efficiency. Leveraging SQL for data querying and Power BI for visualization, we identified key trends, regional disparities, and product performance.

## Introduction

In a competitive retail landscape, data-driven decision-making is critical. This report analyzes SuperStore's transactional data (2015-2018) to uncover trends in sales, customer segments, and shipping efficiency. The findings aim to guide strategic improvements in inventory, marketing, and operations.

## Project Objectives

1. Analyze sales trends across categories, regions, and time.
2. Identify top-performing products and loss-making subcategories.
3. Evaluate customer segment contributions.
4. Assess regional performance and shipping efficiency.

## Dataset Description

Three primary datasets were used:

1. **Orders:** Sales transactions, and shipping details.
2. **Customers:** Segment classifications (Consumer, Corporate, Home Office).
3. **Products:** Category and subcategory hierarchies.




## Data Preparation:

- Cleaned in Excel




Source	✳
Promoted Headers	✳
Replaced Value	✳
Changed Type with Locale	✳
Changed Type with Locale1	✳
Changed Type	
Split Column by Delimiter	✳
Changed Type1	
Inserted Merged Column	✳
Reordered Columns	
Inserted Merged Column1	✳
Reordered Columns1	
Removed Columns	
Renamed Columns	
Changed Type2	
Changed Type with Locale2	✳
Removed Duplicates	
Replaced Value1	✳

Functions for each cleaning step




1. promote headers




```
   = Table.PromoteHeaders(Source, [PromoteAllScalars=true])
```




2. replace null value in postal code column

```
   = Table.ReplaceValue("#Promoted Headers",null,5401,Replacer.ReplaceValue,{"Postal Code"})
```


3. change data type

```
   = Table.TransformColumnTypes("#Replaced Value", {"Order Date", type date}, "en-GB")
```

```
   = Table.TransformColumnTypes("#Changed Type with Locale", {"Ship Date", type date}, "en-GB")
```

```
   = Table.TransformColumnTypes("#Changed Type with Locale1", {"Sales", Currency.Type})
```

4. split column

```
 = Table.SplitColumn(Table.TransformColumnTypes("#Changed Type", {"Order Date", type text}, "en-US"), "Order Date", Splitter.SplitTextByDelimiter("/", QuoteStyle.Csv), {"Order Date.1", "Order Date.2", "Order Date.3"})
```

## 5. change data type

```
fX = Table.TransformColumnTypes("#Split Column by Delimiter",{{"Row ID", Int64.Type}, {"Order ID", type text},
```

## 6. Inserted merged column

```
fX = Table.AddColumn("#Changed Type1", "Merged", each Text.Combine({Text.From([Order Date.3], "en-US"), Text.From([Order Date.1], "en-US")}, "/"), type text)
```

```
fX = Table.AddColumn("#Reordered Columns", "Merged.1", each Text.Combine({[Merged], Text.From([Order Date.2], "en-US")}, "/"), type text)
```

## 7. Remove split column

```
fX = Table.RemoveColumns("#Reordered Columns1",{"Order Date.1", "Order Date.2", "Order Date.3", "Merged"})
```

## 8. Remove duplicated rows

```
= Table.Distinct("#Changed Type with Locale2", {"Row ID"})
```

## 9. Replace values

```
= Table.ReplaceValue("#Removed Duplicates","", ".", Replacer.ReplaceText,{"Product Name"})
```

```
fX = Table.ReplaceValue("#Replaced Value1","", ".", Replacer.ReplaceText,{"Product Name"})
```

- Data preparation using data model in Power BI

## 1. orders table

Table.TransformColumnTypes("#Changed Type",{"Sales", Currency.Type})

	Order_ID	Order_Date	Ship_Date	Ship_Mode	Customer_ID	Product_ID
1	CA-2017-152156	11/8/2017	11/11/2017	Second Class	CG-12520	FUR-BO-1000
2	CA-2017-152156	11/8/2017	11/11/2017	Second Class	CG-12520	FUR-CH-1000
3	CA-2017-138688	6/12/2017	6/16/2017	Second Class	DV-13045	OFF-LA-1000
4	US-2016-108966	10/11/2016	10/18/2016	Standard Class	SO-20335	FUR-TA-1000
5	US-2016-108966	10/11/2016	10/18/2016	Standard Class	SO-20335	OFF-ST-1000
6	CA-2015-115812	6/9/2015	6/14/2015	Standard Class	BH-11710	FUR-FU-1000
7	CA-2015-115812	6/9/2015	6/14/2015	Standard Class	BH-11710	OFF-AR-1000
8	CA-2015-115812	6/9/2015	6/14/2015	Standard Class	BH-11710	TEC-PH-1000
9	CA-2015-115812	6/9/2015	6/14/2015	Standard Class	BH-11710	OFF-BI-1000

Query Settings

PROPERTIES

Name

fOrders

All Properties

APPLIED STEPS

Source

Navigation

Changed Type

Changed Type1

= Table.TransformColumnTypes(fOrders\_Table,({"Order\_ID", type text}, {"Order\_Date", type date}, {"Ship\_Date", type date}, {"Ship\_Mode", type text}, {"Customer\_ID", type text}, {"Product\_ID", type text}, {"Sales", type number}))

	Order_ID	Order_Date	Ship_Date	Ship_Mode	Customer_ID	Product_ID
1	CA-2017-152156	11/8/2017	11/11/2017	Second Class	CG-12520	FUR-BO-1000
2	CA-2017-152156	11/8/2017	11/11/2017	Second Class	CG-12520	FUR-CH-1000
3	CA-2017-138688	6/12/2017	6/16/2017	Second Class	DV-13045	OFF-LA-1000
4	US-2016-108966	10/11/2016	10/18/2016	Standard Class	SO-20335	FUR-TA-1000
5	US-2016-108966	10/11/2016	10/18/2016	Standard Class	SO-20335	OFF-ST-1000
6	CA-2015-115812	6/9/2015	6/14/2015	Standard Class	BH-11710	FUR-FU-1000

Query Settings

PROPERTIES

Name

fOrders

All Properties

APPLIED STEPS

Source

Navigation

Changed Type

Changed Type1

## 2. customers table

= Table.TransformColumnTypes(dCustomers\_Locations\_Table,({"Customer\_ID", type text}, {"Customer\_Name", type text}, {"Segment", type text}, {"Country", type text}, {"City", type text}, {"State", type text}, {"Region", type text}, {"Postal\_Code", Int64.Type}))

	Customer_ID	Customer_Name	Segment	Country	City	State
1	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky
2	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California
3	SO-20335	Sean O.Donnell	Consumer	United States	Fort Lauderdale	Florida
4	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California
5	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina
6	IM-15070	Irene Maddox	Consumer	United States	Seattle	Washington

Query Settings

PROPERTIES

Name

dCustomers\_Locations

All Properties

APPLIED STEPS

Source

Navigation

Changed Type

Changed Type1

Table.TransformColumnTypes("#Changed Type",{"Postal\_Code", type text})

	Customer_ID	Customer_Name	Segment	Country	City	State
1	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky
2	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles	California
3	SO-20335	Sean O.Donnell	Consumer	United States	Fort Lauderdale	Florida
4	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles	California
5	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolina
6	IM-15070	Irene Maddox	Consumer	United States	Seattle	Washington
7	HP-14815	Harold Pawlan	Home Office	United States	Fort Worth	Texas
8	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin
9	AG-10270	Alejandro Grove	Consumer	United States	West Jordan	Utah

Query Settings

PROPERTIES

Name

dCustomers\_Locations

All Properties

APPLIED STEPS

Source

Navigation

Changed Type

Changed Type1

### 3.DimDate table

table source

Query1(#date(2015, 1, 1), #date(2018, 12, 31))

Table.ReorderColumns(Source,{"Date", "Year", "Quarter", "Month Number", "Month", "Week Number", "Day of Week"})

ABC 123	Date	ABC 123	Year	ABC 123	Quarter	ABC 123	Month Number	ABC 123	Month	ABC 123	Week Numl
1	1/1/2015		2015	Q1			1	January			
2	1/2/2015		2015	Q1			1	January			
3	1/3/2015		2015	Q1			1	January			
4	1/4/2015		2015	Q1			1	January			
5	1/5/2015		2015	Q1			1	January			
6	1/6/2015		2015	Q1			1	January			
7	1/7/2015		2015	Q1			1	January			
8	1/8/2015		2015	Q1			1	January			

Query Settings

PROPERTIES

Name

DimDate

APPLIED STEPS

Source

Reordered Columns

Changed Type

Table.TransformColumnTypes("#Reordered Columns",{{"Date", type date}, {"Year", Int64.Type}, {"Quarter", type text}, {"Month Number", Int64.Type}, {"Month", type text}, {"Week Number", Int64.Type}, {"Day of Week", type text}})

ABC 123	Date	123	Year	ABC 123	Quarter	123	Month Number	ABC 123	Month	123	Week Numl
1	1/1/2015		2015	Q1			1	January			
2	1/2/2015		2015	Q1			1	January			
3	1/3/2015		2015	Q1			1	January			
4	1/4/2015		2015	Q1			1	January			

Query Settings

PROPERTIES

Name

DimDate

APPLIED STEPS

Source

Reordered Columns

Changed Type

### 4.Product table

Table.TransformColumnTypes(Custom1,{{"Category", type text}, {"Product\_ID", type text}, {"Sub\_Category", type text}, {"Product\_Name", type text}})

ABC 123	Product_ID	ABC 123	Category	ABC 123	Sub_Category	ABC 123	Product_Name
1	FUR-BO-10001798		Furniture		Bookcases		Bush Somerset Collection Bookcase
2	FUR-CH-10000454		Furniture		Chairs		Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back
3	OFF-LA-10000240		Office Supplies		Labels		Self-Adhesive Address Labels for Typewriters by Universal
4	FUR-TA-10000577		Furniture		Tables		Bretford CR4500 Series Slim Rectangular Table
5	OFF-ST-10000760		Office Supplies		Storage		Eldon Fold .N Roll Cart System

Query Settings

PROPERTIES

Name

dProduct

APPLIED STEPS

Source

Navigation

Changed Type

Removed Duplicates2

Table.Distinct("#Changed Type", {"Product\_ID"})

ABC 123	Product_ID	ABC 123	Category	ABC 123	Sub_Category	ABC 123	Product_Name
1	FUR-BO-10001798		Furniture		Bookcases		Bush Somerset Collection Bookcase
2	FUR-CH-10000454		Furniture		Chairs		Hon Deluxe Fabric Upholstered Stacking Chairs, Rounded Back
3	OFF-LA-10000240		Office Supplies		Labels		Self-Adhesive Address Labels for Typewriters by Universal
4	FUR-TA-10000577		Furniture		Tables		Bretford CR4500 Series Slim Rectangular Table
5	OFF-ST-10000760		Office Supplies		Storage		Eldon Fold .N Roll Cart System
6	FUR-FU-10001487		Furniture		Furnishings		Eldon Expressions Wood and Plastic Desk Accessories, Cherry Wood
7	OFF-AR-10002833		Office Supplies		Art		Newell 322
8	TEC-PH-10002275		Technology		Phones		Mitel 5320 IP Phone VoIP phone
9	OFF-BI-10003910		Office Supplies		Binders		DXL Angle-View Binders with Locking Rings by Samsill
10	OFF-AP-10002882		Office Supplies		Appliances		Belkin F5C706VTE1 6 Outlet Surge

Query Settings

PROPERTIES

Name

dProduct

APPLIED STEPS

Source

Navigation

Changed Type

Removed Duplicates2

- DAXs using Power BI

Name	Expression
Late_Deliveries%	DIVIDE( [LateOrders], [Total Orders], 0 )
Product Contribution %	DIVIDE( [Total Sales], CALCULATE( [Total Sales], ALL(fOrders) ) )
Avg Shipping Time	FORMAT( AVERAGEX( fOrders, DATEDIFF(fOrders[order_Date], fOrders[Ship_Date], DAY) ), "0" ) & " Days"
LateOrders	COUNTROWS ( FILTER ( fOrders, fOrders[Ship_Date] > fOrders[Expected_Delivery_Date] ) )
Total Customers	
% Weekday_Orders	DIVIDE( CALCULATE(COUNT(fOrders[Order_ID]), DimDate[Day Type] = "Weekday"), COUNT(fOrders[Order_ID]) )
% Weekend_Orders	DIVIDE( CALCULATE(COUNT(fOrders[Order_ID]), DimDate[Day Type] = "Weekend"), COUNT(fOrders[Order_ID]) )
On Time Delivery Rate	DIVIDE( COUNTROWS(FILTER(fOrders, fOrders[Delivery Days] <= fOrders[Expected Delivery Days])), COUNTROWS(fOrders), 0 )



Name	Expression
Total_Orders_By_Shipping_Mode	COUNT(fOrders[Order_ID])
Total Customers	DISTINCTCOUNT ( fOrders[Customer_ID] )
Total Orders	COUNTA(fOrders[Order_ID])
Total Sales	SUM(fOrders[Sales])

Name	Expression
Compair customers	IF( ISBLANK(SELECTEDVALUE(DimDate[Year])), "Filter a year to see", IF( SELECTEDVALUE(DimDate[Year]) = 2015, [Total Customers], IF( ISFILTERED(DimDate), FORMAT([var_customers], "#,0") & "   " & FORMAT([Var_customers%], "#0.0%"), [Total Customers] ) ))

Name	Expression
Compair Orders	IF( ISBLANK(SELECTEDVALUE(DimDate[Year])), "Filter a year to see", IF( SELECTEDVALUE(DimDate[Year]) = 2015, [Total Orders], IF( ISFILTERED(DimDate), FORMAT([var_Orders], "#,0") & "   " & FORMAT([Var_Orders%], "#0.0%"), [Total Orders] ) ) )

Name	Expression
Compair sales	IF( ISBLANK(SELECTEDVALUE(DimDate[Year])), "Filter a year to see", IF( SELECTEDVALUE(DimDate[Year]) = 2015, [Total Sales], IF( ISFILTERED(DimDate), FORMAT([var_Sales], "\$#,0") & "   " & FORMAT([Var_Sales%], "#0.0%"), [Total Sales] ) ) )
Total Sales	
var_customers	[Total Customers]-[Ly_customers]
var_Orders	[Total Orders]-[Ly_orders]
var_Sales	[Total Sales]-[Ly_Sales]
Ly_customers	CALCULATE([Total Customers],SAMEPERIODLASTYEAR(DimDate[Date]))
Ly_orders	CALCULATE([Total Orders],SAMEPERIODLASTYEAR(DimDate[Date]))
Ly_Sales	CALCULATE([Total Sales],SAMEPERIODLASTYEAR(DimDate[Date]))
Var_customers%	DIVIDE([var_customers],[Ly_customers])
Var_Orders%	DIVIDE([var_Orders],[Ly_orders])

Name	Expression
Arrow Sales	IF(ISFILTERED(DimDate[Year]), IF([Var_Sales%]>0,UNICHAR(9650),UNICHAR(9660)),"")
Arrow Orders	IF(ISFILTERED(DimDate[Year]), IF([Var_Orders%]>0,UNICHAR(9650),UNICHAR(9660)),"")
Arrow cusromers	IF(ISFILTERED(DimDate[Year]), IF([Var_customers%]>0,UNICHAR(9650),UNICHAR(9660)),"")
Daynamic color for Sales	IF([Var_Sales%]>0, 1,2)
Daynamic color for orders	IF([Var_Orders%]>0, 1,2)
Daynamic color for customers	IF([Var_customers%]>0, 1,2)
Sales subtitle	FORMAT([Total Sales],"\$#,0")& ""
avg_shiptime_subtitle	FORMAT([Total Orders],"#,00")& ""
Orders_subtitle	FORMAT([Total Orders],"#,00")& ""
customers_subtitle	FORMAT([Total Customers],"#,00")& ""
Var_Sales%	DIVIDE([var_Sales],[Ly_Sales])

## Relations in tables

Name	Relationship
0f94bedc-0a5e-4cf9-afc5-29343a4c9aa3	'fOrders'[Expected_Delivery_Date] *[-]1 'LocalDateTable_33043012-3650-4904-a864-daf5b73bfb8d'[Date]
6158be7f-fd6b-4fd0-a269-ce787096265a	'DimDate'[Date] *[-]1 'LocalDateTable_2428c80e-f3f6-42e9-8802-dcad44022ef2'[Date]
a09d9ef9-b9f5-4289-8503-0029f099669f	'fOrders'[Ship_Date] *[-]1 'LocalDateTable_f41685eb-cf52-48f0-9102-6f6e6cbcb3c4'[Date]
AutoDetected_b5ae7fab-cf89-4290-8a39-f9d5069af08c	'fOrders'[Customer_ID] *[-]1 'dCustomers_Locations'[Customer_ID]
bc957cca-9b79-1220-2c71-2a6c40788c82	'fOrders'[Product_ID] *[-]1 'dProduct'[Product ID]
d0f7d0d3-44b6-2fdb-026e-1fa9027e68f5	'fOrders'[Order_Date] *[-]1 'DimDate'[Date]

## Methodology

Our structured approach included:

- **Excel:** Data cleaning and validation.
- **SQL:** Writing targeted queries to answer specific business questions.
- **Power BI:** Building interactive dashboards and forecasting models to visualize patterns and project future trends.

With these tools and methods in place, we addressed the business's core questions as follows:

## Business Questions and Analysis

### 1. Sales Performance:

#### 1.1 What are the total sales trends over time?

SuperStore's annual sales show consistent growth from **\$479,856** in 2015 to **\$722,062** in 2018, with a temporary dip in 2016. This **50.5%** overall growth indicates strong market performance, though the 2016 decline suggests potential operational challenges worth investigating to prevent future downturns.

## SQL Query

```
SELECT  
EXTRACT(YEAR FROM Order_Date) AS year, SUM(sales) AS total_sales  
FROM orders  
GROUP BY year ORDER BY year;
```

## Query Results

#	year	total_sales
1	2015	479856.208100001
2	2016	459436.005400001
3	2017	600192.550000001
4	2018	722052.019200001

### 1.2 Which regions generate the most revenue?

The West region dominates with **\$744,294** in sales (32.4%), followed by the East (**\$606,351**), Central (**\$514,251**), and South (**\$396,641**). The significant underperformance in the South region presents a clear opportunity for targeted improvements in marketing and distribution.

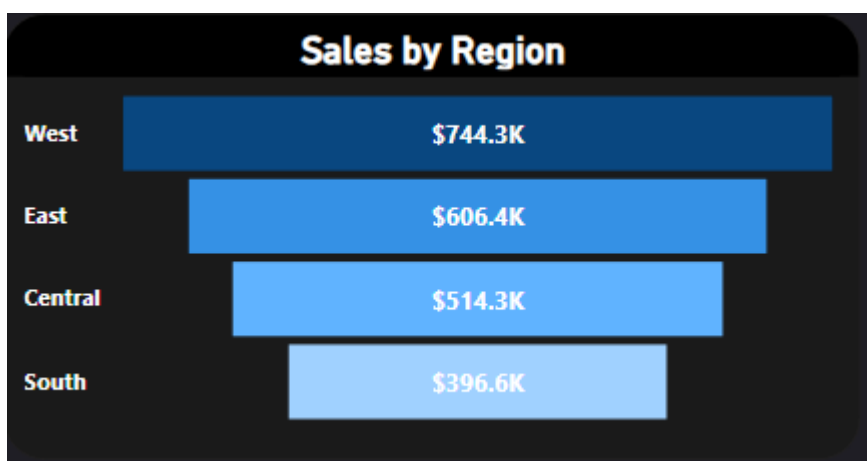
## SQL Query

```
SELECT Region, SUM(sales) AS total_sales  
FROM orders o  
JOIN customers c ON o.Customer_ID = c.Customer_ID  
GROUP BY Region  
ORDER BY total_sales DESC
```

## Query Results

#	region	total_sales
1	West	744293.5516999995
2	East	606351.137500001
3	Central	514251.4722000008
4	South	396640.62130000023

## Sales by region chart



### 1.3 What are the best-selling product categories?

Technology (particularly Phones at \$327,782) and Furniture (Chairs at \$322,823) are the top performers. Office Supplies like Storage (\$219,143) and Binders (\$199,762) also contribute significantly. These categories should receive priority in inventory planning and promotional activities.

#### SQL Query

```
SELECT p.Category, p.Sub_Category, SUM(o.sales) AS
total_sales
FROM orders o JOIN Products p
ON o.Product_ID = p.Product_ID
GROUP BY p.Category, p.Sub_Category
ORDER BY total_sales DESC
limit 5
```

#### Query Results

#	category	sub_category	total_sales
1	Technology	Phones	327782.44800000027
2	Furniture	Chairs	322822.7310000008
3	Office Supplies	Storage	219142.66000000012
4	Furniture	Tables	200132.35100000014
5	Office Supplies	Binders	199761.56699999998

## The best-selling product categories chart

Category	Total Sales ▼
Technology	\$827K
Furniture	\$729K
Office Supplies	\$705K

### 1.4How do monthly sales trends vary?

Sales exhibit strong seasonality, peaking in November (\$117,938) and December (\$83,030) due to holiday shopping. January typically shows the lowest numbers, suggesting the need for post-holiday recovery strategies like clearance sales or new product launches.

#### SQL Query

```
SELECT EXTRACT(YEAR FROM Order_Date) AS year,  
TO_CHAR(Order_Date, 'Month') AS month_name, SUM(sales)  
AS total_sales  
FROM orders  
GROUP BY year, month_name  
ORDER BY year, MIN(Order_Date);
```

## Query Results

#	year	month_name	total_sales
1	2015	January	14205.706999999997
2	2015	February	4519.892
3	2015	March	55205.79700000003
4	2015	April	27906.854999999992
5	2015	May	23644.303
6	2015	June	34322.93560000002
7	2015	July	33781.543
8	2015	August	27117.536499999995
9	2015	September	81623.52679999998
10	2015	October	31453.39299999999
11	2015	November	77907.66070000001
12	2015	December	68167.0585
13	2016	January	18066.957599999994
14	2016	February	11951.411

#	year	month_name	total_sales
15	2016	March	32339.31839999999
16	2016	April	34154.468499999995
17	2016	May	29959.530499999997
18	2016	June	23599.374
19	2016	July	28608.25899999999
20	2016	August	36818.34219999999
21	2016	September	63133.606000000036
22	2016	October	31011.737500000014
23	2016	November	75249.39950000004
24	2016	December	74543.60120000008
25	2017	January	18542.490999999995
26	2017	February	22978.815000000002
27	2017	March	51165.05900000002
28	2017	April	38679.766999999998



#	year	month_name	total_sales
29	2017	May	56656.90799999999
30	2017	June	39724.48599999998
31	2017	July	38320.78299999997
32	2017	August	30542.2003
33	2017	September	69193.39089999997
34	2017	October	59583.03300000001
35	2017	November	79066.4958
36	2017	December	95739.12100000001
37	2018	January	43476.47399999999
38	2018	February	19920.997399999997
39	2018	March	58863.412799999984
40	2018	April	35541.91010000001
41	2018	May	43825.98219999999
42	2018	June	48190.7277
43	2018	July	44825.10400000001
44	2018	August	62837.84799999998
45	2018	September	86152.88800000004
46	2018	October	77448.13119999997
47	2018	November	117938.15500000001
48	2018	December	83030.3888

## Sales trend chart



## 1.5 What are the top-selling individual products?

The Canon imageCLASS copier leads with **\$83,200** in sales, followed by Fellowes binding machines (**\$57,450**) and Cisco teleconferencing systems (**\$22,595**). These high-value office equipment items represent key revenue drivers that could benefit from accessory bundling strategies.

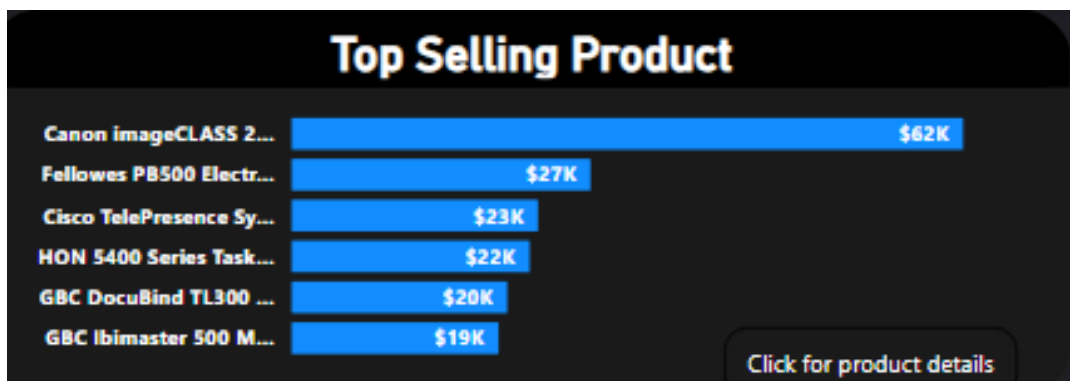
### SQL Query

```
SELECT p.Product_ID, p.Product_Name, SUM(o.sales) AS
total_sales
FROM orders o
JOIN Products p
ON o.Product_ID = p.Product_ID
GROUP BY p.Product_ID, p.Product_Name
ORDER BY total_sales DESC
LIMIT 10;
```

### Query Results

#	product_id	product_name	total_sales
1	TEC-CO-10004722	Canon imageCLASS 2200 Advanced Copier	61599.824
2	OFF-BI-10003527	Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind	27453.384
3	TEC-MA-10002412	Cisco TelePresence System EX90 Videoconferencing Unit	22638.48
4	FUR-CH-10002024	HON 5400 Series Task Chairs for Big and Tall	21870.576
5	OFF-BI-10001359	GBC DocuBind TL300 Electric Binding System	19823.479000000003
6	OFF-BI-10000545	GBC Ibimaster 500 Manual ProClick Binding System	19024.5
7	TEC-CO-10001449	Hewlett Packard LaserJet 3310 Copier	18839.686
8	TEC-MA-10001127	HP Designjet T520 Inkjet Large Format Printer - 24. Color	18374.895
9	OFF-BI-10004995	GBC DocuBind P400 Electric Binding System	17965.068
10	OFF-SU-10000151	High Speed Automatic Electric Letter Opener	17030.311999999998

## Top selling products chart



## Customers Market Insights

### 1.6 Who are our highest-value customers?

Sean Miller (\$26,042) and Tamara Chand (\$19,052) top the customer list, primarily located in major metropolitan areas. These customers represent prime candidates for exclusive loyalty programs and personalized service offerings to enhance retention.

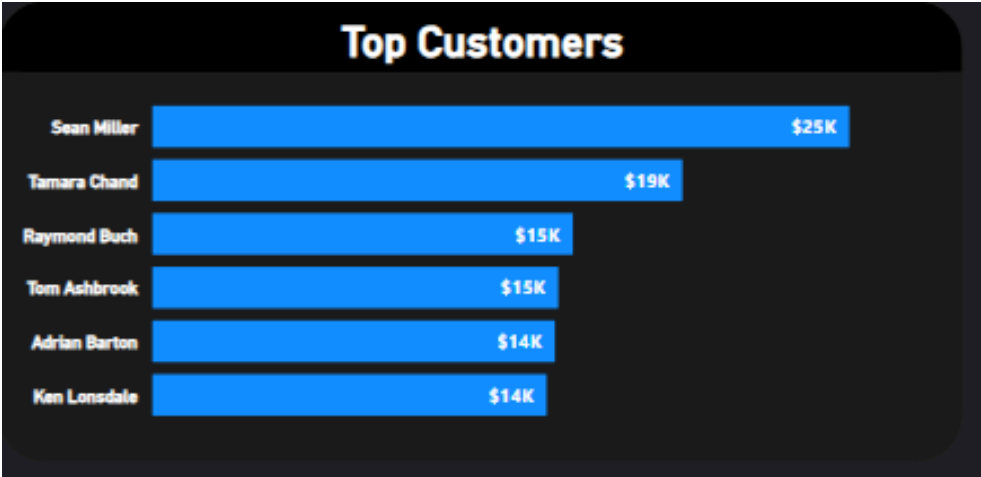
### SQL Query

```
SELECT o.Customer_ID, c.Customer_Name, SUM(o.sales) AS
total_sales
FROM orders o
JOIN customers c
ON o.Customer_ID = c.Customer_ID
GROUP BY o.Customer_ID, c.Customer_Name
ORDER BY total_sales DESC
LIMIT 10;
```

Query Results

#	customer_id	customer_name	total_sales
1	SM-20320	Sean Miller	25043.05
2	TC-20980	Tamara Chand	19052.217999999997
3	RB-19360	Raymond Buch	15117.339
4	TA-21385	Tom Ashbrook	14595.62
5	AB-10105	Adrian Barton	14473.570999999998
6	KL-16645	Ken Lonsdale	14175.229
7	SC-20095	Sanjit Chand	14142.333999999999
8	HL-15040	Hunter Lopez	12873.297999999999
9	SE-20110	Sanjit Engle	12209.438000000002
10	CC-12370	Christopher Conant	12129.072

The highest-value customers chart



## 1.7 How are sales distributed across customer segments?

Consumers drive **57.4%** of total sales (\$1,148,061), significantly outperforming Corporate (**34.4%**) and Home Office (**21.2%**) segments. This distribution suggests that consumer-focused marketing strategies yield the highest returns.

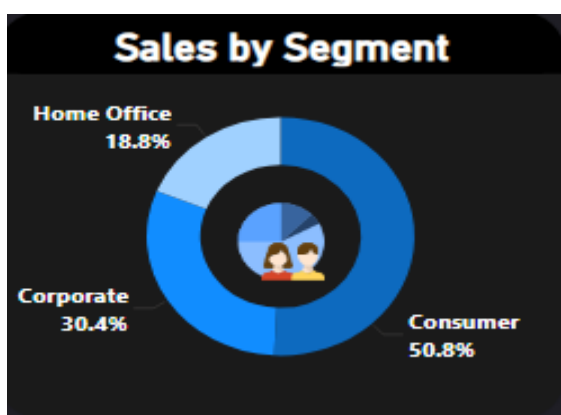
### SQL Query

```
SELECT c.Segment, SUM(o.sales) AS total_sales
FROM orders o
JOIN customers c
ON o.Customer_ID = c.Customer_ID
GROUP BY c.Segment
ORDER BY total_sales DESC;
```

### Query Results

#	segment	total_sales
1	Consumer	1148060.5309999897
2	Corporate	688494.0748000002
3	Home Office	424982.1769000005

### Customer segment sales chart



## 1.8 Which geographic markets perform best?

New York City leads with **\$209,428** in sales, followed by Los Angeles (**\$139,025**) and Philadelphia (**\$129,953**). The concentration of high sales in coastal cities indicates potential for expansion in inland metropolitan areas.

### SQL Query

```
SELECT c.State, c.City, SUM(o.sales) AS total_sales
FROM orders o
JOIN customers c
ON o.Customer_ID = c.Customer_ID
GROUP BY c.State, c.City
ORDER BY total_sales DESC
LIMIT 10;
```

### Query Results

#	state	city	total_sales
1	New York	New York City	209428.65610000017
2	California	Los Angeles	139025.11499999999
3	Pennsylvania	Philadelphia	129953.15500000003
4	California	San Francisco	104842.29150000005
5	Washington	Seattle	104647.1798
6	Texas	Houston	87154.40579999989
7	Illinois	Chicago	61106.194
8	California	San Diego	35344.126999999986
9	Texas	Dallas	29426.380000000005
10	Arizona	Phoenix	26920.074999999997

## Sales by city graph



### 1.9 How do customer segments impact regional sales?

Consumer spending dominates all regions, particularly in the West (\$438,747). Corporate sales are strongest in the East (\$216,597), while the Home Office shows balanced regional distribution. This data supports region-specific segment targeting in marketing campaigns.

### SQL Query

```
SELECT c.Segment, c.Region, SUM(o.sales) AS total_sales
FROM orders o
JOIN customer c ON o.Customer_ID = c.Customer_ID
GROUP BY c.Segment, c.Region
ORDER BY total_sales DESC;
```

## Query Results

#	segment	region	total_sales
1	Consumer	West	438747.96800000063
2	Consumer	East	268016.28770000034
3	Consumer	Central	258031.61130000075
4	Corporate	East	216597.7064000001
5	Corporate	West	188659.49190000017
6	Consumer	South	183264.66399999984
7	Corporate	Central	163684.47419999982
8	Home Office	East	121737.14340000002
9	Corporate	South	119552.4023
10	Home Office	West	116886.09179999994
11	Home Office	South	93823.55500000008
12	Home Office	Central	92535.38669999996



## 2. Operations s Shipping

### 2.1 What are the shipping mode preferences?

Standard Class shipping accounts for **59%** of shipments due to its cost- effectiveness, despite longer delivery times. Second Class (**20%**), First Class (**15%**), and Same Day (**5%**) complete the distribution, suggesting customers prioritize price over speed in most cases.

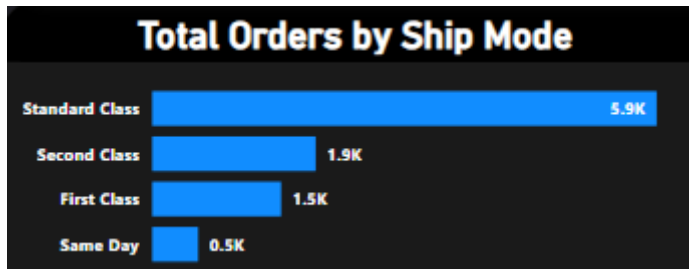
SQL Query

```
SELECT Ship_Mode, COUNT(*) AS total_shipments
FROM orders
GROUP BY Ship_Mode
ORDER BY total_shipments DESC;
```

### Query Results

#	ship_mode	total_shipments
1	Standard Class	5859
2	Second Class	1902
3	First Class	1501
4	Same Day	538

## Sales by ship\_mode chart



### 2.2 What is our average shipping time?

The average order takes **4 days** to deliver, meeting general industry standards. However, this figure masks significant regional variations that require attention to maintain customer satisfaction.

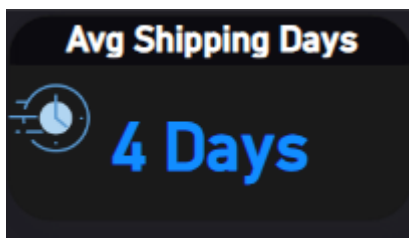
#### SQL Query

```
SELECT round(AVG(Ship_Date - Order_Date),2)
avg_shipping_days
FROM orders;
```

#### Query Results

#	avg_shipping_days
1	3.96

Average days card



### 2.3 What percentage of deliveries experience delays?

18.2% of shipments arrive late (exceeding promised delivery dates). This exceeds acceptable thresholds in e-commerce and risks damaging customer trust and retention rates.

SQL Query

```
SELECT COUNT(*) FILTER (WHERE Ship_Date > Order_Date +  
INTERVAL '5 days') * 100.0 / COUNT(*) AS  
delayed_percentage  
FROM orders;
```

Query Results

#	delayed_percentage
1	18.2142857142857143

## 2.4 How do delays vary by region?

The West region experiences the worst delay rates at **34.2%**, followed by the East (**25.4%**) and Central (**24.1%**). These figures indicate potential issues with regional distribution centers or carrier performance that require immediate operational review.

### SQL Query

```
WITH Shipping_Thresholds AS
( SELECT Ship_Mode, CASE WHEN Ship_Mode = 'Same Day'
THEN 1 WHEN Ship_Mode = 'First Class' THEN 2 WHEN
Ship_Mode = 'Second Class' THEN 3 WHEN Ship_Mode =
'Standard Class' THEN 5 ELSE 4 END AS max_shipping_days
FROM orders )

SELECT c.Region, COUNT(o.Order_ID) AS delayed_shipments,
ROUND(100.0 * COUNT(o.Order_ID) / SUM(COUNT(o.Order_ID))
OVER(), 2) AS delay_percentage
FROM orders o
JOIN customers c ON o.Customer_ID = c.Customer_ID
JOIN Shipping_Thresholds s ON o.Ship_Mode = s.Ship_Mode
WHERE (o.Ship_Date - o.Order_Date) > s.max_shipping_days
GROUP BY c.Region
ORDER BY delayed_shipments DESC;
```

### Query Results

#	region	delayed_shipments	delay_percentage
1	West	4389141	34.10
2	East	3265971	25.37
3	Central	3099952	24.08
4	South	2117339	16.45

Delays varies regions chart

Region	Total Orders	Late_Deliveries%
West	3160	39.9%
East	2709	38.7%
Central	2334	36.5%
South	1597	37.7%

## 2.5 How does shipping mode affect sales performance?

First Class shipments generate the highest average order value (\$238), suggesting customers spending more prefer faster delivery. Standard Class, while most popular, yields lower average values (\$229), indicating an opportunity to upsell faster shipping options.

### SQL Query

```
SELECT o.Ship_Mode,  
SUM(o.Sales) AS total_sales,  
COUNT(o.Order_ID) AS total_orders,  
ROUND(AVG(o.Sales), 2) AS avg_order_value,  
ROUND(AVG(o.Ship_Date - o.Order_Date), 2) AS  
avg_shipping_days  
FROM orders o  
GROUP BY o.Ship_Mode  
ORDER BY total_sales DESC;
```

### Query Results

#	ship_mode	total_sales	total_orders	avg_order_value	avg_shipping_days
1	Standard Class	1340831.30699999823	5859	229	5
2	Second Class	449914.179399999996	1902	237	3
3	First Class	345572.257300000035	1501	230	2
4	Same Day	125219.039	538	233	0

## Key Findings

1. Technology and Furniture categories are profit leaders.
2. The West region outperforms; the South needs intervention.
3. Consumer segment is the most lucrative.
4. Shipping delays are concentrated in the West.
5. Discounts above 20% erode margins (observed in Tables/Bookcases).

## Conclusion and Recommendations

Our analysis journey revealed both the strengths and challenges facing SuperStore. The company's data highlights clear opportunities for optimization.

### **We recommend the following actions:**

1. Product Optimization: Discontinue or repricing loss-making products (e.g., Tables).
2. Regional Focus: Boost marketing in the South; address West's shipping delays.
3. Customer Retention: Reward top customers (e.g., Sean Miller) with loyalty programs.
4. Discount Strategy: Limit discounts to <20% to protect margins.

## Appendix

GitHub Link:

<https://github.com/mariamrg212/superstore-analysis-project>