

# Global Superstore Sales Analysis



**BY: Numaira Zaib**

# Table of Contents

<b>Introduction .....</b>	<b>4</b>
<b>Dataset Overview and Structure .....</b>	<b>4</b>
<b>Key Features of the Dataset.....</b>	<b>4</b>
<b>Initial Observations .....</b>	<b>5</b>
<b>Data Preprocessing in Power BI .....</b>	<b>5</b>
<b>Introduction to Data Pre-processing .....</b>	<b>5</b>
<b>Checking and Handling Missing Values .....</b>	<b>6</b>
<b>Removing Duplicates.....</b>	<b>6</b>
<b>Correcting Data Types .....</b>	<b>6</b>
<b>Error Free data Images with some statistics.....</b>	<b>6</b>
<b>Merging Columns for Enhanced Data Representation.....</b>	<b>8</b>
<b>Feature Engineering: Creating New Columns .....</b>	<b>9</b>
<b>1. Profit Margin Calculation .....</b>	<b>9</b>
<b>2. Identifying High-Value Orders .....</b>	<b>9</b>
<b>Dashboard 1: Sales &amp; Profit Analysis.....</b>	<b>10</b>
<b>Dashboard Overview .....</b>	<b>10</b>
<b>Problem Statement .....</b>	<b>10</b>
<b>Key Performance Indicators (KPIs) .....</b>	<b>10</b>
<b>Dashboard Image .....</b>	<b>11</b>
<b>Data Visualization Explanation.....</b>	<b>11</b>
<b>Dashboard 2: Shipping Cost &amp; Discount Insights .....</b>	<b>13</b>
<b>Dashboard Overview .....</b>	<b>13</b>
<b>Problem Statements.....</b>	<b>13</b>
<b>Key Performance Indicators (KPIs) .....</b>	<b>13</b>
<b>Dashboard Image .....</b>	<b>14</b>
<b>Data Visualization Explanation.....</b>	<b>14</b>
<b>Star Schema: Data Organization .....</b>	<b>16</b>
<b>Five OLAP Operations.....</b>	<b>17</b>

<b>1. Roll-up Analysis: Yearly Summary of Sales and Profit .....</b>	<b>17</b>
<b>2. Drill-down Analysis: Monthly Breakdown of Sales and Profit .....</b>	<b>17</b>
<b>3. Slicing Operation: Sales in Technology Category .....</b>	<b>17</b>
<b>4. Dicing Operation: Profit for Appliances in Germany .....</b>	<b>17</b>
<b>5. Pivoting Table: Category-wise Sales by Region .....</b>	<b>17</b>
<b>Visualizations of OLAP Operations .....</b>	<b>17</b>
<b>Analyzing and Solving Profitability Issues in Table A Hypothesis-Driven Approach .....</b>	<b>18</b>
<b>1. Identifying the Problem.....</b>	<b>18</b>
<b>2. Hypothesis Testing and Findings.....</b>	<b>18</b>
<b>Hypothesis 1: Impact of Discounts on Profitability.....</b>	<b>18</b>
<b>Hypothesis 2: Effect of High Shipping Costs on Profitability .....</b>	<b>19</b>
<b>Hypothesis 3: Relationship Between High-Value Orders and Shipping Costs...</b>	<b>20</b>
<b>3. Proposed Solutions to Improve Table's Profitability .....</b>	<b>20</b>
<b>Conclusion .....</b>	<b>21</b>

# Introduction

In today's competitive business environment, data-driven decision-making is crucial in optimizing sales performance and improving efficiency. This report presents an in-depth analysis of Global Superstore Sales Data, covering various aspects of data preprocessing, visualization in Power BI, statistical hypothesis testing, and business insights.

This project aimed to explore sales data, identify key trends, and address challenges affecting business growth. By leveraging data preprocessing techniques, we ensured data quality and consistency. Interactive dashboards were then developed to visualize sales patterns and enhance business intelligence. Furthermore, hypothesis testing was conducted to validate assumptions and uncover meaningful statistical relationships. Finally, actionable recommendations were proposed to help improve sales, optimize inventory, and enhance customer satisfaction.

This study provides a comprehensive approach to analyzing large-scale retail data and demonstrates the importance of data analytics in making strategic business decisions.

## Dataset Overview and Structure

The Global Superstore Sales Dataset contains transactional data related to sales, customers, shipping, and product details. It provides valuable insights into various aspects of business operations, such as sales trends, customer segments, shipping preferences, and product performance.

### Key Features of the Dataset

The dataset consists of multiple columns, each representing specific information about an order. Below are some of the key attributes:

- **Order ID:** A unique identifier for each transaction.
- **Order Date & Ship Date:** The date an order was placed and the corresponding shipping date.
- **Ship Mode:** The type of shipping service used (e.g., First Class, Standard Class).
- **Customer Details:** Includes Customer ID, Name, and Segment (e.g., Consumer, Corporate).
- **Location Information:** Covers Postal Code, City, State, Country, and Region where the order was placed.
- **Market:** The geographical market segment (e.g., USCA for the United States and Canada).
- **Product Details:** Contains Product ID, Category, Sub-Category, and Product Name.
- **Sales & Quantity:** The total sales amount and the number of units sold.
- **Discount & Profit:** The discount applied and profit earned from the order.
- **Shipping Cost:** The cost incurred for shipping the order.
- **Order Priority:** The priority level assigned to the order (e.g., High, Medium, Low).

## Initial Observations

- The dataset contains a diverse range of products across multiple categories such as Technology, Furniture, and Office Supplies.
- Customers belong to different segments, including Consumer, Corporate, and Home Office.
- Sales data is recorded across various geographical locations, providing opportunities to analyze regional trends.
- Different shipping modes affect delivery time and costs, impacting customer satisfaction.
- The presence of discounts and varying profit margins suggests the need for pricing strategy analysis.

Understanding this dataset forms the foundation for further analysis, including data preprocessing, dashboard creation, and hypothesis testing, to extract meaningful business insights.

Home

global\_sales.xlsx

Sign in

Upgrade now

Menu

Home

Insert

Page Layout

Formulas

Data

Review

View

Tools

Smart Toolbox

Share

Format Painter

Paste

Calibri

11

Font settings icons

General

Conditional Formatting

Data Processing

Smart Toolbox

Settings

AE21

fx

#	A	B	C	D	E	F	G	H	I	J	K	L	
	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Postal C	City	State	Country	Region
1	40998	CA-2014-AB10015140-41954	11/11/2014	13/11/2014	First Class	AB-100151402	Aaron Bergman	Consumer	73120	Oklahoma City	Oklahoma	United States	Central
2	26341	IN-2014-JR162107-41675	05/02/2014	07/02/2014	Second Class	JR-162107	Justin Ritter	Corporate		Wollongong	New South Wales	Australia	Oceania
3	25330	IN-2014-CR127307-41929	17/10/2014	18/10/2014	First Class	CR-127307	Craig Reiter	Consumer		Brisbane	Queensland	Australia	Oceania
4	13524	ES-2014-KM1637548-41667	28/01/2014	30/01/2014	First Class	KM-1637548	Katharine Murray	Home Office		Berlin	Germany	Germany	Western
5	47221	SG-2014-RH9495111-41948	05/11/2014	06/11/2014	Same Day	RH-9495111	Rick Hansen	Consumer		Dakar	Dakar	Senegal	Western
6	22732	IN-2014-JM156557-41818	28/06/2014	01/07/2014	Second Class	JM-156557	Jim Mitchum	Corporate		Sydney	New South Wales	Australia	Oceania
7	30570	IN-2012-TS2134092-41219	06/11/2012	08/11/2012	First Class	TS-2134092	Toby Swindell	Consumer		Porirua	Wellington	New Zealand	Oceania
8	31192	IN-2013-MB1808592-41378	14/04/2013	18/04/2013	Standard Class	MB-1808592	Mick Brown	Consumer		Hamilton	Waikato	New Zealand	Oceania
9	40099	CA-2014-AB10015140-41954	11/11/2014	13/11/2014	First Class	AB-100151402	Aaron Bergman	Consumer	73120	Oklahoma City	Oklahoma	United States	Central
10	36258	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404	Aaron Bergman	Consumer	98103	Seattle	Washington	United States	Western
11	36259	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404	Aaron Bergman	Consumer	98103	Seattle	Washington	United States	Western
12	28879	ID-2013-AJ107801-41383	19/04/2013	22/04/2013	First Class	AJ-107801	Anthony Jacobs	Corporate		Kabul	Kabul	Afghanistan	Southern
13	45794	SA-2012-MM7260110-41269	26/12/2012	28/12/2012	Second Class	MM-7260110	Magdelene Morse	Consumer		Jizan	Jizan	Saudi Arabia	Western
14	4132	MX-2013-VF2171518-41591	13/11/2013	13/11/2013	Same Day	VF-2171518	Vicky Freymann	Home Office		Toledo	Parana	Brazil	South America
15	27704	IN-2014-PF1912027-41796	06/06/2014	08/06/2014	Second Class	PF-1912027	Peter Fuller	Consumer		Mudanjiang	Heilongjiang	China	Eastern
16	13779	ES-2015-BP118545-42216	31/07/2015	03/08/2015	Second Class	BP-118545	Ben Peternan	Corporate		Paris	Ile-de-France	France	Western
17	39519	CA-2012-AB10015140-40958	19/02/2012	25/02/2012	Standard Class	AB-100151402	Aaron Bergman	Consumer	76017	Arlington	Texas	United States	Central
18	12069	ES-2015-PI1883564-42255	08/09/2015	14/09/2015	Standard Class	PI-1883564	Patrick Jones	Corporate		Prato	Tuscany	Italy	Southern
19	22096	IN-2015-JS156857-42035	31/01/2015	01/02/2015	First Class	JS-156857	Jim Sink	Corporate		Townsville	Queensland	Australia	Oceania
20	49463	TZ-2015-RH955129-42343	05/12/2015	07/12/2015	Second Class	RH-955129	Ritsa Hightower	Consumer		Uvinza	Kigoma	Tanzania	Eastern
21	46630	PL-2013-AB600103-41494	08/08/2013	10/08/2013	First Class	AB-600103	Ann Blume	Corporate		Bytom	Silesia	Poland	Eastern
22	36260	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404	Aaron Bergman	Consumer	98103	Seattle	Washington	United States	Western
23	21586	IN-2012-JK1532527-41030	01/05/2012	02/05/2012	First Class	JK-1532527	Jason Klamczynski	Corporate		Suzhou	Anhui	China	Eastern
24	13528	ES-2014-LB16795139-41697	27/02/2014	01/03/2014	Second Class	LB-16795139	Laurel Beltran	Home Office		Edinburgh	Scotland	United Kingdom	Northern
25	1570	US-2015-NP1832582-42216	31/07/2015	01/08/2015	First Class	NP-1832582	Naresj Patel	Consumer		Julirez	Chihuahua	Mexico	Central
26	3484	MX-2015-VD2167039-42252	05/09/2015	08/09/2015	First Class	VD-2167039	Valerie Dominguez	Consumer		Soyapango	San Salvador	El Salvador	Central
27	30191	IN-2012-PB19210127-41259	16/12/2012	19/12/2012	First Class	PB-19210127	Phillip Breyer	Corporate		Taipei	Taipei City	Taiwan	Eastern
28	11645	ES-2012-EB141048-40961	13/03/2012	16/03/2012	Second Class	EB-141048	Eugene Barchas	Consumer		Leipzig	Saxony	Germany	Western
29	38460	CA-2012-AH10030140-41020	21/04/2012	23/04/2012	Second Class	AH-100301406	Aaron Hawkins	Corporate	12180	Troy	New York	United States	Eastern
30	22999	IN-2013-BP1123058-41329	24/02/2013	24/02/2013	Same Day	BP-1123058	Benjamin Patterson	Consumer		Surat	Gujarat	India	Southern
31	330	US-2015-PP1832582-42216	31/07/2015	01/08/2015	First Class	PP-1832582	Patricia Patterson	Consumer		Surat	Gujarat	India	Southern

## Data Preprocessing in Power BI

### Introduction to Data Pre-processing

Data Pre-processing is a crucial step in data analysis that involves cleaning, transforming, and preparing data for analysis and visualization. In Power BI, data preprocessing is performed using Power Query Editor, which allows users to handle missing values, remove duplicates, format data, and more.

#### Why is Preprocessing Important?

- Ensures data consistency and accuracy.
- Removes errors and missing values that can impact analysis.
- Standardizes data for better insights and visualization.
- Improves data quality for decision-making.

## Checking and Handling Missing Values

Different strategies are used for different columns:

Column	Missing Value Strategy
Postal code	Fill with N/A
Sales, profit, Quantity, Shipping cost	Replace with 0
Discount	If missing set to 0%

## Removing Duplicates

To ensure data accuracy, duplicate entries were identified and removed based on Order ID and other relevant attributes. This step prevented redundancy and maintained the integrity of sales records.

## Correcting Data Types

Incorrect data types can lead to computational errors and inconsistencies in visualization. For example:

- **Text** → Customer Name, Order ID
- **Date** → Order Date, Ship Date
- **Number** → Sales, Profit, Quantity
- **Currency** → Sales, Profit, Shipping Cost

## Error Free data Images with some statistics

	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID
1	40098	CA-2014-AB10015140-41954	11/11/2014	13/11/2014	First Class	AB-100151402
2	26341	IN-2014-JR162107-41675	05/02/2014	07/02/2014	Second Class	JR-162107
3	25330	IN-2014-CR127307-41929	17/10/2014	18/10/2014	First Class	CR-127307
4	13524	ES-2014-KM1637548-41667	28/01/2014	30/01/2014	First Class	KM-1637548
5	47221	SG-2014-RH9495111-41948	05/11/2014	06/11/2014	Same Day	RH-9495111
6	22732	IN-2014-JM156557-41818	28/06/2014	01/07/2014	Second Class	JM-156557
7	30570	IN-2012-TS2134092-41219	06/11/2012	08/11/2012	First Class	TS-2134092
8	31192	IN-2013-MB1808592-41378	14/04/2013	18/04/2013	Standard Class	MB-1808592
9	40099	CA-2014-AB10015140-41954	11/11/2014	13/11/2014	First Class	AB-100151402
10	36258	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404
11	36259	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404
12	28879	ID-2013-AJ107801-41383	19/04/2013	22/04/2013	First Class	AJ-107801
13	45794	SA-2012-MM7260110-41269	26/12/2012	28/12/2012	Second Class	MM-7260110
14	4132	MX-2013-VF2171518-41591	13/11/2013	13/11/2013	Same Day	VF-2171518
15	27704	IN-2014-PF1912027-41796	06/06/2014	08/06/2014	Second Class	PF-1912027
16	13779	ES-2015-BP1118545-42216	31/07/2015	03/08/2015	Second Class	BP-1118545
17	39519	CA-2012-AB10015140-40958	19/02/2012	25/02/2012	Standard Class	AB-100151402
18	12069	ES-2015-PJ1883564-42255	08/09/2015	14/09/2015	Standard Class	PJ-1883564
19	22096	IN-2015-JS156857-42035	31/01/2015	01/02/2015	First Class	JS-156857
20	49463	TZ-2015-RH9555129-42343	05/12/2015	07/12/2015	Second Class	RH-9555129
21	46630	PL-2013-AB600103-41494	08/08/2013	10/08/2013	First Class	AB-600103
22	36260	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404
23	21586	IN-2012-JK1532527-41030	01/05/2012	02/05/2012	First Class	JK-1532527

```

= Table.TransformColumnTypes("#Promoted Headers",{{"Row ID", Int64.Type}, {"Order ID", type text}, {"Order Date", type date}, {"Ship Date", type date}, {"Ship Mode", type text}, {"Customer ID", type text}, {"Customer Name", type text}, {"Segment", type text}, {"Postal Code", Int64.Type}, {"City", type text}, {"State", type text}, {"Country", type text}, {"Region", type text}, {"Market", type text}, {"Product ID", type text}, {"Category", type text}, {"Sub-Category", type text}, {"Product Name", type text}, {"Sales", type number}, {"Quantity", Int64.Type}, {"Discount", type number}, {"Profit", type number}, {"Shipping Cost", type number}, {"Order Priority", type text}})

```

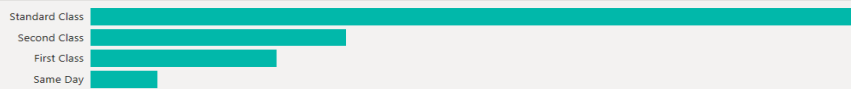
City	State	Country	Region	Market	Product ID	Category
Valid 100%	Valid 100%	Valid 100%	Valid 100%	Valid 100%	Valid 100%	Valid 100%
Error 0%	Error 0%	Error 0%	Error 0%	Error 0%	Error 0%	Error 0%
Empty 0%	Empty 0%	Empty 0%	Empty 0%	Empty 0%	Empty 0%	Empty 0%
Oklahoma City	Oklahoma	United States	Central US	USCA	TEC-PH-5816	Technology
Wollongong	New South Wales	Australia	Oceania	Asia Pacific	FUR-CH-5379	Furniture
Brisbane	Queensland	Australia	Oceania	Asia Pacific	TEC-PH-5356	Technology
Berlin	Berlin	Germany	Western Europe	Europe	TEC-PH-5267	Technology
Dakar	Dakar	Senegal	Western Africa	Africa	TEC-CO-6011	Technology
Sydney	New South Wales	Australia	Oceania	Asia Pacific	TEC-PH-5842	Technology

Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Postal Code
1	11/11/2014	13/11/2014	First Class	AB-100151402	Aaron Bergman	Consumer
2	05/02/2014	07/02/2014	Second Class	JR-162107	Justin Ritter	Corporate
3	17/10/2014	18/10/2014	First Class	CR-127307	Craig Reiter	Consumer
4	28/01/2014	30/01/2014	First Class	KM-1637548	Katherine Murray	Home Office
5	05/11/2014	06/11/2014	Same Day	RH-9495111	Rick Hansen	Consumer
6	28/06/2014	01/07/2014	Second Class	JM-156557	Jim Mitchum	Corporate
7	06/11/2012	08/11/2012	First Class	TS-2134092	Toby Swindell	Consumer
8	14/04/2013	18/04/2013	Standard Class	MB-1808592	Mick Brown	Consumer
9	11/11/2014	13/11/2014	First Class	AB-100151402	Aaron Bergman	Consumer

## Column statistics

Count	51290
Error	0
Empty	0
Distinct	4
Unique	0
Empty string	0
Min	First Class
Max	Standar...

## Value distribution

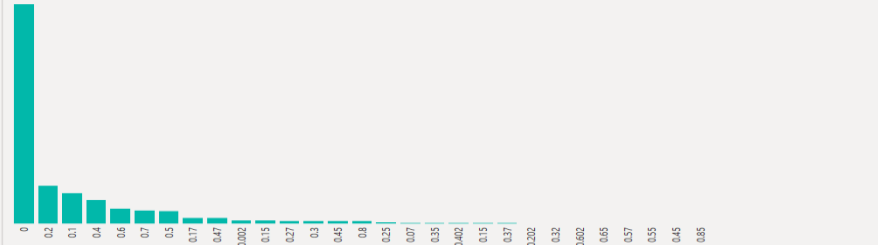


Sales	Quantity	Discount	Profit	Shipping Cost	Order Priority
1	221.98	2	0	62.1544	40.77 High
2	3709.395	9	0.1	-288.765	923.63 Critical
3	5175.171	9	0.1	919.971	915.49 Medium
4	2892.51	5	0.1	-96.54	910.16 Medium
5	2832.96	8	0	311.52	903.04 Critical
6	2862.675	5	0.1	763.275	897.35 Critical
7	1822.08	4	0	564.84	894.77 Critical
8	5244.84	6	0	996.48	878.38 High
9	341.96	2	0	54.7136	25.27 High

## Column statistics

Count	51290
Error	0
Empty	0
Distinct	29
Unique	0
NaN	0
Zero	29009
Min	0
Max	0.85
Average	0.14290...
Standard deviation	0.21227...

## Value distribution

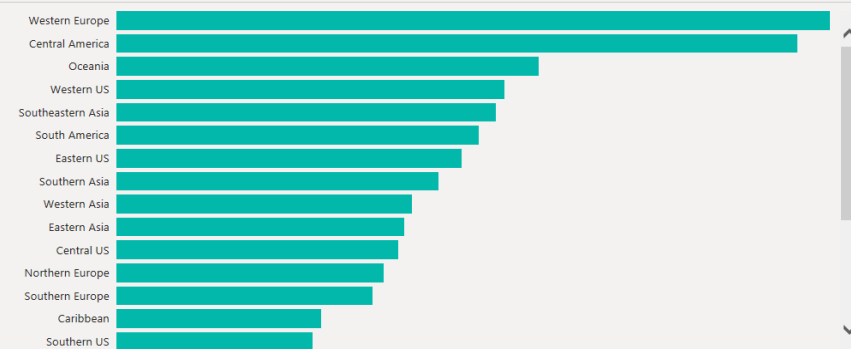


	A <sup>B</sup> <sub>C</sub> State	A <sup>B</sup> <sub>C</sub> Country	A <sup>B</sup> <sub>C</sub> Region	A <sup>B</sup> <sub>C</sub> Market	A <sup>B</sup> <sub>C</sub> Product ID	A <sup>B</sup> <sub>C</sub> Category
1	Oklahoma	United States	Central US	USCA	TEC-PH-5816	Technology
2	New South Wales	Australia	Oceania	Asia Pacific	FUR-CH-5379	Furniture
3	Queensland	Australia	Oceania	Asia Pacific	TEC-PH-5356	Technology
4	Berlin	Germany	Western Europe	Europe	TEC-PH-5267	Technology
5	Dakar	Senegal	Western Africa	Africa	TEC-CO-6011	Technology
6	New South Wales	Australia	Oceania	Asia Pacific	TEC-PH-5842	Technology
7	Wellington	New Zealand	Oceania	Asia Pacific	FUR-CH-5378	Furniture
8	Waikato	New Zealand	Oceania	Asia Pacific	FUR-TA-3764	Furniture
9	Oklahoma	United States	Central US	USCA	FUR-BO-5957	Furniture
10						

## Column statistics

Count	51290
Error	0
Empty	0
Distinct	23
Unique	0
Empty string	0
Min	Canada
Max	Western...

## Value distribution



	<div><div><div>123</div></div></div> Row ID	<div><div><div>ABC</div></div></div> Order ID	<div><div><div>Calendar</div></div></div> Order Date	<div><div><div>Calendar</div></div></div> Ship Date	<div><div><div>ABC</div></div></div> Ship Mode	<div><div><div>ABC</div></div></div> Customer ID	
	<div><div><div>Valid</div><div>Error</div><div>Empty</div></div></div> 100% <div>0%</div> <div>0%</div>	<div><div><div>Valid</div><div>Error</div><div>Empty</div></div></div> 100% <div>0%</div> <div>0%</div>	<div><div><div>Valid</div><div>Error</div><div>Empty</div></div></div> 100% <div>0%</div> <div>0%</div>	<div><div><div>Valid</div><div>Error</div><div>Empty</div></div></div> 100% <div>0%</div> <div>0%</div>	<div><div><div>Valid</div><div>Error</div><div>Empty</div></div></div> 100% <div>0%</div> <div>0%</div>	<div><div><div>Valid</div><div>Error</div><div>Empty</div></div></div> 100% <div>0%</div> <div>0%</div>	
1		40098	CA-2014-AB10015140-41954	11/11/2014	13/11/2014	First Class	AB-100151402
2		26341	IN-2014-JR162107-41675	05/02/2014	07/02/2014	Second Class	JR-162107
3		25330	IN-2014-CR127307-41929	17/10/2014	18/10/2014	First Class	CR-127307
4		13524	ES-2014-KM1637548-41667	28/01/2014	30/01/2014	First Class	KM-1637548
5		47221	SG-2014-RH9495111-41948	05/11/2014	06/11/2014	Same Day	RH-9495111
6		22732	IN-2014-JM156557-41818	28/06/2014	01/07/2014	Second Class	JM-156557
7		30570	IN-2012-TS2134092-41219	06/11/2012	08/11/2012	First Class	TS-2134092
8		31192	IN-2013-MB1808592-41378	14/04/2013	18/04/2013	Standard Class	MB-1808592
9		40099	CA-2014-AB10015140-41954	11/11/2014	13/11/2014	First Class	AB-100151402
10		36258	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404
11		36259	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404
12		28879	ID-2013-AJ107801-41383	19/04/2013	22/04/2013	First Class	AJ-107801
13		45794	SA-2012-MM7260110-41269	26/12/2012	28/12/2012	Second Class	MM-7260110
14		4132	MX-2013-VF2171518-41591	13/11/2013	13/11/2013	Same Day	VF-2171518
15		27704	IN-2014-PF1912027-41796	06/06/2014	08/06/2014	Second Class	PF-1912027
16		13779	ES-2015-BP1118545-42216	31/07/2015	03/08/2015	Second Class	BP-1118545
17		39519	CA-2012-AB10015140-40958	19/02/2012	25/02/2012	Standard Class	AB-100151402
18		12069	ES-2015-PJ1883564-42255	08/09/2015	14/09/2015	Standard Class	PJ-1883564
19		22096	IN-2015-JS156857-42035	31/01/2015	01/02/2015	First Class	JS-156857
20		49463	TZ-2015-RH9555129-42343	05/12/2015	07/12/2015	Second Class	RH-9555129
21		46630	PL-2013-AB600103-41494	08/08/2013	10/08/2013	First Class	AB-600103
22		36260	CA-2012-AB10015140-40974	06/03/2012	07/03/2012	First Class	AB-100151404
23		21586	IN-2012-JK1532527-41030	01/05/2012	02/05/2012	First Class	JK-1532527
24		13528	ES-2014-LB16795139-41697	27/02/2014	01/03/2014	Second Class	LB-16795139
25		15320	US-2015-MB1833563-42216	31/07/2015	01/08/2015	First Class	MB-1833563

## Merging Columns for Enhanced Data Representation

Certain columns were merged to create more comprehensive fields, improving data readability and streamlining analysis. Specifically, the City and State columns were combined into one column, enhancing the clarity of location-based insights.

<pre>= Table.CombineColumns("#Changed Type",{ "City", "State"},Combiner.CombineTextByDelimiter(", ", QuoteStyle.None),"city_state")</pre>					
AB C Segment	123 Postal Code	AB C city_state	AB C Country	AB C Region	AB C Marke

Before

AB C City	AB C State
Oklahoma City	Oklahoma
Wollongong	New South Wales
Brisbane	Queensland
Berlin	Berlin
Dakar	Dakar
Sydney	New South Wales
Porirua	Wellington
Hamilton	Waikato
Oklahoma City	Oklahoma
Seattle	Washington
Seattle	Washington
Kabul	Kabul
Jizan	Jizan
Toledo	Parana
Mudanjiang	Heilongjiang

After

AB C city_state
Oklahoma City,Oklahoma
Wollongong,New South Wales
Brisbane,Queensland
Berlin,Berlin
Dakar,Dakar
Sydney,New South Wales
Porirua,Wellington
Hamilton,Waikato
Oklahoma City,Oklahoma
Seattle,Washington
Seattle,Washington
Kabul,Kabul
Jizan,Jizan
Toledo,Parana
Mudanjiang,Heilongjiang



## Feature Engineering: Creating New Columns

Feature engineering plays a crucial role in enhancing data analysis by deriving new insights from existing variables. In this project, new calculated columns were created using DAX (Data Analysis Expressions) in Power BI to improve the analytical model's performance. Two key features were introduced:

### 1. Profit Margin Calculation

The Profit Margin column was created to analyze the profitability of each order. It is calculated using the formula:

$$\text{Profit Margin} = (\text{Profit} / \text{Sales}) \times 100$$

- A higher profit margin indicates better profitability.
- If the profit margin is negative, the order was sold at a loss.
- This metric helps assess how efficiently the company generates profit from sales.

**Custom Column**

Add a column that is computed from the other columns.

New column name: Profit Margin

Custom column formula: `= [Profit] / [Sales] * 100`

Available columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.

OK Cancel

**Profit Margin (%)**

ABC 123	Profit Margin (%)
	28
	-7.784692652
	17.77662999
	-3.337585695
	10.99627245
	26.66299877
	30.99973656
	18.99924497
	16
	11.25
	26
	13.9976006

### 2. Identifying High-Value Orders

A new column was created to classify orders as "High-Value" if the total sales amount exceeded \$500. This feature helps in:

- Identifying premium customers who place large orders.
- Analyzing purchasing trends of high-spending customers.

**Custom Column**

Add a column that is computed from the other columns.

New column name: High\_value\_orders

Custom column formula: `= if [Sales] > 500 then "Yes" else "No"`

Available columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.

OK Cancel

**High\_value\_orders**

ABC 123	High_value_orders
	No
	Yes
	Yes
	Yes
	Yes
	Yes
	Yes
	Yes
	No
	No

ABC 123	Profit Margin (%)	ABC 123	High_value_orders
Valid	100%	Valid	100%
Error	0%	Error	0%
Empty	0%	Empty	0%
	28	No	
	-7.784692652	Yes	
	17.77662999	Yes	
	-3.337585695	Yes	
	10.99627245	Yes	
	26.66299877	Yes	
	30.99973656	Yes	
	18.99924497	Yes	
	16	No	

This pre-processing ensures that data is clean, structured, and ready for analysis in Power BI. Each step verifies data integrity, making it suitable for visualization and decision-making.

## Dashboard 1: Sales & Profit Analysis

### Dashboard Overview

The Sales & Profit Analysis Dashboard is designed to help businesses track their sales performance, profit trends, and high-value orders. By analyzing key performance indicators (KPIs), the dashboard provides insights into revenue generation, profitability, and product performance.

### Problem Statement

Businesses often face challenges in understanding:

- Which products and categories contribute the most to profit.
- How sales and profit fluctuate over time.
- The impact of high-value orders on business revenue.
- The profitability of different categories and sub-categories.

How to optimize product sales to maximize overall profit. To address these concerns, the dashboard visualizes sales and profit metrics, enabling data-driven decision-making to improve business performance.

### Key Performance Indicators (KPIs)

**Total Sales:** The total revenue generated from all sales.

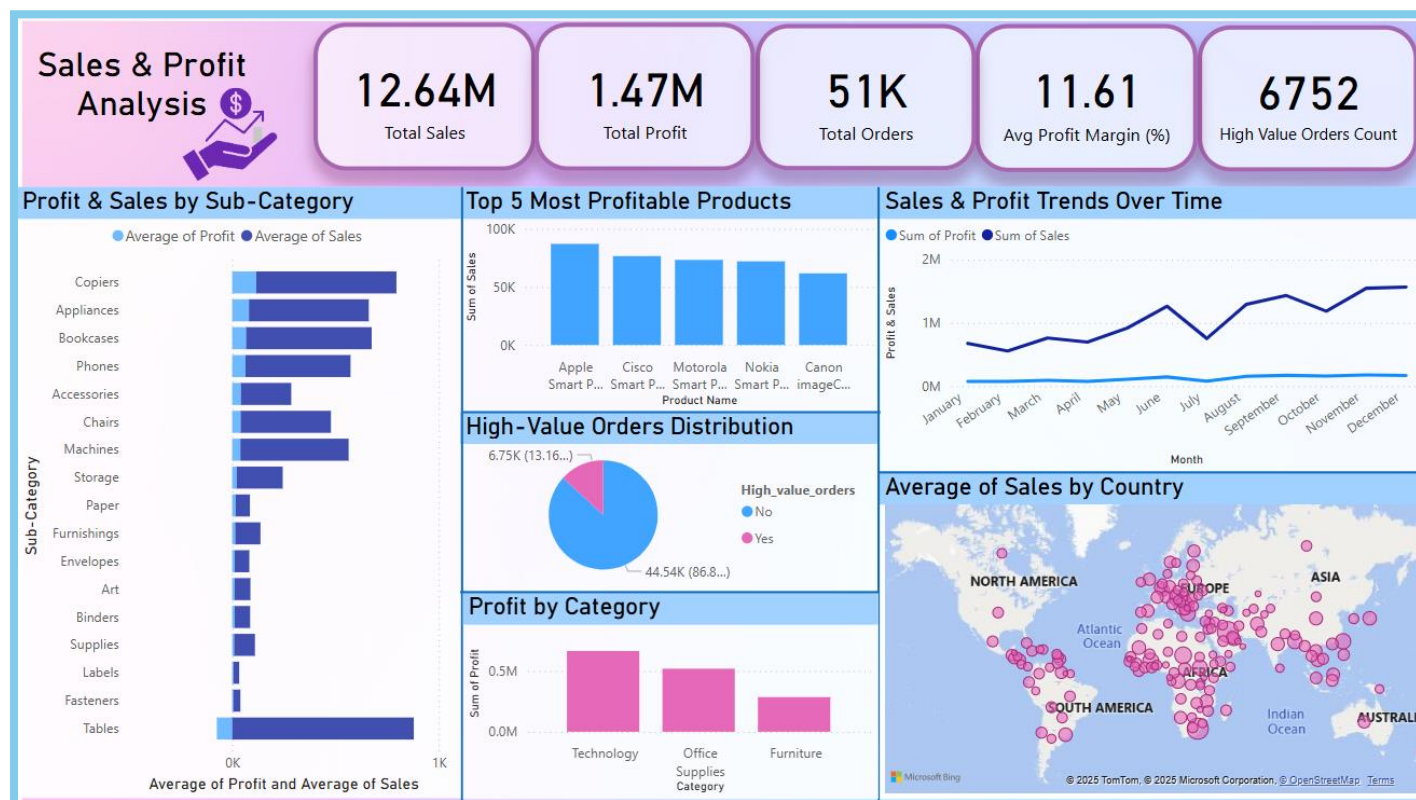
**Total Profit:** The overall profit after deducting costs.

**Total Orders:** total number of orders placed.

**Average Profit Margin (%):**  $(\text{Total Profit} / \text{Total Sales}) \times 100$ , measuring profitability efficiency percentage.

**High-Value Orders Count:** Identifies orders above a certain price threshold.

## Dashboard Image



## Data Visualization Explanation

### 1. Sales & Profit Trend Chart

**Chart Type:** Line Chart

**Metrics:**

- Sum of Sales (Dark Blue Line)
- Sum of Profit (Light Blue Line)

**Findings:**

This line chart displays the trend of total sales and profit over 12 months (January to December).

- Sales show an upward trend, peaking in June, September, and December, indicating revenue growth.
- Profit remains stable with minor fluctuations, highlighting controlled margins.
- The gap between sales and profit reflects operational costs and discounting impacts.
- Seasonal trends suggest higher demand mid-year and year-end, likely due to promotions or holidays.

**Use Case:** Helps identify sales and profit trends over time to optimize business strategy and financial planning.

### 2. Top 5 Most Profitable Products

**Chart Type:** Clustered column chart

**Metric:** Sum of Sales

**Findings:**

- The Apple Smart Phone has the highest sales among all products.
- Other high-performing products include Cisco Smart Phone, Motorola Smart Phone, Nokia Smart Phone, and Canon imageClass
- The sales volume among these top products is relatively consistent, indicating steady demand across high-value items.

**Use Case:** Helps in identifying best-selling products to optimize inventory and marketing strategies.

**3. High-Value Orders Distribution**

**Chart Type:** Pie Chart

**Metric:** High-Value Orders (Yes/No)

**Findings:**

- 86.8% of total orders are not high-value.
- Only 13.16% of orders qualify as high-value orders.
- Even though high-value orders are a smaller percentage, they might contribute significantly to overall revenue.

**Use Case:** Understanding the distribution of high-value orders can help optimize pricing strategies, promotions, and targeted customer engagement.

**4. Profit by Category**

**Chart Type:** Clustered column chart

**Metric:** Sum of Profit

**Findings:**

- Technology is the most profitable category.
- Office Supplies Category follows, with Furniture being the least profitable.
- The profit distribution varies significantly, highlighting the need for category-specific pricing and inventory strategies.

**Use Case:** Helps in profit optimization by identifying high-margin categories to focus on while addressing low-margin areas.

**5. Profit & Sales by Sub-Category**

**Chart Type:** Stacked bar chart

**Metric:** Average Profit (light blue), Average Sales (dark blue)

**Findings:**

- Tables have high sales but low negative profit, indicating pricing inefficiencies.
- Copiers, Appliances, and Machines balance sales and profit well.
- Fasteners, Labels, and Art have minimal sales and profit and need better marketing.

**Use Case:** Identifies profitable and underperforming sub-categories for better pricing and inventory decisions.

## 6. Average of Sales by Country

**Chart Type:** Map bubble size

**Metric:** Average Sales (bubble size & density)

**Findings:**

- North America, Europe, and Asia lead in sales, while Africa and South America show growth potential.
- Larger bubbles indicate major sales hubs, guiding resource allocation.

**Use Case:** Helps identify high-demand regions for market expansion and sales optimization.

This dashboard offers key insights into sales, profitability, product trends, and regional performance. It helps prioritize high-performing products, adjust low-profit items, and optimize inventory and marketing based on seasonal trends. Overall, it supports data-driven decisions for growth and efficiency.

## Dashboard 2: Shipping Cost & Discount Insights

### Dashboard Overview

This dashboard provides key insights into shipping costs and discount distribution and their impact on order fulfillment and profitability. It helps identify trends, inefficiencies, and optimization opportunities in logistics and discounts.

### Problem Statements

- Analyzing shipping expenses over time to identify cost-heavy periods and inefficiencies.
- Examining the impact of discounts on profitability to optimize pricing strategies.
- Evaluating fulfillment times across different shipping methods to enhance efficiency.
- Identifying high-cost regional order volume improves logistics and reduces shipping expenses.

### Key Performance Indicators (KPIs)

**Total Shipping Cost:** Provides an overview of overall shipping expenses

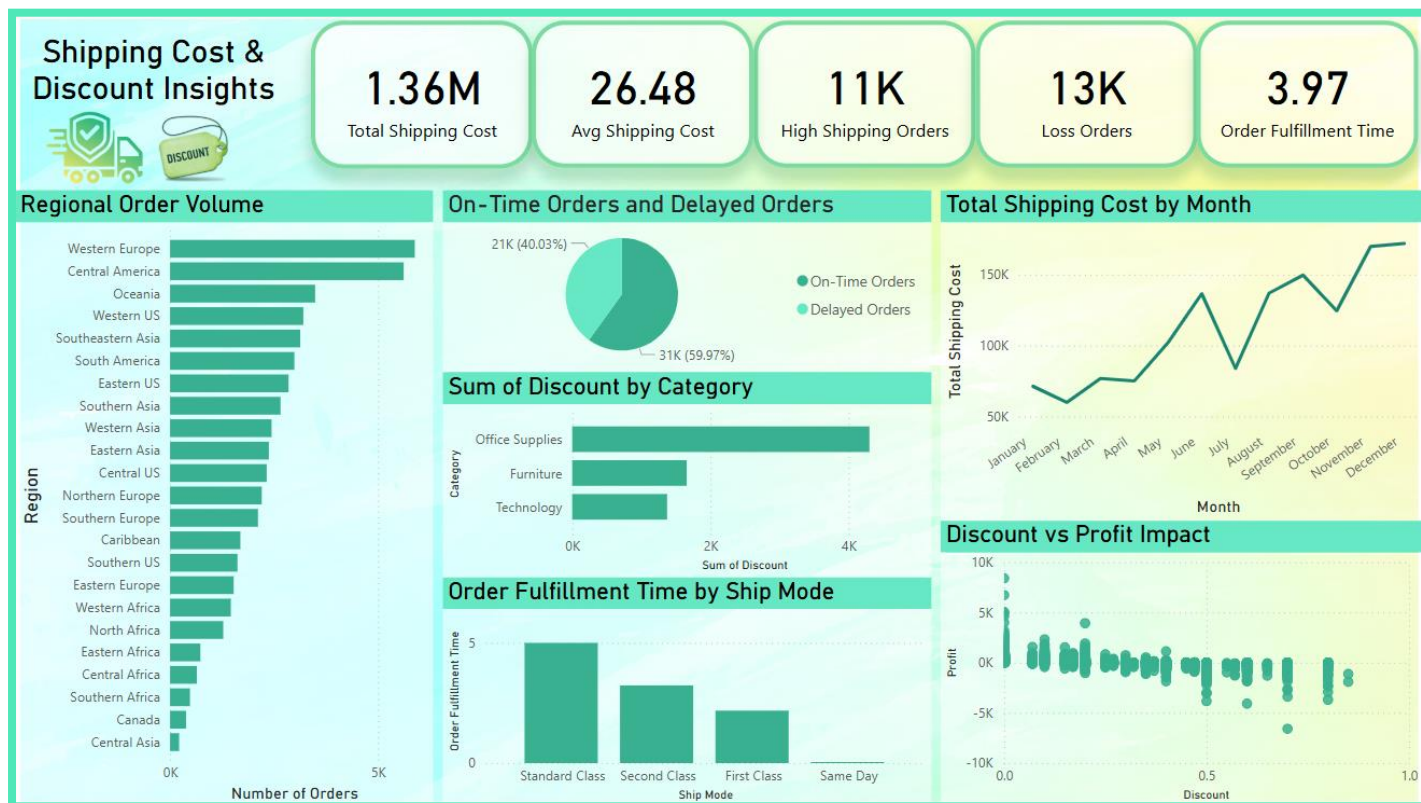
**Average Shipping Cost:** Provides the average cost per order, helping assess logistics efficiency.

**High Shipping Orders:** Identifies orders with excessive shipping costs, highlighting potential areas for cost optimization.

**Loss Orders:** Represents orders where costs outweigh revenue, often due to high shipping fees or large discounts.

**Order Fulfillment Time:** Measures the average time taken to process and deliver orders, influencing customer experience and efficiency.

## Dashboard Image



## Data Visualization Explanation

### 1. Regional Order Volume

**Chart Type:** Bar chart

**Metric:** Number of orders per region.

**Findings:**

- Western Europe and Central America have the highest order volumes.
- Central Asia and Canada have the lowest order volumes.
- Order volume distribution suggests a strong market presence in certain regions, while others may need better logistics or marketing strategies.

**Use Case:** Helps in identifying key markets for expansion and optimizing regional shipping strategies.

### 2. On-Time Orders and Delayed Orders

**Chart Type:** Pie Chart

**Metric:** Percentage of on-time vs. delayed orders.

**Findings:**

- 59.97% of orders are delayed, highlighting potential logistical inefficiencies.
- Only 40.03% of orders are delivered on time, which could impact customer satisfaction.
- A significant portion of delayed orders suggests a need for improved supply chain management.

**Use Case:** Helps in improving delivery performance by addressing bottlenecks in supply chain operations.

### 3. Sum of Discount by Category

**Chart Type:** Bar chart

**Metric:** Total discount provided across product categories.

**Findings:**

- Office Supplies receive the highest discounts, suggesting a competitive pricing strategy.
- Furniture receives moderate discounts, indicating controlled price adjustments.
- Technology has the lowest discounts, possibly due to higher margins or less price competition.

**Use Case:** Helps evaluate discounting strategies and assess profitability by category.

### 4. Total Shipping Cost by Month

**Chart Type:** Line Chart

**Metric:** Monthly total shipping cost trend.

**Findings:**

- Shipping costs increase steadily throughout the year, with peaks in June, September, and December.
- The rise in December suggests higher demand during the holiday season.
- Some months show a dip, indicating possible operational cost optimizations or seasonal fluctuations.

**Use Case:** Assists in budgeting and forecasting shipping expenses for peak periods.

### 5. Order Fulfillment Time by Ship Mode

**Chart Type:** Column chart

**Metric:** Average time taken to fulfill orders based on shipping mode.

**Findings:**

- Standard Class takes the longest time for fulfillment, impacting delivery speed.
- Same-day shipping is the fastest but likely incurs higher costs.
- Faster shipping modes (First Class, Second Class) provide a balance between speed and cost.

**Use Case:** Helps in optimizing shipping methods to improve delivery efficiency.

### 6. Discount vs. Profit Impact

**Chart Type:** Scatter plot

**Metric:** Relationship between discount amounts and profit.

**Findings:**

- Higher discounts tend to reduce profitability, with some transactions showing negative profits.
- Some low-discount orders still result in losses, indicating other cost factors.
- A balanced discounting approach is needed to maintain profitability without compromising sales.

**Use Case:** Helps in balancing discounting strategies to maximize revenue while maintaining profitability.

The dashboard highlights trends in fulfillment time, helping to streamline delivery processes for better customer satisfaction. By analyzing regional shipping cost variations, businesses can refine logistics strategies. These insights enable cost-effective decision-making while balancing efficiency and profitability.

## Star Schema: Data Organization

**Fact Table:** SalesFact — Central transactional data (sales, shipping, discounts, profits)

**Dimension Tables & Relationships:** (one to many)

SalesFact → CustomerData

- Linked by Customer ID (Customer details)

SalesFact → ProductData

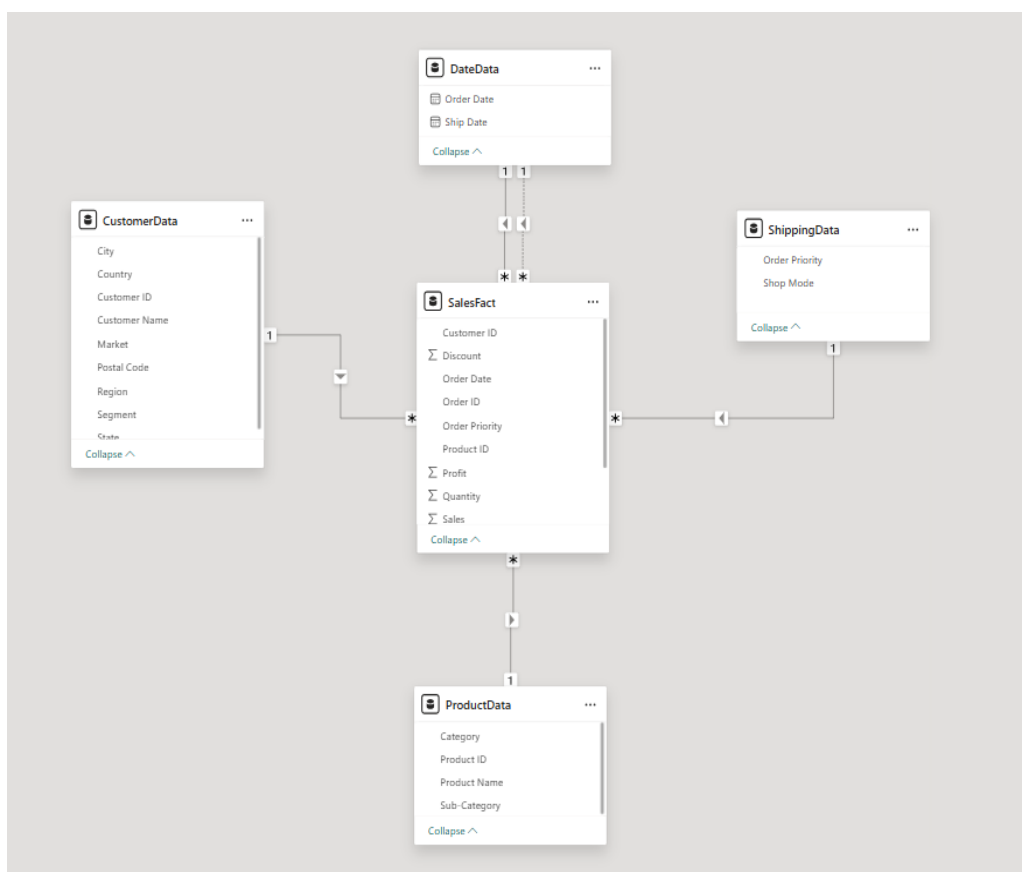
- Linked by Product ID (Product information)

SalesFact → ShippingData

- Linked by Ship Mode (Shipping method details)

SalesFact → DateData

- Linked by Order Date (Order timeline)
- Linked by Ship Date (Shipping timeline)





# Five OLAP Operations

## 1. Roll-up Analysis: Yearly Summary of Sales and Profit

(Chart: "Sum of Profit and Sum of Sales by Year")

→ Aggregates data from lower levels (e.g., months) to a higher time level (years).

## 2. Drill-down Analysis: Monthly Breakdown of Sales and Profit

(Chart: "Sum of Profit and Sum of Sales by Month")

→ Explores deeper time granularity within the year for detailed trend analysis.

## 3. Slicing Operation: Sales in Technology Category

(Chart: "Sum of Sales by Category")

→ Filters the entire dataset to focus only on the Technology category.

## 4. Dicing Operation: Profit for Appliances in Germany

(Chart: "Sum of Profit by Country and Sub-Category")

→ A narrow data slice filtered by Country = Germany and Sub-Category = Appliances.

## 5. Pivoting Table: Category-wise Sales by Region

(Table at the bottom)

→ Rearranges data to compare multiple metrics across Categories and Regions.

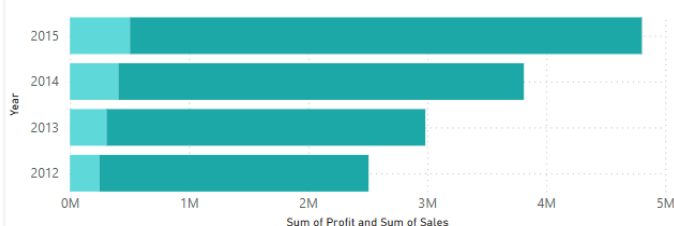
## Visualizations of OLAP Operations

Roll-up, Drill-down, Slicing, Dicing & Pivoting

### OLAP OPERATIONS

#### Roll-up Analysis: Yearly Summary of Sales and Profit

Sum of Profit Sum of Sales



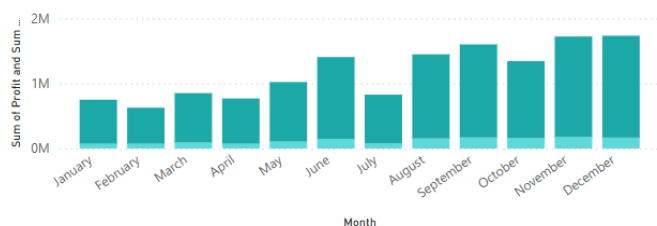
#### Slicing Operation: Sales in Technology Category

Category Technology



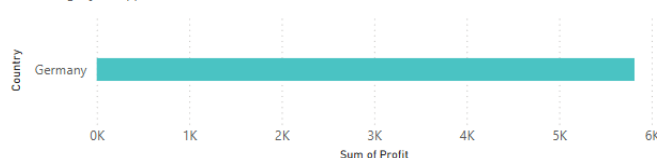
#### Drill-down Analysis: Monthly Breakdown of Sales and Profit

Sum of Profit Sum of Sales



#### Dicing Operation: Profit for Appliances in Germany

Sub-Category Appliances



#### Pivoting Table: Category-wise Sales by Region

Category	Canada	Caribbean	Central Africa	Central America	Central Asia	Central US	Eastern Africa	Eastern Asia	Eastern Europe	Eastern US	North Africa	Northern Europe	Oceania
Furniture	10,595.28	118,372.40	36,594.84	441,987.37	3,167.57	163,692.58	32,395.95	336,384.59	98,121.90	208,202.05	60,977.31	167,682.53	410,468.00
Office Supplies	30,034.08	89,575.42	45,345.18	319,442.98	7,953.15	167,131.00	51,277.47	203,284.03	103,652.61	205,605.21	85,540.41	216,127.54	281,713.60
Technology	26,298.81	116,333.05	61,689.99	461,670.28	8,190.74	170,416.31	44,182.60	315,390.77	108,258.93	264,973.98	86,698.89	252,969.09	408,002.50
<b>Total</b>	<b>66,928.17</b>	<b>324,280.86</b>	<b>143,630.01</b>	<b>1,223,100.63</b>	<b>19,311.46</b>	<b>501,239.89</b>	<b>127,856.02</b>	<b>855,059.39</b>	<b>310,033.44</b>	<b>678,781.24</b>	<b>233,216.61</b>	<b>636,779.16</b>	<b>1,100,184.60</b>

# Analyzing and Solving Profitability Issues in Table A

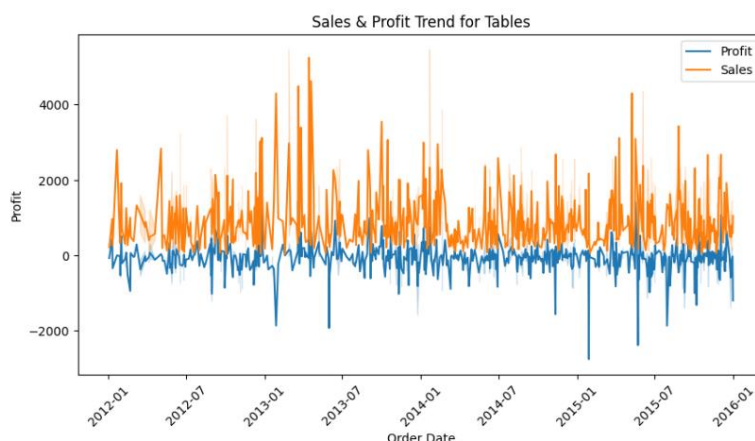
## Hypothesis-Driven Approach

### 1. Identifying the Problem

During our analysis of Table's sales and profitability data, we observed an unexpected trend: under specific conditions, Table is experiencing negative profits. This is a significant issue because profitability is a key measure of financial sustainability. If losses continue, it could impact the long-term viability of the business.

To better understand the root causes of this issue, we conducted a detailed hypothesis-driven analysis focusing on three key factors that could be influencing profitability:

- **High Discounts:** Discounts are meant to attract customers, but they might be reducing profits too much.
- **Shipping Costs:** Higher shipping costs could be eating into profit margins.
- **High-Value Orders and Shipping Costs:** High-value orders are expected to generate more profit, but they might be incurring disproportionately high shipping costs.



### 2. Hypothesis Testing and Findings

#### Hypothesis 1: Impact of Discounts on Profitability

##### Hypothesis Statement:

- Null Hypothesis ( $H_0$ ): Discounts do not significantly reduce profit.
- Alternative Hypothesis ( $H_1$ ): Discounts significantly reduce profit.

##### Analysis:

We categorized the Table data into two groups:

- High-discount orders (Discount > 30%)
- Low-discount orders (Discount ≤ 30%)

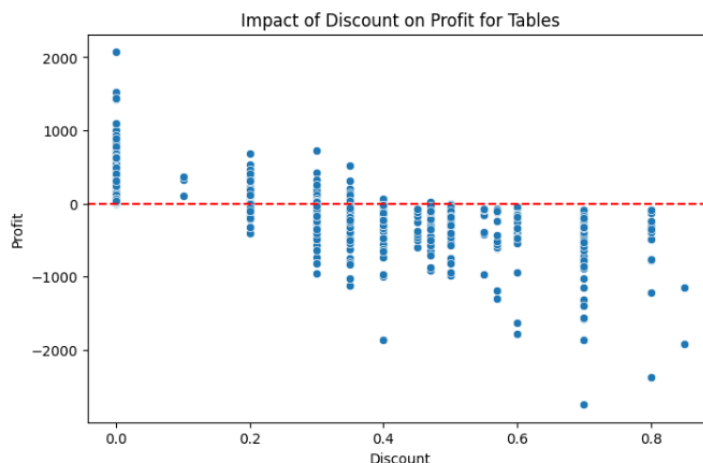
A t-test was performed to compare the profits of these two groups.

**Results:** T-statistic: -18.54 and P-value: 3.44e-61 (very small)

**Conclusion:** Since the p-value is far below 0.05, we reject  $H_0$  and conclude that discounts significantly reduce profit.

### Interpretation:

While discounts are a common strategy to boost sales, the data shows that high discounts are leading to losses rather than increasing profitability. This suggests that Table's discounting strategy needs to be revised.



## Hypothesis 2: Effect of High Shipping Costs on Profitability

### Hypothesis Statement:

- Null Hypothesis ( $H_0$ ): High shipping costs do not significantly reduce profit.
- Alternative Hypothesis ( $H_1$ ): High shipping costs significantly reduce profit.

### Analysis:

We divided the Table data into:

- High-shipping-cost orders (Above median shipping cost)
- Low-shipping-cost orders (Below median shipping cost)

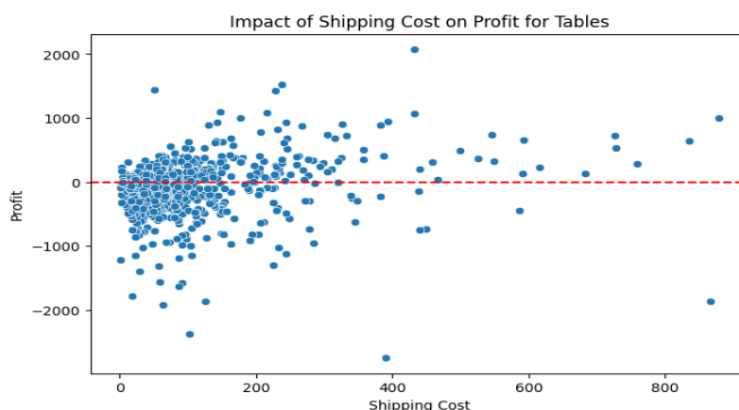
A t-test was conducted to compare profits between these two groups.

**Results:** T-statistic: 2.69 and P-value: 0.0074

**Conclusion:** Since the p-value is below 0.05, we reject  $H_0$  and conclude that high shipping costs significantly reduce profit.

### Interpretation for Table:

Shipping costs are significantly impacting profitability, which means that Table needs to optimize logistics expenses to avoid unnecessary costs.



### Hypothesis 3: Relationship Between High-Value Orders and Shipping Costs

#### Hypothesis Statement:

- Null Hypothesis ( $H_0$ ): There is no significant difference in shipping costs between high-value and low-value orders.
- Alternative Hypothesis ( $H_1$ ): High-value orders have significantly higher shipping costs.

#### Analysis:

- High-value orders were defined as those in the top 25% of total sales.
- We performed a t-test comparing shipping costs for high-value vs. low-value orders.

#### Results:

T-Statistic: 88.20 and P-value: 0.0

**Conclusion:** Since the p-value is extremely small, we reject  $H_0$  and confirm that high-value orders incur significantly higher shipping costs.

#### Interpretation:

Shipping costs should ideally scale proportionally with order value, but the data suggests that Table's high-value orders are disproportionately expensive to ship. This indicates inefficiencies in shipping cost allocation.



### 3. Proposed Solutions to Improve Table's Profitability

#### Optimizing Discount Strategies:

- Implement a tiered discounting model, where discounts are based on total spending rather than high discounts across all products.
- Shift towards loyalty-based discounts instead of excessive markdowns, ensuring that only repeat customers receive high-value incentives.

#### Reducing Shipping Costs:

- Negotiate better rates with shipping providers to lower logistics expenses.
- Introduce a minimum order value for free shipping to offset costs.

- Optimize packaging methods to reduce dimensional weight charges.

**Controlling High Shipping Costs for High-value Orders:**

- Establish regional distribution centers to minimize long-distance shipping expenses.
- Implement batch shipping where multiple orders are consolidated, reducing per-order shipping costs.

**Conclusion**

Our analysis of Table's profitability issues revealed that high discounts and excessive shipping costs are major contributors to negative profits. Through hypothesis testing, we quantified these impacts and identified key areas for improvement.

By optimizing discount strategies and improving shipping cost management, Table can effectively mitigate losses and enhance profitability. These data-driven recommendations ensure that Table's pricing and logistics strategies are aligned with sustainable business growth.