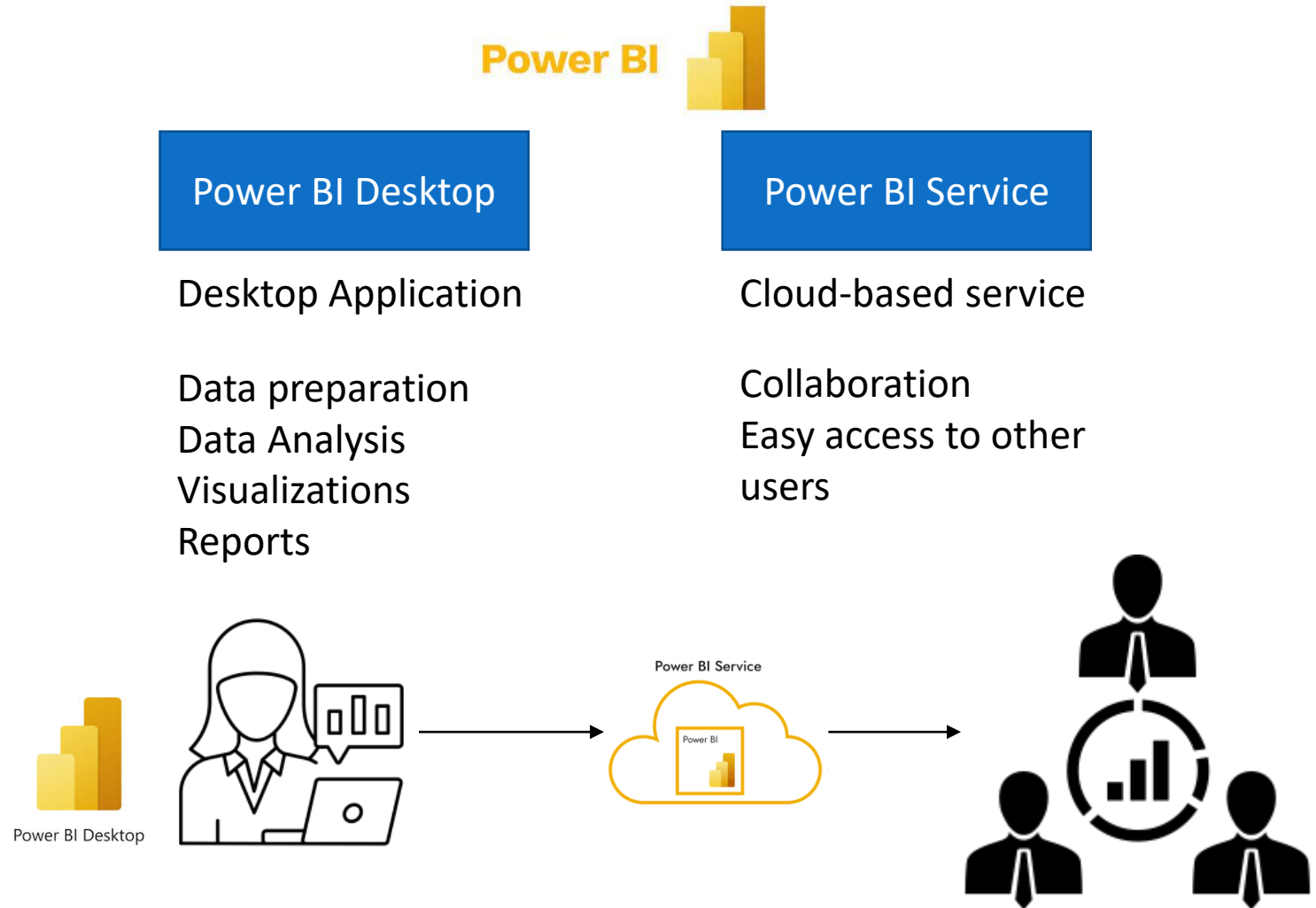




Start-Tech Academy

# Setting up Power BI

## Installation



# About the Datasets

## Three Datasets

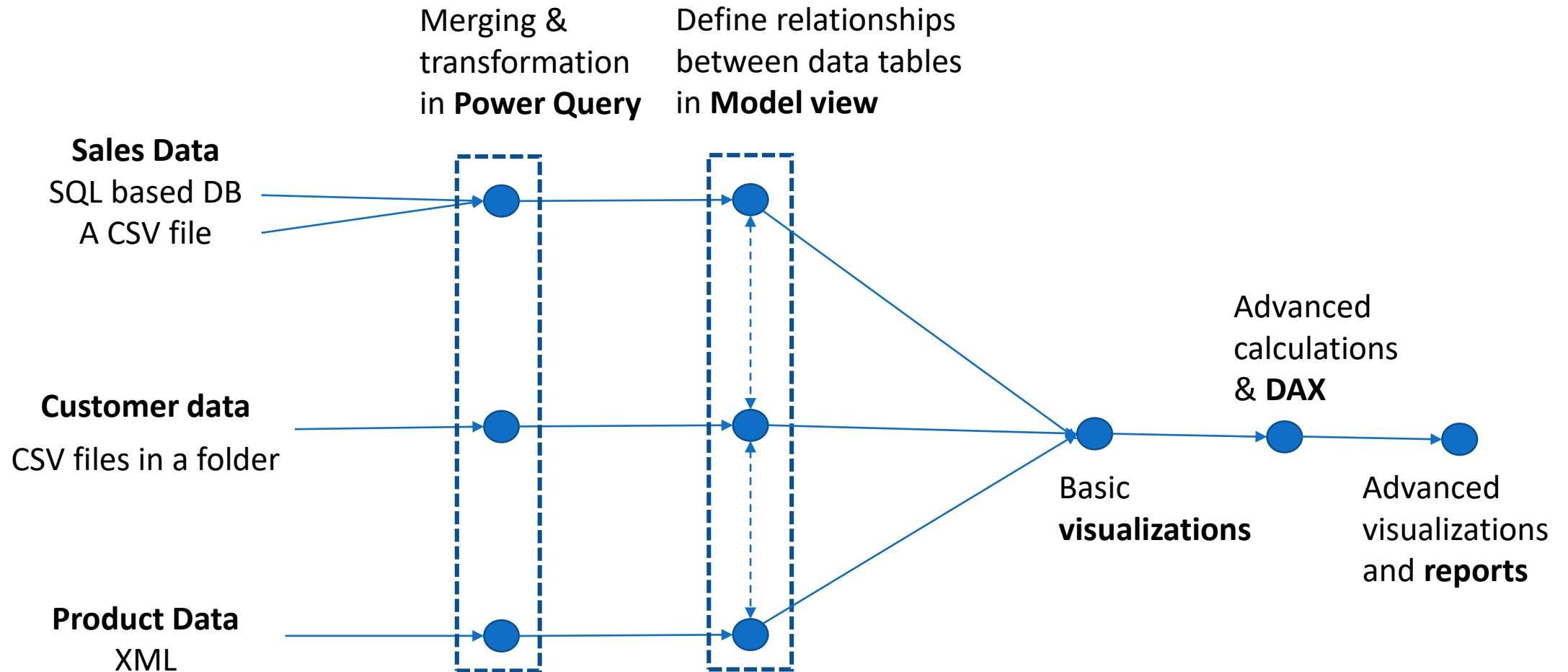
- Sales Data – Divided into two parts
  - Sales before 1<sup>st</sup> Jan 2023 – Kept as a CSV data dump
  - Sales after 1<sup>st</sup> Jan 2023 – Kept in an SQL database (CSV also available)
- Customer Data – Divided into four files region wise kept in a single folder
- Product Data – Saved as an XML file

### Why

- To explore different types of connection in Power BI
- Learn how to merge and append similar data coming from different sources
- **This is how the real-world data is**



# The Process



# Sales Data

## Steps

1. Two CSV files in the resources: **2023sales.csv** and **Salestill2022.csv**
2. Either set up PostgreSQL Database and load **2023sales.csv** file's data into it or you can use the csv file directly in Power BI.
3. Use Power BI to connect and fetch data from both sales tables.



# Setting up PostgreSQL

## CREATE TABLE

```
Create table sales (  
Order_Line int primary key,  
Order_ID varchar,  
Bar_code varchar,  
Order_Date date,  
Ship_Date date,  
Ship_Mode varchar,  
Customer_ID varchar,  
Product_ID varchar,  
Sales numeric,  
Quantity int,  
Discount numeric,  
Profit numeric  
);
```



# Setting up PostgreSQL

## Importing Data from CSV file

Change the location as per your installation directory  
COPY sales from 'C:\Program Files\PostgreSQL\14\data\dataset\2023sales.csv'  
delimiter ',' csv header;

*To Check if the data has been correctly imported, run the select command*  
SELECT \* FROM sales;



# Setting up PostgreSQL

## Summary

1. Install PostgreSQL and PGAdmin on your System
2. Create a new Database called **salesdb**
3. Open Query Tool and run the CREATE TABLE Command
4. Copy the '**2023sales.csv**' file and paste it in the data folder of PostgreSQL
5. Import data from this file using the Copy command





# Joining data from multiple tables

## Scenario 1: Merging two different tables

Sales Table

Order Line	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Product ID	Sales	Quantity	Discount	Profit
1	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	FUR-BO-10001798	261.96	2	0	41.9136
2	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	FUR-CH-10000454	731.94	3	0	219.582
3	CA-2016-138688	12-06-2016	16-06-2016	Second Class	DV-13045	OFF-LA-10000240	14.62	2	0	6.8714
4	US-2015-108966	11-10-2015	18-10-2015	Standard Class	SO-20335	FUR-TA-10000577	957.5775	5	0.45	-383.031
5	US-2015-108966	11-10-2015	18-10-2015	Standard Class	SO-20335	OFF-ST-10000760	22.368	2	0.2	2.5164

Sales values only

Customer Table

Customer ID	Customer Name	Segment	Age	Country	City	State	Postal Code	Region
CG-12520	Claire Gute	Consumer	67	United States	Henderson	Kentucky	42420	South
DV-13045	Darrin Van Huff	Corporate	31	United States	Los Angeles	California	90036	West
SO-20335	Sean O'Donnell	Consumer	65	United States	Fort Lauderdale	Florida	33311	South
BH-11710	Brosina Hoffman	Consumer	20	United States	Los Angeles	California	90032	West

Region values only



# 2 ways of joining data

Merging

Relationships



# Joining data from multiple tables

## Scenario 2: Appending similar tables

Customer ID	Customer Name	Segment	Age	Country	City	State	Postal Code	Region
EB-13870	Emily Burns	Consumer	34	United States	Orem	Utah	84057	West
EH-13945	Eric Hoffmann	Consumer	21	United States	Los Angeles	California	90049	West
TB-21520	Tracy Blumstein	Consumer	48	United States	Philadelphia	Pennsylvania	19140	East
MA-17560	Matt Abelman	Home Office	19	United States	Houston	Texas	77095	Central

Online  
customers



Customer ID	Customer Name	Segment	Age	Country	City	State	Postal Code	Region
ON-18715	Odella Nelson	Corporate	27	United States	Eagan	Minnesota	55122	Central
PO-18865	Patrick O'Donnell	Consumer	64	United States	Westland	Michigan	48185	Central
LH-16900	Lena Hernandez	Consumer	66	United States	Dover	Delaware	19901	East

Offline  
customers



Customer ID	Customer Name	Segment	Age	Country	City	State	Postal Code	Region
EB-13870	Emily Burns	Consumer	34	United States	Orem	Utah	84057	West
EH-13945	Eric Hoffmann	Consumer	21	United States	Los Angeles	California	90049	West
TB-21520	Tracy Blumstein	Consumer	48	United States	Philadelphia	Pennsylvania	19140	East
MA-17560	Matt Abelman	Home Office	19	United States	Houston	Texas	77095	Central
ON-18715	Odella Nelson	Corporate	27	United States	Eagan	Minnesota	55122	Central
PO-18865	Patrick O'Donnell	Consumer	64	United States	Westland	Michigan	48185	Central
LH-16900	Lena Hernandez	Consumer	66	United States	Dover	Delaware	19901	East



# Merge Queries

To merge tables we must know:

1. The names of the tables to be joined
2. The common column based on which we will join them
3. The list of columns from each table

## What's needed

Order Line	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Product ID	Sales	Quantity	Discount	Profit
1	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	FUR-BO-10001798	261.96	2	0	41.9136
2	CA-2016-152156	08-11-2016	11-11-2016	Second Class	CG-12520	FUR-CH-10000454	731.94	3	0	219.582
3	CA-2016-138688	12-06-2016	16-06-2016	Second Class	DV-13045	OFF-LA-10000240	14.62	2	0	6.8714
4	US-2015-108966	11-10-2015	18-10-2015	Standard Class	SO-20335	FUR-TA-10000577	957.5775	5	0.45	-383.031
5	US-2015-108966	11-10-2015	18-10-2015	Standard Class	SO-20335	OFF-ST-10000760	22.368	2	0.2	2.5164

Customer ID	Customer Name	Segment	Age	Country	City	State	Postal Code	Region
CG-12520	Claire Gute	Consumer	67	United States	Henderson	Kentucky	42420	South
DV-13045	Darrin Van Huff	Corporate	31	United States	Los Angeles	California	90036	West
SO-20335	Sean O'Donnell	Consumer	65	United States	Fort Lauderdale	Florida	33311	South
BH-11710	Brosina Hoffman	Consumer	20	United States	Los Angeles	California	90032	West



# Transformation

## Pivoting and Unpivoting

Pivoted data

States	Consumer	Corporate	Home Office
California	86	45	26
New York	44	24	19
Texas	42	25	7
Pennsylvania	23	18	9
Illinois	14	11	16
Ohio	24	9	5
Washington	27	9	2
North Carolina	17	9	4

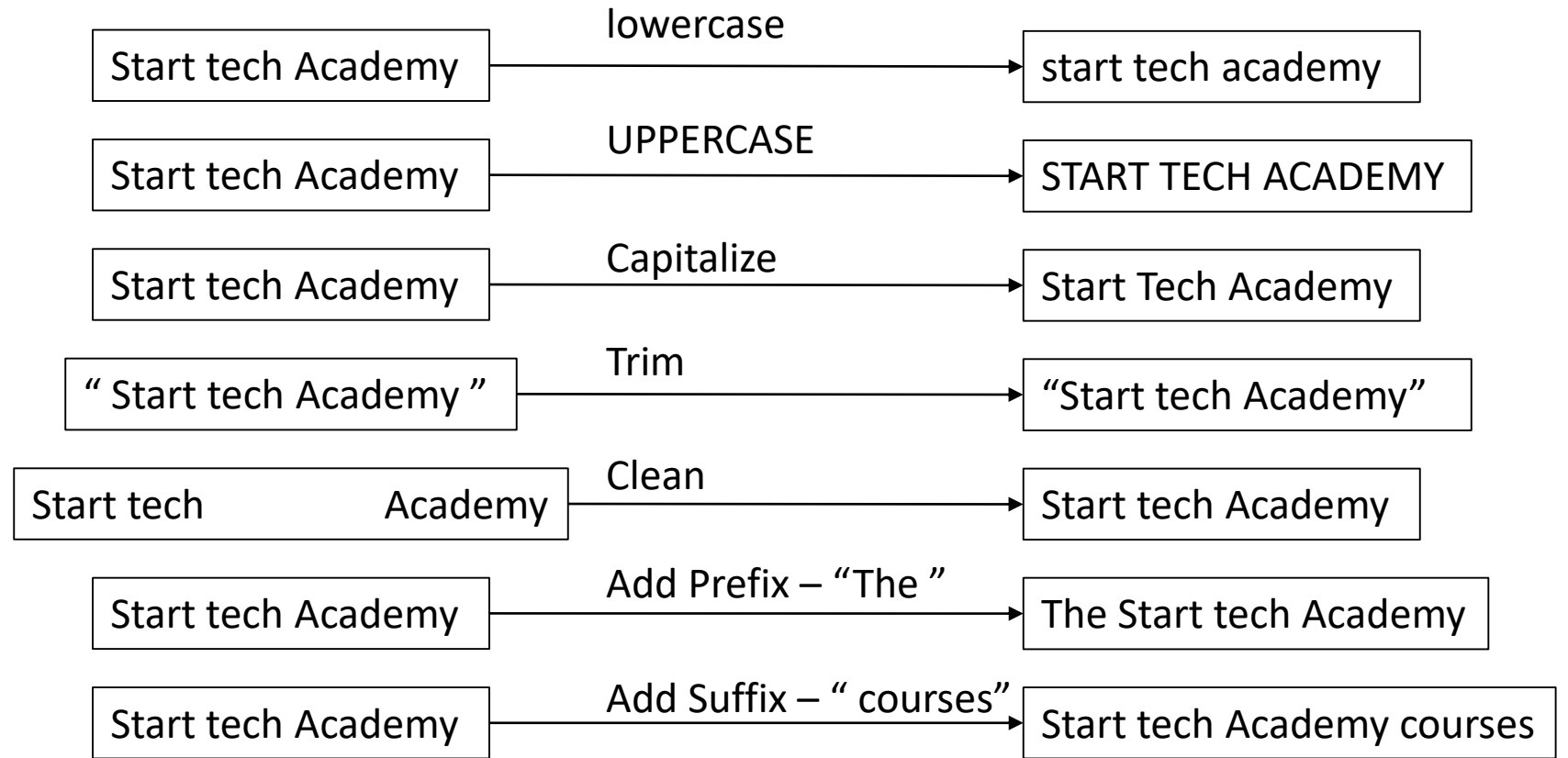
Unpivoted data

State	Segment	Number of Customers
California	Consumer	86
California	Corporate	45
California	Home Office	26
New York	Consumer	44
New York	Corporate	24
New York	Home Office	19
Texas	Consumer	42
Texas	Corporate	25
Texas	Home Office	7
Pennsylvania	Consumer	23
Pennsylvania	Corporate	18
Pennsylvania	Home Office	9
Illinois	Consumer	14
Illinois	Corporate	11
Illinois	Home Office	16
Ohio	Consumer	24
Ohio	Corporate	9
Ohio	Home Office	5



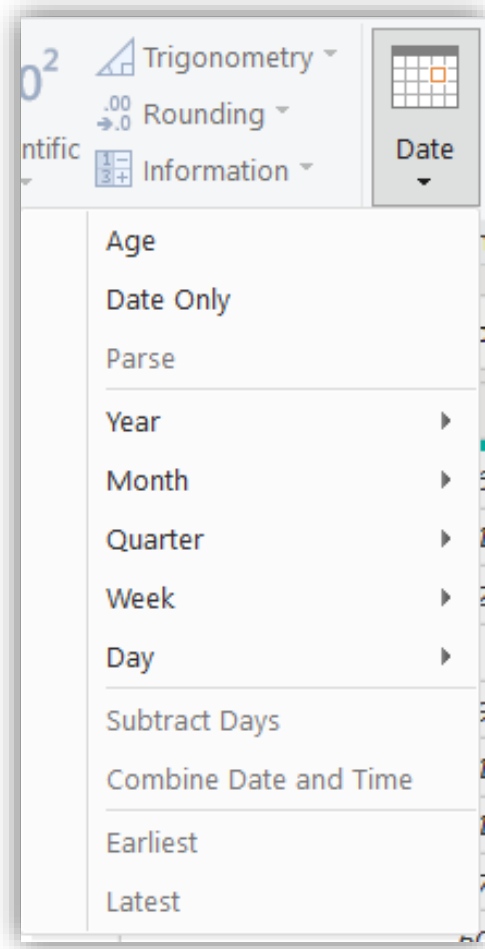
# Text Transformation

## Format options



# Transformations

## Date



**Year** - To extract the year from a date column for year-based trend analysis

**Quarter** - Isolate the quarter from each date for quarterly analysis.

**Month** - Extracts the month from each date, enabling trend analysis on a monthly basis.

**Week** – Extract week of the year for weekly analysis

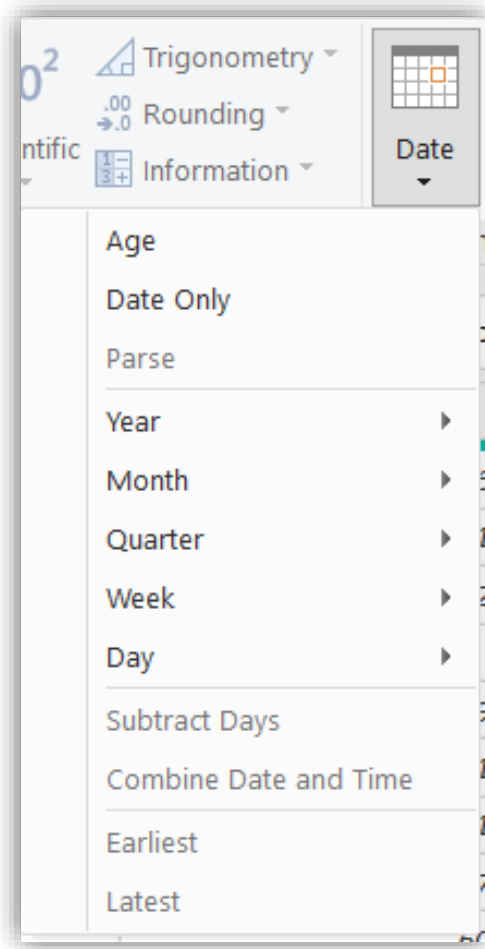
**Day** - Extract daily data from a date for daily sales or usage trend analysis

**Date Only** – Get only dates from date-time data



# Transformations

## Date



**Age** – Calculate the difference between the given date and current date.

**Parse** - Convert dates stored as text into a recognizable date format

**Subtract Days** – Subtract 'N' days from any date column

**Combine date and time** – Combine separate date and time data into one column

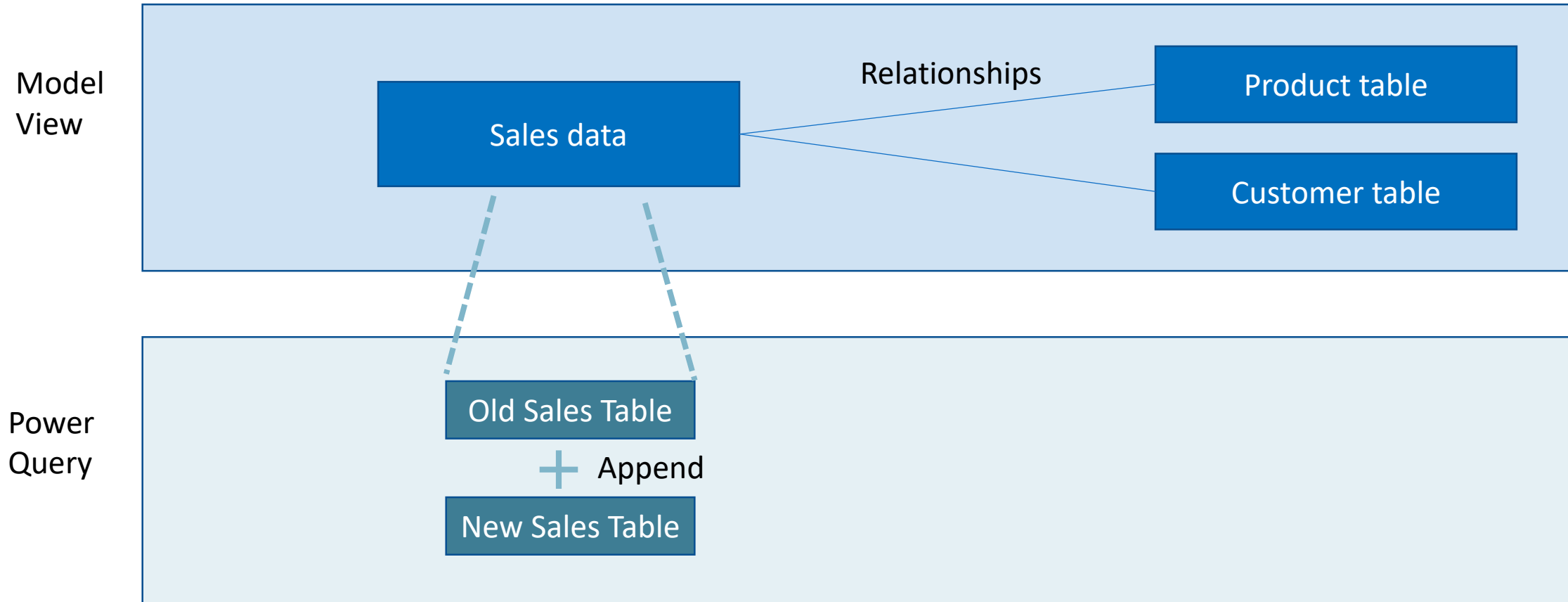
**Earliest** – Find the start or the earliest date in the data

**Latest** – Find the last or the latest date in the data





# Data Model



# Data Model

## Facts and Dimensions

- A **fact table** stores numerical measurements of the business as a quantity of products sold, discounts, taxes, number of invoices, and anything that can be measured.
- These measurements are referred to as **facts**.
- **Dimension tables** contain the textual descriptors of the business.
- Typical dimensions are product, time, customers, and regions.

In our Example,

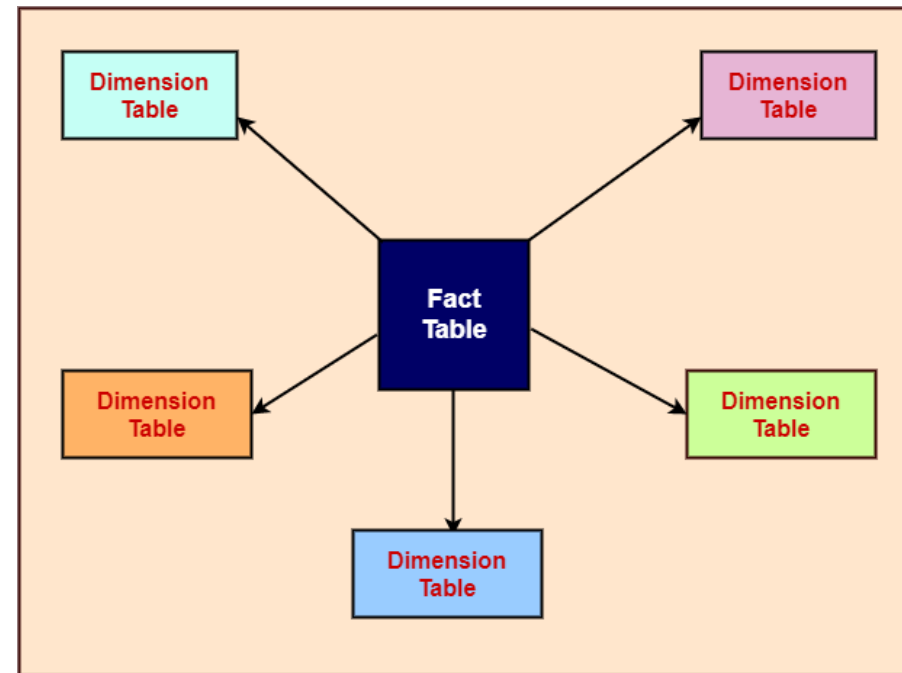
- Sales table is a fact table
- Customer and Product table are Dimension tables



# Data Model

## Star Schema

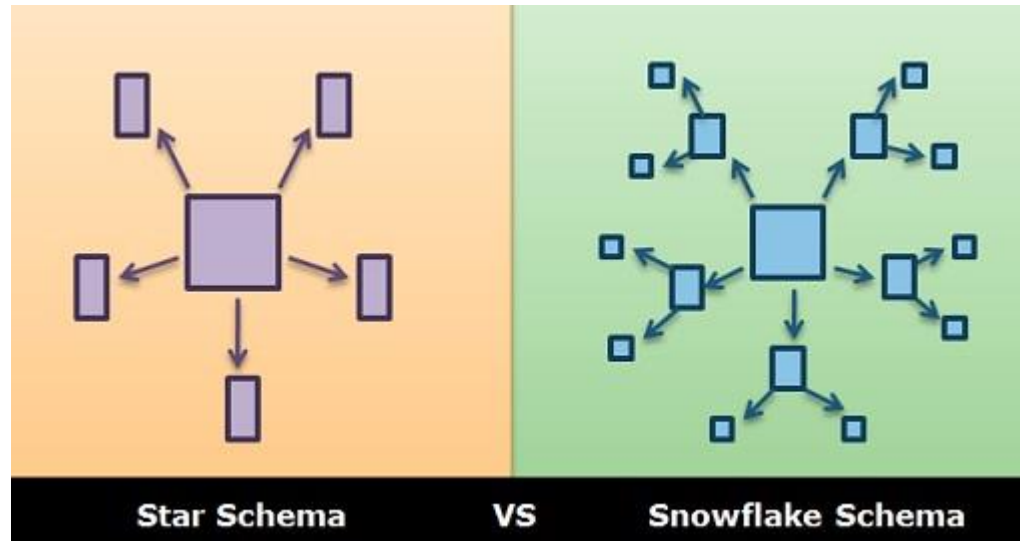
A fact table is at the centre which is connected with several dimension tables



# Data Model

## Snowflake Schema

- Dimension tables are further connected with sub-dimension tables



# Relationship

## Cardinality

Customer table

Customer ID	Postal code
CG-12520	42420
DV-13045	90036
SO-20335	33311
BH-11710	90036

Are postal code values unique in this table?

Not unique -> **'Many'**

Postal code master table

Postal code	City	State	Region
32303	Tallahassee	Florida	South
32725	Deltona	Florida	South
32935	Melbourne	Florida	South
33012	Hialeah	Florida	South
33311	Fort Lauderdale	Florida	South
.	.	.	.
.	.	.	.

Are postal code values unique in this table?

Unique -> **'One'**

**'Many to One'**

*We are joining customer table with postal code master - postal code is the matching key*



# Relationship

## Cardinality

Customer table

Customer ID	Postal code
CG-12520	42420
DV-13045	90036
SO-20335	33311
BH-11710	90036

Is customer ID unique in this table?

Unique -> **'One'**

Reference table

Customer ID	Ref Name	Ref Contact
CG-12520	Cindy Stewart	10897310
CM-11935	Dan Campbell	16589278
CM-12385	Darren Koutras	95721837
CG-12520	Denny Ordway	16507437
CS-12355	Evan Bailliet	76772276
CS-12460	Erica Hackney	69524187

Is customer ID unique in this table?

Not unique -> **'Many'**

**'One to Many'**

*We are joining customer table with reference table – customer ID is the matching key*

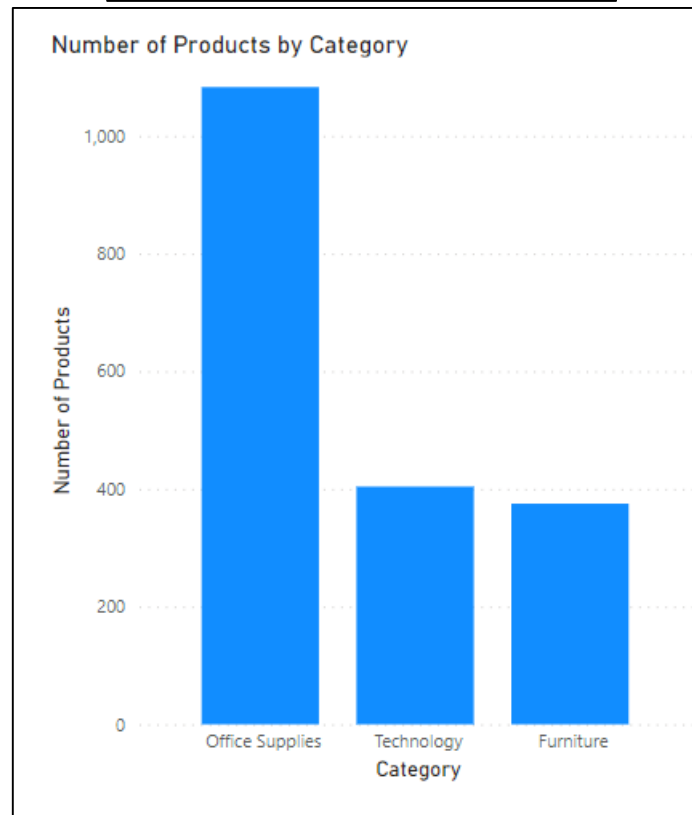


# Data Visualization

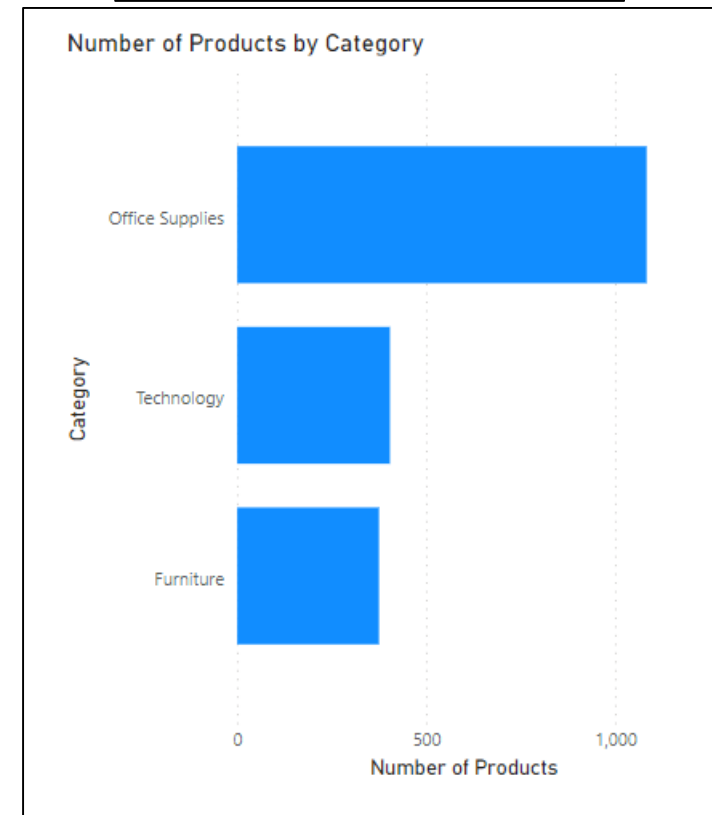
## Bar charts

Best for representing and comparing quantities for different categories.

**Column chart**



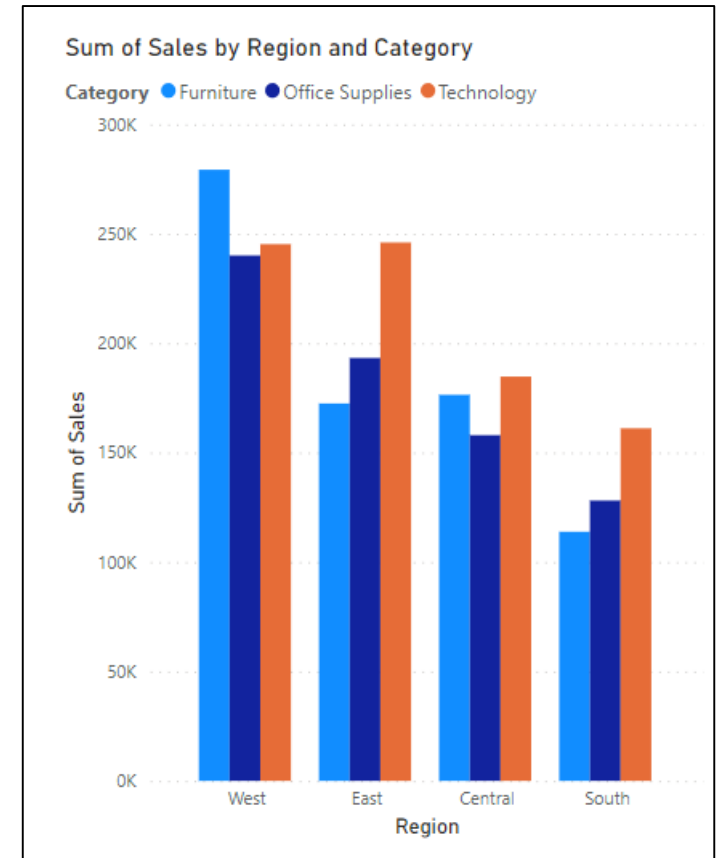
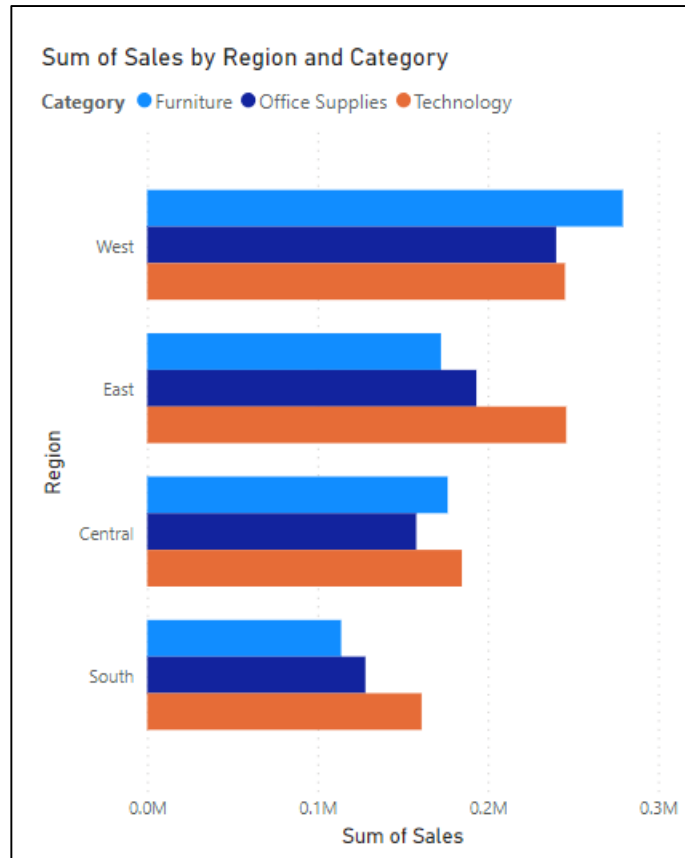
**Bar chart**



# Data Visualization

## Clustered Bar charts

Allows comparison not just between the main categories, but also between individual sub-categories

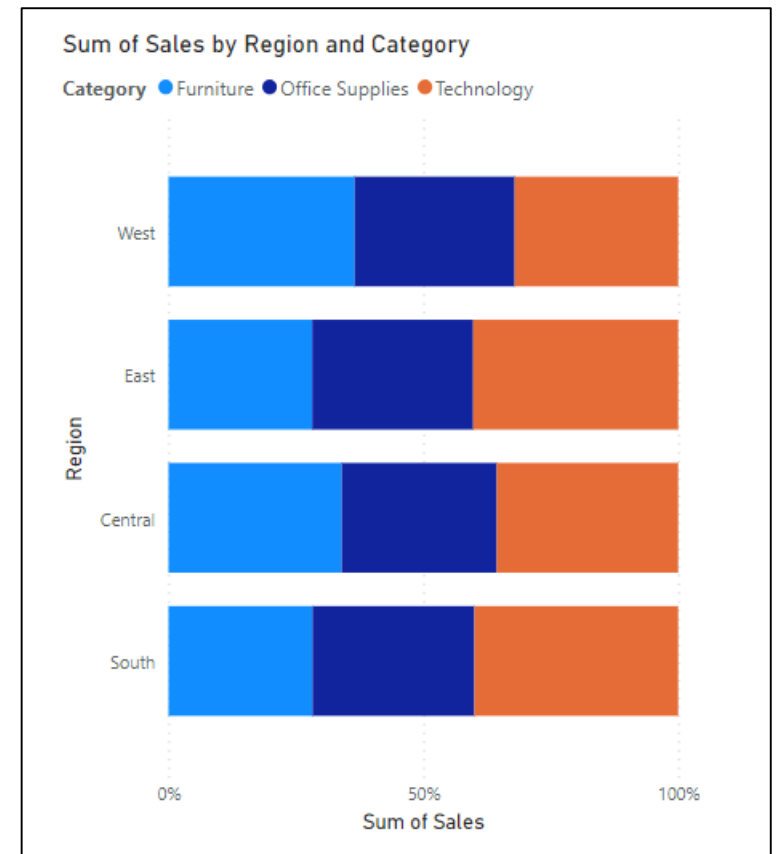
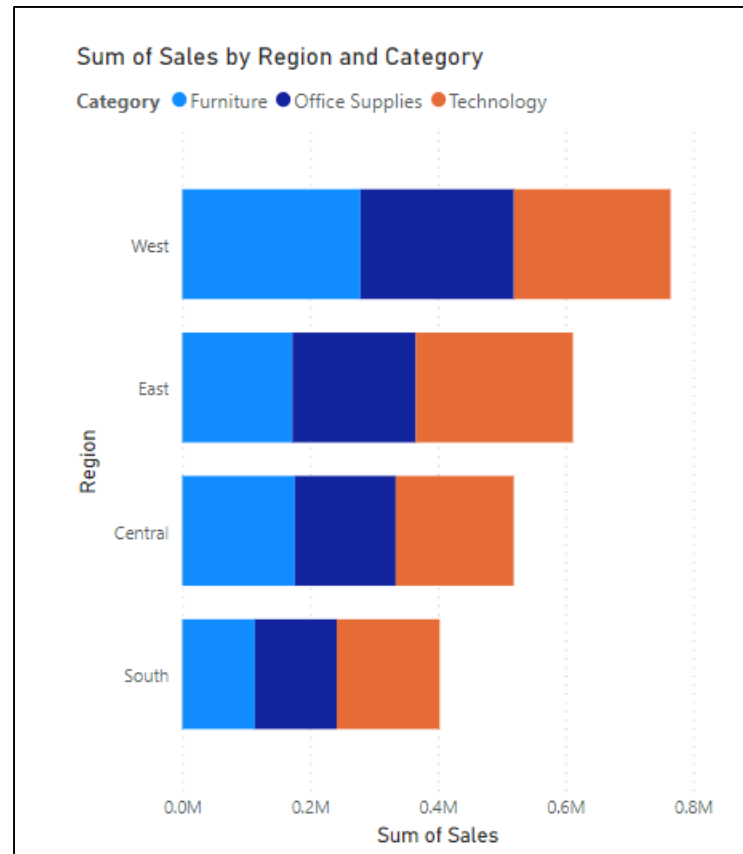




# Data Visualization

## Stacked Bar charts

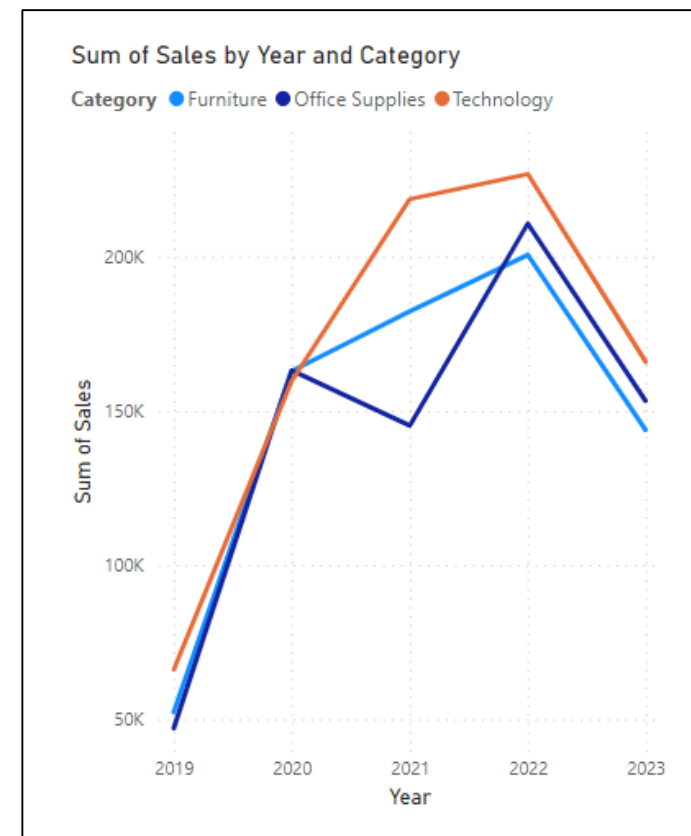
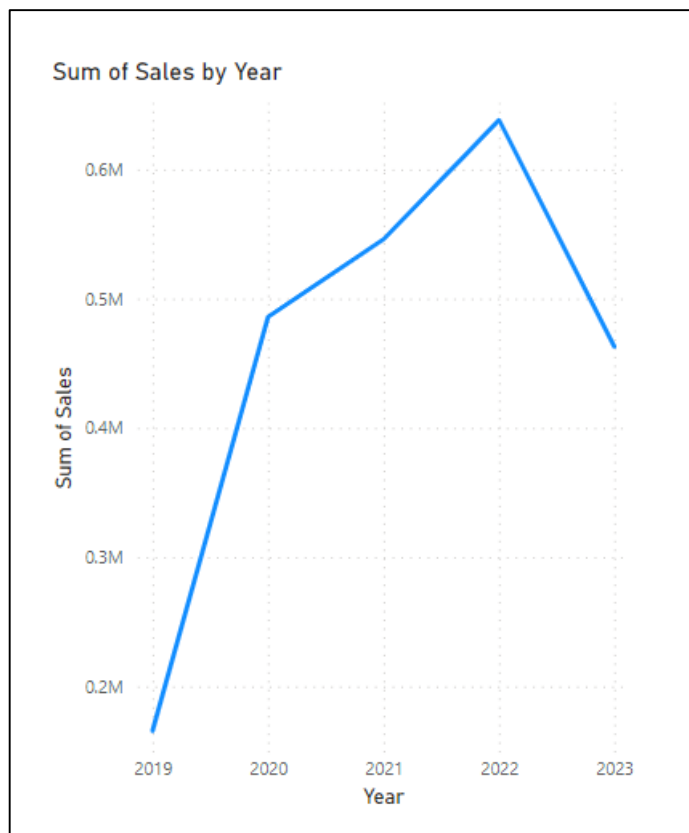
Compare the total sizes of categories, as well as the distribution of sub-categories within those categories.



# Data Visualization

## Line chart

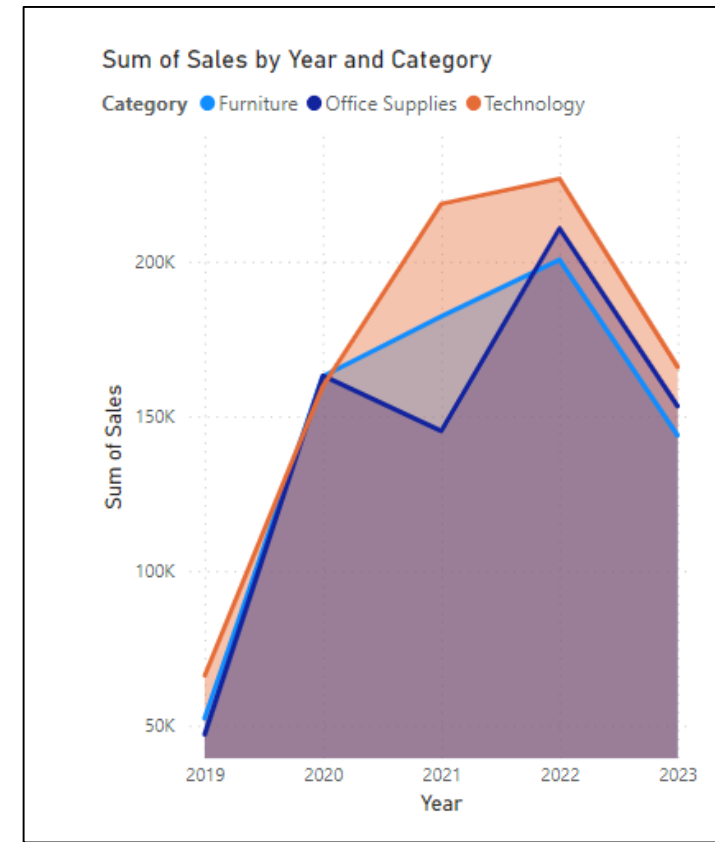
Best for displaying trends over time



# Data Visualization

## Area chart

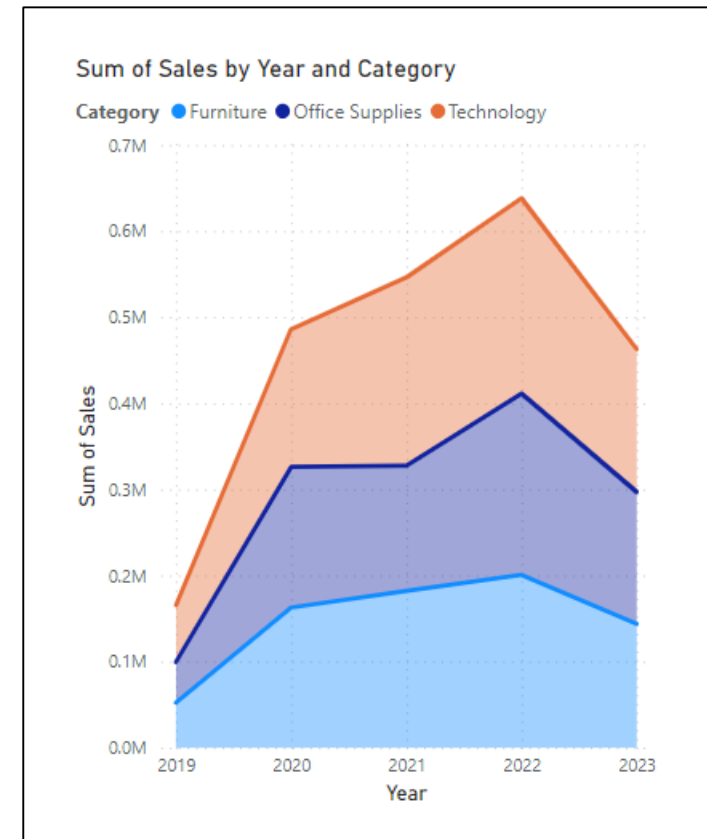
Shows changes over time



# Data Visualization

## Stacked Area chart

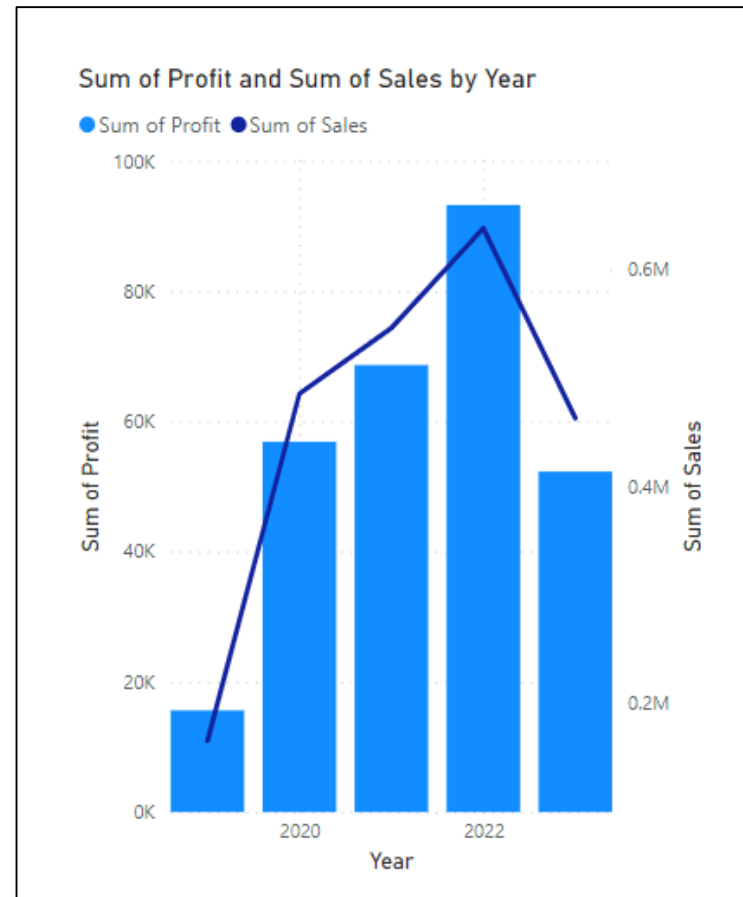
Shows part-to-whole relationships/ contribution and its change over time



# Data Visualization

Used to compare different types of data that may have different scales or units

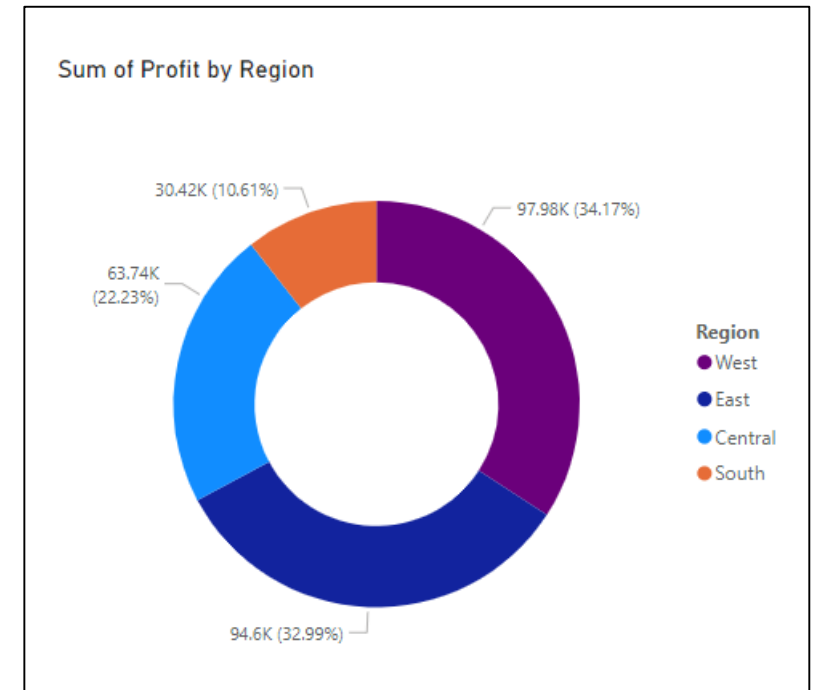
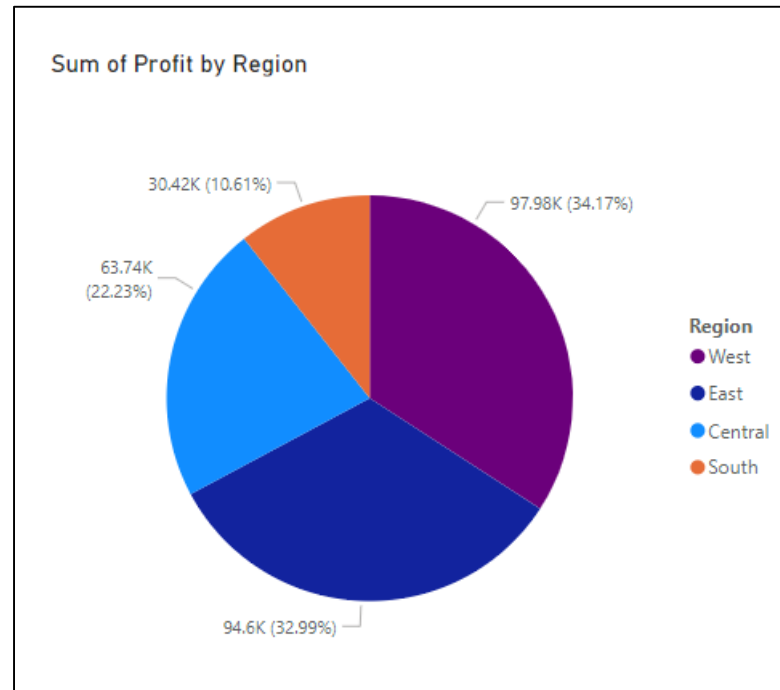
## Combo charts



# Data Visualization

## Pie and Donut

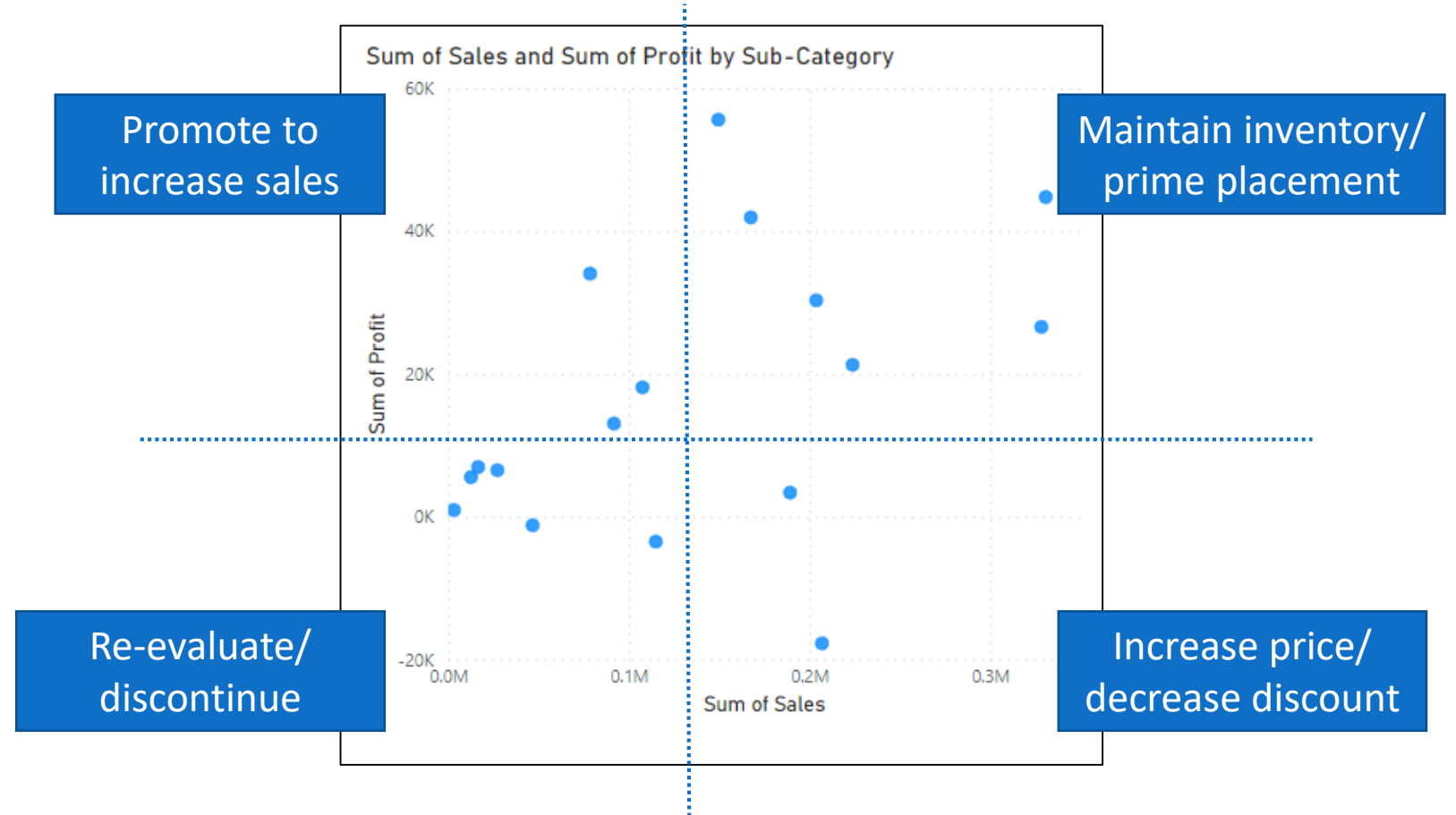
Best used when you want to compare parts of a whole.



# Data Visualization

## Scatterplot

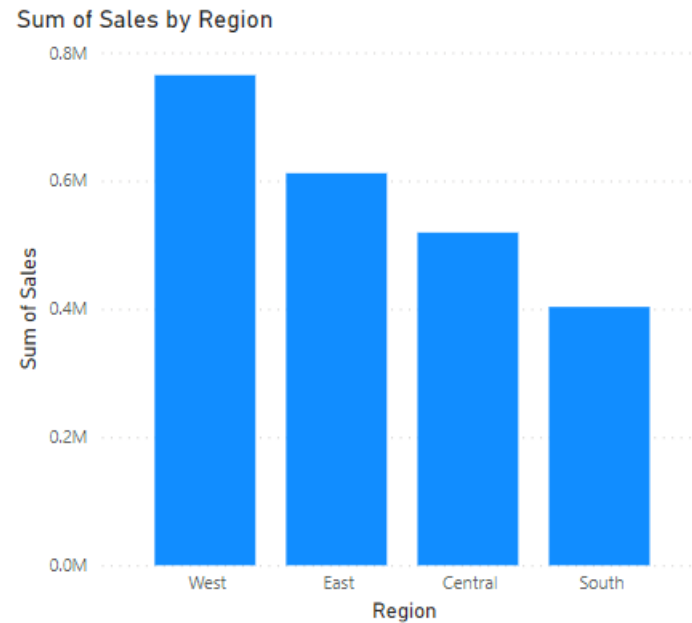
Used to display relationships between two numerical variables



# Data Visualization

## Filters

Used to control the data to be displayed



What are region-wise sales for the year 2021?

What are region-wise sales for the 'Furniture' product category?



# Data Visualization

## Filters

There are three types of filters

Visual level filters

Only applies to a particular chart/ visual

Numerical data

Page level filters

Applies to all charts on a page

Non-numerical data

Report level filters

Applies to all charts on all pages of a report

Date type data

# DAX – Data Analysis Expressions

## Section flow

Introduction to DAX



Understanding  
concepts and key terms



Explore most popular  
DAX functions



# DAX – Data Analysis Expressions

What

Formula language for analysis

Very easy and intuitive

Create new columns

Create new measures

Sales	Discount percentage	Discount value
821	0.2	164.2
894	0.1	89.4
520	0.3	156
550	0.25	137.5
351	0.2	70.2

New calculated column =  
Sales \* Discount percentage

This could have been done using Power Query as well. Why use DAX?

# DAX – Data Analysis Expressions

## What

Formula language for analysis

Very easy and intuitive

Create new columns

Create new measures

This could have been done using Power Query as well. Why use DAX?

If the calculation you want to do is simple, use any.

### Benefits of Power Query

- ✓ Data is always fresh
- ✓ Power Query GUI is easy to use

### Benefits of DAX

- ✓ Better performance
- ✓ Able to do complex calculations
- ✓ Excel-like functions and formulas

# DAX – Data Analysis Expressions

## What

Formula language for analysis

Very easy and intuitive

Create new columns

Create new measures

Aggregate numerical values  
Sum(Sales), Average(profit)

**Implicit measures**  
Automatically  
created by Power Bi

**Explicit measures**  
Created using DAX

- Can be used across different visuals and reports
- Apply complex logic and conditions
- Use them in the creation of other measures

# DAX – Data Analysis Expressions

## Syntax

Name = Calculation formula

Sales	Discount percentage	DiscountValue
821	0.2	164.2
894	0.1	89.4
520	0.3	156
550	0.25	137.5
351	0.2	70.2

DiscountValue = Sales \* Discount percentage

TotalDiscount = Sum(sales[DiscountValue])

# DAX – Data Analysis Expressions

## Row context vs Filter context

### Row context

While creating calculated  
columns

Row-wise evaluated

Sales	Discount percentage	DiscountValue
821	0.2	164.2
894	0.1	89.4
520	0.3	156
550	0.25	137.5
351	0.2	70.2

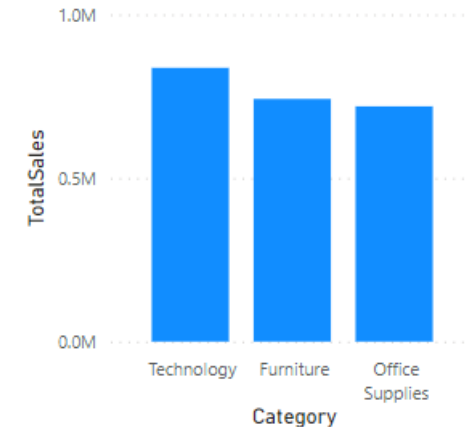
DiscountValue = Sales \*  
Discount percentage

### Filter context

While creating measures

Evaluated based on conditions  
and filters of the visual

TotalSales by Category



TotalSales = SUM('Sales Data'[Sales])

# DAX – Data Analysis Expressions

## Types of operators

Arithmetic operator	Meaning	Example
+ (plus sign)	Addition	3+3
– (minus sign)	Subtraction or sign	3–1–1
* (asterisk)	Multiplication	3*3
/ (forward slash)	Division	99/3
^ (caret)	Exponentiation	16^4

Text operator	Meaning	Example
& (ampersand)	Concatenates two values to produce one text value	[Region] & ", " & [City]



# DAX – Data Analysis Expressions

## Types of operators

Comparison operator	Meaning	Example
=	Equal to	[Region] = "USA"
==	Strict equal to	[Region] == "USA"
>	Greater than	[Sales Date] > "Jan 2009"
<	Less than	[Sales Date] < "Jan 1 2009"
>=	Greater than or equal to	[Amount] >= 20000
<=	Less than or equal to	[Amount] <= 100
<>	Not equal to	[Region] <> "USA"

# DAX – Data Analysis Expressions

## Types of operators

Text operator	Meaning	Examples
&& (double ampersand)	Creates an AND condition between two expressions	([Region] = "France") && ([BikeBuyer] = "yes"))
(double pipe symbol)	Creates an OR condition between two expressions.	(([Region] = "France")    ([BikeBuyer] = "yes"))
IN	Creates a logical OR condition	'Product'[Color] IN { "Red", "Blue", "Black" }

# DAX – Data Analysis Expressions

## Types of functions

Aggregation functions

Date and time functions

Filter functions

Financial functions

Information functions

Logical functions

Math and Trig functions

Parent and Child functions

Relationship functions

Statistical functions

Table manipulation functions

Text functions

Time intelligence functions

Power BI DAX functions reference -

<https://learn.microsoft.com/en-us/dax/dax-function-reference>

# DAX – Data Analysis Expressions

## Count functions

Function	Description
COUNT	Counts the number of rows in the column where the value is not blank.
COUNTA	Counts the number of rows in the column, regardless of data type, where the value is not blank.
COUNTAX	Counts all non-blank results when evaluating the result of an expression over a table.
COUNTBLANK	Counts the number of blank values in a column.
COUNTROWS	Counts the number of rows in a table.
COUNTX	Counts all non-blank results when evaluating an expression for each row in a table.
DISTINCTCOUNT	Counts the number of distinct values in a column.
DISTINCTCOUNTNOBLANK	Counts the number of distinct non-blank values in a column.



# DAX – Data Analysis Expressions

## RELATED

RELATED – Get values from a related table

Syntax – **Product name = RELATED('Product data'[Product\_Name])**

- ✓ There must be an active relationship
- ✓ Only for calculated columns



# DAX – Data Analysis Expressions

## CALCULATE

CALCULATE – Modify the context in which calculation is made

Syntax – **CALCULATE**(<expression>, <filter1>, <filter2>, ...)

Example

WestSales = CALCULATE([TotalSales], 'Customer Data'[Region] = "West")

- ✓ The filter context of the visualization will be ignored



# Reports

## Introduction

Collection of pages containing charts and tables

Goal

Audience

Create charts and format report as per your goals and audience

- ✓ Uniform and clean design
- ✓ Interactivity for the viewer
- ✓ Easy-to-use

Format page report, use branding

Slicers, drill-through

Navigation, Bookmarks



Hotel Count 1192

Profits 72M

Rooms 19K

Bookings 177K

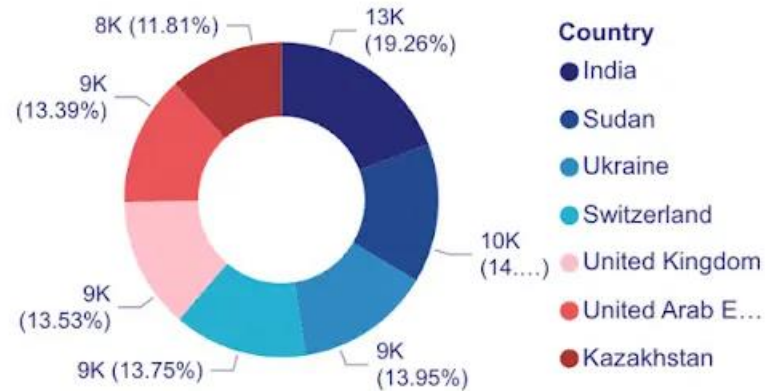
Customer 21

City 24

Country 23

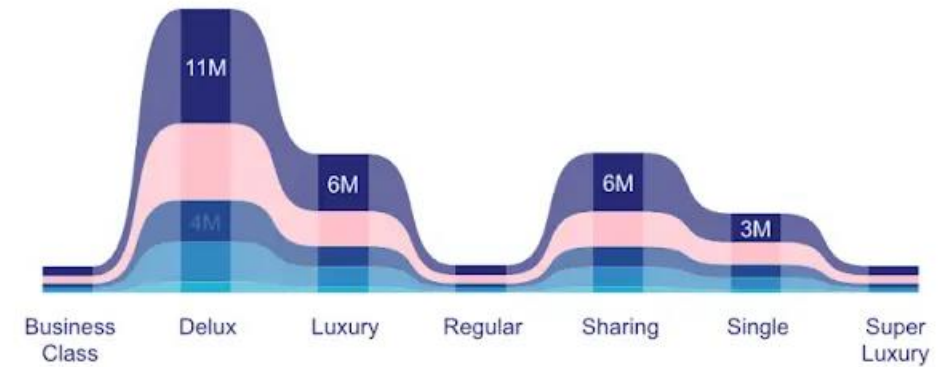
Room Type 7

Room Booked by Country



Price by Room Type and Payment Mode

Payment Mode ● Cash ● Coupon ● Credit Card ● Debit Card ● Online



Hotel Name	Room Type	City	Country	Payment Mode	Price	Room Available	Room Booked
Hotel Delight	Sharing	Baghdad	Iraq	Cash	5,64,800	17	284
Inn Eve	Delux	Reykjavik	Iceland	Cash	5,26,047	1	303
Sea View	Luxury	Madrid	Spain	Cash	4,76,270	30	525
Luxury	Delux	South Tarawa	Kiribati	Online	4,52,001	10	268
Sea View	Delux	Nur-Sultan	Kazakhstan	Online	4,10,430	16	284
Sea View	Delux	London	United Kingdom	Online	3,94,576	6	284
Cool Pool	Luxury	New Delhi	India	Cash	3,85,292	3	134
Hotel Paradise	Delux	South Tarawa	Kiribati	Cash	3,69,000	2	487
Inn Eve	Sharing	Jakarta	Indonesia	Cash	3,68,584	3	134
Total					7,22,70,834	18,556	1,77,076



ABC Apparel Company  
Online Clothing Reviews  
Last 6 Months

23,486

Reviews

1,206

Products Reviewed

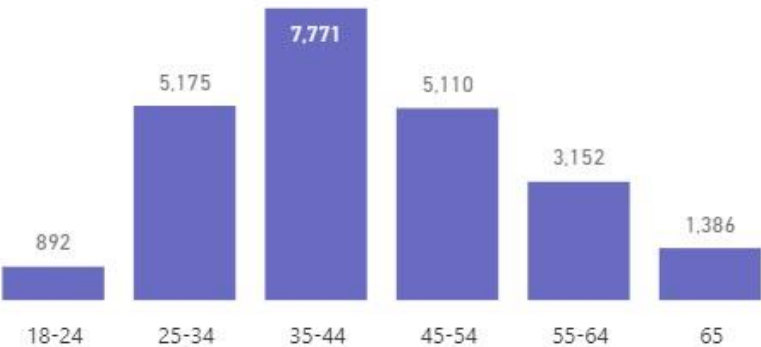
82.24%

% Recommended

4.20

Average Rating (max 5)

Reviews By Age Group



Reviews By Department

Department Name	Reviews	% Recommended
Trend	119	73.95%
Dresses	6,319	80.82%
Tops	10,468	81.52%
Jackets	1,032	83.62%
Intimate	1,735	85.01%
Bottoms	3,799	85.13%
Miscellaneous	14	100.00%
<b>Total</b>	<b>23,486</b>	<b>82.24%</b>

Reviews By Division



Ratings By Product Class





# Sales Dashboard

FISCAL YEAR

2014

2015

2016

2017



REVENUE  
\$2,297,201



PROFIT  
\$286,397



ORDERS  
5009



CUSTOMERS  
793



QUANTITY  
37,873

## SALES PERFORMANCE

REVENUE BY MONTH



## REGION PERFORMANCE

REVENUE SPLIT BY CATEGORY



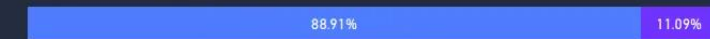
## SUB-CATEGORY RANKING

QUANTITY & SALES BY CATEGORY

Phones	Technology	3,289	\$330,007
Chairs	Furniture	2,356	\$328,449
Storage	Office Supplies	3,158	\$223,844

## PERFORMANCE RATIO

REVENUE VS PROFIT



# GLOBAL SUPERSTORE

## REVENUE AND PERFORMANCE DASHBOARD

Category

All

Segment

All

State

All

Year

All

Month

All

TOTAL REVENUE

286K



vs previous month

TOTAL ORDERS

38K



vs previous month



Region ● Central ● East ● South ● West

Region ● Central ● East ● South ● West



Category ● Furniture ● Office Supplies ● Technology





Total Volume  
IN 2014

50K

Market Share  
LAST 12 MONTHS

32.86%

Our Total Volume  
IN 2014

16K

Sentiment

68

Sentiment Gap

4

% Units Market Share vs. % Units Market Share Rolling 12 Months  
BY MONTH



Total Units Overall  
BY SEGMENT



Total Units YTD Variance %  
BY MONTH, MANUFACTURER



Total Units YTD  
BY MANUFACTURER, REGION



Total Units for 2014  
BY MONTH, MANUFACTURER



Industry % Market Share YOY Change  
BY ROLLING PERIOD, REGION

