Retail Sales & Customer Insights Dashboard

Data Cleaning and Visualization using Power BI

(Retail Sales Analysis)

Name : **Vanitha S**

Batch No : **TN_DA_FNB03**

Contact Number : **9500708068**

Email ID : **svanitha694@gmail.com**

Project Domain : **Data Analytics & Business Intelligence (Retail Sales Analysis)**

Dataset Link :

https://drive.google.com/file/d/1ROtCz3k459pPsCnkuJ0fay9FRvjClELO/view?usp=sharing

Cleaned Dataset Link :

https://docs.google.com/spreadsheets/d/14VssDGAOZtC-4jYM-VXsNX2n99QOTm95/edit?usp=sharing&ouid=109789318516480788346&rtpof=true&sd=true

Mentor Name : **Kumaran M**

Submission Date : **21st August 2025**

I. Project Description:

- This Projects demonstrates the end-to-end process of transforming a raw retails dataset into an interactive and insightful Power BI Dashboard.
- The project focuses on analysing retail sales data to generate insights into revenue, discount patterns, customer behaviour, and regional performance.
- **The dataset contains ~1000 rows and 15 columns including:**

a) **Order Information:** Order ID, Date, City, State, Region
b) **Customer Information:** First Name, Last Name, Age, Age Band
c) **Product Information:** Category, Sub-Category, Quantity, Unit Price, Discount, Discount Band
d) **Financial Metrics:** Total Sales (calculated as Quantity × Unit Price × (1 – Discount))

II. Objective:

- To clean and prepare a raw retail dataset (with missing values, inconsistencies, duplicates).
- To perform data modelling and transformations in Power Query (splitting, merging, data type corrections, creating calculated columns).
- To create interactive dashboards with KPIs, charts, and slicers in Power BI.
- To help management monitor sales performance, customer demographics, and discount impact.

III. Data Cleaning & Transformation (in Power BI):

1) Merging First Name & Last Name into Full Name:

- The dataset originally had two separate columns: *First Name* and *Last Name*.
- To improve readability and reporting, both columns were merged into a single column called Customer Name.
- Avoiding duplicate filters (instead of filtering by two columns, we use one “Full Name”).
- Making the dataset more consistent with business practices where customers are usually referred to by full name.

FirstName	LastName
Eva	Smith
Bob	Miller
Fiona	Davis
Chris	Brown
Eva	Williams
Hannah	Brown
Alice	Brown
Eva	Brown
Fiona	Davis
Chris	Brown
Bob	Jones
Fiona	Davis
Fiona	Jones
Alice	Davis
David	Miller
Chris	Johnson
Jane	Smith
Fiona	Brown
David	Miller
Jane	Johnson

Full Name
Eva Smith
Bob Miller
Fiona Davis
Chris Brown
Eva Williams
Hannah Brown
Alice Brown
Eva Brown
Fiona Davis
Chris Brown
Bob Jones
Fiona Davis
Fiona Jones
Alice Davis
David Miller
Chris Johnson
Jane Smith
Fiona Brown
David Miller
Jane Johnson

2) Removing Leading & Extra Spaces in Region Column:

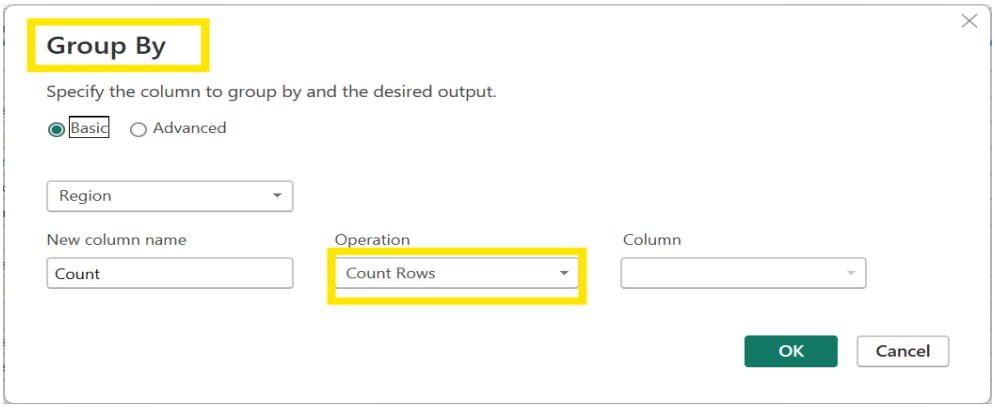
- The Region column contained extra spaces at the beginning or between words (e.g., " East " or "South ").
- These unnecessary spaces can cause mismatches while filtering or grouping.
- Using **Trim** and **Clean** functions in Power Query, extra spaces were removed to ensure consistency.
- Example: " East " → "East".
- These steps made the Region column clean, consistent, and ready for accurate grouping and visualization in Power BI.

Region
west
West
Central
West
central
South
East
South
Central
East
Central
Central
West
East
West
EAST
Central
CENTRAL
South

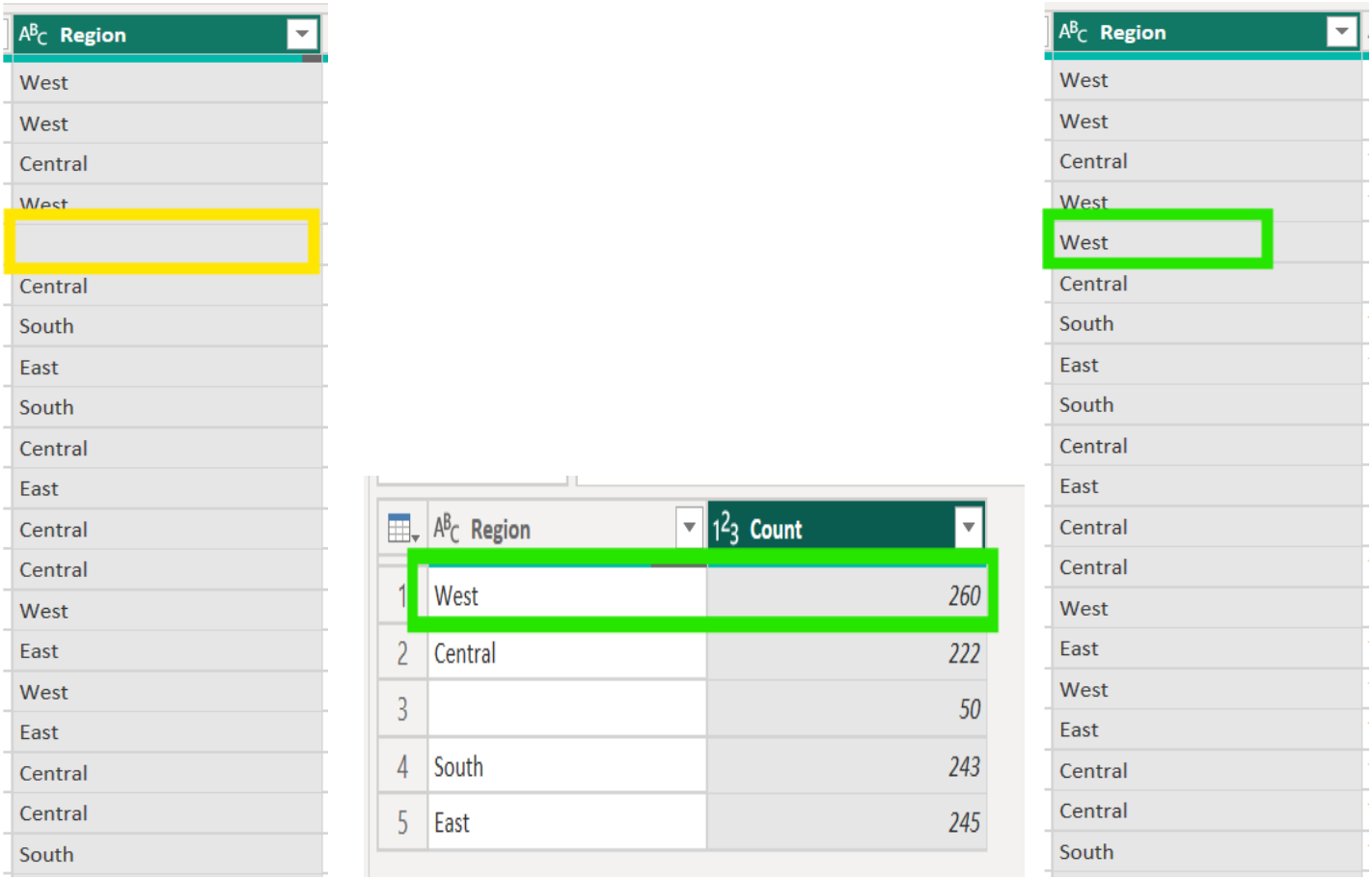
Region
West
West
Central
West
Central
South
East
South
Central
East
Central
Central
West
East
West
East
Central
Central
South

3) <u>Handling Missing Data in Region Column (Using Group By – Mode):</u>

- The Region column had some missing values. Since region is a categorical field, missing values can affect grouping and filtering in reports.
- To resolve this, a **Group By operation** was performed in Power Query:
    - Data was grouped by Region. For each group, the **most frequently occurring Region (Mode)** was calculated.



- The missing Region values were then replaced with this Mode.



4) <u>Correcting Spelling Mistakes in Category Column:</u>

- While reviewing the dataset, a spelling inconsistency was found in the *Category* column:

  "Furniture" was incorrectly written as "Furnture".

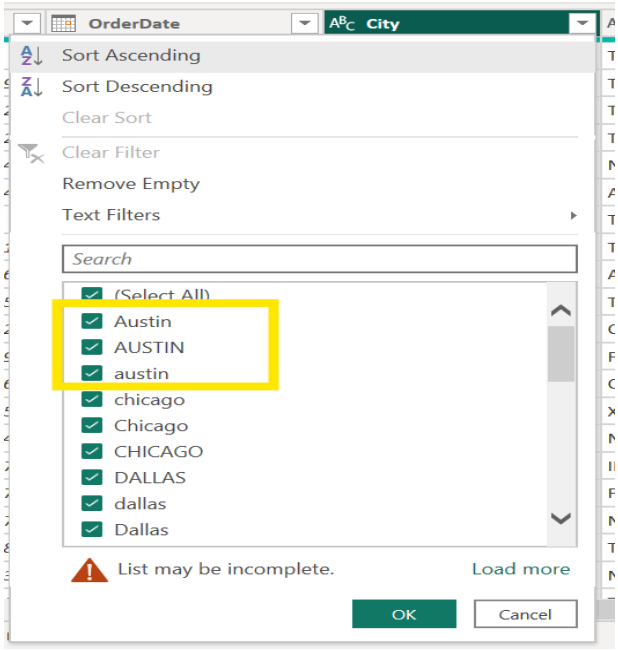
- To ensure consistency and accuracy in analysis, the **Replace Values** option in Power Query was used:

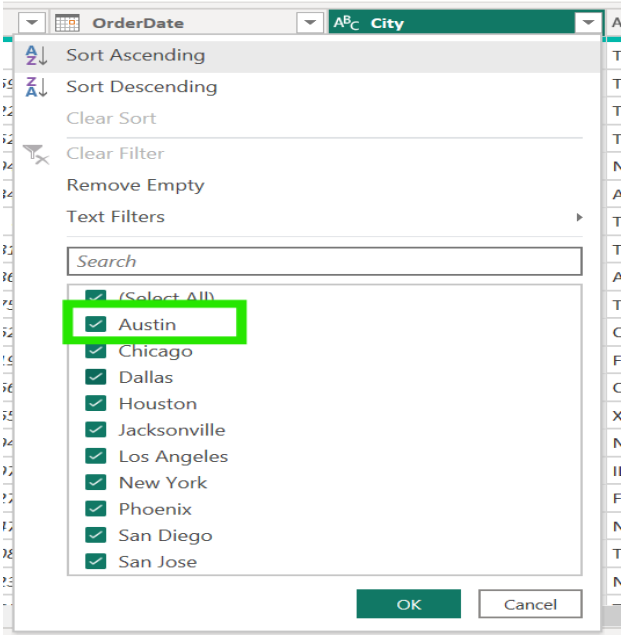
  "Furnture" → replaced with "Furniture".

- This prevents data from splitting into multiple categories during analysis and ensures all sales related to Furniture are grouped correctly.

5) <u>Standardizing City Column (Case Consistency):</u>

- The City column contained inconsistent text formats:
  - Some entries were in **UPPERCASE** (e.g., "AUSTIN").
  - Some in **lowercase** (e.g., "chicago").
- To fix this inconsistency, the **Text.Proper** function in Power Query was applied.
- This converted all city names into a consistent format with the first letter capitalized:
  - "AUSTIN" → "Austin"
  - "chicago" → "Chicago"
- Standardizing the City column improves readability, prevents duplicates caused by case differences, and ensures accurate grouping in reports.



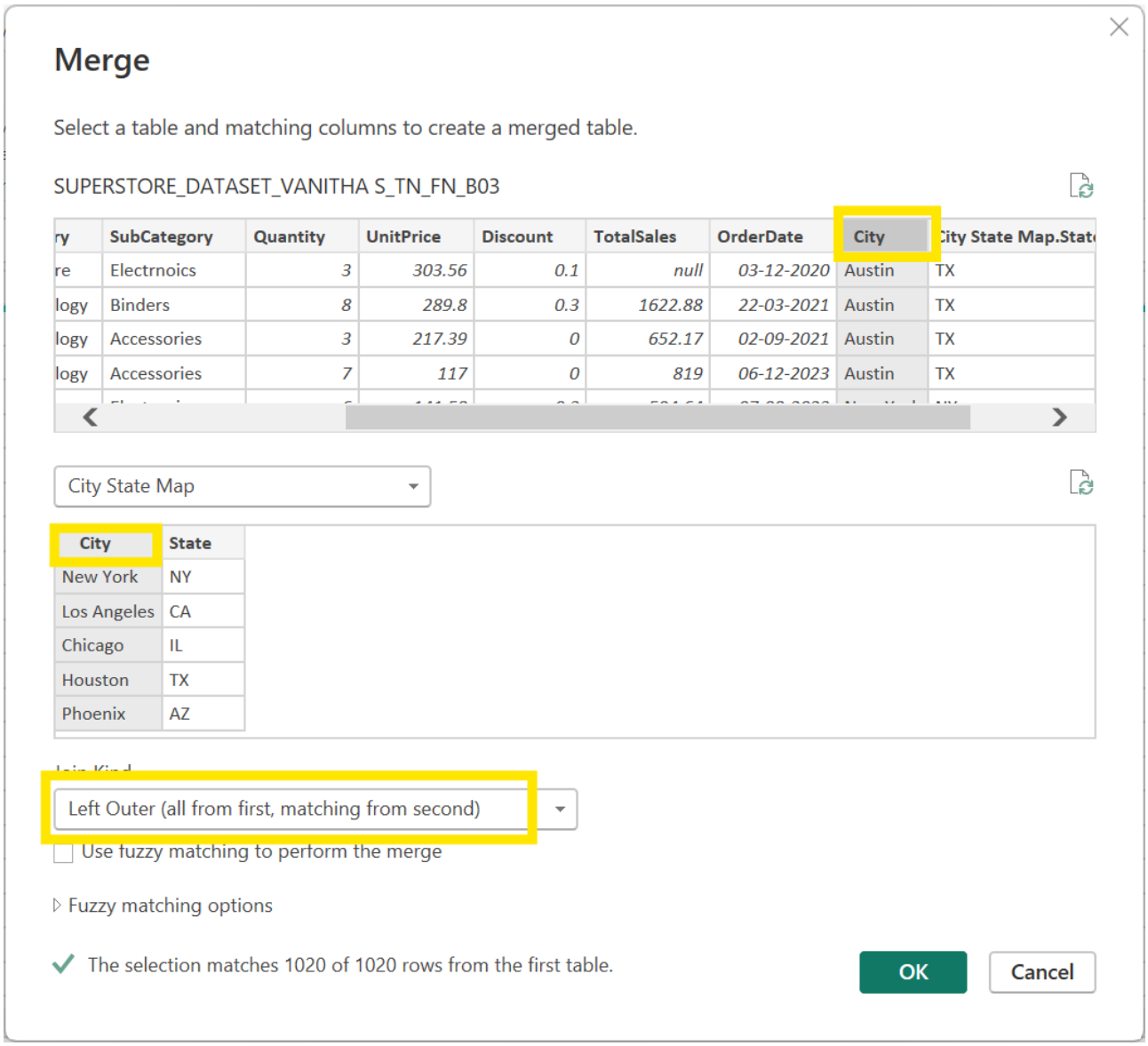


6) Correcting State & City Mismatch (Using Joins):

- During data validation, mismatches were found between the *State* and City columns.
  - Example: "Los Angeles" city incorrectly mapped to "XX" state.
- To correct this, a **reference mapping table** (State–City master table) was created.
- Using **Joins in Power Query**:
  - The main dataset was merged with the reference table based on City.
  - The correct State value from the reference table replaced incorrect entries.
- Example:

a) Before: City = "Los Angeles", State = "XX"
b) After: City = "Los Angeles", State = "CA"

- Ensures geographical accuracy in analysis, improves grouping by State/City, and avoids misleading insights in regional dashboards.

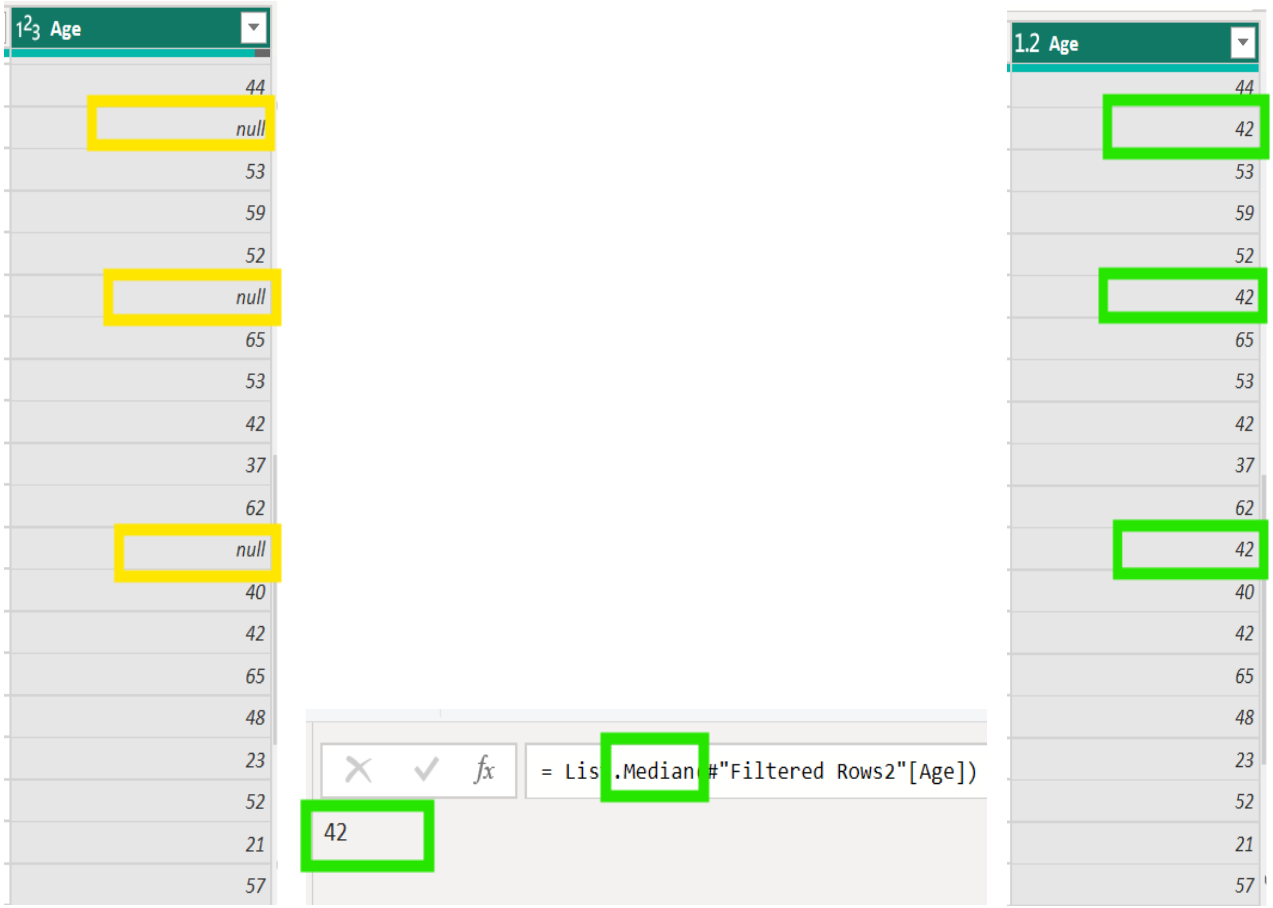


City	State
Austin	TX
Houston	TX
Austin	TX
Austin	TX
New York	NY
Phoenix	AZ
Austin	TX
Houston	TX
Phoenix	AZ
Houston	TX
Los Angeles	CA
Jacksonville	FL
Los Angeles	CA
Los Angeles	XX
New York	NY
Chicago	IL
Jacksonville	FL
New York	NY
Houston	TX
New York	NY

City	City State Map.State
Austin	TX
Austin	TX
Austin	TX
Austin	TX
New York	NY
Houston	TX
Houston	TX
Houston	TX
Los Angeles	CA
Phoenix	AZ
Phoenix	AZ
Jacksonville	FL
Los Angeles	CA
Los Angeles	CA
New York	NY
Chicago	IL
Jacksonville	FL
New York	NY
Houston	TX
New York	NY

7) Handling Missing Data in Age Column (Using Median):

- The Age column contained some missing values. Missing values in age can impact demographic analysis.
- To handle this, the **Median** of the existing Age values was calculated.
    - Median is preferred over Average because it is less affected by extreme values (outliers).
- The missing Age values were then replaced with this Median value.
- Example:
    - Age values present: [18, 19, 20, ….., 63,64,65] → Median = **42**
    - Missing values replaced with **42**.
- Maintains data consistency while ensuring the distribution of ages is not skewed by very high or low values.



8) Handling Missing Data in Quantity Column (Using Formula):

- The Quantity column had missing values, which are critical for calculating sales and analyzing product performance.
- Since Total Sales is calculated from Quantity, Unit Price, and Discount, the missing Quantity was derived using the reverse formula:

➢ **Formula: Quantity_Filled =**

IF(ISBLANK('SUPERSTORE_DATASET'[Quantity]),'SUPERSTORE_DATASET'[TotalSales]/ ('SUPERSTORE_DATASET'[UnitPrice] * (1 - 'SUPERSTORE_DATASET'[Discount])),'SUPERSTORE_DATASET'[Quantity])

- This approach ensures that missing quantities are logically calculated instead of guessed, keeping the dataset mathematically consistent.

```
1 Quantity_Filled = IF(ISBLANK('SUPERSTORE_DATASET_VANITHA S_TN_FN_B03'[Quantity]),'SUPERSTORE_DATASET_VANITHA S_TN_FN_B03'
[TotalSales]/('SUPERSTORE_DATASET_VANITHA S_TN_FN_B03'[UnitPrice]*(1-'SUPERSTORE_DATASET_VANITHA S_TN_FN_B03'[Discount])),
'SUPERSTORE_DATASET_VANITHA S_TN_FN_B03'[Quantity])
```

egory	Quantity	UnitPrice	Discount	TotalSales	OrderDate	City	Age	State	Quantity_Filled	Unit Price_Filled
	5		0.20	941.36	26-04-2023	Austin	27	TX	5	150
		11.77	0.10	21.19	17-08-2020	New York	44	NY	2	11
	6	45.39	0.10	245.11	07-07-2020	San Jose	42	CA	6	45
	8	133.00	0.00	1064.00	09-12-2023	Austin	53	TX	8	133
ries	4	18.79	0.30	52.61	10-01-2020	San Diego	59	CA	4	18
	2	317.53	0.10	571.55	11-02-2023	New York	52	NY	2	317
	5	216.07	0.30	756.24	04-01-2023	Houston	42	TX	5	216
ries	2	276.33	0.10	497.39	06-11-2021	Jacksonville	65	FL	2	276
	2	91.45	0.10	164.61	16-05-2021	New York	53	NY	2	91
	6	151.49	0.00	908.94	25-03-2020	Houston	42	TX	6	151
ries	9	333.59	0.20	2401.85	18-02-2020	Chicago	37	IL	9	333
		482.82	0.20	1545.02	17-04-2022	Los Angeles	62	CA	4	482
	6	29.96	0.10	161.78	04-10-2022	New York	42	NY	6	29
		445.74		3120.18	19-11-2021	Dallas	40	TX	7	445
ics	6	290.56	0.30	1220.35	21-01-2021	San Diego	42	CA	6	290

9) Handling Missing Data in Unit Price Column (Using Formula):

- The Unit Price column had missing values, which are essential for financial calculations and profitability analysis.
- Since Total Sales is calculated from Quantity, Unit *Price, and* Discount, the missing Unit Price was derived using the reverse formula:
  - **Formula: Unit Price_Filled =**

    if(isblank('SUPERSTORE_DATASET'[UnitPrice]),'SUPERSTORE_DATASET'[TotalSales]/'SUPERSTORE_DATASET'[Quantity_Filled]*(1-'SUPERSTORE_DATASET'[Discount]),'SUPERSTORE_DATASET'[UnitPrice])
- This ensures that missing Unit Price values are computed accurately from existing sales data, maintaining consistency in revenue and margin analysis.

```
Unit Price_Filled = if(isblank('SUPERSTORE_DATASET'[UnitPrice]),'SUPERSTORE_DATASET'[TotalSales]/'SUPERSTORE_DATASET'
[Quantity_Filled]*(1-'SUPERSTORE_DATASET'[Discount]),'SUPERSTORE_DATASET'[UnitPrice])
```

ory	Quantity	UnitPrice	Discount	TotalSales	OrderDate	City	Age	State	Quantity_Filled	Unit Price_Filled
	9	305.53	0.20	2199.82	12-01-2023	Chicago	64	IL	9	305.53
	6	197.42	0.10	1066.07	20-10-2023	New York	60	NY	6	197.42
	9		0.30	780.88	04-06-2022	Los Angeles	44	CA	9	60.74
	5	52.01	0.00	260.05	02-07-2021	Dallas	22	TX	5	52.01
	7	160.88	0.30	788.31	27-09-2023	Phoenix	23	AZ	7	160.88
	4	52.84	0.30	147.95	18-04-2022	Chicago	30	IL	4	52.84
		92.56	0.10		01-12-2023	Austin	40	TX		92.56
	1	493.73	0.00	493.73	19-12-2023	San Jose	59	CA	1	493.73
	7	224.83	0.00	1573.81	21-01-2023	Chicago	42	IL	7	224.83
	8	268.28	0.30	1502.37	27-04-2022	Los Angeles	63	CA	8	268.28
	5	437.44	0.10	1968.48	05-09-2021	San Jose	21	CA	5	437.44
	2	497.81	0.00		15-02-2021	Phoenix	49	AZ	2	497.81
	5	293.44	0.30	1027.04	07-03-2021	Austin	61	TX	5	293.44
	6	407.17	0.00	2443.02	08-11-2021	New York	42	NY	6	407.17
s	2		0.10	300.31	07-06-2020	Jacksonville	43	FL	2	135.14
s	6	156.34	0.10	844.24	14-01-2020	Chicago	59	IL	6	156.34
	6	204.34	0.10	1103.44	07-01-2021	Phoenix	28	AZ	6	204.34
s	1	337.89	0.00	337.89	12-08-2020	Los Angeles	26	CA	1	337.89
s	1	342.68	0.20	274.14	01-01-2023	Chicago	28	IL	1	342.68

10) Handling Missing Data in Discount Column (Using Formula):

- The *Discount* column had missing values. Since discounts directly affect revenue, it's important to restore them logically instead of leaving blanks.
- Using the sales formula, missing Discount values were derived:
  - **Discount_Filled = if(isblank(SUPERSTORE_DATASET[Discount]),1-(SUPERSTORE_DATASET[TotalSales]/(SUPERSTORE_DATASET[Quantity_Filled]*SUPERSTORE_DATASET[Unit Price_Filled])), SUPERSTORE_DATASET[Discount])**
- This method ensures missing Discount values are reconstructed mathematically, keeping the dataset internally consistent and avoiding biased analysis.

```
1 Discount_Filled = if(isblank(SUPERSTORE_DATASET[Discount]),1-(SUPERSTORE_DATASET[TotalSales]/(SUPERSTORE_DATASET[Quantity_Filled]
  *SUPERSTORE_DATASET[Unit Price_Filled])),SUPERSTORE_DATASET[Discount])
```

tity	UnitPrice	Discount	TotalSales	OrderDate	City	Age	State	Quantity_Filled	Unit Price_Filled	Discount_Filled	To
7	348.24	0.10	2193.91	28-05-2023	Austin	38	TX	7	348.24	0.10	
3	58.48	0.20	140.35	21-04-2022	Austin	39	TX	3	58.48	0.20	
1	182.52	0.20	146.02	18-01-2022	Austin	63	TX	1	182.52	0.20	
8	172.96	0.00	1383.68	05-12-2021	Austin	61	TX	8	172.96	0.00	
6	85.15		510.90	15-08-2023	Austin	40	TX	6	85.15	0.00	
9	203.35	0.30	1281.10	31-01-2023	Austin	58	TX	9	203.35	0.30	
6	196.70	0.30	826.14	23-05-2023	Austin	23	TX	6	196.70	0.30	
8	85.52	0.00	684.16	21-10-2022	Austin	24	TX	8	85.52	0.00	
5	108.37	0.00	541.85	25-02-2020	Austin	55	TX	5	108.37	0.00	
6	44.24	0.00	265.44	10-09-2021	Austin	45	TX	6	44.24	0.00	
2	470.74		659.04	21-09-2023	Austin	22	TX	2	470.74	0.30	
9	485.62	0.20	3496.46	20-06-2023	Austin	21	TX	9	485.62	0.20	
6	362.42	0.20	1739.62	23-04-2021	Austin	42	TX	6	362.42	0.20	
1	205.68	0.20	164.54	10-04-2020	Austin	57	TX	1	205.68	0.20	
7	243.23	0.20	1362.09	18-12-2022	Austin	26	TX	7	243.23	0.20	
8	408.12	0.30	2285.47	13-08-2023	Austin	65	TX	8	408.12	0.30	
4	143.77		460.06	29-07-2020	Austin	56	TX	4	143.77	0.20	
5	70.43	0.10	316.94	20-02-2023	Austin	39	TX	5	70.43	0.10	
6	451.69	0.30	1897.10	26-09-2020	Austin	23	TX	6	451.69	0.30	
8	279.57	0.20	1789.25	23-03-2022	Austin	28	TX	8	279.57	0.20	

11) Handling Missing Data in Total Sales Column (Using Formula):

- The Total Sales column is the core measure for revenue analysis. A few missing values were identified.
- Since Total Sales depends on Quantity, Unit Price, and Discount, missing values were recalculated using the standard formula:
  - **Total_sales = if(isblank(SUPERSTORE_DATASET[TotalSales]), SUPERSTORE_DATASET[Quantity_Filled]*SUPERSTORE_DATASET[Unit Price_ Filled]*(1-SUPERSTORE_DATASET[Discount_Filled]), SUPERSTORE_DATA SET [TotalSales])**
- This ensures no blank sales values remain, and the dataset maintains perfect financial consistency.

```
1 Total_sales = if(isblank(SUPERSTORE_DATASET[TotalSales]),SUPERSTORE_DATASET[Quantity_Filled]*SUPERSTORE_DATASET[Unit
  Price_Filled]*(1-SUPERSTORE_DATASET[Discount_Filled]),SUPERSTORE_DATASET[TotalSales])
```

ce	Discount	TotalSales	OrderDate	City	Age	State	Quantity_Filled	Unit Price_Filled	Discount_Filled	Total_sales
06.02		593.71	22-01-2021	Austin	21	TX	8	106.02	0.30	593.71
59.38	0.10		28-02-2020	Austin	45	TX	3	59.38	0.10	160.33
61.47	0.30	791.20	08-03-2023	Austin	29	TX	7	161.47	0.30	791.20
24.20	0.30	434.70	14-03-2021	Austin	33	TX	5	124.20	0.30	434.70
77.85	0.20	249.12	27-04-2020	Austin	32	TX	4	77.85	0.20	249.12
52.50	0.20	854.00	23-06-2022	Austin	30	TX	7	152.50	0.20	854.00
	0.30	702.58	10-03-2023	Austin	41	TX	6	81.97	0.30	702.58
32.25	0.10	833.18	25-08-2022	Austin	60	TX	7	132.25	0.10	833.18
75.56	0.00	2253.36	06-12-2022	Austin	44	TX	6	375.56	0.00	2253.36
64.90	0.00	2119.20	29-08-2023	Austin	53	TX	8	264.90	0.00	2119.20
87.83	0.20	1170.79	04-11-2020	Austin	65	TX	3	487.83	0.20	1170.79
49.21	0.00	1796.84	09-10-2023	Austin	47	TX	4	449.21	0.00	1796.84
64.09	0.30	403.77	12-10-2020	Austin	42	TX	9	64.09	0.30	403.77
48.63	0.00	145.89	30-09-2021	Austin	33	TX	3	48.63	0.00	145.89
54.87	0.00		30-03-2021	Austin	59	TX	5	154.87	0.00	774.35
02.29	0.10		06-02-2021	Austin	35	TX	2	302.29	0.10	544.12
35.01	0.00	670.02	03-03-2022	Austin	65	TX	2	335.01	0.00	670.02
03.56	0.20	2582.78	29-04-2023	Austin	60	TX	8	403.56	0.20	2582.78
61.71	0.20	1846.84	11-11-2022	Austin	33	TX	5	461.71	0.20	1846.84

12) <u>Business Interpretation of Discount Band:</u>

From the created **Discount Band** column, the dataset can now be analyzed across different discount levels.

- **No Discount (=0)** → Generally associated with **higher profit margins** while still maintaining stable sales.
- **Low Discount (5% – 15%)** → Provides a balance between **sales volume and profitability**.
- **High Discount (> 15%)** → Attracts **higher sales volume** but reduces overall **profit margins** significantly.

☞ This categorization allows decision-makers to understand the **trade-off between discount strategy and profitability**, and to identify which discount band drives the **best business outcomes**.

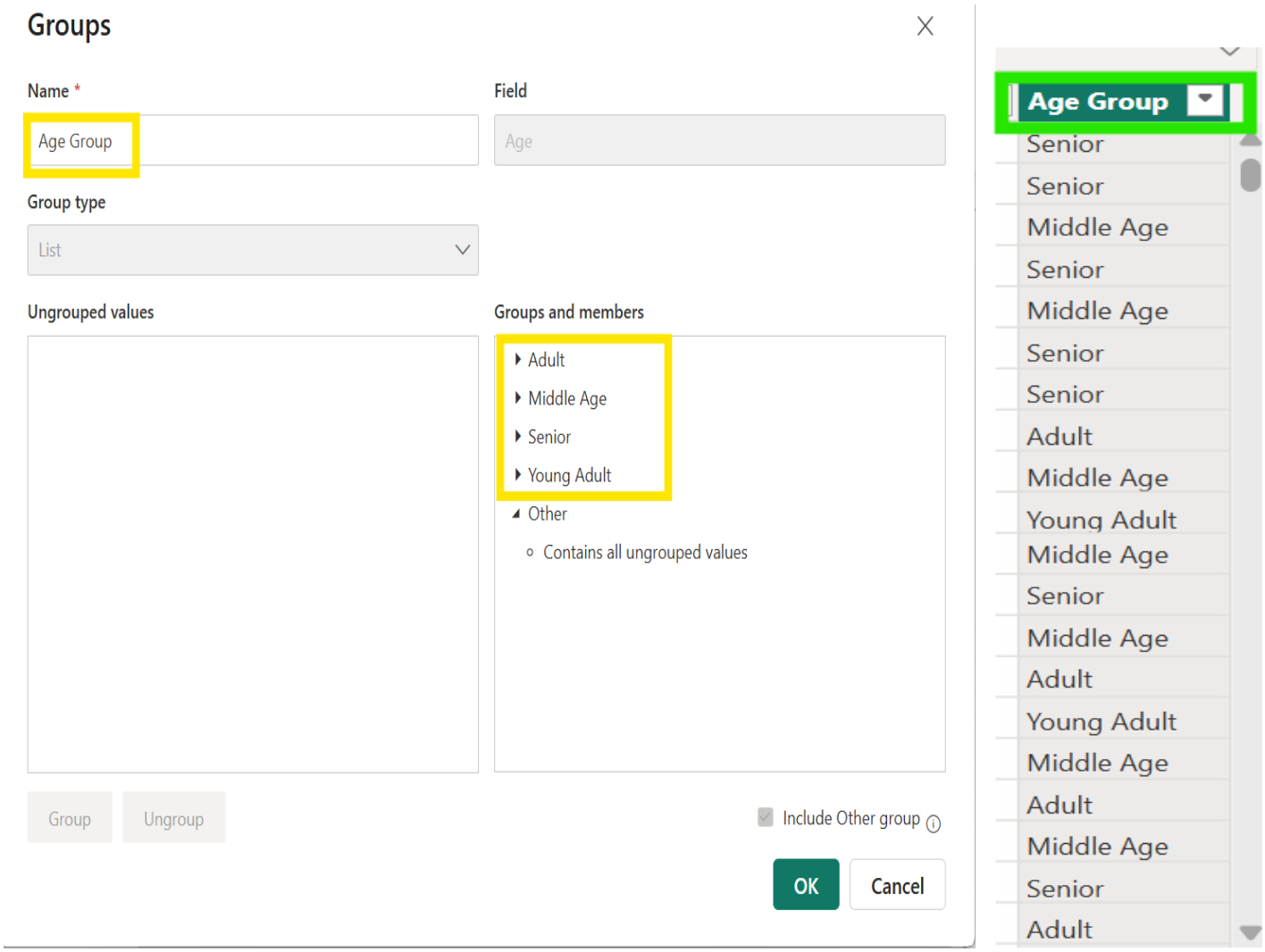
```
= Table.AddColumn(#"Changed Type1", "Discount Band", each if [Discount] = 0 then "No Discount" else
    if [Discount] < 0.15 then "Low Discount" else "High Discount")
```

	City	State	Age	Discount Band
03-12-2020	Austin	TX	28	Low Discount
22-03-2021	Austin	TX	45	High Discount
02-09-2021	Austin	TX	44	No Discount
06-12-2023	Austin	TX	44	No Discount
07-08-2023	New York	NY	36	High Discount
02-05-2020	Houston	TX	43	High Discount
06-09-2021	Houston	TX	37	No Discount
07-10-2020	Houston	TX	33	High Discount
02-12-2022	Los Angeles	CA	56	High Discount
01-11-2021	Phoenix	AZ	62	Low Discount
15-06-2022	Phoenix	AZ	20	Low Discount
18-08-2021	Jacksonville	FL	61	No Discount
03-12-2021	Los Angeles	CA	46	High Discount
10-12-2022	Los Angeles	CA	32	Low Discount
28-07-2020	New York	NY	47	High Discount
05-05-2020	Chicago	IL	57	High Discount
06-05-2021	Jacksonville	FL	18	No Discount

13) Creating Age Groups (Listing in Power BI):

To analyze customer demographics more effectively, the **Age** column was categorized into **Age Groups** using the Group/Listing feature in Power BI.

- **Steps:**
  1. Select the **Age** column in Power BI.
  2. From the Modeling tab, choose **New Group**.
  3. Define ranges (bins) for age values.
- **Logic Used (Example):**
  - Age 18–25 → **Young Adult**
  - Age 26–35 → **Adult**
  - Age 36–50 → **Middle Age**
  - Age51- 65 → **Senior**
- Provides insights into which **age segments contribute the most to sales**.
- Helps in **targeted marketing and customer segmentation**.
- Makes visualizations more meaningful (e.g., comparing sales between Young vs. Senior customers).



14) Creating Measure in Power BI:

a) Count of orders:

- To analyze the number of unique orders in the dataset, a **measure** was created using the **DISTINCTCOUNT()** function.

  - **DAX Formula:**

Count of Orders = DISTINCTCOUNT(SUPERSTORE_DATASET[OrderID])

- **Explanation:**
  - The DISTINCTCOUNT() function ensures that duplicate order IDs are not counted multiple times.
  - This provides an **accurate count of unique orders** rather than simply counting rows.
- **Benefit:**
  - Helps in analyzing **order volume** across customers, products, regions, and discount bands.

b) Average Discount:

❖ To analyze the discount behavior across products and customers, a **measure** was created to calculate the **average discount**.

- **DAX Formula:**

Average Discount = AVERAGE(SUPERSTORE_DATASET[Discount_Filled])

❖ **Explanation:**

  - The AVERAGE() function calculates the mean of all discount values in the dataset.
  - This provides insight into the **typical discount percentage** offered across transactions.

❖ **Benefit:**

  - Helps compare **average discounts** across regions, product categories, or customer segments.
  - Useful for identifying **over-discounting** that may reduce profit margins.
  - Supports visualizations such as **Discount Band Analysis** and **Region-wise Discount Comparison**.

c) Total Sales:

❖ To evaluate overall business performance, a **measure** was created to calculate the **Total Sales** from each order.

- **DAX Formula:**

Total Sales = Sum(SUPERSTORE_DATASET[Total_sales])

❖ **Explanation:**

  - The SUM() function adds up all sales values from the dataset.
  - This provides the **total revenue generated** across all orders.

❖ **Benefit:**

  - Forms the **core KPI** for financial analysis.
  - Useful for comparisons such as **Sales by Region, Sales by Product Category, or Sales by Customer Segment**.

IV. Data Visualisation:

1.Geographical Data Standardization using City-State Lookup Table:

"To remove mismatches between City and State values, a lookup table was created with unique City-State combinations and joined with the Orders dataset in Power BI Model View. This ensured consistency and improved data accuracy in geographical analysis."

Step 1: Create City-State Table

- Make a **separate table** with unique combinations of **City** and **State**.

*(*No duplicates, only unique values)

Step 2: Load into Power BI

- Load this **City-State table** into Power BI.
- Ensure that both Orders[City] and Orders[State] columns match the lookup table

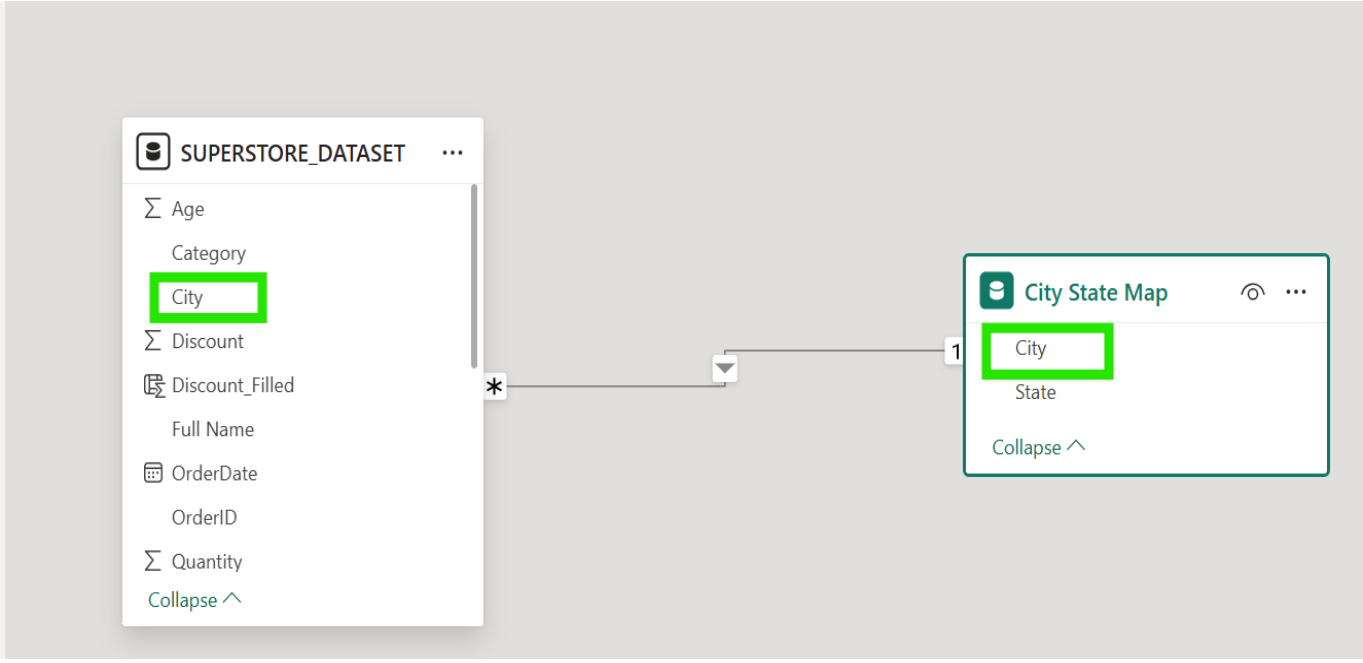
Step 3: Create Relationships

- Go to **Model View** in Power BI.
- Create a relationship:
  - Orders[City] → CityState[City]
  - Or Orders[State] → CityState[State]

*(*Better: use *City* as primary key if unique per state, else use composite key *City* + *State*.)

Step 4: Use in Visuals

- Now, you can use the **cleaned City-State data** from the lookup table in visuals.
- This ensures **consistency** (no spelling errors, no mismatched states).



2. Sales Trend Analysis (Line Chart):

A **Line Chart** was created to analyze the **sales trend over time**.

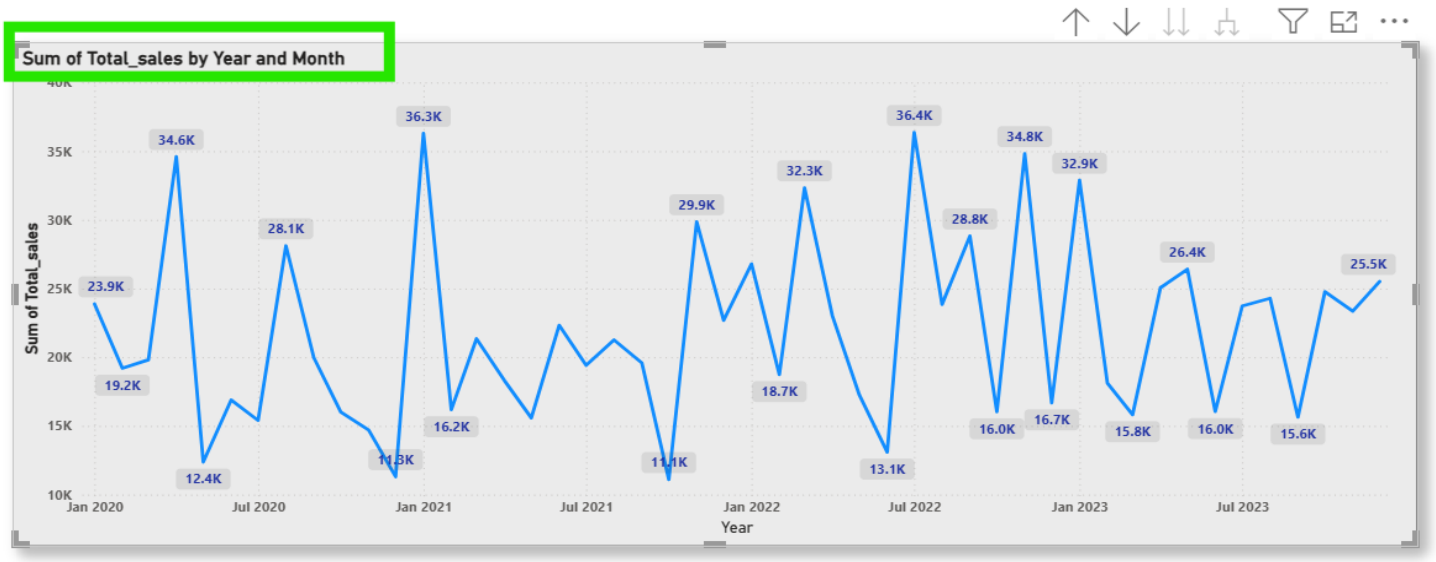
- **Axis (X-Axis):** Order Date (Year, Month hierarchy)
- **Values (Y-Axis):** Total Sales (using the SUM(Sales) measure)

Explanation:

- The line chart helps to observe how **sales fluctuate over different months and years**.
- It provides insights into:
  - **Seasonal patterns** (example: higher sales during festive seasons or year-end).
  - **Growth trends** over multiple years.
  - **Business performance monitoring** month-on-month.

Business Use:

- Managers can identify **peak months** to plan inventory and marketing strategies.
- Helps in detecting **downward trends** early to take corrective actions.



Insights: In 2023, sales peaked in December, showing festive season impact.

3. Sales by Product Sub-Category (Clustered Column Chart):

In this visualization, a **Clustered Column Chart** was used to show **Total Sales** across different **Product Sub-Categories**, broken down by **Category**.

- **Axis (X-Axis):** Sub-Category
- **Values (Y-Axis):** Total Sales (SUM(Total Sales))
- **Legend (Color Split):** Category (Furniture, Office Supplies, Technology)

Explanation:

- This chart shows how each **sub-category contributes to the overall sales**.
- Using **Category as legend**, we can compare sub-categories within each product category (e.g., Technology, Furniture, Office Supplies).
- It highlights **top-performing and low-performing products**.

Business Use:

- Identifies **top-performing products** that contribute significantly to revenue.
- Pinpoints **weaker sub-categories** where strategies like discounts, promotions, or product rebranding may be needed.
- Supports **inventory and marketing decisions** by focusing on profitable sub-categories.

Insights:

- **Binders (75K)**, **Chairs (72K)**, and **Phones (66K)** recorded the **highest sales** among all sub-categories.
- **Electronics (9K–25K)** and **Accessories (43K–61K)** are the **lowest-performing sub-categories**.
- The use of category as legend helps us compare **within-category performance**.
- Binders and Chairs lead sales across categories, while Electronics show significantly lower sales. This suggests potential improvement strategies for Electronics and Accessories, while continuing to strengthen sales in top-performing categories like Furniture and Technology.

4. Total Sales by Region and Category:

This visualization uses a **Clustered Column Chart** to show the distribution of **Total Sales** across different **Regions (West, South, Central, East)**, segmented by **Product Category**.

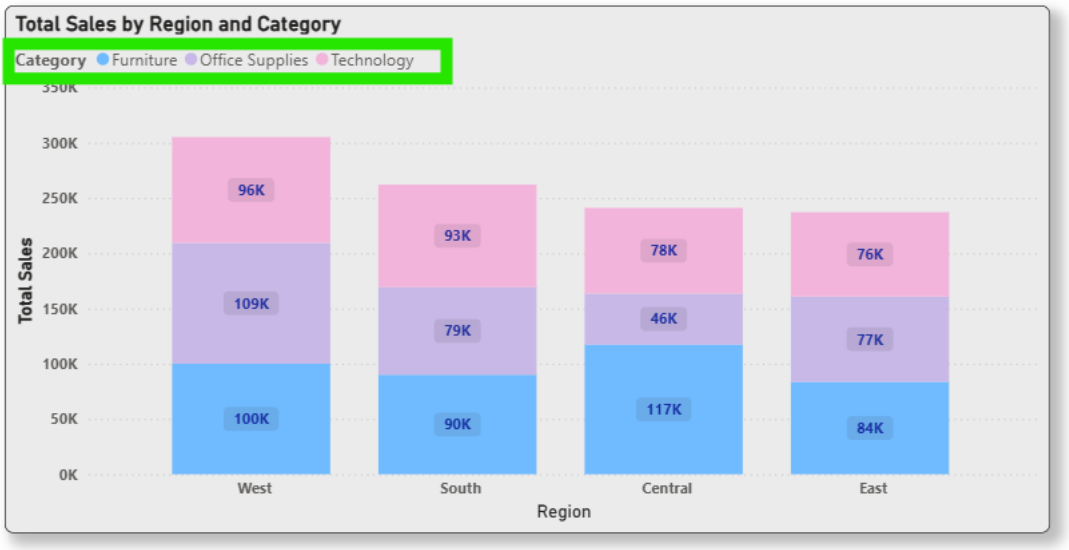
- **Axis (X-Axis):** Region
- **Values (Y-Axis):** Total Sales (SUM(Total Sales))
- **Legend (Color Split):** Category (Furniture, Office Supplies, Technology)

Business Value:

- Highlights **regional strengths** (e.g., West leading overall, Central excelling in Technology).
- Identifies **weaker regions (East)** where marketing and sales strategies may need improvement.
- Supports **region-wise decision-making** for promotions, discounts, and inventory planning.

Insights:

- The **West Region (305K)** recorded the **highest total sales**, with strong contributions from all categories.
- The West region dominates overall sales, while Central leads in Furniture sales. The East region lags behind, indicating an opportunity to focus on targeted strategies to boost performance.



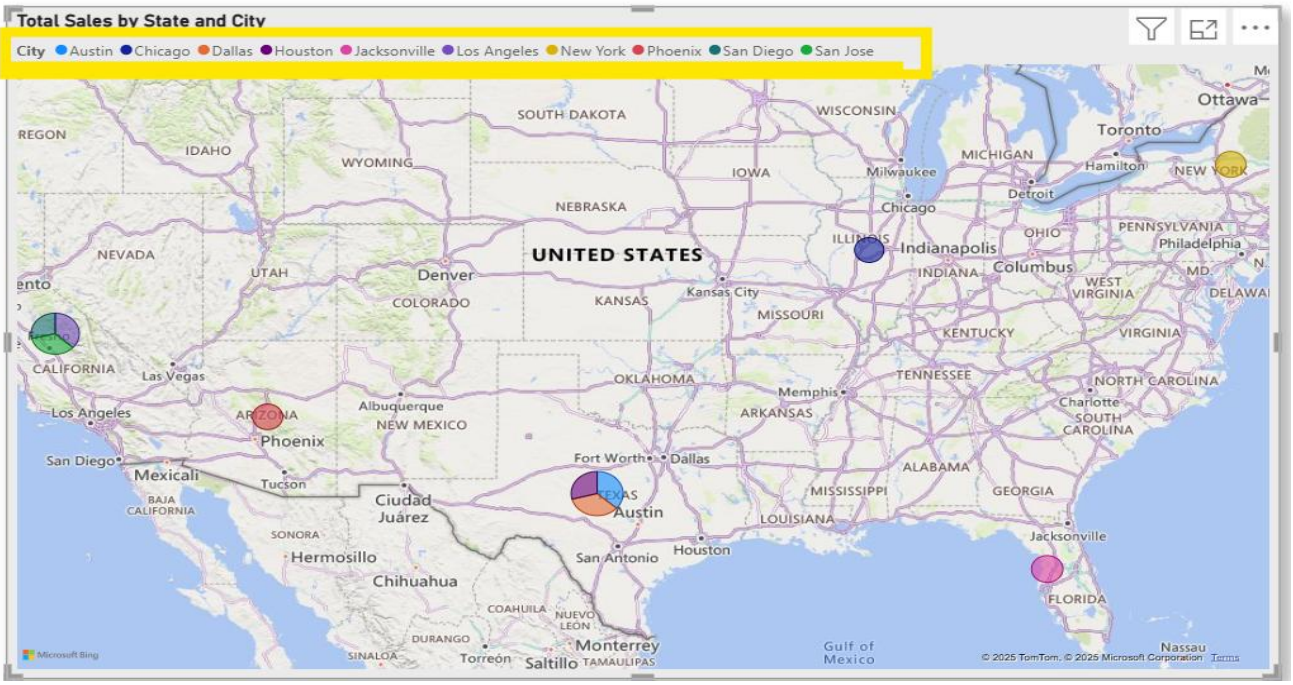
5. Total Sales by State and City:

This visualization uses a **Map Visual** to display **Total Sales** across different **States and Cities** in the United States.

- **Location (Map Field):** State and City
- **Values (Size of Bubble):** Total Sales (SUM(Total Sales))
- **Legend (Color Split):** City

Business Value:

- Helps decision-makers understand **regional sales concentration**.
- Useful for **logistics and supply chain planning** by identifying high-demand locations.
- Supports **regional marketing campaigns** by pinpointing top-performing and low-performing cities.



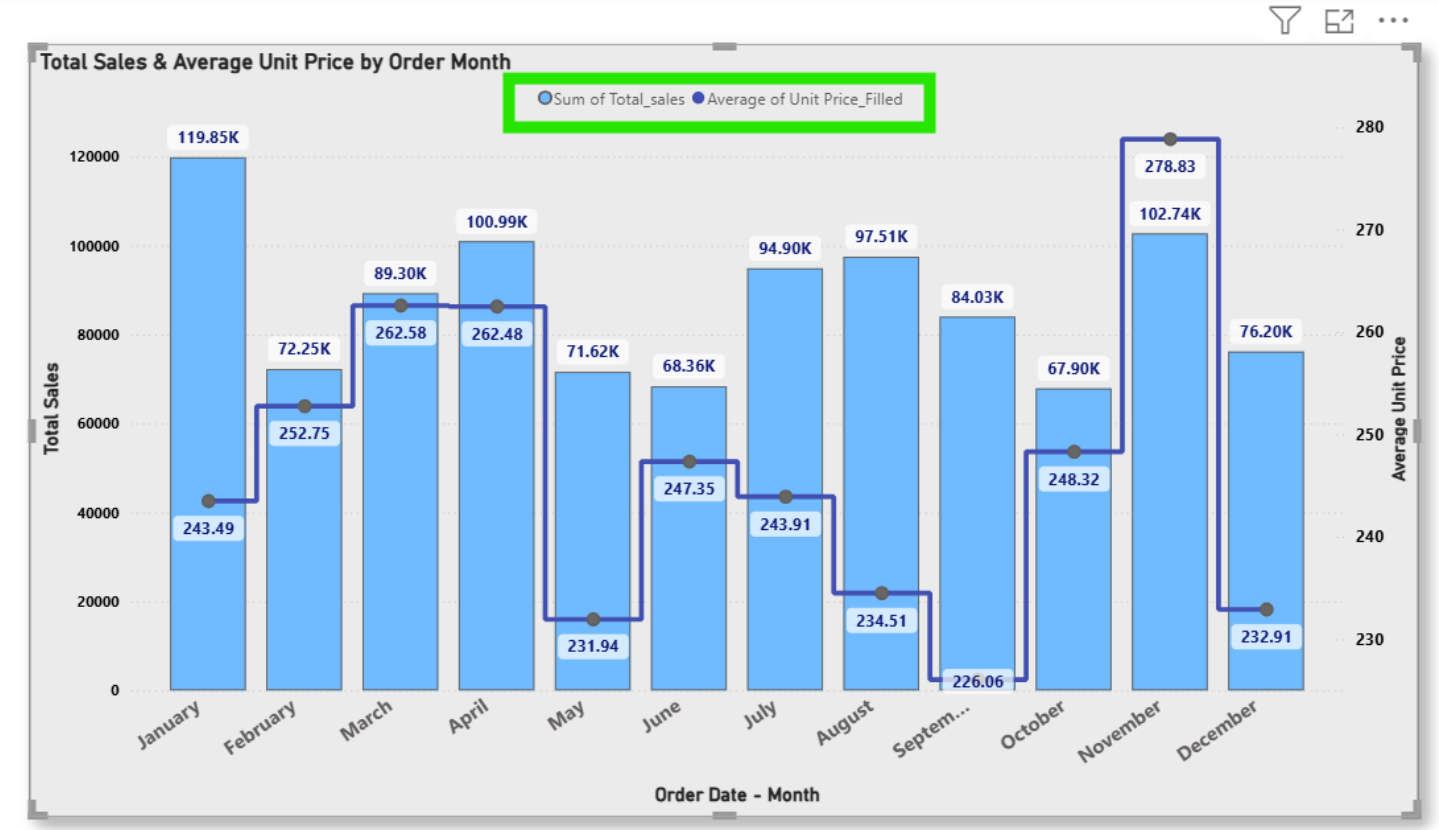
Insights:

The Map visualization highlights that sales are heavily concentrated in Texas and California, while other states show moderate contributions. This suggests regional strengths that can be leveraged for strategic expansion

6. Line and Column Combo Chart: Total Sales vs. Average Unit Price

A **Line and Column Combo Chart** was created to compare **Total Sales** and **Average Unit Price** on a **monthly basis**.

- **Columns (Total Sales):** Show the revenue trend across different months.
- **Line (Average Unit Price):** Represents the variation in average product price during the same period.



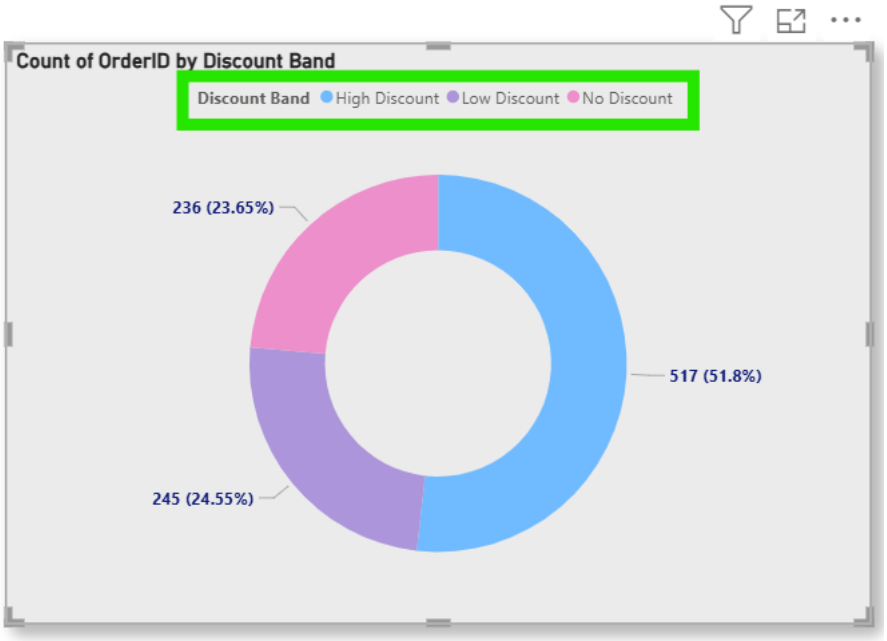
Insights:

- **Sales Performance:** Identifies which months had higher or lower revenue.
- **Price Sensitivity:** Evaluates whether changes in average unit price influenced sales volume.

7. Donut Chart: Order Count by Discount Band

A **Donut Chart** was created to represent the **distribution of orders across different discount bands** – High Discount, Low Discount, and No Discount.

- **High Discount:** 517 orders (51.8%)
- **Low Discount:** 245 orders (24.55%)
- **No Discount:** 236 orders (23.65%)



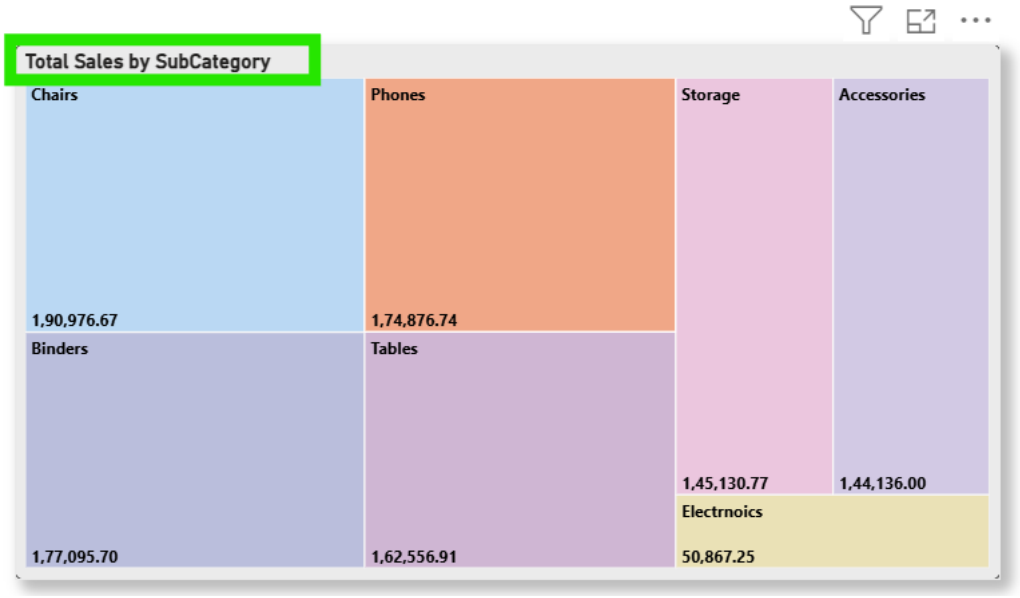
Insights:

Offering **higher discounts** significantly boosts sales volume, while **low discounts are less effective**. However, since No Discount also drives a large share of orders, businesses need to balance between **profit margins and promotional strategies**.

8. Tree Map: Total Sales by Sub-Category

A **Tree Map** was created to represent **Total Sales across different Sub-Categories**.

- Each rectangle in the Tree Map represents a **Sub-Category**, and the **size of the rectangle** corresponds to the **sales value**.
- Sub-Categories with **higher sales** occupy **larger areas**, making it easy to identify top-performing product groups at a glance.
- This visualization is particularly useful to **compare relative contribution** of each Sub-Category without needing exact numbers.



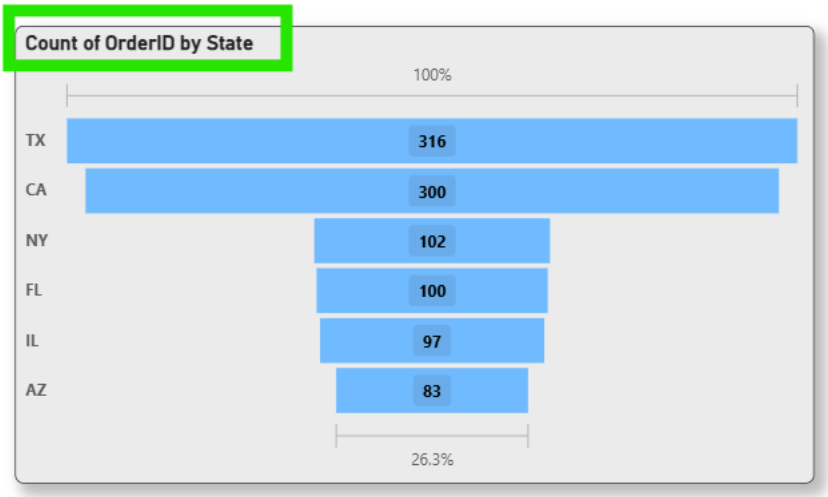
Insight:

The Tree Map clearly highlights which Sub-Categories (e.g., **Chairs, Phones, or Binders**) dominate sales, while smaller rectangles indicate lower-performing Sub-Categories. This allows the business to focus on **high-performing areas for growth** and **low-performing areas for improvement**.

9. Funnel Chart: Order Count by State:

A **Funnel Chart** was created to visualize the **number of orders across different states**.

- The chart arranges states in descending order of **Order Count**, with the **widest part** representing the state with the **highest number of orders** and the **narrowest part** showing the lowest.
- This gives a clear **ranking of states by order volume**, making it easy to identify top contributors.



Insight:

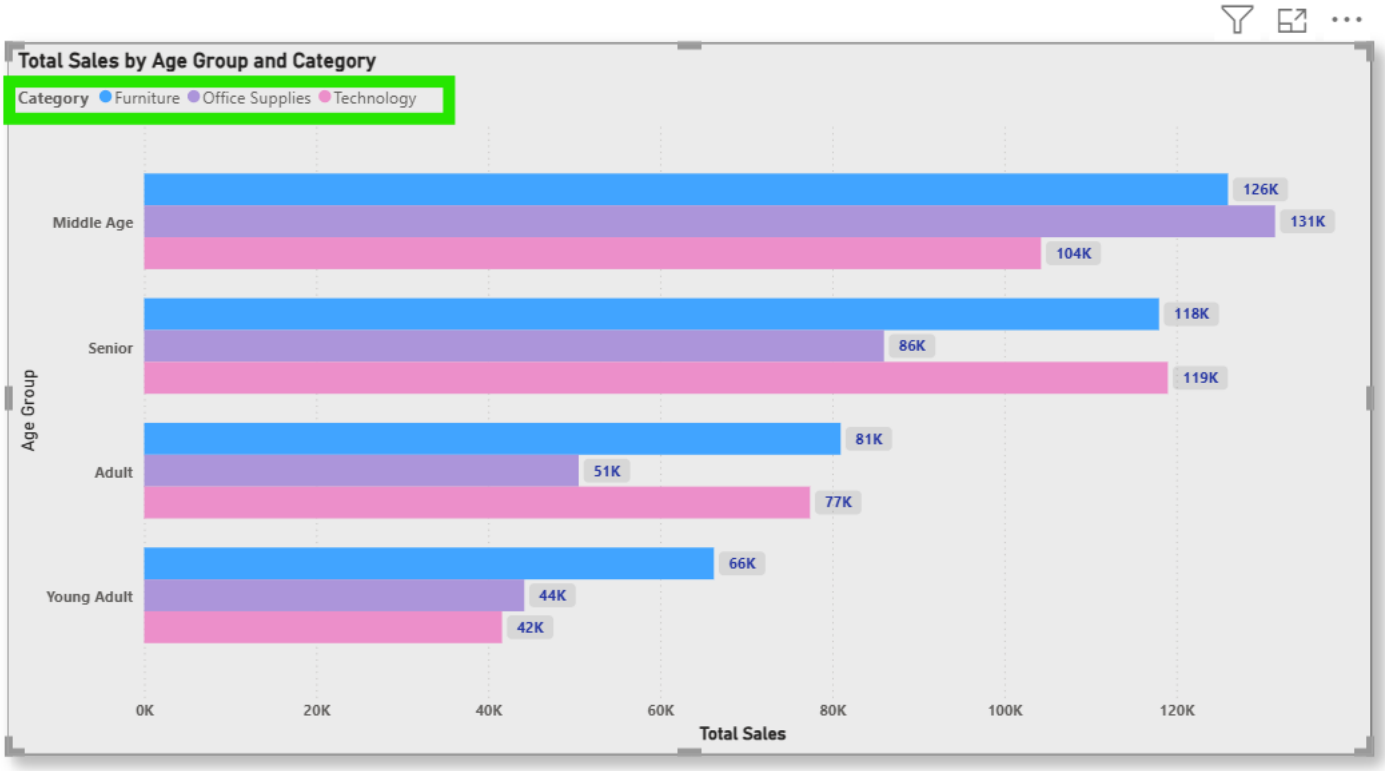
The Funnel highlights that a few states (such as **California, Texas, and New York**) dominate the majority of orders, while several smaller states contribute comparatively less. This suggests that the business should:

- Focus on maintaining performance in high-order states.
- Explore marketing opportunities in low-order states to increase market penetration.

10. Clustered Bar Chart: Total Sales by Age Group and Category:

A **Clustered Bar Chart** was created to analyze **Total Sales across different Age Groups** segmented by **Category**.

- The **X-axis** represents **Total Sales**, while the **Y-axis** represents the **Age Groups**.
- Within each Age Group, bars are clustered by **Category** (e.g., Furniture, Technology, Office Supplies) to show category-wise contribution.
- This helps compare how different categories perform within each customer age segment.



Insight:

- Certain age groups (such as **35–50**) show **higher total sales**, indicating stronger purchasing power.
- Categories like **Technology** may dominate in younger age groups, while **Furniture and Office Supplies** might contribute more in older groups.
- The visualization helps businesses **target promotions more effectively** by identifying which age groups prefer which product categories.

11. KPI Cards: Total Sales, Order Count, and Average Discount:

To provide a **high-level summary**, three **KPI Cards** were created in Power BI:

1. **Total Sales** – Displays the overall revenue generated from all transactions.
2. **Order Count (Distinct Order ID)** – Shows the total number of unique customer orders placed.
3. **Average Discount** – Represents the average percentage of discount offered across all transactions.

These KPI cards act as the **key performance indicators (KPIs)** of the business, enabling decision-makers to quickly track performance without analyzing individual charts.

Insight:

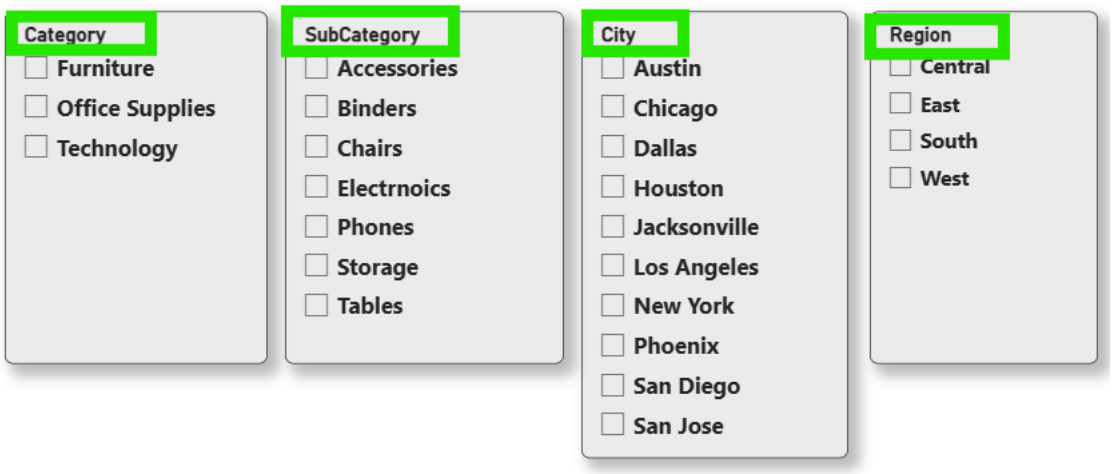
- The **Total Sales card** indicates the company's revenue performance.
- The **Order Count card** shows overall customer activity and demand.
- The **Average Discount card** helps monitor discounting practices and ensures they align with profitability goals.

Together, these KPIs provide an **instant overview of sales, demand, and discounting strategy**, serving as a starting point for deeper analysis in the dashboard.

12. Slicers for Interactive Filtering:

To enhance interactivity and allow users to explore data from different perspectives, **Slicers** were added to the dashboard in Power BI. The following slicers were included:

- **City** – Enables filtering of all visuals by specific city.
- **Category** – Allows users to select product categories (Furniture, Technology, Office Supplies).
- **Sub-Category** – Provides detailed filtering within categories (e.g., Chairs, Phones, Binders).
- **Region** – Filters visuals to focus on specific geographical regions.



Insight:

- Slicers help end-users **drill down into the data** and view performance by location, category, or product type.
- They improve usability by making the report **dynamic and customizable**, allowing stakeholders to answer specific business questions without modifying the report structure.

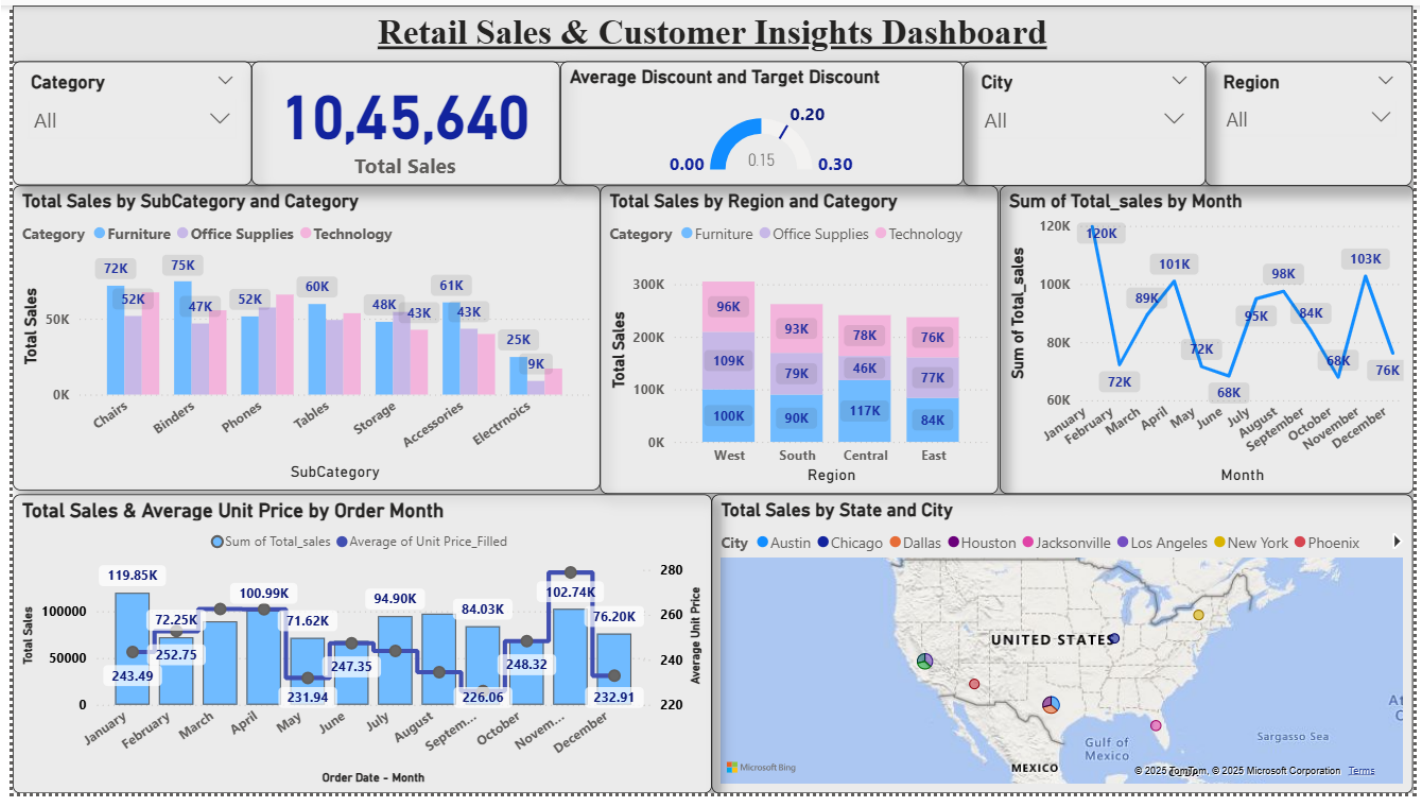
V. Dashboards:

Dashboard 1: Sales & Regional Insights:

This dashboard focuses on **overall sales trends, categories, and regional performance**.

Key Features:

- **KPI Cards**: Total Sales (10,45,640) with Average Discount vs Target Discount.
- **Sub-Category Sales**: Clustered column chart to compare category contribution (e.g., Chairs, Phones, Tables).
- **Regional Sales**: Category-wise breakdown across West, South, Central, and East regions.
- **Trend Analysis**: Line chart showing monthly sales fluctuations, plus a combo chart comparing sales vs. average unit price.
- **Geographical Insights**: Map visualization to track sales concentration by state and city.
- **Interactive Slicers**: Category, Sub-Category, City, and Region filters for dynamic exploration.



Insights:

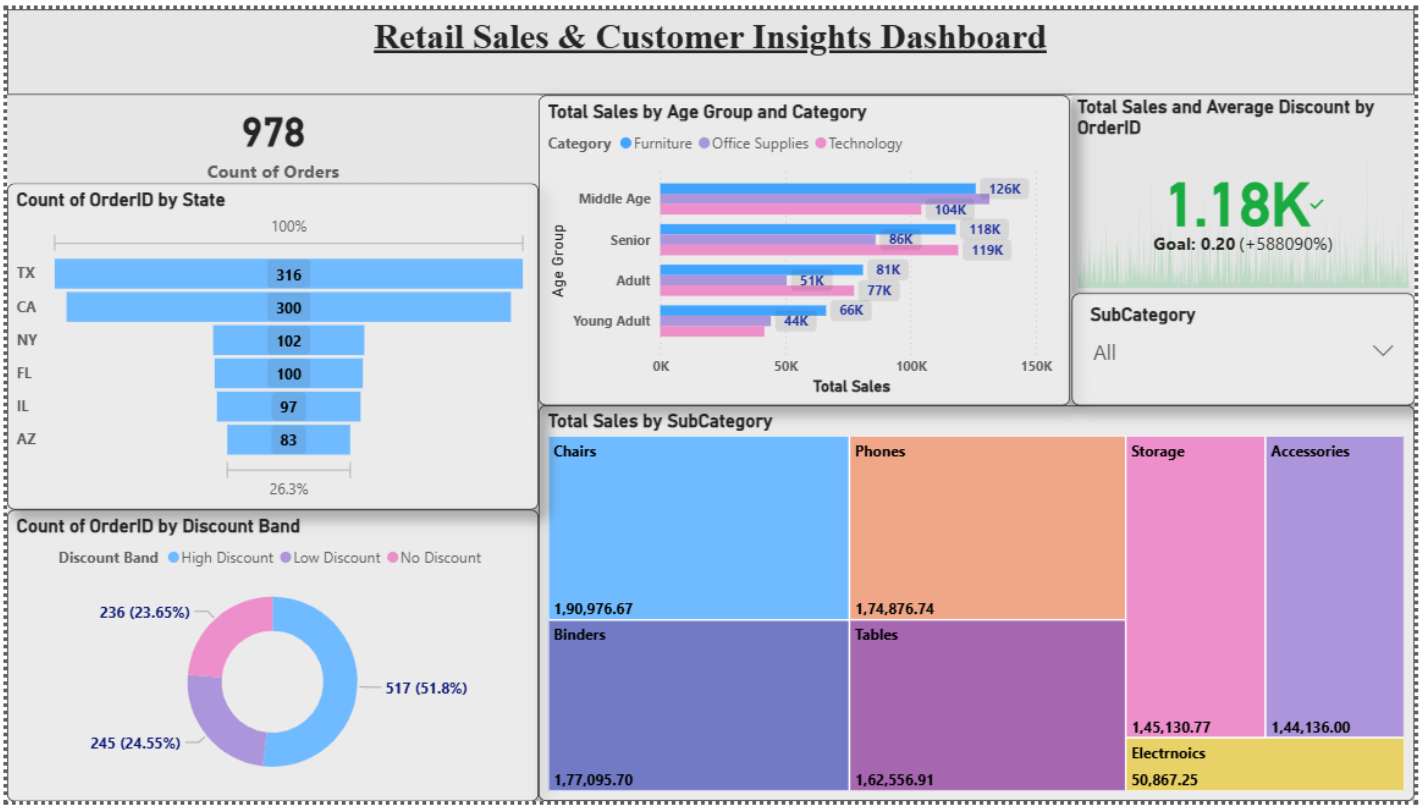
- Chairs and Phones lead in sub-category sales.
- The West region performs strongly across categories.
- Some months show clear peaks in sales, highlighting seasonal demand.
- Discounts are within target ranges, ensuring profitability.
- Cities like New York, Dallas, and Los Angeles are key sales hubs.

Dashboard 2: Customer Demographics & Discount Analysis:

This dashboard highlights **order distribution, customer age group analysis, and discount impact**.

Key Features:

- **KPI Cards**: Count of Orders (978) and Average Discount vs Goal.
- **Orders by State**: Funnel chart ranking states by total order count (TX, CA, NY being top).
- **Orders by Discount Band**: Donut chart showing distribution of High, Low, and No Discount orders.
- **Age Group Analysis**: Bar chart showing sales contribution by customer age group across categories.
- **Sub-Category Sales**: Treemap highlighting top-performing sub-categories like Chairs, Binders, and Phones.



Insights

- Texas, California, and New York generate the highest number of orders.
- Majority of orders (51.8%) come from **High Discount** purchases, **No Discount** is also similar to **Low Discount** purchases.
- Middle Age and Senior groups contribute the highest sales across all categories.
- Chairs, Phones, and Binders are top-performing sub-categories in terms of revenue.

Overall Conclusion:

Together, these dashboards provide a **360° view of retail sales and customer insights**:

- The first dashboard highlights **sales trends, regional performance, and pricing behaviour**.
- The second dashboard adds insights into **customer demographics, discount strategies, and product-level performance**.
- By using slicers and KPIs, users can **drill down into specific business questions** and align strategies with customer demand and regional opportunities.