

FINAL REPORT

NAME : BANUPRIYA N
BATCH NUMBER : TN-DA-FNB03
CONTACT NUMBER : +91 6374056072
EMAIL ID : banupriyamba98@gmail.com
DATASET DOMAIN : RETAIL & E-COMMERCE
PROJECT TITLE : END-TO-END DATA CLEANING AND VISUALIZATION OF RETAIL
STORE SALES THROUGH EXCEL AND POWER BI
SUBMISSION DATE : 23-08-2025
MENTOR NAME : KUMARAN M

RAW DATASET LINK:

https://docs.google.com/spreadsheets/d/1n87mD0aotQniNqhkgH3D2By4mdeG8xU4/edit?usp=drive_link&oid=103974124451059527448&rtpof=true&sd=true

CLEANED DATASET LINK:

https://docs.google.com/spreadsheets/d/1WM7eSYtg0roSz4GYgqm3_ZE0snCFDgCJ/edit?usp=drive_link&oid=103974124451059527448&rtpof=true&sd=true

DATA CLEANING

1) Splitting First Name Column in Customer Details Table:

Convert Text to Columns Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters

- ☐ Tab
- ☐ Semicolon
- ☐ Comma
- ☐ Space
- ☒ Other: .

☐ Treat consecutive delimiters as one

Text qualifier: " ▾

Data preview

First Name
Ms Diya
Mr Krishna
Ms Meera
Mr Arjun
Mr Dinesh
Mr Aditya

Buttons: Cancel, < Back, Next >, Finish

Splitting First Name Column in Customer Details table by Delimiters (.) in Excel using the path Data Tab→Data Tools Group→Text to Columns→delimiters other(.) to split salutation occurs with First Name.

2) Making First Row as Headers in Customer details Table:

	AB_C Column2	AB_C Column3	AB_C Column4	AB_C Column5	AB_C Column6
1	Salutation	First Name	Last Name	Gender	City
2	Ms	Diya	Das	Male	Pune
3	Mr	Krishna	Kumar	Male	Mangalore
4	Ms	Meera	Kumari	Female	Chennai

Making First as headers in customer details table in power query editor in excel by Transform tab→Table Group→Use First Row as Headers→Use first Row as Headers, Changed screenshot attached below,

AB_C Salutation	AB_C First Name	AB_C Last Name	AB_C Gender	AB_C City	AB_C State
Ms	Diya	Das	Male	Pune	Maharashtra
Mr	Krishna	Kumar	Male	Mangalore	Karnataka
Ms	Meera	Kumari	Female	Chennai	TamilNadu
Mr	Arjun	Gupta	Male	Patna	Bihar
Mr	Dinesh	Tulkarni	Male	Bhopa;	Madhya Pradesh
Mr	Aditya	kumar	Male	Kolkata	West Benga;
Mr	Rohan	singh	Male	Jodhpur	Rajasthan
Mr	benze	Emmanuel	Male	Indore	Madhyapadesh
Mr	Dakshina	Moorthy	Male	Kota	Rajasthan

PROPERTIES

Name
Table1

All Properties

APPLIED STEPS

- Source
- Changed Type
- Promoted Headers
- First row as headers

3) Remove Inconsistencies from First Name and Last Name:

Data Type: Text 1 Replace Values 2 Unpivot Columns 3 Split Column 4 Merge Columns 5 Extract 6 Parse

Detect Data Type Fill Move Pivot Column Convert to List Any Column

TransformColumnTypes("#Promoted Headers",{"Customer ID",

AB_C	First Name	AB_C	Last Name	AB_C	Gender	AB_C	City	AB_C	State
	Diya		Das		Male		Pune		Maharashtra
	Krishna		Kumar		Male		Mangalore		Karnataka
	Meera		Kumari		Female		Chennai		TamilNadu
	Arjun		Gupta		Male		Patna		Bihar
	Dinesh		Tulkarni		Male		Bhopal		Madhya Pradesh
	Aditya		kumar		Male		Kolkata		West Bengal
	Rohan		singh		Male		Jodhpur		Rajasthan
	benze		Emmanuvel		Male		Indore		Madhyapadesh
	Dakshina		Moorthy		Male		Kota		Rajasthan
	MOHAN		KUMAR		Male		Madurai		Tamil Nadu
	Naveen		Raj		Male		Tiruchirappalli		Tamil Nadu
	Krishna		Moorthy		Male		Durgapur		West Bengal
	Ananya		Verma		Female		Bhopal		Madhya Pradesh
	Kabir		Singh		Male		Coimbatore		TamilNadu
	Pari		Reddy		Male		Nagpur		Maharashtra
	vimal		kumar		Male		Kolkata		West Bengal
	Ishaan		singh		Male		Mysuru		Karnataka
	Navya		Gupta		FEMALE		Udaipur		Rajasthan
	Arjun		Reddy		Male		Noida		Uttar Pradesh
	Varun		Vangli		Male		Lucknow		Uttar Pradesh

First Name and Last name column having Inconsistencies and uppercase, So Inconsistencies removed in power query editor in excel by selecting two columns and Transform tab→Text Column Group→Format option→Capitalize Each word, Trim and Clean.

AB_C	First Name	AB_C	Last Name	AB_C	Gender	AB_C	City	AB_C	State
	Diya		Das		Male		Pune		Maharashtra
	Krishna		Kumar		Male		Mangalore		Karnataka
	Meera		Kumari		Female		Chennai		TamilNadu
	Arjun		Gupta		Male		Patna		Bihar
	Dinesh		Tulkarni		Male		Bhopal		Madhya Pradesh
	Aditya		Kumar		Male		Kolkata		West Bengal
	Rohan		Singh		Male		Jodhpur		Rajasthan
	Benze		Emmanuvel		Male		Indore		Madhyapadesh
	Dakshina		Moorthy		Male		Kota		Rajasthan
	Mohan		Kumar		Male		Madurai		Tamil Nadu
	Naveen		Raj		Male		Tiruchirappalli		Tamil Nadu
	Krishna		Moorthy		Male		Durgapur		West Bengal
	Ananya		Verma		Female		Bhopal		Madhya Pradesh
	Kabir		Singh		Male		Coimbatore		TamilNadu
	Pari		Reddy		Male		Nagpur		Maharashtra
	Vimal		Kumar		Male		Kolkata		West Bengal
	Ishaan		Singh		Male		Mysuru		Karnataka
	Navya		Gupta		FEMALE		Udaipur		Rajasthan
	Arjun		Reddy		Male		Noida		Uttar Pradesh

PROPERTIES

Name
Table1

APPLIED STEPS

- Source
- Changed Type
- Promoted Headers
- First row as headers
- F&L Name Capitalized Each W...
- F & L Name Trimmed Text
- F&L Name Cleaned Text**

4) Merging First Name and Last Name into a Single Column named "Customer Name":

First Name and Last Name Column merged into a single column using Merge columns in power query editor in excel by Selecting two columns Transform tab→Text Column→Merge Columns→Separator (Space)→New Column Name (Customer Name).

Customer Name	Gender	City	State
Diya Das	Male	Pune	Maharashtra
Krishna Kumar	Male	Mangalore	Karnataka
Meera Kumari	Female	Chennai	TamilNadu
Arjun Gupta	Male	Patna	Bihar
Dinesh Tulkarni	Male	Bhopa;	Madhya Pradesh
Aditya Kumar	Male	Kolkata	West Benga;
Rohan Singh	Male	Jodhpur	Rajasthan
Benze Emmanuvel	Male	Indore	Madhyapradesh
Dakshina Moorthy	Male	Kota	Rajasthan
Mohan Kumar	Male	Madurai	Tamil Nadu
Naveen Raj	Male	Tiruchirappalli	Tamil Nadu
Krishna Moorthy	Male	Durgapur	West Bengal
Ananya Verma	Female	Bhopal	Madhya Pradesh
Kabir Singh	Mal	Coimbatore	TamilNadu

PROPERTIES
Name
Table1
All Properties

APPLIED STEPS
Source
Changed Type
Promoted Headers
First row as headers
F&L Name Capitalized Each W...
F & L Name Trimmed Text
F&L Name Cleaned Text
X Merged First & Last name Col...

5) Removing inconsistencies from Gender, City and State Column:

Gender	City	State
Male	Pune	Maharashtra
Male	Mangalore	Karnataka
Female	Chennai	Tamilnadu
Male	Patna	Bihar
Male	Bhopa;	Madhya Pradesh
Male	Kolkata	West Benga;
Male	Jodhpur	Rajasthan
Male	Indore	Madhyapradesh
Male	Kota	Rajasthan
Male	Madurai	Tamil Nadu
Male	Tiruchirappalli	Tamil Nadu
Male	Durgapur	West Bengal
Female	Bhopal	Madhya Pradesh
Mal	Coimbatore	Tamilnadu
Male	Nagpur	Maharashtra
Male	Kolkata	West Bengal
Male	Mysuru	Karnataka

PROPERTIES
Name
Table1
All Properties

APPLIED STEPS
Source
Changed Type
Promoted Headers
First row as headers
F&L Name Capitalized Each W...
F & L Name Trimmed Text
F&L Name Cleaned Text
Merged First & Last name Col...
Gender,city,state Capitalized E...
Gender,city,state Trimmed Text
X Gender,city,state Cleaned Text

Gender, City and State column having Inconsistencies and uppercase, So Inconsistencies removed in power query editor in excel by selecting two columns and Transform tab→Text Column Group→Format option→Capitalize Each word, Trim and Clean.

6) Remove Spelling Mistakes from Gender, City and State column:

Gender	City	State
Male	Pune	Maharashtra
Male	Mangalore	Karnataka
Female	Chennai	Tamilnadu
Male	Patna	Bihar
Male	Bhopa;	Madhya Pradesh
Male	Kolkata	West Benga;
Male	Jodhpur	Rajasthan
Male	Indore	Madhyapradesh
Male	Kota	Rajasthan
Male	Madurai	Tamil Nadu
Male	Tiruchirappalli	Tamil Nadu
Male	Durgapur	West Bengal
Female	Bhopal	Madhya Pradesh
Mal	Coimbatore	Tamilnadu

Spelling mistakes in gender (Mal, Femal;), city (Bhopa;), State (Tamilnadu, West Benga;, Madhyapradesh) are removed using Transform→Any Column→Replace Values in power query editor in excel.

Gender	City	State
Male	Pune	Maharashtra
Male	Mangalore	Karnataka
Female	Chennai	Tamil Nadu
Male	Patna	Bihar
Male	Bhopal	Madhya Pradesh
Male	Kolkata	West Bengal
Male	Jodhpur	Rajasthan
Male	Indore	Madhya Pradesh
Male	Kota	Rajasthan
Male	Madurai	Tamil Nadu
Male	Tiruchirappalli	Tamil Nadu
Male	Durgapur	West Bengal
Female	Bhopal	Madhya Pradesh
Male	Coimbatore	Tamil Nadu
Male	Nagpur	Maharashtra
Male	Kolkata	West Bengal
Male	Mysuru	Karnataka
Female	Udaipur	Rajasthan
Male	Noida	Uttar Pradesh
Male	Lucknow	Uttar Pradesh

PROPERTIES
Name
Table1
All Properties

APPLIED STEPS
First Row as Headers
F&L Name Capitalized Each...
F & L Name Trimmed Text
F&L Name Cleaned Text
Merged First & Last name ...
Gender,city,state Capitalize...
Gender,city,state Trimmed ...
Gender,city,state Cleaned T...
Filtered Rows
Replaced Value
Gender Replaced Value
Gender Female Replaced V...
Replaced Value in city
Replaced Value in state
Replaced Value in state
X Replaced Value state

7) Data Type Standardization in retail store sales table:

Data Type: Date/Time	Merge Queries	Manage Parameters	Data source settings	New Source
Use First Row as Headers	Append Queries	Combine Files	Recent Sources	Enter Data
Replace Values	Combine	Parameters	Data Sources	New Query

type text}, {"Customer ID", type text}, {"Category", type text},

Location	Transaction Date	Discount Applied
ne	08-04-2024 00:00:00	TRUE
ne	23-07-2023 00:00:00	TRUE
ne	05-10-2022 00:00:00	FALSE
ne	07-05-2022 00:00:00	null
ne	02-10-2022 00:00:00	FALSE
ne	30-11-2023 00:00:00	null
tore	10-06-2023 00:00:00	TRUE
tore	02-04-2024 00:00:00	TRUE
tore	26-04-2023 00:00:00	FALSE
ne	11-03-2024 00:00:00	FALSE

Query Settings
PROPERTIES
Name
Table4 (2)
All Properties
APPLIED STEPS
Source
X Changed Type

Data Type for Date column changed as Date and "DD-MM-YYYY" format using power query editor in excel through Home Tab → Transform group → Data Type → Date.

8) Filling null/Blank values in Discount Applied Column:

The cells which holding null values or Blank are considered as Zero/No Discount was Applied. So as per the Boolean values (TRUE=1, FALSE=0) blank/Null cells replaced as FALSE Using REPLACE VALUES Method in Power Query Editor in Excel. Selecting the Discount Applied Column → Transform Tab → Any Column → Replace Values → Replace Values.

Replace Values

Replace one value with another in the selected columns.

Value To Find

null

Replace With

FALSE

OK

Cancel

Discount Applied

TRUE
TRUE
FALSE
FALSE
FALSE
FALSE
TRUE
TRUE
FALSE
FALSE
TRUE
TRUE
TRUE
FALSE
FALSE
FALSE
TRUE
FALSE
FALSE
FALSE
TRUE
FALSE
FALSE
FALSE

PROPERTIES

Name
Table4 (2)

All Properties

APPLIED STEPS

Source
Changed Type
✕ Replaced Value in Discount ap... ⚙

9) Filling Blank Values in Price per Unit Column:

=IF(ISBLANK([@Price]),[@Total Spent]/[@Quantity],[@Price])									
Price Per Unit	Quantity	Total Spent	Payment Method	Location	Transaction Date	Discount Applied	Price	Price per unit calculated	
18.5	10	185	Digital Wallet	Online	08-04-2024	TRUE	18.5	18.5	18.5
29	9	261	Digital Wallet	Online	23-07-2023	TRUE	29	29	29
21.5	2	43	Credit Card	Online	05-10-2022	FALSE	21.5	21.5	21.5
27.5	9	247.5	Credit Card	Online	07-05-2022	FALSE	27.5	27.5	27.5
12.5	7	87.5	Digital Wallet	Online	02-10-2022	FALSE	12.5	12.5	12.5
	10	200	Credit Card	Online	30-11-2023	FALSE	20	20	20
5	8	40	Credit Card	In-store	10-06-2023	TRUE	5	5	5
33.5			Digital Wallet	In-store	02-04-2024	TRUE			33.5
27.5	1	27.5	Credit Card	In-store	26-04-2023	FALSE	27.5	27.5	27.5
36.5	3	109.5	Cash	Online	14-03-2024	FALSE	36.5	36.5	36.5
8	9	72	Cash	In-store	14-12-2024	TRUE	8	8	8
	8	52	Digital Wallet	In-store	12-01-2025	TRUE	6.5	6.5	6.5
6.5	7	45.5	Cash	Online	09-09-2023	TRUE	6.5	6.5	6.5
39.5	6	237	Digital Wallet	In-store	20-05-2022	FALSE	39.5	39.5	39.5
27.5	2	55	Digital Wallet	Online	22-05-2022	FALSE	27.5	27.5	27.5
24.5			Credit Card	In-store	11-05-2022	FALSE			24.5
29	8	232	Cash	Online	16-11-2024	TRUE	29	29	29
	10	275	Digital Wallet	Online	17-04-2022	FALSE	27.5	27.5	27.5
23	1	23	Digital Wallet	In-store	07-02-2022	FALSE	23	23	23

Price per unit column having blank values and those values are filled by using Formula in two cases,

Case 1: Using Total Spend and Quantity

Price column filled by formula **Price=Total Spend/Quantity**.

Case 2: Using Old price per unit column

In some rows, Quantity or Total Spend is missing but Price per unit already appears, Created one new column to use both cases using formula,

=IF(ISBLANK([@Price]),[@Price Per Unit],[@Price])

After that Copied the New column values and pasted in old column “Price per unit” as values using paste special option.

10) Filling Blank values in ITEM Column:

Item ID	Category	Item	Price Per Unit	Quantity	Total Spent
	Beverages	Item_20_BEV	33.5	5	167.5
	Beverages	Item_20_BEV	33.5	1	33.5
	Beverages	Item_20_BEV	33.5	1	33.5
	Beverages	Item_20_BEV	33.5	6	201
	Beverages	Item_20_BEV	33.5	5	167.5
	Beverages	Item_20_BEV	33.5	1	33.5
	Beverages		33.5		
	Beverages	Item_20_BEV	33.5	7	234.5
	Beverages	Item_20_BEV	33.5	6	201
	Beverages	Item_20_BEV	33.5	10	335
	Beverages	Item_20_BEV	33.5	8	268
	Beverages	Item_20_BEV	33.5	8	268
	Beverages	Item_20_BEV	33.5	1	33.5
	Beverages	Item_20_BEV	33.5	8	268
	Beverages	Item_20_BEV	33.5	4	134
	Beverages	Item_20_BEV	33.5	6	201
	Beverages	Item_20_BEV	33.5	6	201
	Beverages	Item_20_BEV	33.5	7	234.5
	Beverages	Item_20_BEV	33.5	8	268

Missing values are filled in Item column based on “Category”, “Price per unit”, “Quantity” and “Total Spend” Column and using Filter and Find & Replace method.

1)Filtering Category

2)Filtering price per unit as per the blank cells in “Item” Column.

3)Filling values for Blank cells using Find & Replace method

4)Same process followed for each category

Category	Item	Price Per Unit	Quantity	Total Spent	Payment Method	Location
Beverages		27.5			Cash	In-store
Beverages		27.5			Credit Card	Online
Beverages		27.5			Cash	Online
Beverages		27.5			Credit Card	In-store

Find and Replace

Find

Replace

Find what:

Replace with:

Options >>

Replace All

Replace

Find All

Find Next

Close

11) Filling Blank cells in Quantity Column:

Quantity column having Blank cells and blank cells filled by Finding MEDIAN, Because, quantity should not be a decimal number, it should be a whole number, so Median calculated by Category wise using EXCEL.

Formula =IF(ISBLANK(F5), MEDIAN(\$F\$5:\$F\$12573), F5)

	D	E	F
Item	Price per unit	Quantity	Q
Item_10_PAT	18.5	10	
Item_17_MILK	29	9	
Item_12_BUT	21.5	2	
Item_16_BEV	27.5	9	
Item_6_FOOD	12.5	7	
Item_11_PAT	20	10	
Item_1_FOOD	5	8	
Item_20_FUR	33.5		
Item_16_FUR	27.5	1	
Item_22_BUT	36.5	3	
Item_3_BUT	8	9	
Item_2_MILK	6.5	8	
Item_2_FOOD	6.5	7	
Item_24_PAT	39.5	6	
Item_16_MILK	27.5	2	
Item_14_BEV	24.5		
Item_17_PAT	29	8	
Item_16_MILK	27.5	10	
Item_13_EHE	23	1	

And the same process followed for each category. Values pasted in old column using paste special→values option.

=IF(ISBLANK(F6966), MEDIAN(\$F\$5:\$F\$12573), F6966)					
ID	Category	Item	Price per unit	Quantity	Quantity Calculated
	Beverages	Item_16_BEV	27.5	9	9
	Beverages	Item_14_BEV	24.5		6
	Beverages	Item_7_BEV	14	9	9
	Beverages	Item_25_BEV	41	7	7
	Beverages	Item_11_BEV	20	5	5
	Beverages	Item_22_BEV	36.5	7	7
	Beverages	Item_8_BEV	15.5	3	3
	Beverages	Item_4_BEV	9.5	7	7
	Beverages	Item_7_BEV	14	6	6
	Beverages	Item_22_BEV	36.5	6	6
	Beverages	Item_14_BEV	24.5	2	2
	Beverages	Item_24_BEV	39.5		6
	Beverages	Item_23_BEV	38	8	8
	Beverages	Item_19_BEV	32	6	6
	Beverages	Item_20_BEV	33.5	5	5
	Beverages	Item_6_BEV	12.5	2	2
	Beverages	Item_20_BEV	33.5	1	1
	Beverages	Item_22_BEV	36.5	1	1
	Beverages	Item_6_BEV	12.5	5	5

12) Fill Blank Values In total spend column:

Total Spend Column having null values and that are filled by using formula

Total Spend=Price*Quantity

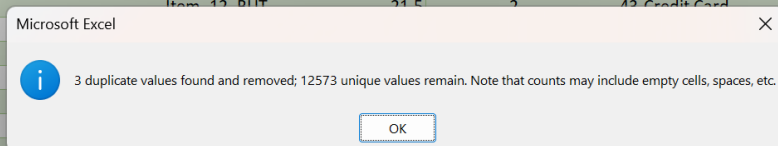
Values pasted in old column using Paste special option in excel.

=E9*F9						
	C	D	E	F	G	H
D	Category	Item	Price per unit	Quantity	Total	Total Spent
	Patisserie	Item_10_PAT	18.5	10	185	185
	Milk Products	Item_17_MILK	29	9	261	261
	Butchers	Item_12_BUT	21.5	2	43	43
	Beverages	Item_16_BEV	27.5	9	247.5	247.5
	Food	Item_6_FOOD	12.5	7	87.5	87.5
	Patisserie	Item_11_PAT	20	10	200	200
	Food	Item_1_FOOD	5	8	40	40
	Furniture	Item_20_FUR	33.5	6	201	
	Furniture	Item_16_FUR	27.5	1	27.5	27.5
	Butchers	Item_22_BUT	36.5	3	109.5	109.5
	Butchers	Item_3_BUT	8	9	72	72
	Milk Products	Item_2_MILK	6.5	8	52	52
	Food	Item_2_FOOD	6.5	7	45.5	45.5
	Patisserie	Item_24_PAT	39.5	6	237	237

13) Duplicate Removal in retail store sales table:

In this table Transaction ID and Customer ID both are considered as Unique Values, After using Remove Duplicates by Transaction ID, 3duplicate values found are removed.

Transaction ID	Customer ID	Category	Item	Price per unit	Quantity	Total Spent	Payment Method	Lead
TXN_6867343	CUST_09	Patisserie	Item_10_PAT	18.5	10	185	Digital Wallet	O
TXN_3731986	CUST_22	Milk Products	Item_17_MILK	29	9	261	Digital Wallet	O
TXN_9303719	CUST_02	Butchers	Item_12_BUT	21.5	2	43	Credit Card	O
TXN_9458126	CUST_06	Beverages	Item_16_BEV	27.5	9	247.5	Credit Card	O
TXN_4575373	CUST_05	Food	Item_6_FOOD	12.5	7	87.5	Credit Card	O
TXN_7482416	CUST_09	Patisserie	Item_11_PAT	20	10	200	Credit Card	O
TXN_3652209	CUST_07	Food	Item_1_FOOD	5	8	40	Credit Card	In
TXN_1372952	CUST_21	Furniture	Item_20_FUR	33.5	6	201	Credit Card	In
TXN_9728486	CUST_23	Furniture	Item_16_FUR	27.5	1	27.5	Credit Card	In
TXN_2722661	CUST_25	Butchers	Item_22_BUT	36.5	3	109.5	Cash	O
TXN_8776416	CUST_22	Butchers	Item_3_BUT	8	9	72	Cash	In



14) Data Type Standardization in Item, Price, Total Spend and other Column in EXCEL in retail_store_sales table:

Price per Unit and Total Spend Column Data type changed from General to Currency.

Data Type for Item and other column showing as General, Changed the data type as Text through Home Tab→Number Group→General- Text

Formula Bar			
Price per unit	Quantity	Total Spent	
AT	10	₹ 185.0	D
ILK	9	₹ 261.0	D
JT	2	₹ 43.0	C
EV	9	₹ 247.5	C
DD	7	₹ 87.5	D
AT	10	₹ 200.0	C
DD	8	₹ 40.0	C
JR	6	₹ 201.0	D
JR	1	₹ 27.5	C
JT	3	₹ 109.5	C
T	9	₹ 72.0	C
ILK	8	₹ 52.0	D
DD	7	₹ 45.5	C
AT	6	₹ 237.0	D
ILK	2	₹ 55.0	D
EV	6	₹ 147.0	C
AT	8	₹ 232.0	C
ILK	10	₹ 275.0	D
HE	1	₹ 23.0	D

15) Data Type standardization in customer details table:

Data Type Showing as General for all column in customer_details table, then it changed to TEXT through Home tab→Number Group→General-TEXT.

Customer ID							
Customer ID	Salutation	Customer Name	Gender	City	State		
CUST_01	Ms	Diya Das	Male	Pune	Maharashtra		
CUST_02	Mr	Krishna Kumar	Male	Mangalore	Karnataka		
CUST_03	Ms	Meera Kumari	Female	Chennai	Tamil Nadu		
CUST_04	Mr	Arjun Gupta	Male	Patna	Bihar		
CUST_05	Mr	Dinesh Tulkarni	Male	Bhopal	Madhya Pradesh		

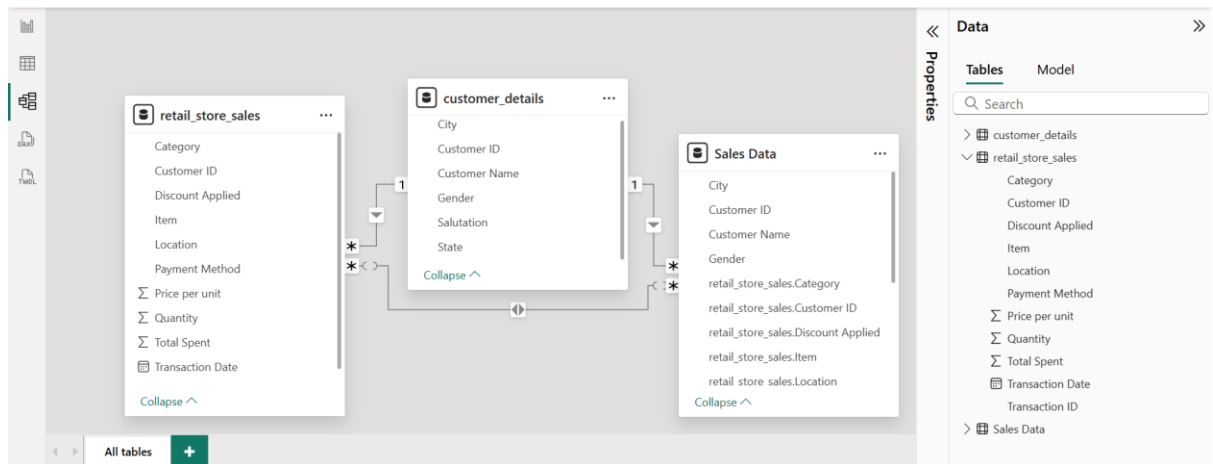
16) Merge Table:

Customer Details Table and Retail store sales table merged as a new table named “Sales Data” using merge queries (Full outer Join) in Power query editor in Power Bi.

City	State	retail_store_sales.Transaction ID	retail_store_sales.Customer ID
1 Pune	Maharashtra	TXN_4206593	CUST_01
2 Kota	Rajasthan	TXN_6867343	CUST_09
3 Kota	Rajasthan	TXN_7482416	CUST_09
4 Kota	Rajasthan	TXN_5422631	CUST_09
5 Kota	Rajasthan	TXN_2490363	CUST_09
6 Mangalore	Karnataka	TXN_9303719	CUST_02
7 Rajkot	Gujarat	TXN_3731986	CUST_22
8 Rajkot	Gujarat	TXN_8776416	CUST_22
9 Patna	Bihar	TXN_3314099	CUST_04
10 Kolkata	West Bengal	TXN_9458126	CUST_06
11 Bhopal	Madhya Pradesh	TXN_4575373	CUST_05
12 Bhopal	Madhya Pradesh	TXN_3481599	CUST_05
13 Jodhpur	Rajasthan	TXN_3652209	CUST_07
14 Gaya	Bihar	TXN_1372952	CUST_21
15 Bhopal	Madhya Pradesh	TXN_9728486	CUST_23
16 Bhopal	Madhya Pradesh	TXN_5874772	CUST_23
17 Bhopal	Madhya Pradesh	TXN_7563311	CUST_23
18 Madurai	Tamil Nadu	TXN_6547964	CUST_10

VISUALIZATION

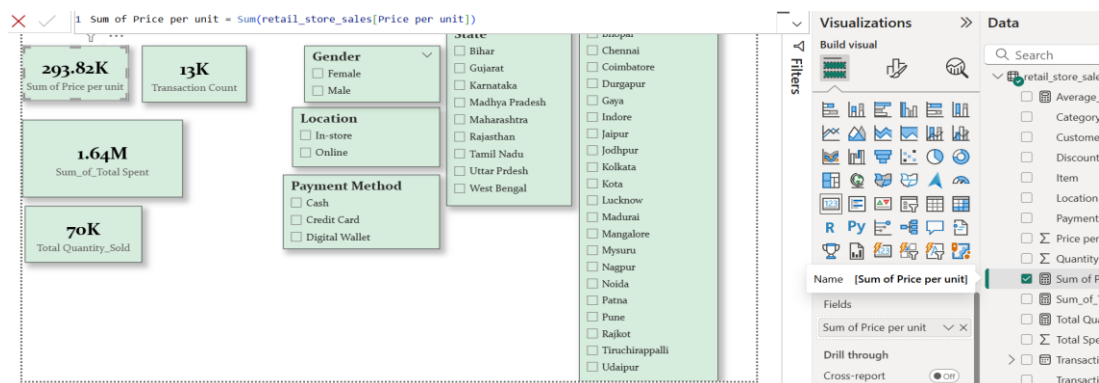
1) Data Modeling:



Data Modeling for three Tables has been created and all the relationships are active.

2) Calculate Measure

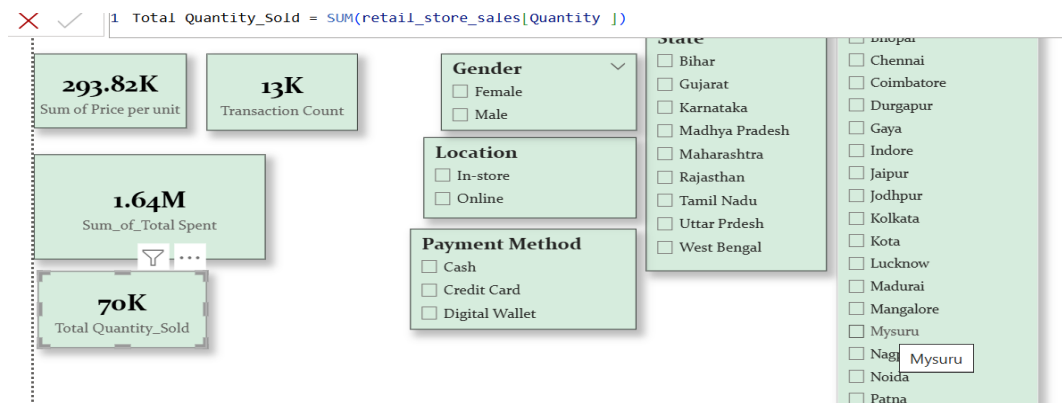
i) Sum of Price per unit:



Total Amount Sold Is calculated using New measure in Power Bi by using formula,

Total Amount Sold = Sum(retail_store_sales[Price per unit])

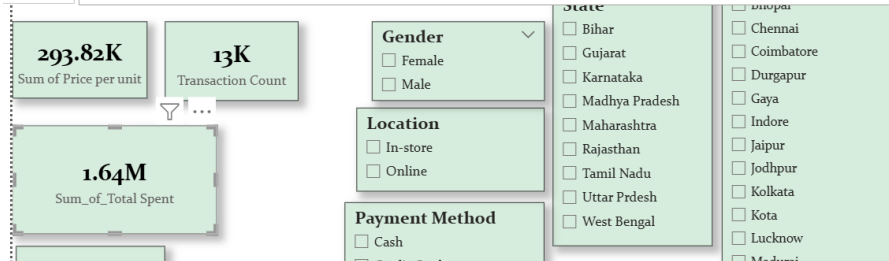
ii) Total Quantity Sold:



Total Quantity_Sold = SUM(retail_store_sales[Quantity])

iii) Sum of Total Spent (Revenue):

✗ ✓ 1 Sum_of_Total Spent = SUM(retail_store_sales[Total Spent])



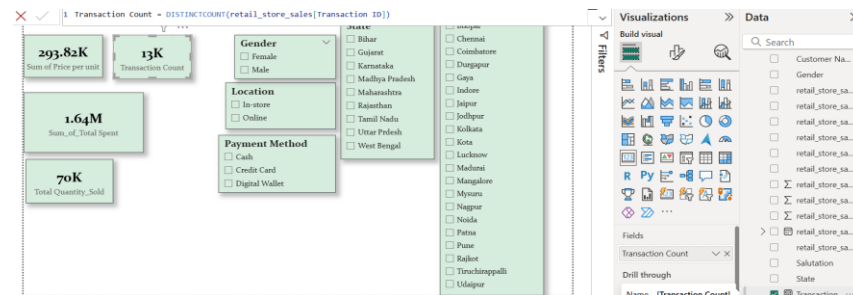
Sum of Total Spent has been Calculated using New measure in Power Bi by Formula,

Sum_of_Total Spent = SUM(retail_store_sales[Total Spent])

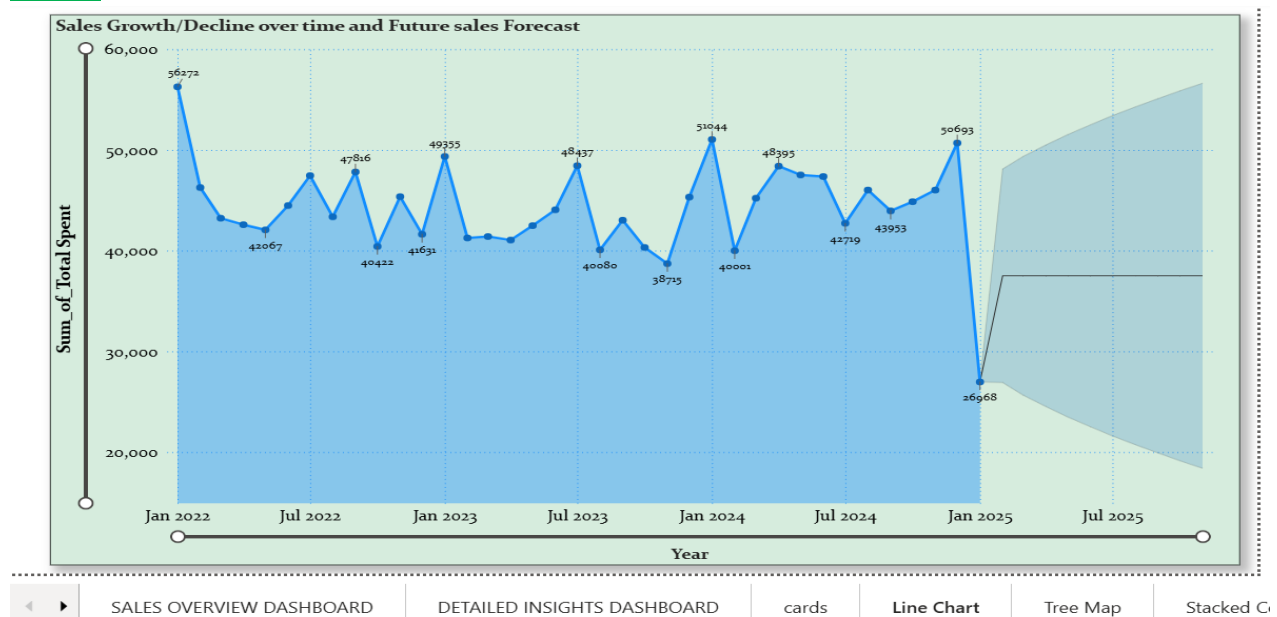
iv) Calculating Transaction Count:

Unique order count has been calculated using New measure formula,

Transaction Count = DISTINCTCOUNT(retail_store_sales[Transaction ID])

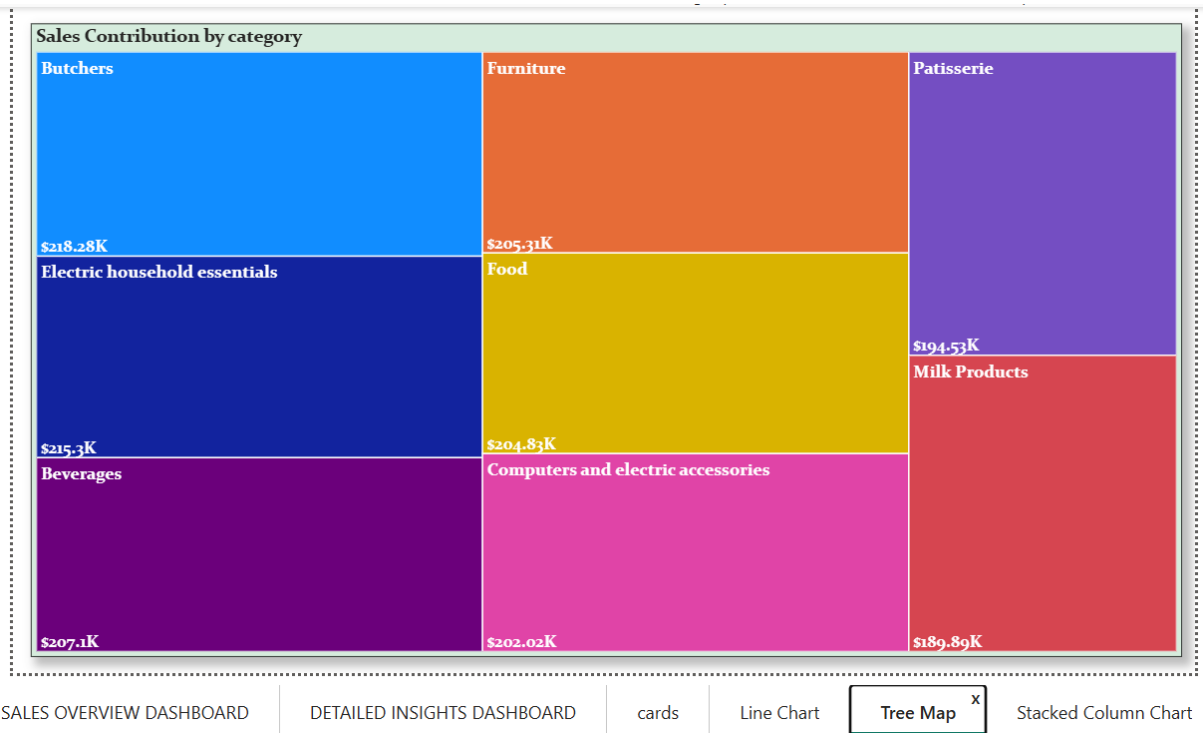


3) Sales Growth/Decline over time and Future Sales Forecast Using Line chart:



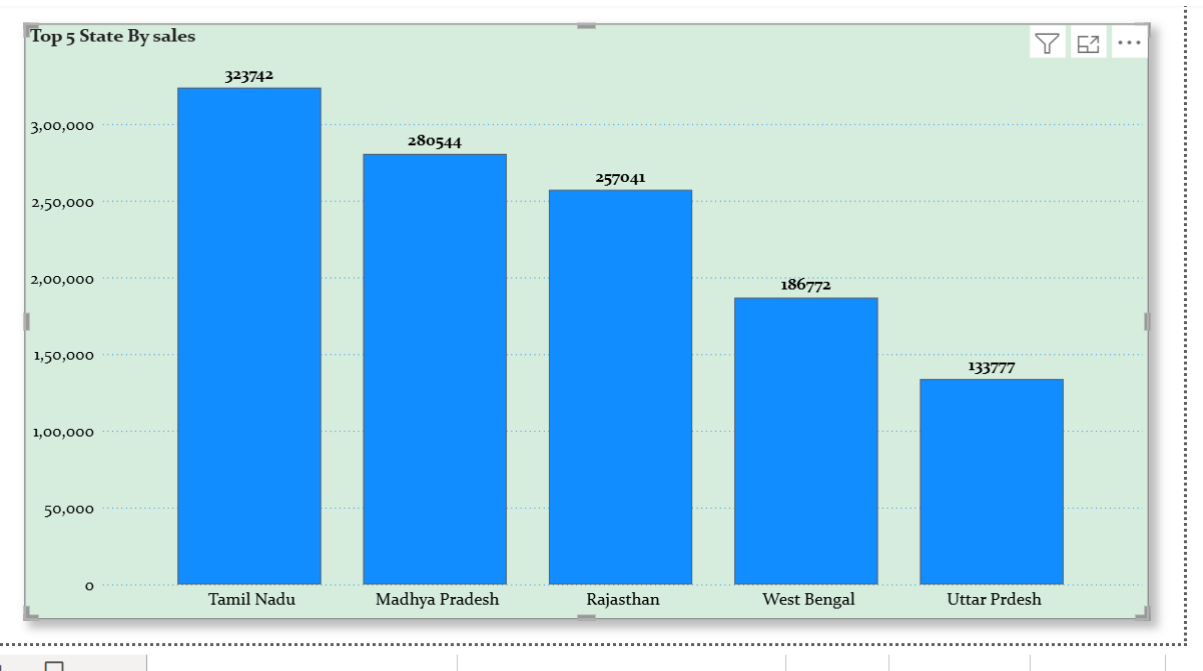
Highest Sales was Happened on 2022 January with the sales amount of 56,272 and the lowest sales was happened on 2025 January with the sales amount of 26,967. And we can expect the sales growth in 2025 November with Forecast amount of 37,508.

4) Sales Contribution by Category using Tree Map:



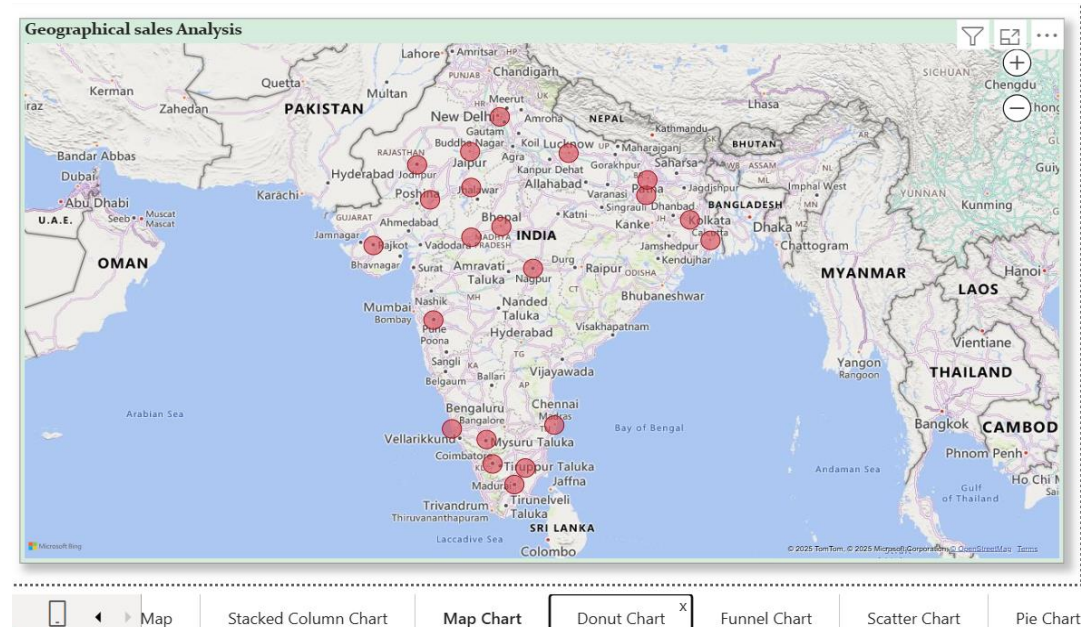
According to the Tree map placed above, Category Butchers Contributed High Sales for 2,18,282 and Milk Products Contributed Low sales for 1,89,892.

5) Top 5 State by sales using Stacked Column Chart:

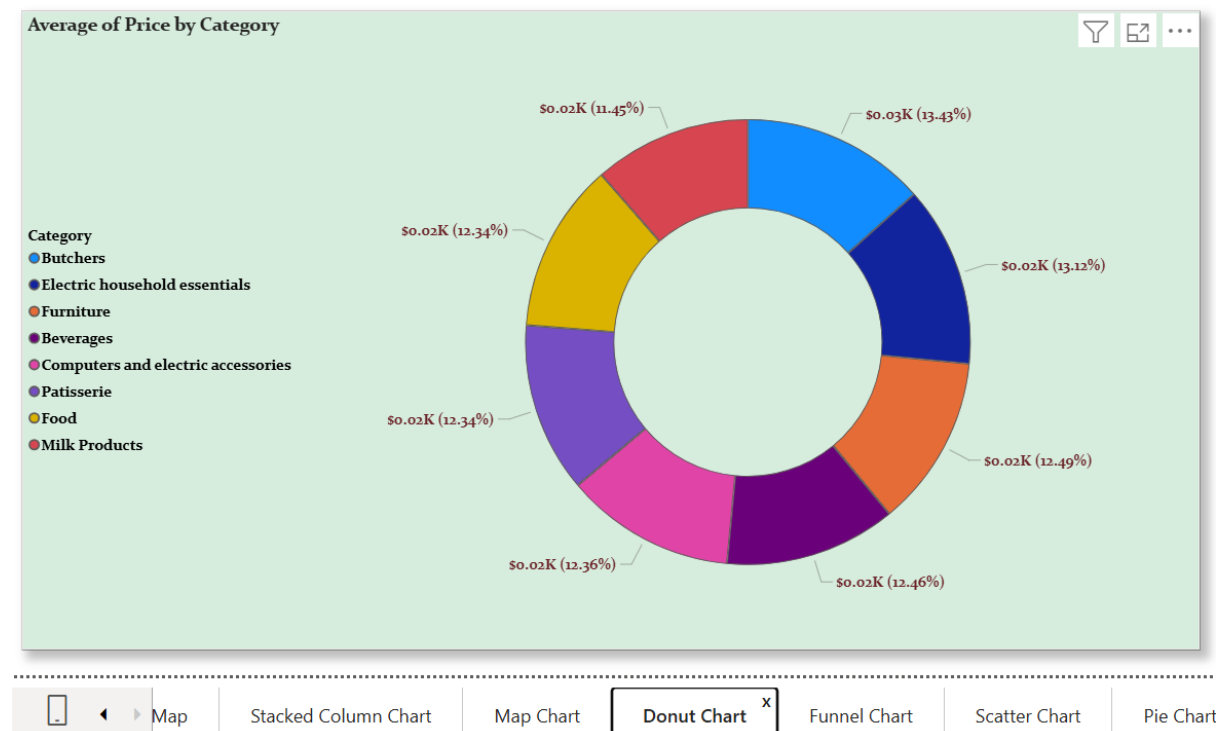


The above chart clearly shows that the Top 5 Staes by sales, Tamil Nadu With Sales Value 3,23,742, Madhya Pradesh 2,80,544, Rajasthan with sales value 2,57,041, West Bengal with sales value 1,86,772 and Uttar Pradesh with 1,33,777.

6) Geographical Sales Analysis Map:

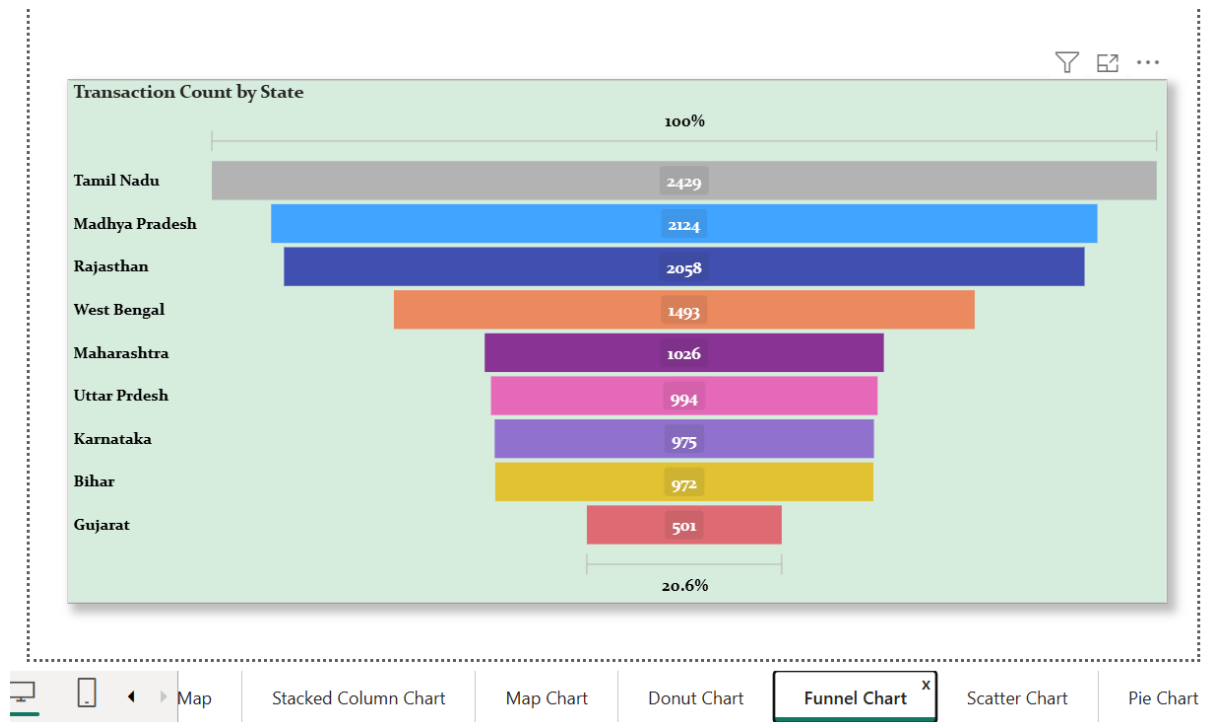


7) Average of Price by Category using Donut Chart:



The above Donut Chart shows the Average price per unit with value of percentage By Category.

8) Transaction Count by State using Funnel Chart:

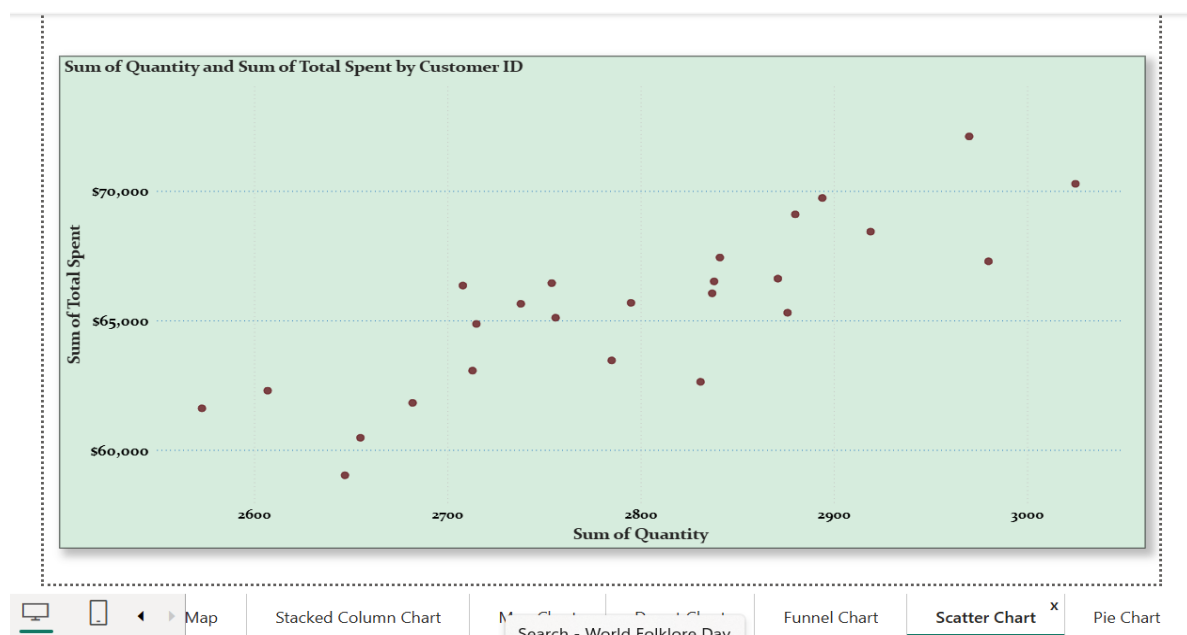


The Highest Transaction Happened In Tamil Nadu, and Lowest Transaction Happened in Gujarat.

Transaction Count of Tamil Nadu: 2429

Transaction Count of Gujarat: 501

9) Sum of Quantity and Sum of Total Spent by Customer ID by Scatter Chart:



The above Chart Shows Sum of Quantity and Total spent by Customer Id. CUST_12 Purchased Multiple Products, Quantity 2647 by spending 59.019 which is low compared to others. And CUST_05 purchased Multiple Products, Quantity of 2970 by spending 72,106, which is High Compared to others. So, this Customer Id CUST_05 Should be most Frequent Buyer in this shop.

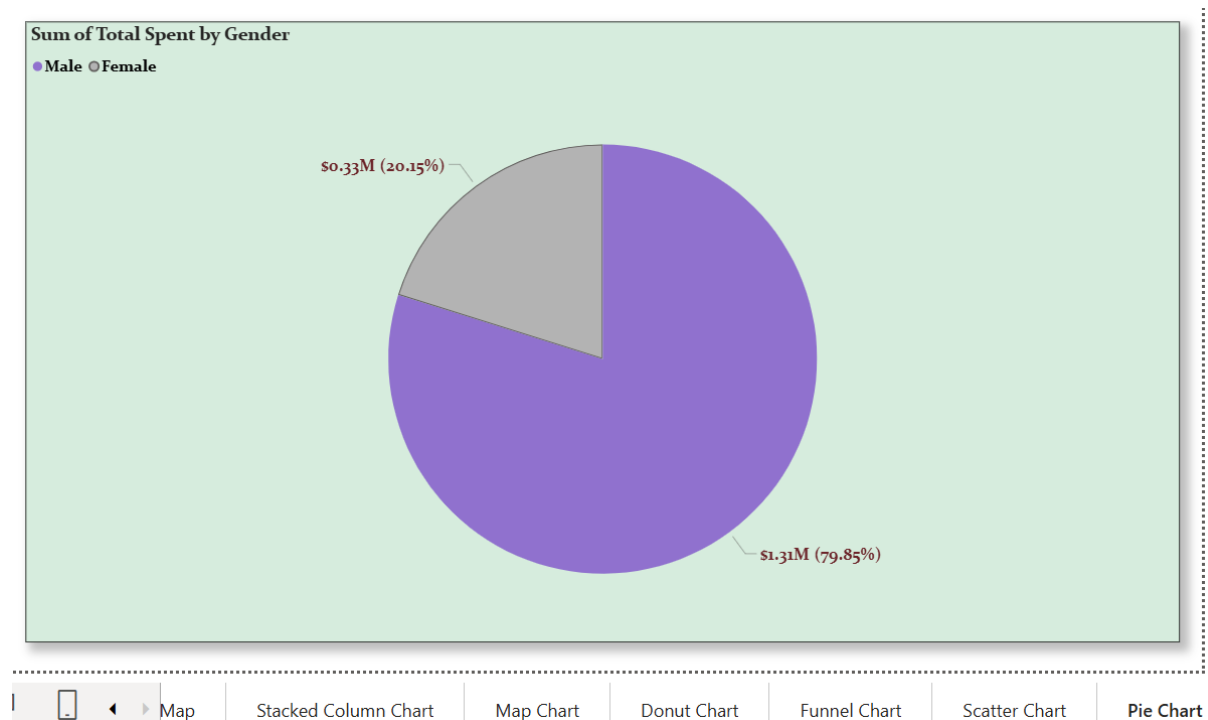
10) Sum of Total Spent by Gender Using Pie Chart:

The Below Pie Chart representing the Sales Contribution By gender.

According to the Chart Mostly Male Customers are Visiting the shop and purchasing the Products Comparing to Female Customers.

Male Customers Sales Contribution is 1.31M which is 79.85%

Female Customers Sales Contribution is 0.33M which is 20.15%



The Following Images is the Screenshot of Dashboard (Report Page)

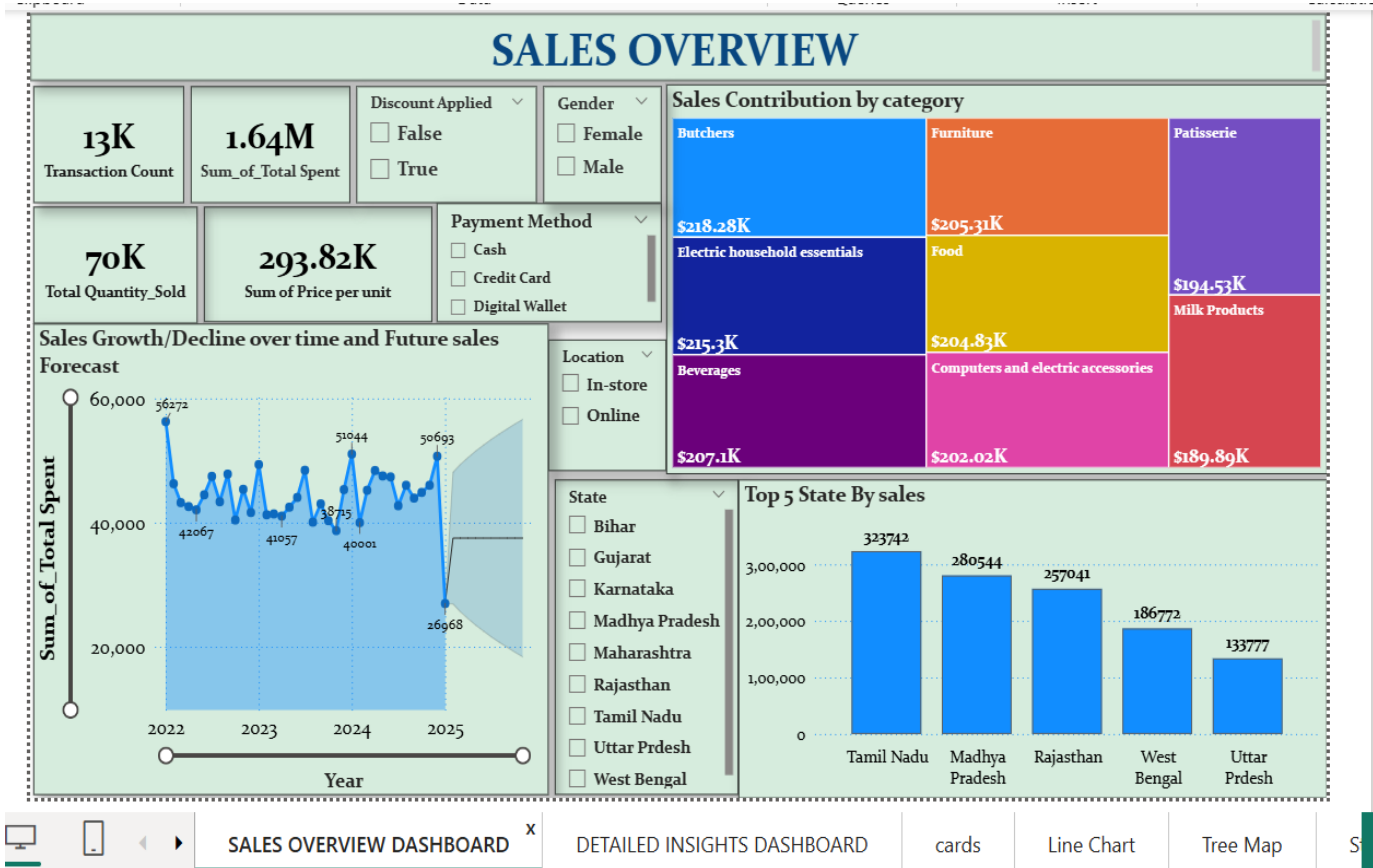
DASHBOARD 1: SALES OVERVIEW DASHBOARD: Sales Performance Overview

Focused on Cards, Slicers, Growth trends, Top State and Categories

DASHBOARD 2: DETAILED INSIGHTS: Sales Insights and Sales Analysis

Focused on Geographical analysis, Price per unit, Total spent, and Transaction Count.

DASHBOARD 1: SALES OVERVIEW DASHBOARD



CONCLUSION:

The Analysis of the Retail store Sales Dataset provided valuable insights into Sales performance, Product demand pattern, and customer buying behavior. After performing End-to-End Data Cleaning and visualization with analytical Techniques few insights identified.

Sales shows Both Growth and Decline over time, however store can expect some growth with the estimate of future performance. Some Categories Contributed well in sales, while few categories Not contributed much. Sales Happened on Multiple States and cities, and Some sates Experienced low sales where marketing and sales strategy need to be improved.

The Product Price, Category and Total spent reflecting Customer buying behavior and product demand. Mostly Male Customers are visiting the shop while female customers are not visited much where Category variety, Safety and environment needs to be improved.

Overall, the study Demonstrate the Importance of Data-Driven Decision-Making in Retail.