Title: To find closest match to value from list of values

Aim: The aim of this project is to utilize Excel functions and tools to

analyse the dataset by performing ETL operations, data preprocessing, visualization, statistical analysis, hypothesis testing, regression analysis, and dashboard creation.

Project Description:

This project will involve answering business questions using the dataset , extracting , transforming , and loading the data into Excel, applying Excel functions to automate computations and create new features, visualizing the data using various charts and pivot tables , conducting descriptive statistics to gain insights into the data , formulating and testing hypotheses , performing regression analysis to and graphs supported by various data filters predict the target feature , and building a dashboard to summarize the data with tables

Dataset name: Supermart Grocery Sales

URL: https://www.kaggle.com/datasets/mohamedharris/supermart-grocery-sales-retail-analytics-dataset?resource=download

Business Questions:

Below are the business questions which can be answered through the dataset

1. What are the top-selling products by category and sub-category?

2. How do sales vary by region, and which regions contribute the most to total sales?

3. How do sales vary over time? (by year)

4. Which city has the highest sales ?

5. What is the total profit generated from each category?

6.What is the average order value for each category?

7. Which region has highest average profit margin?

8. What are the top 5 most selling products?

- Now, let's load data into Excel from source then we will Perform ELT operations and basic pre-processing using Power Query Editor.

> Once the data is loaded, it's important to clean up the data by removing any irrelevant columns, handling missing values, and transforming the data as needed. As we didn't find any missing values or irrelevant columns, we only applied correct formatting to few columns only and there was no need of removing or adding any columns.

- Now let's apply correct formatting to columns. We changed the formatting for 'Order date' column from text to short date, for 'Discount' column from number into Percentage and for 'Sales and

Profit' column we changed into Indian Currency. After applying the formatting, the table looks like this.

	A	B	C	D	E	F	G	H	I	J	K
1	Order ID	Customer Name	Category	Sub Category	City	Order Date	Region	Sales	Discount	Profit	State
2	OD1	Harish	Oil & Masala	Masalas	Vellore	11-08-2017	North	₹1,254	12%	₹401.28	Tamil Nadu
3	OD2	Sudha	Beverages	Health Drinks	Krishnagiri	11-08-2017	South	₹749	18%	₹401.28	Tamil Nadu
4	OD3	Hussain	Food Grains	Atta & Flour	Perambalur	06-12-2017	West	₹2,360	21%	₹401.28	Tamil Nadu
5	OD4	Jackson	Fruits & Veggies	Fresh Vegetables	Dharmapuri	10-11-2016	South	₹896	25%	₹401.28	Tamil Nadu
6	OD5	Ridhesh	Food Grains	Organic Staples	Ooty	10-11-2016	South	₹2,355	26%	₹401.28	Tamil Nadu
7	OD6	Adavan	Food Grains	Organic Staples	Dharmapuri	06-09-2015	West	₹2,305	26%	₹322.70	Tamil Nadu
8	OD7	Jonas	Fruits & Veggies	Fresh Vegetables	Trichy	06-09-2015	West	₹826	33%	₹346.92	Tamil Nadu
9	OD8	Hafiz	Fruits & Veggies	Fresh Fruits	Ramanadhapuram	06-09-2015	West	₹1,847	32%	₹147.76	Tamil Nadu
10	OD9	Hafiz	Bakery	Biscuits	Tirunelveli	06-09-2015	West	₹791	23%	₹181.93	Tamil Nadu
11	OD10	Krithika	Bakery	Cakes	Chennai	06-09-2015	West	₹1,795	27%	₹484.65	Tamil Nadu
12	OD11	Ganesh	Snacks	Chocolates	Karur	06-09-2015	West	₹1,903	13%	₹437.69	Tamil Nadu
13	OD12	Yadav	Eggs, Meat & Fish	Eggs	Namakkal	06-09-2015	West	₹701	10%	₹308.44	Tamil Nadu
14	OD13	Sharon	Snacks	Cookies	Dindigul	4/15/2018	South	₹1,659	19%	₹315.21	Tamil Nadu
15	OD14	Peer	Fruits & Veggies	Fresh Vegetables	Kanyakumari	12-05-2017	West	₹1,277	25%	₹63.85	Tamil Nadu
16	OD15	Sundar	Eggs, Meat & Fish	Chicken	Kanyakumari	11/22/2016	Central	₹831	22%	₹207.75	Tamil Nadu
17	OD16	Ramesh	Oil & Masala	Edible Oil & Ghee	Krishnagiri	11/22/2016	Central	₹1,440	11%	₹100.80	Tamil Nadu
18	OD17	Alan	Bakery	Cakes	Dharmapuri	11-11-2015	Central	₹1,678	28%	₹318.82	Tamil Nadu
19	OD18	Arutra	Beverages	Health Drinks	Bodi	5/13/2015	West	₹1,617	19%	₹113.19	Tamil Nadu
20	OD19	Haseena	Eggs, Meat & Fish	Mutton	Tenkasi	8/27/2015	West	₹1,757	35%	₹386.54	Tamil Nadu
21	OD20	Verma	Beverages	Soft Drinks	Kanyakumari	8/27/2015	West	₹692	29%	₹159.16	Tamil Nadu
22	OD21	Hafiz	Beverages	Health Drinks	Vellore	8/27/2015	West	₹522	19%	₹208.80	Tamil Nadu
23	OD22	Alan	Food Grains	Dals & Pulses	Karur	12-09-2017	Central	₹948	13%	₹47.40	Tamil Nadu
24	OD23	Haseena	Beverages	Soft Drinks	Krishnagiri	12-09-2017	Central	₹707	34%	₹148.47	Tamil Nadu
25	OD24	Alan	Fruits & Veggies	Organic Vegetables	Tenkasi	7/16/2018	East	₹969	29%	₹77.52	Tamil Nadu
26	OD25	Sharon	Eggs, Meat & Fish	Eggs	Ooty	9/25/2016	West	₹1,100	11%	₹495.00	Tamil Nadu
27	OD26	Krithika	Snacks	Chocolates	Tirunelveli	1/16/2017	West	₹2,022	11%	₹202.20	Tamil Nadu
28	OD27	Muneer	Snacks	Cookies	Trichy	1/16/2017	West	₹541	12%	₹43.28	Tamil Nadu
29	OD28	Jackson	Bakery	Biscuits	Viluppuram	9/17/2016	East	₹979	22%	₹401.39	Tamil Nadu
30	OD29	Veronica	Beverages	Soft Drinks	Krishnagiri	9/17/2016	East	₹1,988	22%	₹874.72	Tamil Nadu
31	OD30	Shah	Oil & Masala	Masalas	Kanyakumari	9/17/2016	East	₹989	17%	₹69.23	Tamil Nadu
32	OD31	Ramesh	Beverages	Soft Drinks	Dharmapuri	9/17/2016	East	₹1,845	24%	₹166.05	Tamil Nadu
33	OD32	Mathew	Snacks	Noodles	Vellore	9/17/2016	East	₹1,400	11%	₹420.00	Tamil Nadu
34	OD33	Akash	Oil & Masala	Edible Oil & Ghee	Trichy	9/17/2016	East	₹2,163	26%	₹670.53	Tamil Nadu
35	OD34	Anu	Oil & Masala	Masalas	Madurai	9/17/2016	East	₹1,689	34%	₹337.80	Tamil Nadu
39	OD38	Willams	Fruits & Veggies	Organic Vegetables	Ramanadhapuram	12/27/2016	Central	₹1,354	16%	₹446.82	Tamil Nadu
40	OD39	Jonas	Fruits & Veggies	Fresh Fruits	Chennai	12/27/2016	Central	₹1,751	19%	₹280.16	Tamil Nadu
41	OD40	Sabeela	Eggs, Meat & Fish	Chicken	Cumbum	12/27/2016	Central	₹2,045	23%	₹879.35	Tamil Nadu
42	OD41	Malik	Beverages	Health Drinks	Namakkal	12/27/2016	Central	₹660	19%	₹92.40	Tamil Nadu
43	OD42	Arutra	Beverages	Soft Drinks	Salem	09-10-2018	Central	₹1,868	18%	₹616.44	Tamil Nadu
44	OD43	Sharon	Snacks	Cookies	Bodi	7/17/2017	West	₹1,017	25%	₹345.78	Tamil Nadu
45	OD44	Mathew	Snacks	Cookies	Dharmapuri	9/19/2018	South	₹2,311	20%	₹1,039.95	Tamil Nadu
46	OD45	Amrish	Bakery	Biscuits	Krishnagiri	03-11-2017	Central	₹1,663	17%	₹731.72	Tamil Nadu
47	OD46	Vince	Eggs, Meat & Fish	Eggs	Madurai	03-11-2017	Central	₹1,420	19%	₹355.00	Tamil Nadu
48	OD47	Suresh	Fruits & Veggies	Organic Vegetables	Nagercoil	10/20/2015	Central	₹1,415	31%	₹509.40	Tamil Nadu
49	OD48	Mathew	Snacks	Noodles	Chennai	6/20/2017	East	₹1,488	30%	₹610.08	Tamil Nadu
50	OD49	Ridhesh	Snacks	Noodles	Pudukottai	6/20/2017	East	₹726	18%	₹43.56	Tamil Nadu
51	OD50	Amrish	Bakery	Cakes	Ooty	4/18/2016	Central	₹1,259	15%	₹415.47	Tamil Nadu
52	OD51	Esther	Beverages	Soft Drinks	Salem	4/18/2016	Central	₹902	29%	₹153.34	Tamil Nadu
53	OD52	Shah	Beverages	Soft Drinks	Salem	4/18/2016	Central	₹2,306	30%	₹553.44	Tamil Nadu
54	OD53	Sudha	Bakery	Cakes	Theni	4/18/2016	Central	₹1,417	19%	₹99.19	Tamil Nadu
55	OD54	Yusuf	Fruits & Veggies	Fresh Vegetables	Madurai	12-11-2017	East	₹627	33%	₹181.83	Tamil Nadu
56	OD55	Esther	Fruits & Veggies	Organic Fruits	Nagercoil	12-11-2017	East	₹760	28%	₹60.80	Tamil Nadu

- Apply Excel Functions to automate Computation, simulation of new feature
    - Below we are going to use COUNTIFS function, so that we can get the number of people who have ordered amount greater than 1000 rupees by region wise.
    - Also how you can classify orders into different value segments using IF statements based on the total sales amount in your dataset (low<600, medium<800 and high<1000).
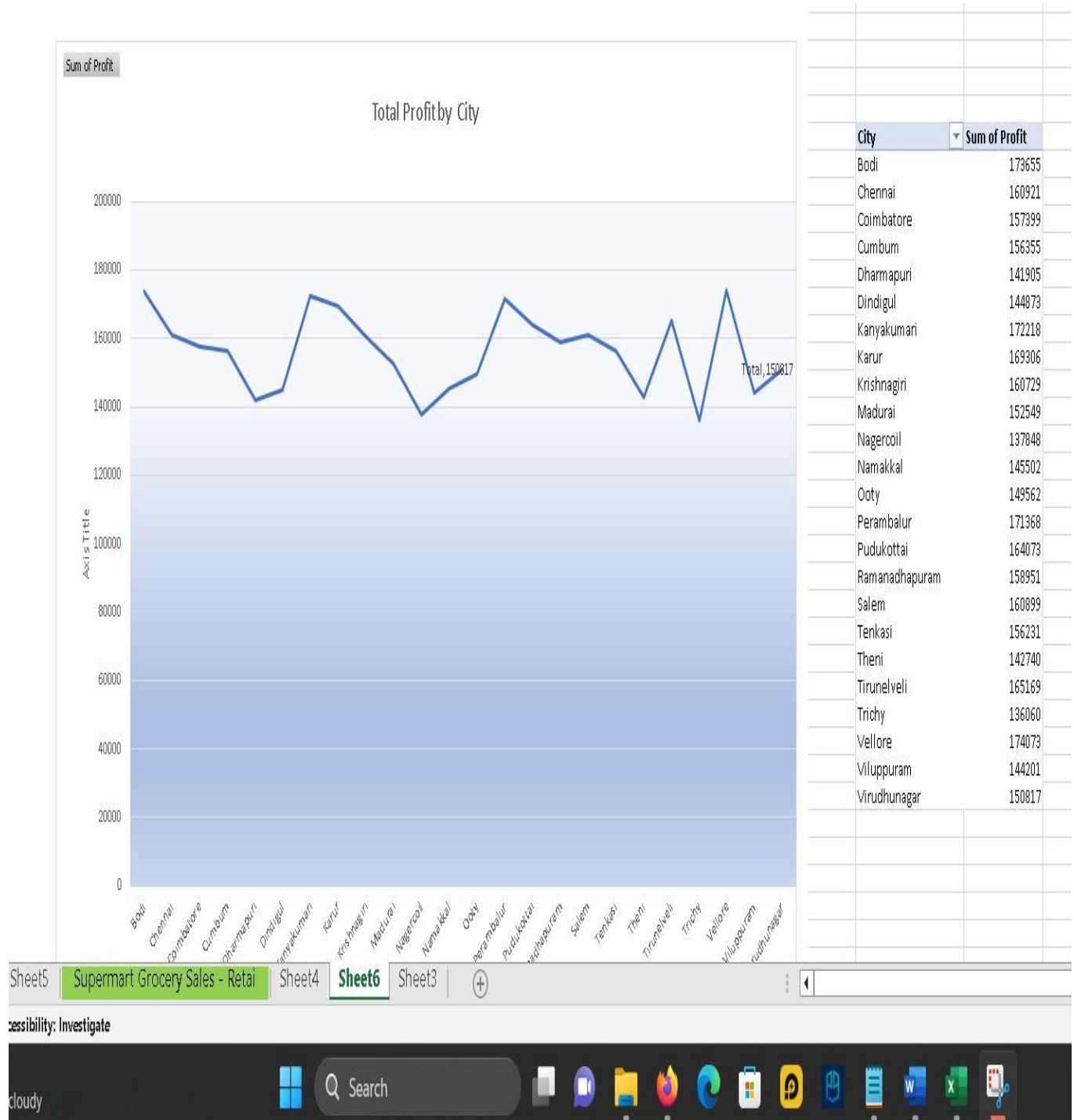

	Order ID	Customer Name	Category	Sub Category	City	Order Date	Region	Sales	Discount	Profit	State	Value Segment		REGIONS	SALES	PEOPLE
1																
2	OD1	Harish	Oil & Masala	Masalas	Vellore	11-08-2017	North	₹791	12%	₹401.28	Tamil Nadu	Medium				
3	OD2	Sudha	Beverages	Health Drinks	Krishnagiri	11-08-2017	South	₹749	18%	₹401.28	Tamil Nadu	Medium				
4	OD3	Hussain	Food Grains	Atta & Flour	Perambalur	06-12-2017	West	₹2,360	21%	₹401.28	Tamil Nadu	High				
5	OD4	Jackson	Fruits & Veggies	Fresh Vegetables	Dharmapuri	10-11-2016	South	₹896	25%	₹401.28	Tamil Nadu	High				
6	OD5	Ridhesh	Food Grains	Organic Staples	Ooty	10-11-2016	South	₹2,355	26%	₹401.28	Tamil Nadu	High				
7	OD6	Adavan	Food Grains	Organic Staples	Dharmapuri	06-09-2015	West	₹2,305	26%	₹322.70	Tamil Nadu	High				
8	OD7	Jonas	Fruits & Veggies	Fresh Vegetables	Trichy	06-09-2015	West	₹826	33%	₹346.92	Tamil Nadu	High				
9	OD8	Hafiz	Fruits & Veggies	Fresh Fruits	Ramanadhapuram	06-09-2015	West	₹1,847	32%	₹147.76	Tamil Nadu	High				
10	OD9	Hafiz	Bakery	Biscuits	Tirunelveli	06-09-2015	West	₹791	23%	₹181.93	Tamil Nadu	Medium		REGIONS	SALES	PEOPLE
11	OD10	Krithika	Bakery	Cakes	Chennai	06-09-2015	West	₹1,795	27%	₹484.65	Tamil Nadu	High		EAST	>1000	2138
12	OD11	Ganesh	Snacks	Chocolates	Karur	06-09-2015	West	₹1,903	13%	₹437.69	Tamil Nadu	High		WEST	>1000	2415
13	OD12	Yadav	Eggs, Meat & Fish	Eggs	Namakkal	06-09-2015	West	₹701	10%	₹308.44	Tamil Nadu	Medium		NORTH	>1000	0
14	OD13	Sharon	Snacks	Cookies	Dindigul	4/15/2018	South	₹1,659	19%	₹315.21	Tamil Nadu	High		SOUTH	>1000	1208
15	OD14	Peer	Fruits & Veggies	Fresh Vegetables	Kanyakumari	12-05-2017	West	₹1,277	25%	₹63.85	Tamil Nadu	High				
16	OD15	Sundar	Eggs, Meat & Fish	Chicken	Kanyakumari	11/22/2016	Central	₹831	22%	₹207.75	Tamil Nadu	High				
17	OD16	Ramesh	Oil & Masala	Edible Oil & Ghee	Krishnagiri	11/22/2016	Central	₹1,440	11%	₹100.80	Tamil Nadu	High				
18	OD17	Alan	Bakery	Cakes	Dharmapuri	11-11-2015	Central	₹1,678	28%	₹318.82	Tamil Nadu	High				
19	OD18	Arutra	Beverages	Health Drinks	Bodi	5/13/2015	West	₹1,617	19%	₹113.19	Tamil Nadu	High				
20	OD19	Haseena	Eggs, Meat & Fish	Mutton	Tenkasi	8/27/2015	West	₹1,757	35%	₹386.54	Tamil Nadu	High				
21	OD20	Verma	Beverages	Soft Drinks	Kanyakumari	8/27/2015	West	₹692	29%	₹159.16	Tamil Nadu	Medium				
22	OD21	Hafiz	Beverages	Health Drinks	Vellore	8/27/2015	West	₹522	19%	₹208.80	Tamil Nadu	Low				
23	OD22	Alan	Food Grains	Dals & Pulses	Karur	12-09-2017	Central	₹948	13%	₹47.40	Tamil Nadu	High				
24	OD23	Haseena	Beverages	Soft Drinks	Krishnagiri	12-09-2017	Central	₹707	34%	₹148.47	Tamil Nadu	Medium				
25	OD24	Alan	Fruits & Veggies	Organic Vegetables	Tenkasi	7/16/2018	East	₹969	29%	₹77.52	Tamil Nadu	High				
26	OD25	Sharon	Eggs, Meat & Fish	Eggs	Ooty	9/25/2016	West	₹1,100	11%	₹495.00	Tamil Nadu	High				
27	OD26	Krithika	Snacks	Chocolates	Tirunelveli	1/16/2017	West	₹2,022	11%	₹202.20	Tamil Nadu	High				
28	OD27	Muneer	Snacks	Cookies	Trichy	1/16/2017	West	₹541	12%	₹43.28	Tamil Nadu	Low				
29	OD28	Jackson	Bakery	Biscuits	Viluppuram	9/17/2016	East	₹979	22%	₹401.39	Tamil Nadu	High				
30	OD29	Veronica	Beverages	Soft Drinks	Krishnagiri	9/17/2016	East	₹1,988	22%	₹874.72	Tamil Nadu	High				
31	OD30	Shah	Oil & Masala	Masalas	Kanyakumari	9/17/2016	East	₹989	17%	₹69.23	Tamil Nadu	High				
32	OD31	Ramesh	Beverages	Soft Drinks	Dharmapuri	9/17/2016	East	₹1,845	24%	₹166.05	Tamil Nadu	High				
33	OD32	Mathew	Snacks	Noodles	Vellore	9/17/2016	East	₹1,400	11%	₹420.00	Tamil Nadu	High				
34	OD33	Akash	Oil & Masala	Edible Oil & Ghee	Trichy	9/17/2016	East	₹2,163	26%	₹670.53	Tamil Nadu	High				
35	OD34	Anu	Oil & Masala	Masalas	Madurai	9/17/2016	East	₹1,689	34%	₹337.80	Tamil Nadu	High				
36	OD35	Anu	Beverages	Health Drinks	Vellore	10/19/2018	Central	₹809	32%	₹80.90	Tamil Nadu	High				
37	OD36	Saboola	Eggs, Meat & Fish	Eggs	Tirunelveli	12-08-2017	Central	₹1,265	22%	₹569.25	Tamil Nadu	High				

Supermart Grocery Sales - Retail   Sheet3

- Visualize the data columns using charts such as bar chart, line chart, pie chart, tree map.
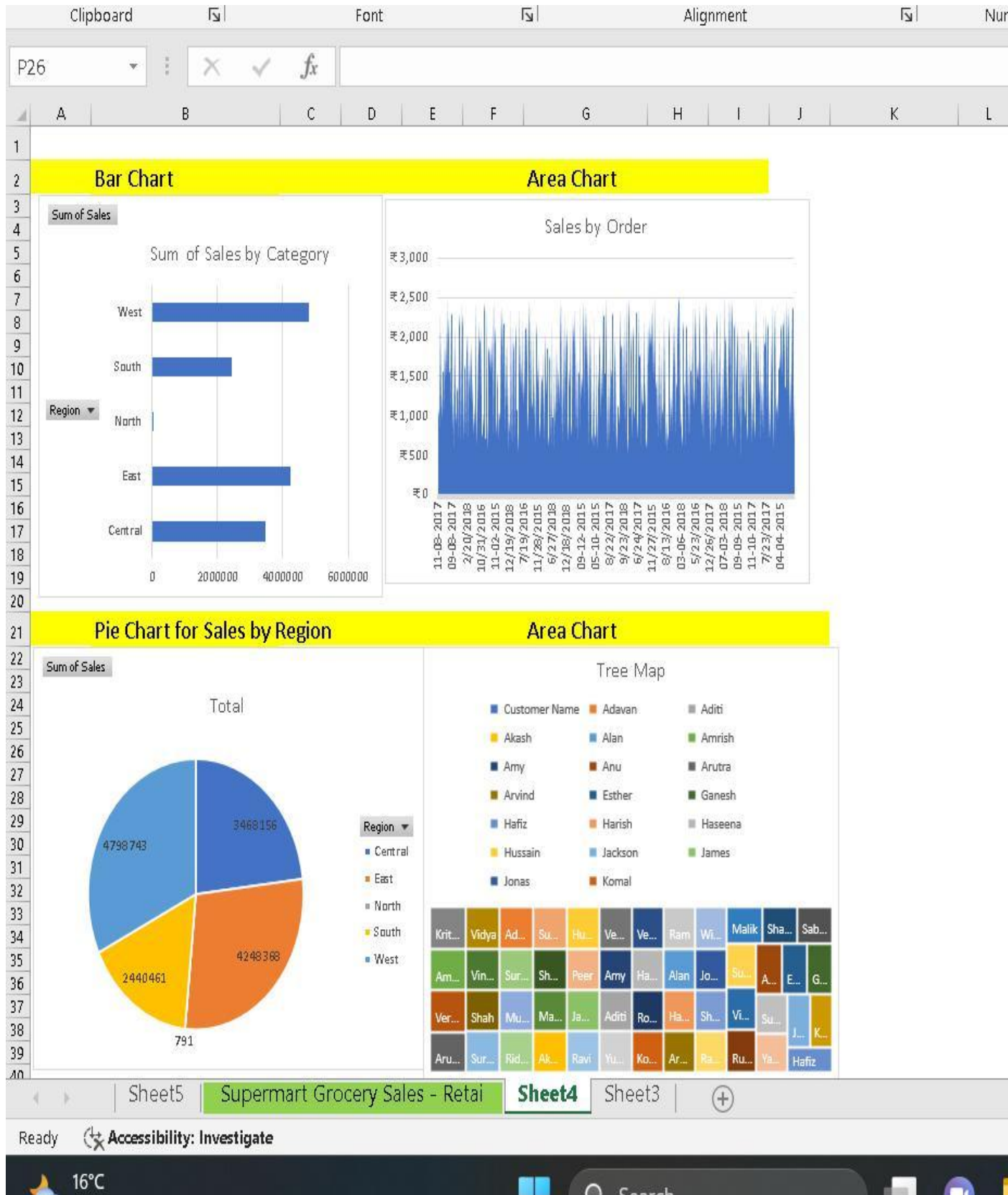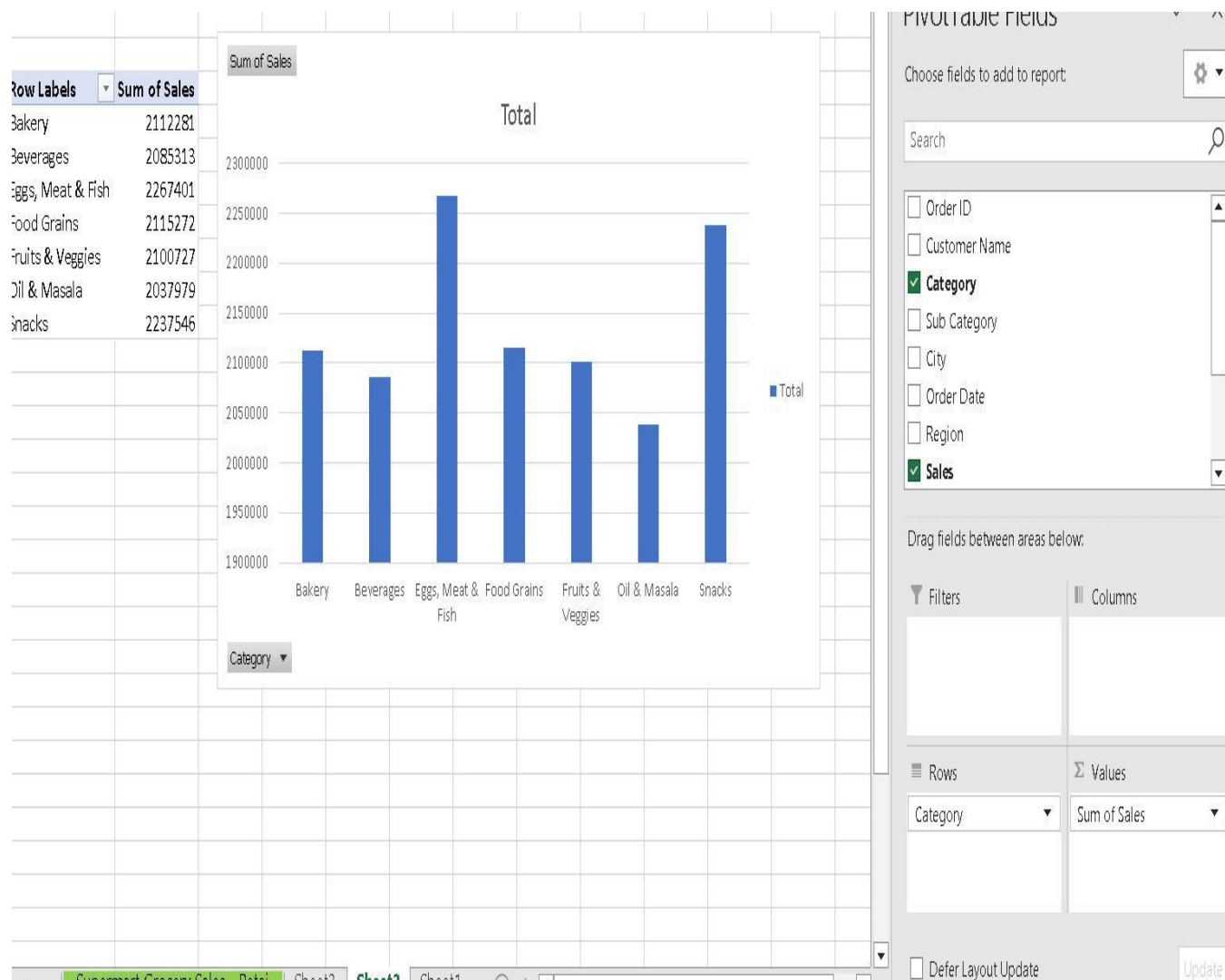
- Prepare different pivot tables to summarize the data, visualize the pivot table using pivot Chart. In the pivot table below we have put Profit in 'Values' and city in 'Column'.

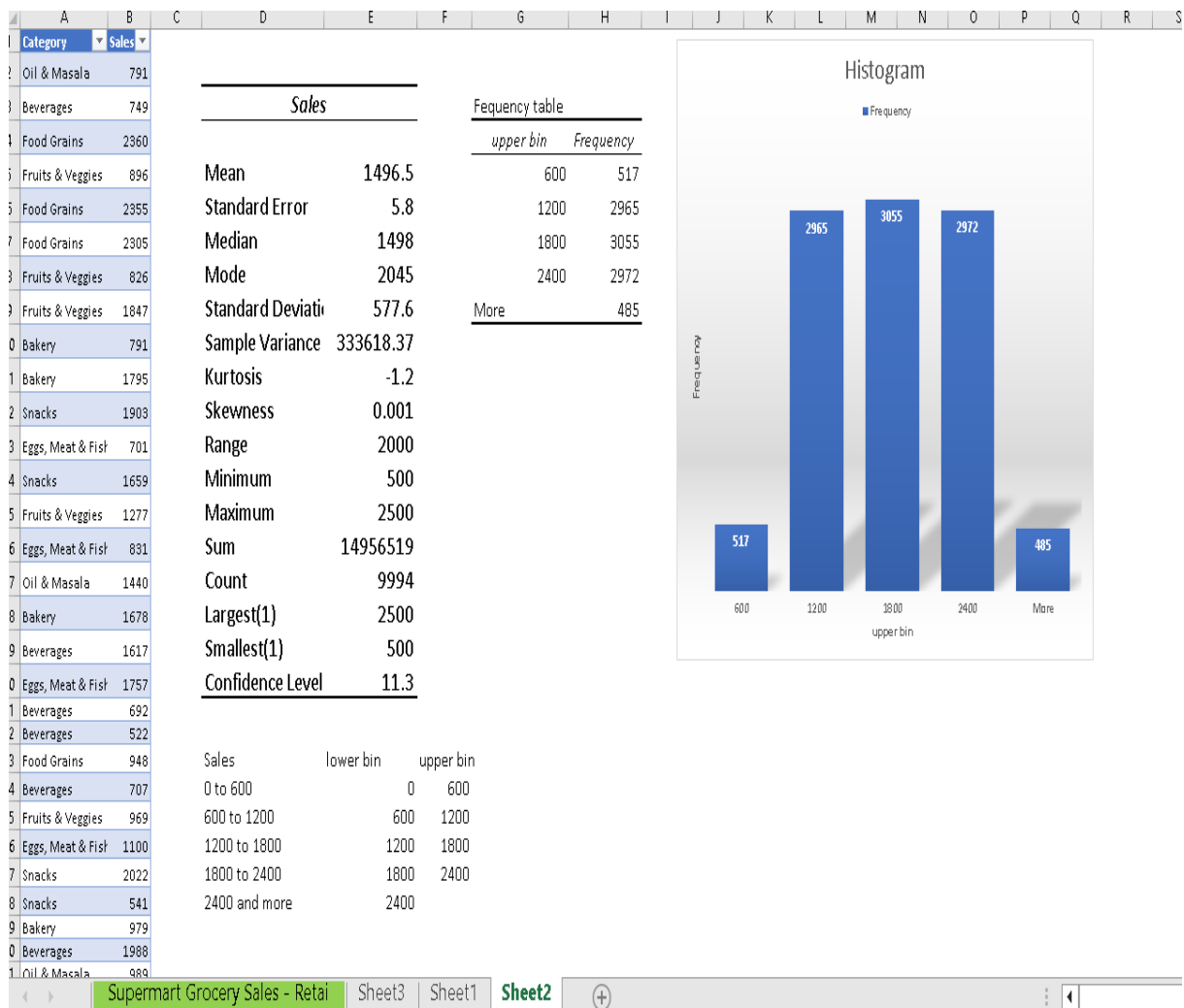Let's create another pivot chart along with pivot table in the next slide.

City	Sum of Profit
Bodi	173655
Chennai	160921
Coimbatore	157399
Cumbum	156355
Dharmapuri	141905
Dindigul	144873
Kanyakumari	172218
Karur	169306
Krishnagiri	160729
Madurai	152549
Nagercoil	137848
Namakkal	145502
Ooty	149562
Perambalur	171368
Pudukottai	164073
Ramanadhapuram	158951
Salem	160899
Tenkasi	156231
Theni	142740
Tirunelveli	165169
Trichy	136060
Vellore	174073
Viluppuram	144201
Virudhunagar	150817

Below is another pivot chart with pivot table for Sales by Category.

Row Labels	Sum of Sales
Bakery	2112281
Beverages	2085313
Eggs, Meat & Fish	2267401
Food Grains	2115272
Fruits & Veggies	2100727
Oil & Masala	2037979
Snacks	2237546

- Perform Descriptive Statistics on the dataset to get basic statistical distribution or insights from data
  - ➢ To perform Descriptive Statistics on the dataset, first we need to install Data Analysis Toolpak. In this section, we are going to create a Summary table, histogram and frequency table. First of all enable Data Analysis Toolpak by going to file->options ->Manage dropdown and select Data Analysis from Excel Add-ins.

Category	Sales
Oil & Masala	791
Beverages	749
Food Grains	2360
Fruits & Veggies	896
Food Grains	2355
Food Grains	2305
Fruits & Veggies	826
Fruits & Veggies	1847
Bakery	791
Bakery	1795
Snacks	1903
Eggs, Meat & Fish	701
Snacks	1659
Fruits & Veggies	1277
Eggs, Meat & Fish	831
Oil & Masala	1440
Bakery	1678
Beverages	1617
Eggs, Meat & Fish	1757
Beverages	692
Beverages	522
Food Grains	948
Beverages	707
Fruits & Veggies	969
Eggs, Meat & Fish	1100
Snacks	2022
Snacks	541
Bakery	979
Beverages	1988
Oil & Masala	989

Sales

Mean	1496.5
Standard Error	5.8
Median	1498
Mode	2045
Standard Deviation	577.6
Sample Variance	333618.37
Kurtosis	-1.2
Skewness	0.001
Range	2000
Minimum	500
Maximum	2500
Sum	14956519
Count	9994
Largest(1)	2500
Smallest(1)	500
Confidence Level	11.3

Fequency table

upper bin	Frequency
600	517
1200	2965
1800	3055
2400	2972
More	485

Sales	lower bin	upper bin
0 to 600	0	600
600 to 1200	600	1200
1200 to 1800	1200	1800
1800 to 2400	1800	2400
2400 and more	2400	

Sheets: Supermart Grocery Sales - Retai | Sheet3 | Sheet1 | **Sheet2**

- Form hypothesis for your data and perform Hypothesis Testing (Analytic Toolpak) between different variables to validate whether it is valid or not.
  - ➢ In this example, we are going to test Null hypothesis between two categorical column 'Region' and 'Category'. We are going to use ChiTest function, so we can Calculate P value in order to validate whether it is valid or not.

Expected Cell Frequency

$$Expected\ cell\ frequency = \frac{(row\ total) \times (column\ total)}{(grand\ total\ of\ all\ cells)}$$

Row Labels	Bakery	Beverages	Eggs, Meat & Fish	Food Grains	Fruits & Veggies	Oil & Masala	Snacks	Grand Total
Central	309	343	331	310	355	300	375	2323
East	392	385	449	393	384	416	429	2848
North						1		1
South	240	227	233	232	234	202	251	1619
West	472	445	477	463	445	442	459	3203
Grand Total	**1413**	**1400**	**1490**	**1398**	**1418**	**1361**	**1514**	**9994**

Row Labels	Bakery	Beverages	Eggs, Meat & Fish	Food Grains	Fruits & Veggies	Oil & Masala	Snacks	Grand Total
Central	328.437	325.415249	346.3348009	324.95037	329.5991595	316.35011	351.913348	2323
East	402.6639984	398.959376	424.6067641	398.389434	404.0888533	387.845507	431.446068	2848
North	0.141384831	0.14008405	0.149089454	0.13988393	0.141885131	0.13618171	0.15149089	1
South	228.9020412	226.796078	241.3758255	226.472083	229.7120272	220.478187	245.263758	1619
West	452.8556134	448.689214	477.5335201	448.048229	454.4580748	436.190014	485.225335	3203
Grand Total	1413	1400	1490	1398	1418	1361	1514	9994

CHITEST VALUE(P)	0.443267571

H0: There is no significant association between Gender and Occupation

P value < 0.05	Reject H0
P Value > 0.05	Accept H0

P Value	0.44326757

As P value is greater , so Null hypothesis(H0) gets Accepted.

- In the above table, we can see that we have two tables, one table consists of actual frequency of sales placed in each region and in second table, it shows expected frequency. Then we Calculated the ChiTest value (p value) and finally we got our conclusion that, there is no significant association between Gender and Region (these two variable).

- Perform regression analysis using the (Analytic Toolpak) input features and predict the target feature in the dataset

  ➢ In this step, first select the data range including headers. In your case, it would be from A1 to K20.Go to the "Data" tab, then click on "Data Analysis" in the "Analysis" group and choose "Regression" from the list of analysis tools and click "OK."In the Regression dialog box:Input Y Range: Select the column containing the target variable you want to predict (in this case, "Sales") & Input X Range: Discount & profit .

J	K	L	M	N	O	P	Q	R	S	T	U	V
ofit	State											
401.28	Tamil Nadu											
401.28	Tamil Nadu											
401.28	Tamil Nadu		SUMMARY OUTPUT									
401.28	Tamil Nadu											
401.28	Tamil Nadu		Regression Statistics									
322.7	Tamil Nadu		Multiple R	0.605125561								
346.92	Tamil Nadu		R Square	0.366176945								
147.76	Tamil Nadu		Adjusted R Square	0.366050066								
181.93	Tamil Nadu		Standard Error	459.8884035								
484.65	Tamil Nadu		Observations	9994								
437.69	Tamil Nadu											
308.44	Tamil Nadu		ANOVA									
315.21	Tamil Nadu			df	SS	MS	F	Significance F				
63.85	Tamil Nadu		Regression	2	1220778413	610389206.4	2886.037223	0				
207.75	Tamil Nadu		Residual	9991	2113069961	211497.3437						
100.8	Tamil Nadu		Total	9993	3333848374							
318.82	Tamil Nadu											
113.19	Tamil Nadu			Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%	
386.54	Tamil Nadu		Intercept	959.4636579	16.382415	58.56668005	0	927.3508242	991.5764916	927.3508242	991.5764916	
159.16	Tamil Nadu		X Variable 1	-41.17873199	61.6391261	-0.66806158	0.504109717	-162.0038366	79.64637259	-162.0038366	79.64637259	
208.8	Tamil Nadu		X Variable 2	1.457271388	0.019181906	75.97114748	0	1.419670988	1.494871787	1.419670988	1.494871787	
47.4	Tamil Nadu											
148.47	Tamil Nadu											
77.52	Tamil Nadu											
495	Tamil Nadu											
202.2	Tamil Nadu											
43.28	Tamil Nadu											
401.39	Tamil Nadu											
874.72	Tamil Nadu											
69.23	Tamil Nadu											
166.05	Tamil Nadu											
420	Tamil Nadu											
670.53	Tamil Nadu											
337.8	Tamil Nadu											
80.9	Tamil Nadu											

Multiple R: The correlation coefficient between the dependent variable (Sales) and the independent variables (Discount and Profit) is approximately 0.605. This indicates a moderate positive linear relationship between the variables.

R Square: The coefficient of determination is approximately 0.366. This means that approximately 36.6% of the variation in sales can be explained by the variation in Discount and Profit.

Adjusted R Square: This is the R Square adjusted for the number of predictors in the model. It's approximately 0.366, similar to R Square.

Standard Error: The standard error of the estimate is approximately 459.89. It measures the average distance between the actual sales values and the predicted sales values by the model.

**Observations**: The number of data points used in the analysis is 9994.

- ## Evaluate the regression mechanism using different measures such as MSE, RMSE, R2 score etc

1. In this step, we added four new columns, so that we can calculate MSE, RMSE and R2 values. **Calculate Residuals**: Subtract the predicted sales values from the actual sales values to compute the residuals.
2. **Square Residuals**: Square each residual value.
3. **Calculate MSE**: Find the average of the squared residuals. Our MSE value is 2,11,433.86
4. **Calculate RMSE**: Take the square root of MSE. Our RMSE value is 459.89
5. **Calculate R2**: And finally, our R2 value is 0.366. The formula to calculate R2 value is 1-(SSR/SST)

6. Where SSR is the sum of squared residuals and SST is the total sum of squares (you which is computed by squared differences between each actual value and the mean of all the actual values, and then sum up these squared differences.

- **Build a Dashboard for the dataset to summarize data with tables, graphs supported with different data filters**
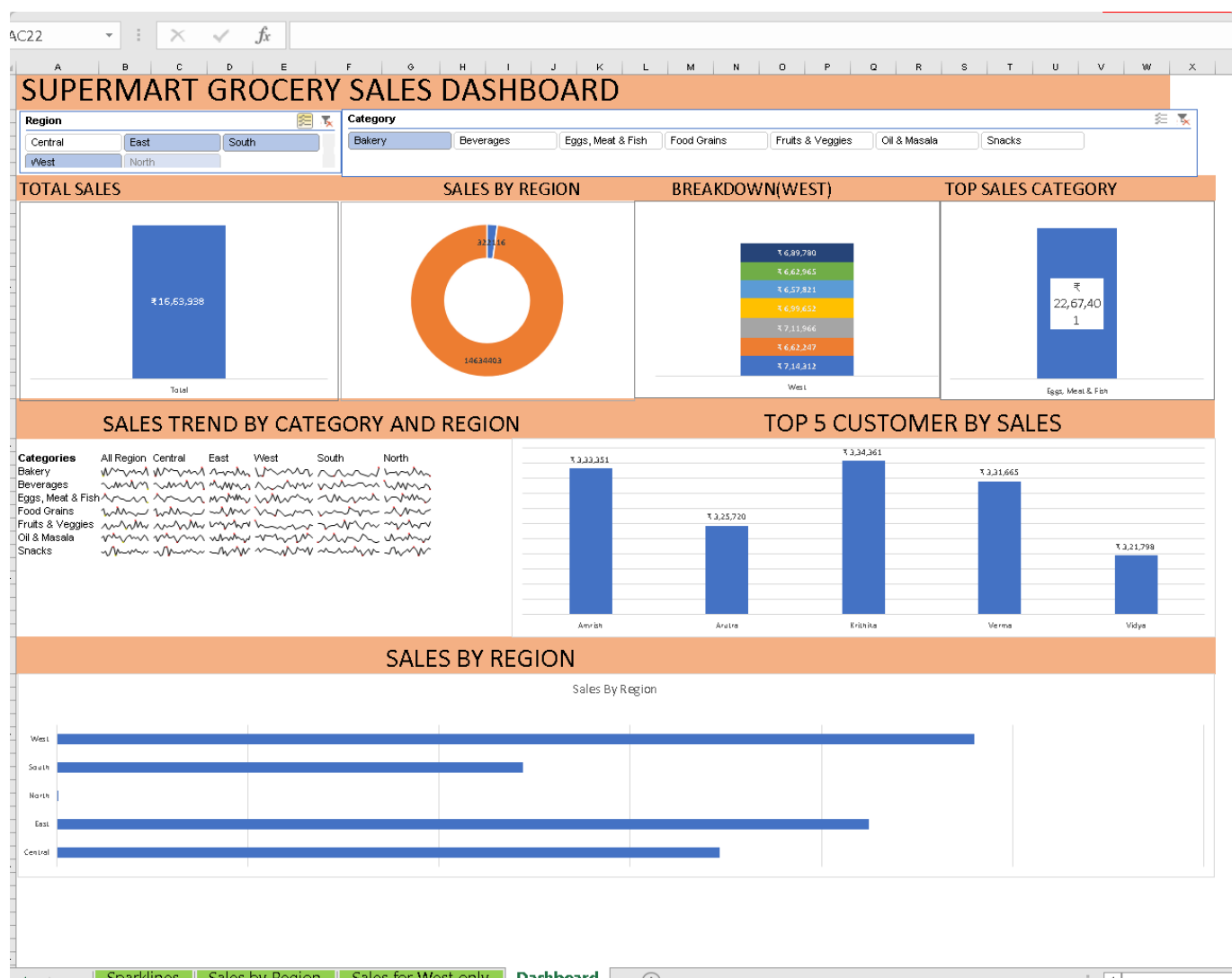
As you can see from above picture, we have created a interactive dashboard which includes of different types of charts and sparklines. We have also used two Slicers i.e. Region and Category, and have also turned on 'Report Connection' from slicers tab, to establish connection through various pivot tables.

1. What are the top-selling products by category and sub-category?

First of all, we are going to set up the PivotTable with the Category and Sub-Category fields in the Rows area and the Sales field in the Values area, you'll see a table displaying the total sales for each combination of Category and Sub-Category.

To identify the top-selling products by Category and Sub-Category, we can use the following steps: Click on any cell within the PivotTable. Go to the "Data" tab in the Excel ribbon. Click on the "Sort Highest to Smallest ". Now we have have sorted sub-categories in Ascending order where higher value comes first and have highlighted them in yellow colour.

	Queries & Connections		Sort & Filter		Data Tools	

684083

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3	Row Labels	Sum of Sales											
4	Bakery	₹ 21,12,281											
5	Breads & Buns	₹ 7,42,586											
6	Cakes	₹ 6,85,612											
7	Biscuits	₹ 6,84,083											
8	Beverages	₹ 20,85,313											
9	Health Drinks	₹ 10,51,439											
10	Soft Drinks	₹ 10,33,874											
11	Eggs, Meat & Fish	₹ 22,67,401											
12	Mutton	₹ 6,11,200											
13	Eggs	₹ 5,75,156											
14	Fish	₹ 5,60,548											
15	Chicken	₹ 5,20,497											
16	Food Grains	₹ 21,15,272											
17	Organic Staples	₹ 5,58,929											
18	Atta & Flour	₹ 5,34,649											
19	Dals & Pulses	₹ 5,23,371											
20	Rice	₹ 4,98,323											
21	Fruits & Veggies	₹ 21,00,727											
22	Fresh Fruits	₹ 5,51,212											
23	Fresh Vegetables	₹ 5,25,842											
24	Organic Vegetables	₹ 5,20,271											
25	Organic Fruits	₹ 5,03,402											
26	Oil & Masala	₹ 20,37,979											
27	Masalas	₹ 6,97,017											
28	Spices	₹ 6,72,876											
29	Edible Oil & Ghee	₹ 6,68,086											
30	Snacks	₹ 22,37,546											
31	Cookies	₹ 7,68,213											
32	Noodles	₹ 7,35,435											
33	Chocolates	₹ 7,33,898											
34													
35													

In the above example, we can see that for bakery Category, Bread & Buns have most sales with the amount figure of 7,42,586 rupees. Similarly , in Beverages category we

have Health Drinks as top selling sub-category .Also Eggs, Meat & Fish is the highest category by sales.

## 2. How do sales vary by region, and which regions contribute the most to total sales?

At first, we create a pivot table for the dataset. In the PivotTable Fields pane that appears on the right, drag the "Region" field to the Rows area and the "Sales" field to the Values area. Then we are going to use a PivotChart to visualize the sales distribution by region.
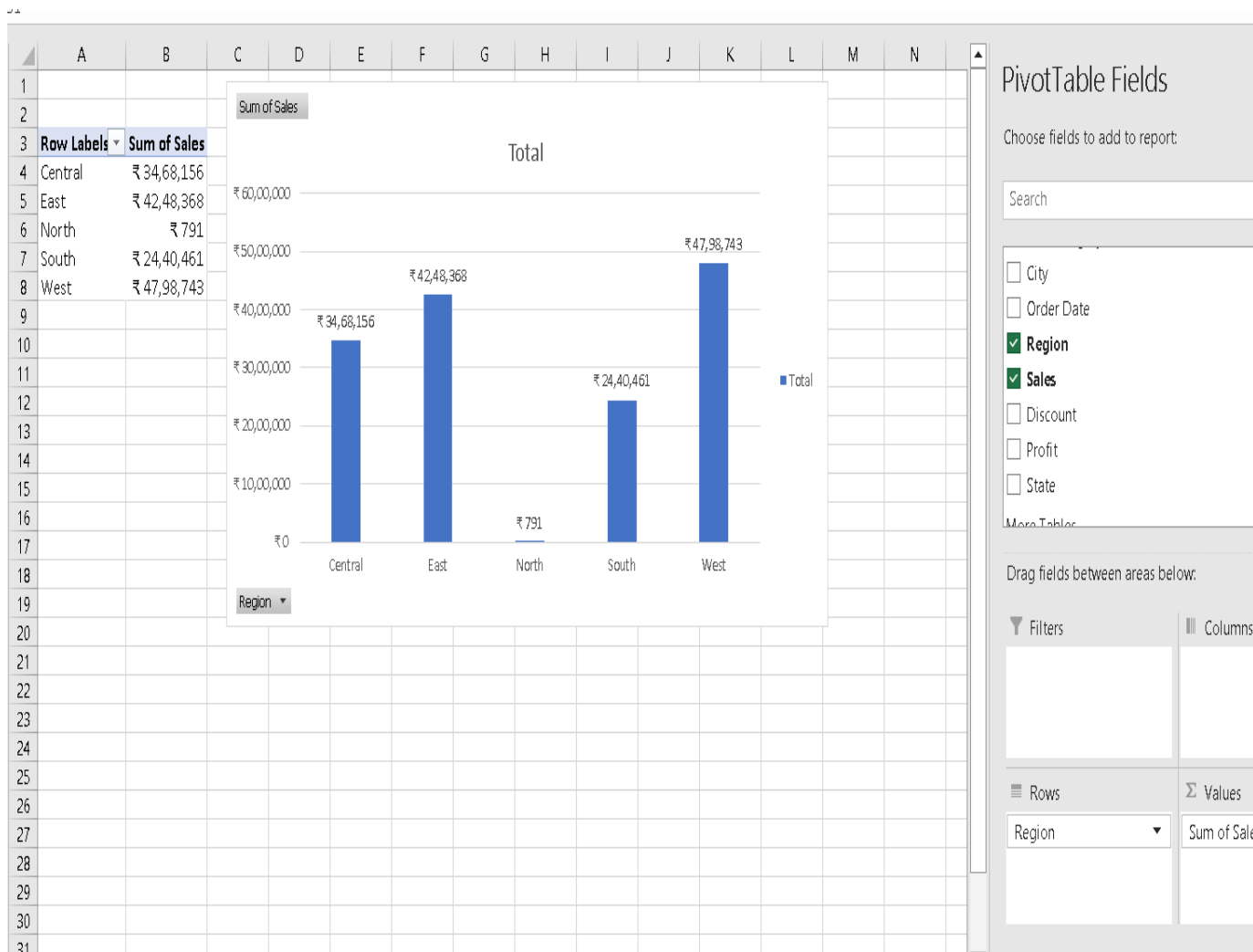
From the above chart we can clearly see that West region has the most sales with the value of ₹ 47,98,743. Likewise, North region has the least sales with the value of ₹ 791.

3. How do Sales vary over time? (Year)

For this we are going to create a line chart.
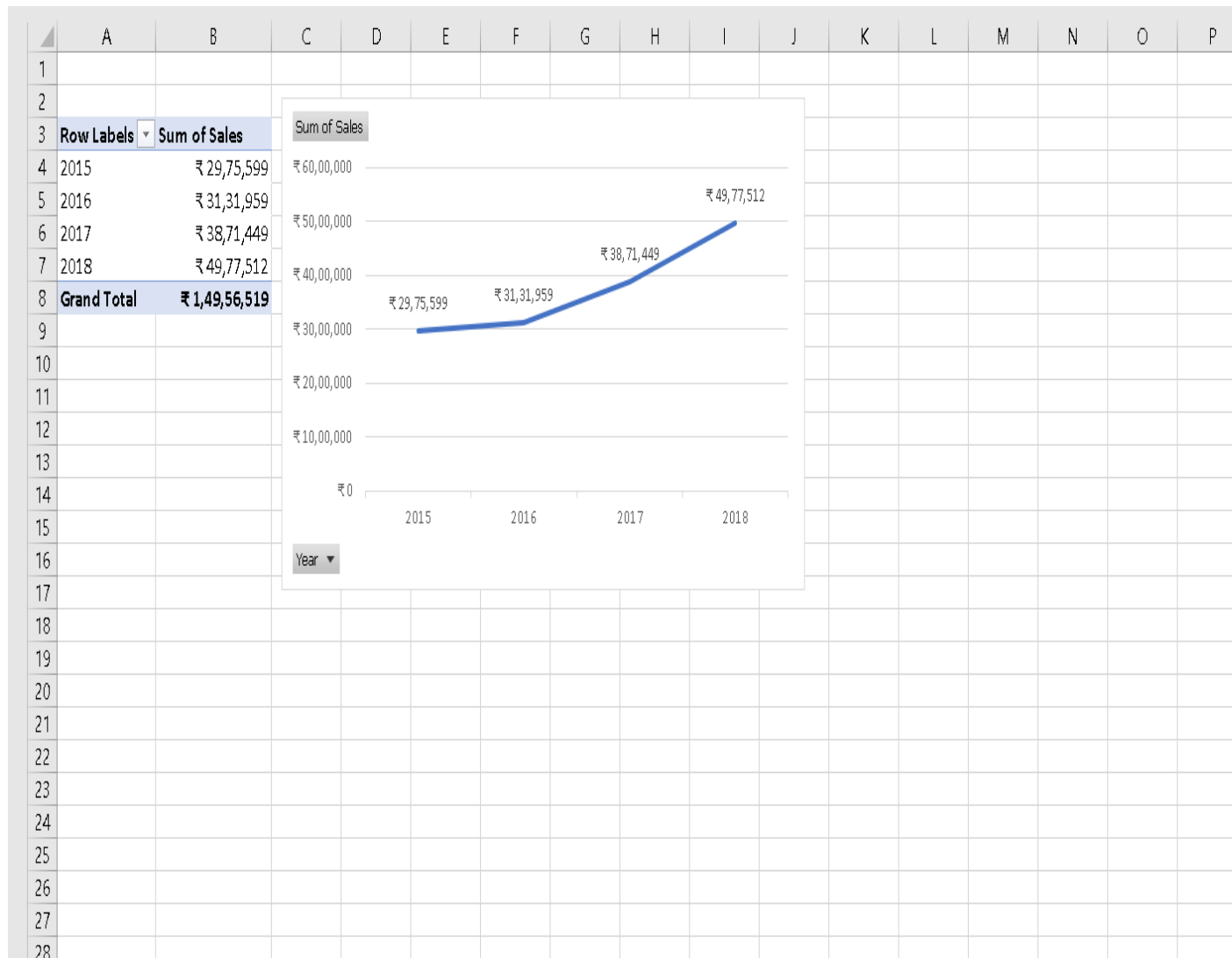
> ➢ Total sales for the year 2015 amounted to ₹2,975,599. This marks the starting point of the observed sales trend. Sales showed a slight increase in 2016, reaching a total of ₹3,131,959. This indicates a modest growth in sales compared to the previous year. The year 2017 experienced a more significant increase in sales, with total sales amounting to ₹3,871,449. This suggests a notable uptrend in sales performance compared to the previous years. Sales continued to rise in 2018, reaching the highest total sales figure of ₹4,977,512 among the observed years. This indicates a substantial growth in sales, demonstrating the highest level of performance during the period.

4. Which city has highest sales?

To determine which city has highest sales, we are going to filter top 10 cities by sum of sales and then we will be representing the data in clustered column.
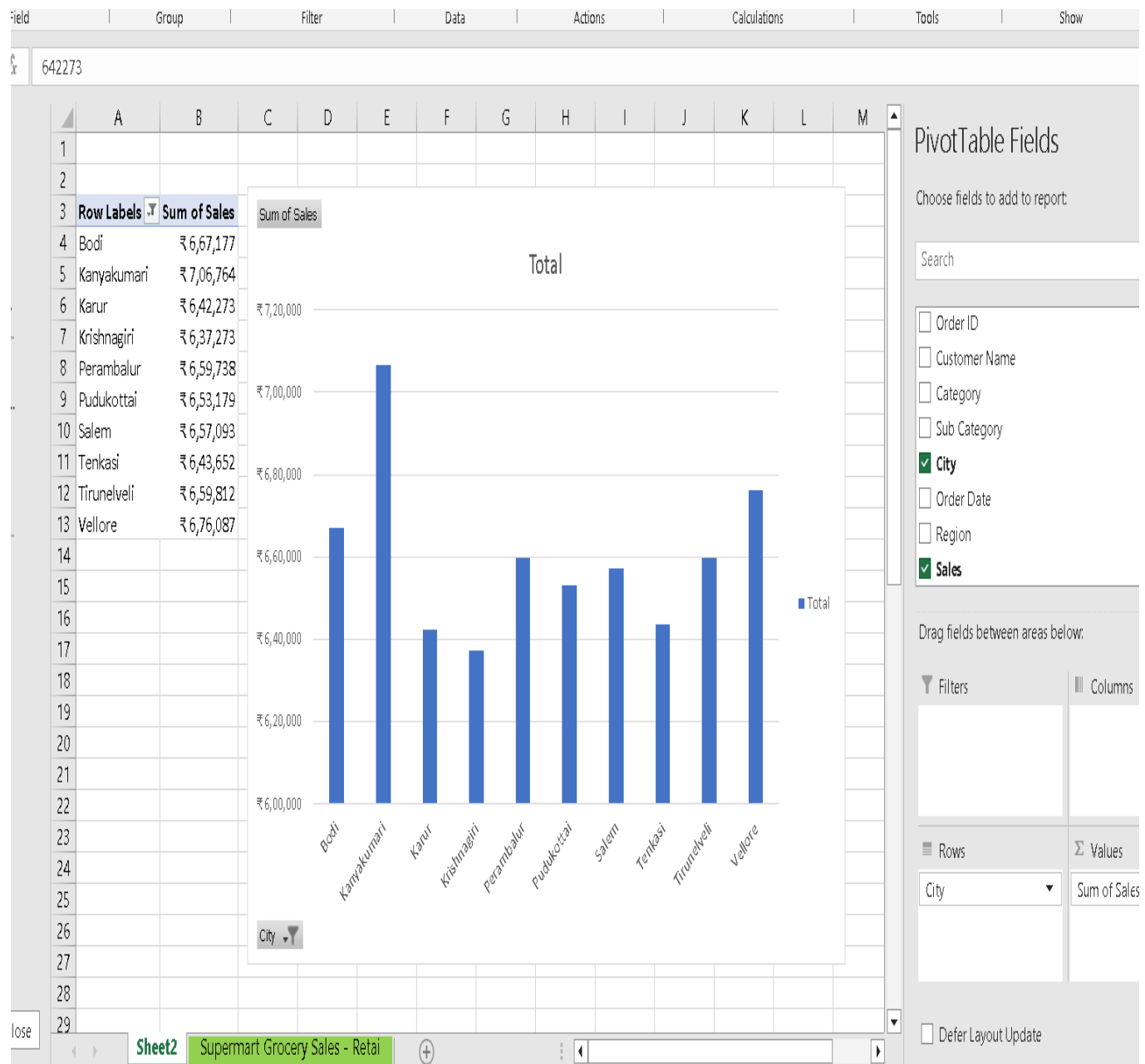
5. What is the total profit generated from each category?

For this we are again going to use pivot table where we place profit in 'Values' and Category in 'Rows'. And finally, we will be replacing the data in pie chart.
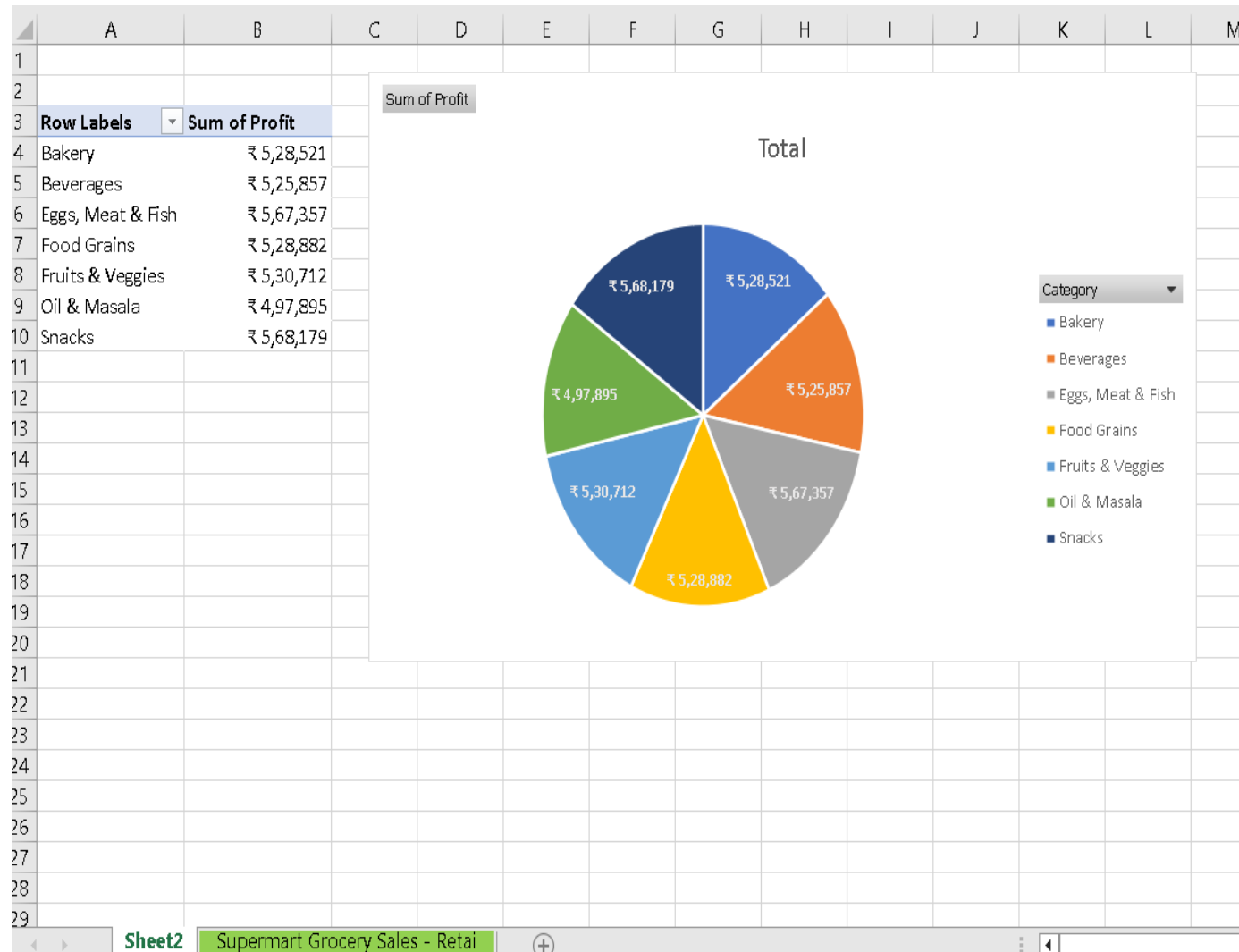
From above we can say that the Bakery category has generated a total profit of 528,521. This indicates the profitability of products such as bread, cakes, and pastries within this category. Beverages have contributed a total profit of 525,857. This includes profits from various drinks such as tea, coffee, juices, and soft drinks. The category of Eggs, Meat & Fish has generated a total profit of 567,357. This category likely includes profits from items such as eggs, poultry, meat, and seafood. Food Grains have contributed a total profit of 528,882. This category encompasses profits from staples such as rice, wheat, grains, and pulses.

6. What is the average order value for each category?

In this step, we are going to create a new column named Average order value and in the first cell we are going to calculate average order by using formula Sales*(1-Dis %) and we are going to drag it down to fill other cells. Finally, we will be placing average order value is 'Values' and Category in 'Rows'. Also, in the "Value Field Settings" dialog box, change the "Summarize Values By" option to "Average".

7. Which region has highest average profit margin?

First of all, we need to create a new column (let's say "Profit Margin") and use the formula =(Profit / Sales) * 100 to calculate the profit margin for each order. This formula calculates profit margin as a percentage of sales. Then we will insert a PivotTable (go to Insert > PivotTable). Drag the "Region" column to the Rows area and the "Profit Margin" column to the Values area. I n the "Value Field Settings" dialog box, change the "Summarize Values By" option to "Average". Click "OK".

8. What are the top 5 most selling products overall?

At first, we need to create a PivotTable using our dataset. Then drag the "Product" column to the Rows area. Drag the "Sales" column to the Values area. And then we insert a new column next to the PivotTable to calculate the rank. In the first cell of the new column, we need to enter the formula =RANK.EQ(B2,$B$2:$B$100,0). At last Filter the rank column to show only the top 5 ranks. This will display the top 5 best-selling products overall.

To find closest match to a value from a list of values

To find the closest match to a specific value in the "Sales" column, we need to Calculate the Differences. In a new column, subtract the specific value from each value in the "Sales" column. This calculates the absolute differences between the specific value and each value in the list. In our example, we are taking 442 Rupees as specific value. Then we will use the MIN function to find the minimum absolute difference from the list using the formula: =MIN(L2:C9550).At last we will use INDEX and MATCH functions to find the closest match based on the minimum difference. In a separate cell, enter the formula: =INDEX(H2:H9995,MATCH(MIN(L2:L9995),L2:L9995,0)) and our closest value is 500.

ID	Customer Name	Category	Sub Category	City	Order Date	Region	Sales	Discount	Profit	State	Value difference			Specific Value	₹ 442
	Harish	Oil & Masala	Masalas	Vellore	08-11-2017	North	₹791	12%	₹ 401.28	Tamil Nadu	349				
	Sudha	Beverages	Health Drinks	Krishnagiri	08-11-2017	South	₹749	18%	₹ 401.28	Tamil Nadu	307			Min Difference	58
	Hussain	Food Grains	Atta & Flour	Perambalur	12-06-2017	West	₹2,360	21%	₹ 401.28	Tamil Nadu	1918				
	Jackson	Fruits & Veggies	Fresh Vegetables	Dharmapuri	11-10-2016	South	₹896	25%	₹ 401.28	Tamil Nadu	454			Closest Match	500
	Ridhesh	Food Grains	Organic Staples	Ooty	11-10-2016	South	₹2,355	26%	₹ 401.28	Tamil Nadu	1913				
	Adavan	Food Grains	Organic Staples	Dharmapuri	09-06-2015	West	₹2,305	26%	₹ 322.70	Tamil Nadu	1863				
	Jonas	Fruits & Veggies	Fresh Vegetables	Trichy	09-06-2015	West	₹826	33%	₹ 346.92	Tamil Nadu	384				
	Hafiz	Fruits & Veggies	Fresh Fruits	Ramanadhapuram	09-06-2015	West	₹1,847	32%	₹ 147.76	Tamil Nadu	1405				
	Hafiz	Bakery	Biscuits	Tirunelveli	09-06-2015	West	₹791	23%	₹ 181.93	Tamil Nadu	349				
	Krithika	Bakery	Cakes	Chennai	09-06-2015	West	₹1,795	27%	₹ 484.65	Tamil Nadu	1353				
	Ganesh	Snacks	Chocolates	Karur	09-06-2015	West	₹1,903	13%	₹ 437.69	Tamil Nadu	1461				
	Yadav	Eggs, Meat & Fish	Eggs	Namakkal	09-06-2015	West	₹701	10%	₹ 308.44	Tamil Nadu	259				
	Sharon	Snacks	Cookies	Dindigul	15-04-2018	South	₹1,659	19%	₹ 315.21	Tamil Nadu	1217				
	Peer	Fruits & Veggies	Fresh Vegetables	Kanyakumari	05-12-2017	West	₹1,277	25%	₹ 63.85	Tamil Nadu	835				
	Sundar	Eggs, Meat & Fish	Chicken	Kanyakumari	22-11-2016	Central	₹831	22%	₹ 207.75	Tamil Nadu	389				
	Ramesh	Oil & Masala	Edible Oil & Ghee	Krishnagiri	22-11-2016	Central	₹1,440	11%	₹ 100.80	Tamil Nadu	998				
	Alan	Bakery	Cakes	Dharmapuri	11-11-2015	Central	₹1,678	28%	₹ 318.82	Tamil Nadu	1236				
	Arutra	Beverages	Health Drinks	Bodi	13-05-2015	West	₹1,617	19%	₹ 113.19	Tamil Nadu	1175				
	Haseena	Eggs, Meat & Fish	Mutton	Tenkasi	27-08-2015	West	₹1,757	35%	₹ 386.54	Tamil Nadu	1315				
	Verma	Beverages	Soft Drinks	Kanyakumari	27-08-2015	West	₹692	29%	₹ 159.16	Tamil Nadu	250				
	Hafiz	Beverages	Health Drinks	Vellore	27-08-2015	West	₹522	19%	₹ 208.80	Tamil Nadu	80				
	Alan	Food Grains	Dals & Pulses	Karur	09-12-2017	Central	₹948	13%	₹ 47.40	Tamil Nadu	506				
	Haseena	Beverages	Soft Drinks	Krishnagiri	09-12-2017	Central	₹707	34%	₹ 148.47	Tamil Nadu	265				
	Alan	Fruits & Veggies	Organic Vegetables	Tenkasi	16-07-2018	East	₹969	29%	₹ 77.52	Tamil Nadu	527				
	Sharon	Eggs, Meat & Fish	Eggs	Ooty	25-09-2016	West	₹1,100	11%	₹ 495.00	Tamil Nadu	658				
	Krithika	Snacks	Chocolates	Tirunelveli	16-01-2017	West	₹2,022	11%	₹ 202.20	Tamil Nadu	1580				
	Muneer	Snacks	Cookies	Trichy	16-01-2017	West	₹541	12%	₹ 43.28	Tamil Nadu	99				

Supermart Grocery Sales - Retai   Sup Gro

CONCLUSION:

In conclusion, our team successfully completed the project work on finding the 1st match to a value in Excel. Through collaborative efforts and dedication, each team member contributed their skills and expertise to achieve our objectives. By working together, we were able to analyze the requirements, develop a systematic approach, and implement the solution efficiently. Each team member played a crucial role in the project's completion, contributing unique perspectives and insights to overcome challenges .

This project has not only enhanced our technical skills but also strengthened our teamwork and communication abilities, which are essential for future endeavors. Working on this project has been an enriching journey for all of us, allowing us to explore new concepts, hone our skills, and collaborate effectively as a team. We would like to extend our heartfelt gratitude to the L&T Edu Tech Team for entrusting us with the exciting opportunity to work on this project. We are immensely grateful for the confidence you have placed in our team and for providing us with this valuable learning experience.