****

**Introduction:**

NewZEN is a computer hardware manufacturer company which has clients across India. The company has a head office in Delhi and also regional offices across India.

The Sales Director of the company is facing couple of challenges like:

* Understanding the dynamics of the market.
* Tracking the Sales and Revenue of all the regions/markets.
* Profit Contribution by the customers.

Implementing the different phases of data analysis process:

**1. Ask:** Ask the questions to identify, understand and define the problem.

*Key tasks –*

**Problem statement:** To understand the sales insights from the visual dashboard that can be apparent for sales team for decision making.

**Stakeholders:** Sales Director, Marketing Team, Customer Service Team, Data Analytics Team, IT.

**2. Prepare:** Collecting the data, understanding the metrics of data and storing data.

Importing the dataset into MySQL Workbench.

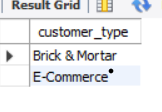
**Inspecting the data using SQL:**

The dump has 4 years history data. Tables in the sales database: **customers**, **date**, **markets**, **products**, **transactions**. Let’s look into the details of each table.

**I) customers**:

Contains customer details like code, name, type. With 38 customers and two distinct customer types as

select distinct customer\_type FROM sales.customers



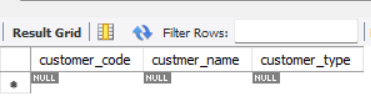
No NULL records found in the table.

select \*

from sales.customers

where custmer\_name is null

OR customer\_type is null



Found inconsistent column name *custmer\_name.*

**II) date**:

Contains sales dates details between 2017 to 2020 with

select min(date), max(date) FROM sales.date;



**III) markets:**

Has details of market like name, code, and zone. There are 17 unique markets.

select COUNT(distinct markets\_code), COUNT(distinct zone) FROM sales.markets



No zones for *market\_name* values ‘New York’ and ‘Paris’. These values indicate the out of scope of our analysis. i.e, not with in India, so no zones are allocated. Need to consider only 15 market codes.

select \*

from sales.markets

where markets\_name = ''

or zone = ''

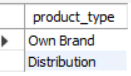


Markets with codes 'Mark007', 'Mark013' has same market name as ‘Bhopal’, we have to check whether ‘Bhopal’ has two different market codes or else entered the incorrect data.



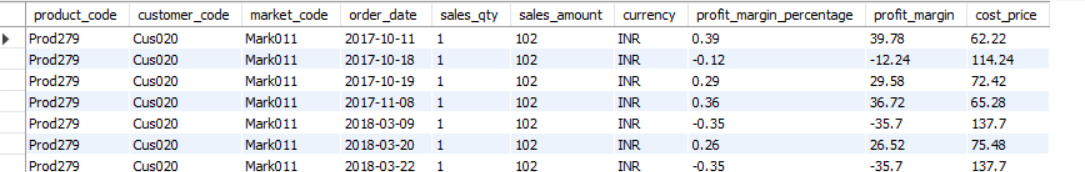
**IV) products:** has 279 products code and 2 product types information.

SELECT distinct product\_type FROM sales.products



**V) transactions:**

Themain transaction details are maintained in this table. It has 148395 transaction details.



Found 338 unique values in *products\_code* column in **transactions** table, where in **products** table*products\_code* column has 279 values.

SELECT COUNT(distinct product\_code) FROM sales.transactions ### 338 records

Need to check the reason for excess product codes in **transaction** table and handle it.

*market\_code* column name is having inconsistent naming compared to *markets\_code* in **markets** table.

When compared the **dates** tablewith the dates in **transactions** table,

**dates** table: details start from June 2017

**transactions** table: *order\_date* details recorded from October 2017

Are the transaction records before October 2017 are not updated, or it has nothing to do with the **dates** table need to check with the technical team.

select min(order\_date), max(order\_date) FROM sales.transactions;

# min(order\_date), max(order\_date)

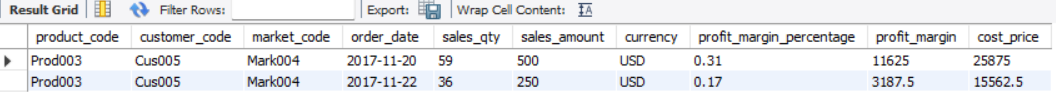
#'2017-10-04', '2020-06-26'

Two records are found with *currency* as ‘USD’

SELECT distinct currency FROM sales.transactions;

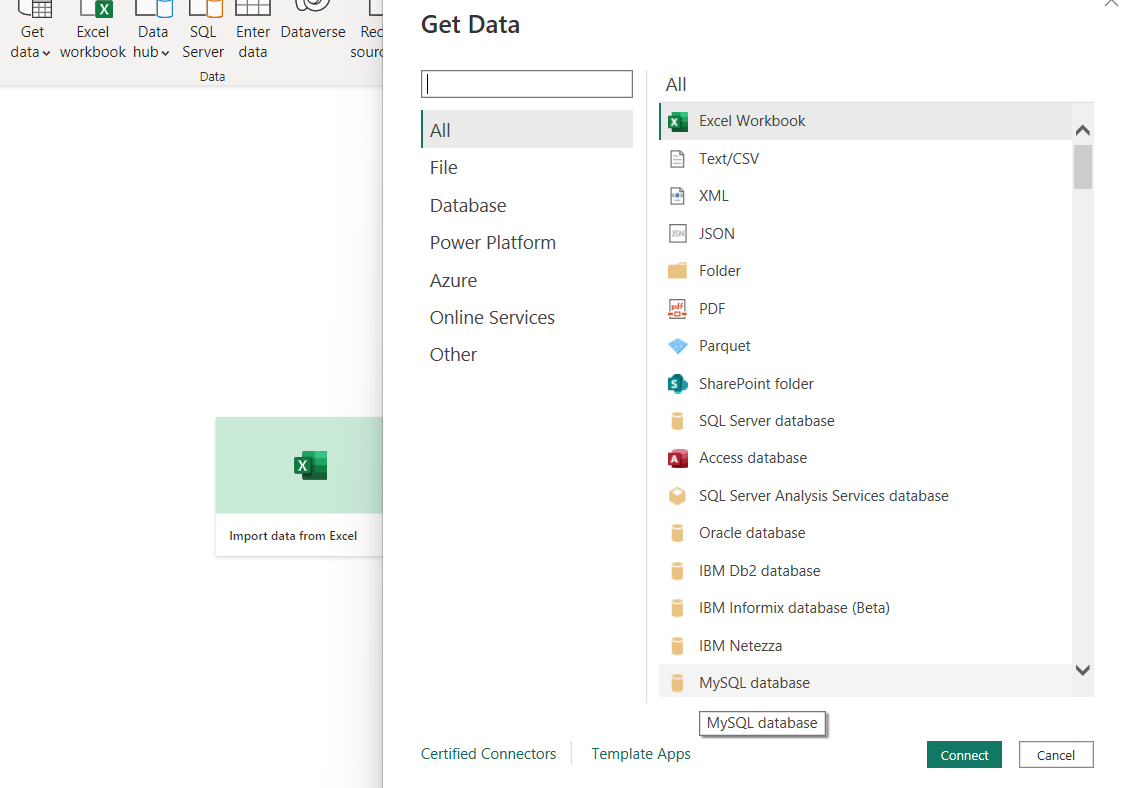


SELECT \* FROM sales.transactions where currency = 'USD';

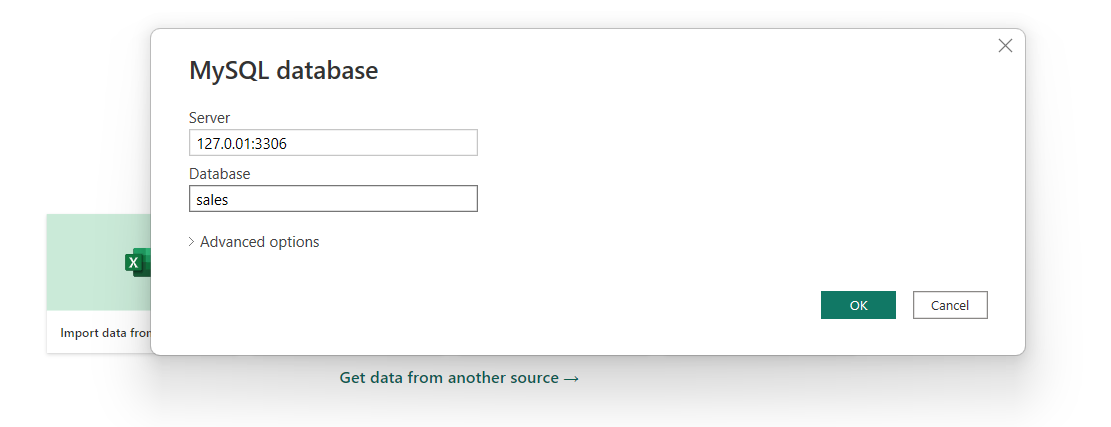
****

**Connecting to Power BI:**

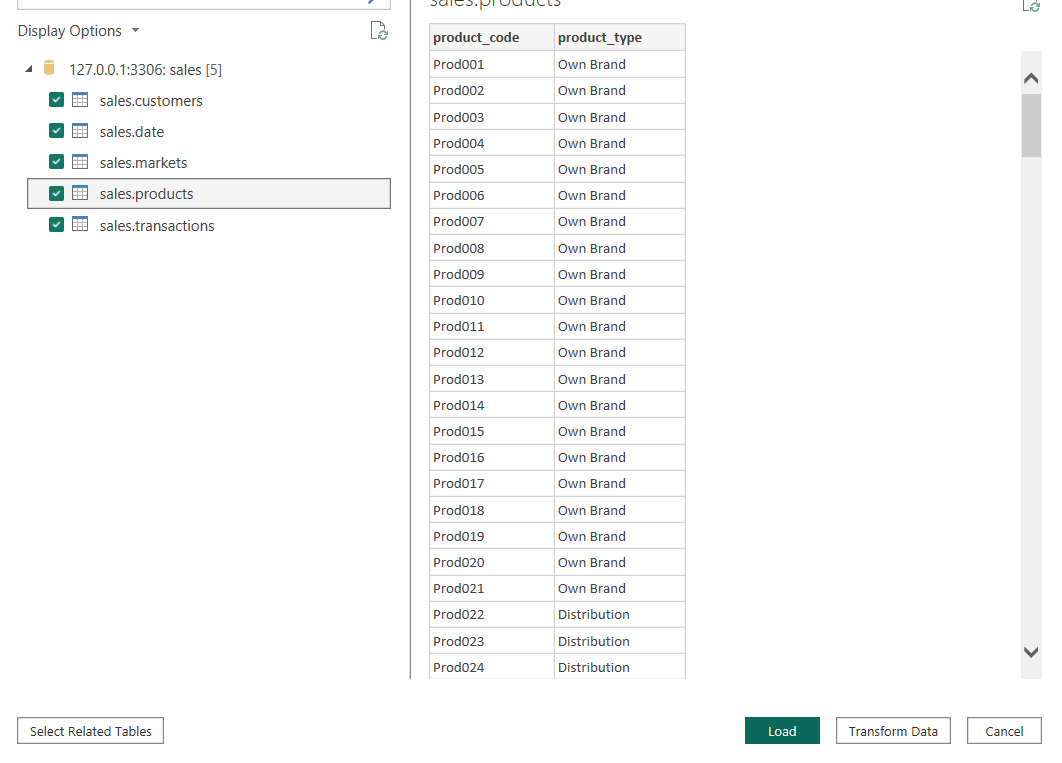
* We are using the Power BI (Desktop) tool for data transformation and analysis.
* Creating the database connection from Power BI.



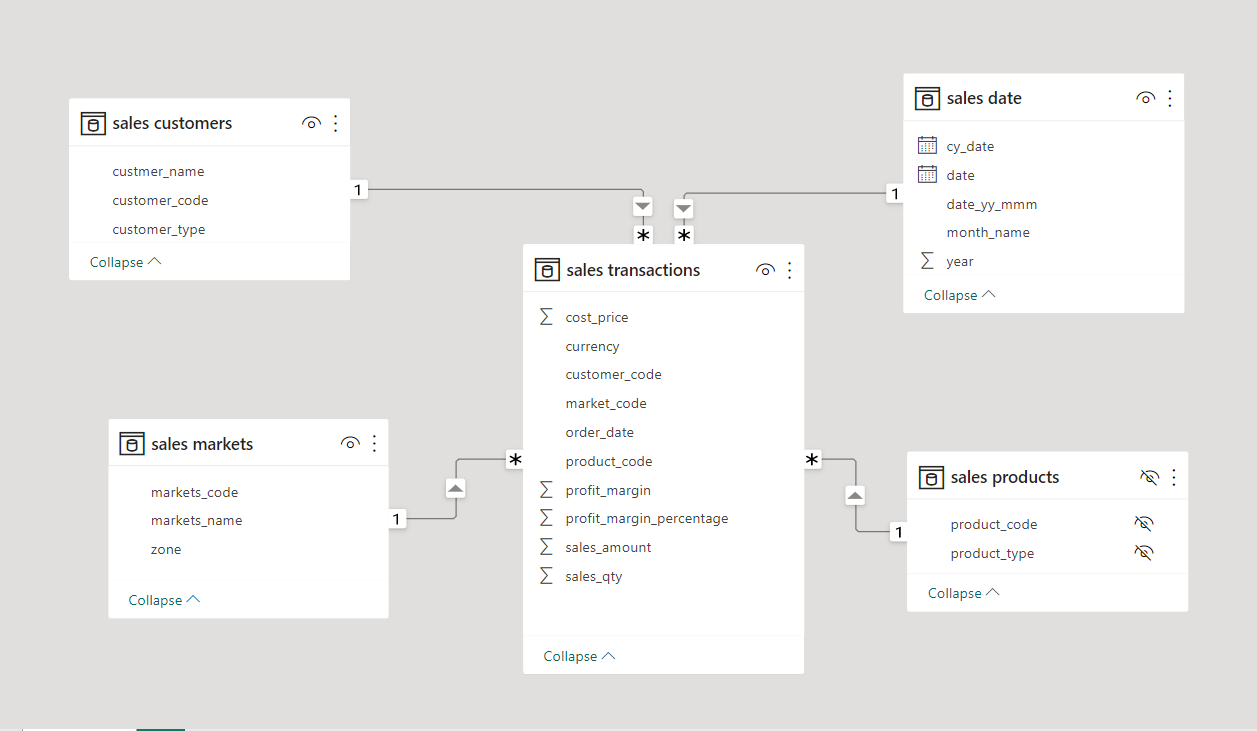
* Enter the server and database details:



* Select all the tables from the Navigator and Load:



* Open the Model tab, it’s the data model which shows the relationships between the tables.



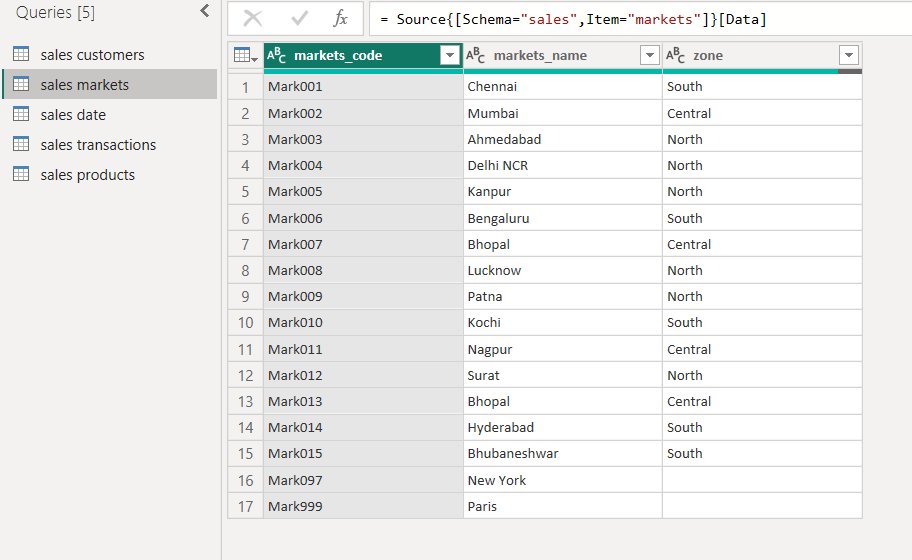
* By default, Power BI builds relationship between the tables based on the column names and datatypes, otherwise we can do it manually by drag and drop.
* Power Query is used to perform ETL operations on data.

**3. Process:** We have to clean the data before we transform and analyze it.

**Power Query for data transformation:**

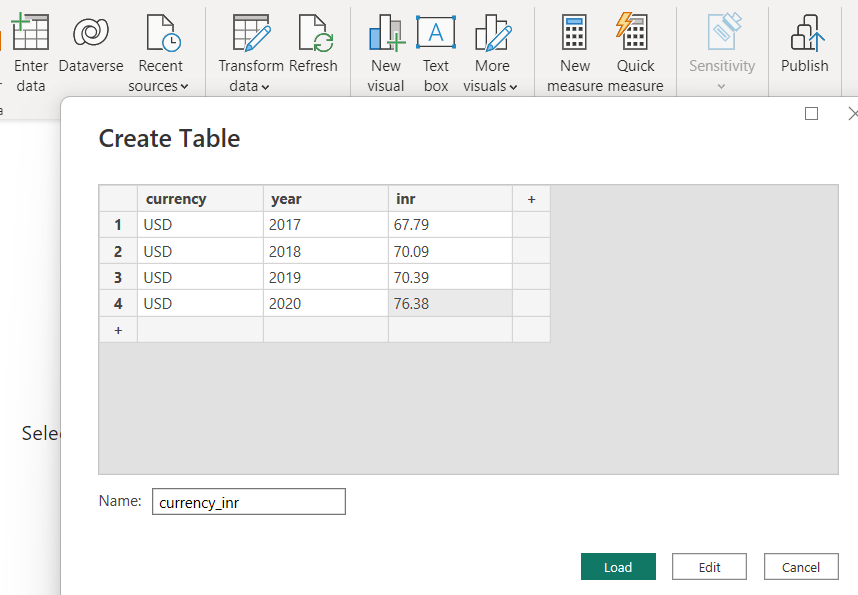
1. ***Remove irrelevant data:***

**markets** table: Deleting the records *market\_codes* ‘Mark097’ and ‘Mark999’ with corresponding markets\_names ‘New York’ and ‘Paris’.

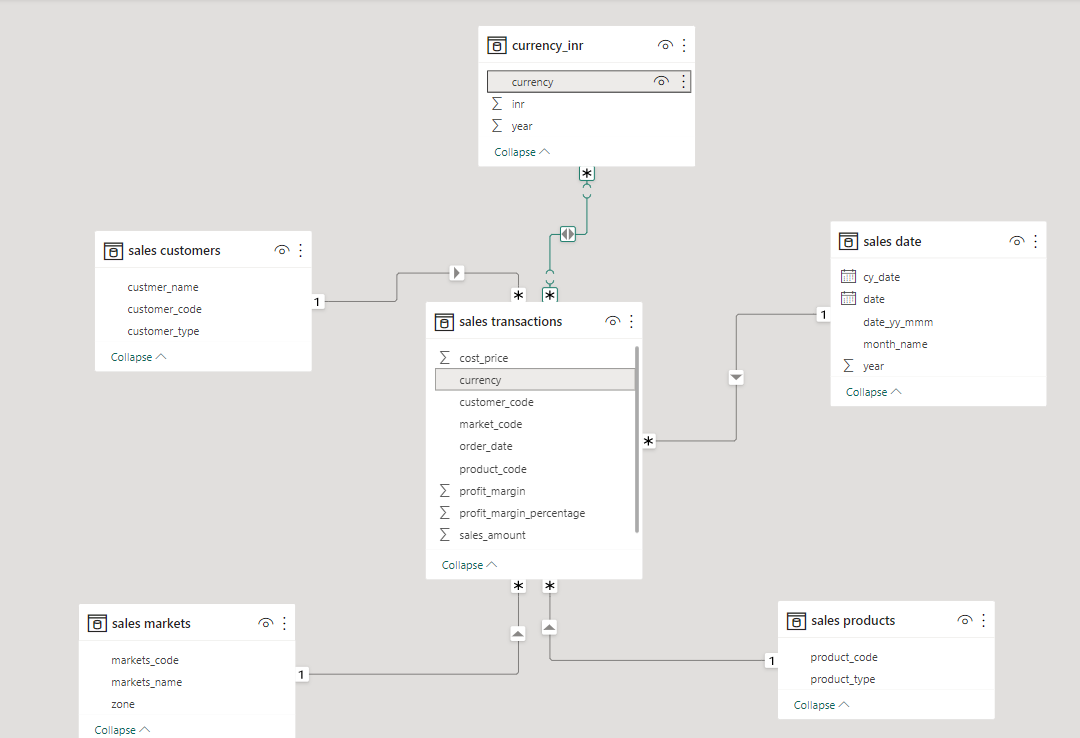


1. ***Converting the currency USD to INR:***

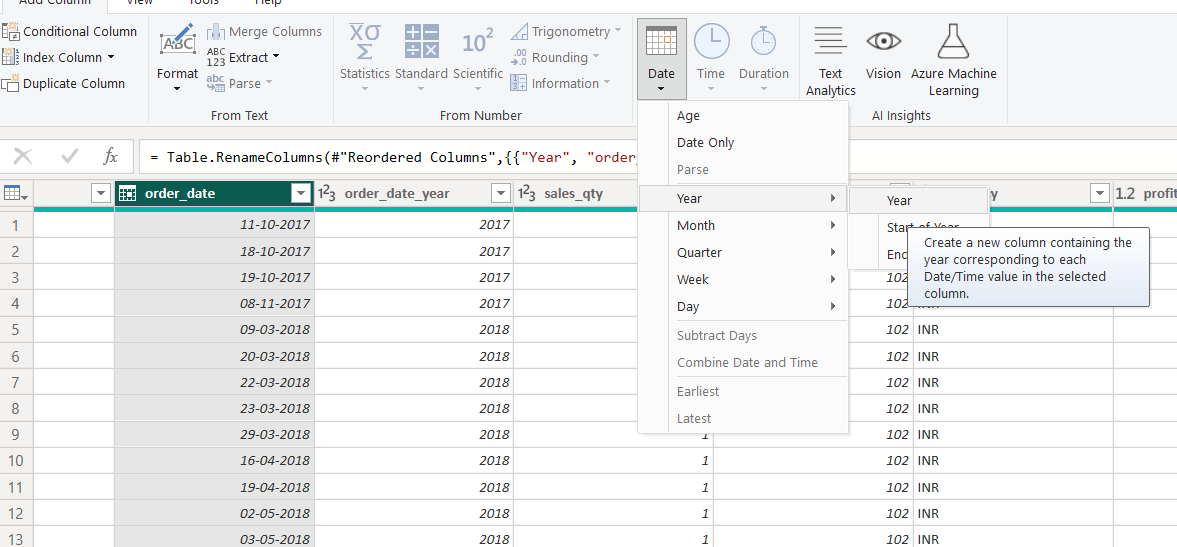
* **transactions** table records with *currency* as‘USD’ has to be converted into ‘INR’ by multiplying the *sales\_amount* with exchange rate of that corresponding transaction year.
* Creating a new table **currency\_inr** with coulmns as *currency, year, inr*



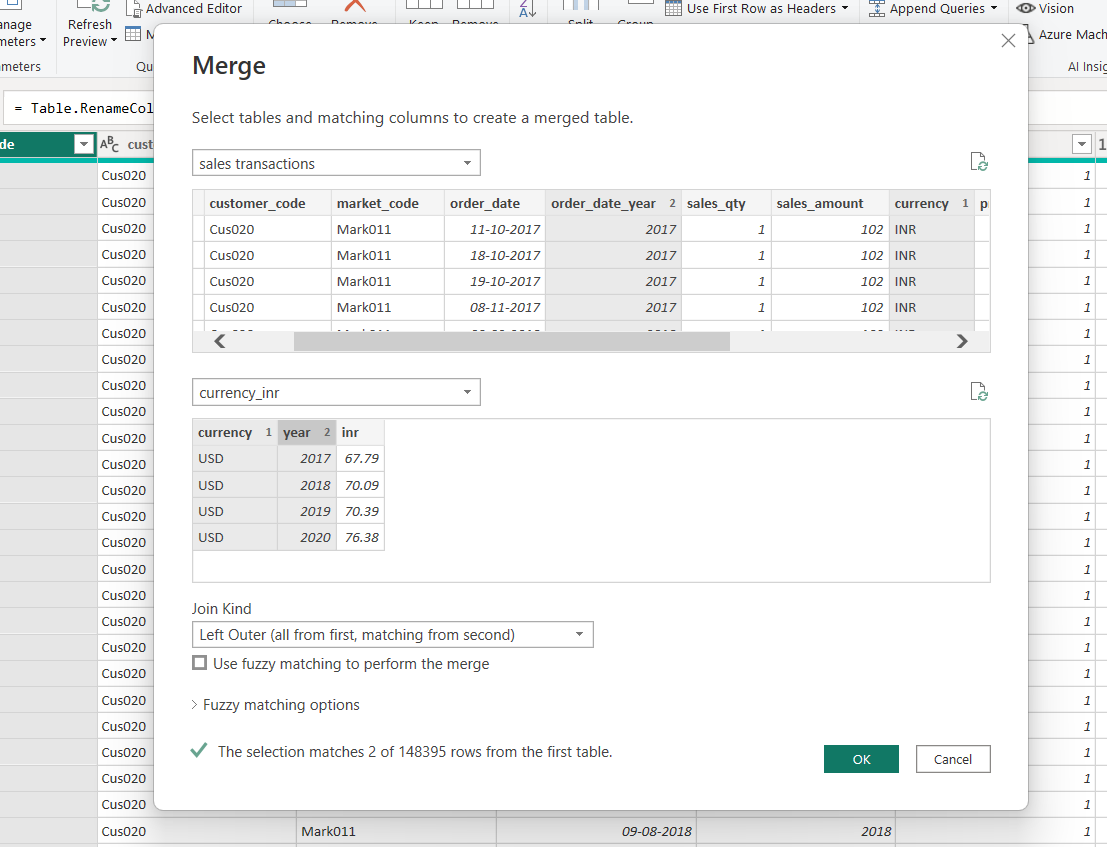
* Updating the data model with new table:



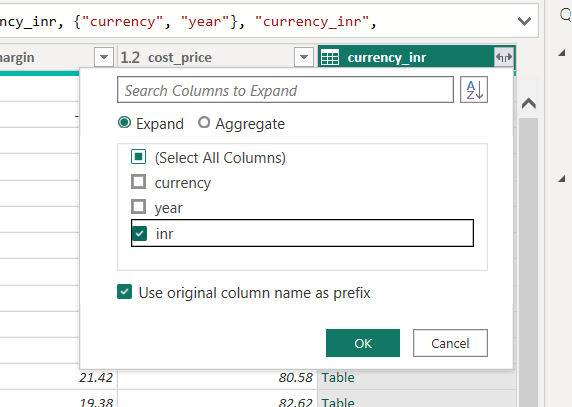
* A new column *order\_date\_year* in **transaction** table is added by extracting the year from *order\_date* column.



* Merging the new table **currency\_inr** with **transactions** tableas **transactions\_inr** by matching the records based on currency and year criteria. 2 records found with that scenario.



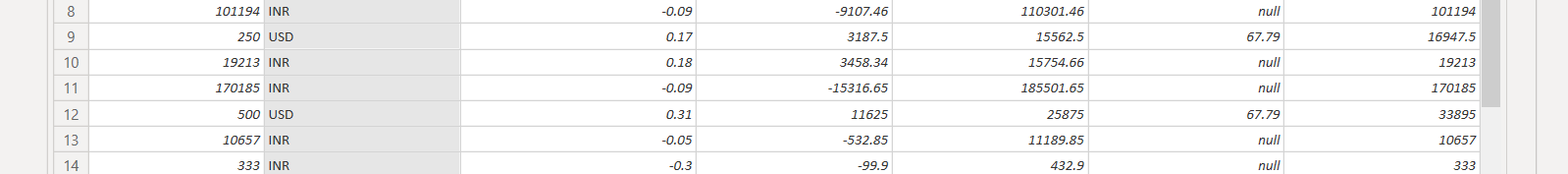
* Expanded the *currency\_inr*



* Adding a conditional column *sales\_amount\_inr* in the merge table **transactions\_inr** using the columns *sales\_amount* and *currency\_inr.*

= Table.AddColumn(#"Expanded currency\_inr", "sales\_amount\_inr", each if [currency] = "USD" then [currency\_inr.inr]\*[sales\_amount] else if [currency] = "INR" then [sales\_amount] else null)

* The new column *sales\_amount\_inr* nowhas all the currency amounts in INR.



1. ***Fix structural errors:*** Modified the two naming convention errors in the tables:

**customers:***custmer\_name to customer\_name*

**markets:** *markets\_code* to *market\_code*

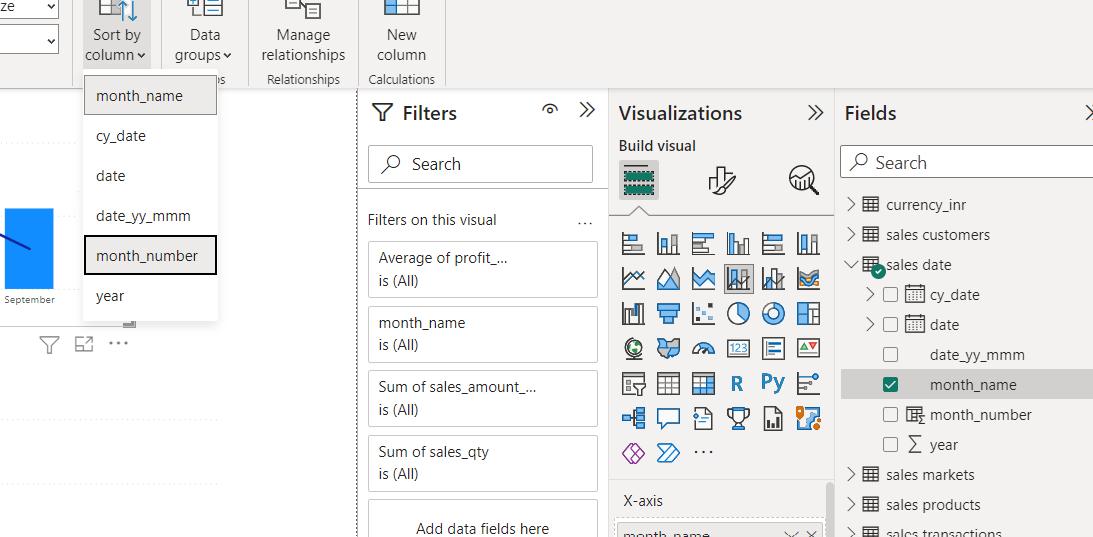
1. ***Adding a column for month sorting during Visualization:***

Creating the month\_number in **date** table using thebelowformula:

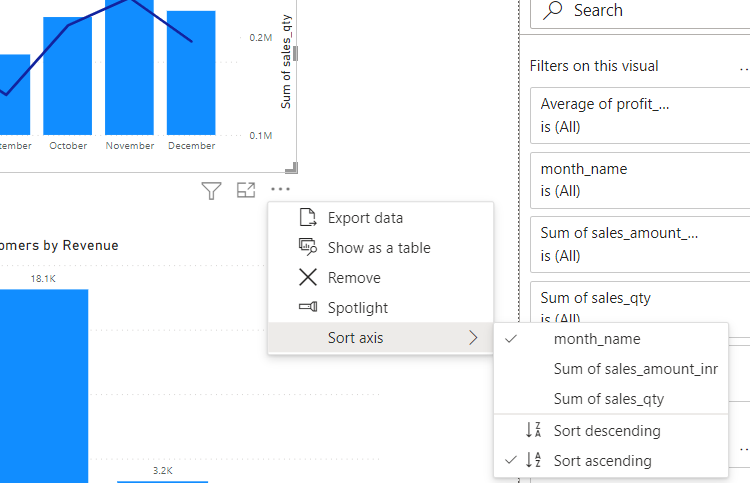
month\_number = MONTH('sales date'[date])

In the chart Revenue and Sales quantity by month, the *month\_name*(x-axis) from the **date** table is not sorted in the chronological order. Because the *month\_name* is in text format.

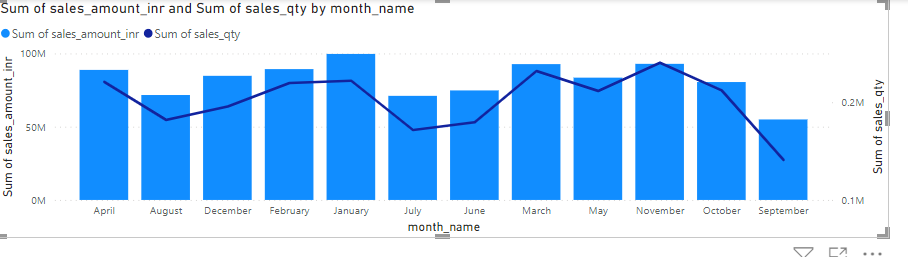
* To fix this issue, we can make use of new column *month\_number*.
* Click on *month\_name* from fields and select **Column tools** tab from the header.
* Select **Sort by column**>*month\_number*.



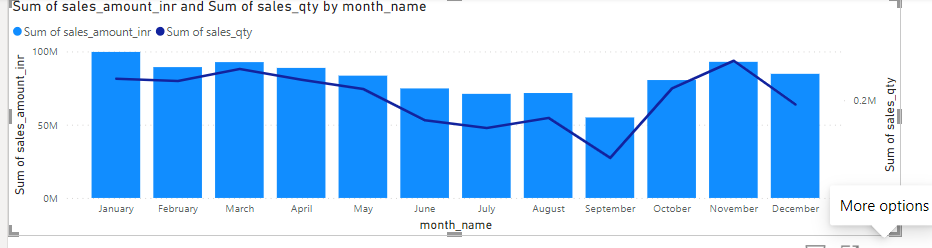
* Now sort the month axis by *month\_name*



*Before sorting:*



*After sorting:*



**4. Analyze:**

**Analyzing the dataset using visualization:**

*Created a new table* ***sales\_metrics*** *for measures:*

Revenue = SUM('sales transactions\_inr'[sales\_amount\_inr])

Total Sales = SUM('sales transactions\_inr'[sales\_qty])

Total Profit Margin = SUM('sales transactions\_inr'[profit\_margin])

Profit Margin % = DIVIDE([Total Profit Margin],[Revenue])

Revenue LY = CALCULATE([Revenue],SAMEPERIODLASTYEAR('sales date'[date] ))

Target Diff = [Profit Margin %]-'Profit Target'[Profit Target Value]

*Measures for titles:*

TopNRevenueCustomers Title = "Revenue by Top "&SELECTEDVALUE(ViewTopN[ViewTopN])&" Customers"

TopNSalesQntyCustomers Title = "Sales Quantity by Top "&SELECTEDVALUE(ViewTopN[ViewTopN])&" Customers"

*Measures for Customer Ranks:*

CustomerRankByRevenue = IF(HASONEVALUE('sales customers'[customer\_name]),RANKX(ALL('sales customers'[customer\_name]),[Revenue],,DESC))

CustomerRankBySalesQty = IF(HASONEVALUE('sales customers'[customer\_name]),RANKX(ALL('sales customers'[customer\_name]),[Total Sales],,DESC))

*Measures for TopN Slicer:*

Created a table ViewTop with numeric values till 10 and adding this table column to a slicer.

SelectedTopNNumber = IF(HASONEVALUE('ViewTopN'[ViewTopN]),MIN('ViewTopN'[ViewTopN]),10)

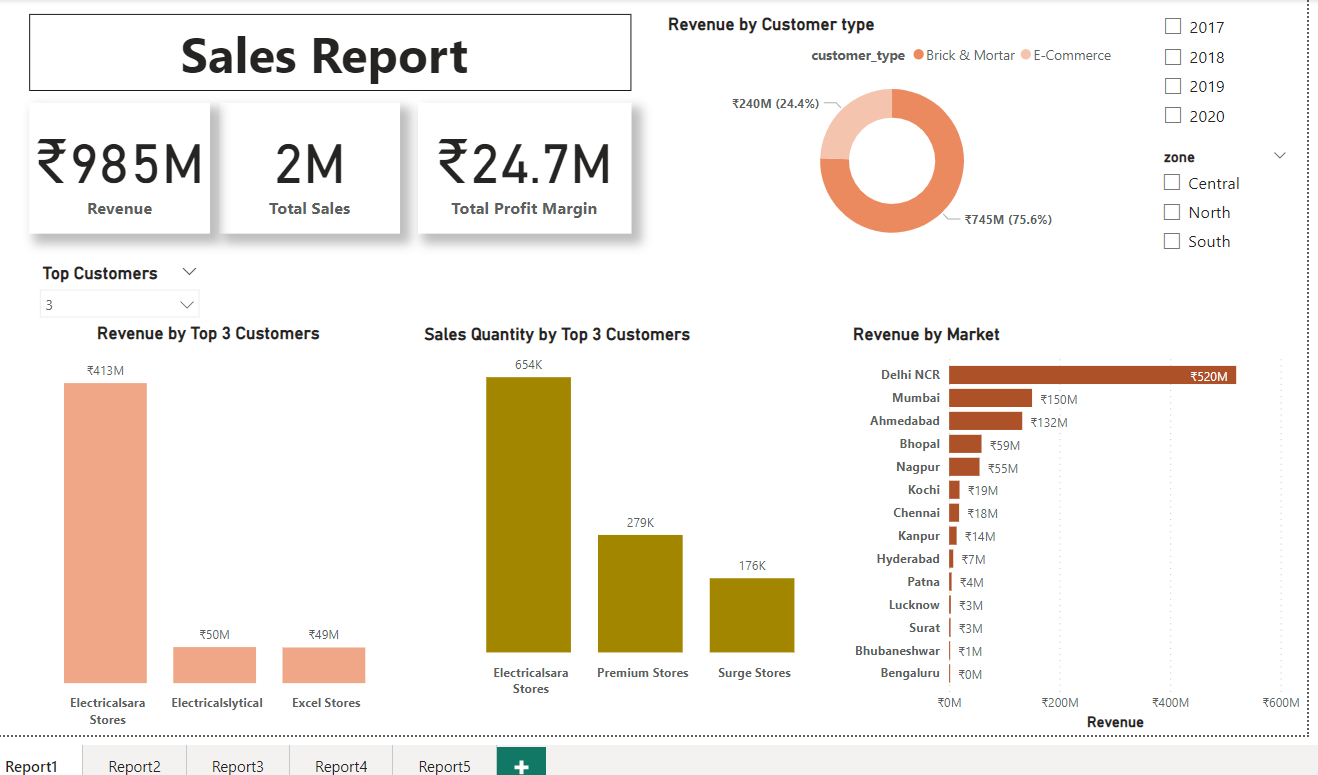
CustomerToBeIncludedInRevn = IF([CustomerRankByRevenue] <= [SelectedTopNNumber],1,0)

CustomerToBeIncludedInSalesQty = IF([CustomerRankBySalesQty] <= [SelectedTopNNumber],1,0)

**Analysis of features based on the metrics:** KPIs, Customers, Order Date, Markets, Products, Profit Margin

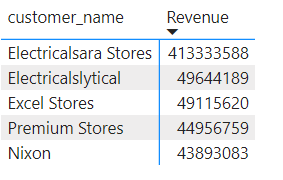
**1. KPIs (Key Performance Indicator):**

* *Total* ***Revenue*** *across India:* **985M INR**
* *Total* ***Sales*** *from all Customers:* **2M**
* *Total Profit Margin:* ***24.7M***
* *Average Profit Margin:* **166.16**

****

**2. Revenue by Customers, Products, Markets:**

* *Top 5 Customers by Revenue:*

﻿

﻿﻿Customer Electricalsara Stores leads the chart with Revenue 413.33M INR and

accounted for 41.97% of Total Revenue.﻿﻿ ﻿﻿

*SQL Query:*

SELECT c.custmer\_name

, (sum(t.sales\_amount)/(select sum(sales\_amount) from sales.transactions))\*100 as percentage

FROM sales.transactions t

INNER JOIN sales.customers c

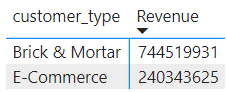
ON t.customer\_code = c.customer\_code

group by c.customer\_code

having custmer\_name = 'Electricalsara Stores' ##'Electricalsara Stores', '413333588'

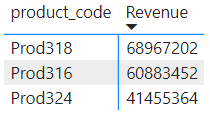


* *Revenue by Customer Type:*

  
﻿Revenue for Brick & Mortar (744.52M INR) was higher than E-Commerce (240.34M INR).﻿﻿ ﻿﻿

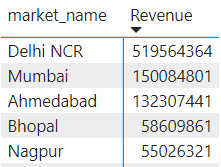
﻿﻿Brick & Mortar accounted for 75.60% of Revenue.﻿﻿ ﻿﻿ ﻿

* *Top 3 performing products:*

  
Product Prod318 has contributed the highest Revenue with 68.97M INR.

Surprisingly these 3 products are not updated in the **products** table, need to check with technical team.

* *Revenue by Market:*



﻿Delhi NCR tops the list with the highest Revenue of 519.56M INR and accounted for 52.75% of Revenue.﻿﻿ ﻿﻿

*SQL query:*

select m.markets\_name mn

, (sum(t.sales\_amount)/(select sum(sales\_amount) from sales.transactions))\*100 as percentage

from transactions t

inner join markets m

on t.market\_code = m.markets\_code

group by mn

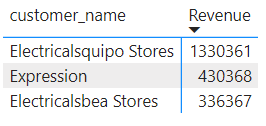
having mn = 'Delhi NCR'



Bengaluru has generated least Revenue 373.12K INR.

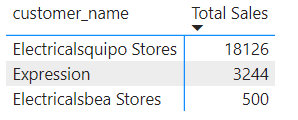
Where in, North zone has generated highest Revenue of 676M INR compared to Central and South zones.

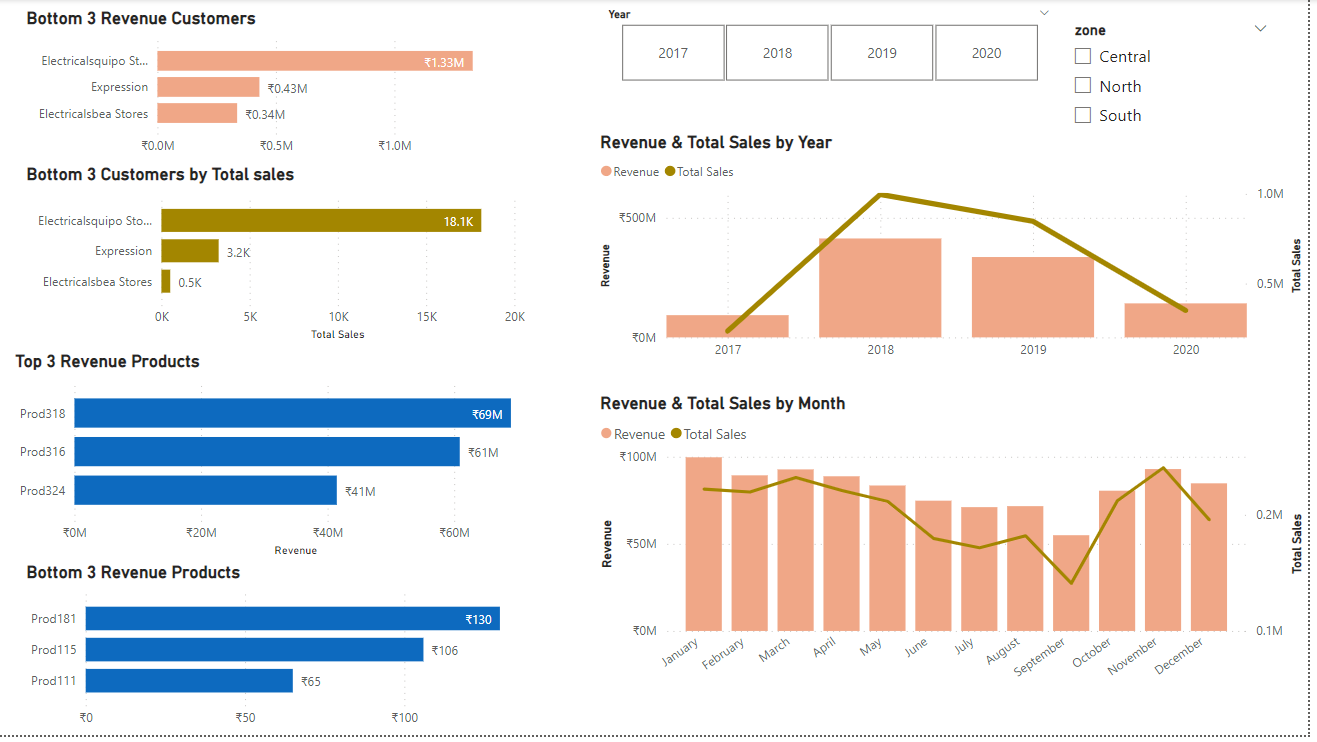
* *Bottom 3 Customers by Revenue:*



Customer Electricalbea Stores has contributed low Revenue 0.34Ms.

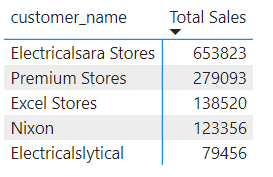
* *Bottom 3 Customers by Total sales:*

  
﻿Electricalsbea Stores has least total sales of 500.

****

**3. Sales by Customers, Products, Markets:**

* *Top 5 Customers by Sales Quantity:*



﻿﻿﻿Customer Electricalsara Stores accounted for 26.91% of Total Sales with sales value of 653823.﻿﻿ ﻿﻿

*SQL query:*

SELECT c.custmer\_name

, (sum(t.sales\_qty)/(select sum(sales\_qty) from sales.transactions))\*100 as percentage

FROM sales.transactions t

INNER JOIN sales.customers c

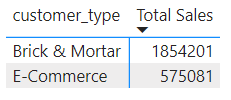
ON t.customer\_code = c.customer\_code

group by c.customer\_code

having custmer\_name = 'Electricalsara Stores'



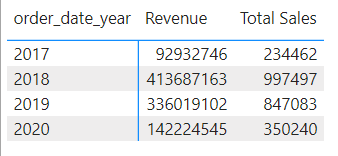
* *Total sales by Customer Type:*

  
﻿Total Sales for Brick & Mortar (1854201) was higher than E-Commerce (575081).﻿﻿ ﻿﻿ ﻿﻿

Brick & Mortar accounted for 76.33% of Total Sales.﻿﻿ ﻿﻿ ﻿

**4. Revenue and Sales over the period:**

* *Revenue & Total Sales by Year:*



﻿Overall year 2018 has outperformed with the highest Revenue of 413.69M INR and was 345.15% higher than 2017, which had the lowest Revenue at 92.93M INR.﻿﻿

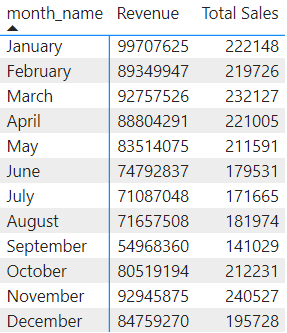
﻿﻿% Increase in Revenue = (413687163-92932746)/(92932746)\*100 = 345.1468%

﻿﻿Revenue and total Total Sales are positively correlated with each other over the years.﻿﻿ ﻿﻿ ﻿﻿

But the Revenue has come down in year 2020 at 142.22M INR which is 57.67% lower than 2019.

% Decrease ﻿﻿in Revenue = (336019102-142224545)/(336019102)\*100 = 57.6737%

* *Revenue & Total Sales by Month:*



﻿Month January had the highest Revenue of 99.71M INR and September, which had the lowest Revenue at 54.97M INR.﻿﻿ ﻿﻿ ﻿﻿

Revenue and Total Sales are positively correlated with each other.﻿﻿ ﻿﻿

﻿﻿January accounted for 10.12% of Revenue.﻿﻿ ﻿﻿

And November contributed 9.9% of Total sales

*SQL Query:*

SELECT MONTH(t.order\_date) month\_no

, (sum(t.sales\_amount)/(select sum(sales\_amount) from sales.transactions))\*100 as percentage

FROM sales.transactions t

group by month\_no

having month\_no = 1



**5. Average Profit margin by Top Customers, Products, Markets:**

* How profitable is the business?

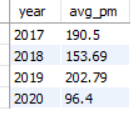
***Profit margin analysis:*** This analysis is as per the Profit Margin data provided in the **transaction** table in column *profit\_margin*.

* *Average profit margin over the period (2017-2020):*

select year(order\_date) year, round(avg(profit\_margin), 2) avg\_pm

from transactions

group by year(order\_date)



Year 2019 is quite profitable over the years with the Average Profit Margin 202 where in during 2020 it has come down to 96.

* *Average profit margin by Customers:*

select c.customer\_code, c.custmer\_name, avg(t.profit\_margin) avg\_pm

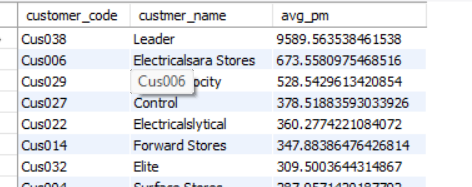
from transactions t

inner join customers c

on t.customer\_code = c.customer\_code

group by customer\_code

order by avg\_pm desc



Business with Customer ‘Leader’ is highly profitable with 9589.5. 'Electricalslance Stores' has very poor average profit margin -12.07.

* *Average profit margin by Market:*

select m.markets\_name, avg(profit\_margin) avg\_pm

from transactions t

inner join markets m

on t.market\_code = m.markets\_code

group by m.markets\_name

order by avg\_pm desc



Patna tops the chart with high average profit margin over all the markets with 453.38, where in Bengaluru (-5538.2) and Kanpur(-23.4) are with negative ratios.

* *Transaction bins based on profit margin:*

A good margin will vary considerably by industry, but as a general rule of thumb,

10% net profit margin is considered average,

20% margin is considered high (or “good”),

5% margin is low.

Again, these guidelines vary widely a variety of other factors.

Based on the above, let’s create bins with count of transactions:

select

SUM(case when profit\_margin <= 5 then 1 else 0 end) as 'Low\_Margin'

, SUM(case when profit\_margin > 5 and profit\_margin <= 10 then 1 else 0 end) as 'Average\_Margin'

, SUM(case when profit\_margin > 10 and profit\_margin < 20 then 1 else 0 end) as 'Good\_Margin'

, SUM(case when profit\_margin >= 20 then 1 else 0 end) as 'High\_Margin'

from transactions



So, majority of transactions ended up with low profit margin.

Let’s find what percentage of these low margins transactions contribute to the whole period transactions:

with temp as(

select

SUM(case when profit\_margin <= 5 then 1 else 0 end) as 'Low\_Margin'

, SUM(case when profit\_margin > 5 and profit\_margin <= 10 then 1 else 0 end) as 'Average\_Margin'

, SUM(case when profit\_margin > 10 and profit\_margin < 20 then 1 else 0 end) as 'Good\_Margin'

, SUM(case when profit\_margin >= 20 then 1 else 0 end) as 'High\_Margin'

from transactions),

temp\_list as(

select 'Low\_Margin' category, Low\_Margin margin\_count from temp

union all

select 'Average\_Margin' category, Average\_Margin margin\_count from temp

union all

select 'Good\_Margin' category, Good\_Margin margin\_count from temp

union all

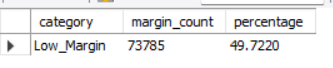
select 'High\_Margin' category, High\_Margin margin\_count from temp)

select category, margin\_count

,(select max(margin\_count)/sum(margin\_count)\*100 from temp\_list) as percentage

from temp\_list

where margin\_count = (select max(margin\_count) from temp\_list)



49.72% of the transactions had low profit margins.

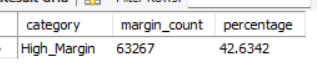
….

select category, margin\_count

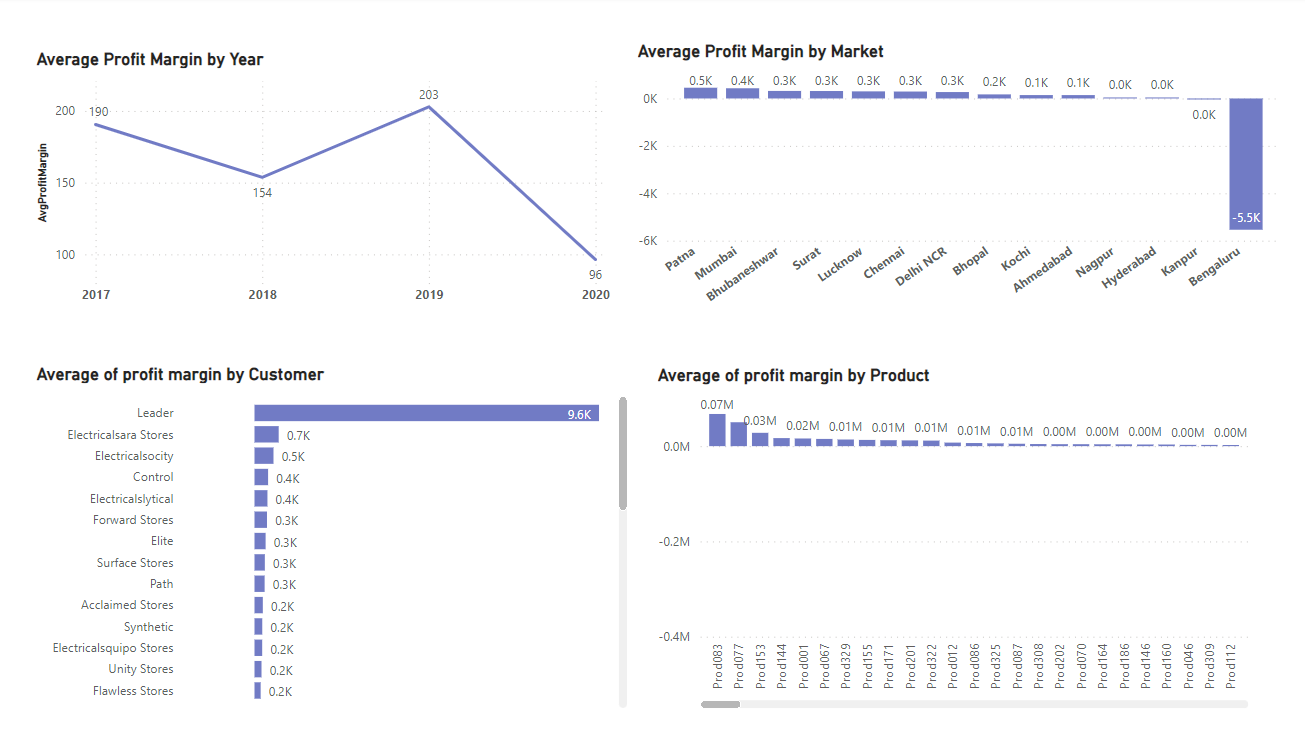
,margin\_count/(select sum(margin\_count) from temp\_list)\*100 as percentage

from temp\_list

where category = 'High\_Margin'



42.63% of transactions are High profit margin.

****

**5. Profit margin % by Customers, Products, Period, Markets:**

* *Profit margin % by Market:*

By adding the new measures,

Total Profit Margin = SUM('sales transactions\_inr'[profit\_margin])

Profit Margin % = DIVIDE([Total Profit Margin],[Revenue])

Visualisation on Profit margin% by market reveals:

Even though New Delhi generates high Revenue of 520M, profit margin % with 2.3% in that market is not remarkable. In contrast, Surat with low Revenue 3M, shows up with 4.9%.

* *Profit margin Contribution % by Market:*

Profit Margin Contribution % = DIVIDE([Total Profit Margin],CALCULATE([Total Profit Margin],ALL('sales products'),ALL('sales customers'),ALL('sales markets')))

The market New Delhi with 48.5% contributes to the total profit margin over all the markets. Wherein Bengaluru and Kanpur markets under perform with negative profit shares.

* *Revenue Contribution % by Market:*

Revenue Contribution % = DIVIDE([Revenue],CALCULATE([Revenue],ALL('sales products'), ALL('sales customers'),ALL('sales markets')))

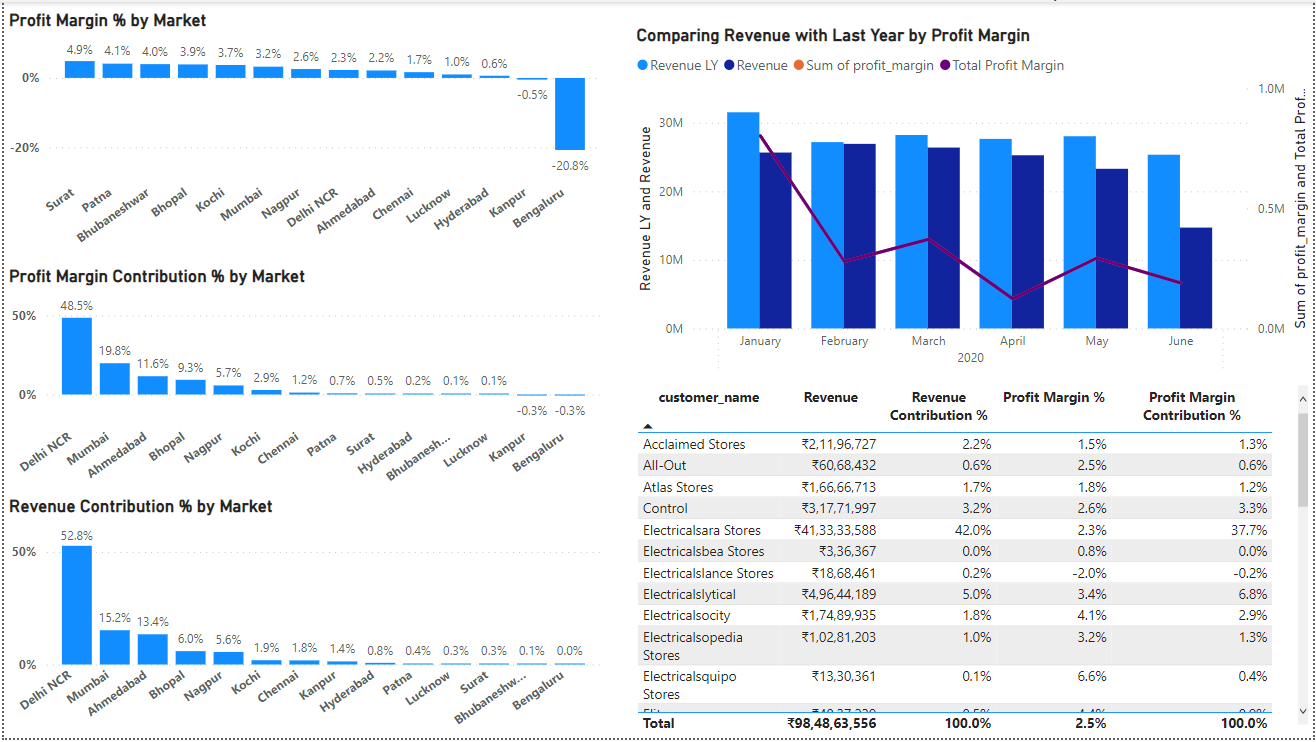
New Delhi is contributing high Revenue percentage of 52.8 accordingly with Revenue.

* *Comparing Revenue of Year 2020 with Last Year by Profit Margin:*

Revenue = SUM('sales transactions\_inr'[sales\_amount\_inr])

Revenue LY = CALCULATE([Revenue],SAMEPERIODLASTYEAR('sales date'[date] ))

From the Clustered chart, it is clearly evident that the Year 2020 Revenue has come down and so the profit margin.

****

* *Customer behaviour using metrics:*

From the table chart when we select the Customer name, accordingly we can identify which customer is giving more business over those periods.

Customer ‘Electricalsara Stores’ though the profit margin is low, its Revenue % and Profit Margin Contribution % is high.

* *Customer Order counts:*

From the Customer count table, we can infer the in the Year 2018 ‘Premium stores’ has given more business with high count of orders 10218 and its overall Revenue contribution is just 4.6%. Where in next year it has drop to 42%.

High Revenue generating Customer ‘Electricalsara Stores’ orders count is 5871 in 2018 and its count has gone low over the next years to less than 50%.

SELECT c.custmer\_name,

SUM(case when YEAR(order\_date) = 2017 then 1 else 0 end) as '2017'

,SUM(case when YEAR(order\_date) = 2018 then 1 else 0 end) as '2018'

,SUM(case when YEAR(order\_date) = 2019 then 1 else 0 end) as '2019'

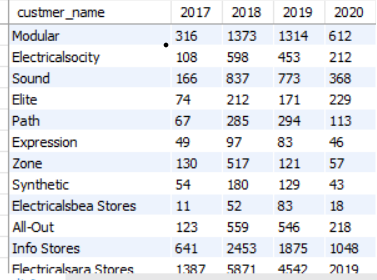
,SUM(case when YEAR(order\_date) = 2020 then 1 else 0 end) as '2020'

FROM sales.transactions t

INNER JOIN sales.customers c

ON t.customer\_code = c.customer\_code

group by c.custmer\_name



* *Customer Vs Profit Target:* Let’s try to understand the customers based on the target profit set.

Creating a New Parameter ‘Profit Target’ and adding a slicer along.

To highlight the market/customer as a Red Alert if the they are not reaching the target profit value set in the slicer.

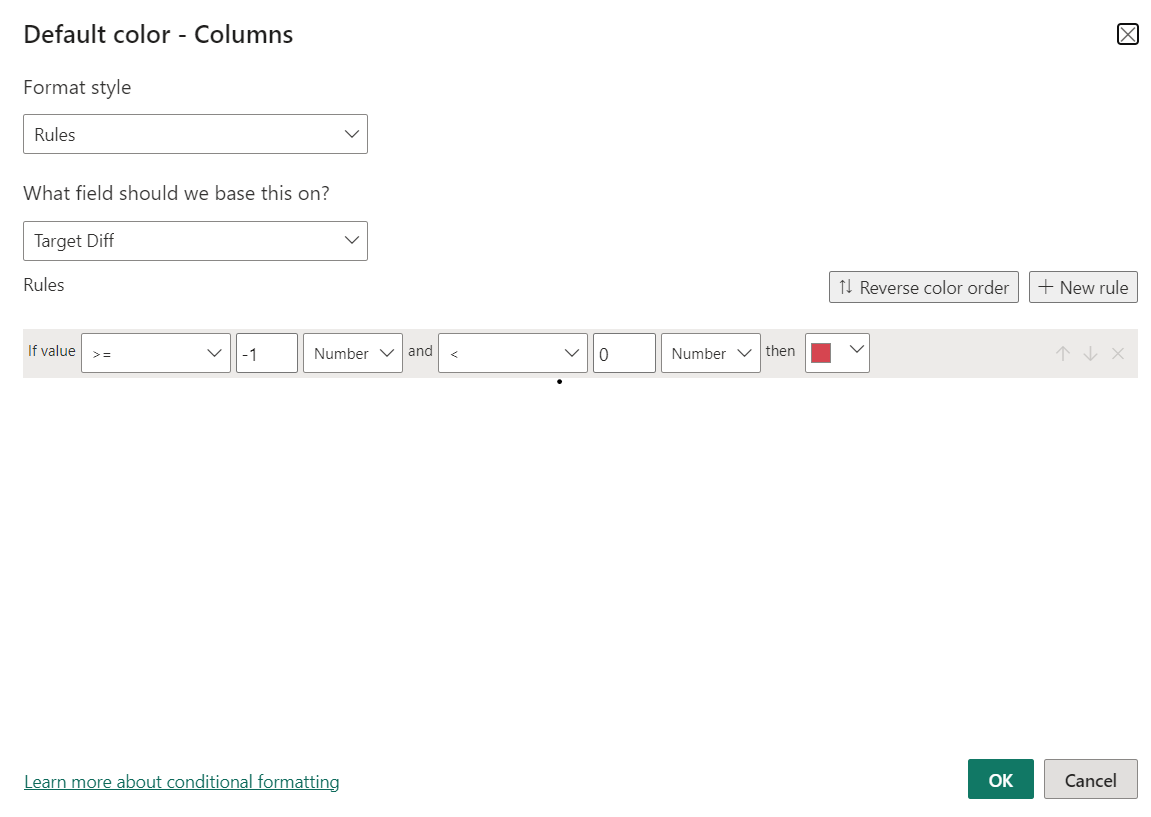
Profit Target = GENERATESERIES(-0.05, 0.15, 0.01)

Profit Target Value = SELECTEDVALUE('Profit Target'[Profit Target])

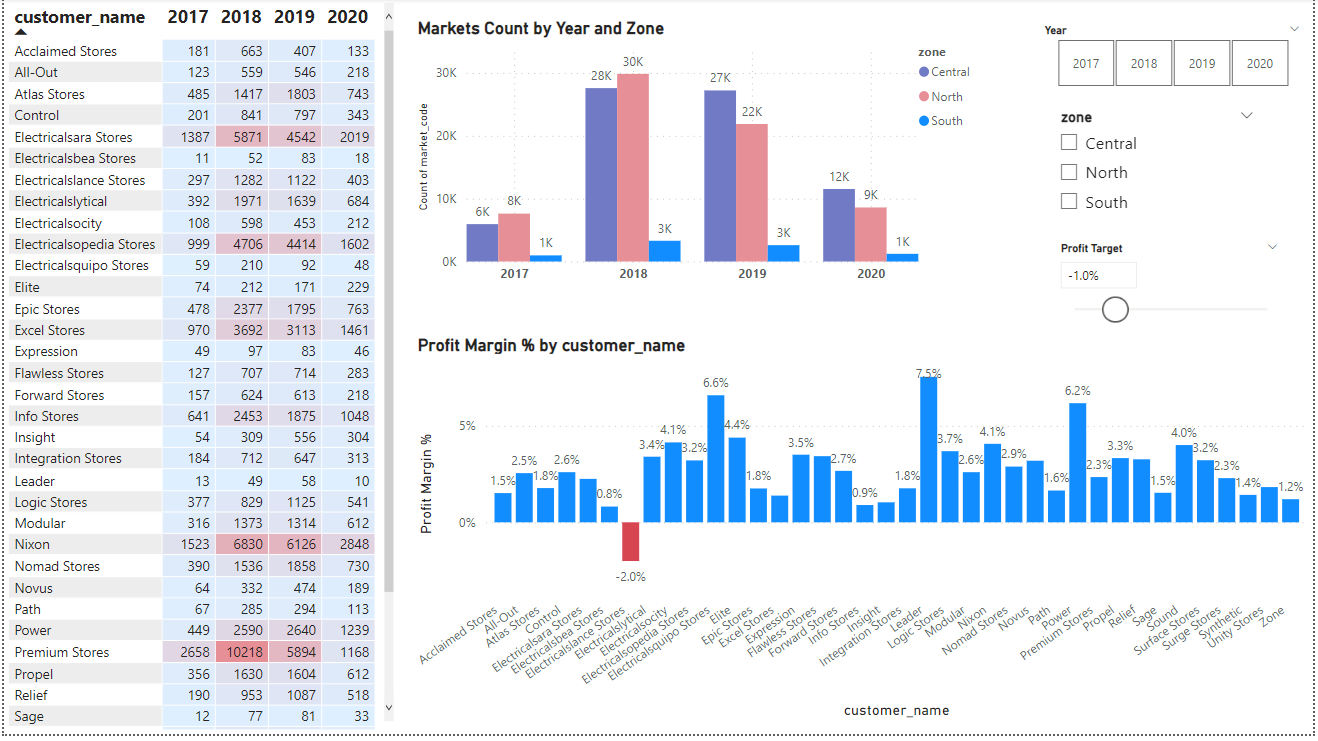
Creating a measure for highlighting the bar: To show up the bar in Red colour if the difference between profit margin and set target value is 0.

Target Diff = [Profit Margin %]-'Profit Target'[Profit Target Value]

Using this measure, we will change the formula to change the colour dynamically.

**

Based on the slicer value(Target profit) we can identify which customers/markets are not able to reach the target profit over the period or zone.

****