

ROWS Column profiling based on top 1000 rows

PREVIEW DOWNLOADED AT 9:23 AM



# ASSIGNMENT



- Create a visual to display no of employees on the basis of seniority level (tenure Years: > 5 Senior, 3-5 Junior, 1-3 Intern, <1 - Temp)
- Grade the employees on the basis of salary (A >300000 / B >100000/ C >50000 /NA <= 50000) and create a visual to get the percentage wise bifurcation of the grade.
- Create a visual to display the promotion percentage.
- Using a funnel chart display City wise bifurcation of salary.
- Create a visual to display city wise count of employees bifurcated on the basis of designation.



# Auto Fill in Power Query

Without writing any code you can fill the data

The FILL Option is available under Transform tab.

The screenshot displays the Microsoft Power Query Editor interface. The 'Transform' tab is active, and the 'Fill' option is highlighted in the ribbon. The main area shows a data table with columns 'State', 'SubCategory', and 'Sales'. The 'State' column has values 'Alabama' and 'Arizona', with many 'null' entries in between. The 'SubCategory' column lists various product categories like 'Accessories', 'Appliances', 'Art', etc. The 'Sales' column contains numerical values. The right sidebar shows 'Query Settings' with 'APPLIED STEPS' including 'Source', 'Navigation', 'Promoted Headers', and 'Changed Type'. At the bottom, a status bar indicates '656 ROWS' and 'Column profiling based on top 1000 rows'.

	State	SubCategory	Sales
1	Alabama	Accessories	2322.83
2		Appliances	208.16
3		Art	301.21
4		Binders	466.82
5		Chairs	3915.54
6		Copiers	899.97
7		Envelopes	105.84
8		Fasteners	3.62
9		Furnishings	40.48
10		Labels	506.49
11		Machines	3040
12		Paper	883.02
13		Phones	2706.28
14		Storage	1249.41
15		Supplies	484.51
16		Tables	2376.46
17	Arizona	Accessories	3396.16
18		Appliances	774.432
19		Art	1389.152
20		Binders	2185.353
21		Bookcases	519.213
22		Chairs	6499.08
23		Envelopes	214.52
24		Fasteners	101.232
25		Furnishings	2511.648



# Grouping in Power Query

**Group By** is used to group the data. This option is available under Transform Menu

The screenshot displays the Microsoft Power Query Editor interface. The 'Transform' tab is selected in the ribbon, and the 'Group By' option is highlighted. A 'Group By' dialog box is open, showing 'SubCategory' as the column to group by and 'Count' as the new column name, with 'Count Rows' as the operation. The background shows a data table with columns 'State', 'SubCategory', and 'Sales'. The 'Query Settings' pane on the right shows the 'APPLIED STEPS' list.

State	SubCategory	Sales
Alabama	Accessories	2322.83
Arizona	Accessories	3396.16
Arizona	Appliances	774.432
Arizona	Art	1389.152
Arizona	Binders	2185.353
Arizona	Bookcases	519.213
Arizona	Chairs	6499.08
Arizona	Envelopes	214.52
Arizona	Fasteners	101.232
Arizona	Furnishings	2511.648



# Transpose Data in Power Query

In **Transpose** rows are converted as columns & columns are converted as Rows. This option is available under Transform Menu

The screenshot displays the Microsoft Power Query Editor interface. The 'Transform' tab is selected in the ribbon, and the 'Transpose' button is highlighted with a red rectangle. A tooltip for 'Transpose' is visible, stating: 'Transpose this table, treating rows as columns and columns as rows.'

The main workspace shows a table with the following data:

	Column2	Column3	Column4	Column5	Column6	
1	Alabama	null	null	null		
2	SubCategory	Accessories	Appliances	Art	Binders	Chairs
3	Sales	2322.83	208.16	301.21	466.82	

The right-hand pane shows the 'Query Settings' for the 'Horizontal' query, with the 'Navigation' step applied.



# Unpivot Data in Power Query

Unpivot data is the way to get detailed data from the summarized data, which can be further used for visualization purpose.

The screenshot displays the Microsoft Power Query Editor interface. The ribbon at the top includes tabs for File, Home, Transform, Add Column, View, Tools, and Help. The 'Transform' tab is active, showing options like 'Unpivot Columns', 'Unpivot Other Columns', and 'Unpivot Only Selected Columns'. A context menu is open over a table, highlighting the 'Unpivot Columns' option. The table contains data for various US states and their corresponding values. The 'Query Settings' pane on the right shows the query name 'unpivot' and the applied steps: 'Source', 'Navigation', 'Promoted Headers', and 'Changed Type'.

State	Value
Alabama	6332.48
Arizona	13525.291
Arkansas	3187.55
California	156064.6015
Colorado	13243.037
Connecticut	5174.987
Delaware	4759.319
District of Columbia	1346.58
Florida	22987.038
Georgia	8321.48
Idaho	2595.482
Illinois	28274.522
Indiana	11496.71
Iowa	2642.31
Kansas	111.12
Kentucky	12126.84
Louisiana	2963.03
Maine	109.48
Maryland	9149.253
Massachusetts	10919.064
Michigan	22321.1
Minnesota	7611.35
Mississippi	4317.85
Missouri	2936.45
Montana	63.98



# Data type in Power Query

Power Query automatically identifies the data type while loading the data. However if required it can be changed.

The screenshot displays the Microsoft Power Query Editor interface. The main area shows a table with three columns: EID, NAME, and DOB. The DOB column is selected, and a context menu is open, listing various data types for conversion. The 'Changed Type' step is visible in the 'APPLIED STEPS' pane on the right.

EID	NAME	DOB
1	1001 RAMESH GUPTA	1.2
2	1002 Sandeep Sharma	\$
3	1003 Rajesh Sharma	123
4	1004 Yogeshwar Sharma	%
5	1005 Manoj Kumar	Date/Time
6	1006 Rohit Gupta	Date
7	1007 Kapil Sharma	Time
8	1008 Archana Sharma	Date/Time/Timezone
9	1009 Ranjeeta Goyal	Duration
10	1010 Komal Singh	Text
11	1011 AMIT KAPOOR	True/False
12	1012 MONIKA ARORA	Binary
13	1013 ROHAN KUMAR	Using Locale...
14	1014 RAJAT KUMAR	
15	1015 ABHISHEK SHARMA	
16	1019 RAJAT KAPOOR	10/11/1983
17	1020 Rohan Kumar	10/10/1989
18	1021 Lalit Sharma	10/12/1989
19	1022 Monika Bajaj	10/12/1992
20	1023 Devender Kapoor	10/10/1982
21	1024 Ravinder Kumar	10/10/1985
22	1025 GAURAV GUPTA	12/10/1987
23	1026 David	12/17/1987
24	1027 KONIKA KAPOOR	10/10/1980
25	1028 RAJAT SHARMA	1/14/1989



# Keep Remove Rows in Power Query

Keep Remove row option is available under Home tab in Power Query. It provide various options to clean the data.

The screenshot displays the Microsoft Power Query Editor interface. The 'Home' tab is selected in the ribbon, and the 'Remove Rows' dropdown menu is open, showing options like 'Remove Top Rows', 'Remove Bottom Rows', 'Remove Alternate Rows', 'Remove Duplicates', 'Remove Blank Rows', and 'Remove Errors'. The main data area shows a table with columns EID, NAME, and DOB. The 'KR-Row' query is selected in the left pane. The right pane shows the 'Query Settings' for 'KR-Row', including 'PROPERTIES' and 'APPLIED STEPS'.

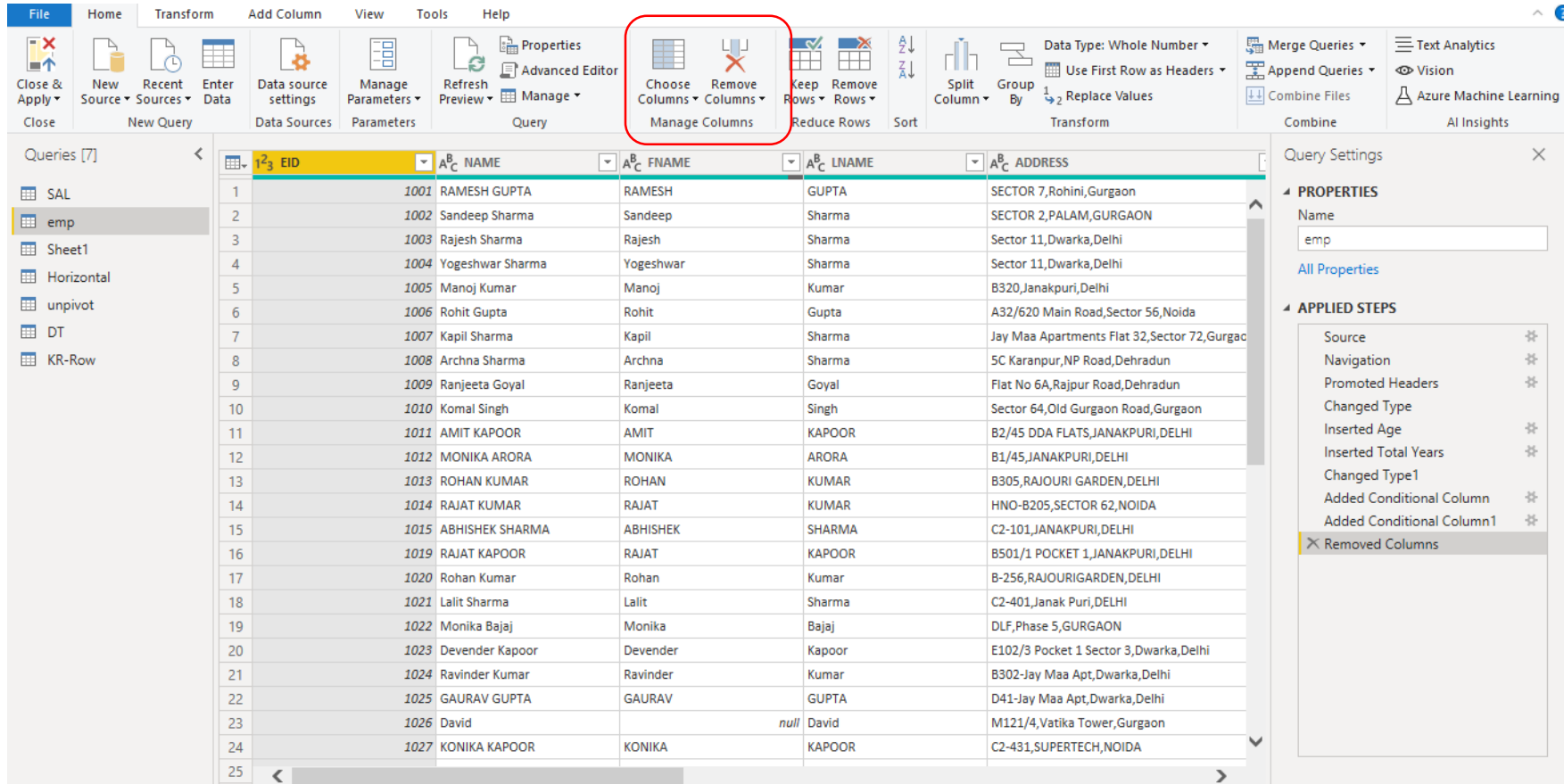
EID	NAME	DOB
1	1001 RAMESH GUPTA	9/1/1990
2	1002 RAMESH GUPTA	9/15/1995
3	1003 Rajesh Sharma	3/16/2001
4	1004 Yogeshwar Sharma	xxx
5	1005 Manoj Kumar	7/1/1985
6	1006 Rohit Gupta	3/31/1992
7	1007 Kapil Sharma	9/28/1987
8	1008 Archana Sharma	5/10/1985
9	1009 Ranjeeta Goyal	12/31/1989
10	1010 Komal Singh	3/31/1990
11	1011 AMIT KAPOOR	1/1/1992
12	1012 MONIKA ARORA	1/1/1982
13	1013 ROHAN KUMAR	10/10/1998
14	1014 RAJAT KUMAR	10/10/1998
15	1015 ABHISHEK SHARMA	10/11/1982
16	1019 RAJAT KAPOOR	10/11/1983
17	1020 Rohan Kumar	10/10/1989
18	1021 Lalit Sharma	10/12/1989
19	1022 Monika Bajaj	10/12/1992
20	1023 Devender Kapoor	10/10/1982
21	1024 Ravinder Kumar	10/10/1985
22	1025 GAURAV GUPTA	12/10/1987
23	1026 David	12/17/1987
24	1027 KONIKA KAPOOR	10/10/1980
25	1028 RAJAT SHARMA	1/14/1989





# Manipulating Columns in Power Query

The Manage Column option under the Home tab allows us to Choose the column, Remove Column, Navigate to a particular column.



The screenshot displays the Microsoft Power Query interface. The 'Home' tab is selected in the ribbon, and the 'Manage Columns' option is highlighted with a red box. The main area shows a table with columns: EID, NAME, FNAME, LNAME, and ADDRESS. The 'Query Settings' pane on the right shows the 'Properties' tab with the query name 'emp' and a list of 'Applied Steps'.

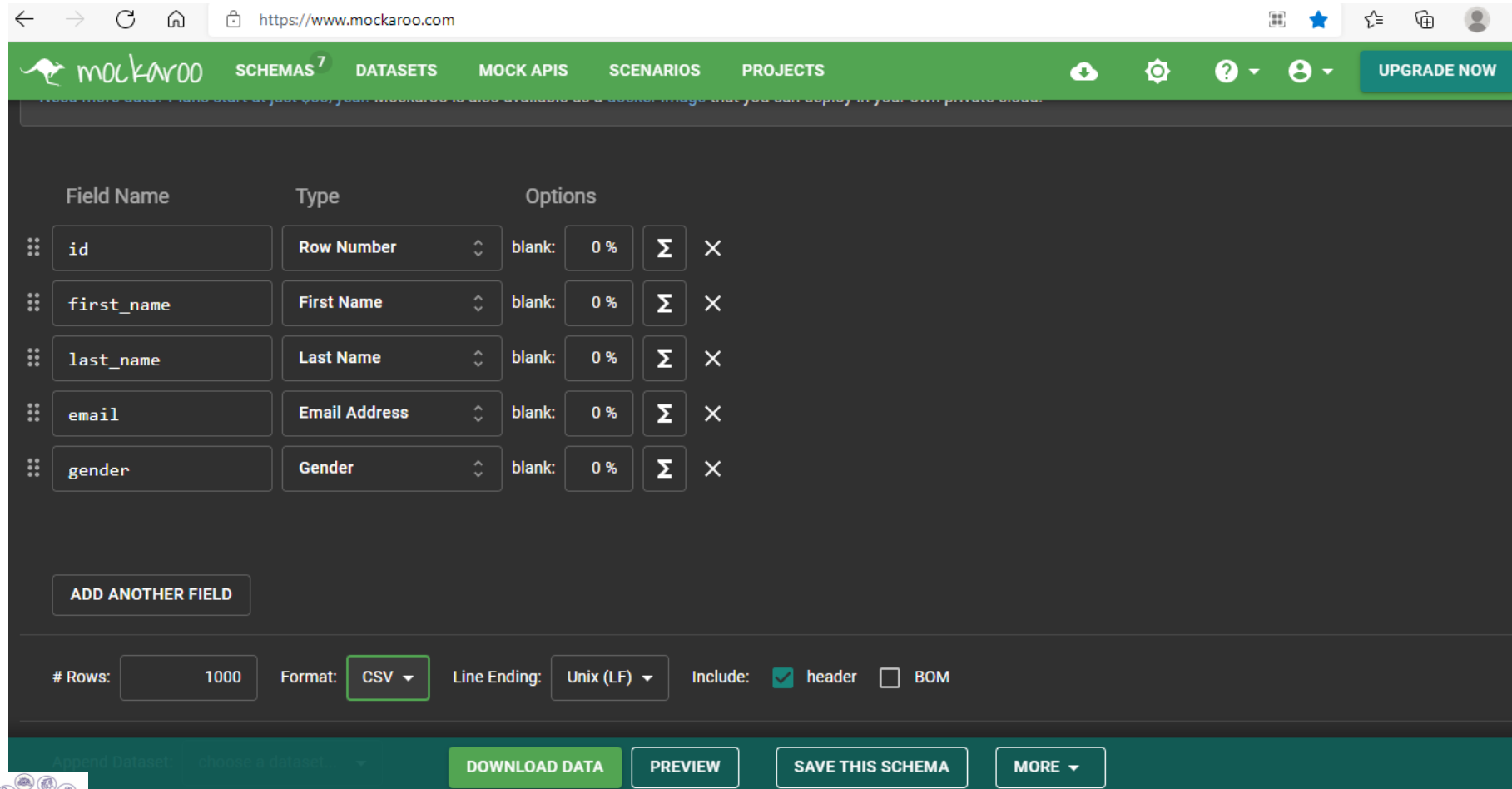
	EID	NAME	FNAME	LNAME	ADDRESS
1	1001	RAMESH GUPTA	RAMESH	GUPTA	SECTOR 7,Rohini,Gurgaon
2	1002	Sandeep Sharma	Sandeep	Sharma	SECTOR 2,PALAM,GURGAON
3	1003	Rajesh Sharma	Rajesh	Sharma	Sector 11,Dwarka,Delhi
4	1004	Yogeshwar Sharma	Yogeshwar	Sharma	Sector 11,Dwarka,Delhi
5	1005	Manoj Kumar	Manoj	Kumar	B320,Janakpuri,Delhi
6	1006	Rohit Gupta	Rohit	Gupta	A32/620 Main Road,Sector 56,Noida
7	1007	Kapil Sharma	Kapil	Sharma	Jay Maa Apartments Flat 32,Sector 72,Gurgaon
8	1008	Archana Sharma	Archana	Sharma	5C Karanpur,NP Road,Dehradun
9	1009	Ranjeeta Goyal	Ranjeeta	Goyal	Flat No 6A,Rajpur Road,Dehradun
10	1010	Komal Singh	Komal	Singh	Sector 64,Old Gurgaon Road,Gurgaon
11	1011	AMIT KAPOOR	AMIT	KAPOOR	B2/45 DDA FLATS,JANAKPURI,DELHI
12	1012	MONIKA ARORA	MONIKA	ARORA	B1/45,JANAKPURI,DELHI
13	1013	ROHAN KUMAR	ROHAN	KUMAR	B305,RAJOURI GARDEN,DELHI
14	1014	RAJAT KUMAR	RAJAT	KUMAR	HNO-B205,SECTOR 62,NOIDA
15	1015	ABHISHEK SHARMA	ABHISHEK	SHARMA	C2-101,JANAKPURI,DELHI
16	1019	RAJAT KAPOOR	RAJAT	KAPOOR	B501/1 POCKET 1,JANAKPURI,DELHI
17	1020	Rohan Kumar	Rohan	Kumar	B-256,RAJOURIGARDEN,DELHI
18	1021	Lalit Sharma	Lalit	Sharma	C2-401,Janak Puri,DELHI
19	1022	Monika Bajaj	Monika	Bajaj	DLF,Phase 5,GURGAON
20	1023	Devender Kapoor	Devender	Kapoor	E102/3 Pocket 1 Sector 3,Dwarka,Delhi
21	1024	Ravinder Kumar	Ravinder	Kumar	B302-Jay Maa Apt,Dwarka,Delhi
22	1025	GAURAV GUPTA	GAURAV	GUPTA	D41-Jay Maa Apt,Dwarka,Delhi
23	1026	David		David	M121/4,Vatika Tower,Gurgaon
24	1027	KONIKA KAPOOR	KONIKA	KAPOOR	C2-431,SUPERTECH,NOIDA



# Mackaroo- Random Data Generator

It is a free data generator tool which lets you create custom datasets as per the specified schema to test and demo your application.

<https://www.mockaroo.com>



The screenshot displays the Mockaroo web application interface. At the top, there is a navigation bar with links for SCHEMAS, DATASETS, MOCK APIS, SCENARIOS, and PROJECTS. Below this, a table defines the schema for a dataset with five fields: id, first\_name, last\_name, email, and gender. Each field has a specific type (Row Number, First Name, Last Name, Email Address, Gender) and options for blank values and a sum symbol. At the bottom, there are controls for the number of rows (1000), the output format (CSV), line ending (Unix (LF)), and whether to include a header or BOM. A 'DOWNLOAD DATA' button is prominently displayed.

Field Name	Type	Options
id	Row Number	blank: 0 % Σ ×
first_name	First Name	blank: 0 % Σ ×
last_name	Last Name	blank: 0 % Σ ×
email	Email Address	blank: 0 % Σ ×
gender	Gender	blank: 0 % Σ ×

ADD ANOTHER FIELD

# Rows: 1000 Format: CSV Line Ending: Unix (LF) Include: ☒ header ☐ BOM

Append Dataset: choose a dataset

DOWNLOAD DATA PREVIEW SAVE THIS SCHEMA MORE



# Mackaroo- Random Data Generator

Download dataset using Mackroo.com to create the following tables:

## **Student**

Student ID

Student Name

Gender

DOB

Birth Country

## **Citizenship**

Status (**Active**/Deferred/Withdrawn)

Reason for withdrawal  
(Homesick /Academically  
Unprepared/Expensive)

## **Feedback with 5 questions**

SID

Q01 - Professors teach well at this university?

Q02 - Teacher encourage us to perform better?

Q03 - I feel job-competent moving forward  
from this university?

Q04- Lecturers were proficient enough?

Q05 - I would like to recommend this  
educational institution to others?

**Responses** : Strongly Agree - 5

Moderately Agree - 4

Neither Agree or Disagree - 3

Moderately Disagree - 2

Strongly Disagree - 1



# Data Modelling in Power BI

Data Modeling is used to connect multiple data sources using relationship and allows us to create visuals based on multiple data sources.

To create a data model we need to switch to Model view.



Report View - Allows us to create the visuals



Data View – to see the data from various sources

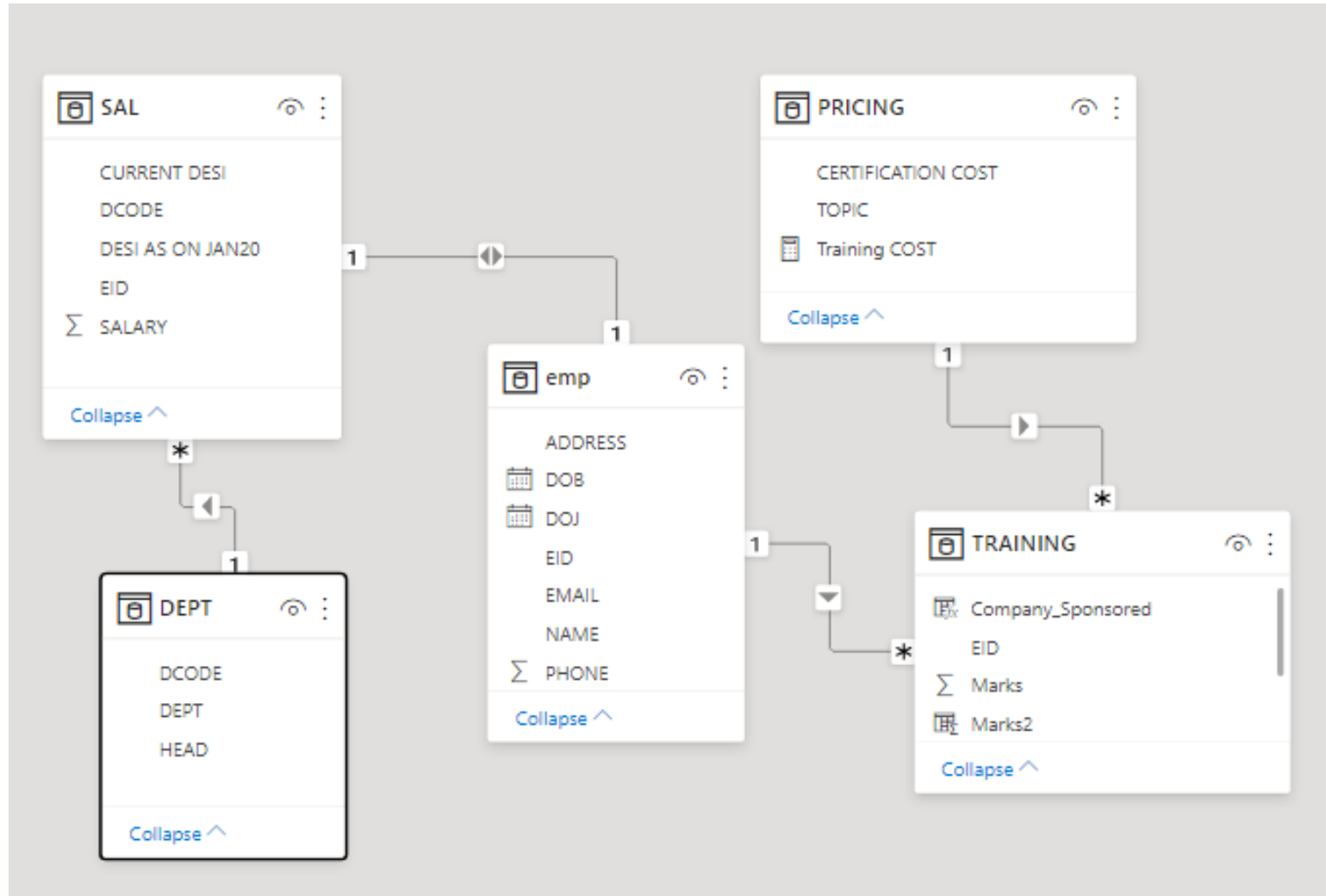


Model View – to manage relationships



# Data Modelling in Power BI

The below model shows the relation ship between Employee, Salary, Dept, Pricing & Training Tables.



# Data Modelling in Power BI

The objective of the model is to analyze the training status on the basis of department head

## Training Analysis:

- No of Training applied under each head

- No of training completed under each head

- Training result of each head

- No. of company Sponsored Trainings under each head (*Marks  $\geq 95$  is CS*)

- Training applied for each module

- Percentage of self & company sponsored training

## Cost Analysis:

- Department wise total salary

- Department wise cost of company sponsored training

- Bifurcation of company & self sponsored training under each module

- Total no of trainings conducted under each module along with the total cost.



# Data Modelling in Power BI



Data View – to see the data from various sources, new columns & measures can also be created in this view.

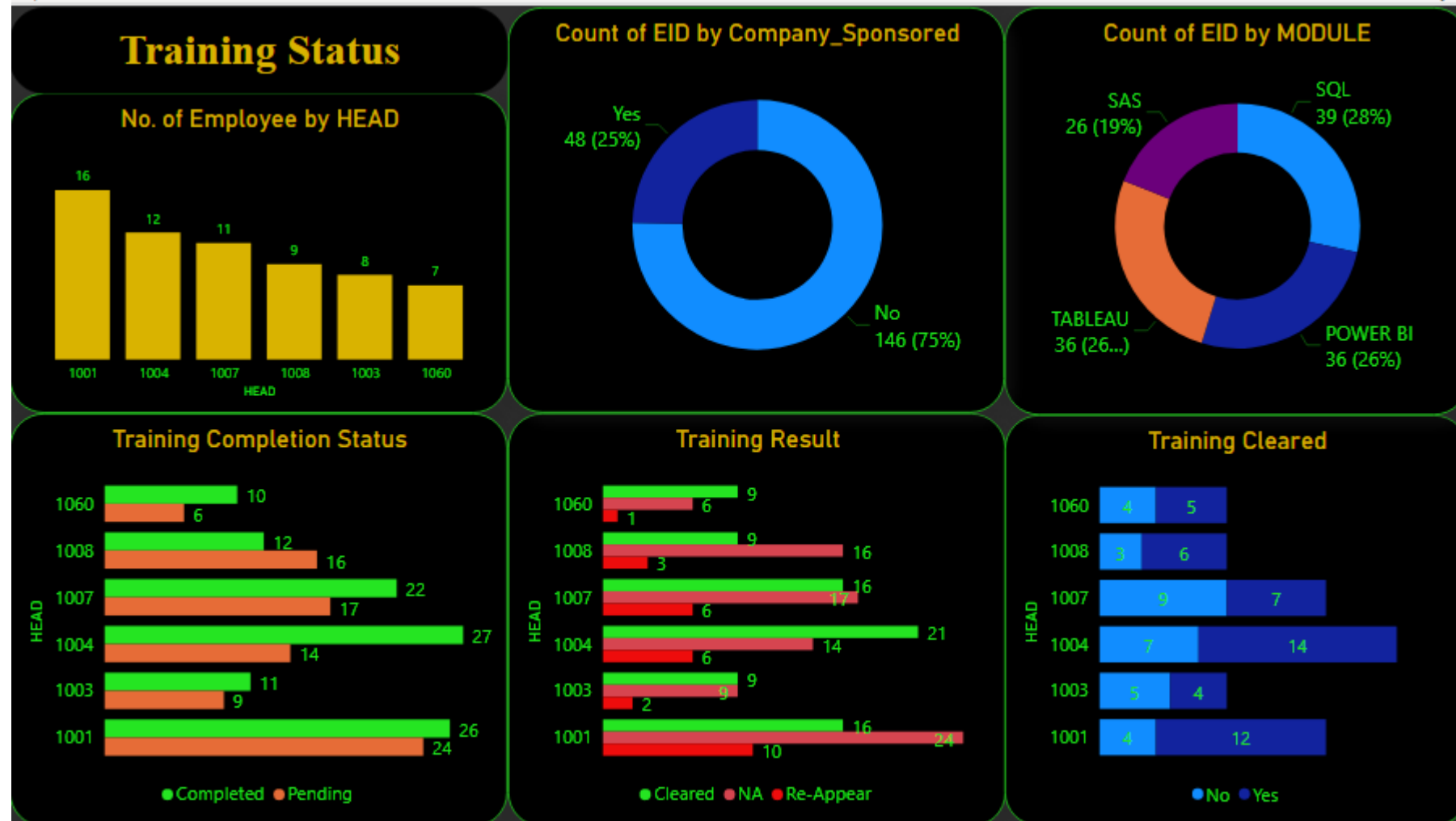
> Column Tools – New Column

The screenshot displays the Power BI Data View interface. The 'Column tools' ribbon is active, showing options for Name, Data type, Format, Summarization, Data category, Sort by column, Data groups, Manage relationships, and New column. The 'New column' button is highlighted. Below the ribbon, a formula bar shows the DAX formula: `1 RESULT = IF(TRAINING[FINALMARKS]>=85, "CLEAR", IF(TRAINING[FINALMARKS]=0, "NA", "RE-APPEAR"))`. The main area shows a table with columns: EID, MODULE, STATUS, Marks, RESULT, and FINALMARKS. The 'RESULT' column is highlighted. The 'Fields' pane on the right shows a search bar and a list of fields: DEPT, DCODE, DEPT, HEAD, emp, PRICING, SAL, TRAINING, EID, FINALMARKS, Marks, MODULE, RESULT, and STATUS. The 'RESULT' field is selected.

EID	MODULE	STATUS	Marks	RESULT	FINALMARKS
E1065	TABLEAU	Pending	81	NA	0
E1053	TABLEAU	Pending	66	NA	0
E1052	TABLEAU	Pending	82	NA	0
E1038	SQL	Pending	83	NA	0
E1003	POWER BI	Pending	70	NA	0
E1003	SQL	Pending	79	NA	0
E1022	TABLEAU	Pending	89	NA	0
E1029	SQL	Pending	97	NA	0
E1040	POWER BI	Pending	62	NA	0
E1041	POWER BI	Pending	75	NA	0
E1056	SAS	Pending	93	NA	0
E1034	TABLEAU	Pending	62	NA	0
E1056	SAS	Pending	68	NA	0
E1028	TABLEAU	Pending	62	NA	0
E1035	POWER BI	Pending	95	NA	0
E1066	POWER BI	Pending	68	NA	0
E1035	TABLEAU	Pending	62	NA	0
E1044	POWER BI	Pending	61	NA	0
E1058	POWER BI	Pending	93	NA	0

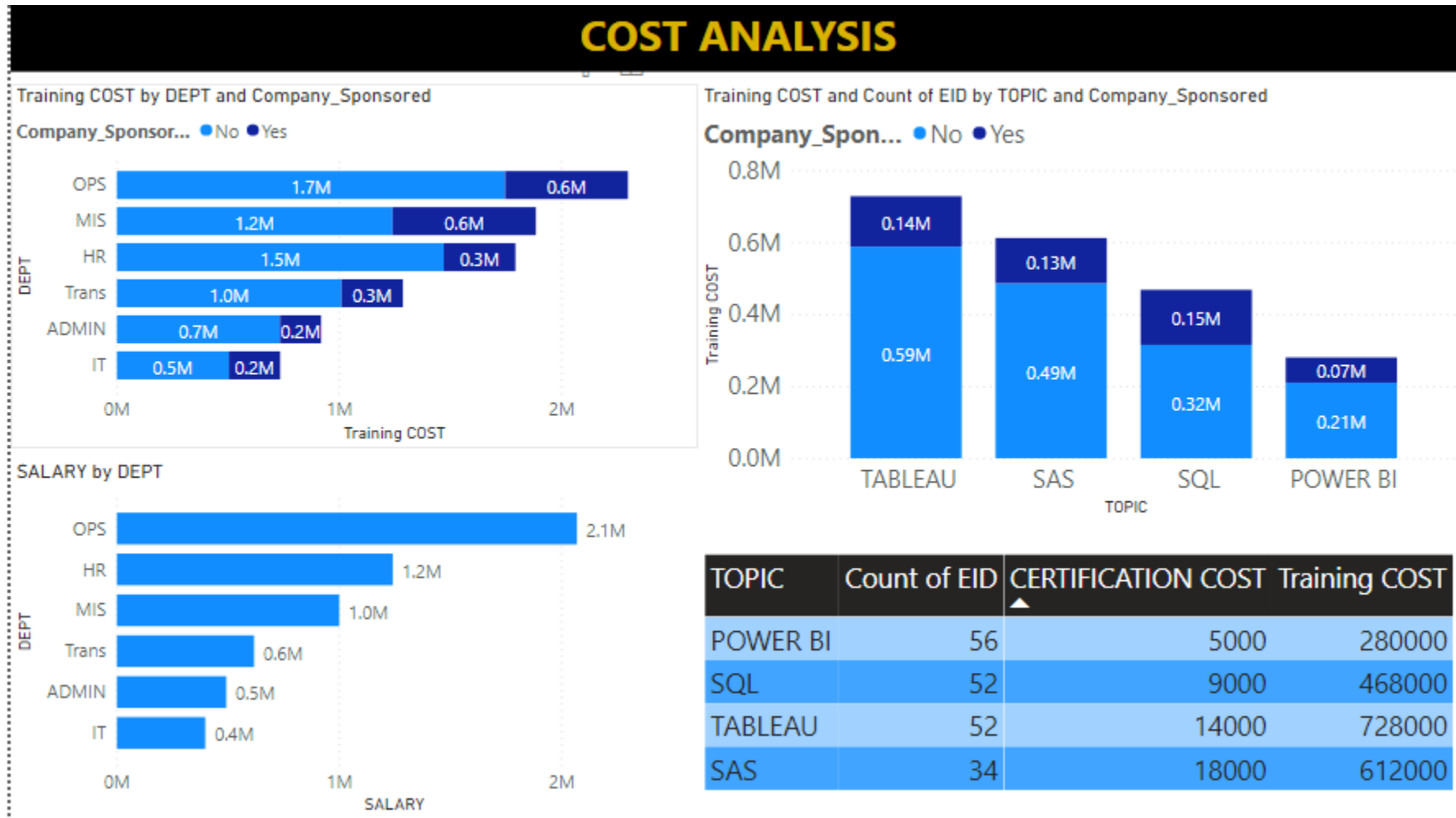


# Data Modelling in Power BI



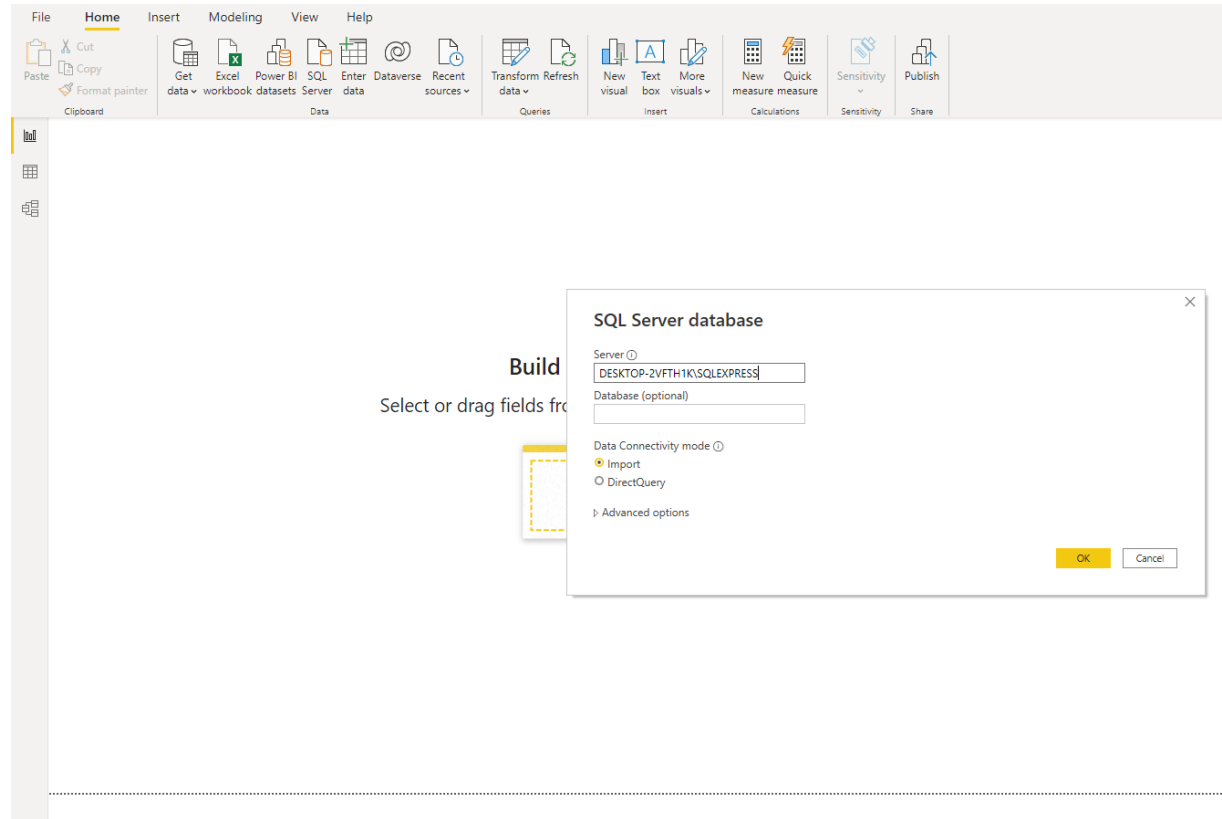


# Data Modelling in Power BI



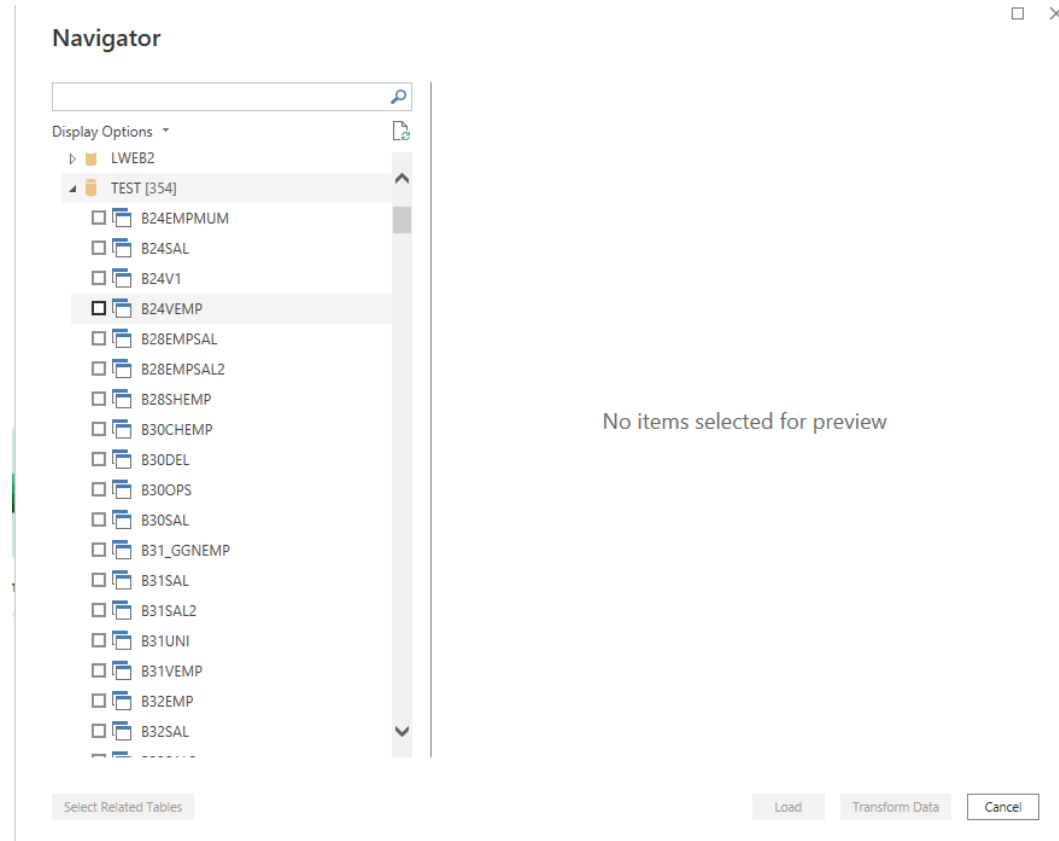
# Connecting to SQL

To connect to SQL we just need to specify the server name and click on OK



# Connecting to SQL

Once connected all the database will be displayed. Go to the desired database, select the table you need to connect and select Load / Transform

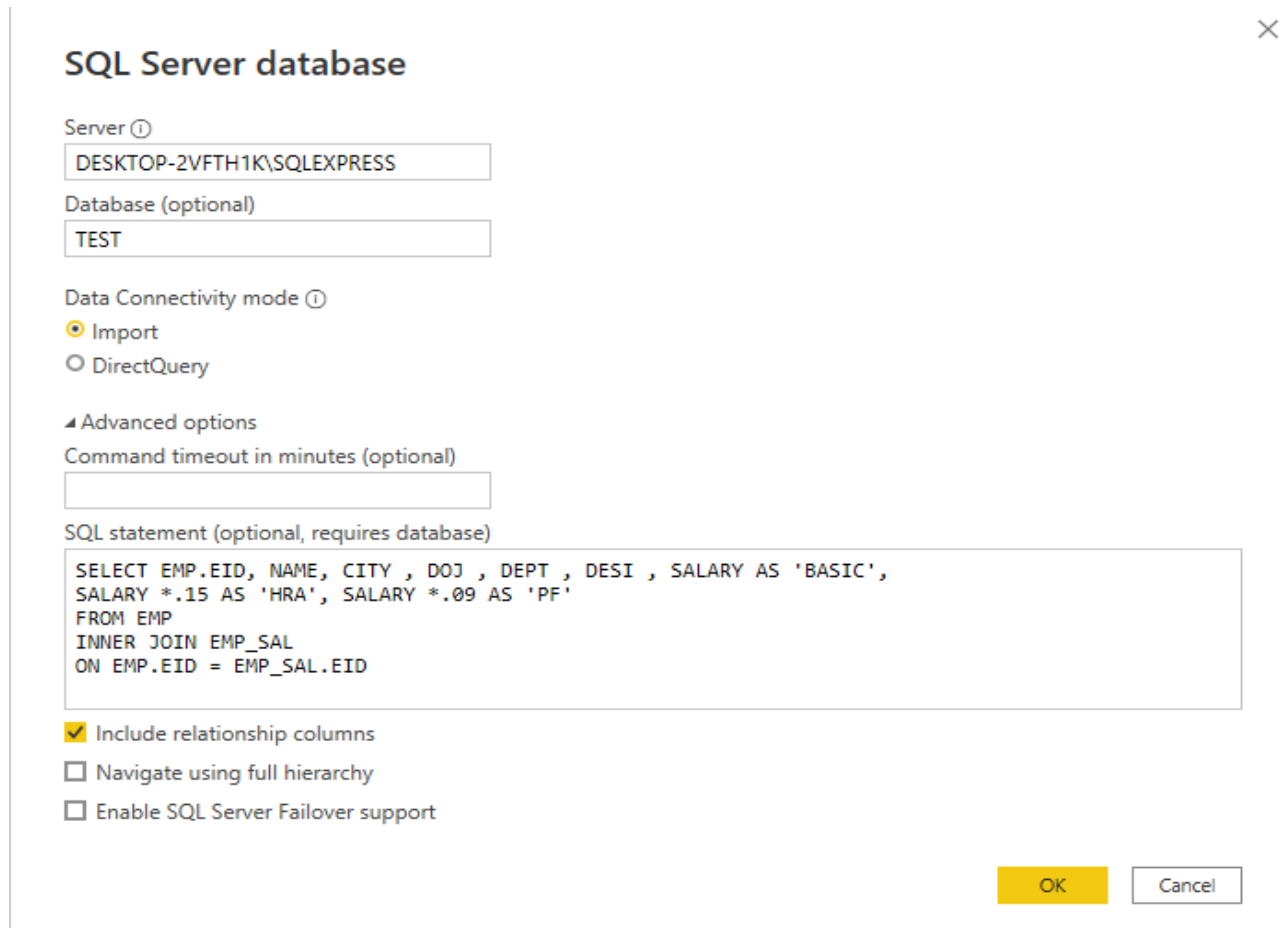


We can get the data from multiple tables and create the model in Power BI



# Connecting to SQL

We can also write the SQL query to get the desired data from multiple tables and create a single query in Power Bi.



**SQL Server database**

Server ①  
DESKTOP-2VFTH1K\SQLEXPRESS

Database (optional)  
TEST

Data Connectivity mode ①  
☒ Import  
☐ DirectQuery

Advanced options

Command timeout in minutes (optional)

SQL statement (optional, requires database)  
SELECT EMP.EID, NAME, CITY , DOJ , DEPT , DESI , SALARY AS 'BASIC',  
SALARY \*.15 AS 'HRA', SALARY \*.09 AS 'PF'  
FROM EMP  
INNER JOIN EMP\_SAL  
ON EMP.EID = EMP\_SAL.EID

☒ Include relationship columns  
☐ Navigate using full hierarchy  
☐ Enable SQL Server Failover support

OK Cancel



Note : Changes done in data source will be automatically reflected in the report

*Thanks!*

***EVERY ENDING  
IS REALLY JUST A  
NEW BEGINNING***



**Rajeev Garg  
Data Analytics Trainer**