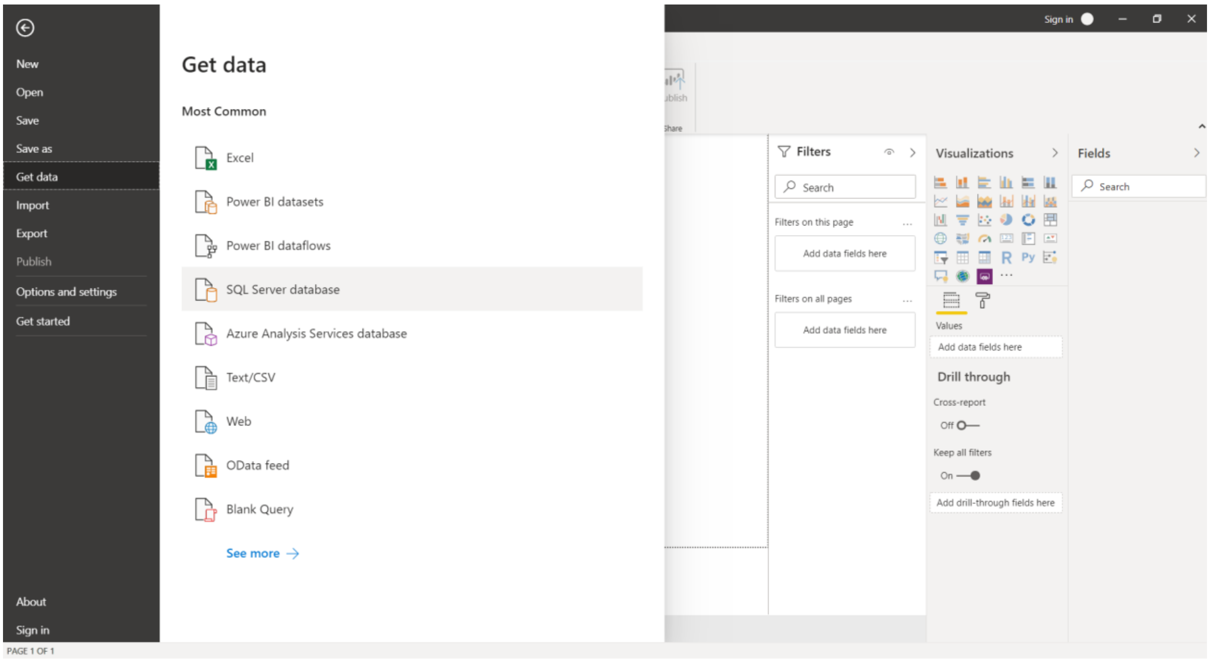
Kishinchand Chellaram College, Mumbai-2020
FY / SY / TY B Sc. (I.T) Semester ________

Practical No: - 1

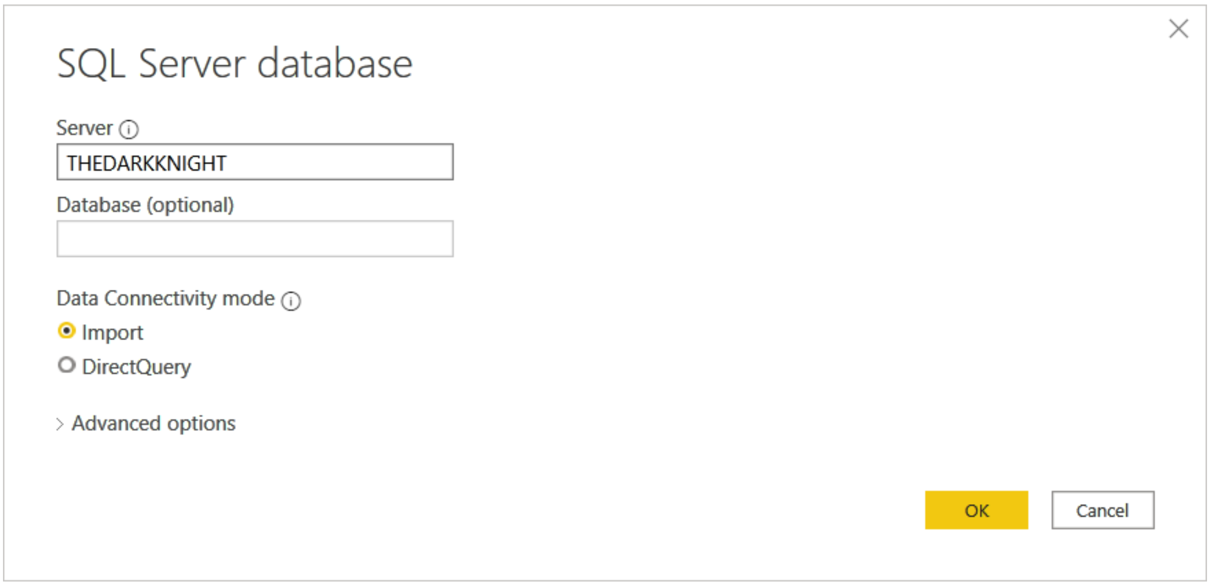
Aim: - Import the legacy data from different sources such as (Excel, SqlServer, Oracle etc.) and load in the target system. (You can download sample database such as Adventureworks, Northwind, foodmart etc.)

Steps: -

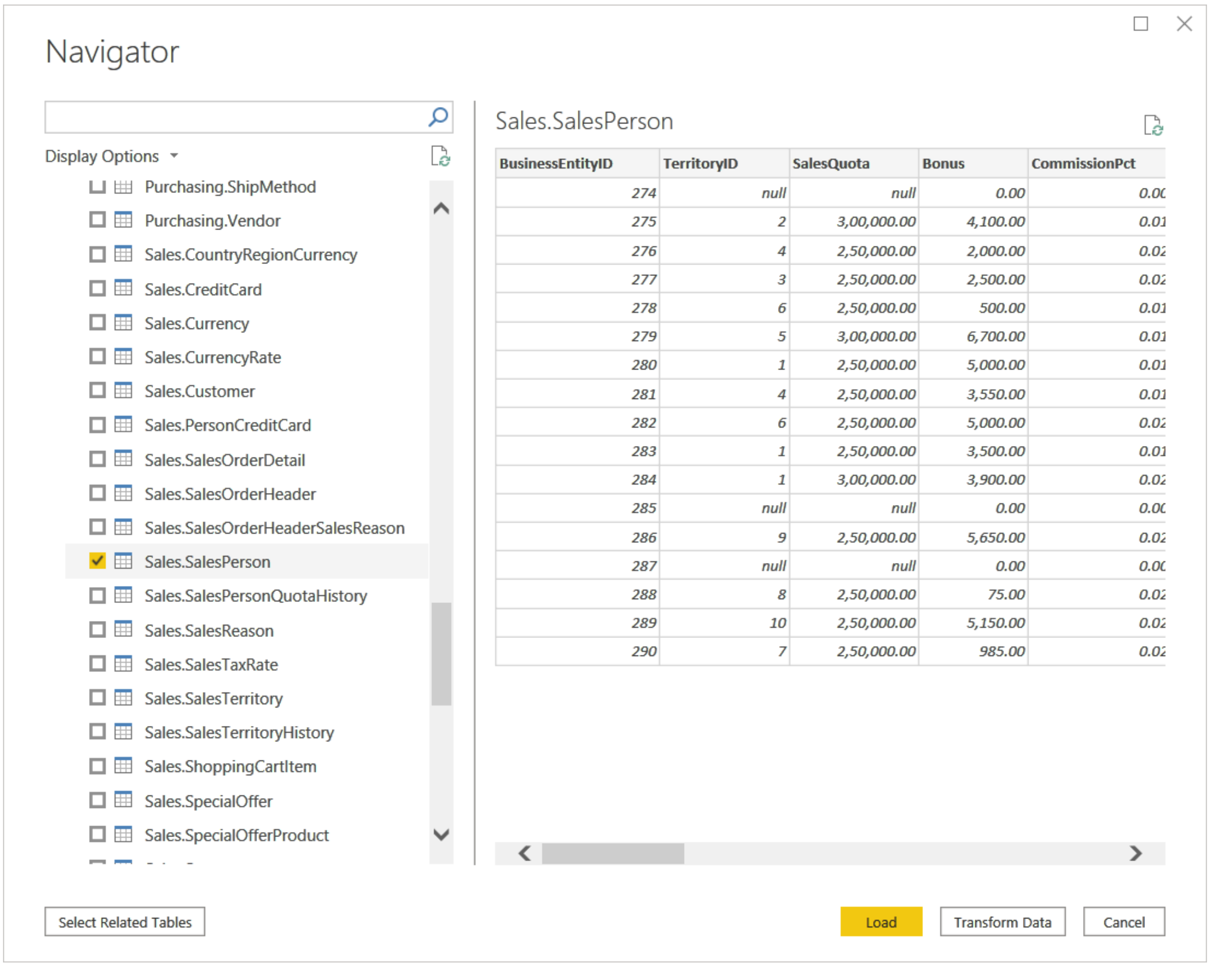
1. Open PowerBi Desktop -> Get Data -> SQL Server Database

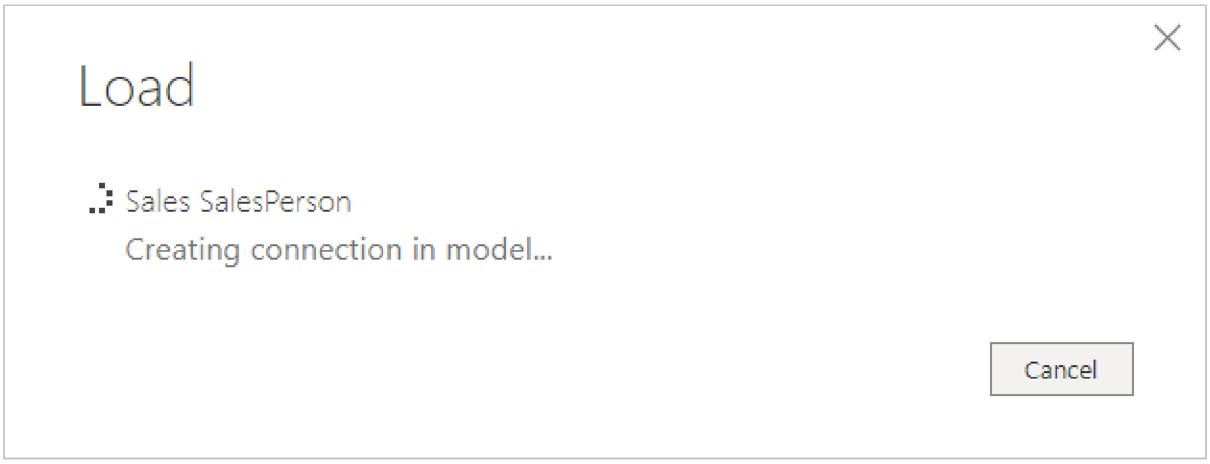


2. Type your server name -> OK

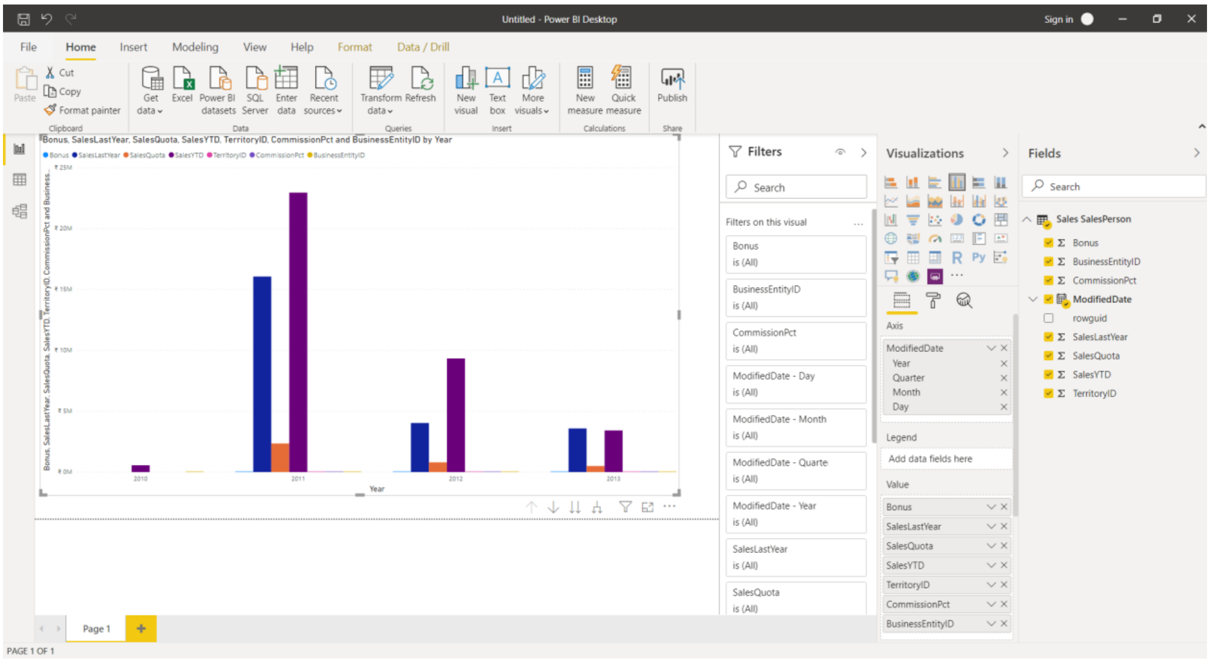


3. Navigator will show your database -> select any table -> Load





4. Finally convert data into graph and display the output.

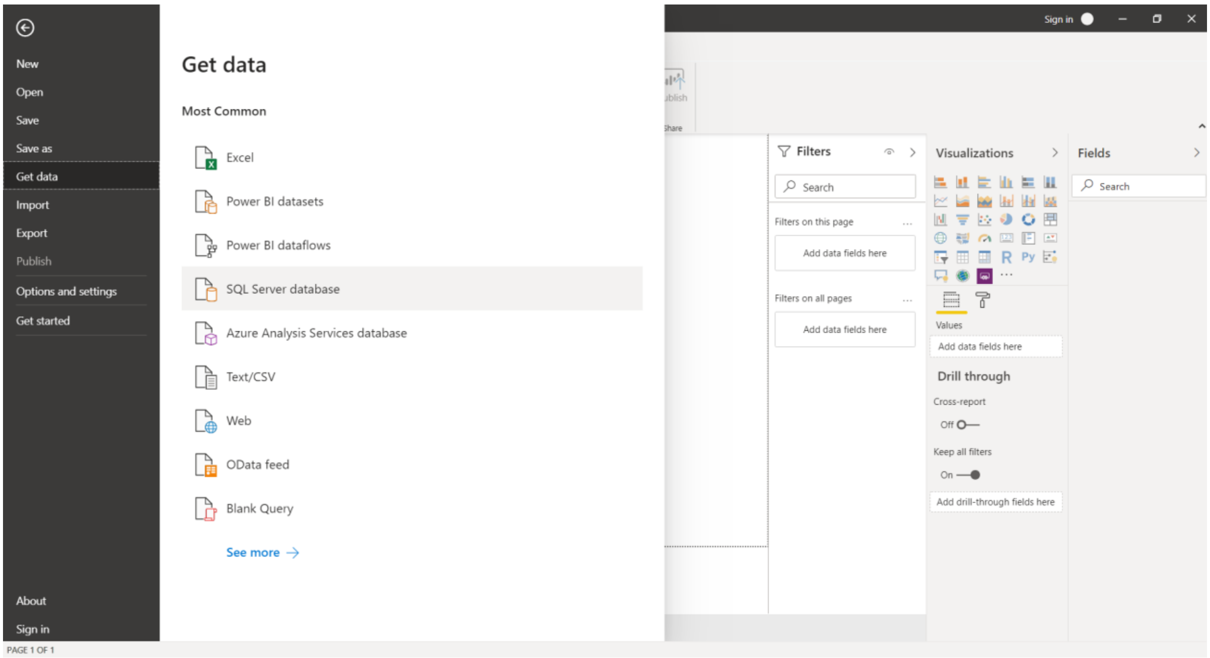


Practical No: - 2

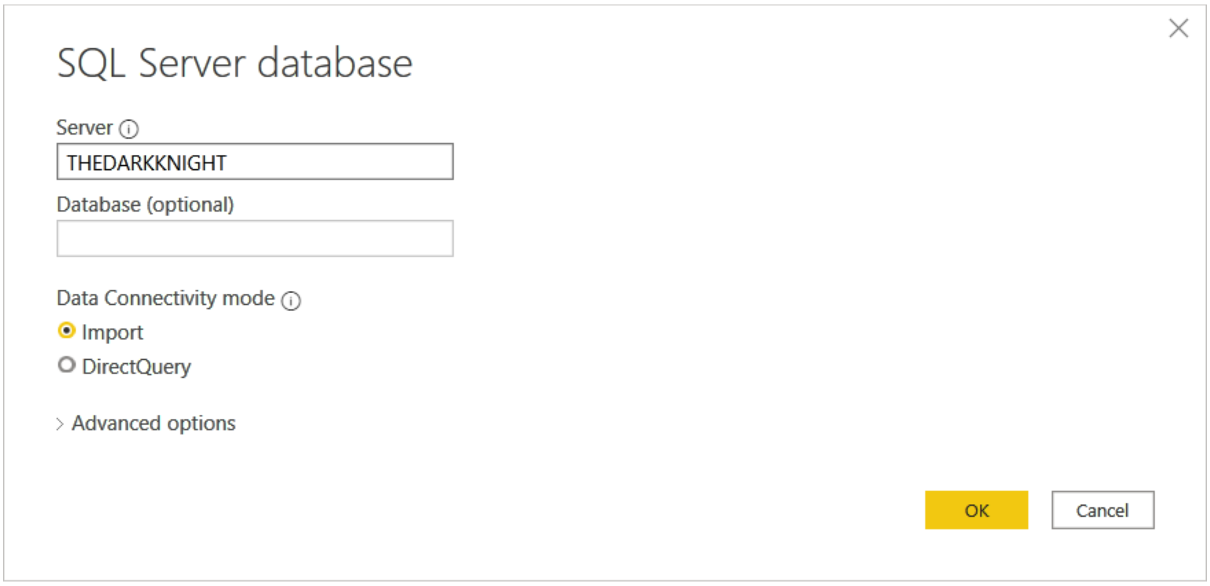
Aim: - Perform the Extraction Transformation and Loading (ETL) process to construct the database in the SqlServer.

Steps: -

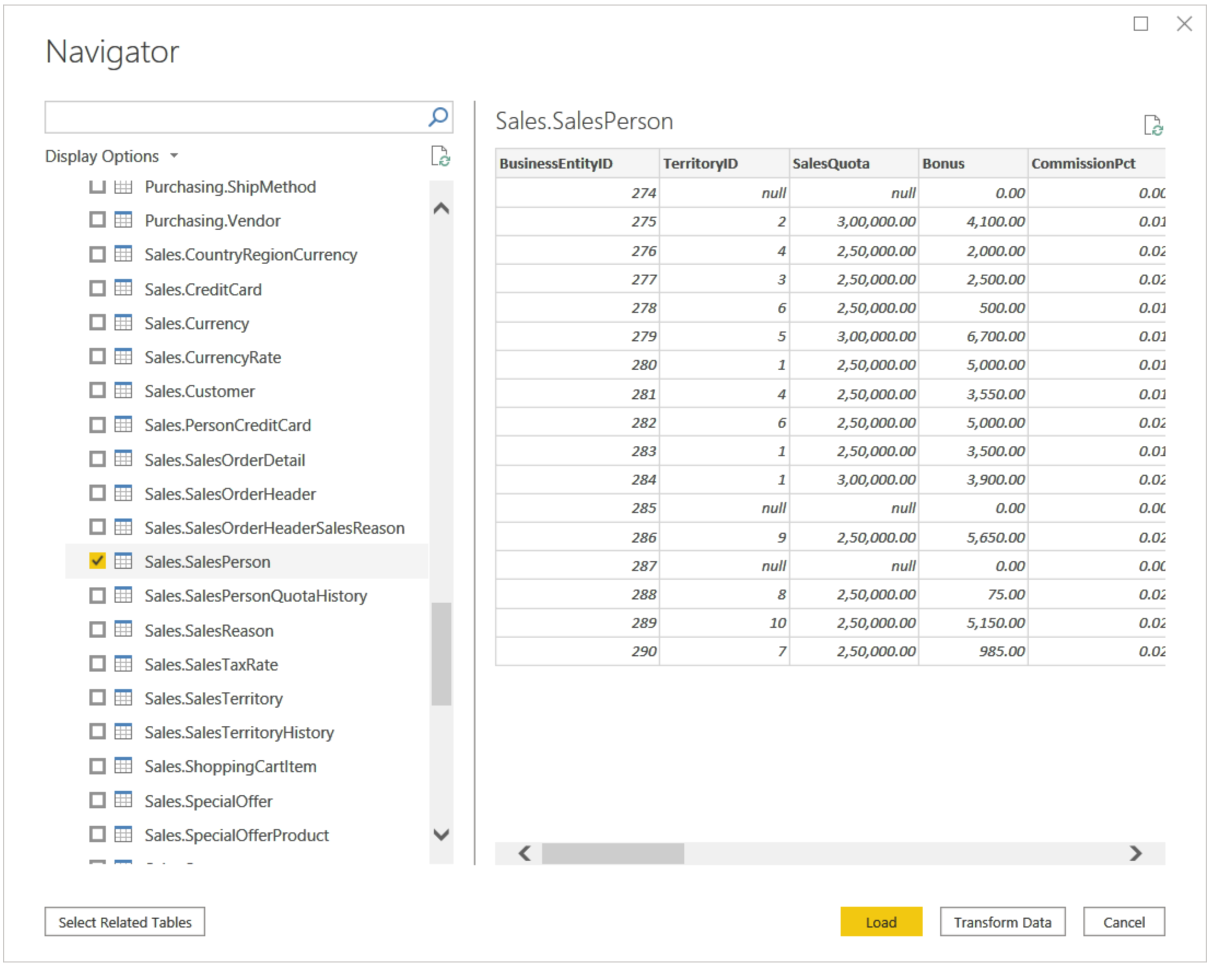
1. Open PowerBi Desktop -> Get Data -> SQL Server Database



2. Type your server name -> OK

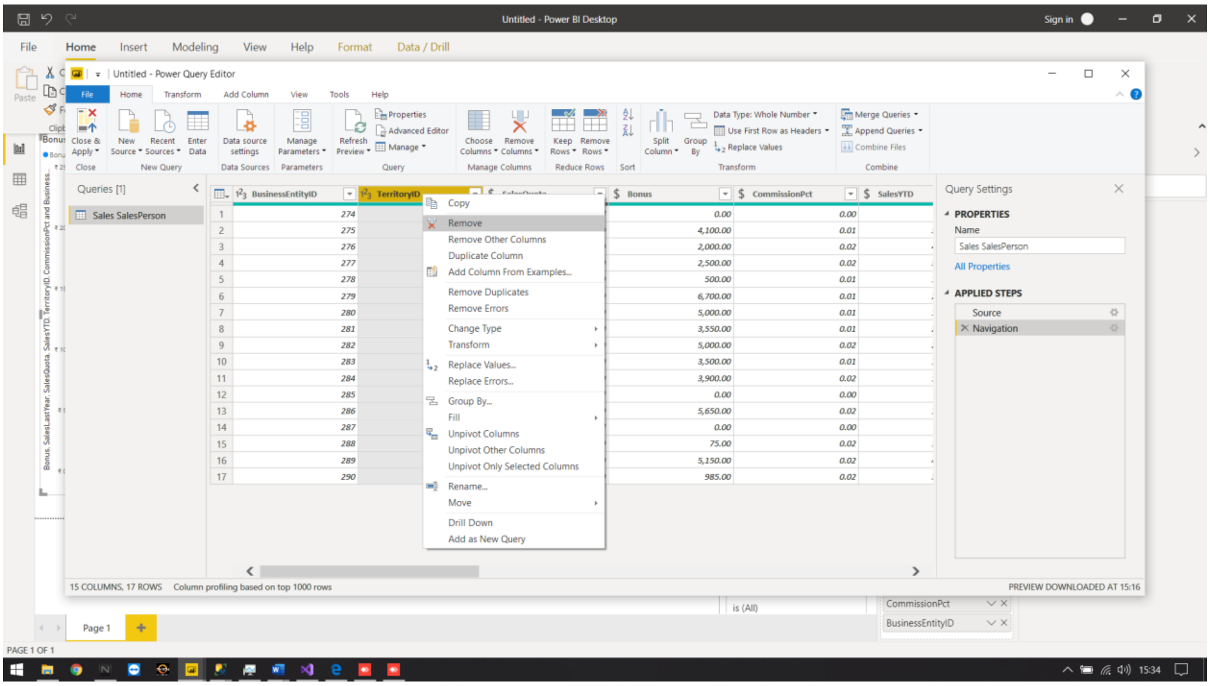


3. Navigator will show your database -> select any table -> Transform Data

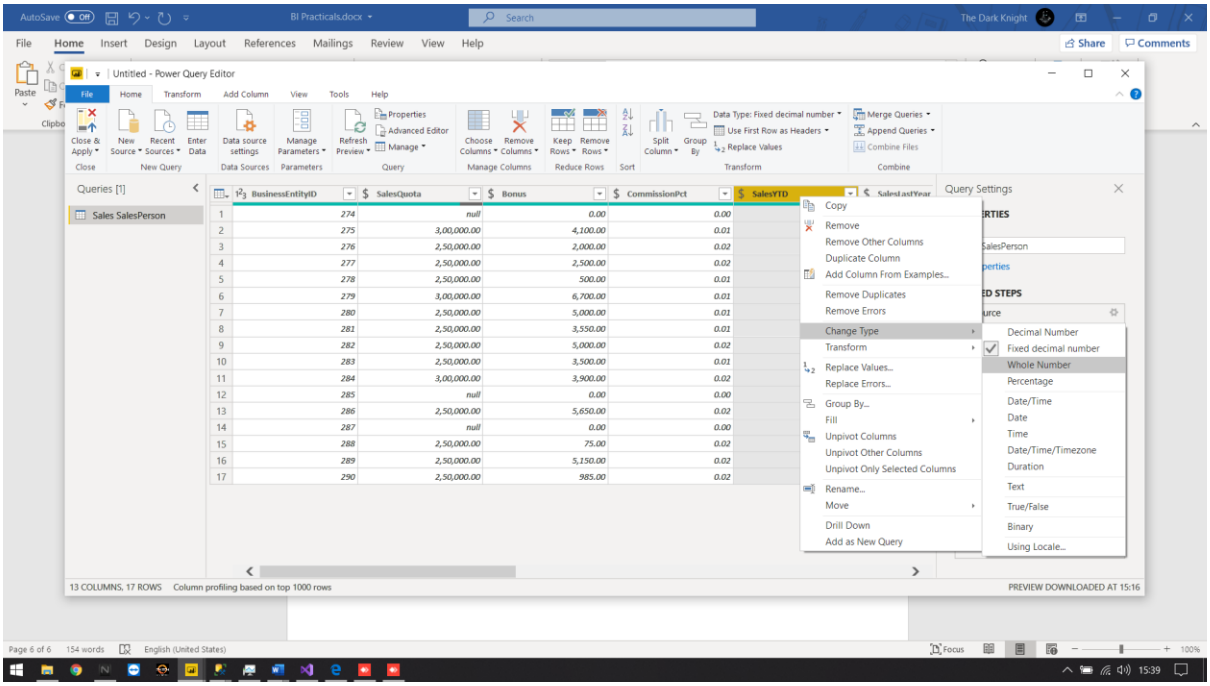


BusinessEntityID	TerritoryID	SalesQuota	Bonus	CommissionPct
274	null	null	0.00	0.00
275	2	3,00,000.00	4,100.00	0.01
276	4	2,50,000.00	2,000.00	0.02
277	3	2,50,000.00	2,500.00	0.02
278	6	2,50,000.00	500.00	0.01
279	5	3,00,000.00	6,700.00	0.01
280	1	2,50,000.00	5,000.00	0.01
281	4	2,50,000.00	3,550.00	0.01
282	6	2,50,000.00	5,000.00	0.02
283	1	2,50,000.00	3,500.00	0.01
284	1	3,00,000.00	3,900.00	0.02
285	null	null	0.00	0.00
286	9	2,50,000.00	5,650.00	0.02
287	null	null	0.00	0.00
288	8	2,50,000.00	75.00	0.02
289	10	2,50,000.00	5,150.00	0.02
290	7	2,50,000.00	985.00	0.02

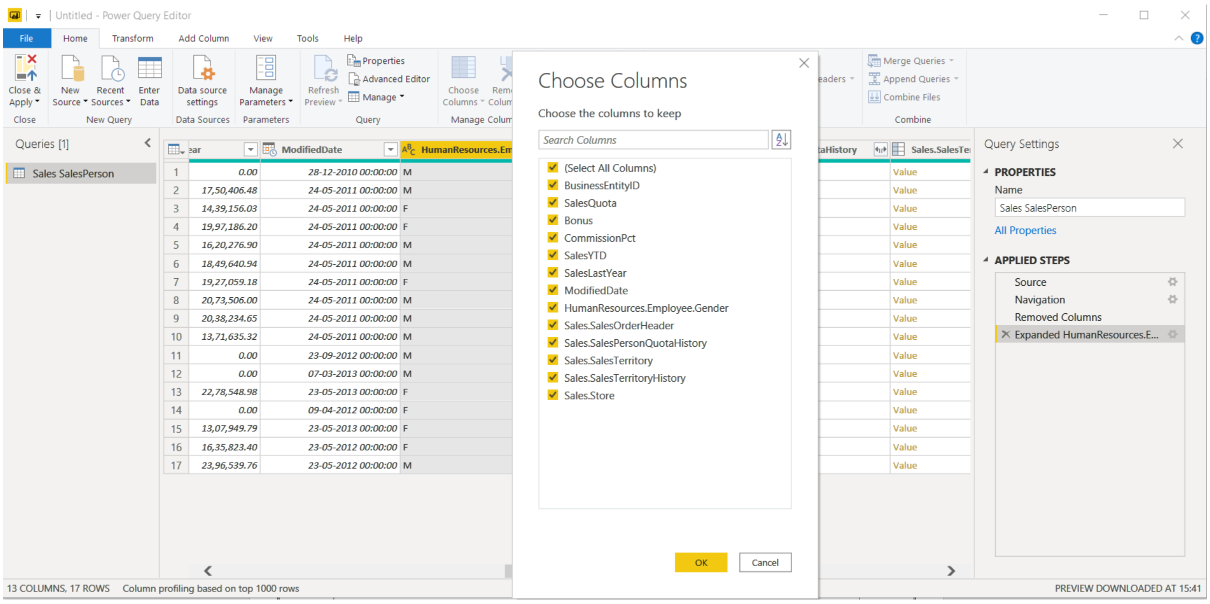
4. Manage Columns -> Remove Columns with null values and unnecessary data to display only what is required.

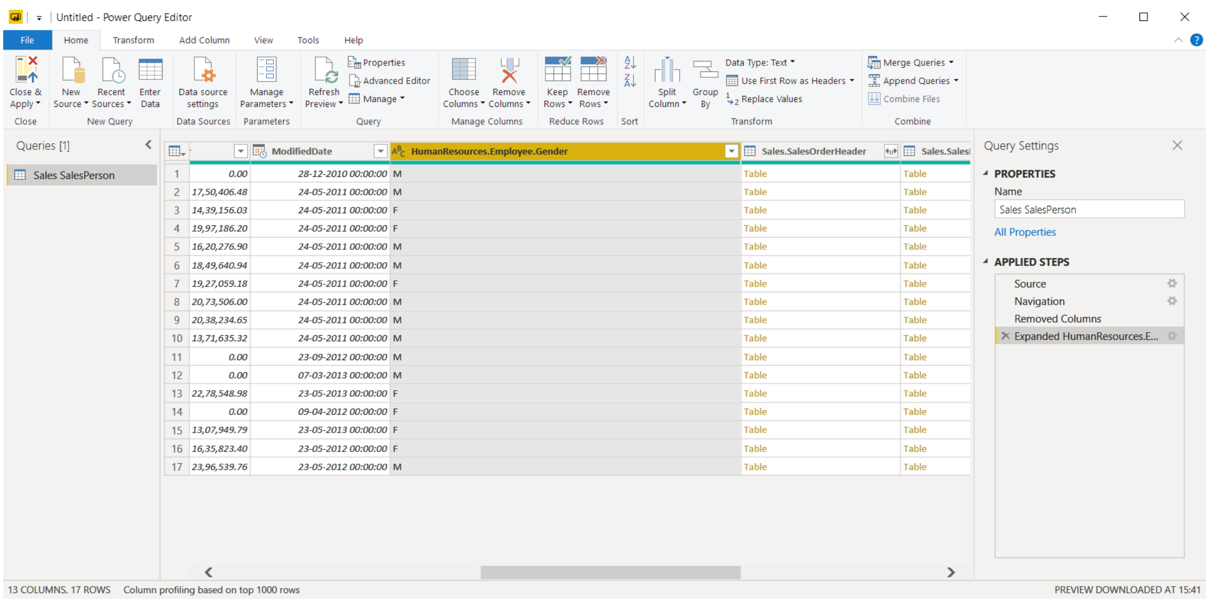


5. Manage Columns -> Convert Column -> Change data unit into Whole Number.

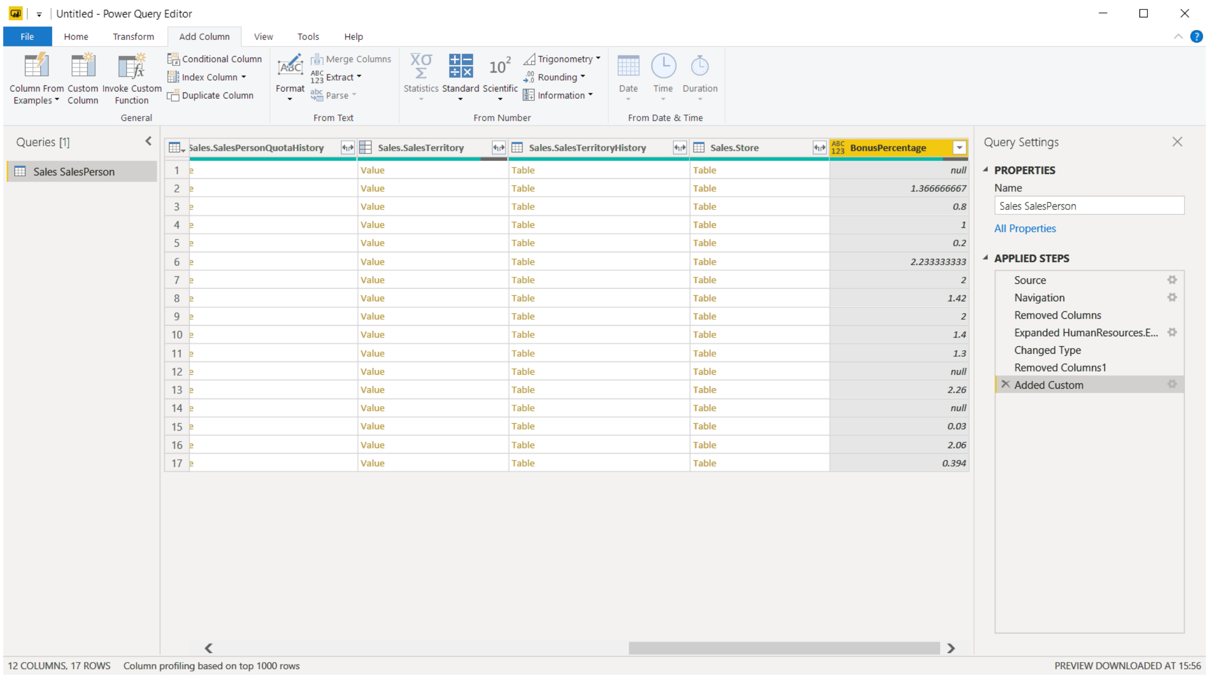


6. Manage Columns -> Expand another table (Human Resources) -> add selected columns of choice.

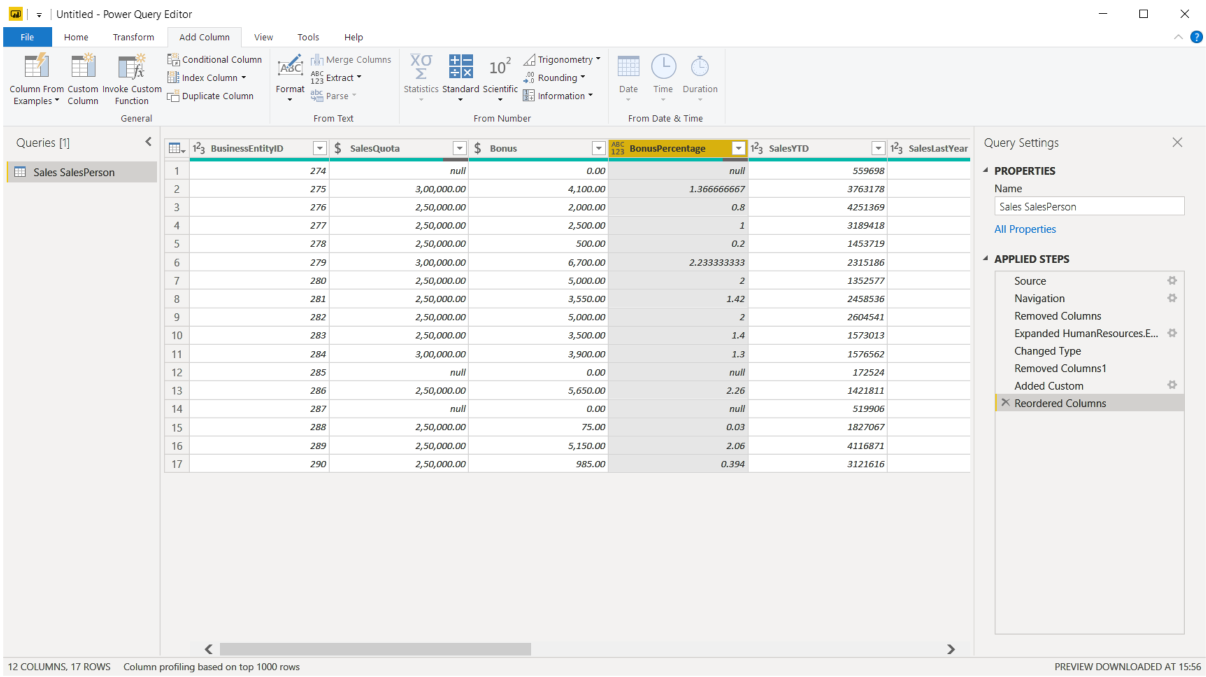


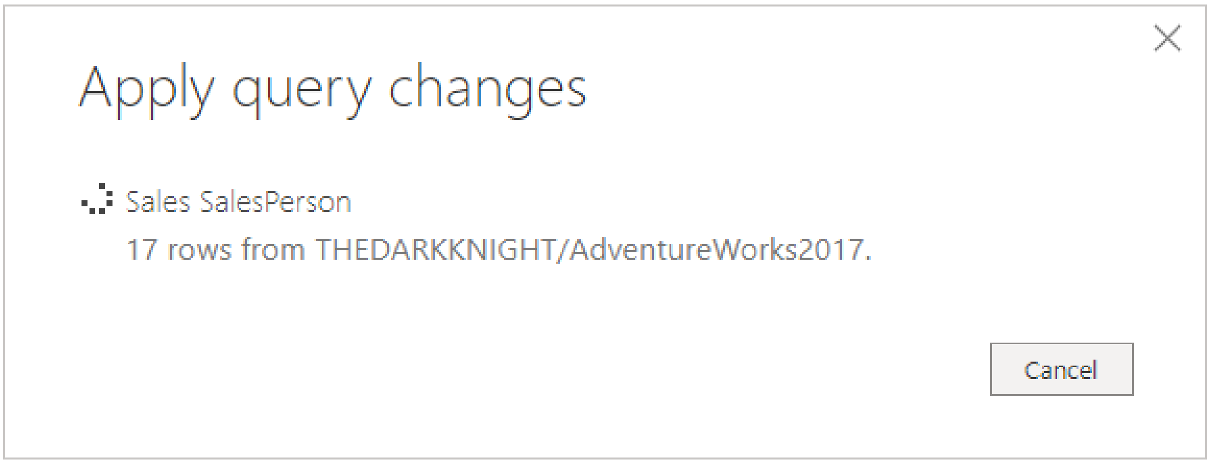


7. Add Custom Column -> Calculate a new column value (Bonus Percentage)

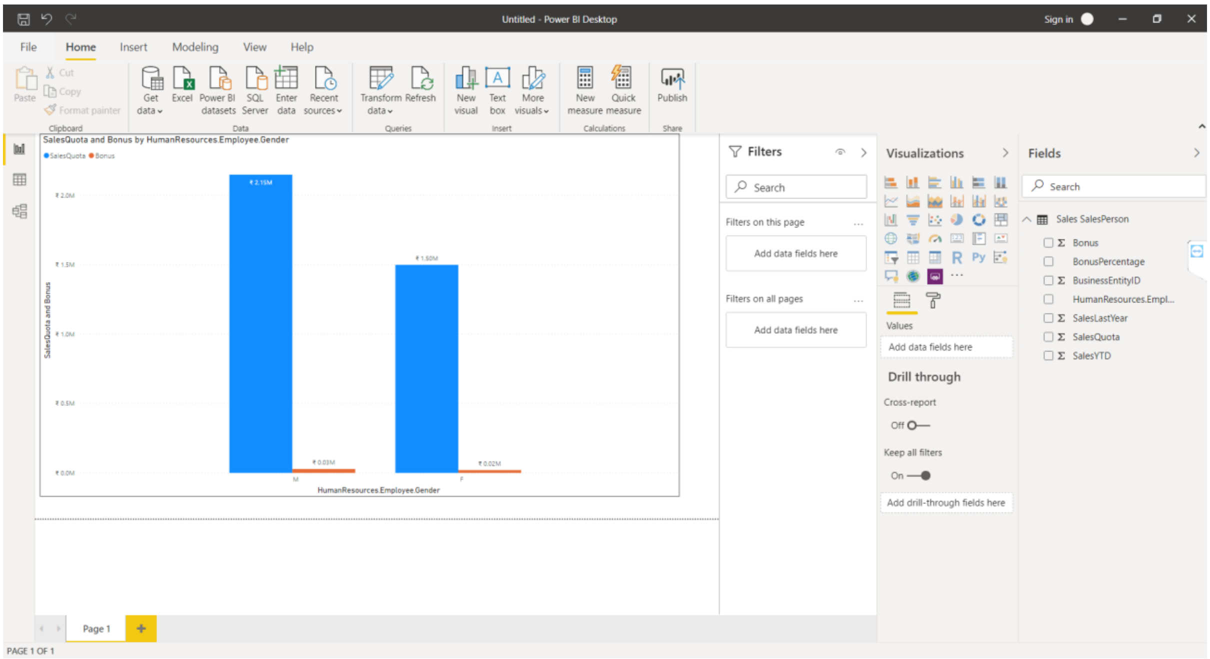


8. Rename & Reorder the columns





9. Show the relationship output (Gender Based Sales & Bonus Relationship)

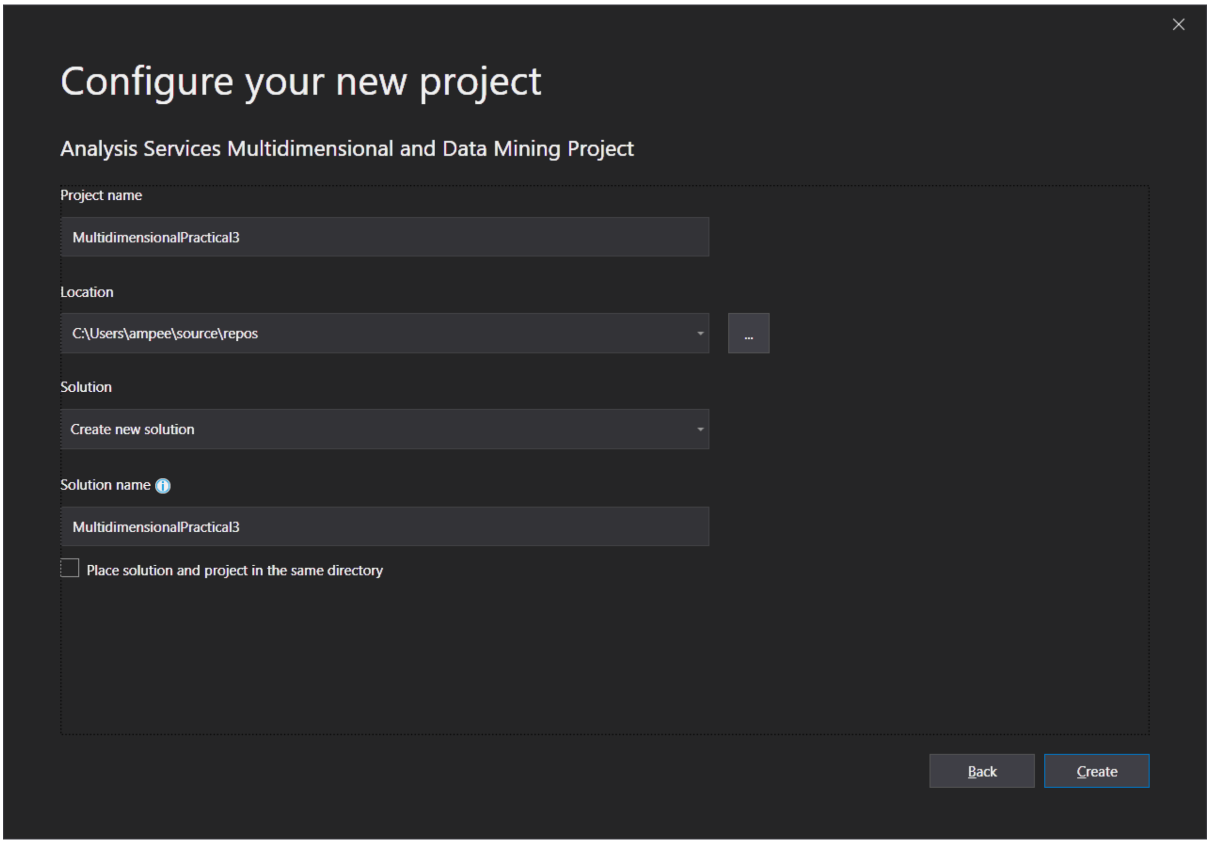


Practical No: - 3

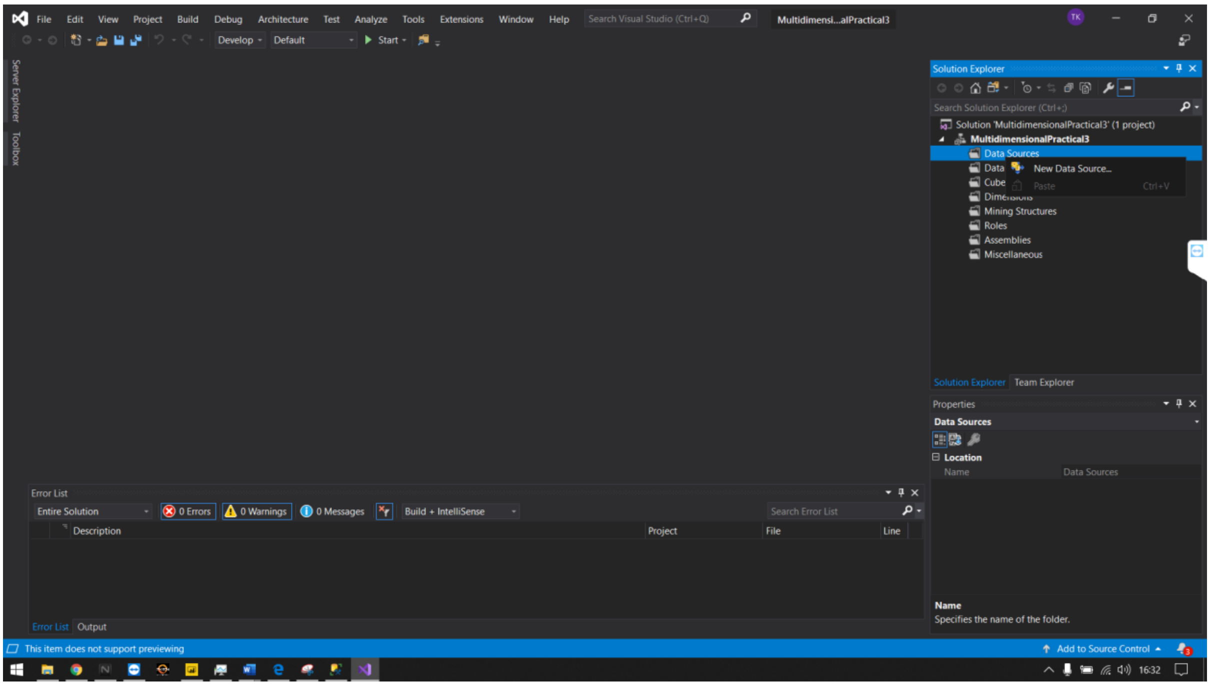
Aim: - Create the Data staging area for the selected database & Create the cube with suitable dimension and fact tables based on ROLAP, MOLAP and HOLAP model.

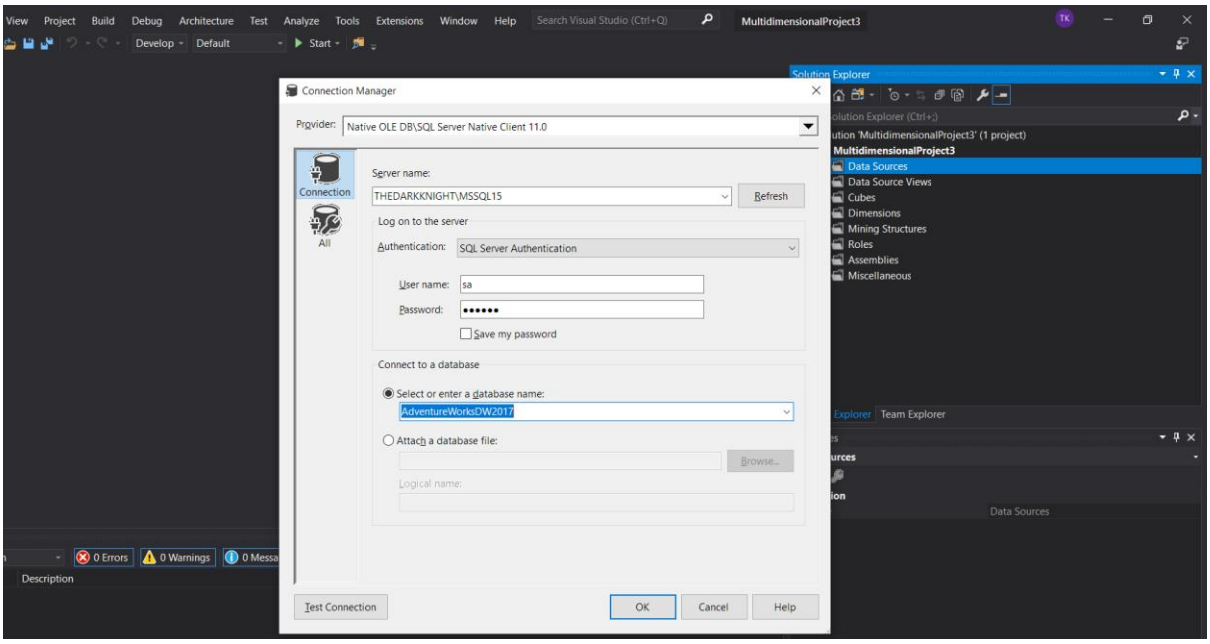
Steps: -

1. Create a new Multidimensional Project.

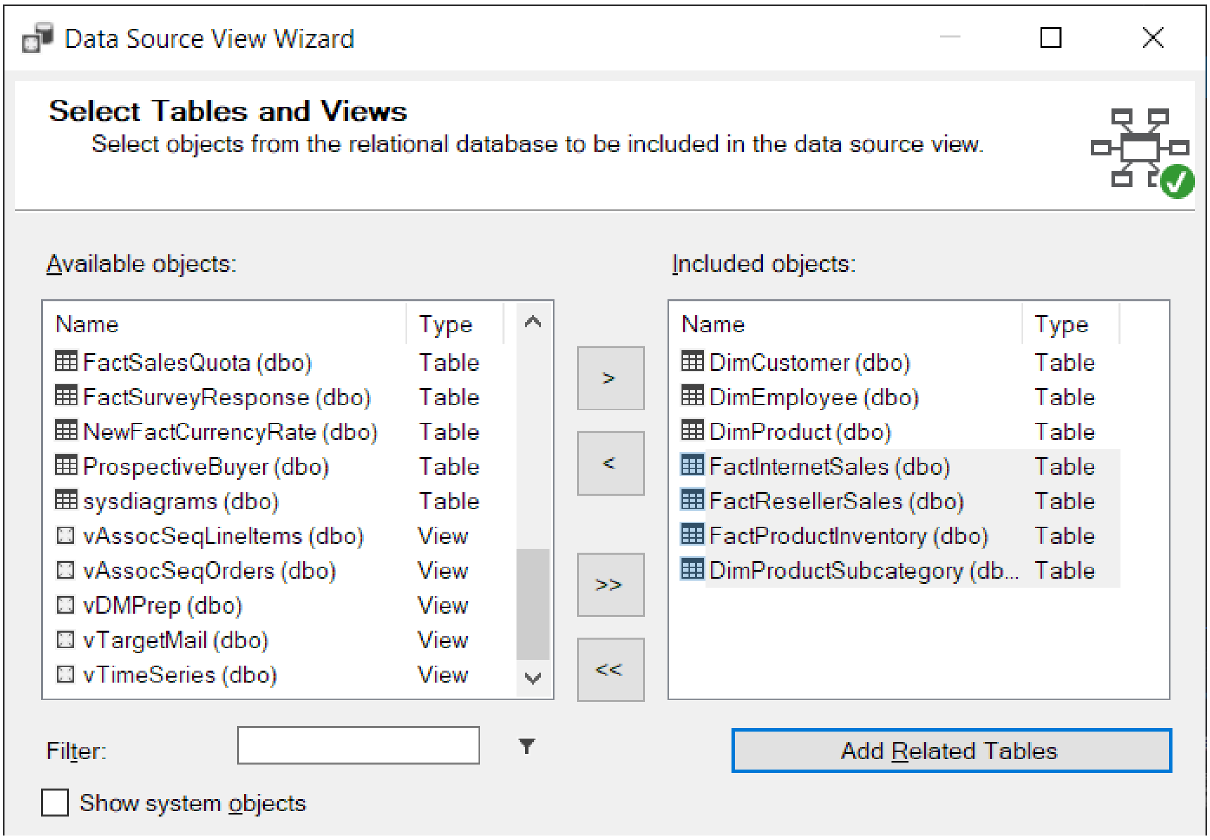


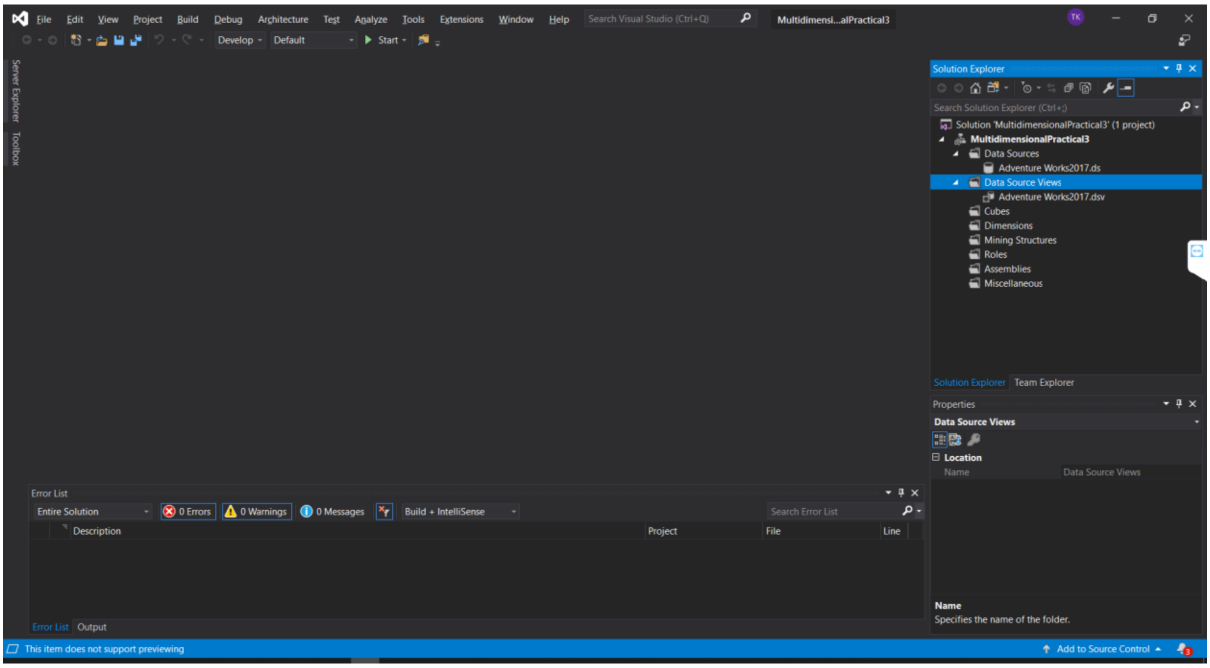
2. Go to DataSource & add a new source.



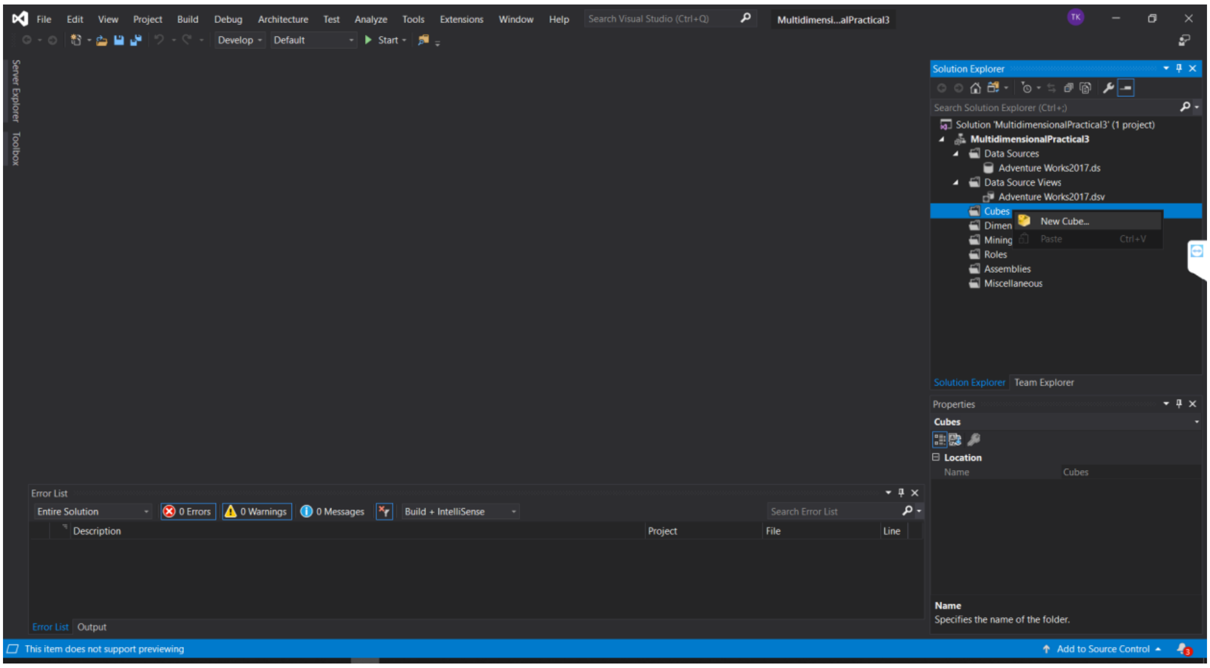


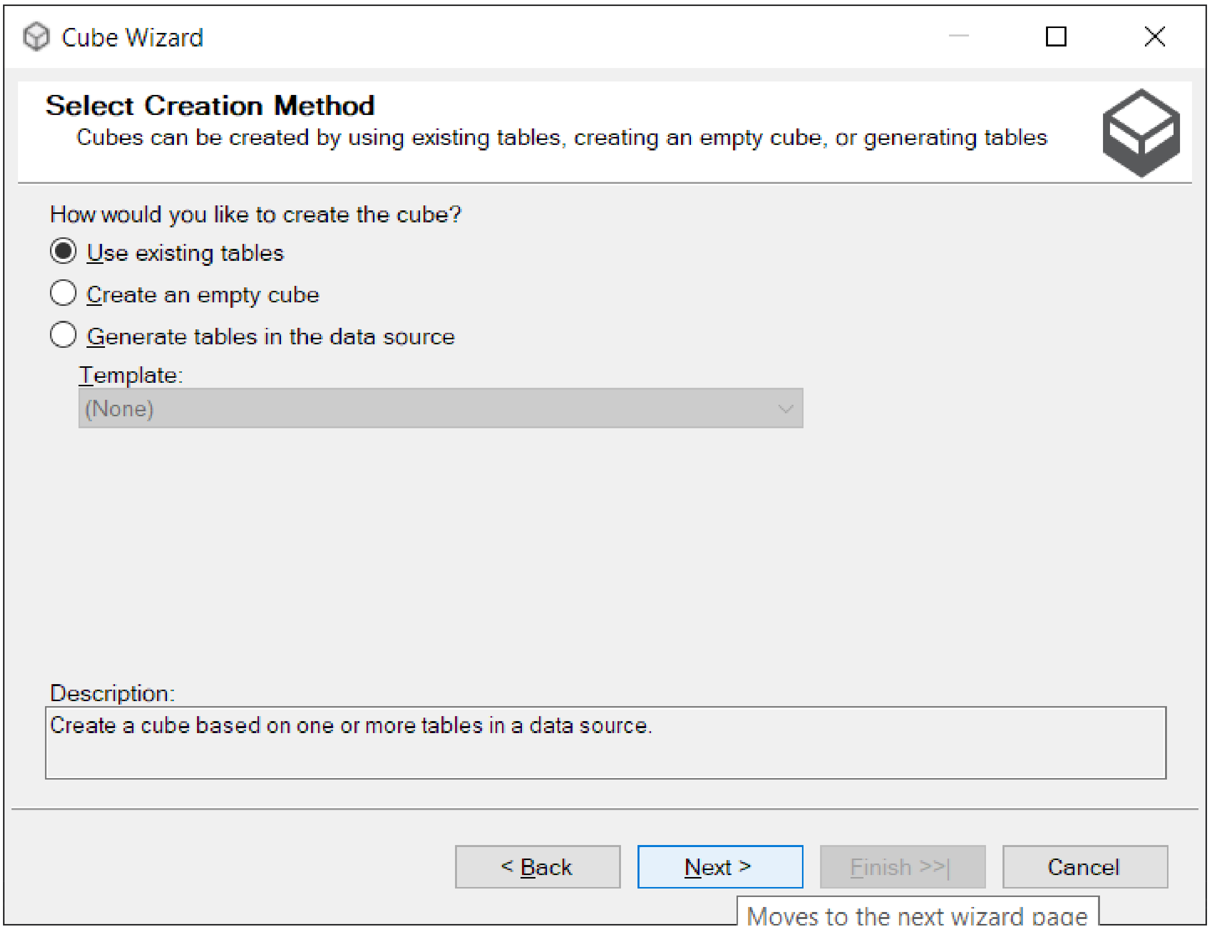
3. Create a view from the added Source.

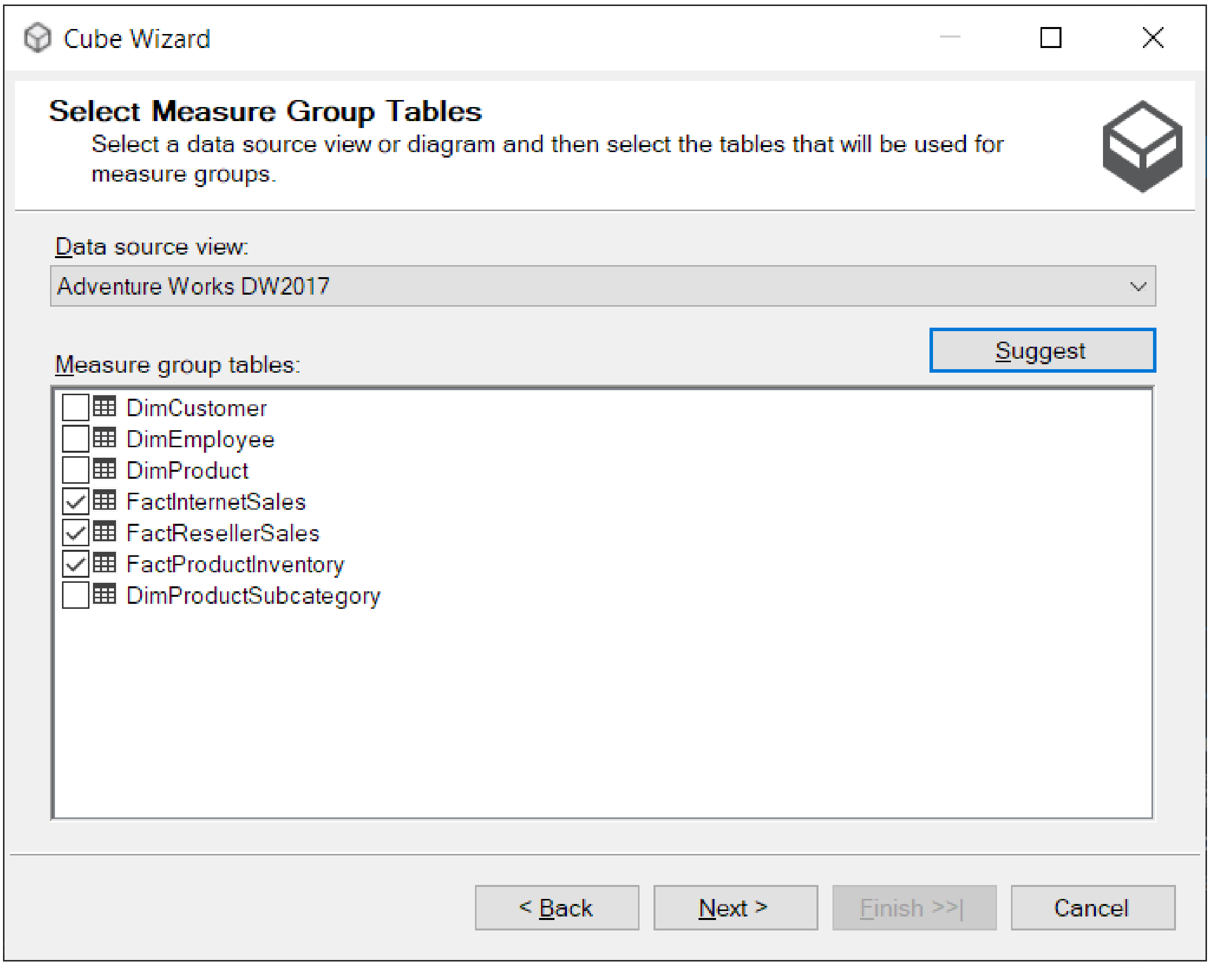




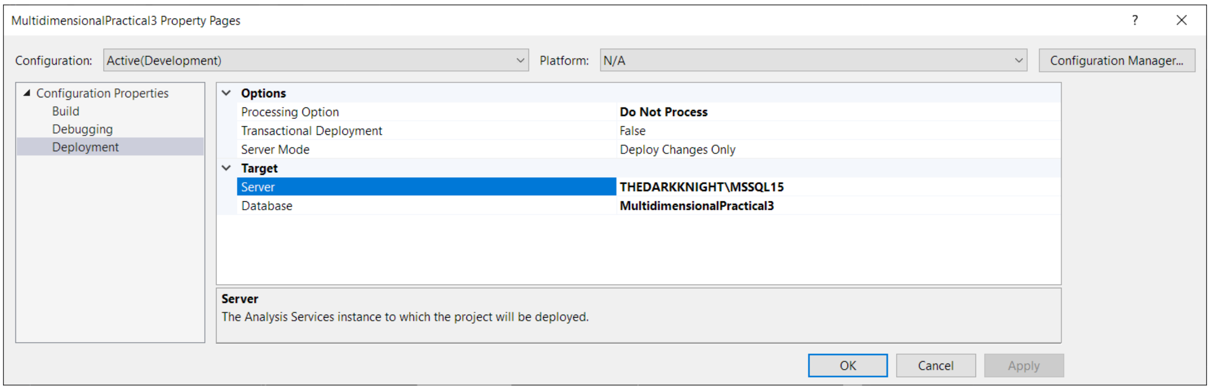
4. Create a new cube to generate data to be displayed.



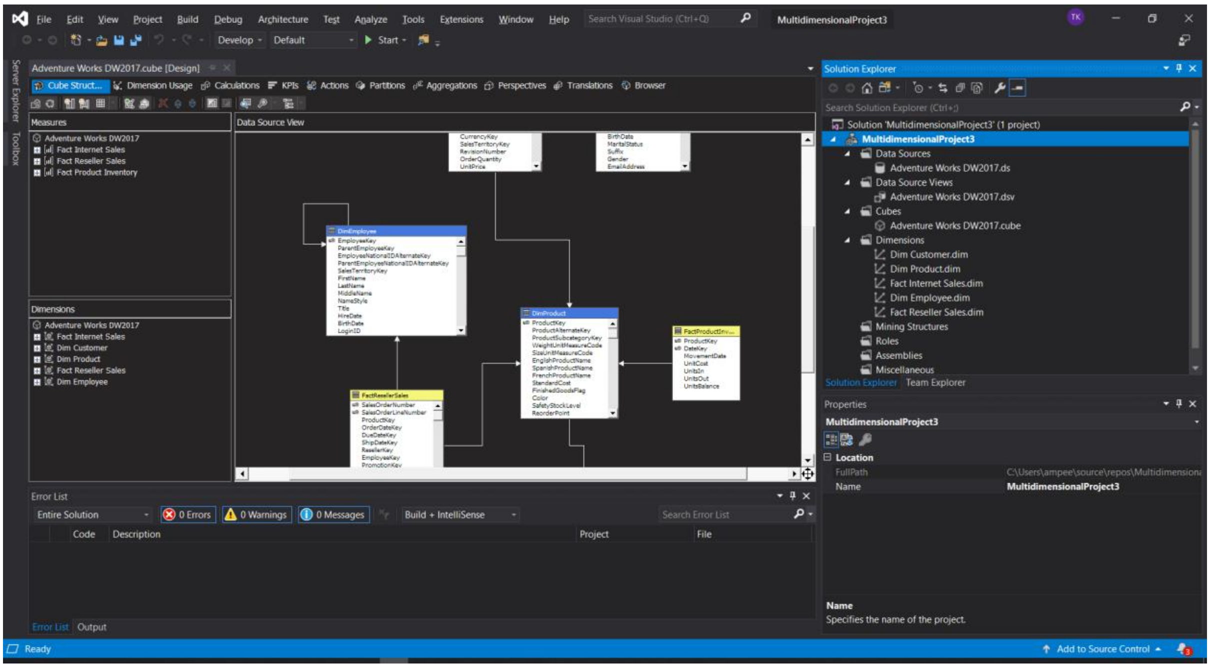


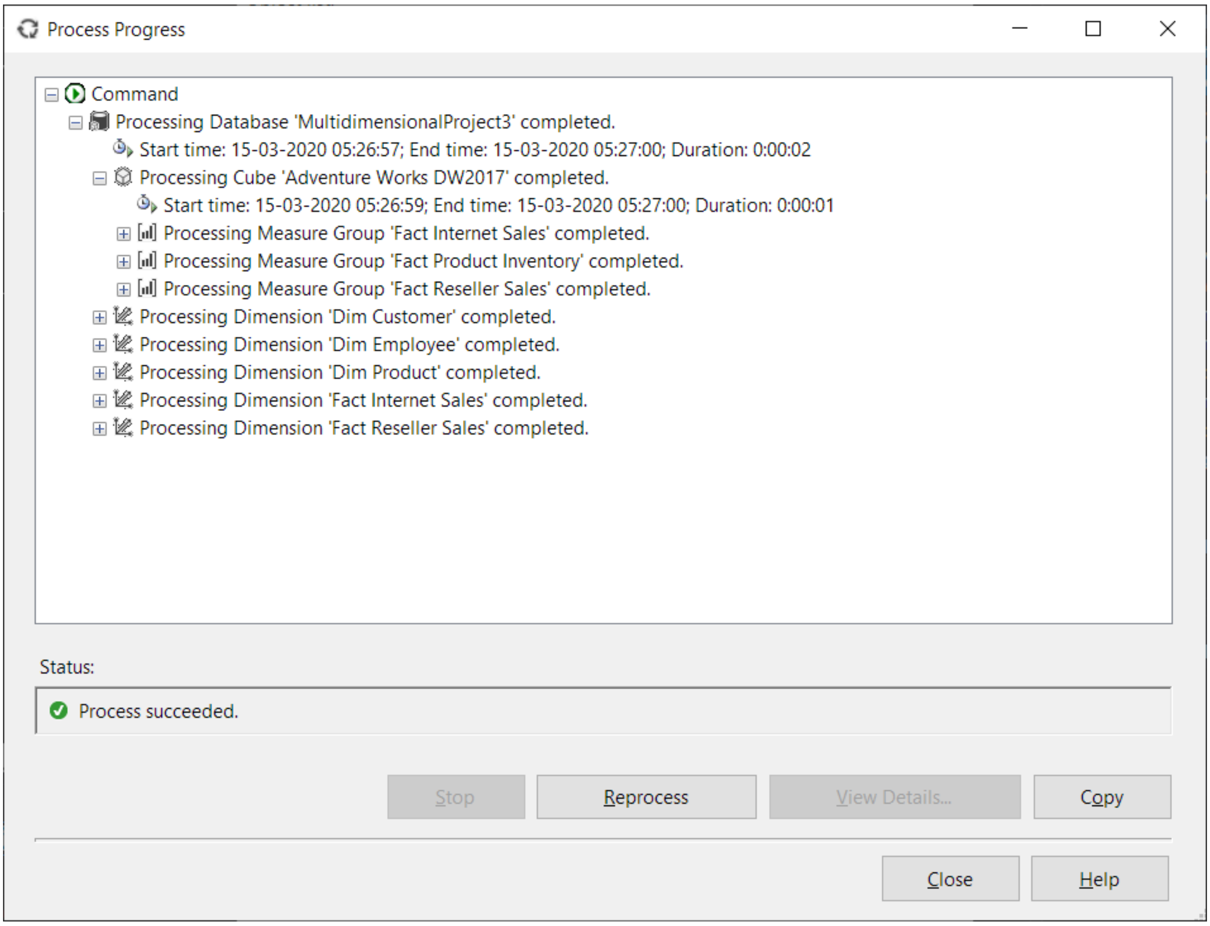


5. Change properties of Multidimensional Project.

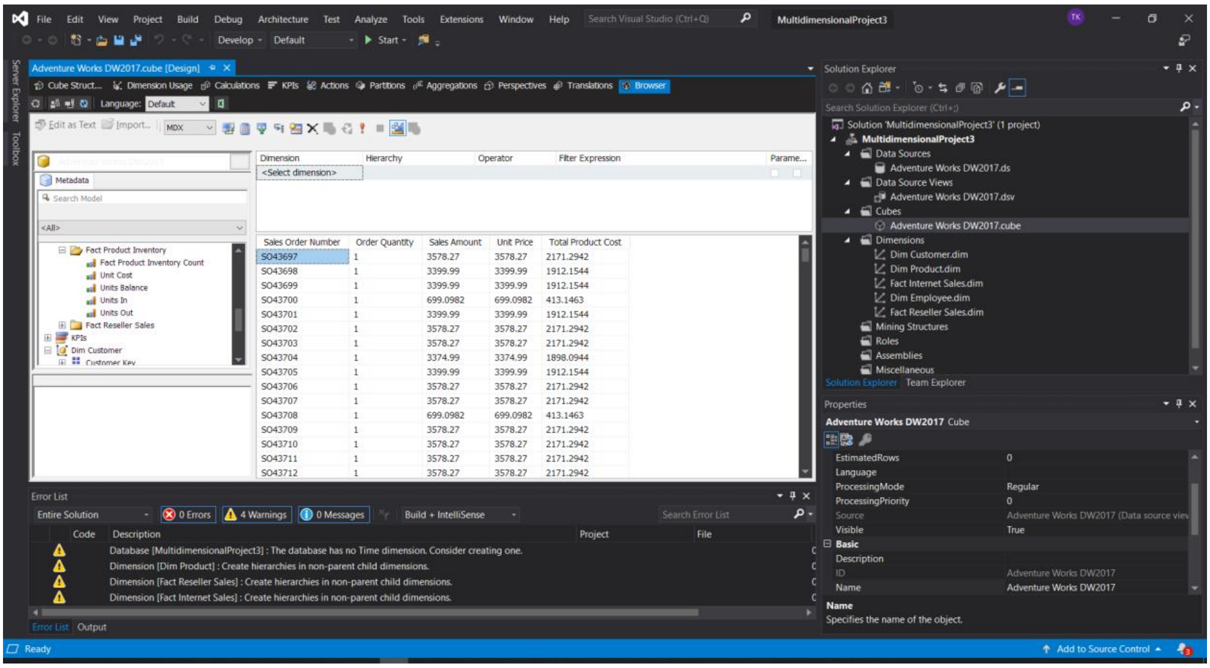


6. Show Deployment Process.





6. Click on browse and fire a query.



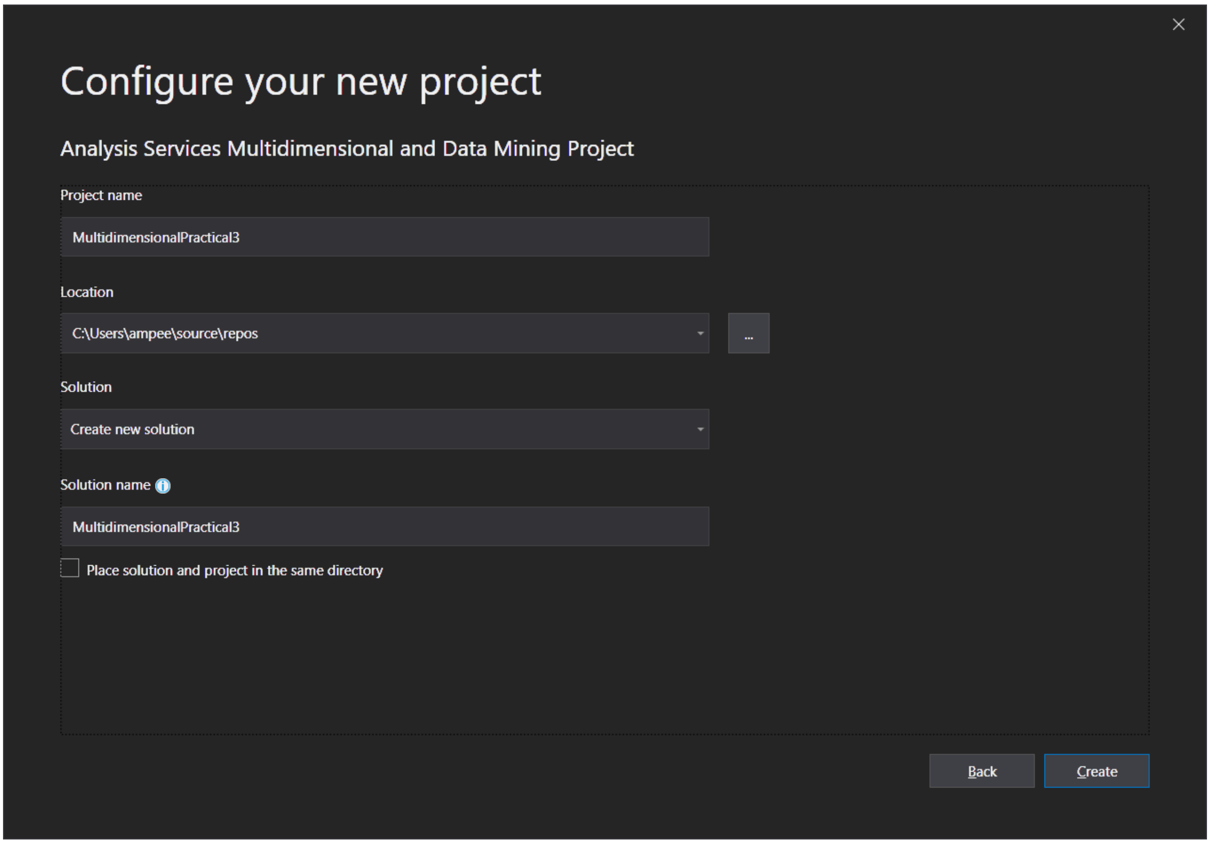
Practical No: - 4

Aim: - a. Create the ETL map and setup for execution. b. Execute the MDX queries to extract the data from the Datawarehouse.

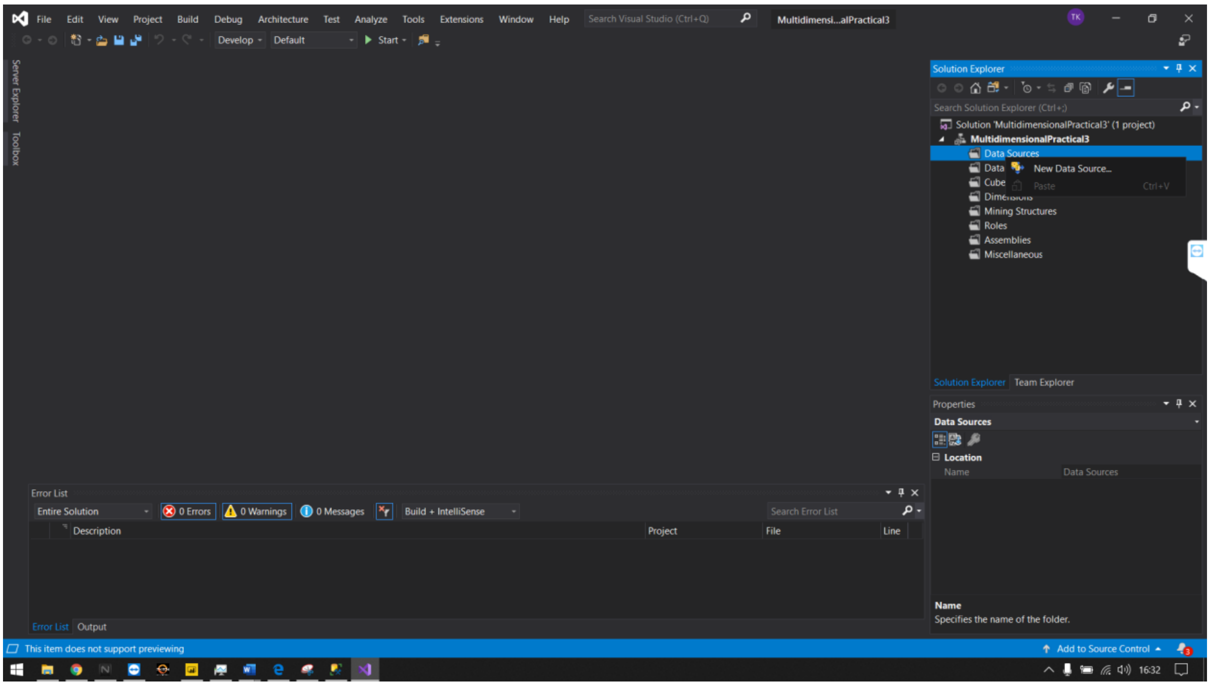
Steps: -

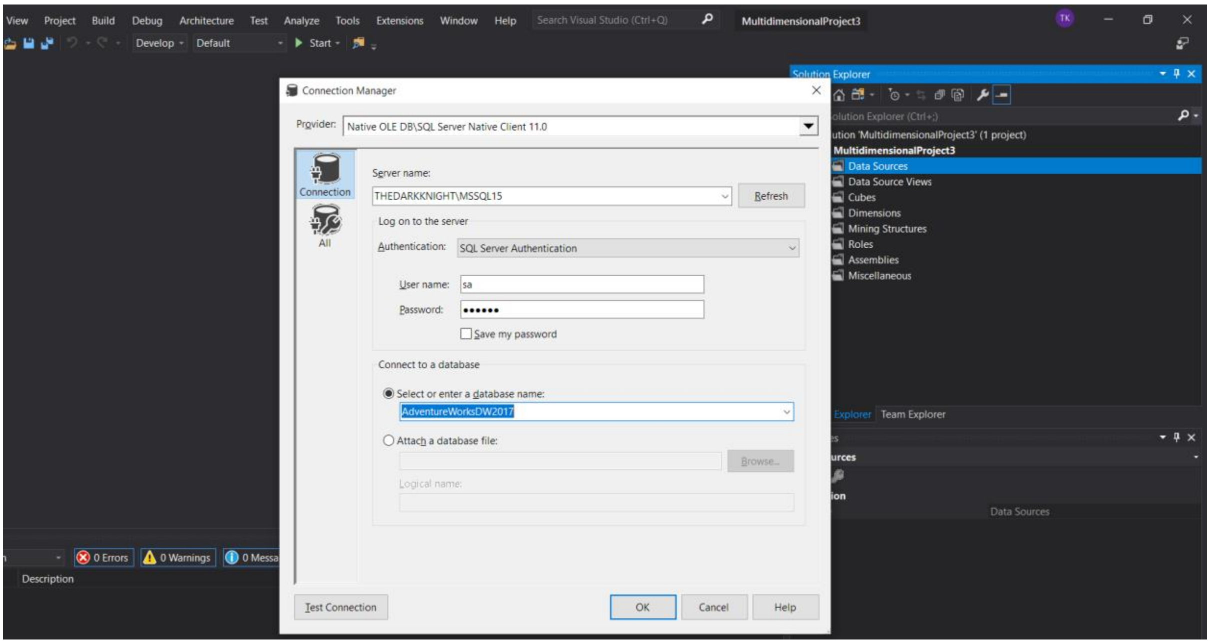
PART A -> Create the ETL map and setup for execution

1. Create a new Multidimensional Project.

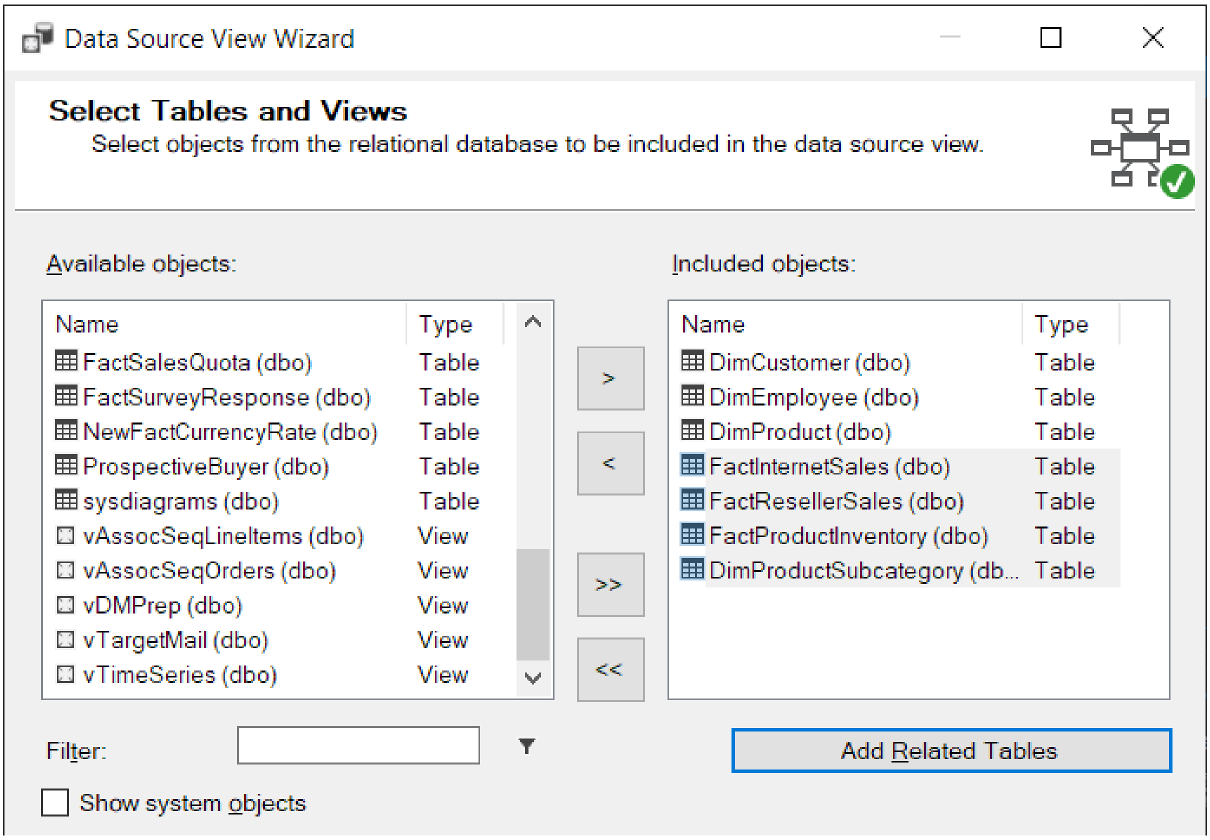


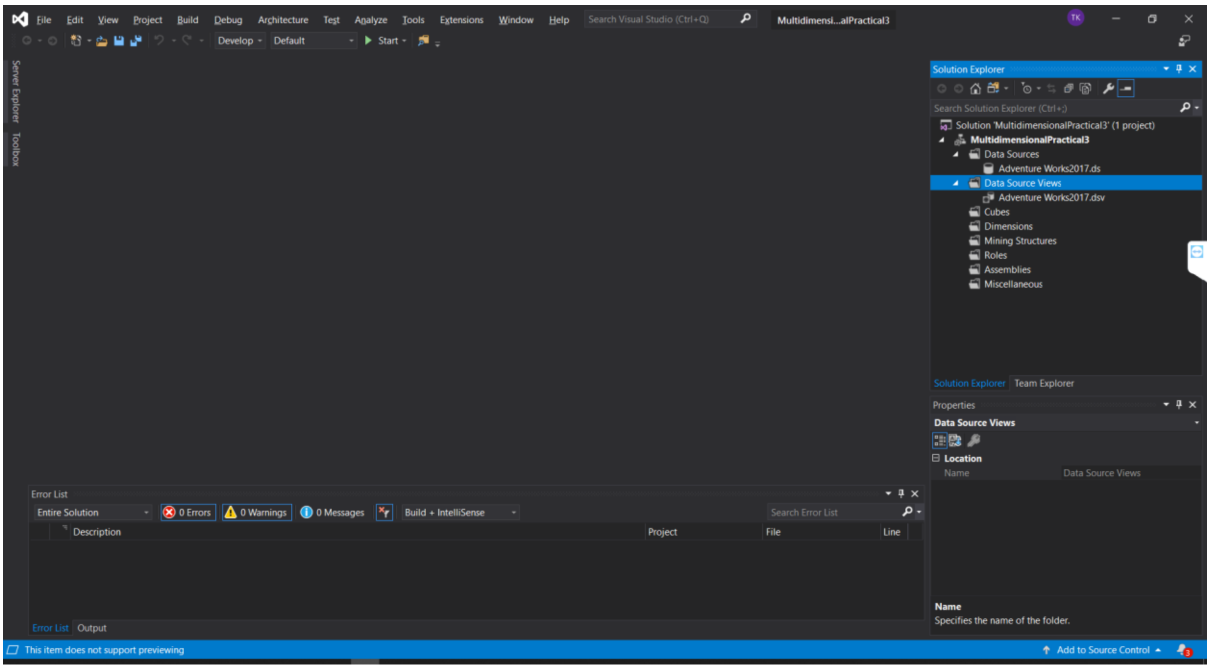
2. Go to DataSource & add a new source.



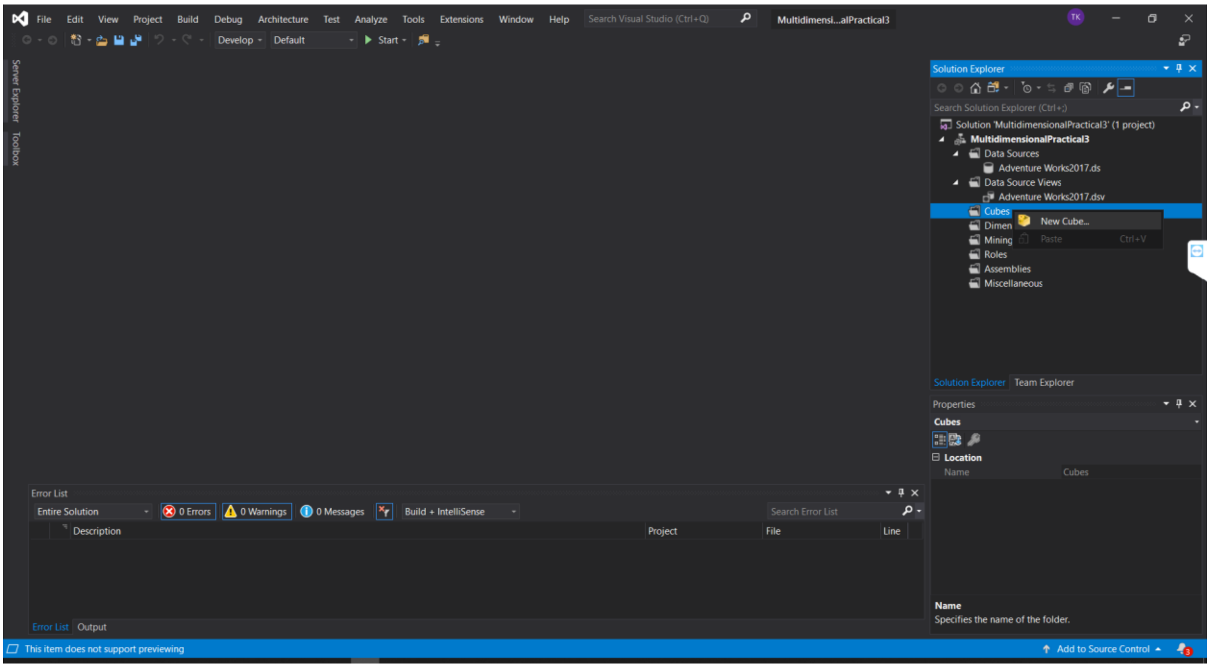


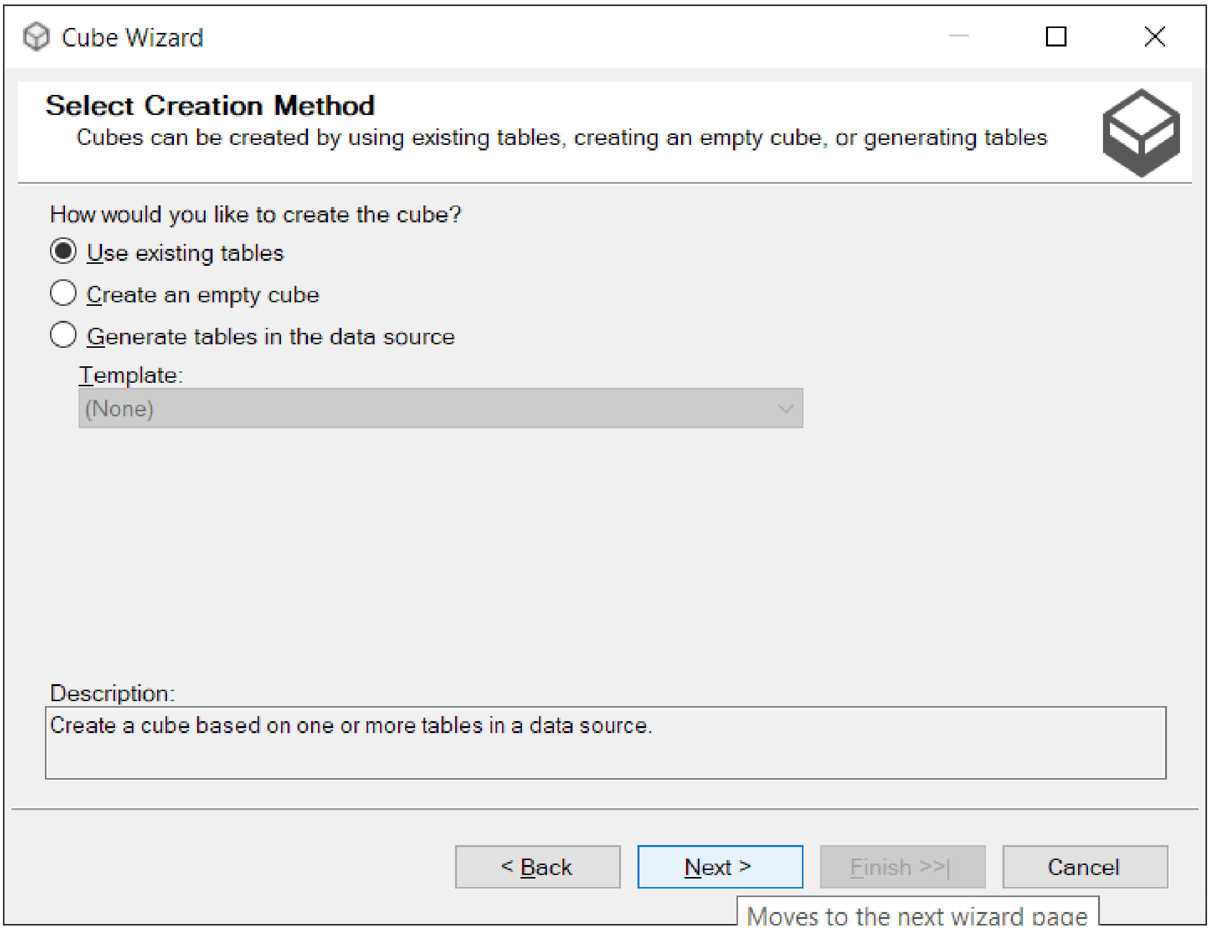
3. Create a view from the added Source.

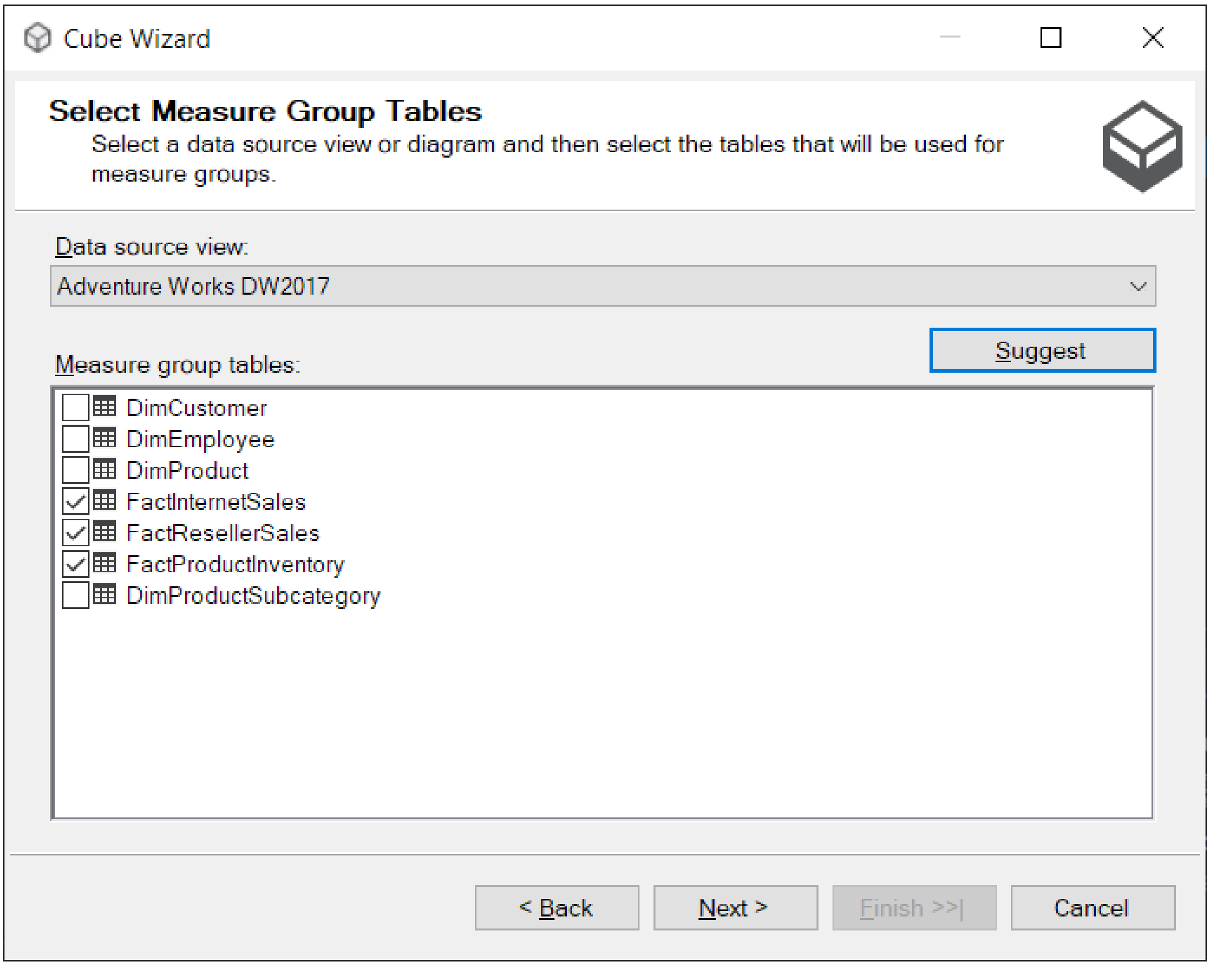




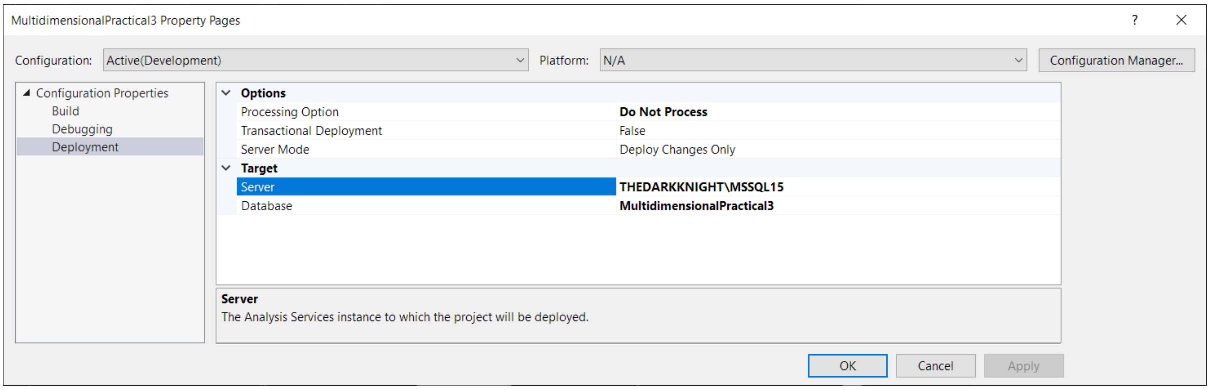
4. Create a new cube to generate data to be displayed.



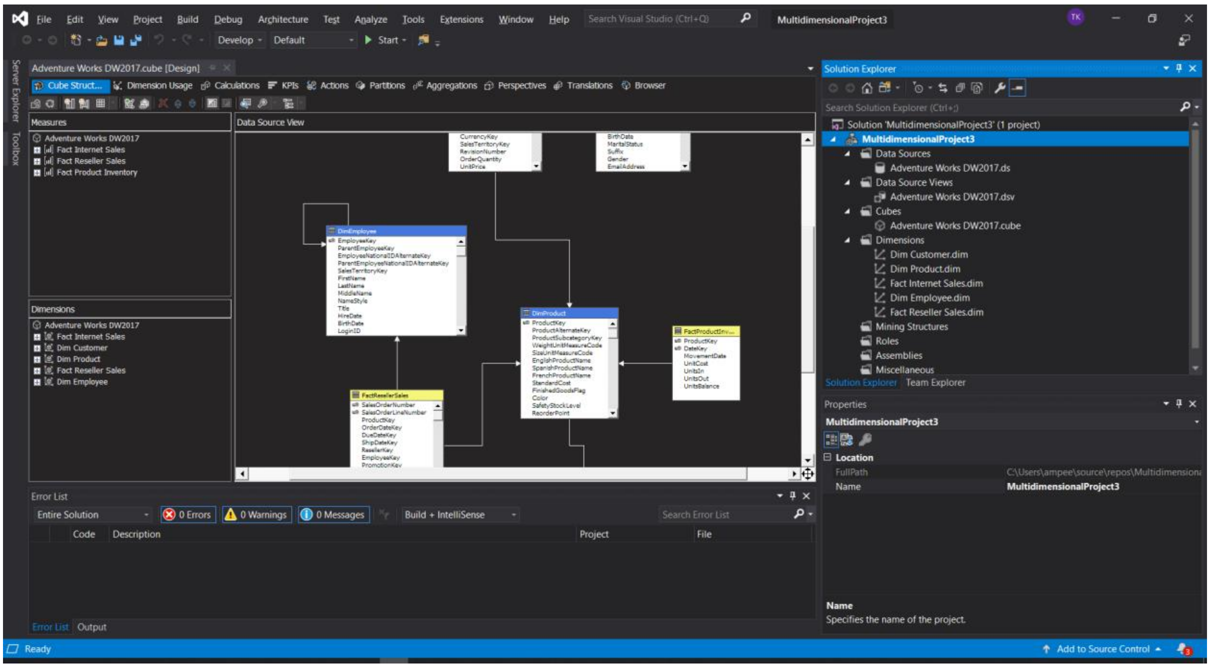


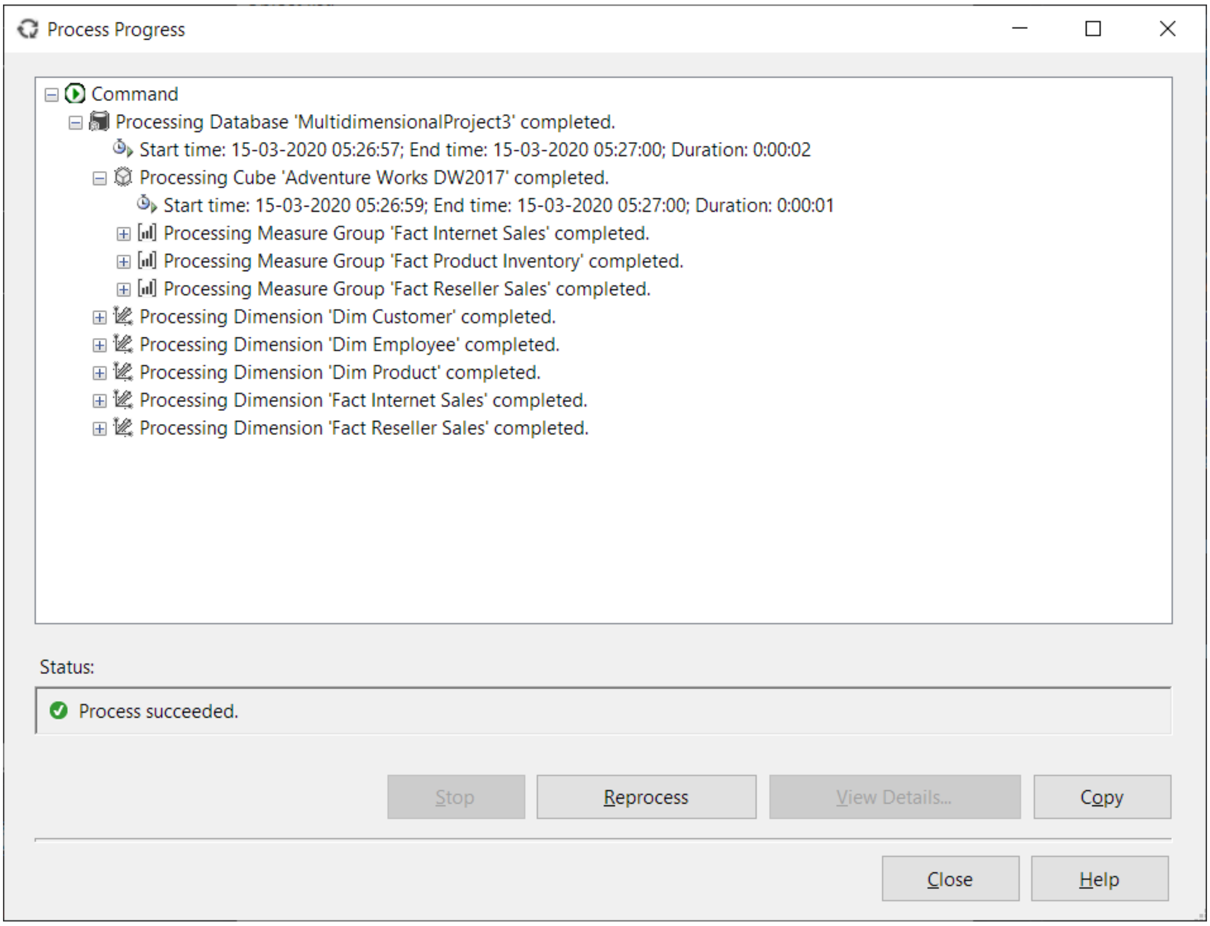


5. Change properties of Multidimensional Project.

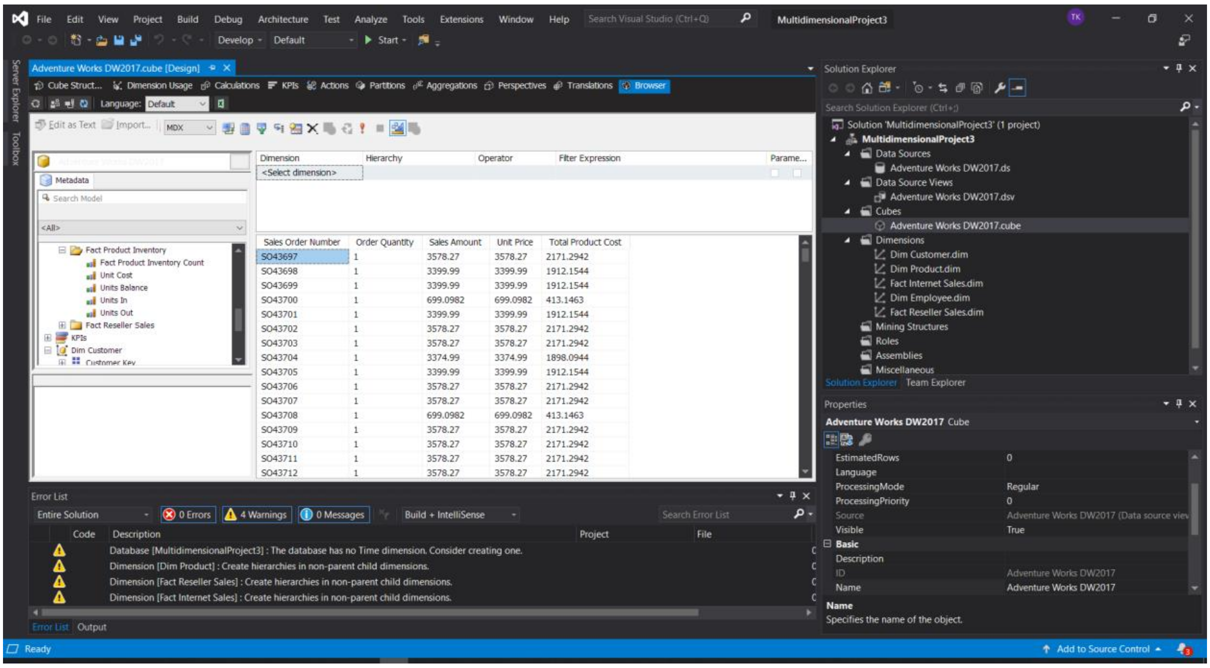


6. Show Deployment Process.



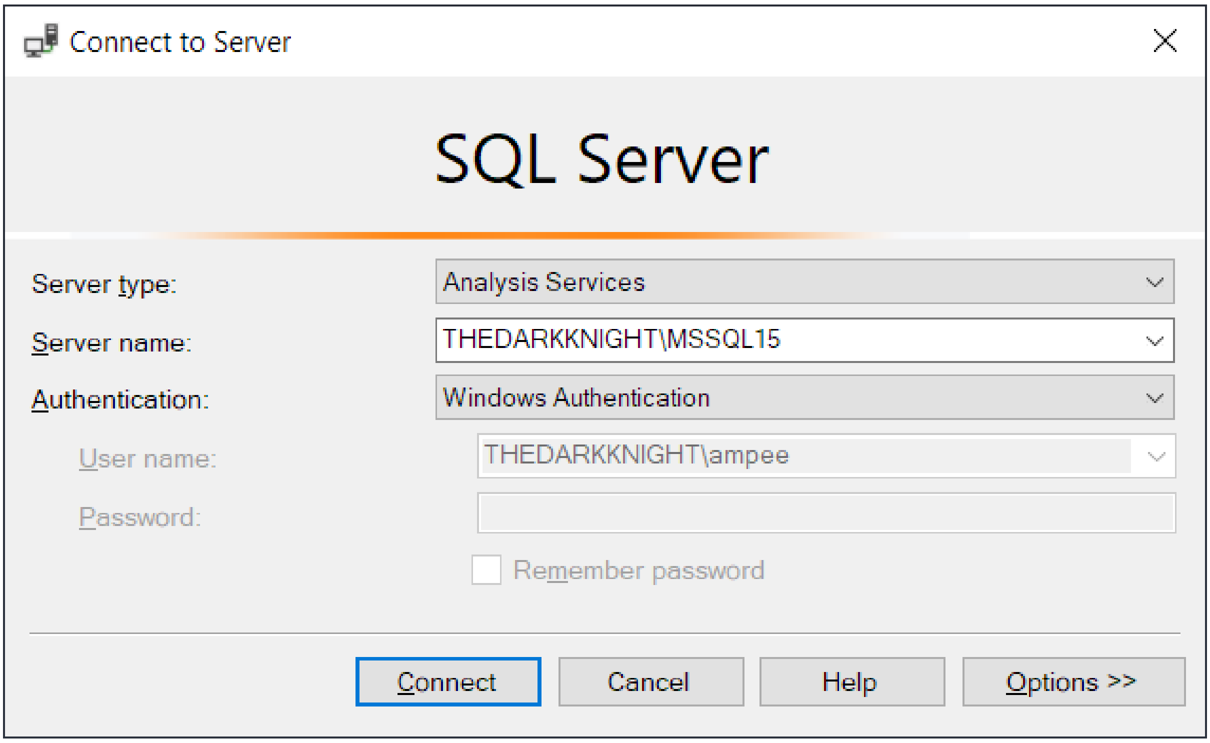


6. Click on browse and fire a query.

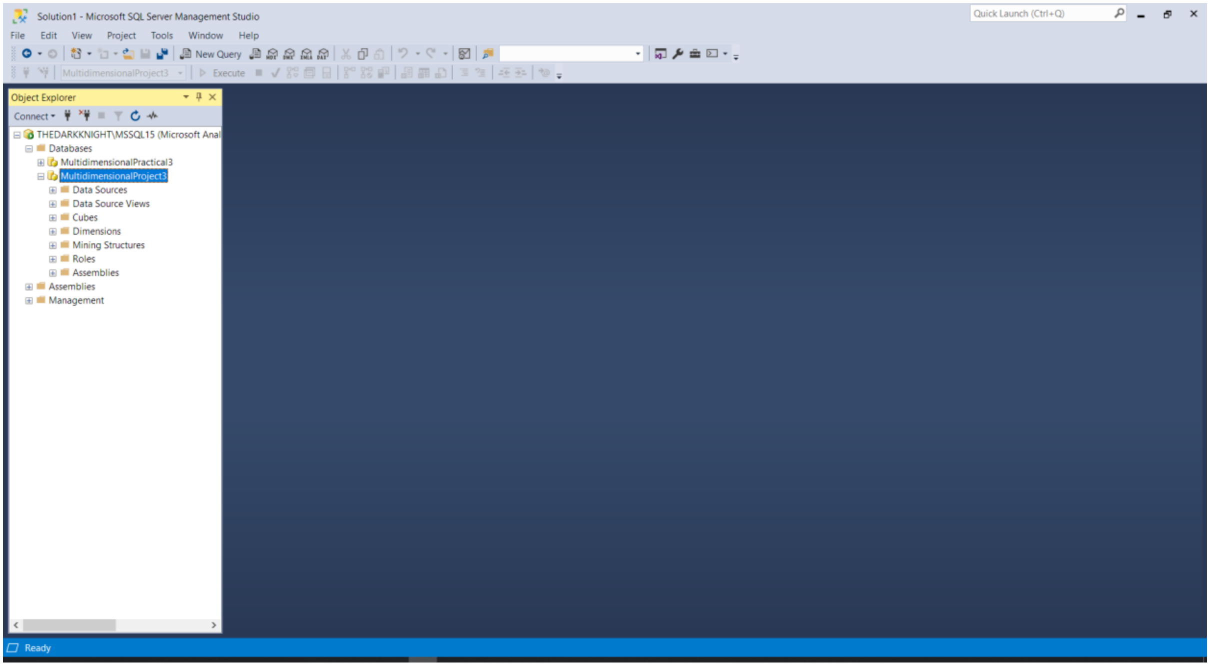


PART B -> Execute MDX Query to extract data from Warehouse.

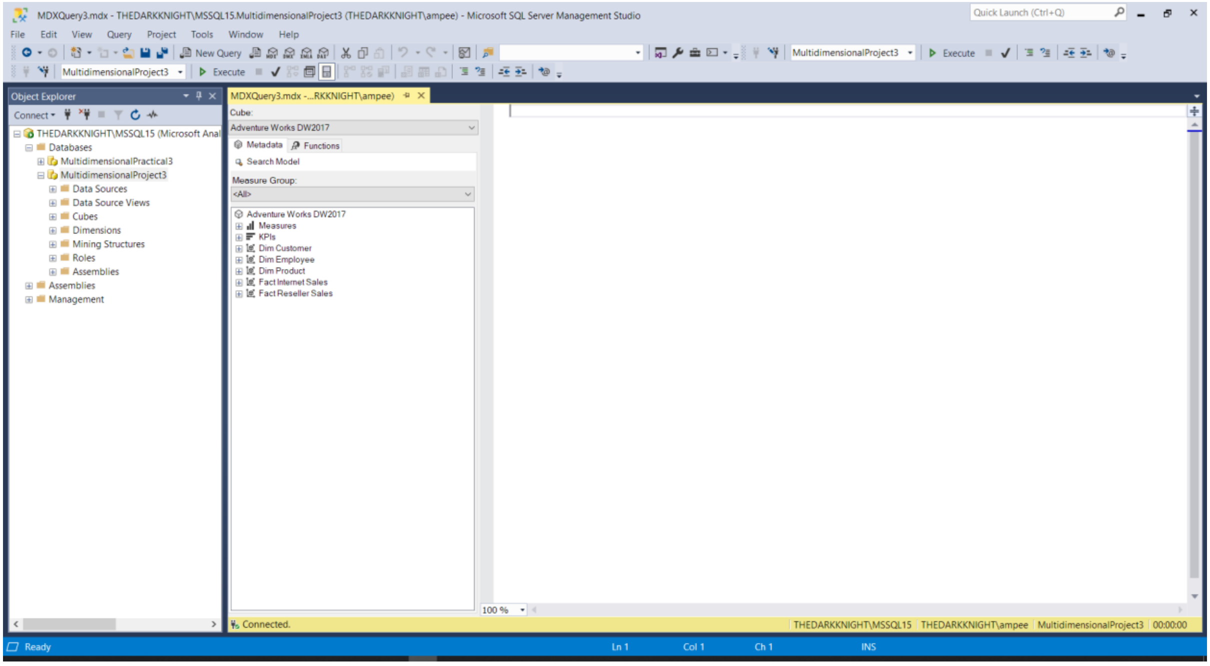
1. Open SQL Server Management Studio -> Select Analysis Server Engine -> Connect.

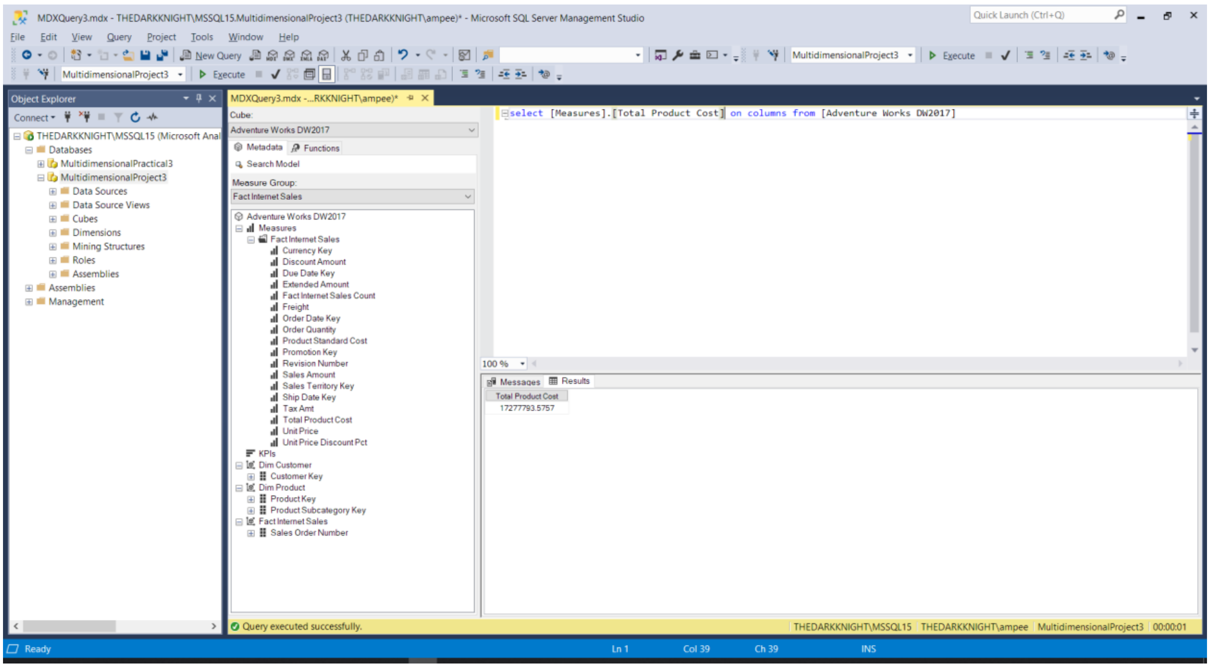


2. Open the Multi Dimension Project from earlier and load it.



3. Execute MDX Query to fetch result from Data Warehouse. (Select total production cost).





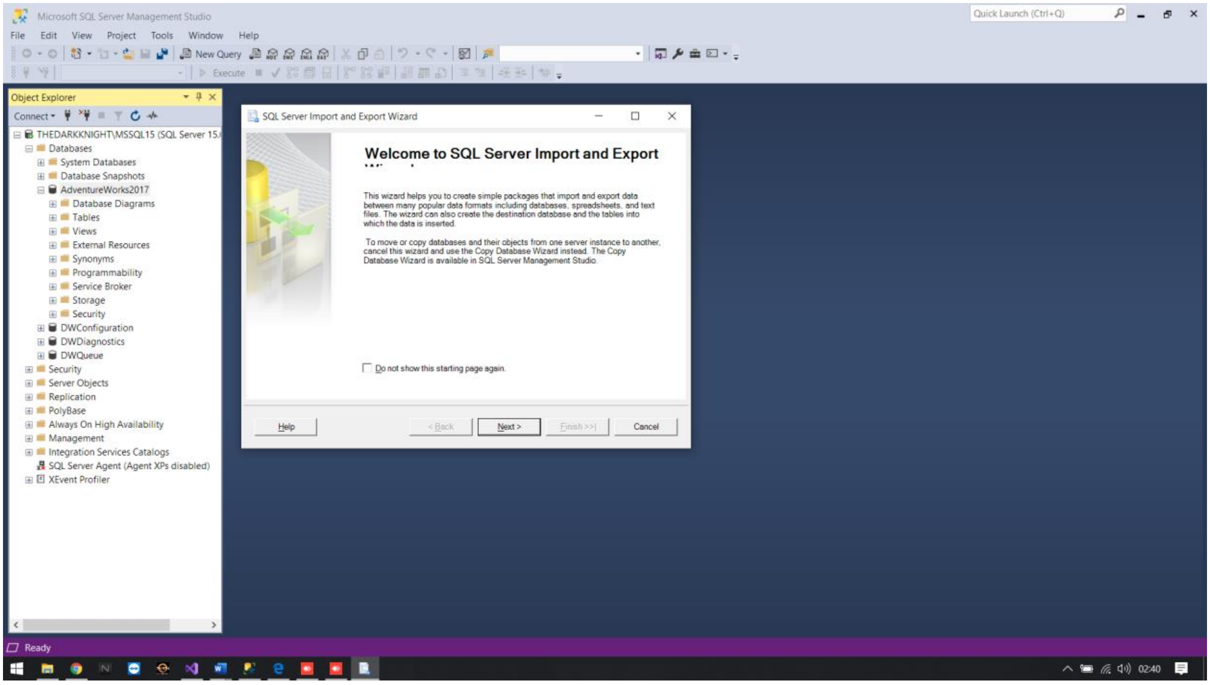
Practical No: - 5

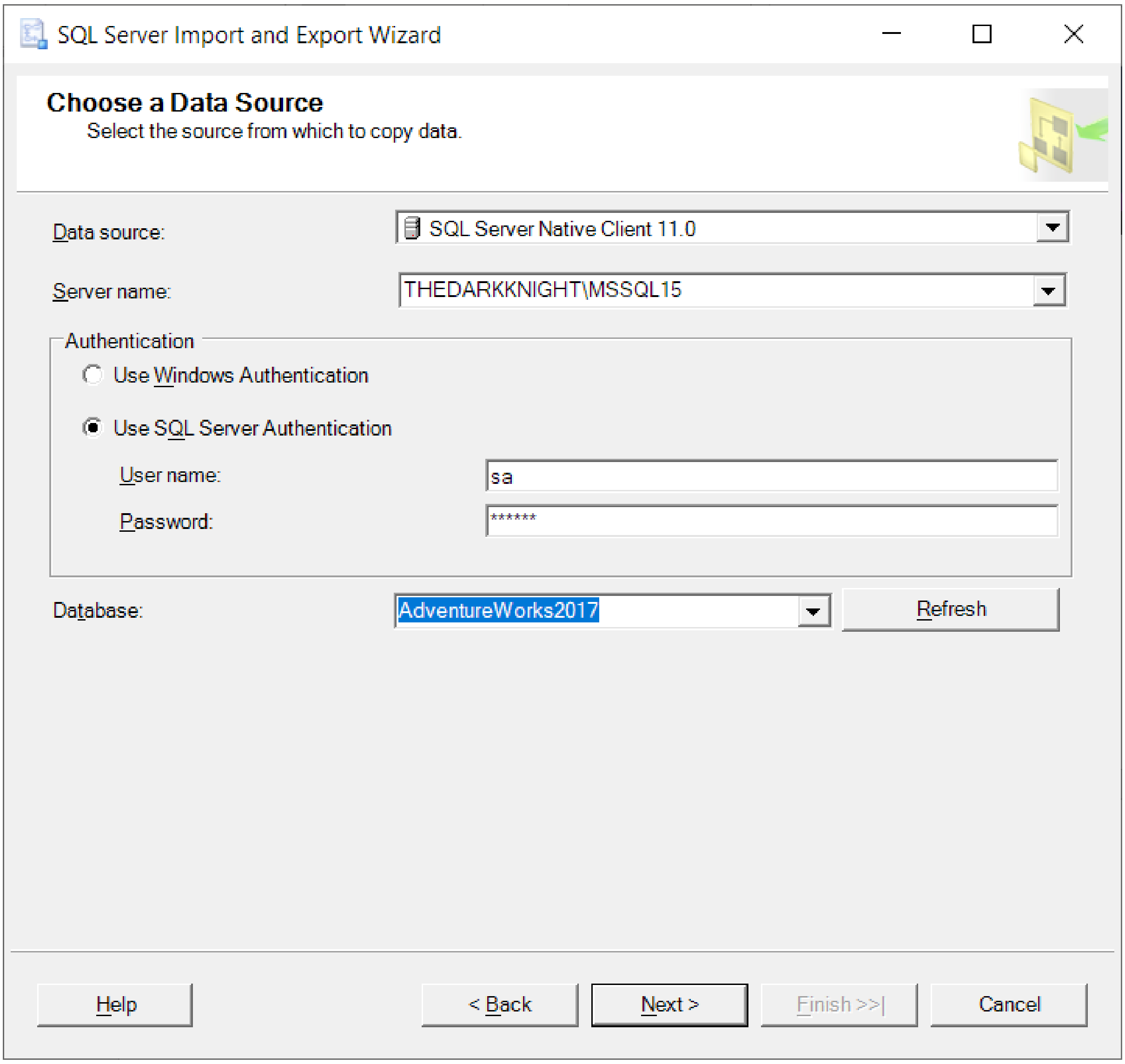
Aim: - a. Import the Datawarehouse data in Microsoft Excel and create the Pivot table and Pivot Chart. b. Import the cube in Microsoft Excel and create the Pivot table and Pivot Chart to perform data analysis.

PART A -> Import the Datawarehouse data in Microsoft Excel and create the Pivot table and Pivot Chart

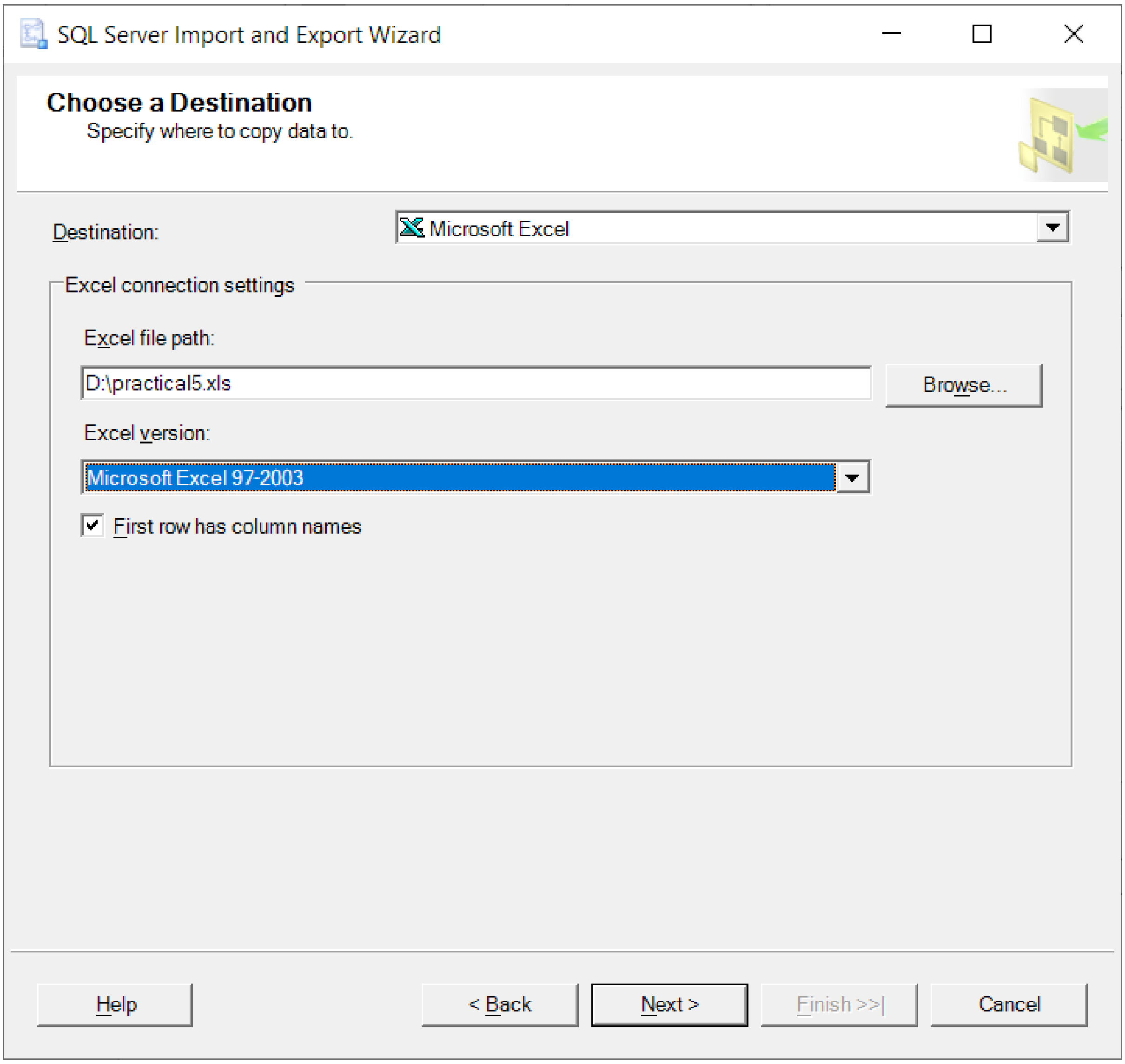
Steps: -

1. Export dB from Microsoft SQL Server Management Studio.

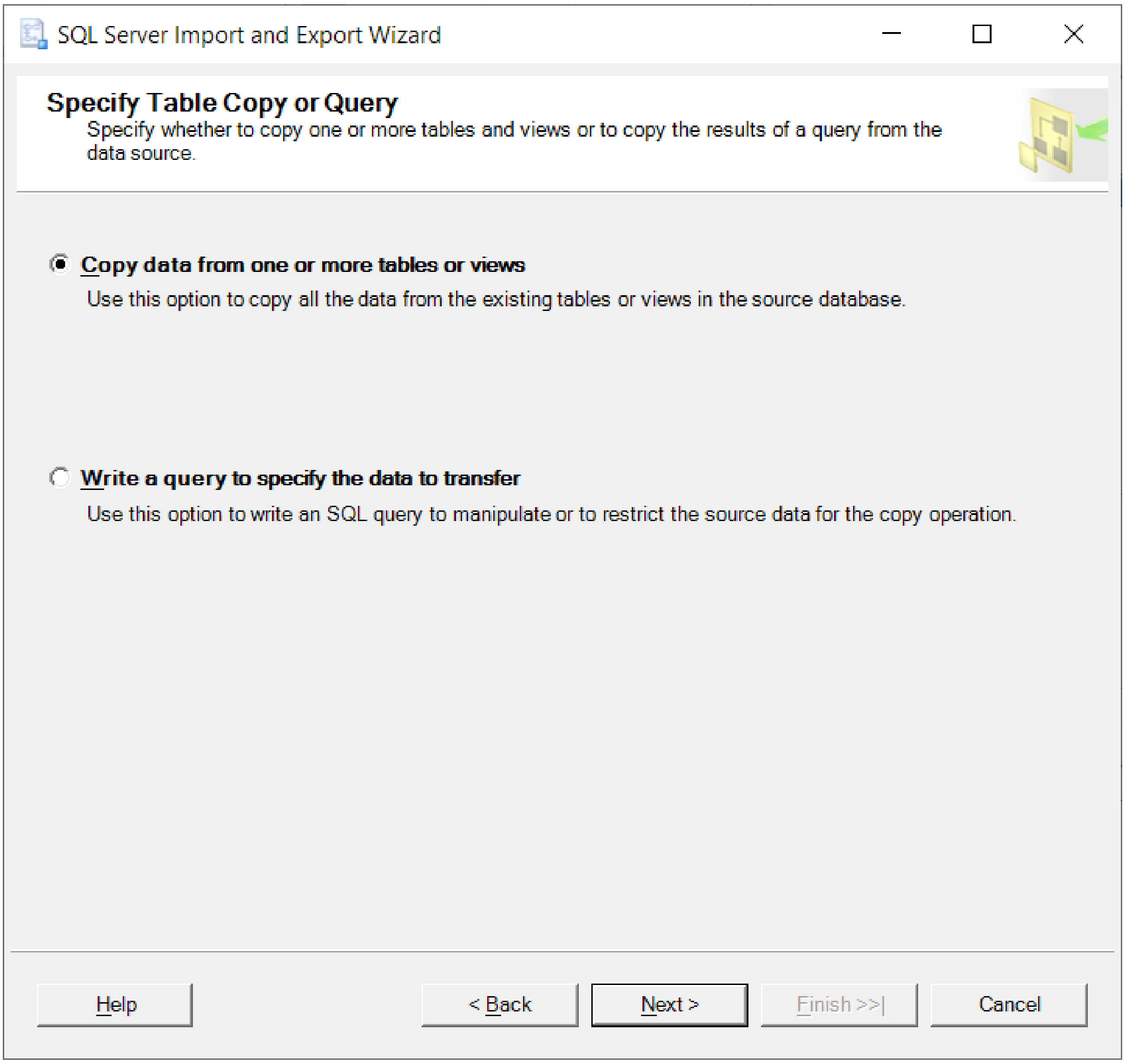




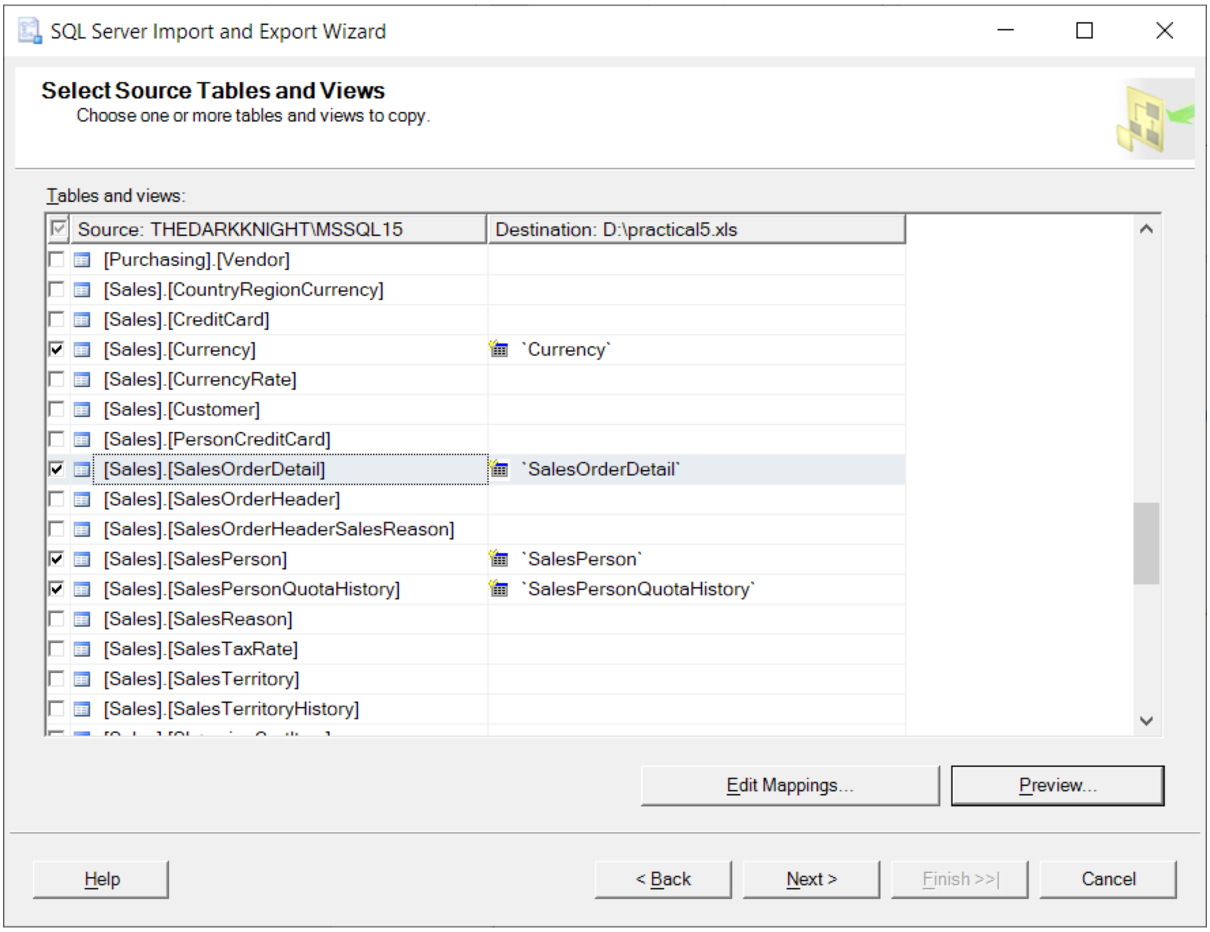
SQL Server Import and Export Wizard
Choose a Data Source
Select the source from which to copy data.
Data source:
SQL Server Native Client 11.0
Server name:
THEDARKKNIGHT\MSSQL15
Authentication
Use Windows Authentication
Use SQL Server Authentication
User name:
sa
Password:
******
Database:
AdventureWorks2017
Refresh
Help
< Back
Next >
Finish >>|
Cancel



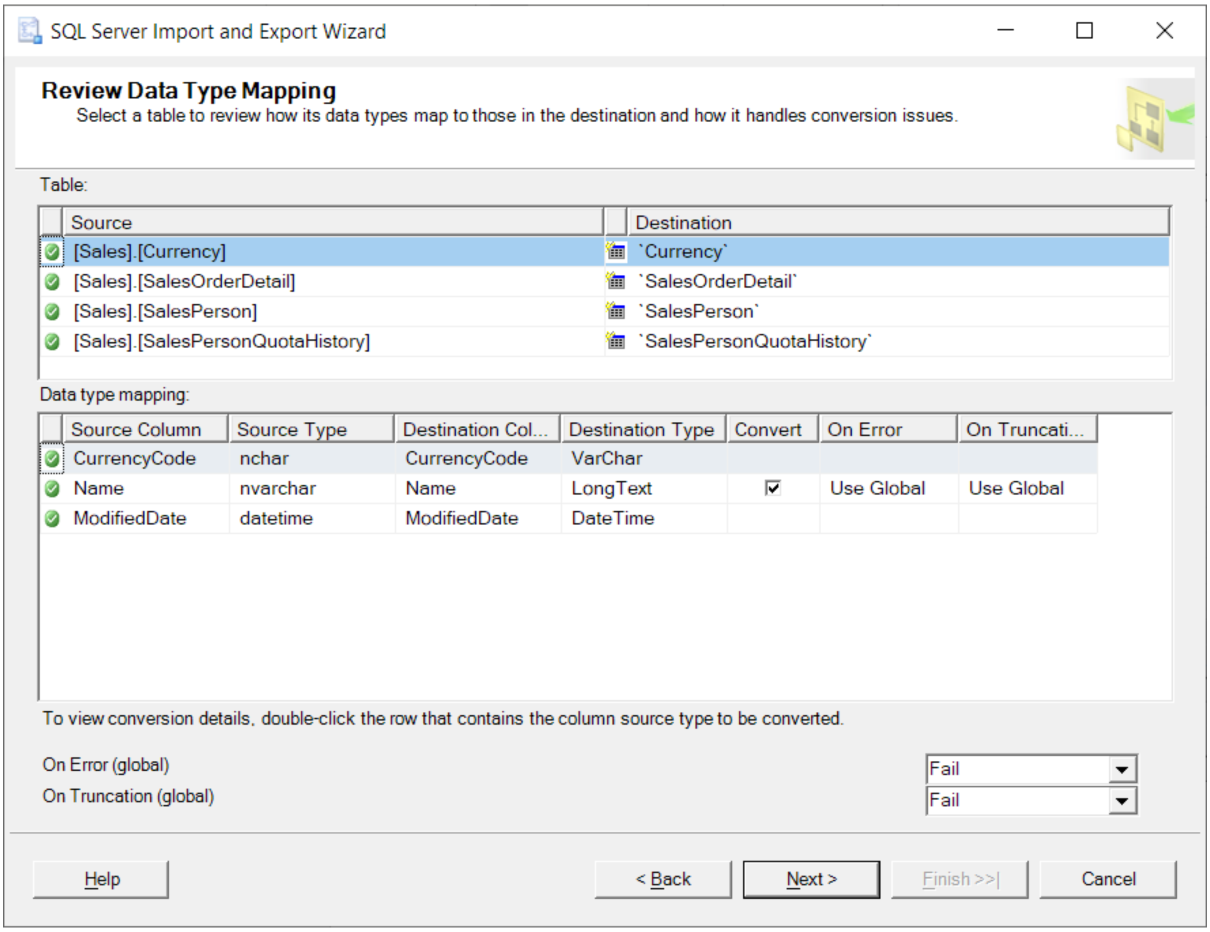
SQL Server Import and Export Wizard
Choose a Destination
Specify where to copy data to.
Destination:
Microsoft Excel
Excel connection settings
Excel file path:
D:\practical5.xls
Browse...
Excel version:
Microsoft Excel 97-2003
First row has column names
Help
< Back
Next >
Finish >>|
Cancel

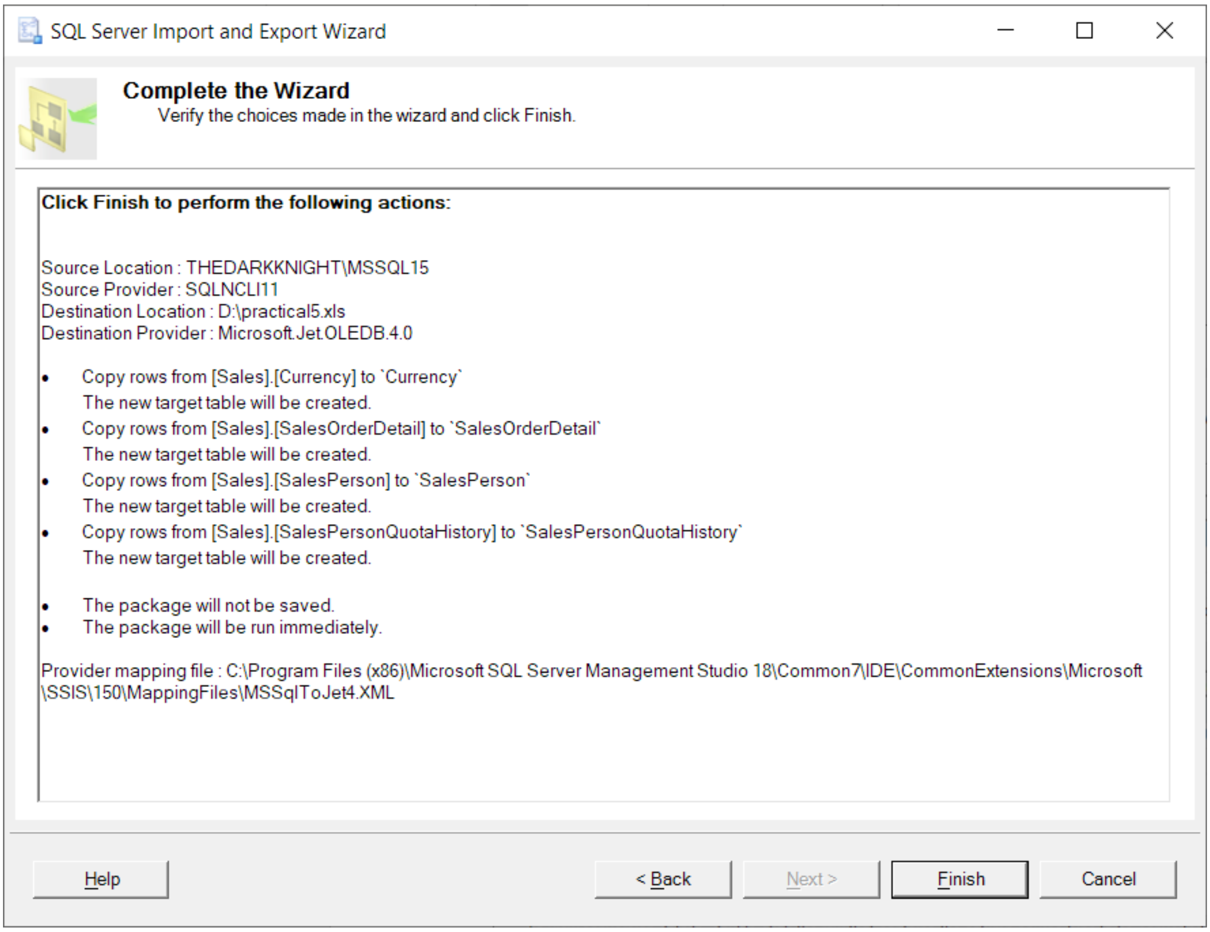


SQL Server Import and Export Wizard
Specify Table Copy or Query
Specify whether to copy one or more tables and views or to copy the results of a query from the data source.
Copy data from one or more tables or views
Use this option to copy all the data from the existing tables or views in the source database.
Write a query to specify the data to transfer
Use this option to write an SQL query to manipulate or to restrict the source data for the copy operation.
Help
< Back
Next >
Finish >>|
Cancel

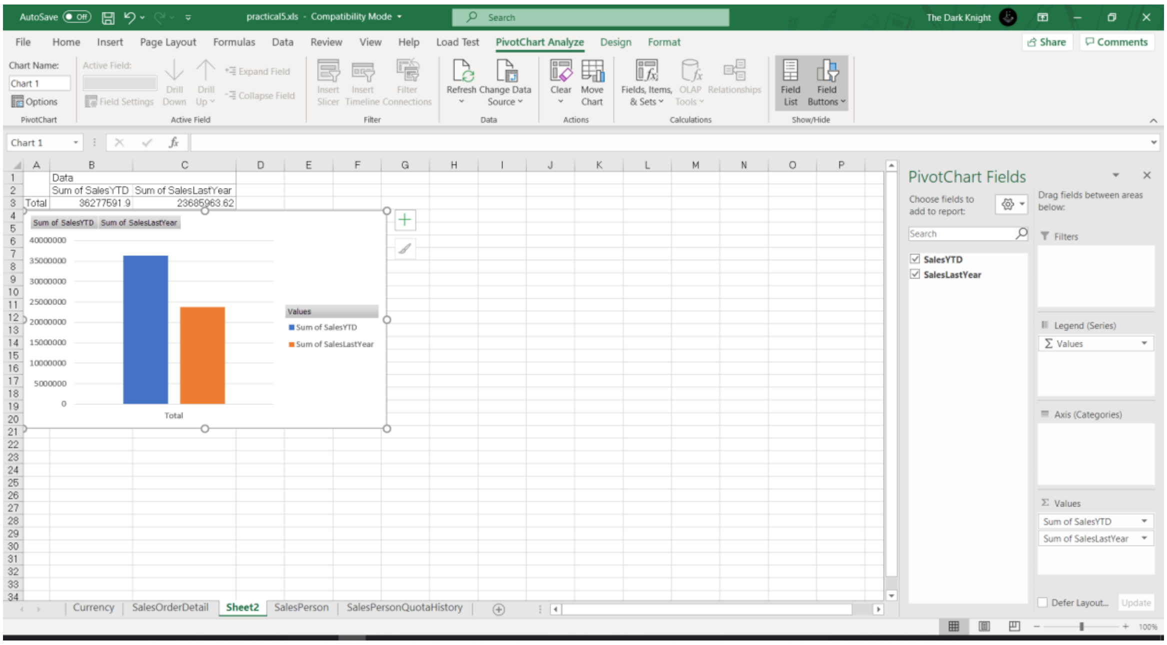


SQL Server Import and Export Wizard
Select Source Tables and Views
Choose one or more tables and views to copy.
Tables and views:
Source: THEDARKKNIGHT\MSSQL15
Destination: D:\practical5.xls
[Purchasing].[Vendor]
[Sales].[CountryRegionCurrency]
[Sales].[CreditCard]
[Sales].[Currency]
`Currency`
[Sales].[CurrencyRate]
[Sales].[Customer]
[Sales].[PersonCreditCard]
[Sales].[SalesOrderDetail]
`SalesOrderDetail`
[Sales].[SalesOrderHeader]
[Sales].[SalesOrderHeaderSalesReason]
[Sales].[SalesPerson]
`SalesPerson`
[Sales].[SalesPersonQuotaHistory]
`SalesPersonQuotaHistory`
[Sales].[SalesReason]
[Sales].[SalesTaxRate]
[Sales].[SalesTerritory]
[Sales].[SalesTerritoryHistory]
Edit Mappings...
Preview...
Help
< Back
Next >
Finish >>|
Cancel


SQL Server Import and Export Wizard
Review Data Type Mapping
Select a table to review how its data types map to those in the destination and how it handles conversion issues.
Table:
Source
Destination
[Sales].[Currency]
`Currency`
[Sales].[SalesOrderDetail]
`SalesOrderDetail`
[Sales].[SalesPerson]
`SalesPerson`
[Sales].[SalesPersonQuotaHistory]
`SalesPersonQuotaHistory`
Data type mapping:
Source Column
Source Type
Destination Col...
Destination Type
Convert
On Error
On Truncati...
CurrencyCode
nchar
CurrencyCode
VarChar
Name
nvarchar
Name
LongText
Use Global
Use Global
ModifiedDate
datetime
ModifiedDate
DateTime
To view conversion details, double-click the row that contains the column source type to be converted.
On Error (global)
Fail
On Truncation (global)
Fail
Help
< Back
Next >
Finish >>|
Cancel



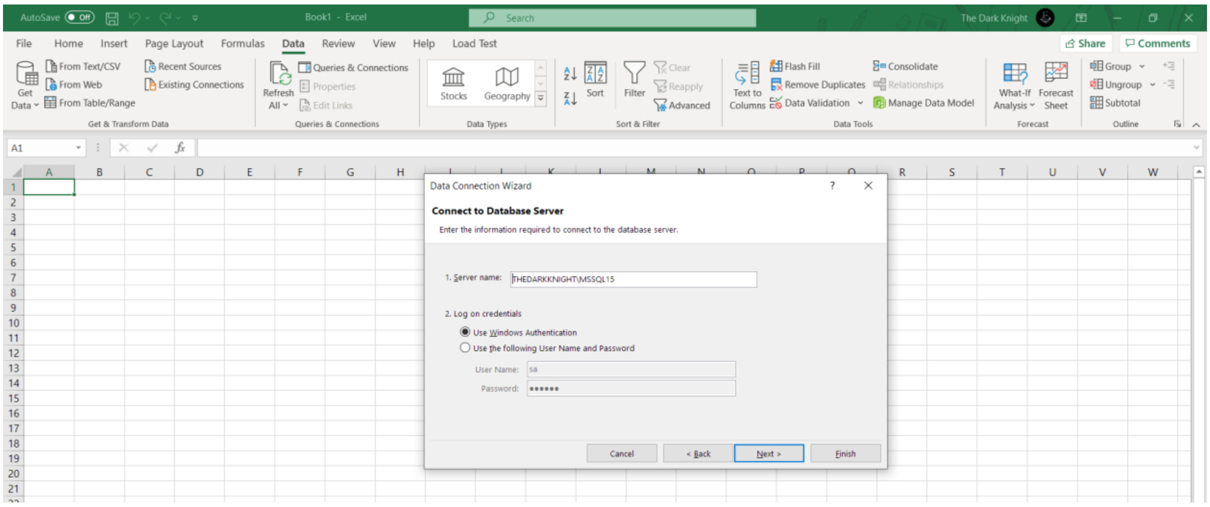
2. Open excel file & generate a graph using the data available.

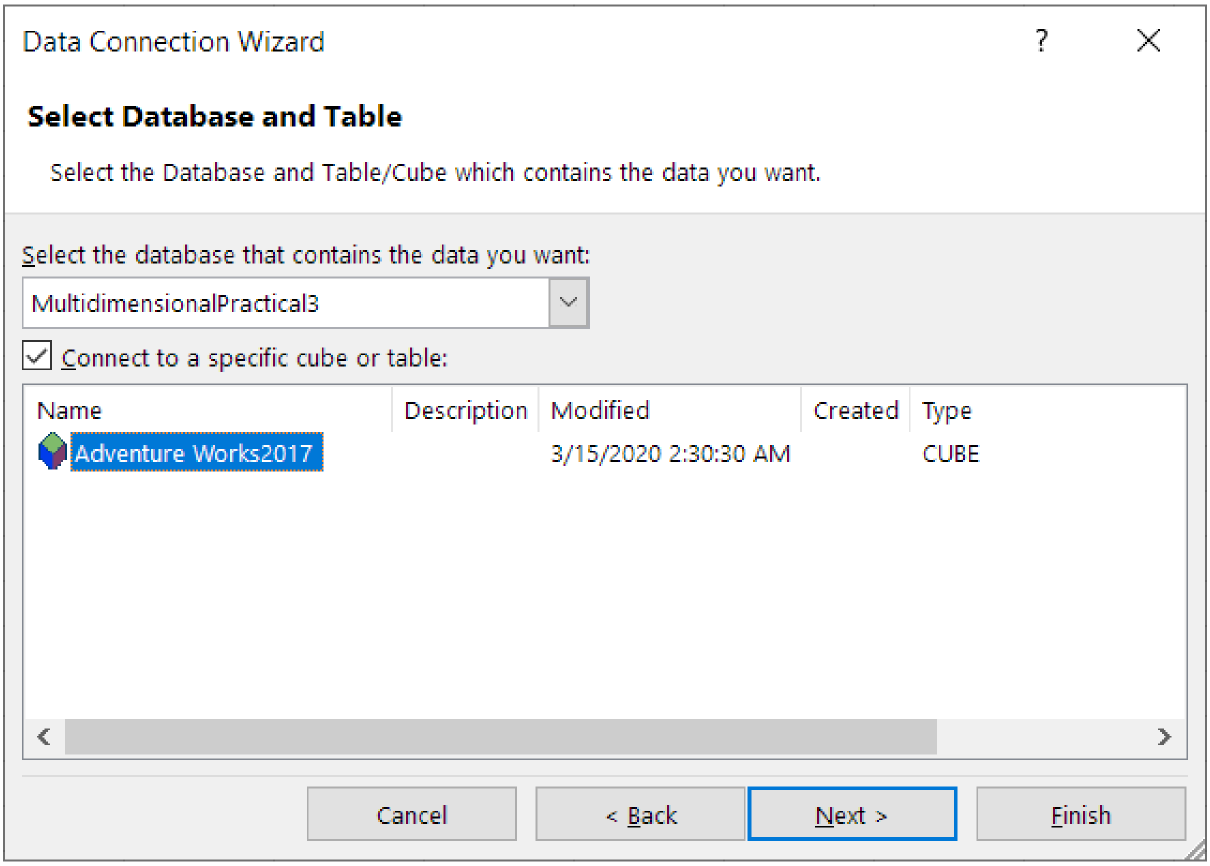


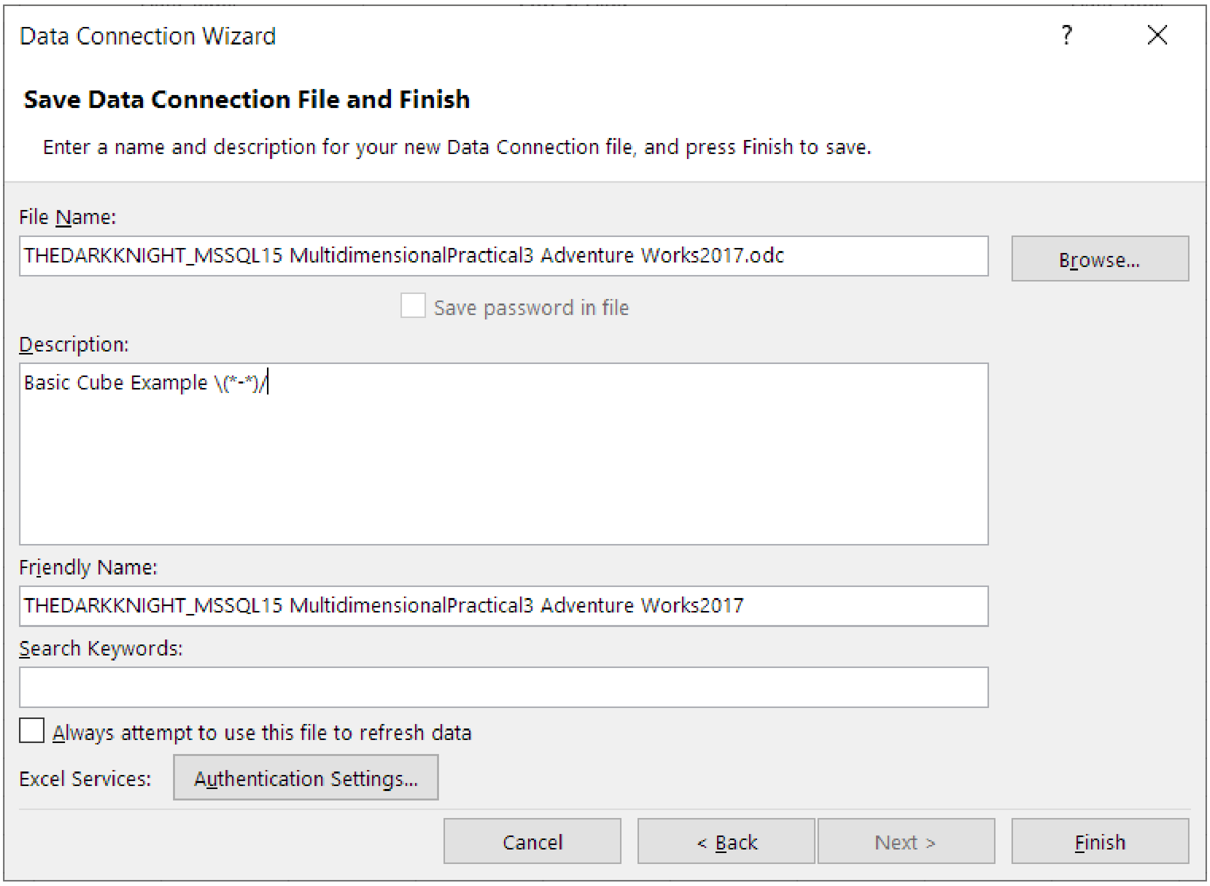
PART B -> Import the cube in Microsoft Excel and create the Pivot table and Pivot Chart to perform data analysis

Steps: -

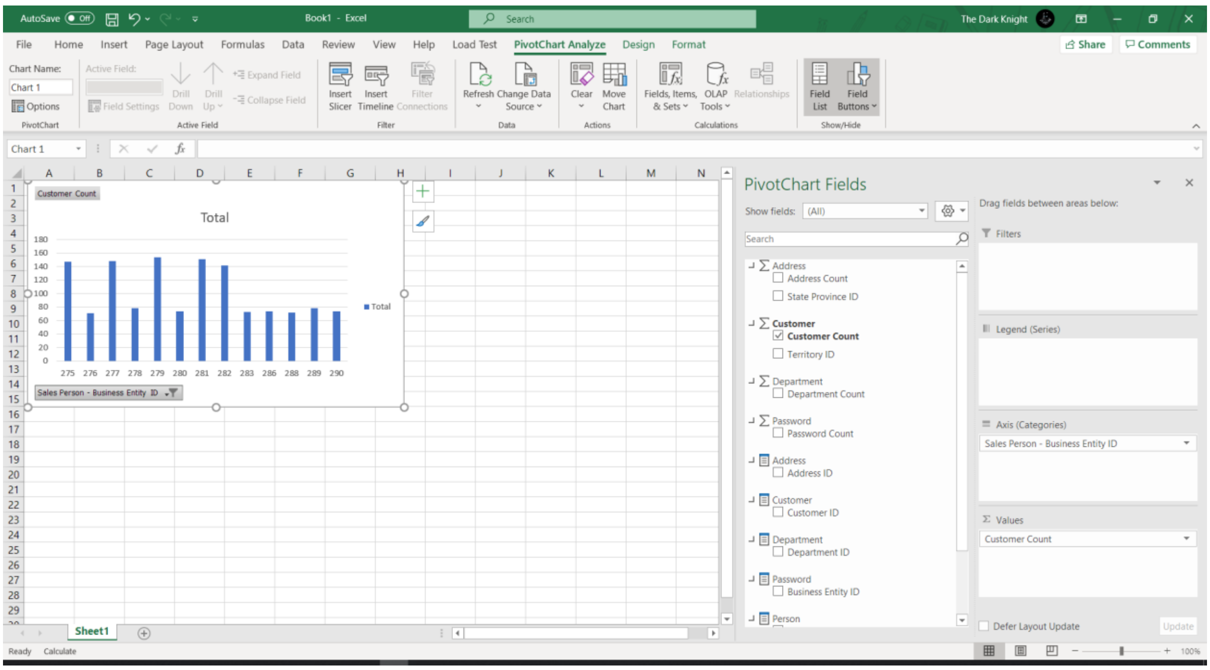
1. Connect to database server and add the dB with cubes.







2. Render the graph



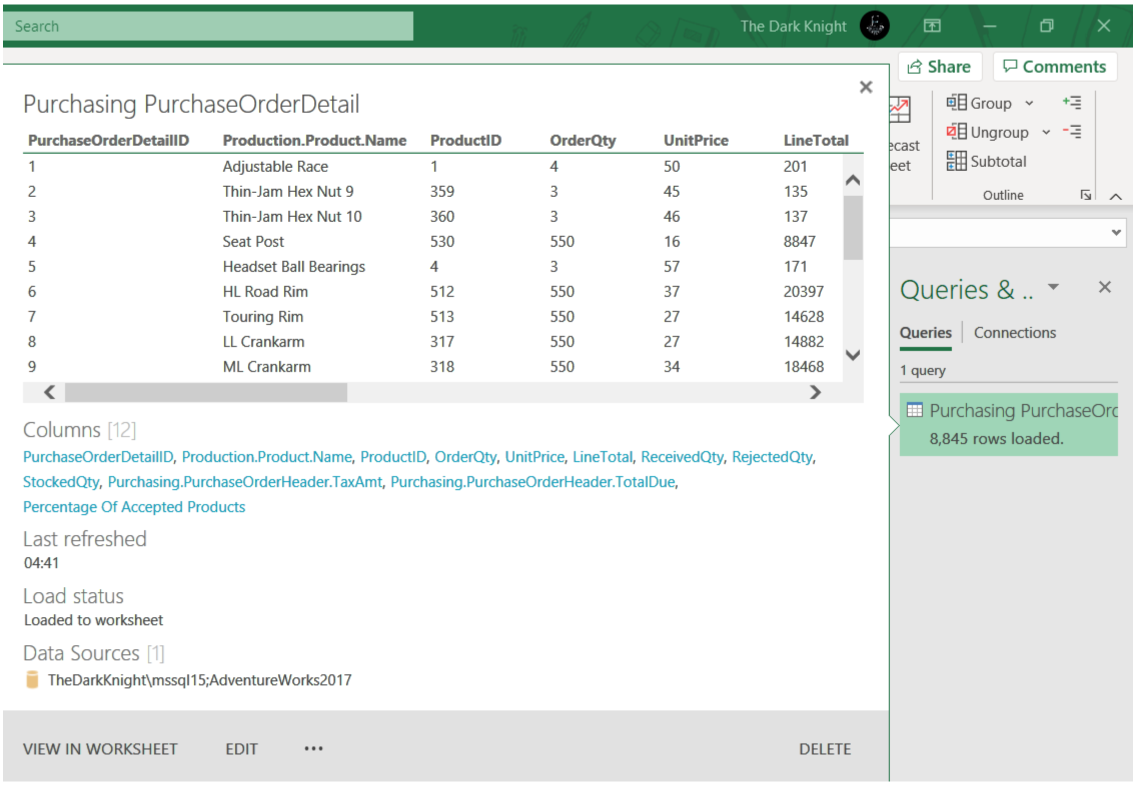
Kishinchand Chellaram College, Mumbai-2020
FY / SY / TY B Sc. (I.T) Semester ________

Practical No: - 6

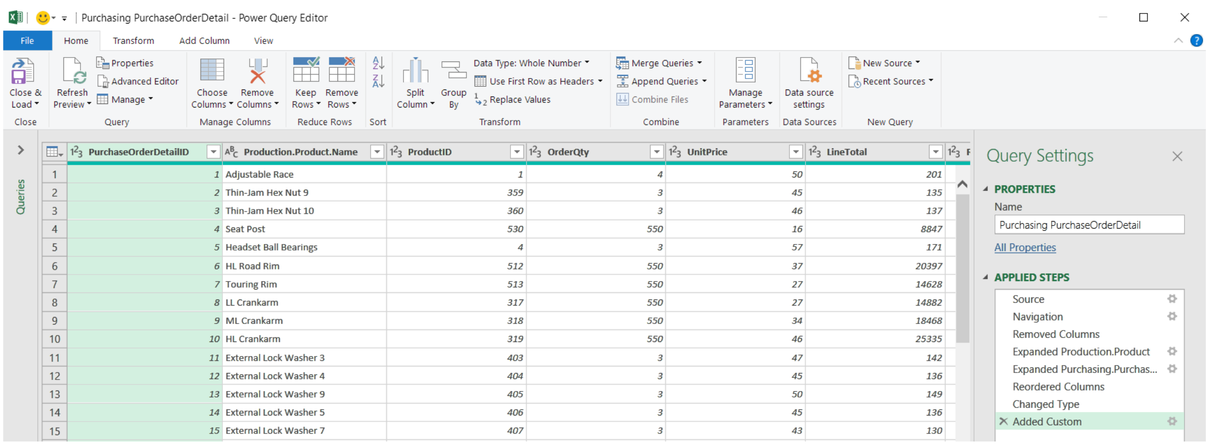
Aim: - Apply the what – if Analysis for data visualization. Design and generate necessary reports based on the data warehouse data.

Steps: -

1. Import data from dB to excel.

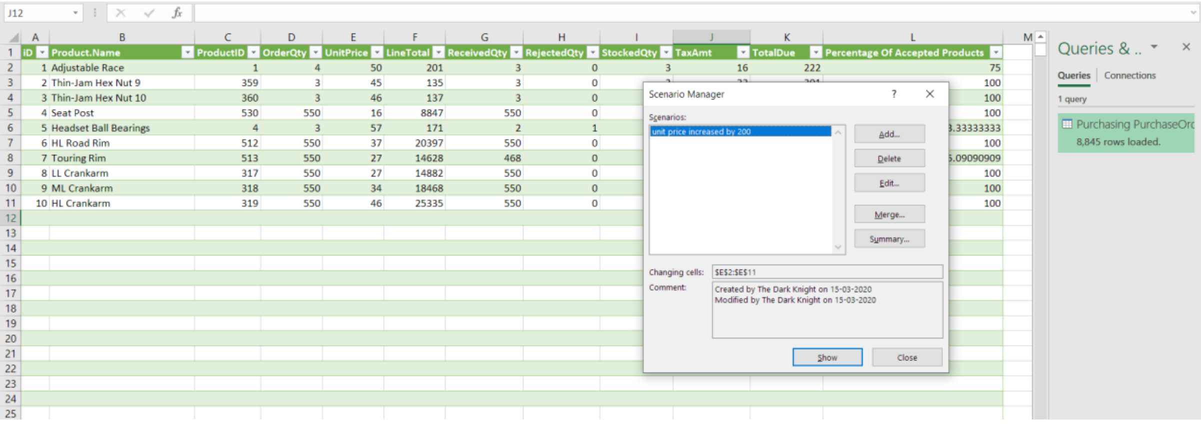


2. ETL the data.



3. Select Data -> What If Analysis -> Scenario Manager -> Add New Scenario.

Scenario: all unit prices increased by 200



4. Click on Summary to check the result.

B	C	D	E
Scenario Summary			
		Current Values:	unit price increased by 200
Changing Cells:			
	$E$2	50	250
	$E$3	45	245
	$E$4	46	246
	$E$5	16	216
	$E$6	57	257
	$E$7	37	237
	$E$8	27	227
	$E$9	27	227
	$E$10	34	234
	$E$11	46	246
Result Cells:			
	$F$2	200	1000
	$F$3	135	735
	$F$4	138	738
	$F$5	8800	118800
	$F$6	171	771
	$F$7	20350	130350
	$F$8	14850	124850
	$F$9	14850	124850
	$F$10	18700	128700
	$F$11	25300	135300

Notes: Current Values column represents values of changing cells at time Scenario Summary Report was created. Changing cells for each scenario are highlighted in gray.

Practical No: - 7

Aim: - Perform the data classification using classification algorithm.

PART A -> Perform logistic regression in the given data warehouse.

Code: -

```
input <- mtcars [, c("am","cyl","hp","wt")]
print(head(input))
am.data=glm (formula=am~cyl+hp+wt, data=input, family=binomial)
print(summary(am.data))
```

R Console

```
Call:
glm(formula = am ~ cyl + hp + wt, family = binomial, data = input)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.17272  -0.14907  -0.01464   0.14116   1.27641

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 19.70288    8.11637   2.428   0.0152 *
cyl          0.48760    1.07162   0.455   0.6491
hp           0.03259    0.01886   1.728   0.0840 .
wt          -9.14947    4.15332  -2.203   0.0276 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 43.2297  on 31  degrees of freedom
Residual deviance:  9.8415  on 28  degrees of freedom
AIC: 17.841

Number of Fisher Scoring iterations: 8

>
```

C:\Users\ampee\Documents\practical9.R - R Editor

```
input <- mtcars[,c("am","cyl","hp","wt")]
print(head(input))
am.data=glm(formula=am~cyl+hp+wt,data=input,family=binomial)
print(summary(am.data))
```

PART B -> Perform the data classification using classification algorithm

Code: -

library(party)
print(head(readingSkills))
input<-readingSkills [c (1:105),]
print(head(input))
png(file="decisiontree.png")
output. tree<-ctree (nativeSpeaker~age+shoeSize+score, data=input)
plot (output. tree)
dev.off ()

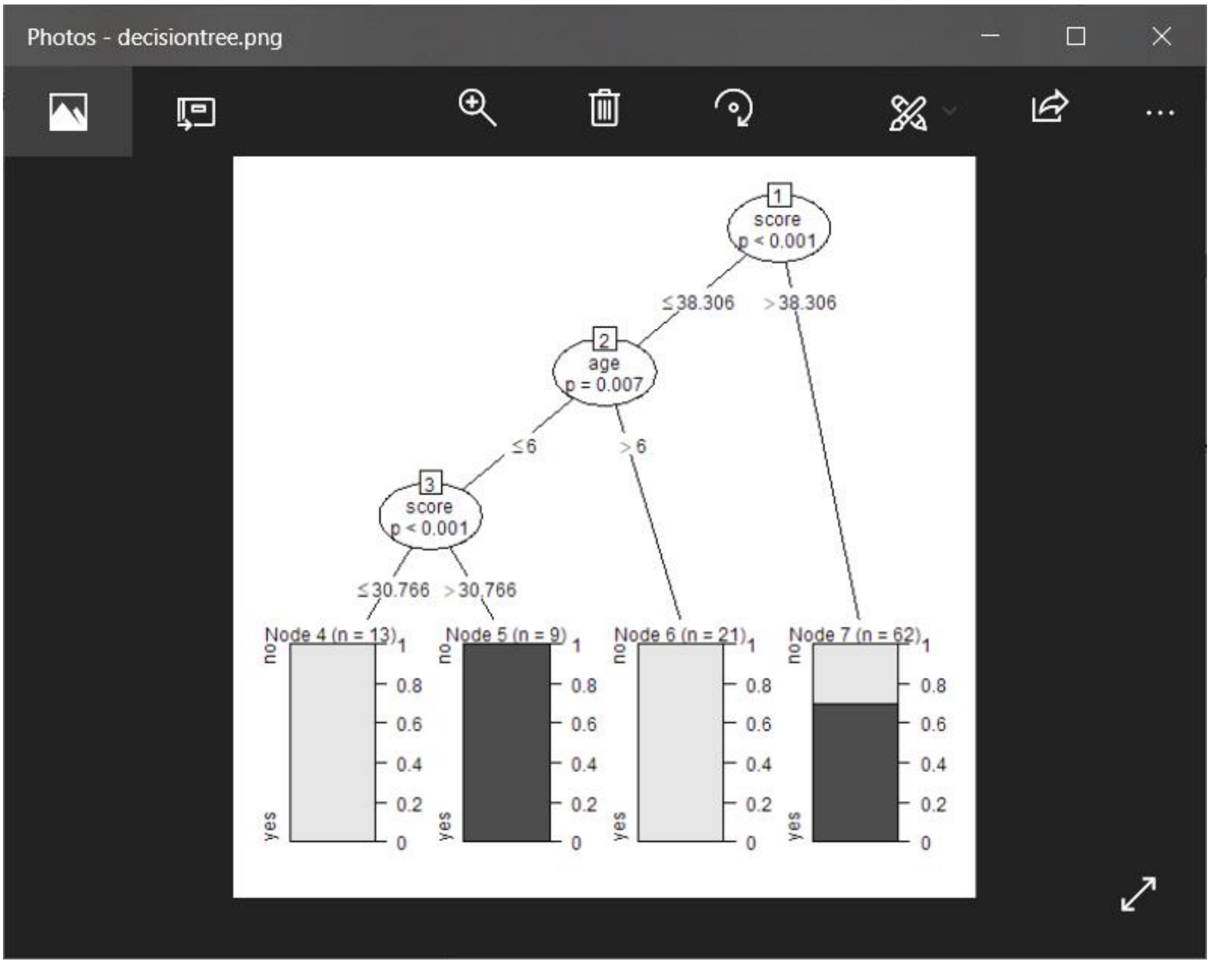
R Console

```
6: package 'sandwich' was built under R version 3.5.3
> print(head(readingSkills))
  nativeSpeaker age shoeSize    score
1           yes   5 24.83189 32.29385
2           yes   6 25.95238 36.63105
3            no  11 30.42170 49.60593
4           yes   7 28.66450 40.28456
5           yes  11 31.88207 55.46085
6           yes  10 30.07843 52.83124
> input<-readingSkills[c(1:105),]
> print(head(input))
  nativeSpeaker age shoeSize    score
1           yes   5 24.83189 32.29385
2           yes   6 25.95238 36.63105
3            no  11 30.42170 49.60593
4           yes   7 28.66450 40.28456
5           yes  11 31.88207 55.46085
6           yes  10 30.07843 52.83124
> png(file="decisiontree.png")
> output.tree<-ctree(nativeSpeaker~age+shoeSize+score,data=input)
> plot(output.tree)
> dev.off()
null device
          1
> 
```

C:\Users\ampee\Documents\practical9.R - R Editor

```
print(head(readingSkills))
input<-readingSkills[c(1:105),]
print(head(input))
png(file="decisiontree.png")
output.tree<-ctree(nativeSpeaker~age+shoeSize+score,data=input)
plot(output.tree)
dev.off()
```

From the decision tree, we can conclude the people with reading skills less than 38.3 and also age is greater than 6 they are not a motive speaker.

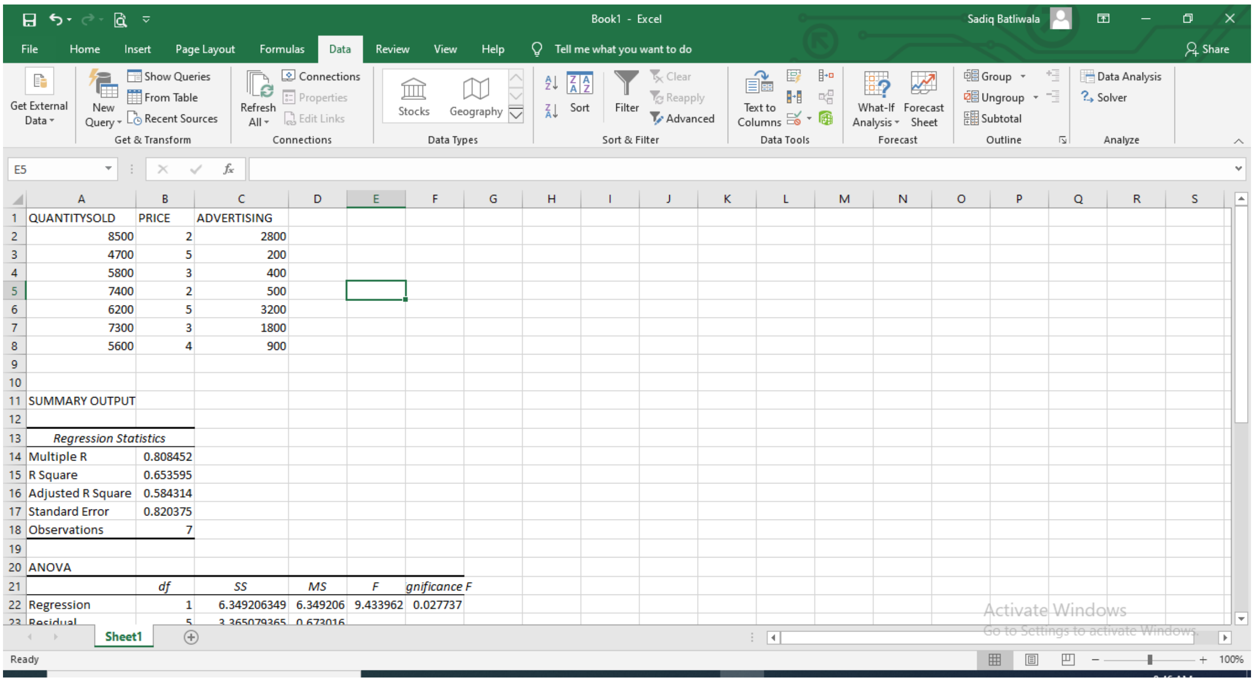


Kishinchand Chellaram College, Mumbai-2020
FY / SY / TY B Sc. (I.T) Semester ________

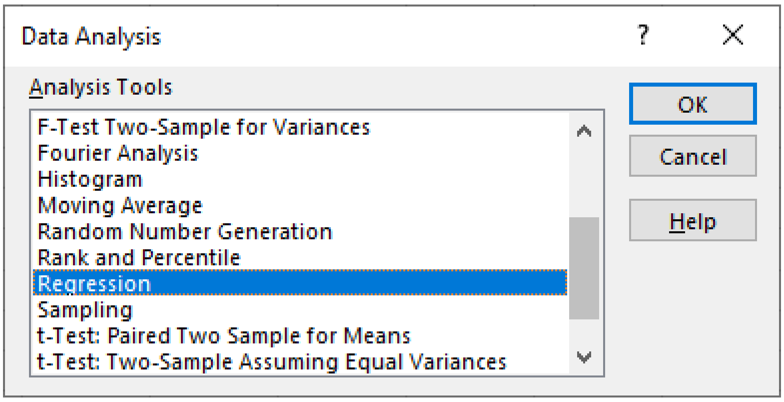
PART C -> Linear Regression with Excel

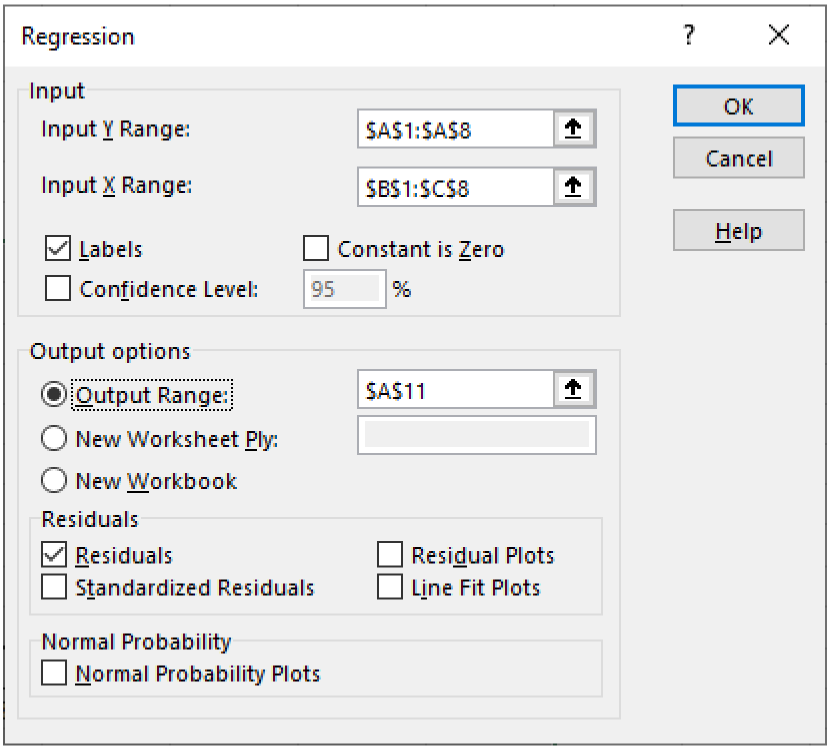
Steps: -

1. Import data into excel

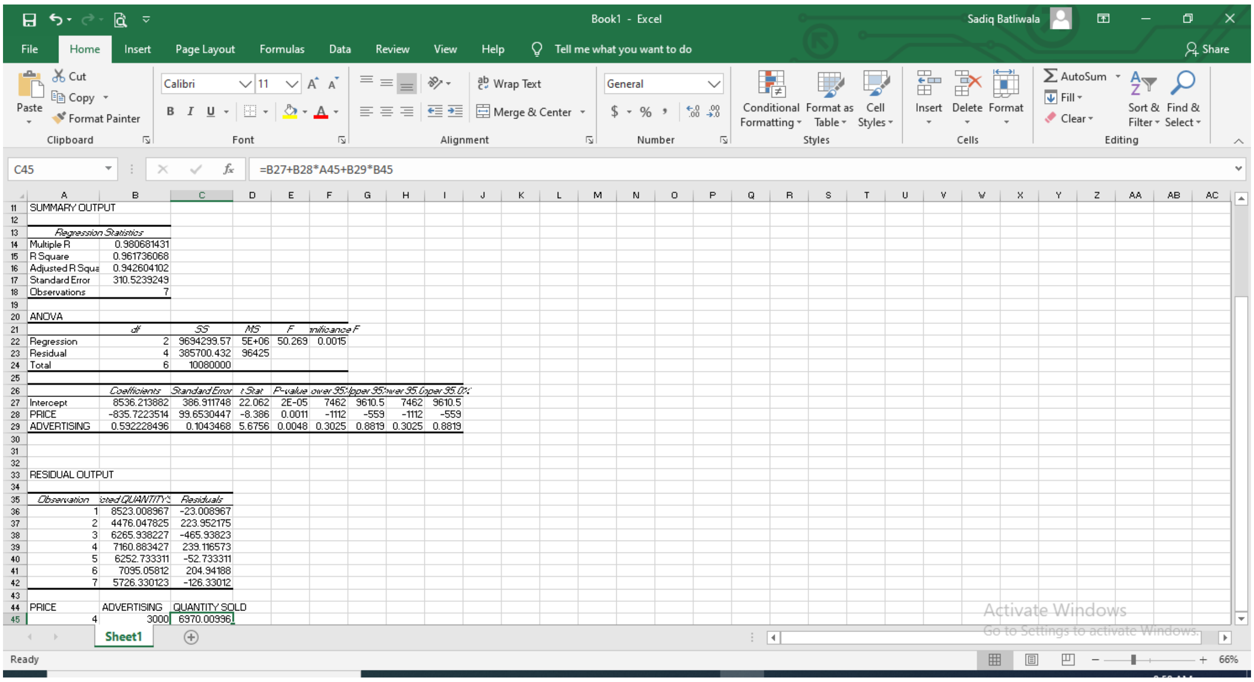


2. Apply Data Analysis -> Regression





Regression
Input
Input Y Range:
$A$1:$A$8
Input X Range:
$B$1:$C$8
Labels
Constant is Zero
Confidence Level:
95
%
OK
Cancel
Help
Output options
Output Range:
$A$11
New Worksheet Ply:
New Workbook
Residuals
Residuals
Residual Plots
Standardized Residuals
Line Fit Plots
Normal Probability
Normal Probability Plots



Book1 - Excel
C45
=B27+B28*A45+B29*B45
SUMMARY OUTPUT
Regression Statistics
Multiple R 0.980681431
R Square 0.961736068
Adjusted R Squa 0.942604102
Standard Error 310.5239249
Observations 7
ANOVA
df SS MS F ignificance F
Regression 2 9694299.57 5E+06 50.269 0.0015
Residual 4 385700.432 96425
Total 6 10080000
Coefficients Standard Error t Stat P-value
Intercept 8536.213882 386.911748 22.062 2E-05 7462 9610.5 7462 9610.5
PRICE -835.7223514 99.6530447 -8.386 0.0011 -1112 -559 -1112 -559
ADVERTISING 0.592228496 0.1043468 5.6756 0.0048 0.3025 0.8819 0.3025 0.8819
RESIDUAL OUTPUT
Observation Residuals
1 8523.008967 -23.008967
2 4476.047825 223.952175
3 6265.938227 -465.93823
4 7160.883427 239.116573
5 6252.733311 -52.733311
6 7095.05812 204.94188
7 5726.330123 -126.33012
PRICE ADVERTISING QUANTITY SOLD
4 3000 6970.00996
Sheet1
Ready

3. Render the Output & Check the result -> APPLY Formula=B17+B18*A36+B19*B36

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.980681
R Square	0.961736
Adjusted R Square	0.942604
Standard Error	310.5239
Observations	7

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	9694300	4847150	50.26854	0.001464
Residual	4	385700.4	96425.11		
Total	6	10080000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	8536.214	386.9117	22.06243	2.5E-05	7461.975	9610.453	7461.975	9610.453
Price	-835.722	99.65304	-8.38632	0.001106	-1112.4	-559.041	-1112.4	-559.041
Advertising	0.592228	0.104347	5.675579	0.004755	0.302515	0.881942	0.302515	0.881942

RESIDUAL OUTPUT

Observation	*Predicted Quantity Sold*	*Residuals*
1	8523.009	-23.009
2	4476.048	223.9522
3	6265.938	-465.938
4	7160.883	239.1166
5	6252.733	-52.7333
6	7095.058	204.9419
7	5726.33	-126.33

Price	advertisement	quantity sold
4	3000	6970.01

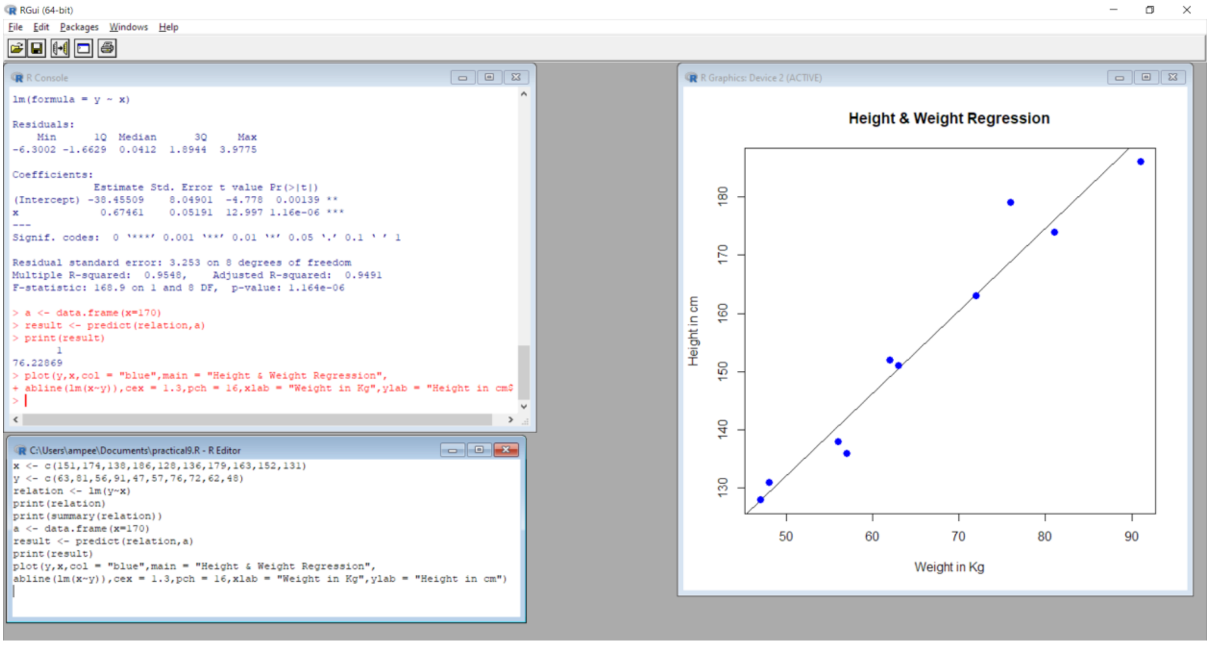
Kishinchand Chellaram College, Mumbai-2020
FY / SY / TY B Sc. (I.T) Semester ________

Practical No: - 9

Aim: - Perform the Linear regression on the given data warehouse data.

Code: -

```
x <- c (151,174,138,186,128,136,179,163,152,131)
y <- c (63,81,56,91,47,57,76,72,62,48)
relation <- lm(y~x)
print(relation)
print(summary(relation))
a <- data. frame(x=170)
result <- predict (relation, a)
print(result)
plot (y, x, col = "blue", main = "Height & Weight Regression",
abline(lm(x~y)), cex = 1.3, pch = 16, xlab = "Weight in Kg", ylab = "Height in cm")
```



Kishinchand Chellaram College, Mumbai-2020
FY / SY / TY B Sc. (I.T) Semester ________

Practical No: - 10

Aim: - Perform the logistic regression on the given data warehouse data.

Code: -

```
bodysize=rnorm (20,30,2)
bodysize=sort(bodysize)
survive=c (0,0,0,0,0,1,0,1,0,0,1,1,0,1,1,1,0,1,1,1)
dat=as.data. frame (cbind (bodysize, survive))
dat
plot (bodysize, survive, xlab="Body size", ylab="Probability of survival")
g=glm (survive~bodysize, family=binomial, dat)
curve (predict (g, data. frame(bodysize=x), type="resp"), add=TRUE)
points (bodysize, fitted(g), pch=20)
```