Business Intelligence and Data Analytics Practical

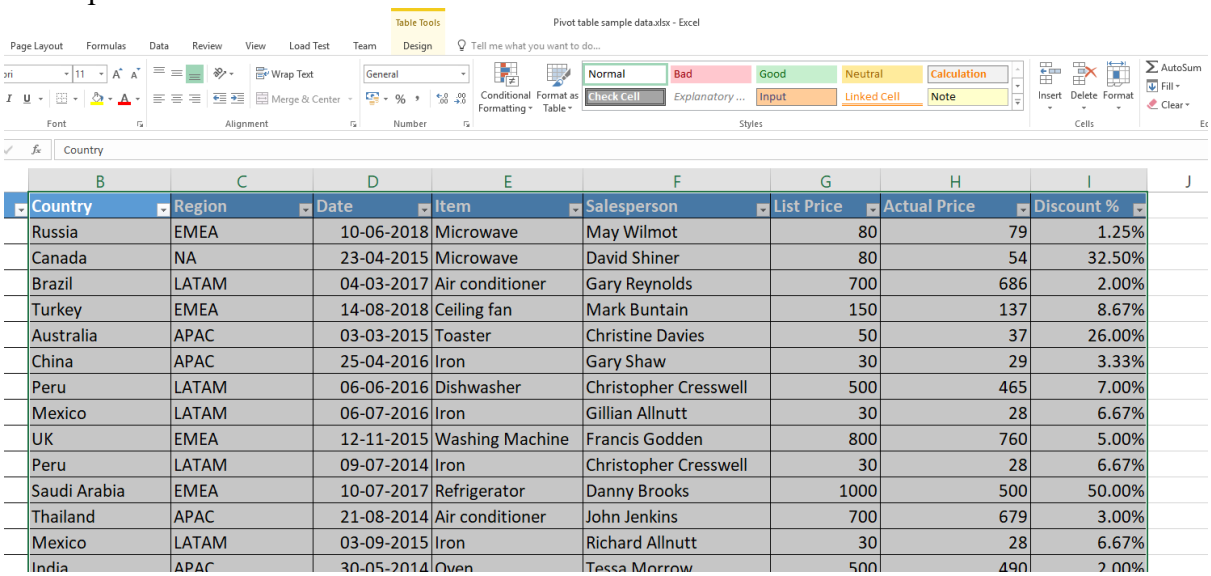
INDEX

Sr. No	Practical Title	Sign
1.	Import the data warehouse data in Microsoft Excel and create the Pivot table and Pivot Chart.	
2.	Apply the what – if Analysis for data visualization. Design and generate necessary reports based on the data warehouse data. Use Excel.	
3.	Perform the data classification using classification algorithm using R/Python.	
4.	Perform the data clustering using clustering algorithm using R/Python.	
5.	Perform the Linear regression on the given data warehouse data using R/Python.	
6.	Perform the logistic regression on the given data warehouse data using R/Python.	
7.	Write a Python program to read data from a CSV file, perform simple data analysis, and generate basic insights. (Use Pandas is a Python library).	
8.	Perform data visualization using Python on any sales data.	
9.	Perform data visualization using Power BI on any sales data.	

1. Import the data warehouse data in Microsoft Excel and create the Pivot table and Pivot Chart.

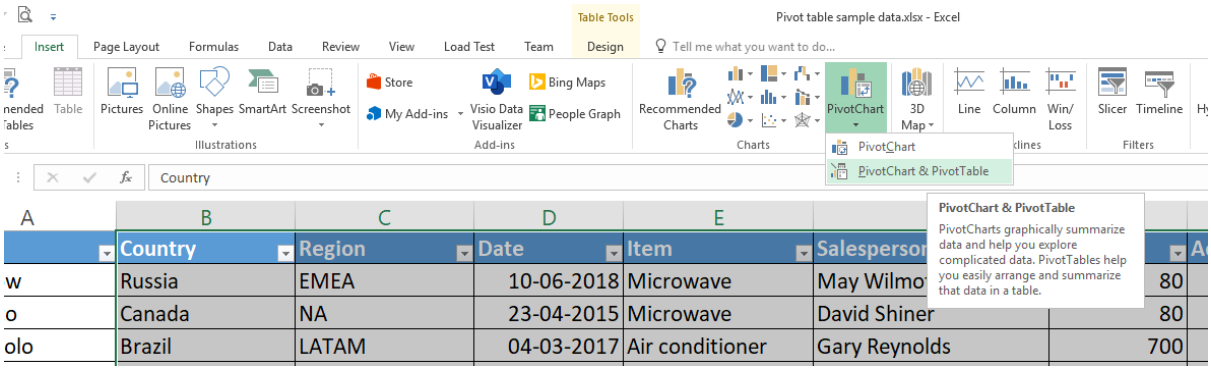
When we want to collect count, sum, and values either in tabular form or in the form of 2-column groupings, we may use an excel pivot table to categorize, sort, filter, and summarise any length of the data table. To insert a pivot table that will automatically locate a table or range, choose the Pivot table option from the Insert menu tab. By taking into account the necessary columns, we may also design a customizable table.

Step 1: Enter all the data in the excel sheet that is already done above as you can see in the above picture.

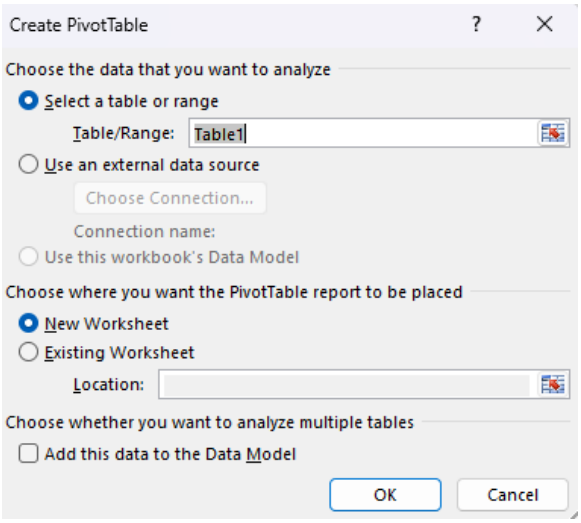


Country	Region	Date	Item	Salesperson	List Price	Actual Price	Discount %
Russia	EMEA	10-06-2018	Microwave	May Wilmot	80	79	1.25%
Canada	NA	23-04-2015	Microwave	David Shiner	80	54	32.50%
Brazil	LATAM	04-03-2017	Air conditioner	Gary Reynolds	700	686	2.00%
Turkey	EMEA	14-08-2018	Ceiling fan	Mark Buntain	150	137	8.67%
Australia	APAC	03-03-2015	Toaster	Christine Davies	50	37	26.00%
China	APAC	25-04-2016	Iron	Gary Shaw	30	29	3.33%
Peru	LATAM	06-06-2016	Dishwasher	Christopher Cresswell	500	465	7.00%
Mexico	LATAM	06-07-2016	Iron	Gillian Allnutt	30	28	6.67%
UK	EMEA	12-11-2015	Washing Machine	Francis Godden	800	760	5.00%
Peru	LATAM	09-07-2014	Iron	Christopher Cresswell	30	28	6.67%
Saudi Arabia	EMEA	10-07-2017	Refrigerator	Danny Brooks	1000	500	50.00%
Thailand	APAC	21-08-2014	Air conditioner	John Jenkins	700	679	3.00%
Mexico	LATAM	03-09-2015	Iron	Richard Allnutt	30	28	6.67%
India	APAC	30-05-2014	Oven	Tessa Morrow	500	490	2.00%

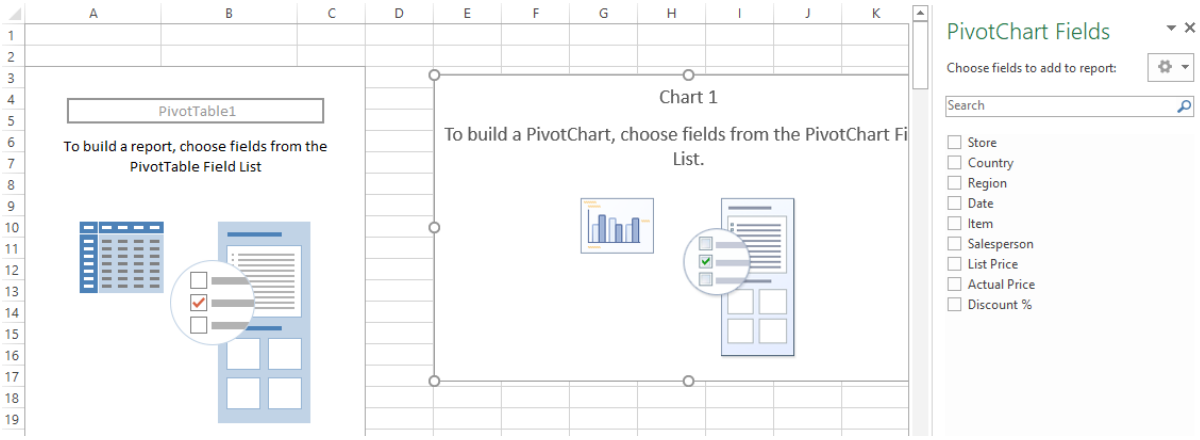
Step 2: Select all the data and then go to the INSERT option and then you will see here the option of PIVOT TABLE.



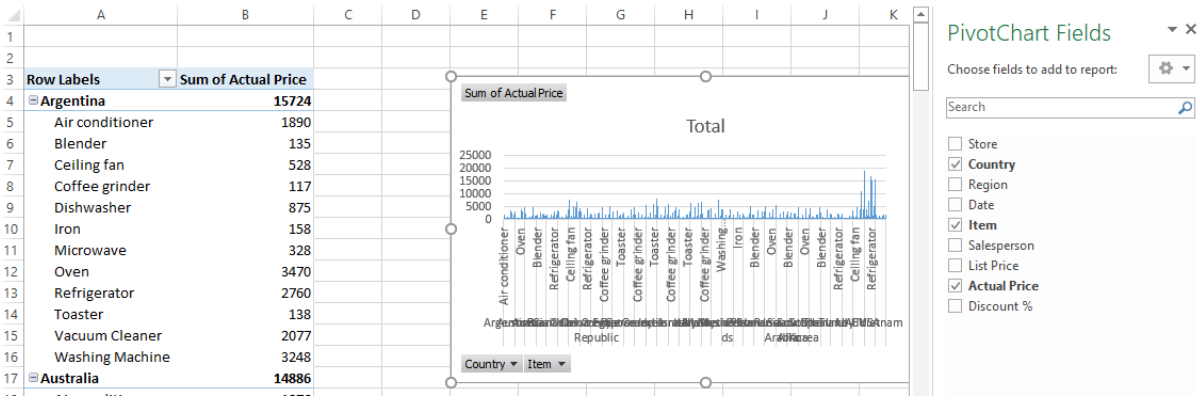
Step 3: After choosing the pivot table a box will open in which you have to enter the range in the location box. Enter the range that is mentioned below in the image.



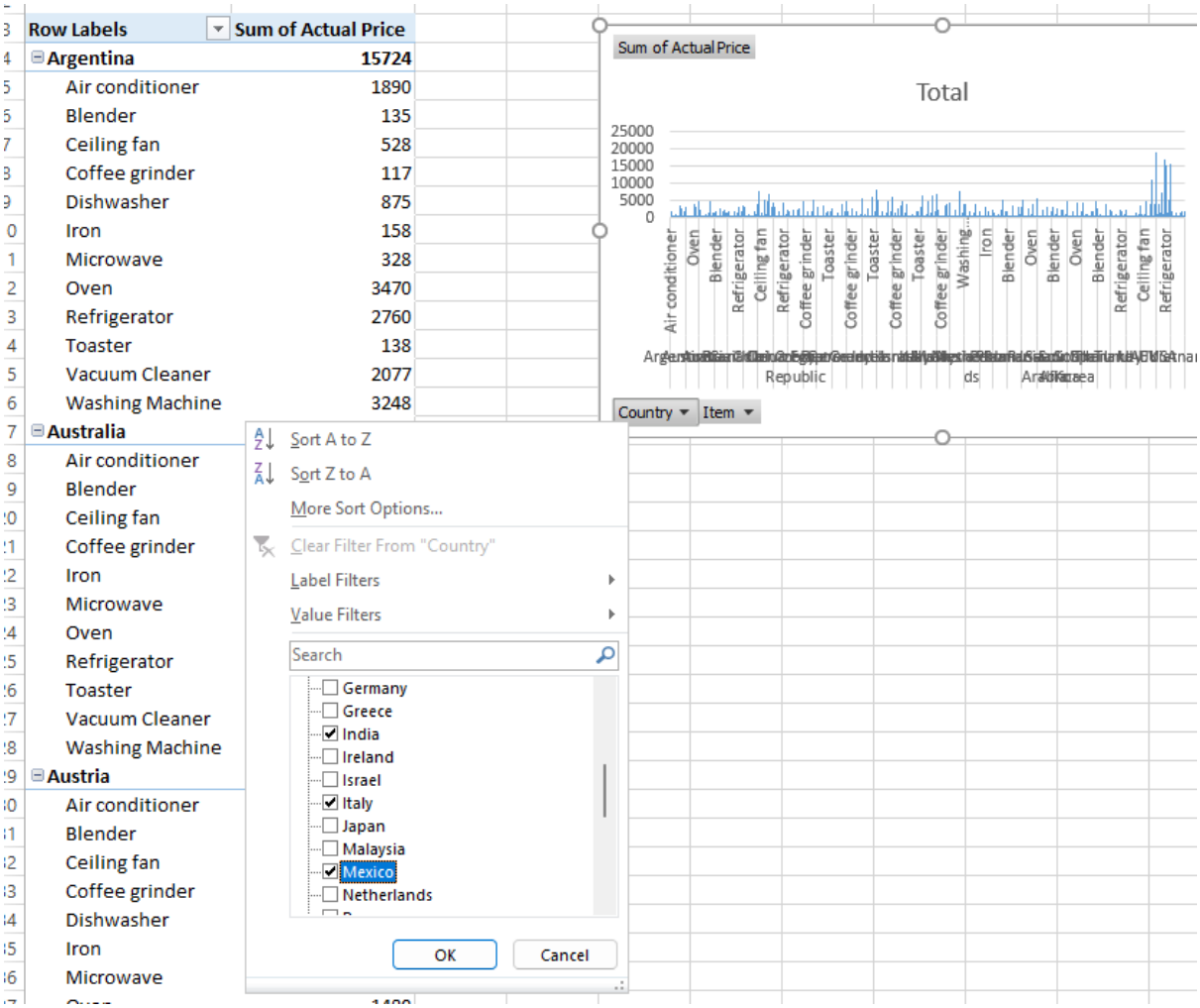
Step 4: Now you will the pivot table with pivot the table field that is shown in the below image and then you have to select Country, Items, and Actual Price in the pivot table field.

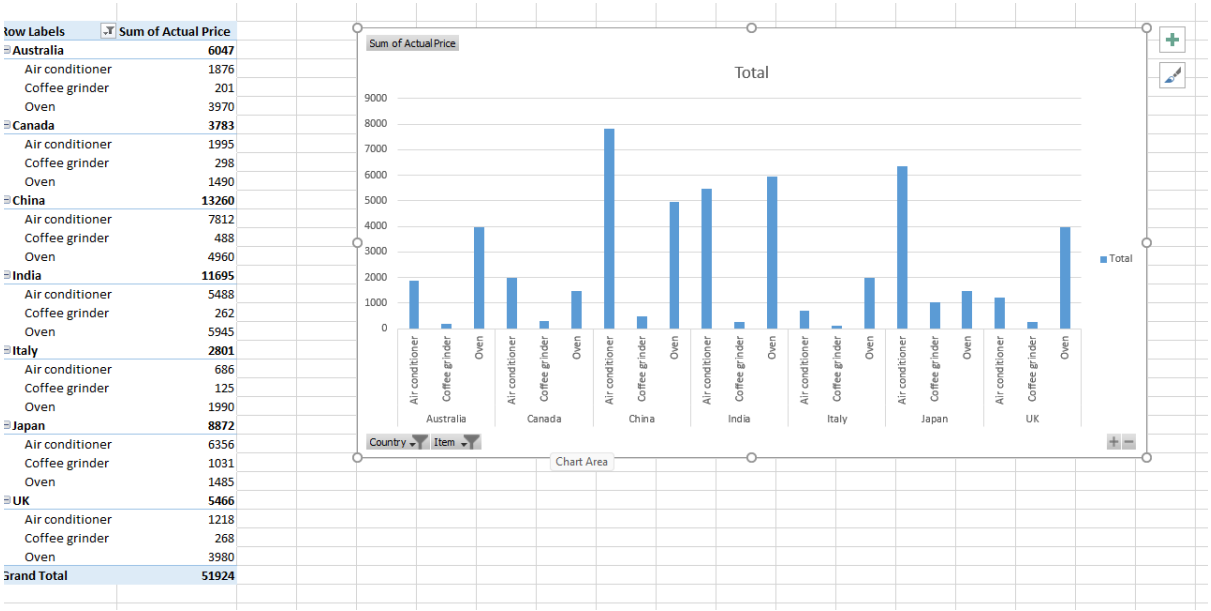


Step 5: After the selection, you will observe that your pivot table will be like in this form that is shown below in the image. This pivot table has both correct and incorrect data for each student.



Step 6: After clicking on the option, now you will see a field where you have to first Change the options from Country.





2. Apply the what – if Analysis for data visualization. Design and generate necessary reports based on the Sample data using Excel.

What-if analysis is a procedure in excel in which we work in tabular form data. In the What-if analysis variety of values have been in the cell of the excel sheet to see the result in different ways by not creating different sheets. There are three tools of what-if analysis.

Tools of what-if analysis

There are three tools in what-if analysis:

- Goal seek
- Scenario manager
- Data Table

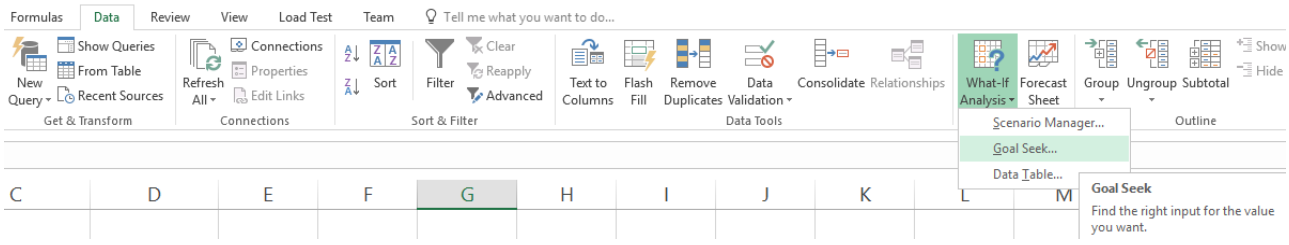
Goal seek

In goal seek we already know our output value we have to find the correct input value. For example, if a student wants to know his English marks and he knows all the rest of the marks and total marks in all subjects.

Step 1: Write all Product name, Quantity, prise and their total in an excel sheet and do the sum by applying the formula sum.

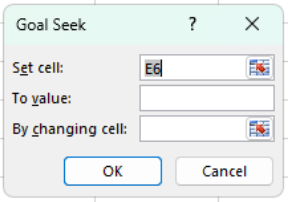
	A	B	C	D	E
1	**Sr.no**	**Product Name**	**Product Qty**	**Price**	**Total**
2	1	A	10	70	700
3	2	B	16	10	160
4	3	C	17	30	510
5	4	D	10		0
6					**1370**

Step 2: Go into the data tab of the Toolbar. Select the **What-if analysis,** drop-down appears. Select the **Goal Seek.**



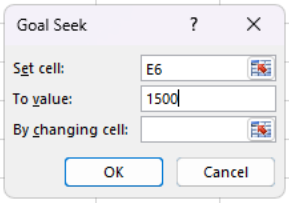
The dialogue box appears in the first column write the name of the cell in which you apply the formula sum. Type E6 in Set cell.

	A	B	C	D	E
1	Sr.no	Product Name	Product Qty	Price	Total
2	1	A	10	70	700
3	2	B	16	10	160
4	3	C	17	30	510
5	4	D	10		0
6					1370



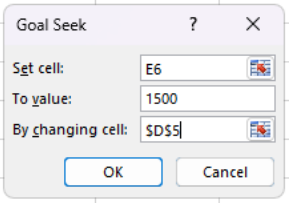
Step 3: In the second column write the value of the target (To Value). The target value for this example is **1500**.

	A	B	C	D	E
1	Sr.no	Product Name	Product Qty	Price	Total
2	1	A	10	70	700
3	2	B	16	10	160
4	3	C	17	30	510
5	4	D	10		0
6					1370



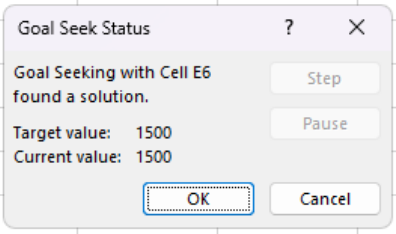
Step 4: In the third column write the name of the cell in which you want to get price for product "**D**". Provide absolute cell reference, i.e. D5.

	A	B	C	D	E
1	Sr.no	Product Name	Product Qty	Price	Total
2	1	A	10	70	700
3	2	B	16	10	160
4	3	C	17	30	510
5	4	D	10		0
6					1370



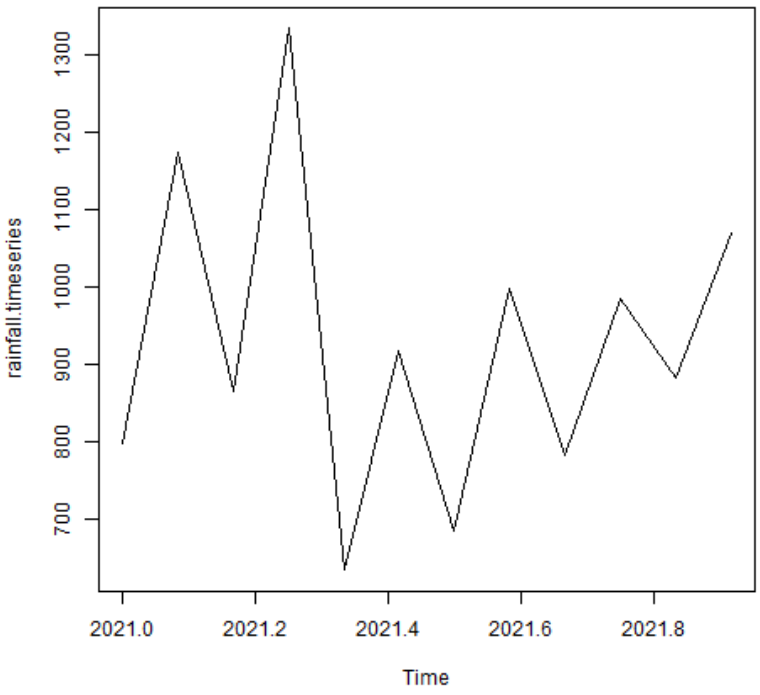
Step 5: Click ok and see the result. The estimated Price for "D" Product is 13.

	A	B	C	D	E
1	Sr.no	Product Name	Product Qty	Price	Total
2	1	A	10	70	700
3	2	B	16	10	160
4	3	C	17	30	510
5	4	D	10	13	130
6					1500



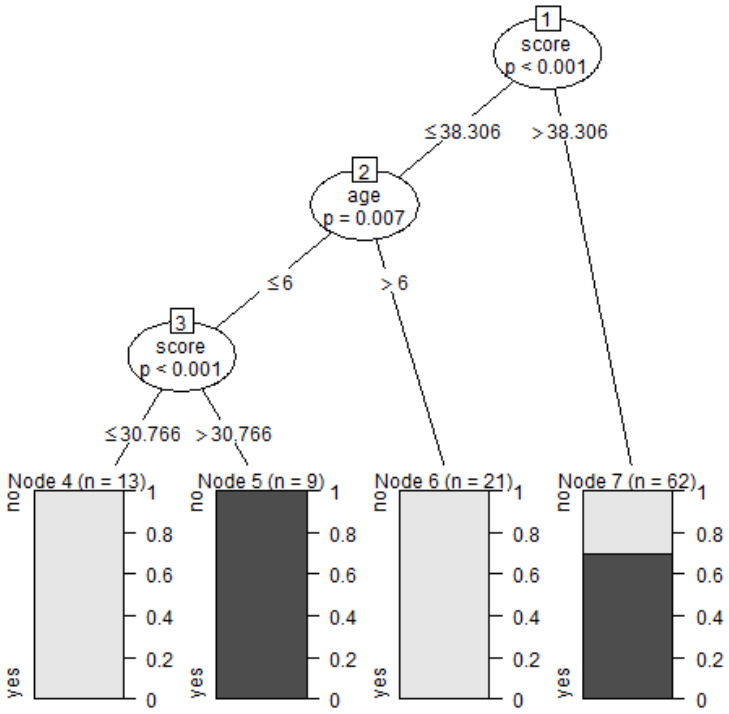
3. Perform the data classification using classification algorithm using R.

```
rainfall<-c(799,1174.8,865.1,1334.6,635.4,918.5,685.5,998.6,784.2,985,882.8,1071)
rainfall.timeseries <- ts(rainfall,start = c(2021,1),frequency = 12)
print(rainfall.timeseries)
png(file="rainfall.png")
plot(rainfall.timeseries)
dev.off()

# Get the data points in form of a R vector.
# Convert it to a time series object.
# Print the timeseries data.
# Give the chart file a name.
# Plot a graph of the time series.
# Save the file.
```



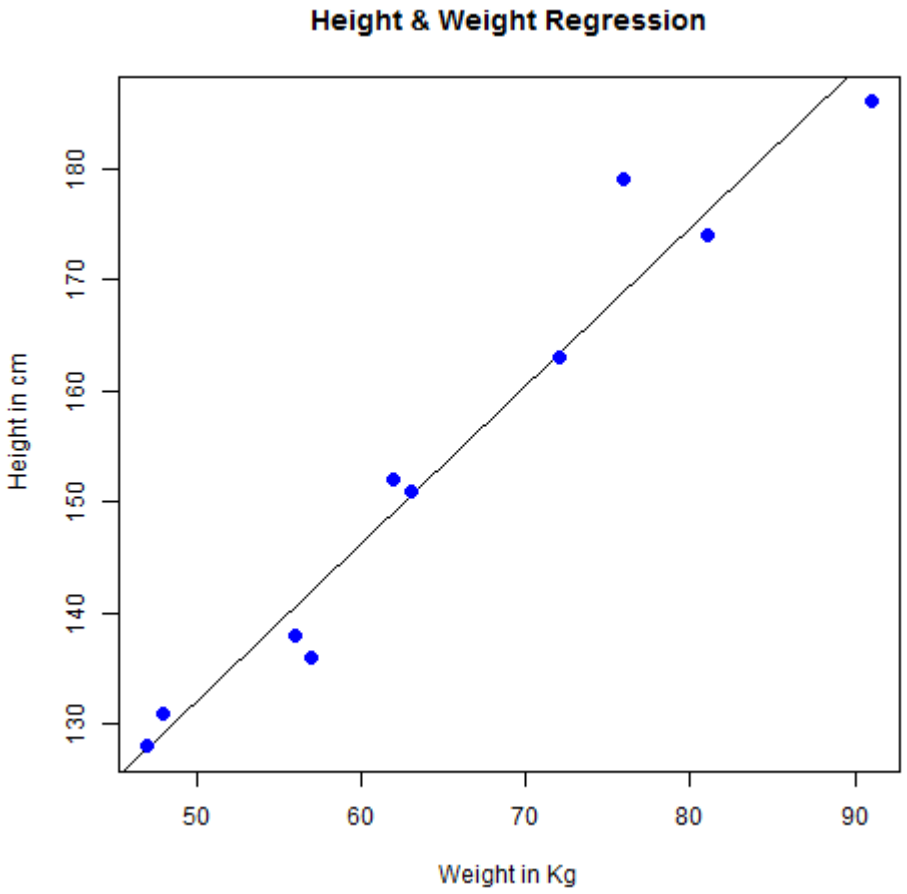
4. Perform the data clustering using clustering algorithm using R.

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

# Load the party package.
# It will automatically load other dependent packages.
# Create the input data frame.
# Give the chart file a name.
# Create the tree.
# Plot the tree.
# Save the file.
```



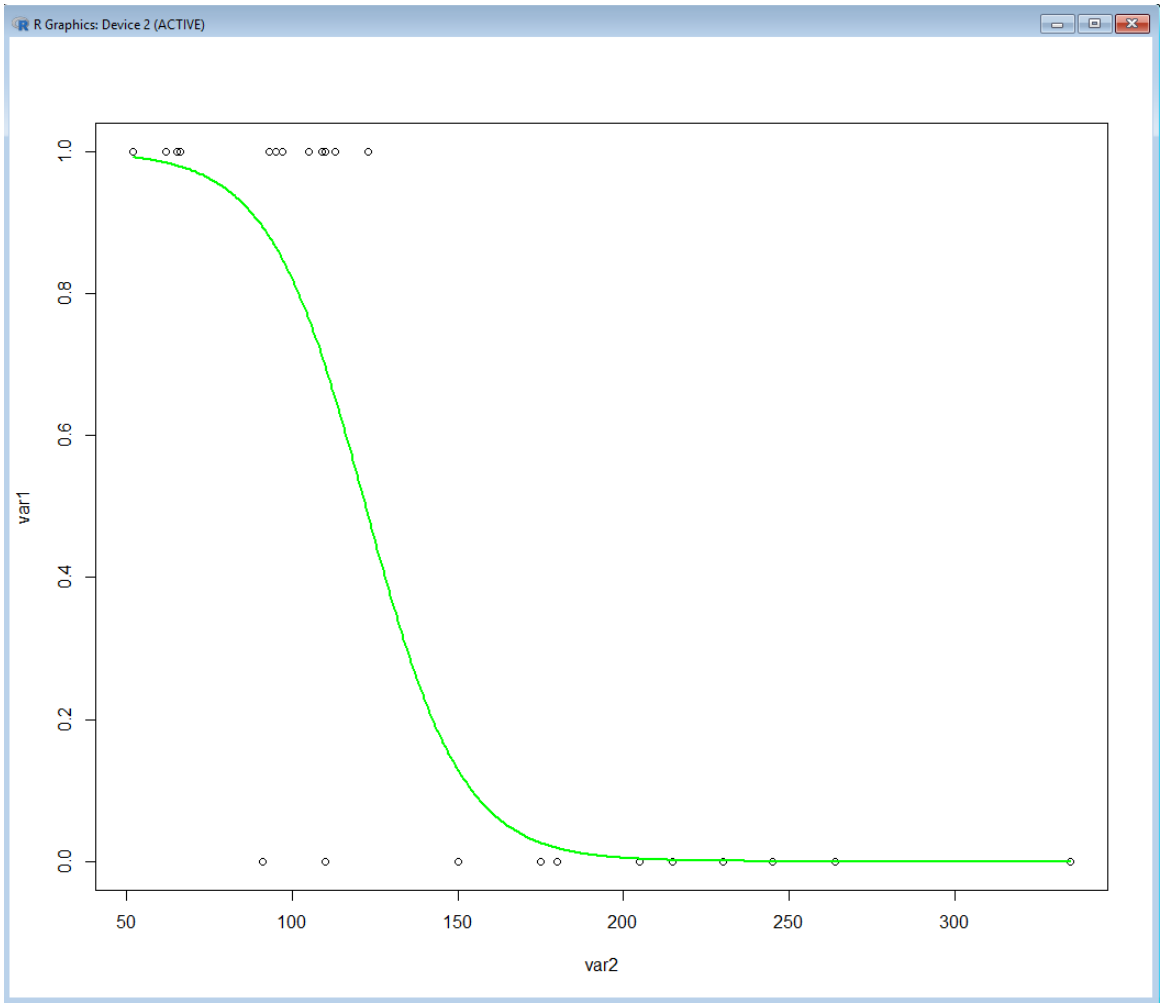
5. Perform the Linear regression on the given data warehouse data using R.

```
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)
png(file = "linearregression.png")
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")
dev.off()

# Create the predictor and response variable.
# Give the chart file a name.
# Plot the chart.
# Save the file.
```



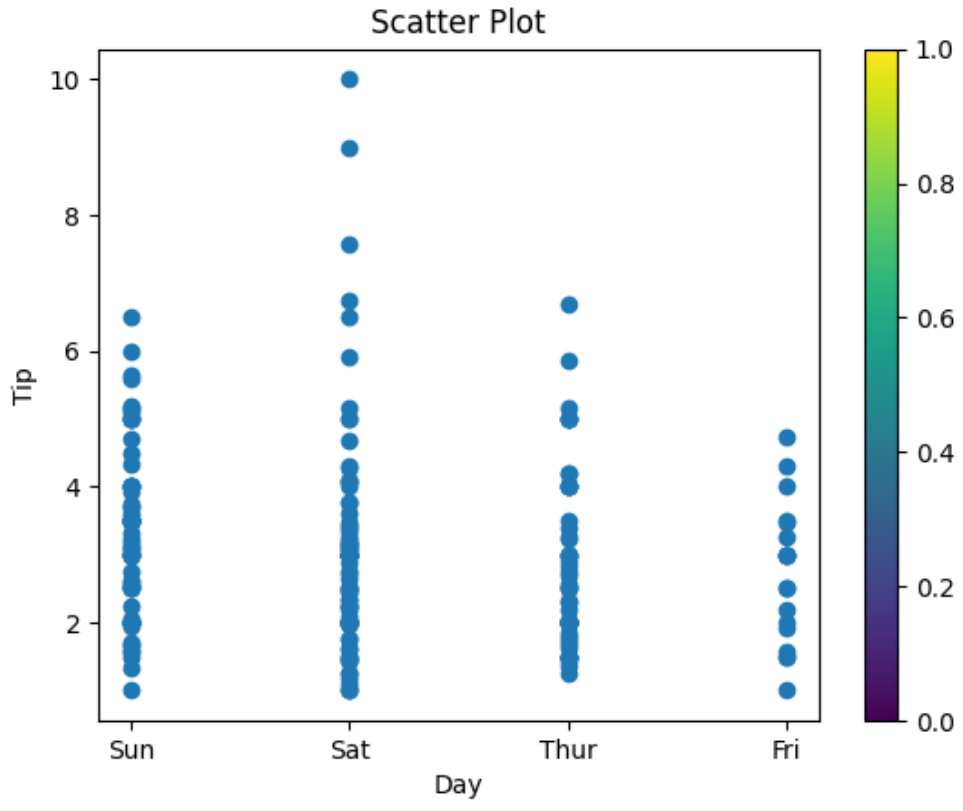
6. Perform the logistic regression on the given data warehouse data using R.

```
# load dataframe
df <- read.csv("h://Sample4.csv")
 # create logistic regression model
logistic_model<-glm(var1 ~ var2, data=df, family=binomial)
 #Data frame with hp in ascending order
Predicted_data<-data.frame(var2=seq(min(df$var2), max(df$var2),len=500))
 # Fill predicted values using regression model
Predicted_data$var1 = predict(logistic_model, Predicted_data, type="response")
 # Plot Predicted data and original data points
plot(var1 ~ var2, data=df)
lines(var1 ~ var2, Predicted_data, lwd=2, col="green")
```



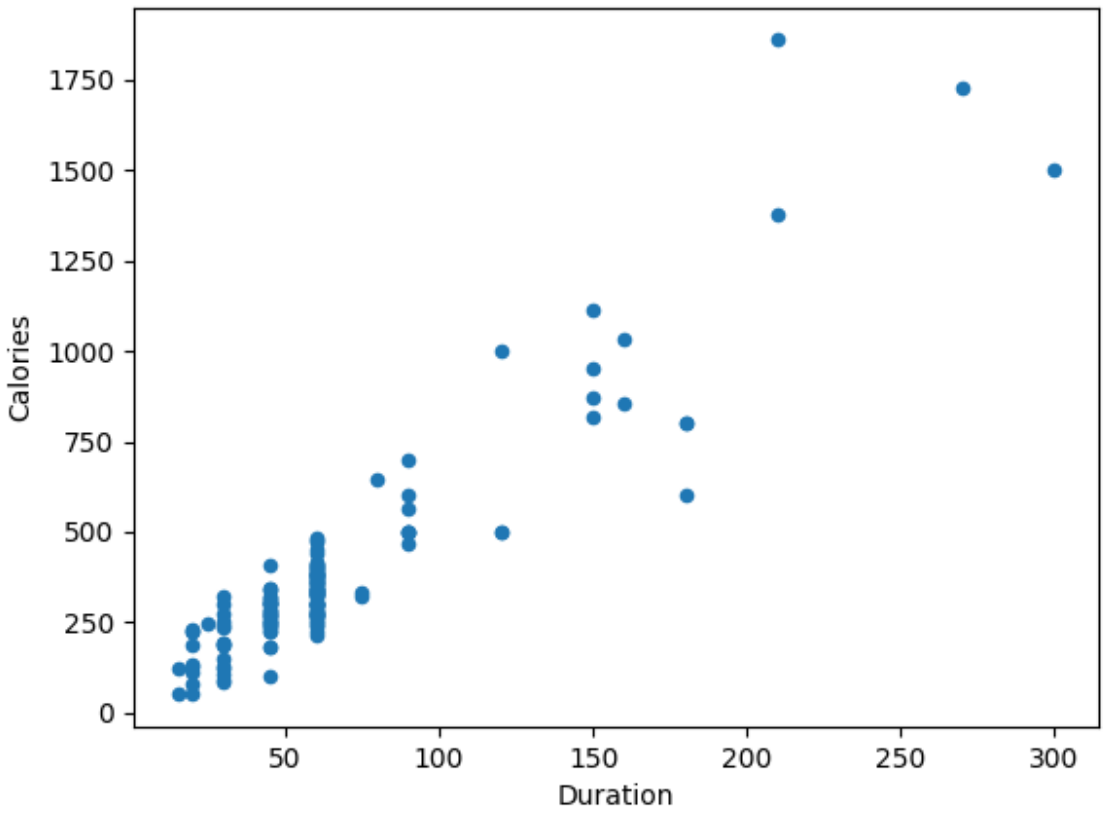
7. Write a Python program to read data from a CSV file, perform simple data analysis, and generate basic insights. (Use Pandas is a Python library).

```
>pip3 install pandas
>pip install matplotlib
>pip3 list

import pandas as pd
import matplotlib.pyplot as plt

# reading the database
data = pd.read_csv("tips.csv")

# Scatter plot with day against tip
plt.scatter(data['day'], data['tip'])

# Adding Title to the Plot
plt.title("Scatter Plot")

# Setting the X and Y labels
plt.xlabel('Day')
plt.ylabel('Tip')
plt.colorbar()
plt.show()
```



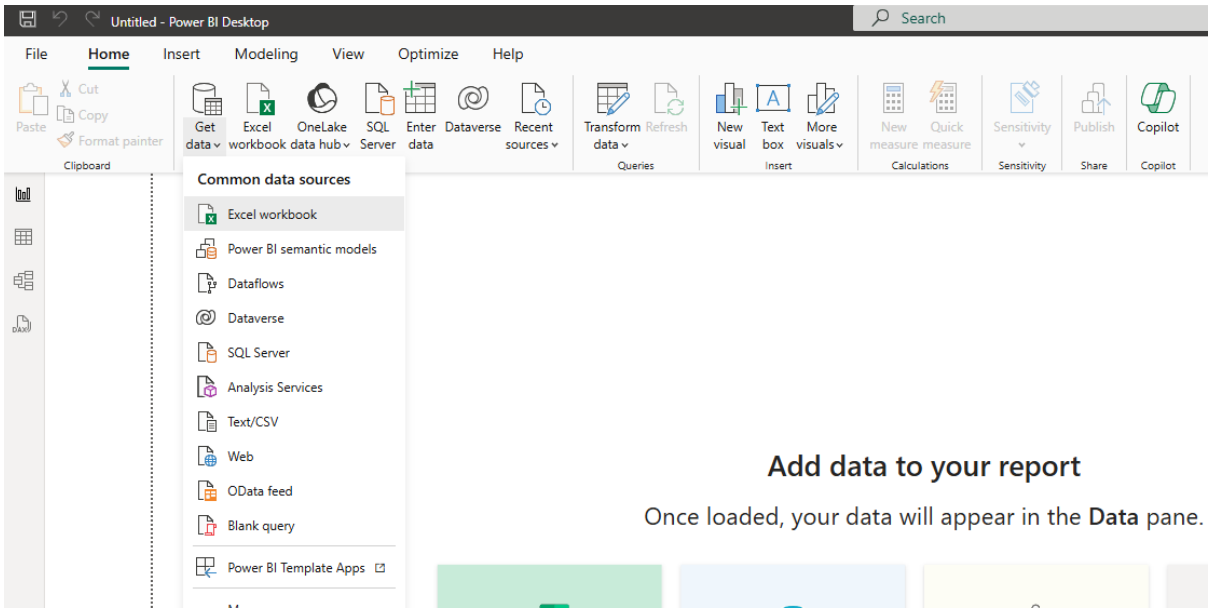
8. Perform data visualization using Python on any data.

```
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('data.csv')
df.plot(kind = 'scatter', x = 'Duration', y = 'Calories')
plt.show()
```

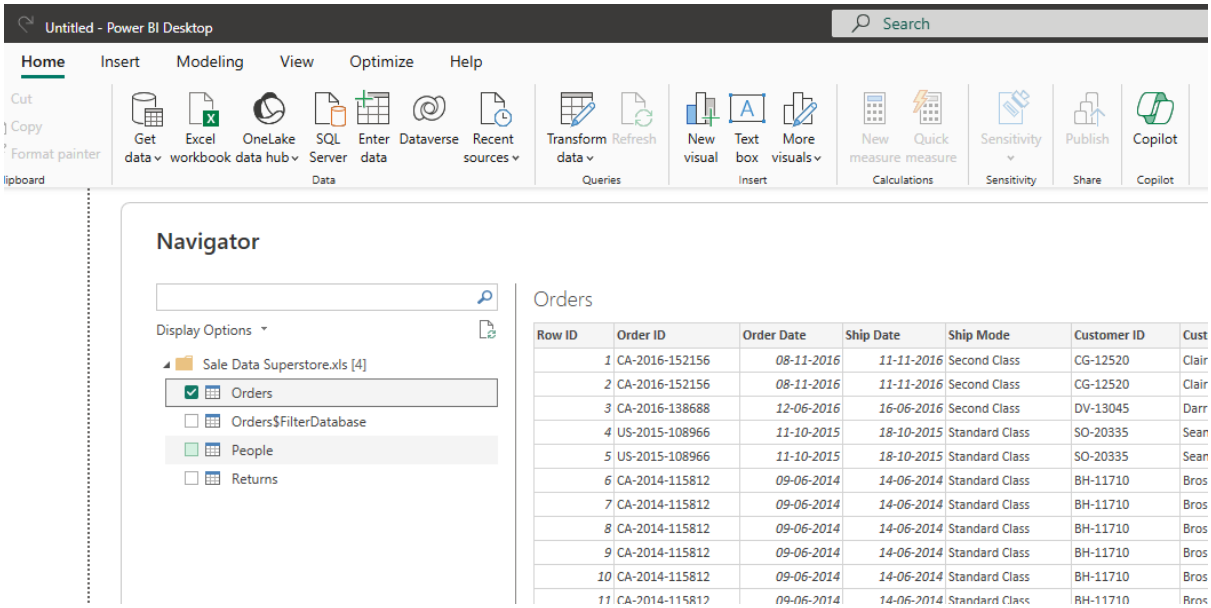


9. Perform data visualization using PowerBI on any sales data.

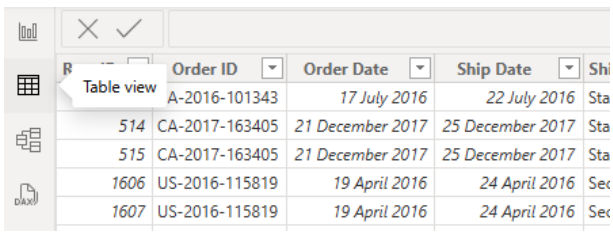
We import data from Excel to Power BI desktop. Please, follow the screenshots from the top menu choose “Get Data” -> Excel Workbook Open the file: Sale Data Superstore.xlsx



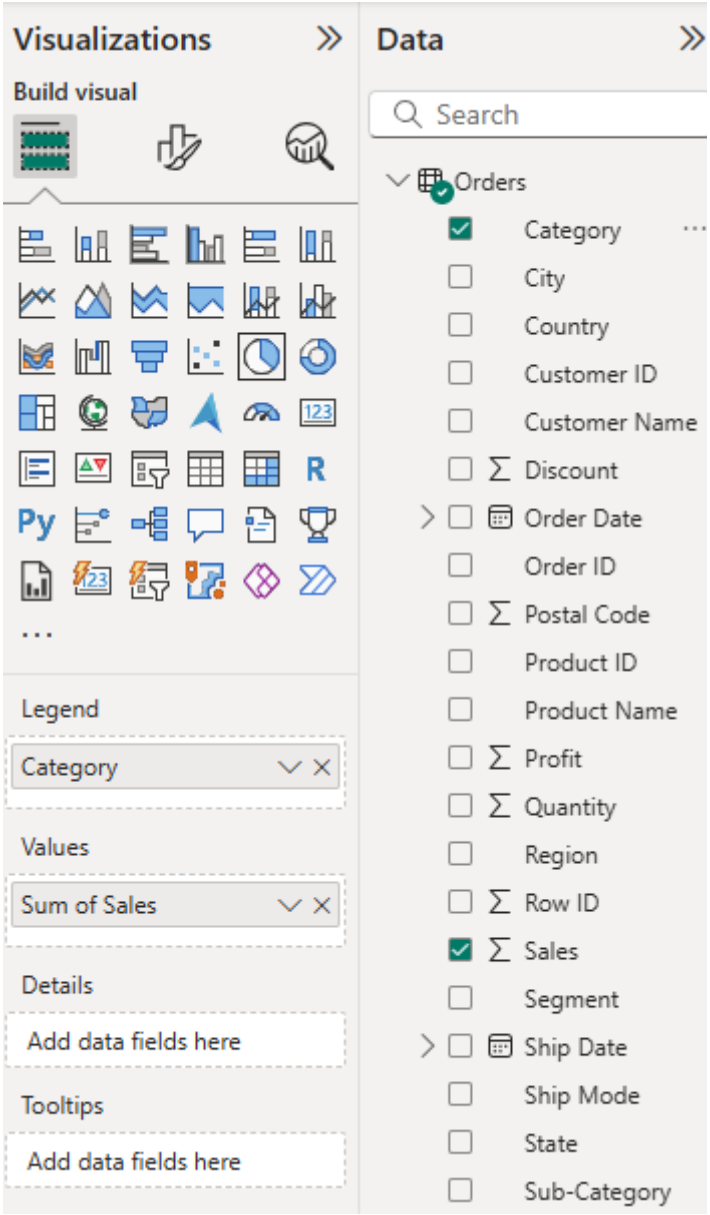
Choose “Order” worksheet (you will see the preview of the worksheet) and click “Load”.



To see your data imported choose the icon on the left side of the screen.



Click on report view and insert pie chart



Try to add a pie chart to your report. Use the right side of the screen (drag drop the fields)