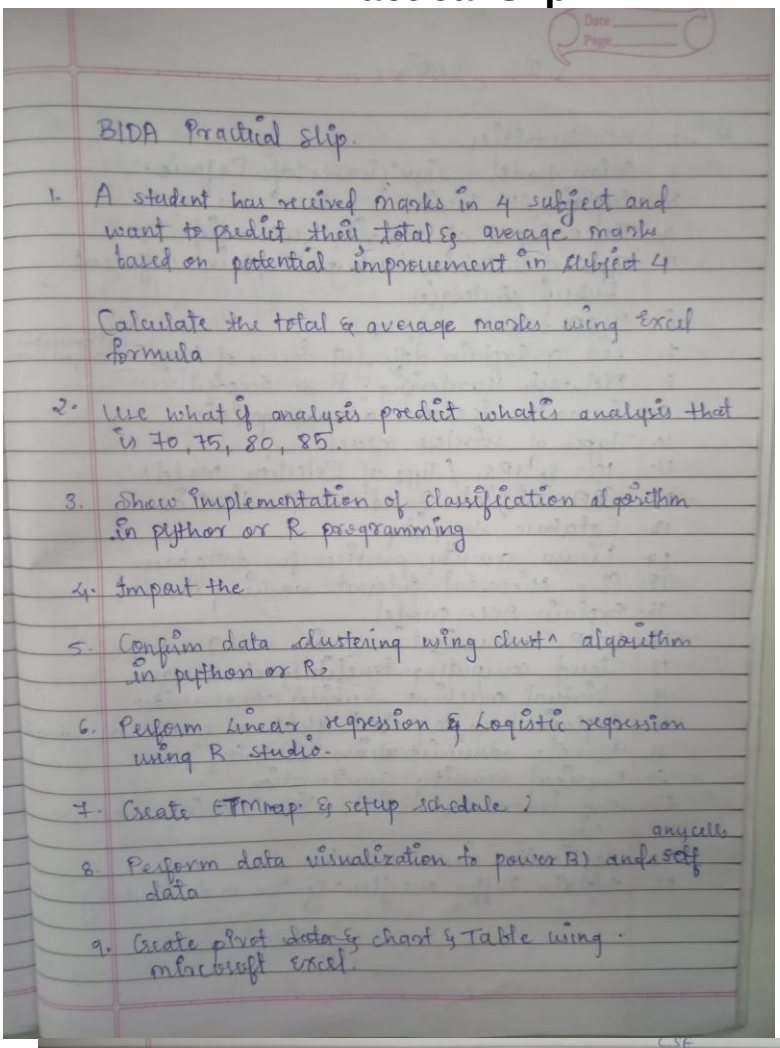
BIDA Practical Slip



BIDA Practical Slip.

1. A student has received marks in 4 subject and want to predict their total & average marks based on potential improvement in subject 4

Calculate the total & average marks using Excel formula

2. Use what if analysis predict what is analysis that is 70, 75, 80, 85.

3. Show implementation of classification algorithm in python or R programming

4. Impart the

5. Confirm data clustering using clust^ algorithm in python or R.

6. Perform Linear regression & Logistic regression using R studio.

7. Create ETLmap & setup schedule.

8. Perform data visualization to power BI and soft data

9. Create pivot data & chart & Table using Microsoft excel.

10. Write pivot programing in read data file in python

11. Implement Kmean clustering using R studio

VVIMP

1. Excel Pivot Table & Chart.
2. K means clustering
3. Linear regression
4. Decision Tree.
5. What if analysis scenario (Student marks analysis)

1. **A student has received marks in 4 subjects and wants to predict their total & average marks based on potential improvement in subject 4.**
   Calculate the total & average marks using Excel formula.

   Use what-if analysis to predict. What's the analysis that is 70, 75, 80, 85?

Ans:

A	B	C	D	E	F	G	H
S1	S2	S3	S4	Total marks	Formula		
				type the formula =sum(A3:D3)	type the formula =average(A3:D3)		
78	85	90	88	341	85.25		

DATA> WHAT IF ANALYSIS > SCENARIO MANAGER -> MODIFY A SCENARIO AND SELECT THE CELLS AND MAKE CHANGES -> SHOW SUMMARY

excel

2. Show implementation of classification algorithm in Python or R programming.

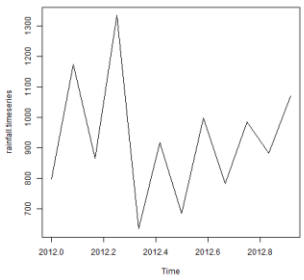
Ans:

Practical 6

```
# Get the data points in form of a R vector.
rainfall <-c(799,1174.8,865.1,1334.6,635.4,918.5,685.5,998.6,784.2,985,882.8,1071)

# Convert it to a time series object.
rainfall.timeseries <- ts(rainfall,start = c(2012,1),frequency = 12)

# Print the timeseries data.
print(rainfall.timeseries)

# Give the chart file a name.
png(file = "rainfall.png")

# Plot a graph of the time series.
plot(rainfall.timeseries)

# Save the file.
dev.off()
```

Output:
When we execute the above code, it produces the following result and chart −

```
Jan Feb Mar Apr May Jun Jul Aug Sep
2012 799.0 1174.8 865.1 1334.6 635.4 918.5 685.5 998.6 784.2
Oct Nov Dec
2012 985.0 882.8 1071.0
```



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

Ans:

Same as question 8(almost)

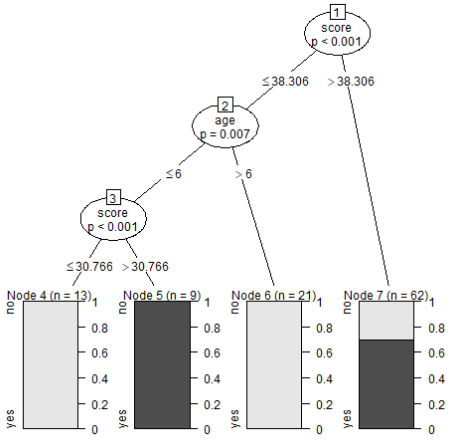
4. **Perform data clustering using cluster algorithms in Python or R.**

Ans:

Practical 7

First import package by click tools- install package- search “party” install

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



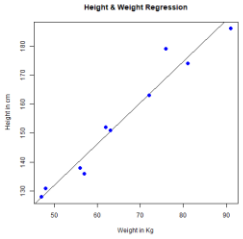
5. **Perform linear & logistic regression using R studio.**

Ans:

Practical 9 -linear

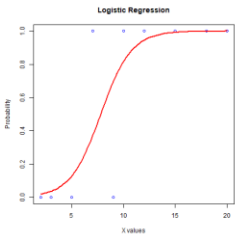
```
# Create the predictor and response variable.
x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)
```

```
y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)
relation <- lm(y~x)
# Give the chart file a name.
png(file = "linearregression.png")
# Plot the chart.
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")
# Save the file.
dev.off()
```



logistic regression

```
# Predictor (x) and response variable (y)
x <- c(2, 3, 5, 7, 9, 10, 12, 15, 18, 20)  # Example predictor
y <- c(0, 0, 0, 1, 0, 1, 1, 1, 1, 1)       # Binary response (0 or 1)

# Fit logistic regression model
model <- glm(y ~ x, family = binomial)

# Plot logistic regression curve
png(file = "logisticregression.png")
plot(x, y, col = "blue", main = "Logistic Regression", xlab = "X values", ylab = "Probability")
curve(predict(model, data.frame(x = x), type = "response"), add = TRUE, col = "red", lwd = 2)
dev.off()
```



6. Create ETL Map & setup schedule.

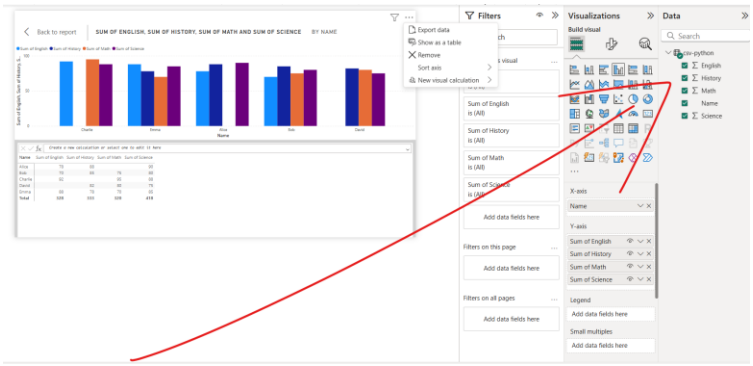
Ans: **nhi sikhaya aarya ne aur isko kuch SQL se krna hai**

7. Perform data visualization in Power BI and Sales data.

Ans:

application:B1

Import data from excel > select excel file and load data > select any chart and double click > select jo bhi chahiye > then on extreme right tick the data you want to display

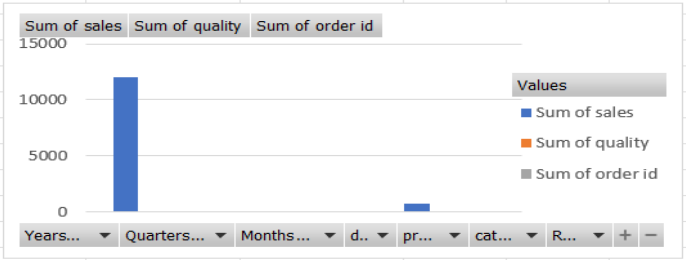


8. Create pivot table & chart & table using Microsoft Excel.

Ans:

order id	date	product	category	sales	quality	Region
101	01-03-2005	laptop	electronic	12005	2	north
102	01-03-2006	mobile	accessories	25	2	west
103	01-03-2007	mouse	electronic	800	1	east

Row Labels	Sum of sales	Sum of quality	Sum of order id
⊞2005	12005	2	101
⊞2006	25	2	102
⊞2007	800	1	103
Grand Total	12830	5	306



Sabse pehle upar wala table banao then usko select karo:

Go to the **"Insert"** tab and click **TABLE** -> **"PivotTable"**.

1. Choose where to place the PivotTable (New or Existing Worksheet). CLICK **EXISTING**
2. Click **OK**.
3. Drag fields into the **Rows, Columns, Values, and Filters** areas as needed. (eg order, category etc.)

2. Create a Pivot Chart

1. Click inside the PivotTable.
2. Go to **"Insert" > "PivotChart"**.
3. Select a table or range: click on any
4. Choose a chart type (e.g., Column, Pie, Line).
5. Click **OK** to generate the chart.

10. Write programming in read data CSV file in Python.

Ans:

Python IDLE

Kya pata CSV file milega ki Nhi

Path acche se copy karo (/ jo hai kabhi ulta copy hota hai toh usko seedha kr dena)

(ye csv file khud se bhi bana skte hai excel me same bas **.cvs extension** se save krna hai)

Name	Math	Science	English	History
Alice	85	90	78	88
Bob	75	80	70	85
Charlie	95	88	92	
David	80	75		82
Emma	70	85	88	78

Ye cmd me type karna hai dono cheez
Cmd 1: -m ensurepip --default-pip
Sabse pehle cmd2: “pip install pandas”

Now use **IDLE and paste the code**

```
import pandas as pd
data = pd.read_csv("C:/Users/dhari/Desktop/csv-python.csv")

print("First 5 rows of the data:")
print(data.head())

print("\nMissing values in each column:")
print(data.isnull().sum())

print("\nSummary statistics for numerical columns:")
print(data.describe())
```

Output: just shows the missing values from the table

11. Implement K-means clustering using R Studio.

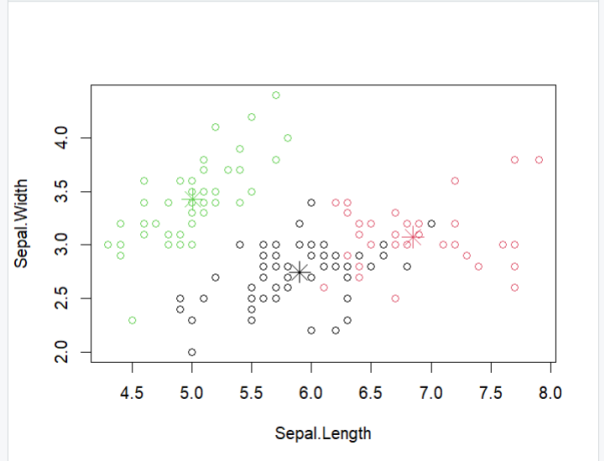
Ans:

RSTUDIO > FILE > NEW > RSCRIPT

Practical 8

```
newiris<-iris
newiris$Species<-NULL
(kc<- kmeans(newiris,3))
table(iris$Species,kc$cluster)
plot(newiris[c("Sepal.Length","Sepal.Width")],col=kc$cluster)
points(kc$centers[,c("Sepal.Length","Sepal.Width")],col=1:3,pch=8,cex=2)
```

Iska **output plot** mai hai



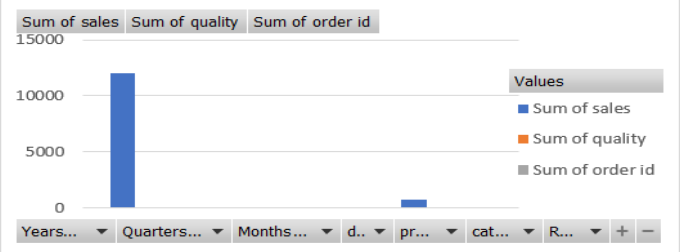
V.V. IMP

1. Excel Pivot Table & Chart.

Ans:

order id	date	product	category	sales	quality	Region
101	01-03-2005	laptop	electronic	12005	2	north
102	01-03-2006	mobile	accessories	25	2	west
103	01-03-2007	mouse	electronic	800	1	east

Row Labels	Sum of sales	Sum of quality	Sum of order id
⊞2005	12005	2	101
⊞2006	25	2	102
⊞2007	800	1	103
Grand Total	12830	5	306



Sabse pahale uper wala table banao then usko select karo:
Go to the **"Insert"** tab and click **"PivotTable"**.

4. Choose where to place the PivotTable (New or Existing Worksheet).
5. Click **OK**.
6. Drag fields into the **Rows, Columns, Values, and Filters** areas as needed.

2. Create a Pivot Chart

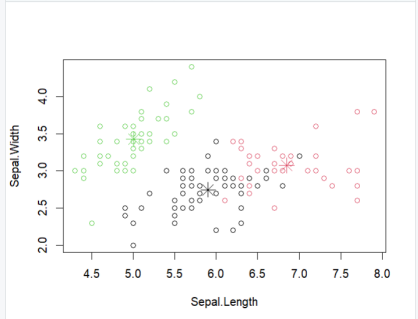
6. Click inside the PivotTable.
7. Go to **"Insert" > "PivotChart"**.
8. Choose a chart type (e.g., Column, Pie, Line).
9. Click **OK** to generate the chart.

2. K-means clustering.

Ans:

Practical 8

```
newiris<-iris
newiris$Species<-NULL
(kc<- kmeans(newiris,3))
table(iris$Species,kc$cluster)
plot(newiris[c("Sepal.Length","Sepal.Width")],col=kc$cluster)
points(kc$centers[,c("Sepal.Length","Sepal.Width")],col=1:3,pch=8,cex=2)
```

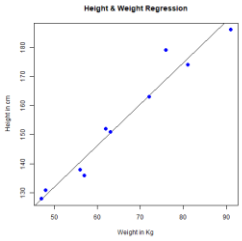


3. Linear regression.

Ans:

Practical 9

```
# Create the predictor and response variable.
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)
# Give the chart file a name.
png(file = "linearregression.png")
# Plot the chart.
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")
# Save the file.
dev.off()
```

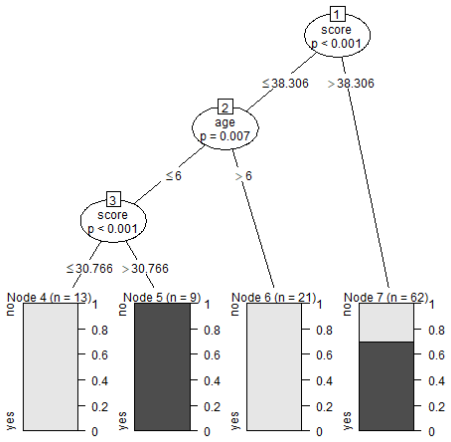


4. Decision Tree.

Ans:

Practical 7

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



5. What-If analysis scenario (Student marks analysis).

Ans:

Refer Q1

Viva Questions:

Excel & Data Analysis

1. **What is What-If Analysis in Excel?**
   → It helps predict outcomes by changing input values to see their effects on formulas.

2. **How do you use Scenario Manager in Excel?**
   → Go to "Data" > "What-If Analysis" > "Scenario Manager," modify scenarios, and generate a summary.

3. **What is a Pivot Table?**
   → A Pivot Table summarizes large datasets by organizing and analyzing data dynamically.

4. **How do you create a Pivot Chart in Excel?**
   → Select a Pivot Table > Go to "Insert" > Click "PivotChart" > Choose chart type > Click OK.

5. **How do you import data from a warehouse in Excel?**
   → Use "Data" > "Get External Data" > Choose source > Load data into Excel.

Programming & Data Science

How do you read a CSV file in Python?
→ Using `pandas`:

python
CopyEdit

```
import pandas as pd
data = pd.read_csv("file_path.csv")
print(data.head())
```

6.
7. **What is K-Means Clustering?**
   → It groups similar data points into 'k' clusters based on distance from cluster centers.

8. **What is the purpose of the `kmeans()` function in R?**
   → It performs K-means clustering on a dataset to group similar observations.

9. **What is the difference between Linear and Logistic Regression?**
   → **Linear Regression** predicts continuous values; **Logistic Regression** predicts binary outcomes (0/1).

How do you perform Linear Regression in R?
→ Using the `lm()` function:

r
CopyEdit

```
relation <- lm(y ~ x)
plot(y, x, main="Regression Line", abline(relation))
```

10.
11. **What is Decision Tree Algorithm?**
    → A supervised learning algorithm used for classification by splitting data into branches based on conditions.

12. **What package is used for Decision Trees in R?**
    → The **party** package, with the `ctree()` function.

Power BI & Data Visualization

13. **What is Power BI used for?**
    → It is a business intelligence tool for data visualization, reporting, and analytics.

14. **How do you create a report in Power BI?**
    → Import data > Select visualization > Drag fields into visualization areas.

ETL & Data Warehousing

15. **What is ETL in data processing?**
    → **Extract** (get data), **Transform** (clean & modify), **Load** (store into

database/warehouse).

16. **How do you schedule an ETL process?**
    → Use SQL jobs, Python scripts with `schedule` library, or automation tools like SSIS.