BI practical

Perform data clustering using clustering algorithm in python or r studio.

Practical 8- k-means clustering

- 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)

practical9- linear regression

> 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)

> 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",ylan = "Height in cm")

> 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()

Now go to files and search for linearregression.png file open it and show to miss.

Decision tree (practical 7)

> install.packages("party")

```
> library(party)
> print(head(readingSkills))
> 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()
```

Search for decision_tree.png in files show output to miss

Implementation of classification algorithm in R programming.

```
> 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(2012,1),frequency = 12)
> print(rainfall.timeseries)
> png(file = "rainfall.png")
> plot(rainfall.timeseries)
> dev.off()
```

Go to files section and search for rainfall.png and show it to miss

Write a python program to read data from csv file

Step 1- open pycharm

Step2- go to file -> settings ->python interpreter download pandas, metapolib and seaborn and then click ok

Create a python file names data_analysis.py and then copy this code

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Read data from a CSV file (Ensure 'data.csv' exists in your project folder)
file_path = "data.csv"
df = pd.read_csv(file_path)

print("\nDataset Information:")
print(df.info())

# Show the first 5 rows
print("\nFirst 5 Rows:")
print(df.head())
```

```
print("\nSummary Statistics:")
print(df.describe())
print("\nUnique Value Counts in Categorical Columns:")
print(df.nunique())

print("\nMissing Values in Each Column:")
print(df.isnull().sum())

print("\nMost Frequent Values in Each Column:")
print(df.mode().iloc[0])

if 'Category' in df.columns and 'Price' in df.columns:
    category_avg_price = df.groupby('Category')['Price'].mean()
    print("\nAverage Price per Category:")
    print(category_avg_price)

if 'Price' in df.columns:
    print("\nTop 5 Most Expensive Items:")
    print(df.sort_values(by='Price', ascending=False).head())

if df.select_dtypes(include=['number']).shape[1] > 1:
    print("\nCorrelation Matrix:")

numeric_df = df.select_dtypes(include=['number'])  # Select only numeric columns

if not numeric_df.empty:  # Check if there are numeric columns
    print("\nCorrelation Matrix:")
    print(numeric_df.corr())  # Compute correlation only for numbers

    import matplotlib.pyplot as plt
    import seaborn as sns

    plt.figure(figsize=(8, 6))
    sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm", linewidths=0.5)
    plt.title("Correlation Matrix Heatmap")
    plt.show()
else:
    print("\nNo numerical data available for correlation analysis.")

    plt.figure(figsize=(8, 6))
    sns.heatmap(df.corr(), annot=True, cmap="coolwarm", linewidths=0.5)
    plt.title("Correlation Matrix Heatmap")
    plt.show()
```

now create a new file name data.csv
and add this code

Product,Category,Price,Rating,Stock
Book1,Fiction,9.99,4.2,50
Book2,Science,12.50,4.5,30
Book3,Fiction,15.99,3.9,20
Book4,History,18.75,4.1,15
Book5,Science,22.99,4.8,10
Now right click on data_analysis file and run the file show pink vala box ka photo to miss

A student has received marks in 4 subjects and total of average marks based on potential improvement in subject 4 .claculate the total marks and average marks using excel formula

Step1- open excel add 4 subject marks and name of the student create total and average marks column and apply the formula to it
Use sum and avg formula from top right corner

Apply what-if analysis for the above improvement in subject 4

Goto excel
Add subject names and marks
Use excel sum and avg formula
For improvement in sub4 follow

1. **goto data**
2. **what-if**
3. **scenario manager**
4. **add a scenario (improvement in sub4)**
5. **changing cell (put subject 4 ka cell)**
6. **ok**
7. **enter value for changing cells**
8. **ok**
9. **show (the value will get updated)**

Data Modelling and Analytics with Pivot Table and pivot chart in Excel

Open excel enter this data

Student Name	Subject	Marks	Semester
John Doe	Math	85	Semester 1
John Doe	Science	78	Semester 1
John Doe	English	90	Semester 1
Jane Smith	Math	88	Semester 1
Jane Smith	Science	80	Semester 1
Jane Smith	English	92	Semester 1
Alex Brown	Math	75	Semester 2
Alex Brown	Science	82	Semester 2
Alex Brown	English	89	Semester 2
Emily White	Math	91	Semester 2
Emily White	Science	84	Semester 2
Emily White	English	95	Semester 2

Step2- select the entire table
Step3- go to insert tavb-> click pivotable
Step4- create pivotable window:

- Ensure the range is correct.
- Select **New Worksheet** (or Existing Worksheet where you want the Pivot Table).
- Click **OK**.

After inserting the PivotTable, a **PivotTable Fields** pane appears on the right.

- **Drag "Student Name" to the Rows section.**
- **Drag "Subject" to the Columns section.**

- **Drag "Marks" to the Values section** (It will automatically sum the marks).
- **Drag "Semester" to the Filters section.**

Create ETL map and setup schedule for exhibition.
And
Data visualization using power BI
Refer manual practical 2 and 3
Okay bye 😉