EX. No. 1(a)	Numerical Operations (MAX, MIN, AVG, SUM, SQRT, ROUND)

AIM

To implement numerical operations using MS-EXCEL.

ALGORITHM

Step 1: Start Ms Excel application in Ms- office.

Step 2: Create datasheet for student marks in Ms Excel application.

Step 3: Calculate the Maximum of the given marks using max function.

Step 4: Calculate the Minimum of the given marks using MIN function.

Step 5: Calculate the average of the given marks using average function.

Step 6: Calculate the sum of the given marks using sum function.

Step 7: Calculate the square root of the given mark using SQRT function.

Step 8: Calculate the Round of the given mark using Roundup function.

Step 9: Display the desired output of all numerical operation in neat format.

Step 10: Save the excel file and Close the Ms Excel application.

OUTPUT

Numerical Operations
(MAX, MIN, AVG, SUM, SQRT, ROUND)

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and	OMD551 Basic of Biomedical Instrumentatio
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK.S	80	72	77	87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	AB	80	AB	80
		MAX MARKS	=MAX(D6:D11)					
		MIN MARKS						
		AVGERAGE MARKS						
		SUM OF THE MARKS						
		SQRT OF ANY						
		ROUND OF THE MARKS						

Numerical Operations

(MAX, MIN, AVG, SUM, SQRT, ROUND)

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and	OMD551 Basic of Biomedical Instrumentatio
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK.S	80	72	77	87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	AB	80	AB	80
		MAX MARKS	92	92	87	90	87	87
		MIN MARKS	=MIN(D6:D11)					
		AVGERAGE MARKS						
		SUM OF THE MARKS						
		SQRT OF ANY						
		ROUND OF THE MARKS						

Numerical Operations

(MAX, MIN, AVG, SUM, SQRT, ROUND)

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and	OMD551 Basic of Biomedical Instrumentatio
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK.S	80	72	77	87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	AB	80	AB	80
		MAX MARKS	92	92	87	90	87	87
		MIN MARKS	34	72	77	80	75	80
		AVGERAGE MARKS	=AVERAGE(D6:D11)					
		SUM OF THE MARKS						
		SQRT OF ANY						
		ROUND OF THE MARKS						

Numerical Operations

(MAX, MIN, AVG, SUM, SQRT, ROUND)

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and	OMD551 Basic of Biomedical Instrumentatio
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK.S	80	72	77	87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	AB	80	AB	80
		MAX MARKS	92	92	87	90	87	87
		MIN MARKS	34	72	77	80	75	80
		AVGERAGE MARKS	74	83	82.2	84	81.2	83.5
		SUM OF THE MARKS	=SUM(D6:D11)					
		SQRT OF ANY						
		ROUND OF THE MARKS						

Numerical Operations

(MAX, MIN, AVG, SUM, SQRT, ROUND)

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and	OMD551 Basic of Biomedical Instrumentatio
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK.S	80	72	77	87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	AB	80	AB	80
		MAX MARKS	92	92	87	90	87	87
		MIN MARKS	34	72	77	80	75	80
		AVGERAGE MARKS	74	83	82.2	84	81.2	83.5
		SUM OF THE MARKS	444	498	411	504	406	501
		SQRT OF ANY	=SQRT(D6)					
		ROUND OF THE MARKS						

Numerical Operations

(MAX, MIN, AVG, SUM, SQRT, ROUND)

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and	OMD551 Basic of Biomedical Instrumentatio
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK.S	80	72	77	87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	AB	80	AB	80
		MAX MARKS	92	92	87	90	87	87
		MIN MARKS	34	72	77	80	75	80
		AVGERAGE MARKS	74	83	82.2	84	81.2	83.5
		SUM OF THE MARKS	444	498	411	504	406	501
		SQRT OF ANY	9.591663047	9.32737905	8.94427	9.32738	9.16515	9.32738
		ROUND OF THE MARKS	=ROUNDUP(D16,2)					

Numerical Operations

(MAX, MIN, AVG, SUM, SQRT, ROUND)

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and	OMD551 Basic of Biomedical Instrumentatio
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK.S	80	72	77	87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	AB	80	AB	80
		MAX MARKS	92	92	87	90	87	87
		MIN MARKS	34	72	77	80	75	80
		AVGERAGE MARKS	74	83	82.2	84	81.2	83.5
		SUM OF THE MARKS	444	498	411	504	406	501
		SQRT OF ANY	9.591663047	9.32737905	8.94427	9.32738	9.16515	9.32738
		ROUND OF THE MARKS	9.6	9.33	8.95	9.33	9.17	9.33

RESULT

The numerical operations were implemented using MS-EXCEL successfully and the desired output was displayed.

Ex. No. 1 (b)	Perform data import/export operations for different file formats

AIM

To perform data import/export operations for different file formats using MS-EXCEL.

ALGORITHM

Step 1 : Start Ms Excel application in Ms- office.

Step 2 : Create datasheet for student marks in Ms Excel application.

Step 3 : Save the excel file.

Step 4 : Export the file into CSV file using file menu and export option.

Step 5: Next , import CSV file using data menu and get data option.

Step 6 : Display the desired output in neat format.

Step 7 : Save the excel file and Close the Ms Excel application.

PROCEDURE

Data Import/Export Operations for Different File Formats

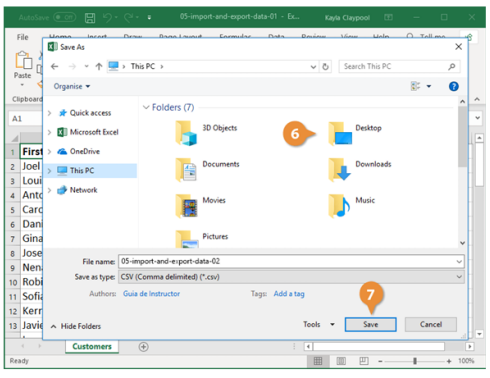
Excel can import and export many different file types aside from the standard .xslx format. If your data is shared between other programs, like a database, you may need to save data as a different file type or bring in files of a different file type.

EXPORT DATA

When you have data that needs to be transferred to another system, export it from Excel in a format that can be interpreted by other programs, such as a text or CSV file.

1. Click the File tab.
2. At the left, click Export.
3. Click the Change File Type.
4. Under Other File Types, select a file type.
    a. Text (Tab delimited): The cell data will be separated by a tab.
    b. CSV (Comma delimited): The cell data will be separated by a comma.
    c. Formatted Text (space delimited): The cell data will be separated by a space.
    d. Save as Another File Type: Select a different file type when the Save As dialog box appears. The file type you select will depend on what type of file is required by the program that will consume the exported data.
5. Click Save As.
6. Specify where you want to save the file.
7. Click Save. A dialog box appears stating that some of the workbook features may be lost.
8. Click Yes.

OUTPUT

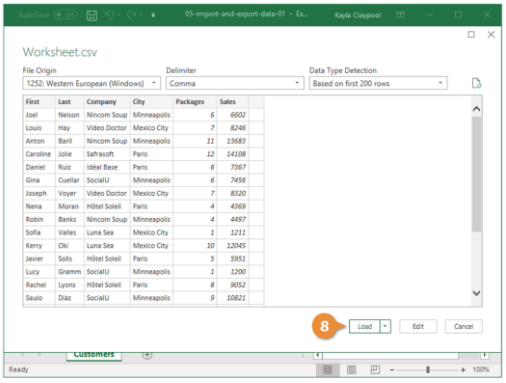


IMPORT DATA

Excel can import data from external data sources including other files, databases, or web pages.

1. Click the Data tab on the Ribbon.
2. Click the Get Data button. Some data sources may require special security access, and the connection process can often be very complex. Enlist the help of your organization's technical support staff for assistance.
3. Select From File.
4. Select From Text/CSV. If you have data to import from Access, the web, or another source, select one of those options in the Get External Data group instead.
5. Select the file you want to import.
6. Click Import. If, while importing external data, a security notice appears saying that it is connecting to an external source that may not be safe, click OK.
7. Verify the preview looks correct. Because we've specified the data is separated by commas, the delimiter is already set. If you need to change it, it can be done from this menu.
8. Click Load.

OUTPUT



RESULT

The data import/export operations for different file formats were preformed successfully using MS-EXCEL.

Ex. No. 2	PERFORM STATISTICAL OPERATIONS [Mean, Median, Mode and Standard Deviation, Variance, Skewness, Kurtosis]

AIM

To Perform statistical operations using MS-EXCEL.

ALGORITHM

Step 1 : Start Ms Excel application in Ms- office.

Step 2 : Create datasheet for student marks in Ms Excel application.

Step 3 : If you haven't already installed the Analysis ToolPak , Click the Microsoft Office button, then click on the Excel Options , and then select Add-Ins , Click Go, check the Analysis ToolPak box, and click Ok

Step 4 : Select Data tab, then click on the Data Analysis option, then selects Descriptive Statistics from the list and Click Ok. [Data tab >> Data Analysis >> Descriptive Statistics]

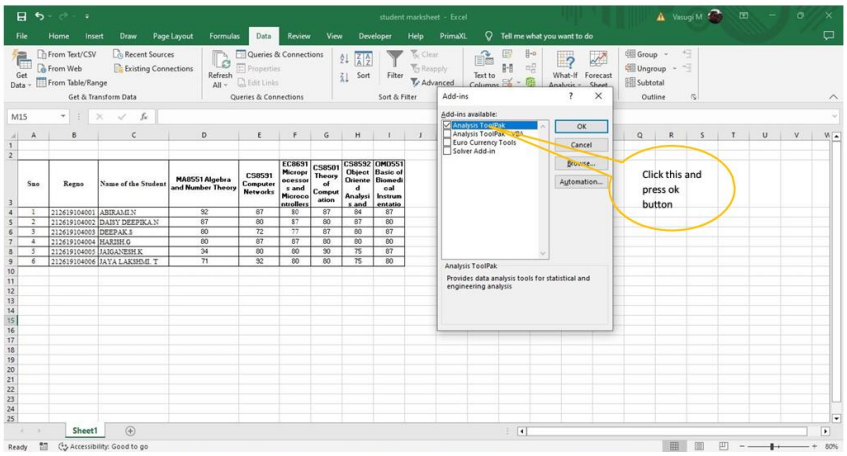
Step 5: In the Input Range we select the data, and then select Output Range where you want the output to be stored. If you don’t specify the output range it will throw output in the new worksheet.

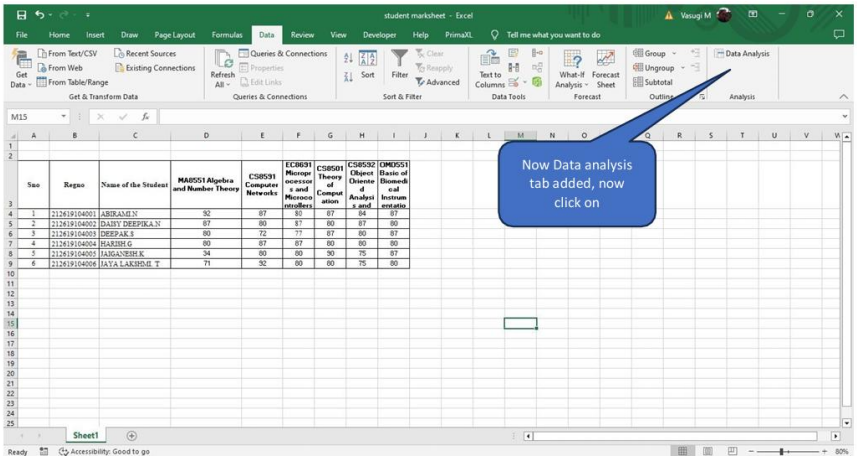
Step 6 : Check Summary Statistics and Confidence Level for Mean options. By default the confidence level is 95%. You can change the level as per the hypothesis standard of study.

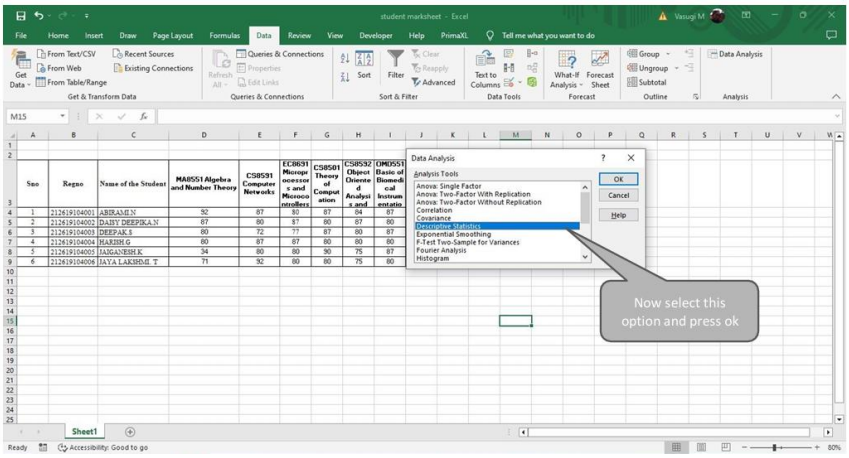
Step 7 : When you click Ok, you will see the result in the selected output range.

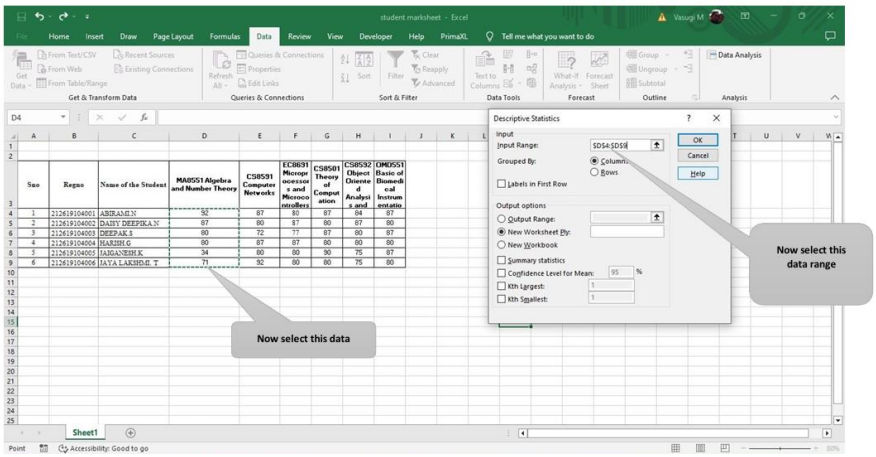
Step 8: Save the excel file and Close the Ms Excel application.

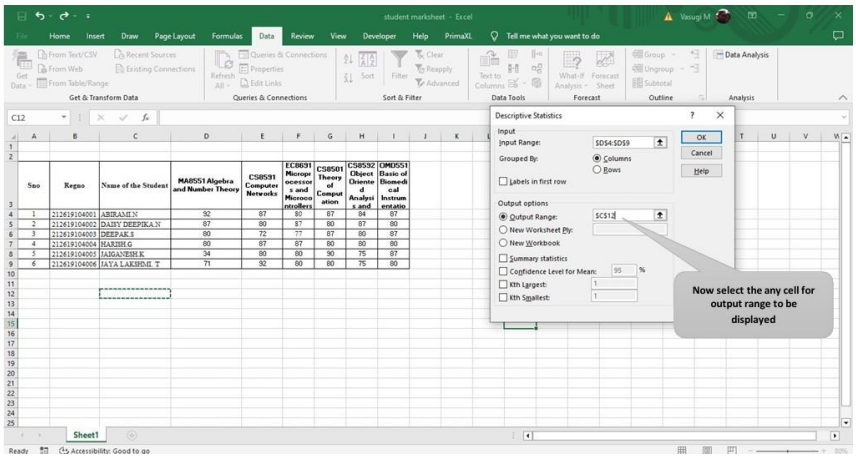
OUTPUT

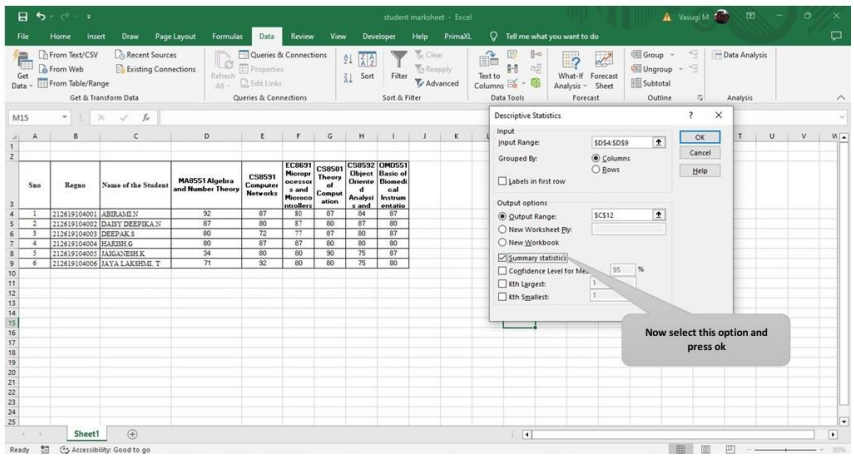





Data Analysis
Analysis Tools
Anova: Single Factor
Anova: Two-Factor With Replication
Anova: Two-Factor Without Replication
Correlation
Covariance
Descriptive Statistics
Exponential Smoothing
F-Test Two-Sample for Variances
Fourier Analysis
Histogram
OK
Cancel
Help
Now select this option and press ok


Descriptive Statistics
Input
Input Range:
$D$4:$D$9
Grouped By:
Columns
Rows
Labels in First Row
Output options
Output Range:
New Worksheet Ply:
New Workbook
Summary statistics
Confidence Level for Mean:
Kth Largest:
Kth Smallest:
OK
Cancel
Help
Now select this data range
Now select this data


Descriptive Statistics
Input
Input Range:
$D$4:$D$9
Grouped By:
Columns
Rows
Labels in first row
Output options
Output Range:
$C$12
New Worksheet Ply:
New Workbook
Summary statistics
Confidence Level for Mean:
Kth Largest:
Kth Smallest:
OK
Cancel
Help
Now select the any cell for output range to be displayed


Descriptive Statistics
Input
Input Range:
$D$4:$D$9
Grouped By:
Columns
Rows
Labels in first row
Output options
Output Range:
$C$12
New Worksheet Ply:
New Workbook
Summary statistics
Confidence Level for Mean:
Kth Largest:
Kth Smallest:
OK
Cancel
Help
Now select this option and press ok

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and Design	OMD551 Basic of Biomedical Instrumentation
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK.S	80	72	77	87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI.T	71	92	80	80	75	80

Column1	
Mean	74
Standard Error	8.512735557
Median	80
Mode	80
Standard Deviation	20.85185843
Sample Variance	434.8
Kurtosis	3.733266953
Skewness	-1.838637384
Range	58
Minimum	34
Maximum	92
Sum	444
Count	6

RESULT

The statistical operations were performed successfully using MS-EXCEL and the desired output was displayed in neat format.

Ex. No 3	Perform Z-test, T-test & ANOVA

AIM

To Perform Z-test, T-test & ANOVA operations using MS-EXCEL.

ALGORITHM

Z-TEST

Step 1 : Start Ms Excel application in Ms- office.

Step 2 : Create datasheet for student marks in Ms Excel application.

Step 3 : If you haven't already installed the Analysis ToolPak , Click the Microsoft Office button, then click on the Excel Options , and then select Add-Ins , Click Go, check the Analysis ToolPak box, and click Ok

Step 4 : Select Data tab, then click on the Data Analysis option, then selects Descriptive Statistics from the list and Click Ok. [Data tab >> Data Analysis >>z-test two sample means]

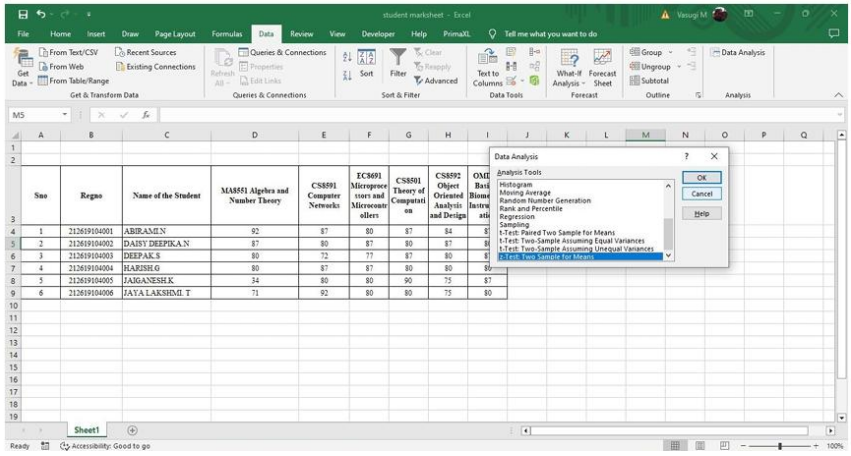
Step 5: In the Input Range we select range of the data for variable 1 and variable 2 and Give variable 1 and variable 2 value as 0.5. then select Output Range where you want the output to be stored. If you don't specify the output range it will throw output in the new worksheet.

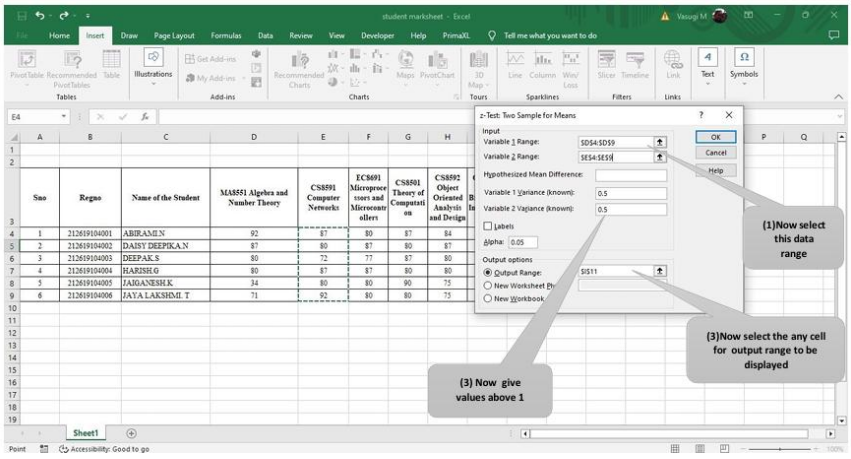
Step 6 : Then select Output Range where you want the output to be stored. If you don't specify the output range it will throw output in the new worksheet.

Step 7 : When you click Ok, you will see the result in the selected output range.

Step 8: Save the excel file and Close the Ms Excel application.

OUTPUT





Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and	OMD551 Basic of Biomedical Instrumentatio
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK.S	80	72	77	87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	80	80	75	80

z-Test: Two Sample for Means		
	Variable 1	*Variable 2*
Mean	74	83
Known Variance	0.5	0.5
Observations	6	6
Hypothesized Mean Di	0	
z	-22.045408	
P(Z<=z) one-tail	0	
z Critical one-tail	1.64485363	
P(Z<=z) two-tail	0	
z Critical two-tail	1.95996398	

T-TEST

Step 1 : Start Ms Excel application in Ms- office.

Step 2 : Create datasheet for student marks in Ms Excel application.

Step 3 : If you haven't already installed the Analysis ToolPak , Click the Microsoft Office button, then click on the Excel Options , and then select Add-Ins , Click Go, check the Analysis ToolPak box, and click Ok

Step 4 : Select Data tab, then click on the Data Analysis option, then selects Descriptive Statistics from the list and Click Ok. [Data tab >> Data Analysis >> T-test Paired two sample for means]

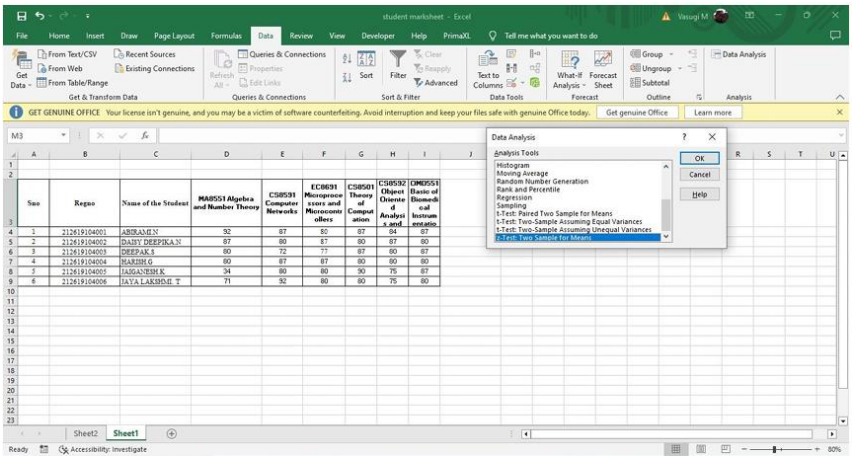
Step 5: In the Input Range we select range of the data for variable 1 and variable 2 and Give alpha value as 0.05. then select Output Range where you want the output to be stored. If you don’t specify the output range it will throw output in the new worksheet.

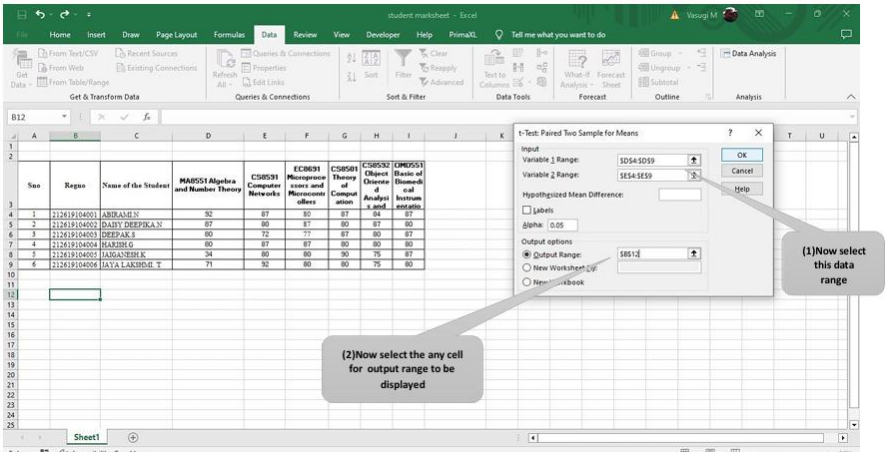
Step 6 : Then select Output Range where you want the output to be stored. If you don’t specify the output range it will throw output in the new worksheet.

Step 7 : When you click Ok, you will see the result in the selected output range.

Step 8: Save the excel file and Close the Ms Excel application.

OUTPUT





Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and	OMD551 Basic of Biomedical Instrumentatio
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK.S	80	72	77	87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	80	80	75	80

t-Test: Paired Two Sample for Means		
	Variable 1	*Variable 2*
Mean	74	83
Variance	434.8	50.4
Observations	6	6
Pearson Correlation	0.113487818	
Hypothesized Mean	0	
df	5	
t Stat	-1.037387876	
P(T<=t) one-tail	0.173548244	
t Critical one-tail	2.015048373	
P(T<=t) two-tail	0.347096488	
t Critical two-tail	2.570581836	

ANOVA TEST

Step 1 : Start Ms Excel application in Ms- office.

Step 2 : Create datasheet for student marks in Ms Excel application.

Step 3 : If you haven't already installed the Analysis ToolPak , Click the Microsoft Office button, then click on the Excel Options , and then select Add-Ins , Click Go, check the Analysis ToolPak box, and click Ok

Step 4 : Select Data tab, then click on the Data Analysis option, then selects Descriptive Statistics from the list and Click Ok. [Data tab >> Data Analysis >> Anova : Single factor]

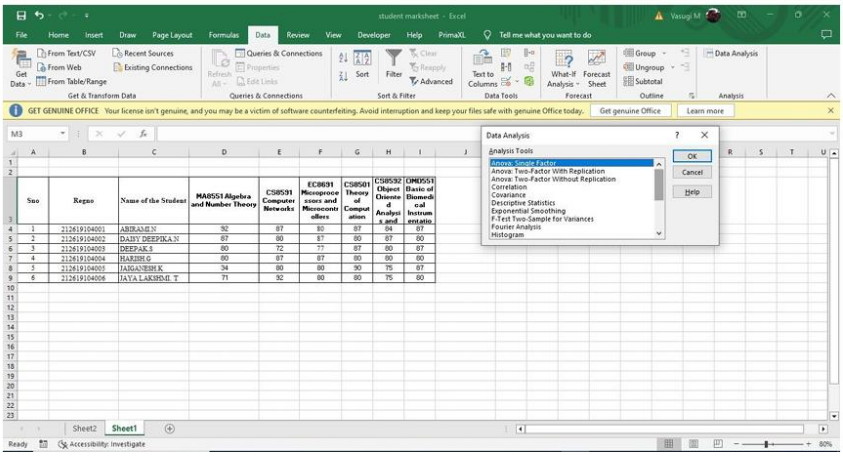
Step 5: In the Input Range we select range of the data and Give alpha value as 0.05. then select Output Range where you want the output to be stored. If you don’t specify the output range it will throw output in the new worksheet.

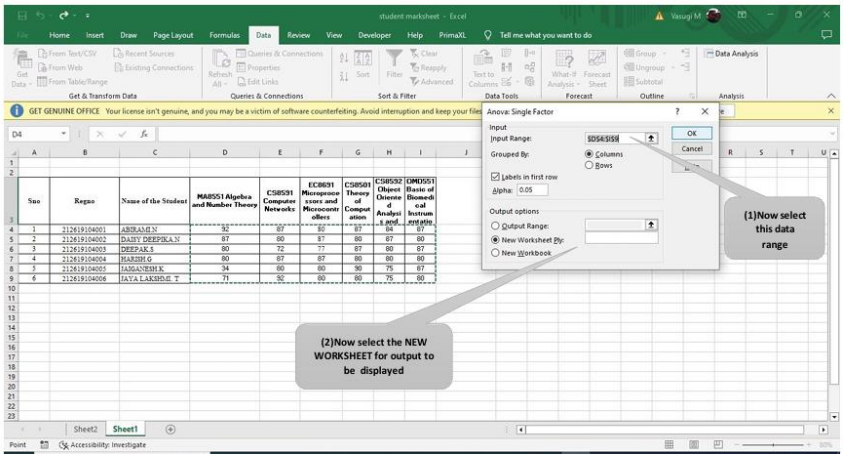
Step 6 : Then select Output Range where you want the output to be stored. If you don’t specify the output range it will throw output in the new worksheet.

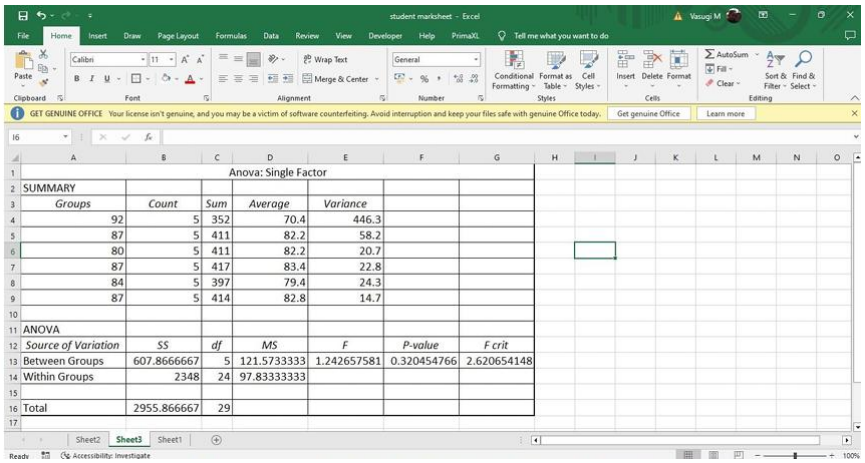
Step 7 : When you click Ok, you will see the result in the selected output range.

Step 8: Save the excel file and Close the Ms Excel application.

OUTPUT







Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
92	5	352	70.4	446.3
87	5	411	82.2	58.2
80	5	411	82.2	20.7
87	5	417	83.4	22.8
84	5	397	79.4	24.3
87	5	414	82.8	14.7

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	607.8666667	5	121.5733333	1.242657581	0.320454766	2.620654148
Within Groups	2348	24	97.83333333			
Total	2955.866667	29				

RESULT

The Z-test, T-test and ANOVA operations was performed successfully using MS-EXCEL and the desired output was displayed in neat format.

Ex. No 4 (a)	Perform data pre-processing operations - Handling Missing data

AIM:

To handle the missing data in data pre-processing operations on the dataset using MS-EXCEL.

ALGORITHM:

Step 1 : Start Ms Excel application in Ms- office.

Step 2 : Create datasheet for student marks in Ms Excel application.

Step 3 : If you haven't already installed the PrimaXL Addin, install it. Click the PrimaXL tab , choose missing

Step 4 : In the Input Range we select marks of all subjects with missing values and select the Choice as “filling of the missing data by taking average” or ” filling of the missing data by random pick”.

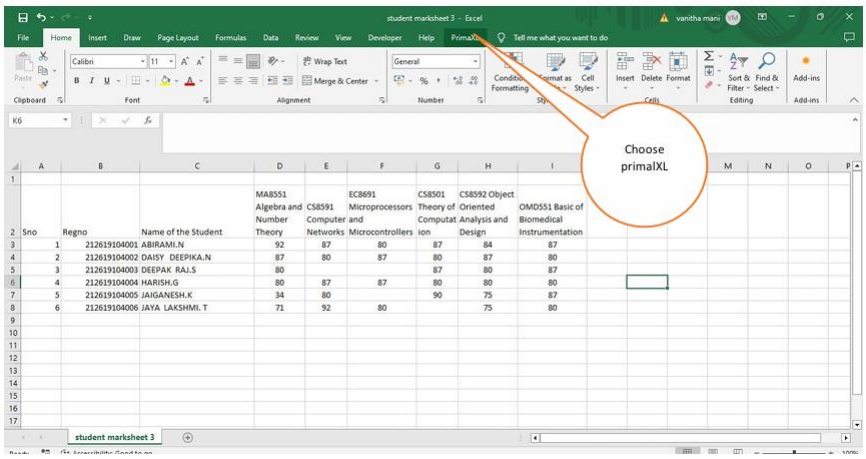
Step 5: Then select Output Range where you want the output to be stored. If you don’t specify the output range it will throw output in the new worksheet.

Step 6 : Then select Output Range where you want the output to be stored. If you don’t specify the output range it will throw output in the new worksheet.

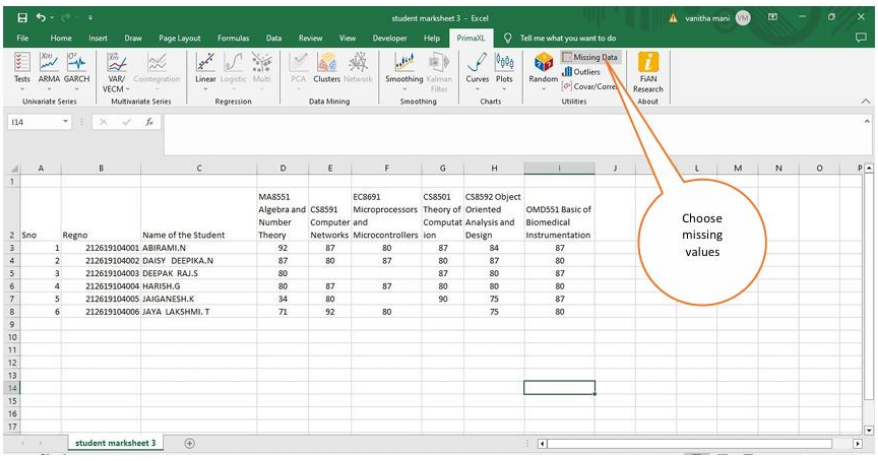
Step 7 : When you click Ok, you will see the result in the selected output range.

Step 8: Save the excel file and Close the Ms Excel application.

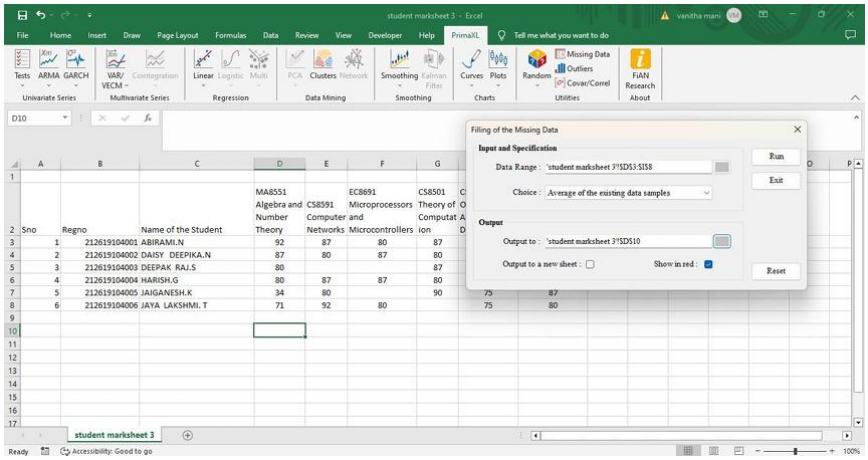
OUTPUT

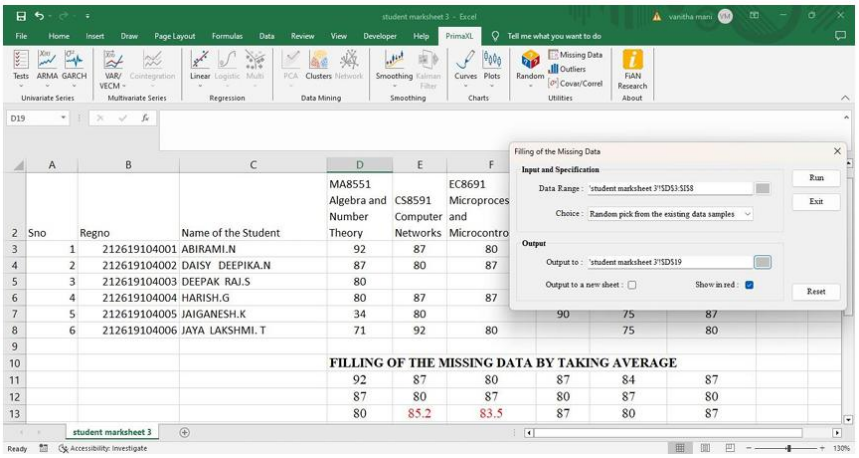


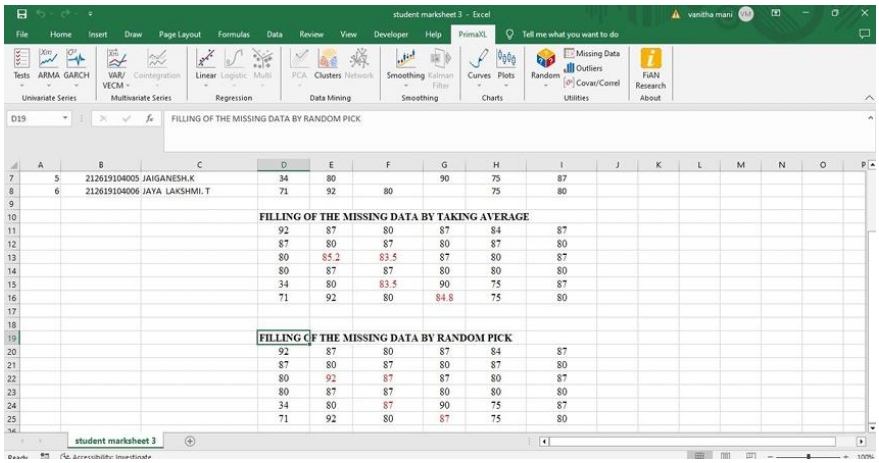
Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and Design	OMD551 Basic of Biomedical Instrumentation
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK RAJ.S	80			87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80		90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	80		75	80



Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theory of Computation	CS8592 Object Oriented Analysis and Design	OMD551 Basic of Biomedical Instrumentation
1	212619104001	ABIRAMI.N	92	87	80	87	84	87
2	212619104002	DAISY DEEPIKA.N	87	80	87	80	87	80
3	212619104003	DEEPAK RAJ.S	80			87	80	87
4	212619104004	HARISH.G	80	87	87	80	80	80
5	212619104005	JAIGANESH.K	34	80		90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	80		75	80







RESULT

The missing data on dataset was handled successfully using MS-EXCEL and the desired output was displayed in neat format.

Ex. No. 4(b)	Perform data pre-processing operations - Normalization

AIM: To normalize in the given dataset using MS-EXCEL.

Normalization (Or Min-Max scaling) data in excel It is the process of scaling data in such a way that all data points lie in a range of 0 to 1. Thus, this technique, makes it possible to bring all data points to a common scale. The mathematical formula for normalization is given as:

$$X' = \frac{X - Xmin}{Xmax - Xmin}$$

where X is the data point, Xmax and Xmin are the maximum and minimum value in the group of records respectively. The process of normalization is generally used when the distribution of data does not follow the Gaussian distribution.

PROCEDURE:

Step 1 : Start Ms Excel application in Ms- office.

Step 2 : Create datasheet for sales data in Ms Excel application.

Step 3 : Find maximum and minimum values of given data set.

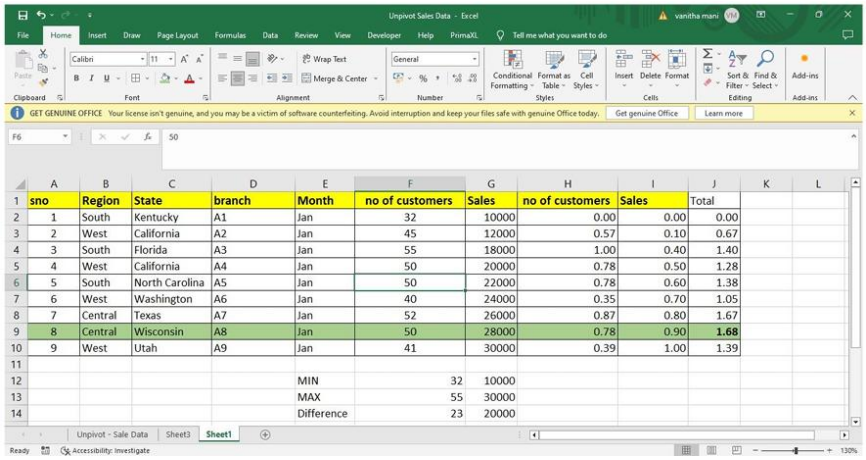
Step 4 : Calculate the difference between maximum and minimum values

Step 5: Apply the normalization formula using maximum value, minimum value and difference value. Step 6 : Find the best value of the normalized data.

Step 7 : Display the normalized data in desired format .

Step 8: Save the excel file and Close the Ms Excel application.

OUTPUT



	A	B	C	D	E	F	G	H	I	J
1	sno	Region	State	branch	Month	no of customers	Sales	no of customers	Sales	Total
2	1	South	Kentucky	A1	Jan	32	10000	0.00	0.00	0.00
3	2	West	California	A2	Jan	45	12000	0.57	0.10	0.67
4	3	South	Florida	A3	Jan	55	18000	1.00	0.40	1.40
5	4	West	California	A4	Jan	50	20000	0.78	0.50	1.28
6	5	South	North Carolina	A5	Jan	50	22000	0.78	0.60	1.38
7	6	West	Washington	A6	Jan	40	24000	0.35	0.70	1.05
8	7	Central	Texas	A7	Jan	52	26000	0.87	0.80	1.67
9	8	Central	Wisconsin	A8	Jan	50	28000	0.78	0.90	1.68
10	9	West	Utah	A9	Jan	41	30000	0.39	1.00	1.39
11										
12					MIN	32	10000			
13					MAX	55	30000			
14					Difference	23	20000			

RESULT

The given dataset was normalized using MS-EXCEL and the desired output was displayed in neat format.

EXP.NO.: 05	PERFORM DIMENSIONALITY REDUCTION OPERATION USING PCA, KPCA & SV

AIM

To perform dimensionality reduction using Principal Component Analysis (PCA), Kernel PCA (KPCA), and Singular Value Decomposition (SVD) on the *mtcars* dataset.

TOOLS REQUIRED

- RStudio / R environment
- stats package (for PCA & SVD) – comes built-in with R
- kernlab package (for Kernel PCA)
- ggplot2 package (for visualization)
- ggfortify package (for enhanced PCA biplot)

ALGORITHM

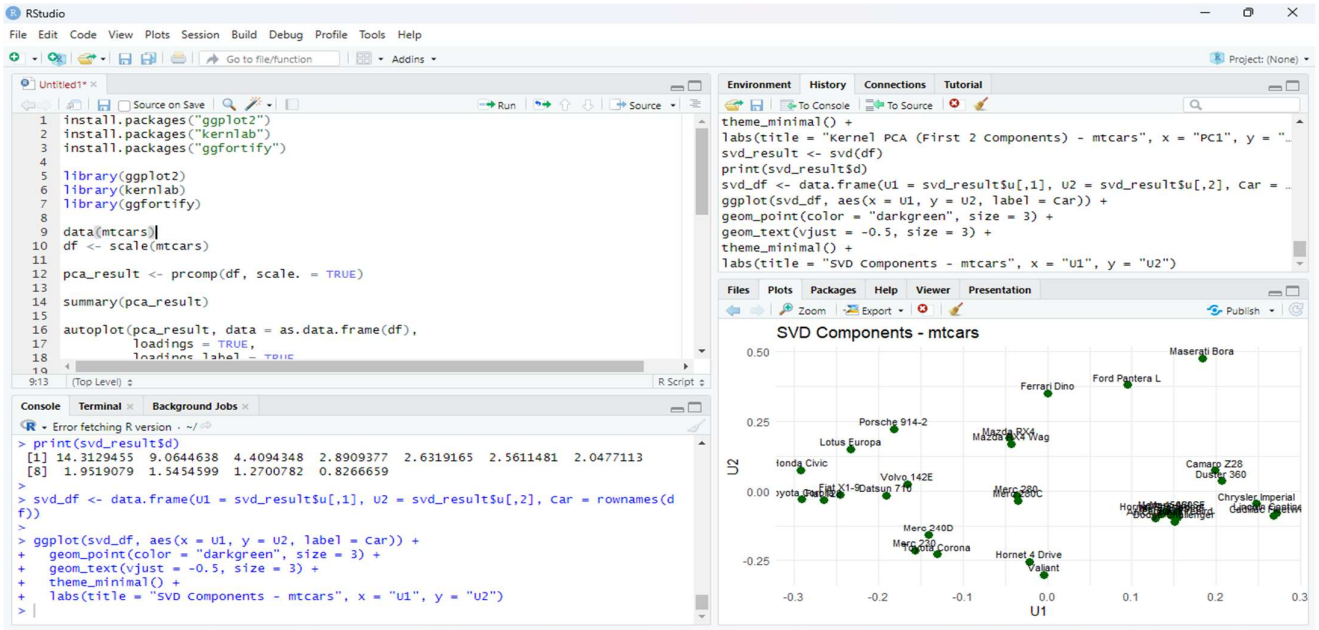
- Install and load required packages (stats, kernlab, ggplot2, ggfortify).
- Load the mtcars dataset and standardize it using scale().
- Apply PCA using prcomp() and examine variance using summary().
- Visualize PCA using autoplot() with variable loadings labeled.
- Apply Kernel PCA (kpca()) with RBF kernel and visualize first two components using ggplot2.
- Perform Singular Value Decomposition (svd()) and plot the first two left singular vectors using ggplot2.
- Compare and interpret results from PCA, Kernel PCA, and SVD.

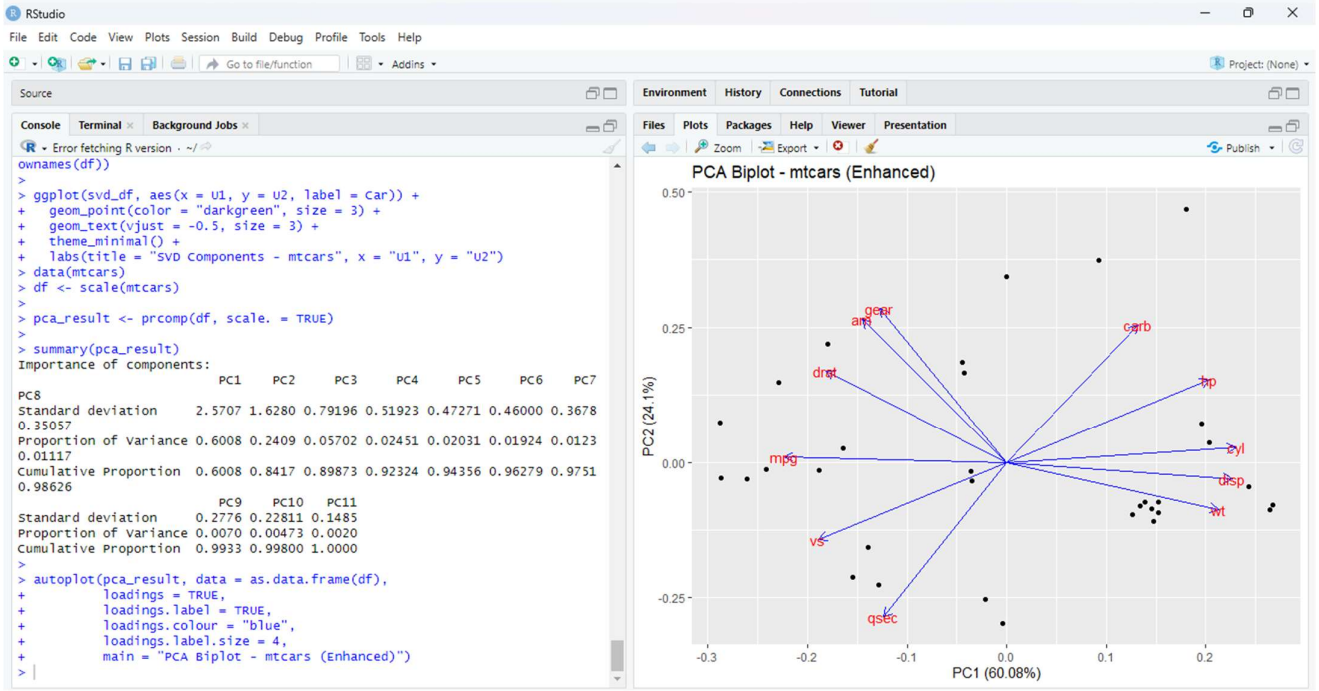
CODING

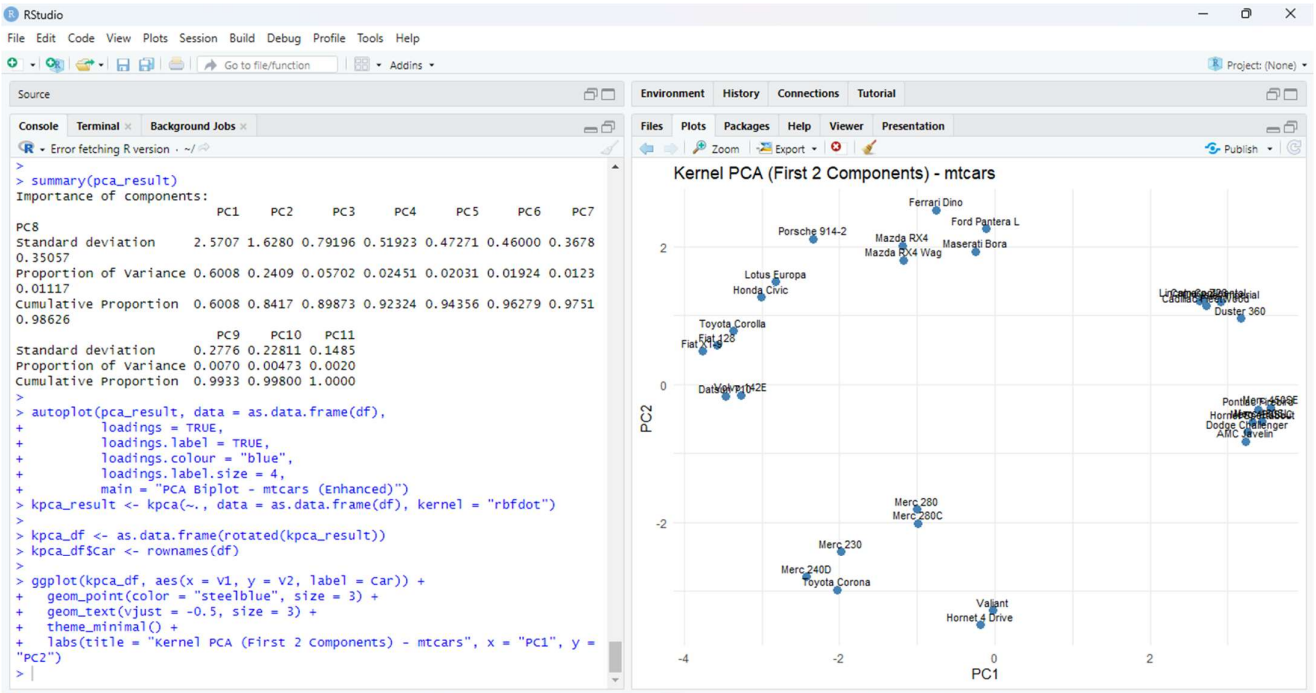
```
install.packages("ggplot2")
install.packages("kernlab")
install.packages("ggfortify")


library(ggplot2)
library(kernlab)
library(ggfortify)


data(mtcars)
df <- scale(mtcars)


pca_result <- prcomp(df, scale. = TRUE)
```

```
summary(pca_result)

autoplot(pca_result, data = as.data.frame(df),
         loadings = TRUE,
         loadings.label = TRUE,
         loadings.colour = "blue",
         loadings.label.size = 4,
         main = "PCA Biplot - mtcars (Enhanced)")

kpca_result <- kpca(~., data = as.data.frame(df), kernel = "rbfdot")

kpca_df <- as.data.frame(rotated(kpca_result))
kpca_df$Car <- rownames(df)

ggplot(kpca_df, aes(x = V1, y = V2, label = Car)) +
  geom_point(color = "steelblue", size = 3) +
  geom_text(vjust = -0.5, size = 3) +
  theme_minimal() +
  labs(title = "Kernel PCA (First 2 Components) - mtcars", x = "PC1", y = "PC2")

svd_result <- svd(df)
print(svd_result$d)

svd_df <- data.frame(U1 = svd_result$u[,1], U2 = svd_result$u[,2], Car = rownames(df))

ggplot(svd_df, aes(x = U1, y = U2, label = Car)) +
  geom_point(color = "darkgreen", size = 3) +
  geom_text(vjust = -0.5, size = 3) +
  theme_minimal() +
  labs(title = "SVD Components - mtcars", x = "U1", y = "U2")
```

OUTPUT


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Untitled1*
1 install.packages("ggplot2")
2 install.packages("kernlab")
3 install.packages("ggfortify")
5 library(ggplot2)
6 library(kernlab)
7 library(ggfortify)
9 data(mtcars)
10 df <- scale(mtcars)
12 pca_result <- prcomp(df, scale. = TRUE)
14 summary(pca_result)
16 autoplot(pca_result, data = as.data.frame(df),
17 loadings = TRUE,
18 loadings_label = TRUE
Console Terminal Background Jobs
> print(svd_result$d)
[1] 14.3129455 9.0644638 4.4094348 2.8909377 2.6319165 2.5611481 2.0477113
[8] 1.9519079 1.5454599 1.2700782 0.8266659
> svd_df <- data.frame(U1 = svd_result$u[,1], U2 = svd_result$u[,2], Car = rownames(df))
> ggplot(svd_df, aes(x = U1, y = U2, label = Car)) +
+ geom_point(color = "darkgreen", size = 3) +
+ geom_text(vjust = -0.5, size = 3) +
+ theme_minimal() +
+ labs(title = "SVD Components - mtcars", x = "U1", y = "U2")
Environment History Connections Tutorial
theme_minimal() +
labs(title = "Kernel PCA (First 2 Components) - mtcars", x = "PC1", y = "...
svd_result <- svd(df)
print(svd_result$d)
svd_df <- data.frame(U1 = svd_result$u[,1], U2 = svd_result$u[,2], Car = ...
ggplot(svd_df, aes(x = U1, y = U2, label = Car)) +
geom_point(color = "darkgreen", size = 3) +
geom_text(vjust = -0.5, size = 3) +
theme_minimal() +
labs(title = "SVD Components - mtcars", x = "U1", y = "U2")
Files Plots Packages Help Viewer Presentation
SVD Components - mtcars
Maserati Bora
Ford Pantera L
Ferrari Dino
Porsche 914-2
Lotus Europa
Mazda RX4
Mazda RX4 Wag
Honda Civic
Camaro Z28
Duster 360
Volvo 142E
Fiat X1-9
Datsun 710
Merc 280
Merc 280C
Chrysler Imperial
Merc 240D
Toyota Corona
Hornet 4 Drive
Valiant
U2
U1


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
ownames(df))
> ggplot(svd_df, aes(x = U1, y = U2, label = Car)) +
+ geom_point(color = "darkgreen", size = 3) +
+ geom_text(vjust = -0.5, size = 3) +
+ theme_minimal() +
+ labs(title = "SVD Components - mtcars", x = "U1", y = "U2")
> data(mtcars)
> df <- scale(mtcars)
> pca_result <- prcomp(df, scale. = TRUE)
> summary(pca_result)
Importance of components:
PC1 PC2 PC3 PC4 PC5 PC6 PC7
PC8
Standard deviation 2.5707 1.6280 0.79196 0.51923 0.47271 0.46000 0.3678
0.35057
Proportion of Variance 0.6008 0.2409 0.05702 0.02451 0.02031 0.01924 0.0123
0.01117
Cumulative Proportion 0.6008 0.8417 0.89873 0.92324 0.94356 0.96279 0.9751
0.98626
PC9 PC10 PC11
Standard deviation 0.2776 0.22811 0.1485
Proportion of Variance 0.0070 0.00473 0.0020
Cumulative Proportion 0.9933 0.99800 1.0000
> autoplot(pca_result, data = as.data.frame(df),
+ loadings = TRUE,
+ loadings.label = TRUE,
+ loadings.colour = "blue",
+ loadings.label.size = 4,
+ main = "PCA Biplot - mtcars (Enhanced)")
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation
PCA Biplot - mtcars (Enhanced)
gear
am
carb
drat
hp
mpg
cyl
disp
wt
vs
qsec
PC2 (24.1%)
PC1 (60.08%)



RESULT

Thus successfully completed dimensionality reduction on the *mtcars* dataset.The reduced components effectively preserved major variance and improved visualization.

EXP.NO.: 06	Perform bivariate and multivariate analysis on the dataset

AIM

To perform bivariate and multivariate analysis on the *iris* dataset and to study correlations, scatterplots, and group comparisons using MANOVA.

TOOLS REQUIRED

- RStudio IDE
- R programming language
- Built-in dataset: iris
- Statistical functions in R

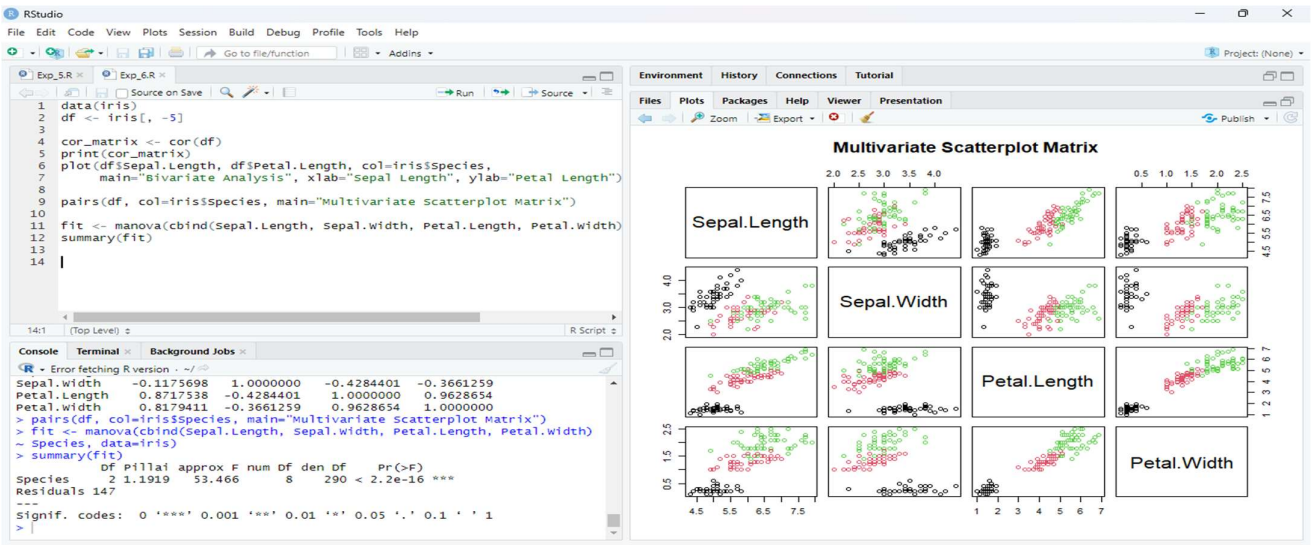
ALGORITHM

- Load the iris dataset and select numerical variables.
- Perform **correlation analysis** for bivariate relationships.
- Plot **scatterplots** for variable comparisons.
- Generate **multivariate scatterplot matrix**.
- Conduct **MANOVA** to check group differences among species.

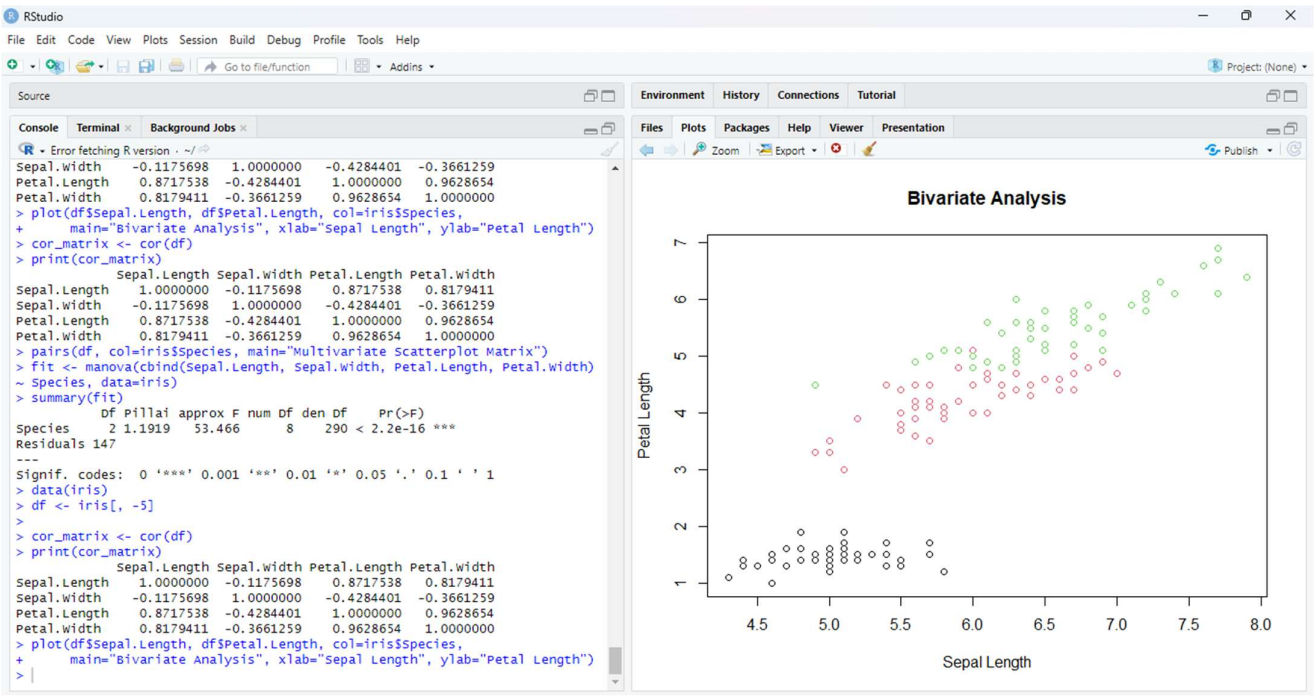
CODING

```
data(iris)
df <- iris[, -5]


cor_matrix <- cor(df)
print(cor_matrix)
plot(df$Sepal.Length, df$Petal.Length, col=iris$Species,
    main="Bivariate Analysis", xlab="Sepal Length", ylab="Petal Length")


pairs(df, col=iris$Species, main="Multivariate Scatterplot Matrix")


fit <- manova(cbind(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width) ~ Species, data=iris)
summary(fit)
```

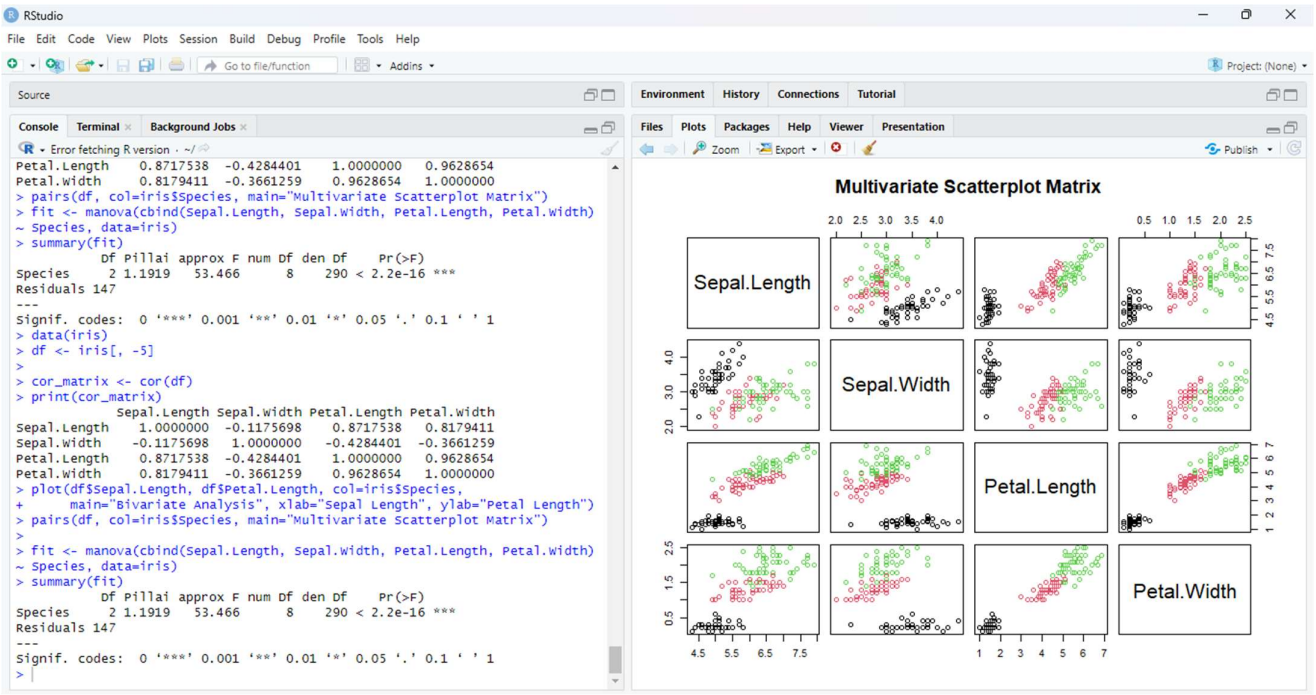
OUTPUT


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Exp_5.R
Exp_6.R
1 data(iris)
2 df <- iris[, -5]
3
4 cor_matrix <- cor(df)
5 print(cor_matrix)
6 plot(df$Sepal.Length, df$Petal.Length, col=iris$Species,
7 main="Bivariate Analysis", xlab="Sepal Length", ylab="Petal Length")
8
9 pairs(df, col=iris$Species, main="Multivariate Scatterplot Matrix")
10
11 fit <- manova(cbind(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width)
12 summary(fit)
13
14
Console Terminal Background Jobs
Sepal.Width -0.1175698 1.0000000 -0.4284401 -0.3661259
Petal.Length 0.8717538 -0.4284401 1.0000000 0.9628654
Petal.Width 0.8179411 -0.3661259 0.9628654 1.0000000
> pairs(df, col=iris$Species, main="Multivariate Scatterplot Matrix")
> fit <- manova(cbind(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width)
~ Species, data=iris)
> summary(fit)
Df Pillai approx F num Df den Df Pr(>F)
Species 2 1.1919 53.466 8 290 < 2.2e-16 ***
Residuals 147

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation
Multivariate Scatterplot Matrix
Sepal.Length
Sepal.Width
Petal.Length
Petal.Width


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
Sepal.Width -0.1175698 1.0000000 -0.4284401 -0.3661259
Petal.Length 0.8717538 -0.4284401 1.0000000 0.9628654
Petal.Width 0.8179411 -0.3661259 0.9628654 1.0000000
> plot(df$Sepal.Length, df$Petal.Length, col=iris$Species,
+ main="Bivariate Analysis", xlab="Sepal Length", ylab="Petal Length")
> cor_matrix <- cor(df)
> print(cor_matrix)
Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length 1.0000000 -0.1175698 0.8717538 0.8179411
Sepal.Width -0.1175698 1.0000000 -0.4284401 -0.3661259
Petal.Length 0.8717538 -0.4284401 1.0000000 0.9628654
Petal.Width 0.8179411 -0.3661259 0.9628654 1.0000000
> pairs(df, col=iris$Species, main="Multivariate Scatterplot Matrix")
> fit <- manova(cbind(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width)
~ Species, data=iris)
> summary(fit)
Df Pillai approx F num Df den Df Pr(>F)
Species 2 1.1919 53.466 8 290 < 2.2e-16 ***
Residuals 147

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> data(iris)
> df <- iris[, -5]
>
> cor_matrix <- cor(df)
> print(cor_matrix)
Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length 1.0000000 -0.1175698 0.8717538 0.8179411
Sepal.Width -0.1175698 1.0000000 -0.4284401 -0.3661259
Petal.Length 0.8717538 -0.4284401 1.0000000 0.9628654
Petal.Width 0.8179411 -0.3661259 0.9628654 1.0000000
> plot(df$Sepal.Length, df$Petal.Length, col=iris$Species,
+ main="Bivariate Analysis", xlab="Sepal Length", ylab="Petal Length")
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation
Bivariate Analysis
Petal Length
Sepal Length




```
Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
> pairs(df, col=iris$Species, main="Multivariate Scatterplot Matrix")
> fit <- manova(cbind(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width)
~ Species, data=iris)
> summary(fit)
             Df Pillai approx F num Df den Df    Pr(>F)
Species      2 1.1919   53.466      8    290 < 2.2e-16 ***
Residuals 147
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> data(iris)
> df <- iris[, -5]
>
> cor_matrix <- cor(df)
> print(cor_matrix)
             Sepal.Length Sepal.Width Petal.Length Petal.Width
Sepal.Length    1.0000000  -0.1175698    0.8717538   0.8179411
Sepal.Width    -0.1175698   1.0000000   -0.4284401  -0.3661259
Petal.Length    0.8717538  -0.4284401    1.0000000   0.9628654
Petal.Width     0.8179411  -0.3661259    0.9628654   1.0000000
> plot(df$Sepal.Length, df$Petal.Length, col=iris$Species,
+      main="Bivariate Analysis", xlab="Sepal Length", ylab="Petal Length")
> pairs(df, col=iris$Species, main="Multivariate Scatterplot Matrix")
>
> fit <- manova(cbind(Sepal.Length, Sepal.Width, Petal.Length, Petal.Width)
~ Species, data=iris)
> summary(fit)
             Df Pillai approx F num Df den Df    Pr(>F)
Species      2 1.1919   53.466      8    290 < 2.2e-16 ***
Residuals 147
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

RESULT

Thus successfully completed bivariate and multivariate analysis on the *iris* dataset.The study showed clear correlations and confirmed species differences through MANOVA.

EXP. NO.: 07	Apply and explore various plotting functions on the data set

AIM

To visualize data using different plotting functions in R and to analyze patterns and distributions in the *airquality* dataset.

TOOLS REQUIRED

- RStudio IDE
- R programming language
- ggplot2 library for advanced plots
- Built-in dataset: airquality

ALGORITHM

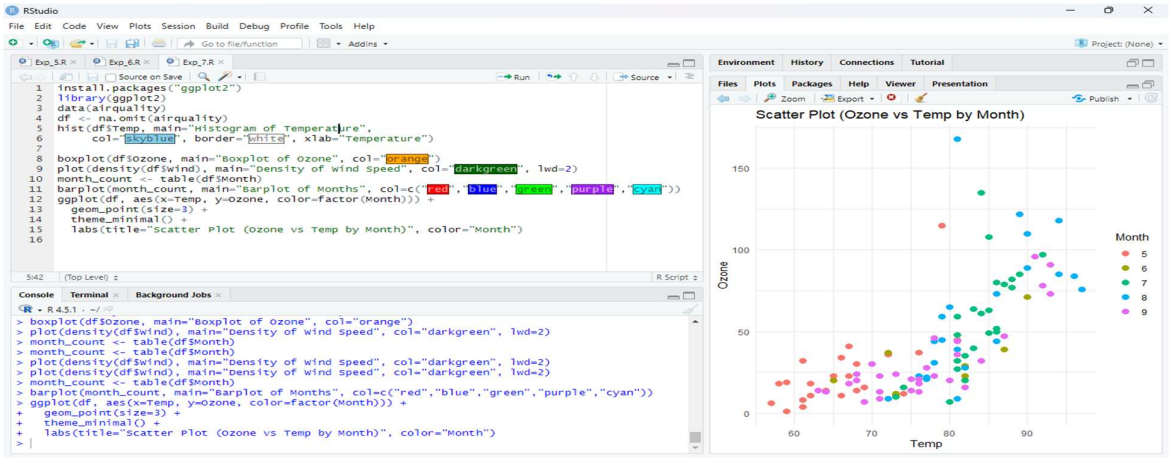
- Install and load necessary packages (ggplot2).
- Load the airquality dataset.
- Handle missing values using na.omit().
- Plot histogram for temperature distribution.
- Draw boxplot for ozone levels.
- Plot density curve for wind speed.
- Plot bar chart for monthly distribution.
- Use ggplot2 for scatter plot visualization of Ozone vs Temperature, colored by Month.

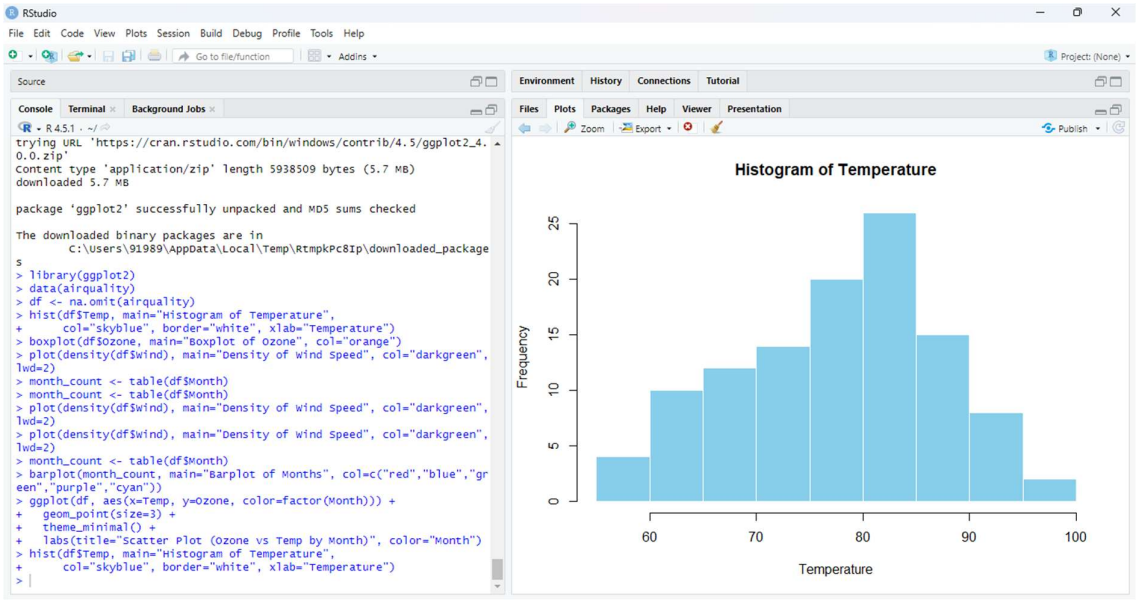
CODING

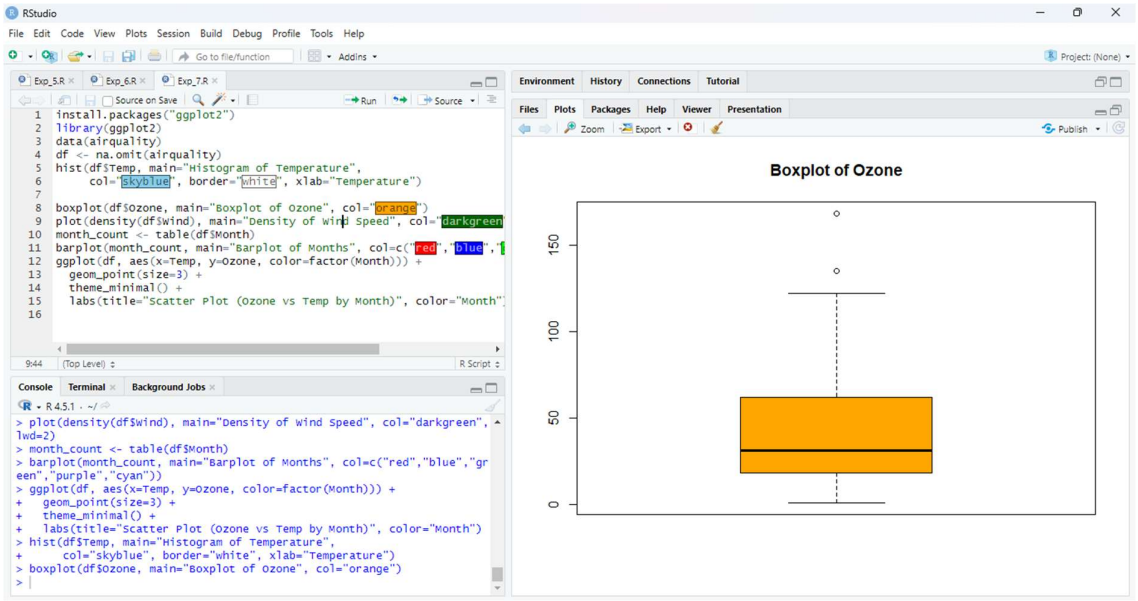
```
install.packages("ggplot2")

library(ggplot2)

data(airquality)

df <- na.omit(airquality)

hist(df$Temp, main="Histogram of Temperature",

    col="skyblue", border="white", xlab="Temperature")
```

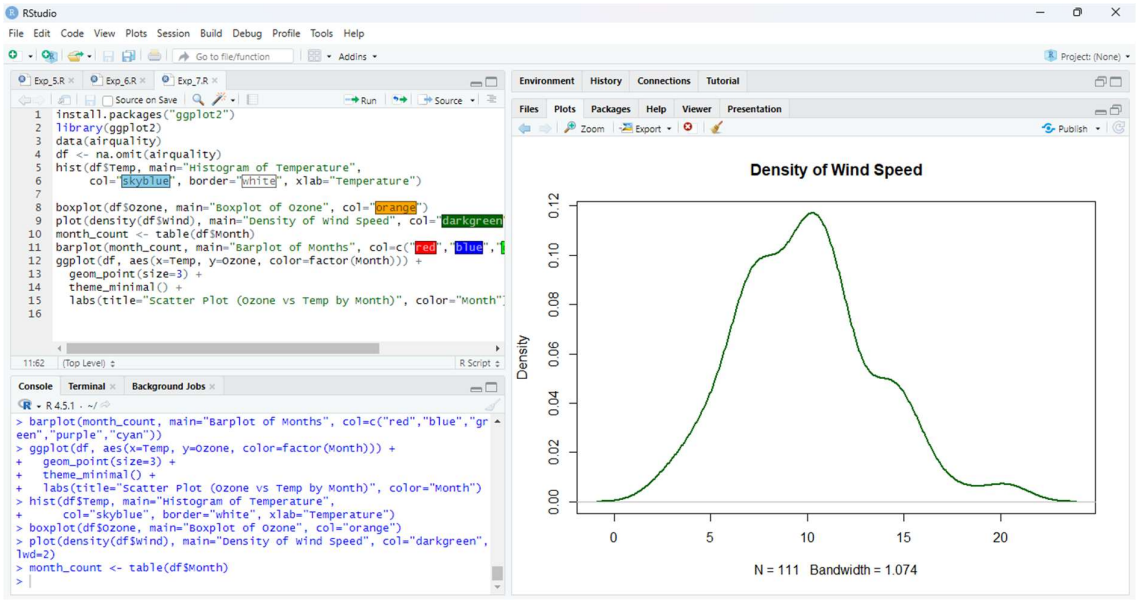
```
boxplot(df$Ozone, main="Boxplot of Ozone", col="orange")

plot(density(df$Wind), main="Density of Wind Speed", col="darkgreen", lwd=2)

month_count <- table(df$Month)

barplot(month_count, main="Barplot of Months",
col=c("red","blue","green","purple","cyan"))

ggplot(df, aes(x=Temp, y=Ozone, color=factor(Month))) +
  geom_point(size=3) +
  theme_minimal() +
  labs(title="Scatter Plot (Ozone vs Temp by Month)", color="Month")
```

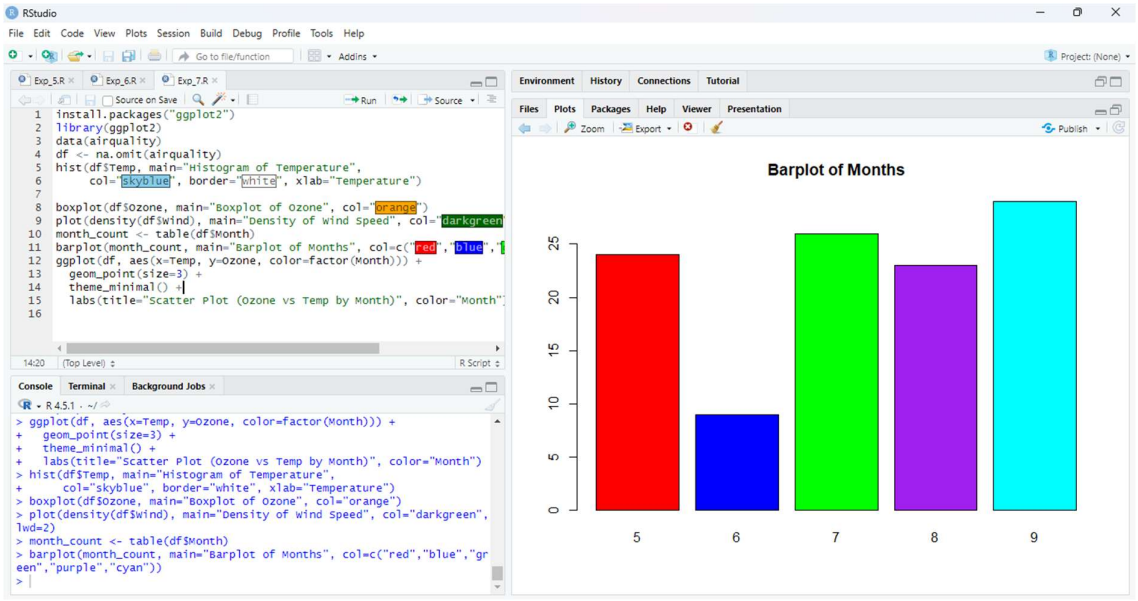
OUTPUT

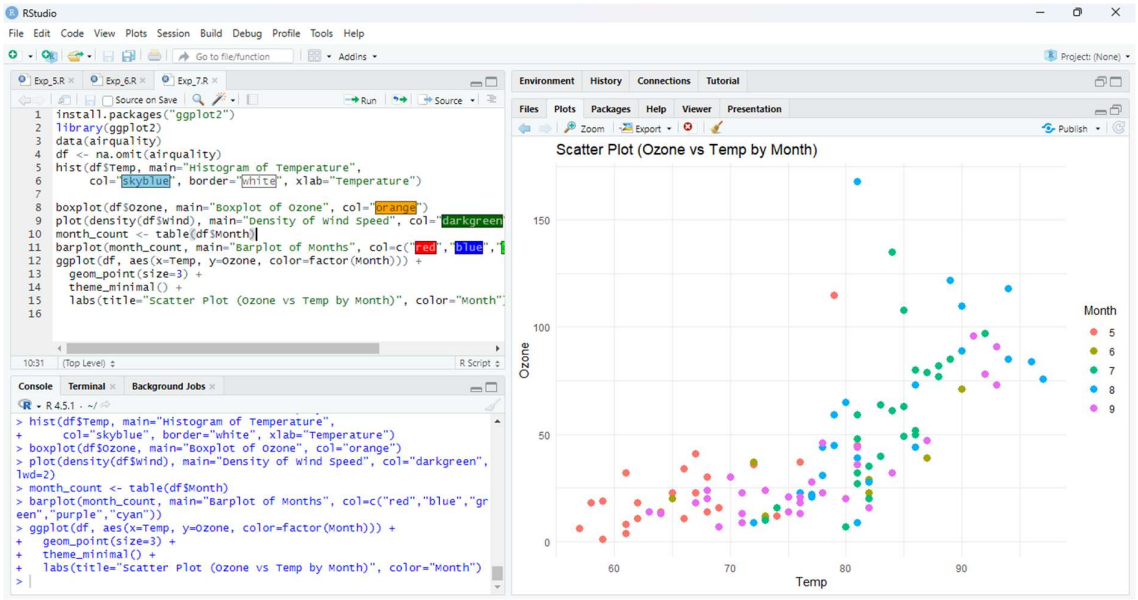



RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source
Console Terminal Background Jobs
R 4.5.1 · ~/
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.5/ggplot2_4.0.0.zip'
Content type 'application/zip' length 5938509 bytes (5.7 MB)
downloaded 5.7 MB
package 'ggplot2' successfully unpacked and MD5 sums checked
The downloaded binary packages are in
C:\Users\91989\AppData\Local\Temp\RtmpkPc8Ip\downloaded_packages
> library(ggplot2)
> data(airquality)
> df <- na.omit(airquality)
> hist(df$Temp, main="Histogram of Temperature",
+ col="skyblue", border="white", xlab="Temperature")
> boxplot(df$Ozone, main="Boxplot of Ozone", col="orange")
> plot(density(df$Wind), main="Density of Wind Speed", col="darkgreen", lwd=2)
> month_count <- table(df$Month)
> month_count <- table(df$Month)
> plot(density(df$Wind), main="Density of Wind Speed", col="darkgreen", lwd=2)
> plot(density(df$Wind), main="Density of Wind Speed", col="darkgreen", lwd=2)
> month_count <- table(df$Month)
> barplot(month_count, main="Barplot of Months", col=c("red","blue","green","purple","cyan"))
> ggplot(df, aes(x=Temp, y=Ozone, color=factor(Month))) +
+ geom_point(size=3) +
+ theme_minimal() +
+ labs(title="Scatter Plot (Ozone vs Temp by Month)", color="Month")
> hist(df$Temp, main="Histogram of Temperature",
+ col="skyblue", border="white", xlab="Temperature")
>
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Histogram of Temperature
Frequency
0 5 10 15 20 25
60 70 80 90 100
Temperature


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Exp_5.R Exp_6.R Exp_7.R
Source on Save Run Source
1 install.packages("ggplot2")
2 library(ggplot2)
3 data(airquality)
4 df <- na.omit(airquality)
5 hist(df$Temp, main="Histogram of Temperature",
6 col="skyblue", border="white", xlab="Temperature")
7
8 boxplot(df$Ozone, main="Boxplot of Ozone", col="orange")
9 plot(density(df$Wind), main="Density of Wind Speed", col="darkgreen
10 month_count <- table(df$Month)
11 barplot(month_count, main="Barplot of Months", col=c("red","blue",
12 ggplot(df, aes(x=Temp, y=Ozone, color=factor(Month))) +
13 geom_point(size=3) +
14 theme_minimal() +
15 labs(title="Scatter Plot (Ozone vs Temp by Month)", color="Month"
16
9:44 (Top Level) R Script
Console Terminal Background Jobs
R 4.5.1 · ~/
> plot(density(df$Wind), main="Density of Wind Speed", col="darkgreen", lwd=2)
> month_count <- table(df$Month)
> barplot(month_count, main="Barplot of Months", col=c("red","blue","green","purple","cyan"))
> ggplot(df, aes(x=Temp, y=Ozone, color=factor(Month))) +
+ geom_point(size=3) +
+ theme_minimal() +
+ labs(title="Scatter Plot (Ozone vs Temp by Month)", color="Month")
> hist(df$Temp, main="Histogram of Temperature",
+ col="skyblue", border="white", xlab="Temperature")
> boxplot(df$Ozone, main="Boxplot of Ozone", col="orange")
>
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Boxplot of Ozone
0 50 100 150


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Exp_5.R Exp_6.R Exp_7.R
Source on Save Run Source
1 install.packages("ggplot2")
2 library(ggplot2)
3 data(airquality)
4 df <- na.omit(airquality)
5 hist(df$Temp, main="Histogram of Temperature",
6 col="skyblue", border="white", xlab="Temperature")
7
8 boxplot(df$Ozone, main="Boxplot of Ozone", col="orange")
9 plot(density(df$Wind), main="Density of Wind Speed", col="darkgreen
10 month_count <- table(df$Month)
11 barplot(month_count, main="Barplot of Months", col=c("red","blue",
12 ggplot(df, aes(x=Temp, y=Ozone, color=factor(Month))) +
13 geom_point(size=3) +
14 theme_minimal() +
15 labs(title="Scatter Plot (Ozone vs Temp by Month)", color="Month"
16
11:62 (Top Level) R Script
Console Terminal Background Jobs
R 4.5.1 · ~/
> barplot(month_count, main="Barplot of Months", col=c("red","blue","gr
een","purple","cyan"))
> ggplot(df, aes(x=Temp, y=Ozone, color=factor(Month))) +
+ geom_point(size=3) +
+ theme_minimal() +
+ labs(title="Scatter Plot (Ozone vs Temp by Month)", color="Month")
> hist(df$Temp, main="Histogram of Temperature",
+ col="skyblue", border="white", xlab="Temperature")
> boxplot(df$Ozone, main="Boxplot of Ozone", col="orange")
> plot(density(df$Wind), main="Density of Wind Speed", col="darkgreen",
lwd=2)
> month_count <- table(df$Month)
>
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Density of Wind Speed
Density
0.00 0.02 0.04 0.06 0.08 0.10 0.12
0 5 10 15 20
N = 111 Bandwidth = 1.074


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Exp_5.R Exp_6.R Exp_7.R
Source on Save Run Source
1 install.packages("ggplot2")
2 library(ggplot2)
3 data(airquality)
4 df <- na.omit(airquality)
5 hist(df$Temp, main="Histogram of Temperature",
6 col="skyblue", border="white", xlab="Temperature")
7
8 boxplot(df$Ozone, main="Boxplot of Ozone", col="orange")
9 plot(density(df$Wind), main="Density of Wind Speed", col="darkgreen
10 month_count <- table(df$Month)
11 barplot(month_count, main="Barplot of Months", col=c("red","blue",
12 ggplot(df, aes(x=Temp, y=Ozone, color=factor(Month))) +
13 geom_point(size=3) +
14 theme_minimal() +
15 labs(title="Scatter Plot (Ozone vs Temp by Month)", color="Month"
16
14:20 (Top Level) R Script
Console Terminal Background Jobs
R 4.5.1 · ~/
> ggplot(df, aes(x=Temp, y=Ozone, color=factor(Month))) +
+ geom_point(size=3) +
+ theme_minimal() +
+ labs(title="Scatter Plot (Ozone vs Temp by Month)", color="Month")
> hist(df$Temp, main="Histogram of Temperature",
+ col="skyblue", border="white", xlab="Temperature")
> boxplot(df$Ozone, main="Boxplot of Ozone", col="orange")
> plot(density(df$Wind), main="Density of Wind Speed", col="darkgreen",
lwd=2)
> month_count <- table(df$Month)
> barplot(month_count, main="Barplot of Months", col=c("red","blue","gr
een","purple","cyan"))
>
Environment History Connections Tutorial
Files Plots Packages Help Viewer Presentation
Zoom Export Publish
Barplot of Months
0 5 10 15 20 25
5 6 7 8 9



RESULT

Thus successfully completed plotting on the *airquality* dataset using multiple visualization techniques.The plots provided insights into data distribution, relationships, and seasonal variations.

EXP. NO.: 08	FEATURES OF POWER BI DESKTOP

AIM

To study and implement the features of Power BI Desktop by preparing & loading data into Power BI, and developing the data model for analysis and visualization.

TOOLS REQUIRED

1. **Power BI Desktop** (latest version).
2. **Dataset** – Sales Data (CSV/Excel from Kaggle).

ALGORITHM

A: Prepare & Load Data

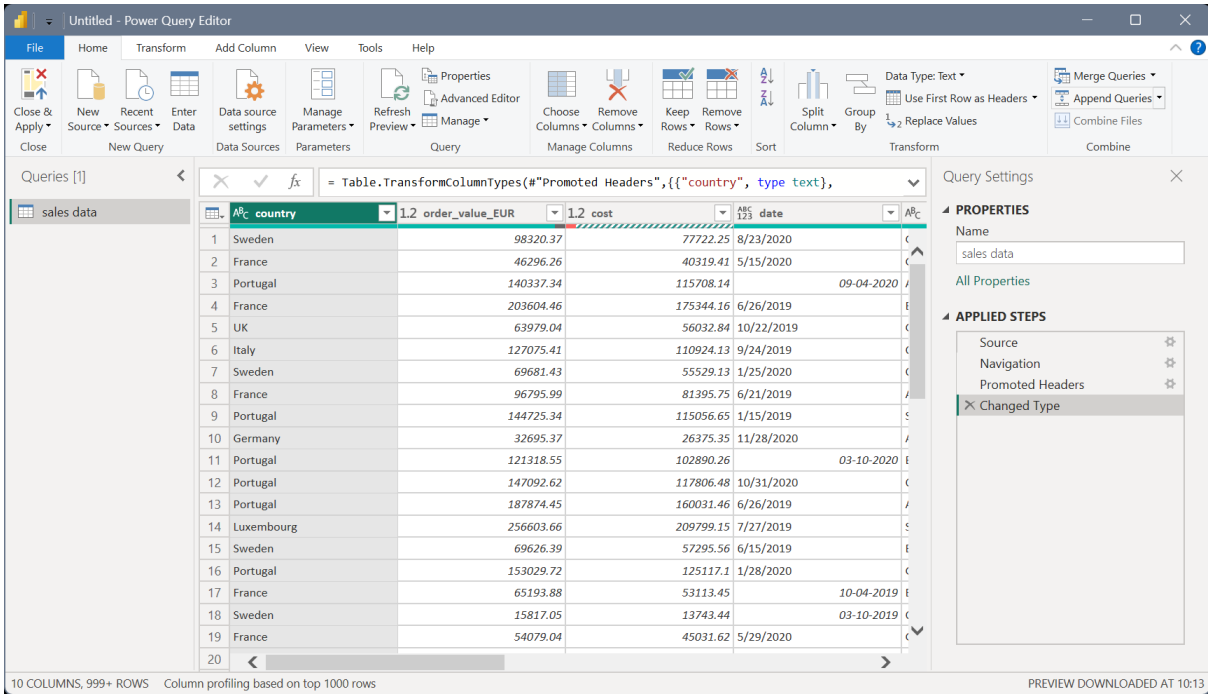
1. Open **Power BI Desktop**.
2. Click **Home → Get Data → Excel/CSV** and select the dataset file.
3. Preview the dataset in the **Navigator** window.
4. Select the required sheet/table and click **Load** (or **Transform Data** if cleaning is needed).
5. In **Power Query Editor**, perform data preparation:
    - Promote headers.
    - Check and change data types (Date, Number, Text).
    - Remove null or duplicate rows if necessary.
    - Rename columns for clarity.
6. Click **Close & Apply** to load the cleaned data into Power BI.

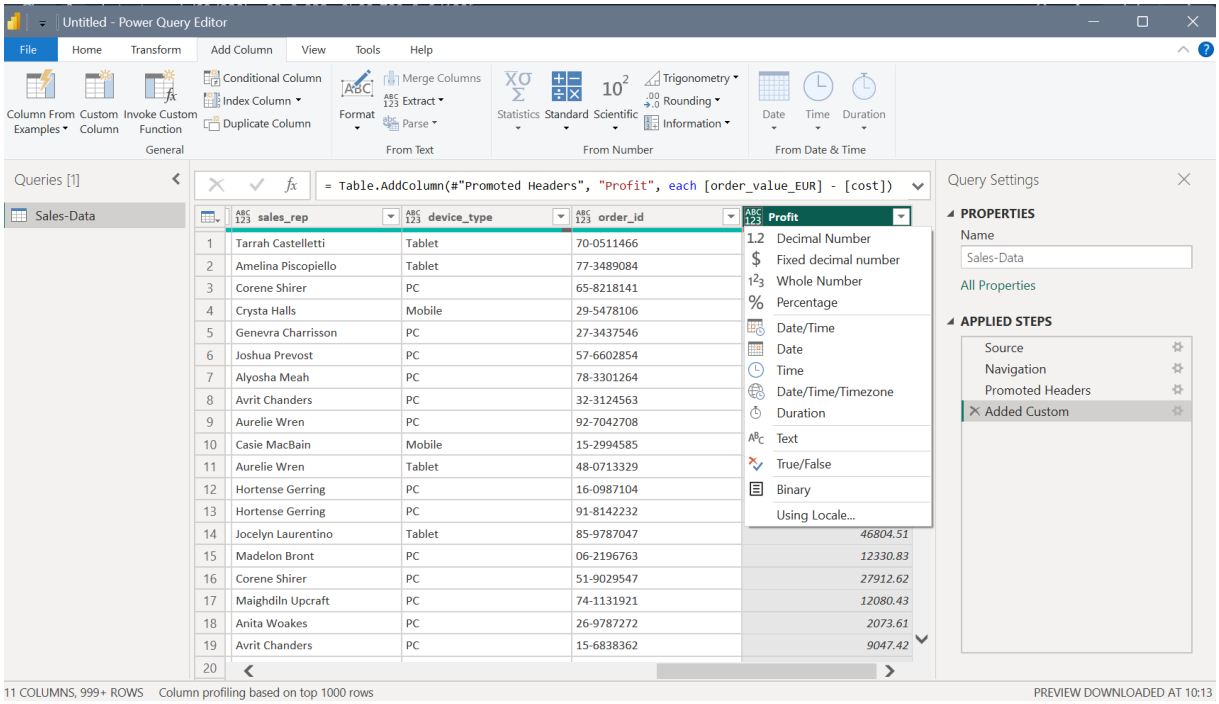
B: DEVELOPING THE DATA MODEL

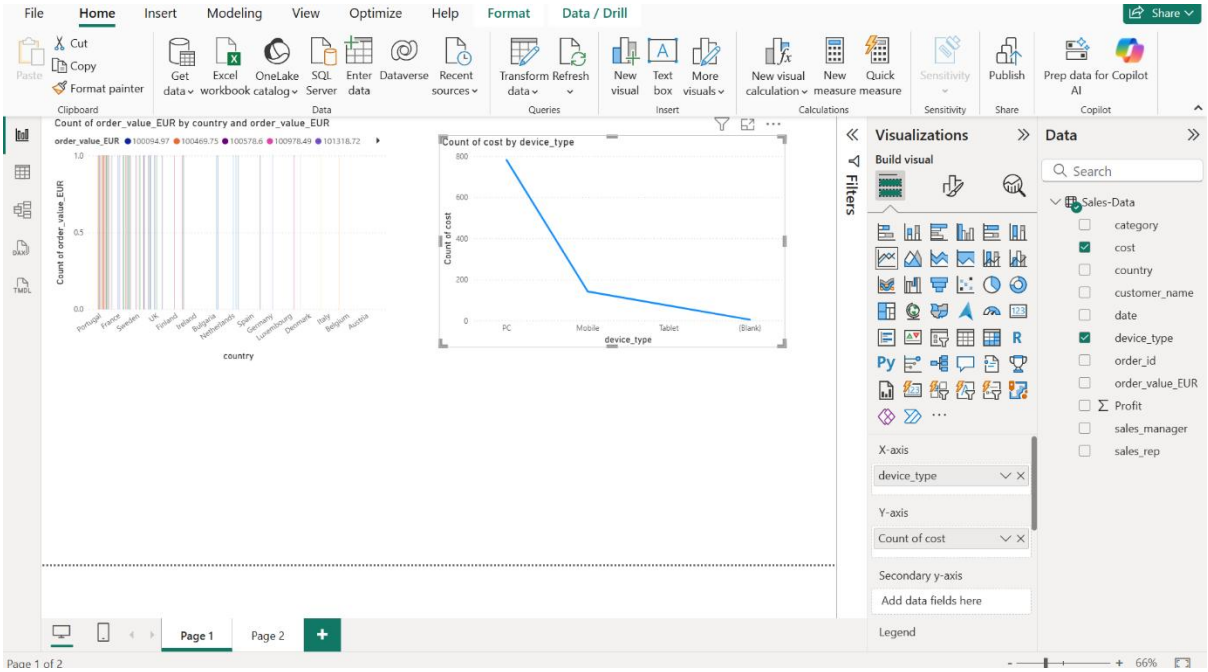
1. Switch to **Model View** in Power BI.
2. Verify that the table(s) are loaded correctly.
3. If multiple tables exist, define **relationships** between them (e.g., Sales ↔ Customers ↔ Products).
4. Ensure data types and relationships are correct to support meaningful analysis.

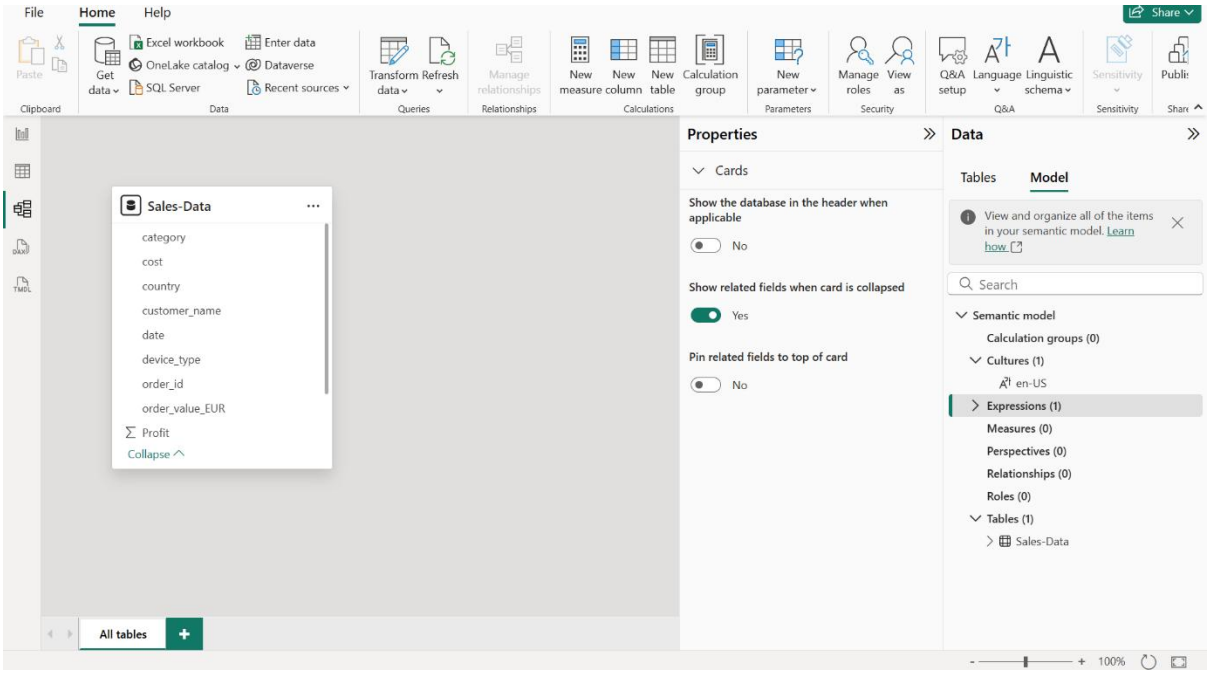
5. The developed data model is now ready for building **visualizations** like charts, cards, and slicers.

OUTPUT









RESULT

The sales dataset was successfully prepared, cleaned, and loaded into Power BI Desktop. A simple data model with calculated measures was developed, making the data ready for visualization and analysis.

EXP NO: 09	PERFORM DAX CALCULATIONS AND DESIGN A REPORT

AIM:

To load a dataset in Power BI, calculate total and average marks for students using DAX and design a report to analyze performance across subjects and gender.

TOOLS REQUIRED:

- Power BI Desktop
- Student marks dataset

ALGORITHM:

1. Load ‘student_data’ dataset into Power BI desktop.
2. Check columns, data types and remove unnecessary columns.
3. Create DAX measures for ‘Total_Marks’, ‘Average_Marks’, ‘Highest_Score’ and such.
4. Design the report by adding appropriate visualizations.
5. Save the dashboard, and export as required.

CODING:

DAX query to calculate Total_Marks:

Total_Marks = Sheet1[English] + Sheet1[Math] + Sheet1[Science] + Sheet1[Social Science] + Sheet1[Tamil]

DAX query to calculate Average_Marks:

Average_Marks = Sheet1[Total_Marks]/ 5

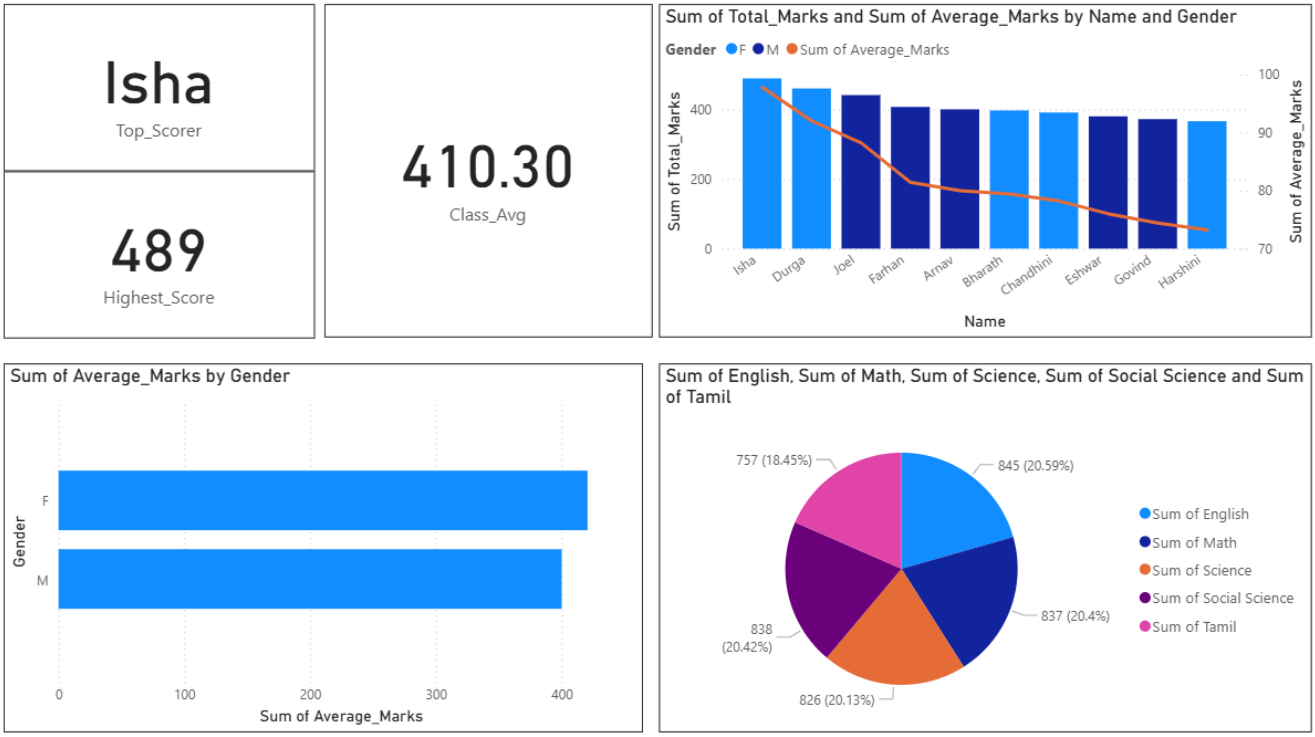
DAX query to calculate Highest_Score:

Highest_Score = MAXX(Sheet1, [Total_Marks])

DAX query to find Top_Scorer:

Top_Scorer = VAR TopStudent = TOPN(1, Sheet1, Sheet1[Total_Marks], DESC) RETURN CONCATENATEX(TopStudent, Sheet1[Name])

OUTPUT



RESULT

DAX measures were successfully applied to calculate Total Marks, Average Marks, and Highest Scorer. The report visually summarized student performance across subjects and gender. Interactive visuals enabled easy exploration and identification of trends and top performers.

EXP NO: 10	CREATE A DASHBOARD AND PERFORM DATA ANALYSIS

AIM:

To load a dataset in Power BI, perform data modelling, apply DAX calculations, design interactive visuals, and create a dashboard for financial analysis.

TOOLS REQUIRED:

- Power BI Desktop
- Financials dataset

ALGORITHM:

1. Load ‘financials’ sample dataset into Power BI desktop.
2. Check columns, data types and remove unnecessary columns.
3. Go to report view and insert visuals as pre requirement to plot necessary graphs.
4. Create new measurements like ‘Profit_Per_Unit’ using DAX queries to display as cards or .
5. Use ‘Analyze’ feature in plotted charts to identify factors that influence the distribution of the selected measure.
6. Arrange all visuals on a single page to make a cohesive report.
7. Publish the report to your workspace and pin the visuals required to the dashboard
8. View the dashboard online in Power BI Service.

CODING:

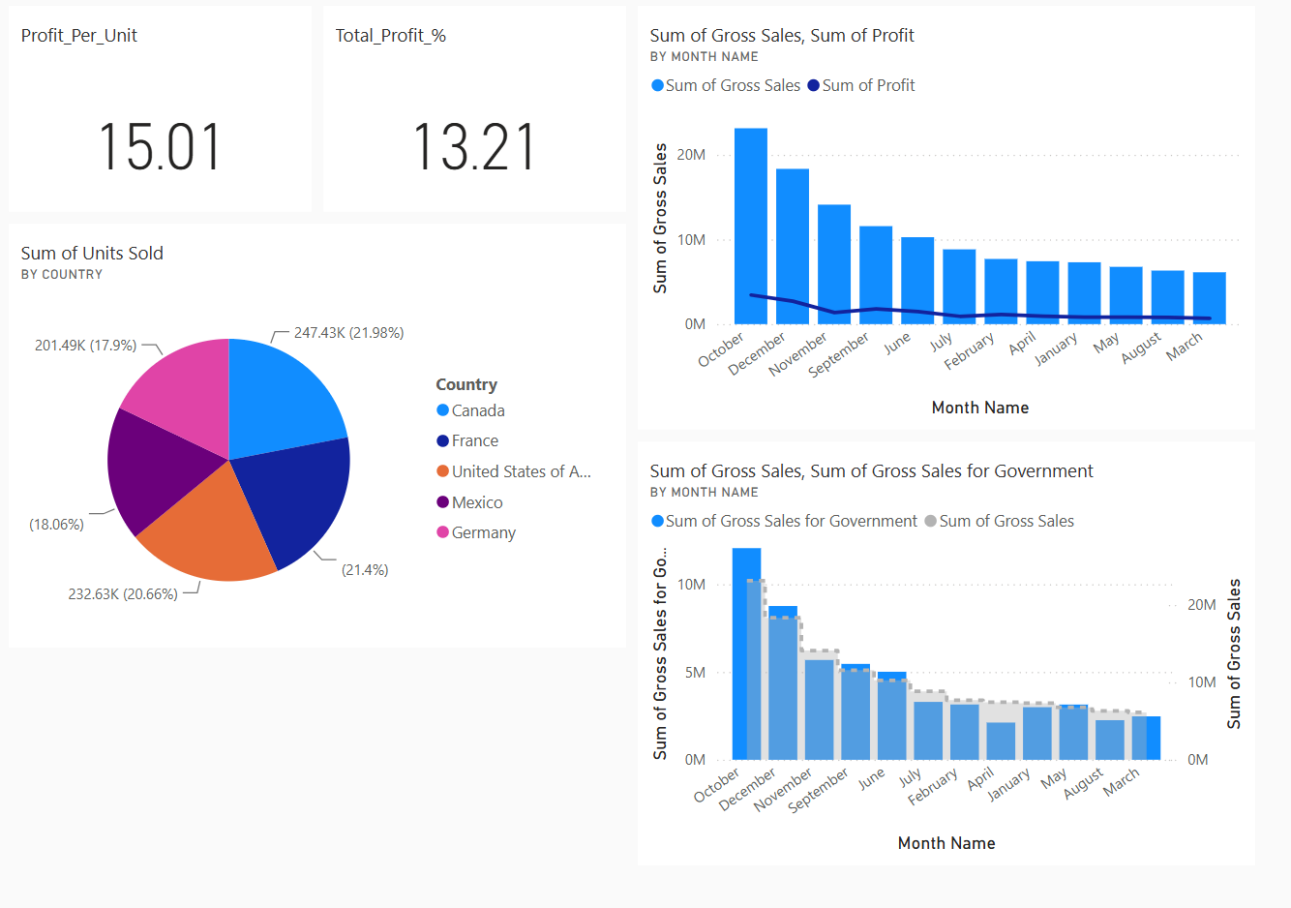
DAX query to calculate Profit_Per_Unit:

```
Profit_Per_Unit = DIVIDE( SUM(financials[Profit]), SUM(financials[Units Sold]), 0)
```

DAX query to calculate Total_Profit:

```
Total_Profit_% = DIVIDE(SUM(financials[Profit]), SUM(financials[Gross Sales]), 0)*100
```

OUTPUT



RESULT

A dashboard was successfully created in Power BI using the financials sample dataset. Key metrics such as Total Profit %, Profit per Unit and Revenue trends were calculated using DAX measures. Interactive visuals allowed clear identification of patterns, segment-wise performance and factors influencing financial outcomes.