EXP NO: 1(A)	PERFORM NUMERICAL OPERATIONS (MAX, MIN, AVG, SUM, SQRT, ROUND).
DATE: 3/6/24	

AIM:

To get the input from user and perform numerical operations (Max, Min, Avg, Sum, Sqrt, Round).using MS-Excel.

THEORY:

Descriptive analytics is a statistical interpretation used to analyze historical data to identify patterns and relationships. Descriptive analytics seeks to describe an event, phenomenon, or outcome. It helps understand what has happened in the past and provides businesses the perfect base to track trends.

PROCEDURE:

Step 1: Start MS-Excel application in MS-Office.
Step 2: Create data sheet for student /Employee/ departmental store in MS-Excel application
Step 3: Calculate the Maximum of the given marks/salary using Max function.
Step 4: Calculate the Minium of the given marks/salary using Min function.
Step 5: Calculate the Average of the given marks/salary using Avg function.
Step 6: Calculate the Sum of the given marks/salary using Sum function.
Step 7: Calculate the Square root of the given marks/salary using SQRT function.
Step 8: Calculate the Round of the given marks/salary using Roundup function.
Step 9: Display the Output of the Desired numerical operation.
Step 10: Save the MS-Excel and close the application.

PROCEDURE: LOAD THE DATASET USING MS-EXCEL

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

1.The MAX function returns the highest value in a set of data.

Firstly, in Cell B12 type the formula given below-

=MAX(B2:B8)

2.The MIN function returns the smallest value from the numbers provided.

Firstly, in Cell B13 type the formula given below-

=MIN(B2:B8)

3.The AVERAGE function calculates the average (arithmetic mean) of a group of numbers.

Firstly, in Cell B14 type the formula given below-

=AVERAGE(B2:B8)

4.The SUM function calculates the SUM of a group of numbers in Excel.

Firstly, in Cell B15 type the formula given below

5.The Excel SQRT function returns the square root of positive number

6.The ROUND function rounds a number to a specified number of digits.

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

OUTPUT:

No	A	B	C
1.	Roll No	Name	Marks
2.	1	Jane	74
3.	2	Lix	80
4.	3	Remy	65
5.	4	Rani	90
6.	5	Keerthi	70
7.	6	Dinesh	80
8.	7	Max	80
9.	8	Min	65
10.		Avg	71.4
11.		Sum	357

EXP NO: 1(B)	DATA IMPORT/EXPORT OPERATIONS FOR DIFFERENT FILE FORMATS.
DATE: 3/6/24	

AIM:

To Perform data import and export operations for different file formats using MS-Excel.

THEORY:

- **Data Importing:** This involves transferring external data into your system.
- **Manual Data Entry:** This basic method is time-consuming but vital when dealing with small datasets.
- **Automated Data Extraction Tools**: These tools can import large quantities of data quickly and accurately.
- **Data Transformation:** This process converts the imported data into a format suitable for analysis.
- **Cleaning Data:** It's essential to remove errors and inconsistencies for accurate analysis.
- **Standardizing Data:** Homogenizing the formats ensures easy comparison and interpretation.
- **Data Exporting**: The final stage involves moving the transformed data to another location for further use or storage.
- **File Formats:** Different types (CSV, JSON, XML) serve different purposes. Choosing wisely affects performance and compatibility.

PROCEDURE:

1. **IMPORT FILE**

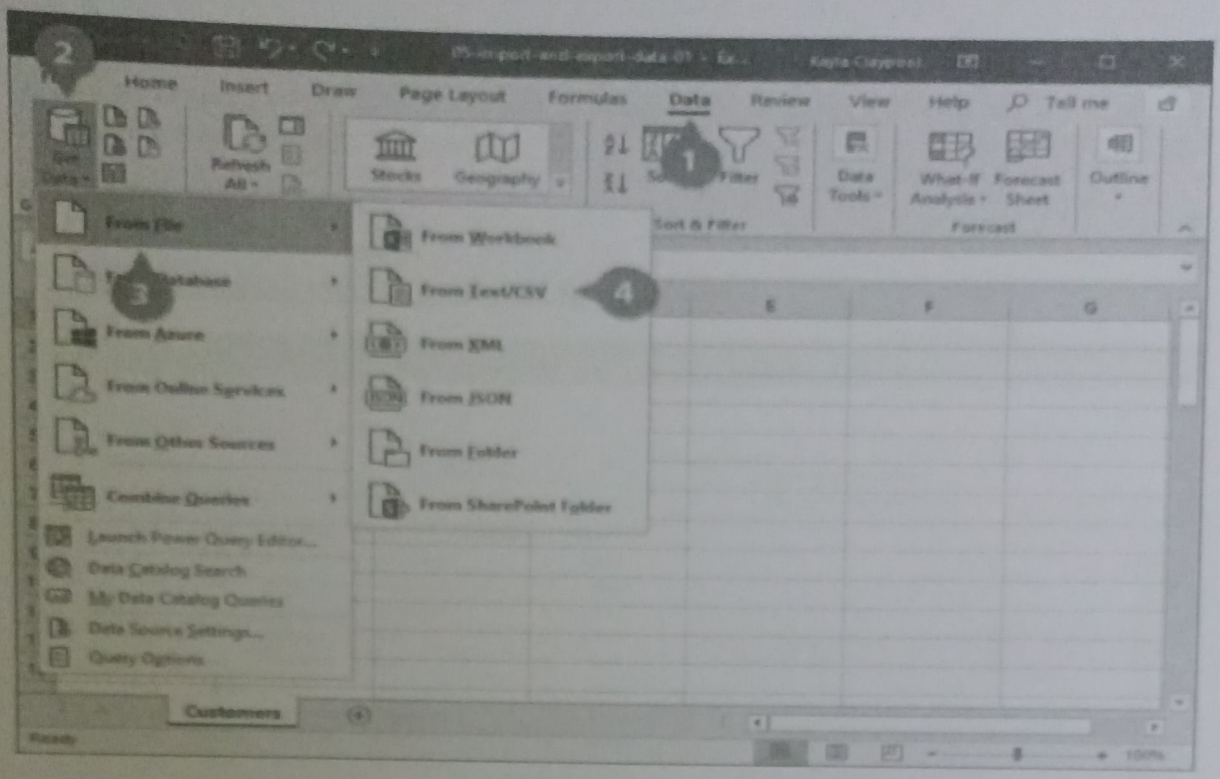
Excel can import data from external data sources including other files, databases, or web pages.

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.

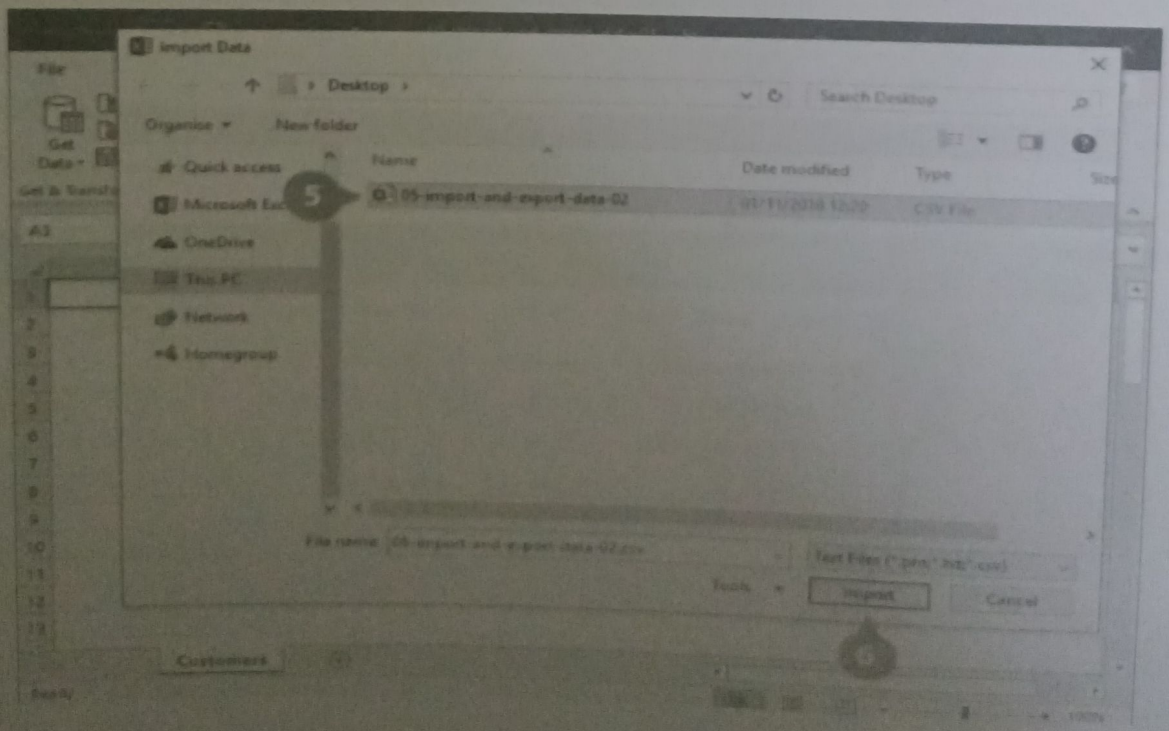
4. Select From Text/CSV



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.

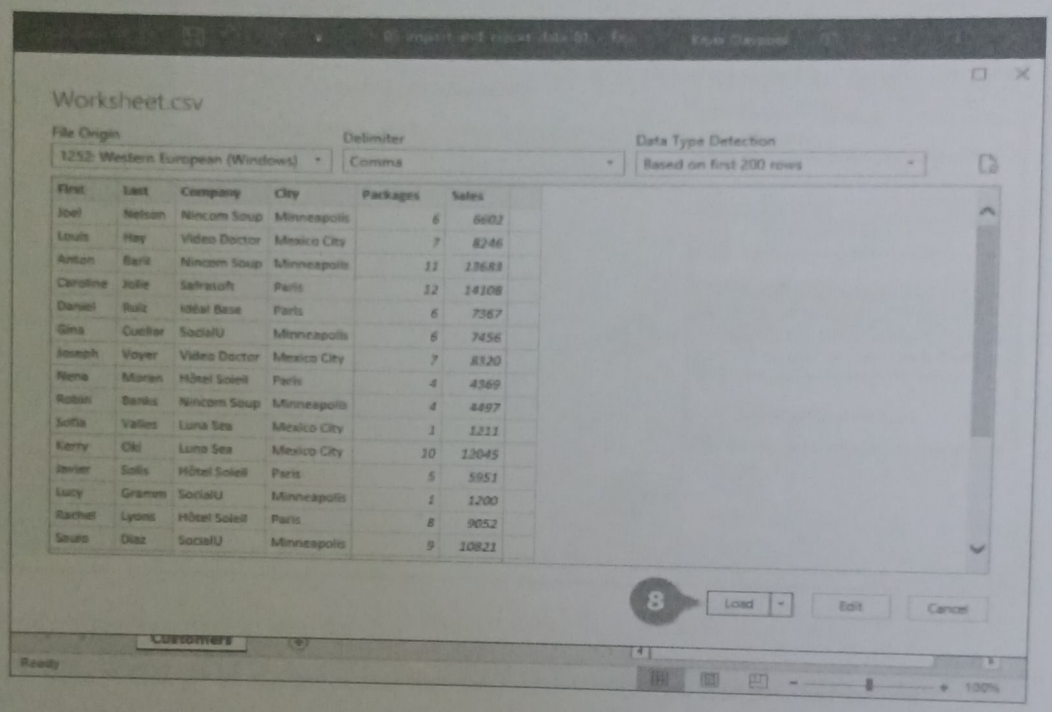


7. Verify the preview looks correct.

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.

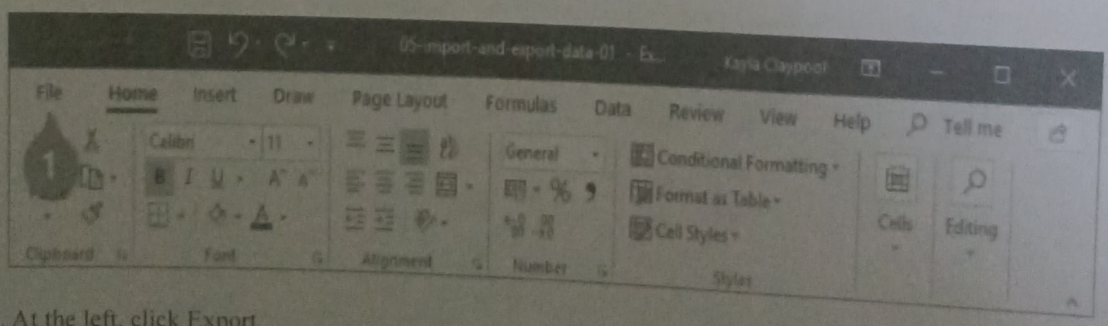
OUTPUT FOR IMPORT THE FILE:



First	Last	Company	City	Packages	Sales
Joel	Nelson	Nincom Soup	Minneapolis	6	6602
Louis	Hay	Video Doctor	Mexico City	7	8246
Anton	Baril	Nincom Soup	Minneapolis	11	13683
Caroline	Jolie	Safrasoft	Paris	12	14108
Daniel	Ruiz	Idéal Base	Paris	6	7367
Gina	Cuellar	SocialU	Minneapolis	6	7456
Joseph	Voyer	Video Doctor	Mexico City	7	8320
Rena	Moran	Hôtel Soleil	Paris	4	4369
Robin	Banks	Nincom Soup	Minneapolis	4	4497
Sofia	Valles	Luna Sea	Mexico City	1	1211
Kerry	Oki	Luna Sea	Mexico City	10	12045
Javier	Solis	Hôtel Soleil	Paris	5	5951
Lucy	Gramm	SocialU	Minneapolis	1	1200
Rachel	Lyons	Hôtel Soleil	Paris	8	9052
Souto	Diaz	SocialU	Minneapolis	9	10821

2.EXPORT FILE

1. Click the File tab.



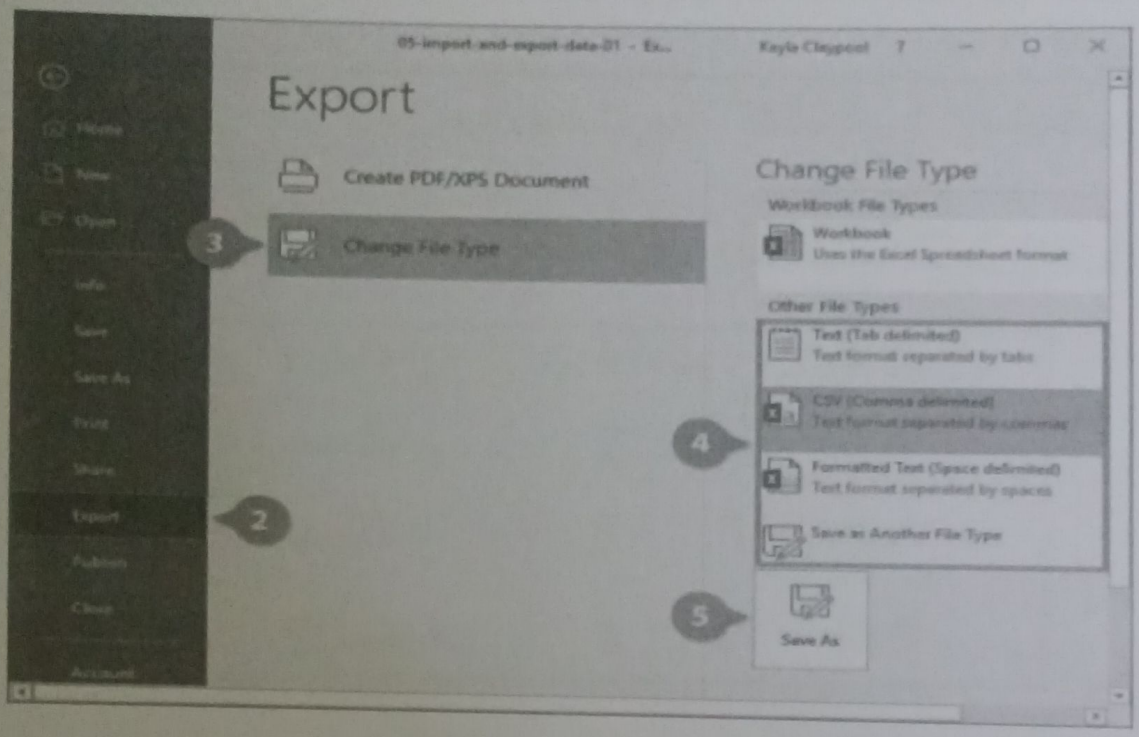
2. At the left, click Export.

3. Click the Change File Type.

4. Under Other File Types, select a file type. o Text (Tab delimited): The cell data will be separated by a tab. o CSV (Comma delimited): The cell data will be separated by a comma. o Formatted Text

(space delimited): The cell data will be separated by a space. o 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.

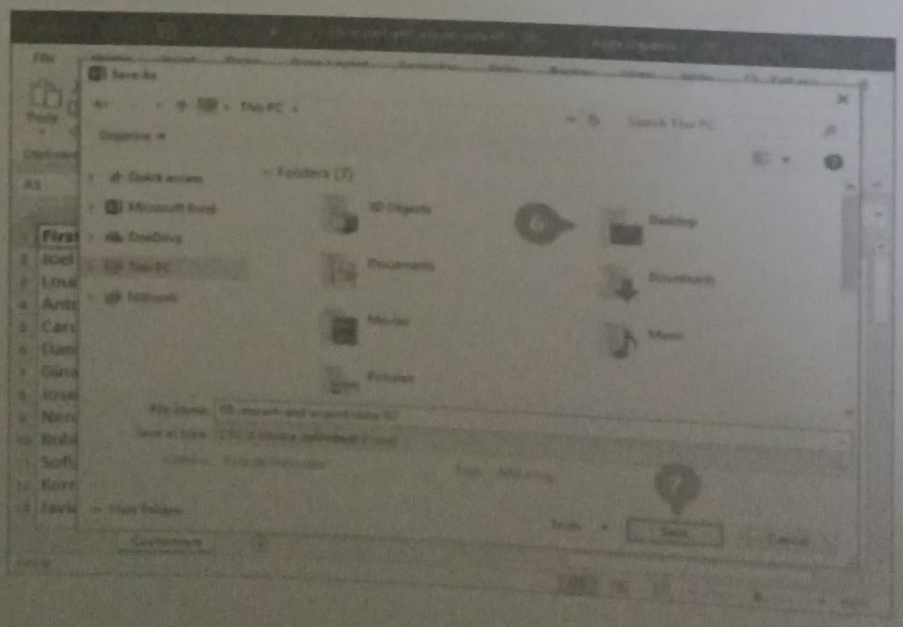


6. Specify where you want to save the file.

7. Click Save. A dialog box appears stating that some of the workbook features may be lost.

8. Click Yes.

OUTPUT FOR EXPORTING THE FILE



OUTPUT:

Auto save [+C++]

Worksheet.csv

File origin | Delimiter | Datatype Collection

Student details | Common | Based on rows

FNAME	LNAME	GRADE	SUBJECT
John	vishnu	II	BA
Ligger	victor	I	DV
Rory	Dinesh	I	HCI

LOAD | EDIT | CANCEL

EXP NO: 2	PERFORM STATISTICAL OPERATIONS [Mean, Median, Mode and Standard Deviation, Variance, Skewness, Kurtosis]
DATE: 10/6/24	

AIM:

To perform statistical operations like Mean, Median, Mode and Standard deviation, Variance, Skewness, kurtosis.

THEORY:

Mean, median, and mode: Mean, median, and mode are different measures of centre in a numerical data set. They each try to summarize a dataset with a single number to represent a "typical" data point from the dataset.

Mean: The "average" number; found by adding all data points and dividing by the number of data points.

Example: The mean of 444, 111, and 777 is (4+1+7)/3=12/3=4(4+1+7)/3=12/3=4left parenthesis, 4, plus, 1, plus, 7, right parenthesis, slash, 3, equals, 12, slash, 3, equals, 4.

Median: The middle number; found by ordering all data points and picking out the one in the middle (or if there are two middle numbers, taking the mean of those two numbers).

Example: The median of 444, 111, and 777 is 444 because when the numbers are put in order (1(1left parenthesis, 1, 444, 7)7)7, right parenthesis, the number 444 is in the middle.

Mode: The most frequent number—that is, the number that occurs the highest number of times.

Variance :

According to layman's words, the variance is a measure of how far a set of data are dispersed out from their mean or average value. It is denoted as 'σ2 '.

Properties of Variance

• It is always non-negative since each term in the variance sum is squared and therefore the result is either positive or zero.

• Variance always has squared units. For example, the variance of a set of weights estimated in kilograms will be given in kg squared. Since the population variance is squared, we cannot compare it directly with the mean or the data themselves.

Standard Deviation:

The spread of statistical data is measured by the standard deviation. Distribution measures the deviation of data from its mean or average position. The degree of dispersion is computed by the method of estimating the deviation of data points. It is denoted by the symbol, 'σ'.

Properties of Standard Deviation

• It describes the square root of the mean of the squares of all values in a data set and is also called the root-mean-square deviation.

• The smallest value of the standard deviation is 0 since it cannot be negative.

• When the data values of a group are similar, then the standard deviation will be very low or close to zero. But when the data values vary with each other, then the standard variation is high or far

Skewness and Kurtosis:

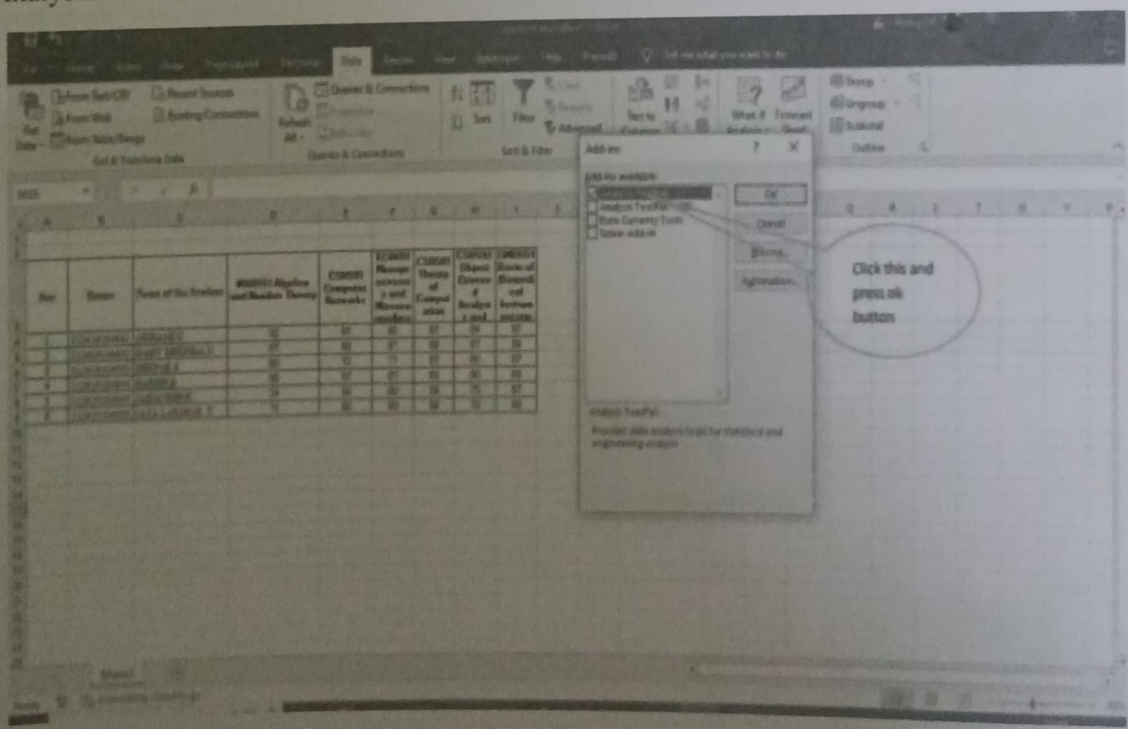
Skewness measures the degree of asymmetry of the distribution, while Kurtosis measures the degree of peaked ness and flatness of a distribution. Skewness is the third measure of moments, while kurtosis is the fourth measure of moments. The value of both Skewness and Kurtosis ranges from -infinity to +infinity.

PROCEDURE:

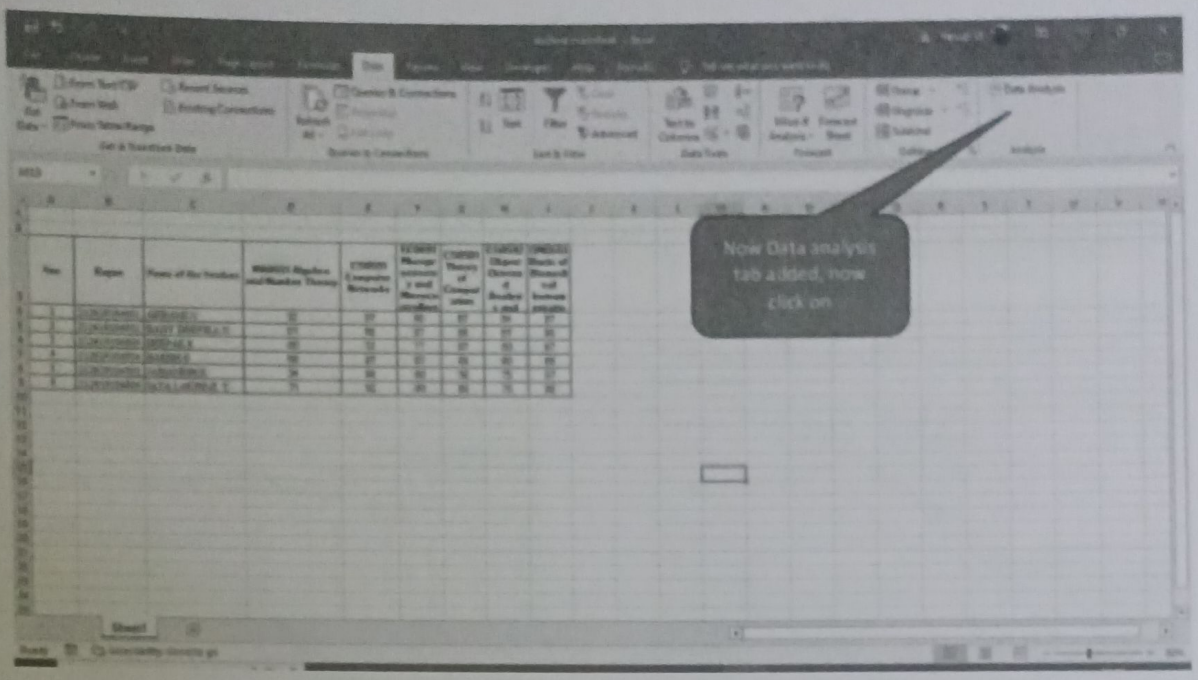
Step 1 : Start Ms Excel application in Ms- office.

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

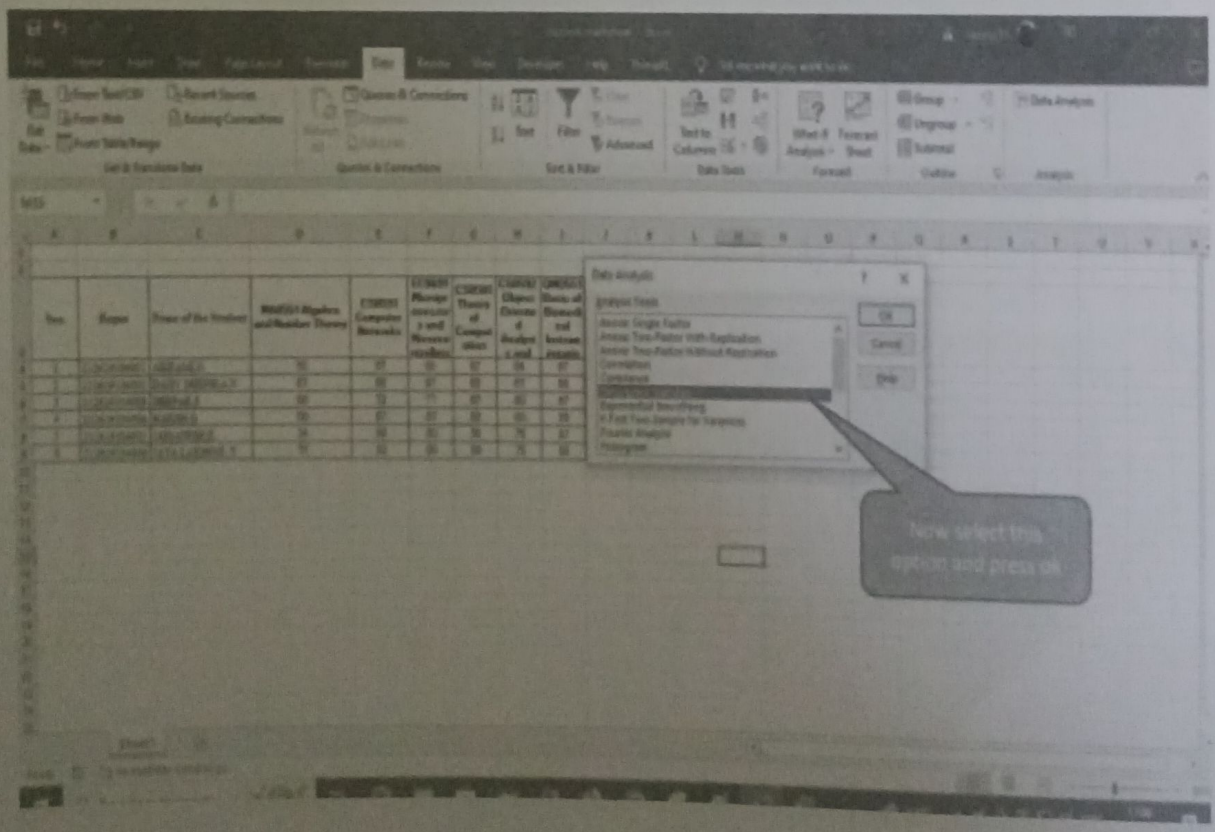
Step 3 : If you haven't already installed the Analysis ToolPak , Click the Microsoft Office button, then click on the Excel Options , and then select Add-Ins , Click Go, check the Analysis ToolPak box, and click Ok



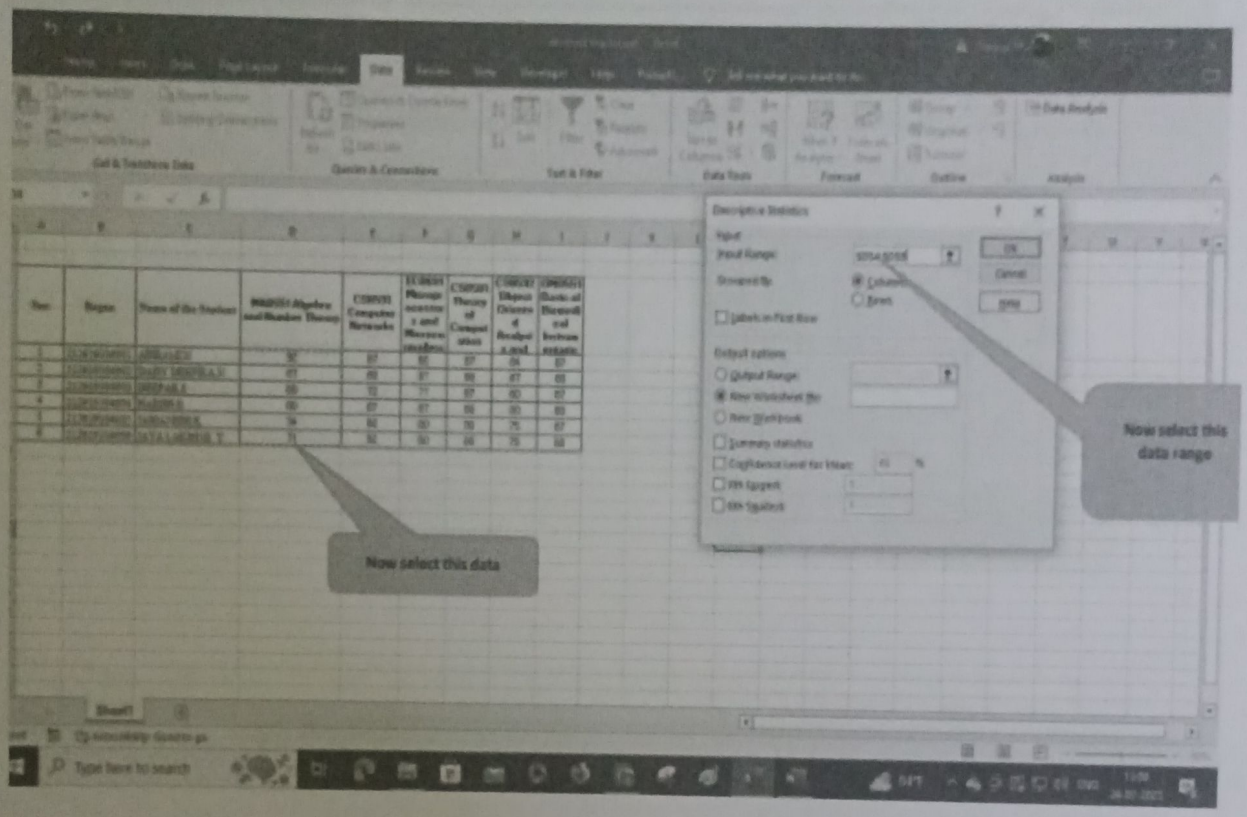
Step 4 : Select Data tab, then click on the Data Analysis option, then selects Descriptive Statistics from the list and Click Ok. [Data tab >> Data Analysis >> Descriptive Statistics]



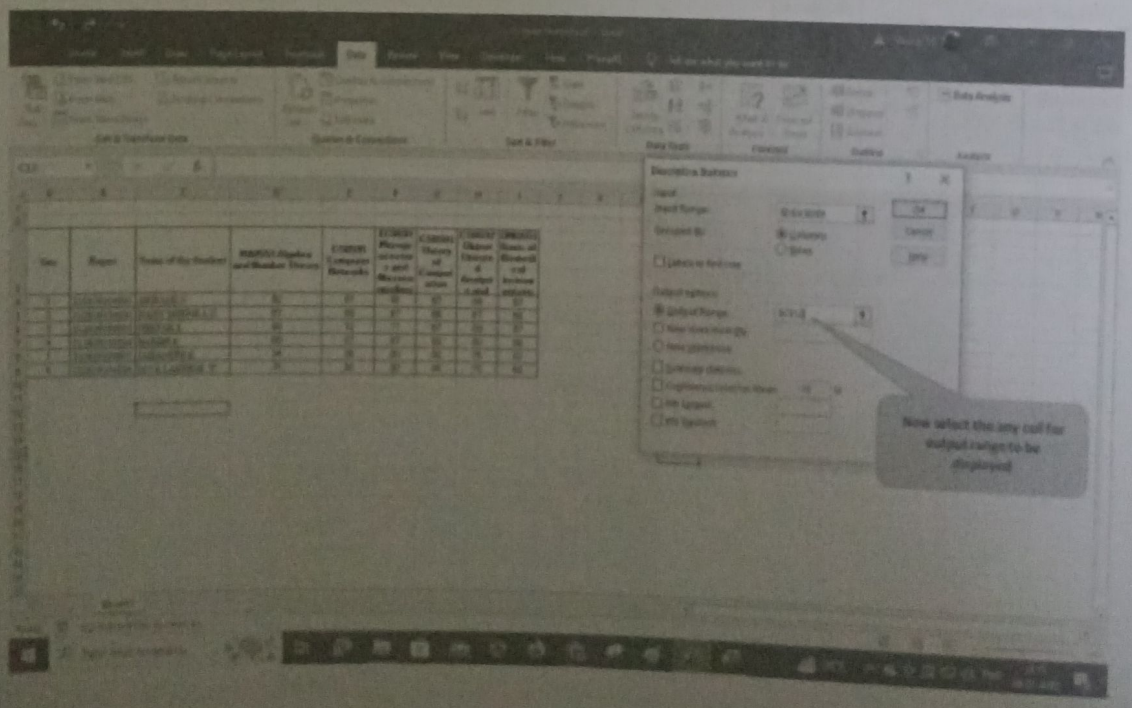
Step 5: In the Input Range we select the data, and then select Output Range where you want the output to be stored. If you don't specify the output range it will throw output in the new worksheet.



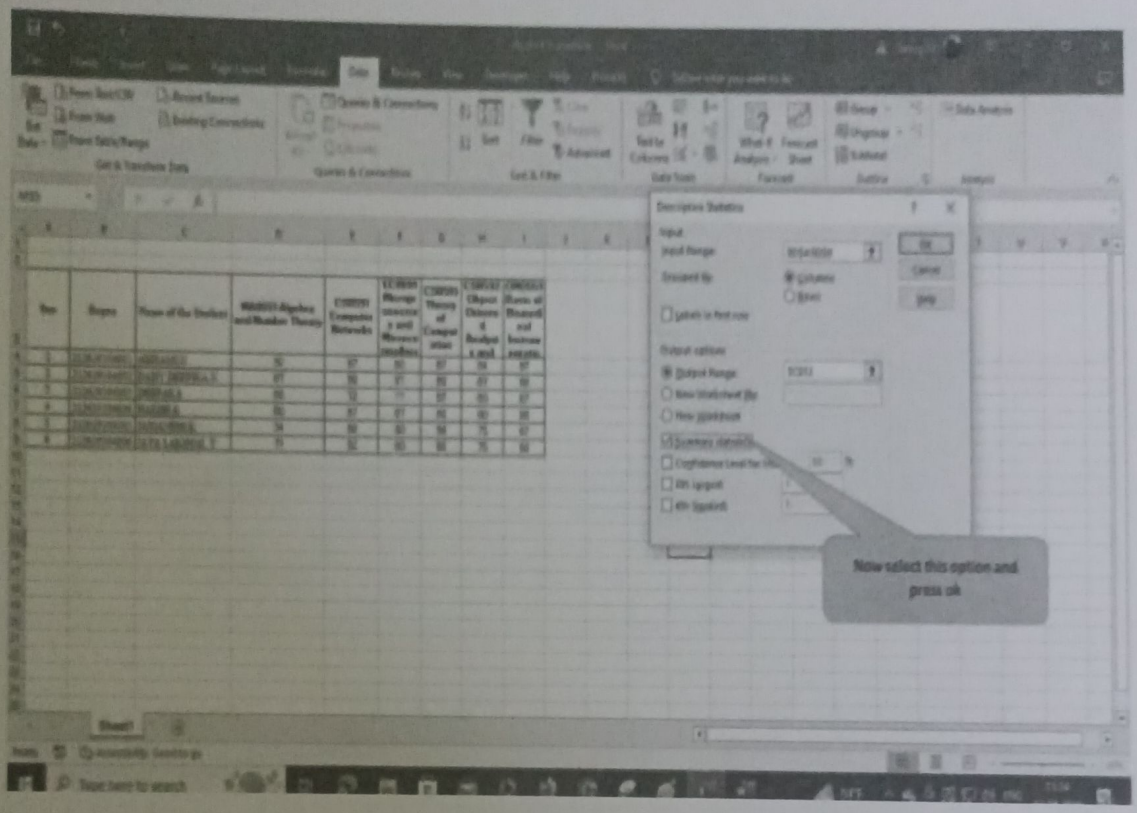
Step 6 : Check Summary Statistics and Confidence Level for Mean options. By default the confidence level is 95%. You can change the level as per the hypothesis standard of study.



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.



SAMPLE:

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

OUTPUT:

MEAN	74
STANDARD ERROR	8.5127
MEDIAN	80
SD	80
SIMPLE VARIANCE	434.8
MINIMUM	34
MAXIMUM	94
SUM	444
COUNT	6

VIVA QUESTIONS:

1. What are the various statics formulas?

   Mean Variance

   Median

   Mode

2. What is Statics?

   Statics is a branch of mathematics that deals with collecting, analyzing, interpreting, presenting.

3. What is analytics?

   Analytics is the systematic computational analysis of data on statistics.

RESULT:

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

EXP NO: 3(A)	PERFORM T-TEST
DATE: 24/6/24	

AIM:

To Perform T-test operations using MS-EXCEL

THEORY:

The t-Test is used to test the null hypothesis that the means of two populations are equal. Below you can find the study hours of 6 female students and 5 male students.

H0: $\mu 1 - \mu 2 = 0$

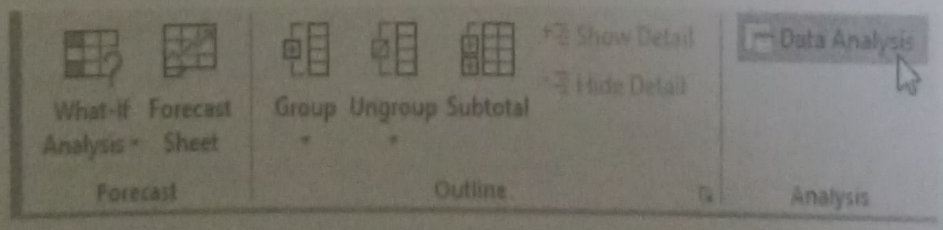
H1: $\mu 1 - \mu 2 \neq 0$

	A	B	C
1	Female	Male	
2	26	23	
3	25	30	
4	43	18	
5	34	25	
6	18	28	
7	52		
8			

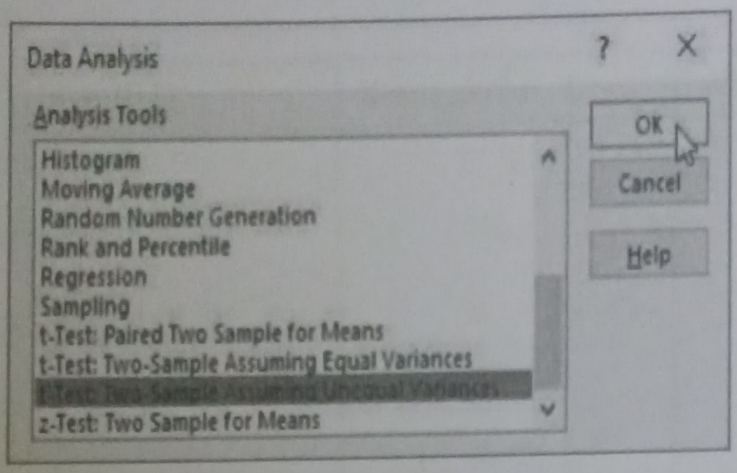
PROCEDURE:

1. First, perform an F-Test to determine if the variances of the two populations are equal. This is not the case.

2. On the Data tab, in the Analysis group, click Data Analysis.



3. Select t-Test: Two-Sample Assuming Unequal Variances and click OK.

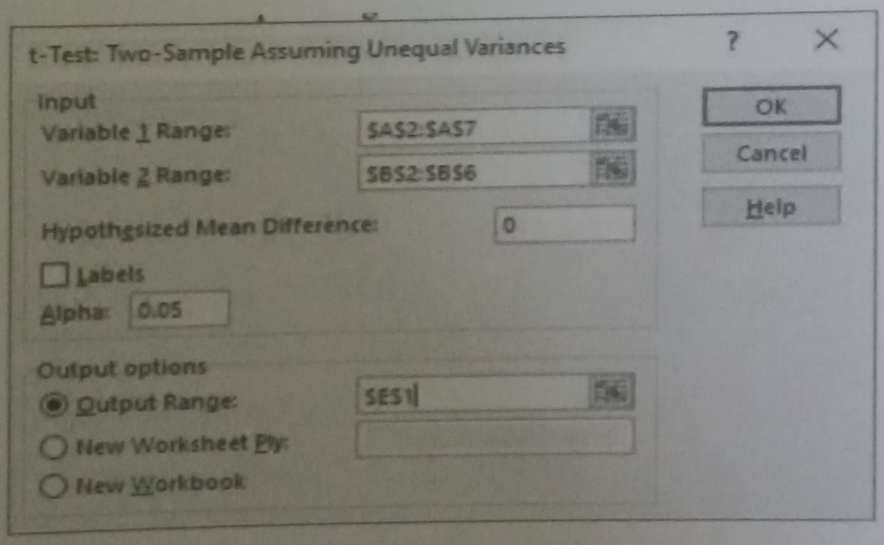


4. Click in the Variable 1 Range box and select the range A2:A7.

5. Click in the Variable 2 Range box and select the range B2:B6.

6. Click in the Hypothesized Mean Difference box and type 0 (H0: $\mu 1 - \mu 2 = 0$).

7. Click in the Output Range box and select cell E1.



8. Click OK.

SAMPLE OUTPUT:

E	F	G
t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	33	24.8
Variance	160	21.7
Observations	6	5
Hypothesized Mean Difference	0	
df	7	
t Stat	1.47260514	
P(T<=t) one-tail	0.092170202	
t Critical one-tail	1.894578605	
P(T<=t) two-tail	0.184340405	
t Critical two-tail	2.364624252	

OUTPUT:

	Variable 1	Variable 2
MEAN	33	24.8
VARIANCE	166	21.7
OBSERVATION	6	5
DF	7	
STAND	1.4726	
P(T<=t) ONETAIL	0.0921	
P(T<=t) TWOTAIL	2.3642	

VIVA QUESTIONS:

1. What is T-test?

A T-test compares the means of two group to see if they are significantly different.

2. What are the types of T-test?

Independent T-test

Paired T-test

One sample T-test

RESULT:

Thus the T-test operations are performed on data using MS-EXCEL.

EXP NO: 3(B)	PERFORM Z-TEST
DATE: 24/6/24	

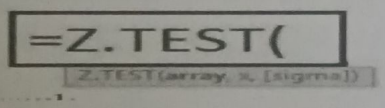
AIM:

To Perform Z-test operations using MS-EXCEL

THEORY:

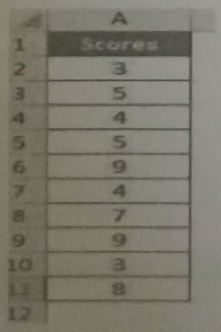
The Z.TEST function is one such hypothesis test function. It tests the mean of the two sample data sets when the variance is known and the sample size is large. The sample size should be >= 30. Otherwise, we need to use T-TEST. To Z.TEST, we need to have two independent data points that are not related to each other or do not affect each other data points. We should normally distributed the data.

Syntax The Z.TEST is the built-in function in Excel. Below is the formula of the Z.TEST function in Excel.



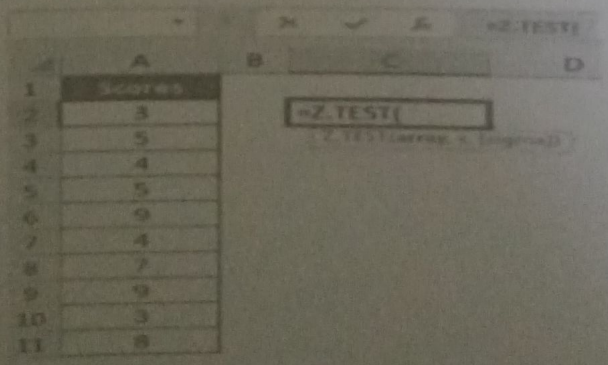
PROCEDURE 1: Using the Z.Test Formula

Step 1: Open data set

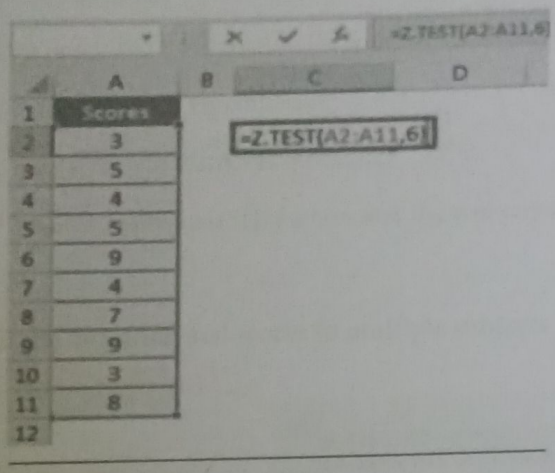


	A
1	Scores
2	3
3	5
4	4
5	5
6	9
7	4
8	7
9	9
10	3
11	8
12	

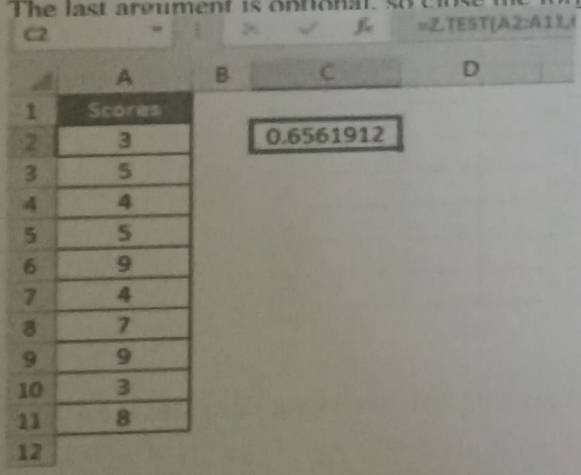
Step 2: Open the Z.TEST formula in an Excel cell.



Step 3: Select the array as scores, A2 to A11. Downloaded by Diana ju (ddianajulie@gmail.com) lOMoARcPSD|43569983 The next argument is "x." Since we have already assumed the hypothesized population mean is 6,apply this value to this argument.

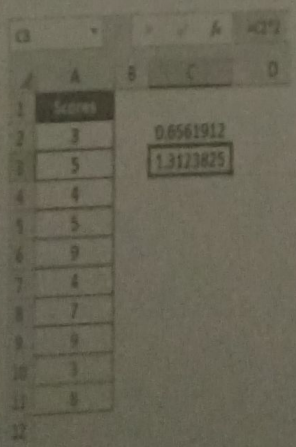


3) The last argument is optional, so close the formula to get the Z.TEST value.



Sample Output:

4) It is a one-tailed Z TEST value to get the two-tailed Z.TEST value to multiply this value by 2.



PROCEDURE 2: Z TEST Using Data Analysis Option

Z.TEST using the "Data Analysis" option in Excel. To compare two means when the variance is known, we use Z.TEST. We can frame two hypotheses here. One is the "Null Hypothesis.

" Another one is the "Alternative Hypothesis" below is the equation of both these hypotheses.

H0: $\mu 1 - \mu 2 = 0$ (Null Hypothesis)

H1: $\mu 1 - \mu 2 \neq 0$ (Alternative Hypothesis)

The alternative hypothesis (H1) states that the two population means are not equal.

We will use two students' scores in multiple subjects for this example.

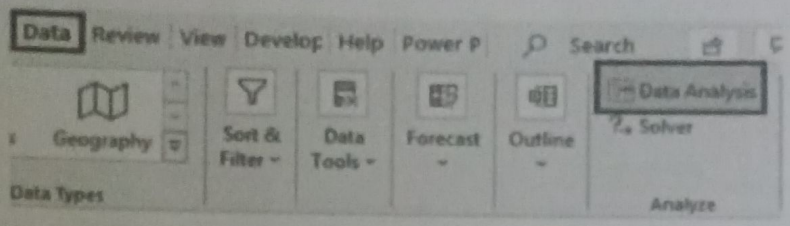
	A	B
1	Student 1	Student 2
2	93	79
3	68	66
4	82	73
5	51	72
6	87	80
7	82	74
8	72	95
9	71	93
10	92	99
11	73	97

Step 1: First, we need to calculate the variables for these two values using the VAR.P function.

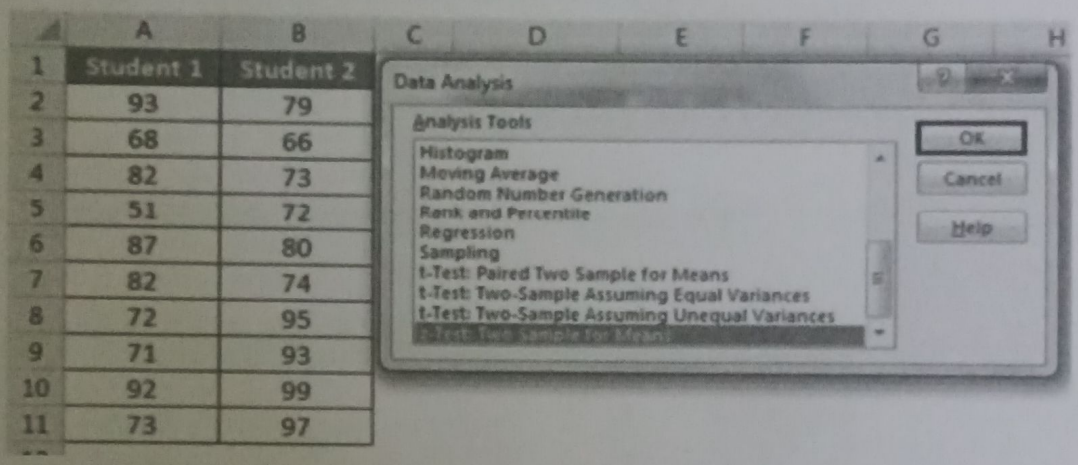
E2 =VAR.P(A2:A11)

	A	B	C	D	E
1	Student 1	Student 2			
2	93	79		Variance 1	146.49
3	68	66		Variance 2	131.16
4	82	73			
5	51	72			
6	87	80			
7	82	74			
8	72	95			
9	71	93			
10	92	99			
11	73	97			

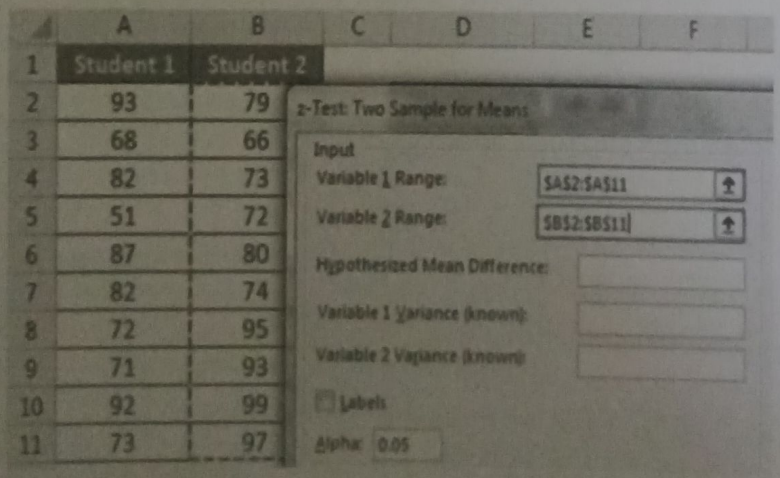
- **Step 2:** Go to the "Data" tab and click "Data Analysis."



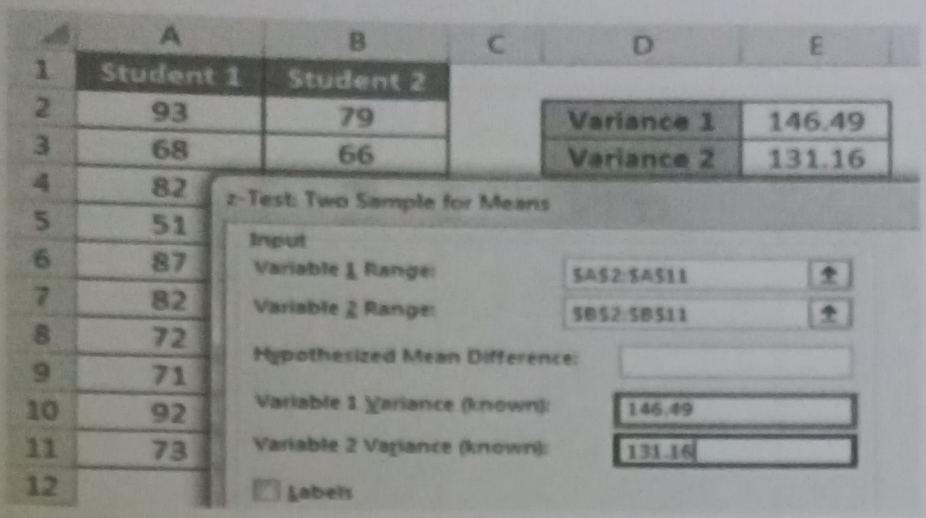
Scroll down and select z-Test: Two Sample for Means" and click on "OK."



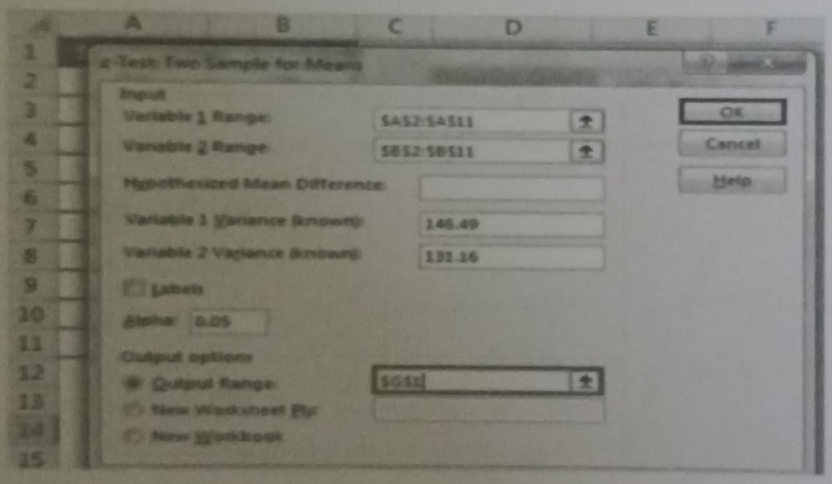
Step 3: For the "Variable 1 Range," select "Student 1" scores. For the "Variable 2 Range," select "Student 2" scores.



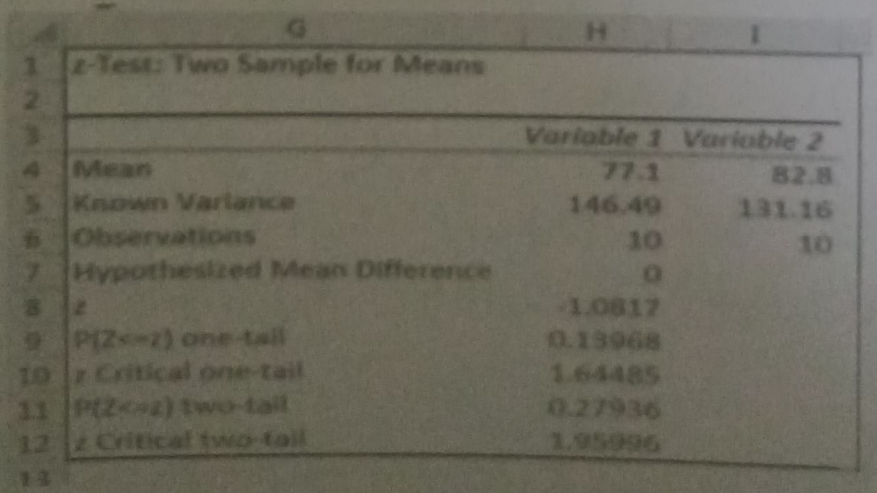
Step 4: For "Variable 1 Variance(known)," select "Student 1" variance score, and for "Variable 2 Variance(known)," select "Student 2" variance score.



Step 5: Select the "Output Range" as a cell and press "OK."



SAMPLE OUTPUT:



z-Test: Two Sample for Means

	Variable 1	Variable 2
Mean	77.1	82.8
Known Variance	146.49	131.16
Observations	10	10
Hypothesized Mean Difference	0	
z	-1.0817	
P(Z<=z) one-tail	0.13968	
z Critical one-tail	1.64485	
P(Z<=z) two-tail	0.27936	
z Critical two-tail	1.95996	

OUTPUT:

	Variable 1	Variable 2
MEAN	771	22.8
KNOWN VARIANCE	146.49	131.6
OBSERVATION	10	10
z	11.0819	
P(z<=2) ONE TAIL	0.136	
P(z<=2) TWO TAIL	0.279	
Z CRITICAL TWO TAIL	1.95996	

VIVA QUESTIONS:

1. **What is Z-score?**

A z-score measure how many standard deviation a data points is from the mean.

$$z = \frac{X - M}{\sigma}$$

2. **What you mean by population and sampling?**

Population :- The entire group of individuals or y terms that you want to study.

Sample :- A subset of the population selected for the actual study.

RESULT:

Thus the Z-test operations are performed on data using MS-EXCEL.

EXP NO: 3(B)	PERFORM ANOVA ON DATA
DATE: 24/6/24	

AIM:

To Perform Anova operations on data using MS-EXCEL

THEORY:

ANOVA is an analysis of variance. A single factor or one-way ANOVA is used to test the null hypothesis that the means of several populations are all equal.

Below you can find the salaries of people who have a degree in economics, medicine or history.

H0: $\mu 1 = \mu 2 = \mu 3$

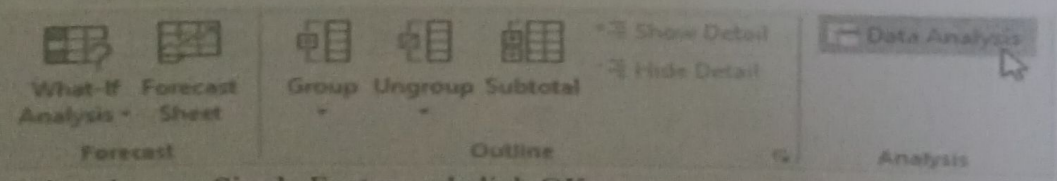
H1: at least one of the means is different.

	A	B	C	D
1	economics	medicine	history	
2	42	69	35	
3	53	54	40	
4	49	58	53	
5	53	64	42	
6	43	64	50	
7	44	55	39	
8	45	56	55	
9	52		39	
10	54		40	
11				

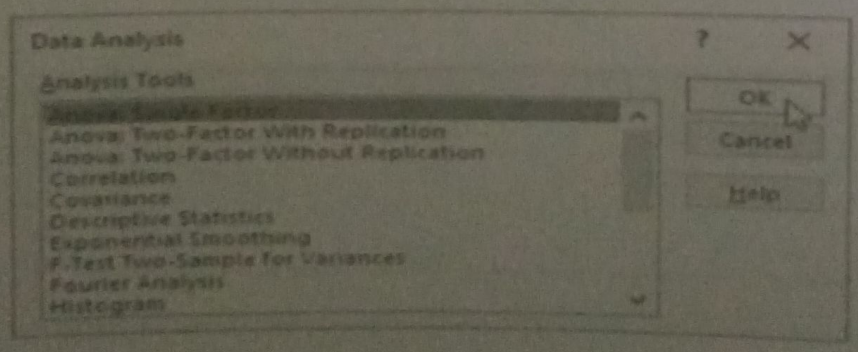
PROCEDURE:

To perform a single factor ANOVA, execute the following steps.

1. **On the Data tab, in the Analysis group, click Data Analysis.**

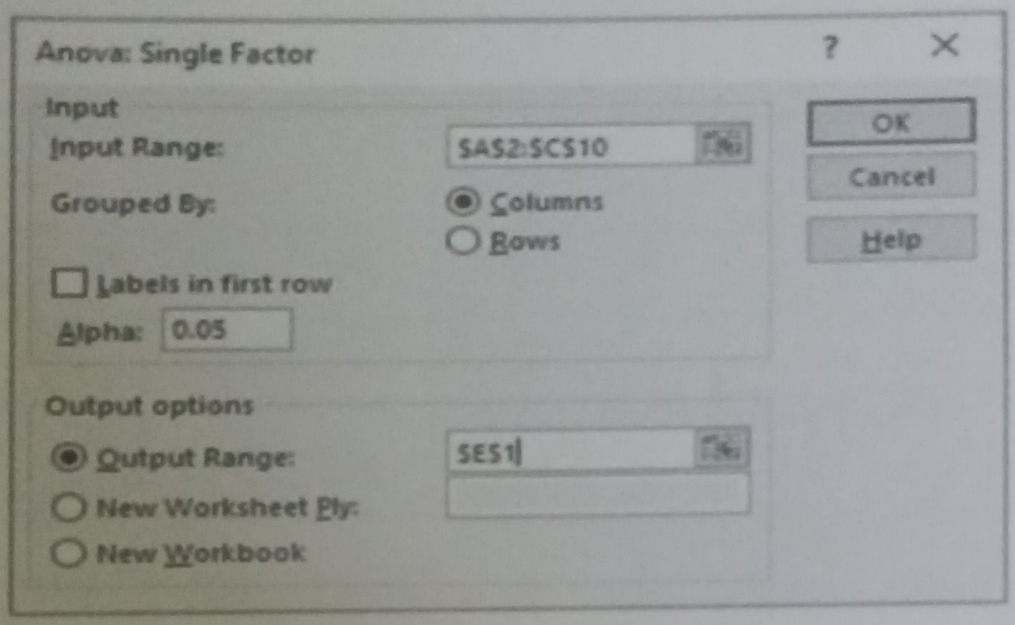


2. **Select Anova: Single Factor and click OK.**



3. Click in the Input Range box and select the range A2:C10.

4. Click in the Output Range box and select cell E1.



5. Click OK. output:

SAMPLE OUTPUT:

E	F	G	H	I	J	K
Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	9	435	48.33333	23.5		
Column 2	7	420	60	32.33333		
Column 3	9	393	43.66667	50.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1085.84	2	542.92	15.19623	7.16E-05	3.443357
Within Groups	786	22	35.72727			
Total	1871.84	24				

OUTPUT:

SUMMARY

GROUP	COUNT	SUM	AVG	VARIANCE
1	9	435	48.33	33.5
2	7	420	486	32.5

ANOVA

SOURCE OF VARIANCE	SS	OF	MS	F	P-VALUE
BETWEEN GROUPS	1035.84	2	54.42	15.19	7.16
WITHIN GROUPS	736	22	25.5		

VIVA QUESTIONS:

1.What is one way Anova?

One way ANOVA (Analysis of variance) is a statistical test used to determine if there are significant different between mean.

2.Differentiate single and two factors variable?

Single factor variable (one way ANOVA) examine the effect of ~~two~~ one independent variable on a dependent variable across different levels.

Two factor variable (Two way ANOVA) examine the effect of two independent variable as well as their interaction on a dependent variable.

RESULT:

Thus the Anova operations are performed on data using MS-EXCEL.

EXP NO: 4(A)	PERFORM DATA PRE-PROCESSING OPERATION HANDLING MISSING DATA
DATE: 1/7/24	

AIM:

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

THEORY:

To handle missing data in excel using some functions such as ISERROR and VLOOKUP Functions Using NOT, ISNUMBER, MATCH Functions

You can see that some persons are missing on the day of taking the vaccine and we want to find out the missing persons. Here we will use ISERROR, VLOOKUP functions to deal with the missing data.

PROCEDURE 1:

Step1: • First of all, create a column and apply the following formula to the selected cell.

Dataset Overview

ID	Registered Person	Present Person
1602	Mike	Adam
1725	Adam	Antony
1235	David	John
1823	Jonathon	Hamlet
1290	Lilly	Lilly
1692	Bob	Collins
1286	Collins	
1427	John	
1735	Hamlet	
1882	Antony	

=ISERROR(VLOOKUP(C5,$D$5:$D$14,1,0))

The VLOOKUP function considers C5 as the lookup_value and $D$5:$D$14 as lookup array,

1 as col_index_num, and 0 as range lookup. The ISERROR function returns TRUE as it finds an error or FALSE where it doesn't find an error.

So, when the LOOKUP function doesn't find a value and makes an error, the ISERROR function returns TRUE, otherwise FALSE.

E5 =ISERROR(VLOOKUP(C5,$D$5:$D$14,1,0))

Using ISERROR and VLOOKUP Functions

ID	Registered Person	Present Person	Missing Person
1602	Mike	Adam	TRUE
1725	Adam	Antony	
1235	David	John	
1823	Jonathon	Hamlet	
1290	Lilly	Lilly	
1692	Bob	Collins	
1286	Collins		
1427	John		
1735	Hamlet		
1882	Antony		

Then, use the Fill Handle tool to Autofill the formula for the next cells and you will get the output for the missing data.

E5 =ISERROR(VLOOKUP(C5,$D$5:$D$14,1,0))

Using ISERROR and VLOOKUP Functions

ID	Registered Person	Present Person	Missing Person
1602	Mike	Adam	TRUE
1725	Adam	Antony	FALSE
1235	David	John	TRUE
1823	Jonathon	Hamlet	TRUE
1290	Lilly	Lilly	FALSE
1692	Bob	Collins	TRUE
1286	Collins		FALSE
1427	John		FALSE
1735	Hamlet		FALSE
1882	Antony		FALSE

Here, you can also apply Conditional Formatting to highlight the missing data. For this, just select the column> go to Conditional Formatting> select Highlight Cells Rules> select Equal to> select the cell (i.e. E6)> choose Fill Color> click OK.

Using ISERROR and VLOOKUP Functions

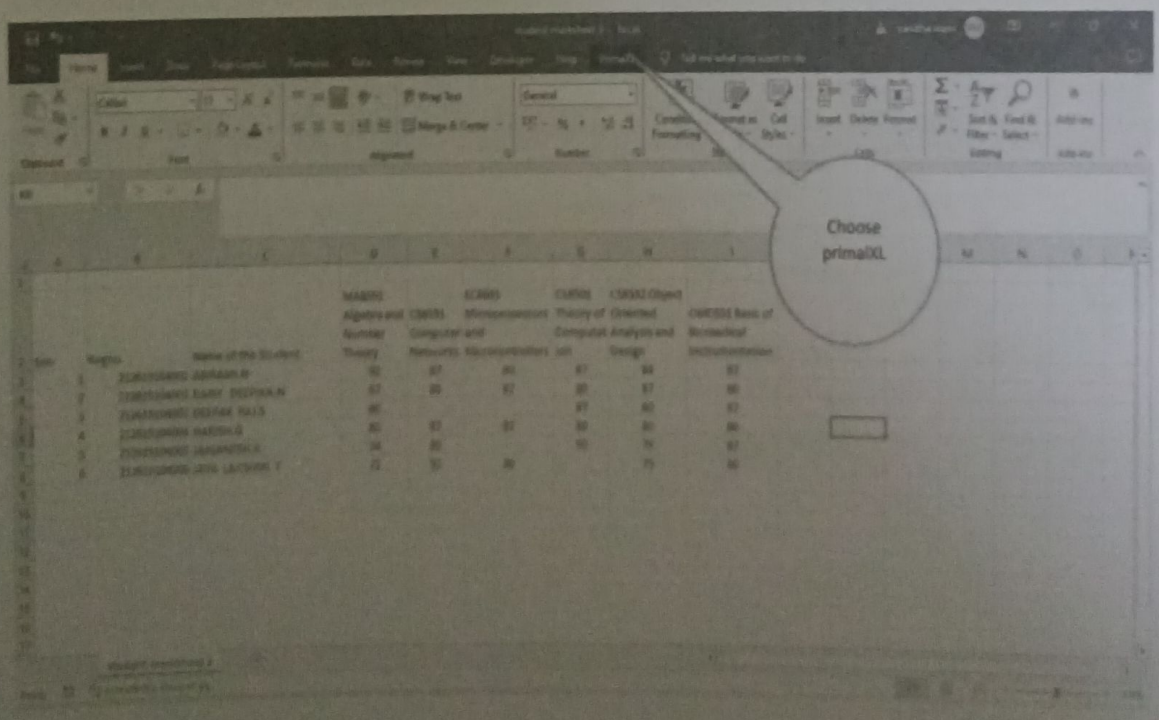
ID	Registered Person	Present Person	Missing Person
1602	Mike	Adam	TRUE
1725	Adam	Antony	FALSE
1235	David	John	TRUE
1828	Jonathon	Hamlet	TRUE
1290	Lilly	Lilly	FALSE
1692	Bob	Collins	TRUE
1286	Collins		FALSE
1427	John		FALSE
1735	Hamlet		FALSE
1882	Antony		FALSE

PROCEDURE 2:

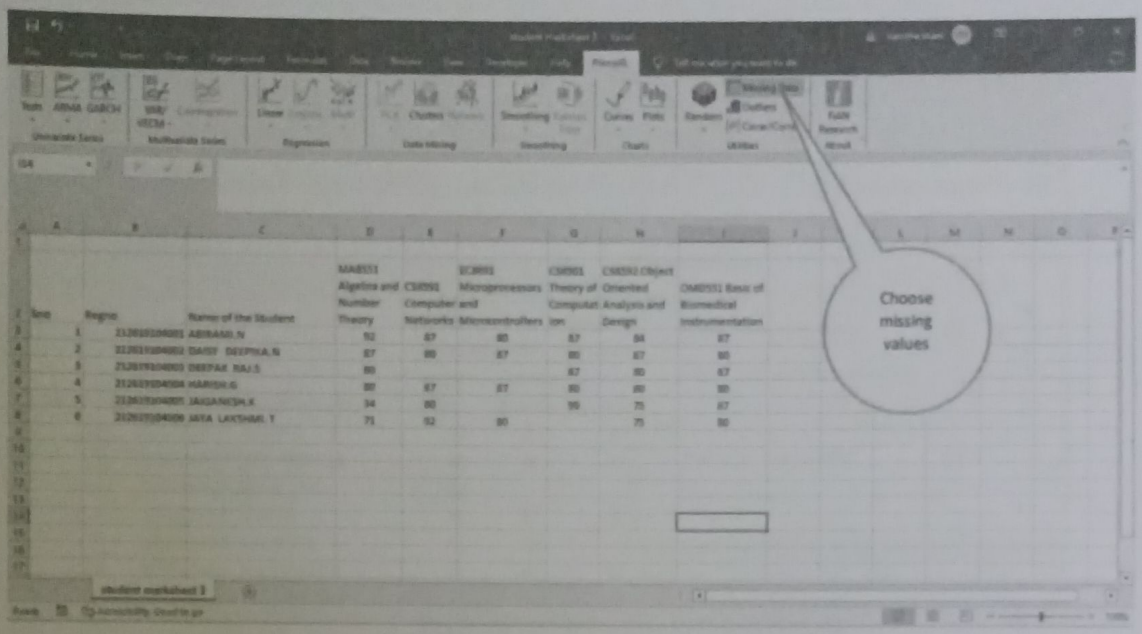
Step 1 : Start Ms Excel application in Ms- office.

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

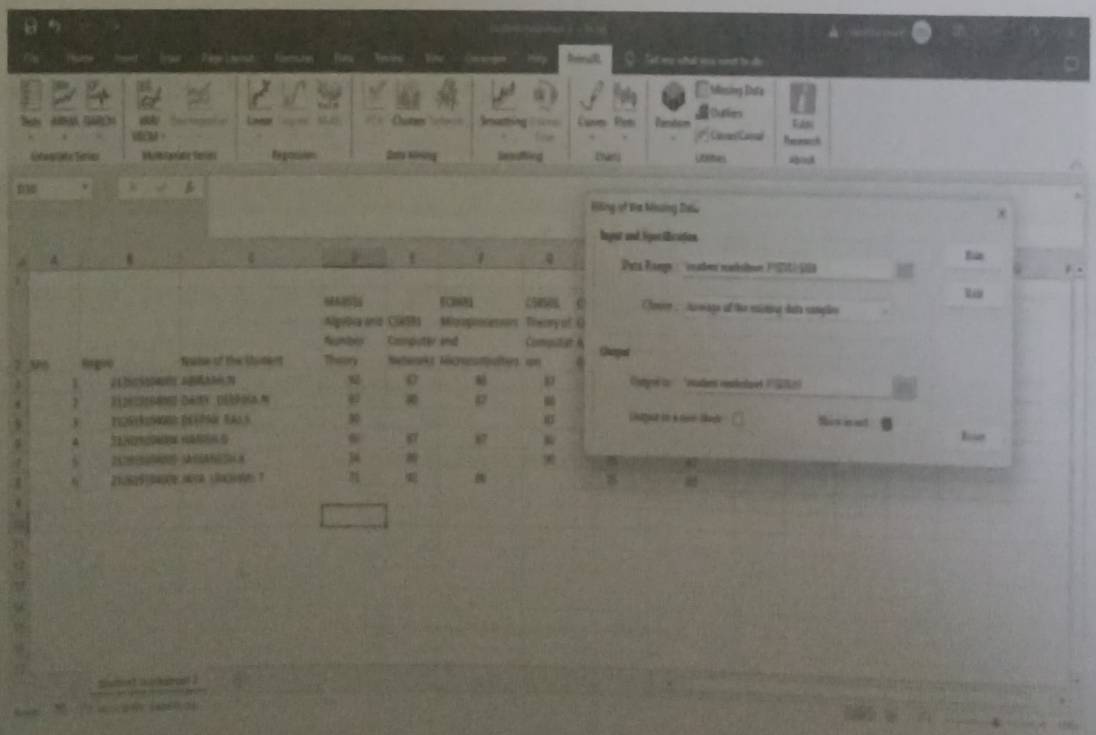
Step 3 : If you haven't already installed the PrimaXL Addin, install it. Click the PrimaXL tab , choose missing



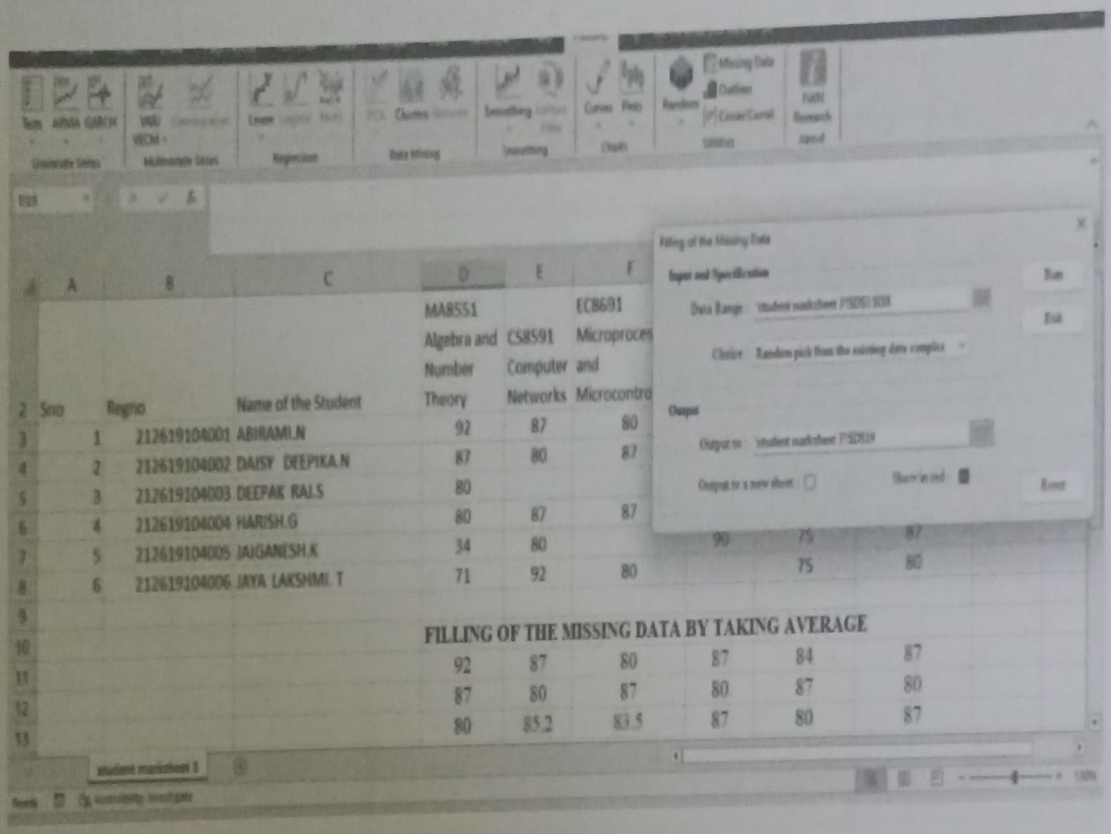
Step 4 : In the Input Range we select marks of all subjects with missing values and select the Choice as "filling of the missing data by taking average" or " filling of the missing data by random pick".



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.

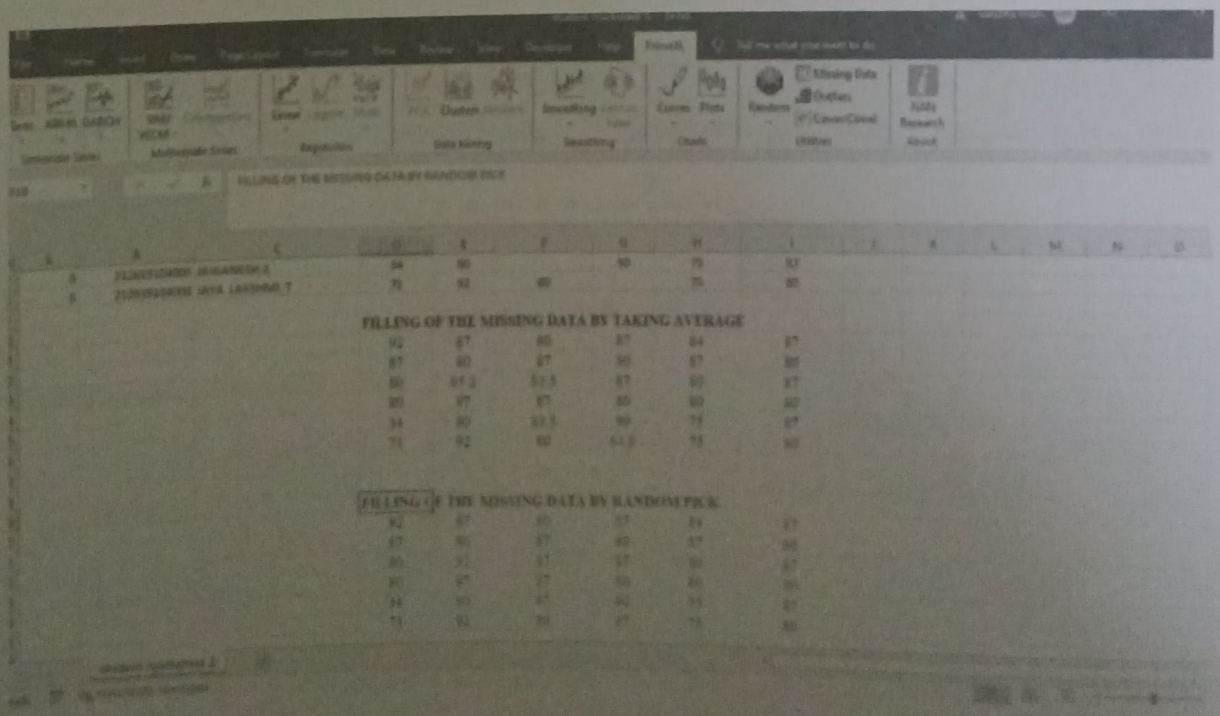


Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microproces and Microcontro			
1	212619104001	ABIRAMI.N	92	87	80			
2	212619104002	DAISY DEEPIKA.N	87	80	87			
3	212619104003	DEEPAK RAJ.S	80					
4	212619104004	HARISH.G	80	87	87			
5	212619104005	JAIGANESH.K	34	80		90	75	87
6	212619104006	JAYA LAKSHMI. T	71	92	80		75	80

FILLING OF THE MISSING DATA BY TAKING AVERAGE

92	87	80	87	84	87
87	80	87	80	87	80
80	85.2	83.5	87	80	87

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:

FILLING THE MISSING DATA BY TAKING EASY

92	87	80	87	84	87
87	80	87	80	80	80
80	85.2	83.5	87	80	87
80	80	83.5	80	75	80

FILLING THE MISSING DATA BY RANDOM PICK

97	87	87	86	84	87
92	92	67	87	80	32
80	35	83	87	34	83
80	34	85	86	80	86

VIVA QUESTIONS:

1.What you mean by data reduction?

The process of reducing the volume of data while processing its efficient information.

2. What is called data reduction.

Technique to simplify database, making them more manageable and efficient for analysis.

EXP NO: 4(B)	PERFORM DATA PRE-PROCESSING OPERATION NORMALIZATION
DATE: 1/7/24	

AIM:

To perform normalization on the dataset using MS-EXCEL.

THEORY:

- Normalization refers to scaling data so that the minimum value is equal to 0 and the maximum value is equal to 1. This process is often referred to as min-max scaling. Standardization is often also referred to as normalization – in this case, the dataset is modified so that the mean is equal to 0 and the standard deviation is equal to 1.
- Standardization and normalization allow you to work with data of different scales more effectively. Imagine having a dataset that compares a person's income and height in centimeters. Likely, these values will be in completely different scales.
- Because of this, looking at their means or standard deviations would not be comparable. Standardizing data is quite simple using built-in operators and functions, such as the MIN() and MAX() functions. Mathematically,

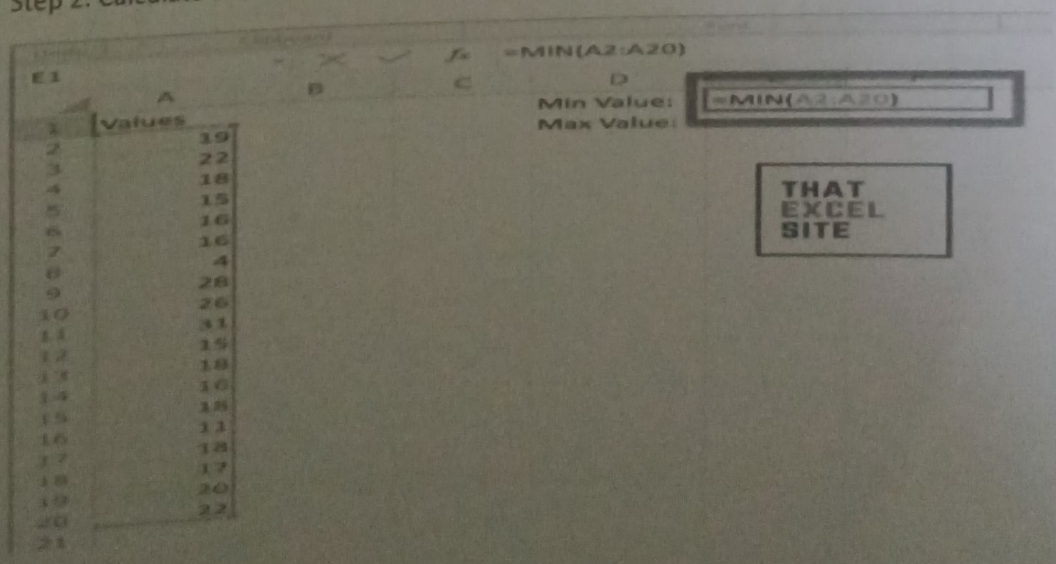
we normalize value using the formula below:

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

PROCEDURE :

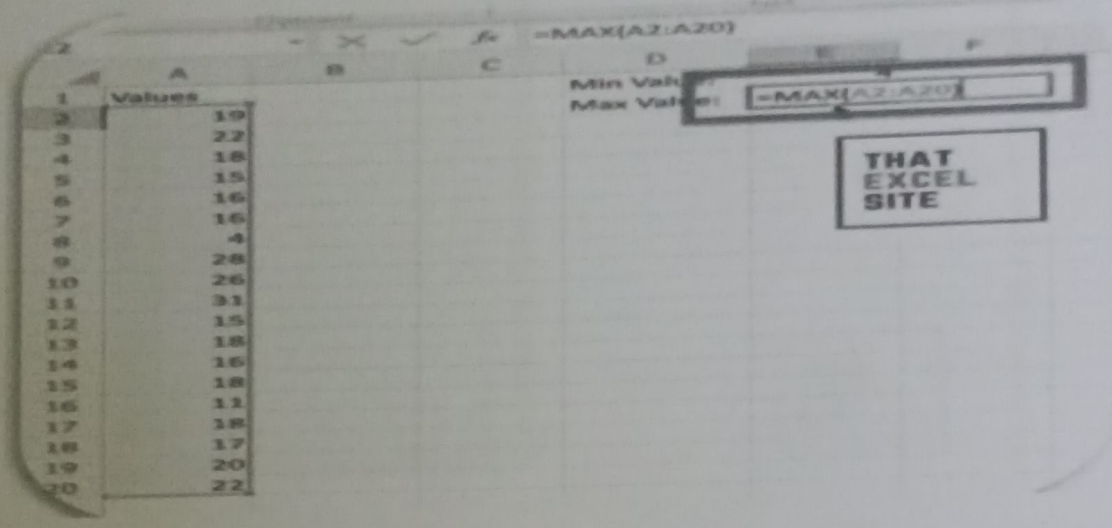
Step 1: Load the dataset you want to normalize

Step 2: Calculate the minimum value in the dataset using the MIN() function

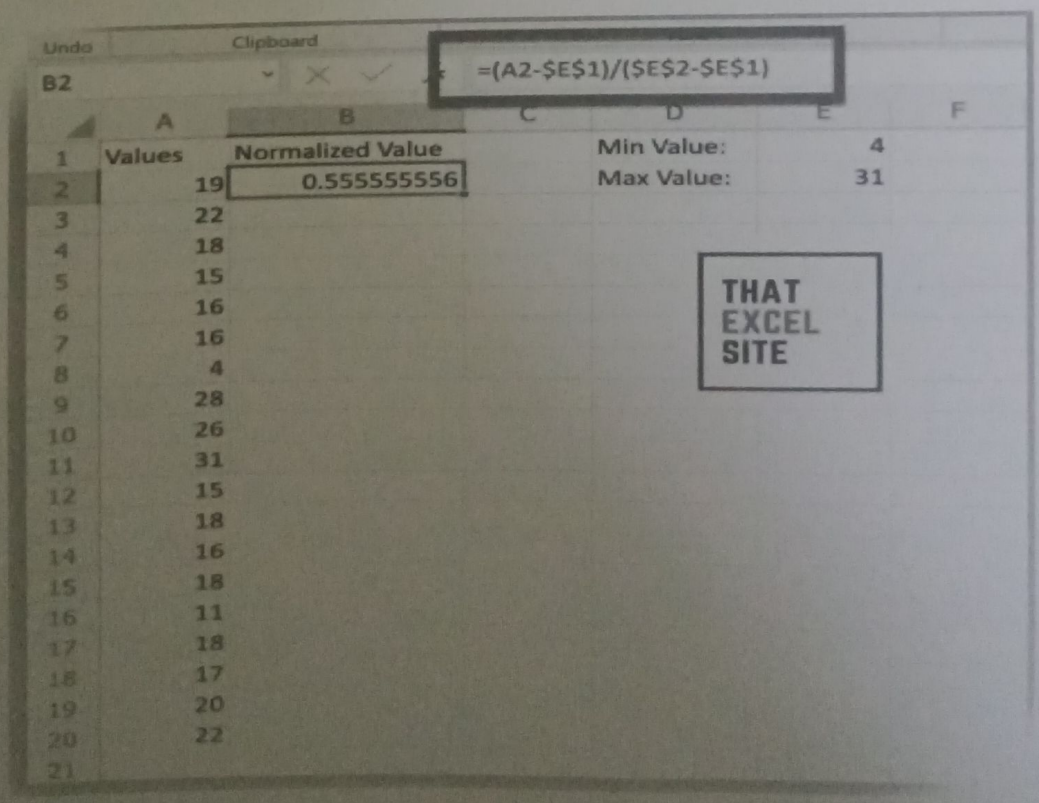


Calculate the minimum value of the dataset Use the MIN() function to calculate the minimum value of the dataset. In this case, this returns 4. Step

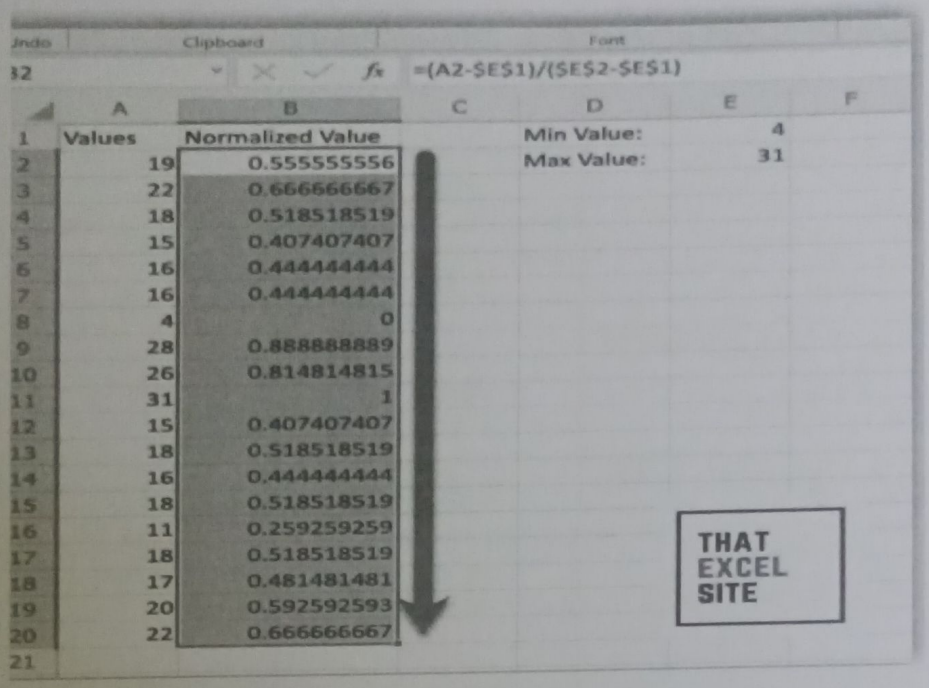
Step 3: Calculate the maximum value in the dataset using the MAX() function



Step 4: Normalize the first record by subtracting the minimum value from the observation and dividing by the range of the dataset



Step 5: Drag the fill handle down to normalize all records in the dataset



Undo | Clipboard | Font

B2 =(A2-$E$1)/($E$2-$E$1)

	A	B	C	D	E	F
1	Values	Normalized Value		Min Value:	4	
2	19	0.555555556		Max Value:	31	
3	22	0.666666667				
4	18	0.518518519				
5	15	0.407407407				
6	16	0.444444444				
7	16	0.444444444				
8	4	0				
9	28	0.888888889				
10	26	0.814814815				
11	31	1				
12	15	0.407407407				
13	18	0.518518519				
14	16	0.444444444				
15	18	0.518518519				
16	11	0.259259259				
17	18	0.518518519				
18	17	0.481481481				
19	20	0.592592593				
20	22	0.666666667				
21						

VIVA QUESTIONS:

1. **What you mean by normalization?**

Normalisation is the process of organising data in a database. Scaling data for a standard range (eg:- 0 to 1).

2.Define 3NF?

A table design where all attribute depends only on the primary key, eliminating transitive dependencies.

OUTPUT:

No	values	Normalized value	
1	18	0.505	
2	10	0.338	
3	18	0.222	MIN VALUE : 4
4	11	0.5125	MAX VALUE : 31
5	18	0.666	
6	17	0.5785	
7	20	0.444	

RESULT:

Thus the normalizing of data in excel was executed successfully.

EXP NO: 5	PLOTTING FUNCTIONS ON THE DATA SET
DATE: 8/7/24	

AIM:

To apply and explore various plotting functions on the data set using MS-EXCEL.

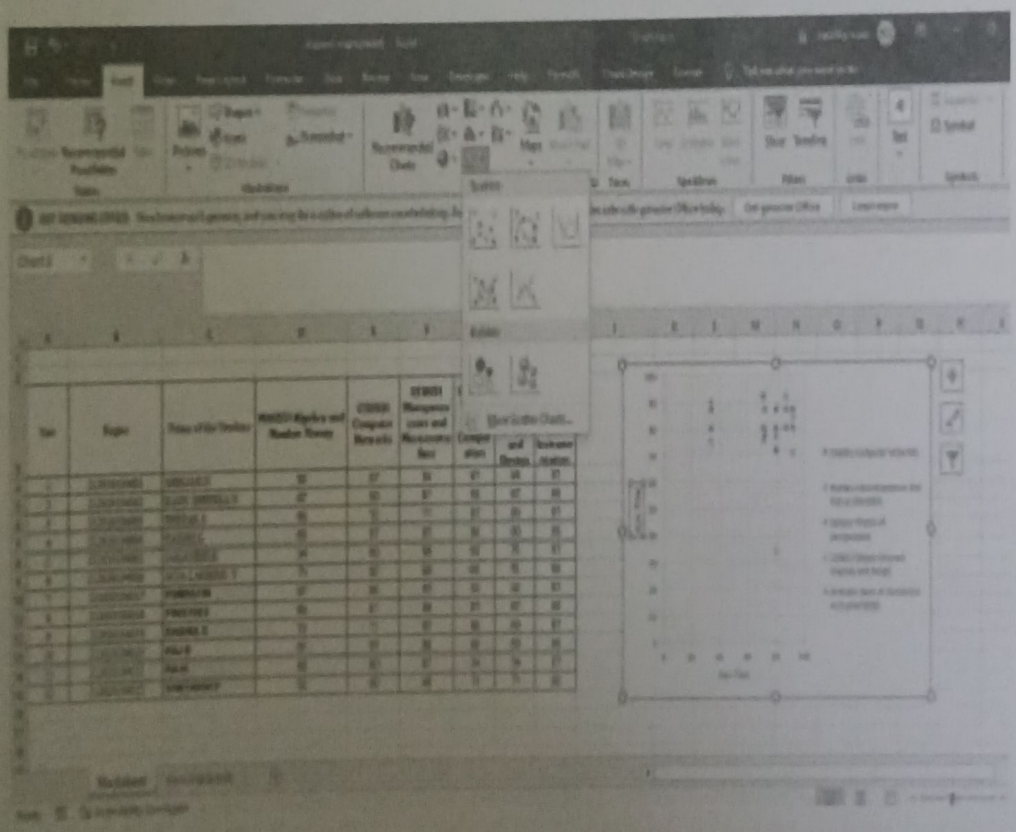
PROCEDURE:

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

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

Step 3 : select the data for which a chart is to be created.

Step 4 : In the INSERT menu, select Recommended Charts.



Step 5 : Choose any chart from the list of charts Excel recommends for your data on the Recommended Charts tab, and click it to preview how it will look with your data.

Step 6 : lick on All Charts if you are unable to locate a chart you like.

Step 7 : Click on the chart that you prefer and then click OK.

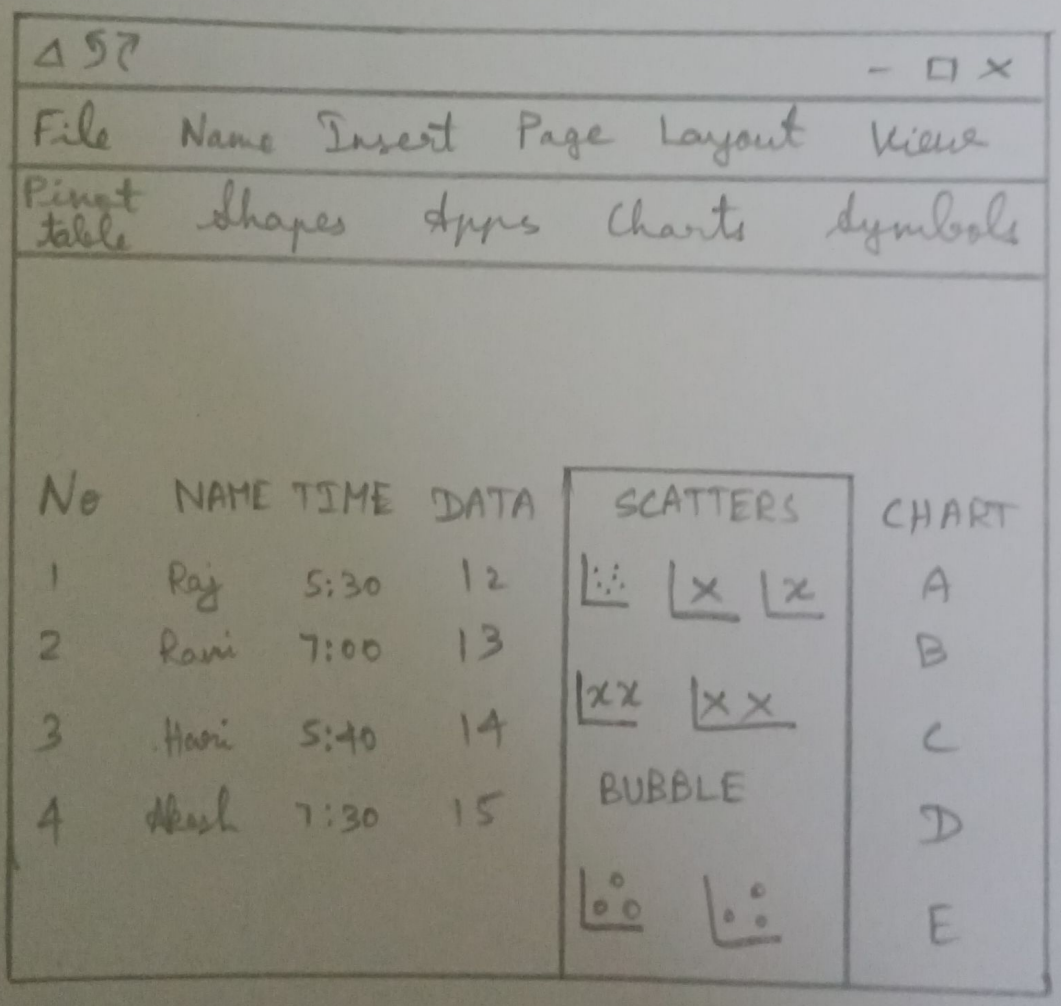
Step 8 : Add chart elements such as axis titles or data labels, customize the appearance of the chart, or change the data displayed in the chart by clicking on Chart Elements, Chart Styles, and Chart Filters in the upper-right corner of the chart.

Step 9 : Click on the chart TOOLS tab on the ribbon to add additional design and formatting capabilities and then click the options you desire under the DESIGN and FORMAT tabs.

Step 10: When you click Ok, you will see the result in the selected output range.

Step 11 : Save the excel file and Close the Ms Excel application.

OUTPUT:



EXP NO: 6	PLOTTING FUNCTIONS ON THE DATA SET USING R-PROGRAMMING
DATE: 20/7/24	

AIM:

To apply and explore various plotting functions on the data set using R-programming.

THEORY:

RStudio is a must-know tool for everyone who works with the R programming language. It's used in data analysis to import, access, transform, explore, plot, and model data, and for machine learning to make predictions on data.

PROCEDURE:

Step 1 : Start the R-studio.

Step 2 : Create datasheet in R-studio application.

Step 3 : select the data for which a chart is to be created.

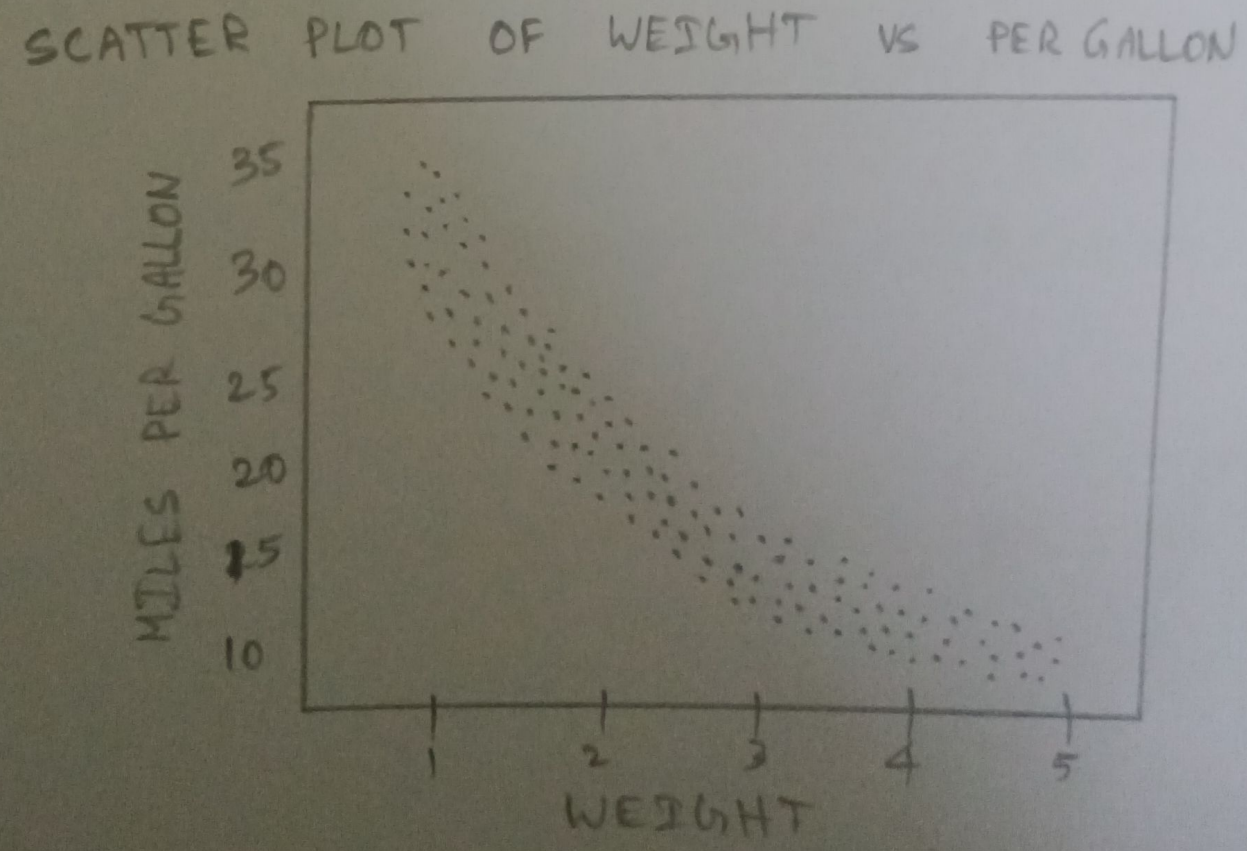
Step 4: Install `library(readxl)`
`> install.packages("ggplot2")`

Step 5 : Write the code to display plots on data set.

Step 6: Draw the various plots and close the application.

PROGRAM:

```
data(mtcars)
plot(mtcars$wt, mtcars$mpg,
  main = "Scatter Plot of Weight vs. Miles Per Gallon",
  xlab = "Weight (1000 lbs)",
  ylab = "Miles Per Gallon",
  pch = 19)
plot(mtcars$wt, mtcars$mpg,
  type = "l",
  main = "Line Plot of Weight vs. Miles Per Gallon",
  xlab = "Weight (1000 lbs)",
```

```
    ylab = "Miles Per Gallon")
barplot(table(mtcars$cyl),
    main = "Bar Plot of Cylinder Counts",
    xlab = "Number of Cylinders",
    ylab = "Count")
hist(mtcars$mpg,
    main = "Histogram of Miles Per Gallon",
    xlab = "Miles Per Gallon",
    ylab = "Frequency",
    breaks = 10)
boxplot(mpg ~ cyl,
    data = mtcars,
    main = "Box Plot of Miles Per Gallon by Cylinder",
    xlab = "Number of Cylinders",
    ylab = "Miles Per Gallon")
```

OUTPUT: