STELLA MARY'S COLLEGE OF ENGINEERING

(Approved by AICTE, New Delhi, Affiliated to Anna University, Chennai & Accredited by NAAC & NBA (CSE & MECH)

Aruthenganvilai, Kallukatti Junction, Azhikal Post Kanyakumari District – 629 202, Tamil Nadu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CCW331 – BUSINESS ANALYTICS

REGULATION – 2021

SEMESTER–V

2022-2023 (ODD/EVEN)

LAB MANUAL

Prepared by,
Mr. Dinesh AP/ Dept. of CSE

Institute Vision:

To emerge as a premiere institution, acknowledged as a center for excellence imparting technical education, creating technocrats who can address the needs of the society through exploration and experimentation and uplift mankind.

Department Vision:

To produce Computer Science professionals who can accomplish path-breaking solutions for a better society, through quality technical education, on gaining the required inter-personal, entrepreneurial and computing skills. .

Institute Mission:

To provide an education that transforms students, through rigorous course-work and by providing an understanding of the needs of the society and the industry.

Department Mission:

- To impart a holistic and experiential learning experience by making use of innovative teaching methodologies.
- To provide optimal technology solutions through collaborative and life-long learning for industry and societal needs with universal ethical values.
- To nurture leadership skills and facilitate various co-curricular and extra-curricular activities to implant the spirit of entrepreneurship.
- To provide industry-institute-interaction opportunities in order to motivate inter-disciplinary research capabilities with an inquiring mind.

Programme Education Objectives (PEOs)

PEO1:

Graduates will be competent in creating innovative technologies through inter-disciplinary research and comprehensive skills sets that are suitable for the global computing industry.

PEO2:

Graduates will be capable of managing leading positions with a broad understanding of the application of ethics in evolving computer-based solutions for the societal needs.

PEO3:

Graduates will imbibe entrepreneurial qualities and develop their career by upgrading their, communication, analytical and professional skills constantly.

Programme Specific Outcomes(PSOs)

PSO1:

Use data management techniques and algorithmic thinking for Software Design and Development practices.

PSO2:

Develop reliable IT solutions based on the expertise in Distributed Applications Development, Web Designing and Networking for various societal needs and entrepreneurial practices ethically.

PSO3:

Manage multidisciplinary environments effectively through their interpersonal and analytical skills and be responsible members and leaders of the society.

PROGRAMME OUTCOMES:

POs

1. Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one‘s own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Sl. No.	Criteria	Marks	Excellent (3) 91% - 100%	Good (2) 71% - 90%	Average (1) 50% - 70%	Poor (0) <50%
1	Observation	3	Gives clear idea about the aim and having the capability of both recording & analyzing the data much easier. (3)	Capability of both recording & analyzing the data much easier but no proper clarification about the objective. (2)	Gives clear idea about the target and has less capability of both recording & analyzing the data. (1)	Gives indistinct idea about the target and has less capability of both recording & analyzing the data & who feel difficult to follow the objectives. (<1)
2	Assessment	3	Have executed the system in an efficient way & make credible and unbiased judgments regarding the conduct of the experiments. (3)	Executed the system with less difficulty & has partial judgements regarding the overall system. (2)	Executed the system with less efficiency and has no judgements regarding the system. (1)	Incomplete system execution & lack of judgments regarding the system. (<1)
3	Submission	4	Followed all the instructions given in the procedure and submitted the observation books in time. (4)	Followed all the instructions given in the procedure with some assisting (3)	Followed some of the instructions given in the procedure & late in submission of note books. (2)	Trying to follow the instructions given in the procedure & late in submission of note books. (<1)

LIST OF EXPERIMENTS:

Use MS-Excel and Power-BI to perform the following experiments using a business data set, and make presentations.

Students may be encouraged to bring their own real-time socially relevant data set.

I Cycle – MS Excel

1. Explore the features of Ms-Excel.
2. (i) Get the input from user and perform numerical operations (MAX, MIN, AVG, SUM, SQRT, ROUND)

ii) Perform data import/export operations for different file formats.

3. Perform statistical operations - Mean, Median, Mode and Standard deviation, Variance, Skewness, Kurtosis
4. Perform Z-test, T-test & ANOVA
5. Perform data pre-processing operations i) Handling Missing data ii) Normalization
6. Perform dimensionality reduction operation using PCA, KPCA & SVD
7. Perform bivariate and multivariate analysis on the dataset.
8. Apply and explore various plotting functions on the data set.

II Cycle – Power BI Desktop

9. Explore the features of Power BI Desktop
10. Prepare & Load data
11. Develop the data model
12. Perform DAX calculations
13. Design a report
14. Create a dashboard and perform data analysis
15. Presentation of a case study

INDEX				
S.NO	**DATE**	**NAME OF EXPERIMENT**	**SIGNATURE**	**REMARK**
1		Explore the features of Ms-Excel.		
2		(i) Get the input from user and perform numerical operations (MAX, MIN, AVG, SUM, SQRT, ROUND)		
2		ii) Perform data import/export operations for different file formats.		
3		Perform statistical operations - Mean, Median, Mode and Standard deviation, Variance, Skewness, Kurtosis		
4		Perform Z-test, T-test & ANOVA		
5.1		Perform data pre-processing operations i) Handling Missing data		
5.2		Perform data pre-processing operations ii) Normalization		
6		Apply and explore various plotting functions on the data set.		
7		Explore the features of Power BI Desktop		
8		Prepare & Load data		
9		Develop the data model		
10		Perform DAX calculations		

Explore the features of Ms-Excel

Ex.No:1

Date:

Aim:

Explore and demonstrate the practical application of key features in Microsoft Excel for business analytics.

Features of Microsoft Excel:

i. Spreadsheets and data organization

Excel is at its core a spreadsheet application, which means it is designed to organize and manage data in a grid-based format. This makes it easy to enter, edit, and format data, as well as to perform calculations and analysis.

Excel also allows you to create multiple worksheets within a single workbook, which can be used to organize your data into different categories or sections. This can be helpful for keeping track of complex data sets or for creating reports that combine data from multiple sources.

ii. Formulas and functions

One of Excel's most powerful features is its ability to perform calculations and operations on data using formulas and functions. Formulas allow you to combine mathematical operators, cell references, and built-in functions to create complex calculations.

Excel includes over 400 built-in functions that can be used to perform a wide variety of tasks, such as:

- Mathematical operations (e.g., SUM, AVERAGE, MEDIAN)
- Logical operations (e.g., IF, AND, OR)
- Text manipulation (e.g., LEFT, RIGHT, TRIM)
- Date and time functions (e.g., TODAY, NOW, DATEDIF)
- Statistical functions (e.g., COUNT, COUNTIF, STDEV)

Formulas and functions can be used to automate a wide variety of tasks, such as:

- Calculating totals and averages
- Applying discounts and taxes
- Generating reports
- Performing data analysis

iii. Data visualization

Excel includes a wide range of data visualization tools, such as charts and graphs, which can be used to visually represent data and trends. This can make it easier to understand and interpret data, as well as to communicate insights to others.

Excel offers a variety of chart types to choose from, including bar charts, line charts, pie charts, and scatter charts. You can also customize the appearance of your charts to suit your needs.

iv. Pivot Tables and Pivot Charts

Pivot Tables and Pivot Charts are two powerful tools that can be used to analyze and summarize data. PivotTables allow you to group, sort, and filter data, as well as to calculate totals and subtotals. Pivot Charts are a visual representation of PivotTables, which can make it easier to identify trends and relationships in data.

Pivot Tables and Pivot Charts are often used by business analysts and data scientists to analyze large and complex data sets. However, they can also be used by anyone who needs to analyze and summarize data, such as accountants, marketers, and sales professionals.

v. Data validation and drop-down lists

Data validation allows you to restrict the types of data that can be entered into cells. This can help to prevent errors and ensure that data is entered in a consistent format.

Drop-down lists are a type of data validation that allows users to select a value from a pre-defined list. This can be helpful for entering data such as product names, customer names, or country codes.

Data validation and drop-down lists can save time and improve the accuracy of data entry.

vi. Conditional formatting

Conditional formatting allows you to automatically apply formatting to cells based on their values. This can be used to highlight important data, identify trends, and make data easier to read and understand.

For example, you could use conditional formatting to highlight cells that contain negative values, or to color-code cells based on their value range.

vii. Macros

Macros are a series of recorded actions that can be played back to automate tasks. This can save time and improve the efficiency of your work.

Macros can be used to automate a wide variety of tasks, such as:

- Formatting data
- Creating reports
- Generating charts
- Importing and exporting data
- Macros can be created and edited using the Visual Basic Editor (VBE).

viii. Collaboration features

Excel includes a number of collaboration features that make it easy to share and work on spreadsheets with others.

For example, you can co-author spreadsheets in real time, and you can also leave comments and notes for others to see.

Excel also allows you to save spreadsheets to the cloud, such as OneDrive or SharePoint, which makes it easy to access them from anywhere.

ix. Add-ins

Add-ins are third-party applications that can be installed to extend the functionality of Excel. There are a wide variety of add-ins available, including add-ins for data analysis, financial modeling, and project management.

Result:

The key features of Microsoft Excel for business analytics are explored successful.

Get the input from user and perform numerical operations (MAX, MIN, AVG, SUM, SQRT, ROUND)

Ex.No:2.1

Date:

Aim:

The aim of this lab is to enable participants to use Microsoft Excel for performing numerical operations (MAX, MIN, AVG, SUM, SQRT, ROUND) with input provided by the user.

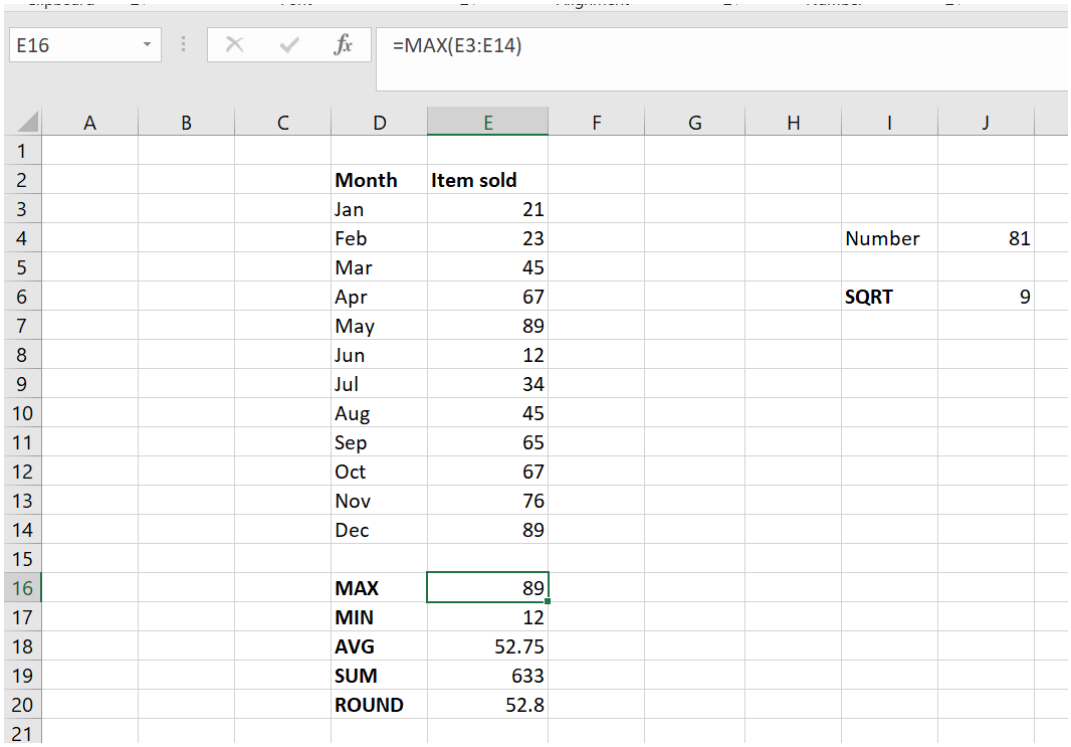
Procedure:

1. Open new Microsoft Excel spreadsheet
2. Type the required numerical data to perform numerical operations
3. Perform the following numerical operation (MAX, MIN, AVG, SUM, SQRT, ROUND) using microsoft Excel functions
4. Verify the output manually.
5. Stop the experiment

MAX function:

The MAX function is a premade function in Excel, which finds the highest number in a range.

- Select a cell (E16)
- Type =MAX
- Double click the MAX command
- Select a range (E3:E14)
- Hit enter

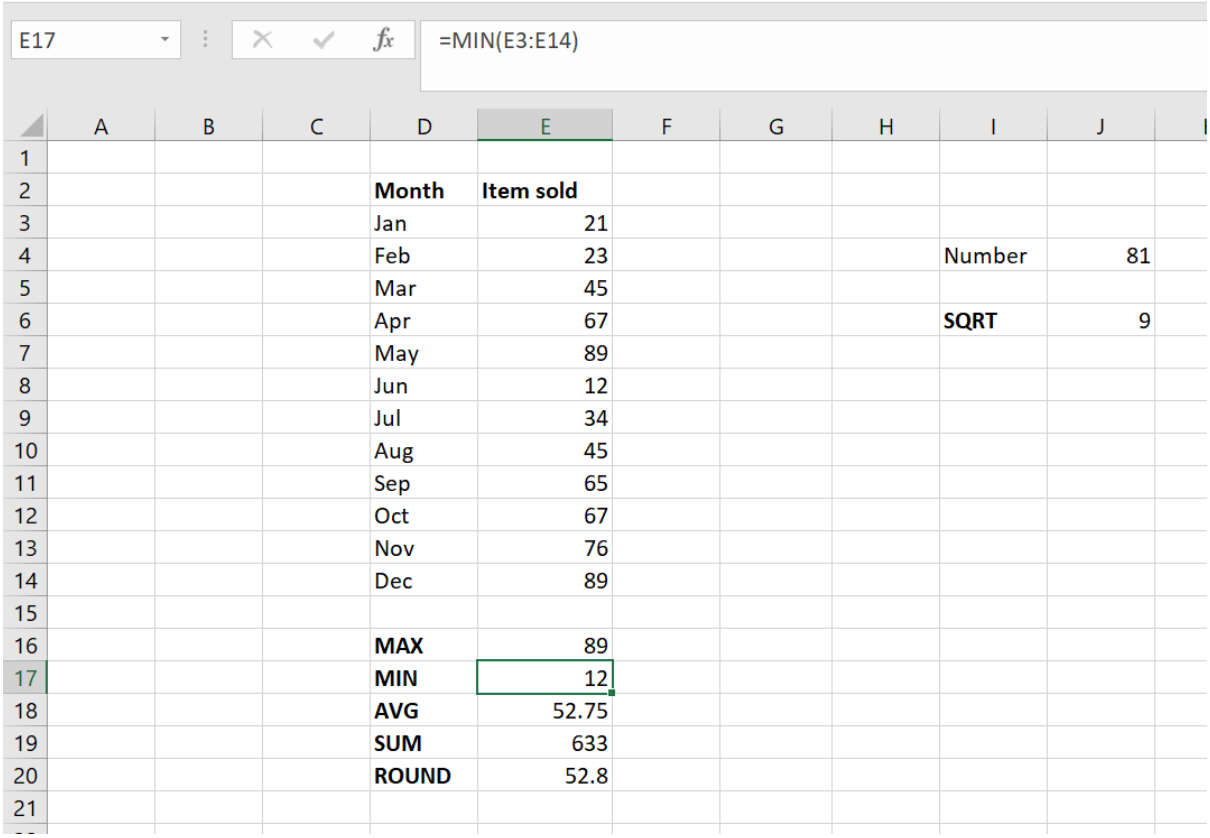


	A	B	C	D	E	F	G	H	I	J
1										
2				**Month**	**Item sold**					
3				Jan	21					
4				Feb	23				Number	81
5				Mar	45					
6				Apr	67				**SQRT**	9
7				May	89					
8				Jun	12					
9				Jul	34					
10				Aug	45					
11				Sep	65					
12				Oct	67					
13				Nov	76					
14				Dec	89					
15										
16				**MAX**	89					
17				**MIN**	12					
18				**AVG**	52.75					
19				**SUM**	633					
20				**ROUND**	52.8					
21										

MIN function:

The MIN function is a premade function in Excel, which finds the lowest number in a range.

- Select a cell (E17)
- Type =MIN
- Double click the MIN command
- Select a range (E3:E14)
- Hit enter

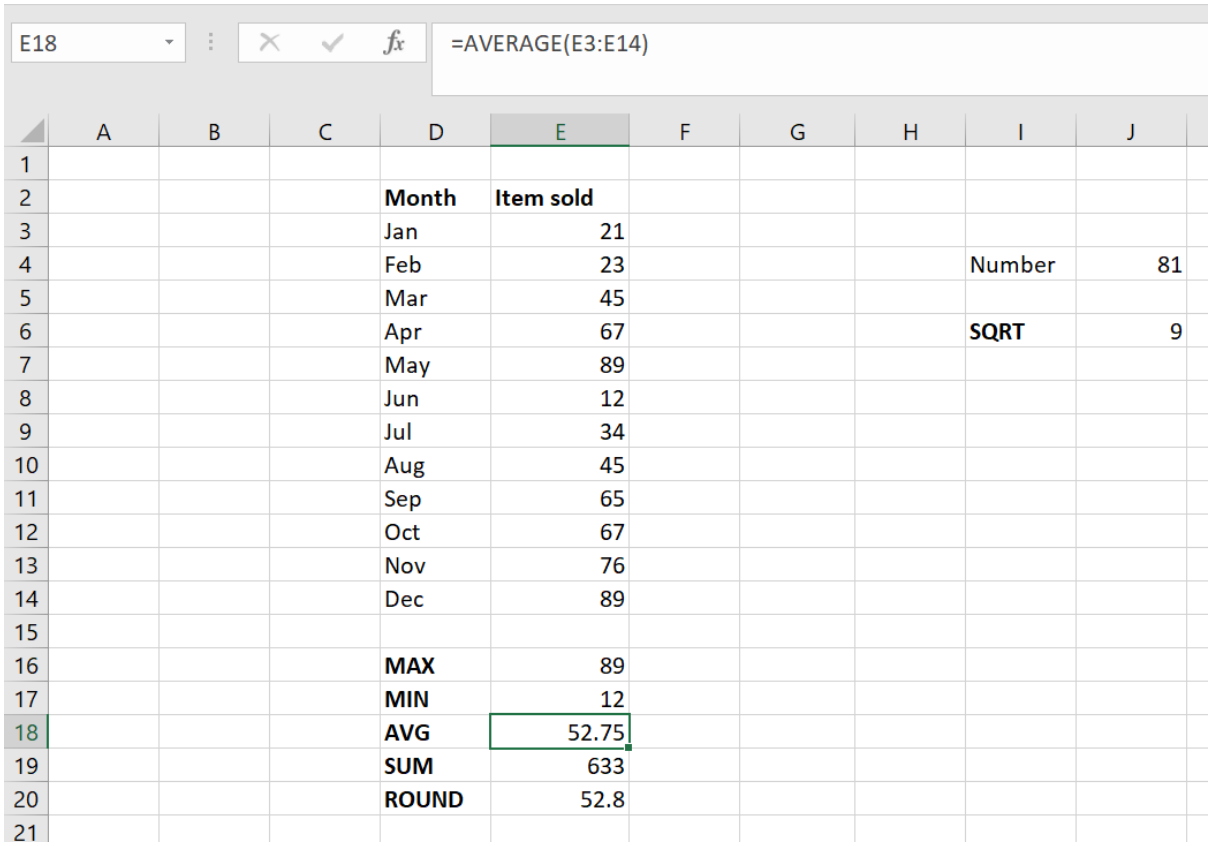


	A	B	C	D	E	F	G	H	I	J
1										
2				**Month**	**Item sold**					
3				Jan	21					
4				Feb	23				Number	81
5				Mar	45					
6				Apr	67				**SQRT**	9
7				May	89					
8				Jun	12					
9				Jul	34					
10				Aug	45					
11				Sep	65					
12				Oct	67					
13				Nov	76					
14				Dec	89					
15										
16				**MAX**	89					
17				**MIN**	12					
18				**AVG**	52.75					
19				**SUM**	633					
20				**ROUND**	52.8					
21										

AVG function:

The AVERAGE function is a premade function in Excel, which calculates the average (arithmetic mean).

- Select a cell (E18)
- Type =AVERAGE
- Double click the AVERAGE command
- Select a range (E3:E14)
- Hit enter

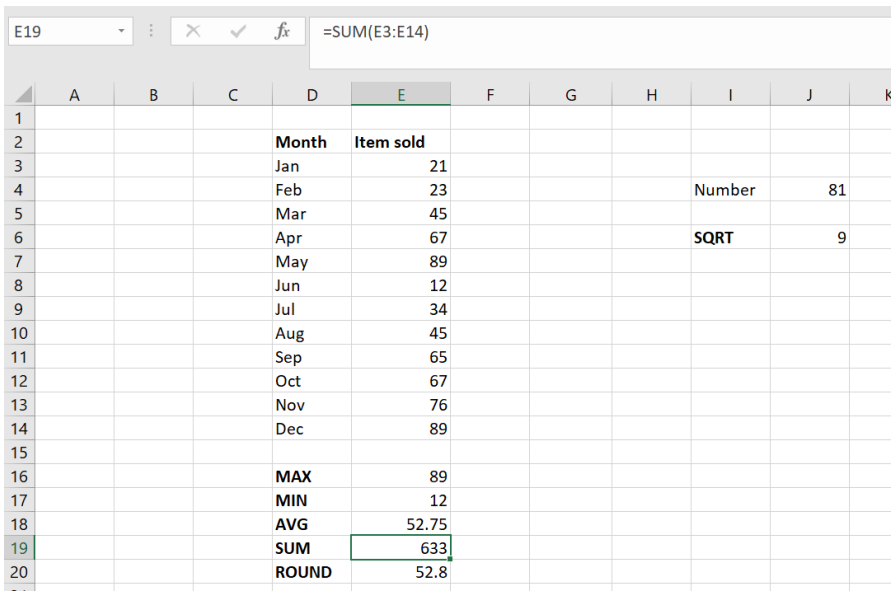


	A	B	C	D	E	F	G	H	I	J
1										
2				**Month**	**Item sold**					
3				Jan	21					
4				Feb	23				Number	81
5				Mar	45					
6				Apr	67				**SQRT**	9
7				May	89					
8				Jun	12					
9				Jul	34					
10				Aug	45					
11				Sep	65					
12				Oct	67					
13				Nov	76					
14				Dec	89					
15										
16				**MAX**	89					
17				**MIN**	12					
18				**AVG**	52.75					
19				**SUM**	633					
20				**ROUND**	52.8					
21										

SUM function:

The SUM function is a premade function in Excel, which adds numbers in a range.

- Select a cell E19
- Type =SUM
- Double click the SUM command
- Select a range E3:E14
- Hit enter

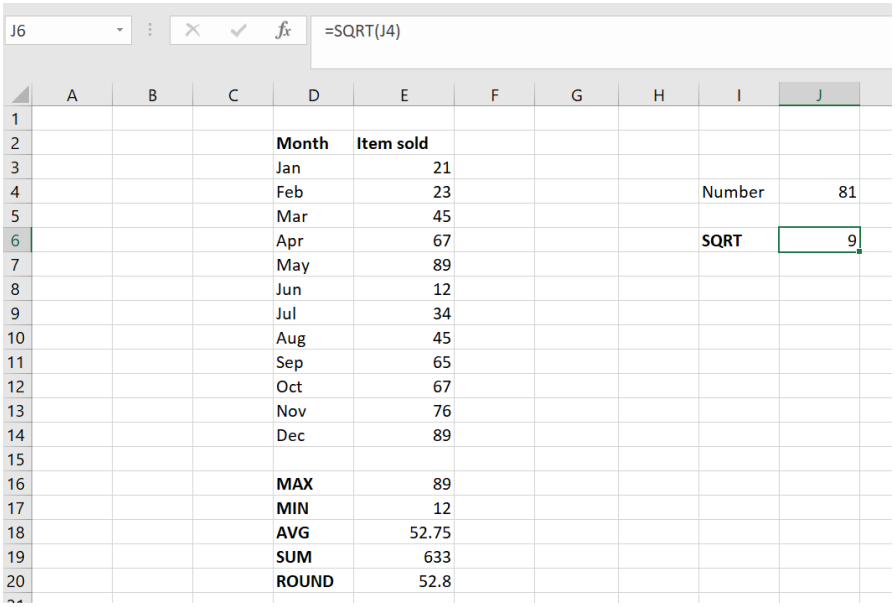


	A	B	C	D	E	F	G	H	I	J	K
1											
2				**Month**	**Item sold**						
3				Jan	21						
4				Feb	23				Number	81	
5				Mar	45						
6				Apr	67				**SQRT**	9	
7				May	89						
8				Jun	12						
9				Jul	34						
10				Aug	45						
11				Sep	65						
12				Oct	67						
13				Nov	76						
14				Dec	89						
15											
16				**MAX**	89						
17				**MIN**	12						
18				**AVG**	52.75						
19				**SUM**	633						
20				**ROUND**	52.8						

SQRT function:

The SQRT function is premade function in excel, used to find square root of the value.

- Select a cell J6
- Type =SQRT
- Double click the SQRT command
- Select a value J4
- Hit enter



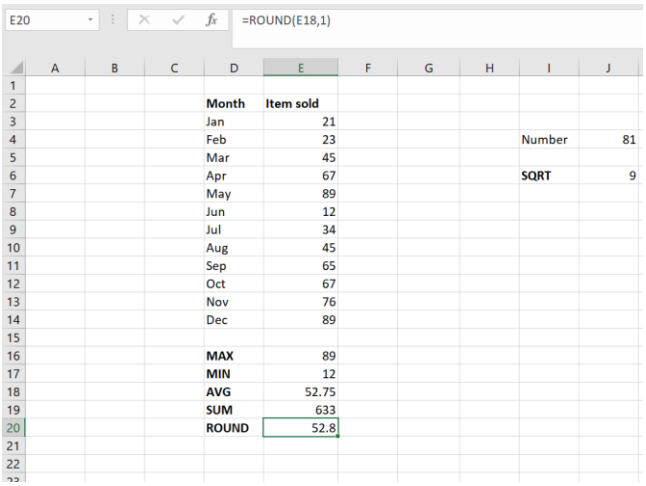
J6 =SQRT(J4)

	A	B	C	D	E	F	G	H	I	J
1										
2				**Month**	**Item sold**					
3				Jan	21					
4				Feb	23				Number	81
5				Mar	45					
6				Apr	67				**SQRT**	9
7				May	89					
8				Jun	12					
9				Jul	34					
10				Aug	45					
11				Sep	65					
12				Oct	67					
13				Nov	76					
14				Dec	89					
15										
16				**MAX**	89					
17				**MIN**	12					
18				**AVG**	52.75					
19				**SUM**	633					
20				**ROUND**	52.8					

ROUND Function:

The ROUND function in Excel rounds a number to a specified number of digits.

- Select a cell J6
- Type =SQRT
- Double click the SQRT command
- Select a value J4
- Hit enter



E20 =ROUND(E18,1)

	A	B	C	D	E	F	G	H	I	J
1										
2				**Month**	**Item sold**					
3				Jan	21					
4				Feb	23				Number	81
5				Mar	45					
6				Apr	67				**SQRT**	9
7				May	89					
8				Jun	12					
9				Jul	34					
10				Aug	45					
11				Sep	65					
12				Oct	67					
13				Nov	76					
14				Dec	89					
15										
16				**MAX**	89					
17				**MIN**	12					
18				**AVG**	52.75					
19				**SUM**	633					
20				**ROUND**	52.8					
21										
22										

Result:

This experiment demonstrates the effective use of programming to collect user input and perform various numerical operations (MAX, MIN, AVG, SUM, SQRT, ROUND).

Perform data import/export operations for different file formats.

Ex.No:2.2

Date:

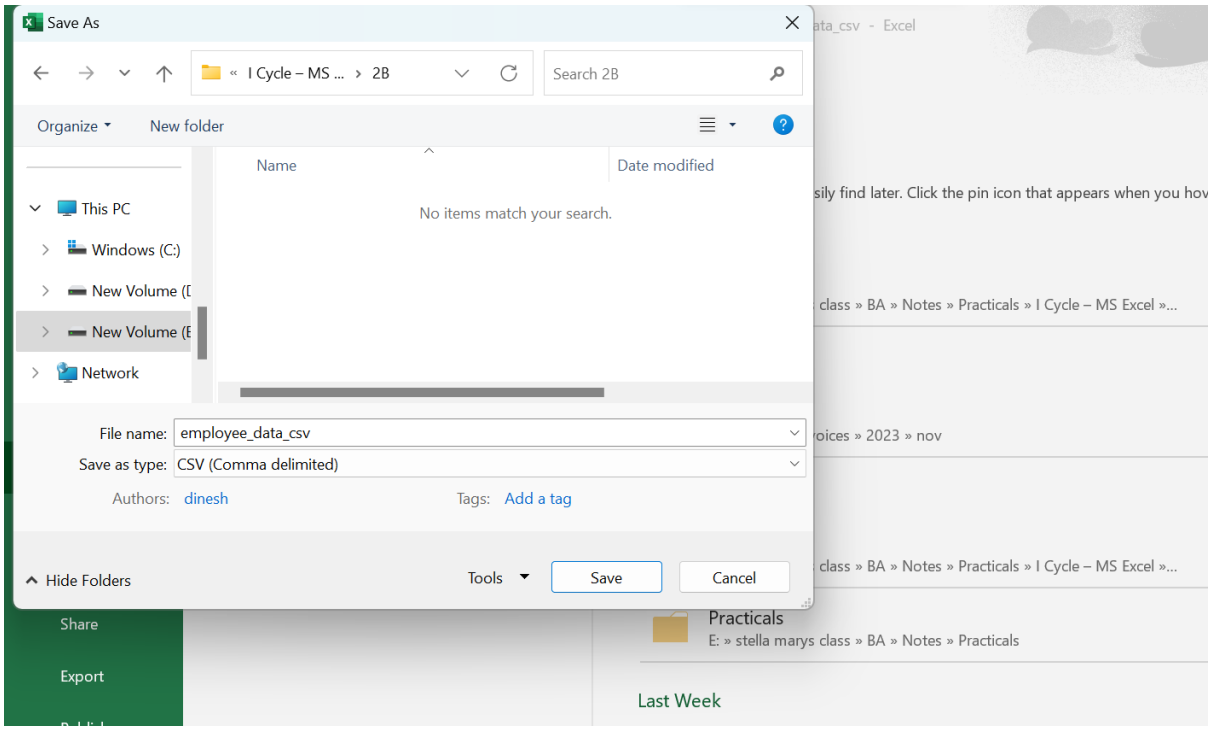
Aim:

To perform data import/export operations for different file formats.

Procedure:

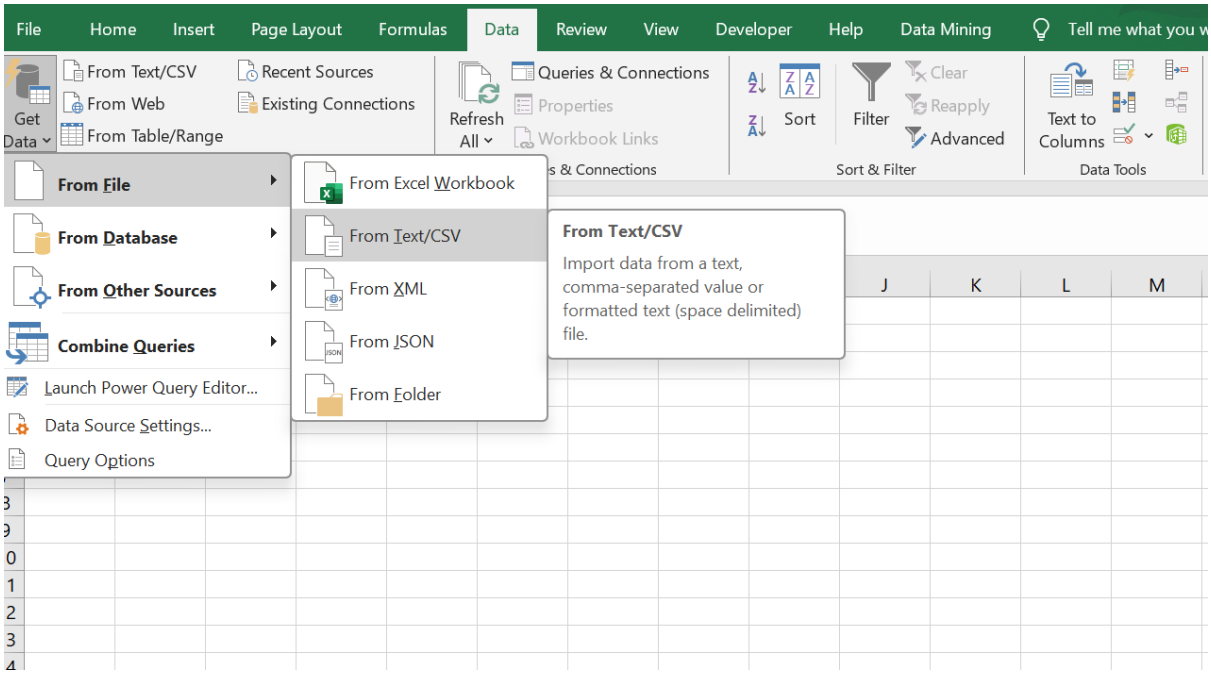
1. Importing Data from CSV File and Export it into Microsoft Excel Worksheet (.xlsx):

a. Open Microsoft Excel, Type sample dataset and save as “employee_data_csv.csv”

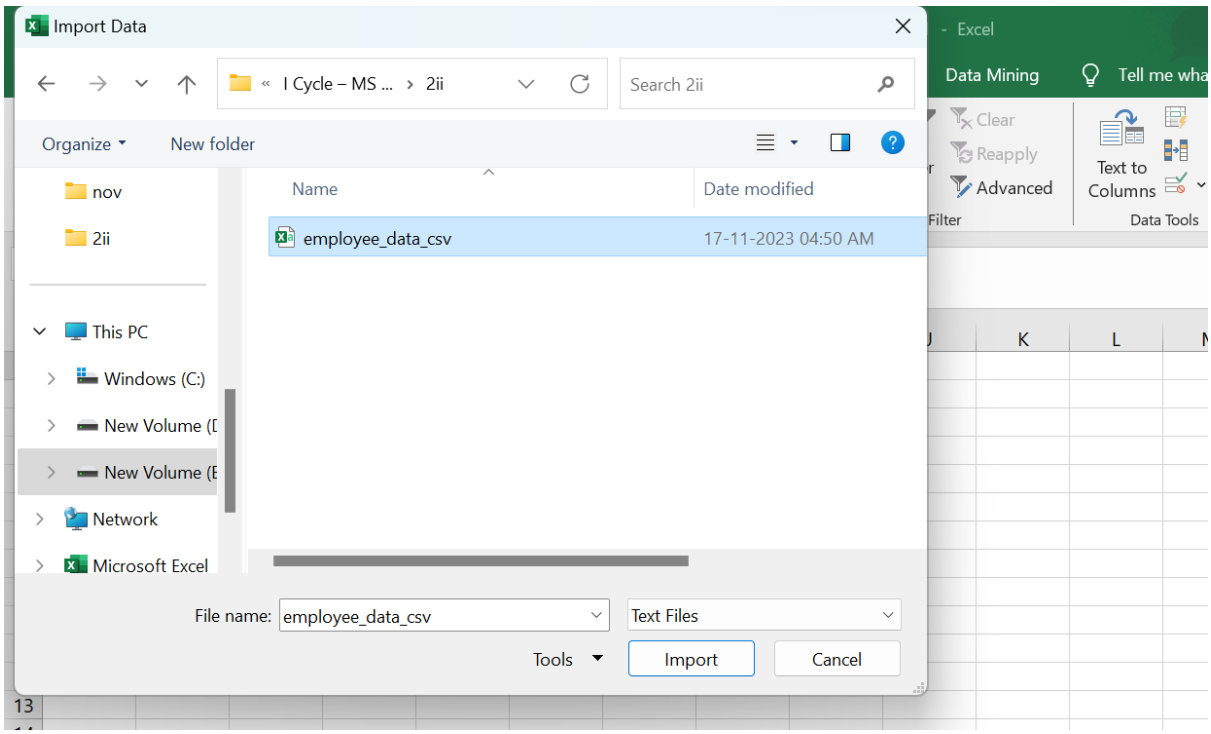


b. Go to the "Data" tab.

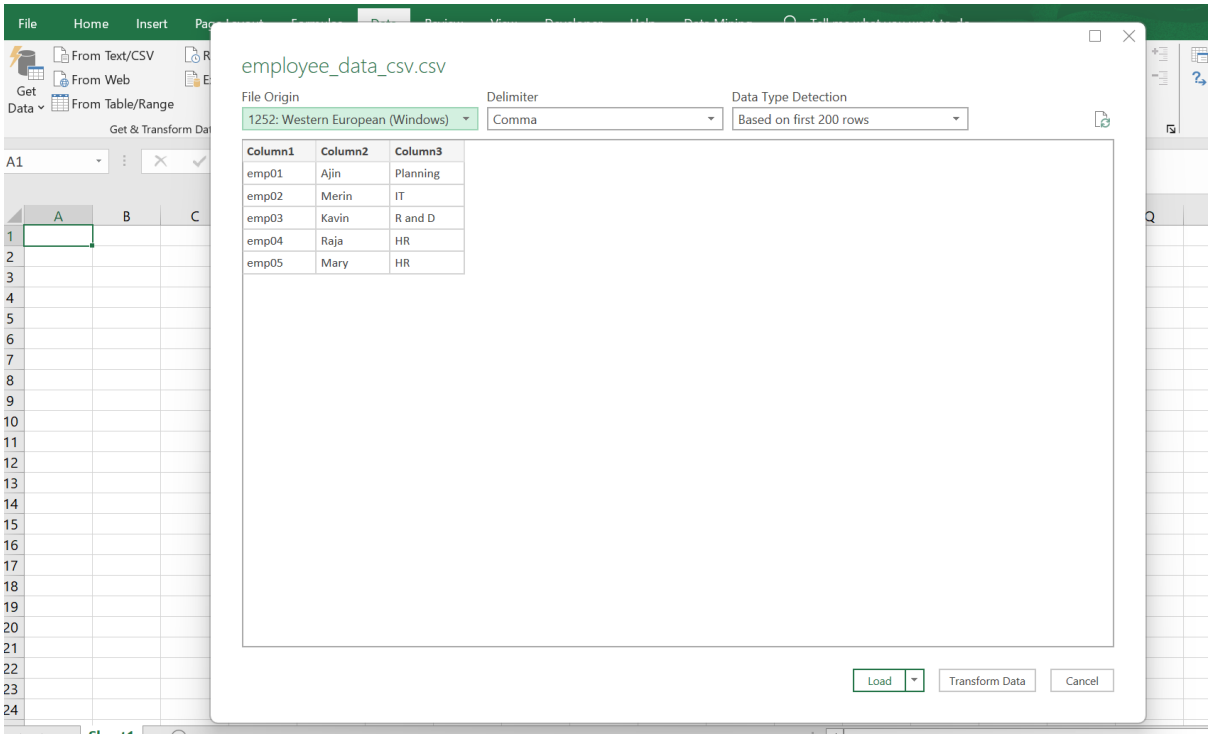
c. Click on "Get Data" and select "From Text/CSV."



d. Choose a sample CSV file and click "Import."

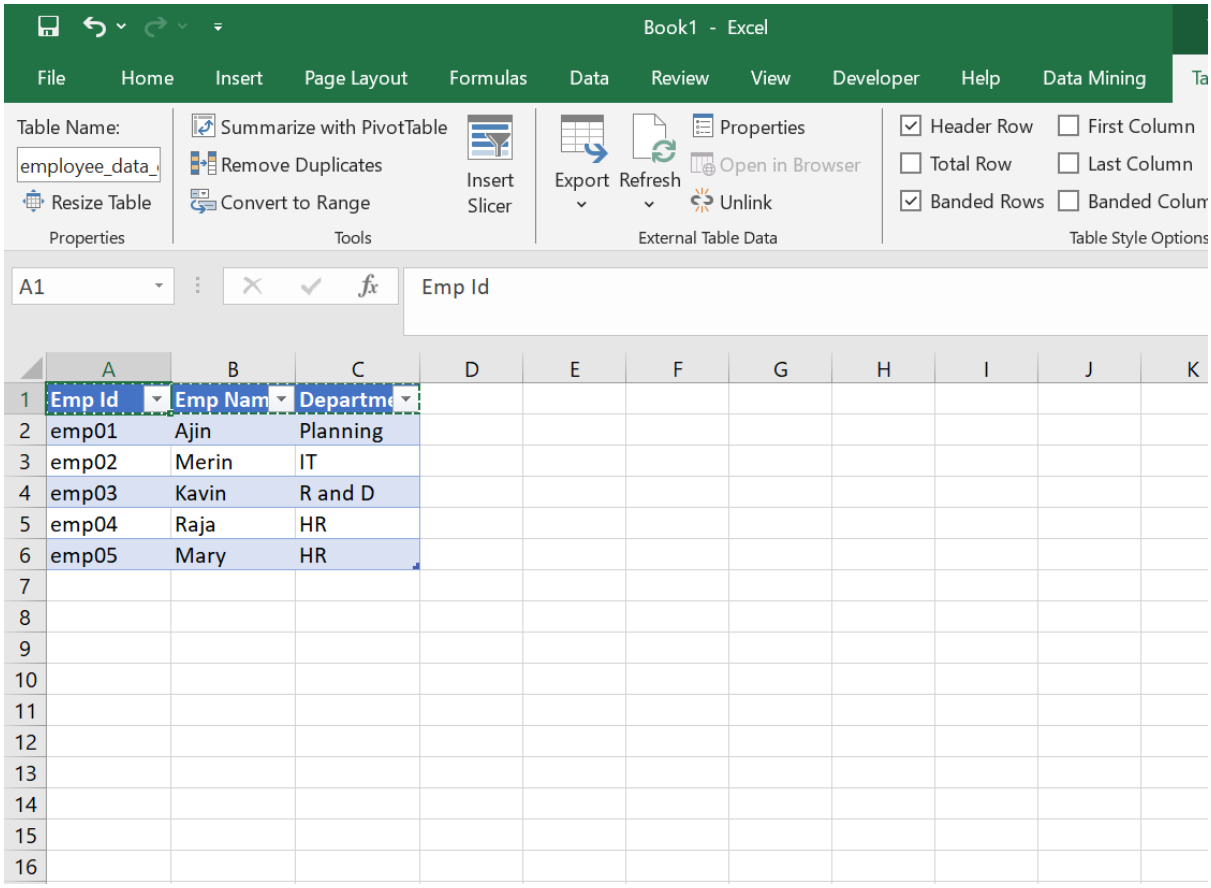


e. Ensure proper delimiter selection and preview the data.

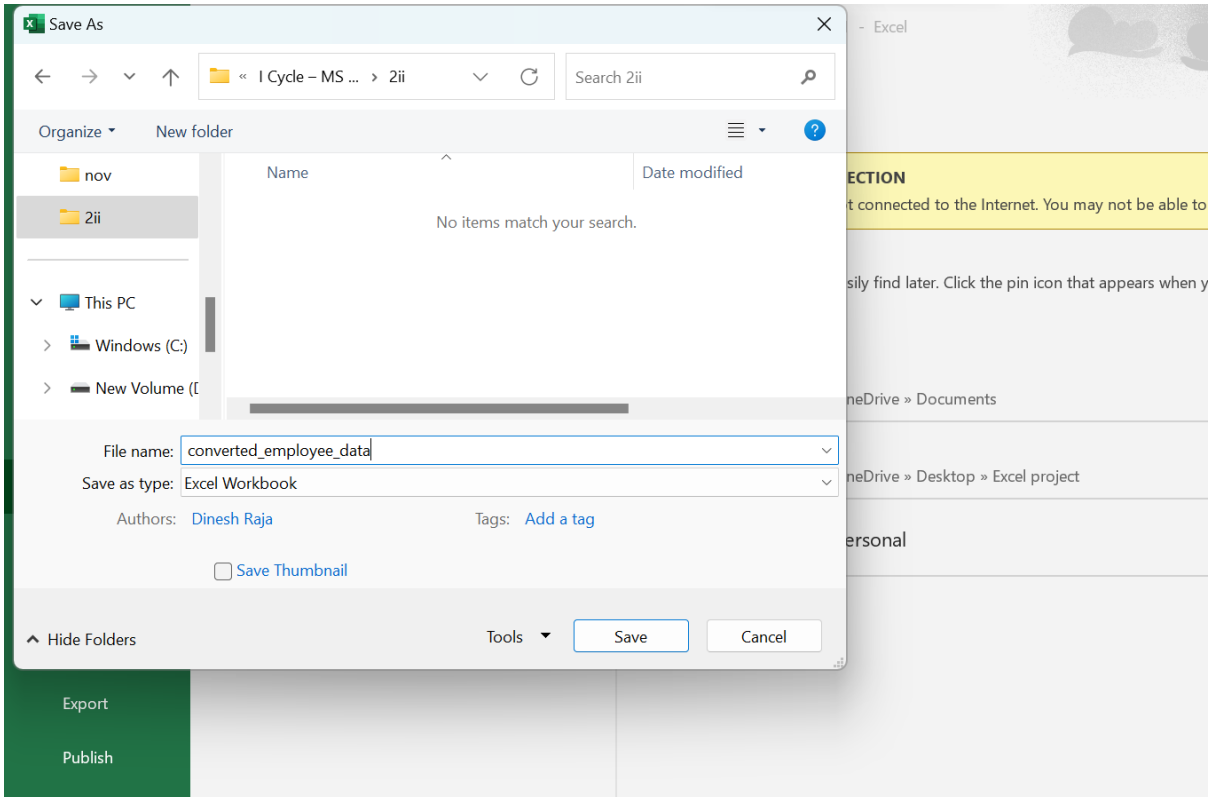


f. Click "Load" to import the data into Excel.

g. Edit the default column names (column1, column2, column3) into Emp Id, Emp Name, Department

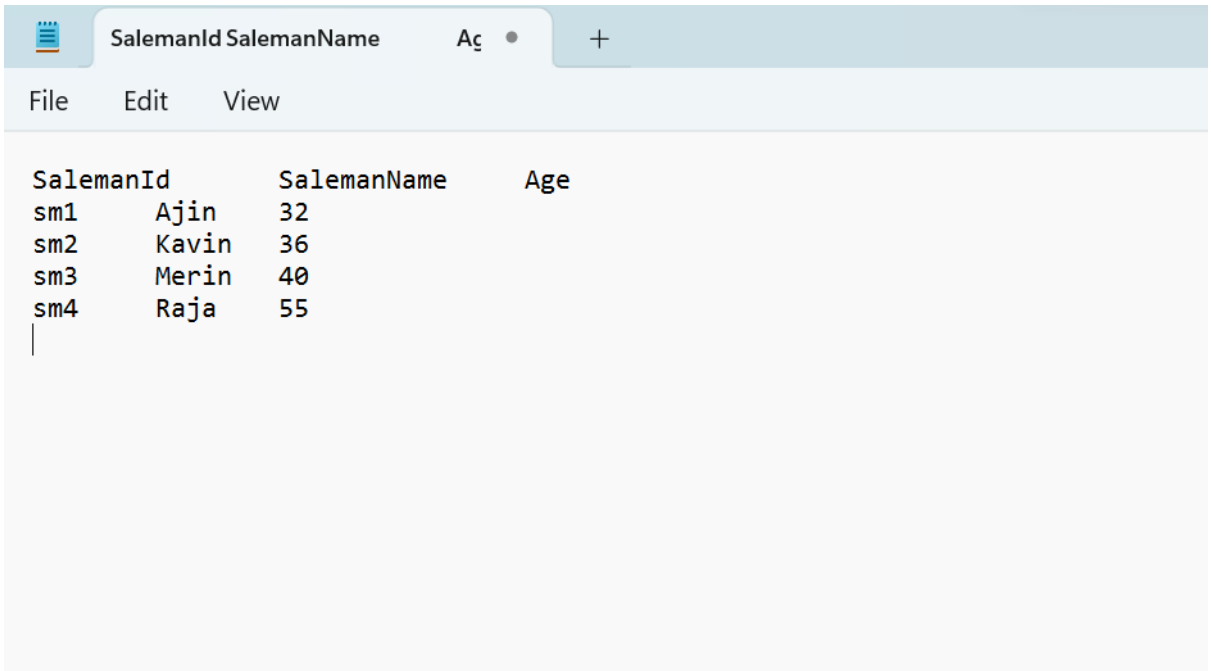


h. Save the file as “converted_employee_data” Microsoft Excel Worksheet (.xlsx)

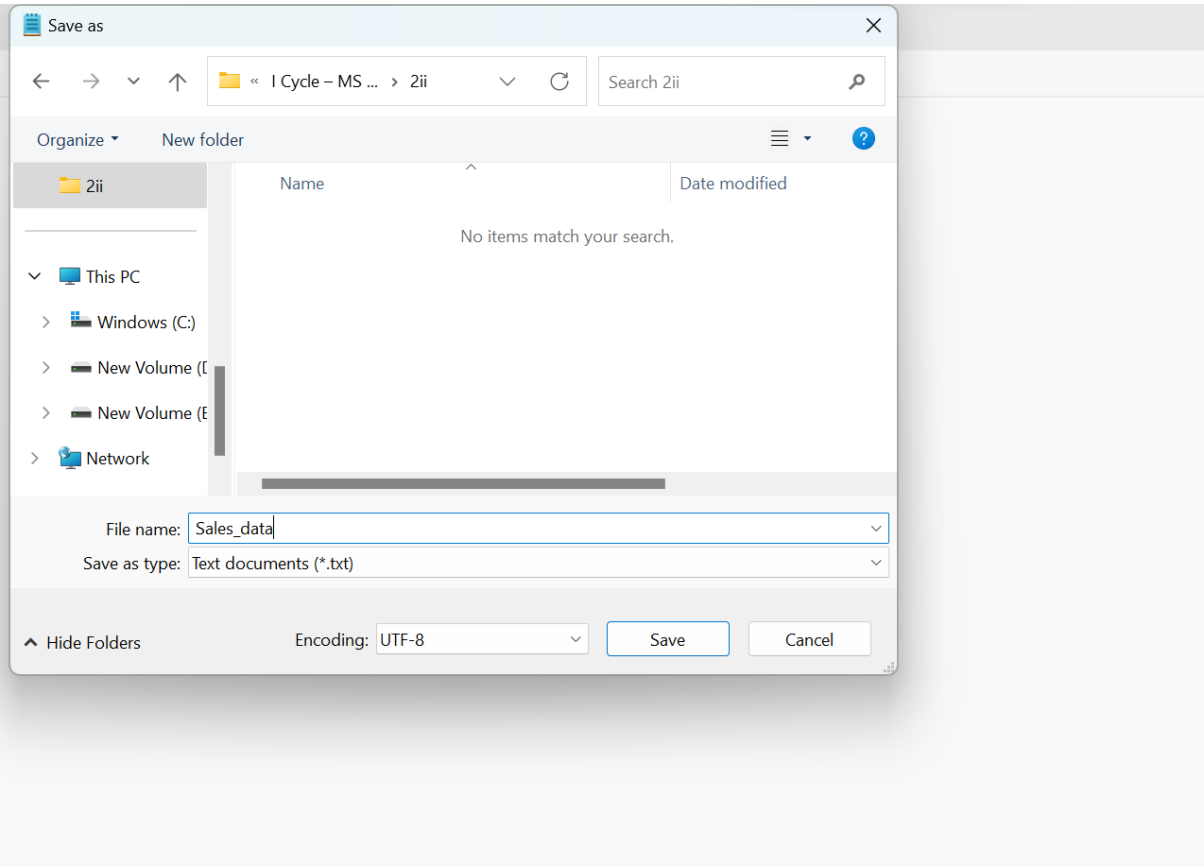


2. Importing Data from Text File and Export into CSV file:

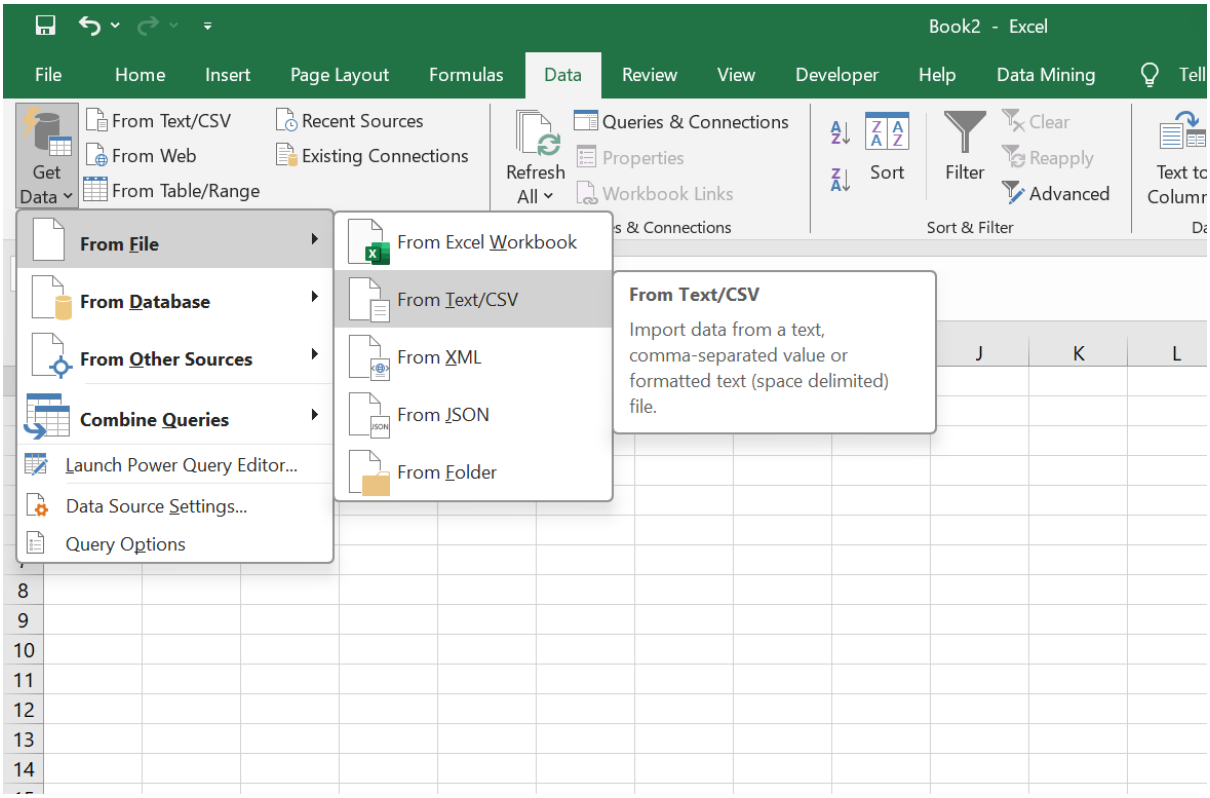
a. Opent Notepad and type sample dataset



b. Then save as “Sales_data.txt”

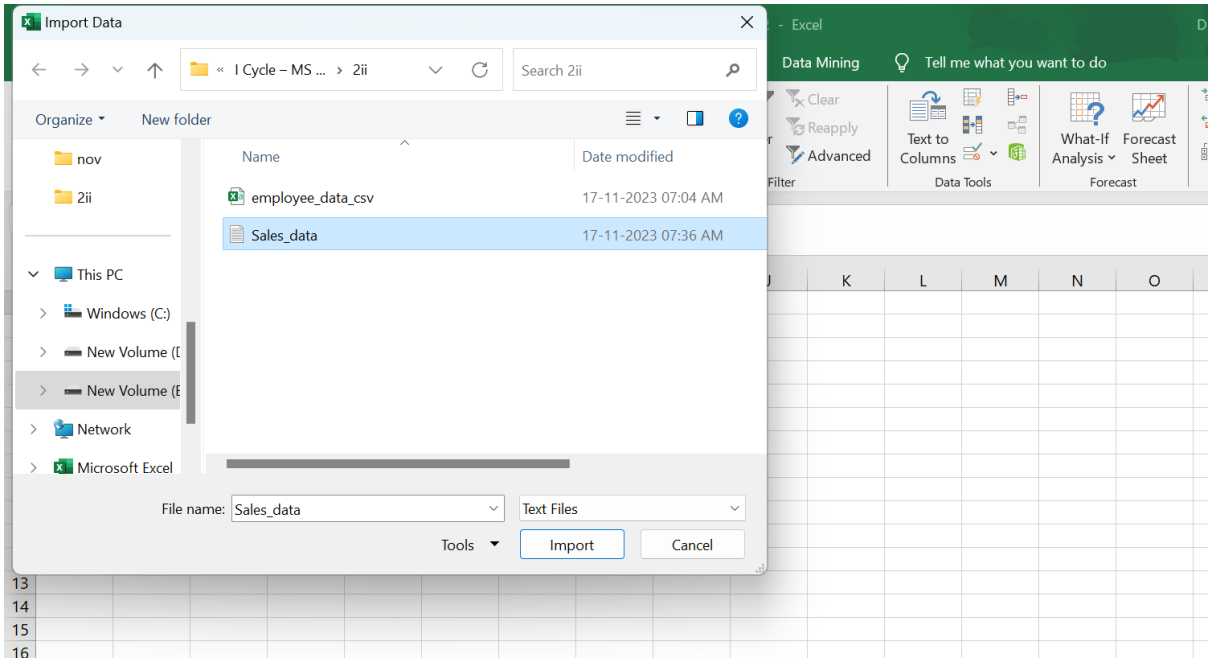


b. Open New MS Excel and Go to the "Data" tab.

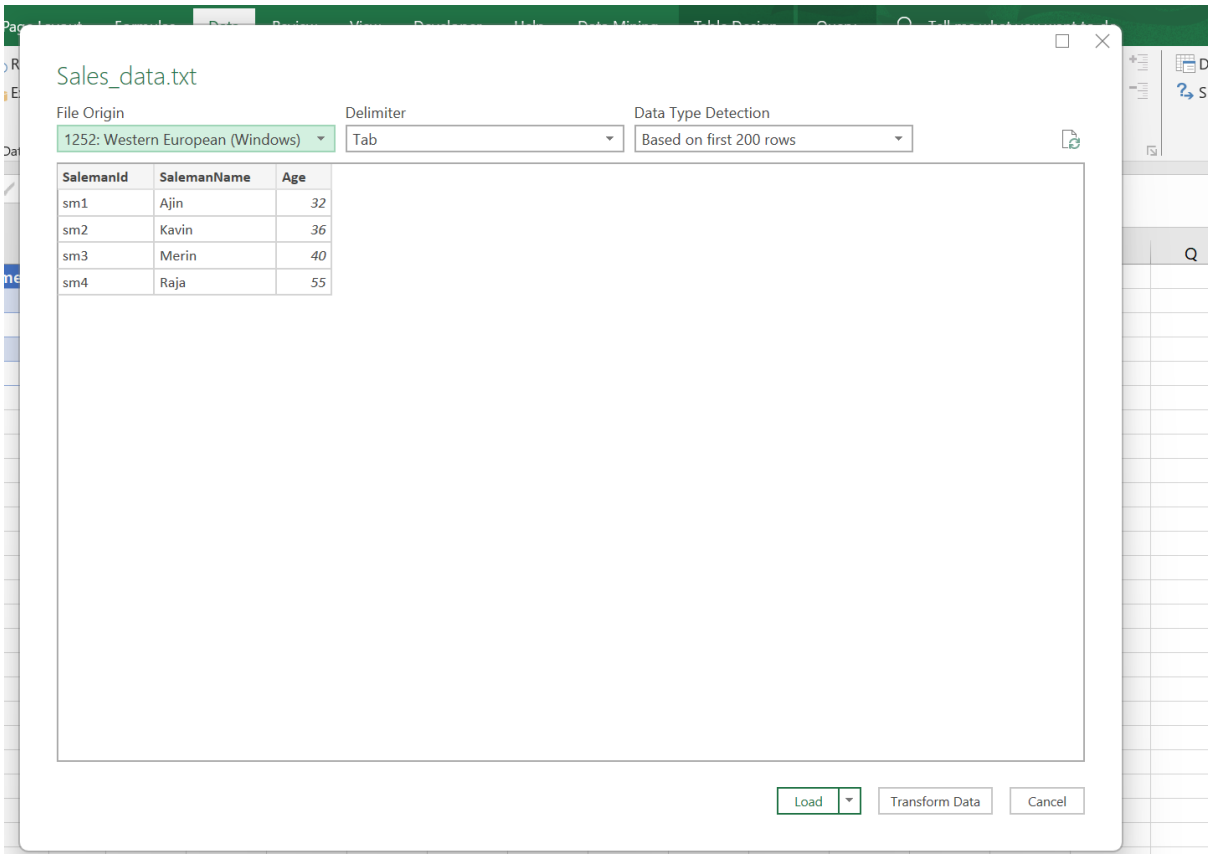


c. Click on "Get Data" and select "From Text."

d. Choose a sample text file (Sales_data.txt) and click "Import."



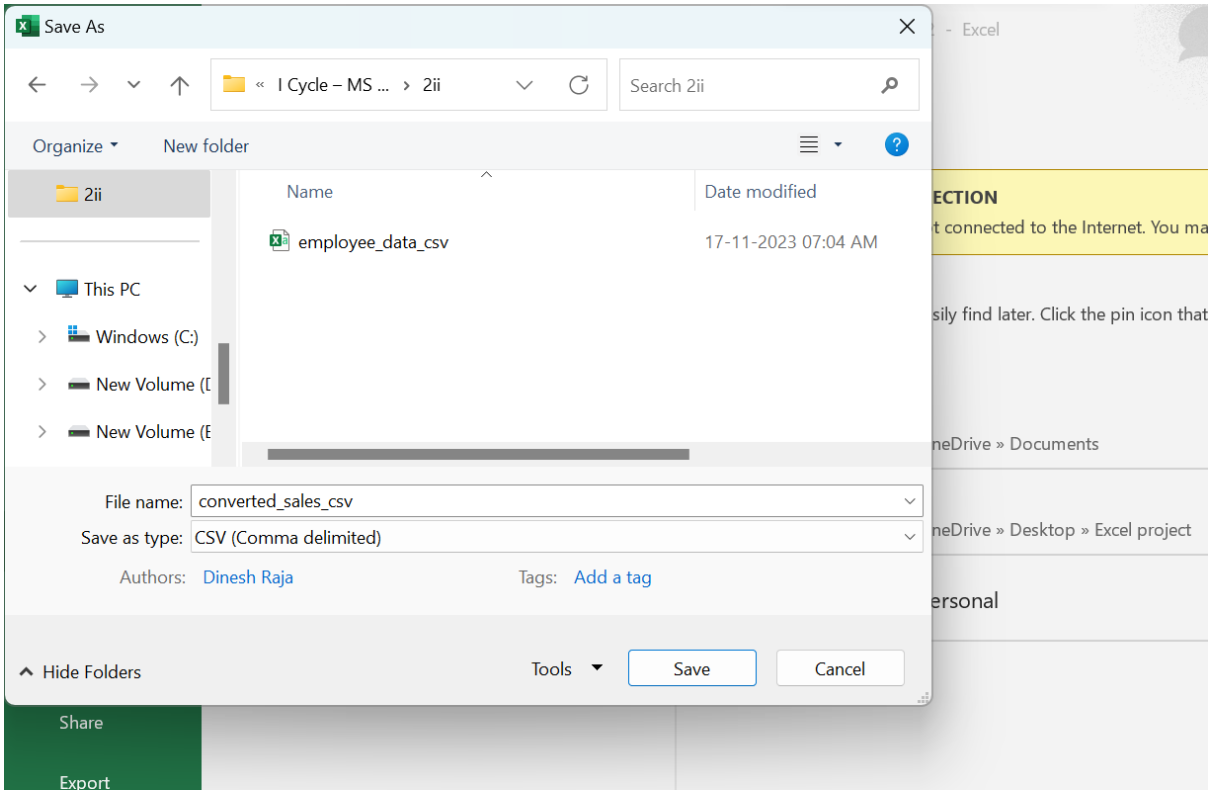
e. Choose required delimiter (Tab) if needed and preview the data.



f. Click "Load" to import the text file data into Excel.

g. Go to the "File" tab, choose "Save As," and select "CSV (Comma delimited) (*.csv)."

h. Save the file with an appropriate name ( ).



Result:

The CSV file is imported and then exported into a Microsoft Excel file (.xlsx). The Text File (.txt) is imported into Excel and then successfully exported into CSV (Comma-Separated Value) files.

Perform statistical operations - Mean, Median, Mode and Standard deviation, Variance, Skewness, Kurtosis

Ex.No:3

Date:

Aim:

This lab aims to introduce participants to statistical operations in Microsoft Excel, allowing them to gain practical experience in calculating and interpreting measures like Mean, Median, Mode, Standard Deviation, Variance, Skewness, and Kurtosis.

Procedure:

Statistical Operations:

Input dataset:

The below dataset is used for perform statistical operation in Microsoft excel. Open New Spreadsheet and type the data from A1 to K1.

Mark 50 45 78 89 90 20 98 100 60 70

Mean:

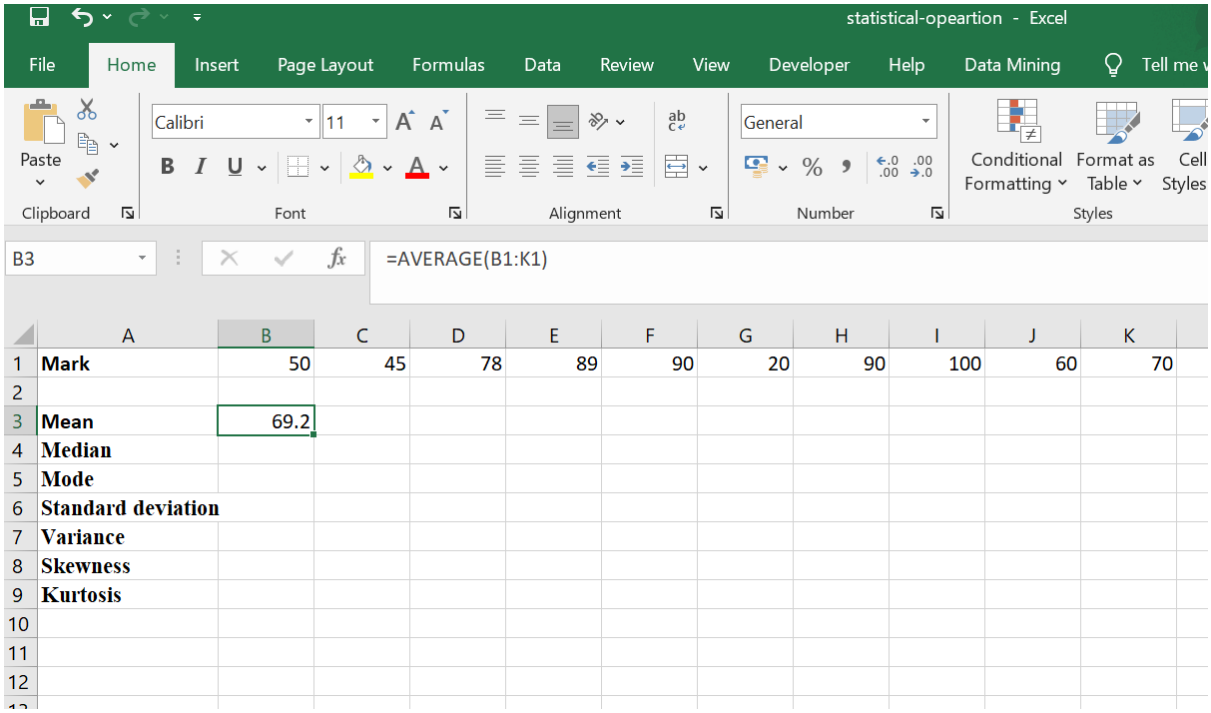
The AVERAGE function in Excel calculates the average (arithmetic mean) of a group of numbers.

a. Select a cell (B3)

b. Type =AVERAGE(

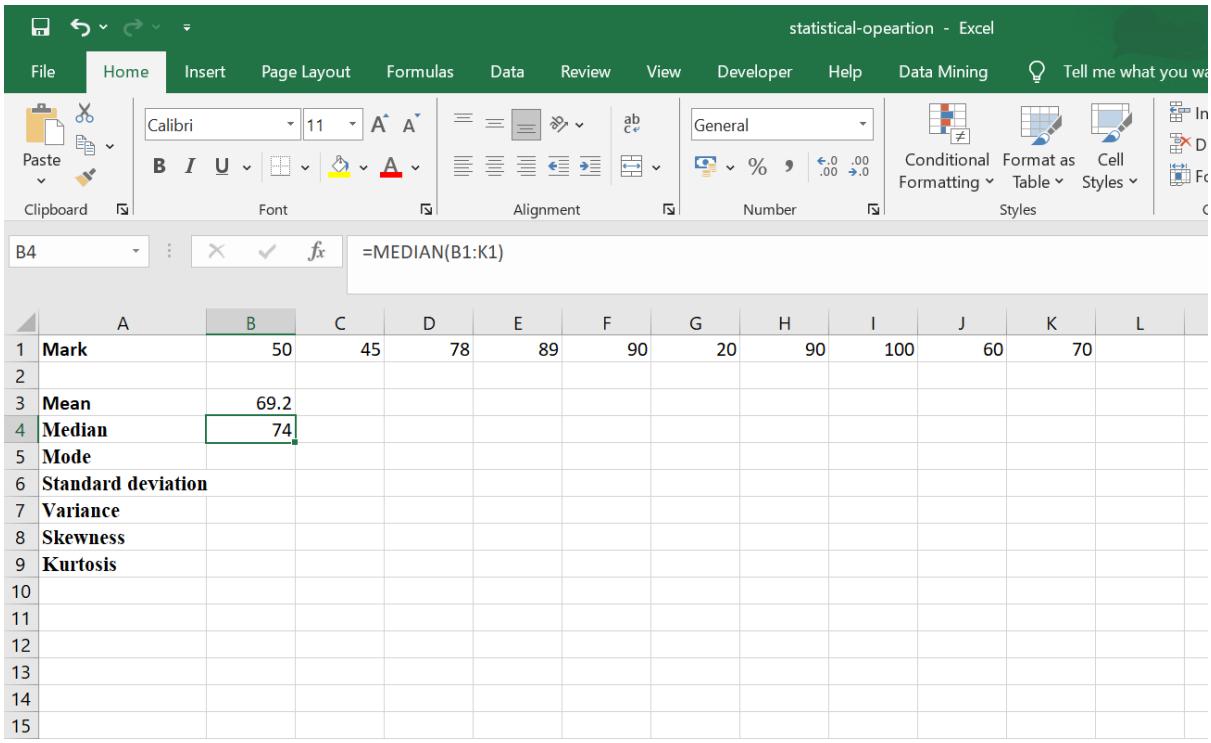
c. Select a range (B1:K1) and type close bracket)

d. And then Hit enter



Median:

To find the median (or middle number), use the MEDIAN function.



	A	B	C	D	E	F	G	H	I	J	K	L
1	**Mark**	50	45	78	89	90	20	90	100	60	70	
2												
3	**Mean**	69.2										
4	**Median**	74										
5	**Mode**											
6	**Standard deviation**											
7	**Variance**											
8	**Skewness**											
9	**Kurtosis**											
10												
11												
12												
13												
14												
15												

a. Select a cell (B4)

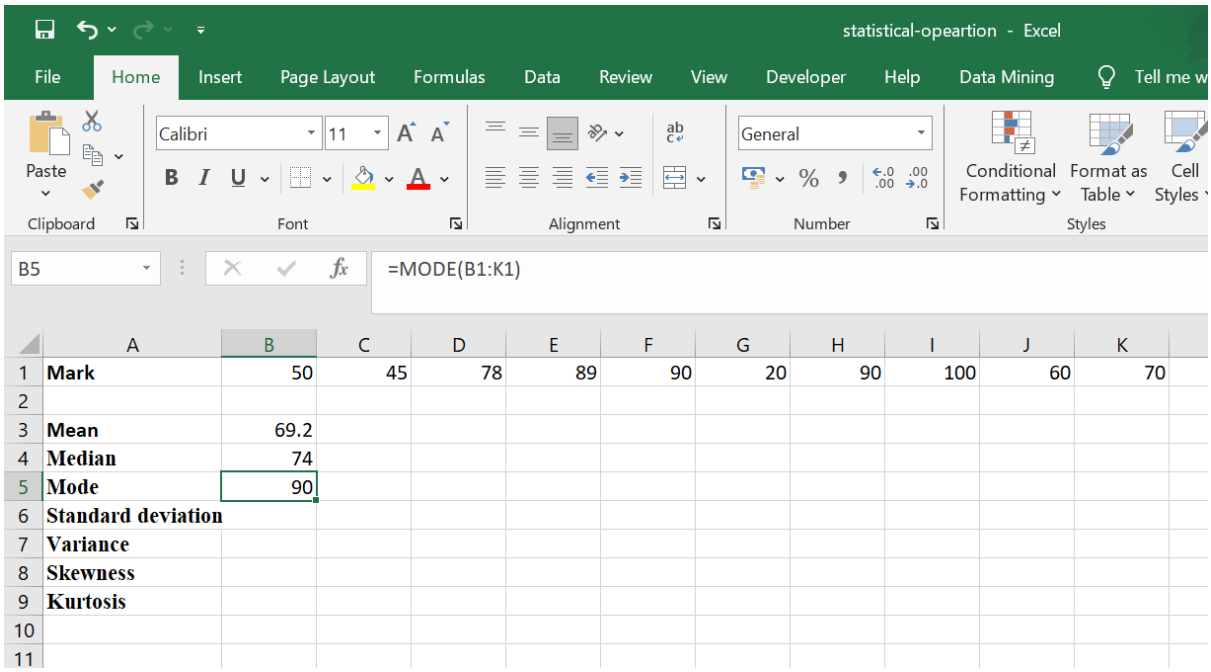
b. Type =MEDIAN(

c. Select a range (B1:K1) and type close bracket)

d. And then Hit enter

Mode:

To find the most frequently occurring number, use the MODE function.



	A	B	C	D	E	F	G	H	I	J	K
1	**Mark**	50	45	78	89	90	20	90	100	60	70
2											
3	**Mean**	69.2									
4	**Median**	74									
5	**Mode**	90									
6	**Standard deviation**										
7	**Variance**										
8	**Skewness**										
9	**Kurtosis**										
10											
11											

a. Select a cell (B5)

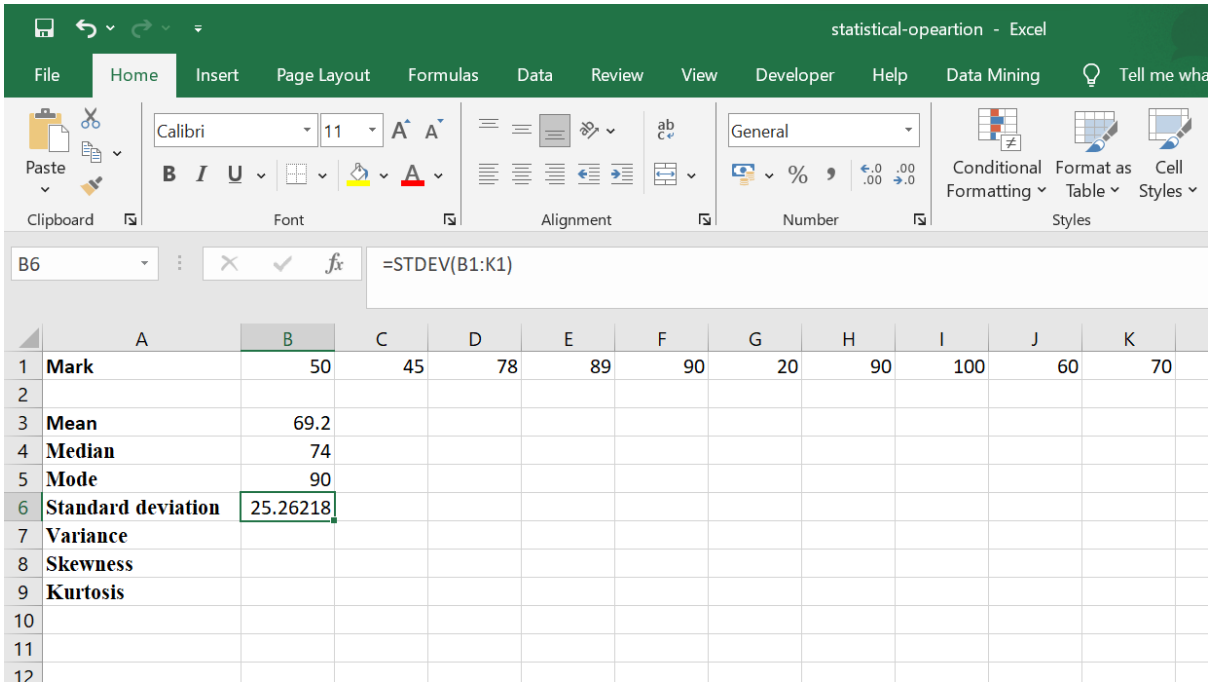
b. Type =MODE(

c. Select a range (B1:K1) and type close bracket)

d. And then Hit enter

Standard Deviation:

Standard deviation is a statistical measure that quantifies the amount of variation or dispersion in a set of data points.



	A	B	C	D	E	F	G	H	I	J	K
1	**Mark**	50	45	78	89	90	20	90	100	60	70
2											
3	**Mean**	69.2									
4	**Median**	74									
5	**Mode**	90									
6	**Standard deviation**	25.26218									
7	**Variance**										
8	**Skewness**										
9	**Kurtosis**										
10											
11											
12											

a. Select a cell (B6)

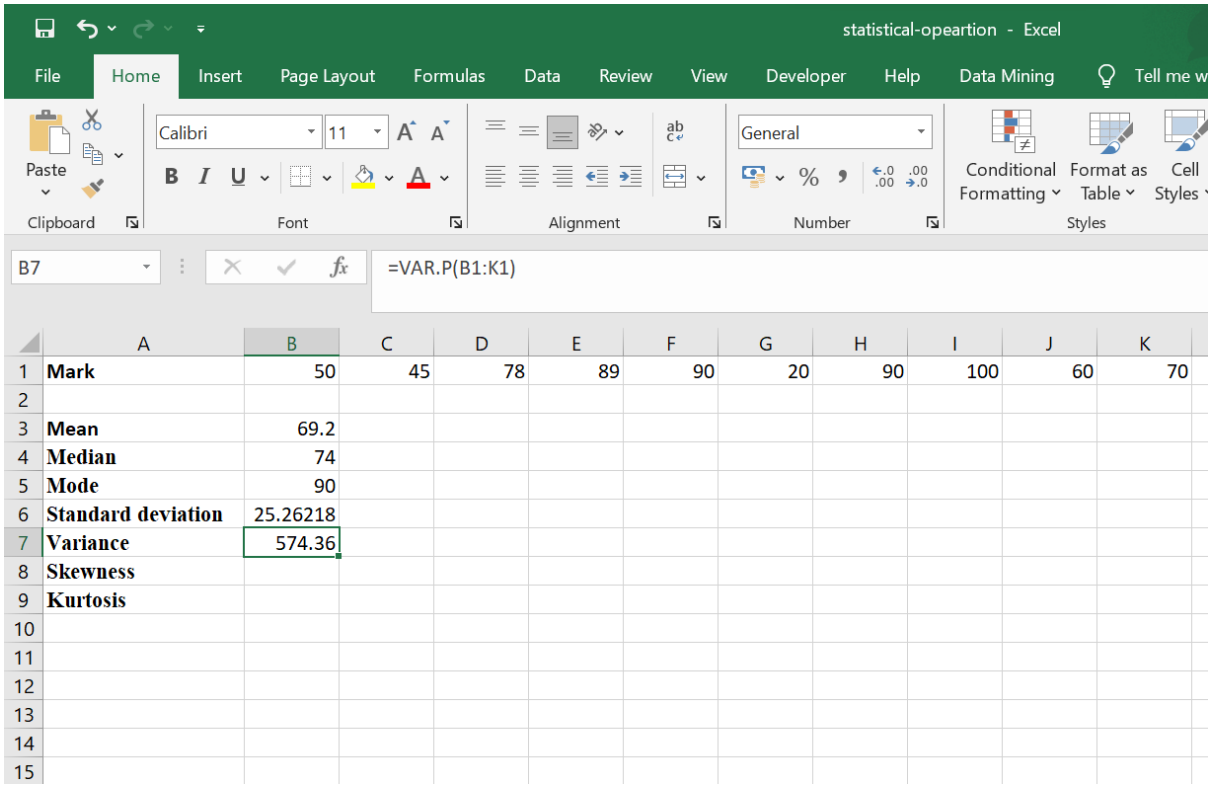
b. Type =STDEV(

c. Select a range (B1:K1) and type close bracket)

d. And then Hit enter

Variance:

Variance is a statistical measure that quantifies the degree of spread or dispersion in a set of data points.



	A	B	C	D	E	F	G	H	I	J	K
1	**Mark**	50	45	78	89	90	20	90	100	60	70
2											
3	**Mean**	69.2									
4	**Median**	74									
5	**Mode**	90									
6	**Standard deviation**	25.26218									
7	**Variance**	574.36									
8	**Skewness**										
9	**Kurtosis**										
10											
11											
12											
13											
14											
15											

a. Select a cell (B7)

b. Type =VAR.P(

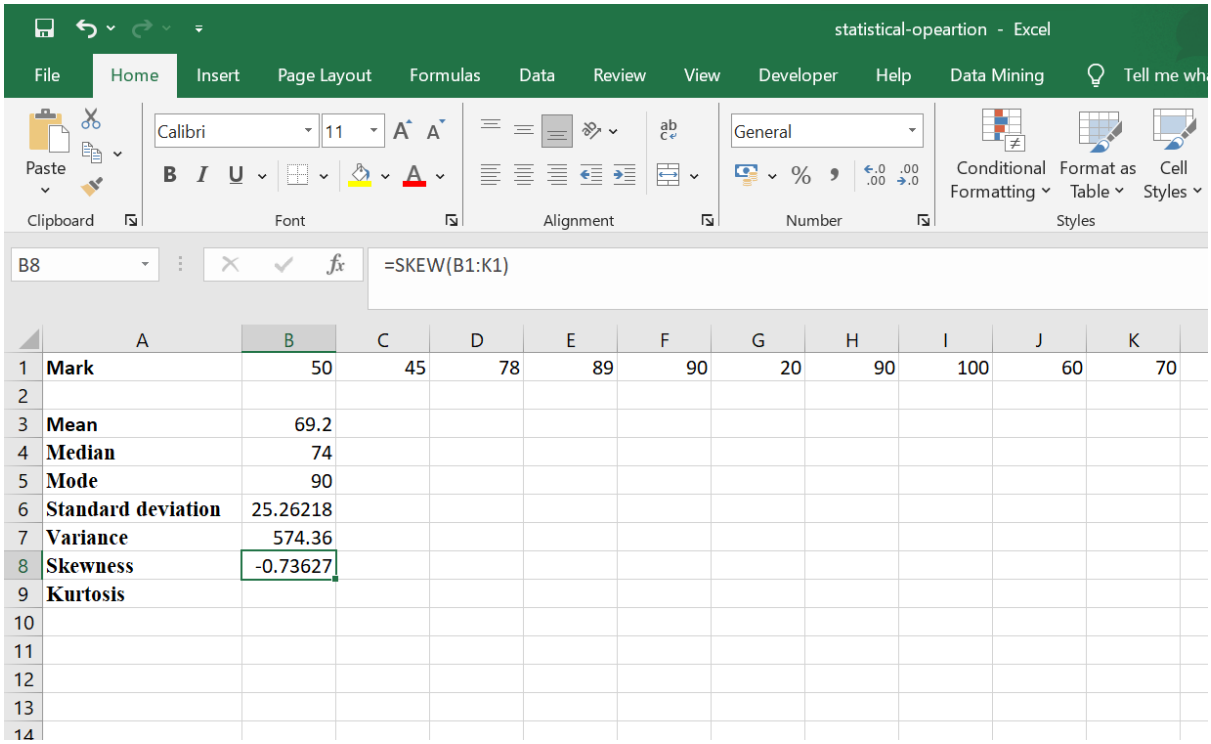
c. Select a range (B1:K1) and type close bracket)

d. And then Hit enter

Skewness:

Skewness is a statistical measure that quantifies the asymmetry or lack of symmetry in a probability distribution or set of data.

Returns the skewness of a distribution. Skewness characterizes the degree of asymmetry of a distribution around its mean. Positive skewness indicates a distribution with an asymmetric tail extending toward more positive values. Negative skewness indicates a distribution with an asymmetric tail extending toward more negative values.



	A	B	C	D	E	F	G	H	I	J	K
1	**Mark**	50	45	78	89	90	20	90	100	60	70
2											
3	**Mean**	69.2									
4	**Median**	74									
5	**Mode**	90									
6	**Standard deviation**	25.26218									
7	**Variance**	574.36									
8	**Skewness**	-0.73627									
9	**Kurtosis**										
10											
11											
12											
13											
14											

a. Select a cell (B8)

b. Type =SKEW(

c. Select a range (B1:K1) and type close bracket)

d. And then Hit enter

Kurtosis:

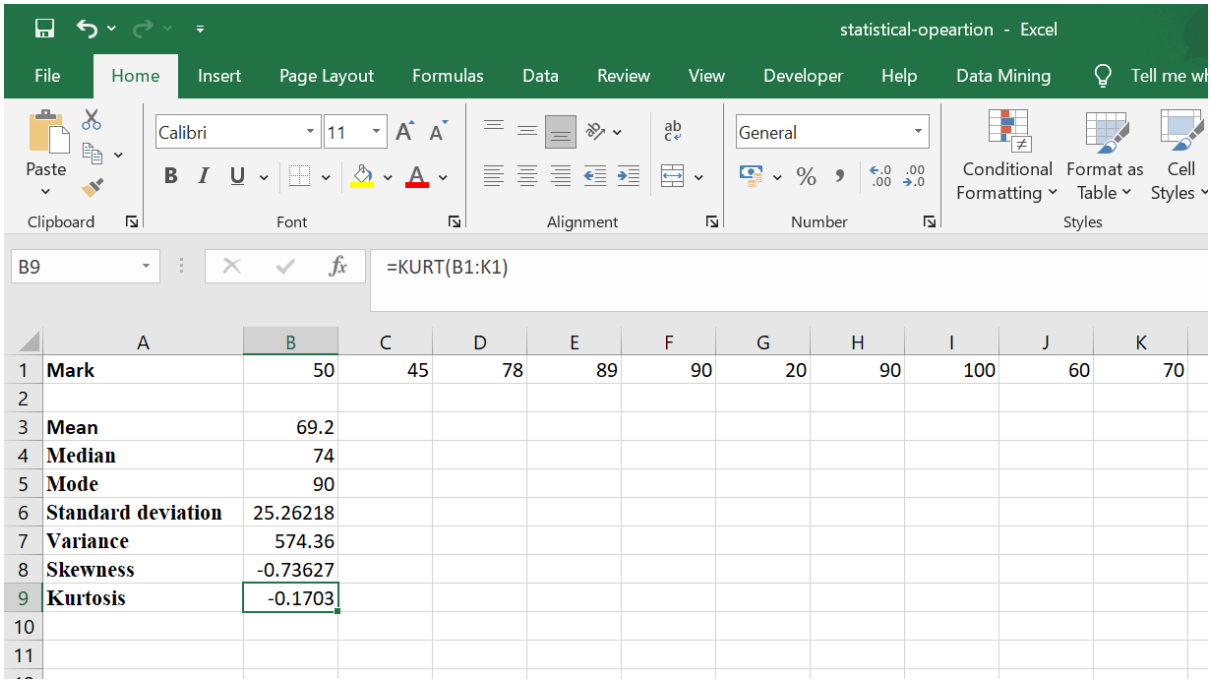
Returns the kurtosis of a data set. Kurtosis characterizes the relative peakedness or flatness of a distribution compared with the normal distribution. Positive kurtosis indicates a relatively peaked distribution. Negative kurtosis indicates a relatively flat distribution.

a. Select a cell (B3)

b. Type =KURT(

c. Select a range (B1:K1) and type close bracket)

d. And then Hit enter



	A	B	C	D	E	F	G	H	I	J	K
1	**Mark**	50	45	78	89	90	20	90	100	60	70
2											
3	**Mean**	69.2									
4	**Median**	74									
5	**Mode**	90									
6	**Standard deviation**	25.26218									
7	**Variance**	574.36									
8	**Skewness**	-0.73627									
9	**Kurtosis**	-0.1703									
10											
11											

RESULT:

The statistical operations - Mean, Median, Mode and Standard deviation, Variance, Skewness, Kurtosis are performed successfully.

Perform ANOVA, t-test, and z-test in Microsoft Excel with sample input data

Ex.No:4

Date:

AIM:

To perform ANOVA, t-test, and z-test in Microsoft Excel with sample input data.

PROCEDURE:

ANOVA

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 the data scientist who have a degree in statistic, computer science or physics.

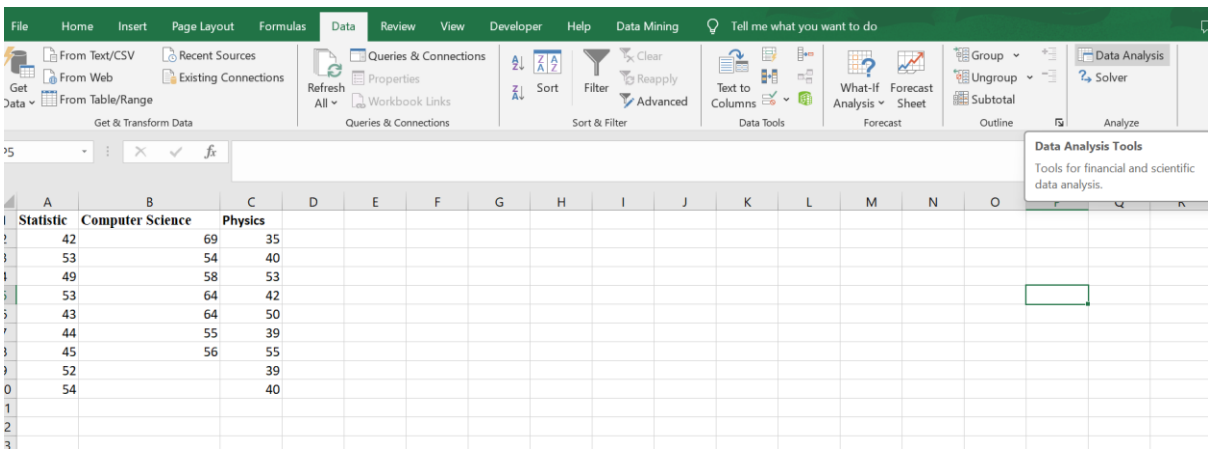
H0: $\mu 1 = \mu 2 = \mu 3$

H1: at least one of the means is different.

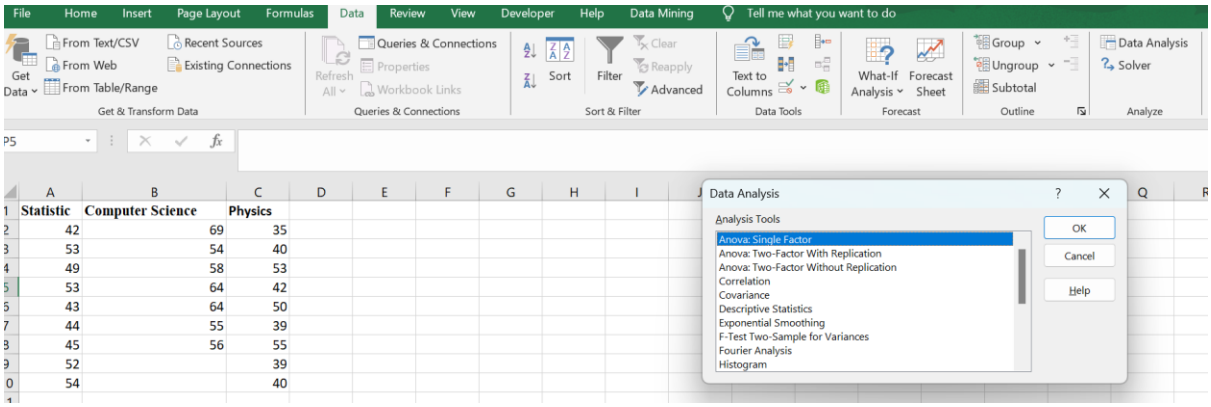
1. Enter your data into a table.

	A	B	C	D
1	**Statistic**	**Computer Science**	**Physics**	
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				

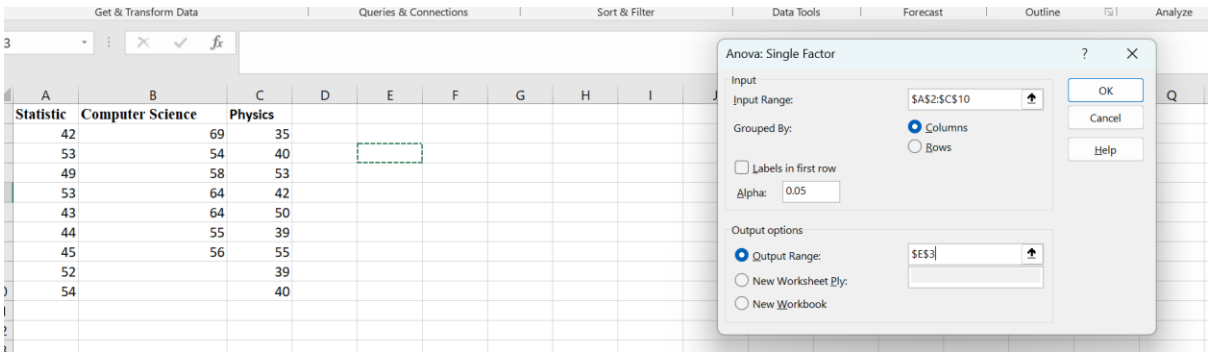
2. Click the Data tab and click Data Analysis.



3. In the Data Analysis dialog box, select ANOVA: Single Factor.



4. Click in the Input Range box and select the range $A$2:$C$10.
5. Click the Output Range box and enter the cell ($E$3) where you want the ANOVA output to appear.



6. Click OK. The Result

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				

Conclusion: if F > F crit, we reject the null hypothesis. This is the case, 15.196 > 3.443. Therefore, we reject the null hypothesis. The means of the three populations are not all equal. At least one of the means is different. However, the ANOVA does not tell you where the difference lies.

T-test

This example teaches you how to perform a t-Test in Excel. The t-Test is used to test the null hypothesis that the means of two populations are equal.

Below you can find the cricket score of 7 team A players and 6 team B players.

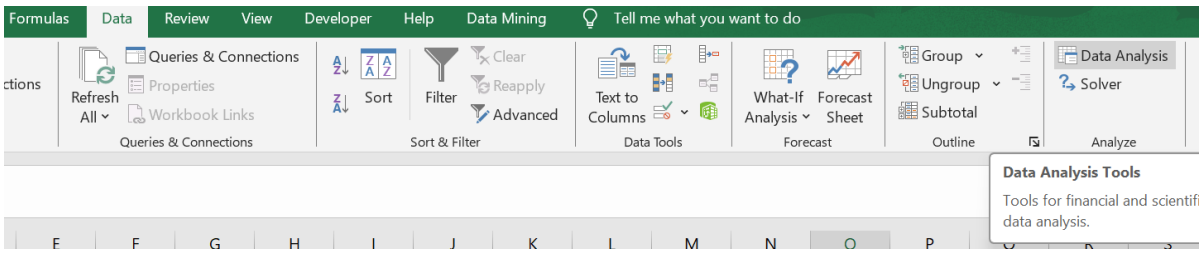
H0: $\mu 1 - \mu 2 = 0$

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

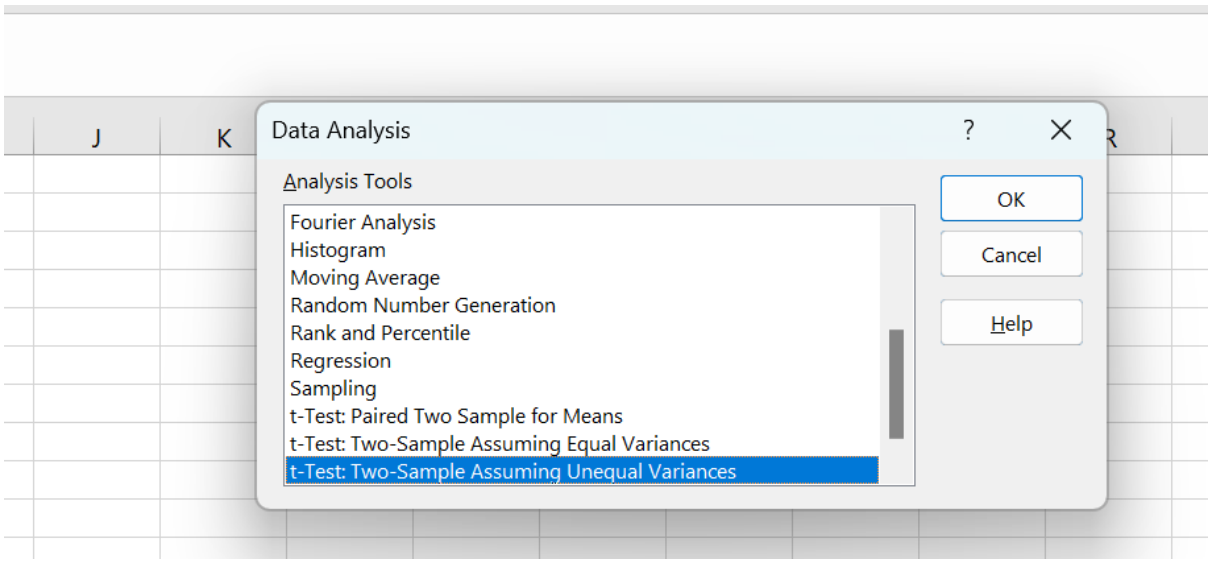
	A	B	C
1	Team A	Team B	
2	45	8	
3	50	30	
4	21	54	
5	35	25	
6	18	34	
7	52	12	
8	23		

To perform a t-Test, execute the following steps.

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

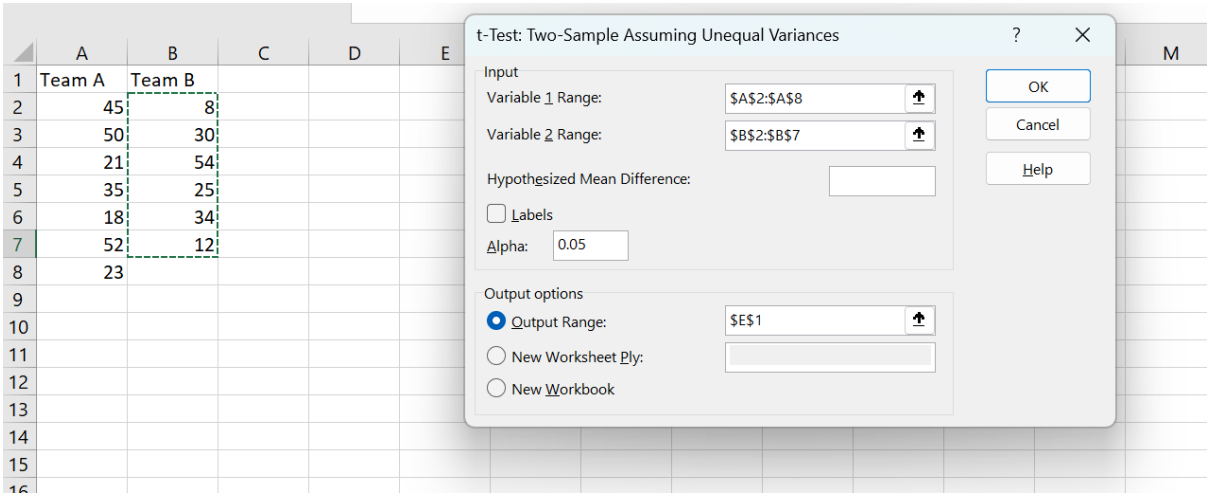


2. Select t-Test: Two-Sample Assuming Unequal Variances and click OK.

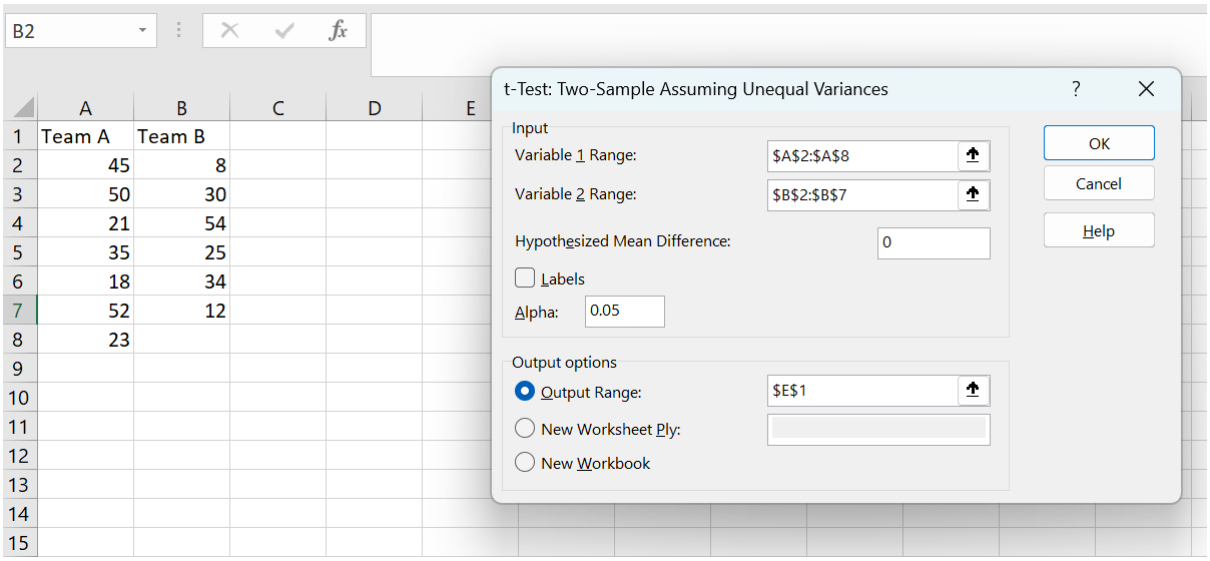


3. Click in the Variable 1 Range box and select the range A2:A8

4. Click in the Variable 2 Range box and select the range B2:B7.



5. Click in the Hypothesized Mean Difference box and type 0 (H0: μ1 - μ2 = 0).



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

7. The result will be.

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	36.83333333	30.2
Variance	215.7666667	275.2
Observations	6	5
Hypothesized Mean Difference	0	
df	8	
t Stat	0.69535803	
P(T<=t) one-tail	0.253261602	
t Critical one-tail	1.859548038	
P(T<=t) two-tail	0.506523204	
t Critical two-tail	2.306004135	

Z-test

While using the Z Test, we test a null hypothesis that states that the two population's mean is equal.

i.e.

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

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

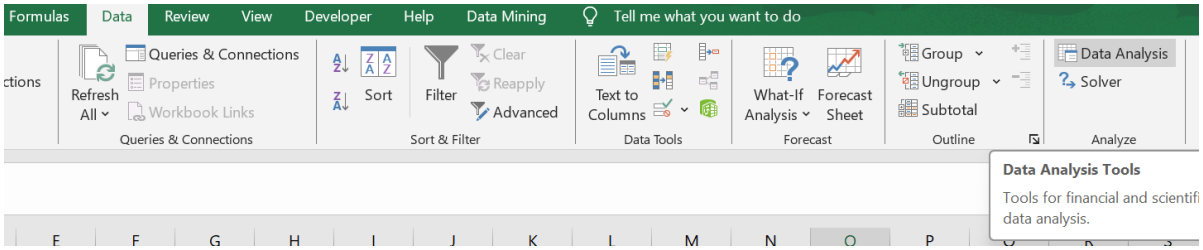
Where H1 is called an alternative hypothesis, the mean of the two populations is not equal.

Let's take the example of student's marks in two different subjects.

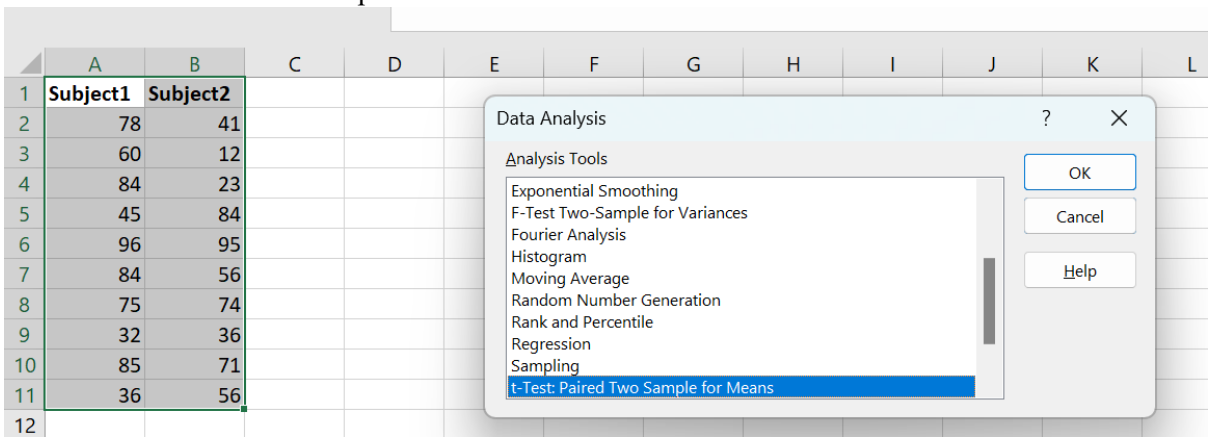
	A	B	C
1	**Subject1**	**Subject2**	
2	78	41	
3	60	12	
4	84	23	
5	45	84	
6	96	95	
7	84	56	
8	75	74	
9	32	36	
10	85	71	
11	36	56	
12			

To perform a t-Test, execute the following steps.

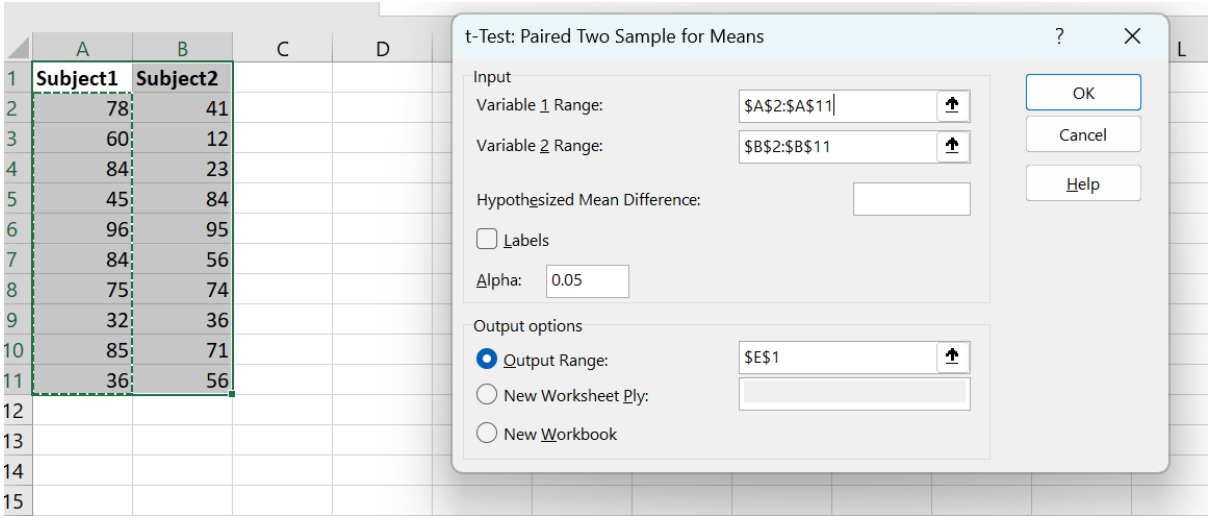
1. On the Data tab, in the Analysis group, click Data Analysis.



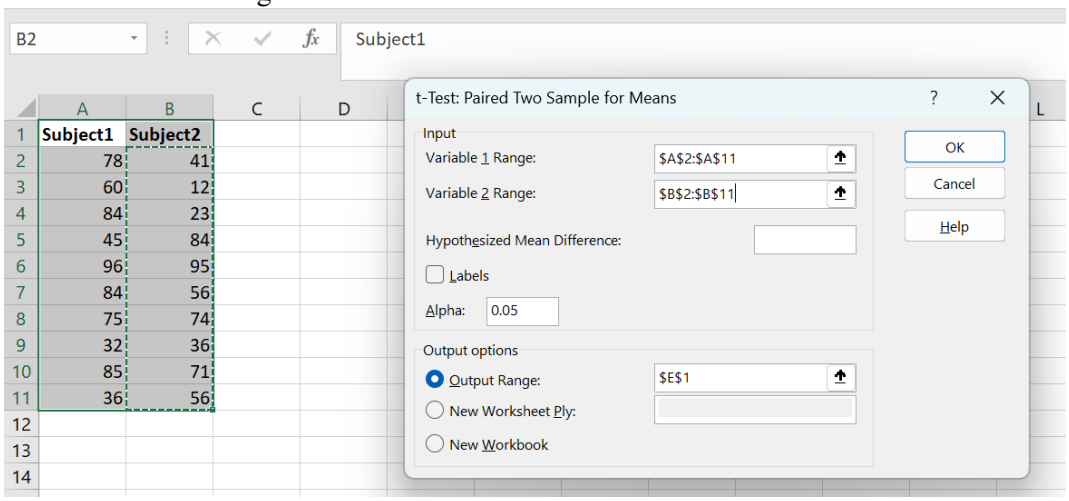
2. Select t-Test: Paired Two Sample for Means Ok



3. Select Variable 1 Range and Select A2:A11



4. Select Variable 2 Range and Select B2:B11



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

6. The output will be

E	F	G
t-Test: Paired Two Sample for Means		
	Variable 1	*Variable 2*
Mean	67.5	54.8
Variance	516.0555556	721.0667
Observations	10	10
Pearson Correlation	0.221672659	
Hypothesized Mean Difference	0	
df	9	
t Stat	1.291703551	
P(T<=t) one-tail	0.114321786	
t Critical one-tail	1.833112933	
P(T<=t) two-tail	0.228643573	
t Critical two-tail	2.262157163	

Result:

The ANOVA, t-test, and z-test in Microsoft Excel with sample input data are performed successfully and the output is verified.

Perform data pre-processing operations i) Handling Missing data

Exp No: 5.1

Date:

AIM:

To perform data pre-processing operations: Handling Missing data in Microsoft Excel

Procedure:

Handling missing data is a crucial step in the data pre-processing phase. In Microsoft Excel, there are several methods to deal with missing data. Here we handle missing value by replace with mean, median and mode .Here's a step-by-step guide using a simple example:

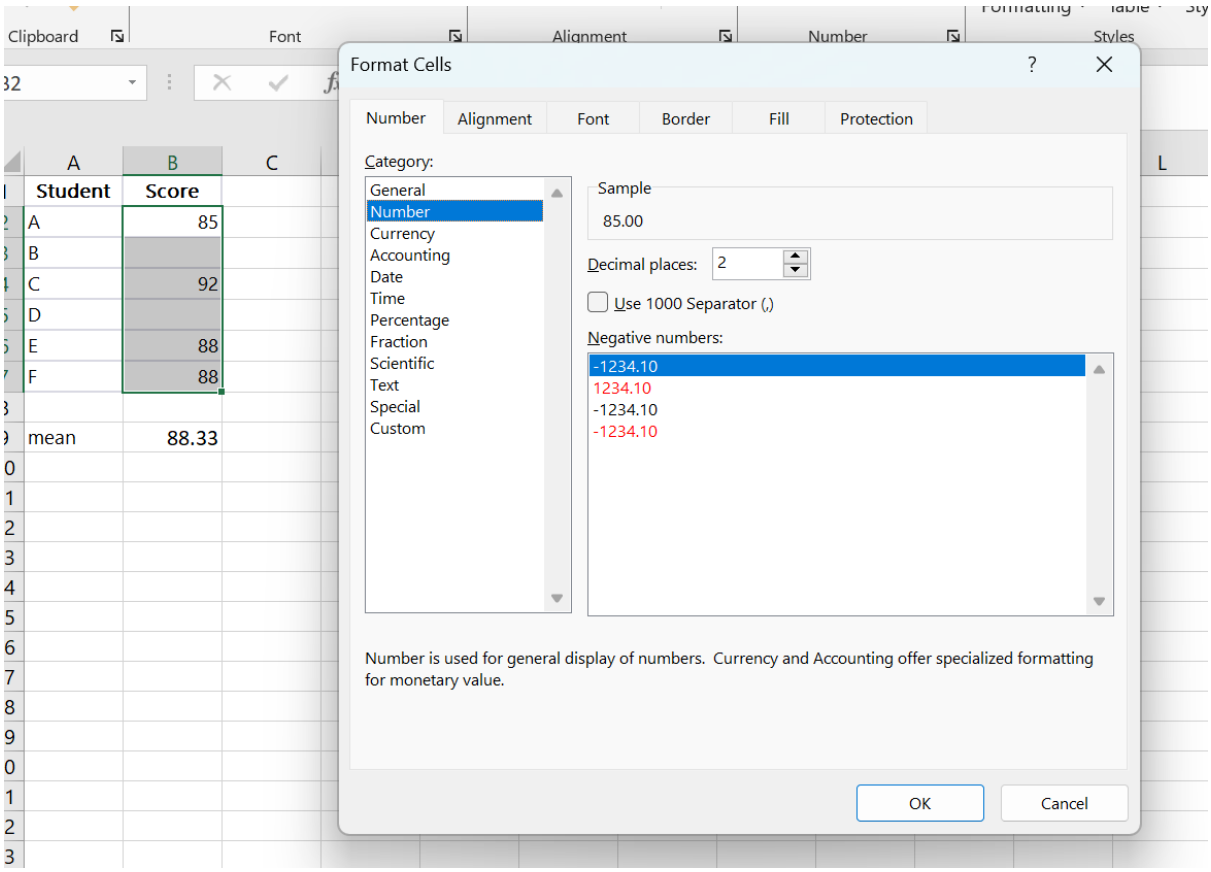
Example Scenario:

Suppose you have a dataset with missing values in the "Score" column.

	A	B
1	**Student**	**Score**
2	A	85
3	B	
4	C	92
5	D	
6	E	88
7	F	88
8		

Method 1: Replace with Mean

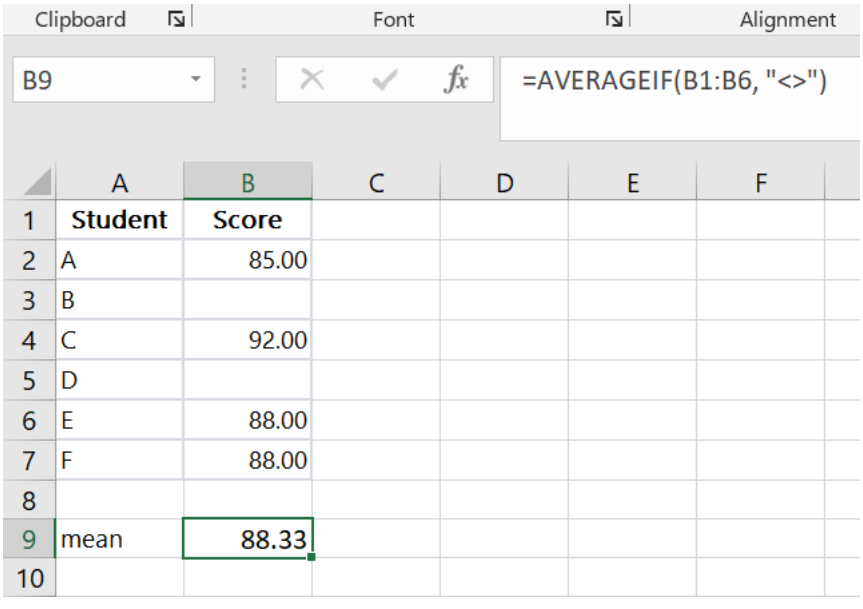
1. Right the range of cell B2:B7 and format cells and choose category number and choose decimal place to 2



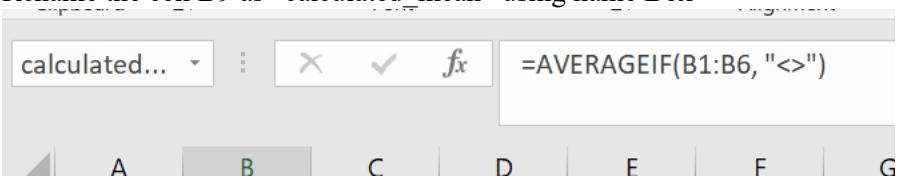
2. Calculate the mean of the existing scores (excluding blanks) using the AVERAGEIF function:

   =AVERAGEIF(B2:B7, "<>")

   This formula calculates the mean of the existing scores.

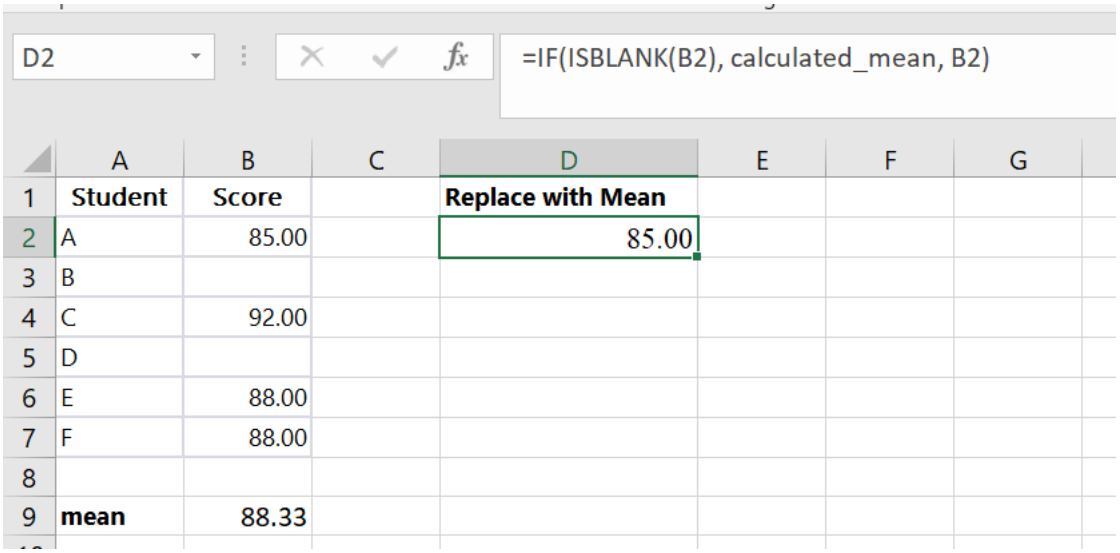


3. Rename the cell B9 as “calculated_mean” using name Box



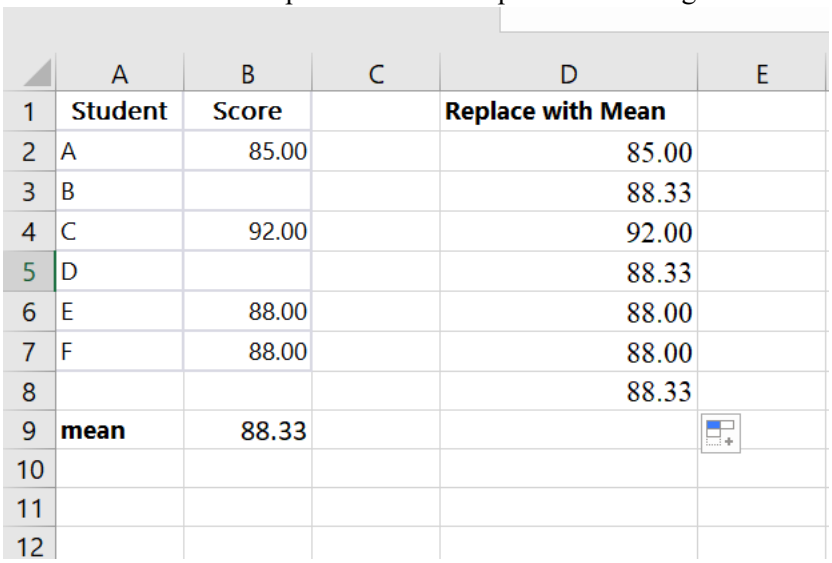
4. Replace the missing values with the calculated mean using the IF and ISBLANK functions:

=IF(ISBLANK(B2), calculated_mean, B2)



	A	B	C	D	E	F	G
1	**Student**	**Score**		**Replace with Mean**			
2	A	85.00		85.00			
3	B						
4	C	92.00					
5	D						
6	E	88.00					
7	F	88.00					
8							
9	**mean**	88.33					

5. And Track the D2 cell up to D7 and do required formatting



	A	B	C	D	E
1	**Student**	**Score**		**Replace with Mean**	
2	A	85.00		85.00	
3	B			88.33	
4	C	92.00		92.00	
5	D			88.33	
6	E	88.00		88.00	
7	F	88.00		88.00	
8				88.33	
9	**mean**	88.33			
10					
11					
12					

Method 2: Replace with Median

1. Calculate the median of the existing scores (excluding blanks) using the MEDIAN function:
   =MEDIAN(IF(B2:B7<>"", B2:B7))
   Enter this as an array formula by pressing Ctrl + Shift + Enter. And rename the B10 with calculated_median

B10 | {=MEDIAN(IF(B2:B7<>"", B2:B7))}

	A	B	C	D	E	F
1	**Student**	**Score**		**Replace with Mean**		
2	A	85.00		85.00		
3	B			88.33		
4	C	92.00		92.00		
5	D			88.33		
6	E	88.00		88.00		
7	F	88.00		88.00		
8				88.33		
9	**mean**	88.33				
10	**Median**	88				
11						

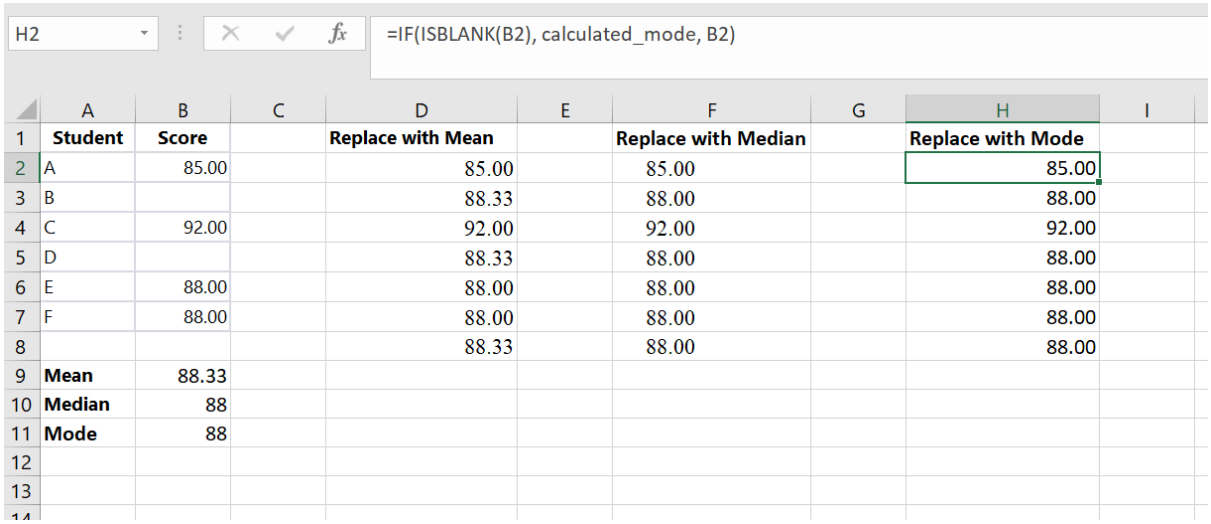
2. Replace the missing values with the calculated median using the IF and ISBLANK functions: =IF(ISBLANK(B2), calculated_median, B2)
3. Drag the formula down to apply it to other cells and do required formatting, the output will be

F2 | =IF(ISBLANK(B2), calculated_median, B2)

	A	B	C	D	E	F	G
1	**Student**	**Score**		**Replace with Mean**		**Replace with Median**	
2	A	85.00		85.00		85.00	
3	B			88.33		88.00	
4	C	92.00		92.00		92.00	
5	D			88.33		88.00	
6	E	88.00		88.00		88.00	
7	F	88.00		88.00		88.00	
8				88.33		88.00	
9	**mean**	88.33					
10	**Median**	88					
11							

Method 3: Replace with Mode

1. Calculate the median of the existing scores (excluding blanks) using the Mode function: =Mode(IF(B2:B7<>"", B2:B7))
   Enter this as an array formula by pressing Ctrl + Shift + Enter. And rename the B10 with calculated_mode
4. Replace the missing values with the calculated median using the IF and ISBLANK functions: =IF(ISBLANK(B2), calculated_mode, B2)
2. Drag the formula down to apply it to other cells and do required formatting, the output will be



	A	B	C	D	E	F	G	H	I
1	**Student**	**Score**		**Replace with Mean**		**Replace with Median**		**Replace with Mode**	
2	A	85.00		85.00		85.00		85.00	
3	B			88.33		88.00		88.00	
4	C	92.00		92.00		92.00		92.00	
5	D			88.33		88.00		88.00	
6	E	88.00		88.00		88.00		88.00	
7	F	88.00		88.00		88.00		88.00	
8				88.33		88.00		88.00	
9	**Mean**	88.33							
10	**Median**	88							
11	**Mode**	88							
12									
13									

Result:

The data pre-processing operations: Handling Missing data by replace data with mean, median, and mode in Microsoft Excel is preformed successfully and the output verified.

Perform data pre-processing operations: Normalization

Ex.No:5.2

Date:

Aim:

The aim of this data pre-processing operation is to normalize numerical features within a dataset using Min-Max Scaling. Normalization ensures that features with different scales are brought to a standard range, typically between 0 and 1, facilitating fair comparisons and improving the performance of certain machine learning algorithms.

Procedure:

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:

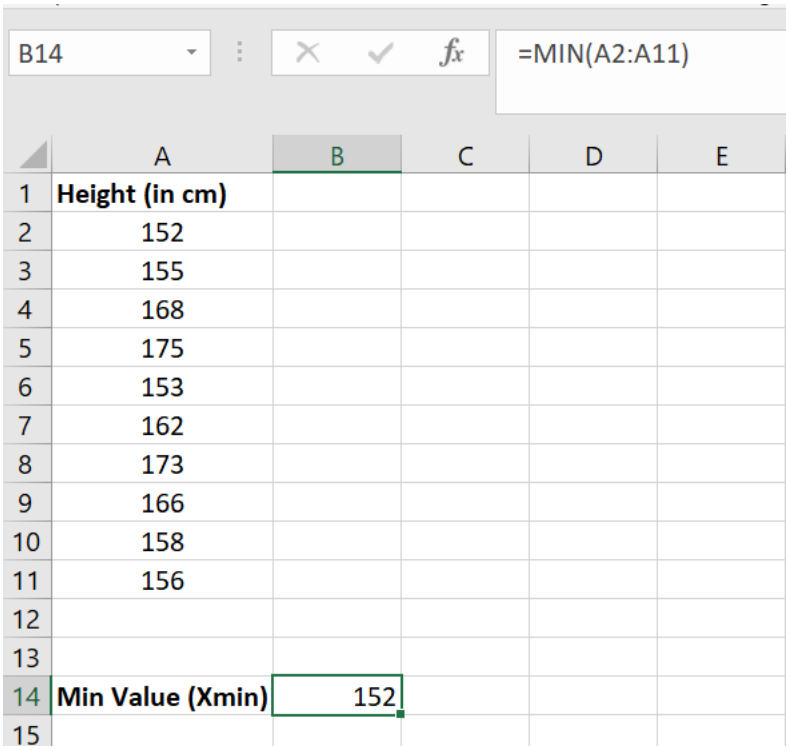
$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

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.

Let's have a look at one example to see how can we perform normalization on a sample dataset. Suppose, we have a record of the height of 10 students inside a class as shown below:

	A	B
1	**Height (in cm)**	
2	152	
3	155	
4	168	
5	175	
6	153	
7	162	
8	173	
9	166	
10	158	
11	156	
12		
13		

Step 1: Calculate the minimum value in the distribution. It can be calculated using the MIN() function. The minimum value comes out to be 152 which is stored in the B14 cell.



	A	B	C	D	E
1	**Height (in cm)**				
2	152				
3	155				
4	168				
5	175				
6	153				
7	162				
8	173				
9	166				
10	158				
11	156				
12					
13					
14	**Min Value (Xmin)**	152			
15					

Step 2: Calculate the maximum value in the distribution. It can be calculated using the MAX() function. The maximum value comes out to be 175 which is stored in the B15 cell.

	A	B	C
1	**Height (in cm)**		
2	152		
3	155		
4	168		
5	175		
6	153		
7	162		
8	173		
9	166		
10	158		
11	156		
12			
13			
14	**Min Value (Xmin)**	152	
15	**Max Value(Xmax)**	175	
16			
17			

Step 3: Find the difference between the maximum and minimum values. Their difference comes out to be 175 – 152 = 23 which is stored in the B16 cell.

B16 =B15-B14

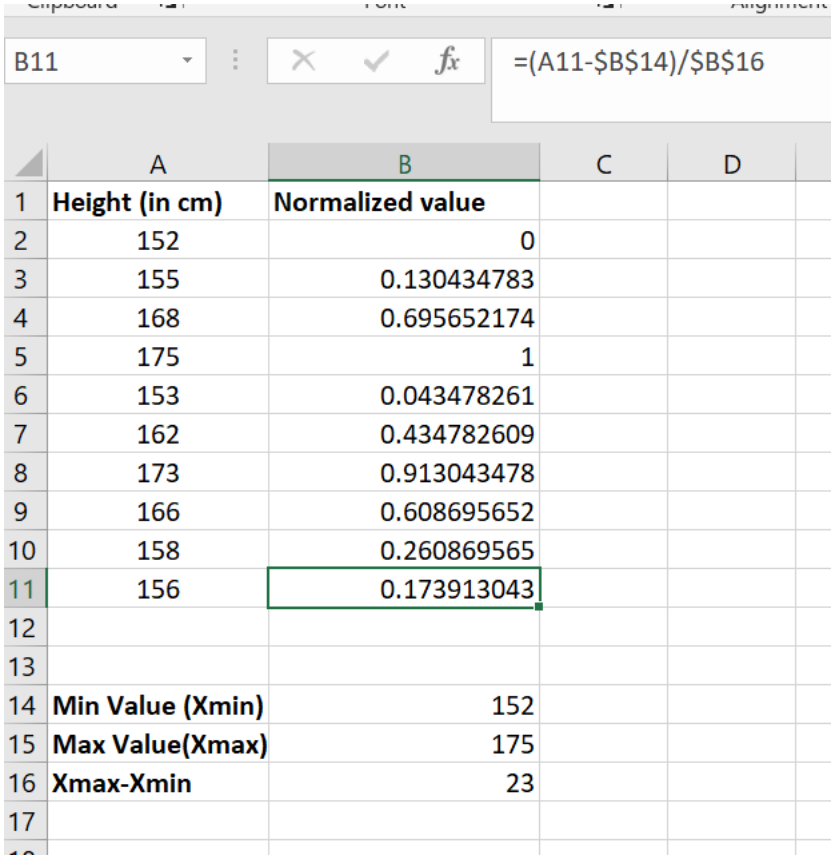
	A	B	C	D
1	**Height (in cm)**			
2	152			
3	155			
4	168			
5	175			
6	153			
7	162			
8	173			
9	166			
10	158			
11	156			
12				
13				
14	**Min Value (Xmin)**	152		
15	**Max Value(Xmax)**	175		
16	**Xmax-Xmin**	23		
17				
18				

Step 4: For the first data stored in the A2 cell, we will calculate the normalized value as shown in the below

B2 =(A2-$B$14)/$B$16

	A	B	C	D	E
1	**Height (in cm)**				
2	152	0			
3	155				
4	168				
5	175				
6	153				
7	162				
8	173				
9	166				
10	158				
11	156				
12					
13					
14	**Min Value (Xmin)**	152			
15	**Max Value(Xmax)**	175			
16	**Xmax-Xmin**	23			
17					
18					

Step 5: We can manually calculate all values one by one for each data record or we can directly get values for all the other cells using the auto-fill feature of Excel. For this, go to the right corner of the B2 cell until a (+) symbol appears, and then drag the cursor to the bottom to auto-populate values inside all the cells.



	A	B	C	D
1	**Height (in cm)**	**Normalized value**		
2	152	0		
3	155	0.130434783		
4	168	0.695652174		
5	175	1		
6	153	0.043478261		
7	162	0.434782609		
8	173	0.913043478		
9	166	0.608695652		
10	158	0.260869565		
11	156	0.173913043		
12				
13				
14	**Min Value (Xmin)**	152		
15	**Max Value(Xmax)**	175		
16	**Xmax-Xmin**	23		
17				

Result:

The data pre-processing operations: Normalization is performed and the output is verified successfully.

Apply and explore various plotting functions on the data set

Ex.No:6

Date:

Aim:

The aim of this data analysis task is to visually explore and analyze the dataset through various plotting functions in Microsoft Excel. This aims to provide a graphical representation of the data, aiding in the identification of patterns, trends, and potential insights.

Procedure:

Step 1: Open Excel and Load Data

Open Microsoft Excel.

Load the dataset you want to explore into a new or existing worksheet.

Let the consider the car brands and car listings as a dataset.

	A	B
1	**Brand**	**Cars Listings**
2	Audi	700
3	BMW	636
4	Mercedes-Benz	820
5	Mitsubishi	320
6	Renault	438
7	Toyota	487
8	Volkswagen	875
9		

Step 2: Select Data for Plotting

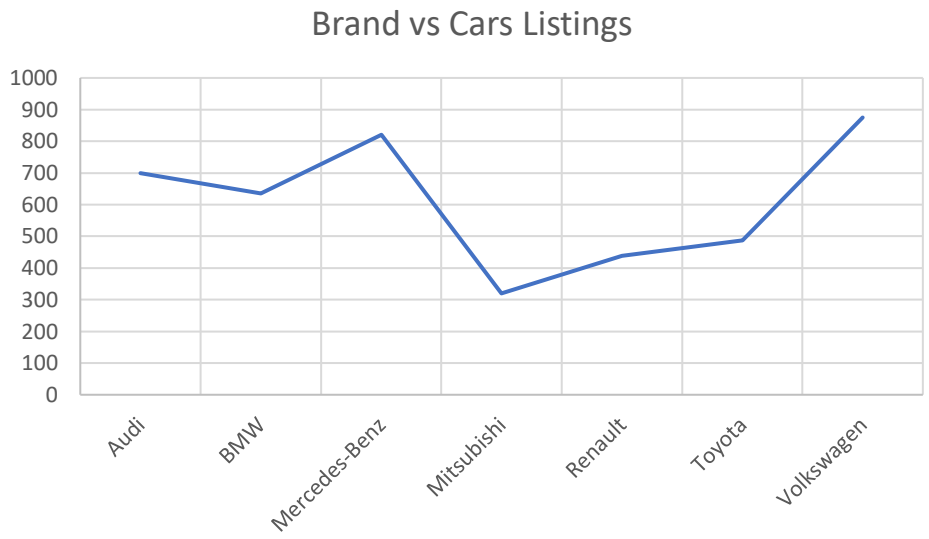
Highlight the range of data you want to visualize in your dataset.

Step 3: Explore Various Plotting Functions

Line Chart:

Create a line chart to show trends over time or across categories.

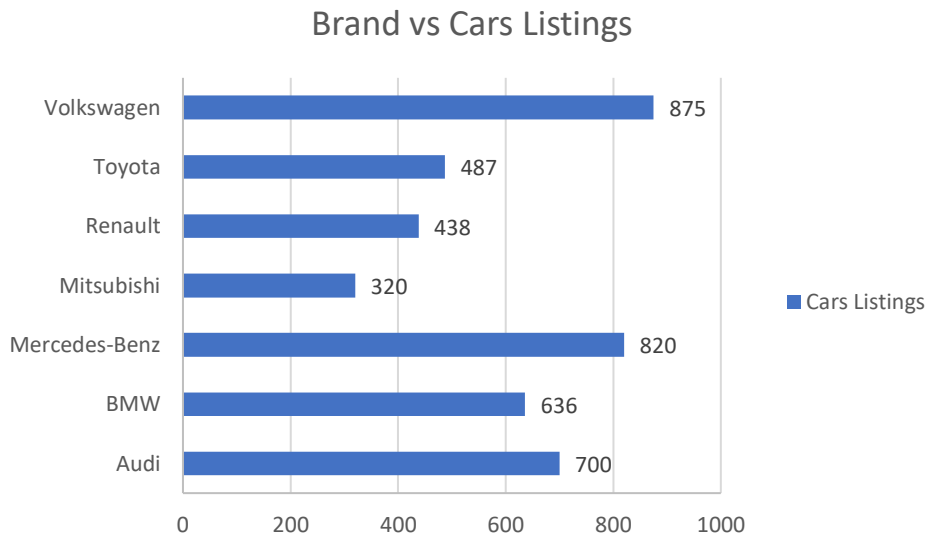
Go to the "Insert" tab, select "Line," and choose the appropriate line chart type.



Bar Chart:

Use a bar chart to compare values across different categories.

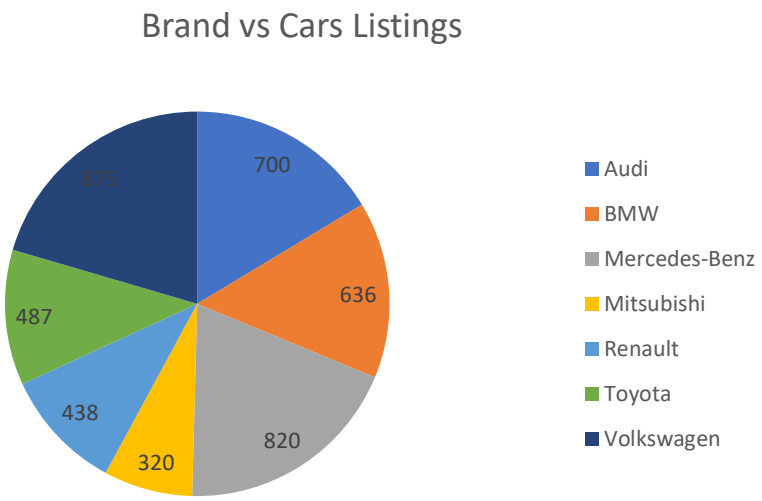
Go to the "Insert" tab, select "Bar," and choose the appropriate bar chart type.



Pie Chart:

Create a pie chart to show the proportion of each category in a dataset.

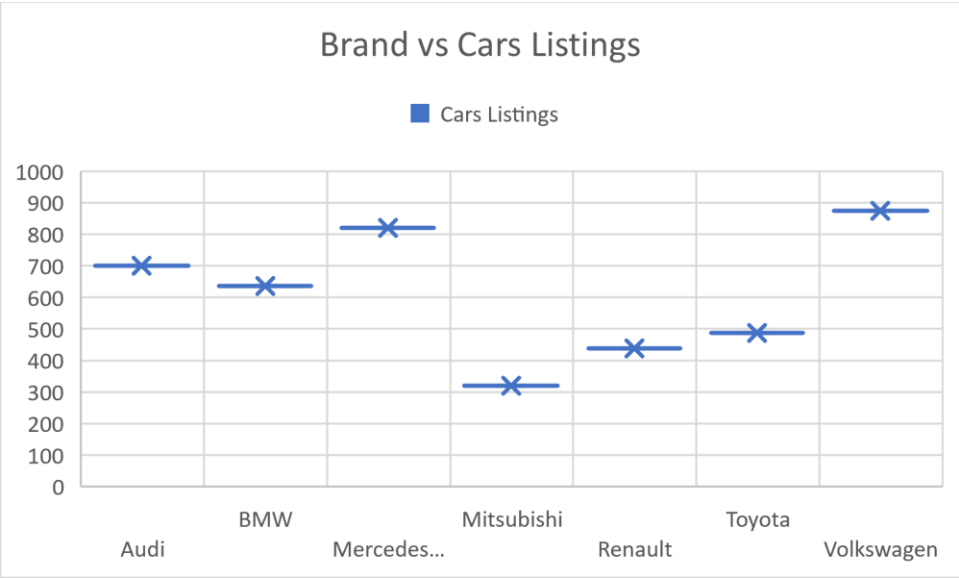
Go to the "Insert" tab, select "Pie," and choose the appropriate pie chart type.



Box-and-Whisker Plot (Box Plot):

Use a box plot to display the distribution and variability of a dataset.

Go to the "Insert" tab, select "Box & Whisker," and choose the appropriate box plot type.



Step 4: Customize and Format

Customize each plot based on your preferences.

Adjust labels, titles, colors, and other formatting options to enhance clarity.

Result:

The data analysis task is to visually explore and analyze the dataset through various plotting functions in Microsoft Excel are performed successfully.

Exploring Power BI Desktop for Business Analytics

Ex.No:7

Date:

Aim:

The aim of this lab is to study features of Power BI Desktop for effective business analytics.

Features of Power BI Desktop:

1. Data Connectivity:

Connect to various data sources, including Excel, databases (SQL Server, MySQL, etc.), cloud services (Azure, Google Analytics), and more.

Import or create a live connection to data.

2. Data Transformation and Cleaning:

Use Power Query Editor to shape and clean data.

Apply transformations, filter rows, merge tables, and handle missing values.

3. Data Modeling:

Create relationships between tables.

Define calculated columns and measures using Data Analysis Expressions (DAX).

Create hierarchies to enhance drill-down capabilities.

4. Data Visualization:

Build interactive and customizable reports.

Drag-and-drop elements onto the canvas to create visuals (charts, tables, maps, etc.).

Utilize a wide range of visualization options and customize formatting.

5. Dashboard Creation:

Combine visuals into dashboards for a comprehensive view.

Pin visuals to dashboards for quick access and easy sharing.

6. Q&A (Natural Language Query):

Use natural language queries to ask questions about your data.

Get instant visualizations based on your queries.

7. Power BI Service Integration:

Publish reports to the Power BI Service for online sharing and collaboration.

Schedule data refreshes for up-to-date reports.

Access reports from web browsers and mobile devices.

8. Advanced Analytics and AI Integration:

Integrate advanced analytics and machine learning models using Power BI's AI capabilities.

Leverage built-in AI visuals and features.

9. Custom Visuals and Extensions:

Import custom visuals from the Power BI marketplace.

Develop and use custom visuals and extensions.

10. Security and Sharing:

Implement row-level security to control access to data.

Share reports and dashboards with specific users or groups.

11. Report Storytelling:

Create report narratives and tell a compelling data-driven story.

Utilize bookmarks and buttons to guide users through a sequence of visuals.

12. Templates and Themes:

Create and apply templates for consistent report formatting.

Customize themes to match corporate branding.

Result:

The features of Power BI Desktop are explored successfully.

Data Preparation and Loading in Power BI Desktop

Ex.No:8

Date:

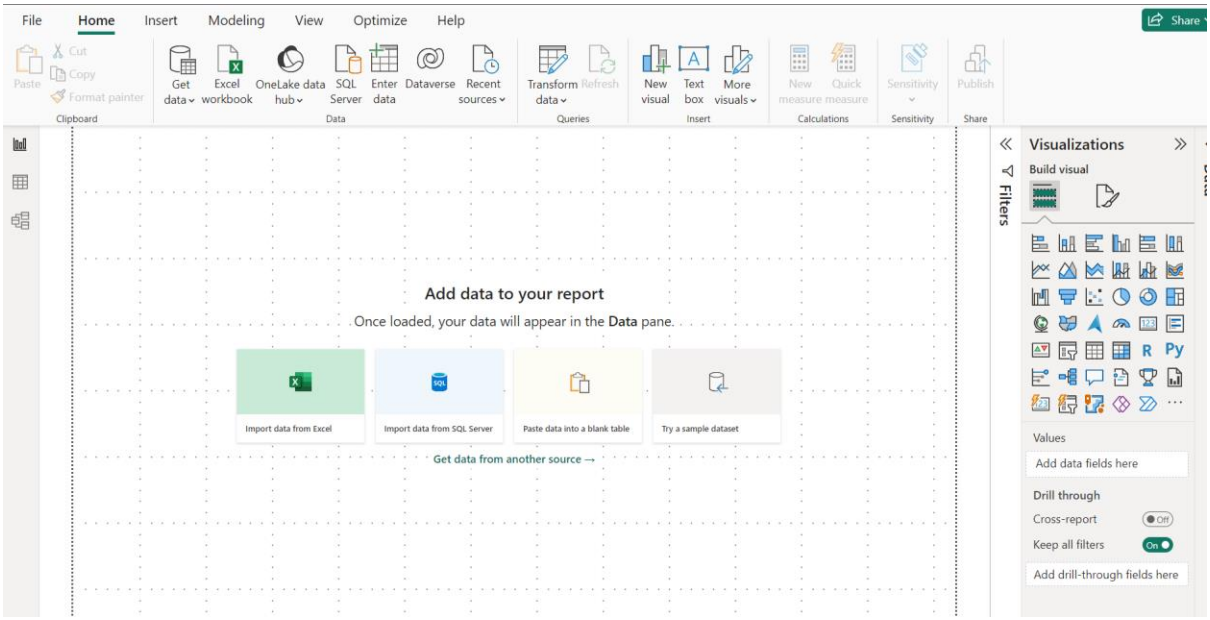
Aim:

The aim of this lab is to guide participants through the process of preparing and loading data into Power BI Desktop. Participants will learn how to import data from various sources, clean and transform it, and create a foundation for analysis.

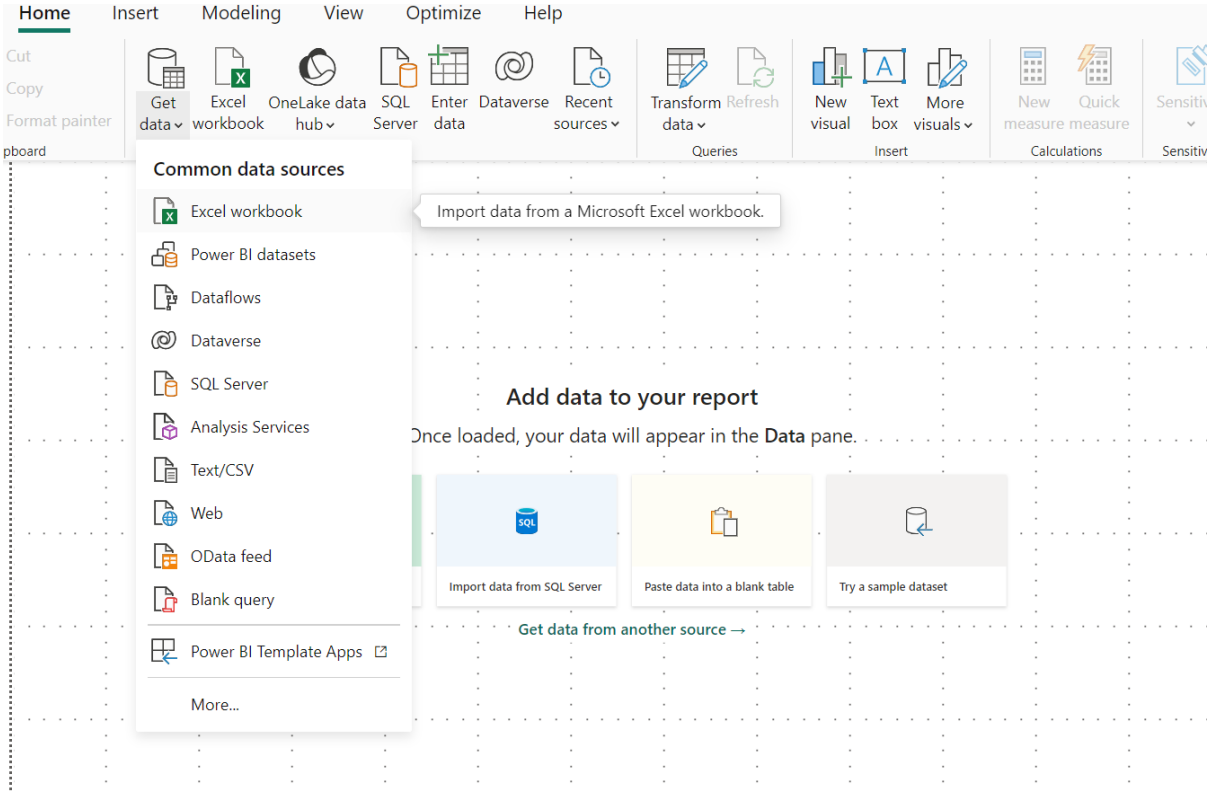
Procedure:

1. Connect to Data Sources:

a) Launch Power BI Desktop and click on the "Get Data" button in the Home ribbon.



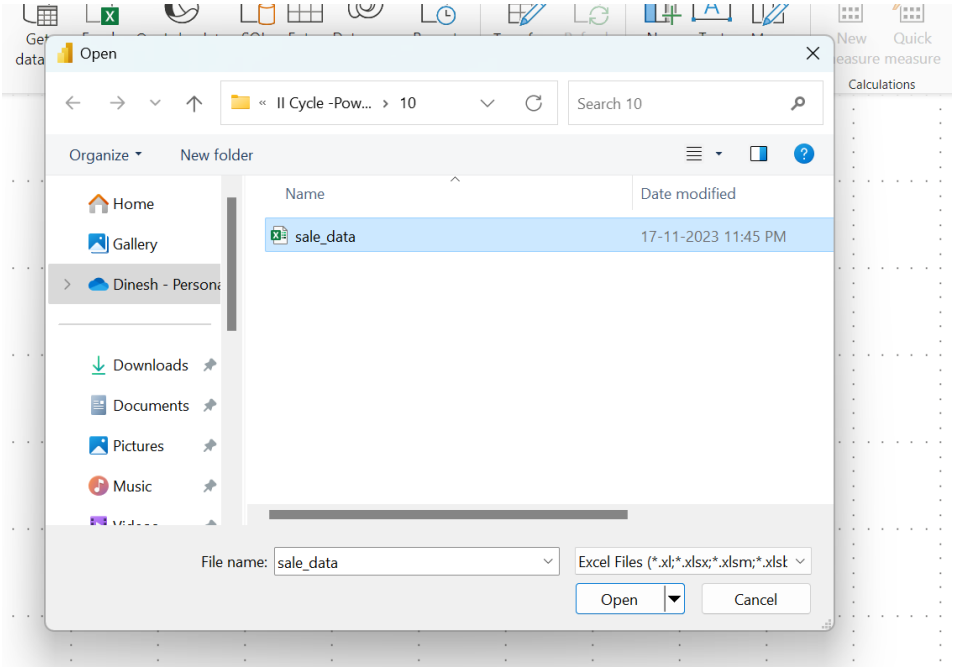
b) Select the appropriate data source from the list of options, such as Excel, SQL Server, or cloud-based services like Azure Data Lake Storage or Power BI Dataflows.



c) Provide the necessary connection details, such as server name, database credentials, or file location.

2. Transform and Clean Data:

a) Once connected to the data source, Power BI Desktop opens the Power Query Editor, a powerful data transformation tool.



b) Inspect the data to identify any inconsistencies, missing values, or data quality issues.

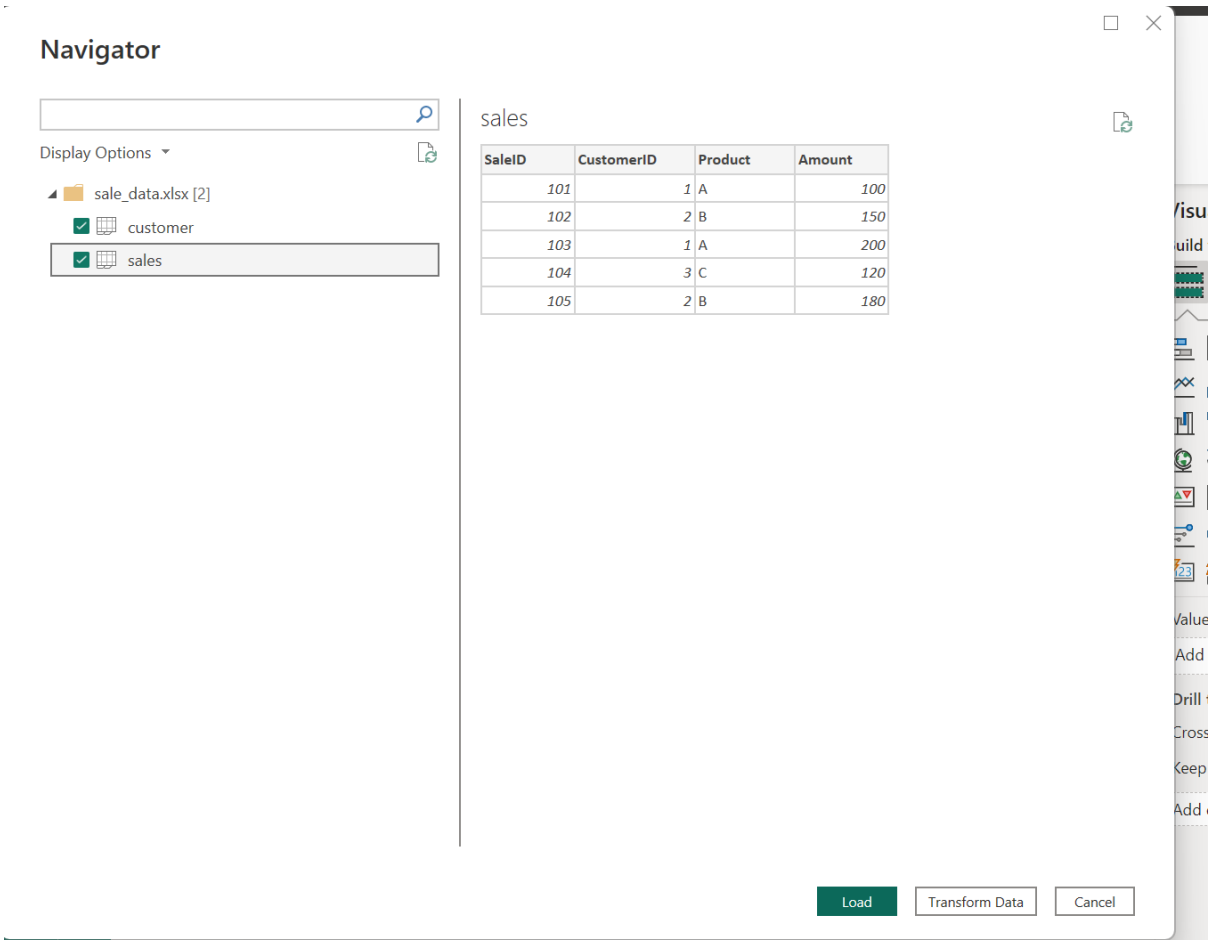
c) Use Power Query Editor's transformation functions to clean and shape the data.

d) Rename columns, remove duplicate rows, split or merge columns, and apply data type conversions as needed.

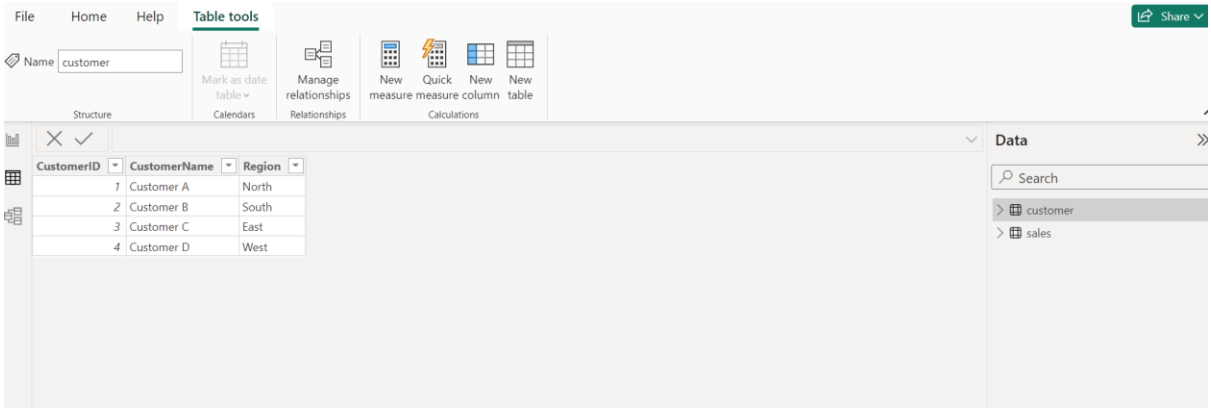
3. Load Data into Power BI Model:

a) Once the data is transformed and cleaned, it's ready to be loaded into the Power BI data model.

b) Click on the "Load" button in the Power Query Editor.



c) Power BI Desktop will import the transformed data into the data model, making it available for visualization and analysis.



4. Manage Data Relationships:

a) If you have multiple tables, establish relationships between them to create a cohesive data model.

b) Drag fields from one table to corresponding fields in another table to establish relationships.

c) Relationships enable Power BI to understand how tables are connected and combine data appropriately for analysis.

5. Data Validation and Refinement:

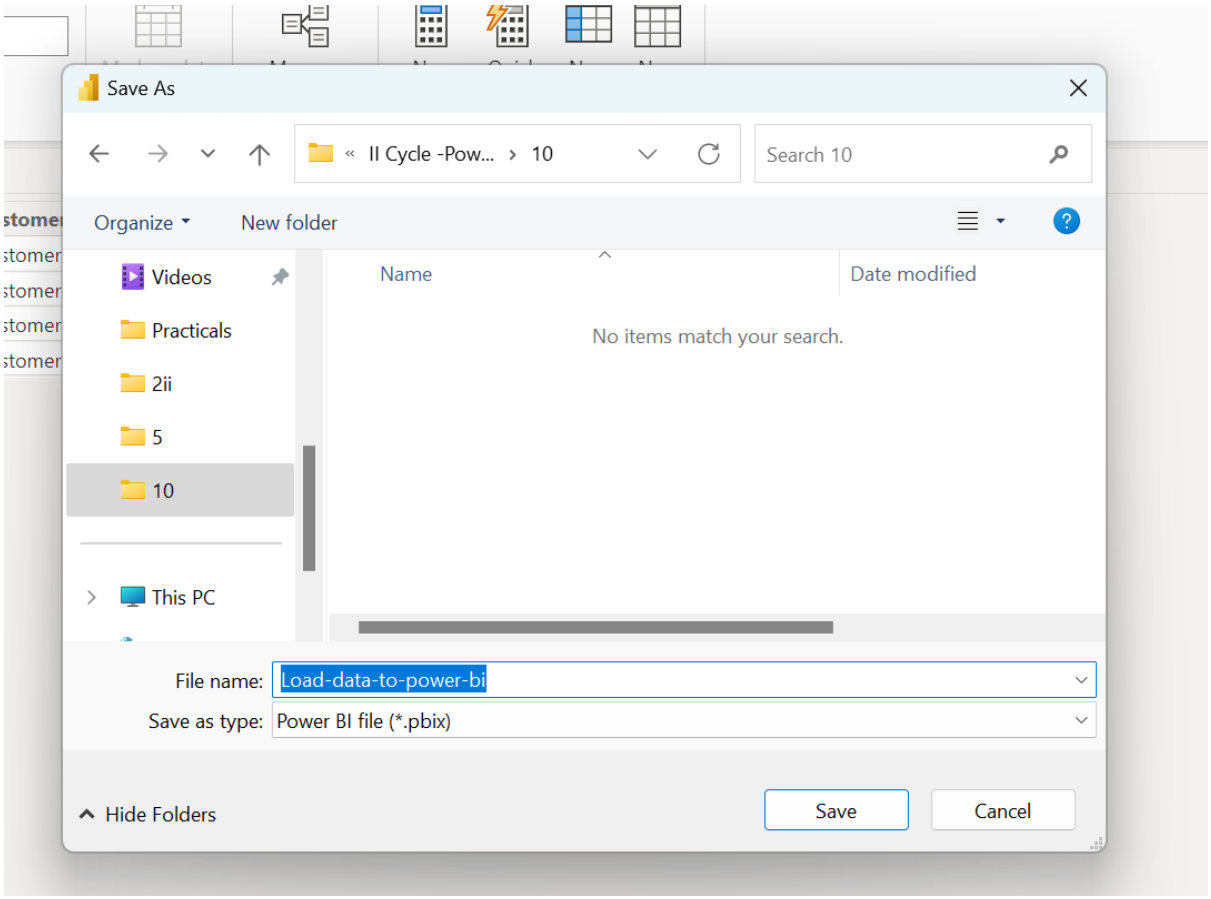
a) Review the data in the Power BI Desktop report view to ensure it's loaded correctly and looks as expected.

b) Use data filtering and sorting options to explore the data further and identify any potential issues.

c) Refine data transformations or relationships if necessary to ensure the data is accurately represented in the Power BI model.

6. Save and Share Your Work:

a) Save your Power BI Desktop file (.pbix) to keep your progress and data modifications.



b) Share your Power BI report with others by publishing it to the Power BI service or exporting it to a format like PDF or PowerPoint.

Result:

Upon completion of this lab, participants will have gained practical experience in preparing and loading data using Power BI Desktop. They will be able to import data from various sources, clean and transform it effectively, and load it into Power BI for further analysis.

Develop the data model using power BI Desktop

Ex.No:9

Date:

Aim:

To develop a simple data model using power BI and create powerful visualizations and gain insights into your data

Procedure:

Suppose you have two tables: one containing information about customers and another containing sales data.

Customers Table:

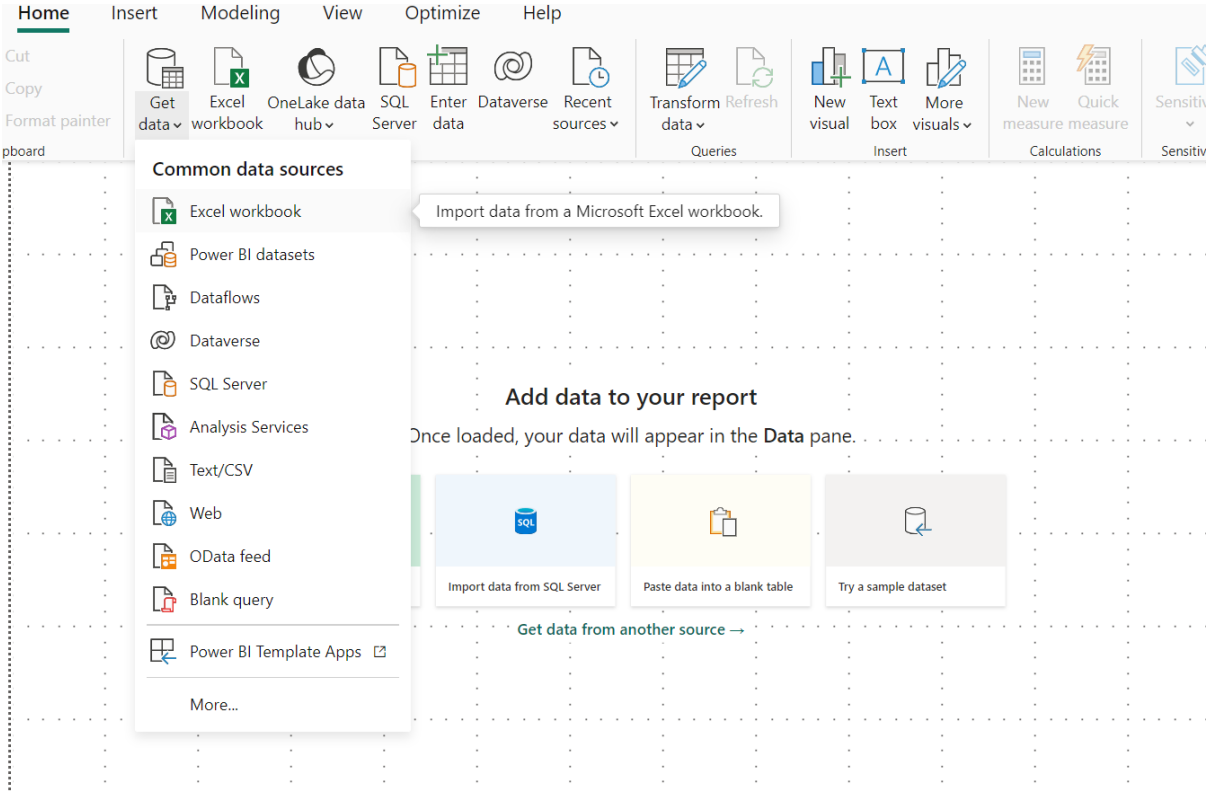
CustomerID	CustomerName	Region
1	Customer A	North
2	Customer B	South
3	Customer C	East
4	Customer D	West

Sales Table:

SaleID	CustomerID	Product	Amount
101	1	A	100
102	2	B	150
103	1	A	200
104	3	C	120
105	2	B	180

Steps to Develop the Data Model:

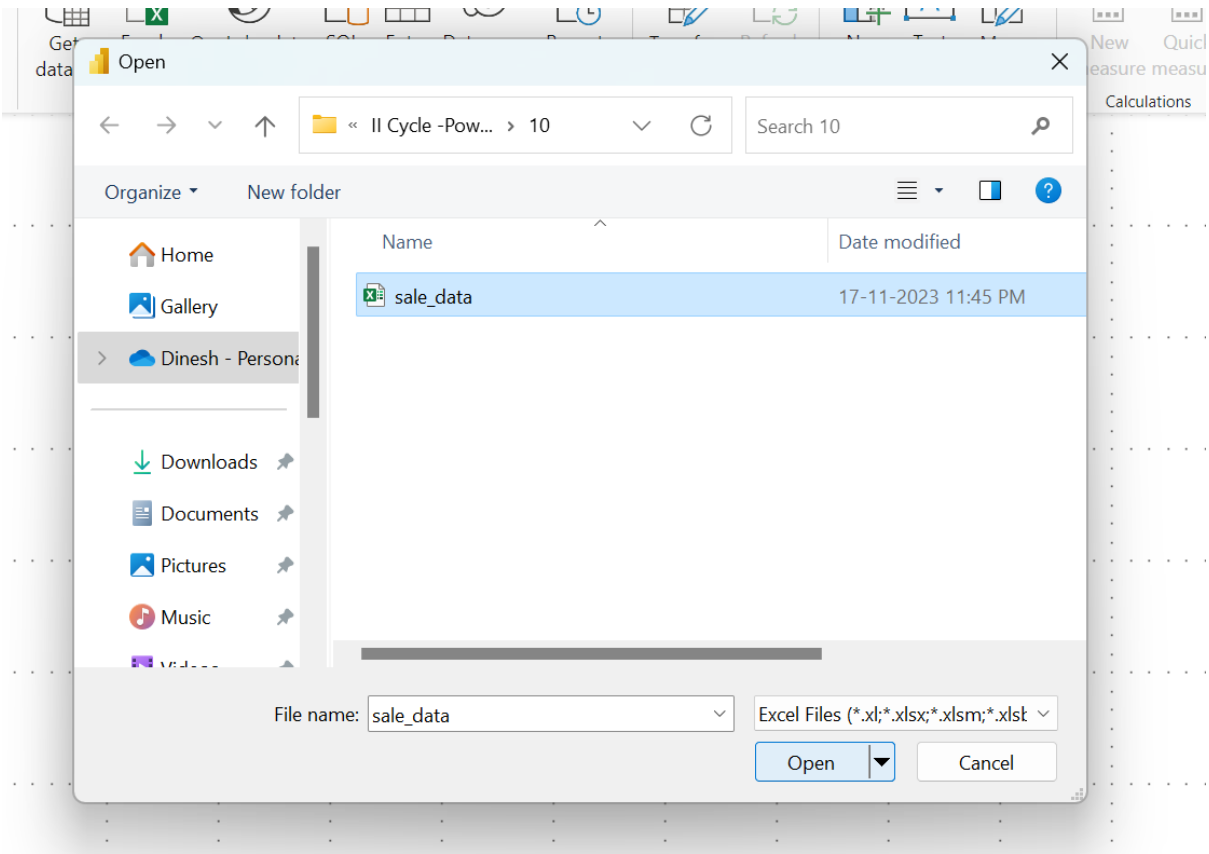
Step 1: Open Power BI Desktop



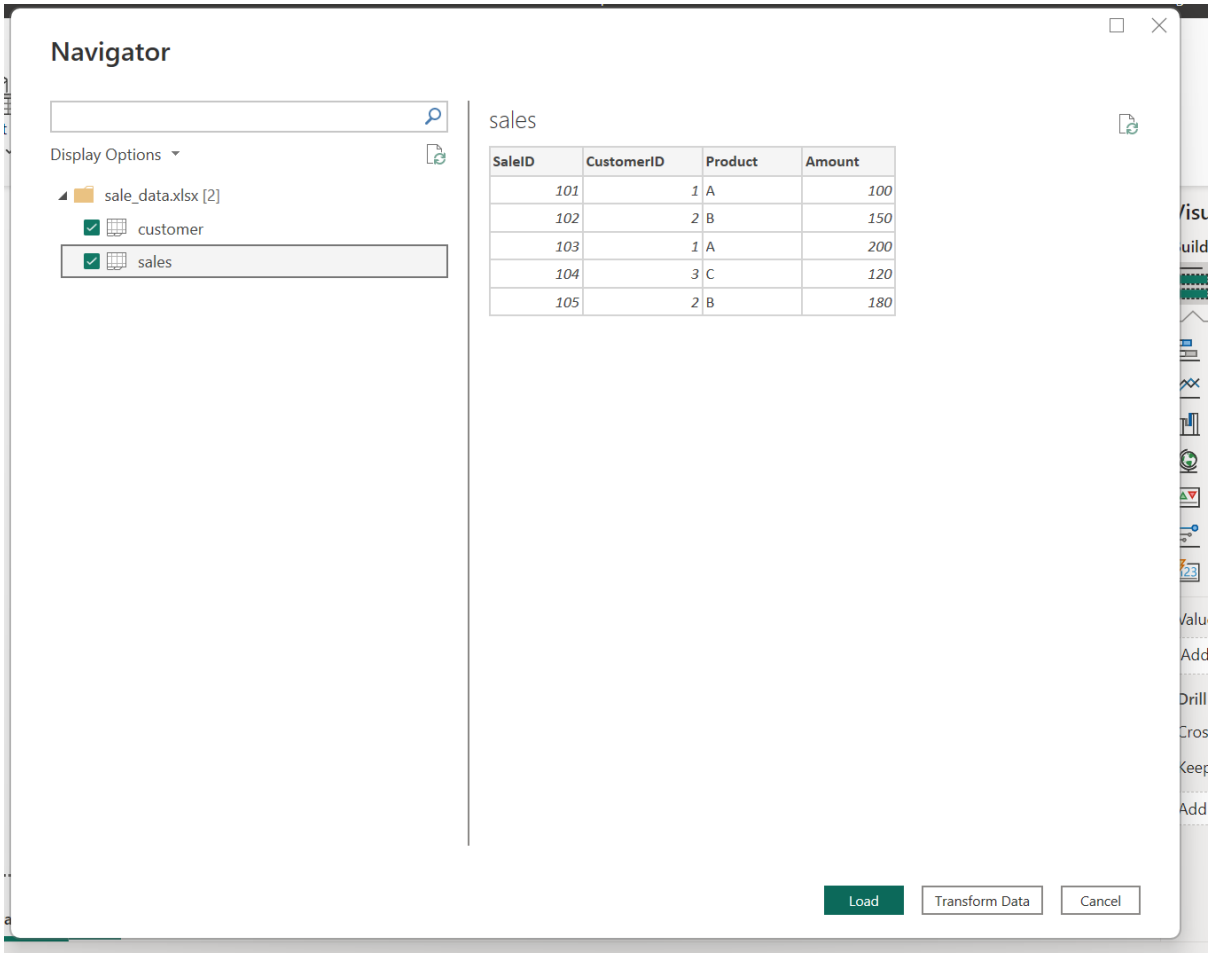
Open Power BI Desktop on your computer.

Step 2: Load Data

Click on "Get Data" and choose the appropriate data source (e.g., Excel, CSV, database).



Load both the Customers and Sales tables into Power BI.



Step 3: Transform and Shape Data

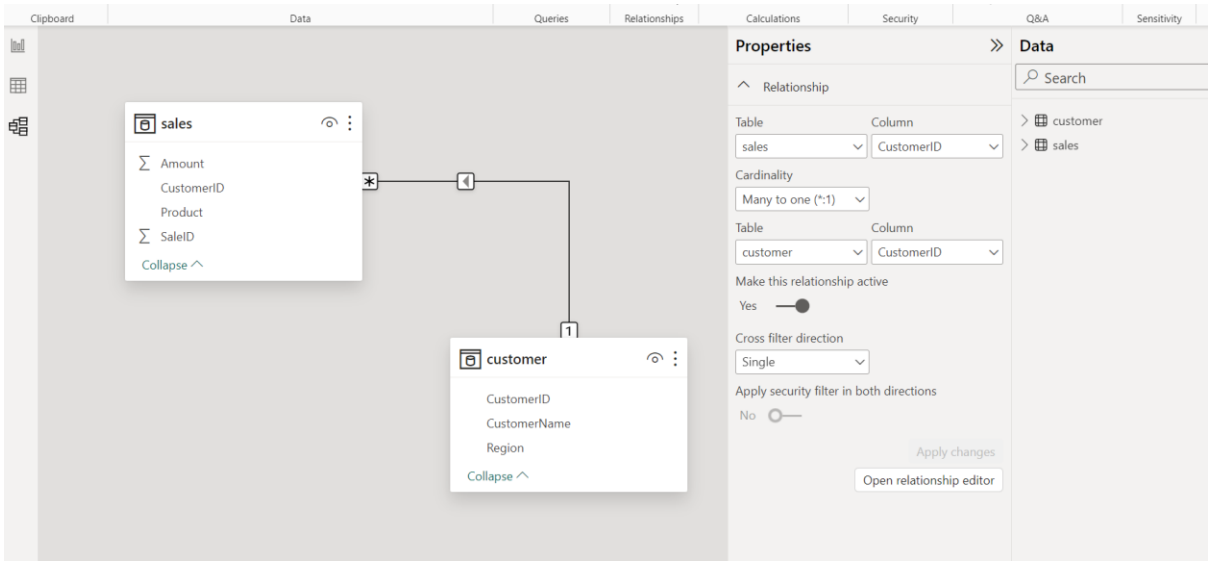
Use Power Query Editor to transform and shape the data.

Rename columns, filter rows, or perform any necessary transformations.

Step 4: Create Relationships

Go to the "Relationships" view by clicking on the "Model" icon in the left sidebar.

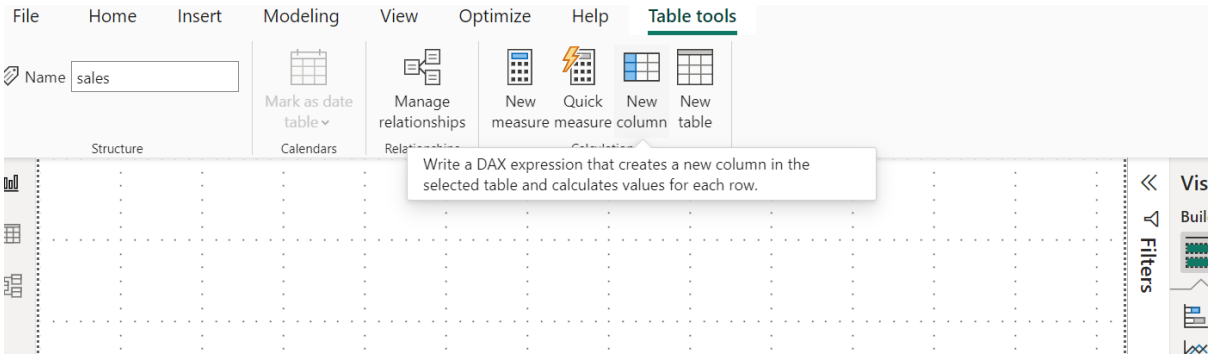
Drag the CustomerID field from the Customers table and drop it onto the CustomerID field in the Sales table to create a relationship.



Step 5: Define Calculated Columns

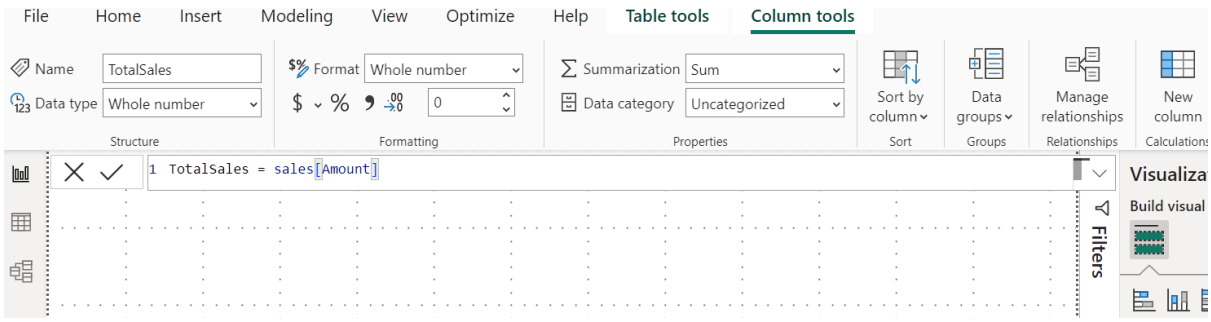
Go back to the "Data" view.

Select New Column



Create a calculated column in the Sales table to calculate total sales for each row:

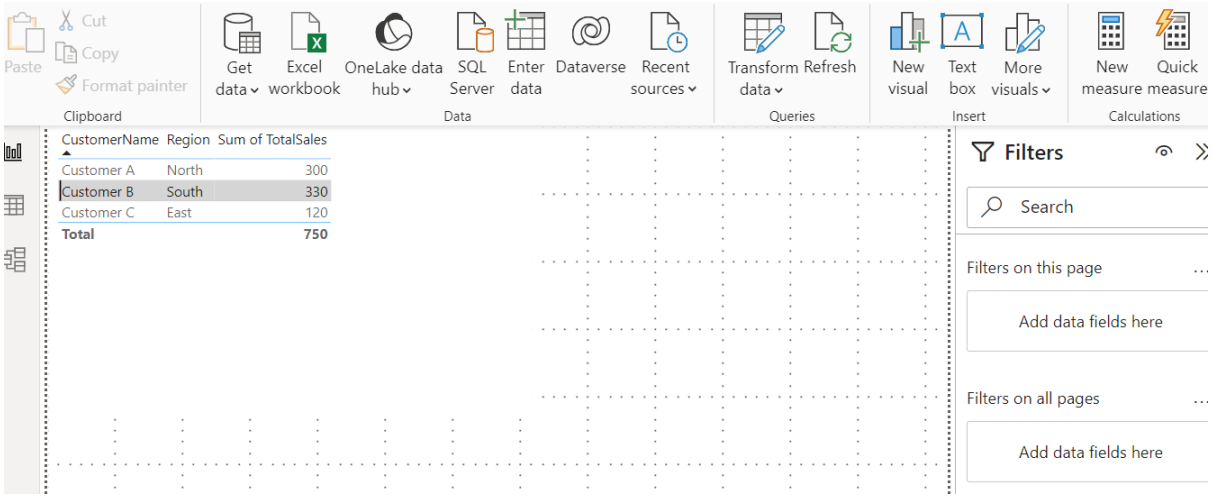
TotalSales = Sales[Amount]



Step 6: Create Visualizations

Go to the "Report" view.

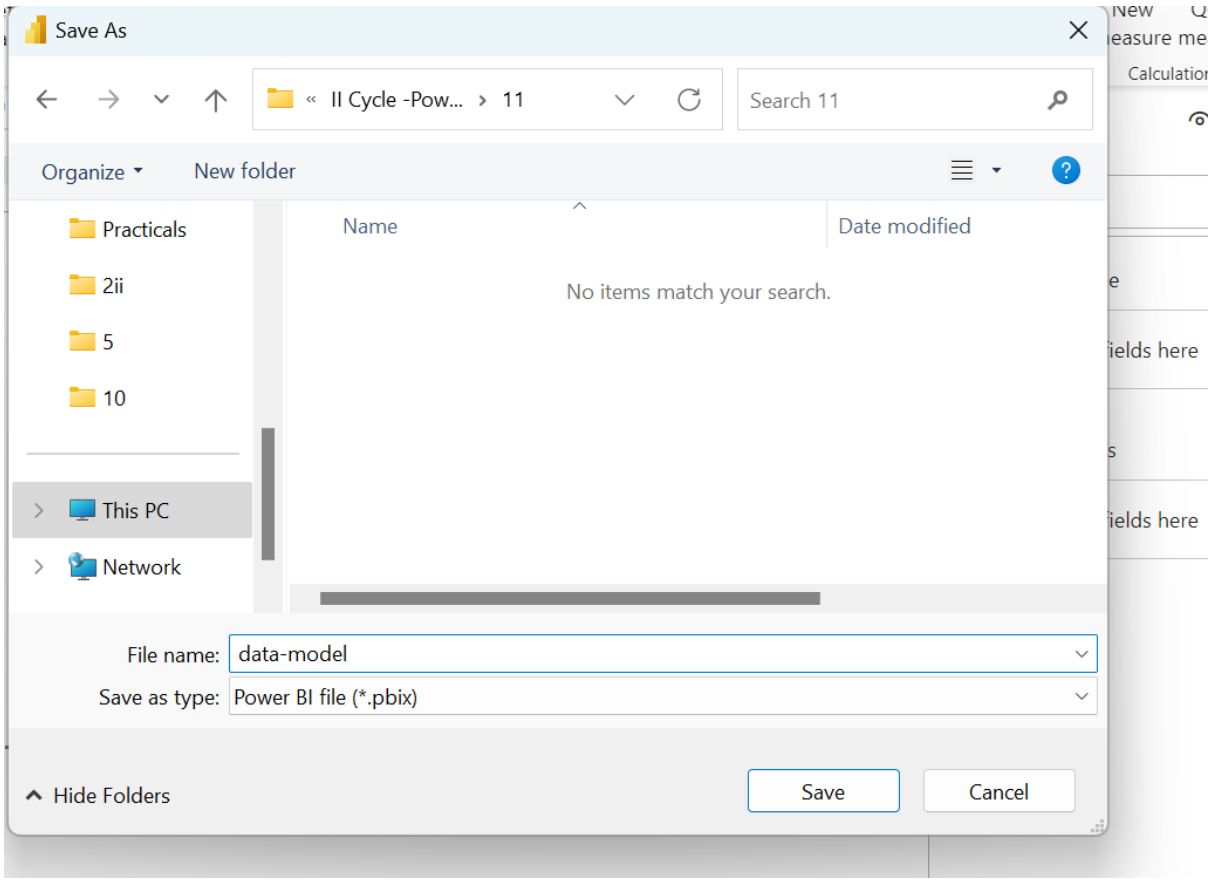
Drag and drop visualizations onto the canvas. For example, create a table showing CustomerName, Region, and TotalSales.



Step 7: Save and Publish

Save your Power BI file.

Publish the file to the Power BI Service if you want to share it online.



Result:

You have now developed a basic data model in Power BI with two related tables and a calculated column. The model allows you to create powerful visualizations and gain insights into your data. As you become more familiar with Power BI, you can explore advanced features like DAX measures, time intelligence functions, and more complex data models.

Perform DAX calculations

Ex.No:10

Date:

Aim:

To perform DAX(Data Analysis Expressions) calculations in PowerBI

Procedure:

Performing Data Analysis Expressions (DAX) calculations in Microsoft Power BI involves creating custom measures and calculated columns to derive insights from your data. Below, I'll outline a simple example:

Example Scenario:

Suppose you have a dataset in Power BI with a Sales table containing information about products, quantities sold, and prices.

Sales Table:

	A	B	C	D
1	**Product**	**Quantity Sold**	**Price per Unit**	
2	A	100	5	
3	B	150	8	
4	C	200	10	
5	D	120	6	
6	E	80	12	
7				
8				
9				

Steps to Create a DAX Measure:

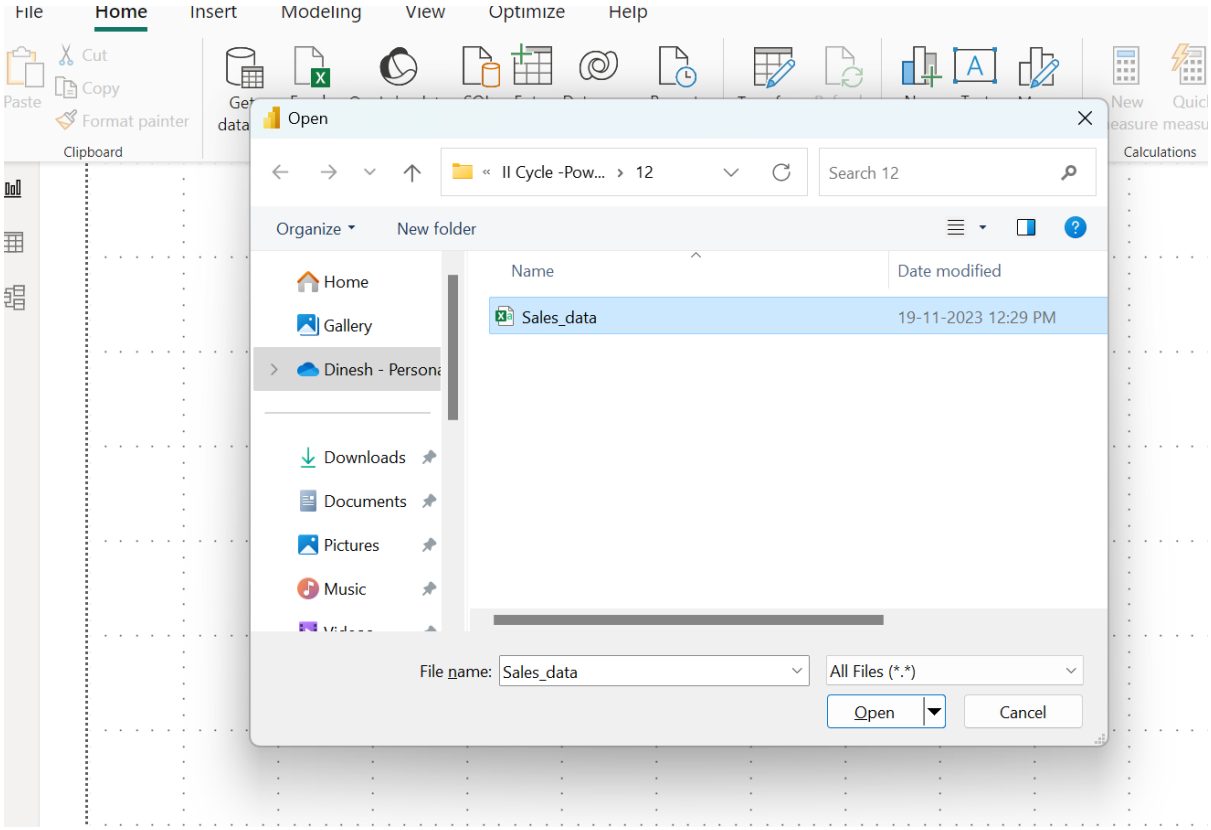
Step 1: Open Power BI Desktop

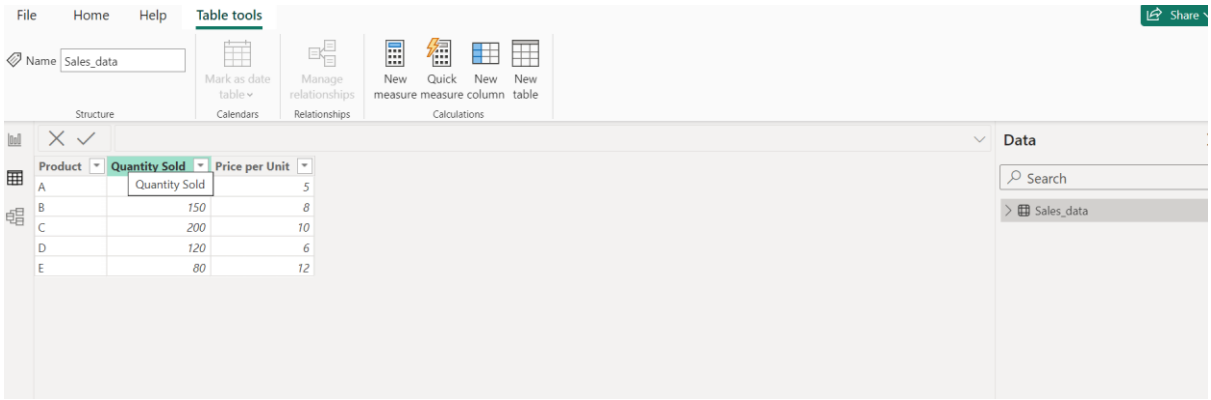
Open Power BI Desktop on your computer.

Step 2: Load Data

Click on "Get Data" and choose the appropriate data source (e.g., Excel, SQL Server, etc.).

Load the Sales table into Power BI.





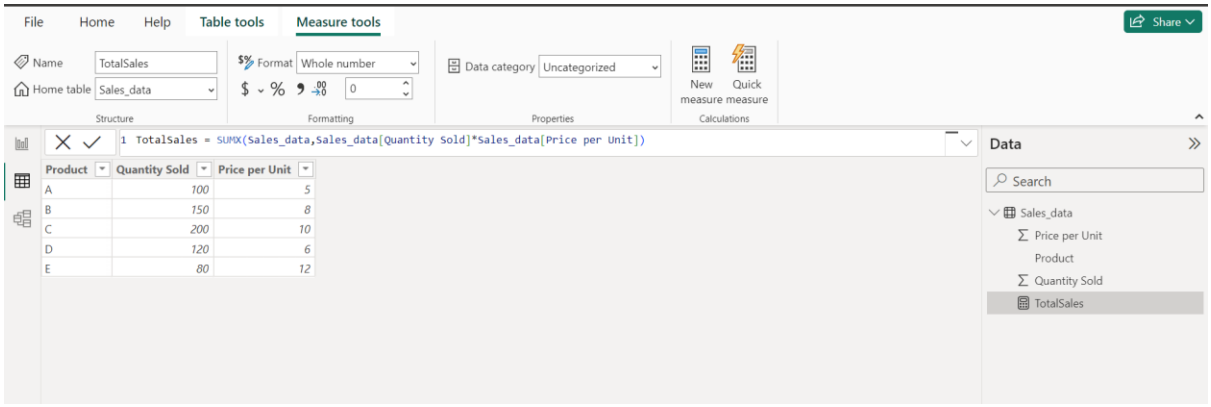
Step 3: Create a DAX Measure

In the "Fields" pane, right-click on the Sales table and choose "New Measure."

Enter a DAX formula for a new measure, for example, to calculate Total Sales:

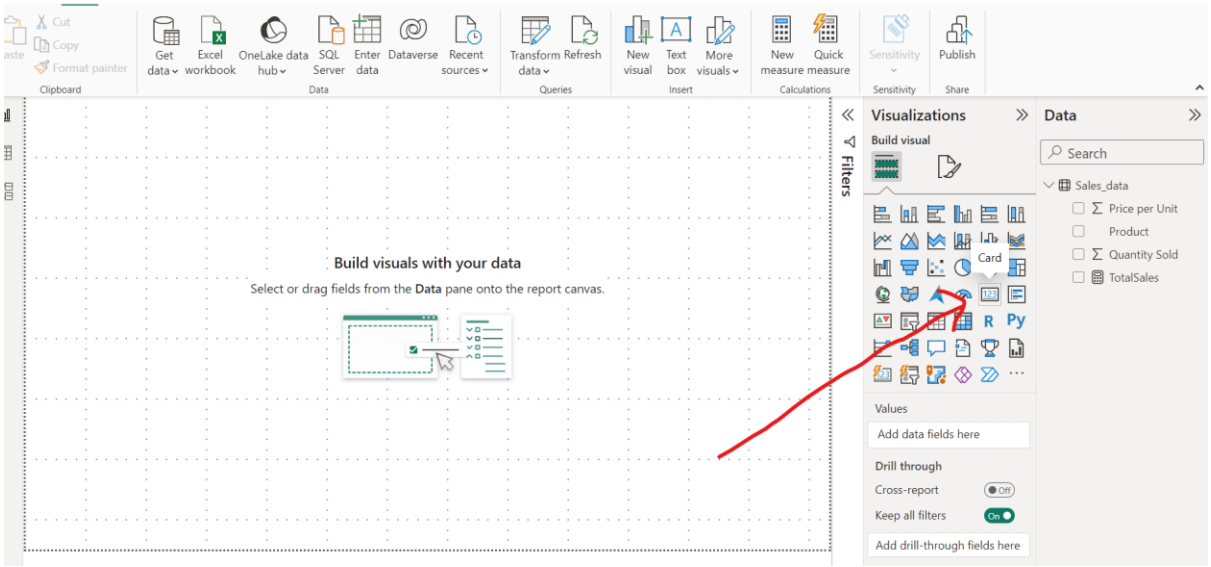
TotalSales = SUMX(Sales, Sales[Quantity Sold] * Sales[Price per Unit])

This formula uses the SUMX function to calculate the sum of the product of "Quantity Sold" and "Price per Unit" for each row in the Sales table.



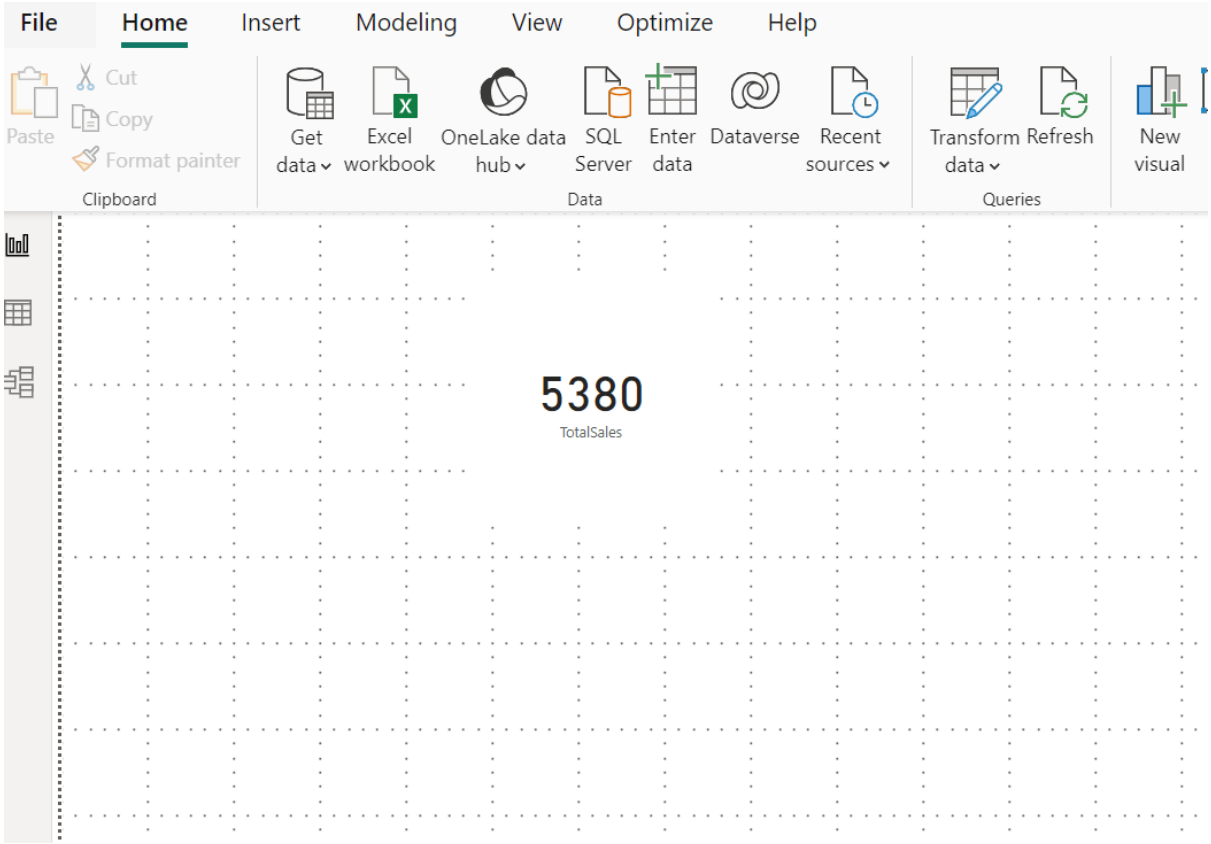
Step 4: Use the Measure in Visualizations

Go to the "Report" view.



Drag and drop a visualization onto the canvas (e.g., a card or a table).

Add the "TotalSales" measure to the Values or Fields area of the visualization.



Result:

You've created a DAX measure, "TotalSales," which calculates the total sales based on the quantity sold and price per unit. This measure can now be used in various visualizations throughout your Power BI report.