



**KIT- KALAINAR KARUNANIDHI INSTITUTE OF TECHNOLOGY**  
(AN AUTONOMOUS INSTITUTION)  
(Accredited by NAAC & NBA with 'A' Grade)  
Kannampalayam Post, Coimbatore -641 402



## **Department of Artificial Intelligence and Data Science**

### **B19ADP701 – BUSINESS ANALYTICS LABORATORY**

Name .....

Batch ..... Reg. No. ....

Branch ..... Year .....



## KIT-KALAIGNARKARUNANIDHI INSTITUTE OF TECHNOLOGY

(An Autonomous Institution)

(Approved by AICTE & Affiliated to Anna University, Chennai)

Coimbatore – 641 402

### **Department of Artificial Intelligence and Data Science**

#### **BONAFIDE CERTIFICATE**

**Student Name:.....**

**Register No:..... Roll No: .....**

**Branch...B.Tech – Artificial Intelligence and Data Science**

**Certified that this is bonafide record work done by Mr/Ms .....**

**of IV-Year Artificial Intelligence and Data Science during the academic year 2024-2025.**

---

**FACULTY IN-CHARGE**

---

**HOD**

**Submitted for the University Practical Examination held on \_\_\_\_\_**

**INTERNAL EXAMINER**

**EXTERNAL EXAMINER**

## **INSTRUCTIONS FOR LABORATORYCLASSES**

1. Enter the lab with record workbook & necessary things.
2. Enter the lab without bags and footwear.
3. Footwear should be kept in the outside shoe rack neatly.
4. Maintain silence during the Lab Hours.
5. Read and follow the work instructions inside the laboratory.
6. Handle the computer systems with care.
7. Shutdown the Computer properly and arrange chairs in order before leaving the lab.
8. The program should be written on the left side pages of the record work book.
9. The record workbook should be completed in all aspects and submitted in the next class itself.
10. Experiment number with date should be written at the top left-hand corner of the record work book page.
11. Strictly follow the uniform dress code for Laboratory classes.
12. Maintain punctuality for lab classes.
13. Avoid eatables inside and maintain the cleanliness of the lab.

## **VISION**

To produce competent professionals to the dynamic needs of the emerging field of Artificial Intelligence and Data Science

## **MISSION**

- To empower students with the knowledge and skills necessary to create intelligent systems and innovative solutions that address societal issues.
- Providing technical knowledge on par with Industry to the students through qualified faculty members having knowledge in recent trends and technologies.
- To produce competent engineers who are both professional and life-skills oriented.
- Providing opportunities for students to improve their research skills in order to address a variety of societal concerns through innovative projects.

## **PROGRAMME OUTCOMES (POs)**

**Students graduating from Artificial Intelligence and Data Science should be able to:**

**PO1 Engineering knowledge :** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex Artificial Intelligence and Data Science problems.

**PO2 Problem analysis:** Identify, formulate, research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and Artificial Intelligence and Data Sciences.

**PO3 Design/development to 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 in the field of Artificial Intelligence and Data Science.

**PO4 Conduct investigations of complex problems:** Using research-based knowledge and Artificial Intelligence & Data Science oriented research methodologies including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

**PO5 Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex Artificial Intelligence and Data Science Engineering activities with an understanding of the limitations.

**PO6 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.

**PO7 Environment and sustainability:** Understand the impact of the professional Artificial Intelligence and Data Science Engineering solutions in societal and environmental contexts, and demonstrate the knowledge, and need for the sustainable development.

**PO8 Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

**PO9 Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams and in multidisciplinary settings.

**PO10 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.

**PO11 Project management and finance:** Demonstrate knowledge and understanding of the Artificial Intelligence and Data Science engineering and management principles and apply these to one's own work, as a member and leader in a team and, to manage projects in multidisciplinary environments.

**PO12 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.

### **PROGRAM EDUCATIONAL OBJECTIVES (PEOs)**

**PEO1:** Graduates will have a strong foundation in mathematics, programming, machine learning, artificial intelligence, and data science, as well as advanced skills in these areas to solve technical problems.

**PEO2:** Graduates will have the capability to apply their knowledge and skills to identify and solve the issues in real world Artificial Intelligence and Data Science related applications.

**PEO3:** Graduates will be able to engage in life-long learning by completing advanced software Technologies, certificates, and/or other professional development.

## **PROGRAM SPECIFIC OUTCOME (PSOs)**

Graduates of Artificial Intelligence and Data Science Programmed should be able to:

**PSO1:** Apply fundamental concepts of Artificial Intelligence and Data Science according to the environmental needs.

**PSO2:** Ability to develop skills to address and solve Artificial Intelligence based social and environmental problem using Data Science to deal multidisciplinary projects using modern tools.

### **COURSEOUTCOMES**

At the end of this course the student will be able to:

CO	Course Outcomes	Knowledge Level
CO1	Illustrate business analytics and apply for simple problems	K2
CO2	Apply data pre-processing operations	K3
CO3	Build statistical operations on data and perform analysis	K3
CO4	Classify data using generative models.	K4
CO5	Evaluate real world applications with business analytics techniques.	K5

### **CO-PO Mapping Table:**

CO/PO & PSO		PO1 (K3)	PO2 (K4)	PO3 (K5)	PO4 (K5)	PO5 (K6)	PO6 (K3) (A3)	PO7 (K2) (A3)	PO8 (K3) (A3)	PO9 (A3)	PO10 (A3)	PO11 (K3) (A3)	PO12 (A3)	PSO1 (K3 A3)	PSO2 (K3 A3)
CO1	K2	3	3	2	1	1	-	-	-	-	-	-	-	2	2
CO2	K3	2	2	2	2	-	-	-	-	-	-	-	-	2	2
CO3	K3	3	3	2	2	1	-	-	-	-	-	-	-	3	3
CO4	K4	2	2	2	2	2	-	-	-	-	-	-	-	3	3
CO5	K5	2	2	2	2	2	-	-	-	-	-	-	-	2	2
Weighted average		2	2	2	2	2	-	-	-	-	-	-	-	2	2

## **SYLLABUS**

### **LIST OF EXPERIMENTS:**

	<b>Experiments using MS Excel</b>
1.	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.
2.	Perform statistical operations - Mean, Median, Mode and Standard deviation, Variance, Skewness, Kurtosis
3.	Perform Z-test, T-test & ANOVA on data.
4.	Perform data pre-processing operations i) Handling Missing data ii) Normalization
5.	Apply and explore various plotting functions on the data set.
6.	Apply and explore various plotting functions on the data set using R-programming.
	<b>Experiments using Power BI Desktop</b>
7.	Prepare & Load data.
8.	Perform statistical operations - Mean, Median, Mode and Standard deviation, Variance, Develop the data model using power BI.
9.	Perform data pre-processing operations using Streamline data analysis.
10.	Design a report .
11.	Create a dashboard and perform data analysis using the tool Performance Analyzer
12.	Perform DAX calculation using the tool DAX studio.

**Total hours: 45**

## CONTENT

S. No	Title of the Experiment	Page No
	<b>Experiments using MS Excel</b>	
1.	(A) Get the input from user and perform numerical operations (Max, Min, Avg, Sum, Sqrt, Round). (B) Perform data import/export operations for different file formats.	
2.	Perform statistical operations - Mean, Median, Mode and Standard deviation, Variance, Skewness, Kurtosis.	
3.	(A) Perform Z-test, (B) Perform T-test (C) Perform ANOVA on data.	
4.	Perform data pre-processing operations (A) Handling Missing data Perform data pre-processing operations (B) Normalization	
5.	Apply and explore various plotting functions on the data set.	
6.	Apply and explore various plotting functions on the data set using R-programming.	
	<b>Experiments using Power BI Desktop</b>	
7.	Prepare & Load data.	
8.	Perform statistical operations - Mean, Median, Mode and Standard deviation, Variance, Develop the data model using power BI.	
9.	Perform data pre-processing operations using Streamline data analysis.	
10.	Design a report .	
11.	Create a dashboard and perform data analysis using the tool Performance Analyzer	
12.	Perform DAX calculation using the tool DAX studio.	

### Content beyond syllabus

S. No	Title of the Experiment	Page No
1.	Presentation of a case study- Campus recruitment analysis	

S.No	Experiment	Prerequisites	Learning Objectives
1	For All Experiments	Analytics formula	<ul style="list-style-type: none"> <li>1. To Illustrate business analytics and apply for simple problems</li> <li>2. To understand data pre-processing operations</li> <li>3. To apply statistical operations on data and perform analysis.</li> <li>4. To analyze data using generative models.</li> <li>5. To determine the real-world applications with business analytics techniques.</li> </ul>

Practical Record Book Index Page

									Sl. No.
									Date
									Name of the Experiment
									Page Number
									Aim & Algorithm (20 Marks)
									Program (30 Marks)
									Output & Inference (15 Marks)
									Viva-Voce (10Marks )
									Total (75Marks)
									Signature of the Faculty Member

### **Signature of the Faculty Member**

## Record Note book Marks (75):

Sl. No.	Date	Name of the Experiment							

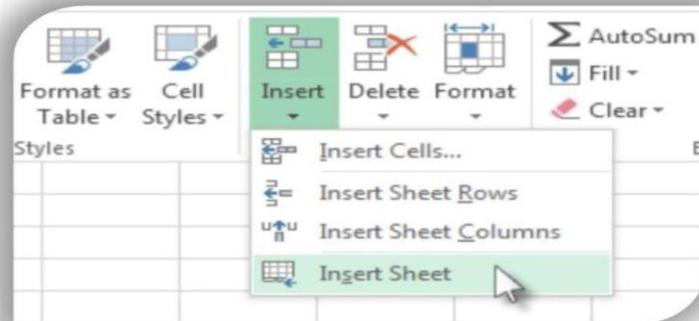
## **EXPLORE THE FEATURES OF MS-EXCEL**

Most professionals use Microsoft Excel to store their data and represent it understandably. It is similar to Google Sheets. To know all the features of Excel which are helpful and time-saving. Here are the top 10 unique features of Microsoft Excel. To analyze the data quickly. Microsoft Excel is used in any type of device like Windows, macOS, Android, and iOS.

### **PROCEDURE:**

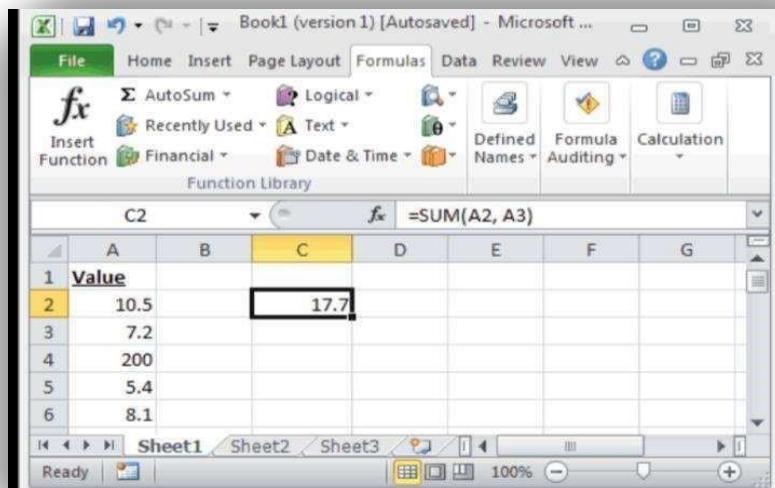
#### **1. Inserting a Spreadsheet**

By default, you will have 1-3 sheets at the beginning. You may need multiple spreadsheets for your data. Inserting and deleting new spreadsheets is quite simple in MS Excel. Click on the “+” button to insert a new sheet. You can also use the shortcut Shift+F11 to insert a new spreadsheet.



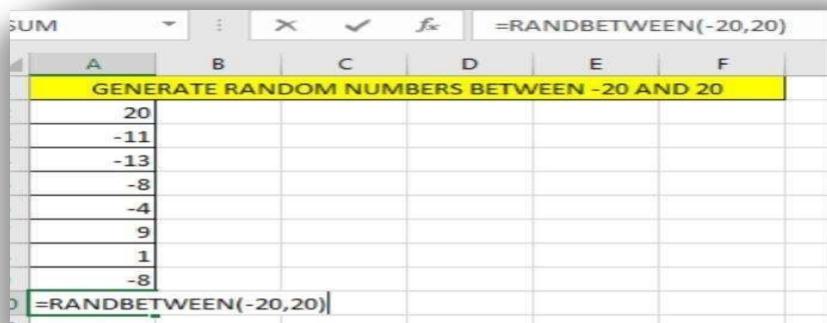
#### **2. Sum Of Numbers :**

To enter a list of numbers in a column and find the sum. Don't search for a Calculator or pen and paper to perform addition. MS Excel has a cool feature to find the sum easily. Select the cell of which you want to find the sum, and press the shortcut Alt+=. Tadaa! The sum is displayed automatically.



### 3. Inserting Random Numbers:

You can find plenty of unique features and functions in Excel. One of them is “RANDBETWEEN.” You can use this function to insert random numbers into the sheet. It takes two arguments. The first one is the least number you are going to insert into your sheet, and the second one is the largest number. With this feature, you don’t have to waste time guessing the numbers. You can finish this within a fraction of a second.



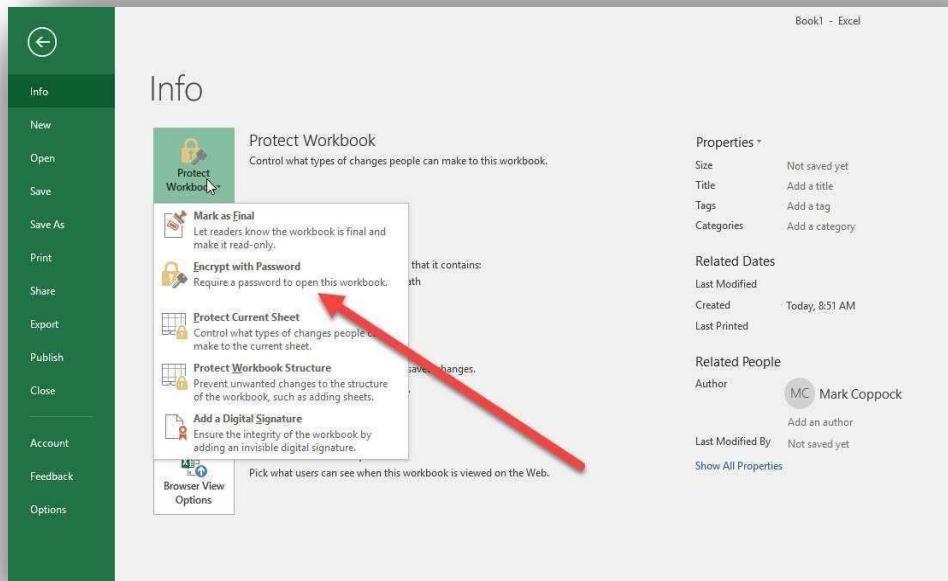
### 4. Shortcut Keys

Excel has made it easy to analyze data. Now it is easier to access the data when you know the shortcut keys. You can perform certain operations without touching the mouse in a fraction of a second. Some of them are



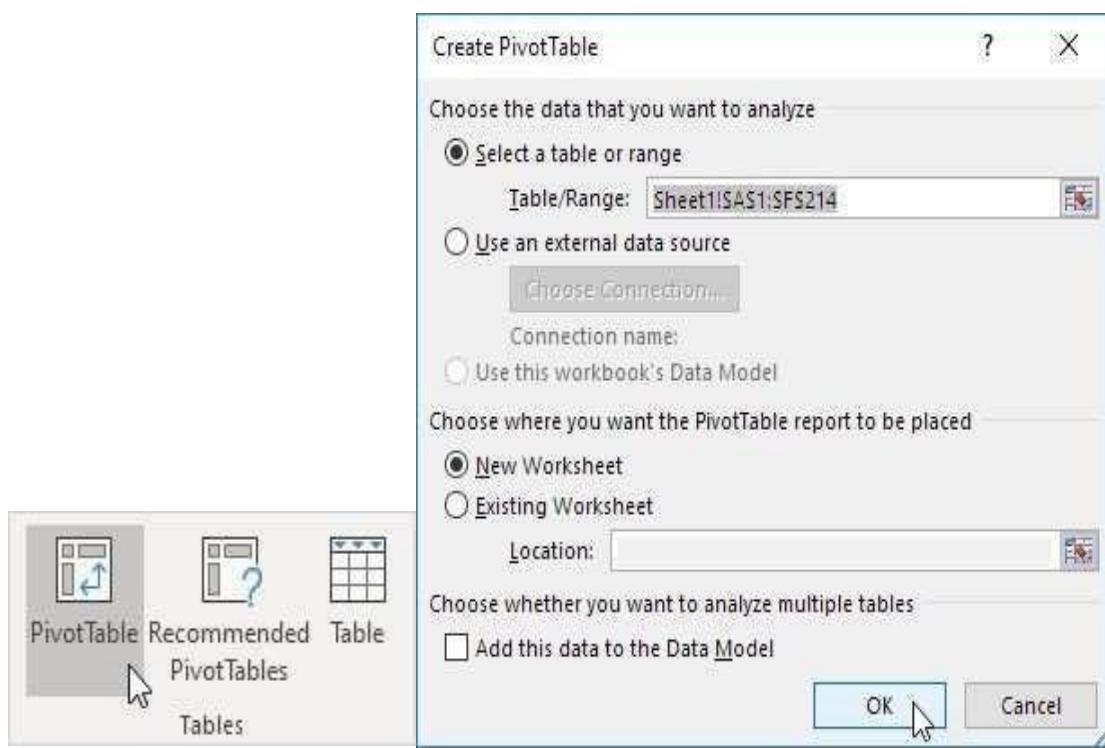
## 5. Password Protection

You may have some private data that you want to keep secured. If you are worried about unauthorized access, you can use password protection to ensure your safety. After enabling this feature, no one can open your document without the password.



## 6. Pivot Tables

Analysing the data in Excel is quite interesting. If you want to create a table that is clear to understand, you can use the pivot table feature.

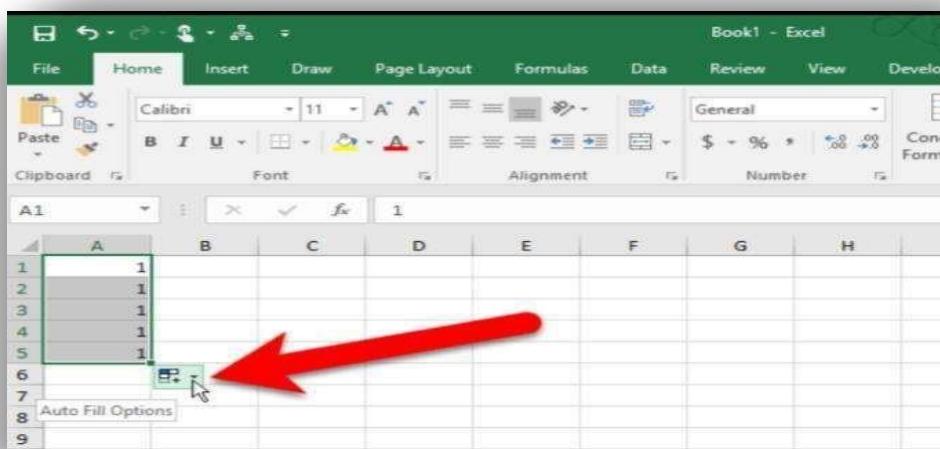


## 7. Auto-Fill

The screenshot shows the 'PivotTable Fields' ribbon. The 'Choose fields to add to report:' section contains checkboxes for 'Order ID', 'Product' (which is checked), 'Category', 'Amount' (which is checked), 'Date', and 'Country' (which is checked). The 'Drag fields between areas below:' section includes 'Filters' (with 'Country' selected) and 'Columns' (empty). It also includes 'Rows' (with 'Product' selected) and 'Values' (with 'Sum of Amount' selected). A 'Defer Layout Update' checkbox is at the bottom left, and an 'Update' button is at the bottom right. To the right of the ribbon is a PivotTable view showing the following data:

	A	B	C
1	Country	(All)	
2			
3	Row Labels	Sum of Amount	
4	Apple	191257	
5	Banana	340295	
6	Beans	57281	
7	Broccoli	142439	
8	Carrots	136945	
9	Mango	57079	
10	Orange	104438	
11	Grand Total	1029734	
12			

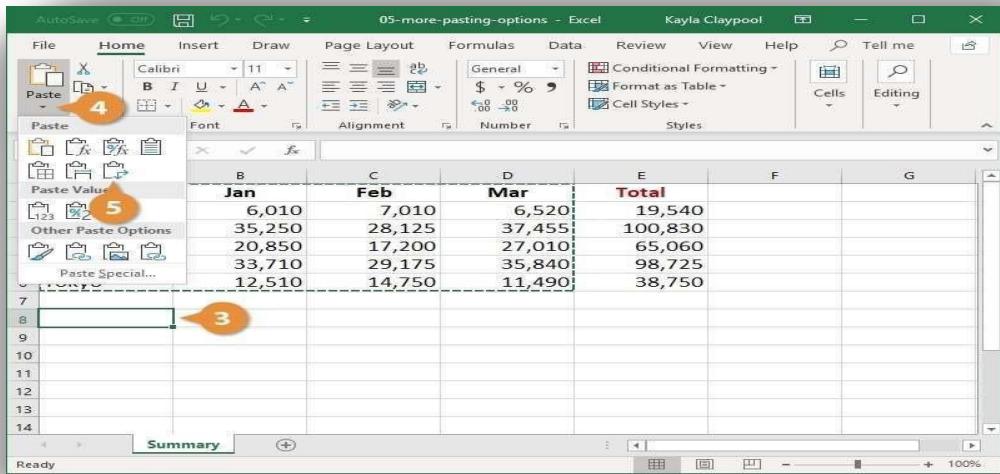
Auto-fill is a simple but useful feature in Excel. With this feature, you can fill the data in series. You can save a lot of time with this feature.



## 8. Paste Special:

When we started to use the copy and paste option, we didn't have to type everything. But sometimes, you may copy formulae, values, and comments. Sometimes you may not want to copy all of them. In such situations, you can use Paste Special feature. It is also a time-saving feature.

A screenshot of the Microsoft Excel interface. The ribbon is visible with the Home tab selected. The Home tab's ribbon group 'Clipboard' is highlighted with an orange circle containing the number '2'. Below the ribbon, a table is displayed with columns A through G. Row 1 contains headers: 'Excursion', 'Jan', 'Feb', 'Mar', 'Total'. Rows 2 through 6 contain data: Beijing (6,010, 7,010, 6,520, 19,540), Las Vegas (35,250, 28,125, 37,455, 100,830), Mexico DF (20,850, 17,200, 27,010, 65,060), Paris (33,710, 29,175, 35,840, 98,725), and Tokyo (12,510, 14,750, 11,490, 38,750). Cell E5 has a red circle with the number '1' around it. The status bar at the bottom says 'Excuse'.



## 9. Index-Match:

Index-Match future comes in handy when you have to handle a massive database with vast data. If you want to find a particular data, you may feel hard. It will seem, like searching for a drop of blood in an ocean. But Index-Match future will find and get you the desired information instantly. But remember the index should be unique like passport number or roll number.

	A	B	C	D	E	F	G	H	I
1									
2	ID	First Name	Last Name	Salary	ID			53	
3	72	Emily	Smith	\$64,901		Salary		\$58,339	
4	66	James	Anderson	\$70,855					
5	14	Mia	Clark	\$188,657					
6	30	John	Lewis	\$97,566					
7	53	Jessica	Walker	\$58,339					
8	56	Mark	Reed	\$125,180					
9	79	Richard	Lopez	\$91,632					
10									

## 10. Rand Function

We can enter random values with the RANDBETWEEN function. But what to do if we want to enter random fraction values? Then, we can use the RAND function.

Using this function, we can enter values between 1 and 0. It does not have any parameters. Just put the parentheses. Note that RANDBETWEEN and RAND functions will vary when we change the sheet.

The screenshot shows a Microsoft Excel window with the formula bar displaying '=RAND()'. The main area contains the title 'RAND' and the function '=RAND()' below it. A descriptive text states: 'The Excel RAND function returns a random number between 0 and 1.' Below this is a table titled 'Values' containing six rows of random numbers. The table has a header row and five data rows. The data rows are as follows:

Values					
0.7887	0.3306	0.5909	0.4733	0.6709	0.9096
0.2997	0.6665	0.7617	0.1732	0.8354	0.1623
0.3653	0.9002	0.1918	0.1201	0.1638	0.2161
0.6797	0.9027	0.8797	0.1848	0.6195	0.3002
0.9157	0.1862	0.8154	0.7980	0.8614	0.8296

The Excel ribbon is visible at the top, and the status bar at the bottom shows 'Sheet1'.

<b>EXP NO: 1(A)</b>
<b>DATE:</b>

## **PERFORM NUMERICAL OPERATIONS (MAX, MIN, AVG, SUM, SQRT, ROUND).**

### **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 Minimum 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

<b>Numerical Operations (MAX, MIN, AVG, SUM, SQRT, ROUND)</b>								
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 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	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 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	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:**

## **VIVA QUESTIONS**

1. List out few useful functions in Excel
2. Difference between COUNT, COUNTA in MS-Excel.
3. what are the special functions in MS-Excel.

## **RESULT:**

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

<b>EXP NO: 1(B)</b>
<b>DATE:</b>

## **DATA IMPORT/EXPORT OPERATIONS FOR DIFFERENT FILE FORMATS.**

### **AIM:**

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

### **THEORY:**

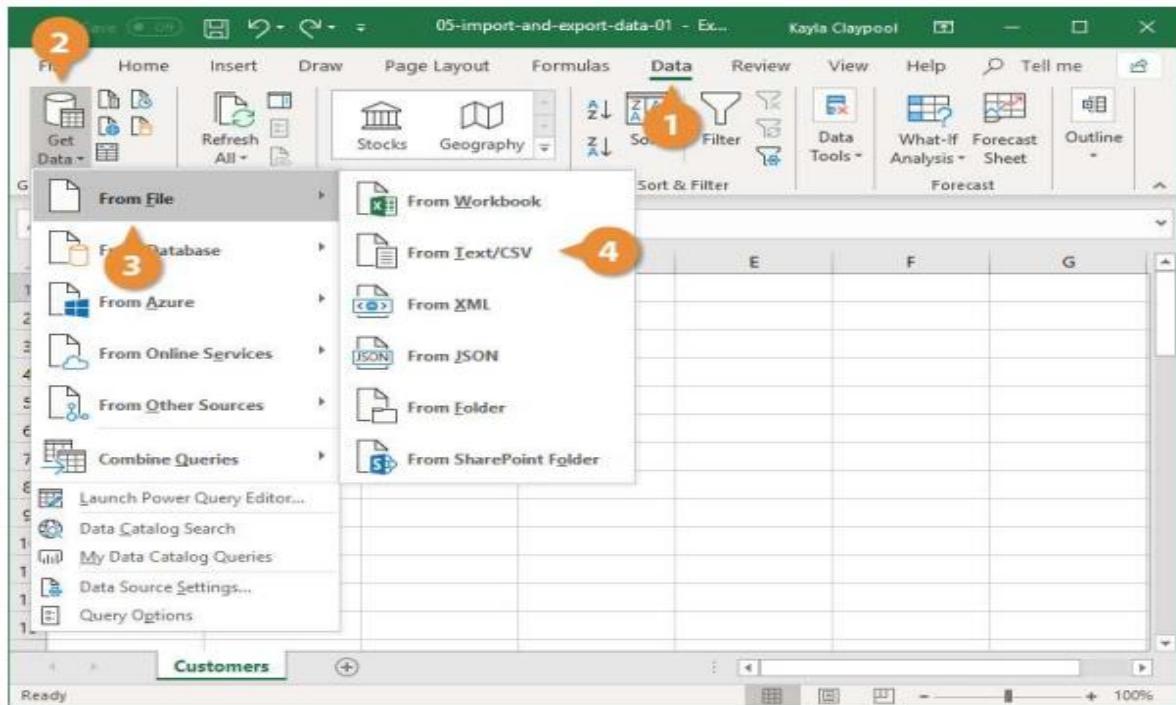
- **Data Importing:** This involves transferring external data into your system.
- **Manual Data Entry:** This basic method is time-consuming but vital when dealing with small datasets.
- **Automated Data Extraction Tools:** These tools can import large quantities of data quickly and accurately.
- **Data 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**

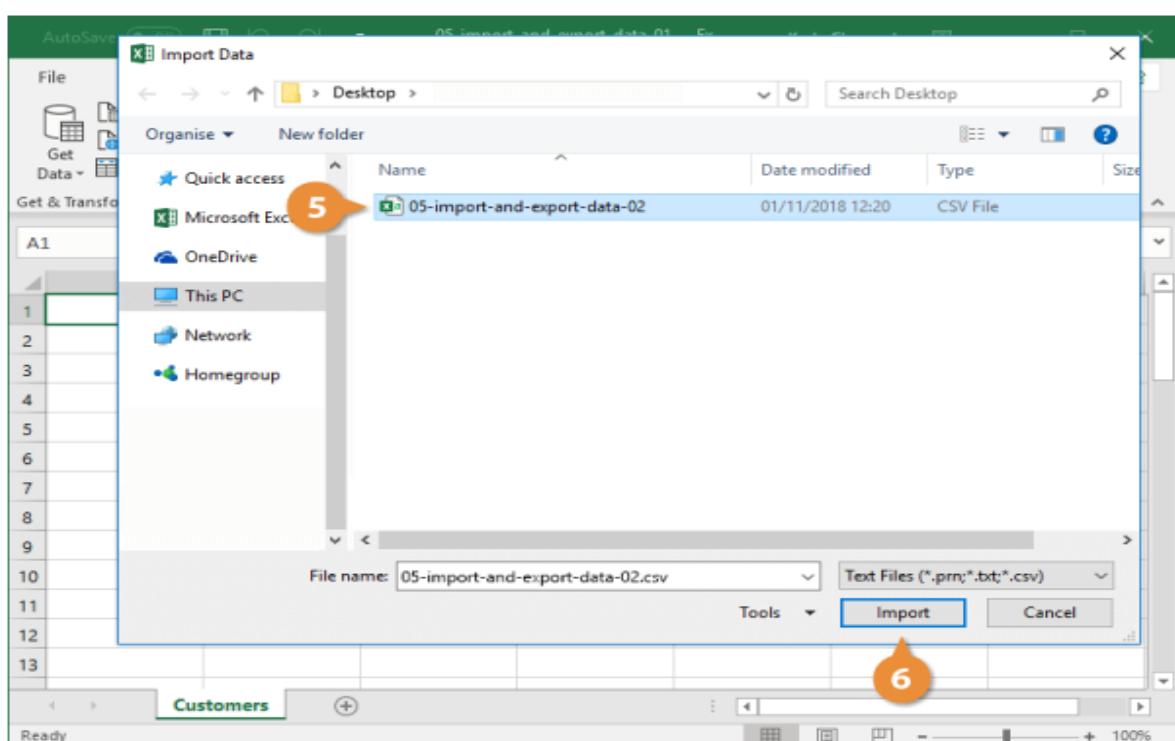
Excel can import data from external data sources including other files, databases, or web pages.

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:**

The screenshot shows the 'Import Data' dialog box in Excel. At the top, there are dropdown menus for 'File Origin' (set to 1252: Western European (Windows)), 'Delimiter' (set to Comma), and 'Data Type Detection' (set to Based on first 200 rows). Below these, a preview grid displays data from the 'Customers' sheet. The columns are labeled 'First', 'Last', 'Company', 'City', 'Packages', and 'Sales'. The data includes entries like Joel Nelson from Nincom Soup in Minneapolis with 6 packages and sales of 6602. At the bottom right of the dialog, there are 'Load', 'Edit', and 'Cancel' buttons. An orange circle with the number '8' is overlaid on the 'Load' button.

## **2.EXPORT FILE**

1. Click the File tab.

The screenshot shows the Excel ribbon with the 'File' tab selected. The ribbon tabs from left to right are: File, Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, Help, and Tell me. An orange circle with the number '1' is overlaid on the 'File' tab icon. The rest of the ribbon is visible, including the font, alignment, and number formats sections.

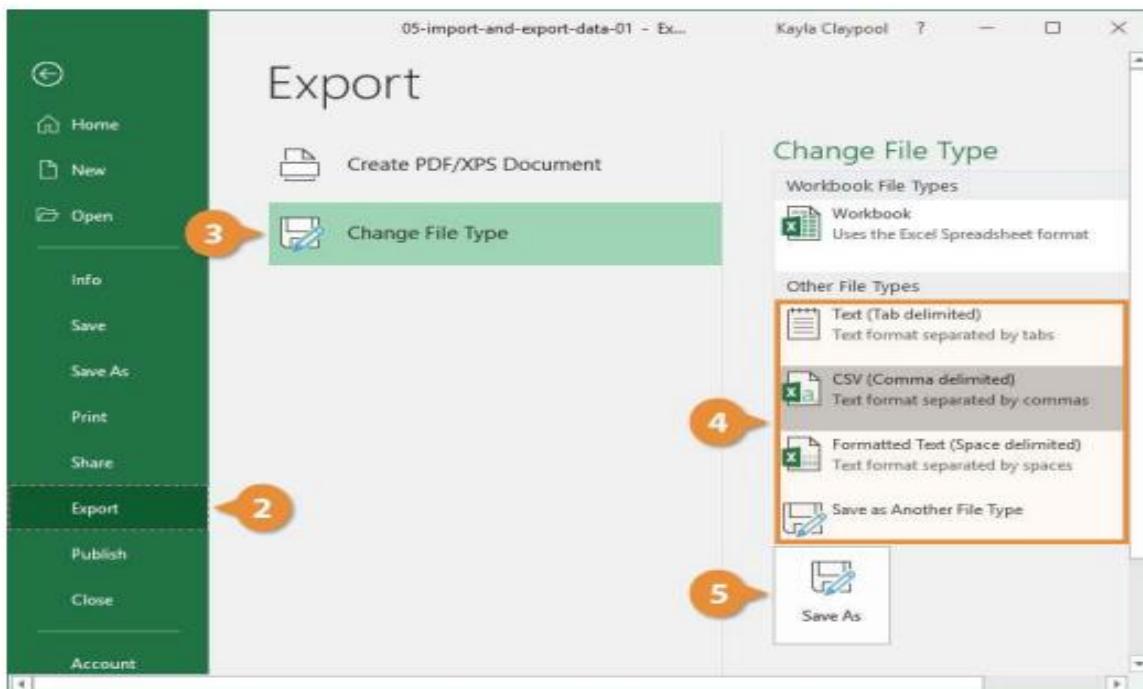
2. At the left, click Export.

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.

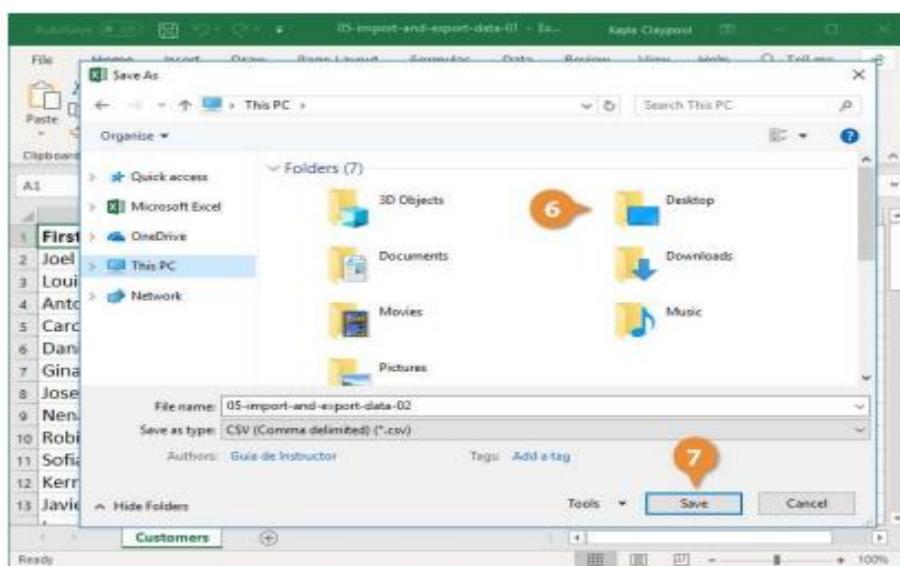


6. Specify where you want to save the file.

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:**

**VIVA QUESTIONS:**

1. Explain a situation you had to import unstructured data. What was your process?
2. Have you ever used batch processing for data import/ what are the pros and cons?
3. Can you explain the role of data mapping in the import process?

**RESULT:**

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

<b>EXP NO: 2</b>	<b>PERFORM STATISTICAL OPERATIONS</b>
<b>DATE:</b>	<b>[Mean, Median, Mode and Standard Deviation, Variance, Skewness, Kurtosis]</b>

### **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=4 left 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(left parenthesis, 1, 444, 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 ' $\sigma^2$ '.

### **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, ' $\sigma$ '.

### **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:**

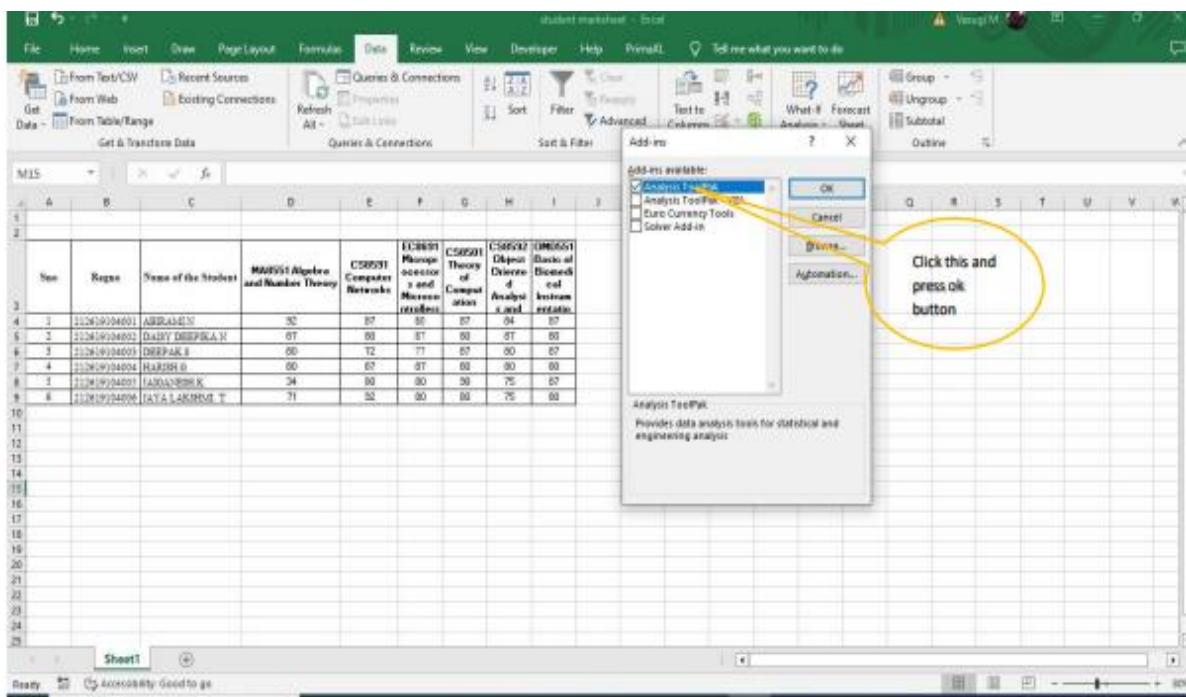
Skewness measures the degree of asymmetry of the distribution, while Kurtosis measures the degree of peakedness and flatness of a distribution. Skewness is the third measure of moments, while kurtosis is the fourth measure of moments. The value of both Skewness and Kurtosis ranges from -infinity to +infinity.

### **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]

The screenshot shows a Microsoft Excel window titled "student marksheet - Excel". The ribbon at the top has tabs for File, Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, Developer, Help, and Print. A blue callout bubble points to the "Data" tab, which now includes a "Data Analysis" option under the "Analysis" section. The main worksheet area displays student marks for various subjects like Mathematics, Computer Networks, etc., across multiple rows and columns.

**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.

The screenshot shows the same Excel window with the "Data Analysis" dialog box open. The dialog box lists various statistical analysis tools: Analysis Tools, ANOVA: Single Factor, ANOVA: Two-Factor With Replication, ANOVA: Two-Factor Without Replication, Correlation, Covariance, Descriptive Statistics, Exponential Smoothing, F-Test Two-Sample for Variances, Fourier Analysis, and Histogram. A blue callout bubble points to the "Descriptive Statistics" option, with a note below it saying "Now select this option and press ok".

**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.

student marksheet - Book1

File Home Insert Draw Page Layout Formulas Data Review View Developer Help PrintL Tell me what you want to do

From File... CSV From Web From Table/Range Get & Transform Data Queries & Connections Sort & Filter Data Tools Forecast Outline Analysis

From File... Properties Advanced

From Table/Range

From Web

From File... CSV

From Table/Range

Get & Transform Data

Queries & Connections

Sort & Filter

Data Tools

Forecast

Outline

Analysis

Descriptive Statistics

Input: Input Range: \$D\$4:\$D\$9

Grouped By:  Columns  Rows

Labels in First Row

Output options:

- Output Range:
- New Worksheet By:
- New Workbook
- Summary Statistics
- Confidence Level for Mean: 95 %
- F1% Largest: 1
- F1% Smallest: 1

OK Cancel Help

Now select this data range

Sl.no	Regno	Name of the Student	MAT551 Algebra and Number Theory	C58531 Computer Networks	EC8891 Managerial Economics and Macroeconomics	C58501 Theory of Computation	C58502 Discrete Mathematics and Combinatorics	OM8551 Basic of Biostatistics and Bioinformatics
1	212610104001	ABRAHAM S	95	87	95	87	94	87
2	212610104002	DAILY DEEPIKA N	87	89	87	88	87	88
3	212610104003	DEEPAK R	86	72	77	87	80	87
4	212610104004	HARISH G	80	87	87	88	80	88
5	212610104005	JAYANSHIKH K	94	88	88	90	75	87
6	212610104006	JATA LAKHMI T	71	52	80	88	75	88

Sheet1

Type here to search 11:00 24-07-2023

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

student marksheet - Book1

File Home Insert Draw Page Layout Formulas Data Review View Developer Help PrintL Tell me what you want to do

From File... CSV From Web From Table/Range Get & Transform Data Queries & Connections Sort & Filter Data Tools Forecast Outline Analysis

From File... Properties Advanced

From Table/Range

From Web

From File... CSV

From Table/Range

Get & Transform Data

Queries & Connections

Sort & Filter

Data Tools

Forecast

Outline

Analysis

Descriptive Statistics

Input: Input Range: \$D\$4:\$D\$9

Grouped By:  Columns  Rows

Labels in First Row

Output options:

- Output Range:
- New Worksheet By:
- New Workbook
- Summary Statistics
- Confidence Level for Mean: 95 %
- F1% Largest: 1
- F1% Smallest: 1

OK Cancel Help

Now select the any cell for output range to be displayed

Sl.no	Regno	Name of the Student	MAT551 Algebra and Number Theory	C58531 Computer Networks	EC8891 Managerial Economics and Macroeconomics	C58501 Theory of Computation	C58502 Discrete Mathematics and Combinatorics	OM8551 Basic of Biostatistics and Bioinformatics
1	212610104001	ABRAHAM S	95	87	95	87	94	87
2	212610104002	DAILY DEEPIKA N	87	89	87	88	87	88
3	212610104003	DEEPAK R	86	72	77	87	80	87
4	212610104004	HARISH G	80	87	87	88	80	88
5	212610104005	JAYANSHIKH K	94	88	88	90	75	87
6	212610104006	JATA LAKHMI T	71	52	80	88	75	88

Sheet1

Type here to search 11:11 24-07-2023

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

The screenshot shows an Excel spreadsheet titled "student marksheets - book1". The data consists of student records with columns for Sno, Regno, Name of the Student, and various subjects. An "Analysis" tab is selected in the ribbon, and a "Descriptive Statistics" dialog box is open. The dialog box has the following settings:

- Input Range: \$D\$4:\$D\$9
- Grouped By: Columns
- Output Range: \$C\$12
- Summary Statistics (checkbox checked)
- Confidence Level for Mean: 95%
- Alpha: 1
- Alpha squared: 1

A callout bubble with the text "Now select this option and press ok" points to the "Summary Statistics" checkbox.

### SAMPLE:

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessor 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	88	80	80
5	212619104005	JAIGANESH.K	34	80	80	90	75	87
6	212619104006	JAYA LAKSHMI.T	71	92	80	89	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:****VIVA QUESTIONS:**

1. What are the various statics formulas?
  
  
  
  
  
  
2. What is Statics?
  
  
  
  
  
  
3. What is analytics?

**RESULT:**

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

**EXP NO: 3(A)**

**DATE:**

## **PERFORM T-TEST**

### **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.

$$H_0: \mu_1 - \mu_2 = 0$$

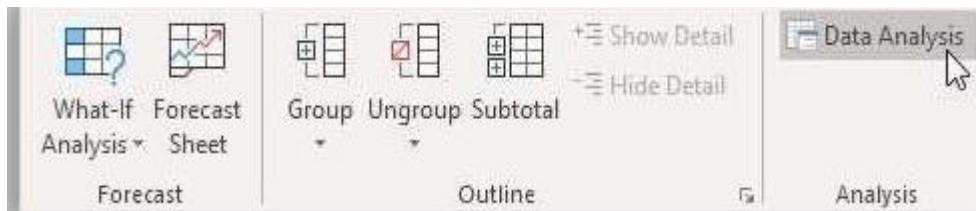
$$H_1: \mu_1 - \mu_2 \neq 0$$

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

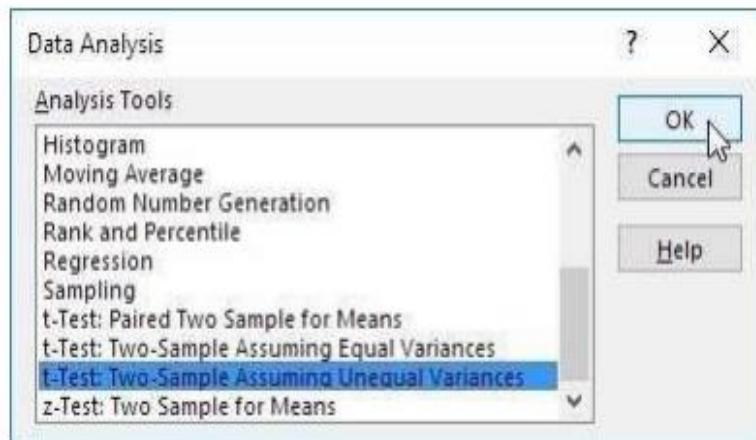
### **PROCEDURE:**

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

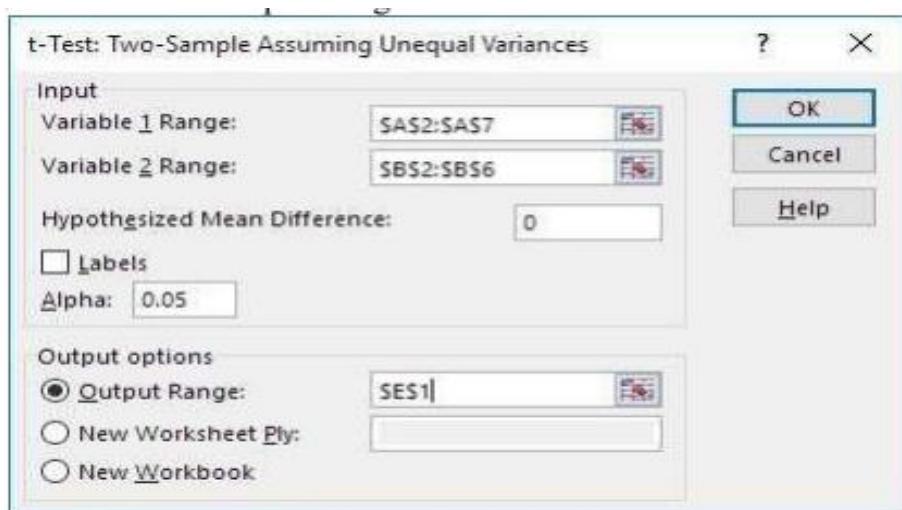
2. On the Data tab, in the Analysis group, click Data Analysis.



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 ( $H_0: \mu_1 - \mu_2 = 0$ ).
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:**

**VIVA QUESTIONS:**

1.What is T-test?

2.What are the types of T-test?

**RESULT:**

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

EXP NO: 3(B)	
DATE:	

## PERFORM Z-TEST

### 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  $\geq 30$ . 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.



=Z.TEST(

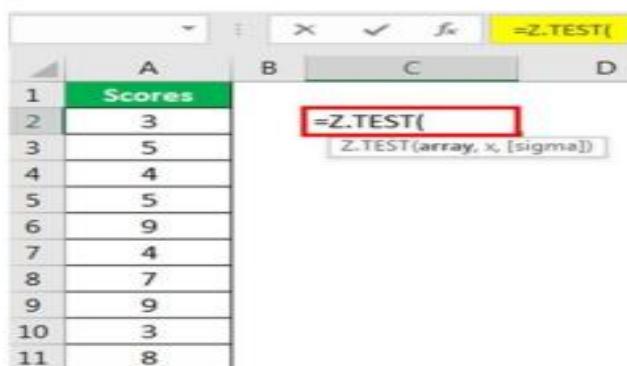
Z.TEST(array, x, [sigma])

### PROCEDURE 1: Using the Z.Test Formula

Step 1: Open data set

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

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



A	B	C	D
1	Scores	=Z.TEST(	
2	3		
3	5		
4	4		
5	5		
6	9		
7	4		
8	7		
9	9		
10	3		
11	8		
12			

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

	A	B	C	D
1	Scores			
2	3	=Z.TEST(A2:A11,6)		
3	5			
4	4			
5	5			
6	9			
7	4			
8	7			
9	9			
10	3			
11	8			
12				

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

	A	B	C	D
1	Scores			
2	3	=Z.TEST(A2:A11,6)		
3	5			
4	4			
5	5			
6	9			
7	4			
8	7			
9	9			
10	3			
11	8			
12				

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.

	A	B	C	D
1	Scores			
2	3	0.6561912		
3	5	1.3123825		
4	4			
5	5			
6	9			
7	4			
8	7			
9	9			
10	3			
11	8			
12				

## 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.

$H_0: \mu_1 - \mu_2 = 0$  (Null Hypothesis)

$H_1: \mu_1 - \mu_2 \neq 0$  (Alternative Hypothesis)

The alternative hypothesis ( $H_1$ ) states that the two population means are not equal.

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

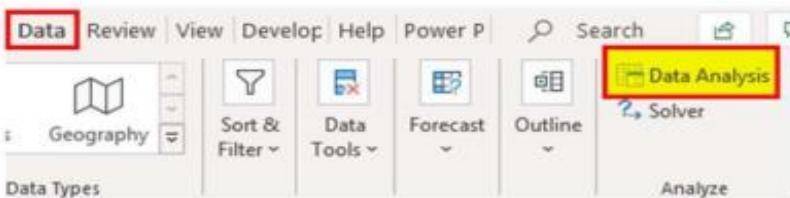
	A	B
1	Student 1	Student 2
2	93	79
3	68	66
4	82	73
5	51	72
6	87	80
7	82	74
8	72	95
9	71	93
10	92	99
11	73	97
12		

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			
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			

**Variance 1** 146.49  
**Variance 2** 131.16

- 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.”

The image shows a Microsoft Excel spreadsheet with two columns: 'Student 1' and 'Student 2'. Below the table is the 'Data Analysis' dialog box. The 'Analysis Tools' list is shown, and 'z-Test: Two Sample for Means' is highlighted with a green box. The 'OK' button is visible in the top right corner of the dialog box.

	A	B
1	<b>Student 1</b>	<b>Student 2</b>
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 3:** For the “Variable 1 Range,” select “Student 1” scores. For the “Variable 2 Range,” select “Student 2” scores.

The image shows the 'z-Test: Two Sample for Means' dialog box. The 'Input' section has 'Variable 1 Range:' set to '\$A\$2:\$A\$11' and 'Variable 2 Range:' set to '\$B\$2:\$B\$11'. Other fields include 'Hypothesized Mean Difference:' (empty), 'Variable 1 Variance (known):' (empty), 'Variable 2 Variance (known):' (empty), 'Labels' (unchecked), and 'Alpha: 0.05'.

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

	A	B	C	D	E
1	<b>Student 1</b>	<b>Student 2</b>			
2	93	79		<b>Variance 1</b>	146.49
3	68	66		<b>Variance 2</b>	131.16
4	82				
5	51				
6	87				
7	82				
8	72				
9	71				
10	92				
11	73				
12					

**z-Test: Two Sample for Means**

**Input**

Variable 1 Range:

Variable 2 Range:

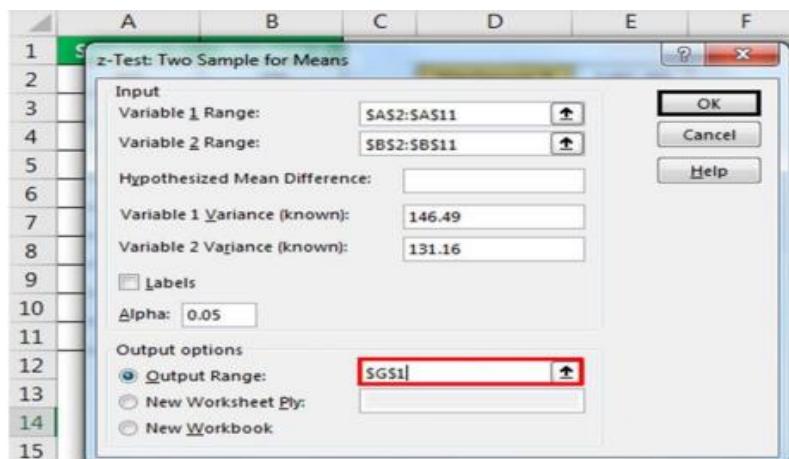
Hypothesized Mean Difference:

Variable 1 Variance (known):  146.49

Variable 2 Variance (known):  131.16

Labels

**Step 5:** Select the “Output Range” as a cell and press “OK.”



**SAMPLE OUTPUT:**

	G	H	I
1	<b>z-Test: Two Sample for Means</b>		
2			
3		<b>Variable 1</b>	<b>Variable 2</b>
4	Mean	77.1	82.8
5	Known Variance	146.49	131.16
6	Observations	10	10
7	Hypothesized Mean Difference	0	
8	z	-1.0817	
9	P(Z<=z) one-tail	0.13968	
10	z Critical one-tail	1.64485	
11	P(Z<=z) two-tail	0.27936	
12	z Critical two-tail	1.95996	
13			

**OUTPUT:**

**VIVA QUESTIONS:**

1. **What is Z-score?**
  
2. **What you mean by population and sampling?**

**RESULT:**

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

**EXP NO: 3(B)**

**DATE:**

## **PERFORM ANOVA ON DATA**

### **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.

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$H_1$ : 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.

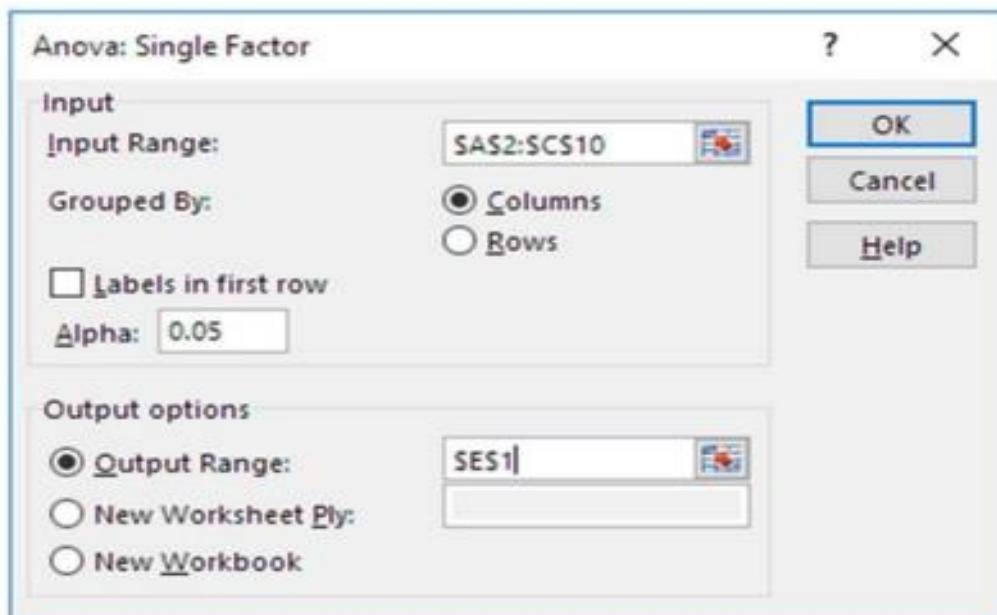


2. Select Anova: Single Factor and click OK.



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:

## VIVA QUESTIONS:

- 1.What is one way Anova?
  - 2.Differentiate single and two factors variable?

## RESULT:

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

<b>EXP NO: 4(A)</b>
<b>DATE:</b>

## **PERFORM DATA PRE-PROCESSING OPERATION HANDLING MISSING DATA**

### **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			
	<b>ID</b>	<b>Registered Person</b>	<b>Present Person</b>
1	1602	Mike	Adam
2	1725	Adam	Antony
3	1235	David	John
4	1823	Jonathon	Hamlet
5	1290	Lilly	Lilly
6	1692	Bob	Collins
7	1286	Collins	
8	1427	John	
9	1735	Hamlet	
10	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.

	A	B	C	D	E	F
1						
2						
<b>Using ISERROR and VLOOKUP Functions</b>						
4	<b>ID</b>	<b>Registered Person</b>	<b>Present Person</b>	<b>Missing Person</b>		
5	1602	Mike	Adam	TRUE		
6	1725	Adam	Antony			
7	1235	David	John			
8	1823	Jonathon	Hamlet			
9	1290	Lilly	Lilly			
10	1692	Bob	Collins			
11	1286	Collins				
12	1427	John				
13	1735	Hamlet				
14	1882	Antony				
15						

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

	A	B	C	D	E
1					
2					
<b>Using ISERROR and VLOOKUP Functions</b>					
4	<b>ID</b>	<b>Registered Person</b>	<b>Present Person</b>	<b>Missing Person</b>	
5	1602	Mike	Adam	TRUE	
6	1725	Adam	Antony	FALSE	
7	1235	David	John	TRUE	
8	1823	Jonathon	Hamlet	TRUE	
9	1290	Lilly	Lilly	FALSE	
10	1692	Bob	Collins	TRUE	
11	1286	Collins		FALSE	
12	1427	John		FALSE	
13	1735	Hamlet		FALSE	
14	1882	Antony		FALSE	
15					

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
1823	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

The screenshot shows a Microsoft Excel spreadsheet titled "student marksheet 3 - Excel". The ribbon at the top has tabs for File, Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, Developer, Help, and PrimaXL. A floating toolbar is visible, with a red arrow pointing from the "PrimaXL" tab in the ribbon to a circled "Choose primaXL" button in the floating toolbar. The main worksheet contains data for six students across nine columns, with column headers including Sno, Regno, Name of the Student, and various subject codes like MA8551, EC8691, CS8501, etc. The rows are numbered from 2 to 8.

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

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”.

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

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.

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

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.

The screenshot shows a Microsoft Excel spreadsheet titled "student marksheets 3". The data table includes columns for Sno, Regno, Name of the Student, MA8551 (Algebra and Number Theory), EC8691 (Microprocessor and Networks), and Microcontrol. Row 8 contains missing values in the EC8691 column. A dialog box titled "Filling of the Missing Data" is overlaid on the spreadsheet. The "Input and Specification" section shows "Data Range: 'student marksheet 3'!\$D\$3:\$G\$8" and "Choice: Random pick from the existing data samples". The "Output" section shows "Output to: 'student marksheet 3'!\$D\$19" and "Output to a new sheet: <input type='checkbox' checked='checked'>" checked. Buttons for "Run", "Exit", and "Reset" are visible.

Sno	Regno	Name of the Student	MA8551	EC8691
1	212619104001	ABIRAMI.N	92	87
2	212619104002	DAISY DEEPIKA.N	87	80
3	212619104003	DEEPAK RAJ.S	80	
4	212619104004	HARISH.G	80	87
5	212619104005	JAIGANESH.K	34	80
6	212619104006	JAYA LAKSHMI.T	71	92
7				80
8				90
9				75
10				87
11			92	87
12			87	80
13			80	85.2

**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
80	87	87	80	80	80
34	80	83.5	90	75	87
71	92	80	84.8	75	80

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

The screenshot shows the same Excel spreadsheet after applying the random pick method. The missing value in row 8, cell EC8691, has been replaced with the value 85.2. The rest of the data remains the same as in the previous step.

Sno	Regno	Name of the Student	MA8551	EC8691
1	212619104001	ABIRAMI.N	92	87
2	212619104002	DAISY DEEPIKA.N	87	80
3	212619104003	DEEPAK RAJ.S	80	85.2
4	212619104004	HARISH.G	80	87
5	212619104005	JAIGANESH.K	34	80
6	212619104006	JAYA LAKSHMI.T	71	92
7				80
8				87
9				80
10			92	87
11			87	80
12			80	85.2
13			80	83.5
14			34	90
15			71	84.8

**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
80	87	87	80	80	80
34	80	83.5	90	75	87
71	92	80	84.8	75	80

**FILLING OF THE MISSING DATA BY RANDOM PICK**

92	87	80	87	84	87
87	80	87	80	87	80
80	92	87	87	80	87
80	87	87	80	80	80
34	80	87	90	75	87
71	92	80	87	75	80

**OUTPUT:**

**VIVA QUESTIONS:**

1.What you mean by data reduction?

2. What is called data reduction.

3.what are the types of data processing?

4.how to clean the data?

5. what is the process of data integration?

**RESULT:**

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

EXP NO: 4(B)	PERFORM DATA PRE-PROCESSING OPERATION
DATE:	NORMALIZATION

## PERFORM DATA PRE-PROCESSING OPERATION NORMALIZATION

### 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

The screenshot shows a Microsoft Excel spreadsheet with data in column A labeled 'Values'. The formula `=MIN(A2:A20)` is entered in cell E1. The status bar at the bottom right displays 'Min Value: =MIN(A2:A20)' and 'Max Value:'. A watermark 'THAT EXCEL SITE' is visible in the center of the sheet.

	A	B	C	D	E
1	Values				
2	19				
3	22				
4	18				
5	15				
6	16				
7	16				
8	4				
9	28				
10	26				
11	31				
12	15				
13	18				
14	16				
15	18				
16	11				
17	18				
18	17				
19	20				
20	22				
21					

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

A screenshot of Microsoft Excel showing a dataset in column A labeled "Values". The range A2:A20 contains the values: 19, 22, 18, 15, 16, 16, 16, 4, 28, 26, 31, 15, 18, 16, 18, 11, 18, 17, 20, 22. The formula =MAX(A2:A20) is entered into cell E2, and the result 31 is displayed. The formula bar also shows =MAX(A2:A20). A green box highlights the formula in the formula bar. A watermark "THAT EXCEL SITE" is visible in the background.

	A	B	C	D	E	F
1	Values					
2	19				=MAX(A2:A20)	
3	22					
4	18					
5	15					
6	16					
7	16					
8	4					
9	28					
10	26					
11	31					
12	15					
13	18					
14	16					
15	18					
16	11					
17	18					
18	17					
19	20					
20	22					

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

A screenshot of Microsoft Excel showing a dataset in column A labeled "Values" and a new column B labeled "Normalized Value". The range A2:A20 contains the values: 19, 22, 18, 15, 16, 16, 16, 4, 28, 26, 31, 15, 18, 16, 18, 11, 18, 17, 20, 22. The formula =(A2-\$E\$1)/(\$E\$2-\$E\$1) is entered into cell B2, and the result 0.5555555556 is displayed. The formula bar also shows =(A2-\$E\$1)/(\$E\$2-\$E\$1). A green box highlights the formula in the formula bar. A watermark "THAT EXCEL SITE" is visible in the background.

	A	B	C	D	E	F
1	Values	Normalized Value		Min Value:	4	
2	19	0.5555555556		Max Value:	31	
3	22					
4	18					
5	15					
6	16					
7	16					
8	4					
9	28					
10	26					
11	31					
12	15					
13	18					
14	16					
15	18					
16	11					
17	18					
18	17					
19	20					
20	22					
21						

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

A	B	C	D	E	F
1	Values	Normalized Value		Min Value:	4
2	19	0.5555555556		Max Value:	31
3	22	0.6666666667			
4	18	0.518518519			
5	15	0.407407407			
6	16	0.4444444444			
7	16	0.4444444444			
8	4	0			
9	28	0.8888888889			
10	26	0.814814815			
11	31	1			
12	15	0.407407407			
13	18	0.518518519			
14	16	0.4444444444			
15	18	0.518518519			
16	11	0.259259259			
17	18	0.518518519			
18	17	0.481481481			
19	20	0.592592593			
20	22	0.6666666667			
21					

THAT  
EXCEL  
SITE

#### VIVA QUESTIONS:

1. What you mean by normalization?

2. Define 3NF?

**OUTPUT:**

**RESULT:**

Thus the normalizing of data in excel was executed successfully.

**EXP NO: 5**

**DATE:**

## PLOTTING FUNCTIONS ON THE DATA SET

### AIM:

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

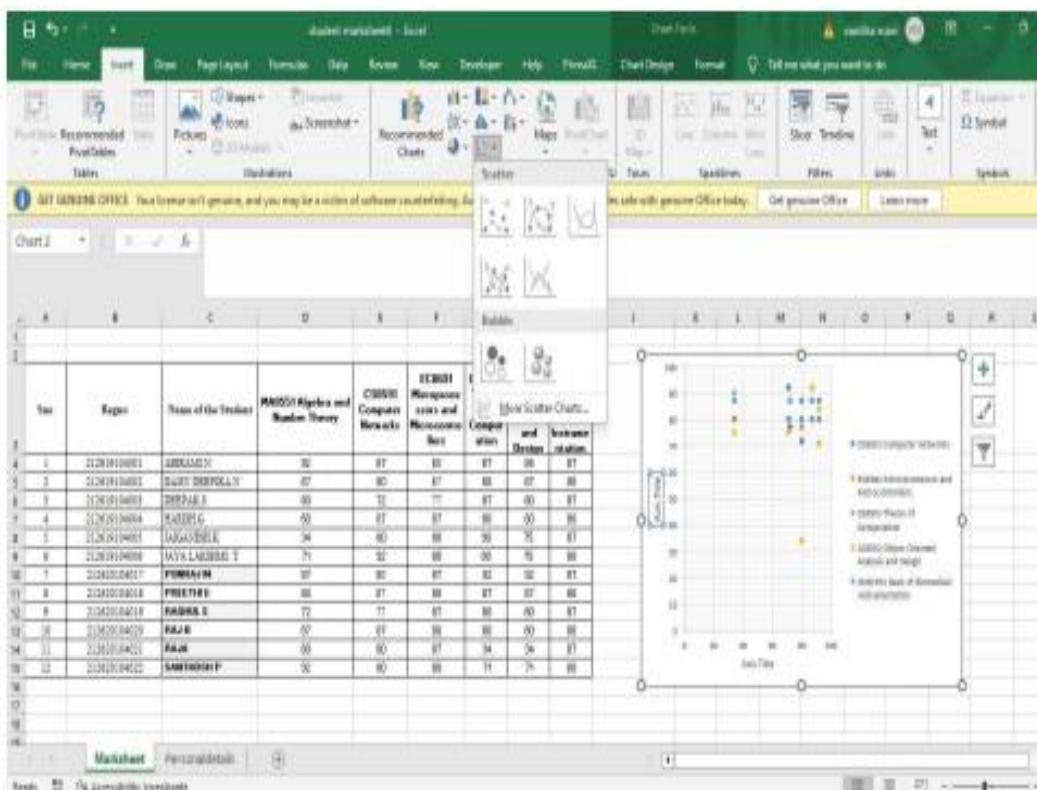
### PROCEDURE:

**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 :** click 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:**



## **VIVA QUESTIONS:**

1. What is the primary purpose of plotting?
  2. What is Matplotlib in python?
  3. Explain scatter plot?
  4. What are the different types of plot you can create?

## **RESULT:**

The plotting functions was applied and explored on the data set successfully using MS-EXCEL and the desired output was displayed in neat format.

<b>EXP NO: 6</b>
<b>DATE:</b>

## **PLOTTING FUNCTIONS ON THE DATA SET USING R-PROGRAMMING**

### **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:



## **VIVA QUESTIONS:**

1. What is the difference between R and RStudio?
  2. Is R or Python better?
  3. What is the advantage of RStudio?

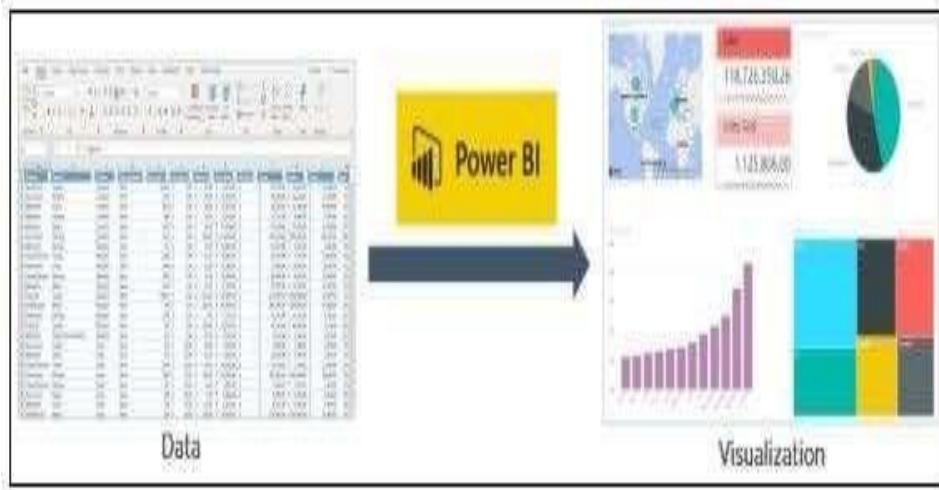
## RESULT:

The plotting functions was applied and explored on the data set successfully using R-programming and the desired output was displayed in neat format.

## EXPLORE THE FEATURES OF POWER BI DESKTOP

### What is Power BI?

Power BI is a business analytics service provided by Microsoft that lets you visualize your data and share insights. It converts data from different sources to build interactive dashboards and Business Intelligence reports.



As you see above, we have some sales data in an excel sheet. Using this data, Power BI helps you build different charts and graphs to visualize the data.

Now that you have learned what is Power BI, let us now understand why Power BI is required.

### Advantages Of Power BI

1. **User-friendly interface:** Power BI has an intuitive interface allowing users to visualize and analyze data easily.
2. **Data integration:** Power BI allows users to easily integrate data from various sources, including Excel, SQL Server, and cloud-based sources like Azure and Salesforce.
3. **Customizable dashboards:** Users can create customized dashboards and reports to display data in a way that is meaningful to them.
4. **Real-time data:** Power BI supports real-time data processing, which means users can view up-to-date data in their dashboards and reports.
5. **Collaboration:** Power BI allows users to share their dashboards and reports with others, making collaborating on data analysis projects easy.

## **Disadvantages Of Power BI**

1. Limited data processing capabilities: Power BI is not designed for heavy-duty data processing and may struggle with large datasets or complex queries.
2. Limited customization options: While Power BI offers a range of customization options, users may find that they are limited in their ability to create truly unique visualizations and reports.
3. Cost: Power BI is not a free tool, and users may need to pay for additional features or storage space. Salient features of Microsoft Power BI

Microsoft Power BI is a collection of BI tools for creating complex visualizations and shareable reports. The following are some of Power BI's outstanding features:

### **1. Analytics and data management**

Microsoft Power BI provides powerful analytics to assist users in gaining valuable insights, transforming data into powerful components capable of providing ideas and breakthroughs to resolve business problems.

Advanced analytics in Power BI allows business users to track key performance parameters. It helps the business figure out the signs which lead to additional opportunities and success. Power BI creates stunning interactive dashboards and is a complex data management solution worthy of being termed. It uses data mining and BI technologies to identify data trends and provides tools to aid advanced analytics.

### **2. Quick Insights**

Using powerful algorithms, users can obtain exciting insights from various subsets of the data.

Quick Insights gives people better and faster access to analytics results with a single click.

### **3. Ask a Question**

With this functionality, users can ask questions and get instant responses in the form of visuals such as charts, tables, or graphs.

#### **4. Intuitive reports**

Power BI reports are a collection of dashboards that are comprehensive and well organized, with various types of visualizations and formats that are relevant to a specific business issue by presenting valuable insights. Users may quickly share Power BI reports with other users.

#### **5. Integration of Azure Machine Learning**

Because of the inclusion of Machine Learning in Power BI, users can now visualize the outcomes of Machine Learning algorithms by dragging, dropping, and joining data modules.

#### **6. Data Analysis Expression**

Data Analysis Expression helps extract one or more values from a data set by analyzing the results generated by applying multiple formulae to the dataset under observation. It works similarly to Microsoft Excel, except without the numbers and rows. Data Analysis reports are simple to comprehend and create.

#### **7. Real-time data analytics**

As data is sent or streamed in, Power BI dashboards refresh in real-time, allowing viewers to solve problems and spot possibilities instantly. Real-time data and images are available in any report or dashboard. Factory sensors, social media sources, and anything else that may gather or transmit time-sensitive data are potential sources of streaming data into Power BI dashboards/reports.

#### **8. Visualization and Reporting**

Using Power BI will result in a more profound knowledge of data and the discovery of business insights in real-time. Power BI has many pre-built visualizations and the possibility to customize existing ones or choose from an ever-growing library of in-built visualizations.

## **9. Lively Dashboards**

Provides a single point of reference with a complete visual picture of business-critical information.

## **10. Pre-existing Ready to use templates**

Microsoft and the community have offered hundreds of out-of-the-box graphics, including charts, cards, KPIs, maps, and matrices.

**EXP NO: 7**

**DATE:**

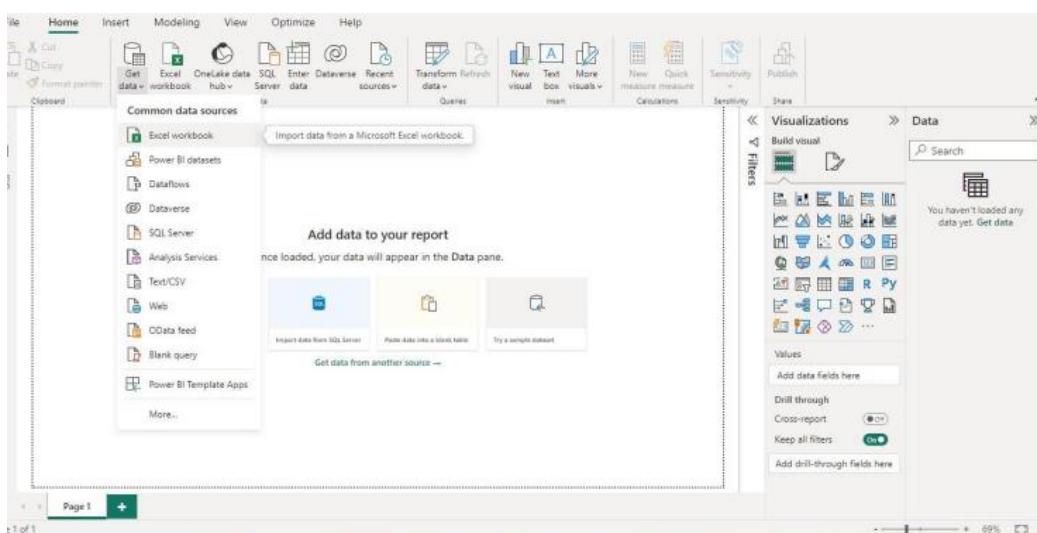
## **PREPARE & LOAD DATA USING POWER BI**

### **AIM:**

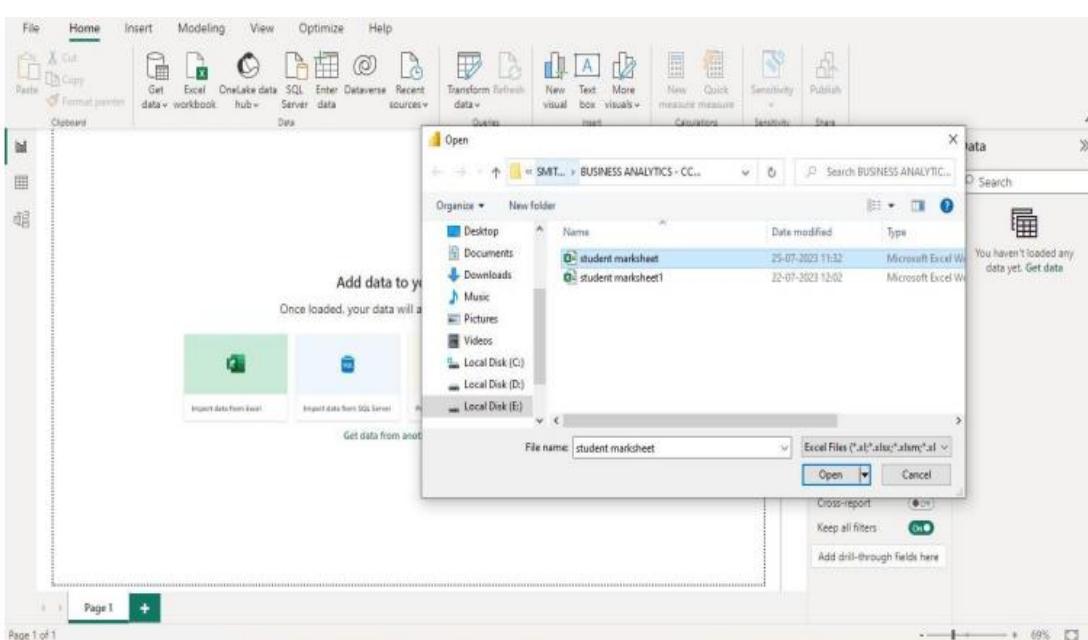
To Prepare & Load data Using Power Bi

### **PROCEDURE:**

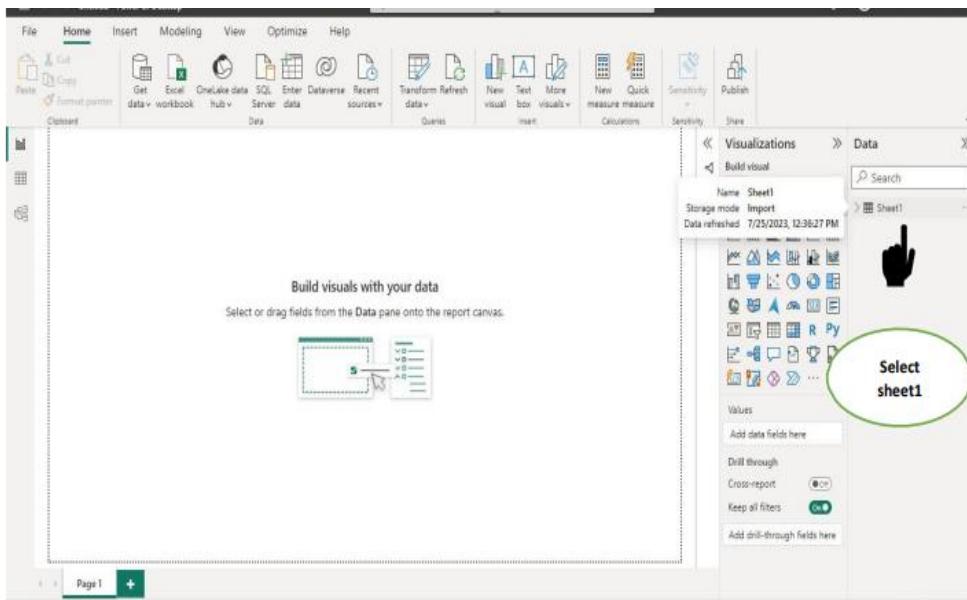
#### **Step 1: Open Power bi application**



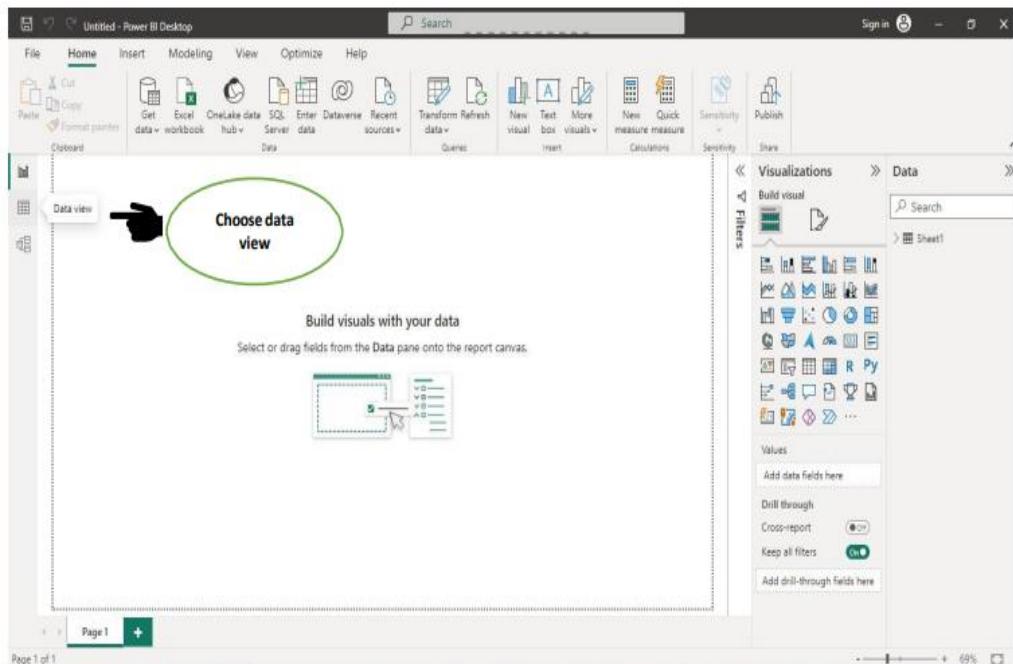
#### **Step 2: Select the data set to load in Power bi application**



**Step 3:** The loaded data set will be displayed on screen.



**Step 4:** choose the data view to display you data set.



**Step 5:** Data set will be displayed and close the application.

The screenshot shows a Microsoft Power BI Desktop interface. The ribbon at the top has 'File', 'Home', 'Help', and 'Table tools' selected. Under 'Table tools', there are tabs for 'Structure' (selected), 'Calendars', 'Relationships', and 'Calculations'. Below the ribbon is a search bar and a 'Sign in' button. The main area displays a table titled 'Sheet1' with the following data:

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8691 Microprocessors and Microcontrollers	CS8501 Theoretical Computer Science
1	212619104001	ABIRAMI.N	92	87	80	
2	212619104002	DAISY DEEPIKA.N	87	80	87	
3	212619104003	DEEPAKS	80	72	77	
4	212619104004	HARISH.G	80	87	87	
5	212619104005	JAGANESH.K	34	80	80	
6	212619104006	JAVA LAKSHMI.T	71	92	80	

## OUTPUT:

## VIVA QUESTIONS:

1. What are the types of data?
  2. Difference between descriptive and diagnostic data analysis?
  3. Give a example for qualitative and quantitative data?

## RESULT:

The Excel data sheet was prepared and loaded successfully into Power BI Desktop and displayed desired output in neat format.

**EXP NO: 8(A)****DATE:**

**PERFORM STATISTICAL OPERATIONS USING POWER BI**  
**[Mean, Median, Mode and Standard Deviation,**  
**Variance]**

**AIM:**

To perform statistical operations like Mean, Median, Mode and Standard deviation, Variance Using Power Bi.

**PROCEDURE:**

**Step 1 :** Start the application Desktop power bi.

**Step 2 :** Create datasheet for student marks in application.

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8631 Microprocessor 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	30	75	87
6	212619104006	JAYA.LAKSHMI.T	71	92	80	80	75	80

**Step 3:** calculate Mean, Median, Mode and Standard deviation, Variance

**Step 4:** your output will be displayed.

Sno	Regno	Name of the Student	MA8551 Algebra and Number Theory	CS8591 Computer Networks	EC8631 Microprocessor 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	30	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:**

**VIVA QUESTIONS:**

1. What are the various statics formulas?
2. What is mean?
3. Write the formula for standard deviation?

**RESULT:**

**To perform statistical operations like Mean, Median, Mode and Standard deviation, Variance Using Power Bi.**

**EXP NO: 8(B)**

**DATE:**

## **DEVELOPING DATA MODEL**

### **AIM:**

To Develop the data model using Power BI Desktop

### **THEORY:**

Data models are visual representations of an enterprise's data elements and the connections between them. By helping to define and structure data in the context of relevant business processes, models support the development of effective information systems.

The predominant data modelling types are hierarchical, network, relational, and entity-relationship. The Data Modelling diagram is used to create or view graphical models of relational database system schemas including a range of database objects.

### **PROCEDURE:**

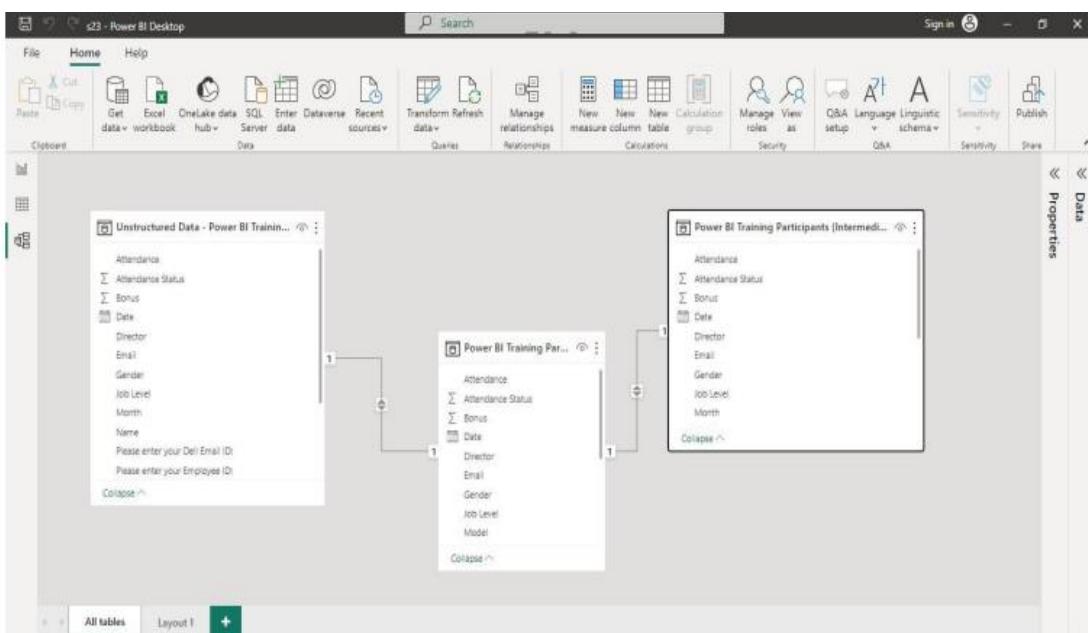
**Step 1 :** Start Power BI Desktop

**Step 2 :** Goto to home menu , select get data and then choose excel worksheet

**Step 3 :** Select your file to load into Power Bi application and press load button

**Step 4 :** Now the file is loaded and now click model view in the left panel

**Step 5:** Now data modeling of the given tables are displayed in output screen



**Step 6 :** Save the file in desired location

**Step 7 :** Close the Power BI Desktop application

**OUTPUT:**

**VIVA QUESTIONS:**

1. What are the five steps of data modelling?

2. Why data modeling?

3. What are the features of data models?

**RESULT:**

The data model was developed and loaded successfully using Power BI Desktop and displayed desired output in neat format.

**EXP NO: 9**

**DATE:**

## **DATA PREPROCESSING OPERATIONS**

### **AIM:**

To perform data preprocessing operations using Power BI Desktop.

### **THEORY:**

Data processing starts with data in its raw form and converts it into a more readable format (graphs, documents, etc.), giving it the form and context necessary to be interpreted by computers and utilized by employees throughout an organization.

Data preprocessing includes

- Removing un relevant data
- Removing null values
- Add needed columns
- Add default values

### **PROCEDURE:**

Step 1: Load the dataset in power bi.

The screenshot shows the Power BI Desktop interface with the 'Day3\_PowerQuery - Power BI Desktop' window open. The 'Home' tab is selected. A data source named 'sales\_w04.csv' is currently being loaded. The 'File Origin' dropdown shows '1259: Western European (Windows)'. The 'Delimiter' dropdown is set to 'Comma'. The 'Data Type Detection' section indicates 'Based on first 200 rows'. The main area displays a preview of the data with columns: Row.ID, Order.ID, Order.Date, Ship.Date, Ship.Mode, Customer.ID, Customer.Name, Segment, City, State. Below the preview, there are numerous rows of data. To the right, the 'Fields' pane lists several measures and calculated columns, including 'Retail Analytics Meas.', 'Sales', 'Sales Amount', 'Sales Cty', and 'Calendar'. Some of these items have yellow highlights applied to them.

Step 2: Select the column and click remove option to remove unwanted column.

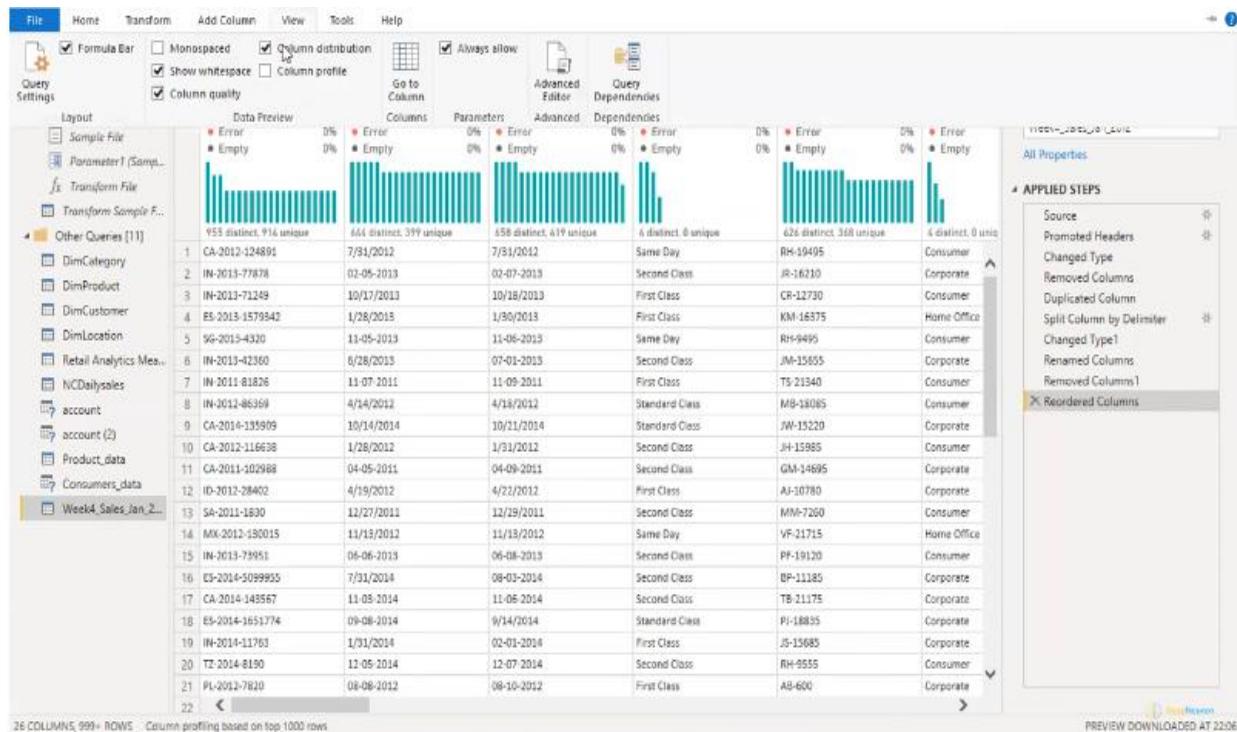
Screenshot of Power BI Data Editor showing the 'Week4\_Sales\_Jan\_2...' query. The 'Row.ID' column is selected, and the context menu shows options like Copy, Remove, and Promoted Headers. The 'Promoted Headers' option is highlighted. The 'APPLIED STEPS' pane shows the step 'Changed Type'.

#### Step 4: make a duplicate column to split a column into two

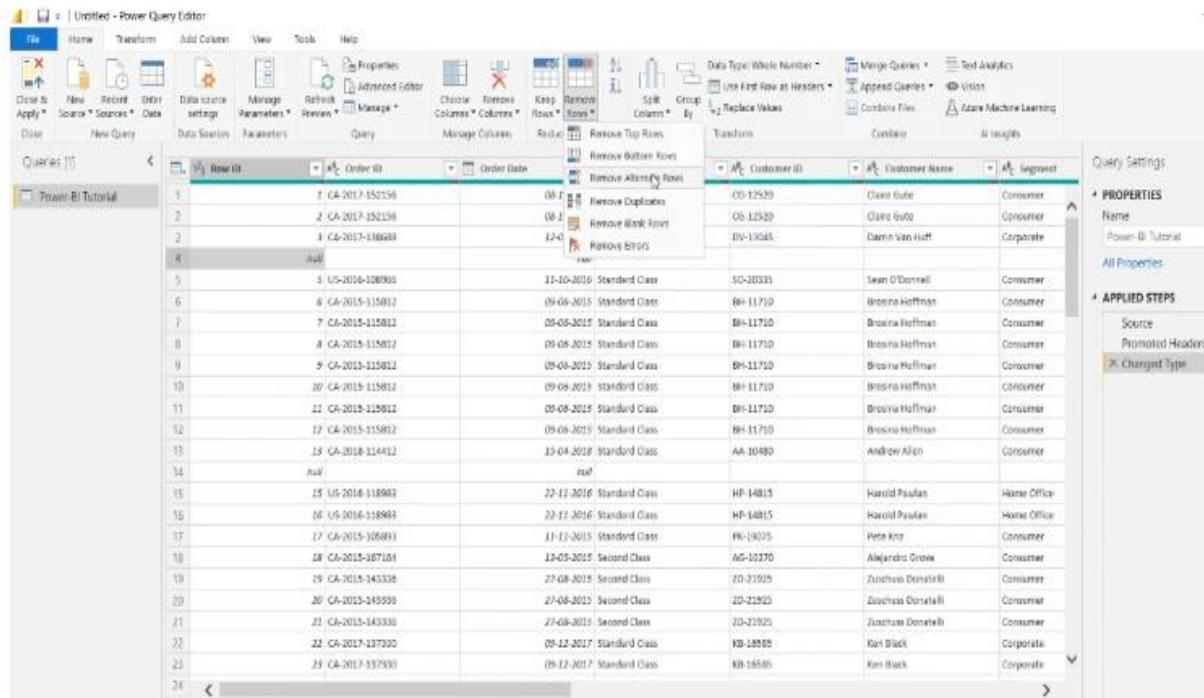
Screenshot of Power BI Data Editor showing the 'Week4\_Sales\_Jan\_2...' query. The 'Customer\_Name' column is selected, and the context menu shows options like Copy, Remove, and Duplicate Column. The 'Duplicate Column' option is highlighted. The 'APPLIED STEPS' pane shows the step 'Removed Columns'.

Step 5: split column using split column by number, delimiter

Step 6: to remove null value select view ribbon,in that select column quilty



Step 7 : to remove null values select remove rows ,in that you can choose the various option.



## OUTPUT:

## VIVA QUESTIONS:

1. What is the ETL process?
  2. What are the four types of data processing?

## **RESULT:**

Thus the data preprocessing operations are executed successfully using Power BI Desktop.

**EXP NO: 10**

**DATE:**

## **DESIGN A REPORT**

### **AIM:**

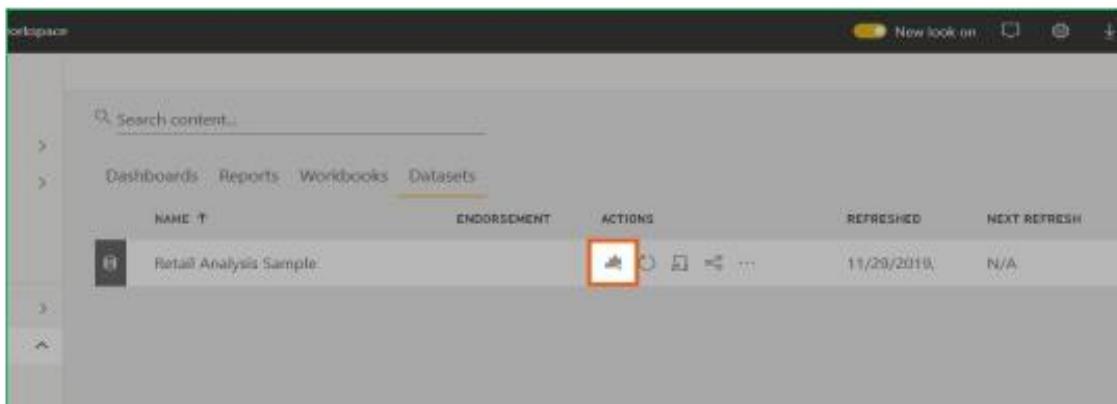
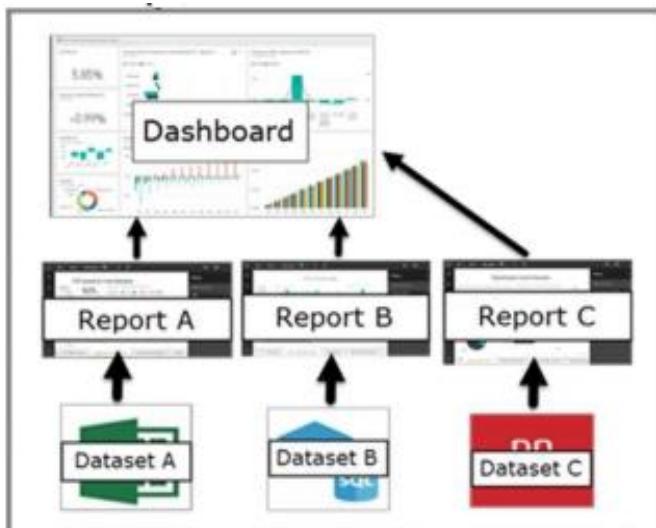
To design a report using Power BI Desktop.

### **THEORY:**

Before you can create any reports in Power BI, you need a dataset and a blank report canvas.

Once you have your dataset, go to the “Datasets” section in your workspace and click the ‘Create report’ icon.

In our example, the available dataset is the “Retail Analysis Sample”:



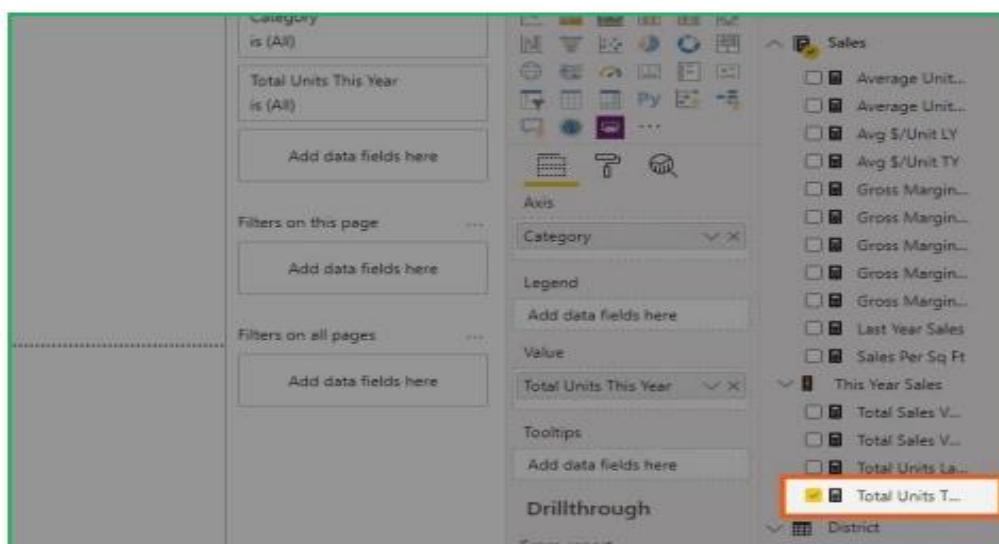
There are three primary sections you have to know:

- Canvas
- Fields pane
- Visualizations pane

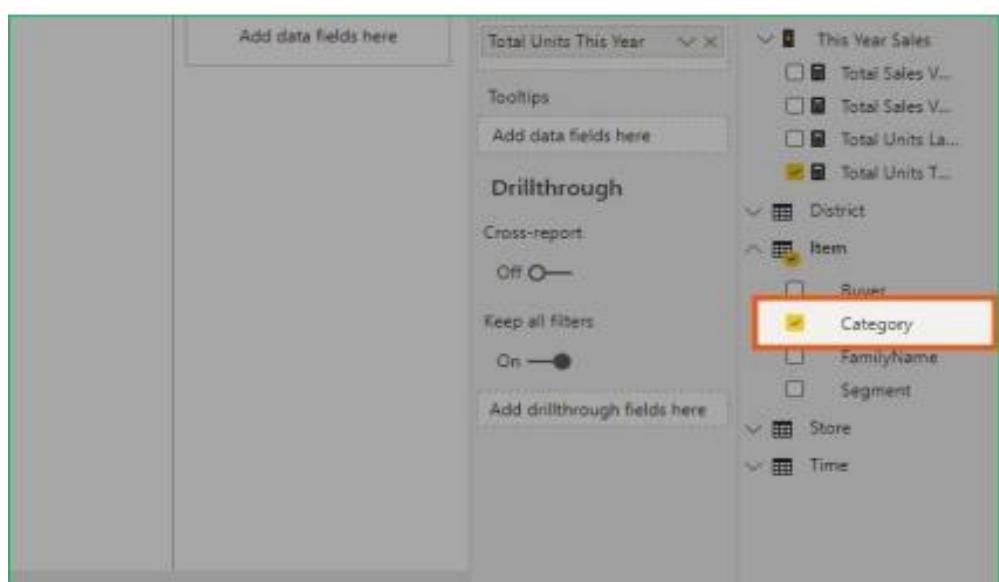
Obviously, the canvas is the empty white space at the center where the visuals will be made. The fields pane contains the different fields of your dataset (think of it as the columns of your dataset). The visualizations pane is where you can edit and modify your visualizations like the type of visualization, the format, and specific options for the values of your visual.

### PROCEDURE:

**Step 1:** Select “Sales” and then “Total Units This Year”



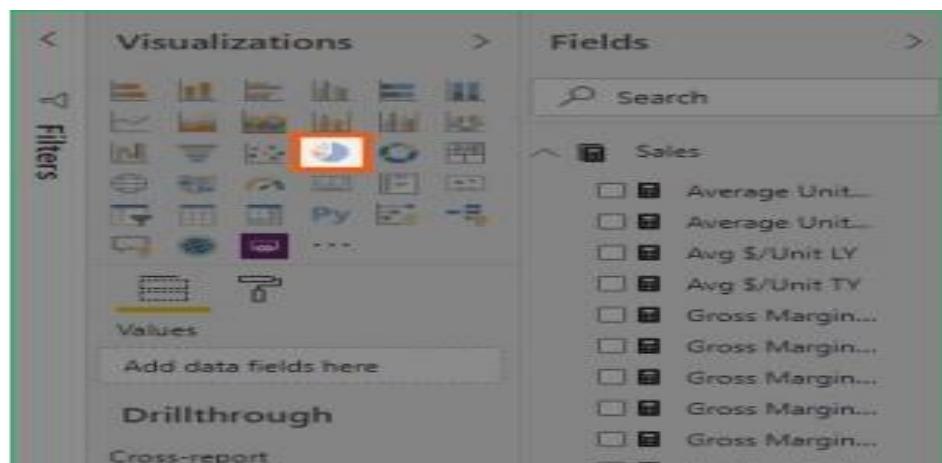
**Step 2:** Select items and then category



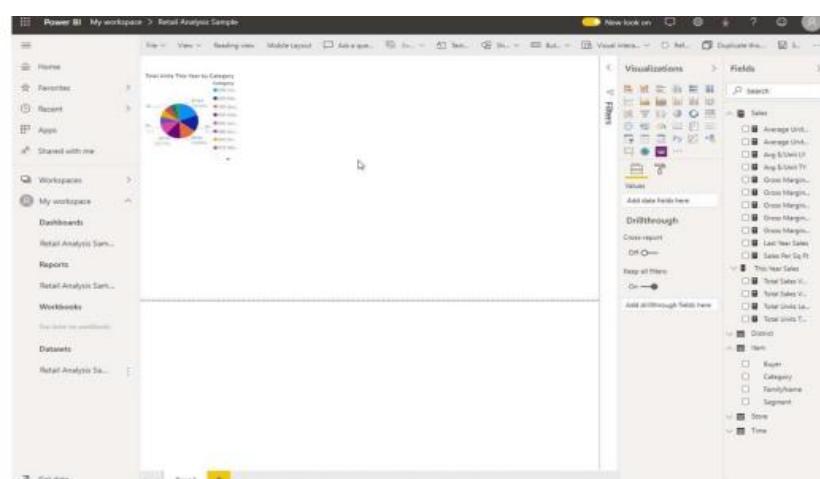
**Step 3: the visuals will appear on canvas.**



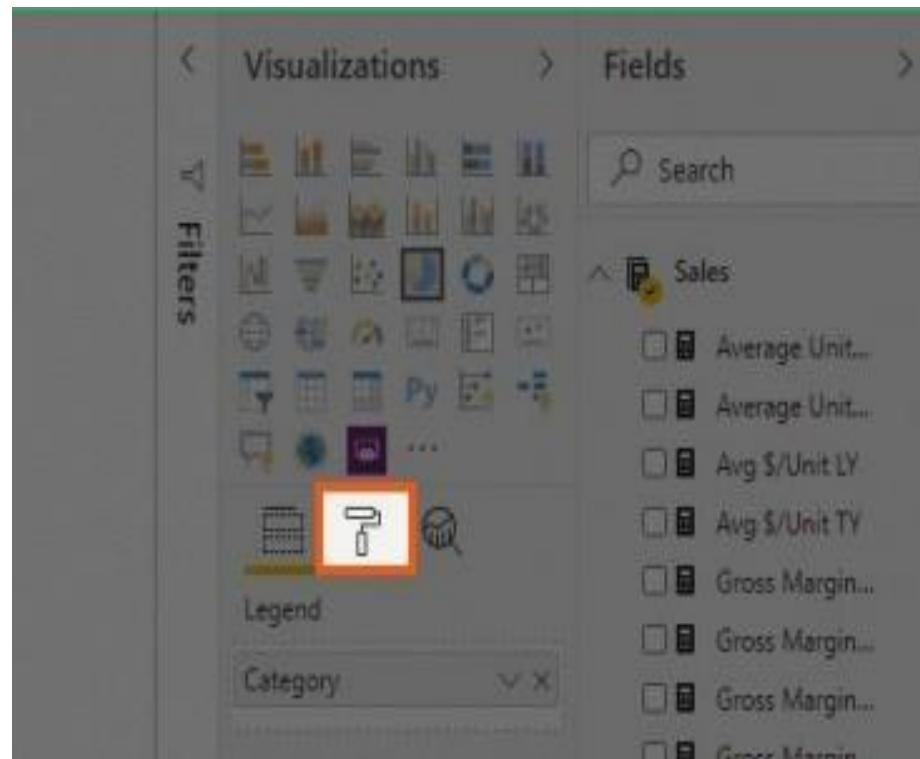
**Step 4: To change the type of visual, just head over to the visualizations pane and click the pie chart icon:**



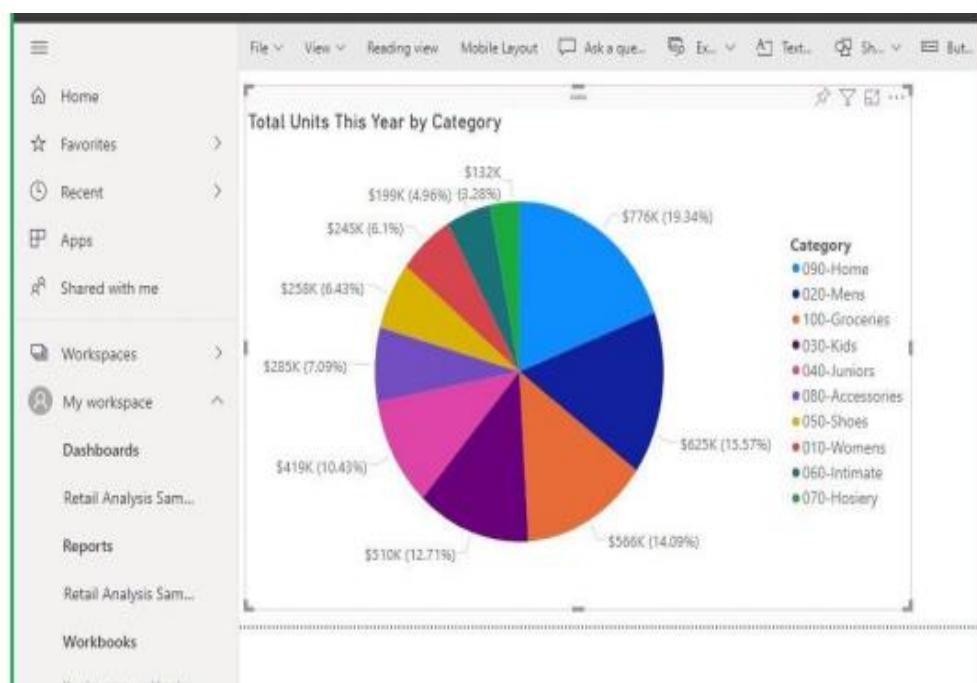
**Step 5: Simply click on the visual and drag its corners:**



**Step 6:** The next part is enlarging the texts — legend, detail labels, and title — of your report. On the visualizations pane, go to the ‘Format’ section:



**Step 7:** you can change the format of your visuals. Each type of visual has different sets of options so an option on one might not be available on another. To enlarge the text, simply go through the ‘Legend’, ‘Detail labels’, and ‘Title’ and then adjust the text



**Step 8:** click ‘File’ from the tab list and select ‘Save’

OUTPUT:

**VIVA QUESTIONS:**

1. List out advanced data-analysis tools.

2. List out Reporting Tools.

3. What is MS in data analysis?

**RESULT:**

Thus the report was designed successfully using power bi tool.

**EXP NO: 11**

**DATE:**

## **CREATE A DASHBOARD AND PERFORM DATA ANALYSIS**

### **AIM:**

To Create a dashboard and perform data analysis using Power BI Desktop.

### **THEORY:**

Power BI offers interactive and dynamic features required for creating interactive dashboards. These dashboards, which are simply a collection of visuals, can be built with a deep level of interactivity and are accessible in various formats to consumers. Since they are usually a single page, Power BI dashboards need to be well-designed highlights of an entire story.

It is also important to note that Power BI dashboards are quite different from Power BI reports. For example:

1. Power BI reports are available on Power BI Desktop and Power BI service, while Power BI dashboards can only be found on Power BI service.
2. Reports can be multi-paged, while dashboards are single-paged highlights.

### **PROCEDURE:**

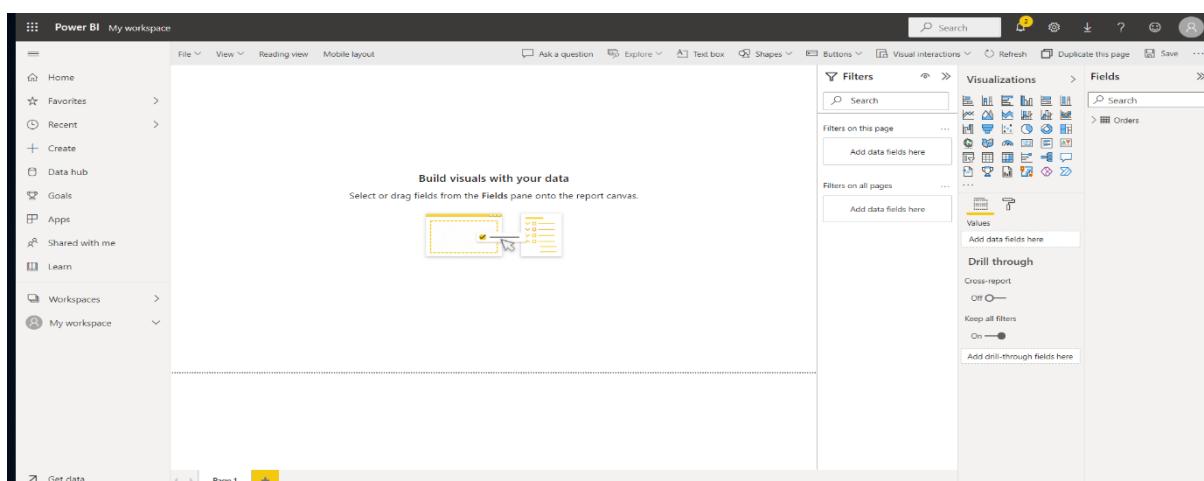
#### **Step 1: load the dataset.**

#### **Step 2: Opening a Report from the Uploaded Data**

Notice that from the workspace image, there are two file types of the same name. One is the dataset and the other is the dashboard.

Click on the Superstore.xlsx dashboard. On the blank canvas that pops up, click on the dataset name ‘Superstore.xlsx’.

Finally, a Reporting canvas will appear, similar to the view on Power BI Desktop.

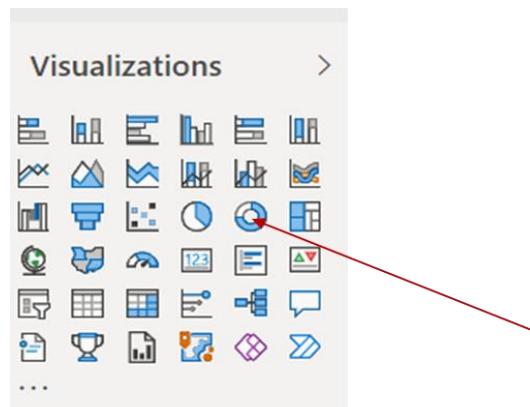


### Step 3: Creating a Tile and ‘Pin to a Dashboard’

You can begin creating your visuals after you have decided on the data that will form the basis of your report, and designed the layout of this report.

We will create a simple donut chart that shows the ‘Sales across the Shipmode’.

- Select the donut chart option from the visualization pane.



Step 4: Drop the Sales in the Values field and Ship Mode to the legend.

**Visualizations**

Legend: Ship Mode

Details: Add data fields here

Values: Sales

Toolips: Add data fields here

Drill through: Cross-report

Cross-report: Off

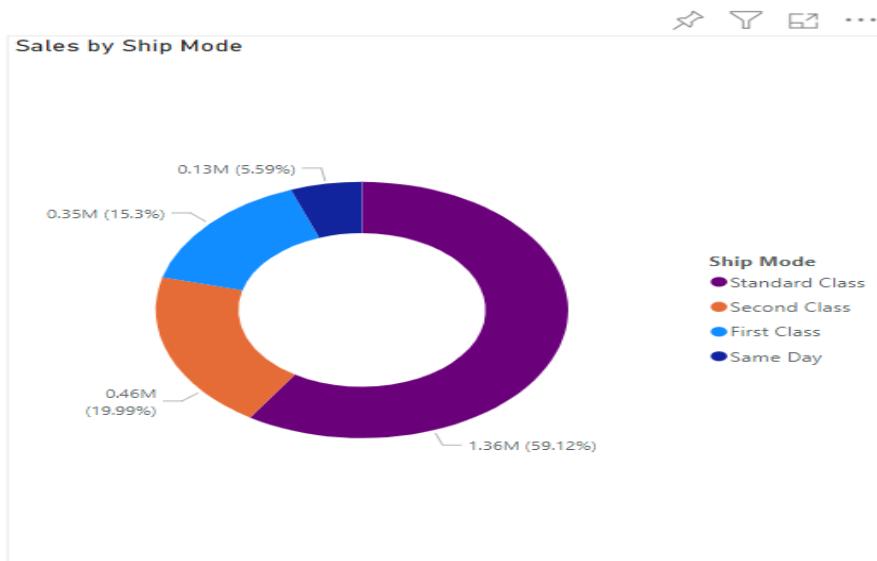
Keep all filters: On

**Fields**

Search: Orders

- Category
- City
- Country
- Customer ID
- Customer Name
- $\Sigma$  Discount
- Order Date
- Order ID
- Order Month
- Order Year
- Postal Code
- Product ID
- Product Name
- $\Sigma$  Profit
- $\Sigma$  Quantity
- Region
- Row ID
- $\Sigma$  Sales
- Segment
- Ship Date
- Ship Mode
- Ship Month
- Ship Year

Step 5: We now have a donut chart.



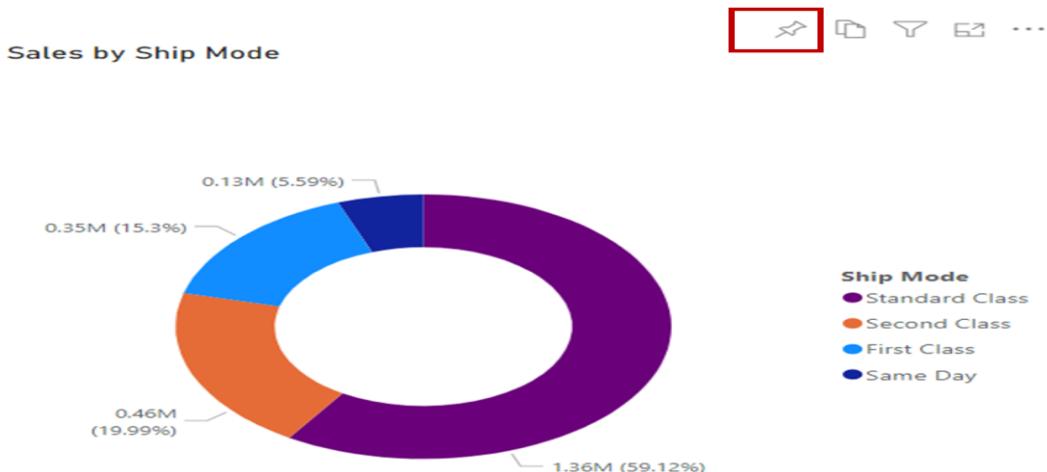
**Step6: Save the report by clicking on the ‘Reading view’ option on the ribbon at the top.**



Step 7: Give the report a name. In this case, we will call it ‘Superstore’.

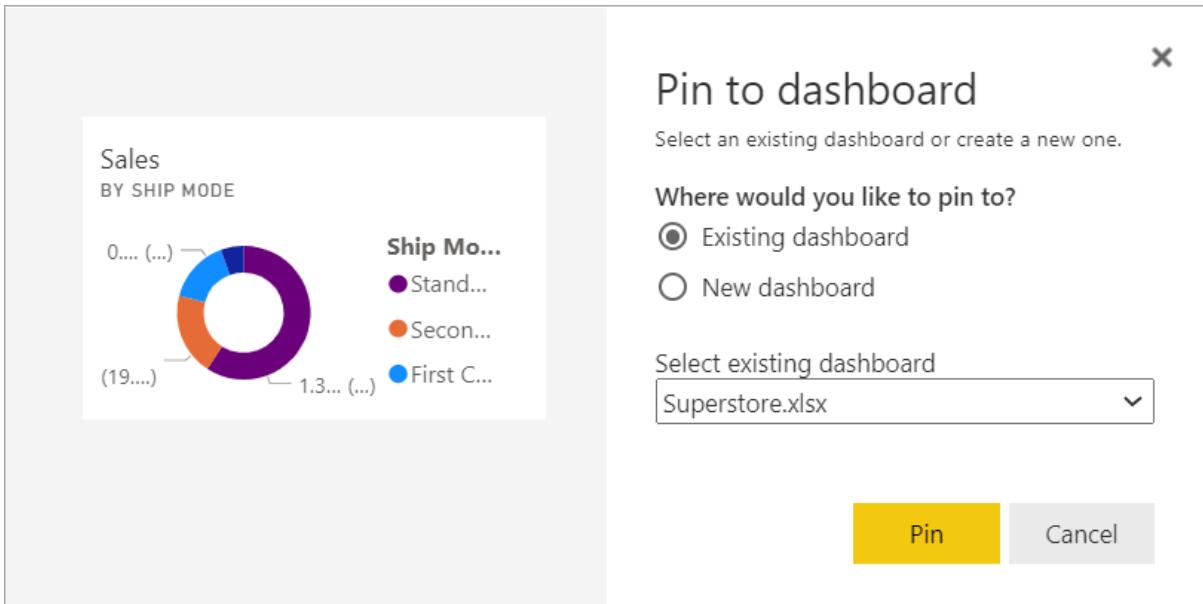
Step 8: Pinning a Single Visual to a Dashboard

- Now that we have created a tile, we can pin this to a dashboard.



Step 9: the recently created donut chart and click on the pin icon that appears.

- A ‘Pin to dashboard’ window will appear. You can pin to an existing dashboard that we created while importing the dataset or create a new one. In our case, we will pin to the existing one and then click on the ‘Pin’ button.



Step 10: A pop-up will appear on the top right. Click on ‘Go to dashboard’.

Step 11: We will be taken to a new window, the **Superstore Dashboard**.

## Step 12: Organizing and Updating Tiles

### Positioning of Tiles

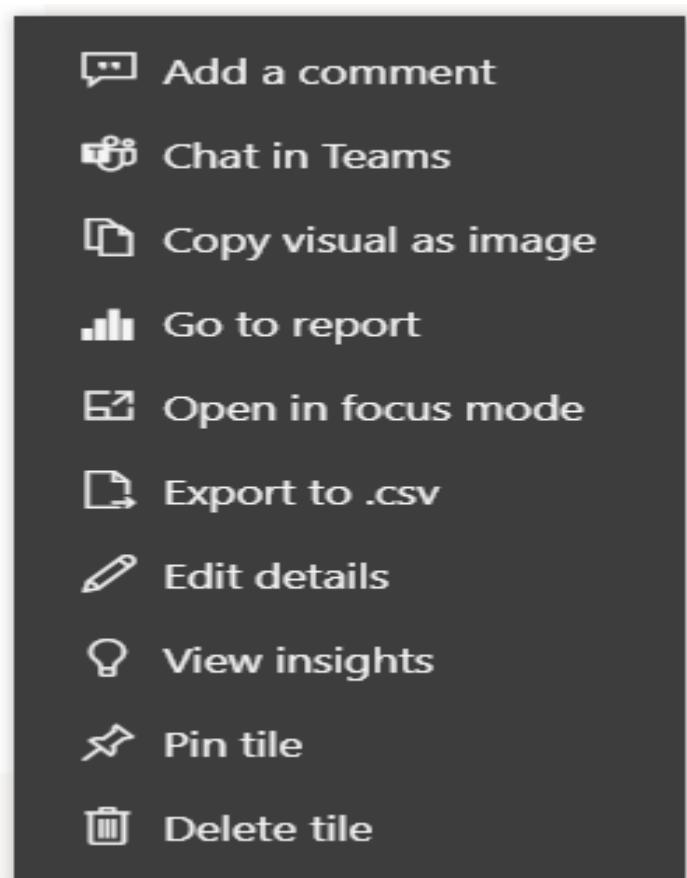
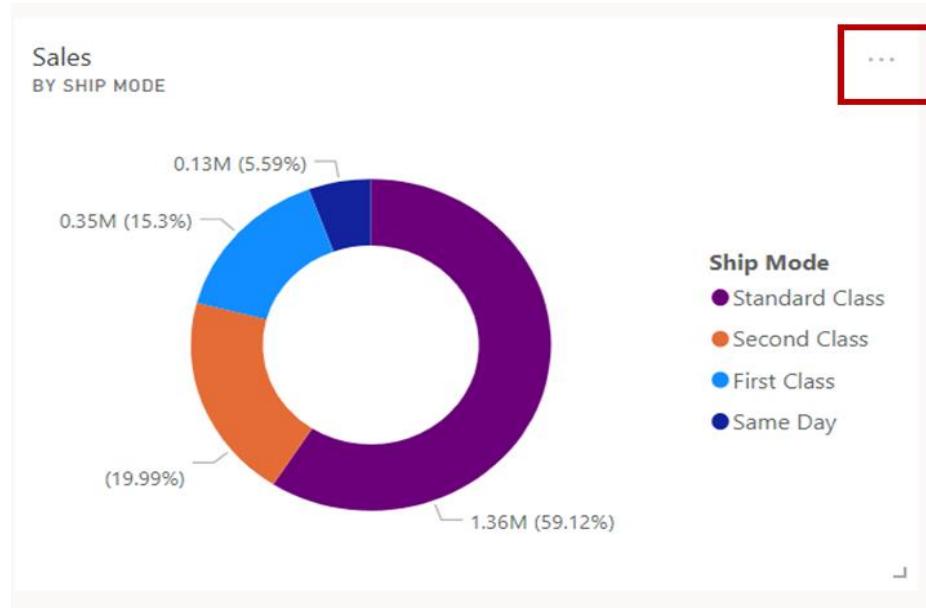
You can move the tile around on your dashboard to a convenient position just by holding a left click on the tile and moving it.

### Tile Size

You can increase or decrease your tile size by dragging the little arrow at the bottom right of the tile.

### Other Tile Options:

By clicking on the kebab (...) menu on the tile, more options are shown around ‘Tile settings’.



### Step 13: Edit Details

- From the option above, we can click on ‘Edit details’.

**Tile details**

\* Required

**Details**

Display title and subtitle

**Title**  
Sales

**Subtitle**  
by Ship Mode

**Functionality**

Display last refresh time

Set custom link

**Link type**

External link

Link to a dashboard or report in the current workspace

**URL \***

**Open custom link in the same tab?**

Yes

No

[Restore default](#)

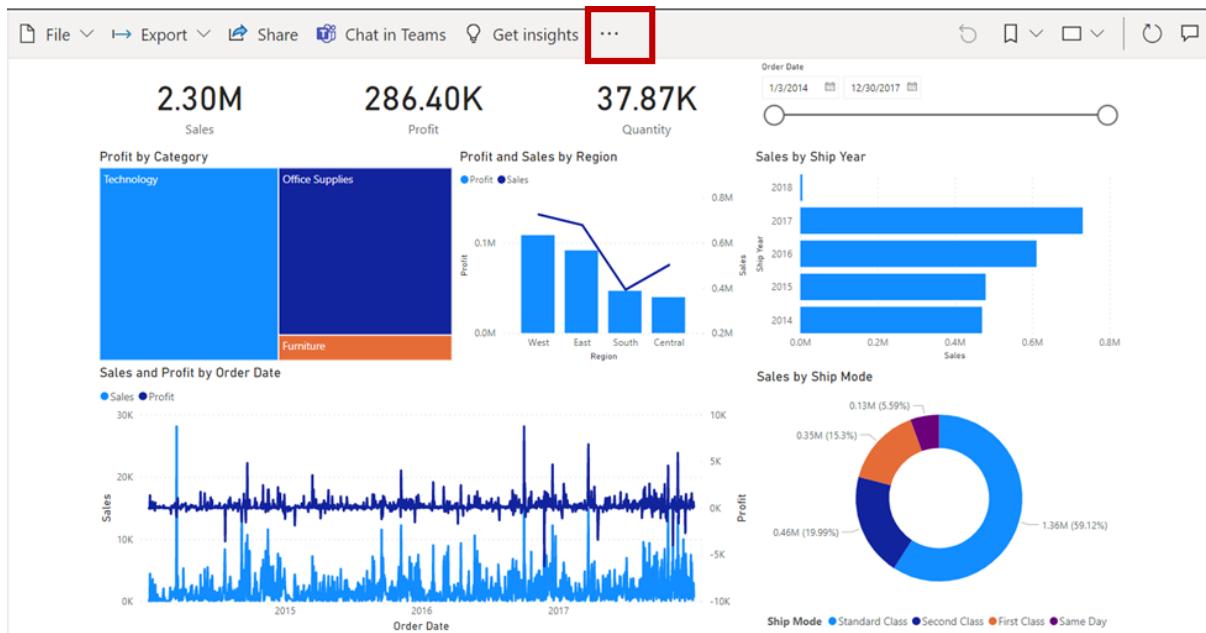
[Technical Details](#)

**Apply** **Cancel**

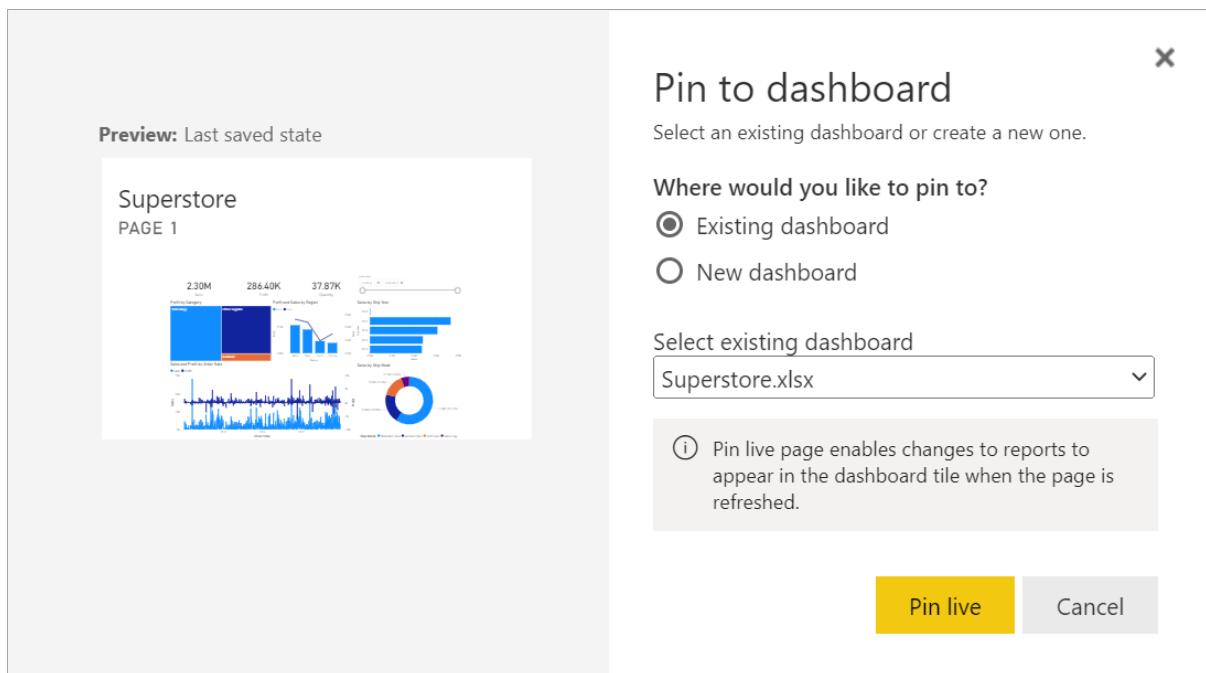
- We can make certain edits, such as changing the tile title and displaying the last refresh time showing when a certain tile was last updated. A custom link can also be associated with your tile.

### Step 14: Pin an Entire Report to a Power BI Dashboard

- We will first need to create more visuals for our Superstore Report.
- Go back to ‘My workspace’. Open the Superstore report type.
- Click on the Edit button.
- Add some more Power BI Visuals.
- Select the option (...) at the top of the report view.



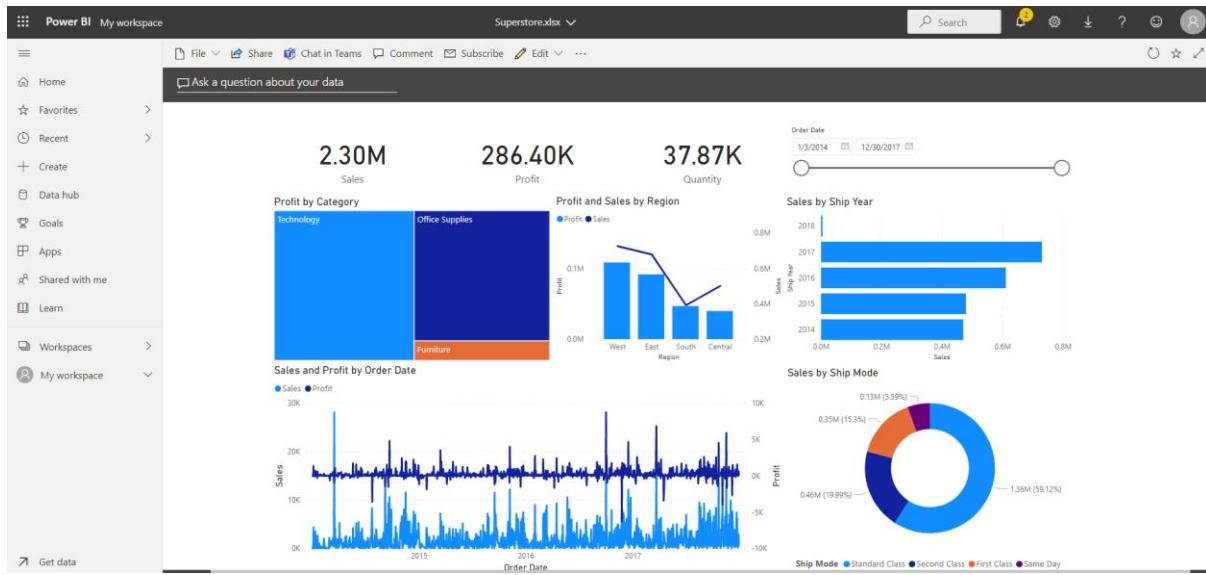
Step 15: Choose the ‘Pin to a dashboard’ option.



Step 16: Click the ‘Pin live’ option. Then select the ‘Go to dashboard’ option.

Note: when pinning at a report level, any changes that happen to the report will also be taken into effect on the dashboard.

- You should now see your report page pinned to a dashboard.



## OUTPUT:

**VIVA QUESTIONS:**

1.What is an example of a data analytics dashboard?

2.What is a KPI dashboard?

3.What are the benefits of a dashboard?

4.What are the four elements of dashboard?

**RESULT:**

Thus the dashboard was created and data analysis was performed successfully.

<b>EXP NO: 12</b>	<b>PERFORM DAX CALCULATIONS</b>
<b>DATE:</b>	

**AIM:**

To Perform DAX calculations using Power BI Desktop.

**THEORY:**

Data Analysis Expressions (DAX) is a formula expression language used in Analysis Services, Power BI, and Power Pivot in Excel. DAX formulas include functions, operators, and values to perform advanced calculations and queries on data in related tables and columns in tabular data models.

If you need to perform an OR operation on multiple expressions, you can create a series of calculations or, better, use the OR operator (||) to join all of them in a simpler expression. The function evaluates the arguments until the first TRUE argument, then returns TRUE.

**PROCEDURE:**

Step 1: You can directly upload the data table to the Power BI file.

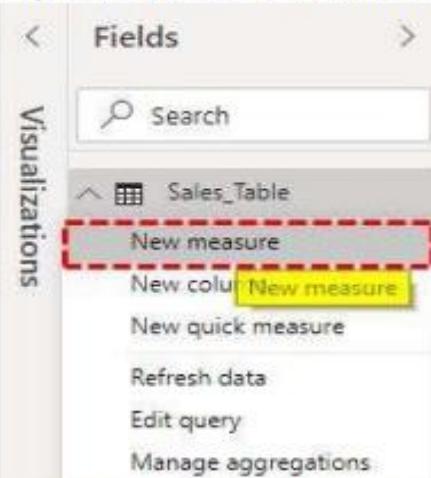
	A	B	C	D	E
1	City	State	Units Sold	Sale Price	Sale Value
2	Auburn	Alabama	\$ 143	\$ 14	\$ 2,002
3	Auburn	New York	\$ 181	\$ 22	\$ 3,982
4	Auburn	Washington	\$ 219	\$ 18	\$ 3,942
5	Columbia	Maryland	\$ 363	\$ 13	\$ 4,719
6	Columbia	South Carolina	\$ 309	\$ 15	\$ 4,635
7	Columbia	South Carolina	\$ 436	\$ 24	\$ 10,464
8	Columbia	Georgia	\$ 154	\$ 17	\$ 2,618
9	Columbus	Georgia	\$ 544	\$ 16	\$ 8,704
10	Columbus	South Carolina	\$ 123	\$ 14	\$ 1,722
11	Columbus	Georgia	\$ 241	\$ 11	\$ 2,651
12	Concord	California	\$ 272	\$ 18	\$ 4,896
13	Concord	New Hampshire	\$ 270	\$ 21	\$ 5,670
14	Concord	California	\$ 433	\$ 10	\$ 4,330
15	Des Moines	Iowa	\$ 473	\$ 16	\$ 7,568
16	Des Moines	Washington	\$ 129	\$ 18	\$ 2,322
17	Des Moines	Iowa	\$ 382	\$ 24	\$ 9,168
18					

Now, we will experiment CALCULATE function to arrive at different sets of results. Arrive at one particular city sales total

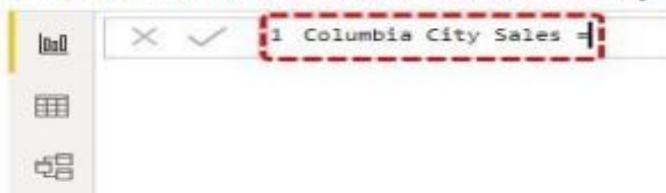
For example, assume you need to create a “New Measure,” which gives one particular city total, for example, “Columbia” city.

The steps to use the DAX calculate function in Power BI is as follows.

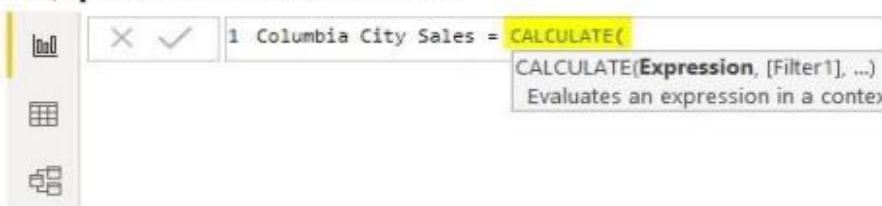
1. Right-click on the table, and choose the “New measure” option.



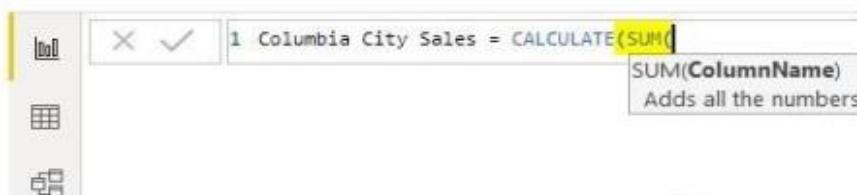
2. Give the name to this measure “Columbia City Sales.”



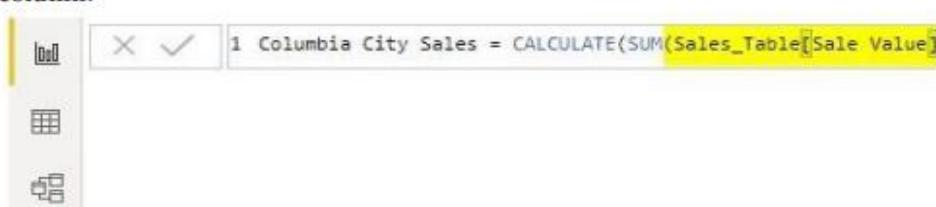
3. Now, open the CALCULATE function.



4. An expression is the first option. We need to add the “Columbia” city total in this example, so open the SUM function.



5. The “ColumnName” we need to SUM is the “Sales Value” column, so choose the respective column.



6. The “ColumnName” we need to SUM is the “Sales Value” column, so choose the respective column.

1 Columbia City Sales = CALCULATE(SUM(Sales\_Table[Sale Value]), FILTER(**Table**, FilterExpression))  
FILTER(Table, FilterExpression)  
Returns a table that has been filtered.

7. The table we refer to is the “Sales\_Table.” So first, choose the table name.

1 Columbia City Sales = CALCULATE(SUM(Sales\_Table[Sale Value]), FILTER(**Sales\_Table**,

8. We need to select the “City” column for Filter Expression and give the criteria as “Columbia.”

1 Columbia City Sales = CALCULATE(SUM(Sales\_Table[Sale Value]),  
FILTER(**Sales\_Table**, **Sales\_Table[City]** = "Columbia")  
FILTER(Table, FilterExpression)  
Returns a table that has been filtered.

9. We are done. Close two brackets and press the “Enter” key to get the new measure.

City	State	Units Sold	Sale Price	Sale Value
Auburn	Alabama	143	14	2002
Auburn	New York	181	22	3982
Auburn	Washington	219	18	3942
Columbia	Maryland	363	13	4719
Columbia	South Carolina	309	15	4635
Columbia	South Carolina	436	24	10464
Columbia	Georgia	154	17	2618
Columbus	Georgia	544	16	8704
Columbus	South Carolina	123	14	1722

10. Drag the “Columbia City Sales” to “Fields” to see the new measure.

Visualizations > Fields >

Search

Sales\_Table

City

Columbia City...

11. Now, this measure only gives the total sales of the city “Columbia.”



12. You can cross-check the “Columbia” city total in Excel as well.

	A	B
1		
2		
3	<b>City</b>	<b>Sum of Sale Value</b>
4	Auburn	9,926
5	Columbia	22,436
6	Columbus	13,077
7	Concord	14,896
8	Des Moines	19,058
9	<b>Grand Total</b>	<b>79,393</b>
10		

Like this, we can use the CALCULATE function to arrive at different results.

## OUTPUT:

**VIVA QUESTIONS:**

1. When should I use DAX?

2. What are the two types of DAX?

3. What is the advantage of DAX?

**RESULT:**

Thus the about DAX calculations was executing successfully using power bi tool.

<b>EXP NO: 13</b>
<b>DATE:</b>

## **PRESENTATION OF A CASE STUDY CAMPUS RECRUITMENT ANALYSIS**

### **AIM:**

To Presentation of a case study for Campus Recruitment Analysis using Power BI Desktop

### The Challenge:

Campus Placements/ Campus Recruitment drives are conducted in various educational institutes for providing job opportunities to the students who are pursuing their particular academic courses. As much as it is important to the students, it is also important to the institute as it gives a chance to contemplate about the process. This data includes students from various colleges.

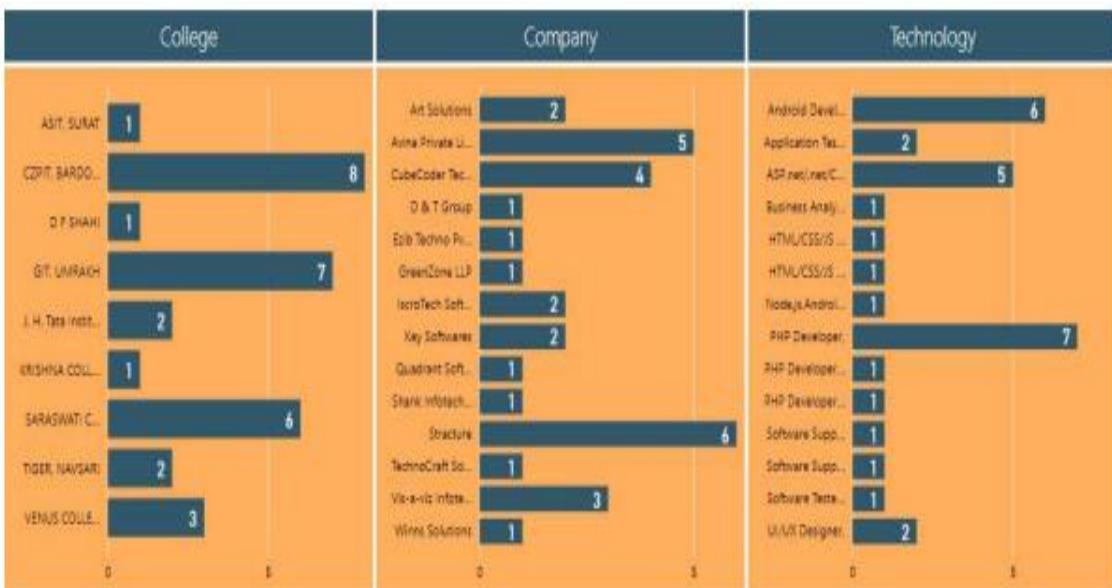
### The Solution What every management team wants to know:

- How many companies appeared?
- How many students appeared?
- Which companies have hired the most of the students and from which college?
- How many students accepted the offer?
- Were the companies beneficial in accordance to their vacancies?
- How many students were selected by the companies?
- What was the ratio between students applied and the vacancies available?
- Which college were the most preferred by the students?
- Which technologies were the most preferred by the students?
- Which colleges were benefitted by this recruitment in terms of the selection?
- Which colleges did not perform up to the mark?
- Which colleges were involved the most in terms of the students registered?
- Which companies were beneficial in accordance to their vacancies?

### The Extra Mile

- We gave users a menu driven page where they can navigate to the page of their interest.

- Overview of companies with respect to the technologies.
- Overview of students with respect to the technologies, companies and colleges.
- Overview of selected students with respect to the technologies, companies and colleges.
- Insights on basis of the students' priority and the offers given to them.
- We have given extra information apart from what is need about every area of interest.



**OUTPUT:****RESULT:**

The Presentation of a case study for Campus Recruitment Analysis was performed Successfully using Power BI Desktop and displayed desired output in neat format.