knowledgehut
upGrad

Data Analysis in Excel

Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.

By the End of This Session:

- You shall learn the fundamentals of Statistical Analysis.
- You shall understand the various measures of central tendency and the measures of spread.
- You will learn how to perform various types of analysis using the above measures.
- You will learn to create and analyze pivot tables.

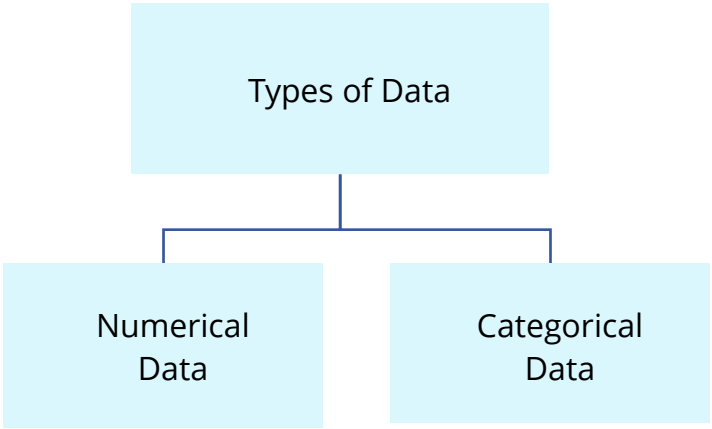
What Have We Learned So Far?

- Fundamentals of Excel
- Multiple components of the Excel toolbar
- Use of Formula Bar
- Creating dynamic formulae using cell references and cell range
- Using and manipulating cell name ranges
- Multiple ways of formatting rows, columns and datatypes
- Variety of functions for text and date columns

Introduction to Statistical Analysis

Introduction to Statistics

Statistics is the science of collecting, analyzing, interpreting, and presenting data.

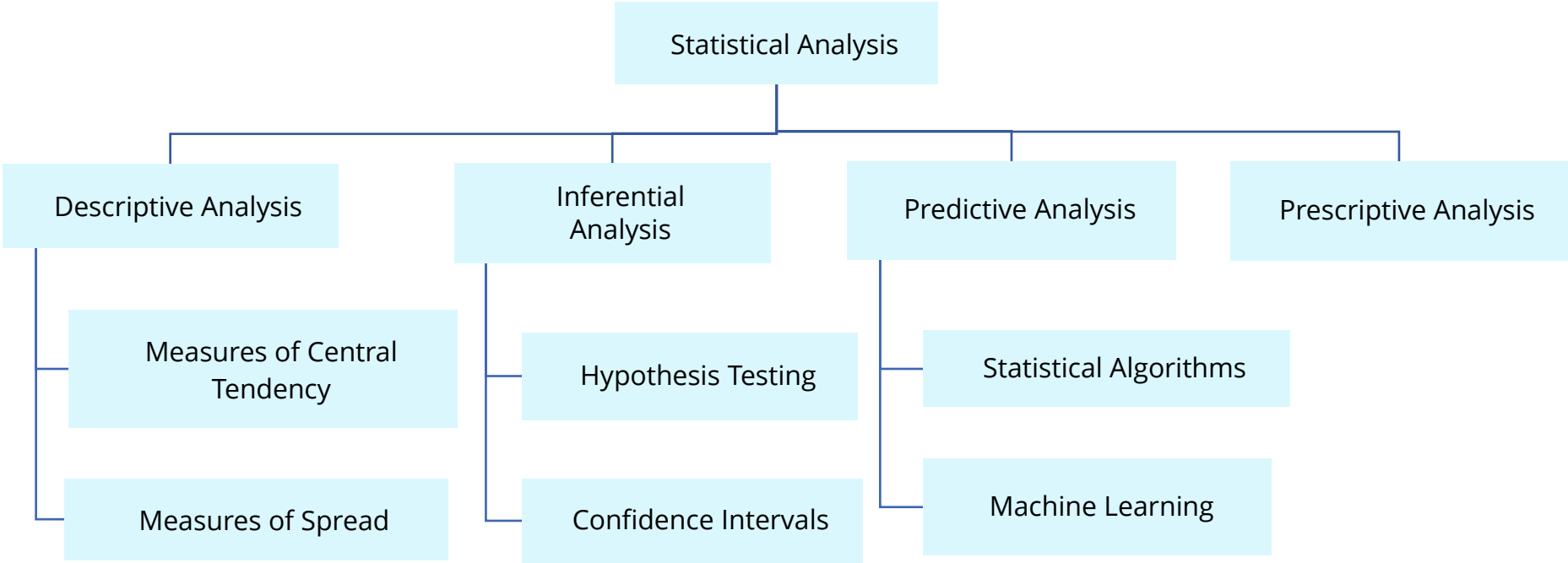


Real-world Application of Statistics

Design and analyze clinical trials to determine the effectiveness of new drugs.

Analyze survey data to understand consumer preferences, market trends, and customer satisfaction.

Types of Statistical Analysis



Poll Time

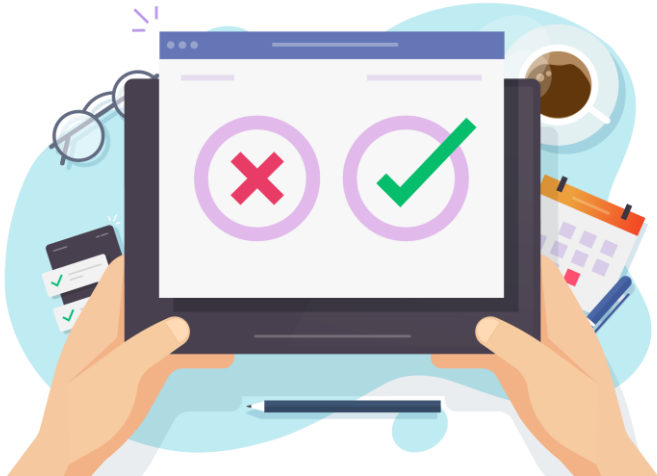
Q. How would you categorize the process of forecasting the flight demand for the next 2 weeks for an airline?

a. Descriptive Analysis

b. Prescriptive Analysis

c. Inferential Analysis

d. Predictive Analysis



Poll Time

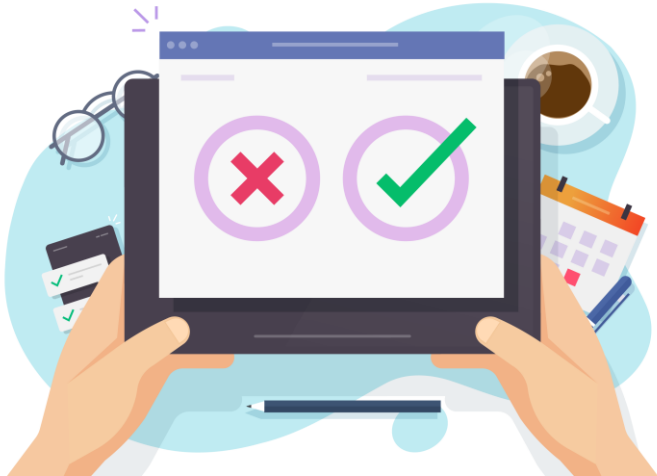
Q. How would you categorize the process of forecasting the flight demand for the next 2 weeks for an airline?

a. Descriptive Analysis

b. Prescriptive Analysis

c. Inferential Analysis

d. Predictive Analysis



Measures of Central Tendencies

Measures of Central Tendencies

Measures of central tendency summarize the center or average value of a dataset.

- Mean
- Median
- Mode

Introduction to Mean

- Mean is defined as the average value of a collection of numbers.
- It considers all values equally.
- It is sensitive to extreme values in the data.

$$Mean = \frac{Sum\ of\ Values}{Count\ of\ Values}$$

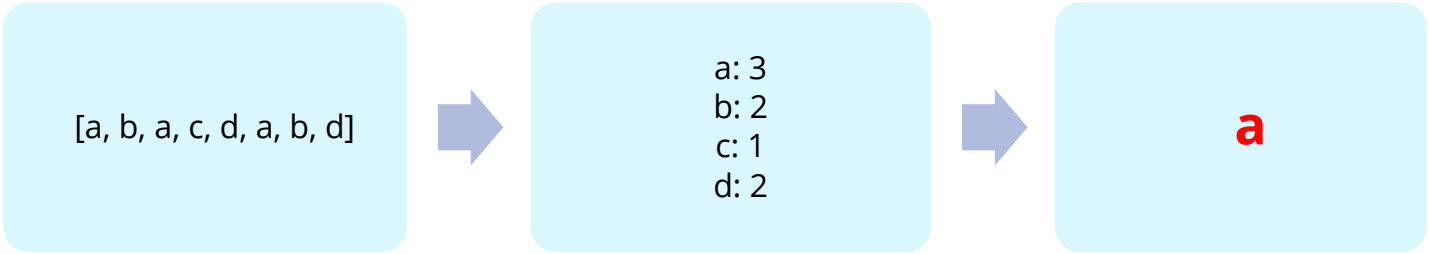
Introduction to Median

- The median is the middle value of a dataset when arranged in ascending or descending order.
- It represents the 50^{th} percentile value of a dataset.
- Unlike the mean, the median is not sensitive to outliers.

[10, 3, 5, 2, 8, 1, 7] → [1, 2, 3, 5, 7, 8, 10] → [1, 2, 3, **5**, 7, 8, 10]

Introduction to Mode

- The mode is the value that appears most frequently in the dataset.
- A dataset can have no mode (no value appears more than once) or multiple modes (multiple values appear with the same frequency).
- Mode is generally used in the case of discreet or categorical columns.



Hands-on – Measures of Central Tendency in Excel

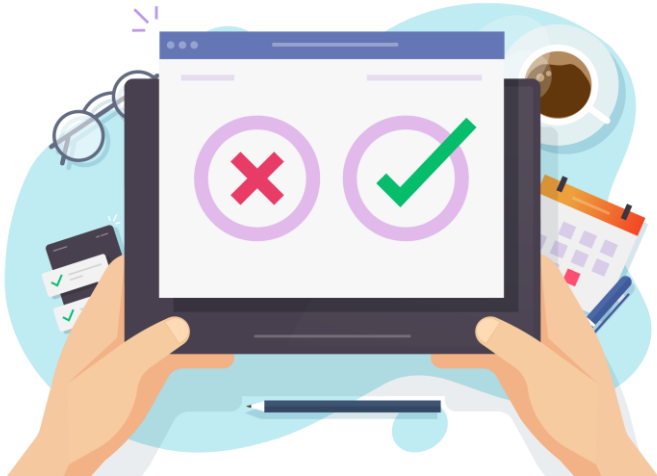
Poll Time

Q. We use the median to summarize the values in a column that contains outliers or extreme values. Is this True or False?

a. True

b. False

c. Depends on the use-case



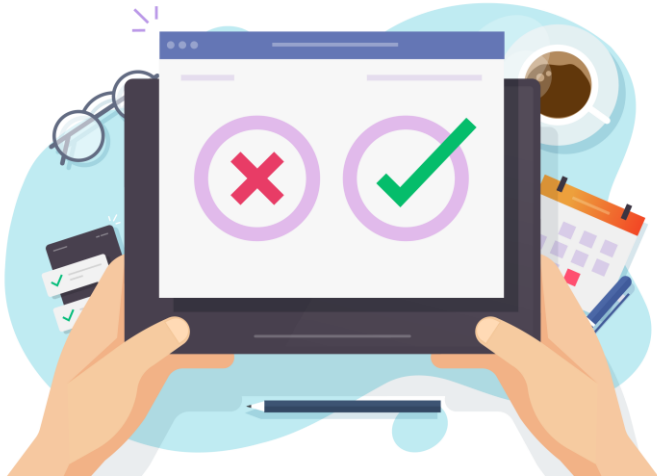
Poll Time

Q. We use the median to summarize the values in a column that contains outliers or extreme values. Is this True or False?

a. True

b. False

c. Depends on the use-case

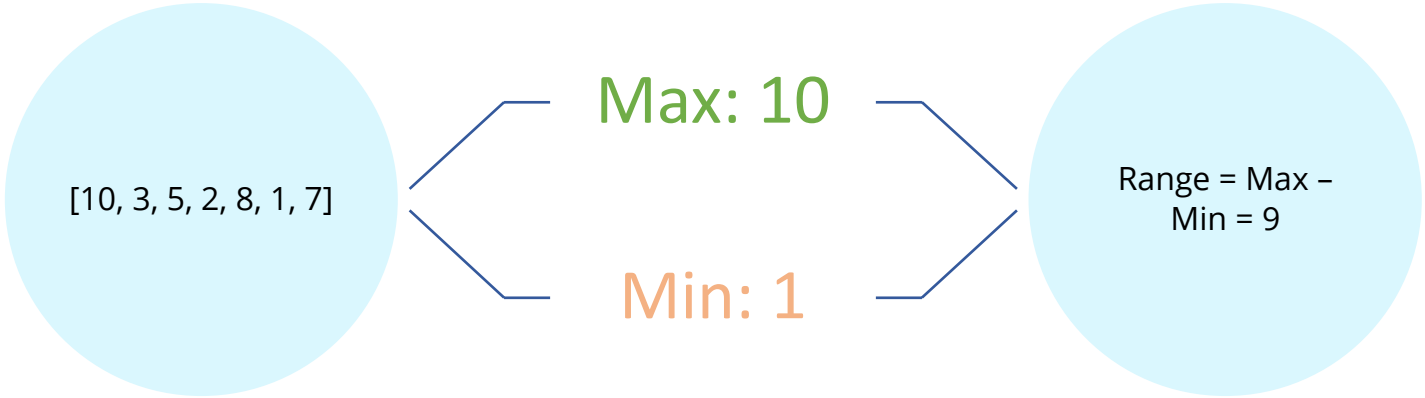


Q&A

Measures of Spread

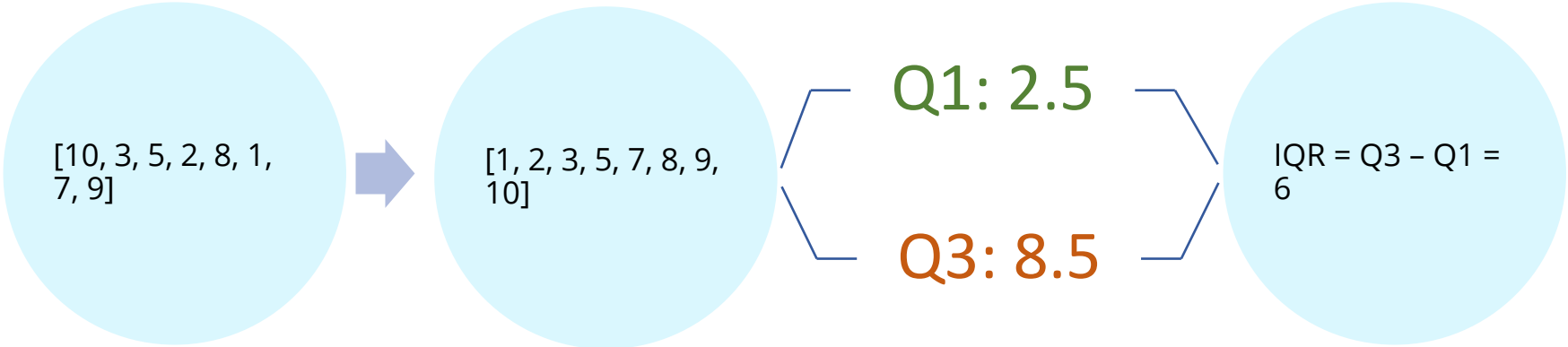
Introduction to Range

- The range is the simplest measure of spread.
- It is calculated by subtracting the minimum value from the maximum value in the dataset.
- The range gives an idea of the spread but is influenced by outliers.



Introduction to Interquartile Range

- The IQR is a measure that focuses on the middle 50% of the data.
- It is calculated as the difference between the third quartile (Q3) and the first quartile (Q1) of the dataset.
- The IQR is resistant to outliers and provides a measure of variability within the central portion of the data.



Introduction to Standard Deviation

- The Standard Deviation measures the average deviation of a data point from it's mean.
- Standard Deviation = Square root of Variance.
- Standard Deviation is influenced by outliers and provides a more detailed measure of dispersion.

$$Standard\ Deviation = \sqrt{\frac{Sum\ of\ squared\ deviations\ from\ the\ mean}{(Number\ of\ observations - 1)}}$$

Poll Time

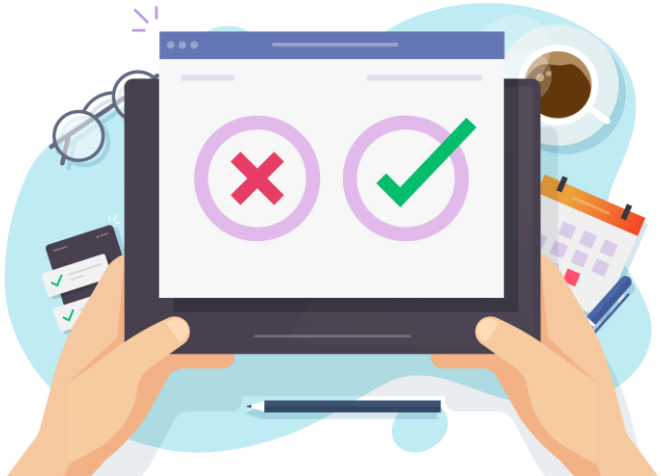
Q. Which of the following is not true about IQR?

a. IQR is not affected by outliers

b. IQR is the difference between Q2 and Q1

c. IQR is the difference between Q3 and Q1

d. IQR is a measure of spread



Poll Time

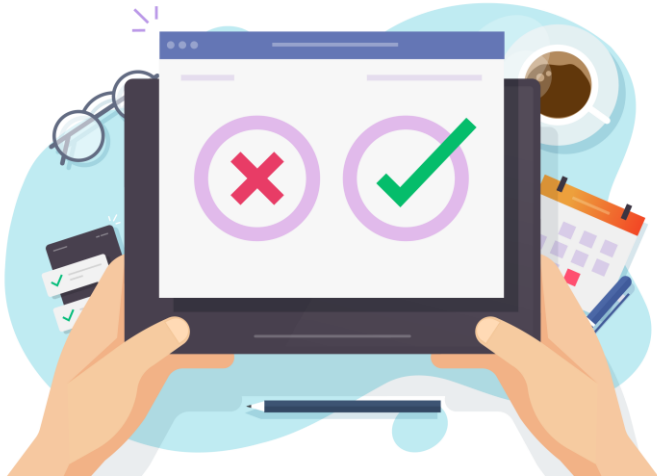
Q. Which of the following is not true about IQR?

a. IQR is not affected by outliers

b. IQR is the difference between Q2 and Q1

c. IQR is the difference between Q3 and Q1

d. IQR is a measure of spread



Hands-on – Calculating Range, IQR, and SD

Exercise

Using the insurance data perform the below analysis:

- Find out the most frequent age of the customers.
- Calculate the standard deviation in the premiums paid by the customers of the above age group.
- Determine the difference in the average premiums between male and female customers.

Poll Time

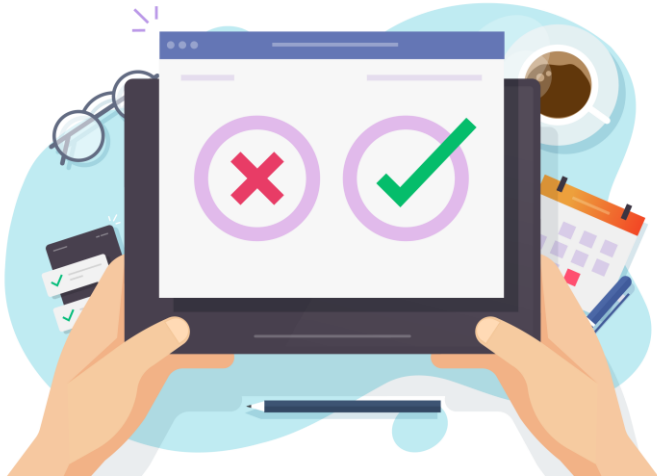
Q. Which measure of spread provides a measure of variability within the central portion of the data?

a. Range

b. IQR

c. Variance

d. Standard Deviation



Poll Time

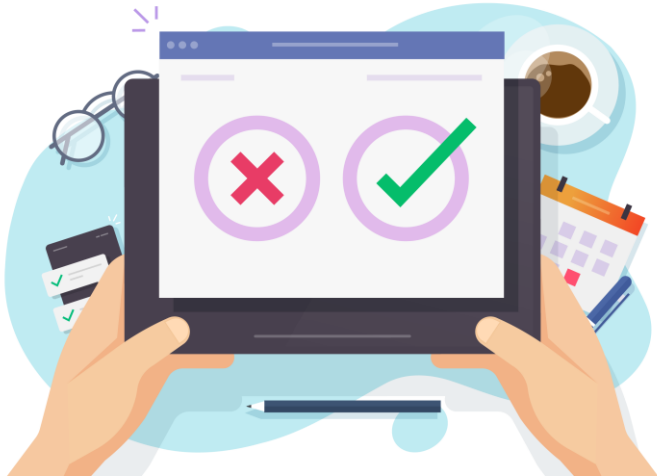
Q. Which measure of spread provides a measure of variability within the central portion of the data?

a. Range

b. IQR

c. Variance

d. Standard Deviation



Q&A

Other Functions in Excel

Introduction to Dependent Dropdown List

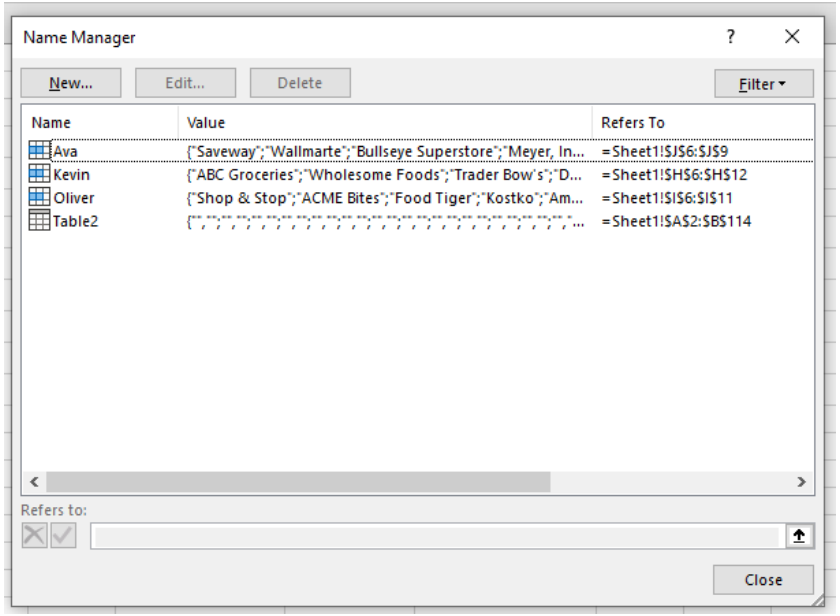
- Dependent dropdown lists in Excel allow you to create a dynamic selection process based on the values chosen in a previous dropdown.
- This is useful when you have related or hierarchical data that you want to organize and select from.

Creating a Dependent Dropdown List

Step 1: Set up the Data

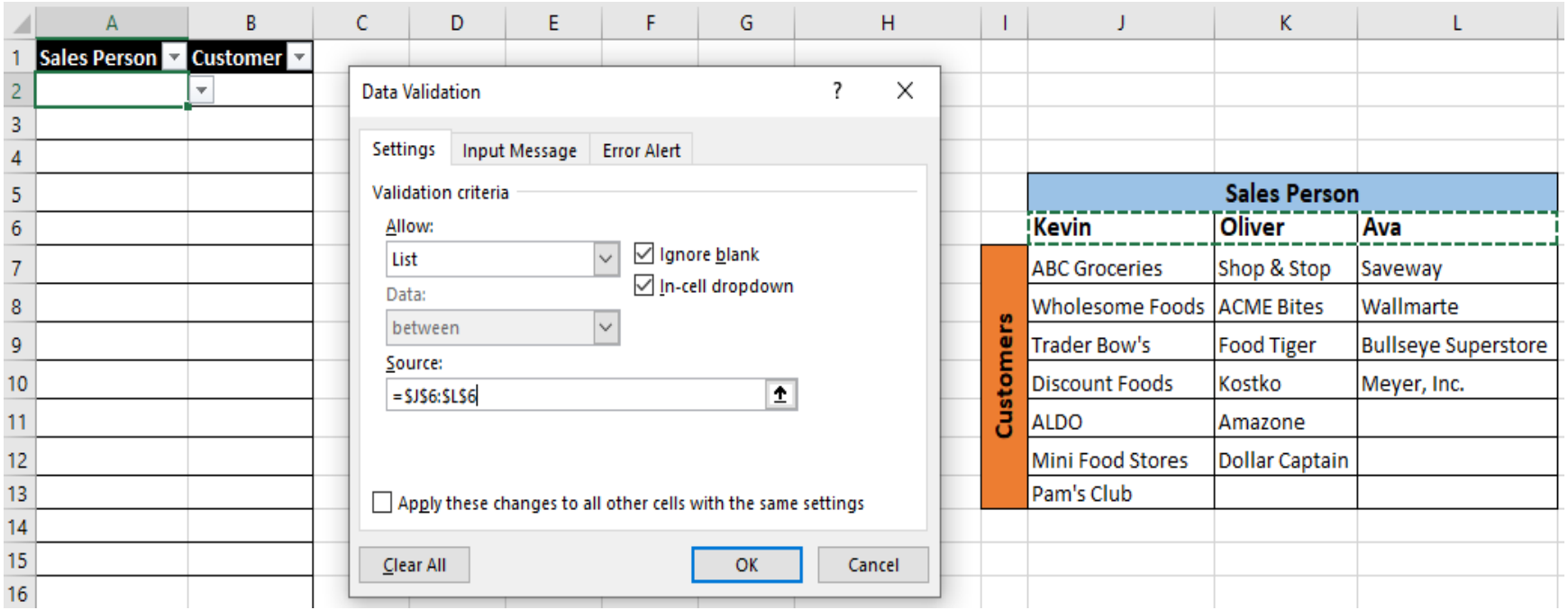
	Sales Person		
	Kevin	**Oliver**	**Ava**
Customers	ABC Groceries	Shop & Stop	Saveway
	Wholesome Foods	ACME Bites	Wallmarte
	Trader Bow's	Food Tiger	Bullseye Superstore
	Discount Foods	Kostko	Meyer, Inc.
	ALDO	Amazone	
	Mini Food Stores	Dollar Captain	
	Pam's Club		

Step 2: Create Name Ranges



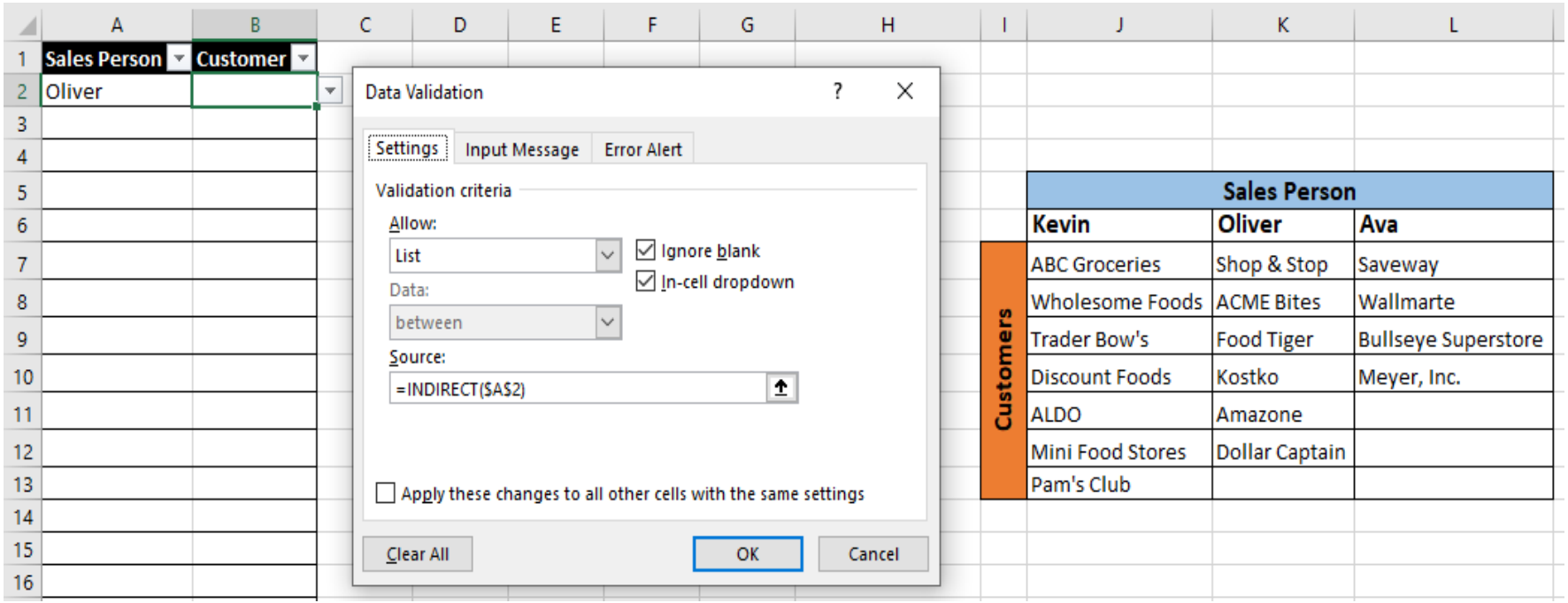
Creating a Dependent Dropdown List

Step 3: Create the First Dropdown



Creating a Dependent Dropdown List

Step 4: Create the Second Dropdown



Poll Time

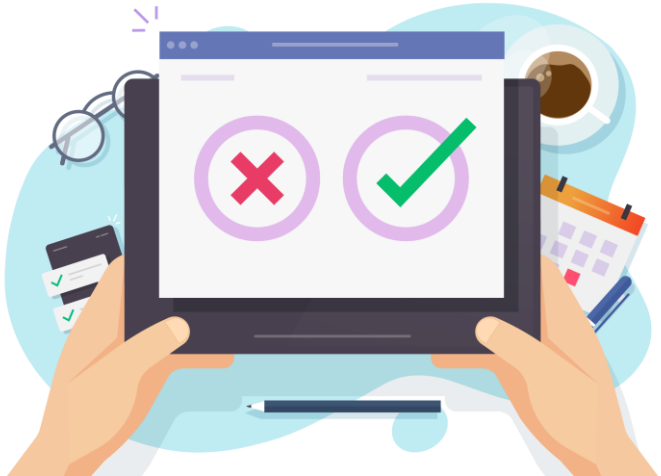
Q. Which of the below function is used to create a dependent dropdown list?

a. Unique()

b. Concatenate()

c. Direct()

d. Indirect()



Poll Time

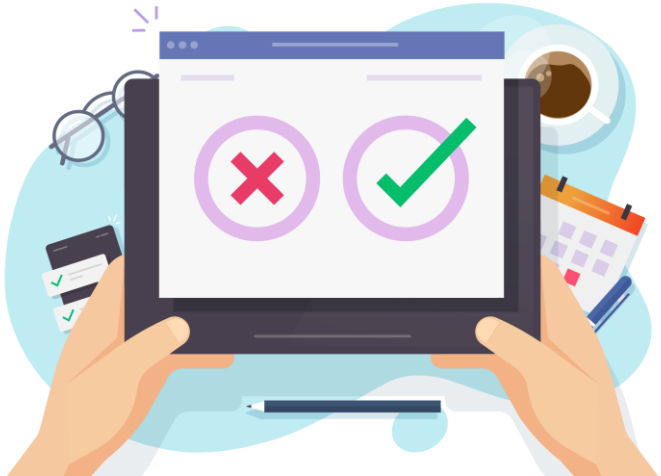
Q. Which of the below function is used to create a dependent dropdown list?

a. Unique()

b. Concatenate()

c. Direct()

d. Indirect()



Introduction to Filters in Excel

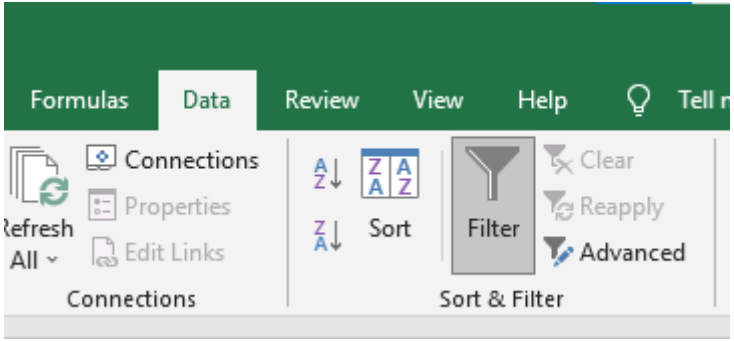
- Filtering in Excel allows you to quickly analyze and manipulate data by displaying only the information you need.
- Filters help you sort and narrow down data based on specific criteria.

Using Filters

Step 1: Select the Data

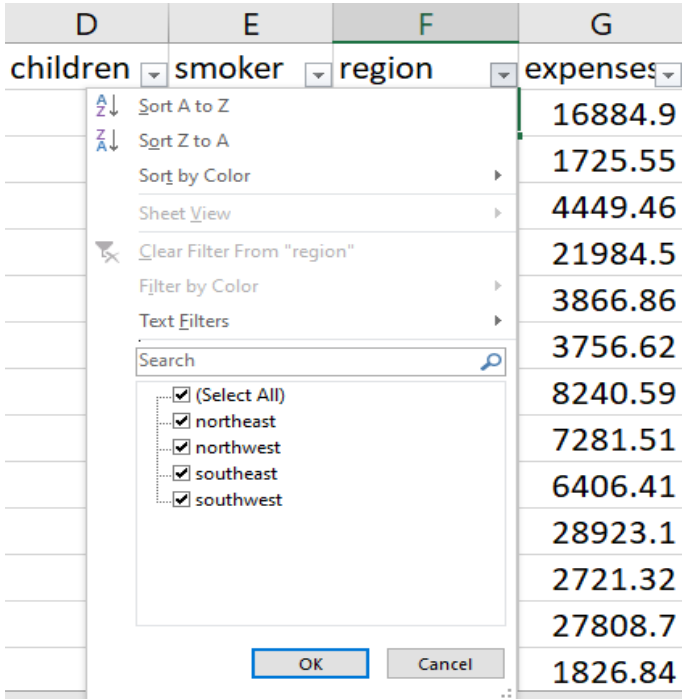
	A	B	C	D	E	F	G
1	age	sex	bmi	children	smoker	region	expenses
2	19	female	27.9	0	yes	southwest	16884.9
3	18	male	33.8	1	no	southeast	1725.55
4	28	male	33	3	no	southeast	4449.46
5	33	male	22.7	0	no	northwest	21984.5
6	32	male	28.9	0	no	northwest	3866.86
7	31	female	25.7	0	no	southeast	3756.62
8	46	female	33.4	1	no	southeast	8240.59
9	37	female	27.7	3	no	northwest	7281.51
10	37	male	29.8	2	no	northeast	6406.41
11	60	female	25.8	0	no	northwest	28923.1

Step 2: Apply the Filter



Using Filters

Step 3: Filter the column



Poll Time

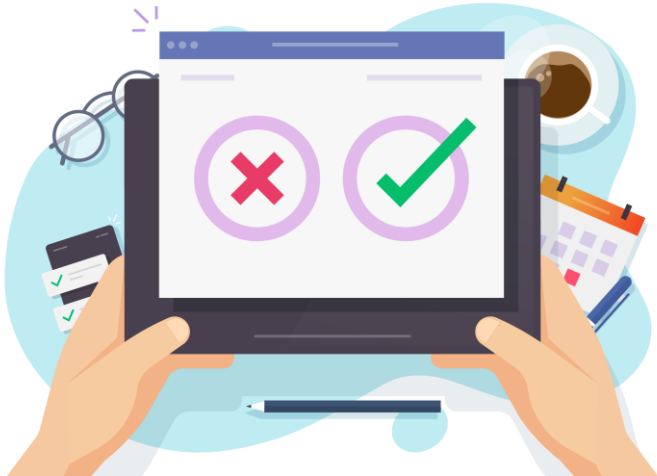
Q. Which of the below options are incorrect about filtering?

a. Filter option is available in the Data Tab

b. Filter option is available in the Home Tab

c. Filter option is available in the Formula Tab

d. Converting raw data to tables sets filters to each column automatically



Poll Time

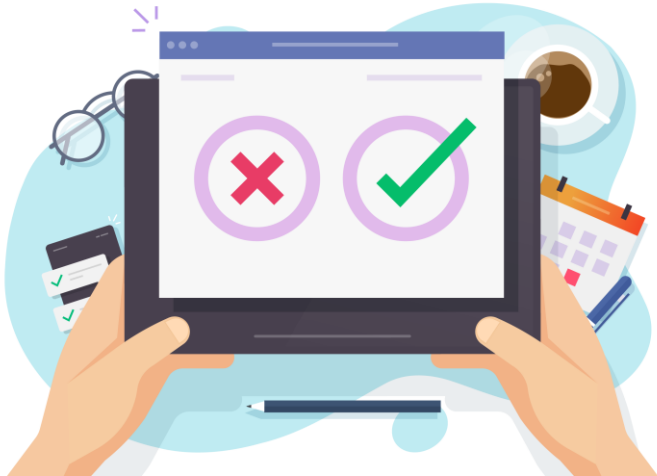
Q. Which of the below options are incorrect about filtering?

a. Filter option is available in the Data Tab

b. Filter option is available in the Home Tab

c. Filter option is available in the Formula Tab

d. Converting raw data to tables sets filters to each column automatically



Pivot Tables in Excel

Introduction to Pivot Tables in Excel

- Pivot tables in Excel allow you to summarize and analyze large amounts of data quickly and efficiently.
- They provide a flexible and powerful tool for data analysis and reporting.

Example of Pivot Table Analysis

Smoking Rate across Age Groups

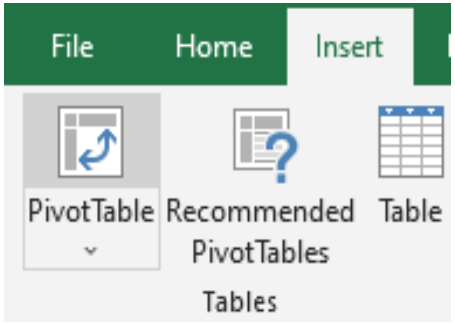
Count of People	Smoker		
Age	no	yes	Grand Total
18	57	12	69
19	50	18	68
20	20	9	29
21	26	2	28
22	22	6	28
23	21	7	28
24	22	6	28
25	23	5	28
26	25	3	28
27	19	9	28
28	25	3	28
29	21	6	27
30	18	9	27
31	22	5	27
32	21	5	26
33	20	6	26
34	21	5	26
35	20	5	25

How to Perform Pivot Table Analysis

Step 1: Select the Data

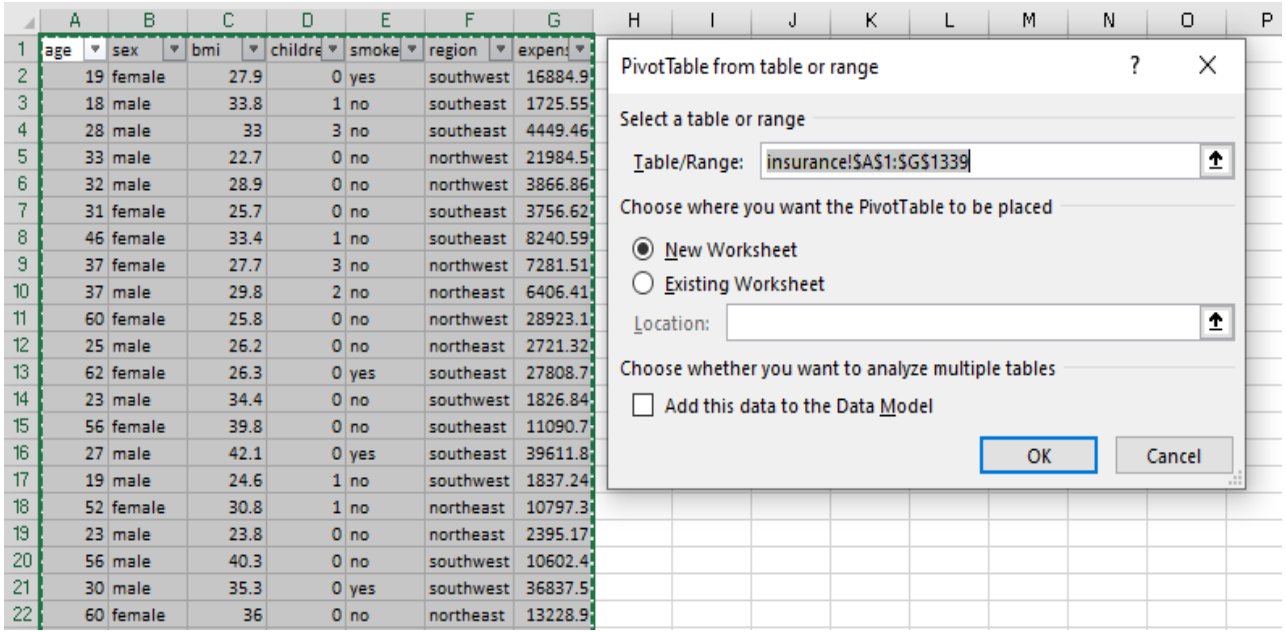
	A	B	C	D	E	F	G
1	age	sex	bmi	children	smoker	region	expenses
2	19	female	27.9	0	yes	southwest	16884.9
3	18	male	33.8	1	no	southeast	1725.55
4	28	male	33	3	no	southeast	4449.46
5	33	male	22.7	0	no	northwest	21984.5
6	32	male	28.9	0	no	northwest	3866.86
7	31	female	25.7	0	no	southeast	3756.62
8	46	female	33.4	1	no	southeast	8240.59
9	37	female	27.7	3	no	northwest	7281.51
10	37	male	29.8	2	no	northeast	6406.41
11	60	female	25.8	0	no	northwest	28923.1

Step 2: Create a Pivot Table



How to Perform Pivot Table Analysis

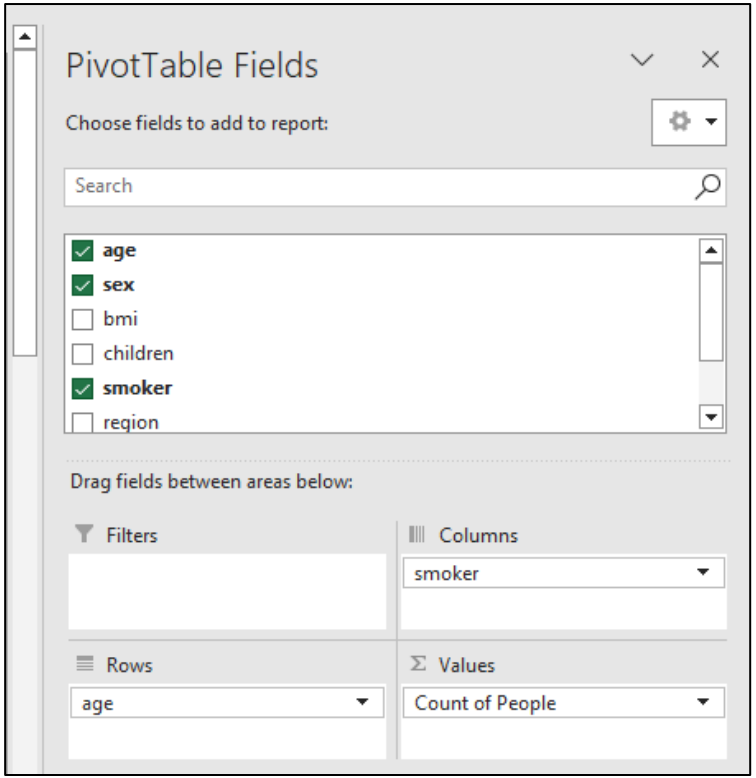
Step 3: Choose the Destination



	age	sex	bmi	childre	smoke	region	expens
2	19	female	27.9	0	yes	southwest	16884.9
3	18	male	33.8	1	no	southeast	1725.55
4	28	male	33	3	no	southeast	4449.46
5	33	male	22.7	0	no	northwest	21984.5
6	32	male	28.9	0	no	northwest	3866.86
7	31	female	25.7	0	no	southeast	3756.62
8	46	female	33.4	1	no	southeast	8240.59
9	37	female	27.7	3	no	northwest	7281.51
10	37	male	29.8	2	no	northeast	6406.41
11	60	female	25.8	0	no	northwest	28923.1
12	25	male	26.2	0	no	northeast	2721.32
13	62	female	26.3	0	yes	southeast	27808.7
14	23	male	34.4	0	no	southwest	1826.84
15	56	female	39.8	0	no	southeast	11090.7
16	27	male	42.1	0	yes	southeast	39611.8
17	19	male	24.6	1	no	southwest	1837.24
18	52	female	30.8	1	no	northeast	10797.3
19	23	male	23.8	0	no	northeast	2395.17
20	56	male	40.3	0	no	southwest	10602.4
21	30	male	35.3	0	yes	southwest	36837.5
22	60	female	36	0	no	northeast	13228.9

How to Perform Pivot Table Analysis

Step 4: Designing the Pivot Table



Step 6: Customize the Pivot Table

Count of People	Smoker		
Age	no	yes	Grand Total
18	57	12	69
19	50	18	68
20	20	9	29
21	26	2	28
22	22	6	28
23	21	7	28
24	22	6	28
25	23	5	28
26	25	3	28
27	19	9	28
28	25	3	28
29	21	6	27
30	18	9	27
31	22	5	27
32	21	5	26
33	20	6	26
34	21	5	26
35	20	5	25

Hands-on: Conducting Data Analysis in Excel

Poll Time

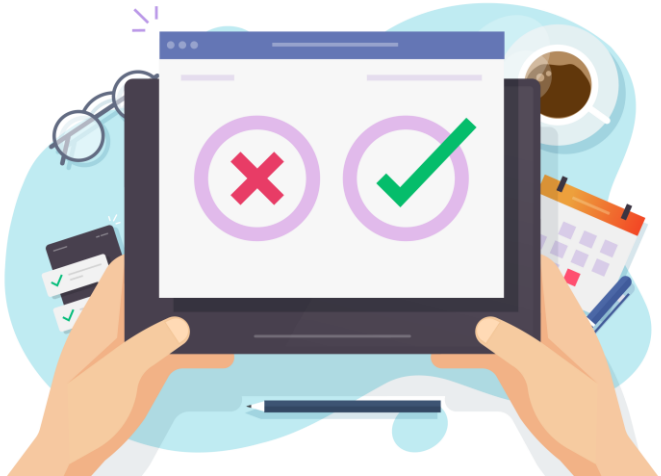
Q. Which of the following shelves are not a part of the Pivot Table window?

a. Cell

b. Column

c. Value

d. Row



Poll Time

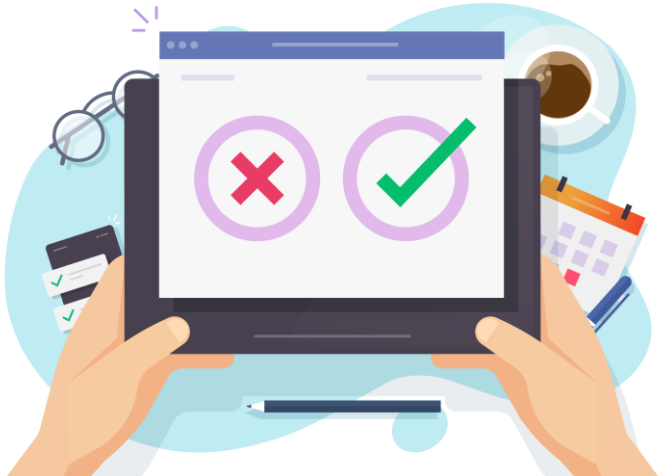
Q. Which of the following shelves are not a part of the Pivot Table window?

a. **Cell**

b. Column

c. Value

d. Row



Summary

- Central tendency summarizes the center or average value of a dataset.
- Measures of Spread quantify the variability or spread of values in a dataset.
- Dependent dropdown lists are useful when you have related or hierarchical data that you want to organize and select from.
- Filters help you sort, and narrow down data based on specific criteria.
- Pivot Tables allow you to summarize and analyze large amounts of data quickly and efficiently.

Activity 1

Pre-requisites:

- MS Excel
- WorkOrders.xlsx

Scenario:

You have been given an Excel file containing details for several work orders taken up by a company. Perform the following analysis on the given dataset:

- Create a new table that contains the valid number of technicians allowed for each service type, as shown below.

Serive Type				
Assess	Deliver	Install	Repair	Replace
1	1	1	1	1
2	2	2	2	2
		3	3	

Activity 2

Pre-requisites:

- MS Excel
- WorkOrders.xlsx

Scenario:

You have been given an Excel file containing details for several work orders taken up by a company. Perform the following analysis on the given dataset:

- Using the table created in the previous activity, create named ranges for each service type such that each range contains the respective technician counts.

Activity 3

Pre-requisites:

- MS Excel
- WorkOrders.xlsx

Scenario:

You have been given an Excel file containing details for several work orders taken up by a company. Perform the following analysis on the given dataset:

- Create a data validation over the existing “Techs” column such that the value in each row is as per the structure defined in the table created in activity 1.

Activity 4

Pre-requisites:

- MS Excel
- WorkOrders.xlsx

Scenario:

You have been given an Excel file containing details for several work orders taken up by a company. Perform the following analysis on the given dataset:

- Create a pivot table that contains the total revenue and count of services done across all service categories for various technician counts.

Q&A

Next Session:
Mastering Advanced Excel: Macros, Data Validation, and Dashboard Creation

THANK YOU

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

knowledgehut
upGrad

knowledgehut
upGrad

Mastering Advanced Excel: Macros, Data Validation, and Dashboard Creation

Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.

By the End of This Session:

- You shall automate repetitive Excel tasks using macros.
- You shall learn the fundamentals of VBA to modify, debug and write macros.
- You will set constraints on your raw data.
- You will create informative dashboards to summarize your analysis.

What Have We Learned So Far?

- We’ve understood the fundamentals of Statistical Analysis.
- We’ve learned the various measures of central tendency and the measures of spread.
- We’ve learned the use of dependent dropdown lists.
- We’ve learned the use of filters in Excel.
- We’ve learned to create and use pivot tables to analyze raw data.

Poll Time

Q. Which of the following options accurately describes dependent dropdown lists in Excel?

a. Dropdown lists that rely on external data sources

b. Dropdown lists that automatically update based on the user's input

c. Dropdown lists that dynamically change based on the selection in another dropdown

d. Dropdown lists that display values randomly selected from a list

Poll Time

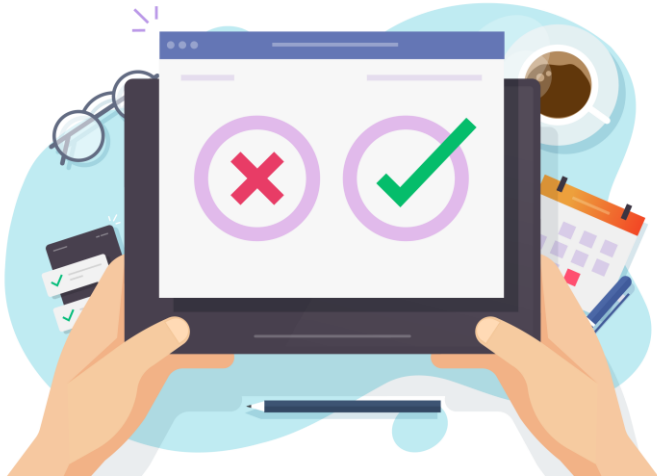
Q. Which of the following options accurately describes dependent dropdown lists in Excel?

a. Dropdown lists that rely on external data sources

b. Dropdown lists that automatically update based on the user's input

c. Dropdown lists that dynamically change based on the selection in another dropdown

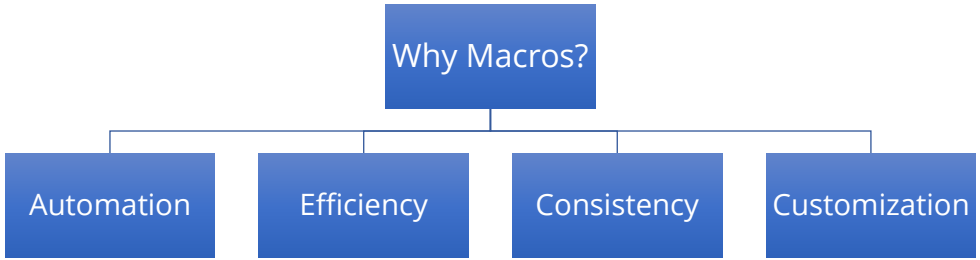
d. Dropdown lists that display values randomly selected from a list



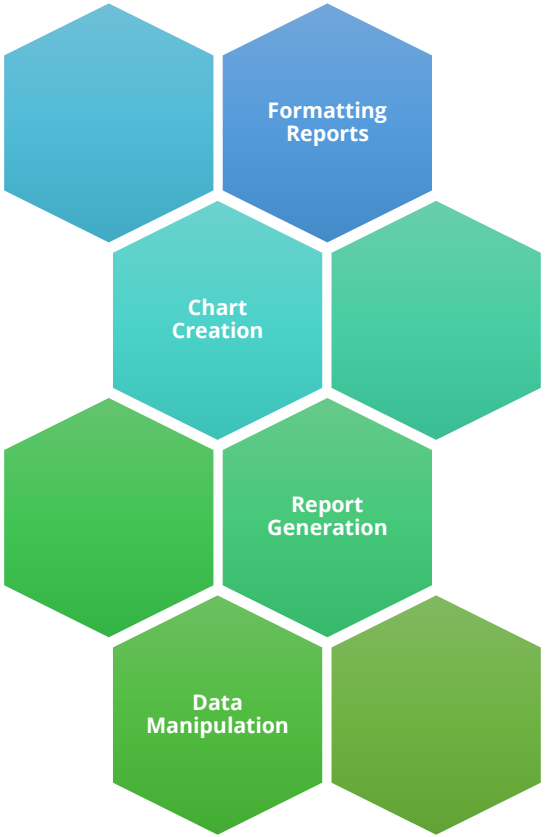
Macros and VBA

Introduction to Macros and VBA

- A macro is a set of instructions or code that automates tasks in Excel.
- It allows you to record and replay a series of actions, making repetitive tasks faster and more efficient.
- Macros can be created using the built-in Macro Recorder or by writing VBA code (Visual Basic for Applications).

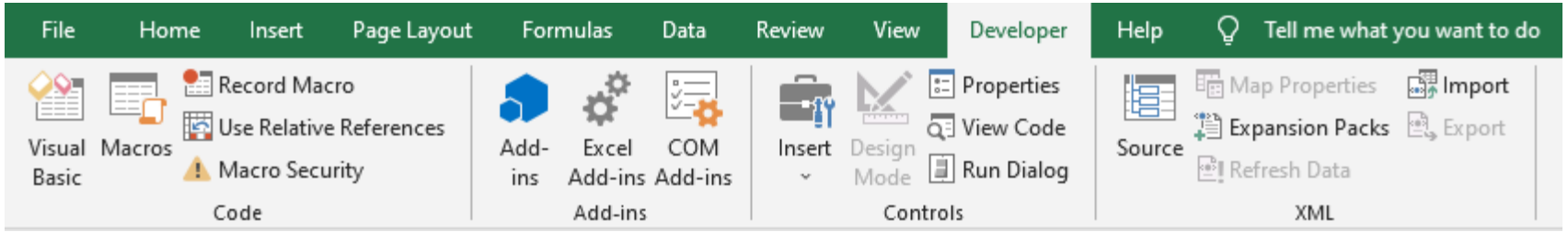


Use Case of Macros



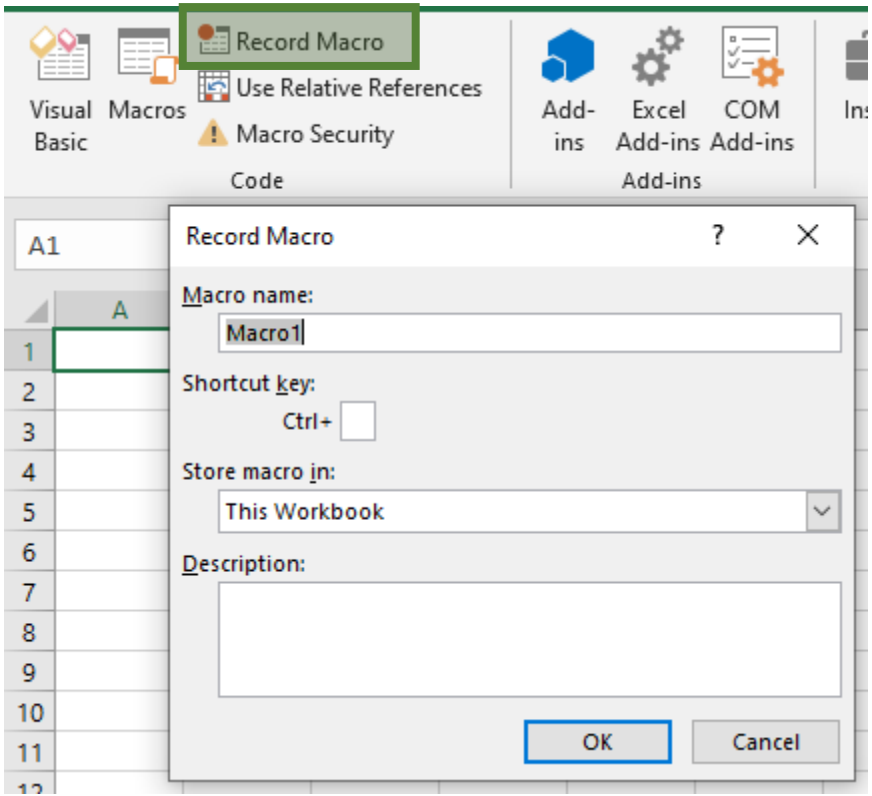
Recording Macros

Step 1: Enable the Developer Tab



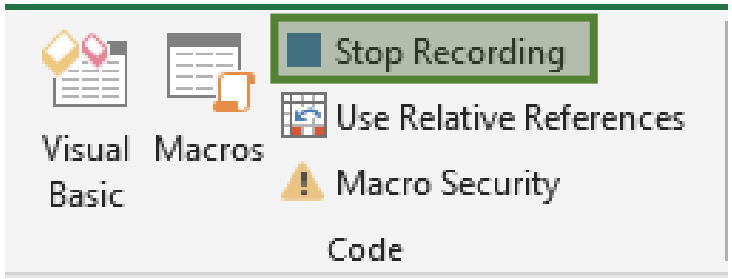
Recording Macros

Step 2: Click on Record Macro button



Recording Macros

Step 3: Click on Stop Recording once done



Pop Quiz

Q. Which of the following macro techniques in Excel allows you to write and customize VBA code for creating advanced automation and functionality?

a. Macro Recording

b. Macro Execution

c. Macro Debugging

d. Custom Macro Creation

Pop Quiz

Q. Which of the following macro techniques in Excel allows you to write and customize VBA code for creating advanced automation and functionality?

a. **Macro Recording**

b. Macro Execution

c. Macro Debugging

d. Custom Macro Creation

Hands-on: Writing and Debugging a Simple VBA Code

Q&A

Advanced Data Validation

Introduction to Advanced Data Validation

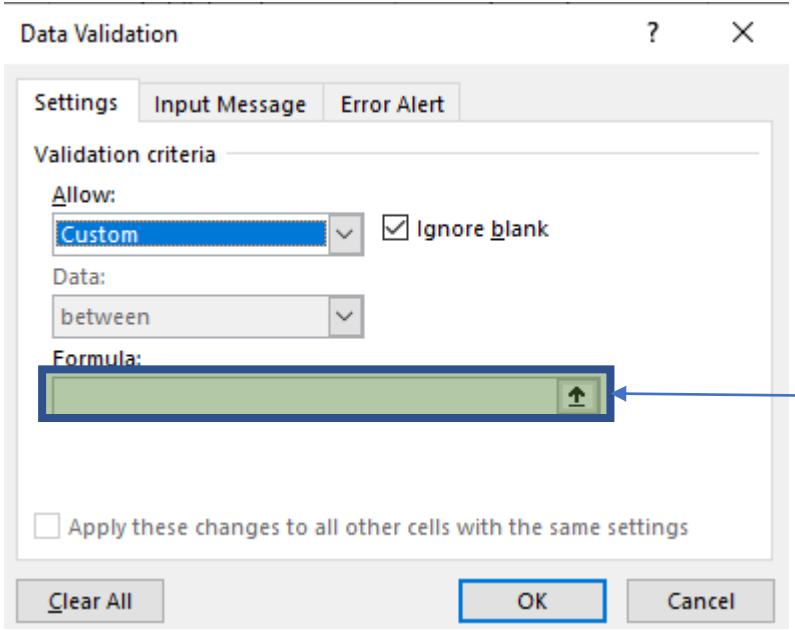
- Data validation is a feature in Excel that helps control and restrict the type of data that can be entered into a cell.
- It ensures data accuracy, consistency, and integrity by setting specific rules and criteria for data entry.

Creating dependent dropdown lists using data validation.

Using formulas in data validation rules for dynamic validation.

Applying data validation across multiple worksheets or workbooks.

Using Formulas in Data Validation Rules



Insert the custom formula in the “Formula” box.

Pop Quiz

Q. Which of the following Excel formulas can be used for creating formula-based data validation?

a. =AVERAGE()

b. =VLOOKUP()

c. =COUNTIF()

d. =SUM()

Pop Quiz

Q. Which of the following Excel formulas can be used for creating formula-based data validation?

a. =AVERAGE()

b. =VLOOKUP()

c. =COUNTIF()

d. =SUM()

Hands-on - Using Formulas in Validation Rules

Poll Time

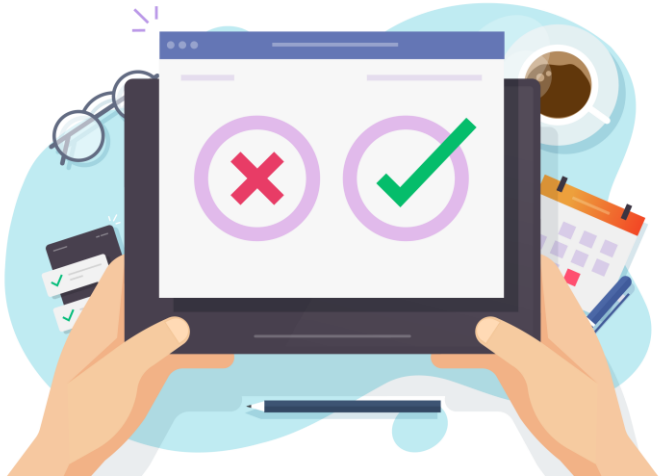
Q. Which of the following statements accurately describes the purpose of data validation in Excel?

a. Data validation ensures data encryption and security in Excel files

b. Data validation helps in generating random data for statistical analysis in Excel

c. Data validation restricts and validates the type and values entered into cells in Excel

d. Data validation automatically calculates and summarizes data in Excel spreadsheets



Poll Time

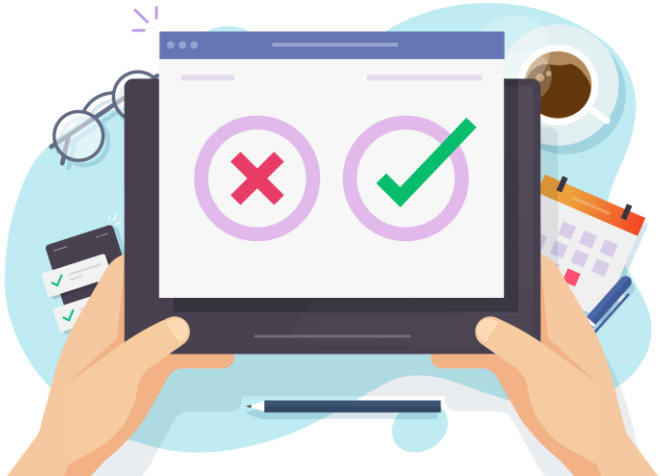
Q. Which of the following statements accurately describes the purpose of data validation in Excel?

a. Data validation ensures data encryption and security in Excel files

b. Data validation helps in generating random data for statistical analysis in Excel

c. Data validation restricts and validates the type and values entered into cells in Excel

d. Data validation automatically calculates and summarizes data in Excel spreadsheets



Exercise

In the “Work Order” dataset, create a formula-based data validation in the PartsCost column to allow only positive values in the column.

PartsCost
$90.04
$35.10
$832.16
$938.00
$240.00
$120.00
$341.00
$155.39
$204.28
$535.62
$43.26

Poll Time

Q. Which of the following Excel formulas is used to count the number of cells in a range that meet criteria?

a. =COUNTIF()

b. =COUNT()

c. =SUMIF()

d. =SUM()

Poll Time

Q. Which of the following Excel formulas is used to count the number of cells in a range that meet criteria?

a. **=COUNTIF()**

b. =COUNT()

c. =SUMIF()

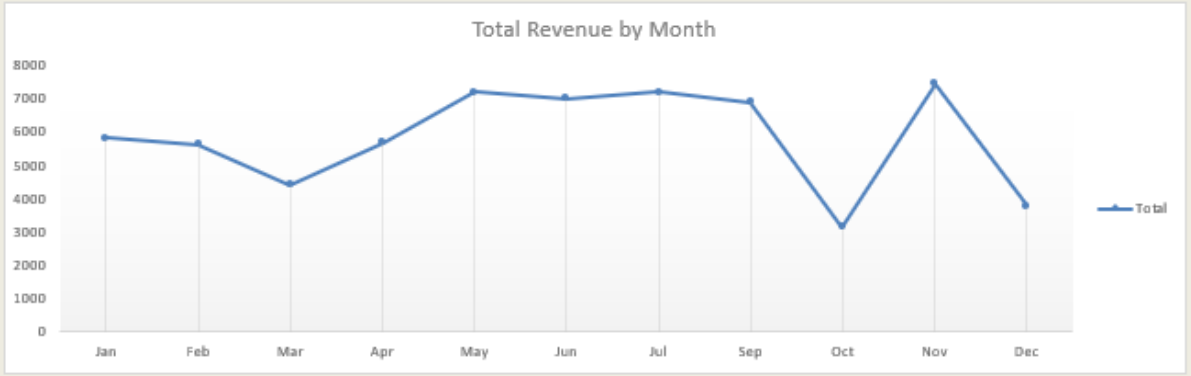
d. =SUM()

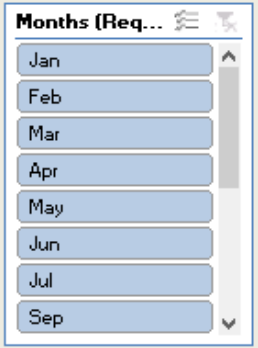
Introduction to Excel Dashboards

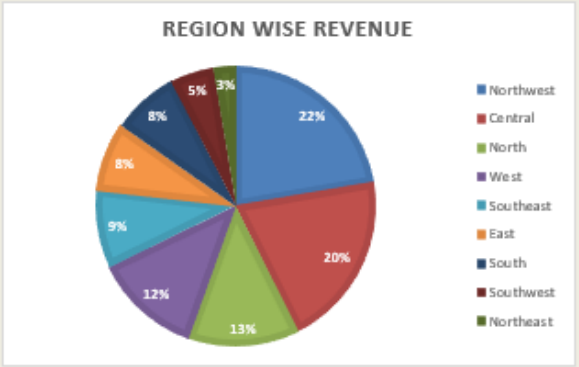
Introduction to Excel Dashboards

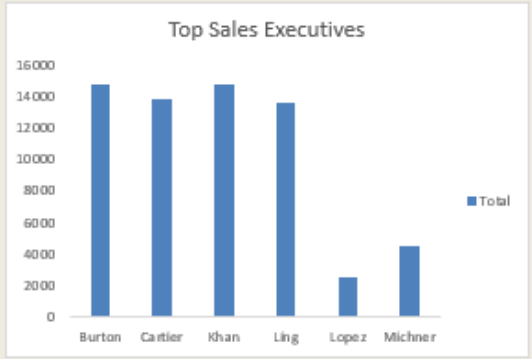
- A dashboard is a visual representation of data that provides a comprehensive overview of key information.
- In Excel, a dashboard is created by consolidating and presenting data in a visually appealing and interactive format.

Business Use Case - Dashboard









Pop Quiz

Q. When creating dashboards in Excel, what is the primary purpose of incorporating data visualizations?

a. To make the dashboard visually appealing

b. To showcase advanced Excel skills

c. To present data in a concise and meaningful way

d. To add interactivity and dynamic features

Pop Quiz

Q. When creating dashboards in Excel, what is the primary purpose of incorporating data visualizations?

a. To make the dashboard visually appealing

b. To showcase advanced Excel skills

c. To present data in a concise and meaningful way

d. To add interactivity and dynamic features

Hands-on – Building Dashboards

Pop Quiz

Q. Which of the following data validation techniques in Excel allows you to create a dropdown list of options based on values from another range?

a. Whole number validation

b. Text length validation

c. List validation

d. Date/Time validation

Pop Quiz

Q. Which of the following data validation techniques in Excel allows you to create a dropdown list of options based on values from another range?

a. Whole number validation

b. Text length validation

c. List validation

d. Date/Time validation

Summary

- We understood the use of Macros and VBA in automating repetitive tasks in excel.
- We learnt some advanced data validation techniques such as dependent dropdown lists and formula-based data validation.
- We created a dashboard to sum up our analysis and present it in a concise manner.

Activity 1

Pre-requisites:

- MS Excel
- Insurance.xlsx

Scenario:

You are given an Excel file containing insurance premiums for multiple customers.

Record a macro while performing the following analysis on the dataset:

- Create a new column called **Total Insured,** which contains the total number of people covered under each policy.
- Calculate the premium paid per head by dividing the **Expense** column and the **Total Insured** column.

Activity 2

Pre-requisites:

- MS Excel
- Insurance.xlsx

Scenario:

You are given an Excel file containing insurance premiums for multiple customers.

Create the below-mentioned data validation on the respective columns:

- Ensure that the Age of customers lies between 18 – 80.
- Ensure that the gender column contains one of **male**, **female,** and **others**.
- Ensure that the BMI and expenses columns contain positive values.
- Ensure that the smoker column contains either a **yes** or a **no**.

Activity 3

Pre-requisites:

- MS Excel
- Insurance.xlsx

Scenario:

You are given an Excel file containing insurance premiums for multiple customers.

Create a dashboard that contains the following charts:

- Line chart depicting the relation between the average premium per head and age.
- Bar chart containing the fraction of male and female population who smoke.
- Pie chart depicting the count of people who smoke in each region.

Q&A

Session Feedback

Next Session:
Excel Case Study

THANK YOU

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

knowledgehut
upGrad