Excel Theoretical Question

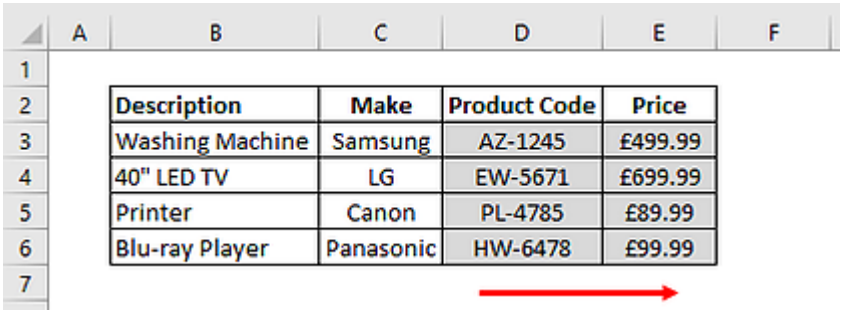
1. What is vlookup function? What are its limitations VLOOKUP is an Excel function to look up data in a table organized vertically?

Ans:- VLOOKUP is a function to lookup up and retrieve data in a table. The "V" in VLOOKUP stands for vertical, which means the data in the table must be arranged vertically, with data in rows.

Limitations:-

(1) VLOOKUP Only Looks to the Right:-

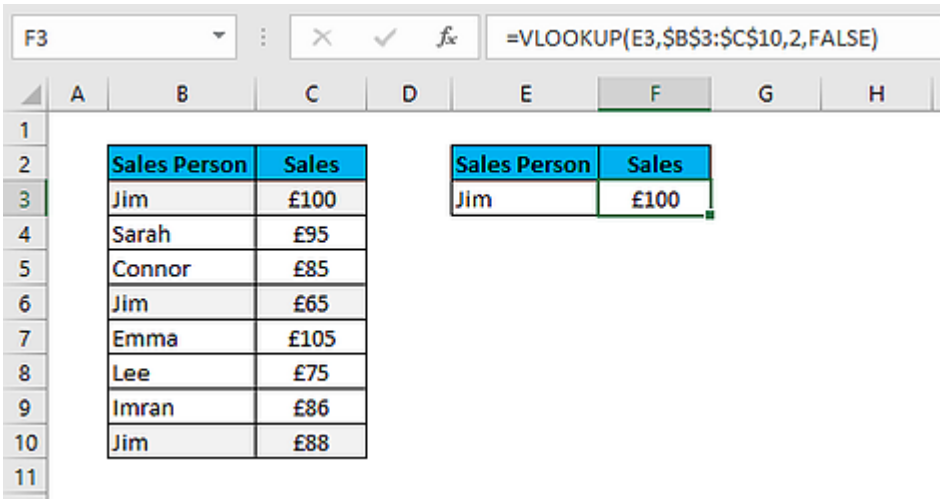
One of the biggest limitations is that the VLOOKUP function can only look at the columns to the right of the lookup value.



Description	Make	Product Code	Price
Washing Machine	Samsung	AZ-1245	£499.99
40" LED TV	LG	EW-5671	£699.99
Printer	Canon	PL-4785	£89.99
Blu-ray Player	Panasonic	HW-6478	£99.99

(2) VLOOKUP Only Finds the First Match:-

If the lookup value column contains duplicate values then VLOOKUP will only extract the first value.

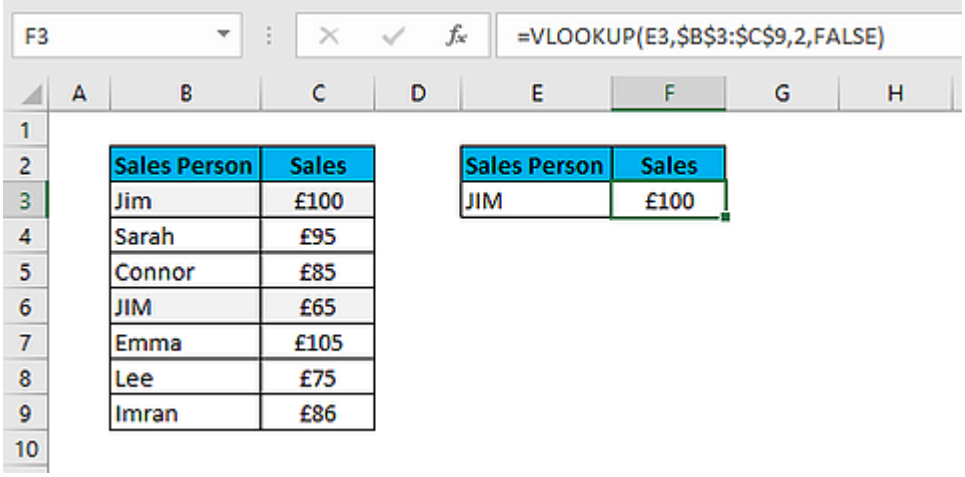


Sales Person	Sales
Jim	£100
Sarah	£95
Connor	£85
Jim	£65
Emma	£105
Lee	£75
Imran	£86
Jim	£88

Sales Person	Sales
Jim	£100

(3) VLOOKUP is not Case Sensitive:-

VLOOKUP does not distinguish between lower and uppercase values. It sees them as the same.



Q-2 What is the importance of index match functions?

Ans:- The INDEX function in Excel is very powerful at the same time a flexible tool that retrieves the value at a given location in a range. In another word, It returns the content of a cell, specified by row and column offset.

Syntax: `=INDEX (reference, [row], [column])`

Q-3 Explain countif() function.

Ans:- COUNTIF is an Excel function to count cells in a range that meet a single condition. COUNTIF can be used to count cells that contain dates, numbers, and text

Q-4 Difference between count and counta() functions.

Ans:-

COUNT	COUNTA
The COUNT function counts the number of cells within a specified range that has numeric values.	The COUNTA function counts the number of cells within a specified range that are non-empty.
The COUNT function only counts dates and numeric entries.	The COUNTA function counts anything as long as a cell has something, regardless of the data types.
Syntax: = COUNT (value1, value2,...)	Syntax: = COUNTA (value1, value2,...)

Q5 What are slicers and how can be used?

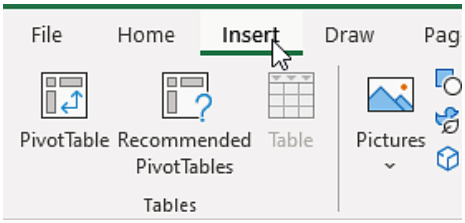
Ans:- Slicers in Excel are software filters used along with excel tables or pivot tables over a large amount of data. Not just filtering out the data, but slicers also help you with an easy understanding of the information being extracted and displayed on the screen.

Steps to use slicers:

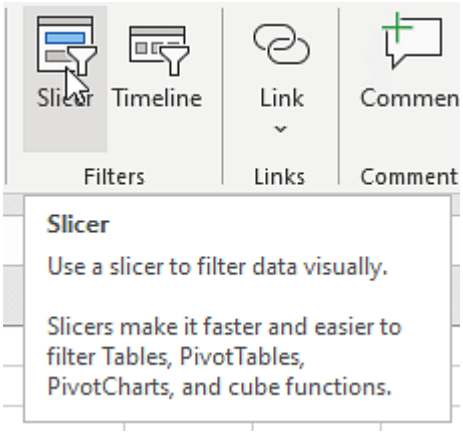
-Select any cell of the Excel table or pivot table and insert slicers.

Now insert the slicers. Inserting slicers is completely the same as inserting slicers in a pivot table.

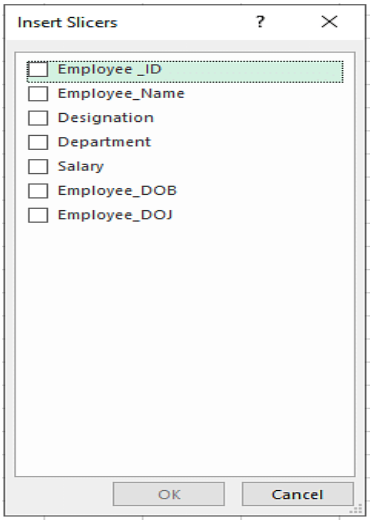
- Select any cell in the table
- Navigate to the insert option on the toolbar



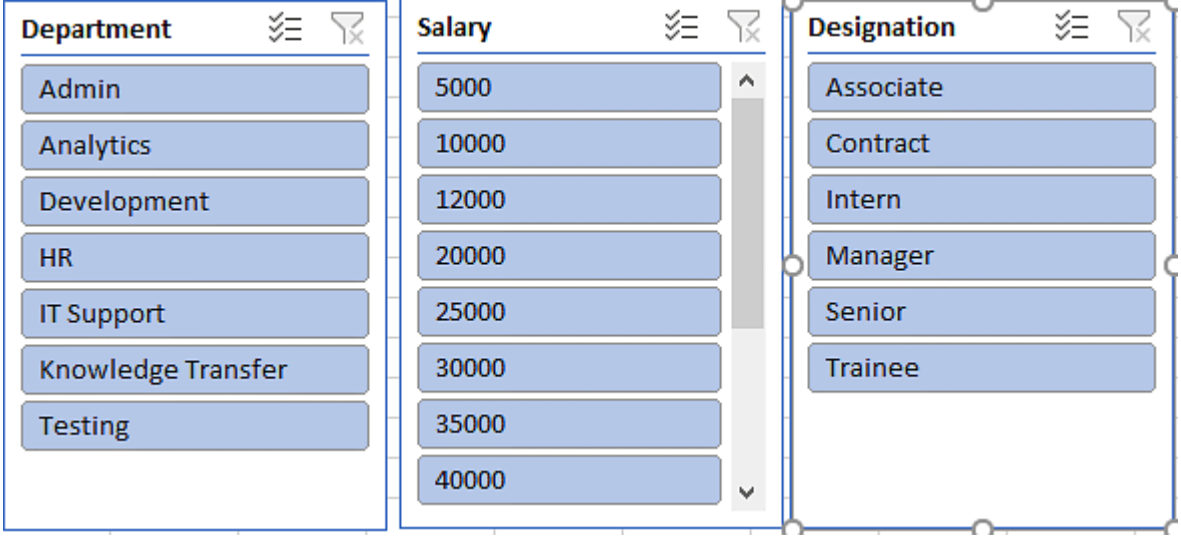
- Select insert slicer



- A new window will appear asking for parameters. Select parameters and press ok



- The slicers will be available on the screen and ready for operations



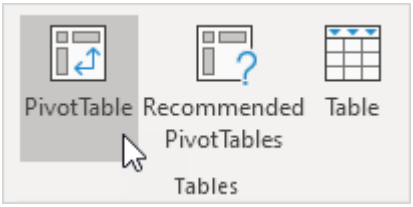
Q6 What is a pivot table? How to create one?

Ans:- A pivot table is a statistics tool that summarizes and reorganizes selected columns and rows of data in a spreasheet or database table to obtain a desired report. The tool does not actually change the spreadsheet or database itself, it simply "pivots" or turns the data to view it from different perspectives.

Insert a Pivot Table:-

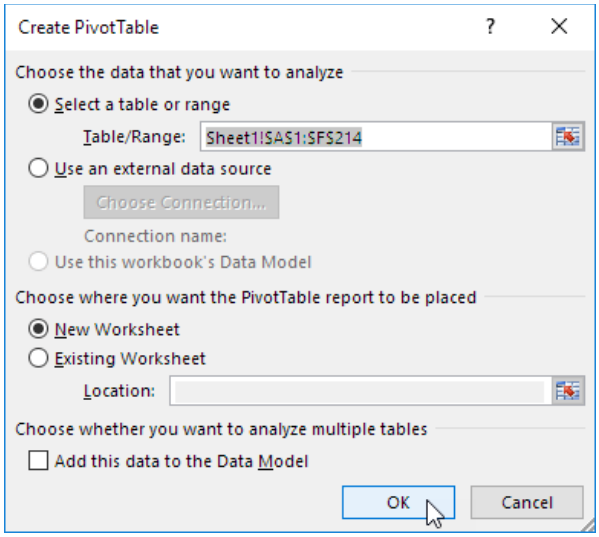
1. Click any single cell inside the data set.

2. On the Insert tab, in the Tables group, click PivotTable.



The following dialog box appears. Excel automatically selects the data for you. The default location for a new pivot table is New Worksheet.

3. Click OK.



Q7 What is the shortcut to add a filter to a table?

Ans:- **Ctrl+Shift+L**

Q8 How does the IF() function in Excel work?

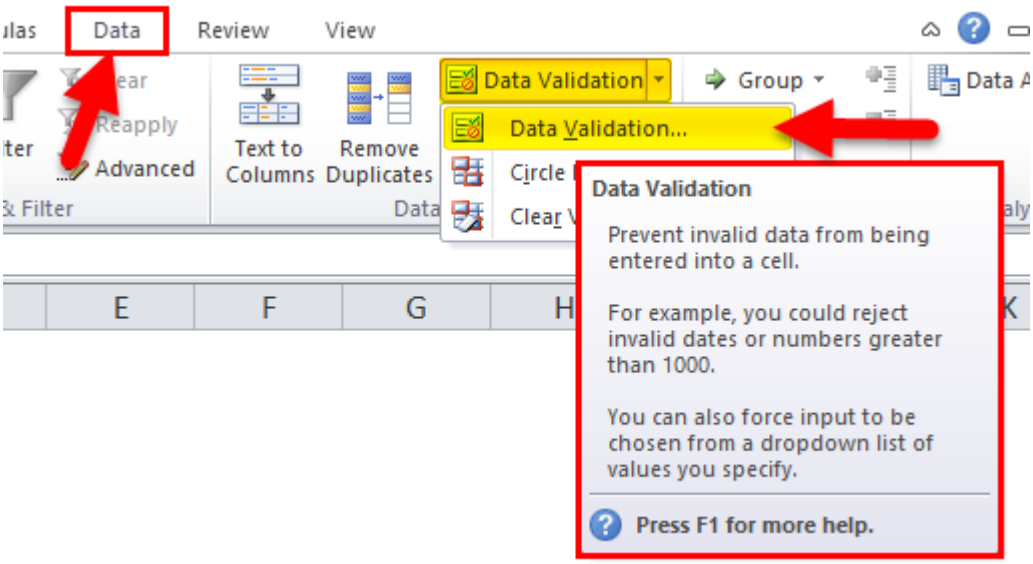
Ans:- The IF Function is an important function for comparing data, performing lookups to group data, and making reports interactive.

=IF (logical_test, [value_if_true], [value_if_false]

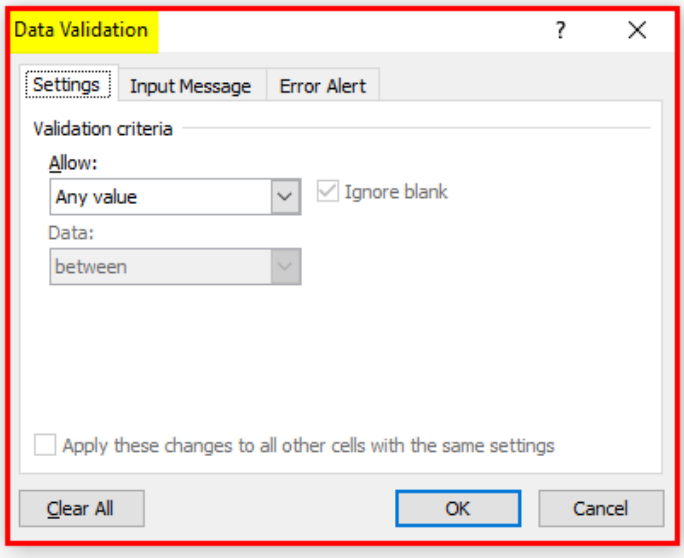
Q9 How to create a drop-down list in Excel?

Ans:-

- Click on the data menu and choose the data validation menu

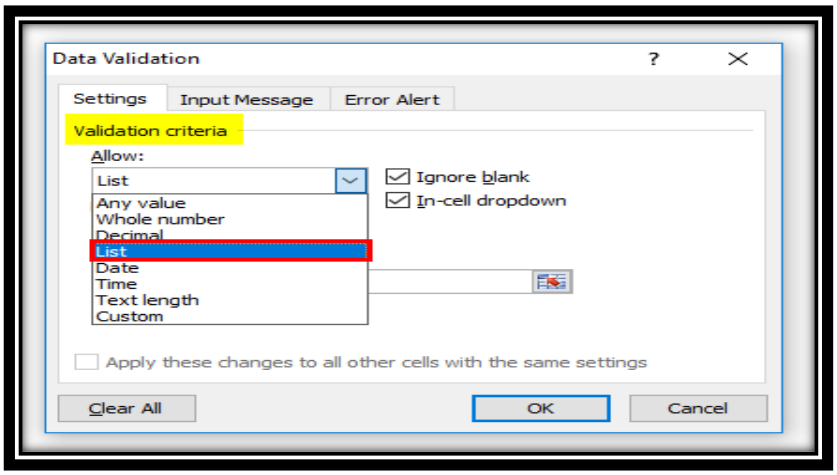


- Click on the Data Validation TAB so that we will get a dialogue box.



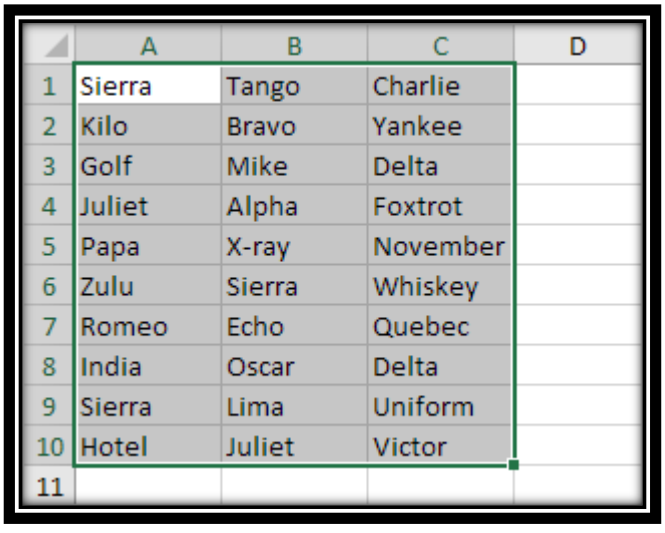
Now we can see the Validation criteria part where it has to allow option click on the dropbox where it will list the several validation parts like the whole number, decimal, etc.

- From the validation, criteria choose the List option to create a Excel Drop Down List, which will create and ask for selecting a cell to display the Drop Down List in Excel.



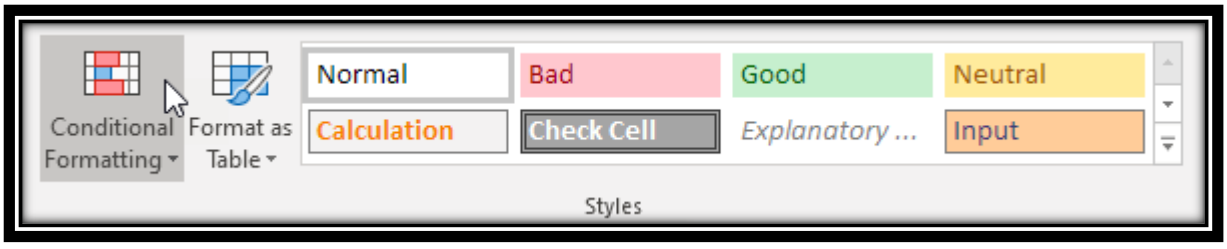
Q10 How do you find duplicate values in a column?

Ans:- 1. Select the range A1:C10.

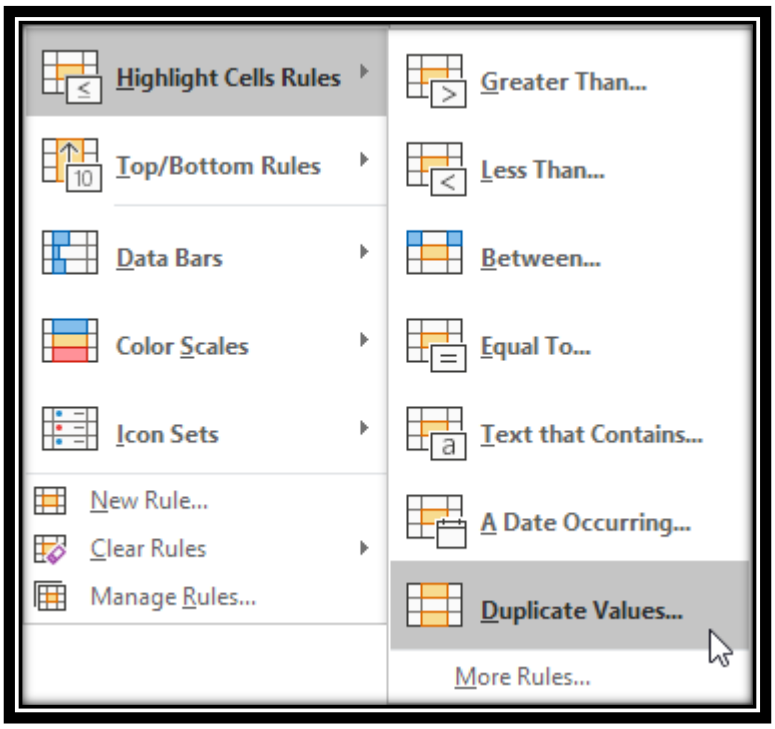


	A	B	C	D
1	Sierra	Tango	Charlie	
2	Kilo	Bravo	Yankee	
3	Golf	Mike	Delta	
4	Juliet	Alpha	Foxtrot	
5	Papa	X-ray	November	
6	Zulu	Sierra	Whiskey	
7	Romeo	Echo	Quebec	
8	India	Oscar	Delta	
9	Sierra	Lima	Uniform	
10	Hotel	Juliet	Victor	
11				

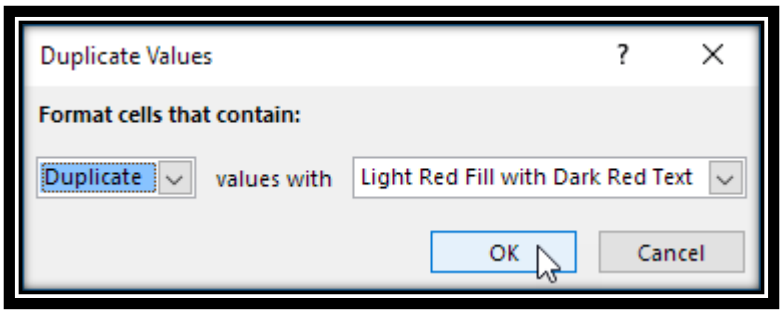
2. On the Home tab, in the Styles group, click Conditional Formatting.



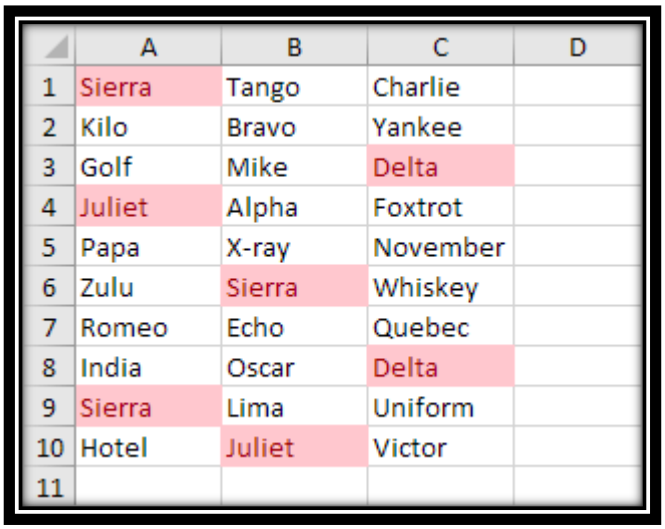
3. Click Highlight Cells Rules, Duplicate Values.



4. Select a formatting style and click OK.



Result. Excel highlights the duplicate names.



	A	B	C	D
1	Sierra	Tango	Charlie	
2	Kilo	Bravo	Yankee	
3	Golf	Mike	Delta	
4	Juliet	Alpha	Foxtrot	
5	Papa	X-ray	November	
6	Zulu	Sierra	Whiskey	
7	Romeo	Echo	Quebec	
8	India	Oscar	Delta	
9	Sierra	Lima	Uniform	
10	Hotel	Juliet	Victor	
11				

Q11. What is the structure of sumif and what are its uses?

SUMIF(range, criteria, [sum_range])

The SUMIF function syntax has the following arguments:

Range Required:- The range of cells that you want evaluated by criteria. Cells in each range must be numbers or names, arrays, or references that contain numbers. Blank and text values are ignored. The selected range may contain dates in standard Excel format (examples below).

criteria Required:-The criteria in the form of a number, expression, a cell reference, text, or a function that defines which cells will be added. Wildcard characters can be included - a question mark (?) to match any single character, an asterisk (*) to match any sequence of characters. If you want to find an actual question mark or asterisk, type a tilde (~) preceding the character.

Use:- The **SUMIF** function is a premade function in Excel, which calculates the sum of values in a range based on a **true** or **false condition**.

Q12. What is Data Validation?

Data validation is a feature in excel that helps you to have a control on what can be entered in your worksheet.

KEY FEATURES OF DATA VALIDATION

- Restrict entries as specified
- Allows creation of custom rules
- Drop down list of items can be created in a cell
- Reduces errors and mistakes

Q13. When trim() function is used?

The trim() function removes whitespace and other predefined characters from both sides of a string.

Related functions:

- ltrim() - Removes whitespace or other predefined characters from the left side of a string
- rtrim() - Removes whitespace or other predefined characters from the right side of a string

Syntax:- trim(*string,charlist*)

Q14. How is a Formula different from a Function in Excel?

S. No.	Function	Formula
1.	It is a pre-defined (built-in) operation used for certain calculations in a certain order.	It is a user-defined operation used for manual calculation in any order.
2.	functions are used to perform complex calculations that cannot be done manually by the user.	Formulas are used to perform simple calculations that can be done manually by the user.
3.	Functions are expressions that operate on small and large values in a range of cells.	Formula is used for quick tasks such as finding the sum, count, average, minimum, maximum, etc. values in a range of cells.
4.	A function can be a formula.	Formula can't be a function.
	A function starts with an equal sign.	You must begin the formula with an equal sign.
5.	Eg: =sum(b2:d4).	Eg: =20+60+40/60 (or), =sum(b2:d4)*256%.
6.	Functions can be updated by the developer.	Formulas are like general calculations, so you can use as per your requirement.
7.	You can either combine formulas in the functions or functions in the formulas for calculations. But it is also treated as a formula (Not a function).	You can either combine formulas in the functions or functions in the formulas for calculations and it is treated as a formula.

Q15. What is concatenating character in Excel?

Concating character is '&' used in Excel.

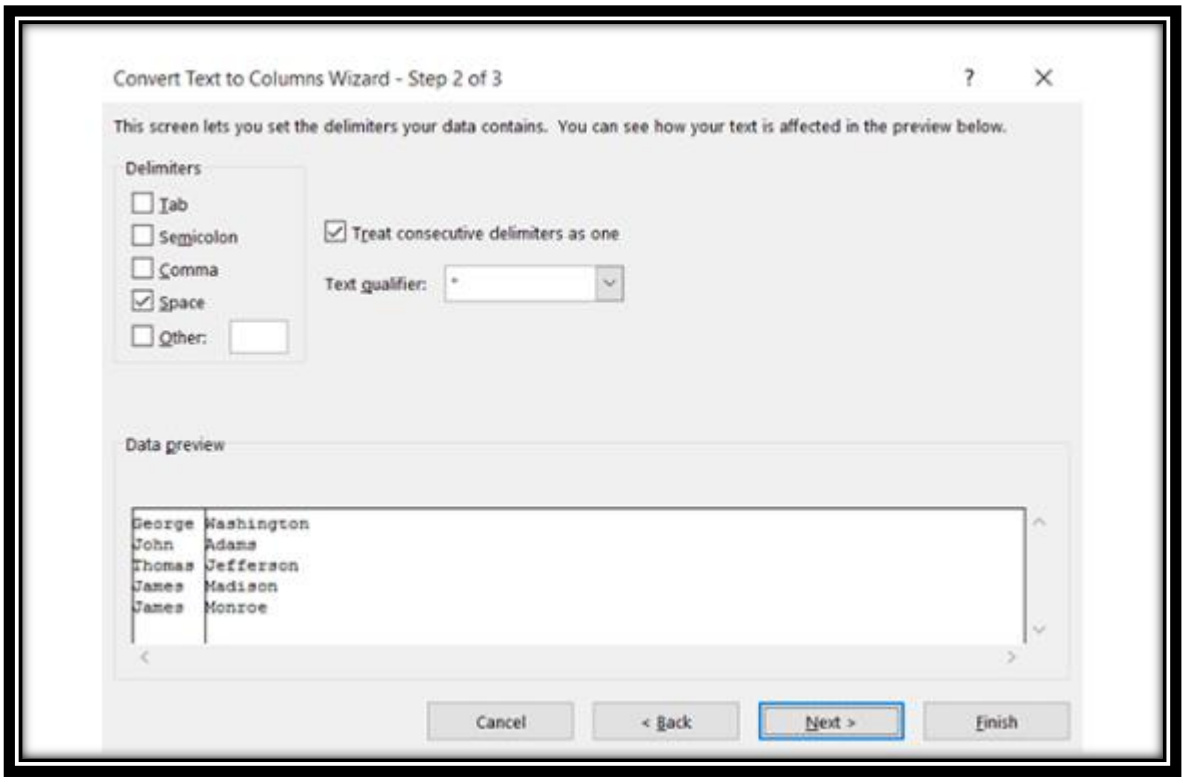
Q16. How can you split a column into 2 or more columns?

(1) Select the data that needs dividing into two columns.

(2) On the Data tab, click the Turn to Columns button. The Convert Text to Columns Wizard appears.

(3) Choose the Delimited option (if it isn't already chosen) and click Next.

(4) Under Delimiters, choose the option that defines how you will divide the data into two columns. Choose Space, for example, to divide a name column into a first name and last name column. The Data Preview box shows what the data will look like when it is divided into two columns.



(5) Click Next.

(6) Click Finish.

Q17. What are the wildcards available in Excel?

Wildcards in Excel are the special characters in excel which takes place of the characters in it, there are three wildcards in excel and they are asterisk, question mark, and tilde, asterisk is used to multiple numbers of characters in excel while question mark is used to represent only a single character whereas tilde is referred to the identification if the wild card character.

Types

There are three types of wildcard characters in excel.

Type #1 – Asterisk (*)

This is to match zero or the number of characters. For example, “Fi*” could match “Final, Fitting, Fill, Finch, and Fiasco,” etc….

Type #2 – Question Mark (?)

This is used to match any single character. For example, “Fa? e” could match “Face” & “Fade,” “?ore” could match “Bore” & “Core,” “a?ide,” which could match “Abide” & “Aside.”

Type #3 – Tilde (~)

This is used to match wildcard characters in the word. For example, if you have the word “Hello*” to find this word, we need to frame the sentence as “Hello~*,” so here, the character tilde (~) specifies the word “Hello” as it is not followed by the wild card character.

18. How pandas are useful in python?

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. The name Pandas is derived from the word Panel Data – an Econometrics from Multidimensional data.

Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data — load, prepare, manipulate, model, and analyze.

Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Key Features of Pandas

- Fast and efficient DataFrame object with default and customized indexing.
- Tools for loading data into in-memory data objects from different file formats.
- Data alignment and integrated handling of missing data.
- Reshaping and pivoting of date sets.
- Label-based slicing, indexing and subsetting of large data sets.
- Columns from a data structure can be deleted or inserted.
- Group by data for aggregation and transformations.
- High performance merging and joining of data.
- Time Series functionality.

19. Explain pd.cut() using example.

Pandas cut() function is used to separate the array elements into different bins . The cut function is mainly used to perform statistical analysis on scalar data.

Syntax: cut(x, bins, right=True, labels=None, retbins=False, precision=3, include_lowest=False, duplicates="raise",)
Parameters:
x: The input array to be binned. Must be 1-dimensional.
bins: defines the bin edges for the segmentation.
Right : (bool, default True) Indicates whether bins includes the rightmost edge or not. If right == True (the default), then the bins [1, 2, 3, 4] indicate (1,2], (2,3], (3,4].
labels : (array or bool, optional) Specifies the labels for the returned bins. Must be the same length as the resulting bins. If False, returns only integer indicators of the bins.
retbins : (bool, default False) Whether to return the bins or not. Useful when bins are provided as a scalar.

20.How can you change the data type of a field in your data using pandas?

We can pass any Python, Numpy or Pandas datatype to change all columns of a dataframe to that type, or we can pass a dictionary having column names as keys and datatype as values to change type of selected columns.

Syntax: DataFrame.astype(dtype, copy = True, errors = 'raise', **kwargs)

21. How can you import CSV files in python?

1. Import the csv library:-

```
import csv
```

2.Open the CSV file:-

The .open() method in python is used to open files and return a file object.

```
file = open('Salary_Data.csv')
type(file)
```

3.Use the csv.reader object to read the CSV file

```
csvreader = csv.reader(file)
```

4. Extract the field names

Create an empty list called header. Use the next() method to obtain the header.

The .next() method returns the current row and moves to the next row.

The first time you run next() it returns the header and the next time you run it returns the first record and so on.

```
header = []
header = next(csvreader)
header
```

5. Extract the rows/records

Create an empty list called rows and iterate through the csvreader object and append each row to the rows list.

```
rows = []
for row in csvreader:
        rows.append(row)
rows
```

6. Close the file

Close () method is used to close the opened file. Once it is closed, we cannot perform any operations on it.

```
file.close()
```

22.What is the difference between list and array?

S.No.	List	Array
1	List is used to collect items that usually consist of elements of multiple data types.	An array is also a vital component that collects several items of the same data type.
2	List cannot manage arithmetic operations.	Array can manage arithmetic operations.
3	It consists of elements that belong to the different data types.	It consists of elements that belong to the same data type.
4	When it comes to flexibility, the list is perfect as it allows easy modification of data.	When it comes to flexibility, the array is not suitable as it does not allow easy modification of data.
5	It consumes a larger memory.	It consumes less memory than a list.
6	In a list, the complete list can be accessed without any specific looping.	In an array, a loop is mandatory to access the components of the array.
7	It favors a shorter sequence of data.	It favors a longer sequence of data.

23.Write code for removing duplicates in lists.

Method 1: Naïve Method

In this method, we traverse the list and then append the first occurrence of the element in the new list, and then all other elements are ignored.

Example:

```
# removing duplicated from the list using naive methods

# initializing list
sam_list = [11, 13, 15, 16, 13, 15, 16, 11]
print ("The list is: " + str(sam_list))

# remove duplicated from list
result = []
for i in sam_list:
    if i not in result:
        result.append(i)

# printing list after removal
print ("The list after removing duplicates : " + str(result))
```

Output:

```
The list is: [11, 13, 15, 16, 13, 15, 16, 11]

The list after removing duplicates: [11, 13, 15, 16]
```

Method 2: Using a list comprehensive

This method is similar to the above method, but this method is shorter than the above one.

Example:

```
# removing duplicated from the list using list comprehension

# initializing list
sam_list = [11, 13, 15, 16, 13, 15, 16, 11]
print ("The list is: " + str(sam_list))


# to remove duplicated from list
result = []
[result.append(x) for x in sam_list if x not in result]

# printing list after removal
print ("The list after removing duplicates: " + str(result))
```

Output:

```
The list is: [11, 13, 15, 16, 13, 15, 16, 11]

The list after removing duplicates: [11, 13, 15, 16]
```

Method 3: Using set()

This method is the most popular method to remove the duplicate from the python list. This is because the set data structure does not allow duplicates. But the drawback of this method is that the order of the elements is lost.

Example:

```
# removing duplicated from the list using set()

# initializing list
sam_list = [11, 15, 13, 16, 13, 15, 16, 11]
print ("The list is: " + str(sam_list))

# to remove duplicated from list
sam_list = list(set(sam_list))

# printing list after removal
# ordering distorted
print ("The list after removing duplicates: " + str(sam_list))
```

Output:

```
The list is: [11, 15, 13, 16, 13, 15, 16, 11]

The list after removing duplicates: [16, 11, 13, 15]
```

Method 4: Using list comprehensive + enumerate()

List comprehensive when merged with enumerate function we can remove the duplicate from the python list. Basically, in this method, the already occurred elements are skipped, and also the order is maintained.

Example:

```
# removing duplicated from the list using list comprehension + enumerate()

# initializing list
sam_list = [11, 15, 13, 16, 13, 15, 16, 11]
print ("The list is: " + str(sam_list))

# to remove duplicated from list
result = [i for n, i in enumerate(sam_list) if i not in sam_list[:n]]

# printing list after removal
print ("The list after removing duplicates: " + str(result))
```

Output:

```
The list is: [11, 13, 15, 16, 13, 15, 16, 11]

The list after removing duplicates: [11, 13, 15, 16]
```

Method 5: Using collections.OrderedDict.fromkeys()

This is the fastest method to achieve the target of removing duplicates from the python list. This method will first remove the duplicates and return a dictionary that has converted to a list. Also, this method works well in the case of a string.

Example:

```python
# removing duplicated from list using collections.OrderedDict.fromkeys()
from collections import OrderedDict

# initializing list
sam_list = [11, 15, 13, 16, 13, 15, 16, 11]
print ("The list is: " + str(sam_list))

# to remove duplicated from list
result = list(OrderedDict.fromkeys(sam_list))

# printing list after removal
print ("The list after removing duplicates: " + str(result))
```

Output:

```
The list is: [11, 15, 13, 16, 13, 15, 16, 11]

The list after removing duplicates: [11, 15, 13, 16]
```

24.Explain Numpy and pandas libraries.

Pandas

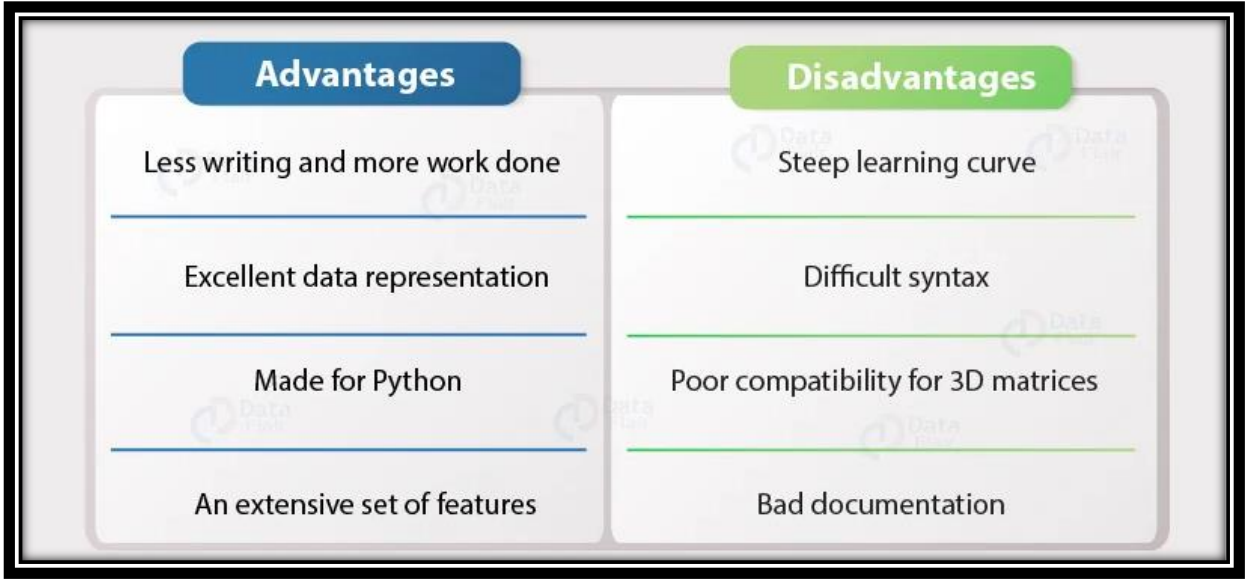
Pandas is a very popular library for working with data (its goal is to be the most powerful and flexible open-source tool, and in our opinion, it has reached that goal). DataFrames are at the center of pandas. A DataFrame is structured like a table or spreadsheet. The rows and the columns both have indexes, and you can perform operations on rows or columns separately.

A pandas DataFrame can be easily changed and manipulated. Pandas has helpful functions for handling missing data, performing operations on columns and rows, and transforming data. If that wasn't enough, a lot of SQL functions have counterparts in pandas, such as join, merge, filter by, and group by. With all of these powerful tools, it should come as no surprise that pandas is very popular among data scientists.

NumPy

NumPy is an open-source Python library that facilitates efficient numerical operations on large quantities of data. There are a few functions that exist in NumPy that we use on pandas DataFrames. For us, the most important part about NumPy is that pandas is built on top of it. So, NumPy is a dependency of Pandas.

25.What are the Advantages of pandas?



1.1. Data representation

Pandas provide extremely streamlined forms of data representation. This helps to analyze and understand data better. Simpler data representation facilitates better results for data science projects.

1.2. Less writing and more work done

It is one of the best advantages of Pandas. What would have taken multiple lines in Python without any support libraries, can simply be achieved through 1-2 lines with the use of Pandas. Thus, using Pandas helps to shorten the procedure of handling data. With the time saved, we can focus more on data analysis algorithms.

1.3. An extensive set of features

Pandas are really powerful. They provide you with a huge set of important commands and features which are used to easily analyze your data. We can use Pandas to perform various tasks like filtering your data according to certain conditions, or segmenting and segregating the data according to preference, etc.

1.4. Efficiently handles large data

Wes McKinney, the creator of Pandas, made the python library to mainly handle large datasets efficiently. Pandas help to save a lot of time by importing large amounts of data very fast.

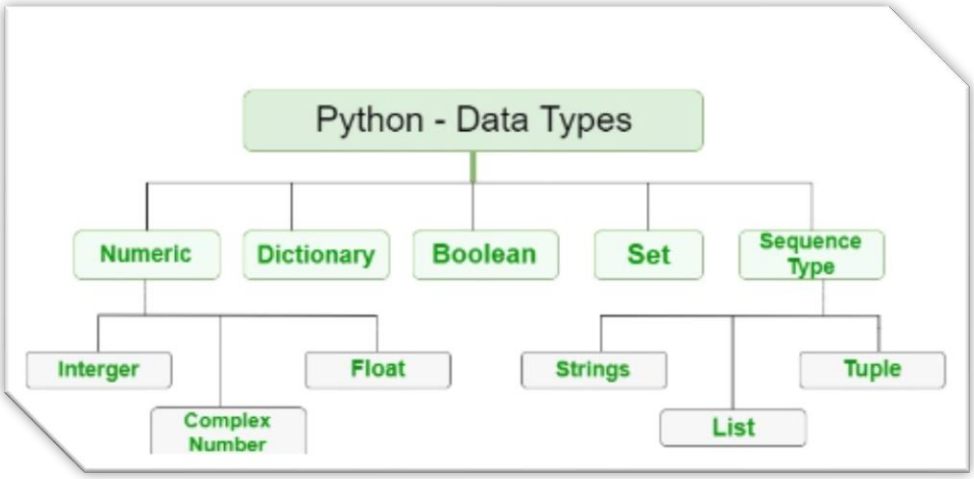
1.5. Makes data flexible and customizable

Pandas provide a huge feature set to apply on the data you have so that you can customize, edit and pivot it according to your own will and desire. This helps to bring the most out of your data.

1.6. Made for Python

Python programming has become one of the most sought after programming languages in the world, with its extensive amount of features and the sheer amount of productivity it provides. Therefore, being able to code Pandas in Python, enables you to tap into the power of the various other features and libraries which will use with Python. Some of these libraries are NumPy, SciPy, MatPlotLib, etc.

26.What are datatypes in python?



27. What is the difference between lists and tuples?

List:- list is just like the arrays, declared in other languages. Lists need not be homogeneous always which makes it the most powerful tool in Python. In Python, the list is a type of container in Data Structures, which is used to store multiple data at the same time. Lists are a useful tool for preserving a sequence of data and further iterating over it.

Syntax:

```
list_data = ['an', 'example', 'of', 'a', 'list']
```

Tuple:-tuple is also a sequence data type that can contain elements of different data types, but these are immutable in nature. In other words, a tuple is a collection of Python objects separated by commas. The tuple is faster than the list because of static in nature.

Syntax:

```
tuple_data = ('this', 'is', 'an', 'example', 'of', 'tuple')
```

SR.NO.	LIST	TUPLE
1	Lists are mutable	Tuples are immutable
2	Implication of iterations is Time-consuming	The implication of iterations is comparatively Faster
3	The list is better for performing operations, such as insertion and deletion.	Tuple data type is appropriate for accessing the elements
4	Lists consume more memory	Tuple consume less memory as compared to the list
5	Lists have several built-in methods	Tuple does not have many built-in methods.
6	The unexpected changes and errors are more likely to occur	In tuple, it is hard to take place.

28.What is the difference between blank and Null?

A null value is a value that doesn't exist: the field does not contain a value of any kind (not even a blank value). By contrast, a blank value is a real value: it just happens to be a string value containing 0 characters.

29.How to create dataframe and how to convert list into dataframe?

DataFrame:-

Pandas is a software library written for the Python programming language for data manipulation and analysis. Pandas Dataframe is a two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A data frame could be a two-dimensional data structure, i.e., knowledge is aligned in a very tabular fashion in rows and columns. Pandas Dataframe consists of 3 principal elements, the data, rows, and columns.

For Example

```
import pandas as pd

# list of strings

lst = ['fav', 'tutor', 'coding', 'skills']
df = pd.DataFrame(lst)
print(df)
```

Output

```
       0
0    fav
1  tutor
2 coding
3 skills
```

Convert List to DataFrame in Python:-

There are many ways to create a data frame from the list. We will look at different 6 methods to convert lists from data frames in Python. Let us study them one by one with an example:

1) Basic method

This is the simplest method to create the data frames from the list.

For example

```
# import pandas as pd

import pandas as pd
# list of strings
lst = ['fav', 'tutor', 'coding', 'skills']
# Calling DataFrame constructor on list
df = pd.DataFrame(lst)
print(df)
```

Output

```
       0
0    fav
1  tutor
2 coding
3 skills
```

2) Using a list with index and column names

We can create the data frame by giving the name to the column and index the rows

For example

```
# import pandas as pd

import pandas as pd
# List1
lst = [['apple', 'red', 11], ['grape', 'green', 22], ['orange', 'orange', 33], ['mango', 'yellow',
df = pd.DataFrame(lst, columns =['Fruits', 'Color', 'Value'], dtype = float)
print(df)
```

Output

```
  tutorial
1      fav
2    tutor
3   coding
4   skills
```

3) Using zip() function

We can create the data frame by zipping two lists.

For example

```
 # import pandas as pd

import pandas as pd
# list of strings
lst1 = ['fav', 'tutor', 'coding', 'skills']
# list of int
lst2 = [11, 22, 33, 44]
# Calling DataFrame after zipping both lists, with columns specified
df = pd.DataFrame(list(zip(lst1, lst2)), columns =['key', 'value'])
print(df)
```

Output

```
      key  value
0     fav     11
1   tutor     22
2  coding     33
3  skills     44
```

6) Using a list in the dictionary

We can create data frames using lists in the dictionary.

For example

```
# import pandas as pd

import pandas as pd
# list of name, degree, score
n = ["apple", "grape", "orange", "mango"]
col = ["red", "green", "orange", "yellow"]
val = [11, 22, 33, 44]
# dictionary of lists
dict = {'fruit': n, 'color': col, 'value': val}
df = pd.DataFrame(dict)
print(df)
```

Output

```
    fruit   color  value
0   apple     red     11
1   grape   green     22
2  orange  orange     33
3   mango  yellow     44
```

SQL Theory Questions

1. What is a Foreign Key? Explain with Example.

-A foreign key is a key used to link two tables together. This is sometimes also called as a referencing key.

-A Foreign Key is a column or a combination of columns whose values match a Primary Key in a different table.

-The relationship between 2 tables matches the Primary Key in one of the tables with a Foreign Key in the second table.

-If a table has a primary key defined on any field(s), then you cannot have two records having the same value of that field(s).

EXAMPLE:-

we have two tables, a ***CUSTOMER*** table that includes all customer data, and an ***ORDERS*** table that includes all customer orders. Business logic requires that all orders must be associated with a customer that is already in the ***CUSTOMER*** table. To enforce this logic, we place a foreign key on the ***ORDERS*** table and have it reference the primary key of the ***CUSTOMER*** table. This way, we can ensure that all orders in the ***ORDERS*** table are related to a customer in the ***CUSTOMER*** table. In other words, the ***ORDERS*** table cannot contain information on a customer that is not in the ***CUSTOMER*** table.The structure of these two tables will be as follows:

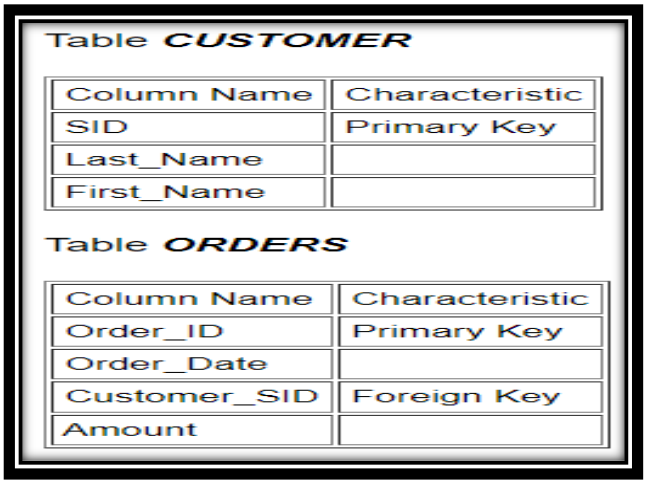


Table ***CUSTOMER***

Column Name	Characteristic
SID	Primary Key
Last_Name	
First_Name	

Table ***ORDERS***

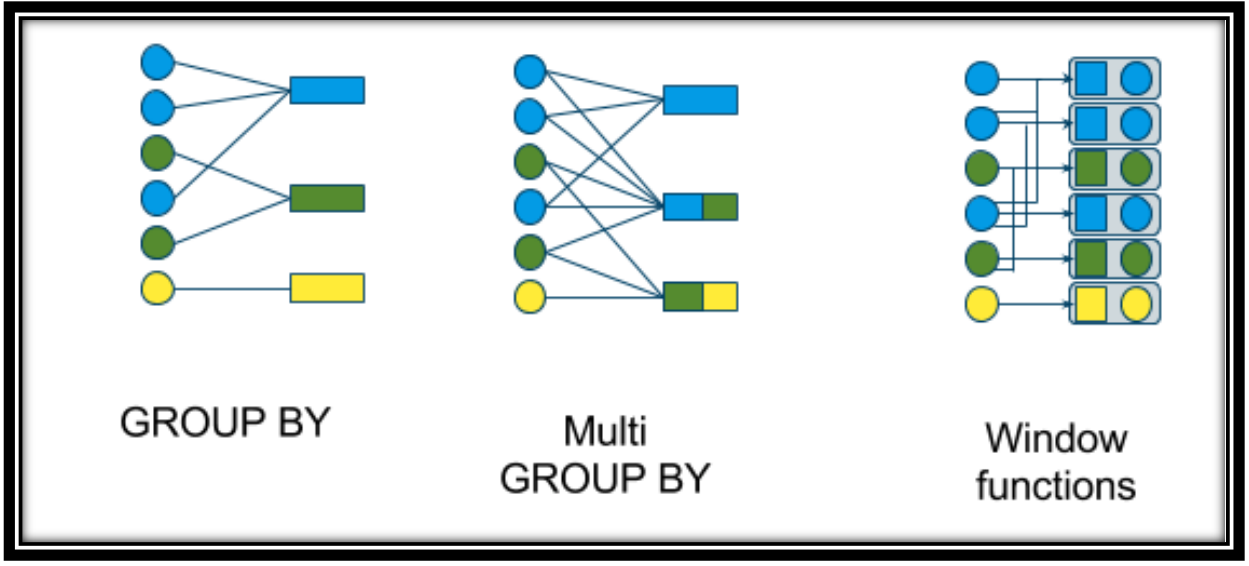
Column Name	Characteristic
Order_ID	Primary Key
Order_Date	
Customer_SID	Foreign Key
Amount	

2. WHAT IS THE DIFFERENCE BETWEEN GROUP BY AND WINDOW FUNCTIONS?

Window Function:- A window function performs an operation across a set of rows in a table, also known as partitions (subgroups). It's called window precisely because the set of rows is called window.

Group By Function:- The GROUP BY statement **groups rows that have the same values into summary rows**, like "find the number of customers in each country". The GROUP BY statement is often used with aggregate functions (COUNT() , MAX() , MIN() , SUM() , AVG()) to group the result-set by one or more columns.

VISUAL THAT SHOWS HOW EFFICIENT AND POWERFUL A WINDOW FUNCTION IS COMPARED GROUP BY. UNLIKE GROUP BY, WINDOW FUNCTIONS CAN CALCULATE RUNNING TOTALS AND MOVING AVERAGES VIA RANKING AND VALUE FUNCTIONALITY.



3. What is the difference between UNION and UNION ALL?

UNION	UNION ALL
It combines the result set from multiple tables and **returns distinct records** into a single result set.	It combines the result set from multiple tables and **returns all records** into a single result set.
Following is the basic **syntax** of UNION operator: SELECT column_list FROM table1 UNION SELECT column_list FROM table2;	Following is the basic **syntax** of UNION ALL operator: SELECT column_list FROM table1 UNION ALL SELECT column_list FROM table2;
It has a default feature to eliminate the duplicate rows from the output.	It has no feature to eliminate the duplicate rows from the output.
Its **performance is slow** because it takes time to find and then remove duplicate records.	Its **performance is fast** because it does not eliminate the duplicate rows.
Most database users prefer to use this operator.	Most database users do not prefer to use this operator.

4. Are NULL values the same as that of zero or a blank space?

A NULL value is not same as zero or a blank space. A NULL value is a value which is 'unavailable, unassigned, unknown or not applicable'. Whereas, zero is a number and blank space is a character.

5. What is the difference between a Where clasue and a Having Clause?

Where Clause in SQL	Having Clause in SQL
Filter table based data catering to specific condition	Group based data under set condition
Applicable without GROUP BY clause	Does not function without GROUP BY clause
Row functions	Column functions
Select, update and delete statements	Only select statement
Applied before GROUP BY clause	Used after GROUP BY clause
Used with single row operations such as Upper, Lower and so on	Applicable with multiple row functions such as Sum, count and so on

6. WHAT IS A SUBQUERY IN SQL? WHAT ARE THE DIFFERENT TYPES OF SUBQUERY?

-SQL SUBQUERIES OR NESTED QUERIES ARE SQL STATEMENTS WHERE WE NEED THE RESULTS FROM OUR DATABASE AFTER USING MULTIPLE FILTERS.

-A SUBQUERY IS PUT TO RESTRICT THE DATA POOL FOR THE MAIN QUERY I.E., THE INNER QUERY GIVES US THE DATA WHICH IS THE POOL FOR THE MAIN QUERY.

-SUBQUERIES ARE COMPATIBLE WITH ALMOST ALL SQL STATEMENTS, FOR EXAMPLE,

INSERT

UPDATE

DELETE

SELECT

Types of SQL Subqueries

We have various subqueries; some of them are as follows:

1. Single Row Subquery
Returns zero or one row in results.

2. Multiple Row Subquery
Returns one or more rows in results.

3. Multiple Column Subqueries
Returns one or more columns

4. Correlated Subqueries
Returns one or more columns according to the main or the outer query, thus called a correlated subquery.

7. What is the difference between Truncate, Delete and Drop statements?

DELETE :

Basically, it is a Data Manipulation Language Command (DML). It is used to delete one or more tuples of a table. With the help of the "DELETE" command, we can either delete all the rows in one go or can delete rows one by one. i.e., we can use it as per the requirement or the condition using the Where clause. It is comparatively slower than the TRUNCATE command. The TRUNCATE command does not remove the structure of the table.

SYNTAX –

If we want to delete all the rows of the table:

```
DELETE from;
```

SYNTAX –

If we want to delete the row of the table as per the condition then we use

```
DELETE from  WHERE  ;
```

DROP :

It is a Data Definition Language Command (DDL). It is used to drop the whole table. With the help of the "DROP" command we can drop (delete) the whole structure in one go i.e. it removes the named elements of the schema. By using this command the existence of the whole table is finished or say lost.

SYNTAX –

If we want to drop the table:

```
DROP table ;
```

TRUNCATE :

It is also a Data Definition Language Command (DDL). It is used to delete all the rows of a relation (table) in one go. With the help of the "TRUNCATE" command, we can't delete the single row as here WHERE clause is not used. By using this command the existence of all the rows of the table is lost. It is comparatively faster than the delete command as it deletes all the rows fastly.

SYNTAX - If we want to use truncate :

```
TRUNCATE TABLE table_name;
```

8. What is a Relationship and what are they?

A relational database collects different types of data sets that use tables, records, and columns. It is used to create a well-defined relationship between database tables so that relational databases can be easily stored. For example of relational databases such as Microsoft SQL Server, Oracle Database, MYSQL, etc.

There are some important parameters of the relational database**:**

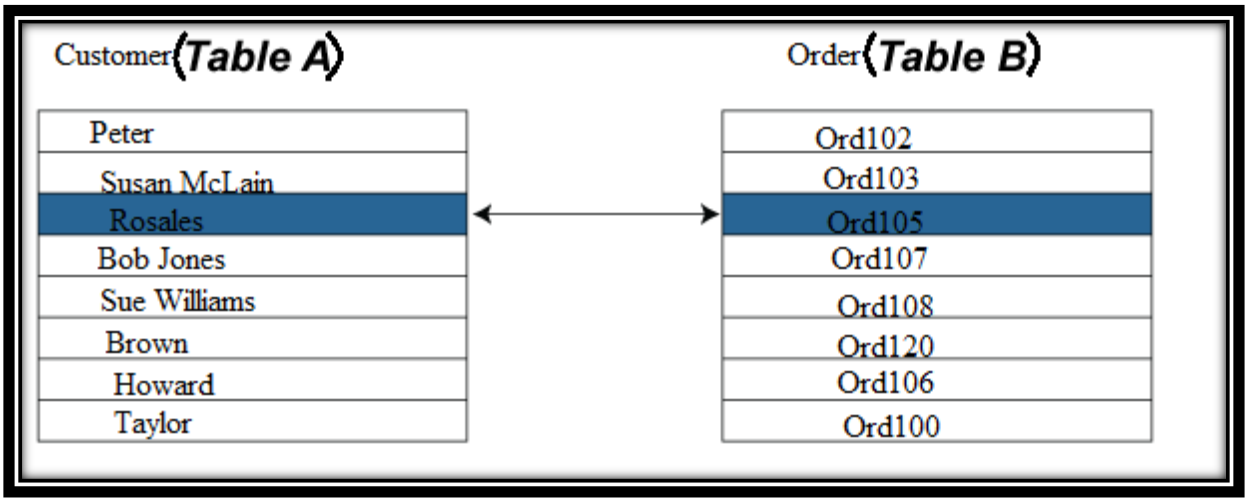
- It is based on a relational model (Data in tables).
- Each row in the table with a unique id, key.
- Columns of the table hold attributes of data.

Following are the different types of relational database tables.

1. One to One relationship
2. One to many or many to one relationship
3. Many to many relationships

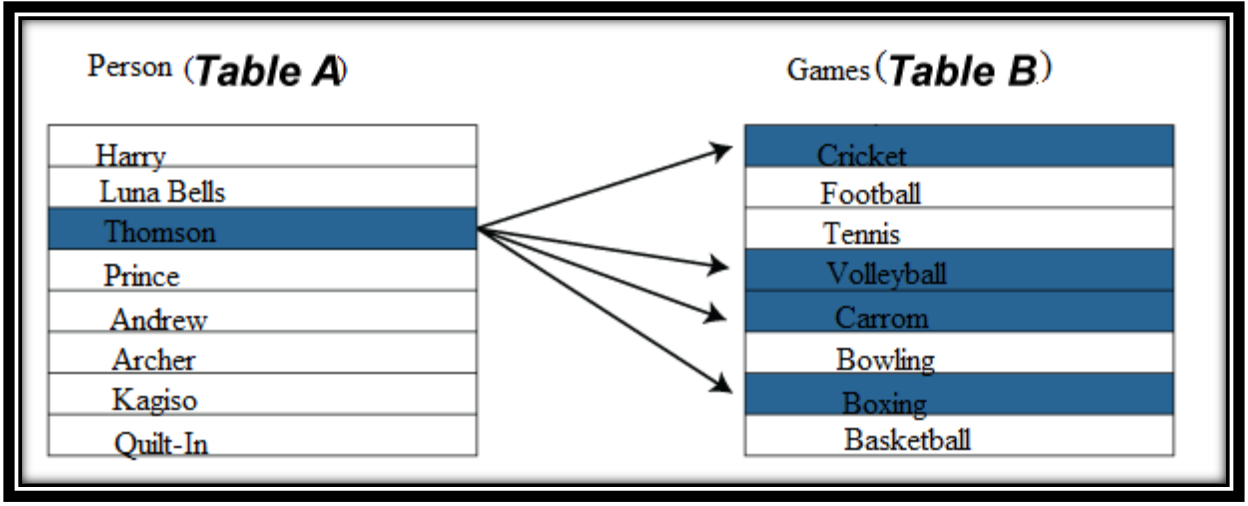
One to One Relationship (1:1): It is used to create a relationship between two tables in which a single row of the first table can only be related to one and only one records of a second table. Similarly, the row of a second table can also be related to anyone row of the first table.

Following is the example to show a relational database, as shown below.



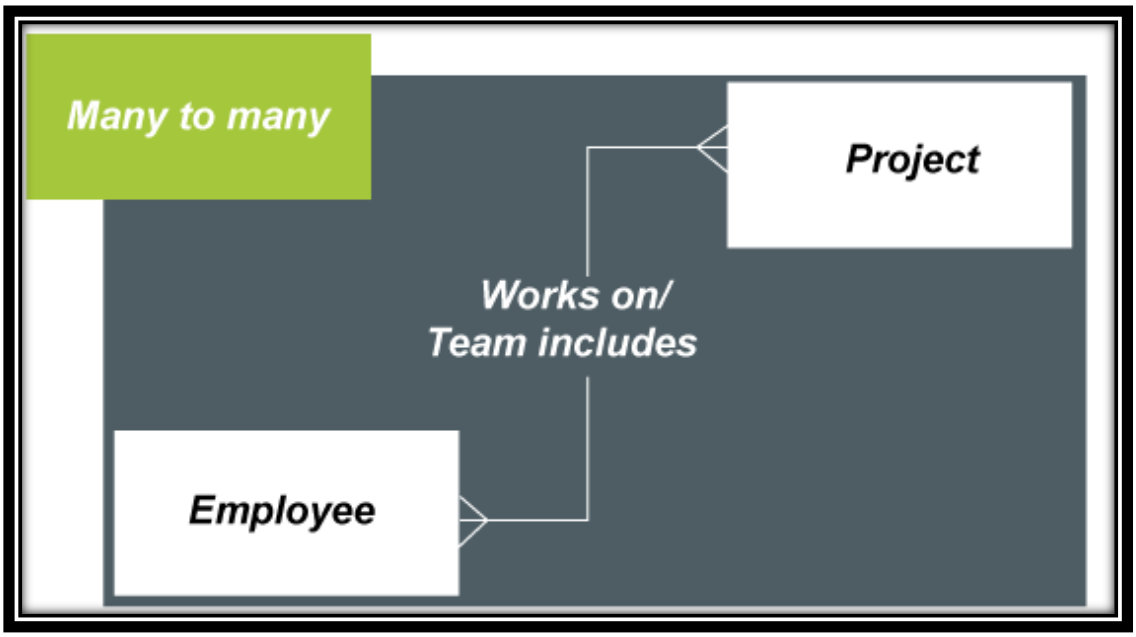
One to Many Relationship: It is used to create a relationship between two tables. Any single rows of the first table can be related to one or more rows of the second tables, but the rows of second tables can only relate to the only row in the first table. It is also known as a **many to one** relationship.

Representation of **One to Many** relational databases:



Many to Many Relationship: It is **many to many** relationships that create a relationship between two tables. Each record of the first table can relate to any records (or no records) in the second table. Similarly, each record of the second table can also relate to more than one record of the first table. It is also represented an **N:N** relationship.

For example, there are **many** people involved in each **project,** and every person can involve more than one project.



9. What do you mean by table and field in SQL?

Tables contain rows and columns, where the rows are known as records and the columns are known as fields.

10. What are constraints in SQL?

SQL constraints are used to specify rules for the data in a table.

Constraints are used to limit the type of data that can go into a table. This ensures the accuracy and reliability of the data in the table. If there is any violation between the constraint and the data action, the action is aborted.

Constraints can be column level or table level. Column level constraints apply to a column, and table level constraints apply to the whole table.

The following constraints are commonly used in SQL

- `NOT NULL` - Ensures that a column cannot have a NULL value
- `UNIQUE` - Ensures that all values in a column are different
- `PRIMARY KEY` - A combination of a `NOT NULL` and `UNIQUE`. Uniquely identifies each row in a table
- `FOREIGN KEY` - Prevents actions that would destroy links between tables
- `CHECK` - Ensures that the values in a column satisfies a specific condition
- `DEFAULT` - Sets a default value for a column if no value is specified
- `CREATE INDEX` - Used to create and retrieve data from the database very quickly