



---

# DATA ANALYTICS INTERVIEW QUESTIONS AND ANSWERS

---



MOHD MUJTABA

# DATA ANALYTICS INTERVIEW QUESTIONS AND ANSWERS

## Q1 What is Data Analysis?

Data analysis is basically a process of analyzing, modeling, and interpreting data to draw insights or conclusions. With the insights gained, informed decisions can be made. It is used by every industry, which is why data analysts are in high demand. A Data Analyst's sole responsibility is to play around with large amounts of data and search for hidden insights. By interpreting a wide range of data, data analysts assist organizations in understanding the business's current state.

## Q2 What are the responsibilities of a Data Analyst?

Some of the responsibilities of a data analyst include:

- Collects and analyzes data using statistical techniques and reports the results accordingly.
- Interpret and analyze trends or patterns in complex data sets.
- Establishing business needs together with business teams or management teams.
- Find opportunities for improvement in existing processes or areas.
- Data set commissioning and decommissioning.
- Follow guidelines when processing confidential data or information.
- Examine the changes and updates that have been made to the source production systems.
- Provide end-users with training on new reports and dashboards.
- Assist in the data storage structure, data mining, and data cleansing.

## Q3 Write some key skills usually required for a data analyst.

Some of the key skills required for a data analyst include:



InterviewBit

- Knowledge of reporting packages (Business Objects), coding languages (e.g., XML, JavaScript, ETL), and databases (SQL, SQLite, etc.) is a must.
- Ability to analyze, organize, collect, and disseminate big data accurately and efficiently.
- The ability to design databases, construct data models, perform data mining, and segment data.
- Good understanding of statistical packages for analyzing large datasets (SAS, SPSS, Microsoft Excel, etc.).
- Effective Problem-Solving, Teamwork, and Written and Verbal Communication Skills.
- Excellent at writing queries, reports, and presentations.
- Understanding of data visualization software including Tableau and Qlik.
- The ability to create and apply the most accurate algorithms to datasets for finding solutions.

#### Q4 What is the data analysis process?

Data analysis generally refers to the process of assembling, cleaning, interpreting, transforming, and modeling data to gain insights or conclusions and generate reports to help businesses become more profitable. The following diagram illustrates the various steps involved in the process:



- **Collect Data:** The data is collected from a variety of sources and is then stored to be cleaned and prepared. This step involves removing all missing values and outliers.
- **Analyse Data:** As soon as the data is prepared, the next step is to analyze it. Improvements are made by running a model repeatedly. Following that, the model is validated to ensure that it is meeting the requirements.
- **Create Reports:** In the end, the model is implemented, and reports are generated as well as distributed to stakeholders.

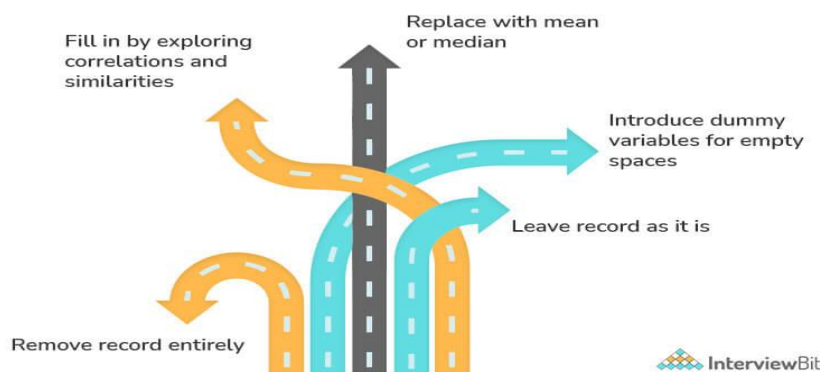
#### Q5 What are the different challenges one faces during data analysis?

While analyzing data, a Data Analyst can encounter the following issues:

- Duplicate entries and spelling errors. Data quality can be hampered and reduced by these errors.
- The representation of data obtained from multiple sources may differ. It may cause a delay in the analysis process if the collected data are combined after being cleaned and organized.
- Another major challenge in data analysis is incomplete data. This would invariably lead to errors or faulty results.
- You would have to spend a lot of time cleaning the data if you are extracting data from a poor source.
- Business stakeholders' unrealistic timelines and expectations
- Data blending/ integration from multiple sources is a challenge, particularly if there are no consistent parameters and conventions
- Insufficient data architecture and tools to achieve the analytics goals on time.

#### Q6 Explain data cleansing.

Data cleaning, also known as data cleansing or data scrubbing or wrangling, is basically a process of identifying and then modifying, replacing, or deleting the incorrect, incomplete, inaccurate, irrelevant, or missing portions of the data as the need arises. This fundamental element of data science ensures data is correct, consistent, and usable.



### Q7 What are the tools useful for data analysis?

Some of the tools useful for data analysis include:

1. R and Python
2. Microsoft Excel
3. Tableau
4. RapidMiner
5. KNIME
6. Power BI
7. Apache Spark
8. QlikView
9. Talend
10. Splunk

### Q8 Write the difference between data mining and data profiling.

**Data mining Process:** It generally involves analyzing data to find relations that were not previously discovered. In this case, the emphasis is on finding unusual records, detecting dependencies, and analyzing clusters. It also involves analyzing large datasets to determine trends and patterns in them.

**Data Profiling Process:** It generally involves analyzing that data's individual attributes. In this case, the emphasis is on providing useful information on data attributes such as data type, frequency, etc. Additionally, it also facilitates the discovery and evaluation of enterprise metadata.

Data Mining	Data Profiling
It involves analyzing a pre-built database to identify patterns.	It involves analyses of raw data from existing datasets.
It also analyzes existing databases and large datasets to convert raw data into useful information.	In this, statistical or informative summaries of the data are collected.
It usually involves finding hidden patterns and seeking out new, useful, and non-trivial data to generate useful information.	It usually involves the evaluation of data sets to ensure consistency, uniqueness, and logic.
Data mining is incapable of identifying inaccurate or incorrect data values.	In data profiling, erroneous data is identified during the initial stage of analysis.
Classification, regression, clustering, summarization, estimation, and description are some primary data mining tasks that are needed to be performed.	This process involves using discoveries and analytical methods to gather statistics or summaries about the data.

### Q9 Which validation methods are employed by data analysts?

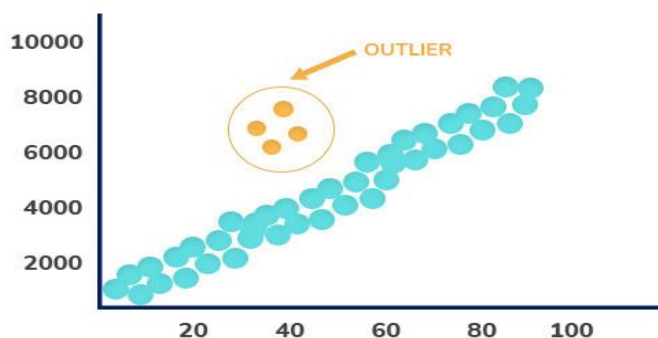
In the process of data validation, it is important to determine the accuracy of the information as well as the quality of the source. Datasets can be validated in many ways. Methods of data validation commonly used by Data Analysts include:

- **Field Level Validation:** This method validates data as and when it is entered into the field. The errors can be corrected as you go.
- **Form Level Validation:** This type of validation is performed after the user submits the form. A data entry form is checked at once, every field is validated, and highlights the errors (if present) so that the user can fix them.

- **Data Saving Validation:** This technique validates data when a file or database record is saved. The process is commonly employed when several data entry forms must be validated.
- **Search Criteria Validation:** It effectively validates the user's search criteria in order to provide the user with accurate and related results. Its main purpose is to ensure that the search results returned by a user's query are highly relevant.

#### Q10 Explain Outlier.

In a dataset, Outliers are values that differ significantly from the mean of characteristic features of a dataset. With the help of an outlier, we can determine either variability in the measurement or an experimental error. There are two kinds of outliers i.e., Univariate and Multivariate. The graph depicted below shows there are four outliers in the dataset.



#### Q11 What are the ways to detect outliers? Explain different ways to deal with it.

Outliers are detected using two methods:

- **Box Plot Method:** According to this method, the value is considered an outlier if it exceeds or falls below  $1.5 \times \text{IQR}$  (interquartile range), that is, if it lies above the top quartile (Q3) or below the bottom quartile (Q1).
- **Standard Deviation Method:** According to this method, an outlier is defined as a value that is greater or lower than the mean  $\pm (3 \times \text{standard deviation})$ .

#### Q12 Write difference between data analysis and data mining.

**Data Analysis:** It generally involves extracting, cleansing, transforming, modeling, and visualizing data in order to obtain useful and important information that may contribute towards determining conclusions and deciding what to do next. Analyzing data has been in use since the 1960s.

**Data Mining:** In data mining, also known as knowledge discovery in the database, huge quantities of knowledge are explored and analyzed to find patterns and rules. Since the 1990s, it has been a buzzword.



Data Analysis	Data Mining
Analyzing data provides insight or tests hypotheses.	A hidden pattern is identified and discovered in large datasets.
It consists of collecting, preparing, and modeling data in order to extract meaning or insights.	This is considered as one of the activities in Data Analysis.
Data-driven decisions can be taken using this way.	Data usability is the main objective.

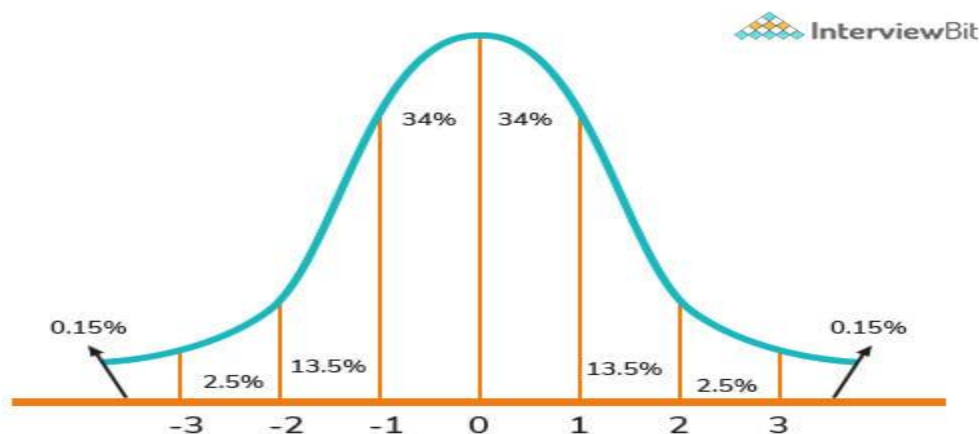
Data Analysis	Data Mining
Data visualization is certainly required.	Visualization is generally not necessary.
It is an interdisciplinary field that requires knowledge of computer science, statistics, mathematics, and machine learning.	Databases, machine learning, and statistics are usually combined in this field.
Here the dataset can be large, medium, or small, and it can be structured, semi-structured, and unstructured.	In this case, datasets are typically large and structured.

### Q13 Explain the KNN imputation method.

A KNN (K-nearest neighbor) model is usually considered one of the most common techniques for imputation. It allows a point in multidimensional space to be matched with its closest k neighbors. By using the distance function, two attribute values are compared. Using this approach, the closest attribute values to the missing values are used to impute these missing values.

### Q14 Explain Normal Distribution.

Known as the bell curve or the Gauss distribution, the Normal Distribution plays a key role in statistics and is the basis of Machine Learning. It generally defines and measures how the values of a variable differ in their means and standard deviations, that is, how their values are distributed.



The above image illustrates how data usually tend to be distributed around a central value with no bias on either side. In addition, the random variables are distributed according to symmetrical bell-shaped curves.

### Q15 What do you mean by data visualization?

The term data visualization refers to a graphical representation of information and data. Data visualization tools enable users to easily see and understand trends, outliers, and patterns in data through the use of visual elements like charts, graphs, and maps. Data can be viewed and analyzed in a smarter way, and it can be converted into diagrams and charts with the use of this technology.

### Q16 How does data visualization help you?

Data visualization has grown rapidly in popularity due to its ease of viewing and understanding complex data in the form of charts and graphs. In addition to providing data in a format that is easier to understand, it highlights trends and outliers. The best visualizations illuminate meaningful information while removing noise from data.

**Q17 Mention some of the python libraries used in data analysis.**

Several Python libraries that can be used on data analysis include:

- NumPy
- Bokeh
- Matplotlib
- Pandas
- SciPy
- SciKit, etc.

**Q18 Write characteristics of a good data model.**

An effective data model must possess the following characteristics in order to be considered good and developed:

- Provides predictability performance, so the outcomes can be estimated as precisely as possible or almost as accurately as possible.
- As business demands change, it should be adaptable and responsive to accommodate those changes as needed.
- The model should scale proportionally to the change in data.
- Clients/customers should be able to reap tangible and profitable benefits from it.

**Q19 Write disadvantages of Data analysis.**

The following are some disadvantages of data analysis:

- Data Analytics may put customer privacy at risk and result in compromising transactions, purchases, and subscriptions.
- Tools can be complex and require previous training.
- Choosing the right analytics tool every time requires a lot of skills and expertise.
- It is possible to misuse the information obtained with data analytics by targeting people with certain political beliefs or ethnicities.

**Q20 What do you mean by Time Series Analysis? Where is it used?**

In the field of Time Series Analysis (TSA), a sequence of data points is analyzed over an interval of time. Instead of just recording the data points intermittently or randomly, analysts record data points at regular intervals over a period of time in the TSA. It can be done in two different ways: in the frequency and time domains. As TSA has a broad scope of application, it can be used in a variety of fields. TSA plays a vital role in the following places:

- Statistics
- Signal processing
- Econometrics
- Weather forecasting
- Earthquake prediction
- Astronomy
- Applied science

**Q21 What do you mean by clustering algorithms? Write different properties of clustering algorithms?**

Clustering is the process of categorizing data into groups and clusters. In a dataset, it identifies similar data groups. It is the technique of grouping a set of objects so that the objects within the same cluster are similar to one another rather than to those located in other clusters. When implemented, the clustering algorithm possesses the following properties:

- Flat or hierarchical
- Hard or Soft
- Iterative
- Disjunctive



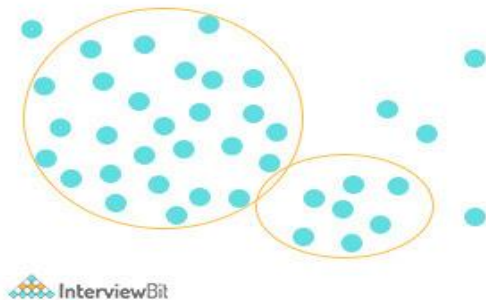
**Q22 Name some popular tools used in big data.**

In order to handle Big Data, multiple tools are used. There are a few popular ones as follows:

- Hadoop
- Spark
- Scala
- Hive
- Flume
- Mahout, etc.

**Q23 Explain Hierarchical clustering.**

This algorithm group objects into clusters based on similarities, and it is also called hierarchical cluster analysis. When hierarchical clustering is performed, we obtain a set of clusters that differ from each other.



This clustering technique can be divided into two types:

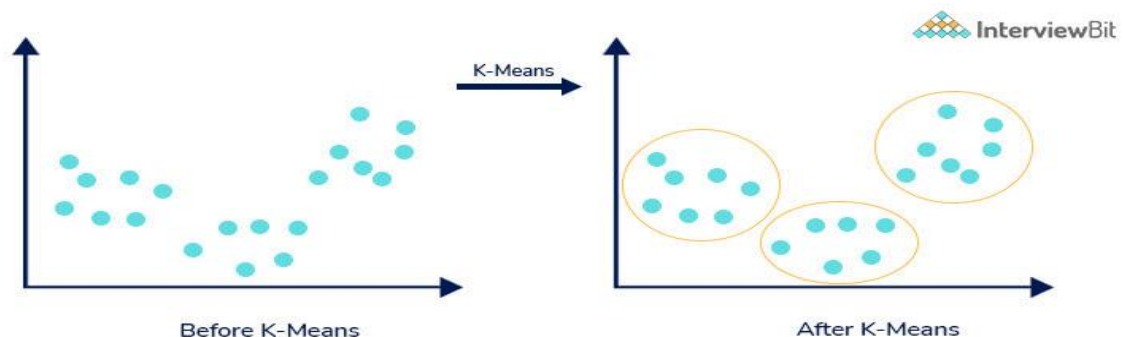
- Agglomerative Clustering (which uses bottom-up strategy to decompose clusters)
- Divisive Clustering (which uses a top-down strategy to decompose clusters)

**Q24 What do you mean by logistic regression?**

Logistic Regression is basically a mathematical model that can be used to study datasets with one or more independent variables that determine a particular outcome. By studying the relationship between multiple independent variables, the model predicts a dependent data variable.

**Q25 What do you mean by the K-means algorithm?**

One of the most famous partitioning methods is K-mean. With this unsupervised learning algorithm, the unlabeled data is grouped in clusters. Here, 'k' indicates the number of clusters. It tries to keep each cluster separated from the other. Since it is an unsupervised model, there will be no labels for the clusters to work with.





**Q26 Write the difference between variance and covariance.**

**Variance:** In statistics, variance is defined as the deviation of a data set from its mean value or average value. When the variances are greater, the numbers in the data set are farther from the mean. When the variances are smaller, the numbers are nearer the mean. Variance is calculated as follows:

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$



Here, X represents an individual data point, U represents the average of multiple data points, and N represents the total number of data points.

**Covariance:** Covariance is another common concept in statistics, like variance. In statistics, covariance is a measure of how two random variables change when compared with each other. Covariance is calculated as follows:

$$\text{COV}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n-1}$$



Here, X represents the independent variable, Y represents the dependent variable, x-bar represents the mean of the X, y-bar represents the mean of the Y, and N represents the total number of data points in the sample.

**Q27 Name the statistical methods that are highly beneficial for data analysts?**

Accurate predictions and valuable results can only be achieved through the right statistical methods for analysis. Research well to find the leading ones used by the majority of analysts for varied tasks to deliver a reliable answer in the analyst interview questions.

- Bayesian method
- Markov process
- Simplex algorithm
- Imputation
- Spatial and cluster processes
- Rank statistics, percentile, outliers detection
- Mathematical optimization

In addition to this, there are various types of data analysis as well, which the data analysts use-

1. Descriptive
2. Inferential
3. Differences
4. Associative
5. Predictive

**Q28 What's the difference between a data lake and a data warehouse?**

- Imputation techniques
- Bayesian methodologies

The storage of data is a big deal. Companies that use big data have been in the news a lot lately, as they try to maximize its potential. Data storage is usually handled by traditional databases for the layperson. For storing, managing, and analyzing big data, companies use data warehouses and data lakes.

**Data Warehouse:** This is considered an ideal place to store all the data you gather from many sources. A data warehouse is a centralized repository of data where data from operational systems and other sources are stored. It is a standard tool for integrating data across the team- or department-silos in mid-and large-sized companies. It collects and manages data from varied sources to provide meaningful business insights. Data warehouses can be of the following types:

- **Enterprise data warehouse (EDW):** Provides decision support for the entire organization.
- **Operational Data Store (ODS):** Has functionality such as reporting sales data or employee data.

**Data Lake:** Data lakes are basically large storage device that stores raw data in their original format until they are needed. with its large amount of data, analytical performance and native integration are improved. It exploits data warehouses' biggest weakness: their incapacity to be flexible. In this, neither planning nor knowledge of data analysis is required; the analysis is assumed to happen later, on-demand.

#### **Q29 What should a data analyst do with missing or suspected data?**

In such a case, a data analyst needs to:

- Use data analysis strategies like deletion method, single imputation methods, and model-based methods to detect missing data.
- Prepare a validation report containing all information about the suspected or missing data.
- Scrutinize the suspicious data to assess their validity.
- Replace all the invalid data (if any) with a proper validation code.
- Model preparation for the missing data
- Predict the missing values

#### **Q30 Name the different data validation methods used by data analysts.**

There are many ways to validate datasets. Some of the most commonly used data validation methods by Data Analysts include:

- **Field Level Validation** – In this method, data validation is done in each field as and when a user enters the data. It helps to correct the errors as you go.
- **Form Level Validation** – In this method, the data is validated after the user completes the form and submits it. It checks the entire data entry form at once, validates all the fields in it, and highlights the errors (if any) so that the user can correct it.
- **Data Saving Validation** – This data validation technique is used during the process of saving an actual file or database record. Usually, it is done when multiple data entry forms must be validated.
- **Search Criteria Validation** – This validation technique is used to offer the user accurate and related matches for their searched keywords or phrases. The main purpose of this validation method is to ensure that the user's search queries can return the most relevant results.

#### **Q31 Define “Collaborative Filtering”.**

Collaborative filtering is an algorithm that creates a recommendation system based on the behavioral data of a user. For instance, online shopping sites usually compile a list of items under “recommended for you” based on your browsing history and previous purchases. The crucial components of this algorithm include users, objects, and their interests. It is used to broaden the options the users could have. Online entertainment applications are another example of collaborative filtering. For example, Netflix shows recommendations basis the user's behavior. It follows various techniques, such as-

1. Memory-based approach
2. Model-based approach

**Q32 Mention the steps of a Data Analysis project.**

The core steps of a Data Analysis project include:

- The foremost requirement of a Data Analysis project is an in-depth understanding of the business requirements.
- The second step is to identify the most relevant data sources that best fit the business requirements and obtain the data from reliable and verified sources.
- The third step involves exploring the datasets, cleaning the data, and organizing the same to gain a better understanding of the data at hand.
- In the fourth step, Data Analysts must validate the data.
- The fifth step involves implementing and tracking the datasets.
- The final step is to create a list of the most probable outcomes and iterate until the desired results are accomplished.

The whole meaning of the data analysis is to help in effective decision-making. The data analysis projects are the steps towards achieving it. For example, while undergoing the above-said process, the analysts use the past data and once the data has been analysed it gets put in a presentable form so the decision-making process can be smoother.

**Q33 What are the problems that a Data Analyst can encounter while performing data analysis?**

A critical data analyst interview question you need to be aware of. A Data Analyst can confront the following issues while performing data analysis:

- Presence of duplicate entries and spelling mistakes. These errors can hamper data quality.
- Poor quality data acquired from unreliable sources. In such a case, a Data Analyst will have to spend a significant amount of time in cleansing the data.
- Data extracted from multiple sources may vary in representation. Once the collected data is combined after being cleansed and organized, the variations in data representation may cause a delay in the analysis process.
- Incomplete data is another major challenge in the data analysis process. It would inevitably lead to erroneous or faulty results.

**Q34 Explain descriptive, predictive, and prescriptive analytics.**

<b>Descriptive</b>	<b>Predictive</b>	<b>Prescriptive</b>
It provides insights into the past to answer “what has happened”	Understands the future to answer “what could happen”	Suggest various courses of action to answer “what should you do”
Uses data aggregation and data mining techniques	Uses statistical models and forecasting techniques	Uses simulation algorithms and optimization techniques to advise possible outcomes

Example: An ice cream company can analyze how much ice cream was sold, which flavors were sold, and whether more or less ice cream was sold than the day before	Example: An ice cream company can analyze how much ice cream was sold, which flavors were sold, and whether more or less ice cream was sold than the day before	Example: Lower prices to increase the sale of ice creams, produce more/fewer quantities of a specific flavor of ice cream
-----------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------

### Q35 Describe univariate, bivariate, and multivariate analysis.

The **univariate analysis** is the simplest and easiest form of data analysis where the data being analyzed contains only one variable.

Example - Studying the heights of players in the NBA.

Univariate analysis can be described using Central Tendency, Dispersion, Quartiles, Bar charts, Histograms, Pie charts, and Frequency distribution tables.

The **bivariate analysis** involves the analysis of two variables to find causes, relationships, and correlations between the variables.

Example – Analyzing the sale of ice creams based on the temperature outside.

The bivariate analysis can be explained using Correlation coefficients, Linear regression, Logistic regression, Scatter plots, and Box plots.

The **multivariate analysis** involves the analysis of three or more variables to understand the relationship of each variable with the other variables.

Example – Analysing Revenue based on expenditure.

Multivariate analysis can be performed using Multiple regression, Factor analysis, Classification & regression trees, Cluster analysis, Principal component analysis, Dual-axis charts, etc.

### Q36 What are the different types of Hypothesis testing?

Hypothesis testing is the procedure used by statisticians and scientists to accept or reject statistical hypotheses. There are mainly two types of hypothesis testing:

- Null hypothesis: It states that there is no relation between the predictor and outcome variables in the population.  $H_0$  denoted it.

Example: There is no association between a patient's BMI and diabetes.

- Alternative hypothesis: It states that there is some relation between the predictor and outcome variables in the population. It is denoted by  $H_1$ .

Example: There could be an association between a patient's BMI and diabetes.

Explain the Type I and Type II errors in Statistics?

In Hypothesis testing, a Type I error occurs when the null hypothesis is rejected even if it is true. It is also known as a false positive.

A Type II error occurs when the null hypothesis is not rejected, even if it is false. It is also known as a false negative.

## EXCEL INTERVIEW QUESTIONS

**Q1 What is the difference between COUNT, COUNTA, COUNTBLANK, and COUNTIF in Excel?**

- COUNT function returns the count of numeric cells in a range
- COUNTA function counts the non-blank cells in a range
- COUNTBLANK function gives the count of blank cells in a range
- COUNTIF function returns the count of values by checking a given condition

**Q2 How do you make a dropdown list in MS Excel?**

- First, click on the Data tab that is present in the ribbon.
- Under the Data Tools group, select Data Validation.
- Then navigate to Settings > Allow > List.
- Select the source you want to provide as a list array.

**Q3 How Many Data Formats Are Available in Excel?**

There are six data formats available in Excel:

- Excel workbook: .xlsx
- Excel macro-enabled workbook: .xlsm
- Excel binary workbook: .xlsb
- Template: .xltx
- Template (code): .xltm
- XML data: .xml

**Q4 What is the function to find the day of the week for a particular date value?**

The get the day of the week, you can use the WEEKDAY() function.

	A	B
1	Date	Day of Week
2	17-12-2021	=WEEKDAY(A2)

The above function will return 6 as the result, i.e., 17th December is a Saturday.  
need to find things in a table or a range by row.

**Q5 What function would you use to get the current date and time in Excel?**

In Excel, you can use the TODAY() and NOW() function to get the current date and time.

=TODAY()	➡	08-05-2020
=NOW()	➡	08-05-2020 18:46

**Q6 Using the below sales table, calculate the total quantity sold by sales representatives whose name starts with A, and the cost of each item they have sold is greater than 10.**

	A	B	C	D	E	F	G
1	Date	Sales Rep	Item	Cost each	Quantity	Sale total	Cost range
2	05-07-2005	A. Yamamoto	J21344A	₹ 19.50	4	\$ 78.00	Low
3	06-07-2005	Q. Ackerman	Q00345B	₹ 39.00	19	\$ 741.00	High
4	26-07-2005	J. Wilson	L98700F	₹ 8.25	2	\$ 16.50	Low
5	12-07-2005	F. Rosenstein	B20011A	₹ 22.15	19	\$ 420.85	Average
6	27-07-2005	J. Wilson	J21344A	₹ 19.50	19	\$ 370.50	Low
7	28-07-2005	Q. Ackerman	Q00345B	₹ 39.00	19	\$ 741.00	High
8	23-07-2005	A. Yamamoto	C55440D	₹ 16.75	2	\$ 33.50	Low
9	01-07-2005	F. Rosenstein	Q00345B	₹ 39.00	5	\$ 195.00	High
10	24-07-2005	A. Mathews	L98700F	₹ 8.25	15	\$ 123.75	Low
11	29-07-2005	D.F. Chang	J21344A	₹ 19.50	11	\$ 214.50	Low
12	21-07-2005	F. Rosenstein	B20011A	₹ 22.15	15	\$ 332.25	Average
13	06-07-2005	J. Wilson	Q00345B	₹ 39.00	18	\$ 702.00	High
14	18-07-2005	S. Muller	C55440D	₹ 16.75	17	\$ 284.75	Low
15	03-07-2005	O. McBride	J21344A	₹ 19.50	9	\$ 175.50	Low
16	28-07-2005	L. Sanchez	J21344A	₹ 19.50	15	\$ 292.50	Low
17	22-07-2005	F. Rosenstein	L98700F	₹ 8.25	11	\$ 90.75	Low
18	04-07-2005	J. Wilson	C55440D	₹ 16.75	4	\$ 67.00	Low
19	07-07-2005	W. Carver	B20011A	₹ 22.15	3	\$ 66.45	Average
20	03-07-2005	A. Symonds	J21344A	₹ 19.50	7	\$ 136.50	Low

You can use the SUMIFS() function to find the total quantity.

For the Sales Rep column, you need to give the criteria as "A\*" - meaning the name should start with the letter "A". For the Cost each column, the criteria should be ">10" - meaning the cost of each item is greater than 10.

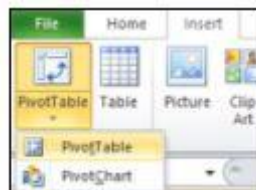
	A	B	C	D	E	F	G	H	I	J
1	Date	Sales Rep	Item	Cost each	Quantity	Sale total	Cost range			
2	05-07-2005	A. Yamamoto	J21344A	₹ 19.50	4	\$ 78.00	Low			
3	06-07-2005	Q. Ackerman	Q00345B	₹ 39.00	19	\$ 741.00	High			
4	26-07-2005	J. Wilson	L98700F	₹ 8.25	2	\$ 16.50	Low			
5	12-07-2005	F. Rosenstein	B20011A	₹ 22.15	19	\$ 420.85	Average			
6	27-07-2005	J. Wilson	J21344A	₹ 19.50	19	\$ 370.50	Low			
7	28-07-2005	Q. Ackerman	Q00345B	₹ 39.00	19	\$ 741.00	High			
8	23-07-2005	A. Yamamoto	C55440D	₹ 16.75	2	\$ 33.50	Low			
9	01-07-2005	F. Rosenstein	Q00345B	₹ 39.00	5	\$ 195.00	High	=SUMIFS(E2:E20,B2:B20,"A*",D2:D20,">10")		
10	24-07-2005	A. Mathews	L98700F	₹ 8.25	15	\$ 123.75	Low			
11	29-07-2005	D.F. Chang	J21344A	₹ 19.50	11	\$ 214.50	Low			
12	21-07-2005	F. Rosenstein	B20011A	₹ 22.15	15	\$ 332.25	Average			
13	06-07-2005	J. Wilson	Q00345B	₹ 39.00	18	\$ 702.00	High			
14	18-07-2005	S. Muller	C55440D	₹ 16.75	17	\$ 284.75	Low			
15	03-07-2005	O. McBride	J21344A	₹ 19.50	9	\$ 175.50	Low			
16	28-07-2005	L. Sanchez	J21344A	₹ 19.50	15	\$ 292.50	Low			
17	22-07-2005	F. Rosenstein	L98700F	₹ 8.25	11	\$ 90.75	Low			
18	04-07-2005	J. Wilson	C55440D	₹ 16.75	4	\$ 67.00	Low			
19	07-07-2005	W. Carver	B20011A	₹ 22.15	3	\$ 66.45	Average			
20	03-07-2005	A. Symonds	J21344A	₹ 19.50	7	\$ 136.50	Low			



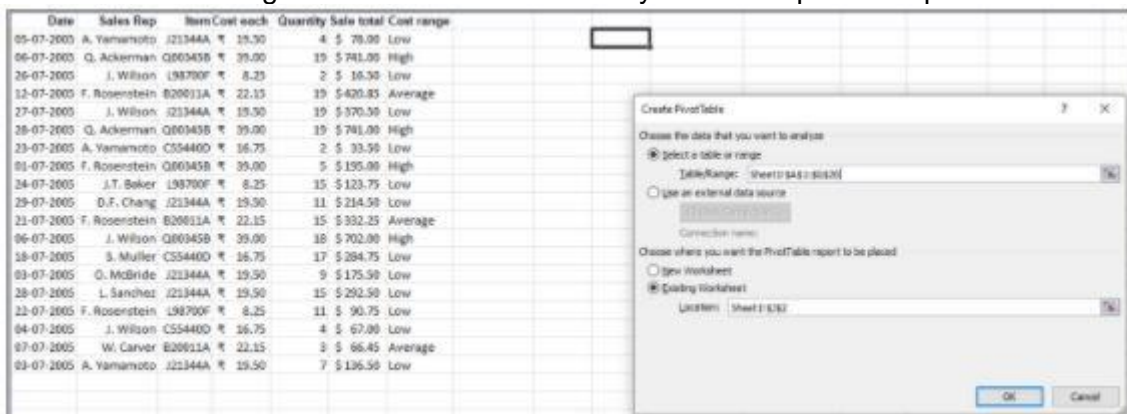
**Q7 Using the data given below, create a pivot table to find the total sales made by each sales representative for each item. Display the sales as % of the grand total.**

	A	B	C	D	E	F	G
1	Date	Sales Rep	Item	Cost each	Quantity	Sale total	Cost range
2	05-07-2005	A. Yamamoto	J21344A	₹ 19.50	4	\$ 78.00	Low
3	06-07-2005	Q. Ackerman	Q00345B	₹ 39.00	19	\$ 741.00	High
4	26-07-2005	J. Wilson	L98700F	₹ 8.25	2	\$ 16.50	Low
5	12-07-2005	F. Rosenstein	B20011A	₹ 22.15	19	\$ 420.85	Average
6	27-07-2005	J. Wilson	J21344A	₹ 19.50	19	\$ 370.50	Low
7	28-07-2005	Q. Ackerman	Q00345B	₹ 39.00	19	\$ 741.00	High
8	23-07-2005	A. Yamamoto	C55440D	₹ 16.75	2	\$ 33.50	Low
9	01-07-2005	F. Rosenstein	Q00345B	₹ 39.00	5	\$ 195.00	High
10	24-07-2005	J.T. Baker	L98700F	₹ 8.25	15	\$ 123.75	Low
11	29-07-2005	D.F. Chang	J21344A	₹ 19.50	11	\$ 214.50	Low
12	21-07-2005	F. Rosenstein	B20011A	₹ 22.15	15	\$ 332.25	Average
13	06-07-2005	J. Wilson	Q00345B	₹ 39.00	18	\$ 702.00	High
14	18-07-2005	S. Muller	C55440D	₹ 16.75	17	\$ 284.75	Low
15	03-07-2005	O. McBride	J21344A	₹ 19.50	9	\$ 175.50	Low
16	28-07-2005	L. Sanchez	J21344A	₹ 19.50	15	\$ 292.50	Low
17	22-07-2005	F. Rosenstein	L98700F	₹ 8.25	11	\$ 90.75	Low
18	04-07-2005	J. Wilson	C55440D	₹ 16.75	4	\$ 67.00	Low
19	07-07-2005	W. Carver	B20011A	₹ 22.15	3	\$ 66.45	Average
20	03-07-2005	A. Yamamoto	J21344A	₹ 19.50	7	\$ 136.50	Low

- Select the entire table range, click on the Insert tab and choose PivotTable



- Select the table range and the worksheet where you want to place the pivot table

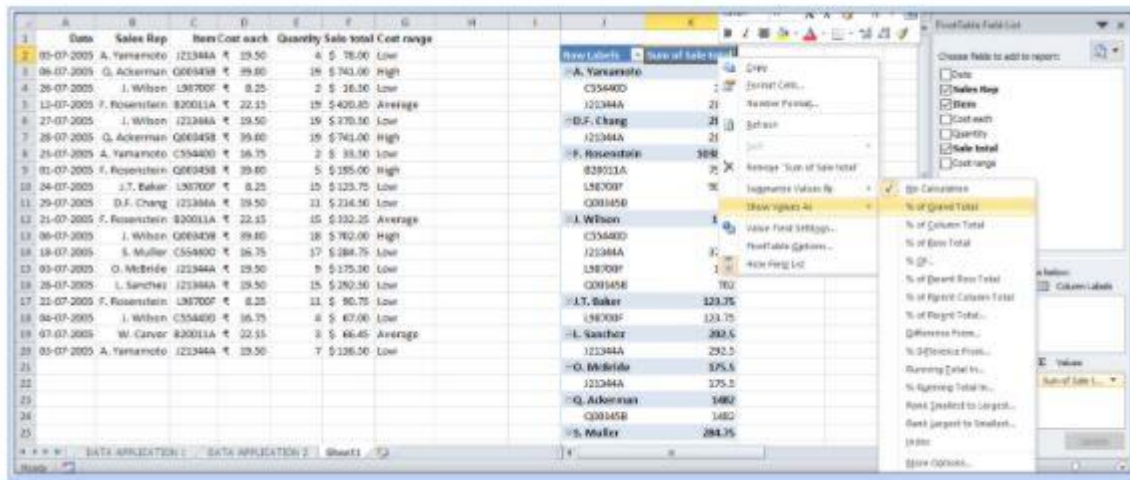


- Drag Sale total on to Values, and Sales Rep and Item on to Row Labels. It will give the sum of sales made by each representative for every item they have sold.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Date	Sales Rep	Item	Cost each	Quantity	Sale total	Cost range						
2	05-07-2005	A. Yamamoto	J21344A	₹ 19.50	4	\$ 78.00	Low						
3	06-07-2005	Q. Ackerman	Q00345B	₹ 39.00	19	\$ 741.00	High						
4	26-07-2005	J. Wilson	L98700F	₹ 8.25	2	\$ 16.50	Low						
5	12-07-2005	F. Rosenstein	B20011A	₹ 22.15	19	\$ 420.85	Average						
6	27-07-2005	J. Wilson	J21344A	₹ 19.50	19	\$ 370.50	Low						
7	28-07-2005	Q. Ackerman	Q00345B	₹ 39.00	19	\$ 741.00	High						
8	23-07-2005	A. Yamamoto	C55440D	₹ 16.75	2	\$ 33.50	Low						
9	01-07-2005	F. Rosenstein	Q00345B	₹ 39.00	5	\$ 195.00	High						
10	24-07-2005	J.T. Baker	L98700F	₹ 8.25	15	\$ 123.75	Low						
11	29-07-2005	D.F. Chang	J21344A	₹ 19.50	11	\$ 214.50	Low						
12	21-07-2005	F. Rosenstein	B20011A	₹ 22.15	15	\$ 332.25	Average						
13	06-07-2005	J. Wilson	Q00345B	₹ 39.00	18	\$ 702.00	High						
14	18-07-2005	S. Muller	C55440D	₹ 16.75	17	\$ 284.75	Low						
15	03-07-2005	O. McBride	J21344A	₹ 19.50	9	\$ 175.50	Low						
16	28-07-2005	L. Sanchez	J21344A	₹ 19.50	15	\$ 292.50	Low						
17	22-07-2005	F. Rosenstein	L98700F	₹ 8.25	11	\$ 90.75	Low						
18	04-07-2005	J. Wilson	C55440D	₹ 16.75	4	\$ 67.00	Low						
19	07-07-2005	W. Carver	B20011A	₹ 22.15	3	\$ 66.45	Average						
20	03-07-2005	A. Yamamoto	J21344A	₹ 19.50	7	\$ 136.50	Low						
21													
22													
23													
24													
25													
26													



- Right-click on “Sum of Sale Total” and expand Show Values As to select % of Grand Total.



- Below is the resultant pivot table.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Date	Sales Rep	Item Cost each	Quantity	Sale total	Cost range							
2	05-07-2005	A. Yamamoto	121344A	2	5 78.00	Low							
3	06-07-2005	Q. Ackerman	Q00345B	19	5 743.00	High							
4	26-07-2005	J. Wilson	138700F	2	5 16.30	Low							
5	12-07-2005	F. Rosenstein	R20011A	22.15	5 420.85	Average							
6	27-07-2005	J. Wilson	121344A	19	5 370.50	Low							
7	26-07-2005	Q. Ackerman	Q00345B	19	5 743.00	High							
8	25-07-2005	A. Yamamoto	C55440D	16.75	2 5 33.50	Low							
9	01-07-2005	F. Rosenstein	Q00345B	19	5 5195.00	High							
10	24-07-2005	J.T. Baker	138700F	8.25	15 5 123.75	Low							
11	29-07-2005	D.F. Chang	121344A	19	5 214.50	Low							
12	21-07-2005	F. Rosenstein	R20011A	22.15	15 5 332.25	Average							
13	06-07-2005	J. Wilson	Q00345B	19	5 703.00	High							
14	18-07-2005	S. Muller	C55440D	16.75	17 5 284.75	Low							
15	03-07-2005	O. McBride	121344A	19	5 175.50	Low							
16	26-07-2005	L. Sanchez	121344A	19	5 292.50	Low							
17	22-07-2005	F. Rosenstein	138700F	8.25	11 5 90.75	Low							
18	04-07-2005	J. Wilson	C55440D	16.75	4 5 67.00	Low							
19	07-07-2005	W. Carver	R20011A	22.15	3 5 66.45	Average							
20	03-07-2005	A. Yamamoto	121344A	19	7 5 136.50	Low							
21													
22													
23													
24													
25													
26													

## Q8 What do you mean by clustering algorithms? Write different properties of clustering algorithms?

Clustering is the process of categorizing data into groups and clusters. In a dataset, it identifies similar data groups. It is the technique of grouping a set of objects so that the objects within the same cluster are similar to one another rather than to those located in other clusters. When implemented, the clustering algorithm possesses the following properties:

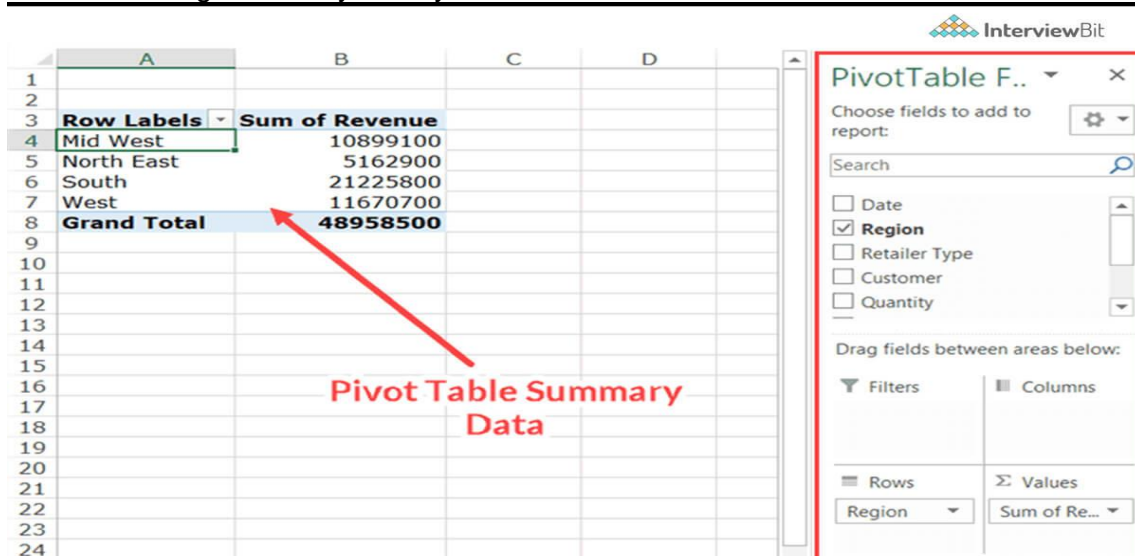
- Flat or hierarchical
- Hard or Soft
- Iterative
- Disjunctive

## Q9 What is a Pivot table? Write its usage.

One of the basic tools for data analysis is the Pivot Table. With this feature, you can quickly summarize large datasets in Microsoft Excel. Using it, we can turn columns into rows and rows into columns. Furthermore, it permits grouping by any field (column) and applying advanced calculations to them. It is an extremely easy-to-use program since you just drag and drop rows/columns headers to build a report. Pivot tables consist of four different sections:

- Value Area:** This is where values are reported.
- Row Area:** The row areas are the headings to the left of the values.

- **Column Area:** The headings above the values area make up the column area.
- **Filter Area:** Using this filter you may drill down in the data set.



#### Q10 Define Excel Charts.

A chart in Excel is a feature that allows you to display data through a range of visually intuitive graphs. These charts and graphs can make it easier and quicker to comprehend data compared to just looking at the numbers on the worksheet. Available charts on Excel include:

- Bar graphs
- Line graphs
- Pie charts
- Area graph
- Scatter graphs
- Surface graphs
- Doughnut graphs
- Radar charts

#### VLOOKUP accepts the following four parameters:

- lookup\_value - The value to look for in the first column of a table
- table - The table from where you can extract value
- col\_index - The column from which to extract value
- range\_lookup - [optional] TRUE = approximate match (default). FALSE = exact match
- Let's understand VLOOKUP with an example.

	A	B	C	D	E
1	First Name	Last Name	Department	City	Date Hired
2	Ben	Zampa	HR	Chicago	10-11-2001
3	Stuart	Carry	Marketing	Kansas	20-06-2002
4	Jenson	Button	Operations	New York	01-12-2004
5	Lucy	Davis	Sales	Los Angeles	25-02-2011
6	Trent	Patinson	IT	Boston	17-08-2015
7	Jhonny	Evans	Sales	Houston	10-01-2018

- If you wanted to find the department to which Stuart belongs to, you could use the VLOOKUP function as shown below:

9	Vlookup				
10	First Name	Last Name	Department	City	Date Hired
11	Stuart		=VLOOKUP(A11,A2:E7,3,0)		

- Here, A11 cell has the lookup value, A2:E7 is the table array, 3 is the column index number with information about departments, and 0 is the range lookup.
- If you hit enter, it will return "Marketing", indicating that Stuart is from the marketing department

#### Q11 What Is VLOOKUP?

E4		fx		=VLOOKUP(D4,B8:F17,4,FALSE)	
	A	B	C	D	E
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					

Lookup value →

ID	Email
622	j.adder@ace.com

Result

1 2 3 4 5

ID	First	Last	Email	Department
610	Janet	Farley	j.farley@ace.com	Fulfillment
798	Steven	Batista	s.batista@ace.com	Sales
841	Evelyn	Monet	e.monet@ace.com	Fulfillment
886	Marilyn	Bradley	m.bradley@ace.com	Fulfillment
622	Jonathan	Adder	j.adder@ace.com	Marketing
601	Adrian	Birt	a.birt@ace.com	Engineering
869	Julie	Irons	j.irons@ace.com	Marketing
867	Erica	Tan	e.tan@ace.com	Fulfillment
785	Harold	Clayton	h.clayton@ace.com	Fulfillment
648	Sharyn	Castor	s.castor@ace.com	Support

Lookup column → Result column

EXCELJET

VLOOKUP is a predetermined function in Excel that allows the user to find data within a table corresponding to a particular row.

For instance, say you have a table of employee information that includes (from column A onward) employee ID, employee name, start date, hours per week, and salary. With VLOOKUP you can specify a row from the first column (i.e an employee number) and look up corresponding data from other columns, like the salary of the employee with that employee ID.

#### Q12 How Do You Use VLOOKUP?

The VLOOKUP syntax is composed of the lookup value, the range of data in which the lookup value is located, and the column number within this range that contains the desired return value. You can also specify whether you want an approximate match or an exact match to be returned, but this step is optional.

In other words, you must first indicate the cell reference of the value you would like to search for. Next, indicate the range of data you would like to search for (this will often be the entire table). You can then specify the column that contains the information you seek and input it as a number (the right-most column selected will be column 1).

To indicate whether the return value should be approximate or exact, finish the formula with TRUE (for approximate) or FALSE (for exact). An example formula would look like this: =VLOOKUP(A7,A1:E10,5,FALSE).

**Q13 What Is the Default Value of the Last Parameter of VLOOKUP?**

If the last parameter is not specified via TRUE or FALSE, the return value will default to TRUE (approximate), and show an approximate match for your request.

**Q14 What Is the Main Limitation of the VLOOKUP Function?**

The VLOOKUP function can only move in one direction, from left to right. Therefore, the information you wish to seek out must be located in a column to the right of the lookup value's location.

In newer versions of Excel, a successor to VLOOKUP has been added, called XLOOKUP. This new function works in any direction and defaults to exact matches rather than approximate. At some point in the future, XLOOKUP will completely replace VLOOKUP, but this will not happen until the majority of users have moved away from using older versions of Excel.

**Q15 Does VLOOKUP Look Up Case-Sensitive Values?**

VLOOKUP is not case-sensitive, and will always return the first value of the match irrespective of the case. In other words, the name Apgar and the acronym APGAR would be viewed as the same by VLOOKUP.

It is, however, possible to manipulate VLOOKUP into returning case-sensitive values by using a helper column.

**Q16 How Do the INDEX and MATCH Functions Work?**

The MATCH function returns the position of a cell in a row or column based on its value and the INDEX function will return the value of a cell based on the cell reference.

**Q17 How Do the INDEX and MATCH Functions Work Together?**

You can use two MATCH functions within an INDEX formula to specify a cell reference and return the value of that cell. The dynamic formula will return the corresponding data of any two MATCH values you input.

For example, if you have a table detailing the price per unit and the number of units sold for a variety of products, you can use the match index function to return a specific piece of information about a specific product.

**Q18 What Is Conditional Formatting? How Can It Be Used?**

Conditional formatting will apply various formatting types onto cells based on specified conditions. For instance, it can be used to apply a highlight to any duplicate cells, or cells with a numeric value under 5.

**Five Excel interview questions for data analysts**

Here are five more questions specific to Excel that you might be asked during your interview:

1. What is a VLOOKUP, and what are its limitations?
2. What is a pivot table, and how do you make one?
3. How do you find and remove duplicate data?
4. What are INDEX and MATCH functions, and how do they work together?
5. What's the difference between a function and a formula?

# SQL INTERVIEW QUESTIONS

## Q1 What is Data?

Data is a collection of a distinct small unit of information. It can be used in a variety of forms like text, numbers, media, bytes, etc. it can be stored in pieces of paper or electronic memory, etc.

Word 'Data' is originated from the word 'datum' that means 'single piece of information.' It is plural of the word datum.

In computing, Data is information that can be translated into a form for efficient movement and processing. Data is interchangeable.

## Q2 What is Database?

A database is an organized collection of data, stored and retrieved digitally from a remote or local computer system. Databases can be vast and complex, and such databases are developed using fixed design and modeling approaches.

A database is a systematic collection of data. They support electronic storage and manipulation of data. Databases make data management easy.

Let us discuss a database example: An online telephone directory uses a database to store data of people, phone numbers, and other contact details. Your electricity service provider uses a database to manage billing, client-related issues, handle fault data, etc.

Let us also consider Facebook. It needs to store, manipulate, and present data related to members, their friends, member activities, messages, advertisements, and a lot more. We can provide a countless number of examples for the usage of databases.

## Q3 What is a Datawarehouse?

Datawarehouse refers to a central repository of data where the data is assembled from multiple sources of information. Those data are consolidated, transformed and made available for the mining as well as online processing. Warehouse data also have a subset of data called Data Marts.

## Q4 What is DBMS?

DBMS stands for Database Management System. DBMS is a system software responsible for the creation, retrieval, updation, and management of the database. It ensures that our data is consistent, organized, and is easily accessible by serving as an interface between the database and its end-users or application software.

## Q5 What is RDBMS? How is it different from DBMS?

RDBMS stands for Relational Database Management System. The key difference [here](#), compared to DBMS, is that RDBMS stores data in the form of a collection of tables, and relations can be defined between the common fields of these tables. Most modern database management systems like MySQL, Microsoft SQL Server, Oracle, IBM DB2, and Amazon Redshift are based on RDBMS.



DBMS	RDBMS
DBMS applications store <b>data as file</b> .	RDBMS applications store <b>data in a tabular form</b> .
In DBMS, data is generally stored in either a hierarchical form or a navigational form.	In RDBMS, the tables have an identifier called primary key and the data values are stored in the form of tables.
<b>Normalization is not</b> present in DBMS.	<b>Normalization is</b> present in RDBMS.
DBMS does <b>not apply any security</b> with regards to data manipulation.	RDBMS <b>defines the integrity constraint</b> for the purpose of ACID (Atomocity, Consistency, Isolation and Durability) property.
DBMS uses file system to store data, so there will be <b>no relation between the tables</b> .	in RDBMS, data values are stored in the form of tables, so a <b>relationship</b> between these data values will be stored in the form of a table as well.
DBMS has to provide some uniform methods to access the stored information.	RDBMS system supports a tabular structure of the data and a relationship between them to access the stored information.
DBMS <b>does not support distributed database</b> .	RDBMS <b>supports distributed database</b> .
DBMS is meant to be for small organization and <b>deal with small data</b> . it supports <b>single user</b> .	RDBMS is designed to <b>handle large amount of data</b> . it supports <b>multiple users</b> .
Examples of DBMS are file systems, <b>xml</b> etc.	Example of RDBMS are <b>mysql, postgre, sql server, oracle</b> etc.

#### Q6 What is SQL?

SQL stands for Structured Query Language. It is the standard language for relational database management systems. It is especially useful in handling organized data comprised of entities (variables) and relations between different entities of the data.

#### Q7 What is the difference between SQL and MySQL?

SQL is a standard language for retrieving and manipulating structured databases. On the contrary, MySQL is a relational database management system, like SQL Server, Oracle or IBM DB2, that is used to manage SQL databases.

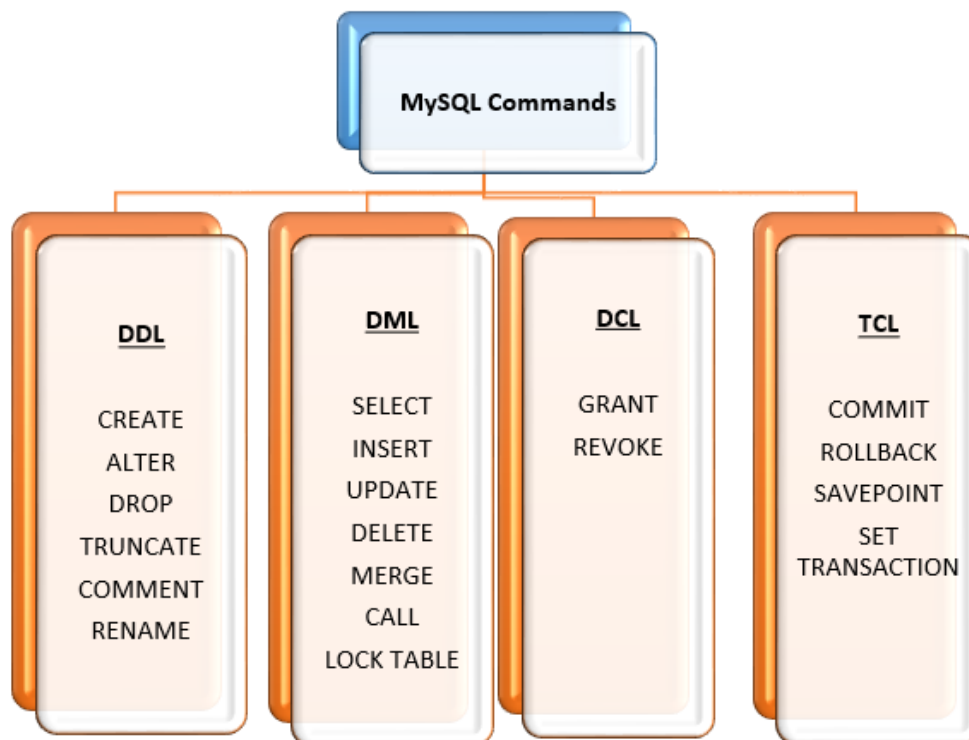




SQL	MySQL
SQL is a query programming language that manages RDBMS.	MySQL is a relational database management system that uses SQL.
SQL is primarily used to query and operate database systems.	MySQL allows you to handle, store, modify and delete data and store data in an organized way.
SQL does not support any connector.	MySQL comes with an in-built tool known as MySQL Workbench that facilitates creating, designing, and building databases.
SQL follows a simple standard format without many or regular updates.	MySQL has numerous variants and gets frequent updates.
SQL supports only a single storage engine.	MySQL offers support for multiple storage engines along with plug-in storage, making it more flexible.
SQL does not allow other processors or even its own binaries to manipulate data during execution.	MySQL is less secure than SQL, as it allows third-party processors to manipulate data files during execution.



## Q8 What are the subsets of SQL?



The following are the four significant subsets of the SQL:

- **Data definition language (DDL)**: It defines the data structure that consists of commands like CREATE, ALTER, DROP, etc.
- **Data manipulation language (DML)**: It is used to manipulate existing data in the database. The commands in this category are SELECT, UPDATE, INSERT, etc.
- **Data control language (DCL)**: It controls access to the data stored in the database. The commands in this category include GRANT and REVOKE.
- **Transaction Control Language (TCL)**: It is used to deal with the transaction operations in the database. The commands in this category are COMMIT, ROLLBACK, SET TRANSACTION, SAVEPOINT, etc.

## Q9 What is the purpose of DDL Language?

DDL stands for Data definition language. It is the subset of a database that defines the data structure of the database when the database is created. **For example**, we can use the DDL commands to add, remove, or modify tables. It consists of the following commands: CREATE, ALTER and DELETE database objects such as schema, tables, indexes, view, sequence, etc.

### Example

```
CREATE TABLE Students (  
Roll_no INT,  
Name VARCHAR(45),  
Branch VARCHAR(30),  
);
```

### Q9 What is the purpose of DML Language?

Data manipulation language makes the user able to retrieve and manipulate data in a relational database. The DML commands can only perform read-only operations on data. We can perform the following operations using DDL language:

- Insert data into the database through the INSERT command.
- Retrieve data from the database through the SELECT command.
- Update data in the database through the UPDATE command.
- Delete data from the database through the DELETE command.

#### Example

```
INSERT INTO Student VALUES (111, 'George', 'Computer Science')
```

### Q10 What is the purpose of DCL Language?

Data control language allows users to control access and permission management to the database. It is the subset of a database, which decides that what part of the database should be accessed by which user at what point of time. It includes two commands, GRANT and REVOKE.

**GRANT:** It enables system administrators to assign privileges and roles to the specific user accounts to perform specific tasks on the database.

**REVOKE:** It enables system administrators to revoke privileges and roles from the user accounts so that they cannot use the previously assigned permission on the database.

#### Example

```
GRANT * ON mydb.Student TO javatpoint@localhsot;
```

### Q11 What is a Query?

An SQL query is used to retrieve the required data from the database. However, there may be multiple SQL queries that yield the same results but with different levels of efficiency. An inefficient query can drain the database resources, reduce the database speed or result in a loss of service for other users. So it is very important to optimize the query to obtain the best database performance.

### Q12 What is a SubQuery?

In SQL a Subquery can be simply defined as a query within another query. In other words, we can say that a Subquery is a query that is embedded in the WHERE clause of another SQL query.

### Q13 What are Tables and Fields?

A table is an organized collection of data stored in the form of rows and columns. Columns can be categorized as vertical and rows as horizontal. The columns in a table are called fields while the rows can be referred to as records.

### Q14 What are Constraints in SQL?

Constraints are used to specify the rules concerning data in the table. It can be applied for single or multiple fields in an SQL table during the creation of the table or after creating using the ALTER TABLE command. The constraints are:

- **NOT NULL** - Restricts NULL value from being inserted into a column.
- **CHECK** - Verifies that all values in a field satisfy a condition.
- **DEFAULT** - Automatically assigns a default value if no value has been specified for the field.
- **UNIQUE** - Ensures unique values to be inserted into the field.
- **INDEX** - Indexes a field providing faster retrieval of records.
- **PRIMARY KEY** - Uniquely identifies each record in a table.
- **FOREIGN KEY** - Ensures referential integrity for a record in another table.

### Q15 What is a Primary Key?

The PRIMARY KEY constraint uniquely identifies each row in a table. It must contain UNIQUE values and has an implicit NOT NULL constraint.

A table in SQL is strictly restricted to have one and only one primary key, which is comprised of single or multiple fields (columns).

```
CREATE TABLE tableName (  
col1 int NOT NULL,  
col2 varchar(50) NOT NULL,  
col3 int,  
.....  
PRIMARY KEY (col1)  
);
```

### Q16 What is a UNIQUE constraint?

A UNIQUE constraint ensures that all values in a column are different. This provides uniqueness for the column(s) and helps identify each row uniquely. Unlike primary key, there can be multiple unique constraints defined per table. The code syntax for UNIQUE is quite similar to that of PRIMARY KEY and can be used interchangeably.

```
CREATE TABLE Persons (  
ID int NOT NULL UNIQUE,  
LastName varchar(255) NOT NULL,  
FirstName varchar(255),  
Age int  
);
```

### Q17 What is a Foreign Key?

A FOREIGN KEY comprises of single or collection of fields in a table that essentially refers to the PRIMARY KEY in another table. Foreign key constraint ensures referential integrity in the relation between two tables.

The table with the foreign key constraint is labeled as the child table, and the table containing the candidate key is labeled as the referenced or parent table.

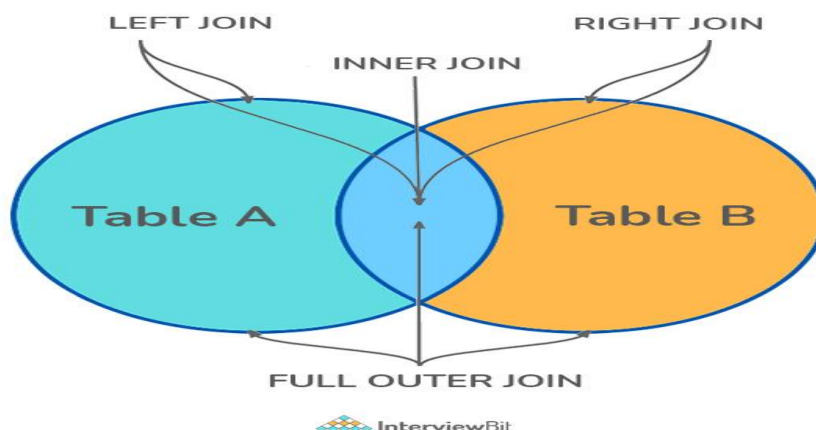
```
CREATE TABLE childTable (  
col1 int NOT NULL,  
col2 int NOT NULL,  
col3 int,  
.....  
PRIMARY KEY (col1),  
FOREIGN KEY (col3) REFERENCES parentTable(parent_Primary_key)  
);
```

### Q18 What is the difference between Primary key and Foreign key?

S.NO.	PRIMARY KEY	FOREIGN KEY
1	A primary key is used to ensure data in the specific column is unique.	A foreign key is a column or group of columns in a relational database table that provides a link between data in two tables.
2	It uniquely identifies a record in the relational database table.	It refers to the field in a table which is the primary key of another table.
3	Only one primary key is allowed in a table.	Whereas more than one foreign key are allowed in a table.
4	It is a combination of UNIQUE and Not Null constraints.	It can contain duplicate values and a table in a relational database.
5	It does not allow NULL values.	It can also contain NULL values.
6	Its value cannot be deleted from the parent table.	Its value can be deleted from the child table.
7	It constraint can be implicitly defined on the temporary tables.	It constraint cannot be defined on the local or global temporary tables.

### Q19 What is a Join? List its different types.

The **SQL Join** clause is used to combine records (rows) from two or more tables in a SQL database based on a related column between the two.



There are four different types of JOINS in SQL:

- **(INNER) JOIN:** Retrieves records that have matching values in both tables involved in the join. This is the widely used join for queries.

```
SELECT *  
FROM Table_A  
JOIN Table_B;
```

```
SELECT *  
FROM Table_A  
INNER JOIN Table_B;
```

- **LEFT (OUTER) JOIN:** Retrieves all the records/rows from the left and the matched records/rows from the right table.

```
SELECT *  
FROM Table_A A  
LEFT JOIN Table_B B  
ON A.col = B.col;
```

- **RIGHT (OUTER) JOIN:** Retrieves all the records/rows from the right and the matched records/rows from the left table.

```
SELECT *  
FROM Table_A A  
RIGHT JOIN Table_B B  
ON A.col = B.col;
```

- **FULL (OUTER) JOIN:** Retrieves all the records where there is a match in either the left or right table.

```
SELECT *  
FROM Table_A A  
FULL JOIN Table_B B  
ON A.col = B.col;
```

#### Q20 What is a Self-Join?

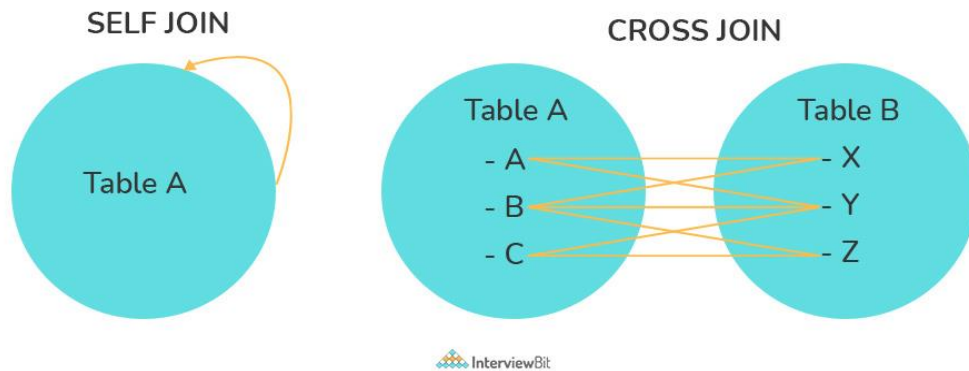
A **self JOIN** is a case of regular join where a table is joined to itself based on some relation between its own column(s). Self-join uses the INNER JOIN or LEFT JOIN clause and a table alias is used to assign different names to the table within the query.

```
SELECT A.emp_id AS "Emp_ID", A.emp_name AS "Employee",  
B.emp_id AS "Sup_ID", B.emp_name AS "Supervisor"  
FROM employee A, employee B  
WHERE A.emp_sup = B.emp_id;
```

#### Q21 What is a Cross-Join?

Cross join can be defined as a cartesian product of the two tables included in the join. The table after join contains the same number of rows as in the cross-product of the number of rows in the two tables. If a WHERE clause is used in cross join then the query will work like an INNER JOIN.

```
SELECT stu.name, sub.subject  
FROM students AS stu  
CROSS JOIN subjects AS sub;
```



## Q22 What is an Index? Explain its different types.

A database index is a data structure that provides a quick lookup of data in a column or columns of a table. It enhances the speed of operations accessing data from a database table at the cost of additional writes and memory to maintain the index data structure.

```
CREATE INDEX index_name /* Create Index */
ON table_name (column_1, column_2);
DROP INDEX index_name; /* Drop Index */
```

There are different types of indexes that can be created for different purposes:

- **Unique and Non-Unique Index:**

Unique indexes are indexes that help maintain data integrity by ensuring that no two rows of data in a table have identical key values. Once a unique index has been defined for a table, uniqueness is enforced whenever keys are added or changed within the index.

```
CREATE UNIQUE INDEX myIndex
ON students (enroll_no);
```

Non-unique indexes, on the other hand, are not used to enforce constraints on the tables with which they are associated. Instead, non-unique indexes are used solely to improve query performance by maintaining a sorted order of data values that are used frequently.

- **Clustered and Non-Clustered Index:**

Clustered indexes are indexes whose order of the rows in the database corresponds to the order of the rows in the index. This is why only one clustered index can exist in a given table, whereas, multiple non-clustered indexes can exist in the table.

The only difference between clustered and non-clustered indexes is that the database manager attempts to keep the data in the database in the same order as the corresponding keys appear in the clustered index.

Clustering indexes can improve the performance of most query operations because they provide a linear-access path to data stored in the database.

## Q23 What are the different types of SQL operators?

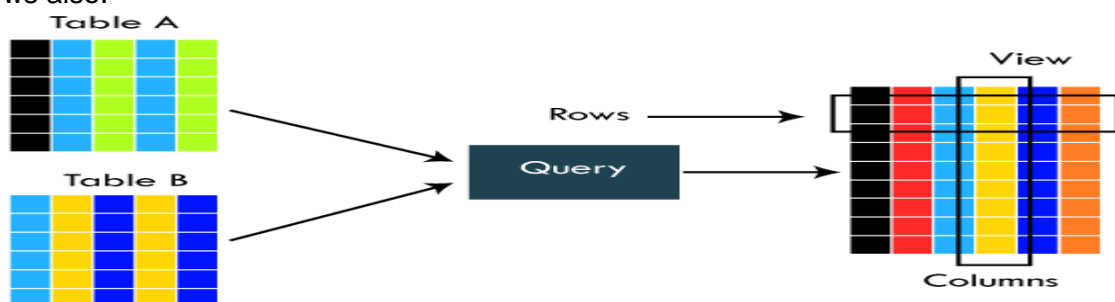
Operators are the special keywords or special characters reserved for performing particular operations. They are also used in SQL queries. We can primarily use these operators within the WHERE clause of SQL commands. It's a part of the command to filters data based on the specified condition. The SQL operators can be categorized into the following types:

- **Arithmetic operators:** These operators are used to perform mathematical operations on numerical data. The categories of this operators are addition (+), subtraction (-), multiplication (\*), division (/), remainder/modulus (%), etc.

- **Logical operators:** These operators evaluate the expressions and return their results in True or False. This operator includes ALL, AND, ANY, ISNULL, EXISTS, BETWEEN, IN, LIKE, NOT, OR, UNIQUE.
- **Comparison operators:** These operators are used to perform comparisons of two values and check whether they are the same or not. It includes equal to (=), not equal to (!= or <>), less than (<), greater than (>), less than or equal to (<=), greater than or equal to (>=), not less than (!<), not greater than (!>), etc.
- **Bitwise operators:** It is used to do bit manipulations between two expressions of integer type. It first performs conversion of integers into binary bits and then applied operators such as AND (& symbol), OR (|, ^), NOT (~), etc.
- **Compound operators:** These operators perform operations on a variable before setting the variable's result to the operation's result. It includes Add equals (+=), subtract equals (-=), multiply equals (\*=), divide equals (/=), modulo equals (%=), etc.
- **String operators:** These operators are primarily used to perform concatenation and pattern matching of strings. It includes + (String concatenation), += (String concatenation assignment), % (Wildcard), [] (Character(s) matches), [^] (Character(s) not to match), \_ (Wildcard match one character), etc.

#### Q24 What is a view in SQL?

A view is a database object that has no values. It is a virtual table that contains a subset of data within a table. It looks like an actual table containing rows and columns, but it takes less space because it is not present physically. It is operated similarly to the base table but does not contain any data of its own. Its name is always unique. A view can have data from one or more tables. If any changes occur in the underlying table, the same changes reflected in the views also.



The primary use of a view is to implement the security mechanism. It is the searchable object where we can use a query to search the view as we use for the table. It only shows the data returned by the query that was declared when the view was created.

We can create a view by using the following syntax:

```
CREATE VIEW view_name AS
SELECT column_lists FROM table_name
WHERE condition;
```



**Q25 What are the differences between SQL, MySQL, and SQL Server?**

The following comparison chart explains their main differences:

SQL	MySQL	SQL Server
SQL or Structured Query Language is useful for managing our relational databases. It is used to query and operate the database.	MySQL is the popular database management system used for managing the relational database. It is a fast, scalable, and easy-to-use database.	SQL Server is an RDBMS database system mainly developed for the Windows system to store, retrieve, and access data requested by the developer.
SQL first appeared in 1974.	MySQL first appeared on May 23, 1995.	SQL Server first appeared on April 24, 1989.
SQL was developed by IBM Corporation.	MySQL was developed by Oracle Corporation.	SQL Server was developed by Microsoft Company.
SQL is a query language for managing databases.	MySQL is database software that uses SQL language to conduct with the database.	SQL Server is also a software that uses SQL language to conduct with the database.
SQL has no variables.	MySQL can use variables constraints and data types.	SQL Server can use variables constraints and data types.
SQL is a programming language, so that it does not get any updates. Its commands are always fixed and remain the same.	MySQL is software, so it gets frequent updation.	SQL Server is also software, so it gets frequent updation.

**Q26 What is the difference between SQL and PL/SQL?**

The following comparison chart explains their main differences:

SQL	PL/SQL
SQL is a database structured query language used to communicate with relational databases. It was developed by IBM Corporations and first appeared in 1974.	PL/SQL or Procedural Language/Structured Query Language is a dialect of SQL used to enhance the capabilities of SQL. Oracle Corporation developed it in the early 90's. It uses SQL as its database language.
SQL is a declarative and data-oriented language.	PL/SQL is a procedural and application-oriented language.
SQL has no variables.	PL/SQL can use variables constraints and data types.

SQL can execute only a single query at a time.	PL/SQL can execute a whole block of code at once.
SQL query can be embedded in PL/SQL.	PL/SQL cannot be embedded in SQL as SQL does not support any programming language and keywords.
SQL can directly interact with the database server.	PL/SQL cannot directly interact with the database server.
SQL is like the source of data that we need to display.	PL/SQL provides a platform where SQL data will be shown.

#### Q27 Is it possible to sort a column using a column alias?

Yes. We can use the alias method in the ORDER BY instead of the WHERE clause for sorting a column.

#### Q28 What is the difference between IN and BETWEEN operators?

The following comparison chart explains their main differences:

BETWEEN Operator	IN Operator
This operator is used to select the range of data between two values. The values can be numbers, text, and dates as well.	It is a logical operator to determine whether or not a specific value exists within a set of values. This operator reduces the use of multiple OR conditions with the query.
It returns records whose column value lies in between the defined range.	It compares the specified column's value and returns the records when the match exists in the set of values.
The following syntax illustrates this operator: SELECT * FROM table_name WHERE column_name BETWEEN 'value1' AND 'value2';	The following syntax illustrates this operator: SELECT * FROM table_name WHERE column_name IN ('value1','value 2');

#### Q29 What is the difference between DELETE and TRUNCATE statements in SQL?

The main difference between them is that the delete statement deletes data without resetting a table's identity, whereas the truncate command resets a particular table's identity. The following comparison chart explains it more clearly:

DELETE	TRUNCATE
The delete statement removes single or multiple rows from an existing table depending on the specified condition.	The truncate command deletes the whole contents of an existing table without the table itself. It preserves the table structure or schema.
DELETE is a <b>DML command</b> .	TRUNCATE is a <b>DML command</b> .

We <b>can</b> use the <b>WHERE</b> clause in the DELETE command.	We <b>cannot</b> use the <b>WHERE</b> clause with TRUNCATE.
DELETE statement is used <b>to delete a row</b> from a table.	TRUNCATE statement is used <b>to remove all the rows</b> from a table.
DELETE is <b>slower</b> because it maintained the log.	TRUNCATE statement is <b>faster</b> than DELETE statement as it deletes entire data at a time without maintaining transaction logs.
You <b>can</b> roll back data after using the DELETE statement.	It is <b>not possible to roll back</b> after using the TRUNCATE statement.
DELETE query <b>takes more space</b> .	TRUNCATE query <b>occupies less space</b> .

### Q30 What is the difference between TRUNCATE and DROP statements?

#### DROP

The DROP command is used to remove table definition and its contents.

In the DROP command, table space is freed from memory.

DROP is a DDL(Data Definition Language) command.

In the DROP command, a view of the table does not exist.

In the DROP command, integrity constraints will be removed.

In the DROP command, undo space is not used.

The DROP command is quick to perform but gives rise to complications.

#### TRUNCATE

Whereas the TRUNCATE command is used to delete all the rows from the table.

While the TRUNCATE command does not free the table space from memory.

Whereas the TRUNCATE is also a DDL(Data Definition Language) command.

While in this command, a view of the table exists.

While in this command, integrity constraints will not be removed.

While in this command, undo space is used but less than DELETE.

While this command is faster than DROP.

### Q31 Why do we use Commit and Rollback command?

#### COMMIT

COMMIT permanently saves the changes made by the current transaction.

The transaction can not undo changes after COMMIT execution.

When the transaction is successful, COMMIT is applied.

#### ROLLBACK

ROLLBACK undo the changes made by the current transaction.

Transaction reaches its previous state after ROLLBACK.

When the transaction is aborted, ROLLBACK occurs.

### Q32 What is the difference between CHAR and VARCHAR?

CHAR is a fixed-length character data type, while VARCHAR is a variable-length character data type.

### Q33 What is the difference between DROP and TRUNCATE commands?

DROP command removes a table and it cannot be rolled back from the database whereas TRUNCATE command removes all the rows from the table.

### Q34 What is the default ordering of data using the ORDER BY clause? How could it be changed?

The ORDER BY clause in MySQL can be used without the ASC or DESC modifiers. The sort order is preset to ASC or ascending order when this attribute is absent from the ORDER BY clause.

### Q35 How do we use the DISTINCT statement? What is its use?

The SQL DISTINCT keyword is combined with the SELECT query to remove all duplicate records and return only unique records. There may be times when a table has several duplicate records.

The DISTINCT clause in SQL is used to eliminate duplicates from a SELECT statement's result set.

### Q36 What are functions and their usage in SQL?

SQL functions are simple code snippets that are frequently used and re-used in database systems for data processing and manipulation. Functions are the measured values. It always performs a specific task. The following rules should be remembered while creating functions:

- A function should have a name, and the name cannot begin with a special character such as @, \$, #, or other similar characters.
- Functions can only work with the SELECT statements.
- Every time a function is called, it compiles.
- Functions must return value or result.
- Functions are always used with input parameters.

SQL categories the functions into two types:

- **User-Defined Function:** Functions created by a user based on their needs are termed user-defined functions.
- **System Defined Function:** Functions whose definition is defined by the system are termed system-defined functions. They are built-in database functions.

SQL functions are used for the following purposes:

- To perform calculations on data
- To modify individual data items
- To manipulate the output
- To format dates and numbers
- To convert data types

### **Q37 What is meant by case manipulation functions? Explains its different types in SQL.**

Case manipulation functions are part of the character functions. It converts the data from the state in which it is already stored in the table to upper, lower, or mixed case. The conversion performed by this function can be used to format the output. We can use it in almost every part of the SQL statement. Case manipulation functions are mostly used when you need to search for data, and you don't have any idea that the data you are looking for is in lower case or upper case.

There are three case manipulation functions in SQL:

**LOWER:** This function is used to convert a given character into lowercase. The following example will return the 'STEPHEN' as 'stephen':

```
SELECT LOWER ('STEPHEN') AS Case_Result FROM dual;
```

**UPPER:** This function is used to convert a given character into uppercase. The following example will return the 'stephen' as 'STEPHEN':

```
SELECT UPPER ('stephen') AS Case_Result FROM dual;
```

**INITCAP:** This function is used to convert given character values to uppercase for the initials of each word. It means every first letter of the word is converted into uppercase, and the rest is in lower case. The following example will return the 'hello stephen' as 'Hello Stephen':

```
SELECT INITCAP ('hello stephen') AS Case_Result FROM dual;
```

### **Q38 Explain character-manipulation functions? Explains its different types in SQL.**

Character-manipulation functions are used to change, extract, and alter the character string. When one or more characters and words are passed into the function, the function will perform its operation on those input strings and return the result.

The following are the character manipulation functions in SQL:

**A) CONCAT:** This function is used to join two or more values together. It always appends the second string into the end of the first string. For example:

**Input:**

```
SELECT CONCAT ('Information-', 'technology') FROM DUAL;
```

**Output:** Information-technology

**B) SUBSTR:** It is used to return the portion of the string from a specified start point to an endpoint. For example:

**Input:**

```
SELECT SUBSTR ('Database Management System', 9, 11) FROM DUAL;
```

**Output:** Management

**C) LENGTH:** This function returns the string's length in numerical value, including the blank spaces. For example:

**Input:**

```
SELECT LENGTH ('Hello Javatpoint') FROM DUAL;
```

**Output:** 16

**D) INSTR:** This function finds the exact numeric position of a specified character or word in a given string. For example:

**Input:**

```
SELECT INSTR ('Hello Javatpoint', 'Javatpoint');
```

**Output:** 7

**E) LPAD:** It returns the padding of the left-side character value for right-justified value. For example:

**Input:**

```
SELECT LPAD ('200', 6, '*');
```

**Output:** \*\*\*200

**F) RPAD:** It returns the padding of the right-side character value for left-justified value. For example:

**Input:**

```
SELECT RPAD ('200', 6, '*');
```

**Output:** 200\*\*\*

**G) TRIM:** This function is used to remove all the defined characters from the beginning, end, or both. It also trimmed extra spaces. For example:

**Input:**

```
SELECT TRIM ('A' FROM 'ABCD CBA');
```

**Output:** BCDCB

**H) REPLACE:** This function is used to replace all occurrences of a word or portion of the string (substring) with the other specified string value. For example:

**Input:**

```
SELECT REPLACE ('It is the best coffee at the famous coffee shop.', 'coffee', 'tea');
```

**Output:** It is the best tea at the famous tea shop.

### Q39 What is the difference between the WHERE and HAVING clauses?

The main difference is that the WHERE clause is used to filter records before any groupings are established, whereas the HAVING clause is used to filter values from a group. The below comparison chart explains the most common differences:

WHERE	HAVING
This clause is implemented in row operations.	This clause is implemented in column operations.
It does not allow to work with aggregate functions.	It can work with aggregate functions.
This clause can be used with the SELECT, UPDATE, and DELETE statements.	This clause can only be used with the SELECT statement.

**Q39 What is the difference between the RANK() and DENSE\_RANK() functions?**

The **RANK** function determines the rank for each row within your ordered partition in the result set. If the two rows are assigned the same rank, then the next number in the ranking will be its previous rank plus a number of duplicate numbers. For example, if we have three records at rank 4, the next rank listed would be ranked 7.

The **DENSE\_RANK** function assigns a unique rank for each row within a partition as per the specified column value without any gaps. It always specifies ranking in consecutive order. If the two rows are assigned the same rank, this function will assign it with the same rank, and the next rank being the next sequential number. For example, if we have 3 records at rank 4, the next rank listed would be ranked 5.

**Q40 What are SQL comments?**

Comments are explanations or annotations in SQL queries that are readable by programmers. It's used to make SQL statements easier to understand for humans. During the parsing of SQL code, it will be ignored. Comments can be written on a single line or across several lines.

- **Single Line Comments:** It starts with two consecutive hyphens (--).
- **Multi-line Comments:** It starts with /\* and ends with \*/.

**Q41 List the different types of relationships in SQL.**

**There are different types of relations in the database:**

**One-to-One** – This is a connection between two tables in which each record in one table corresponds to the maximum of one record in the other.

**One-to-Many and Many-to-One** – This is the most frequent connection, in which a record in one table is linked to several records in another.

**Many-to-Many** – This is used when defining a relationship that requires several instances on each sides.

**Self-Referencing Relationships** – When a table has to declare a connection with itself, this is the method to employ.

**Q42 What is Auto Increment in SQL?**

Autoincrement keyword allows the user to create a unique number to get generated whenever a new record is inserted into the table.

This keyword is usually required whenever PRIMARY KEY in SQL is used.

AUTO INCREMENT keyword can be used in Oracle and IDENTITY keyword can be used in SQL SERVER.

**Q43 Write the SQL query to get the third maximum salary of an employee from a table named employees.**

Employee table

employee_name	salary
A	24000
C	34000
D	55000
E	75000
F	21000
G	40000
H	50000

```
SELECT * FROM(
SELECT employee_name, salary, DENSE_RANK()
OVER(ORDER BY salary DESC)r FROM Employee)
WHERE r=&n;
```

To find 3rd highest salary set n = 3



#### Q44 How to find the nth highest salary in SQL?

The most typical interview question is to find the Nth highest pay in a table. This work can be accomplished using the dense rank() function.

Employee table

employee_name	salary
A	24000
C	34000
D	55000
E	75000
F	21000
G	40000
H	50000

```
SELECT * FROM(  
SELECT employee_name, salary, DENSE_RANK()  
OVER(ORDER BY salary DESC)r FROM Employee)  
WHERE r=&n;
```

#### Q45 Write a SQL query to find the names of employees that begin with 'A'?

To display name of the employees that begin with 'A', type in the below command:

```
SELECT * FROM Table_name WHERE EmpName like 'A%'
```

#### Q46 What is the main difference between 'BETWEEN' and 'IN' condition operators?

BETWEEN operator is used to display rows based on a range of values in a row whereas the IN condition operator is used to check for values contained in a specific set of values.

##### Example of BETWEEN:

```
SELECT * FROM Students where ROLL_NO BETWEEN 10 AND 50;
```

##### Example of IN:

```
SELECT * FROM students where ROLL_NO IN (8,15,25);
```

#### Q47 What is an ALIAS command?

ALIAS command in SQL is the name that can be given to any table or a column. This alias name can be referred in WHERE clause to identify a particular table or a column.

For example-

```
Select emp.empID, dept.Result from employee emp, department as dept  
where emp.empID=dept.empID;
```

#### Q48 What is a Stored Procedure?

A stored procedure is a prepared SQL code that you can save, so the code can be reused over and over again.

So if you have an SQL query that you write over and over again, save it as a stored procedure, and then just call it to execute it.

You can also pass parameters to a stored procedure, so that the stored procedure can act based on the parameter value(s) that is passed.

#### Q49 How to create a stored procedure using SQL Server?

A stored procedure is a piece of prepared SQL code that you can save and reuse again and over.

So, if you have a SQL query that you create frequently, save it as a stored procedure and then call it to run it.

You may also supply parameters to a stored procedure so that it can act based on the value(s) of the parameter(s) given.

#### Stored Procedure Syntax

```
CREATE PROCEDURE procedure_name  
AS  
sql_statement  
GO;  
Execute a Stored Procedure  
EXEC procedure_name;
```

to the 2nd highest salary set n = 2

To find 3rd highest salary set n = 3 and so on.

#### **Q50 List some advantages and disadvantages of Stored Procedure?**

##### **Advantages:**

A Stored Procedure can be used as a modular programming which means create once, store and call for several times whenever it is required. This supports faster execution. It also reduces network traffic and provides better security to the data.

##### **Disadvantage:**

The only disadvantage of Stored Procedure is that it can be executed only in the database and utilizes more memory in the database server.

#### **Q51 What is Triggers?**

A trigger is a set of actions that are run automatically when a specified change operation (SQL INSERT, UPDATE, or DELETE statement) is performed on a specified table. Triggers are useful for tasks such as enforcing business rules, validating input data, and keeping an audit trail.

#### **Q52 How many TRIGGERS are allowed in the MySQL table?**

6 triggers are allowed in the MySQL table:

- BEFORE INSERT
- AFTER INSERT
- BEFORE UPDATE
- AFTER UPDATE
- BEFORE DELETE
- AFTER DELETE

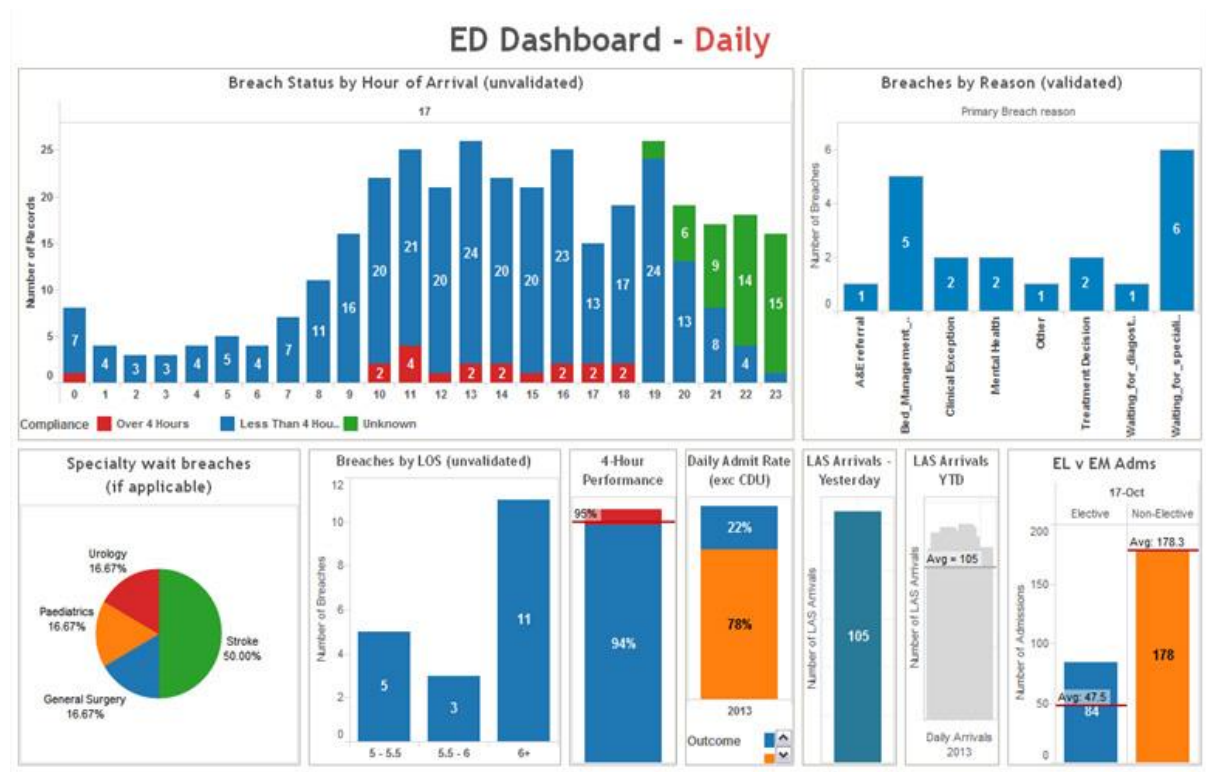
# TABLEAU INTERVIEW QUESTIONS

## Q1 What is data visualization?

Data visualization means the graphical representation of data or information. We can use visual objects like graphs, charts, bars, and a lot more. Data visualization tools provide an accessible way to see and understand the data easily.

## Q2 What is Data Modelling?

Data modeling is the analysis of data objects that are used in a business or other context and also used as identification of the relationships among these data objects. It is the first step of doing object-oriented programming.



## Q3 What is the difference between Traditional BI Tools and Tableau?

Traditional BI Tools	Tableau
1. Architecture has hardware limitations.	1. Do not have dependencies.
2. Based on a complex set of technologies.	2. Based on Associative Search which makes it dynamic and fast
3. Do not support in-memory, multi-thread, multi-core computing.	3. Supports in memory when used with advanced technologies.
4. Has a predefined view of data.	4. Uses predictive analysis for various business operations.

#### Q4 What is Tableau?

- Tableau is a business intelligence software.
- It allows anyone to connect to the respective data.
- Visualizes and creates interactive, shareable dashboards.



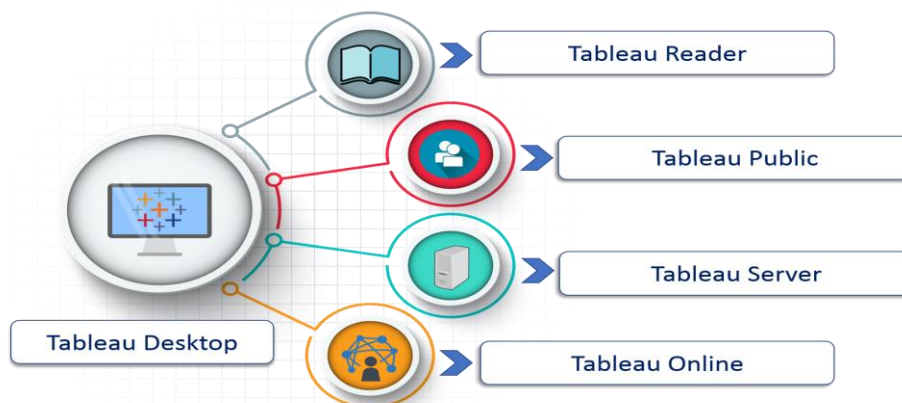
#### Q5 List out Tableau File Extensions.

The below ones are few extensions in Tableau:

- Tableau Workbook (.twb)
- Tableau Data extract (.tde)
- Tableau Datasource (.tds)
- Tableau Packaged Datasource (.tdsx)
- Tableau Bookmark (.tbm)
- Tableau Map Source (.tms)
- Tableau Packaged Workbook (.twbx) – zip file containing .twb and external files.
- Tableau Preferences (.tps)

#### Q6 What are the different Tableau Products and what is the latest version of Tableau?

Here is the Tableau Product family.



##### (i) Tableau Desktop:

It is a self service business analytics and data visualization that anyone can use. It translates pictures of data into optimized queries. With tableau desktop, you can directly connect to data from your data warehouse for live upto date data analysis. You can also perform queries without writing a single line of code. Import all your data into Tableau's data engine from multiple sources & integrate altogether by combining multiple views in a interactive dashboard.

## (ii) Tableau Server:



### Tableau Certification Training Course

Explore Curriculum

It is more of an enterprise level Tableau software. You can publish dashboards with Tableau Desktop and share them throughout the organization with web-based Tableau server. It leverages fast databases through live connections.

## (iii) Tableau Online:

This is a hosted version of Tableau server which helps makes business intelligence faster and easier than before. You can publish Tableau dashboards with Tableau Desktop and share them with colleagues.

## (iv) Tableau Reader:

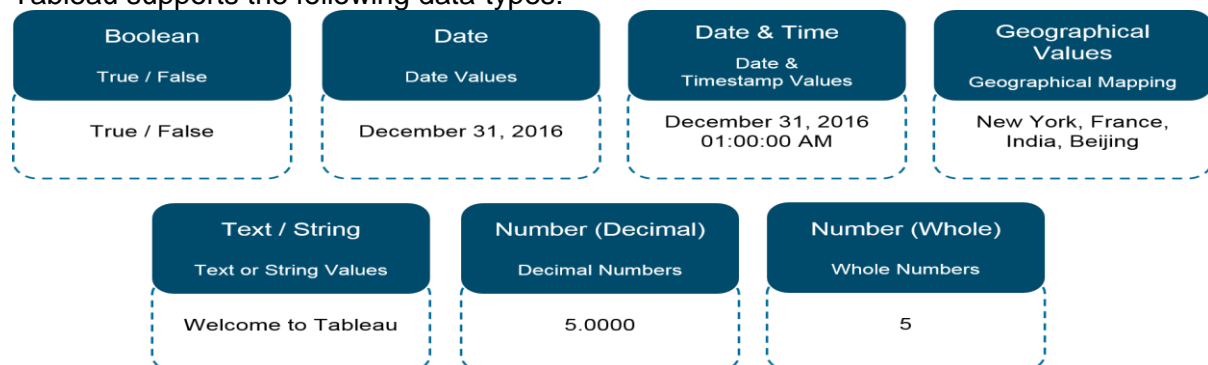
It's a free desktop application that enables you to open and view visualizations that are built in Tableau Desktop. You can filter, drill down data but you cannot edit or perform any kind of interactions.

## (v) Tableau Public:

This is a free Tableau software which you can use to make visualizations with but you need to save your workbook or worksheets in the Tableau Server which can be viewed by anyone.

### Q7 What are the different datatypes in Tableau?

Tableau supports the following data-types:



### Q8 What are Measures and Dimensions?

**Measures** are the numeric metrics or measurable quantities of the data, which can be analyzed by dimension table. Measures are stored in a table that contain foreign keys referring uniquely to the associated dimension tables. The table supports data storage at atomic level and thus, allows more number of records to be inserted at one time. For instance, a Sales table can have product key, customer key, promotion key, items sold, referring to a specific event.

**Dimensions** are the descriptive attribute values for multiple dimensions of each attribute, defining multiple characteristics. A dimension table, having reference of a product key from the table, can consist of product name, product type, size, color, description, etc.

#### Q9 What is the difference between .twb and .twbx extension?

- A .twb is an xml document which contains all the selections and layout made you have made in your Tableau workbook. It does not contain any data.
- A .twbx is a 'zipped' archive containing a .twb and any external files such as extracts and background images.

#### Q10 Define LOD Expression?

LOD Expression stands for Level of Detail Expression, and it is used to run complex queries involving many dimensions at the data sourcing level.

#### Q11 What are the different types of joins in Tableau?

The joins in Tableau are same as SQL joins. Take a look at the diagram below to understand it.



#### Q12 How many maximum tables can you join in Tableau?

You can join a maximum of 32 tables in Tableau.

#### Q13 What are the different connections you can make with your dataset?

We can either connect live to our data set or extract data onto Tableau.

- **Live:** Connecting live to a data set leverages its computational processing and storage. New queries will go to the database and will be reflected as new or updated within the data.
- **Extract:** An extract will make a static snapshot of the data to be used by Tableau's data engine. The snapshot of the data can be refreshed on a recurring schedule as a whole or incrementally append data. One way to set up these schedules is via the Tableau server.

The benefit of Tableau extract over live connection is that extract can be used anywhere without any connection and you can build your own visualization without connecting to database.

#### Q14 What are sets?

**Sets** are custom fields that define a subset of data based on some conditions. A **set** can be based on a computed condition, for example, a **set** may contain customers with sales over a certain threshold. Computed **sets** update as your data changes. Alternatively, a **set** can be based on specific data point in your view.

#### Q15 What are groups?

A group is a combination of dimension members that make higher level categories. For example, if you are working with a view that shows average test scores by major, you may want to group certain majors together to create major categories.

#### Q16 What are the supported data types in Tableau?

The following data types are supported in Tableau:

Data Type	Possible Values
Boolean	True/False
Date	Date Value (December 28, 2016)
Date & Time	Date & Timestamp values (December 28, 2016 06:00:00 PM)
Geographical Values	Geographical Mapping (Beijing, Mumbai)
Text/String	Text/String
Number	Decimal (8.00)
Number	Whole Number (5)

### Q17 What are shelves?

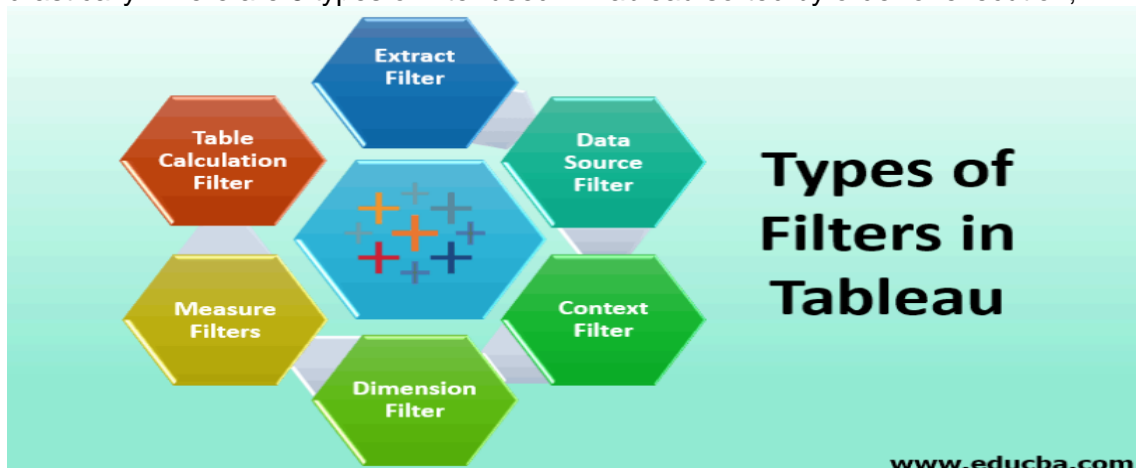
Tableau worksheets contain various named elements like columns, rows, marks, filters, pages, etc. which are called shelves. You can place fields on shelves to create visualizations, increase the level of detail, or add context to it.

### Q18 What is a hierarchical field?

A hierarchical field in tableau is used for drilling down data. It means viewing your data in a more granular level.

### Q19 What are the different filters in Tableau and how are they different from each other?

All the organizations use filters to reduce the size of the dataset and removing irrelevant information to improve the performance or highlight the required information. In Tableau, there are different ways to filter the dataset to increase data efficacy. Each filter is created for different purposes and the order in which they are executed can change the performance drastically. There are 6 types of filter used in Tableau sorted by order of execution;

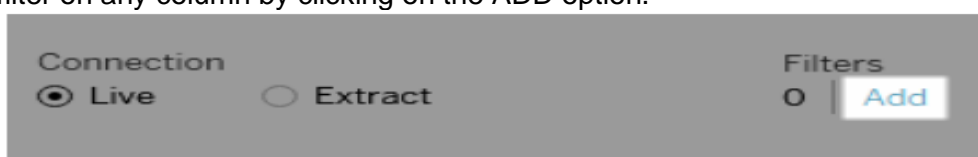


#### 1. Extract Filter

We can use extract filter while loading the dataset into Tableau, so it reduces the number of times Tableau queries for the data source. We can further reduce the size of the data by applying filters to the extract as required.

#### 2. Data Source Filter

This filters any important or sensitive information that we want to control while loading the data into Tableau. It works on both Live and Extract connection. We can add the data source filter on any column by clicking on the ADD option.

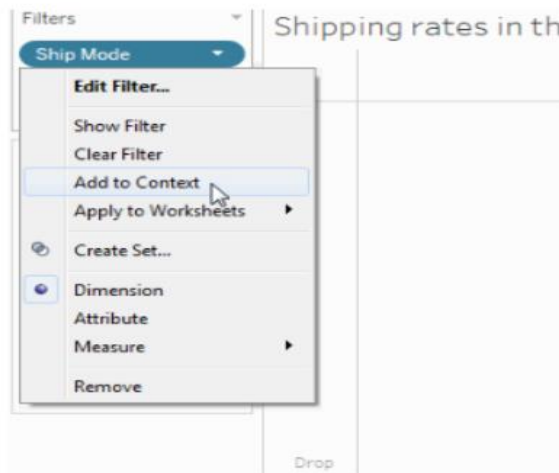




After clicking on the ADD option, the ADD Filter dialog box will appear containing all the fields, then we can select the field that we want to apply the filter on. We can also edit or remove the data source filters as required.

### 3. Context Filters

The filters used in Tableau are normally the independent filters that produce their own result but there are certain filters that are executed to process the records as returned by the first filter. Context filter is an independent filter that creates a different worksheet out of the original dataset and computes the calculation in the filtered dataset. It can be used to improve the performance of large data sources. They can be created by dragging the dimension to the filter section box and clicking on Add to context option. By clicking on this the dimension will change to a grey color which is an indication of the context filter.

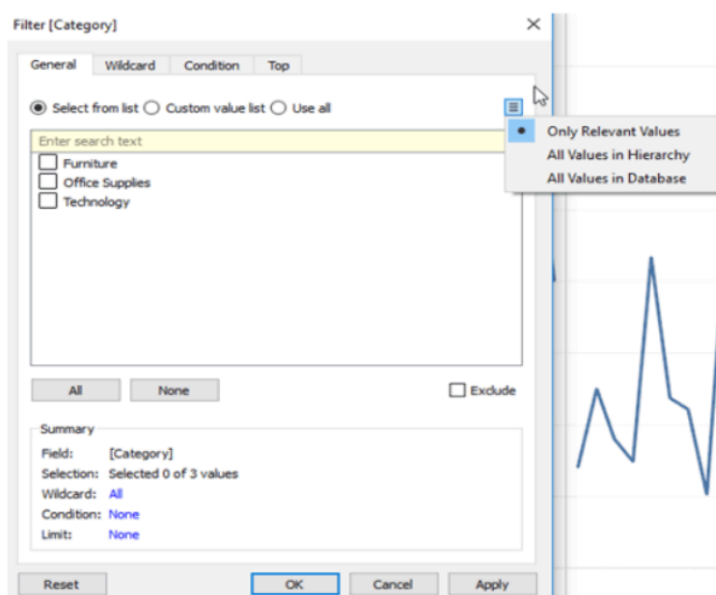


It can also be used to view Top N products in any particular category. We can also remove the Context filter.

### 4. Dimension Filter

Fields in Dimension contain discrete categorical data and we can exclude or include the values that we want to analyze. The process of adding the dimension filter is simple and is given as follows;

- Drag the dimension from the dimension list to the filter section box.
- It will open a Filter Dialog box where we can select the values that we want to analyze.



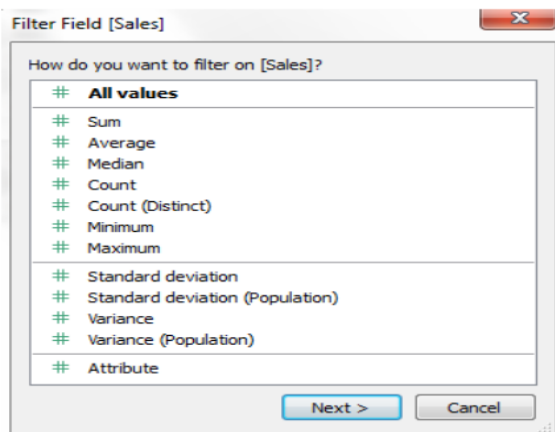
There are four tabs in the Filter dialog box:

1. **General:** To select the members present in the dimension that we want to include or exclude.
2. **Wildcard:** To filter the result on the basis of a particular pattern. For e.g. if we want to filter the email address of a particular domain then we can use the filter that ends with “@yahoo.com” to include those email addresses.
3. **Condition:** To filter the result on the basis of a particular condition.
4. **Top:** To filter the Top N products of a particular category.

## 5. Measure Filters

Measure fields contain quantitative data and these filters are applied to the measure fields. It can be applied by following the below procedure:

Drag the measured field from the Measure box to Filter section and a Filter dialog box will open containing various operations.



Select the operation that needs to be performed and click Next. In subsequent dialog box there are 4 types of filter:

1. **Range:** To select the range of values to include in the result.
2. **At least:** To select the minimum value of a measure to filter the data.
3. **At most:** To select the maximum value of a measure to filter the data.
4. **Special:** To select null or non-null values.

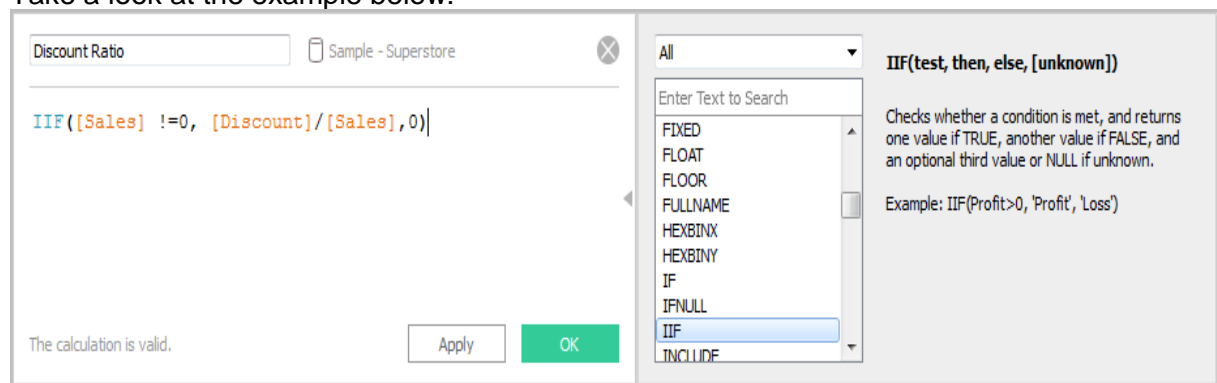
## 6. Table Calculation Filter

These filters are used when we do not want to filter the view without changing the underlying data. Table Calculations are functions used when creating Calculated Fields such as LOOKUP, WINDOW\_SUM, WINDOW\_AVG, etc.

### Q20 How to create a calculated field in Tableau?

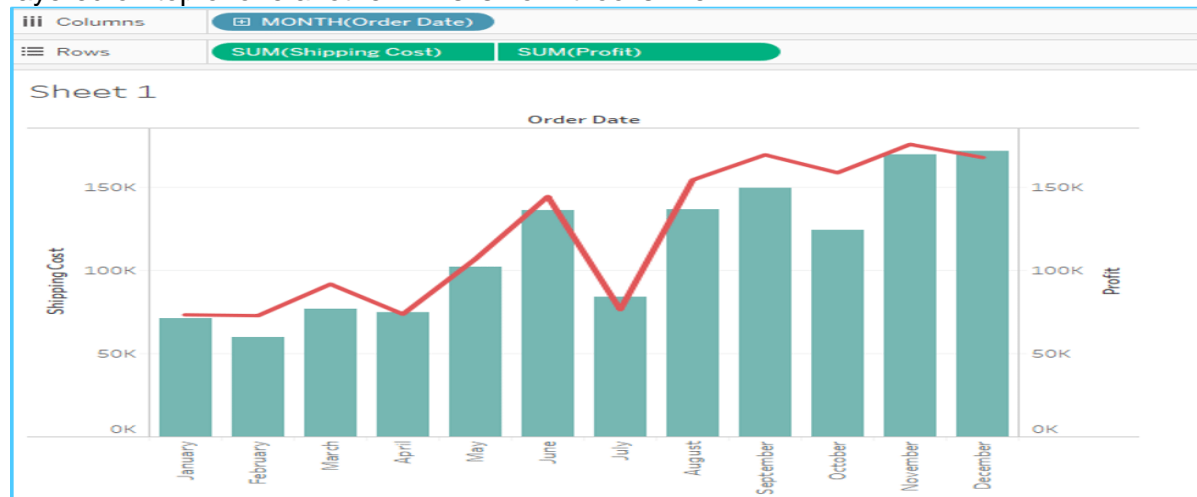
- Click the drop down to the right of Dimensions on the Data pane and select “Create > Calculated Field” to open the calculation editor.
- Name the new field and create a formula.

Take a look at the example below:



### Q21 What is a dual axis?

Dual Axis is an excellent phenomenon supported by Tableau that helps users view two scales of two measures in the same graph. Many websites like Indeed.com and other make use of dual axis to show the comparison between two measures and their growth rate in a septet set of years. Dual axes let you compare multiple measures at once, having two independent axes layered on top of one another. This is how it looks like:



### Q22 What is disaggregation and aggregation of data?

The process of viewing numeric values or measures at higher and more summarized levels of the data is called aggregation. When you place a measure on a shelf, Tableau automatically aggregates the data, usually by summing it. You can easily determine the aggregation applied to a field because the function always appears in front of the field's name when it is placed on a shelf. For example, Sales becomes SUM(Sales). You can aggregate measures using Tableau only for relational data sources. Multidimensional data sources contain aggregated data only. In Tableau, multidimensional data sources are supported only in Windows.

According to Tableau, Disaggregating your data allows you to view every row of the data source which can be useful when you are analyzing measures that you may want to use both independently and dependently in the view. For example, you may be analyzing the results from a product satisfaction survey with the Age of participants along one axis. You can aggregate the Age field to determine the average age of participants or disaggregate the data to determine what age participants were most satisfied with the product.

### Q23 What is the difference between joining and blending in Tableau?

- Joining term is used when you are combining data from the same source, for example, worksheet in an Excel file or tables in Oracle database
- While blending requires two completely defined data sources in your report.

### Q24 What is a blended axis?

Blended axis is the axis where multiple measures are shown in a single axis and all the marks are shown in a single pane. We can blend measures by dragging the 1st measure on one axis and the 2nd on the existing axis.

- Drag a dimension in a column
- Drag the first measure in the column
- Drag the second measure in the existing axis
- [Us/multiplemeasures\\_blendedaxes.html](#)

### Q25 Give a brief about the tableau dashboard?

Tableau dashboard is a group of various views which allows you to compare different types of data simultaneously. Datasheets and dashboards are connected if any modification happens to the data that directly reflects in dashboards. It is the most efficient approach to visualize the data and analyze it.

### Q26 Define the story in Tableau?

The story can be defined as a sheet which is a collection of series of worksheets and dashboards used to convey the insights of data. A story can be used to show the connection between facts and outcomes that impacts the decision-making process. A story can be published on the web or can be presented to the audience.

### Q27 Define Bullet graph?

A bullet graph is a variant of Bar graph. It is responsible for comparing the performance of one measure with other measures.

### Q28 Define Gantt chart?

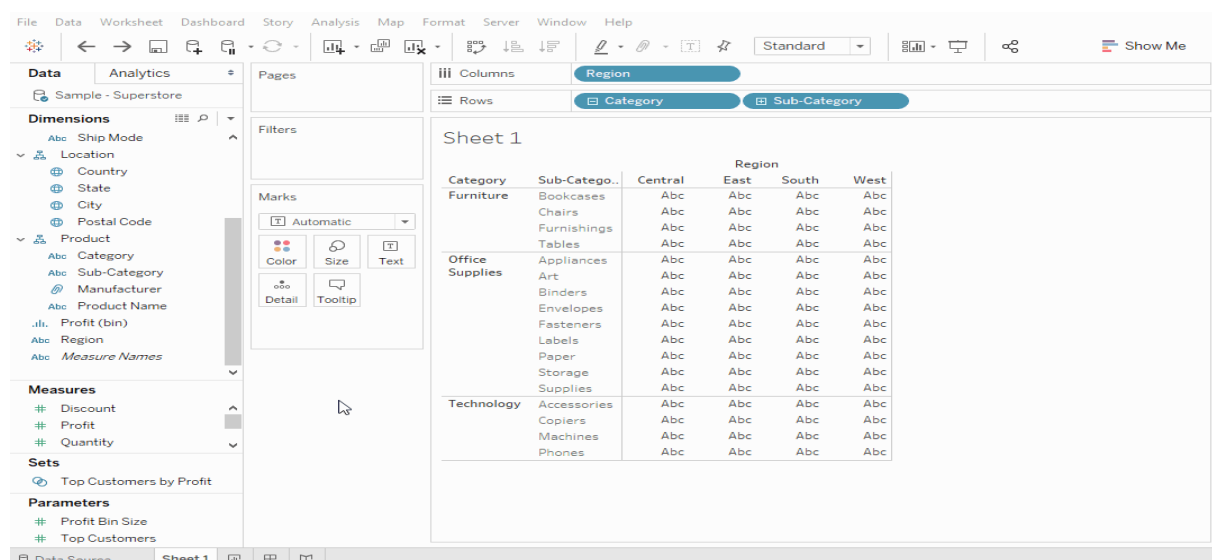
Gantt Chart displays the progress of value over the period. It consists of bars along with the time axis. It is a project management tool. Here, each bar is a measure of a task in the project framework.

### Q29 Define a Histogram chart?

A histogram chart shows the distribution of continuous information over a certain period of time. This chart helps us to find extreme points, gaps, unusual values, and more concentrated values.

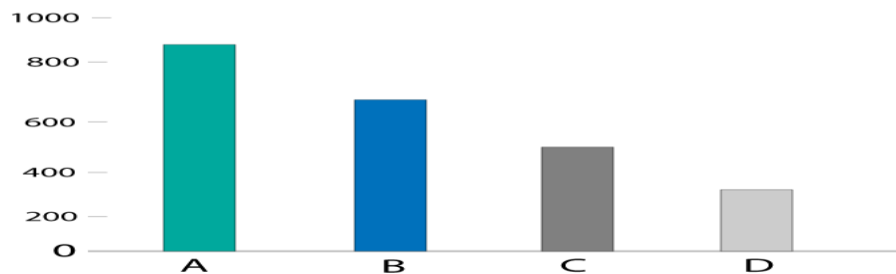
### Q30 What is the Hierarchy in Tableau?

When we are working with large volumes of data, incredibly data may be messed. With Tableau, you can easily create hierarchies to keep your data neat. Even if you don't need it, it is built into your data, which you can easily manage or organize the data and you can track the data easily.



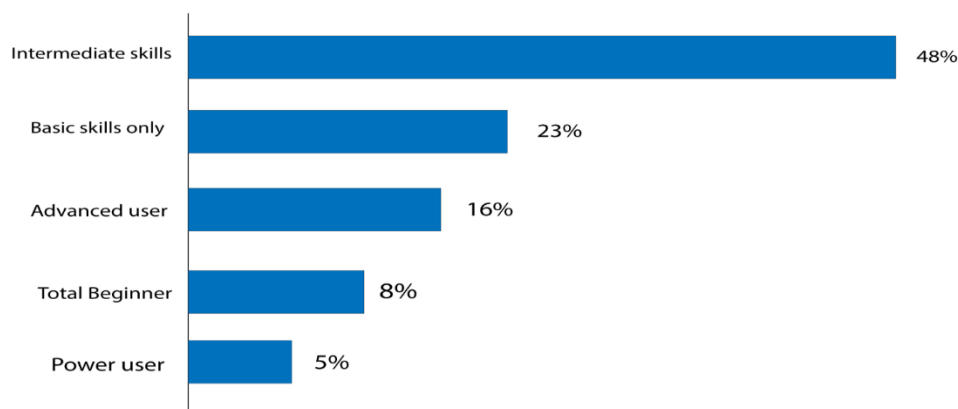
### Q31 What is a Column chart?

A column chart visualizes the data as a set of rectangle columns, as their lengths are proportional to values when they represent the data. The horizontal axis shows the category to which they belong, and the vertical axis shows the values.



### Q32 What is the Bar Chart in Tableau?

The bar chart visualizes the data as a set of rectangle bars, as their values are proportional to lengths when they represent the data. The vertical axis shows the category to which they belong to and the horizontal axis shows the values. So, the bar chart is a vertical version of the Column chart.



### Q33 What is the Line Chart?

The line chart is a popular type of diagrammatic way for visualizing the data, it connects the individual data points to view the data. We can easily visualize the series of values, we can see trends over time or predict future values. The horizontal axis holds the category to which it belongs and the vertical axis holds the values.

### Q34 What is a Stacked Bar chart?

Stacked Bar Chart, composed of multiple bars stacked horizontally, one below the other. The length of the bar depends on the value in the data point. A stacked bar chart makes the work easier, they will help us to know the changes in all variables presented, side by side. We can watch the changes in their total and forecast future values.

### Q35 What is a Stacked Column Chart?

Stacked Column Chart, composed of multiple bars stacked vertically, one on another. The length of the bar depends on the value in the data point. A stacked column chart is the best one to know the changes in all variables. This type of chart should be checked when the number of series is higher than two.

### Q36 What is an Area Chart?

An area chart is nothing but line chart, the area between the x-axis and lines will be color or patterns. These charts are typically used to represent accumulated totals over time and are the conventional way to display stacked lines

### Q37 What is Context Filter and show the steps on how to create the Context Filter Tableau?

Context Filters are applied to the data rows before any other filters. They are limited to views, but they can be applied on selected sheets. They define Aggregation and Disaggregation of data in Tableau

**Step 1:** Drag the subcategory dimensions to the row shelf and measure sales to the column shelf. Now choose the horizontal bar chart as chart type and again drag the sub-category dimensions to the filter shelf. Then we will get the following chart.

**Step 2:** Right-click on the Sub-Category field in the filter shelf and go to the Top fourth tab. Choose the option field, from the next drop-down and choose the option Top 10 by Sales Sum as shown in the following screenshot.

**Step 3:** Drag the dimension Category to the filter shelf. Give right-click on the general tab to edit and under that choose Furniture from the list. As you can see the result shows three subcategories of products.

**Step 4:** Right-click the Category: Furniture filter and select the option Add to Context. This produces the final result, which shows the subcategory of products from the category Furniture which are among the top 10 subcategories across all the products.

### Q38 Differentiate parameters and filters in Tableau

**Filters** are the simpler and straightforward feature in Tableau. It applies to dimensions or measures directly. For example, to only show Gujarat or Karnataka in a State dimension, we can apply the filter on that. In Tableau, there are multiple UI options available for filters like radio buttons, drop-down lists, checkboxes, sliders, and more. Filters on sheets are also available in Tableau.

Filters, on the other hand, are used to restrict the data based on a condition that is mentioned in the filters shelf.

**Parameters** are like variables. They are complex and more powerful. Like a variable, a parameter can be used in calculations. So, that means, it only allows a single value. Parameters have the same UI options except for checkboxes because checkboxes don't have a single value. For example, we can create a parameter for interest rate and period, and then we can use these parameters to calculate interest and principal payments. Parameters are dynamic values that can replace constant values in calculations. Parameters can serve as filters as well.

### Q39 Does a parameter have its own drop-down list?

Yes, it may have its own drop-down list. The entries one makes in a parameter while creating it can be viewed as items in the drop-down list.

### Q40 Tell me different ways to use parameters in Tableau

- Filters
- Calculated fields
- Actions
- Measure-swaps
- Changing views
- Auto-updates

### Q41 Differentiate between Tiled and Floating in dashboards?

In a tiled layout, items don't overlap. The layout will be adjusted according to dashboard size. In the floating layout, items can be placed on some other layers. Floating items can have fixed positions and sizes.

#### Q42 Differentiate discrete and continuous data roles in Tableau

Discrete data roles consist of values that are separate and distinct. Discrete data roles can take individual values within a range. For Example – cancer patients in the hospital, no. of threads in a sheet, state. Discrete values are displayed as blue icons in the data window and blue pills on shelves. Discrete fields can be sorted.

Continuous data roles consist of any value within the finite or infinite intervals. For Example – age, unit price, order quantity. Continuous values displayed as green icons in the data window and green pills on shelves. Continuous fields cannot be sorted.

Discrete	Continuous
Discrete data is the value that is counted as distinct or separate.	Continuous data is used to measure continuous data.
Only It can take individual values within a range.	It can take any values within a finite and infinite range.

#### Q43 Explain the disaggregation and aggregation of data in Tableau?

**Aggregation** → The process of summarizing the data and viewing a single numeric value is called aggregation. Example – sum/avg of salary for each employee

**Disaggregation** → The process of viewing each transaction for analyzing all the measures both dependently and independently. Example – individual salary transactions for each employee.

#### Q44 State the components of the dashboard?

The dashboard consists of 5 components.

- **Web:** it consists of a web page embedded in the dashboard.
- **Horizontal component:** it is a horizontal layout container in which we can add objects.
- **Vertical component:** it is a vertical layout container in which we can add objects.
- **Image Extract:** it allows you to upload an image to the dashboard from a computer.
- **Text:** it is a small Wordpad where we can format and edit the text.

#### Q45 What is the difference between the Tree map and Heat map?

Tree map	Heat map
A treemap also does the same thing as well as it can be used for illustrating hierarchical data and part of whole relationships.	A heat map can compare categories with color and size In the heat map, you can compare two different measures together.

#### Q46 What is the difference between Data Joining and Data blending?

Data joining	Data blending
Data joining is used when you are combining the data from the same source.	Data blending is required two completely defined data sources in a report.



#### Q47 What are the different Tableau files?

- **Bookmarks:** It contains only single worksheet and its easy way to share your work.
- **Workbooks:** Workbook can hold one or more dashboards and worksheets.
- **Packaged workbooks:** It contains the workbook along with any supporting local file data and background images.
- **Data extraction files:** Data extraction files are a local copy of a data source or a subset.
- **Data connection files:** Data connection file is a small XML file that contain various connection information.

#### Q48 What is the difference between the published data source and an embedded data source?

Published data source	Embedded data source
It contains connection information which is independent of any workbook, and multiple workbooks can use it.	It contains connection information, and it is associated with a workbook.

#### Tips to clear an interview

To clear interview for Tableau, follow these tips:

- Focus on the fundamentals: What is Tableau and its working. How calculations work or how a query is processed when visualization is created.
- Thoroughly know about Dimensions and Measures because that is one of the important concepts in Tableau.
- Get acquainted with the best practices of creating dashboards and visualizations and also discrete and continuous views.
- Explain why you like Tableau or how it differs from other similar tools like QlikView or IBM Cognos. Your interest in BI tools will put you ahead in the competition.
- What are the scenarios where you'll use Live connection or Data extract in Tableau?
- How dashboards are deployed on the Server.
- What was the maximum amount of data you have handled in Tableau? If you are learning Tableau, while practicing, check the size of your visualization or the TDE file.
- Create some visualization stories for sample work.
- How you will take requirements before creating a dashboarding application.
- What was your development methodology: Waterfall or Agile?
- How much time it takes you to create a dashboard.

## POWER BI INTERVIEW QUESTIONS

### Q1 What is Power BI?

Power BI is a business analytics tool developed by Microsoft that helps you turn multiple unrelated data sources into valuable and interactive insights. These data may be in the form of an Excel spreadsheet or cloud-based/on-premises hybrid data warehouses. You can easily connect to all your data sources and share the insights with anyone.

### Q2 Difference between Power BI and Tableau

Both Tableau and Power BI are the current IT industry's data analytics and visualization giants. Yet, there are a few significant differences between them. You will now explore the important differences between Tableau and Power BI.

Tableau	Power BI
Tableau uses MDX for measures and dimensions	Power BI uses DAX for calculating measures
Tableau is capable of handling large volumes of data	Power BI is qualified only to handle a limited amount of data
Tableau is best suitable for experts	Power BI is suitable for both experts and beginners
Tableau User Interface is complicated	Power BI User Interface is comparatively simpler
Tableau is capable of supporting the cloud with ease.	Power BI finds it difficult, as its capacity to handle large volumes of data is limited.

### Q3 Difference between Power Query and Power Pivot

The differences between Power Query and Power Pivot are explained as follows:

Power Query	Power Pivot
<u>Power Query</u> is all about analyzing data.	<u>Power Pivot</u> is all about getting and Transforming data.
Power Query is an ETL service tool.	Power Pivot is an in-memory data modeling component

#### Q4 What is Power BI Desktop

Power BI Desktop is an open-source application designed and developed by Microsoft. Power BI Desktop will allow users to connect to, transform, and visualize your data with ease. Power BI Desktop lets users build visuals and collections of visuals that can be shared as reports with your colleagues or your clients in your organization.

#### Q5 What is Power Pivot?

Power Pivot is an add-on provided by Microsoft for Excel since 2010. Power Pivot was designed to extend the analytical capabilities and services of Microsoft Excel.

#### Q6 What are filters in Power BI?

Filters sort data based on the condition applied to it. Filters enable us to select particular fields and extract information in a page/visualization/report level. For example, filters can provide sales reports from the year 2019 for the Indian region. Power BI can make changes based on the filters and create graphs or visuals accordingly. Types of filters are:

- **Page-level filters:** These are applied on a particular page from various pages available within a report.
- **Visualization-level filters:** These are applied to both data and calculation conditions for particular visualizations.
- **Report-level filters:** These are applied to the entire report.

The following are the variety of filters available in Power BI:

- Manual filters
- Auto filters
- Include/Exclude filters
- Drill-down filters
- Cross Drill filters
- Drillthrough filters
- Drillthrough filters
- URL filters–transient
- Pass-Through filters

#### Q7 What is Visualization?

Visualization is a process to represent data in pictorial form like tables, graphs, or charts based on the specific requirement.

#### Q8 What are Custom Visuals in Power BI?

Custom Visuals are like any other visualizations, generated using Power BI. The only difference is that it develops the custom visuals using a custom SDK. The languages like JQuery and JavaScript are used to create custom visuals in Power BI.

#### Q9 What is a Report?

The report is a Power BI feature that is a result of visualized data from a single data set. A report can have multiple pages of visualization.

#### Q10 What is GetData in Power BI?

Get Data is a simple icon on Power BI used to import data from the source.

#### Q11 Mention some advantages of Power BI.

Some of the advantages of using Power BI:

- It helps build an interactable data visualization in data centers
- It allows users to transform data into visuals and share them with anyone
- It establishes a connection for Excel queries and dashboards for fast analysis
- It provides quick and accurate solutions
- It enables users to perform queries on reports using simple English words

**Q12 List out some drawbacks/limitations of using Power BI.**

Here are some limitations to using Power BI:

- Power BI does not accept file sizes larger than 1 GB and doesn't mix imported data accessed from real-time connections.
- There are very few data sources that allow real-time connections to Power BI reports and dashboards.
- It only shares dashboards and reports with users logged in with the same email address.
- Dashboard doesn't accept or pass user, account, or other entity parameters.

**Q13 What are some differences in data modeling between Power BI Desktop and Power Pivot for Excel?**

Power Pivot for Excel supports only single directional relationships (one to many), calculated columns, and one import mode. Power BI Desktop supports bi-directional cross-filtering connections, security, calculated tables, and multiple import options.

**Q14 Name the different connectivity modes available in Power BI?**

There are three main connectivity modes used in Power BI.

**SQL Server Import**

An SQL Server Import is the default and most common connectivity type used in Power BI. It allows you to use the full capabilities of the Power BI Desktop.

**Direct Query**

The Direct Query connection type is only available when you connect to specific data sources. In this connectivity type, Power BI will only store the metadata of the underlying data and not the actual data.

**Live Connection**

With this connectivity type, it does not store data in the Power BI model. All interaction with a report using a Live Connection will directly query the existing Analysis Services model. There are only 3 data sources that support the live connection method - SQL Server Analysis Services (Tabular models and Multidimensional Cubes), Azure Analysis Services (Tabular Models), and Power BI Datasets hosted in the Power BI Service.

**Q15 What are the key differences between a Power BI dataset, a report, and a dashboard?**

Dataset	Report	Dashboard
A series of Power Query queries that have been shaped in a DAX model.	A series of visualizations, filters, and static elements on a canvas.	A way of pulling visualizations together from several reports.
A Power BI dataset can have many data sources.	Each report can have multiple sheets.	A Power BI dashboard is a single page, often called a canvas, that uses visualizations to tell a story.
A data set can have one report, and a report can have one data set.	The data set and your report are going to have a one-to-one relationship.	A dashboard is a tool for pinning visuals from different reports and other sources of data.

**Q16 Describe the components of Microsoft's self-service BI solution.**

Self-service business intelligence (SSBI) is divided into the Excel BI Toolkit and Power BI.

**Q17 What is self-service BI, anyway?**

SSBI is an abbreviation for Self-Service Business Intelligence and is a breakthrough in business intelligence. SSBI has enabled many business professionals with no technical or coding background to use Power BI and generate reports and draw predictions successfully. Even non-technical users can create these dashboards to help their business make more informed decisions.

**Q18 Mention some advantages of Power BI.**

Some of the advantages of using Power BI:

- It helps build an interactable data visualization in data centers`
- It allows users to transform data into visuals and share them with anyone
- It establishes a connection for Excel queries and dashboards for fast analysis
- It provides quick and accurate solutions
- It enables users to perform queries on reports using simple English words

**Q19 What are some differences in data modelling between Power BI Desktop and Power Pivot for Excel?**

Power Pivot for Excel supports only single directional relationships (one to many), calculated columns, and one import mode. Power BI Desktop supports bi-directional cross-filtering connections, security, calculated tables, and multiple import options.

**Q20 What are the various types of refresh options provided in Power BI?**

Four important types of refresh options provided in Microsoft Power BI are as follows:

- Package refresh - This synchronizes your Power BI Desktop or Excel file between the Power BI service and OneDrive, or SharePoint Online.
- Model or data refresh - This refreshes the dataset within the Power BI service with data from the original data source.
- Tile refresh - This updates the cache for tile visuals every 15 minutes on the dashboard once data changes.
- Visual container refresh - This refreshes the visible container and updates the cached report visuals within a report once the data changes.

**Q21 Name the data sources can Power BI can connect to?**

Several data sources can be connected to Power BI, which is grouped into three main types:

- **Files**  
*It can import data from Excel (.xlsx, .xlsm), Power BI Desktop files (.pbix) and Comma-Separated Values (.csv).*
- **Content Packs**  
These are a collection of related documents or files stored as a group. There are two types of content packs in Power BI:  
Content packs from services providers like Google Analytics, Marketo, or Salesforce and Content packs are created and shared by other users in your organization.
- **Connectors**  
*Connectors help you connect your databases and datasets with apps, services, and data in the cloud.*

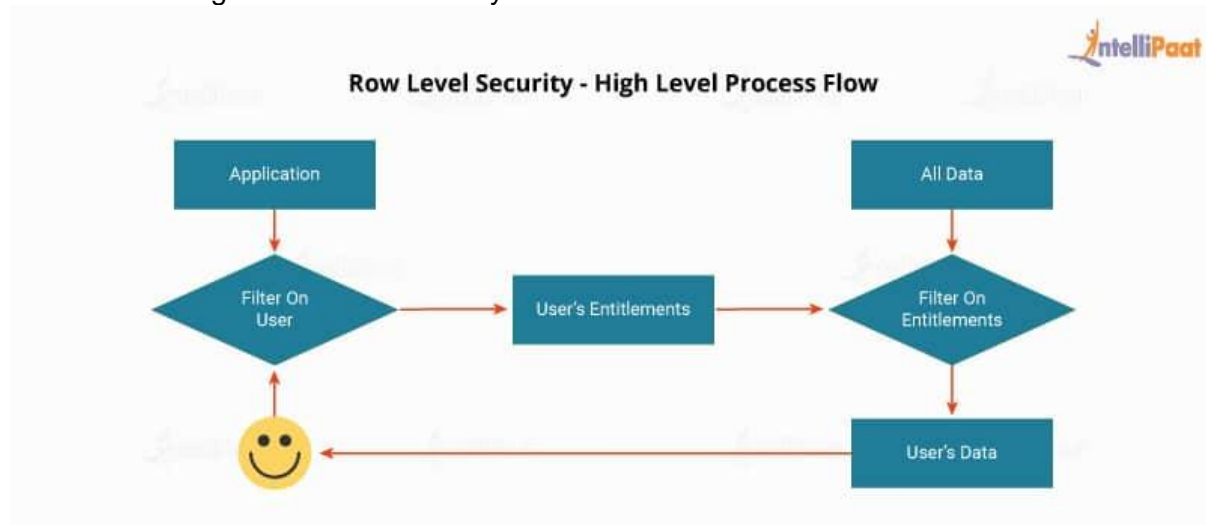
**Q22 Explain how relationships are defined in Power BI Desktop?**

Relationships between tables are defined in two ways:

- Manually - Relationships between tables are manually defined using primary and foreign keys.
- Automatic - When enabled, this automated feature of Power BI detects relationships between tables and creates them automatically

### Q23 What is row-level security?

Row-level security limits the data a user can view and has access to, and it relies on filters. Users can define the rules and roles in Power BI Desktop and also publish them to Power BI Service to configure row-level security.



### Q24 Explain the building blocks of Microsoft Power BI.

The important building blocks of Power BI are as follows:

#### **Visualizations**

Visualization is the process of generating charts and graphs for the representation of insights on business data.

#### **Datasets**

A dataset is the collection of data used to create a visualization, such as a column of sales figures. Dataset can get combined and filtered from a variety of sources via built-in data plugins.

#### **Reports**

The final stage is the report stage. Here, there is a group of visualizations on one or more pages. For example, charts and maps are combined to make a final report.

#### **Dashboards**

A Power BI dashboard helps you to share a single visualization with colleagues and clients to view your final dashboard.

#### **Tiles**

A tile is an individual visualization on your final dashboard or one of your charts in your final report.

### Q25 What are the different views in Power BI Desktop?

There are three different views in Power BI, each of which serves a different purpose:



**Report View:** Users can add visualizations and additional report pages and publish the same on the portal from here.

**Data View:** Data shaping can be performed through Query Editor tools.

**Relationship View:** Users can manage relationships between datasets in this view.

#### Q26 What are the critical components of the Power BI toolkit?

The critical components of Power BI are mentioned below.

- Power Query
- Power Pivot
- Power View
- Power Map
- Power Q&A
  - **Power Query:** It is one of the most important components of PowerBI to transform data. Power Query helps to extract data from different data sources like Oracle, SQL, Text/CSV files, Excel, etc. and even delete data from different sources.
  - **Power Pivot :** It is used for data modeling that uses DAX ( Data Analysis Expression) functions for the calculations. Relationships between different tables can also be created here and we can get values that can be shown in Pivot Tables.
  - **Power View:** The Power View is used for providing an intuitive display of the data and retrieving the metadata for data analysis. The views are interactive in nature and slicers and filters can be used for slicing and dicing the data.
  - **Power BI Desktop:** Power Desktop is an integration tool for Power Query, Power View, and Power Pivot. It helps to create advanced queries, data models, reports and dashboards and helps in developing your BI skills for data analysis.
  - **Power BI Mobile Application:** It is available for the Operating systems Android, iOS and even Windows. The App has an interactive display of the dashboards which can be shared as well.
  - **Power Map:** It presents geo-spatial visualization of the data in 3 Dimensional Mode. The data can be highlighted based on the geographical location which can be continent, state, city or even street address.
  - **Power Q&A :** It is used to provide answers to the questions asked by users. It works with Power View and can be answered with representations by Power Q&A.
- Quick accessibility to data means there is no speed and memory issue

#### Q27 What are the Different Products in the PowerBI family?

Below are different Power BI services/products:

- Power BI Desktop
- Power BI Services
- Power BI Mobile
- Power BI Gateway
- Power BI Premium
- Power BI Report Server
- Power BI Embedded

#### Q28 What do you mean by grouping?

Power BI Desktop helps you to group the data in your visuals into chunks. You can, however, define your groups and bins. For grouping, use Ctrl + click to select multiple elements in the visual. Right-click one of those elements and, from the menu that appears, choose Group. In the Groups window, you can create new groups or modify existing ones.



### Q29 Explain responsive slicers in Power BI.

On a Power BI final report page, a developer can resize a responsive slicer to various sizes and shapes, and the data collected in the container will be rearranged to find a match. If a visual report becomes too small to be useful, an icon representing the visual takes its place, saving space on the report page.

### Q30 What is query folding in Power BI?

Query folding is used when steps defined in the Query Editor are translated into SQL and executed by the source database instead of your device. It helps with scalability and efficient processing.

### Q31 What are the different stages in the working of Power BI?

There are three different stages in working on Power BI, as explained below.

1. Data Integration
2. Data Processing
3. Data Presentation

#### **Data Integration**

The primary step in any business intelligence is to establish a successful connection with the data source and integrate it to extract data for processing.

#### **Data Processing**

The next step in business intelligence is data processing. Most of the time, the raw data also includes unexpected erroneous data, or sometimes a few data cells might be empty. The BI tool needs to interpret the missing values and inaccurate data for processing in the data processing stage.

#### **Data Presentation**

The final stage in business intelligence is analyzing the data got from the source and presenting the insights using visually appealing graphs and interactive dashboards.

### Q32 What is the comprehensive working system of Power BI?

Power BI's working system mainly comprises four steps:

**Data Importing:** The first step is to import the data and convert it into a standard format and store it in a staging area.

**Data Cleaning:** After assembling the data, it requires transformation or cleaning to remove unimportant values.

**Data Visualization:** Now the data is visually represented on the Power BI desktop as reports and dashboards using powerful visualization tools.

**Save and Publish:** Finally when your report is ready you can save and publish these reports that can be shared across users via mobile apps or web.

### Q33 What are the types of visualizations in Power BI?

In PowerBI, we can represent the data in graphs and visualizations. The visualization can be of any type, for example:

**Bar and Column Charts:** It is a standard visualization for looking at a specific value across various categories.

**Area Charts( Basic and Stacked ) :** It is based on the line chart and the area under the line. It depicts the magnitude of change over time.

**Card:** Card shows aggregate value of a certain datapoint, can be one or more but one per row.

**Doughnut and Pie Charts:** They show the relation in parts of a whole. Doughnut charts have a hollow in the centre while pie charts don't.

**Maps:** To show categorical and quantitative data with spatial locations.

**Matrix:** It's a type of table with easier display that shows aggregated data

**Slicers:** Slicer is used to filter other visuals on the page.

There are other visuals like Combo Charts, Decomposition Tree, Funnel charts, Gauge charts, KPIs, Line Charts, Ribbon Chart, Scatter, Q&A, Tables, Treemaps, etc.

#### Q34 What is the difference between a Filter and a Slicer?

Filters are used to restrict users and not allow them to interact with dashboards or reports, while the slicers are used to interact with dashboards and reports.

#### Q35 What is the difference between a new column and a new measure in Power BI?

In Power BI, a new column is an area where the physical data is stored when logic is applied. On the other hand, the measure is where the calculations are performed on the fly based on dimensions. The measure doesn't store any physical data like Column.

#### Q36 What are the different joins in Power BI?

There are mainly two types of joins in Power BI:

- **Horizontal Joins:** Horizontal Joins are used to append data from multiple tables.
- **Vertical Joins:** Vertical Joins are used to merge the data from multiple tables.

#### Q37 What are the various type of users who can use Power BI?

**Ans:** PowerBI can be used by anyone for their requirements but there is a particular group of users who are more likely to use it:

**Report Consumers:** They consume the reports based on a specific information they need

**Report Analyst:** Report Analysts need detailed data for their analysis from the reports

**Self Service Data Analyst:** They are more experienced business data users. They have an in-depth understanding of the data to work with.

**Basic Data Analyst:** They can build their own datasets and are experienced in PowerBI Service

**Advanced Data Analyst:** They know how to write SQL Queries and have hands-on experience on PowerBI. They have experience in Advanced PowerBI with DAX training and data modelling.

measure Name

B- = – indicate beginning of formula

C- DAX Function

D- Parenthesis for Sum Function

E- Referenced Table

F- Referenced column name

#### Q38 What is DAX?

DAX stands for Data Analysis Expressions. It's a collection of functions, operators, and constants used in formulas to calculate and return values. In other words, it helps you create new info from data you already have.

#### Q39 What is the CALCULATE function in DAX?

The CALCULATE function evaluates the sum of the Sales table Sales Amount column in a modified filter context. It is also the only function that allows users to modify the filter context of measures or tables.

Moving ahead, you will step up to the following Power BI Interview Questions from the Intermediate Level.

#### Q40 What are the three fundamental concepts of DAX?

##### **Syntax**

This is how the formula is written—that is, the elements that comprise it. The Syntax includes functions such as SUM (used when you want to add figures). If the Syntax isn't correct, you'll get an error message.

##### **Functions**

These are formulas that use specific values (also known as arguments) in a particular order to perform a calculation, similar to the functions in Excel. The categories of functions are

date/time, time intelligence, information, logical, mathematical, statistical, text, parent/child, and others.

### **Context**

There are two types: row context and filter context. Row context comes into play whenever a formula has a function that applies filters to identify a single row in a table. When one or more filters are applied in a calculation that determines a result or value, the filter context comes into play.

Name some commonly used tasks in the Query Editor.

- Connect to data
- Shape and combine data
- Group rows
- Pivot columns
- Create custom columns
- Query formulas

### **Q41 What are the most common DAX Functions used?**

Below are some of the most commonly used DAX function:

- SUM, MIN, MAX, AVG, COUNTROWS, DISTINCTCOUNT
- IF, AND, OR, SWITCH
- ISBLANK, ISFILTERED, ISCROSSFILTERED
- VALUES, ALL, FILTER, CALCULATE,
- UNION, INTERSECT, EXCEPT, NATURALINNERJOIN, NATURALLEFTOUTERJOIN, SUMMARIZECOLUMNS, ISEMPTY,
- VAR (Variables)
- GEOMEAN, MEDIAN, DATEDIFF

### **Q42 What are the purpose and benefits of using the DAX function?**

DAX or Data Analysis Expression is a functional language which can create calculated columns and/or measures for smarter calculations to limit the data the dashboard has to fetch and visualize.

### **Q43 How is the FILTER function used?**

The FILTER function returns a table with a filter condition applied for each of its source table rows. The FILTER function is rarely used in isolation, it's generally used as a parameter to other functions such as CALCULATE.

- FILTER is an iterator and thus can negatively impact performance over large source tables.
- Complex filtering logic can be applied such as referencing a measure in a filter expression.
  - `FILTER(MyTable,[SalesMetric] > 500)`

### **Q44 What is the CALCULATE function in DAX?**

The CALCULATE function measures the sum of a column from any table and can be modified with Filters.

Syntax:

`CALCULATE ( <Expression> [, <Filter> [, <Filter> [, ... ] ] ] )`

Expression: The expression to be evaluated.

Filter: A boolean (True/False) expression or a table expression that defines a filter.

#### **Q45 What is special or unique about the CALCULATE and CALCULATETABLE functions?**

These are the only functions that allow you modify filter context of measures or tables.

- Add to existing filter context of queries.
- Override filter context from queries.
- Remove existing filter context from queries.

#### **Limitations:**

- Filter parameters can only operate on a single column at a time.
- Filter parameters cannot reference a metric.

#### **Q46 What is the common table function for grouping data?**

##### **SUMMARIZE()**

- Main groupby function in SSAS.
- Recommended practice is to specify table and group by columns but not metrics. You can use ADDCOLUMNS function.

##### **SUMMARIZECOLUMNS**

- New group by function for SSAS and Power BI Desktop; more efficient.
- Specify group by columns, table, and expressions.

#### **Q47 What are some benefits of using Variables in DAX ?**

DAX or Data Analysis Expression is a functional language which can create calculated columns and/or measures for smarter calculations to limit the data the dashboard has to fetch and visualise.

#### **Q48 What is the difference between Distinct() and Values() in DAX?**

Generally, the Distinct() and Values() functions are the same in Power BI. The only difference between them is that the Values() function don't calculate null values, whereas the Distinct() function calculates even the null values.

#### **Q49 What is Power Query?**

Power Query is a business intelligence tool designed by Microsoft for Excel. Power Query allows you to import data from various data sources and will enable you to clean, transform and reshape your data as per the requirements. Power Query allows you to write your query once and then run it with a simple refresh.

#### **Q50 Define Power Query?**

Power Query is a business intelligence tool designed by Microsoft for Excel. Power Query is an ETL tool that allows you to import data from various data sources and will enable you to clean, transform and reshape your data as per the requirements. Power Query allows you to write your query once and then run it with a simple refresh.

With this:

- You can import data from various sources like databases from files
- Append and join data from a wide range of sources
- You can shape data as needed by adding and removing it

**Q51 How are a Power BI Dashboard and Report different from each other?**

To understand the difference between Power BI Dashboard and Report, let's run through some quick points.

Capability	Report	Dashboard
Pages	Can be of one or more pages.	Consists of one page only
Data sources	It has a single dataset per report.	Can have data tiles from one or more datasets or reports.
Filtering	Can perform slicing, filtering, and highlighting.	Cannot filter or slice reports.
Set alerts	No option for setting alerts.	Enable setting email alerts
Featured reports	No option for creating a featured dashboard.	Enables to set only one dashboard as a featured dashboard.
Accessing tables and fields in datasets	Provides options to view dataset tables, values, and fields.	Cannot view or access underlying datasets tables and fields.

**Q52 What are the data destinations for Power Queries?**

There are two destinations for output we get from power query:

1. Load to a table in a worksheet.
2. Load to the Excel Data Model.

**Q53 What is query folding in Power Query?**

Query folding is when steps defined in Power Query/Query Editor are translated into SQL and executed by the source database rather than the client machine. It's important for processing performance and scalability, given limited resources on the client machine.

**Q54 What are some common Power Query/Editor Transforms?**

Changing Data Types, Filtering Rows, Choosing/Removing Columns, Grouping, Splitting a column into multiple columns, Adding new Columns ,etc.

**Q55 Can SQL and Power Query/Query Editor be used together?**

Yes, a SQL statement can be defined as the source of a Power Query/M function for additional processing/logic. This would be a good practice to ensure that an efficient database query is passed to the source and avoid unnecessary processing and complexity by the client machine and M function.

**Q56 What are query parameters and Power BI templates?**

Query parameters can be used to provide users of a local Power BI Desktop report with a prompt, to specify the values they're interested in.

- The parameter selection can then be used by the query and calculations.
- PBIX files can be exported as Templates (PBIT files).
- Templates contain everything in the PBIX except the data itself.

Parameters and templates can make it possible to share/email smaller template files and limit the amount of data loaded into the local PBIX files, improving processing time and experience.

**Q57 Which language is used in Power Query?**

A new programming language is used in power query called M-Code. It is easy to use and similar to other languages. M-code is case-sensitive language.

**Q58 Why do we need Power Query when Power Pivot can import data from mostly used sources?**

Power Query is a self-service ETL (Extract, Transform, Load) tool which runs as an Excel add-in. It allows users to pull data from various sources, manipulate said data into a form that suits their needs and load it into Excel. It is most optimum to use Power Query over Power Pivot as it lets you not only load the data but also manipulate it as per the users needs while loading.

**Q59 Name some commonly used tasks in the Query Editor.**

Some commonly used tasks in the Query Editor are:

Connect to Data: Get Data from various sources and Transform data.

Shape Data: Transform your data according to requirement to clean and shape it

Group Rows: You can group the values of many rows into one single value by summarizing

Pivot Columns: Pivot columns and create a table with aggregated values

Create Custom Columns: You can use custom formulas to create new columns in your table

Advanced Editor: You can make modifications to the data using Advanced Query Editor with query.

## REFERENCE

<https://www.interviewbit.com/data-analyst-interview-questions/>

<https://www.upgrad.com/blog/data-analyst-interview-questions-and-answer/>

<https://www.simplilearn.com/tutorials/data-analytics-tutorial/data-analyst-interview-questions>

<https://www.springboard.com/blog/data-analytics/excel-interview-questions/>

<https://www.javatpoint.com/what-is-database>

<https://www.javatpoint.com/sql-interview-questions>

<https://www.geeksforgeeks.org/sql-interview-questions/>

<https://www.edureka.co/blog/interview-questions/top-tableau-interview-questions-and-answers/>

<https://mindmajix.com/tableau-interview-questions>

<https://www.interviewbit.com/tableau-interview-questions/>

<https://www.simplilearn.com/tableau-interview-questions-and-answers-article>

<https://www.javatpoint.com/tableau-interview-questions>

<https://intellipaat.com/blog/interview-question/tableau-interview-questions/>

<https://www.simplilearn.com/power-bi-interview-questions-and-answers-article>

<https://www.edureka.co/blog/interview-questions/power-bi-interview-questions/>

<https://mindmajix.com/power-bi-interview-questions>

<https://www.javatpoint.com/power-bi-interview-questions>