



CLOUDYML



100+ DATA ANALYST INTERVIEW QNA PDF COLLECTION



Akash Raj
Data Scientist

1. What is Data Integrity?

Data Integrity is the assurance of accuracy and consistency of data over its entire life-cycle and is a critical aspect of the design, implementation, and usage of any system which stores, processes, or retrieves data. It also defines integrity constraints to enforce business rules on the data when it is entered into an application or a database.

2. What is the Difference Between Joining and Blending in Tableau?

Combining the data from two or more different sources is data blending, such as Oracle, Excel, and SQL Server. In data blending, each data source contains its own set of dimensions and measures. Combining the data between two or more tables or sheets within the same data source is data joining. All the combined tables or sheets contain a common set of dimensions and measures.

3. What is slicing in Python?

When data is ingested into Power BI, it is basically stored in Fact and Dimension tables.

Fact tables: The central table in a star schema of a data warehouse, a fact table stores quantitative information for analysis and is not normalized in most cases.

Dimension tables: It is just another table in the star schema that is used to store attributes and dimensions that describe objects stored in a fact table.

4. What is the difference between NOW() and CURRENT_DATE() in SQL?

NOW() returns a constant time that indicates the time at which the statement began to execute. (Within a stored function or trigger, NOW() returns the time at which the function or triggering statement began to execute.

The simple difference between NOW() and CURRENT_DATE() is that NOW() will fetch the current date and time both in format 'YYYY-MM_DD HH:MM:SS' while CURRENT_DATE() will fetch the date of the current day 'YYYY-MM_DD'.

5. What's a Fourier transform?

A Fourier transform is a generic method to decompose generic functions into a superposition of symmetric functions. The Fourier transform finds the set of cycle speeds, amplitudes, and phases to match any time signal. A Fourier transform converts a signal from time to frequency domain—it's a very common way to extract features from audio signals or other time series such as sensor data.

6. What is x-velocity in Power Pivot?

X-Velocity is the in-memory analytics engine behind Power Pivot that loads and handles huge data in Power BI. It stores data in columnar storage that results in faster processing.

7. Gantt chart in Tableau ?

A Tableau Gantt chart illustrates the duration of events as well as the progression of value across the period. Along with the time axis, it has bars. The Gantt chart is primarily used as a project management tool, with each bar representing a project job.

8. What in Excel is a macro?

An Excel macro is an algorithm or a group of steps that helps automate an operation by capturing and replaying the steps needed to finish it. Once the steps have been saved, you may construct a Macro that the user can alter and replay as often as they like.

9. What is the lambda function in Python?

Python Lambda Functions are anonymous function means that the function is without a name. As we already know that the def keyword is used to define a normal function in Python. Similarly, the lambda keyword is used to define an anonymous function in Python.

Eg. `lambda_cube = lambda y: y*y*y`

10. What is the difference between SQL and MySQL?

Autoencoders are artificial neural networks that learn without any supervision. Here, these networks have the ability to automatically learn by mapping the inputs to the corresponding outputs.

Autoencoders, as the name suggests, consist of two entities:

Encoder: Used to fit the input into an internal computation state

Decoder: Used to convert the computational state back into the output

11. What are Filters in Power BI?

The term "Filter" is self-explanatory. Filters are mathematical and logical conditions applied to data to filter out essential information in rows and columns. The following are the variety of filters available in Power BI:

- Manual filters
- Auto filters
- Include/Exclude filters
- Drill-down filters
- Cross Drill filters

12. What is concurrency control in DBMS?

This is a process of managing simultaneous operations in a database so that database integrity is not compromised. The following are the two approaches involved in concurrency control:

Optimistic approach – Involves versioning

Pessimistic approach – Involves locking

13. What is a checkpoint in DBMS and when does it occur?

A checkpoint is a mechanism where all the previous logs are removed from the system and are permanently stored on the storage disk. So, basically, checkpoints are those points from where the transaction log record can be used to recover all the committed data up to the point of crash.

14. What are groups in Tableau?

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.

15. How are nested IF statements used in Excel?

The function IF() can be nested when we have multiple conditions to meet. The FALSE value in the first IF function is replaced by another IF function to make a further test.

16. What are the ways to detect outliers?

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 $\pm (3 \times \text{standard deviation})$.

17. What is a Recursive Stored Procedure?

A stored procedure that calls itself until a boundary condition is reached, is called a recursive stored procedure. This recursive function helps the programmers to deploy the same set of code several times as and when required.

18. What is the shortcut to add a filter to a table in EXCEL?

The filter mechanism is used when you want to display only specific data from the entire dataset. By doing so, there is no change being made to the data. The shortcut to add a filter to a table is **Ctrl+Shift+L**.

19. What is DAX in Power BI?

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.

20. What is the Difference Between a Shallow Copy and Deep Copy in python?

Deepcopy creates a different object and populates it with the child objects of the original object. Therefore, changes in the original object are not reflected in the copy. `copy.deepcopy()` creates a Deep Copy. Shallow copy creates a different object and populates it with the references of the child objects within the original object. Therefore, changes in the original object are reflected in the copy. `copy.copy` creates a Shallow Copy.

21. How can you remove duplicate values in a range of cells?

To delete duplicate values in a column, select the highlighted cells, and press the delete button. After deleting the values, go to the 'Conditional Formatting' option present in the Home tab. Choose 'Clear Rules' to remove the rules from the sheet. 2. You can also delete duplicate values by selecting the 'Remove Duplicates' option under Data Tools present in the Data tab

22. Define shelves and sets in Tableau?

Shelves: Every worksheet in Tableau will have shelves such as columns, rows, marks, filters, pages, and more. By placing filters on shelves we can build our own visualization structure. We can control the marks by including or excluding data.

Sets: The sets are used to compute a condition on which the dataset will be prepared. Data will be grouped together based on a condition. Fields which are responsible for grouping are known assets. For example – students having grades of more than 70%.

23. What data sources can Power BI connect to?

The list of data sources for Power BI is extensive, but it can be grouped into the following:

Files: Data can be imported from Excel (.xlsx, xlsm), Power BI Desktop files (.pbix) and Comma Separated Value (.csv).

Content Packs: It is a collection of related documents or files that are stored as a group. In Power BI, there are two types of content packs, firstly those from services providers like Google Analytics, Marketo, or Salesforce, and secondly those created and shared by other users in your organization.

Connectors to databases and other datasets such as Azure SQL, Database and SQL, Server Analysis Services tabular data, etc.

24. What are the different integrity rules present in the DBMS?

The different integrity rules present in DBMS are as follows:

Entity Integrity: This rule states that the value of the primary key can never be NULL. So, all the tuples in the column identified as the primary key should have a value.

Referential Integrity: This rule states that either the value of the foreign key is NULL or it should be the primary key of any other relation.

25. What are some common clauses used with SELECT query in SQL?

Some common SQL clauses used in conjunction with a SELECT query are as follows:

WHERE clause in SQL is used to filter records that are necessary, based on specific conditions.

ORDER BY clause in SQL is used to sort the records based on some field(s) in ascending (ASC) or descending order (DESC).

GROUP BY clause in SQL is used to group records with identical data and can be used in conjunction with some aggregation functions to produce summarized results from the database.

HAVING clause in SQL is used to filter records in combination with the GROUP BY clause. It is different from WHERE, since the WHERE clause cannot filter aggregated records.

26. What is the difference between count, counta, and countblank in Excel?

The count function is very often used in Excel. Here, let's look at the difference between count, and its variants - counta and countblank.

1. COUNT

It counts the number of cells that contain numeric values only. Cells that have string values, special characters, and blank cells will not be counted.

2. COUNTA

It counts the number of cells that contain any form of content. Cells that have string values, special characters, and numeric values will be counted. However, a blank cell will not be counted.

3. COUNTBLANK

As the name suggests, it counts the number of blank cells only. Cells that have content will not be taken into consideration.

27. What is Density-based Clustering?

Density-Based Clustering is an unsupervised machine learning method that identifies different groups or clusters in the data space. These clustering techniques are based on the concept that a cluster in the data space is a contiguous region of high point density, separated from other such clusters by contiguous regions of low point density.

Partition-based(K-means) and Hierarchical clustering techniques are highly efficient with normal-shaped clusters while density-based techniques are efficient in arbitrary-shaped clusters or detecting outliers.

28. How to create empty tables with the same structure as another table?

To create empty tables:

Using the INTO operator to fetch the records of one table into a new table while setting a WHERE clause to false for all entries, it is possible to create empty tables with the same structure. As a result, SQL creates a new table with a duplicate structure to accept the fetched entries, but nothing is stored into the new table since the WHERE clause is active.

29. What is a Parameter in Tableau? Give an Example

A parameter is a dynamic value that a customer could select, and you can use it to replace constant values in calculations, filters, and reference lines.

For example, when creating a filter to show the top 10 products based on total profit instead of the fixed value, you can update the filter to show the top 10, 20, or 30 products using a parameter.

30. How will you write the formula for the following in Excel?

Multiply the value in cell A1 by 10, add the result by 5, and divide it by 2.

To write a formula for the above-stated question, we have to follow the PEDMAS Precedence.

The **correct answer** is $((A1*10)+5)/2$.

Answers such as $=A1*10+5/2$ and $=(A1*10)+5/2$ are not correct. We must put parentheses brackets after a particular operation.

31. Define the term 'Data Wrangling'

Data Wrangling is the process wherein raw data is cleaned, structured, and enriched into a desired usable format for better decision making. It involves discovering, structuring, cleaning, enriching, validating, and analyzing data. This process can turn and map out large amounts of data extracted from various sources into a more useful format.

32. What are the best methods for data cleaning?

Create a data cleaning plan by understanding where the common errors take place and keep all the communications open. Before working with the data, identify and remove the duplicates. This will lead to an easy and effective data analysis process. Focus on the accuracy of the data. Set cross-field validation, maintain the value types of data, and provide mandatory constraints. Normalize the data at the entry point so that it is less chaotic. You will be able to ensure that all information is standardized, leading to fewer errors on entry

33. Explain 4 steps to use CTE in sql.

All CTE starts with "with" clause.

After with you need to define CTE name and the field names. For instance in the below code snippet I have 3 fields Count, Column and Id. The name of CTE is "MyTemp".

Once you have defined CTE we need to specify the SQL which will give the result for the CTE.

Finally you can use the CTE in your SQL query

34. What are the various types of refresh options provided in Power BI?

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.

35. What are Ensemble Methods?

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model. Random Forest is a type of ensemble method. The number of component classifier in an ensemble has a great impact on the accuracy of the prediction, although there is a law of diminishing results in ensemble construction.

36. What are constraints in SQL?

Constraints are the rules that we can apply on the type of data in a table. That is, we can specify the limit on the type of data that can be stored in a particular column in a table using constraints. NOT NULL, UNIQUE, DEFAULT, PRIMARY KEY, FOREIGN KEY, CHECK are the different constraints in SQL.

37. How do you apply a single format to all the sheets present in a workbook?

To apply the same format to all the sheets of a workbook, follow the given steps:
Right-click on any sheet present in that workbook

Then, click on the Select All Sheets option

Format any of the sheets and you will see that the format has been applied to all the other sheets as well

38. Define the term 'Data Wrangling.

Data Wrangling is the process wherein raw data is cleaned, structured, and enriched into a desired usable format for better decision making. It involves discovering, structuring, cleaning, enriching, validating, and analyzing data. This process can turn and map out large amounts of data extracted from various sources into a more useful format.

39. What are the best methods for data cleaning?

Create a data cleaning plan by understanding where the common errors take place and keep all the communications open. Before working with the data, identify and remove the duplicates. This will lead to an easy and effective data analysis process. Focus on the accuracy of the data. Set cross-field validation, maintain the value types of data, and provide mandatory constraints. Normalize the data at the entry point so that it is less chaotic. You will be able to ensure that all information is standardized, leading to fewer errors on entry

40. What is the difference between HAVING and WHERE clauses?

WHERE Clause is used to filter the records from the table based on the specified condition. HAVING Clause is used to filter records from the groups based on the specified condition. WHERE Clause can be used without GROUP BY Clause. HAVING Clause cannot be used without GROUP BY Clause. WHERE Clause implements in row operations. HAVING Clause implements in column operation. WHERE Clause cannot contain aggregate function. HAVING Clause can contain aggregate function

41. 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.

Become Job-Ready In 5 Steps With CloudyML

01 You will be learning all the technical topics from basics to advanced level and solving assignments to get complete hands-on practical learning experience.



02 After that, you will be doing Industrial Projects that will boost up your confidence and enhance your resume



03 simultaneously with the course, you will be learning LinkedIn Growth Hacks, Pro Resume Building and Proven techniques to find jobs through various online platforms.



04 After learning all the core topics and doing projects, you can do Industrial Internship with Us and get Real-World exposure that will super enhance your skills



05 With our Placement Guarantee Program, which comes with Job Referrals, Unlimited Mock Interviews, Company Tie-up Interviews etc, you will be able to get your Dream Job and you can start your career in Data Science, Analytics & Engineering Domain.



Visit Our Website Now



42. Mention the order of operations used in Excel while evaluating formulas.

The order of operations in Excel is referred to as PEMDAS. Shown below is the order of precedence while performing an Excel operation.

- Parentheses
- Exponentiation
- Division/Multiplication
- Addition
- Subtraction

43. What is map function in Python?

map function executes the function given as the first argument on all the elements of the iterable given as the second argument. If the function given takes in more than 1 arguments, then many iterables are given

44. How many report formats are available in Excel?

There are three report formats available in Excel; they are:

1. Compact Form
2. Outline Form
3. Tabular Form

45. What are sets in Tableau?

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

46. 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.

47. 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.

48. How to fetch unique records from a table in SQL?

SQL DISTINCT clause is used to remove the duplicated columns from the result set.

The distinct keyword is used with the select keyword in conjunction. It is helpful when we avoid duplicate values present in the specific columns/tables. The unique values are fetched when we use the distinct keyword.

49. How are Pivot tables used to filter data in Excel?

You can filter data according to your requirements with Excel Pivot tables. Place the field on which you want the data to be filtered. Then open the drop-down list of the field you put in the Filter area from the pivot table and choose your line.

50. What is R2? What are some other metrics that could be better than R2 and why??

R-squared (R²) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. R-squared does not measure goodness of fit. R-squared does not measure predictive error. R-squared does not allow you to compare models using transformed responses. R-squared does not measure how one variable explains another. Some better metrics that could be better than R² are:

Mean Squared Error (MSE).

Root Mean Squared Error (RMSE).

Mean Absolute Error (MAE)

51. What is the curse of dimensionality??

The curse of dimensionality basically means that the error increases with the increase in the number of features. It refers to the fact that algorithms are harder to design in high dimensions and often have a running time exponential in the dimensions..

52. What are advantages of plotting your data before performing analysis ?

It can help identify obvious errors, as well as better understand patterns within the data, detect outliers or anomalous events, find interesting relations among the variables

53. How would you explain a confidence interval to an engineer with no statistics background? What does 95% confidence mean? ?

A 95% confidence interval, for example, implies that were the estimation process repeated again and again, then 95% of the calculated intervals would be expected to contain the true parameter value.

54. How do you deal with some of your predictors being missing?

Simple approaches include taking the average of the column and use that value, or if there is a heavy skew the median or mode might be better. A better approach, you can perform regression or nearest neighbor imputation on the column to predict the missing values. Then continue on with your analysis/model.

55. What is a sensitivity analysis in the decision making process?

Sensitivity analysis is a method for predicting the outcome of a decision if a situation turns out to be different compared to the key predictions. It helps in assessing the riskiness of a strategy. Helps in identifying how dependent the output is on a particular input value.

56. How do you interpret the data using statistical techniques

Most Important Methods For Statistical Data Analysis Mean.

1. Standard Deviation.
2. Regression.
3. Sample Size.
4. Determination.
5. Hypothesis Testing.

57. 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.

58. What is Map Reduce?

MapReduce facilitates concurrent processing by splitting petabytes of data into smaller chunks, and processing them in parallel on Hadoop commodity servers. In the end, it aggregates all the data from multiple servers to return a consolidated output back to the application.

59. What is a Pivot Table?

A pivot table is a table of grouped values that aggregates the individual items of a more extensive table within one or more discrete categories.

60. Difference between 1-Sample T-test, and 2-Sample T-test?

The 2-sample t-test takes your sample data from two groups and boils it down to the t-value. The process is very similar to the 1-sample t-test, and you can still use the analogy of the signal-to-noise ratio. Unlike the paired t-test, the 2-sample t-test requires independent groups for each sample.

61. variance and covariance difference?

Variance and covariance are mathematical terms frequently used in statistics and probability theory. Variance refers to the spread of a data set around its mean value, while a covariance refers to the measure of the directional relationship between two random variable.

62. What is the statistical power of sensitivity?

The statistical power of an A/B test refers to the test's sensitivity to certain magnitudes of effect sizes. More precisely, it is the probability of observing a statistically significant result at level alpha (α) if a true effect of a certain magnitude (MEI) is in fact present..

63. What is the difference between covariance and correlation?

"Covariance" indicates the direction of the linear relationship between variables. "Correlation" on the other hand measures both the strength and direction of the linear relationship between two variables.

64.What is negative indexing? Why is it needed? Can you give an example for the same in python ?

This means that the index value of -1 gives the last element, and -2 gives the second last element of an array. The negative indexing starts from where the array ends.

example: for list L = [0,2,35,3]; L[-1] will print 3 in Python.

65. What is the condition for using a t-test or a z-test?

z-test is used for it when sample size is large, generally $n > 30$. Whereas t-test is used for hypothesis testing when sample size is small, usually $n < 30$ where n is used to quantify the sample size.

66. What is the main difference between overfitting and underfitting?

Overfitting is a modeling error which occurs when a function is too closely fit to a limited set of data points. Underfitting refers to a model that can neither model the training data nor generalize to new data.

67. Deal with unbalanced binary classification?

Techniques to Handle unbalanced Data:

1. Use the right evaluation metrics
2. Use K-fold Cross-Validation in the right way
3. Ensemble different resampled datasets
4. Resample with different ratios
5. Design your own models

68. What is Activation function?

Activation functions are mathematical equations that determine the output of a neural network model. It is a non-linear transformation that we do over the input before sending it to the next layer of neurons or finalizing it as output.

69. What is Dimension reduction?

Dimensionality Reduction is used to reduce the feature space with consideration by a set of principal features.

70. Why is mean square error a bad measure of model performance?

Mean Squared Error (MSE) gives a relatively high weight to large errors — therefore, MSE tends to put too much emphasis on large deviations.

71. How to create sets in tableau?

Sets are custom fields used to compare and ask questions about a subset of data. For creating a set on dimension, right-click on a dimension in data pane and select create -> set. In general tab select the fields that will be considered for computing the set. Specify the conditions to create set in conditions tab and you also have the option to select top N members in dataset based on any field in the top tab. When a set is created it divides the measure into two parts namely in and out of the set based on the conditions applied by the user.

72. What is live and extract in tableau?

Extract is a snapshot of the data optimized for aggregation. Extracts are loaded into the system and hence improve the performance of tableau. Whereas extracts won't help in situations where data is updated continuously because then we manually need to refresh the data for all the updates but using a live connection might slow the processing but will definitely update the data source itself. So, live connection should be used only when data is continuously updating otherwise extract file is preferred.

73. What is absolute cell reference in excel?

Absolute cell reference is a locked reference and ensures that the rows and columns will not change on copying the cell. We add a '\$' symbol in front of the row and column number. When the cell is \$A\$2, both rows and columns remain unchanged on copying. But if it is \$A2, it means that the column remains unchanged, but the row changes on copying. Similarly, for A\$2, the column changes, but the row remains unchanged on copying.

74. Differentiate between K-Means and KNN algorithms?

KNN algorithms is Supervised Learning where-as K-Means is Unsupervised Learning. With KNN, we predict the label of the unidentified element based on its nearest neighbour and further extend this approach for solving classification/regression-based problems.

K-Means is Unsupervised Learning, where we don't have any Labels present, in other words, no Target Variables and thus we try to cluster the data based upon their coordinates and try to establish the nature of the cluster based on the elements filtered for that cluster.

75. What are the different data types of Tableau.

There are 7 data types in Tableau

- Boolean (True/False)
- Date (Individual Value)
- Date and Time
- Geography
- Text or String
- Decimal Number
- Whole Number

76. What is a dual-axis?

It is a function in Tableau that showcases two scales of two measures in a single graph. This is very similar to the function found on Microsoft Office products where a single graph has line and bar elements. In most cases, it has either two X or two Y axes.

A dual-axis is typically used to show trend lines and historical data. An example would be total revenue vs profit across 12 months.

77. What is the shortcut to add a filter to a table?

The filter mechanism is used when you want to display only specific data from the entire dataset. By doing so, there is no change being made to the data. The shortcut to add a filter to a table is Ctrl+Shift+L.

78. What are the wildcards available in Excel?

Wildcards only work with text data. Excel has three wildcards.

1. * (Asterisk)

This refers to any number of characters.

2.? (Question mark)

It represents one single character.

3.~ (Tilde)

It is used to identify a wildcard character (~, *, ?) in the text

79. What is analysis in tableau?

Tableau comes with inbuilt features to analyze the data plotted on a chart. We have various tools such as adding an average line to the chart which tableau calculates itself after we drop the tool on the chart. Some other features include clustering, percentages, forming bands of a particular range and various other tools to explore and inspect data. All these tools are available in analyze tab on each sheet used to create any chart. The features become visible only when they are applicable to the worksheet.

80. What is quick filter in tableau?

Whenever using a filter in Tableau, it comes with some options to change the functionality of filter very easily, such as using it as a single value drop down or single value list or multiple value list or multiple value drop down and various other options. After we set a filter to a sheet just right click on the sheet and there you can see all the quick filter options. Changes made to these options will also change the aesthetics of filter shown on the sheet.

81. How do you apply a single format to all the sheets present in a workbook?

To apply the same format to all the sheets of a workbook, follow the given steps:

- Right-click on any sheet present in that workbook
- Then, click on the Select All Sheets option
- Format any of the sheets and you will see that the format has been applied to all the other sheets as well

82. Explain SUM and SUMIF functions.

SUM: The SUM function is used to calculate the sum of all the values that are specified as a parameter to it. The syntax of this function is as follows:

`SUM(number1, number2, ...)`

SUMIF: This function is used to calculate the sum of values that comply with a given condition.

SYNTAX:

`SUMIF(range, criteria, [sum_range])`

where,

range specifies the range of cells to be evaluated

criteria provides the condition to be met

sum_range is optional and provides the actual cells to be summed up

83. How do you add a Note to a cell?

To add a Note, select the cell and right-click on the same. then select the New Note option and type in any note that you wish to. In case you want to delete the Note, follow the same procedure and select the Delete Note option. Notes are indicated by a red triangle at the top-right corner of the cell.

84. What are the different types of functions generally used in Tableau?

In Tableau, we have a lot of processing and analytical freedom with the virtue of functions available.

With the help of different types of functions, we can perform a lot of analytical operations on the data.

The main categories of Tableau function are:

String function: These functions like ASCII, CHAR, FIND, ISDATE, LOWER, etc, are known as string functions because they work on the string values or characters to manipulate them.

Date function: We use date functions to apply logical as well as arithmetic operations on date values present at the data source. Using the date functions we can manipulate the date values by changing the old values, creating new ones or searching data on the basis of specific dates.

Some commonly used date functions in Tableau are DATEADD, MAKEDATE, ISDATE, MAKETIME, MONTH, MIN/MAX, TODAY, NOW, etc.

Logical function: We use logical functions to perform logical or relational operations on data in Tableau.

Some commonly used logical functions in Tableau are, CASE, IF, IFNULL, ISNULL, ZN, etc.

Aggregate function: We use aggregate functions to apply aggregation on data values in different ways.

Some important aggregation functions used in Tableau are; AVG, ATTR, MAX, MEDIAN, MIN, PERCENTILE, SUM, STDDEV, etc.

User function: We use functions to manage the users registered on Tableau Server or Tableau Online.

Commonly used user functions are, FULLNAME, ISFULLNAME, ISUSERNAME, USERDOMAIN, USERNAME, etc.

85. What do you understand by context filters?

Context filters are used to apply context on the data under analysis.

By applying a context we set a perspective according to which we can see the charts and graphs.

For example, we have sales data of an electronic store and we want to conduct our analysis only for the corporate sector or segment.

To do this, we have to apply a context filter on our Tableau sheet. Once we add the context for the Corporate segment from the Add to context option, all the charts present on the sheet will only show data relevant to the Corporate segment.

In this way, we can apply a context to our analysis in Tableau.

86. 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.

87. Why and how would you use a custom visual file?

A custom visual file is used when none of the pre existing visuals fit the business needs. Custom visual files are generally created by Developers which can be used in the same way as prepackaged files.

GET 100% PLACEMENT IN DATA SCIENCE DOMAIN

SWITCH FROM ANY NON-CODING BACKGROUND

Join India's #1 Most Affordable Industry Oriented Course With 100% Placement Guarantee

COMPANIES WHERE OUR LEARNERS ARE SHINING



+100
MORE

PRACTICAL
LEARNING
EXPERIENCE

LEARN FROM
BASIC TO
ADVANCED

GET DOUBT
CLEARANCE
SUPPORT

GET
INDUSTRIAL
INTERNSHIP

50+ INDUSTRIAL
PROJECTS &
MORE FEATURES

VISIT OUR WEBSITE NOW



88. What are the various type of users who can use Power BI?

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.

89. Can you name the wildcards in Excel?

There are 3 wildcards in Excel that can be used in formulas.

Asterisk (*) – 0 or more characters. For example, Ex* could mean Excel, Extra, Expertise, etc.

Question mark (?) – Represents any 1 character. For example, R?ain may mean Rain or Ruin.

Tilde (~) – Used to identify a wildcard character (~, *, ?). For example, If you need to find the exact phrase India* in a list. If you use India* as the search string, you may get any word with India at the beginning followed by different characters (such as Indian, Indiana). If you have to look for India" exclusively, use ~.

Hence, the search string will be india~*. ~ is used to ensure that the spreadsheet reads the following character as is, and not as a wildcard.

90. What is cascading filter in tableau?

Cascading filters can also be understood as giving preference to a particular filter and then applying other filters on previously filtered data source. Right-click on the filter you want to use as a main filter and make sure it is set as all values in dashboard then select the subsequent filter and select only relevant values to cascade the filters. This will improve the performance of the dashboard as you have decreased the time wasted in running all the filters over complete data source.

91. What is the difference between .twb and .twbx extension?

A .twb file contains information on all the sheets, dashboards and stories, but it won't contain any information regarding data source. Whereas .twbx file contains all the sheets, dashboards, stories and also compressed data sources. For saving a .twbx extract needs to be performed on the data source. If we forward .twb file to someone else than they will be able to see the worksheets and dashboards but won't be able to look into the dataset.

92. What is Excel BI Toolkit.

Excel BI toolkit is the other half of the Microsoft self-service BI (one half being Power BI). This toolkit is composed of Excel and a few more add-ins such as Power Query, Power Pivot, Power View, Power Map. All the tools in the Excel BI toolkit serve a special purpose like importing, modeling, preparing and visualizing data. Generally, the tools are used to create reports by consolidating data from multiple data sources and to model the datasets. Excel and other add-ins can be used independently or along with each other to optimize BI capabilities of the toolkit.

93. What are the various Power BI versions?

Power BI Premium capacity-based license, for example, allows users with a free license to act on content in workspaces with Premium capacity. A user with a free license can only use the Power BI service to connect to data and produce reports and dashboards in My Workspace outside of Premium capacity. They are unable to exchange material or publish it in other workspaces. To process material, a Power BI license with a free or Pro per-user license only uses a shared and restricted capacity. Users with a Power BI Pro license can only work with other Power BI Pro users if the material is stored in that shared capacity. They may consume user-generated information, post material to app workspaces, share dashboards, and subscribe to dashboards and reports. Pro users can share material with users who don't have a Power BI Pro subscription while workspaces are at Premium capacity.

94. Where is the Power BI data stored?

All the data that we import in Power BI from different data sources get stored in either of the two tables in a data warehouse; Fact tables and Dimension tables. The fact tables are the central/main table of a star schema which contains all the measure (quantitative data) values. It has primary keys and all the dimension tables are linked to the fact table. The fact table is not usually normalized.

While a dimension table is a table in a database which contains all the attribute values (information about data) for the data stored in fact table. Every dimension table in a star schema is linked to a fact table.

95. What are sets and groups?

Sets and groups are used group data based on some specific conditions. The main difference between these two is that a group can divide the dataset into multiple groups whereas a set can have only two options which is either in or out. A user should choose to apply group or sets based on the requirements.

96. What is the difference between a function and a formula in Excel?

A formula is a user-defined expression that calculates a value. A function is pre-defined built-in operation that can take the specified number of arguments. A user can create formulas that can be complex and can have multiple functions in it. For example, =A1+A2 is a formula and =SUM(A1:A10) is a function.

97. How to add custom colors in tableau?

Tableau offers various colour palettes which a user can use to define a legend or to be used in formatting. But sometimes due to client's requirements we need to use some specific colours in our view. So instead of defining the colour each and every time you can create a custom palette which can be used whenever you open a workbook. To create a custom colour palette, go to "My tableau Repository" in the documents folder of your system. Open "Preference.tps" file in a text editor to create the custom palette. By default, the file will be empty with just opening and closing line of workbook. You need to define the colour palette in between these two lines and there are three types of colour categories that can be defined. One of them is categorical which is defined using type = "Regular" and second one is sequential and it comes under the type = "ordered-sequential" and last comes diverging colour which is of the type = "ordered-diverging". Also, kindly note that the colour should be defined in the HTML #RRGGB order.

98. What is a legend in Power BI?

Legends are part of Power BI visuals. They represent categories in a visual and are usually color-coded. In some visuals, we can add a category dimension as legend explicitly. One example can be in stacked bar/column charts, where the stacks represent different categories, and these categories are color-coded. These categories are represented in the legend. And every visual where the legend is applicable has a separate formatting section where we can specify its font and font size, whether we need it to be visible or not, or even its position(top, bottom, left, right)

99. What is Power BI-Embedded?

Power BI Embedded is an analytics solution, provided as a Microsoft Azure Service, a platform-as-a-service (PaaS), wherein the developers & ISVs (Individual Software Vendors) can easily embed their dashboards, Visuals & reports into an application for their customers.

Fully interactive reports & visuals can be embedded into the applications.

The great thing about Power BI Embedded is that your customers are not required to have knowledge about Power BI. Power BI gives you the capability to create an embedded application using two different methods.

The first method is using a Power BI Pro account & the second method is using the service principal.

The Power BI Pro account acts as the master account of your application. This account will help you generate the required embed tokens for providing access to your customers to view the shared Power BI dashboards and reports.

The service principal embeds Power BI Dashboards/Reports/Visuals into an application using an app-only token. Now this will allow you to generate the embed tokens for providing access to your shared application's Power BI dashboards and reports.

100. What is the Ribbon and what does it contain?

The Ribbon refers to the row of buttons and icons at the top of your worksheet. These include common tabs like Home, Insert, Page Layout, and Data.

You can customize the Ribbon and collapse or expand it using CTRL+F1. Some tabs only appear when you select a relevant item, such as a chart or table.

101. Explain Pivot tables along with their features?

Pivot Tables are statistical tables that condense data of those tables that have extensive information. The summary can be based on any field such as sales, averages, sums, etc that the pivot table represents in a simple and intelligent manner.

Features:

Some of the features of Excel Pivot Tables are as follows:

Allow the display of exact data you have to analyze

Provide various angles to view the data

Allow you to focus on important details

Comparison of data is very handy

Pivot tables can detect different patterns, relationships, data trends, etc

They can create instant data

Accurate reports

Serve the base for Pivot charts

102. What is a Horizontal Lookup in Microsoft Excel?

Horizontal Lookup or HLOOKUP looks for a value from the topmost row of the table horizontally and then moves in a downward direction.

Syntax of HLOOKUP is, = HLOOKUP (Val, givetable, row_no, [rnge_look]),

Where,

Val is the value to be searched in the first row of the table.

givetable is the row/rows that are sorted in ascending order.

row_no is the row from which the lookup value is to be recovered.

[rnge_look] is not a mandatory argument where TRUE (default) means inexact match and FALSE means exact match.

103. What is the use of cycle fields in tableau?

Cycle fields help in switching and trying different colour combinations or views in a cyclic order. It will work only if we have a chart that allows more than one measure such as stacked bar chart and we are unable to finalize the visualizations then we can use cycle fields. To use cycle field, click on analysis menu in the toolbar then select cycle fields to take a quick look at an alternative visualization.

104. What is the calculated column in Power BI?

Calculated columns are built to extend the data attributes. They are those columns that are created when the available columns in the data do not serve our purpose or we are not able to generate any useful insight from the same. That is when calculated columns come into the picture. And these are created using different DAX functions as per need. We can have a simple example of a date where we do not want to work with the complete date but individual day, month, or year. For this, in the data part of the Power BI desktop, we have the option of a 'New Column.' When we click that, we get a formula bar on top (like excel), where it asks us to type the new column name with its calculation. So, if we consider the data example, it will be something like this: yearcol = YEAR([Date column]) where YEAR is a DAX function. And 'yearcol' is your calculated column.

105. How to integrate tableau with website?

To integrate Tableau with a website, developer needs to have complete understanding of Javascript API for tableau. It contains all the types of functions required to view and control a tableau worksheet and dashboard directly from the website instead of interacting directly with the worksheet. Tableau has provided a complete list of functions used in Javascript API and implementation of some of the important functions is explained in the tutorial section also which is free for all users. All the tableau dashboards or worksheets when published on tableau public or online or server are integrated with the Javascript API by default developer only needs to call that API in the HTML code and start interacting with it.

106. How to use Power BI in excel?

To use Power BI in Excel, there is an Analyse in Excel option for every report in the Power BI service. To use it, you will need to enable editing and enable content for the report for the first time.

So, what this option provides us is that it gives us the underlying data set of our Power BI report. It comes as a data connection in excel. And we get to play with the data in excel. It is up to us how we analyze the same data, either through pivot tables, charts, etc. By default, when the data is extracted in excel for any report- it gives a Pivot table by default.

107. What is OData feed in Power BI?

OData stands for open data protocol. It is usually used to pull data from websites, SharePoint. It helps in getting data from URLs. It basically helps to extract data from a URL without getting into the details of URL-specific parameters like request, response, HTTP methods, etc. It takes care of these things in the back end, leaving us to focus on pulling the data and performing transformations, and creating useful analysis.

108. What is drill down in Power BI?

Drill down, as the name suggests, is going downwards. And in Power BI, it is going down a hierarchical category or dimension. Basically, when we have data attributes like a date where we have a default hierarchy of year->quarter->month->day, then when using the same field in a visual, in Power BI, it by default takes the hierarchy form of the field. This gives the option of a drill-down feature denoted by up and down arrows. Also, we can right-click on the element in visual and get the option of drill-down.

109. What is the maximum number of tables you can join in Tableau?

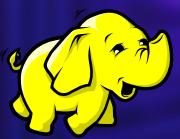
It is not possible to join more than 32 tables together in Tableau. It means you can join a maximum of 32 tables in Tableau.



CLOUDYML

GET PLACED AS A DATA ANALYST DATA SCIENTIST DATA ENGINEER

Become Master Of All The Data Tools



OQOOP

APACHE
Spark

VISIT - WWW.CLOUDYML.COM

**INDIA'S #1 MOST AFFORDABLE INDUSTRY
ORIENTED COURSE PROVIDERS**



**SUBSCRIBE TO
OUR TELEGRAM
CHANNEL TO GET
COMPLETE QNAs PDF**

and more such valuable contents



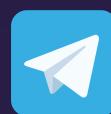
<https://t.me/cloudymlofficial>



107K+



61K+



62K+