**Cumulative QC Review**

**Most important Practical things to be able to implement:**

1. **Excel**
   1. Pivot Tables
   2. Lookup Tables
   3. Chart Creation
2. **BigQuery**
   1. Load data
   2. Export data
   3. Access control
3. **Power BI**
   1. Create calculated measures
   2. Chart creation
   3. Connecting to data
4. **Tableau**
   1. Create calculated measures
   2. Generate reports
   3. Custom filters
   4. Chart Creation

**1.What is the best way of representing this data? (Show a table)**

1. HINT: should you use a bar chart, line chart, pie chart, histogram, other chart, aggregate stats, etc?

**2.When would you choose to use (one visualization tool) instead of (another one)?**

- Tabular format is best used when exact quantities of numbers must be known. Numbers are presented in rows and columns, and may contain summary information, as in PivotTables. This format is not conducive to finding trends and comparing sets of data because it is hard to analyze sets of numbers and the presentation becomes unwieldy with larger datasets.

- Line charts are best used when trying to visualize continuous data over time. Line charts are set against a common scale and are ideal for showing trends in data. You might also add a trend line or a goal line to illustrate performance in a certain period against a set benchmark.

- Bar charts are best used when showing comparisons between categories. Typically, the bars are proportional to the values they represent and can be plotted either horizontally or vertically. One axis of the chart shows the specific categories being compared, and the other axis represents discrete values. Bar charts are ideal when you’re working with limited space.

- Pie charts are best used to compare parts to the whole. Pie charts make it easy for an audience to understand the relative importance of values, but when there are more than five sections, it can become difficult to compare the results. The difference between the sections can become too narrow to effectively interpret. Alternate visual styles include the exploded pie wedge chart, for emphasizing important data, and the donut pie chart, to support information by inserting a design element in the center of the pie.

- Area Charts are best used for showing cumulated totals over time via numbers or percentages. These are basically line charts that are filled in to provide a deeper view of multiple series of data within the chart.

- Bubble Charts are used to show three dimensions of data—comparing entities in terms of their relative values, positions, and sizes. Bubble charts are similar to scatter plots, where the data points are replaced with bubbles.

- Funnel Charts are ideal for showing stages in a particular process (e.g., sales process) or identifying potential problem areas within an organization’s process.

**3.When and why would you as a data analyst need to perform data cleaning?**

Data cleaning is probably the most important part of the data analytics process. Good data hygiene isn't about data analytics, though; it’s good practice to maintain and regularly update your data anyway. Clean data is a core tenet of data analytics and the field of data science more generally.

Clean data is hugely important for data analytics: Using dirty data will lead to flawed insights.

Data cleaning is time-consuming: With great importance comes great time investment. Data analysts spend anywhere from 60-80% of their time cleaning data.

Data cleaning is a complex process: Data cleaning means removing unwanted observations, outliers, fixing structural errors, standardizing, dealing with missing information, and validating your results.

**4.What could be some potential problems you could run into when analyzing a dataset?**

1. ANS: redundant, invalid or corrupt, incomplete, or inaccurate data

**5.Why is it important to identify outliers? Can you identify them in this data? (Show a table or graph)**

**6.Why should you set data alerts?**

Data Alerts can only be set on tiles pinned from report visuals, and only on gauges, KPIs, and cards. Alerts can be set on visuals created from streaming datasets that you pin from a report to a dashboard. Alerts can't be set on streaming tiles created directly on the dashboard by using Add tile-> Custom streaming data.

Only you can see the alerts you set, even if you share your dashboard. Similarly, the dashboard owner can't see alerts you set on your view of their dashboard. Data alerts are fully synchronized across platforms. You can set and view data alerts in the Power BI mobile apps and in the Power BI service. They aren't available for Power BI Desktop. You can automate and integrate alerts with Power Automate.

**7.How does a slicer work in Excel?**

Excel uses slicers to easily filter Pivot Tables. Even more, you can link numerous slicers to numerous pivot tables.

* Choose any cell in the pivot table by clicking it.
* Select Analyze, then Filter, then Insert Slicer.

**8.How do you create a pivot table?Explain it?**

* Make the data ready
* Put your information in the excel sheet.
* To construct a pivot table, highlight your cells.
* In "Row Labels" or "Column Labels," drag and drop a field.
* Drag a field into the "Values" section.
* Adjust your calculations.

While creating a pivot table, keep these in mind:

* Each of the columns' first-row headings should be distinctive.
* There should only be one type of data in each column.
* Rows can only include information for a single recording.
* no empty rows
* Columns shouldn't be left entirely empty.

**9.How can you restrict someone from copying a cell from your worksheet?**

First, choose the data you want to protect.

Hit Ctrl + Shift + F. The Format Cells tab appears. Go to the Protection tab. Check Locked and click OK.

Next, go to the Review tab and select Protect Sheet. Enter the password to protect the sheet.

**10.How to Convert values to phone numbers in a specific format?**

Turning a column of numeric values into telephone numbers may sound like a tricky task, but only until you remember that the Excel TEXT function allows using dashes and parentheses in format codes.

So, to display a number in A2 in a traditional US local 7-digit phone format like 123-456, use this formula:

=TEXT(A2, "###-####")

If some of the original values may contain a domestic prefix (i.e. there can be both 7-digit or 10-digit numbers), include the following conditional format code to display 10-digit numbers in the (123) 456-789 format:

=TEXT(A2,"[<=9999999]###-####;(###) ###-####")

**11.How to use LEFT,RIGHT & MID  function in Excel?**

* The LEFT function in Excel returns the specified number of characters (substring) from the start of a string.The syntax of the LEFT function is as follows:

LEFT(text, [num\_chars])

* The RIGHT function in Excel returns the specified number of characters from the end of a text string.

The syntax of the RIGHT function is as follows:

RIGHT(text, [num\_chars])

* The MID function in Excel is designed to pull a substring from the middle of the original text string. Technically speaking, the MID function returns the specified number of characters starting at the position you specify.

The Excel MID function has the following arguments:

MID(text, start\_num, num\_chars)

**12.How do you create Pivot Tables?**

In order to create a Pivot table, you will first need to prepare the data in a tabular format. Keep the following points in mind while preparing the data:

* Arrange the data into rows and columns
* The first row should contain unique heading for each of the columns
* The columns should have only one type of data
* Rows must have data for a single recording only
* No blank rows
* Columns should not be completely blank
* The data for creating Pivot table should be separate from other data present in the sheet
* select the table and click on the Insert tab. then select Pivot Table command
* Specify where you intend to create the table and then click on OK. Once this is done you will see that an empty pivot table has been created. Also, PivotTables Fields pane will open that will help you configure the Pivot table.

**13.Can you create Pivot tables using multiple tables?**

Yes, you can create Pivot tables using more than one base table. To do this, follow the given steps:

* Press Alt+D and then press P to open up the PivotTable Wizard
* Then select Multiple consolidation ranges option and click on Next
* Select I will create the page fields option and click on Next
* In the next window, you will need to add all the required ranges
* Once that is done, click on Next
* Specify the region where you want to create the table and then click on Finish
* the pivot table has been created by merging both the tables

**14.Explain how to calculate compound interest in Excel?**

To calculate compound interest in Excel, you can use the FV function. FV returns the future value of an investment based on the periodic, constant interest rate and payments.

SYNTAX:

FV(rate, nper, pmt, pv, type)

To find the rate, the number of periods are used to divide the annual rate (annual rate/ periods). nper is obtained by multiplying the no. of years (term) with the periods (term \* periods). Periodic payment (pmt) can be any value (including zero).

**15.How does the VLOOKUP function work?**

The VLOOKUP function, in Excel, a lookup value and begins to look for the same in the leftmost column. When it finds the first occurrence of the given lookup value, VLOOKUP starts to move right i.e in the row where the value was found. It goes on until the column number specified by the user and returns the desired value. This function is used to match exact and approximate lookup values. However, the default match is an approximate match.

Syntax:

VLOOKUP(lookup\_value, table\_array, col\_index\_num, [range\_lookup])

here,

lookup\_value gives the value to be looked out for

table\_index is the range from where the data is to be taken

col\_index\_num specifies the column from which you want to fetch the value

range\_lookup is a logical value i.e TRUE or FALSE (TRUE will find the closest match; FALSE checks for exact match)

**16.How would you highlight cells with negative,duplicate & error values in it?**

We can do this using conditional formatting.

Negative values:

Select the cells in which you want to highlight the cells with negative value.

Go to the Home tab and click on Conditional Formatting option.

Go to Highlight Cell Rules and click on ‘Less Than’ option.

In the ‘Less Than’ dialogue box, specify the value as 0 and the formatting.

Duplicate values:

Select the data in which you want to highlight duplicate cells.

Go to the Home tab and click on Conditional Formatting option.

Go to Highlight Cell Rules and click on ‘Duplicate Values’ option.

Error values:

Select the data in which you want to highlight the cells with errors.

Go to the Home tab and click on Conditional Formatting option.

Click on ‘New Rule’.

In the New Formatting Rule dialog box, select ‘Use a formula to determine which cells to format’ option.

In the formula field, enter =ISERROR(A1), where A1 is the active cell in the selection.

Click on the Format button and specify the color in which you want to highlight the cells.

Click OK.

**17.What are some best practices when creating a dashboard in Excel?**

The  best practices when creating dashboard in Excel:

Convert Tabular Data into Excel Tables: Creating dashboards using an Excel Table as the data source is way easier. Unless you have extremely strong reasons against it, always convert back-end data into an Excel Table.

Numbering your Charts/Section: This helps when you’re presenting the dashboard and referring to different charts/tables. It’s easier to ask people to focus on chart numbered 2, instead of saying the line chart or bar chart on the top left.

Restrict Movement in the dashboard area: While an Excel worksheet is huge, it’s better to remove all the rows/columns except the ones that have your dashboard.

Freeze Important rows/column: In case you want some rows/columns to always be visible, you can freeze these.

Make Shapes/Charts Stick: A user may end up changing row/column width. You don’t want the shapes and charts to get misaligned and move from their original place. So, it’s better to make them stick to their position.

Provide a User Guide: It’s a good idea to include a separate sheet that has the details on how to use the dashboard. This becomes more useful when you have interactive controls in the dashboard.

Save Space with Combination Charts: Since there is limited space in a dashboard (as you need to fit it in one screen), using combination charts can save you space.

Use Symbols & Conditional Formatting: You can make dashboard more visual and easier to read by using symbols and conditional formatting. For example, it’s easier to see the lowest value in a table when it’s highlighted in red, instead of going through all the values one by one.

Use colors wisely to show contrast: When you want to highlight a data point in a table or a bar in the chart, it is better to make it stand out by highlighting it in eye-catching color.

**18.What is a macro in Excel?**

A macro is a set of instructions written in the VBA language that Excel can understand and execute. A macro can be as simple as a single line or can be thousands of line long.

**19.What are the different ways to run a macro in Excel?**

The following ways to run a macro in Excel:

* Assign the Macro to a Shape
* Assign the Macro to a Button
* Run a Macro from the Ribbon (Developer Tab)
* Run a Macro from the VB Editor
* Run a Macro using a Keyboard Shortcut
* Call a Macro from another Macro

**20. How do you distribute your table data into fact and dimension tables?**

Ans: Distributing table data into fact tables and dimensions is a process commonly known as dimensional modeling. Here, we can take several approaches to find out which data can be pushed into facts and which into dimensions:

* Understand the data and thoroughly understand the business process and subject area in which you are modeling.
* **Identify Fact Tables:** Fact tables include key performance indicators (KPIs) which are mostly in the form of numerical or measurable data that captures the events or transactions in the business process.
* **Determine Dimensions:** Identify the dimensions that are associated with the events or transactions captured in the fact tables. It provides the context and descriptive attributes for analyzing the data.
* **Establish Relationships:** Define the relationships between the fact tables and dimension tables using foreign key constraints. Each dimension table should have a primary key column and a fact table should have foreign keys to establish relationships between them.

**21.  Which are the properties a Data Warehouse should follow?**

SINTA

Subject-oriented

 Data is based around main objective of company using data

Integrated

 Data is integrated, summarized, compact

 Non-volatile

 Data will persist and won’t change. Write once, read many times.

 Time-variant

Data should have a time stamp to keep track of period of data retrieval. Data like ""Age"" will become stale.

Archive

Data is archived to remember location of data in DWH and high volumes.

**22. How data can be loaded into BigQuery?**

BigQuery Data Transfer Service is the tool that should be utilized for the most successful loading of data into BigQuery. we will be able to swiftly and efficiently import data into BigQuery from various sources, including other services offered by the Google Cloud Platform.BigQuery supports multiple input formats when receiving data.

BigQuery's web-based user interface is another option for transferring data files. In addition to importing data from a local file or a Google Cloud Storage bucket, the BigQuery command-line tool can do the same for a Google Cloud Datastore bucket. BigQuery's application programming interface (API) then lets you import records from numerous sources.

**23. How to convert a stringified array to an array?**

We can convert a stringified array to an array from a BigQuery Table:

COMMAND:

#standardSQL

WITH k AS (

  SELECT 1 AS id, '["a", "b", "c"]' AS x UNION ALL

  SELECT 2, '["x", "y"]'

)

SELECT

  id,

  ARRAY(SELECT \* FROM UNNEST(SPLIT(SUBSTR(x, 2 , LENGTH(x) - 2)))) AS x

FROM k

**24. How can I determine the BigQuery storage size for a single table?**

We can determine the BigQuery storage size for a single table:

COMMAND:

select

  sum(size\_bytes)/pow(10,9) as size

from

  <your\_dataset>.\_\_TABLES\_\_

where

  table\_id = '<your\_table>'

**25. Write down a command to fetch each user between two dates?**

We may fetch each user between two dates from a BigQuery Table:

COMMAND:

SELECT

  timestamp\_trunc(timestamp, DAY) as Day,

  user\_id,

  count(1) as Number

FROM `table`

WHERE timestamp >= '2023-06-28 00:00:00 UTC'

AND timestamp <= '2023-06-27 23:59:59 UTC'

GROUP BY 1, 2

ORDER BY Day

**26. Your data team is building a new real-time data warehouse for a client. The client wants to use Google Big Query for performing streaming inserts. You get a unique ID and an event timestamp whenever data gets inserted in the row but it is not guaranteed that data will only be sent in once. Which clauses and functions you will use to write a query which ensures that duplicates are not included while interactively querying data?**

To ensure that duplicates are not included, use the ROW\_NUMBER window function with PARTITION BY based on unique ID WHERE row equals to1.

**27. An analytics company handles data processing for different clients. Clients use their own suite of analytics tools. Some clients have allowed direct query access via Google Big Query. You want to ensure that clients cannot see each other’s data. What steps can you perform inside Big Query to ensure the data security of clients?**

To ensure that clients could not see each other’s data, the following steps could be taken:

a. For each client, load data into a different dataset.

b. Restrict a client’s dataset such that only approved users can access their dataset

c. For further security, use the relevant identity and access management (IAM) roles for each client’s users.

**28. A client provides your company with a daily dump of data that flows into Google Cloud Storage as CSV files. How would you build a pipeline that will analyze the data stored in Google Cloud Storage in the Google Big Query when the data may contain rows which are formatted incorrectly or corrupted?**

To build a pipeline for the above scenario follow the below steps:

a. Import the data from Google Cloud Storage to the Big Query by running Google Cloud Dataflow.

b. Push the corrupted rows to another dead-letter table for analysis.

**29. You work as an analyst in an e-commerce company. You use Google Big Query to correlate the customer data with the average prices of the 40 most common products sold, including laptops, mobile phones, television, etc. After every 25 minutes, the average prices of these goods are updated. What steps you should follow to ensure that this average price data stays up to date so that you can easily combine it with other data in Big Query as cheaply as possible?**

Follow the below steps to ensure that this average price data stays up to date so that you can easily combine it with other data in Big Query as cheaply as possible:

a. Create a regional Google Cloud Storage Bucket to store and update the average price data

b. Then, use the Cloud Storage Bucket as a federated data source in Big Query.

**30. How can we delete the duplicate rows in the BigQuery Table?**

Using the following command, we may delete duplicate rows from a BigQuery Table:

COMMAND:

SELECT \*

FROM (

  SELECT

      \*,

      ROW\_NUMBER()

          OVER (partition by Fixed Accident Index)

          row\_number

  FROM Accidents.Cleaned FilledCombined

)

WHERE row\_number = 5

**31. How to migrate servers and virtual machines hosted on-premises or in another cloud to the Compute Engine of the Google Cloud Platform?**

The virtual machines (VMs) can be moved from on-premises data centres, Azure, and Amazon Web Services (AWS) to Google's Compute Engine with the use of the cloud software known as Google Cloud Migrate for Compute Engine. This software does not come with any additional charges or fees attached to it.

**32. What is "Virtual Private Cloud" (VPC) when referring to Google Cloud Platform?**

Through the use of a Virtual Private Cloud, your Google Cloud Platform (GCP) virtual machine (VM) instances, Google Kubernetes Engine (GKE) clusters, and other resources will be able to connect with one another (VPC). The Virtual Private Cloud gives users a great deal of wiggle room in terms of regulating regional and global workload connectivity. Without having to rely on the public internet, virtual private networks (VPCs) make it possible for multiple regions to communicate with one another.

**33. Explain the most significant advantages gained by utilizing Google's cloud services?**

The following is a list of the primary characteristics of GCP:

Using Google Cloud Platform makes it simple to fine-tune the CPU, RAM, and storage capacities of your virtual machine. The virtual machine (VM) rightsizing advice tool clearly demonstrates in a short amount of time whether or not the machines in your environment are utilizing the appropriate quantity of hardware.

You will have access to the Google cloud shell when you utilize GCP. This shell comes pre-loaded with a broad number of helpful tools and makes it possible for you to manage your infrastructure with just a few keystrokes. Docker, Gradle, Make, npm, nvm, and pip, along with a great deal more software, is pre-installed and ready to use.

You'll have the ability to swiftly prototype new kinds of machines with Google Cloud Platform thanks to its fully programmable CPU, RAM, and storage.

The preemptible virtual machines that come with this technology can slash expenses by as much as 70 per cent for fault-tolerant and batch processing.

The Cloud SQL functionality of GCP does a check on the database's available storage once every 30 seconds and adds additional if it's required.

It is possible to alter the size of a persistent disc in real-time and without disrupting service in any way, either by decreasing or increasing the amount of space it occupies.

**34. What does it mean by "Object Versioning"?**

Recovery of unintentionally overwritten or destroyed data is made feasible through object versioning. To secure the safety of objects when they are rewritten or removed, versioning them incurs additional storage expenses. When object versioning is set on in a GCP bucket, anytime an object is removed or replaced, a unique version of the object is created. Generation and meta-generation attributes are used to determine the specific iteration of an object. A generation recognises the production of new content, whereas a metageneration recognises the production of new metadata.

**35. Differentiate between elasticity and scalability.**

One of the most important aspects of cloud computing is its scalability, which enables it to boost the number of resources it can provide in reaction to an increase in demand for those resources. When there is an increase in the demand for traffic, the design can be scaled up to provide the additional resources that are required. Elasticity, on the other hand, is a property that enables the instantaneous assembly and disassembly of enormous amounts of available resources. It is contingent on the quantity and duration of the resources that are accessible.

**36.Who are the system integrators when it comes to cloud computing?**

Because there are so many moving pieces, understanding clouds can be difficult at times. The system integrator is the overarching strategy that enables different cloud-related tasks, such as cloud design and the assembly of necessary elements for a public, private, or hybrid cloud infrastructure. In the cloud, the system integrator is the strategy that enables these tasks.

**37.Differentite between the Google Compute Engine and the Google App Engine?**

The Google App Engine and the Google Compute Engine each have a great deal to contribute to one another. Google Application Engine is a Platform as a Service (PaaS), whereas Google Compute Engine (GCE) offers computing resources. GAE is frequently used to operate a wide variety of applications, the most typical of which are mobile backends, online apps, and bespoke business software.

Compute Engine is an excellent option to go with if we want to have a greater influence over the underlying infrastructure. For example, we could leverage Google Compute Engine to build our very own storage system or to supply specialized business logic.

**38. You have a large dataset stored in BigQuery containing customer information, including their names, email addresses, and purchase history. How can you retrieve a list of customers who have made more than 10 purchases?**

Ans:  You can execute a SQL query like the following:

SELECT customer\_name, email

FROM dataset.table

GROUP BY customer\_name, email

HAVING COUNT(\*) > 10;

**39. What is a factless fact table and what is its importance?**

Ans:  Factless facts are those fact tables that have no measures associated with the transaction.  Factless facts are a simple collection of dimensional keys which define the transactions or describe conditions for the time period of the fact.  These tables are occasionally necessary for capturing important dimensional relationships which are critical to meeting the defined business reporting requirements.

**40. What is Slowly changing dimensions and what are the types of it?**

Ans: SCD is defined as slowly changing dimensions, and it applies to the cases where record changes over time.

There are three types of SCD, and they are as follows:

SCD 1 – The new record replaces the original record.

SCD 2 – A new record is added to the existing customer dimension table.

SCD 3 – Original data is modified to include new data.

**41. Is there any difference between ER Modeling and Dimensional Modeling?**

Ans: ER modeling will have logical and physical models, but Dimensional modeling will have only Physical models.

ER Modeling is used for normalizing the OLTP database design whereas Dimensional Modeling is used for de-normalizing the Relational-OLAP and Multidimensional-OLAP design.

**42. How can you resolve common BigQuery SQL Errors?**

Ans: Use the Query Validator to check the syntax of your query. When you attempt to run a query that has errors, it fails and logs the error in the Job details. When the query is error-free, the query validator displays a green checkmark. Click Run to execute the query and display the results after the green checkmark appears.

**43. What are the best practices for building data warehousing?**

Ans: There could be several points that need to consider while building data warehousing are:

* **Clearly define objectives:** Clearly define the objectives and goals of your data warehousing initiative. Understand the business requirements and identify key metrics and reports that need to be generated from the data warehouse.
* **Plan for scalability:** Design and build the data warehouse with scalability in mind. Anticipate future growth and ensure that the architecture and infrastructure can handle increasing data volumes and user loads over time.
* **Data modeling**: Develop a robust and flexible data model that supports the reporting and analytical requirements of the organization. Utilize techniques such as dimensional modeling to optimize query performance and facilitate ad-hoc analysis.
* **ETL processes:** Implement efficient Extract, Transform, and Load (ETL) processes to move and transform data from source systems into the data warehouse. Focus on data integration, data cleansing, and data transformation to ensure data consistency and reliability.
* **Choose between On-Premises and Cloud-Based Data Warehouse:** An on-premises data warehouse means the customer deploys one of the available data warehouse systems – either open-source or paid systems on his/her own infrastructure. In a cloud-based data warehouse service, the customer does not need to worry about deploying and maintaining a data warehouse at all. The data warehouse is built and maintained by the provider and all the functionalities required to operate the data warehouse are provided as web APIs.
* **Security and access control:** Implement strong security measures to protect sensitive data stored in the data warehouse. Define access controls and user permissions to ensure that only authorized users can access and manipulate the data.

**39. How can we define the relationships between two tables in a data model in the Power BI Desktop?**

There are two approaches:

Manual: by using primary and foreign keys

Automatic: the relationships are identified automatically if the autodetect feature is switched on

To define the relationships between two tables, there shouldn't be any null values or duplicate rows in the data. Also, it's possible to have multiple relationships between tables (represented by dotted lines), but only one of them can be active (represented by a continuous line).

**40. We are provided a situation where a bank needs to join multiple complaints files from different sources such as customer complaints, home loan complaints, credit card complaints, and student loan complaints.How can we get multiple files to one location?**

The folder option as a data source is ideal for multiple files that must be pooled from different sources and compiled at a singular place. The folder option resides under the get data option.

**41.  How can we take a query from one solution and use it in a different solution altogether to save on time and effort?**

There are three discrete ways to implement this:

Copy M-code

Render all the transformations as a template

Data Flows

**42. How to download the Power BI datasets or reports without publishing them to the web?**

We can do it by creating a new report on the Power BI desktop and use the connect to "PBI dataset" option to access the data in the table.

**43. How would you handle data from multiple sources with different granularity levels?**

Handling data from multiple sources with different granularity levels involves the following steps:

Data Integration:

Use Power Query to connect and extract data from multiple sources.

Apply data transformations, clean, and prepare the data for analysis.

Data Modeling:

Create relationships between tables to link data from different sources.

Use common dimensions to establish relationships between tables with different granularity levels.

Granularity Adjustment:

Create new calculated columns or measures to adjust granularity levels.

Use DAX (Data Analysis Expressions) to aggregate or disaggregate data as required.

Data Validation:

Compare aggregated data with source data to ensure accuracy.

Address any discrepancies to maintain data integrity.

**44. Describe the process of creating a dynamic KPI visualization in Power BI?**

The process of creating a dynamic KPI visualization in Power BI involves the following steps:

Prepare the Data:

Import or connect to the data source.

Clean, transform, and model the data as required.

Create Measures:

Use DAX to create measures for KPI calculations, such as the actual value, target value, and percentage difference.

Create KPI Visuals:

Select the appropriate visuals for KPI display, such as cards, gauges, or custom visuals like KPI indicators.

Add the created measures to the respective fields in the visual.

Make KPIs Dynamic:

Add slicers or filters to allow users to interactively change the displayed data based on specific criteria, such as date range, product category, or region.

**45. Explain the steps to create a rolling average or moving average in Power BI?**

Creating a rolling average or moving average in Power BI involves the following steps:

Prepare the Data:

Import or connect to the data source.

Ensure that the data is sorted by date or a relevant time period.

Create a Date Table (if not already present):

Use DAX to create a date table with a continuous date range.

Establish a relationship between the date table and the data source.

Create a Rolling Average Measure:

Use DAX to create a measure that calculates the rolling or moving average.

Incorporate window functions like AVERAGEX, DATESINPERIOD, or DATEADD to define the rolling period.

Add the Measure to a Visual:

Select an appropriate visual, such as a line chart or a bar chart.

Add the rolling average measure along with other relevant fields to the visual.

**46. How do you create a measure that calculates year-to-date sales using DAX?**

To create a measure that calculates year-to-date (YTD) sales using DAX, follow these steps:

Identify the relevant columns:

Locate the columns containing the date and sales data in your data model.

Create the YTD measure:

Use the CALCULATE and TOTALYTD functions to calculate the YTD sales.

For example, if you have a column named 'SalesAmount' and a date column named 'OrderDate', the DAX expression would be:

YTD Sales = CALCULATE(SUM('Sales'[SalesAmount]), TOTALYTD(SUM('Sales'[SalesAmount]), 'Sales'[OrderDate]))

**47. Explain how to calculate a dynamic ranking in Power BI using DAX?**

To calculate a dynamic ranking in Power BI using DAX, follow these steps:

Identify the relevant columns:

Locate the columns containing the data for which the ranking will be calculated.

Create a ranking measure:

Use the RANKX function to create a measure that calculates the ranking based on the desired criteria.

For example, if you want to rank products by sales, the DAX expression would be:

Product Rank = RANKX(ALL('Products'), CALCULATE(SUM('Sales'[SalesAmount])))

**48. Describe the process of creating a cumulative total using DAX in Power BI?**

To create a cumulative total using DAX in Power BI, follow these steps:

Identify the relevant columns:

Locate the columns containing the data for which the cumulative total will be calculated.

Create a cumulative total measure:

Use the CALCULATE and FILTER functions to create a measure that calculates the cumulative total based on the desired criteria.

For example, if you want to calculate the cumulative sales by date, the DAX expression would be:

Cumulative Sales = CALCULATE(SUM('Sales'[SalesAmount]), FILTER(ALLSELECTED('Sales'[OrderDate]), 'Sales'[OrderDate] <= MAX('Sales'[OrderDate])))

**49. How can you use DAX to handle data discrepancies and irregularities?**

DAX can be used to handle data discrepancies and irregularities in various ways, such as:

Data Cleaning:

Use DAX functions like TRIM, SUBSTITUTE, or REPLACE to remove unwanted characters, spaces, or correct typos.

Data Transformation:

Use DAX functions like UPPER, LOWER, or PROPER to standardize text case.

Use DAX functions like DATE, EOMONTH, or DATEADD to manipulate date values.

Handling Missing or Incomplete Data:

Use DAX functions like IF, SWITCH, or COALESCE to replace missing or incomplete data with default values or alternative data.

Error Handling:

Use DAX functions like ISERROR, IFERROR, or DIVIDE to handle errors in calculations or data retrieval.

Data Validation:

Use DAX functions like CONTAINS, LOOKUPVALUE, or RELATED to validate data against other tables or data sources.

**50. Assume you are working with a Finance dataset in PowerBI, which has one column as "currency\_generated" where the values of the columns are like $200, 150$, 100R, and R300. Now you wanted to create a new column in your table that will take only the number values from the "currency\_generated" column. How would you do that?**

Ans: Steps to solve the problem:

* Right-click on the table, it will open a list of options. Choose the "Edit query" option from there.
* It will open a Power Query Editor. In the query editor home tab, click on the "Split Column" option. It will open a dropdown list of multiple options to select.
* From there, select the option "By digit to non-digit". It will create two columns from the original column, one with only numbers and one with "currency symbols".
* Remove the extra column having records of "currency symbols" by right-clicking on the column and choosing the "remove" option from there.

**51. How do you remove the blank rows from your table in PowerBI? If the table looks like this**

|  |  |  |
| --- | --- | --- |
| **ID** | **Name** | **Subject** |
| 100 | Rohan | Maths |
|  |  |  |
| 102 | Rohit | Science |
|  |  |  |
| 103 | Shan | Biology |
| 105 | Mohit | English |

Ans: Steps to remove null from any table:

* Right-click on the table, it will open a list of options. Choose the "Edit query" option from there.
* It will open a Power Query Editor. In the query editor home tab, click on the "Remove Rows" option under “Reduce Rows”. It will open a dropdown list of multiple options to select.
* From there, select the option "Remove Blank Rows". It will remove all the blank rows from your table.

**52. Imagine you have a dataset that has 4 columns in it named Id, Name, Mark1, Mark2. How do you add a new column to your model/table in PowerBI that stores the addition values of two columns Marks1 and Marks2?**

Ans: Steps to add a new column to your model/table are as follows:

* Right-click on the table, it will open a list of options. Choose the **"Edit query"** option from there.
* It will open a Power Query Editor. In the query editor **“Add Column”** tab, select the **“Custom Column”** option. It will open a Custom Column window.
* In the new window, you will see the **“Available columns”** list that has all the columns of your table. A **“Custom column formula”** box, where you can provide the formula to add the calculated values to your new column. And a **“New column name”** field, where you can provide the new name for your column.
* Select the Marks1 and Marks2 columns from the **“Available columns”** list, and then select “**Insert”** below the list to add them to the custom column formula.
* In between two selected columns, that is Mark1 and Marks2, add a **“+”** operator and then click on the **“ok”** button in the last. It will create a new column that will have the sum values of two selected columns.

**53. Suppose you are working with a Sales table in PowerBI having columns such as Id, Name, total sales, and DateTime. In DateTime column, values are presented as “2012-02-25::04:42:18”, “2012-01-21::05:22:10”. Now you want to create two separate columns “Date” and “Time”, that will take the values of date and time respectively from the DateTime column. How will you do that in PowerBI? Note: Date and Time are separated by double colon (::) in DateTime column.**

Ans:  Steps to solve the above problem:

* Right-click on the table, it will open a list of options. Choose the **"Edit query"** option from there.
* It will open a Power Query Editor. In the query editor, right-click on the **“DataTime”** column and select **“Split Column”** option from there.
* After clicking on **“Split Column”** option it will ask you to select between two options, **“By Delimiter”** or **“By Number of Characters”**.  Select the **“By Delimiter”** option.
* It will open a new window, where you need to select the delimiter that you want to split the column with. By default, it is Space.
* From **“Select or enter delimiter”** option, choose the last **“Custom”** option. It will ask you to define the delimiter that will be used to split the column.
* There you need to specify **“::”** and click “OK” button. It will create the two columns with dates and times separately.

**54. What might be the reason for getting “Refresh failed due to gateway configuration issues” while doing the refresh over your dataset in PowerBI?**

Ans: The error message "Refresh failed due to gateway configuration issues" in Power BI typically indicates a problem with the gateway configuration. The Power BI gateway is a tool used to connect on-premises data sources to Power BI, allowing you to refresh data from those sources in the Power BI service.

* A possible reason for this could be the **gateway** might not be installed or configured correctly. Ensure that the gateway is installed on a suitable machine and properly configured to connect to your data sources. Verify that the gateway is running and connected to the Power BI service.
* The other reason could be the gateway requires appropriate connectivity to access your data sources. Verify that the gateway machine has **network connectivity** to the data source and that there are no firewall or security restrictions blocking the connection.
* **Incorrect credentials or authentication** issues can also cause the refresh to fail. Double-check the credentials used to connect to the data source.

**55. What are the steps of doing refresh on datasets in PowerBI?**

Ans: To refresh datasets in Power BI, you typically follow these steps:

* **Configure Refresh Settings**: Once the data is loaded into Power BI Desktop, go to the "Home" tab and click on "Manage Refresh" in the "External Data" section. In the "Scheduled Refresh" settings, specify how often you want the dataset to be refreshed, such as daily, weekly, or manually.
* **Configure Credentials**: If the data sources require authentication, you need to provide the appropriate credentials. Depending on the data source, you can choose different authentication methods such as Windows, database, OAuth, or custom.
* **Publish the Dataset**: Save the dataset in Power BI Desktop and publish it to the Power BI service. This step makes the dataset available for consumption by users and allows for scheduled refreshes in the cloud.
* **Enable Scheduled Refresh in Power BI Service**: After publishing the dataset, go to the Power BI service (app.powerbi.com). Open the dataset and click on the "Settings" (gear) icon in the top-right corner. Under the "Dataset settings" tab, ensure that the scheduled refresh option is enabled and configure the desired refresh frequency.
* **Validate and Monitor Refresh:** Test the refresh process by triggering a manual refresh initially. Monitor the refresh history and verify that the data is being updated correctly.

**56. Are there any problems with PowerBI Dashboards? If yes, tell me what the problems are and how you solve them.**

Ans: **Data Refresh Issues:** Sometimes, Power BI Dashboards may fail to refresh or display outdated data. This can happen due to problems with data source connections, credentials, or gateway configuration. To resolve data refresh issues, you can try the following:

* Verify the data source connections and credentials used for data refresh. Ensure they are correct and up to date.
* Check the scheduled refresh settings in the PowerBI service to ensure they align with your requirements.

**Visual Display Problems**: Occasionally, visuals in a Power BI Dashboard may not display as expected. This can be due to formatting issues, incorrect visual settings, or incompatible visuals with the selected data. Here are a few steps to address visual display problems:

* Check the visual formatting options and ensure they are configured correctly.
* Validate the field mappings and data types used in the visuals.
* Verify that the required data fields are added to the visuals and that they have valid data.

**57. What are the different types of filters can we see in PowerBI?**

Ans: Here are the different types of filters available in Power BI:

* **Visual-Level Filters:** These filters are applied to individual visualizations or charts within a report. Visual-level filters allow you to filter data within a specific visual without affecting other visuals on the report page.
* **Page-Level Filters:** Page-level filters are applied to an entire report page. When you apply a page-level filter, it affects all the visuals on that specific page, allowing you to focus on a subset of data for analysis.
* **Report-Level Filters:** Report-level filters apply to all visuals and pages within a report. When you apply a report-level filter, it acts as a global filter for the entire report, ensuring that the specified filter conditions are applied consistently across all visuals.

**58. You have a dataset with product sales information, including the product name, category, and sales quantity. You want to create a visualization that shows the top-selling products by category. How can you accomplish this in Tableau?**

Ans: To visualize the top-selling products by category in Tableau, you can proceed as follows:

* Import your dataset into Tableau.
* Drag the "Category" field to the Rows shelf and the "Sales Quantity" field to the Columns shelf.
* Change the default chart type to a bar chart or a horizontal bar chart.
* To show only the top-selling products within each category, you can either right-click on the "Sales Quantity" axis and choose "Add Reference Line" > "Top" or use the "Show Me" pane to select the "Top N" option and configure it accordingly.

**59. Differentiate between Power BI personal Gateway and Data Management Gateway?**

Power BI Personal Gateway is used for reports that are deployed in Powerbi.com. On the other hand, data management is an app that installs the gateway on source data machines to deploy reports on Sharepoint and schedule to refresh automatically.

**60.What do you know about Row Level Security and how do you implement it?**

Row-level security restricts the data that users view and access based on filters. To configure row-level security, users can define rules and roles within Power BI Desktop and publish them to Power BI Service. Also, the username() function can restrict data in the table to the current user. However, to enable row-level security, a Power BI Pro subscription account is essential, and Excel sheets can be used when converted to the .pbix file format.

These are the ways to implement Row Level Security in Power BI:

* Define roles and rules in Power BI Desktop
* Validate the roles within Power BI Desktop
* Manage security on your model
* Validating the role within the Power BI service

**61. Explain the different types of gateways available in Power BI? Why we use them?**

The Gateway acts as an extension between azure cloud administration and on-premise data sources. There are three major types of Gateways listed as follows:

On-premises data gateway: Allows multiple users to connect to various on-premises data sources. You can use an on-premises data gateway with all supported services with a single gateway installation. For complex scenarios, this gateway is best suited.

On-premises data gateway (personal mode): This data gateway allows one user to connect to sources that can’t be shared with others. This gateway is best suited to scenarios in which you only create reports and don’t need to share any data sources with others.

The virtual network data gateway: This data gateway allows multiple users to connect various data sources that are secured by virtual networks. For complex scenarios in which multiple people access multiple data sources, this data gateway is highly preferred.

**62. What are Slicers in Power BI?**

Slicers are the visual filters, which are present in the report page of Power BI. It helps us to sort and filter information on a packed report

Slicers unlike filters present as a report visual and you can select values on it while analyzing the report data.

**63. How do you compare Target and Actual Values from a Power BI report?**

A gauge chart is used to compare two different measures where the Target and Actuals are inferred.

**64. How can we use summarize function in PowerBI?**

Summarize is a DAX function that gives an aggregated result from a table.

we can use summarize function like :

Summarize(<table>,<grouping column>,[<name>,<expression>])

Table - a DAX expression that returns a table

grouping column - the column name you want to use for grouping

name - the name of the new aggregated column

expression - generates the aggregated column.

**65. How would you use Power BI to analyze sales data?**

Data: we must first connect to a source that contains sales data. This might be a sales program or a customer relationship management (CRM) suite.

 Integration:  Next, create a report in Power BI with sales data. You might create a top performers list, a year-over-year performance comparison, or sales by product listing.

Analyze: Finally, analyze the sales data to identify trends and insights. You might use bar charts, column charts, or pie charts to visualize sales data.

**66. How would you use Power BI to track website analytics?**

Data: First, connect to a data source with website analytics data, like Google Analytics.

Integration: Next, create a Power BI report with website analytics data. You might create a report that shows website visitors over time or by geographic location.

Analyze :  Analyze the website analytics using charts that display interactions over time. Line charts are usually the best choice for this scenario.

**67. How would you use Power BI to monitor social media metrics?**

Data: Connect to a data source with social media data, like Twitter or Facebook, a social media manager, or an aggregation service.

Integration : For social media metrics, you might create a report that shows how many times a particular hashtag was used, or how many likes a Facebook page has.

Analyze : Analyze the social media data in the report to identify trends and insights. Social media data is intricate; you should start by determining what your organization wants to track.

**68. How would you use Power BI to track KPIs?**

Data.:  Connect to a data source that contains the KPIs you want to track. This could be an Excel file, SQL database, or data warehousing solution. KPIs are highly situation-dependent, so choose one as an example, such as conversions.

Integration :  You might create a report that shows how each KPI is performing over time. For conversions, you might want to match conversions, conversion sources, and even customer lifetime value.

Analyze :  What type of reports would you use to visualize KPIs? For conversion, you might just need a line chart or a graph.

**69. How would you use Power BI to forecast future sales?**

Connect to your sales data silo.

Bring sales data in and analyze year-over-year, quarter-over-quarter, and month-over-month.

Forecast data utilizing past performance and projected growth data.

**70. How would you use Power BI to predict customer churn?**

Connect to your customer data well.

Use Power BI to pull data from customers who have left.

Identify current customers who fit the profile of customers that have left.

Use this to identify the customers that are most likely to churn.

**71. How to share Power BI dashboards?**

Power BI reports/dashboards can be shared in multiple ways.

We are using the share option in reports and dashboards - This option enables access to a report or dashboard to individual users.

Using content packs – Publish your report/dashboard along with the data set as a content pack and then share it with either a group or individual or open it for the entire organization.

Publish your dashboards and reports into App Workspace and share the App link to a group or individual or open it for the entire organization. When publishing as App , we can select the reports and datasets that we want to be included in the App.

The report can be embedded into Web or SharePoint, for which we need the embed code. This embed code is added to the website code or the SharePoint code.

**72.Suppose I want to design a view, without using a line or bar chart, to show the region-wise profit and sales. How should I go about doing it? Explain.**

- Generate a map using cities

- Then, drag the profit and sales to Details

- Add the state as a Quick filter

**73.Design a view in a map such that if a user selects any state, the profit and sales in the cities under that state show up?**

If a user wants to show the sales and profit of each and every city under the states in the same worksheet, then they should first have State, City, Sales, and Profit fields in their dataset.

- Double-click on the State field

- Drag City and drop into the Marks card, which is under the State field

- Drag Sales and drop into Size

- Drag Profit and drop into Color

- Click on Size legend and increase the size (75%)

- Right-click on the State field and select Show Quick filter

- Select any state and check whether you got the required view or not

In this, the view size indicates the number of sales and the color indicates the profit values.

**74.How can we combine a database and the flat file data in Tableau Desktop?**

Connect data twice, once for database tables and then for the flat file. The Data->Edit Relationships

Give a Join condition on the common column from DB tables to the flat file.

**75. How is a workbook published and scheduled in Tableau Server?**

First, create a schedule for a particular time and then create an Extract for the data source and publish the workbook on the server.

Before publishing it, there is an option called Scheduling and Authentication. Click on that and select the schedule from the drop-down and then publish. Also, publish data sources and assign the schedule. This schedule will automatically run for the assigned time and the workbook will get refreshed on a regular basis.

**76.What is a TDE file and how it is created?**

Tableau Data Extract (TDE) file is a data extract or snapshot of the data. TDE files can be created by right-clicking on the data source and selecting Extract Data.

-When an extract is created, the symbol of the data source icon changes to two cylinders.

-Extracts result in better performance as compared to live connections.

-Extracts are useful in offline access to data. Extracts can be saved locally and do not require a live connection to the source data.

-The extract is a snapshot of data, it needs to be refreshed as underlying data changes.

-Filters can be applied while creating an extract.

-While creating an extract data can be aggregated by dimensions.

-An extract can be created for all the rows or incrementally.

-Extract refreshes can be scheduled to run.

-Provide Aliases to give a meaningful description of the values. For example, if the Region is SZ, you can alias it to South Zone.

-Hide the columns, dimensions, and measures which are not used in the calculations or the view.

-Create Calculations to achieve the required functionality. Name calculations in a user-friendly manner.

-Apply appropriate filters to get meaning full data.

-Use extract when possible. Extracts are faster than live connections.

-Save the data source as a TDS file. This file can be shared and other developers can use this file for development.

**77.A data table has 100,000 rows of data. Every month around 10,000 new rows get added to this table. What option you will choose to load only the new rows?**

This can be done by creating an extract and choosing an incremental load option.

**78. How do you calculate daily profit measures using LOD in Tableau?**

To calculate daily profit measures using LOD in Tableau, you can create a calculated field that uses an LOD Expression to calculate the profit at the transaction level, and then aggregate the results by day. For example, the calculated field might be: { FIXED [Order Date], [Product ID] : SUM([Sales]) - SUM([Cost]) }.

**How can you schedule a Tableau Workbook to run automatically after it has been published?**

To schedule a Tableau Workbook to run automatically after it has been published, you can use Tableau Server or Tableau Online. First, you need to create a schedule by navigating to the "Schedules" page in the "Settings" section. Then, you can select the frequency and time when you want the workbook to run. Finally, you can select the workbook you want to schedule and specify the parameters you want to use for the workbook.

**Can you provide a method for making a webpage dynamic in Tableau?**

To make a webpage dynamic in Tableau, you can embed a Tableau dashboard in the webpage using an iframe. Then, you can use Tableau's JavaScript API to interact with the dashboard and modify its behavior based on user actions.

**Design a view to show profit and sales data by region.**

To design a view to show profit and sales data by region, you can use Tableau's "Map" visualization type. You can plot the regions on a map and color-code them based on the profit and sales data. You can also use tooltips to display additional information about each region.

**Why are data servers crucial in Tableau?**

Data servers are crucial in Tableau because they allow users to securely store and manage their data, and access it from anywhere. Tableau connects to data servers to access data sources, perform queries, and extract data for analysis. With a data server, users can collaborate on data analysis, share dashboards and reports, and ensure data consistency across multiple users and teams. Additionally, data servers provide features such as security, backup and recovery, and scalability, which are essential for managing enterprise-level data. Overall, data servers are a critical component of Tableau's functionality, enabling users to effectively analyze and make data-driven decisions.

**Describe a scenario where a Heat Map would be useful in Tableau.**

A Heat Map is a visualization tool that represents data values using color gradients. It is useful in scenarios where the data needs to be quickly and easily understood. For example, a Heat Map can be used to represent the intensity of traffic on different routes in a city. The color gradient can represent the different levels of traffic, allowing users to quickly identify the busiest routes.

what are the differences between aggregation and disaggregation in Tableau,?

Aggregation refers to the process of combining multiple values into a single value. For example, summing up sales data by region. Disaggregation, on the other hand, refers to the process of breaking down a single value into multiple values. For example, breaking down sales data by individual products. In Tableau, aggregation and disaggregation can be applied to the same data set to analyze the data from different perspectives.

**Explain the difference between discrete and continuous in Tableau?**

In Tableau, discrete data refers to data that is made up of distinct values or categories. Examples of discrete data include categorical data, such as gender or product type. Continuous data, on the other hand, refers to data that can take on any value within a range. Examples of continuous data include numerical data, such as sales figures or temperature.

**How can you calculate percentages in Tableau?**

To calculate percentages in Tableau, you can drag a measure to the view, right-click on the measure, choose Quick Table Calculation, and then choose Percent of Total, Percent Difference, or another calculation option.

**Why is data aggregation or disaggregation necessary for creating charts in Tableau?**

Data aggregation or disaggregation is necessary for creating charts in Tableau because it determines the level of detail in the visualization. Aggregating the data can help simplify the chart and show trends and patterns at a higher level, while disaggregating the data can provide more detail and granularity.

**Explain the process of using custom SQL in Tableau?**

To use custom SQL in Tableau, you can select "Connect to Data" and choose the appropriate data source. From there, you can select "Use Custom SQL" and write your SQL query. Tableau will then use the results of the query as the data source for your visualization.