----------**Power BI**----------

1. What is BI?
2. What is Power BI and why?
3. What is a Power BI dashboard?
4. What is Power BI report?
5. What are the Difference between Power BI and Tableau?
6. What are the Component of Power BI?
7. What is data cleaning?
8. What is ETL?
9. What is DAX?
10. What are the types of DAX?
11. What is Measures?
12. What is Row level security?
13. What is CALCULATE function in DAX?
14. Where is the data stored in Power BI?
15. What is Embedded Code?
16. What is Dimension table, Fact table?
17. What is the difference between calculated column and measure?
18. What is M Query/function?
19. What is Power BI Service?
20. What is Cardinality?
21. What is difference between SUM () and SUMX ()?
22. What is the difference between star schema and snowflake schema?
23. What is Power BI Gateway?
24. How do you optimize the performance of Power BI Report?
25. What are the filters and its type?
26. What is Query folding?
27. What is Data Modeling in Power BI?
28. What Technique are used to reduce the size in Power BI.

29. What is Data Masking.

30. What is Various refresh options available in Power BI.

31. What is Drill Through.

32. In How many modes you can get the data in Power BI

33. What is append and merge.

34. Different types of charts available in Power BI?

35. What are the key components in data modeling in Power BI.

36. What is the role of visuals in a Power BI report?

What is BI?

BI is stands for business intelligence it is process for analyzing business data and present that data in graphical way using graphs and charts to make interactive dashboards.

This report is helpful for business owners to take important decision for their organization growth.

POWER BI

TABLEAU

QLIKVIEW

ALTEREX

What is Power BI and why?

Power BI is business intelligence tool which is developed by Microsoft which is used for analyzing your data, clean your data and to make interactive dashboards.

Because from many years of market research power bi is consistently ranked as number one bi tool.

What is a Power BI dashboard?

A dashboard is a single page or canvas in Power BI that shows a consolidated view of key metrics and KPI

It provides high level overview with real time data typically sourced from multiple reports or datasets

Dashboard are best for monitoring performance at a glance.

What is Power BI report?

A Report consist of multiple pages with detailed, in depth data visualization

Reports allow users to Slice, filter and drill down into specific data point for more detailed analysis

Each reports are built from single dataset and can be highly interactive.

What is difference between Power BI and Tableau?

- Power bi is easy to learn and understand where tableau is complex to understand.
- Power bi offers query editor to edit our data before loading it into Power BI, Tableau doesn't provide.
- Before select between these two we need to understand the parameter like requirement of data, volume of data and size of organization or person who will analyze that data

---

What are the Component of Power BI?

Power BI Desktop-It is software or tool which you can download and install on your system.

Power BI Service-It is online platform where we publish our report and share among organization.

Power Query-It is used for Cleaning Process

Power View-It is used for visualization Process.

Power Pivot-It is used for data modeling Purpose.

What is Data Cleaning?

Data cleaning is the process of identifying and correcting errors inconsistencies and inaccuracies in dataset.

What is ETL?

ETL stand for Extract, Transform, Load. It is the process for retrieving data from various sources transforming it into usable format and loading it into database.

What is DAX?

It is data analysis expression is the formula language which is used for create custom calculation and aggregation on data.

What are the Types of DAX?

Calculated columns are DAX expression that are calculated and stored as part of the data model.

They create a new column with a value calculated based on the expression

Measures are DAX expressions that perform dynamic calculations based on the data being analyzed.

Measures aggregate data like sum, average, count etc. Time intelligence expressions are DAX expressions

that perform calculations based on time periods, such as year to date running totals etc.

Text functions are DAX expressions that manipulate text data such as concatenating extracting replacing etc.

Information functions are DAX expressions that retrieve information about the data such as data type, error handling etc.

Filter functions are DAX expressions that filter data based on conditions such as SUMIF, COUNTIF etc.

Statistical functions are DAX expressions that perform statistical calculations such as average, variance etc.

What is Measures?

Measures are the Calculation that perform on the flag during visualization and analysis. They are used for aggregation and to Perform Complex calculation.

What is Row-level security?

Row-level security (RLS) in Power BI restricts data access for specific users by using filters to limit data at the row level.

What is the CALCULATE function in DAX?

The CALCULATE function in DAX is a powerful function that evaluates expressions or formulas by modifying the filter context. It is also the only function that allows users to modify the filter context of measures or tables.

Where is the data stored in Power BI?

Primarily, Power BI has two sources to store data:

Azure Blob Storage: When users upload the data, it gets stored here.

Azure SQL Database: All the metadata and system artifacts are stored here.

They are stored as either fact tables or dimensional tables.

What is Embedded Code?

In Power BI,Embedded code refers to a snippet of code that allows you to integrate Power BI reports, dashboards, or visuals directly into another application, like a website or custom software,

It display Power BI analytics within your own platform instead of requiring users to access Power BI separately; this functionality is primarily achieved through Power BI Embedded, a service that provides APIs to embed content within external applications

In Power bi there is option where we will generate link for power bi report and share that link in organization so that our client or end users can view our report and dashboards..

---**What is Dimension table and Fact table?**

Fact table- "A fact table holds numerical data (facts) for analysis, Fact table is measured that can be Used to find out the Summed, Averaged or Manipulate. A fact table contain two types of data a dimension key and measure.

Dimension Table-Dimension table means master data that doesn't change frequently and the table that describe dimension involved are called dimension table, while a dimension table provides descriptive context for the facts.

What is the difference between Calculated Column and Measure?

Measures are works on the filter context gives result on flag whereas calculated column works on the row context and it gives result row by row.

Measures take the physical space in model view where calculated column doesn't take any physical space in model view.

In Measures you can create relationship between tables but in calculated column you can't create relationship between tables

---

What is M Query/function?

M Query/code is language that behind the scene of power query when you do any kind of changes in power query corresponding M query is written for it

---

What is Power BI Service?

It is online cloud based service which is used to share our report in organization so anyone in our organization can view our dashboards.

What is Cardinality and why is it important to create relationship between tables?

In Power BI, Cardinality refers to the relationship between two tables in a data model. Relationships between those tables are necessary to accurately calculate results and display the correct information in your reports.

Four cardinality choices exist: many-to-one, one-to-one, one-to-many, or many-to-many.

When creating relationships, it is recommended that the joining field contains unique values in at least one of the tables.

This allows you to use the one-to-many or many-to-one options in your data model.

What is difference between SUM() and SUMX()?

SUM ()-It returns the columns of data and aggregate it and give the result

No expression is allowed

SUMX ()-It returns the sum of all equivalent expression for each row in table

Expression is allowed.

What is the difference between Star schema and Snowflake schema?

- Star schema is a type of Multi-dimensional model used in data warehouse.
- Star schema It contains the fact table as well as dimension table
- Design of Star schema is simple to understand
- Snowflake schema is a type of multidimensional model used in data warehouse
- It contains fact table, dimensional table as well as sub dimensional table
- Design of Snowflake schema is complex to understand.

What is Power BI Gateway?

A connection between the cloud service and the local network

Power BI Gateway is a software tool for accessing data located on an on-premises network. This tool acts as a gatekeeper for the data source.

If anyone needs to access data located on an on-premises network from a web-based or cloud-based application, the request goes through this gateway. All connection requests are processed by **Power BI Gateway,** and access is allocated according to need and authentication.

How do you Optimize the performance of Power BI Report?

To optimize the result, we can minimize the number of visual on page to remove the duplicates from dataset, limit the use of complex calculation using aggregation, we can significantly improve the performance of power bi.

What are the Filters and Its type?

Ans- Filters are mathematical and logical conditions applied to data to filter out essential information in rows and columns. Filters in Power BI sorts information and data based on a bunch of parameters.

Users can pick and choose particular fields or values within fields to only see the information that pertains to them.

Or

These filters help you focus on specific data subsets, compare values, and gain insights.

Types of filters:

1) Visual Level Filter

2) Page Level Filter

3) Report Level Filter

4) Drill Through Filter

Visual-level filters: These filters only affect the data on a single visual, such as a table, graph, card, slicer, or other visual.

Page-level filters: These filters are handy for designing pages that focus on certain subsets of your data because they only filter the data on that certain page.

Report-level Filters: Filters that influence all of the data in a report, irrespective of what

you're having a look at, are called report-level filters.

Drill-down filters: These filters are automatically added to the filter panel when users leverage the drill-down functionality for a visual in your report.

What is Query Folding?

Query folding in Power BI is a technique that optimizes query performance by generating a single query statement that retrieves and transforms source data

Query folding is the ability for a Power Query to generate a single query statement that retrieves and transforms source data.

Conversion of M language into native language of the source side directly instead of on our local machine. as a result, performance of our model increases.

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

What is Data Modeling in Power BI?

Ans- Data modeling in Power BI is the process of analyzing and defining data types and relationships, and creating a visual representation of that data

---

What Technique can be used to reduce the Size of Power BI report

Ans-

1. Data Reduction Techniques:
2. Use Data Compression:
3. Use Query Folding in Power Query:
4. Remove Unused Fields:
5. Limit the Use of Calculated Columns:
6. Avoid Repeating Data in Visuals:
7. Optimize Relationships and Joins:
8. Use Incremental Refresh for Large Datasets:
9. Disable Auto Date/Time:

10.Optimize Visual Interaction Settings:

What is Data Masking?

Ans- Data masking in Power BI is a way to hide or cover up sensitive information so that only authorized users can see it.

What are the various refresh options available?

Four main refresh options are available in Power BI:

- **P**ackage/OneDrive refresh: This synchronizes Power BI desktop or Excel file between the Power BI service and OneDrive
- Data/Model refresh: This means scheduling the data import from all the sources based on either refresh schedule or on-demand.
- **T**ile refresh: Refresh the tiles' cache on the dashboard every time the data changes.
- Visual container refresh: Update the reports' visuals and visual container once the data changes

---

What is Drill Through In Power BI.

Ans- Drill through in Power BI is a feature that allows users to click on a specific data point in a report and "drill through" to another page that provides more detailed information about that point. It's a way to explore data in depth by moving from a summary view to a more focused, detailed view.

Drill through helps users dive deeper into the data without cluttering the main report, making it easier to analyze and explore insights efficiently.

In how many modes you can get the data in Power BI?

Ans-In conclusion, Power BI offers two modes for accessing data: Import mode and Direct Query mode.

Import mode:

In this mode, data from the source is imported into Power BI Desktop or Power BI

Service. This mode is suitable for small to medium-sized datasets that do not

frequently change.

It provides fast performance as the data is stored within the Power BI file or dataset.

Direct Query mode:

With Direct Query mode, Power BI sends queries directly to the data source each time a visualization or report is refreshed or interacted with.

This mode is suitable for large datasets or scenarios where you need real-time or near real-time access to data.

However, it's important to note that performance may be slower compared to Import mode, especially for complex queries or large datasets.

Live Connection:

- Similar to Direct Query, but this is used for specific data sources like SQL Server Analysis Services (SSAS), Power BI datasets, and Azure Analysis Services.

Composite Model:

- This mode allows you to combine both Import and Direct Query modes within the same report.

What is difference between Append and Merge?

Append: When you have additional row of data that you add to an existing query, you append query also to append structure of both the queries/tables should be identical

Merge-When you have one or more columns that you would like to ADD to another query. You merge the query

Six types:

Left outer

Inner

Right outer

Full OUTER

Left anti

Right anti

Q. Different types of charts available in Power BI?

Ans-

1. Bar Chart: Bar charts are the standard for looking at a specific value across different categories.

2. Multi row card: Multi row cards display one or more data points, one per row.OR

Single Number: Single number cards display a single fact, a single data point.

3. Combo Chart: A combo chart combines a column chart and a line chart. Combining the two charts into one lets you make a quicker comparison of the data. Combo charts can have one or two Y axes, so be sure to look closely.

4. Funnel Chart: Funnels help visualize a process that has stages, and items flow sequentially from one stage to the next.

5. KPI: A **key performance indicator** is a measurable value that shows how effectively you're meeting your goals. Key performance indicators (KPIs) are targets that help you measure progress against your most strategic objectives..

6. Line Chart: Line charts emphasize the overall shape of an entire series of values, usually over time.

7. Pie Chart: Pie charts show the relationship of parts to a whole.

8. Slicer: A slicer is a standalone chart that can be used to filter the other visuals on the page.

9. Table: A table is a grid that contains related data in a logical series of rows and columns.

10. WaterFall Chart: A waterfall chart shows a running total as values are added or subtracted

What are the key components in data modeling in Power BI?

Ans-

Here are some key components of data modeling in Power BI:

- Tables: The building blocks of a data model, tables are containers for data.
- Relationships: Established using primary and foreign keys, relationships are one of the most essential parts of data modeling in Power BI.
- Cardinality: The relationship between two tables or columns, defined by the number of unique values in each table.
- Cross-filter direction: Determines how filters applied to one table affect related tables.
- Power Query: A component that allows users to gather and redesign data from multiple sources and platforms.
- Hierarchies: Allow users to explore data at different levels of granularity.
- Power Pivot: An Excel add-in that allows users to combine large amounts of data from various sources to create reports and dashboards.
- Data visualization: A crucial aspect of data analysis, data visualization allows users to communicate complex information in a clear and visually appealing way.
- Measurement: Involves understanding data sources, creating relationships between tables, and crafting measures and calculated columns to derive metrics

---

What is the role of visuals in a Power BI report?

Visuals in a Power BI report are graphical representations of data that help users analyze and understand information

---

**********SQL**********

1. What is SQL?
2. What is Database?
3. What is DBMS?
4. What is difference between DBMS and RDBMS?
5. What is difference between SQL v/s MySQL
6. What is difference between SQL v/s NoSQL
7. What is DDL and DML command?
8. What is Where clause?
9. What is Constraints?
10. What is SUB QUERIES?
11. What is CTE?
12. What is Join and its types?
13. What is the difference between having and where clause.
14. What are Primary Keys and Foreign Keys?
15. What is GRANT & REVOKE?
16. What is Group by?
17. What is Case expression?
18. What is commit and rollback?
19. Explain Order of execution in SQL?
20. Explain Window function in SQL?
21. What is Index and their types?
22. What is difference between Union and Union all?
23. What is Query Optimization?
24. What is date, time function in SQL?

Q. What is SQL?

Ans-

SQL is stands for (Structured Query Language).

SQL is a Standard language for storing, manipulating and retrieving data in databases.

SQL is not a database system, It is language which is used by database management system like (Oracle, MySQL, Microsoft SQL Server, MongoDB and more).

It is also pronounced as See-Quell.

Q. What is Database?

Ans-A database is an organized collection of structured information, or data,

---

Q. What is DBMS?

Ans-DBMS (Database Management System) is software used to create, manage, and

---

organize Databases

Q. What is difference between SQL v/s MySQL

Ans-SQL is a language used to perform CRUD operations in Relational DB, while MySQL is a RDBMS that uses SQL.

Q. What is Where clause?

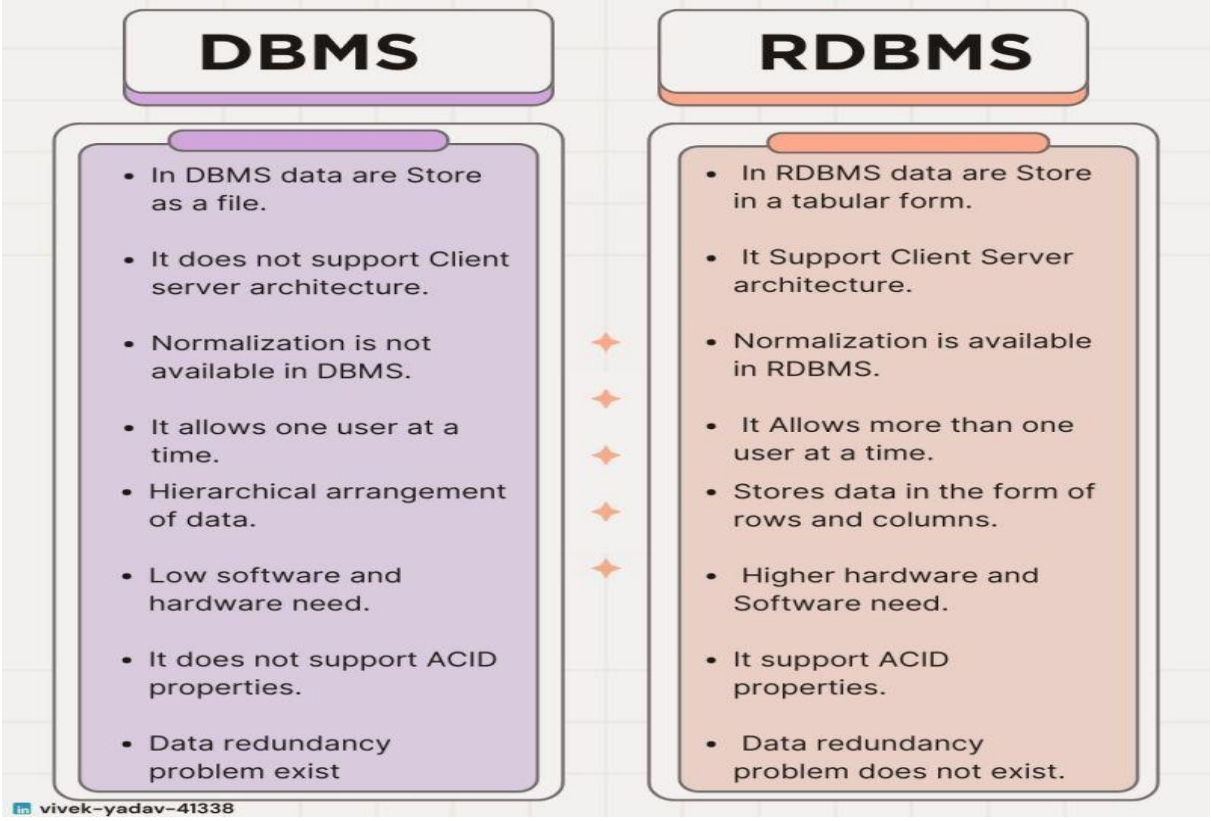
Ans- The WHERE clause in SQL is used to filter records and specify which rows you want to retrieve or modify in a query.

It allows you to define conditions that rows must meet to be included in the result set or to be affected by an operation.

Q. What is the difference between Having and Where clause?

Ans: HAVING is used to specify a condition for a group or an aggregate function used in select statement. The WHERE clause selects rows before grouping. The HAVING clause selects rows after grouping. Because the where clause cannot be used with aggregate functions.

Q. What is difference between DBMS and RDBMS.



Ans-

The main difference between DBMS (Database Management System) and RDBMS (Relational Database Management System) lies in the way they manage and structure data.

Q. What is Constraints?

Ans-Constraints are used to specify rules for data in a table

This ensures the accuracy and reliability of the data in the table

---

Q. What is SUB QUERIES?

Ans-Subqueries, also known as nested queries or inner queries, allow you to use the result of one query (the inner query) as the input for another query (the outer query).

Subqueries are often used to retrieve data that will be used for filtering, comparison, or calculation within the context of a larger query.

Q. What is difference between SQL and NoSQL?

Ans-

Feature	SQL and NoSQL database difference	
	SQL	NoSQL
Data model	Structured	Unstructured or semi-structured
Scalability	Vertical scaling	Horizontal scaling
Performance	Complex queries can be slow	Fast read and write performance
ACID compliance	Yes	No or partial
Schema	Rigid	Flexible
Language	SQL (Structured Query Language)	JSON (JavaScript Object Notation), XML, YAML, or binary schema

---

Q. What is DDL and DML command?

Ans-

Types of DDL command:

CREATE: It is used to create objects in the database, such as tables, views, stored procedures, and more.

ALTER: It is used to modify the structure of an existing database object.

DROP: It is used to delete an entire object or part of an object from the database.

TRUNCATE: Used to delete all records from a table but does not delete the table structure. RENAME: Used to rename an existing database object.

Types of DML command:

INSERT: Used to add new records to a database table.

UPDATE: Used to modify existing records in a database table.

DELETE: Used to delete existing records from a database table.

MERGE: Used to combine data from two or more tables into one.

SELECT: Used to retrieve data from one or more tables in a database.

CALL: Used to call a stored procedure or function.

Q. What is CTE? Why to use CTE?

Ans-A Common Table Expression or CTE is a temporary named given to result set that are created using select statement that can be used in subsequent select statement.

We can define CTE by adding WITH clause directly before Select, Insert, Update, Delete statement.

Because They provide a way to simplify complex queries and make them more

readable. The WITH clause can include one or more CTE

Q. What is Join And its types?

Ans: An SQL Join is used to combine data from two or more tables, based on a common field between them.

Types of Joins:

1) INNER JOIN-Returns records that have matching values in both tables

2) LEFT JOIN-Returns all records from the left table, and the matched records from the right table

3) RIGHT JOIN-Returns all records from the right table, and the matched records from the left table

4) FULL JOIN-Returns all records when there is a match in either left or right table

---

Q. What are Primary Keys and Foreign Keys?

Ans: Primary keys is a unique identifier for each record in a table,

They must contain unique values and cannot be null. A table can have only one primary

key. A FOREIGN KEY is a field in a table, that refers to the PRIMARY KEY of another

table.

Foreign keys establishing a relationship between tables.

Q. What is GRANT & REVOKE?

Ans-The grant command is used to provide specific privilege or permission to user.

The revoke command is used to remove specific privilege that previously granted.

Q. What is Group By?

The group by statement groups row that have the same values into rows.

Q. What is Case Expression?

Ans-

In SQL, CASE is a conditional expression used to create if-then-else logic within queries.

It allows you to return different values depending on whether certain conditions are true, similar to how you might use IF statements in programming.

OR

The case expression goes through condition and return a value when the first condition is met so once a condition is true it will stop reading and return the result, if no condition is true it return the value in the else clause.

Q. What is Commit and Rollback?

The COMMIT command is used to permanently save the changes made during transaction

The rollback command is used to undo changes made during a transaction

---

Q. Explain Order of execution in SQL?

Ans-An SQL query comprises of various clauses like SELECT, FROM, WHERE, GROUPBY, HAVING, and ORDERBY clauses.

Clause Function-

FROM / JOIN -When you write any query, SQL starts by identifying the tables for the data retrieval and how they are connected.

WHERE -It acts as a filter; it filters the record based on the conditions specified by the users.

GROUP BY- The filtered data is grouped based on the specified condition.

HAVING -It is similar to the WHERE clause but applied after grouping the data.

SELECT -The clause selects the columns to be included in the final result.

DISTINCT -Remove the duplicate rows from the result. Once you apply this clause, you are only left with distinct records.

ORDER BY -It sorts (increasing/decreasing/A->Z/Z->A) the results based on the specified condition.

LIMIT / OFFSET -It determines the number of records to return and from where to start

General Order:

SELECT COLUMNS

FROM TABLE NAME

WHERE CONDITION

GROUP BY COLUMNS

HAVING CONDITION

ORDER BY COLUMNS ASC

Q. Explain Window function in SQL?

Ans-A Window function performs a calculation across a set of table rows that are somehow related to the current row.

or

A Window function makes a calculation across multiple rows that are related to the current row.

For example, a window function allows you to calculate:

Running totals (i.e. sum values from all the rows before the current row) 7-day moving averages (i.e. average values from 7 rows before the current row) Rankings

---

Q. What is Index and their types?

Ans- An Index in SQL is a database object used to improve the speed of data retrieval operations on a table.

Think of it as a "lookup guide" that helps the database find rows much faster, similar to how an index in a book helps you quickly find specific topics.

How Indexes Work:

- When you create an index on a column or a set of columns in a table, the database creates an internal structure that allows it to find rows without scanning the entire table.
- Indexes are especially useful for queries that use WHERE clauses, JOINs, or ORDER BY operations, as they significantly improve query performance.

Types of Indexes in SQL:

1.Clustered Index:

o Definition: A clustered index determines the physical order of data in a table. A table can only have one

clustered index because the rows themselves are stored in the order of the clustered index.

o Example: If a table has a clustered index on the ID column, the rows will be stored on disk in the order of ID.

o Impact: Faster retrieval for queries that use the indexed column, but slower for inserts/updates due to the need

to maintain order.

o Usage: Typically created on the primary key of a table.

o Example: CREATE CLUSTERED INDEX idx_name ON table_name(column_name);

2. Non-Clustered Index:

o Definition: A non-clustered index creates a separate structure to store the index and points to the actual data

in the table. It does not affect the physical order of rows.

o Example: A table can have multiple non-clustered indexes.

o Usage: Used to improve performance of SELECT queries where the column(s) in the WHERE clause or JOIN conditions are frequently used.

o Example: CREATE NONCLUSTERED INDEX idx_name ON table_name(column_name);

3. Unique Index:

o Definition: A unique index ensures that the values in the indexed column are unique across the table, preventing duplicate values.

o Usage: It's often created on columns that must have unique values, like email addresses or usernames.

o Example: CREATE UNIQUE INDEX idx_name ON table_name(column_name);

4. Composite Index (Multi-Column Index):

o Definition: A composite index is an index that includes more than one column.

o Usage: Useful when queries frequently filter or sort data based on multiple columns.

o Example: CREATE INDEX idx_name ON table_name (column1, column2);

5. Full**-Text** Index:

o Definition: A full-text index is used for performing full-text searches on large text fields,

such as those found in VARCHAR or TEXT columns.

o Usage: Useful for searching text fields for words, phrases, or patterns.

o Example: CREATE FULLTEXT INDEX idx_name ON table_name(column_name);

6. Filtered Index:

o Definition: A filtered index is an optimized non-clustered index that only includes a subset of rows from the table,

based on a filter condition.

o Usage: This is helpful when you frequently query a small portion of data in a large table,

for example, records with a specific status.

o Example: CREATE INDEX idx_name ON table_name(column_name) WHERE column_name = 'Active';

7. Bitmap Index (Used in Some Databases):

o Definition: A bitmap index stores index entries as bitmaps, which can be very efficient for columns with a small number of distinct values,

such as Gender or Yes/No fields.

o Usage: Typically used in OLAP systems rather than OLTP systems.

o Example: CREATE BITMAP INDEX idx_name ON table_name(column_name); (Note: Supported in databases like Oracle, not in SQL Server or MySQL).

Key Points About Indexes:

- Trade-off: While indexes speed up data retrieval, they can slow down INSERT, UPDATE, and DELETE operations, as the index needs to be updated when data changes.
- Index Selectivity: A highly selective index (where many rows have unique values) is more beneficial than a low-selectivity index (where many rows have the same value).
- Storage: Indexes take up extra storage space in the database.

Indexes are a powerful tool for improving SQL query performance, but it's important to use them wisely to maintain the balance between speed and storage or update costs.

Q. What is difference between Union and Union all?

Ans-

UNION: Removes duplicate rows, ensuring that all rows in the result set are unique.

UNION ALL: Includes all rows from the combined queries, preserving duplicates.

Q. What is Query Optimization?

Ans-Query optimization in SQL is the process of improving the efficiency of a query so that it runs faster and uses fewer

resources (like CPU and memory). When a query is executed, the SQL database engine tries to find the most efficient way to

retrieve or manipulate the requested data. Query optimization aims to ensure that the database runs the query in the best possible way.

Q. What is Date, Time function in SQL?

Ans- Date and time functions in SQL are used to manipulate and format date and time values. These functions help in performing operations like extracting specific parts (year, month, day), calculating differences between dates, adding or subtracting intervals, and formatting date/time into different formats.

**********Statistics**********

What is Mean,Median,and Mode?

What is Central tendancy?

What is Standard deviation?

What is Variance?

What is Probablity?

What is Percentile?

What is Linear regression and Logistics regression?

What is Outliers?

What is Hypothesis Testing?

What is Sampling?

What is Bell curve?

What is Mean, Median, Mode?

Mean-average value of dataset

Median-The median is the middle value in a data set that split dataset in

half Mode-most frequently value in dataset

What is Central tendency?

Ans-Central tendency refers to measure used to determine the "center" of the distribution of data.

What is Standard deviation?

Ans-Standard deviation is measure of amount of variation or dispersion in a set of values

To find Standard deviation we take the square root of the variance.

What is Variance?

Variance is a measure of how data points are differ from the mean.

What is Probablity?

Probability deals with the likelihood of events occurring, assigning a measure between 0 and 1 to quantify uncertainty. It is a branch of mathematics that deals with the occurrence of a random event.

It helps us understand the chances of different outcomes in uncertain situations

---

What is Percentile?

A percentile is a value below which a certain percentage of observation lie

OR

A Percentile is defined as the percentage of values found under the specific values.

In statistics, a percentile is a term that describes how a score compares to other scores from the same set.

OR

A percentile is a measure used in statistics to indicate the value below which a given percentage of data points fall.

In simple terms, it tells you how a particular value compares to the rest of the data.

What is Linear regression and Logistics regression?

Linear Regression-Linear regression is defined as an algorithm that provides a linear relationship between an independent variable and a dependent variable to predict the outcome of future events.

Logistics Regression-Logistic regression, also known as a logit model, is a statistical analysis method to predict a binary outcome, such as yes or no, based on prior observations of a data set.

What is Hypothesis Testing?

Ans-

Hypothesis testing is type of stastical analysis in which you put your assumptions to the test.

It is used for estimate relationship between 2 stastical variables.

Hypothesis testing provides a way to verify whether experiment is valid.

What is Outliers?

Ans-In Statistics, an outlier is a data point that differs significantly from other observations

or

An Outlier is an observation that lies an abnormal distance from other values in a random sample from a population

---

What is Sampling.

Ans-Sampling is a statistical method to understand a subset of data from an entire data set to estimate the characteristics of the whole population.

What is Bell curve?

Ans-In simple terms, a bell curve is a graph that shows how data is spread out. Most of the values in the data are in the middle, and fewer values are at the ends. It gets its name because the graph looks like the shape of a bell.

**********Python**********

What is Python.

Ans-Python is a popular, high-level programming language. It was created by Guido van Rossum, and released in 1991

It is used for:

web development (server-side),

software development,

mathematics,

system scripting.

Features of **Py**thon:

Easy to Understand

Free and Open source

High level Language

Portable: works on windows/Linux

**********NumPy**********

What is NumPy?

Ans- NumPy is a popular Python library used for working with numbers and arrays (collections of numbers).

It is especially useful for scientific computing and handling large amounts of data efficiently.

**********Pandas**********

What is Pandas?

Ans-Pandas is Open source data analysis library in python. It uses Power and Speed of Numpy to make data analysis and preprocessing easy.

It provides rich and high robust data operations.

What is Structure of Pandas?

Ans-

Pandas Has two type of data structure:

- Series: Its one dimensional array with indexes, It store single column or row of data in dataframe:
- Dataframe: Its tabular spreadsheet like structure representing rows each of which contains one or multiple columns

---

**********Projects**********

Project on "E Commerce sales"

Introduction:

In this E-commerce sales analysis project, our goal was to analyze sales data from online store to gain insights into customer behavior, product performance, and overall business trends.

By using Power BI, we provide actionable insights to optimize its sales strategies and improve business performance.

Conclusion:

In conclusion, we successfully demonstrated the power of data driven insights to optimize its efficiency in business growth By using the capabilities of Power BI, we were able to unlock valuable insights that informed strategic decision-making and positioned the business for future success.

Project on "Financial performance analysis"

Introduction:

The financial performance analysis project in Power BI aimed to evaluate the company's financial health and performance through interactive visualizations and in-depth analysis of financial data. How? In Power BI, we created interactive dashboards and reports to visualize financial data.

For example, we used line charts to track revenue trends over time, pie charts to analyze expense breakdowns, and bar charts to compare profitability by product or segment.

Conclusion:

In conclusion, the financial performance analysis project in Power BI provided valuable insights into the company's financial health and performance. By leveraging Power BIs capabilities, we were able to deliver actionable recommendations to drive business growth and optimize financial outcomes.

Project on "Real estate stock analysis"

Introduction:

The stock market project involves analyzing stock data to create dashboard that provide insights into stock performance ,trends, and other key metrics.The project will include data cleaning, analysis using SQL queries and visualization using a dashboard tool like tableau or power bi.

In the end of the project ,you can create a comprehensive dashboard that provides valuable insights into stock performance and market trends,enhancing your data analysis skills and offering practical value to stakeholders.

Project on "HR admin analysis"

Introduction:

The HR Admin Analysis project aimed to provide insights into various HR metrics and processes within the organization.By leveraging Power BI, we sought to analyze employee data to improve decision-making and optimize HR strategies. How? Our Power BI dashboard included visualizations such as turnover trend charts, absenteeism heat maps, performance score distributions, and satisfaction score comparisons.

Through these visualizations, we identified trends, outliers, and correlations within the HR data, enabling better decision-making.

Conclusion:

In conclusion, the HR Admin Analysis project demonstrated the value of leveraging Power BI for HR analytics to drive strategic HR decision-making and improve organizational performance.

Moving forward, I am excited to continue using data-driven insights to enhance HR practices and contribute to organizational success.

Project on "Healthcare system"

Introduction:

The Healthcare System project in Power BI aimed to analyze healthcare data to improve patient care and operational efficiency within healthcare organizations. By leveraging Power BIs capabilities, we sought to extract actionable insights to support strategic decision-making and enhance healthcare outcomes.

Conclusion:

In conclusion, the Healthcare System project in Power BI demonstrated the value of data-driven insights in improving patient care and healthcare operations. -

Project Explanations: as per Resume

1. E-Commerce Sales Analysis:
    - Description: Analyzed e-commerce sales data using Power BI to evaluate growth or decline trends. By comparing current and past sales data, calculated percentage changes and identified areas of improvement or decline, revealing a 15% change in key metrics.
    - Impact: Insights helped businesses strategize for inventory management and marketing campaigns.
2. Financial Performance Analysis:
    - Description: Built a comprehensive Power BI dashboard to evaluate financial performance. This included analysis of trading metrics, balance sheets, and profit/loss data using SQL for database extraction and Power BI for visualization.
    - Impact: Improved the decision-making process by enhancing financial insights by 20% and reducing reporting time by 15%.
3. Real Estate Sector Analysis:
    - Description: Conducted trend analysis in the real estate sector using Python for data preprocessing and Power BI for visualization. Focused on market trends, property price variations, and demographic insights.
    - Impact: Improved trend identification by 25%, aiding in strategic planning and investment decision-making.

4. Sales Analysis (Excel):
    - Description: Conducted a detailed analysis of retail sales data using Excel. Identified seasonal trends and purchasing patterns, enabling actionable recommendations.
    - Impact: Resulted in improved inventory turnover rates by 20%, optimizing stock levels and reducing excess inventory.

Common Data Analyst Interview Questions with Answers:

General Questions:

1. What are the key responsibilities of a data analyst?
    - Answer: Collecting, cleaning, and organizing data; conducting exploratory data analysis (EDA); creating visualizations to communicate insights; supporting decision-making by interpreting trends; and ensuring data quality and accuracy.
2. How do you handle missing data in a dataset?
    - Answer: Use techniques such as:
        - Removing rows/columns with a high percentage of missing values.
        - Imputation using mean/median/mode or predictive modeling.
        - Employing domain knowledge to estimate values or using algorithms that can handle missing data, like decision trees.
3. What tools are you most comfortable with, and why?
    - Answer: I am proficient in Power BI, Python, SQL, and Excel. Each tool has specific strengths:
        - Power BI for interactive dashboards.
        - Python for data preprocessing and advanced analytics.
        - SQL for querying and managing databases.
        - Excel for quick calculations and initial data exploration.

Scenario-Based Questions:

4. Can **y**ou describe a time you worked with a large dataset? How did **you** approach it?
    - Answer: During my internship, I worked on a dataset exceeding 1 million rows. I automated the cleaning process using Python and SQL, reducing processing time by 40%. Steps included:
        - Sampling the data for initial exploration.
        - Identifying and resolving inconsistencies.
        - Writing modular Python scripts for repetitive tasks.
5. How do you ensure data accuracy in **y**our analysis?
    - Answer:
        - Validate data sources and formats.
        - Conduct preliminary checks, like row counts and value ranges.
        - Use SQL queries and Python scripts for duplicate checks and consistency.
        - Apply logical tests and cross-reference outputs with business expectations.

Technical Questions:

6. What is the difference between JOINs in SQL?
    - Answer:
        - INNER JOIN: Returns rows with matching values in both tables.
        - **LEFT JOIN:** Returns all rows from the left table and matching rows from the right.
        - **RIGHT JOIN**: Returns all rows from the right table and matching rows from the left.
        - FULL JOIN: Combines all rows from both tables, including unmatched rows.
7. What is the purpose of a dashboard in data analysis?
    - Answer: Dashboards present data visually to stakeholders, enabling quick insights and informed decision-making. They aggregate data from multiple sources into interactive visualizations for real-time tracking of KPIs.
8. How would you explain a complex analysis to a non-technical stakeholder?

- Answer: Focus on the key business impact rather than technical details. Use simple language and visualizations (charts or dashboards) to illustrate findings, emphasizing actionable insights.

What are Roles and Responsibilities for Each Project

E-Commerce Sales Analysis (Power BI)

- **Roles:**
  - Collaborated with the sales and marketing teams to gather business requirements.
  - Defined KPIs to track performance, such as growth rate, sales volume, and regional performance.
- **Responsibilities:**
  - Extracted sales data from internal systems or databases for analysis.
  - Cleaned and preprocessed data to ensure accuracy and consistency.
  - Created visualizations in Power BI, such as trend lines and bar charts, to showcase year-over-year growth and decline rates.
  - Delivered actionable insights to improve sales strategies and optimize inventory management.

Financial Performance Analysis (Power BI + SQL)

- **Roles:**
  - Worked closely with finance teams to identify key financial metrics.
  - Ensured the accuracy of data sourced from transactional databases using SQL queries.
- **Responsibilities:**
  - Designed SQL queries to extract and join data from multiple tables, such as revenue, expenses, and balance sheets.
  - Developed Power BI dashboards with clear, interactive visualizations of profit margins, revenue trends, and expense breakdowns.
  - Conducted variance analysis to identify deviations from financial targets.
  - Presented reports to stakeholders, enabling faster and more informed decision-making.

Real Estate Sector Analysis (Power BI + Python)

- **Roles:**
  - Partnered with market analysts to understand sector-specific trends.
  - Utilized Python to handle complex data transformations.
- **Responsibilities:**
  - Analyzed real estate market data, such as property prices, demand patterns, and demographic statistics.
  - Automated data cleansing using Python libraries like Pandas and NumPy.
  - Built dashboards in Power BI to visualize property price trends, geographic patterns, and investment opportunities.
  - Recommended strategies based on data trends to aid stakeholders in strategic planning.

Sales Analysis (Excel)

- **Roles:**
  - Coordinated with retail teams to align data requirements with business goals.
  - Used Excel to handle and process large datasets efficiently.
- **Responsibilities:**
  - Imported and structured raw sales data into Excel for analysis.
  - Conducted seasonal trend analysis using pivot tables, charts, and statistical functions.
  - Identified underperforming products and suggested promotional strategies to increase turnover.
  - Prepared summary reports for management, highlighting actionable insights.

General Responsibilities Across All Projects

- Data **P**reparation: Ensured all datasets were cleaned, validated, and formatted for analysis.
- **T**ool Utilization: Leveraged tools like Power BI, SQL, Python, and Excel to extract, transform, and visualize data.
- Stakeholder Communication: Collaborated with cross-functional teams to align analysis outcomes with business objectives.
- Insight Generation: Interpreted data trends to provide actionable recommendations.
- **R**eport Development: Delivered comprehensive dashboards and reports to enhance transparency and decision-making.

What challenges you faced while doing a projects

Challenges Faced in Each Project

1. E-Commerce Sales Analysis (Power BI)

- Challenge:
  - Incomplete or inconsistent sales data: Missing entries in key fields like transaction dates or sales values.
- Solution:
  - Performed data imputation techniques, such as filling missing values with averages or interpolating trends.
  - Standardized data collection processes by collaborating with the sales team to ensure future consistency.
- Challenge:
  - Managing large datasets: Power BI occasionally slowed down due to large data volumes.
- Solution:
  - Optimized data models by reducing the number of fields and pre-aggregating data before importing into Power BI.

2. Financial Performance Analysis (Power BI + SQL)

- Challenge:
  - Complex database structures: Understanding relationships between multiple tables in financial databases.
- Solution:
  - Conducted detailed schema analysis and worked closely with database administrators to clarify table relationships.
  - Used SQL joins and subqueries to extract meaningful insights without overcomplicating queries.
- Challenge:
  - Communicating financial insights: Stakeholders with limited technical knowledge found it hard to understand advanced financial metrics.

- Solution:
    - Simplified dashboards with intuitive visuals like bar charts and KPIs.
    - Provided written summaries alongside dashboards to explain trends in plain language.

---

3. Real Estate Sector Analysis (Power BI + Python)

- Challenge:
    - Data inconsistencies from multiple sources: Different formats and scales in datasets collected from government reports, market surveys, and real estate agencies.
- Solution:
    - Developed Python scripts using Pandas for data standardization (e.g., converting all prices to a uniform currency).
    - Applied statistical techniques to normalize data and ensure compatibility across datasets.
- Challenge:
    - Identifying trends in highly volatile markets: Real estate data often showed fluctuations due to economic conditions or policy changes.
- Solution:
    - Used moving averages and time-series analysis to smooth out short-term volatility.
    - Highlighted long-term patterns to stakeholders for better decision-making.

---

4. Sales Analysis (Excel)

- Challenge:
    - Large data files slowing down Excel: Processing large retail sales datasets caused performance lags.
- Solution:
    - Split data into smaller chunks and used Power Query to preprocess data.
    - Utilized Excel's advanced features like pivot tables with filters to optimize performance.
- Challenge:
    - Uncovering seasonal trends: Difficulty identifying hidden patterns in raw sales data.
- Solution:
    - Applied Excel's statistical functions and charts (e.g., trendlines) to identify seasonal variations and correlations.

Common Challenges Across Projects

1. Balancing Stakeholder Expectations:
    - Issue: Stakeholders often had evolving requirements mid-project.
    - Solution: Held regular review meetings to gather feedback and ensure alignment with changing business needs.
2. **Technical** Roadblocks:
    - Issue: Bugs or limitations in tools like Power BI or Python libraries during data processing.
    - Solution: Researched solutions in community forums, consulted documentation, and tested alternative approaches.
3. **Data Security and Privacy:**
    - Issue: Ensuring sensitive data remained secure while performing analysis.
    - Solution: Followed organizational data security protocols, anonymized data where necessary, and worked in secure environments.
4. **Time Constraints:**
    - Issue: Delivering high-quality analysis within tight deadlines.
    - Solution: Prioritized tasks based on business impact and streamlined processes using automation tools.