

Data Visualization – Power BI

1. Answer the following questions (15 Marks)

a. Explain the different roles in data.

The following sections highlight these different roles in data and the specific responsibility in the overall spectrum of data discovery and understanding:

- Business analyst
- Data analyst
- Data engineer
- Data scientist
- Database administrator

Business analyst

While some similarities exist between a data analyst and business analyst, the key differentiator between the two roles is what they do with data. A business analyst is closer to the business and is a specialist in interpreting the data that comes from the visualization. Often, the roles of data analyst and business analyst could be the responsibility of a single person.

Data Analyst

A data analyst helps businesses maximize the value of their data by using visualization and reporting tools like Microsoft Power BI. Their role involves profiling, cleaning, transforming data, building scalable semantic models, and implementing advanced analytics in reports to deliver meaningful insights. They collaborate with stakeholders to define data and reporting needs, manage Power BI assets such as reports, dashboards, and workspaces, and ensure data security in line with stakeholder requirements. Additionally, data analysts work closely with data engineers and database administrators to access relevant data sources and continuously improve data collection processes for analysis.

Data Engineer

Data engineers provision and manage data platform technologies, both on-premises and in the cloud, ensuring the secure and seamless flow of structured and unstructured data from multiple sources. They work with various data platforms, including relational and nonrelational databases, data streams, and file stores, and are responsible for integrating data services across these platforms. Their primary tasks involve using data services and tools to ingest, egress, and transform data, as well as collaborating with business stakeholders to identify data requirements and design solutions that meet business needs.

While there is some overlap with database administrators, data engineers focus more broadly on managing data pipelines and platforms rather than just operational data management. They play a critical role in accelerating business intelligence and data science projects by wrangling and preparing data, allowing data scientists and analysts to concentrate on their analyses.

Data engineers also support data analysts in accessing various data sources and optimizing semantic models typically served from modern data warehouses or data lakes. Professionals from database administration or business intelligence backgrounds can transition into data engineering roles by learning the relevant tools and technologies needed to process and manage large-scale data.

Data Scientist

Data scientists perform advanced analytics ranging from descriptive analytics, which involves exploratory data analysis (EDA), to predictive analytics that use machine learning models to detect patterns and forecast outcomes. Their work may also include deep learning and developing custom algorithms to solve complex problems through iterative experiments. While a significant part of their role involves data wrangling and feature engineering, collaboration with data engineers can streamline these processes. Despite perceptions of distinct roles, data scientists and data analysts often work closely together—scientists focus on forming hypotheses and experiments, while analysts support data visualization and reporting to communicate insights effectively.

Database Administrator

A database administrator (DBA) manages the operational aspects of cloud-native and hybrid data platform solutions using Microsoft Azure data services and SQL Server. They ensure the availability, performance, optimization, security, and health of databases, including managing data backup and recovery plans in collaboration with stakeholders. Unlike data engineers, who focus on ingesting, transforming, and preparing data for analysis, DBAs are responsible for monitoring databases and the underlying hardware, as well as controlling user access and maintaining data security based on business requirements.

- b. Explain predictive analytics in brief with examples.

Predictive analytics is a branch of advanced analytics that uses historical data, statistical algorithms, and machine learning techniques to forecast future outcomes or trends. By analyzing past data, predictive models identify patterns and relationships to make informed predictions about future events or behaviors.

For example:

- In retail, predictive analytics helps forecast customer demand, optimize inventory, and suggest personalized product recommendations based on past purchase behavior.
- In healthcare, it can predict disease outbreaks or identify patients at high risk for certain conditions, enabling early intervention.
- In finance, predictive models are used to assess credit risk, detect fraudulent transactions, and predict stock price movements.

- In manufacturing, it anticipates equipment failures through predictive maintenance, reducing downtime and repair costs.

By leveraging predictive analytics, organizations can make proactive decisions, improve efficiency, and gain a competitive advantage.

c. What are the various building blocks in PBI?

The building blocks of Power BI are semantic models and visualizations. Create a semantic model and then use visuals to build a report.

Semantic Model

A semantic model consists of all connected data, transformations, relationships, and calculations. To follow the flow of Power BI, you first connect to data, transform data, and create relationships and calculations to create a semantic model.

First, connect to as many data sources you need. Then clean and transform the data to your needs. Add relationships between tables and calculations to extend the semantic model. After all of that, now you can create a report.

Visualizations

In Power BI Desktop, when you create a visualization (also called visual), you add it to the canvas for a report page. Choose your visualizations to build pages in your report. It's ideal to keep each page simple with related data, so consumers can easily see the insights.

Power BI is a low-code solution, which means that you can "drag and drop" data field directly onto the canvas. Power BI will choose a visual for your data field. You can easily change between visuals for the same fields and add or remove data fields to the visual.

Dashboards

One of the most valuable features of Power BI reports is the interactivity between visuals. Consumers can select different data points in the visual and see how that affects the other visuals. Depending on your design, they can also drill through from one visual to more detail or filter based on different fields in the report.

Once you're satisfied with your report, you publish it to the Power BI service.

In the Power BI service, you can also create dashboards after you've published a report. Dashboards consist of a single page made up of tiles. Add tiles to a dashboard by pinning a visual in a report to the dashboard. Tiles aren't interactive like visuals, so when a user interacts with the tile, they go to the underlying report for more information.

Dashboards are an excellent way to provide high-level information to consumers. Like a dashboard in a vehicle, including the most important information in a dashboard. Then consumers can go to the report for more details.

2. Answer the following questions (15 Marks)

a. Explain the various storage modes used in Power BI.

Power BI uses three primary storage modes to manage how data is stored, accessed, and queried for reports and dashboards. Each mode offers different performance, flexibility, and data connectivity options:

Import Mode:

In this mode, data from the source is imported into Power BI and stored in its highly compressed in-memory storage (VertiPaq engine).

It provides the fastest performance since queries run on the cached data.

However, the data is static and needs to be refreshed periodically to stay updated with the source.

Best for scenarios where high performance and fast interactions are required.

Direct Query Mode:

In Direct Query mode, Power BI does not store data locally. Instead, it queries the source system in real time whenever a user interacts with a report.

This ensures data is always up to date, but performance depends on the source system and network.

Suitable for working with large datasets or when up-to-the-minute data is required but can be slower compared to Import mode.

Composite Mode:

Composite mode allows a combination of Import and Direct Query in a single data model.

It provides flexibility by letting you store frequently used data in memory (Import) for fast access, while less critical data can be queried live from the source (Direct Query).

This hybrid approach offers a balance between performance and real-time data access.

Each mode serves different use cases based on performance needs, data size, and real-time requirements in Power BI reports and dashboards.

b. Explain the following concepts

i. PBI Dataset

A dataset in Power BI is a collection of data that you import or connect to from various sources, such as Excel, SQL databases, cloud services, or web APIs. It serves as the foundation for creating reports and visualizations. Once data is loaded into Power BI, you can clean, transform, and shape it using Power Query. The dataset includes tables, columns, and measures that users can use in reports and dashboards. Datasets can be refreshed to keep the data up to date,

and they are often shared across reports and dashboards in the Power BI service.

ii. Data Model

A data model in Power BI is the structured representation of your data, including the relationships between different tables, calculated columns, measures (formulas), and hierarchies. It defines how data is organized and connected to support reporting and analytics. The data model enables efficient querying and analysis by using in-memory storage (VertiPaq engine) for fast performance. A well-designed data model ensures accurate insights and supports complex calculations by defining clear relationships, primary keys, and business rules between data entities.

c. Assign data types for the following variables/fields in Power Query.

- i. Roll No. – Text
- ii. City – Text
- iii. Subject – Text
- iv. Enrolment Date – Date
- v. Marks scored – Whole Number

3. Answer the following questions (15 Marks)

- a. Explain the concept of aggregation while working with any BI tool.

Aggregation refers to the process of summarizing or combining data to provide meaningful insights in business intelligence (BI) tools. It involves applying mathematical operations to data groups to generate higher-level summaries. Common aggregation functions include SUM, AVERAGE, COUNT, MIN, MAX, and MEDIAN.

For example:

- If you have sales data for each day, you can aggregate it to show total sales per month or average sales per region.
- In Power BI (or any BI tool), aggregations allow you to analyze data at different levels of granularity—such as by time periods (daily, monthly, yearly), geography (city, state, country), or product categories.

Aggregations help users:

- Simplify large datasets, making them easier to understand and analyze.
- Improve performance by reducing the amount of data processed at once.
- Highlight trends, patterns, and key metrics, enabling better decision-making.

In BI tools, aggregations can be applied in visualizations (charts, graphs, tables), calculated fields, or data models, depending on the analysis requirements.

- b. Explain the concept of cardinality and its relevance while creating relationships in a data model.

Cardinality refers to the uniqueness of data values in a column and describes the nature of the relationship between two tables in a data model. It defines how many instances of one entity relate to instances of another entity, and it is a key concept when establishing relationships between tables in BI tools like Power BI.

There are three main types of cardinalities used in relationships:

One-to-One (1:1)

- Each row in Table A relates to exactly one row in Table B, and vice versa.
- Example: Each employee has one unique employee ID.

One-to-Many (1:), or Many-to-One (:1)

- A single row in Table A can be related to multiple rows in Table B.
- This is the most common type of relationship in data models.
- Example: One customer can have many orders.

Many-to-Many (:)

- Rows in Table A can be related to multiple rows in Table B, and vice versa.
- Requires special handling in BI tools because it can create ambiguous relationships.
- Example: Students enrolled in multiple courses, and courses having multiple students.

Relevance of Cardinality in a Data Model

- Accuracy: Understanding and setting the correct cardinality ensures data is aggregated and analyzed correctly, preventing issues like duplication or incorrect totals.
- Performance: Proper cardinality helps optimize the data model and improves query performance.
- Filter propagation: It determines how filters flow between related tables, which affects the behavior of slicers and visuals in reports.

In summary, cardinality is crucial in defining how tables interact in a data model, ensuring data integrity and enabling accurate and efficient analysis in BI tools.

- c. Explain with an example what quick measures are.

Quick Measures are pre-built, ready-to-use calculations in Power BI that allow users to create commonly used metrics without writing complex DAX (Data Analysis Expressions) formulas manually. They help users perform calculations like aggregates, percentages, running totals, time-based analysis, and mathematical operations quickly and easily through a user-friendly interface.

Key Points:

- Quick Measures automatically generate the underlying DAX code for you.
- They are ideal for users who may not be familiar with DAX but need to perform advanced calculations.
- Once created, you can view and edit the DAX formula to customize it further.

Example:

Imagine you have a **Sales** table with columns like **Total Sales** and **Product Category**, and you want to calculate the **percentage of total sales by category**.

1. In Power BI, right-click on a table in the Fields pane and select **New Quick Measure**.
2. Choose **Percentage of Grand Total** from the list of quick measure options.
3. Select **Total Sales** as the base value and **Product Category** as the grouping field.
4. Power BI creates the DAX formula automatically to calculate the percentage of total sales for each category.

This quick measure can now be used in reports and visuals, making it easier to show how each category contributes to overall sales without needing to write DAX manually.