Tableau - Basics Interview Questions

(Practice Project)

PW SKILLS

Basic

1. What is Tableau and what are its main features?

Answer: Tableau is a leading data visualization and business intelligence tool designed to help users analyze and visualize data in an interactive and intuitive manner. Its main features include a user-friendly drag-and-drop interface that allows users to create complex visualizations without extensive coding knowledge. Tableau supports a wide range of data sources, both local and cloud-based, and provides real-time data analysis capabilities. It also enables the creation of interactive dashboards and reports, offering advanced features like calculated fields, parameter controls, and robust sharing options.

2. What are the key components of Tableau Desktop?

Answer: Tableau Desktop is a powerful tool for creating and designing visualizations. The key components include: Data Pane, where users connect to and manage their data sources.
Worksheet, where individual charts and graphs are created.
Dashboard, which combines multiple visualizations into a cohesive view for more comprehensive analysis.
Story, which is a sequence of visualizations designed to present a data-driven narrative.

These components work together to enable users to explore data, build insights, and share findings effectively.

3. What is a data source in Tableau?

Answer: In Tableau, a data source refers to the origin of the data that Tableau uses to generate visualizations and reports. Data sources can be files, such as Excel or CSV, databases like SQL Server or Oracle, or cloud-based platforms such as Google BigQuery or AWS Redshift. Connecting to a data source allows Tableau to import and work with data, enabling users to perform analysis, create visualizations, and build interactive dashboards.

4. What is a calculated field in Tableau?

Answer: A calculated field in Tableau is a user-defined field that is created by applying a formula to existing data fields. This allows users to generate new data values based on calculations or expressions, such as sums, averages, or custom metrics. Calculated fields enable more advanced data analysis by allowing users to create new dimensions or measures that are not present in the original data source. This feature is essential for performing complex data manipulations and deriving insights from the data.

5. Explain the difference between a dimension and a measure in Tableau.

Answer: In Tableau, dimensions and measures serve distinct roles in data analysis. Dimensions are categorical fields that define the qualitative aspects of the data, such as product categories, regions, or dates. They are used to slice and dice the data, providing context and segmentation. Measures, on the other hand, are quantitative fields that represent numerical data, such as sales figures, profit margins, or quantities. Measures are used to perform aggregations and calculations, enabling users to analyze and compare numeric values across different dimensions.

6. What is the purpose of filters in Tableau?

Answer: Filters in Tableau are used to control which data is displayed in visualizations by excluding or including specific data points based on defined criteria. They allow users to focus on relevant subsets of data, improving the clarity and relevance of the analysis. Filters can be applied at various levels, including data source, worksheet, or dashboard, and can be configured to support interactive and dynamic data exploration. This functionality helps users drill down into specific segments of data and gain more precise insights.

7. What is Tableau Public?

Answer: Tableau Public is a free version of Tableau designed for sharing and publishing interactive visualizations and dashboards online. It allows users to create visualizations and share them with the public via the Tableau Public server. While it offers many of the same features as Tableau Desktop, the primary distinction is that all visualizations published to Tableau Public are accessible to anyone on the internet. This platform is ideal for individuals and organizations looking to showcase their data insights without incurring software costs.

8. What are the different types of joins available in Tableau?

Answer: Tableau supports several types of joins to combine data from multiple tables or data sources. These include:

i. Inner Join: Returns only the matching records between the joined tables.

ii. Left Join: Returns all records from the left table and the matched records from the right table, with null values where there is no match.

iii. Right Join: Returns all records from the right table and the matched records from the left table, with null values where there is no match.

iv. Outer Join: Returns all records from both tables, with null values where there are no matches. This includes Full Outer Join, which includes all records from both tables, Left Outer Join, and Right Outer Join.

9. What is a story in Tableau?

Answer: A story in Tableau is a sequence of visualizations that work together to convey a narrative or present a data-driven argument. Stories consist of individual story points, each representing a different visualization or dashboard. These points are arranged in a logical flow to guide the audience through a structured analysis, highlighting key insights and findings. Stories are useful for presenting complex data in a coherent and engaging manner, making it easier for viewers to understand the context and implications of the data.

10. What is the purpose of using parameters in Tableau?

Answer: Parameters in Tableau are dynamic values that can be used to control various aspects of a visualization or analysis. They provide a way for users to input or select values that influence filters, calculated fields, reference lines, and other elements within a dashboard. Parameters enable interactive and customizable features, allowing users to adjust the view and explore different scenarios. They are particularly useful for creating what-if analyses, scenario planning, and enhancing user interaction with the data.

Intermediate

11. Given the following sales data:

How would we create a calculated field to show the average profit per sale?

Region	Sales	Profit
North	5000	1200
South	3000	800
East	4000	1000
West	3500	900

Answer: To create a calculated field for the average profit per sale, use the formula: AVG([Profit]) / AVG([Sales]). This calculates the average profit and divides it by the average sales to get the average profit per sale. Applying this calculation to the sample data, we would compute the average profit as (1200 + 800 + 1000 + 900) / 4 = 975, and the average sales as (5000 + 3000 + 4000 + 3500) / 4 = 3875. The average profit per sale would be 975 / 3875 ≈ 0.251 or 25.1%.

12. What is the difference between a table calculation and a basic calculation in Tableau?
Answer: A basic calculation in Tableau is performed at the data source level and involves simple operations like sums, averages, or custom formulas on individual data points. In contrast, a table calculation operates on the results of visualization and is used to perform computations across a table of data, such as running totals, moving averages, or percent of total calculations. Table calculations are useful for analyzing patterns or trends within the displayed data, rather than in the underlying data source.

13. How can you use a Level of Detail (LOD) expression to calculate the average sales per region regardless of the filters applied?
Answer: To calculate the average sales per region regardless of the filters applied, we would use a FIXED LOD expression. The formula would be: FIXED [Region]: AVG([Sales]). This LOD calculation computes the average sales for each region independently of the filters applied to the view. It provides a consistent average for each region that is unaffected by changes in the filter context.

14. Given the following sales data:
How would you create a table calculation to show the running total of sales?

Product	Sales
A	200
B	150
C	300
D	250

Answer: To create a table calculation for the running total of sales, we can use the RUNNING_SUM function. The formula is RUNNING_SUM([Sales]). This calculation aggregates the sales values sequentially, adding each value to the sum of the previous values. For the sample data, the running total would be:

- For Product A: 200
- For Product B: 200 + 150 = 350
- For Product C: 350 + 300 = 650
- For Product D: 650 + 250 = 900

15. Explain the use of the WINDOW_AVG function in a table calculation.
Answer: The WINDOW_AVG function in Tableau calculates the average of a measure over a specified window of data within a visualization. It is used in table calculations to compute the average value of a measure across a range of rows or columns in the view. For example, applying WINDOW_AVG([Sales]) to a dataset can provide the average sales over a defined period or across different segments, allowing users to analyze trends and patterns in the data.

16. What is the role of the EXCLUDE LOD expression, and when would you use it?

Answer: The EXCLUDE LOD expression in Tableau is used to aggregate data by excluding specific dimensions from the calculation. The formula is EXCLUDE [Dimension]: AGG([Measure]). This is useful when we want to calculate a metric without considering certain dimensions that may be present in the view. For example, suppose we want to calculate the total sales across all products without considering the product category. In that case, we can use an EXCLUDE LOD expression to remove the category dimension from the aggregation.

17. How would you use Tableau Prep to clean and prepare data before analysis?

Answer: Tableau Prep is a data preparation tool that allows users to clean, reshape, and organize data before analysis. It provides features for filtering out irrelevant data, correcting errors, and combining data from multiple sources. Users can create data flows to apply transformations such as joining or blending datasets, aggregating values, and creating calculated fields. Tableau Prep also includes data profiling capabilities to assess data quality and consistency, ensuring that the data is ready for accurate analysis in Tableau Desktop.

18. Given a dataset with customer sales data, how would you create a calculated field to identify high-value customers?

Customer	Sales
Alice	12000
Bob	8000
Carol	15000
Dave	5000
Eve	11000

Answer: To identify high-value customers, we could create a calculated field using a formula based on sales thresholds or criteria. For example, if we define high-value customers as those with sales greater than $10,000, we can use the formula: IF [Sales] > 10000 THEN 'High Value' ELSE 'Low Value' END. This calculated field will categorize customers into "High Value" or "Low Value" based on their sales figures, allowing us to filter or analyze high-value customers more easily.

19. What is a dynamic reference line in Tableau, and how do you create one?

Answer: A dynamic reference line in Tableau is a reference line that adjusts based on user interactions or changes in data. For example, it can be set to display the average value of a measure, which updates automatically as data changes. To create a dynamic reference line, we can add a reference line to a chart, select the option to use a calculated field or aggregate function, such as AVG([Sales]), and configure it to update based on the current view or filter context.

20. How do you handle null values in Tableau?

Answer: In Tableau, null values can be handled through several approaches. we can use functions like ZN() to replace null values with zero, or IFNULL() to substitute null values with a specific value. Additionally, we can configure how null values are displayed in visualizations, such as hiding them or displaying them as a separate category. Proper handling of null values ensures that they do not adversely affect the analysis or misrepresent the data.

21. What is the purpose of using sets in Tableau, and how can they be created?
Answer: Sets in Tableau are custom groupings of data that allow users to create dynamic subsets of a data source based on specific criteria. They are useful for analyzing specific segments or categories within the data. Sets can be created by selecting data points in a view and choosing the "Create Set" option, or by defining rules and conditions for membership. Once created, sets can be used in visualizations, calculations, and filters to focus analysis on particular groups.

22. What is the role of data blending versus data joining in Tableau, and when would you use each method?
Answer: Data blending and data joining are methods for combining data from multiple sources in Tableau. Data joining occurs at the data source level, where tables are combined using join operations (inner, left, right, outer). This method is suitable when working with data from the same source or closely related tables. Data blending, on the other hand, happens at the visualization level, where data from different sources are aggregated and merged based on common fields. Blending is used when dealing with disparate data sources or when data cannot be joined directly. Both methods have their use cases depending on the data structure and integration requirements.

Advanced

23. Given the following sales data, how would you use a window function to calculate the percent of total sales for each region?
How would we apply the WINDOW_SUM function to achieve this?

Region	Sales
North	5000
South	3000
East	4000
West	3500

Answer: To calculate the percent of total sales for each region, use the WINDOW_SUM function to get the total sales and then calculate the percentage for each region. First, create a calculated field for total sales: WINDOW_SUM(SUM([Sales])). Then, create another calculated field to find the percent of total sales: SUM([Sales]) / WINDOW_SUM(SUM([Sales])). Applying this calculation to the data will give:

- For North: 5000 / 15500 ≈ 32.3%
- For South: 3000 / 15500 ≈ 19.4%
- For East: 4000 / 15500 ≈ 25.8%
- For West: 3500 / 15500 ≈ 22.6%

24. Describe a scenario where you would use a FIXED LOD expression versus an INCLUDE LOD expression.
Answer: A FIXED LOD expression is used when we want to calculate a metric at a fixed level of granularity, regardless of the filters applied in the view. For instance, calculating the total sales per region regardless of any other dimensions in the view. The formula would be: FIXED [Region]: SUM([Sales]).
An INCLUDE LOD expression is used when we want to include additional dimensions in the aggregation while considering the current filters. For example, calculating average sales per product category within the filtered view. The formula would be: INCLUDE [Product Category]: AVG([Sales]).

25. How would you use Tableau's LOOKUP function in a table calculation to compare current values to previous values?

Answer: The LOOKUP function is used to access data from previous or subsequent rows in a table calculation. To compare current values to previous values, we can create a calculated field using: SUM([Sales]) - LOOKUP(SUM([Sales]), -1). This formula calculates the difference between the current sales value and the sales value from the previous row. For example, if sales in the previous month were 1000 and the current month's sales are 1200, the result would be 200.

26. How can you use Tableau's DATEPART function to analyze sales by month, quarter, or year?

Answer: The DATEPART function extracts specific parts of a date for analysis. To analyze sales by month, quarter, or year, we can use DATEPART to group and aggregate data accordingly. For example:

- To analyze by month: DATEPART('month', [Order Date])
- To analyze by quarter: DATEPART('quarter', [Order Date])
- To analyze by year: DATEPART('year', [Order Date])

These functions help aggregate sales data into the desired time periods for trend analysis and reporting.

27. How would you create a rolling average in Tableau to analyze trends over time?

Answer: To create a rolling average in Tableau, use a table calculation with the WINDOW_AVG function. For example, to calculate a 3-month rolling average, create a calculated field with the formula: WINDOW_AVG(SUM([Sales]), -2, 0). This formula computes the average of sales over the current month and the previous two months. The rolling average smooths out fluctuations and highlights trends over time.

28. Explain how to use the RANK function in Tableau to rank items within a specific category.

Answer: The RANK function assigns a rank to each item within a specified category based on a measure. For example, to rank products by sales within each category, use the formula: RANK(SUM([Sales]), 'desc'). This formula ranks products in descending order of their sales. we can also partition the ranking by category by including the category dimension in the calculation, ensuring that ranks are calculated separately for each category.

29. Given the following dataset, how would you use a WINDOW_AVG function to calculate a 7-day moving average of sales?

Date	Sales
2024-08-01	100
2024-08-02	120
2024-08-03	130
2024-08-04	110
2024-08-05	140
2024-08-06	150
2024-08-07	160

Answer: To calculate a 7-day moving average of sales, create a calculated field with the formula: WINDOW_AVG(SUM([Sales]), -6, 0). This formula calculates the average of sales over the past 7 days, including the current day. The moving average smooths out daily fluctuations and provides a clearer trend of sales performance over a week.

30. How would you use the PREVIOUS_VALUE function to create a cumulative sum of sales in Tableau?

Answer: The PREVIOUS_VALUE function is used to create a running total or cumulative sum by referring to the previous value in the calculation. To create a cumulative sum of sales, use the formula: PREVIOUS_VALUE(0) + SUM([Sales]). This formula starts with an initial value of 0 and adds the sales value of the current row to the cumulative total of the previous rows. The result will be a running total of sales over the dataset.

31. How would you handle performance issues with large datasets in Tableau?

Answer: To handle performance issues with large datasets in Tableau, consider several strategies:

- Extracts: Use data extracts instead of live connections to improve performance by pre-aggregating and optimizing data.
- Aggregation: Aggregate data at a higher level to reduce the volume of data processed.
- Indexing: Create indexes on key fields to speed up query performance.
- Filters: Apply filters to limit the amount of data loaded and displayed.
- Performance Recording: Use Tableau's performance recording tools to identify and address specific bottlenecks.

32. Explain the concept of incremental data refresh and how it can be implemented in Tableau.

Answer: Incremental data refresh in Tableau involves updating only the new or changed records in a data extract, rather than refreshing the entire dataset. This approach improves performance and efficiency. To implement incremental data refresh, configure the extract to use an incremental refresh option by specifying a column that indicates new or updated records (e.g., a date or ID column). Tableau will then update the extract with only the new or modified records, reducing the load and processing time compared to a full refresh.