knowledgehut
upGrad

Introduction to Tableau

Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.

By the end of this session, you will:

- Understand visualizations' importance and Tableau's capabilities.
- Navigate the Tableau interface and work with different file types.
- Comprehend dimensions, measures, data types, aggregation, and granularity.
- Utilize marks card for visualization customization.
- Apply the "Show me" feature for selecting appropriate visualizations.

Key Takeaways from This Session

Tableau Proficiency: Visualize datasets efficiently.

Data Visualization Expertise: Convey insights persuasively.

Efficient Data Analysis: Simplify data exploration.

Visualization Customization: Create impactful visuals.

Pop Quiz

Q. Who is the parent company of Tableau?

A. Salesforce

B. Workday

C. Microsoft

D. Google

Pop Quiz

Q. Who is the parent company of Tableau?

A. Salesforce

B. Workday

C. Microsoft

D. Google

Introduction to Tableau

What is Tableau?

History and Evolution:

- Founded in 2003 by Chris Stolte, Christian Chabot, and Pat Hanrahan.
- Rapidly evolved into a leading data visualization and business intelligence software.
- Acquired by Salesforce in 2019, reinforcing its position in the analytics market.

Purpose/Problem it Solves:

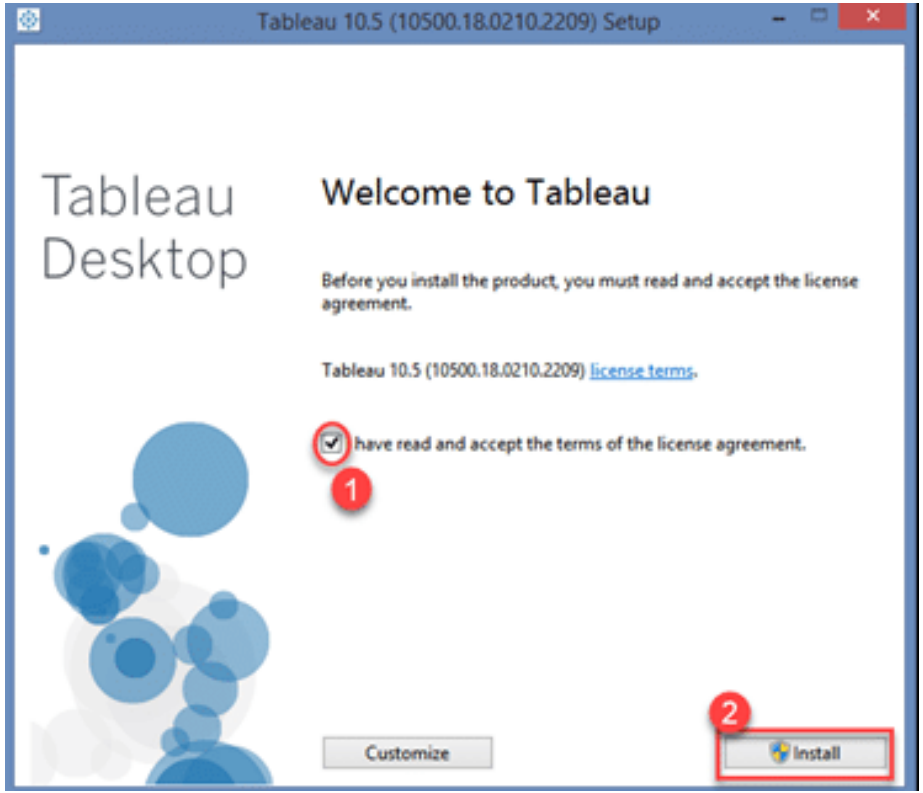
- Tableau bridges the gap between complex data and meaningful insights.
- Enables users to create interactive visualizations without requiring extensive technical expertise.
- Empowers data-driven decision-making for organizations of all sizes.

What is Tableau?

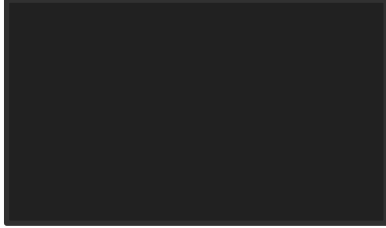
Overview of Tableau's Capabilities and Features:

- **Intuitive Drag-and-Drop Interface:** Effortlessly build visualizations from diverse datasets.
- **Powerful Data Connectivity:** Seamlessly connect to various data sources, both cloud-based and on-premises.
- **Dynamic Dashboards:** Create interactive dashboards for real-time exploration.
- **Rich Visualization Options:** Choose from an array of charts, graphs, and maps for compelling representations.
- **Ad Hoc Analysis:** Analyze data on-the-fly with instant responses to questions.
- **Robust Analytics:** Perform complex calculations and statistical analysis effortlessly.

Installation of Tableau



Perks of using Tableau



- **Ease of Use**: Tableau offers a user-friendly interface with drag-and-drop functionality, enabling users of all levels to create compelling visualizations without extensive technical expertise.
- **Fast Data Visualization**: Tableau's efficient processing and real-time data connectivity allow users to quickly analyze and visualize data, accelerating the decision-making process.
- **Versatile Data Sources**: Tableau supports a wide range of data sources, including spreadsheets, databases, cloud services, and big data platforms, ensuring seamless integration and analysis of diverse datasets.
- **Interactivity and Collaboration**: Tableau empowers users to build interactive dashboards and share insights with others, promoting collaboration and data-driven discussions.
- **Advanced Analytics Capabilities**: With integration into statistical tools and support for complex calculations, Tableau enables advanced analytics and predictive modeling for data-driven insights.

Pop Quiz

Q. What is the Tableau Online?

A. A desktop application for creating and publishing Tableau visualizations

B. A cloud-based platform for sharing and collaborating on Tableau visualizations

C. A mobile app for creating and viewing Tableau visualizations

D. A programming language for data analysis and visualization

Pop Quiz

Q. What is the Tableau Online?

A. A desktop application for creating and publishing Tableau visualizations

B. A cloud-based platform for sharing and collaborating on Tableau visualizations

C. A mobile app for creating and viewing Tableau visualizations

D. A programming language for data analysis and visualization

Product Types and File Types

What is Included in Tableau Product Suite

- **Tableau Desktop**: This is the primary authoring and development tool used to create interactive visualizations, dashboards, and reports. Tableau Desktop allows users to connect to various data sources, transform data, and design compelling data visualizations.

- **Tableau Server**: Tableau Server is a web-based platform that allows organizations to share and collaborate on Tableau workbooks and dashboards securely. It provides central management and governance capabilities, ensuring controlled access to data and maintaining data security.

- **Tableau Online**: Similar to Tableau Server, Tableau Online is a cloud-based platform that offers similar collaboration and sharing capabilities. It allows users to publish Tableau dashboards to the Tableau Online server for easy accessibility from anywhere with an internet connection.

What is Included in Tableau Product Suite

- **Tableau Public**: Tableau Public is a free cloud-based platform where users can share interactive data visualizations and dashboards with the public. It is designed for public data storytelling and allows users to publish visualizations to the Tableau Public server for public consumption.

- **Tableau Reader**: Tableau Reader is a free desktop application that allows users to view and interact with Tableau packaged workbooks (.twbx files) created in Tableau Desktop. However, it does not support data editing or sharing, making it suitable for individual viewing.

File Types in Tableau

Tableau Workbook (.twb):
- Interactive and editable file containing visualization elements, data connections, and metadata.
- Used with Tableau Desktop for creating and saving visualizations.

Tableau Packaged Workbook (.twbx):
- A compressed file that includes both the .twb workbook and its dependent data sources.
- Facilitates easy sharing of visualizations with others, ensuring data consistency.

Tableau Data Source (.tds):
- A connection file that stores the connection information to a specific data source.
- Allows users to share data connections without including the actual data.

File Types in Tableau

Tableau Data Extract (.tde):
- A compressed file that stores a subset or snapshot of data from the original data source.
- Enables faster data access and improves performance for large datasets.

Tableau Data Engine File (.tde, .hyper):
- A high-performance columnar database engine used for data storage and query execution.
- Enhances data processing speed for large-scale datasets.

Tableau Bookmark (.tbm):
- A file that stores a specific view or state of a Tableau visualization.
- Allows users to save and share specific data insights with others.

Demo – Walkthrough of the Tableau Interface

Pop Quiz

Q. The Tableau Product Suite consists of _____.

A. Tableau Desktop, Tableau Public, Tableau Online, Tableau Server, Tableau Measure

B. Tableau Desktop, Tableau Public, Tableau Online, Tableau Server, Tableau Reader

C. Tableau Desktop, Tableau Public, Tableau Dimension, Tableau Server, Tableau Reader

D. Tableau Desktop, Tableau Public, Tableau Online, Tableau Server, Tableau Organic

Pop Quiz

Q. The Tableau Product Suite consists of _____.

A. Tableau Desktop, Tableau Public, Tableau Online, Tableau Server, Tableau Measure

B. Tableau Desktop, Tableau Public, Tableau Online, Tableau Server, Tableau Reader

C. Tableau Desktop, Tableau Public, Tableau Dimension, Tableau Server, Tableau Reader

D. Tableau Desktop, Tableau Public, Tableau Online, Tableau Server, Tableau Organic


Q&A

Understanding Tableau Building Blocks

Dimensions and Measures

Dimensions:

- Dimensions are qualitative or categorical data fields in Tableau.
- They represent the descriptive attributes or characteristics of the data.
- Examples: Names, categories, dates, geographical locations.

Characteristics:

- Discrete: Represents distinct and separate values.
- Can be used to categorize data and create groups.
- Placed on the Rows or Columns shelf to define the level of granularity.

Measures:

- Measures are quantitative or numerical data fields in Tableau.
- They represent the measurable and numerical aspects of the data.
- Examples: Sales, revenue, profit, quantity, temperature.

Characteristics:

- Continuous: Represents a range of numeric values.
- Can be aggregated using functions like SUM, AVG, COUNT, etc.
- Placed on the Rows, Columns, or Marks shelf to visualize and analyze numeric data.

Understanding Auto Generated Fields

Suppose you have a dataset containing information about sales transactions in an e-commerce store. The dataset includes the following columns: Order ID, Product Category, Product Name, Quantity, Unit Price, and Total Price.

Auto-Generated Fields in Tableau:

Number of Records (Number of Records):

- This auto-generated field will provide the count of records in the dataset.

- In this example, if you have 100 sales transactions, the Number of Records field will have a constant value of 100 for each data point.

Understanding Auto Generated Fields

Auto-Generated Fields in Tableau:

Measure Names (Measure Names):
- The Measure Names auto-generated field will list all numeric measures available in the dataset, which in this case are Quantity, Unit Price, and Total Price.
- It enables users to create dynamic views or dashboards that automatically adjust based on the selected measures.

Measure Values (Measure Values):
- The Measure Values auto-generated field contains the numeric values of all the measures in the dataset, which are Quantity, Unit Price, and Total Price.
- By using this field, users can create a single view that simultaneously displays the quantities, unit prices, and total prices of products in the e-commerce store.

Usage of Auto Generated Fields

- Using the auto-generated fields in Tableau, you can create a dashboard that displays the total sales revenue by product category.
- You can place the Product Category dimension on the Rows shelf and the Total Price measure on the Columns shelf.
- Then, you can use the Measure Names filter to allow users to switch between different measures (Quantity, Unit Price, Total Price) and dynamically view the sales performance of products.

Different Data Types in Tableau

String (Text):
- Represents text values or alphanumeric characters.
- Examples: Names, addresses, product descriptions.

Integer (Whole Number):
- Represents whole numbers without decimals.
- Examples: Number of units sold, quantity of items.

Float (Decimal Number):
- Represents numbers with decimal points.
- Examples: Prices, monetary values, percentages.

Date:
- Represents calendar dates (year, month, day).
- Examples: Order dates, registration dates, event dates.

Different Data Types in Tableau

String (Text):
- Represents text values or alphanumeric characters.
- Examples: Names, addresses, product descriptions.

Integer (Whole Number):
- Represents whole numbers without decimals.
- Examples: Number of units sold, quantity of items.

Float (Decimal Number):
- Represents numbers with decimal points.
- Examples: Prices, monetary values, percentages.

Different Data Types in Tableau

Date:
- Represents calendar dates (year, month, day).
- Examples: Order dates, registration dates, event dates.

Date Time:
- Represents dates and times together, including hours, minutes, and seconds.
- Examples: Timestamps, log entries.

Different Data Types in Tableau

Boolean (Logical):
- Represents binary values (True or False).
- Examples: Yes/No responses, binary indicators.

Geographical (Spatial):
- Represents geographical data, such as latitude and longitude coordinates.
- Examples: Geographic locations, map data.

Currency:
- Represents monetary values with specific currency symbols.
- Examples: Dollar amounts, Euro amounts.

Different Data Types in Tableau

Percentage:
- Represents values in percentage form.
- Examples: Growth rates, profit margins.

Ordinal:
- Represents categorical data with a specific order.
- Examples: Ratings (1 to 5), ranking categories.

BLOB (Binary Large Object):
- Represents binary data or large data objects, like images or documents.

Pop Quiz

Q. Which one of the following is not a valid data type in Tableau?

A. String

B. Integer

C. Boolean

D. Complex

Pop Quiz

Q. Which one of the following is not a valid data type in Tableau?

A. String

B. Integer

C. Boolean

D. Complex

Understanding Discrete and Continuous Fields

In Tableau, data fields are categorized into two main types: discrete and continuous. Understanding the distinction between these field types is crucial for data analysis and visualization.

Discrete Fields:

- Discrete fields represent data that is categorical or qualitative in nature.
- They have distinct, separate values and are used for grouping and creating categories.
- Examples: Country names, product categories, gender, and any data that is not continuous.

Characteristics of Discrete Fields:

- The data points are distinct and separate, such as individual categories or labels.
- Discrete fields are typically used for dimensions in Tableau, providing the context for data analysis.
- Discrete data can be represented in bar charts, pie charts, and other discrete visualizations.

Understanding Discrete and Continuous Fields

Continuous Fields:
- Continuous fields represent data that is numeric or quantitative in nature.
- They have a range of values and can take any numeric value within that range.
- Examples: Sales revenue, temperature, time, and any data that represents a continuous measurement.

Characteristics of Continuous Fields:
- The data points are numeric and can be any value within a range, allowing for smooth transitions between points.
- Continuous fields are typically used for measures in Tableau, enabling mathematical calculations and aggregations.
- Continuous data is often visualized using line charts, area charts, and other continuous visualizations.

Understanding Discrete & Continuous Fields

Usage and Impact:

- Understanding the distinction between discrete and continuous fields in Tableau is vital for choosing appropriate visualization types, performing correct aggregations, and accurately interpreting data.

- Using the correct field type ensures that the visualizations represent the data accurately and facilitate meaningful insights for data-driven decision-making.

Introduction to Aggregation and Granularity

Aggregation in Tableau refers to the process of summarizing or combining multiple data points into a single value.

It involves applying a mathematical function (e.g., SUM, AVG, COUNT) to a set of data to obtain a higher-level overview or insight.

Purpose of Aggregation:

- Aggregation helps in simplifying and understanding large datasets by presenting essential information in a concise format.
- It allows users to analyze trends, patterns, and general characteristics of the data without dealing with individual data points.

Examples of Aggregation:

- Summing up sales revenue to find total revenue for a specific period.
- Calculating the average temperature for a month using daily temperature readings.
- Counting the number of orders to determine the total number of transactions.

Introduction to Aggregation and Granularity

Granularity in Tableau refers to the level of detail or the degree of data aggregation in a visualization or dataset.

It defines the level at which individual data points are represented or how finely data is divided.

Purpose of Granularity:

- Granularity allows users to control the level of detail in visualizations and analysis, helping focus on specific insights.
- It enables users to view data at different levels, from the most detailed (fine-grained) to the most summarized (coarse-grained).

Examples of Granularity:

- **Fine-grained granularity**: Visualizing sales data at the individual transaction level, showing each sale separately.
- **Coarse-grained granularity**: Visualizing sales data aggregated at the monthly level, displaying total sales for each month.

Introduction to the Mark Cards

Marks Cards in Tableau are a fundamental part of the Tableau interface used to customize the appearance and behavior of individual data points (marks) in a visualization.

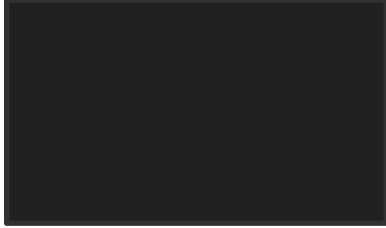
Each data point in a visualization, such as a bar, a point on a line, or a slice in a pie chart, is represented as a mark.

Purpose of Marks Cards:

- Marks Cards allow users to control how data is represented visually, enabling precise customization for better communication of insights.

- By adjusting properties on Marks Cards, users can change colors, sizes, labels, shapes, and other visual attributes of marks.

Customization Option on Mark Cards



- **Color**: Change the color of marks based on data values or specific conditions, highlighting patterns or differences.
- **Size**: Modify the size of marks to emphasize significance or create visual hierarchy.
- **Label**: Display data labels on marks to provide additional context or show specific values.
- **Shape**: Use different shapes for marks to differentiate categories or add unique visual elements.
- **Tooltip**: Configure tooltips to show data details when hovering over marks, enhancing interactivity.

Impact on Visualizations:

- Marks Cards allow users to tailor visualizations to fit specific data storytelling needs and effectively communicate insights.
- Fine-tuning visual attributes on Marks Cards enhances the clarity and impact of visualizations, making them more engaging and informative.

Pop Quiz

Q. What is the difference between a discrete and a continuous field in Tableau?

A. A discrete field represents categorical data, while a continuous field represents numeric data

B. A discrete field represents numeric data, while a continuous field represents categorical data

C. A discrete field is represented by individual data points, while a continuous field is represented by a continuous range of values

D. There is no difference between a discrete and a continuous field in Tableau

Pop Quiz

Q. What is the difference between a discrete and a continuous field in Tableau?

A. A discrete field represents categorical data, while a continuous field represents numeric data

B. A discrete field represents numeric data, while a continuous field represents categorical data

C. A discrete field is represented by individual data points, while a continuous field is represented by a continuous range of values

D. There is no difference between a discrete and a continuous field in Tableau

Demo – Tableau Show Me

Demo – Create a Simple Visualization using a Dataset

Poll Time

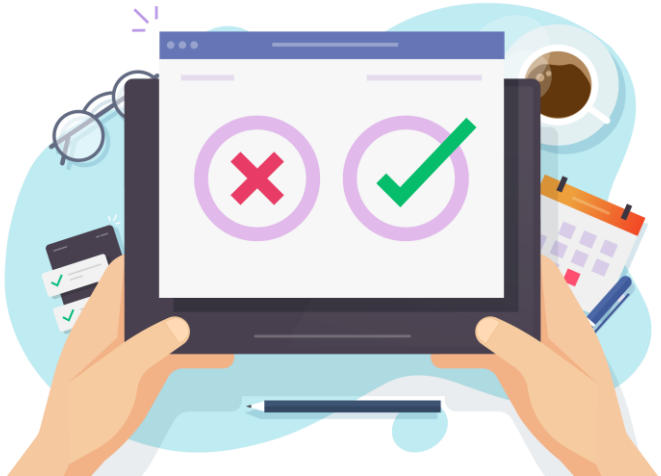
Q. Which one of the following is not a valid aggregation function in Tableau?

A. Sum

B. Count

C. Average

D. Maximize



Poll Time

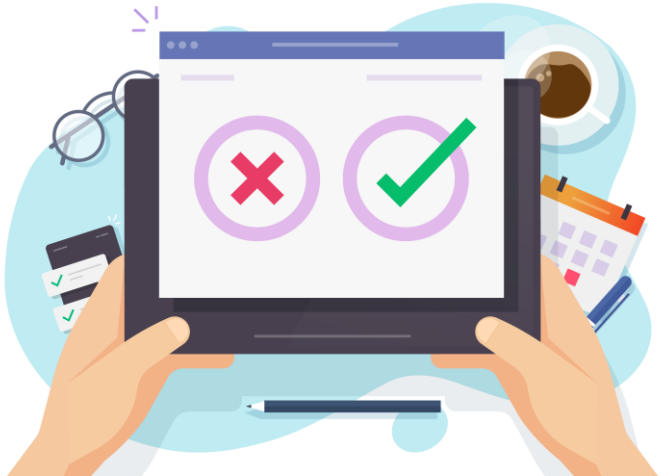
Q. Which one of the following is not a valid aggregation function in Tableau?

A. Sum

B. Count

C. Average

D. Maximize




Q&A

Summary

- Explored the capabilities of Tableau, a robust data visualization and business intelligence software.
- Familiarized yourself with the Tableau Product Suite, including Tableau Desktop, Server, Online, and Mobile.
- Discovered the convenience of Auto-Generated Fields, simplifying data exploration and analysis.

Summary

- Differentiated between discrete and continuous Fields to represent data accurately in visualizations.
- Utilized Aggregation and Granularity to summarize data and control visualization detail.
- Customized visual elements with the Marks Card and leverage the "Show Me" feature to choose appropriate visualizations.

Activity 1

Pre-requisites: A small dataset containing sales data, including product categories and corresponding sales revenue.

Scenario: You are a data analyst for a retail company, and your task is to create a simple bar chart using Tableau to visualize the sales performance of different product categories.

Expected Outcome: The final result should be a bar chart showing the total sales revenue for each product category. The chart should be labeled appropriately and visually appealing.

Activity 1

Steps:

a. Open Tableau and connect to the sales dataset.

b. Drag the "Date" field to the Columns shelf and the "Sales Revenue" field to the Rows shelf.

c. Tableau may automatically aggregate the data by date. If not, right-click on the "Date" field, choose "Measure," and select "Sum" to aggregate the sales revenue.

d. Tableau should create a line chart showing the sales revenue over time.

e. Add a date filter to enable users to choose a specific time period for analysis.

f. Customize the chart by adding appropriate labels, axis titles, and a chart title for clarity.

g. Test the line chart by applying different date filters to observe sales trends over different periods.

Next Session:
Deep Dive into Tableau

THANK YOU

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

knowledgehut
upGrad

knowledgehut
upGrad

Deep Dive into Tableau

Pre-requisites

Hope you have gone through the self-learning content for this session on the PRISM portal.

By the End of This Session, You Will:

- Connect Tableau to Excel, CSV, and PDF files, as well as SQL databases using common file-based connections and techniques.
- Understand and navigate the Tableau Metadata Grid to explore data sources and metadata effectively.
- Apply Data Interpreter to clean and structure messy data for analysis.
- Use Pivot and Split Fields to reshape data for better visualization and analysis.
- Combine data from multiple sources using Joins and Cross Database Joins to create comprehensive analyses.
- Master the usage of Union and Tableau Data Extracts to optimize data sources and enhance performance.
- Apply Data Source Filters to refine data for targeted analysis.

Poll Time

Q. What are the file extensions in Tableau?

A. Tableau Workbook (.twb)

B. Tableau Packaged Workbook (.twbx)

C. Tableau Data Source(.tds)

D. All of the listed

Poll Time

Q. What Are The File Extensions In Tableau ?

A. Tableau Workbook (.twb)

B. Tableau Packaged Workbook (.twbx)

C. Tableau Data Source(.tds)

D. All of the listed

Key Takeaways from This Session

Learning to manage various data connections (Excel, CSV, PDF, SQL databases) equips data scientists with the skills to handle diverse data sources efficiently.

Understanding data preparation techniques such as Data Interpreter, Pivot, and Split Fields enables data scientists to clean and transform raw data effectively.

Proficiency in using Union and Tableau Data Extracts enables data scientists to optimize data sources and enhance performance during analysis and visualization.

Applying Data Source Filters enables data scientists to refine data for specific analyses, eliminating irrelevant or sensitive information.

Managing Data Connections

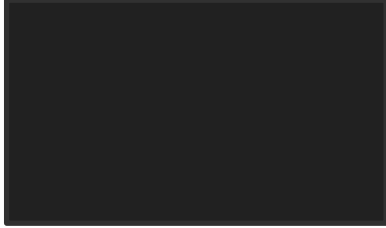
Importance of Data Connections in Tableau

- **Data Accessibility:** Connect to diverse data sources (Excel, CSV, databases) for comprehensive analysis.
- **Real-Time Analysis:** Perform real-time analysis with live data connections for up-to-date insights.
- **Data Integration:** Blend data from different sources to create holistic analyses.
- **Data Transformation:** Use Data Interpreter for efficient data preparation and cleansing.

Importance of Data Connections in Tableau

- **Performance Optimization:** Optimize data sources with Tableau Data Extracts for faster query times.
- **Enhanced Data Exploration:** Access metadata and data details for better data exploration.
- **Cross-Functional Insights:** Foster collaboration by visualizing data from multiple sources.
- **Data Security:** Ensure data security and governance with Tableau's robust features.

Overview of Tableau Metadata Grid



- The Tableau Metadata Grid is a powerful feature that provides users with a comprehensive view of the underlying data structure and metadata associated with the data sources used in Tableau.

- It offers valuable insights into the data source, enabling users to better understand the data and efficiently navigate through complex datasets.

Key Elements of Tableau Metadata Grid

Tables and Fields: The Metadata Grid displays a list of tables and fields present in the connected data source. It presents a clear hierarchy of tables, allowing users to explore the data schema.

Data Types: For each field, the Metadata Grid provides information about the data type, such as string, integer, date, etc. Understanding data types is crucial for accurate data analysis and visualization.

Aggregation: The Metadata Grid indicates whether each field is aggregated or not. This helps users identify how Tableau will handle the field in calculations and visualizations.

Key Elements of Tableau Metadata Grid

Data Source Filters: The Metadata Grid displays data source filters, allowing users to identify any filters applied to the data source. This helps in understanding data restrictions and potential data quality concerns.

Joins and Relationships: Users can view information about joins and relationships between tables, which helps in comprehending how the data is related and blended within Tableau.

Aliasing: The Metadata Grid allows users to alias table names and field names to provide more user-friendly labels in the Tableau interface.

Demo – Connecting to Common File-based Sources

Pop Quiz

Q. What is a data source in Tableau?

A. A way to group related dimensions together

B. A way to limit the data displayed in a chart or table

C. A way to connect to and access data from external sources

D. A way to create a custom calculation in a chart or table

Pop Quiz

Q. What is a data source in Tableau?

A. A way to group related dimensions together

B. A way to limit the data displayed in a chart or table

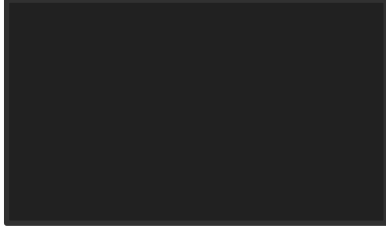
C. A way to connect to and access data from external sources

D. A way to create a custom calculation in a chart or table


Q&A

Preparing Data for Use

Introduction to the Data Interpreter



- The Data Interpreter is a powerful data preparation feature in Tableau that automatically detects and resolves common data issues when connecting to external datasets.
- It streamlines the process of cleaning and structuring data, ensuring that analysts can work with clean and accurate data for analysis and visualization.

Key Features of Data Interpreter

- **Automated Data Cleaning:** The Data Interpreter automatically identifies and cleans common data anomalies, such as extra headers, footers, and irregular spacing, to create a consistent and structured dataset.
- **Handling Data Unions:** When working with datasets that require union (combining similar tables vertically), the Data Interpreter handles varying numbers of columns in each table, making the process seamless.
- **Skip Rows and Columns:** The Data Interpreter intelligently skips irrelevant rows and columns, avoiding unnecessary data inclusion in the analysis.
- **Handling Data Type Inconsistencies:** It detects and corrects inconsistencies in data types, ensuring that numeric fields are treated as numbers and textual fields as strings.
- **Data Preview and Correction:** Users can preview the data before applying the Data Interpreter and make manual corrections if needed.

Joints in Tableau

In Tableau, "joins" refer to the process of combining data from two or more tables based on a common field. It allows users to create a unified dataset that includes information from multiple tables, facilitating more comprehensive analysis and visualization. Tableau supports various types of joins, each serving a specific purpose.

Use Case:
Imagine you have two tables, one containing customer information (Customer Table) and the other with sales data (Sales Table). Both tables have a common field, "CustomerID," which uniquely identifies each customer. By performing an inner join on "CustomerID," you can create a combined dataset that includes customer information along with their sales data. This unified dataset enables deeper analysis, such as identifying customer segments with high sales or customer demographics influencing purchase behavior.

Types of Joints in Tableau

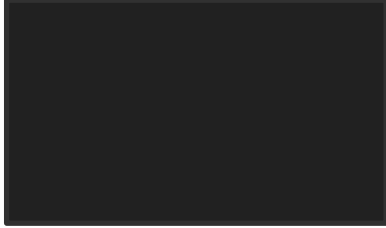
- **Inner Join:** An inner join returns only the rows that have matching values in both tables. It eliminates rows that do not have a corresponding match in the other table. This type of join is used to find common data between tables.

- **Left Join (Left Outer Join):** A left join returns all the rows from the left (primary) table and matching rows from the right (secondary) table. If there is no match in the right table, the result will show null values for the right table columns.

- **Right Join (Right Outer Join):** A right join returns all the rows from the right (secondary) table and matching rows from the left (primary) table. If there is no match in the left table, the result will show null values for the left table columns.

- **Full Outer Join:** A full outer join returns all the rows from both tables and matches them where possible. It includes rows from both tables, even if there are no matches in the other table, filling in null values for unmatched data.

Introduction to Cross-database Clients



Cross-database clients in Tableau refer to the capability of connecting to and analyzing data from multiple data sources simultaneously. This feature empowers Tableau users to combine and blend data from different database platforms, such as SQL Server, Oracle, MySQL, and more, within a single Tableau workbook.

Use Case:
Suppose a retail company stores sales data in an SQL Server database and customer data in an Oracle database. By using Tableau's cross-database clients, analysts can connect to both databases simultaneously. They can then combine customer data with sales data to analyze customer buying behavior, identify high-value customers, and gain insights into sales patterns based on customer demographics.

Poll Time

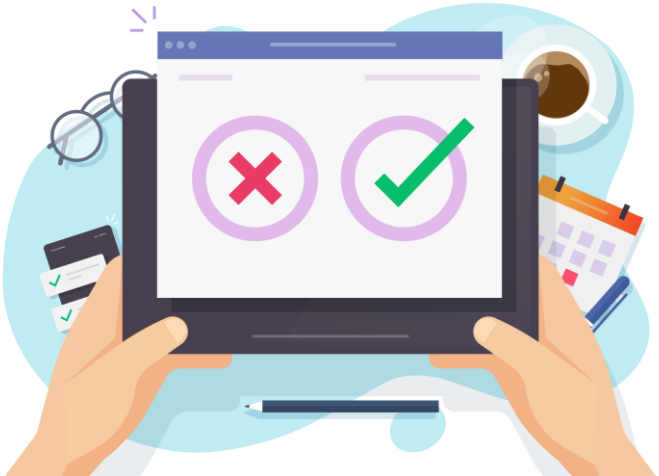
Q. What type of join is used in blending?

A. Right join

B. Left join

C. Full join

D. Inner join



Poll Time

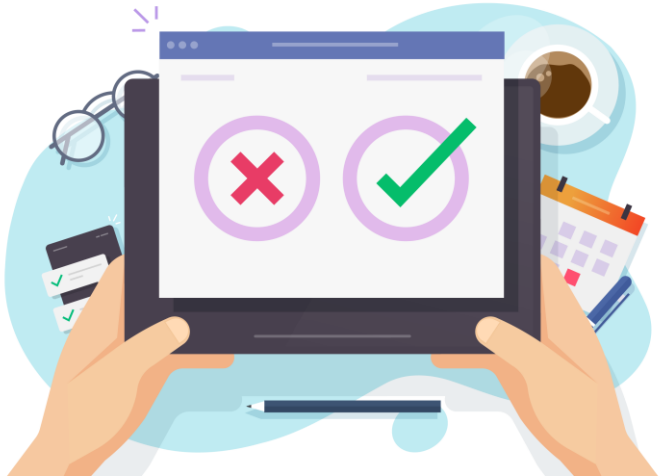
Q. What type of join is used in blending?

A. Right join

B. Left join

C. Full join

D. Inner join



Demo – Data Preparation Techniques

Demo – Applying Data Source Filters

Pop Quiz

Q. How do you find the field is continuous in tableau?

A. # symbol

B. Blue color

C. Green color

D. None of the listed

Pop Quiz

Q. How do you find the field is continuous in tableau?

A. # symbol

B. Blue color

C. Green color

D. None of the listed


Q&A

Summary

- Connected Tableau to various data sources, including Excel, CSV, PDF files, and SQL databases, expanding their data analysis capabilities.
- Navigated the Tableau Metadata Grid, gaining insights into data sources, tables, fields, and relationships, enhancing data exploration proficiency.
- Automated data cleaning and structuring, ensuring accurate data for analysis and visualization.
- Mastered data preparation techniques, including pivot, split fields, and joins (including cross-database joins) for seamless data blending and integration.
- Employed unions and Tableau Data Extracts to optimize data sources, enhancing performance during analysis and visualization.

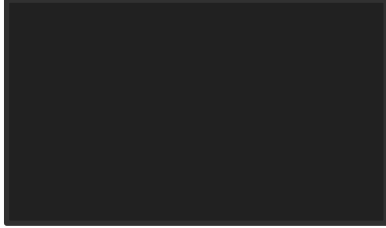
Activity 1

Pre-requisites:
Students should have access to a dataset containing sales data, including product categories, sales revenue, and dates.

Scenario:
Imagine you are working as a data analyst for an e-commerce company. Your task is to prepare the sales data for analysis and visualization in Tableau.

Expected Outcome:
The final result should be a cleaned and structured dataset ready for analysis in Tableau. The dataset should include appropriate field types and no duplicate or irrelevant data.

Activity 1



Steps:

- Use Data Interpreter in Tableau to clean the dataset, resolving any common data issues like extra headers or footers.
- Identify and handle any missing or null values in the dataset.
- Convert data types to appropriate formats, ensuring numeric fields are recognized as numbers, and dates are properly formatted.
- Pivot the data to create a tidy dataset suitable for analysis, with sales revenue and product categories as separate fields.
- Save the cleaned dataset as a Tableau Data Extract for better performance during analysis.

Session Feedback

Next Session: Introduction to Data Visualization using Tableau

THANK YOU

Please complete your assessments and review the self-learning content for this session on the **PRISM** portal.

knowledgehut
upGrad