Branch : MBA

Semester : 3rd

Subject Name : DATA VISUALIZATION

Subject Code : BAY-1

Faculty Name : AVI BRAMHBHATT

Index

UNIT – 1

GETTING STARTED WITH TABLEAU

1 Introduction – Getting Started with Tableau

1.1 What is Tableau?

Tableau is a powerful **Business Intelligence (BI)** and **Data Visualization** tool that enables users to analyze and represent data visually through charts, graphs, and dashboards.
Its primary purpose is to help individuals and organizations **"see and understand their data."**

Instead of relying on static spreadsheets or complex programming, Tableau allows **interactive exploration** of data through a simple **drag-and-drop interface**.

1.1.2 Purpose and Importance of Tableau

Tableau bridges the gap between **raw data** and **decision-making**. It transforms complex numerical tables into intuitive visuals such as bar graphs, line charts, heat maps, and geographic maps.

Feature	Explanation / Benefit
Ease of use	Drag-and-drop interface requires no coding knowledge.
Fast analysis	Processes large datasets quickly.
Data connectivity	Connects to Excel, databases, and online data sources.
Interactive visualization	Enables filters, tooltips, and dashboards for dynamic analysis.
Sharing	Dashboards can be shared via Tableau Server, Tableau Online, or Tableau Public.

1.1.3 Tableau Interface – First Look

When Tableau Desktop opens, you see:

- **Start page:** Connects to data (Excel, CSV, SQL, etc.)
- **Data pane:** Lists all fields (Dimensions & Measures).
- **Shelves:** Rows, Columns, Filters, Marks, etc.

- **View area (Canvas):** Where the visualization is displayed.
- **Show Me panel:** Suggests chart types based on selected data.

1.1.4 Basic Workflow in Tableau

1. **Connect to Data** – import from Excel or database.
2. **Understand Fields** – identify dimensions and measures.
3. **Drag & Drop** – place fields on Rows/Columns shelves.
4. **Choose Visualization Type** – bar, line, pie, etc.
5. **Refine View** – add colors, filters, labels, or tooltips.
6. **Save / Share** – publish dashboard or workbook.

1.2 Dimensions vs Measures

Criteria	Dimensions	Measures
Type of Data	Contain **Qualitative** data	Contain **Quantitative** data
Nature	**Independent** variables	**Dependent** variables
Aggregation	Cannot be directly **aggregated**	Can be **aggregated**
Position in Data Pane	They occupy the **Top half**	They occupy the **Bottom half**
Visual Representation	Most of the time they are **Discrete** fields, they create **Headers/ Labels** in a view and are displayed as **Blue** pills	Most of the time they are **Continuous** fields, they create **Axis/ Axes** in a view and are displayed as **Green** pills
Role in Analysis	Define the **WHO, WHAT, WHERE & WHEN** of the data	Define the **HOW MUCH** of the data
Typical Examples	*Customer Name (Name), Order ID (IDs), Country (Geographical)*	*Sales, Profit (Numeric Values), Profit Ratio (Ratios), Discount (Average or Percentages)*

1.2.1 Concept

All data fields used in Tableau belong to one of two main categories:

	Dimensions	Measures
Nature	Qualitative (descriptive)	Quantitative (numeric)
Purpose	Describe, group, or categorize data	Perform mathematical calculations
Aggregation	Not aggregated	Always aggregated (e.g., SUM, AVG)
Example	Region, Product Name, Category	Sales, Profit, Quantity

	Dimensions	Measures
Effect in Visualization	Defines level of detail (what we are analyzing)	Defines numeric value (how much)

1.2.2 Detailed Explanation

- **Dimensions** are *discrete attributes*—they divide the data into groups or categories.
  - Example: "Region" divides the dataset into East, West, North, and South.
- **Measures** are *quantitative metrics* that Tableau can aggregate (add, average, etc.).
  - Example: "Sales" or "Profit" is aggregated per region.

1.2.3 Example in Practice

Dataset snippet:

Region	Product	Sales	Profit
East	Laptop	30000	2000
West	Printer	20000	1200
South	Tablet	15000	800

In Tableau:

- **Region** → Dimension (x-axis)
- **Sales** → Measure (y-axis)
  → Bar chart of total Sales per Region.

1.2.4 Importance of Distinguishing Between Them

- Ensures correct **aggregation** and **visual type**.
- Avoids misinterpretation (e.g., average profit vs. total profit).
- Provides **granularity control**: you can view data by Region, then by Product Category, etc.

1.3 Discrete vs Continuous Fields

Every field in Tableau—whether a Dimension or a Measure—is also classified as **Discrete** or **Continuous.**

Discrete vs Continuous in Tableau Software

DISCRETE FIELD	CONTINUOUS FIELD
Individually Separate	Unbroken Whole
Represented by **Blue** pill	Represented by **Green** pill
Can filter individual elements	Can filter only by range
Countable	Measurable
Becomes header in a view	Becomes axis in a view
Brings level of detail (or detail) to view	Brings aggregate to view
Can have hierarchy	Cannot have hierarchy
Can be sorted	Cannot be sorted

1.3.1 Discrete (Blue Pills)

- Contain **separate, distinct values**.
- Represent categories or labels.
- Create **headers** in a view (no numeric axis).
- Displayed as **blue** fields in Tableau.

Examples

- Region → East, West, North, South
- Category → Furniture, Technology, Office Supplies
- Order ID, Customer Segment

Visualization Example:
A *bar chart* where each bar represents one Region is created using a **discrete** field.

1.3.2 Continuous (Green Pills)

- Represent **numerical or temporal ranges**.
- Have infinite possible values within an interval.
- Create **axes** (with continuous scale).

- Displayed as **green** fields in Tableau.

Examples

- Sales (₹ 0 → ₹ ∞)
- Profit % (–100 → 100)
- Order Date (used as a timeline)

Visualization Example:
A *line chart* showing "Sales over Time" uses Order Date as a **continuous** field.

1.3.3 Difference Summary

Feature	Discrete Field	Continuous Field
Color Code	Blue	Green
Values	Distinct, separate	Range, uninterrupted
Displayed As	Headers	Axes
Common Use	Category comparison	Trend or distribution
Example Field	Region	Sales, Order Date

1.3.4 Conceptual Understanding

- Discrete = **"Countable"** (e.g., number of regions).
- Continuous = **"Measurable"** (e.g., amount of sales).

(For visual reference, search "Tableau discrete continuous color chart" online; it shows blue vs green field examples.)

1.4 Application of Discrete and Continuous Fields

1.4.1 Using Discrete Fields

Scenario: A company wants to know how sales vary by Region.

- Dimension: Region (Discrete)
- Measure: Sales (Continuous)
  → Visualization: Bar chart comparing total Sales for each Region.

Each Region appears as a **separate category** on the x-axis.

1.4.2 Using Continuous Fields

Scenario: Management wants to see the trend of sales growth.

- Dimension: Order Date (Continuous timeline)
- Measure: Sales
  → Visualization: Line chart showing Sales over time.

The x-axis shows a **continuous date range** (Jan – Dec), and points are connected smoothly.

1.4.3 Combination of Discrete and Continuous

It is common to use both simultaneously.

Example:
"Sales Trend by Region"

- Order Date → Continuous x-axis
- Region → Discrete color dimension
- Sales → Continuous y-axis

Result: Four colored trend lines, one per Region, on a continuous timeline.

1.4.4 Practical Benefits

- **Discrete fields**: Great for **categorical comparisons** (e.g., sales by region, category).
- **Continuous fields**: Ideal for **trend and time-series** analysis (e.g., profit over months).
- Combination enables **multi-dimensional** insights.

1.5 Aggregation in Tableau

1.5.1 Meaning of Aggregation

Aggregation is the process of **summarizing multiple data points** into a single representative value.

When a Measure is placed in the view, Tableau **automatically aggregates** it (by default → SUM).

1.5.2 Common Aggregation Functions in Tableau

Function	Meaning	Example Use
SUM()	Adds all values	Total Sales per Region
AVG()	Finds mean value	Average Profit per Customer
COUNT()	Counts records	Number of Orders
MIN()	Lowest value	Minimum Sales Value
MAX()	Highest value	Maximum Profit
MEDIAN()	Middle value	Median Order Value

1.5.3 Default Aggregation

- Tableau uses **SUM** automatically for most measures.
- You can change aggregation by right-clicking the measure → **Measure (Sum)** → select another (Average, Count, etc.).

1.5.4 Example of Aggregation

Dataset

Region	Sales
East	10 000
East	12 000
West	8 000

Result after Aggregation:

- SUM(Sales) for East = 22 000
- SUM(Sales) for West = 8 000

In the bar chart:

- “East” bar = 22 000
- “West” bar = 8 000

1.5.5 Changing Aggregation Type

- Right-click on Sales → "Measure (Sum)" → choose AVG → Chart now shows average sales per region.
- Same data, different interpretation.

1.5.6 Why Aggregation Matters

- Reduces data size and simplifies visualization.
- Allows computation of Key Performance Indicators (KPIs) like Total Revenue, Average Order Value, etc.
- Enables comparisons between groups (e.g., average profit by region).

1.6 Summary of Concepts

Concept	Core Idea	Example / Output
Getting Started with Tableau	Connect data → Build visual → Share dashboard	Connect Excel "Sales.xlsx" and make bar chart
Dimensions vs Measures	Dimensions = categories; Measures = numerics	Region vs Sales
Discrete vs Continuous	Discrete (blue): headers; Continuous (green): axes	Sales by Region / Sales over Time
Applications	Combine discrete + continuous for multi-view analysis	Sales trend by Region
Aggregation	Summarizing data (SUM, AVG, COUNT, etc.)	SUM(Sales) per Category

---

1.7. Key Points to Remember

- Tableau fields = **Dimensions + Measures.**
- **Blue pill = Discrete field**, **Green pill = Continuous field.**
- Discrete → creates categories; Continuous → creates axes.
- Aggregation summarizes measure values; default is SUM.
- Tableau's workflow: **Connect → Analyze → Visualize → Share.**

UNIT – 2

2. Working with Metadata

2.1 What is Metadata in Tableau?

Metadata refers to the **"data about data"** — in Tableau, it includes **field names, data types, formats, roles, and structures** that define how Tableau interprets the underlying dataset.
When you connect a dataset, Tableau automatically reads its metadata (field names, types, hierarchies) and displays it in the **Data Pane**.

2.2 Components of Metadata in Tableau

1. **Field Names** – The names of columns or variables.
   - Example: *Order Date*, *Customer ID*, *Sales*.
2. **Data Types** – Define the kind of values a field holds:
   - Text (ABC), Number (#), Date (Calendar icon), Boolean (T/F), Geographic (Globe icon).
3. **Field Roles** – Whether the field is a *Dimension* or *Measure*.
4. **Field Properties** – Default aggregation, number format, etc.
5. **Data Source Metadata Grid** – View and edit field names, data types, and roles.

2.3 Editing Metadata

Tableau allows you to modify metadata without altering the original data source.
You can:

- **Rename Fields:** Right-click a field → *Rename* → type a descriptive name.
- **Change Data Type:** Click field icon → choose correct type (e.g., Date instead of String).
- **Change Role:** Move a field between *Dimensions* and *Measures*.
- **Create Aliases:** Provide alternate display names for values (e.g., "N" → "North").
- **Hide Fields:** Right-click → *Hide* (to declutter the data pane).

Example:
Original column name: `Cust_ID`
Renamed in Tableau as: `Customer ID` (for readability).

2.4 Metadata Management Best Practices

- Rename fields to business-friendly terms.
- Set correct data types early (e.g., avoid treating numbers as text).
- Hide unused fields to simplify the view.

- Create hierarchies (e.g., Country → State → City).
- Document field meanings for consistency.

2.2 Filters in Tableau

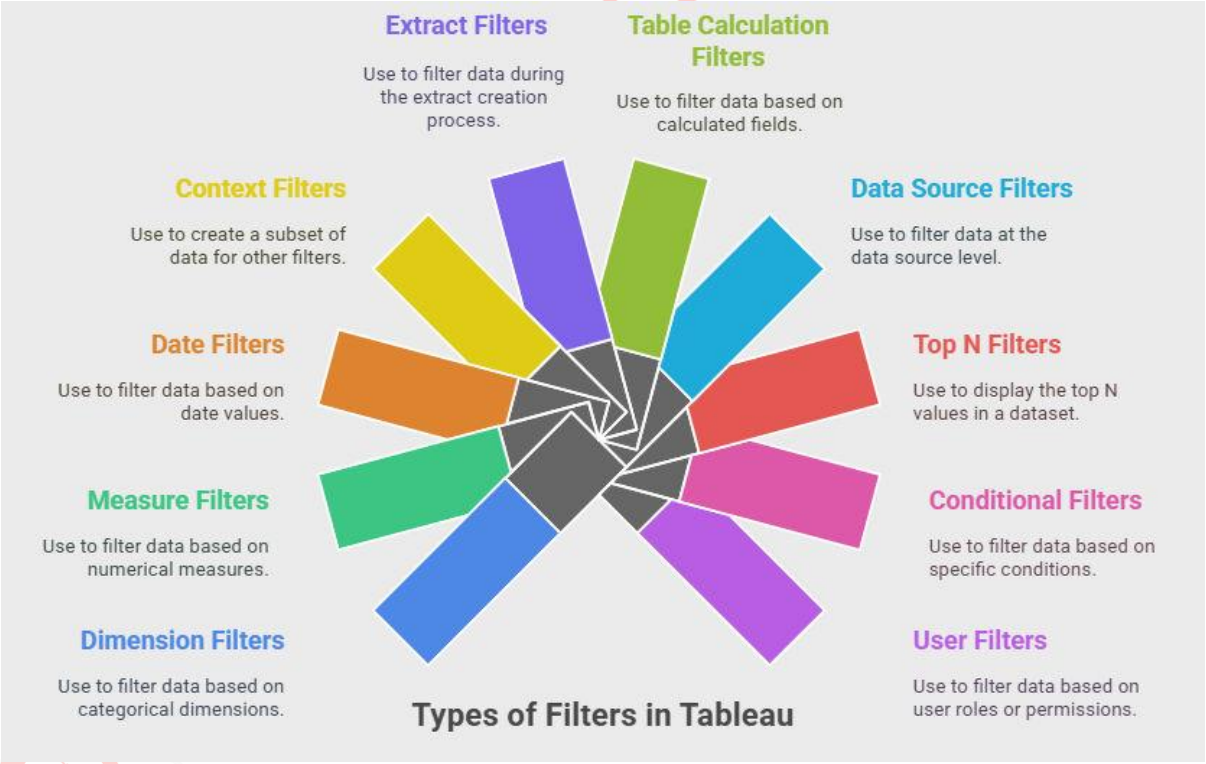
Filters help **focus analysis** by limiting the data visible in the worksheet or dashboard.

2.2.1 What is a Filter?

A **filter** removes unwanted data from the view, allowing analysis of only relevant information.
It acts like a **query condition** applied to visualizations.

Example:
View only "Sales in South Region" or "Orders between Jan–Mar 2024."

2.2.2 Types of Filters in Tableau



Filter Type	Applies To	Example / Use Case

Filter Type	Applies To	Example / Use Case
Extract Filters	Data Extract	Include only 2024 data during extraction.
Data Source Filters	Whole Data Source	Restrict analysis to "India" region.
Context Filters	Base filter for others	Filter first by Region, then by Category.
Dimension Filters	Categorical fields	View only Furniture or Technology category.
Measure Filters	Numerical values	Show records where Sales > ₹10,000.
Top N Filters	Rank-based filtering	Display Top 10 Customers by Sales.
Relative Date Filters	Time-based	View last 3 months of data dynamically.

2.3 Applying a Filter

1. Drag a field (e.g., *Region*) → **Filters Shelf.**
2. Choose filter type (e.g., *Select Values*).
3. Click **Show Filter** to display it as a dropdown on the worksheet.

Example:

- Drag *Category* → Filters → select *Furniture, Office Supplies.*
- The view now only displays those categories.

2.4 Context Filters (Important)

- A **context filter** acts as a *primary filter* that defines a subset of data for other filters to act on.
- Use when applying multiple dependent filters.

Example:
Filter 1 (Context): Region = South
Filter 2 (Dependent): Top 5 Products by Sales → Applies only to the South region.

2.5 Filter Order of Operations

1. Extract Filter
2. Data Source Filter
3. Context Filter
4. Dimension Filter
5. Measure Filter

2.6 Filter Presentation Options

- Dropdown Menu
- Checkboxes
- Slider (for range filters)
- Date Picker
- Search box

These make dashboards interactive.

2.7. Applying Analytics to the Worksheet

2.7.1 Purpose

Once data is visualized, Tableau allows analytical tools to be applied directly to the **worksheet** for deeper insights.

2.7.2 Common Analytical Features

Feature	Description	Example
Reference Line	Add static line showing average or target value.	Add "Average Sales" line to bar chart.
Trend Line	Show direction or trend over time (linear, exponential).	Sales growth over months.
Forecast	Predict future values based on patterns.	Forecast next 6 months' sales.
Average Line	Show mean of all values visually.	Average profit per region.
Box Plot	Visualize data distribution and outliers.	Profit distribution per product category.
Cluster Analysis	Group similar data points.	Segment customers by purchasing patterns.

2.7.3 How to Apply Analytics

1. Open a worksheet.
2. Go to **Analytics Pane** (next to Data Pane).
3. Drag analytics option (e.g., *Trend Line*, *Average Line*) onto the view.
4. Configure settings (e.g., linear regression, forecast period).

Example:
Drag "Trend Line" → drop it on "Sales over Time" line chart → Tableau fits a regression line showing trend.

2.7.4 Example Visualization

Scenario: An analyst wants to forecast sales for upcoming months.

Steps:

1. Create line chart: *Order Date (Continuous)* vs *Sales (Sum)*.
2. Drag *Forecast* from Analytics Pane.
3. Tableau projects future values and shows confidence intervals.

2.7.5 Analytical Insights Supported by Tableau

- **Descriptive Analysis:** What happened? (e.g., Sales by Region)
- **Diagnostic Analysis:** Why did it happen? (e.g., Profit by Discount Rate)
- **Predictive Analysis:** What will happen? (Forecasting trends)
- **Prescriptive Analysis:** What should we do? (Scenario-based visualization)

2.8. Dashboard in Tableau

2.8.1 Definition

A **Dashboard** is a **collection of multiple worksheets** and visuals displayed together on a single canvas to present a unified story or analysis.

2.8.2 Components of a Dashboard

- **Sheets:** Individual charts or maps.
- **Text and Images:** For titles or explanations.
- **Filters:** To make the dashboard interactive.
- **Containers:** Organize layout (horizontal or vertical).
- **Actions:** Add interactivity (filter, highlight, URL).

2.8.3 Creating a Dashboard

Steps:

1. Click the **Dashboard tab** → *New Dashboard.*
2. Drag required worksheets into the dashboard area.
3. Adjust layout and size.
4. Add filters, legends, and text boxes.
5. Add interactivity (Dashboard → Actions → Filter / Highlight).
6. Save or publish.

2.8.4 Example Use Case

A "**Sales Performance Dashboard**" may contain:

- Map showing *Sales by State*.
- Bar chart showing *Top 10 Products*.
- Line chart showing *Sales Trend over Time*.
- Filter for *Region* or *Category*.

(Search "Tableau Superstore Dashboard example" online to view visuals.)

2.8.5 Dashboard Best Practices

- Maintain a clean and consistent design.
- Use interactive filters for better user control.
- Keep color schemes meaningful and minimal.
- Use titles, legends, and tooltips for clarity.
- Ensure responsiveness for different screen sizes.

2.9. Modifications to Data Connections

2.9.1 What is a Data Connection?

When you connect Tableau to a file or database, a **data connection** defines the **link between Tableau and the source**.

Sometimes, analysts need to modify connections (e.g., change data source, update schema, or add new tables).

2.9.2 Modifying Data Connections

1. **Change Data Source:**
   - Go to *Data → Replace Data Source*.
   - Useful when the same structure data comes from a new file.
2. **Refresh Connection:**
   - *Data → Refresh All Extracts* to update data.
3. **Edit Connection:**
   - Right-click data source → *Edit Connection* → change path or database credentials.
4. **Add New Tables or Fields:**
   - Reopen connection editor → drag new tables or join additional data.

Example:
Replacing "Sales_2024.xlsx" with "Sales_2025.xlsx" while maintaining same fields.

2.10. Edit Data Source

2.10.1 Purpose

To clean or customize how Tableau reads and uses data from the source.

Steps

1. Go to *Data* menu → select your data source → *Edit Data Source*.
2. You'll enter the **Data Source Page**, showing tables, joins, and data preview.
3. You can:
    - Rename fields
    - Change data types
    - Create calculated fields
    - Hide unused columns
    - Manage joins/unions

Example

If a column named *Cust_ID* is numeric but used as text, change its type to "String."
Right-click → *Change Data Type* → *String*.

2.11. Unions, Joins, and Data Blending

These are **methods to combine data** from multiple tables or data sources.

Unions

Definition:
A **Union** appends (stacks) data from multiple tables with the **same structure** (same columns).

Example:

Table 1 (Sales_Jan) **Table 2 (Sales_Feb)**
OrderID Sales

After Union → Combined table with both months' data.

In Tableau:
Data Source Page → drag one table → drop another below it → choose *Union*.

Use When: You have monthly or quarterly files with identical structure.

Joins

Definition:
A **Join** combines tables **horizontally** based on a **common field (key)** like Customer ID or Order ID.

Types of Joins:

Join Type	Description	Result Example
Inner Join	Returns only matching records in both tables.	Customers who placed orders.
Left Join	All records from left table + matching ones from right.	All customers + their orders (null if none).
Right Join	All from right + matching from left.	All orders + customer data if available.
Full Outer Join	All records from both, matching or not.	Complete data union.

Example:
Joining "Customer Table" and "Orders Table" on *Customer ID*.

Data Blending

Definition:

Data Blending combines **data from different sources** (e.g., Excel + SQL) that can't be joined directly.

How it Works:

- One data source is the **Primary** (blue checkmark).
- The other is the **Secondary** (orange link icon).
- Linked through a **common field** (like Customer ID).

Example:

- Excel file: *Customer Info*
- SQL database: *Sales Transactions*
  → Blend using *Customer ID* as common field.

Use Case:
Compare "Actual Sales (SQL)" with "Sales Targets (Excel)."

2.11.1 Difference Between Join and Data Blending

Aspect	Join	Data Blending
Data Source	Same source	Different sources
Level	Row-level combination	Aggregated level
Performance	Faster	Slightly slower
Use When	Tables in same database	Data across systems (Excel + SQL)

2.12. Summary

Topic	Core Idea / Function
Working with Metadata	Adjust field names, types, and roles for clarity.
Filters	Restrict data shown in view (Dimension, Measure, Context).
Applying Analytics	Add reference lines, trend lines, and forecasts for insights.
Dashboard	Combine multiple views into one interactive canvas.
Modify Data Connections	Update or replace data sources and tables.
Edit Data Source	Clean or adjust fields, joins, and metadata.
Unions	Append rows from similar tables.
Joins	Merge tables based on keys.
Data Blending	Combine multiple data sources.

SUGGESTED BOOKS

1. **Data Visualization with Tableau** by Praveen Kumar
2. **Tableau 10.0 Best Practices** by Jenny Zhang