

# Tableau Lecture 2

Topics covered

- Basic Visual Analytics

# Basic Visual Analytics

## **Learning Objective:**

- In this module, you will understand the importance of Visual Analytics and explore the basic charts, features and techniques used for Visualization.
- We will learn about the importance of structuring data in Tableau and the various data structuring options available.

# Content

- The Basics
- Basic Charts types
- Pages
- Structuring of Data
  - Data Granularity
  - Aggregation
  - Data Granularity vs Aggregation
  - Various data Structuring options
    - Groups
    - Sets
    - Groups vs Sets
    - Parameters
    - Control sets with parameters
    - Actions
    - Filters and Types of Filters
    - Sorting
    - Hierarchy

# The Basics :

## ❖ What is Visual Analytics and why is it important?

- It is the use of sophisticated tools and processes to analyze datasets using visual representations of the data.
- Users get actionable insights which in turn help organizations make better, data-driven decisions.
- Allows users without data science skills or experience to combine, manipulate, and explore large, dynamic, multi-dimensional, and multi-sourced datasets. In a world in which big data becomes the norm, visual analytics becomes an essential tool.

# The Basics : (conti..)

## ❖ The difference between Visual Analytics and Data Visualization

Data visualization	Visual analytics
<ul style="list-style-type: none"><li>• Graphical/visual depiction of data to help people better understand the patterns, relationships, trends, and other meaningful insights in datasets.</li></ul>	<ul style="list-style-type: none"><li>• The use of an analytics program to perform advanced analysis of complex datasets and allow users to explore and interact with dynamic visualizations.</li></ul>

## The Basics : (conti..)

- ❖ How to use visual analytics for Business Intelligence (BI)?
  - Business users can leverage visual analytics to easily explore and understand large volumes of diverse data and gather the insights they need to solve problems, identify key opportunities, optimize performance, and inform strategic and tactical decision-making.
  - It can also be used to create interactive dashboards and support collaboration amongst teams working with the same datasets.

# Charts

Charts display many forms of analysis in a visual format: comparison, relationship, distribution, and composition. No two charts tell the same story. Next, we will look at few basic charts.

## Basic Charts types

- ❖ **Line chart** - They connect individual data points in a view. They provide a simple way to visualize a sequence of values and are useful when you want to see trends over time
  - Discrete Line chart
    - **Business problem 1** : Find product category that has highest sales for most of the years
  - Continuous Line chart
    - **Business problem 2** : Find the year and month that had the highest and lowest sales
- ❖ **Bar chart** - It is a chart with rectangular bars where lengths and heights are proportional to the value that they represent
  - Horizontal bar chart- it is bar graph that represents data horizontally dimensions are present on vertical axis and data values on horizontal axis
    - **Business problem 3** : Find sales and profit by region for the year 2017



- ❖ **Pie chart** - It represents data as a slice of a circle with different sizes and colour
  - **Business problem 4:** Show relative percentage of sales and profit by region for the year 2017
  
- ❖ **Histogram** - It is a chart that displays the shape of a distribution. A histogram looks like a bar chart but groups values for a continuous measure into ranges, or bins.
  - **Business problem 5:** Find sales by region based on the number of quantity sold (Bins)

# Pages

- The Pages shelf lets you break a view into a series of pages so you can better analyze how a specific field affects the rest of the data in a view.

# Structuring of Data

**Note :** We will show how structuring of data, filters etc are all related to granularity and aggregation of data in view (using tools like basic charts covered above)

# Data Granularity

## What is Data Granularity

- *Data granularity is a measure of the level of detail in a data structure.*
- In time-series data, for example, the granularity of measurement might be based on intervals of years, months, weeks, days, or hours.
- The name field could represent the full name or have separate entries for first name, middle name, and last name.

## Why is Data Granularity Important?

- The level of data granularity determines what analysis can be performed on the data, and whether results from that analysis lead to appropriate conclusions.
- The more granularity, the more information is available for analysis, but at the cost of increased storage, memory, and required computing resources.
- Some analysis may require information to be analyzed at a higher level, which would require aggregating the underlying detail into the higher level of granularity.

# Aggregation

- We can add aggregation to dimension and measures in tableau
- Types of aggregation for measures
- We can convert a Dimension to a measure and perform aggregation on it
- Types of aggregation for dimension
- If you want to remove the default aggregation for fields in the view than click on analysis and clear the aggregate measure option

# Data Granularity vs Aggregation

## Granularity vs Aggregation

- Granularity and aggregation work opposite of each other.
- In Tableau, when you bring dimensions and measure to the view, the measures are aggregated by whatever dimensions are on the view.
- As you remove and add more dimensions you are decreasing and increasing the granularity. The more dimensions/details on the view, the more granular it is.

# Various data structuring options :

- Groups
- Sets
- Group vs Sets
- Parameters
- **Control Sets with Parameters**
- Actions
- Filters and Types of Filters
- Sorting
- Hierarchy

# Groups

- Tableau allows single-dimensional members to be grouped together and automatically creates a new dimension with the Tableau Group at the end of the name.
- Tableau makes no adjustments to the members' original dimensions.

- **Business problem 6:**

Create a group called as small items which includes Binder, Bookcases, Envelope, Fastener, supplies and compare its sale with rest of the sub-categories

Dataset - Superstore Sales



# Sets

- Tableau Sets are **custom fields** that are used to keep a **subset of data** based on a **condition**.
- When you drag a **set** to the **visualization**, tableau displays the set using the **In/Out mode**. This is a binary categorization of sets
  - In-Members inside the set
  - Out- Members that aren't part of the set
- Static sets - **Members** of static set **do not change if** the **underlying data changes**
  - **Business problem 7-** Create a set of customer with highest profit
- Dynamic Sets - **Members** of dynamic data **changes when** the **underlying data changes**
  - **Business Problem 8** - Find top 5 products subcategories by sales
- Combined Sets - Allows us to **compare members** from **2 different sets**
  - **Business problem 9-** Find top customers by sales but low profitability

Data source-Sample superstore

## Groups vs sets

Sets	Groups
(Can be) <b>dynamic</b>	<b>static</b> (must be manually changed)
<b>Binary</b> (In/Out)	<b>No limit</b> to number of groups
Can be <b>created</b> only on <b>dimension</b>	Can be <b>created on</b> both <b>dimensions and measures</b>

# Parameters

- It is like a **workbook variable** that is used to **replace constant value** in a filter, calculation or reference lines
- They are **user generated** values that are **not necessarily attached** to the **dataset**

**Business problem 10**-Allow users to filter out top N subcategories of products by average discount

## Control Sets with Parameters

**Business problem 11**-Have a parameter where the user can choose top N products sub categories by sales

# Actions

- **Filter action** - Use the data from one view to filter data in another to help guide analysis.
  - **Business problem 12** - Filter Out sales of products based on categories and subcategories (superstore dataset)
- **Highlight action** - Call attention to marks of interest by coloring specific marks and dimming all others.
  - **Business problem 13**- Highlight sales based on region for each product category and sub category (superstore dataset)
- **URL action** - Create hyperlinks to external resources, such as a web page, file, or another Tableau worksheet.
  - **Business problem 14** - Create URL action such that when you click on a country it opens up wikipedia page for that country(superstore dataset)

# Actions(conti..)

- **Change Parameter** - Let users change parameter values by directly interacting with marks on a viz.
  - **Business problem 15** - Create a Parameter action to display the sales amount based on the category that is being selected
- **Change set value** - Let users change the values in a set by directly interacting with marks on a viz.
  - **Business problem 16**- Display what proportion did each region contributed to total monthly sales for each year using set action

# Filters

- Filter removes some scope of data from a data set.
- Filters are very helpful to create dashboards in Tableau.
- Filters can help to minimize the size of data sets for efficient use, eliminate irrelevant dimension elements, clean up underlying data, set date ranges and measures as required, simplify and organize data, etc.

## Filtering Order of Operations

Tableau performs actions on our view in a very specific order; this is called the Order of Operations. It helps you avoid filter conflicts and achieve efficiency with your dashboard.

Filters are executed in the following order:

1. Extract filters
2. Data source filters
3. Context filters
4. Filters on dimensions (whether on the Filters shelf or in filter cards in the view)
5. Filters on measures (whether on the Filters shelf or in filter cards in the view)
6. The Table Calculation filter

# Types of Filters

## 1. Extract filters

- Extract filter in Tableau are used to extract a small subset of data from the original data source.
- Tableau then creates a local copy of the data set that is to be stored in the repository.
- These methods reduce Tableau queries.
- The data size can be further reduced by applying the measure or dimension filter to the extract as required.
- Unlike Tableau Desktop(paid) which supports both extract and live connections, Tableau public, by default, only supports extract.



## 2. Data source filters

- Data source filters in Tableau are mainly used to restrict sensitive data from viewers and reduce data feeds.
- Viewers can, however, have certain access rights to view the underlying data.
- Data source filters allow the direct application to source data.
- One important thing to mention is that the extract filter and the data source filter are not linked, and if you happen to go back to a live connection, the data source filter will remain intact.

**Business Problem 17** : Show orders which made profit  $\geq$  1K

### 3. Context filters

- By default, all filters that you set in Tableau are computed independently. That is, each filter accesses all rows in your data source without regard to other filters. However, you can set one or more categorical filters as context filters for the view.
- You can think of a context filter as being an independent filter. Any other filters that you set are defined as dependent filters because they process only the data that passes through the context filter.
- The context filter adds an actionable context to data analysis, but if the data is not reduced enough, the cost of computing can be very high.

**Business Problem 18** : Show top 10 furniture items based on sales

#### 4. Filters on dimensions (whether on the Filters shelf or in filter cards in the view)

- Dimension filters in Tableau are **non-aggregated filters**. The dimensions that are used are mostly blue pills. Blue pills correspond to discrete data.
- **If there are many dimensions, one can search** for them. Dimension filter provides **four options, General, Wildcard, Condition, and Top/Bottom**. You can pick up any of the four options **to select the right data or remove the unwanted data.**
- One can **create their own formula** as well and then **use it in** the **Condition filter and** the **Top/Bottom filter for data selection.** They provide a channel to measure to get the required data.

**Business Problem 19:** Show ship mode and subcategories wrt profit where subcategories Labels and Storage are excluded.

5. Filters on measures (whether on the Filters shelf or in filter cards in the view)

- Using a **Measure filter** in Tableau **allows** for **various operations and aggregate functions** such as sum, median, avg, standard deviation, etc.
- **Aggregated filters** are always applied **after non-aggregated filters**, no matter what the order is on the Filters pane.
- The filters are **applied to Measure fields** consisting of **quantitative data**.

In a subsequent dialog box, you will get **four types of filters**:

1. **Range:** Select the range of values to include in the result
2. **At least:** Select the minimum value of a measure
3. **At most:** Select the maximum value of a measure
4. **Special:** Select null or non-null values

**Business Problem 20** : Show only the subcategories whose average profit is greater than 20.

## 6. The Table Calculation filter

- The Table Calculation filter is the **last filter** that is applied **after the view has been created**.
- If you want to add a filter to the view, the Table Calculation filter will do the job for you **without filtering the underlying data**.
- **Table calculations** are a special **type of calculated field that apply transformations** (i.e. additional math) **on values within a visualization**. Common **examples** of table calculations include **running sum, moving average, and percent of total**. *Calculated field* will be **covered later in the course**.

## 7. Interactive Filter

- When an interactive filter is shown, you can **quickly include or exclude data in the view**.

**Note :** A Demo of working is shown .

## 8. Date Filter

- When you drag a date field from the Data pane to the Filters shelf, Filter Field dialog box appears .
- You can select whether you want to :
  - filter on a relative date;
  - filter between a range of dates; or
  - select discrete dates or individual dates to filter from the view.

**Business Problem 21** : Show Profit wrt Order date(where date is a continuous dimension)

**Extra :**

## Cascading Filter

**Problem Statement 22** : Using a cascading filter to find out the customers of a particular Region, State, City, and Postal Code.

Demonstration Video : <https://www.youtube.com/watch?v=uK8K1C3Pxl0&t=20s>

# Sorting

Sorting of data is a very important feature of data analysis. Tableau allows the sorting of data of the fields, which are called dimensions.

## ❖ Data source sorting :

- In the data source, data can be stored based on the user requirement.
- It can be sorted using data source order such as A to Z ascending, Z to A descending, A to Z ascending per table and Z to A descending per table.
- Once the data is connected with Tableau, data sorting is done using the Sort Fields option. The Sort Fields option is present in the Data Source tab.

## ❖ Computed vs Manual Sorting

- The computed sorting is a sort which is directly applied on the axis using the sort dialog button.
- Manual sorting is a sort that rearranges the order of dimension fields by dragging them next to each other in ad hoc fashion.

**Note :** Demonstration of working of Data Source Filtering, Computed and Manual Sorting have been added.



# Hierarchy

- Hierarchical Data Visualization is used in Business Intelligence and Data Analytics to show how certain items or data are ranked in a system. The level-wise configuration is represented by Hierarchy, which is a tree-like structure.
- Data with relationships can be used to create a Hierarchy in Tableau.
- When you connect to a data source, Tableau automatically separates date fields into hierarchies so you can easily break down the viz.

**Business problem 23** - Build a Hierarchy of Product category and Sub Category and have a plot which displays sales of product by category and drill it down using the hierarchy to display sales by sub category

**Business Problem 24** - Using the Natural Hierarchy of Dates build plots to display total sales for each year,for each quarter and for each month

End