

18CSE490T - Big Data Visualization

Unit 3

Data Visualization?

- ✓ Data visualization is a graphical representation of any data or information.
- ✓ Visual elements such as charts, graphs, and maps are the few data visualization tools that provide the viewers with an easy and accessible way of understanding the represented information.
- ✓ In this world governed by Big Data, data visualization enables you or decision-makers of any enterprise or industry to look into analytical reports and understand concepts that might otherwise be difficult to grasp

Big Data Visualization – Popular Tools

- ✓ **Zoho Analytics:** Focusing on ease of use – a particularly key attribute as data tools grow – Zoho analytics is a self service option. Meaning that users will not need the assistance of IT staff or professional data scientists to glean insight from data.
- ✓ **IBM Cognos Analytics:** Driven by their commitment to Big Data, IBM's analytics package offers various self-service options to more easily identify insight.
- ✓ **QlikSense and QlikView:** The Qlik solution touts its ability to perform the more complex analysis that finds hidden insights.
- ✓ **Microsoft PowerBI:** The Power BI tools enables you to connect with hundreds of data sources, then publish reports on the Web and across mobile devices.

Big Data Visualization – Popular Tools

- ✓ **Oracle Visual Analyzer:** A web-based tool, Visual Analyzer allows creation of curated dashboards to **help discover correlations and patterns in data.**
- ✓ **SAP Lumira:** Calling it “**self service data visualization for everyone,**” Lumira allows you to combine your visualizations into storyboards.
- ✓ **SAS Visual Analytics:** The SAS solution promotes its “**scalability and governance,**” along with dynamic visuals and flexible deployment options.
- ✓ **Tableau:** Tableau’s interactive dashboards allow you to “**uncover hidden insights on the fly,**” and power users can manage metadata to make the most of disparate data sources.
- ✓ **TIBCO Spotfire:** Offers analytics software as a service, and touts itself as a solution that “scales from a small team to the entire organization.”

TOP DATA VISUALISATION TOOLS



Tableau Desktop



Zoho Reports



Microsoft Power BI



MATLAB



Sisense

Data Visualization using Tableau

- ✓ Tableau is the US-based data visualization firm that easily connects to almost any data source; it could either be corporate Data Warehouse, Microsoft Excel or web-based data.
- ✓ Also, the data does not have to be accumulated as Tableau allows analysis on real-time data feed.
- ✓ **Tableau is designed to support how you think, utilizing drag and drop to create visualizations of your data and leverage your natural ability to spot visual patterns quickly**
- ✓ Tableau enables to connect with various data sources, files, and servers.
- ✓ Tableau also encourages to work on different file formats such as CSV, JSON, Txt, and even servers such as Tableau Server, MySQL, Amazon Redshift and more.

Other benefits of Tableau

- ✓ Tableau is a smart tool and even recommends which visualization tool could be used for what purpose. You could navigate through the software, click on the 'Show Me' feature available within the tool, which would show different types of graphs and charts with various attributes.
- ✓ Tableau also supports map, data of which could be easily modified by you, and unlike other BI tools, you don't have to break your head to use them. It is relatively easy to use maps in Tableau.
- ✓ Tableau has an interactive UI that plays an important role in how you could design and develop data visualization charts. The tool offers numerous fields and chart options such as heat map, Scatter Plot, Packed Bubble, and more.

Trifacta (Tool to perform some data manipulations) – Data wrangling

- ✓ Trifacta is a privately owned software company headquartered in San Francisco with offices in Bengaluru, Boston, Berlin and London. The company was founded in October 2012 and primarily develops data-wrangling software for data exploration and self-service data preparation on cloud and on-premises data platforms.
- ✓ There are many options for **data profiling, preprocessing, and manipulation**, A tool that is popular with many Tableau users is a product named Trifacta, which claims that it is the number-one (data) wrangling solution for Tableau.
- ✓ The term wrangling is defined loosely as to round up, herd, or otherwise take charge of, and in this case, **the wrangling being performed focuses on data**
- ✓ Trifacta can be downloaded for your personal evaluation from www.trifacta.com

Trifacta working with example

- Kindly refer to the Book James D. Miller Big Data Visualization – page No: (179 – 194)

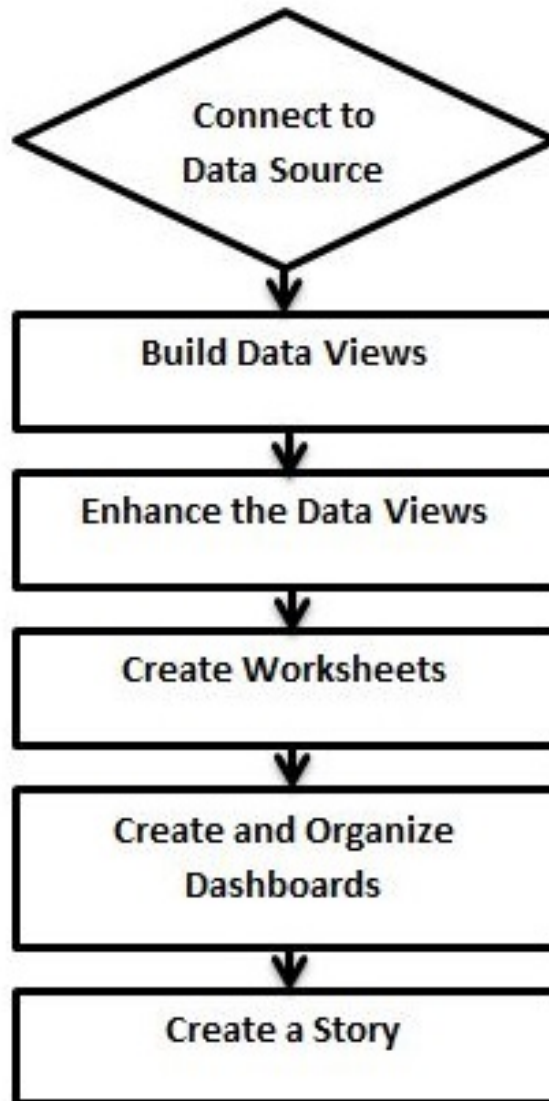
Tableau - Overview

- Tableau is a leading data visualization tool with many desirable and unique features.
- Its powerful data discovery and exploration application allows you to answer important questions in seconds.
- You can use Tableau's drag-and-drop interface to visualize any data, explore different views, and even combine multiple databases easily. It does not require any complex scripting.
- Anyone who understands the business problems can address it with a visualization of the relevant data.
- After analysis, sharing with others is as easy as publishing to Tableau Server.

Tableau - Design Flow

- As Tableau helps in analyzing lots of data over diverse time periods, dimensions, and measures, it needs a very meticulous planning to create a good dashboard or story. Hence, it is important to know the approach to design a good dashboard. Like any other field of human endeavor, there are many best practices to be followed to create good worksheets and dashboards.
- Though the final outcome expected from a Tableau project is ideally a dashboard with story, there are many intermediate steps which needs to be completed to reach this goal.
- Following is a flow diagram of design steps that should be ideally followed to create effective dashboards.

Tableau - Design Flow (Con..)



Connect to Data Source

Tableau connects to all popular data sources. (Text files, relational sources, SQL sources or cloud databases, Tableau connects to nearly every data source)

Build Data Views

After connecting to a data source, you get all the columns and data available in the Tableau environment. You classify them as dimensions and measures and create any hierarchy required.

Enhance the Views

The views created above need to be enhanced further by the use of filters, aggregations, labelling of axes, formatting of colors and borders, etc.

Create Worksheets

Create different worksheets to create different views on the same or different data.

Create and Organize Dashboards

Dashboards contain multiple worksheets which are linked. Hence, the action in any of the worksheets can change the result in the dashboard accordingly.

Create a Story

A story is a sheet that contains a sequence of worksheets or dashboards that work together to convey information.

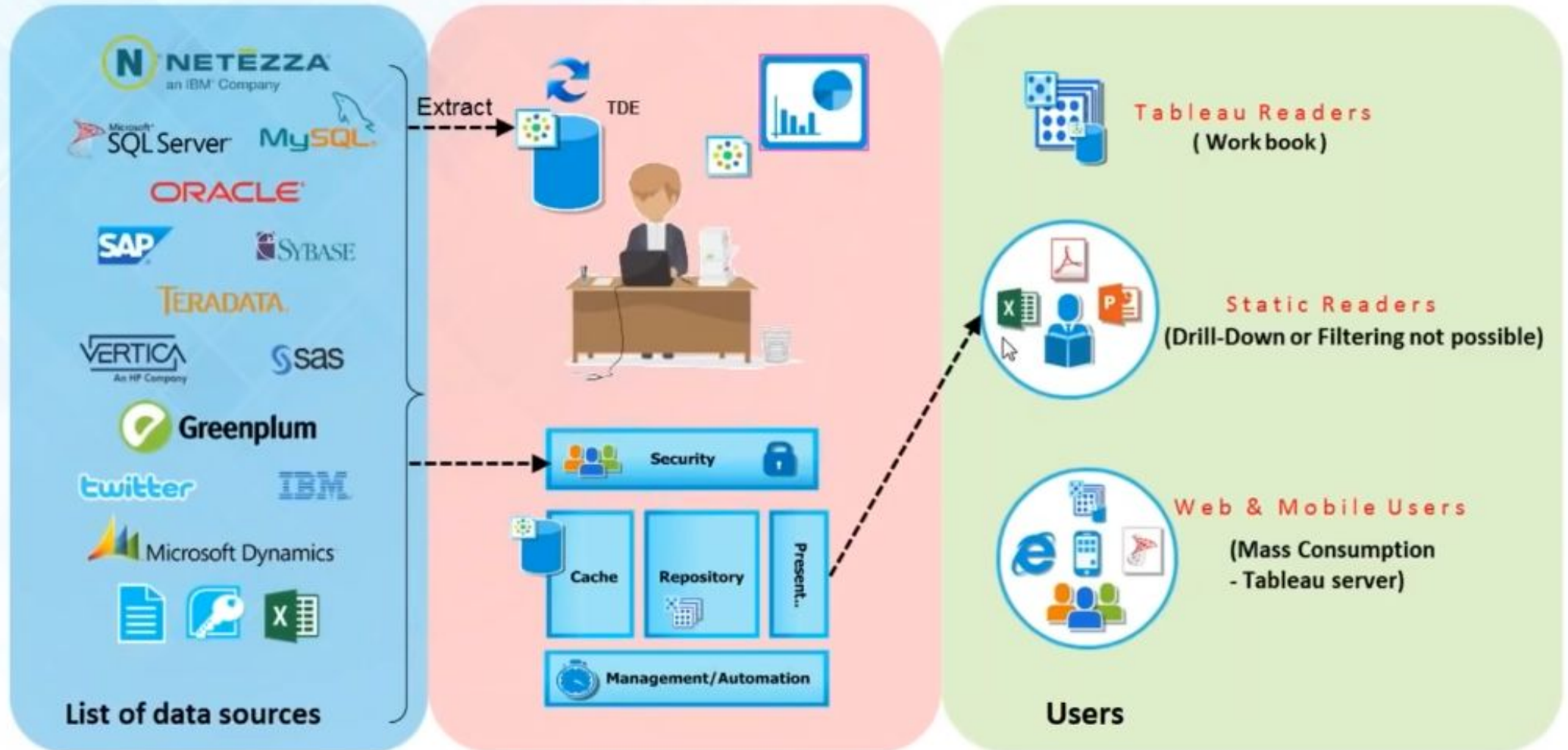
Tableau Product Family

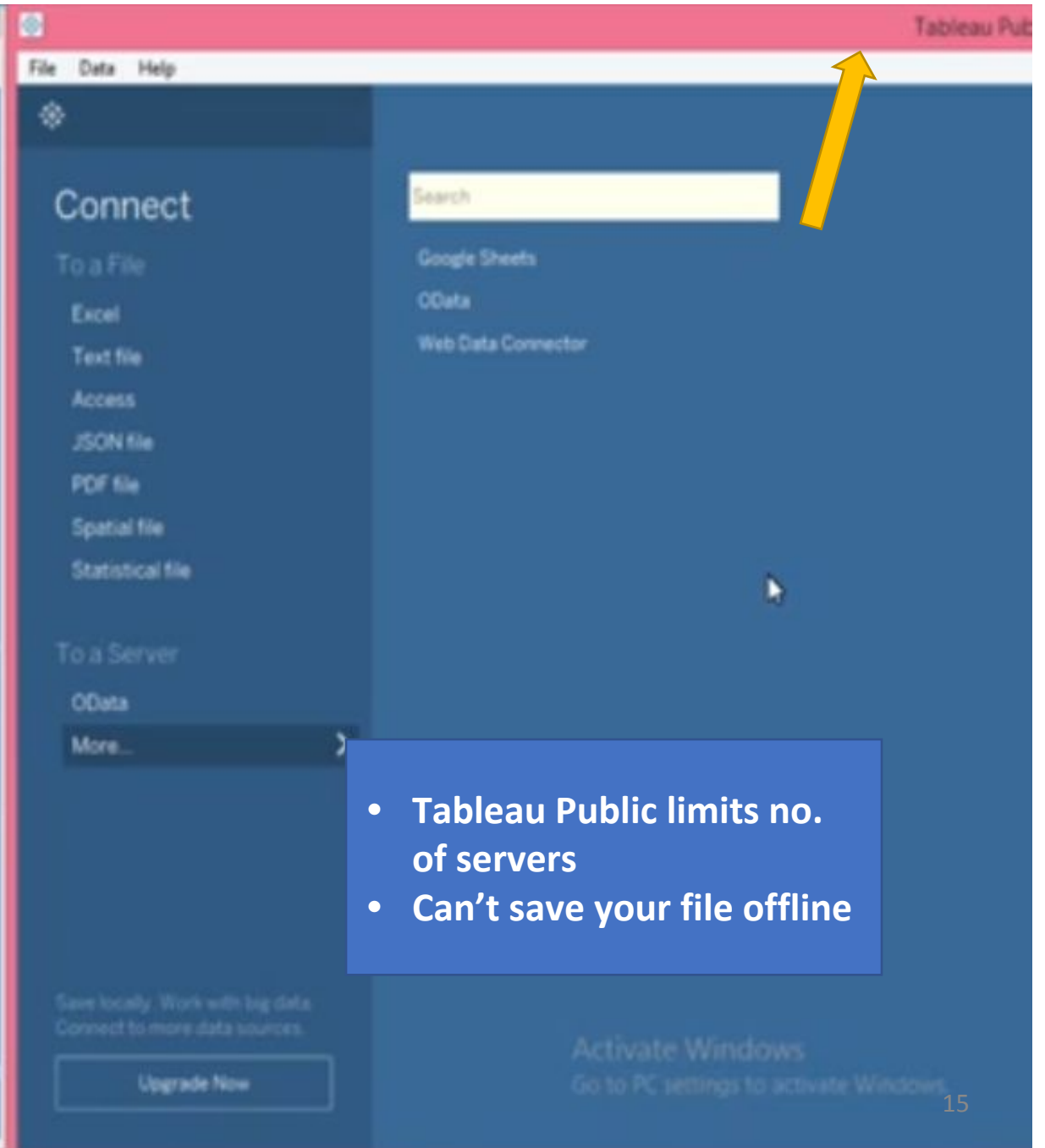
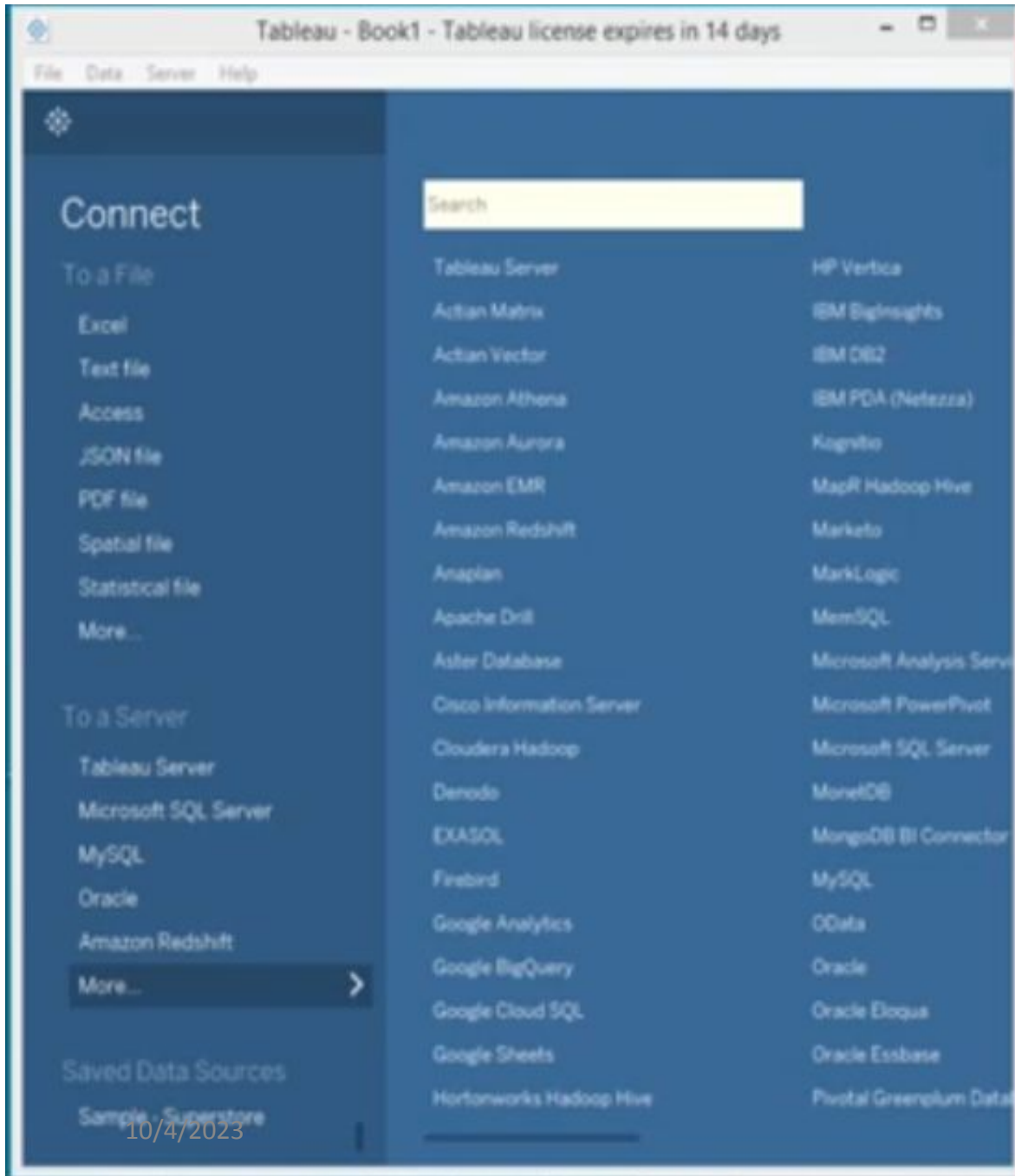


Tableau Public and Tableau Reader are free to use, while both Tableau Server and Tableau Desktop come with a 14-day fully functional free trial period

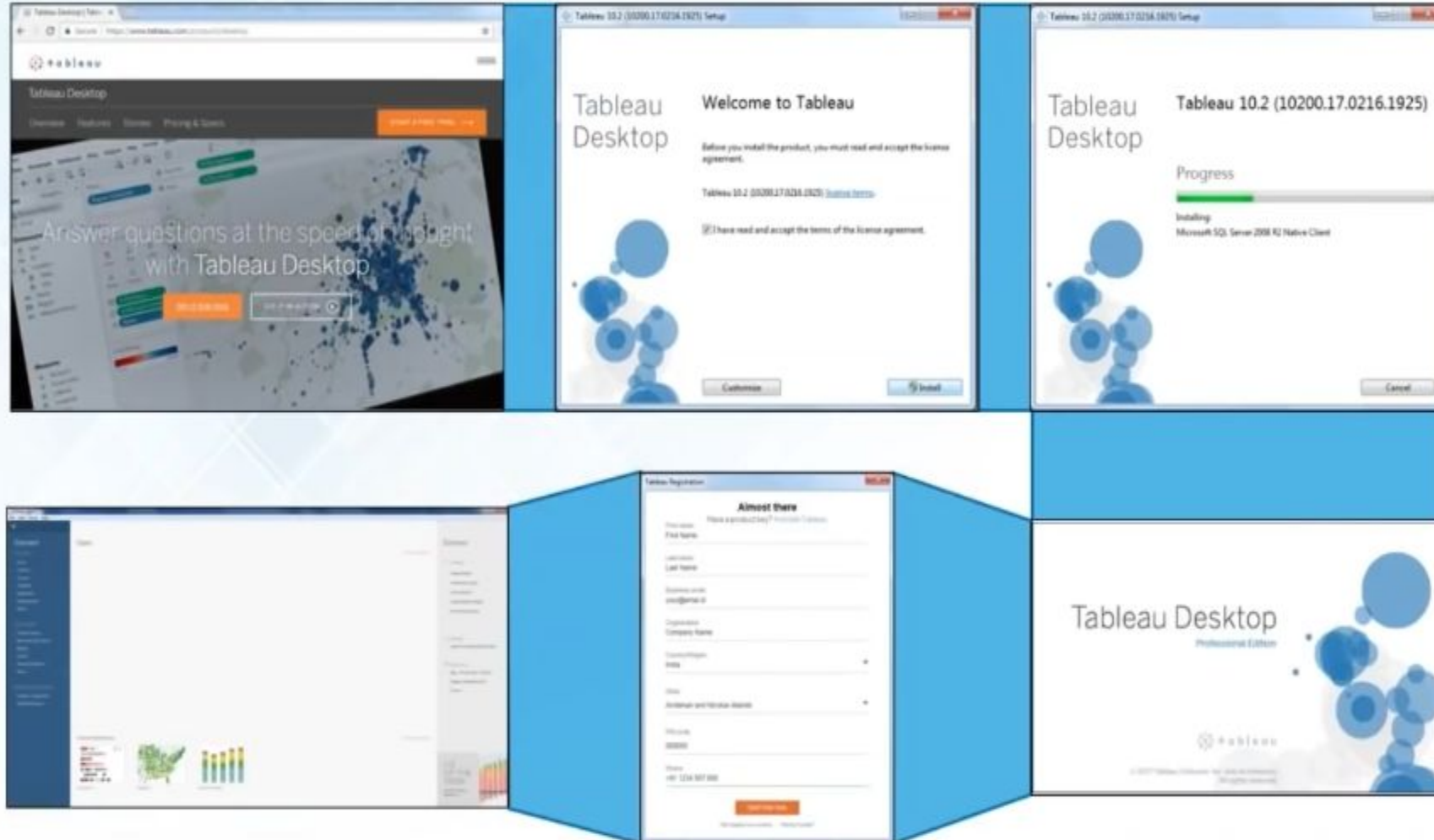
Download the trial version using these link <http://www.tableau.com>
Register to get a year Licence using these link <https://www.tableau.com/academic/students>

Tableau Architecture





Installing Tableau



Data Types

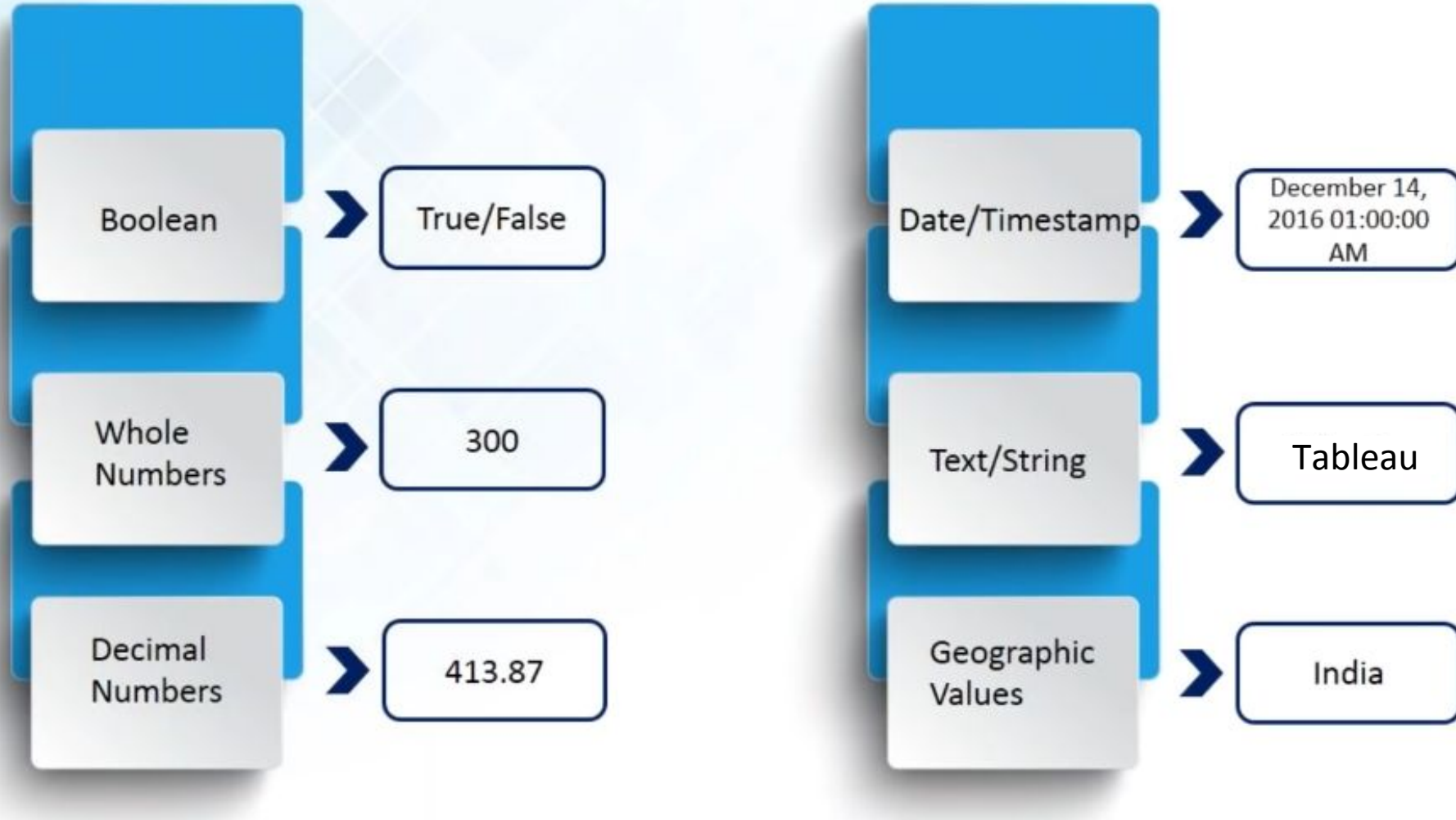


Tableau Datatypes

Icon	Data type
Abc	Text (string) values
📅	Date values
🕒	Date & Time values
#	Numerical values
T/F	Boolean values
🌐	Geographic values

Boolean

True / False

True / False

Date

Date Values

December 31, 2016

Date & Time

Date &
Timestamp Values

December 31, 2016
01:00:00 AM

Geographical Values

Geographical Mapping

New York, France,
India, Beijing

Text / String

Text or String Values

Welcome to Tableau

Number (Decimal)

Decimal Numbers

5.0000

Number (Whole)

Whole Numbers

5

Tableau UI – Connections

File Data Server Window Help

Connections

Sample - Superstore

Orders

Connection: ☒ Live ☐ Extract

Filters: 0 | Add

Connection Established with file: "Sample - Superstore"

Sort fields: Data source order

Show aliases Show hidden fields 1,000 rows

Order ID	Order Date	Ship Date	Ship Mode	Customer Name	Segment
CA-2011-108800	03-01-2013	07-01-2013	Standard Class	Darren Powers	Consumer
CA-2011-112326	04-01-2013	08-01-2013	Standard Class	Phillina Ober	Home Office
CA-2011-112326	04-01-2013	08-01-2013	Standard Class	Phillina Ober	Home Office
CA-2011-112326	04-01-2013	08-01-2013	Standard Class	Phillina Ober	Home Office
CA-2011-141817	05-01-2013	12-01-2013	Standard Class	Mick Brown	Consumer
CA-2011-130813	06-01-2013	08-01-2013	Second Class	Lycoris Saunders	Consumer

A preview of the data set can be seen before opening our worksheet

We have connected to an *Excel* Sheet "*Sample - Superstore*"

This is the *Dashboard* once we load data into *Tableau Desktop*

Tableau Desktop UI

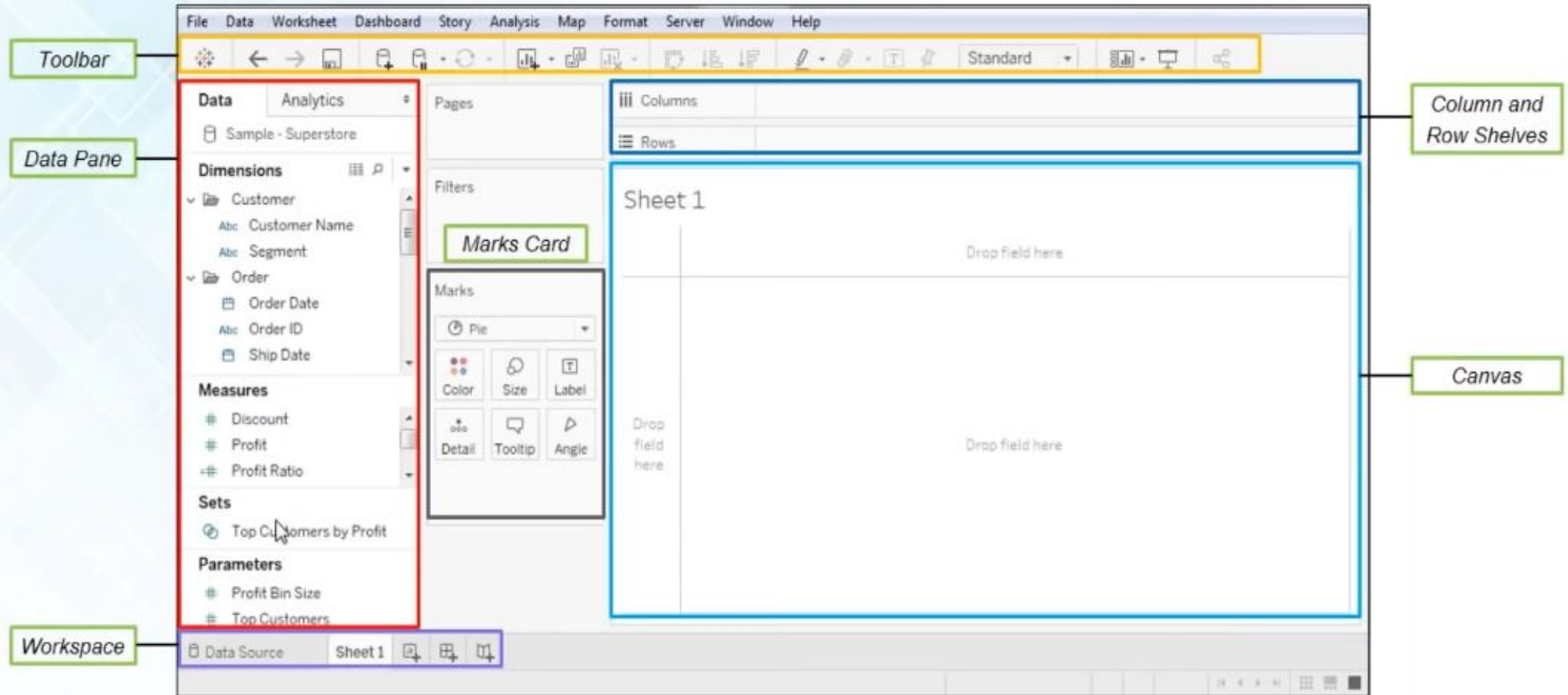


Tableau Dashboard

Measures & Dimensions

Dimensions

Abc

A Dimension is a field that is an independent variable.

A dimension is usually text

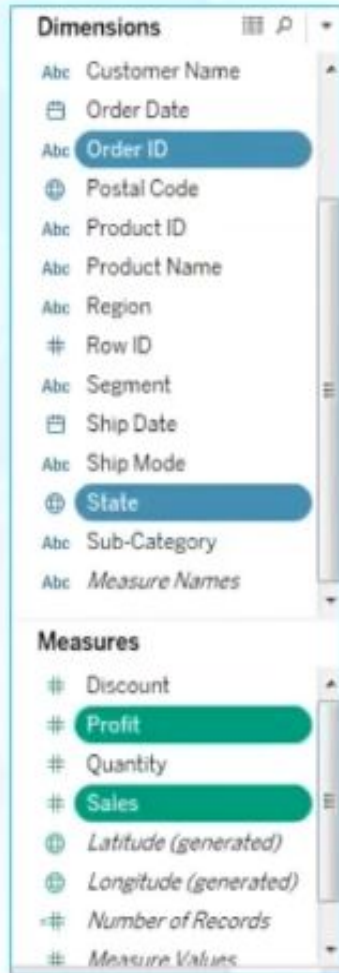
Measures

#

A Measure is a field that is a Dependent Variable and its value is a function of one or more Dimensions.

A measure is usually a number

Tableau UI – Dimensions & Measures



Once we connect to a data source in Tableau, the columns get divided into *Dimensions* and *Measures*.

A *Dimension* is a field that is an independent variable.

A *Measure* is a field that is a *Dependent Variable* and its value is a function of one or more *Dimensions*.

Tableau UI – Show Me



Show Me in Tableau shows all the possible visualizations available for our data source.

The visualizations that are highlighted are the ones that are currently available for the current selection of dimensions and measures.

Tableau – Applying Visualizations

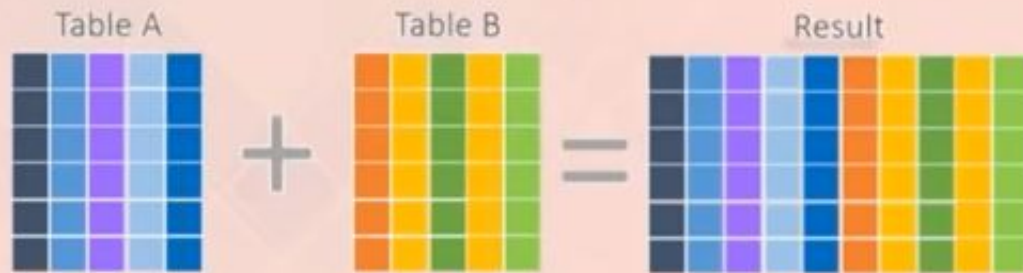
Let us discuss which visualizations are the best for various applications

Visualization	Application
Bar Graph	Used when Dimension is discontinuous
Line Graph	Preferred for continuous Dimensions
Dual Axis Graph	Used to represent two Measures together
Geographical Graph	Used to plot Measures on geographical map
Area Graph – Dual Axes	Provides better comparison amongst Measures
Heat Map	Used to visualize variations across categories
Tree Map	Used to represent quantity in nested rectangles

Tableau – Join & Union

Join and Union functions are used to perform operations to combine data

Join – Combine Columns



Union – Combine Rows

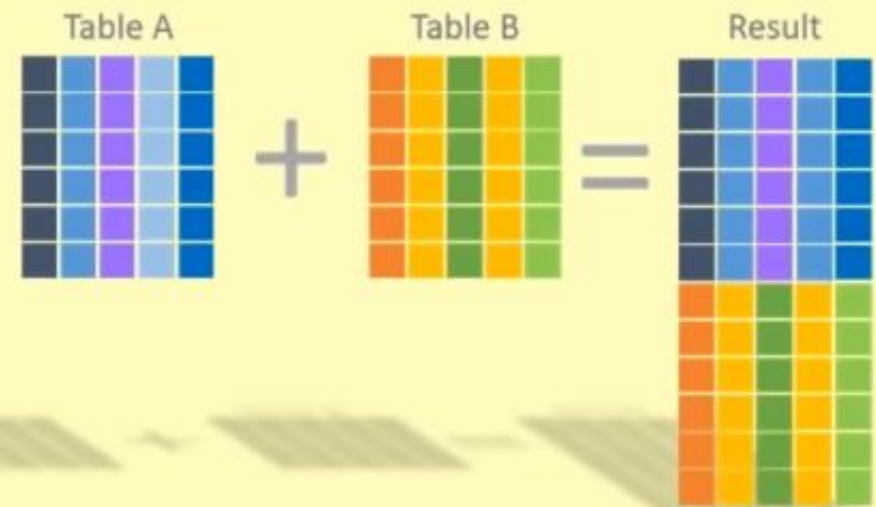
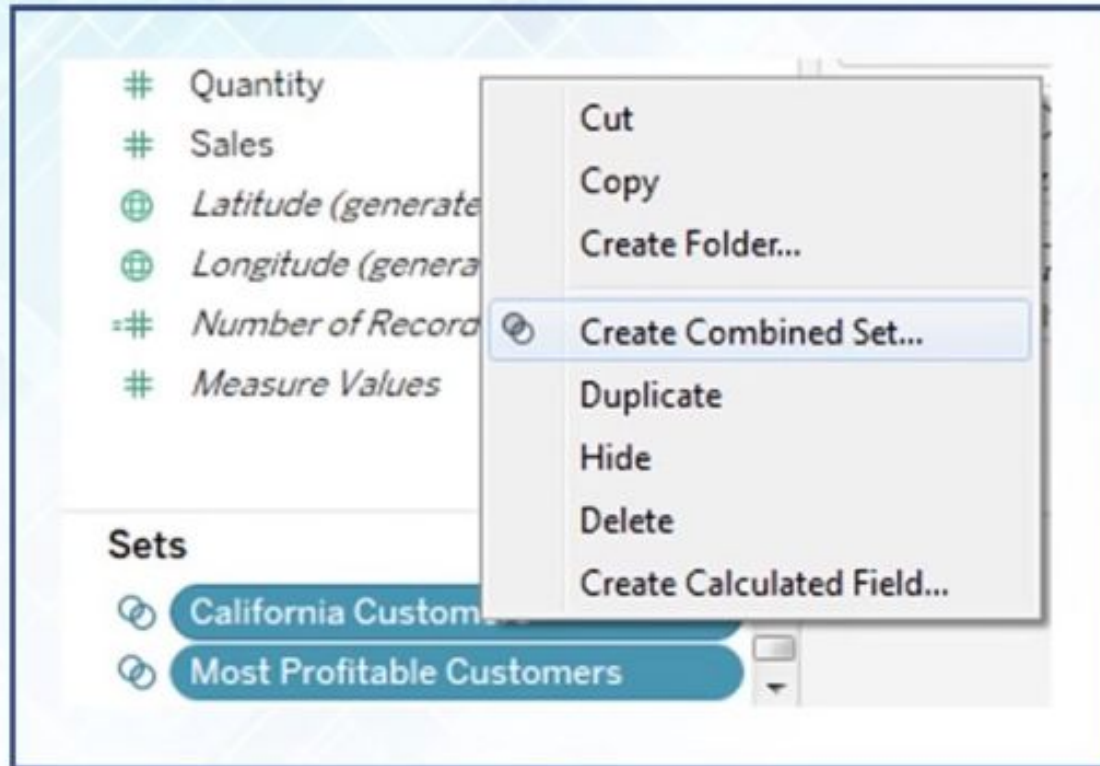


Tableau – Sort

Sort function can be used to sort categories and sub categories based on their measure



Sets



- **Sets** are custom fields that define a subset of data based on some conditions.
- A **set** can be based on a computed condition.
- For example, a **set** may contain customers with sales over a certain threshold.
- Computed **sets** update as your data changes.
- Alternatively, a **set** can be based on specific data point in your view

Tableau – Set

Set is a type of filter where we can set a condition for displaying values, e.g. Discount > 10%

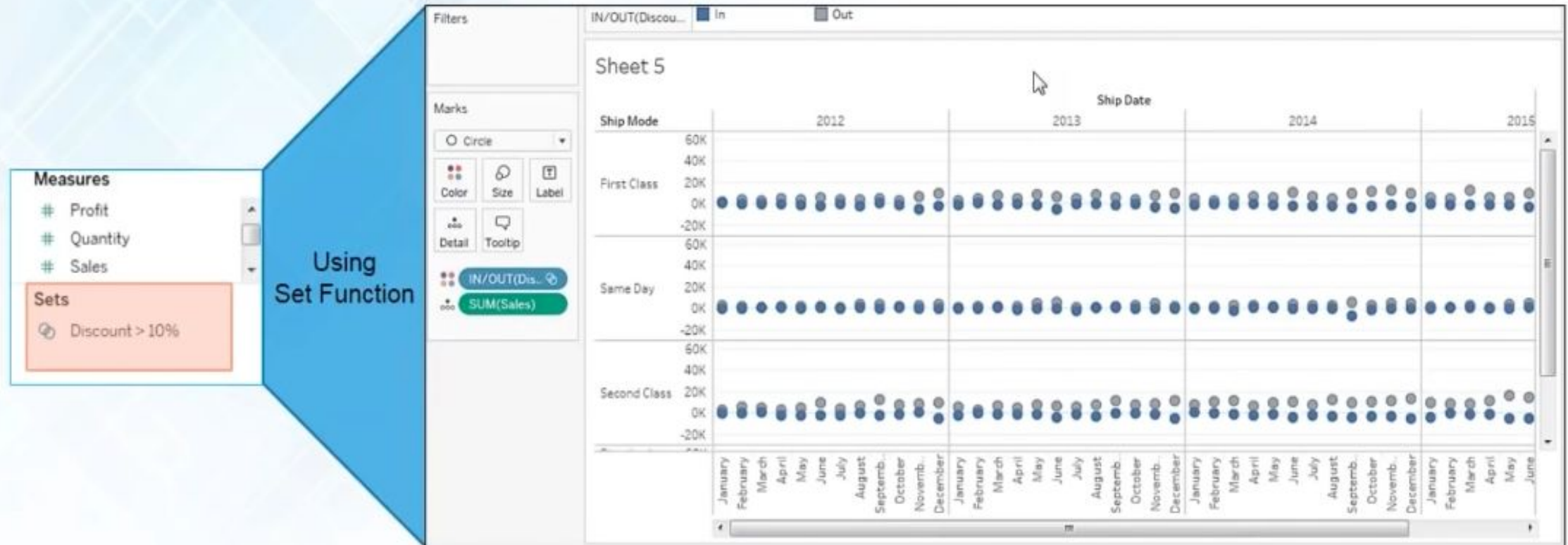


Tableau – Forecasting

Forecasting is used to predict the future trend based on the current values of graphs

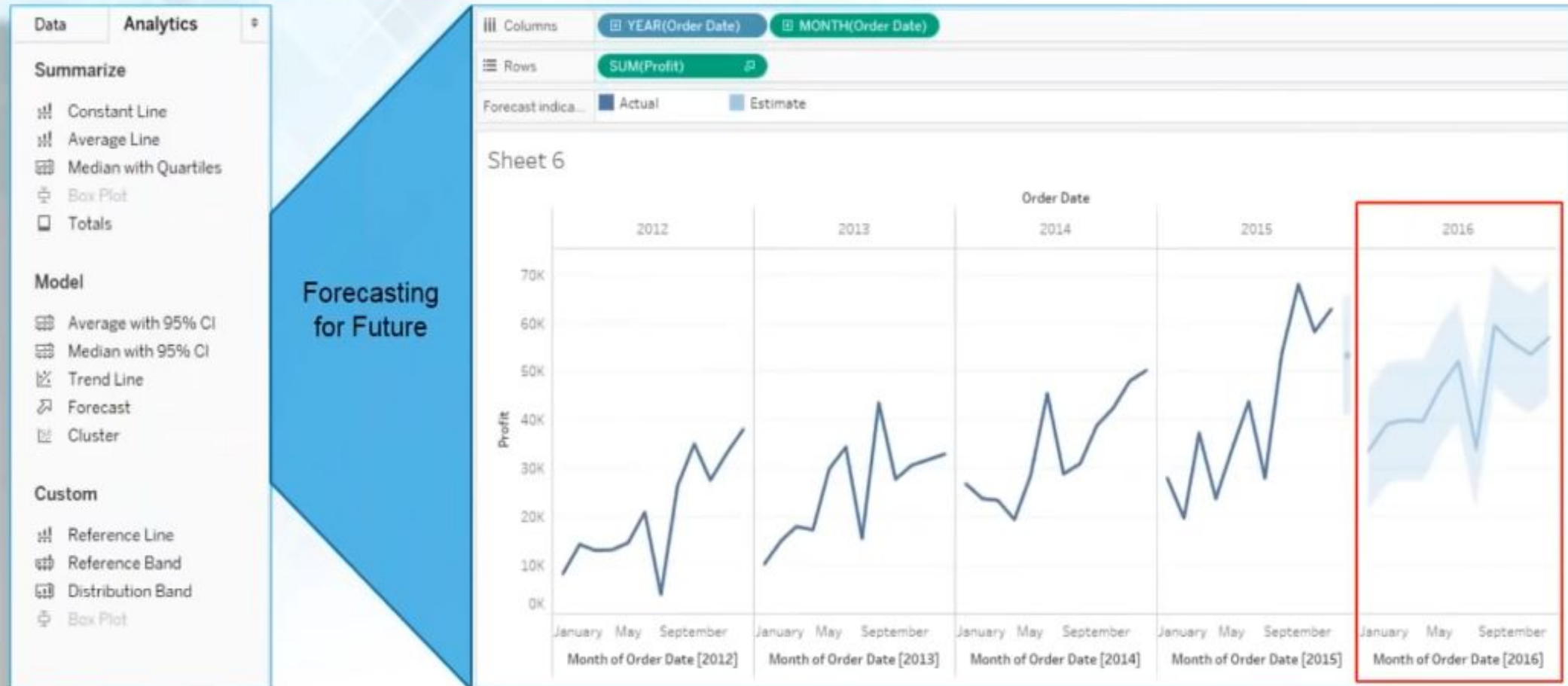


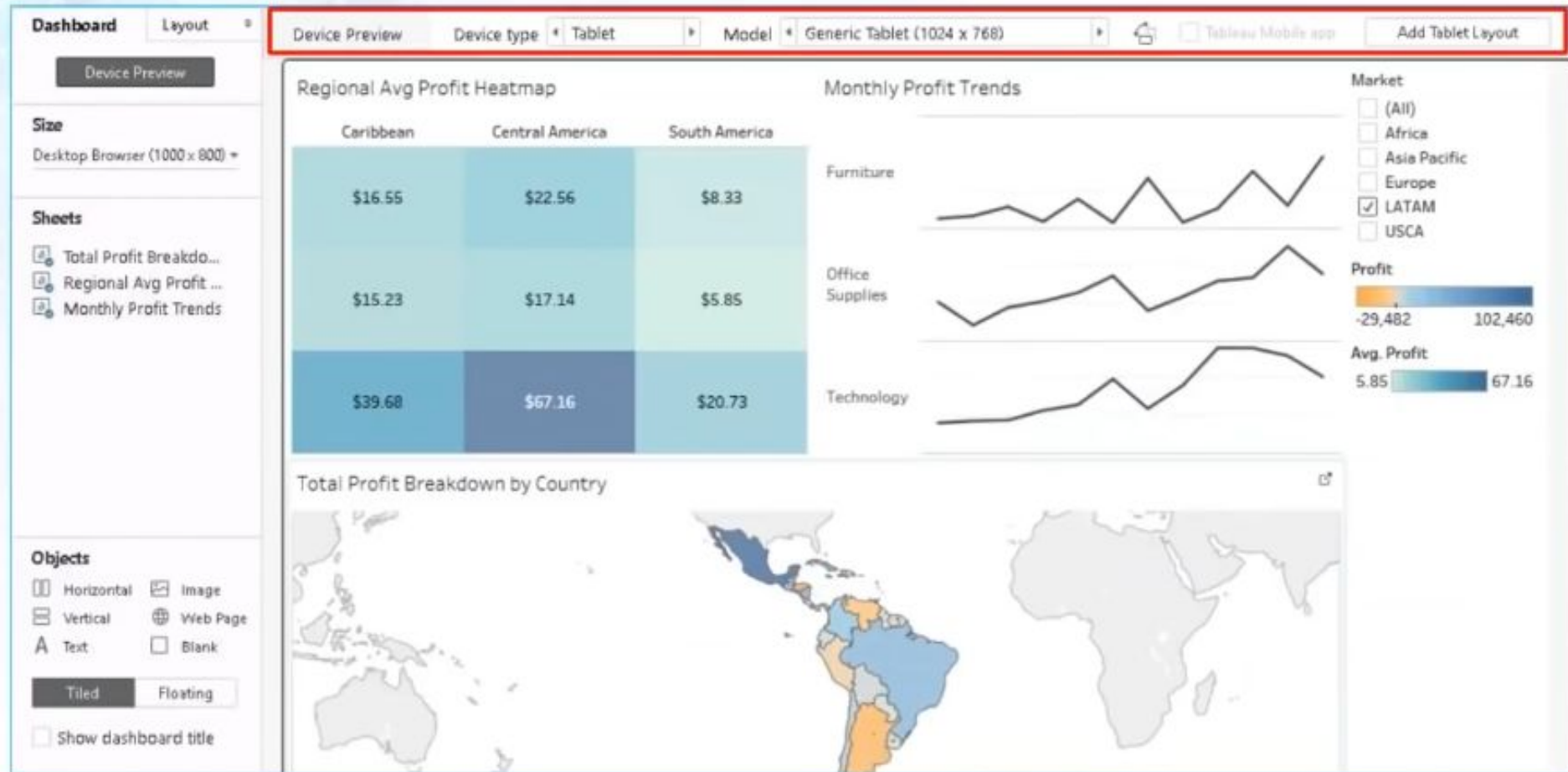
Tableau – Highlighting

Highlighting is used to highlight specific sub categories among all others



Tableau – Device Designer

Device Designer is used to make best visualizations for specific devices



Advantages

Drawbacks

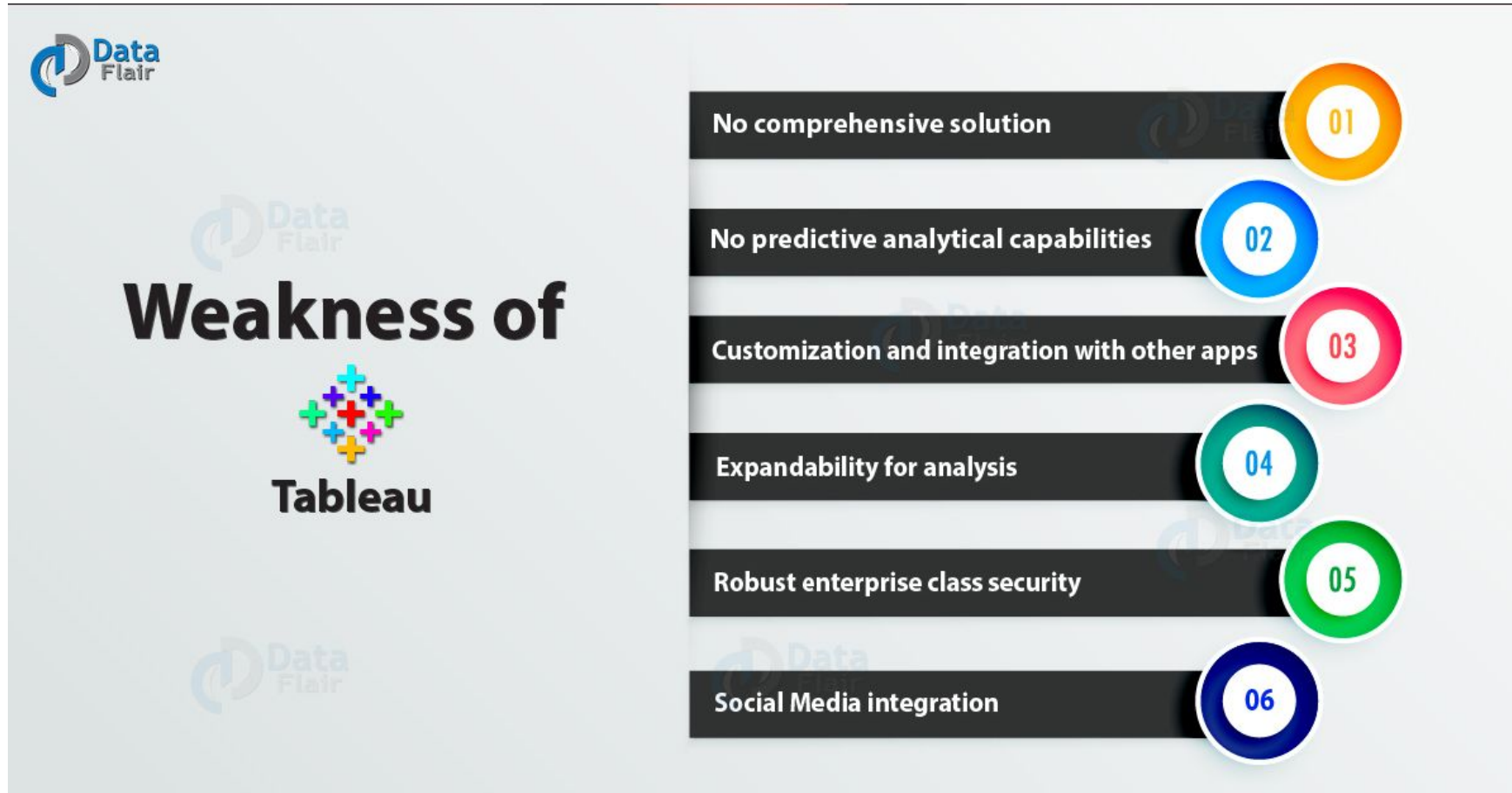


Tableau file types and extensions

- Tableau Workbook (.twb)
- Tableau Packaged Workbook (.twbx)
- Tableau Datasource (.tds)
- Tableau Packaged Datasource (.tdsx)
- Tableau Data Extract (.tde)
- Tableau Bookmark (.tbm)
- Tableau Map Source (.tms)
- Tableau Preferences (.tps)

Tableau Workbook (.twb)

- This file type is probably the most common.
- It is in XML format (try editing it in a text editor) and
- contains all the information on each sheet and dashboard that is contained within your workbook.
- It also includes data source connection information and any metadata you have created for that connection (see more on this below under .tds).

Tableau Packaged Workbook (.twbx)

- Whilst a Tableau Workbook (.twb) as described above holds all the information Tableau requires to draw your viz, it does not include the data itself.
- A Packaged Workbook however, combines the information in a workbook and bundles it with any local data – i.e. data that is not on a server.
- A .twbx will also include any custom images, as well as any custom geocoding you may have used in your work.
- The primary reason you would save your work as a Packaged Workbook is so that you can share it with other Tableau Desktop users, or for others to open using Tableau Reader.

Tableau Datasource (.tds)

- When you connect to your data for the first time, you may have a little bit of data ‘modelling’ to do – setting the right data types, changing default aggregations, setting default colours, creating some custom calculated fields etc etc.
- You are giving Tableau information about the data you will be using – you are setting up its ‘metadata’.
- When you want to connect to this data again, you don’t want to really go through all this data modelling a second time so instead you can save your metadata as a .tds file (again, it is saved in XML format) and connect to your data through this file instead.
- You could also distribute this file so that your colleagues have access to the nice formatting and custom fields you have worked to set up.

Tableau Packaged Datasource (.tdsx)

- Just like the fact that a .twb does not contain any of the data but a .twbx does, a .tds file only contains the information about the data, not the data itself.
- A Tableau Packaged Datasource (.tdsx), however, contains the data too.
- You would create this type of file instead of a .tds if you wanted to share the connection information with someone else who did not have access to the underlying data (for example if it was stored on your local machine)
-

Tableau Data Extract (.tde)

- When you connect to data using Tableau you can either connect ‘Live’ or you can extract the data into a .tde.
- Data extracts are used to radically improve performance, particularly when connecting to slow databases or slow files (e.g. CSVs), as well as enabling additional functionality (try doing a count distinct whilst connected live to Excel) and offline analysis.
- The primary disadvantage to using an extract is that your Tableau viz is no longer pointing to the ‘live’ data source – if that data source updates then your viz will not until you refresh the extract.

Tableau Bookmark (.tbm)

- A slightly lesser known Tableau file type is the Tableau bookmark.
- This file is a bit like an export of one single worksheet, which you can then import into another workbook to save you recreating the view from scratch.
- Tableau 8.1 introduces functionality to help copy and paste worksheets from one workbook to another
- so this file type may become used less but it can still be handy if you regularly use a particular view in many of the workbooks you create (a header page or appendix, for example)

Tableau Map Source (.tms)

- When plotting maps with Tableau, the software will connect to its mapping provider (Urban Mapping) to load the relevant map tiles in the background to plot your data points against.
- From the Map menu in Tableau Desktop, you have the option to add your own WMS server so that images from this source are loaded, rather than images from Urban Mapping.
- After you have added a new mapping source, you can share this set up with others by creating a distributing a Tableau Map Source file.

Tableau Preferences (.tps)

- The Tableau Preferences file can be used to create custom colour palettes so that using consistent colours (e.g your corporate colour schemes) across all your workbooks is made easier.
- This file is kept in your My Tableau Repository directory and is held in XML format.