

Tableau

Module 7

Organizing Data & Visual Analytics

Integration of Tableau with R & Hadoop

Tableau Prep



+ a b | e a u®

Agenda

Organizing Data & Visual Analytics

- Formatting Data
- Formatting Pane
- Trend and Reference Lines
- Forecasting
- K-means Cluster Analysis in Tableau
- Visual Analytics in Tableau
- Reference Lines and Bands
- Confidence Interval

Integration of Tableau with R & Hadoop

- Introduction to R Language
- Applications and Use Cases of R
- Deploying R on Tableau Platform
- Integration with Hadoop
- Learning R Functions in Tableau

Tableau Prep

- Introduction to Tableau Prep
- How does Tableau Prep help to quickly combine, join, shape and clean data for analysis?
- Create smart experiences with Tableau Prep & get deeper insights into your data with great visual experience
- Intergradation of Tableau Prep and Tableau analytical workflow
- Seamless process from data preparation to analysis with Tableau Prep

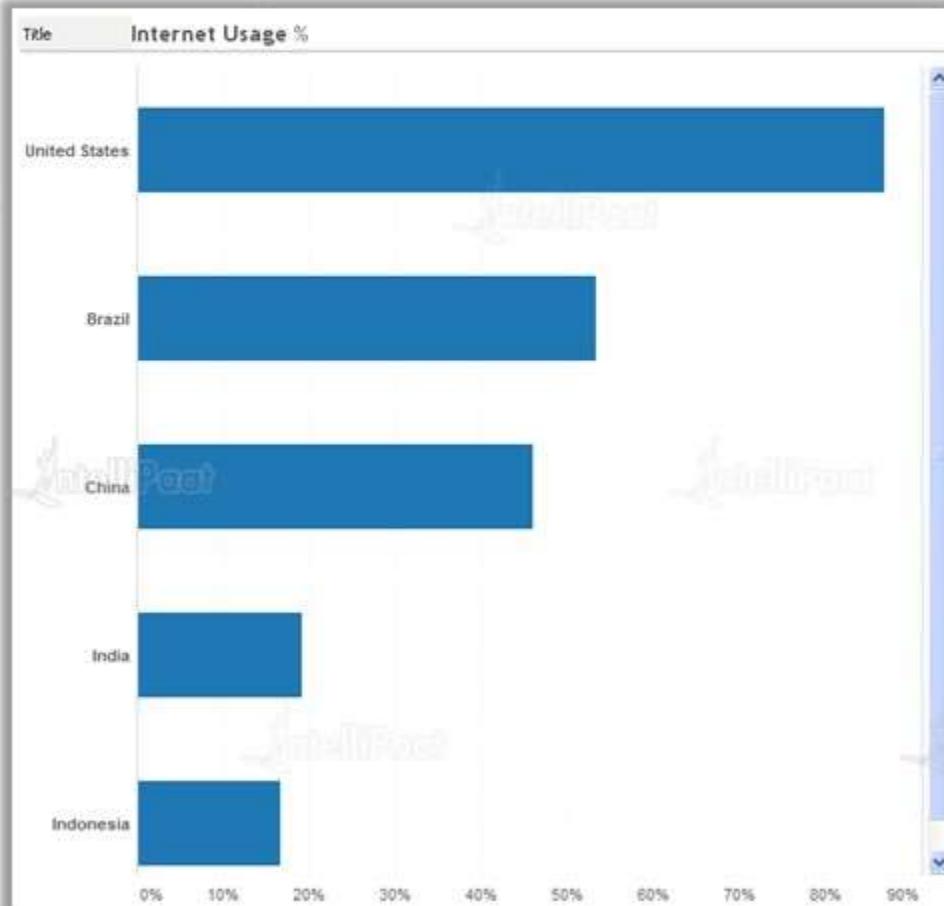


Organizing Data & Visual Analytics

Formatting Data

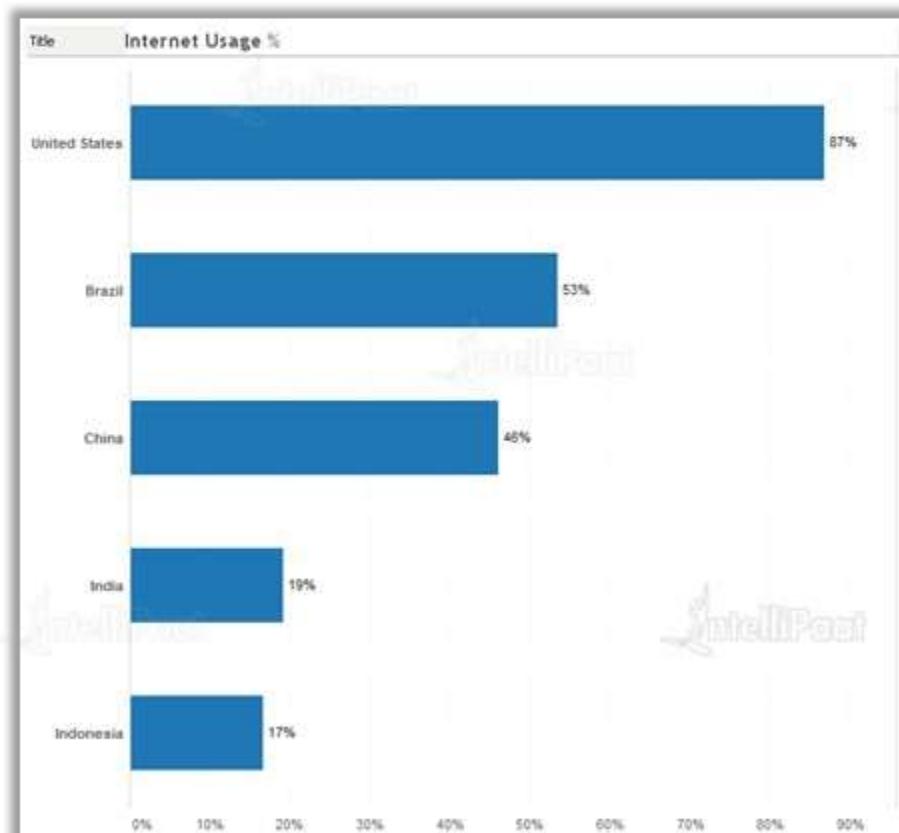
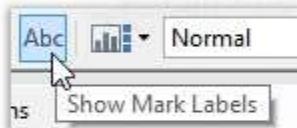
Formatting Labels

- To understand how to format labels, let's consider a simple horizontal bar chart as shown below.



Formatting Labels

- Click on the **Show Mark Labels** button in the toolbar, and by default Tableau will label your marks with the primary measure in the **Row** shelf as shown:



Formatting Labels

- Click on the “Label” button on the Marks card, you will get a label option menu.



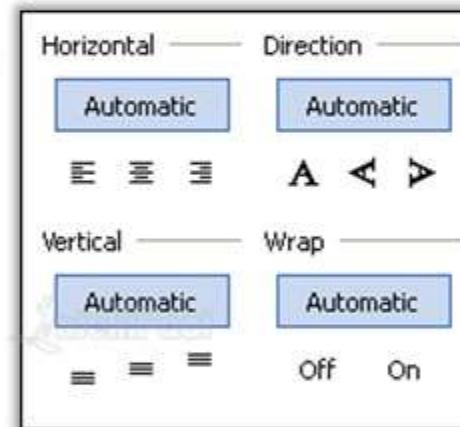
Formatting Labels



Font Menu



Alignment Menu

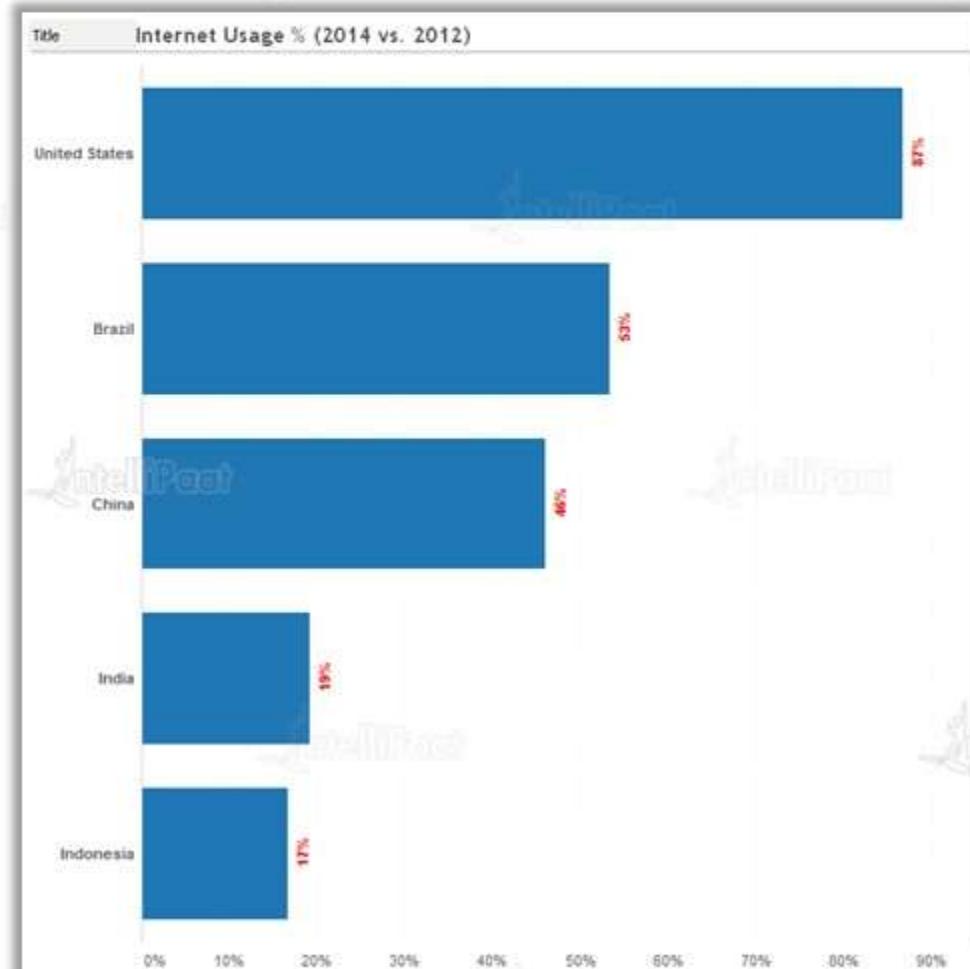


Let's make our labels bold and red, and let's also change the direction.



Formatting Labels

- The output is as shown below:



Formatting Annotations

- Annotations are added to a visualization to call out a specific mark, a specific point, such as a location on a map, or a specific area, such as a cluster of scatter marks.
- There are three types of annotations:**

Mark

- Associated with the selected mark. This option is only available if a data point (mark) is selected.

Point

- Select this option to annotate a specific point in the view.

Area

- Select this option to annotate an area in the viz, such as a cluster of outliers or a targeted region.

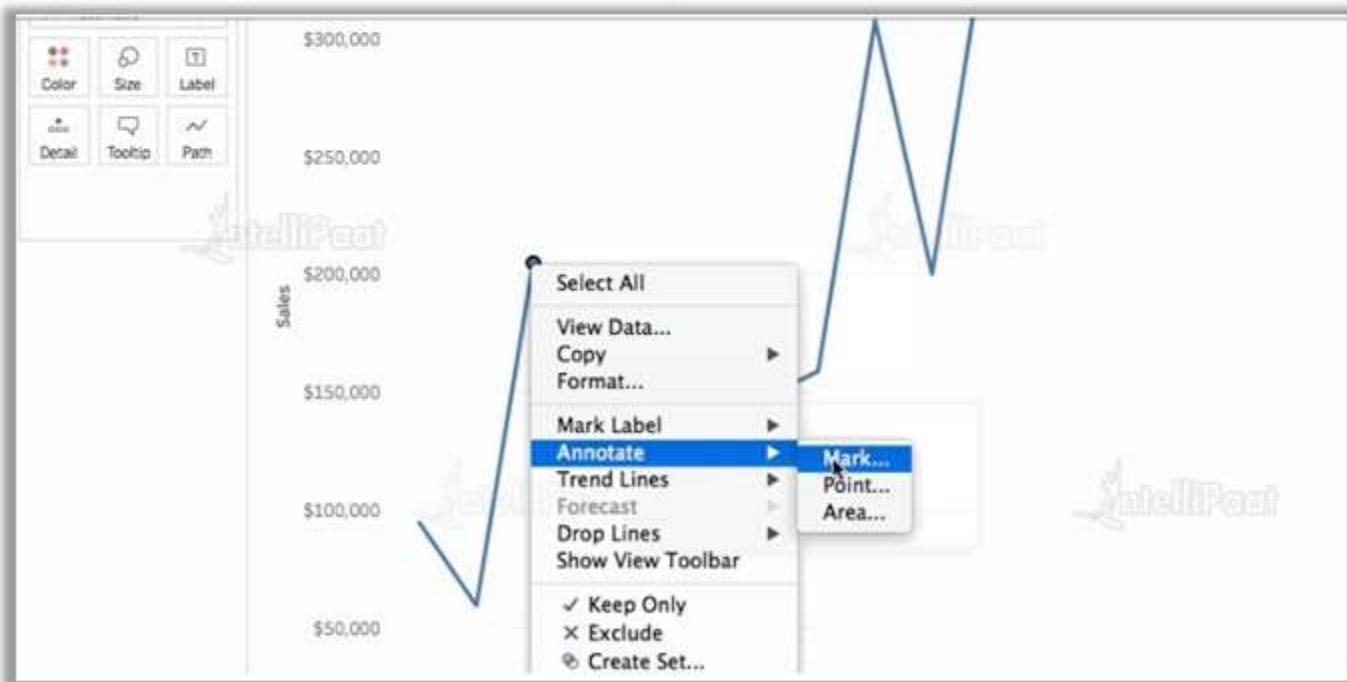
Formatting Annotations

- Follow the below steps to add annotations to visualization:

Step 1:

Right-click on a data point or a spot on the viz where you want to add an annotation.

Select **Annotate**, and then select the type of annotation you want to add.

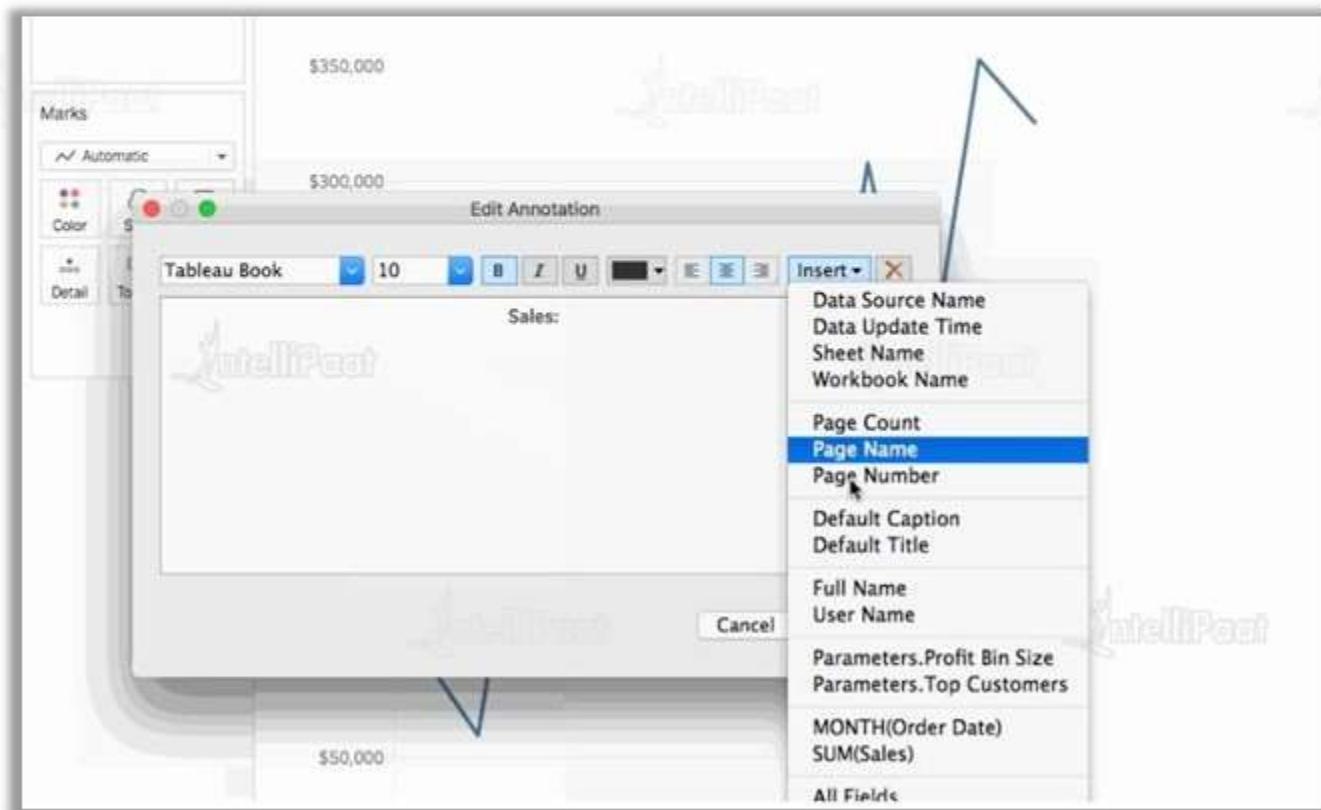


Formatting Annotations

Step 2:

You will get an annotation dialog box. Type the text you want to show in the annotation.

You can insert the dynamic variables using the insert menu.

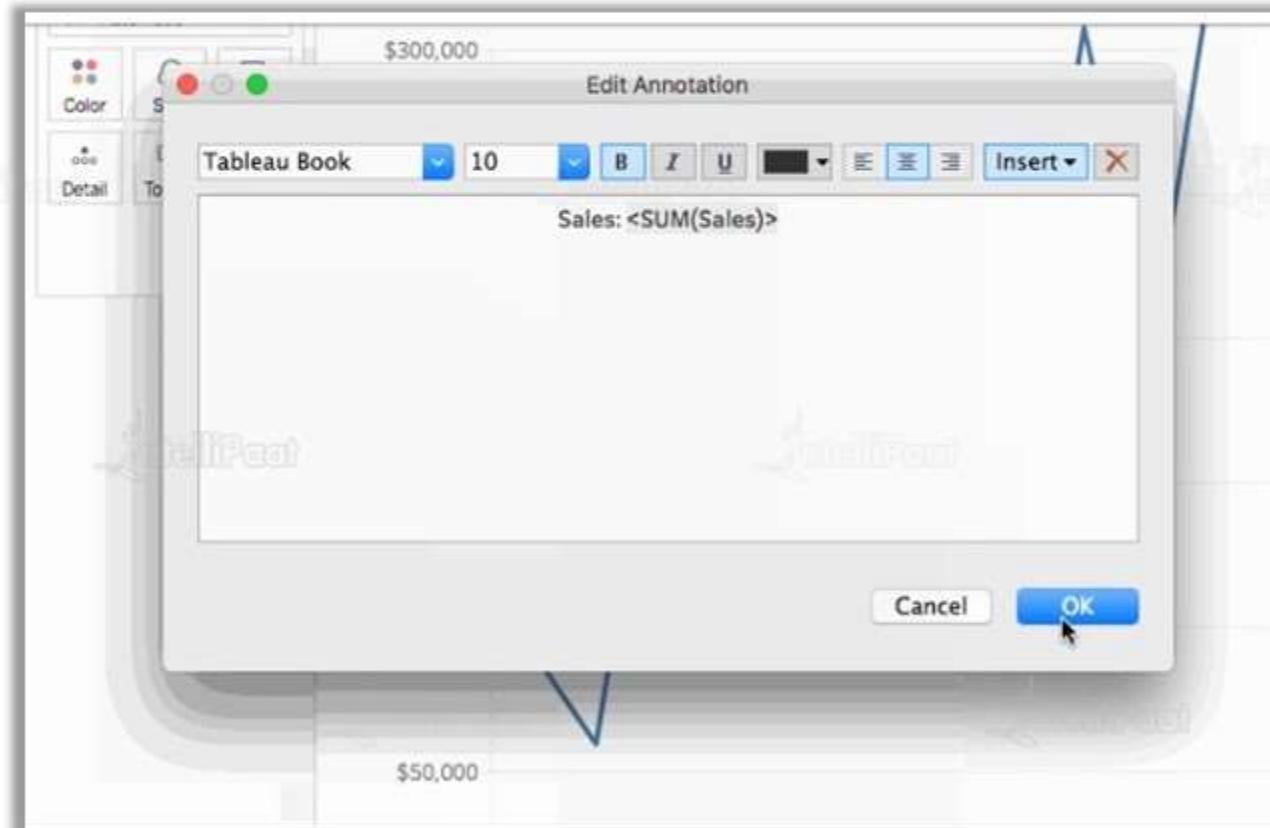


Formatting Annotations



Step 3:

Click on Ok



Formatting Tooltips



- Using the Tooltip, one can add additional information to your view.
- Tooltip can be populated with as much information as needed including dynamic fields.
- These are mainly used in building interactivity and reinforcing the story for views.
- Tooltips also offer convenient tools to quickly filter or remove a selection, select marks that have the same value or view the underlying data.

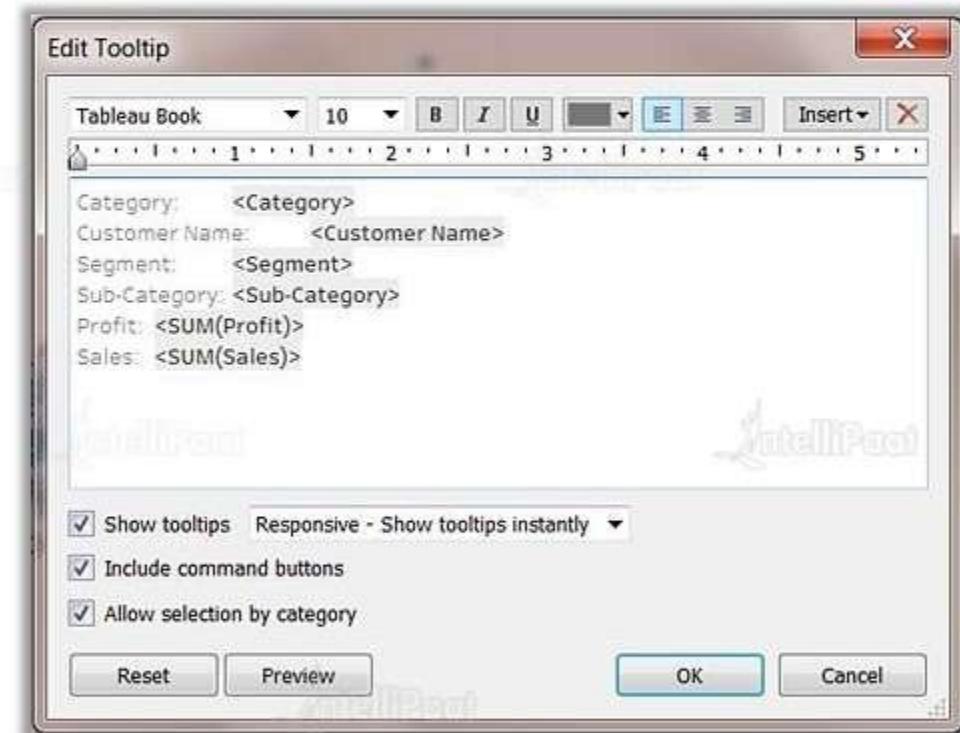
Formatting Tooltips

- Below are the steps to edit the Tooltip:

Step 1:

Select **Worksheet > Tooltip**.

Tooltips are specified for each sheet and can be formatted using the formatting tools provided along the top of the Edit Tooltip dialog box. Use the Insert menu at the top of the dialog box to add dynamic text such as field values, sheet properties and more.



Formatting Tooltips



Step 2:

Use the formatting tools along the top of the Edit Tooltip dialog box.

Step 3:

Use the Insert drop-down list at the top of the dialog box to add dynamic text such as field values, sheet properties and more.

Step 4:

Select the Include command buttons option to show filtering and view data options in the tooltip.

Step 5:

Check the Allow selection by category box to select marks in a view that have the same value by clicking on a discrete field in a tooltip.

Visual Analytics

At Tableau, we believe in using the visual to drive your analysis.

When you get observable feedback as you analyze, the power of exploration is in your hands. When you build an analysis and learn from it simultaneously, opportunities for investigation present themselves.

Tableau's visual analytics makes asking and answering questions of your data intuitive, even as those questions naturally grow in complexity.



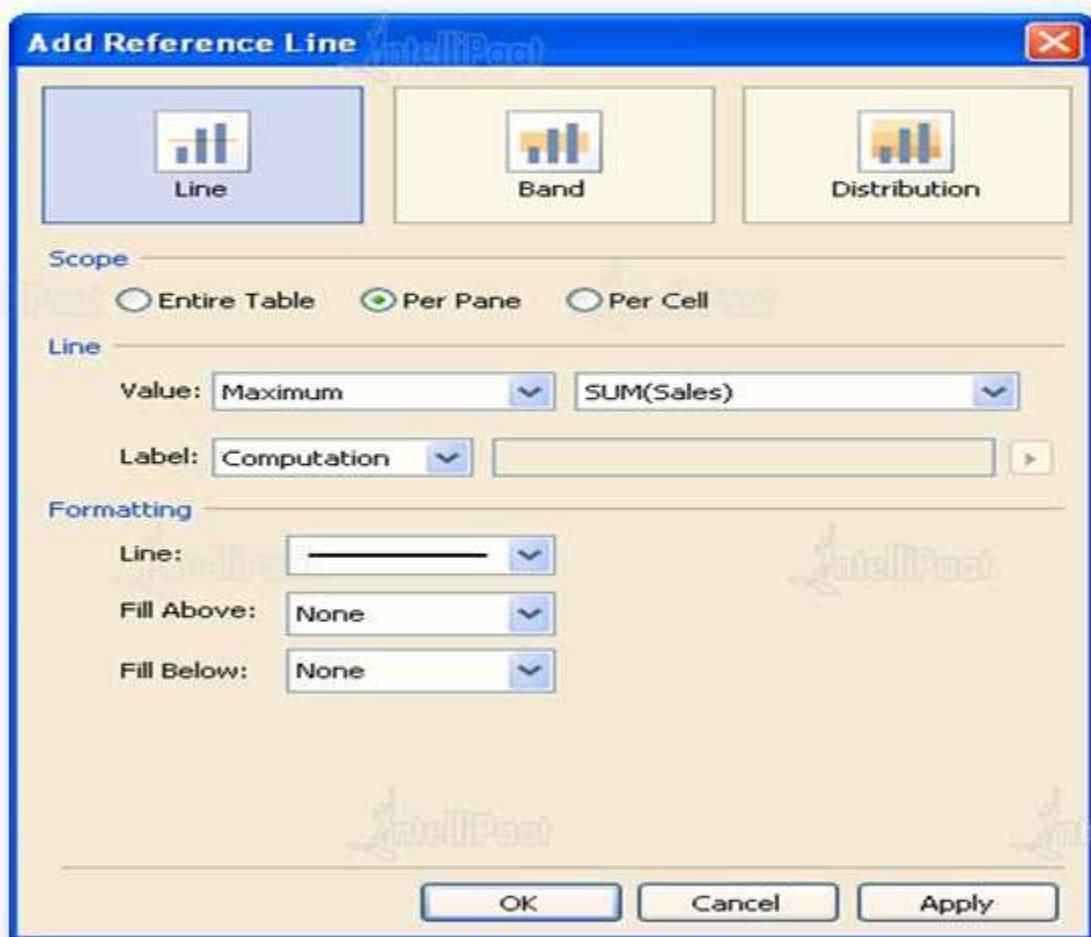
We believe that to help everyone see and understand data, we need to provide rich capabilities for users of all levels of technical ability.

From simple go-to metrics to advanced analytic techniques, Tableau provides a flexible front-end for data exploration with the necessary analytical depth for data scientists.

By leveraging sophisticated calculations, R and Python integration, rapid cohort analysis, and predictive capabilities, data scientists can conduct complex, quantitative analyses in Tableau and share visual results to facilitate better understanding and collaboration with data.

Reference Lines & Bands

- Reference line is typically used to mark a specific value or region on an axis.
- Tableau lets you add an unlimited number of reference lines.
- You can add reference lines using the Add Reference Line dialog box.
- There are three types of reference lines: line, band and distribution.



Reference Lines & Bands



[Picture 1]



[Picture 2]

Line: Adds a line at a constant or computed value on the axis. Computed values can be based on a specified field (Pic 1).

Band: Shades an area behind the marks in the view between two constant or computed values on the axis (Pic 2).

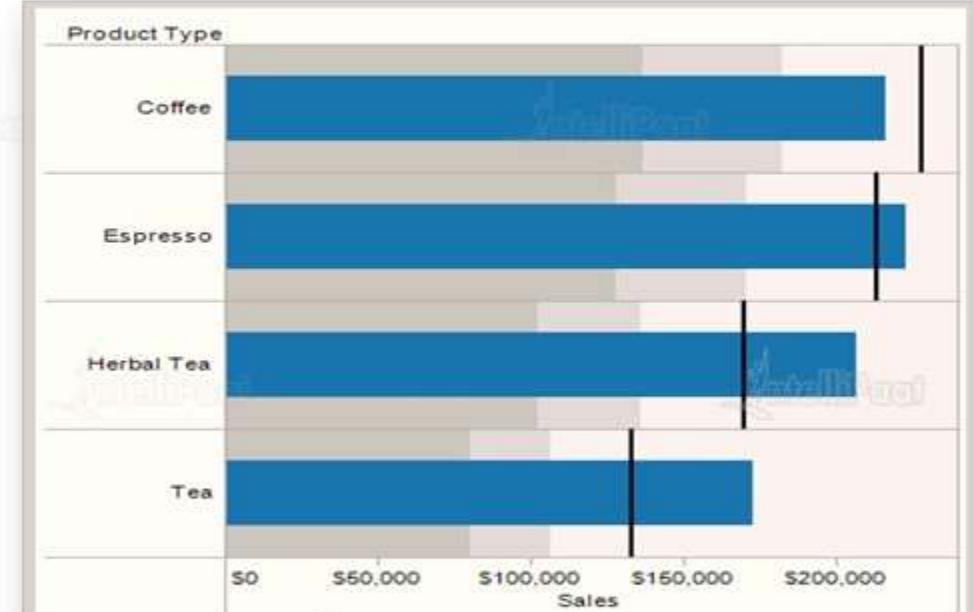
Reference Lines & bands



Adds a gradient of shading to indicate the distribution of values along the axis.

Distribution can be defined by confidence interval, percentages, percentiles, quantities or standard deviation.

This type of reference line is used to create bullet charts.

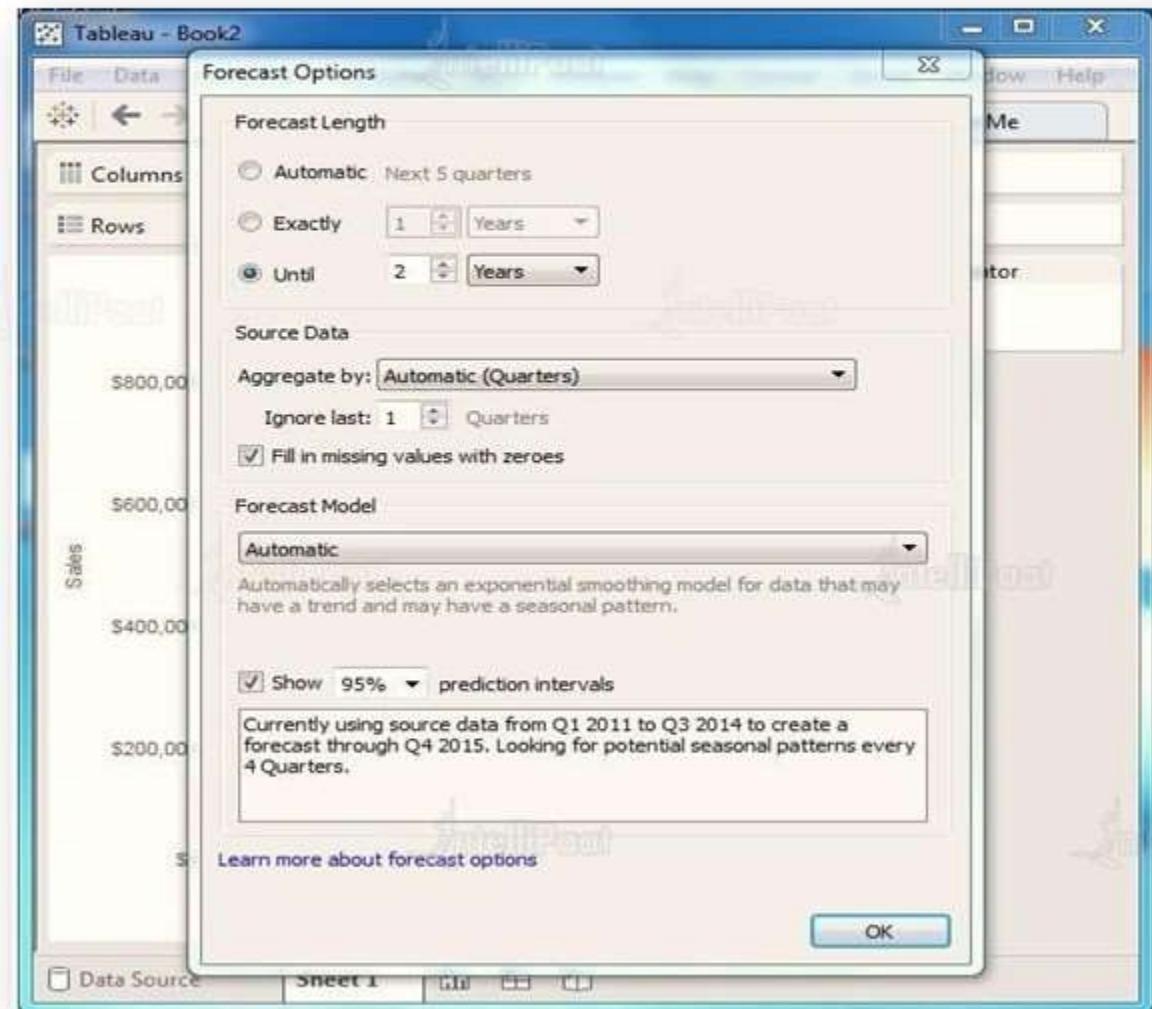
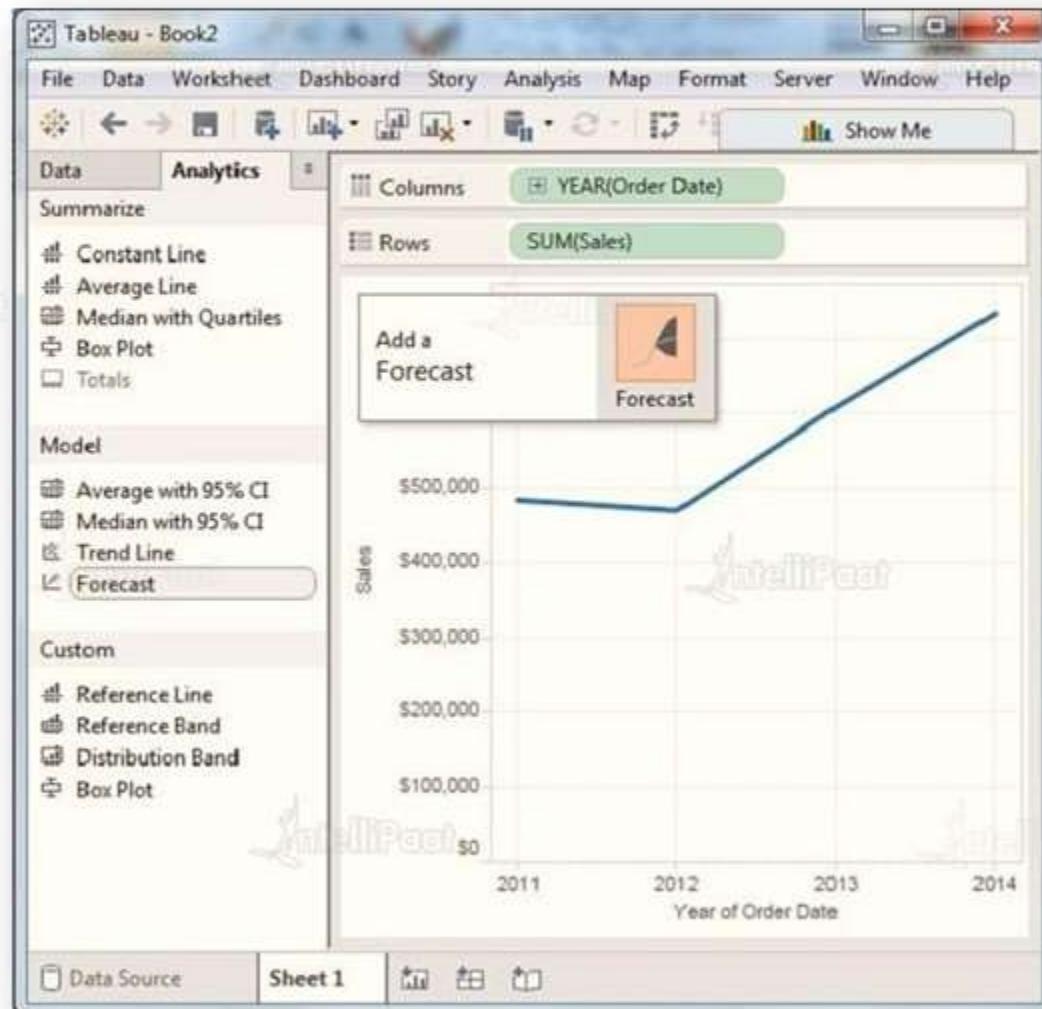


Forecasting

- Forecasting is about predicting the future value of a measure. There are many mathematical models for forecasting. Tableau uses the model known as **exponential smoothing**.
- In exponential smoothing, recent observations are given relatively more weight than older observations. These models capture the evolving trend or seasonality of the data and extrapolate them into the future. The result of a forecast can also become a field in the visualization that is created.
- Tableau takes a time dimension and a measure field to create a forecast.



Forecasting



Forecasting

- 1) Past events + present events --> identify patterns --> applies them to the future
- 2) The more data you have from the past, the more accurate the forecasting will be

**MOVING AVERAGE

- smooths out the fluctuations in data by calculating an average within a given window of time period

Types

- Simple moving average -> each data point is equally weighted
- Exponential moving average -> recent data is given more weight

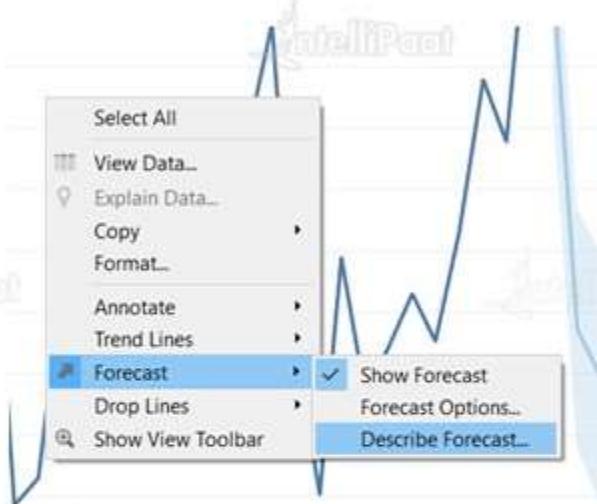
**Create a custom forecasting line

- * The more data you have from the past, the more accurate the forecasting will be

**Tableau's inbuilt forecasting tool

- Exponential Smoothing
- TREND - a direction in which the data goes
- SEASONALITY - a pattern that repeats itself over a given time span

Describe Forecasting



Describe Forecast

Summary Models

Options Used to Create Forecasts

Time series: Month of Order Date
Measures: Sum of Sales
Forecast forward: 13 months (December 2018 – December 2019)
Forecast based on: January 2015 – November 2018
Ignore last: 1 month (December 2018)
Seasonal pattern: 12 month cycle

Sum of Sales

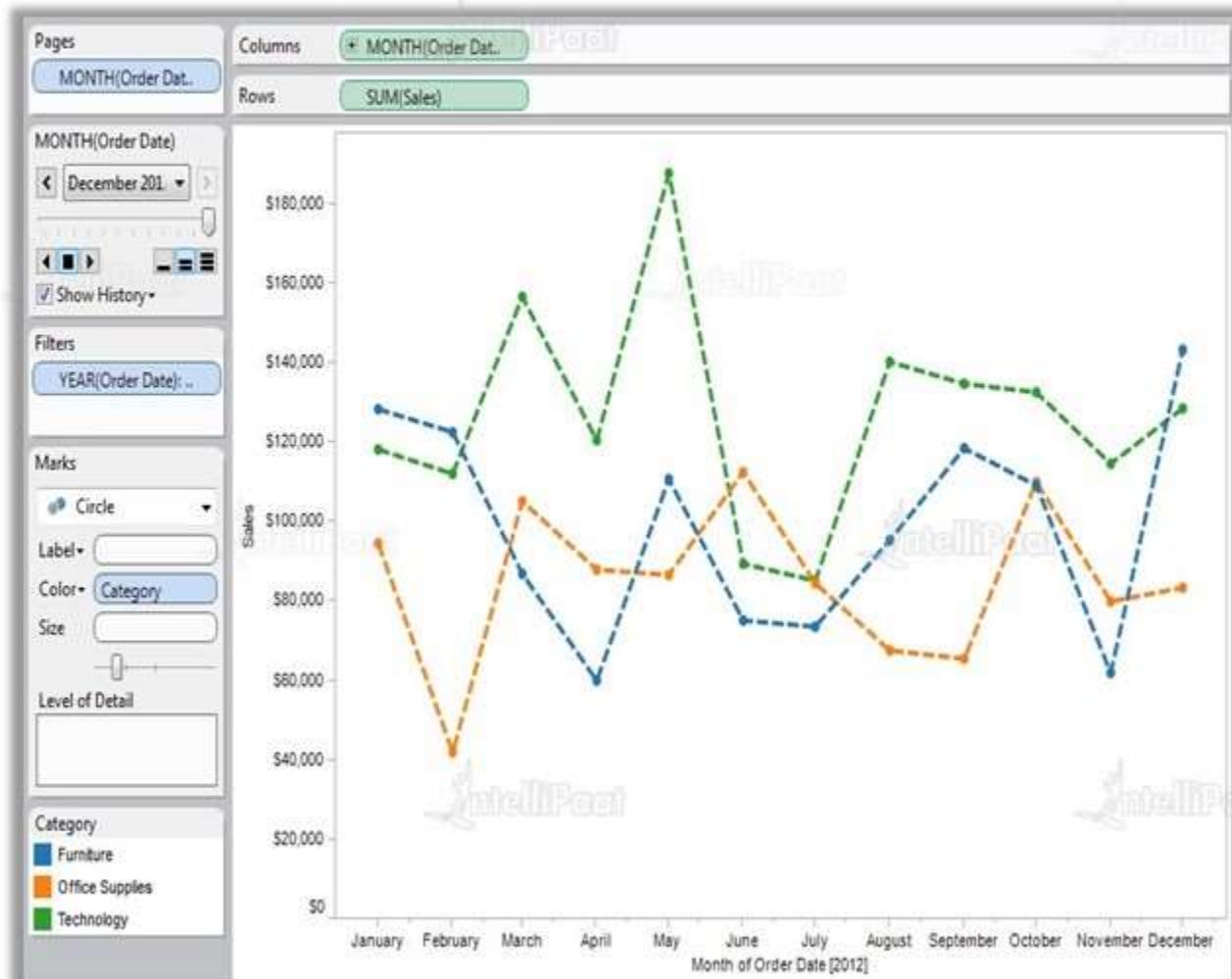
Initial	Change From Initial	Seasonal Effect	Contribution
December 2018	December 2018 – December 2019	High Low	Trend Season Quality
103.260 ± 18.670	0	November 2019 37.518 February 2019 -31.547	0.0% 100.0% Ok

Show values as percentages

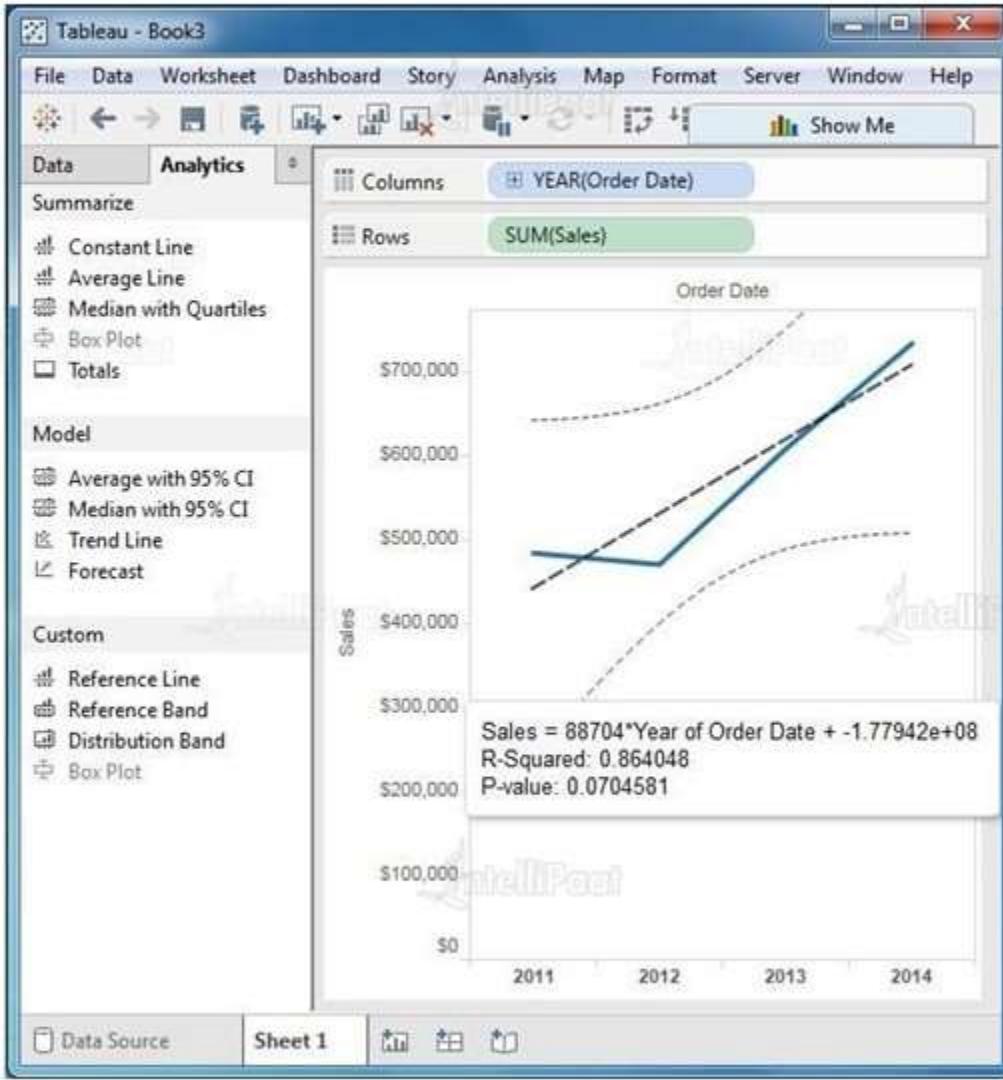
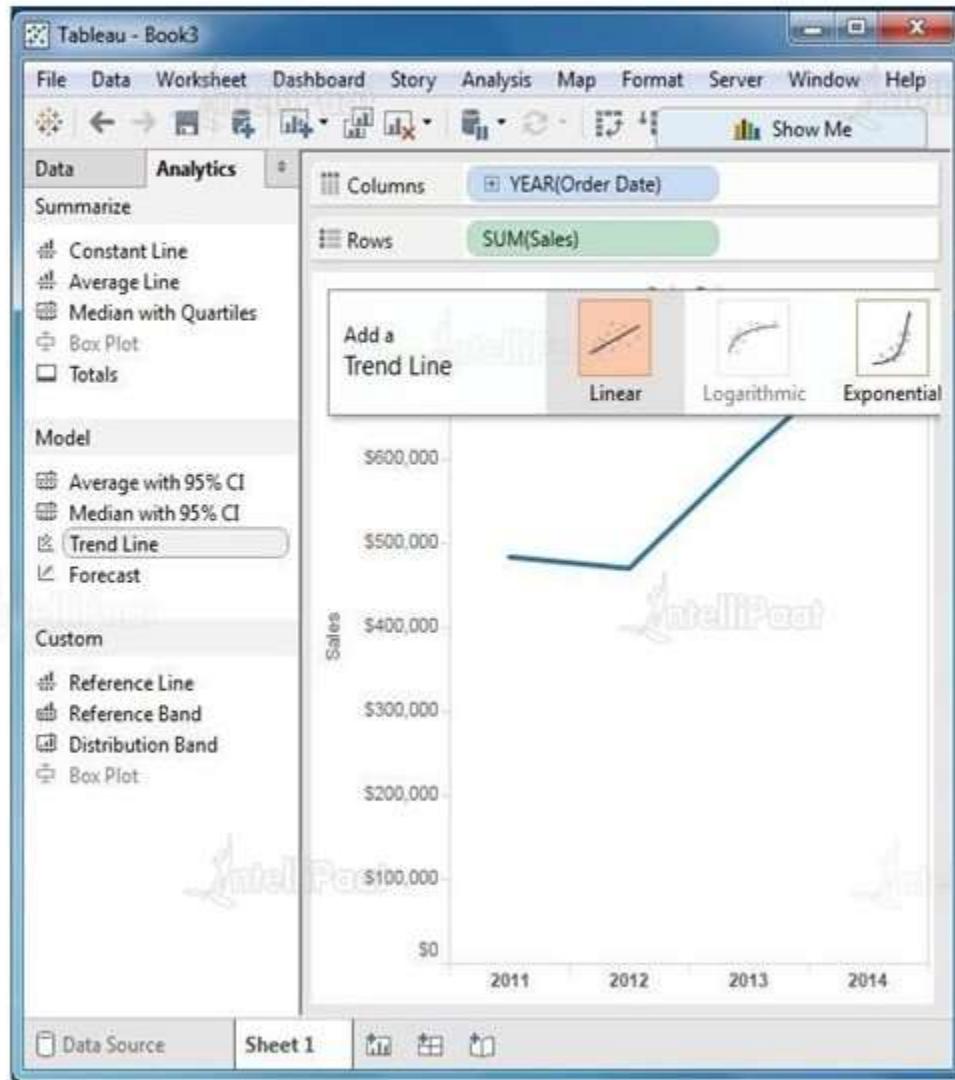
[Copy to Clipboard](#) [Learn more about the forecast summary](#) [Close](#)

Trend Lines

- Trend lines are used to predict the continuation of a certain trend of a variable. It also helps to identify the correlation between two variables by observing the trend in both of them, simultaneously.
- There are many mathematical models for establishing trend lines. Tableau provides four options. They are linear, logarithmic, exponential and polynomial.
- Tableau takes a time dimension and a measure field to create a trend line.



Trend Lines



Clustering

Clustering is a powerful new feature in Tableau 10 that allows you to easily group similar dimension members.

Clustering helps you create statistically-based segments which provide insights into how different groups are similar and how they are performing compared to each other.

The clustering function uses k-means to cluster the dataset based on the measures and dimensions that are dragged into the new *Cluster Variables* card.

The new clustering feature automatically groups together similar data points.

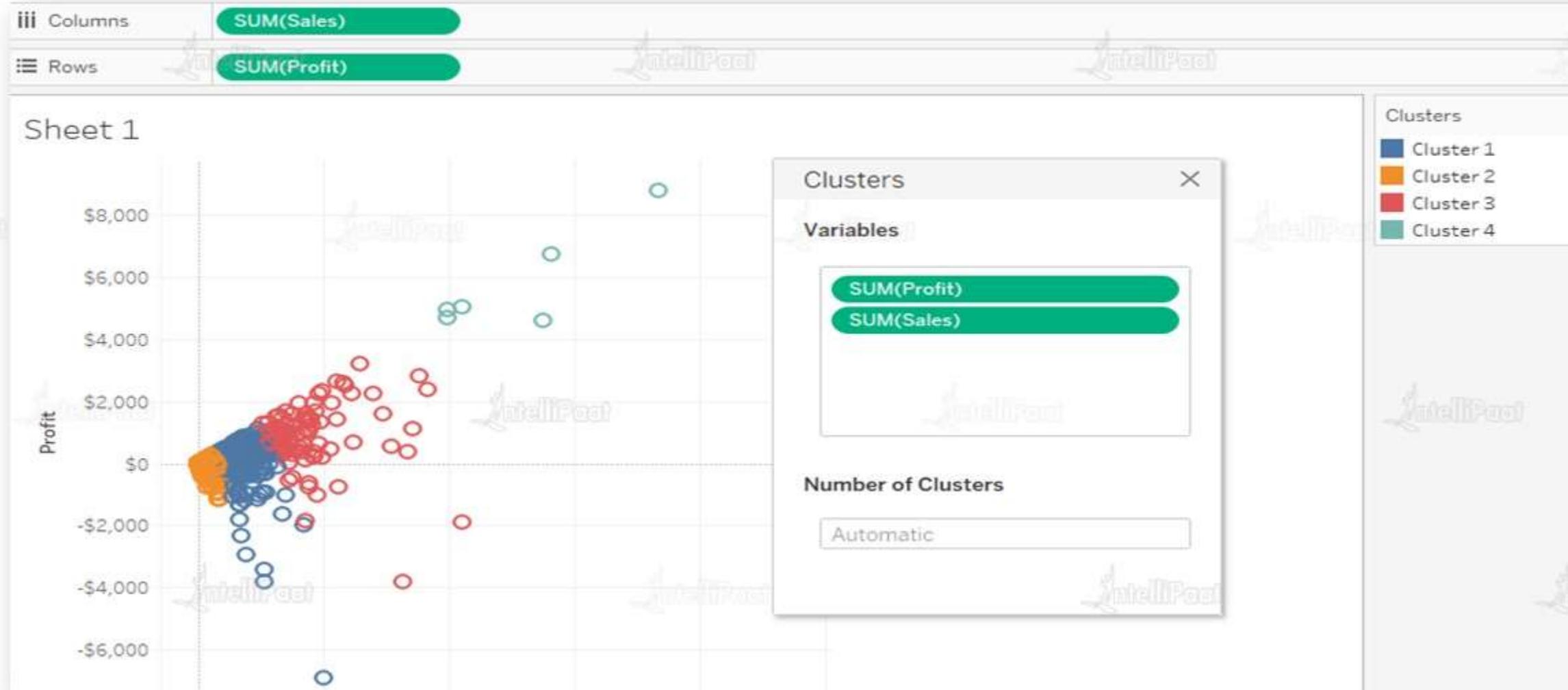
We can use clustering on any type of visualization, from scatter plots to text tables and even on maps.

To create clusters in your sheet, just drag clustering from the Analytics pane into the view.

To see how different inputs change clustering results, we can experiment by dragging them in and out of the dialog and see the results in the real time.

We can specify any number of clusters in the report.

Clustering

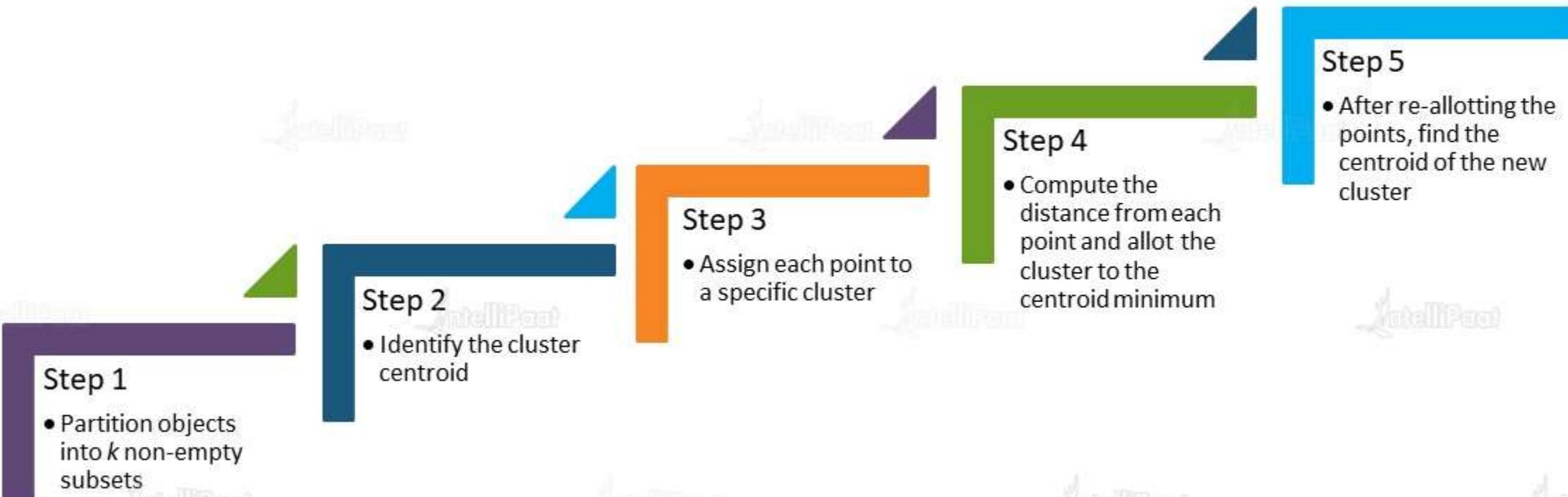


K-means Clustering

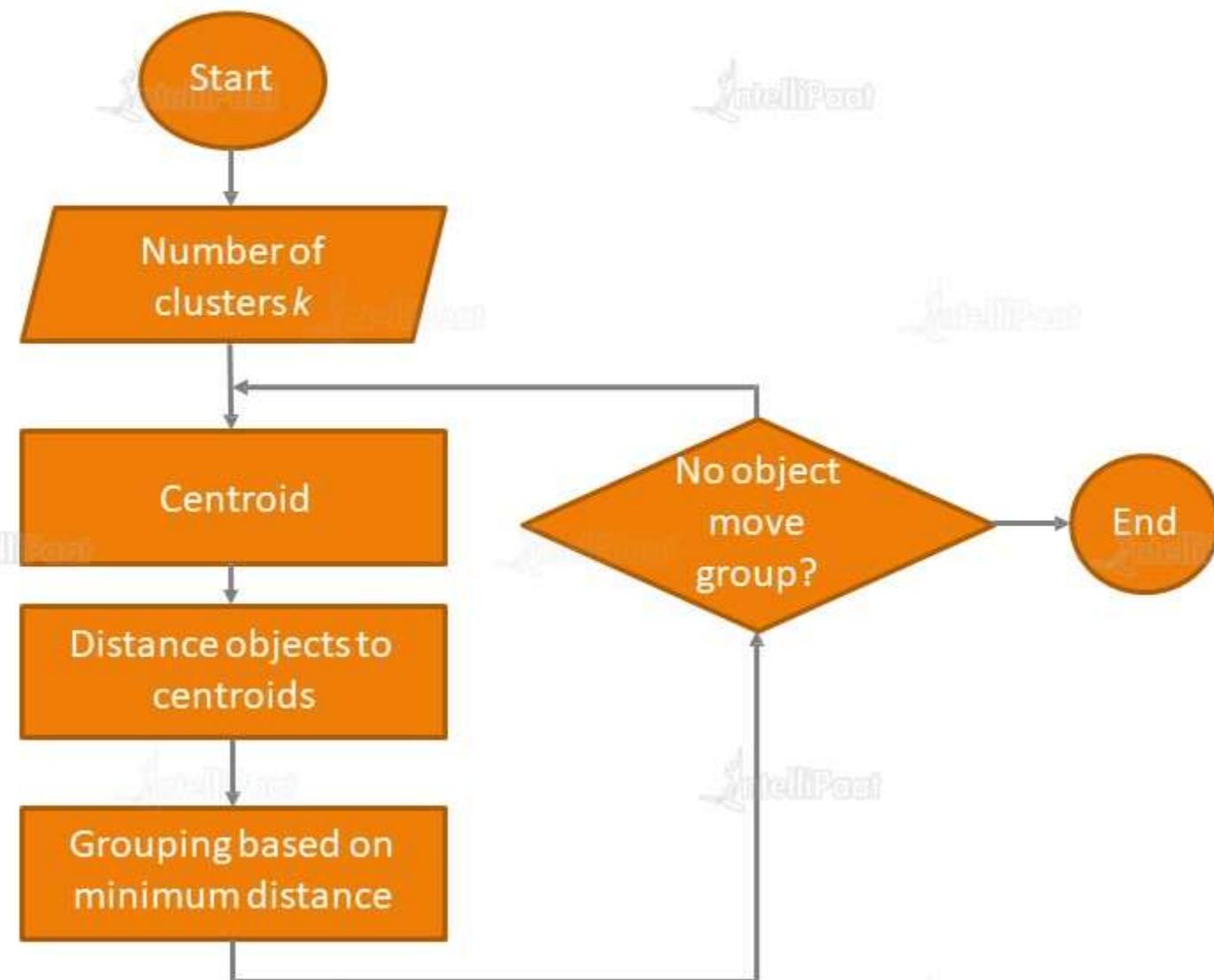


- K-means clustering is an algorithm, the main goal of which is to group similar data points into a cluster.
- The number of groups or clusters are represented by k . It is an algorithm that runs iteratively to form a cluster.
- K-means clustering runs on distance calculations, which uses “Euclidean Distance” to calculate the distance.

K-means Clustering



K-means Clustering



K-means Clustering

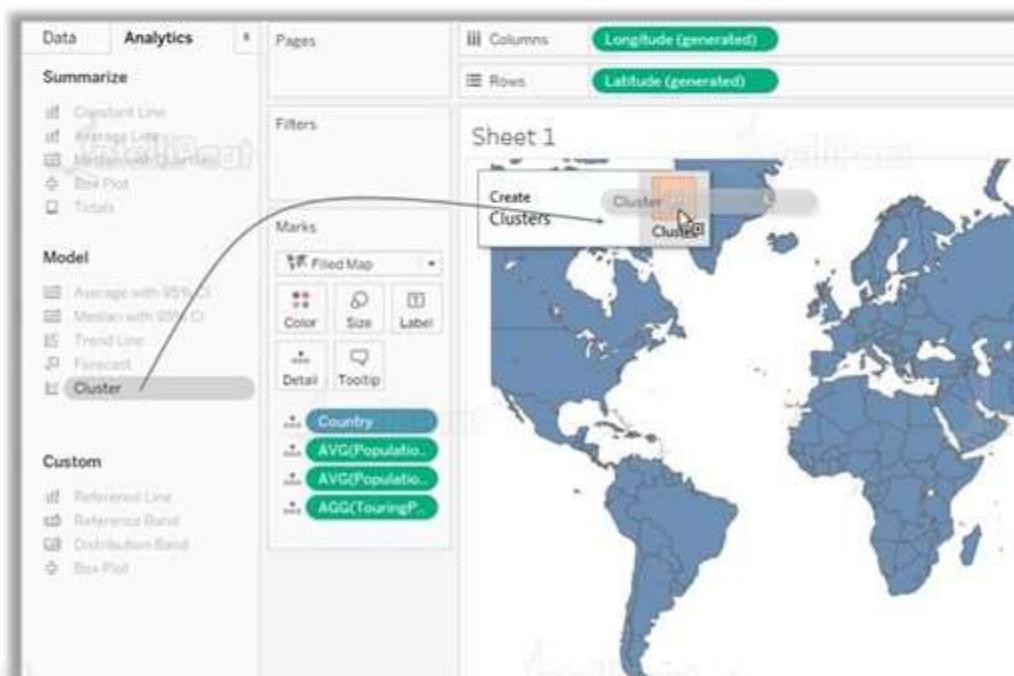
- To find clusters in a view in Tableau, follow these steps:

Step 1:

Create a View.

Step 2:

Drag Cluster from the Analytics pane into the view, and drop it on in the target area in the view.

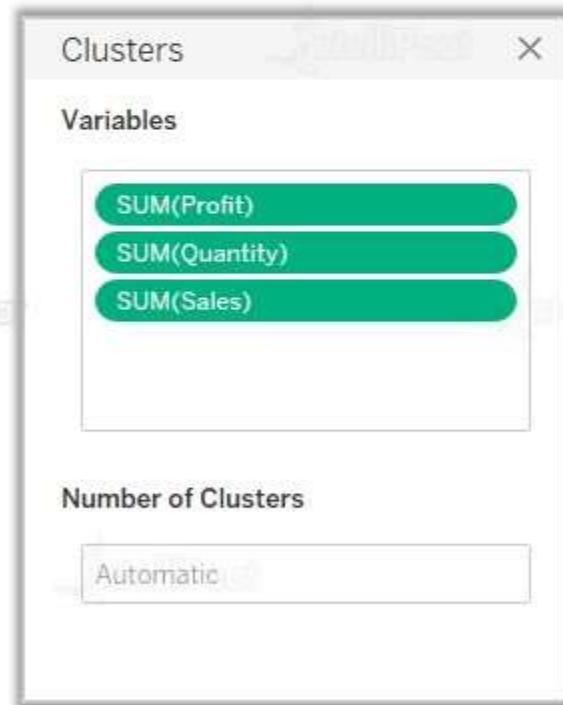


K-means Clustering

Step 3:

When you drop a cluster, Tableau creates a cluster group on color.

Tableau displays the Clusters dialog box, where you can customize the cluster.



K-means Clustering

Step 4:

Drag new fields from the Data pane into the Variables area of the Clusters dialog box.

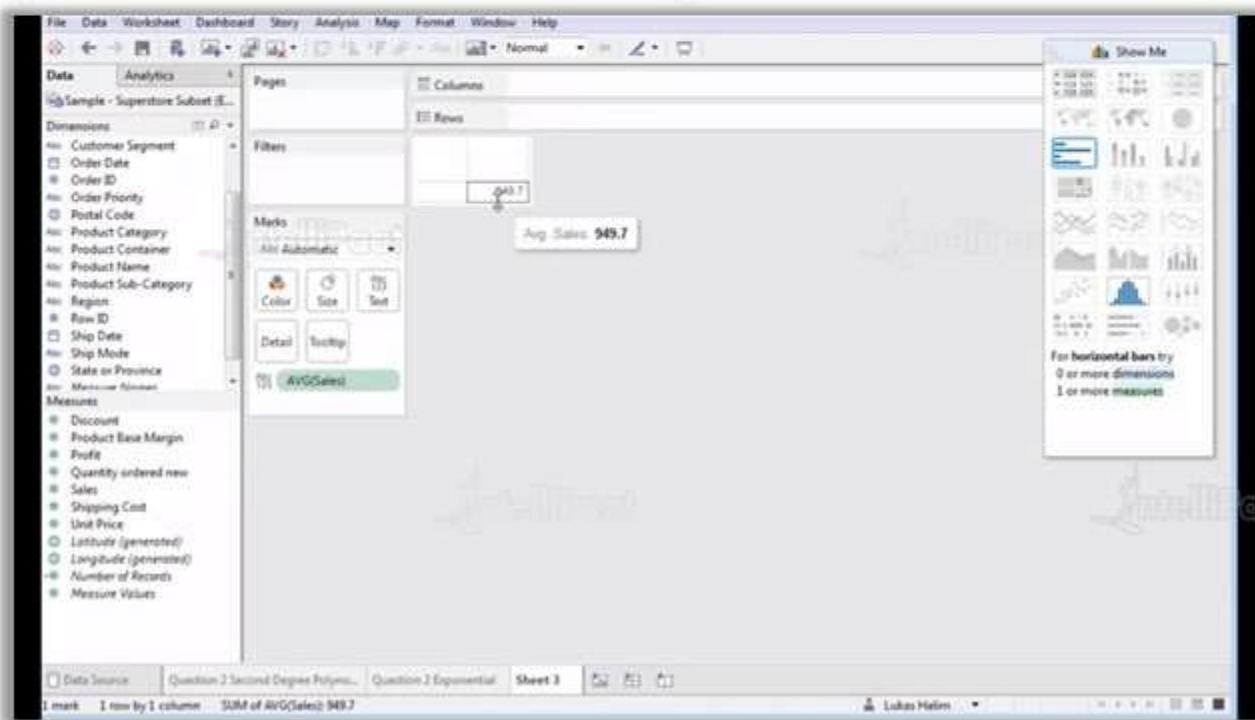
Step 5:

Once customizing is done, click on the “Close” icon.



Confidence Interval

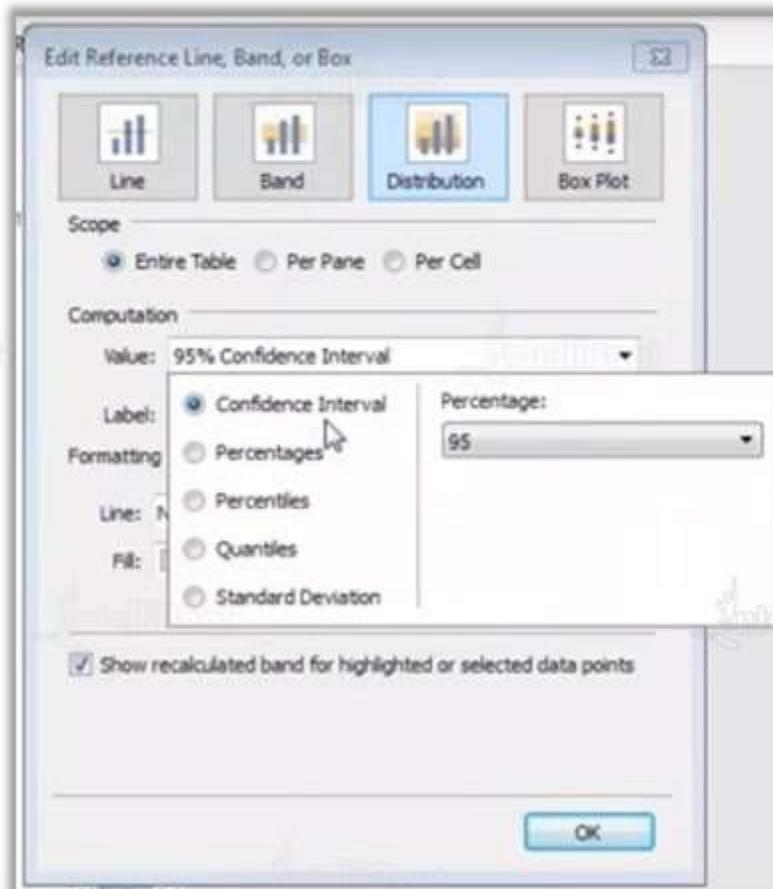
- Tableau can display confidence intervals by turning on the option, and Tableau does the calculation of lower and upper confidence intervals. Let's take an example and see how it works. Consider a sales report for which the true population mean should be within the interval 95%. Here, the average value of sales at present is 949.7.



Change it to bar chart

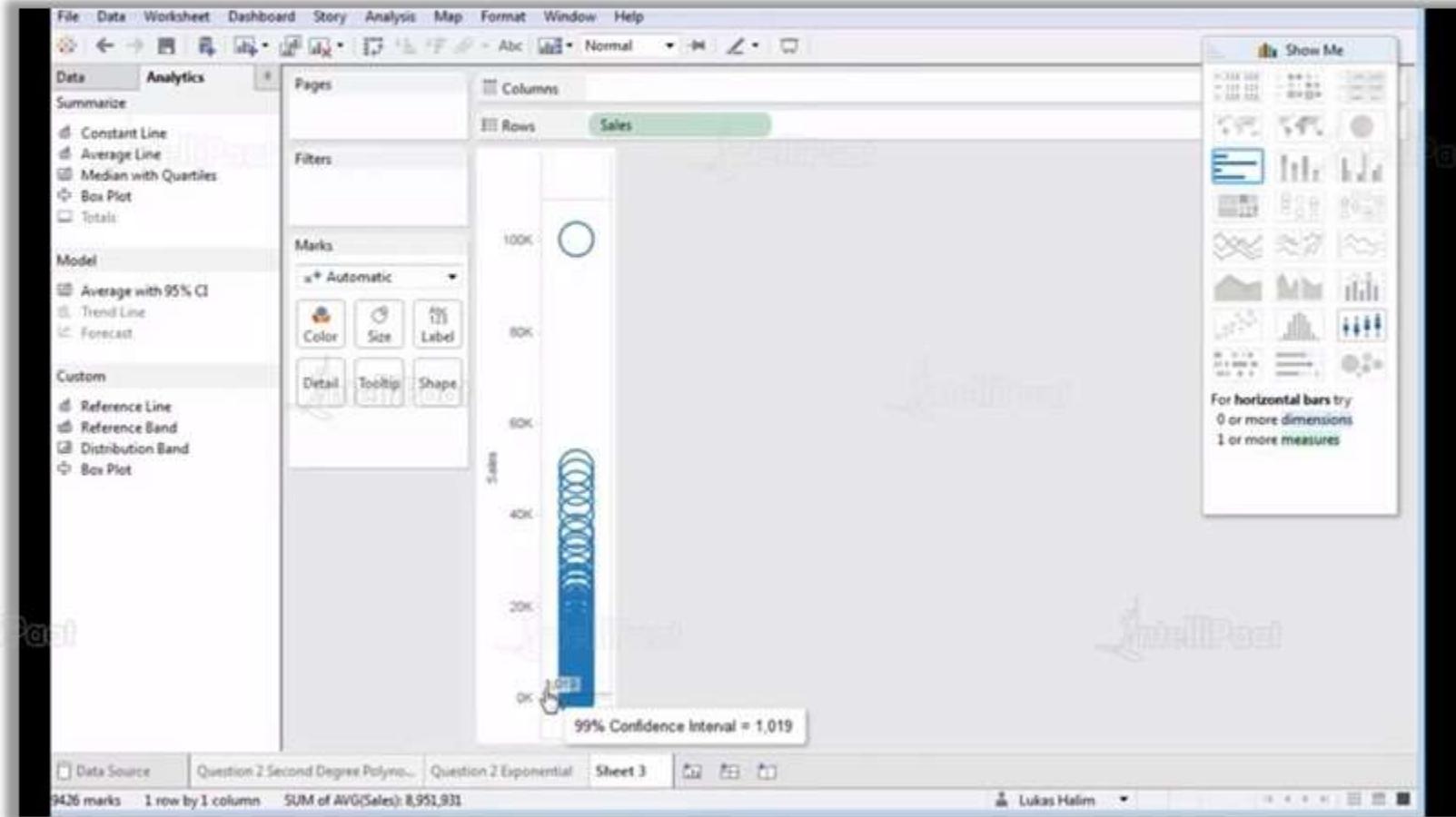
Confidence Interval

- Choose disaggregate from analysis and switch to analytics.
- Next choose Distribution band and select Confidence Interval in the Value column. Also, choose Percentage as 99% and Label as Value.



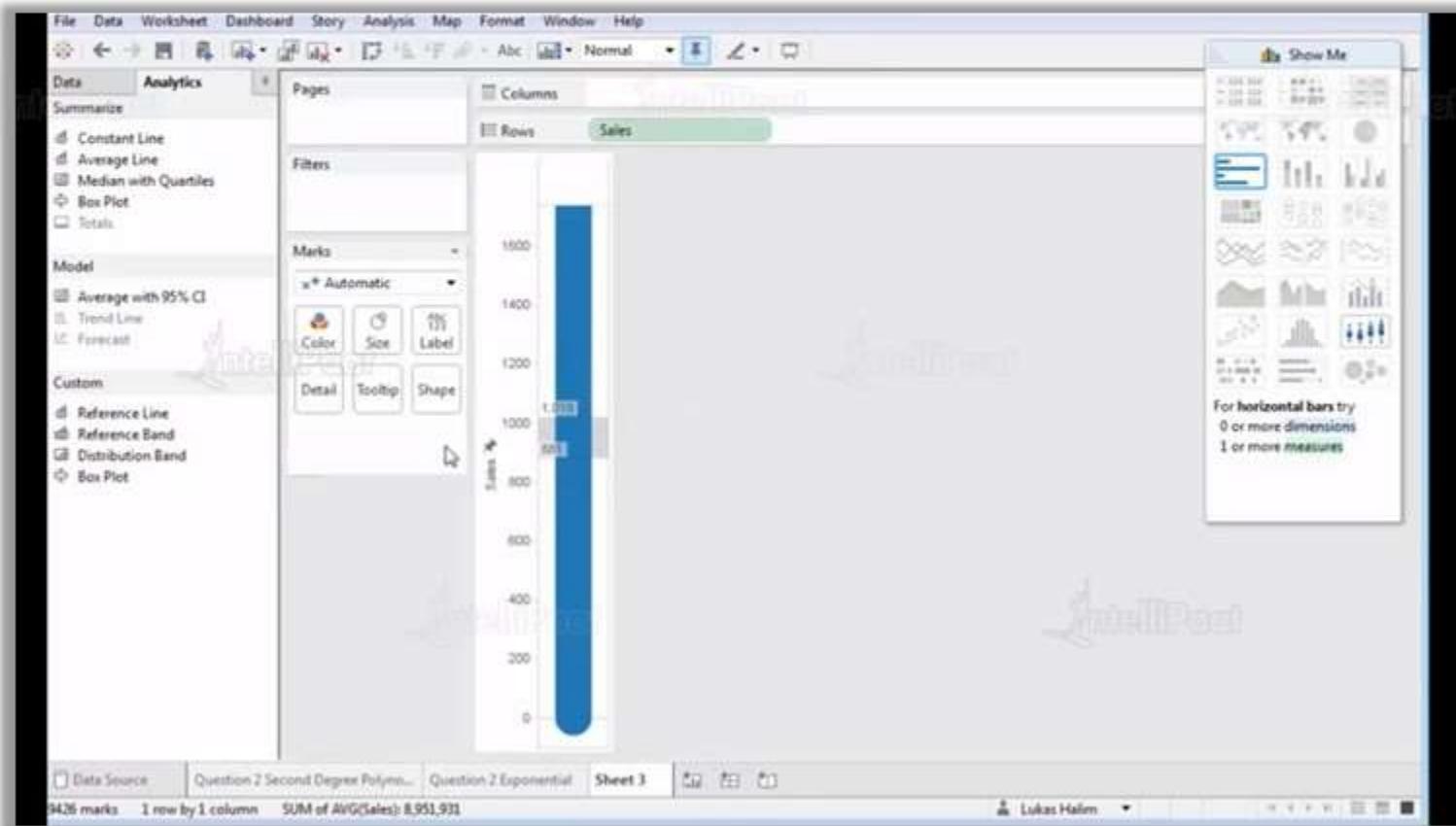
Confidence Interval

- When you mouseover on the graph you can see the confidence interval.



Confidence Interval

- This image shows that 99% of the time, our true population mean time our true population mean should be between 881 and 1019.



Integration of Tableau with R & Hadoop

Introduction to Tableau & R Integration



- What are the benefits of using Tableau and R integration?

Accessing R packages or functions for quantitative analysis



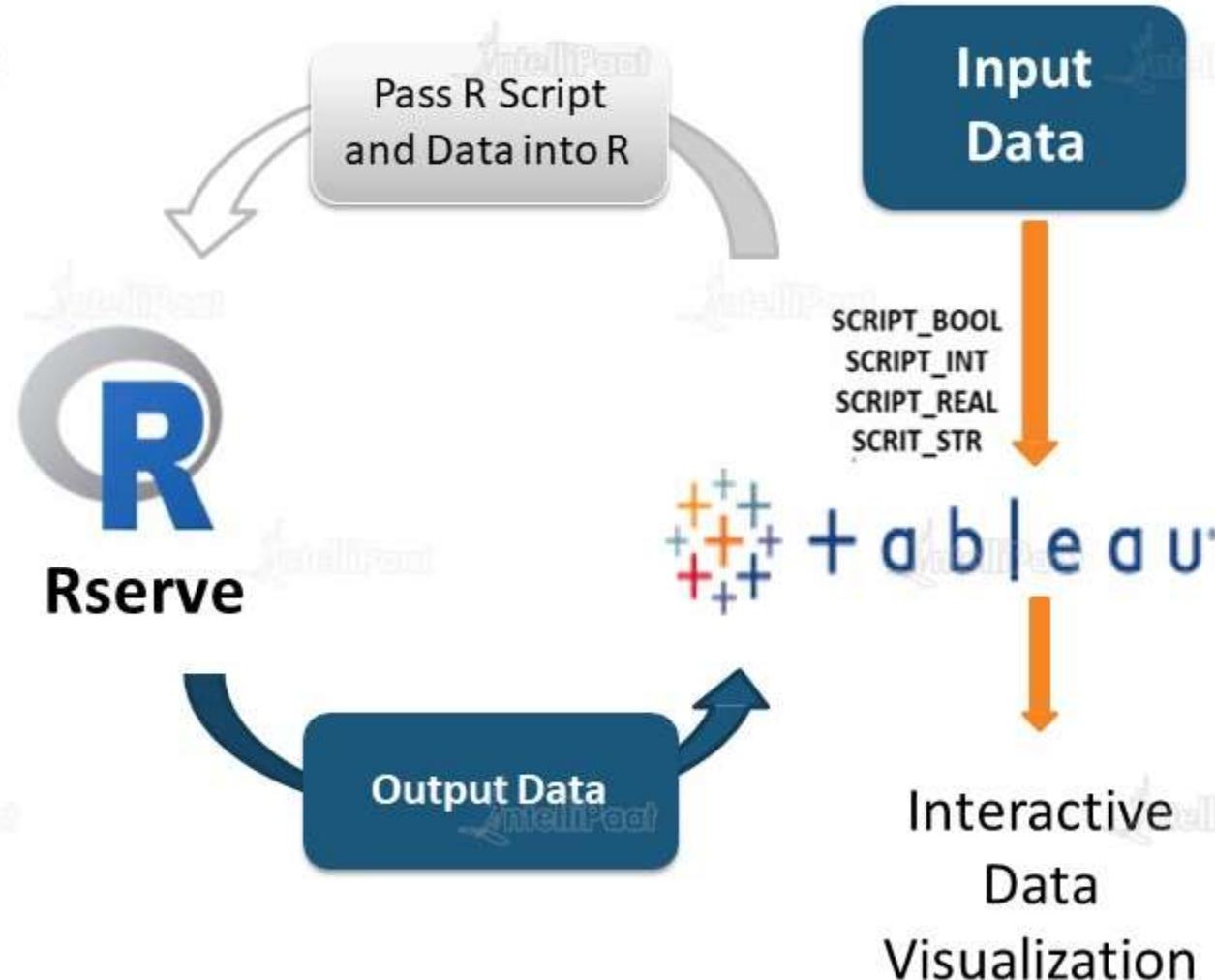
Taking advantage of Tableau's data visualization capacities

Introduction to Tableau & R Integration



- How is Tableau integrating with R?
 - Install Rserve Package
 - Create Calculation Field in Tableau
 - Use one of the four functions:
 - SCRIPT_BOOL: Return a Boolean
 - SCRIPT_INT: Return an Integer
 - SCRIPT_REAL: Return a Real
 - SCRIPT_STR: Return a String

```
install.packages("Rserve")
library(Rserve)
Rserve()
```



R & Python



Why R and
Python?

- R and Python provides a powerful way to do statistical analysis on large sets of data. It is also free, which is a compelling factor to its growth.
- Because it is open source, new functions and packages are created all the time. So, if you can't find a capability initially, you can search for a package that can do it or even create a package of your own.

Why R & Python?



Following are some of the important statistical analysis use cases which can be implemented using R

- Descriptive Analysis
- Predictive Analytics
- Linear and Nonlinear Modeling
- Statistical tests like P Test, T Test, Z Test and Chi-square Test
- Time-series Analysis
- Classification
- Clustering
- Sentiment Analysis
- Text Mining
- Correlation Analysis

Use Cases



Sentiment Analysis

R has libraries that help users parse unstructured text and determine sentiments based on taxonomy comparison.

Prescriptive Analytics

Prescriptive analytics analyzes for the best possible outcome out of potential decisions.

Prescriptive analytics thrives on Big Data, and it helps you decide what should happen.

Example: How do Airlines determine their ever-changing ticket prices?

Decision Trees

Sometimes data is so complex and it is difficult to find an overall regression model or a single predictive formula that can describe it.

One solution is to break up data into smaller subsets until you can explain each subset with a simple model. R allows you to quickly build decision trees to classify complex data.

R-Tableau Integration

Why R-Tableau Integration:

- Pair the easy-to-use interface of Tableau for exploring and manipulating your data with the power of R's statistical libraries and packages.

Method:

- R functions and models can now be used in Tableau by creating new calculated fields that dynamically invoke the R engine and pass values to R. The results are then returned back to Tableau for use by the Tableau visualization engine.

Prerequisites:

- Users must have proficiency in R and Python languages to write appropriate scripts and functional calls they require.
- Users must have access to an R server to access R functions (Rserve to run R function calls) from Tableau Desktop or Tableau Server.

Functions in Tableau



- There are four new built-in functions that are used to call specific R models and functions.

SCRIPT_REAL

Returns a Real Result from the specified expression

SCRIPT_STR

Returns a String Result from a given R expression

SCRIPT_INT

Returns a Numeric Result from a given R expression

SCRIPT_BOOL

Returns a Boolean Result from a given R expression



What is Python?

- Python is a widely used general-purpose programming language, popular among academia and industry alike.
- It provides a wide variety of statistical and machine learning techniques, and is highly extensible.
- Together, Python and Tableau is the Data Science dream team to cover any organization's data analytics needs.
- In 2013, Tableau introduced the R Integration [since Tableau 8.1] and the ability to call R scripts in calculated fields using R Server [Rserve()].
- With the release of Tableau 10.1, you can use Python scripts as part of your calculated fields in Tableau, just as you've been able to do with R.
- Python Integration happens through the Tableau Python Server [**TabPy**].



Installing Python

- **Installing Tableau Python server: TabPy**

You can download TabPy from the following link:

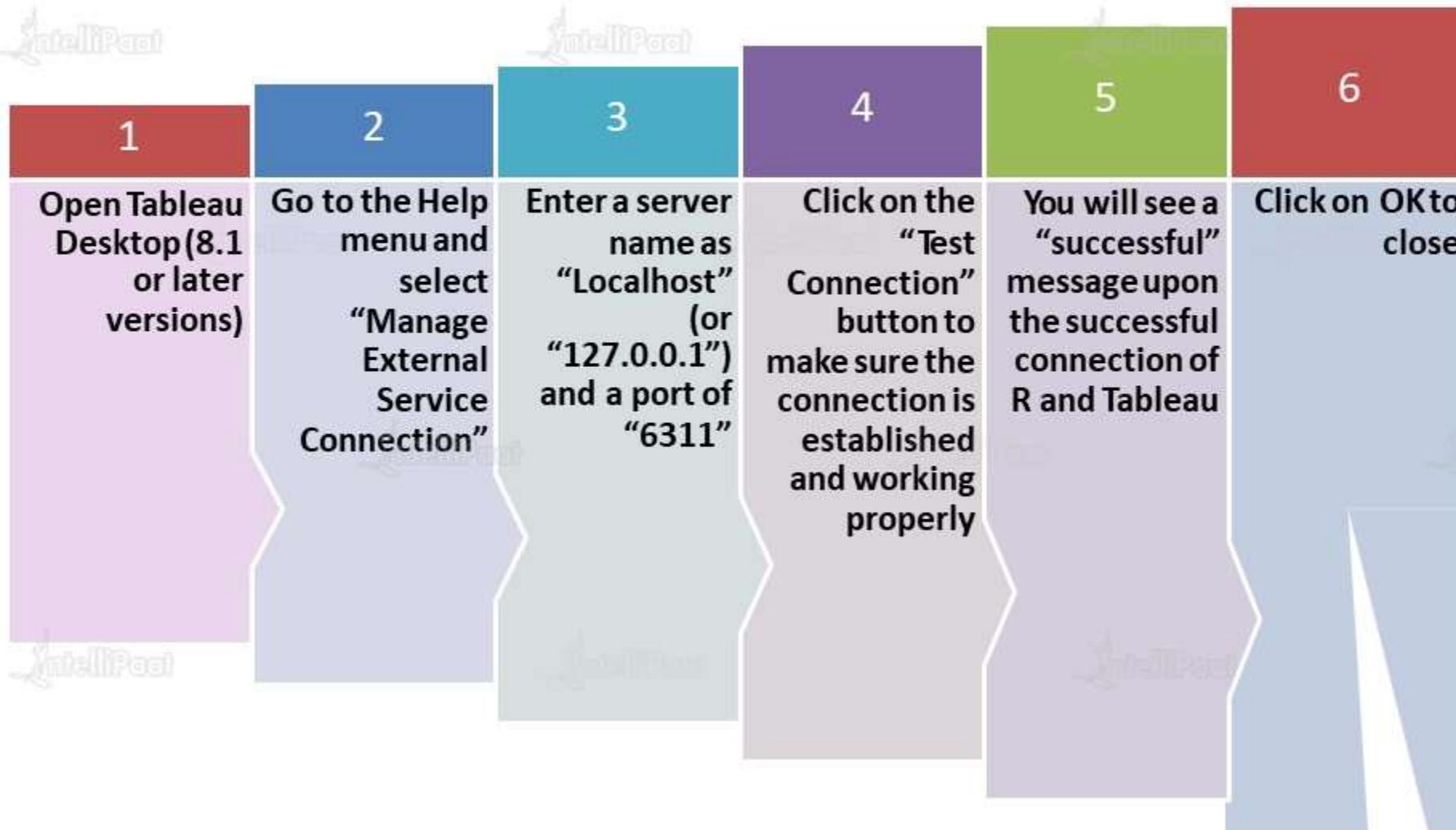
- [GitHub - tableau/TabPy: Execute Python code on the fly and display results in Tableau visualizations](#)

Click on the *Clone or download* button in the upper right corner of the TabPy [repository page](#)

- Download the zip file and extract it.

Connecting to R

To connect Tableau to R, follow the given steps:



Connecting to Python



To Connect Tableau to Python, follow the below steps:

- Open Tableau Desktop (10.2 or later versions)
- Go to the Help menu and select “Manage External Service Connection”
- Enter a server name as “localhost” (or “127.0.0.1”) and a port of “9004”
- Click on the “Test Connection” button to make sure the connection is established and working properly
- You will see a “successful” message upon the successful connection
- Click on OK to close



Hadoop



- Hadoop is an open-source framework that allows to store and process big data in a distributed environment across clusters of computers using simple programming models.
- It is designed to scale up from single servers to thousands of machines, each offering local computation and storage.

Hadoop

- Install Tableau Desktop on the machine
 - To connect MySQL relational database, we need to download its ODBC drivers and install on the machine. ODBC drivers for MySQL can be downloaded from <http://dev.mysql.com/downloads/connector/odbc/>
 - Tableau supports visualization on Hadoop data by connecting to HiveServer or HiveServer2. There are no ODBC drivers available as of now for connecting to Apache Hive directly from Tableau, but Cloudera provides ODBC drivers for Tableau integration with Hadoop via Cloudera HiveServer.
- Cloudera ODBC can be downloaded at

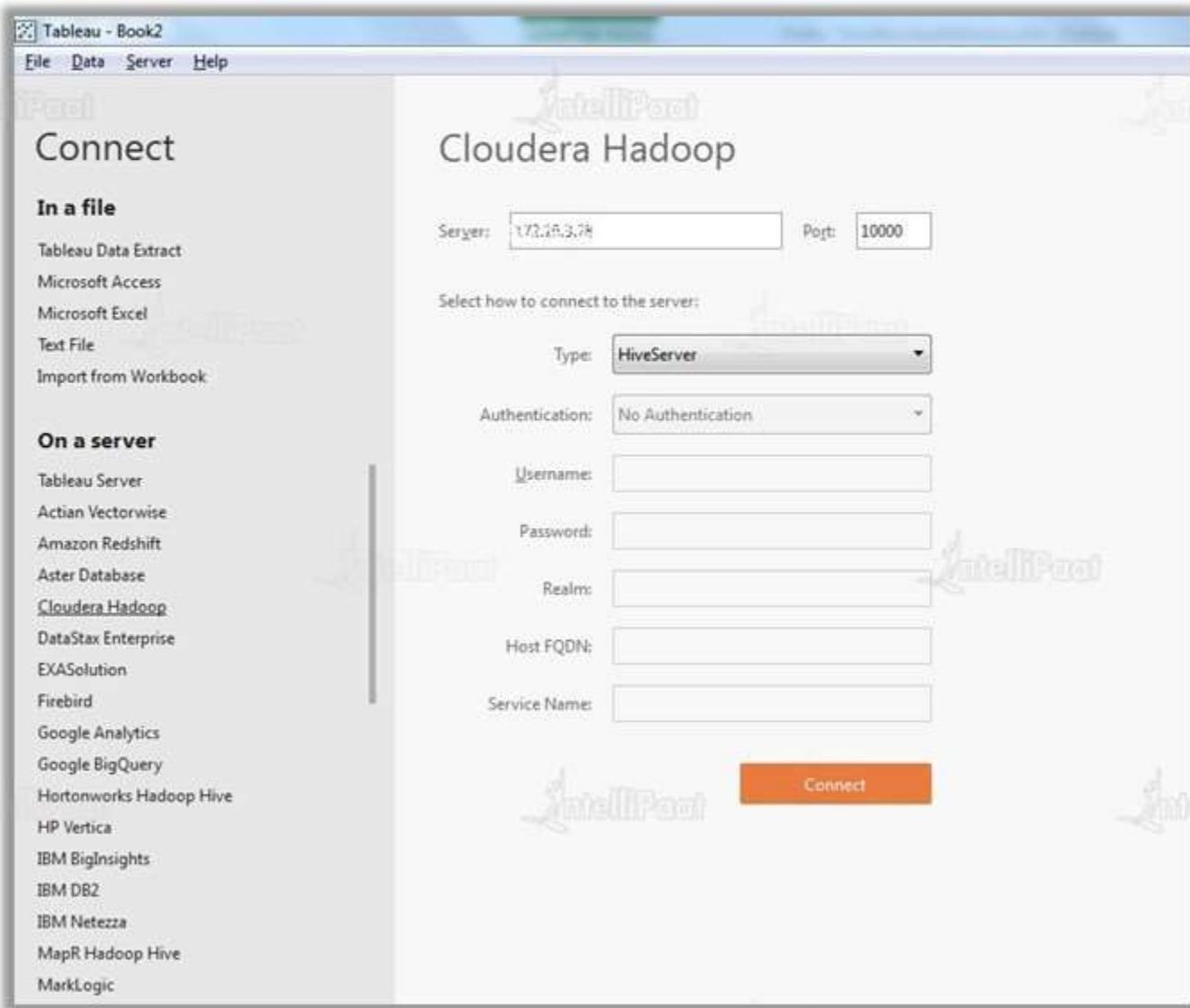
<http://www.cloudera.com/content/cloudera/en/downloads/connectors.html>

- Install the above .msi files in windows machine and restart it to correctly integrate the drivers with each other.

Connecting Tableau with Hadoop



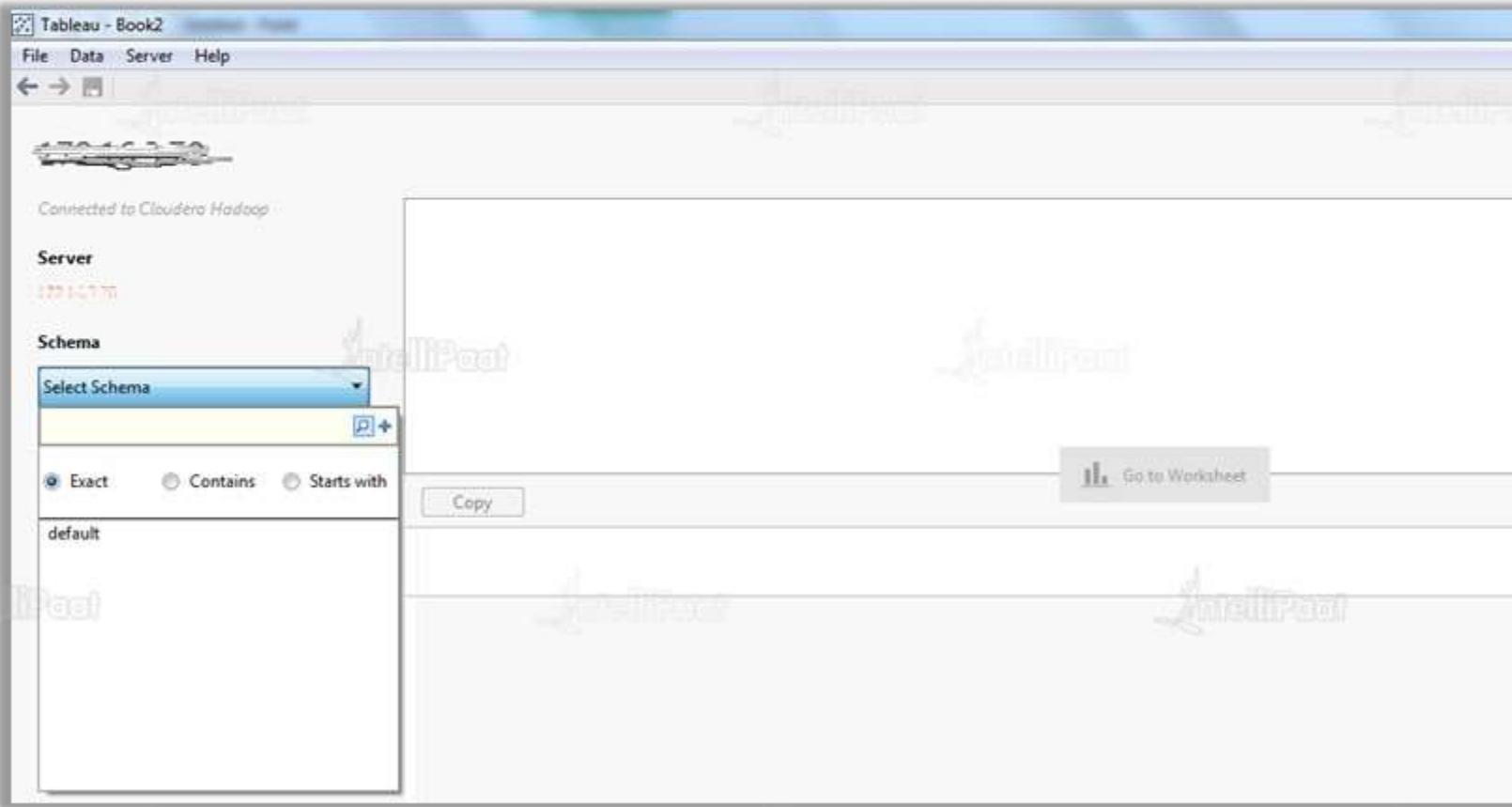
- Start Tableau Desktop software and connect to Cloudera Hive (Data → Connect) and provide the details.
- Connect to HiveServer at port 10000 on the IP address of the machine on which HiveServer is running.



Connecting Tableau with Hadoop



- Once Tableau is connected to HiveServer machine, we need to select the schema (database name) and need to either search tables or select any listed tables.



Connecting Tableau with Hadoop



- After selecting default as schema and clicking on search button against tables, all the tables present in this schema will be listed as shown in the below screenshot.

The screenshot shows the Tableau interface with the following details:

- Title Bar:** Tableau - Book2
- Menu Bar:** File | Data | Server | Help
- Search Bar:** t1
- Connection Status:** Connected to Cloudera Hadoop
- Server:** Cloudera Hadoop
- Schema:** default
- Table:** Enter table name
- Search Criteria:** Exact (radio button selected)
- Results:** A list of tables in the default schema:
 - _asap (default._asap)
 - _asap_sample (..._asap_sample)
 - emp (default.emp)
 - emp2 (default.emp2)
 - hive_sample (...hive_sample)
 - hivetable (default.hivetable)
 - hivetest (default.hivetest)
 - hivetest1 (default.hivetest1)
 - system (default.system)
 - system1 (default.system1)
 - t1 (default.t1)
- Buttons:** Go to Worksheet, Copy, Update Now, Automatically Update



Tableau Prep

Introduction to Tableau Prep



- Tableau Prep is a new Tableau tool that is used to transform, clean and massage the data for quick analysis.
- With simple drag-and-drop features, it simplifies the complicated tasks of joins, unions, pivots, aggregate or create calculate fields for the data quickly.

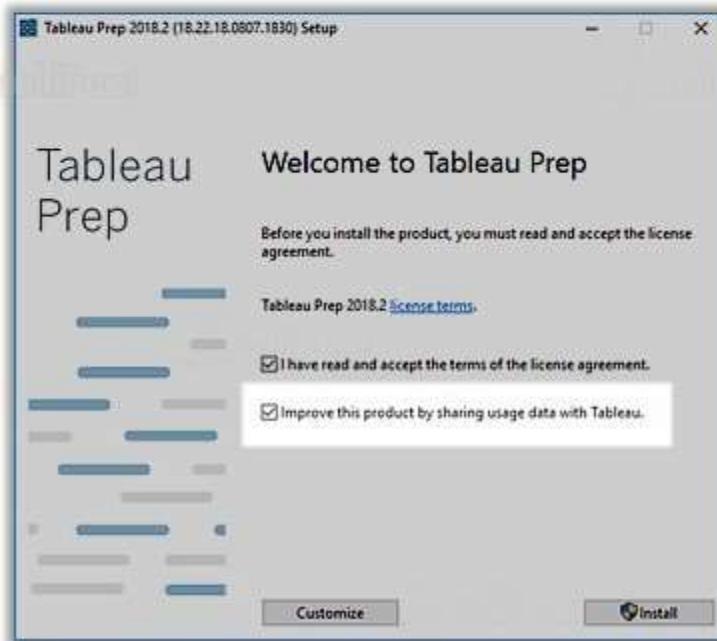
Installing Tableau Prep



Step 1: Download the Installer and install it on your computer by accepting the licensing agreement.

Step 2: To enable or disable usage reporting, complete the following steps. This option allows us to gather usage pattern data to improve the product.

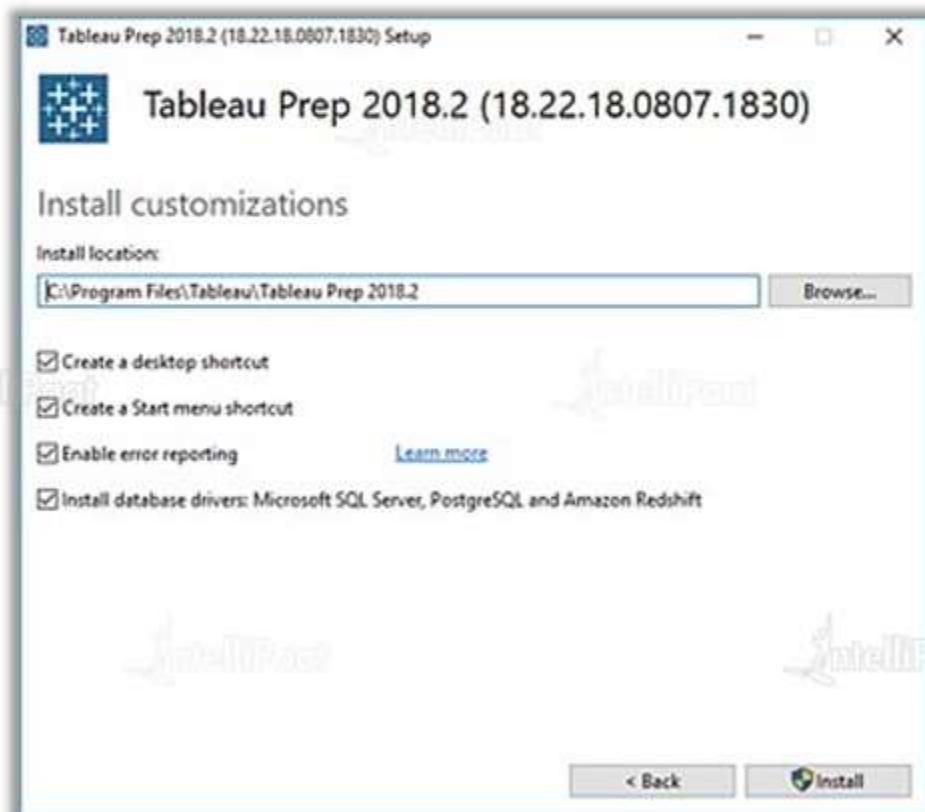
Step 3: For Windows, check the “Improve this product by sharing usage data with Tableau” box.



Installing Tableau Prep



Step 4: To customize the install, on the **Install** welcome screen for Windows or on the **Installation Type** step for the Mac, click on **Customize** and change any of the following options.



Step 5: Click on **Install**.

Cleaning Data in Tableau Prep



- In Tableau Prep, data can be cleaned at any time except the output step.
- This can be done by applying cleaning operations such as filtering, adding, renaming, splitting, grouping or removing fields.

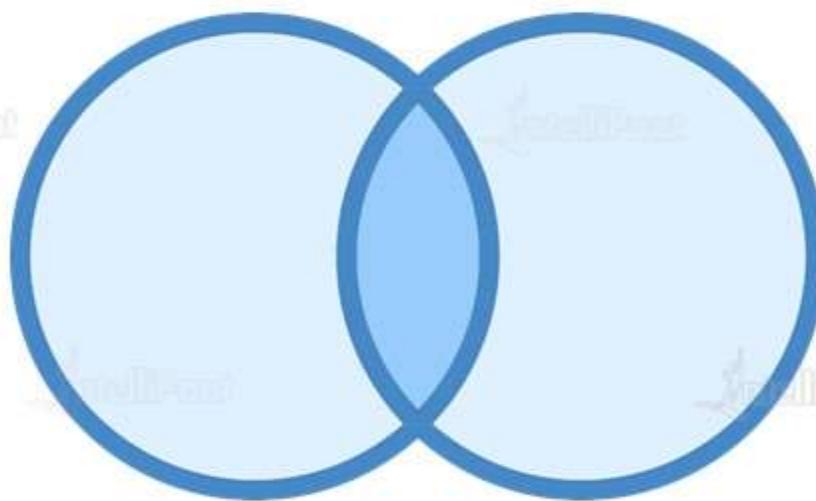
Cleaning Data in Tableau Prep: Input Step



One can make any of the following changes in the Input field list. The changes are tracked in the **Changes** pane and annotations are added to the left of the Input step in the **Flow** pane and in the Input field list.

- **Filter:** Click **Filter Values** in the toolbar and enter your filter criteria in the calculation editor.
- **Rename Field:** In **Field Name**, double-click (Ctrl + click on Mac) on the field name and enter a new field name.
- **Change Data Type:** Click on the data type for the field and select a new data type from the menu.
- **Remove Field:** Uncheck the boxes next to the fields that you don't want to include in your flow.

Joining Data in Tableau Prep



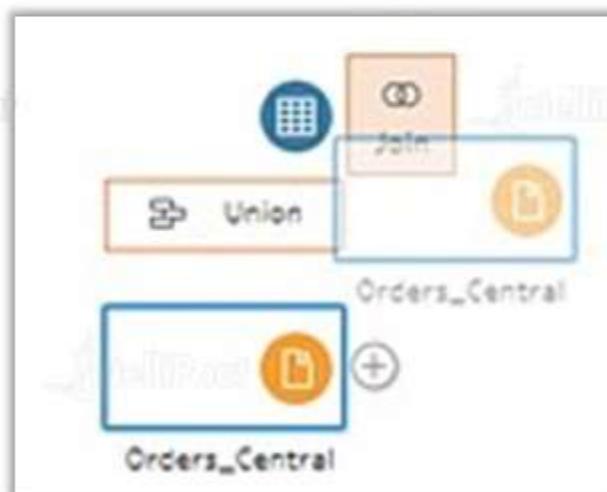
- Join is nothing but combining the related data on common fields.
- In general, the data will be in the form of collection of tables that are related by specific fields.
- Joining is an operation you can do anywhere in the flow. Joining early in a flow can help you understand your datasets and expose areas that need attention right away.

Joining Data in Tableau Prep



To create a join, do the following:

Step 1: Add at least two tables on pane, and select and drag the related table to the other table until the Join option displays. You can also click on the “plus” icon and select Add Join from the menu. A new join step is added to the flow, and the profile pane updates to show the join profile.



Joining Data in Tableau Prep



Step 2: Review the summary of joins and specify the type of Venn diagram.

Under Applied Join Clauses, click on the “plus” icon. Or, on the field chosen for the default join condition specify or edit the join clause. The fields you selected in the join condition are the common fields between the tables in the join.

A screenshot of the Tableau Prep interface showing the 'Applied Join Clauses' section. It displays two tables: 'sets' and 'lego_sales'. A join clause is defined between 'sets.year' and 'lego_sales.year' using an equals sign (=). Other columns listed include 'descr', 'pieces', 'set_id', 't1', 't2', 't3', 'qty_sold', 'set_family', 'set_family_member', and 'year'. The 'year' column from the 'lego_sales' table is highlighted with a gray background.

Joining Data in Tableau Prep



Alternatively, you can click on the recommended join clauses shown under Join Clause Recommendations to add the clause to the list of applied join clauses.

The screenshot shows the Tableau Prep interface with the following sections:

- Applied Join Clauses:** Shows joins between "Clear_Notes_Approver" and "All_Orders".
 - Product ID: Product ID
 - Order ID: Order ID
- Join Type:** Right join (selected). A diagram shows the join type as a right outer join.
- Summary of Join Results:** Includes a table showing the count of included and excluded values.

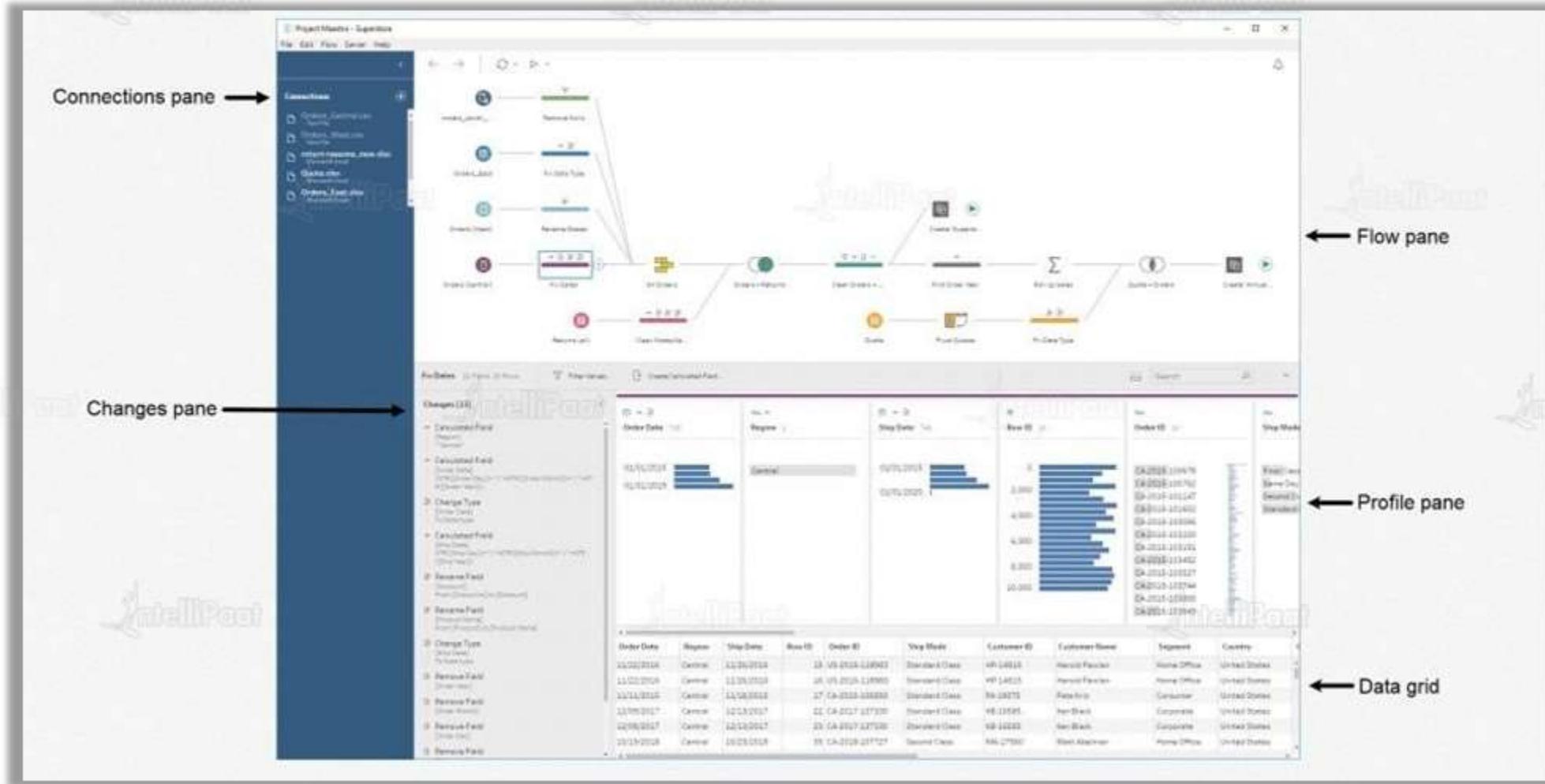
Included	Excluded
272	1
All_Orders	16,302
Join Result:	16,302
- Join Clause Recommendations:** A table listing recommended join clauses between "Product ID" and "Product Name".
- Join Clauses:** A table showing join clauses between "Clear_Notes_Approver" and "All_Orders".

Join Clauses	Show only mismatched values
Product ID	Product ID
Order ID	Order ID
- Join Results:** A table showing the results of the joined data, including columns for Table Names, Approver, Return Notes, Order ID, Product ID, and Returns Reason.

Creating Smart Experience



Tableau has various functions that can make smart experience and visualizations.



Creating Smart Experience

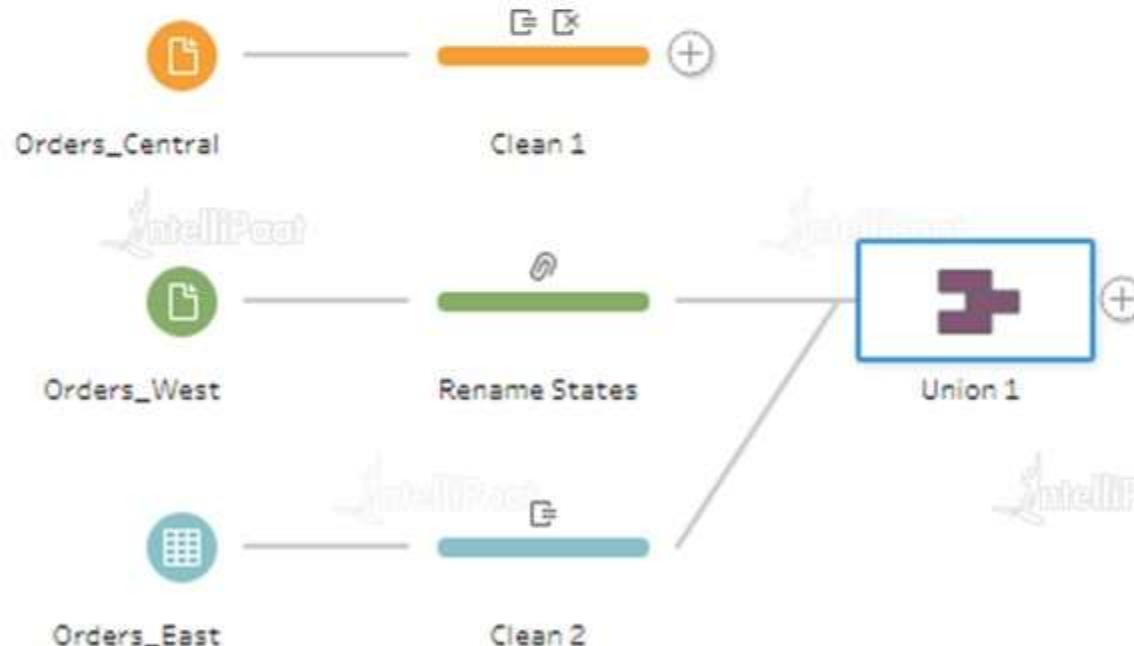
- **Connections pane:** Shows the databases and files you are connected to. Add connections to one or more databases, and then drag the tables you want to work with into the flow pane.
- **Flow pane:** As you clean, shape and combine your data, steps will appear in the flow. This visual indication will allow you to see an overview of your changes. You can accomplish a variety of data cleaning tasks in moments, such as fuzzy clustering and other smart features.
- **Profile pane:** Displays a summary of each field in your data and allows you to see the shape of your data and begin to identify any issues with your data.
- **Changes pane:** Tableau Prep keeps track of the changes you make, in the order you to make them, so that you can always go back and review or edit those changes.
- **Data grid:** Lets you see row-level details and verify individual records.

Combining Data



If you want to combine the files together to add the rows from each file into a single table, follow the below steps:

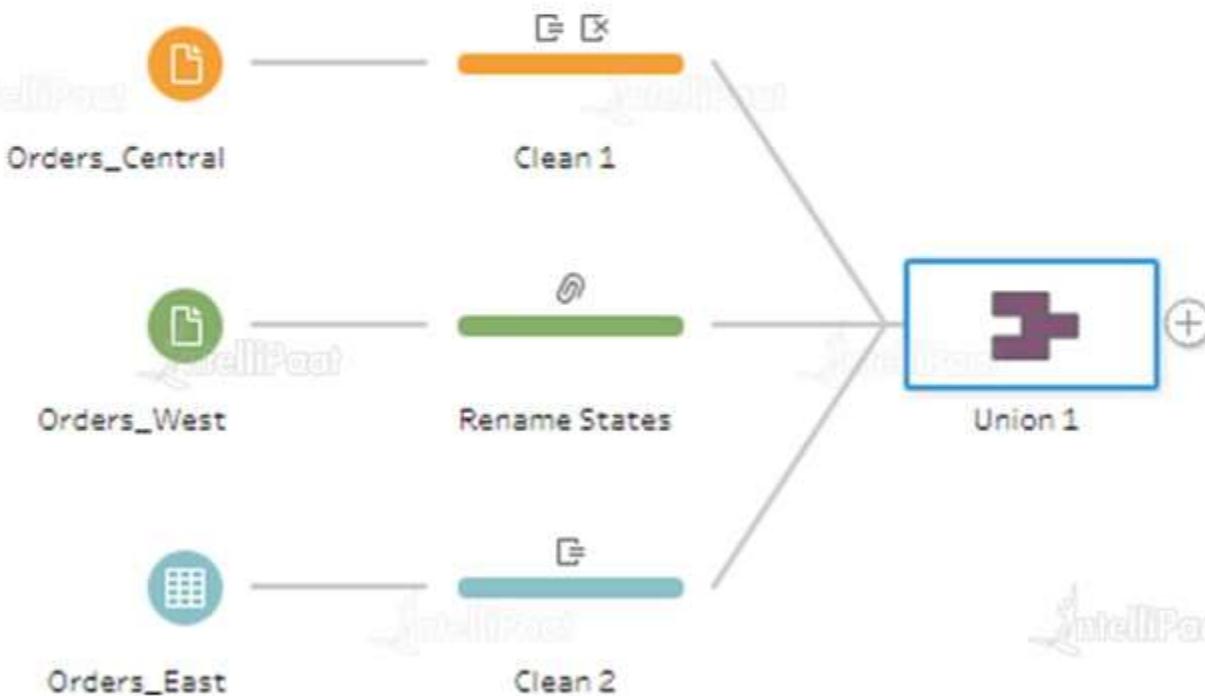
Step 1: Drag Orders_West file to Orders_East file and use Union step.



Combining Data



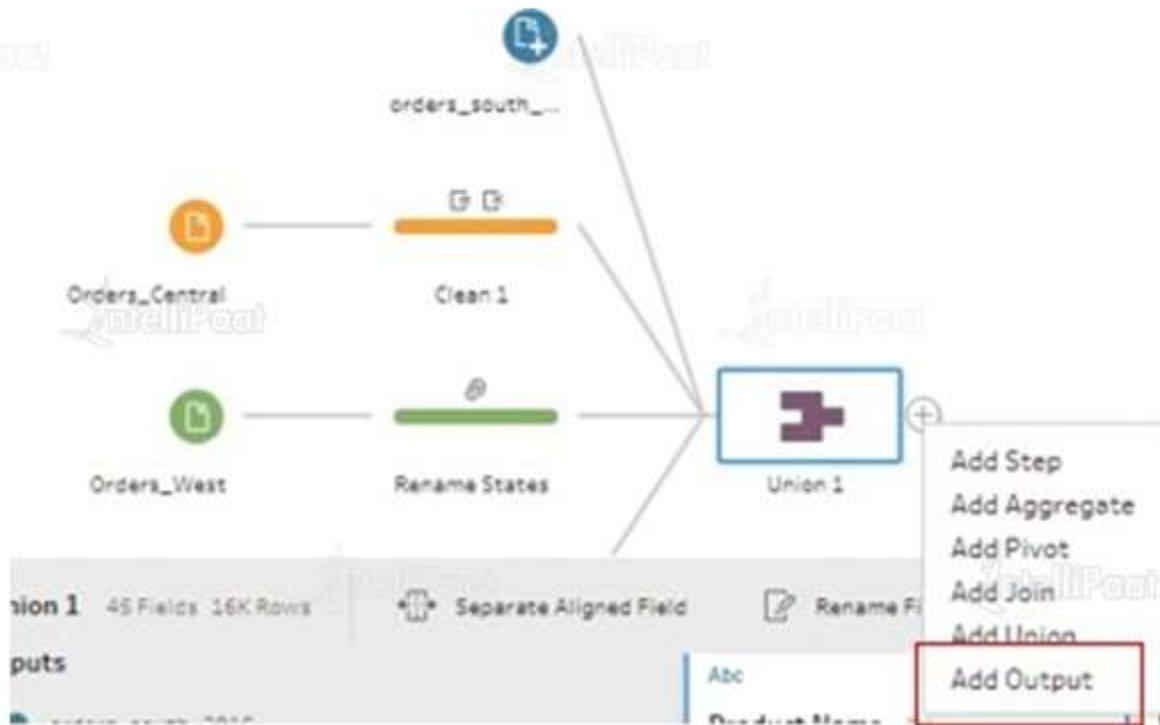
Step 2: Add Orders_Central's cleaned extract to the Union 1.



Combining Data



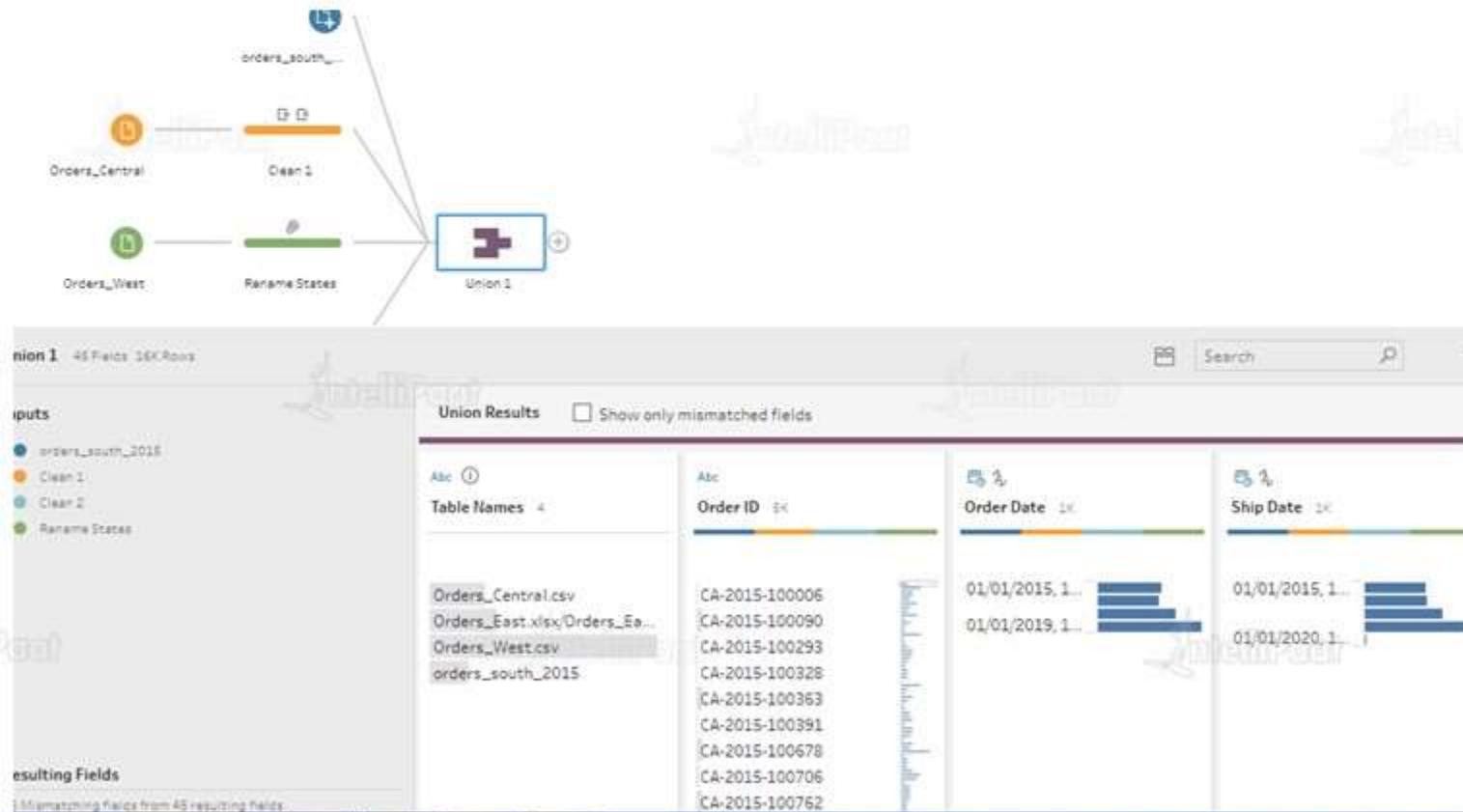
Step 3: Drag the orders_south_2015 step to the new Union step. Drop it on Add to add it to the existing union.



Combining Data



Step 4: Now, all our files are combined into a single table. In the Flow pane, select the Union step to see results.



Integrating Tableau Prep with Tableau



- Tableau Prep is integrated into the Tableau analytical workflow to help the users to move data quickly from data prep to analysis.
- This can be done by just opening your Tableau Prep flow in Tableau Desktop at any time or sharing it with your colleagues over Tableau Server or Tableau Online.
- Tableau Prep uses the Tableau data connectors, calculation language, permissions and data source certifications.

Integrating Tableau Prep with Tableau

