Hands on Tableau Training for Data Science:

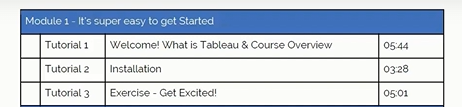
# Section -1- It’s super easy to get started.

## What is Tableau and Course Overview

Tableau:

Tableau is a very simple, yet powerful tool for everything to do with data. The company's mission is to help people see and understand data. Tableau is a completely drag and drop software and using Tableau, it is possible to create visuals sometimes ten times faster than what could be created in other programs.

And overall, Tableau is a new and innovative approach to business intelligence.



Table

Description automatically generated

Graphical user interface, application

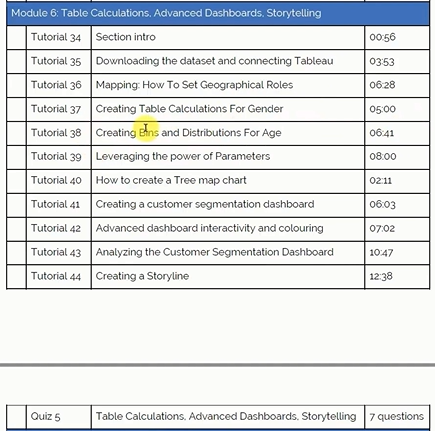
Description automatically generated with medium confidence

Table

Description automatically generated

Table

Description automatically generated

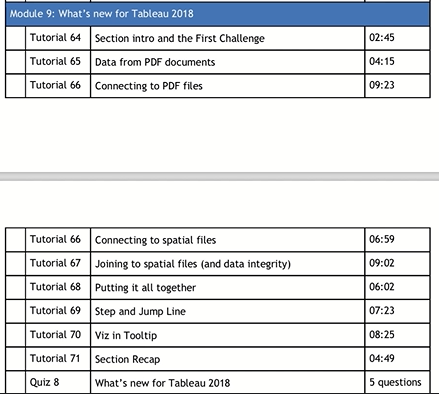


Table

Description automatically generated

Table

Description automatically generated



## Installation

Direct download of Tableau which is a 14-day trial

<https://www.tableau.com/>

Tableau Public – FREE

<https://public.tableau.com/app/discover>

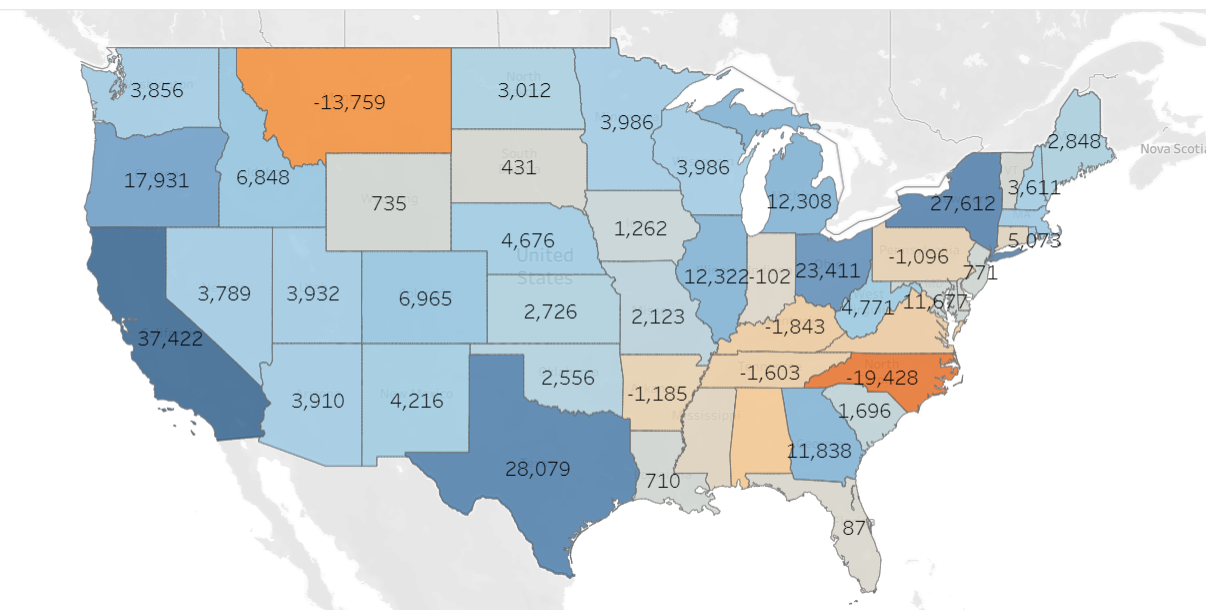
## Exercise

To get the Datasets for this course, click the below link

<https://www.superdatascience.com/pages/tableau>

* A Map to show which state has the greater and lesser profits

<https://public.tableau.com/app/profile/sayeesudha.senthil.velan.sayee./viz/SalesprofitinUS/Stateprofits>



INSIGHTS:

* It is clear that the least profitable state with most losses were incurred in North Carolina
* The most profitable state is California.
* According to the map, these southern eastern states are not doing well except for Georgia and the northern states.
* The western states are doing better when compared to Eastern State except for Montana.

## Get the Datasets here:

Please download the materials required for each section through the link below:

<https://www.artofvisualization.com/pages/tableau>

## Extra Resources:

<https://sdsclub.com/wp-content/uploads/2022/10/WHAT-CHART-TYPE-TO-CHOOSE-FOR-WHAT-DATA.pdf>

<https://sdsclub.com/wp-content/uploads/2022/10/7-Reasons-why-Tableau-Top-BI-tool.pdf>

# Section-2- Tableau Basics: First Bar Chart.

## The Business Challenge – Who gets the annual bonus?

BUSINESS PROBLEM:

It's end of financial year, and that means. It means it's time for annual bonuses.

The store operates in three regions, and only the top performing employee in each region qualifies for a bonus.

Find out which three employees are eligible to get a bonus to get bonuses for this year.

**Employees are measured on the total number or total value of sales.**

|  |
| --- |
| **CSV – Comma Separated Values** |

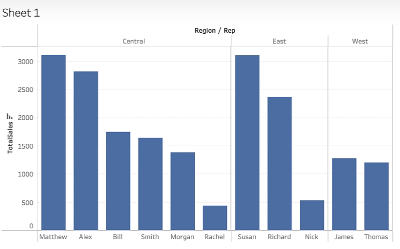
## Connecting Tableau to a Data File – CSV File

* CSV is opened as text file
* A single Tableau can many datasets and we form relationship between them in the connection manager screen
* Data -> New Data Source -> csv file

## Navigating Tableau

* Data table – it’s on the left and it has two divisions and they are
* Dimensions – quantitative data, independent variables
* Measures- qualitative data, dependent variable
* Worksheet – A single sheet where data analysis is performed
* Dashboards – A combination of worksheets
* Story – A combination of worksheets and dashboards. This is mostly the latest update of Tableau
* Analysis – How you want to perform analysis on the current sheet. Ex- Map
* Format – deals with formatting
* Worksheet
* Column
* Rows
* Show me

## Creating Calculated Fields

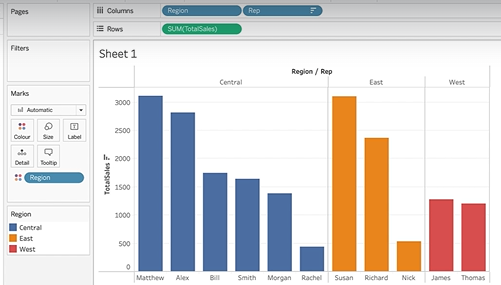
* 
* First the chart is broken as regions and following it the Rep data is broken
* Calculated fields – Right click on the unit section 🡪 Create Calculated fields
* By total sales
* 
* Bonus – Mathew, Susan and James
* By Unit Sales Result

Chart, bar chart

Description automatically generated

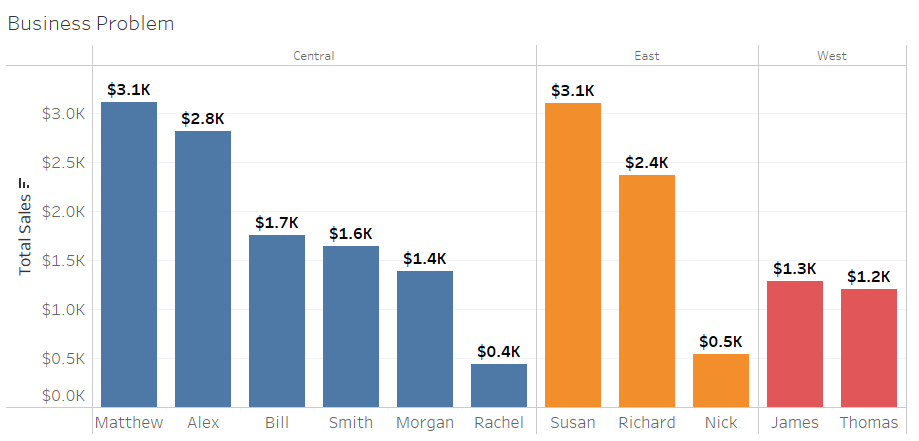
Bonus - Alex, Richard, James

## Adding Colours

* The Color button in Mark Section is where we could use to change color
* More color patterns can be provided by selecting: Colors 🡪 Edit Colors 🡪 Select Color Palette 🡪 Select anything 🡪 Assign Pallete 🡪 Apply 🡪 Ok
* TIPS – use Ctrl + select the field data in the Columns and Rows for which the colors are needed to be visualized. It is better than pulling the data from the Data sections of the Worksheet area
* 

## Adding Labels and Formatting

* How much Bill made?
* Here we can use the Label in Mark section
* Formatting – can be done using the format tab in the Menu bar and this format will be overridden by the Label format in Mark section



## Exporting Your Worksheet

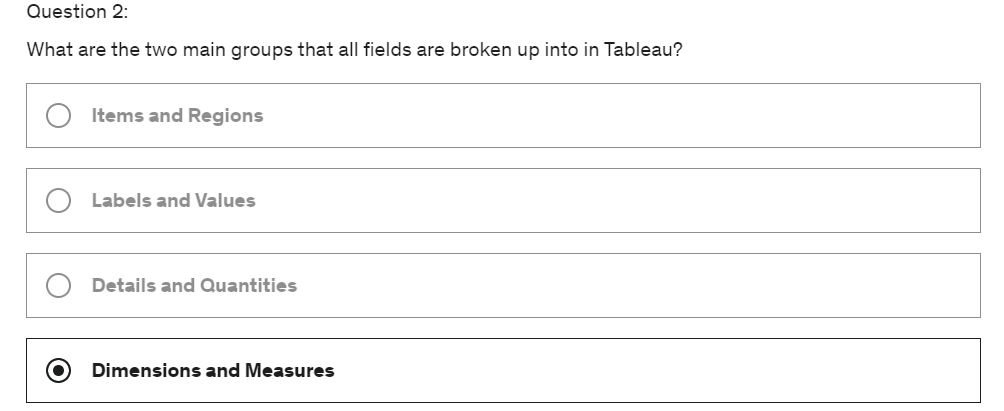
This cannot be done public version, it can be done in full Tableau Desktop version

* Worksheet 🡪 Copy 🡪 image
* Right click 🡪 copy 🡪 image

## Get the Viz

* Once the Visualized data is saved in public, then immediately your respective Viz would be created
* File 🡪 Save to Tableau Public

## Quiz 1:Tableau Basics



Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Text, application

Description automatically generated

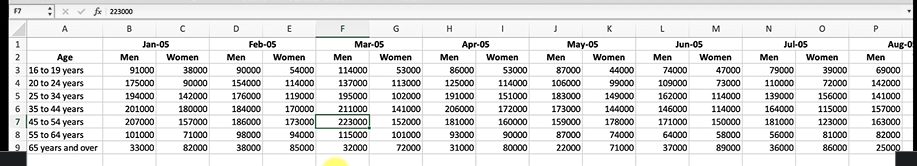
# Section-3- Time Series, Aggregation and Filters

## Section Intro

* Data Set – It is about long-time unemployment rate in the US for the past couple of years
* Time Series Data visualization
* Focus on Granularity and aggregation – concepts that underpins whole of Tableaus works
* How Granularity works, how level of details works in tableau in-order to do good visualization
* Filters and Quick Filters
* Area Chart and format it

## Working with Data Extracts in Tableau

* The Dataset has many repeated values or duplicates
* This more natural for an individual to understand the dataset



* The unstructured data is what Tableau or any BI tools preferred to modifying for the dataset to structured
* Extract Data (Paid version)– right click the data set in Data section 🡪 Click Extract data

This creates an extract for the tableau to work from

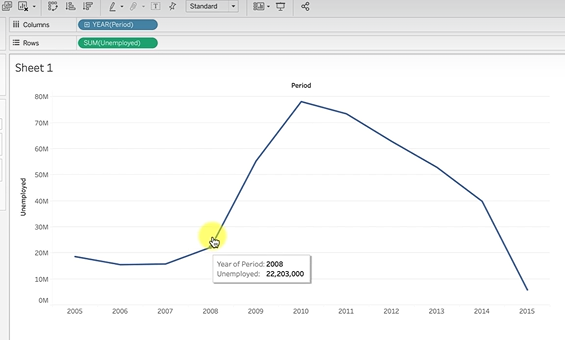
* Why do you use extract?

Using the live database will not be reliable as the data might be updated in the database and so extract the existing dataset and then evaluate the values and so if so any changes done in the extracted dataset. It will not affect the actual live dataset.

* To refresh the extract dataset: right click the data 🡪 click Extract 🡪 Refresh
* To return to the live connection: right click 🡪 uncheck Extract

## Working with Time Series

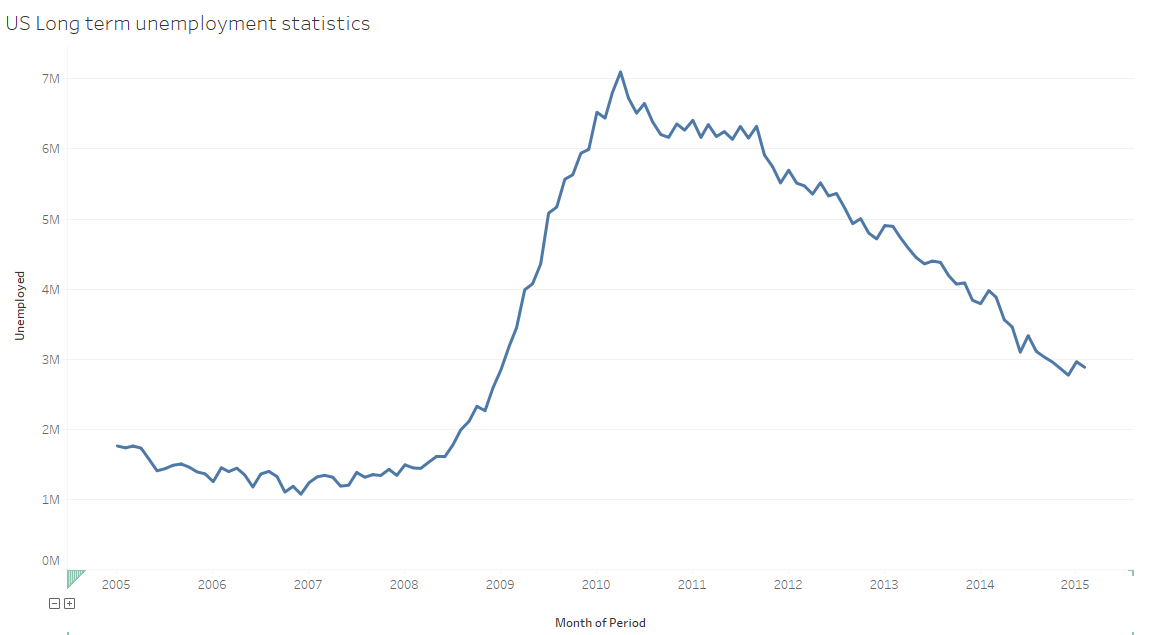
* Tableau automatically sums the number unemployment for each year (2007 – 2015)



* Now how to get the data of months (Granularity concept is utilized)
* TIPS: Dimensions are – ‘BLUE’ and measures – ‘GREEN’
* TIPS: double click the column or row data to see what calculation is performed by tableau



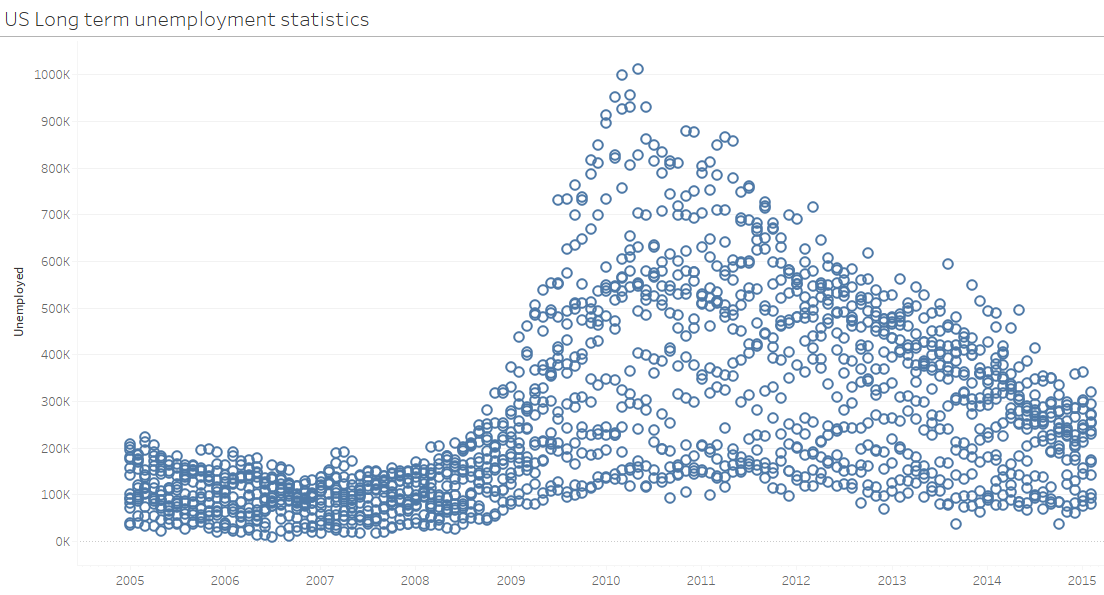
* In this case, the timeline must be evaluated as measures and not dimensions

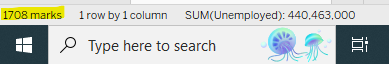


## Understanding Aggregation, Granularity, and Level of Details

* Month of period is the variable that governs granularity
* Measures get aggregated and dimensions mentions the level of granularity
* If you switch of the aggregation (Analysis 🡪 Untick the Aggregate measures)

Every single row of the data set is plotted in this graph. So it I possible to have lots of value on the same month and year



* In the bottom - 

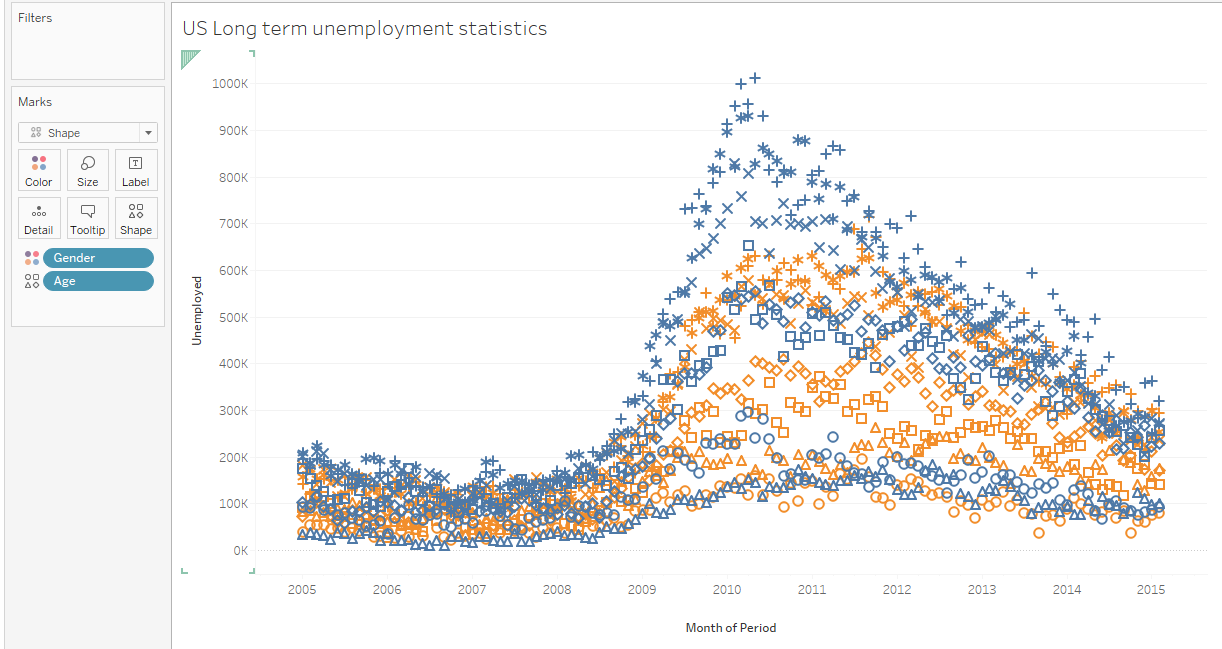
This is total number of rows that the dataset has, as we know our data set in monthly levels

* Gender – M and F who were unemployed in the specific months

Graphical user interface, chart, scatter chart

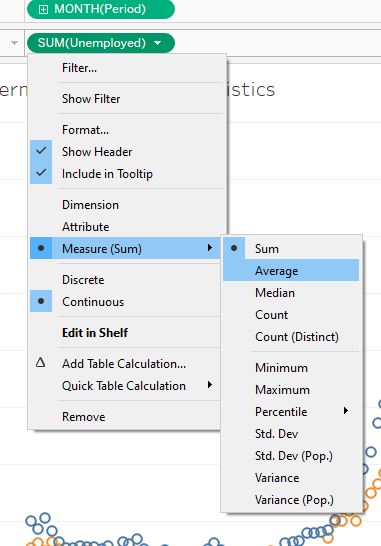
Description automatically generated

* Age is also granularized by using shape inf mark section



* Graphical user interface, application

  Description automatically generated
* Different types of aggregations:
* Avg



* Median

Chart, scatter chart

Description automatically generated

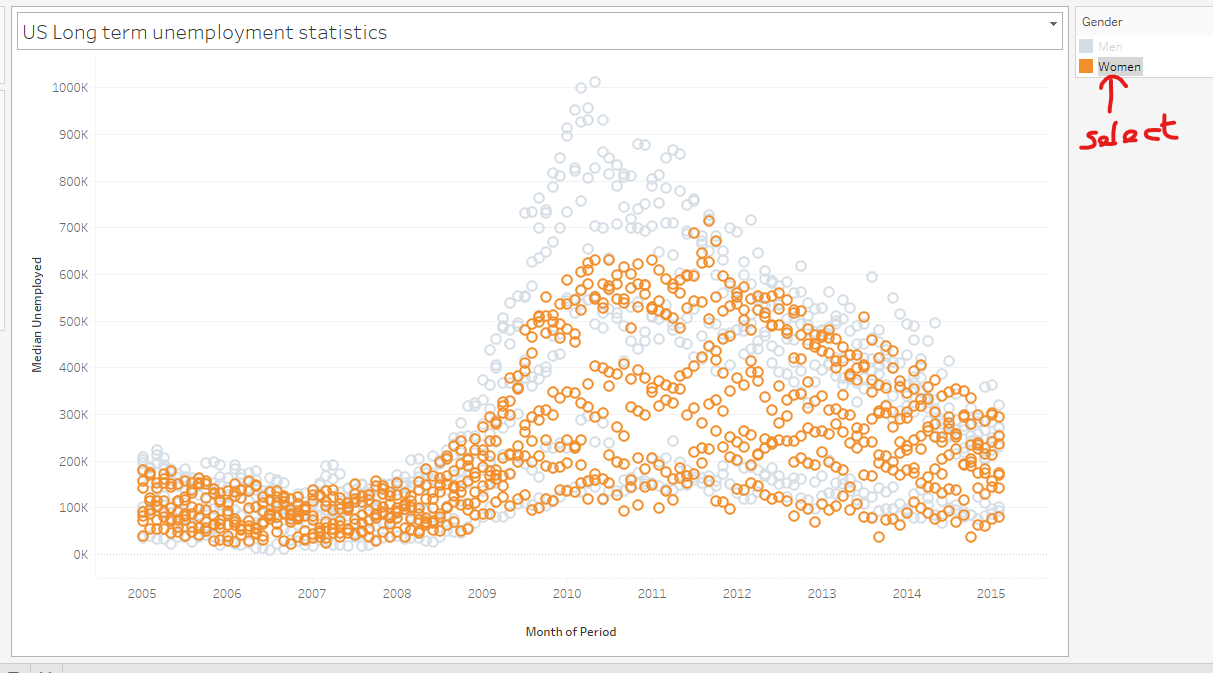
* More Details to be provided for granularity.
* Select option ‘Detail’ in mark section

Chart, scatter chart

Description automatically generated

## Creating an Area Chart & Learning about highlighting

* Need to know more about a particular Gender granularity.



* You can view the visualization of data of a particular group. For this Highlight select item option can be used

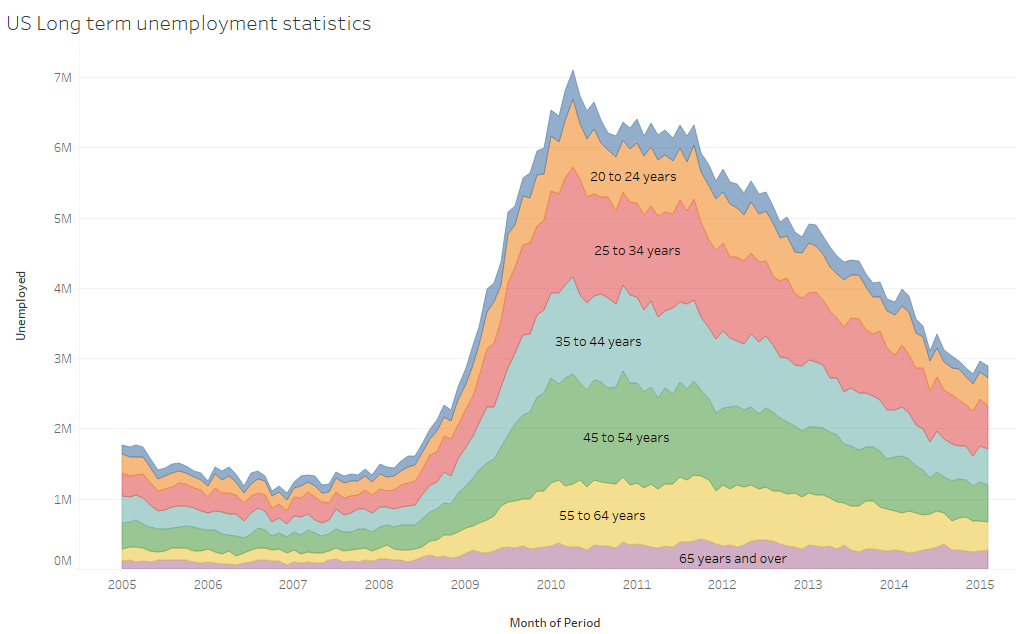
Graphical user interface, text, application

Description automatically generatedif we click on any of the data only that can be highlighted and visualized

Graphical user interface, chart, scatter chart

Description automatically generated

* Area Chart



## Adding a Filter and Quick Filter

* To do filter add the data field value to the filter section of the workseet

Chart, histogram

Description automatically generated

* Quick Filter – right click of the field value 🡪 show filter

Chart, histogram

Description automatically generated

* Multiple filters and quick filters

Chart, histogram

Description automatically generated

* Types of filters

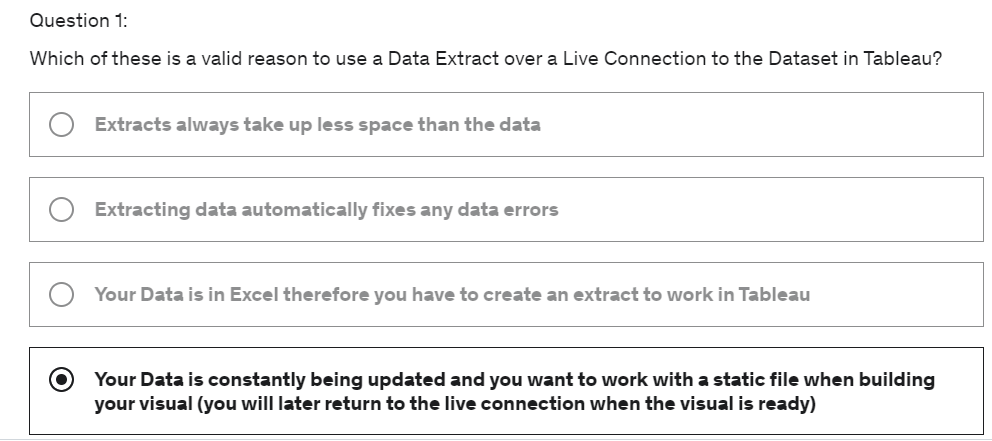
Graphical user interface, application

Description automatically generated

Chart

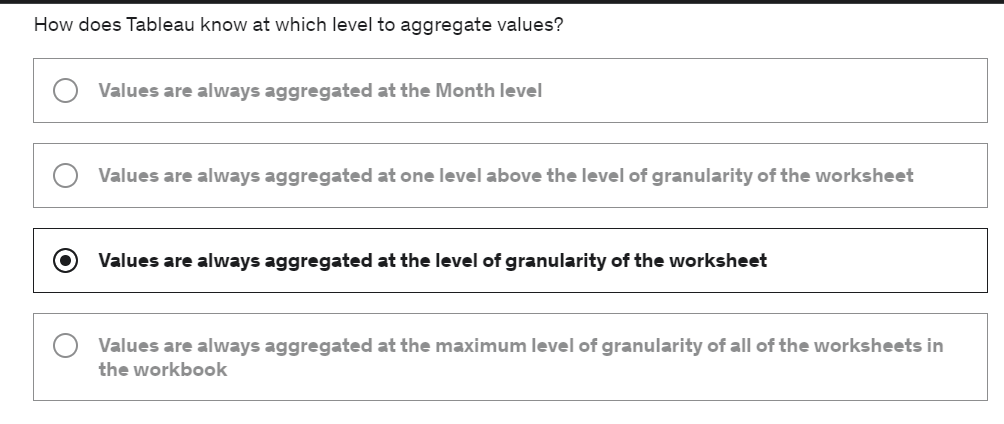
Description automatically generated

## QUIZ 2: Timeseries, Aggregation and Filters



Graphical user interface, text, application

Description automatically generated



Text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

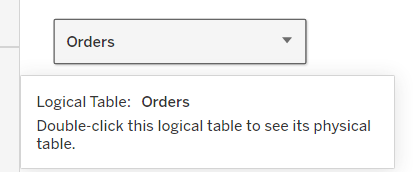
# Section-4- Maps, Scatterplots and your first Dashboard

## Section Intro

* We will doing Maps and Scatter plots
* Datasets – Sales data of a store in Europe and we will create a map to determine how it is performing across different regions and scatter plots for customers
* Create our first Dashboard which is interactive

## Relationships vs Joins

* V.2020.2
* Data Model - Where one dataset is connected with another dataset
* Relationships – in Logical Layer Tables are connected for relationships
* Joins – It is created by double clicking the table data set in the Logical layer, so it creates the physical layer



Physical layer

Graphical user interface, application

Description automatically generated

Drag the other data set in this layer in-order to create a join

Diagram

Description automatically generated

Once you close the physical layer, the logical layer will have the joined data set

Table

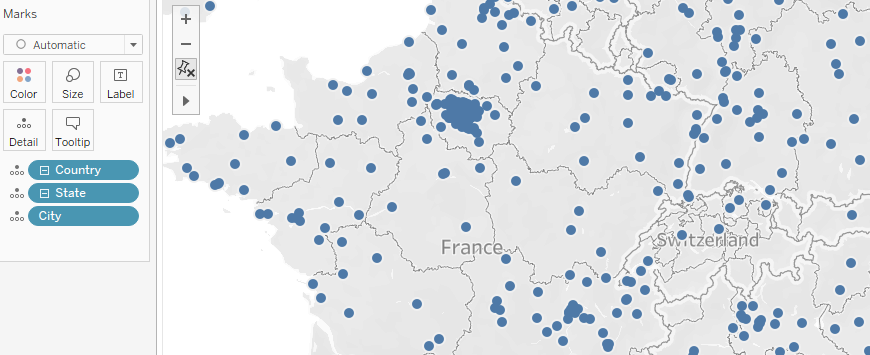
Description automatically generated with low confidence

## Joining Data in Tableau

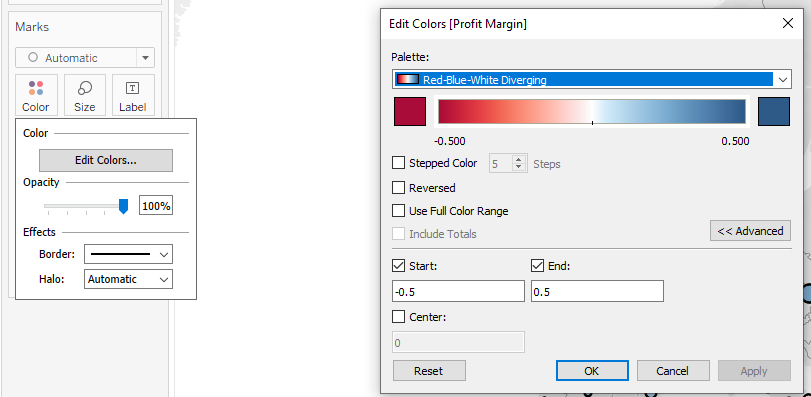
* We need to join ‘List of Orders’ and ‘Order Breakdown’ tables

## Creating a Map, Working with Hierarchies

* Challenge – Look at the profit margin at every single state
* Geographical data of the dataset are – Country, State and City
* We can create the natural hierarchy by dragging city from the dataset fields and place them next to country. Tableau will prompt us saying a hierarchy to be created 🡪 then create the hierarchy
* When we visualize. Country, State and City. The level of granularity visualization is clearly shown in the analysis

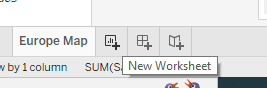


* Profit Margin – Sum of profit / Sum of Sales
* Color Advance section helps us to set the type of color to be highlighted for this range

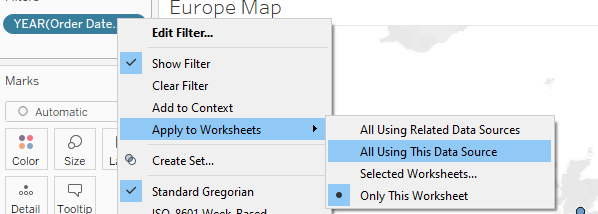


## Creating Scatter plot, Applying filters to Multiple Worksheets

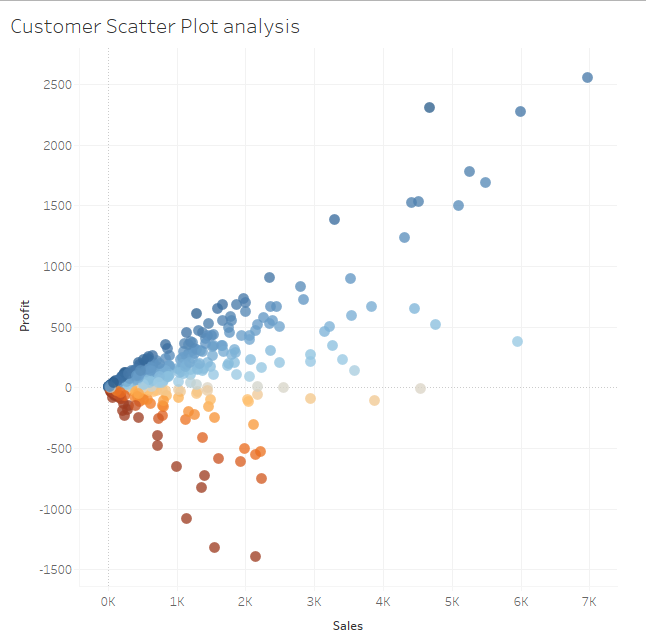
* To create new worksheet



* Apply a filter across many different worksheet



* Customer Scatter plot analysis

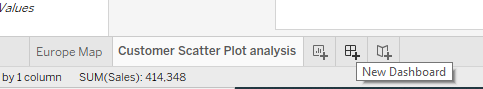


INSIGHTS:

* Least profitable customers are red – do surveys, get reviews and analyse
* Most profitable customers are blue – Give them some offers, reward points

## Lets create our first Dashboard

* Two worksheets which we created can be made as a dashboard, so that this allows a user to look into all the worksheets as the same project
* To create a new dashboard. Select the following option

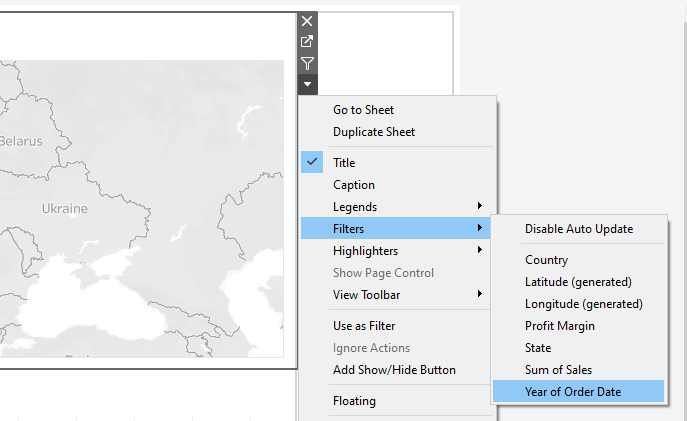


* All the sheets are available here

Graphical user interface, application

Description automatically generated

* How to add the filter in the vertical container



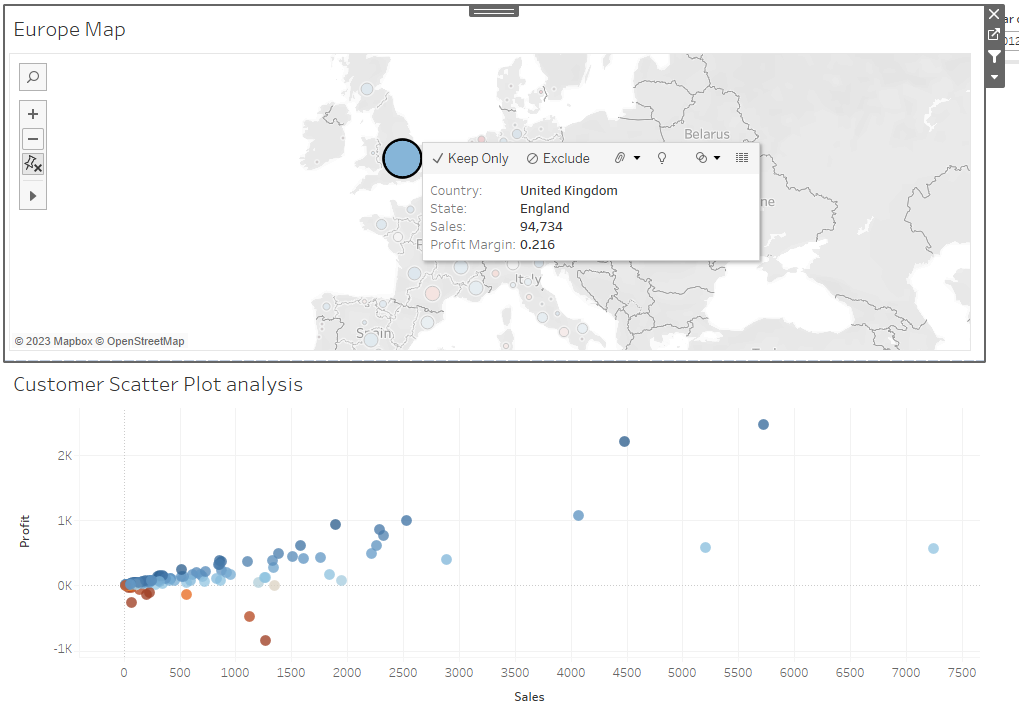
## Adding an Interactive Action – Filter

* To add an interactive action – Dashboard 🡪 Actions
* To do automatic action

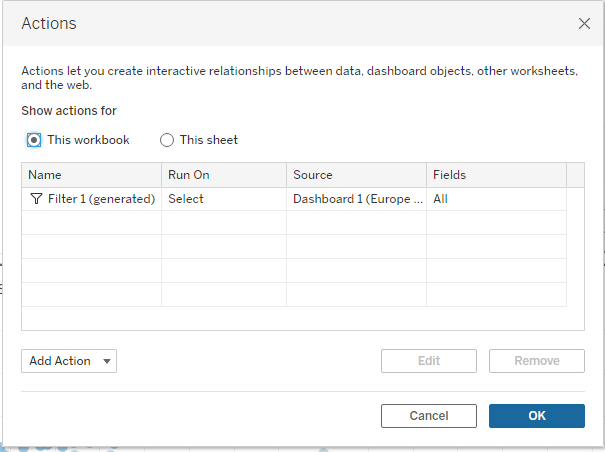
Graphical user interface, application, Word

Description automatically generated

* This helps the scatter plot to be adjusted for the relative country which was selected in Europe map



* The action is updated in the Action window

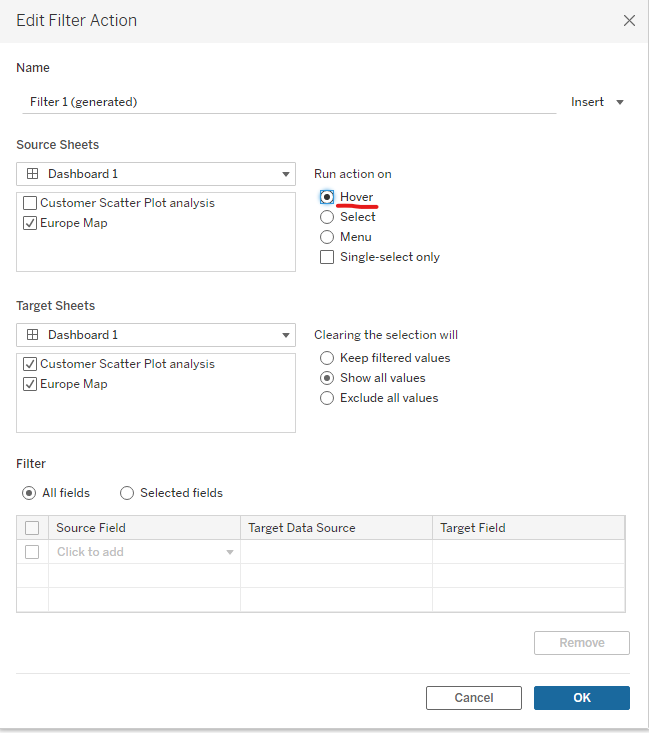


* Now we can edit it

Table

Description automatically generated

* Changed the Action



* How to add this filter on your own –

Graphical user interface, text, application

Description automatically generated- Add the filter 🡪 click ok🡪ok

- To do multiple selections – Select ctrl 🡪 Select the states or use

Graphical user interface

Description automatically generated with medium confidence🡪 Map

Description automatically generated

## Adding an interactive Action – Highlighting

* To add highlights 🡪 Dashboards 🡪 Actions 🡪 Highlights 🡪 Make the changes 🡪 ok 🡪ok
* The highlights will not update the scatter plot once we select a state in Europe. Instead it highlights the plots of the customer who has purchased in that state
* What if a Customer has purchased in different states of Europe. So Now update the state in the scatter plot must be updated. This is because, now the highlight feature will not be effectively utilized
* Now the highlights will be done effectively



## Quiz 3: Maps, Scatterplots and Your First Dashboard

Text

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

A picture containing table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

# Section-5- Joining, Blending and Relationships; PLUS: Dual Axis Charts

## Section Intro

* When multiple sources of datasets are provided, then we must use these concepts. In tableau three ways for handling these datasets – Joining, blending and relationships
* We will create dual axis chart
* Calculated field to be used across blend

## Understanding how LEFT, RIGHT, INNER and OUTER Joins work

* **INNER JOIN** – matching rows of both the tables will be mapped
* **LEFT OUTER or LEFT** – Primary table is left, rows which don’t match the content of left table will be removed from right table
* RIGHT OUTER or RIGHT – Primary table is right, rows which don’t match the content of right table will be removed from left table
* FULL OUTER JOIN – all the rows of both the table will be kept after the join

## Joins with duplicate values

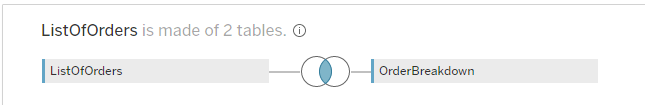
* What happens when cols have duplicate values
* After doing the join - The rows will be duplicated based on the value of duplication from another table

## Joining on multiple fields

* When it is necessary to join on multiple fields and show what happens when you accidently forget to join on multiple fields. How to rectify this
* Operators are used for providing multiple conditions for joins – AND, OR, NOT is utilized

## The Showdown: Joining Data vs Blending Data in Tableau

* We need to compare whether the departments are meeting or not meeting their Targets
* Created the inner join



* Later blend this data source to new data Source (Data 🡪 New data source).
* Create this data source for Sales target.
* Graphical user interface, text, application

  Description automatically generated
* This concept of creating two data source creation is called data blending and this is done when we find JIONING data sets are not going to give effective results
* Other reason for Blending; is when the data sources are different. Example - Excel and csv file.

## Data Blending in Tableau

* This is a feature which is not often much used at the same time it is very controversial where it must be used often
* While working with the datasets, the data of two different sources can be interlinked and blended. These blended data values in Tableau can be represented as follows:

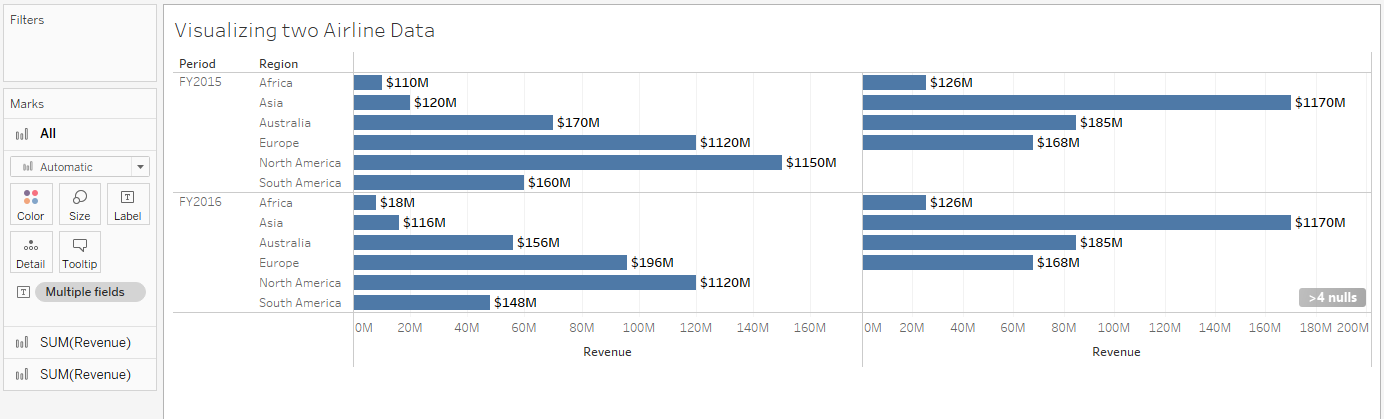


* Each blended data of revenue can be modified based on our needs

Graphical user interface, application

Description automatically generated

* ALL will allow us to make the changes to be common for both the Revenue charts of Airline 1 and Airline 2



* To blend on 2 fields in tableau .i.e. to join on two different data sources field

Data 🡪Edit Blend Relationships 🡪 Custom 🡪 Add🡪 Select the fields which needs to be blended

Graphical user interface

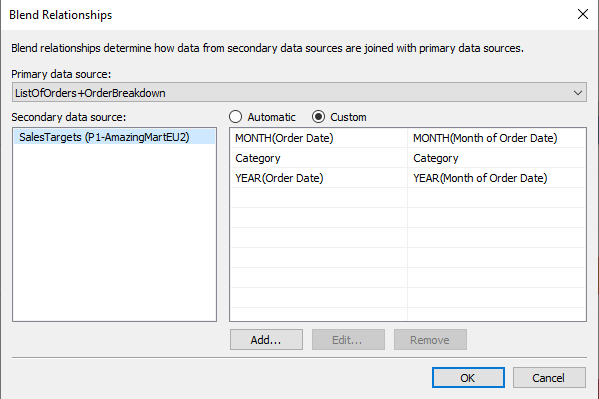
Description automatically generated

A picture containing graphical user interface

Description automatically generated Blending is LEFT JOIN

* Other way we can join multiple fields is by re-naming the field name in one of the dataset which matches the same as the field name of the other dataset
* Blue data base is primary and orange is secondary table
* Blending is done based on per worksheet basis. So when we create a new worksheet, the previous blended worksheets will not be there. So here we can do new blending of datasets

## Dual Axis Chart

* Advance Blending Concepts of visualization and do dual axis chart
* 

A screenshot of a graph

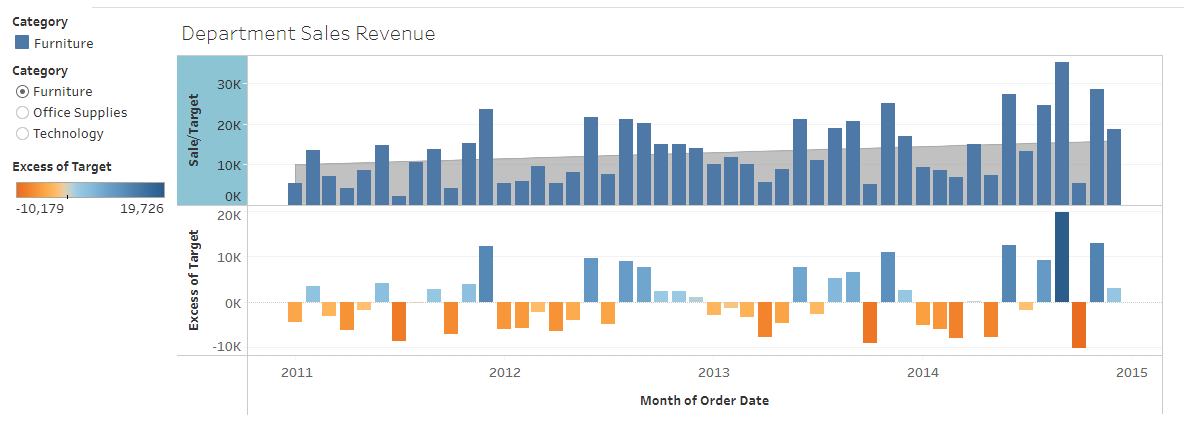
Description automatically generated with low confidence

* The Target sales is differing for each category. Some were having a standard raise on their target sales for each year were as for others it was raising in high level based on their demands.
* Dual Axis – Right click on the first chart 🡪 Dual Axis 🡪 Right click again in the right end 🡪 Synchronize the axis
* It is important to the axis as the Target axis has different values when compared to sales and so once we have synchronised the axis; both target and sales will have similar values.
* Chart, bar chart

  Description automatically generated

## Creating Calculated fields in Blend (Advanced Topic)

* How to create a calculated field with in a blend and what that means is how to create a calculated field that requires data elements from both of the tables which are being blended
* Actual sale – target sale; helps to determine whether the department was performing below bar or above bar

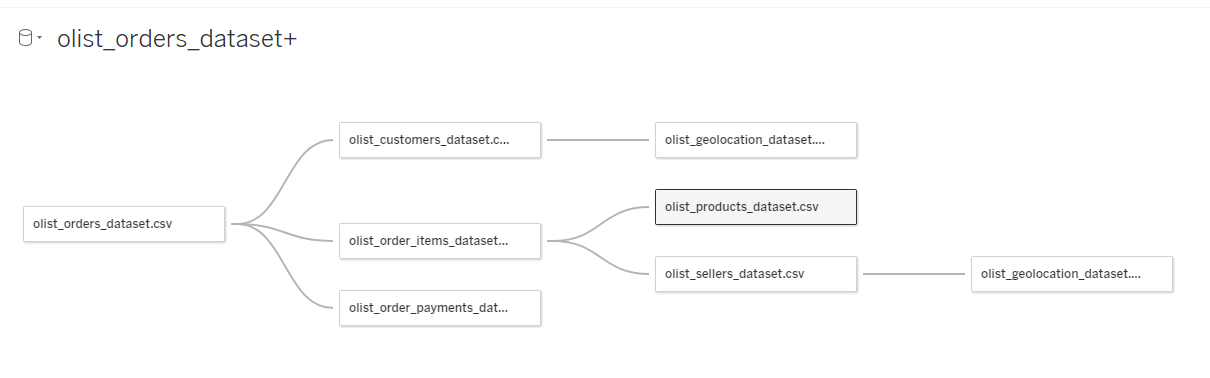


## New Challenge and Dataset

* 2020.2 V Tableau has Data Model and concept of relationship
* We will learn about data Model and understand the analysis of Brazilian e-commerce datasets

## The Data Model (v2020.2 update)

Follow the data schema pattern in Kaggle and replicate the same in tableau



## Working with multiple Relationships in Tableau (v2020.2 update)

* You are a data scientist at the Brazilian e commerce store. Your manager has asked you to create a map showing sellers in their geo locations with bubbles to illustrate how much money and total payments they have received from customers.

## Section Recap

## QUIZ 4: Joining, Blending and Relationships; PLUS: Dual Axis Charts

# Section-6- Table Calculation, Advanced Dashboards, Storytelling

## Section Intro

## Downloading the Dataset and Connecting to Tableau

## Mapping: how to Set Geographical Roles

## Creating Table calculation for Gender

## Creating Bins and Distribution for Age

## Leveraging the Power of Parameters

## How to Create a Tree Map Chart

## Creating a Customer Segmentation Dashboard

## Advanced Dashboard Interactivity

## Analysing the Customer Segmentation Dashboard

## Creating a Storyline

## QUIZ 5: Table Calculation, Advanced Dashboards, Storytelling

# Section-7- Advanced Data Preparation

## Section Intro

## What format your data should be in

## Data Interpreter

## Pivot

## Splitting a Column into Multiple Columns

## Meta Data Grid

## Fixed Geographical Data Errors in Tableau

## QUIZ 6: Advanced Data Preparation

# Section-8- Clusters, Custom Territories, Design Features

## The Challenge: Startup Expansion Analytics

## Custom Territories via Groups

## Custom Territories via Geographical Roles

## Adding Highlighter

## Clustering in Tableau

## Cross-Database Joins

## Modelling with Clusters

## Saving your Clusters

## New Mobile Features

## Section Recap

# Section-9- What is new in Tableau

## Section Intro

## Data From PDF files

## Connecting to PDF

## Connecting to Spatial Files

## Joining to Spatial Files

## Putting it All Together

## Step and Jump Line Chart

## Viz in Tooltip

# Section-10- Conclusion

## Course overview and Roadmap

## EXTRA – Interview with a graduate of this course

## Some Additional Resources