EXERCISE 01: Connecting to a Data Source

In this exercise, you will connect to a data source for the first time, which is the very first step when analyzing data in Tableau.

There are many types of data sources that you can connect to, but for the purposes of this exercise, you will work with an Excel file—in this case, Sample-Superstore.xls, which comes in-built with Tableau and contains sales and profit data for a company.

Perform the following steps to complete the exercise:

1) Select the **Microsoft Excel** option from the **To a File** option under **Connect** on the left-hand side of the landing page. You should see the following screen:

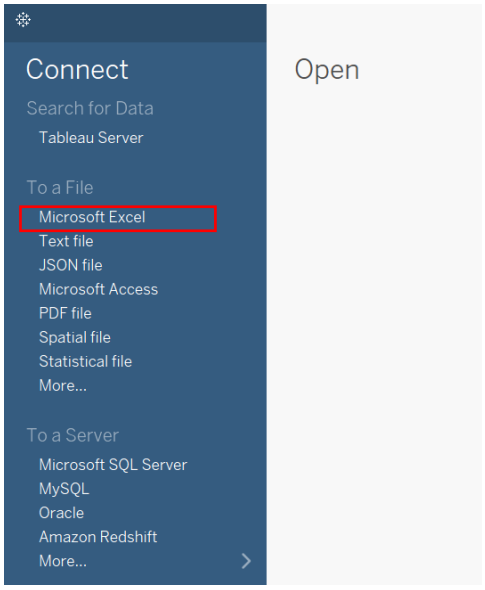

Figure 1: A screenshot showing the Connect to Microsoft Excel option

2) Once you have selected this option, it will ask you to browse the Excel file that you wish to connect to. To do this, connect to **Sample-Superstore.xls**, which can be found in **Documents>My Tableau Repository>Datasources**, or can also be downloaded from **E-Learning**, Refer to the following screenshot:

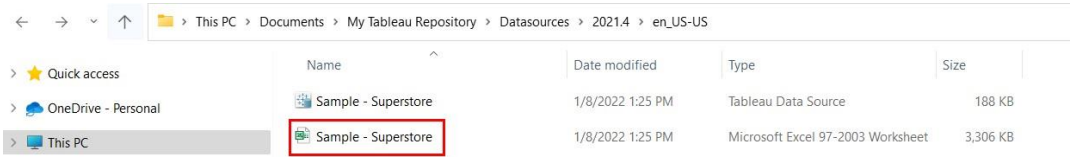

Figure 2: A screenshot showing the Sample - Superstore.xls data under My Tableau Repository

3) Once you have connected to this data source, you will see the ***data connection page*** of Tableau Desktop, as shown in the following screenshot. Review the following notes to better understand what you're looking at:

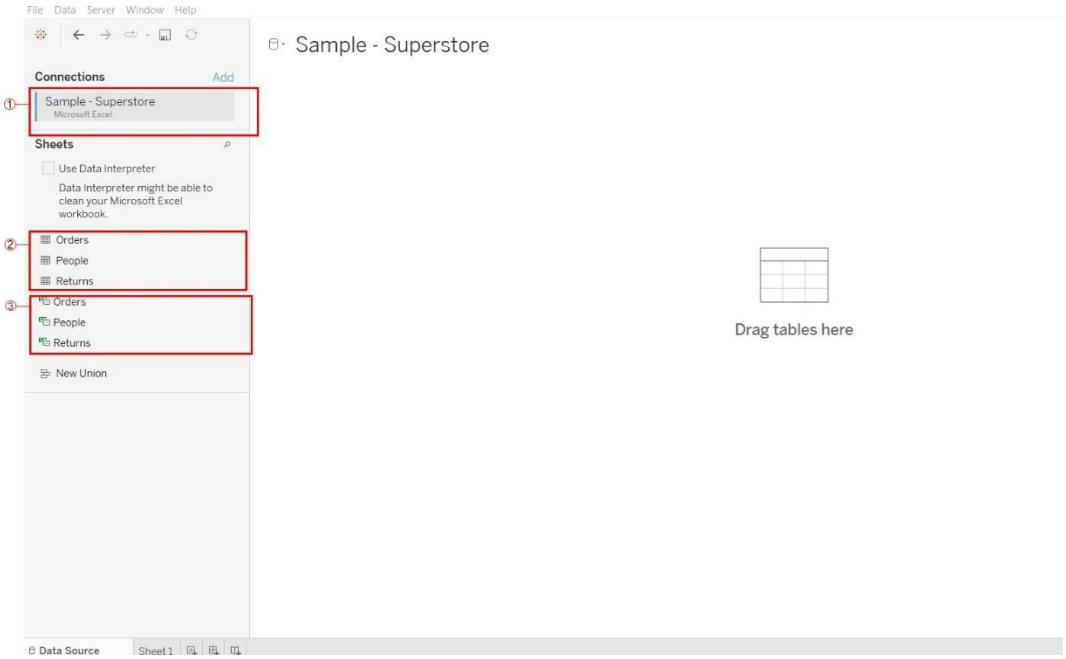

Figure 3: A screenshot showing the data connection page in Tableau Desktop

Section 1: This highlights the *data source* that you have connected to. This is the **Sample - Superstore.xls** file that you just established a connection with. One point to note here is that just because you have established a connection to this Excel file does not mean that you have connected to the data.

Section 2: These are the tables/worksheets in your **Sample - Superstore. xls** file, which is where the actual data resides. The **Orders** table contains the list of all transactions from this retail superstore and contains data at an order level. This order level contains details of the day, product, and customer levels. Refer to the following figure to take a glance at the **Orders** table:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit
1	CA-2018-152156	08-11-2018	11-11-2018	Second	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-10001	Furniture	Bookcases	Bush Somers	261.96	2	0	41.9136
2	CA-2018-152156	08-11-2018	11-11-2018	Second	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-10000	Furniture	Chairs	Hon Deluxe f	731.94	3	0	219.582
3	CA-2018-138688	12-06-2018	16-06-2018	Second	DV-13045	Darrin Van Huff	Corporate	United States	Los Angele	California	90036	West	OFF-LA-10000	Office Suppl	Labels	Self-Adhesiv	14.62	2	0	6.8714
4	US-2017-108966	11-10-2017	18-10-2017	Standar	SO-20335	Sean O'Donnell	Consumer	United States	Fort Laude	Florida	33311	South	FUR-TA-10000	Furniture	Tables	Bretford CR4	957.5775	5	0.45	-383.031
5	US-2017-108966	11-10-2017	18-10-2017	Standar	SO-20335	Sean O'Donnell	Consumer	United States	Fort Laude	Florida	33311	South	OFF-ST-10000	Office Suppl	Storage	Eldon Fold 'N	22.368	2	0.2	2.5164
6	CA-2016-115812	09-06-2016	14-06-2016	Standar	BH-11710	Brosina Hoffman	Consumer	United States	Los Angele	California	90032	West	FUR-FU-10001	Furniture	Furnishings	Eldon Expres	48.86	7	0	14.1694
7	CA-2016-115812	09-06-2016	14-06-2016	Standar	BH-11710	Brosina Hoffman	Consumer	United States	Los Angele	California	90032	West	OFF-AR-10002	Office Suppl	Art	Newell 322	7.28	4	0	1.9656

Figure 4: A screenshot showing a glimpse into the Orders worksheet of Sample - Superstore.xls

The **People** table contains just two columns: **Region** and **Person**. The **Person** column is the list of managers for each **Region**. Refer to the following screenshot to take a glance at the **People** table:

	A	B
1	**Person**	**Region**
2	Anna Andreadi	West
3	Chuck Magee	East
4	Kelly Williams	Central
5	Cassandra Brandow	South
6		

Figure 5: A screenshot showing a glimpse into the People worksheet of Sample - Superstore.xls

The **Returns** table contains the list of all the transactions/orders that were returned. So, again, only two columns: **Returned** and **Order ID**. Refer to the following screenshot to take a glance at the **Returns** table:

	A	B
1	**Returned**	**Order ID**
2	Yes	CA-2016-100762
3	Yes	CA-2016-100762
4	Yes	CA-2016-100762
5	Yes	CA-2016-100762
6	Yes	CA-2016-100867
7	Yes	CA-2016-102652
8	Yes	CA-2016-102652
9	Yes	CA-2016-102652
10	Yes	CA-2016-102652
11	Yes	CA-2016-103373
12	Yes	CA-2016-103744
13	Yes	CA-2016-103744

Figure 6: A screenshot showing a glimpse into the Returns worksheet of Sample - Superstore.xls

Section 3: This is the list of *Named Ranges* that were created on the aforementioned tables/ worksheets (that is, **Orders**, **People**, and **Returns**) of the **Sample - Superstore.xls** data source. *Named Ranges* are a feature in *Microsoft Excel*, and Tableau gives you the option of reading data from these predefined *Named Ranges*. To understand more about these *Named Ranges in Excel*, please refer to the following link: https://support.microsoft.com/en-us/office/ define-and-use-names-in-formulas-4d0f13ac-53b7-422e-afd2-abd7ff379c64?ui=en- us&rs=en-us&ad=us.

4) So, at this point, you have made a connection to the **Sample - Superstore. xls** file; however, you are yet to establish a connection to the data to be able to read it in Tableau for your analysis. To do so, drag the **Orders** worksheet from the left-hand side list and drop it into the top blank section, which reads **Drag sheets here**. (If you are working with a version later than 2020.1, this may instead read **Drag tables here**.) Please note that you need to use the **Orders** worksheet and not the **Orders** named range since the data in the named range could be limited compared to the data in the **Orders** worksheet. Refer to the following screenshot:

Figure 7: A screenshot showing how to read data from the Orders worksheet of Sample -Superstore.xls

5) Once you drag and drop the **Orders** worksheet into the **Drag sheets here** section, you will see the view update for you, as shown in the following screenshot:

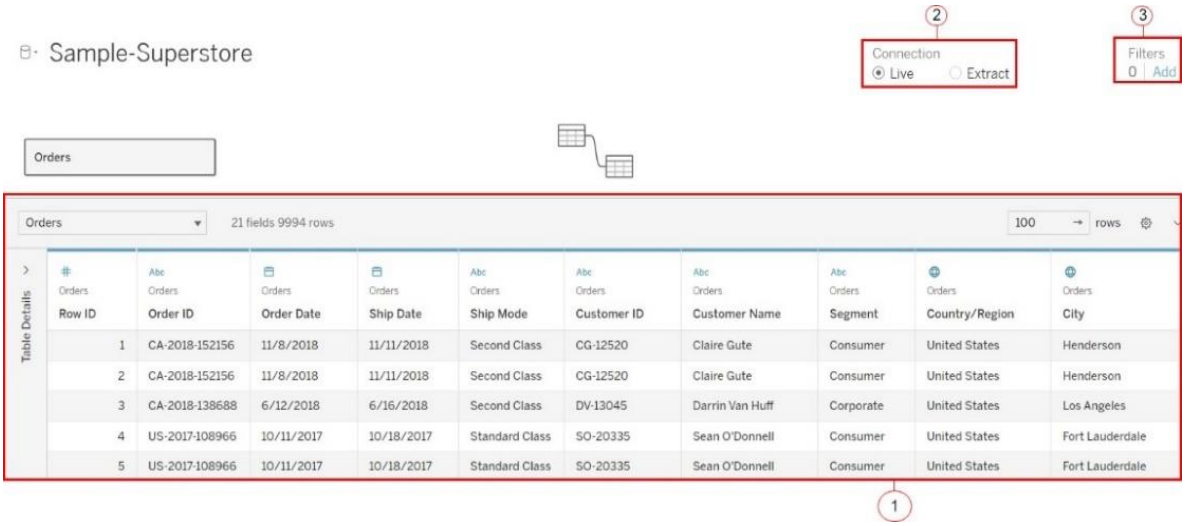

Figure 8: A screenshot showing the view after dragging and dropping the Orders worksheet

The preceding figure shows the view after fetching the **Orders** worksheet into the **Drag sheets here** section. Review the highlighted sections in the screenshot and the corresponding notes below to understand more.

Section 1: This is the *preview section* where you get to see a quick preview (about 100 rows) of your **Orders** data. This is where you can quickly take stock of your data and make sure you have all the necessary columns to work with.

Section 2 *(Only Tableau Desktop)*: This is the **Connection** option. It has two options to choose from, **Live** and **Extract**. A **Live** connection is the option that you use when you want to connect to data in real time. This means that basically any changes at the data end will be reflected in Tableau. However, a quick point to note here is that the **Live** connection option relies on the data sources to process all the queries, and this could lead to

performance issues in Tableau if the backend data source is a slow-performing data source. The **Extract** connection, on the other hand, is a snapshot of your data stored in a Tableau propriety format called **Tableau Data Extract**, which uses the file extension **.hyper**. Since the **.hyper** file only has a snapshot of the data, it will have to be refreshed if you need to see and use the updated data.

Section 3: This is the **Filters** option, which is used to limit the amount of data that is read and used in Tableau. This works for both the **Live** and **Extract** options mentioned earlier.

6) Connect **to** your **Orders** data from **Sample - Superstore.xls**. Refer to the following screenshot:

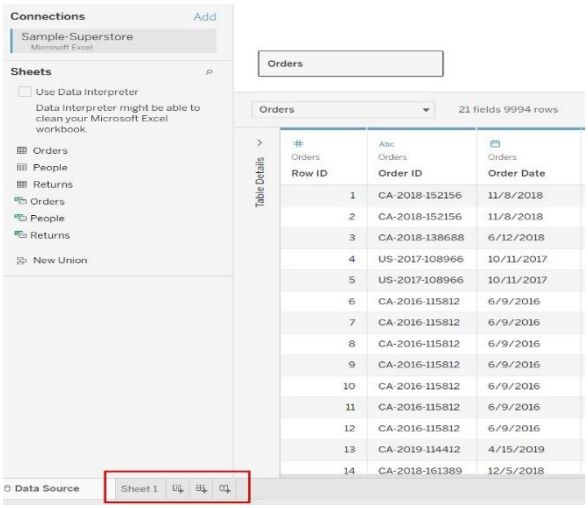

Figure 9: A screenshot showing the Go to Worksheet option

7) Now, the final step for fetching the data for your analysis is to click on **Sheet1**, and from there, select **Go to Worksheet**. With this, you will have read the data into Tableau Desktop and will now be able to start using it. Refer to the following screenshot:

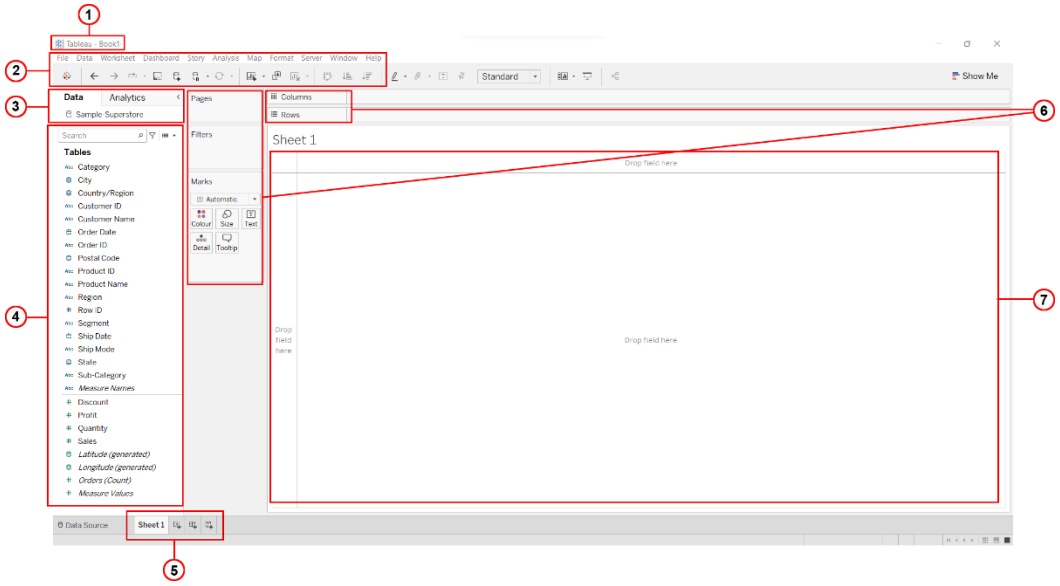

Figure 10: A screenshot showing the workspace of Tableau

The preceding screenshot shows the *Tableau workspace*. This is the space in which you will create your visualizations going forward. Let's quickly go through the highlighted sections in the screenshot to understand the workspace in more detail.

Section 1: This is the *workbook name*. As mentioned previously, a *workbook* in Tableau is a file that consists of multiple worksheets and/or dashboards and/or storyboards. By default, it is named **Book1** (as shown in the image). However, you can assign any new name you like when you save the workbook.

Section 2: This is the *toolbar section*, and this consists of various options that help you explore the various features and functionalities available in Tableau.

Section 3: This is the *side bar* area, which contains the **Data** pane and the **Analytics** pane. The **Data** pane shows the details of the fields coming from the data, which are classified as either **Dimensions** or **Measures**. The **Analytics** pane, on the other hand, shows the various analyses, such as constant line, average line, median with quartiles, totals, trend line, forecast line, and clusters that can be performed on the view that you create. To begin with, the **Analytics** pane is disabled or grayed out and will only start appearing when you create a view or visual.

Section 4: This is the **Dimensions** and **Measures** section, which technically is part of the **Data** pane. **Dimensions** are all the fields from the data that are categorical, descriptive, or qualitative in nature, such as **Customer Name**, **Product Name**, **Order ID**, and **Region**. These, when fetched in the view, will result in each data member of that field being displayed in the view. **Measures**, on the other hand, are fields from the data that are quantitative in nature and can be aggregated as either sum, average, minimum, maximum, standard deviation, variance, and so on. These, when fetched in the view, will result in aggregated values being displayed. Examples of **Measures** are fields such as **Sales**, **Profit**, and **Quantity**, which will be aggregated for the purpose of your analysis. Refer to the following screenshot for more clarity:

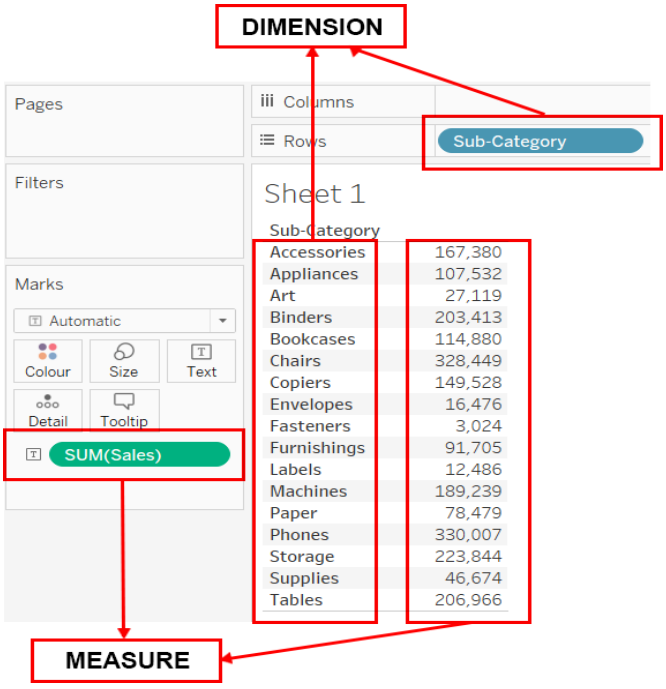

Figure 11: A screenshot showing the difference between Dimensions and Measures

Section 5: This is the **Sheet** tab. Here you get the option to create either a new worksheet, dashboard, or storyboard.

Section 6: These are the various *cards and shelves* available for use in Tableau. Here you can see various shelves such as the **Columns** shelf, **Rows** shelf, **Pages** shelf, and **Filters** shelf, along with the **Marks** card, which contains shelves such as the **Color** shelf, **Size** shelf, **Text** shelf, **Detail** shelf, and the **Tooltip** shelf. These shelves are used to change the appearance and details of your view.

Section 7: This is the **View** section. This is where you will create your visualizations. It can be referred to as the canvas for creating your views and visualizations.