

Batch: D2

Roll No.: 16010122323

Experiment 03

Grade: AA / AB / BB / BC / CC / CD / DD

Title: Tableau- Importing Data

Objective:

- 1. To learn how to import dataset from various file format**
 - 2. To learn how to import dataset from various server**
-

Course Outcome:

CO1: Learn how to locate and download datasets, extract insights from that data and present their findings in a variety of different formats.

Books/ Journals/ Websites referred:

Sosulski, K. - Data Visualization Made Simple: Insights into Becoming Visual
Cole Nussbaumer Knaflie - Storytelling with Data

Resources used:

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

Theory (About Data Preprocessing):

Data preprocessing is a data mining technique which is used to transform the raw data in a useful and efficient format.

Why is Data preprocessing important?

Preprocessing of data is mainly to check the data quality. The quality can be checked by the following

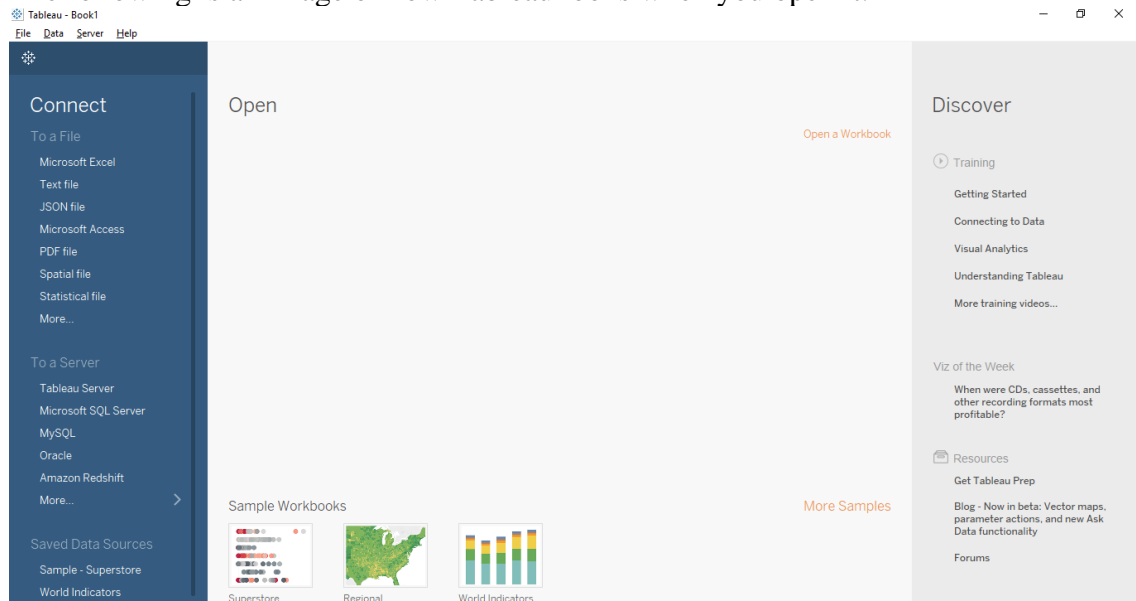
- Accuracy: To check whether the data entered is correct or not.
- Completeness: To check whether the data is available or not recorded.
- Consistency: To check whether the same data is kept in all the places that do or do not match.

- Timeliness: The data should be updated correctly.
- Believability: The data should be trustable.
- Interpretability: The understandability of the data.

Different approaches of importing dataset:

- Import from various file format (pdf, excel, csv, text etc)

The following is an image of how Tableau looks when you open it.



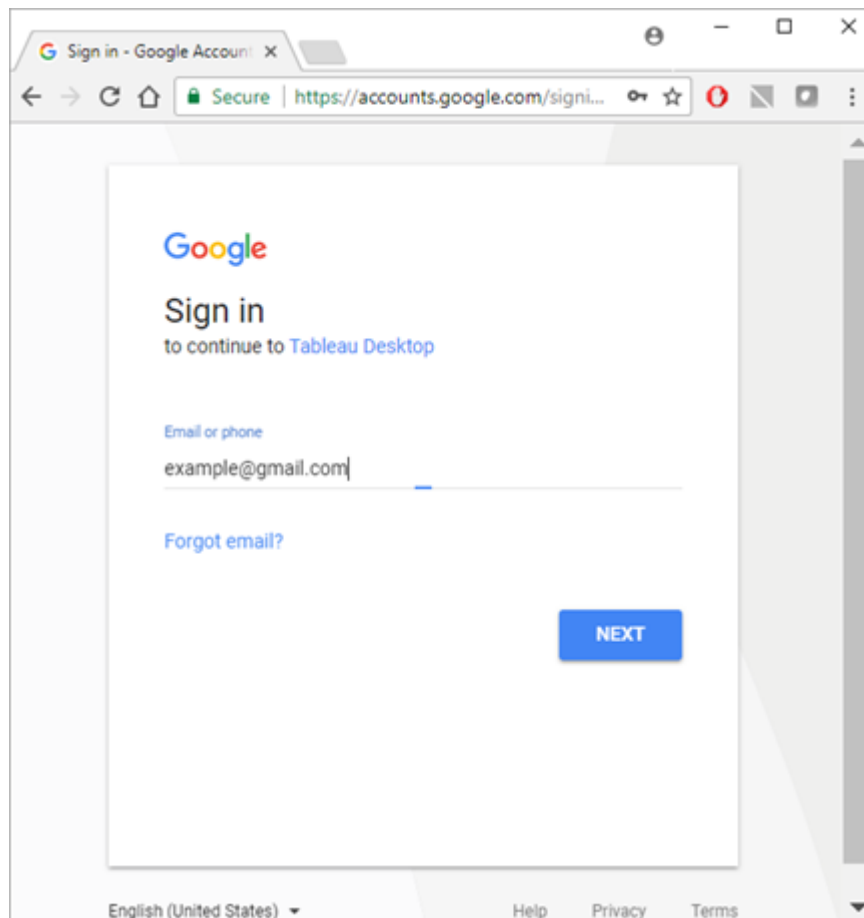
To import data from an excel file, first, click on "Microsoft Excel" under the connect tab. It will open a dialog box where you can navigate to the Excel file in your machine from which you want to import the data. Then, click on the file and click on 'Open'. If there are multiple sheets in your Excel workbook, all of them will be imported automatically and they will be listed as sheets on the left-hand side panel of Tableau.

Remember that any formulas you have written in the Excel won't be imported into Tableau. Also, Tableau queries the file for data and imports the data only. So, if any data is updated in the source file, the connection needs to be refreshed each time the data is modified to keep the data up-to-date in Tableau. Also, remember that pivot tables in excel are not supported by Tableau. File extensions, in this case, would be [.xls](#) and [.xlsx](#).

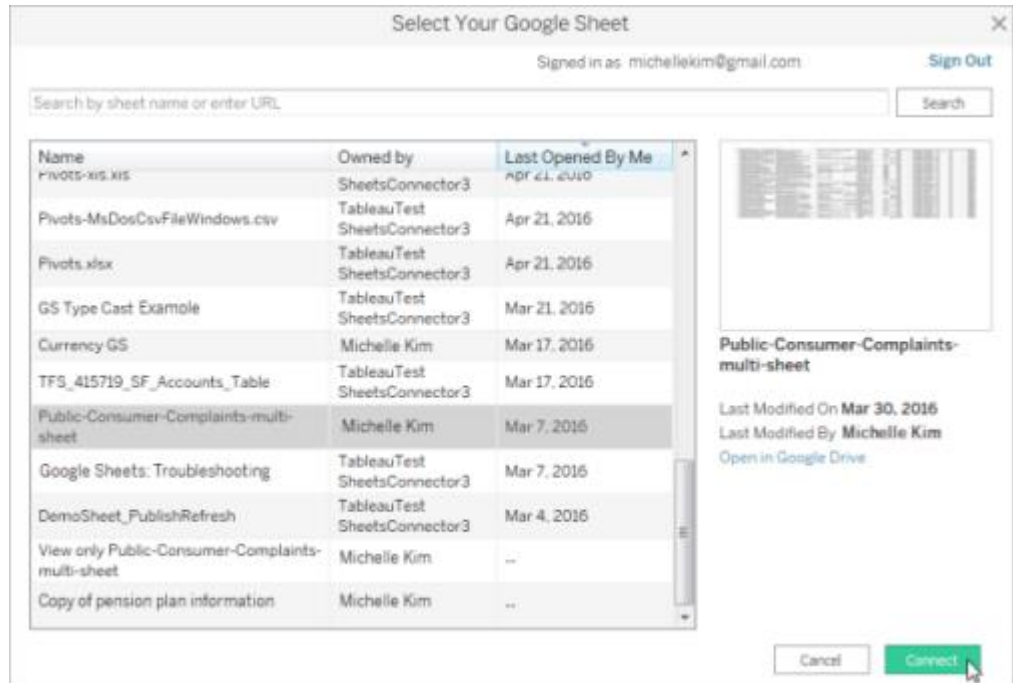
To add more data (which can be from any source) into the existing workbook, just click on "add" which is present beside connections on the left pane. Sometimes Tableau recommends using the inbuilt data interpreter to use while importing the data. There is no need to worry about it; it just formats the data. Adding more data and using the data interpreter are the two most common steps that you will use while importing data from any sources or file formats.

- **Import from server**

1. Start Tableau and under **Connect**, select **Google Sheets**. For a complete list of data connections, select **More** under **To a Server**. In the tab Tableau opens in your default browser, do the following:
 - a. Sign in to Google Sheets using your email or phone, and then select **Next** to enter your password. If multiple accounts are listed, select the account that has the Google Sheets data you want to access and enter the password, if you're not already signed in.



- b. Select **Allow** so that Tableau Desktop can access your Google Sheets data.
- c. Close the browser window when notified to do so.
- d. Select a Google Sheet from the list or use the text box to search for a Google Sheet by name or by URL, and then select **Connect**.



2. On the data source page, do the following:
 - a. (Optional) Select the default data source name at the top of the page, and then enter a unique data source name for use in Tableau. For example, use a data source naming convention that helps other users of the data figure out which data source to connect to.
 - b. If your Google Sheets file has one table, select the sheet tab to start your analysis.
3. Select Your Google Sheet dialog box functionality

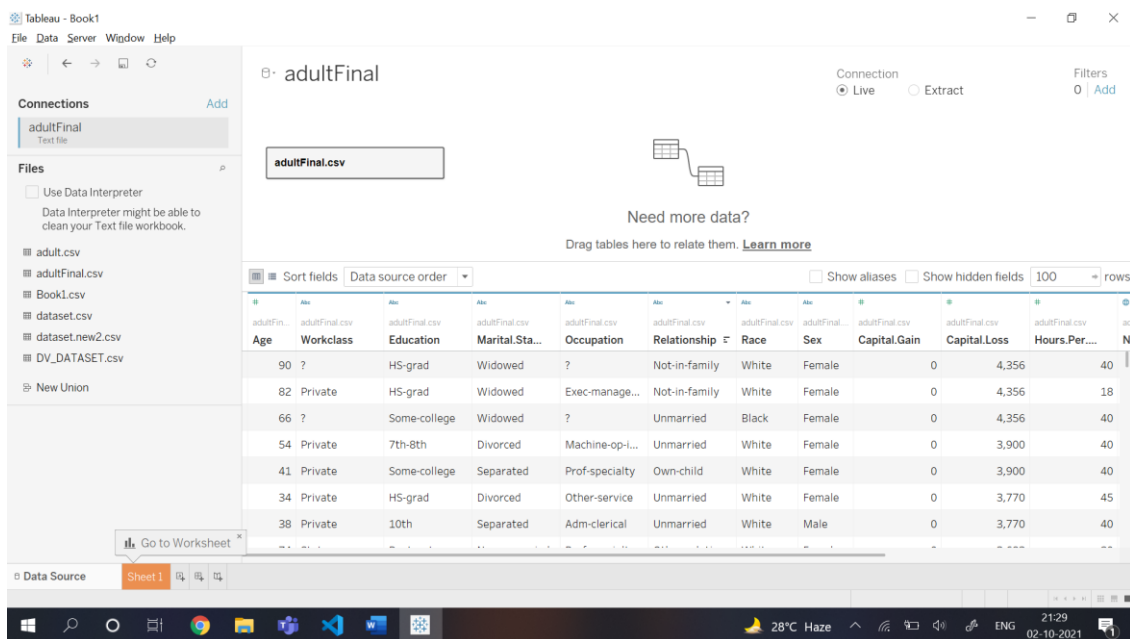
The Select Your Google Sheet dialog box includes the following functionality:

- The list of sheets that you can select from includes your private sheets, sheets shared with you, and the public sheets that you've accessed in the past.
- If you search by URL and the URL doesn't exist or you don't have access to it, an error displays.
- You can select the Name and Last opened by me column names to sort the Google Sheets, and when you select a sheet you can preview it in the right pane. You cannot sort by Owned by.

Platform used by the student:

Excel and Jupyter Notebook

Working:



Conclusion:

Thus, through this experiment we have learnt how to use data visualisation software, namely Tableau and how to import data into the software using various file formats and servers as well

Post Lab Question:

1. List down types of data Tableau can import?

i) String Data type: The collection of characters give rise to the string data type. A string is always enclosed within a single or double inverted comma. The samples of the string are — “Vansh”, “Hi! How are you?”, and “GeeksforGeeks”, etc.

We can divide String data type into two types, Char and Varchar.

- **Char string type-** Char data type normally stores alphanumeric data values having fixed lengths. If the user enters a string value which is greater than the fixed length of the Char data type, then the system returns an error.
- **Varchar string type-** Varchar data type also stores alphanumeric data values. As the name suggests, Varchar stores data values having a variable length. So, the user can enter as many string values as they want, without facing any restriction from the system.

ii) Numeric Data type: This data type consists of both integer type or floating type. Out of which users prefer to use integer type over floating type, as it is difficult to accumulate the decimal point after a certain limit. It also contains a function known as the Round() function which can be used in rounding up float values.

iii) Date and Time Data type: Tableau supports all forms of date and time like dd-mm-yy, or mm-dd-yyyy, etc. And the time data values can be in the form of a decade, year, quarter, month, hour, minutes, seconds, etc. Whenever the user enters data and time values, Tableau automatically registers it under Date data type and Date & Time data value.

iv) Boolean Data type: As a result of relational calculations, boolean data type values are formed. The boolean data values are either True or False. Many a time the result of a relational calculation is unknown, in this situation Null data values are used.

v) Geographic Data type: All values that are used in maps, comes under geographic data type. The example of geographic data values is country name, state name, city, region, postal codes, etc.

vi) Cluster or Mixed Data type: Sometimes data set contains values having a mixture of data types. Such values are known as cluster group values or mixed data values. In such a situation, users have the option either to handle it manually or allow Tableau to operate on it.

2. What is significance of Measures and Dimensions in dataset stored in Tableau?

Data fields are made from the columns in your data source. Each field is automatically assigned a data type (such as integer, string, date), and a role: Discrete Dimension or Continuous Measure (more common), or Continuous Dimension or Discrete Measure (less common).

- Dimensions contain qualitative values (such as names, dates, or geographical data). You can use dimensions to categorize, segment, and reveal the details in your data. Dimensions affect the level of detail in the view.
- Measures contain numeric, quantitative values that you can measure. Measures can be aggregated. When you drag a measure into the view, Tableau applies an aggregation to that measure (by default).

Date:

Signature of faculty in-charge