**Batch:H2-4 Roll No.:16010122257**

**Experiment 03**

|  |
| --- |
| **Title:** Importing Data and exploring the data |

# Objective:

# 1. To learn how to import dataset from various file format

# *Text, csv, pdf, excel, word*

# 2. To learn how to import dataset from various server

# Example (MySQL, MSSQL, Oracle, DB2, Google spreadsheets, Google drive, AWS, other)

# Minimum One connection with Server (Student choice mentioned in Objective 2)

# 3. Explore the data over platform

# Live data and Extracted data

# Data types

# Combining two data sources

# View data

# Sort option

# Measures and dimensions

# Splitting the column

# Discrete and continues values

# Drill down and Hierarchies

# Grouping

# 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:

Tableau Online Help Documentation

# Resources used:

* <https://www.kaggle.com>
* <https://archive.ics.uci.edu>
* <https://www.tableau.com/trial/tableau-cloud?utm_campaign_id=2017049&utm_campaign=APAC_IND_FY24Q3_DATABUYER_Google_Brand_Exact_Base_BR&utm_medium=Paid+Search&utm_source=Google+Search&utm_language=EN&utm_country=IND&kw=tableau&adgroup=https://www.tableau.com/trial/tableau-cloud?utm_campaign_id=2017049&utm_campaign=APAC_IND_FY24Q3_DATABUYER_Google_Brand_Exact_Base_BR&utm_medium=Paid+Search&utm_source=Google+Search&utm_language=EN&utm_country=IND&kw=tableau&adgroup=YYY&adused=669595729426&matchtype=e&placement=&nc=7013y0000029zDoAAI&d=7013y000000vYhH&adused=669595729426&matchtype=e&placement=&nc=7013y0000029zDoAAI&d=7013y000000vYhH&gclid=Cj0KCQjw9rSoBhCiARIsAFOiplnHJlKu7vwMo8uzesRpb1axrcmYpPtr4C822_MG8HLfsQxEyo3J46AaAi3eEALw_wcB&gclsrc=aw.ds>

# Theory (About Data Preprocessing):

# Data preprocessing involves the cleaning, transformation, and organization of raw data into a usable format. It includes tasks like handling missing values, removing duplicates, transforming data types, and more.

# 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.

# Following points should be written by students

# Different approaches of importing dataset:

# Import from various file format (PDF, Excel, .CSV, .txt)

# 

# (Tableau on opening)

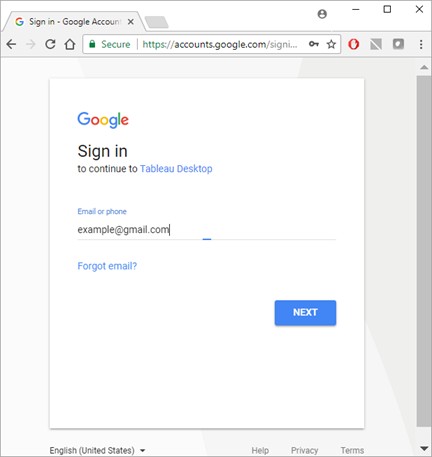
# To import data from excel file in Tableau:

* Click the “Microsoft Excel” in the “To a file” of “Connect” section, ( Connect -> To a file -> Microsoft Excel.)
* On clicking, we get the file selector. You can change the file location or can open any Excel file from any location.
* Select an Excel file and open it.
* After opening an Excel file in tableau you will get a screen as shown below with the connected data.
* Here, you will get your dataset (Excel file content) with all sheets it has.
* You can view the data/sheet by dragging and dropping it and further operations can be performed on it.

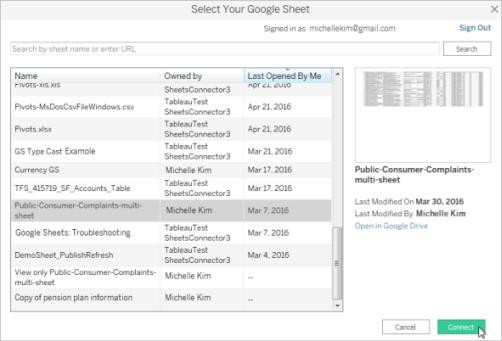
Remember that any formulas you have written in the Excel won't be imported into Tableau. Also, Tableau queries the file for data and imports the data only.So, if any data is updated in the source file, the connection needs to be refreshed each time the data is modified to keep the data up-to-date in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



* + 1. Select **Allow** so that Tableau Desktop can access your Google Sheets data.
    2. Close the browser window when notified to doso.
    3. Select a Google Sheet from the list or use the text box to search for a GoogleSheet 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 GoogleSheets file has one table, select the sheet tab to start your analysis.

3. Select Your GoogleSheet dialog box functionality

Then select,your GoogleSheet 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.

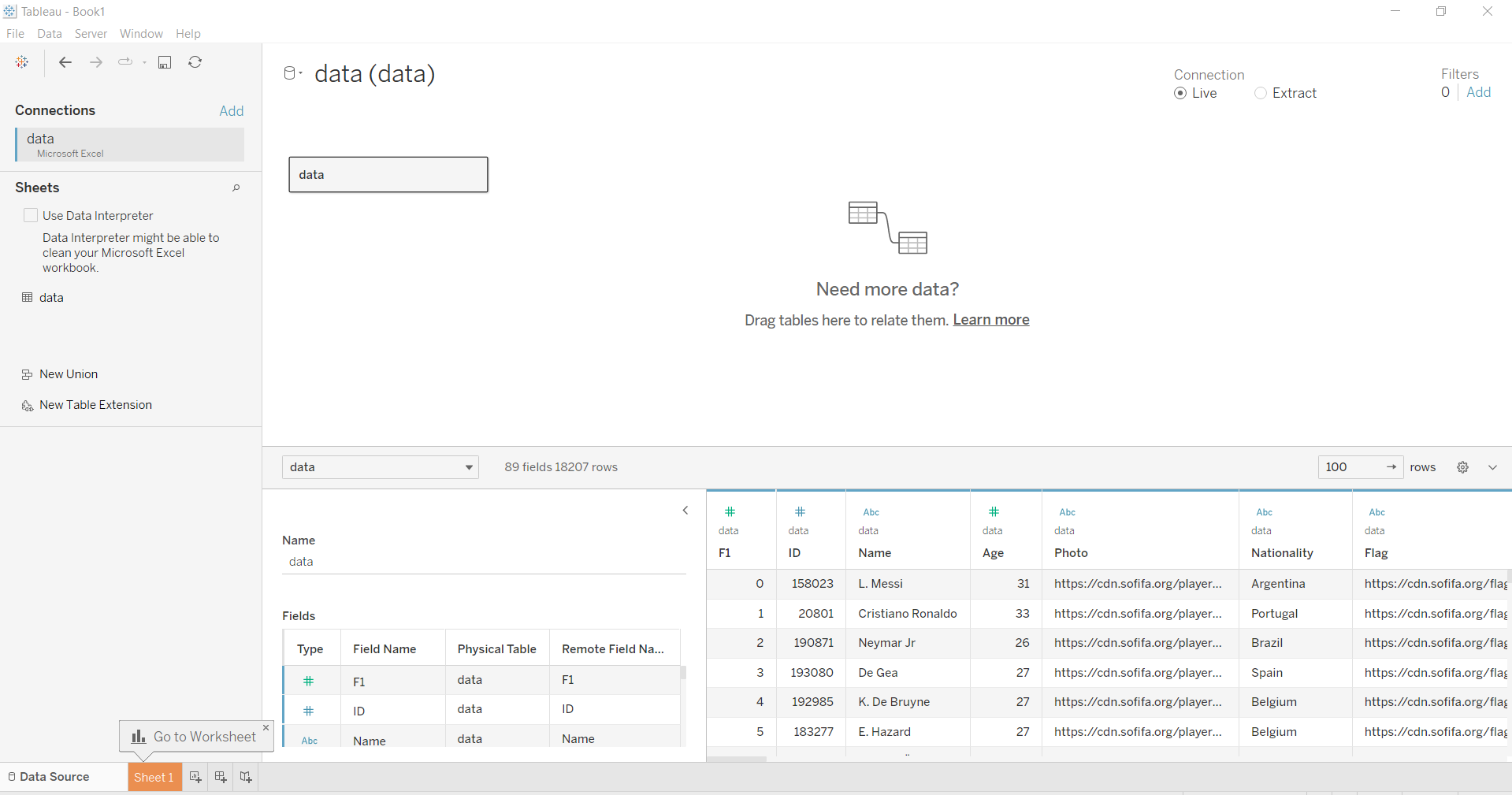
• If you search by URL and the URL doesn't exist or you don't have access to it,an error displays.

• 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 (Students should write in their own words):

# Through this experiment, I have learnt how to successfully import a wide range of datasets from different formats and servers into Tableau. I also gained insights into exploring of data, understanding data types, working with measures and dimensions, and creating meaningful visualizations. I believe that this hands-on experience will help me develop necessary skills in the field of data analysis.

**Date: \_\_\_\_\_\_\_\_\_\_\_\_\_ Signature of faculty in-charge**

# Post Lab Question:

# List down types of data Tableau(any other you use) can import?

# Tableau can import various data types including Excel, Access, Text, PDF, JSON, statistical files, and more.

# Basic data types 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 — “Boy”, “Hi! Hello man!”, and “Clouds”, 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(any other you use)?

Measures are quantitative variables used for analysis, while dimensions are categorical variables used for grouping and filtering. In Tableau, measures are typically used in numerical calculations, while dimensions are used for creating hierarchies and organizing data for visualization. Each field is automatically assigned a data type (such as integer, string, date), and a role: Discrete Dimension orContinuous Measure (more common), or Continuous Dimension or Discrete Measure(lesscommon).

* Dimensions contain qualitative values (such as names, dates, or geographical data). They can be used to categorize, segment, and reveal the details in the data. Dimensions affect the detail’s level in the view.

# Measures contain numeric,quantitative values that can be measured,hence the name.They can be aggregated.On dragging a measure into the view,by default, Tableau applies an aggregation to that measure.