

Batch: H2-1 Roll No.: 16010122151

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:

(Students should write)

Google
Kaggle

Resources used:

Kaggle

Tableau

Theory (About Data Preprocessing):

Data preprocessing is an important step in the data mining process. It refers to the cleaning, transforming, and integrating of data in order to make it ready for analysis. The goal of data preprocessing is to improve the quality of the data and to make it more suitable for the specific data task.

Following points should be written by students

Different approaches of importing dataset:

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

Platform used by the student:

Tableau

Working: (Screenshots of various file format imported in software)

Tableau - Book3

File Data Server Window Help

Connections [Add](#)

Sample - Superstore
Microsoft Excel

Sheets [p](#)

☐ Use Data Interpreter
Data Interpreter might be able to clean your Microsoft Excel workbook.

Orders
People
Returns
People
Returns
New Union
New Table Extension

Orders (Sample - Superstore)

Connection ☒ Live ☐ Extract

Filters 0 [Add](#)

Orders

Need more data?
Drag tables here to relate them. [Learn more](#)

Orders 21 fields 10194 rows 100 rows

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country/Region
1	US-2019-103800	03-01-2019	07-01-2019	Standard Class	DP-13000	Darren Powers	Consumer	United States
2	US-2019-112326	04-01-2019	08-01-2019	Standard Class	PO-19195	Phillina Ober	Home Office	United States
3	US-2019-112326	04-01-2019	08-01-2019	Standard Class	PO-19195	Phillina Ober	Home Office	United States
4	US-2019-112326	04-01-2019	08-01-2019	Standard Class	PO-19195	Phillina Ober	Home Office	United States
5	US-2019-141817	05-01-2019	12-01-2019	Standard Class	MB-18085	Mick Brown	Consumer	United States
6	US-2019-167199	06-01-2019	10-01-2019	Standard Class	ME-17320	Maria Etezadi	Home Office	United States
7	US-2019-167199	06-01-2019	10-01-2019	Standard Class	ME-17320	Maria Etezadi	Home Office	United States
8	US-2019-106054	06-01-2019	07-01-2019	First Class	JO-15145	Jack O'Briant	Corporate	United States
9	US-2019-167199	06-01-2019	10-01-2019	Standard Class	ME-17320	Maria Etezadi	Home Office	United States

DATA TYPES:

Number (decimal)

✓ Number (whole)

Date & Time

Date

String

Boolean

✓ Default

Geographic Role ▶

SORTED BY ROW ID

# Orders	Abc Orders	📅 Orders
Row ID	Order ID	Order Date
1	US-2019-103800	03-01-2019
2	US-2019-112326	04-01-2019
3	US-2019-112326	04-01-2019
4	US-2019-112326	04-01-2019
5	US-2019-141817	05-01-2019
6	US-2019-167199	06-01-2019
7	US-2019-167199	06-01-2019
8	US-2019-106054	06-01-2019
9	US-2019-167199	06-01-2019
10	US-2019-167199	06-01-2019
11	US-2019-167199	06-01-2019
12	US-2019-130813	06-01-2019
13	US-2019-167199	06-01-2019

SORTED BY SHIP DATE

 Orders Order Date	 Orders Ship Date 	 Orders Ship Mode
03-01-2019	07-01-2019	Standard Class
06-01-2019	07-01-2019	First Class
04-01-2019	08-01-2019	Standard Class
04-01-2019	08-01-2019	Standard Class
04-01-2019	08-01-2019	Standard Class
06-01-2019	08-01-2019	Second Class
06-01-2019	10-01-2019	Standard Class
06-01-2019	10-01-2019	Standard Class
06-01-2019	10-01-2019	Standard Class
06-01-2019	10-01-2019	Standard Class
06-01-2019	10-01-2019	Standard Class
06-01-2019	10-01-2019	Standard Class
06-01-2019	10-01-2019	Standard Class

#	Abc
Orders	Orders
Row ID	Order ID
1	US-2019-103800
2	US-2019-112326
3	US-2019-112326
4	US-2019-112326
5	US-2019-141817
6	US-2019-167199
7	US-2019-167199
8	US-2019-106054
9	US-2019-167199
10	US-2019-167199
11	US-2019-167199
12	US-2019-130813
13	US-2019-167199

Conclusion (Students should write in their own words):

I have learned how to use tableau to view datasets sort them hide columns split columns and the difference between live and extracted data.

Date: _____

Signature of faculty in-charge

Post Lab Question:

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

Text, date, date and time, numerical values, Boolean value, geographic value

2. What is significance of Measures and Dimensions in dataset stored in Tableau(any other you use)?

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).