

**Batch: A2**

**Roll No.: 16010122041**

**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:**

<https://www.postgresql.org>  
Postgres Documentation

**Resources used:**

<https://www.tutorialspoint.com/postgresql/index.html>  
[help.tableau.com/current/pro/desktop/en-us/examples\\_postgresql.html](https://help.tableau.com/current/pro/desktop/en-us/examples_postgresql.html)

**Theory (About Data Preprocessing):**

**Steps Involved in Data Pre-processing:**

**1. Data cleaning**

The data can have many irrelevant and missing parts. To handle this part, data cleaning is done. It involves handling of missing data, noisy data etc.

**a) Missing Data:**

This situation arises when some data is missing in the data. It can be handled in various ways. Some of them are:

**1. Ignore the tuples:**

This approach is suitable only when the dataset we have is quite large and multiple values are missing within a tuple.

## **2. Fill the missing values:**

There are various ways to do this task. You can choose to fill the missing values manually, by attribute mean or the most probable value.

### **b) Noisy data:**

Noisy data is a meaningless data that can't be interpreted by machines. It can be generated due to faulty data collection, data entry errors etc. It can be handled in following ways:

#### **1. Binning method:**

This method works on sorted data in order to smooth it. The whole data is divided into segments of equal size and then various methods are performed to complete the task. Each segment is handled separately. One can replace all data in a segment by its mean or boundary values can be used to complete the task.

#### **2. Regression:**

Here data can be made smooth by fitting it to a regression function. The regression used may be linear (having one independent variable) or multiple (having multiple independent variables).

#### **3. Clustering:**

This approach groups the similar data in a cluster. The outliers may be undetected, or it will fall outside the clusters.

## **2. Data transformation:**

This step is taken to transform the data in appropriate forms suitable for mining process. This involves following ways:

- **Normalization:**

It is done to scale the data values in a specified range (-1.0 to 1.0 or 0.0 to 1.0)

## **3. Data Reduction:**

Since data mining is a technique that is used to handle huge amount of data. While working with huge volume of data, analysis became harder in such cases. To get rid of this, we use data reduction technique. It aims to increase the storage efficiency and reduce data storage and analysis costs.

## Different approaches of importing dataset:

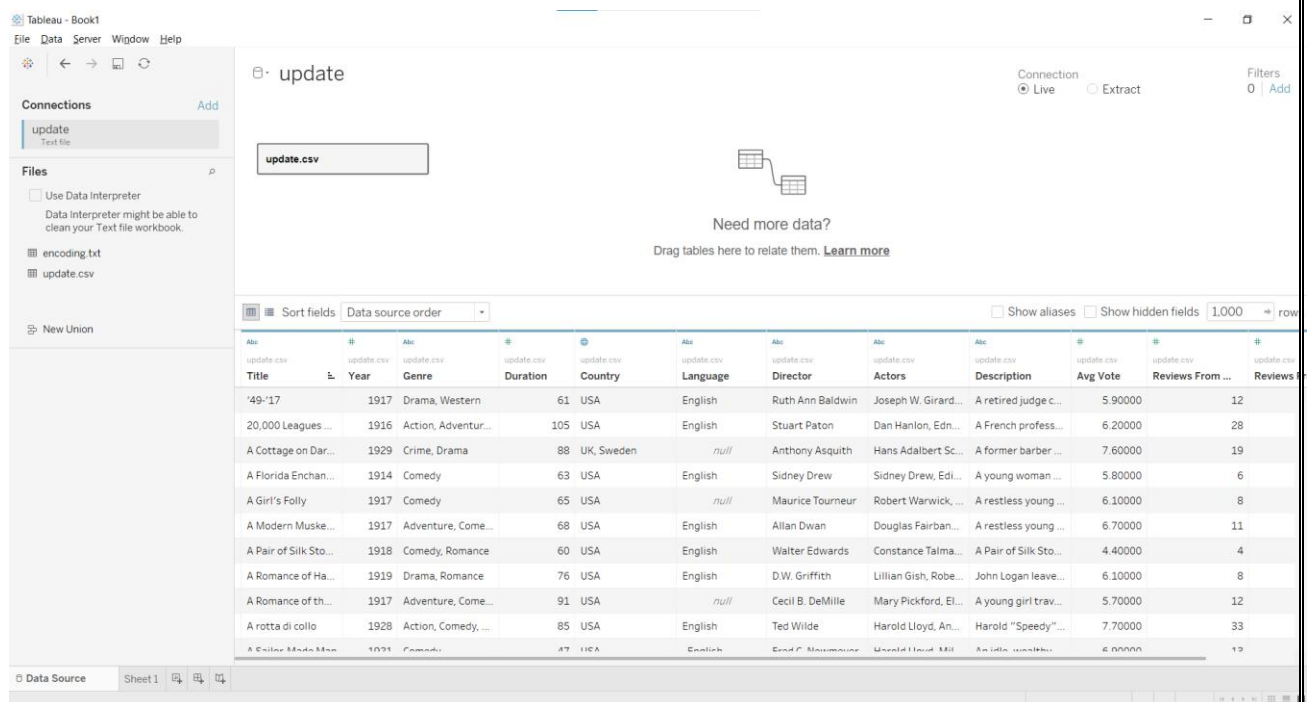
- Import from various file format (pdf, excel, csv, text etc)
- Import from server (min one server of choice)

Platform used by the student:

### Approach 1:

Importing CSV file through import

Tableau



Title	Year	Genre	Duration	Country	Language	Director	Actors	Description	Avg Vote	Reviews From	Reviews To
'49-'17	1917	Drama, Western	61	USA	English	Ruth Ann Baldwin	Joseph W. Girard...	A retired judge c...	5.90000		12
20,000 Leagues ...	1916	Action, Adventur...	105	USA	English	Stuart Paton	Dan Hanlon, Edn...	A French profess...	6.20000		28
A Cottage on Dar...	1929	Crime, Drama	88	UK, Sweden	null	Anthony Asquith	Hans Adalbert Sc...	A former barber ...	7.60000		19
A Florida Enchan...	1914	Comedy	63	USA	English	Sidney Drew	Sidney Drew, Edi...	A young woman ...	5.80000		6
A Girl's Folly	1917	Comedy	65	USA	null	Maurice Tourneur	Robert Warwick, ...	A restless young ...	6.10000		8
A Modern Muske...	1917	Adventure, Come...	68	USA	English	Allan Dwan	Douglas Fairban...	A restless young ...	6.70000		11
A Pair of Silk Sto...	1918	Comedy, Romance	60	USA	English	Walter Edwards	Constance Talma...	A Pair of Silk Sto...	4.40000		4
A Romance of Ha...	1919	Drama, Romance	76	USA	English	D.W. Griffith	Lillian Gish, Robe...	John Logan leave...	6.10000		8
A Romance of th...	1917	Adventure, Come...	91	USA	null	Cecil B. DeMille	Mary Pickford, El...	A young girl trav...	5.70000		12
A rotta di collo	1928	Action, Comedy, ...	85	USA	English	Ted Wilde	Harold Lloyd, An...	Harold "Speedy" ...	7.70000		33
A Sailor Made Man	1921	Comedy	87	USA	English	Frank Lloyd	Harold Lloyd, Mi...	A sailor makes a ...	6.00000		12

### Approach 2:

Connecting a live server to Tableau

Server used: PostgreSQL

Platform: Tableau

Step 1: Getting the server started

```
SQL Shell (psql)
Server [localhost]:
Database [postgres]:
Port [5433]:
Username [postgres]:
Password for user postgres:
psql (12.8)
WARNING: Console code page (437) differs from Windows code page (1252)
        8-bit characters might not work correctly. See psql reference
        page "Notes for Windows users" for details.
Type "help" for help.

postgres=#
```

Step 2:

On Tableau, establish remote access to the live server

PostgreSQL ×

General

Initial SQL

Server

localhost

Port

5433

Database

postgres

Authentication

Username and Password ▼

Username

postgres

Password

\*\*\*\*\*

☒ Require SSL

Sign In

Note: SSL certificate not supported

Step 3: Load the CSV on the platform

Tableau - Book1

File Data Server Window Help

Connections **update** Add

Files

- Text file
- encoding.txt
- update.csv

Need more data?

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

Sort fields Data source order

Show aliases Show hidden fields 1,000 rows

update.csv	update.csv	update.csv	update.csv	update.csv	update.csv	update.csv	update.csv	update.csv	update.csv	update.csv	update.csv
Title	Year	Genre	Duration	Country	Language	Director	Actors	Description	Avg Vote	Reviews From ...	Reviews Fr
'49-'17	1917	Drama, Western	61	USA	English	Ruth Ann Baldwin	Joseph W. Girard...	A retired judge c...	5.90000		12
20,000 Leagues...	1916	Action, Adventur...	105	USA	English	Stuart Paton	Dan Hanlon, Edn...	A French profess...	6.20000		28
A Cottage on Dar...	1929	Crime, Drama	88	UK, Sweden	null	Anthony Asquith	Hans Adalbert Sc...	A former barber ...	7.60000		19
A Florida Enchan...	1914	Comedy	63	USA	English	Sidney Drew	Sidney Drew, Edi...	A young woman ...	5.80000		6
A Girl's Folly	1917	Comedy	65	USA	null	Maurice Tourneur	Robert Warwick, ...	A restless young ...	6.10000		8
A Modern Muske...	1917	Adventure, Come...	68	USA	English	Allan Dwan	Douglas Fairban...	A restless young ...	6.70000		11
A Pair of Silk Sto...	1918	Comedy, Romance	60	USA	English	Walter Edwards	Constance Talm...	A Pair of Silk Sto...	4.40000		4
A Romance of Ha...	1919	Drama, Romance	76	USA	English	D.W. Griffith	Lillian Gish, Robe...	John Logan leave...	6.10000		8
A Romance of th...	1917	Adventure, Come...	91	USA	null	Cecil B. DeMille	Mary Pickford, El...	A young girl trav...	5.70000		12
A rotta di collo	1928	Action, Comedy, ...	85	USA	English	Ted Wilde	Harold Lloyd, An...	Harold "Speedy"...	7.70000		33
A Silent Stride...	1931	Comedy	87	USA	English	Frank Capra	Harold Lloyd, ...	A silent comedy...	6.00000		13

Data Source Sheet 1

### Conclusion (Students should write in their own words):

From the above lab experiment we learnt how to import files of different file formats on to our tableau dashboard Used PostgreSQL as a live server and connected it to Tableau and viewed datasets.

Date:

Signature of faculty in-charge

## Post Lab Question:

### 1. List down types of data Tableau can import?

Tableau can handle a variety of file types.

- Excel. xlsx
- Text File .txt
- PDF .pdf
- CSV .csv,
- JSON .json
- Spatial File
- Statistical File
- Tableau workbook .twb
- Zip file .zip
- Local Cube files .cub

It also has many pre-installed connectors for different servers

Installed Connectors (76)			Additional Connectors (14) ⓘ
Action Matrix	Google Analytics	Oracle Eloqua	Action ODBC by Actian
Action Vector	Google BigQuery	Oracle Essbase	Elasticsearch by Elastic
Alibaba AnalyticDB for MySQL	Google Cloud SQL	Pivotal Greenplum Database	Incorta Connector by Incorta
Alibaba Data Lake Analytics	Google Drive	PostgreSQL	Kyligence Connector by Kyligence
Alibaba MaxCompute	Google Sheets	Presto	MarkLogic by MarkLogic
Amazon Athena	Hortonworks Hadoop Hive	Progress OpenEdge	Ocient JDBC by Ocient
Amazon Aurora for MySQL	IBM BigInsights	Qubole Presto	Oracle NetSuite by Tableau
Amazon EMR Hadoop Hive	IBM DB2	Salesforce	Qubole Hive by Qubole
Amazon Redshift	IBM PDA (Netezza)	SAP HANA	Salesforce CDP by Salesforce
Anaplan	Impala	SAP NetWeaver Business Warehouse	SAP SuccessFactors by Tableau
Apache Drill	Intuit QuickBooks Online	SAP Sybase ASE	SQream DB by SQream Technologies
Aster Database	Kognitio	SAP Sybase IQ	Starburst Enterprise by Starburst
Azure Data Lake Storage Gen2	Kyvos	ServiceNow ITSM	Stratio Crossdata by Stratio BD
Azure SQL Database	LinkedIn Sales Navigator	SharePoint Lists	Yellowbrick by Yellowbrick Data
Azure Synapse Analytics	MapR Hadoop Hive	SingleStore	
Box	MariaDB	Snowflake	
Cloudera Hadoop	Marketo	Spark SQL	
Databricks	MarkLogic	Splunk	
Datorama	Microsoft Analysis Services	Teradata	
Denodo	Microsoft PowerPivot	Teradata OLAP Connector	
Dremio	Microsoft SQL Server	TIBCO Data Virtualization	
Dropbox	MonetDB	Vertica	
Esri ArcGIS Server	MongoDB BI Connector	Web Data Connector	
Exasol	MySQL		
Firebird 3	OData	Other Databases (JDBC)	
Google Ads	OneDrive	Other Databases (ODBC)	
	Oracle		

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

Using measurements as our primary example: In Tableau there are four possible modes when dealing with your measurement fields which result in a different layout or result. They are a) continuous aggregate measure, b) discrete aggregate measure, c) continuous disaggregate measure, d) discrete disaggregate measure. B and D are considered dimensions by Tableau.

- 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).

Tableau represents data differently in the view depending on whether the field is discrete (blue), or continuous (green). Continuous and discrete are mathematical terms. Continuous means "forming an unbroken whole, without interruption"; discrete means "individually separate and distinct."