**Batch: C3 Roll No.: 16010120193**

**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

# Theory (About Data Preprocessing):

# Steps Involved in Data Pre-processing:

# 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.

1. **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.

1. **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.

1. **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 segmented is handled separately. One can replace all data in a segment by its mean or boundary values can be used to complete the task.

1. **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).

1. **Clustering:**

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

# 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)

# 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

# Graphical user interface, application, table Description automatically generated

# Approach 2:

# Connecting a live server to Tableau

# Server used: PostgreSQL

# Platform: Tableau

# Step 1: Getting the server started

# Text Description automatically generated

# Step 2:

# On Tableau, establish remote access to the live server

# Graphical user interface, text, application, email Description automatically generated

# Note: SSL certificate not supported

# Step 3: Load the CSV on the platform

# Graphical user interface, application, table Description automatically generated

# 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: 12-10-2021 Signature of faculty in-charge**

# Post Lab Question:

# List down types of data Tableau can import?

# Tableau can handle a variety of file types.

# Excel. xslx

# 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

# Graphical user interface Description automatically generated with low confidence

# 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."