B.E / B.Tech. PRACTICAL END SEMESTER EXAMINATIONS, NOVEMBER/DECEMBER 2022

Third Semester

AD3301 – DATA EXPLORATION AND VISUALIZATION

(Regulations 2021)

Time : 3 Hours   Answer any one Question   Max. Marks 100

Aim and Algorithm	Program	Output and Results	Viva-Voce	Record	Total
20	30	30	10	10	100

1.   A) Install the Data Analysis and Visualization Tool Python.

     B) Install Pandas Package in Python and execute the Program for simple Data frame attributes.

2.   Create a Program using Numpy package functions and 2D or 3D array to perform simple matrix operation.

3.   To combine Numpy and Pandas data frame to create dataset and perform the following

     A) Color Variation each column data

     B) To Display NAN values

     C) Highlight Max and Min values.

     D) To generate Background gradient color variation

4.   Explore multivariable dataset, To perform any four data cleaning method and visualize Bar chart.

5.   Explore using seaborn to load the dataset three variable(Username, Tweet, Location) tweets comment review for #tag Jallikattu Protest.

     A) Perform scatterplot using different location tweet

     B) Perform Bubble chart for # and @ tag.

     Ex: #Tamilnadu, #Jallikattu, etc..,

         @SaveTNfarmers, @Marina, etc..,

6. Create a Pie chart for Student Result Analysis by using pie plot in python. Plot segregation will be Distinction (Greater than or Equal to 8.5 CGPA) and First Class (Greater than 6.5 CGPA).

7. Create a Lollipop chart for Festival Shopping dataset of your own(20 rows and 5 to 10 columns)

8. To perform the following data transformation techniques of your own dataset. (20 Rows and 5 Columns)

    A) Removing Null Values (NaN)

    B) Drop Columns

    C) Merging database style dataframes

9. To perform dataframe merge function (inner, left and outer join) using simple dataset.

10. Explore simple dataset and perform Transformation techniques such as data deduplication, Replace values, Handling missing Data, Backward and Forward filling.

11. To perform hypothesis testing using stats library of your own dataset Explore T test.

12. Explore and visualize of your own dataset/ data frame and perform numerical summaries and spread level.

    A) Floating values into two columns from single variable.

    B) Perform Descriptive Analysis

    C) Perform Percentage Table both row and column.

13. Perform Time Series Analysis and apply various visualization methods for Internet Traffic Time Dataset. (Create own data with minimum 5 columns and 20 rows)

14. Perform EDA for Water quality dataset.

    All attributes are numeric variables and they are listed bellow:

    aluminium - dangerous if greater than 2.8

    ammonia - dangerous if greater than 32.5

    arsenic - dangerous if greater than 0.01

    barium - dangerous if greater than 2

    cadmium - dangerous if greater than 0.005

15. Perform EDA on map using various map dataset to find the nearest Sports Shop from your Location with mouse rollover effect.

16. Build a Cartographic visualization for two datasets involving states and district in India.

17. Perform EDA for Price of petroleum products in India from the year 2013 to 2023.

    (Create dataset with minimum 5 columns and 20 rows.)

18. Explore and visualize women empowerment in India 2025 and compare every five year from 2010. (Create dataset with minimum 5 columns and 20 rows.)

19. Perform EDA and Visualization for COVID-19 dataset.

    A) State wise Bar chart

    B) Recovered from COVID-19 District wise Bar chart

    C) Descriptive analysis for different age group.

    (Create dataset with minimum 5 columns and 20 rows.)

20. Perform EDA for Ticket Booking(Bus/Train/Flights)