**B.Sc. Mathematical Sciences / III Semester**

**LAB WORK ASSESSMENT**

**DATA ANALYSIS USING PYTHON (SEC)**

**Deen Dayal Upadhyaya College, University of Delhi**

Date: **30th November 2020** Time: **10am-12noon INSTRUCTIONS**

∙ The Lab-Work assessment consists of 3 sets to be solved through Python Programming. The set number against each student is given in the table at the end of this document. Every student must submit the answers for his own set only.

∙ Total Marks allocated for Lab Work assessment is 50 and the duration of today’s task is 2 hours.

∙ Students are required to use Google Colab Notebook for Python Programming. ∙ The **submission must be done via Google Forms on Google Classroom** regarding this assessment on 30th November 2020 strictly by 12 Noon. Separate emails or WhatsApp files will not be considered for evaluations.

∙ Clearly state all the assumptions and test cases used in the code logic explanation. ∙ For any confusion, clarification or concern, contact Dr. Rajan Gupta - 9818236739 ∙ The submission form must contain the following information.

(a) Full Name:

(b) Class Roll Number:

(c) Name of Course: B.Sc. Mathematical Sciences

(d) Semester: 3rd Semester (2nd Year)

(e) Name of Subject: Data Analysis Using Python (SEC)

(f) Assessment: B.Sc. Mathematical Sciences, 3rd Semester Lab-Work Assessment (g) Assessment Date: 30th November 2020

(h) Assessment Time: 10am – 12noon

(i) Compiler Used:

(j) Email ID used for assessment submission:

(k) Submitted to: Dr. Rajan Gupta

(l) College/University: Deen Dayal Upadhyaya College, University of Delhi -----x-----

1

**SET-1**

1) Import Iris dataset.

Using appropriate data analysis commands, find the mean, median, standard deviation, quartiles, range, correlation coefficients, covariance, scatter plot, and box plot for the dataset. 2) Create a series of size 20 with random elements in any range and perform the following operations

(a) Change the index from 'a' to 't'

(b) Assign the values of 4th and 5th element of the series at 'm' and 'n' index position (c) Assign 10 value to the last 3 elements of the series

(d) Display the series values from index 'e' to 'r'

(e) Using loc function, display first 10 elements of series

(f) Using iloc function, display last 10 elements of series

(g) Sort the values of the series

3) Mention different techniques to fill the missing data in continuous data and categorical data, with example.

**SET-2**

1) Create a 1-D array of the following shoe brand names in the given particular order - (Nike, Adidas, Reebok, Puma, Adidas, Nike, Reebok, Reebok). Now create a random array of 8 rows and 4 columns with only positive numbers and multiply each element by 10,000. This array determines the price of the shoes of various brands having four different variants.

Display the price of all Adidas Shoes, Reebok Shoes, PUMA Shoes and Nike Shoes, individually. Then display the Price of only 2nd and 3rd variant of all Nike or Adidas shoes, and the price of 1st and 2nd variant of shoes other than Nike or Puma Brand. Also, display the price of all the shoes which are more than 1500 and then change the price of all the Reebok shoes to 5000. Then print only 1st and 2nd variant prices of first four brands of the shoes.

2) Create a dataframe of 10 students for their complete graduation (3 years) with 3 columns - Name, Year, Percentage

e.g. (Ram, 2017, 75), (Ram, 2018, 77), (Ram, 2019, 80), (Reeta, 2017, 83), (Reeta, 2018, 89), (Reeta, 2019, 85), etc. (This will make 30 entries in the dataframe for the 10 students across 3 years [2017-2018-2019])

Perform the following operations.

(a) Display the first 8 entries of the dataframe

(b) Display the last 5 entries of the dataframe

(c) Change the index values of the dataframe with numbers written as text. e.g. 1 replaced as One, 2 replaced as Two, etc.

(d) Add a grade column with no values

2

(e) Display the student details only for the year 2018

(f) Display the student details having more than 75% marks in any year

(g) Find the transpose of the dataframe

3) Explain the GROUPBY feature with 3 examples.

**SET-3**

1) Create two dataframes - First having 4 rows and 4 columns and Second with 3 rows and 4 columns, with random data values of your choice, and perform following operations (a) Perform additions & differences of the two dataframes

(b) Change the diagonal values in the first dataframe to NaN and then perform addition and substraction of the two dataframes. While adding the two dataframes, fill 100 in place of Nan values.

(c) Write a function to print difference between highest and lowest values of each column of both the dataframes

(d) Write a function to print difference between highest and lowest values of each row of both the dataframes

(e) Sort the index of both the dataframes columnwise and row-wise

(f) Sort the values of both the dataframes column-wise and row-wise

(g) Find the rank of elements in 2nd column of both the dataframes. Check whether a number entered by the user through standard input is a prime number or not.

2) Explain the following concepts under Data Transformation with working examples – Dummy Variable, Outlier detection, Discretization, Binning, Replacing Functions, Removing duplicates

3) Explain the concept of saving files in Numpy in different file formats.

3

**Set Information for Various Students**

|  |  |  |  |
| --- | --- | --- | --- |
| **S.No** | **Roll No** | **Student Name** | **SET** |
| 1 | 19MTS5602 | AADESH | **3** |
| 2 | 19MTS5604 | ANUPAM | **2** |
| 3 | 19MTS5605 | ANVI | **1** |
| 4 | 19MTS5606 | ARSHPREET | **3** |
| 5 | 19MTS5608 | CHAHAT | **2** |
| 6 | 19MTS5609 | DHEER JAIN | **1** |
| 7 | 19MTS5611 | JYOTI YADAV | **3** |
| 8 | 19MTS5613 | MANISHA | **2** |
| 9 | 19MTS5614 | MAYANK BHURIA | **1** |
| 10 | 19MTS5615 | MD ABID | **3** |
| 11 | 19MTS5616 | MOHIT KUMAR | **2** |
| 12 | 19MTS5617 | NEETU YADAV | **1** |
| 13 | 19MTS5618 | NIKHIL TEOTIA | **3** |
| 14 | 19MTS5620 | RAGHAV BAJAJ | **2** |
| 15 | 19MTS5621 | RAHUL | **1** |
| 16 | 19MTS5622 | RIDHI GOEL | **3** |
| 17 | 19MTS5623 | SACHIN PATEL | **2** |
| 18 | 19MTS5625 | SANVI | **1** |
| 19 | 19MTS5626 | SHASHWAT | **3** |
| 20 | 19MTS5627 | SIDDHANT JAIN | **2** |
| 21 | 19MTS5628 | SURENDRA BHINCHER | **1** |
| 22 | 19MTS5629 | UTKARSH TRIPATHI | **3** |
| 23 | 19MTS5630 | ANKIT | **2** |
| 24 | 19MTS5632 | ANUPRIYA | **1** |
| 25 | 19MTS5633 | DEEPA YADAV | **3** |
| 26 | 19MTS5634 | HIMANSHU YADAV | **2** |
| 27 | 19MTS5635 | ROSHAN LAL MEENA | **1** |

4