

Name:

Enrolment No:



UNIVERSITY OF PETROLEUM AND ENERGY STUDIES

End Semester Examination, July 2020

Course: Data Visualization for Analytics

Semester: 4

CourseCode: CSBA2007

Time: 2hr

Programme: B Tech CSE+BAO

Max. Marks: 100

Instructions: All Questions are compulsory.

Q.1 EDA: (Unit 5)

Total Marks: 14

Write python code for EDA operations. Write outputs if possible (each carries 2 marks)

1. Importing all necessary libraries
2. Reading CSV file
3. Showing first 5 rows from the give data set
4. Getting information about each column
5. Dropping the duplicates by keeping the very first data
6. Getting the shape of given data-set
7. Getting count, mean, SD, min, max, 1st, 2nd, 3rd quantile

Q.2 Seaborn: (unit 5)

Total Marks: 16

1. Importing necessary library
2. Getting correlation table
3. Getting heatmap visualization
4. Getting uni-variable analysis by visualization by "joinplotting" and "Dist plot"
5. Getting multivariable analysis visualization by "pair plotting"
6. Getting bi-variable analysis visualization by "bar plot"
7. Getting bi-variable analysis visualization by "box plot"
8. Getting bi-variable analysis visualization by "violin plot"

Q.3 Represent the following data using box-plot visualization technique: **Unit 2 (10 marks)**

{5, 2, 10, 6, 2, 8, 8, 1, 2, 5, 7, 3, 3, 4}, Calculate Q1, Q2 and Q3

Q.4 A hospital switch board receives an average of 4 emergency calls in 10 minutes interval, what is probability that 1) there are at most 2 emergence calls.

2) which distribution will be applicable for above problem?

Unit 4 (20 marks)

Q.5 Consider an exam that contains 10 MCQ with 4 possible choice for each questions and only one of them is correct. What is the probability to get no answer correct? which distribution will be applicable for above problem? **Unit 4 (20 marks)**

Q.6 Covariance are used for which type of analysis? Identify the limitation of covariance and how it is overcome by Pearson Co-relation co-efficient. **Unit 3 (10 marks)**

Q.7 Z-score values are used for finding the proportion data distributed under Standard Normal Distribution; justify that statement. **Unit 3 (5 marks)**

Q.8 We have recorded the heights of 10 students (in cm): {168, 170, 150, 160, 182, 140, 175, 180, 170, 190}. Find the 1st, 2nd and 3rd quartile for the give data set. **Unit 2 (5 marks)**