

**PARUL UNIVERSITY**  
**FACULTY OF ENGINEERING & TECHNOLOGY**  
**B.Tech. Summer 2021 - 22 Examination**

**Semester: 8<sup>th</sup>**  
**Subject Code: 203105483**  
**Subject Name: Data Science**

**Date: 30/03/2022**  
**Time: 10:30 AM to 1:00 PM**  
**Total Marks: 60**

**Instructions:**

1. All questions are compulsory.
2. Figures to the right indicate full marks.
3. Make suitable assumptions wherever necessary.
4. Start new question on new page.

**Q.1 Do as Directed for given sub questions (All are compulsory) (Each of one mark) (15)**

**Write Correct Answer option for Q-1's sub questions 1 to 5.**

1. Which of the following is one of the key data science skills?  
 a) Statistics    b) Machine Learning    c) Data Visualization    d) All of the mentioned
2. Point out the wrong statement.  
 a) Merging concerns combining datasets on the same observations to produce a result with more variables  
 b) Data visualization is the organization of information according to preset specifications  
 c) Sub-setting can be used to select and exclude variables and observations  
 d) All of the mentioned
3. Which of the following is not an application for data science?  
 a) Recommendation Systems    b) Image & Speech Recognition  
 c) Online Price Comparison    d) Privacy Checker
4. Data can be categorized into \_\_\_\_\_ groups.  
 a) Structured data and Unstructured data    b) Categorical data and Non-categorical data  
 c) Summarized data and Scattered data    d) Labeled data and Unlabeled data
5. The kernels used in Support Vector Machine are \_\_\_\_\_.  
 a) Gaussian Kernel    b) Polynomial Kernel    c) Sigmoid Kernel    d) All of the above

**Fill in the Blanks of Q-1's sub questions 6 to 10.**

6. The square root of the variance is called the \_\_\_\_\_ deviation.
7. Machine learning algorithms are built a model based on sample data, which is known as \_\_\_\_\_ data.
8. \_\_\_\_\_ is a multidisciplinary which involves extraction of knowledge from large volumes of data that are structured or unstructured.
9. ETL refers to Extract, \_\_\_\_\_, & Load.
10. Bokeh library is used for \_\_\_\_\_ type of data visualization.

**Answer the Following of Q-1's sub questions 11 to15 in one word/one line.**

11. Name the data which is recorded as the rank order of scores?
12. What is "variance" in data analytics?
13. Which type of distribution describes the amount of time between events occurring at random moments?
14. Name the Sources of Primary Data?
15. Explain different of Data Visualization.

**Q.2 Answer the following questions. (Attempt any three) (15)**

- A) What methods are used to collect data?
- B) What is API? Explain APIs with applications with their use in data science.
- C) Explain Linear Regression with its applications?
- D) Explain use of Agile framework in Data Science

**Q.3 A) What is Data Visualization? Explain it with scenarios and show how does it help in data science? (07)**

**B) Explain Bernoulli distribution along with Example in details? (08)**

**OR**

**B) What is Naïve Bayes algorithms? Explain with its advantages and disadvantages. (08)**

**Q.4 A) What are the python libraries used for data science? Explain with their purposes. (07)**

**OR**

**A) Demonstrate the use of "sklearn" or "scikit" library to import iris dataset & using the bokeh library also show steps to generate a chart to visualize the dataset. (07)**

**B) Explain the various types of Data used in Data Science and also show their applications in recent trends of Data Science? (08)**