# GUJARAT TECHNOLOGICAL UNIVERSITY
## BE - SEMESTER–V (NEW) EXAMINATION – SUMMER 2021

**Subject Code:3150713**                                      **Date:07/09/2021**
**Subject Name:Python for Data Science**
**Time:10:30 AM TO 01:00 PM**                        **Total Marks: 70**
**Instructions:**
1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**
4. **Simple and non-programmable scientific calculators are allowed.**

|       |     |                                                                                                                                        | MARKS |
|-------|-----|----------------------------------------------------------------------------------------------------------------------------------------|-------|
| **Q.1** | **(a)** | Differentiate the list and dictionary data types of python by their characteristics along with example in brief.                      | **03** |
|       | **(b)** | What do you mean by slicing operation in string of python? Write an example of slicing to fetch first name and last name from full name of person and display it. | **04** |
|       | **(c)** | Which are the basic activities we performed as a part of data science pipeline? Summarize and explain in brief.                       | **07** |
| **Q.2** | **(a)** | What is the core competencies needed to become a data scientist? Explain in brief.                                                    | **03** |
|       | **(b)** | Compare and summarize four different coding styles supported by Python language.                                                      | **04** |
|       | **(c)** | Summarize the characteristics of NumPy, Pandas, Scikit-Learn and matplotlib libraries along with their usage in brief.                | **07** |
|       |     | **OR**                                                                                                                                 |       |
|       | **(c)** | What do you mean by prototyping? List the phases of prototyping and experimentation process and explain in brief.                     | **07** |
| **Q.3** | **(a)** | Compare the numpy and pandas on the basis of their characteristics and usage.                                                         | **03** |
|       | **(b)** | For what purpose sampling is used. Demonstrate random sampling with example.                                                          | **04** |
|       | **(c)** | What is the need of streaming the data? Explain data uploading and streaming data with example.                                       | **07** |
|       |     | **OR**                                                                                                                                 |       |
| **Q.3** | **(a)** | How XPath is useful for analysis of html data? Explain in brief.                                                                      | **03** |
|       | **(b)** | Define term n-gram. Explain the TF-IDF techniques.                                                                                    | **04** |
|       | **(c)** | List the techniques to handle missing data. Explain various techniques with example.                                                 | **07** |
| **Q.4** | **(a)** | List various types of graph/chart available in the pyplot of matplotlib library for data visualization. Explain any two of them in brief. | **03** |
|       | **(b)** | What kind data is analyzed with Bag of word model? Explain it with example.                                                           | **04** |
|       | **(c)** | What do you mean by time series data? How can we plot it? Explain it with example to plot trend over time                             | **07** |

**OR**

**Q.4** **(a)** Compare bar graph, box-plot and histogram with respect to their applicability in data visualization. **03**

**(b)** Define stemming. Explain the concept of stemming with example. **04**

**(c)** What is the use of scatter-plot in data visualization? Can we draw trendline in scatter-plot? Explain it with example. **07**

**Q.5** **(a)** Define the term Data wrangling. Explain the steps needed to perform data wrangling. **03**

**(b)** Why we need to perform Z-score standardization in EDA? Justify it with example. **04**

**(c)** What is the use of hash function in EDA? Express various hashing trick along with example. **07**

**OR**

**Q.5** **(a)** What do you mean by Exploratory Data Analysis (EDA)? How t-test is useful for EDA? **03**

**(b)** What do you mean by covariance? What is the importance of covariance in data analysis? Explain it with example. **04**

**(c)** List different way for defining descriptive statistics for Numeric Data. Explain them in brief. **07**

**\*\*\*\*\*\*\*\*\*\*\*\***