# Guru Tegh Bahadur Institute of Technology, New Delhi

# Foundations of Data Science (Question Bank)

# Course Name: B.Tech (AIML) Semester: III SUB CODE: AIML203

**Unit-1**

Q1. Why do we need Python for Data Science? Explain its importance depicting various features of python useful for data science.

Q2. Explain the term “Domain Knowledge” in context of Data science and python.

Q3. Differentiate between the following:

1. Structured and Unstructured data
2. Data Scientist, Data Analyst and Data Engineer

Q4. Explain the importance of data preprocessing and data cleaning in data science. What are the different ways of handling missing values in python?

Q5. Write a code in python for the following:

1. Find the sum of series upto n terms (where the value of n is taken from the user)
2. Display all prime numbers in a given range
3. Create a calculator in python
4. Count the number of digits in a number

Q6. Describe synthetic dataset and how it is different from natural datasets? Justify with an instance. How do you create a synthetic dataset in python? Which packages and libraries are used for the same?

Q7. Interpret the life cycle of data science with the help of a diagram and a case study highlining the main procedural steps. Differentiate between data cleansing and data munging.

Q8. Justify the statement “Information is a processed version of Data”.

Q9. What is data science and how does it relate to other disciplines like statistics and machine learning?

Q10. Differentiate between Data mining and Data Science.

**Unit-2**

Q1. Summarize the Data leaning life cycle with the help of a real-life case study.

Q2. Explain imputation. Demonstrate any two imputation techniques with the help of an example.

Q3. What is a delimiter? Highlight the steps taken to import the data (CSV) from a given folder.

Q4. Differentiate between bitwise left shift and bitwise right shift operator with the help of an example for a negative number.

Q5. How does data visualization enhance the initial exploration of a dataset? Can you provide a specific example illustrating its significance in uncovering insights from the data?

Q6. Compare and contrast the efficiency of calculating factorial of number in Python using both recursive and iterative approaches.

Q7. Analyze the importance of handling and identifying missing values in the field of data science. Discuss the strategies and methods for managing missing data effectively and their significance in ensuring data quality and meaningful analysis.

Q8. Differentiate between list and dictionary data types of python by their characteristics with examples kin brief.

**Unit-3**

Q1. Describe in detail about Exploratory Data Analysis?

Q2. Explain essential python libraries required for data processing and data visualization in python. Illustrate the usage with the help of a case study.

Q3. List various types of graphs/charts available in the Pyplot of matplotlib library for data visualization. Explain any two of them in brief.

Q4. What are the advantages of using Seaborn over Matplotlib for certain types of visualizations?

Q5. What is a heat map, and how can you generate one using Seaborn?

Q6. Explain the difference between plt.show() and plt.savefig() in Matplotlib.

**Unit-4**

Q1. Explain the differences between supervised and unsupervised machine learning algorithms.

Q2. Explain recommender system. State some of its applications. Illustrate recommender system with the help of real-life case study.

Q3. What are the various mathematical and scientific applications for data analysis. Perform data analysis on any one of the listed applications to get a useful insight of data.

Q4. How can you visualize trends in a dataset using Python? Provide examples of relevant Python libraries for the same. Illustrate the concept with a real-life case study.

Q5. What do you mean by predictive analysis? Explain various steps involved in predictive analysis. What are the applications of predictive analysis? Illustrate predictive analysis for health care application