SHRI GNANAMBICA DEGREE COLLEGE, MADANAPALLE
(AUTONOMOUS)
PROGRAMME: B. Sc (DATA SCIENCE)
II YEAR-IV-SEM
COURSE: DATA VISIUALIZATION USING PYTHON
QUESTION BANK

UNIT I: Introduction and Essential Python Libraries
Short Answer Questions

1. Define **Data Science** and explain its components.
2. What is Exploratory Data Analysis (EDA)? Explain its purpose.
3. What are the advantages of using Python for Data Science?
4. Explain the features and advantages of Jupyter Notebook.
5. Write short notes on NumPy and pandas library and its applications.
6. Write short notes on matplotlib and seaborn libraries.
7. Explain the role of scikit-learn and SciPy libraries.

Long Answer Questions

1. Define Data Science and explain the Data Science Process in detail with diagram.
2. Explain Exploratory Data Analysis (EDA) techniques and its importance in Data Science.
3. Explain why Python is preferred for Data Science. Describe the role of Jupyter Notebook.
4. Explain the following libraries: NumPy, pandas, matplotlib, seaborn.
5. Explain the role and applications of SciPy, scikit-learn, and statsmodels in Data Science.

UNIT II: Pandas, Data Loading, Cleaning and Preparation
Short Answer Questions

1. Explain vectorized computation and its advantages.
2. Describe the Series data structure in pandas.
3. Describe the DataFrame data structure in pandas.
4. Explain the essential functionality of pandas.
5. Explain methods of summarizing descriptive statistics.
6. Explain techniques for handling missing data.
7. Explain string manipulation functions in pandas.

Long Answer Questions

1. Explain pandas data structures (Series and DataFrame) and their operations with examples.
2. Explain different methods of data loading, storage, and file formats in pandas.
3. Explain data cleaning techniques including handling missing data and transformation.
4. Explain methods for reading and writing data in text and binary formats.
5. Explain how pandas is used for web scraping, interacting with Web APIs, and databases.

UNIT III: Data Wrangling and Visualization

Short Answer Questions

1. Define Data Wrangling and explain its importance.
2. Explain Hierarchical Indexing in pandas.
3. Explain the concept of data reshaping.
4. What is pivoting? Explain with example.
5. Explain basic features of matplotlib.
6. Explain the role of seaborn library in visualization.
7. Explain different types of categorical plots.

Long Answer Questions

1. Explain Data Wrangling and discuss hierarchical indexing, merging, and combining datasets.
2. Explain reshaping and pivoting techniques in pandas with examples.
3. Explain matplotlib library and its plotting functions in detail.
4. Explain how pandas and seaborn are used for data visualization.
5. Explain advanced categorical and numeric visualization techniques.

UNIT IV: Data Aggregation and Time Series Analysis

Short Answer Questions

1. Explain the concept of GroupBy operation.
2. Explain data aggregation techniques.
3. Define split-apply-combine strategy.
4. Explain date and time data types in pandas.
5. Explain time series basics.
6. Explain date ranges and frequency.
7. Explain time zone handling.

Long Mark Questions

1. Explain GroupBy operations and data aggregation techniques with examples.
2. Explain split-apply-combine strategy and its applications.
3. Explain pivot tables and cross tabulation in pandas.
4. Explain time series data analysis including date ranges, shifting, and frequency conversion.
5. Explain moving window functions and time zone handling with examples.

UNIT V: Advanced Pandas

Short Answer Questions

1. Define categorical data in pandas. Explain advantages of categorical data.
2. Explain methods for cleaning categorical data.
3. Explain visualization techniques for categorical data.
4. Explain advanced GroupBy methods.
5. Define method chaining. Explain advantages of method chaining.
6. Explain techniques used in advanced data cleaning.
7. Explain the importance of data visualization after cleaning.

Long Answer Questions

1. Explain categorical data handling and its advantages in pandas.
2. Explain data cleaning and visualization techniques for categorical data.
3. Explain advanced GroupBy methods and their applications.
4. Explain method chaining techniques with examples.
5. Explain advanced pandas operations for data analysis and visualization.