Question Bank : Unit IV & V

Getting Started with Pandas

1. What are the two primary data structures in pandas?
   Illustrate with examples how to create a Series and a DataFrame.
2. Explain how indexing and slicing works in a pandas DataFrame.
   Write a short program to demonstrate reindexing and label-based selection using .loc and .iloc.
3. How do you compute basic descriptive statistics (mean, count, std) in pandas?
   Create a DataFrame with missing values and use .describe() and .info() to summarize it.
4. Differentiate between NaN and None in pandas.
   Demonstrate the use of .isna(), .fillna(), and .dropna() with appropriate examples.

NumPy Basics: Arrays and Vectorized Computation

5. How do you create NumPy arrays using np.array, np.arange, and np.zeros?
   Show how to reshape and slice a 2D NumPy array.
6. What are universal functions in NumPy?
   Use np.sqrt(), np.exp(), and np.maximum() in examples comparing two arrays.
7. What is broadcasting in NumPy? Provide a use-case example.
   Use vectorized operations to compute the row-wise mean and column-wise standard deviation of a 2D array.

Data Wrangling with pandas

8. Compare pd.concat(), merge(), and join() with code examples.
   What are the types of joins in pandas and how do they work?
9. Explain the use of stack(), unstack(), and melt() in reshaping data.
   Write code to pivot a DataFrame and explain the difference between pivot() and pivot_table().

Data Visualization using matplotlib and pandas

10. Write a program to draw a line graph using matplotlib.pyplot.
    Customize the plot with labels, title, and legend.
11. How can you create plots directly from a pandas Series or DataFrame?
    Generate and customize the following using pandas' plotting methods:
    - Line Plot
    - Bar Chart
    - Histogram
    - Box Plot

Additional Questions Chapter wise:

Getting Started with Pandas

1. What are the key data structures in pandas? Explain their differences with examples.

2. Write a program to create a pandas Series and DataFrame. Explain each step.

3. How is data indexed in a DataFrame? Differentiate between label-based and position-based indexing.

4. Demonstrate the use of loc[] and iloc[] in pandas with examples.

5. What are some key methods for summarizing data in pandas?

6. Write code to compute mean, median, mode, and standard deviation of numeric columns in a DataFrame.

7. What is missing data? How is it represented in pandas?

8. Show how to detect missing data in a DataFrame. Use isnull(), notnull().

9. Demonstrate fillna() and dropna() functions with different parameters.

10. How can you replace missing values conditionally in pandas? Write an example.

NumPy Basics – Arrays and Vectorized Computation

1. How do you create a 1D and 2D array using NumPy? Give examples using array(), arange(), linspace(), zeros().

2. What is the importance of the ndarray object in NumPy?

3. Explain slicing and indexing in NumPy arrays. Illustrate with examples.

4. What are universal functions (ufuncs) in NumPy? Explain with at least three examples.

5. Show how to use np.maximum(), np.sqrt(), and np.exp() for element-wise operations.

6. Explain broadcasting with examples. When is it useful?

7. Write a program to calculate the row-wise and column-wise mean of a 2D array.

8. How do you perform conditional logic using NumPy arrays? Use np.where() in an example.

9. Illustrate any five aggregate functions in NumPy: sum, mean, std, var, min.

10. How is vectorized computation better than loops in Python? Demonstrate with timing comparison.

Data Wrangling – Combining, Merging, Reshaping

1. Differentiate between concat(), merge(), and join() in pandas.

2. Write a program to combine multiple DataFrames using concat().

3. How do you merge two DataFrames based on a key column? Show inner, outer, left, right joins.

4. What are the use cases of merge() with multiple keys? Illustrate with an example.

5. Explain reshaping using stack(), unstack(), melt(), and pivot().

6. How is pivot_table() more powerful than pivot()? Give a real-life use case.

7. Write a code snippet to create a tidy DataFrame using melt().

8. Demonstrate reshaping data using wide_to_long() method.

9. What is hierarchical indexing? How is it used in reshaping?

10. Illustrate reshaping of time series data into wide and long formats using real or dummy data.

Data Visualization – matplotlib and pandas

1. Write a program to create a simple line plot using matplotlib.

2. How can you customize a plot with title, axis labels, legend, and grid?

3. What are the different plot types supported by pandas? Write code to generate a bar plot and histogram.

4. Create a box plot and explain how it helps in visualizing data distribution.

5. Demonstrate a scatter plot between two variables using pandas and matplotlib.

6. How can you plot multiple subplots in one figure? Write example code.

7. Compare pyplot vs pandas plotting interface with examples.

8. Write code to generate a pie chart using matplotlib. Show labels and percentages.

9. How do you change the style, color, and marker of lines in matplotlib?

10. Show how to plot a DataFrame with a datetime index. Format the x-axis with rotation.