**Mid-II**

**STATISTICAL ANALYSIS AND COMPUTING**

**SAQ's**

1. How is a DataFrame different from a Series in pandas?
2. Write the difference between loc [ ] and iloc[ ] in pandas with examples
3. Differentiate Correlation and Covariance
4. How do you read a text file in Python?
5. What is the significance of crosstabulations (crosstab) in data analysis?
6. How do you iterate over groups in pandas?
7. Write a program in Python to summarize and compute descriptive statistics for a pandas Series or data frame.
8. Describe the role of regular expressions in data cleaning and string pattern matching using pandas.
9. How do you remove duplicates from a dataset in pandas?
10. What is discretization and binning in the context of data preparation?
11. What is the purpose of group operations in data analysis?
12. How do you group data using dictionaries and Series in pandas?
13. What are the different types in the datetime module
14. How to create time series object
15. How Generate Date Ranges using pandas
16. What is Shifting in the context of Time series
17. Differentiate Downsampling and Upsampling
18. How do you Transforming Data Using a Function or Mapping
19. How do you compute Indicator/Dummy Variables
20. List Some pandas.read\_csv function arguments

**LAQ’s**

1. Write a program in Python to summarize and compute descriptive statistics for a pandas Series or data frame.
2. Discuss the various ways of indexing, selection, and filtering in pandas.
3. Explain Regular expression methods with examples
4. Discuss the techniques for detecting and filtering outliers in a dataset.
5. Explain Optimized groupby methods with examples
6. Explain the concept of pivot tables and cross-tabulation in pandas.
7. Explain the concept of reindexing in pandas with an example.
8. Create a Python program to write data to a text file and then read it back to display the contents.
9. Suppose you have an SQLite database containing a large table of sales data with columns for date, product, quantity, and price. Design a panda’s code to read this data into a data frame, filter it for a specific date range, calculate the total revenue for each product, and then sort the products based on their total revenue.
10. Write a Python program to create a panda DataFrame with the following columns Name, age and gender

Implement the following tasks:

* 1. Identify the number of missing values in each column.
  2. Filter out any rows that have missing values in the Age column.
  3. Fill in the missing values in the Age column with the median of the values in the column.
  4. Print the resulting DataFrame.

1. Explain the concept of reindexing in pandas with an example.
2. Discuss the methods for handling missing data in pandas, including filtering out and filling in missing values.
3. Discuss Python Built-In String Object Methods
4. Explain the concept of split-apply-combine in pandas with an example.
5. Discuss the use of data aggregation in pandas with examples.
6. Write a Python program to read a text file in pieces and display its contents.
7. Describe the concept of discretization and binning in data preparation. Provide an example.
8. Suppose you have an SQLite database containing a large table of sales data with columns for date, product, quantity, and price. Design a panda’s code to read this data into a data frame, filter it for a specific date range, calculate the total revenue for each product, and then sort the products based on their total revenue.
9. Discuss the use of correlation and covariance in analyzing data with pandas
10. Write a python program to read data from an Excel file and display its contents and perform some operations on it.
11. Discuss Base time series frequencies with examples
12. Explain Time Zone Handling in
13. Discuss Periods and Period Arithmetic