Data Wrangling Question Bank

Unit 1

1. Define Data Science and explain the applications of data science in various sectors.

2. Define Data Wrangling. Explain it's importance and various steps involved in it.

3. Write about the essential python libraries.

4. Write about the built in data structures used in python. Explain the following:

      a) List Comprehension b) Set Comprehension c) Dictionary Comprehension

5. a)Write about Functions. Explain about scope and local functions.

      b) Write about anonymous lambda functions. Explain with 2 examples.

Unit 2

1. What is Numpy. What are the main areas of functionalities used for data analysis applications.

2. a) Write about numpy ndarray.

      b) What are the arithmetic operations performed using numpy arrays. Explain with examples.

3. Write about different array creation methods in numpy. Explain with examples.

4. a) Explain about the basic data structures used in pandas.

      b) Write the syntax for creating a Series with list, and a DataFrame with list and dictionary.

5. Write about the essential functionalities in pandas. Explain the following in detail using suitable examples:

 a) Re indexing b) Dropping entries from an axis. c) Indexing d) Selection e) Filtering

Unit 3

1. Explain reading and writing data in text format and write about parsing functions in pandas.

2. Write about function arguments used for reading csv/excel files.
3. How to read, write and append data to files? Explain with syntax and examples.

4. Explain about web scraping with an example.

5. Write about Interacting with API's with an example.

Unit 4 1.Write

about handling missing data

a) Filtering out missing data

b) Filling in missing data

2.Write about data transformation

a) removing duplicates

b) transforming data using a function or mapping

c) replacing values

d) renaming index

3.Write about detecting and filtering outliers

4.Explain about String Manipulation Techniques.

5.Write about regular expression and write the code to retrieve pattern on email address

Unit 5

1.What is Hierarchial Indexing? Give examples. Explain about indexing with data frame's columns?

2.Explain about combining and merging Datasets? Write about different join type?

3.Explain about Concatenation along Axis? Explain with example about concat function arguments?

4.What is data aggregation? What are the different aggregation functions. Explain with examples.

5.Write about Groupby mechanics? Explain about

a) Iterating over groups

b) Selecting a column or subset columns.

c) Grouping with functions.