DATAFRAMES
VIVA QUESTIONS

1. **What is a DataFrame?**
   A DataFrame is a 2D labeled heterogeneous, data-mutable and size-mutable array. The data in DataFrame is aligned in a tabular fashion in rows and columns therefore has both a row and column labels.

2. **What is a difference between DataFrames and Numpy Array?**

   The basic difference between a Numpy array and a Dataframe is that Numpy array contains homogenous data while Dataframe contains heterogenous data.

3. **How can we create customized labels in DataFrames?**
   The index and columns parameters can be used to change the default row and column labels.

4. **What happens when list of dictionaries is used to create a DataFrame?**
   The dictionary keys are treated as column labels and row labels take default values starting from zero. The values corresponding to each key are treated as rows. The number of rows is equal to the number of dictionaries present in the list.

5. **What values are filled in the dataFrame for missing keys?**
   NaN

6. **What happens when you try to add a column by the label which already exist?**
   If the column already exists by the same label then the values of that column are updated by the new values.

7. **Name the method which is used to add a new row and column.**
   loc() method

8. **Name the method which is used to delete a row/column?**
   DataFrame.drop() method

9. **Name the method is used to rename the labels of rows and columns in a DataFrame.**
   DataFrame.rename() method
   The parameter *axis=columns* to reset column labels and *axis=index* to reset row labels.

10. **What is the difference between loc() and iloc() methods?**

The **loc[]** method is used to retrieve a group of rows and columns by labels or a boolean array present in the DataFrame.

Whereas, the **iloc[]** method takes only integers and is used retrieve a group of rows and columns at particular positions in the index.

11. How do we iterate over DataFrame?

There are 3 ways to iterate over Pandas dataframes. These are-

a) **iteritems():** Helps to iterate over each element of the set, column-wise.
b) **iterrows():** Helps to iterate over each element of the set, row-wise.
c) **itertuple():** Helps to iterate over each row and form a tuple out of them.

12. Name the functions used to perform the following operations on DataFrame:

Operations on DataFrame	Functions
To add all values present in a DataFrame	sum()
To find the mean of all values present in a DataFrame	mean()
To find the minimum of all values present in a DataFrame	min()
To find the maximum of all values present in a DataFrame	max()
To count the number of values, present in a DataFrame	count()
To return the sorted DataFrame(ascending=True/False)	sort_values()
To check for missing data on DataFrame object	isnull()