Programming Language Learning Series
Mastery of Python Language
(Interview Questions/Assignment-Series&DataFrame)

Q1: Find the output produced by following expressions.

> *import pandas as pd*
> *s = pd.Series([10,20,30,40], index=[1,4,6,8])*
> > a. *s[1]*
> > b. *s[1:4]*

Q2: Find the output produced by following expressions.

> *import pandas as pd*
> *s1 = pd.Series([1,2,3,4])*
> *s2 = pd.Series([1,2,3,4], index=range(1, 5))*
> *s1+s2*

Q3: How would you convert the following series to float type?

> *s = pd.Series([1,2,3,4])*

Q4: Find the output produced by following expressions.

> *import pandas as pd*
> *data = pd.DataFrame(2d array 4 by 4, columns=['C1',..,'C4'], index=['r1',..'r4'])*
> > a. *data.loc['r1':'r3', 'c2':'c4']*
> > b. *data.loc[:'r4', :'c3']*
> > c. *data.iloc[0:2,1:3]*
> > d. *data.iloc[0:2,::2]*

Q5: Find the output of following expressions.

> *df = pd.DataFrame([10,20,30,40], columns=['c1'], index=[1,4,6,8])*
> > a. *df[1]*
> > b. *df[1:4]*
> > c. *df[df.c1>30]*

Programming Language Learning Series
Mastery of Python Language
(Interview Questions/Assignment-Series&DataFrame)

Q6: Find the output of following code. How do you handle duplicate names when merge?

```
df1 = pd.DataFrame(d1, columns=list('ABC'))
df2 = pd.DataFrame(d1, columns=list('ACD'))
pd.merge(df1, df2, on='A')
```

Q7: Find the output of following code. How do you get the index right after concatenation?

```
df1 = pd.DataFrame(d1, columns=list('ABC'))
df2 = pd.DataFrame(d1, columns=list('ACD'))
pd.concat([df1,df2], axis = 0)
```

Q8: What is the difference between merge and pjoin operations in pandas?

Q9: What is the output of the following code snippet?

```
import numpy as np
import pandas as pd
array = np.array([[1,2], [3,4]])
df = pd.DataFrame(array, columns=list('QR'))
df - df.iloc[0]
```

Q10: Using the iris dataset from http://goo.gl/3b3439, answer the following questions:

a) Querying Structural Properties
  - How many rows and columns are there?
  - What is the type of each column?
  - Show only the columns whose type is numeric
  - Convert all column names to UPPERCASE

b) Querying Data
  - Show the first 3 rows
  - Show the last 3 rows
  - Show 3 random rows without repetition
  - Show all unique values for the species column
  - Show rows 5 to 10 (inclusive)

c) Statistical Queries
- Count the number of flowers of each Species
- Count the number of observations where *Petal.Length* is longer than *Sepal.Width*
- Sort the observations by *Sepal.Width* in decreasing order.
- Find the mean, min and max values of all four measurements *(sepal.length, sepal.width, petal.length, petal.width) for* each species
- Find the average *petal.width* for rows where the *petal.length* is less than the *sepal.width*
- Find the Species with the most number of observations where the *Sepal.Length* is less then the mean *Sepal.Length* of all observations

d) Reshape
- Create a new data frame df that has only 3 columns (Species, Measure, Value) where Measure takes on the values *Sepal.Length, Sepal.Width, Petal.Length or Petal.Width*. Show the first 5 rows.
- Show the mean value and counts for each *Species* and Measure of df

e) Rearrange/Drop/Create Columns
- Rearrange the columns in the order *sepal_length, sepal_width, petal_length, petal_width, species*
- Create a new column *square_length* that is the square of *sepal_length*
- Create a new column *toal_length* that is the square of *sepal_length* and *petal_length*
- Drop the *species* column

Q11: The *heart* dataframe at https://goo.gl/CbJwQM contains information about the survival of patients on the waiting list for the Stanford heart transplant program.

> *start, stop, event: Entry and exit time and status for this interval of time*
> *age: age-48 years*
> *year: year of acceptance (in years after 1 Nov 1967)*
> *surgery: prior bypass surgery 1=yes*
> *transplant: received transplant 1=yes*
> *id: patient id*

Answer the following questions with respect to the heart data set:
- Sort the data frame by age in descending order (oldest at top) without making a copy
- How many patients received a transplant?
- What is the average age for transplanted patients under the age of 70?
- Find the mean and standard deviation of age for each value of the transplant variable.

Q12: Read the flights data at https://raw.githubusercontent.com/mwaskom/seaborn-data/master/flights.csv into a pandas data frame. Find the average number of passengers per quarter (Q1, Q2, Q3, Q4) across the years 1950-1959 (inclusive of 1950 and 1959), where

- Q1 = Jan, Feb, Mar   Q2 = Apr, May, Jun
- Q3 = Jul, Aug, Sep   Q4 = Oct, Nov, Dec

Q13: Read the following data sets into DataFrames.
- url1 = https://raw.github.com/vincentarelbundock/Rdatasets/master/csv/DAAG/hills.csv
- url2 = https://raw.github.com/vincentarelbundock/Rdatasets/master/csv/DAAG/hills2000.csv

Create a new DataFrame only containing the names present in both DataFrames. Drop the `timef` column and have a single column for `dist` , `climb` and `time` that shows the average value of the two DataFrames. The final DtataFrame will thus have 4 columns (name, dist, climb, time).