**Data Science Foundations – Assessment**

Ali hasan

This test will assess your knowledge of the usage of the Python NumPy and Pandas libraries. Each student should insert their solution below each question

**Part A**

**Q 1:** Write the code to create a **Pandas Series** containing these values [2, 4, 6, 8, 10].

Import numpy as np

Import pandas as pd

s = pd.Series([2,4,6,8,10])

print(s)

**Q 2:** Write the **NumPy code** to add, subtract, multiply and divide two arrays.

a = [2, 4, 6, 8, 10]

b = [1, 3, 5, 7, 9]

import numpy as np

in\_arr1 =np.array([2,4,6,8,10])

In\_arr2=np.array([1,3,5,7,9])

out\_arr=np.add(in\_arr1, in\_arr2)

print(“added array : ”, out\_arr)

out\_num = np.subtract(in\_arr1, in\_arr2)

print (“subtracted array : ”, out\_num)

out\_div = np.divide(in\_arr1, in\_arr2)

print (“divided array : ”, out\_div)

Out\_mul = in\_arr1\*in\_arr2

Print (“Multiplied array : ”, out\_mul )

**Q3:** Write the **Pandas code** to Read Purchases.csv file as a DataFrame and display first three rows of this DataFrame.

**Import pandas as pd**

**A = pd.read\_csv(“user/Aloosh/downloads/purchase.csv”)**

**Print(A.head(3))**

**Q4:** Write the **code** to find the column names of the data file exported as DataFrame in Q3.

Import pandas as pd

**A = pd.read\_csv(“user/Aloosh/downloads/purchase.csv”)**

**For col in A.columns:**

**Print(col)**

**Q5:** Explore the dataset in Q3 through *info*() and also find the shape of the data. Write the results here.

The code

import pandas as pd

A = pd.read\_csv("/Users\Aloosh\Downloads\purchases.csv")

Shape=A.shape

print(A)

print("Shape={}")

A.info()

The results

Unnamed: 0 apples oranges

0 June 3 0

1 Robert 2 3

2 Lily 0 7

3 David 1 2

Shape={}

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 4 entries, 0 to 3

Data columns (total 3 columns):

Unnamed: 0 4 non-null object

apples 4 non-null int64

oranges 4 non-null int64

dtypes: int64(2), object(1)

memory usage: 224.0+ bytes

**PART B:** Consider the code below which creates and display a Pandas DataFrame from the dictionary data and index labels below.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| name | Anastasia | Dima | Katherine | James | Emily | Michael | Matthew | Laura | Kevin | Jonas |
| score | 12.5 | 9 | 16.5 | np.nan | 9 | 20 | 14.5 | np.nan | 8 | 19 |
| attempts | 1 | 3 | 2 | 3 | 2 | 3 | 1 | 1 | 2 | 1 |
| qualify | yes | No | Yes | no | no | yes | yes | no | no | yes |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| labels | a | b | c | d | e | f | g | h | i | j |

**Dictionary data and list labels**

import pandas as pd

import numpy as np

exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'], 'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19], 'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}

labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']

df = pd.DataFrame(exam\_data , index=labels)

print(df)

**Q6:** Write the Pandas code to display a summary of the basic information about the DataFrame (df) created above.

**import pandas as pd**

**import numpy as np**

**exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'], 'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19], 'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}**

**labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']**

**df = pd.DataFrame(exam\_data , index=labels)**

**''' the answer '''**

**print(df.shape)**

**print(df.dtypes)**

**print(df.size)**

**print(df.ndim)**

**Q7:** Write the Pandas code to get the first three rows of df.

import pandas as pd

import numpy as np

exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'], 'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19], 'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}

labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']

df = pd.DataFrame(exam\_data , index=labels)

''' the answer '''

print(df.head(3))

**Q8:** Write the Pandas code to retrieve the 'score' and 'attempts' columns of df.

**import pandas as pd**

**import numpy as np**

**exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'], 'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19], 'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}**

**labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']**

**df = pd.DataFrame(exam\_data , index=labels)**

**''' the answer '''**

**subset=df[['score','attempts']]**

**print(subset)**

**Q9:** Write the Pandas code to select 1, 3, 5, 6 rows of 'attempts' and 'score' column of df.

**import pandas as pd**

**import numpy as np**

**exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'], 'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19], 'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}**

**labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']**

**df = pd.DataFrame(exam\_data , index=labels)**

**''' the answer '''**

**print(df.iloc[[0,2,4,5],:].loc[:,['attempts','score']])**

**Q 10:** Write a program to find the rows for which the score is between 15 and 20 (inclusive) from df.

**import pandas as pd**

**import numpy as np**

**exam\_data = {'name': ['Anastasia', 'Dima', 'Katherine', 'James', 'Emily', 'Michael', 'Matthew', 'Laura', 'Kevin', 'Jonas'], 'score': [12.5, 9, 16.5, np.nan, 9, 20, 14.5, np.nan, 8, 19], 'attempts': [1, 3, 2, 3, 2, 3, 1, 1, 2, 1],'qualify': ['yes', 'no', 'yes', 'no', 'no', 'yes', 'yes', 'no', 'no', 'yes']}**

**labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']**

**df = pd.DataFrame(exam\_data , index=labels)**

**''' the answer '''**

**for ele in df.iterrows():**

**if 15 <=float(ele[1]['score']) <=20.0:**

**print(ele)**