Fourth Industrial Revolution (4IR) Summer School

Data Preparation – Day 1 exercises

Pandas Data Structures

Question 1

Write a Pandas program to add, subtract, multiple and divide two Pandas Series.

Sample Series: [2, 4, 6, 8, 10], [1, 3, 5, 7, 9]

Question 2

Write a Pandas program to compare the elements of the two Pandas Series.
Sample Series: [2, 4, 6, 8, 10], [1, 3, 5, 7, 9]

Question 3

Write a Python program to convert a NumPy array to a Pandas series, then convert a the resulted Series into an array again.

Sample array np.array([10, 20, 30, 40, 50])

Question 4

Write a Pandas program to add some data to an existing Series.

Sample Series: [2, 4, 6, 8, 10]

Question 5

Write a Pandas program to create a subset of a given series based on value and condition.

Sample Series: [2, 4, 6, 8, 10]

Sample condition: the value > 5.

Question 6

Write a Pandas program to create a DataFrame using the data given in the following:

Sample DataFrame:

```
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']
```

a) calculate the sum of the examination attempts by the students.
b) change the score in row 'd' to 11.5.
c) return the row having the maximum score
d) append a new row 'k' to DataFrame with given values for each column.
e) delete the new row and return the original data frame.
f) change the name 'James' to 'Sam' in name column of the data frame.
g) insert a new column in existing DataFrame.
h) get list from DataFrame column headers.
i) get the datatypes of columns of a DataFrame