Data Analysis and Visualization Using Python
Internal Assessment 1 25.05.2023

Q1. From 2 numpy arrays a and b, Write the code to extract the indexes in which the elements in the 2 arrays match. 1 mark

```
a = np.array([1,2,3,4,5])

b = np.array([1,3,2,4,5])
```

Q2. Write a Pandas program to convert a dictionary to a Pandas series. Where the dictionary has indices a,b,c,d,e with values 100, 200,300,400,500 respectively. 1 mark

Q3. Write a Pandas program to convert the first column of a DataFrame as a Series. Where col1 has values 1, 2, 3, 4, 7, 11, col2 has values 4, 5, 6, 9, 5, 0 and col3 has values 7, 5, 8, 12, 1,11 1 mark

Q4. Write a Pandas program to create a series having values 0 to 9. Display the subset of the given series having all values greater than 6. 1 mark

Q5. Write a Pandas program to change the order of the index of a given series. 1 mark

Q6. Create a series with an index from 0 to 34 and values between 1 to 9 using a random function. Then reshape the series into a dataframe with 7 rows and 5 columns. 2 mark

Q7. From `ser1` remove items present in `ser2`. 1 mark

Q8. Let obj = pd.series(range(3), index=['a', 'b', 'c'])

```
index = obj.index

index[1] = 'd'
```

Considering the above code snippet. What would be the resultant index.

Q9. Answer the following queries using data frame Test Data. 5 marks

Test Data:

	ord_no	purch_amt	ord_date	customer_id	salesman_id
0	70001.0	150.50	2012-10-05	3002	5002.0
1	NaN	270.65	2012-09-10	3001	5003.0
2	70002.0	65.26	NaN	3001	5001.0
3	70004.0	110.50	2012-08-17	3003	NaN
4	NaN	948.50	2012-09-10	3002	5002.0
5	70005.0	2400.60	2012-07-27	3001	5001.0
6	NaN	5760.00	2012-09-10	3001	5001.0
7	70010.0	1983.43	2012-10-10	3004	NaN
8	70003.0	2480.40	2012-10-10	3003	5003.0
9	70012.0	250.45	2012-06-27	3002	5002.0
10	NaN	75.29	2012-08-17	3001	5003.0
11	70013.0	3045.60	2012-04-25	3001	NaN

a) Detect missing values of a given DataFrame. Display True or False.
b) Identify the column(s) of a given DataFrame which have at least one missing value.
c) Count the number of missing values in each column of a given DataFrame.
d) Drop the rows where at least one element is missing in a given DataFrame.
e) Drop the rows where all elements are missing in a given DataFrame.
f) To keep the rows with at least 2 NaN values in a given DataFrame.
g) Total number of missing values of the said DataFrame.
h) To replace NaNs with a single constant value in specified columns in a DataFrame.
i) To replace NaNs with the value from the previous row or the next row in a given DataFrame.
j) To replace NaNs with median or mean of the specified columns in a given DataFrame.