

▼ Chapter 4 - Exercise 2: Hãy thực hiện những yêu cầu liên quan tới Data Frame

```
import numpy as np
import pandas as pd
```

```
# Câu 1: Cho dictionary như sau:
```

```
dic_1 = {'X':[78,85,96,80,86], 'Y':[84,94,89,83,86], 'Z':[86,97,96,72,83]}
```

```
# Tạo dataframe df1 từ dic_1
```

```
df1 = pd.DataFrame(dic_1)
```

```
# In nội dung của dataframe df1
```

```
df1
```

```
# Câu 2:
```

```
# Cho dictionary exam_data như sau:
```

```
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']}
```

```
# Cho list labels như sau:
```

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

```
# Câu 2a: Tạo dataframe df2 từ exam_data, với index của dataframe là labels
```

```
df2 = pd.DataFrame(exam_data, index = labels)
```

```
# In nội dung của dataframe df2
```

```
df2
```

```
# Câu 2b: Xem thông tin (info()) của dataframe df2
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 10 entries, a to j
Data columns (total 4 columns):
name          10 non-null object
score         8 non-null float64
attempts      10 non-null int64
qualify       10 non-null object
dtypes: float64(1), int64(1), object(2)
memory usage: 400.0+ bytes
```

```
# Câu 3: Tạo dataframe df3 từ df2, chỉ chứa 2 cột là name và score
df3 = df2[['name', 'score']]
# Xem kiểu dữ liệu (type) và kích thước (shape) của df3
print(type(df3))
print(df3.shape)
# Hiển thị các dòng dữ liệu đầu tiên (head) của df3
df3.head()
```

```
# Câu 4: Hiển thị thông tin thống kê chung (describe) của dataframe df2
df2.describe(include = 'all')
```

```
# Câu 5: Tạo dataframe df4 từ df2, chỉ chứa 2 cột là name và score, và chỉ có các dòng 1, 3, 5, 6
df4 = df2.iloc[[1, 3, 5, 6], [0, 1]]
```

```
# In nội dung của dataframe df4
df4
```

```
# Câu 6: Từ dataframe df2, in các dòng có dữ liệu score bị null
df2[df2['score'].isnull()]
```

```
# Câu 7: Từ dataframe df2, in các dòng có score > 15 và <=20
df2[(df2['score'] > 15) & (df2['score'] <= 20)]
```

```
# Câu 8: Cập nhật điểm (score) ở dòng 'd' thành 18
df2.loc['d', 'score'] = 18
df2.head()
```

```
# Câu 9: Cho biết điểm (score) nào có tần suất xuất hiện nhiều nhất trong df2, và in ra những dòng có điểm là tần suất xuất l
mark = df2['score'].mode()
print(mark[0])
```

```
score = df2[(df2['score']==mark[0])]
score
```

```
# Cho biết có bao nhiêu dòng thỏa điều kiện?
#score.shape[0]
```

```
# Câu 10: Thêm dòng k có nội dung như sau: ['Suresh', 15.5, 1, 'yes'] vào df2
df2.loc['k'] = ['Suresh', 15.5, 1, 'yes']
# Hiển thị 5 dòng cuối cùng (tail) của df2.
df2.tail()
```

```
# Câu 11a: Thêm dòng l có nội dung như sau: ['Janny', 12.5, 2, 'yes'] vào df2.
```

```
df2.loc['1'] = ['Janny', 12.5, 2, 'yes']  
# Hiển thị 5 dòng cuối cùng (tail) của df2.  
df2.tail()
```

```
# Câu 11b: Xóa bỏ dòng 1 của df2.  
df2 = df2.drop(['1'])  
df2.tail()  
# Hiển thị lại 5 dòng cuối cùng (tail) của df2.  
df2.tail()
```

```
# Câu 12: Sắp xếp df2 tăng dần theo điểm (score)  
df2 = df2.sort_values(by='score')  
df2
```

```
# Câu 13: Thêm cột result vào df2,  
# dựa vào dữ liệu của cột'score', nếu dòng nào có điểm >=10 thì giá trị của cột result = 1, ngược lại = 0  
df2['result'] = df2['score'].map(lambda x: 1 if x >= 10 else 0)  
# In nội dung của dataframe df2  
df2
```

```
# Câu 14: Trong df2, thay tên 'Emily' thành 'Samantha'  
df2['name'] = df2['name'].replace('Emily', 'Samantha')  
# In lại nội dung của dataframe df2  
df2
```

```
# Câu 15: Duyệt df2, in name, score, result: nếu nếu giá trị cột result = 1 thì in 'Pass', ngược lại thì in 'Fail'
```

```
for index, row in df2.iterrows():
    print(row['name'], ' - Score:', row['score'], ' - Result:', "Pass" if row['result'] == 1 else "Fail")

Kevin   - Score: 8.0   - Result: Fail
Dima    - Score: 9.0   - Result: Fail
Samantha - Score: 9.0   - Result: Fail
Anastasia - Score: 12.5 - Result: Pass
Matthew - Score: 14.5 - Result: Pass
Suresh  - Score: 15.5 - Result: Pass
Katherine - Score: 16.5 - Result: Pass
James   - Score: 18.0 - Result: Pass
Jonas   - Score: 19.0 - Result: Pass
Michael - Score: 20.0 - Result: Pass
Laura   - Score: nan  - Result: Fail
```