

DATA SCIENCE

Bài tập Chương 5

1. Tạo 1 pandas DataFrame có 40 phần tử ngẫu nhiên, giá trị <200 và có 4 cột ABCD

1a. In ra dataframe

1b. In ra hàng có giá trị tăng dần

1c. Thêm 1 cột E có giá trị ngẫu nhiên < 300 vào DataFrame và in ra DataFrame mới

1d. In ra 3 cột B, C và E

```
import pandas as pd
import numpy as np

df = pd.DataFrame(np.random.randint(200, size=(10, 4)), columns=list('ABCD'))
#a
print(df)

#b
dong_tang= df[(df.A < df.B) & (df.B < df.C) & (df.C < df.D)]
print(dong_tang)

#c
cot =np.random.randint(1,300)
df["E"]=cot
print(df)

# d
print(df.drop(["A", "D"], axis=1))
```

2. Nhập dữ liệu từ file k2020.csv (file kèm)

2.a. In dữ liệu ra màn hình

```
import pandas as pd
d = pd.read_csv("k2020.csv", index_col = 0)
# câu 1: in dữ liệu ra màn hình
print(d) |
```

2.b. In 5 dòng đầu tiên và 5 dòng cuối cùng của dữ liệu ra màn hình

```
import pandas as pd
d = pd.read_csv("k2020.csv", index_col = 0)
# câu 1: in dữ liệu ra màn hình
print(d)
#in 5 dòng đầu tiên
print(d.head(5))
#in 5 dòng cuối cùng
print(d.tail(5))
```

2.c. Thống kê xem lớp có bao nhiêu bạn điểm loại giỏi (điểm từ 8 trở lên)

```
import pandas as pd
d = pd.read_csv("k2020.csv", index_col = 0)

#thống kê loại giỏi
print("Sinh viên đạt loại giỏi gồm ", len(d[d.Diem >= 8]), "sinh viên") #
```

2.d. Thông kê xem lớp có bao nhiêu bạn trượt môn (điểm dưới 4 hoặc không có điểm)

```
d = pd.read_csv("k2020.csv", index_col = 0)
print("Sinh viên rớt môn gồm ",len(d[(d.Diem < 4) | (d.Diem.isnull())]))
```