

Lab Data Cleaning01

ให้นักศึกษาทำการเขียนโค้ดต่อไปนี โดยใช้ google colab

1. ทำการเรียกใช้ pandas

```
import pandas as pd
```

2. ทำการสร้าง data series ซึ่งมีลักษณะคล้ายกับ column ของ table

```
a = [1, 7, 2]
```

```
print(a)
```

```
myvar = pd.Series(a)
```

```
print(myvar)
```

##ดูข้อแตกต่างของ print(a) ที่เป็น List และ print(myvar) ที่เป็น Pandas data series

3. การกำหนด Index ให้กับข้อมูลใน data series โดยใช้ attribute index

```
myData=pd.Series(["dog","cat","fish","cow"],index=["a","b","c","d"])
```

```
print(myData)
```

```
print(myData["b"])
```

*** ให้เขียน code เพิ่มเติมเพื่อกำหนด index ให้กับ myvar ในข้อ 2 และ print เพื่อแสดงผลลัพธ์

4. ทำการสร้าง data frame ซึ่งมีลักษณะคล้าย table ประกอบด้วย row และ column โดยใช้ข้อมูล dictionary

```
student={
```

```
    "name":["David","Micheal","Tim","Alice","John"],
```

```
    "Age":[18,24,19,25,19],
```

```
    "Gender":["Male","Male","Male","Female","Male"],
```

```
    "GPA":[2.55,3.15,3.58,2.32,1.98]
```

```
}
```

```
myStudent=pd.DataFrame(student)
```

```
print(myStudent)
```

5. การใช้คำสั่ง Loc เพื่อแสดงผลข้อมูลบางส่วน

```
print(myStudent.loc[2]) ## แสดงข้อมูลแถว 2
```

```
print(myStudent.loc[[2, 4]])
```

```
print(myStudent.loc[1,"name"])
```

```
print(myStudent.loc[[3, 4],"Gender"]) ##แสดงเฉพาะ Gender แถว 3,4
```

```
print(myStudent.loc[myStudent["GPA"]>3]) ##กำหนดเงื่อนไข GPA มากกว่า 3
```

6. ทำการสร้าง data frame โดยกำหนด index

```
data = {
```

```
"calories": [420, 380, 390,300,330,400,450,375,380,405],
```

```
"duration": [50, 40, 45,50,60,54,56,61,48,53]
```

```
}
```

```
df = pd.DataFrame(data, index=["day1", "day2", "day3","day4", "day5","day6", "day7",  
"day8","day9", "day10"])
```

```
print(df)
```

```
print(df.loc["day2"])
```

7. การเพิ่มข้อมูลใหม่เข้าสู่ data frame ด้วย insert

```
df.insert(1, "heartRate", [101, 102,103,104,105, 101, 100,103,107,106])
```

```
print(df)
```

8. การแสดงผลข้อมูลโดยใช้คำสั่ง head() , tail() , info() , shape ให้อธิบายว่าแต่ละคำสั่งให้ผลลัพธ์ต่างกันอย่างไร

```
print(df.head())
```

```
print(df.tail(3))
```

```
print(df.info())
```

```
print(df.shape)
```

9. การหาค่าทาง สถิติ อย่างง่าย mean , median , mode

```
mean_calories=df["calories"].mean()
```

```
print("Mean of Calories :: ",mean_calories)
```

```
median_calories=df["calories"].median()
```

```
print("Median of Calories :: ",median_calories)
```