เแนวคำตอบใบกิจกรรมที่ 7.1

ตามหาข้อมูล

|  |
| --- |
| ชื่อ - นามสกุล …………………………………………….. ชั้น ………… เลขที่ …….. |

1. ศึกษาหัวข้อ 2.1.2 โครงสร้างข้อมูลของ Pandas 2.1.3 ไฟล์ข้อมูลในรูปแบบซีเอสวี และ 2.1.4 การเชื่อมโยงข้อมูล แล้วทดลองเขียนโปรแกรมตัวอย่างที่ 2.1 -2.4 จากหนังสือเรียนในบทที่ 2 และให้นักเรียนสรุปสิ่งที่ได้

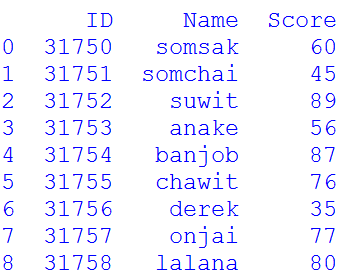
………………………………………………………………………………………………………………………………………………………………...

1. เขียนโปรแกรมชื่อ ex\_71.py เพื่อแสดงข้อมูลรหัสประจำตัวนักเรียน ชื่อ และคะแนนสอบ จากไฟล์ score.csv โดยให้มีหัวตารางเป็น ID, Name, Score

ไฟล์ score.csv มีรายละเอียดดังนี้

|  |
| --- |
| 31750,somsak,60  31751,somchai,45  31752,suwit,89  31753,anake,56  31754,banjob,87  31755,chawit,76  31756,derek,35  31757,onjai,77  31758,lalana,80 |

ตัวอย่างผลลัพธ์



แนวการตอบ

|  |
| --- |
| import pandas as pd  stdData = pd.read\_csv('score.csv', names= ['ID', 'Name', 'Score'])  print(stdData) |

1. เขียนคำสั่งโดยใช้ DataFrame จากข้อ 2 ให้แสดงผลลัพธ์ดังต่อไปนี้

|  |  |
| --- | --- |
| ผลลัพธ์ | คำสั่ง |
| 4.1 แสดง DataFrame ทั้งหมด | print(stdData) |
| 4.2 แสดง ชื่อกับคะแนน | print(stdData[['Name','Score']]) |
| 4.3 แสดงชื่อ 5 คนแรก | print(stdData['Name'].head(5)) |
| 4.4 แสดงข้อมูลนักเรียนที่มีคะแนนตั้งแต่ 50 คะแนนขึ้นไป | print(stdData[stdData.Score>=50]) |
| 4.5 แสดงจำนวนนักเรียนทั้งหมด | print(stdData['Name'].count()) |
| 4.6 แสดงคะแนนเฉลี่ยของนักเรียนทั้งหมด | print(stdData.Score.mean()) |
| 4.7 แสดงคะแนนเฉลี่ยของนักเรียน 4 คนสุดท้าย | print('Mean',stdData.Score.tail(4).mean()) |

**ใบกิจกรรมที่ 7.2**

**ประมวลผลข้อมูล**

1. เขียนโปรแกรมตัวอย่างที่ 2.5 - 2.8 จากหนังสือเรียนในบทที่ 2
2. ให้นักเรียนดาวน์โหลดไฟล์ [Data1.csv](https://drive.google.com/file/d/1wpwmIfWwV_hatdnaFjIxLC0A0NVaFX6e/view?usp=sharing) และไฟล์ [Data2.csv แล้วเขียนโปรแกรม](https://drive.google.com/open?id=1btzUMa6-EcQZSOrIGx4jO9e02-XE5IKs)รวมข้อมูลทั้ง 2 ไฟล์ ชื่อ ex\_72.py ตามคำสั่งดังนี้

|  |
| --- |
| import pandas as pd  stdData1=pd.read\_csv('Data1.csv')  stdData2=pd.read\_csv('Data2.csv')  mData = stdData1.merge(stdData2) |

3. เขียนคำสั่งโดยใช้ DataFrame จากข้อ 2 ให้แสดงผลลัพธ์ดังต่อไปนี้

|  |  |
| --- | --- |
| ผลลัพธ์ | คำสั่ง |
| 3.1 แสดงข้อมูลทั้งหมดของการรวมข้อมูล | print(mData) |
| 3.2 แสดงข้อมูล ชื่อชั้นเรียน ชื่อนักเรียน และหมู่เลือด | print(mData[['Class','Name','BloodType']]) |
| 3.3 แสดงข้อมูล ชื่อชั้นเรียน ชื่อนักเรียน ความสูง และเลขประจำตัว โดยเรียงลำดับตามความสูง จากมากไปน้อย | print(mData.sort\_values(by=['Height'],ascending=False)[['Class','Name','Height','ID']]) |
| 3.4 แสดงข้อมูล ชื่อชั้นเรียน เลขที่ ชื่อนักเรียน และน้ำหนัก เรียงตามชื่อชั้นเรียนและเลขที่ จากน้อยไปมาก | print(mData.sort\_values(by=['Class','Weight'])[['Class','Number','Name','Weight']]) |
| 3.5 แสดงค่าน้ำหนักเฉลี่ยแยกตามหมู่เลือด | print(mData.groupby('BloodType').mean()['Weight']) |
| 3.6 แสดงค่าความสูงของคนที่สูงที่สุดในแต่ละชั้นเรียน | print(mData.groupby('Class').max()['Height']) |

4. ให้นักเรียนเขียนโปรแกรมเพื่อแสดงข้อมูลใน Data1.csv โดยลบข้อมูลชื่อ Adebayor ให้เหลือเป็นช่องว่าง เมื่อรันโปรแกรมจะได้ผลลัพธ์อย่างไร

จะไม่แสดงข้อมูลแถวที่มีการลบชื่อออกไป/ชื่อเป็นช่องว่าง

………………………………………………………………………………………………………………………………………..

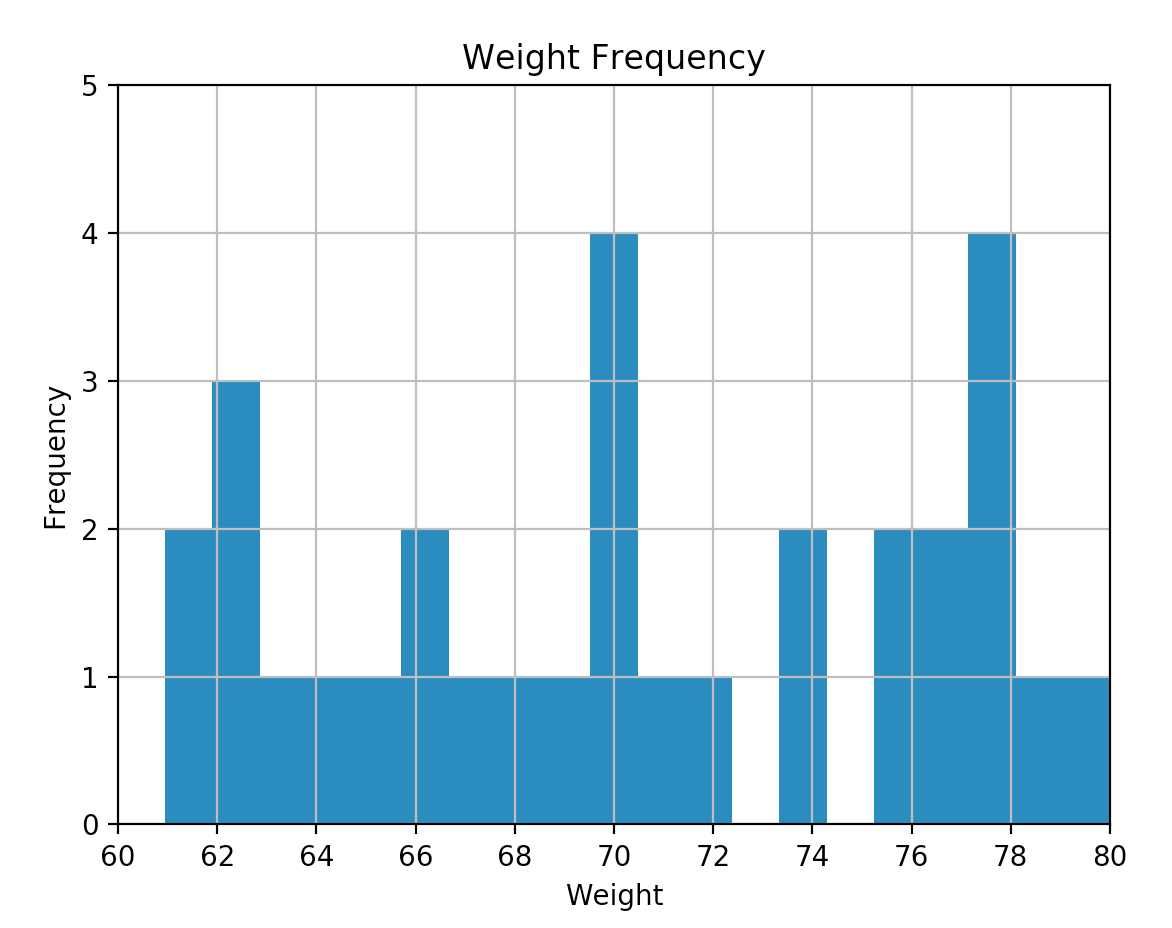
ใบกิจกรรมที่ 7.3

ทำข้อมูลให้เป็นภาพ

|  |
| --- |
| ชื่อ - นามสกุล …………………………………………….. ชั้น ………… เลขที่ …….. |

1. ศึกษาหัวข้อ 2.1.5 การนำเสนอข้อมูลด้วยภาพ แล้วเขียนโปรแกรมตัวอย่างที่ 2.9 จากหนังสือเรียนในบทที่ 2

2. เขียนโปรแกรมชื่อ ex\_73.py แสดงฮิสโทแกรมของน้ำหนัก ที่มีชื่อเป็น Weight Frequency โดยแสดงน้ำหนัก (Weight) ที่มีค่าอยู่ระหว่าง 60-80 กิโลกรัม แบ่งเป็น 21 ช่วง บนแกน x และแสดงจำนวนคน (Frequency) บนแกน y ที่มีค่าอยู่ระหว่าง 0-5 และแสดงเส้นกริดตามแนวแกน x ทุก 2 กิโลกรัม และตามแนวแกน y ทุก 1 ค่าความถี่ โดยมีตัวอย่างผลลัพธ์ดังนี้

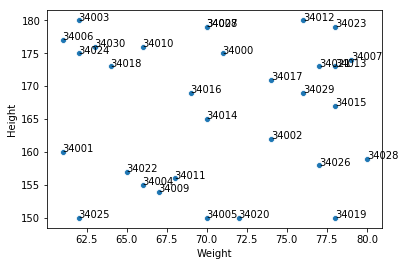


แนวการตอบ

|  |
| --- |
| import matplotlib.pyplot as plt  import pandas as pd  stdData1=pd.read\_csv('Data1.csv')  stdData2=pd.read\_csv('Data2.csv')  mData = stdData1.merge(stdData2)  mData.Weight.hist(range=[60,80],bins=21)  plt.title('Weight Frequency')  plt.ylabel('Frequency')  plt.xlabel('Weight')  plt.axis([60,80,0,5])  plt.xticks(range(60,81,2))  plt.show() |

3. เขียนโปรแกรมแสดงแผนภาพการกระจายของความสัมพันธ์ระหว่างน้ำหนักและส่วนสูง โดยให้แสดงเลขประจำตัวของเจ้าของข้อมูลตรงตำแหน่งที่แสดงบนแผนภาพ

ตัวอย่างการแสดงผลลัพธ์มีดังนี้



แนวคำตอบ

|  |
| --- |
| import pandas as pd  import matplotlib.pyplot as plt  stdData1=pd.read\_csv('Data1.csv')  stdData2=pd.read\_csv('Data2.csv')  mData = stdData1.merge(stdData2)  plt.scatter(mData.Weight, mData.Height)  for i in range(len(mData)):  plt.text(mData.Weight[i],mData.Height[i],mData.ID[i])  plt.show() |

ใบกิจกรรมที่ 7.4

หาค่าดัชนีมวลกาย

|  |
| --- |
| ชื่อ - นามสกุล …………………………………………….. ชั้น ………… เลขที่ …….. |

1. ดาวน์โหลดไฟล์ข้อมูลน้ำหนักและส่วนสูงของทุกคนจากไฟล์ bmi.csv แล้วเขียนโปรแกรมในกิจกรรมที่ 2.1 ข้อ 1-2 จากหนังสือเรียนในบทที่ 2 ตามขั้นตอนการแก้ปัญหา แล้วบันทึกชื่อไฟล์เป็น act21.py

**การวิเคราะห์และกำหนดรายละเอียดของปัญหา**

ข้อมูลเข้า คือ ไฟล์ bmi.csv ซึ่งประกอบไปด้วยข้อมูลน้ำหนักและส่วนสูง

ข้อมูลออก คือ ค่าดัชนีมวลกาย

วิธีในการตรวจสอบข้อมูล มีดังนี้

ถ้าน้ำหนักมีค่า 58 กิโลกรัม และส่วนสูง 158 เซนติเมตร แสดงค่าดัชนีมวลกายเท่ากับ 23.23

ถ้าน้ำหนักมีค่า 60 กิโลกรัม และส่วนสูง 164 เซนติเมตร แสดงค่าดัชนีมวลกายเท่ากับ 22.30

ถ้าน้ำหนักมีค่า 55 กิโลกรัม และส่วนสูง 150 เซนติเมตร แสดงค่าดัชนีมวลกายเท่ากับ 24.44

**การวางแผนการแก้ปัญหา**

|  |
| --- |
| อ่านไฟล์ bmi.csv  คำนวณ BMI ← น้ำหนัก (กิโลกรัม) /ส่วนสูง2 (เมตร)  แสดง BMI |

การดำเนินการแก้ปัญหา โดยเขียนโปรแกรมได้ดังนี้

|  |
| --- |
|  |

แนวคำตอบ

import pandas as pd

bmidata=pd.read\_csv('bmi.csv',delimiter=',')

df = pd.DataFrame(bmidata)

df['BMI']=df['Weight']/((df['Height']/100)\*\*2)

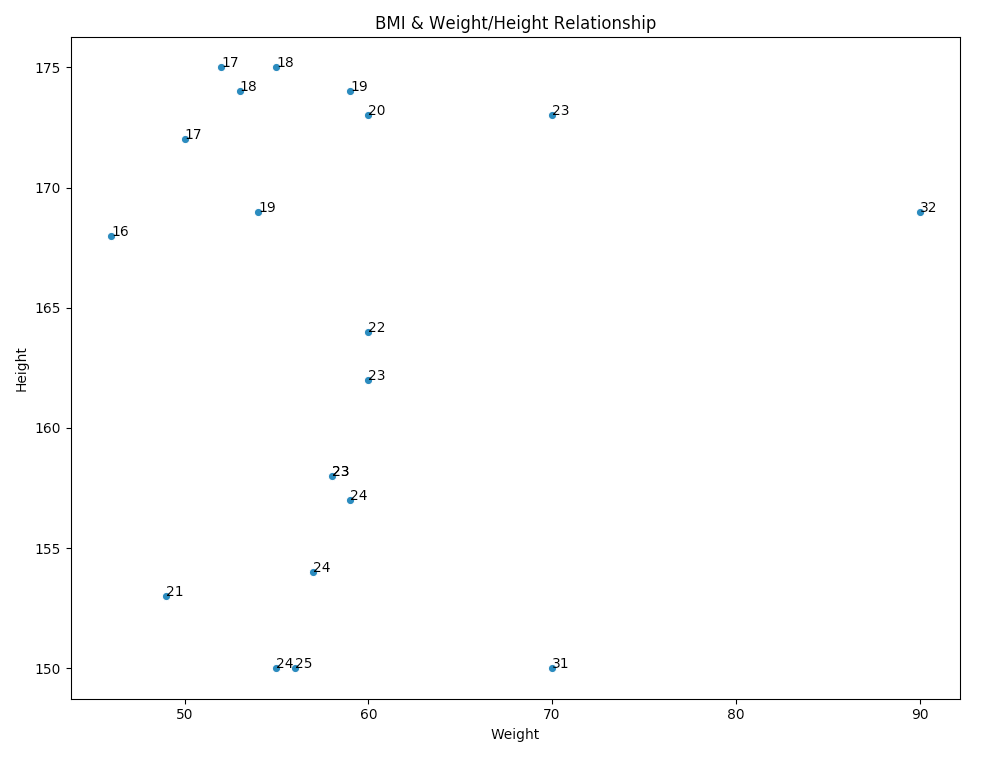
print(df)

2. ถ้านักเรียนต้องการเห็นความสัมพันธ์ของดัชนีมวลกายของทุกคนในภาพรวม ข้อมูลที่ได้จากการรันโปรแกรมข้อ 1 เหมาะสมหรือไม่ หรือควรจะแสดงข้อมูลในรูปแบบใดจึงจะแสดงได้ชัดเจนขึ้น

……………ไม่เหมาะสม เนื่องจากเห็นข้อมูลทุกคนในภาพรวมไม่ชัดเจน ควรแสดงเป็นแผนภาพการกระจาย ……………………………………………..

3. เขียนโปรแกรมในกิจกรรมที่ 2.1 ข้อ 3 จากหนังสือเรียน แล้วบันทึกชื่อไฟล์เป็น act21\_3.py

ตัวอย่างผลลัพธ์ที่ได้มีดังนี้



แนวคำตอบ

|  |
| --- |
| import pandas as pd  import matplotlib.pyplot as plt  bmidata=pd.read\_csv('bmi.csv',delimiter=',')  df = pd.DataFrame(bmidata)  df['BMI']=df['Weight']/((df['Height']/100)\*\*2)  plt.scatter(df.Weight, df.Height)  plt.xlabel('Weight')  plt.ylabel('Height')  plt.title('BMI & Weight/Height Relationship')  for i in range(len(df)):  plt.text(df.Weight[i],df.Height[i],str(int(round(df.BMI[i]))))  plt.show() |

NOTE: ไฟล์ bmi.csv เวอร์ชันภาษาอังกฤษ

ID,Gender,Weight,Height

5884,2,58,158

5885,2,60,164

5887,1,55,150

5889,1,46,168

5890,2,49,153

5893,1,50,172

5895,1,60,162

5896,1,59,157

5897,2,90,169

5900,2,53,174

5903,1,56,150

5905,2,70,150

5906,1,54,169

5908,1,55,175

5910,1,60,173

5912,1,70,173

5915,1,52,175

5916,2,57,154

5919,1,59,174

5921,2,58,158