**Pandas**

**การ Plot Graph ในรูปแบบต่างๆ สำหรับการทำ Data Visualization ตอนที่ 2**

import pandas as pd

df **=** pd**.**DataFrame({

 'name':['john','mary','peter','jeff','bill','lisa','jose'],

 'age':[23,78,22,19,45,33,20],

 'gender':['M','F','M','M','M','F','M'],

 'state':['california','dc','california','dc','california','texas','texas'],

 'num\_children':[2,0,0,3,2,1,4],

 'num\_pets':[5,1,0,5,2,2,3]

})

**ให้นักเรียนสร้างข้อมูลตามตัวอย่างข้างล่าง (เหมือนกับ worksheet ที่แล้ว) และพิมพ์ df กด run เพื่อดูข้อมูลที่สร้างขึ้นมา ตามตัวอย่างด้านล่าง (ข้อมูลจะมีทั้งหมด 6 columns และ index อีกหนึ่ง column)**



1. **จาก exercise ในสัปดาห์ที่แล้ว ให้นักเรียน plot กราฟแบบ pie สำหรับค่า num\_pets และกำหนด label ตามชื่อให้ถูกต้อง รวมถึงให้แสดงตารางควบคู่กับกราฟ pie ด้วย (นักเรียนสามารถดูคำสั่งได้จาก worksheet4)**

หากนักเรียนทำได้ถูกต้องจะได้กราฟ pie ดังต่อไปนี้

****

ในกรณีที่นักเรียนต้องการเน้นข้อมูลบางส่วนในกราฟ pie (แสดงดังภาพด้านล่าง) นักเรียนสามารถกำหนด argument เพิ่มเติมในคำสั่ง plot เช่นกำหนดให้ explode = (0,0.1,0,0,0,0) ในกรณีนี้เราต้องการ explode ค่าที่สอง ซึ่งเป็นจำนวนบุตรของ Mary (การกำหนด explode ให้ทำในฟังก์ชัน .plot( ) )



1. **ให้นักเรียนลองเปลี่ยนค่าใน explode เพื่อดูความแตกต่างของกราฟ pie**
2. **ให้นักเรียน plot กราฟ pie สำหรับ num\_children และ­แสดงกราฟในส่วนของข้อมูลของ jeff และ jose ถูก explode ออกมา**

จากตัวอย่างการ plot กราฟ ใน worksheet ที่แล้ว ในกรณีที่เรา plot กราฟแล้วการกำหนด label ในแกน x และแกน y จะเป็นค่า default ซึ่งอาจจะมีรายละเอียดหรือไม่มี ในตัวอย่างด้านล่างเราทำการ plot กราฟแบบ histogram ได้ดังภาพด้านล่าง กราฟนี้ขาดรายละเอียดทั้งในส่วนของชื่อกราฟ label ในแกน x และ y เราสามารถกำหนดรายละเอียดดังกล่าวได้โดยใช้วิธีการดังต่อไปนี้



ทำการกำหนดตัวแปรขึ้นมาหนึ่งตัว เพื่อรับค่าจากคำสั่ง plot ในตัวอย่างด้านล่างเราใช้ตัวแปรชื่อ ax (สามารถกำหนดชื่อตัวแปรเป็นชื่ออื่นได้) หลังจากนั้นใช้ฟังก์ชัน .set( ) เพื่อกำหนด title, lable ของแกน x และ y เช่น ax.set(title = 'Histogram Graph', xlable = 'Range of Age', ylable = 'Number of People ') เช่น



1. **ในกราฟ pie ด้านบนให้นักเรียนกำหนด title ของกราฟเป็น Pets Owned by Each Person' และกำหนด label ในแกน y เป็น Percentage of Pets'**
2. **ในแบบฝึกหัดนี้ให้นักเรียนทำการโหลดข้อมูลจากไฟล์ scottish\_hills.csv (ดาวน์โหลดได้จาก portal ของอาจารย์นะครับ ข้อมูลในไฟล์นี้ประกอบด้วยชื่อภูเขา ความสูง เส้นรุ้ง (latitude) เส้นแวง (longitude) ของภูเขาแต่ละลูกในประเทศสก๊อตแลนด์)**

****

**และทำการ plot กราฟแบบ histogram สำหรับข้อมูล height (ความสูงของภูเขา) โดยกำหนดค่า bin เป็น 900,950,1000,1050,1100,1150,1200,1250,1300,1350 และค่า rwidth = 0.9 หากนักเรียนทำได้ถูกต้องจะได้กราฟดังต่อไปนี้**

****

1. **ให้นักเรียนกำหนด title, label ในแกน x และ y เพื่อให้ได้กราฟตามภาพด้านล่าง**

****

ต่อเนื่องจากข้อมูลในไฟล์ scottish\_hills.csv เนื่องจากมีข้อมูลเส้นรุ้ง (latitude) เราสามารถ plot ตำแหน่งของภูเขาใน Scotland ได้โดยใช้การ plot แบบ scatter จะสังเกตเห็นได้ว่า label แกน x และ y จะเป็นชื่อ column ข้อมูลที่เรานำมา plot เช่น plot ความสูงกับเส้นรุ้ง



1. **ให้นักเรียนเพิ่ม title ของกราฟโดยกำหนดเป็น Mountains in Scotland และเปลี่ยน label แกน y เป็น Mountain Height**

ในการ plot กราฟเราสามารถกำหนดเงื่อนไข เพื่อให้ทำการ plot เฉพาะข้อมูลที่ต้องการ เช่นหากเราต้องการplot เฉพาะภูเขาที่มีความสูงมากกว่า 1,200 เมตร เราสามารถกำหนดเงื่อนไขได้ดังนี้

ax = sh[sh.Height > 1200].plot(kind='scatter', x='Height',y='Latitude',figsize=(10,10))

โดย syntax คือ ใน dataframe ที่เราต้องการ (ในกรณีนี้คือ sh) เราใส่เงื่อนไขภายใต้ [ ] ว่า sh.Height > 1200 คือเป็นการบอกว่าข้อมูล sh ใน column ความสูง เลือกเอาเฉพาะที่มากกว่า 1,200

1. **ให้นักเรียนลองกำหนดความสูงน้อยกว่า 1,100**
2. **ให้นักเรียนลองกำหนดเส้นรุ้งต่ำกว่า 57**

ในกรณีที่เราต้องการกำหนดเป็นช่วงของข้อมูล เช่น มากกว่า 900 และน้อยกว่า 1100 เราสามารถทำได้โดยการกำหนดให้ sh[sh.Height > 900] และนำไปเก็บไว้ที่ตัวแปรประเภท dataframe อีกตัวหนึ่งก่อน เช่น

sh1 = sh[sh.Height > 900]

และทำอีกแบบเดียวกันอีกครั้งสำหรับเงื่อนไข < 1100 โดยเก็บไว้ตัวแปร dataframe ตัวที่สอง เช่น sh2 = sh1[sh1.Height < 1100] หลังจากนั้นทำการ plot กราฟจากข้อมูล dataframe ตัวที่สอง

1. **ให้นักเรียนลอง plot กราฟโดยกำหนดความสูงระหว่าง 1200 ถึง 1300 หากทำได้ถูกต้องจะได้กราฟแสดงดังภาพต่อไปนี้**

****

1. **ให้นักเรียน plot กราฟโดยกำหนดเส้นรุ้ง (Latitude) ระหว่าง 57 และ 58**

เนื่องจากข้อมูลนี้มีค่า latitude และ longitude เราสามารถ plot ค่าเหล่านี้ลงบนแผนที่จริงได้ โดยใช้ package cartopy อันดับแรกให้นักเรียนทำการติดตั้ง โดยใช้คำสั่งต่อไปนี้ที่ command prompt

conda install -c conda-forge cartopy

และให้ใช้ python code ด้านล่าง เพื่อทำการ plot

import cartopy.crs **as** ccrs

from cartopy.mpl.ticker import LongitudeFormatter, LatitudeFormatter

import cartopy.feature **as** cfeature

plt**.**figure(figsize**=**(20,10))

ax **=** plt**.**axes(projection**=**ccrs**.**Mercator())

ax**.**coastlines('10m')

ax**.**xaxis**.**set\_visible(True)

ax**.**yaxis**.**set\_visible(True)

ax**.**set\_yticks([56,57,58,59], crs**=**ccrs**.**PlateCarree())

ax**.**set\_xticks([**-**8, **-**6, **-**4, **-**2], crs**=**ccrs**.**PlateCarree())

lon\_formatter **=** LongitudeFormatter(zero\_direction\_label**=**True)

lat\_formatter **=** LatitudeFormatter()

ax**.**xaxis**.**set\_major\_formatter(lon\_formatter)

ax**.**yaxis**.**set\_major\_formatter(lat\_formatter)

ax**.**set\_extent([**-**8, **-**1.5, 55.3, 59])

plt**.**scatter(sh['Longitude'],sh['Latitude'],

 color**=**'red', marker**=**'^', transform**=**ccrs**.**PlateCarree())

plt.show()

#plt**.**savefig("munros.png")

หากทำได้ถูกต้องจะได้กราฟบนแผนที่ดังต่อไปนี้



ให้นักเรียนโหลดข้อมูลจากไฟล์ NGSA\_field\_pH.csv ซึ่งประกอบด้วยข้อมูลค่า pH ในดิน TOS (Top of Soil) คือชั้นดินด้านบน 10 cm และ BOS (Bottom of Soil) คือชั้นดินที่ต่ำกว่า 10 cm ในแต่ละพื้นที่

1. **ให้นักเรียนทำการ plot กราฟ scatter สำหรับ TOS\_field\_pH และกราฟ scatter สำหรับ BOS\_field\_pH**
	1. **นอกจากข้อมูลภาพรวม ให้ลองกำหนดค่า pH ระหว่าง 7 – 8**
	2. **ค่า latitude ระหว่าง -26 ถึง -30**
	3. **Plot ค่าลงบนแผนที่ (ให้นักเรียนหาค่า latitude และ longitude ที่สูงสุดและต่ำสุดเพื่อกำหนดค่าในคำสั่ง python)**