

แบบฝึกหัด pandas ครั้งที่ 2

อ.ดร.เสกขวิทย์ เกติผล

1. โหลดไฟล์ที่ชื่อ olympics_cleaned.csv ขึ้นมาใน DataFrame ที่ชื่อ medals
2. โหลดไฟล์ชื่อ drinks.csv ขึ้นมาใน DataFrame ชื่อ drinks
3. ใน drinks มีคอลัมน์ไหนที่ไม่มีข้อมูลบ้าง และไม่มีข้อมูลกี่ตำแหน่งในแต่ละคอลัมน์
4. ลองเช็คชื่อประเทศของตาราง medals ใดที่ไม่มีอยู่ในตาราง drinks
 1. ลองใช้คำสั่ง isin() ในการเช็คว่ามีประเทศใดใน medals ไม่อยู่ใน drinks บ้าง
5. แสดงชื่อประเทศของตาราง drinks ที่ไม่มีอยู่ในตาราง medals
 1. นิสิตจะเห็นความแตกต่างของชื่อประเทศในสองตารางนี้ว่าเรียกชื่อประเทศไม่เหมือนกัน ตอนทำงานจริงนั้นต้องปรับชื่อให้ตรงกัน (เช่น USA กับ United States) แต่ในการบ้านนี้ไม่ต้องทำ
6. นำคอลัมน์ที่ชื่อ continent ของ drinks ไปต่อกับตาราง medals ตามชื่อประเทศของ medals
 1. การที่จะ join ตารางได้นั้นต้องมีคอลัมน์ที่ตรงกันในทั้งสองตาราง ในที่นี้ให้ใช้ชื่อประเทศ
 2. อย่าลืมว่าการ join ข้อมูลนั้นเราต้องการคงข้อมูลของประเทศที่มีในตาราง medals ไว้ ตารางที่ join แล้วจะมีจำนวนแถวเท่ากับตาราง medals
7. ในตาราง medals จงแสดงชื่อประเทศที่ยังไม่มีข้อมูลในคอลัมน์ continent
8. นิสิตจะเห็นแนวทางการแก้ไขข้อมูลหรือรวมประเทศบางประเทศเข้าด้วยกัน แต่ในการบ้านนี้ยังไม่ต้องทำทั้งหมด แค่เติมข้อมูลทวีปที่ยังขาดหายไป ในบางประเทศดังนี้ Australasia: OC, Canada/United States/Mexico: NA, Cuba/Jamaica: SA, Czech/Yugoslavia/Germany(ทั้งหมด)/Great Britain/Russia(ทั้งหมด)/Soviet: EU ประเทศที่เหลือกำหนดให้เป็น 'OT' ไป
 1. ตารางนี้จะเป็นตารางแบบ wide ที่สมบูรณ์ ควรเก็บไว้เพื่อใช้ในการคำนวณ
9. ตั้งให้ ชื่อทวีปเป็น index ชั้นนอกและชื่อประเทศเป็น index ชั้นใน
10. เรียงลำดับชื่อทวีปจากมากไปน้อยและในทวีปเรียงลำดับชื่อประเทศจากน้อยไปมาก
11. จัดคอลัมน์ให้เป็นสองระดับ โดยระดับแรกคือฤดู และระดับที่สองคือจำนวนเกมส์และเหรียญที่ได้ (ดูตัวอย่างจากเอกสารการสอน) ดังนั้นตอนนี้ข้อมูลจะมีสองระดับทั้ง index (ทวีปและประเทศ) และ column (ฤดูและจำนวนเกมและเหรียญ)
12. ย้ายระดับของฤดูมาเป็น index ชั้นในสุด
13. ปรับตารางให้เป็นแบบ long โดยหนึ่งแถวจะมีคอลัมน์คือ
ทวีป ประเทศ ฤดูแข่งขัน จำนวนเกม ประเภทเหรียญ จำนวนเหรียญ

| | continent | Country | Season | Game | Medal | Count |
|-----|-----------|---------|--------|------|--------|-------|
| 12 | EU | Austria | Summer | 26 | Bronze | 35 |
| 13 | EU | Austria | Winter | 22 | Bronze | 81 |
| 304 | EU | Austria | Summer | 26 | Gold | 18 |
| 305 | EU | Austria | Winter | 22 | Gold | 59 |
| 596 | EU | Austria | Summer | 26 | Silver | 33 |
| 597 | EU | Austria | Winter | 22 | Silver | 78 |

ตัวอย่างตารางผลลัพธ์ของข้อ 13 แสดงข้อมูลประเทศ Austria

ตั้งแต่นี้เป็นต้นไปสามารถใช้การคำนวณต่อจากตารางแบบ wide หรือ long ที่ได้ทำไว้แล้ว

14. แสดงจำนวนเหรียญทอง โอลิมปิกฤดูร้อนรวมของแต่ละทวีปจากมากไปน้อย

| continent | |
|-----------|------|
| EU | 2682 |
| NA | 1048 |
| AS | 531 |
| OC | 184 |
| SA | 141 |
| OT | 120 |
| AF | 103 |

15. นับว่าแต่ละทวีปมีกี่ประเทศ (ลองใช้คำสั่ง size() ต่อจากการ groupby())

| continent | |
|--------------|----|
| AF | 26 |
| AS | 30 |
| EU | 46 |
| NA | 3 |
| OC | 4 |
| OT | 24 |
| SA | 13 |
| dtype: int64 | |

16. แสดงตารางรวมเหรียญฤดูร้อนแต่ละประเภทของทุกประเทศในทวีปเดียวกัน แสดงผลแยกตามทวีป

| SummerGold SummerSilver SummerBronze | | | |
|--------------------------------------|------|------|------|
| continent | | | |
| AF | 103 | 112 | 132 |
| AS | 531 | 485 | 533 |
| EU | 2682 | 2830 | 3094 |
| NA | 1048 | 877 | 815 |
| OC | 184 | 175 | 221 |
| OT | 120 | 123 | 131 |
| SA | 141 | 173 | 204 |

17. แสดงตารางสรุปรวมเหรียญแต่ละประเภทของทุกประเทศแยกตามฤดูแข่งขัน

| Season | Summer | Winter |
|--------|--------|--------|
| Medal | | |
| Bronze | 5130 | 948 |
| Gold | 4809 | 959 |
| Silver | 4775 | 958 |

18. หาค่าเฉลี่ยของจำนวนเหรียญทองฤดูร้อนต่อประเทศที่เข้าร่วมแข่งขัน แยกตามทวีป

| continent | |
|-----------|------------|
| AF | 3.961538 |
| AS | 17.700000 |
| EU | 58.304348 |
| NA | 349.333333 |
| OC | 46.000000 |
| OT | 5.000000 |
| SA | 10.846154 |

19. หาจำนวนประเทศที่เคยเข้าร่วมการแข่งขันโอลิมปิก แยกตามทวีปและฤดูแข่งขัน แสดงในตารางเดียวกัน

| Season | Summer | Winter |
|-------------------------|--------|--------|
| Increase code font size | | |
| continent | | |
| AF | 26 | 11 |
| AS | 30 | 17 |
| EU | 46 | 45 |
| NA | 3 | 3 |
| OC | 4 | 3 |
| OT | 24 | 13 |
| SA | 13 | 9 |

20. จำนวนเหรียญรวม(ฤดูร้อนกับฤดูหนาว) แยกตามทวีปและฤดูแข่งขัน

| Medal | Bronze | Gold | Silver |
|-----------|--------|------|--------|
| continent | | | |
| AF | 132 | 103 | 112 |
| AS | 584 | 581 | 545 |
| EU | 3823 | 3415 | 3550 |
| NA | 951 | 1206 | 1035 |
| OC | 225 | 189 | 179 |
| OT | 159 | 133 | 139 |
| SA | 204 | 141 | 173 |