

## แบบฝึกหัด pandas ครั้งที่ 2

อ.ดร.เสกขวิทย์ เกติผล

1. โหลดไฟล์ชื่อ olympic.csv ขึ้นมาใน DataFrame ที่ชื่อ medals เปลี่ยนชื่อคอลัมน์ของ medals ให้สื่อความหมาย (ตามใจนิสิต)
2. โหลดไฟล์ชื่อ drinks.csv ขึ้นมาใน DF ชื่อ drinks
3. drinks มีข้อมูลที่คอลัมน์ไหนที่เป็น NaN บ้าง และแต่ NaN ก็ตัวในแต่ละคอลัมน์
4. ลองเช็คชื่อประเทศของตาราง medals ใดที่ไม่มีอยู่ในตาราง drinks
  1. การที่จะ join ตารางได้นั้นต้องมีคอลัมน์ที่มีข้อมูลเหมือนกัน ในทั้งสองตาราง ในที่นี้ให้ใช้ชื่อประเทศ
  2. สกัดเอาชื่อประเทศจากตาราง medals ออกมาโดยใช้คำสั่ง str.extract("(.)s\\") มาสร้างเป็นคอลัมน์ใหม่
  3. ลองใช้คำสั่ง isin() ในการเช็คว่ามีประเทศใดใน medals ไม่อยู่ใน drinks บ้าง
5. แสดงชื่อประเทศของตาราง drinks ที่ไม่มีอยู่ในตาราง medals
  1. นิสิตจะเห็นความแตกต่างของชื่อประเทศในสองตารางนี้ว่าเรียกชื่อประเทศไม่เหมือนกัน ตอนทำงานจริงนั้นต้องปรับชื่อให้ตรงกัน (เช่น USA กับ United States) แต่ในการบ้านนี้ไม่ต้องทำ
6. นำคอลัมน์ชื่อ continent ของ drinks ไปต่อกับตาราง medals ตามชื่อประเทศของ medals
7. ในตาราง medals จงแสดงชื่อประเทศที่ยังไม่มีข้อมูลในคอลัมน์ continent
8. นิสิตจะเห็นแนวทางการแก้ไขข้อมูลหรือรวมประเทศบางประเทศเข้าด้วยกัน แต่ในการบ้านนี้ยังไม่需要做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 (ฤดูและจำนวนเกมส์และเหรียญ)
12. ย้ายระดับของฤดูมาเป็น index ชั้นในสุด
13. ปรับตารางให้เป็นแบบ long โดยหนึ่งแถวจะมีคอลัมน์คือ  
ทวีป ประเทศ ฤดูแข่งขัน จำนวนเกมส์ 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. แสดงจำนวนเหรียญทองของ โอลิมปิกฤดูร้อนของแต่ละทวีปจากมากไปน้อย
15. นับว่าแต่ละทวีปมีกี่ประเทศ (ลองใช้คำสั่ง size() ต่อจากการ groupby() )
16. แสดงตารางรวมเหรียญฤดูร้อนแต่ละประเภทของทุกประเทศในทวีปเดียวกัน แสดงผลแยกตามทวีป
17. แสดงตารางสรุปรวมเหรียญแต่ละประเภทของทุกประเทศแยกตามฤดูแข่งขัน
18. หาค่าเฉลี่ยของจำนวนเหรียญทองฤดูร้อนต่อประเทศที่เข้าร่วมแข่งขัน แยกตามทวีป
19. หาจำนวนประเทศที่เคยเข้าร่วมการแข่งขัน โอลิมปิก แยกตามทวีปและฤดูแข่งขัน แสดงในตารางเดียวกัน
20. หาค่าเฉลี่ยของจำนวนเหรียญทองต่อประเทศที่เข้าร่วมแข่งขัน แยกตามทวีปและฤดูแข่งขัน