

# ข้อสอบปฏิบัติการวิชา 01418323

## Introduction to Data Science

ภาคปลาย ปีการศึกษา 2562

อ.ดร.เสกฐวิทย์ เกิดผล

### ข้อกำหนดในการสอบ

1. เข้าสอบให้ตรงเวลา ถ้าเข้าสายกว่า 15 นาทีหลังเริ่มสอบจะหมดสิทธิ์สอบ
2. ให้นำเอกสารชี้ที่แล็บเข้าได้ สามารถบันทึกโน้ตในชีทแล็บได้
3. ไม่ให้รับส่งไฟล์อิเล็กทรอนิกส์ใดๆ ยกเว้นไฟล์ที่จะให้ดาวน์โหลดก่อนเริ่มสอบเท่านั้น

### การบันทึกไฟล์คำตอบ

1. ให้ตั้งชื่อไฟล์คำตอบด้วยรหัสนิสิต
2. ให้บันทึกไฟล์ไว้ที่ไดรฟ์ D: เพื่อป้องกันไม่ให้ไฟล์หาย

### การตอบคำถาม

1. ให้ใส่ชื่อและรหัสนิสิตเป็นคำตอบข้อ 0
2. การตอบคำถามแต่ละข้อให้ใส่คอมเมนต์เบอร์ข้อไว้ก่อนเริ่มตอบ
3. ต้องแสดงผลทุกครั้งที่จะตอบ
4. ให้แสดงผลตามที่ถาม เช่นถามจำนวนก็ตอบเป็นตัวเลข ถามว่ามีอะไรบ้างก็ต้องตอบเป็นชื่อ
5. ให้นำไฟล์ข้อมูลมาใส่ไว้ใน directory เดียวกันกับไฟล์คำตอบ
6. นิสิตควรศึกษาข้อมูลในไฟล์ที่จะนำมาวิเคราะห์ก่อนจะเริ่มตอบคำถาม

### การส่งคำตอบ

1. เมื่อหมดเวลาจะเปิดระบบเน็ตเวิร์คให้ส่งไฟล์คำตอบไปยังเซิร์ฟเวอร์ ให้นิสิตส่งไฟล์คำตอบและออกจากห้องสอบทันที
2. ห้ามไม่ให้ส่งไฟล์ไปยังระบบเซิร์ฟเวอร์หลังจากการสอบเสร็จสิ้น นิสิตจะได้ 0 คะแนนทันที และอาจมีการตั้งกรรมการพิจารณาการทุจริตต่อไป

### คำถาม

1. ลองเปิดไฟล์ air\_crash.csv ด้วย Excel เพื่อพิจารณาข้อมูลก่อน จากนั้นโหลดไฟล์ air\_crash.csv เข้าไปใน Jupyter Notebook แล้วแสดงข้อมูล 5 บรรทัดแรก
2. ชุดข้อมูลนี้มีจำนวนกี่แถว และกี่คอลัมน์
3. คอลัมน์ 'Location' มีจำนวนข้อมูลกี่ค่า ไม่นับค่าว่าง
4. แสดงข้อมูลแถวเบอร์ 25-30
5. แสดงข้อมูลของ 5 แถวแรกเฉพาะคอลัมน์ Date, Operator และ AC Type
6. แสดงอุบัติเหตุที่มีคนตายรวม (Fatalities) เกิน 500 คน
7. อุบัติเหตุเครื่องบินตกที่เกิดขึ้นในประเทศไทยมีกี่ครั้ง (ใช้คอลัมน์ Location)

8. แสดงข้อมูลเครื่องบินตกที่เกิดขึ้นในประเทศไทย \*หมายเหตุ: ข้อนี้ยากกว่าที่คิด พิจารณาคำถามข้อ 3 ประกอบด้วย
9. แสดงอุบัติเหตุที่เกิดกับเครื่องบินโดยสารที่มีผู้โดยสาร (Aboard Passengers) มากกว่า 100 คน และจำนวนลูกเรือที่เสียชีวิต (Fatalities Crew) สูงกว่าจำนวนผู้โดยสารที่เสียชีวิต (Fatalities Passengers)
10. จำนวนคนตายทั้งหมดจากอุบัติเหตุเครื่องบินตกคือเท่าไร
11. แสดงอุบัติเหตุที่มีคนตายรวมมากที่สุด
12. สร้างคอลัมน์ใหม่ชื่อ Fatalities Percent แสดงสัดส่วนจำนวนคนตายรวมต่อจำนวนคนในเครื่องบิน (Aboard) แสดง 5 แถวแรกของตารางใหม่
13. แสดงข้อมูล 5 แถวแรกของอุบัติเหตุที่มีคนตายมากที่สุดเรียงจากมากไปน้อย
14. แสดงข้อมูล 5 แถวแรกของอุบัติเหตุที่มี Fatalities Percent ที่สุดเรียงจากมากไปน้อย ถ้าเท่ากันให้เรียงจากจำนวนคนในเครื่องบินจากมากไปน้อย
15. คัดเลือก 7 สายการบิน (Operator) ดังนี้ออกมา All Nippon Airways, EVA Air, Scandinavian Airlines (SAS), Singapore Airlines, Swissair, Thai Airways, United Air Lines (ชื่อทั้งหมดอยู่ในไฟล์ star\_alliance\_7.csv) ตั้งชื่อว่า star\_7 แล้วแสดงข้อมูลในตารางนี้
16. สรุปรวมจำนวนอุบัติเหตุที่เกิดขึ้นของแต่ละสายการบินใน star\_7 \*ระวัง: เช็คว่ามี 7 สายการบิน
17. นำอุบัติเหตุของการบินไทย (Thai Airways) มาเรียงลำดับจากล่าสุดย้อนกลับ
18. แสดงข้อมูลของ 7 สายการบินเรียงตาม Operator จากน้อยไปมาก และในแต่ละ Operator เรียงจากอุบัติเหตุที่เกิดขึ้นก่อนมาครั้งล่าสุด
19. คัดเลือกอุบัติเหตุของสายการบิน Thai Airways ในทศวรรษที่ 90 (ปี 1990–1999) ออกมา
20. ในตาราง star\_7 สกัดค่าแรกของคอลัมน์ 'AC Type' ออกมาเป็นชื่อบริษัทที่สร้างเครื่องบิน จากนั้นนับจำนวนเครื่องบินที่ตกแยกตามสายการบินและบริษัทเครื่องบิน (ไม่ต้องสนใจข้อมูลที่อาจผิดพลาด)
21. นับจำนวนคนตายของ 7 สายการบินแยกตามทศวรรษ (1930, 1940, 1950, ... , 2000) และสายการบิน
22. เปลี่ยนคอลัมน์ Operator ของ star\_7 ให้กลายเป็น category แบบเรียงลำดับ โดยลำดับคือลำดับตามตัวอักษร (ตามข้อ 15)
23. แสดงแผนภาพ Histogram และ Rug plot ของ Fatalities Percent ของอุบัติเหตุจาก 7 สายการบิน
24. แสดงกราฟแท่งแสดงจำนวนคนตายรวมของ 7 สายการบิน
25. จากตาราง star\_7 แสดงแผนภาพแบบ scatterplot แสดงจำนวนครั้งที่เครื่องบินตกเทียบกับจำนวนคนตายเฉลี่ยต่ออุบัติเหตุหนึ่งครั้ง และแยกสีตามสายการบิน ขนาดของจุดแปรผันตามจำนวนคนตายรวมของสายการบิน