

ข้อสอบปฏิบัติการวิชา 01418565

Data Science with Big Data Analytic

ภาคต้น ปีการศึกษา 2561

อ.ดร.เสกฐวิทย์ เกิดผล

ข้อกำหนดในการสอบ

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

การบันทึกไฟล์คำตอบ

1. ให้ตั้งชื่อไฟล์ด้วยรหัสนิสิต
2. นิสิตสามารถทำไฟล์คำตอบเป็นไฟล์ .py (ไพธอน) หรือ .ipynb (Jupyter Notebook) ก็ได้
3. ให้บันทึกไฟล์ไว้ในไดเรกทอรี D: เพื่อป้องกันไม่ให้ไฟล์หาย

การตอบคำถาม

1. เริ่มต้นให้นิสิต print() ชื่อและรหัสนิสิตเป็นบรรทัดแรก
2. การตอบคำถามแต่ละข้อให้ใส่คอมเมนต์เบอร์ข้อไว้ก่อนเริ่มตอบ
3. ต้องแสดงผลทุกครั้งที่จะตอบ ผ่านคำสั่ง print() หรือเรียกชื่อตัวแปรที่เก็บคำตอบ
4. ให้แสดงผลตามที่ถาม เช่นถามจำนวนก็ตอบเป็นตัวเลข ถามว่ามีอะไรบ้างก็ต้องตอบเป็นชื่อ
5. การโหลดไฟล์ให้นำไฟล์มาใส่ไว้ใน directory เดียวกันกับไฟล์คำตอบ
6. นิสิตควรศึกษาข้อมูลในไฟล์ที่จะนำมาวิเคราะห์ก่อนจะเริ่มตอบคำถาม

คำถาม

1. พิมพ์ชื่อและรหัสนิสิตไว้ที่บรรทัดแรก
2. โหลดไฟล์ชื่อ accidents.csv เข้ามาอยู่ใน DF จากนั้นตัดคอลัมน์รหัสจังหวัด รหัสโรงพยาบาล และการนำส่งทิ้งไป เปลี่ยนชื่อคอลัมน์ให้เหมาะสม แล้วแสดง 5 บรรทัดแรกของ DF
3. จำนวนอุบัติเหตุที่เกิดขึ้นทั้งหมดมีกี่ครั้ง
4. จำนวนอุบัติเหตุที่มีผลการรักษาเป็น “ทุเลา/หาย” มีกี่ครั้ง
5. ถนนที่เกิดอุบัติเหตุมีการจำแนกเป็นประเภทใดบ้าง (hint: ลองศึกษาการใช้ method unique())
6. อุบัติเหตุที่ส่งผลให้ต้องรักษาผู้ประสบอุบัติเหตุมากกว่า 7 วันมีกี่ครั้ง
7. อุบัติเหตุอย่างเบาที่ให้การรักษาผู้ประสบภัยน้อยกว่าหนึ่งวันก็ทุเลามีกี่ครั้ง
8. เรียงลำดับข้อมูลตามอายุผู้เกิดเหตุ โดยให้อายุมากมาก่อน เลือกแสดงแต่คอลัมน์ปี เพศ และอายุ
9. วันที่เกิดเหตุนั้นมีตั้งแต่ 27 (ธันวาคม) จนถึง 5 (มกราคม) แต่การตั้งค่าวันแบบนี้จะทำให้การเรียงลำดับวันผิดพลาด ให้แก้วันของเดือนธันวาคมให้เป็นค่าติดลบย้อนหลัง เช่น 31 จะเป็น -1 วันที่ 30 จะเป็น -2 ไปเรื่อยๆ วันของเดือนมกราคมให้คงเดิม (1-5) ส่วนวันที่เป็น 0 ให้ทิ้งแถวนั้นไป จากนั้นแสดงจำนวนแถวที่เหลือและค่าของวันที่เป็นไปได้ทั้งหมด
10. เรียงลำดับข้อมูลตามปี วัน และเวลาที่เกิดเหตุจากน้อยไปมาก ถ้าวันเวลาปีตรงกันให้เรียงจากอายุผู้ประสบอุบัติเหตุจากมากไปน้อย

11. * สร้างคอลัมน์ใหม่ที่เก็บค่า True เมื่อผู้ประสบอุบัติเหตุเสียชีวิต และ เก็บค่า False เมื่อผู้ประสบอุบัติเหตุหายจากการบาดเจ็บ (hint: สังเกตว่าคอลัมน์"ผลการรักษา"มีข้อมูลอะไรบ้าง)
12. [2] สรุปรวมจำนวนครั้งที่เกิดอุบัติเหตุแยกตามปี
13. [2] แสดงจำนวนคนตายแยกตามมาตรการรักษาความปลอดภัยและการดื่มสุรา
14. [2] บางปีนั้นบางจังหวัดไม่มีบันทึกการเกิดอุบัติเหตุเลย จงแสดงชื่อจังหวัดและจำนวนปีที่ไม่มีการเกิดอุบัติเหตุในจังหวัดนั้น
15. [2] จงแสดงชื่อจังหวัดที่เกิดอุบัติเหตุแต่ไม่มีคนตายในปี 52
16. [2] แสดงจำนวนจังหวัดที่เกิดอุบัติเหตุแต่ไม่มีคนตายในแต่ละปี (hint: ลองศึกษาการใช้ method any/all)
17. [2] แสดงจำนวนปีที่แต่ละจังหวัดมีอุบัติเหตุแต่ไม่มีคนตาย (ไม่ต้องแสดงจังหวัดที่มีคนตายทุกปี)
18. [2] แสดงจำนวนโรงพยาบาลในแต่ละจังหวัด
19. [2] แสดงจำนวนอุบัติเหตุตามประเภทรถผู้บาดเจ็บสูงสุด 5 อันดับแรก
20. [2] กรณีที่รถผู้บาดเจ็บเป็นจักรยานยนต์ ให้แสดงจำนวนรถคู่กรณีแต่ละประเภทเรียงจากมากไปน้อย
21. [2] แสดงอายุเฉลี่ยของผู้ประสบอุบัติเหตุแยกตามประเภทรถผู้บาดเจ็บ
22. [2] อายุของคนที่เกิดอุบัติเหตุบ่อยที่สุดคือเท่าไร
23. [2] แสดงอายุของคนที่เกิดอุบัติเหตุบ่อยที่สุดแยกตามปี
24. [2] แสดงอายุของผู้ที่เกิดอุบัติเหตุบ่อยที่สุด 5 ลำดับ แยกตามปี
25. [2] โหลดไฟล์ชื่อ provSize.csv เข้ามาอยู่ใน DF ชื่อ prov_area จัดการ clean ข้อมูลให้เรียบร้อย จากนั้น join ตารางนี้ด้วยชื่อจังหวัดเข้ากับตารางข้อมูลผู้เสียชีวิตแยกรายจังหวัดและรายปี แสดงตารางที่ join แล้ว
26. [2] แสดง 5 จังหวัดแรกที่มีคนตายต่อพื้นที่ที่สูงที่สุดในปี 58