

Microsoft Data Science Workshop

Workshop 1: Exploring Data in Excel

1. ให้สร้างเอกสารเปล่าจาก Excel และให้ทำการตั้งค่า Regional Format สำหรับการใช้งาน Excel บนเว็บโดยเลือกรูปแบบเป็น English (United States)
2. ให้ทำการนำเข้าข้อมูลจากไฟล์ English Exam.csv
3. ให้ทำการจัดรูปแบบข้อมูลทั้งหมดให้เป็นตาราง โดยมีแถวแรกเป็นส่วนหัวของตาราง
4. ให้ทำการตรวจสอบข้อมูลวันที่ เพื่อหาข้อมูลที่หายไปโดยเติมข้อมูลให้ถูกต้อง
5. ให้ทำการตรวจสอบข้อมูลของอายุ เพื่อหาข้อมูลที่หายไปโดยเติมข้อมูลที่เป็นค่าเฉลี่ยของอายุทั้งหมด
6. ที่คอลัมน์ Gender ให้ทำการกรองข้อมูล (Filter) เพื่อหาข้อมูลที่ผิดปกติ แล้วทำการแก้ไขให้ถูกต้อง
7. ให้ตรวจสอบข้อมูล Candidate ที่ซ้ำกันแล้วทำการลบข้อมูลออกให้เหลือเฉพาะหมายเลข Candidate ที่ไม่ซ้ำกัน
8. ให้เพิ่มคอลัมน์ Result เพื่อหาผลลัพธ์ของค่าเฉลี่ยผลการสอบของแต่ละ Candidate โดยแสดงผลในรูปแบบ %
9. ให้ใช้การจัดรูปแบบอย่างมีเงื่อนไข (Conditional Formatting) เพื่อเน้นผลสอบ (Result) ที่มียอดสูงสุด 10 อันดับแรกด้วยโทนสีเขียว (Green Fill with Dark Green Text)
10. สร้างแผนภูมิแบบ Box and Whisker เพื่อแสดงข้อมูลค่ามัธยฐาน (Median) ของข้อมูล Gender และ Age พร้อมกับแสดงค่าสูงสุดและต่ำสุด ย้ายแผนภูมิไปวางไว้ด้านล่างตารางข้อมูล

Workshop 2: Using Statistics in Excel

ในแบบฝึกหัดนี้ จะต้องใช้ XLMiner Analysis ToolPak สำหรับ add-in ของ excel บนเว็บ หรือ
เพิ่ม Analysis ToolPak สำหรับ add-in ของ excel บนคอมพิวเตอร์ (Desktop) เพื่อที่จะได้ทำตาม
คำสั่งด้านล่างนี้ได้อย่างสมบูรณ์

1. ที่เซลล์ L1 ให้แสดงผลของสถิติเชิงบรรยาย (descriptive statistics) สำหรับผลคะแนนของ Reading, Writing, Listening, Speaking
2. ให้ทำการสอบ t ใน Excel โดยใช้ t-Test เพื่อทดสอบสมมติฐานที่ค่าเฉลี่ยของประชากร 2 กลุ่มที่เท่ากัน ของข้อมูล Listening และ Speaking แสดงผลของการทดสอบไว้ที่เซลล์ L24
3. ให้วิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) ในคอลัมน์ Reading, Writing, Listening, Speaking โดยวิเคราะห์ความแปรปรวนแบบ 1 ตัวประกอบ (One-way analysis of variance) เพื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ย (Mean) ของประชากรหลายกลุ่มที่มีค่าเท่ากันทั้งหมด แสดงข้อมูลผลลัพธ์ไว้ที่เซลล์ L40
4. ให้ทำการสร้างแผนงานใหม่ชื่อ Age แล้วทำการคัดลอกคอลัมน์ Age มาวางไว้ที่เซลล์ A1 และทำการหาความถี่ของข้อมูล (Frequency) ช่วงอายุ 0-30, 31-40, 41-50 แสดงข้อมูลผลลัพธ์ไว้ที่เซลล์ D1