

ปฏิบัติการที่ 1

การเตรียมแฟ้มข้อมูลและการนำเข้าข้อมูลสู่เครื่องมือวิเคราะห์ข้อมูล

วัตถุประสงค์

- เพื่อให้สามารถจัดเตรียมข้อมูลให้อยู่ในรูปแบบที่เหมาะสมสำหรับการนำไปวิเคราะห์ด้วยเครื่องมือวิเคราะห์ข้อมูลทางวิทยาการข้อมูล
- เพื่อให้สามารถนำข้อมูลเข้าสู่เครื่องมือวิเคราะห์ข้อมูลเพื่อประมวลผลต่อไปได้

ชุดข้อมูลปฏิบัติการ

- ชุดข้อมูล Comic Characters (สำหรับการสาธิต)
- ชุดข้อมูล 2004 New Car and Truck (สำหรับการฝึกปฏิบัติการ)

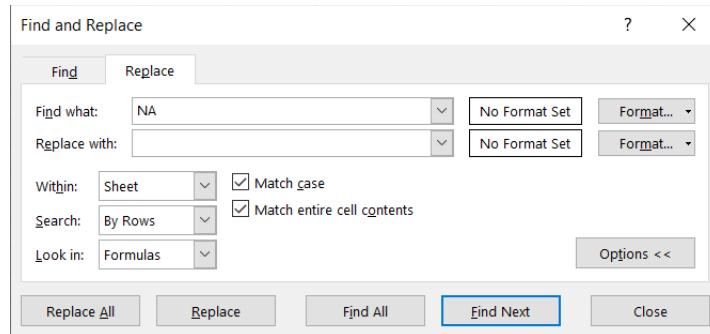
ขั้นตอนปฏิบัติการ

ขั้นตอนปฏิบัติการ มีดังนี้

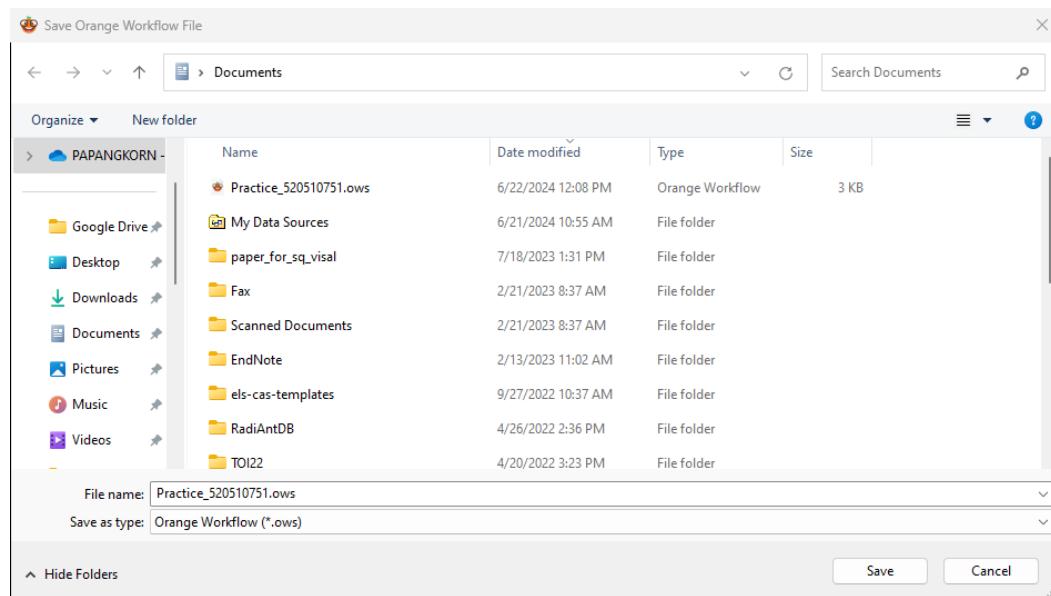
- เปิดแฟ้มข้อมูล comics.csv ด้วยโปรแกรม Microsoft Excel และข้อมูลแรกรอบบุชีคอลัมน์ (ตัวแปร) ทราบ ข้อมูลที่ 2 ถึง 23,273 เป็นระเบียนข้อมูลของตัวละครแต่ละตัว

	A	B	C	D	E	F	G	H	I	J	K
1	name	id	align	eye	hair	gender	gsm	alive	appearanc	first_appe	publisher
2	Spider-Man (Peter Parker)	Secret	Good	Hazel Eyes	Brown Hair	Male	NA	Living Chai	4043	Aug-62	marvel
3	Captain America (Steven Rogers)	Public	Good	Blue Eyes	White Hair	Male	NA	Living Chai	3360	Mar-41	marvel
4	Wolverine (James "Logan" Howl)	Public	Neutral	Blue Eyes	Black Hair	Male	NA	Living Chai	3061	Oct-74	marvel
5	Iron Man (Anthony "Tony" Stark)	Public	Good	Blue Eyes	Black Hair	Male	NA	Living Chai	2961	Mar-63	marvel
6	Thor (Thor Odinson)	No Dual	Good	Blue Eyes	Blond Hair	Male	NA	Living Chai	2258	Nov-50	marvel
7	Benjamin Grimm (Earth-616)	Public	Good	Blue Eyes	No Hair	Male	NA	Living Chai	2255	Nov-61	marvel
8	Reed Richards (Earth-616)	Public	Good	Brown Eye	Brown Hair	Male	NA	Living Chai	2072	Nov-61	marvel
9	Hulk (Robert Bruce Banner)	Public	Good	Brown Eye	Brown Hair	Male	NA	Living Chai	2017	May-62	marvel

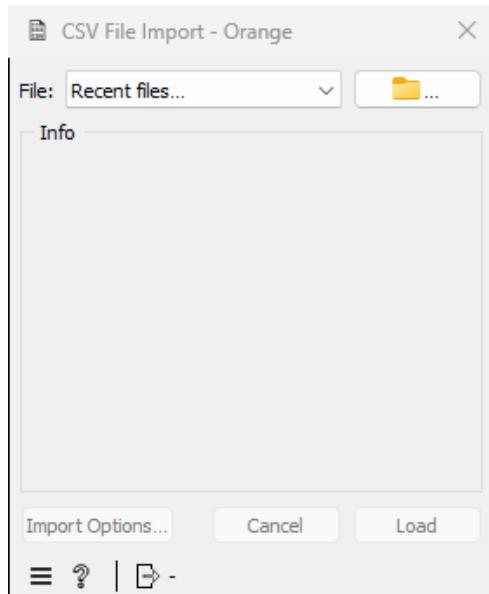
- ทำการแทนที่ข้อมูลในเซลล์ข้อมูลที่มีค่า NA ด้วยค่า ว่างเปล่า เนื่องจาก โปรแกรม Orange กำหนดให้เซลล์ข้อมูลได้ที่เป็นข้อมูลสูญหาย ต้องเป็นเซลล์ว่างเปล่า



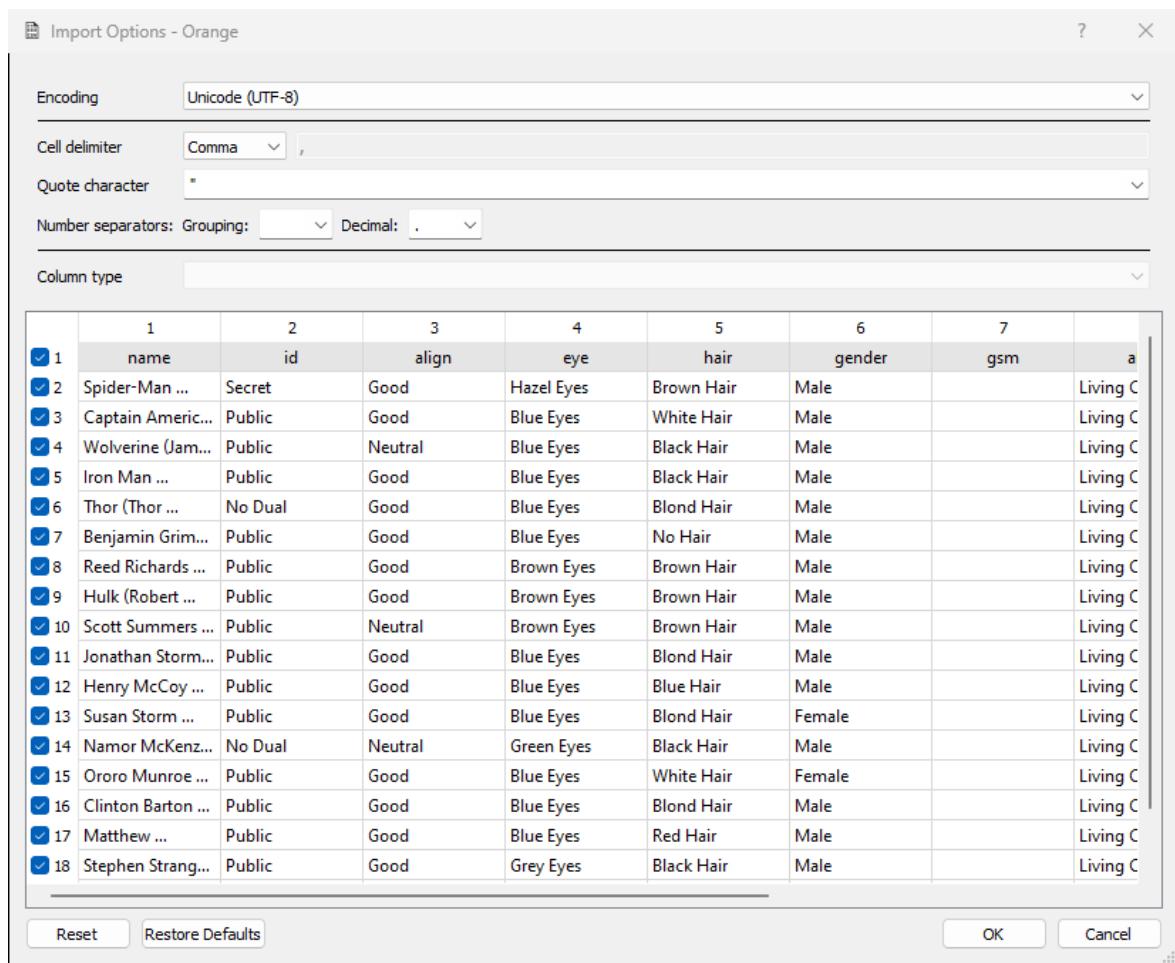
3. บันทึกแฟ้มข้อมูล comics.csv
4. เปิดโปรแกรม Orange
5. ทำการบันทึก workspace โดยไปที่เมนู File เลือก Save จากนั้นทำการตั้งชื่อไฟล์ในรูปแบบ Practice_01_id.ows โดยแทน id ด้วยรหัสนักศึกษา และกดปุ่ม Save



6. นำชุดข้อมูลจากแฟ้มข้อมูล comics.csv เข้าสู่โปรแกรม Orange โดยคลิกเลือกโมดูล CSV File Import จะปรากฏโมดูลใน workspace จากนั้นดับเบิลคลิกที่โมดูล จะปรากฏหน้าต่าง ดังรูป



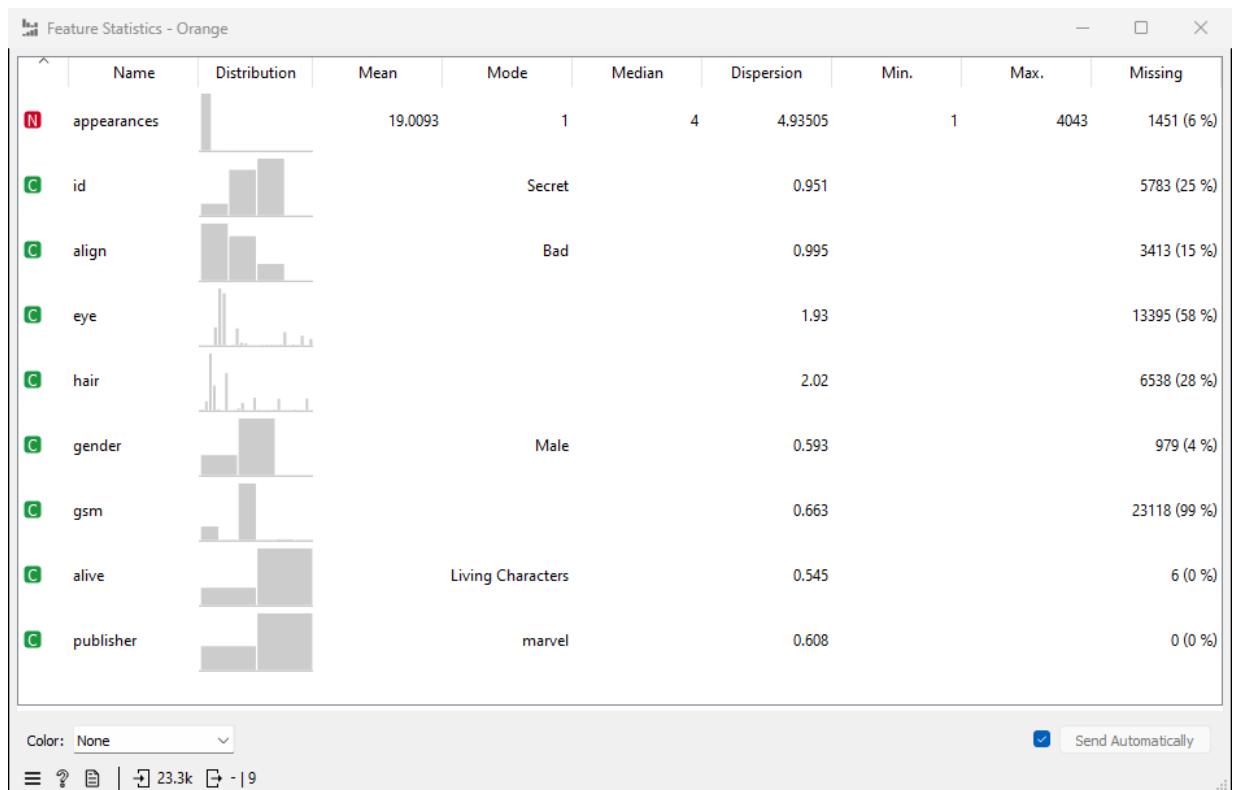
7. คลิกที่ปุ่มรูปไฟล์เดอร์ แล้วเลือกแฟ้มข้อมูล comics.csv จากนั้นจะปรากฏหน้าต่าง ดังรูป



8. ทำการตรวจสอบความถูกต้องให้เรียบร้อย และคลิกปุ่ม OK
9. คลิกเลือกโมดูล Feature Statistics จะปรากฏโมดูลใน workspace จากนั้นคลิกเชื่อมโมดูล CSV File Import จากด้าน output เข้าสู่โมดูล Feature Statistics ด้าน input ดังรูป



10. ดับเบิลคลิกที่โมดูล Feature Statistics จากปรากฏหน้าต่างแสดงผลลัพธ์ค่าสถิติเชิงพรรณของแต่ละตัวแปรในชุดข้อมูล ดังรูป



แบบฝึกปฏิบัติการ

ให้นักศึกษานำชุดข้อมูล 2004 New Car and Truck จากแฟ้มข้อมูล cars04.csv เข้าสู่โปรแกรม Orange และสร้างการทดลองโดยให้นำข้อมูลจากแฟ้มข้อมูล cars04.csv เข้าสู่การ workspace และใช้โมดูล Feature Statistics ในการคำนวณค่าสถิติเชิงพรรณนาของแต่ละตัวแปรในชุดข้อมูล

สิ่งที่ต้องส่งเป็นการบ้าน ทำการบันทึกไฟล์ workspace ของนักศึกษา โดยตั้งชื่อไฟล์ในรูปแบบ Lab_01_id.ows โดยแทน id ด้วยรหัสนักศึกษา ส่งผ่านเว็บไซต์ <http://hw.cs.science.cmu.ac.th>