**Video Game Sales**

ชื่อกลุ่ม…….

จัดทำโดย

นายชนกันต์ ไชยวงค์ 610610576

นาย กิตตินันท์ พรหมเผ่า 610610569

นางสาว ไอริณ แท่นประมูล 610110604

เสนอ

ผู้ช่วยศาสตราจารย์ เบญจมาศ ปัญญางาม

อาจารย์ ดร.สุธาสินี โทวุฒิกุล

รายงานฉบับนี้เป็นส่วนหนึ่งของวิชาหลักการเขียนโปรแกรมสำหรับวิเคราะห์ข้อมูล (Programming for Data Science : 229223)

ภาควิชาวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

*วัตถุประสงค์*

เพื่อนำข้อมูลที่มีจำนวนมากมาวิเคราะห์ จัดการ เพิ่มข้อมูลที่เป็นปัจจุบันและแสดงผลผ่านการใช้โปรแกรมวิเคราะห์ข้อมูล

*หลักการและเหตุผล*

การรู้จักข้อมูลและรู้วิธีการจัดการข้อมูล เป็นสิ่งที่สำคัญในการนำข้อมูลที่มีมาใช้ให้เกิดประโยชน์ในส่วนที่ต้องการ ซึ่งสามารถนำข้อมูลเหล่านั้นไปต่อยอดสำหรับการใช้งานในด้านอื่นๆต่อไป เช่น ด้านอุตสาหกรรม , การแพทย์ , ธุรกิจ , การศึกษา , การเมืองการปกครอง เป็นต้น นอกจากนี้แล้วความชำนาญในการใช้โปรแกรมก็มีส่วนสำคัญเช่นกัน เพื่อนำความสามารถในการทำงานของโปรแกรมมาใช้ในการจัดการข้อมูลที่มีได้อย่างเหมาะสม สุดท้าย คือ ข้อมูลเหล่านี้ต้องมีความเป็นปัจจุบัน ซึ่งจะสามารถนำไปใช้ประโยชน์สำหรับการทำนายหรือแก้ไขปัญหาในอนาคตที่มีความแม่นยำ

Video Game Sales คือฐานข้อมูลที่แสดงข้อมูลโดยทั่วไปของวิดีโอเกมส์ ยอดขายจากการจัดจำหน่าย และอันดับความนิยมของผู้ใช้ทั่วโลก จากการเก็บข้อมูลวิดีโอเกมส์ทั้งหมด 16,500 เกมส์ โดยข้อมูลเหล่านี้สามารถนำไปวิเคราะห์และต่อยอดได้ในด้านทางธุรกิจและอุตสาหกรรมสิ่งบันเทิง ว่าในแต่ละเกมส์ได้รับความนิยมและมียอดขายแตกต่างกันเท่าใด เกมส์ประเภทใดที่ได้รับความนิยมและมียอดขายมาก เกมส์ประเภทใดได้รับความนิยมและมียอดขายน้อย ในแต่ละทวีปมียอดขายของเกมส์แต่ละประเภทแตกต่างกันอย่างไร จากข้อมูลเหล่านี้ผู้ประกอบการสามารถวางแผนได้ว่า ในอนาคตเกมส์ประเภทใดที่จะได้รับความนิยม และควรมีการผลิตจัดจำหน่ายอย่างไร

*แหล่งที่มาของข้อมูล*

Gregory Smith. (2016). Video Game Sales (Online).

Available : https://www.kaggle.com/gregorut/videogamesales [2016 , October 26]

*คุณลักษณะของข้อมูล*

ประกอบไปด้วยข้อมูลชนิด String หรือ Categories ได้แก่ Name (ชื่อวิดีโอเกมส์) , Platform (เครื่องที่รองรับเกมส์) , Genre (ประเภทของเกมส์) , Publisher (ผู้จัดจำหน่าย)

ข้อมูลชนิด Integerได้แก่ Rank (อันดับของวิดีโอเกมส์จากยอดขายทั้งหมด) , Year (ปีที่จัดจำหน่าย)

ข้อมูลชนิด Floating Point ได้แก่NA\_Sales (ยอดขายในทวีปอเมริกาเหนือ) , EU\_Sales (ยอดขายในทวีปยุโรป) , JP\_Sales (ยอดขายในประเทศญี่ปุ่น) , Other\_Sales (ยอดขายจากสถานที่อื่น) , Global\_Sales (ยอดขายทั่วโลก) โดยยอดขายดังกล่าวทั้งหมดจะมีค่าเป็นหลักล้านดอลล่าสหรัฐ

*ตัวอย่าง field ภายใน Dataset*

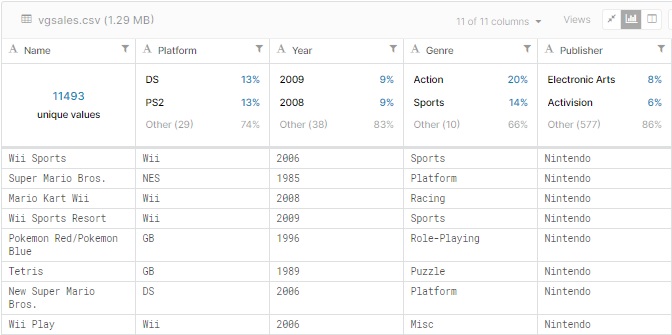


Figure (1) แสดงตัวอย่าง Field : Name , Platform , Year , Genre และ Publisher

*การทำงานในโปรแกรม*

ภายในโปรแกรมจะประกอบไปด้วยฟังค์ชันหลัก และฟังค์ชันสำหรับการใช้งานซึ่ง

ประกอบด้วย 6 ตัวเลือก มีดังนี้

*1.ฟังค์ชันหลัก*

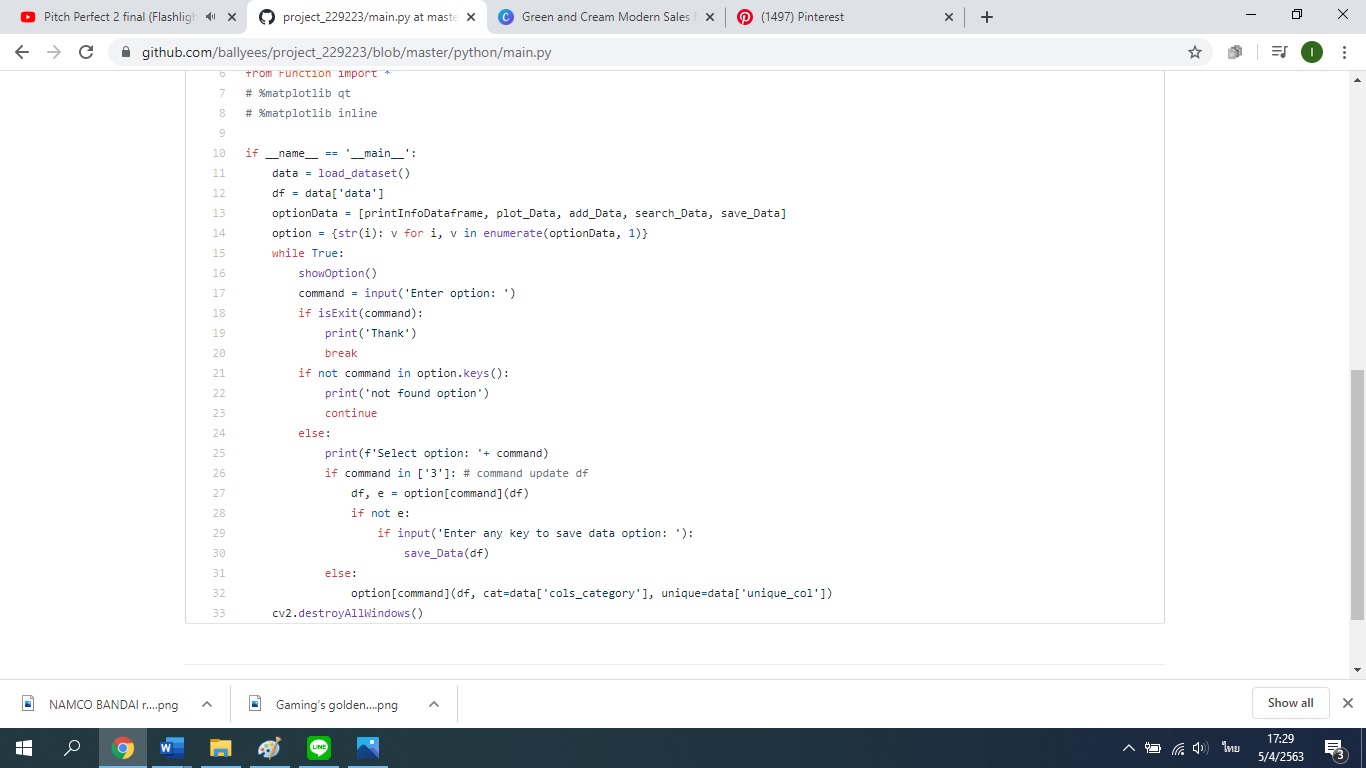


Figure (1)

*2.ฟังค์ชันสำหรับการใช้งาน*

*2.1 Show informations of data*

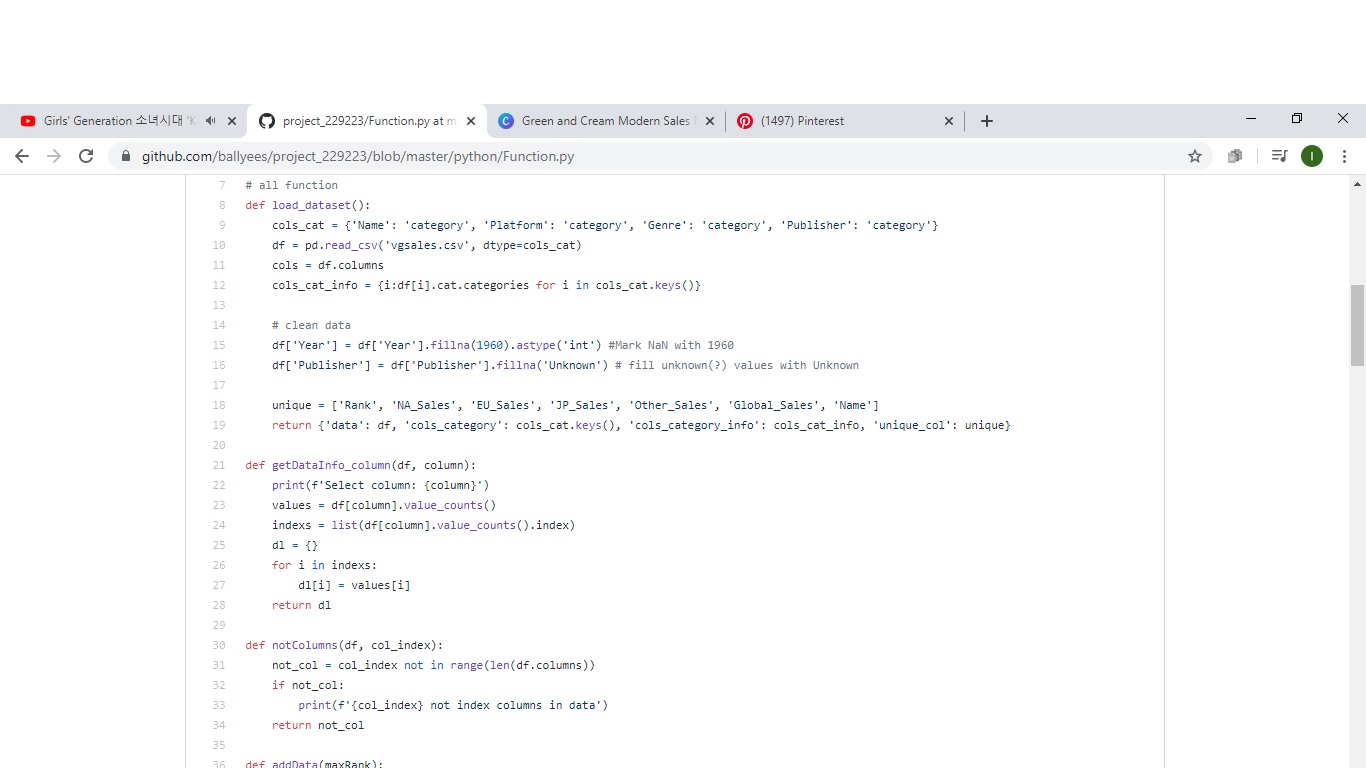


Figure (2)

*2.2 Plot Data (Histogram)*

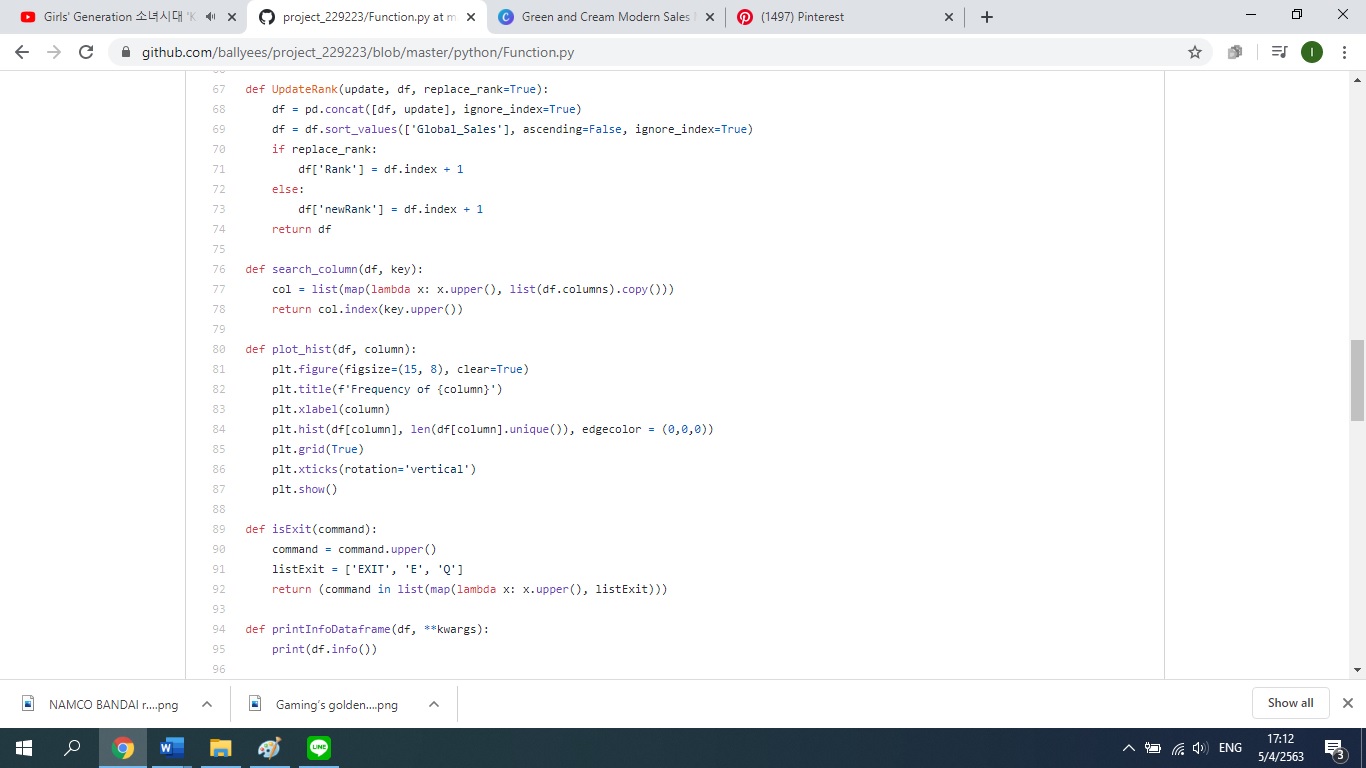


Figure (3.1)

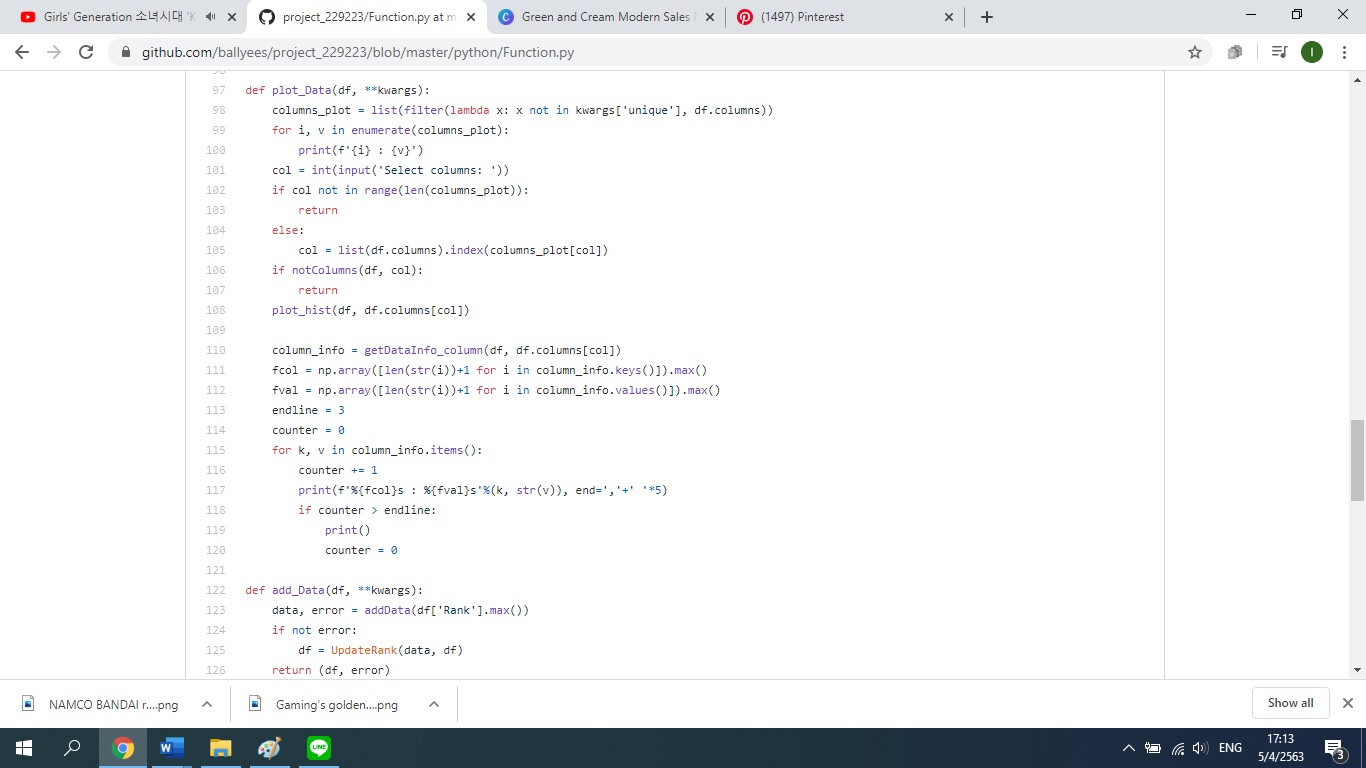


Figure (3.2)

*2.3 Add Data*

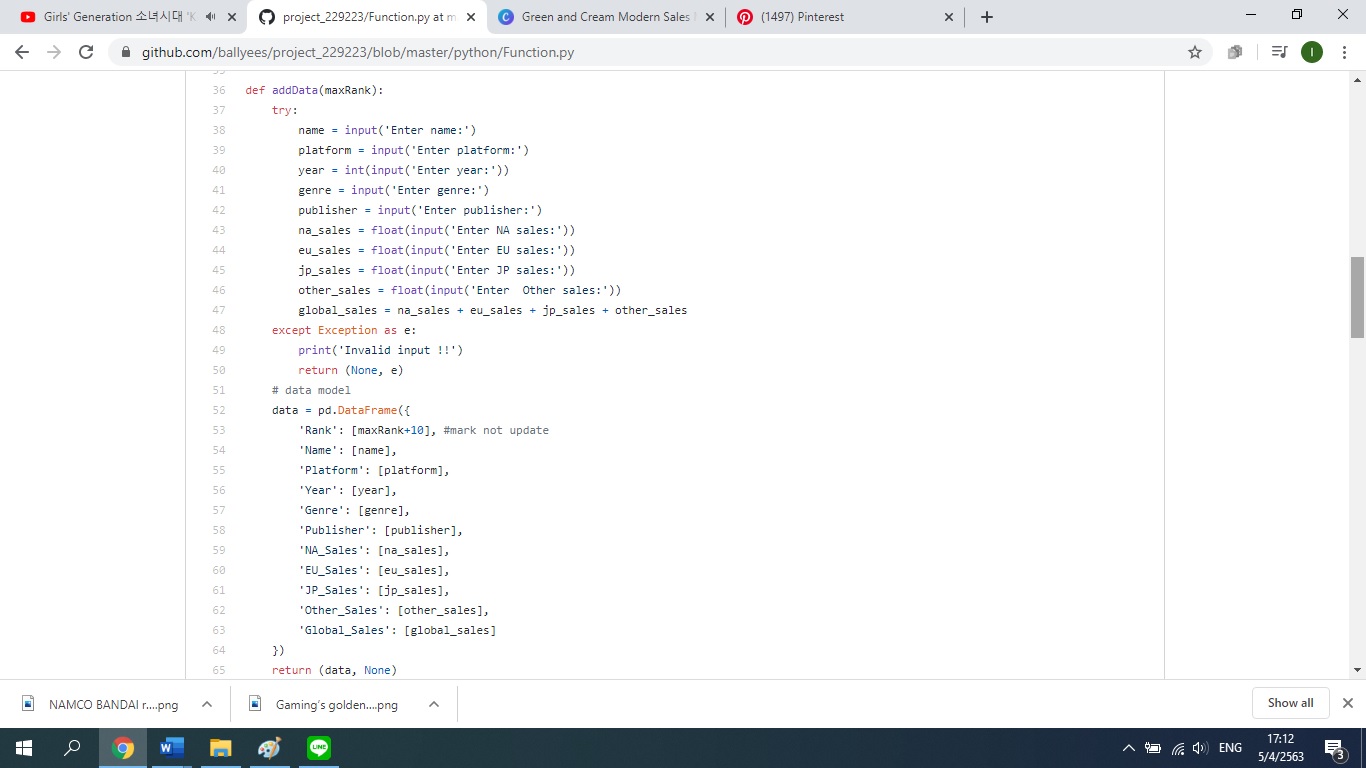


Figure (4.1)

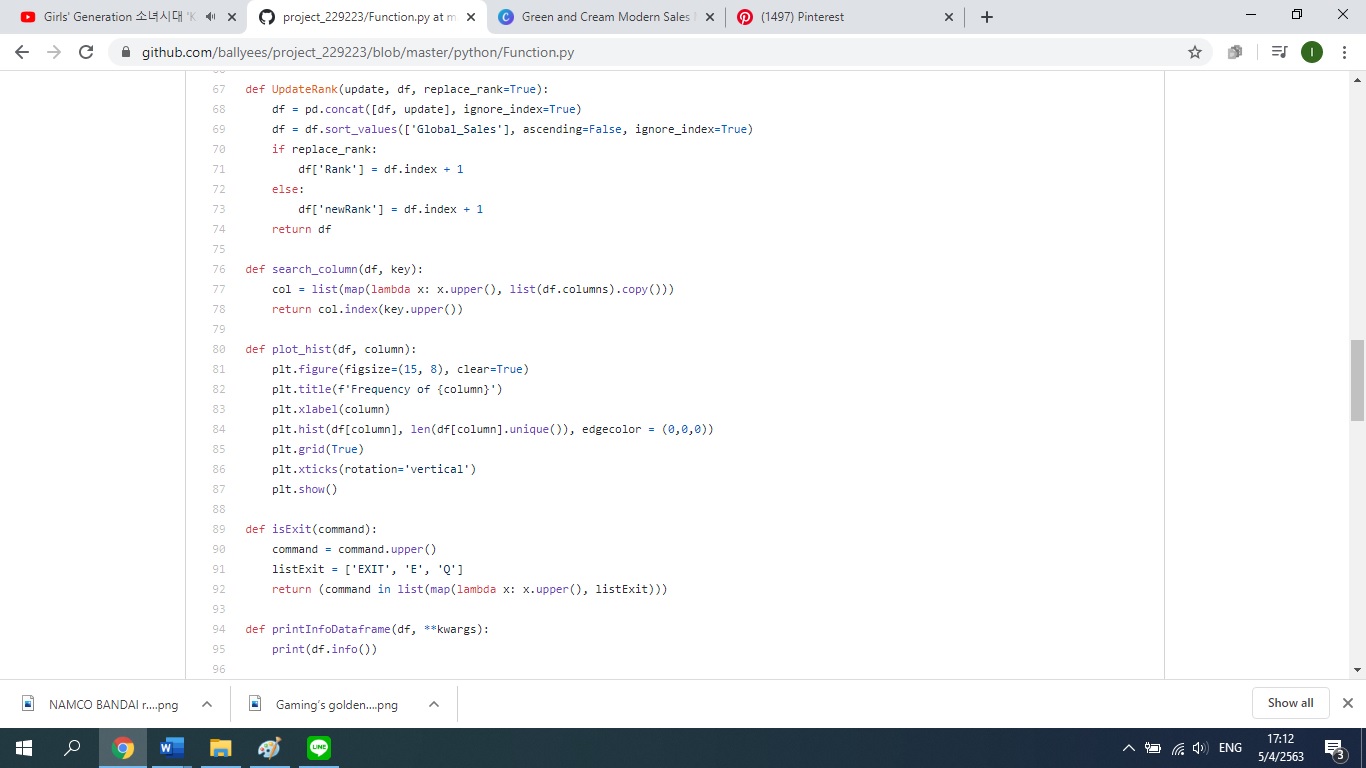


Figure (4.2)

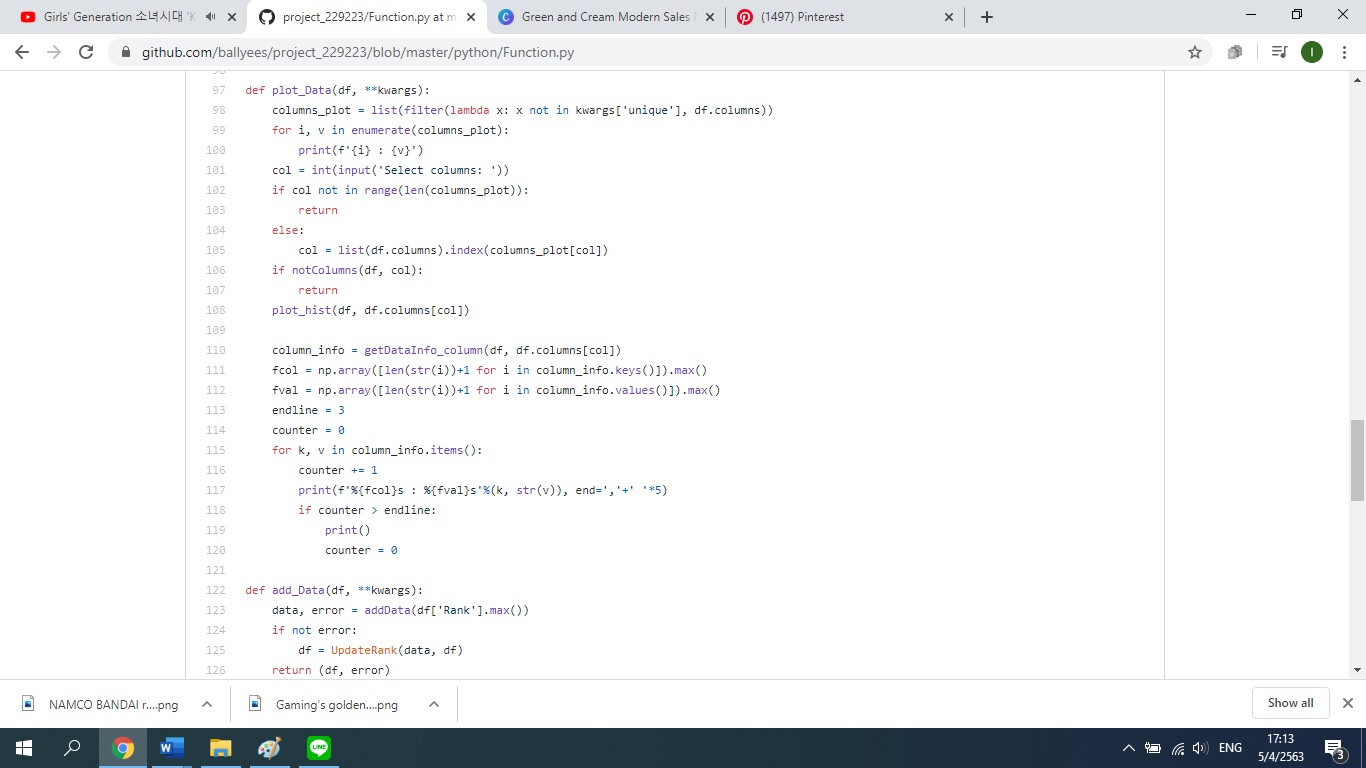


Figure (4.3)

*2.4 Search Data*

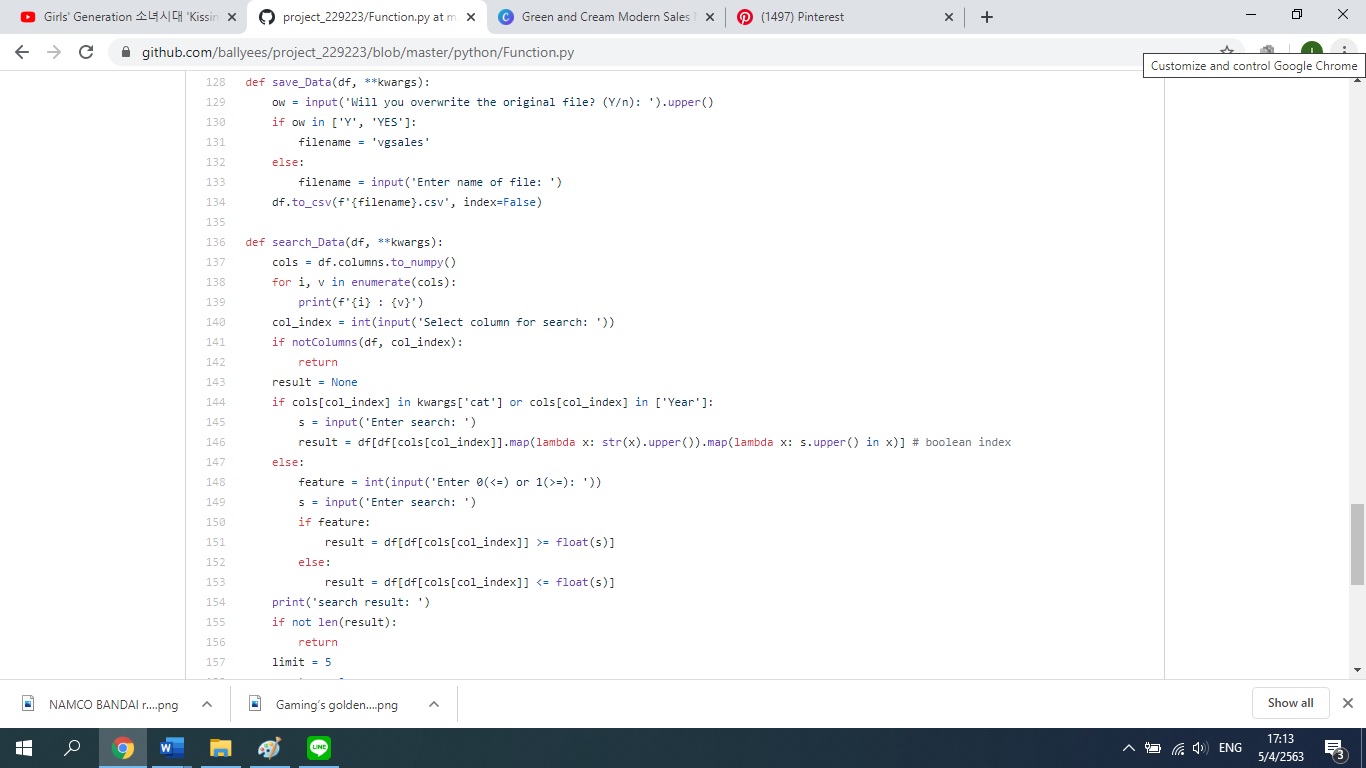
**

Figure (5.1)

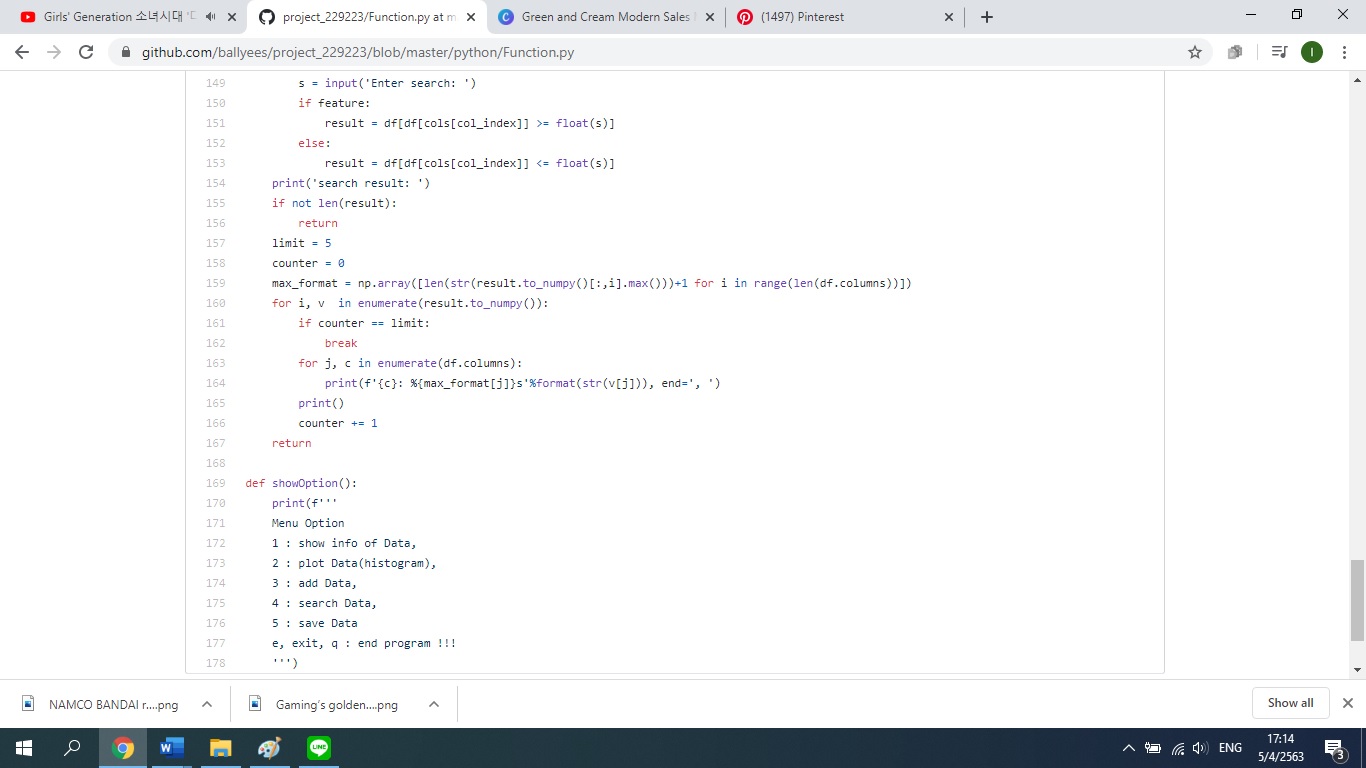
**

Figure (5.2)

หมายเหตุ : Figure 5.1 และ 5.2 เป็นโค้ดที่ต่อเนื่องกัน

*2.5 Save Data*

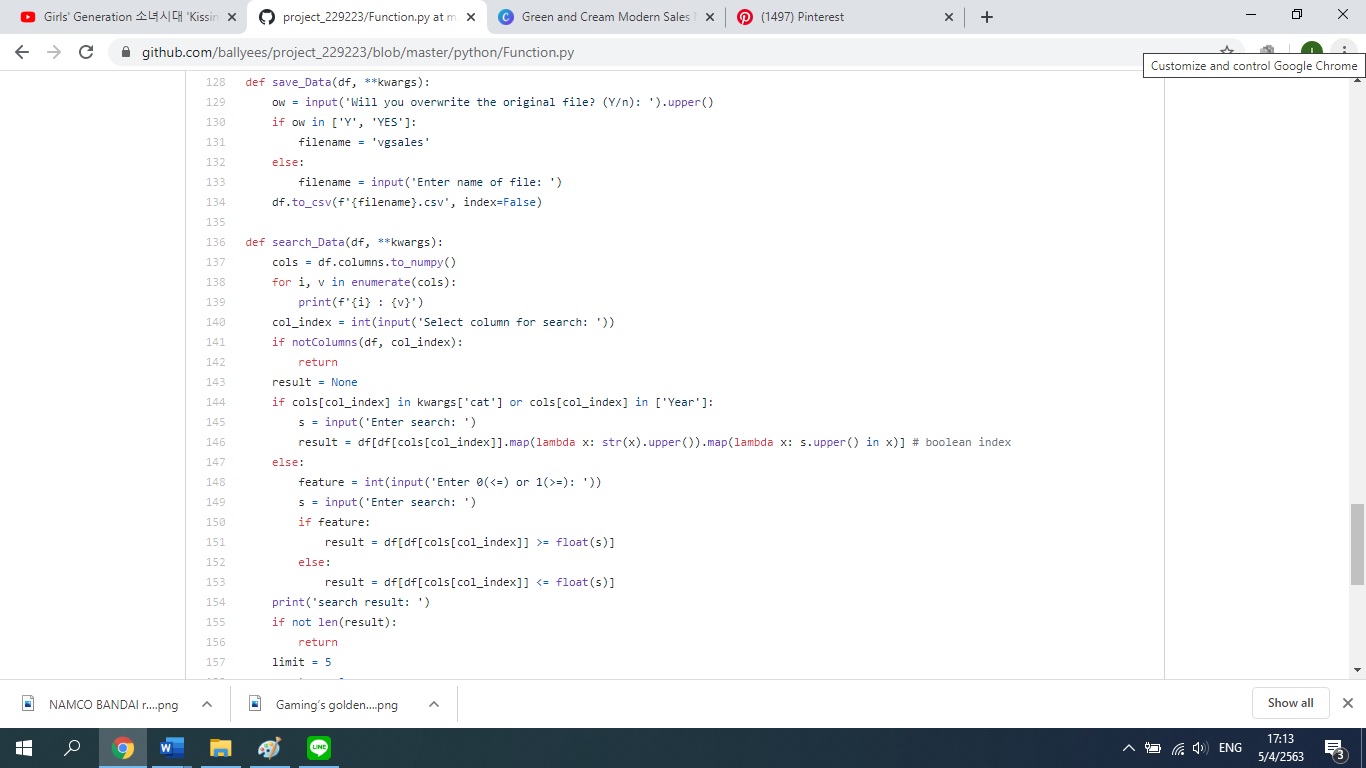
**

Figure (6)

*2.6 Exit and End the program*

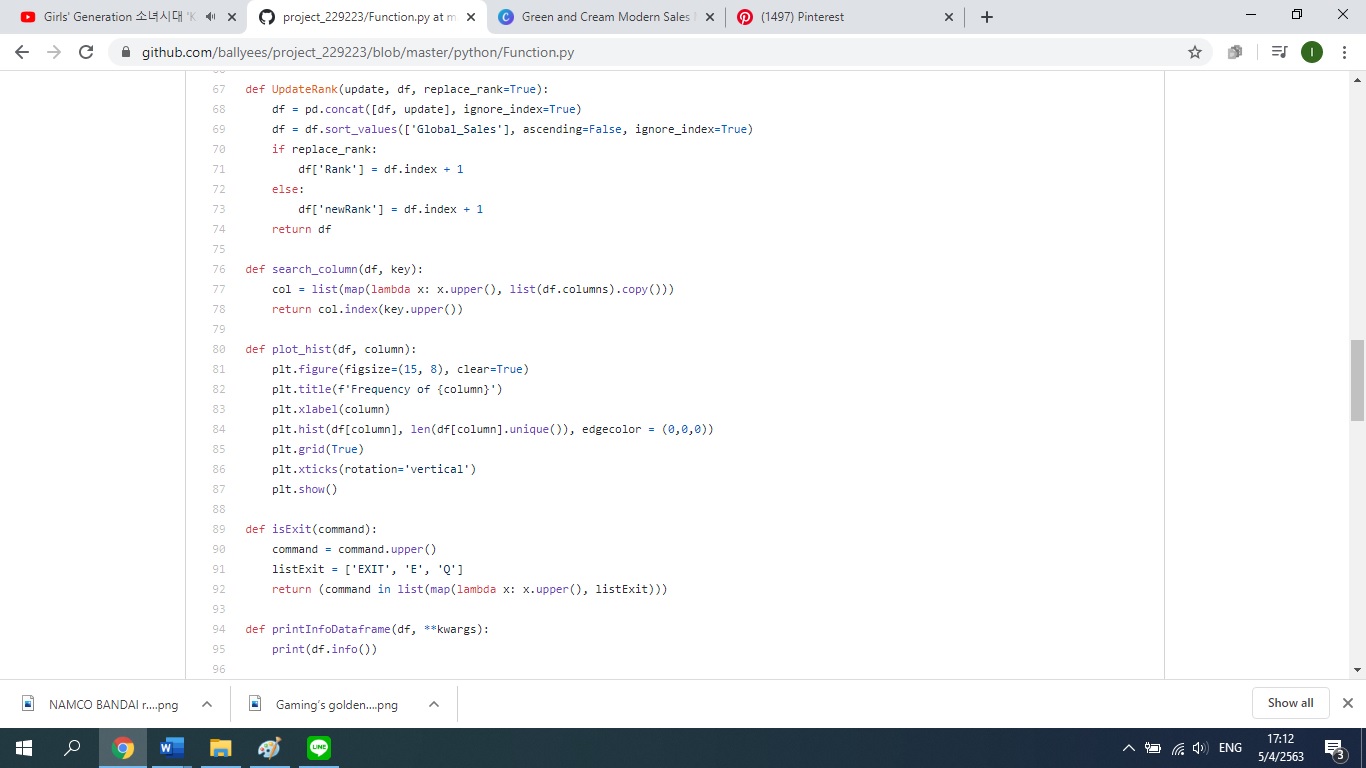
**

Figure (7)

*การเลือกใช้โครงสร้างข้อมูลและการเลือก Plot กราฟ*

*String: การใช้ string upper case ในคำสั่งต่างๆ*

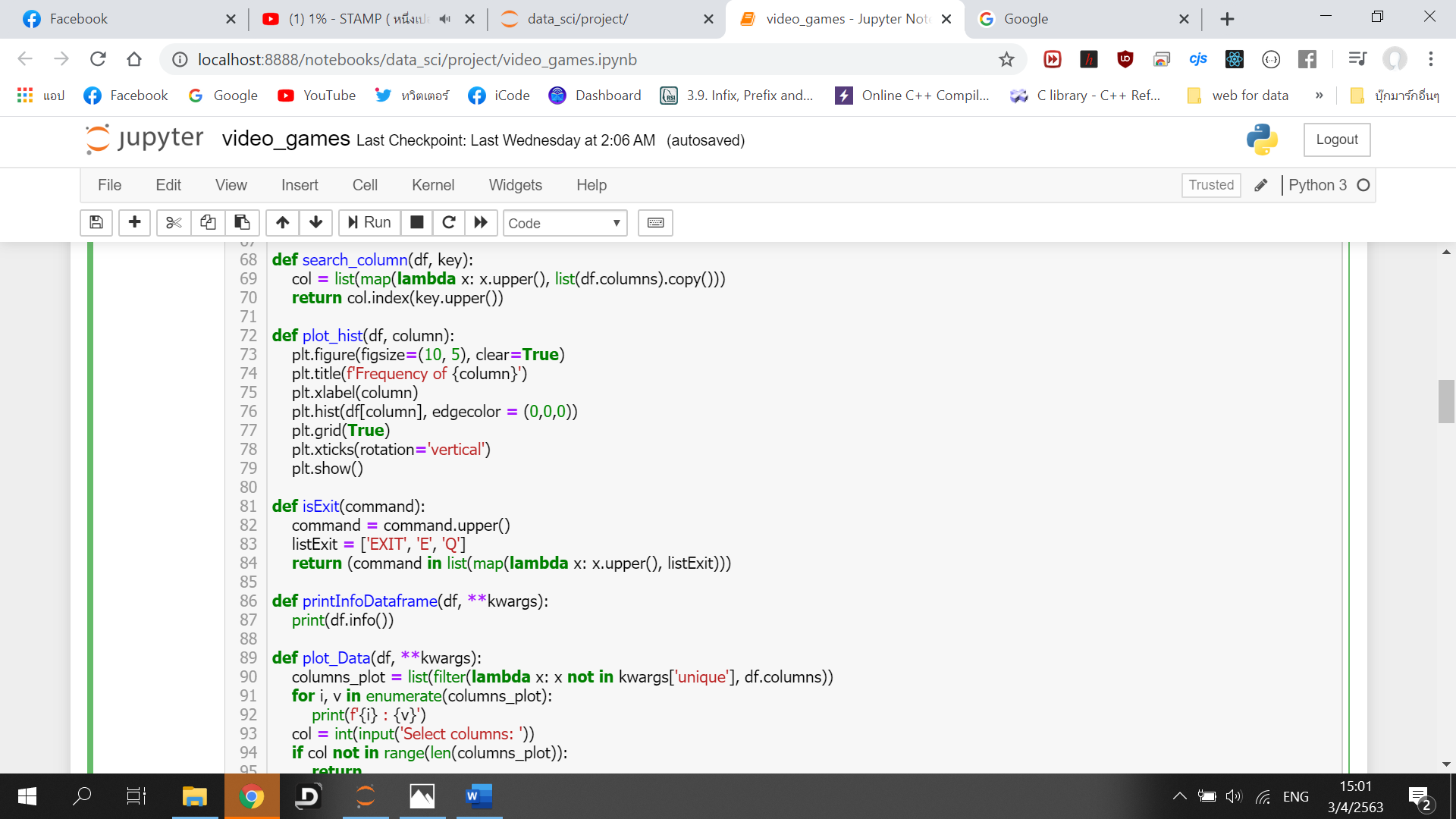


Figure (1.1)

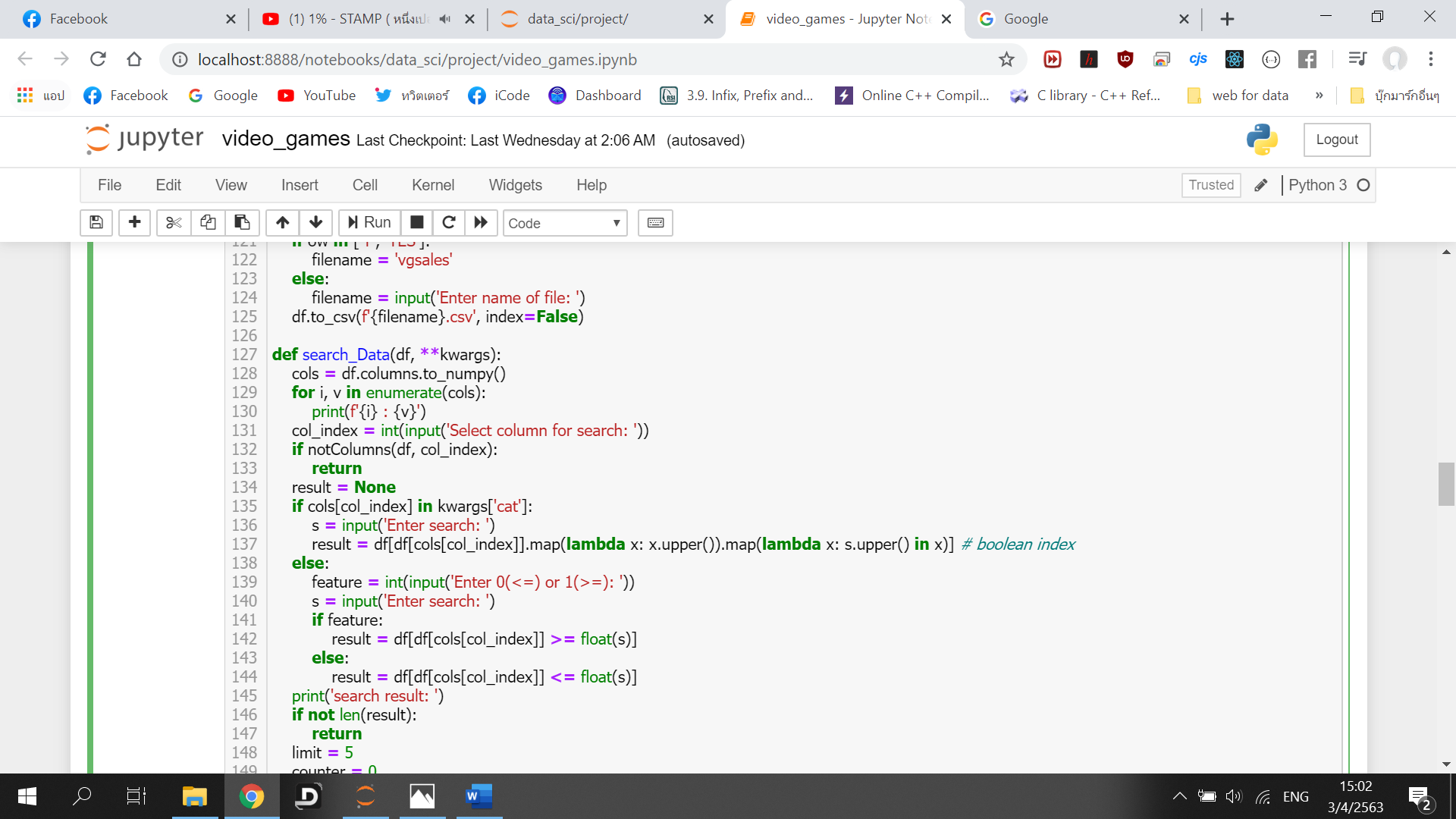


Figure (1.2)

1) ใน figure 1.1 แสดงให้เห็นว่าใช้ string upper case ร่วมกับการใช้ mapping function ในการตรวจสอบว่าคำสั่งที่รับมานั้นควรจะเป็นคำสั่งในการออกโปรแกรมหรือไม่

2) ใน figure 1.2 แสดงให้เห็นว่าใช้ในการ mapping ผลลัพธ์ของ DataFrame ที่ส่งมาให้เป็นตัวใหญ่ทั้งหมดและ filter เอาเฉพาะคำที่มีคำในตัวแปร s มาแสดง

*List & Dictionary*

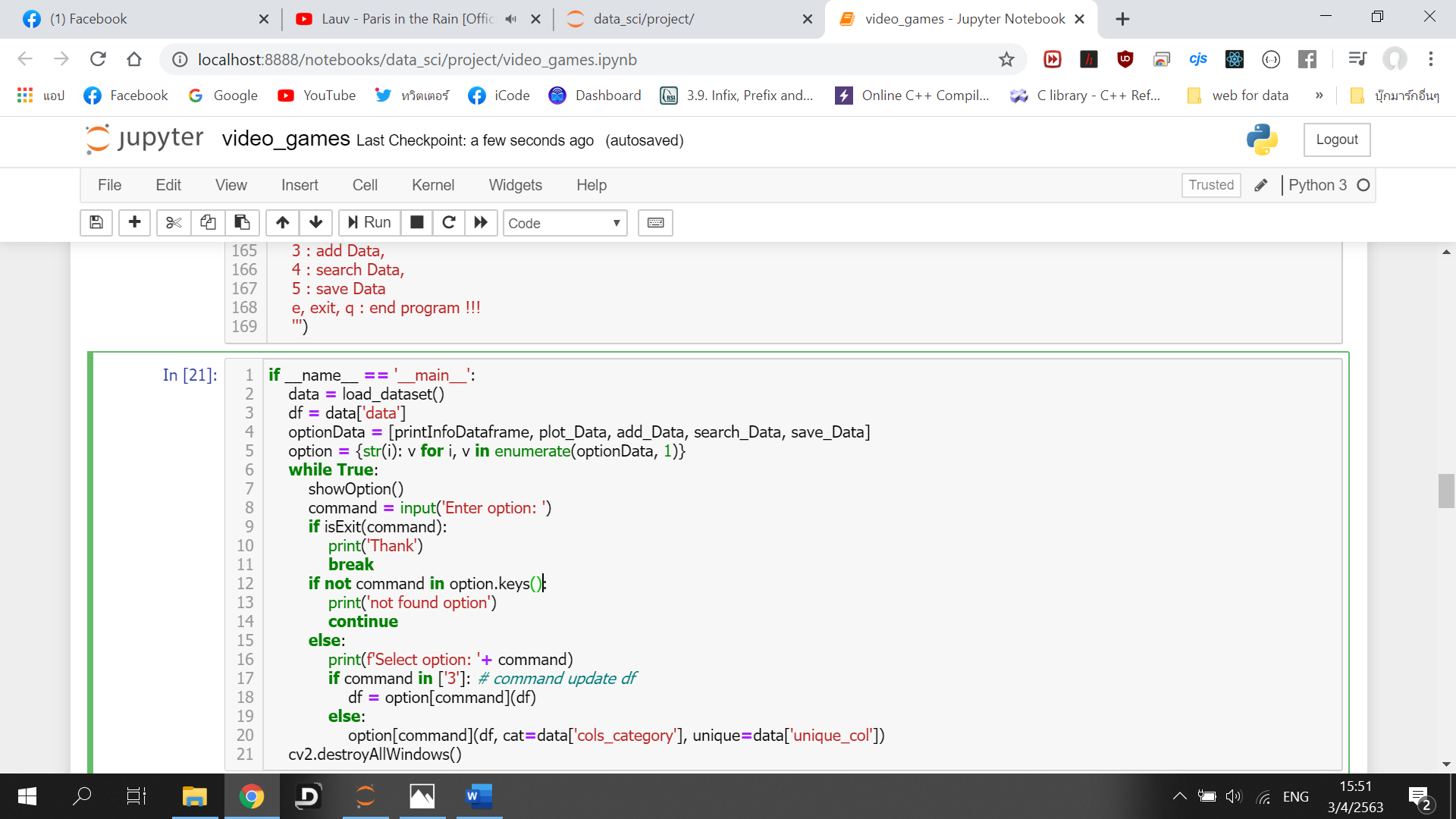


Figure (2)

1) ใน figure 2 จะแสดงให้เห็นว่าใช้ List ในการเก็บค่า location ของ method ต่างๆไว้ใน List เพื่อที่จะเก็บค่าต่างๆไว้ใน Dictionary ต่อและง่ายในการเขียนเมื่อต้องการเพิ่มค่าต่าง

2) ใช้ Dictionary ในการเก็บค่า Option ต่างๆแล้วนำค่ามา map เข้ากับ “1” -> “sizeof(Option)”

*Numpy Array*

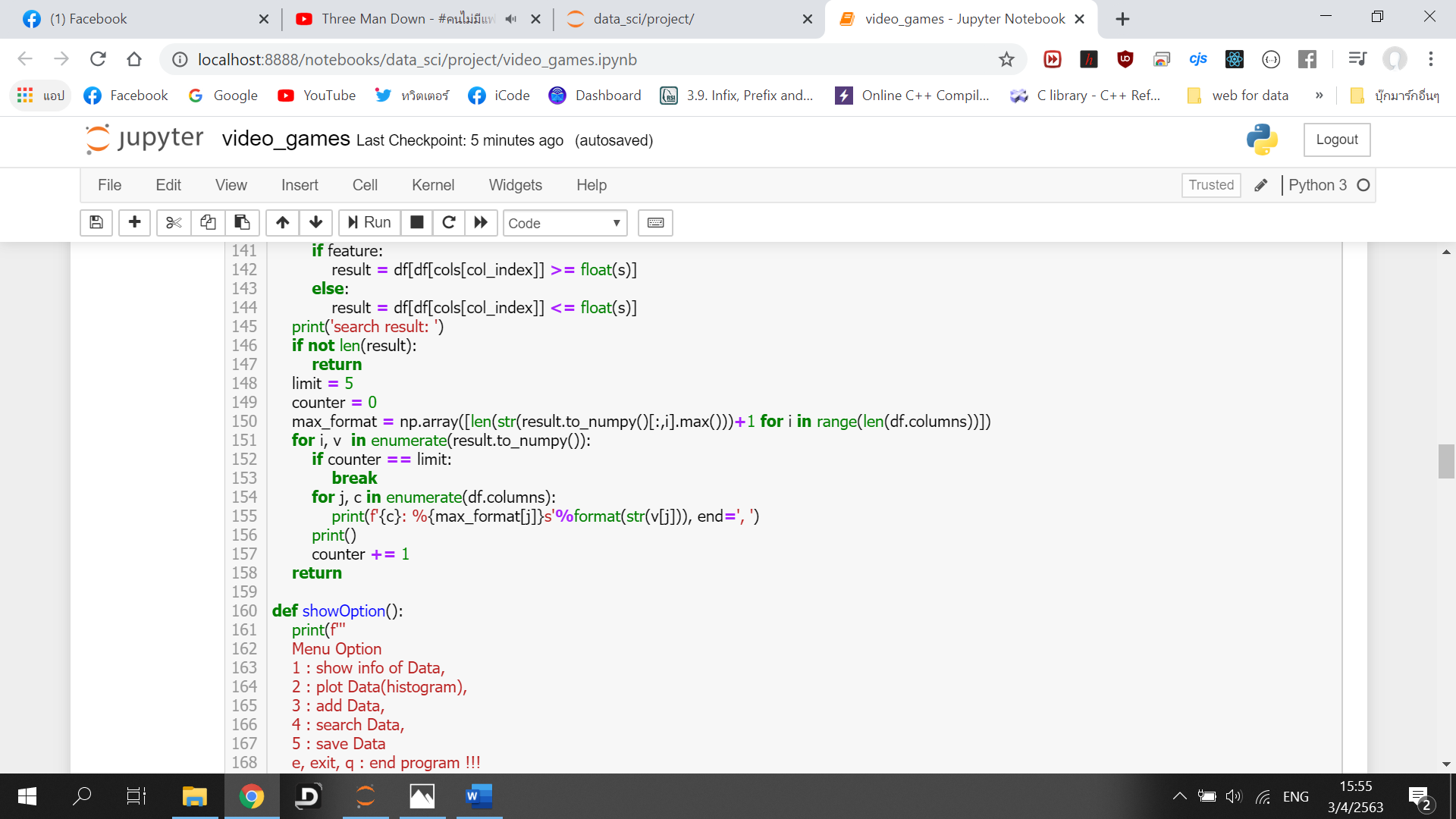


Figure (3)

1) ใน DataFrame จะมี method ที่ชื่อว่า to\_numpy() ซึ่งเราต้องการนำค่า MAX\_LENGTH\_OF\_STRING มาเพื่อที่จะแสดง format ให้สวยงาม และใน numpy array จะมี method max() ที่ต้องการแล้วนำขนาดที่ได้มาบวกกับ 1

*หมายเหตุ* : method max ใน numpy array ถ้าเป็นค่า max ใน string จะเป็นการนำ ค่า string ที่มีความยาวมากที่สุดมากแสดง

*Graph: histogram & data infomation*

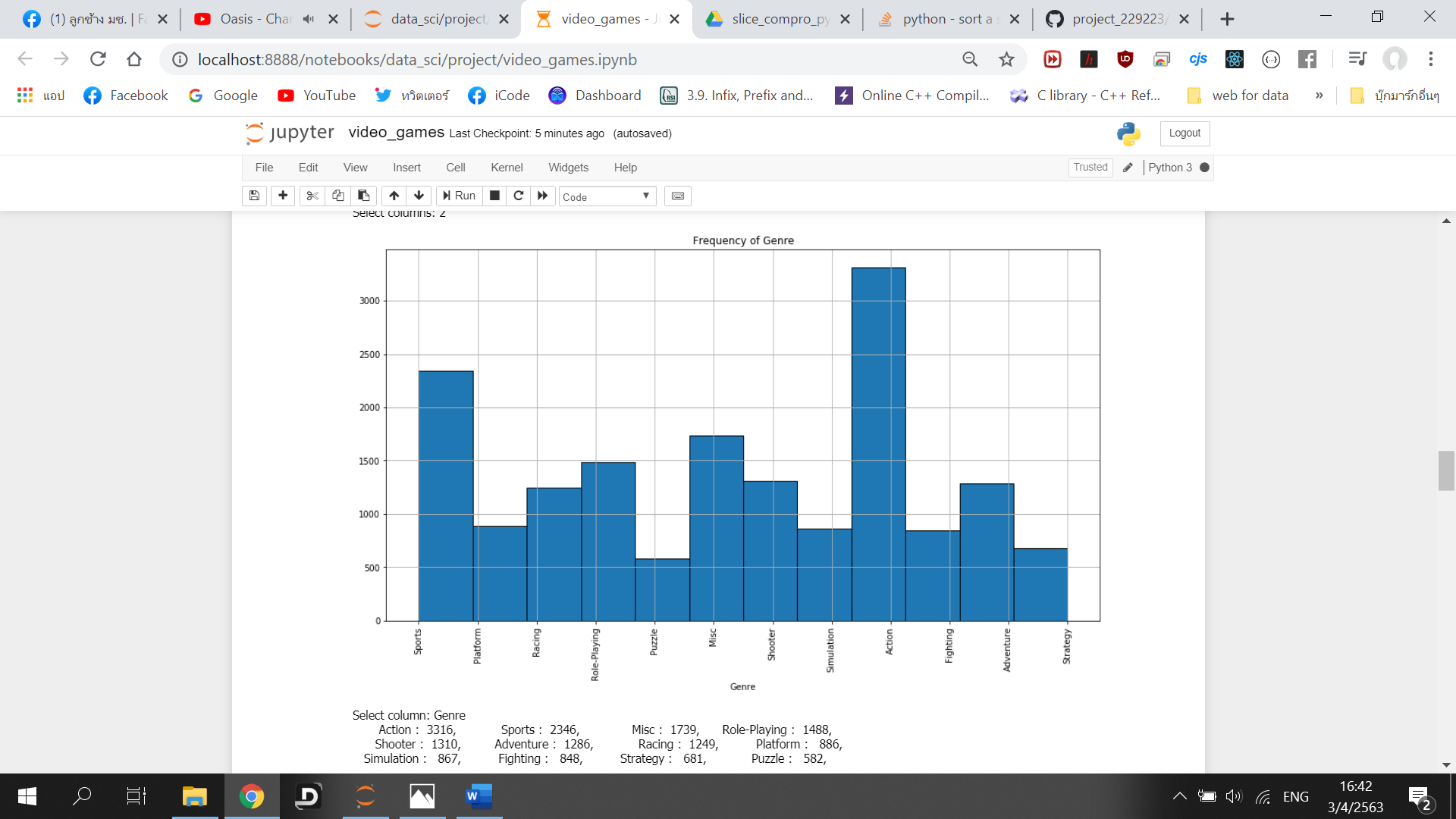


Figure (4) แสดงผลประเภทของวิดีโอเกมส์ (Gemre) ในรูปแบบกราฟฮิสโทแกรม

1) แสดงเป็นกราฟความถี่ใน column นั้นๆในแกน x จะแสดงข้อมูลของใน columns และ ในแกน y จะแสดงข้อมูลคสามมูลความถี่ของการเกิดในแต่ละตัว

2) ในบรรทัดต่อไปจะเป็นการแสดงข้อมูลแบบละเอียดว่าใน column นั้นความถี่ของแต่ละตัวเป็นแบบไหน

*Edit File & Save File:*

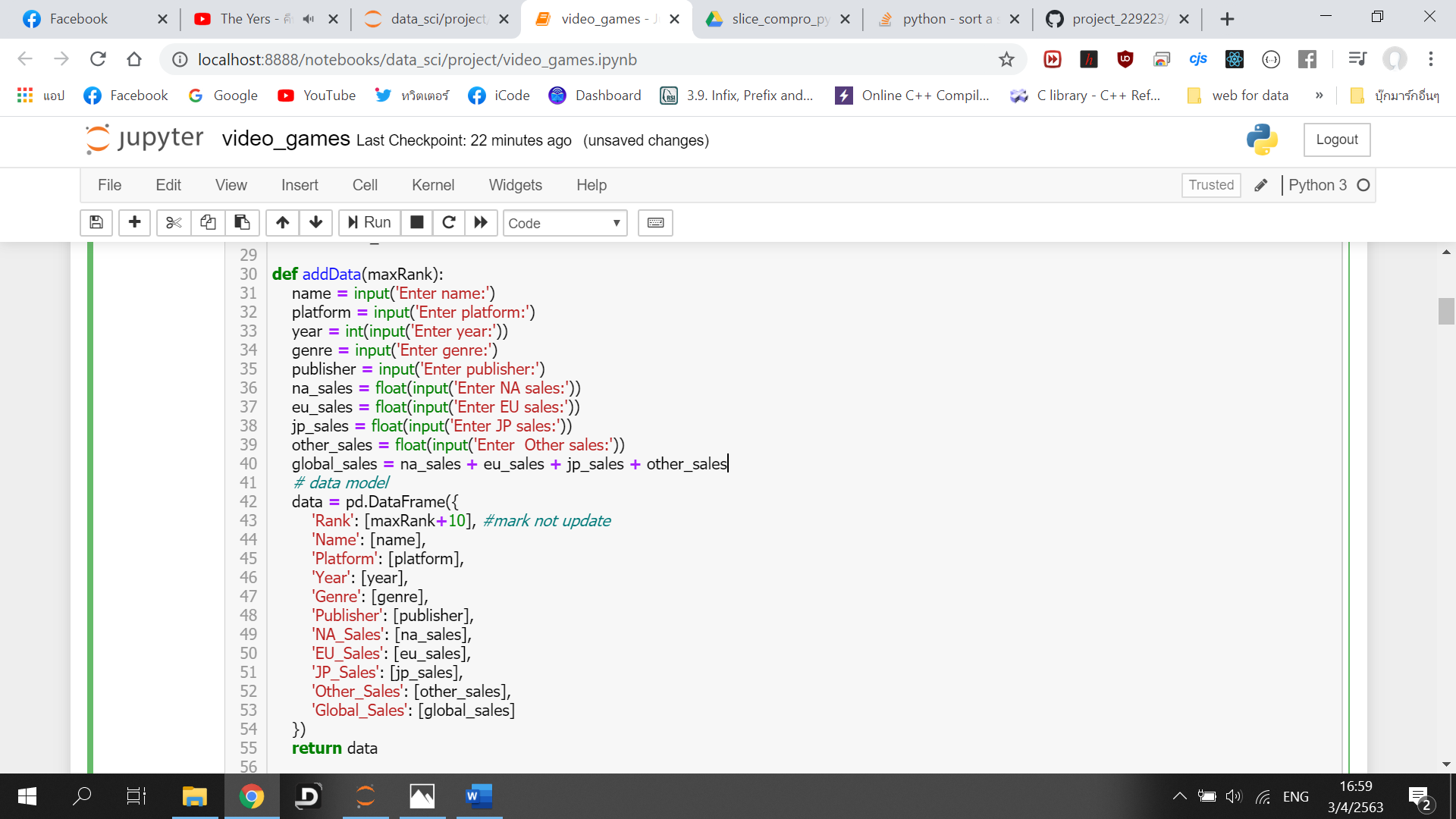


Figure (5.1)

1) ขั้นตอนใน Edit file จะมีขั้นตอนดังนี้ AddData -> UpdateData -> SaveData (Optional)

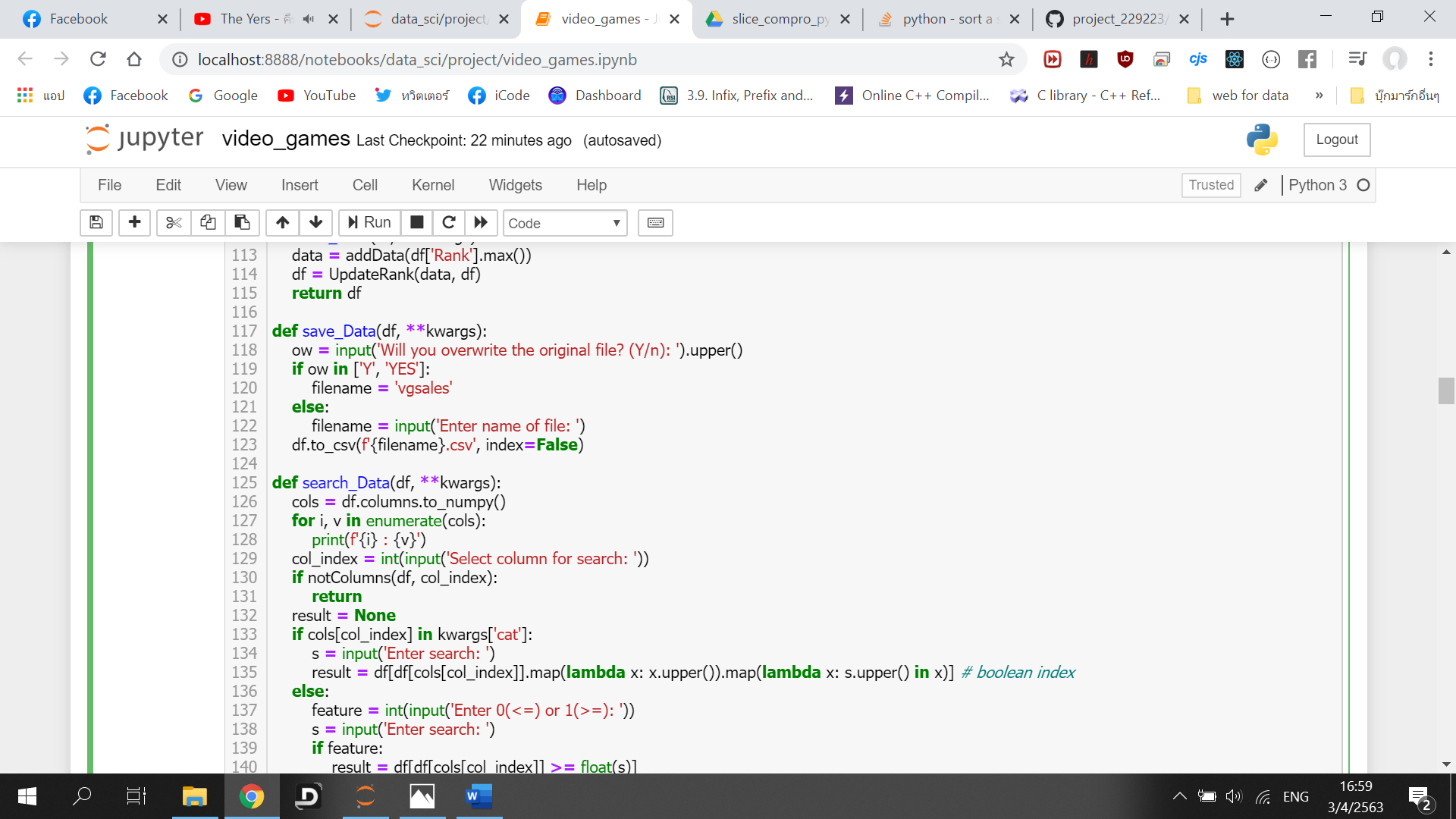


Figure (5.2)

2) ใน figure 5.2 เป็นการ save file ที่ให้เลือกว่าจะ save ที่ file เดิมหรือไม่

*List comprehension*

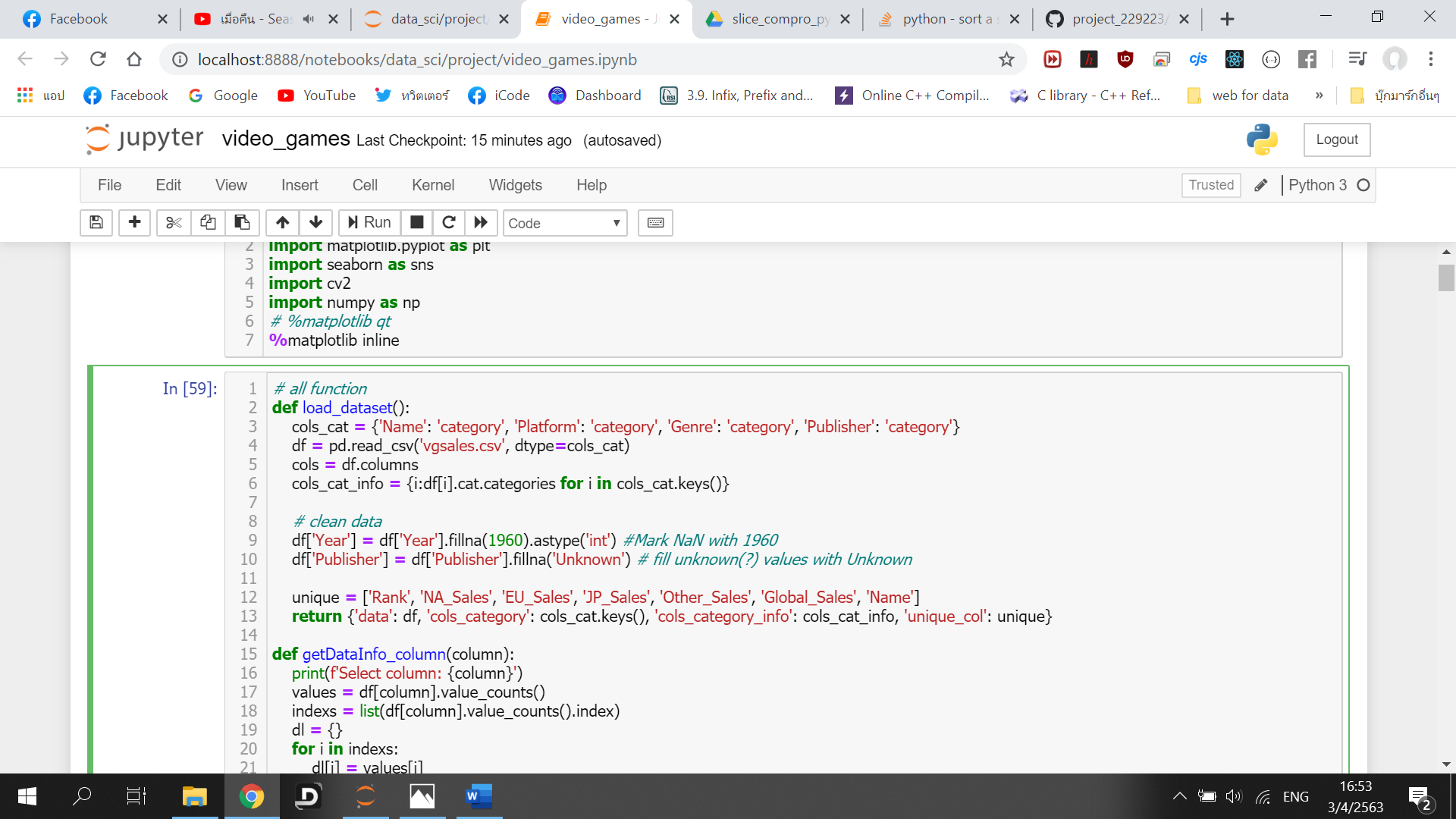


Figure (6)

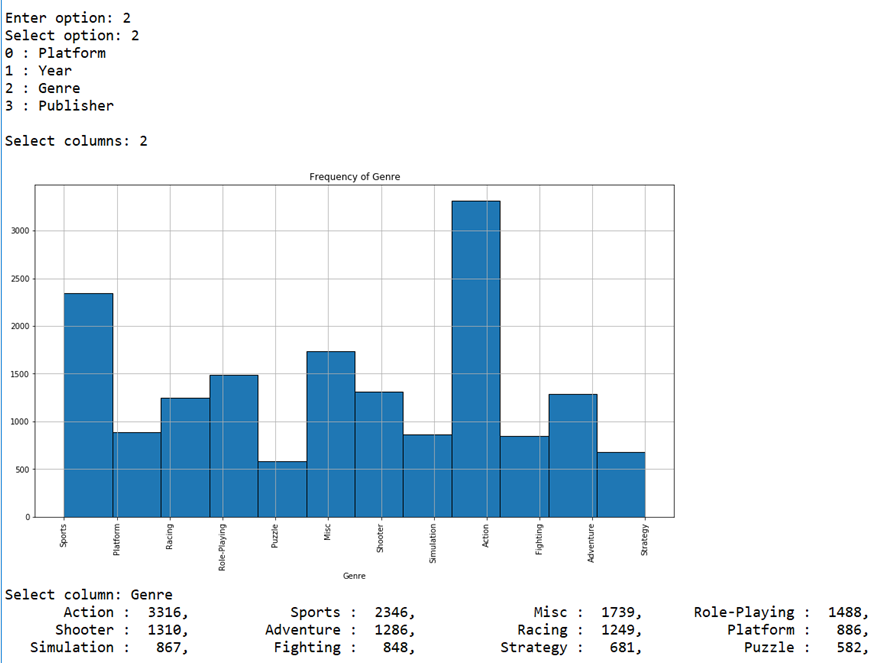
ใน figure 6 จะเห็นว่าใช้ for-loop ในการดึงค่า keys ใน cols\_cat มาใช้งานโดยที่ให้ index เป็น key ของแต่ละตัว แล้ว values เป็น ค่าของ categories ที่ map ไป

*Scipy Function:*

การใช้ pandas DataFrame ใน อ่าน data ใน file csv (figure 6) และการใช้ boolean index ของ pandas Data Frame ในการเข้าถึง (figure 1.2)

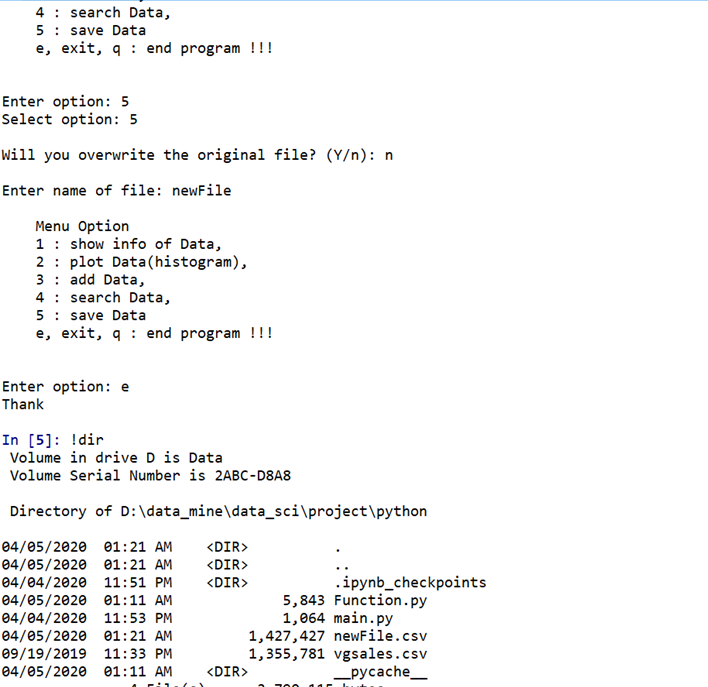
*ตัวอย่างการทำงานของโปรแกรม*

*1.การแสดงผลเมื่อเลือกตัวเลือกที่ 2 : Plot Data (Histogram)*

*.*

เป็นการเลือก option2 (plot histogram) แล้วเลือก column Genre โดยใช้ 2 เป็นตัวอ้างถึงแล้วระบบจะทำการ plot histogram ค่าความถี่ของประเภทเกมแต่ละประเภทมาให้แล้วจะสามารถดูค่าแบบละเอียดได้ในด้านล่างของ histogram

*2. การแสดงผลเมื่อเลือกตัวเลือกที่ 5 : Save Data*

**

เลือก option 5 เพื่อที่จะ save data เก็บไว้โดยที่ไม่ได้ save ไว้ใน file เดิมแต่เป็นไฟล์ใหม่ที่ชื่อ newFile.csv