**Homework #2**

**วิเคราะห์สถิติและกราฟ คุณภาพของไวน์แดง**

ผู้จัดทำเลือกทั้งหมด **2 จาก 3 คอลัมน์** แต่ละคอลัมน์มีทั้งหมด **1599 แถว**:

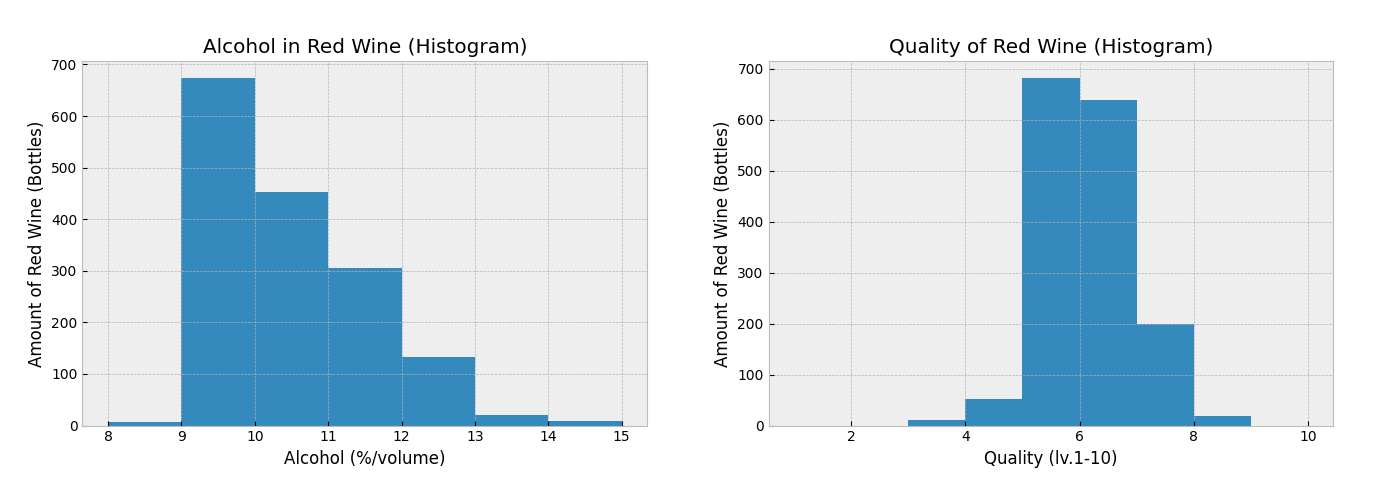
1. Alcohol (g/liter) 2. Quality (lv. 1-10)

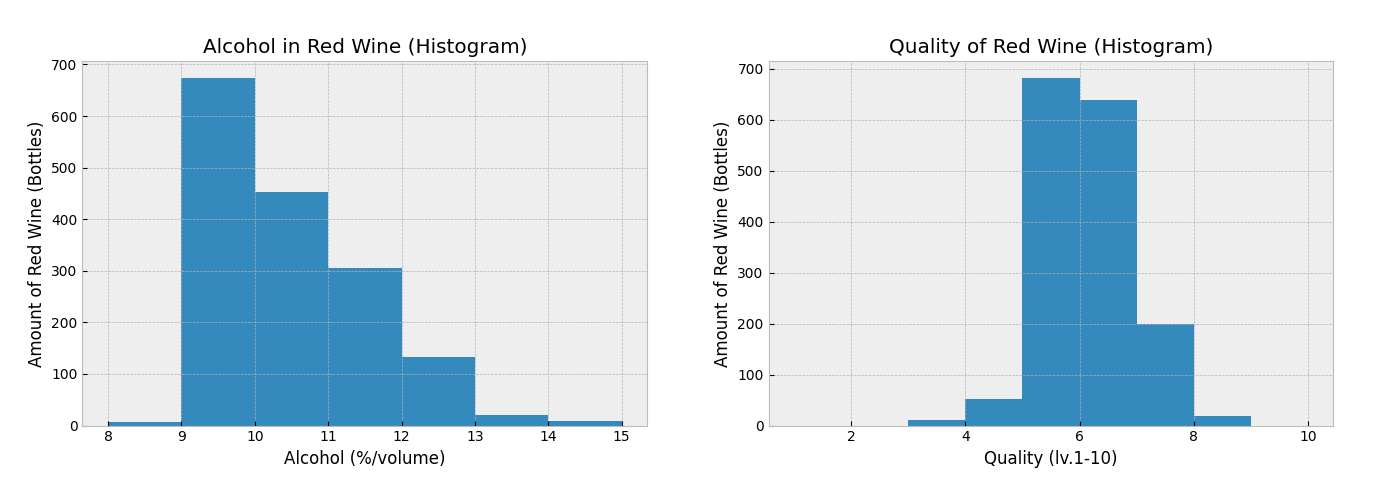
**คำนวณค่าสถิติพื้นฐาน**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Min** | **Mean** | **Median** | **Max** | **Mode** | **Sample Standard Deviation** | **Sample**  **Variation** |
| **1.Alcohol**  **(%/volume)** | 8.4 | 10.42 | 10.20 | 14.9 | 9.5 | 1.0656 | 1.1356 |
| **2.Quality**  **(lv.1-10)** | 3 | 5.6 | 6.0 | 8 | 5 | 0.8075 | 0.6521 |

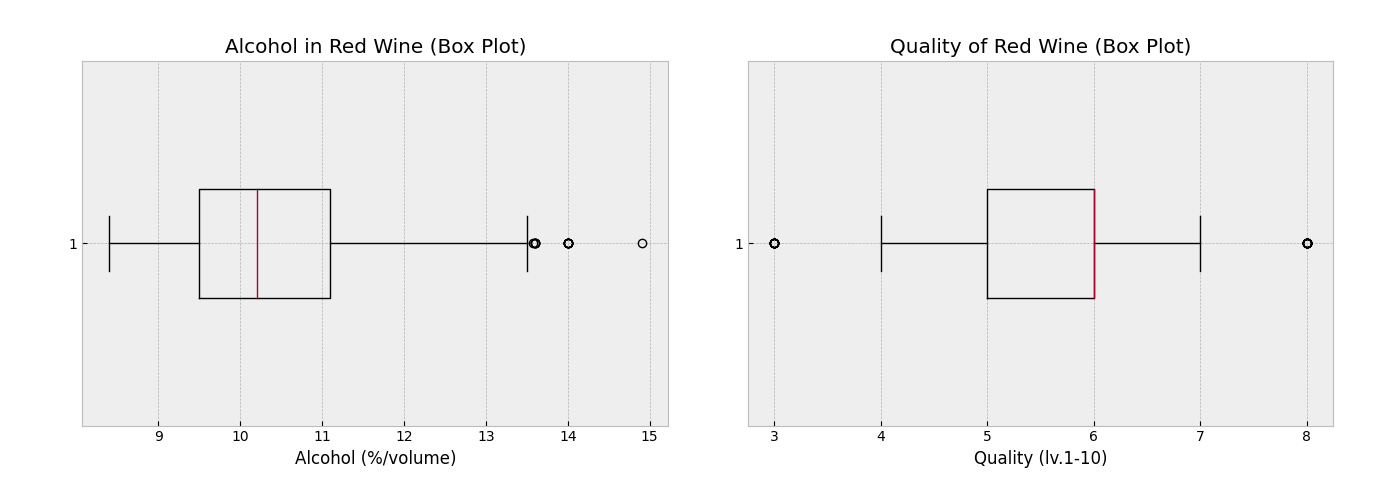
อ้างอิงจากการคำนวณในโปรแกรม WineGraph.py

**Histogram Graphs**

****

****

**Box Plot Graphs**



**การคำนวณหา Outliers**

**สามารถดูได้จากกราฟ Box Plot หรือจากการคำนวณหา Outliers โดยใช้สูตรดังนี้**

* 1. **Q1 = (N+1) \* 1 /4** (คิดแบบสูตรการไม่แจกแจงความถี่ / เห็นข้อมูลเป็นตัว ไม่ใช่ช่วง)
  2. **Q3 = (N+1) \* 3 /4** (คิดแบบสูตรการไม่แจกแจงความถี่ / เห็นข้อมูลเป็นตัว ไม่ใช่ช่วง)
  3. **IQR = Q3 – Q1** (คิดแบบ Inclusive เพื่อทำให้ IQR มีค่าแคบลงเพื่อตัด Outliers ได้มากขึ้น)
  4. **Extreme Outlier Lower Boundary** = Q1 – IQR \* 3
  5. **Mild Outlier Lower Boundary** = Q1 – IQR \* 1.5
  6. **Mild Outlier Upper Boundary** = Q3 + IQR \* 1.5
  7. **Extreme Outlier Upper Boundary** = Q3 + IQR \* 3
  8. **Extreme Outlier (Lower)** = if (value < Ext. Outlier Lower B.)
  9. **Mild Outlier (Lower)** = if (Ext. Outlier Lower B. <= value < Mild Outlier Lower B.)
  10. **Mild Outlier (Upper)** = if (Mild Outlier Upper B. < value <= Ext. Outlier Upper B.)
  11. **Extreme Outlier (Upper)** = if (value >= Ext. Outlier Upper B.)

1. **Alcohol in Red Wine : มีจำนวนรวมจุด Outliers ทั้งหมด 13 ค่า ดังนี้** (หน่วย: %/Volume)
   1. **Extreme Outlier (Lower) = []**
   2. **Mild Outlier (Lower) = []**
   3. **Mild Outlier (Upper) = [13.57, 13.6, 13.6, 13.6, 13.6, 14.0, 14.0, 14.0, 14.0, 14.0, 14.0, 14.0, 14.9]**
   4. **Extreme Outlier (Upper) = []**
2. **Quality of Red Wine : มีจำนวนรวมจุด Outliers ทั้งหมด 28 ค่า ดังนี้** (หน่วย: - (lv.1-10))
   1. **Extreme Outlier (Lower) = []**
   2. **Mild Outlier (Lower) = [3, 3, 3, 3, 3, 3, 3, 3, 3, 3]**
   3. **Mild Outlier (Upper) = [8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8, 8]**
   4. **Extreme Outlier (Upper) = []**

**(รายละเอียดในแต่ละค่าเพิ่มเติม สามารถดูได้ใน OUTPUT ของโปรแกรม WineGraph.py ในหน้าที่ 10)**

**Stem And Leaf Graph**

**Alcohol in Red Wine Stem-And-Leaf**

**รูปภาพประกอบด้วย ข้อความ

คำอธิบายที่สร้างโดยอัตโนมัติ**

**Quality of Red Wine Stem-And-Leaf**

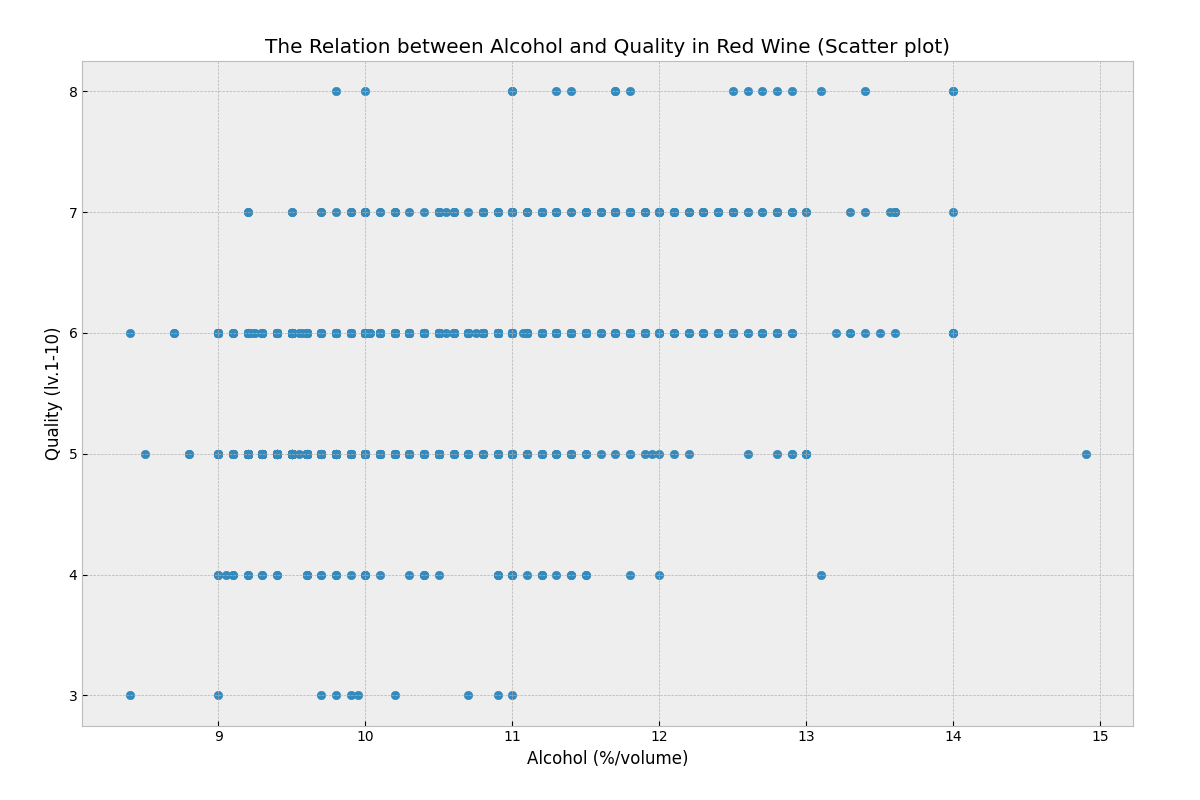
**รูปภาพประกอบด้วย โต๊ะ

คำอธิบายที่สร้างโดยอัตโนมัติ**

**\*หมายเหตุ** เนื่องจากข้อมูลมีจำนวนมาก ภาพอาจจะใหญ่เกินไปสำหรับกราฟแบบ Stem-And-Leaf จึงไม่สามารถเก็บมาหมดได้ สำหรับ Quality สามารถซูมเพื่อดู Sum (เก็บแบบ Aggregation) จากด้านหน้า Stem ได้ จะเป็นการรวมกันเป็นลำดับขึ้นไปเรื่อยๆ จากล่างขึ้นบน รวมทั้งหมด 1599 แถวข้อมูล

**Scatter Plot Graph**

**ความสัมพันธ์ระหว่างปริมาณแอลกอฮอล์และคุณภาพของไวน์แดง**



กำหนดให้ **แนวแกน x** เป็น**ตัวแปรต้น** (Independent Variable) = **ปริมาณแอลกอฮอล์ (%/ปริมาตร)**

กำหนดให้ **แนวแกน y** เป็น**ตัวแปรตาม** (Dependent Variable) = **คุณภาพของไวน์ (ระดับ 1-10)**

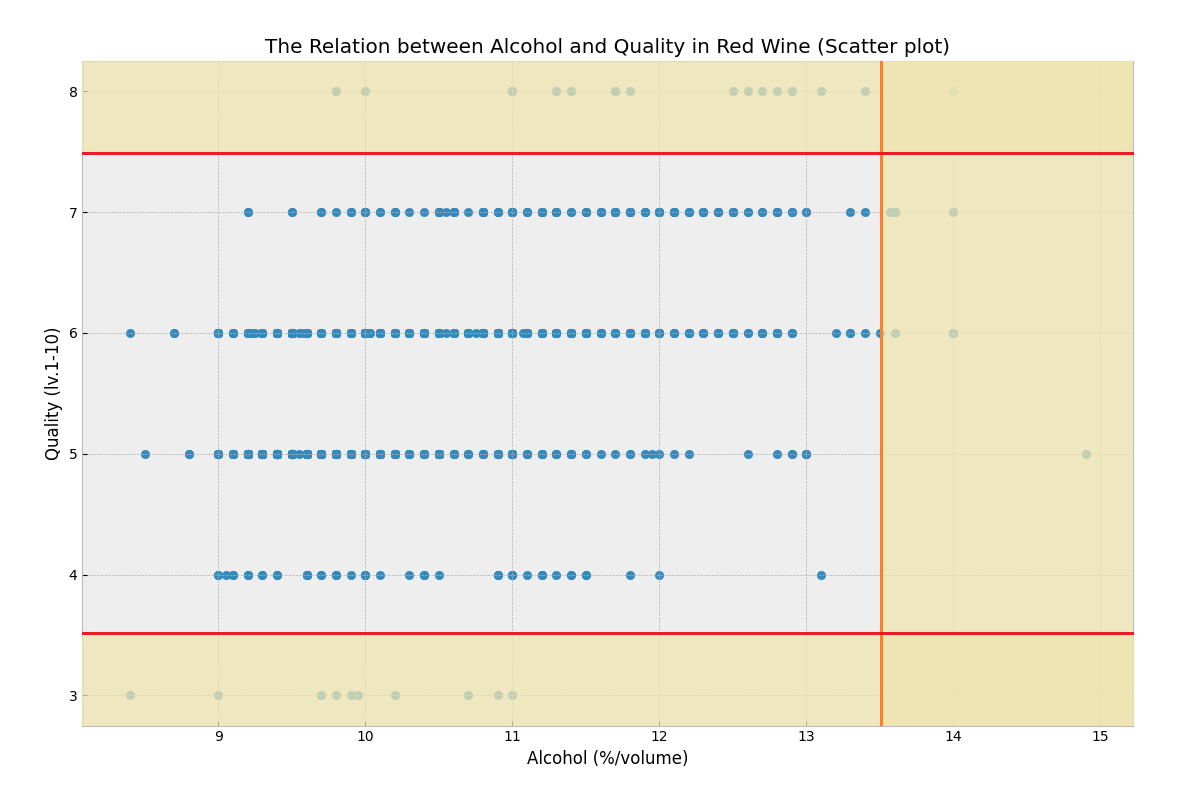
เหตุผลที่ใช้ปริมาณแอลกอฮอล์เป็นตัวแปรต้น เพราะว่าต้องการที่จะศึกษาเกี่ยวกับปริมาณแอลกอฮอล์ในไวน์แดง ว่ามีผลต่อคุณภาพของไวน์แดงมากแค่ไหน เช่น หากเรามีปริมาณแอลกอฮอล์จำนวน %/ปริมาตร ในไวน์แดงที่มาก อาจทำให้คุณภาพของไวน์แดงมากขึ้นตามด้วย จึงกำหนดให้ปริมาณแอลกอฮอล์เป็นตัวแปรต้น และคุณภาพของไวน์เป็นตัวแปรตาม

**บทวิเคราะห์ข้อมูลจากกราฟทั้งหมด**

เนื่องจากกราฟ Scatter ความสัมพันธ์ระหว่างปริมาณแอลกอฮอล์และคุณภาพของไวน์แดงในปัจจุบัน มีจำนวน Outliers อยู่จำนวนหนึ่ง ผมจึงทำการลอง plot กราฟ Scatter ใหม่ โดยการตีเส้นและไม่นำ Outliers มาคิดในการวิเคราะห์ จะได้ออกมาเป็นกราฟดังนี้ (อ้างอิงจากการคำนวณในหน้าที่ 2)

**Outlier Alcohol [ value < 7.1, value > 13.5] (เส้นสีส้ม)**

**Outlier Quality [ value < 3.5, value > 7.5] (เส้นสีแดง)**

****

จากกราฟที่ได้ เราจะสนใจแค่เพียงในส่วนสีขาวเท่านั้น (ที่ไม่ใช่แรเงาสีเหลือง) จะเห็นได้ว่า ข้อมูลที่ได้ค่อนข้างมีความกระจุกตัวอยู่บริเวณตรงกลางเป็นส่วนมาก และมีแนวโน้มเอียงขึ้นไปทางบนขวาเล็กน้อย

ซึ่งถ้าหากเราสังเกตกราฟโดยละเอียด จะพบว่า

1. ช่วงปริมาณแอลกอฮอล์ตั้งแต่ **8.0-9.5 %/volume** คุณภาพของไวน์ส่วนใหญ่จะอยู่ที่**ระดับ 4-6**
2. และช่วงปริมาณแอลกอฮอล์ตั้งแต่ **12.0-13.0 %/volume ขึ้นไป** จะมีคุณภาพของไวน์ตั้งแต่**ระดับ 6-7** เป็นส่วนมาก

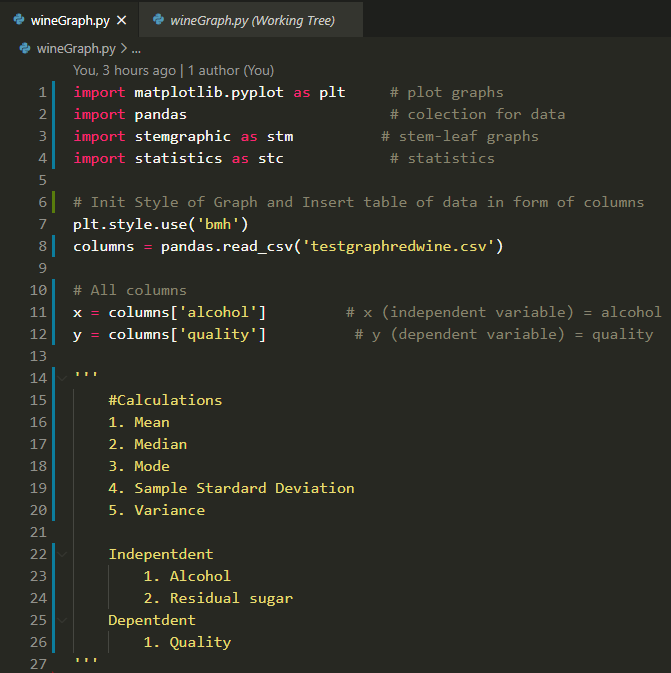
ซึ่งจากการวิเคราะห์ที่ได้ หากเรามีปริมาณแอลกอฮอล์ในไวน์แดงในปริมาณน้อย คุณภาพของไวน์แดงจะมีแนวโน้มที่จะน้อยตามไปด้วย และถ้าหากเรามีปริมาณแอลกอฮอล์ในไวน์แดงที่มาก คุณภาพของไวน์แดงก็จะมีแนวโน้มมากขึ้นตามไปด้วย โดยข้อมูลทั้งหมดนี้ถูกเก็บรวบรวมและอ้างอิงมาจากโรงงานผลิตไวน์ จังหวัด Minho ทางตอนเหนือของประเทศโปรตุเกส

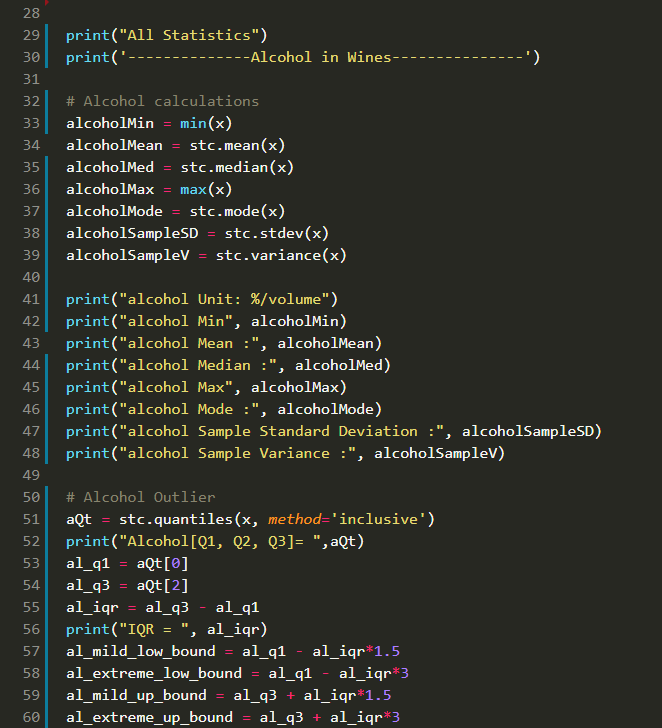
**กล่าวโดยสรุปคือ ปริมาณแอลกอฮอล์ในไวน์แดงที่มากขึ้น อาจมีแนวโน้มที่จะทำให้คุณภาพของไวน์แดงเพิ่มขึ้นตามไปด้วย โดยความสัมพันธ์ระหว่างปริมาณแอลกอฮอล์ในไวน์แดง และคุณภาพของไวน์แดงเป็นความสัมพันธ์แบบแปรผันตรงหรือคล้อยตามกัน**

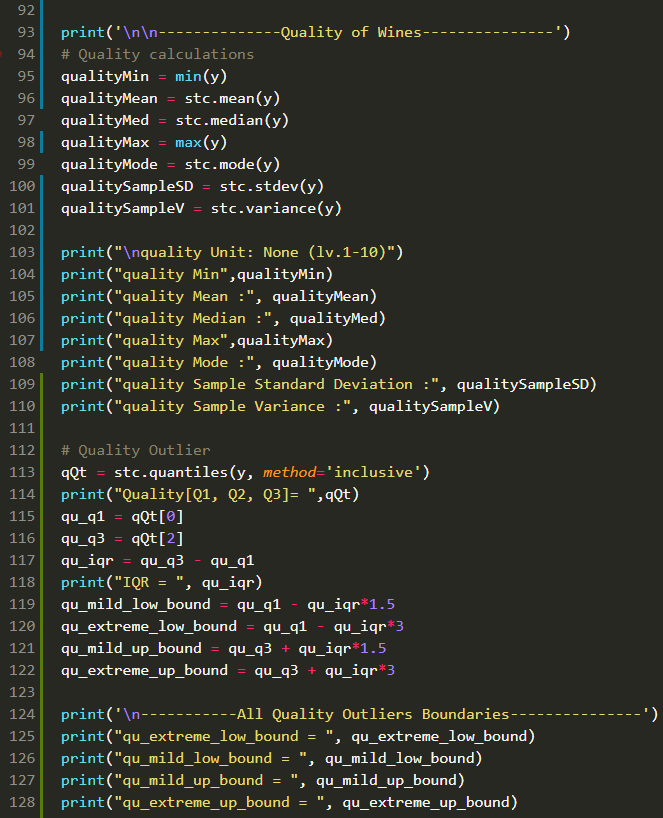
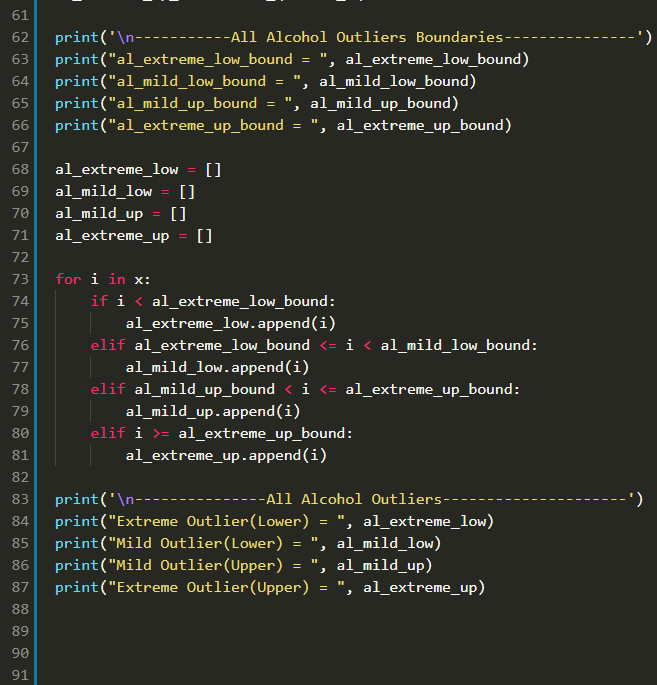
**อย่างไรก็ตาม ทั้งหมดนี้ยังไม่สามารถกล่าวได้อย่างชัดเจน 100% เป็นเพียงแค่แนวโน้มเท่านั้น เนื่องจากในการผลิตไวน์จริง จะมีส่วนผสมอื่นๆ และมีอีกหลายปัจจัยในการกำหนดคุณภาพของไวน์แดง เช่น ค่าความเป็นกรด, น้ำตาลคงค้างที่เหลือในไวน์แดง, ระยะเวลาการผลิตไวน์แดง, คุณภาพขององุ่นที่นำมาใช้ในการผลิต และ เกณฑ์การวัดคุณภาพของไวน์แดง เป็นต้น ซึ่งเกณฑ์การวัดคุณภาพของไวน์แดง(ระดับ 1-10)ในครั้งนี้ อ้างอิงมาจากโรงงานผลิตไวน์ ในจังหวัด Minho ทางตอนเหนือของประเทศโปรตุเกสเท่านั้น**

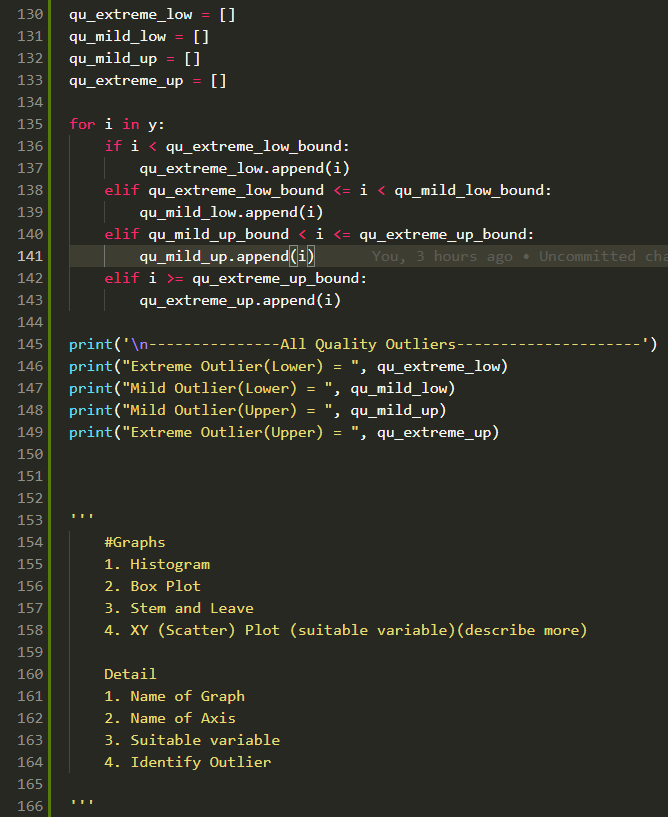


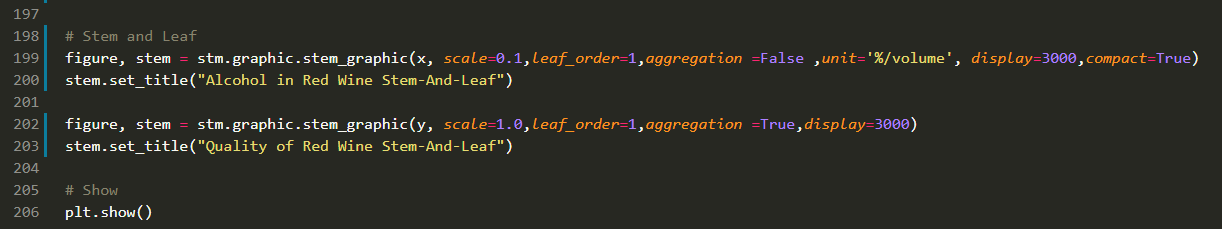
**รายละเอียด Source Code และ OUTPUT ของโปรแกรม WineGraph.py**

****

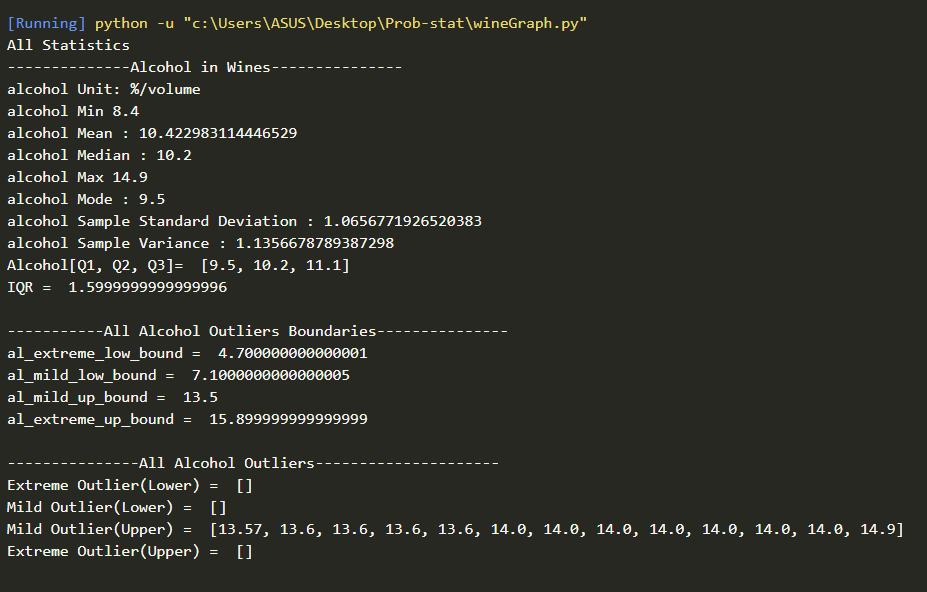
****

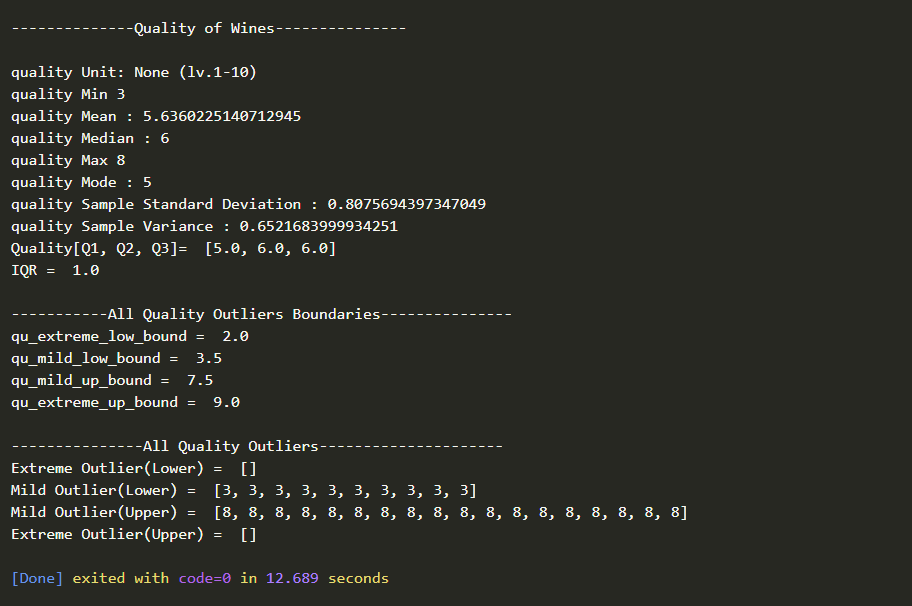
****

****

****

**OUTPUT ของโปรแกรม WineGraph.py**

****

****

**แหล่งที่มาของชุดข้อมูล (Reference/URL) :**

**- ที่มาของชุดข้อมูล** Winequality-red.csv

<https://www.kaggle.com/uciml/red-wine-quality-cortez-et-al-2009>

**- ที่มาคำอธิบายแต่ละส่วนประกอบของไวน์**

<https://archive.ics.uci.edu/ml/datasets/wine+quality>

<https://rstudio-pubs-static.s3.amazonaws.com/57835_c4ace81da9dc45438ad0c286bcbb4224.html>

<https://waterlibrary.com/th-รู้ไหมว่า-ระดับปริมาณแ/#:~:text=ปัจจุบันมีแอลกอฮอล์อยู่ใน,สูงขึ้นด้วยเช่นกัน>

**- วิธีการทำไวน์**

<https://www.youtube.com/watch?v=7gquYRxLMFI&ab_channel=Insider>

**- ประเภทของไวน์**

<https://www.unlockmen.com/terrazas-unlock-wine-101-1/>

<https://thewinelist.shop/blog/news/wine-101>

**- รายละเอียดอื่นๆ เกี่ยวกับคุณภาพและวิธีรับรสที่ดีของไวน์**

<https://www.blockdit.com/posts/5e5f68d77b00780ed6462939>

<https://www.dummies.com/food-drink/drinks/wine/the-special-technique-for-tasting-wine/>

<https://www.quickanddirtytips.com/house-home/entertaining/wine/4-ways-to-know-if-your-wine-is-good>