**Unsupervised learning กับ Supervised learning คืออะไร แตกต่างกันอย่างไร**

- Supervised learning (การเรียนรู้แบบมีผู้สอน) คือ การนำ data มา train ใน model เพื่อทำนายผลอะไรบางอย่างโดย data ที่นำมา train นั้นจะมีการระบุผลลัพธ์อยู่แล้ว

- Unsupervised learning (การเรียนรู้แบบมีผู้สอน) คือ การนำ data มา train ใน model เพื่อให้เห็นรูปแบบหรือลักษณะของกลุ่มข้อมูลนั้น โดยจะแตกต่างกับ supervised learning ในส่วน data ที่นำมา train นั้น จะเป็น data เปล่าๆอย่างเดียวไม่มีการระบุผลลัพธ์ที่ต้องการ

**จงเปรียบเทียบการจัดกลุ่ม แบบ centroid based clustering (เช่น k-means) กับ การจัดกลุ่มแบบ density-based clustering (เช่น DBSCAN) ว่ามีจุดเด่น จุดด้อย และเหมาะกับข้อมูลรูปแบบใด**

|  |  |  |  |
| --- | --- | --- | --- |
| ชนิด | จุดเด่น | จุดด้อย | เหมาะกับ |
| centroid based clustering | สามารถแบ่งกลุ่มให้ข้อมูลทุกตัว | ไม่สามารถแบ่งกลุ่มโดยมีกลุ่มหนึ่งอยู่ภายในอีกกลุ่มหนึ่ง | ข้อมูลวางตัวเกาะกลุ่มกันอย่างหนาแน่นเป็นวงกลม |
| density-based clustering | สามารถแบ่งกลุ่มโดยมีกลุ่มหนึ่งอยู่ภายในอีกกลุ่มหนึ่ง | ไม่สามารถแบ่งกลุ่มให้ข้อมูลทุกตัว (ข้อมูลบางตัวจะไม่ถูกจัดอยู่ในกลุ่มไหนเลย) | ข้อมูลวางตัวเกาะกลุ่มกันอย่างหนาแน่นและไม่เป็นรูปลักษณ์ที่ตายตัว |

**สรุปขั้นตอนการสร้างกราฟด้วย Matplotlib และ Seaborn**

- import library

- กำหนดชนิดของกราฟ set ค่า x และ y และอื่นๆ เช่น ลักษณะของเส้น, สี ฯลฯ

- .show() เพื่อแสดงผล (สำหรับ Matplotlib อย่างเดียว)

**จงเปรียบเทียบไลบารี่ Matplotlib และ Seaborn ในแง่ของการใช้งานและความสามารถ ว่ามีความเหมือนหรือแตกต่างกันอย่างไร**

ความเหมือน

ขั้นตอนการสร้างกราฟเหมือนกัน

ความต่าง

Seaborn ไม่จำเป็นต้องกำหนดรายละเอียดเองทั้งหมด สามารถแสดงการกระจายของข้อมูล

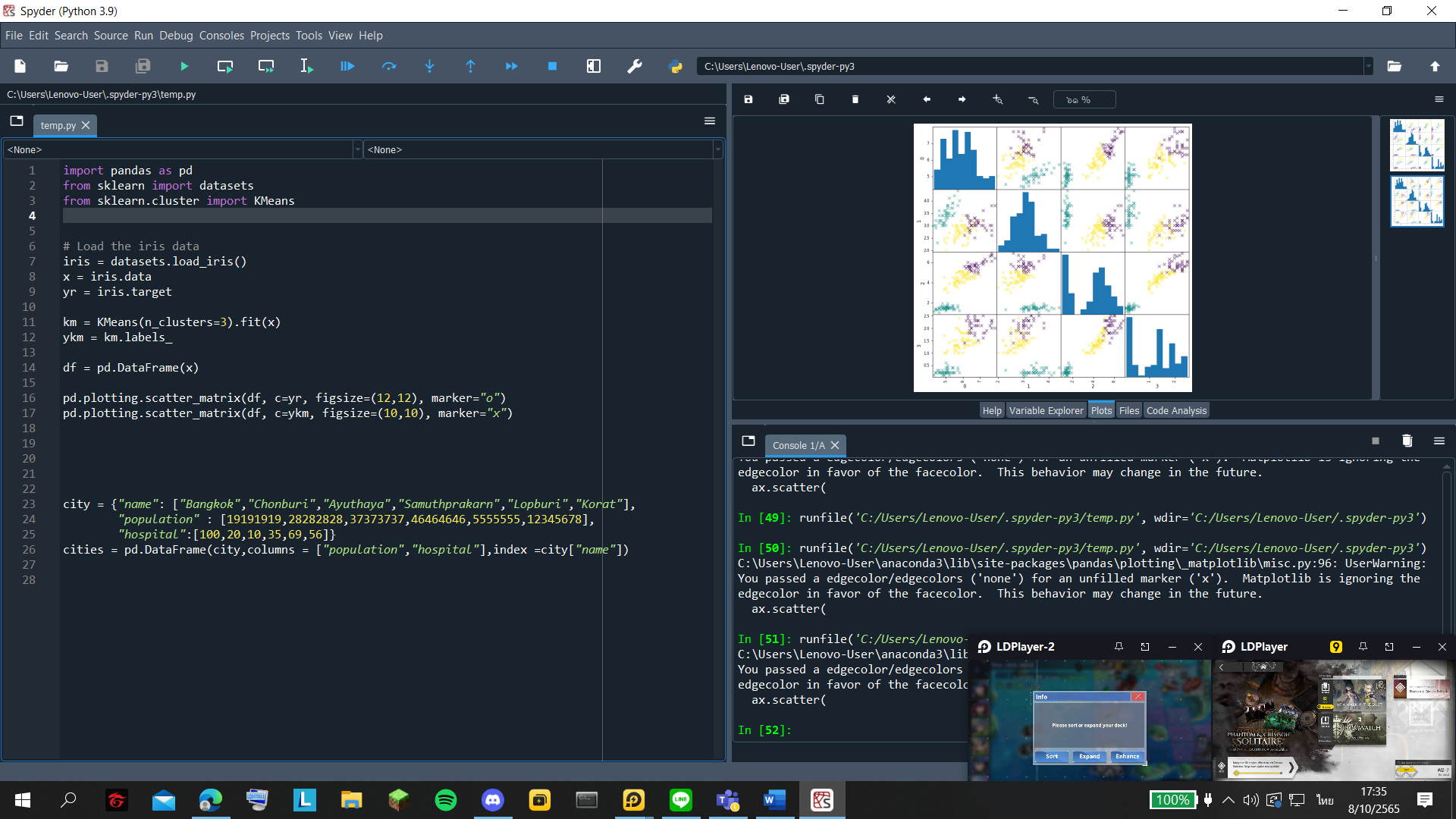
Matplotlib จำเป็นต้องกำหนดรายละเอียดเองทั้งหมดทำให้การใช้งานจะมีความละเอียดที่มากกว่า

**จงสรุปแนวคิดหลักของอัลกอริทึมการจัดกลุ่มแบบ centroid based clustering ที่ชื่อ k-Means พร้อมยกตัวอย่างการจัดกลุ่มข้อมูล iris dataset ด้วยอัลกอริทึม k-Means โดยใช้ไลบารี่ sklearn**

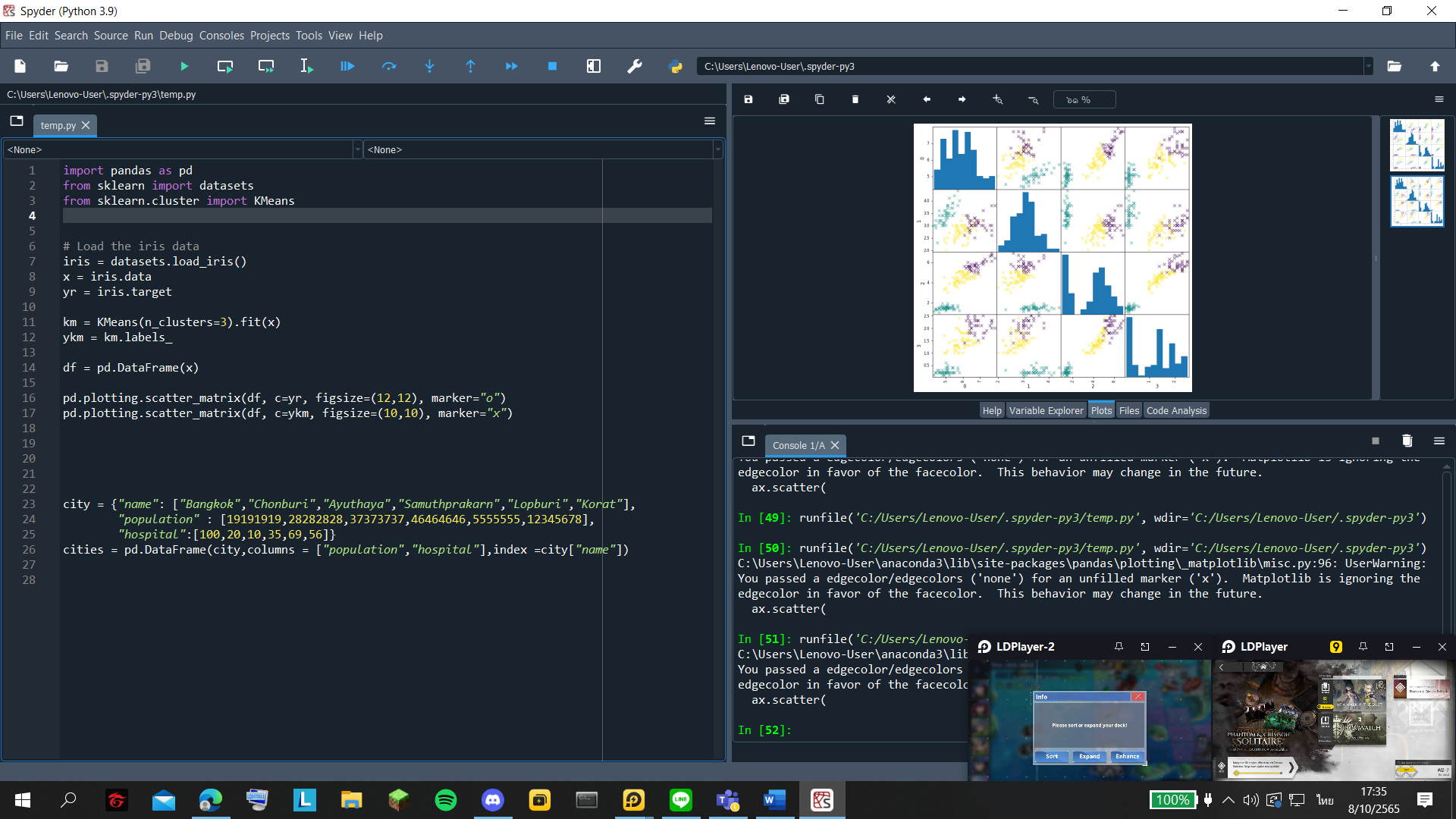
- หาจุดศูนย์กลางของกลุ่มแล้วจัดกลุ่มข้อมูลโดยให้ข้อมูลอยู่ในกลุ่มที่ข้อมูลอยู่ใกล้กับจุดศูนย์กลางนั้นมากที่สุด

การจัดกลุ่มข้อมูล iris dataset

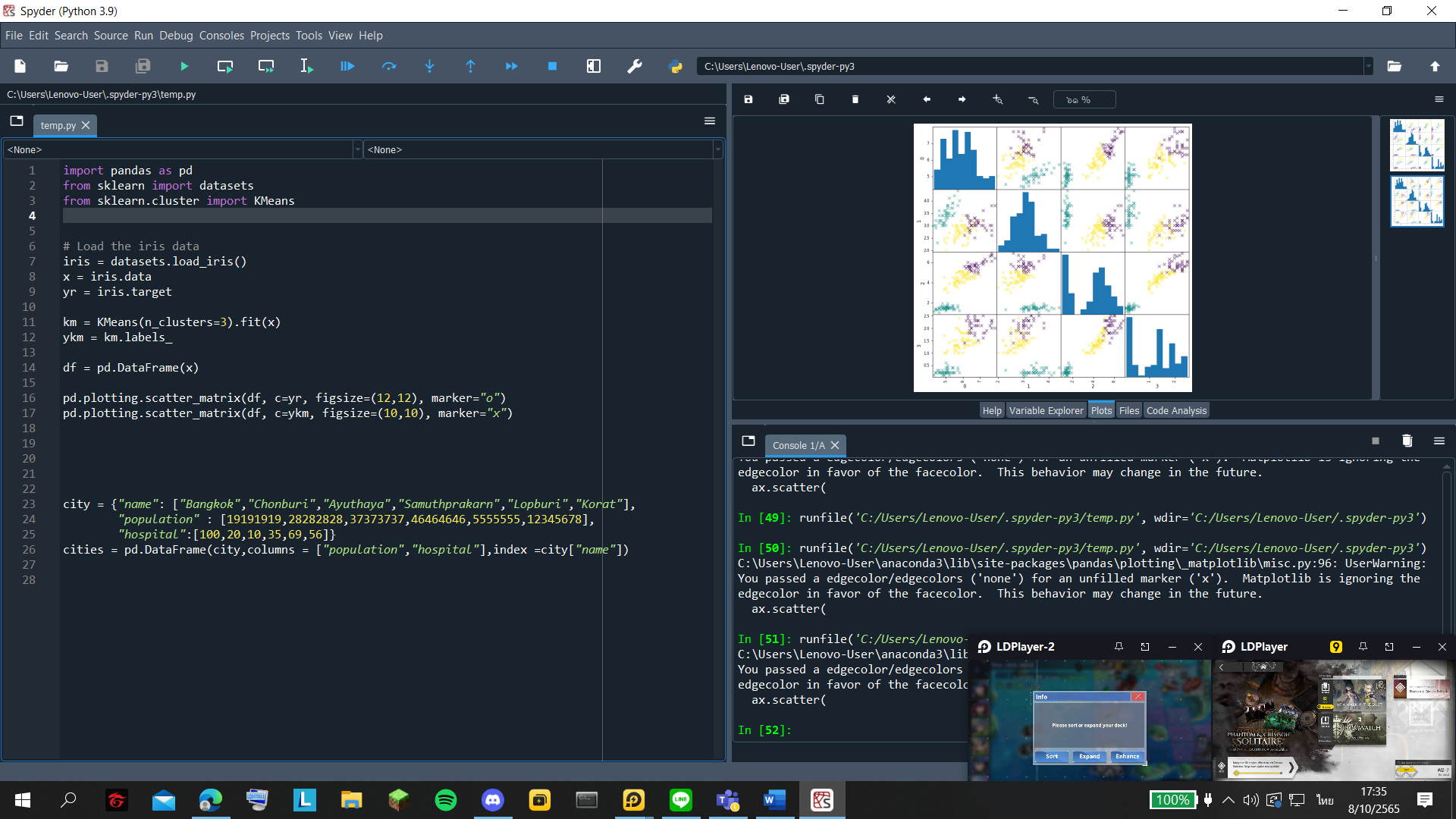
1. import dataset

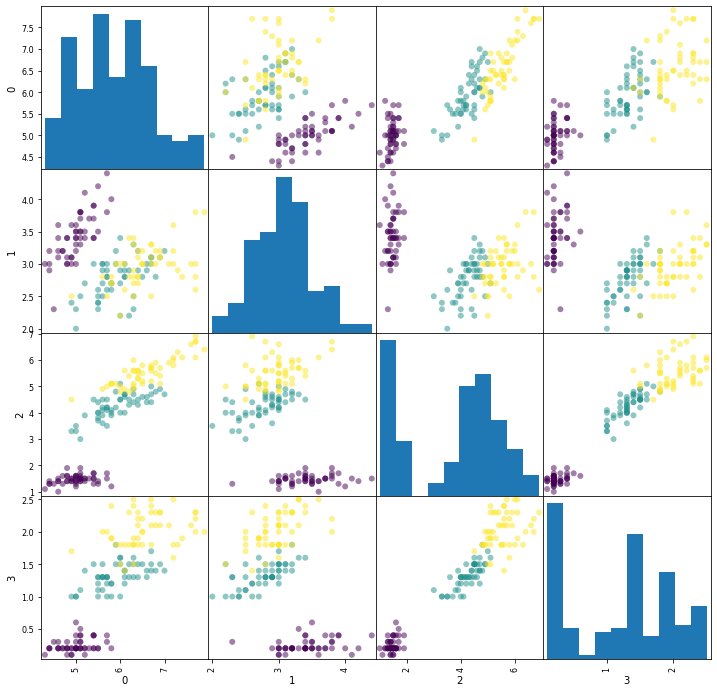
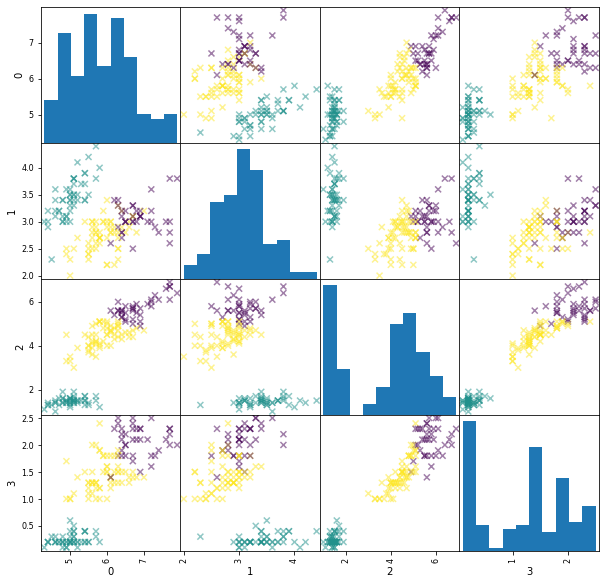


2. ใช้อัลกอริทึม k-Means แบ่งกลุ่ม



3. แสดงผลการจัดกลุ่ม





ภาพข้อมูลที่ใช้อัลกอริทึม k-Means ในการแบ่งกลุ่ม และข้อมูลจริง