

COS3302 LAB#5

# การวิเคราะห์ทางสถิติและ การนำเสนอภาพข้อมูล

จากชุดข้อมูล

## IRIS DATASET

จัดทำโดย

นายศุภณัฐ แซ่เตีย

ID: 6505000270



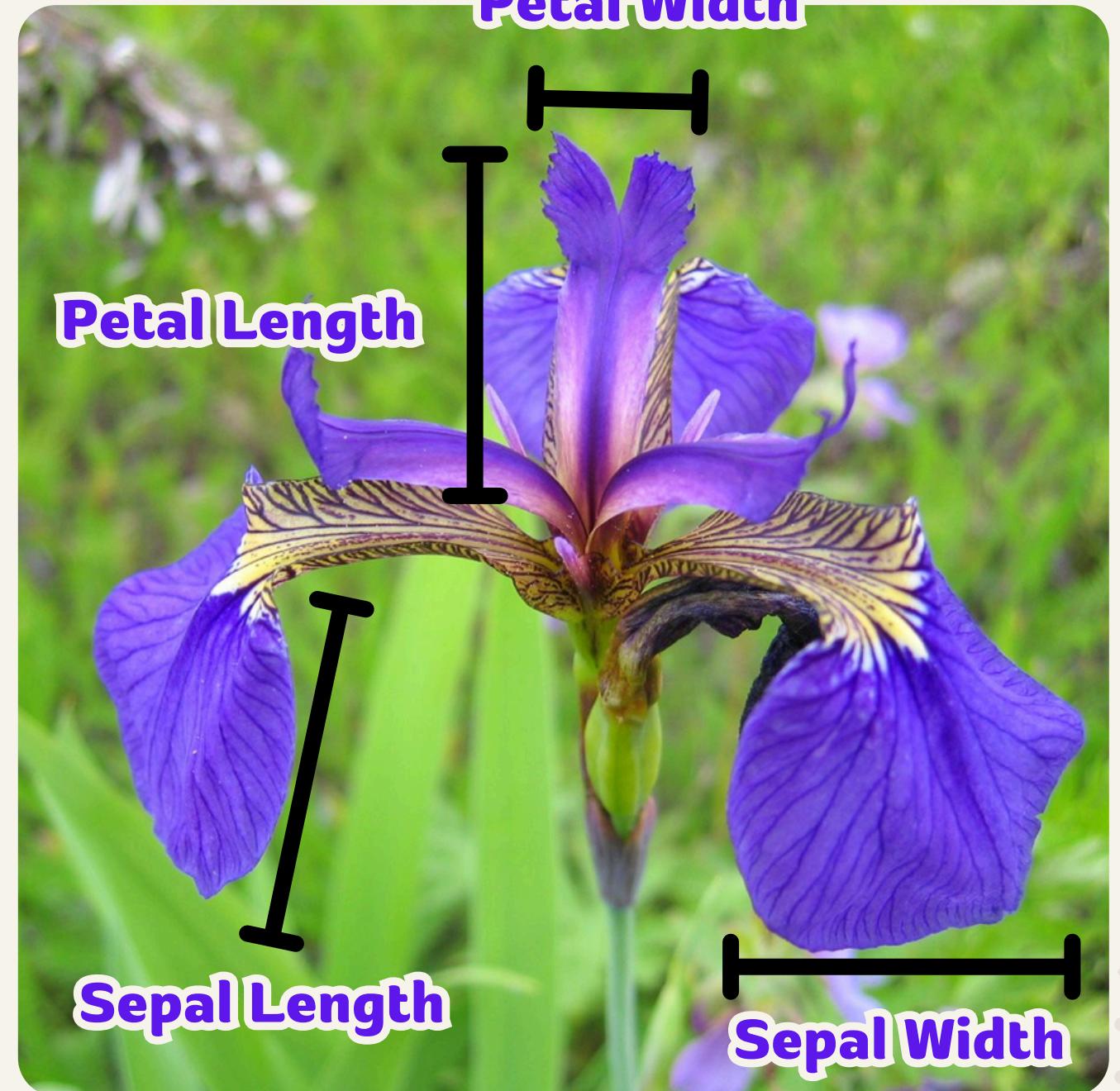
# AGENDA

- ที่มาของและลักษณะของชุดข้อมูล
- คุณสมบัติของแต่ละ Feature
- ความสัมพันธ์ของแต่ละ Feature
- Data Preprocess

# ລັກເບດນະຂອງຫຼຸດຂ້ອມູລ

# ลักษณะของชุดข้อมูล

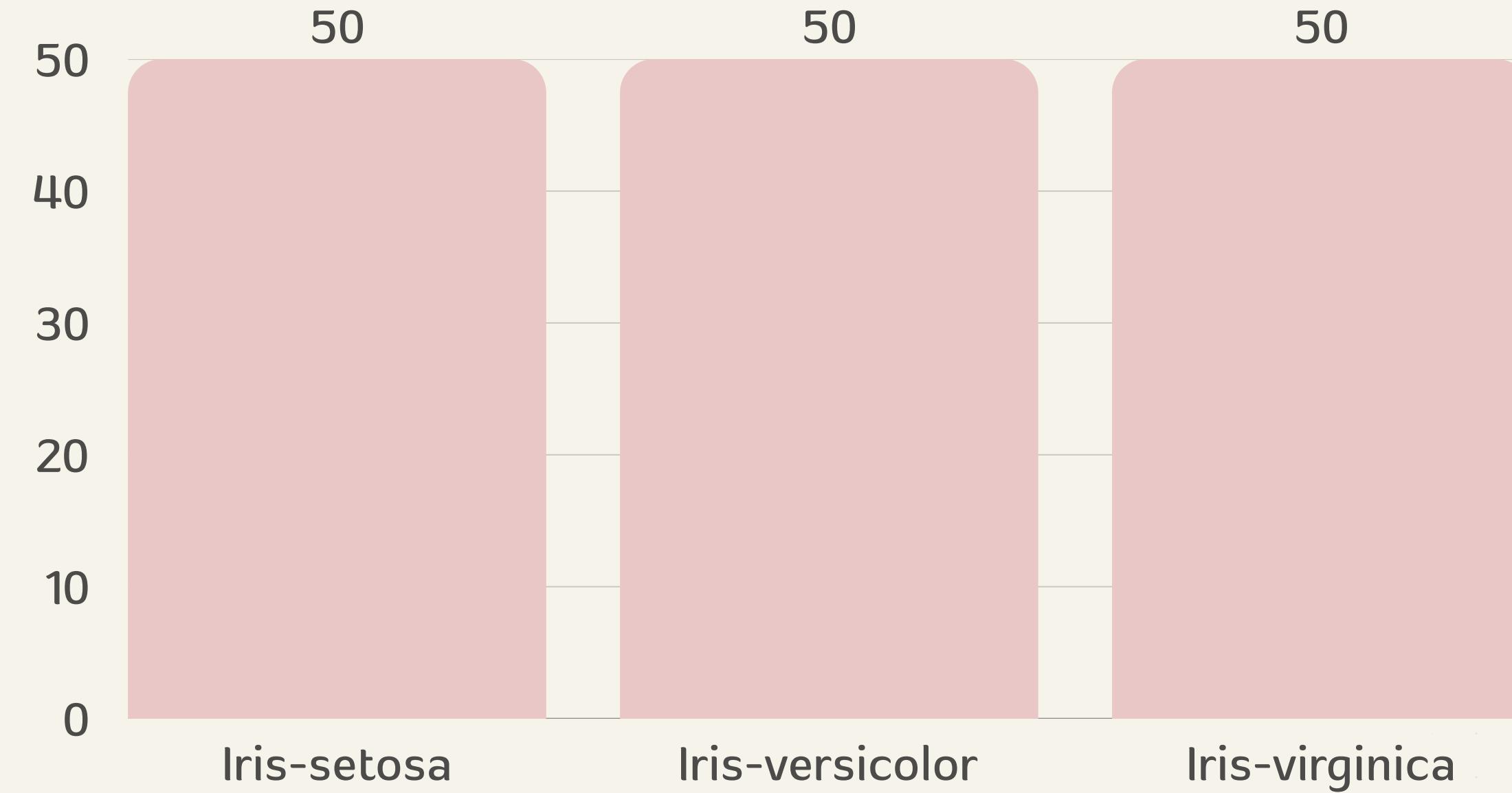
- เป็นข้อมูลของดอกไม้สกุล Iris 3 สายพันธุ์ จำนวนทั้งหมด 150 ดอกซึ่งถูกเก็บรวบรวมจากภาคสมุทรภาคแป ประเทศไทยและแคนาดา โดยคุณเออดการ์ แวนเดอร์ลัน
- ชุดข้อมูลนี้ประกอบด้วย 5 Feature ได้แก่
  - Class
  - Sepal Length
  - Sepal Width
  - Petal Length
  - Petal Width



# คุณสมบัติของแต่ละ FEATURE

# ชนิดของดอก (CLASS)

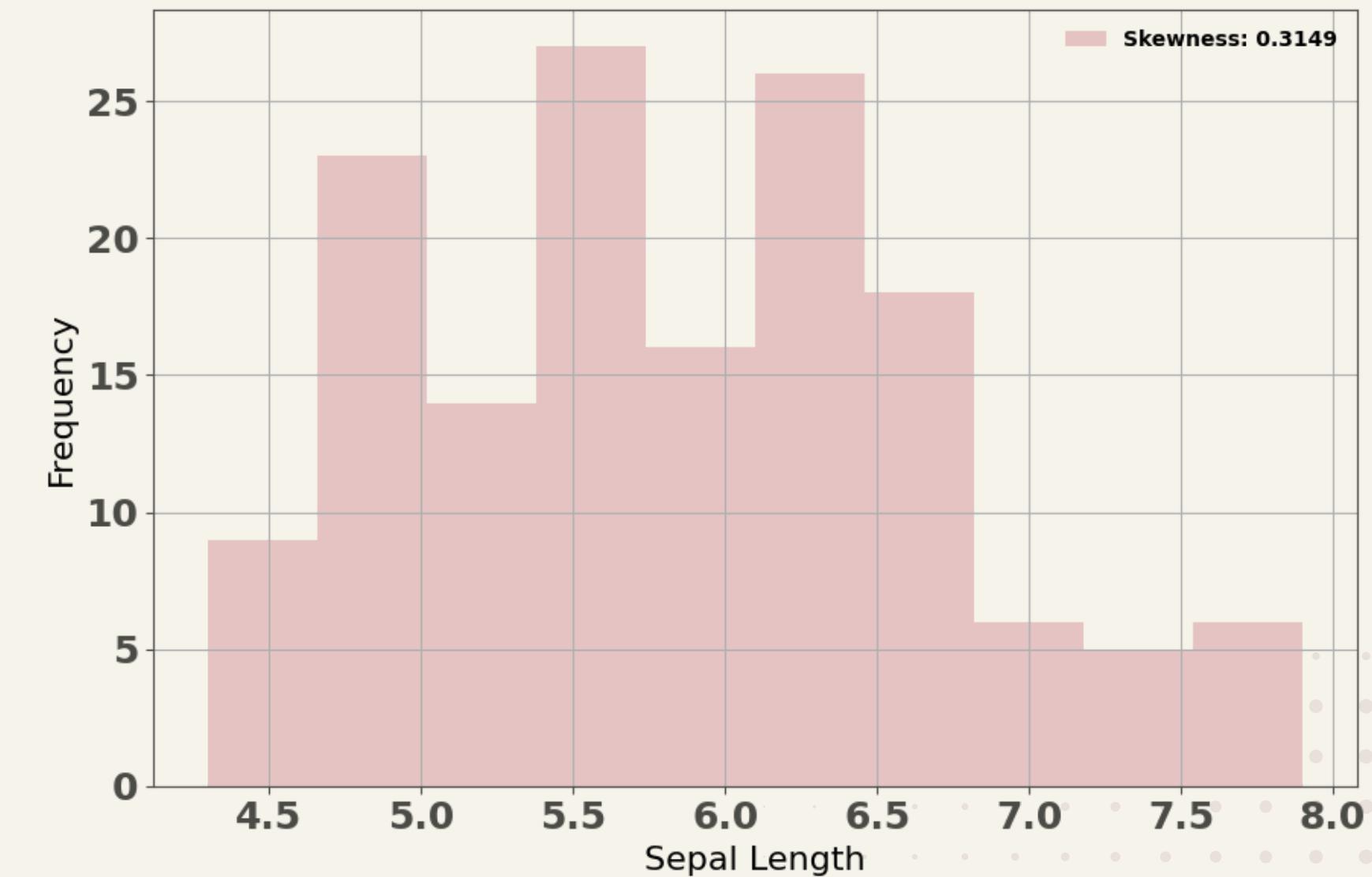
- เป็น Feature ที่แสดงเกี่ยวกับชนิดของดอกไม้ในชุดข้อมูล
- Feature นี้เป็นข้อมูลประเภทเชิงคุณภาพ (Categorical) แบบไม่มีลำดับ (Nominal)
- โดยประกอบด้วย 3 หมวดหมู่ ได้แก่ Iris-setosa, Iris-versicolor และ Iris-virginica



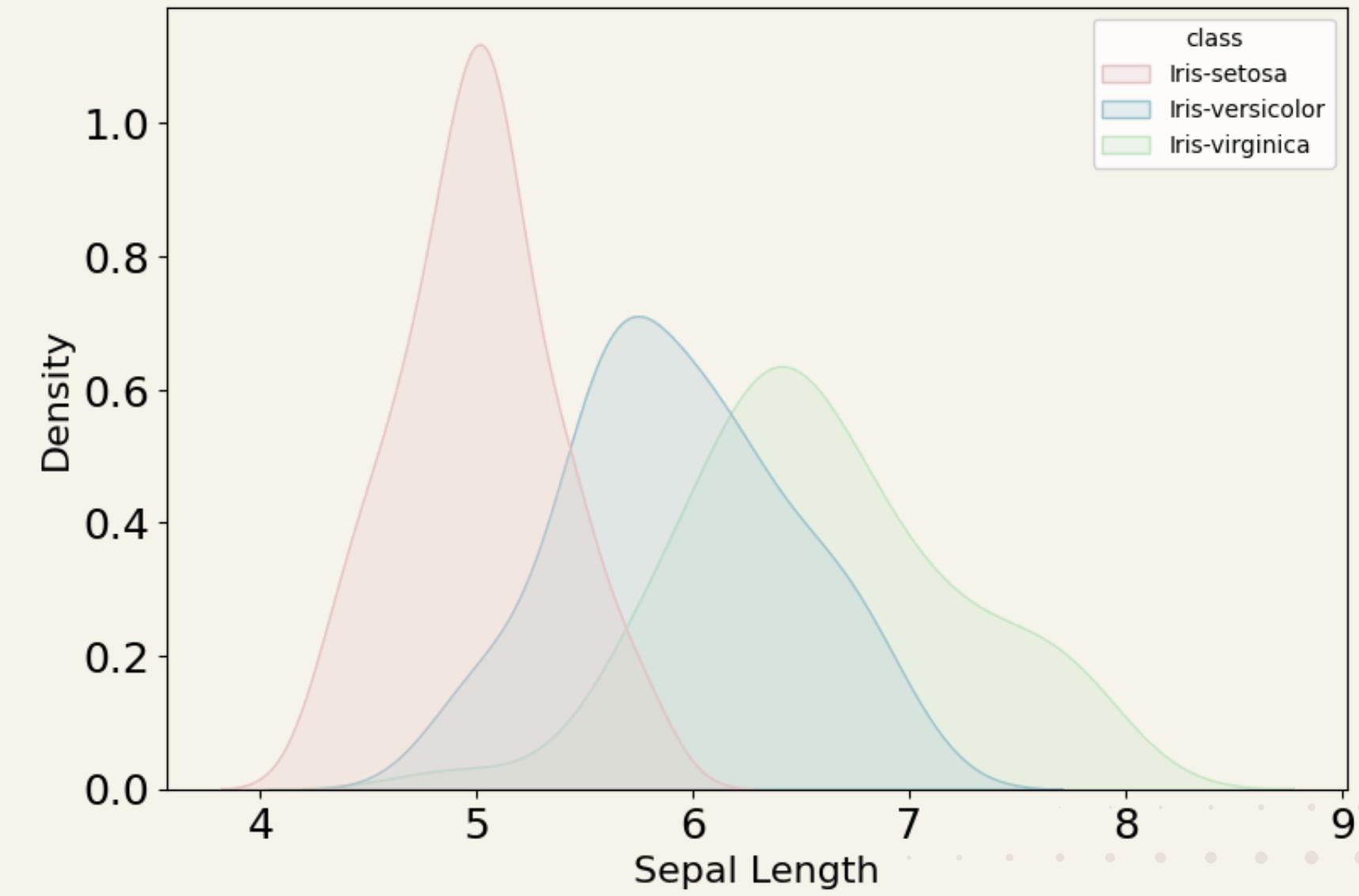
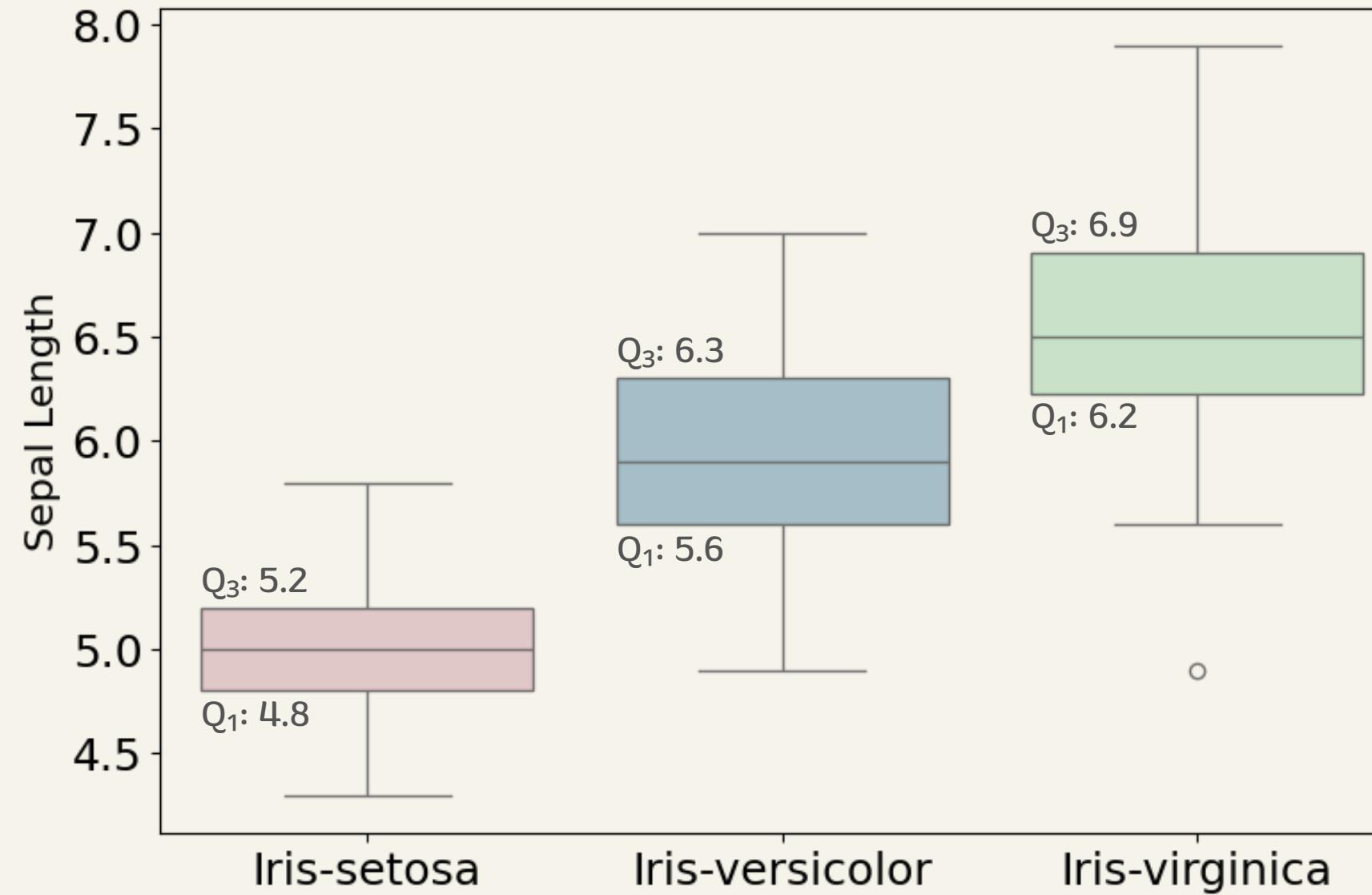
# SEPAL LENGTH

- เป็น Feature ประเภทเชิงปริมาณ (Numerical) ชนิดอัตราส่วน (Ratio)
- กำหนดว่าที่แสดงความยาวของกลีบเลี้ยงโดยมีหน่วยเป็นเซนติเมตร
- ข้อมูลมีลักษณะเบ้าไปทางขวาเล็กน้อย (Right-skewed)

Mean	5.84	$Q_1$	5.1
Median	5.8	$Q_2$	5.8
S.D.	0.83	$Q_3$	6.4
Min	4.3	Max	7.9



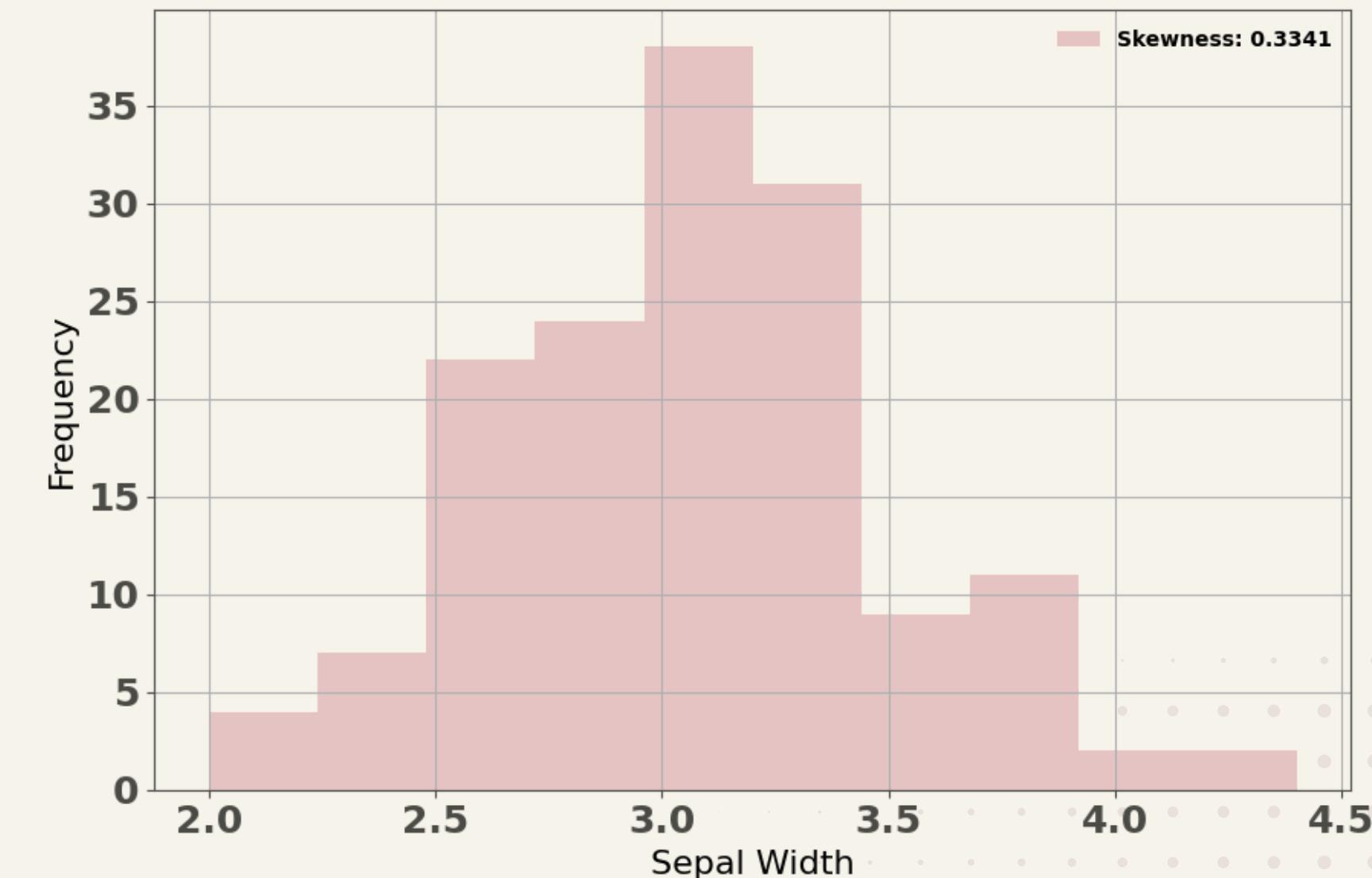
# SEPAL LENGTH



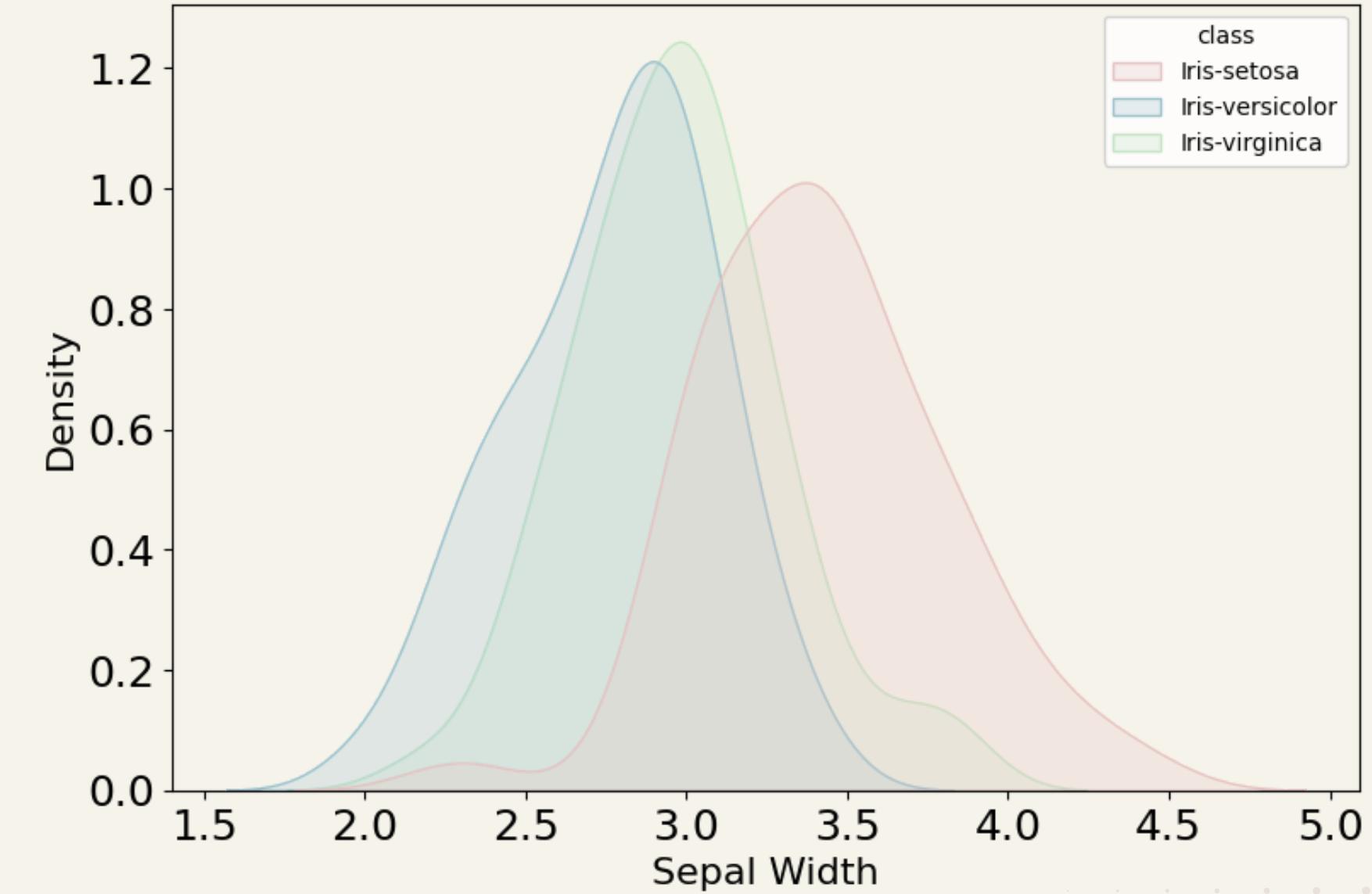
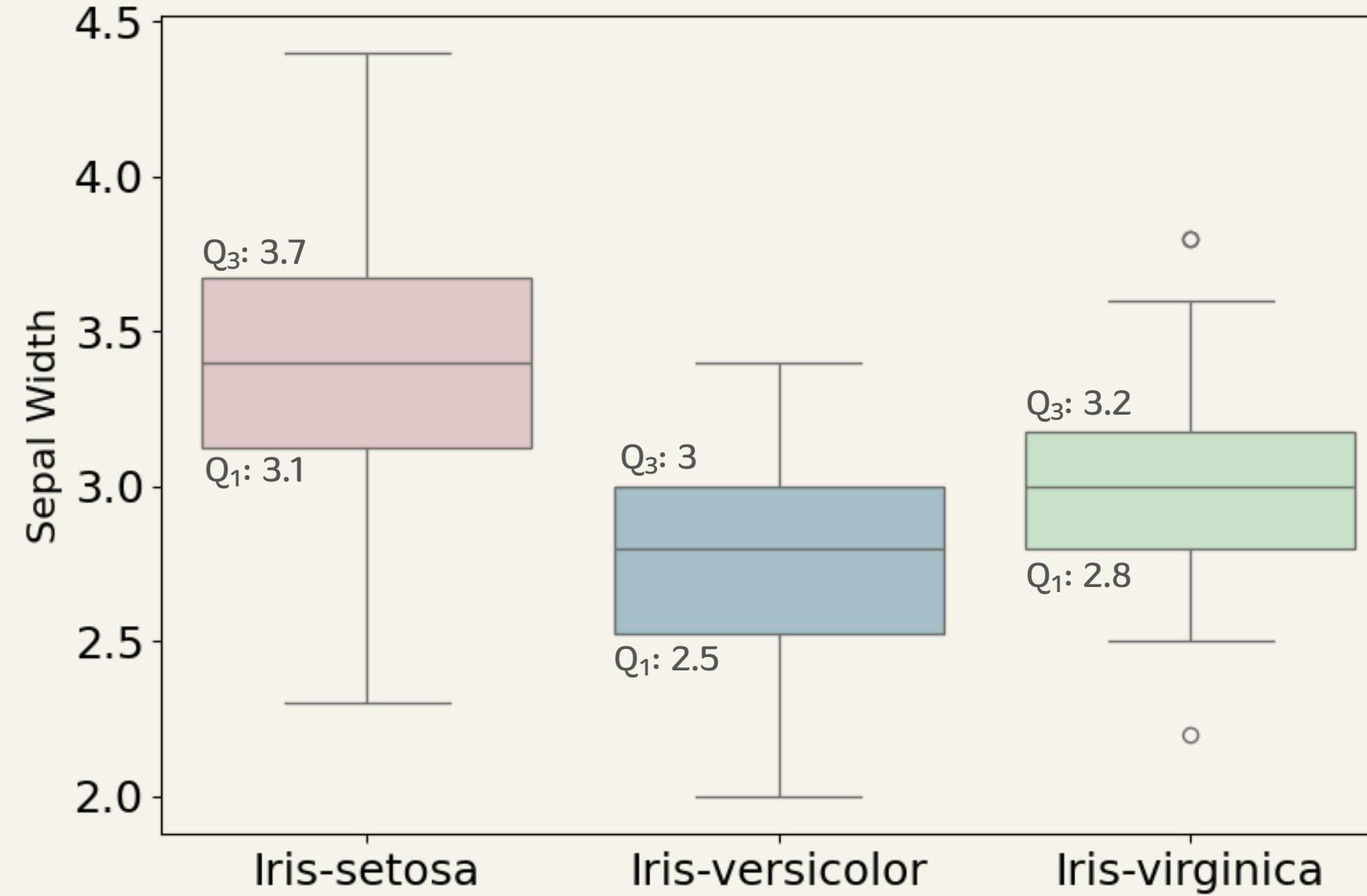
# SEPAL WIDTH

- เป็น Feature ประเภทเชิงปริมาณ (Numerical) ชนิดอัตราส่วน (Ratio)
- กำหนดให้แสดงความกว้างของกลีบเลี้ยงโดยมีหน่วยเป็นเซนติเมตร
- ข้อมูลมีลักษณะเบ้าไปทางขวาเล็กน้อย (Right-skewed)

Mean	3.05	$Q_1$	2.8
Median	3	$Q_2$	3
SD.	0.43	$Q_3$	3.3
Min	2	Max	4.4



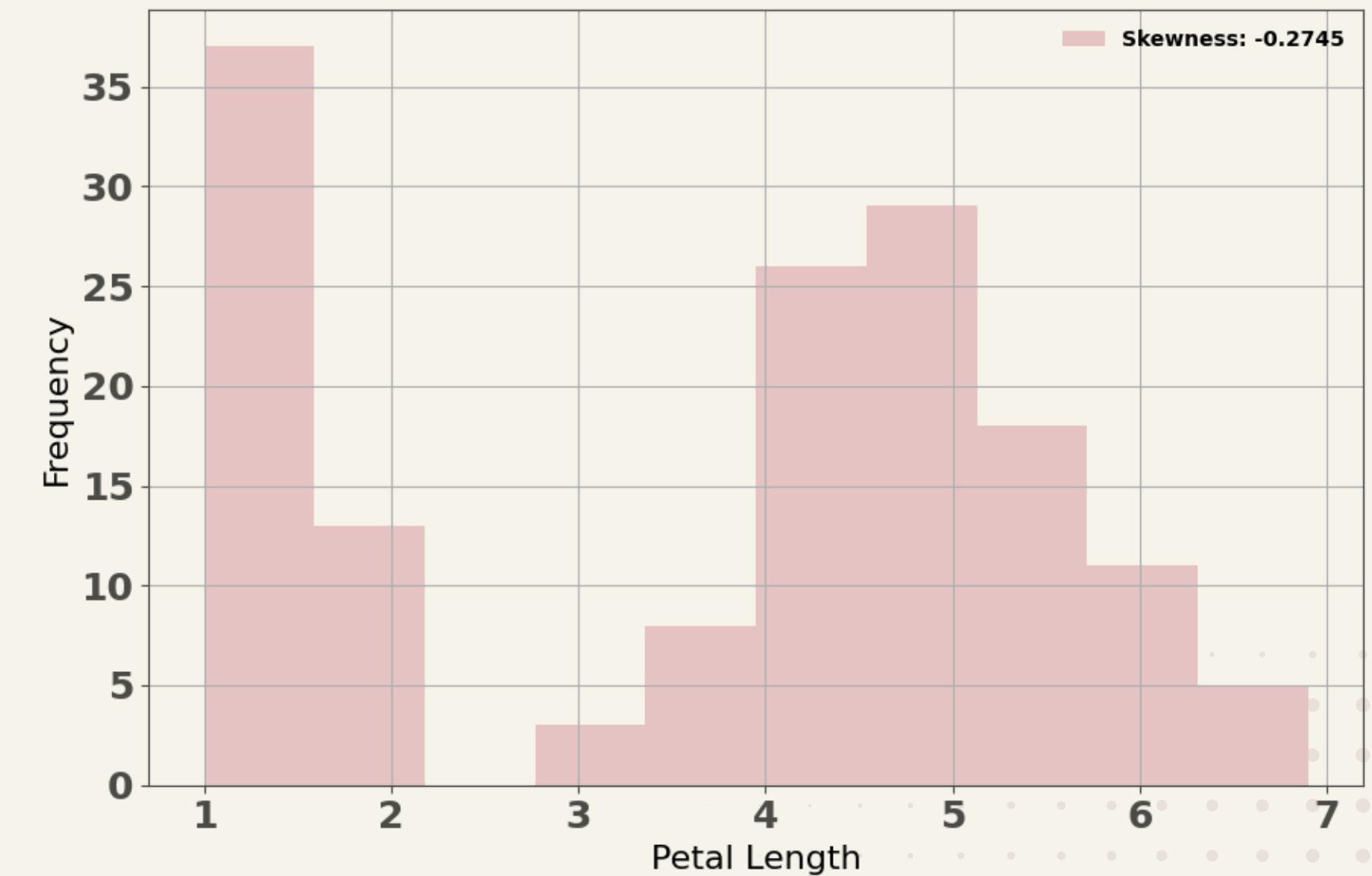
# SEPAL WIDTH



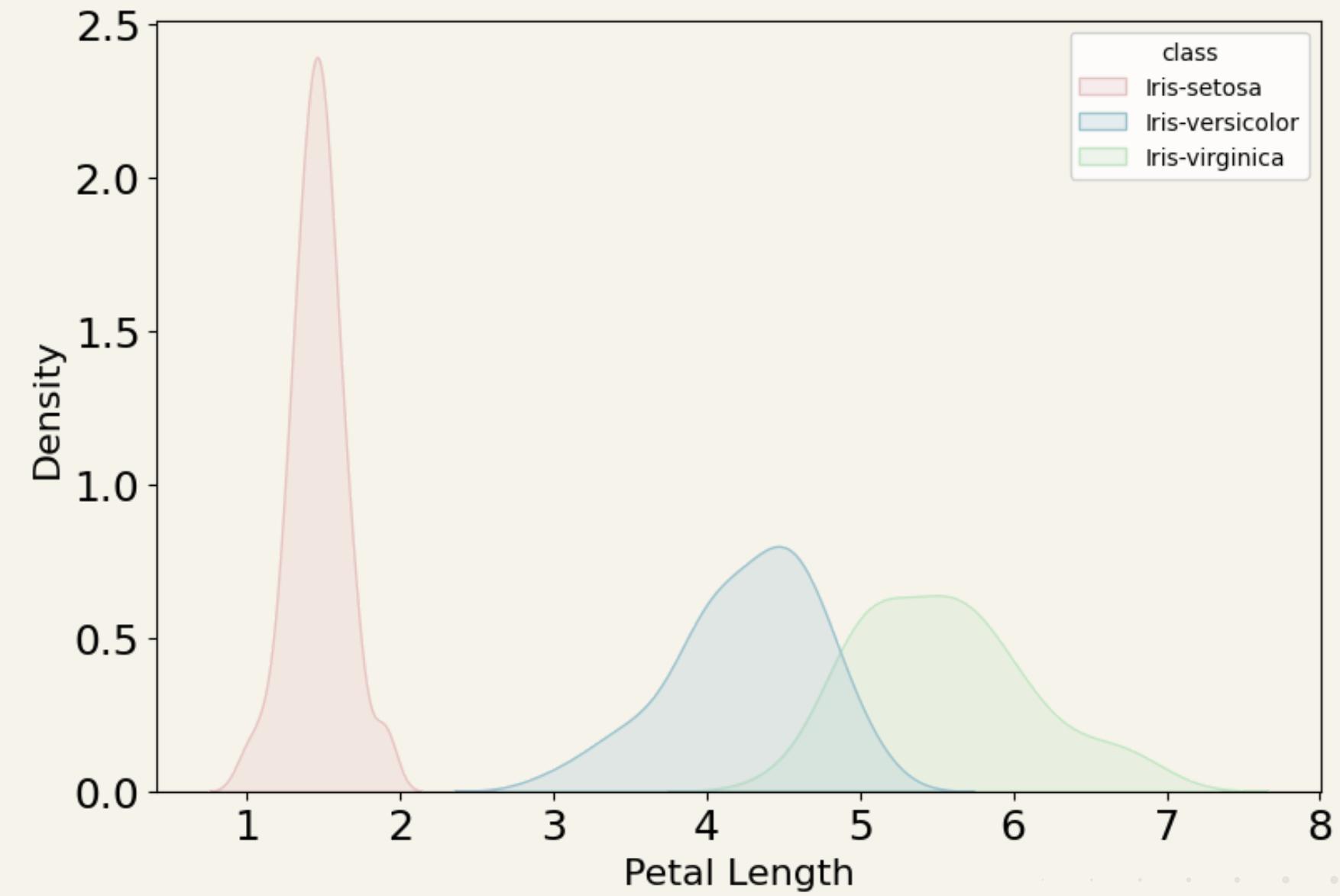
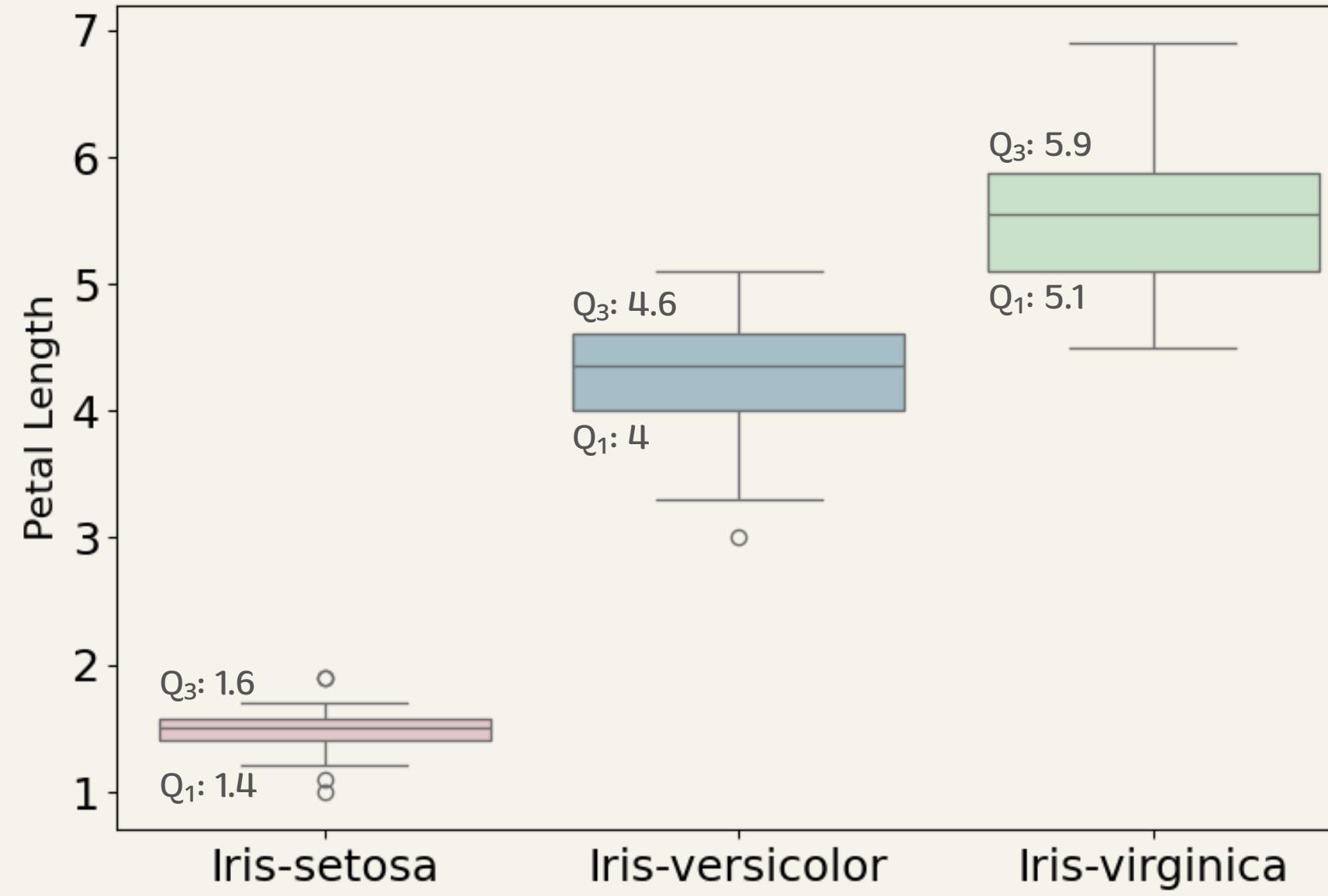
# PETAL LENGTH

- เป็น Feature ประเภทเชิงปริมาณ (Numerical) ชนิดอัตราส่วน (Ratio)
- กำหนดว่าที่แสดงความยาวของกลีบดอกโดยมีหน่วยเป็นเซนติเมตร
- ข้อมูลมีลักษณะเบ้าไปทางซ้ายเล็กน้อย (Left-skewed)

Mean	3.7587	$Q_1$	1.6
Median	4.35	$Q_2$	4.35
SD.	1.7644	$Q_3$	5.1
Min	1	Max	6.9



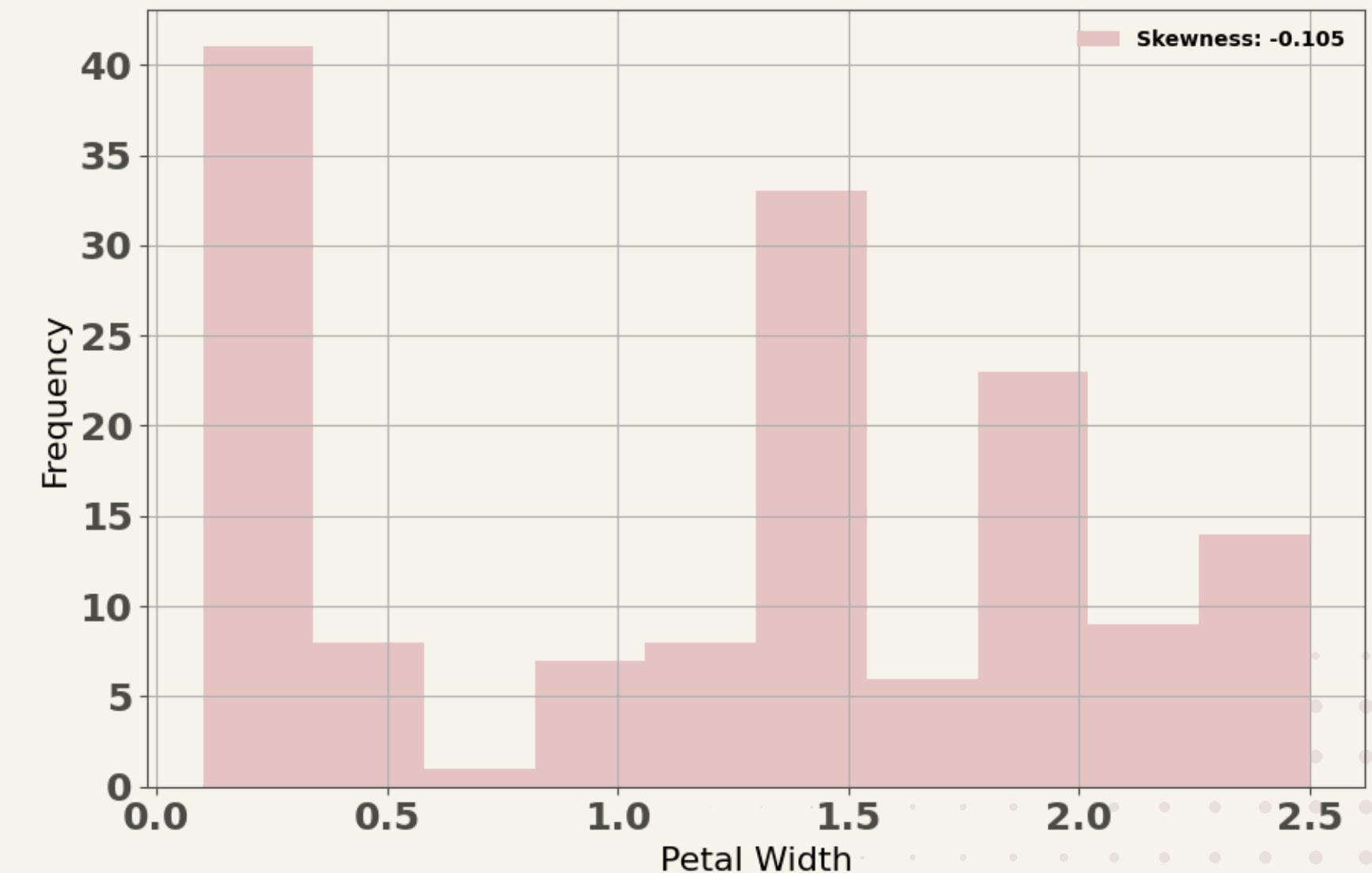
# PETAL LENGTH



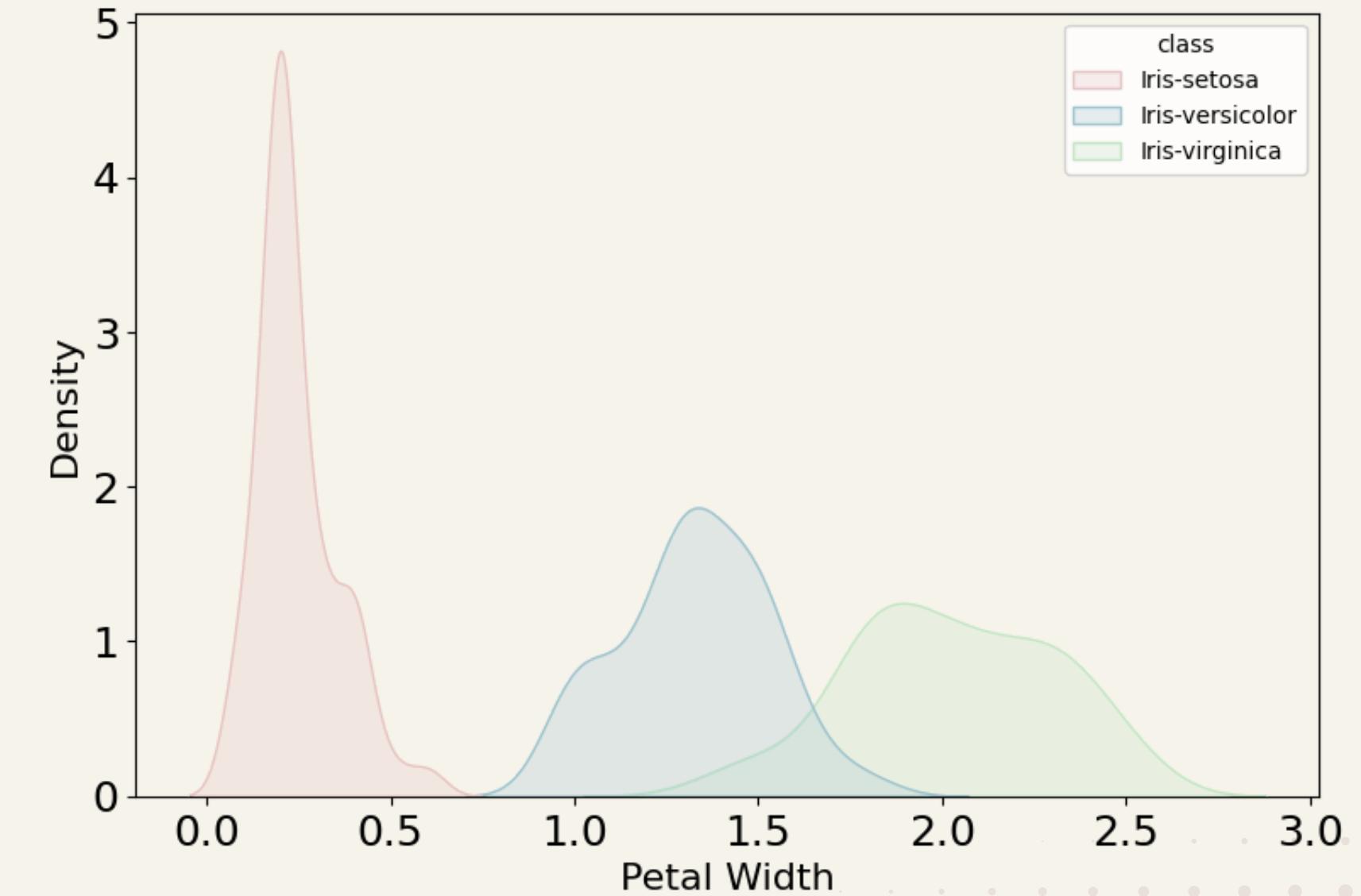
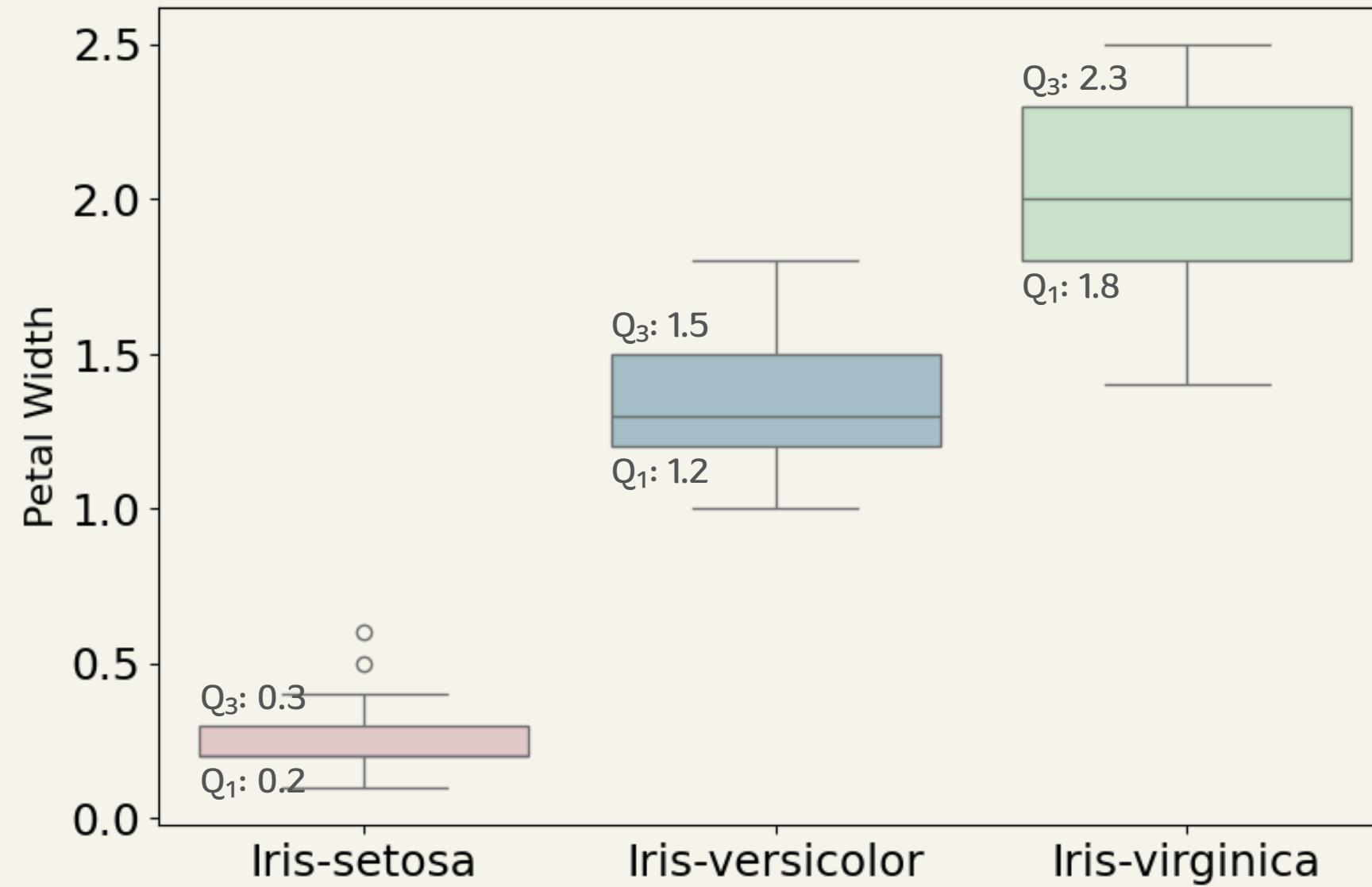
# PETAL WIDTH

- เป็น Feature ประเภทเชิงปริมาณ (Numerical) ชนิดอัตราส่วน (Ratio)
- กำหนดว่าที่แสดงความกว้างของกลีบดอกโดยมีหน่วยเป็นเซนติเมตร
- ข้อมูลมีลักษณะเบ้าไปทางซ้ายเล็กน้อย (Left-skewed)

Mean	1.1987	$Q_1$	0.3
Median	1.3	$Q_2$	1.3
SD.	0.7632	$Q_3$	1.8
Min	0.1	Max	2.5

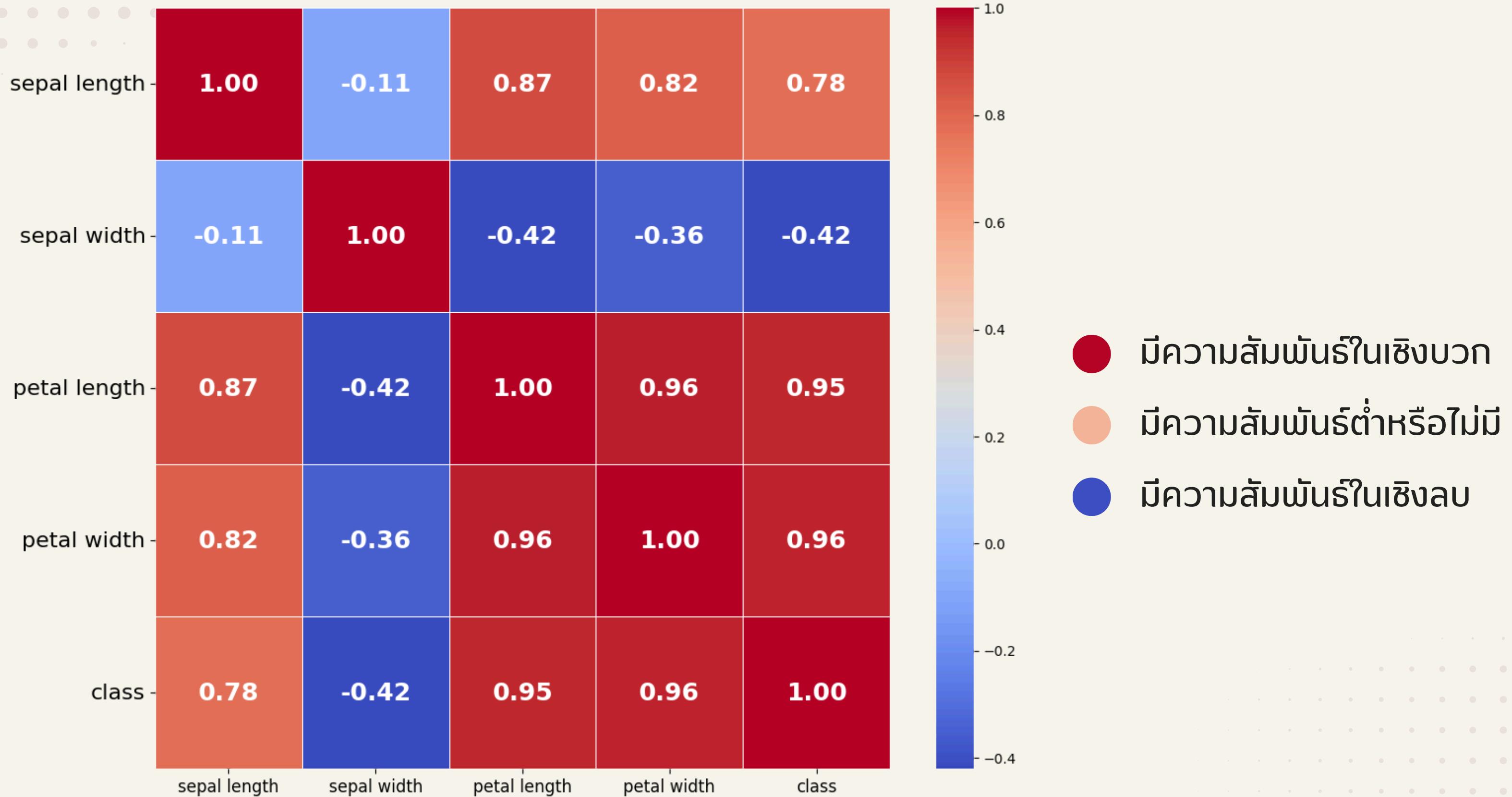


# PETAL WIDTH



# ความสัมพันธ์ของแต่ละ FEATURE

# ความสัมพันธ์ของแต่ละ FEATURE



# DATA PREPROCESS

# FEATURE SCALING

- ทำให้ข้อมูลในแต่ละ Feature มีความใกล้เคียงกัน โดยใช้ Min-Max Scaling ทำให้ข้อมูลอยู่ในช่วง 0 - 1

	sepal length	sepal width	petal length	petal width	class
0	5.1	3.5	1.4	0.2	Iris-setosa
1	4.9	3.0	1.4	0.2	Iris-setosa
2	4.7	3.2	1.3	0.2	Iris-setosa
3	4.6	3.1	1.5	0.2	Iris-setosa
4	5.0	3.6	1.4	0.2	Iris-setosa

ข้อมูลก่อนทำ Feature Scaling



	sepal length	sepal width	petal length	petal width	class
0	0.222222	0.625000	0.067797	0.041667	Iris-setosa
1	0.166667	0.416667	0.067797	0.041667	Iris-setosa
2	0.111111	0.500000	0.050847	0.041667	Iris-setosa
3	0.083333	0.458333	0.084746	0.041667	Iris-setosa
4	0.194444	0.666667	0.067797	0.041667	Iris-setosa

ข้อมูลหลังทำ Feature Scaling



# CLASS ENCODING

- ใช้ Label Encoding ในการแปลง Feature “Class” ให้เป็นประเภท Numerical
- โดย 0 แทน Iris-setosa, 1 แทน Iris-versicolor และ 2 แทน Iris-virginica

	sepal length	sepal width	petal length	petal width	class
0	0.222222	0.625000	0.067797	0.041667	Iris-setosa
1	0.166667	0.416667	0.067797	0.041667	Iris-setosa
2	0.111111	0.500000	0.050847	0.041667	Iris-setosa
3	0.083333	0.458333	0.084746	0.041667	Iris-setosa
4	0.194444	0.666667	0.067797	0.041667	Iris-setosa

ข้อมูลหลังทำ Class Encoding

ข้อมูลก่อนทำ Class Encoding

	sepal length	sepal width	petal length	petal width	class
0	0.222222	0.625000	0.067797	0.041667	0
1	0.166667	0.416667	0.067797	0.041667	0
2	0.111111	0.500000	0.050847	0.041667	0
3	0.083333	0.458333	0.084746	0.041667	0
4	0.194444	0.666667	0.067797	0.041667	0

# ขอบคุณครับ

**6505000270 นายคุกันฐ แซ่เตีย**

# อ้างอิง

- รศ. ดร. เชณูเนติ ศรีสอ้าน. (2564). ตัวอย่างสุด Classic ของ นักวิทยาศาสตร์ข้อมูล ตอน การวิเคราะห์แบ่งกลุ่มของดอกไม้สกุล Iris ด้วยการเรียนรู้ของเครื่อง. สืบค้นเมื่อ 28 มกราคม 2568, จาก <https://chetnetisrisaan.medium.com/ตัวอย่างสุด-classic-ของ-นักวิทยาศาสตร์ข้อมูล-ตอน-การวิเคราะห์แบ่งกลุ่มของดอกไม้สกุล-iris-1e33da940957>
- OpenAI. (2567). ChatGPT (Version 4) [Large language model]. สืบค้นเมื่อ 28 มกราคม 2568 , จาก <https://chatgpt.com/>