

COS3302 LAB#5

การวิเคราะห์ทางสถิติและ การนำเสนอภาพข้อมูล

จากชุดข้อมูล

Iris Dataset

จัดทำโดย

นายศุภณัฐ แซ่เตีย

ID: 6505000270



OVERVIEW

- ลักษณะของชุดข้อมูล
- คุณสมบัติของแต่ละ Feature
- การทำ Data Preprocess
 - Normalization
 - Class Encoding

ລັກບະນະຂອງຫຼຸດຂ້ອມູລ

ลักษณะของชุดข้อมูล

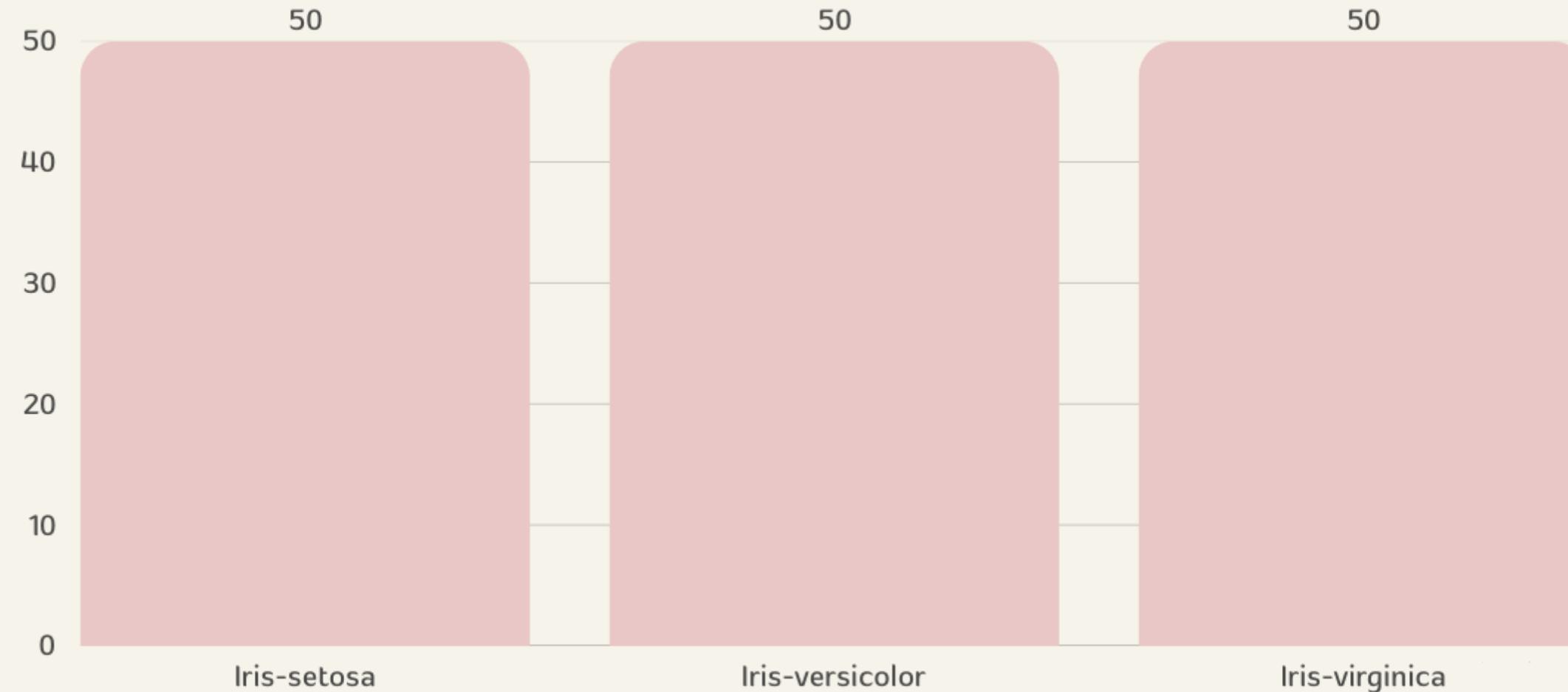
- เป็นข้อมูลของดอกไม้สกุล Iris 3 สายพันธุ์ จำนวนทั้งหมด 150 ดอกซึ่งถูกเก็บรวบรวมจากภาคสมุทรภาคแป ประเทศไทยและแคนาดา โดยคุณเอ็ด-การ์ แวนเดอร์สัน
- ชุดข้อมูลนี้ประกอบด้วย 5 Feature ดังนี้
 - สายพันธุ์ของดอก (Class)
 - ความยาวของกลีบเลี้ยง (Sepal Length)
 - ความกว้างของกลีบเลี้ยง (Sepal Width)
 - ความยาวของกลีบดอก (Petal Length)
 - ความกว้างของกลีบดอก (Petal Width)

	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
...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

គុណសមប័តិថរុងនៃព័ត៌មាន FEATURE

ชนิดของดอก (CLASS)

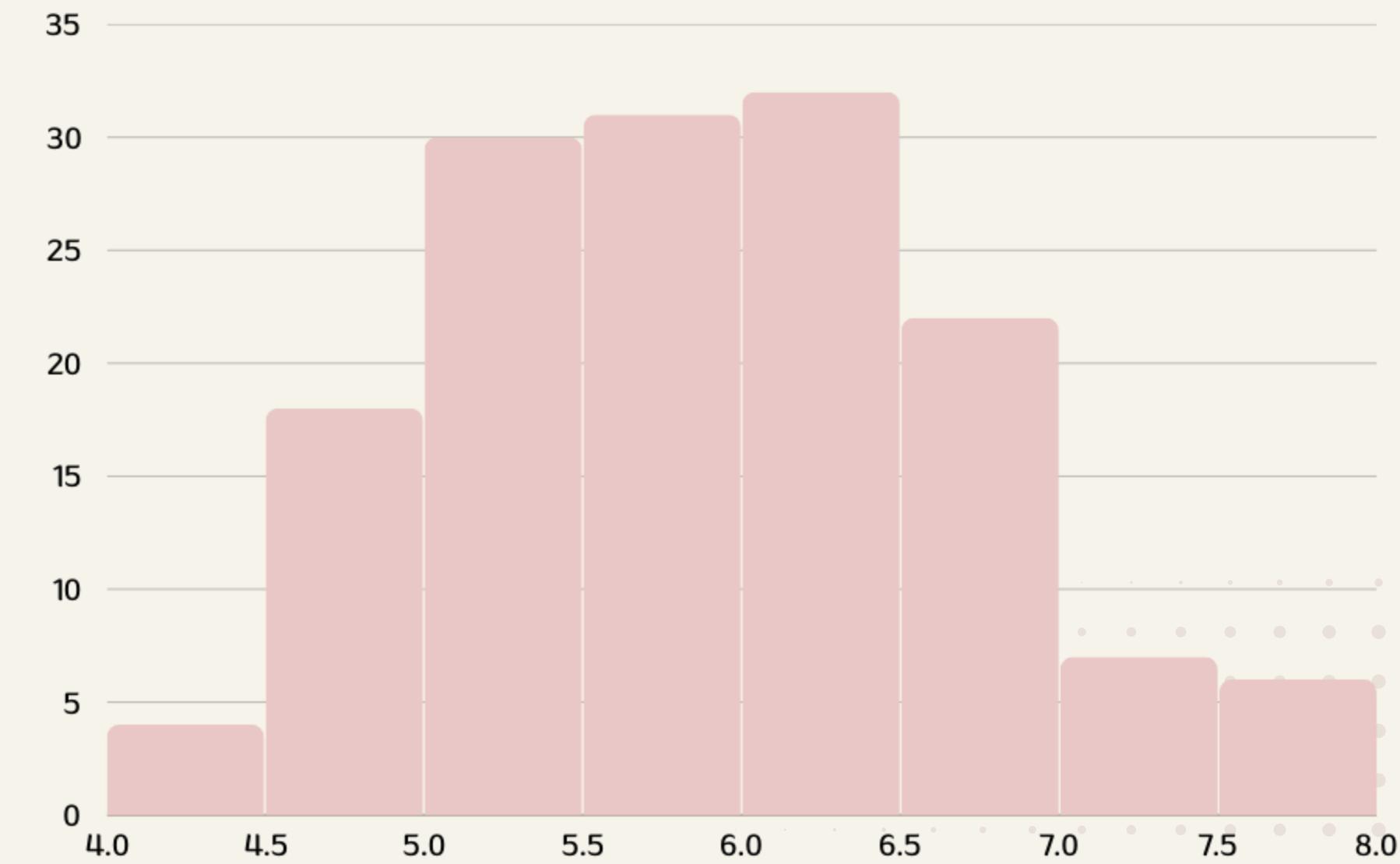
- เป็น Feature ที่แสดงเกี่ยวกับชนิดของดอกไม้ในชุดข้อมูล
- Feature นี้เป็นข้อมูลประเภทเชิงคุณภาพ (Categorical) แบบไม่มีลำดับ (Nominal)
- โดยประกอบด้วย 3 หมวดหมู่ ได้แก่ Iris-setosa, Iris-versicolor และ Iris-virginica



ความยาวของกลีบเลี้ยง (SEPAL LENGTH)

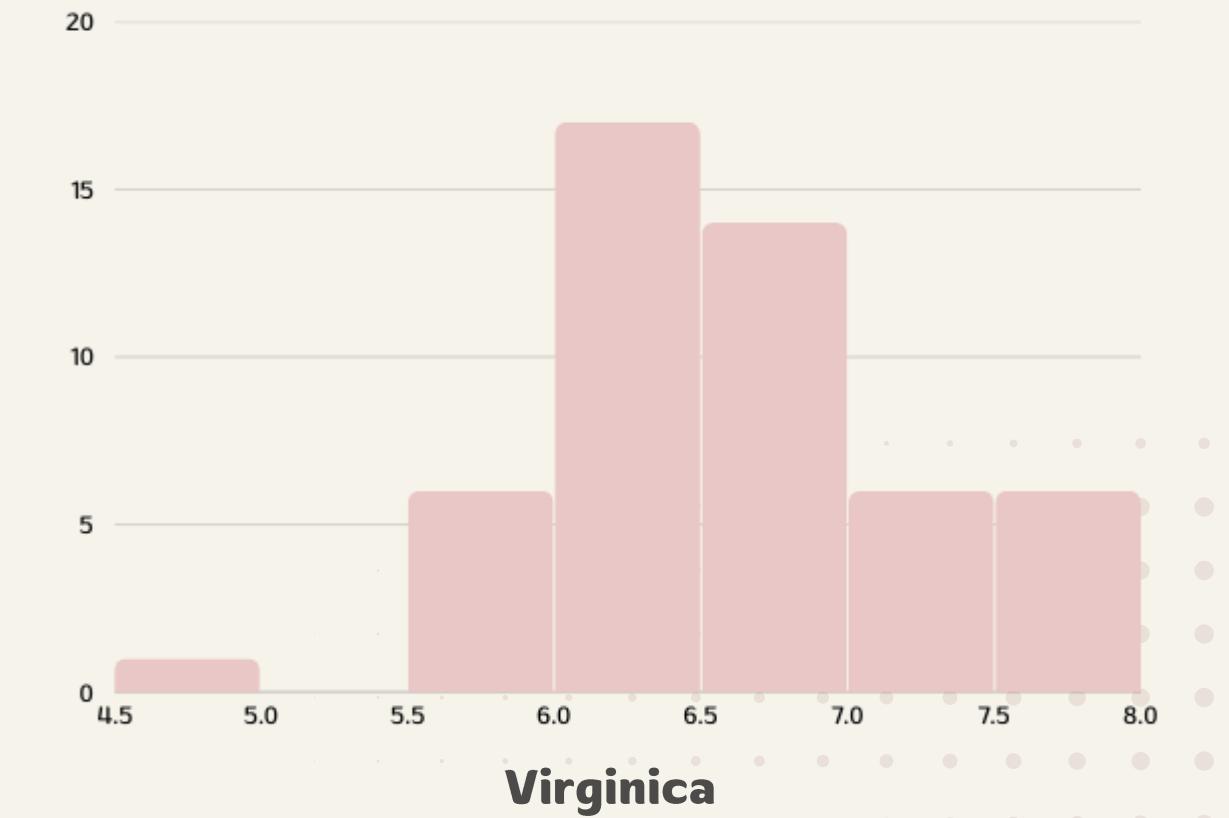
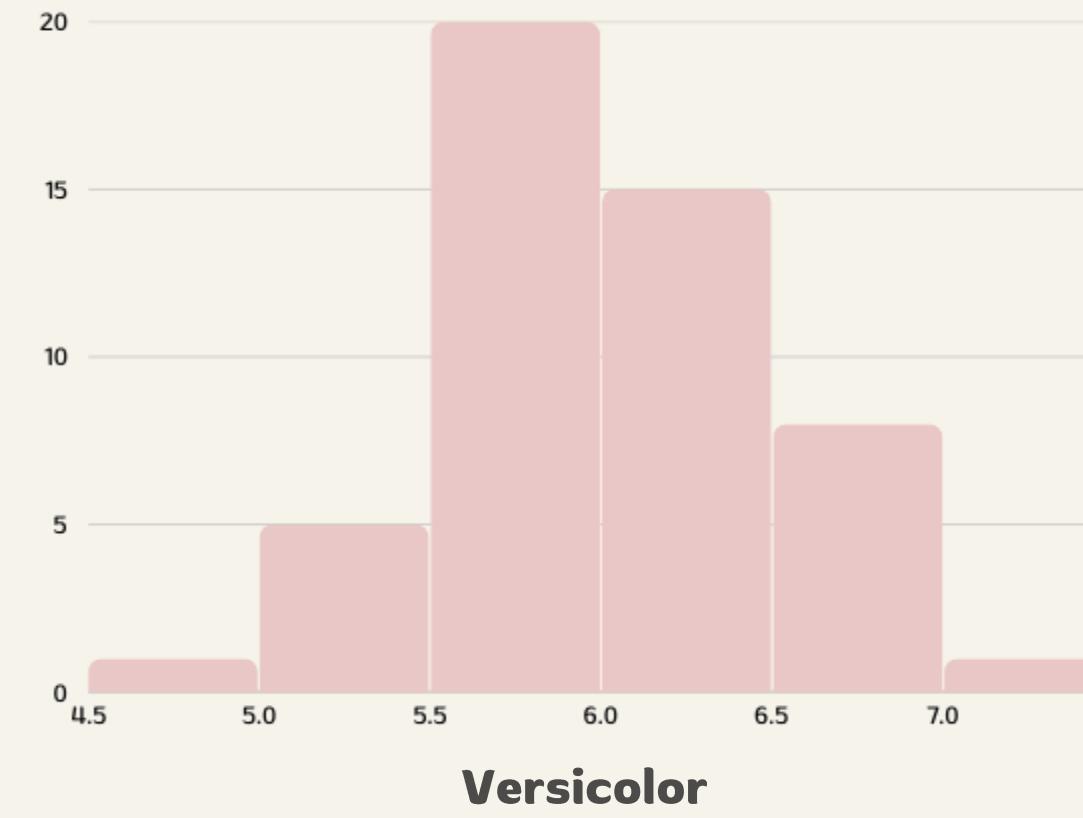
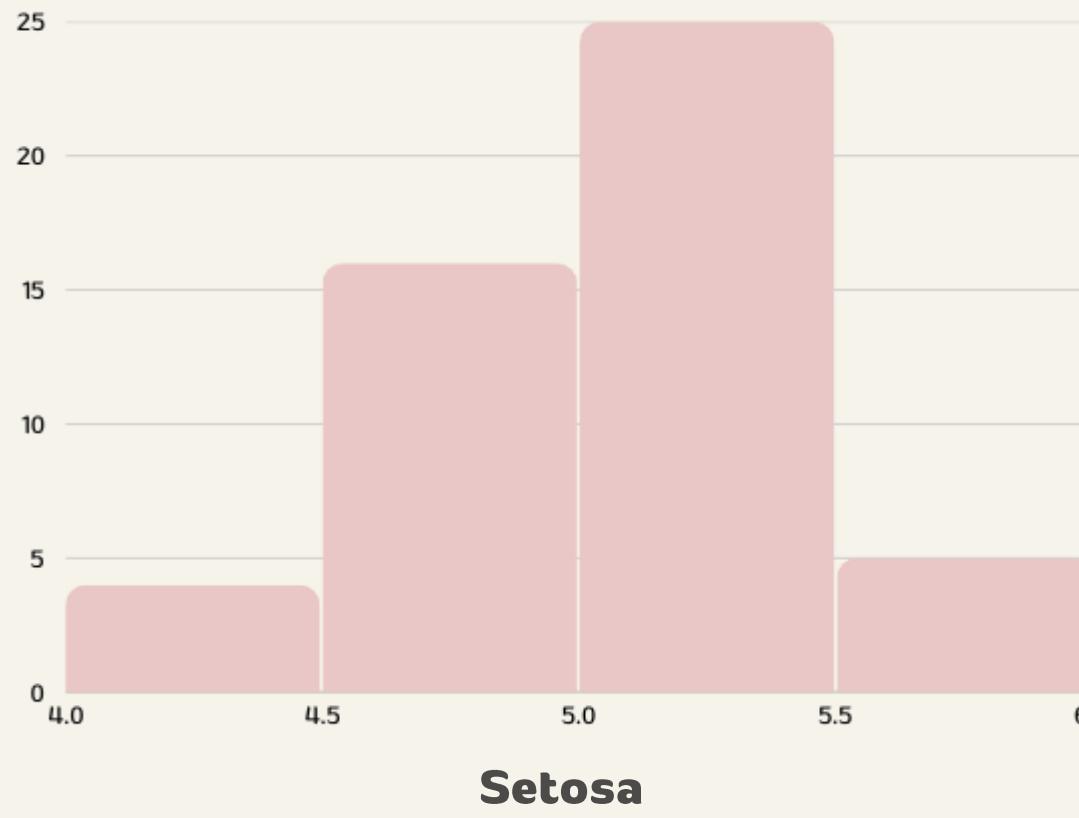
- เป็น Feature ที่แสดงเกี่ยวกับความยาวของกลีบเลี้ยงโดยมีหน่วยเป็นเซนติเมตร
- Feature นี้เป็นข้อมูลประเภทเชิงปริมาณ (Numerical) แบบอัตราส่วน (Ratio)
- โดยข้อมูลมีลักษณะเบ้าไปทางขวา (Right-skewed) เล็กน้อย

Mean	5.8433	Min	4.3
Median	5.8	Q ₁	5.1
Mode	5	Q ₂	5.8
S.D.	0.8281	Q ₃	6.4
Var.	0.6857	Max	7.9



ความยาวของกลีบเลี้ยง (SEPAL LENGTH)

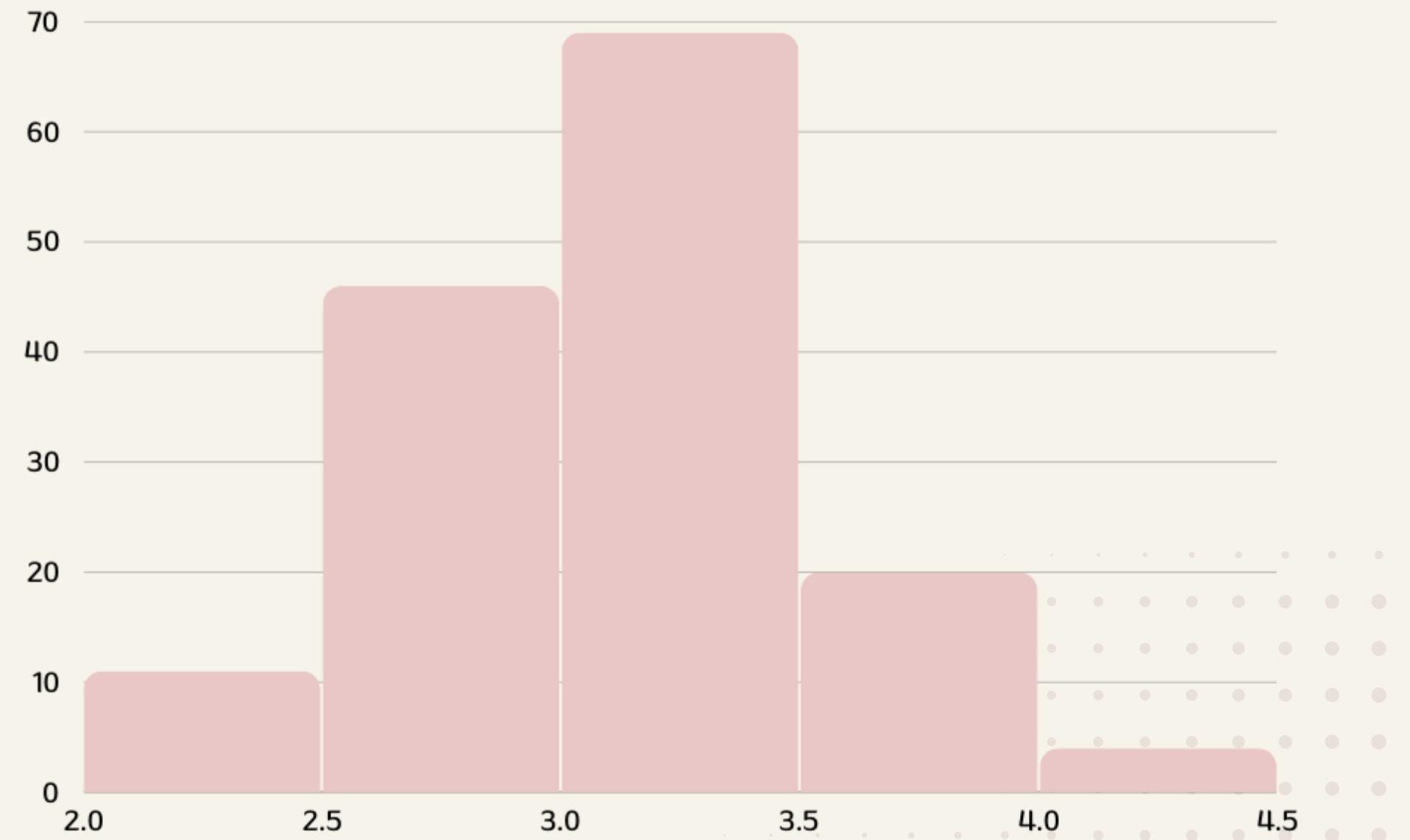
Class	Mean	Median	Mode	S.D.	Var.	Min	Q ₁	Q ₂	Q ₃	Max
Setosa	5.006	5	5	0.3525	0.1242	4.3	4.8	5	5.2	5.8
Versicolor	5.936	5.9	5.5	0.5162	0.2664	4.9	5.6	5.9	6.3	7
Virginica	6.588	6.5	6.3	0.6359	0.4043	4.9	6.225	6.5	6.9	7.9



ความกว้างของกลีบเลี้ยง (SEPAL WIDTH)

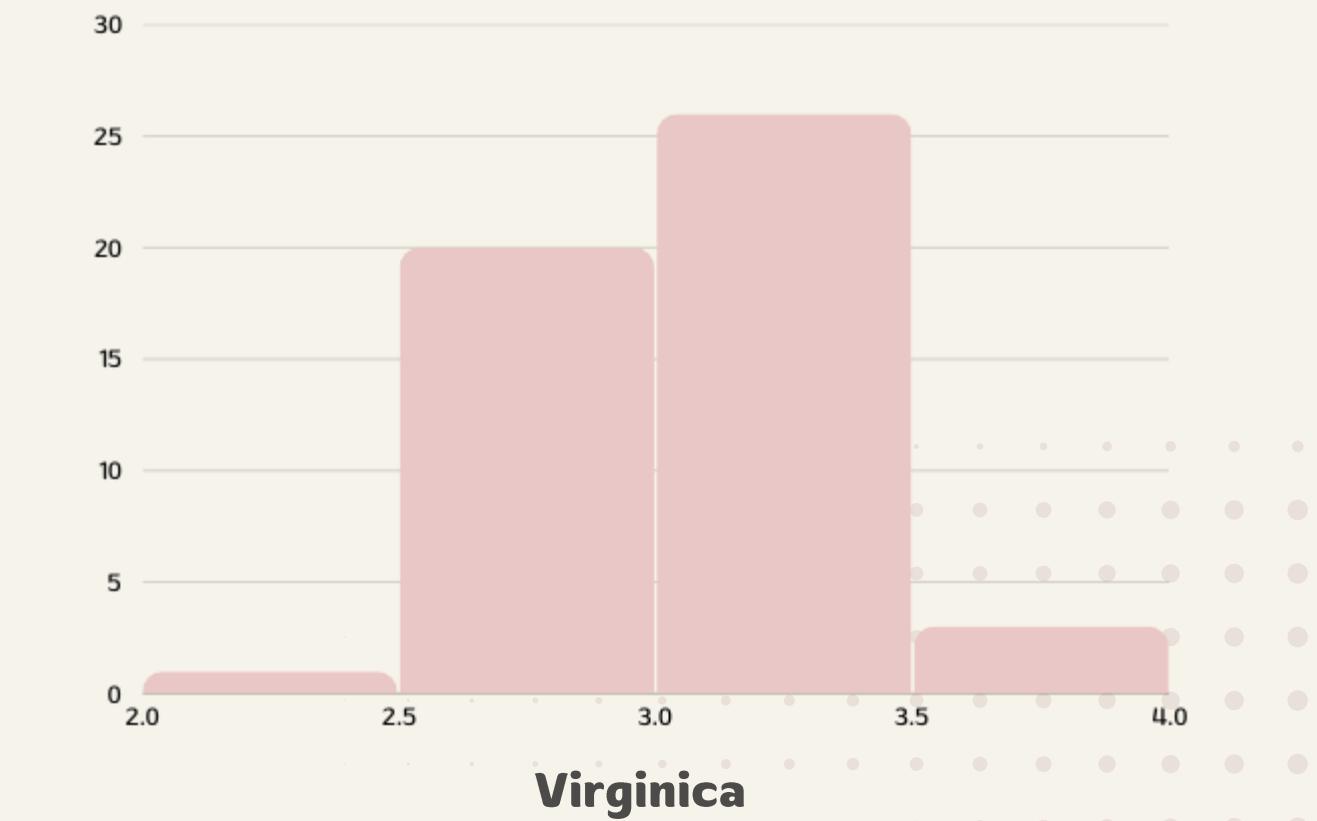
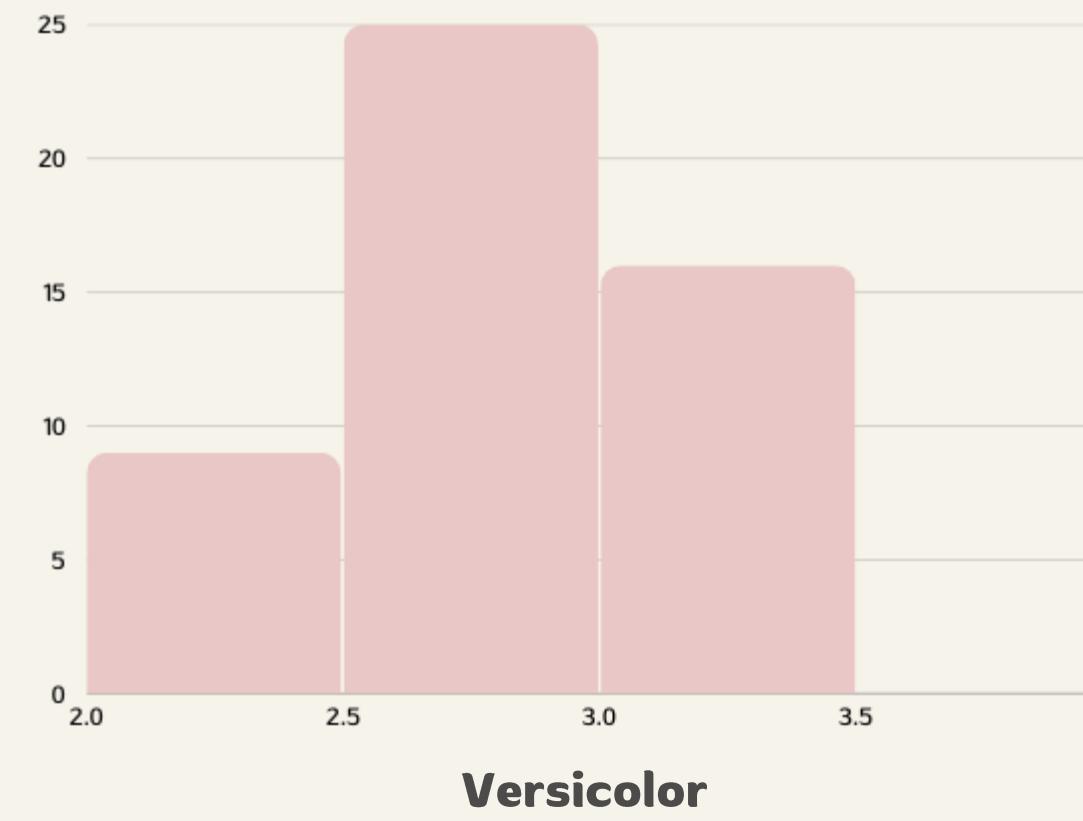
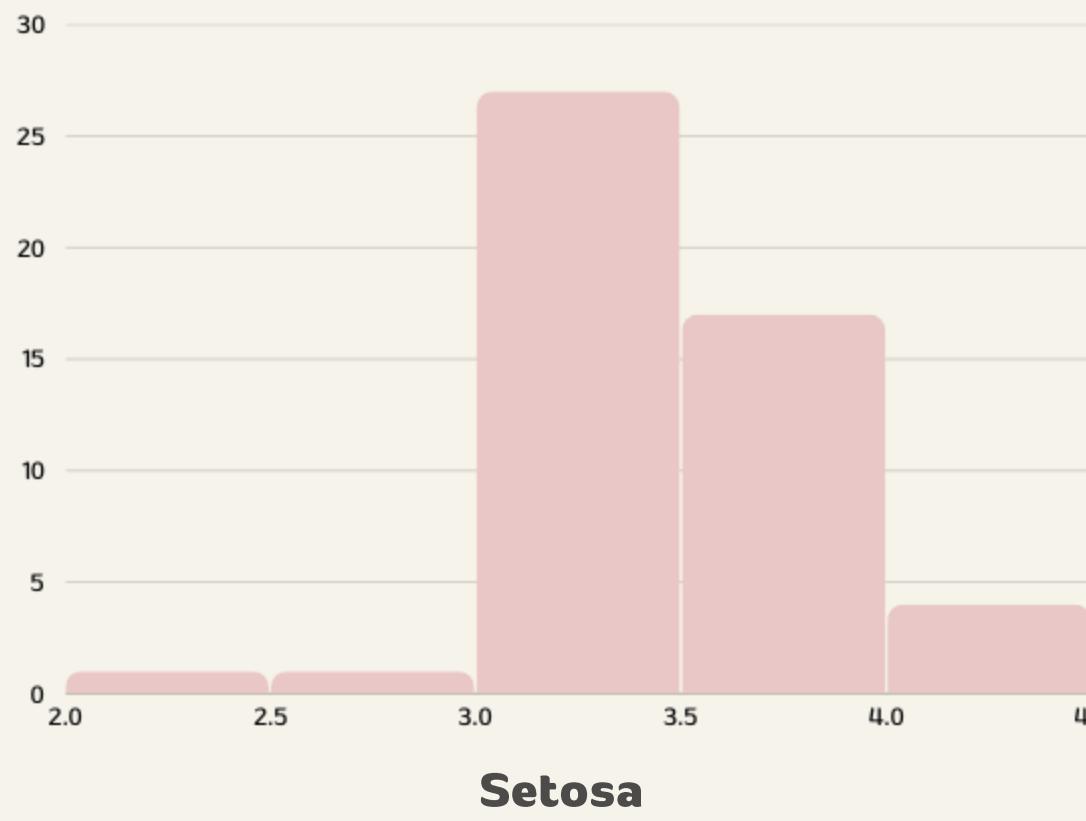
- เป็น Feature ที่แสดงเกี่ยวกับความกว้างของกลีบเลี้ยงโดยมีหน่วยเป็นเซนติเมตร
- Feature นี้เป็นข้อมูลประเภทเชิงปริมาณ (Numerical) แบบอัตราส่วน (Ratio)
- โดยข้อมูลมีลักษณะเบ้าไปทางขวา (Right-skewed) เล็กน้อย

Mean	3.054	Min	2
Median	3	Q ₁	2.8
Mode	3	Q ₂	3
SD.	0.4336	Q ₃	3.3
Var.	0.188	Max	4.4



ความกว้างของกลีบเลี้ยง (SEPAL WIDTH)

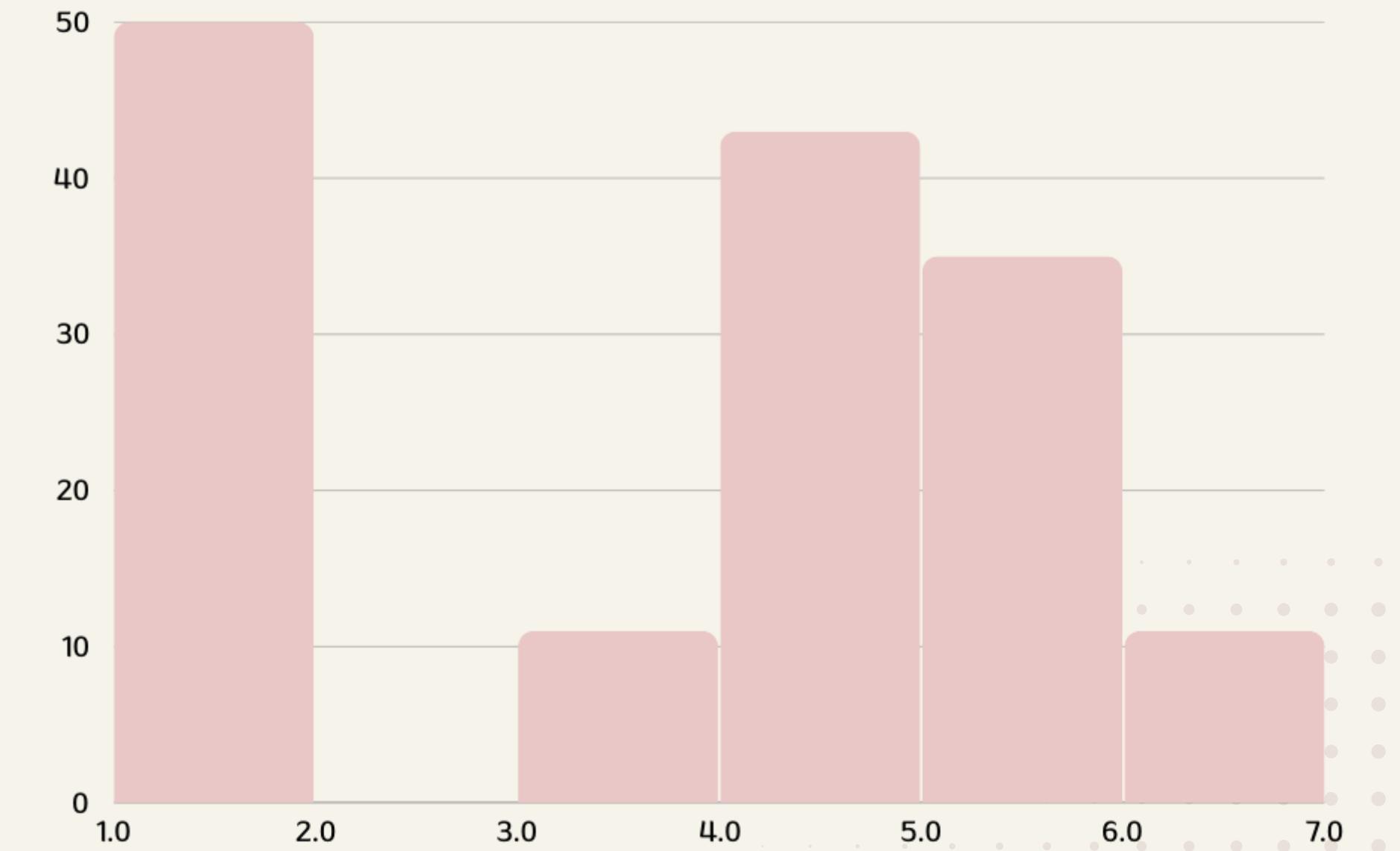
Class	Mean	Median	Mode	S.D.	Var.	Min	Q ₁	Q ₂	Q ₃	Max
Setosa	3.418	3.4	3.4	0.381	0.1452	2.3	3.125	3.4	3.675	4.4
Versicolor	2.77	2.8	3	0.3138	0.0985	2	2.525	2.8	3	3.4
Virginica	2.974	3	3	0.3225	0.104	2.2	2.8	3	3.175	3.8



ความยาวของกลีบดอก (PETAL LENGTH)

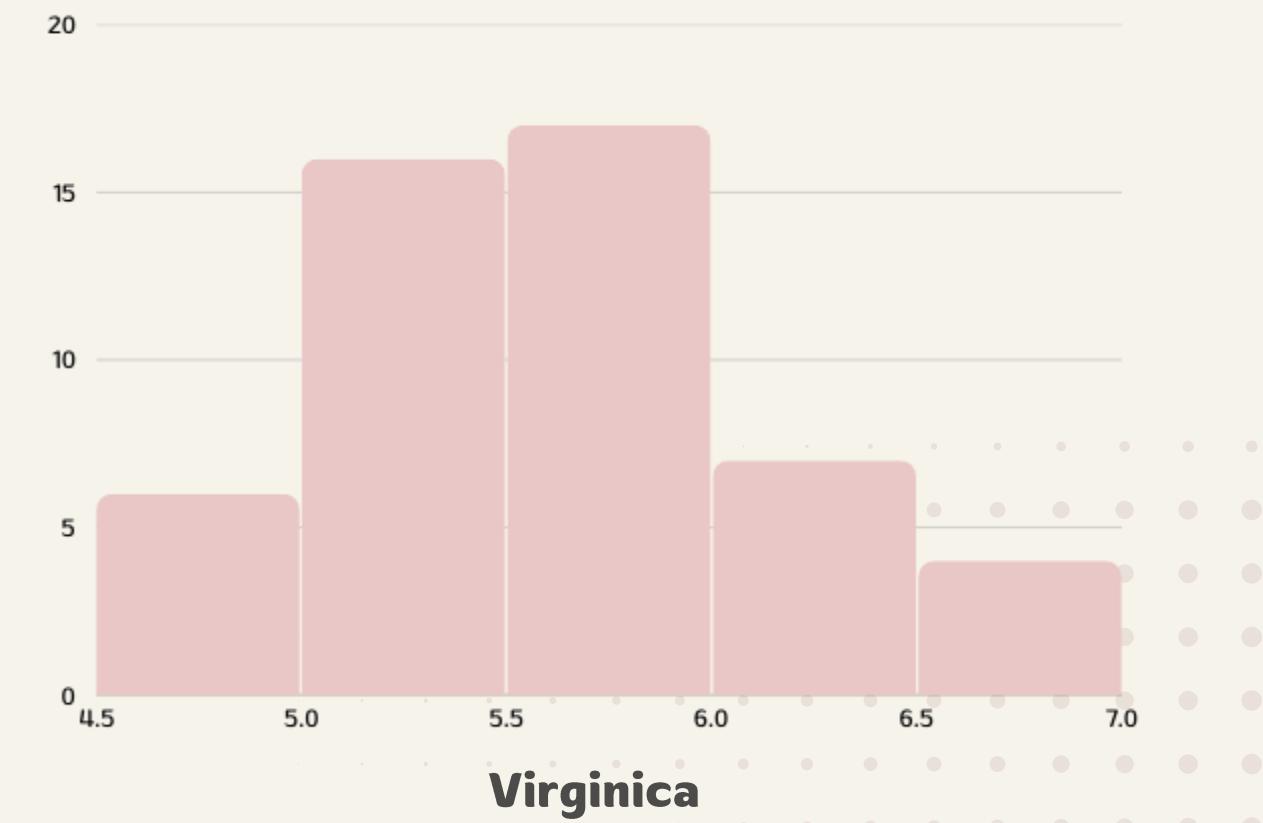
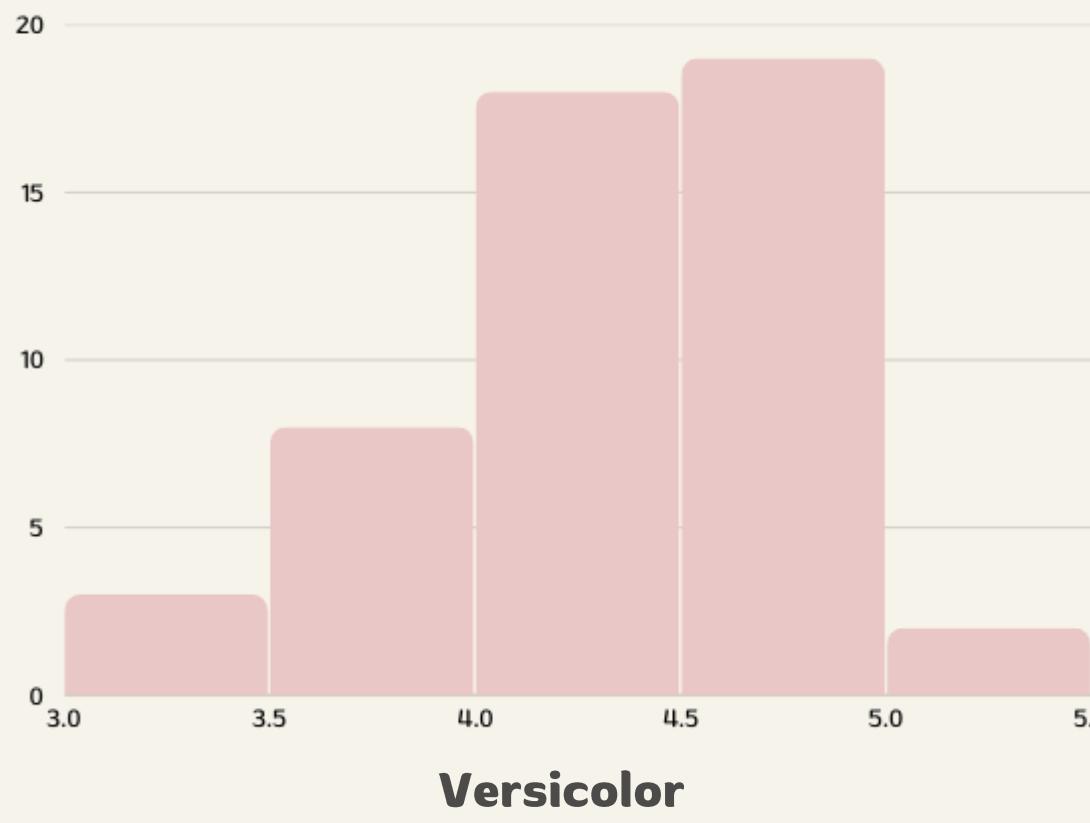
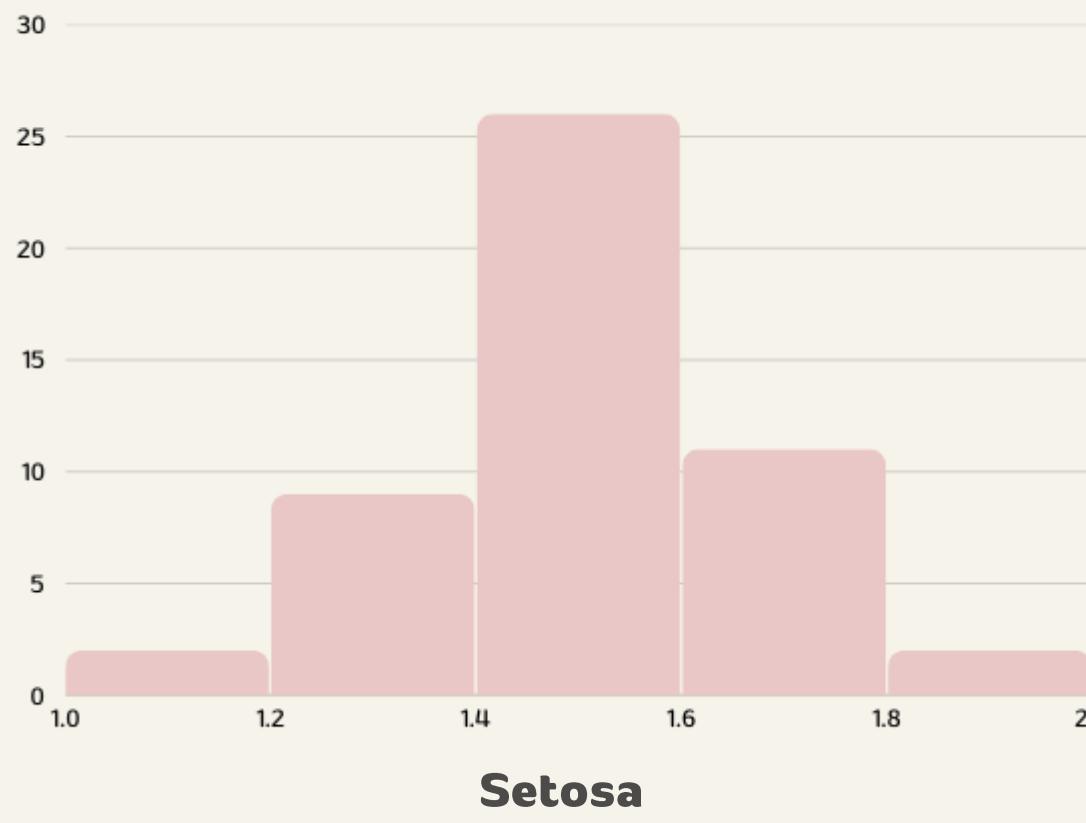
- เป็น Feature ที่แสดงเกี่ยวกับความยาวของกลีบดอกโดยมีหน่วยเป็นเซนติเมตร
- Feature นี้เป็นข้อมูลประเภทเชิงปริมาณ (Numerical) แบบอัตราส่วน (Ratio)
- โดยข้อมูลมีลักษณะเบ้าไปทางซ้าย (Left-skewed) เล็กน้อย

Mean	3.7587	Min	1
Median	4.35	Q ₁	1.6
Mode	1.5	Q ₂	4.35
SD.	1.7644	Q ₃	5.1
Var.	3.1132	Max	6.9



ความยาวของกลีบดอก (PETAL LENGTH)

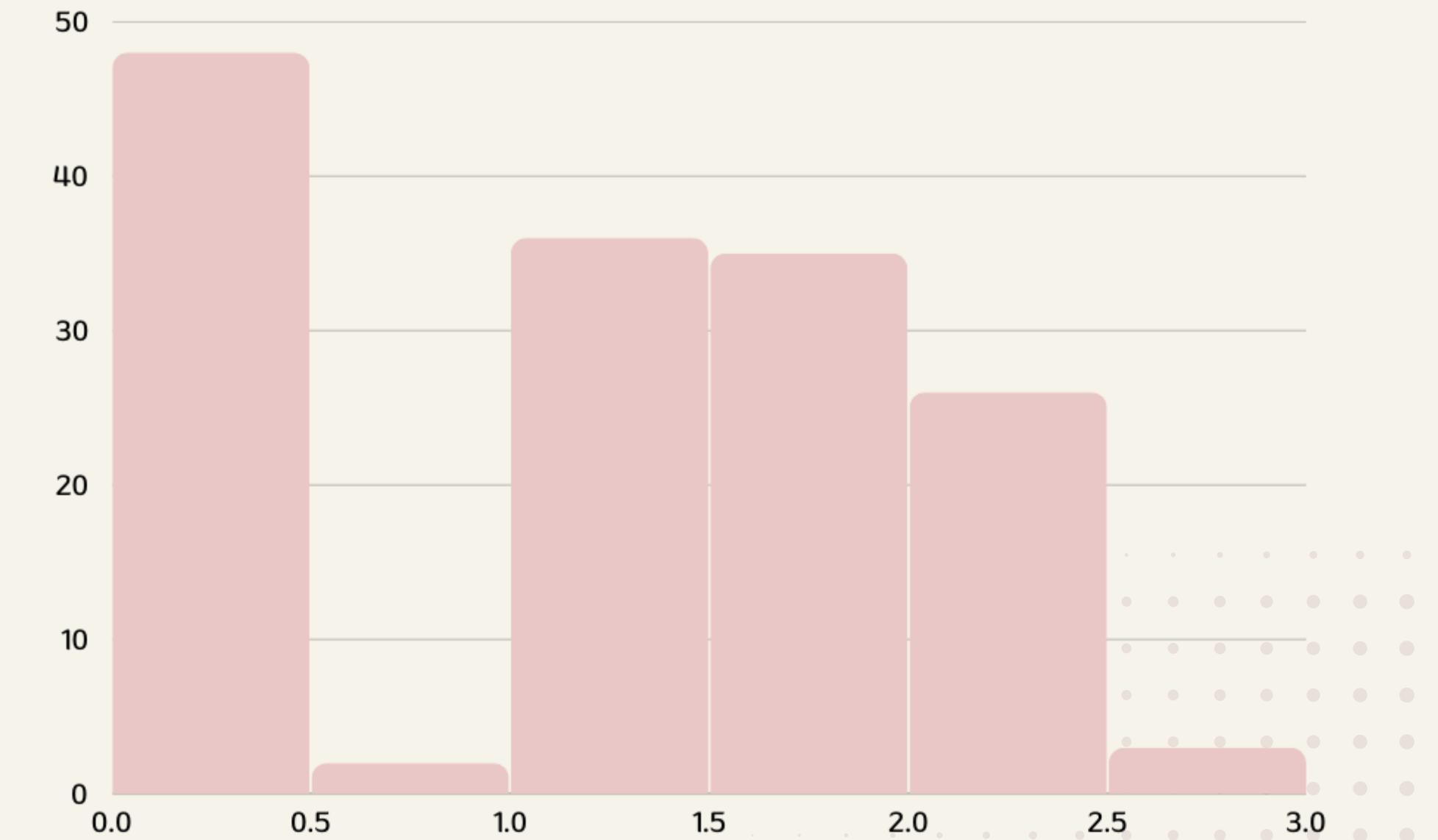
Class	Mean	Median	Mode	S.D.	Var.	Min	Q ₁	Q ₂	Q ₃	Max
Setosa	1.464	1.5	1.5	0.1735	0.0301	1	1.4	1.5	1.575	1.9
Versicolor	4.26	4.35	4.5	0.4699	0.2208	3	4	4.35	4.6	5.1
Virginica	5.552	5.55	5.1	0.5519	0.3046	4.5	5.1	5.55	5.875	6.9



ความกว้างของกลีบดอก (PETAL WIDTH)

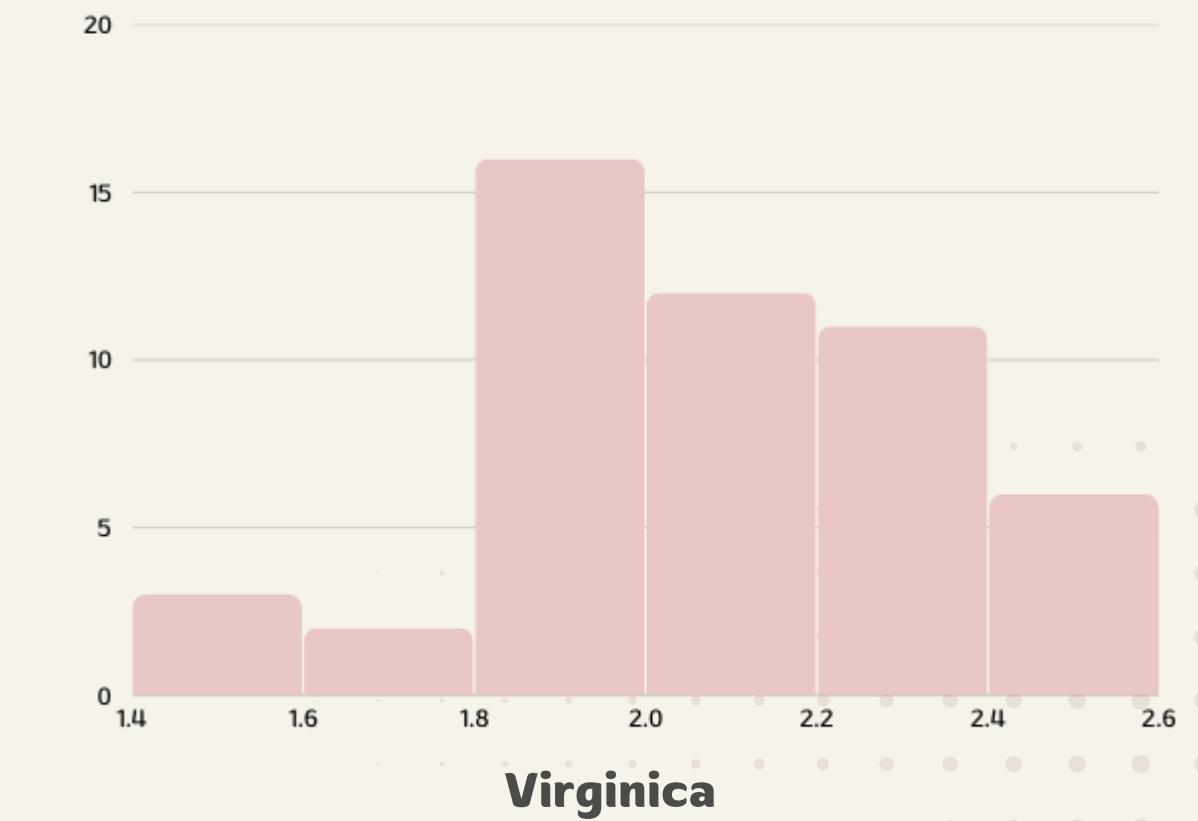
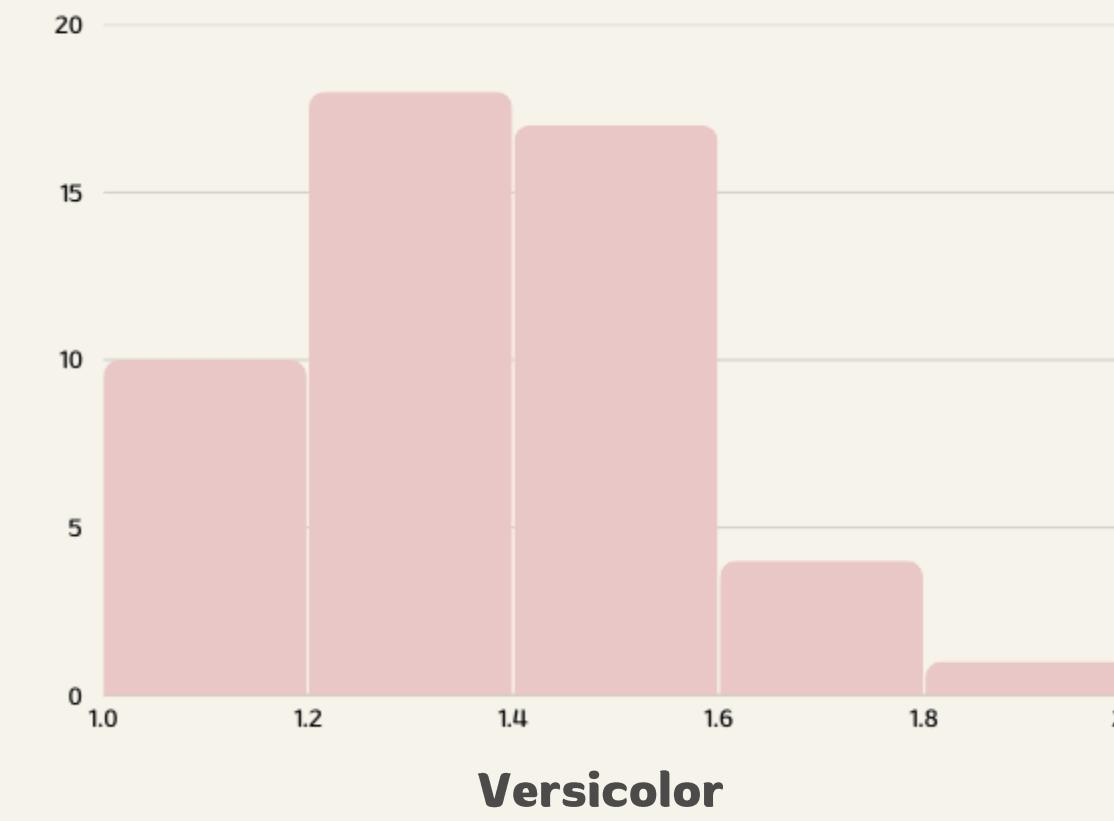
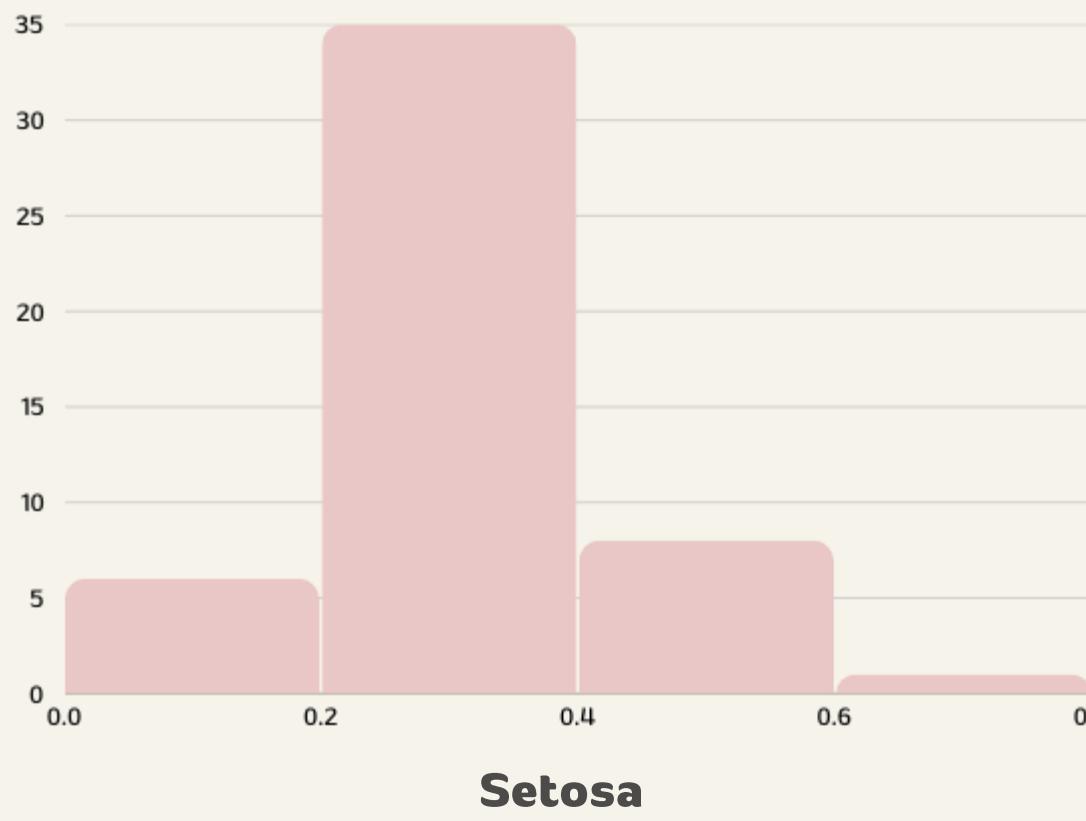
- เป็น Feature ที่แสดงเกี่ยวกับความกว้างของกลีบดอกโดยมีหน่วยเป็นเซนติเมตร
- Feature นี้เป็นข้อมูลประเภทเชิงปริมาณ (Numerical) แบบอัตราส่วน (Ratio)
- โดยข้อมูลมีลักษณะเบ้าไปทางซ้าย (Left-skewed) เล็กน้อย

Mean	1.1987	Min	0.1
Median	1.3	Q ₁	0.3
Mode	0.2	Q ₂	1.3
SD.	0.7632	Q ₃	1.8
Var.	0.5824	Max	2.5



ความกว้างของกลีบดอก (PETAL WIDTH)

Class	Mean	Median	Mode	S.D.	Var.	Min	Q ₁	Q ₂	Q ₃	Max
Setosa	0.244	0.2	0.2	0.1072	0.0115	0.1	0.2	0.2	0.3	0.6
Versicolor	1.326	1.3	1.3	0.1978	0.0391	1	1.2	1.3	1.5	1.8
Virginica	2.026	2	1.8	0.2747	0.0754	1.4	1.8	2	2.3	2.5



DATA PREPROCESS

NORMALIZATION (การปรับขนาดข้อมูล)

- ใช้ Min-Max Scaling เพื่อกำหนดค่าของแต่ละฟีเจอร์มีความใกล้เคียงและอยู่ในช่วงเดียวกัน

	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
...
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

ข้อมูลหลังทำ
Normalization



ข้อมูลก่อนทำ
Normalization

	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
...
145	0.666667	0.416667	0.711864	0.916667	Iris-virginica
146	0.555556	0.208333	0.677966	0.750000	Iris-virginica
147	0.611111	0.416667	0.711864	0.791667	Iris-virginica
148	0.527778	0.583333	0.745763	0.916667	Iris-virginica
149	0.444444	0.416667	0.694915	0.708333	Iris-virginica

CLASS ENCODING

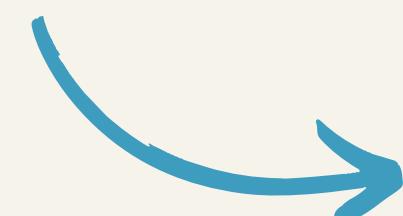
- ใช้ One-Hot Encoding ในการแปลง Feature “Class” จากที่เป็นประเภท Numerical

	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
..
145	0.666667	0.416667	0.711864	0.916667	Iris-virginica
146	0.555556	0.208333	0.677966	0.750000	Iris-virginica
147	0.611111	0.416667	0.711864	0.791667	Iris-virginica
148	0.527778	0.583333	0.745763	0.916667	Iris-virginica
149	0.444444	0.416667	0.694915	0.708333	Iris-virginica

ข้อมูลก่อนทำ
Class Encoding



ข้อมูลหลังทำ
Class Encoding



	sepal length	sepal width	petal length	petal width	class_Iris-setosa	class_Iris-versicolor	class_Iris-virginica
0	0.222222	0.625000	0.067797	0.041667	True	False	False
1	0.166667	0.416667	0.067797	0.041667	True	False	False
2	0.111111	0.500000	0.050847	0.041667	True	False	False
3	0.083333	0.458333	0.084746	0.041667	True	False	False
4	0.194444	0.666667	0.067797	0.041667	True	False	False
..
145	0.666667	0.416667	0.711864	0.916667	False	False	True
146	0.555556	0.208333	0.677966	0.750000	False	False	True
147	0.611111	0.416667	0.711864	0.791667	False	False	True
148	0.527778	0.583333	0.745763	0.916667	False	False	True
149	0.444444	0.416667	0.694915	0.708333	False	False	True

ขอบคุณครับ

6505000270 นายคุกันฐ แซ่เตีย

ว้างอิง

- 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/>