

Wheat Seeds Dataset

วิเคราะห์ชุดข้อมูลเมล็ดข้าวสาลี

6605006698 จุฬานนท์ สมเพชร

Introduction to Dataset

- Source: UCI Machine Learning Repository (Seeds Data Set)
- Created by Institute of Agrophysics, Polish Academy of Sciences
- Total samples: 210 (Kama, Rosa, Canadian)
- No missing values

Dataset Features

1. Area

3. Compactness

5. Width of kernel

7. Length of kernel
groove

2. Perimeter

4. Length of kernel

6. Asymmetry coefficient

Class: 1=Kama, 2=Rosa,
3=Canadian

Statistical Measures (Concepts)

1. Range

2. Min

3. Max

4. Mean

5. Median

6. Mode

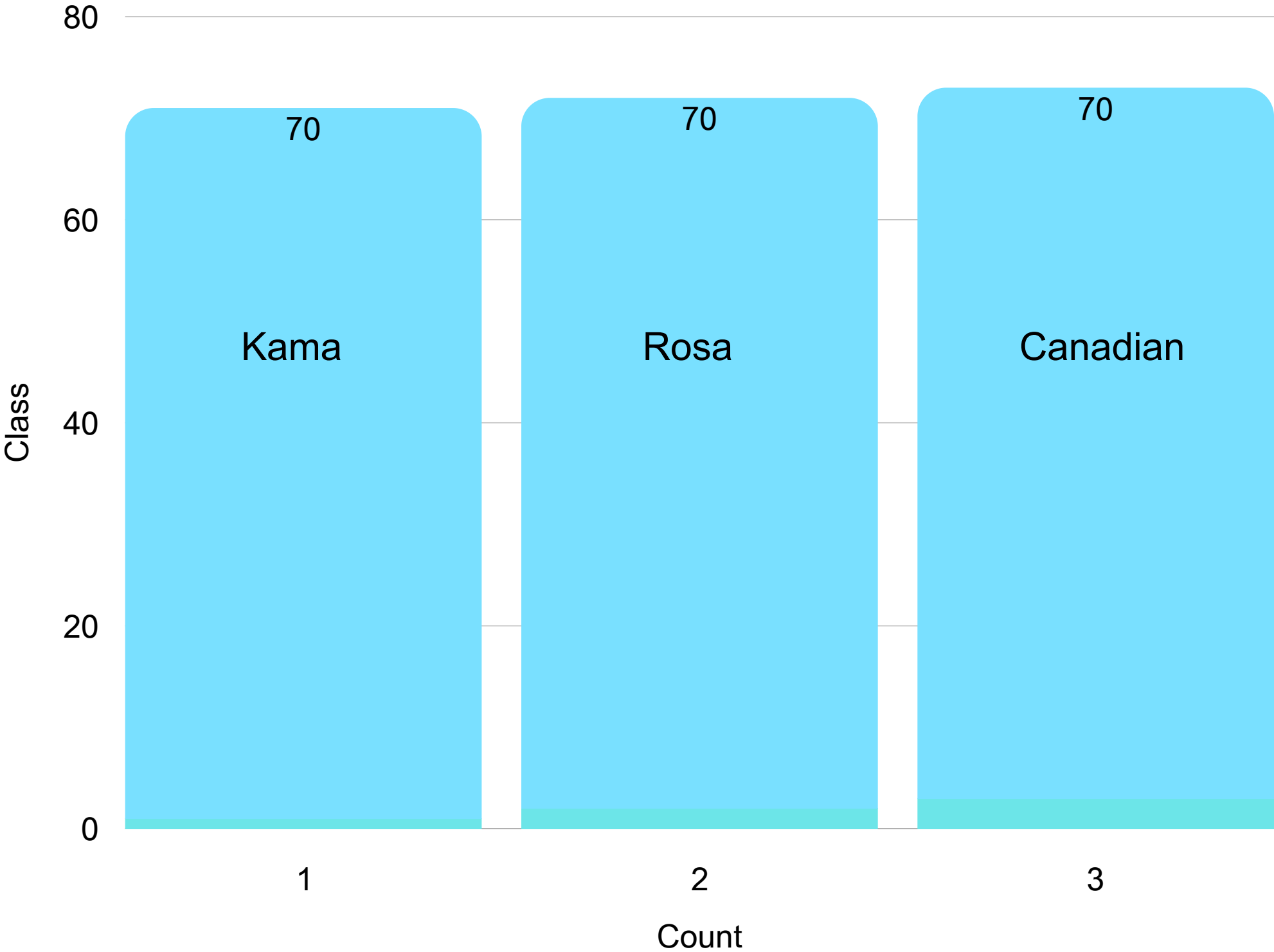
7. Standard deviation

Statistical Summary (All Features)

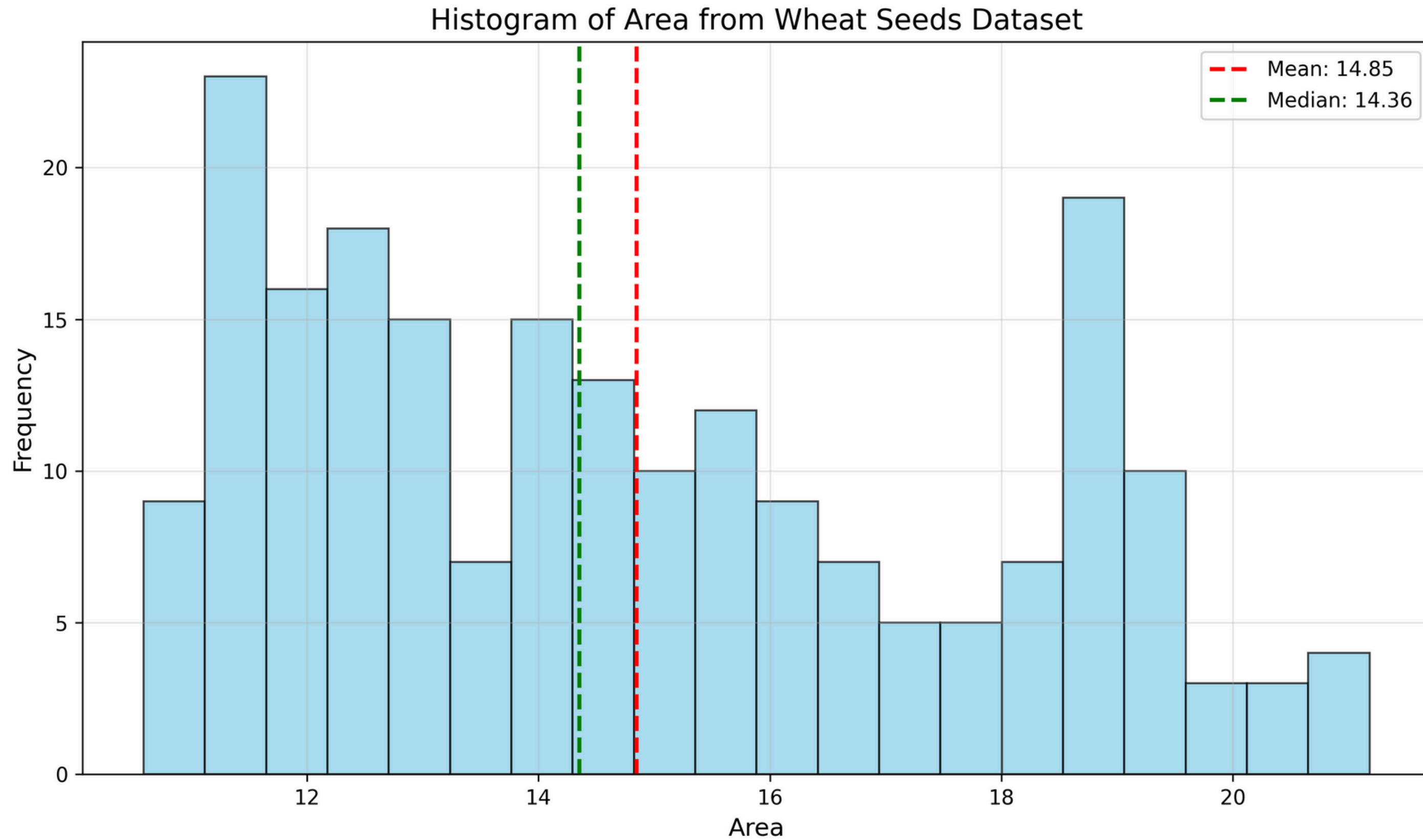
Feature	Range	Min	Max	Mean	Median	Mode	Std Dev
Area	10.59	10.59	21.18	14.85	14.36	11.23	2.91
Perimeter	4.8	12.41	17.25	14.56	14.32	13.47	1.31
Compactness	0.11	0.81	0.92	0.87	0.87	0.88	0.02
Length of kernel	1.78	4.9	6.68	5.63	5.52	5.23	0.44
Width of kernel	1.4	2.63	4.03	3.26	3.23	3.03	0.4
Asymmetry coefficient	7.7	0.77	8.46	3.7	3.6	2.13	1.5
Length of kernel groove	2.03	4.52	6.55	5.4	5.22	5	0.49

Statistics

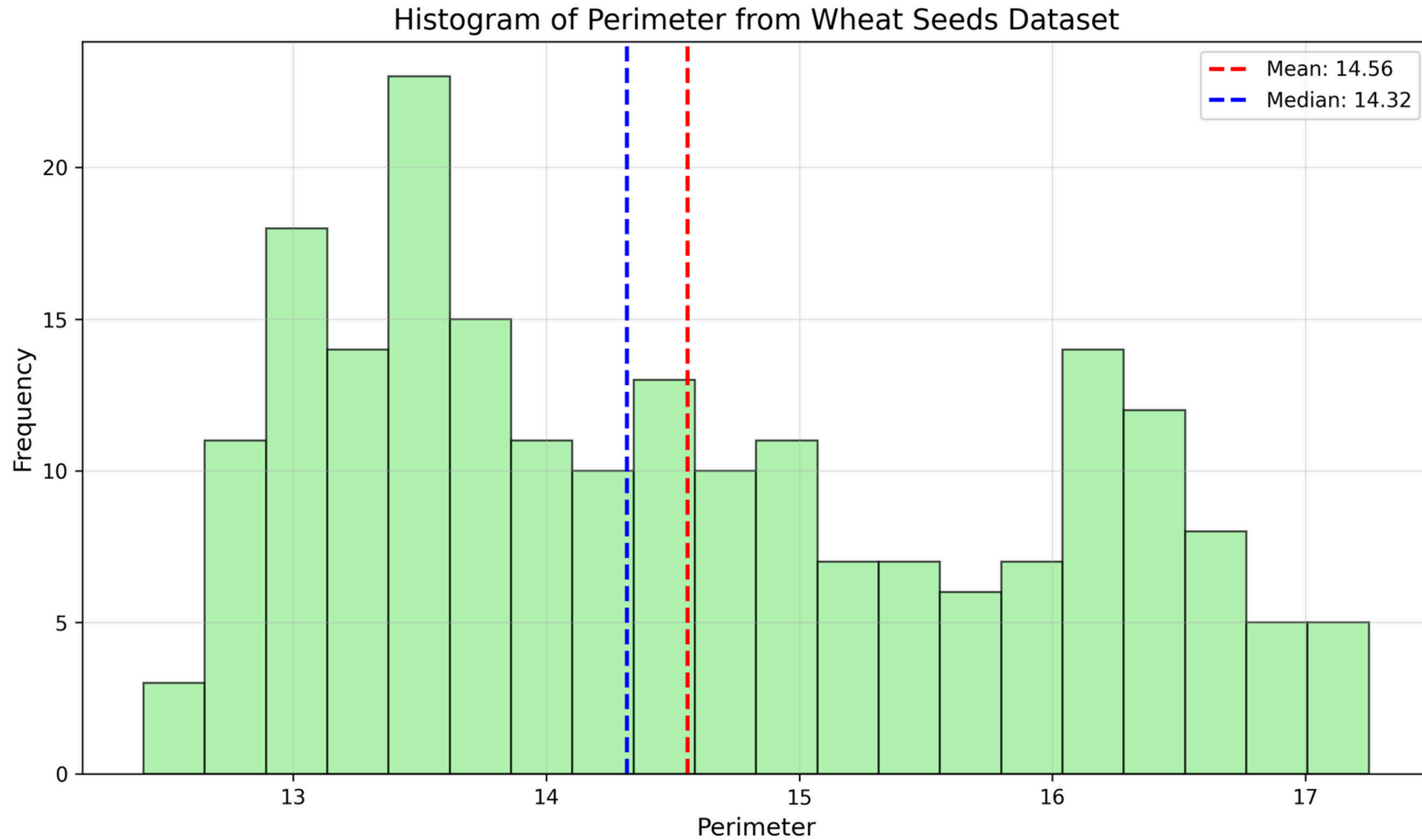
คลาสและการกระจาย



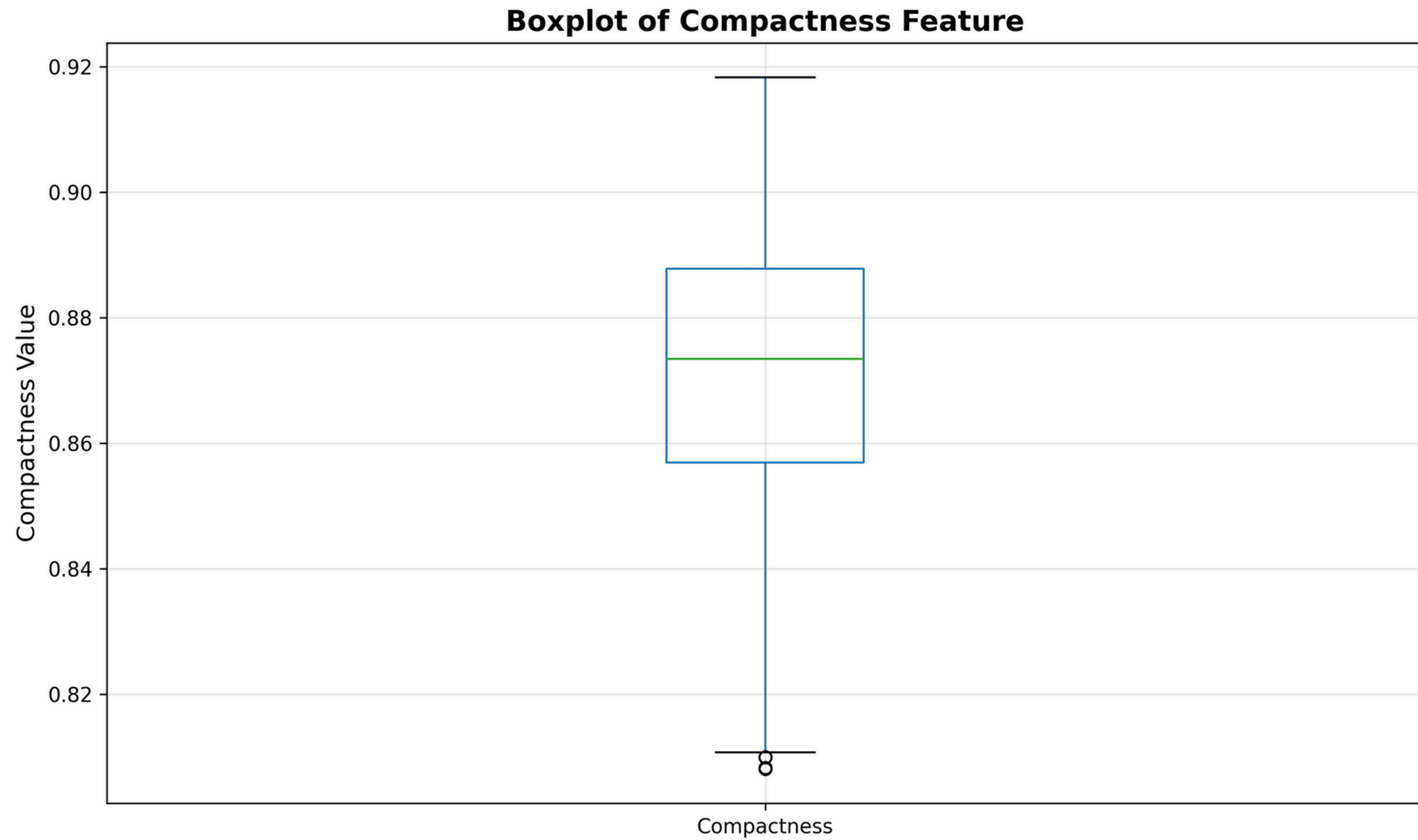
Histogram - Area



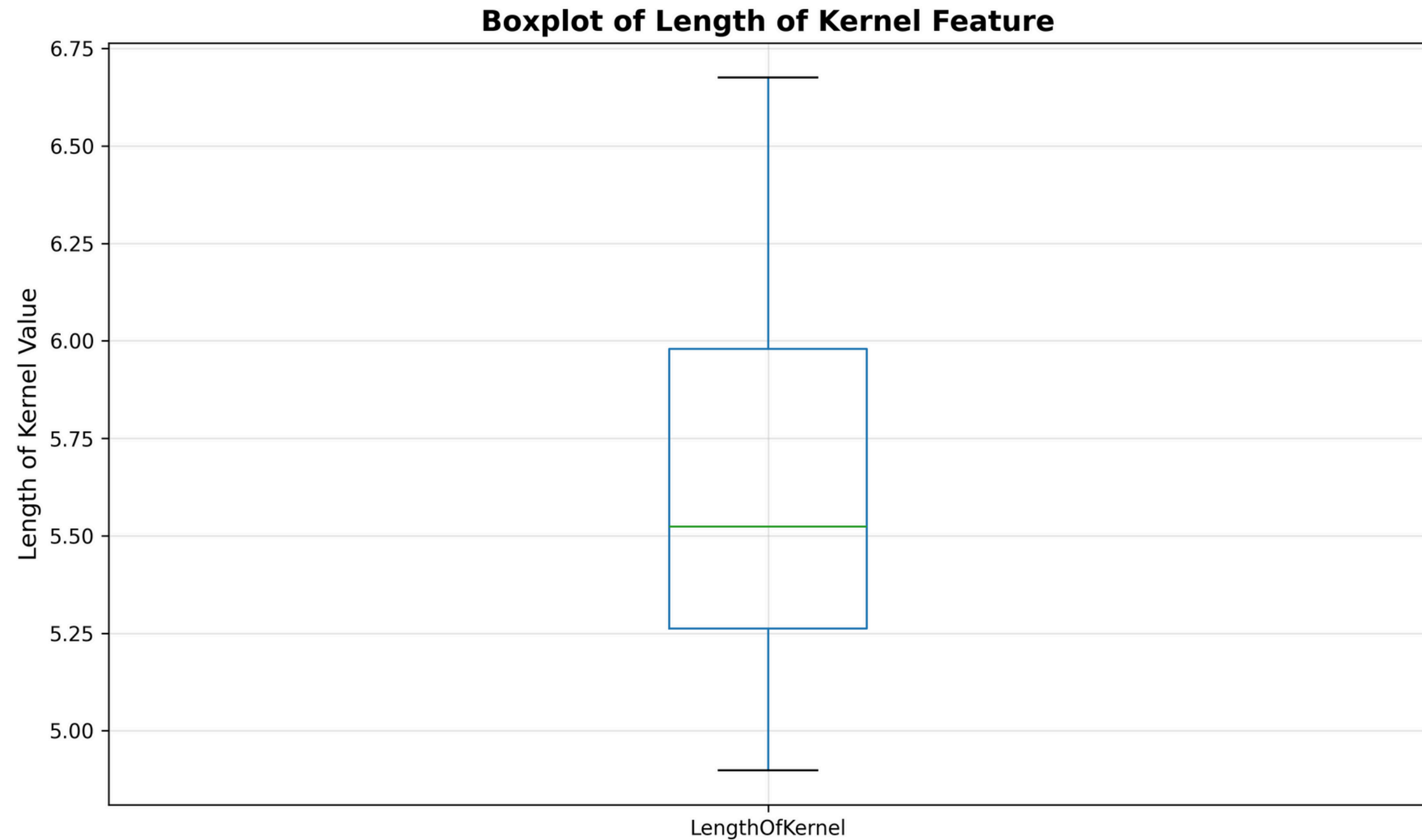
Histogram – Perimeter



Boxplot – Compactness



Boxplot – Length of Kernel



Conclusion

- ชุดข้อมูลเหมาะกับงานจำแนกสายพันธุ์ด้วยลักษณะทางกายภาพ
- ค่าสถิติเบื้องต้นช่วยตีความการกระจายและตรวจจับค่าผิดปกติ

อ้างอิง

<https://www.kaggle.com/code/maniadapala/wheat-seed-quality-detection>

<https://github.com/jbrownlee/Datasets/blob/master/wheat-seeds.csv>

Thank you