Set up the notebook

Her zaman olduğu gibi, kodlama ortamını ayarlayarak başlıyoruz. (Bu kod gizlidir, ancak bu metnin hemen altında, sağdaki "Kod" düğmesine tıklayarak gizliliğini kaldırabilirsiniz.)

import pandas as pd

pd.plotting.register\_matplotlib\_converters()

import matplotlib.pyplot as plt

%matplotlib inline

import seaborn as sns

print("Setup Complete")

Setup Complete

# Select a dataset

Üç farklı süsen türünden (Iris setosa, Iris versicolor ve Iris virginica) 50'şer adet olmak üzere toplam 150 farklı çiçekten oluşan bir veri setiyle çalışacağız.



# Load and examine the data[¶](https://www.kaggle.com/code/alexisbcook/distributions" \l "Load-and-examine-the-data" \t "_self)

Veri setindeki her satır farklı bir çiçeğe karşılık geliyor. Dört ölçüm bulunuyor: çanak yaprağı uzunluğu ve genişliği ile taç yaprağı uzunluğu ve genişliği. Ayrıca, ilgili türleri de takip ediyoruz.

*# Path of the file to read*

iris\_filepath = "../input/iris.csv"

*# Read the file into a variable iris\_data*

iris\_data = pd.read\_csv(iris\_filepath, index\_col="Id")

*# Print the first 5 rows of the data*

iris\_data.head()

| Sepal Length (cm) | Sepal Width (cm) | Petal Length (cm) | Petal Width (cm) | Species |
| --- | --- | --- | --- | --- |
| Id |  |  |  |  |  |
| 1 | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
| 2 | 4.9 | 3.0 | 1.4 | 0.2 | Iris-setosa |
| 3 | 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| 4 | 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
| 5 | 5.0 | 3.6 | 1.4 | 0.2 | Iris-setosa |

# Histograms

İris çiçeklerindeki taç yaprağı uzunluğunun nasıl değiştiğini görmek için bir **histogram** oluşturmak istediğimizi varsayalım. Bunu **sns.histplot** komutuyla yapabiliriz.

*# Histogram*

sns.histplot(iris\_data['Petal Length (cm)'])

<AxesSubplot:xlabel='Petal Length (cm)', ylabel='Count'>

Yukarıdaki kod hücresine, çizim yapmak istediğimiz sütunu içeren komutu girmemiz gerekiyordu (bu durumda, 'Yaprak Uzunluğu (cm)' seçeneğini seçtik).

# Density plots

Bir sonraki grafik türü, **çekirdek yoğunluk tahmini (KDE)** grafiğidir. KDE grafiklerine aşina değilseniz, onları pürüzsüzleştirilmiş bir histogram olarak düşünebilirsiniz.

Bir KDE grafiği oluşturmak için **sns.kdeplot** komutunu kullanırız. **shade=True** ayarı, eğrinin altındaki alanı renklendirir ve **data=** ile hangi sütunu çizeceğimizi seçeriz.

*# KDE plot*

sns.kdeplot(data=iris\_data['Petal Length (cm)'], shade=True)

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:2: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.

This will become an error in seaborn v0.14.0; please update your code.

<AxesSubplot:xlabel='Petal Length (cm)', ylabel='Density'>

# 2D KDE plots

KDE grafiği oluştururken tek bir sütunla sınırlı değiliz. **sns.jointplot** komutuyla **iki boyutlu (2D) bir KDE grafiği** oluşturabiliriz.

Aşağıdaki grafikte, renk kodlaması bize çanak yaprağı genişliği ve taç yaprağı uzunluğunun farklı kombinasyonlarını görme olasılığımızı gösterir; figürün daha koyu kısımları daha olasıdır.

*# 2D KDE plot*

sns.jointplot(x=iris\_data['Petal Length (cm)'], y=iris\_data['Sepal Width (cm)'], kind="kde")

<seaborn.axisgrid.JointGrid at 0x786187a48f50>

Merkezdeki 2D KDE grafiğine ek olarak, figürün üst kısmındaki eğrinin x eksenindeki veriler (bu durumda, **iris\_data['Petal Length (cm)']**) için, figürün sağındaki eğrinin ise y eksenindeki veriler (bu durumda, **iris\_data['Sepal Width (cm)']**) için bir KDE grafiği olduğunu unutmayın.

# Color-coded plots

Eğitimin bir sonraki bölümü için, türler arasındaki farklılıkları anlamak üzere grafikler oluşturacağız.

Yukarıdaki gibi **sns.histplot** komutunu kullanarak taç yaprağı uzunluğuna göre (her tür için bir tane olmak üzere) üç farklı histogram oluşturabiliriz.

* **data=**: Veriyi okumak için kullandığımız değişkenin adını sağlar.
* **x=**: Çizmek istediğimiz verileri içeren sütunun adını ayarlar.
* **hue=**: Veriyi farklı histogramlara ayırmak için kullanacağımız sütunu ayarlar.

# Histograms for each species

sns.histplot(data=iris\_data, x='Petal Length (cm)', hue='Species')

# Add title

plt.title("Histogram of Petal Lengths, by Species")

Text(0.5, 1.0, 'Histogram of Petal Lengths, by Species')

Türler için de, yukarıdaki gibi **sns.kdeplot** kullanarak bir KDE grafiği oluşturabiliriz. **data**, **x** ve **hue** işlevleri, yukarıda **sns.histplot** kullandığımız zamankiyle aynıdır. Ayrıca, her eğrinin altındaki alanı renklendirmek için **shade=True** ayarını yapıyoruz.

*# KDE plots for each species*

sns.kdeplot(data=iris\_data, x='Petal Length (cm)', hue='Species', shade=True)

*# Add title*

plt.title("Distribution of Petal Lengths, by Species")

/opt/conda/lib/python3.7/site-packages/ipykernel\_launcher.py:2: FutureWarning:

`shade` is now deprecated in favor of `fill`; setting `fill=True`.

This will become an error in seaborn v0.14.0; please update your code.

Text(0.5, 1.0, 'Distribution of Petal Lengths, by Species')

Grafiklerde görülebilecek ilginç bir örüntü, bitkilerin iki gruba ayrılmış gibi görünmesidir: **Iris versicolor** ve **Iris virginica** taç yaprağı uzunluğu için benzer değerlere sahipken, **Iris setosa** kendi başına bir kategoriye aittir.

Aslında, bu veri setine göre, herhangi bir iris bitkisini sadece taç yaprağı uzunluğuna bakarak **Iris setosa** (Iris versicolor veya Iris virginica yerine) olarak sınıflandırabiliriz: eğer bir iris çiçeğinin taç yaprağı uzunluğu 2 cm'den azsa, büyük olasılıkla Iris setosa'dır!