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

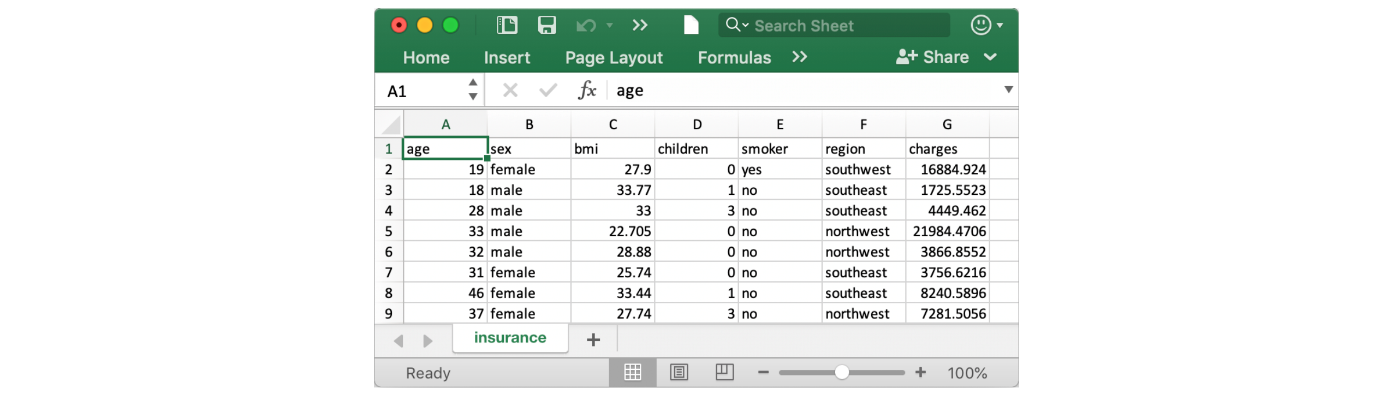
import seaborn as sns

print("Setup Complete")

Setup Complete

# Load and examine the data

Bazı müşterilerin neden diğerlerinden daha fazla ödeme yaptığını anlayabilmek için sigorta ücretlerini içeren (sentetik) bir veri kümesi üzerinde çalışacağız.



*# Path of the file to read*

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

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

insurance\_data = pd.read\_csv(insurance\_filepath)

Her zamanki gibi, veri setinin düzgün yüklendiğini ilk beş satırı yazdırarak kontrol ediyoruz.

insurance\_data.head()

| age | sex | bmi | children | smoker | region | charges |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | 19 | female | 27.900 | 0 | yes | southwest | 16884.92400 |
| 1 | 18 | male | 33.770 | 1 | no | southeast | 1725.55230 |
| 2 | 28 | male | 33.000 | 3 | no | southeast | 4449.46200 |
| 3 | 33 | male | 22.705 | 0 | no | northwest | 21984.47061 |
| 4 | 32 | male | 28.880 | 0 | no | northwest | 3866.85520 |

# Scatter plots

Basit bir **nokta grafiği** (**scatter plot**) oluşturmak için **sns.scatterplot** komutunu kullanırız ve şu değerleri belirtiriz:

* Yatay x ekseni için (**x=insurance\_data['bmi']**),
* Dikey y ekseni için (**y=insurance\_data['charges']**).

sns.scatterplot(x=insurance\_data['bmi'], y=insurance\_data['charges'])

<AxesSubplot:xlabel='bmi', ylabel='charges'>

Yukarıdaki dağılım grafiği, **vücut kitle indeksi** (BMI) ve sigorta masraflarının **pozitif ilişkili** olduğunu gösteriyor. Bu, BMI'sı yüksek olan müşterilerin genellikle sigorta masraflarının da daha yüksek olduğu anlamına geliyor. (Bu ilişki mantıklı, çünkü yüksek BMI genellikle kronik hastalık riskinin artmasıyla ilişkilidir.)

Bu ilişkinin gücünü kontrol etmek için, verilere en iyi uyan çizgi olan bir **regresyon çizgisi** ekleyebilirsiniz. Bunu, komutu **sns.regplot** olarak değiştirerek yaparız.

sns.regplot(x=insurance\_data['bmi'], y=insurance\_data['charges'])

<AxesSubplot:xlabel='bmi', ylabel='charges'>

# Color-coded scatter plots

Nokta grafiklerini, (iki değil) **üç** değişken arasındaki ilişkileri göstermek için kullanabiliriz! Bunu yapmanın bir yolu, noktaları renk kodlamaktır.

Örneğin, sigaranın BMI ve sigorta masrafları arasındaki ilişkiyi nasıl etkilediğini anlamak için, noktaları **'smoker'** sütununa göre renk kodlayıp, diğer iki sütunu ('bmi', 'charges') eksenlerde çizebiliriz.

sns.scatterplot(x=insurance\_data['bmi'], y=insurance\_data['charges'], hue=insurance\_data['smoker'])

<AxesSubplot:xlabel='bmi', ylabel='charges'>

Bu nokta grafiği, sigara içmeyenlerin artan BMI ile biraz daha fazla ödeme eğiliminde olduğunu, sigara içenlerin ise **ÇOK daha fazla** ödeme yaptığını gösteriyor.

Bu gerçeği daha da vurgulamak için, sigara içenlere ve içmeyenlere karşılık gelen iki regresyon çizgisi eklemek üzere **sns.lmplot** komutunu kullanabiliriz. (Sigara içenler için olan regresyon çizgisinin, sigara içmeyenler için olana göre çok daha dik bir eğime sahip olduğunu fark edeceksiniz!)

sns.lmplot(x="bmi", y="charges", hue="smoker", data=insurance\_data)

<seaborn.axisgrid.FacetGrid at 0x7c1b45c4d510>

Yukarıdaki **sns.lmplot** komutu, şimdiye kadar öğrendiğiniz komutlardan biraz farklı çalışır:

* **insurance\_data** içindeki 'bmi' sütununu seçmek için **x=insurance\_data['bmi']** yerine, yalnızca sütun adını belirtmek için **x="bmi"** kullanırız.
* Benzer şekilde, **y="charges"** ve **hue="smoker"** da sütun adlarını içerir.
* Veri setini **data=insurance\_data** ile belirtiriz.

Son olarak, öğrenmeniz gereken bir grafik türü daha var; bu grafik, nokta grafiklerini görmeye alıştığınızdan biraz farklı görünebilir. Genellikle, nokta grafikleri iki sürekli değişken (tıpkı "bmi" ve "charges" gibi) arasındaki ilişkiyi vurgulamak için kullanılır. Ancak, nokta grafiğinin tasarımını ana eksenlerden birinde kategorik bir değişkeni (tıpkı "smoker" gibi) gösterecek şekilde uyarlayabiliriz. Bu grafik türüne **kategorik nokta grafiği** diyeceğiz ve bunu **sns.swarmplot** komutuyla oluşturacağız.

sns.swarmplot(x=insurance\_data['smoker'],

y=insurance\_data['charges'])

/opt/conda/lib/python3.7/site-packages/seaborn/categorical.py:3544: UserWarning: 37.4% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.

warnings.warn(msg, UserWarning)

<AxesSubplot:xlabel='smoker', ylabel='charges'>

/opt/conda/lib/python3.7/site-packages/seaborn/categorical.py:3544: UserWarning: 60.8% of the points cannot be placed; you may want to decrease the size of the markers or use stripplot.

warnings.warn(msg, UserWarning)

Bu grafik bize şunu gösteriyor:

Ortalama olarak, sigara içmeyenler sigara içenlerden daha az ücretlendiriliyor ve

en çok ödeme yapan müşteriler sigara içenler; en az ödeme yapan müşteriler ise sigara içmeyenler.