Introduction

Bu micro-kurs boyunca, veri analizi için en popüler Python kütüphanesi olan **pandas** hakkında her şeyi öğreneceksin.

Bu süreçte, gerçek dünya verileriyle çeşitli uygulamalı alıştırmalar yapacaksın. İlgili rehberleri okurken alıştırmalar üzerinde çalışmanı öneririz.

İlk alıştırmaya başlamak için lütfen buraya tıkla.

Bu rehberde, kendi verini nasıl oluşturacağını ve halihazırda mevcut olan verilerle nasıl çalışacağını öğreneceksin.

# Getting started[¶](https://www.kaggle.com/code/residentmario/creating-reading-and-writing" \l "Getting-started" \t "_self)

Pandas'ı kullanmak için genellikle aşağıdaki kod satırıyla başlarsınız.

import pandas as pd

# Creating data[¶](https://www.kaggle.com/code/residentmario/creating-reading-and-writing" \l "Creating-data" \t "_self)

Pandas'ta iki temel nesne vardır: **DataFrame** ve **Series**.

### **DataFrame[¶](https://www.kaggle.com/code/residentmario/creating-reading-and-writing" \l "DataFrame" \t "_self)**

Bir **DataFrame** bir tablodur. Her biri belirli bir **değere** sahip olan bireysel **girdilerden** oluşan bir dizi içerir. Her bir girdi, bir **satıra** (veya **kayda**) ve bir **sütuna** karşılık gelir.

Örneğin, aşağıdaki basit DataFrame'i ele alalım:

pd.DataFrame({'Yes': [50, 21], 'No': [131, 2]})

| Yes | No |
| --- | --- |
| 0 | 50 | 131 |
| 1 | 21 | 2 |

Bu örnekte, "0, No" girdisinin değeri 131'dir. "0, Yes" girdisinin değeri 50'dir ve bu şekilde devam eder.

DataFrame girdileri tamsayılarla sınırlı değildir. Örneğin, değerleri metin dizileri olan bir DataFrame aşağıdadır:

pd.DataFrame({'Bob': ['I liked it.', 'It was awful.'], 'Sue': ['Pretty good.', 'Bland.']})

| Bob | Sue |
| --- | --- |
| 0 | I liked it. | Pretty good. |
| 1 | It was awful. | Bland. |

Pandas'taki bu **DataFrame** nesnelerini oluşturmak için **pd.DataFrame()** yapıcısını kullanıyoruz. Yeni bir DataFrame tanımlamanın söz dizimi, anahtarları sütun adları (bu örnekte **Bob** ve **Sue**) ve değerleri giriş listeleri olan bir sözlüktür. Bu, yeni bir DataFrame oluşturmanın standart yoludur ve büyük olasılıkla en sık karşılaşacağınız yöntemdir.

Sözlük-liste yapıcısı, değerleri **sütun etiketlerine** atar, ancak **satır etiketleri** için 0'dan başlayan (0, 1, 2, 3, ...) artan bir sayım kullanır. Bazen bu yeterli olur, ancak genellikle bu etiketleri kendimiz atamak isteriz.

Bir DataFrame'de kullanılan satır etiketleri listesine **Dizin (Index)** adı verilir. Yapıcımızda bir **index** parametresi kullanarak bu etiketlere değerler atayabiliriz:

pd.DataFrame({'Bob': ['I liked it.', 'It was awful.'],

'Sue': ['Pretty good.', 'Bland.']},

index=['Product A', 'Product B'])

| Bob | Sue |
| --- | --- |
| Product A | I liked it. | Pretty good. |
| Product B | It was awful. | Bland. |

### **Series**[**¶**](https://www.kaggle.com/code/residentmario/creating-reading-and-writing#Series)

Buna karşın, bir **Series** bir veri değerleri dizisidir. Eğer bir DataFrame bir tabloysa, bir Series bir listedir. Hatta sadece bir listeyle bir Series oluşturabilirsiniz:

pd.Series([1, 2, 3, 4, 5])

0 1

1 2

2 3

3 4

4 5

dtype: int64

Bir Series, esasen bir DataFrame'in tek bir sütunudur. Bu nedenle, daha önce olduğu gibi, **index** parametresini kullanarak Series'e satır etiketleri atayabilirsin. Ancak, bir Series'in sütun adı yoktur, yalnızca genel bir **name** (isim) değeri vardır.

pd.Series([30, 35, 40], index=['2015 Sales', '2016 Sales', '2017 Sales'], name='Product A')

2015 Sales 30

2016 Sales 35

2017 Sales 40

Name: Product A, dtype: int64

Series ve DataFrame birbirleriyle yakından ilişkilidir. Bir DataFrame'i aslında bir araya "yapıştırılmış" bir grup Series olarak düşünmek faydalı olacaktır. Bu rehberin bir sonraki bölümünde bunu daha fazla göreceğiz.

# Reading data files[¶](https://www.kaggle.com/code/residentmario/creating-reading-and-writing" \l "Reading-data-files" \t "_self)

Kendi DataFrame veya Series'ini elle oluşturabilmek kullanışlıdır. Ancak, çoğu zaman kendi verilerimizi elle oluşturmayız. Bunun yerine, zaten mevcut olan verilerle çalışırız.

Veriler, birkaç farklı form ve formatta saklanabilir. Bunların açık ara en temeli, mütevazı CSV dosyasıdır. Bir CSV dosyasını açtığınızda şöyle görünen bir şeyle karşılaşırsınız:

Product A,Product B,Product C,

30,21,9,

35,34,1,

41,11,11

Dolayısıyla bir CSV dosyası, virgülle ayrılmış bir değer tablosudur. Adı da buradan gelir: 'Virgülle Ayrılmış Değerler' (Comma-Separated Values) veya CSV.

Şimdi bu örnek veri kümelerimizi bir kenara bırakalım ve gerçek bir veri kümesinin, bir DataFrame'e okunduğunda nasıl göründüğüne bakalım. Veriyi bir DataFrame'e okumak için **pd.read\_csv()** fonksiyonunu kullanacağız. Bu işlem şu şekilde yapılır:

wine\_reviews = pd.read\_csv("../input/wine-reviews/winemag-data-130k-v2.csv")

Sonuçta ortaya çıkan DataFrame'in ne kadar büyük olduğunu kontrol etmek için şekil niteliğini kullanabiliriz:

wine\_reviews.shape

(129971, 14)

Böylece, yeni DataFrame'imiz 14 farklı sütunda 130.000 kayda sahip. Bu neredeyse 2 milyon girdi demek!

Oluşan DataFrame'in içeriğini, ilk beş satırı getiren **head()** komutunu kullanarak inceleyebiliriz:

wine\_reviews.head()

| Unnamed: 0 | country | description | designation | points | price | province | region\_1 | region\_2 | taster\_name | taster\_twitter\_handle | title | variety | winery |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 0 | Italy | Aromas include tropical fruit, broom, brimston... | Vulkà Bianco | 87 | NaN | Sicily & Sardinia | Etna | NaN | Kerin O’Keefe | @kerinokeefe | Nicosia 2013 Vulkà Bianco (Etna) | White Blend | Nicosia |
| 1 | 1 | Portugal | This is ripe and fruity, a wine that is smooth... | Avidagos | 87 | 15.0 | Douro | NaN | NaN | Roger Voss | @vossroger | Quinta dos Avidagos 2011 Avidagos Red (Douro) | Portuguese Red | Quinta dos Avidagos |
| 2 | 2 | US | Tart and snappy, the flavors of lime flesh and... | NaN | 87 | 14.0 | Oregon | Willamette Valley | Willamette Valley | Paul Gregutt | @paulgwine | Rainstorm 2013 Pinot Gris (Willamette Valley) | Pinot Gris | Rainstorm |
| 3 | 3 | US | Pineapple rind, lemon pith and orange blossom ... | Reserve Late Harvest | 87 | 13.0 | Michigan | Lake Michigan Shore | NaN | Alexander Peartree | NaN | St. Julian 2013 Reserve Late Harvest Riesling ... | Riesling | St. Julian |
| 4 | 4 | US | Much like the regular bottling from 2012, this... | Vintner's Reserve Wild Child Block | 87 | 65.0 | Oregon | Willamette Valley | Willamette Valley | Paul Gregutt | @paulgwine | Sweet Cheeks 2012 Vintner's Reserve Wild Child... | Pinot Noir | Sweet Cheeks |

pd.read\_csv() fonksiyonu, belirtebileceğiniz 30'dan fazla isteğe bağlı parametre ile oldukça zengindir. Örneğin, bu veri kümesinde, pandas'ın otomatik olarak algılamadığı, CSV dosyasında yerleşik bir dizin olduğunu görebilirsiniz. Pandas'ın o sütunu dizin olarak kullanmasını sağlamak için (sıfırdan yeni bir tane oluşturmak yerine), bir **index\_col** belirtebiliriz.

wine\_reviews = pd.read\_csv("../input/wine-reviews/winemag-data-130k-v2.csv", index\_col=0)

wine\_reviews.head()

| ountry | description | designation | points | price | province | region\_1 | region\_2 | taster\_name | taster\_twitter\_handle | title | variety | winery |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | Italy | Aromas include tropical fruit, broom, brimston... | Vulkà Bianco | 87 | NaN | Sicily & Sardinia | Etna | NaN | Kerin O’Keefe | @kerinokeefe | Nicosia 2013 Vulkà Bianco (Etna) | White Blend | Nicosia |
| 1 | Portugal | This is ripe and fruity, a wine that is smooth... | Avidagos | 87 | 15.0 | Douro | NaN | NaN | Roger Voss | @vossroger | Quinta dos Avidagos 2011 Avidagos Red (Douro) | Portuguese Red | Quinta dos Avidagos |
| 2 | US | Tart and snappy, the flavors of lime flesh and... | NaN | 87 | 14.0 | Oregon | Willamette Valley | Willamette Valley | Paul Gregutt | @paulgwine | Rainstorm 2013 Pinot Gris (Willamette Valley) | Pinot Gris | Rainstorm |
| 3 | US | Pineapple rind, lemon pith and orange blossom ... | Reserve Late Harvest | 87 | 13.0 | Michigan | Lake Michigan Shore | NaN | Alexander Peartree | NaN | St. Julian 2013 Reserve Late Harvest Riesling ... | Riesling | St. Julian |
| 4 | US | Much like the regular bottling from 2012, this... | Vintner's Reserve Wild Child Block | 87 | 65.0 | Oregon | Willamette Valley | Willamette Valley | Paul Gregutt | @paulgwine | Sweet Cheeks 2012 Vintner's Reserve Wild Child... | Pinot Noir | Sweet Cheeks |