



Özet

- Pandas, temel olarak ilişkisel veya etiketli verilerle hem kolay hem de sezgisel olarak çalışmak için yapılmış açık kaynaklı bir kütüphanedir. Sayısal verileri ve zaman serilerini işlemek için çeşitli veri yapıları ve işlemler sağlamaktadır. Bu kütüphane, NumPy kütüphanesinin üzerine inşa edilmiştir. Pandas kütüphanesi hızlıdır ve kullanıcılar için yüksek performans ve üretkenliğe sahiptir.
- Pandas kütüphanesi ilk olarak Wes McKinney tarafından 2008 yılında AQR Capital Management'ta çalışırken geliştirilmiştir. Daha sonra aynı şirkette çalışan Chang She, 2012 yılında Pandas kütüphanesine ikinci büyük katkıda bulunan kişi olarak katılmıştır.
- Pandas serisi, her türden veriyi (tamsayı, dize, float, python nesneleri vb.) tutabilen tek boyutlu etiketli bir dizidir. Eksen etiketleri topluca dizin olarak adlandırılır. Etiketlerin benzersiz olması gerekmez, ancak dizinlenebilir ve anlaşılabilir olmalıdır. Pandas kütüphanesindeki nesneler hem tamsayı hem de etiket tabanlı dizinlemeyi destekler ve dizini içeren işlemleri gerçekleştirmek için bir dizi yöntem sağlar. Pandas serisi, mantıksal olarak bir excel sayfasına oldukça benzemektedir.
- Şimdi gerçekte dünyadan veri setleri yüklenerek bir Pandas serisi oluşturmaya çalışalım. Depolama SQL Veritabanı, CSV dosyası, bir Excel dosyası olabilir. Pandas serileri; listelerden, sözlükten ve skaler bir değerden oluşturulabilir.
- Pandas DataFrame yapısı, etiketli eksenlere (satırlar ve sütunlar) sahip iki boyutlu değişkenlere sahip, potansiyel olarak heterojen şekilde bir tablo veri yapısıdır. DataFrame iki boyutlu bir veri yapısıdır, yani veriler satırlar ve sütunlar halinde tablo şeklinde hizalanır. Pandas DataFrame yapıları; veriler, satırlar ve sütunlar olmak üzere üç ana bileşenden oluşur.
- Pandas kütüphanesi genellikle veri bilimi için kullanılmaktadır. Bunun nedeni, Pandas kütüphanesinin veri bilimi için kullanılan diğer kütüphanelerle birlikte kullanılabilir olmasıdır. NumPy kütüphanesi üstüne inşa edilmiştir. Bu durum, birçok NumPy yapısının Pandas'ta kullanıldığı veya çoğaltıldığı anlamına gelir. Pandas tarafından üretilen veriler genellikle Matplotlib'in fonksiyonlarını çizmek, SciPy'de istatistiksel analiz yapmak, Scikit-learn'de makine öğrenme algoritmaları için girdi olarak kullanılmaktadır. Pandas kütüphanesi herhangi bir metin düzenleyiciden çalıştırılabilir, ancak Jupyter Notebook tüm dosyayı yürütmek yerine belirli bir hücrede kod yürütme yeteneği verildiğinden, bunun için Jupyter Notebook kullanılması önerilir. Jupyter ayrıca Pandas kütüphanesinin DataFrame'lerini ve çizimlerini görselleştirmenin kolay bir yolunu sunar.
- Nddarray/Dict çiftinden DataFrame oluşturmak için öncelikli şart olarak tüm dizi aynı uzunlukta olmalıdır. Herhangi bir indeks seçimi durumunda, bu indeks dizinin uzunluğu kadar olmalıdır.
- Bazı durumlarda sütunların isimleri tam olarak bilinmiyor olabilir. Bu durumda bir DataFrame üzerindeki sütunların isimleri öğrenmek için "columns" komutu kullanılmaktadır.
- Elemanların indeksleri de bilinmiyor olabilir. Bu durumda bir DataFrame üzerindeki verilerin indekslerinin öğrenilmesi için "index" komutu kullanılmaktadır.

DEĞERLENDİRME SORULARI

- Aşağıdakilerden hangisi Pandas kütüphanesinin özelliği değildir?
 - a) İlişkisel yapıda olması
 - b) Etiketli veriler üzerinde işlem yapabilmesi
 - c) Sezgisel yapıda olması
 - d) Web tabanlı olması
 - e) Açık kaynaklı olması
- Pandas kütüphanesinin ortaya çıkış yılı aşağıdakilerden hangisinde doğru verilmiştir?
 - a) 2005
 - b) 2008
 - c) 2011
 - d) 2015
 - e) 2020
- Aşağıdakilerden hangisi Pandas kütüphanesinin avantajlarından biridir?
 - a) Sadece DataFrame üzerindeki tek boyutlu nesneler silinir.
 - b) Sütün eklenemez sadece silinebilir.
 - c) Birleştirme işlemi gerçekleşmez.
 - d) Sadece şekillendirilen nesneler döndürülür.
 - e) Zaman serilerinde kullanılabilir.
- Aşağıdakilerden hangisi Pandas kütüphanesinin kullanmış olduğu veri türlerinden biri değildir?
 - a) Float
 - b) Integer
 - c) List
 - d) Short
 - e) Dict
- Listelerden
II. Sözcüklerden
III. Skaler Değerlerden
- Bir Pandas serisi yukarıda verilenlerden hangisi veya hangilerinden oluşabilir?
 - a) Yalnız I
 - b) Yalnız II
 - c) I ve II
 - d) I ve III
 - e) I, II ve III

6. Aşağıdakilerden hangisinde Pandas DataFrame yapısının ana bileşenleri doğru verilmiştir?
 - a) Tablolar, Demetler, Listeler
 - b) Tablolar, Satırlar, Sütunlar
 - c) Listeler, Satırlar, Sütunlar
 - d) Demetler, Listeler, Veriler
 - e) Veriler, Satırlar, Sütunlar
7. Aşağıdakilerden hangisinde Pandas tarafında üretilen verilerin genellikle çiziminde kullanılan kütüphanedir?
 - a) Scikit-learn
 - b) Matplotlib
 - c) SciPy
 - d) NumPy
 - e) PyQt
8. Aşağıdakilerden hangisi DataFrame üzerindeki sütunların isimlerini öğrenmek için kullanılan komuttur?
 - a) read
 - b) columns
 - c) index
 - d) head
 - e) tail
9. Aşağıdakilerden hangisi bir DataFrame'in merkezi eğilimini tespit etmek için kullanılan komuttur?
 - a) index
 - b) memory_usage
 - c) astype
 - d) loc
 - e) describe
10. Aşağıdakilerden hangisi bir Python nesnesini belirli bir veri türüne dönüştürmek için kullanılır?
 - a) astype
 - b) value_counts
 - c) merge
 - d) fillna
 - e) sort_values

Cevap Anahtarı

1.d, 2.b, 3.e, 4.d, 5.e, 6.e, 7.b, 8.b, 9.e, 10.a