

# Data Science Boot Camp

Learn via: **Classroom / Virtual Classroom / Online**

Duration: **5 Gün**

## Overview

### Data Science Boot Camp Sertifikasyon Programı

Introductory Python, Data Science with Python: Data Analysis and Visualization, Big Data with Hadoop and Spark eğitimlerinden oluşan toplam 5 günlük hızlandırılmış bir programdır.

### Introductory Python (1 gün)

Bu eğitim, programcılık konusunda bir geçmişi olmayan ancak temel Python programcılığını öğrenmek isteyen, bilgisayar konusunda bilgi sahibi kişilere yöneliktir. Eğitim, indirilen dosyaları analize uygun hale getirme anlamını taşıyan "veri düzenlemesini" öğrenmek isteyenler içindir. Liste ve dizi düzenlemesi, kontrol yapıları, basit veri analizi paketleri gibi bu dilin temel noktalarına yoğunlaşmakta ve web'den veri indirmek için kullanılacak modülleri tanıtmaktayız.

Bu Python'a Giriş eğitimi, programcılık konusunda bir geçmişi olmayan ancak temel Python programcılığını öğrenmek isteyen, bilgisayar konusunda bilgi sahibi kişilere yöneliktir. Eğitim, indirilen dosyaları analize uygun hale getirme anlamını taşıyan "veri düzenlemesini" öğrenmek isteyenler içindir. Liste ve dizi düzenlemesi, kontrol yapıları, basit veri analizi paketleri gibi bu dilin temel noktalarına yoğunlaşmakta ve web'den veri indirmek için kullanılacak modülleri tanıtmaktayız.

### Data Science with Python: Data Analysis and Visualization (2 gün)

Bu eğitim, Python programlama diliyle veri bilimine kapsamlı bir giriş niteliğindedir. Eğitim, temel programlama bilgisine sahip olan ve bu bilgisini bir sonraki seviyeye çıkartmak isteyen kişilere yöneliktir. Python'da farklı veri yapılarıyla nasıl çalışılacağını tanıtmakta ve numpy, scipy, pandas, matplotlib ve seaborn'un da aralarında yer aldığı en popüler veri analitiği ve görselleştirme modüllerini kapsamaktadır. Kodların sonuçlarını göstermek ve kurs boyunca kodları interaktif bir şekilde değiştirmek için Lpython Notebook'u kullanmaktayız.

### Genel Bilgi

Bu eğitim, Veri Analizi ve Görselleştirmesi için Python'a kapsamlı bir giriş niteliğindedir. Eğitim, temel programlama bilgisine sahip olan ve bu bilgisini bir sonraki seviyeye çıkartmak isteyen kişilere yöneliktir. Python'da farklı veri yapılarıyla nasıl çalışılacağını tanıtmakta ve numpy, scipy, pandas, matplotlib ve seaborn'un da aralarında yer aldığı en popüler Python veri analizi ve görselleştirme modüllerini kapsamaktadır. Kodların sonuçlarını göstermek ve kurs boyunca kodları interaktif bir şekilde değiştirmek için Lpython Notebook'u kullanmaktayız.

### Ön Koşullar

Temel veri tipleri (örn. dizi, nümerik), veri yapıları (örn. liste, çok öğeli, sözlük) konusunda iyi bir bilgiye sahipseniz ve listeleri kavrama ve devam eden döngüler ile ilgili kavramlara aşinaysanız Veri Analizi ve Görselleştirme için Python kursuna katılacak düzeydesiniz demektir. Bu temel Python programcılık konularını kurs sırasında da ele alacağız ancak bu konuları göreceli olarak daha hızlı bir şekilde geçeceğiz.

### 3.Big Data with Hadoop and Spark (2 gün)

Bu, Büyük Veri teknolojilerinin Hadoop ve Spark ekosistemlerine uygulamalı giriş imkanı sunan bir programdır. Kursta Apache Hadoop'un HDFS, veri yayını ile MapReduce, Hive ve Spark gibi temel bileşenleri ele alınacaktır. Programlama Python'da yapılacaktır. Kurs, örneklerimiz için ihtiyaç duyulan Python kavramlarının gözden geçirilmesi ile başlayacaktır. Eğitim formatı interaktiftir. Öğrenciler kursa kendi dizüstü bilgisayarlarını getirmelidir. Çalışmamızı AWS'de (Amazon Web Services) yapacağız ve AWS'ye nasıl bağlanılacağı ve bir nasıl bir hesap alınacağı konusundaki talimatlar da önceden iletilecektir.

### Genel Bilgi

Bu, Büyük Veri teknolojilerinin Hadoop ve Spark ekosistemlerine uygulamalı giriş imkanı sunan ve 6 hafta süren bir programdır. Eğitimde Apache Hadoop'un HDFS, veri yayını ile MapReduce, Hive ve Spark gibi temel bileşenleri ele alınacaktır. Programlama Python'da yapılacaktır. Eğitim, örneklerimiz için ihtiyaç duyulan Python kavramlarının gözden geçirilmesi ile başlayacaktır. Eğitim formatı interaktiftir. Öğrenciler eğitime kendi dizüstü bilgisayarlarını getirmelidir. Çalışmamızı AWS'de (Amazon Web Services) yapacağız ve AWS'ye nasıl bağlanılacağı ve bir nasıl bir hesap alınacağı konusundaki talimatlar da önceden iletilecektir.

## Prerequisites

Bu eğitimden en iyi şekilde faydalanabilmek için Linux dosya sistemlerine, Linux komut satırı arayüzüne (CLI) ve cd, ls, cp gibi temel Linux komutlarına

aşına olmalısınız. Ayrıca Python konusunda temel programlama becerileriniz olmalı ve örneğin bir dizi listesini içiçe bir liste halinde bölmek için eşleştirme() fonksiyonunun nasıl kullanılacağı gibi fonksiyonel programlama konusunda da bilgili olmalısınız. Python'da nesne yönelimli programlama (OOP) gerekli değildir.

## **Outline**

### **Introductory Python (1 gün)**

#### **1. Ünite: Liste düzenleme**

- Basit değerler ve ifadeler
- Fonksiyonları tanımlama, olağan söz dizimlerini ve Lambda söz dizimlerini kullanma
- Listeler
- Bütünleşik fonksiyonlar ve alt simgeleme
- İçiçe geçmiş listeler
- Fonksiyonel operatörler: eşleştirme ve filtreleme
- Listeleri kavrama
- Çoklu liste işlemleri: eşleştirme ve sıkıştırma
- Fonksiyonel operatörler: azaltma

#### **2. Ünite: Diziler ve basit I/O**

- Karakterler
- Karakter listeleri olarak diziler
- Bütünleşik dizi işlemleri
- Dizi listeleri olarak girdi dosyaları
- Baskı tümcesi
- Web'den veri okuma
- İstek paketini kullanma
- Dizi tabanlı web düzenleme (örn. csv dosyalarını ele alma)

#### **3. Ünite: Kontrol yapıları**

- Tümcelere karşı ifadeler
- Döngüler için
- Döngülerdeki değişkenler
- 'Eğer' tümceleri
- Basit ve içiçe geçmiş 'Eğer' tümceleri
- Lambda fonksiyonlarındaki şartlı ifadeler
- Devam eden döngüler
- Kes ve devam et

#### **4. Ünite: Veri Analizi Paketleri**

- NumPy
- Numpy
- Alt simgeleme ve dilimleme
- İşlemler
- Pandas
- Veri Yapısı
- Veri Düzenleme
- Gruplandırma ve Birleştirme

### **Data Science with Python: Data Analysis and Visualization (2 gün)**

#### **1. Ünite: Python'a Giriş**

Python üst düzey bir programlama dilidir. Python'daki temel söz dizimlerini ve veri yapılarını öğreneceksiniz. Kodları, etkileşimli ve keşifçi bir bilişim ortamı için güçlü ve verimli bir ortam sağlayan mükemmel bir araç olan Lpython Notebook'da gösteriyor ve çalıştırıyoruz.

- Lpython Notebook'a giriş
- Python'daki temel nesnelere
- Değişkenler ve kendi kendini tanımlayan fonksiyonlar
- Kontrol akışı
- Veri yapıları

#### **2. Ünite: Python ile Daha Derinleri Keşif**

Python, nesne yönelimli bir programlama (OOP) dilidir. OOP konusunda temel seviyede bilgi sahibi olunması, Python kodlarının nasıl çalıştığını anlamanıza

yardımcı olacaktır. Ayrıca kirliliği ve yapılandırılmamış verilerle de ilgilenmeniz gerekecektir. Düzenli ifadelerin uygulanması gibi verilerinizi temizlemenin bir sürü yolunu öğreneceksiniz.

- Nesne yönelimli programlamaya giriş
- Dosyalarla nasıl ilgilenilir
- Python komut metinlerini çalıştırma
- Dizileri ele alma ve işleme

### 3. Ünite: Bilimsel Bilişim Araçları

Python'u veri analizi için güçlü kılan iki bilimsel bilişim modülü vardır: Numpy ve Scipy. Numpy, Python'da bilimsel bilişim için temel pakettir. SciPy ise bilimsel bilişimi ele alan ve her geçen gün daha da genişleyen bir paketler topluluğudur.

- Numpy
- Scipy

### 4. Ünite: Veri Görselleştirme

Python, ayrıca "Matplotlib" ve "Seaborn" kullanarak grafik de oluşturabilir. Matplotlib, çizim ve diğer iki boyutlu veri görselleştirmelerinin üretilmesi için kullanılan en popüler Python kitaplığıdır. Seaborn ise Matplotlib'i esas alan bir Python görselleştirme kitaplığıdır. İstatistiksel grafiklerin çizilmesi için üst düzeyde bir arayüz sunar

- Seaborn
- Matplotlib

### 5. Ünite: Pandas ile veri manipülasyonu

Pandas, yapılandırılmış verilerle çalışmak için zengin veri yapıları ve fonksiyonları sunar. Pandas içindeki "DataFrame" nesnesi, R'deki "data.frame" nesnesi gibidir. Pandas da veri manipülasyonunu (filtreleme, seçme, gruplandırma, birleştirme, vs.) R'deki kadar kolay bir hale getirir.

- Pandas

## 3. Big Data with Hadoop and Spark (2 gün)

### 1. Ünite - Giriş: Hadoop, MapReduce, Python

- Büyük Veri ve Hadoop ekosistemi hakkında genel bilgi
- MapReduce kavramı
- HDFS – Hadoop Dağıtık Dosya Sistemi
- MapReduce için Python

### 2. Ünite – MapReduce

- MapReduce için daha fazla Python
- Python veri yayını ile MapReduce'un uygulanması

### 3. Ünite – Hive: Büyük Veri için veritabanı

- Hive kavramları, Hive sorgulama dili (HiveQL)
- Python'daki kullanıcı tanımlı fonksiyonlar (veri yayını kullanılarak)
- Python'dan Hive'a erişme

### 4. Ünite – Pig: MapReduce Kullanılarak Büyük Veri Kümelerini Analiz Platformu

- Apache Pig'e giriş
- Pig'de Veri Tipleri
- Pig Latin
- Pig'in MapReduce'da derlenmesi

### 5. Ünite – Spark

- PySpark kullanılarak Spark'a giriş
- Temel Spark kavramları: RDD'ler, dönüştürmeler, işlemler
- PairRDD'ler ve dönüştürmelerin birleştirilmesi
- Gelişmiş Spark: partisyonlar; ortak değişkenler
- SparkSQL

### 6. Ünite – Proje

- Örnek çalışmalar / Final projeleri