Veri Analizi

Hafta-2 02.03.2021

Değişken nedir?

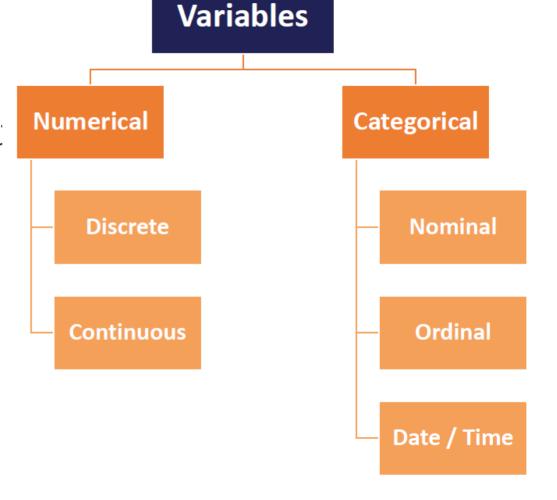
 Değişken: Ölçülebilen veya sayılabilen herhangi bir özellik, sayı veya miktardır.

Değişken Türleri

- Yaş (21, 35, 62, ...)
- Cinsiyet Erkek Kadın)
- Gelir (2000 TL, 3500 TL, 4500 TL, ...)
- Ev fiyatı (100000TL, 350000 TL, 570000 TL, ...)
- Göz rengi (kahverengi, yeşil, mavi, ...)
- Araç markası (Ford, Volkswagen, ...)

Değişken Türleri

- Değişkenler temelde iki türdü
 - Kategorik
 - Sayısal (Numerik)



Ayrık (discrete) değişkenler

Ayrık Değişkenler Değerleri tam sayı (sayılar) olan bir değişken, ayrık olarak adlandırılır.

Örnekler:

- Bir süpermarkette bir müşteri tarafından satın alınan ürün sayısı (10, 50,...)
- Bir borçlunun aktif banka hesaplarının sayısı (1, 4, 7, ...)
- Ailedeki çocuk sayısı

Sürekli (Continuous) değişkenler

- Bir aralıktaki herhangi bir değeri içerebilen bir değişkene sürekli denir. Örnekler:
- Bir süpermarkette bir müşteri tarafından ödenen tutar (32.50 TL, 12 Tl, 5.20 TL,...)
- Ev fiyatı (350.000 , 57.000 , 100.000, ...)
- Bir web sitesinde gezinmek için harcanan süre (3,4 saniye, 5,10 saniye, ...)
- Son aydaki toplam gelirin yüzdesi olarak toplam borç (0.2, 0.001, 0, 0.75,...)

Sıralı (Ordinal) Değişkenler

- Kategorilerin anlamlı bir şekilde sıralanabildiği kategorik değişkenler sıralı olarak adlandırılır. Örnekler:
- Öğrencinin bir sınavdaki notu (A, B, C veya Başarısız)
- Haftanın günleri (Pazartesi = 1 ve Pazar = 7)
- Kategorilerle birlikte eğitim seviyesi: İlkokul, Lise, Kolej mezunu ve
 1'den 4'e kadar olan Doktora

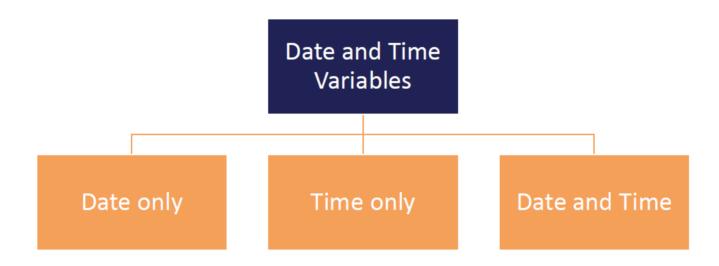
Nominal Değişkenler

- Kategorilerin aldığı değerlerin kendi içerisinde herhangi bir sırasının olmadığı değişken tipidir.
- Örnekler:
- Doğduğu ülke (Türkiye, Arjantin, İngiltere, Almanya)
- Posta kodu
- Araç markası (Ford, Skoda, ...)

Özel Durumlar

- Veri kümelerinde özel durumlar olabilir:
- Kategorilerin sayı olarak kodlandığı kategorik değişkenler (örneğin cinsiyet, erkekler için 0 ve kadınlar için 1 olarak kodlanabilir)
- ID değişkenler: bir gözlemi benzersiz şekilde tanımlayan sayı

Tarih ve Zaman Değişkenleri



Örnek

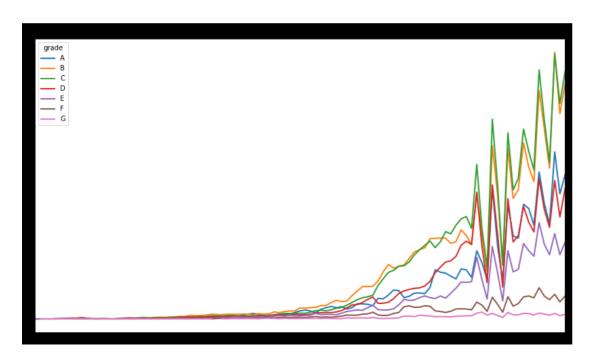
- Tarih ve saat veya tarih saat değişkenleri tarih ve / veya saati değer olarak alır.
- Doğum tarihi ('29 -08-1987 ', '12-01-2012')
- Başvuru tarihi ('2016-Aralık', '2013-Mart')
- Kaza zamanı (12:20:45)
- Ödeme tarihi ('29 -08-1987 15: 20.20 ')

Tarih ve Saat Bilgilerin önişlenmesi

- Tarih ve Saatin Ön İşlenmesi
- Tarih ve saatten bilgileri çıkararak veri setini önemli ölçüde zenginleştirebiliriz.

Örnek

0	5000.0	В	Dec-2011	credit_card	Jan-2015
1	2500.0	С	Dec-2011	car	Apr-2013
2	2400.0	С	Dec-2011	small_business	Jun-2014
3	10000.0	С	Dec-2011	other	Jan-2015
4	3000.0	В	Dec-2011	other	Jan-2016



Karışık (Mixed) Değişkenler -1

- Karışık Değişkenler
- Gözlemler, değerlerinde hem sayıları hem de kategorileri gösterir
- Kabin (Titanic) (A15, B18,...)
- Bilet (Titanik) (A103349)
- Araç tescili (AB500)
- Posta kodu (SE18)

Karışık (Mixed) Değişkenler -2

- Gözlemler, değerleri arasında sayıları veya kategorileri gösterir
- Hesaplarının sayısı (1-100, U, T, M)
- U = bilinmeyen, T = doğrulanmamış, M = eşleşmemiş)
- Devamsız ders sayısı (1-3, M)
- M= mazeretli

Karışık (Mixed) Değişkenler

 Etiketteki bilgileri ve karışık değişkenin sayısını çıkararak veri kümesini önemli ölçüde zenginleştirebiliriz.

Kodların alındığı Kaynak Kitap

Python for Data Science Handbook (Jake VanderPlas)

https://jakevdp.github.io/PythonDataScienceHandbook/index.html

