SF Suç Oranları

Benim üzerinde çalıştığım proje SF suç oranları üzerineydi.

Bu veri setinde farklı nümerik değerler üzerinde çalıştım.

Bunun için kullanacağım fonksiyon KMeans fonksiyonu oldu.

Öncelikle kütüphanemi import ettim.

Veri setimde birçok alan mevcut fakat benim üzerine yoğunlaşacağım kısım enlem ve boylam kısımları.

Enlem ve boylam değerlerim farklı aralıklarda idi bunları aynı aralığa çekmem için önce scale etmem gerekti. (MinMaxScaler)

Daha sonrasında datasetimi projeme ekledim.

Eklediğim datasetindeki verilerimi görüntülemek için kullandığım

dataset.head(1000) kodu ile ilk 1000 satırı listeledim.

metin, yazı tipi, sayı, numara, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

Projemde kullandığım birbirine çok benzeyen ama önemli bir temel fark içeren iki fonksiyondan bahsetmek istiyorum.

train.head ve test.head

Bu iki fonksiyon birçok açıdan birbirine benziyor. İkisi de temel olarak verilerimi listelememe yardımcı oluyor ikisini de kullandım çünkü aralarında önemli bir fark mevcut bunu göstermek istedim.

* train.head = hedef değişkenleri içerir
* test.head = hedef değişkeni içermez

Bu tablodan hedef değişkenlerimi çıkardıktan sonra şu şekil göründü:

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

Projemi açıklarken enlem ve boylam bilgileri üzerinde çalışacağımı belirtmiştim. Burada X Boylam, Y Enlem

Veri setimde herhangi bir null ifadesi var mı diye kontrol ediyorum.

Varsa bunu düzeltmem gerekiyor çünkü sonrasında problem ile karşılaşıyorum.

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturulduVerim temiz çıktı, hepsi null olduğu

devam edebilirim.

Bu dataset 2003-2015 yılları arasındaki verileri içeriyor. Ben sadece 2014 yılları arasındaki verileri inceleyeceğim bunun için filtreleme işlemine ihtiyacım var. Filtreleme işlemini dates sütununa göre yapıyorum.

metin, yazı tipi, sayı, numara, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

Verilerimi daha kolay bir şekilde analiz edebilmek, tablomu daha anlaşılır bir biçime getirebilmek amacı ile ölçeklendirme yaparak X-Y sütunlarını

0 ve 1 aralığına dönüştürdüm.

metin, yazı tipi, sayı, numara, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu Bu işlemi yaptıktan sonra X\_scaled ve Y\_scaled sütunlarımı tabloma ekliyorum.

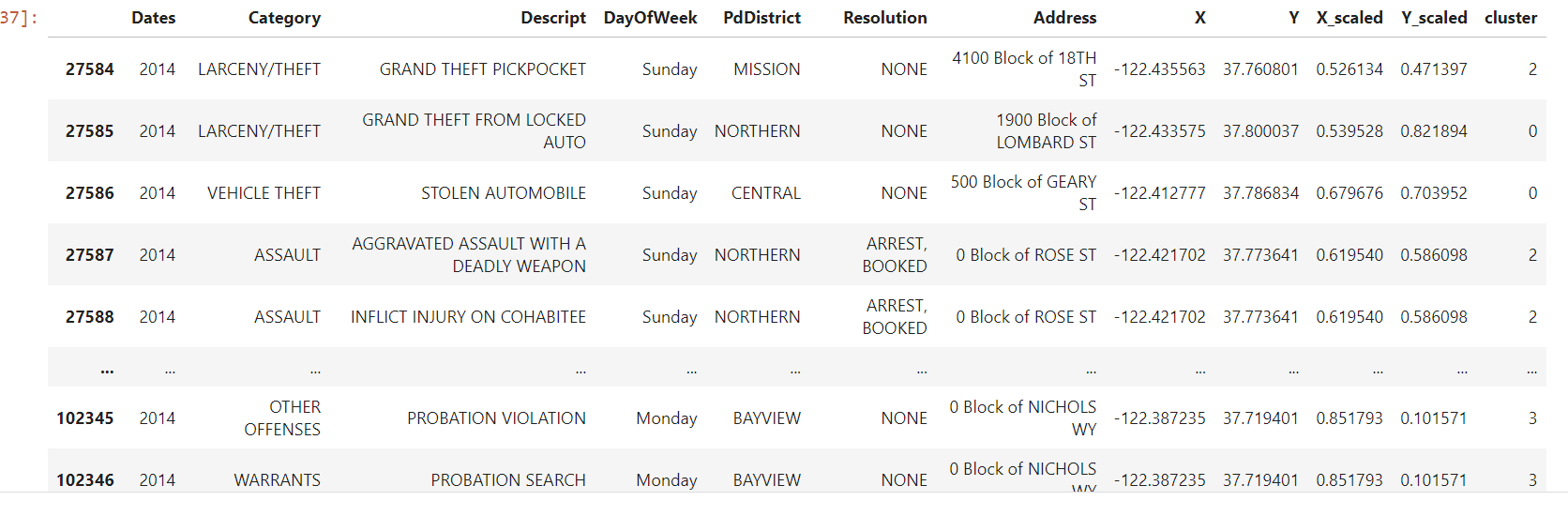
Bu aşamadan sonra yapacağım şey ELBOW Metodunu kullanmak oldu.

Bu metodu kullanmamın temel sebebi bir veri kümesi için en uygun küme sayısını belirlemeye yarayan bir metod olması idi.

Bu kullandığım metod ile aldığım sonuçları bir grafik üzerinde görselleştirdim.

metin, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

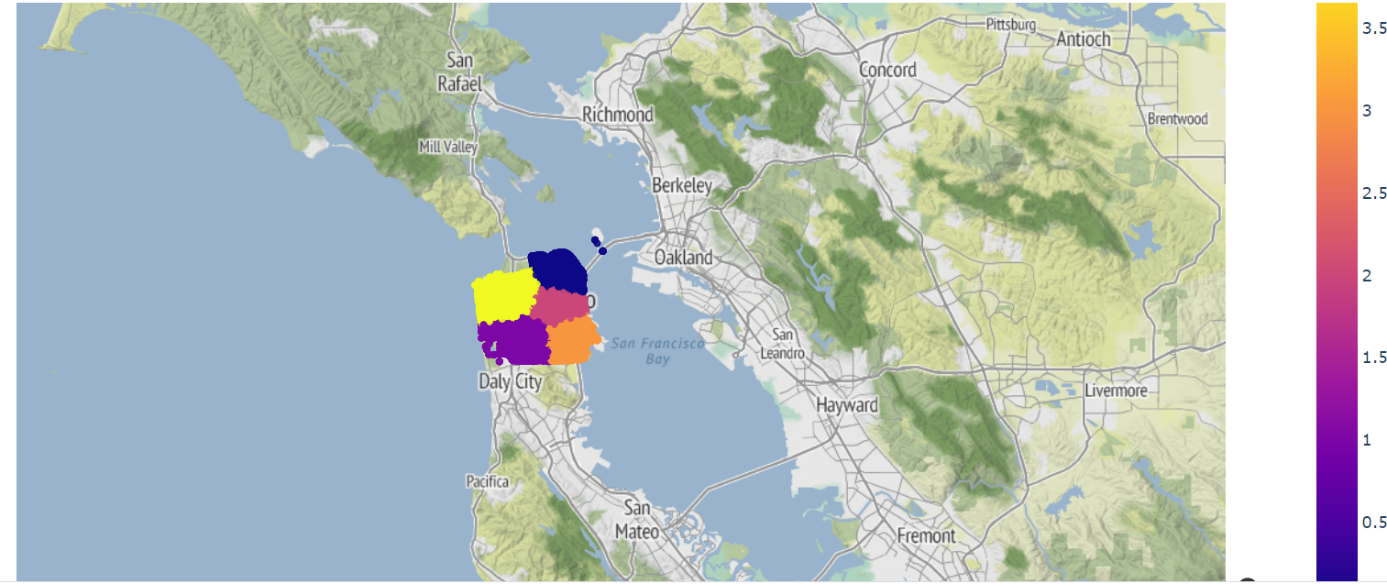


df\_2014 sütunuma cluster adında yeni bir sütun ekledim.

Yaptığım kümeleme sonucu işlemlerini cluster sütunuma ekleyerek veri çerçevesini döndürdü.

Ben suç oranlarını gösteren bir haritalama işlemi yapmak istiyorum. Bunun için de görselleştirme işine yarayan bir kütüphane kullanacağım. Bu kütüphanem ‘plotly’

Haritamın rengine, boyutuna, saydamlığına karar verip kodlarımı yazdım.

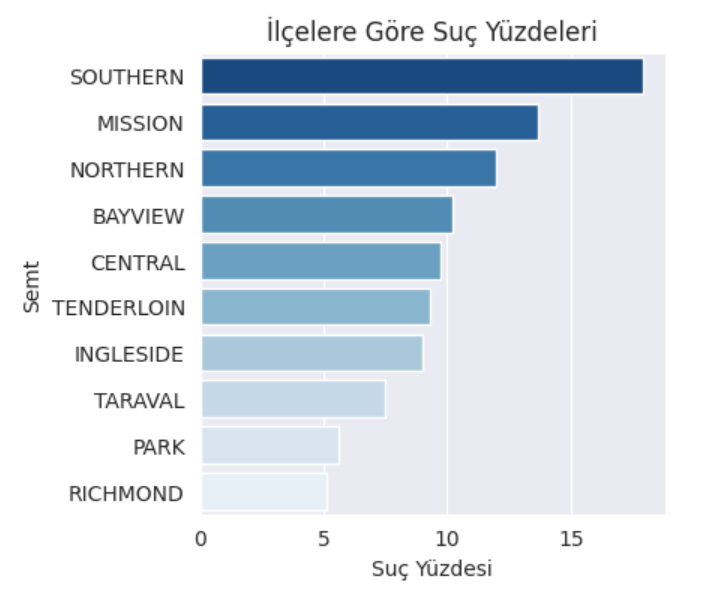


Böyle bir harita elde ettim. (Bunu yapmak 2 günümü aldı)

Sağdaki renklerle belirlenmiş sütun cluster değerlerimi ifade ediyor.

Bir sonraki adımda çeşitli oranlarda grafikler oluşturdum.

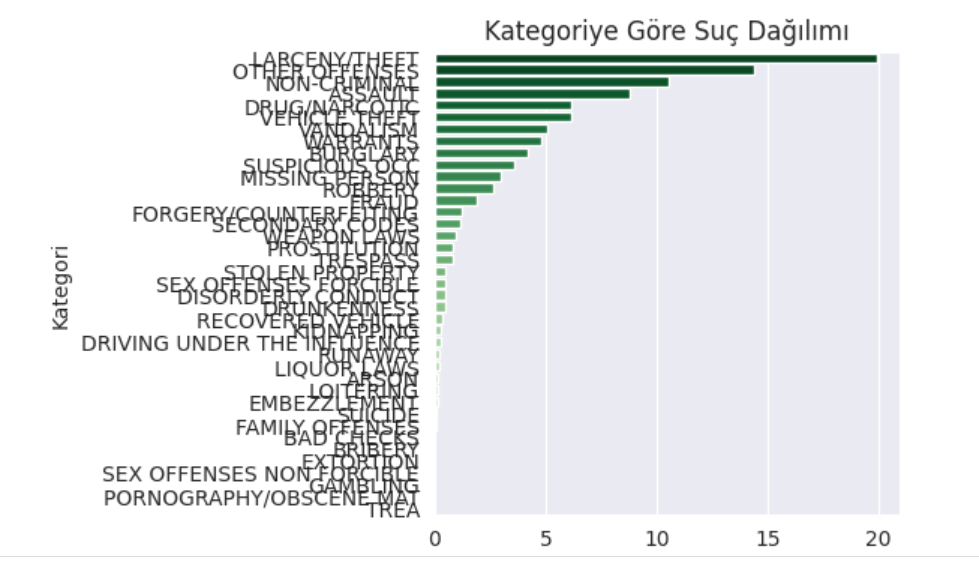
İlk grafiğim suç yüzdeleri grafiğiydi:



Bu grafiklerimi train veri setimden çekerek yaptım.

Bu tabloyu PdDistrict sütunuma göre oluşturdum.

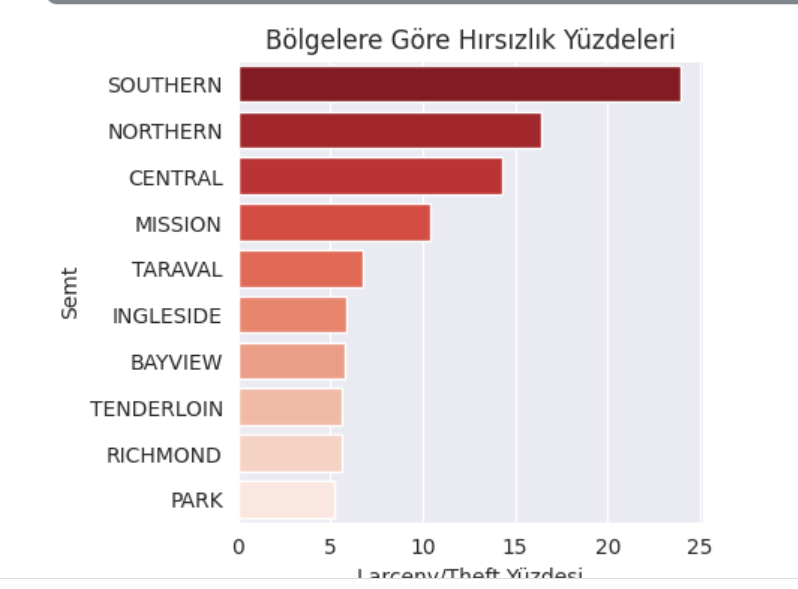
Her bölgenin suç sayısını toplam suç sayısına oranlayarak sonuç çıkardım.



Yine train veri setimden

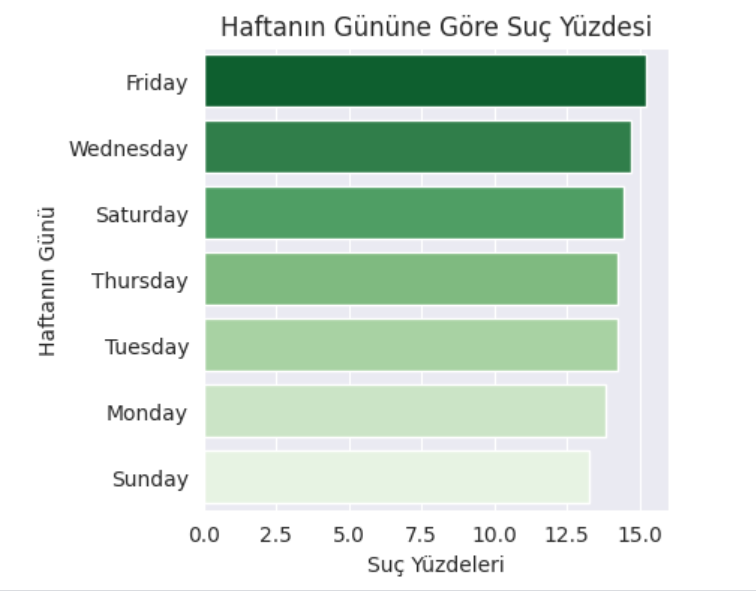
bilgilerimi çekip bu sefer kategoriye göre

bir grafik çıkardım.



'LARCENY/THEFT' suçlarına ait suçların her bölgeye göre yüzde dağılımını gösterecek

böylece hangi bölgenin bu tür suçlarda daha fazla veya az etkilendiğini anlayacağız.



Burada ise DayOfWeek sütunundan verilerimi çekerek

Haftanın gününe göre suç yüzdelerini oluşturdum.

En son adımda ise modelimi oluşturdum,eğittim ve sonuçları CSV dosyasına kaydettim.