"Veri Görselleştirme Dersleri'nin bitirme projesi olan 'Bisiklet Talep Görselleştirme Projesi'ne hoş geldiniz. Bildiğiniz gibi, son zamanlarda kısa mesafe seyahatlerde motorlu toplu taşıma veya özel araçlara alternatif olarak bisikletlere ücretsiz veya uygun erişim sağlanmaktadır. Bu şekilde, trafik sıkışıklığı, gürültü ve hava kirliliğini azaltmayı hedeflemektedir.

Bu proje, Londra bisiklet paylaşımlarının geçmiş verilerini görselleştirme araçlarıyla göstererek verideki mevcut desenleri ortaya çıkarmayı amaçlamaktadır.

Bu, bir makine öğrenme modeli kurmadan ya da gelecege dair tahminlerde bulunmadan önce EDA sürecinin bir parçası olarak veriyi röntgenlememizi sağlayacaktır."

1 = Açık; genellikle açık, ancak sis/hafif sis/yer yer sis/bulutlu bölgelerle birlikte sisli olabilir.

2 = Dağınık bulutlar / az bulutlu

3 = Parçalı bulutlu

4 = Bulutlu

7 = Yağmur / hafif yağmur sağanakları / hafif yağmur

10 = Şimşekli fırtınalı yağmur

26 = Kar yağışı

1. Donmuş sis
2. Veriseti ayni folder içinde değilse, okutmak için path’ini girmem gerekir. Bu, havada bulunan nispi nem miktarını, havanın belirli bir sıcaklıkta tutabileceği maksimum nem miktarına kıyasla temsil eder.
3. Timestamp sutununun veri tipini object olarak görüyoruz bunun datetime olması gerekir. İleride bunu değiştireceğiz.
4. Datasetinin columnlarini tanitacak olursak, timestamp zaman verisi, t1 hava sıcaklığı, t2, hissedilen sıcaklık, hum--> nemlilik, wind\_speed-->rüzgâr hızı, weather code--> havanın durumu (yağmurlu, açık, buzlu.), is\_holiday(resmi veya dini bayramlar dolayısıyla tatil olan hafta içleri), is-weekend-->haftaici mi sonu mu, season-->mevsimler
5. Null veriler olup olmadigina bakiyoruz--> hic yok
6. Duplicate veri olup olmadigina bakiyoruz--> bu rowlarda(satirlarda) duplicate(tekrar) olup olmadiginia bakıyor
7. Duplicate veri olsaydi, drop\_dupliactes ile duplicate olan verileri droplayabiliyoruz. Keep=last dersen mesela sonuncusunu korur, ilkini droplar
8. Discrete verilerin distribution’ina bakmamizi istiyor, bunlar kategorik veriler, tatil mi degil mi, mevsimler gibi. Bunun icin OOM methodunda, for dongusu icin enumarate kullaniyoruz. Df\_discrete olarak bir df tanimlayip, benden istedigi columnlari icine atiyorum, enumarate ile de tum bu columlari gezdirip, hepsini ayri ayri axes yaptiriyorum.
9. Enumarate ile yapmasaydim bunu, tum axis’leri ayri ayri indeksleriyle tanimlyarak, ax[0][1] seklinde mesela ayri ayri cizdirmem gerekecekti
10. Histplot ile cizdirdigimde, y ekseninde frequency getirir
11. Countplot ile cizdirdigimde y ekseninde count getirir
12. For dongusuyle bir kedplot cizdirdigimde ise, bana y ekseninde density’i veriyor.
13. Hue parameter kullandigimda ise palette yerine color kullanmam gerekiyor
14. Grafikleri yorumlayacak olursak, beklenildigi uzere, seasonlar esit dagilim, haftaici gunleri daha fazla, tatil olmayan gunler daha fazla, acik havalar en fazla, ancak 2 de bulutlu, 3 de bulutlu oldugu icin, toplam bulutlu gun sayisi daha fazla denebilir.
15. Timestamp columnunun datatipini datetime’a ceviriyoruz. Bunu yaptiktan sonra da set\_index kodu ile timestamp’i index olarak atiyoruz. Kalici yapiyoruz
16. 6. taskte, bir de bizden timestampin icerdigi yil, ay, gun ve saat bilgilerinden ayri ve yeni columnlar yaparak feature engineering yapmamizi istiyor
17. Bunun icin de bize bir kod vermis kullanmamizi istedigi, bunu kullanarak, bircok zaman verisini iceren timstampten ayri ayri columlar olusturuyoruz
18. Heatmap yapmamizi istiyor 7. soruda. Bu columnlarin birbirleriyle olan korelasyonunu veriyor bize. Kendileri ile olan korelasyonu zaten 1. t1 ile t2 arasindaki ilski 0.99, cunku sicaklik ile hissedilen sicaklik arasinda fazla bir fark yok. Normalde modelim icin bunlardan birini drop etmem gerekiyor.
19. Heatmap sonuclarina baktigimizda ise, cnt ile yani bisiklet kiralama ile en fazla korrelasyon humidity ile yani nemlilik ile cikiyor (-0.46). yani nemlilik arttikca bisiklet kiralama azaliyor.
20. 8. taskte, bizden target variable ile, yani bisiklet kiiralama ile, yani cnt ile, diger sutunlar arasindaki korelasyonu barplot ile gostermemi istiyor.
21. df1.corr() ile bir korelasyonuna bakiyorum.
22. df1.corr().cnt kodu ile de sadece target’in digerleriyle korelasyonuna bakabiliyorum

Bu bize ilk satirda cnt’nin cnt ile korelasyonunu getiriyor. O yuzden ilk satiri drop ederek, cnt’nin diger feature’lar ile korelasyonunu elde ediyorum.

Elde ettigim bu yeni df’yi t\_corr diye bir df’e atiyorum. Bunun uzerinden gosellestirme yapiyorum. Barplot kullanarak column’larin cnt ile korelasyonunu gorsellestirdim. Sonra bir de bunu sort ederek sirasiyla gormek istiyorum ve sort\_values kullanarak value’arina gore siralatiyorum. Daha anlamli cikiyor barplot ve en cok pozitif korelasyon ile en cok negatif korelasyonu bir bakista gorebiliyorum.