**Proje Raporu**

Projede Diabet Dataset i kullandık.8 sorgu,7 görselleştirme kodları kullanarak veritabanından sonuçları çıkarttık. Matplotlib,Pandas,Seaborn ve Matplotlib kütüphanelerini kullandık. Kodlarımızdan başlık olarak bahsedecek olursak veri setinin yapısı, kaç adet veri olduğu, yapısal form hakkında bilgi, eksik veri kontrolü, tekrar eden değer anlama, verileri tablo halinde yazdırma, tablodan istenen değerleri çekme ve 2 değerin ortaklaşa sayısını yazdırma, veriler eksikse sütunların medyan değeri ile doldurma, vücut kitle endeksi değerlendirme listesi oluşturma, yaş ve vücut kitle endeksi görselleştirmesi, diyabeti olan ve olmayan görselleştirmesi, vücut kitle endeksinin görselleştirilmesi, deri kalınlığı ve kan basıncı arasındaki ilişkinin görselleştirilmesi, glukoz histogram görselleştirilmesi, glukoz ve yaş arasındaki ilişkinin görselleştirilmesi, kan basıncının histogram görselleştirilmesini kullandık.

**Kütüphanelerin Kullanımı ve Kütüphanelerin Kullanım Amacı**

**Pandas Kullanımı ve Amacı**

* **Pandas** , kullanımı kolay veri yapıları sunan ve veri üzerinde analiz yapmayı sağlayan açık kaynak kodlu Python kütüphanelerinden birisidir.
* Bu kütüphane temel olarak zaman etiketli serileri ve sayısal tabloları işlemek için bir veri yapısı oluşturur ve bu şekilde çeşitli işlemler bu veri yapısı üzerinde gerçekleştirilebilir.
* Hızlı, güçlü, esnek bir şekilde veri analizi ve işleme yapılabilmesi pandasın genel kullanım amacıdır.
* Biz de pandası
  + csv dosyası yüklemek,
  + veri setinin yapısı hakkında bilgi almak,
  + kaç veri var öğrenmek,
  + eksik veri var mı diye kontrol etmek ,
  + veri tipleri neler görmek,
  + tekrar eden veri var mı öğrenmek ,
  + sütunları yeniden isimlendirmek,
  + Verileri isteğimize uygun filtreleyip listelemek,
  + Yeni sütün ekleyip filtreye uygun doldurtmak

için kullandık.

**MatPlotLib Kullanımı ve Amacı**

* Matplotlib; veri görselleştirmesinde kullandığımız temel python kütüphanesidir.
* Python komut dosyalarında, kabukta, web uygulama sunucularında ve diğer grafik kullanıcı arayüzü araç setlerinde kullanılabilir.
* Biz de matplotlibi
* Yaş ve vücut kitle endeksi arasındaki ilişkiyi tespit etmek için dağılım grafiği oluşturmak,
* Diyabeti olanlarla olmayanların sayı oranını öğrenmek için çubuk grafiği oluşturmak,
* Vücut kitle endeksinin dağılımını öğrenmek için pasta grafiği oluşturmak,
* Deri kalınlığı ve kan basıncının çizim grafiğini görmek,
* Glikozun her iki sınıf özelinde histogramını almak,
* Ortaya çıkan görselleştirmeleri biçimlendirmek

için kullandık.

**Numpy Kullanımı ve Amacı**

* NumPy, Python programlama diline ait çok boyutlu dizilerle ve matrislerle çalışmalarımızda kullanabileceğimiz ileri düzey matematiksel işlemler yapabileceğimiz bir kütüphanedir.
* Biz de NumPy kütüphanesini
* Deri kalınlığı ve kan basıncının grafiğini çizerken bunların değerlerini x ve y ekseninde belirtmek için dizi olarak almada

kullandık.

**Seaborn Kullanımı ve Amacı**

* Verileri görselleştirmek için kullanılan bir Python programa dili kütüphanesidir.
* Kan basıncının histogram grafiğini çizerken Seaborndan yararlandık.

**Veritabanı incelemesi, Veritabanının ne hakkında olduğu, Verilerin tiplerinin belirlenmesi**

Veri tabanımız bazı insanların gebelik, glukoz, kan basıncı, cilt kalınlığı, insülin, vücut kitle endeksi, diyabet soyağacı, yaş ve kişinin diyabet hastası olup olmadığı bilgisini vererek bu bilgiler ışığında verilen bilgilerin diyabet hastalığı ile ilişkisi hakkında bilgi veriyor.

Veri tabanımızda bulunan kolonlar ve açıklamaları:

Gebelikler: Hamilelik Oranı

Glukoz: Oral glukoz tolerans testinde 2 saatlik plazma glukoz konsantrasyonu

Kan Basıncı: Diyastolik kan basıncı (mm Hg)

Cilt Kalınlığı: Triceps cilt kıvrım kalınlığı (mm)

İnsülin: 2 saatlik serum insülini (mu U/ml)

BMI: Vücut kitle indeksi (kg cinsinden ağırlık/ (m cinsinden boy) ^2)

DiabetesPedigreeFunction: Diyabet soyağacı işlevi

Yaş: Yaş (yıl)

Sonuç: Sınıf değişkeni (0 veya 1)

Veri tabanımızda bulunan veri sayısını bulmak için df.info () kodundan faydalanabiliriz:



Sonuçlara göre toplamda 768 tane verimiz ve 9 tane kolonumuz bulunmaktadır.

Veri tabanımızda bulunan verilerin tiplerinin belirlenmesi için yine df.info () kodunu kullanıyoruz.

Bu kod veri tabanımızın yapısal formu hakkında bize bilgi veriyor.metin içeren bir resim

Açıklama otomatik olarak oluşturuldu