

***DERS : VERİ MADENCİLİĞİ***

***ÖDEVİ HAZIRLAYAN ADI : BEYZA***

***ÖDEVİ HAZIRLAYAN SOYADI : TOPCUOĞLU***

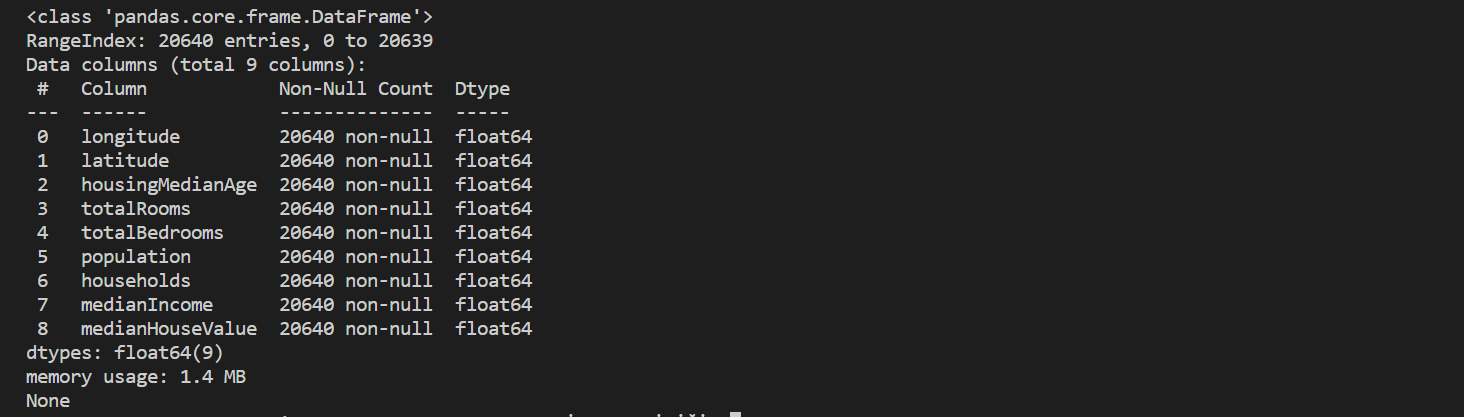
***ÖĞRENCİ NUMARASI : AA220028***

***DERSİN HOCASI : AYŞE CANSEV ÖZDEMİR***

***BÖLÜMÜ:BİLGİSAYAR PROGRAMCILIĞI***

**“datamining.csv” veri setini öncelikle:**

* **Veri türlerini tespit edip aynı veri türünde mi hepsi diye analiz ettim.**

****

**Yorum:Data typelarında herhangi bir farklılık rastlanmamıştır.Hepsinin türü aynıdır.**

* **Data setinde temizleme yapılacak işlem var mı diye analiz ettim.**

**Yorum:Silinecek veri bulunamamıştır.**

***1.SORU (DATA ANALİZ)***

***metin, iç mekan, ekran görüntüsü, gümüş içeren bir resim

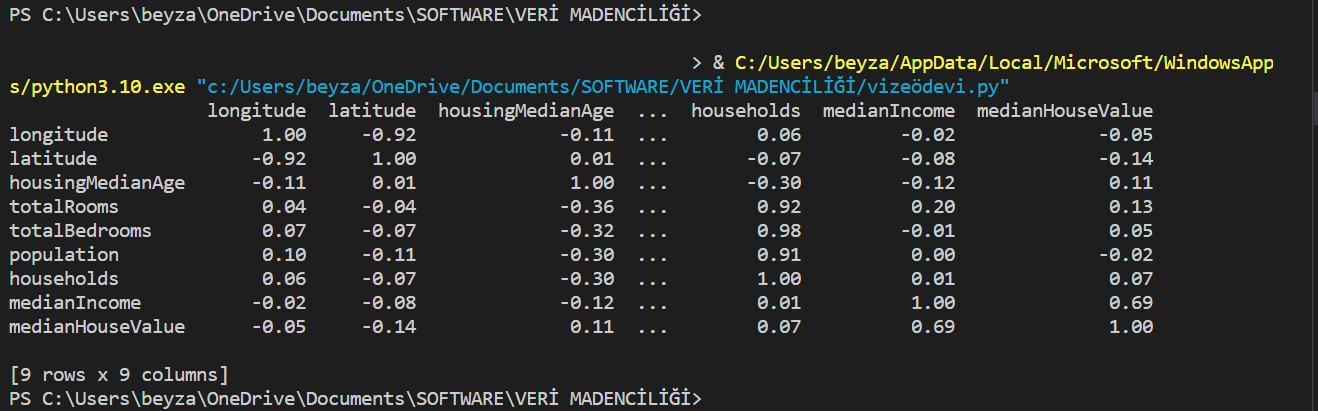
Açıklama otomatik olarak oluşturuldu***

***2.SORU***

***metin, ekran görüntüsü, monitör içeren bir resim

Açıklama otomatik olarak oluşturuldu***

* ***Yukardaki ekran görüntüsünde Kendall,spearmen,pearson hesaplanmıştır.***
* ***Correlasyon ekran matrix görüntüsü***

******

***Yorum: Korelasyon değişkenlerin birbiri ile ilişkisini gösterir. -1 ile 1 arasında değer alır. Bu değerin negatif çıkması iki değişken arasında ters yönlü bir ilişki olduğunu gösterir. Yani bir değişken artarken diğeri azalır. Korelasyonun 1 e yakın olması incelenen iki değişken arasında yüksek anlamlılıkta bir ilişki olduğunu gösterir. Korelasyon değeri ne kadar 1 e yakınsa değişkenler birbiri ile o kadar ilişkilidir.***

***0.06 nın altı ise kötü bir korelasyondur.***

***Örneğin longitude ve latitude değişkenleri arasında negatif yönlü yüksek bir ilişki vardır. Yani bu değerlerden biri artarken diğeri azalacaktır.***

***Pozitif yönde en yüksek korelasyon değeri 0.98 ile totalrooms ve households değişkenleri arasındadır.***

***Bana göre anlamlı ilişkisi olan ikili veri türü totalrooms X households ’dur. Çünkü iyi bir korelasyondur.***

***Başka bir anlamlı ikili veri türü totalbedrooms X households ‘dur.***

***Başka bir anlamlı ikili veri türü population X households ‘dur.***

***CORRELASYONDA ANLAMLI ÇIKAN HER İKİ VERİ TÜRÜ İÇİN;***

***3.LİNEER REGRASYON SLOPE-İNTERCEPT DEĞERİ***

******

* ***GRAFİĞİ***

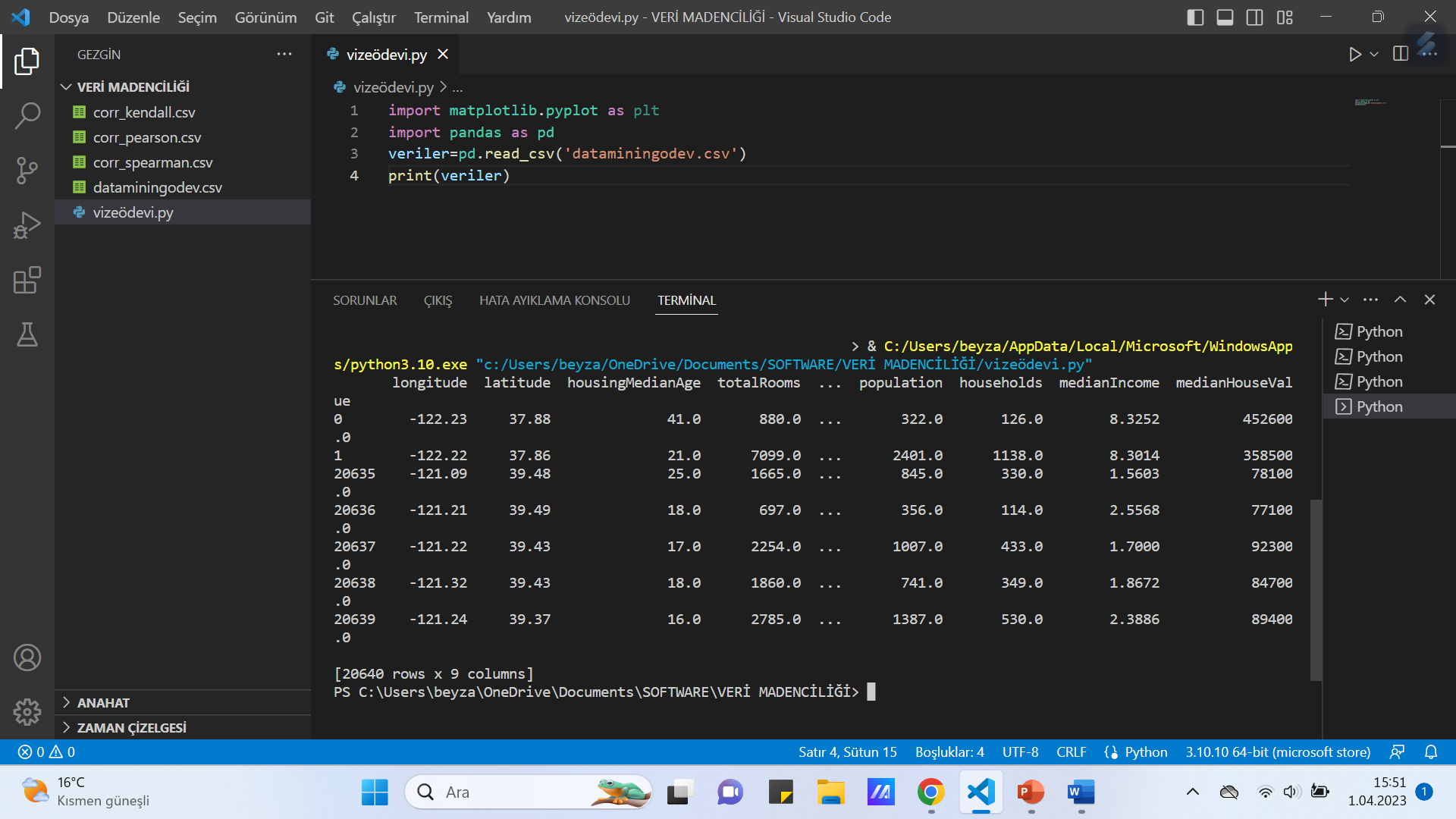
***çizelge içeren bir resim

Açıklama otomatik olarak oluşturuldu***

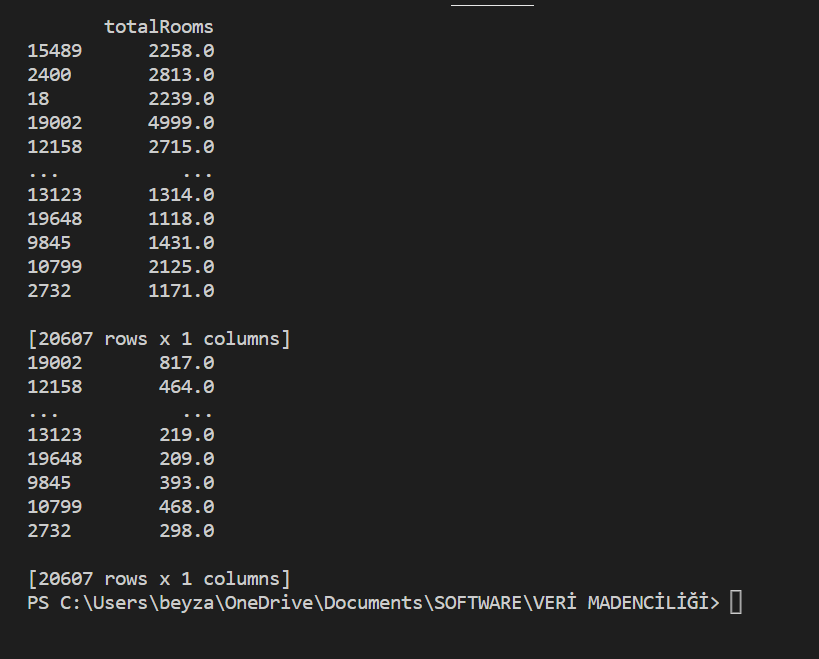
* ***F(X) FONKSİYONU***

***F(x)= 0.16096511+ 75.27377369x1 (Burada x1 ilgilenilen değişkendir.)***

***4.LİNEER REGRASYON YÖNTEMİ***

******

* ***TRAİN-TEST DATA HESAPLAMASI***

******

* ***PLOTTİNG GRAFİK***

***çizelge içeren bir resim

Açıklama otomatik olarak oluşturuldu***

* ***R2 DEĞERİ HESAPLAMASI***

******

* ***R2 değeri açıklama yüzdesidir. Yani kurduğumuz regresyon modelinde bağımsız değişkenler bağımlı değişkenin ne kadarını açıklıyor sorusuna cevap verir. Bu değerin yüksek olması istenir. Bizim örneğimiz için %84 olarak hesaplanmıştır. Bu iyi bir açıklama yüzdesidir.***
* ***Bu sonuç tespitimizin gayet iyi olduğu sonucunu gösteriyor.***
* ***YORUMLAR***

***Veri bir bütün olarak düşünüldüğünde ilk olarak veri setindeki korelasyon matrisi oluşturularak incelenmiştir. Bu inceleme sonucu aralarındaki ilişki yüksek olan totalrooms ve households değişkenleri ele alınarak bir lineer regresyon modeli oluşturulmuştur. Ardından bu iki değişken arasındaki ilişki çizdirilen grafik ile incelenmiştir ve bu inceleme sonucunda aralarında doğrusal bir bağlantı olduğu sonucuna ulaşılmıştır. Veri test ve training olarak iki parçaya ayrılmıştır. Bunlar sonucunda lineer regresyon denklemi olarak f(x) fonksiyonu oluşturulmuştur. Bu fonksiyon ileride veri setine yeni bir değişken eklendiğinde bağımlı değişken (y) değerini bulmakta kullanılacaktır. Ardından R2 değeri incelenmiştir. Bu değer açıklama yüzdesidir. İstatistik literatüründe %70 üstü oldukça iyi bir açıklayıcılık olarak kabul edilmektedir. Bizim değerimiz %84 olarak bulunmuştur. Yani seçilen iki değişken birbirini oldukça iyi açıklamaktadır.***