

**T.C Maltepe University**

**Faculty of Engineering and Natural Sciences**

**Group 9 – Data Science**

**SE 383 01 Project Report**

|  |  |
| --- | --- |
| 150709001 | Ayşe CANDAN |
| 170706906 | Burak Tugay SÜR |
| 180706016 | Ege Nur YEŞİL |
| 180704320 | Mehmet GÜRÜN |
| 170706901 | Melih CÜNEYTER |
| 170706306 | Recai Can SERDAR |

**Advisor  
Dr. Öğr. Üyesi Hüseyin AYGÜN**

**04.01.2021**

1. INTRODUCTION

Depremlerin analizi için çeşitli kurum ve kuruluşların oluşturduğu verilerden yararlanılacaktır. Bu veriler, “.csv” uzantılı bir dosya içerisinde saklanmıştır. İçerisinde ; deprem id, tarih, saat, enlem, boylam, şehir, ülke, deprem mesafesi, deprem derinliği, deprem süresi, yerel büyüklük(Richter), moment büyüklüğü, yüzey dalgası, düzlem dalgası büyüklüğü, diğerleri dışındaki en büyük büyüklük şeklinde veriler saklanmaktadır.

Veri bilimi projeleri, verinin keşfedilmesi ve temizlenmesi ile başlar ve bu işlemler projelerin en çok zaman alan kısımlarıdır. Dolayısıyla verinin keşfi ve temizlenmesi sırasında işleri kolaylaştıracak bir takım kütüphanelere ihtiyaç duyulur. Bu verilerin işlenmesi ve grafik üzerinde gösterilmesi için Python dili kullanılacaktır. Ancak tek başına python yetersiz kalacağı için birkaç kütüphane ve koleksiyon entegre etmek durumundayız. Bu kütüphanelerden aşağıda kısaca bahsedilmiştir.

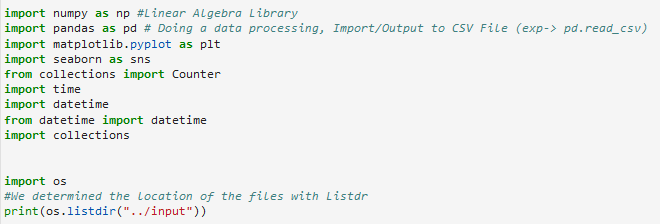
* Numpy : Büyük boyutlu diziler ve matrisler için kullanılır.
* Pandas: Pandas, numpy ın bir uzantısı gibi düşünülebilir. Numpy’ın yetersiz kaldığı noktalarda, Pandas çok kullanışlıdır. Pandas, veri analizi yapılırken kullanılacak temel seriler ve dataframe leri içerisinde barındırır.
* Maplotlib: Çizim kitaplığıdır.
* Seaborn: İstatistiksel grafikler yapmak için kullanılan bir kütüphanedir. Numpy ve Pandas ile tam entegre çalışmaktadır.
* Os: Dosya ve dizinlere kolaylıkla erişim sağlamak ve üzerlerinde değişiklik yapmak için kullanılan kütüphanedir. Kurulum gerektirmez, Python ile birlikte hazır halde gelir. Import edilip kullanmaya başlanılabilir.

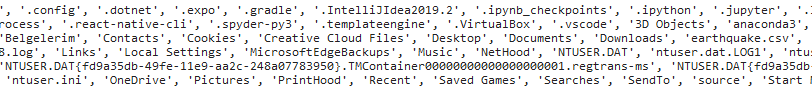
1. PROJECT DATASET, CODE AND SAMPLE OUTPUTS

Dataset içerisinde yer alan veriler, Boğaziçi Üniversitesi tarafından tutulmuştur. Kayıtlı olan bu veriler, 25 -50 enlemlerinde ve 15 – 60 boylamlarındaki tüm depremleri kapsamaktadır. Veriler içerisindeki arama filtresi ise, 27.09.1910 ile 27.09.2017 tarihleri arasındadır. Yoğunluğu 4.0 dan küçük olan çok fazla deprem olduğu için depremler için yoğunluk filtresi en az 3, en fazla 9 olarak ayarlanmıştır. 9 şiddetinin üzerindeki hiçbir veri kaydedilmemiştir. Dosya içerisinde ortalama 20 – 25 bin arası veri vardır.

Bu veriler ham veriler olduğu için üzerinde hiçbir işlem yapılmadan anlaşılması pek mümkün değildir. Bu verilerin kısaca şöyle gösterebiliriz;

|  |
| --- |
|  |

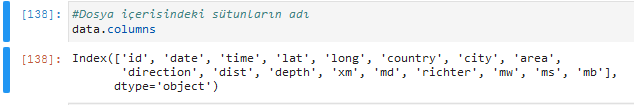
Sistemin çalışması için gerekli olan kütüphaneler uygulama içerisine import edilir. Bu kütüphanelerin tanımları giriş kısmında yapılmıştır.  
  


Dizin altında olan tüm dosyalar, listdir komutu ile listelenir.

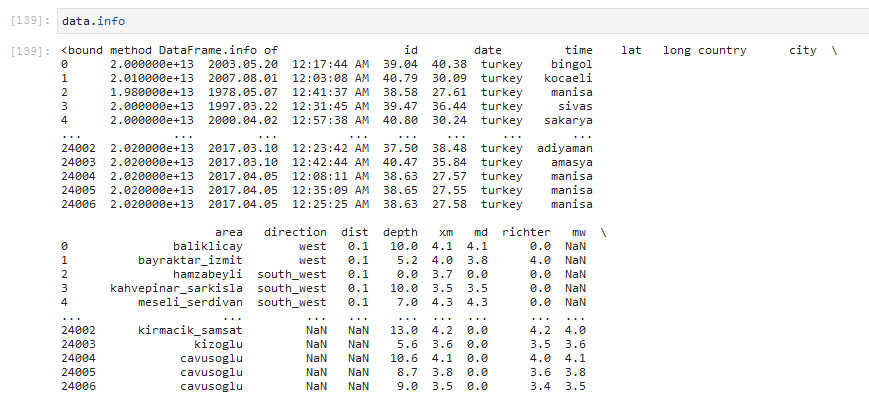
Dizin altındaki tüm dosyalar listelendikten sonra Pandas kütüphanesi ile verilerin bulunduğu .csv uzantılı dosyayı okuyabiliriz. Bunun için;



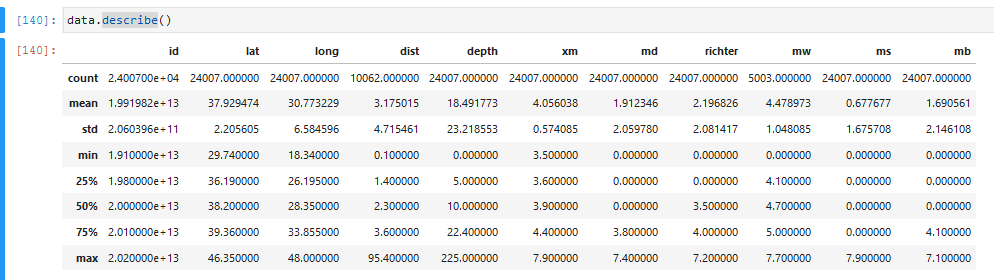
komutu yeterli olacaktır.

Dosya içerisinde bulunan sütunların isimlerini öğrenmek için;

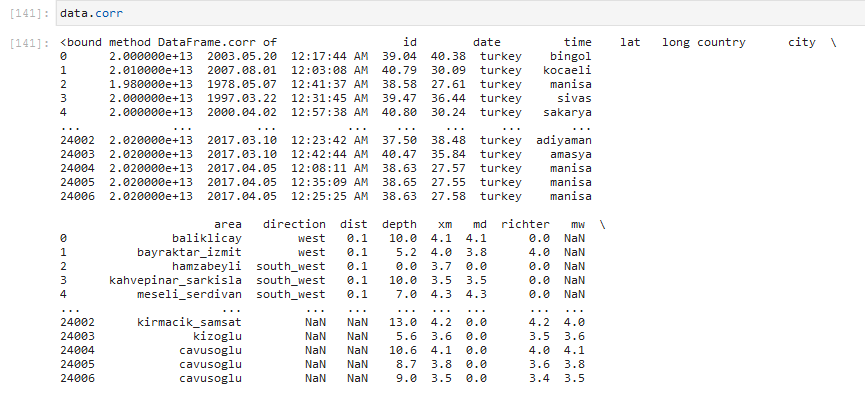
Verilerin detay bilgisi için kullanılan kod ve çıktısı aşağıdaki gibidir.



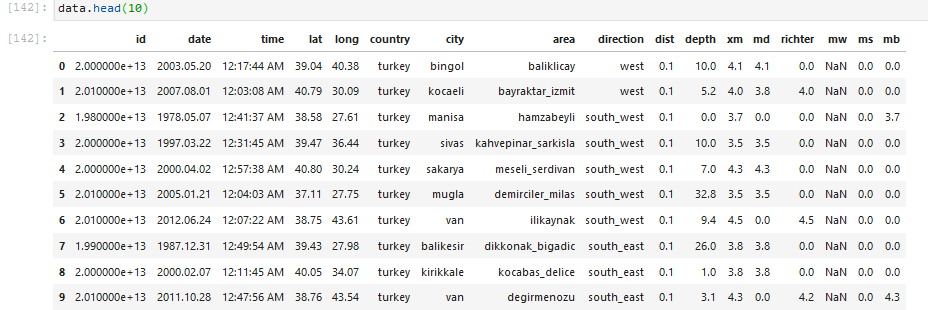
Panda kütüphanesinde bulunan bir diğer kod ise, Describe komutudur. Bir veri bütünü üzerinde yüzdelik dilim, ortalam vs. bazı temel istatistiksel ayrıntıları veya dizi sayısal değerlerini görüntülemek için kullanılır.



Corr komutu ise, veri bütünü içerisindeki tüm sütunların ikili korelasyonunu bulmak için kullanılır. Boş olan değerler hariç tutulur ve sayısal olmayan veriler dikkate alınmaz.



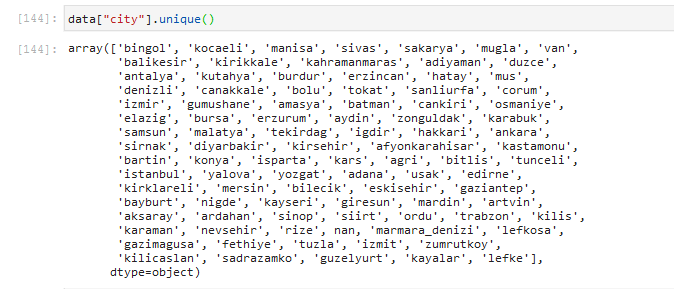
Head, veri bütünü içerisindeki ilk ‘n’ satırını döndürür. Bulunan veri bütünü içerisindeki verilerin doğruluğunu kontrol etmek içinde hızlı bir yoldur. Veri bütünündeki ilk 10 veri listelenmiştir.



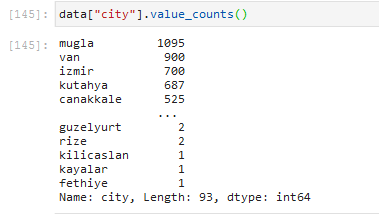
Tail, veri bütünü içerisindeki son ‘n’ satırını döndürür. Satırları sıraladıktan veya ekledikten sonra, verileri hızlı bir şekilde kontrol etmek için kullanılır. Burada 10 adet veri listelenmiştir.

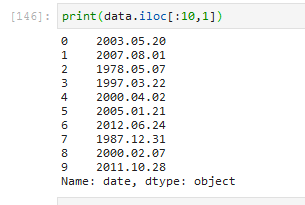


Unique, veri bütünü içerisindeki karma tablodan benzersiz ve görünüm sırasına göre liste çıkarır. Köşeli parantezler içerisinde verilen değişkenin içerisindeki verileri benzersiz şekilde sıralıyor. Aşağıdaki görselde, csv dosyasının içerisinde bulunan veriler ayıklanarak hepsi yalnızca bir kere listelenmiştir.

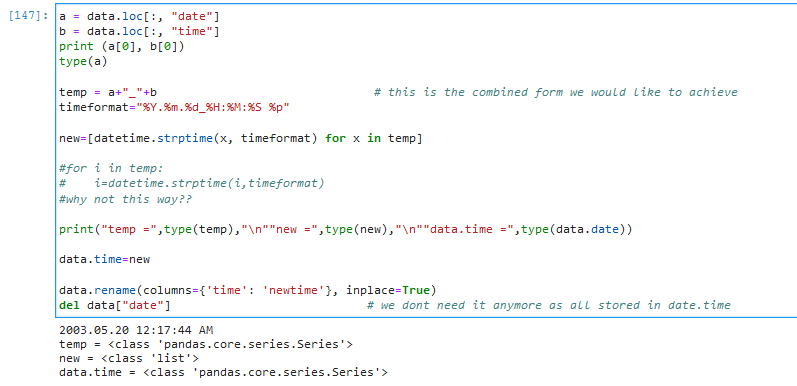


Value\_counts, veri kümesi içerisinde ki verileri analiz edip, bu veriler içerisinden kent bazlı kaç adet deprem kaydı olduğunu listeliyor.



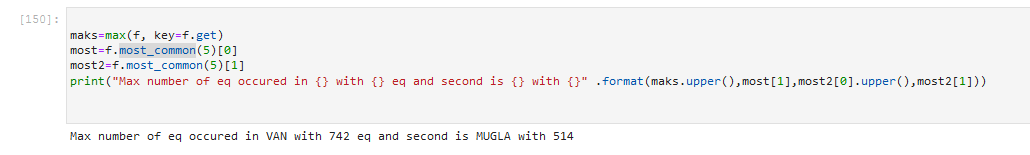
 Iloc, tamsayı konum tabanlı indeksleme için kullanılır.

Bu süreçlerde düzenli listelemek için belirtilen tarih – saat formatına ihtiyac vardır. Aşağıdaki kod satırında ise, istenilen tarih – saat formatında düzenlemeler yapılmaktadır. Tarih, a değişkenine atanıyor. Zaman ise b değişkenine. Daha sonra tarih ve saat biçimi belirtilerek düzenleme yapılıyor. Önceden saklanan verilere ihtiyacımız olmadığı için son satırda önceki verileri silme işlemi yapılıyor.



Counter nesnesi kullanılarak, ülke bilgisi ‘turkey’ olan ve 1’den yüksek deprem verilerinin sayısı listelenmiştir. Analiz edilen veri bütünlerinde 5080 adet veriye ulaşılmış ve 85 adet benzersiz veri oluşturulmuştur.



Most\_common komutu ile en çok deprem yaşanan şehirler listelenmiştir. 

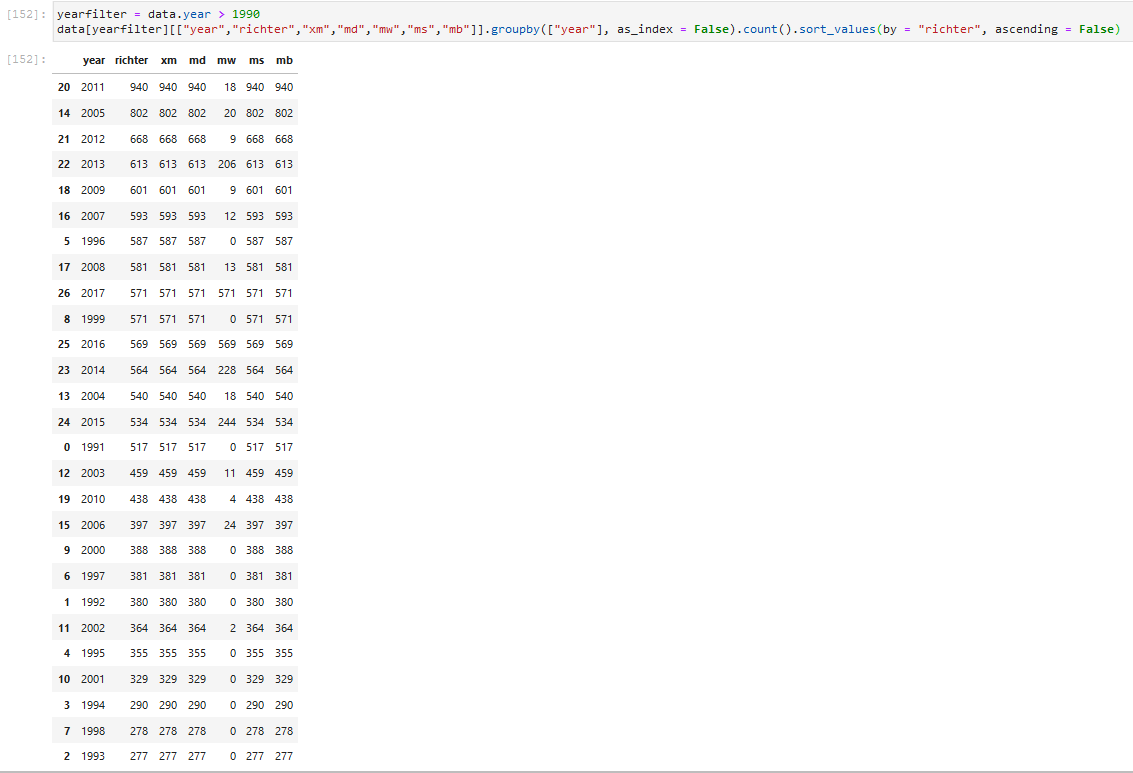
Tarih koşulu ile kullanıcı tarafından girilen tarihten sonra ki yaşanan depremler; yıl, ülke,şehir,bölge,derinlik olarak yıl olarak yakından geçmişe doğru listelenmiştir.



1990 yılından sonra yaşanan depremler filtrelenmiştir. Bu filtre içerisindeki sütun başlıkları aşağıda açıklanmıştır.

* Year – Yıl
* Richter - Yerel büyüklük yada Richter büyüklüğü
* Xm - MD,ML,Mw,MS ve MB dışındaki en büyük büyüklük
* Md - Deprem süresi
* Mw - Moment büyüklüğü
* Ms - Yüzey dalgası büyüklüğü
* Mb - Düzlem dalgası büyüklüğü

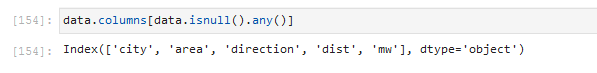
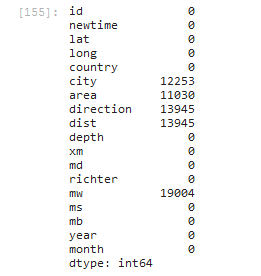
Bu listeleme, yerel büyüklük(Richter) baz alınarak yapılmıştır.



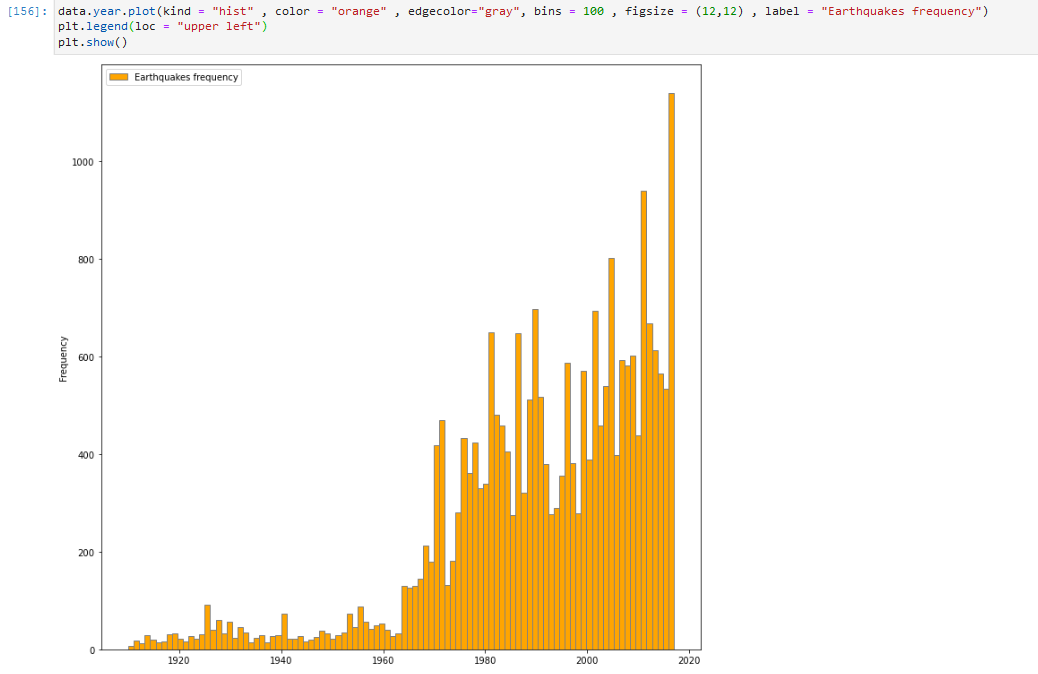
1990 yılından sonra yaşanan depremlerin yıl baz alınarak listelenmesi yapılmıştır. Yıl, Richter büyüklüğü ve boylam baz alınarak en büyük 27 deprem listelenmiştir.



Veri bütünü içerisinde bulunan dizide, herhangi bir değerin eksik olup olmadığını kontrol etmek için isnull() kullanılır. En hızlı yöntem ise .values.any() yöntemidir. Bu analiz içerisinde bulunan sütunlarda, city, area, direction, dist ve mw değerlerinin eksik olup olmadğı kontrol ediliyor.

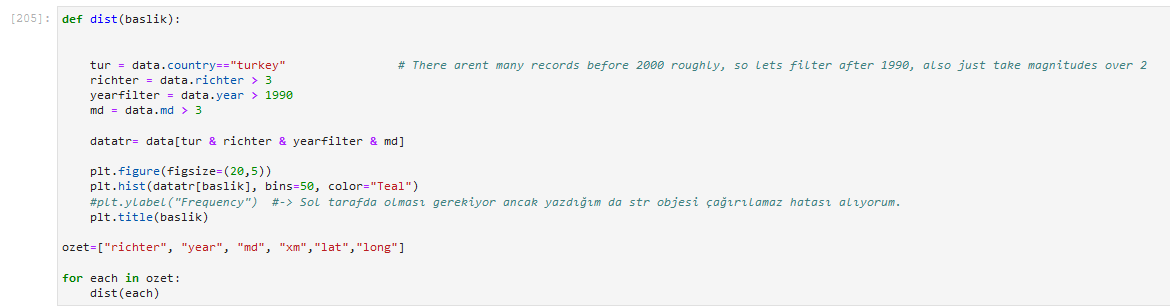
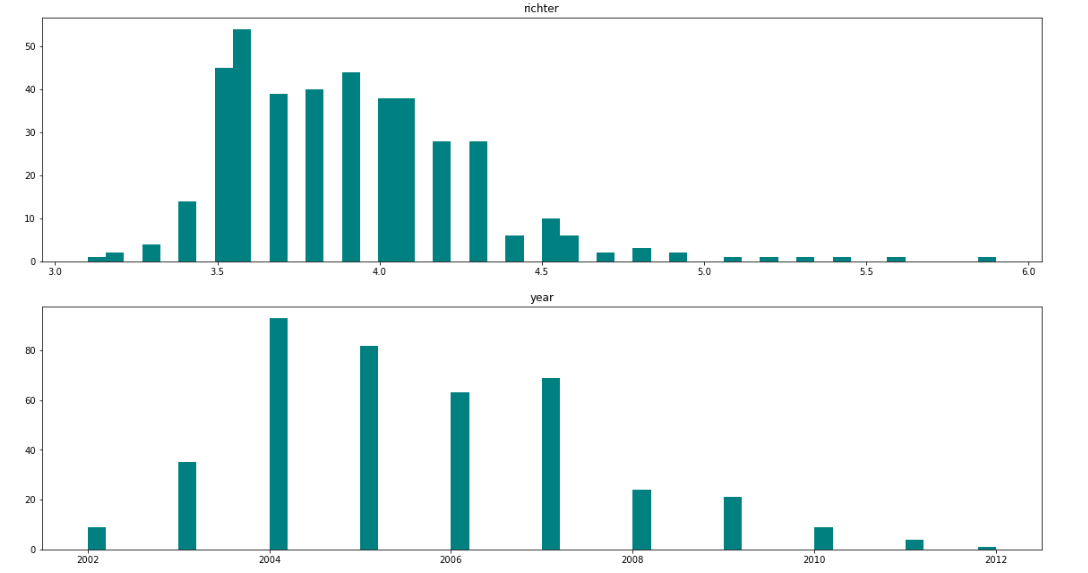


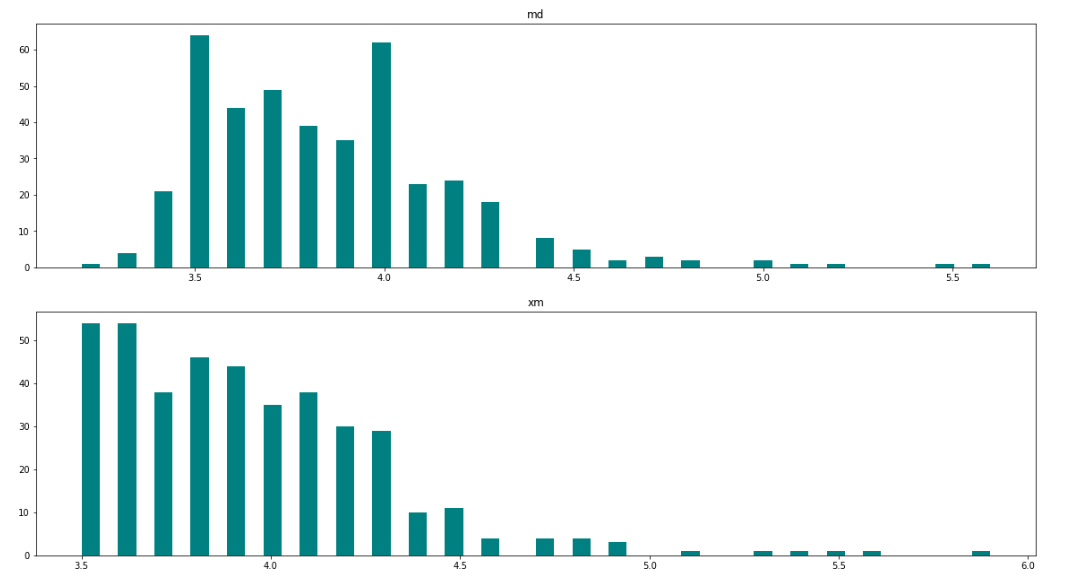
Isnull.sum() fonksiyonu ise, veri kümesindeki eksik değerlerin sayısını döndürür. Eksik değerler ile başa çıkmanın en basit yolu bu veri kümesinde bulunan ve eksik değerler içeren satırları atlamaktır.

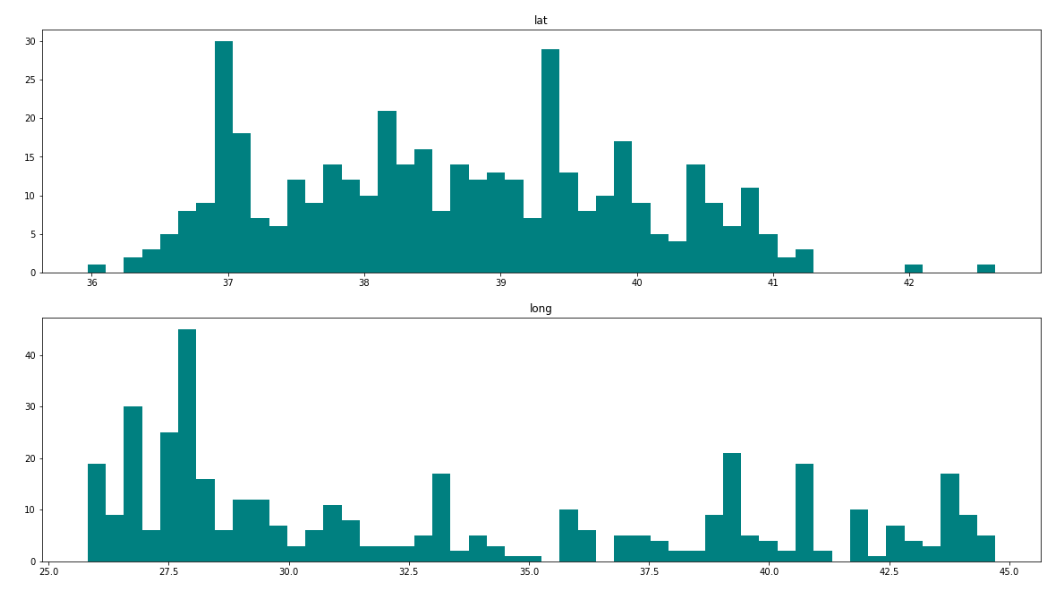
Verilerin grafik üzerinde gösterilmesinde, grafik ile ilgili birçok özelleştirmeyi yapabiliyoruz. Kütüphane içerisinde bulunan grafikler üzerinde seçim yapıp, istediğimiz gibi kullanabilir ve özelleştirebiliriz. Bu değişiklikler aşağıda da görebileceğiniz üzere, renk, çerçeve rengi, grafiğin X-Y düzlemindeki toplam boyutu, grafik açıklamasının ekranın neresinde bulunacağı vs.  
  
Bu grafikte verilerin saklanmaya başladığı tarihten, 2017 yılına kadar olan ve yukarıda gerekli ayarlarını yaptığımız sayıda veri listelenmiştir. 

Farklı bir grafik üzerinde ise, başlıklar üzerinden görselleştirilme yapılmıştır. 2000 den daha fazla veri olduğu için, 1990 yılından itibaren olmak üzere Richter büyüklüğü 3 den büyük olan depremler listelenmiştir.

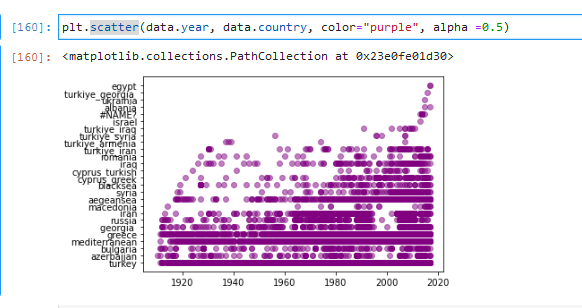
Bu grafiklerin gösterdiği değerler ise sırası ile; Richter, Year, Md, Xm, Lat ve Long grafikleridir.



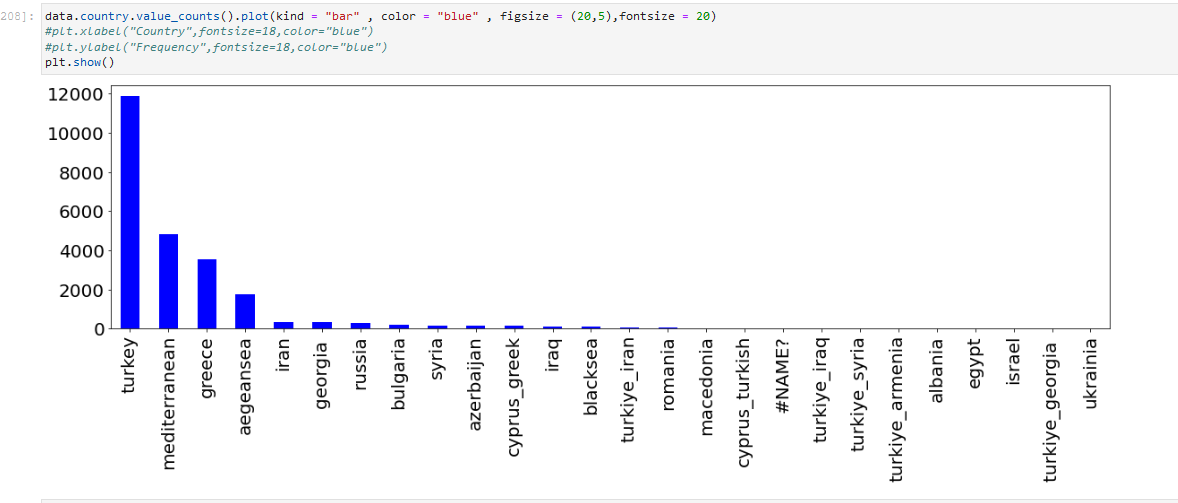




Yuvarlak içerikli grafikler için scatter plot kullanılır. Deprem olma sıklığı olan yerler koyu, az olan yerlerde ise solgun renkler ağırlıklıdır. Yıl ve ülke bazlı sıralama yapılmıştır.



Türkiye sınırlarında yaşanan toplam deprem sayıları ise aşağıda gösterilmiştir.

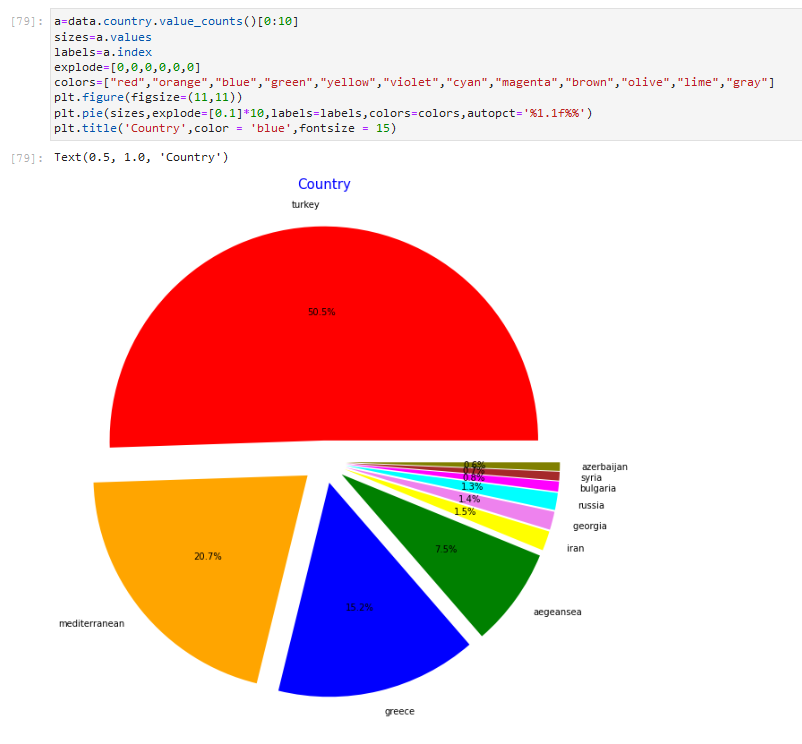


Bu grafikte ise, ülke bazlı sıralama yapılmıştır. Grafik üzerindeki sayılarda ülkesel olarak 100 yıllık süreçte yaşanan depremlerin, bu çemberin % kaçını doldurduğu listelenmiştir.

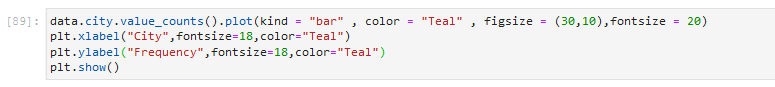
En çok deprem yaşanan 10 ülke baz alınmıştır.

Value\_counts()[0:10]

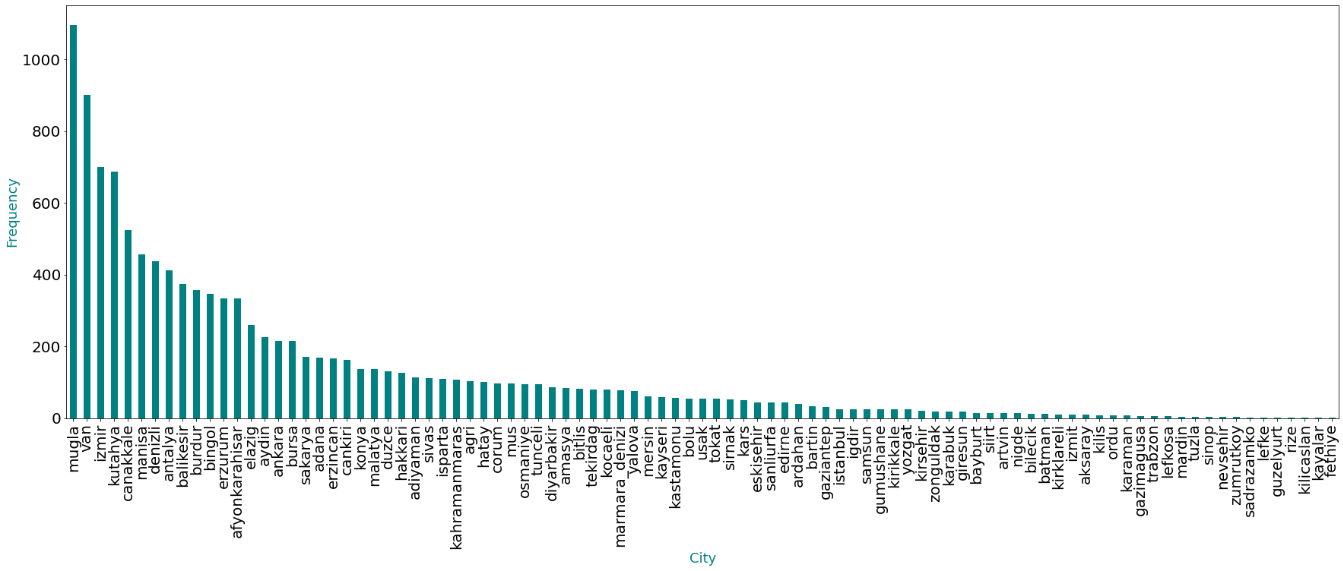
Colors sekmesi altında ise, en büyükten en küçüğe doğru pasta dilimlerine renkler verilmiştir.



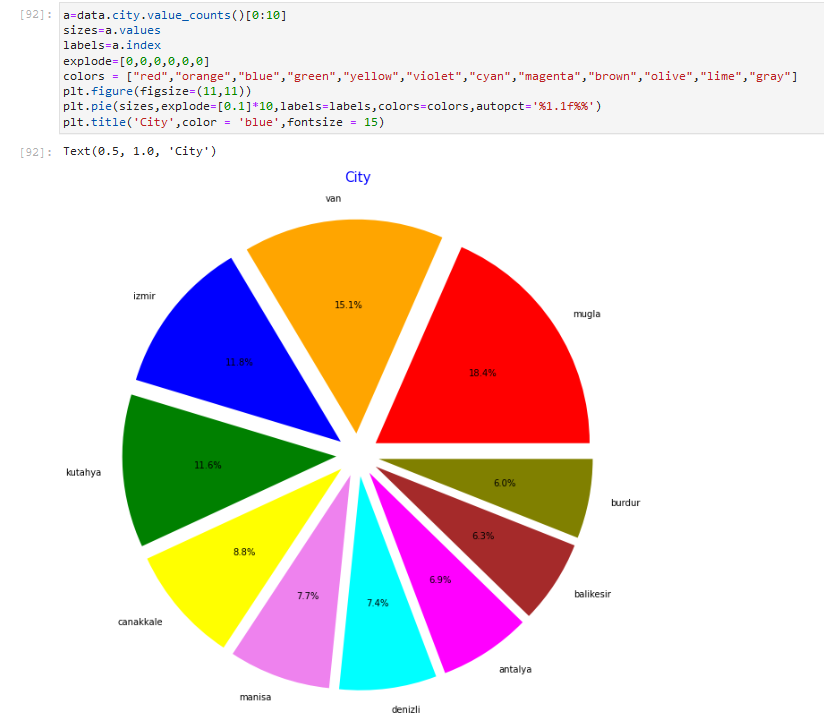
Türkiye sınırları içerisinde bulunan 81 il üzerinde yaşanan depremler ise aşağıda ki kod aracılığıyla listelenmiştir. Toplam deprem sayıları gösterilmektedir ve diğer grafikler gibi yalnızca özelleştirme işlemi yapılmıştır. Value\_count içerisinde bulunan değerler üzerinden veri çekme işlemi gerçekleştirilmektedir. Yalnızca fonksiyonun öncesinde listelemek istenilen değişkenin yazılması yeterlidir.

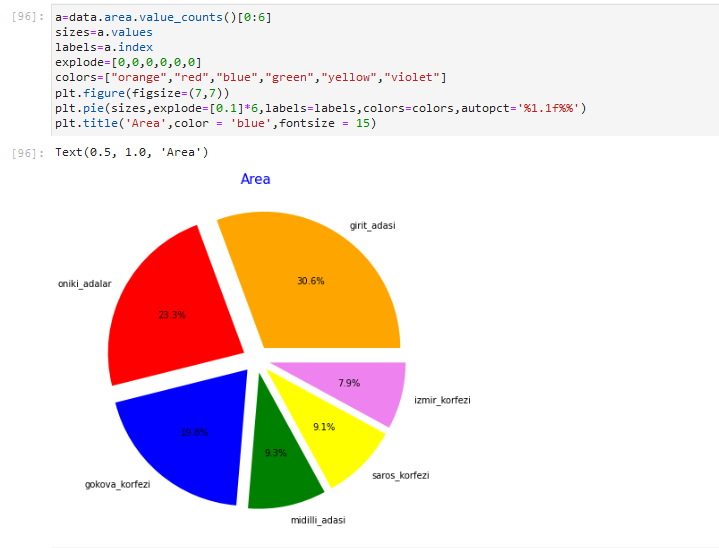


Yukarıda da göreceğiniz gibi, city.value\_count komutu ile şehir bazlı sınıflandırma yapılmıştır. Çıktısı ise aşağıdaki gibidir.

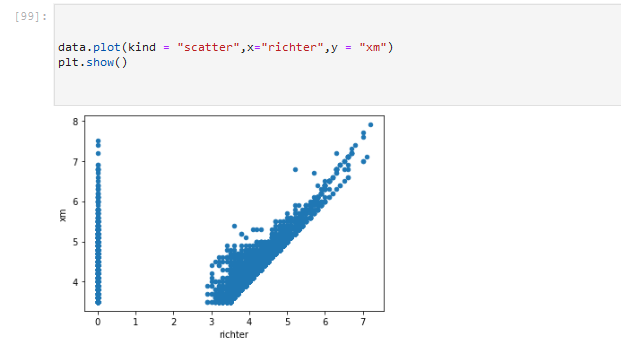


Aşağıda ise, pasta dilimi üzerinde şehir bazlı olarak % sel ne kadar deprem olduğu listelenmek istenmiştir. Toplamda 10 şehir baz alınmıştır.

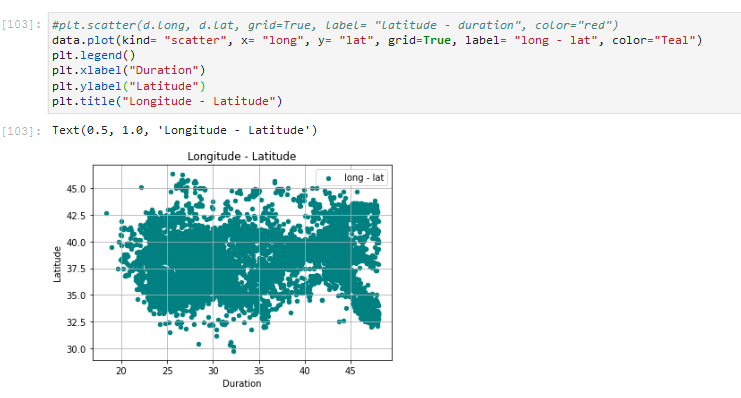


Area.value\_counts komutu ile ise, en çok deprem olan 6 bölge grafikte gösterilmiştir. Bölge adı ve % sel değeri görünmektedir. 

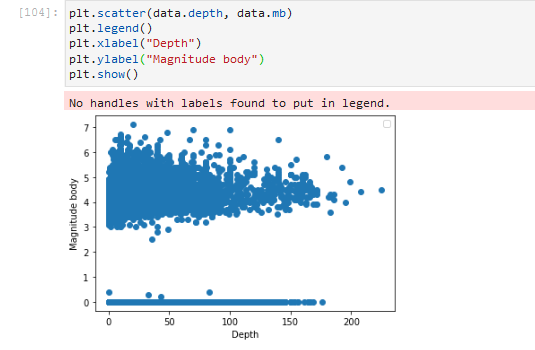
Ritcher ölçeğine göre en büyük deprem büyüklükleri listelenmiştir.



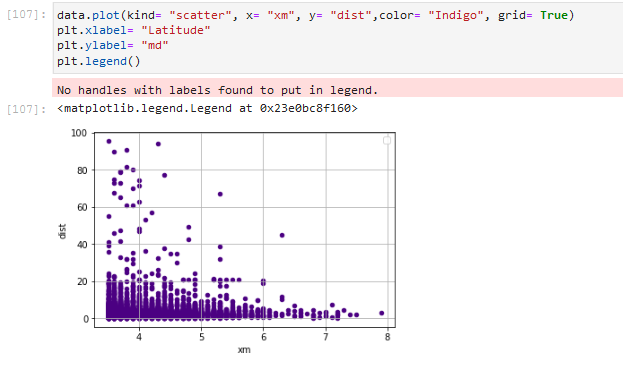
Enlem ve boylam üzerinde meydana gelen depremler yoğunluk haritası şeklinde gösterilmiştir.



Derinlik ve düzlem dalgası, grafik üzerinde yoğunluk haritası olarak gösterilmiştir.



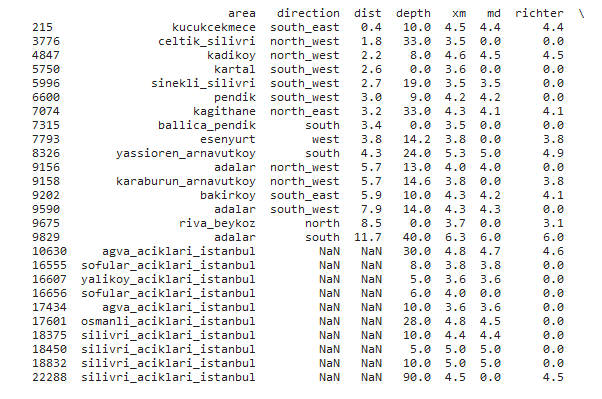
Diğer tüm büyüklüklerin dışındaki yaşanan en büyük büyüklük ise deprem süresi ve enlem şeklinde grafikte gösterilmiştir.

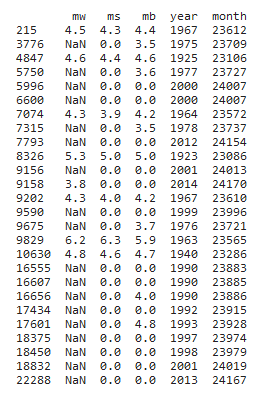


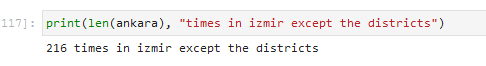
Şehir bazlı tekel listelemeler de yapılmaktadır. Bunlara örnek olarak sadece İstanbul için meydana gelen deprem verilerini listeleyebiliriz.



Yalnız veriler üzerinde işlem yapılmadığı için draft veriler olarak gelecektir.





Örnek; İzmirde kaç adet deprem olduğunu ise;  


Olarak gösterebiliriz.