

# Veri Bilimi İçin Açık Kaynak Veri Setleri

## 1- <http://Kaggle.com/datasets>

Kaggle binlerce veri seti bulundurmaktadır. Kaggle, veri bilimi hedeflerinize ulaşmanıza yardımcı olacak güçlü araçlara ve kaynaklara sahip dünyanın en büyük veri bilimi topluluğudur. Kaggle, kullanıcıların veri kümelerini bulup yayınlamasına, web tabanlı bir veri bilimi ortamında modeller keşfetmesine ve oluşturmaya, diğer veri bilimcileri ve makine öğrenimi mühendisleriyle birlikte çalışmasına ve veri bilimi zorluklarını çözmek için yarışmalara katılmasına olanak tanır. İsterseniz indirip localinizde kullanabilir ister kaggle notebook üzerinde çalışabilirsiniz. Kaggle yarışmalardaki profesyonel çalışmaları inceliyor fikir edinebilirsiniz.

## 2- <https://archive.ics.uci.edu/ml/datasets.php>

Makine öğrenimi topluluğu tarafından makine öğrenimi algoritmalarının deneysel analizi için kullanılan veri tabanları ve veri oluşturuculardan oluşan bir koleksiyondur.

## 3- <http://data.ibb.gov.tr>

İstanbul Büyükşehir Belediyesinin data alanında çalışanlara destek amaçlı veri sunan platformdur. Ekonomi, Afet Yönetimi, Enerji, Yaşam, Mobilite gibi kategoriler altında veriler bulunmaktadır.

#### 4- <http://data.tuik.gov.tr>

Türkiye'ye dair istatistiksel verilerinin paylaşıldığı sitedir. İstatistiksel tablolar, görseller ve çeşitli konulara ulaşabilirsiniz.

#### 5- <http://data.gov>

Data.gov, Mayıs 2009'un sonlarında kullanıma sunulan bir ABD hükümeti web sitesidir. Data.gov, Federal Hükümetin Yürütme Birimi tarafından oluşturulan yüksek değerli, makine tarafından okunabilir veri kümelerine kamu erişimini iyileştirmeyi amaçlamaktadır. Site, halkın kullanımına sunulan federal, eyalet, yerel ve kabile hükümeti bilgilerini içeren veri setlerinin bulunduğu bir havuzdur. Sitede çeşitli alanlarda ve formatlarda 200 bine yakın veri seti bulunuyor. Farklı alanlarda gerçek verilere ulaşmak için güzel bir kaynak.

#### 6- <http://data.gov.in>

Hindistan hükümetinin açık veri girişimini destekleyen bir platformudur. Hindistan hükümetinin India.gov.in ve ABD hükümetinin data.gov projesinin en iyi özelliklerini birleştirir ve genişletir. 400 binden fazla kaynak bulundurmaktadır ve kaynaklar 1 veya daha fazla veri seti içermektedir. Üstelik API ve görsel erişim imkanı sağlamaktadır. Merak ettiğiniz veri setleri için öneride bulunabilmektesiniz.

**7- <http://registry.opendata.aws>**

AWS kullanıcısı iseniz kullanışlı büyük veri setleri ulaşabilirsiniz. AWS hizmetleri aracılığıyla herkese açık hale getirilen veri kümelerini bulmayı kolaylaştırır. Üstelik kullanım örnekleri de bulunmaktadır.

**8- <http://cloud.google.com/bigquery/public-data>**

Google BigQuery'yi atlamak olmaz. AWS gibi Google BigQuery de farklı alanlarda büyük veriler sağlamaktadır.

**9- <http://data.worldbank.org>**

Dünya Bankası veri paylaşım sitesidir. Çeşitli konularda zaman serisi verilerinin koleksiyonlarını içeren bir analiz ve görselleştirme araçları sunar. Ekonomi ile ilgili çalışmalar yapmak istiyorsanız kesinlikle göz atmanız gereken bir kaynak. Üstelik zaman serileri verisi ihtiyacınızı da karşılayabilir.

**10- <http://FiftyEight.com>**

Kamuoyu yoklaması analizi, politika, bilim, ekonomi ve spor bloglarına odaklanan Amerikan web sitesidir. Çeşitli çalışmalarda/projelerde kullanmak için kullanışlı veri setleri sağlamaktadır. Verilerine <http://github.com/fivethirtyeight> üzerinden erişebilirsiniz

**11- <http://research.google.com/youtube8m>**

Youtube'un düzenli güncellediği verilerinin bulunduğu yerdir. Kendi değişleriyle "YouTube'da harika video içeriklerini bulmayı ve keşfetmeyi kolaylaştırmak için büyük ölçekte bilgisayarla görme ve video anlama sistemleri oluşturmak için çalışıyoruz". Sizlerde kendi makine öğrenmelerinizde bu verileri kullanabilirsiniz.

**12- <http://reddit.com/r/datasets>**

Reddit üzerindeki veri paylaşım platformudur. Discussion bölümünde diğer kullanıcılarla rahatlıkla iletişime geçebilir ve yardımlaşabilirsiniz.

**13- <http://wunderground.com>**

Weather Underground, İnternet üzerinden gerçek zamanlı hava durumu bilgileri sağlayan bir hava durumu hizmetidir. API desteğiyle verileri direkt olarak projenize çekebilir ya da csv olarak geçmişini ya da tahmini verilerini indirerek kullanabilirsiniz.

**14- <https://datahub.io/collections>**

Birçok veri analistinin amacı, akıllıca iş kararları alınmasına yardımcı olmaktır. Bu nedenle, portföy projeniz için ekonomik veya ticari veri kümelerini kullanmak dikkate değer olabilir. Datahub, iklim değişikliğinden eğlenceye kadar birçok konuyu kapsarken, ağırlıklı

olarak borsa verileri, emlak fiyatları, enflasyon ve lojistik gibi alanlara odaklanıyor. Portaldaki verilerin birçoğu aylık olarak (hatta günlük olarak) güncellendiğinden, geniş zaman ölçeklerini kapsayan verilerin yanı sıra her zaman üzerinde çalışacak yeni bir şeye sahip olacaksınız.

**15-**<https://datasetsearch.research.google.com/>

Görünüşe göre bugünlerde her şey için Google'a dönüyoruz ve veriler de bir istisna değil. 2018'de piyasaya sürülen Google Veri Kümesi Araması, Google'ın standart arama motoru gibidir, ancak kesinlikle veri içindir. Göz atmayı tercih ediyorsanız en iyi araç olmasa da, aklınızda belirli bir konu veya anahtar kelime varsa, sizi hayal kırıklığına uğratmayacaktır.

**16-)**[www.earthdata.nasa.gov](http://www.earthdata.nasa.gov)

1994'ten beri halka açık olan bu havuz, küçük mavi gezegenimiz için NASA'nın tüm uydu gözlem verilerine erişim sağlar. Tahmin edebileceğiniz gibi, hava ve iklim ölçümlerinden atmosferik gözlemlere, okyanus sıcaklıklarına, bitki örtüsü haritalamasına ve daha fazlasına kadar incelenecek çok şey var. Dünya tabanlı veriler size göre değilse, NASA'nın Gezegensel Veri Sistemi, (2004'ten 2017'ye kadar Satürn'ün yörüngesinde dönen) Cassini sondası gibi gezegenler arası görevlerden elde edilen verilerle işleri bir adım öteye taşıyor. Kim bilir, bilimsel bir keşif bile yapabilirsiniz

**17-<http://opendata.cern.ch/>**

Son derece karmaşık veri kümeleriyle çalışma yeteneğinizi göstermek ister misiniz? CERN Açık Veri Portalına gidin. Büyük Hadron Çarpıştırıcısı parçacık hızlandırıcısından alınan veri kümeleri de dahil olmak üzere iki petabayttan fazla bilgiye erişim sunar.

**18-<https://apps.who.int/gho/data/node.home>**

Global Health Observation veri deposu, BM DSÖ'nün dünyanın dört bir yanından sağlıkla ilgili istatistiklere açılan kapısıdır. Sağlık sektörüne girmek istiyorsanız (özellikle makine öğrenimi alanında birçok veri bilimcisi için temel odak noktası), bu veri kümeleri portföyünüz için iyi bir seçenektir. Sıtmadan HIV/AIDS'e, antimikrobiyal direnç ve aşılanma oranlarına kadar her şeyi kapsayan portalın, veri tablolarını indirmeden önce önizlemenizi sağlayan hoş bir küçük özelliği bile var.

**19- <https://www.bfi.org.uk/industry-data-insights>**

Birleşik Krallık gişe rakamlarından seyirci demografisine, ev eğlencesine, film yapım maliyetlerine ve daha fazlasına kadar her konuda veri toplar ve yayınlar. Yine de en iyi bölüm, yıllık istatistiksel yıllıkları. Bu, bazı mükemmel istatistiksel analizler ve görsel raporlarla yılın verilerini ayrıştırır; veri analitiğinde yeniyseniz ve çalışmanızı gerçeğe karşılaştırmak istiyorsanız harikadır.

**20- <https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>**

2009'dan beri NYC Taksi ve Limuzin Komisyonu, New York şehrinin dört bir yanından ulaşım verilerini topluyor. Alış/bırakma saatleri ve konumları, yolculuk mesafeleri, ücretler, oran ve ödeme türleri, yolcu sayıları ve daha fazlasını kapsayan veri kümelerini bulun. Site ayrıca kullanıcı kılavuzları, taksi bölgesi haritaları, veri sözlükleri (elektronik tablo etiketlerini açıklamak için) ve yıllık endüstri raporları dahil olmak üzere bazı ek araçlar sağlar.

Kaynakça:

<https://careerfoundry.com/en/blog/data-analytics/where-to-find-free-datasets/>

[Linkedin Açık Veri Kaynakları](#)

