COVİD-19 VERİ ANALİZİ VE GÖRSELLEŞTİRİLMESİ

Onur KAPLAN

*Özet---Büyük veri analizine giriş dersinde şimdiye kadar öğrendiğimiz veri işleme ve görselletirme tekniklerini 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE veri setini kullanarak Covid-19’un dünya üzerindeki etkilerini analiz edip görselleştirdim.*

# GİRİŞ

# 2019 yılının Aralık ayında, Çin’de birdenbire sebebi açıklanamayan zatürre olguları ortaya çıkmaya başladı. Yapılan araştırmalar, bu zatürre vakalarının daha önceden tanımlanmamış yeni bir tip coronavirüs (coronavirus) olduğunu ortaya çıkardı. Virüsün bu formuna, 2019 yılında ortaya çıktığı için, Coronavirus 2019 yani COVID-19 denildi. Coronavirüs aslında, hayvanlarda bolca rastlanan bir virüstür. Son zamanlarda hastalığa neden olan virüsün kaynağının, ise Çin'in Wuhan kentinde bulunan Huanan deniz ürünleri pazarı olduğu düşünülmektedir. Önce hayvandan insana bulaşan virüsün insandan insana da yayılabildiği zamanla anlaşılmıştır.

Gündeme son salgın ile birlikte gelse de esasında coronavirüsler, farklı tipleri olan geniş bir virüs gurubudur. Genetik materyal olarak RNA taşır, hatta RNA taşıyan virüsler arasında en büyük genoma sahip olan virüs gurubudur. Coronavirüs, zarflı bir virüstür ve zarfının etrafından dışarıya doğru diken (spike) şeklinde çıkıntılara sahiptir. Bu yapı, ona elektron mikroskobu altında bakıldığında kraliyet tacı şeklinde bir görüntü verir. Bu nedenle de, virüse Latince kraliyet tacı anlamına gelen Corona ismi verilmiştir.

Coronavirüslerin farklı tipleri, farklı rahatsızlıklara neden olur. Örneğin bazı coronavirüsler, sindirim sistemi (gastrointestinal) rahatsızlıklarına neden olurken, bazıları ise solunum sistemi ile ilgili şikayetleri beraberinde getirir. Solunum rahatsızlıklarına neden olan coronavirüsler de kendi aralarında farklılık gösterir. Bazı enfeksiyonlarda sadece nezle kadar hafif semptomlar gözlenirken, bazı kişilerde zatürreye varan sonuçlar ortaya çıkar. Bununla beraber, coronavirüsler, tarihteki üç örneği dışında genelde hafif semptomlar gösterir. Örneklerden ilki yine Çin’de 2003 senesinde gözlenen İngilizce 'Severe Acute Respiratory Syndrome - Ciddi Akut Solunum Sendromu' kelimelerinin kısaltması olan SARS coronavirüsüdür. 2012 senesinde ise, bu sefer Suudi Arabistan’da 'Middle East Respiratory Syndrome - Orta Doğu Solunum Sendromu' olarak adlandırılan MERS coronavirüsü ortaya çıkmıştır. Son olarak da, şu anda dünya üzerinde hızla yayılmakta olan 2019 n-CoV (2019 Novel Coronavirus, 2019 Yeni Coronavirüsü) ortaya çıkmıştır. Bütün bu örneklenen coronavirüsleri ciddi semptomlara yol açar, hatta ölümle sonuçlanabilir.[1]

Dünyada büyük çapta etki yaratan covid-19 bir çok ülkede binlerce ölüme neden oldu ve hala insandan insana hızlı bir şekilde yayılmaktadır. Hastalığın dünya üzerindeki etkisini takip eden bazı veri setleri bize günlük olarak dünyadaki bütün ülkelerin vatandaşlarından kaç tanesinin hastalığa yakalandığı, öldüğü ya da iyileştiği bilgisini veriyor.

Çalışmadaki asıl amacım insanların merak ettikleri şeylere kolayca erişebilmesi için bu günlük verileri basit bir ara yüzle görselleştirmek ve bazı analizler yapıp hastalığın ülke ülke gidişatını gözlemlemektir.

Çalışmamı Excel, Tableau ve Jupiter Notebook kullanarak gerçekleştirdim.

Tableau Stanford Üniversitesi Bilgisayar Bilimleri bölümünde Profesör Pat Hanrahan ve veri tabanları ve veri küplerinin analizini ve görselleştirme teknikleri konusunda uzman ve çok boyutlu ilişkisel veritabanının tablo tabanlı gösterimi çalışmalarının öncüsü Chris Stolte tarafından 1997-2002 yılları arasında geliştirilmiştir. Birlikte VizQL denilen veritabanı görselleştirme ve grafik işleme dili icat edip bunu bir SQL ile birleştirdiler. VizQL, Polaris sisteminin çekirdiğini oluşturmuştur, çok boyutlu ve büyük veritabanlarını keşfetmek için bir arayüzdür. 2003'te Stolte eski iş arkadaşı Christian Chabot'ı CEO olarak işe aldıktan sonra Tableau Stanforddan kullanılabilir bir uygulama olarak ayrıldı. Ürün ilişkisel veritabalarını, küpleri, bulut veritabanlarını ve tabloları sorgular ve değişik grafik seçenekleriyle birlikte görsel yaparak internet veya kurumsal ağda paylaşılmasına olanak sağlar.[2]

Microsoft Excel, Microsoft tarafından Microsoft Windows ve Apple Macintosh işletim sistemleri tabanında çalışmak üzere yazılan ve dağıtımı yapılan bir tablolama programıdır (spreadsheet). İçinde bulunan detaylı finansal çözümlerin yapılabildiği tablolama, grafik oluşturma başarısı ve uygulamalarda kullanılabilecek Visual Basic makro programlama dili sayesinde kendi türünde şu anda dünyadaki en popüler yazılımdır.[3]

Jupyter Notebook, çeşitli programlama dilleri için etkileşimli bir ortam sağlayan açık kaynak kodlu bir programdır.[4]

# YÖNTEM

## Veri Seçimi

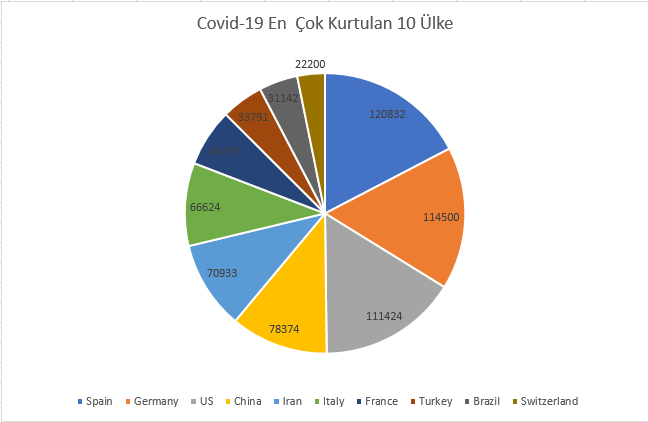
Çalışmada 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE veri seti içinden csse\_covid\_19\_time\_series bölümündeki time\_series\_covid19\_confirmed\_global.csv, time\_series\_covid19\_deaths\_global.csv, time\_series\_covid19\_recovered\_global.csv verilerini kullanmaya karar verdim. Bu verileri seçmemdeki amaç yapmak istediğim çalışmaya en uygun olan verilerdi ve en güncel veri setiydi. Veri setinde sırayla bütün ülkelerin gün gün onaylanan, ölen ve iyileşen hastalarının verileri vardı.

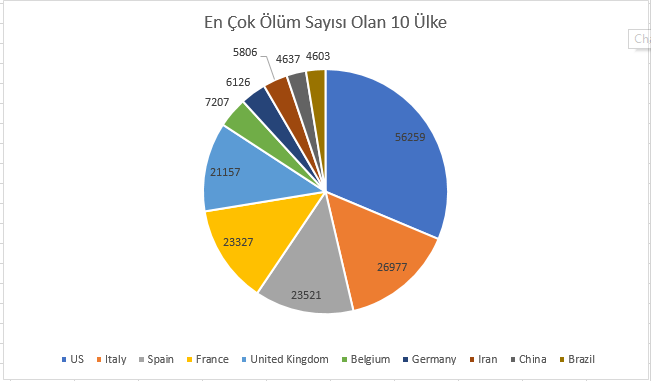
## Veri Önişleme ve Görselleştirme

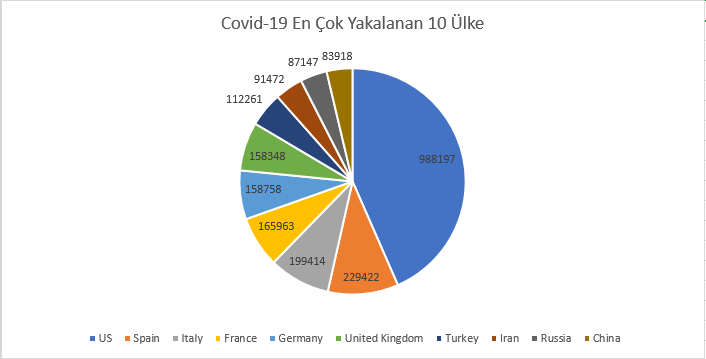
İlk aktivitem için ayrı ayrı csv dosyalarında olan verileri sütun sütun ayırıp tek bir excel dosyasında yan yana ekleyerek birleştirdim ve tableauda görselleştirdim. İkinci aktivitem için tarihleri tek bir sütun altında birleştirdiğim ve diğer sütunlarda da ülke bilgisi, onaylananların bilgisi, ölenlerin bilgisi, iyileşenlerin bilgisi olduğu yeni bir formatta excel dosyayı oluşturdum ve tableau ile görselleştirdim. Üçüncü aktivitem için ikinci aktivitede oluşturmuş olduğum excele ülkelerin her gün yüzde kaç oranda arttığı ve genel olarak günde yüzde kaç arttığı bilgisini ekledim ve =IFERROR((C2-C1)/C1\*100;0), =AVERAGEIF(B:B;J2;F:F) gibi bazı formüller kullandım ve tableau ile görselleştirdim. Dördüncü aktivitem için ikinci aktivitede kullanmış olduğum veriyi tekrardan tek bir sütun haline getirmek için =A:A&","&B:B&","&C:C&","&D:D&","&E:E formülünü kullandım ve Jupyter Notebookta analiz ettim.Beşinci aktivitemde ise time\_series\_covid19\_confirmed\_global.csv, time\_series\_covid19\_deaths\_global.csv, time\_series\_covid19\_recovered\_global.csv verilerini sütunlara ayırıp pivot table atıp pie charts ile bazı görselleştirmeler yaptım.

# ÖRNEK SONUÇLAR

## EXCEL



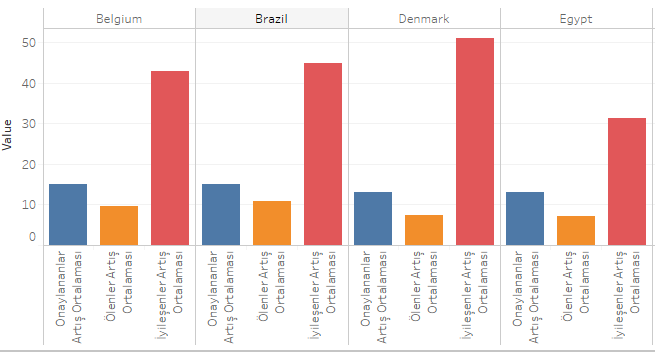




## TABLEAU

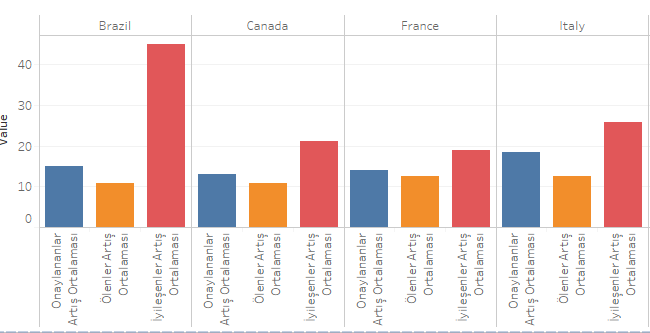
İyileşme Yüzdesine Göre Sıralanmış Verilerin Bir Kısmı

Aktivite 3



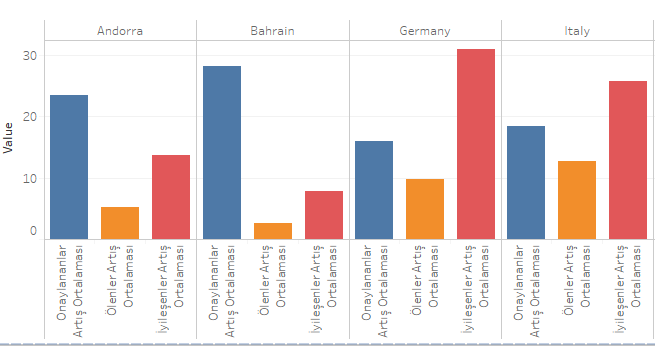
Ölüm Yüzdesine Göre Sıralanmış Verilerin Bir Kısmı

Aktivite 3



Yakalanma Yüzdesine Göre Sıralanmış Verilerin Bir Kısmı

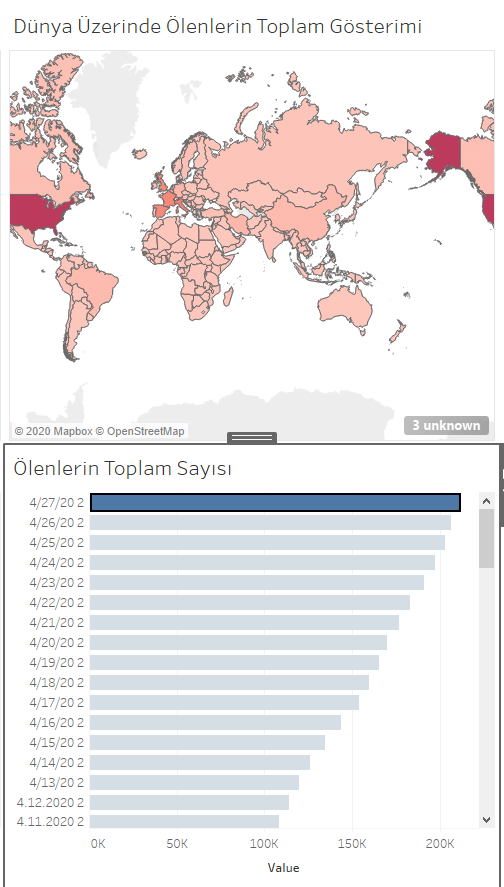
Aktivite 3



Aktivite 2

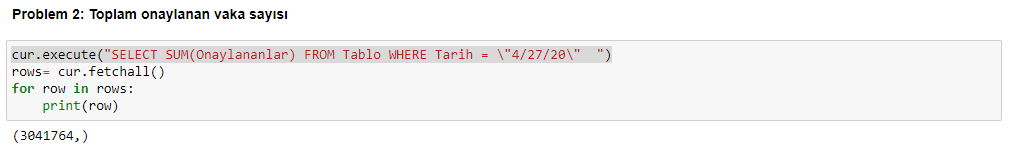
##### 

Aktivite 1



## JUPYTER NOTEBOOK





# KAYNAKÇALAR

1. https://www.medicalpark.com.tr/coronavirus/hg-2287
2. https://tr.wikipedia.org/wiki/Tableau
3. https://tr.wikipedia.org/wiki/Microsoft\_Excel
4. https://medium.com/datarunner/jupyter-notebook-nedir-502c024c0076
5. <https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series>

**.**