

# KOVİDOR (KORONAVİRÜS PANDEMİ KORİDORU) PROJE RAPORU

Yeni vaka ve toplam vaka sayıları arasındaki ilişkiyi inceleyerek ülkenin pandemi sürecindeki yerini anlamaya yönelik bir çalışma

## Amaç

Bu projenin amacı; başta Türkiye ve COVID19 olmak üzere, bir ülkenin pandemideki durumunu dünya geneli verilerde görülen ortak trendle karşılaştırarak tespit etmek ve bu durumu herkes tarafından kolay anlaşılabilir halde ortaya koymaktır. Pek çok kişinin her gün takip ettiği:

- Günlük vaka sayısı
- Toplam vaka sayısı
- Gün / Vaka grafiği
- Lojistik regresyon tahminleri

gibi bilgiler tek başına olduklarında pandeminin ülkedeki ilerleyişinin anlaşılmasına pek katkı sağlamıyor.

### Örneğin:

"Bugün Güney Korede 242 yeni COVID19 vakası tespit edildi"

Haberlerde bu cümleyi görseniz size ne ifade ederdi? 242 kişinin sağlığından olması tabii ki kötü bir durum, ama sadece bu cümleye bakarak Güney Kore'nin pandemi sürecinde nerde olduğu hakkında fikir edinmek pek mümkün değil.

Bu proje sonucunda tek bir cümleyle herhangi bir ülkenin pandemi sürecinde hangi noktada olduğunu ifade edebilmeyi amaçlıyoruz.

## Varsayımlar

- Dünya sağlık örgütü ülkelerin raporladığı vaka sayılarını online olarak paylaşıyor ve bu veriler devamlı güncelleniyor.
- Ülkelerin Dünya Sağlık Örgütüne raporladığı veriler hükümetler tarafından çarpıtılmış, günlere yayılmış, eksik hesaplanmış olabilir.
- Hükümetler tarafından kasten çarpıtılmış ve günlere yayılmış veriler muhtemelen takip eden günlerde ortaya çıkacaktır. Bu tür çarpıtılmış veriler gün yüzüne çıkmasa bile ülkelerin pandemi süreçlerindeki benzerlikleri görmemize engel olmayacaktır.
- Yapılan test sayısı tespit edilen yeni vaka sayısının artmasına çarpan etkisi yapacaktır. Fakat yapılan test sayılarındaki değişim de pandeminin ilerleyişindeki süreçlerden biri olduğu için vaka verilerini test sayısı verilerine göre normalize etmek sürece geniş bir pencereden bakmamıza engel olacaktır.
- COVID19'un *inkübasyon süresi*<sup>1</sup> ortalama 5 gündür.

(**Kaynak:** <https://annals.org/aim/fullarticle/2762808/incubation-period-coronavirus-disease-2019-covid-19-from-publicly-reported>)

6. Enfeksiyon hastalıkları uzmanı olmadan pandeminin karakteristiklerini tümüyle anlamak mümkün değildir. Buna rağmen veriden yola çıkarak yapılan analizler ülkenin pandemi sürecinin anlaşılmasını kolaylaştıracaktır.

*inkübasyon süresi*<sup>1</sup>: Virüsün vücuda girişiyle hastalık semptomlarının meydana çıkışı arasındaki süre.

## Veri Analizi

Verinin indirilmesi, işlenmesi ve görselleştirilmesinde faydalı olacak `tidyverse` ve `plotly` kütüphanelerini yüklüyoruz

```
#load_packages("tidyverse"); load_packages("plotly")
library(tidyverse); library(plotly)
```

### Hazır Veri Seti: Ülkeler ve Toplam Vaka Sayıları

İnternette hazır bulunan verisetlerinden sadece birini kullanacağız. Bu veri seti üzerinde yapacağımız hesaplamalarla kendi özgün veri setimizi oluşturacağız. İndireceğimiz bu veri seti Dünya Sağlık Örgütü açıklamaları doğrultusunda her gün güncellenen **ülkelerin tespit edilen toplam Covid19 vakası sayılarını** içeriyor.

(Bu projede başka hazır veri seti kullanılmamasının sebebi vaka sayısı üzerine bir analiz yapılıyor olması ve buna ilişkin tek veri kaynağının Dünya Sağlık Örgütü olmasıdır.)

```
conf_cases <- read_csv("https://covid.ourworldindata.org/data/ecdc/total_cases.csv")
```

Veri seti hakkında bilgi edinmek için satır/sütun sayılarını ve ilk 10 sütunun ismini bastıralım:

```
paste(c("Bu veri setinde", ncol(conf_cases), "sütun ve", nrow(conf_cases), "satır var. İlk 10 sütunun isimleri:",
as.character(colnames(conf_cases)[1:10])), collapse = " ")
```

'Bu veri setinde 208 sütun ve 116 satır var. İlk 10 sütunun isimleri: date World Afghanistan Albania Algeria Andorra Angola Anguilla Antigua and Barbuda Argentina'

10 satır ve 10 sütununu bastırarak verinin formatına bakalım:

```
conf_cases[78:88, 1:10]
```

	date	World	Afghanistan	Albania	Algeria	Andorra	Angola	Anguilla	Antigua and Barbuda	Argentina
2020-03-17	180094	21	51	58	14	NA	NA	NA	NA	65
2020-03-18	194836	22	55	60	14	NA	NA	NA	NA	79
2020-03-19	213149	22	59	72	53	NA	NA	NA	NA	97
2020-03-20	242372	22	70	90	75	NA	NA	NA	NA	128
2020-03-21	271115	24	70	102	75	NA	NA	1	158	
2020-03-22	305234	24	76	139	88	2	NA	1	225	
2020-03-23	338232	34	89	201	113	2	NA	1	266	
2020-03-24	377959	40	100	230	133	2	NA	1	301	
2020-03-25	416878	42	123	264	164	2	NA	3	387	
2020-03-26	468087	75	146	302	188	2	NA	3	502	
2020-03-27	527829	75	174	367	224	3	2	7	589	

Bu veri setinde:

- Veriler tarihlerle eşleştirilmiş fakat analizimizde verinin hangi ayın kaçınıcı gününe ait olmadığı önemli olmayacağı için bir ülkeye ait verileri alırken verilerin sıralı olması için tarih yerine gün numarası sütunu ekleyeceğiz
- Ülkelerin toplam vaka sayısını raporlamadığı günler `NA` olarak belirtilmiş. Bir ülkeye ait verileri alırken `NA` satırlarını sileceğiz.
- Her satırda sadece o güne ait toplam vaka sayıları verilmiş. Ülkelerin günlük yeni vaka sayılarının bulunduğu sütunları oluşturmak için toplam vaka verileri arasındaki farkı hesaplayacağız.

## Özgün Veri Seti: Gün numaraları,Toplam & Günlük Vaka

Her farklı ülkelere ait veriyi ulaşmak istediğimizde uzun işlemleri tekrar etmemek için `daily_df()` fonksiyonunu oluşturuyoruz. Bu fonksiyon:

- Ülke adı ve toplam vaka verisi olmak üzere 2 argüman alacak.
- Çıktı olarak *Gün*, *Toplam Vaka* ve *Günlük Vaka* sütunlarından oluşan bir tablo verecek.
- Toplam vaka sayısının sıfırdan farklı tüm günler bu tabloya dahil olacak.

```
daily_df <- function(country = "Turkey", data = conf_cases) # Varsayılan değeri Türkiye yapıyoruz
{
  cases <- data[country] %>% drop_na() # Ülkeye ait veri bulunan satırlar
  cases <- cases[pull(cases[, 1]) > 0, 1] # İlk vaka görüldüğü günden itibaren filtreliyoruz
  cases[, 2] <- 1:nrow(cases[country]) # Tarih yerine gün numarası tercih ediyoruz
  cases <- cases[, c(2, 1)]; colnames(cases) <- c("Gün", "Toplam Vaka") # Sütunları sıralama & adlandırma
  daic <- cases[2:nrow(cases[, 2]), 2] - cases[1:length(pull(cases[, 2])) - 1, 2] # Günlük vaka hesabı
  daic <- tibble(c(pull(cases[, 2]), pull(daic)))
  colnames(daic) <- "Günlük Vaka"
  return(cases %>% cbind(daic))
}
```

Analize dahil etmek için kullanacağımız veri sayısı yüksek olan birkaç ülkenin gün numarası ve günlük vaka sayılarını içeren tablolarını kaydedelim ve örnek olarak Türkiye'ye ait verilere bakalım.

```
tr_cases <- daily_df()
ch_cases <- daily_df("China")
it_cases <- daily_df("Italy")
us_cases <- daily_df("United States")
sk_cases <- daily_df("South Korea")
au_cases <- daily_df("Australia")
```

```
tr_cases %>% head(4)
```

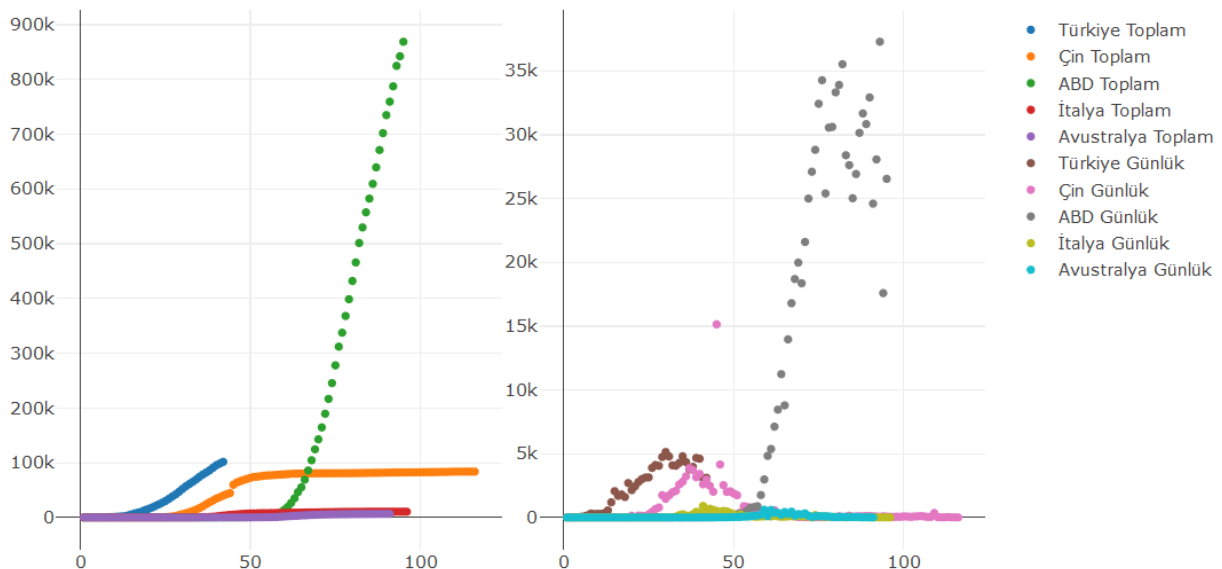
Gün	Toplam Vaka	Günlük Vaka
1	1	1
2	2	1
3	18	16
4	47	29

## Veri Setinin Görselleştirilmesi

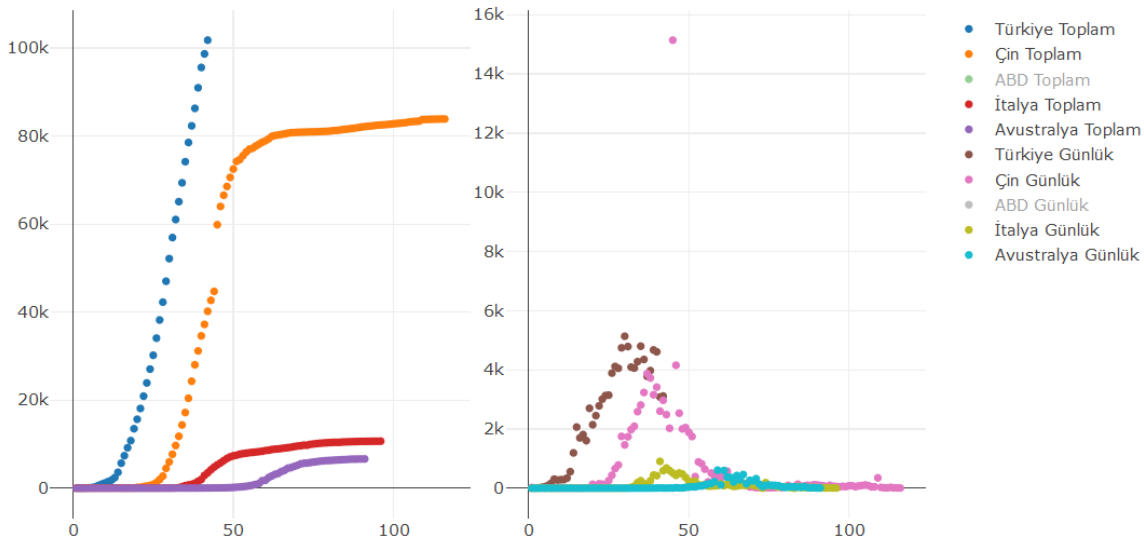
Pandeminin farklı ülkelerdeki gelişimindeki benzerlikleri anlamamız için verilerin **doğru şekilde** görselleştirilmesi çok önemli. Bu yüzden eksenlerdeki değişkenleri değiştirerek bu benzerlikleri yakalamaya çalışacağız.

Öncelikle en sık kullanılan 2 grafiğe, yani **Gün - Toplam Vaka** ve **Gün - Günlük Vaka** grafiklerine bakalım.

```
fig0_1 <- plot_ly(x = ~tr_cases[, 1], y = ~tr_cases[, 2], name = "Türkiye Toplam",
  type = "scatter", mode = "markers", height = "500px") %>%
  layout(yaxis = list(type = "linear", title = "Toplam Vaka"),
    xaxis = list(type = "linear", title = "Gün"))
fig0_1 <- fig0_1 %>% add_trace(x = ~ch_cases[, 1], y = ~ch_cases[, 2], name = "Çin Toplam") %>%
  add_trace(x = ~us_cases[, 1], y = ~us_cases[, 2], name = "ABD Toplam") %>%
  add_trace(x = ~sk_cases[, 1], y = ~sk_cases[, 2], name = "İtalya Toplam") %>%
  add_trace(x = ~au_cases[, 1], y = ~au_cases[, 2], name = "Avustralya Toplam")
fig0_2 <- plot_ly(x = ~tr_cases[, 1], y = ~tr_cases[, 3], name = "Türkiye Günlük",
  type = "scatter", mode = "markers", height = "500px") %>%
  layout(yaxis = list(type = "linear", title = "Günlük Vaka"),
    xaxis = list(type = "linear", title = "Gün"))
fig0_2 <- fig0_2 %>% add_trace(x = ~ch_cases[, 1], y = ~ch_cases[, 3], name = "Çin Günlük") %>%
  add_trace(x = ~us_cases[, 1], y = ~us_cases[, 3], name = "ABD Günlük") %>%
  add_trace(x = ~sk_cases[, 1], y = ~sk_cases[, 3], name = "İtalya Günlük") %>%
  add_trace(x = ~au_cases[, 1], y = ~au_cases[, 3], name = "Avustralya Günlük")
fig0 <- subplot(fig0_1, fig0_2)
```



ABD'nin toplam vaka grafiğinde üstel artışı çok net bir şekilde görüyor olsak da bu haliyle ülkelerin grafikleri arasında tekrar eden bir benzerlik görmek (özellikle de ABD'yi eklemiş olmamız nedeniyle) pek mümkün durmuyor. ABD'yi çıkartıp tekrar inceleyelim.



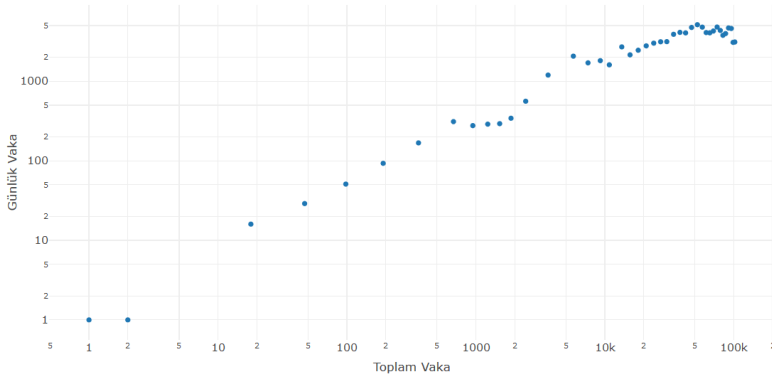
Şimdiyse Türkiye'nin grafiğinde tıpkı ABD'de olduğu gibi üstel bir artışı net bir şekilde görüyoruz. Aynı zamanda diğer ülkelerin de başlangıçta üstel görünen ve sonradan S şeklini (ve lojistik fonksiyonu) andıran grafikleri olduğunu görüyoruz. Yine de bu grafiklere bakarak pandeminin ilerleyişinde ortak bir yön tespit etmek zor gözüküyor. Bunun sebepleri:

- Ülkelerin gün numaralarına göre karşılaştırılıyor olması. Örneğin Covid19 salgınının farklı aşamalarındaki ülkelerin 10.günlerindeki vaka sayılarını karşılaştırmamız bize pek yardımcı olmuyor.
- Lineer ölçekli grafikte vaka sayısı diğerlerinden birkaç kat fazla olan ülkelerin davranışlarını diğer ülkelerle karşılaştırmanın zor olması. Örneğin ABD ve Türkiye verileri çok benzer şekillerde olsa da aynı grafikte gösterildiklerinde bu benzerlik belli olmadı. Ancak ABD verilerini çıkartıp Türkiye grafiğine baktığımızda bu benzerliğin farkına varabildik.

Bu sorunları ortadan kaldırmak için gün eksenini devre dışı bırakarak ülkelerin toplam ve günlük vaka sayılarını tek bir grafikte karşılaştıracamız. Aynı zamanda her bir eksenin logaritmik ölçekli çizerek salgının vaka sayıları birbirinden farklı olan ülkelerdeki davranışındaki benzerlikleri gözlemlemek daha kolay olacaktır. Bu yöntemle üstel artıştan lojistik fonksiyon şekline geçiş noktalarını da daha net görmeyi bekleyebiliriz.

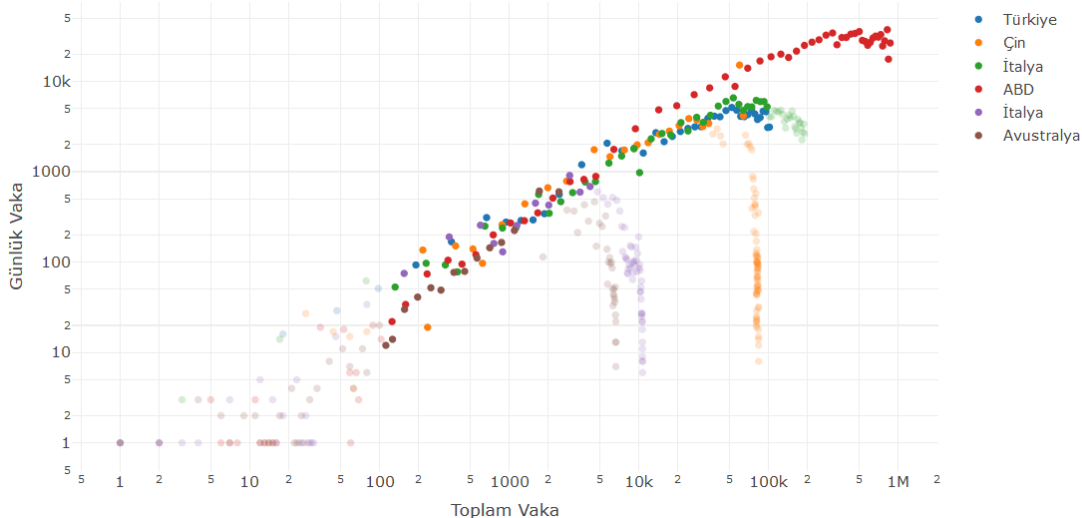
Öncelikle Türkiye'nin grafiğini inceleyelim:

```
fig1_1 <- plot_ly(x = ~tr_cases[, 2], y = ~tr_cases[, 3], name = "Türkiye",
  type = "scatter", mode = "markers", height = "500px") %>%
  layout(yaxis = list(type = "log", title = "Günlük Vaka"),
  xaxis = list(type = "log", title = "Toplam Vaka"))
embed_notebook(fig1_1)
```



Az nokta bulunmasına rağmen küçük dalgalanmalar haricinde çoğunlukla doğrusal bir ilerleyiş olduğu açıkça görülüyor. Şimdi diğer ülkeleri de ekleyerek grafiklerin benzer yönlerini inceleyelim.

```
fig1 <- fig1_1 %>% add_trace(x = ~ch_cases[, 2], y = ~ch_cases[, 3], name = "Çin") %>%
  add_trace(x = ~it_cases[, 2], y = ~it_cases[, 3], name = "İtalya") %>%
  add_trace(x = ~us_cases[, 2], y = ~us_cases[, 3], name = "ABD") %>%
  add_trace(x = ~sk_cases[, 2], y = ~sk_cases[, 3], name = "İtalya") %>%
  add_trace(x = ~au_cases[, 2], y = ~au_cases[, 3], name = "Avustralya")
```



Birçok ülkenin verilerini aynı grafikte logaritmik eksenler kullanarak görselleştirdiğimizde çok ilginç bir görüntü ortaya çıkıyor. Özellikle toplam vaka sayısı 100-1000 arasıdayken tüm ülkelerin grafikleri bir araya toplanmaya başlıyor ve adeta hepsini içine alan bir koridora girmiş gibi birlikte hareket ediyor. Bu koridordan çıkınlarsa doğrudan aşağı iniyor ve kınıldıkları bölgeyi çok net gözlemleyebiliyoruz. Bu koridorun içinde veya dışında olmak ülkelerin pandemideki ilerleyişlerinin güncel durumunu özetlemek için iyi bir ölçüt olabilir.

Günlük dalgalanmaların etkisini ortadan kaldırmak için Y ekseninde 1 günlük yeni vaka sayısı yerine son birkaç günde tespit edilen yeni vaka sayısı kullanmak mantıklı olacaktır. Bu noktada kritik olansa kaç gün toplamının kullanılacağı. Son 1 günde yeni vaka sayısını aldığımızda sapmalar fazla olduğu gibi, bunu son 14 gün toplamını kullanmak da toplam vaka sayısı ile yeni vaka sayısı arasındaki korelasyonu olduğundan çok fazla gösterecektir.

Bu yüzden Y eksenini **son 5 günde tespit edilen yeni vaka sayısı** olarak değiştirmek (5. varsayımımız olan ortalama inkübasyon süresi de göz önünde bulundurulduğunda) uygun olacaktır.

## Özgün Veri Seti: Son 5 Günlük Yeni Vaka

Ülkelerin son 5 günlük yeni vaka toplamalarını elde etmek için `fived_df()` fonksiyonunu oluşturuyoruz. Bu fonksiyon:

- İki argüman alacak: Ülke ve veri seti. Varsayılan değerleri *"Türkiye"* ve hazır aldığımız veri setini içeren **conf\_cases** olacak.
- Argümanlarını `daily_df()` fonksiyonuna verecek ve `daily_df()` çıktısı üzerinden 5 günlük vaka verilerini hesaplayacak.
- Dört günlük veya daha az veri içeren ülkeler için *"Hata: En az 5 veri gerekli"* çıktısını verecek.
- En az beş günlük veri içeren ülkeler için 4.sütunu *"5 Günlük Vaka"* olan bir tablo verecek.

```
fived_df <- function(country = "Turkey", data = conf_cases)
{
  df <- daily_df(country, data)
  if (nrow(df) < 5) return("Hata: En az 5 veri gerekli")
  fid <- replicate(nrow(df), NA)
  i <- 1
  while (i <= nrow(df) - 4)
  {
    rg <- i+4
    fid[rg] <- sum(df[i:rg, 3])
    i <- i + 1
  }
  df <- cbind(df, fid)
  colnames(df)[4] <- "5 Günlük Vaka"
  return(df)
}
```

Türkiye'nin 5 günlük vaka verisini içeren tabloya bakalım:

```
ch_fived <- fived_df("China")
it_fived <- fived_df("Italy")
us_fived <- fived_df("United States")
sk_fived <- fived_df("South Korea")
au_fived <- fived_df("Australia")
tr_fived <- fived_df()

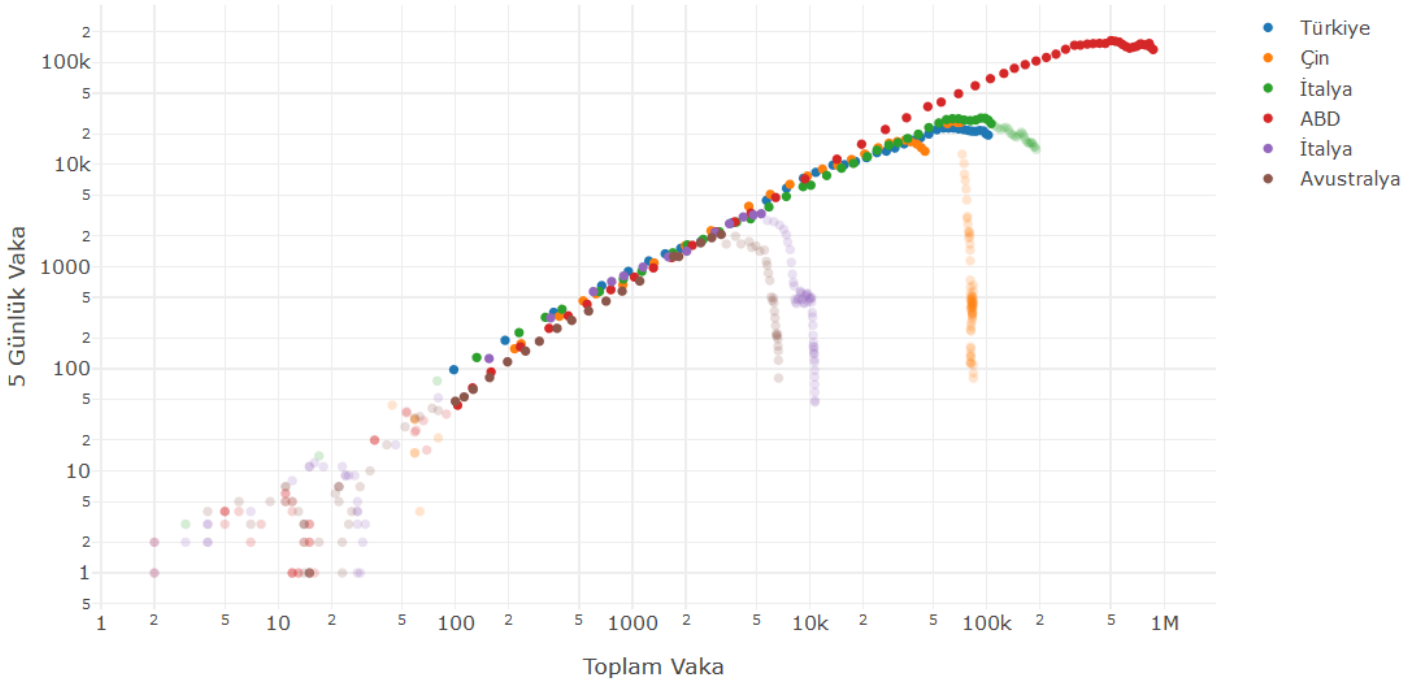
tr_fived %>% head(7)
```

Gün	Toplam Vaka	Günlük Vaka	5 Günlük Vaka
1	1	1	NA
2	2	1	NA
3	18	16	NA
4	47	29	NA
5	98	51	98
6	191	93	190
7	359	168	357

```
fig2 <- plot_ly(x = ~tr_fived[, 2], y = ~tr_fived[, 4], name = "Türkiye",
  type = "scatter", mode = "markers", height = "500px") %>%

  layout(yaxis = list(type = "log", title = "5 Günlük Vaka"),
    xaxis = list(type = "log", title = "Toplam Vaka"))

fig3 <- fig2 %>% add_trace(x = ~ch_fived[, 2], y = ~ch_fived[, 4], name = "Çin") %>%
  add_trace(x = ~it_fived[, 2], y = ~it_fived[, 4], name = "İtalya") %>%
  add_trace(x = ~us_fived[, 2], y = ~us_fived[, 4], name = "ABD") %>%
  add_trace(x = ~sk_fived[, 2], y = ~sk_fived[, 4], name = "İtalya") %>%
  add_trace(x = ~au_fived[, 2], y = ~au_fived[, 4], name = "Avustralya")
```



Görselleştirmedeki son düzenlemelerden sonra pandemi koridoru grafik üzerinde çok rahat farkedilebilir hale geldi. Şimdi "Koridordan çıkmayı" matematiksel olarak tanımlamamız ve sadece *Toplam Vaka* ve *5 Günlük Vaka* verilerinden yola çıkarak bir ülkenin koridor dışına çıkıp çıkmadığını belirleyecek bir fonksiyon oluşturmamız gerekiyor.

Bunu yapmadan önce ülkelerin pandemideki durumunu açıklamak için kullanılan / kullanılabilecek diğer yöntemleri ele alıp koridor modelinden farklılaşan yönlerini açıklayalım.



## Diğer Modeller & Yaklaşımlar

- **"Eğrinin düzleşmesi"**:: Bu ölçüt son aylarda adını en çok duyduğumuz model/yaklaşım olabilir. Temel olarak x ekseninde zaman y ekseninde aktif vaka sayısı verilerek, eğrinin y ekseninde işaretlenen yoğun bakım kapasitesinin üstüne geçmemesi için "düzleşmesinin" amaçlandığı bir gösterebilir. Bu yaklaşım daha çok tedavi olanaklarının yeterliliği üzerine bilgi vermektedir ve virüsün yayılma süreci veya ülkelerin bu süreçteki benzerlikleri hakkında pek bilgi vermez. Kovidor modelinin bir alternatifi değil tamamlayıcıdır.
- **Lojistik regresyon modelleri**: Bu modeller  $x = \text{zaman}$ ,  $y = \text{toplam vaka}$  eksenleriyle oluşturulan grafikte lojistik regresyona benzer bir şekil oluşacağı varsayımıyla büküm noktasını (2.türevin 0 olduğu nokta) tespit edip vaka sayısının kaç gün sonra sabitleneceğini tahmin etmeyi amaçlamaktadır. Geç raporlama, vaka sayılarının günlere dağıtılması, eksik bildirimler gibi verideki düzensizliklerden çok fazla etkilendikleri için bu modellerin başarı oranı düşüktür. Kovidor modeli 5 günlük toplamların son 5 gündeki değişimi *zamana karşı değil, toplam vaka sayısına karşı* grafiğe döktüğü ve incelediği için veride bulunması muhtemel olan geç raporlama ve vaka sayılarının günlere dağıtılması gibi düzensizliklerden minimum seviyede etkilenecektir. Ayrıca lojistik fonksiyonu büküm noktasına ilk ulaştığında üstel fonksiyona çok benzediği için gözle anlam çıkartmak çok zor olmasına karşın kovidor modelini oluşturmak için kullandığımız logaritmik eksenli grafiklerde kovidordan çıkma anı çok belirgin ve anlaşılması kolaydır.
- **Vaka sayılarına sınır konulması**: Vaka sayılarına sınır konup yeni/toplam vaka sayısı  $n$  üstündeyse riskli bir aralıktır diyerek ülkeleri 4e ayırmak mümkündür. Bu yöntem kullanıldığında ülkelerin pandemi süreçlerinde gösterdikleri benzerlik tamamen gözardı edilmiş olur. Çünkü ülkelerin kovidordan farklı vaka sayıları ve toplam vaka sayılarına sahipken çıktıkları açıkça görülmektedir. Ayrıca bu şekilde sınırlar konularak yapılan sınıflandırmalar ülkelerin çok benzer eğime sahip doğruları takip eden davranışlarını modelin dışına bıraktığından başarılı olamayacaktır.

## COVID19 Pandemi Koridoru (Kovidor) Modeli

Sadece 5 ülkenin verilerinin grafiğine bakarak koridorun sınırlarını eksiksiz bir şekilde tanımlamamız mümkün değil. Fakat şunları biliyoruz:

- 5 günlük vaka sayısı toplam vaka sayısından fazla olamayacağı için koridorun üst sınırı  $5\text{GünlükVaka} = \text{ToplamVaka}$  olacaktır.
- Koridorun alt sınırı, bilinmeyen  $m$  ve  $a$  gerçek sayıları kullanılarak şu şekilde ifade edilebilir:

$$\log_{10}(5\text{GünlükVaka}) = m * \log_{10}(\text{ToplamVaka}) + \log_{10}a$$

veya

$$5\text{GünlükVaka} = a * \text{ToplamVaka}^m$$

Yukarıdaki matematiksel modelin  $m$  ve  $a$  parametreleri her ülkenin verileri üzerinde regresyon analizi uygulanarak tahminlenebilir. Özellikle  $m$  parametresindeki ani değişimler, grafikte "çizgideki kırılma" olarak karşımıza çıkan koridordan çıkışı gözümüzle kontrol etmek yerine bir  $\mathbb{R}$  fonksiyonuyla tespit etmemizi sağlayacaktır.

Oluşturacağımız **kovidor** modelinde;

- Grafiklerde gördüğümüz gibi, toplam 100 vakaya ulaşmamış ülkeleri henüz koridora girmemiş kabul edeceğiz.
- Koridora girmiş ülkelerin, hesaplanan son 5 "*5 Günlük Vaka*" sayısı ve son 5 günden önceki "*5 Günlük Vaka*" verilerine 2 ayrı regresyon uygulayarak  $m$  parametresindeki değişimi inceleyeceğiz. Bunu ülke koridordan çıkana kadar tekrarlayacağız.
- Koridordan çıkmamış ülkelerin grafiği hep artış halinde bir doğru şekline yaklaştığı için, bir ülkede  $m$ 'in yaklaşık değeri (grafikte gördüğümüz eğim) -5'ten küçük olduğunda bu ülkeyi koridordan çıkmış kabul edeceğiz.
- Son 5 gün verilerine uygulanan regresyondan elde edilen yaklaşık  $m$  değerini, son 5 günden önceki verilere uygulanan regresyondan elde edilen  $m$  değerine göre kovidordan çıkmaya ne kadar yaklaştığını hesaplayacağız ve buna **sapma oranı** diyeceğiz.
  - **Sapma oranı %0**: Ülke son 5 günde kovidordan çıkmaya yaklaşmadı.
  - **Sapma oranı %100**: Ülke kovidordan çıktı.

Adı verilen ülkenin Covid19 Pandemi Koridoruna (kısaca kovidorda) olma durumunu incelemek için şu 4 fonksiyonu oluşturacağız:

- `covidor_r()` : Adı verilen ülkenin kovidora girip girmediğine karar verdikten sonra; eğer koridora girdiyse 14.günden itibaren her gün için, son 5 güne kadarki ve son 5 güne ait verilere regresyon uygulayarak  $m$  parametresinin yaklaşık değerini hesaplayacak. Çıktı olarak her satırın bir gözlem günü, sütunlarına son 5 günün ve önceki günlerin yaklaşık  $m$  değerleri olduğu bir tablo verecek.
- `who_entered()` : Hazır olarak aldığımız `conf_cases` veri setindeki her ülke için `covidor_r()` fonksiyonunu çalıştıracak ve ülke adlarını *Koridora girmiş olanlar / Koridora girmemiş olanlar* başlıkları altına toplayacak.
- `divi_p()` : `who_entered()` fonksiyonunu çalıştırıp *Koridora girmiş olanlar* başlığı altındaki ülkelerin **sapma oranlarını** hesaplayacak.
- `covidor_model()` : Oluşturduğumuz `covidor_r`, `who_entered()` ve `divi_p()` fonksiyonlarını kullanarak her ülke için şu 3 tür çıktıdan birini içeren bir vektör verecek:
  - *"Kovidora hiç girmedir"*
  - *"Hala kovidorda. Sapma oranı: %[Sapma Oranı]"*
  - *"Kovidordan çıktı."*

```
covidor_r <- function(country = "Turkey")
{
  df <- fived_df(country)
  df <- df %>% drop_na() # Eksik veri içeren satırlar çıkartılıyor
  if(!is.data.frame(df)) return("Hata: En az 10 günlük veri gerekli") # fived_df() 'te hata verenler
  if(nrow(df) <= 10) return("Hata: En az 10 günlük veri gerekli")

  df <- drop_na(df[df[, 4] >= 2, ])
  prev_m <- c()
  current_m <- c()
  if(nrow(df) < 14) return("Koridora girmedir.")
  if(df[nrow(df), 2] < 100) return("Koridora girmedir.")

  for (i in 14:nrow(df))
  {
    prev_t <- df[5: (i - 5), 2]
    prev_g <- df[5: (i - 5), 4]
    current_t <- df[(i - 4) : i, 2]
    current_g <- df[(i - 4) : i, 4]

    prev_m <- c(prev_m, coef(lm(log(prev_g)~log(prev_t), subset = c(prev_t > 1 & prev_g > 1)))[2])
    current_m <- c(current_m, coef(lm(log(current_g)~log(current_t)))[2])
  }
  ndf <- cbind(prev_m, current_m)
  return(ndf)
}
```

```

who_entered <- function()
{
  reg_a <- lapply(colnames(conf_cases)[3:ncol(conf_cases)], covidor_r)
  mode_s <- c()
  for (i in 1:length(reg_a))
  {
    mode_s <- c(mode_s, mode(reg_a[[i]])) # her elemanın türünü sırayla kaydediyoruz
  }
  enteredc <- colnames(conf_cases)[3:ncol(conf_cases)][mode_s == "numeric"]
  didnt_enterc <- colnames(conf_cases)[3:ncol(conf_cases)][mode_s != "numeric"]

  return(list("Kovidora Girmiş Olanlar" = enteredc, "Kovidora Girmemiş Olanlar" = didnt_enterc))
}

```

```

divi_p <- function(country = "Turkey")
{
  df <- country %>% covidor_r()
  for (i in 1:nrow(df))
  {
    pre <- df[i, 1] #son 5 güne kadarki verilerin yaklaşık m değeri
    cur <- df[i, 2] #son 5 günün yaklaşık m değeri

    if (is.na(pre)) pre <- 0
    if (is.na(cur)) cur <- 0
    if (pre < cur) divi <- 0
    if (pre > cur & cur > -5) divi <- round(abs(pre - cur)*100/abs(pre - 5), 2)
    if (cur < -5)
    {
      divi <- 100
      break
    }
  }

  return(divi)
}

```

```

covidor_model <- function()
{
  en_list <- who_entered()
  cov_list <- en_list$`Kovidora Girmiş Olanlar` %>% sapply(divi_p)
  names(cov_list) <- en_list$`Kovidora Girmiş Olanlar`

  n_list <- en_list$`Kovidora Girmiş Olanlar` %>%
    sapply(function(x) paste("Hala kovidorda. Sapma oranı: %", cov_list[x]))

  l_list <- en_list$`Kovidora Girmemiş Olanlar` %>% sapply(function(x) return("Kovidora hiç girmedir."))
  names(l_list) <- en_list$`Kovidora Girmemiş Olanlar`

  lasput <- c(n_list, l_list)[sort(names(c(n_list, l_list)))]
  lasput[lasput == "Hala kovidorda. Sapma oranı: % 100"] <- "Kovidordan Çıktı."
  return(lasput)
}

```

```
covidor_results <- covidor_model()
```

```
covidor_results["Turkey"] # Umarım siz bunu çalıştırdığınızda "Kovidordan çıktı." görürsünüz
```

Turkey: 'Hala kovidorda. Sapma oranı: % 41.81'

covidor\_results

Afghanistan	Hala kovidorda. Sapma oranı: % 0
Albania	Kovidordan Çıktı.
Algeria	Hala kovidorda. Sapma oranı: % 0
Andorra	Kovidordan Çıktı.
Angola	Kovidora hiç girmedir.
Anguilla	Kovidora hiç girmedir.
Antigua and Barbuda	Kovidora hiç girmedir.
Argentina	Hala kovidorda. Sapma oranı: % 0
Armenia	Kovidordan Çıktı.
Aruba	Kovidordan Çıktı.
Australia	Kovidordan Çıktı.
Austria	Kovidordan Çıktı.
Azerbaijan	Hala kovidorda. Sapma oranı: % 0
Bahamas	Kovidora hiç girmedir.
Bahrain	Kovidordan Çıktı.
Bangladesh	Kovidordan Çıktı.
Barbados	Kovidora hiç girmedir.
Belarus	Kovidordan Çıktı.
Belgium	Kovidordan Çıktı.
Belize	Kovidora hiç girmedir.
Benin	Kovidora hiç girmedir.
Bermuda	Kovidordan Çıktı.
Bhutan	Kovidora hiç girmedir.
Bolivia	Hala kovidorda. Sapma oranı: % 26.3
Bonaire Sint Eustatius and Saba	Kovidora hiç girmedir.
Bosnia and Herzegovina	Hala kovidorda. Sapma oranı: % 0
Botswana	Kovidora hiç girmedir.
Brazil	Hala kovidorda. Sapma oranı: % 0
British Virgin Islands	Kovidora hiç girmedir.
Brunei	Kovidordan Çıktı.
...	...

...	...
Suriname	Kovidora hiç girmedir.
Swaziland	Hala kovidorda. Sapma oranı: % 0
Sweden	Hala kovidorda. Sapma oranı: % 28.44
Switzerland	Kovidordan Çıktı.
Syria	Kovidora hiç girmedir.
Taiwan	Kovidordan Çıktı.
Tajikistan	Kovidora hiç girmedir.
Tanzania	Kovidordan Çıktı.
Thailand	Kovidordan Çıktı.
Timor	Kovidora hiç girmedir.
Togo	Kovidordan Çıktı.
Trinidad and Tobago	Kovidordan Çıktı.
Tunisia	Kovidordan Çıktı.
Turkey	Hala kovidorda. Sapma oranı: % 41.81
Turks and Caicos Islands	Kovidora hiç girmedir.
Uganda	Kovidora hiç girmedir.
Ukraine	Hala kovidorda. Sapma oranı: % 81.63
United Arab Emirates	Hala kovidorda. Sapma oranı: % 14.31
United Kingdom	Hala kovidorda. Sapma oranı: % 24.95
United States	Hala kovidorda. Sapma oranı: % 25.15
United States Virgin Islands	Kovidora hiç girmedir.
Uruguay	Kovidordan Çıktı.
Uzbekistan	Kovidordan Çıktı.
Vatican	Kovidora hiç girmedir.
Venezuela	Kovidordan Çıktı.
Vietnam	Kovidordan Çıktı.
Western Sahara	Kovidora hiç girmedir.
Yemen	Kovidora hiç girmedir.
Zambia	Hala kovidorda. Sapma oranı: % 1.31
Zimbabwe	Kovidora hiç girmedir.

## Sonu

- zellikle toplam vaka 3 basamakla-ı sayılara ulařtıktan sonra btn lkelerde pandemi sreci ok benzer davranıř gsteriyor. Bu davranıřa *Pandemi Koridoru* veya *Kovidor* diyoruz.
- Kovidorun etkisini gzle grmek iin lkenin toplam ve 5 gnlk vaka sayıları logaritmik leklerde grafie dklebilir.
- **lkeler kovidordan ıkmadığı srece kovidordaki diğerkm lkelerle ok yakın toplam, gnlk ve 5 gnlk vaka sayılarını tecrbe edecektir.**
- Tm lkelerin kovidorda olma durumlarını hesaplayıp **sapma oranlarıyla** birlikte listeleyen fonksiyon oluřturulmuřtur. Toplam vaka verileri gncellendiğe bu fonksiyon alıřtırılıp lkelerin pandemideki durumları gzlemlenebilir.
- Kovidor modeli **iyimser değildir**. lkenin Kovidor'un dıřında olduėu hesaplandıysa dıřında olduėu kesinleřmiřtir. Kovidor'un dıřına ıkmiř olmak 5 gnlk verilere gre kontrol edildiği iin, bir lkenin kovidordan ıktığının kesinleřmesi 4 gne kadar gecikebilir.

## Sonuçların nemi

- Pandemi kovidoru lkenin pandemideki durumunu en kısa ve net řekilde ifade eden gstergedir. Diğerkm gsterge ve tahminlerin olumlu olması lke pandemi koridorunda olduėu srece sreci anlamamız aıksından pek nemli değildir. **nk kovidordan ıkmadığı srece lke kovidordaki tm diğerkmlerle beraber ok daha kt bir senaryoya srklenmeye devam edecektir.**
- Pandemi koridoru iin oluřturduğumuz model iyimser tahminlere dayanmadığı iin **halkı erken umutlandırmaz**.
- Anlařılır ve temkinli bir gsterge olması nedeniyle *"lke hala kovidorda mı?"* sorusunun cevaplanması halkın pandemi sreci hakkında bilgilendirilmesi aısından nemlidir.

mer Demirtař 24 Nisan2020 .