

## COVID-19 PANDEMİSİ HAKKINDA VERİ ANALİZİ

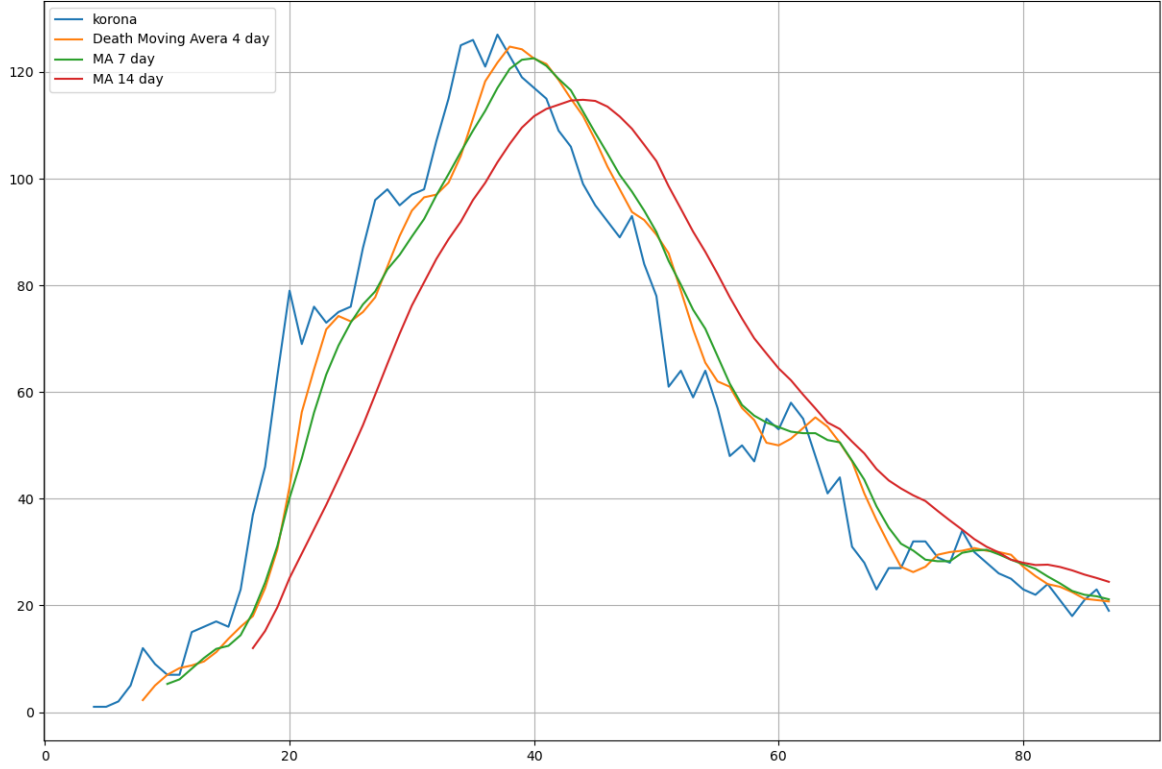
Çalışmamızda güncel bir konu olan ve günümüz gündemimizin büyük bir payına sahip olan Covid-19 hakkında bir veri analizi gerçekleştirilmiştir. Bu veri analizinin Covid-19 hakkında yürütülmesinin temel nedeni dünya genelinde pandemik bir vaka olması ve verilerin değerlendirilmesinde yapılacak ufak bir hatanın büyük yanlışlara yol açabilecek olmasıdır. Bu nedenle ulaşılabilmesi mümkün resmi veriler ile geçmişe yönelik yapılacak bir analiz verimli kılındığı takdirde gelecek için de bir öngörü tablosu önümüze serecektir. Bu veri analizinin de gerçekleştirilmesinin nedeni bunu temel alarak insanların bilinçlenmelerinin ve alacakları tedbirlerin sayısal verilere dayanmasını sağlamaktır. Bu sayede sayısal veriler ile kulaktan dolma bilgilerin de önüne geçilmesi sağlanabilmektedir.

Yapılan analizde Covid-19 verileri arasında büyük farklılıklar görülmektedir ve bu farklılıklar bazı sorunları içermektedir. Verinin düzensiz bir formda ve farklı formatlarda olması, verilerin formatlara aktarılırken eksik, düzensiz ve yetersiz aktarılması bazı sorunlardır. Örneğin csv formatına sahip bir veri çekilmek istendiğinde verilerin düzensiz yapısı veri çekme aşamasında büyük zorluklara neden olmuştur. Bu nedenle çalışmamızda veri çekilmesi aşamasında zorluklar yaşanılmasından dolayı excel formatında daha düzenli veriler ile çalışma yürütülmüştür. Örneğin düzensiz format kapsamı olarak verinin depolanmış olduğu dosya uzantısının ve depolama türünün göstermiş olduğu farklılıklar ve bu doğrultuda yapılacak bir veri çekimi aşamasında ayrı kütüphanelerce gerçekleştirilmesi ve bu kapsamda Pandas kütüphanesi ile ilgili yüklenmiş paketlerin ve paketlerin kullanım bilgisinin kişilerce eksikliği söylenebilmektedir.

Verilerin çekilmesi aşaması sonrasında verilerde kayıp değerlerin var olduğu görülmüştür ve verinin Python IDE programı içerisinde daha iyi sonuçlar ile değerlendirilmesini sağlamak adına kayıp değerler doldurulmuştur. Bu aşama sonucunda istatistiksel hesaplamalar aşamasına geçilmiştir.

Covid-19 pandemisi hakkında ilk olarak hareketli ortalama ile öngörü modeli kullanılmıştır. Hareketli ortalama hesaplamaları doğrultusunda ;

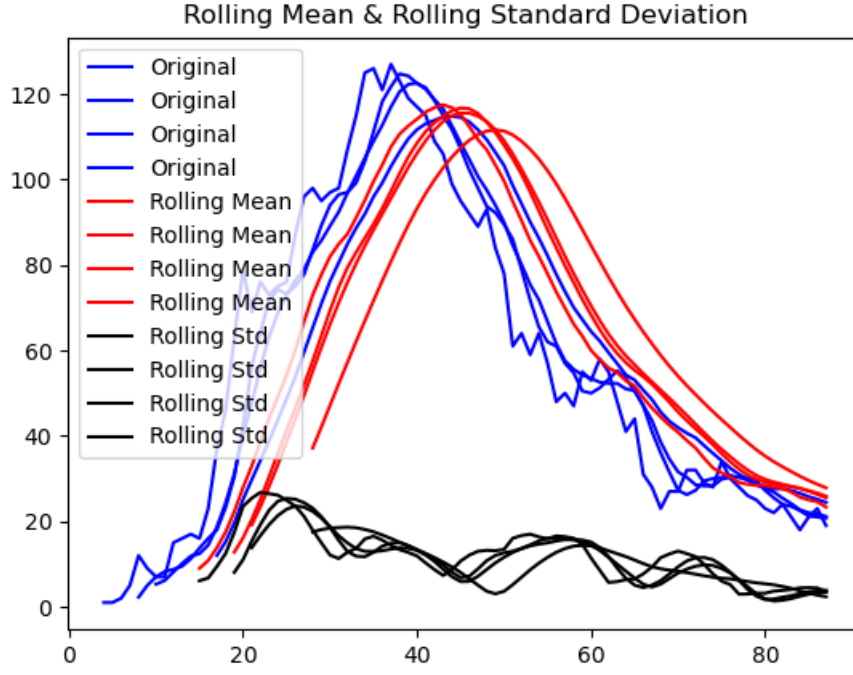
```
Moving average estimation for last 4 date test 42093.0
case 928.0
death 20.0
recovered 2400.0
Name: 87, dtype: float64
test 42093.0
case 928.0
death 20.0
recovered 2400.0
Name: 87, dtype: float64
Moving average estimation for last 7 date test 43891.0
case 907.0
death 21.0
recovered 1776.0
Name: 87, dtype: float64
test 43891.0
case 907.0
death 21.0
recovered 1776.0
Name: 87, dtype: float64
Moving average estimation for last 14 date test 37443.0
case 950.0
death 24.0
recovered 1526.0
Name: 87, dtype: float64
test 37443.0
case 950.0
death 24.0
recovered 1526.0
Name: 87, dtype: float64
```



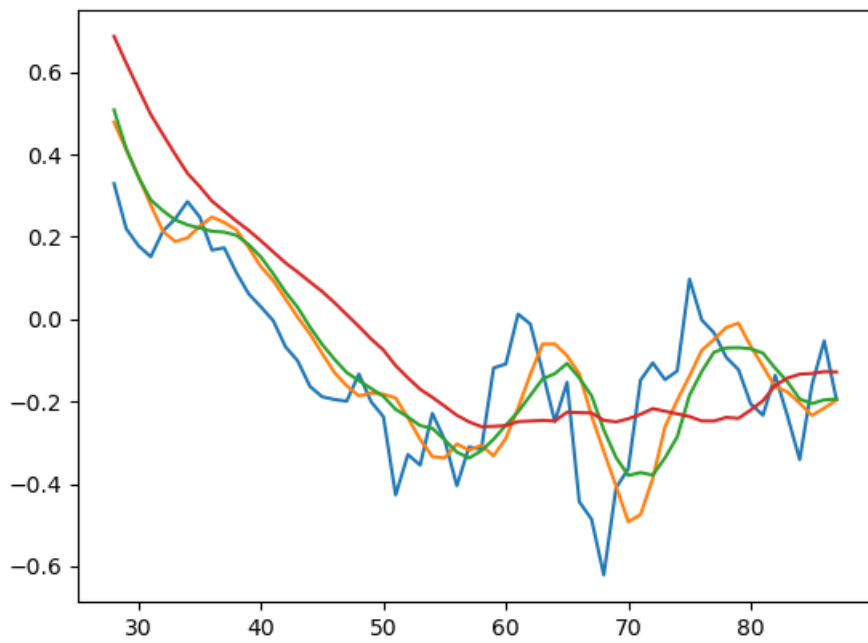
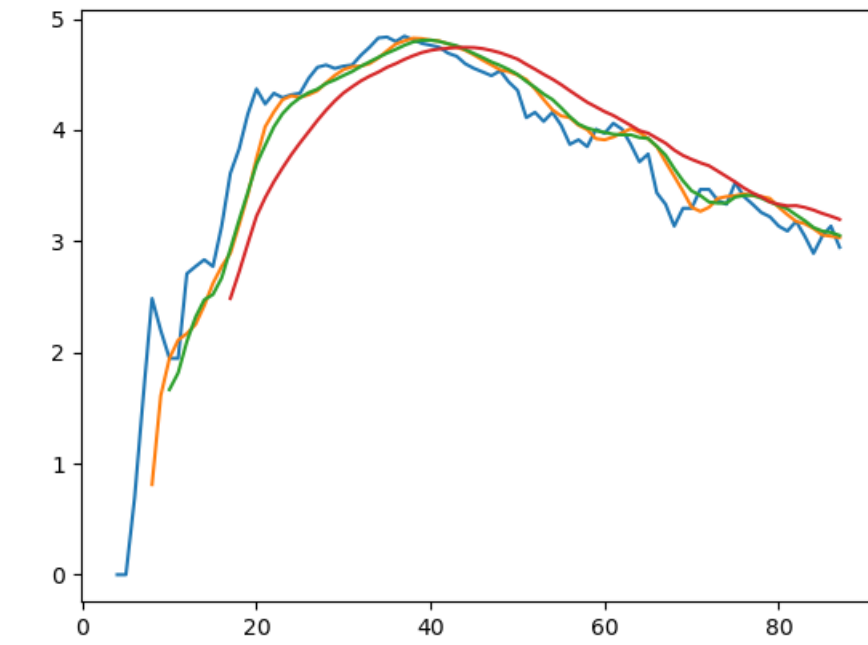
Hareketli ortalama ile geçmişe doğru bir tahmin değeri oluşturulmak ve grafik yapısı ortaya konmak istendiğinde bu doğrultuda son 4,7 ve 14 günlük değerler baz alınmıştır.Grafikte asıl değerler mavi ile gösterilirken; 4 günlük hareketli ortalama turuncu , 7 günlük hareketli ortalama mavi ve 14 günlük hareketli ortalama kırmızı ile gösterilmiştir.Tüm hareketli ortalamalar birlikte ele alındığında ölüm sayılarının ilk veri itibariyle artan bir trend gösterdiği ancak peak noktasından sonra azalan bir trend sergilemiştir.Elde edilen sonuçlara bakıldığında 4 günlük hareketli ortalamanın tahmini gerçek değerlere en yakın sonuçları vermektedir.Baz alınan 4 , 7 ve 14 gün değerleri Covid-19 hastalık belirtilerinin kuluçka sürelerini temsil etmektedir.

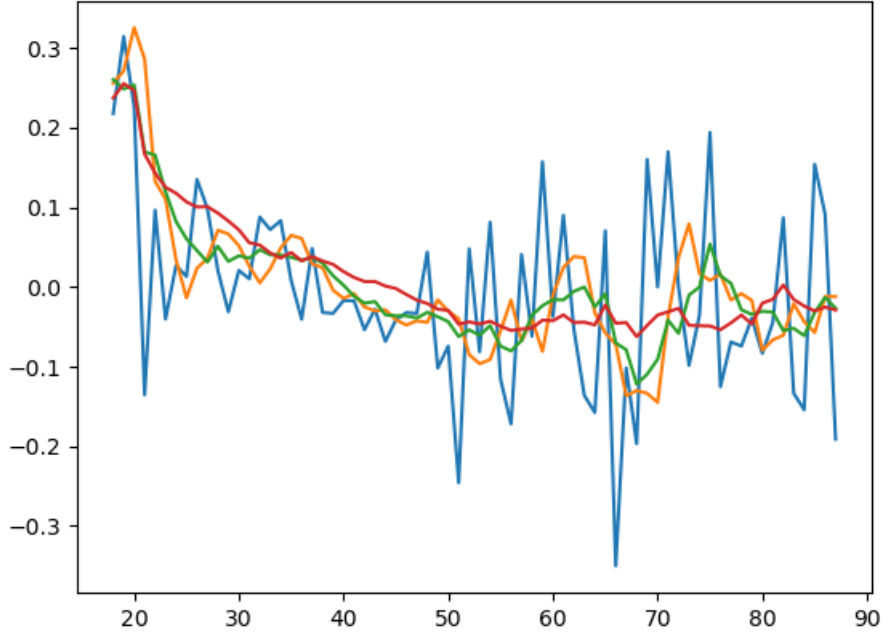
Covid-19 verilerinin bu aşama sonrasında durağanlık (stationarity) durumu analiz edilmiştir.

```
ADF Statistic: -1.4621770929085534
p-value: 0.5521319858354112
Critical Values:
 1%: -3.5097356063504983
 5%: -2.8961947486260944
10%: -2.5852576124567475
```



Elde edilen sonuçlara bakıldığında analizimizle elde edilmiş p değeri :  $0.55 > p$  değeri ( 0.05) anlamlılık düzeyi olması nedeniyle ve ADF istatistik değerinin negatif sonuç vermesi nedeniyle verilerin durağan olmadığı anlaşılmaktadır. Bu sonuçlara dayanarak veriler bazı düzenlemelere tabi tutulmuştur. Düzenlemelerin tahmin grafikleri aşağıda belirtildiği gibidir ve daha iyi bir sonuç ortaya koyamamaktadır.





Yukarıda grafikler ve açıklamalarla belirtilmiş olan gerekçeler dahilinde elde edilen sonuçların durağanlık arz etmemesi analizimizi ARIMA testine yönlendirmiştir. Durağan veriler olmaması ARIMA testini geçerli kılmaktadır. Test sonuçları aşağıda belirtildiği gibi olup grafik ile de gösterilmiştir.

ARIMA Model Results						
Dep. Variable:	D.death	No. Observations:	87			
Model:	ARIMA(1, 1, 2)	Log Likelihood	-272.872			
Method:	css-mle	S.D. of innovations	5.565			
Date:	Sun, 21 Jun 2020	AIC	555.744			
Time:	20:19:11	BIC	568.074			
Sample:	1	HQIC	560.709			
	coef	std err	z	P> z	[0.025	0.975]
const	0.1916	0.932	0.206	0.837	-1.636	2.019
ar.L1.D.death	0.2415	0.366	0.660	0.509	-0.475	0.958
ma.L1.D.death	-0.0481	0.362	-0.133	0.894	-0.757	0.661
ma.L2.D.death	0.2424	0.116	2.088	0.037	0.015	0.470
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	4.1400	+0.0000j	4.1400	0.0000		
MA.1	0.0992	-2.0286j	2.0311	-0.2422		
MA.2	0.0992	+2.0286j	2.0311	0.2422		
count	87.000000					
mean	0.001790					
std	5.596976					
min	-16.323827					
25%	-3.253479					
50%	-0.560168					
75%	4.166903					
max	12.814897					

Sonuçlar incelendiğinde ve p değerleri ile 0.05 anlamlılık düzeyi p değeri kıyaslandığında ma.L1.D.death değişkeninin anlamsız olduğu ve ayrıca durağanlık olmadığı görülmektedir. Bu nedenle sonuçların bir öngörü modeli olarak kullanılamayacağı kanısına varılmıştır.

