



T.C.

FIRAT ÜNİVERSİTESİ  
TEKNOLOJİ FAKÜLTESİ

# YAZILIM MÜHENDİSLİĞİ GÜNCEL KONULAR

**ADI :** Yunus

**SOYADI :** OLGUN

**BÖLÜMÜ :** Yazılım Mühendisliği

**NUMARA :** 14542521

**PROJE KONUSU :** Veri Madenciliği

**DERS HOCASI :** Doç. Dr. Fatih Özkaynak

# ÖNSÖZ

Bu döküman Yazılım Mühendisliği Güncel Konular dersi için yapılan proje için hazırlanmıştır.

Hava kirliliği, gelişmekte olan çoğu ülkelerde en önemli çevresel sorunların başında gelmektedir. Sanayileşme, hızlı nüfus artışı ve kentsel yapılaşma nedeniyle Türkiye’de hava kalitesinin belirlenmesi önem arz etmektedir.

Bu projede hedef; Çevre ve Şehircilik Bakanlığı’ nın [www.havaizleme.gov.tr/](http://www.havaizleme.gov.tr/) sitesi aracılığıyla paylaşmış olduğu bölgelere, şehirlere, ilçelere göre hava kirliliği verileri üzerinden veri toplama, verinin görselleştirilmesi ve sunumu, verilerin karakteristiklerinin analizi, model kurma, geleceğe yönelik veri tahmini, faydalı bilgi çıkarımı aşamalarını sırasıyla gerçekleştirebilmektir.

## Verilerin Toplanması

Hava kirliliđi insan sađlıđını dođrudan veya dolaylı olarak büyük ölçüde etkileyebilmektedir. Hava kirliliđi canlılarda bazı durumlarda kalıcı hasarlara ve hatta bazı durumlarda ölümlere neden olabilmektedir. Bu çerçevede veri istasyonu olarak Türkiye'nin en büyük sanayi şehirlerinden biri olan ve Türkiye'de nüfus yoğunluğu en fazla olan ilk 10 şehrin içerisinde ki Kocaeli'yi seçtim. İstasyon olarak Kocaeli ilini seçmemin asıl nedeni; 2017 yılında yapılan araştırmalarda kanser nedeni ile gerçekleşen ölümlerin dünyada %12.5, Türkiye'de 12.9, Kocaeli'de ise % 18.9 olmasıydı.

Şuan içinde bulunduğumuz pandemi sürecinde belli kısıtlamalar ve işletmelerin kendi inisiyatifleriyle; okullar uzaktan eğitime geçmiş, haftasonları sokađa çıkma kısıtlamaları yapılmış, bazı işletmeler faaliyetlerine ara vermiş veya azaltmıştır.Bu durumun hava kirliliđine etkisini incelemek ve analiz edebilmek için veri seti olarak pandemi süreci öncesi ve pandemi süreci sonrası olmak üzere iki ayrı veri seti belirledim.Bu veri setlerinde parametreler ; PM10, SO2, NO2 , NOX , O3, tarih ve saat bilgileridir. Veri setlerinden birisinin tarih

parametresi pandemi sürceinin başlangıcından günümüze, diğer veri setinin tarih parametresi ise geçen yılın aynı aylarını içermektedir.

**PM10(Partiküler Madde)** : Partiküler madde, havada asılı katı veya sıvı maddelerin mikroskopik parçacıkları olup en tehlikeli[1] hava kirliliğidir.

**SO2(Kükürt dioksit)** : Hava kirliliğine ve asit yağmurlarına sebep olur (SO<sub>2</sub>), renksiz, keskin kokulu reaktif bir gaz olup kömür, fuel-oil gibi kükürt içeren yakıtların yanması sırasında, metal eritme işlemleri ve diğer endüstriyel işlemler sonucu oluşur.

**O3(Ozon gazı)** : Ozon gazı (O<sub>3</sub>), 3 oksijen atomundan oluşan molekülleriyle renksiz bir gazdır ve atmosferin üst katmanlarında yer alır.

**NO2(Azot Dioksit)**: Azot dioksit, NO<sub>2</sub>, kırmızımsı kahve renkli ve zehirli bir gazdır. Keskin bir kokuya sahip olan bu gazın kokusu klor gazına benzemektedir.

**NOX(Azot oksit)**:Azot oksit, insanlar da dahil olmak üzere memelilerin vücutlarında önemli bir sinyal molekülü olmasının yanı sıra kimyasal endüstride de önemli bir ara üründür.

İki veri setimizde de eksik veriler olduğu için ve verilerin değer aralıkları fazla olduğu için eksik verileri gidermemiz ve veri setini normalize etmemiz gerekmektedir.

Eksik verileri gidermek ve veri setini normalize etmek için RapidMiner programında Replace Missing Values ve Normalize metotları kullanılmıştır.

Geçen seneki hava kirliliğine ait ve bu sene pandemi sürecindeki hava kirliliğine ait veri setleri :

	A	B	C	D	E	F	G	H	I
1	Tarih	PM10 ( µg )	SO2 ( µg/r )	NO2 ( µg/ )	NOX ( µg/ )	O3 ( µg/m³ )			
2	2019-03-11 16:00:56	50,5	15,5	83,1	123,2	53,0			
3	2019-03-11 17:00:56	50,5	5,9	100,9	170,0	35,7			
4	2019-03-11 18:00:56	71,7	6,2	115,2	193,1	28,3			
5	2019-03-11 19:00:56	62,2	24,1	141,5	237,6	15,8			
6	2019-03-11 20:00:56	70,2	21,4	143,2	303,6	4,5			
7	2019-03-11 21:00:56	78,0	19,2	51,6	81,2	4,9			
8	2019-03-11 22:00:56	125,9	10,9	98,4	155,7	54,0			
9	2019-03-11 23:00:56	121,6	7,8	91,4	183,3	3,3			
10	2019-03-12 00:00:56	93,0	4,5	89,2	220,2	3,0			
11	2019-03-12 01:00:56	92,9	4,7	85,5	251,9	2,7			
12	2019-03-12 02:00:56	89,8	2,5	69,0	96,9	5,3			
13	2019-03-12 03:00:56	53,0	1,9	52,3	80,3	23,2			
14	2019-03-12 04:00:56	51,7	1,4	56,2	80,0	16,2			
15	2019-03-12 05:00:56	42,1	,8	60,5	79,1	7,7			
16	2019-03-12 06:00:56	46,0	,7	49,0	63,6	18,4			
17	2019-03-12 07:00:56	45,2	,1	64,0	137,1	5,0			
18	2019-03-12 08:00:56	54,3	2,5	83,9	357,8	1,7			
19	2019-03-12 09:00:56	98,0	6,1	105,5	534,2	2,7			
20	2019-03-12 10:00:56	113,1	6,1	88,8	291,9	4,7			
21	2019-03-12 11:00:56	117,6	13,7	127,5	680,1	4,4			
22	2019-03-12 12:00:56	173,9	15,5	120,5	434,5	6,7			
23	2019-03-12 13:00:56	168,6	15,8	127,8	375,4	4,7			
24	2019-03-12 14:00:56	119,8	32,7	124,7	363,5	7,8			
25	2019-03-12 15:00:56	50,5	28,3	89,3	81,2	14,3			
26	2019-03-12 16:00:56	50,5	7,6	25,9	34,1	52,8			
27	2019-03-12 17:00:56	33,0	5,2	18,1	22,1	68,9			
28	2019-03-12 18:00:56	32,8	4,0	22,8	29,8	63,1			
29	2019-03-12 19:00:56	24,9	3,8	54,8	97,0	32,0			
30	2019-03-12 20:00:56	23,0	3,7	73,8	122,4	13,6			
31	2019-03-12 21:00:56	21,7	3,4	75,7	137,7	12,4			
32	2019-03-12 22:00:56	12,9	2,3	51,6	81,2	57,5			

Retrieve replace\_gecensene

Hazır

	A	B	C	D	E	F	G	H
1	Tarih	PM10 ( $\mu\text{g}$ )	SO2 ( $\mu\text{g}/\text{l}$ )	NO2 ( $\mu\text{g}/\text{l}$ )	NOX ( $\mu\text{g}/\text{l}$ )	O3 ( $\mu\text{g}/\text{m}^3$ )		
2	2020-03-11 16:00:56	50,9	43,0	60,1	104,3	42,0		
3	2020-03-11 17:00:56	49,3	45,0	57,4	93,3	44,8		
4	2020-03-11 18:00:56	57,8	37,0	68,0	112,4	31,7		
5	2020-03-11 19:00:56	50,8	11,1	63,9	100,8	29,9		
6	2020-03-11 20:00:56	45,3	18,5	79,8	136,1	13,6		
7	2020-03-11 21:00:56	56,0	13,1	94,7	287,8	1,7		
8	2020-03-11 22:00:56	88,8	8,9	87,6	258,9	1,2		
9	2020-03-11 23:00:56	81,8	6,1	71,1	193,9	,8		
10	2020-03-12 00:00:56	77,5	4,6	55,5	194,9	1,9		
11	2020-03-12 01:00:56	92,4	3,1	46,3	158,7	,9		
12	2020-03-12 02:00:56	74,7	2,3	43,3	154,7	,9		
13	2020-03-12 03:00:56	64,8	1,1	38,6	90,2	,6		
14	2020-03-12 04:00:56	56,9	,8	36,0	92,2	,6		
15	2020-03-12 05:00:56	53,9	,3	33,4	71,4	,5		
16	2020-03-12 06:00:56	49,7	,2	29,4	42,2	5,7		
17	2020-03-12 07:00:56	40,3	,5	34,5	66,5	1,6		
18	2020-03-12 08:00:56	48,5	,9	40,7	107,6	2,3		
19	2020-03-12 09:00:56	63,1	4,1	50,3	219,1	2,7		
20	2020-03-12 10:00:56	95,5	6,2	48,0	125,1	11,0		
21	2020-03-12 11:00:56	80,0	10,7	63,9	145,3	28,6		
22	2020-03-12 12:00:56	90,0	13,1	56,2	81,0	62,4		
23	2020-03-12 13:00:56	88,8	8,2	24,6	42,4	72,4		
24	2020-03-12 14:00:56	49,6	6,0	22,2	34,3	74,2		
25	2020-03-12 15:00:56	35,8	3,3	16,8	27,5	84,0		
26	2020-03-12 16:00:56	28,0	3,8	10,1	12,5	88,9		
27	2020-03-12 17:00:56	29,2	4,6	13,4	16,2	83,3		
28	2020-03-12 18:00:56	36,9	3,1	13,8	15,9	79,1		
29	2020-03-12 19:00:56	32,8	2,6	19,1	21,0	67,7		
30	2020-03-12 20:00:56	27,8	2,9	49,7	54,2	35,1		
31	2020-03-12 21:00:56	52,5	2,5	61,2	70,9	22,3		
32	2020-03-12 22:00:56	69,0	1,8	65,8	83,6	9,3		

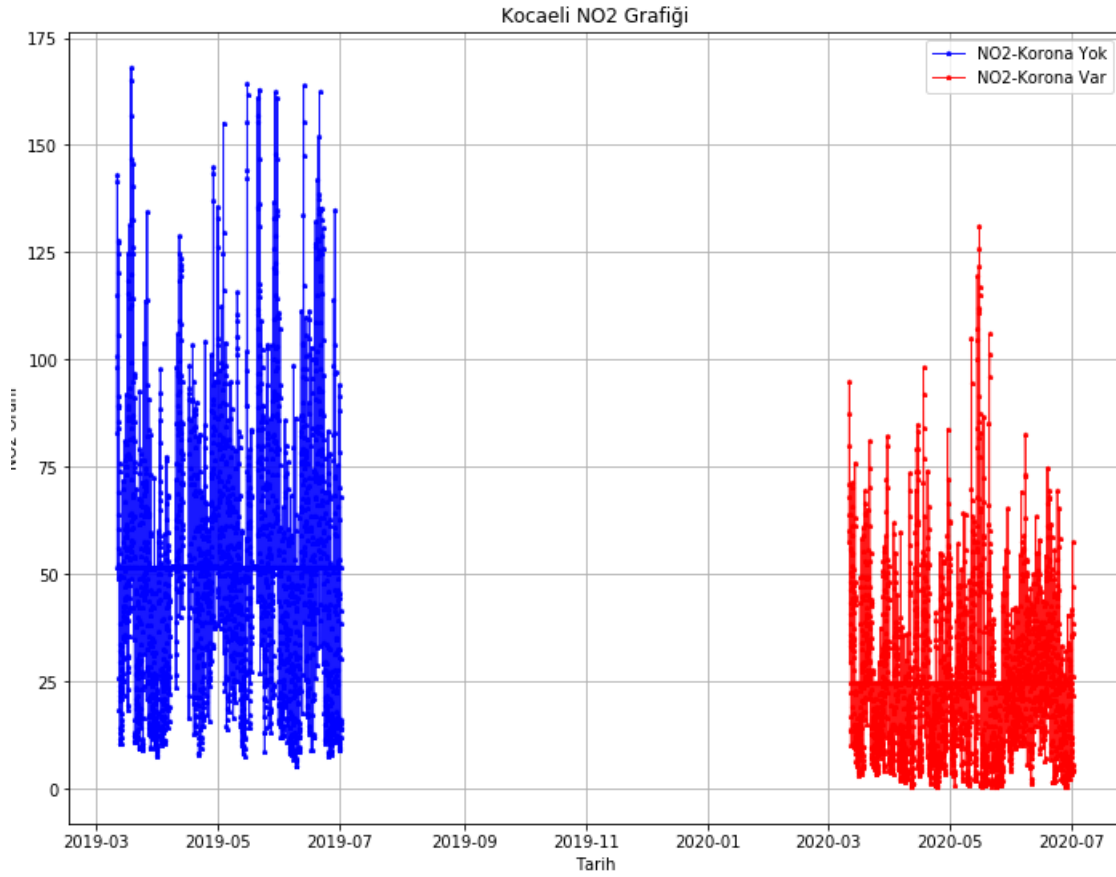
## Verilerin Görselleştirilmesi

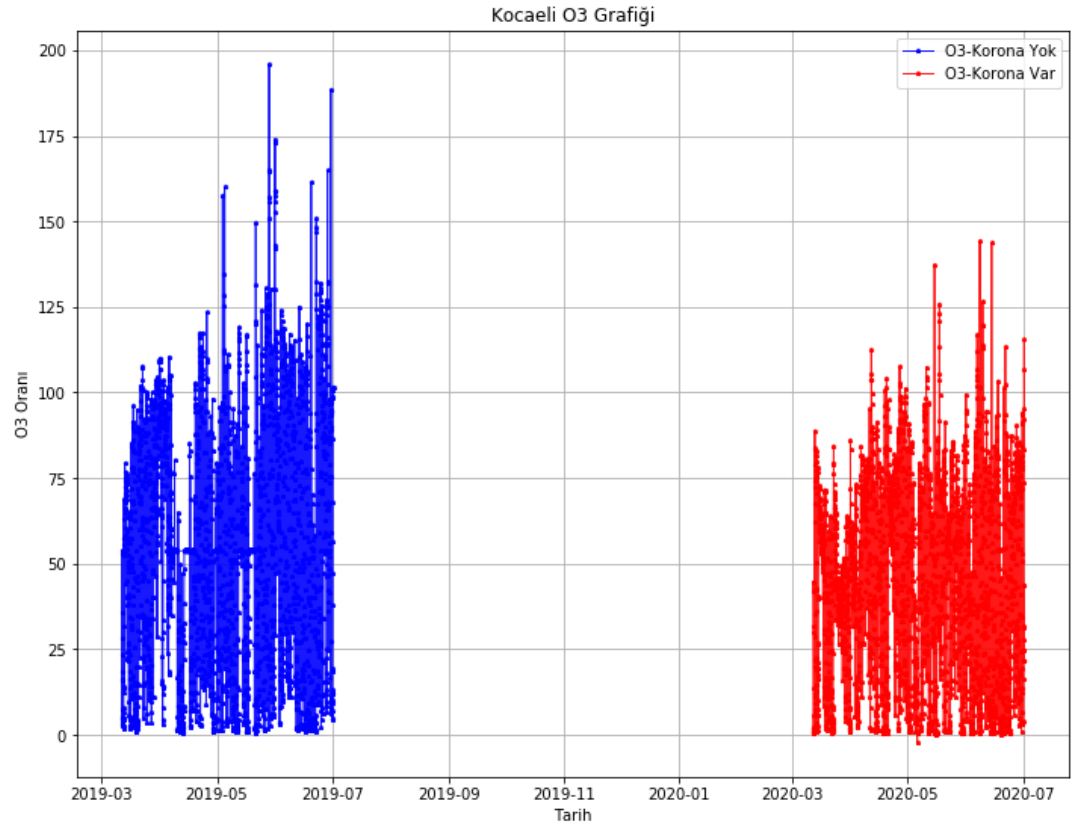
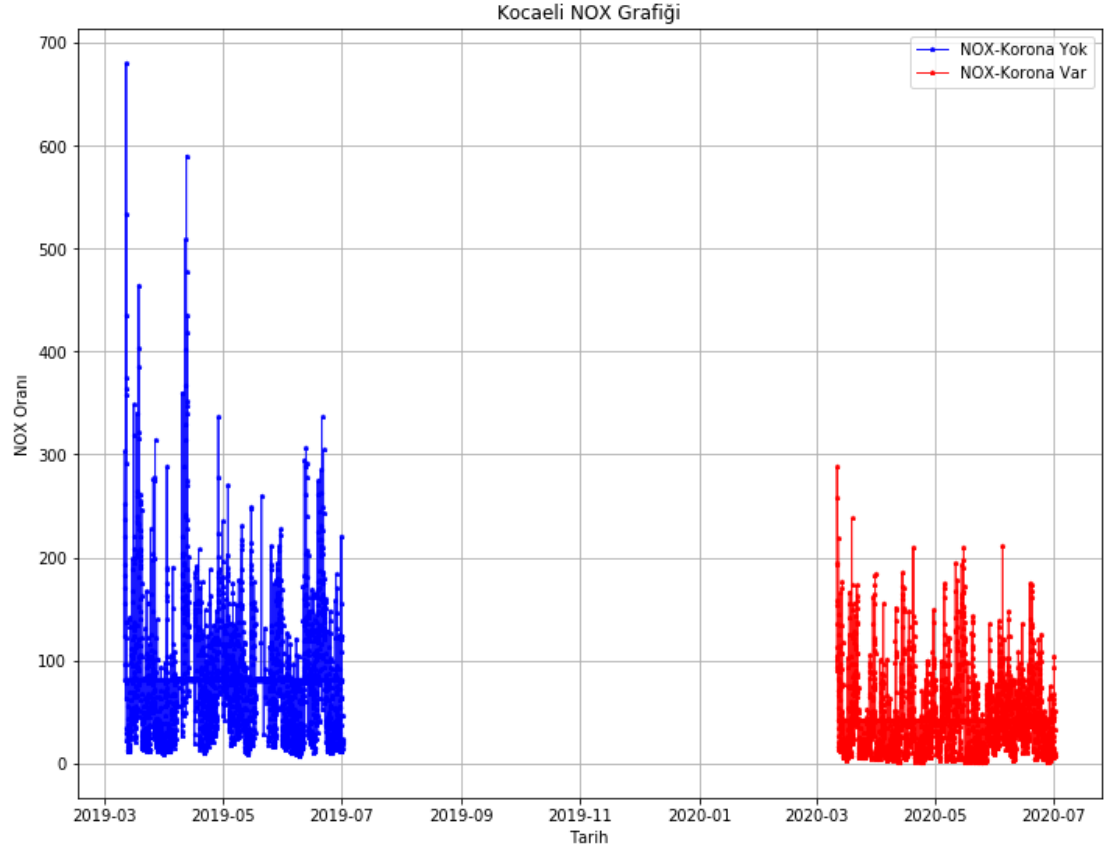
Elimizde bulunan herhangi bir büyük veri setinde değerlere bakarak yorum yapmak veya o veri setinden herhangi bir çıkarımda bulunmak kolay değildir. Veri setinden bir sonuç çıkarmak ve o veri setini yorumlayabilmek için verilerin görselleştirilmesi gerekmektedir.

Verilerin görselleştirilerek verilerin analiz edilebilmesi için Python programla dilini kullanabiliriz. Bunun için Pandas ve matplotlib.pyplot kütüphanelere işimize yarayacaktır.

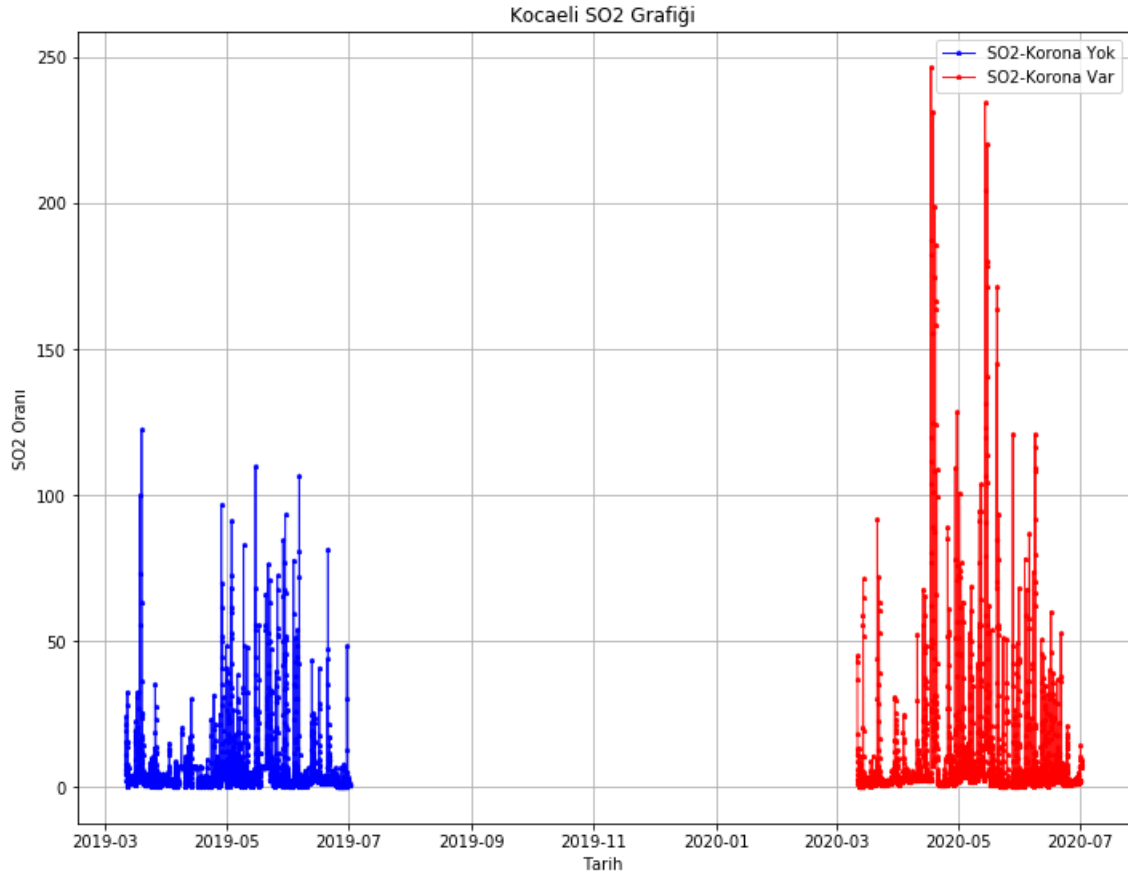
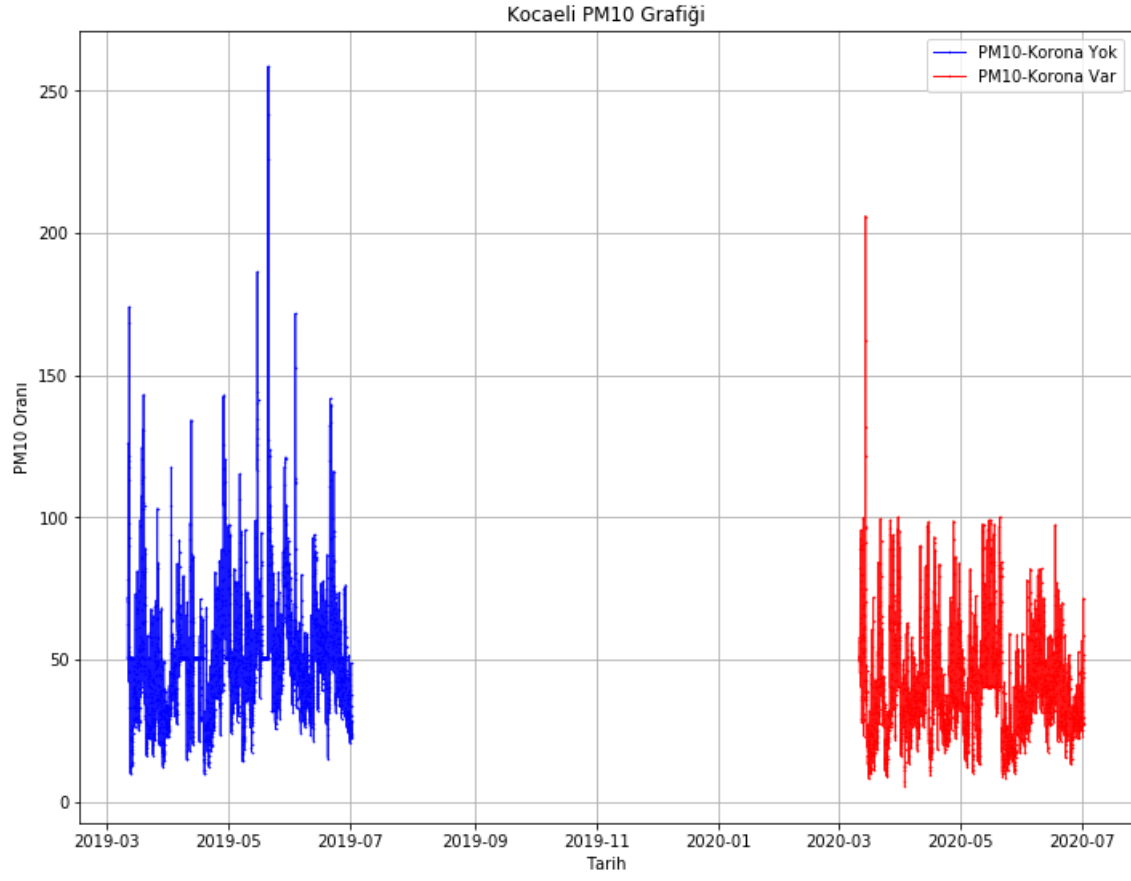
## Verilerin Karakteristik Analizi

Python programlama dili kullanılarak verilerin görselleştirilmiş durumları:4









Yukarıdaki grafiklerden anlaşıldığı gibi kırmızı çizgiler korona virüsün ülkemize girdiği tarih ile günümüze kadar olan süreci kapsamaktadır. Mavi çizgiler ise geçtiğimiz sene için aynı tarihleri kapsamaktadır.

## Sonuç

Şekillere baktığımız zaman genel olarak kırmızı çizgilerin (korona virus sonrası), mavi çizgilere (geçen sene aynı tarihler) göre daha düşük seviyelerde olduğunu görüyoruz. Bu durumda rahatlıkla Kocaeli ilinde pandemi süreci boyunca hava kirliliğinin azaldığını söyleyebiliriz. Buna bağlı olarak fabrikaların faaliyetlerinin azalmasının, okulların açık olmamasıyla motorlu taşıtların kullanımının azalmasının hava kirliliği ile doğru orantılı olduğunu anlıyoruz.

Kanser nedeni ile gerçekleşen ölümlerin Kocaeli’de % 18.9 gibi büyük bir oran olmasının nedenini fabrika sayısının çokluğu, nüfus kalabalığı ve araç sayısının fazlalığı ile ilişkilendirebiliriz.

## Kaynakça

- <https://www.havaizleme.gov.tr/>
- <https://www.kocaelibarisgazetesi.com/guncel/her-100-kisiye-1-fabrika-dusuyor-2-h105666.html>
- <https://tr.wikipedia.org/wiki/Partik%C3%BCl#PM10>
- [https://tr.wikipedia.org/wiki/K%C3%BCk%C3%BCrt\\_dioksit#:~:text=K%C3%BCk%C3%BCrt%20dioksit%20form%C3%BCl%C3%BCnde%20bir%20k%C3%BCk%C3%BCrt%20iki%20oksijen%20yer%20al%C4%B1r.&text=Hava%20kirlili%C4%9Fine](https://tr.wikipedia.org/wiki/K%C3%BCk%C3%BCrt_dioksit#:~:text=K%C3%BCk%C3%BCrt%20dioksit%20form%C3%BCl%C3%BCnde%20bir%20k%C3%BCk%C3%BCrt%20iki%20oksijen%20yer%20al%C4%B1r.&text=Hava%20kirlili%C4%9Fine)

[%20ve%20asit%20ya%C4%9Fmurlar%C4%B1na,di%C4%9Fer%20end%C3%BCstriyel%20i%C5%9Flemler%20sonucu%20olu%C5%9Fur.](#)

- [https://tr.wikipedia.org/wiki/Ozon#:~:text=Ozon%20gaz%C4%B1%20\(O3\)%2C,allotr opudur%2C%20trioksijen%20olarak%20da%20tan%C4%B1mlan%C4%B1r.](https://tr.wikipedia.org/wiki/Ozon#:~:text=Ozon%20gaz%C4%B1%20(O3)%2C,allotr opudur%2C%20trioksijen%20olarak%20da%20tan%C4%B1mlan%C4%B1r.)
- [https://tr.wikipedia.org/wiki/Azot\\_dioksit#:~:text=Azot%20dioksit%2C%20NO2%2C%20k%C4%B1rm%C4%B1z%C4%B1ms%C4%B1,4%20%C5%9Feklinde%20gaz%20haline%20gelmektedir.](https://tr.wikipedia.org/wiki/Azot_dioksit#:~:text=Azot%20dioksit%2C%20NO2%2C%20k%C4%B1rm%C4%B1z%C4%B1ms%C4%B1,4%20%C5%9Feklinde%20gaz%20haline%20gelmektedir.)
- <http://www.okotec.com/azot-oksit-nedir-nox>
- <https://www.youtube.com/channel/UCzR0evbMxyeZV2VF53qqRcw>