

İstanbul İlçe Bazında Aylık Doğalgaz Tüketim Tahmini

Satıgöl ÖZTÜRK 121519044

Emre TORTUMLU 121519063

Feyza Yağmur MUMCU 121519012

İrem Gözde HACI 121519015

3 TEMMUZ 2023

Contents

TEŞEKKÜRLER	2
ÖZET	3
ABSTRACT	4
GİRİŞ	5
BÖLÜM 1	6
1.1 Doğalgaz Tüketimi	6
BÖLÜM 2	9
2.1 Kullanılan Yöntemler	9
BÖLÜM 3	19
3.1 Veri Tanıtımı	19
3.2 Veri Analizi	19
3.3 Modellemeye Hazırlık Aşamaları	26
3.4 Arima Modeli	30
3.5 Auto.arima Modellerinin Oluşturulması	31
3.6 ARIMA(2,0,0)(2,1,1)[12] İle 12 Adımlık Tahmin Değerleri	33
3.7 Sıcaklık İle Doğalgaz Tüketimi Verisi Arasındaki İlişki	33
3.8 2022 Yılı İçin Doğalgaz Tüketimi İle Abone Sayısı Verisinin İlişkisi	36
3.9 Çeyreklik veri İle Zaman Serisi	39
SONUÇ	42
KAYNAKÇA	43

TEŞEKKÜRLER

Lisans eğitimimiz boyunca değerli bilgilerini esirgemeyen Prof. Dr. Müjgan Tez, tez sürecinde bizlere desteklerini esirgemeyen Prof. Dr. Birsen Eygi Erdoğan ve bölüm başkanımız Prof. Dr. Deniz İnan hocalarımıza sonsuz teşekkürler.

ÖZET

Doğalgaz, diğer gelişmiş ve gelişmekte olan ülkelerde olduğu gibi ülkemiz ekonomisinin en önemli girdilerinden olup aynı zamanda hane halkının ısınma ihtiyacını karşılayan temel enerji kaynağıdır. Doğalgaz temini noktasında tamamen dışa bağımlı bir politika izleyen ve doğalgaz stok kapasitesi toplam tüketiminin ancak %5 ini karşılayan ülkemiz için tüketilen doğalgaz miktarının tahmini, son derece önemlidir. Ülkemizde enerji arz kaynaklarının çeşitlendirilmesi amacıyla ithal edilmeye başlanan doğalgazın yıllık tüketimi 50 milyar metreküp seviyelerine yaklaşmış ve Türkiye birincil enerji kaynakları pastasındaki en büyük paya sahip hale gelmiştir. Doğalgazın, sanayi sektörünün en önemli enerji girdisi olması, ulusal elektrik üretiminin yaklaşık yarısının kaynağını teşkil etmesi, ısınma ihtiyacını karşılayan temel yakıt olması gibi gerçekleri arkasına alarak son yıllarda hızla artan önemini perçinlemesi ve dünya genelindeki artışa da paralel olarak ülkemizdeki doğal gaz tüketiminin daha da ileri seviyelere gideceği beklenmektedir. Bu noktada, gerek makroekonomik düzeyde yapılacak planlamalara esas teşkil etmesi gerekse arz güvenliğinin sağlanması hususlarında ülkemizin gelecek yıllardaki doğal gaz tüketiminin doğru şekilde tahmin edilebilmesi büyük önem taşımaktadır. Bu çalışmada enerji kaynağı olarak doğalgazın İstanbul'daki tüketim değerleri incelenmiş, Duragan Olmayan Doğrusal Stokastik Model (ARIMA Modeli) kullanılarak İstanbul'un gelecek aylardaki ve mevsimsel dönemlerdeki doğal gaz tüketimi tahmin edilmeye çalışılmıştır. Veriler 2015 yılı ile 2023 Nisan ayı ve 2022 abone sayılarını kapsamaktadır. İstanbul İlçe Bazında Aylık Doğalgaz Tüketim verileri ve Abone Sayıları İstanbul Büyükşehir Belediyesi Açık Veri Portalından tedarik edilmiştir.

ABSTRACT

Natural gas, like in other developed and developing countries, is one of the most important inputs to our country's economy and serves as the primary energy source for meeting household heating needs. For our country, which follows a completely dependent policy in terms of natural gas supply and where the natural gas stock capacity covers only 5% of total consumption, estimating the amount of natural gas consumed is of great importance. With the aim of diversifying energy supply sources in our country, the annual consumption of imported natural gas has approached the level of 50 billion cubic meters, and Turkey has become the largest share in the pie of primary energy sources. Taking into account the fact that natural gas is the most important energy input in the industrial sector, constitutes about half of the national electricity generation, and serves as the primary fuel for meeting heating needs, it is expected that its importance, which has rapidly increased in recent years, will continue to rise in parallel with the global increase, and natural gas consumption in our country will reach even higher levels. At this point, it is of great importance for our country to be able to accurately predict the future natural gas consumption in the coming years, both in terms of serving as the basis for macroeconomic planning and ensuring supply security. In this study, the consumption values of natural gas in Istanbul as an energy source were examined, and the Non-Stationary Linear Stochastic Model (ARIMA Model) was used to attempt to predict the natural gas consumption in Istanbul in the coming months and seasonal periods. The data covers the period from 2015 to April 2023 and the subscriber numbers for 2022. The Istanbul District-based Monthly Natural Gas Consumption data and Subscriber Numbers were obtained from the Istanbul Metropolitan Municipality Open Data Portal.

GİRİŞ

Havayı kirletmeyen, petrol ve katı yakıtla kıyaslandığında çevre dostu bir enerji kaynağı olan doğalgaz, yüksek yanma verimi, kolaylıkla depolanabilir olması, kurum ve is gibi atık madde bırakmaması, kontrol edilebilir bir enerji kaynağı olması ve fiyat avantajlarından dolayı, başta ısınma olmak üzere, elektrik üretimi ve sanayide birçok alanda kullanılmaktadır. Çevre dostu olan bu enerji kaynağının temini noktasında tamamen dışa bağımlı bir politika izleyen Türkiye, doğalgaz ihtiyacının büyük bölümünü boru hatları yardımıyla Rusya, İran ve Azerbaycan'dan karşılamaktadır. Ayrıca Marmara Ereğlisi'ne Nijerya ve Cezayir'den gemilerle sıvılaştırılmış doğal gaz (LNG) tedarik edilmektedir. Doğalgazın çıkarıldığı kaynaktan tüketim noktalarına nakli, uluslararası antlaşmalarla güvence altına alınmasına karşın çeşitli nedenlerle, doğalgaz ithal edilen ülkelerden gelen doğalgaz akışı kesilebilmekte ve bunun sonucunda özellikle kış aylarında konutlarda ısınma ve sanayide üretimin durması gibi problemler yaşanmaktadır. Doğalgaz tüketim tahmininde doğrusal regresyon ve zaman serileri kullanılmaktadır. Geçmiş dönemlere ilişkin gözlem değerleri yardımıyla geleceğe yönelik tahminler yapmayı amaçlayan zaman serisi modelleri tıp, mühendislik, işletme, ekonomi ve finans gibi birçok alanda yaygın bir şekilde kullanılmaktadır. Zaman serileri yardımıyla tahmin yapmak için değişik yöntemler kullanılarak oluşturulan farklı modeller bulunmaktadır. Bu modeller arasında en çok bilinen ve yaygın olarak kullanılan ARIMA modelleridir. Seriyi oluşturan veriler arasında doğrusal bir ilişkinin olduğunu varsayan ve bu doğrusal ilişkiyi modelleyebilen ARIMA modelleri durağan ya da çeşitli istatistiksel yöntemlerle durağan hale getirilen zaman serilerine başarıyla uygulanabilmektedir. Bu çalışmada ise İstanbul iline ilişkin ilçelere ait aylık doğalgaz tüketimi tahmin edilmeye çalışılmıştır. Bu amaçla zaman serileri kullanılarak tahmin modelleri oluşturulmuş aylık ve mevsimsel veriler kullanılarak kısa dönemli tahmin sonuçları elde edilmiştir. Ayrıca, 2022 yılı için İstanbul'a ait İstanbul Büyükşehir Belediyesi Açık Veri Portalından alınan abone sayıları verilerine göre ayrı bir çalışma yapılmıştır.

BÖLÜM 1

1.1 Doğalgaz Tüketimi

1.1.1 Doğalgaz Nedir?

Fosil yakıtlar grubundan hidrokarbon esaslı doğalgaz, yer altında gözenekli kayaların boşluklarına sıkışmış olarak yada petrol yataklarının üzerinde gaz halinde büyük hacimler şeklinde bulunur. Doğalgaz; %95 metan, az miktarda da etan, propan atom, bütan ve karbondioksitten oluşan renksiz, kokusuz ve havadan hafif bir gazdır. Doğalgaz kokusuz olduğundan kaçakların farkedilebilmesi için özel olarak kokulandırılır. Bu amaçla THT(tetra hidro teofen) ve / veya TBM (tersiyer bütıl merkaptan) kullanılır. Karışımın içinde %95 yada daha yüksek bulunan metan gazının özelliği kimyasal yapısı en basit ve karbon içeriği en düşük olan hidrokarbon gazı olmasıdır. Metan molekülü 1 karbon 4 hidrojen atomundan oluşur. Kimyasal yapısının basit olması nedeniyle yanma işlemi kolaydır ve tam yanma gerçekleşir. Dolayısıyla; duman, is,kurum,ve kül oluşturmaz. Yanması en kolay ayarlanabilen ve yanma verimliliği en yüksek olan yakıttır. Bu özelliği kullanım kolaylığı ve ekonomisi sağlar. Karbon içeriğinin düşük olması nedeniyle atmosferde sera etkisi oluşturan ve insan sağlığı bakımından zehirleyici olan karbondioksit gazı emisyonu, katı yakıtlara göre 1/3 ve sıvı yakıtlara göre 1/2 oranındadır.

1.1.2 Doğalgaz Tüketiminin Önemi

Doğalgaz kullanımı itibariyle hem kömürden hem de petrolden çok daha kolaydır. Diğer enerji kaynaklarına göre çok daha temiz olan ve çevreye kirlilik yaymayan doğal gaz, devletler için önemli bir enerji kaynağı haline gelmiştir. Enerjinin çağdaş uluslararası sistemdeki yeri ve insanların gündelik yaşamındaki önemi çok büyüktür. İnsanların evlerindeki kullandıkları elektrikten, sokaklarda kullandıkları arabaya kadar hepsi için enerji gerekmektedir. Dünya'mız üzerindeki enerji talebini çok fazla arttırmıştır. Ancak talep ve arz arasındaki dengesizlik bu enerjinin paylaşımı konusunda bazı sorunlara neden olmaktadır. Talep ve arz arasında yaşanan dengesizliğin en önemli ürünleri ise yine petrol, doğalgaz ve kömürdür. 2023 yılında bile bu üç enerji kaynağı, dünya enerji talebinin %80'ini karşılamaktadır. Bu kaynakların Dünya çapındaki enerji ihtiyacının %80'inini karşılamasının yanında, enerji kaynaklarının kıtlığını da eklersek bu kaynakların kullanım amaçlarının ne kadar önemli olduğu görülmektedir. Son zamanlarda yaşanan Rusya-Ukrayna arasındaki gerginliğin ülkemizdeki doğalgaz kullanımını büyük ölçüde etkilemiştir.

1.1.3 Kontrollü Doğalgaz Tüketimi

Doğal gaz, sınırlı bir kaynak olduğu için sürdürülebilir bir şekilde kullanılmalıdır. Kontrollü tüketim, doğal gazın gelecek nesillere de sağlanabilmesini ve kaynakların sürdürülebilir bir şekilde kullanılmasını sağlar.

İyi Yalıtım: Ev veya işyerinin iyi yalıtılmış olması, ısı kaybını azaltarak doğalgaz tüketimini düşürebilmektedir. Pencere, kapı, çatı ve duvarların izole edilmiş olması ısı kayıpları minimuma indirir ve enerji verimliliğini artırır.

Enerji Verimli Isıtma Cihazları: Eski ve verimsiz ısıtma cihazları yerine, modern ve enerji verimli modeller kullanılmalıdır. Yeni nesil kazanlar, kombi sistemleri ve ısı pompaları daha az doğalgaz tüketir ve enerji verimliliğini artırır.

Sıcak Su Kullanımı: Sıcak su tüketimini kontrol altında tutmak önemlidir. Daha kısa duş almak, muslukları sıkıca kapatmak ve sıcak su kullanımını gereksiz yere azaltmak enerji tasarrufuna yardımcı olmaktadır.

Enerji Verimli Cihazlar: Doğalgazla çalışan cihazları seçerken enerji verimliliği sağlayan araçlar tercih edilmelidir.

Düzenli Bakım: Doğalgazla çalışan cihazların düzenli bakımını yaptırmak, performanslarını ve verimliliklerini artırır. Böylece, enerji israfını önler ve tüketimi kontrol altında tutar.

1.1.4 Doğalgaz Tüketimini Etkileyen Faktörler

Doğalgaz tüketimi birçok farklı faktörden dolayı değişim gösterebilir.

Hava koşulları: İklim şartları doğalgaz tüketimini büyük ölçüde etkiler. Soğuk hava koşullarında, özellikle kış aylarında, doğalgaz tüketimi artar çünkü evlerin ve işyerlerinin ısıtılması için daha fazla enerjiye ihtiyaç duyulur.

Bina izolasyonu: İyi bir bina izolasyonu, ısı kaybını önleyerek doğalgaz tüketimini azaltır. İyi yalıtımlı binalar, ısıyı daha iyi korur ve daha az enerjiyle ısınır, bu da doğalgaz tüketimini düşürmektedir.

Bina tipi: Bina tipi doğalgaz tüketimini etkiler. Örneğin, apartman daireleri genellikle daha az doğalgaz tüketirken, müstakil evler daha fazla doğalgaz kullanabilir.

Isıtma ve su ısıtma sistemleri: Kullanılan ısıtma ve su ısıtma sistemleri de doğalgaz tüketimini etkiler. Verimli ve modern sistemler, daha az doğalgaz kullanırken, eski ve verimsiz sistemler daha fazla doğalgaz tüketebilir.

Endüstriyel kullanım: Doğalgaz, endüstriyel faaliyetlerde geniş çapta kullanılır. İşletmelerin büyüklüğü, sektörü ve faaliyetleri doğalgaz tüketimini etkileyebilir.

Popülasyon yoğunluğu: Nüfus yoğunluğu da doğalgaz tüketimini etkiler. Yoğun nüfuslu bölgelerde daha fazla doğalgaz tüketimi olabilir.

Ekonomik faktörler: Ekonomik durum da doğalgaz tüketimini etkiler. Ekonomik büyüme ve gelir seviyelerindeki değişiklikler, doğalgaz tüketiminde farklılıklara neden olabilir.

Fiyatlar: Doğalgaz fiyatları, tüketicilerin tüketim düzeylerini etkileyen önemli bir faktördür. Düşük fiyatlar, tüketimi artırabilirken, yüksek fiyatlar tüketimi azaltabilir.

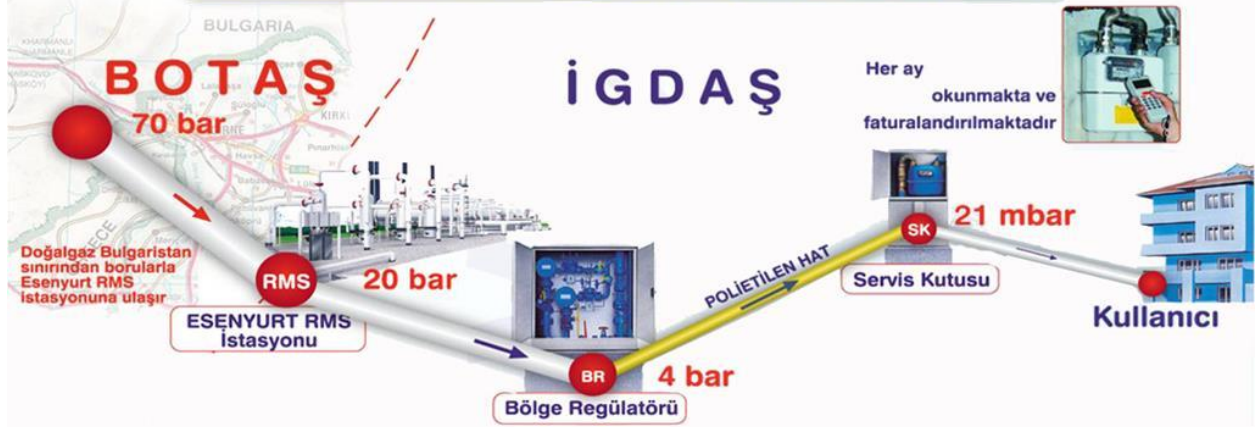
Teknolojik gelişmeler: Daha verimli doğalgaz yakıt sistemleri ve cihazlar, tüketim düzeylerini etkileyebilir. Gelişmiş teknolojilerle daha az doğalgaz kullanımı sağlanabilir.

Enerji politikaları ve teşvikler: Hükümetin enerji politikaları, teşvikler ve düzenlemeler, doğalgaz tüketimini etkileyebilir. Enerji verimliliğini artırmak veya alternatif enerji kaynaklarına yönlendirmek için yapılan politikalar, doğalgaz talebini değiştirebilir.

1.1.5 Doğalgaz'ın İstanbul Serüveni

Üst Isıl değeri maksimum 10.427 kcal/m³ minimum 8.100 kcal/m³ yoğunluğu yaklaşık olarak 0,72 kg/m³ olan doğal gaz, Rusya'dan çıkartıldıktan sonra, Ukrayna, Moldova, Romanya ve Bulgaristan üzerinden boru hatlarıyla Türkiye'ye taşınır. Malkoçlar Ölçüm ve Filtreleme İstasyonu'nda Türkiye'ye giren doğal gazın miktarının ölçülmesi ve filtrelenmesi işlemi yapılır. Malkoçlar ve Kırklareli Kompresör İstasyonları'nda basıncı 35 bara kadar düşmüş doğalgazın basıncı tekrar 70 bara çıkartılır. Marmara Ereğlisi LNG İstasyonu'nda Cezayir ve Nijerya gibi ülkelerden gelen sıvılaştırılmış doğalgaz (LNG), gazlaştırılarak sisteme gaz halinde verilir. Esenyurt'taki RMS İstasyonu'nda İstanbul'a giren doğalgazın basıncı 70 bardan 20 bar seviyesine düşürülerek doğalgazın filtrasyonu, ölçümlendirilmesi, gazın analizi, kokulandırılması ve ısıtılması işlemleri gerçekleştirilir. Buradan İstanbul-Avrupa yakasının doğalgaz dağıtımı yapılır. Marmara Denizi geçişi ile, yaklaşık 100 m derinlikte 106 km uzunluğunda 2 adet 30 inçlik boru ile doğalgaz. Avcılar Ambarlı'dan Pendik sahiline çıkarak İstanbul Anadolu yakasına taşınır.

KAYNAKTAN EVE DOĞALGAZIN SERÜVENİ



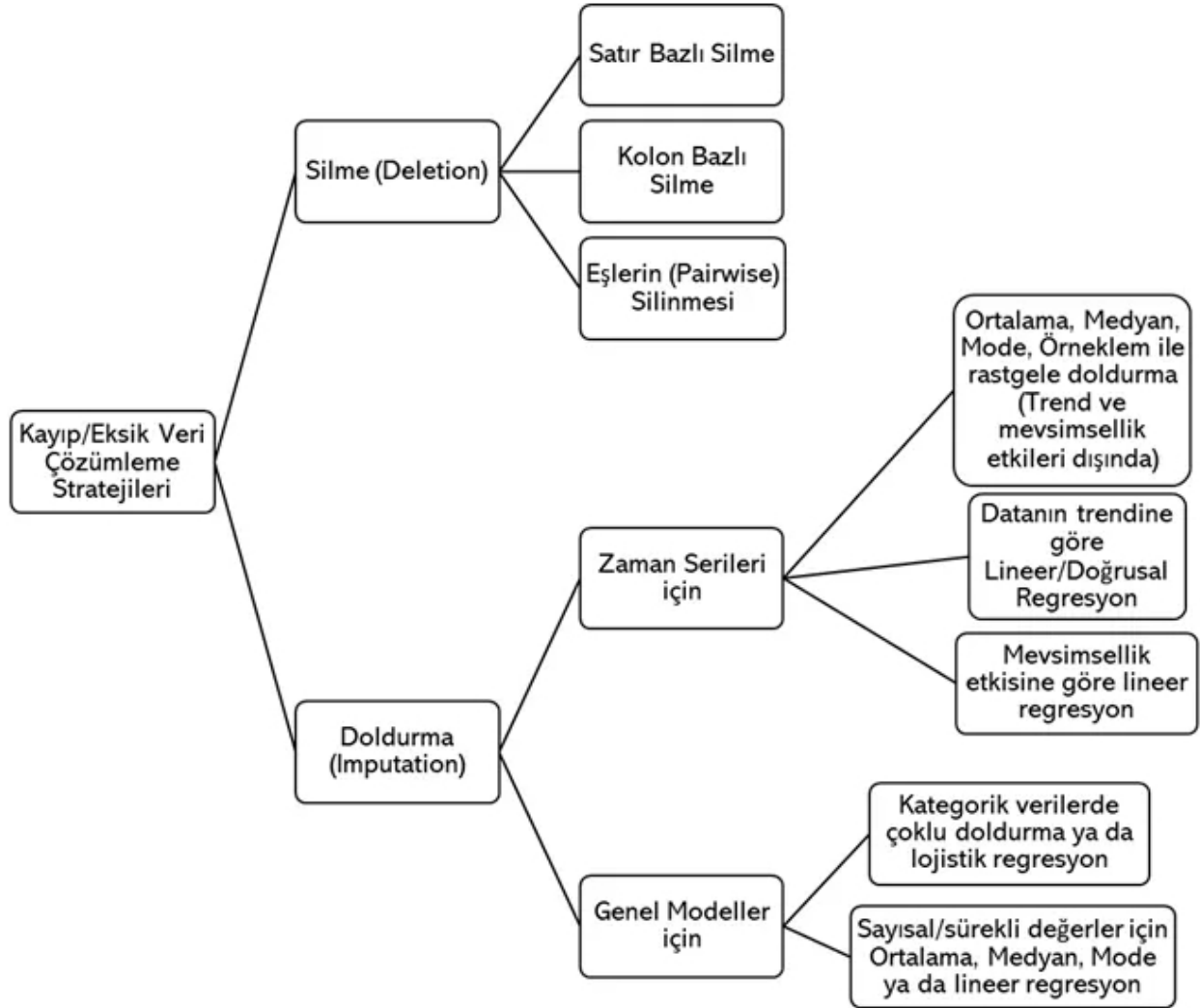
BÖLÜM 2

2.1 Kullanılan Yöntemler

Bu bölümde tüketim tahmini için kullanılan istatistiki teknikler açıklanacaktır.

2.1.1 Eksik Gözlemlerin Kontrolü

Eksik gözlem, veri setindeki herhangi bir hücrede değer olmamasıdır. Bu sorunla karşılaşıldığında aşağıdaki tabloda verilen çözümleme stratejileri uygulanmaktadır.

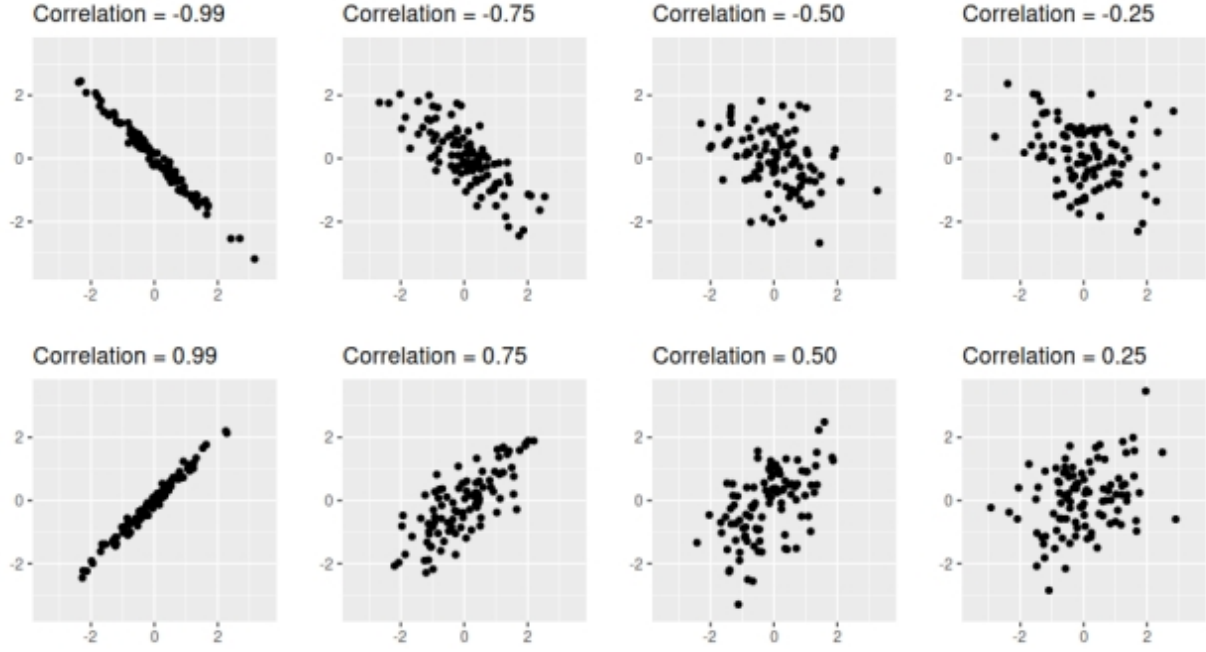


2.1.2 Korelasyon

İki değişken arasındaki ilişkinin gücünü ölçmek için korelasyon katsayılarının hesaplanması yaygındır. x ve y değişkenleri arasındaki korelasyon;

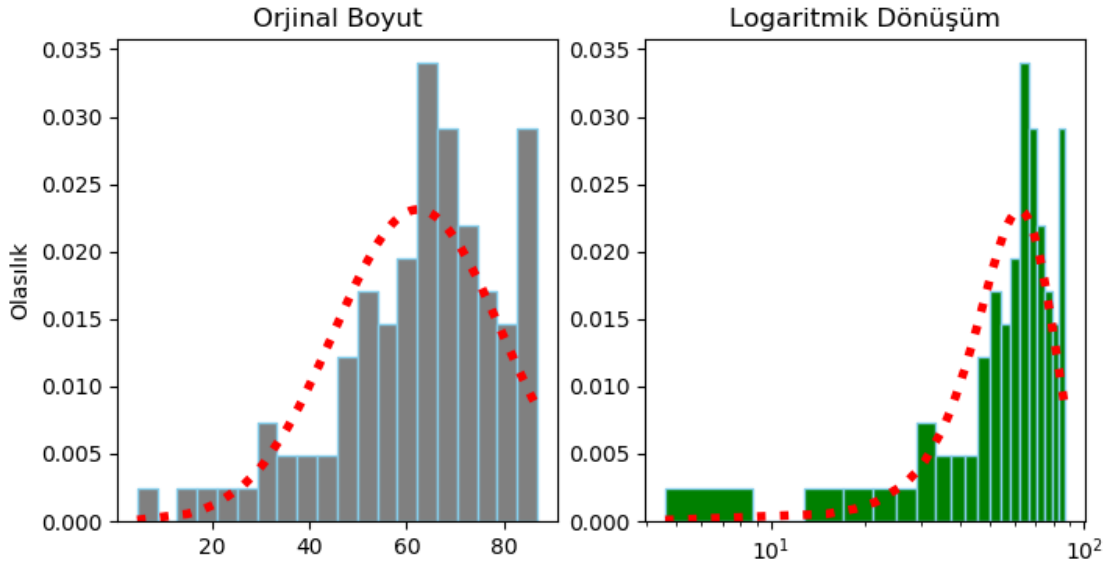
$$r = \frac{\sum (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum (x_t - \bar{x})^2} \sqrt{\sum (y_t - \bar{y})^2}}$$

Korelasyon katsayısı olan “r” her zaman -1 ile 1 arasındadır; negatif değerler negatif bir ilişkiyi belirtir ve pozitif değerler pozitif bir ilişkiyi gösterir. Aşağıdaki şekilde grafikler, farklı korelasyon seviyelerine sahip veri setlerinin örneklerini göstermektedir.



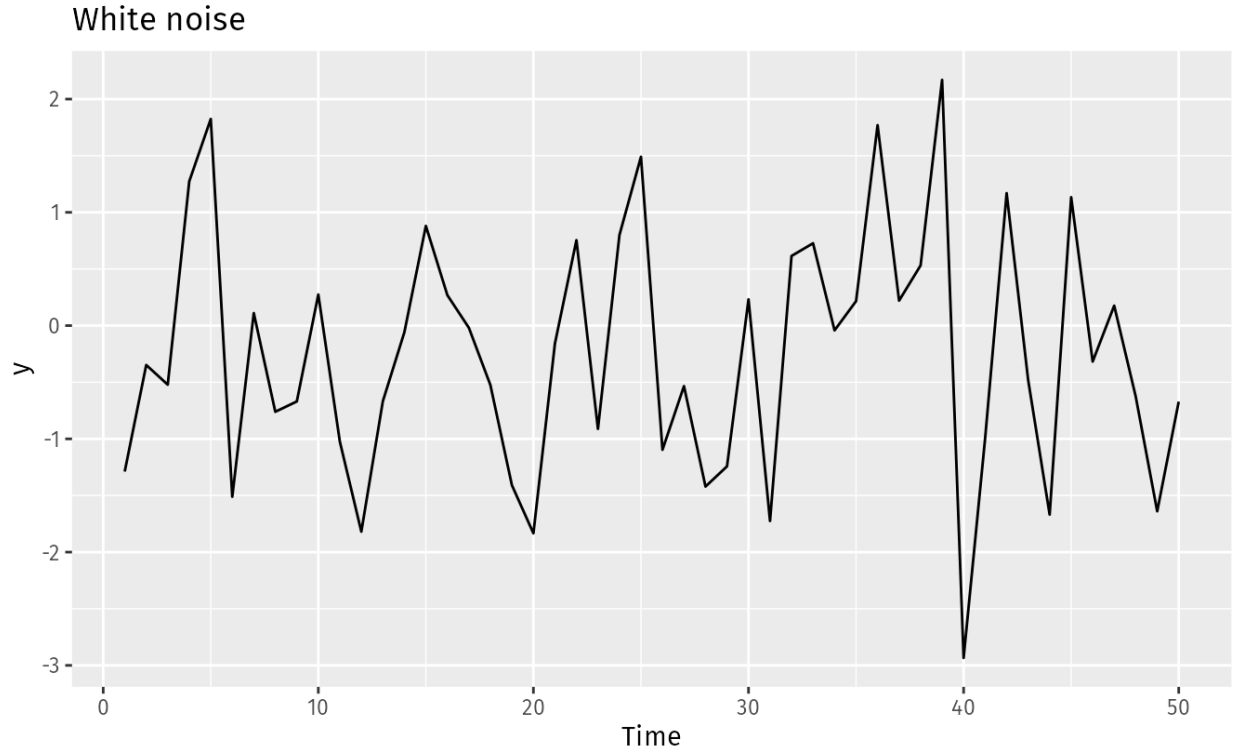
2.1.3 Logaritmik Dönüşüm

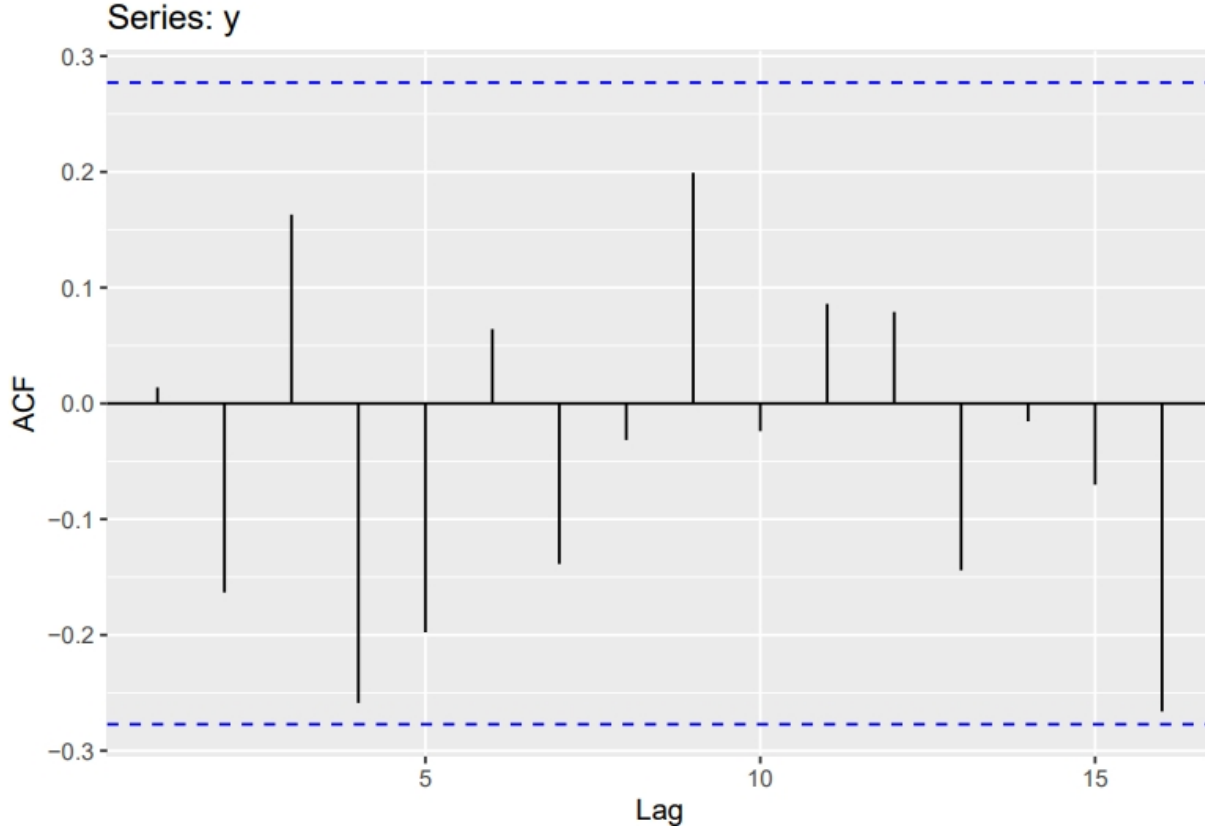
Logaritmik dönüşüm, her x değişkenini bir $\log(x)$ ile değiştirdiği bir veri dönüştürme yöntemidir. Doğal logaritma “ln” olarak gösterilir. Logaritmik dönüşümün amacı, verilerin normal dağılım özelliklerini iyileştirmek veya değişkenler arasındaki ilişkiyi düzeltmektir. Pozitif değere sahip değişkenler üzerinde uygulanır. Logaritmik dönüşüm, özellikle sağa çarpık dağılımlara sahip verilerde yaygın olarak kullanılır. Sağa çarpık bir dağılım, ortalamadan daha büyük değerlere doğru uzayan bir kuyrukla karakterizedir. Bu tür verilerde logaritmik dönüşüm uygulandığında, veri noktaları daha simetrik bir dağılıma yaklaşır ve analizler için daha uygun hale gelir.



2.1.4 Beyaz Gürültü

Otokorelasyon göstermeyen zaman serilerine beyaz gürültü denir. Beyaz gürültü serileri için her bir otokorelasyonun sıfıra yakın olması beklenmektedir.





Beyaz gürültü serisi için, ACF'deki ani artışların % 95'inin $\pm 2 / \sqrt{T}$ aralığında olmasını bekliyoruz; burada T, zaman serisinin uzunluğudur. Bu sınırları ACF'nin bir grafiğinde (yukarıdaki mavi kesikli çizgiler) çizmek yaygındır. Bir veya daha fazla büyük sivri uç bu sınırların dışındaysa veya sivri uçların önemli ölçüde % 5'inden fazlası bu sınırların dışındaysa, seri muhtemelen beyaz gürültü değildir.

2.1.5 Dinamik Regresyon

Tek değişkenli zaman serileri analizinde başka değişkenlerin katkısı ile bir model kurmaya izin veren ARIMA modellerinin genişletilmiş halidir. Dinamik regresyon modelleri, genellikler otoregresif (AR) veya hareketli ortalama (MA) terimlerini içeren zaman serisi modelleridir. AR terimleri, geçmiş dönem değerlerinin mevcut dönem değerini tahmin etmede kullanılmasını sağlar. MA terimleri ise, geçmiş hataların mevcut dönem değerini tahmin etmede kullanılmasını sağlar. Bu terimler, modelin zaman içindeki değişkenlikleri ve trendleri yakalamasına yardımcı olur.

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \varepsilon_t,$$

y_t : Mevcut dönemde tahmin edilmek istenen bağımlı değişkenin değerini temsil eder.

$x_{1,t}, \dots, x_{k,t}$: Mevcut dönemdeki bağımsız değişkenleri temsil eder.

$\beta_0 + \beta_1 + \dots + \beta_k$: Regresyon katsayılarıdır. Modelin tahmin edilen değişken üzerindeki etkisini gösterir.

ε_t : Hata terimidir. Gözlemlenen değer ile tahmin edilen değer arasındaki farkı temsil eder.

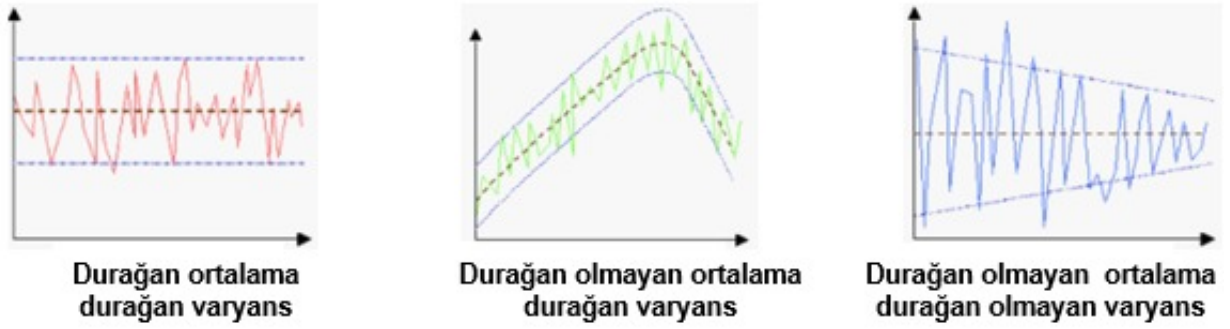
2.1.6 Durağanlık ve Fark Alma

Durağan olmayan serilerle analiz, tahmin ve öngöründe bulunmak tutarsız ve yanıltıcı sonuçlara neden olur. Zaman serilerini modelleyebilmek için serilerin durağan olması, değilse durağanlaştırılması gerekir. Zaman

serisi verilerinin belirli bir zaman sürecinde sürekli artma veya azalmanın olmadığı, verilerin zaman boyunca bir yatay eksen boyunca saçılım gösterdiği biçimde tanımlanır. Genel bir tanımlama ile, sabit ortalama, sabit varyans ve seriye ait ikideğer arasındaki farkın zamana değil, yalnızca iki zaman değeri arasındaki farka bağlı olması şeklinde ifade edilir. Kısaca; Bir zaman serisinde peş peşe gelen iki veri arasındaki fark, zamanın kendisinden kaynaklanmamakta sadece zaman aralığından kaynaklanmakta ise durağanlık söz konusudur. Bu nedenle, trendleri olan veya mevsimsellik içeren zaman serileri durağan değildir denir.

Eğer bir zaman serisinin ortalaması, varyansı ve kovaryansı zaman boyunca sabit kalıyorsa, serinin durağan olduğu söylenebilir. Durağanlık durumu farklı testler ile test edilebilir. Bunların yanı sıra ACF ve PACF grafiklerine bakmanın da durağanlık konusunda karar almada etkileri vardır.

Fark alma ise durağanlığın sağlanmadığı zamanlarda, uygulanan işlemlerden birisidir. Zaman serisinde peşpeşe gelen verilerden son değerden bir önceki değerin çıkarılması ile yapılır. Fark alma işlemi ile trend ve mevsimsel dalgalanmalar yok edilebilmektedir. Ayrıca serideki değişimin yönünü de görebilmeye fırsat verir. (pozitif veya negatif yönü)



2.1.7 SARIMA

Mevsimsel eğilimleri modelleyebilmemizi sağlayan SARIMA modeli, ARIMA modelinin uyarlanmış bir versiyonudur..

SARIMA modeli, ARIMA modelindeki mevsimsel olmayan bileşenlere (p,d,q) ek olarak, mevsimsel bileşenlere (P,D,Q,s) sahiptir.

P: Mevsimsel AR (Autoregressive) modelinin otoregresif terim sayısıdır.

D: Mevsimsel I (Integrated) modelinin fark derecesidir.

Q: Mevsimsel MA (Moving Average) modelinin hareketli ortalama terim sayısıdır.

s: Mevsimsel periyot uzunluğunu temsil eder.

$$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d(Y_t - \mu) = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)(1 - B)^D \epsilon_t$$

$Y(t)$: Zaman serisinin t zamanındaki gözlem değerini temsil eder.

B : Gecikme operatörünü simgeler.

μ : Zaman serisinin ortalamasını ifade eder.

$\phi_1, \phi_2, \dots, \phi_p$: Otoregresif (AR) terimlerin katsayılarıdır.

$\theta_1, \theta_2, \dots, \theta_q$: Hareketli ortalama (MA) terimlerin katsayılarıdır.

d : Zaman serisinin fark derecesidir.

D : Mevsimsel fark derecesidir.

ϵ_t : Zaman serisinin hatalarını temsil eder.

2.1.8 ARIMA

ARIMA modeli, zaman serilerindeki trendleri, mevsimsellikleri ve durağanlık özelliklerini yakalamak için kullanılır. ARIMA modeli üç bileşenden oluşur:

AR (AutoRegressive) bileşeni: Bu bileşen, geçmiş gözlemlerin lineer kombinasyonu ile mevcut gözlemi tahmin etmeye çalışır. AR bileşeni, bir zaman serisindeki geçmiş değerlerin etkisini modellemek için otoregresif terimler kullanır.

I (Integrated) bileşeni: Bu bileşen, zaman serisinin durağanlık düzeyini yakalamak için gerekli olan fark operasyonlarını ifade eder. İntegrasyon, zaman serisindeki trendin kaldırılmasını sağlar ve veriyi durağan hale getirir.

MA (Moving Average) bileşeni: Bu bileşen, geçmiş tahmin hatalarının lineer kombinasyonu ile mevcut gözlemi tahmin etmeye çalışır. MA bileşeni, bir zaman serisindeki geçmiş tahmin hatalarının etkisini modellemek için hareketli ortalama terimlerini kullanır.

ARIMA(p, d, q) formülüyle gösterilen bir ARIMA modelinde:

p: AR bileşenin derecesini temsil eder. AR bileşeninde kaç adet otoregresif terim olduğunu belirtir.

d: İntegrasyon derecesini temsil eder. İntegrasyon, trendin kaldırılması için kaç defa fark alındığını belirtir.

q: MA bileşenin derecesini temsil eder. MA bileşeninde kaç adet hareketli ortalama terimi olduğunu belirtir.

$$Y(t) = c + \phi_1 Y(t-1) + \phi_2 Y(t-2) + \dots + \phi_p Y(t-p) - \theta_1 \varepsilon(t-1) - \theta_2 \varepsilon(t-2) - \dots - \theta_q \varepsilon(t-q) + \varepsilon(t)$$

$Y(t)$: Mevcut zaman noktasındaki gözlem değeridir.

c : Sabit terimdir.

$\phi_1, \phi_2, \dots, \phi_p$: AR terimlerinin katsayılarıdır.

$\varepsilon(t-1), \varepsilon(t-2), \dots, \varepsilon(t)$: MA terimlerinin hatalarıdır.

R'deki `auto.arima()` fonksiyonu, bir ARIMA modeli elde etmek için birim kök testlerini, AICc ve MLE'nin en aza indirilmesini birleştiren Hyndman-Khandakar algoritmasının (Hyndman & Khandakar, 2008) bir varyasyonunu kullanır. `auto.arima()` opsiyonları, algoritmanın birçok varyasyonunu sağlar.

Otomatik ARIMA modellemesi için Hyndman-Khandakar algoritması şu şekildedir:

1) Önce birim kök testleri uygulanarak kaç fark alınacağı belirlenir. ($0 \leq d \leq 2$)

2) Daha sonra AICc yi minimum kılacak olan model ve parametre tahminleri belirlenir. Bu minimizasyon yapılırken arama uzayını daraltabilmek için algoritma adimsal bir yaklaşım kullanır.

a) 4 başlangıç modeli fit edilir.

ARIMA(0,d,0)

ARIMA(2,d,2)

ARIMA(1,d,0)

ARIMA(0,d,1)

$d=2$ durumu dışında modeller sabit terim içerir. $d \leq 1$ ise ayrıca sabit terim içermeyen ARIMA(0,d,0) modeli de fit edilir.

b) a adımında belirlenen modellerden en küçük AICc değerine sahip olan model mevcut model olarak seçilir.

c) mevcut modelin çeşitli varyasyonları denenir.

2.1.9 AUTO ARIMA

R'deki `auto.arima()` fonksiyonu, bir ARIMA modeli elde etmek için birim kök testlerini, AICc ve MLE'nin en aza indirilmesini birleştiren Hyndman-Khandakar algoritmasının (Hyndman & Khandakar, 2008) bir varyasyonunu kullanır. `auto.arima()` opsiyonları, algoritmanın birçok varyasyonunu sağlar.

Otomatik ARIMA modellemesi için Hyndman-Khandakar algoritması şu şekildedir:

- 1) Önce birim kök testleri uygulanarak kaç fark alınacağı belirlenir. ($0 \leq d \leq 2$)
- 2) Daha sonra AICc yi minimum kılacak olan model ve parametre tahminleri belirlenir. Bu minimizasyon yapılırken arama uzayını daraltabilmek için algoritma adimsal bir yaklaşım kullanır.

a) 4 başlangıç modeli fit edilir.

ARIMA(0,d,0)

ARIMA(2,d,2)

ARIMA(1,d,0)

ARIMA(0,d,1)

d=2 durumu dışında modeller sabit terim içerir. $d \leq 1$ ise ayrıca sabit terim içermeyen ARIMA(0,d,0) modeli de fit edilir.

b) a adımı belirlenen modellerden en küçük AICc değerine sahip olan model mevcut model olarak seçilir.

c) mevcut modelin çeşitli varyasyonları denenir.

2.1.10 Decomposition

Bir şeyin parçalara ayrılması veya dağılması anlamına gelir. İstatistiksel yöntemlerde, karmaşık bir sorunu daha küçük, daha yönetilebilir parçalara bölme sürecini ifade eder. Bu, problemin çözülmesini daha kolay hale getirebilir ve daha verimli bir şekilde programlama veya tasarım yapmayı sağlayabilir. Bir zaman serisinin trend, mevsimsellik ve düzensizlik bileşenlerine ayrıştırılması sürecini ifade eder. Bu süreç, zaman serisindeki yapısal örüntüleri anlamak ve analiz etmek için kullanılır.

2.1.11 Toplamsal Ayrıştırma Yöntemi

Trend, mevsimsellik ve hata bileşenlerinin toplanmasıyla oluşup şu şekilde formüle sahip modeldir:

$$Z_t = T_t + M_t + \varepsilon_t$$

Burada hata teriminin(ε_t) rasgele bir harekete sahip olması gerektiğine dikkat edilmelidir.

2.1.12 Çarpımsal Ayrıştırma Yöntemi

rend, mevsimsellik bileşenlerinin çarpılması ve hata bileşeninin toplanmasıyla oluşup şu şekilde formüle sahip modeldir:

$$Z_t = T_t * M_t + \varepsilon_t$$

Bu yöntemin uygulanışı bazı noktalarda toplamsal ayrıştırma yönteminden farklılık gösterir ama temel mantık olarak bu iki model arasında pek bir farklılık yoktur.

$Z(t)$: Zaman serisinin t zamanındaki gözlem değerini temsil eder. Yani, t zamanındaki gerçek veri noktasını ifade eder.

$T(t)$: Trend bileşenini temsil eder. Trend, zaman serisinin genel eğilimini ifade eder. Trend bileşeni, zamanla değişen düzenli bir modeli yansıtır.

$M(t)$: Mevsimsellik (seasonality) bileşenini temsil eder. Mevsimsellik, belirli bir periyotta tekrar eden yapısal desenleri yansıtır. Bu desenler genellikle yılın mevsimleri veya aylar gibi düzenli periyotlarda ortaya çıkar.

ε_t : Hata veya düzensizlik bileşenini temsil eder. Hata bileşeni, trend ve mevsimsellik bileşenlerinin dışında kalan rastgele varyasyonları ifade eder. Bu bileşen, zaman serisindeki beklenmeyen olaylar, geçici etkiler veya diğer faktörler nedeniyle meydana gelen dalgalanmaları temsil eder.

2.1.13 Hareketli Ortalamalı Modeller

Hareketli ortalamalı modeller, zaman serisindeki düzenli desenleri ve trendleri yakalamak için hareketli ortalamaları kullanır. Hareketli ortalama, belirli bir zaman dilimindeki değerlerin ortalamasını hesaplayan bir yöntemdir. Bu ortalama, zaman serisinin düzeltilmesini sağlar ve gürültülü veri noktalarının etkisini azaltır. Hareketli ortalamalı modeller, trend ve mevsimsellik desenlerini yakalamak ve düzensizlik bileşenini azaltmak için kullanılır. Yaygın olarak kullanılan hareketli ortalamalı modeller arasında Basit Hareketli Ortalama (Simple Moving Average - SMA) ve Ağırlıklı Hareketli Ortalama (Weighted Moving Average - WMA) bulunur.

Basit Hareketli Ortalama: Her veri noktasını eşit bir ağırlıkla değerlendirir. Belirli bir zaman dilimindeki değerlerin toplamının, o zaman dilimindeki veri noktalarının sayısına bölünmesiyle elde edilir. Zaman serisinin düzeltilmesini sağlar ve gürültülü veri noktalarının etkisini azaltır. Ayrıca, trendi belirlemek için kullanılabilir. SMA'nın kullanılmasıyla, zaman serisi daha yumuşak bir şekilde görüntülenir ve daha uzun vadeli eğilimlerin belirlenmesine yardımcı olur.

$$SMA(t) = X(t) + X(t-1) + X(t-2) + \dots + X(t-n+1)/n$$

$SMA(t)$: t zamanındaki basit hareketli ortalama değeridir.

$X(t)$: t zamanındaki gözlem değeridir.

n : Kullanılan zaman aralığını belirtir.

Ağırlıklı Hareketli Ortalama: Bu yöntem, zaman serisindeki gözlem değerlerine ağırlık vererek ortalama hesaplar.

$$WMA(t) = W1 * X(t) + W2 * X(t-1) + \dots + Wn * X(t-n+1)/(w1 + w2 + w3 + \dots Wn)$$

$WMA(t)$: t zamanındaki ağırlıklı hareketli ortalama değeridir.

$X(t)$: t zamanındaki gözlem değeridir.

$w1, w2, w3, \dots, wn$: Her gözlem değeri için kullanılan ağırlıkları temsil eder.

n : Kullanılan zaman aralığını belirtir.

2.1.14 Eğilim Etkisi

Verilerin zamanla değişmesini veya trend göstermesi eğilimini ifade eder. Verilerin belirli bir yönde sürekli artması, azalması veya durağanlık eğilimine sahip olması durumunda ortaya çıkar. Eğilim etkisi, zaman serisinin genel davranışını ve gelecekteki trendlerin tahmin edilmesinde önemli rol oynar. Eğilim, genelde verideki uzun vadeli değişimleri ifade eder ve bu değişim, verinin doğrusal veya doğrusal olmayan bir modelle ifade edilebilir. Doğrusal eğilim durumunda, veriler düz bir çizgi boyunca artar veya azalır. Doğrusal olmayan

eğilim durumunda ise, verilerin artış veya azalış hızı zaman içerisinde değişebilir. Eğilim etkisini anlamak, gelecekteki değerleri tahmin etmek ve zaman serisinin davranışını analiz etmek için önemlidir. Eğilimi dikkate alarak yapılan tahminler, trendin sürekliliğini ve gelecekteki olası değişiklikleri hesaba katmış olur. Eğilim çizgisi doğrusal regresyon analiziyle hesaplanabilmektedir. Doğrusal regresyon, zaman serisindeki eğilimi en iyi şekilde temsil eden doğruyu bulmayı hedeflemektedir.

Doğrusal regresyon analizi için genel bir formül:

$$Y = a + b * X$$

Y : Zaman serisi verilerini temsil eden gözlemlenen değerler vektörüdür.

X : zamanın geçmesine karşılık gelen bağımsız değişken vektörüdür.

a : Doğrusal regresyon doğrusunun kesme noktasını (başlangıç noktasını) temsil eden sabittir.

b : Doğrusal regresyon doğrusunun eğimini (eğilim etkisini) temsil eden katsayıdır.

Not: Zaman serilerindeki eğilim etkisini daha karmaşık modeller ile de bulabiliriz. Doğrusal olmayan eğilimleri yakalamak için polinomik regresyon, lojistik regresyon veya zaman serisi analizi gibi yöntemler kullanılabilir.

2.1.15 Sezonallık Etkisi

Mevsimsellik periyodik dalgalanmaları ifade eder. Mevsimsel bir seri, N 'nin belirli bir değeri için her N periyodunu tekrarlayan bir desene sahip olan bir seridir. Desenin sezonun uzunluğu olarak tekrarlanmaya başlamasından önceki dönemlerin sayısını belirtir. Mevsimsel modeli kullanmak için, Sezonun uzunluğunu belirtebilir. Zaman serilerinde mevsimsellikten bahsederken sadece mevsimsel etkilerden kaynaklanan değişikliklerden bahsedemeyiz elbetteki. Dondurma satışlarının yazın yüksek oluşu dışında, Blackfriday gibi her yıl kasım ayının belirli günlerinde yapılan kampanyalar da mevsimsel dalgalanmalara yol açabilir.

Sabit Seriler(Stationary) İçin Mevsimsel Faktörler:

1. Tüm verilerin örnek ortalamasını hesaplayın.
2. Her bir gözlemi örnek ortalamasına bölün. Bu, gözlemlenen verilerin her dönemi için mevsimsel faktörler verir.
3. Her mevsimdeki benzer dönemler için ortalama faktörler, yani bir sezonun ilk dönemine karşılık gelen tüm faktörler, sezonun ikinci dönemine karşılık gelen tüm faktörler ve diğerleri.

Sezonallık etkisinin tahmini için aşağıdaki formül kullanılır:

$$Y(t) = T(t) + S(t) + E(t)$$

$Y(t)$: zaman serisindeki gözlemlenen değerlerin vektörüdür.

$T(t)$: trend bileşenini ifade eden fonksiyondur. Trend, zaman serisinin genel eğilimini temsil eder.

$S(t)$: sezonsal bileşeni ifade eden fonksiyondur. Sezonsal bileşen, periyodik desenleri veya tekrarlanan modelleri temsil eder.

$E(t)$: hata veya rastgele bileşeni. Trend ve sezonsal bileşenin dışındaki diğer faktörleri ve rastgele değişimleri yakalar.

Bazı durumlarda, çarpansal (multiplicative) bir model de kullanılabilir, bu durumda formül şu şekilde ifade edilir:

$$Y(t) = T(t) * S(t) * E(t)$$

2.1.16 Düzensiz Hareketler

üzensiz(beklenmedik) hareketler,zaman serisi verilerinin unsurlarından biridir. Düzensiz hareketler, tesadüfi bir değişken niteliğindedir. Uzun dönem eğilimi, devresel hareketler ve mevsimlik dalgalanmalar zaman serisi verilerinden uzaklaştırıldığında geriye kalan kısım “düzensiz (beklenmedik) değişim” dir. Bu dalgalanmalar, grev, deprem, fırtına, sel vb gibi tahmin edilemeyen belirsiz olayların gelişmesinden kaynaklanmaktadır. Bu tip belirsiz olaylar malların üretimi ve dağıtımı gibi birçok şeyi etkilemektedir. Düzensiz hareketler, belirsiz olaylar olduğundan bunu tahmin etmek mümkün değildir. Bununla birlikte, düzensiz hareketlerin genellikle zaman serisi analizindeki modellere eklenen hata terimi veya rastgele bileşen tarafından temsil edildiğini söyleyebiliriz. Bu hata terimi, trend veya sezonsallıkla ilişkili olmayan, sistemli olmayan ve tahmin edilemeyen varyasyonları yakalar. Genel olarak, bir zaman serisindeki gözlemlenen değerler, trend bileşeni, sezonsal bileşen ve hata terimi (düzensiz hareketler) olmak üzere üç bileşene ayrılabilir:

$$Y(t) = T(t) + S(t) + E(t)$$

$Y(t)$: zaman serisindeki gözlemlenen değerlerin vektörüdür.

$T(t)$: trend bileşenini ifade eden fonksiyondur.

$S(t)$: sezonsal bileşeni ifade eden fonksiyondur.

$E(t)$: hata veya rastgele bileşeni ifade eden fonksiyondur.Düzensiz hareketleri yakalar ve tahmin edilemeyen dalgalanmaları temsil eder.

2.1.17 ARIMA İçin Forecasting

ARIMA hataları içeren bir regresyon modeli kullanarak forecasting yapmak için, modelin regresyon kısmını ve modelin ARIMA kısmını tahmin etmemiz ve sonuçları birleştirmemiz gerekir. Sıradan regresyon modellerinde olduğu gibi, forecastleri elde etmek için önce açıklayıcı değişkenleri tahmin etmemiz gerekir. Açıklayıcı değişkenler gelecekte bilindiğinde (örneğin, zaman, haftanın günü vb. Takvimle ilgili değişkenler), bu basittir. Ancak açıklayıcı değişkenlerin gelecek değerleri bilinmediğinde, onları ayrı ayrı modellemeliyiz veya her bir tahminci için gelecekteki varsayılan değerleri kullanmalıyız.

BÖLÜM 3

3.1 Veri Tanıtımı

Bu çalışma kapsamında doğalgaz tüketim verisi İstanbul ilinde hizmet veren İstanbul Gaz Dağıtım Sanayi ve Ticaret Anonim Şirketi (İGDAŞ) veritabanları üzerinden hazırlanmış ve İstanbul Büyükşehir Belediyesine iletilmiş olup, onlar tarafından halka açık olan İBB Veri Portalı sitesinden elde edilmiştir. Veri seti İstanbulun 39 ilçesi baz alınarak oluşturulmuştur. 2023 Yılı hariç, diğer bütün dönemlerde ilçelere ait 12 aylık harcama verileri mevcuttur.

Veri seti saf halinde 4 satırdan oluşmaktadır. Bunlar Yıl, Ay No, İlçe ve Doğalgaz Tüketim Miktarı (m3).

`Yıl` <- Doğalgazın harcandığı yıla ait bilgi.

`AyNo` <- Doğalgazın harcandığı yılın ay değerlerine ait bilgi.

`İlçe` <- Doğalgaz harcamalarının hangi ilçelere ait olduğunu temsil eden değişken.

`DogalgazTuketimMiktar(m3)` <- Metre küp cinsinden belirlenmiş dönemlere ait harcanan doğalgaz değerleri.

Veriler İGDAŞ tarafından 2015 Ocak ayı ile 2023 Nisan ayı arasında listelenmiştir.

3.2 Veri Analizi

```
## # A tibble: 3,900 x 4
##   Yıl 'Ay No' İlçe   'Dogalgaz Tüketim Miktarı (m3)'
##   <dbl>   <dbl> <chr>         <dbl>
## 1 2015     1 ADALAR      370334.
## 2 2015     2 ADALAR      334257.
## 3 2015     3 ADALAR      353314.
## 4 2015     4 ADALAR      286269.
## 5 2015     5 ADALAR      128613.
## 6 2015     6 ADALAR      132890.
## 7 2015     7 ADALAR      140175.
## 8 2015     8 ADALAR      141850.
## 9 2015     9 ADALAR      121670.
## 10 2015    10 ADALAR      204709.
## # i 3,890 more rows

## # A tibble: 10 x 4
##   Yıl 'Ay No' İlçe   'Dogalgaz Tüketim Miktarı (m3)'
##   <dbl>   <dbl> <chr>         <dbl>
## 1 2022     7 ZEYTİNBURNU  3823445.
## 2 2022     8 ZEYTİNBURNU  4246652.
## 3 2022     9 ZEYTİNBURNU  4474070.
## 4 2022    10 ZEYTİNBURNU  5745271.
## 5 2022    11 ZEYTİNBURNU  8260587.
## 6 2022    12 ZEYTİNBURNU 13614112.
## 7 2023     1 ZEYTİNBURNU 15374084.
## 8 2023     2 ZEYTİNBURNU 17116321.
## 9 2023     3 ZEYTİNBURNU 14844388.
## 10 2023     4 ZEYTİNBURNU  9465694.

##   Yıl   Ay No   İlçe
```

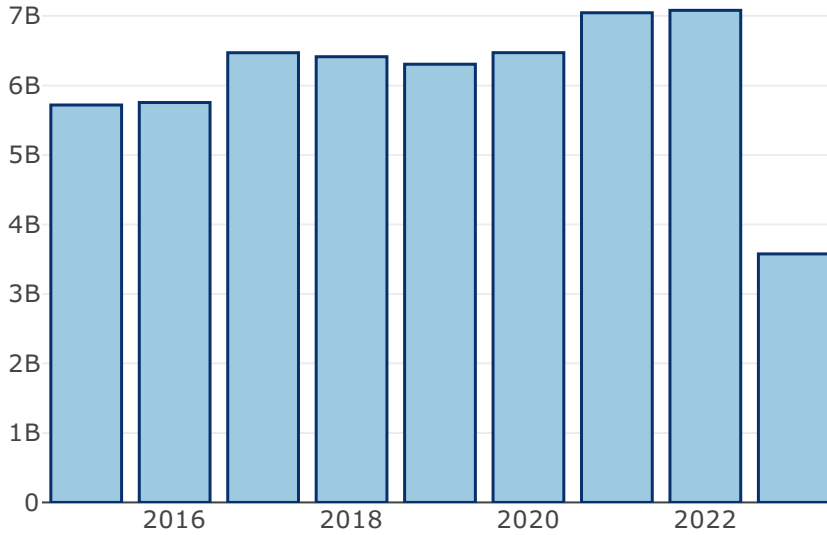
```

## Min. :2015 Min. : 1.00 Length:3900
## 1st Qu.:2017 1st Qu.: 3.00 Class :character
## Median :2019 Median : 6.00 Mode :character
## Mean :2019 Mean : 6.34
## 3rd Qu.:2021 3rd Qu.: 9.00
## Max. :2023 Max. :12.00
## Doğalgaz Tüketim Miktarı (m3)
## Min. : 28832
## 1st Qu.: 4361597
## Median : 9760357
## Mean :14063417
## 3rd Qu.:20994545
## Max. :79742863

```

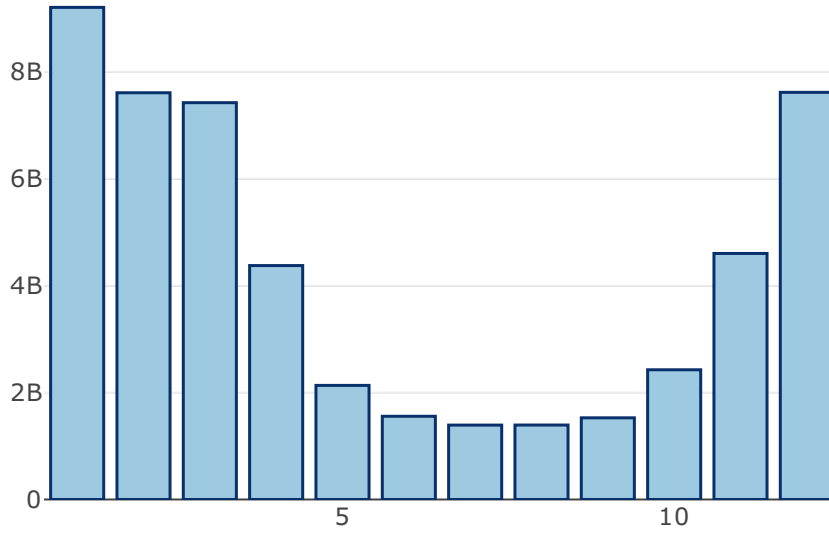
Verimizin ilk 10 ve son 10 gözleminden hangi tarihler arasında olduğu gösterilmiştir. Görüldüğü gibi veri seti 3900 satır ve 4 sütundan oluşmaktadır. Doğalgaz Tüketim değişkenimizin özelliklerini inceleyecek olursak, [28832, 79742863] değerleri arasında, 14063417 ortalamaya sahip bir değişkendir.

3.2.1 Yıllara Göre Toplam Doğalgaz Tüketiminin Bar Grafiği



Yıllara göre toplam tüketim miktarı incelendiği zaman 2022 yılında ve 2021 yılında en yüksek harcamalar olduğu söylenebilir. Bunun sebebi farklı konular olabilir. Sıcaklık, doğalgaz fiyatı, abone sayısı gibi konulardan ziyade bu yıllar için pandemi etkisi göz önünde bulundurulabilir. Daha detaylı düşünecek olursak, 2020 yılından sonra çalışma stili olarak remote çalışma sistemi daha çok tercih edilmeye başlamıştır ve bu durum insanların doğalgaz tüketim harcamasını etkilediği düşünülebilir. 2023 yılının ilk 4 ayı veri setinde dahil olduğu için diğer yıllara göre düşük bir değer gözükmektedir.

3.2.2 Aylara Göre Toplam doğalgaz Tüketiminin Bar Grafiği



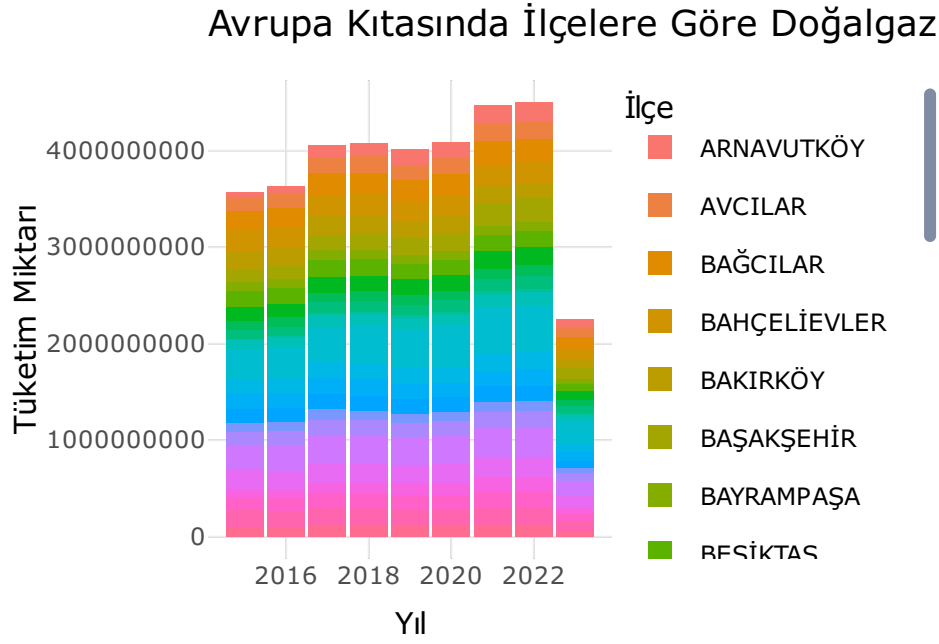
Yıllara göre 2023 yılı ayrı tutularak daha detaylı inceleyecek olarsak eğer aylara göre harcamalarda ilk 3 ay ve son 1 ayda yüksek doğalgaz tüketimi mevcuttur, bunun sebebi kış ayları olduğu için olabilir.

3.2.3 İstanbul'un İlçeleri İçin Toplam Doğalgaz Tüketiminin Harita Gösterimi



İstanbul geneli için 2015 yılından 2023 Nisan ayına kadar olan toplam tüketim verileri ile yapılan görselleştirmede, hangi ilçenin ne kadar çok harcama yaptığını sahip oldukları noktaların büyüklüğünden anlaşılabilir. Daha detaylı inceleme için imleç ile harita üzerine gidildiği zaman hangi ilçe ve ne kadar toplam tüketime sahip olduğu görülmektedir. Sağ tarafta bulunan ilçe listesinden ilçeleri çıkartıp ekleyebiliriz.

3.2.4 Avrupa Kıtası İçin Doğalgaz Tüketiminin Yıllara Göre Bar Grafiği



İlçeleri *Avrupa* ve *Asya* kıtasına ayırarak inceleyecek olursak, hangi kıta da hangi ilçe daha çok doğalgaz tüketim harcamasına sahip görülebilmektedir.

Yıllara göre ayrı ayrı Avrupa kıtası incelendiği zaman ilçelere göre bar grafiğinde en fazla alana sahip olan ilçe en yüksek tüketime sahip ilçedir denilebilir. En yüksek doğalgaz tüketimi 2022 yılında Esenyurt ilçesi olarak görülmektedir.

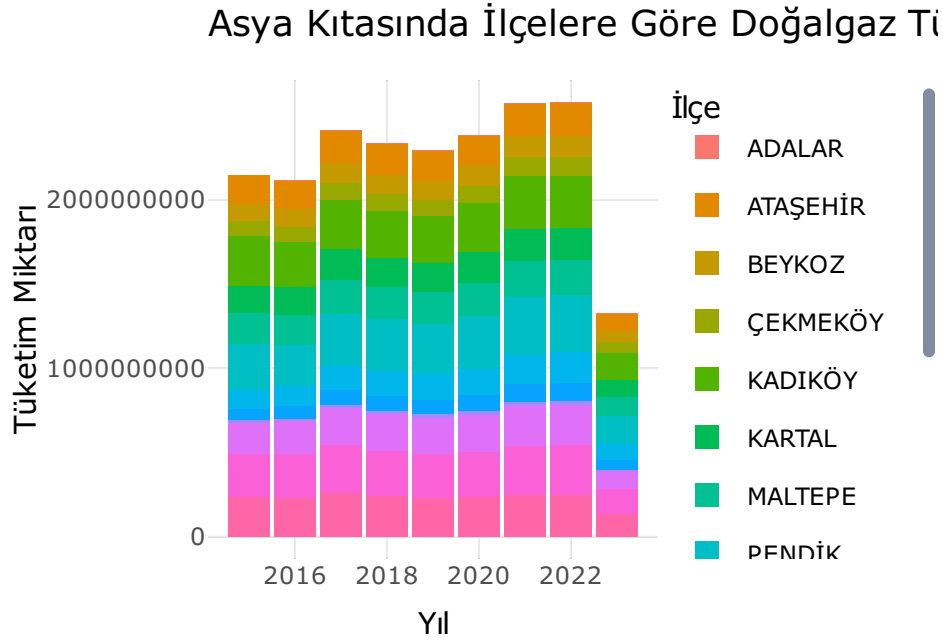
3.2.5 Avrupa Kıtası İçin Doğalgaz Tüketiminin Pasta Grafiği

Avrupa Kıtasında İlçelere Göre Doğalgaz Tüketim Miktarı



Genel olarak Avrupa kıtası için toplam doğalgaz harcamalarını inceleyecek olursak eğer, pasta grafiğine Esenyurt ilçesi %9.83'luk bir yer kaplayarak en yüksek doğalgaz harcamasına sahip ilçe olduğu tespit edilmiştir. En düşük ilçe ise %0,625'luk bir alan kaplayarak Çatalca ilçesi olmuştur.

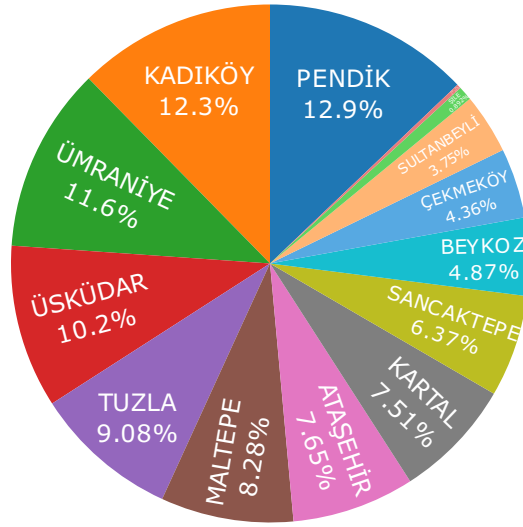
3.2.6 Asya Kıtası İçin Doğalgaz Tüketiminin Yıllara Göre Bar Grafiği



Asya kıtası için ilçeleri ayrı ayrı yıllara göre incelenecek olursa Kadıköy ve Pendik ilçesi en yüksek doğalgaz tüketimine sahip ilçelerdir. Daha detaylı inceleme için imleci grafiğin üzerine getirdiğimiz zaman en yüksek doğalgaz tüketiminini 2022 yılında Pendik ilçesi gerçekleştirmiştir.

3.2.7 Asya kıtası için doğalgaz tüketiminin Pasta Grafiği

Asya Kıtasında İlçelere Göre Doğalgaz Tüketim Miktarı



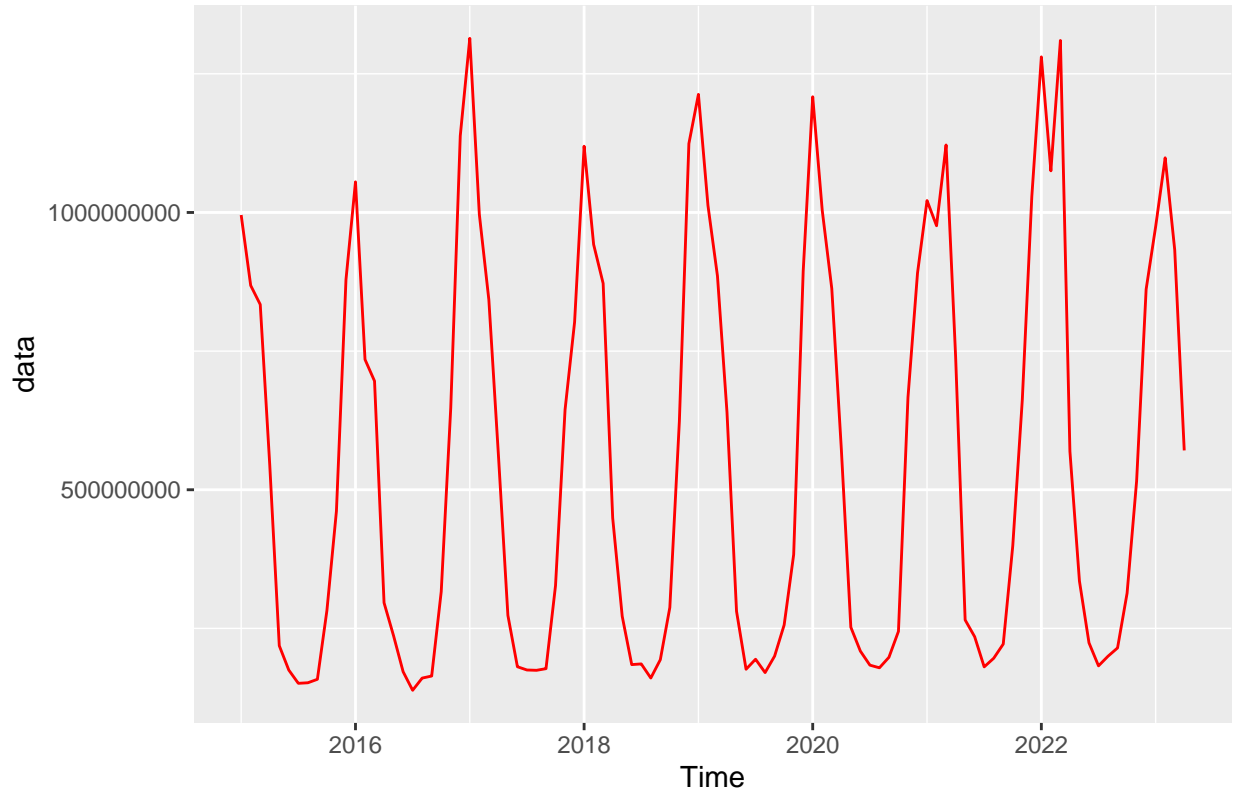
Genel olarak Asya kıtası için toplam doğalgaz harcamalarını incelenecek olursa, pasta grafiğine Pendik ilçesi %12.9'luk bir yer kaplayarak en yüksek doğalgaz harcamasına sahip ilçe olduğu tespit edilmiştir. En düşük ilçe ise %0,259'luk bir alan kaplayarak Adalar ilçesi olmuştur.

Asya kıtasındaki pasta dağılımı Avrupa kıtasına göre daha yüksek yüzdelik değerlere sahiptir, sebebi ise Asya kıtasında Avrupa kıtasına kıyasla daha az ilçe bulunmasından kaynaklıdır.

3.3 Modellemeye Hazırlık Aşamaları

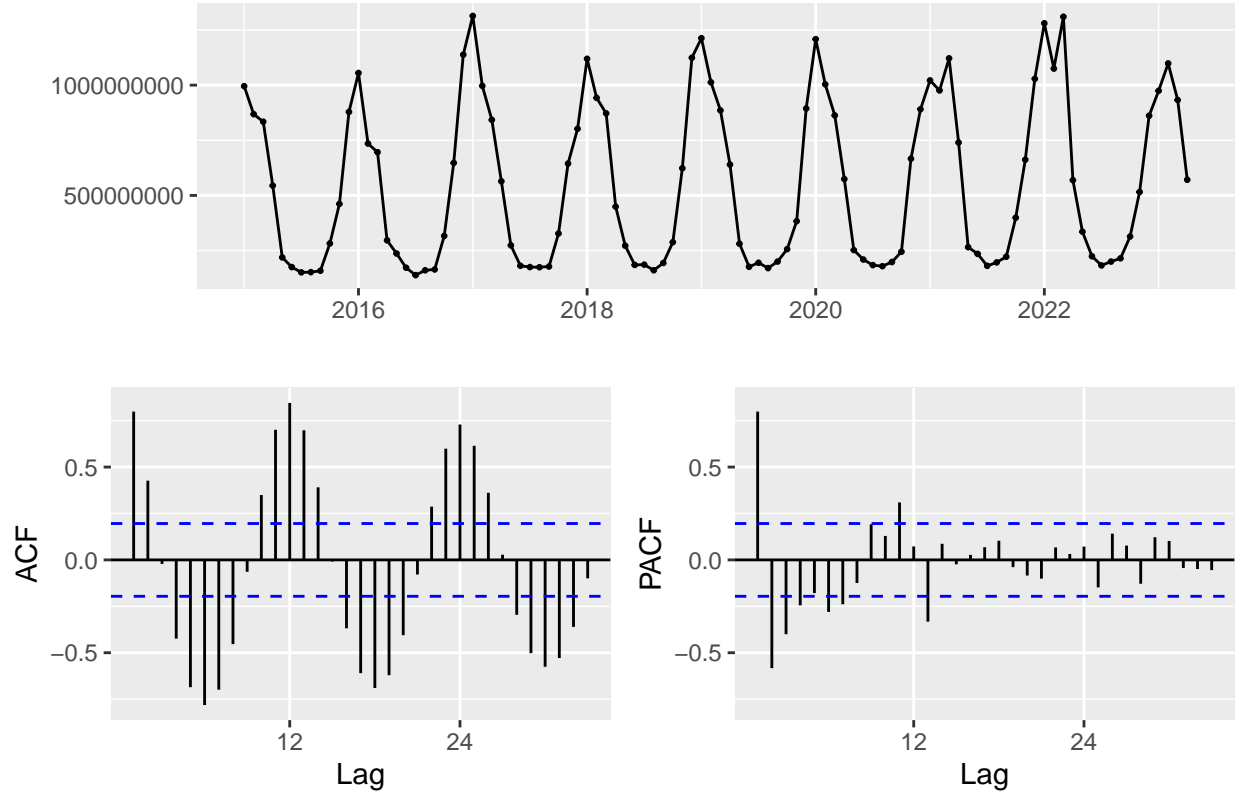
3.3.1 Zaman Serisi Nesnesine Dönüştürme İşlemi

Veri setimizi 2015'in ilk ayından başlayacak şekilde bir zaman serisi nesnesine dönüştürüp, dağılım grafiğine inceliyoruz.



Verinin dağılımına bakıldığı zaman mevsimsellik var denilebilir. Fakat trendden net bir şekilde bahsedilemez.

3.3.2 Durağanlık kontrolü



Veri setinde ACF ve PACF grafiklerine bakıldığı zaman seri durağan değildir denilmesi mümkündür. Seri de otokorelasyon gözükmemektedir. Bu durumu da Ljung-Box test ile kontrol etmek gerekir.

```
##
## Box-Ljung test
##
## data: data
## X-squared = 65.712, df = 1, p-value = 5.551e-16
```

H0: Otokorelasyon yoktur.

H1: Otokorelasyon vardır.

p-value değeri 5.551e-16 olarak bulunmuştur. 0.05 anlamlılık düzeyinden daha küçük olduğundan H0 hipotezi red edilir. Dolayısıyla otokorelasyon olduğu sonucuna varılır.

3.3.3 Birim Kök Testi

Durağanlık durumu birim kök test ile de kontrol edilir.

```
##
## #####
## # KPSS Unit Root Test #
## #####
```

```
##
## Test is of type: mu with 4 lags.
##
## Value of test-statistic is: 0.0611
##
## Critical value for a significance level of:
##          10pct  5pct 2.5pct  1pct
## critical values 0.347 0.463  0.574 0.739
```

Birim kök test testinin çıktısına göre Test-İstatistik değeri 0,0611 tespit edilmiştir. Bu değer tüm kritik değerlerden düşük olduğu için seri durağan gözükmemektedir.

Mevsimsel gözüken bir serinin aynı zamanda durağan olması söz konusu olan bir durum değildir. Bu yüzden bu adımda test kriteri değil de grafikler dikkate alınarak ilerlenmiştir.

3.3.4 Mevsimsel Fark Alma

Serimiz durağan çıkmadığı için bir fark alma işlemi yapılması gereklidir. Seri de güçlü bir mevsimselliği varlığı ya da yokluğu aşağıdaki gibi test edilir.

```
nsdiffs(data)
```

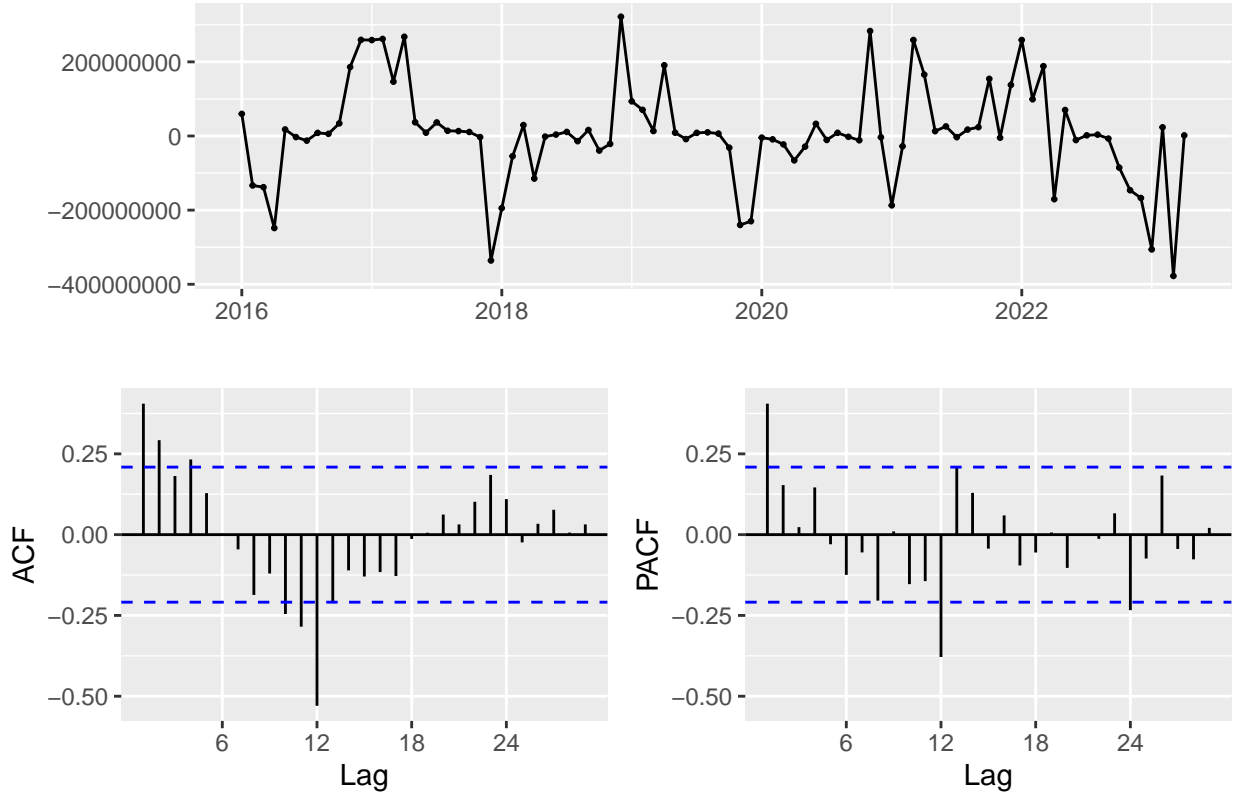
```
## [1] 1
```

Çıkan sonuç 1 olduğundan dolayı seri de güçlü bir mevsimsellik vardır denir. Yani mevsimsel bir fark almamız gereklidir.

```
data_diff <- diff(data, lag = 12)
nsdiffs(data_diff)
```

```
## [1] 0
```

1. Dereceden fark alma işlemi yapıldığı zaman, çıkan sonuç 0 olduğundan dolayı seri durağan hale getirilmiştir.



3.3.5 Test Seti ve Train Seti Ayrımı

```
train <- window(data, end= c(2022,12))
test  <- window(data, start = c(2023,1))
```

2023 yılının Nisan ayına kadar olan kısmı test seti ve geri kalan kısmı train seti olarak ayrılmıştır.

3.4 Arima Modeli

Arima için aday modelleri araştırarak olursak, grafiklerden ve fark alma işlemlerinden elde edilen bilgilere göre aday modeller ARIMA(2,0,0)(2,1,1), ARIMA(2,0,0)(2,1,1), ARIMA(1,0,0)(2,1,1) ve ARIMA(0,0,0)(2,1,1) modelleri çıkmıştır.

```
model_arima = Arima(train, order = c(2,0,0),
                      seasonal = c(2,1,1),
                      lambda = "auto")
model_arima
```

```
## Series: train
## ARIMA(2,0,0)(2,1,1)[12]
## Box Cox transformation: lambda= 0.2559781
##
```

```
## Coefficients:
##          ar1      ar2      sar1      sar2      sma1
##      0.3633  0.2031  0.0574  0.1651 -0.8894
## s.e.  0.1236  0.1080  0.4549  0.3055   0.8828
##
## sigma^2 = 602.7:  log likelihood = -392.2
## AIC=796.39   AICc=797.48   BIC=810.98
```

```
fcast = forecast(model_arima,
                  h = length(test))
accuracy(test,fcast$mean)
```

```
##              ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
## Test set -11544330 99831044 77508857 -1.743467 8.089575 -0.624541 0.3751901
```

Aralarında AIC_c değerlerine göre en iyi sonuç veren model ARIMA(2,0,0)(2,1,1)_12 modeli tespit edilmiştir.

3.5 Auto.arima Modellerinin Oluşturulması

```
## Series: train
## ARIMA(1,0,0)(0,1,1)[12] with drift
##
## Coefficients:
##          ar1      sma1      drift
##      0.4187 -0.7439 1308837.6
## s.e.  0.1014   0.1431  596566.9
##
## sigma^2 = 8007741237295279:  log likelihood = -1660.55
## AIC=3329.1   AICc=3329.6   BIC=3338.82
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 578872.2 82198124 53471116 -3.258011 10.95964 0.6464764
##              ACF1
## Training set -0.004829693
```

```
##              ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
## Test set -71386191 123444311 114069643 -8.785157 12.67055 -0.6728928 0.3662936
```

Burada auto.arima ile fonksiyonu ile uygun model seçim işlemini gerçekleştirmesini sağlanmıştır. AIC_c değeri 3329,6 olan ARIMA(1,0,0)(0,1,1)[12] modelini seçmiştir. Bu modelin test seti üzerindeki performansına bakılacak olursa MAPE değeri 12.67055 çıkmıştır. Burada MAPE değerine bakılmasının sebebi serideki değerlerin çok yüksek değerler olmasıdır.

Auto arimanın modeli daha ayrıntılı tahmin etmesi için *approximation = FALSE, stepwise = FALSE* parametrelerinin bu şekilde düzenlenmesi gereklidir.

```
mod1<-auto.arima(train,lambda = "auto",approximation = FALSE, stepwise = FALSE) #Burada auto.arima mode
```

```
## Series: train
```

```

## ARIMA(1,0,0)(1,1,1)[12] with drift
## Box Cox transformation: lambda= 0.2559781
##
## Coefficients:
##          ar1      sar1      sma1  drift
##          0.3181 -0.1215 -0.7877  0.432
## s.e.  0.1115   0.2101   0.3064  0.120
##
## sigma^2 = 560.1: log likelihood = -389.82
## AIC=789.64  AICc=790.41  BIC=801.8
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
## Training set 4975944 78923765 48897125 -0.9272549 9.383152 0.5911759 0.09354245

##              ME      RMSE      MAE      MPE      MAPE      ACF1 Theil's U
## Test set -80801072 126052029 107721458 -9.661325 12.11183 -0.5667868 0.2857279

```

Gerekli düzenleme yapıldıktan sonra seçilen model ARIMA(1,0,0)(1,1,1)[12] ve AIC_c değeri 790.41 olan model seçilmiştir. İlgili modelin test seti üzerinde ki performansının MAPE değeri 12.11183 olarak hesaplanmıştır. Bu adımda hesaplanan model, önceki modele göre daha iyi bir modeldir.

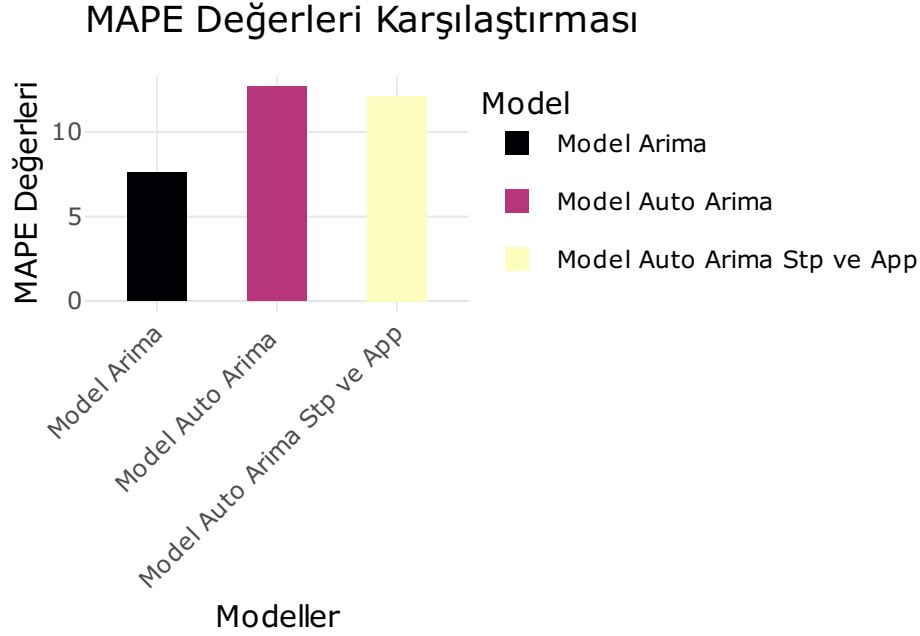
```

## [1] "ARIMA(1,0,0)(1,1,1)[12] modeli için MAPE değeri 12.1118333871363"

## [1] "ARIMA(1,0,0)(0,1,1)[12] modeli için MAPE değeri 12.6705451449269"

## [1] "ARIMA(2,0,0)(2,1,1)[12] modeli için MAPE değeri 7.61768815345162"

```

Genel olarak Arima yöntemleriyle oluşturulan 3 modelden, test seti üzerinde en iyi performansı veren model $ARIMA(2,0,0)(2,1,1)[12]$ modeli seçilmiştir.

3.6 $ARIMA(2,0,0)(2,1,1)[12]$ İle 12 Adımlık Tahmin Değerleri

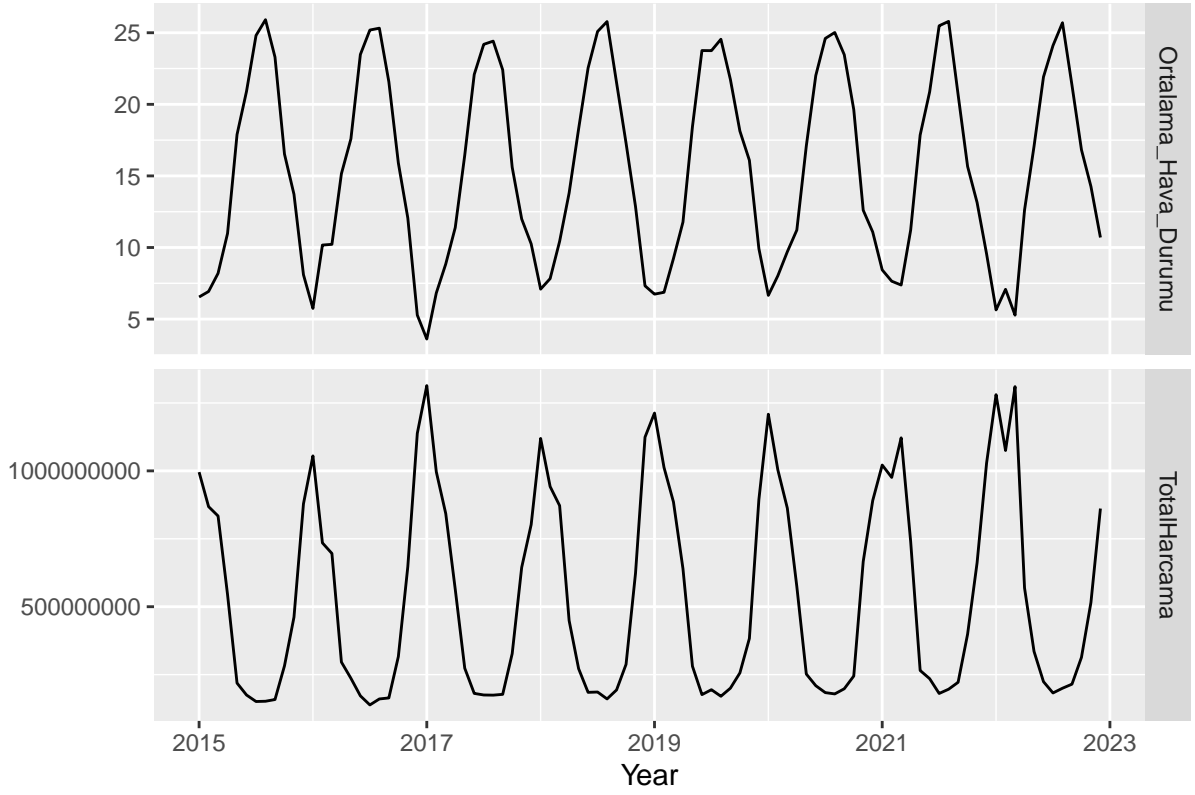
##	Jan	Feb	Mar	Apr	May	Jun
## 2023	1069468952	926988884	969534918	564411013	267499486	202407901
##	Jul	Aug	Sep	Oct	Nov	Dec
## 2023	173686435	179411558	197429834	318193216	574921489	946522746

2023 Yılı'nın tahmin değerleri bu şekilde çıkmıştır.

3.7 Sıcaklık İle Doğalgaz Tüketimi Verisi Arasındaki İlişki

##	Ortalama_Hava_Durumu	TotalHarcama
## 1	6.544355	995387838
## 2	6.929762	868369090
## 3	8.194624	834060447
## 4	10.982500	544521298
## 5	17.890457	218834719
## 6	20.911806	174883424

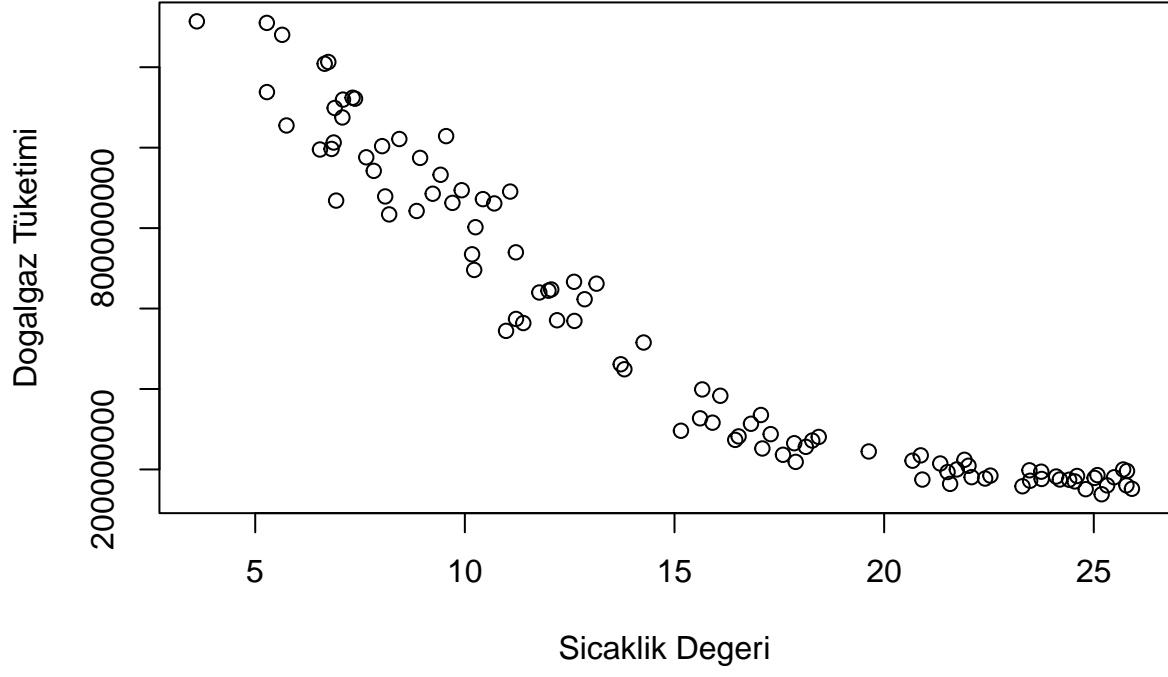
Hava durumu ve dogalgaz tüketimi



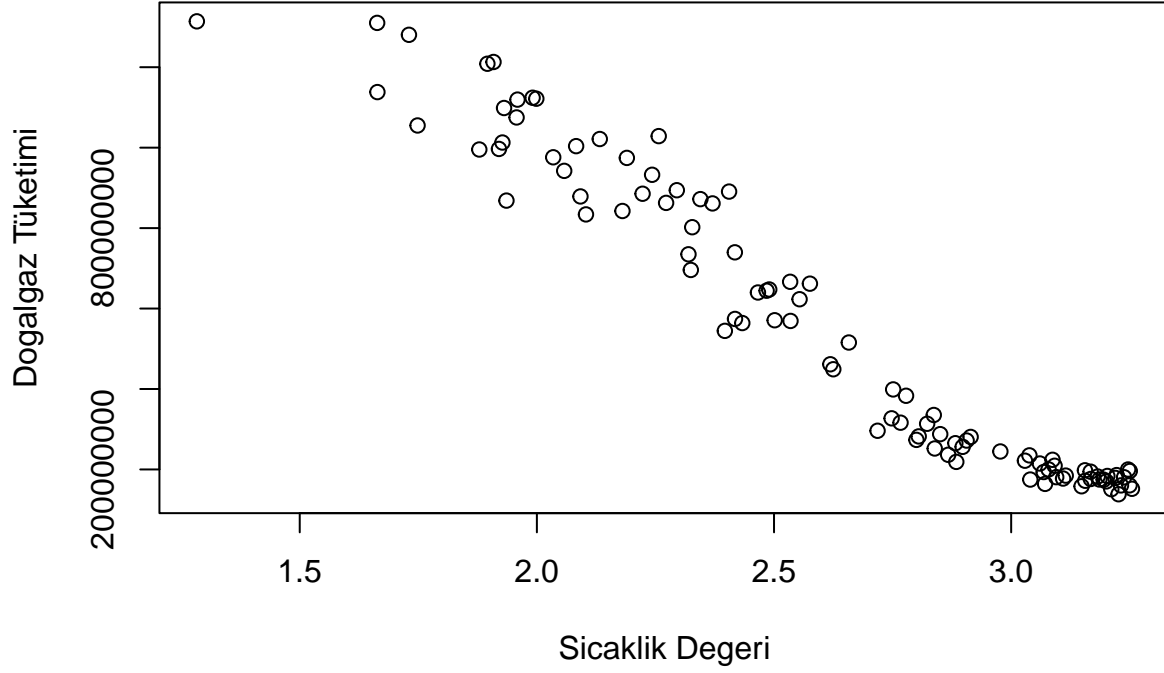
Hava durumu serisinin autoplot grafiğine bakıldığı zaman bir mevsimsellik olduğu gözükmemektedir. 2 grafiğe bakılacak olursa aralarında ters bir ilişki olduğu görülebilir. Bunun için aralarındaki korelasyon katsayısına ve nokta grafiğine bakarak inceleyebiliriz.

```
## [1] -0.9405073
```

2 Değişken arasında negatif yönde -0.9405073 değerlik güçlü bir ilişki olduğu saptanmıştır.



İlgili nokta grafiği incelendiği zaman aralarındaki ilişkinin daha da lineer olması beklenir. Bu işlemi sıcaklık verisinin logaritmik dönüşümü kullanarak elde edebiliriz.



İlgili işlem yapıldıktan sonra ilişkinin daha da lineer olduğu gözükmemektedir.

```
## [1] -0.9705788
```

Dönüşüm yapıldıktan sonra aradaki korelasyon değerinin daha da arttığı gözükmemektedir.

3.8 2022 Yılı İçin Doğalgaz Tüketimi İle Abone Sayısı Verisinin İlişkisi

3.8.1 Doğalgaz Tüketim Verisi İle Abone Sayısı Verisinin Birleştirilmesi

```
## # A tibble: 10 x 2
##   İLÇE      'ABONE SAYI'
##   <chr>      <dbl>
## 1 ADALAR      11064
## 2 ARNAVUTKÖY  147578
## 3 ATAŞEHİR    198619
## 4 AVCILAR     179260
## 5 BAĞCILAR    274645
## 6 BAĞCELİEVLER 237913
## 7 BAKIRKÖY    137930
## 8 BAŞAKŞEHİR  195323
## 9 BAYRAMPAŞA  118167
## 10 BEŞİKTAŞ   138419
```

```
## # A tibble: 39 x 2
```

```
## İlçe ToplamTüketim
## <chr> <dbl>
## 1 ADALAR 6874142.
## 2 ARNAVUTKÖY 205340364.
## 3 ATAŞEHİR 192340421.
## 4 AVCILAR 173403243.
## 5 BAĞÇELİEVLER 209975435.
## 6 BAKIRKÖY 165817672.
## 7 BAYRAMPAŞA 99011317.
## 8 BAĞCILAR 246959146.
## 9 BAŞAKŞEHİR 238544575.
## 10 BEYKOZ 123969009.
## # i 29 more rows
```

Abone sayısı verisi 2022 yılını kapsadığı için Doğalgaz Tüketim verisi için de sadece 2022 yılı baz alınmıştır. İlçe değişkenini kullanarak iki veri setini birleştirelim.

```
## # A tibble: 39 x 3
## İlçe ToplamTüketim 'ABONE SAYI'
## <chr> <dbl> <dbl>
## 1 ADALAR 6874142. 11064
## 2 ARNAVUTKÖY 205340364. 147578
## 3 ATAŞEHİR 192340421. 198619
## 4 AVCILAR 173403243. 179260
## 5 BAĞÇELİEVLER 209975435. 237913
## 6 BAKIRKÖY 165817672. 137930
## 7 BAYRAMPAŞA 99011317. 118167
## 8 BAĞCILAR 246959146. 274645
## 9 BAŞAKŞEHİR 238544575. 195323
## 10 BEYKOZ 123969009. 112513
## # i 29 more rows
```

İstanbul'un 39 ilçesine ait 2022 yılı Doğalgaz tüketim miktarı ve abone sayılarından oluşan tablo görülmektedir.

```
##
## Call:
## lm(formula = ToplamTüketim ~ 'ABONE SAYI', data = veri)
##
## Coefficients:
## (Intercept) 'ABONE SAYI'
## -440217 1012

##
## Call:
## lm(formula = ToplamTüketim ~ 'ABONE SAYI', data = veri)
##
## Residuals:
## Min 1Q Median 3Q Max
## -33708970 -17233946 -6188558 12272547 91149766
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)    -440216.55 10506273.31  -0.042    0.967
## 'ABONE SAYI'      1011.81      53.09  19.060   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 27340000 on 37 degrees of freedom
## Multiple R-squared:  0.9076, Adjusted R-squared:  0.9051
## F-statistic: 363.3 on 1 and 37 DF,  p-value: < 2.2e-16
```

Toplam tüketim'i bağımlı değişken olarak abone ve tüketim miktarı arasında regresyon modeli oluşturulmuştur. Anlamlılık düzeyleri, regresyon katsayıları, p-değerleri, R-kare değeri ve diğer istatistikleri içerir. p-value değeri 2.2e-16 çıkmıştır. Bu değer 0.05 anlamlılık düzeyinden küçük olduğu için model anlamlı çıkmıştır. Adj R-kare 0.9051 çıkmıştır. Bu değer 1'e yakın olması modelin bağımlı değişkendeki varyansın büyük bir kısmını açıkladığını ve aşırı uyumlu olmadığını gösterir. Ancak, ayarlanmış R-kare değeri, tek başına bir modelin iyiliğini tam olarak belirtmez ve başka model değerlendirme yöntemleriyle birlikte kullanılmalıdır.

3.8.2 Doğalgaz Tüketim verisi İle Abone Sayısı Verisinin Korelasyonu

```
# Korelasyonu hesaplama
correlation <- cor(veri$`ABONE SAYI`, veri$`ToplamTüketim`)

# Korelasyon değerini görüntüleme
print(correlation)
```

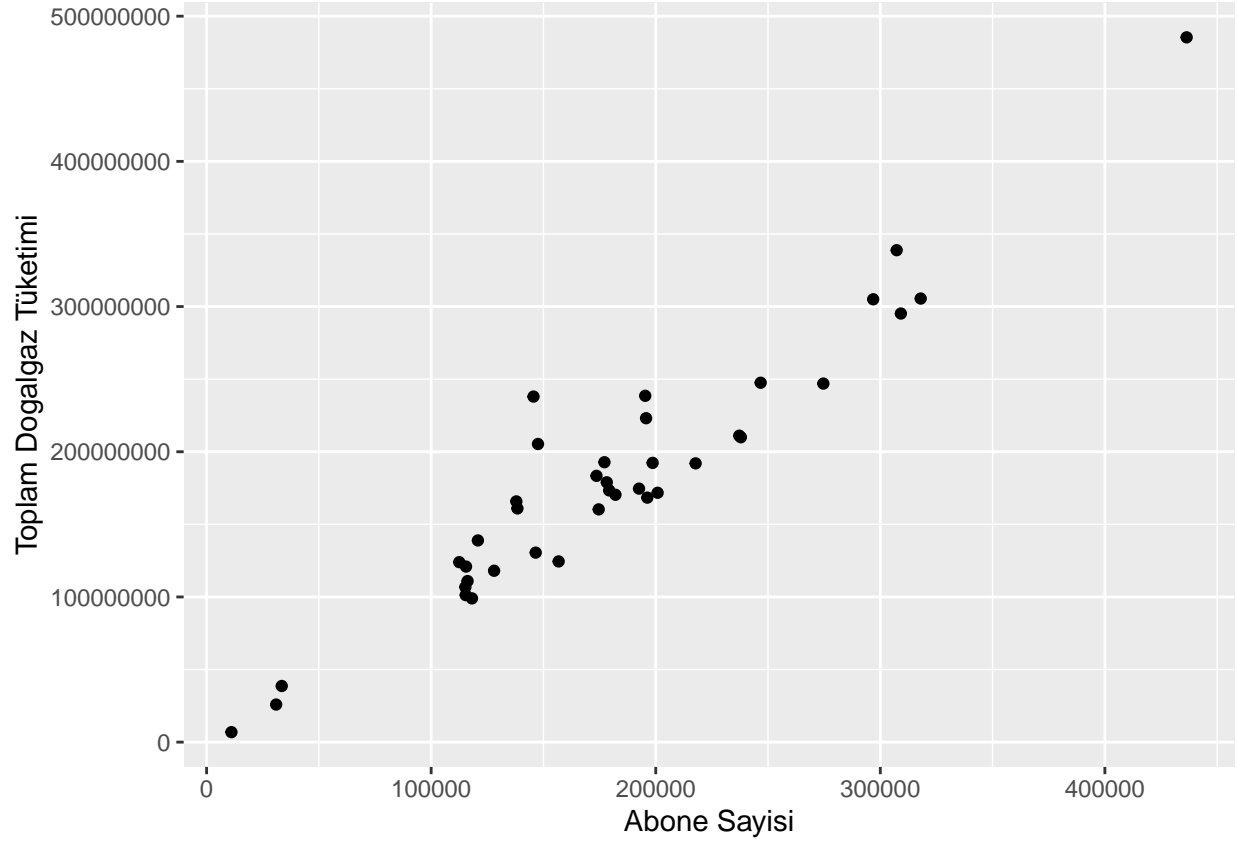
```
## [1] 0.9526627
```

Bu şekilde, “veri” veri setindeki “Abone Sayı” ve “Toplam Tüketim” sütunları arasındaki korelasyonu ölçülmüştür. Korelasyon değeri -1 ile 1 arasında bir değer alır, -1 negatif bir ilişkiyi, 1 pozitif bir ilişkiyi ve 0 ilişkisizliği ifade eder. Sonuç olarak 0.9526627 çıktığı için yüksek değer olarak değerlendirebilir. Abone ile Toplam Doğalgaz Tüketimi arasında güçlü yönde pozitif bir ilişki vardır.

3.8.3 Doğalgaz Tüketim Verisi İle Abone Sayısı Verisinin Korelasyon Grafiği

```
# ggplot2 paketini yükleme
library(ggplot2)

# Korelasyon grafiğini oluşturma
ggplot(veri, aes(x = `ABONE SAYI`, y = `ToplamTüketim`)) +
  geom_point() +
  labs(x = "Abone Sayısı", y = "Toplam Doğalgaz Tüketimi")
```



Dağılım: Grafikteki noktaların dağılımı, “Abone Sayısı” ve “Toplam Tüketim” arasındaki ilişki hakkında bize bilgi verir. Eğer noktalar genel olarak yukarı doğru bir eğilim gösteriyorsa, yani “Abone Sayısı” arttıkça “Toplam Tüketim” de artıyorsa, pozitif bir ilişkiden bahsedebiliriz. Ve korelasyon katsayımız da yüksek olduğu için (0.9526) pozitif yönlü güçlü bir ilişkiden bahsetmiştik. Noktalar rastgele dağılmamıştır ve eğim nettir.

Yoğunluk: Grafikte noktaların yoğun olduğu bölgeler ilişkinin gücünü gösterir. Eğer noktalar yoğun bir şekilde bir araya gelmişse, bu belirli bir “Abone Sayısı” aralığındaki “Toplam Tüketim” değerlerinin benzer olduğunu gösterebilir. Bu durumda ilişki daha güçlü olabilir. Eğer noktalar dağılımı daha seyrek ise, ilişki daha zayıf olabilir veya ilişkisizlik söz konusu olabilir. Bu grafikte noktalar birbirine yakındır.

Aykırı Değerler: Grafikteki aykırı noktalar, genel eğilimden belirgin şekilde sapmış veri noktalarını temsil eder. Bu grafikte 2 değerin sapmış olabileceğini söyleyebiliriz.

Veri görselleştirmede Abone ve Toplam Doğalgaz Tüketimi arasında pozitif negatif veya zayıf bir ilişki olup olmadığını grafikten de anlayabilir ve yorumlayabiliriz. Aykırı değerlerin varlığını ve bunun analize etkisine dikkat etmek gereklidir.

Bu adıma kadar seriler hep aylık şekilde incelenmiştir.

Verinin formatını çeyreklik olarak değiştirip, inceleme yapılmıştır.

3.9 Çeyreklik veri İle Zaman Serisi

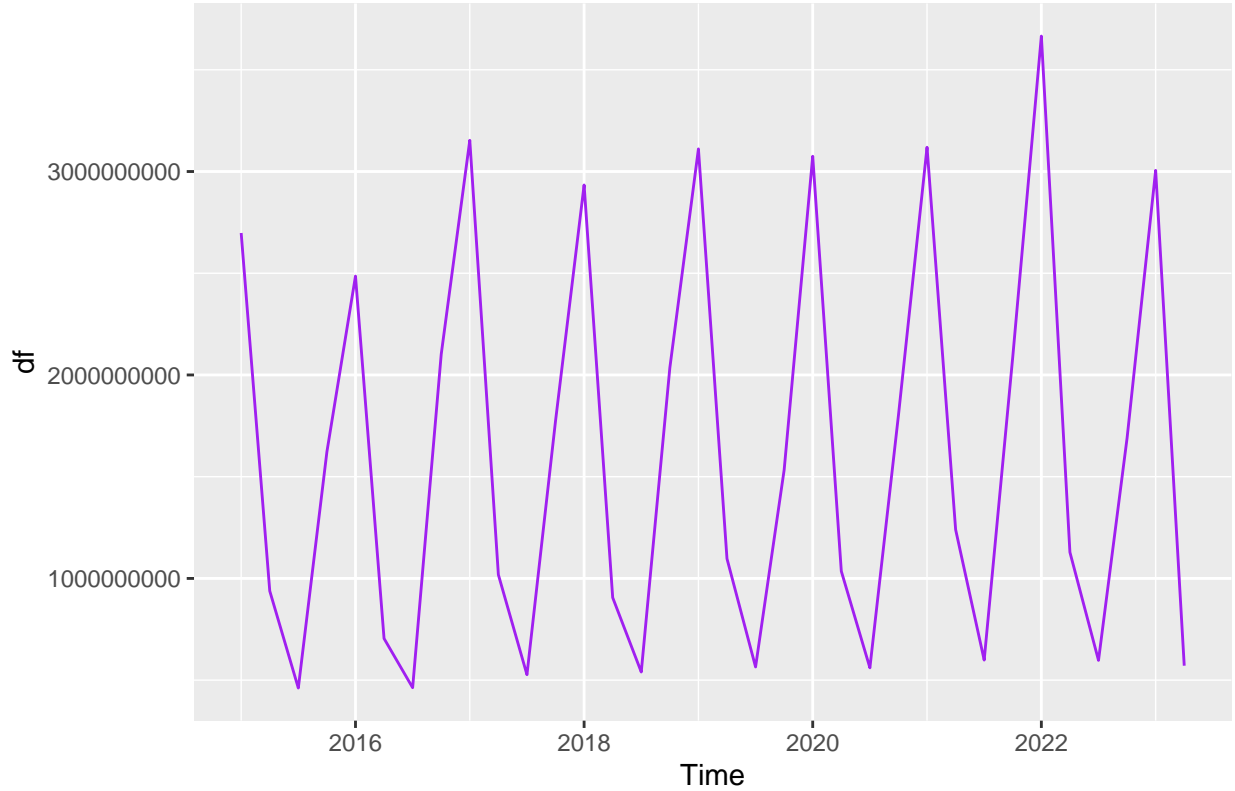
```
## # A tibble: 34 x 2
##   Mevsim toplamHarcama
##   <chr>         <dbl>
## 1 2015_Q1  2697817375.
```

```
## 2 2015_Q2 938239441.
## 3 2015_Q3 461198537.
## 4 2015_Q4 1622346748.
## 5 2016_Q1 2486025528.
## 6 2016_Q2 704636695.
## 7 2016_Q3 462907155.
## 8 2016_Q4 2101789970.
## 9 2017_Q1 3153279778.
## 10 2017_Q2 1018379209.
## # i 24 more rows
```

Toplamsal bir şekilde veri çeyreklik olarak ayrıştırılmıştır.

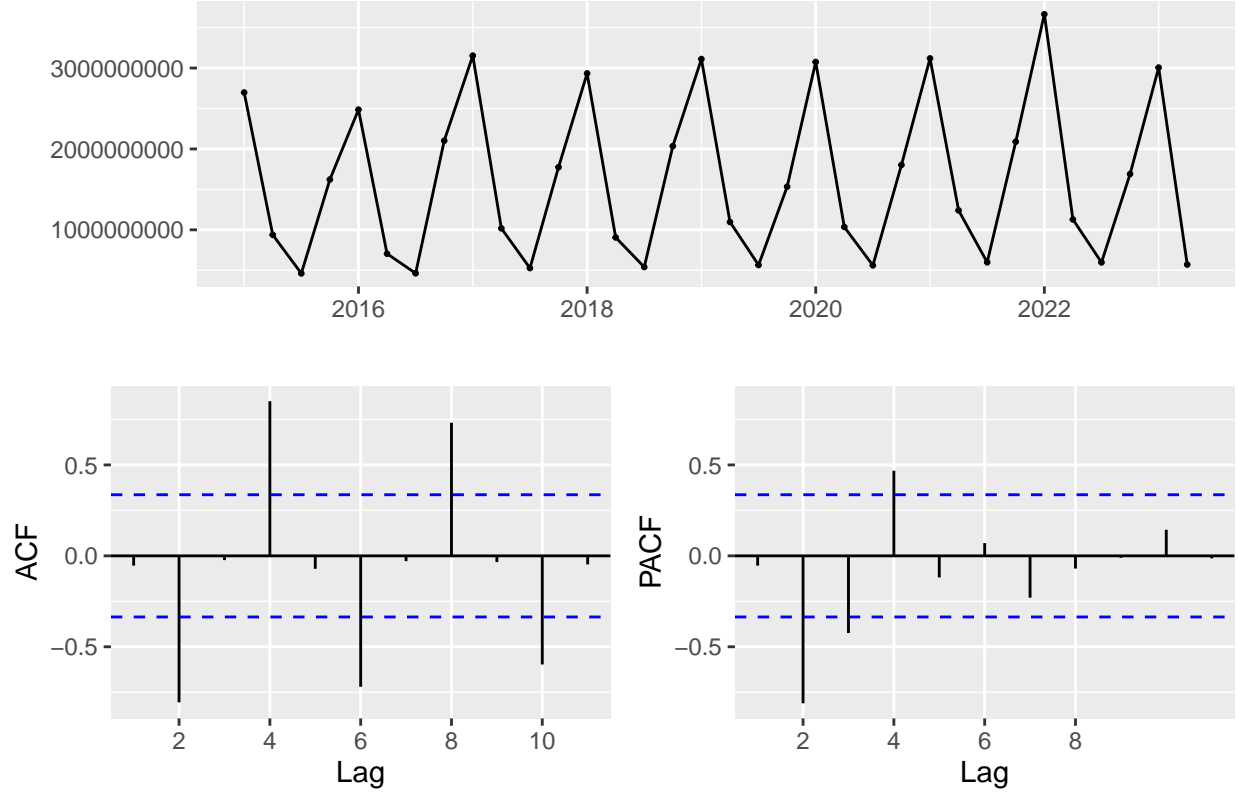
##	Qtr1	Qtr2	Qtr3	Qtr4
## 2015	2697817375	938239441	461198537	1622346748
## 2016	2486025528	704636695	462907155	2101789970
## 2017	3153279778	1018379209	526843706	1773728426
## 2018	2933737909	905734716	540066688	2035043891
## 2019	3111131982	1097100731	564842180	1533593273
## 2020	3075008725	1035433241	560580298	1801997199
## 2021	3119184790	1239994817	598611727	2089486452
## 2022	3665725319	1128654674	597151423	1690470946
## 2023	3005638250	570942837		

2015'in ilk döneminden 2023'ün ilk dönemine kadar veri zaman serisine dönüştürülmüştür.



Serinin dağılımına bakıldığı zaman dönemsel olarak da bir mevsimsellik gözükmektedir.

3.9.1 Durağanlık Kontrolü



Grafik üzerinden otokorelasyon net bir şekilde anlaşılmamaktadır. Bundan dolayı istatistiksel testler ile incelenmelidir.

```
##  
## Box-Ljung test  
##  
## data: df  
## X-squared = 0.10681, df = 1, p-value = 0.7438
```

H0: Otokorelasyon yoktur.

H1: Otokorelasyon vardır.

p-value değeri 0.7438 olarak bulunmuştur. 0.05 Anlamlılık düzeyinden daha büyük olduğundan H0 hipotezi red edilemez Dolayısıyla otokorelasyon olmadığı sonucuna varılır.

Bundan dolayı serinin modellemeye uygun olmadığı sonucuna varılır.

SONUÇ

Bu çalışmada ARIMA zaman serisi analiz yöntemleri kullanılarak İstanbul ilinin doğal gaz tüketimi tahmin edilmiştir. Veriler 2015 yılı ile 2023 Nisan ayı ve 2022 abone sayılarını kapsamaktadır. İstanbul İlçe Bazında Aylık Doğalgaz Tüketim verileri ve Abone Sayıları İstanbul Büyükşehir Belediyesi Açık Veri Portalından temin edilmiştir. Bu yöntem yardımıyla elde edilen tahmin sonuçları MAPE performans kıstası yardımıyla karşılaştırılmıştır. Bu performans ölçütünü değerlendirmek için 3 adet ARIMA modeli oluşturulmuştur. Bu ARIMA modellerinden, test seti üzerinde en iyi performansı ARIMA(2,0,0)(2,1,1)[12] modeli sergilemiştir. Abone sayısı ve sıcaklık değişkeni ile doğalgaz tüketim miktarı arasında yüksek bir korelasyon mevcuttur. Doğalgaz tüketimini etkileyen faktörler arasında doğalgazın fiyatı, alternatif enerji kaynaklarının (petrol, kömür elektrik gibi) fiyatları, hava sıcaklığı, abone sayısı gibi değişkenlerde sayılabilir. Bu farklı değişkenlere ait verilerin olmayışı ve modellere dahil edilememesi çalışmanın kısıtlarından birini oluşturmaktadır. Ülkemizde Aylık bazda doğalgaz istatistiklerine ait veri sayısının az oluşu aylık tahmin yapacak modellerin oluşturulmasına engel teşkil etmiştir. İleriki çalışmalarda farklı değişkenlere ilişkin veriler temin edilerek bu değişkenleri kullanan regresyon modelleri oluşturulup tahmin sonuçları karşılaştırılabilir.

KAYNAKÇA

<https://www.igdas.istanbul/kaynaktan-eve-dogal-gaz/#:~:text=This%20opens%20in%20a%20new,boru%20hatlar%C4%B1>

<https://data.ibb.gov.tr/dataset/ilce-bazinda-aylik-dogalgaz-tuketim-miktari>

<https://data.ibb.gov.tr/dataset/ilcelere-gore-abone-sayilari>

<https://dergipark.org.tr/en/download/article-file/30457>

<https://dergiler.ankara.edu.tr/xmlui/bitstream/handle/20.500.12575/31130/gorkemyusuftopcu.pdf?sequence=1&isAllowed=y>

<https://acikerisim.sakarya.edu.tr/bitstream/handle/20.500.12619/76340/T06894.pdf?sequence=1&isAllowed=y>

<http://nek.istanbul.edu.tr:4444/ekos/TEZ/43349.pdf>

<https://dergipark.org.tr/en/download/article-file/1793885#:~:text=Tahmin%20sonu%C3%A7lar%C4%B1na%20g%C3%B6r>

<https://dergipark.org.tr/en/download/article-file/30457>

https://open-meteo.com/en/docs/historical-weather-api#latitude=41.07&longitude=28.95&hourly=temperature_2m&timezone=auto

<https://www.kaggle.com/code/mykeysid10/natural-gas-price-prediction-using-time-series>

<https://dergipark.org.tr/en/download/article-file/30457>

<https://dash.plotly.com/r/interactive-graphing>

<https://www.kaggle.com/code/soraka/usa-electricity-generation-from-natural-gas-with-r>

<https://yigitsener.medium.com/veri-biliminde-eksik-kay%C4%B1p-verilere-yakla%C5%9F%C4%B1m-stratejileri-ve-python-pandas-uygulamas%C4%B1-501fbf643795>

<https://www.proquest.com/docview/2689295472?fromopenview=true&pq-origsite=gscholar>