Fiyat analizi yapılacaktır. Bir gün içerisinde ne kadar sipariş gelmiş ise o kadar satış yapılmış mıdır? Bir ürünün fiyatı arttığında satın alınma miktarları azalır mı? Bu konular hakkında bir analiz gerçekleştirilecektir.

Kullanmış olduğum veri setinde toplamda 541910 adet veri vardır. Bu verilerin içerisinde hatalı girilmiş, kullanılamayacak satırlar kaldırıma yöntemine gidilerek 477482 adet veri üzerinden analiz yapılacaktır.

Aykırı değer analizi fiyat sütununda yapılmıştır. Kutu grafiğinde aykırı değerlerin olduğu öğrenilmiştir. En yüksek aykırı değerlerin toplamda 33956 adet olduğu en küçük aykırı değerlerin ise olmadığı görülmüştür bunun sebebi bozuk ve hatalı verileri temizlerken fiyat bilgisinin eksili bir değer veya sıfır TL olunamayacağını düşündüğümden eksili ve sıfır değerine sahip olan verileri sildim. Aykırı değerleri ortalama fiyat bilgisine çektim ve bu kutu grafiği öncesi ve sonrası sırasıyla;Aykırı değerlerin öncesi




A blue rectangular object with a black line

Description automatically generated

2010 ile 2011 yılları arasındaki bütün sipariş kayıtları olan bir veri setidir. Bu tarihleri arasında günde satılan ürünlerin toplam fiyatlarını ve toplam satış adetleri, analizimize bu iki veri kümesinden devam edeceğiz. Toplam fiyatların normallik sınaması ile başlarsak;

Öncelikle verilerimizde parametrik testler mi yoksa nan-parametrik testler mi kullanacağımızı anlayalım. Basıklık 0,75 yani-1,5 ile 1,5 arasında değerde olduğu için basıklık normal dağılama uymaktadır. Çarpıklık 0,87 olduğu için bu değerler arasında yer alır ve çarpıklık açısından da normal dağılıma uyuyor olduğu yorumu yapılabilir. Proplot ile kontrol edersek;A graph with a blue line

Description automatically generated

Grafiğe göre verilerin çoğunluğu çizgiye yakın olduğundan normal dağıldığı yorumu yapılabilir ancak istatistiksel test olan shapiro-wilk ve kolmogrow smirnov testlerini kullandığımızda shapiro test sonucu 0.0000000087 çıkıyor %95 güven düzeyinde 0.05den küçük olduğundan boş hipotez reddedilir normal dağılım olmadığı sonucu çıkıyor kolmogrow smirnov testinin sonucu 0.000690 çıkıyor bu test sonucunda da boş hipotez reddedilir normal dağılım olmadığı sonucu ortaya çıkar.

Toplam satış adetlerinin normallik sınamasından devam edersek;

Kurtosis değeri 0.62 skew değeri 0.85 bu iki değer -1,5 ile1,5 arasında olduğu için normallik olduğu düşünülüyor

A graph with a line

Description automatically generated

Grafiğe göre normal dağılıma uyduğu düşünülebilir. Shapiro testinde göre 0.0000000079 kolmogrow smirnov testine göre 0.0047 sonucu çıkmıştır normallik sınamasında test sonucunda da geçememiştir.

Gün bilgilerine göre Toplam fiyatlar ve Toplam satış adetlerinin arasında bir ilişki testi yapılacak verilerin normal dağılmadığını anladığımız doğrusallık testini uygulamamıza gerek yok ancak doğrusallığına bakmak istersek;

A blue dotted line with white text

Description automatically generated

Veriler doğrusal dağılmıştır doğrusallık şartından geçmiştir ancak normallikten kaldığı için Pearson korelasyon testini uygulayamayacağız onun yerine Spearman korelasyon testi uygulanması gerekiyor. Spearman korelasyon testi uygulandık ve bunu ısı grafiğinde gösterdiğimizde;

A graph of a number of numbers

Description automatically generated with medium confidence

Isı grafiğine baktığımızda grafik 0,97 kat sayısı ile pozitif yönde korelasyon olduğunu gösteriyor. Bu gerçek güçlü bir katsayı olduğu için fiyatlar ile satış adetleri arasında güçlü bir korelasyon olduğunu söyleyebiliriz.

Bir gün içerisinde ne kadar çok sipariş gelmiş ise o kadar çok satış yapılmıştır ilişkisi doğrulanmıştır.

Benim araştırmak istediğim asıl konu bir ürünün fiyatı arttıkça ürünün satış miktarı azalır mı?

Bunun için öncelikle birim fiyatların çarpıklık derecesine baktım bu değer 2.16 olduğu için sağa çarpık olduğunu söyleyebiliriz. Aralık -1.5 ile 1.5 arasında bir değerde kalmadığı için normal dağılıma uygun olmadığını düşünebiliriz verilerin kutu grafiği sonucunda aykırı değerlerinin olduğu anlaşılmıştır Aşağıda bu yapı kutu grafiğinde gösterilmiştir;

A blue rectangular object with a line

Description automatically generated

Bu değerlerin ortalamayı etkileyeceği ve analizde sorunlar çıkartacağından dolayı verileri kaldırmak yerine karekök dönüşümü uygulanacaktır. Verilerin aykırı değerlerinin etkisini azaltmak amacıyla yapılmaktadır. Karekök dönüşümden sonra çarpıklık 1.23 düşmüş, normal dağılıma uygun olduğu düşüncesini doğrulamak için shapiro testini uygulamamız daha uygun bir sonuç verebileceği düşünülmüştür. Shapiro testine göre değer %95 güven düzeyinin epey bir altında kalmışta, bundan dolayı normallik kabul edilememektedir. Sıra miktarların çarpıklık derecesinde çarpıklık derecesi 6.78 olduğu için sağa çarpık diyebildiğimizden dolayı normal dağılmadığı yorumunu da çıkartabiliriz. Tabi kesinleştirmek için shapiro testi uygulanmalıdır. Test sonucu 0.05’den çok küçük bir değer çıkmasından dolayı normal dağılmadığını anlayabiliriz. Miktarlarda da aykırı değerler mevcut olduğu için ve sağa çarpık olduğu için karekök dönüşümü uygulayabiliriz. Fiyat verileri ile miktar verileri arasında ilişkiyi kontrol etmek için spearman korelasyon testi uygulanmalıdır. Isı grafiği ile spearman korelasyon gösterimi aşağıdaki gibidir;

A diagram of a number of blue squares

Description automatically generated

Bu test sonucunda p değeri 0.000004’tür, bu %95 güven düzeyinde H1 hipotezi kabul görür ve aralarında bir bağlantı olduğu anlaşılır. Spearman katsayısı(r) -0.24 İlişkinin yönü negatif yönlü olduğu için bu sonucun bize negatif yönlü bir ilişki varlığını söylemektedir. Ürünlerin fiyatları artarken miktarları azalır ürünlerin fiyatları azalırken miktarları artar hipotezi kanıtlanmış oldu.