Medium Search Get app Write

Home
Library
Profile
Stories
Stats

Following

Find writers and publications to follow.
See suggestions

Welcome Offer Access to everything. Now 30% off. Upgrade now

Data Analysis + Eda + Python ✓ Data Visualization + Pandas +

Bir E-Ticaret Veri Seti Bize Gerçekten Ne Anlatıyor?

Fatma Aydın 11 min read · Jul 17, 2026

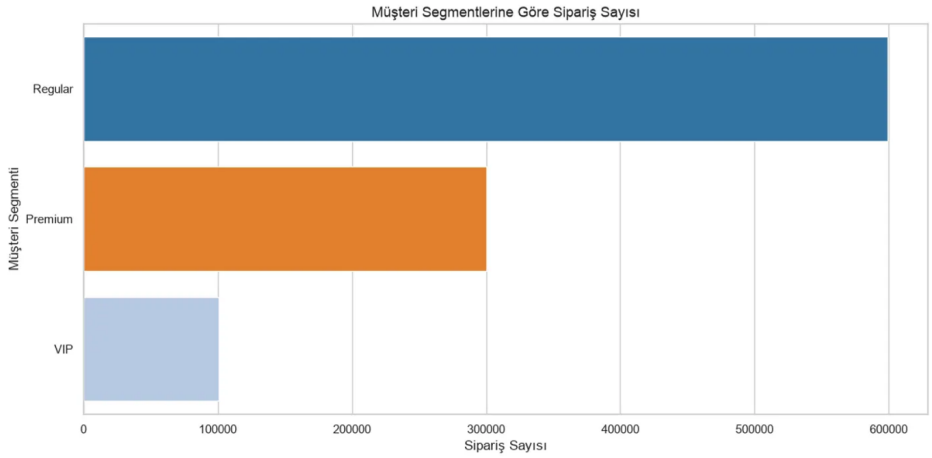
Veriler çoğu zaman yalnızca sayı gibi görünür. Binlerce satır, onlarca sütun ve ilk bakışta birbirinden bağımsız görünen bilgiler... Oysa doğru sorular sorulduğunda bu veriler, işletmelerin güçlü ve zayıf yönlerini ortaya çıkaran önemli ipuçları sunar.

Bu çalışmada bir e-ticaret veri setini Python kullanarak adım adım inceledim. Amacım yalnızca grafikler oluşturmak değil; satışların, kârlılığın, müşteri davranışlarının ve ürün kategorilerinin arkasındaki hikâyeyi keşfetmekti.

Keşifsel Veri Analizi (Exploratory Data Analysis — EDA) süreci boyunca veri setini tanımaktan başlayarak eksik değerleri, kategorik ve sayısal değişkenleri, ürün kategorilerini, markaları ve ülkeleri farklı açılardan analiz ettim.

Siparişlerin Büyük Bölümü Hangi Müşterilerden Geliyor?

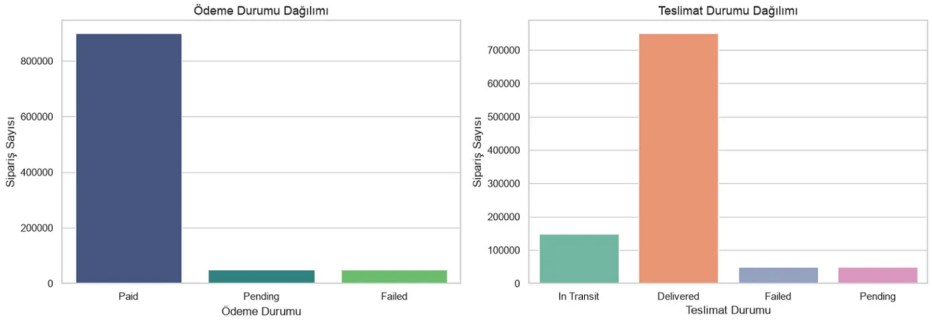
Her müşteri aynı alışveriş davranışını göstermiyor. Bu nedenle ilk olarak siparişlerin müşteri segmentlerine göre nasıl dağıldığını incelemek istedim. Böylece işletmenin hangi müşteri grubuna daha fazla bağımlı olduğunu görmek mümkün oldu.



Grafik incelendiğinde siparişlerin büyük çoğunluğunun Regular müşteri segmentinden geldiği görülüyor. Yaklaşık 600 bin sipariş ile bu grup, toplam sipariş hacminin önemli bir bölümünü oluşturuyor. Onu 300 bin sipariş ile Premium müşteriler takip ederken, VIP segmenti yaklaşık 100 bin sipariş ile listenin son sırasında yer alıyor.

Siparişler Sorunsuz Tamamlanıyor mu?

Bir e-ticaret sürecinin başarıyla sonuçlanması için yalnızca sipariş alınması yeterli değildir. Siparişin ödeme aşamasından sorunsuz geçmesi ve müşteriye zamanında ulaştırılması da en az satış kadar önemlidir. Bu nedenle veri setinde hem ödeme durumlarını hem de teslimat durumlarını birlikte incelemek istedim.



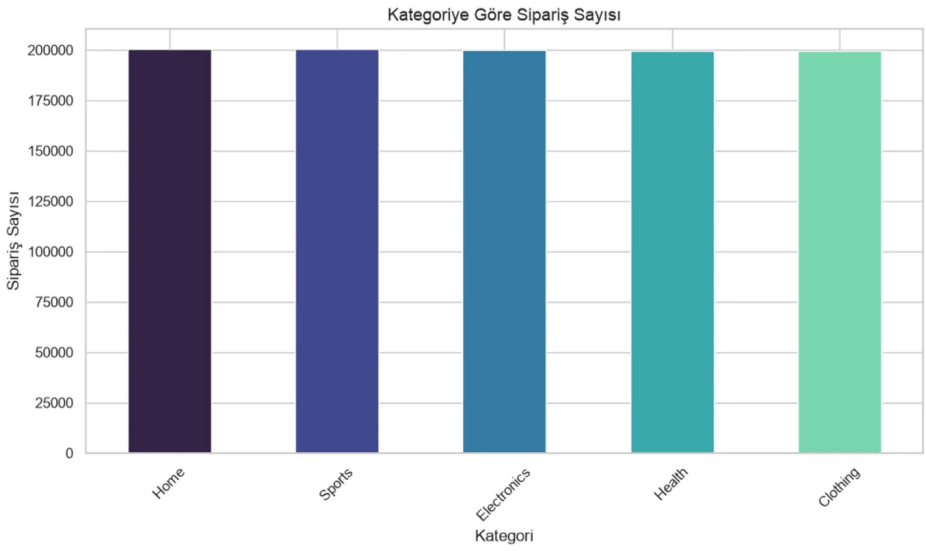
İlk grafik incelendiğinde siparişlerin büyük çoğunluğunun Paid durumunda olduğu görülüyor. Pending ve Failed ödeme sayıları ise oldukça düşük seviyede kalıyor. Bu durum, ödeme süreçlerinin genel olarak sorunsuz ilerlediğini ve müşterilerin büyük bölümünün satın alma işlemini başarıyla tamamladığını gösteriyor.

Teslimat tarafında da benzer bir tablo dikkat çekiyor. Siparişlerin önemli bir kısmı Delivered durumunda yer alırken, daha az sayıda sipariş hâlâ In Transit, Pending veya Failed olarak görünüyor. Bu dağılım, siparişlerin büyük bölümünün müşterilere başarıyla ulaştığını ve lojistik sürecinin genel anlamda istikrarlı işlediğini düşündürüyor.

İlk bakışta hem ödeme hem de teslimat süreçleri başarılı görünse de, tamamlanamayan siparişler tamamen göz ardı edilmemelidir. Özellikle başarısız ödemeler veya teslim edilemeyen siparişler düşük oranlarda gerçekleşse bile müşteri memnuniyetini ve işletmenin gelirini doğrudan etkileyebilir. Bu nedenle bu kayıtların neden oluştuğunu incelemek ve gerekli iyileştirmeleri yapmak, operasyonel verimliliği artırmak açısından önemli bir adım olacaktır.

Kategoriler Gerçekten Fark Yaratıyor mu?

Bir sonraki adımda siparişlerin ürün kategorilerine göre dağılımını incelemek istedim. İlk beklentim, bazı kategorilerin diğerlerine göre belirgin şekilde daha fazla sipariş aldığını görmekti. Ancak grafik farklı bir tablo ortaya koydu.



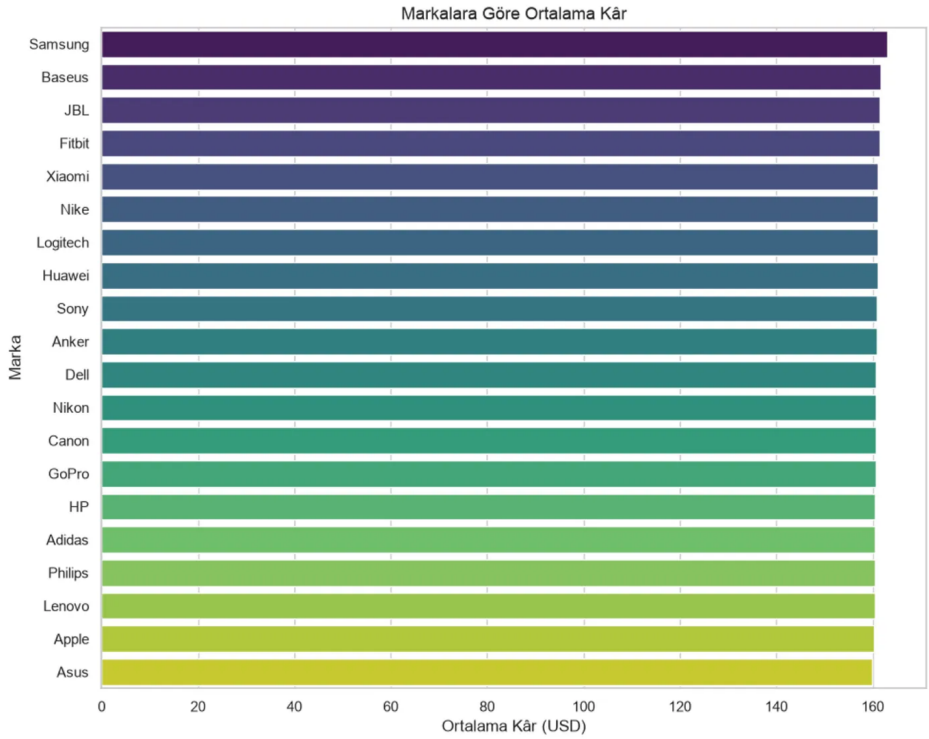
Grafiğe bakıldığında Home, Sports, Electronics, Health ve Clothing kategorilerinin sipariş sayılarının neredeyse aynı seviyede olduğu görülüyor.

Kategoriler arasında dikkat çeken bir fark bulunmadığı için bu grafik tek başına anlamlı bir içgörü sunmuyor.

Aslında bu da önemli bir bulgu. Çünkü her oluşturulan grafik mutlaka yorumlanabilir veya karar vermeye yardımcı olacak bilgiler içermek zorunda değildir. Bu dağılım, veri setindeki siparişlerin kategoriler arasında oldukça dengeli dağıtıldığını gösteriyor.

Markalar Arasında Belirgin Bir Kârlılık Farkı Var mı?

Bir sonraki adımda markaların ortalama kârlarını karşılaştırmak istedim. Amacım, belirli markaların diğerlerine göre daha yüksek kârlılık sağlayıp sağlamadığını görmekti.



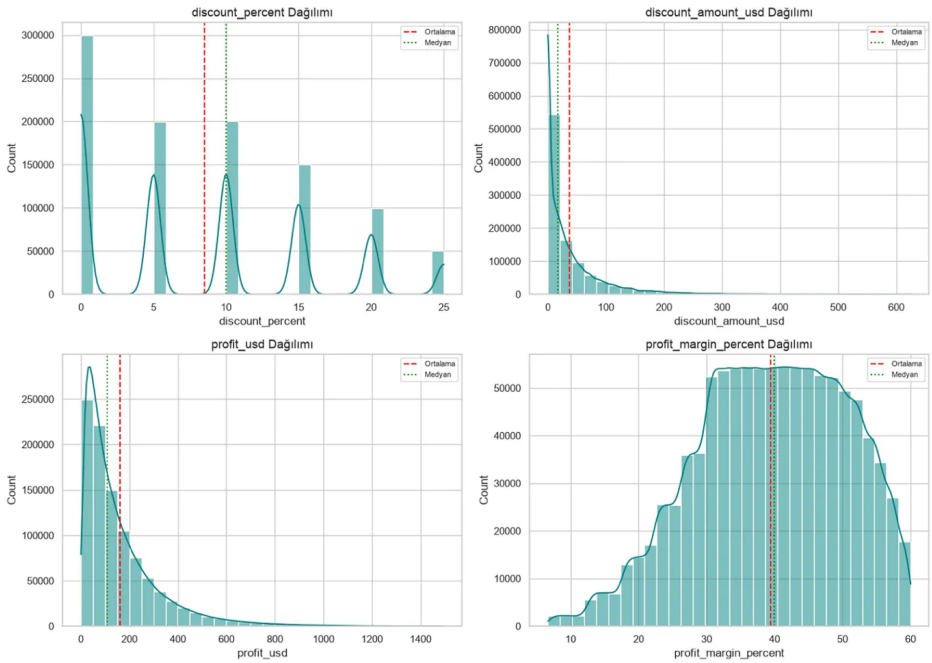
Grafik incelendiğinde Samsung en yüksek ortalama kâra sahip marka olarak görünse de, diğer markalarla arasındaki fark oldukça sınırlı. Baseus, JBL, Fitbit ve Xiaomi gibi markalar da benzer seviyelerde ortalama kâr değerlerine sahip. Listenin alt sıralarında yer alan markaların ortalama kârları da üst sıralardaki markalara oldukça yakın.

Bu nedenle grafik, markalar arasında belirgin bir performans farkı olduğunu söylemek için yeterli bir kanıt sunmuyor. Aksine, veri setindeki markaların ortalama kârlılık açısından birbirine oldukça benzer bir dağılım sergilediğini gösteriyor.

Sayılar Bize Ne Anlatıyor?

Veri setini tanıdıktan sonra sayısal değişkenlerin nasıl dağıldığını incelemek istedim. Çünkü ortalama değerler tek başına yeterli olmayabilir; verinin hangi aralıkta yoğunlaştığını, aykırı değerlerin bulunup bulunmadığını ve dağılımın simetrik olup olmadığını görmek daha doğru yorumlar yapmayı sağlar.

Bu nedenle indirim oranları , indirim tutarı, kâr ve kâr marjı yüzdesi değişkenlerinin dağılımlarını görselleştirdim. Grafiklerde kırmızı kesikli çizgi ortalamayı, yeşil kesikli çizgi ise medyanı göstermektedir.



Grafikler incelendiğinde her değişkenin farklı bir dağılım yapısına sahip olduğu görülüyor.

İndirim oranı dağılımında indirim oranlarının belirli seviyelerde toplandığı dikkat çekiyor. Özellikle %0, %5, %10, %15, %20 ve %25 gibi standart kampanya oranlarının daha sık kullanıldığı görülüyor. Bu durum, işletmenin rastgele indirimler yerine belirli kampanya politikaları uyguladığını düşündürüyor.

İndirim tutarı grafiğinde dağılım sağa çarpık bir yapı sergiliyor. Siparişlerin büyük bölümünde düşük tutarlı indirimler uygulanırken, az sayıdaki siparişte oldukça yüksek indirimler dikkat çekiyor. Ortalama değerin medyandan daha yüksek olması da bu yüksek indirimlerin ortalamayı yukarı çektiğini gösteriyor.

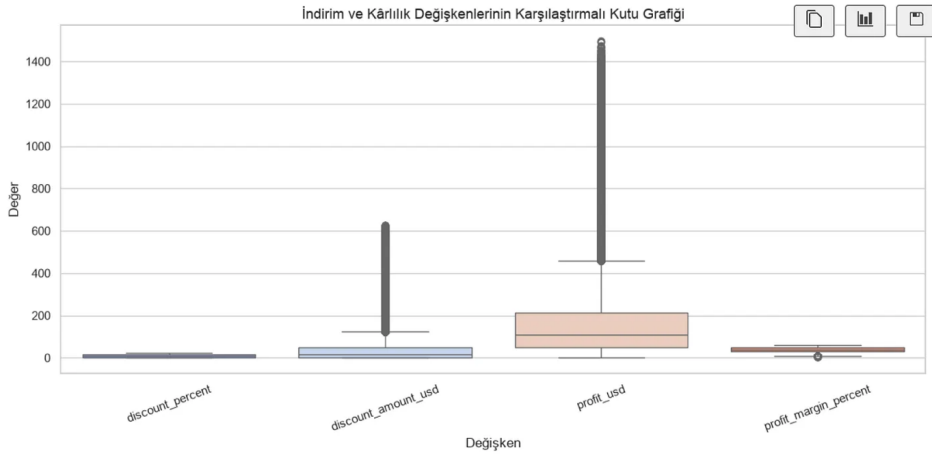
Benzer durum Profit dağılımında da görülüyor. Siparişlerin büyük kısmı düşük ve orta seviyelerde

kâr üretirken, az sayıdaki yüksek kârlı sipariş dağılımın sağ tarafında uzun bir kuyruk oluşturuyor. Bu nedenle ortalama kâr değeri, tipik bir siparişi temsil etmek yerine yüksek kârlı işlemlerden etkilenmiş olabilir.

Buna karşılık kâr marjı yüzdelerinin dağılımı diğer grafiklere göre daha dengeli görünüyor. Ortalama ve medyan değerlerinin birbirine oldukça yakın olması, kâr marjı yüzdelerinin daha simetrik dağıldığını ve bu değişkende aykırı değer etkisinin daha sınırlı olduğunu gösteriyor.

Bu analiz, veri setindeki tüm sayısal değişkenlerin aynı davranışı göstermediğini ortaya koyuyor. Özellikle indirim tutarı ve kâr değişkenlerinde sağa çarpık dağılımlar bulunurken, kâr marjı daha dengeli bir yapı sergiliyor. Bu nedenle sonraki analizlerde yalnızca ortalama değerlere odaklanmak yerine, medyan ve dağılım özelliklerini de dikkate almak daha sağlıklı sonuçlar elde edilmesini sağlayacaktır.

Dağılımları histogramlarla inceledikten sonra, değişkenlerdeki aykırı değerleri ve yayılımı daha net görebilmek için kutu grafiğinden yararlandım.



Kutu grafiği, histogramlarda dikkat çeken bulguları destekler nitelikte sonuçlar ortaya koyuyor. Özellikle indirim tutarı ve kâr değişkenlerinde kutunun üst kısmında çok sayıda aykırı değer bulunduğu görülüyor. Bu durum, veri setindeki siparişlerin büyük bölümünün benzer seviyelerde gerçekleştiğini, ancak az sayıdaki yüksek indirimli ve yüksek kârlı siparişlerin dağılımı önemli ölçüde etkilediğini gösteriyor.

Buna karşılık indirim oranı değişkeninin daha dar bir aralıkta dağıldığı dikkat çekiyor. Bu da işletmenin indirim oranlarını belirli kampanya seviyelerinde tuttuğunu ve çok farklı indirim oranları uygulamadığını düşündürüyor.

Kâr marjı yüzde değişkeni ise diğer değişkenlere göre daha dengeli bir görünüm sergiliyor. Kutunun daha kompakt olması ve aykırı değer sayısının sınırlı kalması, kâr marjlarının belirli bir aralıkta yoğunlaştığını gösteriyor. Bu da histogram analizinde gözlemlediğimiz, ortalama ile medyanın birbirine

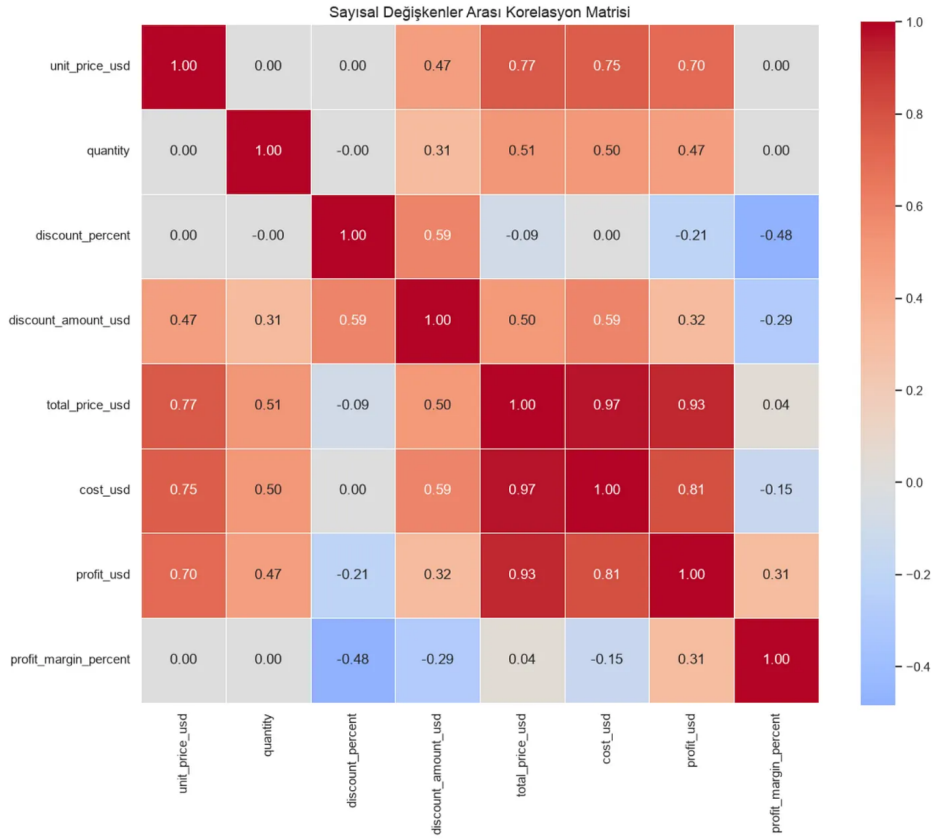
yakın olduğu dağılım yapısını destekliyor.

Kutu grafiği sayesinde yalnızca ortalama değerlere bakmanın yeterli olmadığı bir kez daha görülüyor. Özellikle kâr ve indirim tutarı değişkenlerinde yer alan aykırı değerler, ortalama istatistikleri önemli ölçüde etkileyebileceği için analizlerde medyan ve dağılım özelliklerini de dikkate almak daha doğru sonuçlar elde edilmesini sağlayacaktır.

Değişkenler Birbirini Ne Kadar Etkiliyor?

Veri setindeki bazı değişkenleri inceledikten sonra, bazı sayısal değişkenlerin birbirleriyle nasıl bir ilişki içerisinde olduğunu görmek istedim. Bunun için sayısal değişkenler arasındaki korelasyonları gösteren bir ısı haritası oluşturdum.

Korelasyon katsayısı 1'e yaklaştıkça güçlü pozitif, -1'e yaklaştıkça güçlü negatif, 0'a yaklaştıkça ise zayıf veya ilişkisiz bir bağlantıyı ifade eder.

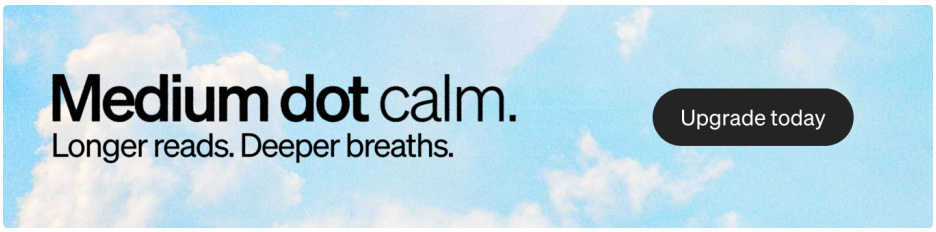


İlk dikkat çeken nokta, satış tutarı ile maliyet arasındaki 0.97 seviyesindeki oldukça güçlü pozitif ilişkidir. Benzer şekilde satış tutarı ile kâr arasında da 0.93 gibi yüksek bir korelasyon bulunuyor. Bu sonuç, satış tutarı arttıkça maliyetin ve elde edilen kârın da doğal olarak arttığını gösteriyor.

Bir diğer dikkat çekici ilişki ise birim fiyat ile satış tutarı arasındaki 0.77 ve Kâr arasındaki 0.70 seviyesindeki pozitif korelasyonlar oldu. Daha yüksek fiyatlı ürünlerin genel olarak daha yüksek satış tutarı ve daha yüksek kâr oluşturduğu söylenebilir.

İndirim değişkenleri incelendiğinde ise farklı bir tablo ortaya çıkıyor. indirim oranı ile kâr marjı yüzdesi arasında -0.48 seviyesinde orta düzeyde

negatif bir ilişki bulunuyor. Bu durum, indirim oranı arttıkça kâr marjının azalma eğiliminde olduğunu gösteriyor. Benzer şekilde indirim oranı ile kâr arasında da zayıf negatif bir ilişki (-0.21) dikkat çekiyor.



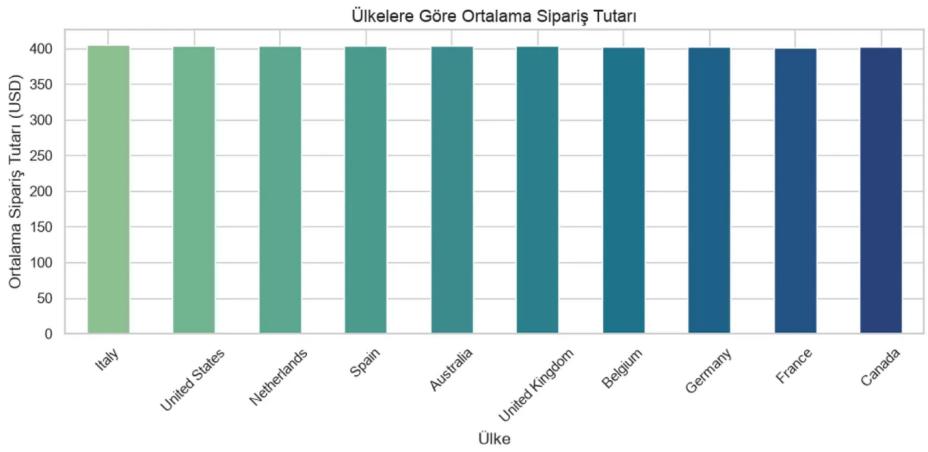
Öte yandan miktar değişkeninin sipariş tutarı (0.51) ve kâr (0.47) ile pozitif ilişki göstermesi beklenen bir sonuç olarak değerlendirilebilir. Daha fazla ürün satıldıkça toplam satış ve elde edilen kârın da artması doğal bir durumdur.

Genel olarak korelasyon matrisi, veri setindeki ilişkilerin büyük ölçüde beklendiği yönde ilerlediğini gösteriyor. Ancak burada önemli bir noktayı da unutmamak gerekir: Korelasyon, nedensellik anlamına gelmez. İki değişken arasında güçlü bir ilişki bulunması, birinin diğerine doğrudan neden olduğu anlamına gelmez. Bu nedenle korelasyon analizi, değişkenler arasındaki ilişkileri keşfetmek için güçlü bir araç olsa da, sonuçların farklı analizlerle desteklenmesi gerekir.

Ülke Faktörü Satışları Ne Kadar Etkiliyor?

Veri setini ülke bazında incelediğimde, farklı ülkelerdeki sipariş davranışlarının değişip değişmediğini görmek istedim. İlk olarak ortalama

sipariş tutarlarını karşılaştırdım. Ardından yalnızca sipariş tutarına odaklanmak yerine kâr, teslimat süresi ve müşteri sadakati gibi metrikleri de birlikte değerlendirdim.

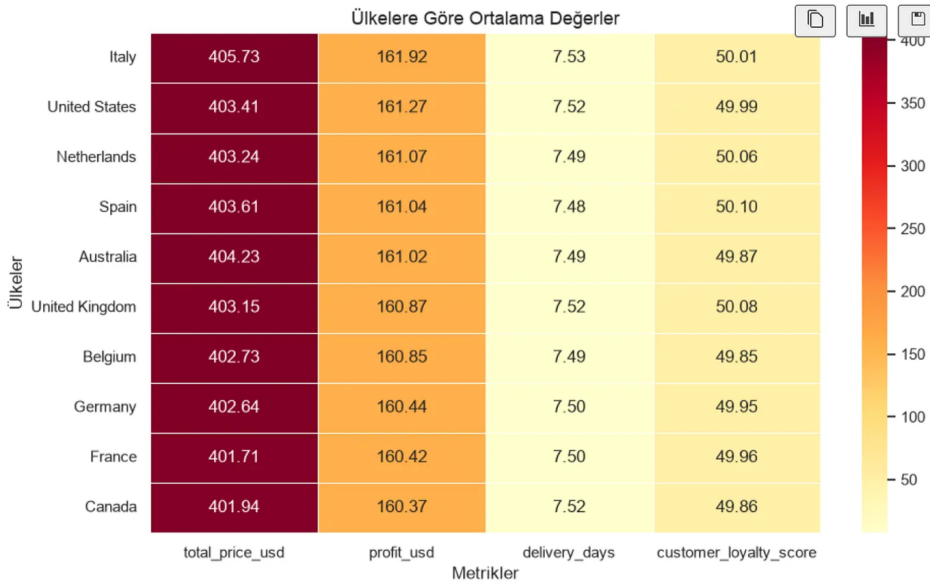


İlk grafik incelendiğinde ülkeler arasındaki ortalama sipariş tutarlarının birbirine oldukça yakın olduğu görülüyor. İtalya listenin ilk sırasında yer alırken, onu ABD, İspanya ve Hollanda çok küçük farklarla takip ediyor. Listenin son sırasında bulunan Fransa ve Kanada gibi ülkeler ile en yüksek değere sahip ülke arasındaki fark ise gerçekten düşük seviyelerde kalıyor.

Bu durum, veri setindeki sipariş tutarlarının ülkelere göre önemli ölçüde değişmediğini gösteriyor. Başka bir ifadeyle, tek başına ülke bilgisi ortalama sipariş tutarını açıklayan güçlü bir değişken gibi görünmüyor.

Bu gözlemi desteklemek amacıyla ülkeleri yalnızca satış tutarı açısından değil, farklı performans

göstergeleri üzerinden de karşılaştırdım.



Isı haritası incelendiğinde de benzer bir tablo ortaya çıkıyor. kâr, teslimat süresi ve müşteri sadakat puanı değerleri ülkeler arasında oldukça dengeli bir dağılım sergiliyor. Örneğin ortalama kâr yaklaşık 160–162 USD, teslimat günü 7,5 gün, müşteri sadakat puanı ise 50 seviyelerinde yoğunlaşıyor.

Bu sonuçlar, veri setinde yer alan ülkelerin operasyonel performans açısından birbirine oldukça benzer özellikler taşıdığını gösteriyor. Dolayısıyla bu veri seti üzerinde yalnızca ülke değişkenine bakarak “en iyi pazar” veya “en kârlı ülke” gibi güçlü çıkarımlar yapmak doğru olmayacaktır.

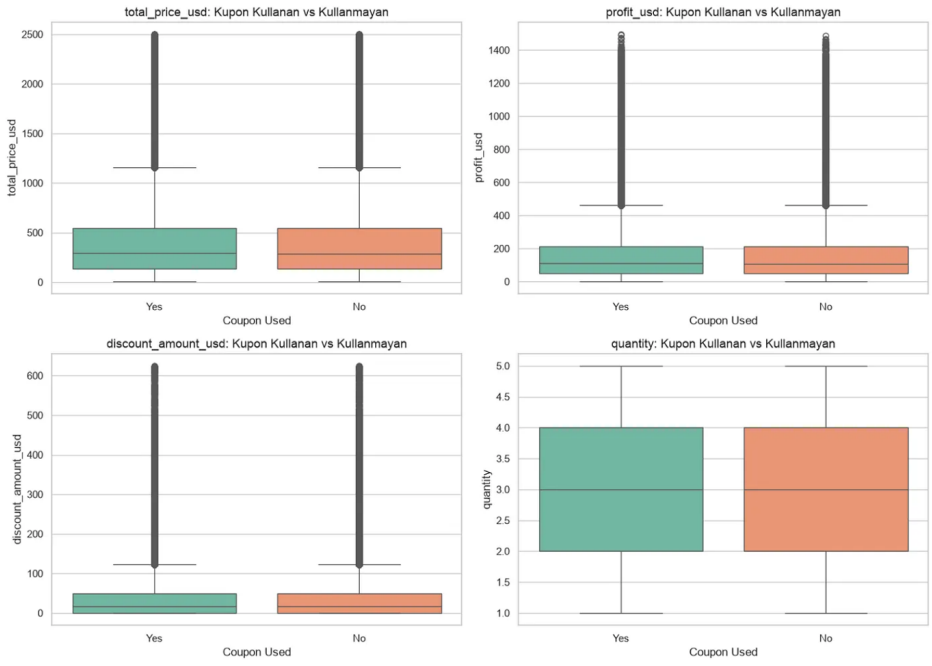
Aslında bu da önemli bir bulgu. Çünkü keşifsel veri analizinin amacı yalnızca büyük farklılıkları ortaya çıkarmak değil, hangi değişkenlerin anlamlı bir ayrışma oluşturmadığını da belirlemektir. Bu analiz

sonucunda ülke değişkeninin, incelenen metrikler açısından belirleyici bir faktör olmadığı görülmektedir.

Kupon Kullanımı Müşteri Davranışını Değiştiriyor mu?

E-ticaret platformlarında kupon kampanyaları, müşterileri daha fazla alışveriş yapmaya teşvik etmek amacıyla sıkça kullanılan stratejilerden biridir. Bu nedenle veri setinde kupon kullanan ve kullanmayan müşteriler arasında anlamlı bir farklılık olup olmadığını incelemek istedim.

Karşılaştırmayı yaparken yalnızca sipariş tutarına değil; sipariş tutarı, kâr, indirim tutarı ve miktar değişkenlerini birlikte değerlendirdim.



Grafikler incelendiğinde kupon kullanan ve kullanmayan müşterilerin dağılımlarının büyük

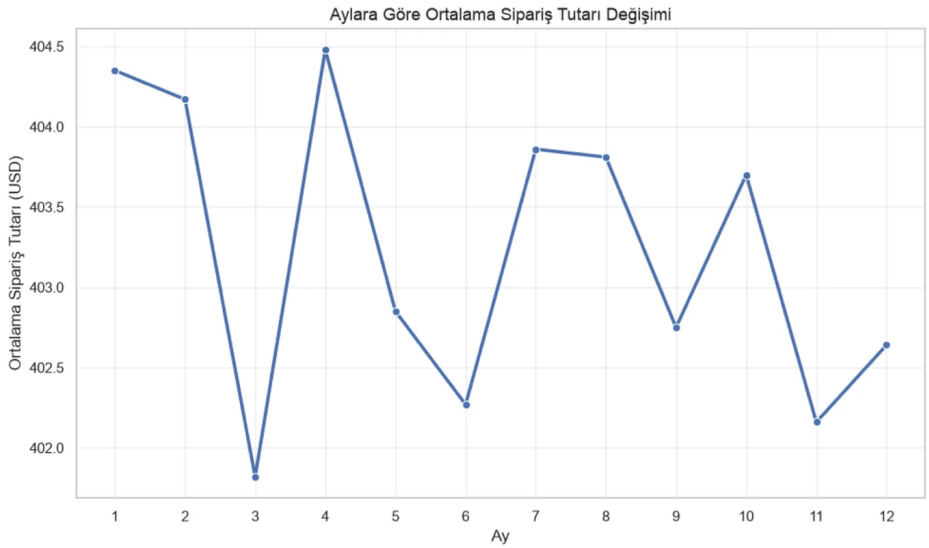
ölçüde birbirine benzediği görülüyor. Sipariş tutarı, elde edilen kâr ve satın alınan ürün adedi açısından kutu grafiklerinin medyan değerleri ile çeyrekler arası dağılımları oldukça yakın seviyelerde yer alıyor.

İndirim tutarında da benzer bir durum dikkat çekiyor. Kupon kullanan müşterilerin daha yüksek indirim aldığı beklense de, veri setinde her iki grubun dağılımı büyük ölçüde benzer görünüyor. Ayrıca tüm değişkenlerde çok sayıda aykırı değer bulunması, bazı siparişlerin genel dağılımdan önemli ölçüde ayrıldığını gösteriyor.

Bu sonuçlar, incelenen veri setinde kupon kullanımının tek başına sipariş tutarı, kârlılık veya satın alınan ürün miktarı üzerinde belirgin bir farklılık oluşturmadığını düşündürüyor. Elbette bu, kupon kampanyalarının etkisiz olduğu anlamına gelmez. Kuponların başarısı; kampanyanın türü, müşteri segmenti, ürün kategorisi veya kampanya dönemi gibi birçok farklı faktörden etkilenebilir.

Sipariş Tutarları Yıl Boyunca Değişiyor mu?

Bir sonraki adımda ortalama sipariş tutarının aylar içerisinde nasıl değiştiğini incelemek istedim. Böylece belirli dönemlerde satışların arttığı veya azaldığı sezonluk bir eğilim olup olmadığını gözlemlemeyi amaçladım.



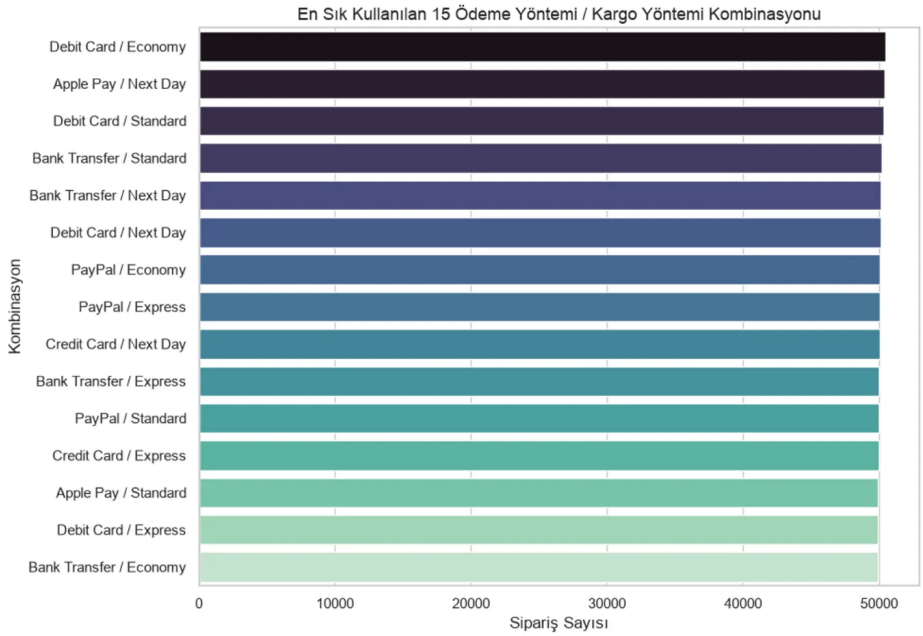
İlk bakışta çizgi grafiği aylar arasında belirli dalgalanmalar varmış gibi görünse de, değerler yakından incelendiğinde değişimin oldukça sınırlı olduğu görülüyor. Y ekseni dar bir aralıkta (yaklaşık 402–405 USD) gösterildiği için küçük değişimler görsel olarak daha büyük dalgalanmalar şeklinde algılanabiliyor. Bu nedenle grafiği yorumlarken yalnızca çizginin hareketine değil, eksen değerlerine de dikkat etmek gerekiyor.

Genel olarak grafik, veri setinde belirgin bir mevsimsellik veya aylık satış trendi bulunmadığını gösteriyor. Ortalama sipariş tutarı yıl boyunca oldukça istikrarlı bir seyir izliyor ve herhangi bir ayın diğerlerinden belirgin şekilde ayrışmadığı görülüyor.

Bu sonuç, incelenen veri setinde sipariş tutarlarının zaman içinde büyük değişiklikler göstermediğini ve ayların tek başına satış performansını açıklayan güçlü bir değişken olmadığını düşündürüyor.

Müşterilerin En Çok Tercih Ettiği Ödeme ve Kargo Kombinasyonları Hangileri?

Bir e-ticaret platformunda müşterilerin yalnızca ne satın aldığı değil, satın alma sürecini nasıl tamamladığı da önemlidir. Bu nedenle ödeme yöntemi ile teslimat seçeneğini birlikte ele alarak en sık tercih edilen kombinasyonları incelemek istedim.



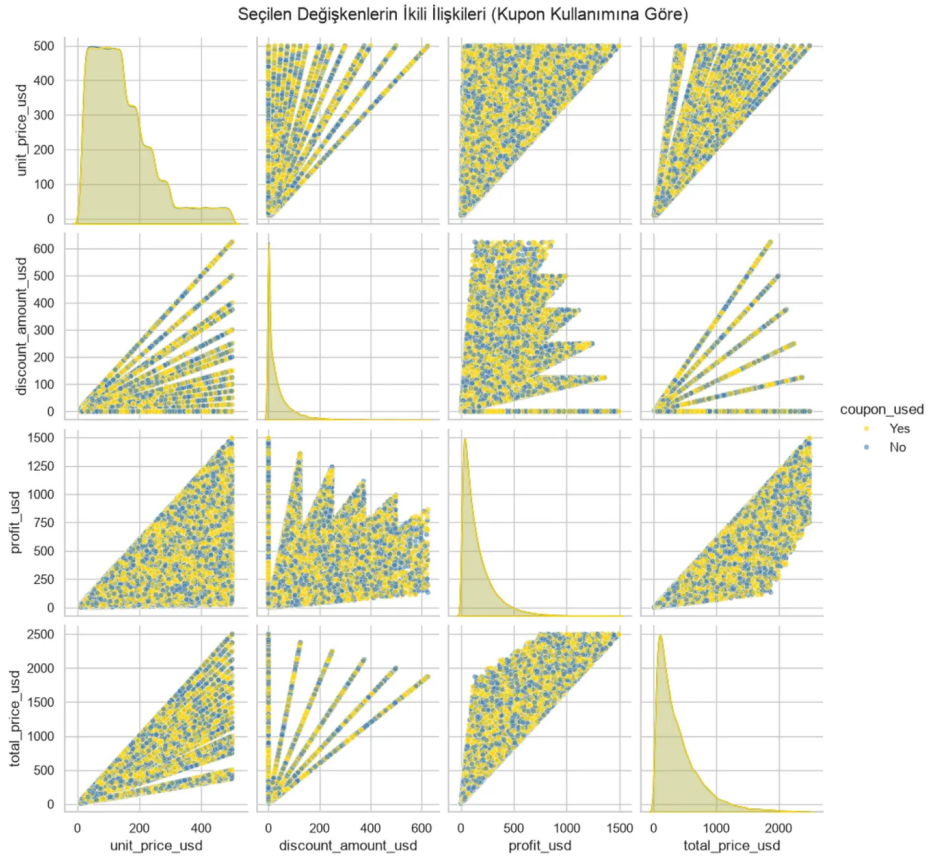
Grafik incelendiğinde ilk 15 ödeme yöntemi ve kargo kombinasyonunun sipariş sayılarının birbirine oldukça yakın olduğu görülüyor. Listenin başında Debit Card / Economy yer alsa da, onu takip eden Apple Pay / Next Day, Debit Card / Standard, Bank Transfer / Standard ve diğer kombinasyonlar arasında yalnızca küçük farklar bulunuyor.

Bu durum, müşterilerin belirli bir ödeme yöntemi veya teslimat seçeneğine yoğun şekilde yönelmediğini gösteriyor. Aksine, ödeme ve kargo

tercihleri veri setinde oldukça dengeli bir dağılım sergiliyor.

Değişkenler Birlikte Nasıl Davranıyor?

Veri setindeki değişkenleri inceledikten sonra, bu kez aralarındaki ilişkileri aynı anda gözlemlemek istedim. Bunun için İkili İlişki Grafiği kullanarak seçilen sayısal değişkenlerin birbirleriyle olan dağılımlarını ve kupon kullanımına göre nasıl farklılaştıklarını analiz ettim.



Grafik incelendiğinde ilk dikkat çeken nokta, kupon kullanan ve kupon kullanmayan siparişlerin neredeyse tamamen iç içe geçmiş olmasıdır. Sarı ve mavi noktaların belirgin kümeler oluşturmaması,

kupon kullanımının tek başına sipariş davranışını ayıran güçlü bir değişken olmadığını gösteriyor.

Değişkenler arasındaki ilişkiler incelendiğinde ise sipariş tutarı ile kâr darasında güçlü bir pozitif ilişki olduğu görülüyor. Sipariş tutarı arttıkça elde edilen kârın da arttığı açıkça gözlemlenebiliyor. Bu bulgu, önceki bölümde incelediğimiz korelasyon matrisiyle de uyumludur.

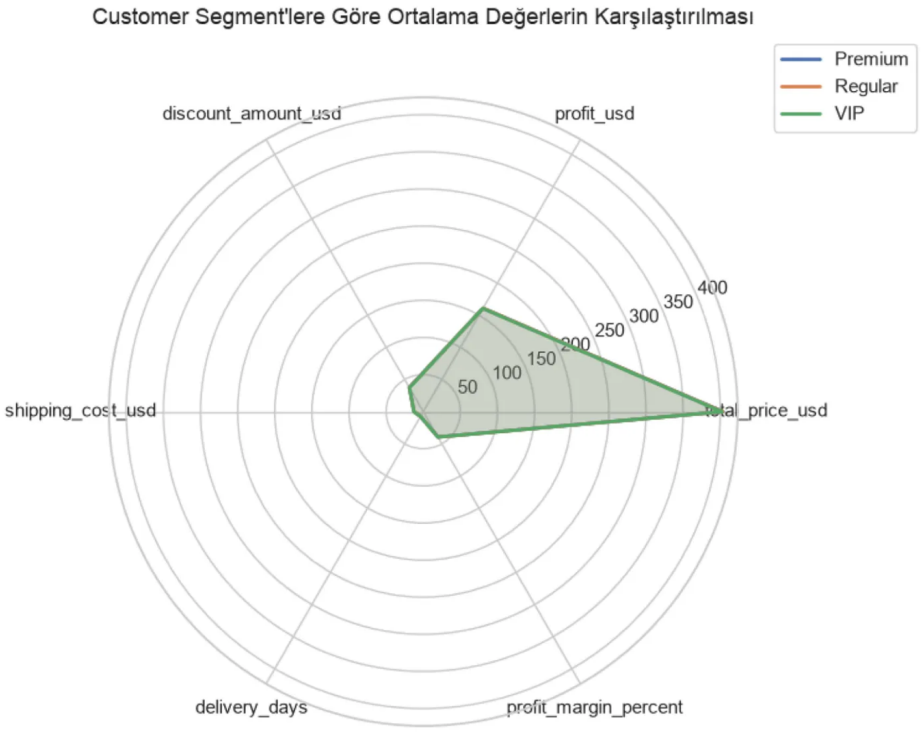
Benzer şekilde birim fiyat yükseldikçe hem toplam sipariş tutarının hem de kârın artma eğiliminde olduğu görülüyor. Buna karşılık indirim tutarı ile diğer değişkenler arasında doğrusal olmayan, ancak sipariş büyüklüğüne bağlı olarak artan bir yapı dikkat çekiyor. Daha yüksek tutarlı siparişlerde indirim miktarının da artması beklenen bir durum olarak değerlendirilebilir.

Grafiğin köşegeninde yer alan dağılımlar ise önceki histogram analizlerini destekliyor. Özellikle sipariş tutarı, indirim tutarı ve kâr değişkenlerinin sağa çarpık dağılıma sahip olduğu görülürken, düşük değerlerde yoğunlaşan ve az sayıdaki yüksek değerlerden oluşan uzun kuyruk yapısı tekrar gözlemleniyor.

Müşteri Segmentleri Birbirinden Gerçekten Farklı mı?

Son olarak müşteri segmentlerini tek bir grafik üzerinde karşılaştırabilmek amacıyla radar

grafiğinden yararlandım. Bu grafikte her segment için ortalama sipariş tutarı, kâr, indirim tutarı, kargo ücreti, teslimat süresi ve kâr marjıbirlikte gösterilmektedir.



İlk bakışta üç müşteri segmentinin neredeyse aynı şekli oluşturduğu görülüyor. Bu da Premium, Regular ve VIP müşterilerin ortalama değerler açısından birbirine oldukça benzer performans sergilediğini düşündürüyor.

Ancak bu grafiği yorumlarken önemli bir noktaya dikkat etmek gerekiyor. Radar grafiğinde kullanılan değişkenler farklı ölçeklere sahip olduğu için sipariş tutarı diğer metriklere göre çok daha büyük değerler alıyor ve grafiği baskılıyor. Bu nedenle teslimat süresi, kargo ücreti veya indirim tutarı gibi daha

küçük ölçekli değişkenler görsel olarak yeterince ayırt edilemiyor.

Dolayısıyla bu grafik, müşteri segmentleri arasında belirgin farklılıklar olduğunu göstermekten ziyade, kullanılan metriklerin ölçek farkı nedeniyle segmentlerin birbirine oldukça benzer göründüğünü ortaya koyuyor.

Bu analiz, veri görselleştirmede yalnızca doğru grafik türünü seçmenin değil, grafiğin hangi veri ölçeğiyle oluşturulduğunun da sonuçları doğrudan etkilediğini gösteriyor.

Bu çalışma boyunca bir e-ticaret veri seti üzerinde keşifsel veri analizi gerçekleştirerek müşteri davranışlarını, sipariş süreçlerini ve işletme performansını farklı açılardan inceleme fırsatı buldum.

Analizler sonucunda bazı değişkenler arasında güçlü ilişkiler gözlemlenirken, bazı grafiklerin ise beklenen ölçüde anlamlı ayrışmalar sunmadığı görüldü. Bu da veri analizinde yalnızca dikkat çekici sonuçlara değil, anlamlı olmayan sonuçların da doğru şekilde yorumlanmasının önemli olduğunu gösterdi.

Bu proje sayesinde veri temizleme, görselleştirme, dağılım analizi, korelasyon analizi ve farklı grafik türlerinin yorumlanması konularında pratik yapma fırsatı elde ettim.

İlerleyen çalışmalarda bu veri seti üzerinde makine öğrenmesi modelleri geliştirerek satış tahmini, müşteri segmentasyonu ve kârlılık analizi gibi daha ileri düzey çalışmalar gerçekleştirmeyi hedefliyorum.

Data Analysis + Eda + Python ✓ Data Visualization + Pandas +

Written by Fatma Aydın

0 followers · 1 following

Edit profile

No responses yet

Fatma Aydın

What are your thoughts?

Recommended from Medium