

Bir Tekstil Perakendecisinin Müşterileri İçin RFM Modeli ile Müşteri Segmentasyonu

Customer Segmentation Analysis Based on RFM for The Customers of A
Retailer

Özet

Günümüzde teknolojinin gelişimi ile birçok sektörde olduğu gibi perakende sektöründe de gelişim ve dijital dönüşüm yaşanmaktadır. Bu çalışma, teknolojinin hızla geliştiği ve veri biliminin her sektörde önem kazandığı bir dönemde, perakende sektöründe müşteri segmentasyonunun önemini vurgulamaktadır. Özellikle tekstil perakendeciliği alanında, müşteri satın alma davranışlarının doğru bir şekilde analiz edilmesi ve segmente edilmesi, işletmelerin müşteri ilişkilerini yönetme ve pazarlama stratejilerini belirleme süreçlerinde kritik bir rol oynamaktadır. Bu çalışma, bir tekstil perakendecisinin maskelenmiş verilerini kullanarak, RFM (Recency, Frequency, Monetary) modeli ile müşteri segmentasyonu yapmayı hedeflemektedir. Veriler, ön işleme tabi tutulmuş ve RFM değerleri hesaplanmıştır. Ardından, K-means ve Fuzzy C-means algoritmaları kullanılarak müşteri kümeleri oluşturulmuştur. Oluşturulan bu kümelerin sonuçları değerlendirilmiş ve müşteri gruplarına yönelik yorumlar yapmak için kullanılmıştır. Sonuçlar, müşteri segmentasyonunun, perakende sektöründe müşteri davranışlarını anlamak, müşteri ilişkilerini yönetmek ve etkili pazarlama stratejileri geliştirmek için ne kadar önemli olduğunu göstermektedir. Bu çalışma, perakende sektöründeki işletmelerin, müşteri verilerini etkin bir şekilde kullanarak, müşteri memnuniyetini artırmak ve işletme performansını iyileştirmek için nasıl stratejiler geliştirebileceğine dair değerli bilgiler sunmaktadır.

Abstract

With the development of technology, as in many sectors, the retail sector is also experiencing development and digital transformation. This study emphasizes the importance of customer segmentation in the retail sector in a period where technology is rapidly developing, and data science is gaining importance in every sector. Especially in the field of textile retailing, the correct analysis and segmentation of customer purchasing behaviors play a critical role in managing customer relationships and determining marketing strategies for businesses. This study aims to perform customer segmentation using the masked data of a textile retailer with the RFM (Recency, Frequency, Monetary) model. The data has been preprocessed and RFM values have been calculated. Then, customer clusters were created using K-means and Fuzzy C-means algorithms. These clusters were evaluated to make comments on customer groups. The results show how important customer segmentation is to understand customer behaviors in the retail sector, manage customer relationships, and develop effective marketing strategies. This study provides valuable information on how businesses in the retail sector can develop strategies to increase customer satisfaction and improve business performance by effectively using customer data.

Giriş

Müşteri İlişkileri Yönetimi, bir şirketin müşterileriyle karşılıklı fayda sağlayan ve güçlü bağlantılar kurma yeteneğini ifade eder. Bu, bir şirketin tüketicilerle etkileşime geçerken benimsediği stratejileri ve prosedürleri kapsar. Müşterinin bir ürünü satın almadan önce, satın alma sırasında ve sonrasında deneyimlediği tüm aşamaları bu kavram içerisinde değerlendirebiliriz.

Birçok şirket, müşteri talepleri ve ihtiyaçları hakkında bilgi edinmek için Müşteri İlişkileri Yönetimi (CRM) gibi, şirketin rekabetçi pazarını güçlendirecek bilgileri maksimum düzeyde

Samet Kanca

Yüksek Lisans Öğrencisi, İstanbul Aydın
Üniversitesi, İstanbul, Türkiye
samet.kanca@gmail.com
Orcid No: <https://orcid.org/0009-0004-0027-2978>

Tuncay Özcan

Doç. Dr., İstanbul Teknik Üniversitesi, İstanbul,
Türkiye, tozcan@itu.edu.tr
Orcid No: <https://orcid.org/0000-0002-9520-2494>

Yakup Çelikkilek

Doç. Dr., İstanbul Aydın Üniversitesi, İstanbul,
Türkiye, yakupcelikkilek@aydin.edu.tr
Orcid No: <https://orcid.org/0000-0003-0585-1085>

Article Type / Makale Türü

Research Article / Araştırma Makalesi

Keywords

Retail, RFM Analysis, Customer Segmentation.

Anahtar Kelimeler

Perakende, RFM Analizi, Müşteri Segmentasyonu.

JEL Codes: C38, C40, M31

Submitted: 08 / 08 / 2023

Accepted: 27 / 09 / 2023

kullanılmamaktadır. Ayrıca, kullanıcı verilerini madencilik ve hesaplamalı yöntemlerin kombinasyonu kullanılmış olmasına rağmen, bu yaklaşımların bazı sınırlılıkları vardır. Ayrıca, sadece birkaç çalışma kullanıcıyı anlama konusunda RFM modelini kullanmaktadır. (Bachtiar, 2018)

Pazarlama çalışmaları, yoğun rekabet ortamında ayakta kalmayı hedefleyen işletmeler açısından müşterilerin önemine sıklıkla dikkat çekmiştir. Müşteri İlişkileri Yönetimi (CRM), müşterilerle ilişkileri geliştirmeyi amaçlayan öne çıkan bir pazarlama yönetimi yaklaşımı olmuştur. CRM yaklaşımının pratik bir uygulaması, işletmeler ve müşteriler için değer çıkarmak üzere müşteri verilerinin analizidir. Bu bağlamda, müşteri segmentasyonu, benzer niteliklere sahip müşteri gruplarını belirlemeye ve müşteri grupları için daha iyi uyarlanmış pazarlama stratejilerini belirlemeye yardımcı olan bir görev olmuştur. Müşteri segmentasyonu için ortaya konulmuş çeşitli yaklaşımlar arasında RFM Modeli, etkin ve kolay uyarlanabilir olmasıyla öne çıkmaktadır. Müşterilerin satış verisine ilişkin 3 farklı boyut üzerinden sıralanmasına dayanan yöntem, sıralamada kullanılan puanlama biçimine göre çeşitli yaklaşımlara konu olmaktadır (Köse ve Arslan, 2020).

Teknolojinin ilerlemesi ve İnternet'in yaygınlaşmasıyla, adını sıkça duymaya başladığımız büyük veri kavramı ortaya çıkmıştır. Büyük veri, yapılandırılmamış veri yığını olarak kısaca tanımlanabilir. Farklı kaynaklardan toplanan veriyi anlamlı ve işlenebilir bir formata dönüştürmeyi amaçlar. Bu yöntemlerden biri RFM analizidir. Erpolat ve arkadaşları tarafından yapılan çalışmada, RFM analizinin önemine vurgu yapılmıştır. RFM analizinin tahmin modellerinde nasıl kullanılacağı detaylı olarak açıklanmıştır. RFM analizinin geri dönüşüm projesine uygulanabilirliği gösterilmiştir. Modelin işleyişi ve RFM'nin geri dönüşüm projelerine uygulanması, metal sektöründe faaliyet gösteren bir şirketin orijinal perakende verilerinde gösterilmiştir. Bu sayede, geri dönüşüme katılımı teşvik edilmiştir. Geri dönüşümün katkı oranına göre, müşterilere indirimler uygulayarak müşteri ve şirket arasında karlı bir ilişki kurulması hedeflenmiştir. (Erpolat Taşabat ve Akca, 2020) Perakende sektörünün yapısı her zamankinden daha hızlı bir şekilde evrilmektedir. Hem yerel hem de çok uluslu işletmelerden artan bir rekabet, şirket birleşmeleri, yeni edinimler ve değişimler, alışveriş deneyimleri konusunda beklentileri olan daha donanımlı ve zorlu müşteriler gözlemlemeye başladık (Dabholkar et al., 1996). Bunun doğal bir sonucu olarak, perakendeciler bugün rakiplerinden daha üstün bir hizmet sağlama konusunda kendilerini çeşitlendirecek yeni bir stratejik yönlendirmeyi benimsemelidir. Bu genellikle rekabette öne çıkmanın en etkili perakende stratejilerinden biri olarak ifade edilir (Reichheld ve Sasser, 1990).

RFM analizi üzerine yapılan başka bir çalışmada segmentasyon gerçekleştirilmiş ve sonrasında K-Means kümeleme, Fuzzy C-Means ve RM K-Means algoritmaları kullanılarak müşteriler kümelendi. Bu algoritmaların çalışma süreleri ve etkinlikleri analiz edilmiştir. Sonuç olarak, RM K-Means algoritmasının diğer tekniklere göre daha az zaman aldığı ve iterasyon sayısını azalttığı gözlemlenmiştir." (Christy, 2018)

Pazarlamacıların çoğu, başarılı kampanyalar için doğru müşterileri belirleme konusunda zorluk çeker. Şimdiye kadar, bir lansman kampanyası için uygun müşterileri seçmek için kullanılan popüler bir yöntem müşteri segmentasyonudur. Ne yazık ki, müşteri segmentasyonu ve pazarlama kampanyası arasındaki bağlantı eksiktir. Başka bir problem, veritabanı pazarlamacılarının genellikle müşteri segmentasyonunu ve müşteri hedeflemeyi gerçekleştirmek için farklı modeller kullanmasıdır. Chan (2008) tarafından yapılan bir çalışmada kampanya stratejileri için müşteri hedeflemeyi ve müşteri segmentasyonunu birleştiren yeni bir yaklaşım sunulmaktadır. Chan araştırmasında (RFM) modeli kullanarak müşteri davranışını belirler ve ardından önerilen segmente edilmiş müşterileri değerlendirmek için bir Müşteri Yaşam Boyu Değer (LTV) modeli kullanmıştır. Önerilen yöntemin verimliliğini göstermek için, bu çalışma, 4000'den fazla müşteriyi segmente etmek için bir Nissan otomobil perakendecisinin ampirik bir çalışmasını gerçekleştirmektedir. Deneysel sonuçlar, önerilen yöntemin değerli müşterileri rastgele seçime göre daha etkili bir şekilde hedefleyebileceğini göstermektedir. (Chan, 2008)

Bilgi teknolojisinin büyük bir gelişme göstermesi, veri miktarında bir patlamaya neden olmuştur, ancak bu verinin faydalı içgörüler elde etmek için işlenmesi gerekmektedir. Müşteri İlişkileri Yönetimi (CRM) olarak da adlandırılan daha iyi ilişkiler kurmak için müşterinin

ihtiyaçlarını, davranışlarını ve değerlerini incelemek için verinin kullanılması gerekmektedir. Şirket büyüdükçe, veri artmakta ve müşterilerle doğrudan etkileşimde bulunmak daha zor hale gelmektedir. Bu nedenle, daha az etkili pazarlama kampanyaları gibi sorunlar, hemen ele alınmazsa zararlara neden olabilir. Bu bağlamda, recency, frequency ve monetary (RFM) değişkenleri kullanılarak ve dirsek yöntemi ve siluet skoru ile küme sayısını belirleyerek K-Means kümeleme ile müşteri segmentasyonu gerçekleştirilmiştir. Analiz sonuçlarına göre, en iyi müşteriler, kaybedilmeyebilecek müşteriler ve ortalama müşteriler olmak üzere üç tür küme bulunmuştur. (Barus et al., 2023)

Müşteriler, son zamanlarda alışveriş yaptıkları bir şirketi ve markasını gelecekteki alışverişler için daha kolay hatırlarlar. Son zamanlarda bir şirketten ödeme yapmış olan tüketicilerin, aylar hatta daha uzun süre boyunca şirketten alışveriş yapmamış olan müşterilere kıyasla gelecekte başka bir alışveriş yapma olasılıkları daha yüksektir. Bu tür bilgiler, son müşterileri geri dönmeye ve daha fazla para harcamaya teşvik etmek için kullanılabilir. Eski müşterilerin izini kaybetmemek için, onlara son işleminden bu yana bir süre geçtiğini hatırlatırken aynı zamanda başka bir alışveriş yapmaları için bir teşvik sağlamak üzere pazarlama çabaları yapılmalıdır. Miglautsch'a göre (2000), müşteri kazanma ve iletişimde bilimsel yöntem, doğrudan pazarlamanın kalbidir. Ana nokta, müşterinin önceki alışveriş geçmişine dayanarak bir takip çağrısına layık olup olmadığıdır. Bu soru, doğrudan posta, kataloglar, telefon, saha ve internet dahil olmak üzere her tür iletişim için geçerlidir. Bu seçimi yapma prosedürüne müşteri segmentasyonu denir. (Miglautsch, 2000).

RFM analizi, müşteri segmentasyon sürecinin merkezinde yer alarak, şirketlerin geçmiş ve değişen satın alma davranışlarına duyarlı bir şekilde belirli hedef kitlelere ulaşmalarını sağlar. Bu yaklaşım, kampanyaların etkinliğini ve müşteri memnuniyetini artırır. RFM segmentasyonu, ekstra dikkat ve odaklanma gerektiren kitle gruplarını belirlemek için güvenilir bir yöntemdir, böylece maksimum fayda ve şirket karı artırılabilir.

RFM Analizi ile özelinde yapılan bir çalışmada, klasik RFM modeline PESTEL analizinin bir bileşeni olan "ekonomik" değişkeni eklenerek RFMS modeli önerilmiştir. Bu model, şirketlerin verimlilik süreçlerini iyileştirmeyi ve müşteri ilişkilerini daha doğru yönetecek bir sınıflandırma yöntemi geliştirmeyi amaçlamaktadır. Böylece, zaman ve maliyet tasarrufu sağlanarak müşteri ile şirket arasında karlı bir ilişki kurulması hedeflenmiştir. Önerilen modelin etkileri, BORUSANCAT Makine ve Güç Sistemleri'nin müşteri veritabanı kullanılarak analiz edilmiştir. En iyi sonucu elde etmek için farklı müşteri grupları için farklı modeller oluşturulmuş ve puanlar alınmıştır. Analiz sonucunda tekliflere olumlu potansiyele sahip olan müşteriler belirlenmiştir. Böylece, zaman ve maliyet tasarrufu sağlanmıştır. (Taşabat et al., 2023)

Perakende sektöründe kampanya yönetimi üzerine Martinez ve arkadaşları tarafından yapılan bir vaka çalışmasında geleneksel RFM modeline ve bir müşteri veritabanına uygulanan 2-tuple RFM modeline dayalı bir müşteri segmentasyonunun ne kadar kolay, doğru ve açıklanabilir olabileceği konusunda gerçek bir örnek gösterilmektedir. Bu, geleneksel model yerine 2-tuple modelin uygulanmasının faydalarını daha iyi anlamamızı sağlayacaktır. 2-tuple modelin üzerine k-ortalama kümelemesi uygulayarak, segmentlerin iş açısından büyük bir uygulanabilirliğe sahip olduğu görülmektedir. Bu örneği geliştirmenin ana hedefi, bir doğrudan kampanya iletişimi için en iyi hedefi tanımlamaktır. (González Martínez et al., 2019)

Bu çalışma, Türkiye'de ve yurt dışında 200'den fazla mağaza ile hizmet veren bir hazır giyim perakende firmasının müşteri verilerini kullanarak RFM Analizi ile müşteri segmentasyonu yapmayı hedeflemektedir. Bu analiz, firmaların müşteri tabanını daha iyi anlamalarını ve onlara daha etkili bir şekilde hizmet vermek için stratejiler geliştirmelerini sağlar. RFM Analizi, müşterilerin son alışveriş tarihlerini (Recency), alışveriş sıklıklarını (Frequency) ve toplam harcamalarını (Monetary) dikkate alarak bir segmentasyon modeli oluşturur. Bu model, müşterilerin geçmiş alışveriş davranışlarına dayalı olarak belirli kategorilere ayrılmasını sağlar. Bu çalışmanın bir parçası olarak, oluşturulan segmentler ve müşterilerin bu segmentlere nasıl dağıldığı detaylı bir şekilde incelenecek ve anlamlandırılacaktır. Bu analiz, firmanın müşteri portföyünün daha derinlemesine anlaşılmasını sağlayacak ve ayrıca pazarlama stratejilerinin daha etkin bir şekilde uygulanmasına yardımcı olacaktır. Ayrıca, bu çalışma, bir dönem firmanın müşterisi olan

ancak yakın zamanda alışveriş yapmayı bırakan ve dolayısıyla "kaybedilmiş" olarak kabul edilen müşterileri belirleme potansiyeli sunar. Bu müşterilerin belirlenmesi, firmanın bu müşterileri geri kazanma stratejileri geliştirmesine ve uygulamasına olanak sağlar. Sonuç olarak, bu çalışma, perakende firmasının müşteri ilişkileri yönetimini daha etkin bir şekilde gerçekleştirmesine yardımcı olacak ve aynı zamanda müşteri memnuniyetini ve firmanın genel performansını artırmada önemli bir rol oynayacaktır.

1. RFM Analizi

RFM (Recency, Frequency, Monetary) Analizi, müşteri segmentasyonu ve doğrudan pazarlama için kullanılan bir tekniktir. RFM, bir müşterinin ne zaman (Recency), ne sıklıkla (Frequency) ve ne kadar para harcadığını (Monetary) ölçer. Bu üç ölçüm, bir müşterinin bir işletme için ne kadar değerli olduğunu belirlemek için birlikte kullanılır. RFM analizi, müşterileri bu üç ölçüme göre sıralar ve her birini belirli bir puanla değerlendirir. Bu puanlar daha sonra bir müşteri değerlendirme skoru oluşturmak için birleştirilir. Bu skor, müşterileri segmentlere ayırmak ve her bir segmente özelleştirilmiş pazarlama stratejileri uygulamak için kullanılır. RFM, büyük miktarda veriden önemli müşterileri ayırmak için üç özelliği kullanan bir modeldir (Sarvari, 2016)

Wan ve arkadaşları tarafından yapılan çalışmada RFM Analizi'nin müşteri segmentasyonu ve müşteri yaşam değeri tahmini gibi alanlarda etkili bir araç olduğunu belirtmiştir. Çalışmalarında, RFM Analizi'nin, müşteri davranışlarını daha iyi anlamak ve müşteri ilişkileri yönetimini (CRM) geliştirmek için kullanılabileceğini göstermişlerdir. Ayrıca, RFM Analizi'nin, müşteri sadakatini ve satışları artırmak için pazarlama stratejilerini geliştirmeye yardımcı olabileceğini belirtmişlerdir. Sonuçlarına göre, RFM Analizi'nin, müşteri segmentasyonu ve hedefleme stratejilerinin geliştirilmesinde önemli bir rol oynayabileceği sonucuna varmışlardır. (Wan et al., 2022)

RFM modeli, genellikle doğrudan gelecekteki sonuçları tahmin etmek yerine çıkarımsal bilgiler sağlamak için kullanılır (Yang, 2004).

Yang'a (2004) göre RFM Analizi'nin güçlü yönleri aşağıdaki gibidir:

1. Basitlik ve Eyleme Geçirilebilirlik: RFM, müşteri davranışlarını anlamak ve gelecekteki eylemleri tahmin etmek için basit ve eyleme geçirilebilir bir yöntemdir. Bu, RFM'nin genellikle pazarlama ve satış ekipleri tarafından kolayca anlaşılabilen bir model olmasını sağlar (Yang, 2004).
2. Esneklik: RFM kodlaması oldukça keyfidir ve hem artıları hem de eksileri vardır. İyi bir yanı, ampirik yöntemin genellikle esnek bir kodlama sistemi ile her zaman çalışabilir olmasıdır. Bu, RFM'nin terk edilmemesi gerektiğini, ancak pazar çabalarının gücünü artırmak için önemli bir öngörücü olarak kullanılması gerektiğini gösterir (Yang, 2004).
3. Gelecekteki Yanıt Olasılığını Çıkarabilme: RFM prensibine uygun olarak, V (Değer) geçmiş değeri temsil eder ve gelecekteki yanıt olasılığını çıkarır. Bu, RFM'nin müşteri davranışlarını ve gelecekteki eylemleri tahmin etme yeteneğini güçlendirir (Yang, 2004).

2. Materyal ve Metot

Bu çalışma, Türkiye'de aktif olarak faaliyet gösteren ve sektörün önde gelen isimlerinden biri olan bir hazır giyim perakende firmasının gerçek verilerini kullanmaktadır. Firma, geniş bir coğrafi alanda, 200'e yakın mağaza ile hizmet vermektedir ve aynı zamanda güçlü bir internet mağazası ile de online alışveriş imkânı sunmaktadır.

Firmadan alınan bu veriler hem fiziksel mağazalarda hem de internet mağazasında gerçekleşen satışları kapsamaktadır. Bu veriler, hangi müşterinin hangi ürünü, kaç adet ve hangi fiyatla satın aldığı gibi temel satış bilgilerinin içermektedir. Bu bilgiler, müşteri davranışlarını anlamak ve segmentasyon yapmak için oldukça değerlidir. Verilerin gizliliği ve müşteri bilgilerinin korunması açısından, tüm veriler maskelenmiştir ve sadece firma tarafından bilinen benzersiz kodlar ile paylaşılmıştır. Bu hem müşteri gizliliğini korurken hem de verilerin analiz edilmesi için gerekli olan bilgileri sağlar.

Bu çalışmanın amacı, bu verileri kullanarak müşteri segmentasyonunu oluşturmak, oluşturulan segmentlere ait müşterileri analiz etmek ve bu analizlerin sonuçlarını kampanya ve pazarlama

faaliyetlerinde kullanılmaktadır. Bu, firmanın müşteri ilişkilerini daha etkin bir şekilde yönetmesine ve pazarlama stratejilerini daha hedef odaklı hale getirmesine yardımcı olacaktır.

Tablo 1. Firmadan Alınan Ham Müşteri Satış Verileri

Müşteri Kodu	Ürün Kodu	Miktarı	Fatura Numarası	Hesap Tarihi	Fatura Tarihi	Mağaza Kodu	Satış Fiyatı	ANA GRUP	ALT GRUP
M1	U1	1	F1	2015-06-10	2021-01-14 00:00:00	Store1	89,99	D0	DJ
M2	U2	1	F2	2015-06-04	2021-01-11 00:00:00	Store2	23,85	C1	C1
M3	U3	1	F3	2015-06-05	2021-01-12 00:00:00	Store3	223,99	C0	CB
M4	U4	1	F4	2008-06-30	2021-01-15 00:00:00	Store4	200,45	D0	DD
M5	U5	1	F5	2015-06-05	2021-01-12 00:00:00	Store5	223,99	C0	CB
M6	U6	1	F6	2010-07-22	2021-01-07 00:00:00	Store6	89,99	D0	DF
M7	U7	1	F7	2010-07-22	2021-01-08 00:00:00	Store7	59,99	N0	NC
M8	U8	1	F8	2010-07-23	2021-01-04 00:00:00	Store8	79,99	40	4A
M9	U9	1	F9	2007-02-20	2021-01-06 00:00:00	Store9	89,99	C0	CB

Tablo 2. Müşteri Ham Verisine Ait Kolon İsimleri Veri Açıklamaları

Kolon Adı	Açıklama
Müşteri Kodu	Her bir müşteri özelinde verilen tekil kod.
Ürün Kodu	Her bir ürün özelinde verilen tekil kod.
Miktar	Satılan ürün adeti.
Fatura Numarası	Siparişe ait fatura numarası.
Hesap Tarihi	Müşterinin hesap oluşturma tarihi.
Fatura Tarihi	Fatura tarihi.
Mağaza Kodu	Satın alım yapılan mağaza kodu.
Satış Fiyatı	Müşteri satış fiyatı.
ANA GRUP	Satılan ürünün bulunduğu hiyerarşiye ait ana grup kodu.
ALT GRUP	Satılan ürünün bulunduğu hiyerarşiye ait alt grup kodu.

Firma tarafından paylaşılan ham verilere ait örnek bir tablo aşağıda paylaşılmıştır. Bu tablo, analiz sürecinde kullanılan verilerin bir örneğini sunmaktadır ve bu verilerin nasıl kullanıldığını ve analiz edildiğini daha iyi anlamamıza yardımcı olacaktır.

Yukarıda detayları paylaşılan veriler ile yapılacak bu çalışma, 3 Ocak 2021 tarihinden başlayarak 12 Mart 2023 tarihine kadar olan, toplamda iki yılı aşkın bir süreyi kapsayan veriler üzerinden gerçekleştirilecektir. Bu zaman dilimi, firma tarafından sağlanan ve analizlerimiz için kullanacağımız veri setinin kapsamını oluşturmaktadır. Tüm analizler, bu belirli zaman dilimine ait satışları ve müşteri davranışlarını referans alarak yürütülecektir.

2.1. Veri Ön İşleme

Müşteri segmentasyonu süreci, verinin bir dizi ön işleme adımından geçirilmesiyle başlayacaktır. Bu ön işleme adımları, verinin analize uygun bir formata getirilmesini ve tutarsızlığın giderilmesini içerir. Bu aşama, verinin temizliğini ve kalitesini sağlamak için kritik öneme sahiptir, çünkü bu faktörler sonuçların doğruluğunu ve güvenilirliğini doğrudan etkileyecektir.

Ön işleme kapsamında, her bir müşteri için Sıklık (Recency), Frekans (Frequency) ve Harcama (Monetary) değerleri de hesaplanmıştır. Sıklık, müşterinin son satın almasının ne kadar önce gerçekleştiğini; Frekans, müşterinin belirli bir zaman dilimi içinde ne kadar sıklıkla satın alma yaptığını; Harcama ise müşterinin toplamda ne kadar harcama yaptığını gösterir. Çalışma kapsamında Frekans değeri için veri içerisindeki OrderCount yani sipariş verme miktarı kullanılmıştır. Verideki müşterilerin satın alma sıklığı düşük olduğundan frekans değerinin kümeleme içerisinde etkisi istenilen katkıyı modele sağlayamamaktadır. Bununla birlikte yine ön işleme kapsamında her bir müşterinin satın alımı için Harcama değerini tutarlı hesaplayabilmek için geçmişten günümüze aylık olarak TÜİK tarafından açıklanan enflasyon oranları bulunmuş ve bulunan bu enflasyon oranları satış fiyatları üzerine geçmişe doğru yansıtılmıştır. Bu sayede daha tutarlı bir Harcama değeri oluşturulması hedeflenmiştir. Ön işlemeden geçirildikten sonra verinin deseni aşağıdaki gibi olmuştur.

Tablo 3. Ön İşlemeden Geçirilmiş Olan Veri Deseni

Müşteri Kodu	recency	monetary	frequency	receipt_date_max	receipt_date_min	order_count	time_gap
M1	330	3248,482127	0,01754386	2022-04-16 00:00:00	2021-08-31 00:00:00	4	228
M2	1	57114,66313	0,042721519	2023-03-11 00:00:00	2021-06-17 00:00:00	27	632
M3	2	22969,85177	0,041450777	2023-03-10 00:00:00	2021-01-27 00:00:00	32	772
M4	20	22598,49523	0,023578363	2023-02-20 00:00:00	2021-03-01 00:00:00	17	721
M5	141	14213,04377	0,029644269	2022-10-22 00:00:00	2021-06-03 00:00:00	15	506
M6	546	1560,907832	0,020661157	2021-09-12 00:00:00	2021-01-13 00:00:00	5	242
M7	589	7185,102242	0,082125604	2021-07-31 00:00:00	2021-01-05 00:00:00	17	207
M8	113	502,8570031	0,03030303	2022-11-19 00:00:00	2022-09-14 00:00:00	2	66
M9	6	22378,97907	0,051175657	2023-03-06 00:00:00	2021-03-13 00:00:00	37	723
M10	443	4604,979074	0,020080321	2021-12-24 00:00:00	2021-04-19 00:00:00	5	249

Ön işleme süreci sonrasında yukarıda deseni verilen veri için K-ortalama ve Bulanık C-ortalama kümeleme algoritmaları kullanılarak müşteriler farklı kümelerle ayrılacaktır. Kümelerle ayırma işlemi sonrasında ise kümelerin başarımları değerlendirilecektir. Yapılacak değerlendirme de Silhouette indeks kullanılacaktır. Bu algoritmalara ve yöntemlere ilişkin bilgiler alt başlıklar halinde verilmiştir.

2.2. K-Ortalama

K-ortalama algoritması, yaygın bir kümeleme yöntemidir ve en az iki parametre gerektirir: küme sayısı (k) ve uzaklık ölçüm metodu. Algoritma, veri noktalarının küme merkezlerine olan uzaklıklarını hesaplar ve kümeleri belirler. Bu işlem, belirlenen durma kriterlerinden biri karşılanana kadar tekrarlanır. Durma kriterleri genellikle küme merkezlerinde bir değişiklik olmaması veya belirli bir iterasyon sayısına ulaşılmasıdır. K-ortalama, hızlı ve etkili bir yöntem olmasına rağmen, başlangıç merkezlerinin rastgele seçilmesi ve küme sayısının doğru belirlenmesi gibi faktörler sonuçları etkileyebilir. Bu nedenle, birden fazla çalışma yapılması ve en iyi sonucun seçilmesi gerekmektedir (James et al., 2013).

$$E = \sum_{i=1}^k \sum_{p \in C} d(M_i, x)^2 \quad (1)$$

2.3. Bulanık C-Ortalama

Bulanık C-ortalama (FCM) algoritması, genellikle bulanık kümeleme problemlerini çözmek için kullanılan bir yöntemdir. Bu algoritma, K-ortalama algoritmasının bir varyasyonu olup, veri noktalarının birden fazla kümeye ait olabileceğini kabul eder. FCM algoritması, en az iki parametre gerektirir: küme sayısı (c) ve bulanıklık derecesi (m). Algoritma, her bir veri noktasının her bir kümeye aidiyet derecesini hesaplar ve bu aidiyet derecelerinin ortalamasını alarak küme merkezlerini belirler. Ardından, veri noktalarının küme merkezlerine olan uzaklıkları hesaplanır ve aidiyet dereceleri güncellenir. Bu işlem, belirlenen durma kriterlerinden biri karşılanana kadar tekrarlanır. FCM algoritmasında genellikle iki farklı durma kriteri kullanılır. Birincisi, küme merkezlerinin yeniden hesaplanması ve aidiyet derecelerinin güncellenmesi sonucunda bir önceki iterasyonla veya belirlenen iterasyon boyunca değişiklik olmaması durumunda algoritmanın durdurulmasıdır. İkincisi, belirli bir iterasyon sayısına ulaşıldığında algoritmanın durdurulmasıdır. FCM algoritması, genellikle etkili bir kümeleme yöntemi olmasına rağmen, başlangıç merkezlerinin ve bulanıklık derecesinin seçimi sonuçları etkileyebilir. Bu nedenle, farklı başlangıç noktaları ve bulanıklık dereceleri ile birden fazla çalışma yapılması ve en iyi sonucun seçilmesi gerekmektedir (Baykasoğlu et al., 2018).

$$J_m = \sum_{i=1}^N \sum_{j=1}^c u_{ij}^m \|x_i - c_j\|^2, 1 < m < \infty \quad (2)$$

2.4. Silhouette İndeksi

Hem K-ortalama hem de Bulanık C-ortalama algoritmalarında olduğu gibi en iyi sonucu belirlemek için birden çok çalışma yapılması ve yapılan bu çalışmalarının

sonucunun sistematik olarak değerlendirilmesi gerekmektedir. Bu nedenle Silhouette indeksi ile kümelerin performansı değerlendirilecektir. Silhouette indeksi, kümeleme performansını ölçmek için kullanılan bir metriktir. Hem K-ortalama hem de Bulanık C-ortalama gibi kümeleme algoritmalarının sonuçlarını değerlendirmek için yaygın olarak kullanılır. Silhouette indeksi, bir veri noktasının kendi kümesi içindeki diğer noktalara olan uzaklığına (bireysel uyum) ve en yakın diğer kümeye olan uzaklığına (bireysel yabancıklık) dayanır. Silhouette Index, her bir veri noktası için bir değer hesaplar. Bu değer, -1 ile 1 arasında bir aralıktadır. Bir veri noktasının Silhouette indeksi değeri 1'e yakınsa, bu, veri noktasının kendi kümesine iyi bir şekilde uyduğunu ve diğer kümelere uzak olduğunu gösterir. Eğer değer -1'e yakınsa, bu, veri noktasının kendi kümesine iyi bir şekilde uymadığını ve belki de başka bir kümeye ait olması gerektiğini gösterir. Eğer değer 0'a yakınsa, bu, veri noktasının kümeler arasında sınırda olduğunu ve hangi kümeye ait olduğunun belirsiz olduğunu gösterir (Starczewski ve Krzyżak, 2015).

$$S(x) = \frac{b(x) - a(x)}{\max(a(x), b(x))} \quad (3)$$

3. Analizler

3.1. Çıktılar ve Öneriler

Bu çalışma kapsamında, veri ön işleme sürecinden geçirilmiş ve müşterilere ait RFM değerleri oluşturularak kümeleme için hazırlanmış bir veri seti üzerinde durulacaktır. Bu bölümde, ön işleme sürecinden geçirilen verilerin K-ortalama ve bulanık C-ortalama algoritmaları kullanılarak nasıl kümelere ayrıldığı ve bu kümelerin Silhouette Index metriği ile nasıl değerlendirildiği detaylı bir şekilde anlatılacaktır. Bu değerlendirme süreci, en başarılı kümeleme sonucunu belirlememize yardımcı olacaktır.

Firmadan alınan veri seti, 3 Ocak 2021 ve 12 Mart 2023 tarihleri arasındaki satış ve müşteri bilgilerini içermektedir. Bu veri seti incelendiğinde, toplamda 1.907.413 tekil müşteri olduğu görülmektedir. Ancak, daha etkili ve anlamlı kümeler oluşturabilmek için, bu iki yıllık dönem içerisinde sadece bir veya iki kez alışveriş yapan müşteriler veri setinden çıkarılmıştır. Bu dönem içerisinde sadece bir kez alışveriş yapmış olan tekil müşteri sayısı 1.260.324, sadece iki kez alışveriş yapmış olan müşteri sayısı ise 363.451'dir. Bu müşterilerin çıkarılmasının ardından, RFM verileri oluşturulan ve kümeleme analizine dahil edilen müşteri sayısı 283.638 tekil kişiye düşmüştür. Bu çalışmada gerçekleştirilen tüm analizler, bu 283.638 tekil müşteri üzerinden yapılmıştır. Bu süreç, veri setinin daha yönetilebilir bir boyuta indirgenmesini ve kümeleme analizinin daha anlamlı ve yorumlanabilir sonuçlar üretmesini sağlamıştır.

283.638 tekil müşteri verisiyle çalıştırılan k-ortalama ve bulanık c-ortalama algoritmalarının kümeleme başarımları Silhouette indeksi ile değerlendirilmiştir. Her iki algoritmada farklı sayıda kümeler için tekrar çalıştırılmış ve küme sayılarındaki değişimin indeksi değeri değerlendirilmiştir. Her iki algoritmanın da farklı küme sayıları ile alınan en iyi sonuçlar aşağıdaki tabloda verilmiştir.

Tablo 4. K-ortalama ve Bulanık C-Ortalama Algoritmalarının Küme Başarımlarının Karşılaştırılması

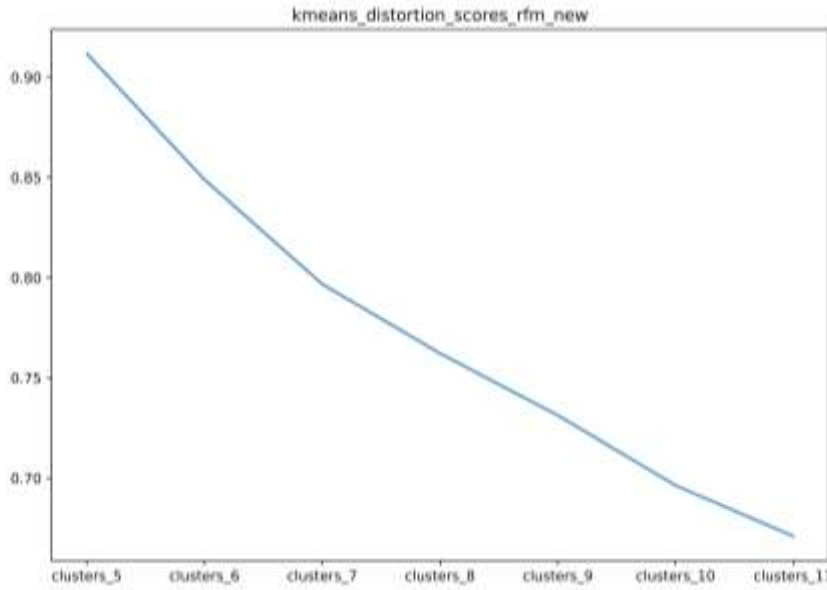
	5	6	7	8	9	10	11
K-Ortalama	0,295437	0,302827	0,306449	0,309285	0,305203	0,310002	0,310525
Bulanık C-Ortalama	0,300608	0,290042	0,299928	0,299751	0,298535	0,294106	0,299999

Tabloda belirtilen değerlerin incelenmesi sonucunda, K-ortalama algoritması için en uygun küme sayısının 8 olduğu sonucuna varılmıştır. Bu sonuç, başarımların 8 kümelik çözüm sonrasında düşüş göstermesi temel alınarak belirlenmiştir. Başka bir deyişle, 8 kümelik çözüm, algoritmanın başarımlarını maksimize ederken, daha fazla küme sayısı algoritmanın başarımlarını düşürmüştür. Bu nedenle, K-ortalama algoritması için en uygun küme sayısı 8 olarak belirlenmiştir.

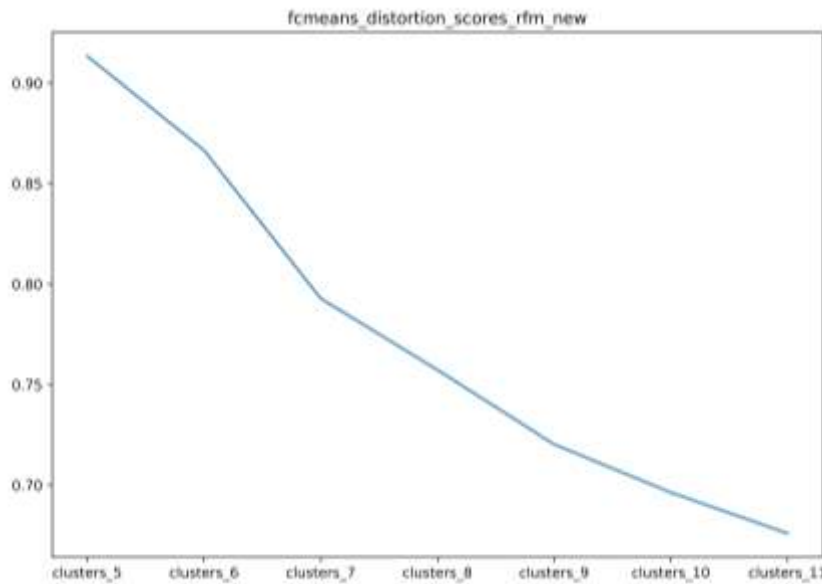
Öte yandan, Bulanık C-ortalama algoritması için en uygun küme sayısı 5 olarak belirlenmiştir. Bu karar, başarımların 5 kümelik çözüm sonrasında düşüş göstermesi temel alınarak verilmiştir. Yani, 5 kümelik çözüm, Bulanık C-ortalama algoritmasının başarımlarını maksimize ederken, daha fazla küme sayısı algoritmanın başarımlarını düşürmüştür. Bu nedenle, Bulanık C-ortalama algoritması için en uygun küme sayısı 5 olarak belirlenmiştir. Bu seçimler, her iki algoritmanın başarımlarını maksimize etmek için yapılmıştır. Başarımların düşüş gösterdiği küme sayısı, her iki algoritma için en uygun küme sayısını belirlememize yardımcı olmuştur. Bu yaklaşım, hem K-ortalama hem de Bulanık C-ortalama algoritmalarının en etkili şekilde kullanılmasını sağlamıştır.

Küme sayısının seçimi noktasında küme içi uzaklık değeri de değerlendirilmiştir. Hem K-ortalama hem de bulanık C-ortalama için küme içi uzaklık değerleri aşağıda verilmiştir.

Başarımların optimum olarak belirlendiği k-ortalama 8 ve bulanık c-ortalama 5 kümeleri üzerinden yapılmasına bu performans kriterlerine bakılarak karar verilmiştir.



Şekil 1. K-Ortalama Algoritmasının Farklı Küme Sayılarının Toplam Küme İçi Uzaklıkları



Şekil 2. Bulanık C-Ortalama Algoritmasının Farklı Küme Sayılarının Toplam Küme İçi Uzaklıkları

3.2. K-ortalama ile RFM Analizi

Kümeleme analizinin başarımını değerlendirmek için kullanılan Silhouette indeksi ve kümelerin genel başarım değerleri dikkate alındığında, K-ortalama algoritması için en uygun küme sayısının sekiz olduğu sonucuna varılmıştır. Bu sonuç, hem Silhouette indeksinin en yüksek değerini veren küme sayısı hem de genel başarım değerlerinin en yüksek olduğu küme sayısı olarak belirlenmiştir. Bu sonuca dayanarak, sekizli küme çözümü üzerinde daha detaylı analiz ve değerlendirme çalışmaları gerçekleştirilmiştir. Bu analizler, sekizli küme çözümünün her bir kümesinin karakteristik özelliklerini, müşteri davranışlarını ve alışveriş alışkanlıklarını daha iyi anlamamızı sağlamıştır. Sekizli küme çözümüne ait veriler ve bu verilerin analiz sonuçları aşağıdaki tabloda sunulmuştur. Bu tablo, her bir kümenin RFM değerlerini, müşteri sayısını ve diğer önemli özelliklerini içermektedir.

Tablo 5. K-Ortalama Algoritması Sekizli Küme Sonuçları

No	monetary	monetary_rank	order_count	order_count_rank	recency	recency_rank
0	6.258,32	5	6,061	7	298,545	5
1	2.152,45	2	3,470	2	439,112	8
2	8.876,04	7	3,673	5	361,866	6
3	2.952,34	3	3,480	3	405,021	7
4	2.103,06	1	3,516	4	273,812	3
5	4.008,94	4	3,771	6	70,516	1
6	7.160,83	6	3,326	1	289,083	4
7	18.410,77	8	9,522	8	98,572	2

K-ortalama algoritması ile yapılan RFM kümelemede ortaya çıkan müşteri kümeleri özelinde müşterilerin davranışlarına ve firma ile olan alışveriş süreçlerine yönelik yorumlar yapılabilmektedir. Bu kümelerin en dikkat çekenlerinden biri de 7 numaralı küme olmuştur. Bu küme 18.410 birimlik harcama ortalamasına sahip ve sıklığı 98 olarak görülen sadık ve yüksek harcama yapan müşteri grubunu ifade etmektedir. Bu grup aynı zamanda en yüksek sipariş sıklığı ortalaması olan 9,5 oranına sahiptir.

1 numaralı küme, dikkatlice incelendiğinde, sıklık değeri 439 olan bir müşteri grubunu temsil etmektedir. Bu değer, bu grubun potansiyel olarak kaybedilmiş müşterileri içerdiğini göstermektedir. Ancak, bu grubun harcama ortalamasına baktığımızda, oldukça düşük bir değer olan 2.152 birimi görüyoruz. Aynı zamanda, bu gruptaki müşterilerin sipariş ortalaması da 3.470 olarak belirlenmiştir. Bu, bu grubun düşük harcama yapan ve nadiren sipariş veren müşterileri içerdiğini göstermektedir. Bu küme özelinde müşterilerin verdikleri siparişler ürün grubu bazında incelenebilir ve müşterilerin hangi ürün gruplarına eğitim gösteriyor bu yüzden daha düşük bir katkı sağlayan grupta bulunuyorlarsa bu ürünler özelinde pazarlama faaliyetleri yürütülmesi faydalı olabilir.

Ancak, 1 numaralı kümeden daha kritik bir grup olan 2 numaralı küme de bulunmaktadır. 2 numaralı küme, sıklık değeri 361 olan bir müşteri grubunu temsil etmektedir. Bu grup, son alışverişlerinin üzerinden oldukça uzun bir süre geçmiş olan ve bu nedenle potansiyel olarak kaybedilmiş müşterileri içerir. Ancak, bu grubun harcama ortalamasına baktığımızda, 8.876 birimlik bir değer görüyoruz. Bu, bu grubun yüksek harcama yapan müşterileri içerdiğini göstermektedir. Bu iki grup arasındaki bu önemli fark, pazarlama stratejilerinin belirlenmesinde önemli bir rol oynar. 2 numaralı grubun yüksek harcama ortalaması, bu grubun tekrar alışveriş yapmalarını sağlamak için özel pazarlama stratejileri geliştirilmesi gerektiğini gösterir. Eğer bu müşterilerin tekrar alışveriş yapmaları sağlanabilirse, firmanın potansiyel kazancı 1 numaralı kümeye göre çok daha yüksek olacaktır. Bu, müşteri ilişkileri yönetimi ve müşteri geri kazanma stratejilerinin önemini vurgular.

K-ortalama algoritması tarafından oluşturulan kümeler arasında, özellikle 5 numaralı grup dikkat çekmektedir. Bu müşteri grubu, firma bünyesinde en yakın zamanda alışveriş yapmış olan müşterileri temsil etmektedir. Bu grubun son alışverişlerinin üzerinden geçen ortalama süre sadece 70 gün olarak belirlenmiştir. Bu, bu grubun müşterilerinin firma ile son derece güncel bir ilişkisi olduğunu göstermektedir. Aynı zamanda, bu müşteri grubunun sipariş ortalaması da 3,771 olarak belirlenmiştir. Bu, bu müşterilerin analiz dönemi boyunca en az ortalama 3,771 kez alışveriş yaptığını göstermektedir. Ancak, bu grubun harcama ortalamasının 4.008 birim olduğu görülmektedir. Bu, bu müşteri grubunun sadık ve yakın zamanda mağazalara gelmiş olmasına rağmen, harcama tutarlarının ortalamasının altında olduğunu göstermektedir. Bu analiz sonuçları, bu müşteri grubuna özel pazarlama stratejileri geliştirilmesi gerektiğini göstermektedir. Bu müşteriler, sadık ve yakın zamanda mağazalara gelmiş olmalarına rağmen, harcama tutarlarının ortalamasının altında olması nedeniyle, harcamalarını artırmaya yönelik özel pazarlama faaliyetleri ile hedeflenmelidir. Bu, firmanın bu müşteri grubunun harcama potansiyelini daha etkin bir şekilde değerlendirmesine ve bu müşterilerden daha yüksek bir gelir elde etmesine yardımcı olabilir.

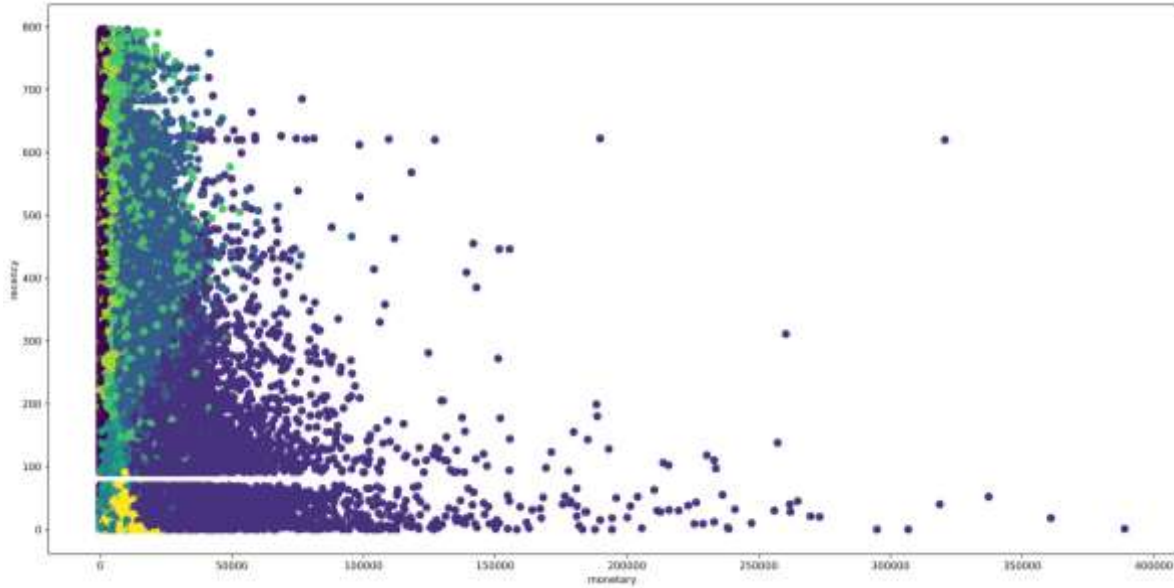
K-ortalama algoritması kullanılarak oluşturulan 6 numaralı kümeye detaylı bir inceleme yaptığımızda, bu segmentin firmanın değerli müşteri grubunu temsil ettiğini gözlemleyebiliriz. Bu segmentteki toplam harcama 7.160 birim olarak belirlenmiştir. Bu, ilk bakışta ortalama bir harcama gibi görünse de bu harcamanın gerçekleştiği toplam sipariş sayısının 3,326 olduğunu göz önünde bulundurduğumuzda durum değişiklik gösteriyor. Bu, bu müşteri grubunun mağazayı ortalama 3,326 kez ziyaret ederek toplamda 7.160 birim harcama yaptığı anlamına gelir. Bu segmenti, benzer bir harcama değerine sahip olan 0 numaralı kümeye kıyasladığımızda, en belirgin farkın toplam harcamayı oluşturan sipariş sayısında olduğunu belirleyebiliriz. Ayrıca, bu müşteri grubunun son alışverişlerinden bu yana 289 gün geçtiğini gözlemliyoruz. Bu, sepet derinliği yüksek ve sipariş başına harcama miktarı oldukça fazla olan bu müşteri grubunun, veri ortalamasına göre uzun bir süre boyunca alışveriş yapmadığı anlamına gelir. Bu durum, bu potansiyel açısından zengin müşteri grubu için özel bir pazarlama stratejisi uygulanmazsa, ilerleyen dönemlerde bu müşterilerin firmadan uzaklaşabileceği riskini ortaya koymaktadır. Alternatif bir yaklaşım olarak, 0 numaralı müşteri grubunu ele alabiliriz. Bu grup, 6.258 birimlik ortalama bir harcama tutarına sahip olup, analiz dönemi boyunca ortalama 6,061 sipariş vermiştir. Bu, bu grubun yüksek bir sipariş sıklığına sahip olduğunu göstermektedir. Ancak, bu müşterilerin son alışverişlerinin üzerinden ortalama 298 gün geçmiştir. Bu durum, bu müşteri grubunun alışveriş sıklığının yüksek olmasına rağmen, son zamanlarda firma ile etkileşime geçmediğini göstermektedir. Bu müşteri grubu da özellikle incelenmesi gereken bir grup olarak görülmektedir. Müşterilerin hem harcama tutarları ortalamaya yakın olduğu gibi hem de sipariş verme oranları yüksektir. Yani bu müşteri grubu ortalama 6 kez mağazalar veya internet mağazası üzerinden sipariş vermiş olmasına rağmen son 298 gündür herhangi bir alışveriş yapmamıştır.

3 ve 4 numaralı kümeler arasındaki benzerlikleri ve farkları derinlemesine incelemek, stratejik pazarlama kararları için oldukça aydınlatıcı olabilir. İlk olarak, bu iki kümeye genel bir bakış attığımızda, her iki grubun da ortalama harcama tutarlarının genel veri setinin ortalamasının altında olduğunu görebiliriz. Bu, her iki grubun da daha tutucu veya seçici alışveriş yaptığını gösteriyor olabilir. 3 numaralı kümenin ortalama harcama toplamı 2.952 birimken, 4 numaralı küme için bu değer 2.103 birimdir. Bu, 4 numaralı kümeyi, veri setindeki en düşük harcamaya sahip grup olarak öne çıkartmaktadır. Ancak, her iki grubun sipariş ortalamaları oldukça benzerdir: 3,48 ve 3,51. Bu, her iki grubun da benzer alışveriş sıklığına sahip olduğunu, ancak harcadıkları miktarın farklı olduğunu gösteriyor. Bu iki grubun harcama alışkanlıkları genel olarak birbirine yakın olsa da, mağazalardan yaptıkları son alışverişten bu yana geçen süre, bu iki grubu birbirinden ayıran en belirgin özelliktir. 3 numaralı gruptaki müşterilerin son alışverişlerinden bu yana 405 gün geçmiştir. Bu süre, genel veri setinin ortalamasının üzerinde olduğundan, bu müşteri grubunun potansiyel olarak kaybedilmiş müşterileri temsil ettiği söylenebilir. Öte yandan, 4 numaralı grupta bu süre 273 gün olarak belirlenmiştir. Bu, bu grubun daha yakın bir zamanda alışveriş yaptığını, ancak yine de potansiyel bir risk taşıdığını göstermektedir. Bu analizler ışığında, 3 numaralı grup için firmanın acil bir şekilde müşteri geri kazanım stratejileri geliştirmesi ve bu müşterilere özel teklifler sunması

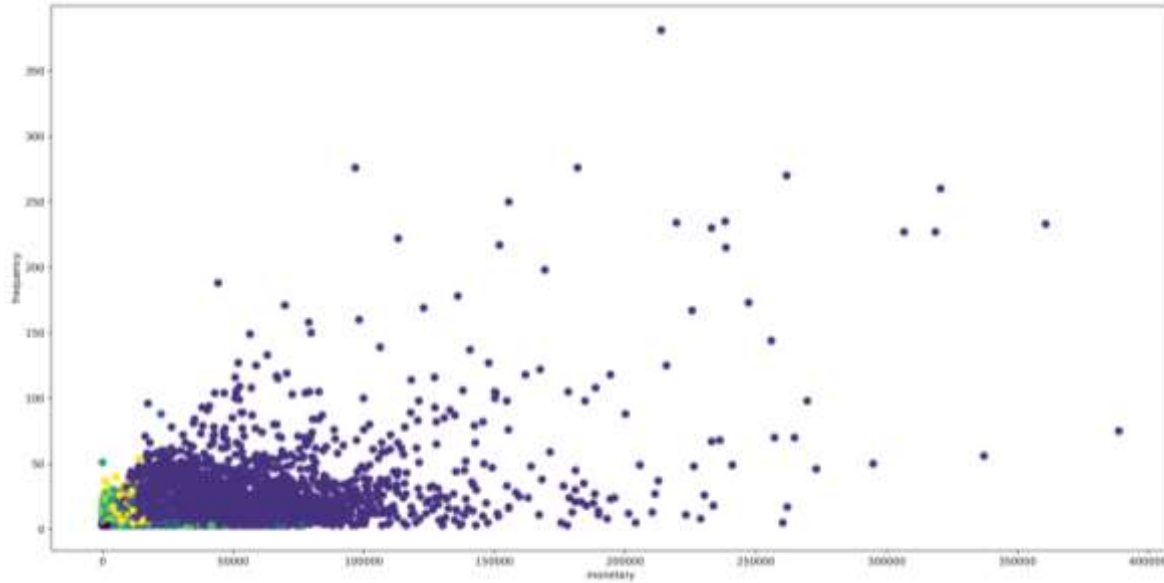
önerilir. 4 numaralı grup içinse, bu müşterilerin ilgisini sürdürebilmek ve onları daha sık alışveriş yapmaya teşvik edebilmek için özelleştirilmiş pazarlama kampanyaları ve teklifler oluşturulması faydalı olacaktır. Eğer bu adımlar atılmazsa, her iki grup da zamanla firmanın müşteri tabanından kaybolabilir.

Kümelerin oluşturulmasında RFM (Sıklık, Frekans, Harcama Tutarı) değerleri temel alınarak bir segmentasyon stratejisi benimsenmiştir. RFM değerleri, müşterilerin alışveriş sıklığını (Sıklık), en son ne zaman alışveriş yaptıklarını (Frekans) ve genel olarak ne kadar harcama yaptıklarını (Harcama Tutarı) temsil eder. Bu üç değer, müşteri davranışlarını ve alışveriş alışkanlıklarını daha iyi anlamak için kritik öneme sahiptir. Bulanık c-ortalama yönteminde olduğu gibi, k-ortalama algoritması da bu RFM değerlerini kullanarak müşterileri belirli segmentlere ayırmıştır. Bu segmentasyon süreci, her bir müşterinin hangi kümeye daha yakın olduğunu belirlemek için RFM değerlerinin birbirleriyle olan ilişkilerini dikkate almıştır.

Segmentasyon sonucunda oluşturulan kümelerin dağılımlarını daha iyi anlamak ve görsel olarak temsil etmek için çeşitli görselleştirme teknikleri kullanılmıştır. Bu görselleştirmeler, kümelerin birbirlerine göre konumlandırılmasını, hangi kümelerin benzer veya farklı özelliklere sahip olduğunu ve genel olarak segmentasyonun ne kadar başarılı olduğunu değerlendirmek için oldukça faydalıdır. Bu görselleştirmeler, metnin ilerleyen bölümlerinde detaylı bir şekilde sunulmuştur ve bu görseller, kümelerin dağılımını anlama ve yorumlama sürecinde kritik bir rol oynamaktadır.



Şekil 3. Sıklık (Recency) ve Harcama Tutarı (Monetary) Değerlerinin Küme Dağılımı



Şekil 4. Frekans (Frequency) ve Harcama Tutarı (Monetary) Değerlerinin Küme Dağılımı

3.3. Bulanık C-ortalama ile RFM Analizi

Başarımların değerlendirmesi sonrasında karar verilen 5 kümeli bulanık c-ortalama algoritması ile üretilen kümelere ait detay aşağıdaki tabloda verilmiştir.

Tablo 6. Bulanık C-Ortalama Algoritması Beşli Küme Merkezleri

No	monetary	monetary_rank	order_count	order_count_rank	recency	recency_rank
0	2.210,24	1	3,671	3	317,527	3
1	5.417,75	4	3,545	1	399,086	5
2	13.049,09	5	7,587	5	130,964	1
3	5.395,86	3	3,864	4	181,430	2
4	5.172,79	2	3,562	2	321,103	4

Bu beşli küme çözümüne özel olarak müşteri dağılımını incelediğimizde, 2 numaralı kümenin en değerli müşteri grubunu oluşturduğunu gözlemleyebiliriz. 2 numaralı küme, müşteri harcamaları ve sipariş sayısı bakımından diğer kümelere belirgin bir şekilde ayrılıyor. 2 numaralı küme içerisinde yer alan müşterilerin, incelenen veri süresince ortalama 13.049 birimlik bir harcama yaptığını görüyoruz. Ayrıca, bu müşteri grubunun ortalama olarak bu süre zarfında 7,5 sipariş verdiği belirlenmiştir. Bu, bu kümenin hem yüksek harcama yapan hem de sık sipariş veren müşterilerden oluştuğunu göstermektedir. Bununla birlikte, aynı müşteri grubunun sıklık değerine baktığımızda, en düşük değer olan 130'u görüyoruz. Bu, 2 numaralı kümenin, firmanın en sadık müşteri grubunu oluşturduğunu göstermektedir. Yani, bu müşteri grubu, firmanın en değerli müşterilerini oluşturuyor. Hem yüksek miktarda harcama yapmaları hem de sık sipariş verme eğilimleri, bu grubun firmanın gelirlerine önemli bir katkı sağladığını göstermektedir. Bu bulgular, firmanın pazarlama ve satış stratejilerini şekillendirmede önemli bir rol oynayabilir. Özellikle, 2 numaralı küme gibi yüksek değerli müşteri gruplarının ihtiyaçlarına ve tercihlerine odaklanmak, firmanın gelirlerini artırmada etkili olabilir.

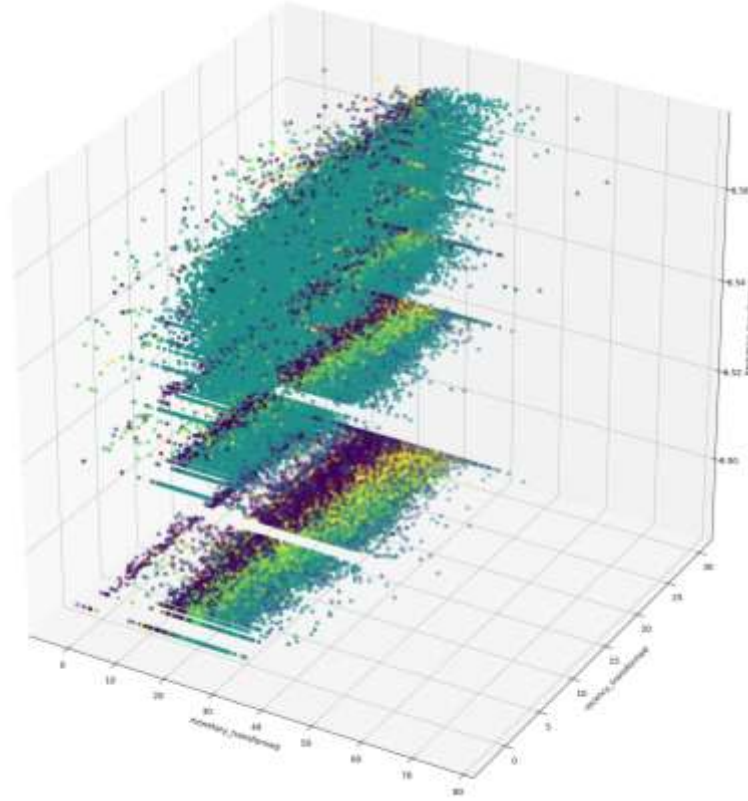
Beşli küme çözümü üzerinde dikkatle durduğumuzda, 1 numaralı kümenin de özellikle odaklanmamız gereken bir diğer önemli grup olduğunu belirtmek gerekir. 1 numaralı küme, harcama tutarı bakımından ortalama bir değere sahip olan bir müşteri grubunu temsil eder. Bu kümenin müşterileri, incelenen veri süresince ortalama 5.417 birimlik bir harcama yapmışlardır. Bu miktar, bu grubun firmanın gelirlerine ortalama bir katkı sağladığını göstermektedir. Ancak, bu kümenin sıklık (recency) değerine baktığımızda, ortalama 399 değerini görüyoruz. Bu değer, bu müşteri grubunun son alışverişlerinin üzerinden oldukça uzun bir süre geçtiğini göstermektedir. Bu

durum, bu müşterilerin potansiyel olarak kayıp müşteri olarak kabul edilebileceğini göstermektedir. Yani, bu müşteriler bir süre önce firmanın ürünlerini satın almayı bırakmış olabilirler. Bu bulgular, firmanın pazarlama stratejilerini şekillendirmede önemli bir rol oynayabilir. Özellikle, 1 numaralı küme gibi potansiyel kayıp müşteri gruplarına odaklanmak, firmanın gelirlerini artırmada etkili olabilir. Bu müşteri grubuna yönelik özel pazarlama faaliyetleri yürütmek ve bu müşterileri tekrar firmanın ürünlerini satın almaya teşvik etmek, firmanın gelirlerini artırmada önemli bir rol oynayabilir. Şekil 5'te bu beşli küme grubu özelinde üç boyutlu bir gösterim ile kümelerin dağılımı bulunmaktadır.

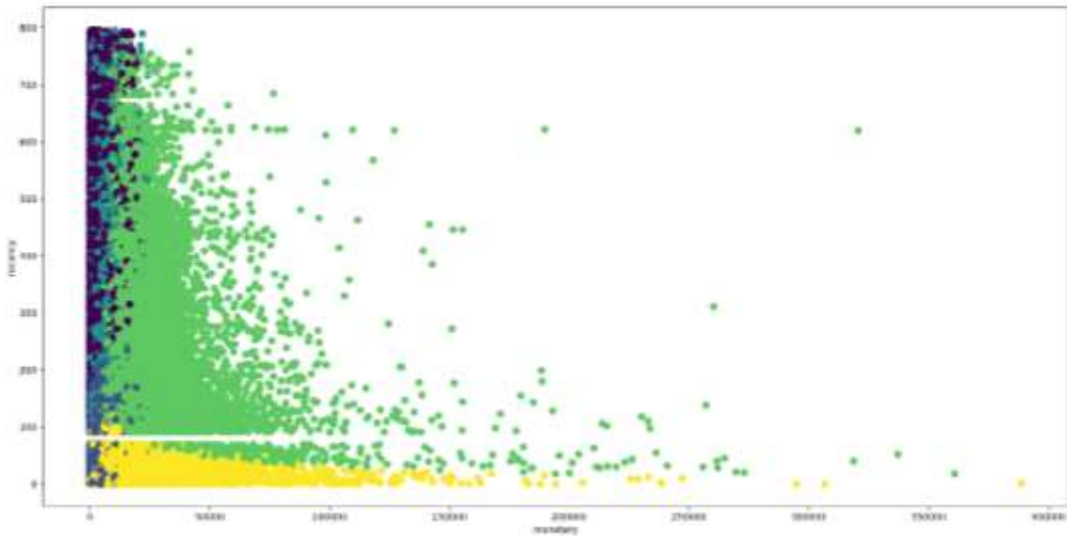
3 numaralı kümeye odaklandığımızda, bu grubun ortalama 5.395 birim harcadığını gözlemliyoruz. Bu değer, en yüksek harcamaya sahip müşteri grubunu takip eden oldukça iyi bir rakamdır. Aynı zamanda, bu müşteri grubunun son alışverişlerinden bu yana geçen süre ortalama 181 gün olarak belirlenmiştir. Bu, onları ikinci en sadık müşteri grubu yapmaktadır. Bu müşteri grubunun sipariş ortalamasını incelediğimizde, 3,86 gibi yüksek bir değerle karşılaşırız. Bu, en yüksek sipariş ortalamasına sahip olan gruptan sonraki en yüksek değeri temsil eder. Özetle, bu segment, en az 181 gün önce mağazayı ziyaret edip alışveriş yapan aktif müşterilerden oluşmaktadır. Ayrıca, bu grubun yüksek bir sipariş ortalaması vardır. Bu nedenle, bu müşteri grubunun geçmiş alışveriş trendlerini dikkate alarak, harcamalarını artırmak için özel pazarlama stratejileri geliştirmek önerilir. 4 numaralı müşteri grubunu incelediğimizde, harcama tutarı ve sipariş ortalaması açısından 3 numaralı gruba oldukça benzer olduklarını görüyoruz. Ancak, bu iki gruba birbirinden ayıran kritik bir fark bulunmaktadır: recency değeri. 4 numaralı grubun son alışverişlerinden bu yana geçen süre tam 321 gündür. Bu grup, 5.172 birimlik iyi bir harcama ortalamasına ve 3,52 sipariş ortalamasına sahiptir. Bu değerler, bu grubun firmanın değerli bir müşteri segmenti olduğunu işaret edebilir. Ancak, bu müşterilerin son 321 gündür alışveriş yapmamış olmaları, onları potansiyel olarak kaybedilmiş müşteri kategorisine taşımaktadır. Bu nedenle, bu müşteri grubunu tekrar aktif hale getirebilmek için özel stratejiler geliştirmek önemlidir.

Segmentasyonumuzun son kümesi 0 numaralı kümedir. Bu küme, harcama tutarı açısından diğerlerinden belirgin bir şekilde ayrılarak en düşük harcamalı müşteri grubunu oluşturmaktadır. Bu grubun harcama tutarı 2.210 birimdir, bu da veri setindeki en düşük harcamaya işaret eder. Bu müşterilerin son alışverişlerinden bu yana 317 gün geçmiştir. Sipariş ortalamaları 3.67 ile diğer segmentlere benzer olsa da, bu değer ve harcama tutarının kombinasyonu, bu müşterilerin ziyaret başına düşük sepet ortalamalarına sahip olduğunu gösterir. Bu müşteri grubu, hem potansiyel olarak kaybedilmiş ya da kaybedilmeye yakın müşteriler olarak görülebilir hem de düşük harcamalı müşteri kategorisindedir. Bu müşterilere yönelik olarak firmanın hem onları tekrar alışverişe teşvik eden pazarlama stratejileri geliştirmesi hem de sepet değerlerini artırmaya yönelik çalışmalar yapması önerilir. Özellikle bu müşterilerin tercih ettiği ürünlere odaklanarak, bu ürünler için özel kampanyalar ve yüksek indirimli ürünlerle müşteriyi tekrar firmanın bünyesine çekmek etkili bir yaklaşım olabilir.

Kümeleme analizinin bir parçası olarak, kümeler içerisinde hesaplanan RFM değerlerinin birbirleriyle olan ilişkileri de detaylı bir şekilde incelenmiştir. Bu değerlerin birbirleriyle olan ilişkileri, verinin yorumlanması ve müşteri dağılımının daha derinlemesine anlaşılması açısından büyük önem taşımaktadır. RFM değerleri, müşteri davranışlarının ve alışveriş alışkanlıklarının daha iyi anlaşılmasına yardımcı olur ve bu da pazarlama stratejilerinin daha etkili bir şekilde belirlenmesine olanak sağlar. Bu beşli küme çözümüne özel olarak, müşteriler için hesaplanan RFM değerlerinin birbirleriyle olan ilişkileri de görselleştirilmiştir. Bu görselleştirme, her bir kümenin karakteristik özelliklerini ve bu özelliklerin nasıl bir araya geldiğini daha iyi anlamamızı sağlar. Bu kapsamda, aşağıda RFM değerlerinin birbirleriyle olan ilişkilerini gösteren görsellere yer verilmiştir.



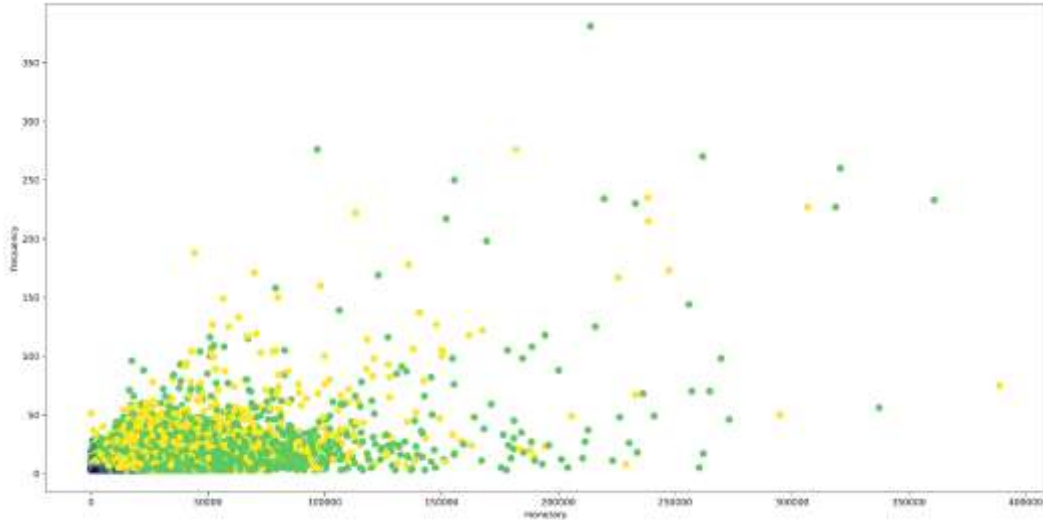
Şekil 5. Bulanık C-Ortalama Kümelerinin Üç Boyutlu Şekillendirmesi



Şekil 6. Sıklık (Recency) ve Harcama Tutarı (Monetary) Değerlerinin Küme Dağılımı

Yukarıdaki görselde görüldüğü üzere, müşterilerin büyük bir çoğunluğu 0 ile 50.000 birim arasında harcama yapmıştır. Ancak, bu değerin üzerinde harcama yapan ve alışveriş sıklığı yüksek olan müşterilerin de olduğunu gözlemlemekteyiz.

Küme dağılımlarının incelenmesinin faydalı olduğu bir başka başlık da frekans ve harcama tutarı verilerinin küme dağılımı olacaktır. Bu dağılım da aşağıdaki görselde verilmiştir.



Şekil 7. Frekans (Frequency) ve Harcama Tutarı (Monetary) Değerlerinin Küme Dağılımı

Sonuç ve Değerlendirme

Bu çalışmada, 1.907.413 adet tekil müşteriye ait 2 yıllık döneme ilişkin gerçek veriler üzerinden detaylı bir RFM analizi gerçekleştirilmiştir. Bu analiz sürecinde, müşteri verileri k-ortalama ve bulanık c-ortalama algoritmaları kullanılarak belirli kümeler halinde segmentlere ayrılmıştır. Bu iki algoritmanın kümeleme performansları dikkatlice değerlendirilmiş ve analiz edilmiştir. Her iki algoritma için de en optimum küme sayıları belirlenmiştir. Bu optimum küme sayılarına dayanarak, belirlenen sonuçlar üzerinde detaylı analizler yapılmıştır. Bu analizler, her iki algoritmanın da firmanın müşteri tabanını başarılı bir şekilde segmentlere ayırdığını ve elde edilen verinin, müşteri satış davranışlarını analiz etme ve pazarlama faaliyetlerini yönlendirme açısından son derece değerli olduğunu ortaya koymuştur. Bununla birlikte, bulanık c-ortalama algoritması ile belirlenen beşli kümeleme grubu ile k-ortalama algoritması tarafından üretilen sekizli küme karşılaştırıldığında, sekizli kümenin müşterilere yönelik daha derinlemesine bir analiz sunduğu görülmüştür. Her iki algoritma tarafından üretilen kümeler detaylı bir şekilde incelendiğinde, firmanın hem kaybettiği varsayılan hem de kaybetmeye yakın olan müşterileri ile sadık müşterilerini analiz etme noktasında önemli sonuçlar elde edildiği görülmektedir. Bu kümeler üzerinden müşterilere odaklanan ve müşteri deneyimini iyileştirmeye yönelik pazarlama stratejileri geliştirilmesi, firmanın kısa sürede önemli geri dönüşler elde etme potansiyelini artırabilir. Bu durum, müşteri odaklı pazarlama ve müşteri deneyimi süreçlerinin firmalar için ne kadar kritik olduğunu bir kez daha göstermektedir.

Gelecekteki çalışmalar için, Isolation Forests ve DBSCAN gibi yeni aykırı değer tespiti ve çıkarma teknikleri kullanılabilir. Perakende sektöründe, özellikle moda alanında, satın alma davranışı diğer perakende pazarlarına göre daha ağır olduğu için veri noktaları genellikle minimum değere yakınsar. Bu nedenle, ön işleme ve veri dönüştürme işlemlerini genişletmek kritik öneme sahiptir. Ayrıca, konu diğer kümeleme algoritmaları veya bu çalışmada kullanılan algoritmaların varyantları, örneğin k-medoids ile test edilebilir. Son öneri olarak, yeni çalışmalarda kümeleme performansını değerlendirmek için daha fazla iç kriter eklenerek geliştirilebilir. Bu tür indeks kriterleri Dunn veya Tao İndeksleri olabilir.

References

- Bachtiar, F. A., (2018) Customer Segmentation Using Two-Step Mining Method Based on RFM Model, 2018 International Conference on Sustainable Information Engineering and Technology (SIET), Malang, Indonesia, pp. 10-15, doi: 10.1109/SIET.2018.8693173.
- Barus, O. P., Nathasya, C., & Pangaribuan, J. J. (2023). The Implementation of RFM Analysis to Customer Profiling Using K-Means Clustering. *Mathematical Modelling of Engineering Problems*, 10(1).

- Baykasoğlu, A., Gölcük, İ., & Özsoydan, F. (2018). Improving fuzzy c-means clustering via quantum-enhanced weighted superposition attraction algorithm. *Hacettepe Journal of Mathematics and Statistics*, 48(3), 859-882. <https://doi.org/10.15672/hjms.2019.657>.
- Chan (2008) Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer, *Expert Systems with Applications*, Volume 34, Issue 4, May 2008, Pages 2754-2762 <https://doi.org/10.1016/j.eswa.2007.05.043>.
- Christy, A.J. (2018). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University – Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2018.09.004>
- Dabholkar, P.A., Thorpe, D.I., Rentz, J.O. (1996). A measure of service quality for retail stores: Scale development and validation. *Journal of the Academy of Marketing Science*, Vol. 24 (1), pp. 3-16.
- Erpolat Taşabat, S., & Akca, E. (2020). Recycling Project with RFM Analysis in Industrial Material Sector. *Sigma Journal of Engineering and Natural Sciences*, 38(4), 1681-1692.
- González Martínez, Carrasco, García-Madariaga, Porcel Gallego, Herrera-Viedma, (2019). A comparison between Fuzzy Linguistic RFM Model and traditional RFM model applied to Campaign Management. Case study of retail business. *Procedia Computer Science* Volume, 162, 2019, Pages 281-289 <https://doi.org/10.1016/j.procs.2019.11.286>.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning, Vol. 112, p. 18. New York: Springer.
- Köse, U. & Arslan, A. (2020). A Novel Customer Segmentation Approach Based on RFM and Clustering: A Case Study in the Retail Industry. *Gaziantep University Journal of Social Sciences*, 19(4), 1229-1248. <https://doi.org/10.17671/gazibtd.570866>.
- Miglautsch, J. R. (2000). Thoughts on RFM scoring. *Journal of Database Marketing & Customer Strategy Management*, 8(1), 67-72.
- Reichheld, F. & Sasser, W. (1990). Zero defections: quality comes to services. *Harvard Business Review*, Vol. 68 (5), pp. 105-111.
- Sarvari, P. A., Ustundag, A., & Takci, H. (2016). Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis. *Kybernetes*, 45(7), 1129-1157. <https://doi.org/10.1108/k-07-2015-0180>.
- Starczewski, A., & Krzyżak, A. (2015). Performance Evaluation of the Silhouette Index. *Artificial Intelligence & Soft Computing: 14th International Conference, ICAISC 2015, Zakopane, Poland, June 14-28, 2015, Proceedings, Part II*, 49-58. https://doi.org/10.1007/978-3-319-19369-4_5.
- Taşabat, S. E., Özçay, T., Sertbaş, S., & Akca, E. (2023). A New RFM Model Approach: RFMS. In *Industry 4.0 and the Digital Transformation of International Business* (pp. 143-172). Singapore: Springer Nature Singapore.
- Wan, S., Chen, J., Qi, Z., Gan, W., & Tang, L. (2022, April). Fast RFM model for customer segmentation. In *Companion Proceedings of the Web Conference 2022* (pp. 965-972).
- Yang, A. X. (2004). How to develop new approaches to RFM segmentation. *Journal of Targeting, Measurement and Analysis for Marketing*, 13(1), 50-60.

Extended Abstract

Aim and Scope

This study is conducted to emphasize the importance of customer segmentation in the retail sector and to evaluate the effectiveness of the methods used in this field. Particularly in the field of textile retailing, the accurate analysis and segmentation of customer purchasing behaviors play a critical role in the processes of managing customer relationships and determining marketing strategies. This study aims to perform customer segmentation using the RFM (Recency, Frequency, Monetary) model with the anonymized data of a textile retailer.

Methods

In this study, the real data of a fashion retail company actively operating in Turkey was used. The data was preprocessed and RFM values were calculated. Then, customer clusters were created

using K-means and Fuzzy C-means algorithms. The results of these clusters were evaluated by comparing the silhouette index scores and used to make comments on customer groups.

Findings

The analysis results show how important customer segmentation is for understanding customer behaviors in the retail sector, managing customer relationships, and developing effective marketing strategies. Especially, the use of K-means and Fuzzy C-means algorithms has helped to carry out customer segmentation more effectively. As a result of the analyses performed on 1,907,413 unique customers within the scope of this study, 8 clusters determined with the K-means algorithm and 5 clusters determined with the Fuzzy C-means algorithm. The identification of these segments has enabled the company to better understand its customer base and determine its marketing strategies more effectively. In particular, the segments representing potentially lost customers have played a significant role in determining the company's strategies to regain these customers.

Conclusion

In conclusion, this study has demonstrated the importance of customer segmentation in the retail sector and the effectiveness of the methods used in this field. Especially, the use of K-means and Fuzzy C-means algorithms has helped to carry out customer segmentation more effectively. However, when comparing the five-cluster group determined by the fuzzy c-means algorithm with the eight-cluster produced by the k-means algorithm, it was observed that the eight-cluster provides a more in-depth analysis of customers. When the clusters produced by both algorithms were examined in detail, it was seen that significant results were obtained in analyzing both the assumed lost and near-lost customers as well as loyal customers of the company. The results of this study can help businesses develop strategies to increase customer satisfaction and improve business performance. Also, this study can serve as a foundation for future research and contribute to further development of customer segmentation in the retail sector.