



# RFM Analizi İle Müşteri Segmentasyonu

Yazılım Mühendisliği Ana Bilim Dalı

Dönem Projesi

Ahmet Er

ORCID 0000-0000-0000-0000

Proje Danışmanı: Dr. Öğr. Üyesi Emre Şatır

Haziran 2024

# RFM Analizi İle Müşteri Segmentasyonu

## ÖZ

RFM (Recency, Frequency, Monetary) analizi, müşteri segmentasyonunu daha verimli ve hedef odaklı hale getiren bir yöntemdir. Literatürde, RFM analizi, şirketlerin ürünleriyle en aktif etkileşimde bulunan ve satın alma olasılığı en yüksek olan tüketicilere ulaşmaları konusunda rehberlik etmektedir. Bu yöntem, müşterilerin alışveriş alışkanlıklarını analiz ederek onların geçmişteki ve mevcut satın alma davranışlarına göre stratejik kararlar almalarını sağlamaktadır. RFM analizinin yanı sıra, veri madenciliği ve kümeleme algoritmaları da müşteri segmentasyonu süreçlerinde sıkça kullanılmaktadır.

Bu çalışma, online perakende veri setini kullanarak müşterilerin alışveriş davranışlarını analiz etmekte ve onları çeşitli segmentlere ayırmaktadır. RFM değerleri hesaplanarak belirlenen segmentlerin özellikleri incelenmiş ve K-Means algoritması ile müşteri grupları oluşturulmuştur. Bu segmentlerin yüzdelik dağılımları görselleştirilmiş olup, elde edilen sonuçlar işletmelere müşteri sadakati ve değerini artırmak için stratejik kararlar almada önemli bilgiler sağlamaktadır.

**Anahtar Sözcükler:** Mühendislik, teknoloji, analitik yöntem, fen bilimleri, araştırma

# Customer Segmentation with RFM Analysis

## Abstract

RFM (Recency, Frequency, Monetary) analysis is a method that makes customer segmentation more efficient and targeted. In the literature, RFM analysis guides companies on how to reach consumers who interact most actively with their products and are most likely to make purchases. This method enables strategic decision-making by analyzing customers' shopping habits based on their past and present purchasing behaviors. In addition to RFM analysis, data mining and clustering algorithms are also frequently used in customer segmentation processes.

This study analyzes customers' shopping behaviors using an online retail dataset and segments them into various groups. RFM values were calculated, and the characteristics of the identified segments were examined. Customer groups were created using the K-Means algorithm, and the percentage distributions of these segments were visualized. The results provide valuable insights for businesses to make strategic decisions to enhance customer loyalty and value.

**Keywords:** Engineering, technology, analytical method, sciences, research

# İçindekiler

Öz .....	i
Abstract .....	ii
Şekiller Listesi.....	iv
Tablolar Listesi.....	v
Kısaltmalar Listesi .....	vi
<b>1 Giriş .....</b>	<b>1</b>
<b>2 RFM Analizi ile Müşteri Segmentasyonu .....</b>	<b>3</b>
2.1 Müşteri İlişkileri Yönetimi .....	3
2.2 Veri Madenciliği .....	7
2.3 Veri Madenciliği Teknikleri .....	10
2.3.1 Müşteri ilişkileri Yönetiminde Veri Madenciliği .....	11
2.4 Müşteri Segmentasyonu .....	13
2.5 Recency / Frequency / Monetary Kavramları .....	14
2.6 RFM Analizi Nasıl Çalışır .....	16
2.6.1 RFM Analizi Kullanım Amaçları .....	16
2.6.2 RFM Analizi Metodu.....	18
2.7 Kümeleme Analizi .....	19
2.7.1 K-ortalama Kümeleme Analizi.....	20
<b>3 RFM Analizi ile Müşteri Segmentasyonu Uygulaması .....</b>	<b>22</b>
3.1 Veri Seti ve Veri Ön İşleme.....	23
3.2 RFM Skoru Hesaplama ve Müşteri Segmentasyonu .....	25
3.3 K-ortalama Analizi Uygulanması .....	30
<b>4 Sonuç .....</b>	<b>33</b>
<b>Kaynaklar .....</b>	<b>34</b>

# Şekiller Listesi

Şekil 2.1	Müşteri İlişkileri Yönetim Süreci .....	5
Şekil 2.2	Veri Tabanı Bilgi Keşfi Sürecinin Aşamaları .....	5
Şekil 3.1	Recency Veri Grafiği .....	16
Şekil 3.2	Frequency Veri Grafiği .....	16
Şekil 3.2	Monetary Veri Grafiği.....	16
Şekil 3.4	RFM Analizi ile Müşteri Segmentasyonu Pasta Grafiği.....	20
Şekil 3.5	Elbow Grafiği.....	20

# Tablolar Listesi

Tablo 3.1	Ham Müşteri Satış Verisi .....	10
Tablo 3.2	Ham Müşteri Verisine Ait Kolon Adı Ve Açıklamaları .....	91
Tablo 3.3	Ön İşlemeden Geçirildikten Sonra Müşteri Veri Deseni .....	92
Tablo 3.4	RFM Değer Tablosu.....	93
Tablo 3.5	RFM Skor Tablosu .....	93
Tablo 3.6	RFM Skoruna Göre Müşteri Segmentlerinin Analitik Özeti .....	95
Tablo 3.7	RFM Analizi Sonucunda Belirlenen Müşteri Segmentleri .....	97
Tablo 3.8	RFM Analizi Segment İlişkileri ve Normalize Değerler .....	98

# Kısaltmalar Listesi

M.İ.Y.	Müşteri İlişkileri Yönetimi
RFM	Recency Frequency Monetary
CRM	Customer Relationship Management
ORCID	Open Researcher and Contributor ID

# Bölüm 1

## Giriş

Toplumların alışveriş davranışları tarih boyunca değişiklik göstermiş ve çevresel faktörlerden birinci derece etkilenmiştir. Özellikle yeni nesil teknolojik gelişmeler alışveriş kanallarını dijital ve çevrimiçi platformlara taşımış bu da ürüne ulaşılabilirliği hızlandırırken müşteri memnuniyetini oldukça önemli kılmıştır (Erarslan, 2023). İşletmeler müşterilerini tanımadığı, onların ilgi alanlarını, ihtiyaçlarını, tercih ve taleplerini anlamadığı sürece onlarla kalıcı ilişkiler kuramaz ve onların işletmeye bağlılığını sağlayamazlar.

Müşterilerin tercihleri anlaşıldığında, işletmeler müşterilerine özel pazarlama stratejileri geliştirebilir, müşterilerinin sunulan hizmetlerden elde ettikleri memnuniyeti artırabilir ve onları tekrar satın alım yapmaya yönlendirebilirler. Müşterilerle kalıcı ilişkiler kurmak, sadık olmayan müşterilere kıyasla çok daha kârlı olan “sadık müşterileri” kazanmanın anahtarıdır (Mohammed ve Rashid, 2012).

Müşteri ilişkileri yönetimi, bir şirketin müşterileriyle derinlemesine ve karşılıklı faydalı ilişkiler kurma kapasitesini yansıtan kritik bir işlemdir. Bu, bir şirketin tüketicilerle nasıl etkileşimde bulunduğunu, hangi stratejileri benimsediğini ve hangi prosedürleri takip ettiğini belirtir. Müşterinin bir ürünü keşfetme, satın alma ve sonrasında yaşadığı deneyimin her aşaması, müşteri ilişkileri yönetiminin kapsamı içindedir (Kanca, 2023).

Günümüzde işletmeler artık fayda sağlayamayan kitlesel pazarlamadan uzaklaşarak güçlü müşteri ilişkileri geliştirmeye yönelik pazarlama tekniklerine yönelmektedirler. Müşteri İlişkileri Yönetimi (M.İ.Y.) müşterilerle ilgili detaylı bilgileri ve müşterilerin işletme içerisinde etkileşime geçtiği tüm noktaları müşteri sadakatini artırmak adına özenli ve dikkatli bir şekilde yönetme sürecidir (Kotler ve Keller, 2009). Müşterilerle uzun süreli ilişkiler kurmak, onları yakından tanımak, her biri için kişiselleştirilmiş hizmet sunmak elbette kolay değildir. Bu bağlamda işletmeler bilgi teknolojilerinin desteğini almakta ve müşteri veri tabanları oluşturmaktadır. Etkili bir M.İ.Y., müşteri



veri tabanı oluşturmaya, müşterilere yönelik trendleri, segmentleri ve bireysel ihtiyaçları ortaya çıkarmada veri madenciliği kullanmayı gerektirir (Kotler ve Keller, 2009).

Veri Madenciliği, günümüzde tüm dünyada popüler bir kavram haline gelmiştir. Fakat bu kavram tek başına bir anlam ifade etmemektedir. Toplanan verilerden anlamlar çıkartmak için farklı amaçlara yönelik birçok yöntem ve algoritma mevcuttur. Müşteri analitiği de bu verilerin kullanıldığı alanlardan biridir (Kurtaran, 2022).

RFM analizi, müşteri segmentasyonunu daha verimli ve hedef odaklı hale getiren bir araçtır. Bu analiz, şirketlere, ürünleriyle en aktif etkileşimde bulunan ve satın alma olasılığı en yüksek olan tüketicilere nasıl ulaşacakları konusunda rehberlik eder. RFM'nin müşteri segmentasyon sürecinde merkezi bir role sahip olması, firmaların, müşterilerin geçmişteki ve mevcut satın alma davranışlarına daha duyarlı stratejilerle yanıt vermesini sağlar. RFM ile müşteri segmentasyonunun gücü, şirketlere, özellikle dikkate alınması gereken ve potansiyel olarak en yüksek getiriyi sağlayabilecek müşteri gruplarını tanımlama yeteneği sunar (Kanca, 2023).

Çalışmanın birinci bölümünde yapılan çalışmanın konusuna ait genel bilgiler verilmiştir. İkinci bölümünde ise RFM analizi ile Müşteri Segmentasyonu; Müşteri ilişkileri yönetimi, veri madenciliği, müşteri segmentasyonu, RFM analizi ve kullanıldığı alanlar anlatılacak ve üçüncü bölümde RFM analizi ile müşteri segmentasyon uygulaması örneği ile devam edilecektir. Son olarak, yapılan analiz sonucunda ulaşılan sonuçtan bahsedilecek ve gelecek çalışmalar için önerilere yer verilecektir.

## Bölüm 2

# RFM Analizi İle Müşteri Segmentasyonu

Bu bölümde öncelikle müşteri ilişkileri yönetimi, veri madenciliği, müşteri segmentasyonu, RFM analizi tanımlanmakta ve kavramların öneminden bahsedilmektedir. Ardından RFM analizi ve kümeleme analizi metotları açıklanmaktadır. Bu kavramlar başlıklar halinde verilecektir.

### 2.1 Müşteri İlişkileri Yönetimi

II. Dünya Savaşı sonrası giderek küreselleşmeye başlayan dünya üzerinde işletmeler arasında hızla artan rekabet sebebiyle müşterilere devamlı suretle ekonomik, kaliteli, kişisel ürün, servis ya da hizmetler sunulması, müşteri sadakatinin ciddi biçimde azalmasına sebep olmuştur. Müşterilerini elde tutmak isteyen işletmeler satıştan elde ettikleri karı azaltma yoluna gitmişlerdir. İşletmeler için karlılığı artırmanın başlıca yolu ürün, servis ya da hizmetleri dahilinde farklılık yaratmak olsa da 20. yy sonlarında hızlanan teknolojik gelişmeler işletmelerin çoğu tarafından kabullenildiği için ürün, servis ya da hizmetlerde bunu sağlamak giderek zorlaşmıştır. Bu noktadan hareketle işletmeler için farklılık yaratarak karlılığı artırmanın en doğru yolu birebir ve doğrudan pazarlama ile müşteri sadakatini artırmak olarak karşımıza çıkmaktadır. Bu durum müşteri ilişkileri yönetiminin çıkış noktası olarak ele alınmaktadır (Yıldız, 2010).

1970'lerde bir grup Kuzey Avrupalı araştırmacı endüstriyel ve hizmet pazarlaması konularında çalışmalar yapmış; satıcı ve müşteri arasındaki ilişkiyi vurgulayan yeni bir pazarlama kavramına ihtiyaç duyulduğunu ortaya çıkarmışlardır. Yapılan çalışmalarda müşteri ilişkilerini kurmanın, güçlendirmenin ve geliştirmenin önemi vurgulanmış; bireysel ve kurumsal amaçları takip ederek müşteri ilişkilerini karlı bir şekilde ticarileştirmeye odaklanılmıştır (Takala ve Uusitalo, 1996).

90'lı yılların ikinci yarısına gelindiğinde CRM uygulaması / yazılımı geliştiren yazılım firmaları müşterilerine bütünleştirilmiş bilgi sistem çözümleri önermeye başladılar. Bu firmalar ürünlerini kullanan nihai son kullanıcılardan daha çok girdi alıp, işletmeler kullandıkları yazılım çözümlerinde müşterileri ile ilgili çok daha fazla girdi gördükçe mevcut uygulamalar / yazılımlar daha kullanışlı hale gelmiştir (Karahan, 2010). 20. yy'ın sonunda işletme, tedarikçi ve müşteri ekosisteminde geniş ölçekte internet kullanımının yaygınlaşması sebebiyle mevcut CRM uygulamaları / yazılımları işletmenin tüm paydaşları tarafından beslenen girdilerle işlev kazanmaya başlamış ve bu süreçte tespit edilen ihtiyaçlar yeni eklenen fonksiyon ve modüller yardımıyla tamamlanmaya çalışılmıştır (Yurdakul ve Karahan, 2010).

İşletmeler arası artan rekabet, müşteriye tatmin etmede yaşanan zorluklar, müşteri ile mutlak bir ilişki geliştirme gerekliliği ve sadık müşteri yaratabilme gibi gereksinimlerin tamamının işletmeler için bir iş stratejisi içerisinde görmek kaçınılmazdır. Bu noktada devreye müşteri ilişkileri yönetimi ya da CRM adı verilen iş stratejisi işletmenin her müşterisi ile öğrenme ilişkisi kurarak müşteri değerini artırmayı hedefleyen bir strateji olarak girmiştir (Rich, 2000).

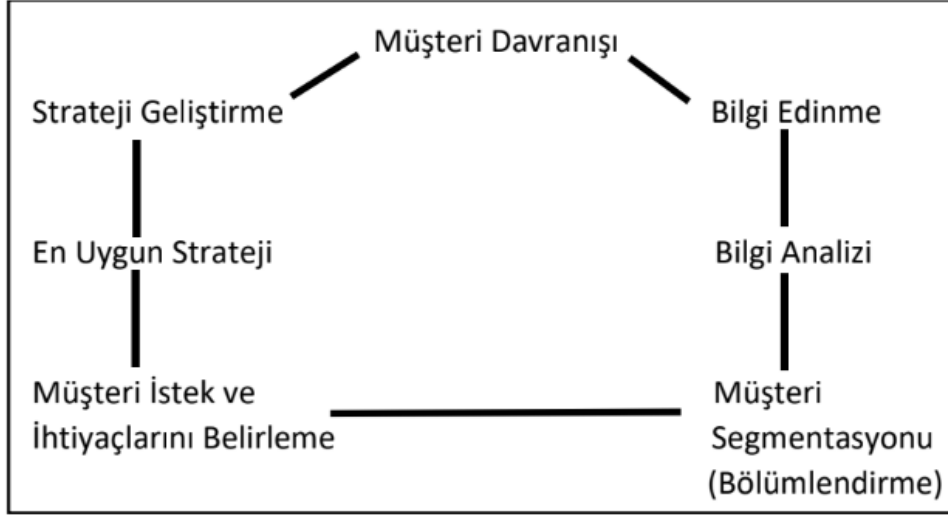
Müşteri davranışlarını daha iyi anlama ihtiyacı ve hangi müşterilerin uzun dönemde daha kârlı olacağına odaklanılması pazarlamacıların bakış açısını değiştirmiştir. Geleneksel pazarlama anlayışında ya rakip firmaların müşterileri ya da henüz satın alma yapmamış müşteriler elde edilmeye çalışılırken ve bu amaçla büyük ölçüde kitlesel reklam ve fiyat odaklı promosyonlar gerçekleştirilirken, günümüzde “müşteri edinme” hedefinden “en iyi müşterileri elde tutma” çabasına doğru bir geçiş yaşanmıştır (Winer, 2001) Bu yaklaşım “Müşteri İlişkileri Yönetimi” olarak ifade edilen yönetim felsefesinin temelini oluşturmaktadır (Aktepe, Baş, ve Tolon, 2009).

Odabaşı (2004), müşteri ilişkilerini, işletme ile müşteri arasında kurulan satış öncesi, satış anı ve sonrasındaki tüm faaliyet ve işlemleri kapsayan bir süreç olarak tanımlarken, Couldwell (1999), müşteri ilişkileri yönetimini var olan müşteri bilgilerinin kullanılması ile işletmenin kârlılığını arttırmasına ve müşteri hizmetlerini geliştirmesine olanak sağlayan bir uygulama olarak tanımlamaktadır. Bir diğer tanıma göre müşteri ilişkileri yönetimi, müşteriye odaklanan yenilikçi bir pazarlama stratejisi ve kapsamlı bir yönetim ve teknoloji metodolojisidir (Fang, Ma, 2009). Müşteri

ilişkileri yönetimi en yalın haliyle bir işletmenin sunduğu ürün, servis ya da hizmetleri doğru müşteriye, doğru zamanda, doğru şekilde ve doğru fiyat ile ulaştırmak amacıyla işletmenin sadık ve karlı müşterilerini tespit etme, kazanma, geliştirme ve uzun vadede elde tutma yolunda sergilediği faaliyetler bütünüdür (Erk, 2009). Müşteri ilişkileri yönetiminin temel mantığı, esas ilişkinin satıştan sonra kurulması, satışın tamamlanmasıyla müşterilerle olan ilişkinin bitirilmemesidir (Hacıfendioğlu ve Çolular, 2008).

CRM, işletmenin ana iş süreçlerinin otomasyonu, teknoloji altyapısı ve bilgi kaynakları vasıtasıyla satış, pazarlama, finans, girişim, hizmet, kaynak planlaması ve üretim yönetimi fonksiyonlarını her bir müşterisi ile ilişkilerini maksimum düzeye çıkarmak amacıyla entegre eder ki bu şekilde işletmenin müşterileri, iş ortakları, girişimciler, satıcılar ve iş verenler arasındaki ilişkiler düzenlenmiş olur. Bu noktada ana hedefimiz doğru bir strateji ile sadık ve karlı bir müşteri portföyü oluşturmaktır (Erk, 2009).

Müşteri ilişkileri yönetimini uygulamadan önce şirketler müşteri portföylerindeki müşterileri tanımak ve kümelere ayırmak ile işe başlamalıdır. Bu adımda önemli olan değerli ve zarar getiren müşteri kümelerinin tanımlanmasıdır. Bu uygulamadaki amaç değerli müşterileri kaybetmemek, potansiyel müşterileri kazanmak ve zarar getiren müşterileri ise işletme için karlı hale getirmeye çalışmaktır. Bu da yalnız şirketlerin müşteri ile sürekli etkileşim halinde olması ile gerçekleşebilir. Müşterilerden elde edilen verilerle bir veri ambarı oluşturulur. Veri ambarının oluşturulması ile çeşitli analizler yapılarak müşteri gereksinimlerine göre doğru yerde, doğru müşteriye, doğru zamanda ve doğru teklifler sunularak müşteri talepleri karşılanmaktadır. Bu amaçla etkili bir müşteri ilişkileri yönetimi oluşumu sağlanmaktadır (Oğur, 2023). Şekil 2.1' de müşteri ilişkileri yönetimi süreci sunulmuştur.



Şekil 2.1: Müşteri İlişkileri Yönetim Süreci (Oğur, 2023)

İşletmelerin uzun uğraş ve revizyonlar sonrası oluşturduğu müşteri ilişkilerinin taklit edilmesi zor ve maliyetli olacağı için ilerleyen yıllarda işletmeler arası rekabetin son teknolojik gelişmelerden çok müşteri ilişkileri üzerinden sürdürüleceği öngörülmektedir (Odabaşı, 1997).

Müşteri ilişkileri yönetimi, müşteriye değen tüm noktalar arasında kusursuz koordinasyon sağlayan kapsamlı bir yaklaşımdır (Goldenberg, 2002). Bu koordinasyon internet ile desteklendikçe daha da kusursuz hale gelmektedir. Müşteri ilişkileri yönetimi anlayışının işletmenin tüm bilgi sistemleri yapısı ile uyumlu bir şekilde çalışması, müşteri profillerinin elde edilmesini ve müşteri ile ilişkilerin kişiselleştirilmesini mümkün kılar. Müşteri ilişkileri yönetimi uygulamalarının başarılı olabilmesi için, müşteri ilişkileri için önemli üç ögenin birbiri ile entegre edilmesi gerekmektedir. Bu ögeler insan, süreç ve teknolojidir (Chen, Popovich, 2003). İnsan, değişkenliği nedeni ile en zor anlaşılabilir ve müşteri ilişkileri yönetimi sisteminin temelinde anlamaya çalışıldığı ama en çok hata yaptığı ögedir. Süreç ise işi yapma şeklini anlatır, müşteri ilişkileri yönetimi süreci, bilgi toplama, fiyatlandırma, satış, satış sonrası hizmetler aşamalarından oluşur ve bu aşamalar, birbirini takip eder ve süreklidir. Teknoloji ise insan ögesinin 'müşteri' olarak davranış ve kararlarını, sürecin içindeki doğru, yanlış, eksik alanları anlayabilmek için verilerin toplanmasına ve analiz edilmesine yarayan ögedir (Chen, Popovich, 2003).

Müşteri ilişkileri yönetimi uygulama süreci; müşterileri tanımlamak, müşterileri gruplamak, müşterilerle etkileşime girmek, müşterilere bire-bir hizmet sunulması olarak özetlenebilir (Yiğit, 2017).

Müşteri ilişkileri yönetiminin başarıya ulaşabilmesi için, müşteri ilişkileri yönetiminin bir iş stratejisi haline gelmesi, organizasyonun tüm birimlerince kabul edilmiş ve uygulanıyor olması gerekmektedir. Müşteri ilişkileri yönetiminin girdi olarak kullanabileceği tüm veriler, her seviyede çalışanlar tarafından bu sisteme dahil edilebilmelidir. Kullanılan otomatik araçlar kullanıcıların rahat kullanabileceği ve anlayabileceği yapıda olmalıdır. Gelişmeyi ve başarıyı gösterecek raporların yeterli ve sadece gerekli veriyi içerdiğinden emin olunması ve sistemin, organizasyonun hedefine uygun en temel ve ama basit şekilde çalıştığında emin olunması gerekir. (Anderson, Kerr, 2002)

Müşteri ilişkileri yönetimi., bilgi teknolojileri ile doğrudan ilişkili bir kavramdır (Yurdakul, 2002). Çünkü bilgi ve iletişim (bilişim) teknolojileri kullanılarak müşterilerle daha kişisel etkileşim kurmayı sağlayan bilgilerin üretilmesini sağlayan veri ve analiz sistemi bir araya getirilmektedir (King ve Burgess, 2008). Müşterilerle işletmeler arasındaki etkileşim yoluyla müşteriler hakkında toplanan bilgiler, iki taraf arasında kurulacak uzun dönemli ilişkiler için oldukça önemlidir. İşletmelerin müşterileri hakkında bilgi toplama ve analiz etmede teknoloji destekli araçlardan faydalanmaları müşterilerin kişisel zevk ve alışkanlıklarına uygun ürün geliştirmelerinde yardımcı olacaktır. Bunun için de iyi bir müşteri veri tabanı oluşturulması gerekmektedir. (Selvi, 2007).

## 2.2 Veri Madenciliği

Veri madenciliği terimi altın ve kömür madenciliğine gönderme yapılarak türetilmiş bir terim olup, veri içindeki korelasyonları ve örüntüleri keşfetmek için istatistiksel algoritmalar kullanan veri araştırma becerisi olarak tanımlanmaktadır. Esas olarak, verinin anlaşılması ve kullanılmasında destek sağlayan bilgileri, yani veri külçelerini bulmaya ve ortaya çıkarma süreci olarak da açıklanabilir (Rygielski, Wang, ve Yen, 2002). Diğer bir ifadeyle veri madenciliği veri ambarında tutulan, net olarak anlaşılamayan veriler arasındaki gizli kalmış bilgileri keşfetmektir (Ünal, 2011).

Büyük veri yığınları çeşitli adımlardan geçirilip kullanılabilir bilgiye çevrilmesi aşamasının hepsi, 1989 yılında sunulan Knowledge Discovery in Databases çalışması ile ‘Veri tabanından bilgi keşfi’ olarak tanıtılmıştır (Fayyad vd., 1996).

Veri madenciliği; veri yığınlarında saklanmış, geçmiş zamanda tanınmayan, önceden kazanılmış deneyimlerle değerlendirilemeyecek, öğrenilmesi durumunda ise şirkete fayda getirecek bilginin ortaya çıkarılması işlemidir (Silahtaroglu, 2016). Veri madenciliği, çok fazla miktarda bilginin depolandığı veri tabanlarından, amaçlar doğrultusunda, gelecekle ilgili tahminler yapabilmeyi sağlayacak anlamlı verilere ulaşma ve bu veriyi kullanma işi olarak da açıklanabilir (Savaş, Topaloğlu, ve Yılmaz, 2012).

Veri madenciliğinin temel kavramları olan veri yığını, veri tabanı ve veri ambarı kavramları ise şöyle tanımlanmaktadır.

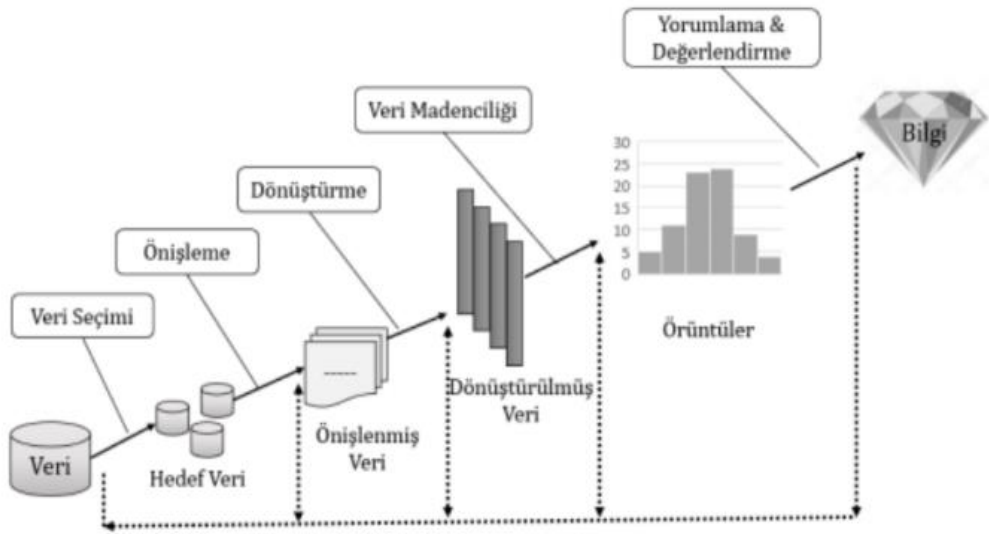
Veri yığını; Daha önce işlenmemiş, değiştirilebilen, denetlenebilen veri topluluğu olarak ifade edilmektedir (Ergün, 2008). Veri yığınları daha önce işlenmedikleri için gereksiz veri bulundurabilirler. Bundan dolayı veri yığınlarına doğrudan veri madenciliği yöntemleri kullanılamaz. Veri madenciliği yöntemi, uygulamaya ilk olarak veri yığınlarındaki gereksiz verileri eleyip kullanılabilir durumdaki verileri tespit ederek başlar. Elde edilen verilerde gereksiz bilgi miktarı çok fazla ise uygun veri madenciliği uygulansa bile arzu edilen sonuç alınamayabilir (Ergün,2008).

Veri tabanı; bağlantılı verilerin bulunduğu, kullanıcıya düzenli ulaşım olanağı sağlayan, taşınılabilen, değiştirilebilen, yönetilebilen, uygun formatta düzenlenmiş bilgi depolarıdır. Veri tabanı; birçok işletmenin uygulama ve programlarının sorunsuz bir şekilde kullanıldığı, birçok verinin tutulduğu yerdir (Döşlü, 2008).

Veri ambarı ise; veri madenciliği tekniklerinin uygulanabilmesi için verileri uygun formata getirip düzenleyen bir veri tabanıdır. Çeşitli kaynaklardan gelen bu veriler çoğunlukla farklı biçimdedirler. Farklı olan bu verilerin, aynı amaç doğrultusunda kullanılması ile veri ambarları oluşturulur. Veri ambarları sebebiyle çeşitli kaynaklardan gelen farklı biçimdeki veriler, bir başlık altında analiz edilebilir (Fayyad vd., 1996). Veri Ambarı, temelde karar vermede kullanılan, konu yönelimli, zaman

değişimli, birleştirilmiş, kararlı veri depolarıdır. Veri ambarının konu yönelimli olması; belirlenmiş konular için düzenlenmiş olması demektir, zaman değişkenli olması; ortalama 5-10 yıllık mevcut olan verileri saklayabilme özelliğidir, birleştirilmiş olması; çeşitli kaynaklardan gelen verilerin birbirleri arasında bağıntılı bir ilişki olmasıdır (Şentürk, 2006).

Verinin, bilgiye dönüştürülürken geçtiği aşamalar Şekil 2.2’de sunulmuştur.



Şekil 2.2: Veri Tabanı Bilgi Keşfi Sürecinin Aşamaları (Fayyad vd., 1996)

Veri madenciliğinin işletmeler açısından önemi büyüktür. Farklı alanlarda bulunan büyük ölçekli veri tabanları, değerli veriyi saklayan madenler gibi düşünülebilir. Büyük hacimli verilerin analiz edilmesi, analizler sonucunda anlamlı bilgilere ulaşılmasını ve elde edilen bu bilgilerin yorumlanarak işletme çalışanlarının ve ilişkisel veri tabanlarının yeteneklerinin arttırılmasına yol açmaktadır. Özellikle dijital veri miktarında yaşanan hızlı artışa karşılık, toplanan bu veri üzerinde araştırma yapabilecek yeterlilikte çalışan sayısının yetersiz kalması, işletmeleri veri madenciliği faaliyetlerine yöneltmiştir. Bu amaçla otomasyonel ve akıllı veri tabanı analizleri için yeni kuşak teknikler ortaya konmuş ve veri madenciliği giderek önemini artıran bir araştırma alanı haline gelmiştir (Savaş, Topaloğlu, ve Yılmaz, 2012).

İçerisinde çok büyük miktarda veri tutan veri tabanları, bir işletmeye kattığı değer açısından gizli kalmış bir maden gibi adlandırılabilir. Gelişen teknolojinin işletmelere



sunduđu en byk avantajlardan biri olan veri madenciliđi sayesinde kaybedilen mşterilerin geri kazanılması, mşterilerin elde tutulma oranı, mşterilerin deđerlerinin belirlenmesi ve gelecekteki hareketlerinin tahmin edilebilmesine ynelik alıřmalar yapılabilir. Tm bu adımlar sayesinde iřletmedeki krlılık oranı, rekabet gcnn arttırılması sađlanacaktır (Altan, 2019)

Mşteri iliřkileri ynetimi srecinin veri madenciliđi noktasında mşteri veri tabanlarını ve diđer bilgi depolama sistemlerini sađlayan bilgi teknolojileri ile harmanlanması; bařarılı, mşteriyi elde tutan ve yeni mşteriler kazandıran bir strateji ortaya konulmasında etkilidir (Mohammed ve Rashid, 2012).

## 2.3 Veri Madenciliđi Teknikleri

Veri madenciliđi teknikleri, dođası geređi mşteri iliřkileri ynetimi stratejilerine taban oluřturmaktadır. Veri tabanı oluřturulması, oluřturulan bu veri tabanının analizini gerekleřtirmek, bu dođrultuda hedef kitlenin tespit edilmesi, hedef kitleye sunulacak tekliflerin belirlenmesi ve mşteri iliřkileri ynetimi programının bařarısını lmek, efektif bir mşteri iliřkileri ynetimi srecidir. Veri madenciliđi teknikleri sayesinde iřletmeler mşterilerin anlamlı trendlerini, iřletme ile iliřkilerini anlayabilmekte; riskler minimize edilmekte, kayıplar azaltılmaya alıřılmaktadır (Cheng ve Chen, 2009). İřletmeler, mşterilerinin eđilimlerini tespit ettiklerinde; buna uygun yol izebilir, tatmini arttırarak yeniden satın alım oranlarını arttırabilirler (Wei, Lee, Chen ve Wu, 2013).

Temel veri madenciliđi teknikleri arasında ise sınıflandırma, kmeleme, iliřki kuralları, regresyon analizleri, ardıřıklık analizi gibi teknikler bulunmaktadır. Bunların yanı sıra kural temelli akıl yrtme yaklařımı, genetik algoritmalar, karar ađaları, bulanık mantık, endktif đrenim sistemleri, istatistiksel metodlar gibi teknikler de mevcuttur (Cheng ve Chen, 2009). Veri madenciliđi tekniklerinden řu konu bařlıklarında bilgi edinmek amacıyla yararlanılmaktadır (Olmeda ve Sheldon, 2001):

- Sınıflandırma: Veri iindeki bir nesne ya da mşterinin ait olduđu grubu tanımlayan rntleri ayırt eder.

- Kümeleme: Veri içindeki değişik gruplanmaları keşfeder ve aynı şekilde örüntüleri ayırt etmek üzerine çalışır. Kümeleme analizi ayrıca tahmin edici modelleme için bir temel sağlar.
- Sıralama ve İlişkiler: Sıralama teknikleri zamana bağlı olayları tanımlamada, ilişkiler ise tek bir olaya bağlı oluşumları tanımlamada kullanılır.
- Tahminleme: Gelecekle ilgili değerleri öngörme yöntemidir.

Hastie (2004), veri madenciliği tekniklerini 3 ana başlıkta toplamaktadır.

- Sınıflandırma ve Regresyon
- Kümeleme
- Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler

Yukarıda bahsi geçen modellerden; sınıflandırma, denetimli öğrenme grubunda yer alırken kümeleme, birliktelik kuralları ve ardışık zamanlı örüntüler denetimsiz öğrenme grubunda bulunmaktadır (Hastie vd., 2004).

Son yıllarda, müşteri ilişkileri yönetiminde, müşteri veri tabanlarını daha etkili bir şekilde segmente etmeyi mümkün kılan veri madenciliği tekniklerinin kullanılmasıyla daha verimli hale geldiği gözlemlenmiştir. Bu konuda mevcut literatürde çeşitli analitik teknikler ele alınmaktadır. Örneğin, RFM uzun yıllardır mevcut olan ve popülerliğini basitliğine borçlu olan bir tekniktir. Bunun yanında, son yıllarda daha sofistike yöntemlerin geliştirildiği görülmüştür. Ancak, RFM'nin kullanımı hala geniş çapta devam etmektedir (McCarty, J., ve Hastak, M. 2007)

### 2.3.1 Müşteri ilişkileri Yönetiminde Veri Madenciliği

Günümüzde işletmeler müşteri yaşam döngüsü analizleri aracılığıyla müşteri değerini artırmakla ilgilenmektedirler. Bilgi depolama, veri madenciliği ve diğer müşteri ilişkileri yönetimi teknikleri işletmelere müşteri ilişkileri yönetimine dair faaliyetleri uygulamaya koyma fırsatı sunmaktadır (Rygielski, Wang, ve Yen, 2002).

İşletmeler bütün müşterilerini tatmin edecek ürünler üzerine odaklanmaktan ziyade müşteri satın alma davranışlarını izleyerek, her bir pazar bölümü için uygun ürünler

sunmaya çalışabilir. Başka bir deyişle, müşteri segmentasyonunda satın alma davranışının baz alınması başarılı müşteri ilişkileri yönetimi açısından önem arz etmektedir (Qiasi, Baqeri-Dehnavi, Minaei-Bidgoli, ve Amooee, 2012).

Winer'a (2001) göre etkin bir M.İ.Y. için izlenmesi gereken adımlar şunlardır:

1. Müşteri aktivitelerini içeren bir veri tabanı oluşturmak
2. Veri tabanını analiz etmek
3. Analizlere bağlı olarak hedef müşterileri belirlemek
4. Bu müşterilere sağlanacak olan ayrıcalıkları tespit etmek
5. Müşteri İlişkileri Yönetimi programının başarısını ölçümlemek

Artan rekabetle başa çıkabilmek için işletmeler müşteri verisine oldukça önem vermeye başlamışlardır. Müşterilerin profil ve işlem bilgilerinin doğru bir şekilde toplanması, güvenli bir şekilde saklanması ve bu bilgilerin müşterinin bir sonraki adımını tahmin etme gibi veri temelli çalışmalar işletmeler açısından değer kazanmaktadır. Bilgi teknolojileri, müşteri ilişkileri yönetimi kapsamında işletmelere aşağıda sunulan konularda katkı sağlamaktadır (Berry, 1995).

1. Mevcut müşterilerin satın alma modellerinin takip edilmesi
2. Ürün, hizmet, teklif ve ücretleri her müşteriye, o müşterinin talebi doğrultusunda uyarlanarak sunulması
3. İletişim akışının çift yönlü olmasını sağlanması
4. Çapraz-yukarı satışa destek olunması

Başarılı bir müşteri ilişkileri yönetimi çözümü için ilk adım bir müşteri veri tabanı yaratılmasıdır. Yeni kurulan işletmeler için bu konu bir adım daha kolaydır fakat daha eski işletmelerde; müşterilere ait geçmiş veriyi bulmakta zorluklarla karşılaşılması muhtemeldir. Hizmet sektöründeki işletmelerse, işin doğası gereği müşterileri ile daha fazla etkileşim içinde olduklarından veri tabanı oluşturmaları perakende sektörüne göre daha kolaydır. Yine perakende sektöründe çoğunlukla araçların kullanılıyor olması veri tabanı oluşturulmasının önünde bir kısıt olarak yer almaktadır. Bu noktada önemli olan; müşteri ile temas noktalarının artırılması ve bu noktalarda bilgi toplanmasıdır (Altan, 2019).

Müşteri ilişkileri yönetimini pazarlama faaliyetlerinde kullanan işletmelerin; müşterilerin profil ve işlem bilgilerini tuttuğu büyük veri tabanları bulunmaktadır ve müşteri değerinin ölçülebilmesi için farklı müşteri ilişkileri yönetimi araçları mevcuttur. Veri madenciliği, müşterilerin davranışlarını, değerini anlamaya yardımcı olacak verinin bulunması süreci olarak açıklanabilir (Rygielski, Wang ve Yen, 2002). Ünal (2011) ise veri madenciliğini “veri ambarında tutulan, net anlaşılmayan veri arasında kalan bilgileri keşfetmek” olarak tanımlamaktadır. Veri tabanlarından geleceğe yönelik tahminler yapılmasını sağlayan anlamlı bilgilere ulaşma ve bu bilgileri kullanma işidir (Savaş, Toplaoğlu ve Yılmaz, 2012).

## 2.4 Müşteri Segmentasyonu

İşletmeler müşteri segmentasyonu yöntemi ile müşterilerini belirli gruplara ayırabilmektedir. Belirlenen her bir grubun karakteristik yapıları tespit edilerek bu gruplara özel ürün ve hizmet önerileri sunulabilmektedir. Bu sayede uzun vadede müşteri aidiyetinin sağlanması ve işletmenin karlılığının artması sağlanabilmektedir (Ergun, 2023).

Pazarlama çalışmaları, yoğun rekabet ortamında ayakta kalmayı hedefleyen işletmeler açısından müşterilerin önemine sıklıkla dikkat çekmiştir. Müşteri ilişkileri yönetimi, müşterilerle ilişkileri geliştirmeyi amaçlayan öne çıkan bir pazarlama yönetimi yaklaşımı olmuştur. Müşteri ilişkileri yönetimi 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 (Köse ve Arslan, 2020).

Müşterilere özel ürün ve hizmetler sunmanın birkaç farklı yöntemi vardır. Örneğin, müşterilerin geçmiş siparişlerine bakarak, bir kullanıcı profili çıkarılarak gelecekte alacağı ürünler tahmin edilmeye çalışılabilir. Yaygın kullanılan bir diğer yöntem ise, benzer alışveriş alışkanlıkları olan müşterileri kendi aralarında kümelere (müşteri segmentasyonu) ayırmaktır. Bu sayede, küme içerisinde bulunan kullanıcılar benzer alışveriş alışkanlıkları sergiledikleri için bu kullanıcıların satın aldığı ürünler küme içerisindeki diğer kullanıcılara önerilebilir. Ayrıca, kümelerin karakteristik özellikleri

çıkarılarak o kümelere yönelik pazarlama çalışmalarında bulunulabilir. Makine öğrenmesi yöntemleri sayesinde tavsiye sistemleri kurarak bu işlemler otomatik bir şekilde gerçekleştirilebilmektedir. Özetle, bir tavsiye sistemi müşterilerini öğrenir ve 2 müşteriye mevcut ürünler arasından en değerli bulacağı ürünleri önerir (Schafer vd., 1999).

## 2.5 Recency / Frequency / Monetary Kavramları

RFM 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).

İşletmeler, hangi müşterilerini hedefleyeceklerine, hangi müşteri bölümünün daha fazla getirisi olduğuna karar verebilmek için bazı ileri analiz yöntemleri uygularlar. RFM (Recency- Frequency- Monetary) analizi yaygın şekilde kullanılan bir müşteri değer analizi yöntemidir (Dursun, 2015).

RFM Analizi ilk kez 1994 yılında Arthur Hughes tarafından ortaya konmuş ve günümüze kadar istatistiksel analizlerde etkinliğini sürdürmüştür. Geçmişe yönelik müşteri verilerinin 3 temel başlıkta modellenerek sonrasında müşterilerin segmentlere ayrılması işlemlerini barındırır. Makine öğrenmesi algoritmalarının ve yeni nesil yapay zekâ çözümlerinin de veri analizine dahil olması ile RFM (Recency- Frequency- Monetary) analizi artık tek başına yeterli görülmemektedir ancak müşteri kullanıcı deneyimi alanında ve pazarlama stratejilerinin belirlenmesinde yine etkilidir ve müşterilerin segmente edilmesi aşamasında bu analiz yöntemi tercih edilir.

RFM analizi pazarlama sektöründe müşteri kümelerinin belirlenmesinde sıkça kullanılan bir analizdir. RFM analizinin amacı, şirketlerin yapmış oldukları pazarlama

kampanyasına cevap verme olasılığı en yüksek olan müşteri kümelerini belirlemektir. Şirketler, belirlenen müşterileri hedef alarak ve kampanyalarına geri dönüş yapmayacak müşterilerden kaçınarak, daha hedefli pazarlama stratejileri geliştirmektedir. RFM analizi, anlaşılabilir ve hızlı uygulanıyor olması, şirketler tarafından anlaşılması ve yorumlanmasının daha kolay olması sebebiyle daha fazla tercih edilmektedir (McCarty ve Hastak, 2007).

RFM Analizinin temelini 3 ana metrik oluşturur. Bunlar aşağıdaki gibidir (Erarslan, 2023).

- Recency – Güncellik
- Frequency - Sıklık
- Monetary – Tutar

Recency; RFM Analizi'nde "Recency" (R), bir müşterinin en son satın alma işleminin ne kadar "yakın" olduğunu ifade eder. Bu, genellikle bir müşterinin en son satın alma işleminden itibaren geçen zaman aralığı olarak ölçülür. Recency, müşterinin işletmeyle olan etkileşiminin tazeliğini gösterir ve genellikle müşterinin gelecekteki satın alma olasılığını tahmin etmek için kullanılır (Kanca, 2023).

Frequency; RFM Analizi'ndeki "F" yani "Frequency" (Sıklık), bir müşterinin belirli bir zaman diliminde ne kadar sık alışveriş yaptığını ifade eder. Bu, genellikle müşterinin bir işletmeyle olan etkileşim sıklığını gösterir ve müşterinin işletmeye olan bağlılığının bir göstergesi olarak kabul edilir. Daha sık alışveriş yapan müşteriler genellikle daha değerli kabul edilir çünkü bu, onların işletmeyle daha güçlü bir ilişkisi olduğunu ve gelecekte de alışveriş yapma olasılıklarının daha yüksek olduğunu gösterir (Kanca, 2023).

Monetary; RFM analizindeki "M", "Monetary" kelimesinin baş harfidir ve "Parasal" anlamına gelir. Bu, bir müşterinin belirli bir zaman diliminde bir şirkete ne kadar para harcadığını ifade eder. Yani, bir müşterinin toplam harcamalarını temsil eder. Bu ölçüm, müşterinin şirket için ne kadar değerli olduğunu belirlemeye yardımcı olur. Genellikle, daha yüksek parasal değere sahip müşteriler daha değerli kabul edilir çünkü daha fazla gelir sağlarlar (Kanca, 2023).

## 2.6 RFM Analizi Nasıl Çalışır

Bir veri madenciliği yöntemi olan RFM analizinde veri tabanında bulunan bazı müşteri işlemlerinin analiz edilerek müşterilerin gelecekteki davranışları tahmin edilmeye çalışılmaktadır. (Wei, Lee, Chen, ve Wu, 2013). Bu sayede işletmeler daha çok kâr getiren müşterilerini belirleyebilmekte, bu müşterilere ve kaybedilen müşterilere yönelik pazarlama stratejilerini belirleyebilmektedirler (Wei, Lee, Chen ve Wu, 2013).

Analizde her müşteri geçmiş davranışlarına göre puanlanmaktadır. Ardından müşteriler, RFM puanlarında göre benzer gruplara ayrılmaktadır. RFM analizi aynı zamanda müşteri sadakatini ölçen bir yöntem olarak da değerlendirilmektedir (Birant, 2011). RFM analizinde bir işletmenin müşteriye ait verisi; güncellik (Recency), sıklık (Frequency) ve parasallık (Monetary) değişkenleri temel alınarak analiz edilmektedir (McCarty ve Hastak, 2007). Güncellik, son satın alımdan sonraki zamanı ifade etmektedir (gün, ay vb.). Diğer bir ifadeyle, Güncelliği ifade eden “R” en son tüketim davranışının olduğu zamanla şimdiki zaman arasındaki süredir. Süre ne kadar kısaysa “R” o kadar büyük değer alır. Sıklık (F), belirli bir zaman aralığında gerçekleştirilen satın alımların sayısıdır. Örneğin “bir yılda iki defa”, “ayda iki defa” gibi. Sıklık ne kadar fazlaysa, “F” o kadar büyük değer alır. Parasallık ise belirli bir zaman aralığında harcanan toplam para miktarını ifade etmektedir. Tutar ne kadar yüksekse “M” o kadar yüksek değer alır. (Wei vd., 2013; Hosseini, Maleki ve Gholamian, 2010).

Günümüzde müşteri hareketlerinin numaralara dökülmesi; büyük işlemsel veri tabanlarının oluşturulması ve kullanılması; teknolojinin de desteği ile gelişme ve başarı göstermeye başlamaktadır. Veri tabanı üzerinden gerçekleştirilen pazarlama araştırmaları ve faaliyetleri de buna dayanarak kendisine daha fazla uygulama alanı bulmuştur (Coussement vd.2014).

### 2.6.1 RFM Analizi Kullanım Amaçları

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. RFM skorlamasının amacı, gelecekteki davranışı öngörmektir. Müşteri verilerini istatistiksel dile getirmek çok önemlidir. Segmentasyona başlamak için öncelikle yakınlık, sıklık ve parasal değerlerin alınması gerekmektedir. Bundan sonra müşteriler azalan sırayla sıralanır ve bu tür bir sıralama, değerlerin ölçümünü sağlar. 1'den 4'e kadar her değer dağılımı puanlanır ve bu, istatistiklere dayanır (Miglautsch 2000).

Yang' a 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 (Value - Değer) geçmiş değeri temsil eder ve gelecekteki yanıt olasılığını çıkarır. Bu, RFM'nin müşteri davranışlarını ve gelecekteki eylemleri tahmin etme yeteneğini güçlendirir (Yang, 2004).



## 2.6.2 RFM Analizi Metodu

Müşteriler, yaptıkları alışveriş sayısı, alışveriş tutarı ve son alışverişlerinden baz alınan tariha kadar geçen süre metriklerine göre segmente edilir. Recency, Frequency ve Monetary kümeleri özelinde müşteriler hem skorlanabilir hem de buldukları kümelere göre kitlelere ayrılabilir (Oğur, 2023).

Müşteri davranışını analiz etmede RFM modelini uygulamanın yaygın bir yolu, RFM'nin her bir boyutuyla müşteri verisini analiz etmek ve bu verileri 5 eşit bölüme ayırarak kodlamaktır. Kodlama yönteminde ilgili parametreler bazında tüm müşterilerle her bir müşteri kıyaslanır ve her müşteri için RFM skoru oluşturulur. Güncellik için müşteri veri tabanı satın alma tarihleri azalan şekilde sıralanır. Müşterilerin firmadan en güncel olarak alışveriş yapan %20'si 5, diğer %20'si 4 olacak şekilde numaralandırılarak devam edilir. Sıklık ve tutar için müşterinin ziyaret sıklığı verisi ve satın alma verisi de yine azalan şekilde sıralanır. Tüm müşteriler 555,554,553,...111 şeklinde 125 mümkün şekilde kodlanırlar (5 x 5 x 5). Böylece veri tabanı 125 eşit kümeye bölünür. Müşteriler içerisinde en yüksek RFM skoruna sahip olanlar genellikle en kârlı olanlardır (Hosseini, Maleki, ve Gholamian, 2010; Wei, Lee, Chen, ve Wu, 2013).

Hughes (1994) tarafından önerilen RFM modelinde,

- Güncellik son satın alımdır. Güncelliği ifade eden “R” en son tüketim davranışının olduğu zamanla şimdiki zaman arasındaki süredir. Süre ne kadar kısaysa “R” o kadar büyük değer alır.
- “F” sıklığı ifade eder ve belirli bir periyoddaki işlemlerin sayısını belirtir. Örneğin “bir yılda iki defa”, “ayda iki defa” gibi. Sıklık ne kadar fazlaysa, “F” o kadar büyük değer alır.
- “M” tutarı ifade eder ve belirli bir periyotta harcanan para tutarını belirtir. Tutar ne kadar yüksekse “M” o kadar yüksek değer alır (Cheng ve Chen, 2009; Hughes, 1994).

Bir diğer yöntem ise kodlanmış numaralar yerine orijinal verinin kullanılmasıdır. Orijinal verinin kullanıldığı yöntemde güncellik son satın alımın yapıldığı zamandan

bu yana geçen süreyi; sıklık bu periyotta yapılan satın alımların sayısını; tutar ise tüm satın alımlarda harcanan toplam parayı ifade etmektedir. “R” hesaplanırken seçilen tarih periyodu en uzak tarih 1’den başlayacak şekilde en yakın tarihe kadar bir sayı arttırılarak numaralandırılır. Örneğin 1 Ocak 2008 ve 31 Mart 2008 tarihlerinin seçilen periyod olduğunu varsayarsak; 1 Ocak 2008 bir, 2 Ocak 2008 iki ve 31 Mart 2008 doksanbir değerini alacak şekilde numaralandırılır. “F” değeri müşterinin bu tarihler arasındaki satın alım sayısıdır. “M” değeri ise müşterinin bu tarihler arasında toplam harcadığı tutardır. Bu değerlerin ortalamaları alınarak, kümelerin ortalamasının üstünde kalan RFM değerleri “↑” işareti ile kümelerin ortalamasının altında kalan RFM değerleri ise “↓” işaretiyle belirtilir. Böylece R↑F↑M↑ müşteriler “sadık müşteriler” olarak tanımlanır. R↓F↓M↓ müşteriler, “kayıp müşteriler” dir. R↑F↓M↓ müşteriler “yeni müşteriler” olarak tanımlanırken R↑F↑M↓ müşteriler ise “potansiyel müşteriler” olarak tanımlanırlar (Wei, Lee, Chen, ve Wu, 2013).

RFM analizi uygulayıcılar tarafından farklı amaçlarla kullanılmakta olduğundan farklı yöneticiler için farklı yönetsel faydalar elde edilebilir (Hosseini, Maleki, ve Gholamian, 2010). Bir diğer teknikte, yine müşteri listesi satın alım tarihlerine göre ters kronolojik şekilde sıralanır ve beş eşit parçaya bölünür. En yakın tarihte alışveriş yapan müşteriler “1” rakamıyla, en uzak tarihtekiler ise “5” rakamıyla kodlanır. Bu şekilde 5 eşit bölüm oluşturmak “beştebirliklemek” olarak adlandırılır. Sonrasında müşteriler frekans sayısı düzenlerine göre sıralanarak aynı metod ve etiketleme işlemi uygulanır. Son olarak aynı metod “ortalama harcama” tutarlarına göre tekrarlanır (Kahan, 1998).

Yukarıda ifade edildiği gibi müşterilerin satın alma davranışlarını kullanarak benzer davranış gösteren müşterileri aynı kümeye toplayan RFM analizi, müşterileri tanınmasını sağlarken, doğru stratejiler geliştirip gerekli aksiyonların alınmasına olanak sağlar (Oğur, 2023).

## 2.7 Kümeleme Analizi

RFM analizinin çıktılarının değerlendirilebilmesi amacıyla, benzer skoru alan kişilerin, benzer gruplarda değerlendirilmesi, bölümlendirmenin bu doğrultuda gerçekleştirilebilmesi amacıyla kümeleme analizi kullanılmaktadır (Altan, 2019).

Kümeleme, sınıflandırma yönteminde olduğu gibi veri tabanındaki birbirine benzeyen verileri gruplara ayırma işlemidir. Kümeleme yönteminin sınıflandırma yönteminden farkı ise; kümeleme yönteminde sınıfların önceden bilinmiyor olmasıdır. Verilerin birbirine olan benzerliklerine göre ayrılacakları sınıflar belirlenir (Dumhan, 2003). Kümeleme yönteminde amaç, belirlenen verileri kullanarak benzerlik ve farklılıklarına göre sınıflara ayrılmasıdır. Belirlenen sınıfların her biri küme olarak ifade edilir. Kullanım amacına göre; model kurma, tahmin etme, veri inceleme, hipotez testi, hipotez oluşturma gibi çeşitli amaçlar için tercih edilir (Çakmak vd., 2005)

Kümeleme analizinde amaç, bir veri grubunun farklı bölümler içerip içermediğinin, içeriyorsa bu bölümlerin belirlenmesidir (Tekin, 2018). Diğer tüm analizlerde küme sınıflarının özellikleri önceden tanımlanırken, kümeleme analizinde sınıfların tanımları aslında birer çıktıdır (Hajizadeh ve Shahrabi, 2010).

Kümeleme yöntemleri iki gruba ayrılır. Bunlar aşamalı kümeleme yöntemleri ve aşamalı olmayan kümeleme yöntemleridir. Aşamalı kümeleme yöntemleri, değişkenlerin birbirlerine olan benzerliklerini dikkate alarak küme uzaklıklarının birbiri ile birleştirmeyi amaçlayan tekniktir. Aşamalı kümeleme yöntemi ile en sık kullanılan yöntemler, en yakın komşu ve en uzak komşu yöntemleridir. Aşamalı olmayan kümeleme yönteminde  $n$  tane düğüm, önceden bilinen  $k$  küme sayısına göre kümeleri belirlenir. Aşamalı olmayan kümeleme yönteminde en sık kullanılan algoritma ise  $k$ -ortalama algoritmasıdır (Oğur, 2023).

### 2.7.1 K – Ortalama Kümeleme Analizi

$K$ -ortalama, bir veri kümesini otomatik olarak  $k$  gruplarına bölmek için yaygın olarak kullanılan bir yöntemdir.  $K$  ilk küme merkezini seçerek ve ardından bunları yineleyerek ilerlemektedir. Her örnek en yakın küme merkezine atanmaktadır. Her bir küme merkezi, kurucu örneklerin ortalaması alınarak güncellenmektedir (Wagstaff vd., 2001).

Analiz, hiçbir verinin küme merkezini oynatmadığı noktaya kadar devam eder. Küme benzerliği, değerlerin ortalamaya uzaklığına göre belirlenir (Altan, 2019).

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, G. vd., 2013).

K-ortalama, en basit kümeleme algoritmasıdır. Bu algoritma, girdi olarak tanımlanmış sayıda küme kullanır. Ortalama, belirli bir kümenin tüm üyelerinin ortalama bir konumu anlamına gelir. Bir örneğin her bir niteliğinin değeri, nitelik eksenleri boyunca orijinden uzak mesafeyi temsil eder. Tabii ki, bu geometriyi verimli bir şekilde kullanmak için, veri kümesindeki değerlerin hepsinin sayısal bir değer taşıması gerekmektedir (Altan, 2019). Bu yüzden bu çalışmada kullanılacak olan değerler numerik hale dönüştürülmektedir.

K-ortalama algoritması genellikle hızlı ve etkili bir kümeleme yöntemi olmasına rağmen, başlangıç merkezlerinin rastgele seçilmesi sonuçları etkileyebilir. Bu nedenle, farklı başlangıç noktalarıyla birden fazla çalışma yapılması ve en iyi sonucun seçilmesi önerilir. Ayrıca, küme sayısının doğru bir şekilde seçilmesi ve veri setine uygun bir değer belirlenmesi de önemlidir. Bu faktörlerin dikkate alınması, K-ortalama algoritmasının daha etkili ve doğru sonuçlar vermesini sağlar (Kanca, 2023).

# Bölüm 3

## RFM Analizi İle Müşteri

### Segmentasyonu Uygulaması

Bu çalışma, Birleşik Krallık merkezli bir çevrimiçi perakende hediyelik eşya mağazasının verilerini kullanarak, şirketin müşteri satın alma davranışlarını derinlemesine ve doğru bir şekilde incelemesini amaçlamaktadır. Bu analizler, şirketin pazarlama stratejilerini ve operasyonlarını daha bilinçli ve etkili bir şekilde yönlendirebilmesi için son derece önemlidir.

Bu bölüm, kullanılan veri setinin detaylı bir tanıtımını, veri seti üzerinde gerçekleştirilen ön işleme süreçlerini, kümeleme algoritmalarının nasıl uygulandığını, bu algoritmaların sonuçlarının nasıl değerlendirildiğini ve elde edilen müşteri segmentlerine dair yapılan analizleri ve yorumları içermektedir.

Çalışmada; RFM analizi ve K-ortalama Python programlama dili ile geliştirilmiştir. Bu süreçte kullanılan Python kütüphaneleri aşağıda verilmiştir:

- numpy: Veri üzerinde matematiksel işlemler gerçekleştirmek için kullanılacaktır.
- pandas: Veri setlerini işlemek, temizlemek ve analiz etmek için kullanılacaktır. Tablo formatında veri manipülasyonu sağlar.
- sklearn: Kümeleme algoritmalarının uygulanması ve model değerlendirme metriklerinin elde edilmesi için kullanılacaktır.
- seaborn ve matplotlib: Veri görselleştirme işlemleri için kullanılacak ve elde edilen sonuçların grafiksel olarak sunulmasını sağlayacaktır.
- skfuzzy: Bulanık mantıkla çalışmak ve bulanık C-ortalama algoritmasını uygulamak için kullanılacaktır.
- scipy: İleri düzey matematiksel işlemler ve istatistiksel analizler için kullanılacaktır.

- plotly: Dinamik ve etkileşimli grafikler oluşturmak için kullanılacaktır.

### 3.1 Veri Seti ve Veri Ön İşleme

Bu çalışma, Birleşik Krallık merkezli bir çevrimiçi perakende hediyelik eşya mağazasının verilerini kullanarak, şirketin müşteri satın alma davranışlarını derinlemesine ve doğru bir şekilde incelemesini amaçlamaktadır.

Şirketten alınan veriler perakende satışlar için gerçekleşen tüm işlemleri içerir. Bu veriler, hangi müşterinin hangi ürünü, kaç adet ve hangi fiyatla satın aldığı gibi temel satış bilgilerini içermektedir. Bu bilgiler, müşteri davranışlarını anlamak ve segmentasyon yapmak için oldukça değerlidir. Verilerin gizliliği ve müşteri bilgilerinin korunması açısından, tüm veriler maskelenmiştir ve sadece şirket tarafından bilinen benzersiz kodlar ile paylaşılmıştır. Bu hem müşteri gizliliğini korurken hem de verilerin analiz edilmesi için gerekli olan bilgileri sağlar.

Şirket tarafından sağlanan ham verilere ait bir örnek tablo aşağıda sunulmuştur. Bu tablo, analiz sürecinde kullanılan verilerin bir örneğini göstermekte olup, bu verilerin nasıl kullanıldığını ve analiz edildiğini daha iyi anlamamıza yardımcı olacaktır.

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	12/1/2009 7:45	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	12/1/2009 7:45	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	12/1/2009 7:45	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	12/1/2009 7:45	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	12/1/2009 7:45	1.25	13085.0	United Kingdom

Tablo 3.1: Ham Müşteri Satış Verileri

Kolon Adı	Kolon Açıklama
Invoice	Fatura numarası
StockCode	Her bir ürün özelinde verilen tekil kod
Description	Ürün Açıklaması
Quantity	İşlem başına düşen her bir ürünün miktarı
InvoiceDate	Fatura tarihi ve saati
Price	Birim başına düşen ürün fiyatı
Customer Id	Her müşteri özelinde verilen tekil kod
Country	Müşterinin satın alma yaptığı ülkenin adı

Tablo 3.2: Ham Müşteri Verisine Ait Kolon Adı Ve Açıklamaları

Müşteri segmentasyonu süreci, verinin bir dizi ön işleme adımından geçirilmesiyle başlayacaktır. Bu aşama, verinin temizliğini ve kalitesini sağlamak için kritik öneme sahiptir, çünkü bu faktörler sonuçların doğruluğunu ve güvenilirliğini doğrudan etkileyecektir.

Ön işleme süreci kapsamında, müşteri analizi için oldukça önemli olan Recency, Frequency ve Monetary değerleri hesaplanmıştır. Bu değerler, bir müşterinin bir işletmeyle olan etkileşimini ve değerini ölçmek için kullanılır.

Ön işlemeden geçirildikten sonra verinin deseni aşağıdaki gibi olmuştur.

Fatura Tarihi	Fatura No	Ülke	Müşteri ID	Ürün Kodu	Ürün Açıklaması	Adet	Birim Fiyat	Tutar
2009-12-01 07:45:00	489434	United Kingdom	13085.0	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	6.95	83.4
2009-12-01 07:45:00	489434	United Kingdom	13085.0	79323P	PINK CHERRY LIGHTS	12	6.75	81.0
2009-12-01 07:45:00	489434	United Kingdom	13085.0	79323W	WHITE CHERRY LIGHTS	12	6.75	81.0
2009-12-01 07:45:00	489434	United Kingdom	13085.0	22041	RECORD FRAME 7" SINGLE SIZE	48	2.10	100.8
2009-12-01 07:45:00	489434	United Kingdom	13085.0	21232	STRAWBERRY CERAMIC TRINKET BOX	24	1.25	30.0

Tablo 3.3: Ön İşlemeden Geçirildikten Sonra Müşteri Veri Deseni

Ön işleme süreci sonrasında yukarıda deseni verilen veri için K-ortalama kümeleme algoritması kullanılarak müşteriler farklı kümelere ayrılacaktır. Kümelere ayırma işlemi sonrasında ise kümelerin başarımları değerlendirilecektir. Yapılacak değerlendirme de Silhouette indeks kullanılacaktır.

## 3.2 RFM Skoru Hesaplama ve Müşteri Segmentasyonu

Recency, Frequency ve Monetary değerlerinin belirlenmesinde 5'li ölçüm kullanılmıştır. Müşteriler R, F, M değerlerine göre sıralanmış ve %20'lik kategorilere ayrılarak; hangi müşterinin hangi skoru alacağı belirlenmiştir.

Her bir müşteri için Recency, Frequency ve Monetary değerleri hesaplanır. Bu değerler, belirli yüzdelik dilimlere göre (quantiles) segmentlere ayrılır. Her bir müşteri için üç ayrı skor belirlenir (1'den 5'e kadar) ve bu skorlar birleştirilir.

RFM analizine gidilmiş ve sırayla Recency – Frequency- Monetary için kolonlar oluşturulmuş daha sonra rfm\_table isimli dataframede biraraya getirilmiştir.

data\_x : Toplam harcama tutarı verisi müşteri bazında groupby ile aggregate edilerek kişi başına düşen tutar üzerinden alışverişin recency değeri belirlenmiştir.

data\_y: Harcama sıklığını ifade eder. Bu değer doğrudan belirlenmesi için herhangi bir veri yok bu nedenle önce invoice verisi müşteri ve invoice no bazında groupby ile aggregate edilerek veri setinde ilgili müşterinin o işleme ait kaç ürün aldığı bilgisi tutulmuştur.

Daha sonra ilgili ürün adedi bilgisi unique değer olarak dikkate alınmış ve buradan lambda x: len(x) methodu ile frekans değeri çıkarılmıştır.

data\_z: Toplam harcama tutarını (TotalPrice) ifade eder ve RFM matrisinin son parametresidir. Toplam harcama tutarı verisi müşteri no bazında groupby ile aggregate edilerek veri setindeki ilgili müşterinin toplam harcama tutarı hesaplanmıştır.

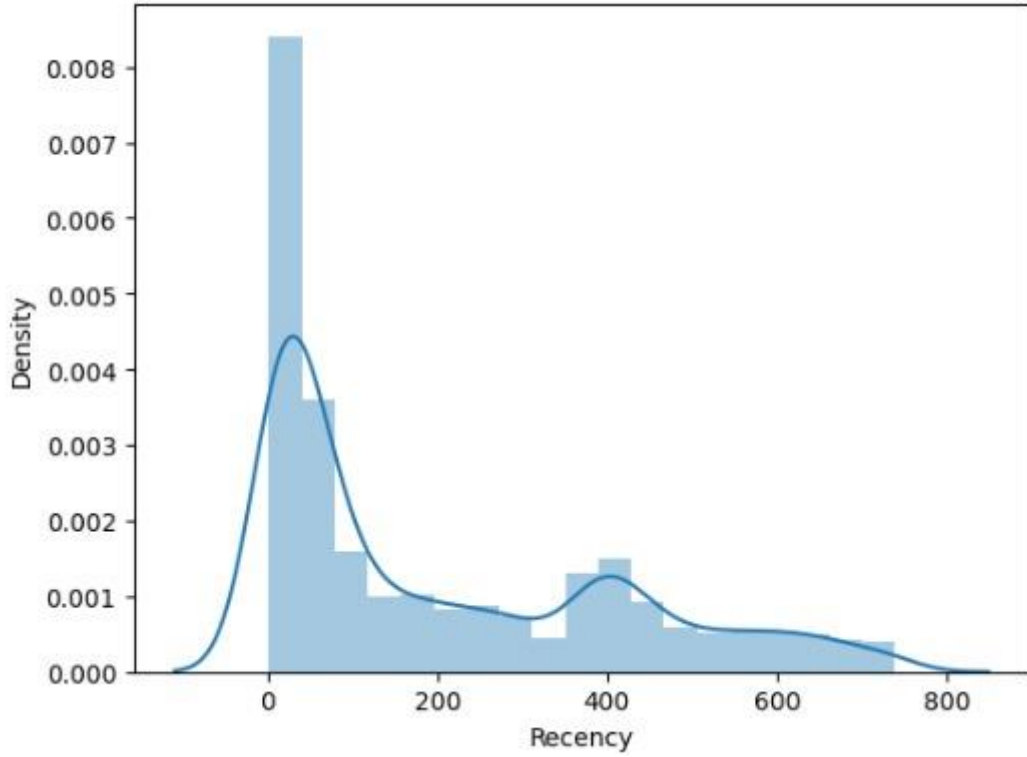
RFM analizi için oluşturulan 3 ayrı parametrenin dataframei mergelenerek tabloya son hali verilmiştir. Böylece rfm\_table aşağıdaki gibidir. Tablodan çıkarım yapılacak olursa 12346.0 id li müşteri 1 yıla yakın süredir alışveriş yapmamış, 12 kez alışverişte bulunmuş ve totalde 77556,46 tutarında harcaması olmuştur.



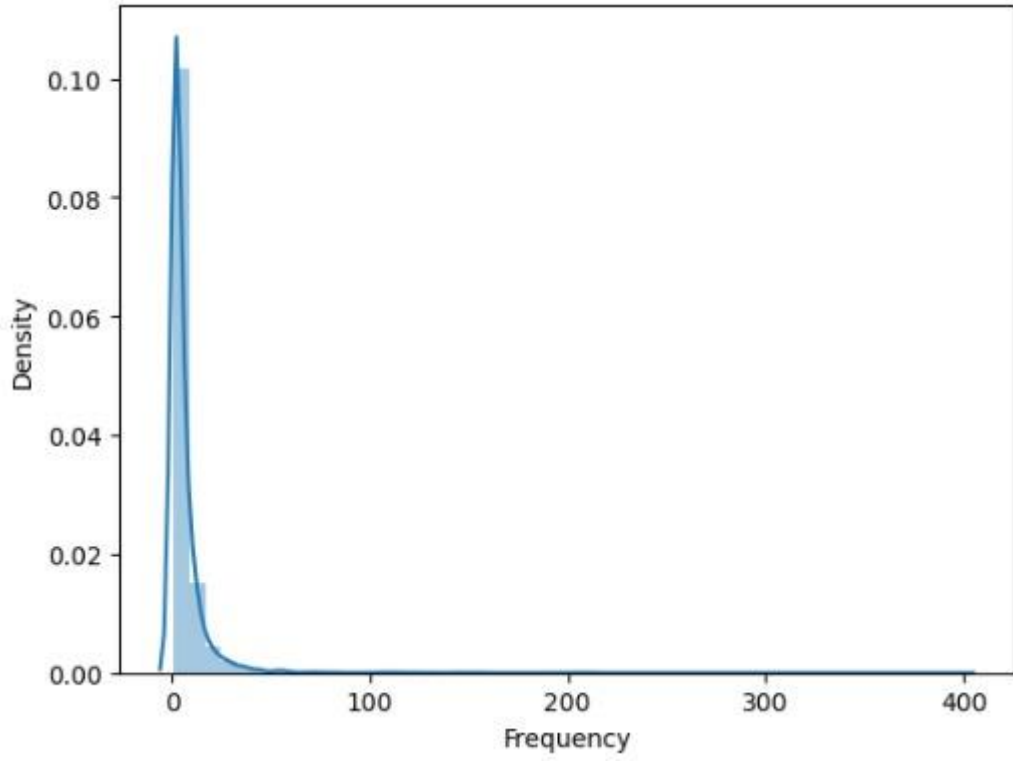
Customer ID	Monetary	Recency	Frequency
12346.0	77556.46	326	12
12347.0	5633.32	2	8
12348.0	2019.40	75	5
12349.0	4428.69	19	4
12350.0	334.40	310	1

Tablo 3.4: RFM Değer Tablosu

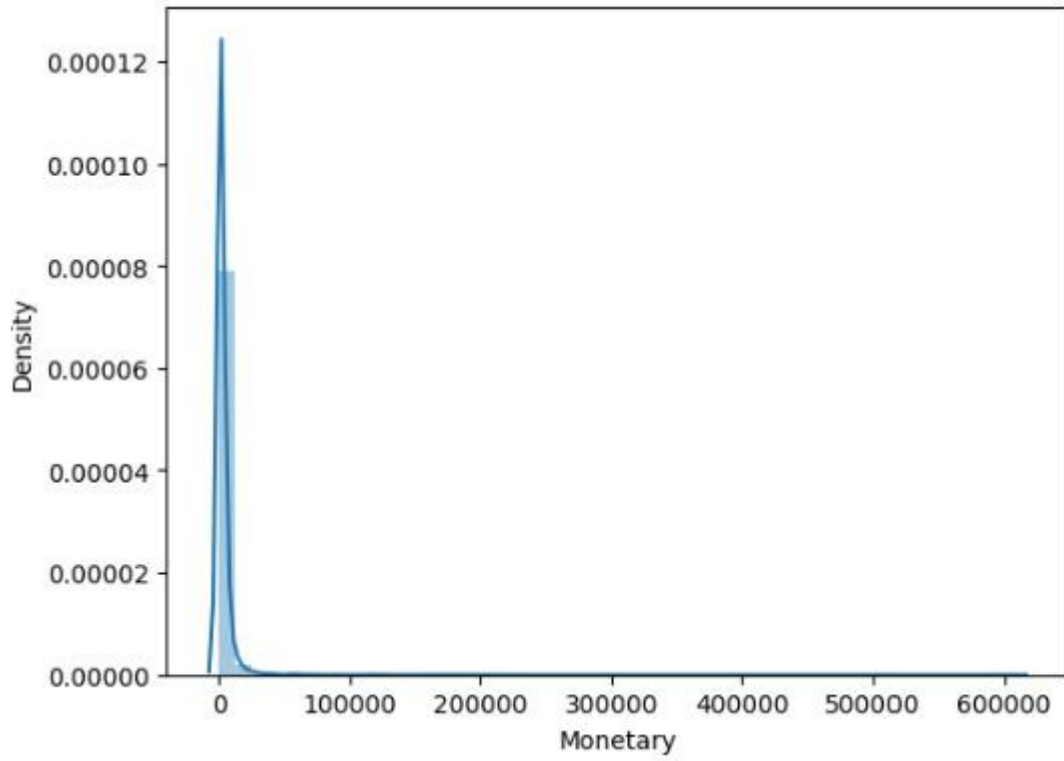
RFM metriklerine ait veri kolonlarının Seaborn veri görselleştirme Python kütüphanesi kullanılarak aşağıdaki grafikler elde edilmiştir.



Şekil 3.1: Recency Veri Grafiği



Şekil 3.2: Frequency Veri Grafiği



Şekil 3.3: Monetary Veri Grafiği

RFM analizi skoru bazında müşterileri 5 eşit parçaya ayırarak çeyreklere bölünmüştür. Bunun için 2 ayrı yöntem kullanılabilir. İlki fonksiyon içerisinde bölümlenmek ikincisi ise qcut() methodudur. Pandas kütüphanesi içerisinde yer alan qcut fonksiyonu ilgili veriyi eşit parçacıklara ayırır. Bu projede çeyrekler %20 lik dilimler olarak ayrılmıştır. 1 den 5 e kadar segmentlere ayırırken pd.qcut(rfm\_table['Recency']) kullanılmıştır. Tabloda sıfır değeri yer almasın istenirse, bulunan değerler 1 artırılır. Bunu bir iş birimi kararı olarak düşünebiliriz tamamen tercih sebebidir. Bulunan değerler birleştirilerek tek bir skor elde edilir. RFM skoru string tutulur ve buna göre müşteriler segmentlere ayrılır. Bu işlem sonucu RFM skoru aşağıdaki gibi listelenmiştir.

Customer ID	Monetary	Recency	Frequency	Freq_Tile	Rec_Tile	Mone_Tile	RFM Score
17001.0	6640.80	1	19	5	1	5	155
15156.0	961.49	1	3	3	1	3	133
12433.0	20581.26	1	10	5	1	5	155
13890.0	2927.38	1	14	5	1	4	154
13298.0	360.00	1	1	1	1	2	112

Tablo 3.5: RFM Skor Tablosu

17001.0 id li müşteriye tabloda bakılacak olursa RFM skoru 155 dir. 1 gündür herhangi bir işlem yapmamış toplamda 19 kez alışveriş yapmış ve toplam 6640,80 tutarında ödeme yapmıştır. Frequency veri setinde alışveriş sıklığı 10 üzeri olduğunda Frequency puanının hemen 5 e yükseldiği görülür.

Bu şekilde, segmentasyon işlemi tamamlanmıştır. Örneğin 111 segmentini yorumlayacak olursak, bu segmentte 40 kişi bulunmakta, ortalama olarak en son alışverişleri 11 gün önce gerçekleşmiş, alışveriş sıklıkları 1, yani 1 alışverişleri var, 170 TL harcamaları olmuştur.

RFM Score	Recency				Frequency				Monetary			
	mean	min	max	count	mean	min	max	count	mean	min	max	count
111	11.6	2	19	40	1.0	1	1	40	170.8	43.2	285.6	40
112	10.9	1	19	29	1.0	1	1	29	380.4	301.3	546.0	29
113	10.0	1	19	5	1.0	1	1	5	895.5	681.0	1201.5	5
115	3.0	3	3	1	1.0	1	1	1	3861.0	3861.0	3861.0	1
121	8.8	1	19	21	2.0	2	2	21	215.4	77.4	284.5	21

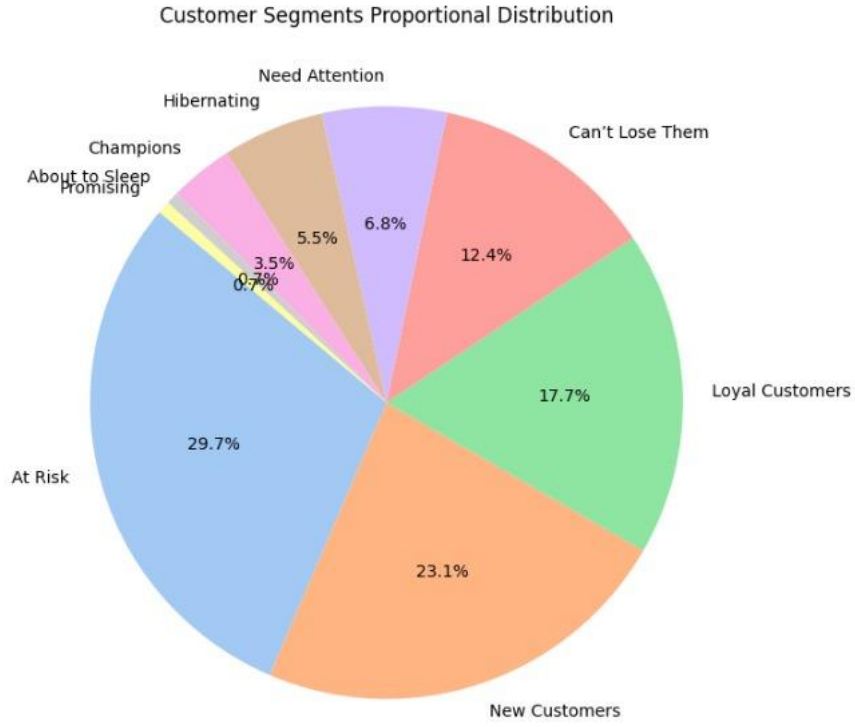
Tablo 3.6: RFM Skoruna Göre Müşteri Segmentlerinin Analitik Özeti

Kimi zaman bu metrikler arasında ağırlıklandırma yapıldığı görülmüştür. Bir şirket için parasal değer daha önemli olabilir, bu durumda %60 parasal değer, %20 recency değeri, %20 frequency değeri ile çarpıp, ortak bir skor elde edilebilir. Ya da, bu skorlar ile kmeans yapıp, daha tutarlı segmentler elde edilebilir. Bunun gibi kararlar şirketteki ilgili kişiler ile tartışılıp ortak bir karara varılması gereken durumlardır. Bazı kabuller, kurallar belirlenerek, segmentasyon kuralları dönüştürülebilir (Erarslan, 2023). RFM skorları belirlenmesi ardından müşteriler Champions, Loyal Customers, Potential Loyalists, Need Attention, At Risk, Hibernating, About to Sleep, Promising, New Customers, Can't Lose Them olmak üzere 10 kategoride etiketlenmiştir.

	Segment	Recency	Frequency	Monetary	Customer Count
0	About to Sleep	11.650000	1.000000	170.750250	40
1	At Risk	21.063647	13.786124	7281.816483	1744
2	Can't Lose Them	136.371901	2.196970	323.048196	726
3	Champions	338.768473	9.512315	4778.991882	203
4	Hibernating	26.825153	1.687117	1210.080920	326

Tablo 3.7: RFM Analizi Sonucunda Belirlenen Müşteri Segmentleri

RFM skoru analizi sonucunda belirlenen müşteri segmentleri, harcama alışkanlıklarına göre gösterilen pasta grafiği şu şekildedir. Bu görselde, kaybedilme potansiyeli yüksek olan müşteri gruplarının sayısının fazla olduğunu gözlemliyoruz. Dolayısıyla, bu müşteri segmentine özel pazarlama kampanyaları düzenlenerek, satış ve pazarlama stratejileri ile geri kazanım sağlanabilir.



Şekil 3.4: RFM Analizi ile Müşteri Segmentasyonu Pasta Grafiği

### 3.3 K-Ortalama Analizi Uygulanması

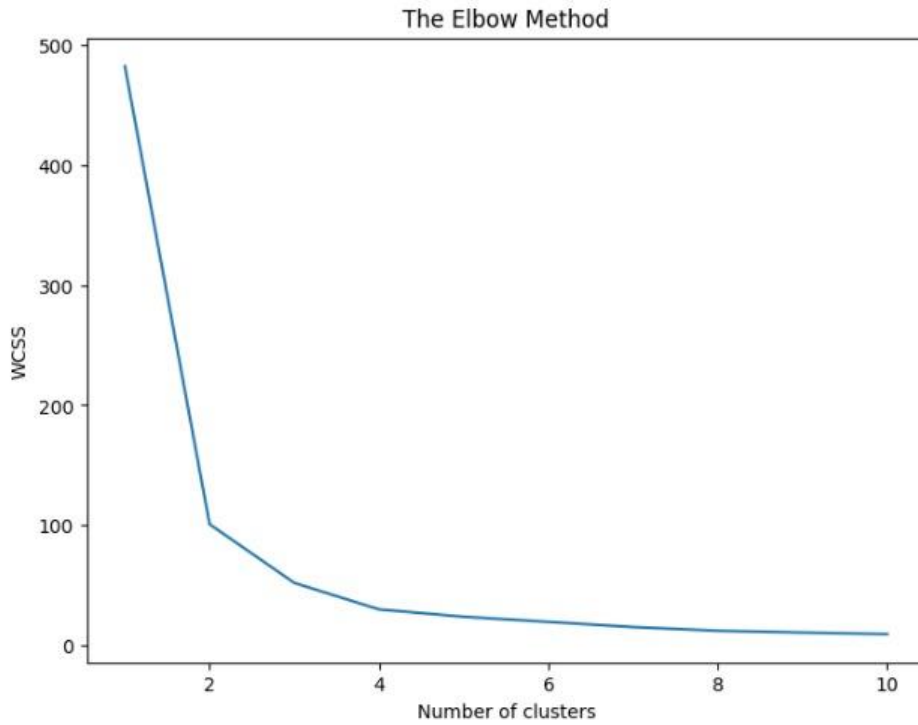
RFM analizi ile segmente edilen müşteri popülasyonuna ek olarak, gözetimli makine öğrenmesi algoritmalarından K-means kümeleme algoritması da uygulanmış ve elde edilen çıktılar karşılaştırılmıştır. Kümeleme algoritmaları genellikle uzaklık temelli çalıştıkları için, verinin tutarlı olması için normalizasyon işlemi yapılmıştır.

Normalleştirilmiş RFM veri tablosunun nihai hali aşağıdaki gibidir. Üç metrik için olan kolonlardaki tüm değerler 0 ve 1 aralığına yeniden atanmış, böylece veriler birbirine daha yakın hale getirilmiştir.

	0	1	2
0	0.127384	0.440379	0.027708
1	0.009248	0.001355	0.017632
2	0.003312	0.100271	0.010076
3	0.007269	0.024390	0.007557
4	0.000544	0.418699	0.000000

Tablo 3.8: RFM Analizi Segment İlişkileri ve Normalize Değerler

Kümele öncesinde elbow methodu kullanılarak kaç kümenin daha tutarlı çıktılar üretebileceği fikrine varılmıştır. Bu method sırasında Within Clusters Sum of Square (WCSS) metriği kullanılır ve kümeler içi kareler toplamını ifade eder. Kümelerin her birinin bir merkez noktası vardır ve WCSS değeri kümeye dahil her noktanın bu merkeze uzaklıklarının kareleri toplamı alınarak hesaplanır. Elbow methodunda WCSS verisinin değişim periyodu takip edilir ve değişimin azaldığı nokta grafikteki dirsek kırılımı noktası kabul edilir (Erarslan, 2023). Mevcut veri seti için en uygun küme adedinin grafiğe bakılarak k=4 olduğu kararı alınmıştır.



Şekil 3.5: Elbow Grafiği

Silhouette Skoru dirsek metodu ile ideal küme sayısı belirlenen modelin başarısı hakkında bir sağlama yöntemidir. Literatürde silüet katsayısı olarak da geçer. Modellemenin ardından kümeleme algoritmasının başarı değerlendirmesi için bu katsayı hesaplanmıştır (Erarslan, 2023).

```
In [43]: from sklearn.metrics import silhouette_score
inertia_list=[]
silhouette_score_list=[]
for i in range (2,10):
    #kmeans = KMeans(init='k-means++', n_clusters=i, n_init=10, max_iter = 300)
    kmeans.fit(data_scaled2)
    silhouette_score_list.append(silhouette_score(data_scaled2,kmeans.labels_))
    print(silhouette_score_list)

[0.6536996186926983]
[0.6536996186926983, 0.6537695009513484]
[0.6536996186926983, 0.6537695009513484, 0.6537604105586737]
[0.6536996186926983, 0.6537695009513484, 0.6537604105586737, 0.6537266357943025]
[0.6536996186926983, 0.6537695009513484, 0.6537604105586737, 0.6537266357943025, 0.653735553910113]
[0.6536996186926983, 0.6537695009513484, 0.6537604105586737, 0.6537266357943025, 0.653735553910113, 0.6537695009513484]
[0.6536996186926983, 0.6537695009513484, 0.6537604105586737, 0.6537266357943025, 0.653735553910113, 0.6537695009513484, 0.6537695009513484]
[0.6536996186926983, 0.6537695009513484, 0.6537604105586737, 0.6537266357943025, 0.653735553910113, 0.6537695009513484, 0.6537695009513484, 0.6537695009513484]
[0.6536996186926983, 0.6537695009513484, 0.6537604105586737, 0.6537266357943025, 0.653735553910113, 0.6537695009513484, 0.6537695009513484, 0.6537695009513484, 0.6537695009513484]
```

Silhouette Skoru, kümeleme performansını ölçmek için kullanılan bir metriktir. Silhouette Skoru, bir veri noktasının kendi kümeleri içindeki diğer noktalara olan uzaklığına (bireysel uyum) ve en yakın diğer kümeye olan uzaklığına (bireysel yabancıklık) dayanır. Silhouette Skoru, her bir veri noktası için bir değer hesaplar. Bu değer, -1 ile 1 arasında bir aralıktadır. Bir veri noktasının Silhouette indeks 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, A., ve Krzyżak, A. 2015).

```
In [42]: kmeans.cluster_centers_

Out[42]: array([[0.00774245, 0.04969146, 0.0209541 ],
                [0.00100921, 0.84067281, 0.00139109],
                [0.00168016, 0.55405092, 0.00435911],
                [0.00235395, 0.2827051 , 0.006979 ]])
```

Inertia cluster, küme içindeki noktaların mesafelerini temsil eder ve modelleme yapılırken bu değer en aza indirilmesi amaçlanır. Yukarıdaki merkezlere olan uzaklıklara bakıldığında, 4 veya 5 kümenin en uygun çözüm olabileceği görülmektedir. K-Means kümeleme algoritmasının temel amacı, k sayıda kümenin ağırlık merkezlerini en iyi şekilde belirlemektir.

# Bölüm 4

## Sonuç

Bu çalışmanın amacı, müşteri segmentasyonu sürecinde RFM (Recency, Frequency, Monetary) analizi ve K-Means kümeleme algoritmalarının etkinliğini değerlendirmektir. Müşteri davranışlarını analiz ederek, pazarlama stratejilerinin daha hedef odaklı ve verimli hale getirilmesine katkı sağlamayı hedeflemiştir.

Bu tez çalışmasının kapsamı, online perakende veri seti kullanılarak gerçekleştirilmiştir. Müşterilerin alışveriş alışkanlıkları analiz edilerek, RFM değerleri hesaplanmış ve bu değerlere göre müşteriler çeşitli segmentlere ayrılmıştır. K-Means kümeleme algoritması kullanılarak müşteri grupları oluşturulmuş ve segmentlerin yüzdelik dağılımları görselleştirilmiştir. Bu analizler sonucunda, müşteri sadakati ve değeri artırma stratejileri için işletmelere önemli bilgiler sağlanmıştır.

RFM analizi aşamalarında, öncelikle veri seti incelenmiş ve müşterilerin alışveriş davranışlarına göre RFM skorları hesaplanmıştır. Recency, Frequency ve Monetary değerlerine göre müşteriler çeşitli kategorilere ayrılmıştır. Bu kategoriler arasında sadık müşteriler, kaybedilme riski taşıyan müşteriler ve yeni müşteriler yer almaktadır. K-Means algoritması ile yapılan kümeleme analizi sonucunda en uygun küme sayısının dört olduğu belirlenmiş ve bu kümeler detaylı olarak incelenmiştir. Silhouette skoru kullanılarak modelin başarısı değerlendirilmiş ve kümeler arasındaki ayrımın net olduğu görülmüştür.

Sonraki çalışmalara yönelik öneriler arasında, RFM analizine ek olarak daha gelişmiş makine öğrenmesi algoritmalarının kullanılması ve müşteri segmentasyonunda farklı veri kaynaklarının entegrasyonunun sağlanması bulunmaktadır. Ayrıca, müşteri davranışlarının daha dinamik bir şekilde izlenmesi ve bu verilerin gerçek zamanlı olarak analiz edilmesi önerilmektedir. Bu yaklaşımlar, müşteri sadakati ve değerini artırmada daha etkili stratejiler geliştirilmesine yardımcı olacaktır.



# Kaynaklar

- Aktepe, C., Bař, M., ve Tolon, M. (2009). *Müşteri İlişkileri Yönetimi*. Detay Yayıncılık.
- Altan, D. (2019). Güncellik/Sıklık/Parasallık (Rfm) Analizi İle Hedef Kitle Seçimi: Hava Yolu Sektöründe Bir Uygulama. [Yüksek Lisans Tezi, Hacettepe Üniversitesi]
- Anderson K., Kerr C. (2002). *Customer Relationship Management*. McGraw-Hill.
- Berry, L. L. (1995). Relationship Marketing of Services- Growing Interest, Emerging Perspectives, *Journal of the Academy of Marketing Science*, 1995, 23(4). 236-245.
- Çakmak, Z., Uzgören N., Keçek G. (2005). Kümeleme Analizi Teknikleri İle İllerin Kültürel Yapılarına Göre Sınıflandırılması ve Değişimlerinin İncelenmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 2005, 12(8), 15–36.
- Chen, I.J., Popovich, K. (2003). Understanding Customer Relationship Management (CRM): People Processes and Technology. *Business Process Management Journal*, 2003, 9(5), 672-688.
- Cheng, C.-H., ve Chen, Y.-S. (2009). Classifying The Segmentation Of Customer Value Via GFP Model And RS Theory. *Expert Systems with Applications*, 2009, 36(4), 4176–4184.
- Couldwell, C. (1999). Loyalty Bonuses. *Marketing Week*, 1999, 2(1), 14.
- Coussement K., Bossche F.A.M, Bock K.W.D. (2014). Data Accuracy's Impact On Segmentation Performance: Benchmarking GFP Analysis, Logistic Regression, And Decision Trees. *Journal of Business Research*, 2014, 67(11) 2751–2758.
- Döşlü, A. (2008). Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi. [Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi]

- Dunham, M.H. (2003). *Data Mining Introductory and Advanced Topics*. Pearson Education Inc.
- Dursun A. (2015). Otel İşletmelerinde Müşteri Profili Oluşturmada Veri Madenciliği Tekniklerinin Kullanılması. [Yüksek Lisans Tezi. Akdeniz Üniversitesi]
- Erarslan, H. G. (2023). Bir E\_Ticaret Firması için RFM Analizi ve Kümeleme Algoritmaları Kullanılarak Müşteri Segmentasyonu ve Analizi [Yüksek lisans tezi, İstanbul Medeniyet Üniversitesi].
- Ergün, E. (2008). Ürün Kategorileri Arasındaki Satış İlişkisinin Birliktelik Kuralları Ve Kümeleme Analizi İle Belirlenmesi Ve Perakende Sektöründe Bir Uygulama. [Doktora tezi, Afyon Kocatepe ÜniversitesiGalatasaray Üniversitesi]
- Ergun, O. (2023). Makine Öğrenmesi Algoritmaları İle Müşteri Segmentasyonu Ve Hepsiburada E-Ticaret Platformu Üzerine Bir Uygulama. [Yüksek Lisans Tezi, Uludağ Üniversitesi]
- Erk, Ç. (2009). Müşteri İçin Değer Yaratma, Müşteri Sadakati Oluşum Süreci ve Şirket Performansına Etkileri Üzerine Araştırma. [Yüksek lisans tezi, Trakya Üniversitesi]
- Fang, B., Ma, S. (2009). *Data Mining Technology and its Application in CRM of Commercial Banks*. First International Workshop on Database Technology and Applications.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *American Association for Artificial Intelligence, AI Magazine, 1996, 17(3), 37-54*.
- Goldenberg, B. (2003). *CRM Automation*. Prentice Hall.
- Hacıfendioğlu, Ş., ve Çolular, N. (2008). İlişkisel Pazarlamada Güven Unsuru ve Otel İşletmelerinde Uygulama. *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 2008, 16(2), 107-126*.

- Hajizadeh E., Shahrabi J. (2010). Application of data mining techniques in stock markets: A survey. *Journal of Economics and International Finance*, 2010, 2(7), 109-118.
- Hastie, T., Tibshirani, R., Friedman, J. (2004). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer Series in Statistics.
- Hosseini, S. M., Maleki, A., ve Gholamian, M. R. (2010). Cluster Analysis Using Data Mining Approach To Develop CRM Methodology To Assess The Customer Loyalty. *Expert Systems with Applications*, 2010, 37(6), 5259–5264.
- Hughes, A. M. (1994). *Strategic Database Marketing*. Probus Publishing.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.
- Kahan, R. (1998). Using Database Marketing Techniques To Enhance Your One-To-One Marketin Initiatives. *Journal Of Consumer Marketing*, 1998, 15(5), 491-493.
- Kanca, S. (2023). Parakende Sektöründe Kümeleme Algoritmaları ile Müşteri Segmentasyonu ve RFM Modeli (Yayın No. Y2213.010019) [Yüksek lisans tezi, İstanbul Aydın Üniversitesi]. YÖK Ulusal Merkezi. <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Karahan, M. O. (2010). Müşteri İlişkileri Yönetimi (CRM) ve Ulaşım Sektöründe Bir Uygulama: Firma Çalışanlarının CRM'ye ve Müşterilerine Bakış Açılarının Belirlenmesi. [Yüksek lisans tezi, Dumlupınar Üniversitesi]
- King, S. F., ve Burgess, T. F. (2008). Understanding Success And Failure İn Customer Relationship Management. *Industrial Marketing Management*, 2008, 37(11), 421–431.
- 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*, 2020, 19(4), 1229-1248.

- Kurtaran, G. D. (2022). RFM Analizi (Recently Frequence Monetary Value, Güncellik Sıklık Para), *YBS Ansiklopedi*, 2022, 10 (1), 20-35.
- MCCARTY, J., & HASTAK, M. (2007). Segmentation Approaches in Data- Mining: A Comparison of RFM, CHAID, and Logistic Regression. *Journal of Business Research* 2007, 7(1), 656-662.
- Miglautsch, J. R. (2000). Thoughts On RFM Scoring. *Journal Of Database Marketing & Customer Strategy Management*, 2000, 8(1), 67–72.
- Mohammed, A. A., ve Rashid, B. (2012). Customer Relationship Management (CRM) in Hotel Industry: A Framework Proposal On The Relationship Among CRM Dimensions, Marketing Capabilities And Hotel Performance. *International Review of Management and Marketing*, 2012, 2(4), 220-230.
- Odabaşı, Y. (1997). *Satış ve Pazarlamada Müşteri İlişkileri*. Der Yayınları.
- Odabaşı, Y. (2004). *Satış ve Pazarlamada Müşteri ilişkileri Yönetimi*. Sistem Yayıncılık.
- Oğur, G. (2023). Müşteri Profili Ve Alışveriş Hareketlerini Belirlemede Rfm Analizi Ve Birliktelik Kurallari Analizi: Perakende Sektöründe Bir Uygulama. [Yüksek lisans tezi, İstanbul Ticaret Üniversitesi]
- Olmeda, I., ve Sheldon, P. J. (2001). Data Mining Techniques and Applications for Tourism Internet Marketing. *Journal of Travel ve Tourism Marketing*, 2001, 11(2/3), 1-20.
- Qiasi, R., baqeri-Dehnavi, M., Minaei-Bidgoli, B., ve Amooee, G. (2012). Developing A Model For Measuring Customer’s Loyalty And Value With GFP Technique And Clustering Algorithms. *The Journal of Mathematics and Computer Science*, 2012, 4(2), 172 - 181.
- Rich, M. (2000). The Direction of Marketing Relationships. *The Journal of Business& Industrial Marketing*, 2000, 15 (2/3), 170 – 191.

- Rygielski , C., Wang, J.-C., ve Yen, D. C. (2002). Data Mining Techniques For Customer Relationship Management. *Technology In Society*,2002, 24(2), 483-502.
- Sarvari, P. A., Üstündağ, A., & Takci, H. (2016). Performance Evaluation Of Different Customer Segmentation Approaches Based On RFM And Demographics Analysis. *Kybernetes*, 2016, 45(7), 1129–1157. <https://doi.org/10.1108/k-07-2015-0180>
- Savaş, S., Topaloğlu, N., ve Yılmaz, M. (2012). Veri Madenciliği Ve Türkiye'deki Uygulama Örnekler. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 2012, 11(21), 1-23.
- Schafer, J. Ben, Konstan, J., & Riedl, J. (1999). Recommender Systems in E-commerce. *Proceedings of the 1st ACM conference on Electronic commerce*, Minnesota, Minneapolis, 158-166. <https://dl.acm.org/doi/10.1145/336992.337035>
- Selvi, M. S. (2007). Müşteri Sadakati, Detay Yayıncılık.
- Şentürk A. (2006). *Veri Madenciliği Kavram ve Teknikleri*. Ekin Yayınevi.
- Silahtaroglu, G. (2016). *Veri Madenciliği: Kavram ve Algoritmaları*, Papatya Yayıncılık.
- Starzewski, A., & Krzyżak, A. (2015). Performance Evaluation of the Silhouette Index. *Artificial Intelligence & Soft Computing: 14th International Conference, Zakopane, Poland, June 14-28, 2015, Proceedings, Part II*, 49–58. [https://doi.org/10.1007/978-3-319-19369-4\\_5](https://doi.org/10.1007/978-3-319-19369-4_5).
- Takala, T., ve Uusitalo, O. (1996). An Alternative View Of Relationship Marketing: A Framework For Ethical Analysis. *European Journal of Marketing*, 1996, 30(2), 45-60.
- Tekin B. (2018). Ward, K-Ortalamlar Ve İki Adımlı Kümeleme Analizi Yöntemleri İle Finansal Göstergeler Temelinde Hisse Senedi Tercihi. *Balıkesir University The Journal of Social Sciences Institute*, 2018, 21(40), 401 – 436.

- Ünal, S. (2011). *Yeni Ekonomide Müşteri İlişkileri ve Bilgi Teknolojileri*. Beta Basım Yayım.
- Wagstaff K., Cardie C., Rogers S., Schroedl S. (2001). Constrained K-means Clustering with Background Knowledge. *Proceedings of the Eighteenth International Conference on Machine Learning*, San Francisco, California, 577–584.
- Wei, J.-T., Lee, M.-C., Chen, H.-K., ve Wu, H.-H. (2013). Customer Relationship Management in the Hairdressing Industry: An Application of Data Mining Techniques. *Expert Systems with Applications*, 2013, 40(7), 7513- 7518.
- Winer, R. S. (2001). A Framework for Customer Relationship Management. *California Management Review*, 2001, 43(4), 89-105.
- Yang, A. X. (2004). How To Develop New Approaches To Rfm Segmentation. *Journal Of Targeting, Measurement And Analysis For Marketing*, 2004, 13(1), 50–60. <https://link.springer.com/content/pdf/10.1057/palgrave.jt.5740131.pdf>
- Yiğit, D. (2017). Müşteri İlişkileri Yönetimi, Veri Madenciliğinin Müşteri İlişkileri Yönetiminde Kullanımı Ve Bir Uygulama. [Yüksek lisans tezi, Galatasaray Üniversitesi]
- Yıldız, D. (2010). Müşteri İlişkileri Yönetimi (CRM) Uygulamalarının İşletme Performansına Etkileri ve Bir Uygulama. [Yüksek lisans tezi, Dumlupınar Üniversitesi]
- Yurdakul, M. (2002). Yeni Bir Pazarlama Stratejisi Olarak Müşteri İlişkileri Yönetimi (CRM)'in Sektörel Bazda Uygulanabilirliği. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 2002, 7(3), 193-202.
- Yurdakul, M. ve Karahan, M. O. (2010). Müşteri İlişkileri Yönetimi (CRM) Ve Ulaşım Sektöründe Bir Uygulama: İşletme Çalışanlarının Müşteri İlişkileri Yönetimi'ne Ve Müşterilerine Bakış Açılarının Belirlenmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 2010, 28(1), 133–144.