



İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
UYGULAMALI MATEMATİK VE HESAPLAMALI BİLİMLER
ANABİLİM DALI

BİR E TİCARET FİRMASI İÇİN RFM ANALİZİ VE
KÜMELEME ALGORİTMALARI KULLANILARAK
MÜŞTERİ SEGMENTASYONU VE ANALİZİ

Yüksek Lisans Tezi

Havva Gülsüm Erarslan

Danışman

Dr. Öğr. Üyesi RÜYA ŞEN

Temmuz 2023



İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
UYGULAMALI MATEMATİK VE HESAPLAMALI BİLİMLER
ANABİLİM DALI

BİR E TİCARET FİRMASI İÇİN RFM ANALİZİ VE
KÜMELEME ALGORİTMALARI KULLANILARAK
MÜŞTERİ SEGMENTASYONU VE ANALİZİ

Yüksek Lisans Tezi

Havva Gülsüm Erarslan

Danışman

Dr. Öğr. Üyesi RÜYA ŞEN

Temmuz 2023

TEZ JÜRİSİ ONAYI

Havva Gülsüm Erarslan tarafından hazırlanan " Bir e-ticaret firması için RFM analizi ve kümeleme algoritmaları kullanılarak müşteri segmentasyonu ve analizi" başlıklı bu Yüksek Lisans Tezi, Uygulamalı Matematik ve Hesaplamalı Bilimler Anabilim Dalı'nda hazırlanmış ve jürimiz tarafından kabul edilmiştir.

JÜRİ ÜYELERİ

İMZA

Tez Danışmanı:

Dr. Öğr. Üyesi RÜYA ŞEN

İstanbul Medeniyet Üniversitesi

Üyeler:

Prof.Dr. Rahmet SAVAŞ

İstanbul Medeniyet Üniversitesi

Doç. Dr. Gülden GÜN POLAT

Gebze Teknik Üniversitesi

Tez Savunma Tarihi: 10/07/2023

BEYANLAR

Yazım ve Kaynak Gösterme Kılavuzu Beyanı

Danışmanlığımda yazılan bu tezin Amerikan Psikologlar Derneği (APA) yazım ve kaynak gösterme kılavuzunda belirtilen kurallara uygun olarak yapılandırıldığı ve bu kılavuzun APA kaynak gösterme standartlarının bu tezde tutarlı olarak uygulandığı tarafımdan incelenerek teyit edilmiştir.

İmza

Dr. Öğr. Üyesi RÜYA ŞEN

Etik İlkelerine Sadakat Beyanı

Hazırladığım bu tezin tamamen kendi çalışmam olduğunu, akademik ve etik kuralları gözeterek çalıştığımı ve her alıntıya kaynak gösterdiğimi beyan ederim.

İmza

Havva Gülsüm Erarslan

ÖZET

***Bir E-ticaret Firması İçin RFM Analizi Ve Kümeleme Algoritmaları
Kullanılarak Müşteri Segmentasyonu Ve Analizi***

Erarslan, Havva Gülsüm

Yüksek Lisans Tezi,

Uygulamalı Matematik ve Hesaplamalı Bilimler Anabilim Dalı

Danışman: Dr. Öğr. Üyesi RÜYA ŞEN

Temmuz 2023

Günümüzde çevrimiçi(online) dolaşım kanallarının yaygınlaşması ile erişim olanakları kolaylaşırken müşteri memnuniyeti farklı bir boyut kazanmıştır. Bu anlamda şirketler müşteri verilerini satış ve pazarlama karar süreçlerinde aktif olarak kullanmaya başlamışlardır. Müşteri kitlesinde yer alan kullanıcıların davranış verilerinin analiz edilerek çıkarımlarda bulunulması ve bu bağlamda satış ve pazarlama hizmetlerine yön verilmesi hususunda iş süreçleri içinde şirketler, müşteri analitiğine yönelmiş ve kullanıcı verileri bu sayede anlam kazanmıştır. Bu tez

alışmasında müşteri kitlesinin bilgisayar yazılımı ile RFM analizi edilmesi, makine öğrenmesi ile yorumlanabilir hale gelmesinin yanı sıra müşteri sadakati ve sürdürülebilirlik bakış açısı ile piyasa değerlendirmesinin yapılabilmesi ve özellikle pandemi sürecinde yeniden şekillenen e-ticaret sektöründeki müşteri profillerinin segmentasyonu amaçlanmaktadır.

Bu kapsamda büyük çapta potansiyel müşteri kitlesine sahip şirketler mevcut ve potansiyel müşterilerle etkileşimini efektif yönetebilmek adına müşteri ilişkileri yönetimine dayanan CRM yaklaşımı benimsenmektedir. Geçmiş satın alma davranışlarının analiz edilerek müşterilerin gruplanması ve kategorize edilmesi ise pazarlama departmanında yer alan veri analistlerince RFM analizi uygulanarak sürece dahil edilmektedir. RFM analizinin temel amacı en doğru hedef kitleye hitap edebilmek ve müşterilerin gelecekteki satın alma davranışlarına yönelik tutarlı öngöründe bulunabilmektir. Müşterilere sunulan ürün ve hizmetler için memnuniyetin en üste taşınabilmesi adına süreç deneyimine yönelik veriler toplanır, bunlar analiz edilerek sürece dair iyileştirmelerde bulunulur. Bu çalışmada da benzer bir süreç uygulanacak olup bir e-ticaret şirketine ait müşteri verileri ve bu kişilere dair harcama davranışları baz alınarak Python programlama yazılımı ile RFM analizinin gerçekleştirilmesi ve makine öğrenmesi kümeleme algoritması kullanılmıştır.

Son zamanlarda çevrimiçi iletişim kanallarının yaygınlaşması ile paralel olarak pazarlama stratejileri de dinamik hale gelmiştir. Veri biliminin de olgunluk kazanması ile bu gibi soru işaretlerine yapay zekâ ile yazılımsal çözümler getirilmesi günümüz tüketim toplumuna objektif ve gerçekçi bir yorum getirmiştir. Bu çalışma çerçevesinde analiz çıktıları ile e-ticaret pazarına ışık tutulması planlanmıştır. RFM analizi aşamaları uygulanmış ve e-ticaret pazarında gerçek veri seti ile çalışılmıştır. Ayrıca literatüre bu alan ile ilgili Türkçe kaynak kazandırılması hedeflenmiştir.

Anahtar Kelimeler: RFM Analizi, Müşteri Segmentasyonu, E-ticaret, Birliktelik Kuramı, CRM, Veri Madenciliği, Öznitelik Çıkarımı

ABSTRACT

Customer Segmentation Using RFM Analysis and Customer Segmentation and Analysis with Using Clustering Algorithms

Erarslan, Havva Gulsum

Master's Thesis, Applied Mathematics and Computational Sciences

Supervisor: Asst. Prof. RUYA SEN

July 2023

Nowadays, with the widespread use of online circulation channels, access opportunities have become easier and customer satisfaction has gained a different dimension. In this sense, companies have started to use customer data actively in their sales and marketing decision processes. Companies have included customer analytics in their company's processes in order to analyze the behavior data of the users in the customer group and make inferences, and in this context, direct sales and marketing services. User data has gained meaning in this way. In this thesis, it is aimed to analyze the RFM with computer software, to make it interpretable with machine learning, as well as to be able to evaluate the market with a customer loyalty and sustainability perspective, and to segment the customer profiles in the e-commerce sector, which has been reshaped especially during the pandemic process.

In this context, companies with a significant number of potential customer base adopt a CRM approach based on customer relationship management in order to effectively manage their interactions with both existing and potential customers. Grouping and categorizing customers based on analyzing their past purchasing behavior is included in the process by applying RFM analysis by data analysts in the marketing department. The main purpose of RFM analysis is to appeal to the most accurate target audience and to make consistent predictions about the future purchasing behavior of customers. In order to maximize satisfaction with the goods and services offered to customers, data on process experience is collected, these are analyzed and improvements are made regarding the process. In this study, a similar process will be applied, and RFM analysis with Python programming software and machine learning algorithms will be used based on the customer data of an e-commerce company and the spending behaviors of these people.

Recently, marketing strategies have also become dynamic in parallel with the widespread use of online communication channels. With the development of data science, artificial intelligence and software solutions to such question marks have brought an objective and realistic interpretation to today's consumer society. Within the scope of this study, it is planned to shed light on the e-commerce market with the analytical results. By applying the stages of RFM analysis, real data set will be studied in the e-commerce market and it is desired to bring Turkish resources to the literature.

Keywords: RFM Analysis, Customer Segmentation, E-commerce, Association Theory, CRM, Data Mining, Attribute Extraction

TEŞEKKÜR

Tez çalışmam boyunca özellikle literatür takibinde ve yazılım projesi geliştirmesi aşamasında bana kıymetli vakitlerini ayıran ve beni yönlendiren Danışman Hocam Dr. Öğr. Üyesi Rüya Şen'e ve veri analitiği alanındaki teknik konularda değerli yönlendirmelerini aldığım Dr. Volkan Oban hocama, tez çalışmasında kullanılmak üzere veri seti arayışında iken iş birimleri ile görüşerek akademik çalışma çerçevesinde gerçek veri edinmemizi sağlayan Abdül Meral'e sundukları katkılarından dolayı teşekkürlerimi sunarım. Bu süreçte desteğini esirgemeyen aileme minnettarım.

İÇİNDEKİLER

ÖZET	ii
ABSTRACT	iii
TEŞEKKÜR	iv
İÇİNDEKİLER	1
1. GİRİŞ	2
1.1. Yeni Nesil Alışveriş Kültürü	2
1.2. Maslow İhtiyaçlar Hiyerarşisi	3
1.3. Pandeminin Pazarlama Sektörüne Etkileri	6
1.4. Pazarlama sektöründe Veri Analitiği	8
2. RFM ANALİZİ	12
2.1. Müşteri Kullanıcı Deneyimi	17
2.2. Kullanılan Veri Setinin Öznitelik Çıkarımı	19
2.3. Müşteri Segmentasyonu	20
2.4. RFM Analizinin Sektörde Kullanım Alanları	22
3. PYTHON PROGRAMLAMA İLE RFM ANALİZİ ÖRNEĞİ	25
3.1. Kod Tabanında Kullanılan Python Kütüphaneleri	27
3.2 Kullanılan Veri Seti Özellikleri	31
3.3 Keşifsel Veri Analizis	33
3.4. RFM tablosu oluşturma	40
3.5. RFM Skoru Hesaplama ve Müşteri Segmentasyonu	45
3.7. KMeans Kümeleme Algoritmasının Uygulanması	48
4. TARTIŞMA VE SONUÇ	55
KAYNAKLAR	69
TABLolar LİSTESİ	Error! Bookmark not defined.
ÖZGEÇMİŞ	Error! Bookmark not defined.

1. GİRİŞ

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. Bu bölümde alışveriş süreçlerinin sosyokültürel yansımaları, müşteri metriklerinin anlamlandırılmasının ve analiz edilerek verinin bilgiye dönüşümü işlenecek, sonraki bölümlerde bu bilgiler ışığında RFM analizinin sektördeki kullanım alanlarına ve müşteri segmentasyonu ile ihtiyaç haritasının belirlenmesi uygulamalarına yer verilecektir.

1.1. Yeni Nesil Alışveriş Kültürü

Önceleri ürün veya hizmete başvurma alışkanlıkları ihtiyaç analizi üzerine şekillenirken şimdilerde ihtiyaç olmaksızın tüketiciyi etkisi altına alan reklam sunumları ile satın almaya sevk etmekte ve pazarlama teknikleri bu etki alanı üzerine şekillenmektedir.

Son yıllara damgasını vuran çevrimiçi alışveriş platformları müşteri kitlesini giderek artırmakta ve ürün çeşitliliğini en üstte tutarak özendirici kampanyalar ile bunları tıpkı bir ihtiyaçmışçasına tüketicilerin beğenisine sunmaktadır. Benzer politika fiziki mağazalarda ve alışveriş merkezlerinde de karşımıza çıkar. Sadakat kartları ile süreklilik hedefleyen mağazalar çeşitli banka anlaşmaları ile kurdukları marka puan uygulamalarını hayata geçirerek müşterilerinin alışveriş deneyimlerine alternatifler getirmektedir. Zaman kısıtı ile sınırlandırma getirilen bu marka puanları kısa sürede tüketilmesi hedeflenen yeni satın alma davranışlarına teşvik eder. Bu da sizi o markaya bağlar ve alışveriş döngüsünün sürekliliğini garantiler.

Özellikle de gelenekselleşmiş özel günleri ön planda tutarak hediye alımını sık sık tetikleyen periyodik kampanya sezonları yine sonsuz alışveriş döngüsünün bir parçasıdır. Anneler veya babalar günü gibi spesifik özel günlerde belirli ürün gruplarında yönelim artarken Avrupa’da “Black Friday” olarak kitleleri harekete geçiren alışveriş sezonu Kasım-Aralık aylarında çılgın alışveriş günleri olarak ülkemizde etkisini göstermekte ve elektronik eşyalar başta olmak üzere hemen her ürün grubunda talebi en üst seviyede tutmaktadır. Bu yansımaları hem fiziki mağazalarda hem de çevrimiçi platformlarda en net şekilde görebiliriz. Markalar tarafından birbiri ardına gelen agresif pazarlama atakları tüketicileri heyecanlandırıp kampanya teşvikleri ile satın almaya tetikler. Bu da ihtiyaç analizi gerekmeksizin direkt olarak aksiyona dönüşür. Bu sirkülasyon ürün veya hizmet sunan pazar payına sahip her üreticinin elde etmek istediği canlı ortamdır ve müşterilerini segmente ederek doğru teknikler ile en verimli üretimi ve en çok satışı hedefler.

1.2. Maslow İhtiyaçlar Hiyerarşisi

Abraham Harold Maslow 1908-1970 yıllarında yaşamış psikoloji profesörü bir akademisyen ve psikologdur. İnsan psikolojisi üzerine derin araştırmalar yapan Maslow, hümanist bir bakış açısı benimsemiş ve çalışmalarında insanın ihtiyaçları doğrultusunda motivasyonunun ne denli etkili olabileceği üzerinde durmuştur. Bu konseptte 1943 yılında “İnsan Motivasyonuna Yönelik Bir Teori” adlı makalesini sunmuş ve insan psikolojisinde tercihler ve davranışlar ile ihtiyaçlar arasında pozitif bir korelasyon olduğunu öne sürmüştür.



Tablo 1: Maslow İhtiyaçlar Hiyerarşisi Piramidi[]

Maslow İhtiyaçlar Hiyerarşisi olarak da adını sıkça duyduğumuz bu bakış açısı Tablo 1 de yer alan piramit görsel ile sık sık makalelerde yer bulur. Bu anlayışa göre ihtiyaçlarımız tercihlerimizi doğrudan ilgilendirir. Piramidin tabanında fiziksel ihtiyaçlar yer alır. Bunlar ekmek su kadar çok temel ve hayatta kalmamıza olanak sağlayan 1. dereceden önemli ihtiyaçlardır. Açlık anında diğer kademelerdeki ihtiyaçlarımıza henüz sıra gelmez ve önceliğimiz mutlaka yiyecek ve barınma gibi temel ihtiyaçlar olur. Bunların sağlanması halinde tokluk anında hayatta kalacağımıza nispeten psikolojik olarak ikna oluruz güvende hissetmek isteriz. Bu da 2. Kademe olarak karşımıza çıkar. Kendimizi koruma altına almak ve tehlikelerden arınmak gibi yine temel ve hayati önem taşıyan ihtiyaçlar davranışlarımıza yön verir.

Fiziksel ve güvenlik ihtiyacının karşılanması neticesinde temel ihtiyaçlarda doyuma ulaşılmıştır ve ihtiyaç hiyerarşisi yerini sosyal

ihtiyaçlara bırakır. Bunlardan ilki aidiyet ve sevgi ihtiyacı olarak karşımıza çıkar. İnsanoğlu tıpkı temel ihtiyaçlar gibi sevmek ve sevilme ihtiyacını da bünyesinde barındırır, bunun eksikliğini hissederek ve tamamlanma duygusu ile sosyal bağlar kurar ilişkilerimizde bu duyguları yaşatırız. Maslow İhtiyaçlar Hiyerarşisinde bir diğer kademe değer ihtiyacıdır. Kurduğumuz ilişkilerde prestijli olduğumuzu hissettiren her bağ bize değer duygusu olarak yansır. Benzer tatmin başarılı olduğumuz ve bunun neticesinde ödüllendirildiğimizde de hissedilmektedir. Bir tebrik veya terfi durumunda gerçekleşen başarıma duygusunun size ne kadar da değerli ve prestijli hissettirdiğini kolayca anımsayabilirsiniz. Kendinize duyduğunuz özsaygı ve size duyulan saygı, bu değer duygusunu pekiştirir. Markalar da özendirme kampanyalarında size bizzat bu hazzı tattırmak ister. Size özel sunulan kampanya teklifleri, birebir danışmanlık alabildiğiniz hizmetler ve yine size özel ürün sunumları aslında etki alanını artırarak o markaya veya hizmete tekrar tekrar başvurmanıza teşvik eder. Bankacılık uygulamalarında yapay zeka robotlarının müşterilere ismi ile hitap etmesinin pozitif yansımaları yine pazarlama sektöründe de gözlemlenmektedir. İhtiyaç piramidinde son kademe kendini gerçekleştirme olarak karşımıza çıkar. İnsan doğasında çok ciddi bir potansiyeli barındırır. Özde yer alan bu potansiyelin açığa çıkması ancak diğer ihtiyaçların doğru ve dozunda sağlanması ile mümkündür. Kendini gerçekleştirme hazzı kişiyi gelişmeye ve potansiyelini en üstte değerlendirmeye sevk eder. Dijital pazarlama sektöründe de tüm bu ihtiyaçlara hitap ederek tüketicinin motivasyonunu canlı tutmak hedeflenir. Hiyerarşide aşağıdan yukarı olarak kademelendirme temel ihtiyaçlardan psikolojik ihtiyaçlara doğru ilerler. Bu da motivasyonumuzu odaklar ve ihtiyaçlarımız buna göre şekillenir. İhtiyaç hiyerarşisi kuralına göre bir kademeden diğerine geçiş için mutlaka alt kademelerde doyuma ulaşılmalıdır. Temel yaşamsal

ihtiyaçlar tam olarak karşılanamadan üst kademelerde yer alan psikolojik beklentilere henüz ihtiyaç hissetmeyiz. Kademelerde yukarı doğru çıkıldıkça değer ve saygınlık motivasyonu içerisinde oluruz. Pazarlama sektörü tam olarak müşterilerinin beklentilerine hitap edebilmek ve doğru hizmet sunabilmek için size daima spesifik olarak size özel kampanya ve tekliflerle hizmet sunarak memnuniyetinizi en üste taşır bu da tüketim davranışlarına teşvik eder. Fiziksel ve çevrimiçi alışveriş platformları artık farklı bir boyuta taşınmış ve ihtiyaç hiyerarşisinden bağımsız olarak tüketiciyi etkisi altına alarak ihtiyaç hissettirme politikasına gitmiştir. Çevrimiçi kanalları düşünürsek sepete eklediğiniz her ürün için veya öncesinde favorilere eklediğiniz ürün kategorisinde benzerleri de size listeleyerek bununla ilgili size bildirimlerde bulunarak sizi sahip olmaya motive eder. Herhangi bir konuda farklı amaçla bir web site ziyaretinizde sizin tarayıcı geçmişinizden ürün kategorilerinizi saptayarak Banner reklamları olarak karşınıza çıkar. Bu da o ürün veya hizmete ziyaretçi trafiği sağlamak adına kullanılan en yaygın pazarlama tekniklerinden biridir.

1.3. Pandeminin Pazarlama Sektörüne Etkileri

Birkaç yılı etkisi altına alan pandemi ile hemen her sektör gibi pazarlama sektörü de yakından etkilenmiştir. Çevrimiçi platformların önem kazandığı bu dönemde yeme-içme ve giyim sektöründe ürüne ulaşma hızı en üst seviyeye ulaşmış, iade politikaları da olabildiğince esneklik kazanmıştır. Böylece alışveriş süreçleri çok daha çevik bir hal almış ve kitleler hızla çevrimiçi platformlara yönelmiştir. Salgın sürecinde izolasyon devam ederken fiziki mağazalara ilgi olabildiğince azalırken en temel ihtiyaçlar dahi dijital platformlara taşınmış ve bu da alışkanlıkları oldukça değişime uğratmıştır. Önceleri çevrimiçi alışverişe mesafeli duran birçok kullanıcı biraz da mecbur kalarak

çareyi sürece adapte almakta bulmuş ve hemen herkes bu izolasyon sürecinde dijital platformları deneyimlemiştir. Sürecin başlarında korku ve endişe ile kıtlık psikoloji içerisine giren tüketim toplumu tercihini sağlık ve hijyen ürünlerine yöneltirken sonrasında birçok sektör evden çalışma kararı aldığından yerini ev giyim ürünlerine bırakmıştır. Yine aynı dönem mutfağa olan yönelim nedeniyle gıda ve mutfak gereçlerine olan talep artarken birkaç ay sonunda izolasyon koşullarının ve önlemlerin hafifletilmesi ile birlikte yerini ev spor aletleri ve kamp ürünlerine bırakmıştır. Bu geçişlerden de anlaşılacağı üzere alışveriş trafiğine ve talep gören ürün kategorilerine bakarak hayatın nabzını tutabiliriz.

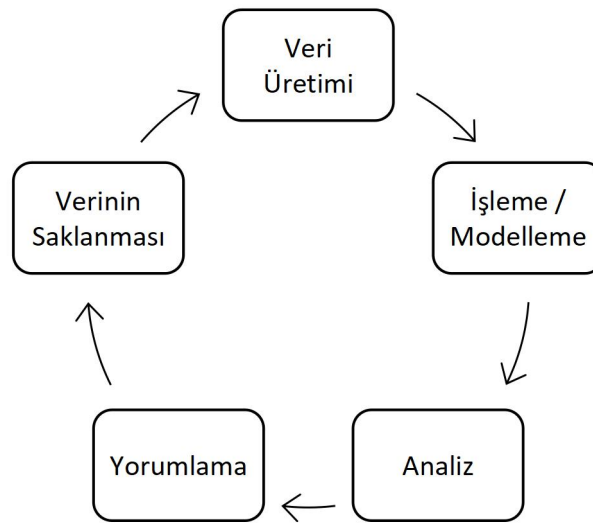
Hem dijital platformlar hem de markalar bu süreci fırsata çevirerek pazarlama faaliyetlerinde iş birliğine gitmiştir. Fiziki mağazalarında ciddi müşteri popülasyonu kaybı yaşayan zincir mağazalar bu tehdidi dijital platformlarla iş birliğine giderek avantaja çevirmiştir. Yüksek meblağlara ulaşan mağaza kira giderlerinde ve personellerine ayırdığı maaş bütçesinde küçülmeye giderek oluşan bütçe payını üretime ve ürün çeşitliliğine adreslemiş ve çevrimiçi platformlarda tüketicilerinin beğenisine sunmuştur. Toptan ve perakende zincirlerinin birçoğu mobil ve web sitelerine ciddi yatırımlar yaparak işlem kapasitesini artırmış kitlelere hitap edebilmek adına veri saklama ve sunucu kapasitelerinde de yayılma politikası izlemiştir. Bunun yanı sıra çoğu marka dijital çağın gelişmelerine ayak uydurarak müşteri kullanıcı deneyimini iyileştirebilmek adına sanal gerçeklik ile ürünü fiziki olarak edinmeden fiziksel haline en yakın deneyimi tüketicilerine sunmuştur. Bunu artırılmış gerçeklik (AR) teknolojisi ile hayata geçiren bazı yenilikçi markalar o ürüne ulaşmadan gerçeğine en yakın renk ve beden karşılaştırması imkânı sunmaktadır. Aynı zamanda yapay zeka ve görüntü işleme özelliklerini web sitelerine entegre eden markalar görsel ile arama deneyimini kullanma aşarken görüntü işlemeden

yararlanarak kategoriler arası filtreleme veya tüketicinin beğenisine göre tarzına en yakın ürünleri listeleme gibi fonksiyonları da hem mobil hem web sitesi uygulamalarında kullanıma açmıştır. Bu gibi alınan aksiyonlar tüketicinin alışveriş deneyimini zenginleştirirken sektördeki rekabet ortamına da yeni bir soluk getirmektedir.

1.4. Pazarlama sektöründe Veri Analitiği

Üretim zincirleri tüketicinin taleplerine en doğru şekilde hitap edebilmek ve arz talep dengesini optimize edebilmek adına kapsamlı stratejiler geliştirmektedir. Planlamadan üretim hattına ve oradan da son kullanıcıya ulaşan ürün veya hizmet en hızlı şekilde tam olarak tüketicinin talebini karşılamalı ki beğeni oluştursun, bu sayede müşteri sadakatini de beraberinde getirsin. Tüm bu beklentiler ancak son tüketiciyi doğru anlamaktan geçiyor tabii bu da ancak müşteri verilerinin sistematik bir şekilde tutularak ve anlamlandırılıp analiz edilmesi ile mümkündür. Pandeminin etkisi ile çevrimiçi platformlardaki pazar payının nasıl etkilendiği hakkında önceki bölümde bilgi verilmiştir. Artan pazar hacmi fırsatları barındırırken birçok marka için de tehdit oluşturmuştur. Bu süreçte veri analitiğinin değeri bir kez daha anlaşılmış ve pazarlama stratejilerinde bu analizlere yer verilmiştir. Bu rekabet ortamında teknolojinin gerisinde kalmayan zincir mağazalar iş zekâsı ve veri analitiği gibi ekipler kurmuş, işlem verilerini muhafaza ederek bunların analizine gitmiştir. Bir sonraki bölümde değinilecek olan RFM analizi de tam olarak bu ihtiyaca istinaden yapılmaya başlanmıştır. Günlük işlem hacmi çok geniş çevrimiçi alışveriş platformları ve büyük zincir mağazaların her veriyi tutabilmesi elbette ki verimli olmayacaktır. Bu nedenle ilk etapta verilere karar verilerek bunların kirli veri dediğimiz analizi olumsuz etkileyebilecek veya sapmalara yol açarak çıktıları yanıltabilecek veriler veri setinden arındırılarak işe başlanır. Sonrasında ortaya çıkan daha

temiz veri anlamlandırılmaya çalışılır. Veri analitiği çok disiplinli bir alan olduğundan birçok branştan uzmanlık sahibi kişilerce verinin analizi yorumlanması işlemleri mümkündür. Sürecin hem yazılım aşaması varken işin içerisine istatistik ve finans da dahil olduğunda birçok disiplinden uzman kadro aynı veri üzerinde aşama aşama birlikte çalışırlar. Uzman kadro tarafından veri analiz edilirken iş süreçleri aşağıdaki diyagramdaki adımlarla takip edilir.



Tablo 2: Veri Analitiği İş Akışı Diyagramı

Bu diyagramdaki ilk iş birimi verinin üretimi ve muhafaza edilmesidir. Veri akışını anlık olarak takip etmek ve bunlardan analize girecek olanları belirleyerek veri tabanına aktarılmasını, kalıcı olarak tutulmasını sağlamak adımlarını içerir. Bu aşamada kişisel verileri koruma konunun el verdiği ölçüde veri tutulabilir veya maskeleye başvurulur. Kurumlar yalnızca müşterinin beyanı olması halinde verileri işleme yetkisini edinebilir. Müşteriye ait işlem verisi bir ID ye atanır ve kişisel veriler müşteri numarası ile maskelenmiş olur. Veri analitiğinde hukuki bir sorumluluk altına girmemek adına buna sıkça başvurulur.

Bir sonraki aşama verinin işlenmesi adımı büyük veriden analize girmeyecek olan veya standart sapmayı etkileyecek yanıltıcı değerlerin

çıkarılmasını içerir. Verinin temizlenmesi yani modellemeye hazır hale getirilmesidir. İşlenen veriyi modellemek her zaman için daha tutarlı sonuçlar almamızı sağlar. Örneğin bir veri setinde standart sapma çok yüksek ise veya farklı nitelikte birimler içeriyorsa bunu modele dahil etmeden önce normalizasyon işlemi uygulanır ve tüm değerler normalize edilerek 1 ve 0 değerleri arasında bir sayısal değer alması sağlanır. Bunun akabinde veri seti modellenir ve bu aşamada ihtiyaç halinde veri setinin niteliklerine uygun olan makine öğrenmesi algoritmaları uygulanır.

Analiz aşamasında modellemenin ardından istatistiksel yöntemlere başvurularak çıktılar detaylandırılır. Python programlama ile de analiz raporlarını edinebilmek mümkündür. Çok sayıda python kütüphanesi istatistiksel analiz raporu üretebilmektedir. Örneğin bu tez çalışmasında SweetViz Python kütüphanesi 2.1.4 versiyonu kullanılarak bir rapor üretilmiş ve çalışmada buna yer verilmiştir. Bunun yanı sıra veri görselleştirmeye yarayan iş zekâsı araçları da iş birimlerince yaygın olarak kullanılmaktadır. Nicel ve nitel verilerin analiz edilerek istatistiksel metriklerinin çıkarılması ve bunun görselleştirilmesi işlemleri Keşifsel Veri Analizi (EDA) olarak geçer ve verinin yorumlanabilmesi için ilk adımdır.

Verinin yorumlanması aşamasında analiz sonrası edinilen metrikler yeterli ise bunları anlamlandırılmaya gidilir. Daha karmaşık ve farklı dinamikler içeren bir alanda bu gibi bir analiz gerçekleştiriliyorsa keşifsel veri analizine, istatistiksel verilerin detaylandırılmasına ve öznitelik çıkarımına gidilebilir. Bu tamamen proje kararıdır ve sahadan alınan verinin yön vermesi ile ihtiyaçlara bağlı olarak planlanır. Çıktılar sonrasında yorumlanarak ilgili probleme dayanarak bizleri çözüme ulaştıracak çıkarımlarda bulunulur. Birçok şirket bu analizleri ileriye dönük tahmin verileri ile de pekiştirerek sonraki iş süreçlerini

planlarken yer vermektedir. Veri analizinin çok disiplinli bir alan olduğundan söz etmiştik, analiz sonrası yorumlanırken ekipteki farklı uzmanlık alanlarına sahip uzman kişilerin görüşleri alınarak yorumların çeşitliliği artırılır bu da ileriye dönük çıkarımlarda tutarlılığı artırarak probleme başka bakış açılarından da çözüme gidilmesine olanak tanır. Şirketler bu inisiyatif çerçevesinde vizyonunu geliştirmek adına akademik danışmanlıklara başvurmaktadır.

Verinin muhafaza edilmesi sürecinde yasal regülasyonlar gözetilmek durumundadır ve o verinin saklanıp işlenebilmesi için yine müşteri beyanına ihtiyaç duyar. Kişisel verilerin korunması kanunu son yıllarda bunu gözeterek ciddi önlemler almıştır. Zamanla dijitalleşen platformlar veri tabanlarında yer verdikleri her türlü özel nitelikli verinin güvenliğinden sorumludur ve bunları ticari iş birlikleri için müşteri izni olmadan üçüncü kişilere aktaramaz. Alınan güvenlik önemlerine rağmen zincir mağazaların veya dijital platformların veri tabanlarına yapılan siber saldırıları zaman zaman duyuyoruz. Kullanıcı bilgilerine izinsiz erişmeye yönelik bu saldırılar veri gizliliğini ihlal ederken hukuki yaptırımlara da tabiidir. Buna engel olmak adına ek yazılımlar ile ve sistem altyapılarını fiziki sunucular yerine bulut tabanlı sistemlere taşıyarak şirketler veri tabanlarını daha güvenli hale getirmek için önlemler almaktadır.

Veri analizi iş akışından edinilen çıkarımlar verinin bilgiye dönüşmesi sürecidir. Sayısal verileri görselleştirmek ve yorumlamak üzere sorumlulukları bünyesinde barındıran iş zekâsı birimleri proje yönetimi ekipleri ile yakın çalışırlar. Bu bilgiler ışığında proje planlarında geleceğe dönük tutarlı tahminlerde bulunulabilir. Perakende sektöründe bir giyim markasından yola çıkacak olursak, işlem trafiğine bakılarak en çok trafiği mağazalardan mı yoksa çevrimiçi kanallardan

mı yakalayabildiği hakkında bilgi edinebilir. Çevrimiçi kanallar kendi web sitesi, mobil uygulaması veya iş birliği içerisinde olduğu diğer platformlar olabilir. Buradaki trafiğe bakılarak bir sonraki sezon için iş birliklerine karar verilir. Pazara sunulan ürün yelpazesi için de yine veri analitiği çalışmaları önem arz eder. Bir ürünün ne derece talep gördüğü bir sonraki üretim hattındaki kaderini belirler. Buna istinaden kimi ürün kategorisine üretimde daha çok ağırlık verilirken, yeterince talep görmeyen bazı ürünler için de tekrar değerlendirmeye gidilir. Bunun kaynağı ürünün yeterince reklamı yapılmamış ve henüz keşfedilmemiş oluşu da olabilir. Bu gibi öngörüler ve alınacak aksiyonları pazarlama ekipleri gerçekleştirebilir. Sahadan alınan veri analizi raporlarına bakılarak ilgili ürün kategorisi için daha fazla özendirme kampanyalarına yer verilir. Bu gibi kampanyaları size doğrudan ulaştırabilmek adına markalardan sıkça izninize başvurulur. Müşterilere kısa mesaj, elektronik posta gibi iletişim kanallarından kişisel kampanyalar sunarak alışveriş trafiğini canlı tutmak hedeftir. Üretim zincirinde benzer ürün kategorilerine hizmet eden diğer marka zincirlerini de düşünürsek, bu satış hedefleri ciddi bir rekabet ortamı ortaya koyar. Markalar geride kalmamak adına rakip firmaların pazarlama stratejileri ve ürün yelpazesine de hâkim olmak durumundadır. Tüm bu sebeplerle çağın gerisinde kalmayan ve teknolojik gelişmelere ayak uydurabilen üreticilerin pandemi gibi en zor koşulları dahi çevik bir şekilde fırsata çevirdiğini ve bu sayede sektöre tutunabildiklerini görülmektedir.

2. RFM ANALİZİ

Satış ve pazarlama süreçlerindeki yoğun rekabet ortamı şirketleri harekete geçirmiş ve bu da planlama yaparken sahayı doğru ön izlemek ve ileriye dönük tutarlı öngörülerde bulunabilmeyi bir ihtiyaç haline getirmiştir. Böylece müşteriye ürün veya hizmet ulaştıran söz konusu şirketler veri analitiği alanında en yeni teknoloji trendlerini yakından takip ederek sahadan veriler toplayıp bunların anlamlandırılması ile potansiyel fırsatları değerlendirme şansı bulmuştur. Dijitalleşmenin hız kazandığı günümüzde ürün veya hizmete ulaşmak kolaylaşırken alternatif alışveriş platformları da nicelik olarak artış göstermiş ancak pandeminin etkisiyle piyasada takip edilen agresif büyüme müşteri perspektifinden alternatiflerin artması ile birlikte yerini çoğu firma için müşteri kaybına bırakmıştır.

Sektöre tutunmada belirleyici bir faktör olan ciddi müşteri sirkülasyonu kimi şirketin sonu olurken teknolojiden geri kalmamanın avantajını değerlendiren şirketler için de avantaja dönüşmüştür. Burada ciddi bir potansiyel olduğunu fark eden pazarlama sektörü de müşterileri tek bir çatı altında toplamaktansa profillere ayırma yöntemine gitmiştir. Müşteri segmentasyonu olarak karşımıza çıkan bu çalışmalarda en yaygın kullanıma sahip analiz yöntemlerinden biri de RFM analizidir. Bu analiz modelinde müşterilerin geçmişe dönük satın alma davranışları incelenir ve buradaki nihai amaç parametrik olarak tutulan veriler ışığında kategorize edilen müşteri profillerini görünür kılmak ve bu segmentlerin beklentilerini doğru analiz ederek her müşteri profiline en uygun ürün veya hizmeti sunmaktır. Yine bu bilgiler ışığında kayıp olarak görülebilecek memnuniyet oranı asgari olan müşteri popülasyonunu özel olarak değerlendirip çeşitli kampanyalar ve özendirme politikaları ile onları geri kazanmak mümkündür. Müşteri memnuniyetinin sürekliliği de çok önemlidir bu da müşteri sadakati olarak ele alınır. Bir önceki segment kadar tehdit olarak algılanmasa da bu segment de müşteri popülasyonunda geniş yer tutar ve genelde konfor alanını gözetken aynı tüketim eğilimi gösteren profiller için geçerlidir. Bu gibi müşteriler için de sadakat

kartları üzerinde alıřılır. Bylece markaya ait harcamalarınız size puan olarak dner bu da belirli bir sre ile sınırlandırılır ve bylece o sre zarfında ihtiyacınız olsun ya da olmasın aynı rn veya hizmetten tekrar faydalanma eęilimi gsterirsiniz. Dięer bir kategori de potansiyel mřterilerdir ve pazarlama stratejilerinde byme hedefi olan oęu řirket iin bunun keřfi byk nem tařır. Rekabet ortamındaki hızlı mřteri sirklasyonunu bylece avantaja evirebilir ve trafięi kendine ekebilir. Genelde sahadaki reklam ve pazarlama alıřmaları yeni mřteri edinme gdsn barındırır. Bu da dijitalleřen aę ile birlikte sosyal medya platformlarında veya ana akım medyada sıka karřımıza ıkar. Kendimizden yola ıkacak olursak daha ncesinde tercih ettięimiz bir rne reklamlarda rastlamak eęer yeni bir versiyonu deęil ise sırf reklamını grdęmz iin o rn tekrar edinme eęilimi gstermeyiz ancak kampanyalı ise ve cazip bir teklif sunulmuř ise aynı rn ilgimizi ekebilir. Onun dıřında ilk kez reklamlarda karřılařtıęımız bir rn merak etmek, arařtırmak, rn yorumlarını incelemek ve satın alma davranıřı gstermemiz ok daha olasıdır.

Mřteri poplasyonunu geniřletmek adına sık sık reklam kampanyalarına bařvurulur. İlgili rn veya hizmet iin ilk alıřveriřinize zel indirimler tanımlanır veya hoř geldin kampanyaları sunulur bylece yeni kiřiler de ilk kez o rn veya hizmeti deneyimleme eęilimi gsterir. TV programlarındaki rn yerleřtirmeler, reklam kuřakları ve sosyal medyada influencer olarak nitelendirilen ve kitleleri etkisi olarak nlenen kiřilerce sosyal medyada tanımlar yapılarak lansmanı yapılır. řirketler piyasaya sunmuř olduęu rn veya hizmeti dijital kanallar ile geniř kitlelere ulařtırma fırsatı yakalar ve potansiyel mřterilere de bu sayede ulařılır.

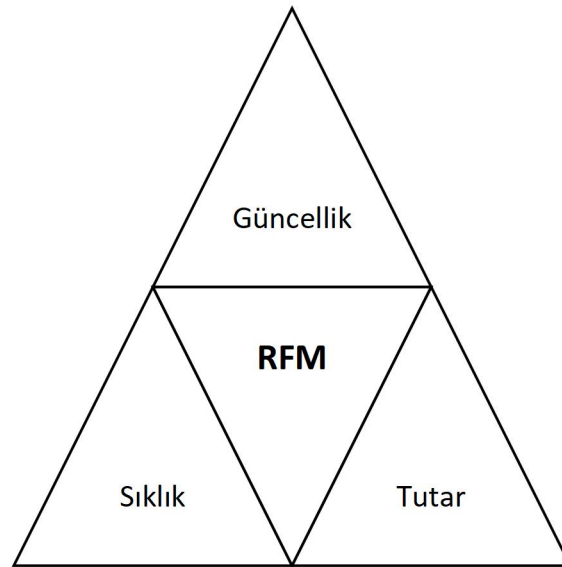
Mevcut mřterilerden alınan veriler ile yapılan analizler gemiře dnk veriler kullanılarak yapılırken aslında hem srecin ncesine hem de bugne ıřık tutar. Buna ek ileriye dnk ngrde bulunmak ve satıř tahminleri neticesinde pazarlama hedeflerini řekillendirmek de

mümkündür. RFM analizi tam olarak burada karşımıza çıkar. Müşterilerin geçmişe dönük satın alma davranışları ile müşteriler segmentlere ayrılır ve istatistiksel analizler sonrası geleceğe dönük pazarlama stratejileri geliştirilir.

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 Analizinin temelini 3 ana metrik oluşturur. Bunlar aşağıdaki gibidir.

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



Tablo 3: RFM Analizi Bileşen Metrikleri

Müşterilerin geçmişe dönük verilerinden elde edilen ilgili metrikler ilk etapta sayısal veri olarak ele alınır sonrasında her başlık 1'den 5'e kadar sınıflandırılır ve 111 den 555 e kadar elde edilen 3 haneli değerler analiz sonunda RFM skoru olarak belirlenir ve nitel veri olarak değerlendirilir.

Güncellik; müşterilerin o ürün veya hizmeti en son ne zaman tercih ettiğine dayanan tarih formatında karşılaştırma yapılarak bulunan bir nicel veridir. Eğer veri setindeki işlemler günümüze kıyasla çok daha eski tarihe dayalı ise bunun için bir referans tarih belirlenir ve her bir işlem verisinin o tarihe uzaklığı gün bazında hesaplanır. Böylece o ürün veya hizmeti en son kaç gün önce tercih ettiği saptanır. Böylece RFM analizine Güncellik (Recency) parametresi olarak bu değer dahil edilir. Örneğin bir e-ticaret sitesinden yola çıkacak olursak günlük kullanıcı trafiği ne kadar önemli ise güncellik değeri de bir o kadar önem taşır. İlgili platformu en yeni ziyaret eden müşteri profili değer kazanırken güncelliğini yitirmiş yani uzunca süredir o platformu tercih etmeyen müşteri profili için kayıp olmaması adına tekrar kazanmak için ekstra kampanyalar sunulması ve geri kazandırılması gerekir.

Sıklık; müşterilerin o ürün veya hizmeti tercih etme frekansına bakılarak hesaplanır. Yine e-ticaret örneğinden gidecek olursak müşteri ID si başına düşen işlemler gruplanarak bunlar tarih formatında karşılaştırılır ve toplamda o periyotta kaç kez ilgili platformun müşterisi olduğu bilgisi hesaplanarak RFM analizinin sıklık(frequency) parametresini oluşturur.

Tutar; müşteri verilerine dair işlemlerdeki toplam harcama tutarını ifade eder. Bir müşteriye ait birden fazla işlem girdisi olması halinde yine aynı müşteri ID ye ait tüm işlemlerin toplam sepet tutarı hesaplanır ve harcama tutarı (Monetary) metriği başlığı altında analize katılır. E-ticaret verisi örneğinde aynı müşteri ID ye ait toplam sepet tutarları dikkate alınır.

Metriklerin hesaplanması ardından normalizasyon işlemine gidilir. Bunun nedeni farklı birimlerin ve sayısal aralıkların işleme katılmasıdır. Analizi tutarlı kılabilmek adına bu gibi veri çeşitliliği olan veri setlerinde modelleme öncesi normalizasyon işlemi yapılarak tablodaki her bir hücreye 0 ve 1 arasında bir değer ataması yapılır.

Detaylı modelleme adımları Python kodlama ile bir sonraki bölümde gerçek müşteri veri verisi üzerinde uygulanmıştır. Bu nedenle teknik detaylara orada yer verilecektir. Bu başlıkta yalnızca analize dair teorik bilgi ve genel uygulama adımlarına yer verilmiştir. 3 ana metriğin veri setinden elde edilen bilgiler ışığında hesaplanması ardından tablo haline getirilir ve böylece her bir müşteri ID ye ait RFM skoru belirlenmiş olur.

RFM skoru belirlenmesi ardından sonraki aşama müşteri segmentasyonudur. Müşteriler segmentlere ayrılarak bu gruplandırmalar üzerine pazarlama stratejileri değerlendirilebilir. Makine öğrenmesi kümeleme algoritmaları ile bu analizi pekiştirmek segmentler hakkında detay çıkarımlarda bulunabilmek mümkündür.

2.1. Müşteri Kullanıcı Deneyimi

Son teknolojik gelişmeler zamanla müşterilere ürün ve hizmeti ulaştıran kanalları çeşitlendirmiş ve bu alternatifler geliştikçe alışveriş deneyiminde beklentiyi üst seviyelere taşımıştır. 2000'li yıllarda dijitalleşme hız kazanırken müşteri kullanıcı deneyimi de ürün geliştirme ve pazarlama süreçlerinde yerini almış böylece bu deneyimin izlenimlerini sahadan edinmek ve sonrasında bu verileri analiz ederek bilgiye dönüştürmek şirketlerin bir iş geliştirme süreci haline gelmiştir. Kimi zaman müşteri kullanıcı deneyimi olarak birlikte değerlendirilse de müşteri ve kullanıcı deneyimi olmak üzere ayrı ayrı ele alındığı yapılan literatür taramalarında gözlenmiştir. Müşteri deneyimi (CX) müşterinin o marka ile etkileşimini ifade eder kullanıcı deneyimi için ise o ürün veya hizmetten yararlandıktan sonraki deneyimi konu

edinir. Müşteri deneyimi bünyesinde kullanıcı deneyimini de kapsar ve daha geniş bir bakış açısı ile durum tespiti yapılmasına imkân sunar. Örneğin ilgili müşterinin o ürün veya hizmetten memnuniyet oranı, tekrar tercih edip etmeme durumu veya başkalarına tavsiye etmesi gibi konular müşteri deneyimi başlığında incelenir. Sahadan alınan metrikler ölçümlenir ve süreç öncesi reklam kampanyaları ile özendirme ve sadakat çalışmaları yapılır, süreç sonrası sunulan ek hizmetler ile de müşteri deneyiminin iyileştirilmesi amaçlanır.

Kullanıcı deneyimi (UX) genelde dijital platformlar için özellikle e-ticaret sektöründe sunulan yazılım çözümlerinin müşteri ile etkileşimi sonucu deneyimlemesi ile başlar. O platformun kullanıcısı ilgili yazılım çözümünde kullandığı tüm senaryolar için fikir sahibi olur ve iyileşme noktaları belirlenir. Anketlerle veya pilot müşteriler ile yapılan kullanıcı deneyimi testleri ile sürece dair sahadan geri bildirimler alınır bunlar yazılım çözümü için iş analistleri tarafından analiz edilir ve iş parçacıkları haline getirilir. Ardından ürün geliştirme, tasarım ve yazılım ekipleri ile paylaşılır. Geliştirmenin ardından son güncellemeler tekrar test sürecine dahil olur testlerden geçmesi halinde yazılım güncellemesi olarak ürüne yansıtılır ve müşterilerinin kullanımına sunulur. E-ticaret platformları geliştirmelerini yazılım yaşam döngüsü çerçevesinde gerçekleştirir ve kullanıcı deneyimini iyileştirmek müşteri memnuniyetinin en önemli parametresidir. Web siteleri ve mobil uygulamaların arayüzleri tasarım ekiplerince en kullanışlı hale getirilmeye çalışılır yazılım ekipleri ile koordine çalışılır. Müşteri kullanıcı deneyimi çok disiplinli bir çalışma konusudur. Kullanılan renk tonlarının müşteri algısında yapacağı çağrışımlardan arayüzdeki butonların doğru servislere yönlendirmesi veya burada geçen sürenin optimizasyonu gibi çok çeşitli kapsamaları bulunmaktadır. Müşteri memnuniyetini en üste taşımak adına dijital platformlar birbiriyle yarışırken hız faktörünü de gözeterek platform içerisinde arama motoru ile ürün araştırması yapabilmesi favorilere eklemesi veya sepete atıp satın almayı gerçekleştirilmesi artık e-ticaret sitelerinde

saniyeler almaktadır. Kullanıcı ara yüzü tasarımı, renk seçimi ve işlevselliği ile ilgili platform daha çok tercih sebebi olabilir. Daha fazla müşteri popülasyonu kazanmak için şirketler müşteri kullanıcı deneyimini iyileştirmek adına sahadan geri bildirim alır ve bu bilgiler ışığında iyileşme noktalarını belirler.

2.2. Kullanılan Veri Setinin Öznitelik Çıkarımı

Sahadan toplanan veri seti yorumlanabilmek için tek başına genellikle yeterli değildir. Bu verinin bilgiye dönüşmesi sürecinde analize başlamadan önceki ilk ve en önemli adım veri setinin hazırlanması işlemidir. Böylece modellemeye hazır bir veri seti analiz edildiğinde ve makine öğrenmesi adımları uygulandığında daha tutarlı çıktılar verir. Veri seti hazırlanırken ilgili problem için geçerli olan kolonlar dikkate alınır, işe yarar parametreler belirlenir böylece öznitelik çıkarımı işlemi uygulanır. Bunun amacı büyük ve kirli veriden anlamlı ve analiz sırasında ihtiyacımız olan nitelikli verilerin çıkarımı ve analize dahil edilmesidir. Veri seti içerisinde kolonların sayısal veri aralığı arasında çok fazla fark olması halinde standart sapması kontrol edilir. Standart sapma çok yüksek ise sapmayı etkileyebilecek uç değerler belirlenir. Bunların hangi müşteri ID sine ait olduğuna bakılır. Eğer standardın dışında seyreden bir veri olur ise analiz çıktılarını yanıltıcı etkilememesi adına o müşteriye ait veriler dışarıda tutulabilir. Bu gibi çıkarımlar için problemi veya analizin yapılacağı çalışma konusunu doğru belirlemek nitelikli verilerin çıkarımında oldukça önem taşır. RFM analizi uygulanacak veri setinden analize dahil olacak kolonlar güncellik, sıklık ve harcama tutarı olmalıdır. Bu kolonlar da kendileri içerisinde taranır ve sapması yüksek olan bir veri görülürse satır olarak o müşteri analiz dışında tutulabilir.

Ham veriden edinebilecek öznitelik çıkarımları her zaman kolonlar bazında olmaz. Bazen birden fazla parametreden tek bir parametre de elde edilebilir bu da öznitelik çıkarımı işleminin bir parçasıdır. Örneğin

bu tez çalışmasında bir e-ticaret şirketine ait müşteri verileri kullanılmıştır ve bu veri setinde işlem verileri periyodik olarak tutulmuştur, haliyle aynı müşteriye ait birden çok işlem satırı olabilmektedir. Bu durumda belirleyici kolon müşteri ID olarak seçilir ve tıpkı veri tabanlarındaki anahtar değer gibi ilgili ID ye ait tüm işlemler gruplanır. Gruplamanın ardından artık her bir satır o müşteriye ait farklı zamanlardaki alışverişleri de listeler. Her bir satır hangi ürünü, ilgili ürün tutarını ve kaç adet alındığını içerir. Ürün fiyatı ve sepetteki alınan adedi çarpımı bize toplam sepet tutarını vermektedir. Böylece öznitelik çıkarımı yapılarak sepet tutarına gidilebilmektedir. Her müşteriye ait ne kadar işlem var ise toplam harcama tutarı alınarak RFM analizine dahil edilmiştir. Yine aynı şekilde güncellik bilgisi hesaplanırken tarih formatında tutulan işlem tarihi GG\AA\YY şeklinde tek bir format haline getirilmiştir. Ardından bir referans tarih dikkate alınmış ve işlemlerin bu tarihe uzaklıkları gün bazında hesaplanarak öz nitelikli veri halini almıştır. Frekans verisi de ham veri içerisinde yoktu ve bu veri de öznitelik çıkarımı ile hesaplanmıştır. Veri seti Müşteri ID si bazında gruplandırılarak her müşteriye ait alışveriş işlemleri listelenir ve toplamda kaç kez ilgili e-ticaret platformundan alışveriş işlemi gerçekleştirdiği bilgisine bu sayede ulaşılmıştır. Elde edilen sayılan veri sıklık parametresi olarak RFM analizine dahil edilmiş ve böylece analizin 3 temel parametresi öznitelik çıkarımı ile ham veri dönüştürülerek elde edilmiş sonrasında uygulanacak makine öğrenmesi algoritmaları için de hazır hale getirilmiştir.

2.3. Müşteri Segmentasyonu

Alışveriş sektöründeki hızlı dönüşümler artan rekabet ortamı ile birlikte şirketleri müşteri ilişkilerini belirlerken aksiyon almaya zorlamıştır. Önceleri alışveriş kanalları geleneksel yöntemler ile ilerliyordu. Fiziki olarak mağazada bulunulur ve birebir etkileşime geçilirdi. Sunumlar, ürün veya hizmet kalitesi ve makul bir satış

pazarlama süreci iyi bir izlenim ve olumlu müşteri deneyimi için yeterliydi. Buradan edinilen olumlu intiba aynı müşterinin çevresine de vereceği tavsiyelerle potansiyel farklı müşterileri de bu ürün veya hizmetten yararlanmaya davet etmekteydi. Günümüzde ise hız kazanan teknolojik gelişmeler, sosyal medya faktörü ve tüm dünyayı etkisi altına alan pandemi süreci ile birlikte ürünü müşteriye ulaştıran kanallar da çeşitlenmiş ve alternatifler rekabet ortamını kaçınılmaz kılmıştır. Bu rekabet ortamını avantaja çeviren şirketler ürün yelpazesinde yaptığı çeşitlilikle ve müşteriye ulaştırma biçimleri ile öne çıkmış bu sayede satış hedeflerini katlayabilme imkânı bulmuş ve müşteri hacmini giderek artırdığı izlenmiştir. Marka zincirlerinden yola çıkacak olursak sektöre tutunabilmeleri, sürekli üretim otomasyonu ve arz talep ilişkisini karşılayabilmeleri yalnızca sunum ve kalite parametrelerine bağlanamaz. Burada müşteriye yaklaşımları da oldukça etkili bir parametre haline gelmiş ve yıllık satış pazarlama hedeflerine ulaşmada etkin rol oynamıştır.

Ürün veya hizmeti henüz üretime geçmeden fikir aşamasında dahi fizibilite çalışmaları yapılarak pazar hakkında öngörülerde bulunulur. Bu çalışmalar sırasında artık müşteriye tek bir profil olarak değerlendiren geleneksel yaklaşımdan vazgeçilmiş, müşteri segmentasyonuna gidilmiştir. Bu yöntem potansiyel müşteri kitlesini segmentlere ayırarak benzer satın alma davranışı takip eden müşteri profillerini guruplarken bir araya getirip her segmentin ihtiyaçlarını doğru belirleyerek onlara özgü ayrı pazarlama stratejileri geliştirilmesi çalışmalarını kapsar. Müşteriye tek tip kabul etmek ve herkese aynı yaklaşmak artık günümüzde etkili değildir o nedenle ürün veya hizmeti piyasaya sunarken kitlenin demografik yapısına (cinsiyet, yaş vs.), satın alma davranışlarına, popülasyonun coğrafi isteklerine hatta psikolojik yansımalarına dikkat edilir. Müşteri segmentasyonu da tıpkı veri analitiği gibi çok disiplinli bir uygulama alanıdır ve ekonomist, psikolog, toplum bilimci gibi birçok farklı branştan uzmanın görüşleri alınarak ortak paydada değerlendirilir. İstatistiksel yöntemler ve yazılım

alanında uygulanan yapay zekâ çözümleri ile müşteri analitiği çerçevesinde analizlere gidilir. Özellikle müşteri segmentasyonunda RFM analizi en yaygın kullanılan çözümlerden biridir. İstatistiksel analizle de pazar payı ve müşteri kitlesi hakkında çıkarımlarda bulunulur. Bu analizler çerçevesinde müşterilerin verilerine bakarak onları profillere ayırmak mümkündür. Pazar hacminin psikolojik yansımaları da satın alma davranışında belirleyici bir faktör olmaktadır. Örneğin pandemi sürecinin hemen öncesi yapılan alışverişler hazır giyim, makyaj ürünü veya eğlence sektörü gibi Maslow hiyerarşisinde olduğu gibi kendini gerçekleştirme kademesindeki kalemlerinde seyrederken bir anda pandemi başlangıcı ile birlikte temel ihtiyaçlara dönmüş ve yerini maske, dezenfektan, kolonya gibi hijyen ürünlerine ve temel gıda ürünlerine bırakmıştır.

2.4. RFM Analizinin Sektörde Kullanım Alanları

RFM analizi uygulanabilirliği ve müşterileri segmentlere ayırma performansı gözetilerek hala yaygın olarak kullanılır. Pazara sunulması planlanan ürünler müşteri profillerine göre çeşitlilik gösterir ve bu trendler gözetilerek satış pazarlama süreçleri tasarlanır. Bir diğer örnek olarak birkaç ay önce ülkemizin yaşadığı deprem felaketi ve yansımaları değerlendirilebilir. Çok sayıda vatandaşımızın hayatını kaybettiği üzücü olayın psikolojik ve ekonomik yansımaları hemen her sektörde derinden hissedilmiştir. Olayın ilk anından itibaren temel gıda, temizlik, ısınma ve barınma ürünlerine olan talep üst seviyeye çıkmış ve dijital alışveriş kanalları çevik bir yaklaşım göstererek birçoğu ulaştırma bedeli almaksızın deprem bölgesine transferler gerçekleştirmiş ürünleri ulaştırmıştır. Bu olayın derin etkileri ve psikolojik yansımaları pazarlama davranışlarına da yansımış müşterilerin ürün tercihlerinde normal rutininin dışına çıkarak temel ihtiyaçlara, yardım için oluşturulan paketlere, deprem ve ilk yardım çantası gibi önlem ürünlerine yöneldiği izlenmiştir. Bu gibi derin etkiler perakende sektöründe satış ve pazarlama stratejilerini yeniden

şekillendirir. İstatistiksel analizler neticesinde segmentlere ayrılan müşteri profillerinin sergilediği harcama davranışları da çevresel faktörlerden birinci derece etkilenir. Bu gibi dinamikleri yakından takip eden ve müşterilerini segmente ederek farklı stratejiler geliştiren şirketler daima pazar hacmini koruyarak daha geniş kitlelere hitap edebilmiştir.

Segmentlere ayrılan müşteri profillerine özgü reklam kampanyaları, ürün sunumları ve özendirme çalışmaları satın alma eğilimindeki her potansiyel müşteri nezdinde de tercih sebebidir. Bazen farkında olmasak da bize özgü sunumlar algıda seçicilik yaratır ve seçimlerdeki yönelimimizi belirler. Örneğin film ve dizi arşivi olarak kullanıcılarının beğenisine sunulan Netflix, kullanıcı verilerini işleyip analiz ederek akışta yer alan film ve dizilerin kapak fotoğraflarını dahi kullanıcılarının beğenisine göre şekillendirmiş ve kişiye özgü bir akış yakalamıştır. Öneri sistemi de yine sizin önceki izlemelerinize göre şekillenir ve benzer içerikler için tavsiyede bulunur. Dijital alışveriş platformlarında da ürün öneri sisteminin bir benzerini görürüz. Favorilerinizdeki ürünler ve ziyaret ettiğiniz ürün sayfaları incelenerek benzer satın alma davranışı seyreden müşteri verileri ile kombinlerin ve ürün kategorileri öneri olarak sunulur. Böylece sepete eklediğiniz bir ürün ardından benzer kullanıcılar bu ürünleri de aldı veya o ürünün yanı sıra bu ürünler de çok satılıyor gibi anonslarla öneri ürün kategorileri de beğeninize sunulur. Müşterileri farklı segmentlerde ele almak beklentileri ve ihtiyaçları da spesifik olarak belirlediğinden müşterilerin beğenilerine daha çok hitap eden ve tam olarak beklenen ürün/hizmeti almalarını sağlar bu da müşteri memnuniyeti ve müşterilerin beğenilerini çevresine tavsiye etmesi ile potansiyel yeni müşteri anlamına gelmektedir. Rekabet ortamında alternatiflerin tek tıkla listelendiği günümüzde sadece fiyat avantajı ve kalite yeterli gelmez, müşteri segmentasyonu ile kişiye özgü çözümlerle beklenti karşılanır. Daha çok kısa mesaj veya mail ile bu kampanyalardan haberdar oluruz markalar bizlere kendini hatırlatır ve yeni ürün

yelpazesinden haberdar eder. Bazen de sosyal medyada takip ettiğimiz bir ünlüden veya televizyon programındaki bir ürün yerleştirmesinde o ürüne rastlar böylece o ürünün tercih sebebi olduğu algısına kapılır ‘influence’ oluruz ve sahip olmak isteriz. Sosyal medyada içerik üreticileri kendini yeni bir akımla ‘influencer’ olarak niteler ve paylaşımlarındaki ürün tanıtımları ile kullanıcılarında merak uyandırarak satın almaya teşvik eder.

Dijitalleşen dünyada farklılaşan bu gibi pazarlama ve alışveriş deneyimleri ile amaçlanan kitleye olabildiğince ulaşabilmek ve müşterilerce tercih sebebi olmak için RFM analizi ile müşteri segmentasyonuna başvurulur ve analiz sonrası edinilen bilgiler ışığında ürün yelpazesi, kampanya çalışmaları ve satış hedefleri belirlenerek şirketlerin ileriye dönük planlarında tutarlı çıkarımlarda bulunmak amaçlanır. Bu analiz ve segmentasyon yöntemi müşteri sirkülasyonu yüksek ve sektörde rekabet oranı yüksek sektörlerde daha yaygın tercih edilir. Bunların başında bankacılık sektörü ve perakende satış pazarlama sektörü gelir. Müşteri popülasyonun yoğun olduğu ve sirkülasyon trafiğinin de bir o kadar hızlı geçişler içerdiği bu iki değerli sektör müşteriye en doğru hizmet veya ürünü ulaştırmak, mevcut kitlesini korumak ve potansiyel müşteri trafiğini de kendine çekmek için stratejiler geliştirir. Pazarlama stratejileri segmente edilen müşteri kitlelerine göre farklılık gösterir ve bu sayede sahada daha etkili olur. Bankacılık sektöründen yola çıkacak olur isek bir müşterinin genelde birden fazla banka hesabı bulunur ve bankalar ilgili müşterinin en çok işlem hacmine sahip olabilmek için birbiri ile yarış halindedir. Müşterilerin işlem verileri genelde maskelenerek ve kişisel verileri koruma kanunu çerçevesinde veri saklama ve işleyebilme izin beyanına tabii olarak analiz aşamalarına dahil edilir ve ham verinin bilgiye dönüşmesi sağlanır. RFM analizi ve sonrasında RFM skoru neticesinde değerlendirilen müşteri segmentleri için özel olarak kampanya ve reklam hizmetleri üzerine çalışılır. Bu nedenle bankacılık işlemleri sonrası örneğin harcamalarınız kart limitine yakın veya limitin

üstünde seyrediyorsa daha üst limitlere yönlendirilirsiniz. Bu da harcama hacminizi genişletir ve bankaya daha fazla borçlanmanız demektir. Ödeme yöntemleri olarak da asgari de seyrediyor veya altın kalıyorsanız size özel ödeme çözümleri sunulur. Kredi kartı ile harcama davranışı sergileyen müşterilerin hesap hareketlerine bakılırsa harcamalarının kategorize edildiği görülür. Böylece ilgili öne çıkan harcama kanallarına özel müşteri spesifik kampanyalar sunulur. Kredi aidatı ve hesap kesim ücreti almama puan tanımlama veya ek taksitlendirme yaparak müşteri sadakati üzerine çalışılır bu da işlem hacmini tek kanalda toplama imkânı sunar. Yeni nesil bankacılık uygulamaları özellikle kart tasarımı ve mobil bankacılık uygulamalarında sesli asistan gibi yapay zekâ çözümlerine giderek rekabet ortamında öne çıkma amacı güder. RFM skoru yüksel olan haliyle gelir düzeyi yüksek müşteriler için kişiye özel bankacılık hizmetleri sunulur. RFM skoru daha aşağıda kalan segmentler için de özendirme kampanyalarına gidilir. RFM skoru 3 ayrı parametre ışığında değerlendirildiğinden güncellik veya işlem frekansı düşük seviyede seyreden müşteriler için kayıp edilebilecek müşteri gözü ile bakılır ve yine müşteri spesifik çalışmalar ile geri kazanım amaçlanır. Tutar parametresi yüksek seyreden müşteriler kredibilitesi yüksek kabul edilerek özel faiz oranları sunulur. Tüm bu pazarlama stratejileri RFM analizi ile müşteri odaklı bir segmentasyon ile mümkün kılınmış ve sahada veri analizi aşaması olarak etkin kullanılmaktadır.

3. PYTHON PROGRAMLAMA İLE RFM ANALİZİ ÖRNEĞİ

Bu tez çalışmasında konunun belirlenmesi ardından veri seti arayışına girilmiş danışman hocamın da yönlendirmeleri doğrultusunda birkaç kurum ile iş birliği için görüşülmüştür. Bankalar dışarı veri aktarma konusunda oldukça temkinli olduklarından ve Bankacılık Düzenleme ve Denetleme Kurulu kurallarına tabii olduklarından özel nitelikli verileri hassasiyetle korumak eğilimindeler ve akademik çalışma çerçevesinde

dahi olsa veri paylaşımına mesafeli oldukları görülmüştür. Kaggle gibi platformlardan hazır veri alınarak yapılan modellemelerin çıktıları ve analizi kısıtlı olabileceğinden sentetik veri edinmek yerine sahadan gerçek veri seti bulmayı amaçladık ve görüşmelerimiz neticesinde perakende giyim sektöründe geniş müşteri kitlesine ve pazar payına sahip LC Waikiki şirketi ile iş birliğine gidilmiştir. İlgili kurumdan akademik çalışma çerçevesinde kullanmak üzere müşteri alışveriş işlemlerine ait veri seti alınmıştır. Veri seti tamamen gerçek müşteri verilerine ait ve maskelenmiş alışveriş işlemi metriklerini tutar. Veri seti birçok kaynaktan beslendiğinden ilk hali analize dahil olmayacak birçok parametre içermekte ve aynı kolonda dahi farklı formatta veri türlerinin tutulduğu karmaşık yapıda idi bu nedenle analiz işlemi öncesinde veri temizliği ve nitel-nicel veri türlerinin Python kodunda atanması, standart sapması yüksek aykırı verilerin ayıklanması veri türlerinin standart hale getirilmesi gibi veri setinin hazırlanması işlemleri uygulanmış veri seti modellemeye uygun hale getirilmiş sonrasında analiz adımları takip edilmiştir.

Veri setinin hazırlık aşaması çoğu projede modelleme veya makine öğrenmesi algoritmalarının uygulanması aşamasına nazaran daha fazla zaman almaktadır. Verinin anlamlı hale getirilmesi ve ham halinden arındırılıp modellemeye en uygun standart formunu kazanması için tüm veri setleri öncelikle hazırlık aşamasından geçer. Analizi yapılacak olan veri tutarlı çıktılar üretebilmesi için literatürde kirli veri olarak da geçen standart sapması yüksek mevcut veri setini çoğu zaman yansıtmayan verilerden arındırılır. Bu gibi bozuk verilere genelde yanlış kayıt alınması sonucu rastlanır. Kayda geçerken yapılan bir noktalama hatası veya veri türü dönüşümleri sırasındaki yuvarlamalar kimi zaman ilgili kolondaki diğer verileri tutmaz ve standardın dışında kalarak analiz sırasında sapmaya yol açar. Bu yanıltıcı verilerin mevcut veri setinden uzaklaştırılması yani komple silinmesi veya veri

setinin boyutu sınırlı ise silmek yerine sapma yaratmayacak şekilde ortalama tutacak yeni değerlerin atanması gibi aksiyonlar veri seti hazırlığı aşamasında temizlik işlemi kapsamında gerçekleşir. Bunların yanı sıra birden fazla kolonun birleştirilmesi ile de yeni bir metrik elde edilebilir. Örneğin tez çalışmasında kullanılan veri setinin ilk halinde sepetteki ürünlerin adet ve fiyat bilgisi mevcuttu. Bunu RFM analizi için daha uygun hale getirebilmek adına bu iki değer çarpılarak sepetteki tüm ürünlerin toplam harcama tutarı hesaplanmış ve ayrı bir kolonda tutulmuştur. Bu gibi silme, dönüştürme veya birleştirme işlemleri verinin hazırlık aşamasında analistler tarafından takip edilir.

3.1. Kod Tabanında Kullanılan Python Kütüphaneleri

Tez çalışmasında gerçekleştirilen RFM analizi Python yazılım dili ile Jupyter Notebook uygulaması üzerinde çalışılmış ve çok sayıda Python kütüphanesinden yararlanılmıştır. Bunlar hakkında bu bölümde detaylı bilgi verilecek ve kullanım alanları detaylandırılacaktır.

```
import warnings
warnings.filterwarnings("ignore")

import pandas as pd
import datetime as dt
import dateutil.parser
import numpy as np
import sklearn

from matplotlib import pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
```

Pandas kütüphanesi; projenin henüz başında veri setinin koda dahil edilmesi verinin dönüştürülmesi ve ham verinin temizlik işlemlerinin yapılması gibi hazırlık aşanalarında kullanılır. Büyük çapta veri setlerinde genelde daha az yer kaplaması için “csv” formatında tutulur. Csv format virgülle ayrılan değerleri içeren metin dosyalarıdır. Tercihe bağlı olarak veri setleri txt veya excell formatında da tutulabilir. Pandas kütüphanesi tüm bu formatları destekler.

Pandas kütüphanesi aracılığı ile veri seti projeye dataframe formatta dahil edilmiştir. Dataframeler ile 2 boyutlu matrislerdir ve tıpkı veri yönetiminde sık tercih edilen SQL tabloları gibi tüm veriyi [SatırxSütun] matrisine dönüştürür böylece dönüşüm işlemlerinde görsel olarak veriyi okumak kolaylaşır.

Datetime kütüphanesi; veri seti içerisinde zaman, saat, tarih bilgisi içeren kolonlarda formatlama yapmaya yarar. Tarih formatlarının çok çeşitli gösterimleri vardır ve kullandığımız veri setinin ham halinde müşteriye ait işlem bilgisi tutulurken farklı formatta zaman dilimi formatlarında kayıt altına alındığı gözlenmemiş formatlama aşamasında bu kütüphaneden yararlanılmıştır.

Zaman serisi analizinde belirli zaman aralığındaki işlem verisi analiz edilerek ileriye dönük tahminlerde bulunulur. Bu çalışmalarda da yine Datetime kütüphanesi gibi tarih formatlaya yardımcı kütüphanelere ihtiyaç duyulur.

Dateutil.parser kütüphanesi; tıpkı Datetime kütüphanesi gibi yine tarih bilgisi içeren verilerin formatlanması aşamasında gün, ay, yıl, saat, dakika, saniye, salise kimi zaman çok daha hassas değişim gösteren zaman serisi verilerinde ise mili saniye ve milisalise kırılımında ilgili zaman verisini alt kırılımlara ayırma işlemi için kullanılır.

Tez çalışmasında kullanılan müşteri veri seti içerisinde her bir müşteriye ait işlem verisi işlem tarihi başlığı altında kayıt altına alınmıştır. İlgili kolonda sıralı tarih formatları bu kütüphane aracılığı ile gün-ay-yıl formatında ayrıştırılarak aynı kolonda tek bir standart tarih formatı halini almıştır.

Numpy kütüphanesi; Python ile çalışırken çok boyutlu diziler ile çalışmamıza olanak sağlar. Veriyi matris formatına getirir ve bu sayede toplama, çıkarma, gruptama gibi birçok matematiksel

operasyon veri seti üzerinde uygulanabilmektedir. Kullanılan müşteri verisinin modellemeye hazırlanması aşamasında gruplama işlemine başvurulmuştur. Groupby yöntemi uygulanırken numpy kütüphanesine ait gruplama ve toplama operasyonları kullanılmıştır.

Sklearn kütüphanesi; bir diğer isimlendirme ile **scikit-learn** olarak da karşımıza çıkar ve denetimli ve denetimsiz tüm makine öğrenmesi algoritmalarını bünyesinde barındıran çok zengin bir kütüphanedir. Hazırlık aşaması tamamlanan ve modellemeye geçilen veri setlerinde ilgili veri setine uygun makine öğrenmesi algoritması belirlenerek bu kütüphane aracılığı ile proje koduna dahil edilir. Tez çalışması RFM analizi ile müşteri segmentasyonu odaklı olduğundan performans karşılaştırması yapabilmek ve çıktıları karşılaştırmalı olarak yorumlayabilmek adına Sklearn kütüphanesi ile makine öğrenmesi algoritmalarından Kmeans kümeleme algoritması koda entegre edilerek tüm adımları uygulanmış ve oluşan kümeler sonrasında rf m skorları ile karşılaştırılarak analiz edilmiştir.

Matplotlib kütüphanesi; veri görselleştirme için kullanılan en yaygın Python kütüphanesidir. Verinin analizi aşamasında sadece csv formatındaki ham veri yorumlaması zor ve analiz için tek başına yeterli değildir. Veriyi daha anlaşılır kılmak ve yorumlamaya açık hale getirmek adına Matplotlib kütüphanesinin zengin grafik içeriklerinden yararlanılır. Tez çalışması sırasında da keşifsel veri analizi ve modelleme sonrası grafikler oluşturulmuş ve bunlar detaylandırılarak çıkarımlarda bulunulmuştur. Görselleşen veri daha anlaşılır bir hal aldığından veri analistleri tarafından yazılım dilleri veya iş zekâsı veri görselleştirme bilgisayar programları kullanılarak veri hakkında daha fazla öngörü elde edilmesi amaçlanır. Daha önceleri yalnızca 2 boyutlu grafikler kullanılırken artık çok sayıda veri görselleştirme Python kütüphanesi geliştirilmiş bu sayede çok boyutlu ve dinamik grafikler

oluřturma imkânı doęmuřtur. Verinin grselleřtirilmesi ařamasında veri setine ve probleme hâkim olmak nemlidir. ok sayıda farklı grafik ierisinden(ubuk, izgi, pasta grafikler vs.) nicel ve nitel metriklerin kombinasyonu ile elde edilen ve grselleřtirmek istedięimiz veri iin en uygun grafik tercih edilir. Bu noktada her grafik her veri seti iin uygun deęildir. Uygun stile karar verilmesi ancak veriyi ve problemi iyi bilen veri analistleri, iř zekası uzmanları ve veri grselleřtirme uzmanları tarafından mmkndr. Gnmzde veri grselleřtirme alıřmaları olduka revata olup hemen her iř kolunda veriyi hem anlamak hem de dięer iř birimlerine anlatabilmek iin kullanılmaktadır. Python dili de bu operasyonların gerekleēebilmesi iin olduka kullanıřlı olduęunda bnyesinde birok grselleřtirme ktphanesi barındırır ve bu ktphaneler yeni zellikler ile geliřmeye devam eder.

Grselleřtirme ktphanelerinden ne ıkan bir dięeri seaborn Python ktphanesidir. Grsel olarak daha ok alternatif ierir ve ok boyut grafiklerin izdirilmesinde kullanılır. Ktphanenin kendi bnyesinde hazır veri setleri de bulunur. Bylece rnek grafikleri python ile jupyter notebook zerinde ařaęıdaki gibi birkaç rnek kod satırı ile izdirmek mmkndr.

In [113]:

```
#example chart

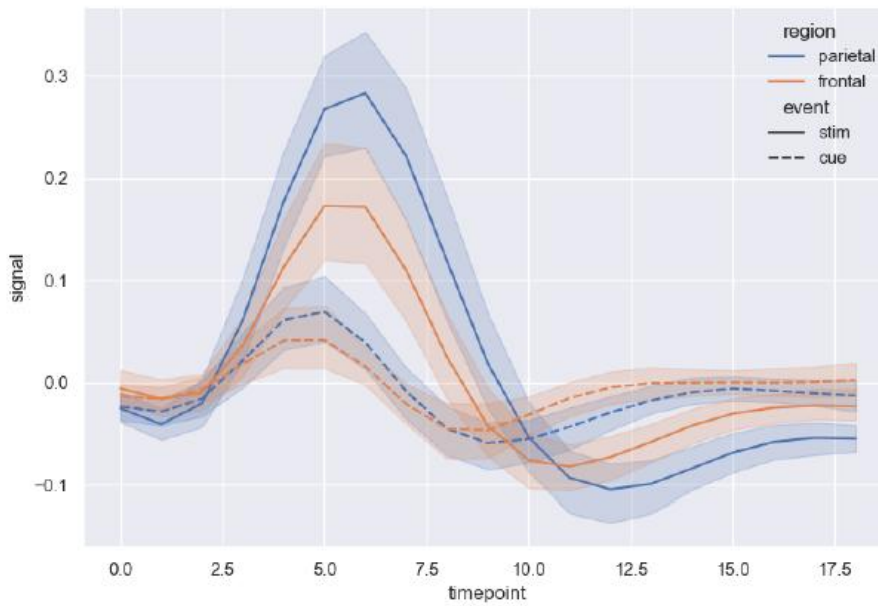
import seaborn as sns
sns.set_theme(style="darkgrid")

# Load an example dataset with long-form data
fmri = sns.load_dataset("fmri")

# Plot the responses for different events and regions
sns.lineplot(x="timepoint", y="signal",
             hue="region", style="event",
             data=fmri)
```

Out[113]:

<AxesSubplot: xlabel='timepoint', ylabel='signal'>



3.2 Kullanılan Veri Seti Özellikleri

Gerçek müşteri verilerine ait alışveriş metriklerini tutan csv formatında edinilen veri seti Jupyter Notebook bilgisayar programına dahil edilerek isimlendirme standardında değişikliğe gidilmiş kolon isimleri Türkçeleştirilmiş gereksiz kolonlar işleme alınmamış ve aynı kolonda veri tipleri tek standart haline getirilmiştir. Verinin hazırlığı ve analizi safhasında referans olarak iş birimleri ile görüşülmüştür sahadan alınan bilgiler ışığında analizler gerçekleştirilmiştir.

LCW zincir perakende mağazalarına ait maskelenmiş müşteri alışveriş işlemlerini barındıran veri seti toplamda 8 kolondan oluşmaktadır.

Veri setinin işlem görmemiş ve veri tabanından ilk çekilmiş ham hali aşağıda göselde yer aldığı gibidir. Bu veri seti Pandas kütüphanesine ait dataframe formatında projeye dahil edilmiştir. Bunun amacı veriyi 2 boyutlu dizi yani matris formatına sokmak bu şekilde veriye SQL tablosu formatında görsellik kazandırmak ve veri hazırlık işlemlerinde satır/sütun takibini kolaylaştırmaktır. Veriyi içeren matris dataframe dönüşümü sonrası 8 farklı kolon 4876 adet satırdan oluşmaktadır. 2016-2018 yıllarına ait müşteri kayıtları tutulmaktadır.

In [97]:

```
data = pd.read_csv (r'C:\Users\████████\Desktop\Tez\rfm_data.csv',delimiter=';')
data.head(10)
```

Out[97]:

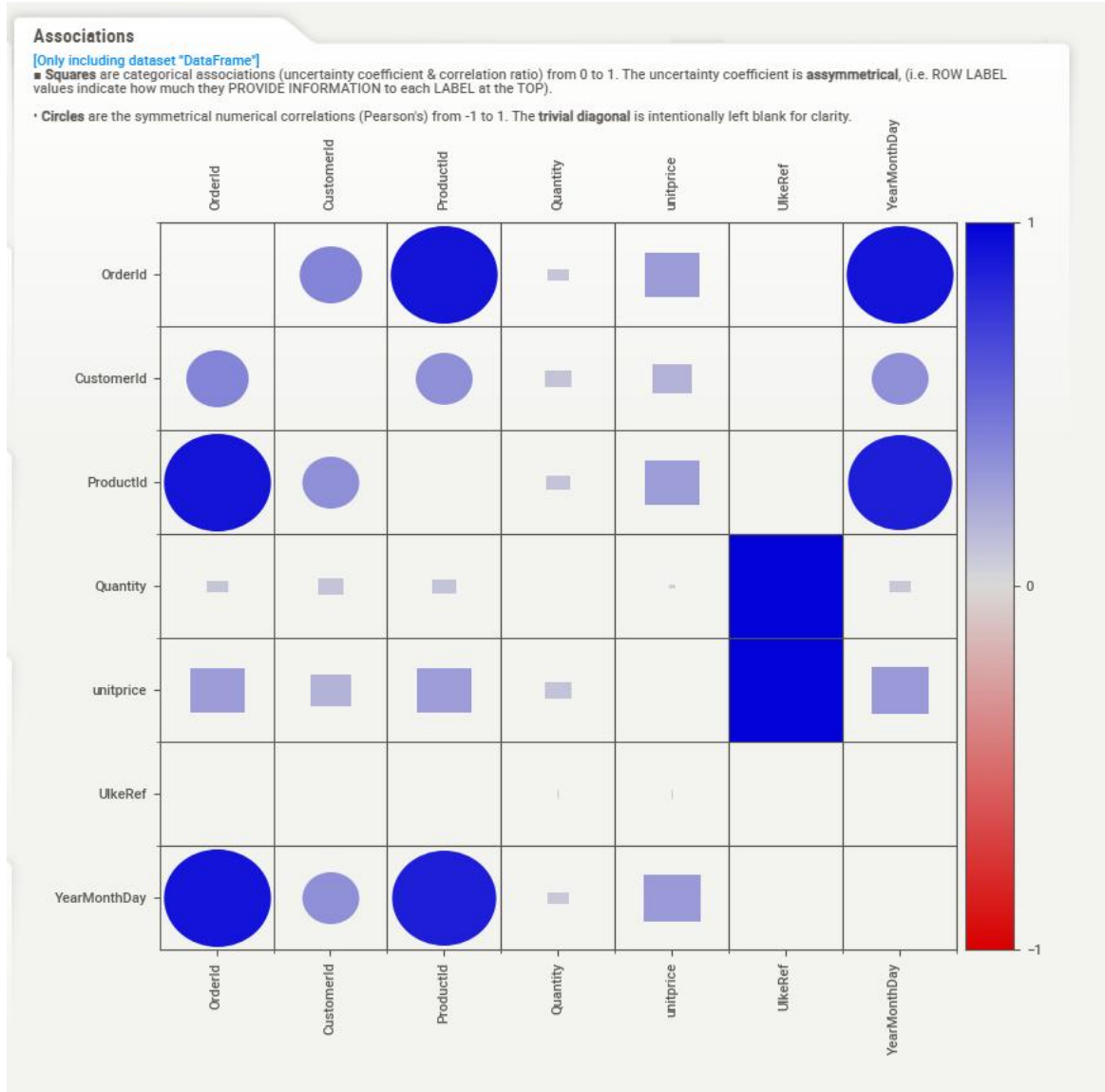
	Orderid	Customerid	OrderDate	Productid	Quantity	unitprice	UlikeRef	YearMonthDay
0	4154205	2126662	4.09.2016	1669027	1	12,95	48	20160904
1	4154205	2126662	4.09.2016	1669794	1	24,95	48	20160904
2	4154205	2126662	4.09.2016	1737107	1	32,95	48	20160904
3	4154205	2126662	4.09.2016	1746673	1	12,95	48	20160904
4	4154205	2126662	4.09.2016	1751216	1	29,95	48	20160904
5	4846981	2224954	13.12.2016	1747611	1	12,95	48	20161213
6	4846981	2224954	13.12.2016	1761517	1	11,95	48	20161213
7	4846981	2224954	13.12.2016	1761604	1	12,95	48	20161213
8	4846981	2224954	13.12.2016	1787959	1	8,95	48	20161213
9	5782088	2846339	25.03.2017	1824747	1	19,95	48	20170325

Veri seti içerisinde sıralı tarih formatlarının farklılık gösterdiği gözlemlenmiştir. Bir müşteriye ait işlem tarihi örneğin bir operasyon için %Y-%m-%d %H:%M:%S.%f formatında tutulurken bir diğer işlem için %Y.%m.%d formatında kayıtlara girmiştir. İlk iş bunları datetime ve dateutil.parser kütüphaneleri kullanılarak standart bir formatta biraraya getirmek olmuştur. Bu kütüphanelerin işlevleri ve kullanım alanları önceki bölümde detaylandırılmıştır. Formatlama aşaması sonrası ilgili işlem tarihi kolonu %d.%m.%Y formatında yani sırayla gün, ay ve yıl verilerinden oluşan forma getirilmiştir. RFM analizinin Recency-Güncellik metriği için gerekli değer bir referans tarih seçilerek o tarihe gün bazında uzaklığı hesaplanarak integer formatında RFM

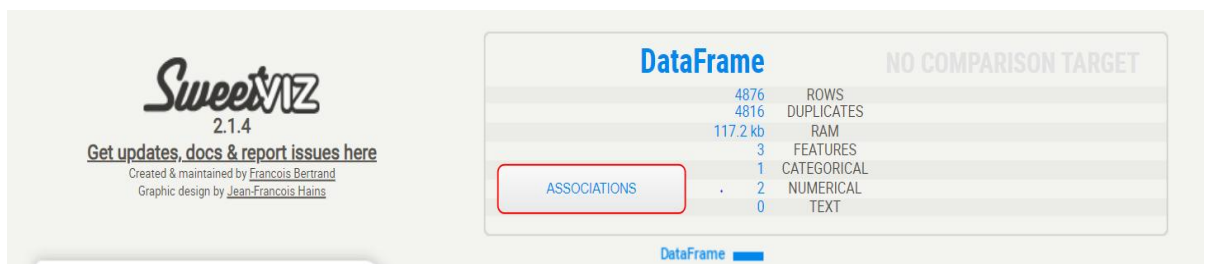
analizine dahil edilmiştir. Bu da verinin hazırlık aşamasında dönüştürülmesi işlemlerinden birine örnek olarak gösterilebilir.

3.3 Keşifsel Veri Analizi

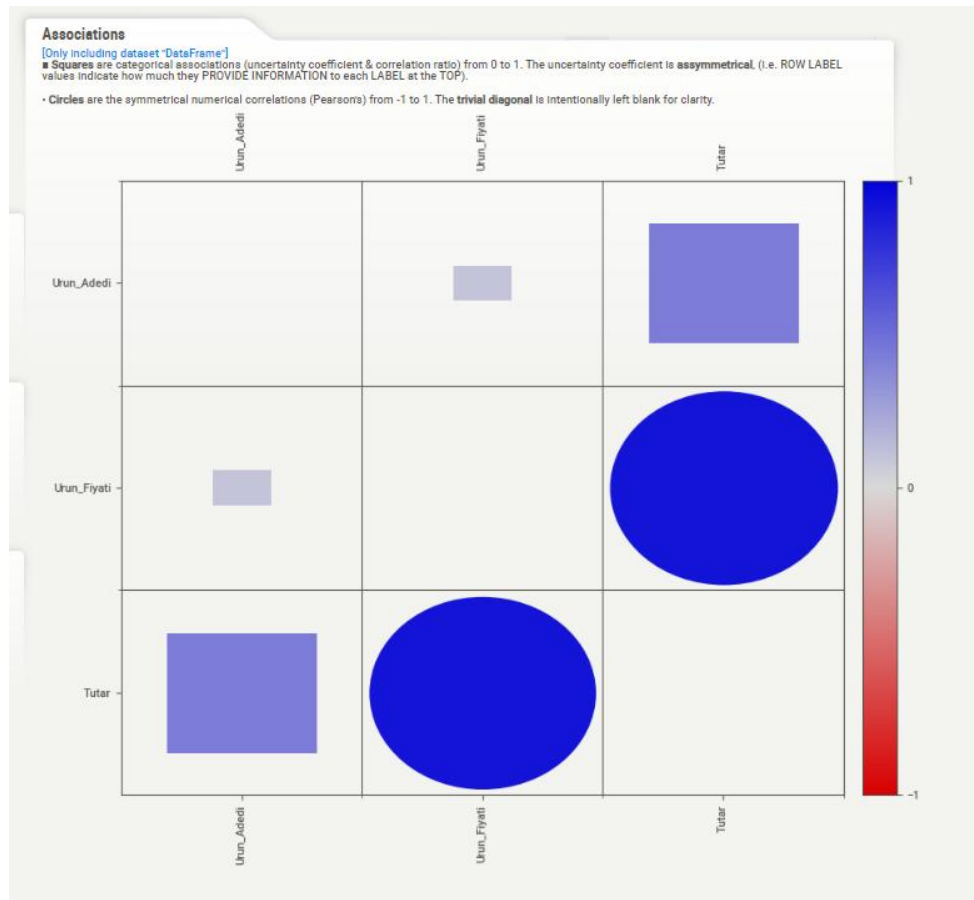
Veri analitiği üzerine çalışmaların hız kazanması ve analiz aşamasında istatistiksel ve yeni nesil yazılım çözümlerinin kullanılması ile birlikte veri analizi artık farklı bir boyuta taşınmıştır. Önceleri kısıtlı yöntemlerle verilerden çıkarım yapmak mümkün iken şimdilerde keşifsel veri analizinin de olgunluk kazanması ile birlikte çok daha etkili keşifsel analizler yapılabilmekte ve ileri dönük tutarlı tahminler yapılabilmektedir. EDA (Exploratory Data Analysis) olarak literatürde yer alan bu çalışmalar istatistiksel yöntemler ile veriden çıkarımlar yapılması ile ilgilenir. Python yazılım kütüphaneleri içerisinde de keşifsel veri analizi yapan ve istatistiksel çıktıları raporlayan kütüphaneler bulunur. Bu tez çalışmasında SweetViz Python kütüphanesi 2.1.4 versiyonu kullanılarak bir analiz raporu üretilmiş ve çalışmada bu verilerden yararlanılmıştır.

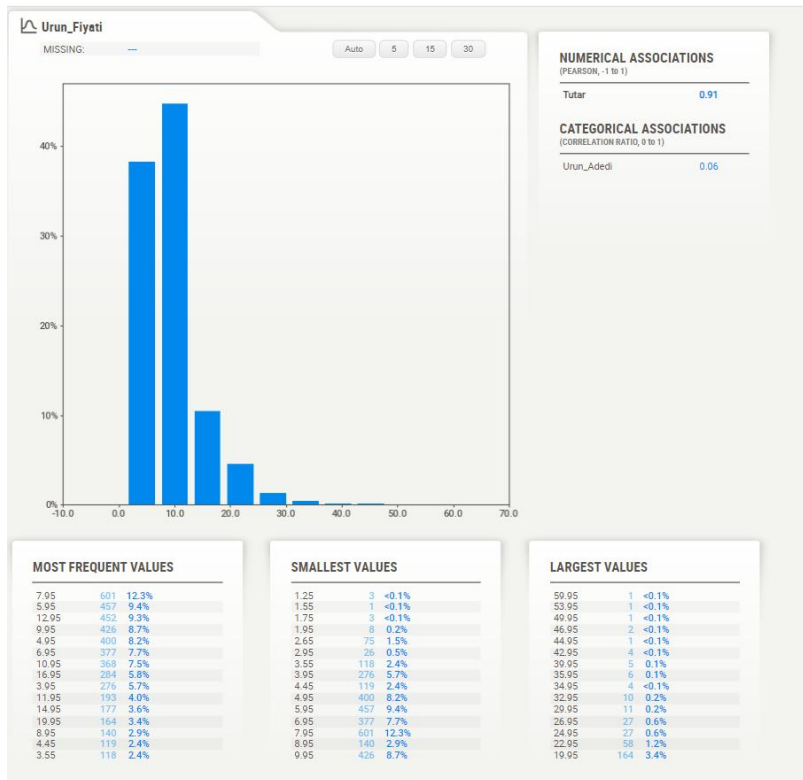
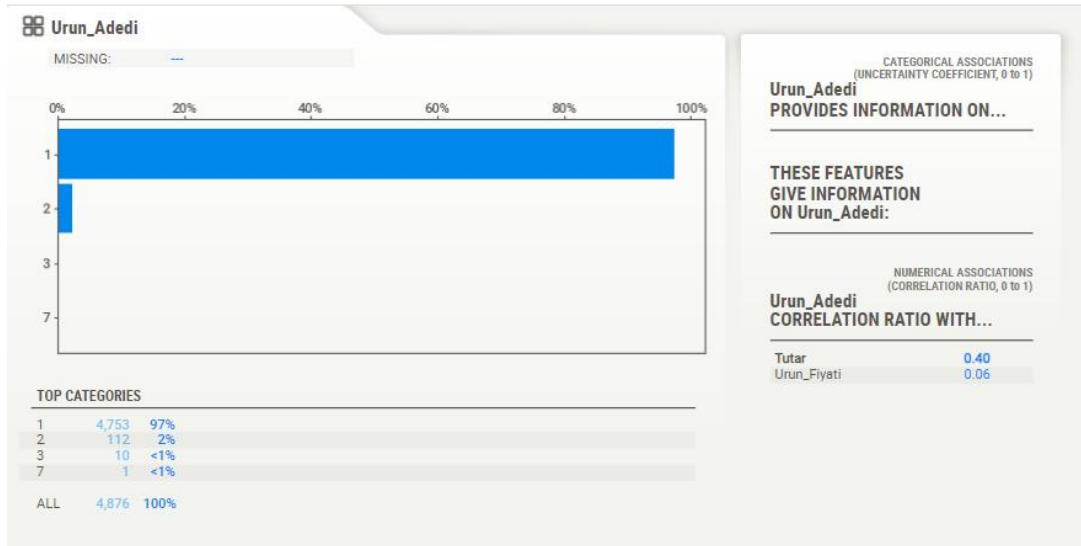


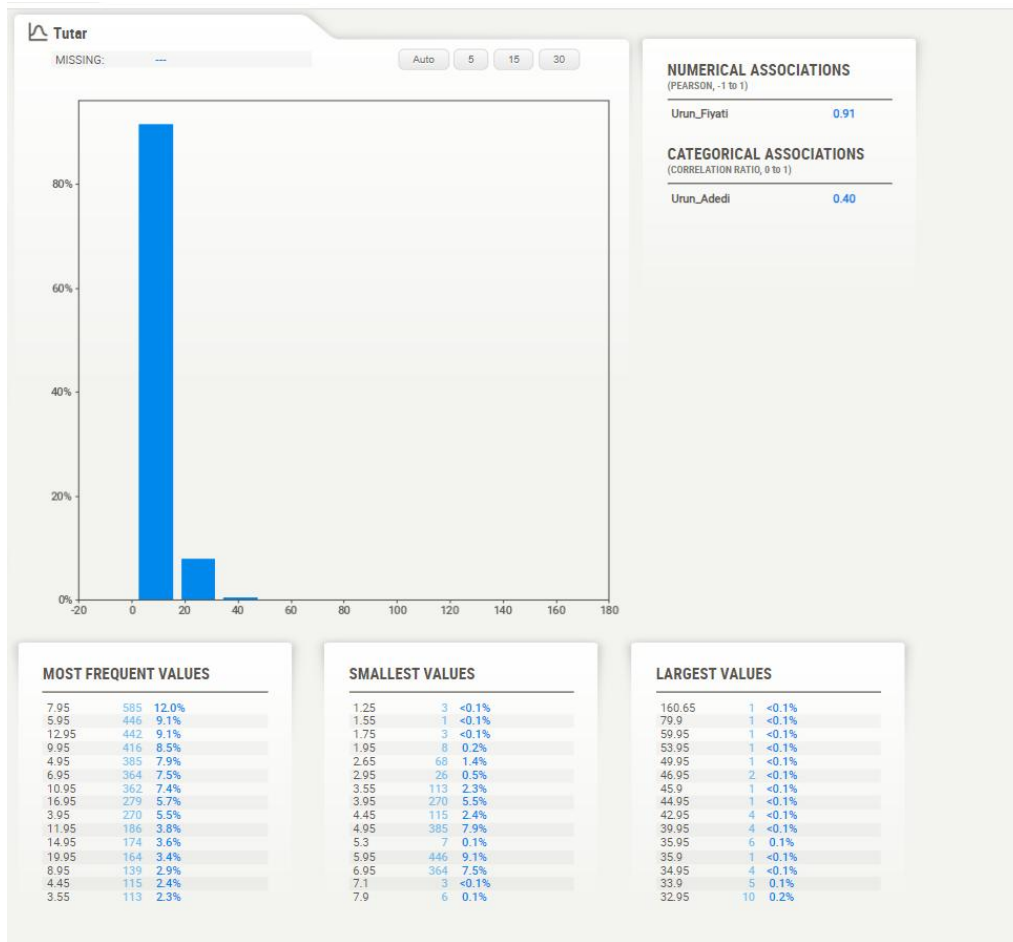
Analiz raporu aynı zamanda tüm metriklerin olduğu kolonları işleme olarak veri değişken tipleri, toplam veri girdisi adedi, satır ve sütun bilgileri gibi metriklerin bilgisini de aşağıdaki formatta kullanıcıların bilgisine sunmaktadır.



Mevcut dataframe ile oluşturulan rapor çıktıları ve korelasyon matrisi incelendiğinde nitel ve nicel verilerin korelasyon karşılaştırması istatistiksel olarak bize bir çıktı vermemektedir. Üretilen analiz raporu iş birimleri ve danışman hocalar ile incelenmiş ve çıkarımların daha gerçekçi olması adına müşteri id, işlem id gibi nitel verilerin analiz dışında bırakılmasına karar verilmiştir. Yalnızca nicel verilerin dikkate alındığı yeni veri setinden üretilen EDA istatistiksel grafikler ve korelasyon matrisi aşağıdaki gibidir.







Raporda yer verilen istatistiksel çıktıları pandas dataframe methodları ile python kodu çalıştırarak da alabilmek mümkündür. İstatistiksel değerler `info()` ve `describe()` metodları ile çıkarılmıştır.

In [104]: `data.describe()`

Out[104]:

	Urun_Adedi	Urun_Fiyati	Tutar
count	4876.000000	4876.000000	4876.000000
mean	1.028302	9.878774	10.129122
std	0.194327	5.543403	6.195273
min	1.000000	1.250000	1.250000
25%	1.000000	5.950000	5.950000
50%	1.000000	7.950000	8.950000
75%	1.000000	12.950000	12.950000
max	7.000000	59.950000	160.650000

```
In [103]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4876 entries, 0 to 4875
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   IslemNo                4876 non-null   string  
1   MusteriNo              4876 non-null   string  
2   Islem_Tarihi           4876 non-null   string  
3   Urun_Kodu              4876 non-null   string  
4   Urun_Adedi             4876 non-null   int64   
5   Urun_Fiyati            4876 non-null   float64  
6   Alan_Kodu              4876 non-null   string  
7   Yil_Ay_Gun             4876 non-null   string  
8   Tutar                  4876 non-null   float64  
dtypes: float64(2), int64(1), string(6)
memory usage: 343.0 KB
```

Çok yüksek ve düşük değerlerde medyana bakmak daha güvenilir olabilir. Yukarıdaki verilerden çıkarım yapacak olursak standart sapmanın ürün fiyatında daha yüksek olduğu görülmektedir. Bu da çeşitliliğin fazla olduğunu gösterir. Veri işlenirken standart sapmayı etkileyen bazı spesifik bozuk veriler olabilir ve bu veriler için ilgili kolondaki değerlerin ortalaması alınarak ortalama değer ataması yapılabilir veya direkt bu veriyi barındıran işlem satırı silinerek gerçeği yansıtmayan yanlış verinin modeli olumsuz etkilemesi önlenir. Projede kullanılan verinin sentetik veri olmaması ve tamamen gerçek müşteri verisi içermesinden kaynaklı olarak bu tarz bozuk veriye rastlanmamıştır. Haliyle veri seti boyutunda değişiklik yapma ihtiyacı duyulmamıştır. Keşifsel veri analizi uygulanmış istatistiksel çıkarımlar elde edilmiş ve metrikler kombinlenerek yeni kolonlar oluşturulmuş bu sayede veri seti rfm analizine ve modellemeye hazır hale getirilmiştir.

Veri seti içerisindeki numerik kolonların kovaryans matrisi çıkarılmıştır. Kovaryans matrisi kolonların 2'li kombinasyonu ile pozitif ilişkinin var olup olmadığını bizlere sunar. Böylece veriler üzerinde teorik de olsa bilgi sahibi olabiliriz ve bu da sonraki aşamalarda yorumlamamıza olanak tanır. Mevcut veri setinde pozitif bir ilişki görülmemiştir.

```
In [106]: data.select_dtypes(include=np.number).cov()
```

```
Out[106]:
```

	Urun_Adedi	Urun_Fiyati	Tutar
Urun_Adedi	0.037763	-0.029245	0.375894
Urun_Fiyati	-0.029245	30.729311	31.359069
Tutar	0.375894	31.359069	38.381402

“Korelasyon katsayısı 1'e yaklaştıkça ilişki düzeyi artarken, -1'e yaklaştıkça ilişki düzeyi zayıflamaktadır. Korelasyon katsayısı “-1” ve “+1” arasında değişen değerler alır. Kovaryans ve Korelasyon , birbirinin tam tersi iki terimdir, ikisi de istatistikte ve regresyon analizinde kullanılır, kovaryans bize iki değişkenin birbirinden nasıl farklı olduğunu gösterirken korelasyon bize iki değişken arasındaki ilişkiyi ve bunların nasıl ilişkili olduğunu gösterir.”

<https://tr.wikipedia.org/wiki/Korelasyon> [Erişim 21.02.22].

```
In [114]: data['Harcama_Skoru'] = data['Urun_Fiyati']*data['Urun_Adedi']  
data.head()
```

```
Out[114]:
```

	IslemNo	MusteriNo	Islem_Tarihi	Urun_Kodu	Urun_Adedi	Urun_Fiyati	Alan_Kodu	Yil_Ay_Gun	Tutar	Yil	Ay	Harcama_Skoru
0	4154205	2126662	4.09.2016	1669027	1	12.95	48	20160904	12.95	2016	3	12.95
1	4154205	2126662	4.09.2016	1669794	1	24.95	48	20160904	24.95	2016	3	24.95
2	4154205	2126662	4.09.2016	1737107	1	32.95	48	20160904	32.95	2016	3	32.95
3	4154205	2126662	4.09.2016	1746673	1	12.95	48	20160904	12.95	2016	3	12.95
4	4154205	2126662	4.09.2016	1751216	1	29.95	48	20160904	29.95	2016	3	29.95

Ürün adedine bağlı olarak toplam harcama tutarı ayrı bi kolonda toplanmıştır ve hangi üründen kaç adet aldı ise toplam tutar data['Harcama_Skoru'] kolonunda tutulmuştur. Daha sonra bu kolon RFM analizinde Tutar (Monetary) metriği olarak kullanılacaktır.

Keşifsel veri analizinin önemli bir adımı da veri setindeki değerler arasında ilişkisel çıkarımlarda bulunmaktır. Böylece verinin anlamlandırılması ve analizi sağlanır. Gruplandırma yapılırken veri setinden yalnızca numerik kolonlar dahil edilmiştir. Bu aşamada groupby() metodu kullanılırken kolonları önceliklendirmenin önemli olduğu farkedilmiştir, aşağıda aynı 2 kolonun farklı sırayla

groupby() çıktıları listelenmiştir.Aynı işlem no veya müşteriye ait verilerin gruplandırılması için aggregate (np.sum) kullanılmıştır.

```
data_groupby = data.groupby(['MusteriNo','IslemNo']).aggregate(np.sum)
data_groupby.head()
```

		Urun_Adedi	Urun_Fiyati	Tutar	Yıl	Ay	Harcama_Skoru
MusteriNo	IslemNo						
1011679	8904752	10	53.50	53.50	20180	30	53.50
1030597	8768212	6	81.70	81.70	12108	18	81.70
1149513	8432773	4	73.80	73.80	8072	12	73.80
1214440	8522163	5	59.75	59.75	10090	15	59.75
123482	8157301	3	65.85	65.85	6054	9	65.85

```
data_groupby = data.groupby(['IslemNo','MusteriNo']).aggregate(np.sum)
data_groupby.head()
```

		Urun_Adedi	Urun_Fiyati	Tutar	Yıl	Ay	Harcama_Skoru
IslemNo	MusteriNo						
4069408	2209933	11	97.50	105.45	20160	30	105.45
4085360	2216191	2	26.90	26.90	4032	6	26.90
4100762	2069158	5	37.75	37.75	10080	15	37.75
4130333	2240836	4	21.80	21.80	8064	12	21.80
4147365	2246658	4	83.80	83.80	8064	12	83.80

3.4. RFM tablosu oluşturma

Bu aşamadan sonra RFM analizine gidilmiş ve sırayla Güncellik(Recency) - Tutar(Monetary) - Islem Adedi(Frequency) için kolonlar oluşturulmuş daha sonra rfm_table isimli dataframede biraraya getirilmiştir.

data_R : Toplam harcama tutarı verisi müşteri bazında groupby ile aggregate edilerek kişi başına düşen tutar üzerinden alışverişin güncellik değeri belirlenmiştir.

```
In [123]: data_R = data.groupby('MusteriNo').agg({'Harcama_Skoru': lambda x: x.sum(),
        'Islem_Tarihi': lambda x: (today - x.max()).days})
data_R.rename(columns={'Harcama_Skoru': 'Tutar', 'Islem_Tarihi': 'Güncellik'}, inplace=True)
data_R.head()
```

```
Out[123]:
```

	Tutar	Güncellik
MusteriNo		
1011679	53.50	117
1030597	81.70	219
1149513	73.80	28
1214440	59.75	252
123482	65.85	90

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

```
In [124]: data_F = data.groupby(['MusteriNo', 'IslemNo']).agg({'Urun_Adedi': lambda x: x.count()})
data_F.head(10)
```

```
Out[124]:
```

		Urun_Adedi
MusteriNo	IslemNo	
1011679	8904752	10
1030597	8768212	6
1149513	8432773	4
1214440	8522163	5
123482	8157301	3
1330031	8211139	25
1360570	6174009	6
1370387	8954132	6
1407333	8357988	7
1433012	8467593	7

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

data_M: Toplam harcama tutarını 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.

```
In [31]: data_M = data.groupby(['MusteriNo']).agg({'Tutar': lambda x: x.sum()})
data_M.head()
```

Out[31]:

	Tutar
MusteriNo	
1011679	53.50
1030597	81.70
1149513	73.80
1214440	59.75
123482	65.85

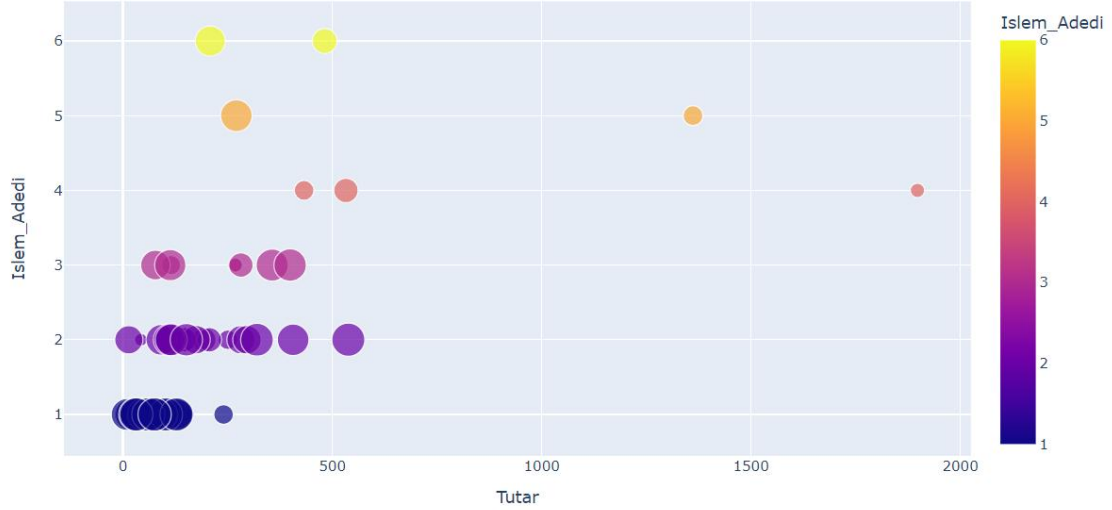
RFM analizi için oluşturulan 3 ayrı parametrenin dataframe i mergelenerek tabloya son hali verilmiştir. Böylece rfm_table aşağıdaki gibidir. Tablodan çıkarım yapılacak olursa 172464 id li müşteri 1 yıldan fazla süredir alışveriş yapmamış, 5 kez alışverişte bulunmuş ve totalde 309,95 tutarında harcaması olmuştur.

```
In [100]: rfm_table= pd.merge(data_R,data_F, on='MusteriNo')
rfm_table = rfm_table[['Güncellik', 'Islem_Adedi', 'Tutar']]
rfm_table.head(30)
```

Out[100]:

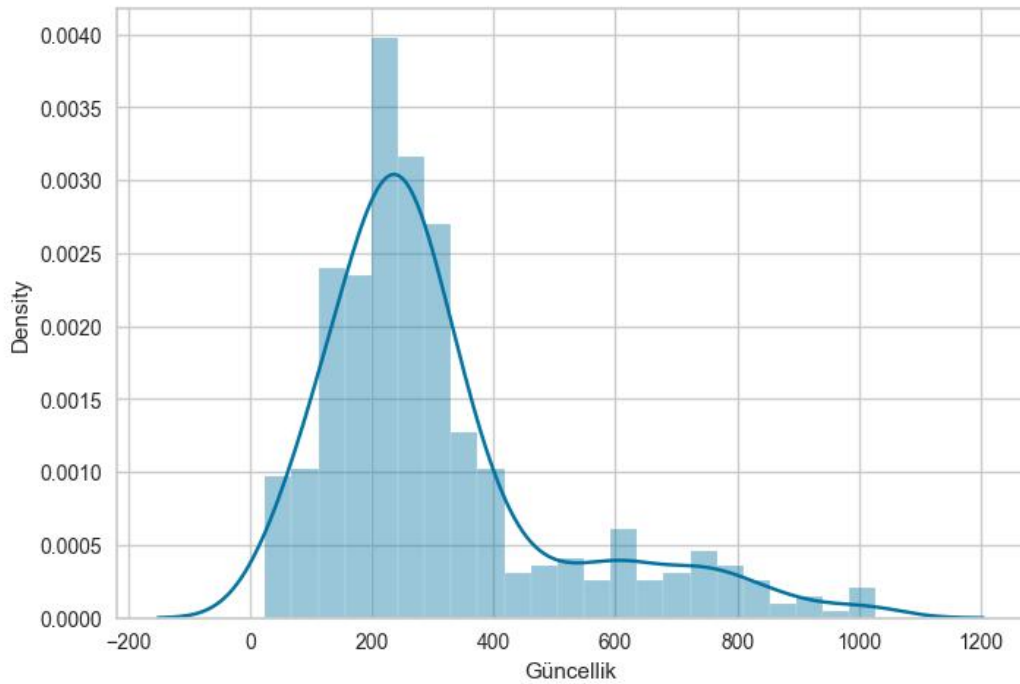
	Güncellik	Islem_Adedi	Tutar
MusteriNo			
1011679	117	1	53.50
1030597	219	1	81.70
1149513	28	1	73.80
1214440	252	1	59.75
123482	90	1	65.85
1330031	289	1	182.45
1360570	545	1	79.70
1370387	199	1	61.70
1407333	303	1	81.65
1433012	280	1	80.65
1463737	524	2	231.80
1484115	57	2	250.50
1493335	559	1	69.80
1507651	292	1	50.75
1510143	252	2	211.15
1524932	321	1	147.50
1528284	805	2	85.50
1550173	750	1	29.95
1599491	220	1	34.75
1691865	277	1	43.75
172464	581	5	309.95
1742304	505	8	443.50
1771850	322	1	50.80
1821926	729	1	8.95
1881392	140	2	91.70
1932878	226	1	108.45
1979908	149	5	270.80
2013642	681	1	38.80
2029800	663	2	281.65
2069158	858	1	37.75

RFM tablosu metriklerinden işlem adedi ve tutar bilgisine dair verilerin grafik görünümü aşağıdaki gibidir. Bu görselden işlem adedinin genelde sınırlı seyrettiği bilgisine ulaşılabilir haliyle standart sapması yüksek bir metriktir.

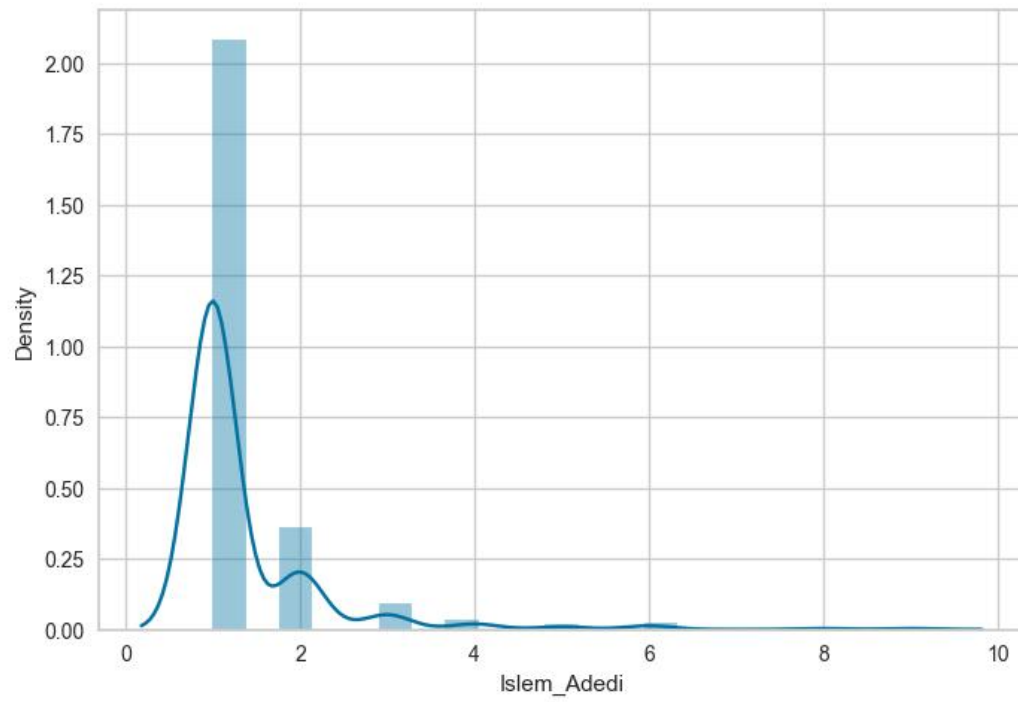


RFM metriklerine ait veri kolonlarının Seaborn (v0.12.2) veri görselleştirme Python kütüphanesi kullılarak aşağıdaki grafikler elde edilmiştir.

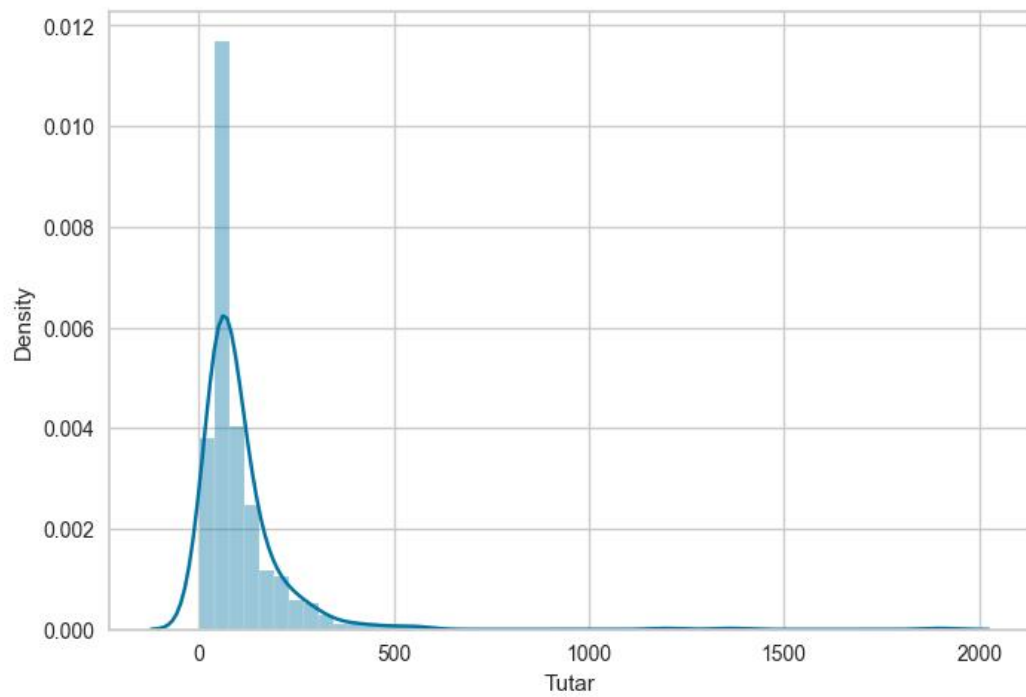
```
sns.distplot(rfm_table["Güncellik"])  
plt.show()
```



```
: sns.distplot(rfm_table["Islem_Adedi"])  
plt.show()
```



```
sns.distplot(rfm_table["Tutar"])  
plt.show()
```



3.5. RFM Skoru Hesaplama ve Müşteri Segmentasyonu

RFM analizi skoru bazında müşterileri 5 eşit parçaya ayırarak çeyreklere bölünmüştür. Bunun için 2 ayrı yöntem de kullanılabilir. İlki fonksiyon içerisinde bölümlmek 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['Güncellik'])` kullanılmıştır. Bu işlem sonucu RFM sonucu aşağıdaki gibi listelenmiştir.

```
In [129]: def Frekans_fonk(x,p,d):
            if x <= d[p][0.20]:
                return 0
            elif x <= d[p][0.40]:
                return 1
            elif x <= d[p][0.60]:
                return 2
            elif x <= d[p][0.80]:
                return 3
            else:
                return 4

            quantiles = rfm_table.quantile(q=[0.20,0.40,0.60,0.80])
            quantiles = quantiles.to_dict()

            quantiles

Out[129]: {'Güncellik': {0.2: 159.0,
                        0.4: 225.0,
                        0.6: 282.79999999999995,
                        0.8: 385.20000000000005},
           'Islem_Adedi': {0.2: 1.0, 0.4: 1.0, 0.6: 1.0, 0.8: 2.0},
           'Tutar': {0.2: 50.8, 0.4: 60.78, 0.6: 82.7, 0.8: 139.52000000000004}}
```

```
In [130]: #Recency bulma
rfm_table = rfm_table.sort_values('Güncellik',ascending=True)
rfm_table['Güncellik_Ceyrek'] = pd.qcut(rfm_table['Güncellik'],5,labels=False)

#Frequency bulma
rfm_table['Islem_Adedi_Ceyrek'] = rfm_table['Islem_Adedi'].apply(Frekans_fonk, args=('Islem_Adedi',quantiles,))

#Monetary bulma
#rfm_table.sort_values('Monetary',ascending=False)
rfm_table['Tutar_Ceyrek'] = pd.qcut(rfm_table['Tutar'],5,labels=False);
```

Tabloda sıfır değeri yer almamasın 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.

```
In [131]: rfm_table['Güncellik_Ceyrek'] = rfm_table['Güncellik_Ceyrek'] + 1
rfm_table['Islem_Adedi_Ceyrek'] = rfm_table['Islem_Adedi_Ceyrek'] + 1
rfm_table['Tutar_Ceyrek'] = rfm_table['Tutar_Ceyrek'] + 1

rfm_table['RFM Skoru'] = rfm_table['Güncellik_Ceyrek'].map(str) + rfm_table['Islem_Adedi_Ceyrek'].map(str) + rfm_table['Tutar_Ceyrek'].map(str)
rfm_table.head()
```

Out[131]:

	Güncellik	Islem_Adedi	Tutar	Güncellik_Ceyrek	Islem_Adedi_Ceyrek	Tutar_Ceyrek	RFM Skoru
MusteriNo							
3933530	25	2	42.85	1	4	1	141
3850993	26	1	10.95	1	1	1	111
3779930	27	1	150.20	1	1	5	115
1149513	28	1	73.80	1	1	3	113
3706499	28	1	37.80	1	1	1	111

3933530 id li müşteriye tabloda bakılacak olursa RFM skoru 141 dir. 25 gündür herhangi bir işlem yapmamış toplamda 2 kez alışveriş yapmış ve toplam 42,85 tutarında ödeme yapmıştır. İşlem adedi veri setinde genelde 1 olduğu için 1+ alışverişte Frequency puanının hemen 4 e yükseldiği görülür. Sayısal verilerde virgülden sonra 1 basamak için round() methodu kullanılmıştır.

Bu şekilde, segmentasyon işlemi tamamlanmıştır. Örneğin 111 segmentini yorumlayacak olursak, bu segmentte 17 kişi bulunmakta, ortalama olarak en son alışverişleri 112 gün önce gerçekleşmiş, alışveriş sıklıkları 1, yani 1 alışverişleri var, 32 TL harcamaları olmuştur.

```
In [132]: rfm_table.groupby('RFM Skoru').agg({
'Güncellik': ['mean', 'min', 'max', 'count'],
'Islem_Adedi': ['mean', 'min', 'max', 'count'],
'Tutar': ['mean', 'min', 'max', 'count'] }).round(1).head()
```

Out[132]:

	Güncellik				Islem_Adedi				Tutar			
	mean	min	max	count	mean	min	max	count	mean	min	max	count
RFM Skoru												
111	112.8	26	159	17	1.0	1	1	17	32.0	10.0	50.8	17
112	115.0	30	152	9	1.0	1	1	9	54.5	51.8	57.8	9
113	118.4	28	154	15	1.0	1	1	15	72.6	62.8	82.6	15
114	109.1	59	156	14	1.0	1	1	14	107.1	88.6	132.4	14
115	42.5	27	58	2	1.0	1	1	2	195.2	150.2	240.3	2

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. RFM skorları belirlenmesi ardından müşteriler ideal, ortalama ve kaybedilebilir olmak üzere 3 kategoride etiketlenmiştir.

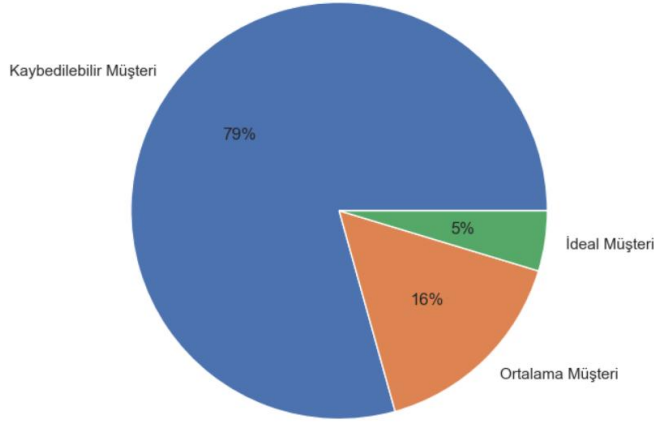
```
In [141]: rfm_table["Müşteri_segmenti"] = np.where(rfm_table['RFM_Skor_Norm'] >
                                                    4.5, "İdeal Müşteri",
                                                    (np.where(
                                                        rfm_table['RFM_Skor_Norm'] > 4,
                                                        "Ortalama Müşteri",
                                                        (np.where(
                                                            rfm_table['RFM_Skor_Norm'] > 3,
                                                            "Ortalama Müşteri",
                                                            np.where(rfm_table['RFM_Skor_Norm'] > 1.6,
                                                                'Kaybedilebilir Müşteri', 'Kaybedilebilir Müşteri'))))))
rfm_table[['RFM_Skor_Norm', 'Müşteri_segmenti']].head()
```

Out[141]:

	RFM_Skor_Norm	Müşteri_segmenti
MusteriNo		
3933530	4.42	Ortalama Müşteri
3850993	2.44	Kaybedilebilir Müşteri
3779930	2.44	Kaybedilebilir Müşteri
1149513	2.43	Kaybedilebilir Müşteri
3706499	2.43	Kaybedilebilir Müşteri

RFM skoru neticesinde etiketlenmiş müşteri verisinin kategorilere bağlı olarak pasta grafiği görünümü aşağıdaki gibidir. Bu görselde harcama davranışlarına bakılarak kaybedilme potansiyeli yüksek müşteri popülasyonunun yüksek olduğunu görüyoruz. Böylece bu kitleye özgü pazarlama kampanya çalışmaları yapılarak kitlenin geri kazanımı alınabilecek satış pazarlama aksiyonları ile mümkündür.

```
In [143]: plt.pie(rfm_table.Müşteri_segmenti.value_counts(),
                labels=rfm_table.Müşteri_segmenti.value_counts().index,
                autopct='%0.0f%%')
plt.show()
```



3.7. KMeans Kümeleme Algoritmasının Uygulanması

RFM analizi ile segmente edilen müşteri popülasyonu ek olarak gözetimli makine öğrenmesi algoritmalarından Kmeans kümeleme algoritması uygulanarak da modellenmiş ve çıktıların karşılaştırılması sağlanmıştır. Kümeleme algoritmaları uzaklık bazlı çalışır bu nedenle verinin aynı ölçekte olması için normalizasyon işlemi kullanılmıştır.

```
In [43]: rfm_table['R_rank'] = rfm_table['Güncellik'].rank(ascending=False)
rfm_table['F_rank'] = rfm_table['Islem_Adedi'].rank(ascending=True)
rfm_table['M_rank'] = rfm_table['Tutar'].rank(ascending=True)

rfm_table['R_rank_norm'] = (rfm_table['R_rank']/rfm_table['R_rank'].max())*100
rfm_table['F_rank_norm'] = (rfm_table['F_rank']/rfm_table['F_rank'].max())*100
rfm_table['M_rank_norm'] = (rfm_table['M_rank']/rfm_table['M_rank'].max())*100

rfm_table.drop(columns=['R_rank', 'F_rank', 'M_rank'], inplace=True)
rfm_table.head()
```

```
Out[43]:
```

	Güncellik	Islem_Adedi	Tutar	Güncellik_Ceyrek	Islem_Adedi_Ceyrek	Tutar_Ceyrek	RFM Skoru	R_rank_norm	F_rank_norm	M_rank_norm
MusteriNo										
3933530	25	2	42.85	1	4	1	141	100.000000	86.333333	86.333333
3850993	26	1	10.95	1	1	1	111	99.777778	39.777778	39.777778
3779930	27	1	150.20	1	1	5	115	99.555556	39.777778	39.777778
1149513	28	1	73.80	1	1	3	113	99.111111	39.777778	39.777778
3706499	28	1	37.80	1	1	1	111	99.111111	39.777778	39.777778

Normalize edilmiş olan RFM veri tablosunun son hali aşağıdaki gibi görünmektedir. 3 metrik için olan kolondaki tüm değerler 0 ve 1 bandından yeniden atanmış böylece noktaların birbirine yaklaştırılması sağlanmıştır.

```
In [53]: from sklearn.preprocessing import MinMaxScaler
x_scaled = MinMaxScaler().fit_transform(clus)
data_scaled2 = pd.DataFrame(x_scaled)

data_scaled2.head()
```

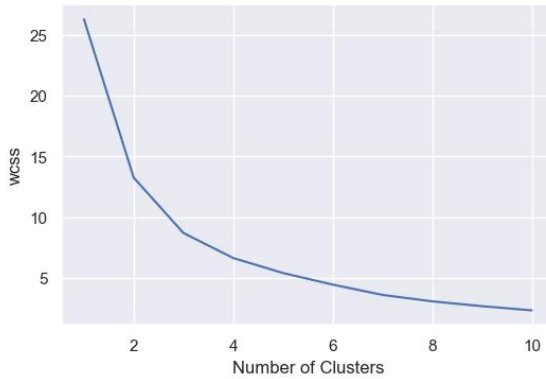
Out[53]:

	0	1	2
0	0.000000	0.125	0.020024
1	0.000997	0.000	0.003170
2	0.001994	0.000	0.076742
3	0.002991	0.000	0.036377
4	0.002991	0.000	0.017356

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. Mevcut veri seti için en uygun küme adedinin grafiğe bakılarak k=4 olduğu kararı alınmıştır.

```
In [54]: #initialize kmeans parameters
kmeans_kwargs = {
    "init": "random",
    "n_init": 10,
    "random_state": 1,
}

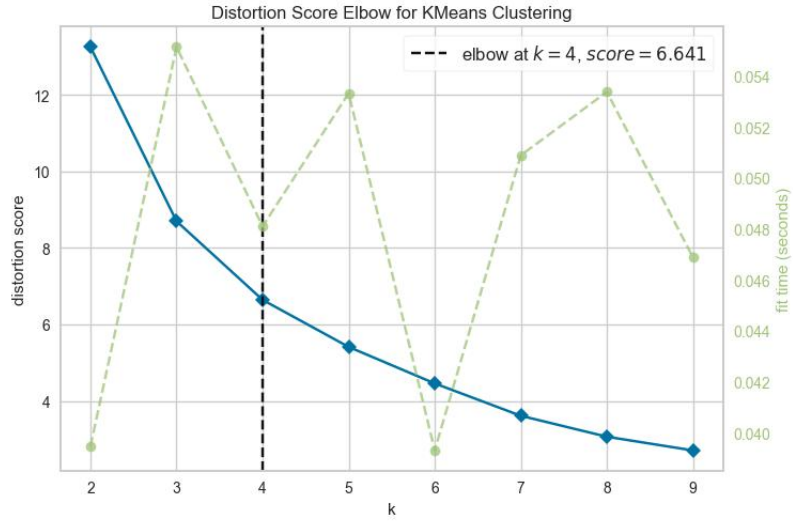
wcss = []
for i in range(1,11):
    kmeans = KMeans(n_clusters=i, init='k-means++', n_init=10, max_iter=100)
    kmeans.fit(data_scaled2)
    wcss.append(kmeans.inertia_)
plt.plot(range(1,11),wcss)
plt.xlabel("Number of Clusters")
plt.ylabel("WCSS")
plt.show()
```



Normalize edilen RFM tablosu bir diğer yaygın kullanılan dirsek methodu kütüphanesi KElbowVisualizer ile modellendiğinde de optimum küme sayısı 4 olduğu gözlemlenmiştir.

```
In [60]: from yellowbrick.cluster import KElbowVisualizer
```

```
model = KMeans()  
visualizer = KElbowVisualizer(model, k=(2,10))  
visualizer.fit(rfm_normalized)  
visualizer.show()  
visualizer.elbow_value_
```

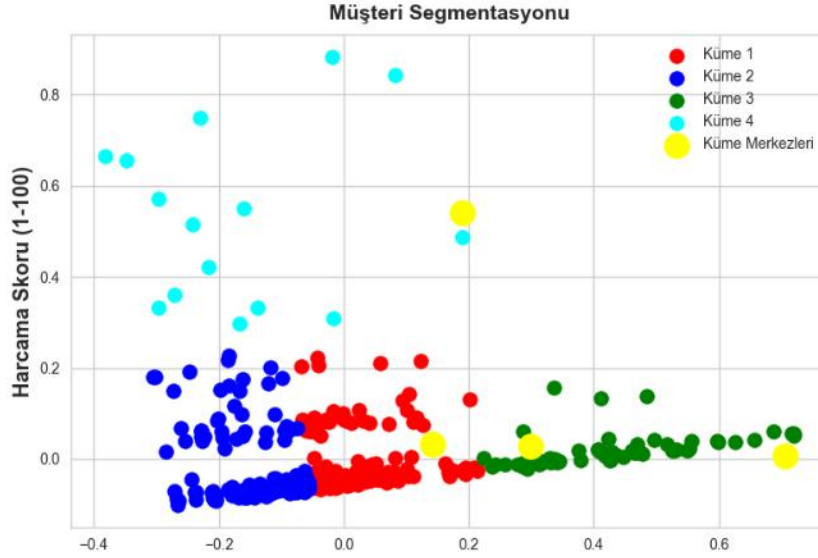


```
Out[60]: 4
```

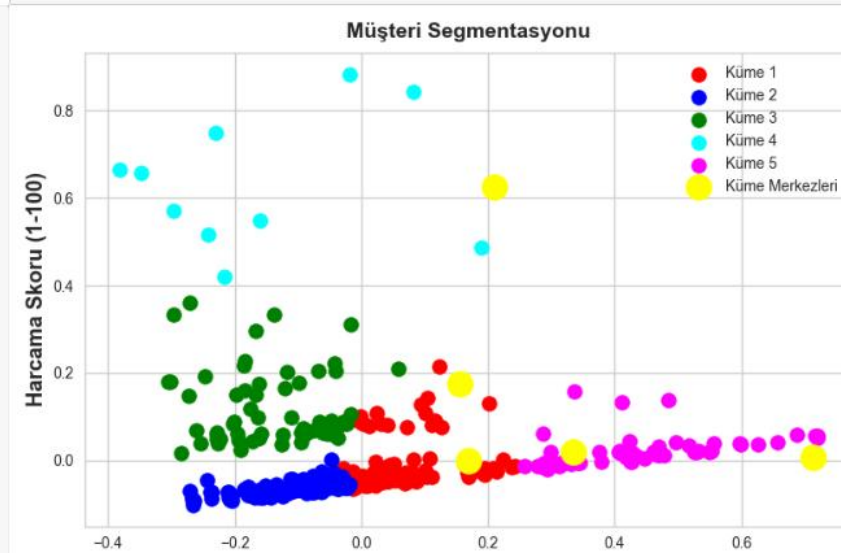
Dirsek methodu sonrası 4 kümenin optimum çözüm olması kararından sonra kmeans algoritması uygulanmıştır. Kümelerin dağılım grafiği aşağıdaki gibidir.

```
In [113]: plt.scatter(rfm_2d[y_kmeans == 0, 0], rfm_2d[y_kmeans == 0, 1], s = 100, c = 'red', label = 'Küme 1')
plt.scatter(rfm_2d[y_kmeans == 1, 0], rfm_2d[y_kmeans == 1, 1], s = 100, c = 'blue', label = 'Küme 2')
plt.scatter(rfm_2d[y_kmeans == 2, 0], rfm_2d[y_kmeans == 2, 1], s = 100, c = 'green', label = 'Küme 3')
plt.scatter(rfm_2d[y_kmeans == 3, 0], rfm_2d[y_kmeans == 3, 1], s = 100, c = 'cyan', label = 'Küme 4')

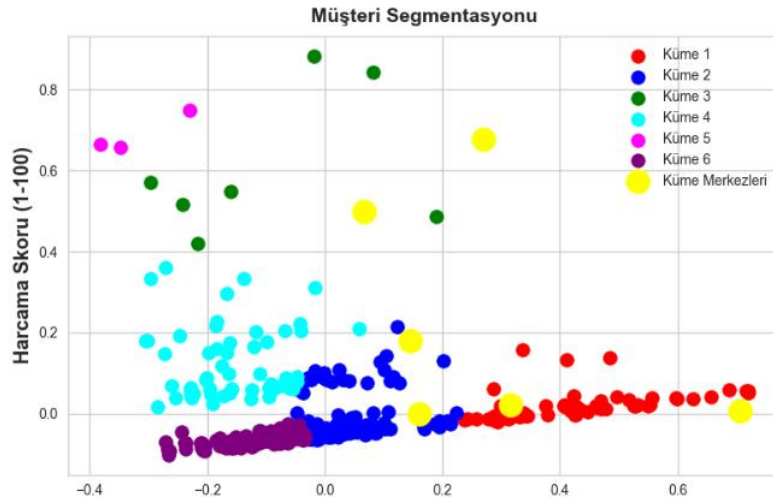
plt.scatter(kmeans.cluster_centers[:, 0], kmeans.cluster_centers[:, 1], s = 300, c = 'yellow', label = 'Küme Merkezleri')
plt.title('Müşteri Segmentasyonu')
plt.ylabel('Harcama Skoru (1-100)')
plt.legend()
plt.show()
```



k=5 alınması halinde ise dağılım aşağıdaki gibi olmuştur.



k=6 alınması halinde ise dağılım aşağıdaki gibi olmuştur.



Sklearn kütüphanesinin kümeleme kütüphanelerinde MiniBatchKMeans ile kümelendiğinde ise çıktı aşağıdaki gibi gözlemlenmiştir.

```
In [77]: from sklearn.cluster import MiniBatchKMeans

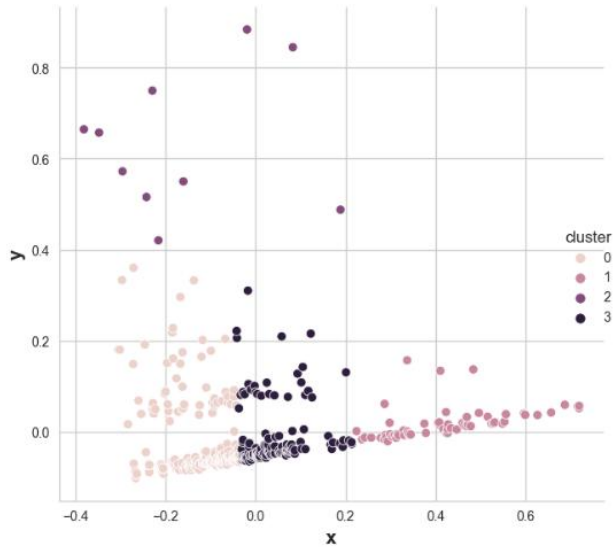
kmeans = MiniBatchKMeans(n_clusters=4, random_state=0, batch_size=6)
data = pd.DataFrame(clus)
data['cluster'] = kmeans.fit_predict(data_scaled2)

pca = PCA(2)

#Transform the data
pca_df = pca.fit_transform(data_scaled2)

data = pd.DataFrame(pca_df, columns = ['x','y'])
data['cluster'] = kmeans.fit_predict(data_scaled2)

plt.style.use('seaborn-whitegrid')
plt.rc("figure", autolayout=True)
plt.rc("axes", labelweight='bold', labelsz='large', titleweight='bold', titlesize=14, titlepad=10)
sns.relplot(x='x', y='y', hue='cluster', data=data, height=6)
plt.show()
```



RFM analizi tablosu 450 tane satırdan oluşur ve tüm verilerin kümelere dağılımı pandas kütüphanesine ait `value_counts()` fonksiyonu ile hesaplanmıştır.

```
In [177]: d_frame['cluster'].value_counts()
Out[177]: 0    199
          1    175
          3     61
          2     15
          Name: cluster, dtype: int64
```

Kümelerin 3 metrik için değer ortalamaları pandas kütüphanesine ait `mean()` fonksiyonu ile hesaplanmıştır.

```
In [169]: d_frame.groupby('cluster').mean()
Out[169]:
```

	Güncellik	Islem_Adedi	Tutar
cluster			
0	166.949749	1.266332	96.086181
1	322.834286	1.217143	97.392571
2	214.800000	5.333333	554.506867
3	729.836066	1.049180	80.445082

Kümelerden birine ait örnek veri dağılımı aşağıdaki gibidir.

```
In [180]: d_frame[d_frame['cluster']==2]
```

Out[180]:

MusteriNo	Güncellik	Islem_Adedi	Tutar	cluster
2440067	31	4	1897.65	2
3702981	58	4	432.55	2
2244384	59	5	1361.60	2
2442917	88	4	532.55	2
2236656	92	6	481.80	2
2755259	135	6	208.45	2
1979908	149	5	270.80	2
2909858	180	4	194.90	2
856703	188	6	1197.75	2
3156195	216	4	331.80	2
3327400	222	6	304.75	2
2473009	331	4	162.00	2
2514734	407	9	187.55	2
1742304	505	8	443.50	2
172464	561	5	309.95	2

3.8. Kümeleme Başarı Değerlendirmesi

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 değerlendirilmesi için bu katsayı hesaplanmıştır.

```
In [188]: from sklearn.metrics import silhouette_score
inertia_list=[]
silhouette_score_list=[]
for i in range(2,10):
    kmeans.fit(rfm_normalized)
    silhouette_score_list.append(silhouette_score(data_scaled2,kmeans.labels_))
    print(silhouette_score_list)

[0.6083762306752909]
[0.6083762306752909, 0.6083762306752909]
[0.6083762306752909, 0.6083762306752909, 0.6083762306752909]
[0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.6083762306752909]
[0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.6083762306752909]
[0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.603007715638089]
[0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.603007715638089, 0.6082521937882489]
[0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.6083762306752909, 0.603007715638089, 0.6082521937882489, 0.6083762306752909]
```

”Silhouette Score, temel gerçekliğe sahip olmadığımız durumlarda en çok kullandığımız ölçüm metriklerden birisidir. Bir veri noktası için

silüet katsayısı $Sil(i) = (b(i) - a(i)) / \max(a(i), b(i))$ dir. Bir veri noktası ile kümedeki diğer tüm veri noktaları arasındaki ortalama mesafe ($a(i)$), bu veri noktasına en yakın kümede ki diğer tüm veri noktalarına olan ortalama uzaklık ($b(i)$) dir. “Davies Bouldin Score” ise yine kümeler arası ortalama benzerliği ölçmek için kullanılır. Bu oran için düşük değerler tercih edilir ve minimum skor değeri 0 dır.” Bektaş, Muhammet.

“Kümeleme Algoritmaları ile Müşteri Segmentasyonu”

<https://medium.com/@mbektas> [Erişim 16.10.2021].

```
In [165]: kmeans.cluster_centers_  
Out[165]: array([[0.14191426, 0.033125 , 0.04803165],  
                [0.29856146, 0.02714286, 0.0491431 ],  
                [0.1892323 , 0.54166667, 0.29035593],  
                [0.70606514, 0.00625 , 0.03926974]])
```

“Inertia cluster içi noktaların mesafesini temsil eder. Modelleme yapılırken bu değerin minimize edilmesine çalışılır. Yukarıdaki merkeze uzaklıklara bakılırsa 4 veya 5 kümenin optimum çözüm olabileceği bilgisine ulaşılır. K-Means kümeleme algoritmasının amacı, basitçe k adet kümenin ağırlık merkezlerini olabilecek en iyi şekilde seçmektir. Bu ağırlık merkezlerine centroid adı verilir. Olabilecek en iyi şekilde seçildiğinde, kayıp fonksiyonu(yani inertia değeri) minimize edilmiş olur.” Güner, Göker. <https://medium.com/software-development-turkey> [Erişim 16.10.2021].

4. TARTIŞMA VE SONUÇ

RFM analizi ve makine öğrenme algoritması olan kümüleme yönteminin modellenmesi neticesinde sonuçların birbiri ile tutarlı olduğu ve örtüştüğü görülmüştür. İlgili veri seti üzerinde çalışırken hem

veri analitiği aşamaları adım adım uygulanmış hem de keşifsel veri analizi yapılmıştır. Bu çalışma neticesinde verinin ham veriden işlenmiş veriye dönüşümü esnasında karşılaşılabilecek zorluklar deneyimlenmiş gerçek veri ile çalışırken her zaman aşamaların sorunsuz çıktılar üretemeyeceği gözlemlenmiştir. Bu nedenle veri işlenirken istatistiksel analizi ve veri dağılımı çıkarılmış, kolonlar içerisindeki veri türleri tek tip olması için ortak formatta tekrar düzenlenmiştir. Veri okunurluğu için de sık sık Python dataframe formatına ve grafiklere başvurulmuş, verilerin görselleştirilmesi sağlanmıştır. RFM metrikleri olan güncellik, sıklık ve tutar değerleri doğrudan veri setinde hazır olarak bulunmuyordu. Bu veriler farklı kolonların kombinasyonu ile elde edilmiştir. Örneğin frekans değeri için referans bir tarih bilgisi atanmış ve bu tarihe uzaklıklar gün bazında yeniden hesaplanarak güncellik metriği kolonu oluşturulmuştur. Aynı şekilde müşterilerin ilgili mağazadan alışveriş sıklığı veri seti içerisinde groupby() fonksiyonu ile gruplandırılarak bulunmuştur. Toplam harcama tutarı için her müşteriye ait işlem id leri biraraya getirilmiş aynı işlem içerisindeki ürün adedi ve ürün fiyatı çarpılarak toplam tutar kolonu elde edilmiştir. Makine öğrenmesi algoritması uygulanırken de yine RFM tablosu kullanılmış fakat veri değer aralıkları birbirinden çok bağımsız olduğundan noktaları birbirine yaklaştırmak adına normalizasyon işlemi uygulanmıştır.

Segmente edilen müşteri popülasyonunun verileri kontrol edildiğinde kaybedilebilir müşteri segmentinde yer alan ve potansiyel kayıp niteliğinde bakılan müşteri profilinin sahiden de alışveriş sıklığının seyreltiği, harcama tutarlarının düştüğü ve/veya güncellik parametresinin artık günceli yansıtmadığı bilgisine ulaşılmıştır. İdeal müşteri profili segmentine bakıldığında ise alışveriş periyodunun sık güncellik parametresinin oldukça düşük değerler seyrettiği görülmüş ve harcama tutarları da yüksektir. Bu bilgiler ışığında müşteri

popülasyonunda kayıp vermemek adına kaybedilebilir müşteri segmenti adına satış ve pazarlama çalışmalarına ağırlık verilip iletişim kanalları ile kampanya bilgilendirmelerinde bulunmak şirket nezdinde satış trendlerini olumlu etkileyecektir. Buna ek olarak ideal ve ortalama müşteri popülasyonunun tercih ettiği ürün kategorileri en çok tercih edilen ürünler olarak değerlendirilebilir ve üretim fazında, ürün geliştirme süreçlerinde bu ürün gruplarına ağırlık verilebilir. RFM analizi sonrası segmentasyon ve buna ek bu veriler ışığında yapılabilecek ürün kategorizasyonları şirketlere iş geliştirme süreçlerinde rehber niteliği taşır. Böylece geçmiş müşteri harcama verilerine bakılarak geleceğe dönük tutarlı tahminlemeler yapılabileceği bilgisine bu tez çalışması sonucu ulaşılmıştır.

İŞLENMİŞ VERİ SETİ ÖRNEKLEMİ

IslemNo	MusteriNo	Islem_Tarihi	Urun_Kodu	Urun_Adedi	Urun_Fiyati	Tutar
4154205	2126662	2016-04-09	1669027	1	12,95	12,95
4154205	2126662	2016-04-09	1669794	1	24,95	24,95
4154205	2126662	2016-04-09	1737107	1	32,95	32,95
4154205	2126662	2016-04-09	1746673	1	12,95	12,95
4154205	2126662	2016-04-09	1751216	1	29,95	29,95
4846981	2224954	2016-12-13	1747611	1	12,95	12,95
4846981	2224954	2016-12-13	1761517	1	11,95	11,95
4846981	2224954	2016-12-13	1761604	1	12,95	12,95
4846981	2224954	2016-12-13	1787959	1	8,95	8,95
5782088	2846339	2017-03-25	1824747	1	19,95	19,95
5782088	2846339	2017-03-25	1840632	1	5,95	5,95
5782088	2846339	2017-03-25	1841460	1	5,95	5,95
5782088	2846339	2017-03-25	1849029	1	6,95	6,95
5782088	2846339	2017-03-25	1851319	1	6,95	6,95
5851166	2868027	2017-03-31	1713729	1	7,95	7,95
5851166	2868027	2017-03-31	1734283	1	4,95	4,95
5851166	2868027	2017-03-31	1821279	1	5,95	5,95
5851166	2868027	2017-03-31	1829484	1	3,55	3,55
5851166	2868027	2017-03-31	1854587	1	6,95	6,95
5851166	2868027	2017-03-31	1856316	1	8,95	8,95
5851166	2868027	2017-03-31	1864294	1	8,95	8,95

5851166	2868027	2017-03-31	1866427	1	9,95	9,95
5851166	2868027	2017-03-31	1866917	1	4,45	4,45
5851166	2868027	2017-03-31	1866977	1	4,45	4,45
5851166	2868027	2017-03-31	1867612	1	4,95	4,95
5851166	2868027	2017-03-31	1868200	1	5,95	5,95
6072940	2944164	2017-04-24	1774814	1	14,95	14,95
6072940	2944164	2017-04-24	1816562	1	35,95	35,95
6072940	2944164	2017-04-24	1885462	1	24,95	24,95
7027877	888428	2017-09-21	1864324	1	6,95	6,95
7027877	888428	2017-09-21	1939724	1	5,95	5,95
7027877	888428	2017-09-21	1942323	1	32,95	32,95
7027877	888428	2017-09-21	1953150	1	12,95	12,95
7211004	357608	2017-10-24	1823097	1	6,95	6,95
7211004	357608	2017-10-24	1965451	1	12,95	12,95
7211004	357608	2017-10-24	1974115	1	9,95	9,95
7211004	357608	2017-10-24	1990143	1	6,95	6,95
7211004	357608	2017-10-24	1990236	1	14,95	14,95
7214968	2909858	2017-10-24	1853591	1	3,95	3,95
7214968	2909858	2017-10-24	1947391	1	14,95	14,95
7214968	2909858	2017-10-24	1959162	1	12,95	12,95
7214968	2909858	2017-10-24	1973343	1	4,95	4,95
7214968	2909858	2017-10-24	1992685	1	9,95	9,95
7305907	2755259	2017-09-11	1782752	1	14,95	14,95
7305907	2755259	2017-09-11	1802965	1	4,45	4,45
7305907	2755259	2017-09-11	1953668	1	12,95	12,95
7732024	3482370	2018-03-01	2010866	1	16,95	16,95
7732024	3482370	2018-03-01	2021415	1	19,95	19,95
7732024	3482370	2018-03-01	2025260	1	16,95	16,95
7728331	3482091	2018-03-01	1968381	1	19,95	19,95
7728331	3482091	2018-03-01	1984658	1	19,95	19,95
7728331	3482091	2018-03-01	1981833	1	16,95	16,95
7728331	3482091	2018-03-01	1981841	1	9,95	9,95
7734346	3198795	2018-03-01	2029424	1	3,55	3,55
7734346	3198795	2018-03-01	2013353	1	3,55	3,55
7734346	3198795	2018-03-01	2015978	1	11,95	11,95
7734346	3198795	2018-03-01	1984795	1	12,95	12,95
7734346	3198795	2018-03-01	1969901	1	11,95	11,95
7734346	3198795	2018-03-01	1969906	1	11,95	11,95
7734346	3198795	2018-03-01	1973829	1	11,95	11,95
7762796	2123553	2018-08-01	1955858	1	6,95	6,95
7762796	2123553	2018-08-01	1972096	1	12,95	12,95
7762796	2123553	2018-08-01	1989813	1	12,95	12,95
7762796	2123553	2018-08-01	2022592	1	9,95	9,95
7762796	2123553	2018-08-01	2027241	1	16,95	16,95
7791096	3484782	2018-01-13	1987421	1	9,95	9,95
7791096	3484782	2018-01-13	1991327	1	7,95	7,95

7813730	3495564	2018-01-16	1993566	1	19,95	19,95
7815519	2295602	2018-01-16	1994450	1	7,95	7,95
7815519	2295602	2018-01-16	2033628	1	6,95	6,95
7815519	2295602	2018-01-16	2034813	1	6,95	6,95
7815519	2295602	2018-01-16	2036051	1	6,95	6,95
7815519	2295602	2018-01-16	2036053	1	6,95	6,95
7813730	3495564	2018-01-16	1992397	2	12,95	25,9
7813730	3495564	2018-01-16	2023836	7	22,95	160,65
7815519	2295602	2018-01-16	2036057	3	6,95	20,85
7842272	2236656	2018-01-21	2030592	1	3,95	3,95
7842272	2236656	2018-01-21	2030635	1	4,95	4,95
7842272	2236656	2018-01-21	2033753	1	3,95	3,95
7842272	2236656	2018-01-21	2033764	1	4,95	4,95
7842272	2236656	2018-01-21	2034588	1	5,95	5,95
7842272	2236656	2018-01-21	2035167	1	5,95	5,95
7842272	2236656	2018-01-21	2035178	1	4,95	4,95
7842272	2236656	2018-01-21	2035189	1	4,95	4,95
7842272	2236656	2018-01-21	2038656	1	4,95	4,95
7842272	2236656	2018-01-21	2039881	1	5,95	5,95
7842272	2236656	2018-01-21	2041184	1	4,95	4,95
7842272	2236656	2018-01-21	2041580	1	4,95	4,95
7844423	3110499	2018-01-21	1770890	1	16,95	16,95

RFM TABLOSU VERİ SETİ ÖRNEKLEMİ

MusteriNo	Güncellik	Islem_Adedi	Tutar
1011679	117	1	53,5
1030597	219	1	81,7
1149513	28	1	73,8
1214440	252	1	59,75
123482	90	1	65,85
1330031	289	1	182,45
1360570	545	1	79,7
1370387	199	1	61,7
1407333	303	1	81,65
1433012	260	1	80,65
1463737	524	2	231,8
1484115	57	2	250,5
1493335	559	1	69,8
1507651	292	1	50,75
1510143	252	2	211,15
1524932	321	1	147,5
1528284	805	2	85,5
1550173	750	1	29,95
1599491	220	1	34,75
1691865	277	1	43,75

172464	561	5	309,95
1742304	505	8	443,5
1771850	322	1	50,8
1821926	729	1	8,95
1881392	140	2	91,7
1932878	226	1	108,45
1979908	149	5	270,8
2013642	681	1	38,8
2029800	663	2	261,65
2069158	858	1	37,75
2111929	152	2	115,3
2111936	823	1	107,05
2121960	346	2	243,05
2123553	153	3	356,4
2126662	997	1	113,75
2133567	509	1	73,8
2138841	315	1	56,85
2158954	120	2	174,45
2167360	318	2	165,3
2174957	592	1	51,75
2178201	293	3	216,8
2206696	538	1	66,75
2208335	270	1	137,1
2209933	864	1	105,45
2216191	222	2	98,45
2223582	140	1	15,9
2224954	151	2	110,55
2225796	302	1	75,1
2236445	85	1	100,8
2236656	92	6	481,8
2240836	853	1	21,8
2244260	309	2	199,4
2244384	59	5	1361,6
2246658	1028	1	83,8
2247705	30	1	51,8
2251167	87	2	193,3
2261234	906	1	72,75
2268336	726	1	8,95
2268810	287	1	70,35
2274062	86	2	106,35
2280670	836	1	44,85
2283232	156	3	399,2
2286400	833	1	35,95
2295602	347	2	133,15
2296800	829	1	36,75
2297495	587	1	39,05

2297885	313	1	114,5
2307530	276	1	56,8
2307568	386	1	54,8
2314922	145	3	112,95
2316848	1027	1	61,75
2320469	381	1	113,5
2320774	732	2	117,55
2324383	151	1	57,8
2326427	601	1	368,1
2326720	328	1	75,65
2334938	782	1	41,7
2349256	765	1	82,65
2350952	805	1	154,6
2351287	258	1	187,15
2367571	212	2	575,15
2373558	417	2	285,35
2376575	284	1	91,6
2386902	377	1	98,65
2406608	904	1	85,9
2418078	125	1	88,65
2426029	777	1	133,3
2426204	777	1	58,8
2438847	293	2	164,3
2439986	773	1	47,85
2440067	31	4	1897,65
2440581	295	3	284,65
2442917	88	4	532,55
2450686	369	1	56,85
2455191	768	1	58,75
2473009	331	4	162
2478473	221	1	22,9
2481123	288	1	65,7
2486673	1025	1	79,8
2489750	302	1	54,85
2493940	964	1	59,75
2496284	933	1	54,75
2508500	31	3	267,95
2514734	407	9	187,55
2555908	426	2	74,8
2565035	738	1	49,8
2580528	733	1	207,35
2580549	733	1	89,6
2591115	730	1	49,85
2596143	242	1	75,05
2606773	240	1	34,05
2614761	721	1	49,85

2617484	457	3	157,9
2621915	156	2	151,3
2622203	387	2	80,65
2637662	714	1	121,4
2640105	714	1	97,85
2651773	711	1	116,95
2694037	339	1	54,55
2707541	431	2	340,75
2740979	681	1	128,6
2745606	149	1	75,1
2755259	135	6	208,45
2789931	159	2	320,05
2822362	180	1	58,45
2823633	90	1	55,75
2825910	156	1	92,35
2846339	647	1	45,75
2850578	362	1	51,65
2854290	91	2	145,3
2859226	643	1	44,85
2860672	643	1	55,85
2868027	641	1	77
2875366	148	1	128,15
2909668	219	2	128,65
2909858	180	4	194,9
2912953	422	2	198,2
2926484	169	1	4,95
2935716	620	1	54,8
2938330	618	1	67,8
2944164	231	3	132,85
2947289	307	1	50,7
2952521	614	1	34,9
2955650	611	1	61,8
2970969	367	1	56,85
2972587	606	1	50,85
2975301	333	1	51,8
2980182	594	1	13,9
2981609	514	1	37,9
3011769	591	1	75,75
3013958	248	1	51,75
3014624	211	2	280,1
3017373	251	1	51,7
3019339	275	1	51,75
3025741	270	1	88,6
3029556	225	1	65,7
3031295	149	2	406,45
3031842	580	1	54,55

3039228	151	1	75,25
3046207	605	1	183,85
3091266	215	1	24,9
3092735	28	3	270,1
3097666	392	3	171,7
3110499	344	1	121,55
3113173	305	2	254,9
3114896	120	2	296,15
3120080	225	1	176,2
3154074	529	1	150,6
3156195	216	4	331,8
3158913	217	3	246,1
3164932	376	1	94,65
3165732	135	1	20,85
3175074	514	1	53,8
3175208	633	1	47,85
3198795	306	1	67,85
3202181	165	1	75,6
3202678	497	1	82,7
3207834	493	1	21,9
3208116	275	1	203,05
3209613	224	1	116,45
3212462	322	1	140
3215270	293	1	71,8
3215368	304	1	58,8
3217482	488	1	57,85
3221303	388	1	238,85
3221906	602	1	77,2
3232776	473	1	76,8
3236112	318	1	65,75
3237061	374	1	52,9
3239983	468	1	16,95
3244083	465	1	144,1
3264609	171	1	60,7
3269616	151	1	9,95
3273915	440	2	42,9
3288436	290	2	187,1
3297919	152	1	90,2
3313532	55	3	115
3314864	328	1	250,95
3319741	209	1	61,7
3322326	329	2	127,5
3322964	232	1	53,9
3327400	222	6	304,75
3327699	255	2	167,5
3336995	270	1	134,6

3348338	223	2	117,7
3349362	259	2	89,55
3352089	414	1	239,1
3358088	370	1	64,85
3361951	410	1	51,8
3377975	357	1	58,7
3389110	402	1	125,5
3390180	402	1	58,7
3402750	337	1	110,5
3405344	321	2	149,75
3407620	385	1	22,95
3408848	287	1	23,85
3432030	384	1	29,9
3434198	383	1	73,8
3437714	190	2	123,45
3440309	380	1	74,7
3442435	379	1	55,25
3442493	270	1	68,7
3453565	376	1	50,8
3463641	372	1	32,85
3472441	318	2	228,35
3479343	365	1	19,95
3482091	306	1	66,8
3482370	306	1	53,85
3484118	275	1	52,45
3484782	275	2	39,8
3485287	275	1	72,8
3493678	122	1	52,85
3495564	349	1	206,5
3496596	89	1	13,9
3499201	61	1	117,75
3499470	354	1	51,9
3505482	352	1	71,75
3506268	351	1	56,75
3516595	60	1	113,2
3535308	305	1	34,95
3540574	333	1	201,1
3544355	274	1	25,9
3567028	320	1	25,8
357608	434	2	109,55
3583457	310	1	87,75
3586164	273	1	26,95
3589537	307	1	33,8
3597468	273	1	31,9
3600023	212	1	54,2
3609939	254	1	64,6

3614309	59	1	120,65
3619727	294	1	76,05
3636986	287	1	150,45
3645674	284	2	163,3
3653290	282	1	14,95
3662899	87	1	53,75
3665593	279	1	54,9
3669604	119	1	132,45
3670986	276	1	223,85
3673555	114	1	117,4
3676921	331	1	45,85
3681697	181	1	67,7
3681864	303	1	40,85
3682145	224	1	60,75
3683691	117	2	281,65
3685201	258	1	24,95
3685891	263	1	89,6
3686183	242	1	93,75
3687783	246	1	52,7
3688799	263	1	29,95
3688823	253	1	100,5
3691449	259	1	21,9
3697288	119	2	110,7
3698496	263	1	147,35
3699894	119	1	50,05
3700241	179	1	63,6
3700596	89	1	62,85
3702981	58	4	432,55
3703827	260	1	56,8
3703888	58	1	240,3
3704790	58	2	120,8
3705920	58	1	48,85
3706499	28	1	37,8
3706869	260	1	314,65
3709633	222	1	53,6
3712825	263	1	59,75
3716018	261	1	55,6
3716609	261	1	76,8
3720750	260	1	52,55
3721875	260	1	51,85
3723747	259	1	139,4
3727532	219	1	59,65
3728469	209	1	132,4
3729031	256	1	73,5
3729877	253	2	108,3
3731866	256	1	56,75

3732166	256	1	63,95
3732232	115	1	79,65
3732522	256	1	59,9
3734802	254	1	74,6
3735545	88	1	82,65
3736532	300	1	50,7
3740477	252	1	39,85
3744388	250	1	99,8
3748241	249	1	56,05
3750471	220	1	68,95
3753205	245	1	50,75
3755104	330	1	143
3756220	302	1	31,85
3757373	302	1	89,65
3757399	330	1	75,7
3757835	241	1	138,4
3759803	302	1	82,7
3760394	271	2	217
3760694	302	1	128,25
3760732	302	1	77,55
3761061	241	1	53,45
3761702	271	1	69,6
3767722	88	2	206,45
3768876	180	1	78,05
3771291	149	1	19,95
3771922	118	2	13,9
3773921	230	2	122,9
3776179	88	3	281,7
3779930	27	1	150,2
3781422	231	1	118,6
3782182	233	1	51,8
3782350	232	1	68,65
3782387	232	1	52,55
3782429	231	1	50,85
3783239	232	1	12,95
3783659	232	1	64,6
3783731	231	1	70,7
3785029	231	1	61,25
3786785	228	1	51,8
3793367	226	1	46,75
3796627	225	1	108,45
3796735	225	2	166,4
3797240	225	1	40,85
3797730	224	1	59,8
3798712	224	1	113,3
3799023	223	1	62,65

3800757	223	1	80,45
3801406	222	1	61,8
3801712	223	1	52,25
3801978	221	1	135,45
3802142	221	1	18,9
3805892	220	1	66,6
3816639	218	1	22,95
3816934	219	1	111,45
3817129	301	1	54,7
3817468	219	1	52,8
3818594	216	1	19,95
3819267	215	2	33,9
3821768	216	1	55,8
3826368	216	1	88
3827873	215	1	31,85
3830934	213	1	87,65
3832444	360	1	107,4
3832905	329	1	75,6
3833880	240	1	207,1
3834845	211	1	66,2
3836101	301	1	73,8
3836149	301	1	55,65
3836635	301	1	63,25
3836655	117	1	65,7
3836873	270	2	71,55
3842378	270	1	124,5
3843463	240	1	58,9
3843818	240	1	59,85
3845108	209	1	17,9
3846336	184	1	63,75
3846698	208	1	59,15
3849119	209	1	150,4
3850101	179	1	27,85
3850993	26	1	10,95
3852857	179	1	52,85
3853255	148	1	76,75
3854379	148	1	75,8
3855343	148	1	64,25
3855804	87	1	105,6
3857299	117	1	69,8
3859685	87	1	80,5
3863517	87	1	28,8
3867698	202	1	61,8
3874549	200	1	53,9
3878125	198	1	64,7
3879449	197	1	57,75

3881211	197	1	70,55
3882187	196	1	80,7
3882420	196	1	60,8
3884891	195	1	54,6
3894053	191	1	59,35
3894144	191	1	165
3894874	191	1	59,8
3895249	190	1	54,75
3895648	190	1	114,55
3897522	188	1	45,8
3900092	185	2	33,9
3900532	188	1	167,4
3902418	188	1	53,9
3906544	186	1	50,85
3913287	269	2	157,4
3917434	269	1	57,8
3918800	172	1	50,6
3920174	239	1	90,5
3921342	208	1	117,25
3925177	163	1	102,65
3926515	147	1	74,75
3930728	86	1	49,9
3933530	25	2	42,85
3938214	163	2	538,25
3939419	130	3	77,5
3942475	168	3	158,25
3942985	168	1	87,65
3951397	165	1	128
3952347	164	1	60,7
3954260	164	1	31,8
3956412	162	1	28,85
3958644	161	1	74,7
3960469	160	1	75,75
3964064	159	1	49,75
3969990	134	1	50,8
3971039	154	1	65,6
3971826	85	2	144,35
3971928	155	1	32,85
3976449	358	2	113,5
3981696	268	1	34,9
3982950	238	1	91,2
3983433	238	1	57,75
3990716	146	1	56,75
3992988	115	1	93,4
4006108	140	1	55,8
4008443	139	1	39,15

4009791	138	1	36,9
4018129	135	1	91,8
4030479	125	1	27,75
4039527	357	1	211,2
4042662	326	1	148,4
4051275	206	1	55,8
4059203	112	1	107,55
641832	633	1	92,6
748356	241	2	132,2
772191	267	3	225,15
79304	342	1	52,9
828080	371	1	120,5
847948	257	1	76,75
856703	188	6	1197,75
888428	467	1	58,8
947536	152	1	52,7

KAYNAKLAR

AYTEKİN, Hasan Tahsin. MAKİNE ÖĞRENİMİNİN ARAŞTIRMACILARIN VERİ ANALİZİ BAĞLAMINDA POTANSİYEL ÖNEMİ. *Ufuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 2021, 10.19: 85-106.

BAŞKOL, M. (2020). DETERMINING CUSTOMER SEGMENTATION BY USING RFM AND CORRESPONDENCE ANALYSIS. *Business & Management Studies: An International Journal*, 8(4), 902.

CAO, Shengjia, et al. Research on Python data visualization technology. In: *Journal of Physics: Conference Series*. IOP Publishing, 2021. p. 012122.

ERDEN, Caner. *PYTHON İLE VERİ MADENCİLİĞİ*. KODLAB YAYIN DAĞITIM YAZILIM LTD. ŞTİ., 2021.

KOHAVI, Ron; PAREKH, Rajesh. Visualizing RFM segmentation. In: *Proceedings of the 2004 SIAM international conference on data mining*. Society for Industrial and Applied Mathematics, 2004. p. 391-399.

KURTARAN, Gizem Duygu. RFM Analizi (Recently Frequence Monetary Value, Güncellik Sıklık Para).

MORGENTHALER, Stephan. Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2009, 1.1: 33-44.

Oban, Volkan. “K means clustering with R ”

https://www.slideshare.net/VolkanOBANMsc/kmeans-clustering-and-custergram-with-r?from_search=0 [Erişim 08.07.2023].

SABUNCU, İbrahim; TÜRKAN, Edanur; POLAT, Hilal. Customer segmentation and profiling with RFM analysis. *Turkish Journal of Marketing*, 2020, 5.1: 22-36.

SÖNMEZ, Oğuzhan; ZENGİN, Kenan. Yiyecek ve içecek işletmelerinde talep tahmini: yapay sinir ağları ve regresyon yöntemleriyle bir karşılaştırma. *Avrupa Bilim ve Teknoloji Dergisi*, 2019, S. 302-308.

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018, April). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP conference series: materials science and engineering* (Vol. 336, p. 012017). IOP Publishing.

Tsiptsis, K. K., & Chorianopoulos, A. (2011). *Data mining techniques in CRM: inside customer segmentation*. John Wiley & Sons.

Tukey, J. W. (1977). *Exploratory data analysis* (Vol. 2, pp. 131-160).

ZIAFAT, Hasan; SHAKERI, Majid. Using data mining techniques in customer segmentation. *Journal of Engineering Research and Applications*, 2014, 4.9: 70-79.

Araştırma Verileri Yönetimi Eğitim Portalı <https://acikveri.ulakbim.gov.tr/acik-veri-acik-bilim/bolum-3-veri-isleme/3-3-veri-kontrolu-ve-veri-temizleme/> [Erişim 15.05.23].

<https://towardsdatascience.com> [Erişim 15.05.23].

<https://scikit-learn.org/> [Erişim 16.10.22].

<https://neptune.ai/blog/k-means-clustering> [Erişim 08.07.2023].

ÖZGEÇMİŞ VE İLETİŞİM BİLGİSİ

Adı Soyadı : Havva Gülsüm Erarslan

Doğum Yeri : Tokat/Niksar

Doğum Tarihi : 30.09.1994

Yabancı Dili : İngilizce

E mail : hgulsume@gmail.com

Öğrenim Durumu

Lise : Eyüp Anadolu Lisesi (2008 - 2012)

Lisans : Sakarya Üniversitesi Bilgisayar ve Bilişim Bilimleri Fakültesi Bilgisayar Mühendisliği Bölümü (2013 - 2017)

Çalıştığı Kurumlar

Waylay NV (Belçika, 2017 - 2018)

İş Bankası Softtech A.Ş. (İstanbul, 2018 - 2021)

Siemens AG (İstanbul, 2021 - Halen)