

<https://dergipark.org.tr/bemarej>

Research Article

Müşteri segmentasyonu ve davranış analizi: Random forest algoritması kullanılarak gelir ve harcama davranışlarının incelenmesi

Bilge Doğanlı¹

ÖZET

Bu çalışma, bir müşteri veri setinin kullanımı ile müşteri segmentasyonu ve davranış analizi yapılarak gerçekleştirilmiştir. Veri seti, 1000 müşteriden oluşmakta ve 9 farklı değişken içermektedir. Çalışma ile Random Forest algoritması kullanılarak; gelir, harcama skoru, üyelik süresi gibi özelliklerin müşteri davranışlarını nasıl etkilediği araştırılmıştır. Gerçekleştirilen özellik önemi analizi ile gelirin ve satın alma sıklığının, müşteri davranışlarının tahmin edilmesinde en etkili faktörler olduğu, yaş, harcama skoru ve üyelik süresi değişkenlerinin ise daha düşük öneme sahip oldukları belirlenmiştir. Ayrıca, cinsiyet, tercih edilen kategori ve gelir dağılımı gibi demografik faktörlerin de; müşteri segmentasyonuna etkileri bulunmaktadır. Çalışma; müşteri değerlendirme ve pazarlama stratejilerinin geliştirilmesinde kullanılabilecek önemli içgörüler sunmaktadır. Segmentasyon analizi sonucunda, yüksek gelirli ve yüksek harcama gerçekleştiren müşteri gruplarına yönelik olarak özel stratejiler geliştirilmesi gerekliliği sonucuna ulaşılmıştır. Bu tür analizlerin, işletmelerin müşteri kitlelerini daha iyi anlamalarına ve stratejik kararlar almalarına yardımcı olabileceği düşünülmektedir.

Anahtar Kelimeler:

Müşteri Segmentasyonu, Random Forest Algoritması, Özellik Önemi Analizi, Gelir ve Harcama Davranışları, Korelasyon Analizi

Customer segmentation and behavior analysis: Examining income and spending behaviors using random forest

ABSTRACT

This study was conducted through customer segmentation and behavioral analysis using a customer dataset. The dataset consists of 1,000 customers and includes nine different variables. By utilizing the Random Forest algorithm, the study examines how features such as income, spending score, and membership duration influence customer behavior. The feature importance analysis revealed that income and purchase frequency are the most significant factors in predicting customer behavior, whereas age, spending score, and membership duration have relatively lower importance. Additionally, demographic factors such as gender, preferred category, and income distribution also have an impact on customer segmentation. This study provides valuable insights that can be utilized for customer evaluation and the development of marketing strategies. As a result of the segmentation analysis, it has been concluded that tailored strategies should be developed for high-income, high-spending customer groups. Such analyses are expected to help businesses better understand their customer base and make more strategic decisions.

Keywords: Customer Segmentation, Random Forest Algorithm, Feature Importance Analysis, Income and Spending Behaviors, Correlation Analysis

1. Giriş

Müşteri segmentasyonu ve davranış analizleri, işletmelerin müşteri tabanlarını daha iyi anlamalarını ve müşteri odaklı stratejiler geliştirmelerini sağlamaktadır (Doğanlı & Çelik, 2024). Müşteri davranışlarının ve segmentasyonunun anlaşılması, etkili pazarlama stratejileri ve iş stratejileri geliştirilmesi için çok önemlidir. Büyük veri kümelerini analiz etmek, işletmelerin farklı müşteri segmentlerini tanımlamasına ve kişiselleştirilmiş pazarlama

¹ Asst. Prof. Dr., Aydın Adnan Menderes University, Nazilli Faculty of Economics and Administration, Department of International Trade and Finance, Türkiye, bdoganli@adu.edu.tr, ORCID: 0000-0002-1985-0430

Academic Editor: Assoc. Prof. Dr. Engin ÇAKIR

Received: 25.02.2025

Acceptance: 07.03.2025

Published: 10.03.2025

Citation: Doğanlı, B. (2025). Müşteri segmentasyonu ve davranış analizi: Random forest algoritması kullanılarak gelir ve harcama davranışlarının incelenmesi. *Business, Economics and Management Research Journal*, 8(1), 52-66. <https://doi.org/10.58308/bemarej.1646966>



Copyright: ©2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives (CC-BY-NC-ND) license (<https://creativecommons.org/licenses/by-nc-nd/4.0>).

abalarına da olanak tanımaktadır (Madhiraju vd., 2024; Wong vd., 2024). Bu alıřma, demografik ve davranıřsal faktörlerin müşteri harcama alışkanlıkları üzerindeki etkilerini inceleyerek, müşteri segmentasyonunda etkili deđiřkenleri belirlemeyi amaçlamaktadır. Bu amaçla da; bir müşteri veri seti kullanılarak, müşteri gelir düzeyi, harcama skoru, yař ve üyelik süresi gibi demografik ve davranıřsal faktörlerin; müşterilerin harcama alışkanlıkları üzerindeki etkileri incelenmiştir. Random Forest algoritması kullanılarak, müşteri gelir düzeyi, yař, harcama skoru ve üyelik süresi gibi faktörlerin, müşterilerin harcama davranıřlarını nasıl etkilediđi analiz edilmiştir. Hangi özelliklerin müşterilerin davranıřlarını daha iyi tahmin edebildiđinin belirlenmesi de amaçlanmıştır. Bu alıřma, pazarlama stratejilerini belirlemeyi ve müşteri deđerlemesi için önemli bilgiler sunmayı hedeflemektedir. Arařtırmanın kapsamı, büyük veri analitiđi ve makine öğrenmesi tekniklerinin müşteri davranıř analizinde nasıl kullanılabileceđinin örneklerle açıklanması olarak özetlenebilir. Bu bağlamda elde edilen bulguların, iřletmelerin pazarlama stratejilerini optimize etmelerinde, müşteri memnuniyetini artırmalarında ve kârlılıklarını maksimize etmelerinde bir yol gösterici olabileceđi düşünölmektedir.

Teknolojik ilerlemeler ve dijital dönüşömlle birlikte, iřletmeler müşteri verilerini daha etkin bir şekilde toplama ve analiz etme imkânına kavuşmuşlardır. Ancak, büyük veri kümelerinin dođru şekilde analiz edilmemesi durumunda, bu veriler iřletmeler için anlamlı deđerler oluşturamazlar (Dođanlı & elik, 2024; Gürsakal & elik, 2021). Random Forest gibi güçlü makine öğrenmesi algoritmaları, büyük veri kümelerinden deđerli bilgiler elde edilmesini sađlayarak müşteri davranıřlarının daha dođru tahmin edilmelerine olanak tanımaktadır (Liu vd., 2024; Miao & Xu, 2024; Salman vd., 2024). Bu alıřma, özellikle perakende, finans ve hizmet sektörlerinde faaliyet gösteren iřletmeler için müşteri segmentasyonu ve davranıř analizleri konularında açıklayıcı özellikler taşımaktadır. Arařtırma sonuçlarının; detaylıca hedeflenmiş pazarlama kampanyalarının oluşturulması ve müşteri yařam boyu deđerinin (Customer Lifetime Value - CLV) artırılması için önemli stratejik girdiler sađlayacađı düşünölmektedir.

2. Kavramsal ereve

2.1. Literatür Taraması

Müşteri segmentasyonu, iřletmelerin farklı müşteri gruplarının, özel ihtiyalarını ve tercihlerini karřılamak için uyarlamalarına izin verdiđi pazarlama stratejilerinin temel bir yönüdür. Demografik ve davranıřsal verilerin entegrasyonu, segmentasyon modellerinin dođruluđunu önemli ölçüde artırarak müşteri memnuniyetini ve karlılıđını artırmaktadır. Makine öğrenimi teknikleri, özellikle kümeleme algoritmaları, bu segmentasyon modellerinin iyileştirilmesinde etkin rol oynamaktadır (Panda vd., 2024; Tripathi vd., 2024). Makine öğrenimi algoritmalarının müşteri segmentasyonu üzerindeki etkisi çok önemlidir, Random Forest algoritması, yüksek dođruluk payı ve deđerken önemini belirleme yeteneđi nedeniyle oldukça önemlidir. Bu; özellikle müşteri gelir düzeyinin ve harcama puanının segmentasyonda kritik faktörler olduđu, perakende sektöründe geçerlidir. Random Forest algoritmasının temel göstergeleri belirlemedeki etkinliđi belgelenmiştir ve bu da onu, müşteri segmentasyon stratejilerini geliřtirmeyi amaçlayan iřletmeler için deđerli bir araç haline getirmektedir (Chen, 2024; Cheng vd., 2024; Liu vd., 2024; Tripathi vd., 2024).

Cheng vd. (2024) alıřmalarında, Random Forest dâhil olmak üzere, ağaç tabanlı algoritmaların e-ticarette müşteri deđerini sınıflandırmadaki etkinliđini tartışmaktadırlar. Bu alıřma, algoritmaların, mevcut ve potansiyel müşteri deđerlerini, ayrıntılı sınıflandırma kriterleri aracılıđıyla entegre edebileceklerini vurgulamaktadır. Gelir seviyesi ve harcama puanı gibi müşteri segmentasyon faktörleri özel olarak ele alınmasa da; veri madenciliđi, verimliliđini artırma ve müşteri yönetiminde akıllı kararlar verilmesinin desteklenmesi için Random Forest gibi algoritmaların kullanılmasının önemi vurgulanmaktadır.

Sumarlin ve Qosidah (2024) alıřmalarında, makine öğrenimi algoritmalarının müşteri segmentasyonu üzerindeki etkilerini veya bu bağlamda müşteri gelir düzeyinin ve harcama puanının etkinliđini özellikle ele almaktadırlar. Üretim şirketlerinde satış performansını tahmin etmede ve envanter yönetiminde makine öğreniminin, özellikle sinir ağlarının ve rastgele ormanların uygulanmasına odaklanmışlardır. Satış tahmin dođruluđu ve operasyonel verimlilikteki geliřmeler vurgulanmış, veri kalitesi ve sistem entegrasyonunun zorlukları da tartışılmıştır.

Dodda vd. (2024) alıřmalarında, özellikle Random Forest modelinin dođruluđuna vurguda bulunarak ve makine öğrenimi algoritmalarını kullanarak müşteri kaybını tahmin etmeye odaklanmışlardır. alıřma; sıralı modelin, zamansal bađımlılıkları yakalama yeteneđini vurgularken, Random Forest ve Karar Ağacı Sınıflandırıcıları yorumlanabilirlikleri ve basitlikleri ile de dikkat ekmektedir.

Khalid vd. (2024) alıřmalarında, Random Forest sınıflandırıcısının müşteri segmentasyonunda, diđerlerinden daha iyi bir performans gösterdiğini ve 0,96'lık bir doğruluk payı elde ettiğini vurgulamaktadırlar. Arařtırma, havacılık endüstrisindeki stratejik müşteri segmentasyonunda, makine öğreniminin uygulanmasına odaklanmaktadır ve özellikle memnun müşterileri etkin bir şekilde belirleyerek ve bölümlere ayırarak, müşteri tutma oranını artırmak için Havayolu Yolcu veri kümesini kullanmaktadır.

Deniz ve Bülbül (2024) alıřmalarında, topluluk modellerinin, özellikle Random Forest, müşteri satın alma davranışını tahmin etmede, yüksek doğruluk ve ROC (Receiver Operating Characteristic) AUC (Area Under the Curve) göstermede mükemmel olduğunu göstermektedirler. Müşteri segmentasyonunu özel olarak ele almasalar da, müşteri eğilimlerini anlamada yıllık gelir gibi demografik faktörlerin önemini vurgulamaktadırlar. Arařtırma, makine öğrenimi modellerinin pazarlama stratejilerini geliřtirmedeki etkinliğine vurguda bulunarak, bu tür modellerden elde edilen içgörülerin, perakende sektöründeki segmentasyon abalarını bilgilendirebileceğini öne sürmektedir.

Bu bağlamda, alıřma literatürdeki bilgileri genişleterek ve Random Forest algoritmasını kullanarak, müşteri davranışlarının daha doğru tahmin edilmesini ve segmentasyon modellerinin etkinliğinin artırılmasını hedeflemektedir.

ROC (Receiver Operating Characteristic) Eğrisi: Makine öğrenimi modellerinin sınıflandırma başarısını deđerlendirmek için kullanılan bir grafikdir. Y-ekseninde True Positive Rate (Duyarlılık, Sensitivity), X ekseninde ise False Positive Rate (1- Özgüllük, Specificity) yer alır. ROC eğrisi, modelin pozitif ve negatif sınıfları ne kadar iyi ayırt edebildiğini gösterir. **AUC (Area Under the Curve - Eğri Altındaki Alan):** ROC eğrisinin altındaki alanı ifade eder. AUC deđerı 0.5 ile 1 arasında deđerışir. 1'e yakın bir AUC, modelin yüksek doğrulukla tahmin yaptığını gösterir. 0.5'e yakın bir AUC, modelin rastgele tahmin yaptığını gösterir. ROC AUC ne kadar yüksekse, modelin müşteri satın alma davranışını doğru tahmin etme gücü o kadar iyidir. Random Forest gibi topluluk modellerinin, müşteri satın alma davranışını tahmin etmede yüksek doğruluk göstermesi ve ROC AUC deđerlerinin yüksek olması, modelin güvenilir ve etkili olduğunu göstermektedir (Alaca, 2023).

Gerçekleřtirilen bu arařtırma, řletmelerin müşteri tabanını daha iyi anlamalarına, stratejik kararlar almalarına ve hedeflenmiş pazarlama kampanyaları geliřtirmelerine rehberlik etmeyi hedeflemektedir. Elde edilen bulgular, veri odaklı karar verme süreçlerini destekleyerek řletmelerin daha sürdürülebilir ve verimli stratejiler oluřturmasına katkı sađlayacaktır diye de düşünülebilir.

2.2. Veri Seti

Bu alıřmada, <https://www.kaggle.com/datasets/fahmidachowdhury/customer-segmentation-data-for-marketing-analysis> adresinden alınan "digital_marketing_campaign_dataset" veri seti kullanılmıřtır. Veri seti toplamda 1000 müşteriye ait verileri içermektedir. Bu veri setinde 9 sütun bulunmaktadır: id, age, gender, income, spending_score, membership_years, purchase_frequency, preferred_category, last_purchase_amount. Tüm sütunlar eksiksiz veriye sahiptir.

Veri setindeki deđerışkenler ve özellikleri ařađıdaki gibidir:

Tablo 1. Veri setindeki Deđerışkenler ve Özellikleri

id: Her müşteri için benzersiz bir kimlik numarası.
age: Müşterinin yaşı (18-69 yaş aralığında).
gender: Müşterinin cinsiyeti (Erkek, Kadın, Diđer).
income: Yıllık gelir (USD cinsinden, ortalama: 88,500 USD, min: 30,004 USD, max: 149,973 USD).
spending_score: Harcama davranışını ve sadakatini belirten bir skor (1-100 arası, ortalama: 50.7).
membership_years: Müşterinin üyelik süresi (1-10 yıl arası, ortalama: 5.5 yıl).
purchase_frequency: Son bir yılda yapılan alışveriş sayısı (ortalama: 26.6 alışveriş/yıl).
preferred_category: Müşterinin en çok alışveriş yaptığı kategori (Electronics, Clothing, Groceries, Home & Garden, Sports).
last_purchase_amount: Son alışverişte harcanan tutar (ortalama: 492.35 USD, max: 999.74 USD).

alıřmada kullanılan veri seti ve Python kodları açık kaynaklı GitHub- <https://github.com/BilgeDoganli09/M-teri-Segmentasyonu-ve-Davran-Analizi>- üzerinden paylařılmıřtır.

3. Arařtırma Yöntemi

Bu çalışmanın analizi, veri seti üzerinde Random Forest algoritması kullanılarak gerçekleştirilmiştir. Random Forest, yüksek doğruluk oranı ve değişkenlerin önem düzeyini belirleyebilme yeteneđi sayesinde müşteri davranış analizi için ideal bir yöntem olarak öne çıkmaktadır.

Arařtırma süreci ařađıdaki adımlardan oluşmaktadır:

- Veri setinin hazırlanması ve temizlenmesi,
- Demografik ve davranışsal değişkenlerin seçilmesi,
- Random Forest algoritması ile modelin eğitilmesi ve değerlendirilmesi,
- Deđişkenlerin müşteri davranışı üzerindeki etkilerinin analiz edilmesi.

Random Forest: Topluluk öğrenme (ensemble learning) yöntemine dayanan bir makine öğrenmesi algoritmasıdır. Random Forest algoritması, çoklu karar ağaçlarını toplayarak, tahmin doğruluđunu artıran sağlam bir topluluk öğrenme yöntemidir. Ařırı uyumu azalttıđı ve dengesiz verileri verimli bir şekilde işlediđi için hem sınıflandırma hem de regresyon görevlerinde özellikle etkilidir (Alfajr & Defiyanti, 2024; Dai vd., 2023; Kumar vd., 2024; Mrg & Hasibuan, 2024; Salman vd., 2024). Random Forest'ın önemli özelliklerinden birisi; her özelliđin, modelin tahminlerine katkısını anlamaya yardımcı olan özellik önemini hesaplama yeteneđidir. Bu, özelliklerin karar ağaçlarındaki bölünmeleri nasıl etkilediđi analiz edilerek ve Gini impurity veya entropi gibi ölçümler kullanılarak elde edilir (Ignatenko vd., 2024; Miao & Xu, 2024). Özelliklerin çok önemli olduđu durumlarda, modelin karar verme süreci daha hızlı ve doğru gerçekleşmektedir. Önemli özellikler, daha erken bölünmelerde yer almakta ve modelin tahmin gücünü artırmaktadırlar.

Özellik Önem Skorları: Random Forest, modellerindeki özellik önem puanları, her özelliđin modelin tahmin doğruluđuna katkısını anlamak için çok önemlidir. 0'dan 1'e kadar deđişen bu puanlar, hangi özelliklerin hedef deđişkeni önemli ölçüde etkilediđini belirlemeye yardımcı olmakta ve bu sayede de model verimliliđini ve doğruluđunu artırmaktadırlar. Random Forest algoritması, topluluk öğrenme tekniđi aracılıđıyla, sadece model güvenilirliđini geliřtirmekle kalmaz, aynı zamanda temel özellikleri vurgulayarak stratejik kararlar verilmesine de yardımcı olur. Bu süreç, sonuçları etkileyen en etkili faktörlere ilişkin içgörüler sağladığından, spor veri analizlerinden, iklim etki çalışmalarına kadar çeřitli alanlarda kullanılabilirler (Li & Mu, 2024; Miao & Xu, 2024). Örneđin, gelir, harcama skoru gibi deđişkenlerin yüksek önemli skorlar alması, bu faktörlerin müşteri segmentasyonu veya başka bir analizde güçlü etkilere sahip olduđunu göstermektedir.

3.1. Arařtırmanın Amacı, Sorusu ve Kapsamı

Bu çalışmanın temel amacı; demografik ve davranışsal faktörlerin müşteri harcama alışkanlıkları üzerindeki etkilerinin incelenerek, müşteri segmentasyonunda etkili deđişkenlerin belirlenmesidir. Bu çalışmanın amacı dođrultusunda geliřtirilen arařtırma soruları da ařađıdadır:

Müşteri gelir düzeyi, harcama skoru, yař ve üyelik süresi gibi demografik ve davranışsal faktörlerin; müşterilerin harcama alışkanlıkları üzerindeki etkileri nelerdir? Hangi özellikler, müşterilerin davranışlarını daha iyi tahmin edebilmektedir? Pazarlama stratejileri nasıl belirlenebilir ve müşteri deđerlemeleri için önemli bilgiler nelerdir?

Arařtırmanın kapsamı; büyük veri analitiđi ve makine öğrenmesi tekniklerinin müşteri davranışlarının analizinde nasıl kullanılabileceđinin örneklerle açıklanması olarak özetlenebilir. Bu bağlamda elde edilen bulguların, işletmelerin pazarlama stratejilerini optimize etmelerinde, müşteri memnuniyetini artırmalarında ve kârlılıklarını maksimize etmelerinde bir yol gösterici olabileceđi düşünülebilir.

4. Bulgular

Tanımlayıcı istatistikler, veriyi anlamak, özetlemek ve karar verme süreçlerini desteklemek için kritik öneme sahiptirler. Ham verilerden doğrudan anlam çıkarmak zor olabilir, bu yüzden bu teknikler veriyi daha anlamlı ve kullanışlı hale getirirler. Tanımlayıcı istatistikler ařađıdaki tablo da verilmiştir.

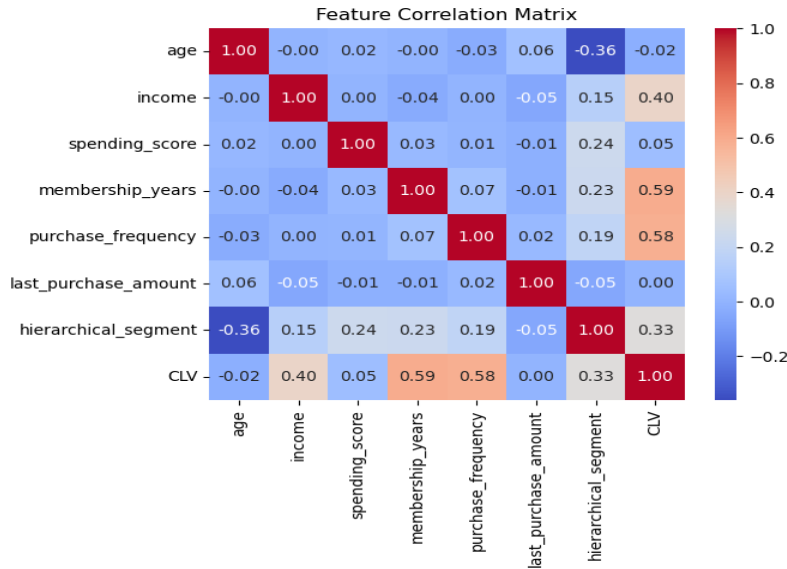
Tablo 2. Tanımlayıcı İstatistikler

	age	income	spending_score	membership_years	purchase_frequency	last_purchase_amount	hierarchical_segment	CLV
count	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000	1000.000000e+03
mean	43.783000	88500.800000	50.685000	5.46900	26.596000	492.348670	1.21300	1.305510e+07
std	15.042213	34230.771122	28.955175	2.85573	14.243654	295.744253	0.986207	1.235574e+07
min	18.000000	30004.000000	1.000000	1.000000	1.000000	10.400000	0.000000	7.972100e+04
25%	30.000000	57911.750000	26.000000	3.00000	15.000000	218.762500	0.000000	3.911475e+06
50%	45.000000	87845.500000	50.000000	5.00000	27.000000	491.595000	1.000000	8.950122e+06
75%	57.000000	116110.250000	76.000000	8.0000	39.000000	747.170000	2.000000	1.911382e+07
max	69.000000	149973.000000	100.000000	10.00000	50.000000	999.740000	3.000000	6.724350e+07

Bu çalışma kapsamında incelenen müşteri veri seti, demografik ve davranışsal değişkenler aracılığıyla, müşteri segmentasyonuna yönelik önemli içgörüler sunmaktadır. Yaş dağılımı (ort: 43,8 ve std: 15,04) orta yaş grubunun baskın olduğunu gösterirken, gelir seviyeleri (ort: 88500 USD ve std: 34230 USD) müşteri kitlesi arasında önemli farklılıklar sergilemektedir. Harcama skoru (ort: 50,7 ve std: 28,96) ve satın alma sıklığı (ort: 26,6 ve std: 14,24) müşteri sadakati ve davranış farklılıklarını vurgulamaktadır. Üyelik süresi (ort: 5,5 yıl ve std: 2,86) müşteri bağlılığının analizinde kritik bir metriktir. Son alışveriş tutarı (ort: 492,35 USD ve std: 295,74) ve müşteri yaşam boyu değeri (ort: 13055,100 USD ve std: 12355,740 USD) gibi finansal göstergeler, müşteri segmentleri arasındaki değer farklılıklarını ortaya koymaktadır. Bu istatistiksel bulgular, hedef odaklı pazarlama stratejileri geliştirmek, müşteri sadakatini artırmak ve kaynakların etkin dağıtımını sağlamak için değerli bir temel oluşturmaktadır.

4.1. Korelasyon Analizi:

Özellikler arasındaki korelasyonları kontrol etmek, hangi değişkenlerin birbirine güçlü bir şekilde bağlı olduğunu görmek, sonuçları yorumlayabilmek için oldukça faydalıdır. Bu, gelecekteki segmentasyon veya modellemeler için anlamlı bağıntıları ortaya çıkarabilir. Aşağıda, Şekil 1’de değişkenlere ait korelasyon matrisi verilmiştir.



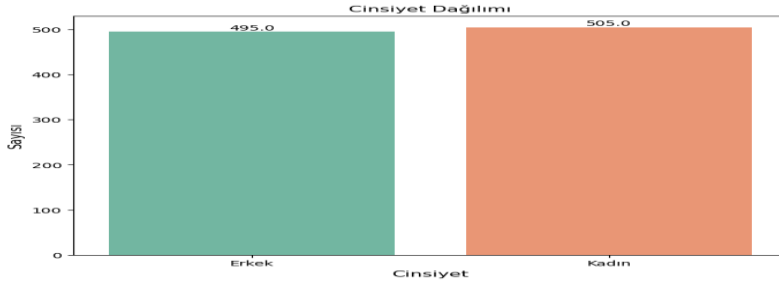
Şekil 1. Korelasyon Matrisi

Gerçekleştirilen korelasyon analizi, müşteri yaşam boyu değeri (CLV) ile üyelik süresi ($r=0,592$) ve satın alma sıklığı ($r=0,581$) arasında güçlü pozitif ilişkiler olduğunu ortaya koymaktadır. Bu durum, uzun süreli üyelik ve sık satın alma davranışlarının müşteri değerini önemli ölçüde artırdığını göstermektedir. Ayrıca, gelir düzeyi ($r=0,396$) ve hiyerarşik segment ($r=0,325$) değişkenleri CLV ile orta düzeyde pozitif ilişki sergilerken, yaş ($r=-0,024$) ve son alışveriş tutarı ($r=0,0009$) ile anlamlı bir ilişki gözlenmemiştir. Hiyerarşik segmentasyon, harcama skoru ($r=0,239$), üyelik süresi ($r=0,232$) ve satın alma sıklığı ($r=0,186$) ile pozitif korelasyon gösterirken, yaş ($r=-0,359$) ile negatif

ilişkilidir, bu da genç müşterilerin segmentasyon açısından daha değerli olabileceğine işaret etmektedir. Gelir, CLV üzerinde orta düzeyde bir etkiye sahipken, diğer değişkenlerle zayıf ilişkiler sergilemiştir. Elde edilen bu bulgular, müşteri değerlemesi ve segmentasyon stratejilerinde özellikle üyelik süresi, satın alma sıklığı ve gelir düzeyine odaklanmanın daha etkili sonuçlar sağlayabileceğini göstermektedir.

4.2. Cinsiyet Dağılımı Analizi

Cinsiyet dağılımı analizi, müşteri kitlesinin demografik yapısını anlamak ve cinsiyete dayalı farklı tüketim davranışlarını belirlemek amacıyla gerçekleştirilmiştir. Elde edilen dağılım, müşteri kitlesinde hangi cinsiyet grubunun baskın olduğunu ve bu grupların pazarlama stratejileri üzerindeki potansiyel etkilerini ortaya koymaktadır.

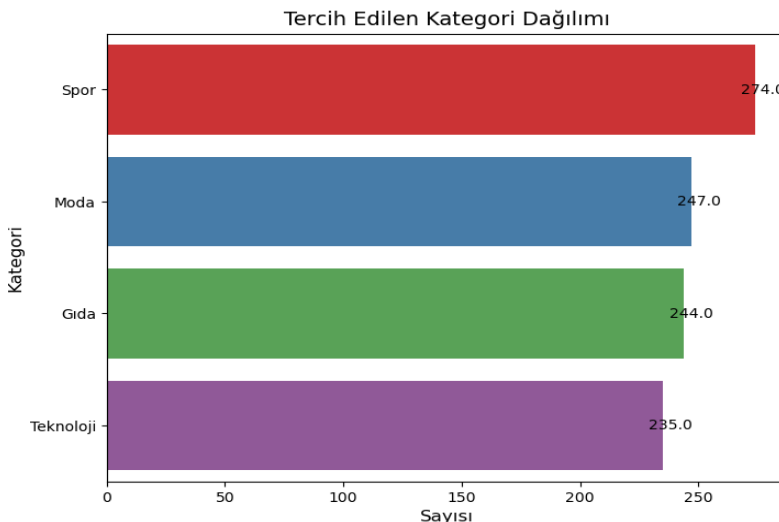


Şekil 2. Cinsiyet Dağılımı

Şekil 2’de gösterilmekte olan verilere göre, müşteri kitlesi %50,5 Kadın (505 müşteri) ve %49,5 Erkek (495 müşteri) olarak iki cinsiyet arasında oldukça dengeli bir dağılım göstermektedir. Bu durum, müşteri tabanının cinsiyete dayalı olarak homojen bir yapıya sahip olduğunu ve herhangi bir cinsiyet grubunun baskın olmadığını göstermektedir. Dengeli bir cinsiyet dağılımı, pazarlama stratejilerinin cinsiyete özgü segmentasyon yerine daha geniş ve kapsayıcı bir şekilde tasarlanmasına olanak tanımaktadır. Ancak, bu homojen yapı içerisinde belirli ürün kategorilerinin veya hizmetlerin cinsiyete göre farklı rağbet görüp görmediği daha derinlemesine incelenmelidir. Sonuç olarak, bu dağılım pazarlama kampanyalarının her iki cinsiyeti de eşit derecede hedeflenmesi gerektiğine işaret etmektedir.

4.3. Tercih Edilen Kategori Dağılımı Analizi

Tercih edilen kategori dağılımı analizi, müşterilerin alışveriş tercihlerindeki eğilimleri belirlemek ve popüler kategorilere yönelik stratejik kararlar almak amacıyla gerçekleştirilmektedir. Bu analiz, hangi ürün kategorilerinin müşteri talebi açısından öne çıktığını ve hangi kategorilerin daha fazla yatırım veya pazarlama faaliyetleri gerektirdiğini göstermektedir. Sonuçlar, belirli kategorilere olan talebin daha yüksek olduğunu ortaya koyduğunda, bu kategorilere yönelik kampanyalar ve stok yönetimi stratejileri önceliklendirilebilir olarak algılanabilir. Örneğin, elektronik ürünlerin diğer kategorilere kıyasla daha fazla tercih edildiği belirlenmişse, bu alanda promosyonların artırılması ve ürün çeşitliliğinin zenginleştirilmesi satışları önemli ölçüde artıracaktır.



Şekil 3. Tercih Edilen Kategoriler

Şekil 3'te gösterilmekte olan, müşterilerin tercih edilen kategori dağılımı incelendiğinde, en çok tercih edilen kategori Spor (%27,4 ile 274 müşteri) olarak öne çıkarken, onu Moda (%24,7 ile 247 müşteri), Gıda (%24,4 ile 244 müşteri) ve Teknoloji (%23,5 ile 235 müşteri) kategorileri takip etmektedir. Bu dengeli dağılım, işletmelerin her kategori için farklı stratejik yatırımlar yapmasını gerektirirken, özellikle yüksek talep gören kategorilere yönelik daha spesifik stratejiler geliştirilmesi gerektiğini ortaya koymaktadır.

Spor kategorisinin en çok tercih edilen kategori olması, müşterilerin spor giyim, ekipman ve sağlıklı yaşam trendlerine olan ilgisini yansıtmaktadır. Bu doğrultuda işletmelerin aşağıdaki stratejilere odaklanması faydalı olacaktır:

- **Stok Yönetimi:** Talebin yüksek olduğu ürün gruplarının belirlenmesi ve stok seviyelerinin dinamik olarak yönetilmesi gerekmektedir. Örneğin, spor giyim ve ayakkabı alt kategorileri yüksek satış hacmine sahipse, bu ürünlerde stok tükenmelerini önlemek için tedarik zinciri süreçleri güçlendirilmelidir.
- **Hedefli Kampanyalar:** Spor ile ilgili sezonluk kampanyalar (örneğin, yaz aylarında açık hava sporlarına yönelik ürünlerde indirimler, kış sezonunda iç mekân sporlarına yönelik kampanyalar) düzenlenebilir. Ayrıca, influencer ve sosyal medya iş birlikleriyle, spor trendlerine duyarlı müşteri segmentlerine ulaşılabilir.
- **Sadakat Programları:** Spor ürünlerine ilgi duyan müşterilerin tekrar alışveriş yapmalarını sağlamak adına sadakat programları oluşturulabilir. Örneğin, belirli bir harcama tutarını aşan müşterilere indirim kuponları veya özel üyelik avantajları sunulabilir.

Moda kategorisinin güçlü bir talebe sahip olması, tüketicilerin trendleri yakından takip ettiğini ve kişisel tarzlarını ön planda tuttuğunu göstermektedir. Bu doğrultuda, işletmelerin aşağıdaki stratejilere odaklanması faydalı olacaktır:

- **Sezonluk Koleksiyonlara Yatırım:** Moda sektörü, hızlı değişen trendlerden etkilendiği için, yeni sezon ürünlerinin önceden planlanarak piyasaya sürülmesi büyük önem taşımaktadır. Özellikle, mevsimsel değişimlere bağlı olarak stok yönetiminin dikkatlice yapılması gerekir.
- **Kişiselleştirilmiş Öneriler:** Müşteri alışveriş verilerine dayalı olarak kişiselleştirilmiş moda önerileri sunulabilir. Örneğin, geçmiş alışverişlerine göre müşterilere özel kombin önerileri veya indirimler sağlanarak müşteri sadakati artırılabilir.
- **Gıda kategorisinin yüksek talep görmesi, müşterilerin günlük ihtiyaçlarını bu kanaldan karşıladığını ve sağlıklı beslenme eğilimlerinin etkili olduğunu göstermektedir. İşletmelerin bu doğrultuda aşağıdaki stratejilere yönelmeleri faydalı olabilir:**
- **Taze ve Organik Ürünlere Yatırım:** Organik ve sağlıklı gıdalara olan talep arttığından, tedarik zincirinde bu ürünlere özel bir planlama yapılması rekabet avantajı sağlayabilir.
- **Hızlı Teslimat Seçenekleri:** Gıda alışverişlerinde hızlı teslimat büyük önem taşımaktadır. Aynı gün veya ertesi gün teslimat seçenekleri sunulursa müşteri memnuniyeti artırılabilir.

Teknoloji kategorisinin diğer kategorilere kıyasla nispeten daha düşük bir oranda yer alması, bu alandaki rekabetin yoğun olabileceğini ve talebin artırılması için yeni stratejiler geliştirilmesi gerektiğini göstermektedir. Bu doğrultuda, işletmelerin aşağıdaki stratejilere odaklanması faydalı olacaktır:

- **Ürün Çeşitliliğinin Artırılması:** Teknoloji kategorisinde öne çıkan alt segmentler belirlenerek (örneğin, akıllı saatler, kablosuz kulaklıklar vb.), müşteri taleplerine uygun yeni ürünler eklenebilir.
- **Taksit ve Finansal Çözümler:** Teknolojik ürünlerin fiyat seviyeleri yüksek olabileceğinden, müşterilere özel taksit seçenekleri veya finansman çözümleri sunulması satın almaları teşvik edilebilir.
- **Eğitici İçerikler ve Deneyim Alanları:** Teknolojik ürünler genellikle teknik özellikler açısından karmaşık olabilir. Bu nedenle, müşterilere yönelik eğitim videoları, incelemeler ve interaktif deneyim alanları oluşturularak güven artırılabilir.

Kategori bazında gerçekleştirilecek yatırımların işletmeye sağlayacağı katkılar aşağıdaki şekilde özetlenmiştir:

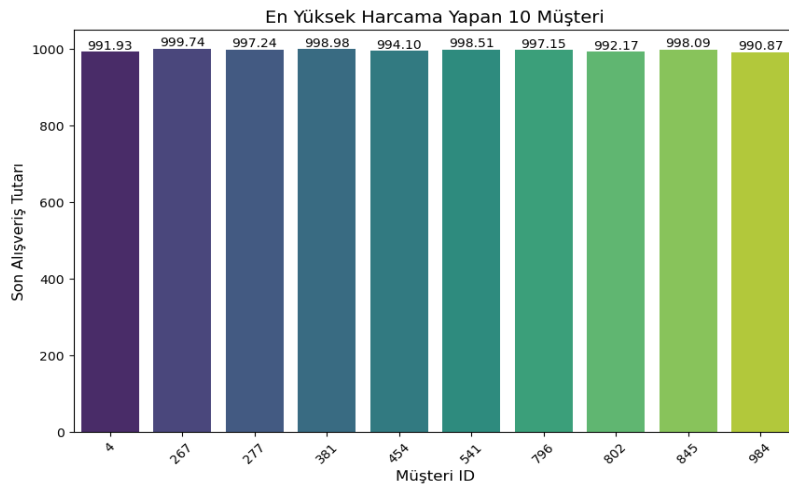
- Doğru stok yönetimi ile tedarik zinciri verimliliği artırılarak satış kayıpları önlenir.
- Hedefli pazarlama stratejileri sayesinde müşteri sadakati artırılarak uzun vadeli gelir elde edilebilir.

- Talepleri karşılayacak yenilikçi kampanyalar ile rekabet avantajı sağlanarak işletmenin pazar payı genişletilebilir.

Sonuç olarak, her kategoriye özgü stratejik yatırımlar gerçekleştirilerek müşteri eğilimleriyle uyumlu hareket eden işletmeler, hem müşteri memnuniyetini artırabilirler hem de sürdürülebilir büyüme sağlayabilirler.

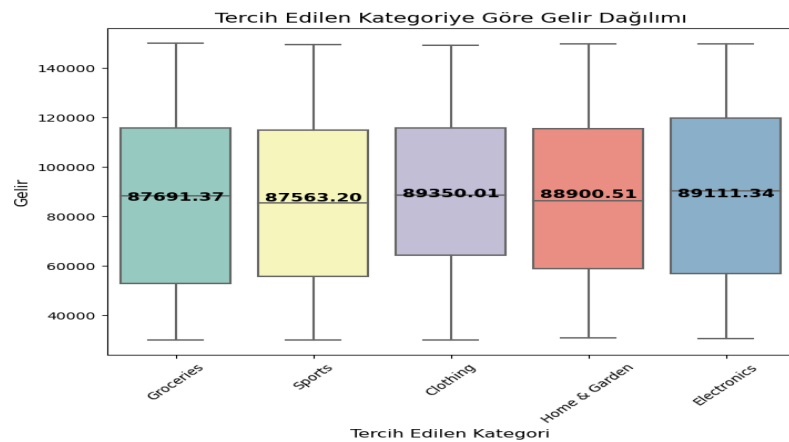
4.4. Müşterilere Özel Analizler

Müşterilere özel analizler, müşteri tabanındaki en değerli bireyleri belirlemek ve bu müşterilerin davranışlarını daha derinlemesine anlamak amacıyla gerçekleştirilmektedirler. Bu analizlerin temel hedefi, yüksek değerli müşterilerin harcama alışkanlıklarını, tercihlerini ve sadakat düzeylerini tespit ederek, işletmelerin stratejik kararlarını bu doğrultuda şekillendirmelerini sağlamaktır. Özellikle en yüksek harcamayı gerçekleştiren müşteriler, işletme gelirlerinin önemli bir kısmını oluşturabilmekte ve bu müşterilerin elde tutulması, işletmenin sürdürülebilir büyümesi için kritik bir faktörü oluşturmaktadır. Bu tür analizler, müşteri sadakat programlarının tasarımı, kişiselleştirilmiş kampanyaların geliştirilmesi ve özel teklifler sunulması gibi stratejilere rehberlik etmektedirler. Ayrıca, yüksek değerli müşterilerin alışveriş desenleri incelenerek, bu davranışların diğer müşteri gruplarına nasıl aktarılabilirliği üzerine çalışmalar da yapılabilir. Sonuç olarak, müşterilere özel analizler, işletmelerin kârlılıklarını artırmaları, müşteri memnuniyetlerini iyileştirmeleri ve uzun vadeli müşteri ilişkileri kurmaları açısından kritik bir role sahiptirler.



Şekil 4. En yüksek Harcamayı Gerçekleştiren Müşteriler

Şekil 5'teki, Veri setinde tercih edilen kategoriye göre gelir dağılımını görsel olarak incelemek için Boxplot kullanılarak her kategori için gelir dağılımı görselleştirilmiştir. Bu, gelirlerin medyanı, çeyrekler arası aralıkları (IQR) ve outlier (aykırı) değerleri gibi önemli istatistiksel bilgilerin hızlıca anlaşılmasını sağlamaktadır.



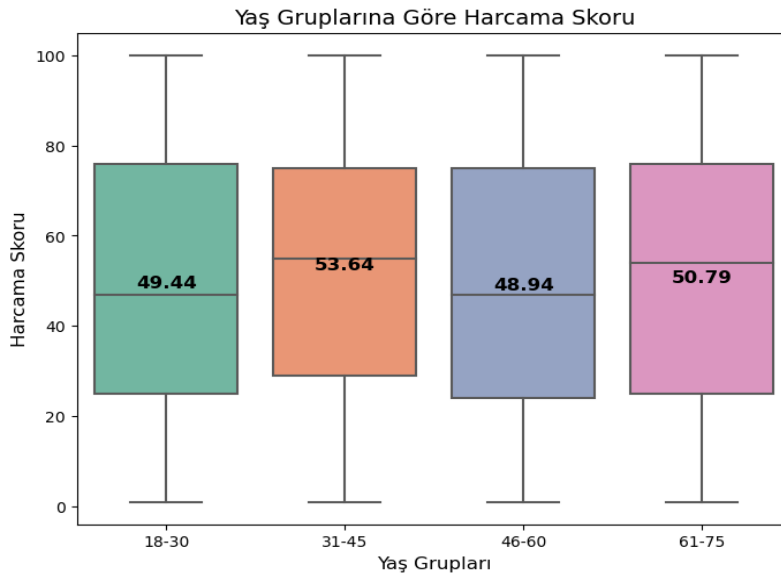
Şekil 5. Tercih Edilen Kategoriler Göre Gelir Dağılımı

Gelir dağılımı verileri, farklı kategorilerdeki ortalama gelirlerin, benzer seviyelerde olduğunu göstermektedir. Clothing (Giyim) kategorisi 89350,01 TL ile en yüksek gelire sahipken, Sports (Spor) kategorisi 87563,20 TL ile en

düşük geliri elde etmiştir. Bu fark, yalnızca birkaç yüz TL civarındadır ve tüm kategorilerin gelirleri birbirine oldukça yakın seviyelerde kalmaktadır. Bu durum, her bir kategorinin farklı pazar segmentlerine hitap etmesine rağmen, gelirlerinin genel olarak homojen olduğunu göstermektedir. Özellikle Electronics (Elektronik) ve Home & Garden (Ev ve Bahçe) gibi yüksek talep gören ve genellikle daha pahalı ürünlerin bulunduğu kategorilerde benzer gelir düzeylerine sahip olup, bu kategorilerdeki yüksek tüketim talebinin ve geniş müşteri kitlesinin gelirleri dengede tutan etmenler arasında yer aldığı söylenebilir. Öte yandan, Sports kategorisinin biraz daha düşük bir ortalama gelire sahip olması, bu alanın daha niş bir pazar segmentine hitap etmesi ve yüksek talep görmeyen ürünlere odaklanması ile açıklanabilir. Genel olarak, bu veriler, tüketici harcamalarının geniş bir yelpazeye yayıldığını, ancak farklı kategoriler arasında belirgin bir gelir farkı olmadığını göstermektedir. Bu tür gelir dağılımı analizleri, pazar araştırmaları, fiyatlandırma stratejileri ve sektör karşılaştırmaları yapmak için önemli bir kaynağı oluşturmaktadır.

4.5. Yaş Gruplarına göre harcama skorları

Yaş gruplarına göre harcama skorlarını görselleştirmek amacıyla, bu analiz gerçekleştirilmiştir. Boxplot grafiđi, her bir yaş grubunun harcama skorlarındaki dağılımının ve merkezi eğilimlerinin (ortalama, medyan, çeyrekler arası aralık gibi) anlaşılmasına yardımcı olmaktadır. Burada kullanılan "Harcama Skoru" verisi, belirli bir yaş grubunun genel harcama alışkanlıklarını veya finansal davranışlarına dair bilgileri sağlar. Analiz, farklı yaş gruplarının harcama alışkanlıklarındaki farkları görselleştirerek, her bir yaş grubunun finansal davranışlarını anlamaya yöneliktir.



Şekil 6. Yaş Gruplarına Göre Harcama Skorları

Şekil 6'daki verilere göre, 18-30 yaş arası bireylerin ortalama harcama skoru 49,44, 31-45 yaş arası bireylerin ortalama harcama skoru ise 53,64'tür. Daha ileri yaş gruplarında ise 46-60 yaş arası ve 61-75 yaş arası gruplarının harcama skorları sırasıyla 48,94 ve 50,79 olarak gözlemlenmektedir. Bu sonuçlar; harcama skorlarının yaş grupları arasında belirgin bir dalgalanma göstermediğini, ancak orta yaş grubunun diğer gruplara kıyasla daha yüksek bir harcama skoruna sahip olduğunu ortaya koymaktadır.

31-45 arası yaş grubu 53,64 ile en yüksek ortalama harcama skorunu elde etmiştir. Bu yaş aralığı, genellikle kariyerin olgunlaşmaya başladığı ve finansal istikrarın arttığı bir dönemdir. Bu dönemde bireyler, gelirlerini daha fazla harcamaya yönlendirebilirler, özellikle aile kurma, ev alımı veya daha büyük tüketim harcamaları gibi ihtiyaçlar doğrultusunda artan harcamalar da görülebilir.

18-30 yaş arası grubu ise 49,44'lük harcama skoru ile diğer gruplardan düşük bir ortalamaya sahipken, bu yaş grubundaki bireylerin genellikle kariyerlerinin başında olmaları, finansal bağımsızlıklarını yeni kazanmaya başlamaları ve eğitim gibi masraflarının olması harcama eğilimlerini etkileyebilir. Bu yaş grubunun harcamaları, genellikle daha genç ve eğlence odaklı tüketim alışkanlıkları doğrultusunda şekillenmiş olabilir.

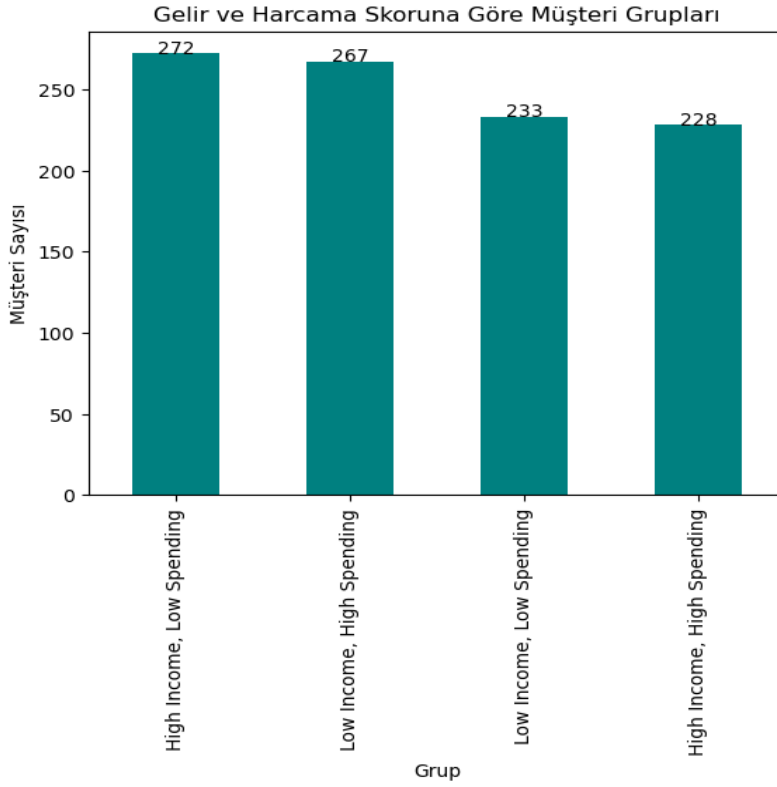
46-60 yaş arası grubu, 48,94 ile en düşük ortalama harcama skoruna sahiptir ve bu yaş grubunda, harcama eğilimlerinin kısıtlanmış olabileceği gözlemlenmektedir. Bu yaş grubundakilerin, emekliliğe yakın dönemde, finansal planlamalarını gözden geçiren ve harcamalarını daha temkinli yapan bireyler olduğu düşünülebilir. Bu dönemde bireyler, emekliliğe yaklaşmanın getirdiği belirsizlikler nedeniyle finansal durumlarını daha sıkı kontrol etme eğilimine girmiş olabilirler. Özellikle bu yaş dönemindeki bireylerde, geleceğe yönelik tasarruf yapma eğilimleri gözlemlenmektedir. Emeklilik dönemine yaklaşmakta olan bireyler, gelirlerinin azalacağını öngördükleri için harcamalarını daha bilinçli bir şekilde yönetmeye başlamaktadırlar. Bu durum, harcama alışkanlıklarını kısıtlamalarına yol açabilir. Bu yaş aralığındaki kişiler, aktif çalışma hayatlarında edindikleri varlıkları koruma ve mevcut borçlarını kapatma yönünde hareket edebilirler. Büyük harcamalardan kaçınarak, tasarruflarını artırmayı hedefleyebilirler. Yaş ilerledikçe sağlık harcamalarının artabileceği beklentisi, bireyleri harcamalarını daha ihtiyatlı yapmaya yönlendirebilir. Emeklilik sigortası, özel sağlık sigortası gibi ek mali yükler, diğer harcama kategorilerinde daha dikkatli davranmalarına sebep olabilir.

Son olarak, 61-75 yaş arası grubu 50,79 ile ortalama harcama skorunun diğer gruplara yakın olduğunu göstermektedir. Bu yaş grubundaki bireyleri, genellikle emekli olmuş ya da emekliliğe yakın dönemde olan kişiler oluşturduğu için, harcama alışkanlıkları gelir düzeyine ve yaşam standartlarına bağlı olarak değişkenlikler gösterebilir. Emeklilik maaşı veya sabit gelirle yaşamaları, harcama tutarlarını daha kontrollü bir seviyede tutmalarına neden olabilir. Bu yaş grubunda, emeklilikle birlikte düzenli maaş gelirinin yerini emekli maaşı, bireysel emeklilik birikimleri veya yatırım getirileri almaktadır. Bu değişim, harcama esnekliğini etkileyebilir. Eğer bireyler mali durumlarını güçlü tutmuşlarsa, belirli harcama kategorilerinde (sağlık, seyahat vb.) artışlar görülebilir. Çalışma hayatının sona ermesiyle birlikte, özellikle zorunlu harcamalar dışındaki tüketim alışkanlıklarında değişiklikler gözlemlenebilir. Örneğin, iş hayatına bağlı masraflar azalırken (ulaşım, profesyonel giyim vb.), hobi ve sağlık odaklı harcamalar artabilir. Bu yaş grubundaki bireylerin sosyal yaşam tarzları da harcama eğilimlerini şekillendirebilir. Örneğin, torunlarına destek olma, hayır işleri veya miras planlaması gibi finansal hareketlilikleri olabilir.

Bu analiz; yaş gruplarına göre farklı harcama alışkanlıklarının belirginleştiğini, ancak genellikle harcama eğilimlerinin orta yaş grubunda zirveye ulaşarak daha sonra azalmaya başladığını ortaya koymaktadır.

4.6. Gelir ve Harcama Skoruna Göre Müşteri Grupları

Müşteri segmentasyonu yapmak amacıyla gelir ve harcama skorları dikkate alınarak müşteri gruplarının oluşturulması hedeflenmektedir. İki ana değişken, gelir ve harcama skoru, müşteri davranışlarını anlamak ve pazarlama stratejilerini hedeflemek için kullanılmaktadır. Bu tür bir segmentasyon, farklı müşteri tiplerinin özelliklerini anlamaya, hangi grupların daha değerli olduğunu belirlemeye ve bu gruplara özel stratejiler geliştirmeye yardımcı olmaktadır.



Şekil 7. Gelir ve Harcama Skoruna Göre Müşteri Grupları

Verilere göre, müşteri segmentasyonunda en büyük grup High Income, Low Spending olan 272 müşteri ile öne çıkmaktadır. Bu grup, yüksek gelirli ancak düşük harcama grubundaki müşterileri temsil etmektedir. İkinci sırada Low Income, High Spending grubu yer almaktadır. Bu grupta, gelirleri düşük, ancak harcamaları yüksek olan 267 müşteri yer almaktadır. Üçüncü sırada Low Income, Low Spending grubu yer almaktadır. Bu grupta, gelirleri de harcamaları da düşük olan 233 müşteri yer almaktadır. Bu grup düşük katkılı müşteri segmenti olarak değerlendirilebilir. Son olarak, High Income, High Spending grubu 228 müşteri ile en küçük segmenti oluşturursa da; yüksek gelirli ve yüksek harcamalı müşterilerden oluşmaktadır; dolayısıyla potansiyel olarak en değerli segmenttir. Gelir ve harcama skorları üzerinden gerçekleştirilen müşteri segmentasyonu, farklı müşteri gruplarının harcama davranışlarının anlaşılmasına yardımcı olmaktadır. Ancak, gelir seviyesinin yalnızca harcama eğilimleri üzerindeki etkisini değil, aynı zamanda müşteri sadakati ve markaya olan bağlılık üzerindeki etkisini incelemek de önemlidir. Gelir seviyesi ve müşteri sadakati arasındaki ilişki, pazarlama stratejilerinin daha hedefli bir şekilde oluşturulmasını sağlayabilir. Aşağıda bu ilişkiler açıklanmaya çalışılmıştır.

High Income, Low Spending (Yüksek Gelirli, Düşük Harcamalı) Müşteriler: Bu grup, yüksek gelir seviyesine sahip olmasına rağmen, harcamalarını kısıtlayan müşterilerden oluşmaktadır. Ancak, bu müşterilerin harcama yapma konusunda çekimser olmalarının nedeni yalnızca tasarruf alışkanlıkları olmayabilir. Yüksek gelirli müşteriler genellikle bilinçli tüketicilerdir ve sadakat programlarına daha duyarlı olabilirler. Kaliteli hizmet ve ayrıcalıklı deneyimler sunulduğunda, bu grup daha fazla harcama yapmaya teşvik edilebilir. Örneğin, premium üyelikler, özel indirimler ve kişiselleştirilmiş öneriler, bu segmentin markaya olan bağlılığını artırabilir. Bu segment için ileri sürülebilecek stratejik öneriler; kişiselleştirilmiş lüks hizmetler ve özel teklifler, VIP müşteri programları ile bu gruba özel ayrıcalıklar, premium içerikler ve deneyimler (özel etkinlikler, öncelikli müşteri desteği vb.) olabilir.

Low Income, High Spending (Düşük Gelirli, Yüksek Harcamalı) Müşteriler: Bu gruptaki müşteriler, gelir seviyelerine kıyasla daha fazla harcama yapma eğilimindedir. Fiyat odaklı kararlar alabilen bu grup, fırsatları ve indirimleri takip ederek alışverişlerini şekillendirmektedir. Bu segmentin sadakat düzeyi genellikle fiyat avantajlarına bağlıdır. Eğer rekabetçi fiyatlandırma ve avantajlı kampanyalar sunulursa, bu müşteriler markaya daha bağlı hale gelebilirler. Ancak, fiyat dışındaki faktörlere (hizmet kalitesi, müşteri deneyimi vb.) daha az duyarlı olabilirler. Bu segment için ileri sürülebilecek stratejik öneriler; sadakat programları kapsamında indirim kuponları, puan sistemleri ve özel kampanyalar, uygun fiyatlı ödeme planları veya taksit seçenekleri ile müşteri harcama deneyimlerinin kolaylaştırılması, sezonluk kampanyalarla bu segmentin alışveriş eğilimleri teşvik edilmesi olabilir.

Low Income, Low Spending (Düşük Gelirli, Düşük Harcamalı) Müşteriler: Bu grup, en düşük harcama ve gelir seviyesine sahip müşteri segmentidir ve genellikle markaya olan bağlılıkları daha düşüktür. Bu segment, fiyat duyarlılığı yüksek olduğundan genellikle indirimler ve promosyonlar yönünde hareket eder. Ancak, belirli bir marka ile sürekli etkileşim içinde olmaları durumunda sadakat kazanılabilir. Bu segment için ileri sürülebilecek stratejik öneriler; düşük fiyatlı ürün alternatifleri sunularak erişilebilirliklerin artırılması, sadakat programları ile daha fazla kazanç elde etmelerinin teşvik edilmesi (örneğin, alışveriş yaptıkça puan biriktirme), sosyal medya ve topluluk bazlı etkileşimlerle marka bilinirliğinin ve bağlılığın artırılması olabilir.

High Income, High Spending (Yüksek Gelirli, Yüksek Harcamalı) Müşteriler: Bu grup, potansiyel olarak en değerli müşteri segmentidir; çünkü hem gelir düzeyleri, hem de harcama eğilimleri yüksektir. Yüksek gelirli ve yüksek harcama yapan müşterilerden oluşmuş bu segmentte, markalara olan bağlılıkların yüksek olması beklenmektedir. Eğer markaya duyulan güven ve alınan hizmet tatmin edici ise, uzun vadeli sadakat geliştirme potansiyeli oluşacaktır. Bu segment için ileri sürülebilecek stratejik öneriler; kişiselleştirilmiş VIP hizmetleri ile müşteri memnuniyetinin artırılması, özel etkinlikler ve prestijli iş birlikleri ile marka prestiji güçlendirilerek bu grubun ilgisinin çekilmesi, lüks ve premium ürün segmentlerine özel pazarlama kampanyaları düzenlenmesi olabilir.

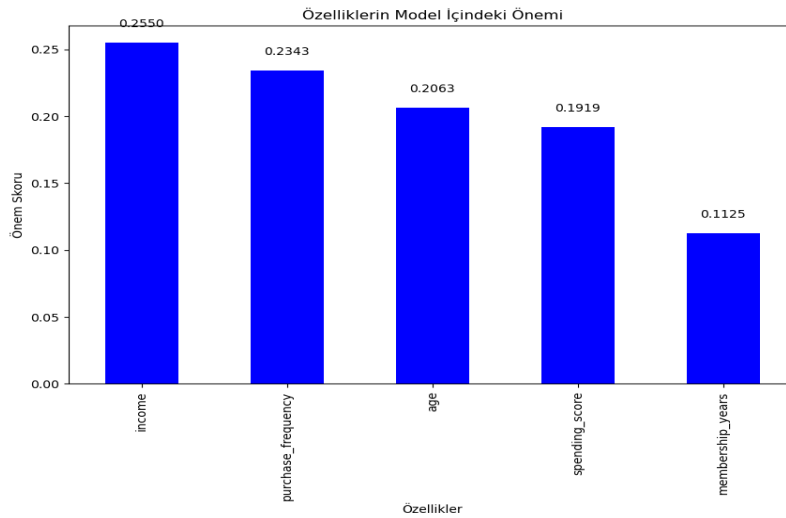
Gerçekleştirilen bu analiz, gelir seviyesinin yalnızca harcama eğilimleri üzerindeki etkisini değil, aynı zamanda müşteri sadakati ve markaya bağlılık üzerindeki etkisini de ortaya koymaktadır. Özetlemek gerekirse:

- Yüksek gelirli müşteriler (High Income, Low Spending & High Spending), genellikle kalite ve deneyime odaklıdır. Bu nedenle bu gruplardaki müşterilere, sadakatlerini artırmak için premium hizmetler, kişiselleştirilmiş öneriler ve özel ayrıcalıklar sunulmalıdır.
- Düşük gelirli müşterilerin (Low Income, High Spending & Low Spending), fiyat duyarlılıkları yüksek olduğundan, indirim kampanyaları, ödeme kolaylıkları ve sadakat programlarıyla markaya bağlılıkları artırılabilir.

Müşteri bağlılığını artırmak için, her segmentin farklı motivasyonlarına uygun stratejiler geliştirilmelidir.

4.7. Özellik Önem Analizi

Müşteri davranışlarını etkileyen en önemli özellikleri tespit etmek için özellik önemi analizlerinin yapılması tercih edilmektedir. Özellik önemi (feature importance) analizi; belirli bir modelle nasıl kararlar alındığını daha iyi kavrayabilmek amacıyla gerçekleştirilir. Özelliklerin modeldeki önemi; her bir özelliğin (değişkenin) hedef değişken üzerindeki etkisini göstermektedir. Özellikle makine öğrenmesi modellerinde, özelliklerin hangi derecede önemli olduğunu belirlemek, modelin doğruluğunu ve güvenilirliğini artırmak için kritik bir adımdır.



Şekil 8. Özellik Önem Analizi

Şekil 8'deki özellik önem analizi, Random Forest algoritması kullanılarak gerçekleştirilmiştir. Random Forest, çok sayıda karar ağacının birleşiminden oluşan bir anlamsal yöntemdir. Her bir ağaç, verinin farklı alt kümeleri üzerinde

eğitilir ve tahminler, ağaçların çoğunluk oyu ile gerçekleştirilirler. Bu yöntem, her bir özelliğin modelin doğruluğuna katkısını belirleyerek, önemli değişkenleri vurgulamaktadır.

Sonuçlara göre, gelir (0,2550) en önemli özellik olup, satın alma sıklığı (0,2343) ve yaş (0,2063) da önemli etkiye sahiptir. Harcama skoru (0,1919) ve üyelik yılları (0,1125) ise daha az önemli faktörlerdir. Bu, gelir ve alışveriş sıklığının müşteri davranışlarını tahmin edilmesinde güçlü roller oynadığını, diğer değişkenlerin ise daha sınırlı etkiler sağladığını göstermektedir. Random Forest algoritması, bu özelliklerin etkilerini doğru bir şekilde değerlendirerek modelin doğruluğunu artırmaktadır.

5. Sonuç ve Öneriler

Gerçekleştirilen analiz sonuçlarına göre, gelir ve satın alma sıklığı, müşteri davranışlarını tahmin etmede en önemli faktörler olarak öne çıkmaktadır. Özellikle yüksek gelirli ve yüksek harcama yapan müşteriler, işletmeler için en değerli segmenti oluşturmaktadır. Cinsiyet, yaş ve tercih edilen kategori gibi demografik faktörler de segmentasyon üzerinde etkili olmakla birlikte, gelir ve harcama davranışları kadar belirleyici olmamışlardır. Bunun yanı sıra, müşteri yaşam boyu değeri ile üyelik süresi ve satın alma sıklığı arasındaki güçlü korelasyon, uzun süreli üyeliklerin ve sık alışverişlerin müşteri değerini artıran önemli faktörler olduğunu göstermektedir. Bu bulguların, daha hedef odaklı pazarlama stratejileri geliştirilmesi ve müşteri bağlılığının artırılması için yol gösterici olabileceği söylenebilir. Ancak, müşteri davranışları statik değildir; zaman içinde ekonomik koşullar, tüketici alışkanlıkları ve dijitalleşme gibi faktörlerle değişiklik gösterebilirler. Bu nedenle, müşteri segmentasyonunun etkinliğini sürdürülebilmesi için sürekli güncellenmesi gerekmektedir. Bu bağlamda aşağıdaki önerilerin faydalı olabileceği düşünülebilir:

- **Düzenli Veri Analizi:** Satın alma verileri, demografik değişiklikler ve pazar eğilimleri, belirli aralıklarla analiz edilmeli ve segmentasyon modelleri güncellenmelidir.
- **Makine Öğrenimi ve Tahmine Dayalı Modeller:** Müşteri davranışlarını daha iyi anlamak için yapay zekâ ve makine öğrenimi algoritmaları kullanılabilir. Bu sayede, müşteri hareketlerindeki değişimler erken tespit edilerek proaktif stratejiler geliştirilebilir.
- **Geri Bildirim Mekanizmaları:** Müşteri anketleri ve etkileşim verileri değerlendirilerek müşteri tercihleri zaman içinde nasıl değiştiği izlenebilir.
- **A/B Testleri ve Dinamik Kampanyalar:** Farklı segmentler için test edilen pazarlama stratejileri sayesinde, müşteri beklentilerine daha hızlı adapte olunabilir.

Sonuç olarak, müşteri segmentasyonu ve davranış analizi, işletmeler için değerli içgörüler sunmakla birlikte, bu analizlerin sürdürülebilir olması için dinamik bir yaklaşım benimsenmelidir. Zaman içinde değişen müşteri eğilimlerini yakından takip etmek, işletmelerin rekabet avantajı elde etmesini ve müşteri memnuniyetini artırmasını sağlayacaktır.

Kaynakça

- Alaca, Y. (2023). Siber güvenlikte CIC-Darknet2020 veri seti kullanarak VPN/NoVPN ve Tor/NoTor sınıflandırması: Basit ve karmaşık modellerin kullanımı. *Fırat Üniversitesi Müh. Bil. Dergisi*, 35(2), 569-579.
- Alfajr, N. H., & Defiyanti, S. (2024). Prediksi penyakit jantung menggunakan metode random forest dan penerapan principal component analysis (PCA). *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(3S1). <https://doi.org/10.23960/jitet.v12i3S1.5055>.
- Dai, Z., Hu, Z., Shen, T., & Zhang, Y. (2023). Risk prediction of diabetes based on spark and random forest algorithm. *2023 IEEE 2nd International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, 535-539. <https://doi.org/10.1109/EEBDA56825.2023.10090801>
- Doğanlı, B., & Çelik, S. (2024). *Pazarlama stratejileri için veri bilimi ve Python* (1st ed.). All Sciences Academy.
- Gürsakal, N., & Çelik, S. (2021). *Büyük veri ve pazarlama* (Birinci Baskı). Dora Yayınevi.
- Ignatenko, V., Surkov, A., & Koltcov, S. (2024). Random forests with parametric entropy-based information gains for classification and regression problems. *PeerJ Computer Science*, 10, e1775. <https://doi.org/10.7717/peerj-cs.1775>

- Kumar, D., Kothiyal, A., Kumar, R., Hemantha, C., & Maranan, R. (2024). Random Forest approach optimized by the Grid Search process for predicting the dropout students. *2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET)*, 1–6. <https://doi.org/10.1109/ICICET59348.2024.10616372>
- Li, Y., & Mu, Y. (2024). Research and performance analysis of random forest-based feature selection algorithm in sports effectiveness evaluation. *Scientific Reports*, *14*(1), 26275. <https://doi.org/10.1038/s41598-024-76706-1>
- Madhiraju, B., Reddy, S., & Sasikala, G. (2024). Customer Segmentation Using RFM Analysis. *EPRA International Journal of Economic and Business Review*, 15-22. <https://doi.org/10.36713/epra17685>
- Miao, Y., & Xu, Y. (2024). Random forest-based analysis of variability in feature impacts. *2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA)*, 1130–1135. <https://doi.org/10.1109/ICIPCA61593.2024.10708791>
- Mrg, R. A., & Hasibuan, M. S. (2024). Best student classification using ensemble random forest method. *Sistemasi*, *13*(3), 1188. <https://doi.org/10.32520/stmsi.v13i3.4101>
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. <https://doi.org/10.58496/BJML/2024/007>
- Wong, C.-G., Tong, G.-K., & Haw, S.-C. (2024). Exploring customer segmentation in e-commerce using RFM analysis with clustering techniques. *Journal of Telecommunications and the Digital Economy*, *12*(3), 97–125. <https://doi.org/10.18080/jtde.v12n3.978>

ETHICAL AND SCIENTIFIC PRINCIPLES STATEMENT OF RESPONSIBILITY

The author declare that ethical rules and scientific citation principles were complied with throughout the preparation process of this study.

STATEMENT OF RESEARCHERS' CONTRIBUTION RATE TO THE ARTICLE

1st author contribution rate: 100%

ETİK VE BİLİMSEL İLKELER SORUMLULUK BEYANI

Bu çalışmanın tüm hazırlanma süreçlerinde etik kurallara ve bilimsel atıf gösterme ilkelerine riayet edildiđini yazar beyan eder. Bu çalışmada, <https://www.kaggle.com/datasets/fahmidachowdhury/customer-segmentation-data-for-marketing-analysis> adresinden alınan “digital_marketing_campaign_dataset” veri seti kullanılmıştır. Bu veri seti herkese açık ve ikincil bir kaynak olduğundan, bu çalışma için etik komite onayı gerekmemiştir. Bu veri seti büyük ölçekli analizler için kullanılabilir ve etik izin gerektirmeyen ikincil bir kaynaktır.

ARAŞTIRMACILARIN MAKALEYE KATKI ORANI BEYANI

1. Yazar Katkı Oranı: %100