

# Pazarlama Kampanyası Yanıt Analizi

Adım İclal Akkuş, numaram 1323130007 bu araştırmayı Python ile Veri Bilimi dersi özelinde gerçekleştirdim. Çalışmamın bitiş tarihi 16/06/2025'tir.

## 1. Projenin Konusu ve Amacı

Günümüz rekabetçi pazarlama ortamında, doğru hedef kitleye doğru kampanyaların sunulması, şirketlerin verimliliğini ve kârlılığını doğrudan etkileyen unsurlar arasındadır. Bu çalışmanın amacı, bir şirketin geçmiş müşteri verilerini kullanarak gelecekteki pazarlama kampanyalarına verilecek yanıtları öngörmektir.

Amaç elde edilen bulgular, pazarlama stratejilerinin daha etkin hedeflenmesine katkı sağlamayı ve müşteri yanıt oranlarını artıracak karar destek mekanizmalarının oluşturulmasını amaçlamaktadır.

## 2. Veri Seti Tanıtımı

Kullanılan veri seti, Kaggle platformunda yayımlanmış olan “Marketing Campaign” adlı veri kümesidir. Veri seti, bir perakende şirketinin müşterilerine ilişkin demografik bilgiler, geçmiş satın alma davranışları ve farklı pazarlama kampanyalarına verdikleri tepkilerden oluşmaktadır. Toplamda 29 öznitelik ve 2.200'ün üzerinde müşteri kaydını içeren veri seti, kampanya başarısının öngörülmesine ve müşteri segmentlerinin belirlenmesine olanak tanımaktadır.

Veri setinde yer alan değişkenler genel olarak beş ana kategoriye ayrılabilir:

- Demografik Bilgiler:** Müşteri yaşı (Year\_Birth), eğitim durumu (Education), medeni hali (Marital\_Status), evdeki çocuk sayısı (Kidhome, Teenhome) ve gelir düzeyi (Income) gibi nitelikler.
- Satın Alma Davranışları:** Son iki yıl içinde farklı ürün kategorilerine yapılan harcamalar (MntWines, MntFruits, MntMeatProducts vb.).
- Etkinlik ve Ziyaret Bilgileri:** Web sitesi, mağaza ve katalog kanallarından yapılan alışveriş sayıları (NumWebPurchases, NumStorePurchases, NumCatalogPurchases) ve web ziyareti sıklığı (NumWebVisitsMonth).
- Kampanya Yanıtları:** Şirketin geçmişte yürüttüğü beş ayrı kampanyaya verilen yanıtlar (AcceptedCmp1–AcceptedCmp5) ile son kampanyaya verilen yanıtı gösteren Response değişkeni.

- **Zaman Bilgileri:** Müşterinin sisteme kayıt tarihi (Dt\_Customer) ve son satın alma işleminden bu yana geçen gün sayısı (Recency).

### 3. Veri Ön İşleme Adımları

Çalışmaya başlamadan önce veri setinin kalitesini ve analiz için uygunluğunu artırmak amacıyla kapsamlı bir ön işleme süreci uygulanmıştır. Bu süreçte aşağıdaki temel adımlar gerçekleştirilmiştir:

#### Eksik Veri Analizi ve Temizliği:

Veri setinde yer alan eksik değerler tespit edilmiş, özellikle Income değişkenindeki eksik gözlemler uygun yöntemlerle tamamlanmış ya da veri kaybını en aza indirecek şekilde işlenmiştir.

Bu süreçte, bilgi içeriği düşük veya analiz açısından gereksiz bulunan **ID**, **Dt\_Customer**, **Z\_CostContact** ve **Z\_Revenue** değişkenleri veri setinden çıkarılmıştır.

#### Yeni Özellik Türetme:

Analizlerin gücünü artırmak için mevcut değişkenlerden yeni anlamlı özellikler oluşturulmuştur. Veri setine analiz açısından anlamlılık katması için eklenen yeni değişkenler:

**Customer\_Age:** Müşteri yaşı (Year\_Birth değişkeninden hesaplandı).

**Customer\_Since\_Days:** Müşterinin kayıt süresi (Dt\_Customer üzerinden).

**Children:** Evdeki çocuk sayısı (Kidhome + Teenhome).

**TotalAcceptedCmp:** Kabul edilen kampanya sayısı (AcceptedCmp1–5 toplamı).

**TotalSpent:** Toplam harcama miktarı (ilgili kategorilerin toplamı).

#### Kategorik Değişkenlerin Kodlanması:

Analiz ve modelleme aşamalarında kullanılmak üzere, Education ve Marital\_Status gibi kategorik değişkenler sayısal forma dönüştürülmüştür. Bu işlem için etiket kodlama (Label Encoding) yöntemi tercih edilmiştir.

#### Aykırı Değerlerin Tespiti ve Temizliği:

Yaş, gelir ve toplam harcama gibi kritik değişkenlerde gözlenen aşırı uç değerler, verinin geneline uyumsuzluk oluşturmaması için Interquartile Range (IQR) yöntemi kullanılarak belirlenmiş ve veri setinden çıkarılmıştır.

## Veri Ölçeklendirme:

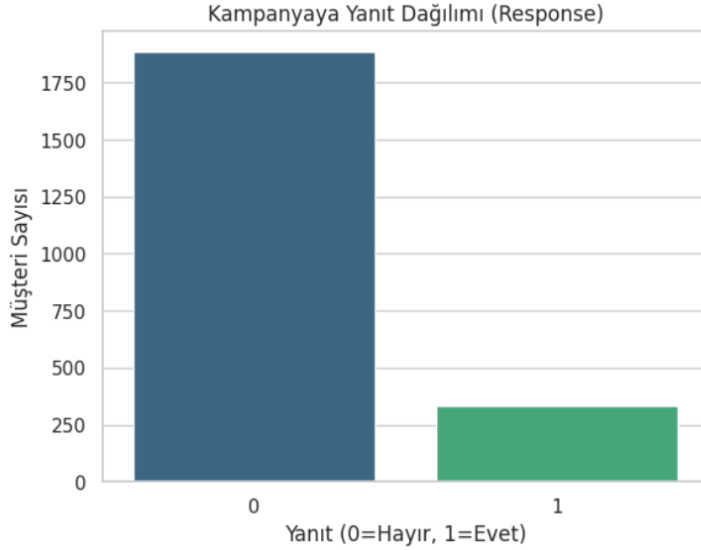
Modelleme ve kümeleme algoritmalarının performansını artırmak amacıyla, sayısal değişkenler Min-Max Normalizasyon yöntemiyle 0-1 aralığına ölçeklendirilmiştir.

Bu ön işleme adımları, veri setinin tutarlı, dengeli ve analizlere uygun hale gelmesini sağlayarak, elde edilen sonuçların doğruluğunu ve güvenilirliğini artırmıştır.

## 4. Keşifsel Veri Analizi (EDA)

Keşifsel veri analizini kullanma amacım, veriyle ilgili genel yapıyı, örüntüleri, aykırı değerleri, eksik verileri ve ilişkileri ortaya çıkarmaktır.

### Kampanyaya Yanıt Dağılımı

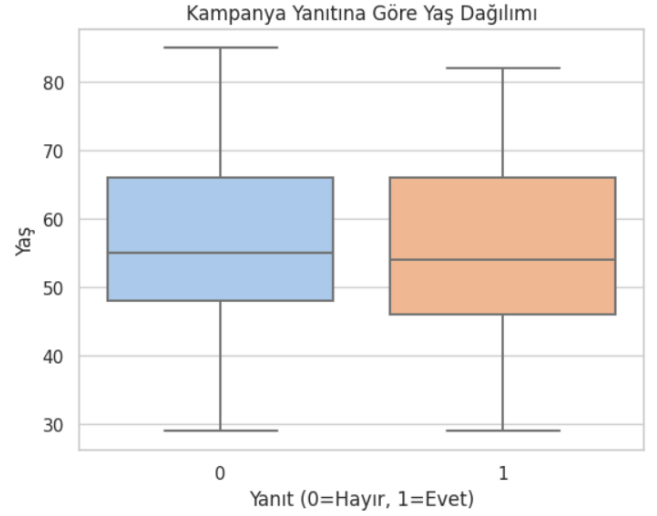
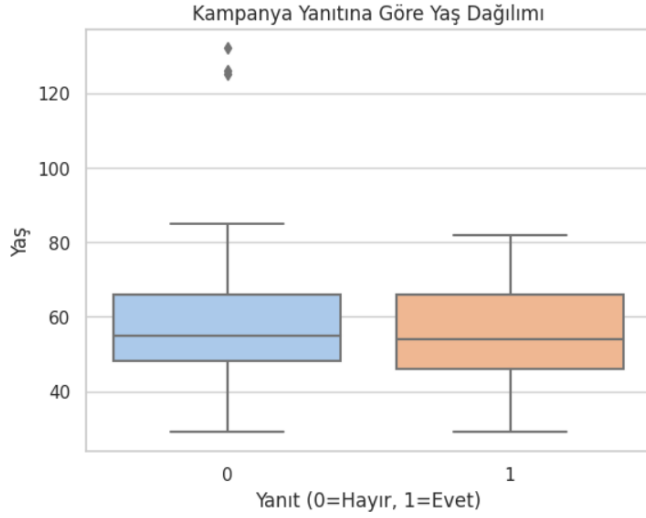


- Sol sütun (0) çok daha yüksek → Bu, **kampanyaya yanıt vermeyen müşteri sayısının çok fazla olduğunu** gösteriyor.
- Sağ sütun (1) çok daha düşük → Kampanyaya yanıt verenler **azınlıkta**.
- Model, çoğunluk sınıfı (yanıt vermeyenlere) odaklanıp, azınlık sınıfı (yanıt verenleri) göz ardı edebilir.

Bu durumda model kurulduğunda yüksek doğruluk gösterebilir ama model azınlık sınıfı iyi tahmin edemez.

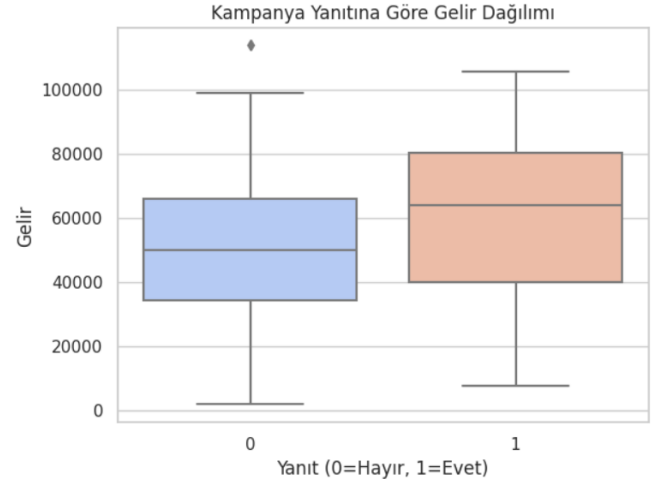
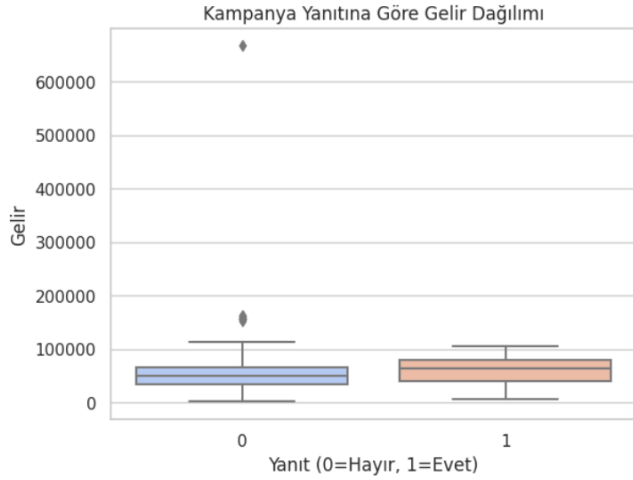
Keşifsel veri analizinin devamında ilişkili öznitelikleri grafiklerle gözlemledim. Aykırı gözlemin etkisini göstermek için ilk grafikte aykırı değerli halini, ikinci grafikte aykırı değerden arınmış halini sunuyorum.

## Kampanya Yanıtına Göre Yaş Dağılımı



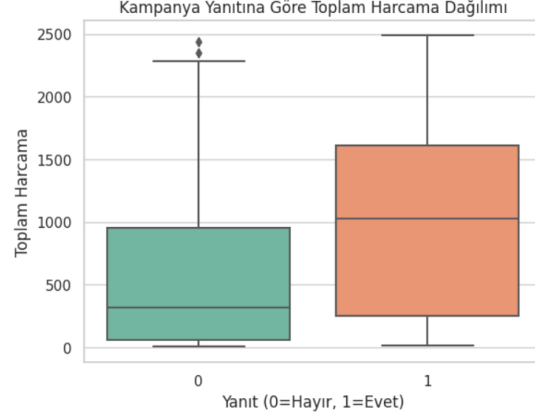
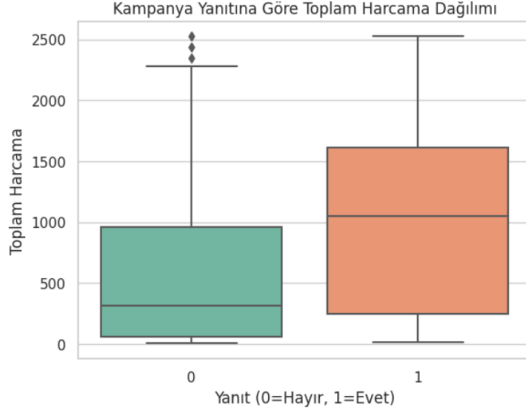
- Yaş, kampanya yanıtını **tek başına güçlü şekilde belirlemediği** görünüyor.
- Ama 50 yaş ve üzeri bireyler, her iki grupta da baskın durumda.

## Kampanya Yanıtına Göre Gelir Dağılımı



- Gelir, kampanya yanıtını etkileyen bir faktör olabilir.
- Daha yüksek gelir grubundaki müşterilerin kampanyaya **daha ilgili** oldukları görülüyor.

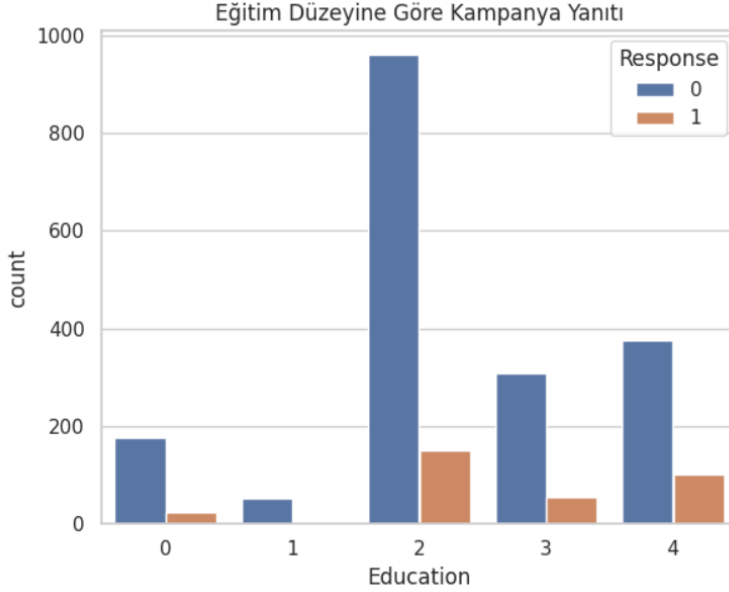
## Kampanya Yanıtına Göre Toplam Harcama Dağılımı



**Toplam harcama (TotalSpent)**, kampanyaya yanıt verme olasılığıyla **doğrudan ilişkili**.

Harcama düzeyi yüksek müşterilere yeni kampanyalar yöneltmek, pazarlama yatırımının geri dönüşünü artırabilir.

## Eğitim Düzeyine Göre Kampanya Yanıtı (Education vs Response)



En yüksek müşteri sayısı  
“**Graduation**” (kod 2) kategorisinde  
(lisans mezunları).

Bu kategoride yanıt verenlerin sayısı  
da diğer gruplara kıyasla daha fazla.

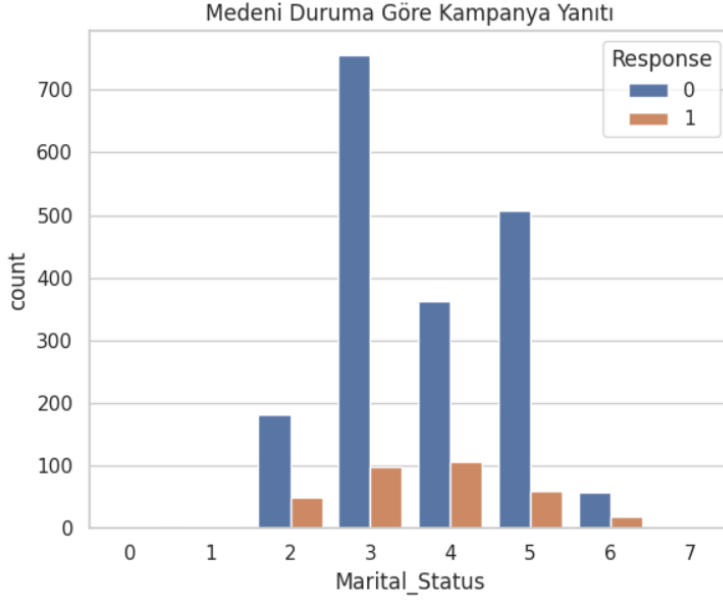
Genel olarak eğitim düzeyi  
yükseldikçe kampanyaya yanıt  
verenlerin oranı artıyor gibi  
gözüküyor (Master ve PhD sahipleri  
dahil).

Eğitim düzeyi arttıkça müşterilerin  
kampanyaya yanıt verme eğilimi  
yükselmektedir. Bu, pazarlama

stratejilerinde eğitim seviyesine göre hedeflemeler yapılabileceğini gösterir.

Eğitim Kodları: 0 = Basic (Temel Eğitim), 1 = 2n Cycle (İkinci Kademe Eğitim - Lise),  
2 = Graduation (Lisans Mezunlu), 3 = Master (Yüksek Lisans), 4 = PhD (Doktora)

## Medeni Duruma Göre Kampanya Yanıtı



Single (Bekâr), 5 = Together (Birlikte Yaşayan), 6 = Widow (Dul), 7 = YOLO (Sadece Bir Kez Yaşa - Veri setine özgü, anlamı belirsiz)

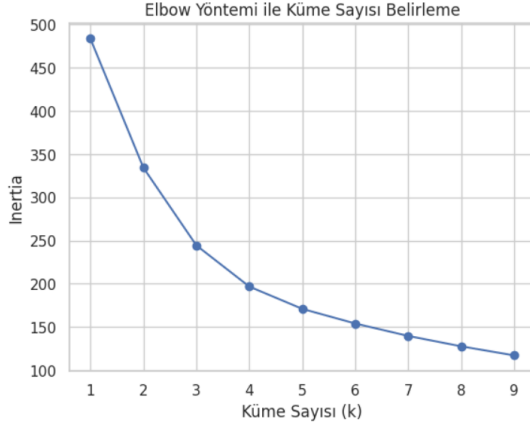
En fazla müşteri sayısı “**Together**” ve “**Married**” kategorilerinde. Bu iki grup kampanyaya yanıt verenlerin de önemli kısmını oluşturuyor.

Evli veya birlikte yaşayan müşterilerin kampanyaya yanıt verme oranları daha yüksek. Medeni durum, pazarlama kampanyalarının segmentasyonunda dikkate alınabilir.

Medeni Durum Kodları:

0 = Alone (Yalnız), 1 = Absurd (Anlamsız / Tutarsız), 2 = Divorced (Boşanmış), 3 = Married (Evli), 4 =

## 5. Müşteri Segmentasyonu



K-means algoritması ile yapılan segmentasyonda en uygun küme sayısı 4 olarak belirlenmiştir. Segmentler; düşük gelirli ilgisiz müşteriler, yüksek harcama yapan sadık kullanıcılar, pasif kullanıcılar ve yüksek potansiyelli ancak düşük katılım gösteren bireyler şeklinde sınıflandırılmıştır.

	Income	Customer_Age	TotalSpent	Customer_Since_Days	Response
Segment					
0	36179.60	53.78	208.90	4522.75	0.15
1	69962.65	58.58	1289.20	4541.22	0.28
2	39254.44	54.98	144.18	4179.41	0.05
3	73296.75	58.63	1149.67	4181.69	0.15

Her segment, gelir düzeyi, toplam harcama miktarı, müşteri yaşı ve kayıt süresi gibi temel göstergeler açısından anlamlı farklılıklar göstermiştir. Ayrıca, segmentlerin pazarlama kampanyalarına yanıt verme oranları da önemli çeşitlilikler içermektedir.

- **Segment 0:** Düşük gelir ve düşük harcama profiline sahip müşterilerden oluşmaktadır. Yaş ortalaması yaklaşık 54 olup, bu grup uzun süredir müşteridir ancak kampanyalara olan ilgisi sınırlıdır. Bu nedenle, bu segment pazarlama kaynaklarının etkin kullanımı açısından düşük öncelik taşıyan bir grup olarak değerlendirilebilir.
- **Segment 1:** En yüksek gelir ve toplam harcamaya sahip segmenttir. Müşteri sadakati de yüksek olan bu grup, kampanyalara en yüksek yanıt oranını (%28) göstermektedir. Bu nedenle, hedefli e-posta kampanyaları, VIP avantajları ve çapraz satış stratejileriyle özel olarak desteklenmesi önerilmektedir.
- **Segment 2:** Orta yaş grubuna ait, düşük gelirli ve düşük harcama yapan müşterilerden oluşur. Bu segment, %5 gibi oldukça düşük bir yanıt oranına sahiptir. Pazarlama maliyetlerinin geri dönüşünün düşük olduğu bu grup için yeniden ilgilendirme stratejileri geliştirilmesi ya da kaynakların alternatif segmentlere yönlendirilmesi uygun olacaktır.
- **Segment 3:** En yüksek gelir seviyesine sahip müşterilerden oluşmasına rağmen kampanyaya yanıt oranı orta düzeydedir (%15). Bu segment potansiyel olarak değerlidir ancak mevcut kampanya yaklaşımlarının bu gruba yeterince çekici gelmediği düşünülebilir. Daha yaratıcı ve prestijli kampanya dili ile bu segmentteki yanıt oranının artırılması hedeflenmelidir.

Bu analiz, müşterilerin harcama ve demografik özelliklerine göre farklı segmentlerde gruplandırılabilmesini ve her segmentin kampanya tepkisinin farklı olduğunu göstermektedir. Pazarlama stratejilerinin bu segmentasyon doğrultusunda şekillendirilmesi, kaynakların daha verimli kullanılmasını ve kampanya başarısının artırılmasını sağlayacaktır.

## 6. Sınıflandırma Modeli

Random Forest algoritması ile müşterilerin kampanyaya yanıt verip vermeyeceği tahmin edilmiştir. Model genel doğrulukta %87 başarı sağlamıştır. Ancak azınlık sınıf olan yanıt verenler için başarı düşüktür.

## Model Performans Değerlendirmesi (Random Forest)

Bu çalışmada kullanılan Random Forest sınıflandırma modeli, müşterilerin pazarlama kampanyasına yanıt verip vermeyeceğini tahmin etmek amacıyla değerlendirilmiştir. Model performansı aşağıdaki ölçütler ışığında incelenmiştir:

	precision	recall	f1-score	support
0	0.89	0.97	0.92	375
1	0.61	0.29	0.39	66
accuracy			0.87	441
macro avg	0.75	0.63	0.66	441
weighted avg	0.84	0.87	0.85	441

### 1. Genel Başarı

Modelin genel doğruluk oranı %87 olarak belirlenmiştir. Bu yüksek doğruluk, modelin büyük çoğunlukla doğru tahminler yaptığı anlamına gelmekle birlikte, veri setindeki sınıf dengesizliği göz önünde bulundurulmalıdır. Dengesiz veri yapısı, modelin yaygın sınıfı (kampanyaya yanıt vermeyenleri) daha iyi öğrenmesine yol açabilir.

### 2. Kampanyaya Yanıt Vermeyenler (Sınıf 0)

- **Precision (Kesinlik):** 0.89
- **Recall (Duyarlılık):** 0.97
- **F1-Score:** 0.92

Bu değerler, modelin kampanyaya yanıt vermeyen müşterileri yüksek doğrulukla tespit ettiğini göstermektedir. Yani, model “yanıt vermedi” tahmininde büyük oranda başarılıdır.

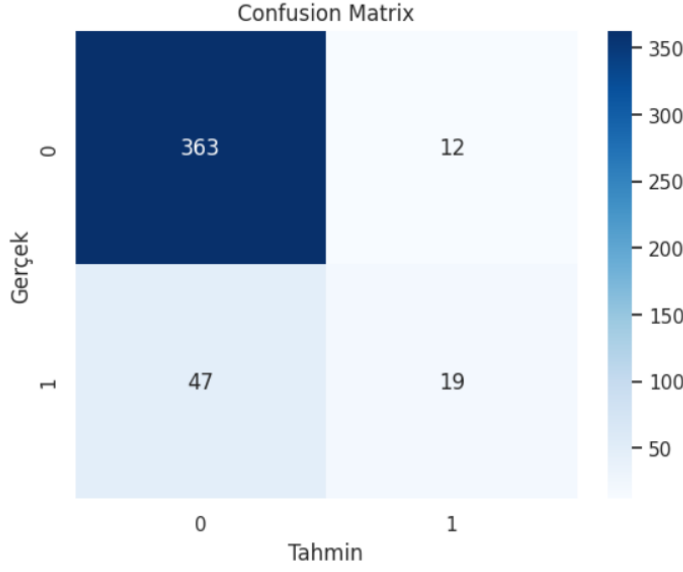
### 3. Kampanyaya Yanıt Verenler (Sınıf 1)

- **Precision (Kesinlik):** 0.61
- **Recall (Duyarlılık):** 0.29
- **F1-Score:** 0.39

Bu metrikler, modelin yanıt veren müşterileri tanımlamakta zorlandığını ortaya koymaktadır. Özellikle düşük recall değeri (%29), modelin gerçek yanıt verenlerin ancak yaklaşık üçte birini doğru tahmin ettiğini göstermektedir. Bu durum, kampanyaya yanıt veren müşterilerin önemli bir kısmının model tarafından “yanıt vermedi” olarak yanlış sınıflandırıldığı anlamına gelir.



## Confusion Matrix (Karmaşıklık Matrisi) Analizi



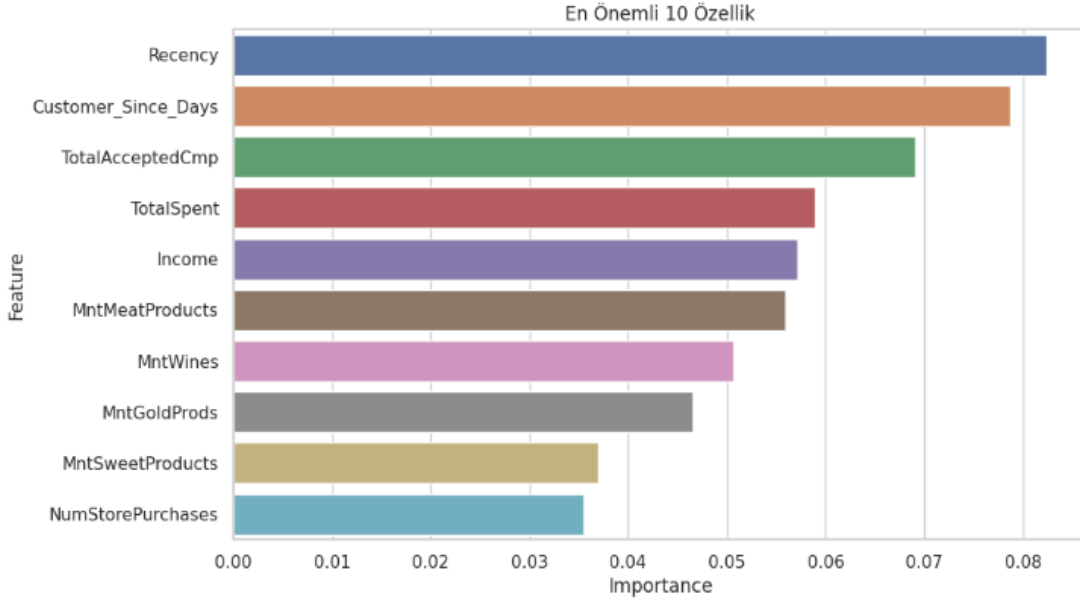
Confusion Matrix, modelin gerçek ve tahmin ettiği sınıflar arasındaki karşılaştırmayı dört kategori üzerinden gösterir.

- **True Negative (TN):** Modelin doğru olarak “yanıt vermeyen” sınıfı tahmin ettiği örnek sayısıdır (363).
- **False Positive (FP):** Modelin yanlış olarak “yanıt veren” dediği ama gerçekte “yanıt vermeyen” olan örneklerdir (12).
- **False Negative (FN):** Modelin yanlış olarak “yanıt vermeyen” dediği ama gerçekte “yanıt veren” olan örneklerdir (47).
- **True Positive (TP):** Modelin doğru olarak “yanıt veren” sınıfı tahmin ettiği örneklerdir (19).

Model, kampanyaya yanıt vermeyen müşterileri oldukça başarılı bir şekilde tahmin etmektedir (yüksek TN sayısı). Ancak, yanıt veren müşterileri tespit etme başarısı düşüktür; özellikle 47 gerçek yanıt veren müşteri model tarafından gözden kaçırılmıştır (FN). Bu durum, modelin azınlık sınıf üzerindeki performansının zayıf olduğunu ve iyileştirme gerektirdiğini göstermektedir. Yanlış negatiflerin yüksek olması, potansiyel değerli müşterilerin kampanya kapsamı dışında kalmasına neden olabilir.

Modelin genel başarısını artırmak için azınlık sınıfın daha iyi tanınması hedeflenmeli, sınıf dengesizliği giderilerek veya farklı algoritmalar kullanılarak performans iyileştirilmelidir.

## 7. Random Forest Modeline Dayalı Özellik Önemi Analizi



Model tarafından belirlenen en önemli 10 özellik, müşterilerin pazarlama kampanyasına yanıt verme olasılığını etkileyen kritik faktörleri göstermektedir.

- **Recency (Son Satın Alma Zamanı):**  
En yüksek önem skoruna sahip değişken olup, müşterinin son satın alma işleminden itibaren geçen gün sayısını ifade eder. Bu, yakın zamanda alışveriş yapan müşterilerin kampanyaya yanıt verme olasılığının daha yüksek olduğunu göstermektedir.
- **Customer\_Since\_Days (Müşterinin Kayıt Süresi):**  
İkinci en etkili özellik olarak, müşterinin şirkete kayıtlı olduğu sürenin uzunluğu yanıt tahmininde belirleyicidir. Uzun süreli müşterilerin kampanyalara daha duyarlı olduğu yorumlanabilir.
- **TotalAcceptedCmp (Kabul Edilen Kampanya Sayısı):**  
Önceki kampanyalara verilen olumlu yanıtların toplamı, müşterinin gelecekteki kampanya tepkisini güçlü şekilde etkiler.
- **TotalSpent (Toplam Harcama):**  
Müşterinin tüm ürün kategorilerindeki toplam harcaması da kampanya yanıtında önemli bir belirleyicidir. Yüksek harcama yapan müşteriler, kampanyalara daha fazla yanıt vermektedir.
- **Income (Gelir Düzeyi):**  
Gelir seviyesi yüksek müşterilerin kampanyaya yanıt verme eğilimleri daha

fazladır, bu da ekonomik durumun pazarlama başarısı üzerindeki etkisini ortaya koyar.

- **MntMeatProducts, MntWines, MntGoldProds, MntSweetProducts (Kategori Bazlı Harcamalar):**  
Belirli ürün gruplarına yapılan harcamalar, müşterinin satın alma alışkanlıkları hakkında ayrıntılı bilgi sunar ve kampanya hedeflemesinde dikkate alınmalıdır.
- **NumStorePurchases (Mağaza Satın Alımları):**  
Fiziksel mağazalardan yapılan alışveriş sayısı, müşterinin genel alışveriş davranışını yansıtarak modelin karar sürecinde yer almaktadır.

Bu analiz, müşteri davranışlarını etkileyen faktörlerin çeşitliliği ve kapsamını ortaya koymaktadır. Modelin en önemli belirleyicileri arasında, müşterinin şirketteki geçmişi (Customer\_Since\_Days), son alışveriş zamanı (Recency) ve önceki kampanya etkileşimleri (TotalAcceptedCmp) yer almaktadır.

Ayrıca, toplam harcama miktarı ve gelir düzeyi gibi ekonomik göstergeler de kampanyaya yanıt verme olasılığında etkili faktörler arasında bulunmuştur. Bu bulgular, müşteri sadakati ve geçmiş etkileşimlerin pazarlama kampanyalarının başarısı üzerindeki doğrudan etkisini net şekilde göstermekte olup, pazarlama stratejilerinin kişiselleştirilmesi ve hedef müşteri segmentlerinin belirlenmesinde önemli girdiler sağlamaktadır.

## 8. Sonuç ve Öneriler

### Keşifsel Veri Analizi Sonucunda Elde Ettiğim Bulgular:

Keşifsel veri analizi sonucunda, kampanyaya yanıt veren müşterilerin genellikle daha yüksek gelir seviyesine ve toplam harcama tutarına sahip olduğu görülmüştür. Ayrıca eğitim ve medeni durum gibi kategorik değişkenlerin de yanıt üzerinde etkili olduğu saptanmıştır.

### Kümeleme ve Sınıflandırma Analizi Sonuçları:

Müşteri segmentasyonu için K-Means algoritması kullanılarak dört farklı müşteri grubu belirlenmiştir. En yüksek gelir ve harcamaya sahip segment, kampanyalara en yüksek yanıt oranını göstermiştir. Düşük gelirli ve düşük harcama yapan segmentler ise kampanyaya daha az ilgi göstermiştir.

Random Forest sınıflandırma modeli, genel doğrulukta %87 başarımla sağlamasına rağmen, azınlık sınıf olan yanıt veren müşterilerin tahmininde düşük recall değerleri göstermiştir. Bu durum, sınıf dengesizliğinin model performansını sınırladığını ortaya koymaktadır.

## **İşletmeye Katkı Sunacak Detaylı Sonuçlar**

### **1. Müşteri Sadakati ve Kampanya Yanıtı Arasındaki Kuvvetli Bağlantı:**

- Customer\_Since\_Days değişkeni, yani müşterinin şirkette kayıtlı olduğu süre, kampanyaya yanıt verme olasılığını anlamada en belirleyici faktörlerden biri olarak ortaya çıkmıştır. Uzun süredir müşterisi olanların kampanyalara yanıt verme oranı daha yüksektir.
- Bu durum, sadık müşterilere özel kampanya stratejilerinin geliştirilmesini zorunlu kılmaktadır.

### **2. Son Satın Alma Zamanının Önemi (Recency):**

- Müşterilerin son satın alma zamanları, kampanya yanıtında kritik rol oynamaktadır. Daha yakın zamanda alışveriş yapan müşterilerin yanıt verme olasılığı anlamlı şekilde daha yüksektir.
- Bu bilgi, yeniden pazarlama (retargeting) ve yakın zamanda aktif müşterilere yönelik kampanyaların önceliklendirilmesi için kullanılabilir.

### **3. Harcamaların Kampanya Tepkisine Etkisi:**

- Toplam harcama (TotalSpent) ile kampanya yanıt oranı arasında pozitif bir ilişki tespit edilmiştir. Yüksek harcama yapan müşterilerin kampanyaya yanıt verme olasılığı daha fazladır.
- Bu nedenle, yüksek harcama yapan segmentlere yönelik ayrıcalıklı kampanyalar tasarlanabilir.

### **4. Demografik Özelliklerin (Yaş, Eğitim, Medeni Durum) Etkisi:**

- Eğitim seviyesi yükseldikçe kampanya yanıt oranında artış gözlenmiştir; özellikle lisans ve üzeri eğitim düzeyine sahip müşteriler daha ilgili.
- Medeni durum açısından, evli veya birlikte yaşayan müşteriler kampanyalara daha olumlu yanıt vermektedir.
- Bu demografik bilgiler pazarlama mesajlarının hedef kitleye göre uyarlanmasında yol gösterici olabilir.

## 5. Müşteri Segmentasyonunun Stratejik Katkısı:

- Kümeleme analiziyle oluşturulan dört segment, gelir, harcama, yaş ve sadakat açısından birbirinden belirgin şekilde farklılaşmaktadır.
- En yüksek yanıt oranı, yüksek gelir ve harcama yapan sadık müşterilerden oluşan segmentte görülmüştür.
- Bu segmentin pazarlama yatırımlarında önceliklendirilmesi kârlılığını artırabilir.
- Düşük yanıt veren segmentler için ise daha düşük maliyetli ve ilgiyi artırmaya yönelik farklı stratejiler geliştirilebilir.

## 6. Modelin Azınlık Sınıfı Yakalamadaki Zayıflığı:

- Azınlık olan yanıt veren sınıf için düşük recall, modelin bu müşteri grubunu yeterince iyi tanımlamadığını gösterir.
- Bu durum, gerçek dünyada önemli bir müşteri grubunun göz ardı edilmesi riskini doğurur.
- İleri tekniklerle bu durumun düzeltilmesi gereklidir.

## İşletmeye Yönelik Öneriler

- Sadık ve yüksek harcama yapan müşteriler için özel sadakat programları ve kişiselleştirilmiş kampanyalar oluşturulmalıdır.
- Kampanyaların zamanlaması, müşterilerin son satın alma tarihleri dikkate alınarak planlanmalıdır.
- Eğitim ve medeni durum gibi demografik özellikler pazarlama stratejilerine entegre edilerek daha etkili hedeflemeler yapılabilir.
- Düşük yanıt veren segmentler için yeniden ilgilendirme yöntemleri ve uygun maliyetli kampanyalar tasarlanmalıdır.

# KAYNAKÇA

## Veri Seti Kaynağı

KAGGLE – Marketing Campaign

<https://www.kaggle.com/datasets/rodsaldanha/arketing-campaign/data>

## Kullanılan Kütüphaneler ve Algoritmalar

### Python:

Pandas — Veri işleme ve manipülasyonu,

NumPy — Sayısal hesaplamalar,

Matplotlib & Seaborn — Veri görselleştirme,

Scikit-learn — Makine öğrenmesi ve veri madenciliği:

- RandomForestClassifier — Rastgele Orman sınıflandırma modeli
- KMeans — K-means kümeleme algoritması
- MinMaxScaler — Veri ölçeklendirme
- train\_test\_split — Veri setinin eğitim ve test olarak bölünmesi

## Referans Alınan Akademik ve Teknik Kaynaklar

1. Breiman, L. (2001). *Random Forests*. Machine Learning, 45(1), 5–32.
2. MacQueen, J. (1967). *Some Methods for Classification and Analysis of Multivariate Observations*. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.
3. Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
4. Pedregosa, F. et al. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.
5. Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). *Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning*. Journal of Machine Learning Research, 18(17), 1-5.