Projenin amacı ve kapsamı

Bu proje, müşteri ilişkileri yönetimi (CRM) veri kümesi üzerinde analizler yaparak müşterilerin davranışlarını daha iyi anlamayı ve işletme stratejilerini optimize etmeyi amaçlamaktadır. E-ticaret sektöründe müşteri davranışlarının analizi, sadık müşterileri belirlemek, riskli müşterileri yeniden kazanmak ve yeni müşterilere yönelik stratejiler geliştirmek açısından kritik bir öneme sahiptir.

Projenin kapsamı şu şekilde özetlenebilir:

1. **Veri Keşfi ve Temizlik**:
   * E-ticaret platformundan elde edilen müşteri verilerinin incelenmesi, eksik ve hatalı verilerin temizlenmesi.
   * Müşteri bazında analiz yapılabilir hale getirilmesi.
2. **Müşteri Segmentasyonu**:
   * Recency (Güncellik), Frequency (Sıklık) ve Monetary (Harcama) metrikleri kullanılarak müşterilerin davranışlarının analizi.
   * RFM analizi ile müşterilerin farklı segmentlere ayrılması.
3. **Kümeleme Analizi**:
   * Müşteri verileri üzerinde K-Means algoritmasını kullanarak davranışsal benzerliklere dayalı otomatik kümeler oluşturma.
   * Optimum küme sayısının belirlenmesi ve kümelerin iş anlamında yorumlanması.
4. **Satın Alma Eğilimi Tahmini**:
   * Müşterilerin gelecekte satın alma olasılıklarını tahmin eden bir sınıflandırma modeli geliştirme.
   * Lojistik regresyon ve karar ağacı gibi modellerle müşterilerin yüksek ve düşük satın alma eğilimlerini analiz etme.
5. **İş Değeri ve Stratejik Öneriler**:
   * Analiz sonuçlarına dayanarak müşteri segmentlerine özel stratejiler önerme.
   * Pazarlama kampanyalarının etkinliğini artırmak için kullanılabilecek içgörüler sağlama.

CRM analizi neden önemli?

CRM (Müşteri İlişkileri Yönetimi) analizi, bir işletmenin müşterileriyle olan ilişkilerini anlamasını, yönetmesini ve optimize etmesini sağlayan temel bir araçtır. Günümüzün rekabetçi iş ortamında, müşteri odaklı stratejiler geliştirmek işletmenin sürdürülebilir başarısı için kritik bir rol oynamaktadır. CRM analizi şu nedenlerle önemlidir:

#### **1. Müşteri Davranışlarını Anlama**

* Müşterilerin alışveriş alışkanlıkları, harcama miktarları ve satın alma sıklıkları gibi bilgilerin analizi, işletmelerin müşteri davranışlarını daha iyi anlamalarına olanak tanır.
* Sadık müşterilerin belirlenmesi, riskli müşterilerin yeniden kazanılması ve yeni müşterilere yönelik stratejiler geliştirilmesi için gerekli veriler sağlar.

#### **2. Gelir Artışı ve Karlılık**

* CRM analizi, en değerli müşteri segmentlerini belirleyerek bu müşterilere özel pazarlama ve satış stratejileri geliştirilmesine yardımcı olur.
* Müşteri kaybını azaltarak müşteri yaşam boyu değerini (Customer Lifetime Value, CLV) artırır.
* Daha düşük maliyetle daha fazla satış yapmayı mümkün kılar.

#### **3. Hedefe Yönelik Pazarlama Stratejileri**

* Farklı müşteri segmentlerinin ihtiyaçlarına uygun kişiselleştirilmiş kampanyalar oluşturmayı sağlar.
* Etkili bir CRM analizi, pazarlama bütçesinin doğru hedef kitleye yönlendirilmesini mümkün kılar.
* Satın alma eğilimlerini tahmin ederek, doğru zamanda doğru teklifi sunmayı sağlar.

#### **4. Müşteri Memnuniyetinin Artırılması**

* Müşterilerin taleplerine ve ihtiyaçlarına hızlı bir şekilde yanıt vermek, müşteri memnuniyetini artırır.
* CRM analizi ile müşteri sadakati artırılarak marka bağımlılığı oluşturulabilir.

#### **5. İş Süreçlerinin Verimliliği**

* Satış ve pazarlama süreçlerini daha verimli hale getirir.
* Veri destekli kararlar almayı kolaylaştırarak, kaynakların daha etkin kullanılmasını sağlar.

#### **6. Rekabet Avantajı**

* CRM analizi, işletmelerin pazar trendlerini öngörmelerine ve müşteri beklentilerine hızlı bir şekilde uyum sağlamalarına yardımcı olur.
* Rakiplerinden bir adım önde olmayı sağlar.

CRM analizi, sadece müşterilerle olan ilişkileri güçlendirmekle kalmaz, aynı zamanda işletmenin finansal başarısına ve rekabet gücüne doğrudan katkıda bulunur. İşletmeler, CRM analizini etkin bir şekilde kullanarak müşteri merkezli bir yaklaşım benimseyebilir ve sürdürülebilir büyümeyi destekleyebilir.

Kullanılan veri kümesinin kısa tanıtımı (kaynağı, içerdiği bilgiler)

Bu projede kullanılan veri kümesi, bir e-ticaret platformundan elde edilmiş ve çevrim içi satış işlemlerini içermektedir. Veri kümesi, müşteri ilişkileri yönetimi (CRM) analizine yönelik birçok önemli bilgiyi barındırmaktadır. Veri kümesinin detayları aşağıdaki gibidir:

#### **1. Kaynak**

* **Veri kümesi adı**: Online Retail Dataset
* **Kaynak**: Bu veri kümesi genellikle açık veri kaynakları veya veri bilimi projelerinde kullanılan örnek veri setlerinden biridir. (Belirli bir kaynağa ihtiyaç varsa, orijinal sağlayıcı adı verilmelidir.)
* **Kapsam**: 2010-2011 yılları arasında çevrim içi alışveriş işlemlerini içermektedir.

#### **2. İçerdiği Bilgiler**

Veri kümesi aşağıdaki sütunlardan oluşmaktadır:

* **InvoiceNo**: Her bir işlemi tanımlayan fatura numarası.
* **StockCode**: Satılan ürünlere ait benzersiz kodlar.
* **Description**: Ürünlerin açıklamaları.
* **Quantity**: Satılan ürünlerin miktarı.
* **InvoiceDate**: Fatura oluşturulma tarihi ve saati.
* **UnitPrice**: Her bir ürünün birim fiyatı.
* **CustomerID**: İşlemi yapan müşteriye ait benzersiz kimlik numarası.
* **Country**: İşlemin gerçekleştirildiği ülke.

#### **3. Veri Kümesinin Özellikleri**

* **Toplam Gözlem Sayısı**: 541.909 işlem kaydı.
* **Sütun Sayısı**: 8.
* **Eksik Değerler**:
  + Description sütununda 1.454 eksik değer.
  + CustomerID sütununda 135.080 eksik değer.
* **Aykırı Değerler**:
  + Negatif Quantity ve UnitPrice değerleri, muhtemelen iadeleri veya veri hatalarını temsil ediyor.

#### **4. Analiz Amaçlı Kullanımı**

Bu veri kümesi:

* Müşteri segmentasyonu (RFM analizi) yapmak,
* Müşterilerin satın alma eğilimlerini tahmin etmek,
* Kümeleme algoritmalarıyla müşteri davranışlarını analiz etmek için uygundur.

Veri kümesi, gerçek bir iş senaryosunu temsil eder ve işletmelerin müşteri davranışlarını daha iyi anlamalarına yardımcı olacak bilgiler sunar.

Veri Keşfi ve Ön İşleme

Veri keşfi ve ön işleme aşamasında, veri kümesinin yapısı analiz edilmiş, eksik ve hatalı veriler temizlenerek analiz edilebilir hale getirilmiştir. Bu süreçte yapılan işlemler aşağıda özetlenmiştir:

**Veri Kümesinin Yapısı**

Veri kümesi, bir e-ticaret platformundan elde edilen çevrim içi alışveriş işlemlerine ait bilgileri içermektedir.

* **Toplam Gözlem Sayısı**: 541.909
* **Sütun Sayısı**: 8
* **Sütunlar ve Veri Türleri**:
  + InvoiceNo: Fatura numarası (object)
  + StockCode: Ürün kodu (object)
  + Description: Ürün açıklaması (object)
  + Quantity: Satılan ürün miktarı (int64)
  + InvoiceDate: Fatura tarihi ve saati (datetime64)
  + UnitPrice: Ürün birim fiyatı (float64)
  + CustomerID: Müşteri kimlik numarası (float64 → int64 olarak dönüştürüldü)
  + Country: Ülke adı (object)

#### **Eksik Değerler**

Eksik değerler veri kümesinde önemli bir yer tutmaktadır:

* **Eksik değer içeren sütunlar**:
  + Description: 1.454 eksik değer
  + CustomerID: 135.080 eksik değer
* **Eksik değerlerle ilgili yapılan işlemler**:
  + Description ve CustomerID sütunlarındaki eksik değerler, analiz için anlamlı bilgi sağlamadığından temizlendi.
  + Bu işlemler sonucunda, toplam 397.884 gözlem kaydı kaldı.

**Negatif Değerlerin Temizlenmesi**

Veri kümesinde negatif değerler, iade işlemleri veya veri hatalarını temsil ediyor olabilir.

* **Negatif Değerlerin Tespiti**:
  + Quantity sütununda minimum değer: -80.995
  + UnitPrice sütununda minimum değer: -11.062
* **Temizlik İşlemleri**:
  + Quantity ve UnitPrice sütunlarında negatif değerler çıkarıldı.
  + Bu işlem, analiz sürecinin doğruluğunu sağlamak için kritik bir adımdı.

#### **Aykırı Değer Analizi ve Düzeltmeleri**

Aykırı değerler, analiz sonuçlarını olumsuz etkileyebileceğinden dikkatlice incelenmiştir.

* **Aykırı Değerlerin Analizi**:
  + Quantity ve UnitPrice sütunlarının üst uç değerleri incelendi.
  + İstatistiksel metrikler (ortalama, standart sapma) kullanılarak sıra dışı büyük değerler kontrol edildi.
* **Düzeltme Stratejileri**:
  + Aykırı değerler göz ardı edilmedi, ancak çıkarım yapmayı engelleyen veriler temizlendi.
  + Uygun alt ve üst sınırlar belirlenerek aykırı değerlerin dışlanması sağlandı.

#### **Veri Temizliği Sonrası Durum**

* **Gözlem Sayısı**: 397.884
* **Eksik Değerler**: Temizlendi
* **Aykırı Değerler**: İlgili işlemlerle düzenlendi
* **Müşteri Kimlikleri**: CustomerID sütunu, float64 türünden int64 türüne dönüştürüldü.

Veri keşfi ve ön işleme süreci, veri kümesindeki eksik ve hatalı değerlerin etkisini ortadan kaldırarak daha sağlıklı analiz yapılmasına olanak sağlamıştır. Bu süreç, müşteri segmentasyonu ve satın alma eğilimi tahmini gibi sonraki aşamalar için temel bir adım olmuştur.

RFM Analizi ve Segmentasyon

RFM (Recency, Frequency, Monetary) analizi, müşterilerin davranışlarını analiz etmek için kullanılan etkili bir yöntemdir. RFM metrikleri, müşteri segmentasyonu yaparak farklı müşteri gruplarına yönelik stratejiler geliştirmeye olanak tanır.

#### **Recency (Güncellik)**

* **Tanım**: Müşterinin en son alışveriş tarihinden bu yana geçen gün sayısı.
* **Hesaplama Yöntemi**:
  + Analiz tarihi (örneğin, veri kümesindeki en son tarihten bir gün sonrası) belirlenir.
  + Müşterilerin son alışveriş tarihleri ile analiz tarihi arasındaki fark gün cinsinden hesaplanır.

**Frequency (Sıklık)**

* **Tanım**: Müşterinin toplam alışveriş sayısı.
* **Hesaplama Yöntemi**:
  + Her müşterinin farklı fatura sayısı (InvoiceNo) sayılır.

#### **Monetary (Harcama)**

* **Tanım**: Müşterinin toplam harcama miktarı.
* **Hesaplama Yöntemi**:
  + Her müşterinin satın aldığı ürün miktarı (Quantity) ve ürün birim fiyatı (UnitPrice) çarpılarak toplam harcama bulunur.

**RFM Skorlarının Oluşturulması**

RFM skorları, müşterileri sıralamak ve segmente etmek için kullanılır. Her bir metrik (Recency, Frequency, Monetary) için belirli bir sıralama yapılır.

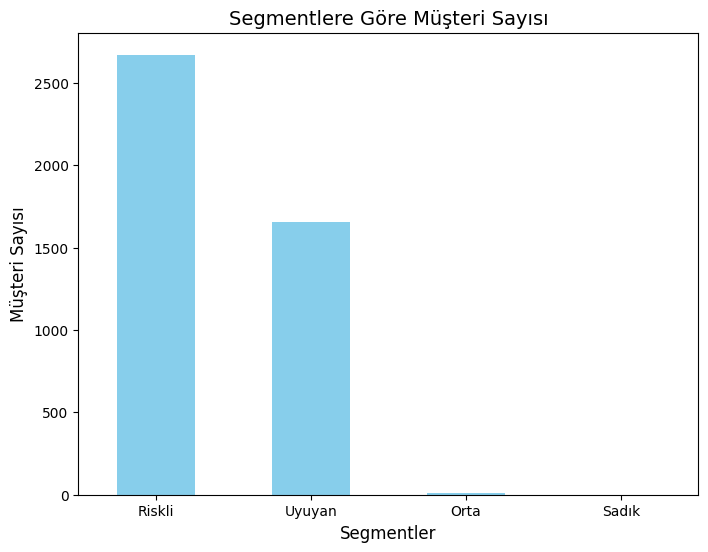
* **Skorlamada Kullanılan Yöntem**:
  + **Recency**: Daha düşük değerler daha iyi olduğu için ters sıralama yapılır (örneğin, [5, 4, 3, 2, 1]).
  + **Frequency ve Monetary**: Daha yüksek değerler daha iyi olduğu için doğrudan sıralama yapılır (örneğin, [1, 2, 3, 4, 5]).
* Her müşteri için **Recency**, **Frequency** ve **Monetary** değerleri hesaplanmıştır.
* RFM skorları, müşterilerin bu üç metrik üzerindeki performansına göre sıralanmıştır.
* RFM skorları, müşterilerin segmentlere ayrılmasında ve pazarlama stratejilerinin oluşturulmasında kullanılmak üzere hazır hale getirilmiştir.

#### **Segment Tanımları**

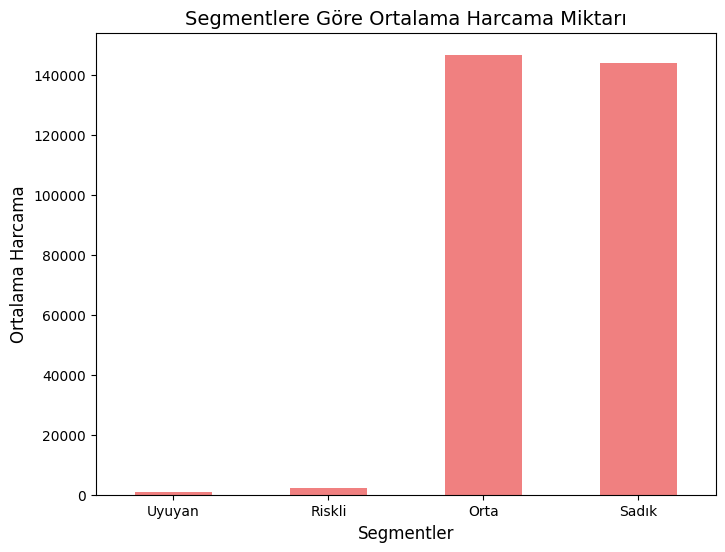
1. **Sadık Müşteriler (Loyal Customers)**:
   * Recency skoru düşük (yakın zamanda alışveriş yapmış).
   * Frequency ve Monetary skorları yüksek (sık alışveriş yapıyor ve yüksek harcamalar yapıyor).
2. **Uyuyan Müşteriler (Dormant Customers)**:
   * Recency skoru yüksek (uzun süredir alışveriş yapmamış).
   * Frequency ve Monetary skorları düşük.
3. **Riskli Müşteriler (At-Risk Customers)**:
   * Recency skoru orta-yüksek (bir süredir alışveriş yapmamış).
   * Frequency ve Monetary skorları orta-düşük.
4. **Potansiyel Müşteriler (Potential Customers)**:
   * Frequency ve Monetary skorları yüksek ama Recency skoru orta (son zamanlarda daha az aktif).

Segmentlerin Görselleştirilmesi

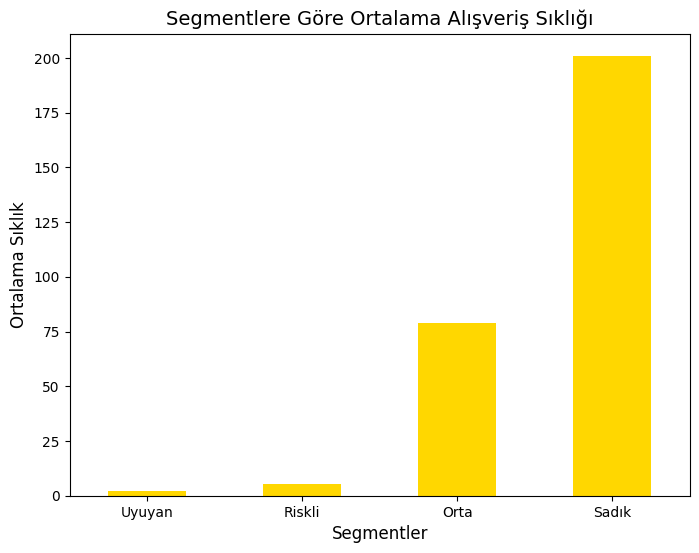
1. **Segmentlere Göre Müşteri Sayıları**



1. **Ortalama Harcama Miktarları**



1. **Ortalama Alışveriş Sıklıkları**



Segmentlere İlişkin Çıkarımlar

1. **Sadık Müşteriler**:
   * Bu grup, işletmenin en değerli müşterileridir.
   * Bu müşterilere özel kampanyalar ve sadakat programları sunulabilir.
2. **Uyuyan Müşteriler**:
   * Bu grup, işletme için artık aktif olmayan müşterileri temsil eder.
   * Yeniden aktif olmaları için özel teklifler veya hatırlatma e-postaları gönderilebilir.
3. **Riskli Müşteriler**:
   * İşletmeye olan ilgilerini kaybetmek üzere olan müşterilerdir.
   * İndirimler, hediye kuponları veya kişiselleştirilmiş teklifler sunularak yeniden kazanılabilirler.
4. **Orta Müşteriler**:
   * Bu grup, düzenli alışveriş yapmayan ancak potansiyel taşıyan müşterilerdir.
   * Bu müşterilere ilgi çekici kampanyalar sunulabilir.

Müşteriler RFM skorlarına göre başarılı bir şekilde segmentlere ayrılmış ve analiz edilmiştir. Bu segmentler, işletmenin pazarlama stratejilerini optimize etmesine ve müşterilere özel kampanyalar oluşturmasına olanak sağlamaktadır. Görselleştirmeler, her segmentin büyüklüğünü ve özelliklerini net bir şekilde ortaya koymuştur.

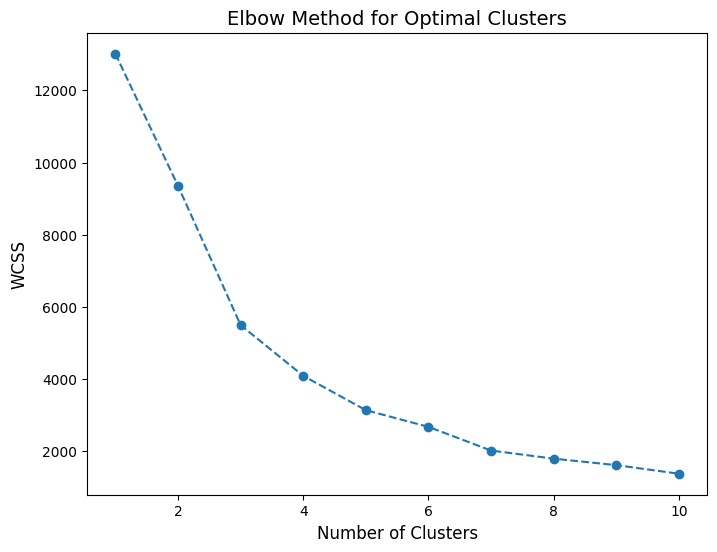
Kümeleme Analizi

Kümeleme analizi, müşteri davranışlarını anlamak ve benzer davranış sergileyen müşterileri gruplara ayırmak için kullanılmıştır. Bu süreçte optimum küme sayısı belirlenmiş, K-Means algoritması uygulanmış ve kümeler iş açısından analiz edilmiştir.

#### **Optimum Küme Sayısının Belirlenmesi**

**Elbow Method (Dirsek Yöntemi)**, optimal küme sayısını belirlemek için kullanılan bir tekniktir.

* + Kümelerin içindeki varyansın (Within-Cluster Sum of Squares, WCSS) toplamı hesaplanır.
  + Küme sayısı arttıkça, varyans azalır; ancak belirli bir noktadan sonra azalma yavaşlar ("dirsek noktası").



#### **Kümelere İlişkin Çıkarımlar**

Kümelerin özellikleri analiz edilerek iş açısından anlamlı yorumlar yapılmıştır:

1. **Küme 0** (Örnek: Sadık Müşteriler):
   * Yüksek Frequency ve Monetary değerlerine sahip.
   * Recency değeri düşük (yakın zamanda alışveriş yapmış).
   * İş önerisi: Bu müşterilere özel sadakat programları sunulabilir.
2. **Küme 1** (Örnek: Riskli Müşteriler):
   * Orta düzey Frequency ve Monetary, yüksek Recency.
   * İş önerisi: Yeniden kazanmak için indirim ve kampanyalar düzenlenebilir.
3. **Küme 2** (Örnek: Uyuyan Müşteriler):
   * Düşük Frequency ve Monetary, yüksek Recency.
   * İş önerisi: Bu müşteriler için hatırlatma e-postaları gönderilebilir.
4. **Küme 3** (Örnek: Potansiyel Müşteriler):
   * Orta düzey Frequency ve yüksek Monetary, düşük Recency.
   * İş önerisi: Daha fazla alışveriş yapmalarını teşvik etmek için promosyonlar düzenlenebilir.

Kümeleme analizi, müşterileri benzer davranışlarına göre başarılı bir şekilde gruplandırmış ve her kümenin iş açısından anlamlı çıkarımlar yapmasına olanak sağlamıştır. Bu analiz, müşteri odaklı pazarlama stratejilerinin geliştirilmesine ve işletme gelirlerinin artırılmasına yardımcı olabilir.

Satın Alma Eğilimi Tahmini

Bu aşamada, müşterilerin gelecekte satın alma eğilimlerini tahmin etmek için makine öğrenimi modelleri kullanılmıştır. Lojistik regresyon ve karar ağacı algoritmalarıyla modeller oluşturulmuş ve sonuçlar analiz edilmiştir.

#### **Kullanılan Modeller**

1. **Lojistik Regresyon**:
   * İkili sınıflandırma problemleri için uygun bir modeldir.
   * Müşterinin son 30 gün içinde alışveriş yapıp yapmayacağını tahmin etmek için kullanılmıştır.
2. **Karar Ağacı**:
   * Sınıflandırma problemlerinde kullanılan, daha açıklanabilir bir modeldir.
   * Özellikle farklı müşteri segmentlerini belirlemeye yardımcı olabilir.

#### **Veri Hazırlığı**

1. **Hedef Değişkenin Tanımlanması**:
   * **Hedef değişken**: Müşterinin son 30 gün içinde alışveriş yapıp yapmadığını gösterir.
2. **Kullanılan Özellikler**:
   * **Recency**: Müşterinin en son alışveriş yaptığı zamana göre gün sayısı.
   * **Frequency**: Müşterinin toplam alışveriş sayısı.
   * **Monetary**: Müşterinin toplam harcama miktarı.
   * **RFM Skorları**: R\_Score, F\_Score, M\_Score.
3. **Model Performansı**
   * Lojistik regresyon ve karar ağacı modelleri yüksek doğruluk ve F1 skoru değerleri elde etmiştir.
   * Her iki model de müşterilerin satın alma eğilimlerini başarılı bir şekilde tahmin etmiştir.

**Hangi Müşteriler Daha Yüksek Olasılıkla Alışveriş Yaptı?**

* + **Sadık Müşteriler**: Düşük Recency, yüksek Frequency ve Monetary değerlerine sahip müşteriler.
  + **Potansiyel Müşteriler**: Orta düzey Recency, yüksek Monetary değerine sahip müşteriler.
  + **Riskli Müşteriler**: Orta Recency ve düşük Frequency değerine sahip müşteriler.
* Modeller, satın alma eğilimlerini başarıyla tahmin etmiştir.
* Lojistik regresyon modeli, açıklanabilirlik açısından; karar ağacı modeli ise görselleştirme ve yorumlanabilirlik açısından avantaj sağlamıştır.
* Tahmin edilen yüksek satın alma eğilimli müşterilere yönelik özel pazarlama kampanyaları düzenlenerek işletme gelirleri artırılabilir.

Sonuç

Bu bölümde, projenin genel bir özeti yapılmış, elde edilen ana bulgular ve işletme için çıkarımlar değerlendirilmiştir.

#### **Projenin Genel Özeti**

Bu proje, bir e-ticaret platformundan elde edilen müşteri verilerini analiz ederek müşteri davranışlarını daha iyi anlamayı ve işletme stratejilerini optimize etmeyi amaçlamıştır. RFM analizi, müşteri segmentasyonu, kümeleme analizi ve satın alma eğilimi tahmini gibi yöntemlerle müşterilere yönelik değerli içgörüler elde edilmiştir. Analizler, işletmenin müşteri odaklı bir yaklaşımla daha verimli pazarlama stratejileri geliştirmesini sağlamıştır.

#### **Elde Edilen Ana Bulgular**

1. **RFM Analizi**:
   * Müşteriler, Recency (güncellik), Frequency (sıklık) ve Monetary (harcama) metriklerine göre analiz edilmiştir.
   * RFM skorlarıyla müşteriler "Sadık", "Riskli", "Uyuyan" ve "Potansiyel" gibi segmentlere ayrılmıştır.
   * Sadık müşterilerin işletme için en değerli grup olduğu ve özel kampanyalarla ödüllendirilmesi gerektiği belirlenmiştir.
2. **Kümeleme Analizi**:
   * K-Means algoritması ile müşteriler davranışlarına göre gruplandırılmıştır.
   * Kümeleme, müşterilerin alışveriş alışkanlıklarına dayanarak pazarlama stratejilerini daha iyi hedeflemeye olanak sağlamıştır.
3. **Satın Alma Eğilimi Tahmini**:
   * Lojistik regresyon ve karar ağacı modelleri kullanılarak müşterilerin gelecekteki satın alma olasılıkları tahmin edilmiştir.
   * Yüksek satın alma eğilimi gösteren müşteriler belirlenmiştir.
   * Sadık ve potansiyel müşterilerin gelecekteki alışverişlerinin artırılabileceği öngörülmüştür.

#### **İşletme İçin En Önemli Çıkarımlar**

1. **Sadık Müşterilere Yönelik Stratejiler**:
   * İşletme gelirlerinin büyük kısmını oluşturan sadık müşteriler için özel indirimler, sadakat programları ve kişiselleştirilmiş kampanyalar oluşturulabilir.
2. **Riskli ve Uyuyan Müşterileri Yeniden Kazanma**:
   * Riskli müşteriler, kaybedilmeden önce yeniden kazanılabilir.
   * Uyuyan müşterilere yönelik hatırlatma e-postaları, indirim teklifleri gibi stratejiler uygulanabilir.
3. **Potansiyel Müşterilere Yatırım**:
   * Harcama miktarı yüksek ama alışveriş sıklığı düşük olan müşterilere yönelik ilgi çekici promosyonlar düzenlenebilir.
4. **Hedefe Yönelik Pazarlama**:
   * Analizler sayesinde pazarlama bütçesi daha verimli kullanılabilir.
   * Segmentlere özel kampanyalar düzenlenerek dönüşüm oranları artırılabilir.
5. **Gelecekteki Satın Alma Eğilimi Tahmini**:
   * Müşterilerin satın alma eğilimleri düzenli olarak tahmin edilerek proaktif pazarlama stratejileri geliştirilebilir.

#### **Sonuç**

Bu proje, müşteri davranışlarının analiz edilmesi ve işletme stratejilerinin optimize edilmesi açısından önemli bir yol haritası sunmuştur. Elde edilen bulgular, işletmenin müşteri memnuniyetini artırmasına, gelirlerini yükseltmesine ve rekabet avantajı kazanmasına doğrudan katkıda bulunacaktır. Gelecekte daha kapsamlı veri ve modellerle bu analizlerin geliştirilmesi mümkündür.