

BTK Akademi Datathon

Eylül 2023

İçindekiler

- [Hakkımda](#)
- [Keşifsel Veri Analizi](#)
- [Ön İzlenim](#)
- [İlişkisel Veri Analizi](#)
- [Kümeleme Analizi](#)
- [Hiyerarşik Kümeleme Yöntemi](#)
- [3D Görselleştirme Yöntemi](#)
- [K-Means Kümeleme Yöntemi](#)
- [Hedef Değişken Analizi](#)
- [Öbek 1: Mavi Yakalı İşçiler](#)
- [Öbek 2: Ev Hanımları](#)
- [Öbek 3: Profesyonel Erkekler](#)
- [Öbek 4: Beyaz Yakalılar](#)
- [Öbek 5: Teknoloji ve Spor Severler](#)
- [Öbek 6: Yaşılı Elit Alışverişçiler](#)
- [Öbek 7: Akademisyenler](#)
- [Öbek 8: Öğrenciler](#)
- [Pazarlama Stratejileri](#)
- [Genel Stratejiler](#)
- [Değerli Madenleri İşleyelim](#)
- [Tahminleme Çalışması](#)
- [Tahminleme Sonuçları & En İyi Model](#)
- [ML Sonuçları](#)
- [Hiperparametre Ayarı](#)
- [Teşekkür Ederim](#)

Hakkımda



Merhabalar ben Ramazan, istatistik, veri bilimi ve yapay zeka alanlarına meraklı birisiyim.

Hacettepe Üniversitesi İstatistik Bölümünü şeref öğrencisi olarak bitirdim. Eğitim hayatım boyunca yazılar yazdım ve projeler üreterek kendimi bu alanlarda geliştirdim.

Hakkımda detaylı bilgi için aşağıdaki bağlantıları ziyaret edebilirsiniz...

Kişisel Web Sayfam

LinkedIn

GitHub

Keşifsel Veri Analizi

İlişkisel Veri Analizi

Veriler kategorik ve numerik veriler birlikte kullanılarak veriler arası gizli ilişkilerin çıkarılması amaçlanmıştır.



Ön İzlenim

Veriler tek tek ele alınmış olup veri seti hakkında genel bir yargı elde etmek için gerçekleştirilmiştir.

Kümeleme Analizi

Hangi müşterilerin küme oluşturduklarını, hangi verilerin birbirine daha çok benzediğini bulmak için yapılmıştır.

Ön İzlenim

Bu kısmında her bir özellik tek başına eline alınmış olup veri seti hakkında genel bir yargı elde edilmeye çalışılmıştır.



Demografik Yapı

1. Cinsiyet dağılımı **eşit**.
2. Medeni durum dengesi **eşit**.
3. Müşterilerimiz çoğunlukla **eğitimli**.
4. **Düzenli** bir iş sahibi.
5. Küçük-büyük şehirler.
6. Gelir dağılımı geniş.



Alışveriş Tutumu

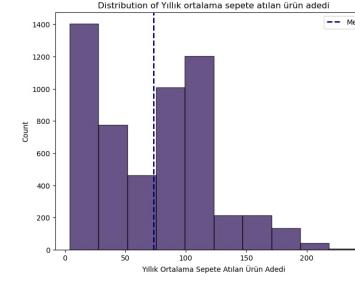
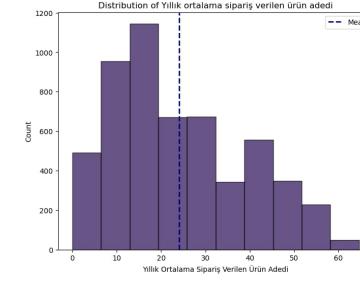
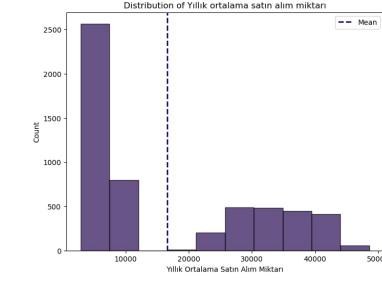
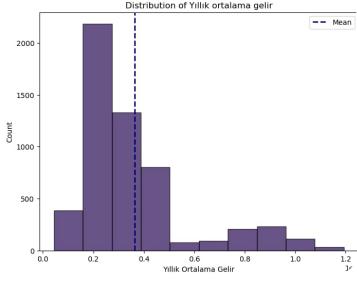
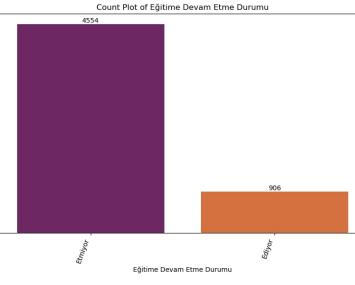
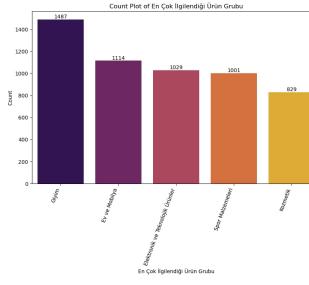
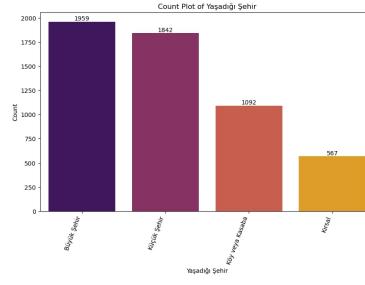
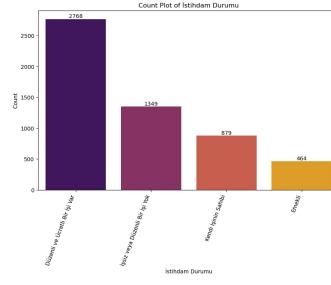
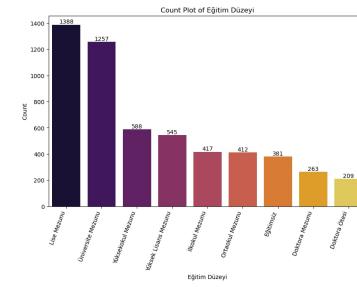
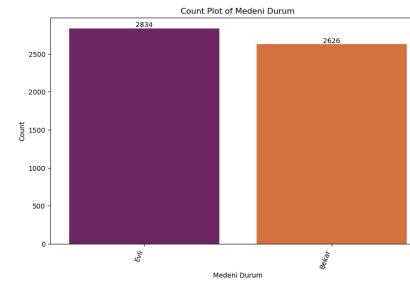
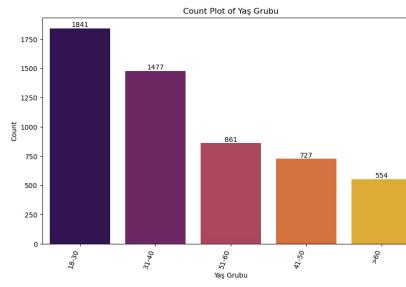
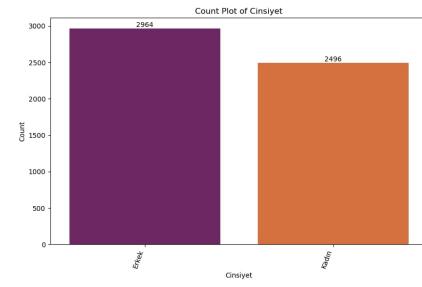
1. Alışveriş kategorisinde **her kategoriden** satışlar gerçekleşiyor. **En fazla giyim** en az ise kozmetik kategorisi.
2. Yıllık ortalama **satın alım** miktarı **iki ucta** yer alıyor. Perakende satın alan müşterilerin varlığından söz edilebilir.



Alışveriş Alışkanlıkları

1. Sipariş edilen ürün adedi **Ki-Kare** dağılımına benziyor.
2. Sepete atılan ürün adedine bakıldığıda bazı müşterilerimiz **internet sitemizde çok geziniyor**.

Galeri



İlişkisel Veri Analizi

Bu kısımda kategorik veriler numerik veriler ile birlikte analiz edilerek veriler arasındaki ilişkiler analiz edilmiştir.

Gençler

Genç nüfus **sepetine çok fazla ürün** eklese de satın **alım miktarı çok düşük**. Gençler ekonomik olarak sıkıntılı olduğu için alamıyor olabilir.

Bekarlık Sultanlık Mıdır?

Medeni durum dengesi eşit olmasına karşın bekarlar evlilere nazaran daha fazla sayı ve tutarda alışveriş yapıyorlar. **Bekarlar daha gözde** müşteriler denilebilir.

Öğren, Kazan ve Satın Al

Korelasyon analizine gerek duymaksızın, **eğitim düzeyi arttıkça** gelir seviyesi de artmaktadır. Satın **alım miktarı** ise **gelir düzeyiyle** doğru orantılı olarak artmaktadır.

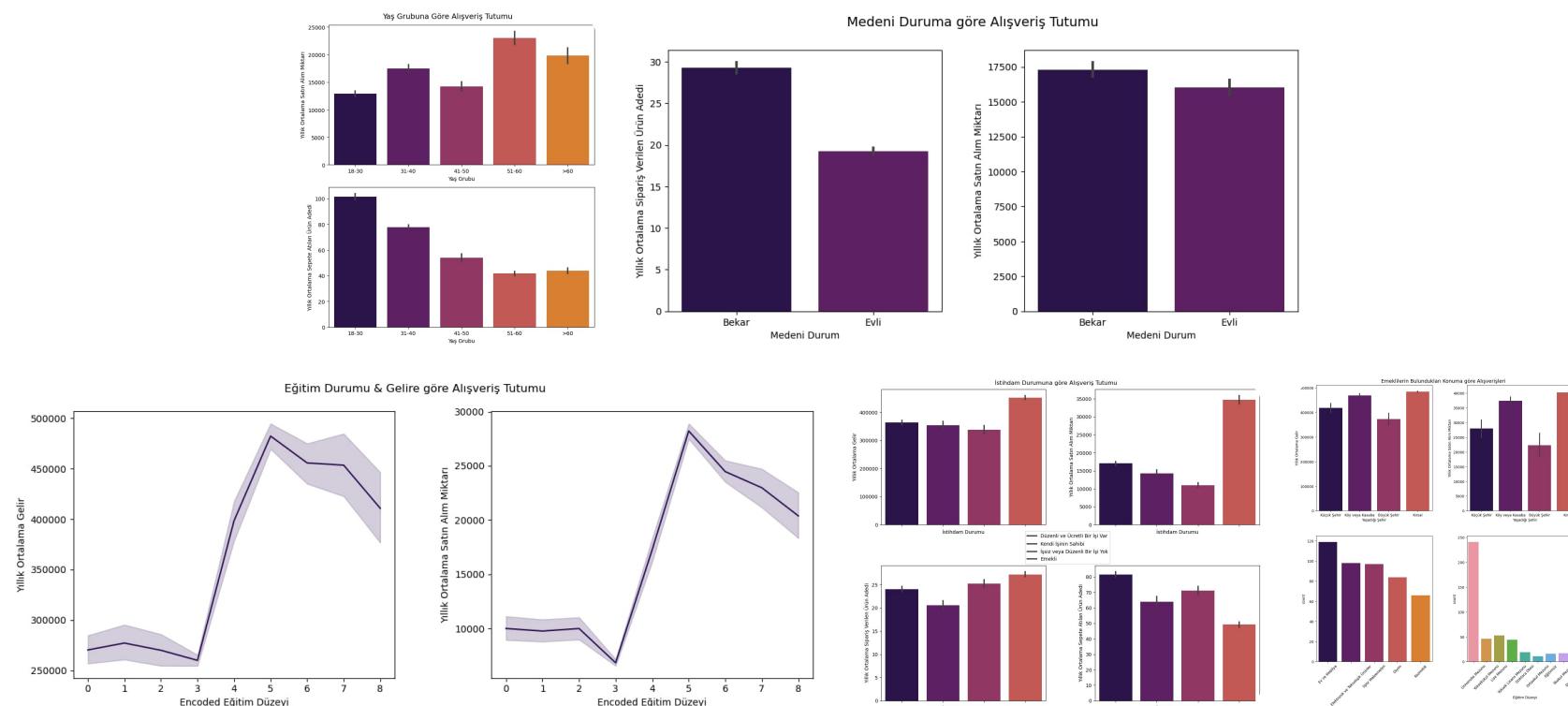
İstihdamda Emekliler

Diğer istihdam düzeyleri arasında **sayıları en az** olmasına rağmen emekliler, yıllık ortalama gelirine oranla **satın alım miktarı en fazla** olan müşterilerimiz.

Ah Şu Emekliler

En fazla alım yapan emekli müşterilerimiz **kırsal, köy ve kasabalarda** yaşamaktalar. Eğitim düzeyleri yüksek. Ev ve mobilya başta olmak üzere her kategoriden alışveriş yapıyorlar.

Galeri



Kümeleme Analizi

Hangi müşterilerin birbirine benzediğini bulmak için gerçekleştirilmiş bu analizde, amaca giden yolda üç yöntem kullanılmıştır. Küme sayısı en iyi tahmin skorunu sağlayacak şekilde ayrıca denenecektir.



Hiyerarşik Kümeleme

- Öklit uzaklığı ve **ward** bağlantı tipi.
- Küme sayısı 2 ve 4.



3D Görselleştirme

Öklit uzaklığı ile elde edilen uzaklık matrisi kullanılarak elde edilen görselleştirmelerden en iyi **küme sayısı 4** olarak belirlendi.

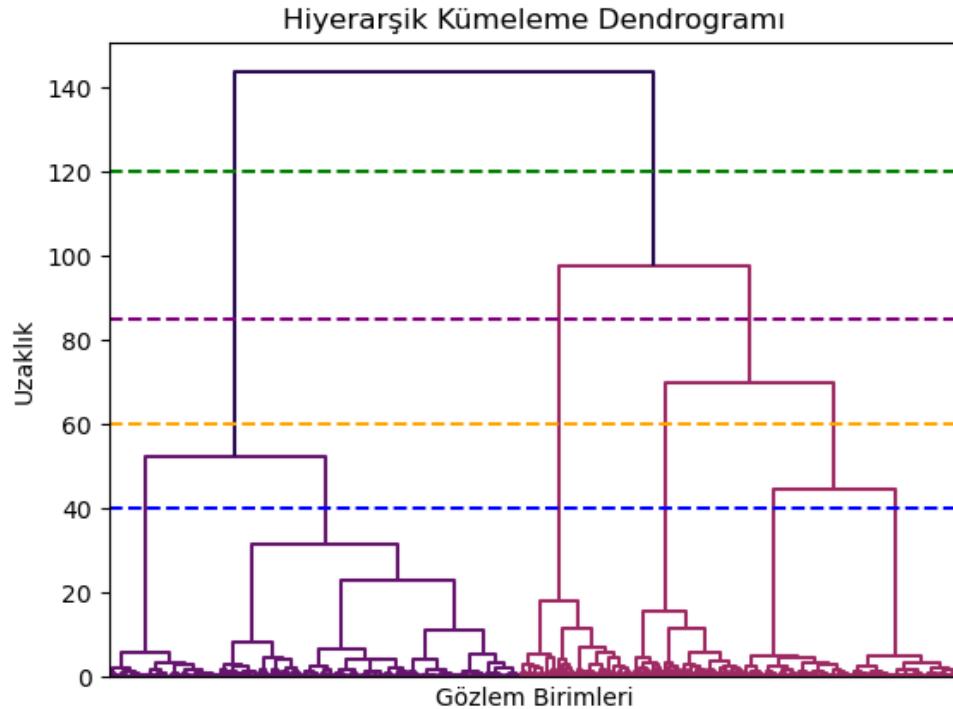


K-Means

Algoritma loop içerisinde bırakılarak elde edilen sonuçlar *Distortion Score* metriği ile görselleştirildi ve **küme sayısı 7** elde edildi.



Hiyerarşik Kümeleme Yöntemi



Neden **ward** bağlantı tipi?

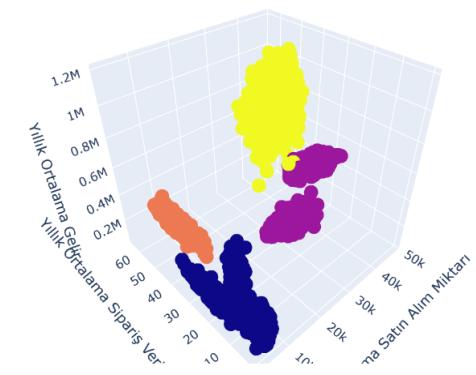
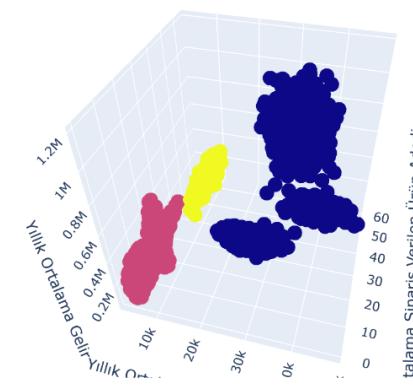
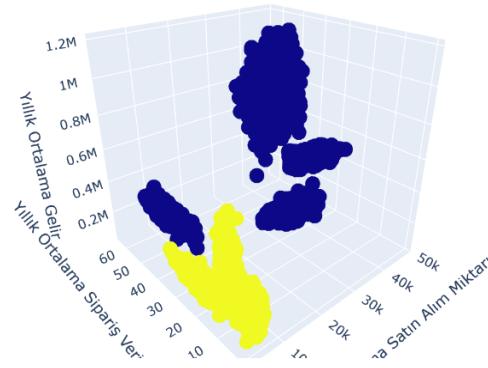
1. Küme içi **varyansın artışını minimize** ederek birleştirme yapar. Bu bizim **daha homojen ve dengeli kümelere** sahip olmamızı sağlar.
2. Varyans artışını minimize ederek bağlantı yapılması, birbirine **daha çok benzer kümeler** oluşturur.
3. Eğer küme içi varyans düşükse, aynı kümeye ait **veri noktaları birbirine yakın** ve benzer özelliklere sahiptir

3D Görselleştirme Yöntemi

Bu yöntem ile inceleme yapılırken **Agglomerative Clustering** ile öklit uzaklığı kullanılarak elde edilen distance matrix verilerinin 3D görselleştirmeleri ile küme sayısı belirlenmeye çalışılmıştır.

- **Son görselleştirmedeki mor küme** 2 farklı küme gibi görünse de k=5 olacak şekilde denedığımde mor küme içerisindeki iki küçük küme arası mesafe önemli olarak bulunmadı. Tahmin modellerinde de k=4 k=5,6 ve 7'den çok daha iyi performans sergiledi.

Galeri

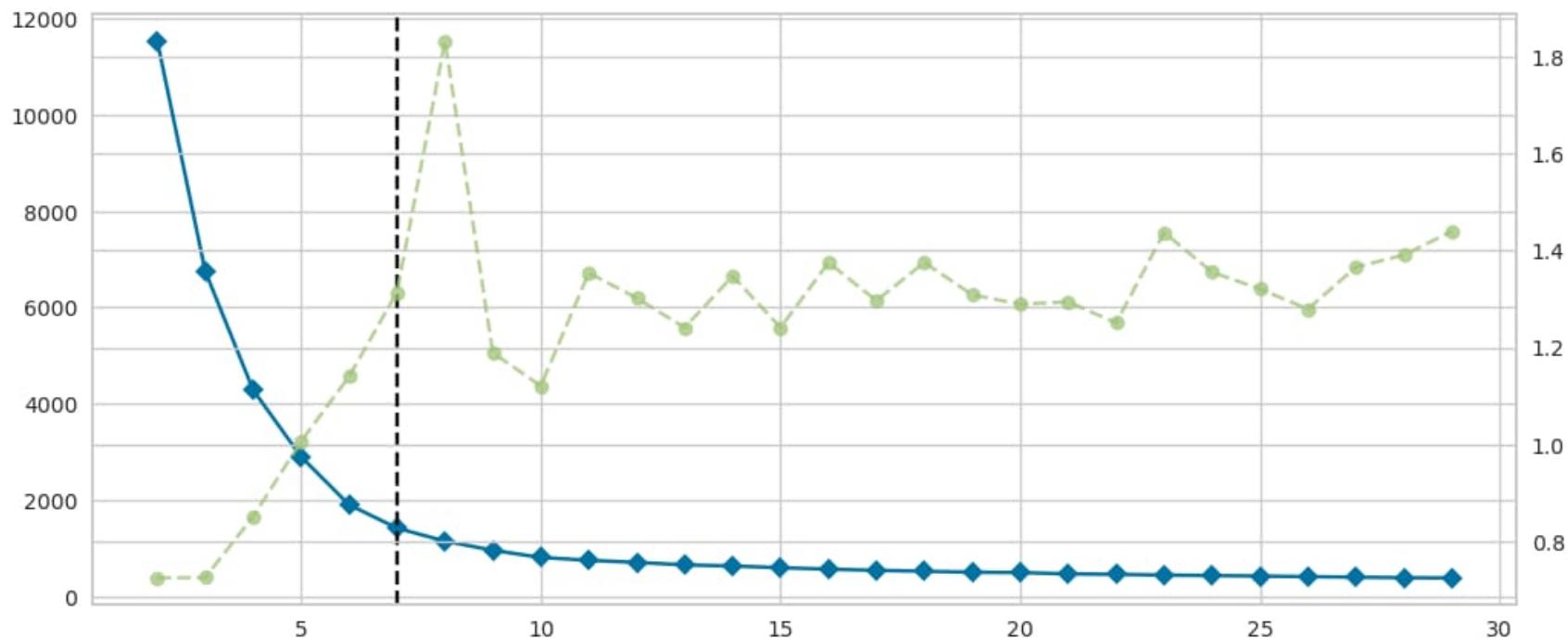


K-Means Kümeleme Yöntemi

K-Means algoritması ile küme sayısını 2'den 30'a kadar olacak şekilde döngüye bırakıldı ve elde edilen sonuçların değerlendirilmesinde başarı metriği olarak **Distortion Score** kullanıldı.

$$\text{Distortion Score} = \sum_{i=1}^n \min_{j=1}^k \|x_i - c_j\|^2$$

Distortion Score: Her bir veri noktasının küme merkezine olan uzaklığını bulur, dolayısıyla ne kadar küçükse o kadar iyidir.



[Kümeleme Yöntemlerine Geri Dön](#)

Hedef Değişken Analizi

Hedef değişken olan müşteri öbeklerinin analiz edilmesinde her bir öbek görselleştirilerek öbek hakkında genel kavramlar çıkarıldı ve yorumlandı.

Öbeklerin analizi sonucunda her bir öbeğe ilişkin alınacak aksiyonlar belirlenirken en değerli üç müşteri sınıfı için nitelikli stratejiler gerçekleştirılmıştır.

Genel Yapı



Ortalama Gelir

363.571 ₺



Ortalama Satın Alım Miktarı

16.617 ₺



Ortalama Sipariş Adedi

24 Adet



Sepete Atılan Ürün Adedi

73 Adet

[Anlamlandırmalara Git](#)

[Stratejilere Atla](#)

Öbek 1: Mavi Yakalı İşçiler



Ortalama Gelir

229.514 ₺

Ortalama Satın Alım Miktarı

6.047 ₺

Ortalama Sipariş Adedi

6 Adet

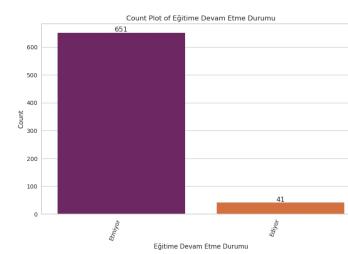
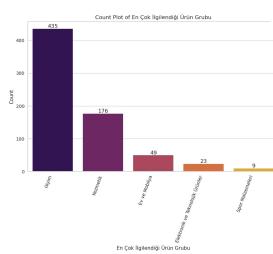
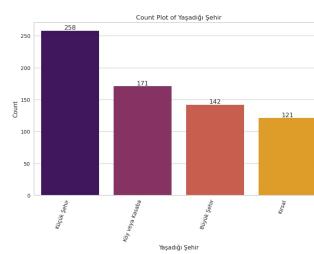
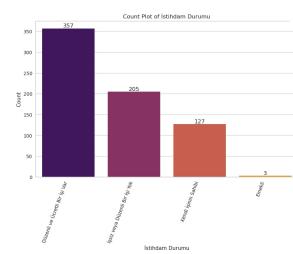
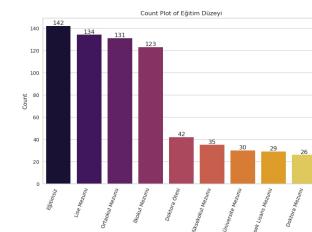
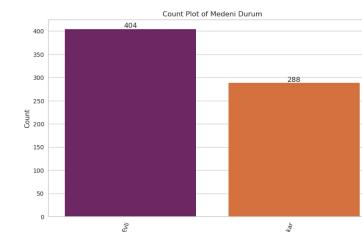
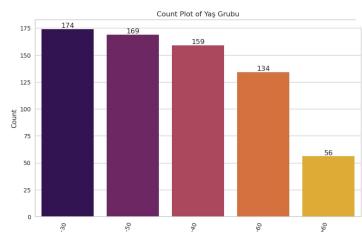
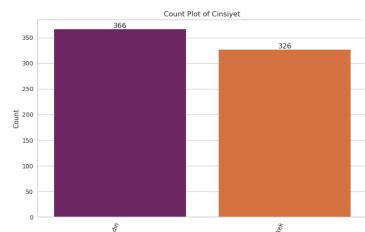
Sepete Atılan Ürün Adedi

13 Adet

Bu öbek, **cinsiyet dağılımı eşit** olan mavi yakalı işçi sınıfına ait. **Genç ve orta yaşılı** işçilerden oluşuyorlar. **Evli** sayısı bekar sayısından nispeten fazla, bu da bu grubun **ailelerden** olduğunu düşündürebilir. **Eğitim seviyesi düşük** ve çoğunluğu düzenli ve ücretli bir işte çalışıyor. **Küçük şehir, köy ve kasabalarda** yaşamayı tercih ediyorlar ve en çok **giyim kategorisine** para harcıyorlar.



Galeri



Öbek 2: Ev Hanımları



Ortalama Gelir
172.828 ₺



Ortalama Satın Alım Miktarı
4.996 ₺



Ortalama Sipariş Adedi
20 Adet

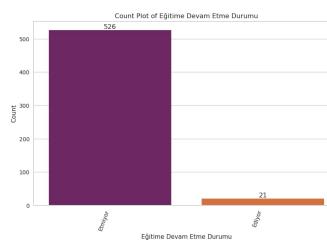
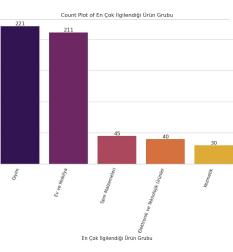
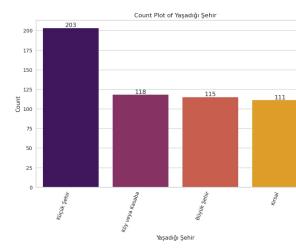
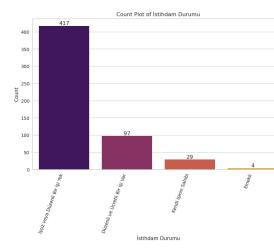
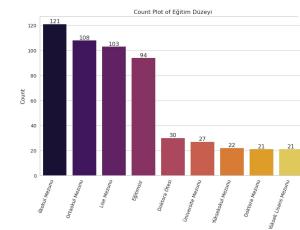
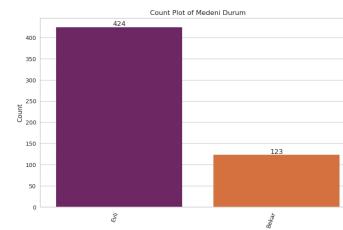
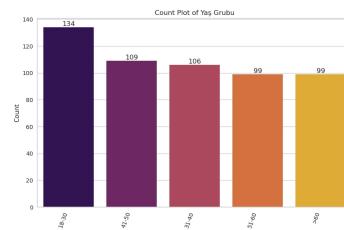
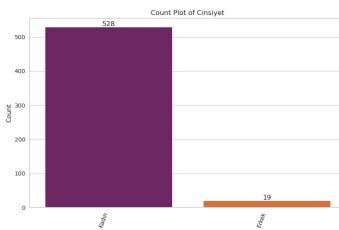


Sepete Atılan Ürün Adedi
51 Adet

Bu öbek, büyük çoğunluğu **kadınlardan** oluşan ve geniş yaş aralığına dağılmuş bir grup ev hanımını temsil ediyor. **Eğitim seviyesinin düşük** olması ve **işsiz veya düzenli bir işleri olmaması**, bu grubun **ekonomik olarak kısıtlı** kaynaklara sahip olduğunu düşündürebilir. **Küçük şehir, köy ve kasabalarda** yaşıyorlar ve **giyim, ev ve mobilya** alışverişini yapıyorlar. Ayrıca, eğitime devam etmiyorlar.



Galeri



Öbek 3: Profesyonel Erkekler



Ortalama Gelir
327.899 ₺



Ortalama Satın Alım Miktarı
7.361 ₺



Ortalama Sipariş Adedi
15 Adet

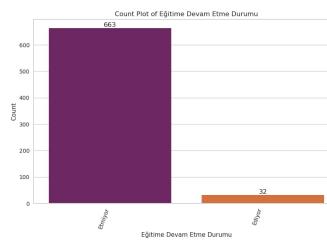
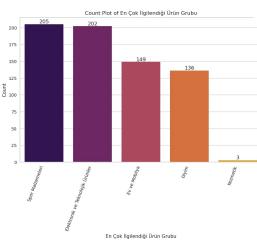
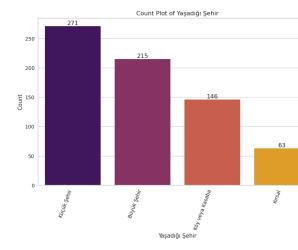
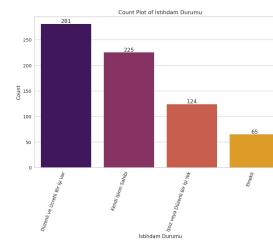
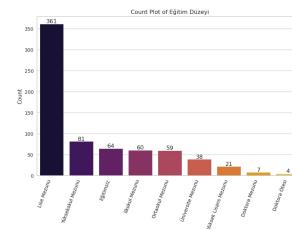
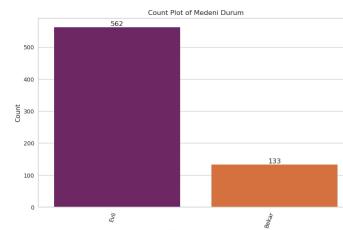
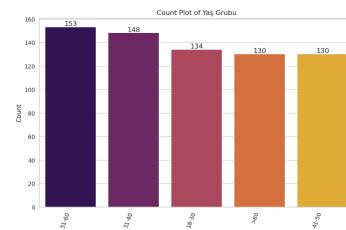
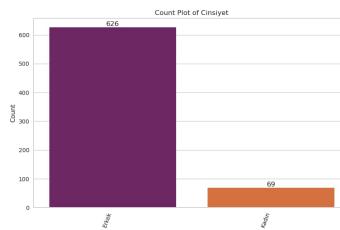


Sepete Atılan Ürün Adedi
22 Adet

Bu öbek, **erkeklerin** ağırlıkta olduğu, **farklı yaş** gruplarından, **evli ve lise mezunu** bir grup müşteriyi temsil ediyor. Çoğunluğu **düzenli bir işte** çalışıyor veya **kendi işine** sahip. Hem küçük hem de büyük **şehirlerde** yaşayan bu müşteriler, **spor malzemeleri ve teknoloji ürünlerini** alışverişine ilgi gösteriyorlar. Kozmetik ürünlerine olan ilgileri çok düşük. Eğitimlerini sürdürmeme eğilimindeler.



Galeri



Öbek 4: Beyaz Yakalılar



Ortalama Gelir
817.976 ₺



Ortalama Satın
Alım Miktarı
32.405 ₺



Ortalama Sipariş
Adedi
39 Adet

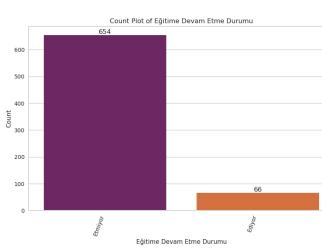
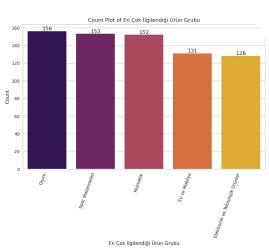
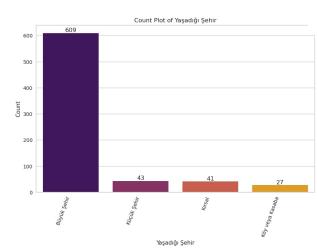
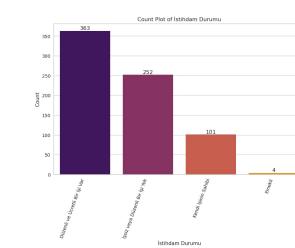
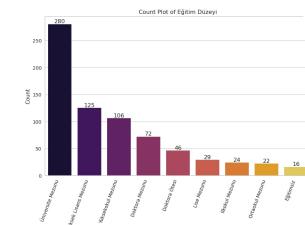
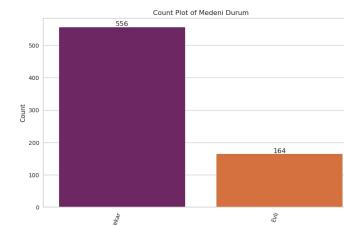
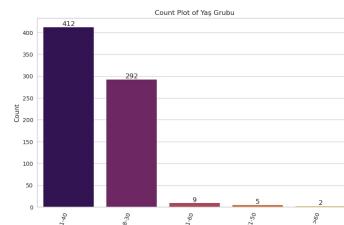
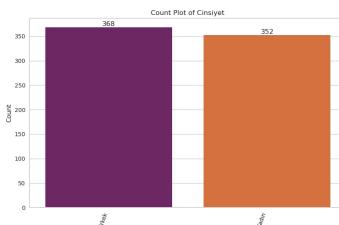


Sepete Atılan Ürün
Adedi
98 Adet

Cinsiyet dağılımı **eşit** olan bu öbek, **genç ve orta yaş** aralığındaki bireylerden oluşuyor. Çoğunluğu **bekar** ve **yüksek eğitim** seviyesine sahip. **İş durumu karmaşık**; bir kısmı **düzenli bir işte** çalışırken, diğerleri **işsiz** veya düzenli bir işleri yok. **Büyük şehirlerde** yoğunlaşıyorlar ve **çok çeşitli ürünler** kategorilerinden alışveriş yapıyorlar. Eğitimlerine devam etmeme eğilimindeler.



Galeri



Öbek 5: Teknoloji ve Spor Severler



Ortalama Gelir
222.224 ₺



Ortalama Satın
Alım Miktarı
4.542 ₺



Ortalama Sipariş
Adedi
11 Adet

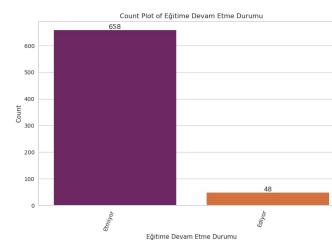
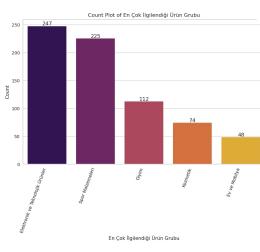
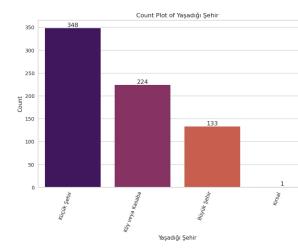
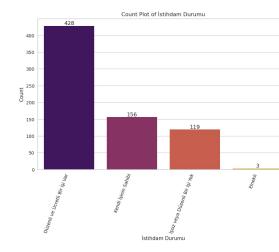
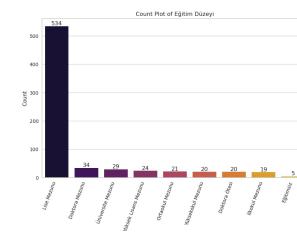
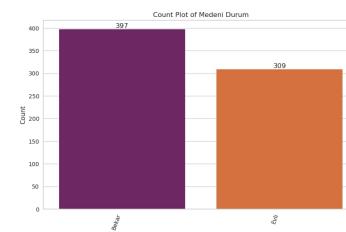
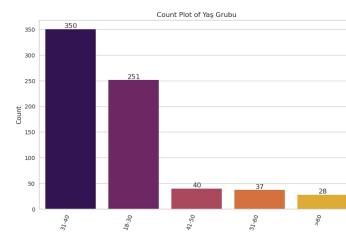
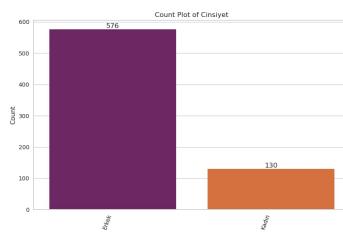


Sepete Atılan Ürün
Adedi
100 Adet

Bu öbek, **erkeklerin** ağırlıkta olduğu **orta yaşılı** bireylerden oluşuyor. Evlilik durumu neredeyse eşit dağılıyor ve genellikle **lise mezunlarından** oluşuyorlar. **Düzenli iş sahibi** olan çokunluk, **küçük şehirlerde** ve **kasabalarda** yaşıyor, ancak kırsal bölgelerde yaşayan yok. **Elektronik ve teknoloji aletleri ile spor malzemelerine** olan ilgileri öne çıkıyor. Eğitime devam etme eğiliminde değiller.



Galeri



Öbek 6: Yaşlı Elit Alışverişçiler



Ortalama Gelir
480.118 ₺



Ortalama Satın
Alım Miktarı
39.440 ₺



Ortalama Sipariş
Adedi
30 Adet

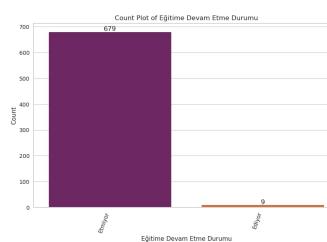
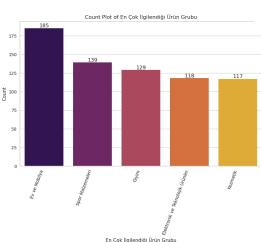
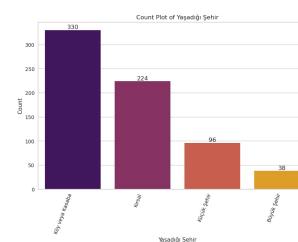
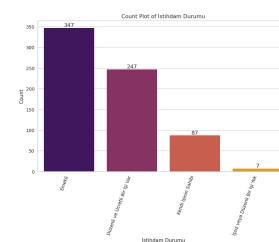
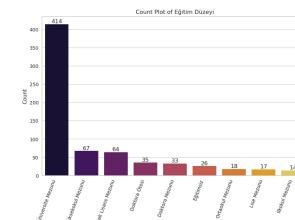
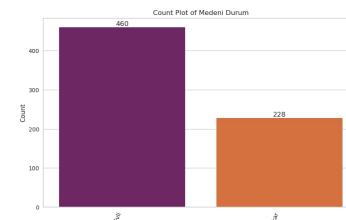
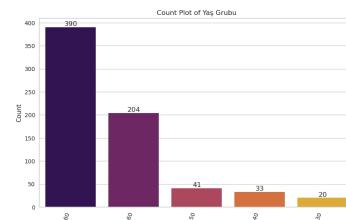
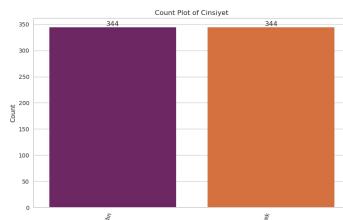


Sepete Atılan Ürün
Adedi
51 Adet

Cinsiyet dağılımı **eşit** olan bu öbek, büyük bir çoğunlukla **yaşlı** bireylerden ve **evlilerden** oluşuyor. **Eğitim düzeyleri yüksek, emekli** ve genellikle **köy ve kasabalarda** yaşıyorlar. **Geniş ürün** yelpazesiyle alışveriş yaparken, eğitimlerine devam etmiyorlar. **Yüksek gelirleri ve satın alma miktarları** öne çıkıyor.



Galeri



Öbek 7: Akademisyenler



Ortalama Gelir
351.950 ₺



Ortalama Satın
Alım Miktarı
26.429 ₺



Ortalama Sipariş
Adedi
20 Adet

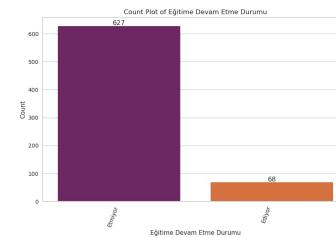
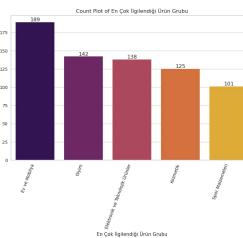
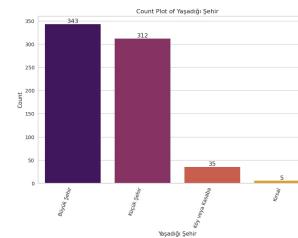
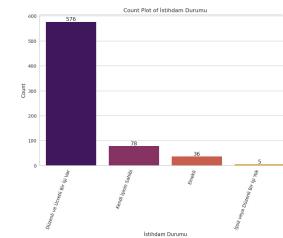
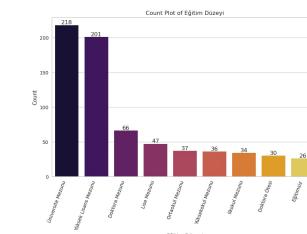
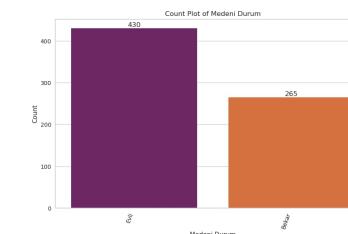
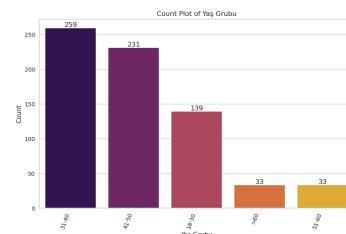
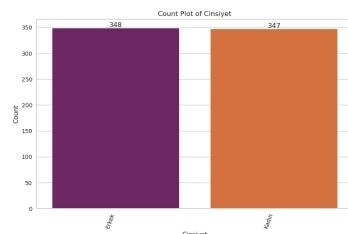


Sepete Atılan Ürün
Adedi
98 Adet

Cinsiyet dağılımı **eşit, orta-üst yaşlarda ve üst düzey eğitimli** bireylerden oluşan bu grup, çoğunluğu **evli** ve **düzenli işleri** var. **Büyük ve küçük şehirlerde** yaşayan bu müşteriler, **ev ve mobilya** çoğunlukta olmak üzere **her kategoriden** alışveriş yapmaktadır. Eğitimlerini sürdürmeme eğilimindeler.



Galeri



Öbek 8: Öğrenciler



Ortalama Gelir

255.355 ₺



Ortalama Satın Alım Miktarı

9.277 ₺



Ortalama Sipariş Adedi

49 Adet



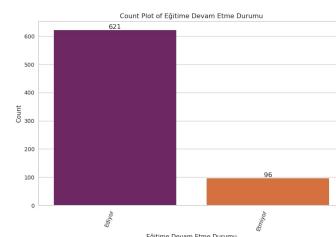
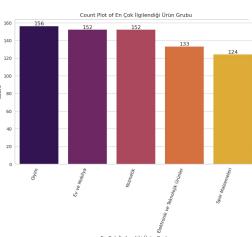
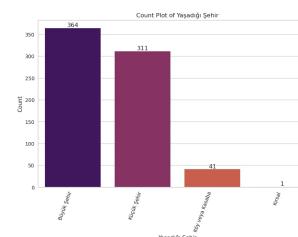
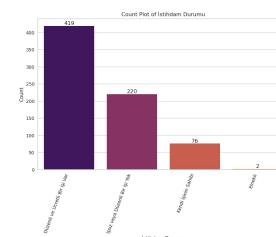
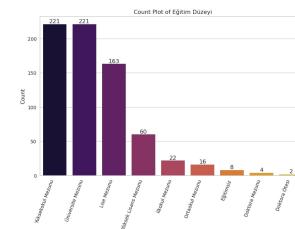
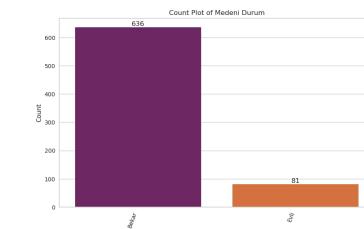
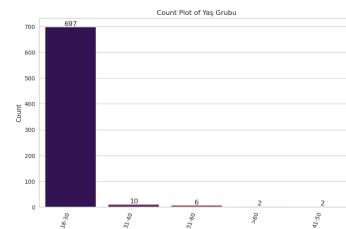
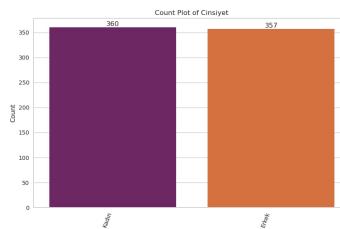
Sepete Atılan Ürün Adedi

146 Adet

Cinsiyet dağılımı **eşit** olan bu öbek, **genç** ve **yüksek eğitimli**. Genellikle **bekarlar** ve **büyük şehirlerde** yaşıyorlar. Ürün çeşitliliği onlar için önemli çünkü **her kategoriden** alım yapıyorlar ve **eğitimlerine devam ediyorlar**. Gelirleri ortalamanın **altında**, ancak alışveriş miktarları ve sepete attıkları ürün adeti dikkat çekici şekilde **yuksek**.

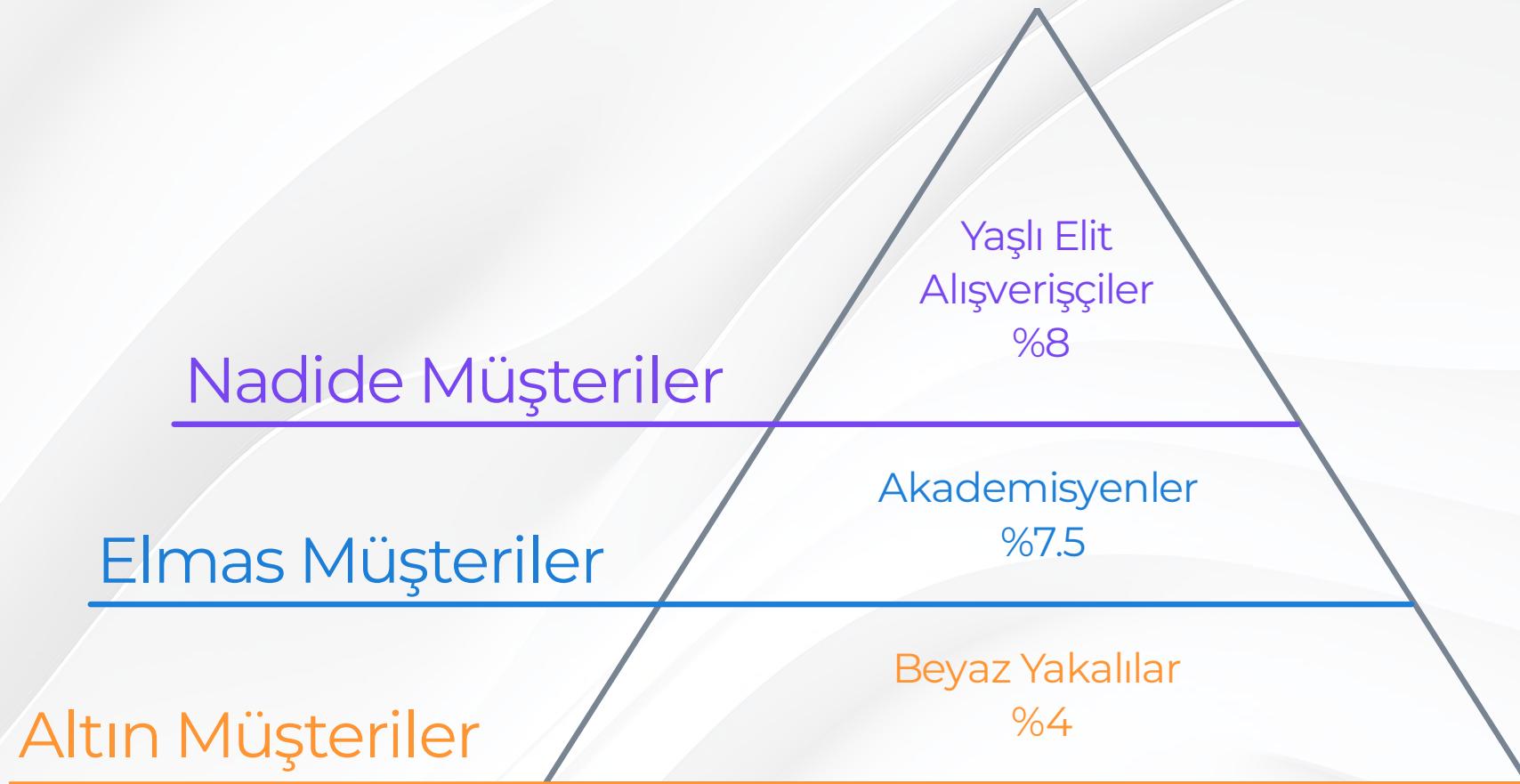


Galeri



Pazarlama Stratejileri

Bu kısımda her bir müşteri öbegine ilişkin alınacak pazarlama stratejilerine ek olarak en değerli üç müşteri sınıfına özel olarak stratejiler geliştirilmiştir.



Genel Stratejiler



Mavi Yakalılar

- Daha Fazla Sosyal Medya Varlığı:
 - Farklı yaş grupları ve cinsiyetler.
- Kişiselleştirilmiş İndirimler:
 - Bütçe dostu giyim.
- Yerel Reklamlar:
 - Küçük şehir, köy ve kasaba.
- Sadakat Programları:
 - Ortalama satın alım miktarı düşük
 - Tekrarlı alışveriş teşvik edebiliriz.

Ev Hanımları

- Daha Fazla Sosyal Medya Varlığı:
 - Farklı yaş grubundan kadınlar.
- Kişiselleştirilmiş İndirimler & Teşvikler:
 - Düşük gelirliler, bütçe dostu.
 - Sağlamlık ve güvenilirlik.
- Yerel Reklamlar.
- Sadakat Programı:
 - Düşük satın alım, tekrarlı alışveriş.

Profesyonel Erkekler

- Dijital reklam stratejisi:
 - Büyük şehir çalışanı. Dijital reklamcılık.
- Özel Kategorilerde Kampanya:
 - Spor ve teknoloji çoğunlukta.
- Sadakat Programı:
 - Ortalama satın alım miktarı düşük.
- Mobil Uygulama:
 - Teknolojiye ilgililer.



Teknoloji ve Spor Severler

- Çoklu Kanal Pazarlama:
 - Online ve offline (billboard vs.) reklamcılık.
- Özel Ürün Kategorisi
- Ürün İncelemeleri ve Kullanıcı Yorumu:
 - Ürün inceleme videoları ve kullanıcı yorumunu destekleme stratejisi.
- Etkinlik Sponsorluğu ve İşbirliği:
 - Spor etkinlikleri kurmak veya spor markaları ile işbirliğine gitmek bu öbeğin dikkatini çekebilir.
- Sadakat Programı:
 - Ortalama geliri sınırlı ancak sık alışveriş yapma potansiyeli var. Sepete atılan ürün adedi yüksek.

Öğrenciler

- Eğitim Odaklı İçerik Üretimi:
 - Eğitim düzeyi yüksek ve devam ediyor.
- Kampanyalı Eğitim Ürünleri
- Sosyal Sorumluluk Projelerine Katılım:
 - Genç nesil sosyal sorumluluk projelerine ilgi gösterir.
 - Marka bilinirliği, itibar ve güven artar.
- Öğrenci İndirim Kartları:
 - Ortalama gelirleri düşük.
- Sosyal Medya ve Interaktif Eğitimler:
 - Genç nesil sosyal medyayı aktif kullanıyor.
 - Etkinlik ve etkileşimlerle dikkat çekilebilir.

Değerli Madenleri İşleyelim



Elit Yaşlı Alışverişçiler



%8.2 → 39.440₺

1. Kişisel Alışveriş Deneyimi: Bu müşteri sınıfı yüksek gelire sahip ve özel hizmetlere ihtiyaç duyabilir. Mağaza içi kişisel alışverişin danışmanları veya **özel sipariş hizmetleri** sunarak onların taleplerine cevap verebiliriz.
2. Özel Ürün Koleksiyonları: Bu müşteri sınıfı için **özel ve yüksek kaliteli ürün** koleksiyonları oluşturarak onların taleplerine cevap verebiliriz.
3. Etkinlik ve Topluluk Oluşturma: Bu müşteri sınıfı yaşlılık dönemlerinde **sosyal etkinliklere ve topluluklara** ilgi gösterebilir. Bu tür etkinlikler ve topluluklar oluşturarak onların ilgisini çekebiliriz.
4. Sosyal Sorumluluk Projeleri: Yaşlı müşteriler **sosyal sorumluluk projelerine** ilgi gösterebilir. Bu tür projeleri destekleyerek marka **itibarımızı** artırabiliriz.
5. Eğitim ve Kültürel Destek: Bu müşteri sınıfının eğitim düzeyi yüksek. Belirli satış miktarını geçenlere **eğitim seminerleri, kültürel etkinlikler** ve sanatsal destek programları ile onların ilgisini çekebiliriz.

Akademisyenler



%7.5 → 26.429₺

1. Eğitim ve Araştırma Kaynakları: Akademik kitaplar, dergiler, özel araştırma yazılımları ve **öğretim materyalleri** gibi ürünlerle bu müşteri sınıfının ilgisini çekebiliriz.
2. Özel Seminerler ve Etkinlikler: Ünlü konuşmacılar veya akademisyenlerle düzenlenen **seminerler ve etkinliklerle** müşteri sınıfını markamiza çekebiliriz.
3. Danışmanlık Hizmetleri: Bu müşteri sınıfının ortalama satın alma miktarı yüksek. **Özel indirimler ve üyelik programları** sunarak onları teşvik edebiliriz.
4. Sorumluluk Projeleri ve Akademik Destek: Akademik bir kitlenin sosyal sorumluluk projelerine veya **akademik destek faaliyetlerine** ilgi gösterebileceğini göz önünde bulundurarak, bu tür projeleri destekleyerek müşteri sınıfına ulaşabiliriz.

Beyaz Yakalılar



%4 → 32.405₺

1. Premium Hizmet ve Ürün Sunumu: ihtiyaçlarına daha özel çözümler bulmak için **premium üyelik** vs. sağlanabilir.
2. Özel ve Sınırlı Ürün Kampanyaları: **Sınırlı sayıda ve özel tasarılanmış** ürünlerde kampanyalar düzenleyerek ilgileri çekilebilir.
3. Etkin Network Oluşturma: Büyük şehirlerde yaşıyorlar ve **iş dünyasında** aktif rol alıyorlar. İş dünyasında daha fazla görünürlük ile daha fazla müşteri, tanınırlık ve itibar.
4. Kişiselleştirilmiş Alışveriş Deneyimi: Yüksek gelirliler ve bekłentileri yüksek. Kişiselleştirilmiş alışveriş deneyimi ile **özel hizmet** sağlanabilir.
5. Sorumluluk ve Lüks Marka İşbirliği.

Tahminleme Çalışması

Bu kısımda çeşitli makine-öğrenmesi modelleri karşılaştırılmış olup en iyi iki model **Random Forest** ve **Gradient Boosting** olarak seçilmiştir. Detaylı model analizi ve temel model seçimi için aşağıdaki bağlantıyı ziyaret edebilirsiniz...

Detaylı Model Seçimi (Kaggle)



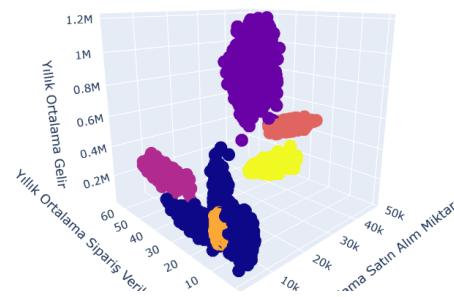
Random Forest

Bu model, veri setimizdeki **çoklu özelliklerin etkileşimlerini yakalama** yeteneği nedeniyle seçilmiştir. Aynı zamanda, veri setimizdeki **potansiyel gürültüyü düzeltme** yeteneği ve **hata toleransı da avantajlıdır.**



Gradient Boosting

Gradient Boosting, veri setimizdeki **doğrusal olmayan ilişkileri ve yüksek etkili özelliklerini öğrenme** yeteneği nedeniyle seçilmiştir. Gradient Boosting, güçlü bir tahmin yeteneğine sahip olup **overfitting'e** karşı dayanıklıdır.

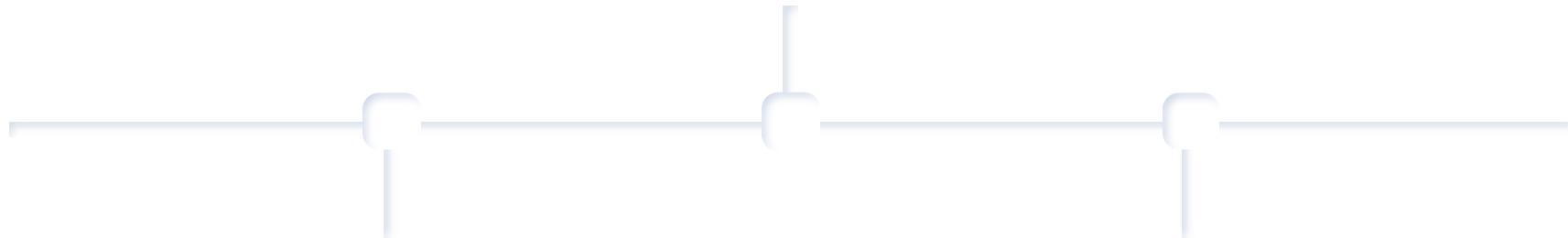


- Ayrıca **turuncu** ve **lacivert** kümelerin olduğu kisim yüksek küme sayılarına çıkıldığında karmaşıklıyor. İleride veri setinin o kismi harici olarak tutulup **Gradient Boosting** ile tahmin edilmeye çalışılabilir. Çünkü, gradient boosting veri setimizdeki **karmaşıklıkları yakalayabilir.**

Tahminleme Sonuçları & En İyi Model

Hiperparametre Ayarı

Modellerin overfit (ezberleme) etmemesi için hiperparametreleri en optimal hale getirildi.



Default Sonuçlar

Ön işleme ve encoding dışında her şey default olarak denendi ve sonuçlar tatmin edici düzeyde yüksek çıktı.

Özellik Çıkarımı

Yeni özellik olarak kümeleme analizi sonucu elde edilen küme bilgileri iki farklı başlıkta eklenerek denendi

ML Sonuçları

Sonuçlar alınırken herhangi bir şekilde modellerin parametreleri oynanmadı ve herhangi bir şekilde özellik çıkarımı eklenmedi. Aynı modeller üzerinde çalıştığımızı bilmek adına `random_state=101` belirtildi.

- ❑ Özellik çıkarımı olan kısımlarda kümeleme sonucu tek bir feature olarak değil dallandırma olarak eklenerek denendi. Tek bir feature olarak kümeleme sonucunu eklemek train ve test accuracy'lerini değiştirmedi.

Bkz: Dallandırma Yöntemi

Bu işlem, veri setinin daha fazla içgörü kazanmasına yardımcı olabilir ve model performansını artırabilir. Veri setinin daha ayrıntılı bir şekilde incelenmesine ve farklı kümeleme sonuçlarının modelin performansına etkisinin değerlendirilmesine yardımcı olabilir.

Değişken 1	Değişken 2	Kümeleme Dal 1	Kümeleme Dal 2
veri	veri	küme 1	Küme 1.1
veri	veri	küme 2	küme 2.1
veri	veri	küme 1	küme 1.2
veri	veri	küme 2	küme 2.2
veri	veri	küme 2	küme 2.1
veri	veri	küme 2	küme 2.2
veri	veri	küme 1	küme 1.2
veri	veri	küme 1	küme 1.1

Model	Veri Seti	Test Accuracy	Train Accuray	Özellik Çıkarımı
Random Forest	Tek	%95.53	%99.98	<input checked="" type="checkbox"/>
Gradient Boosting	Tek	%95.09	%98.83	<input checked="" type="checkbox"/>
Random Forest	Tek	%95.53	%100	<input checked="" type="checkbox"/>
Gradient Boosting	Tek	%95.09	%100	<input checked="" type="checkbox"/>
Random Forest	Temiz	%95.63	%100	<input checked="" type="checkbox"/>
Gradient Boosting	Karışık	%93.74	%100	<input checked="" type="checkbox"/>

Hiperparametre Ayarı

GridSearchCV ile overfit etmeyecek şekilde ancak doğruluk oranını da en yüksek seviyede tutacak şekilde hiperparametre ayarı yapıldı.

▼ Random Forest En İyi Parametreleri

```
param_grid = {  
    'max_depth': None,  
    'max_features': 'auto',  
    'min_samples_leaf': 4,  
    'min_samples_split': 2,  
    'n_estimators': 200  
}
```

▼ Gradient Boosting En İyi Parametreleri

```
param_grid = {  
    'learning_rate': 0.01,  
    'max_depth': 3,  
    'min_samples_leaf': 3,  
    'min_samples_split': 2,  
    'n_estimators': 200  
}
```

Model	Test Accuracy	Train Accuracy	Eğitim Süresi (sn)
Random Forest	%95	%96	1.13
Gradient Boosting	%96	%95	11.11

Grid Search CV Detaylı Sonuçlar

- ❑ Bir diğer teknik olan tüm numeric değerleri **balyalayıp** tüm feature değerlerini kategorik hale getirip **Bayes Teoremi** ve **CatBoost** ile tahmin etmeyi deneyecektim.

Ancak bu kısım için vaktim yetmedi.

Teşekkür Ederim

Ramazan Erduran

