Veri Bilimi Projesi

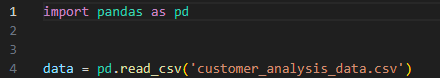
Konu :Bir özel firmanın veri setini kullanarak farklı şubelerindeki müşterileri kümelendirme analizi yapacağım.

Veri setimiz, bir perakende şirketine ait müşteri bilgilerini içermektedir. Her müşteri için yaş, gelir, harcama miktarı, satın alma sıklığı, müşteri ID, şube ID, cinsiyet ve şehir bilgileri bulunmaktadır.

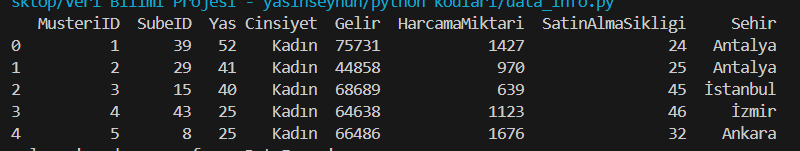
Veri Seti Özellikleri

* Yas: Müşterinin yaşı
* Gelir: Müşterinin yıllık geliri
* HarcamaMiktari: Müşterinin aylık harcama miktarı
* SatinAlmaSikligi: Müşterinin aylık satın alma sıklığı
* MusteriID: Müşterinin ID numarası
* SubeID: Şubenin ID numarası
* Cinsiyet: Müşterinin cinsiyeti (Erkek/Kadın)
* Sehir: Müşterinin ikamet ettiği şehir

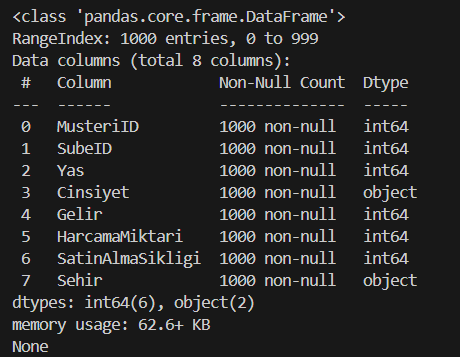
Veri setimizi Python üzerinden sahte bir şekilde oluşturuyoruz. Bunun için csv dosyası üretiyoruz.



Veri Setinin Örnek Satırları



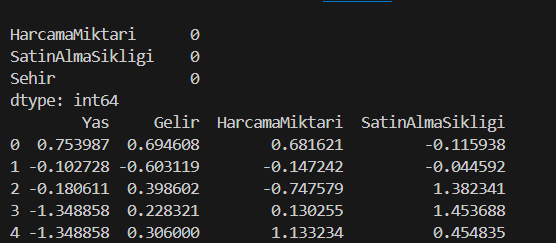
Veri Setinin Genel Bilgisi



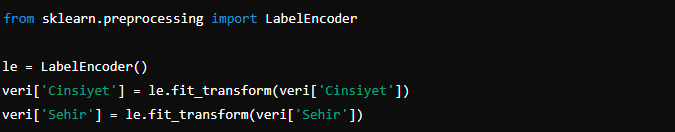
Veriler Arası Korelasyon



Veri setimizi oluşturduktan sonra verilerin doğruluğunu kontrol ediyoruz.

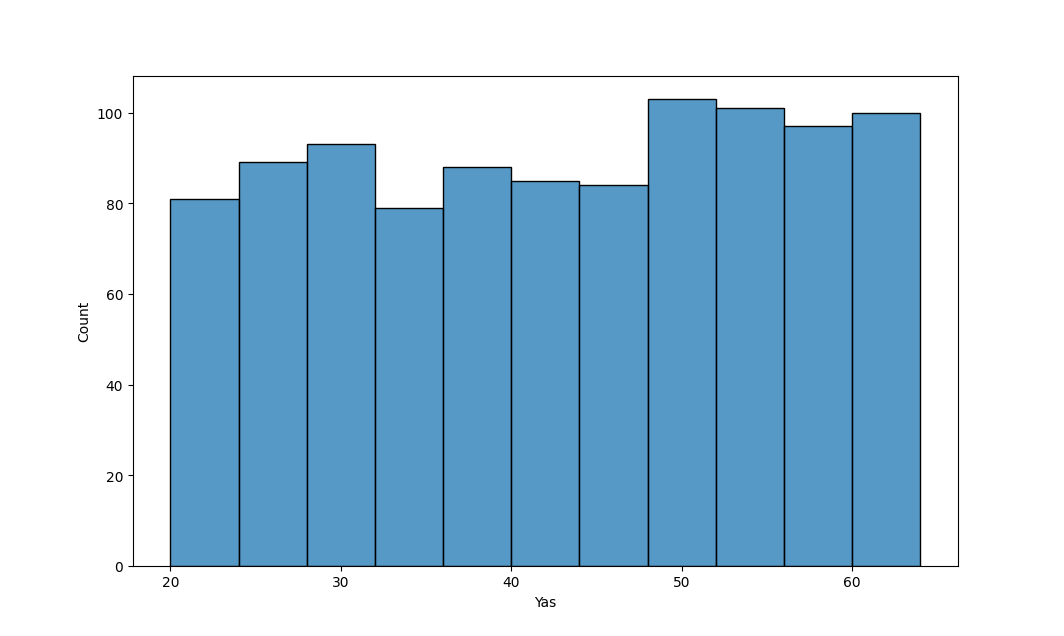


Eksik veya fazla veri yoksa kategorik verilerin kodlanması yapılır.



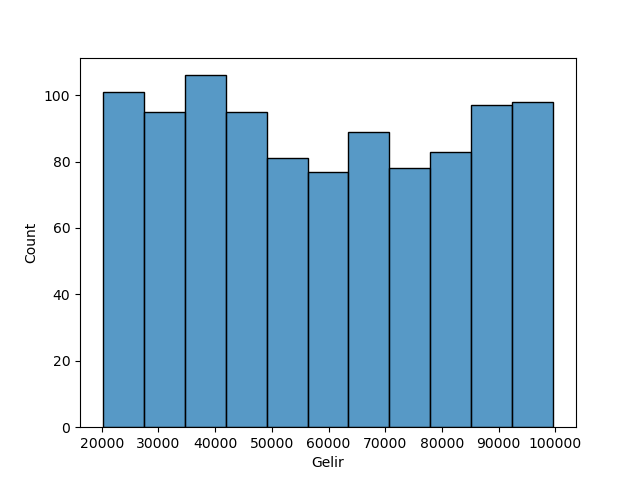
Dağılımlar

Yas Verisi

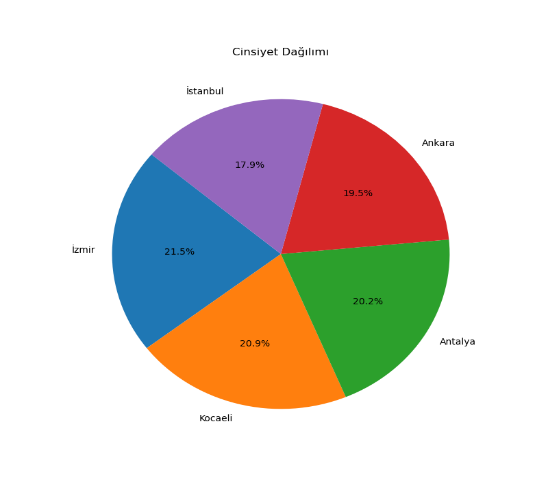


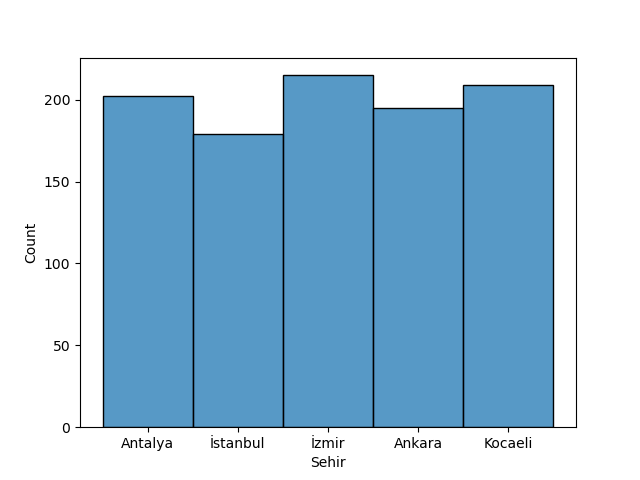
Buradaki yaş verisi sonucunda bu firmada 50-60 yaş aralığının diğer yaş aralıklarına oranla daha çok alışveriş yapıldığı görülüyor.

Gelir Seviyesi

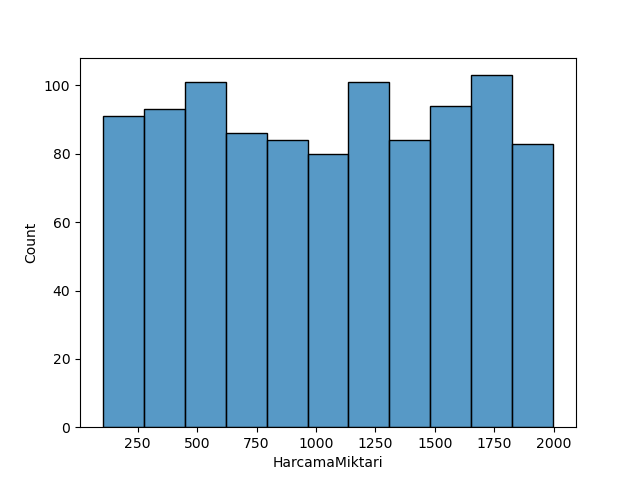


Buradaki sonuç gelir seviyesinin ortalama olduğu görülüyor.

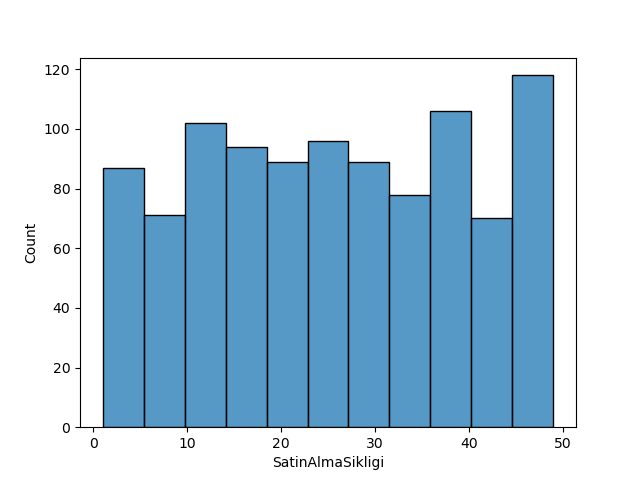
Sehir



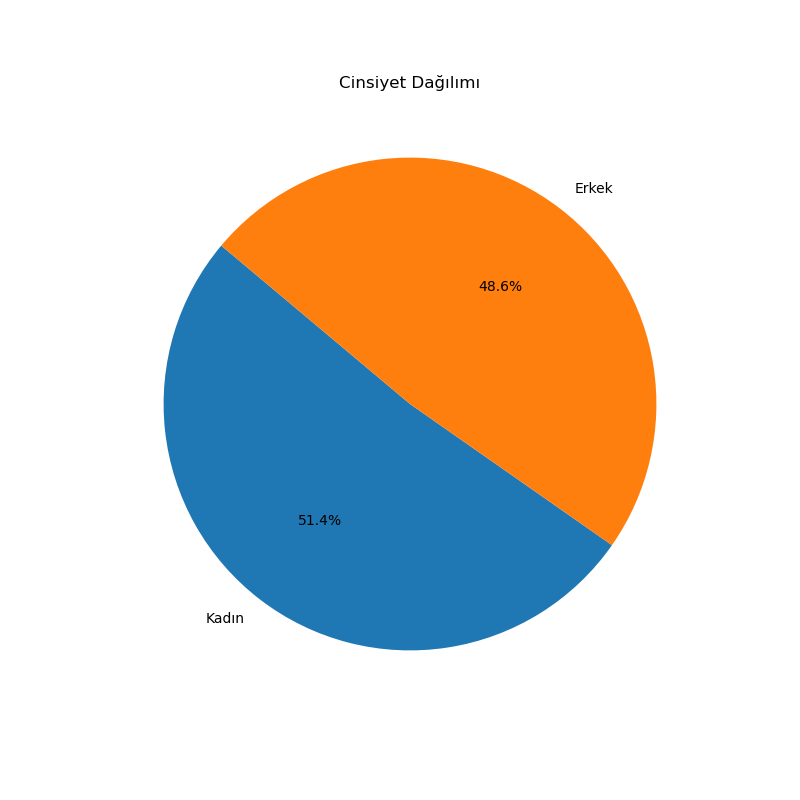
Buradaki sonuç bu firmadan izmir subesinin daha çok olduğu yönündedir.

Harcama Miktarı 

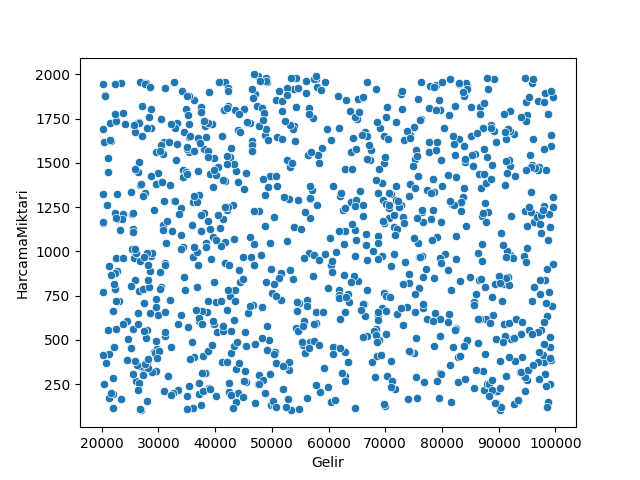
Satın Alma Sıklığı



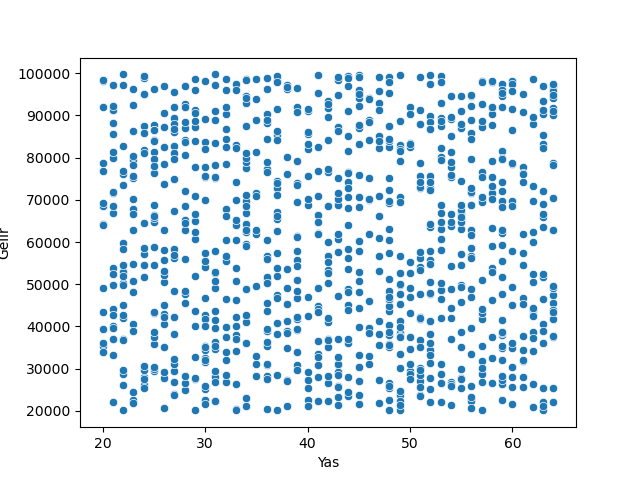
Cinsiyet



Gelir ve Harcama Miktarı

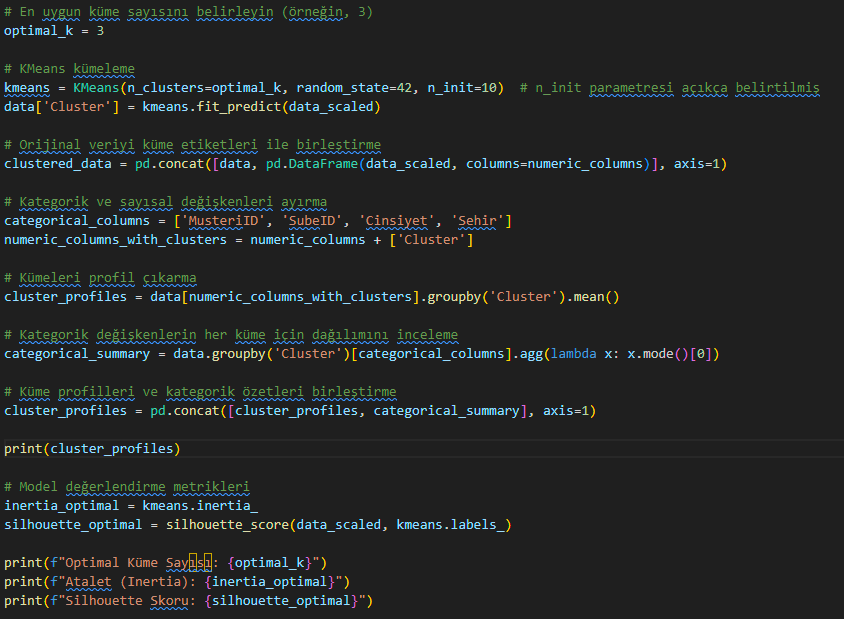


Yas ve Gelir Seviyesi



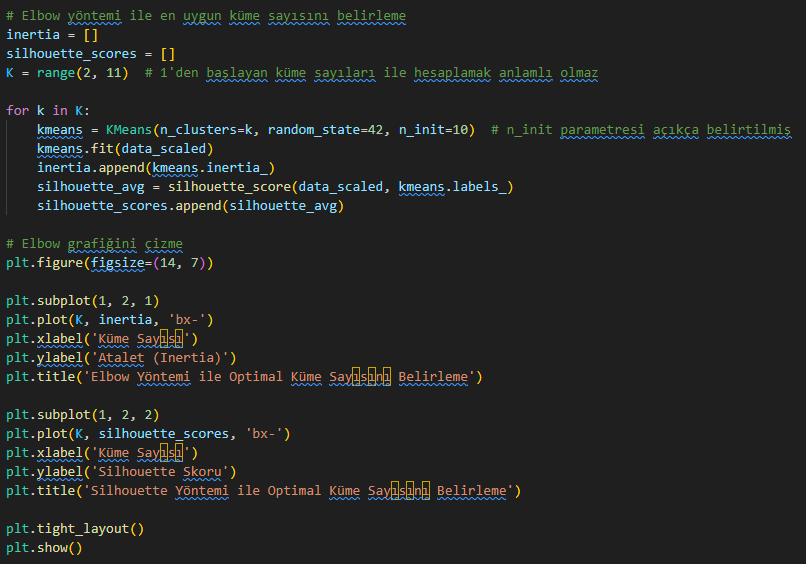
KMeans Kümeleme Modeli

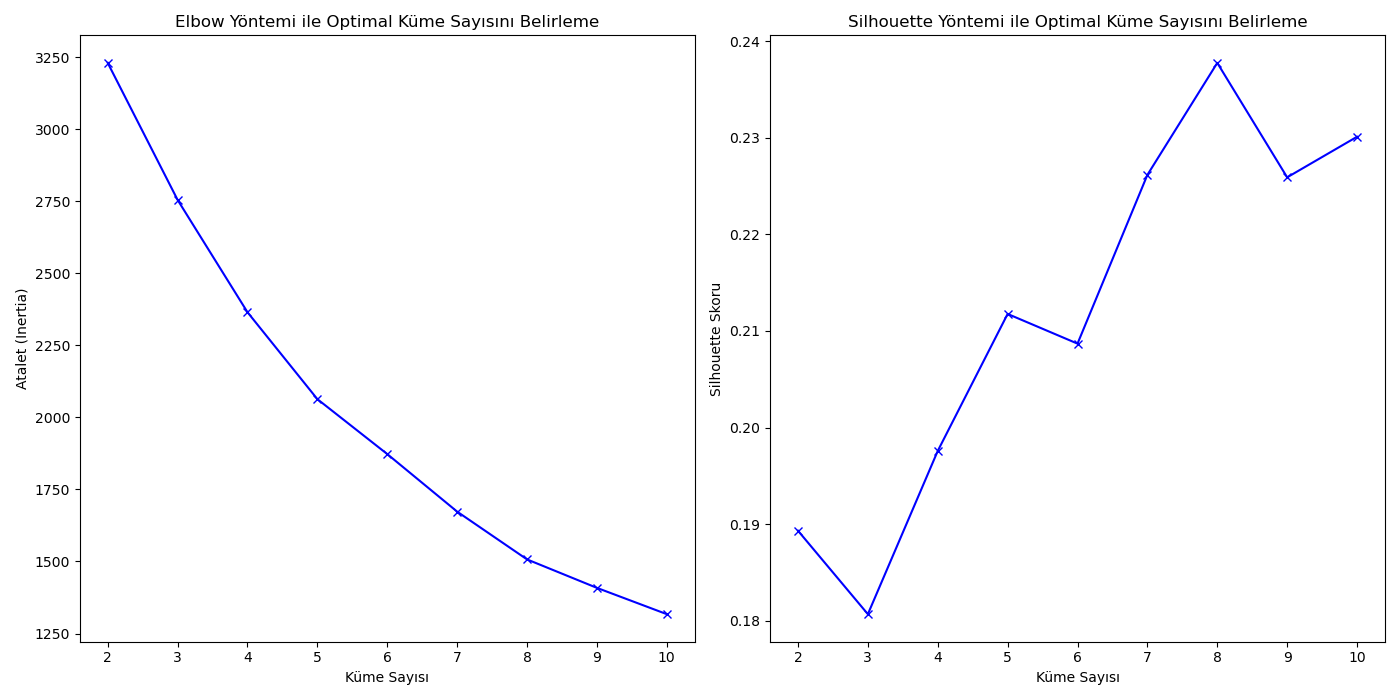
Kmeans algoritması ile kümeleme yaparken ilk adım küme sayısını belirlemektir. Ben bu projemde bunu sayıyı 3 olarak aldım. Bu sayıyı belirlemek için dirsek yöntemi yani Elbow Metodu kullandım. Bu metod farklı küme sayıları için atalet değerlerini hesaplayarak optimal küme sayısını belirler. Grafik üzerinde atalet değerinin azalmayı bıraktığı noktada optimal küme sayısı bulunur. Optimal küme sayısını bulduktan sonra model bu küme sayısı ile eğitilir ve her veri noktasına küme etiketi atanır. Bu işlemlerden sonra sonuçlar değerlendirilir.



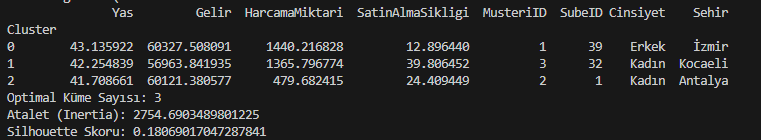
Atalet (Inertia): Kümelerin içindeki verilerin toplam karesel mesafesini gösterir. Daha düşük atalet değerleri, kümelerin daha sıkı ve iyi ayrıldığını gösterir.

Silhouette Skoru: Her bir veri noktasının, kendi kümesine olan yakınlığı ile en yakın komşu kümeye olan uzaklığının farkını ölçer. Bu skor, -1 ile 1 arasında bir değerdir ve 1'e yakın değerler daha iyi kümeleme olduğunu gösterir.





Modelin Sonuçlandırılması



**Atalet (Inertia): 2754.69**

**Silhouette Skoru: 0.1807**

Atalet, kümelerin içindeki verilerin toplam karesel mesafesini gösterir. Düşük değerler, kümelerin daha sıkı olduğunu gösterir. Silhouette skoru, kümelerin ne kadar iyi ayrıldığını ve içindeki elemanların ne kadar benzer olduğunu gösterir. 0.1807'lik bir değer, kümelerin birbirinden çok iyi ayrılmadığını göstermektedir.

**Kümelerin Özellikleri**

* Küme 0: Genellikle İzmir'de yaşayan, erkek müşterilerden oluşur. Yaş ortalaması 43, yıllık gelir ortalaması 60,327 TL ve aylık harcama miktarı 1,440 TL'dir.
* Küme 1: Kocaeli'de yaşayan, kadın müşterilerden oluşur. Yaş ortalaması 42, yıllık gelir ortalaması 56,964 TL ve aylık harcama miktarı 1,366 TL'dir.
* Küme 2: Antalya'da yaşayan, kadın müşterilerden oluşur. Yaş ortalaması 41, yıllık gelir ortalaması 60,121 TL ve aylık harcama miktarı 480 TL'dir.

Bu çalışmada, müşteri verileri kullanılarak KMeans kümeleme algoritmasıyla müşteri segmentasyonu gerçekleştirilmiştir. Optimal küme sayısı üç olarak belirlenmiş ve her küme için ortalama değerler hesaplanmıştır. Sonuçlar, her kümenin demografik ve davranışsal özelliklerini ortaya koymuştur.

**210404056**

**Yasin Seyhun**

**Bilgisayar Mühendisliği 3.Sınıf**