

Siyasal Bilgiler Fakültesi



MÜŞTERİ SEGMENTASYONU SINIFLANDIRMASI

İstatistik 2

Burcu Kara
21080767
30 Mart 2024

MÜŞTERİ SEGMENTASYONU SINIFLANDIRMASI

Giriş

Müşteri Segmentasyonu, denetimsiz öğrenmenin en önemli uygulamalarından biridir. Kümeleme teknikleri sayesinde şirketler, potansiyel müşterilerini hedeflemelerine olanak tanıyan çeşitli müşteri segmentlerini belirleyebilirler. Bu makine öğrenimi projesinde, temel bir yöntem olan k-ortalama Kümeleme yöntemini kullanacağız.

Kapsam

Müşteri segmentasyonu, işletmelerin en değerli müşterilerini belirlemelerine ve onlarla daha etkili bir şekilde etkileşimde bulunmalarına olanak tanır. Bu projede, R programlama dili kullanarak müşteri segmentasyonunu gerçekleştireceğiz, bu da makine öğreniminin temel uygulamalarından biridir. Bu analiz sayesinde, işletmeler hedef kitlelerini daha iyi anlayabilir ve onlara daha özgün pazarlama stratejileri geliştirebilirler.

R Studio

R, özellikle istatistiksel analiz için tasarlanmıştır, bu da onu veri bilimi uygulamaları için son derece uygun hale getirir. R ile programlama yapmanın öğrenme eğrisi dik olabilir, özellikle önceki programlama deneyimi olmayanlar için, ancak R'de metin analizi yapmak için kullanılan araçlar, yalnızca birkaç basit komut kullanarak güçlü ve son teknoloji metin analitiği gerçekleştirmeyi kolaylaştırır. stratejileri geliştirebilirler.

Paketler

- ☐ tidyverse
- ☐ plotrix
- ☐ purrr
- ☐ cluster
- ☐ gridExtra
- ☐ grid
- ☐ NbClust
- ☐ factoextra

R Studio Versiyonu

- ☐ 2023.12.1+402

Veri Seti

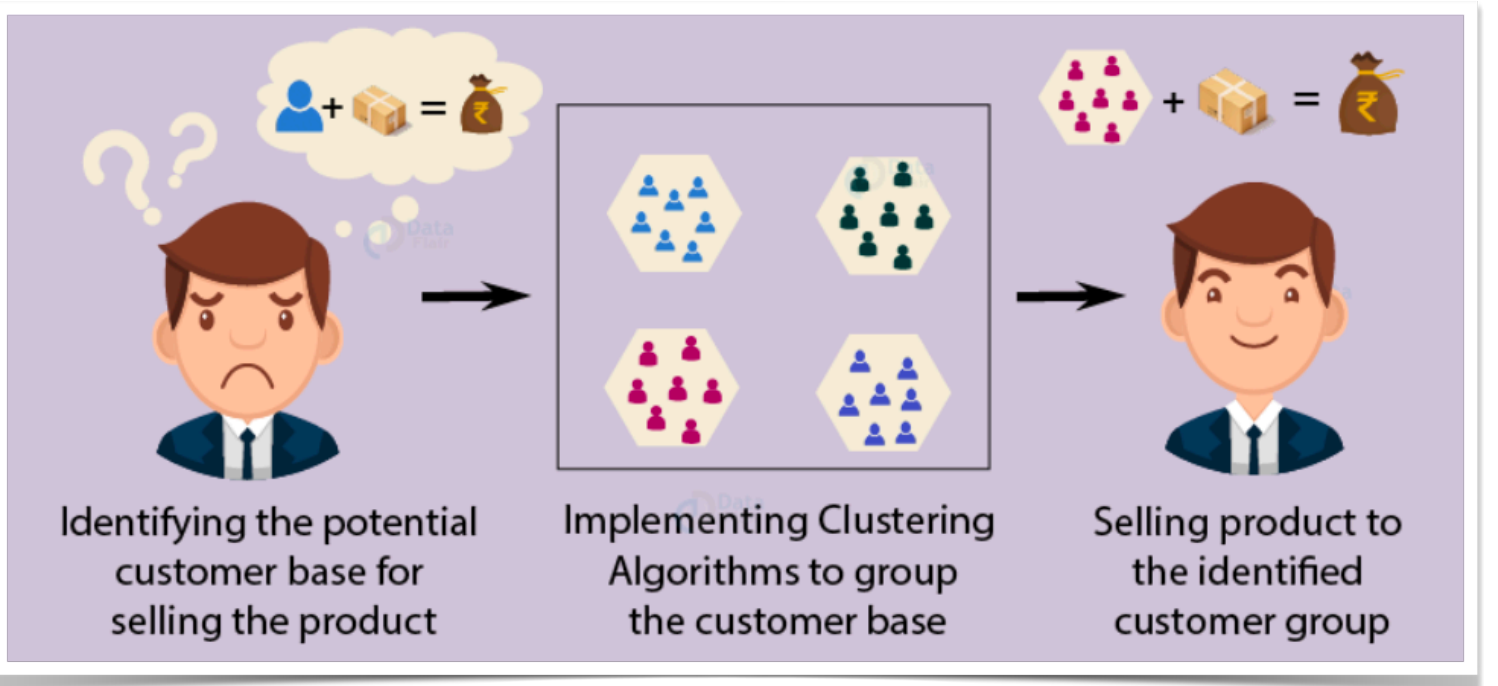
- ☐ Mall_Customers.csv

BÖLÜM: 1

Müşteri Segmentasyonu Nedir?

Müşteri segmentasyonu, müşteri tabanının çeşitli birey gruplarına bölünmesi sürecidir. Bu bölünme, müşterilerin cinsiyet, yaş, ilgi alanları gibi pazarlama ile ilgili farklı şekillerde benzerlik gösteren özelliklerine dayanır. Ayrıca, çeşitli harcama alışkanlıkları da segmentasyon sürecinde önemli bir rol oynar.

Müşteri segmentasyonu uygulayan şirketler, her müşterinin farklı özelliklere sahip olduğu düşüncesiyle hareket eder ve bu gereksinimleri uygun şekilde ele almak için özel bir pazarlama çabası gerektirir. Şirketlerin amacı, hedefledikleri müşteriye daha derin bir yaklaşım sağlamaktır. Bu nedenle, amaçları spesifik olmalı ve her bir müşterinin gereksinimlerini karşılamak üzere uyarlanmış olmalıdır. Ayrıca, veriler aracılığıyla toplanan şirketler, müşteri tercihlerinin yanı sıra daha derin bir anlayış kazanarak maksimum kâr sağlayacak değerli segmentleri keşfedebilirler. Bu şekilde, kendi stratejilerini ve pazarlama tekniklerini daha verimli bir şekilde kullanabilirler ve yatırımlarının riskini en aza indirebilirler.



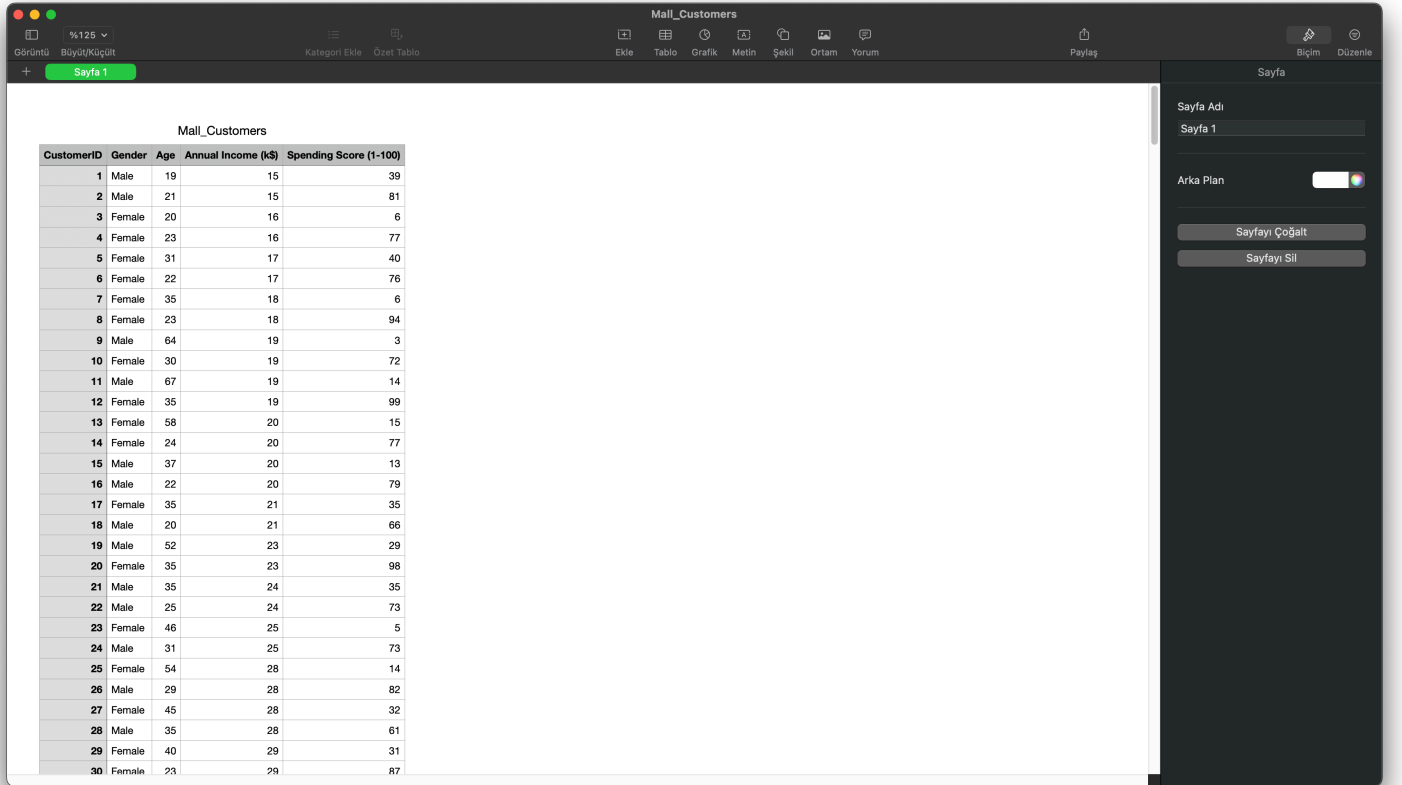
Uygulama

Bu veri bilimi projesinin ilk adımında veri keşfi gerçekleştireceğiz. Temel veri paketlerini inceleyecek ve ardından verilerimizi okuyacağız. Son olarak, girdi verilerinin üzerinden geçerek aşağıdaki bilgilere

Veri Seti - Mall_Customers.csv

Müşteri segmentasyon analizine geçmeden önce, ilk adım analiz yapılacak verinin R tarafından okunmasıdır. Veriler Mall_Customers.csv olarak adlandırılan veri kümesinde kaydedilir. Bu veri kümesi 400 kayıt içerir. Veri "read.csv" komutu ile veri R içerisine aktarılır.

```
customer_data=read.csv("/Users/burcukara/Desktop/R-studio\proje/Mall_Customers.csv")
```

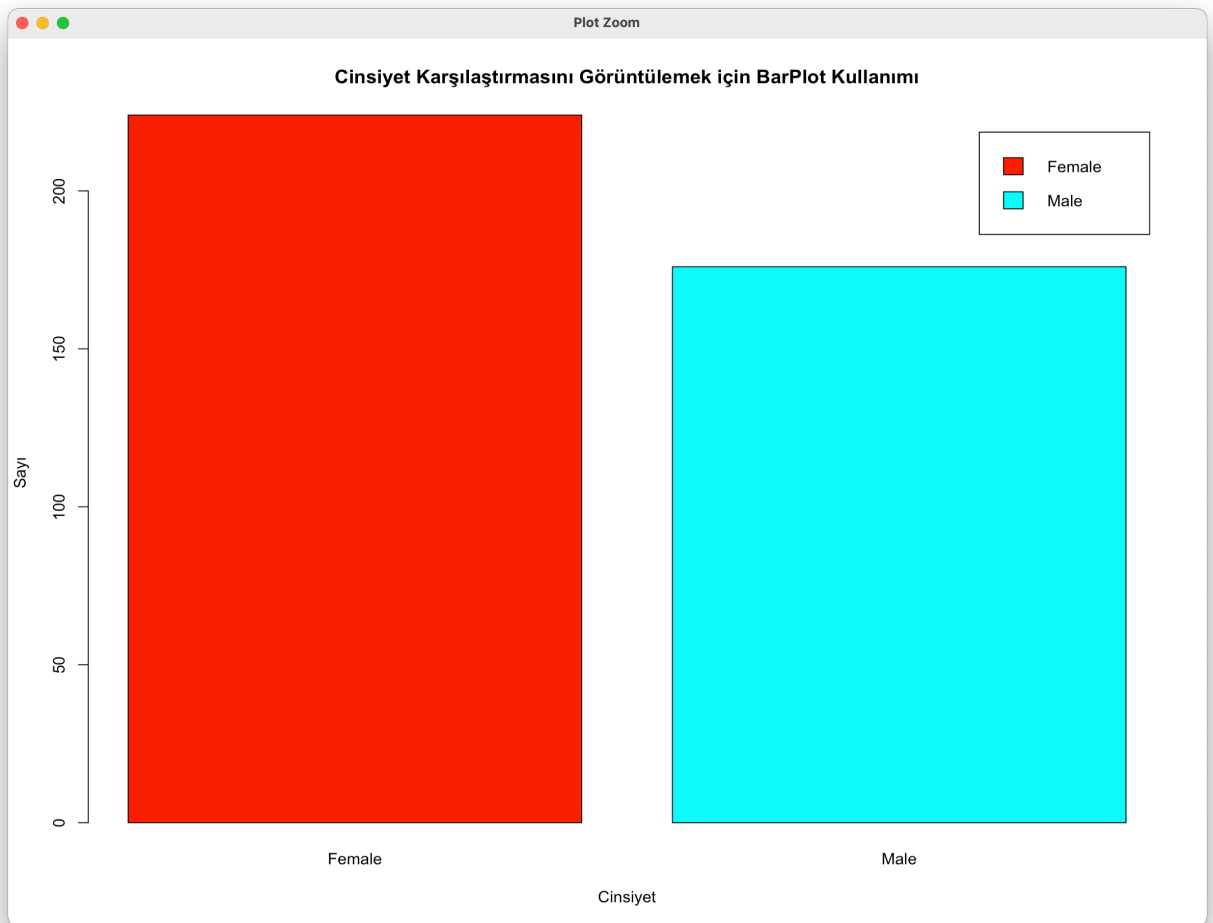


CustomerID	Gender	Age	Annual Income (k\$)	Spending Score (1-100)
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40
6	Female	22	17	76
7	Female	35	18	6
8	Female	23	18	94
9	Male	64	19	3
10	Female	30	19	72
11	Male	67	19	14
12	Female	35	19	99
13	Female	58	20	15
14	Female	24	20	77
15	Male	37	20	13
16	Male	22	20	79
17	Female	35	21	35
18	Male	20	21	66
19	Male	52	23	29
20	Female	35	23	98
21	Male	35	24	35
22	Male	25	24	73
23	Female	46	25	5
24	Male	31	25	73
25	Female	54	28	14
26	Male	29	28	82
27	Female	45	28	32
28	Male	35	28	61
29	Female	40	29	31
30	Female	23	29	87

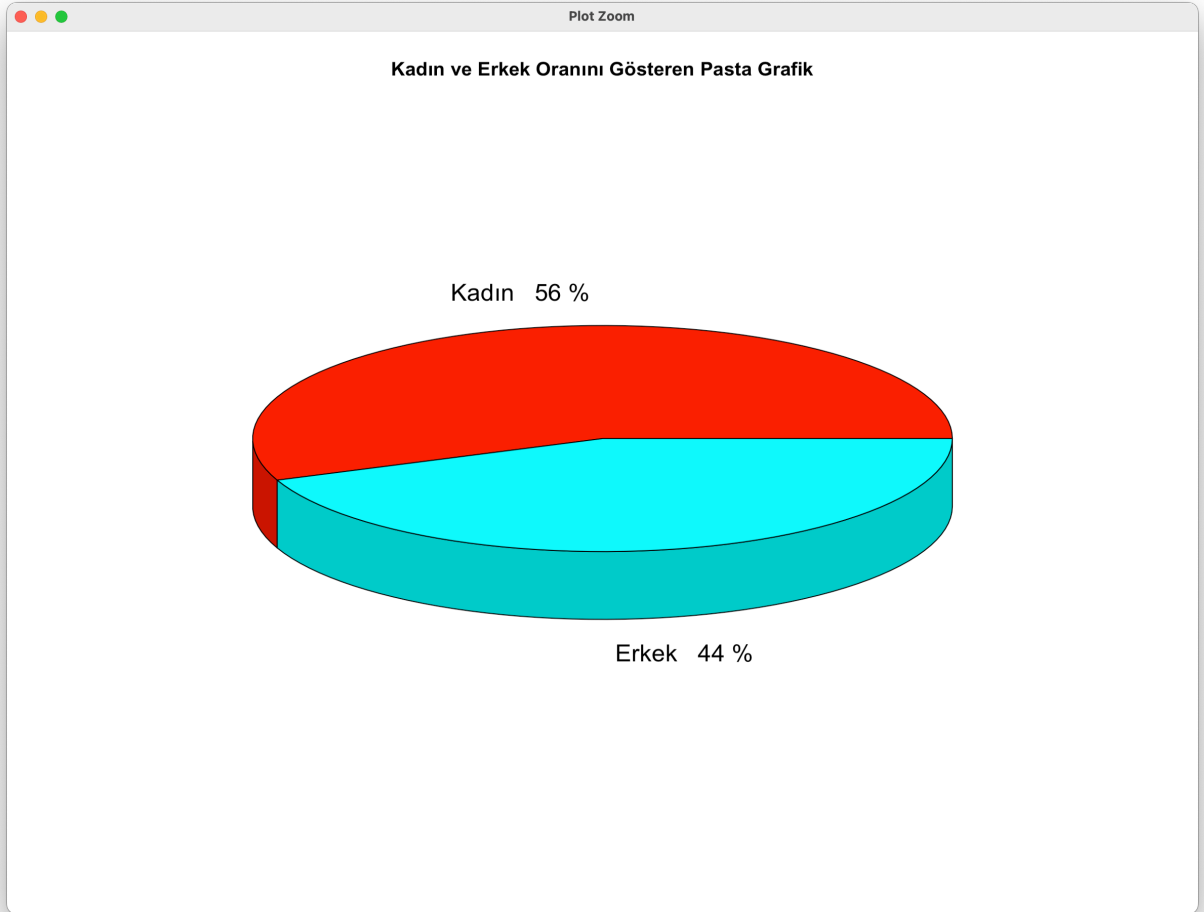
BÖLÜM: 2

Müşteri Cinsiyeti Görselleştirme

Bu bölümde, customer_data veri kümemizdeki cinsiyet dağılımını göstermek için bir çubuk grafik ve bir piechart oluşturacağız. Bir çubuk grafik, verileri dikdörtgen çubuklar halinde temsil eder ve çubuğun uzunluğu değişkenin değeriyle orantılıdır. R, çubuk grafikler oluşturmak için barplot() fonksiyonunu kullanır. R, çubuk grafikte hem dikey hem de yatay çubuklar çizebilir. Çubuk grafikte çubukların her birine farklı renkler verilebilir.



Aşağıdaki grafikten, kadınların yüzdesinin **%56**, erkeklerin yüzdesinin ise müşteri veri kümesindeki oranı **%44**'tür.



Yaş Dağılımının Görselleştirilmesi

Müşteri yaşlarının frekansını çizmek üzere dağılımını görüntülemek için bir histogram çizelim. İlk olarak, yaş değişkeninin özetini alarak devam edeceğiz.

Yaş Dağılımının Görselleştirilmesi

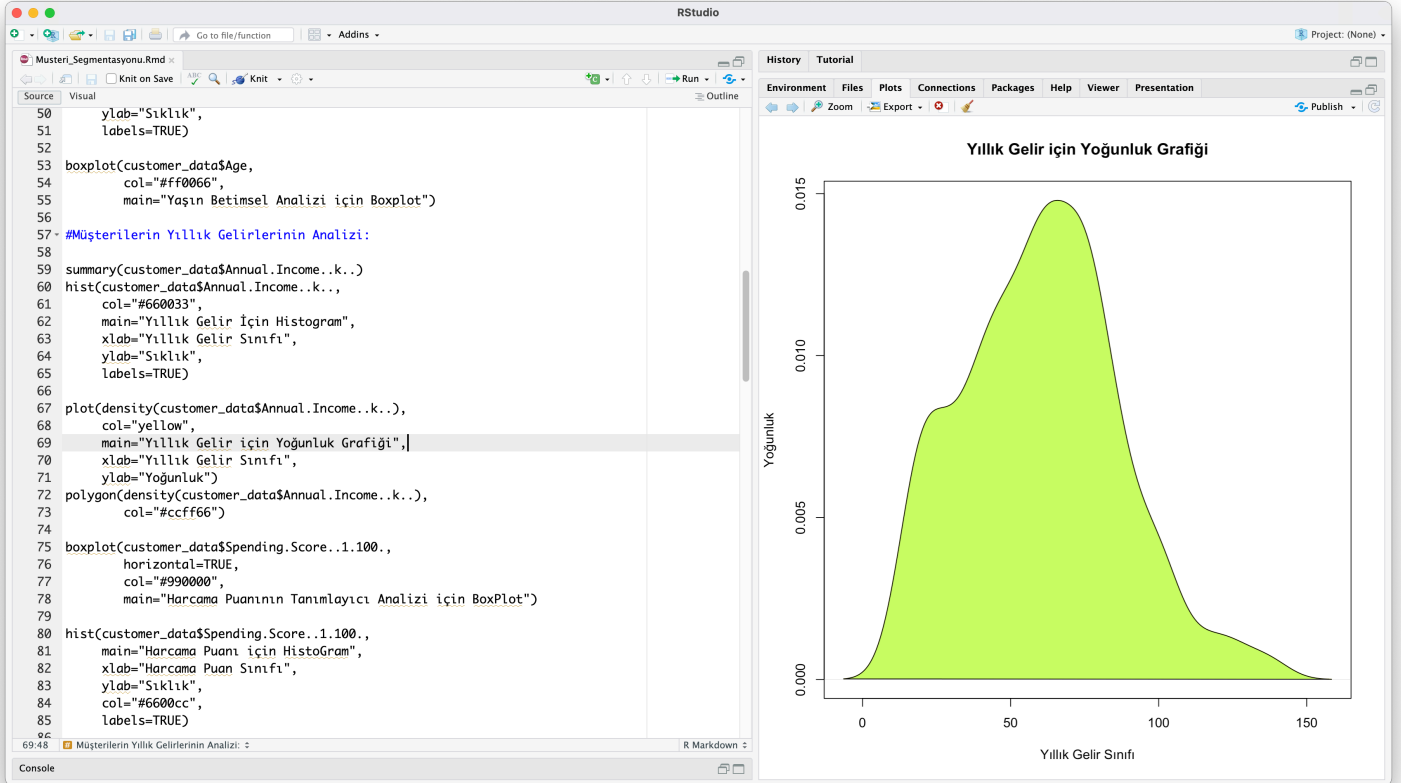
```
summary(customer_data$Age)
hist(customer_data$Age,
col="blue",
main="Yaş Sınıfı Sayısını Gösteren Histogram",
xlab="Yaş Sınıfı",
ylab="Sıklık",
labels=TRUE)
```

```
> summary(customer_data$Age)
  Min. 1st Qu.  Median    Mean 3rd Qu.
18.00  28.75   36.00   38.85  49.00
  Max.
70.00
```

Yukarıdaki görselden maksimum müşteri yaşlarının yukarıdaki aralıklarda olduğu sonucuna varıyoruz.

Müşterilerin Yıllık Gelirlerinin Analizi

R projesinin bu bölümünde, müşterilerin yıllık gelirlerini analiz etmek için görselleştirmeler oluşturacağız. Bir histogram çizeceğiz ve ardından bu verileri bir yoğunluk grafiği kullanarak incelemeye devam edeceğiz.



Yukarıdaki betimleyici analize göre, asgari yıllık gelire sahip müşteri sayısı 15 ve maksimum gelir 137'dir. Ortalama gelir 70 olarak hesaplanmıştır ve histogram dağılımında en yüksek frekans bu değere denk gelmektedir. Tüm müşterilerin ortalama geliri ise 60,56'dır. Gösterilen Kernel Yoğunluk Grafiği'nde, müşteri sayısının yıllık geliri normal bir dağılıma sahiptir.

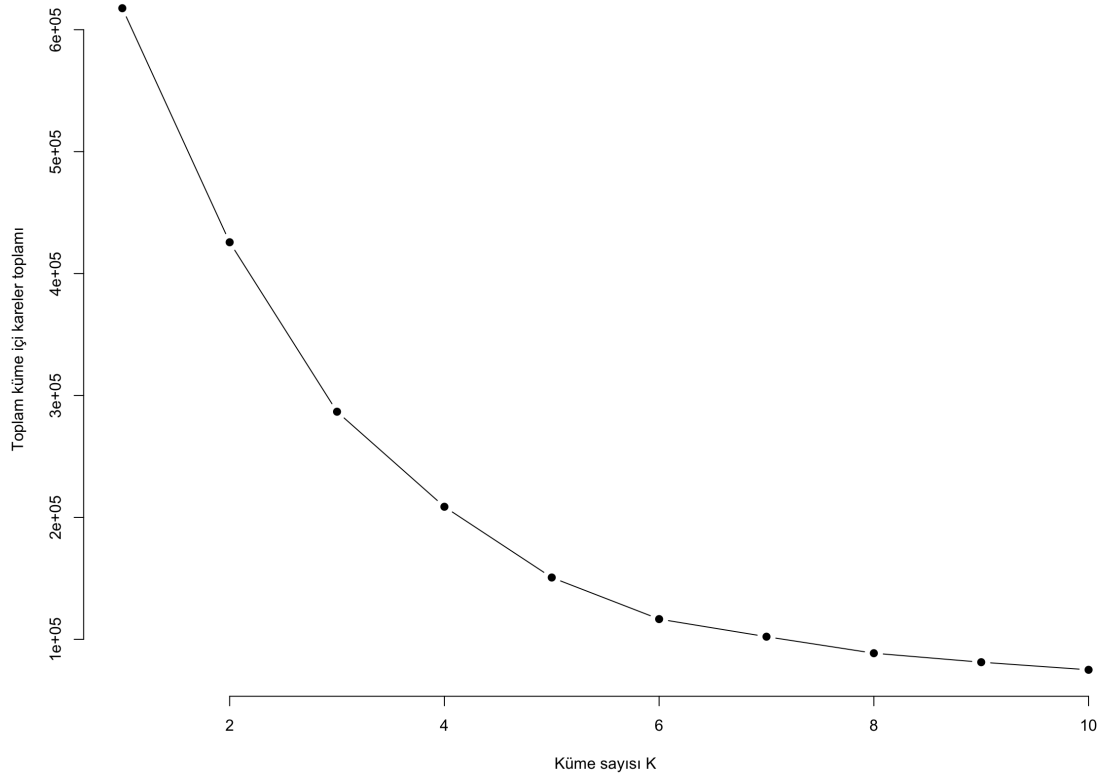
BÖLÜM: 3

```
library(purrr)
set.seed(123)
# function to calculate total intra-cluster sum of square
iss <- function(k) {
  kmeans(customer_data[,3:5],k,iter.max=100,nstart=100,algorithm="Lloyd" )$tot.withinss
}

k.values <- 1:10

iss_values <- map_dbl(k.values, iss)

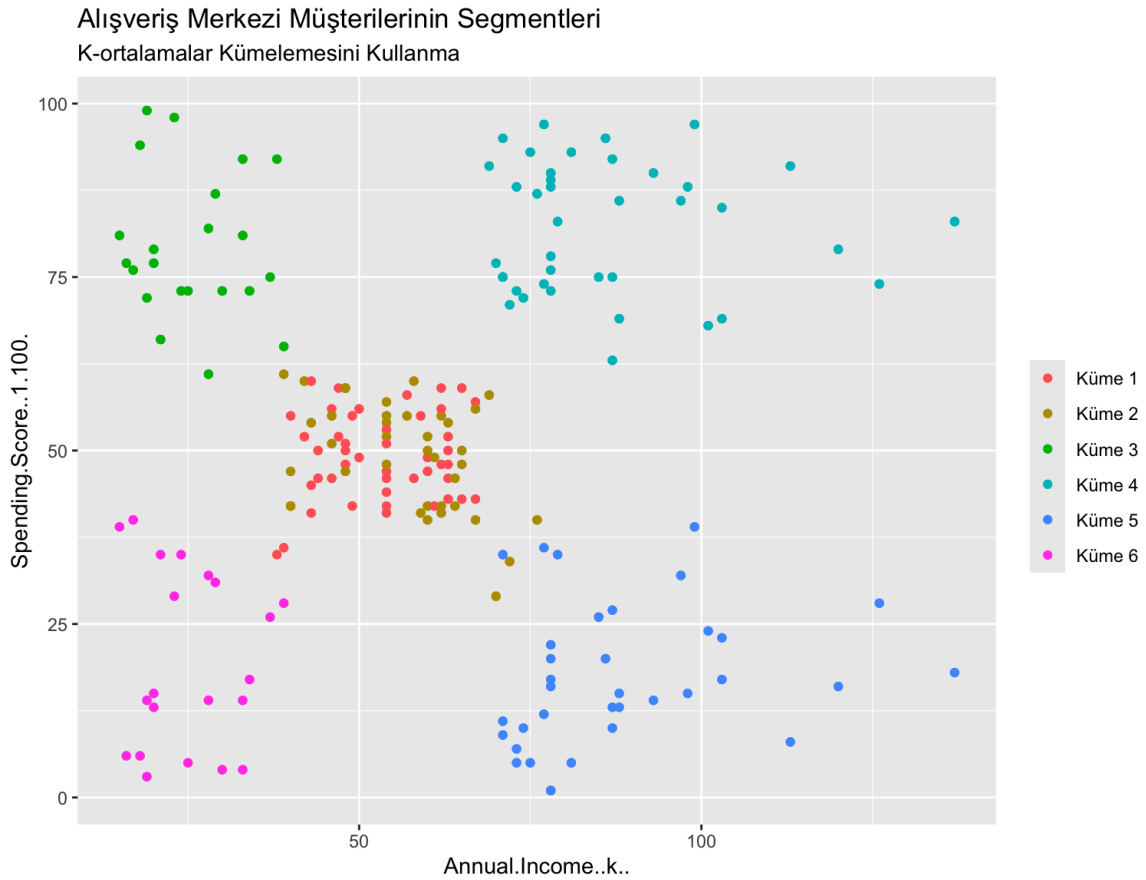
plot(k.values, iss_values,
     type="b", pch = 19, frame = FALSE,
     xlab="Küme sayısı K",
     ylab="Toplam küme içi kareler toplamı")
```



BÖLÜM: 4

Çizgi grafiği veya çizgi çizimi veya çizgi grafiği veya eğri grafiği, bilgileri aşağıdaki gibi gösteren bir grafik türüdür Düz çizgi parçalarıyla birbirine bağlanan 'işaretleyiciler' adı verilen veri noktaları serisi. Temel bir grafik türüdür birçok alanda yaygındır. Birçok alanda kullanılan bu tür grafikler, aşağıdaki durumları göstermede oldukça yardımcı olabilir zaman içinde değerlerdeki değişiklikler. Çizgi grafiğini göstermek için ggplot kullanacağız.

```
set.seed(1)
ggplot(customer_data, aes(x =Annual.Income..k., y = Spending.Score..1.100.)) +
  geom_point(stat = "identity", aes(color = as.factor(k6$cluster))) +
  scale_color_discrete(name=" ",
                        breaks=c("1", "2", "3", "4", "5","6"),
                        labels=c("Küme 1", "Küme 2", "Küme 3", "Küme 4", "Küme 5","Küme
6")) +
  ggtitle("Alışveriş Merkezi Müşterilerinin Segmentleri", subtitle = "K-ortalamalar
Kümelemesini Kullanma")
```



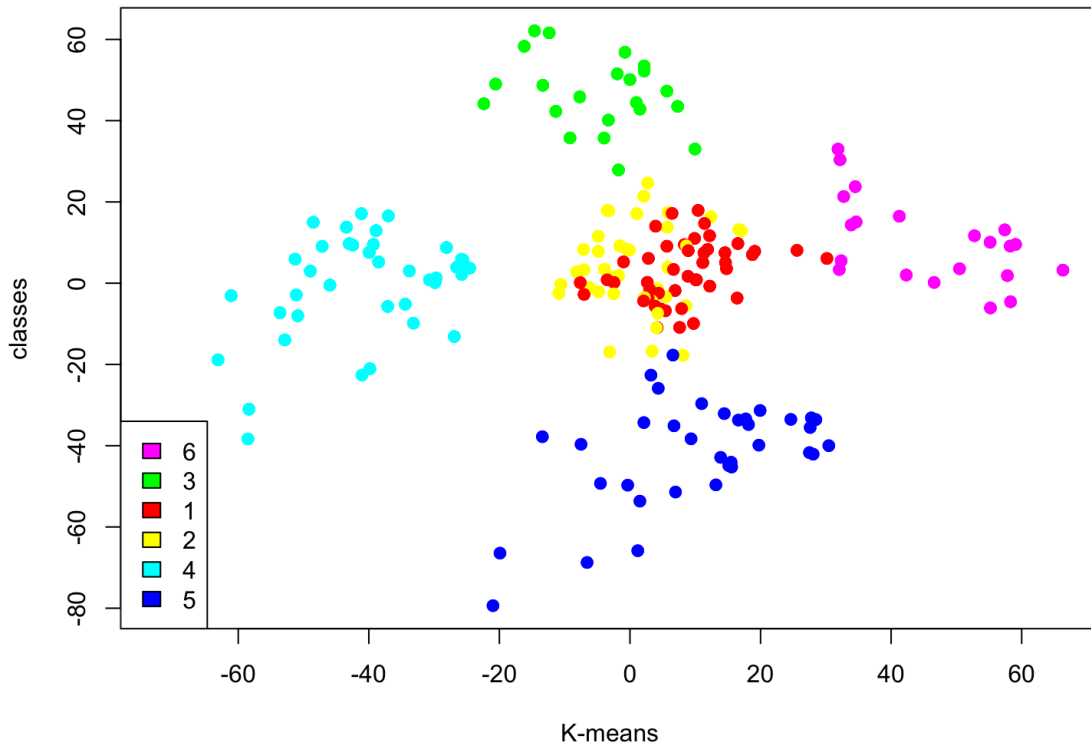
Yukarıdaki görselleştirmeden, aşağıdaki gibi 6 kümeden oluşan bir dağılım olduğunu gözlemliyoruz.

Küme 6 ve 4- Bu kümeler orta gelir maaşına sahip customer_data temsil etmektedir yanı sıra orta düzeyde yıllık maaş harcaması.

Küme 1 - Bu küme, yüksek yıllık gelire sahip müşteri_verilerini ve aynı zamanda yüksek bir yıllık harcama.

Küme 3 - Bu küme, düşük yıllık gelire sahip müşteri_verilerinin yanı sıra düşük yıllık gelir harcaması.

Küme 2 - Bu küme yüksek yıllık geliri ve düşük yıllık harcamayı ifade eder.



Küme 4 ve 1 - Bu iki küme orta PCA1 ve orta PCA1'e sahip müşterilerden oluşmaktadır. PCA2 puanı.

Küme 6 - Bu küme yüksek PCA2 ve düşük PCA1 puanına sahip müşterileri temsil etmektedir.

Küme 5 - Bu kümede orta PCA1 ve düşük PCA2 skoruna sahip müşteriler bulunmaktadır.

Küme 3 - Bu küme yüksek PCA1 gelirine ve yüksek PCA2'ye sahip müşterilerden oluşmaktadır.

Küme 2 - Bu küme, yüksek PCA2 ve orta düzeyde yıllık harcamaya sahip müşterilerden oluşmaktadır.

KAYNAKÇA

- Aydınonat, N. E. (2007). İktisat Öğrencileri için Ödev Yazma Kılavuzu. [http://iktisat.cu.edu.tr/tr/Belgeler/Formlar/Bitirme%20Projesi%20Ödev%20Hazırlama%20Rehberi/N.%20Emrah%20AYDINONAT%20\(2006\)%20Ödev%20Rehberi.pdf](http://iktisat.cu.edu.tr/tr/Belgeler/Formlar/Bitirme%20Projesi%20Ödev%20Hazırlama%20Rehberi/N.%20Emrah%20AYDINONAT%20(2006)%20Ödev%20Rehberi.pdf) adresinden erişildi.
- <https://www.namogoo.com/blog/consumer-behavior-psychology/customer-segmentation/>
- <https://www.datacamp.com/cheatSheet/getting-started-r>
- <https://posit.co/products/open-source/rpackages/>