



# Analisi del comportamento del consumatore

---

Dataset: Mall\_Customers

Source: [kaggle.com](https://www.kaggle.com/datasets/abhishek1999/mall-customers)



# Descrizione del dataset

**Customer\_ID:** codice identificativo del cliente del centro commerciale

**Gender:** genere del cliente

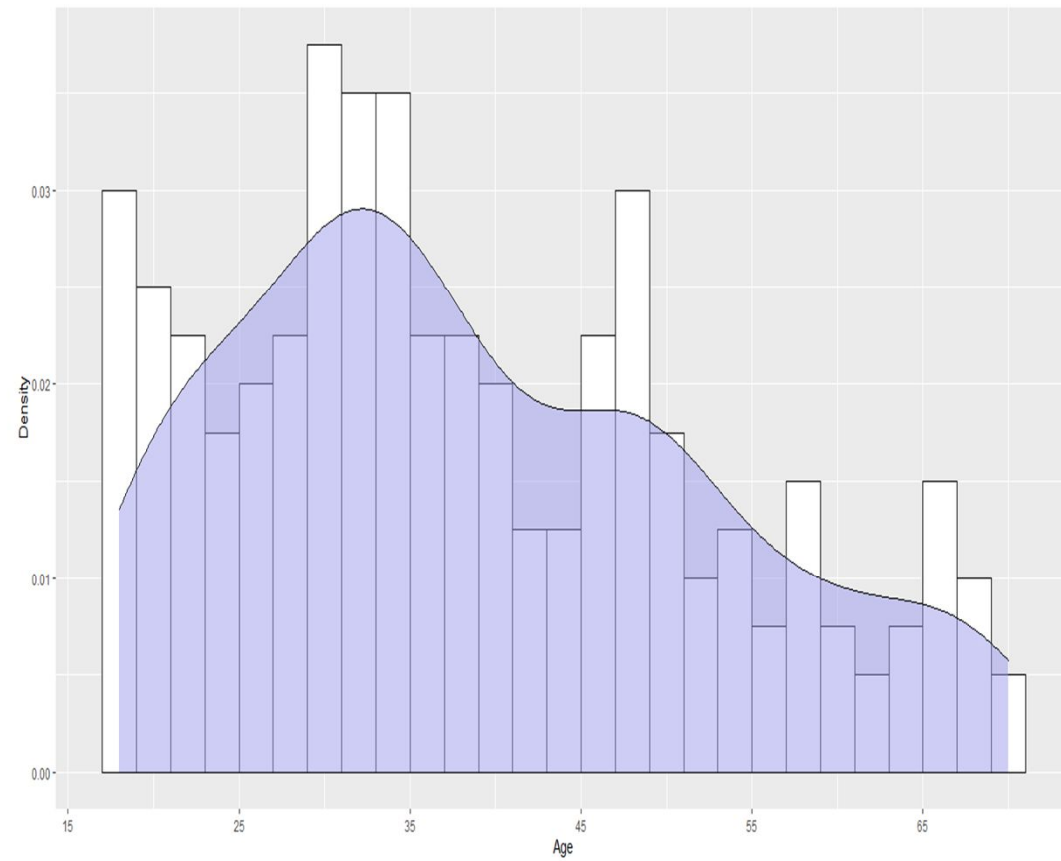
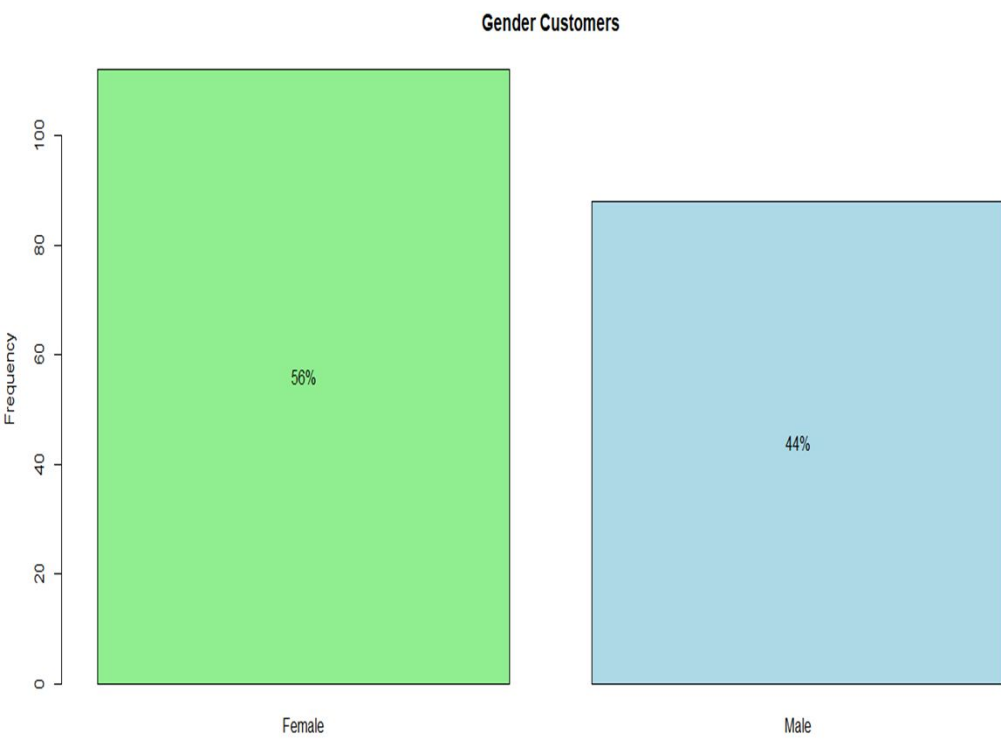
**Age:** Età del cliente

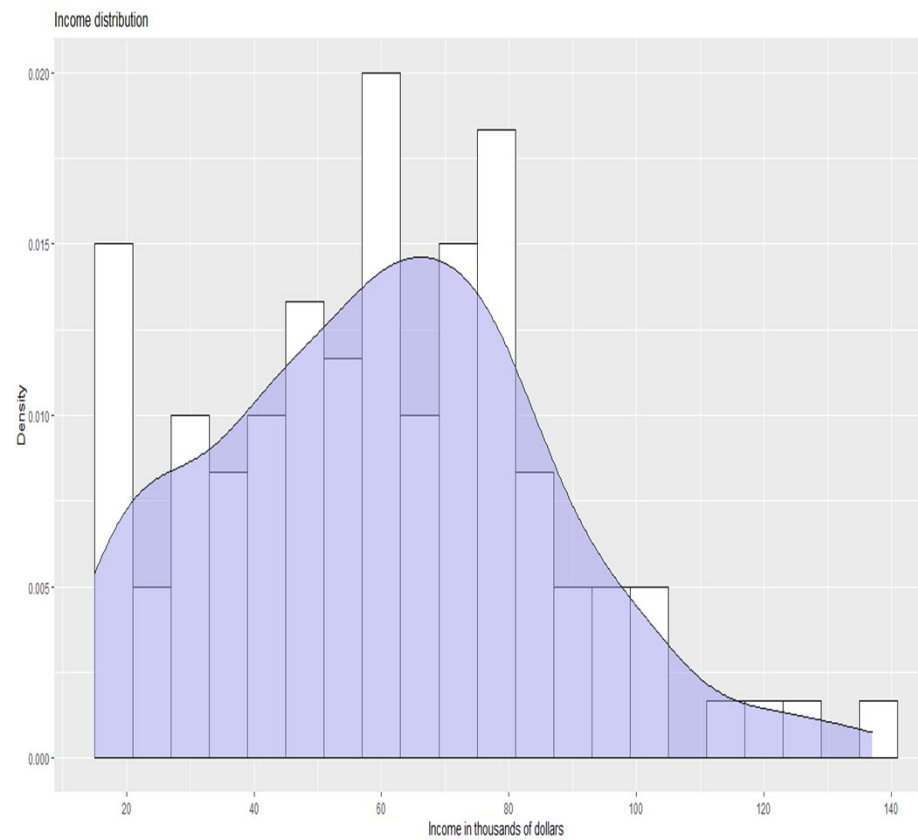
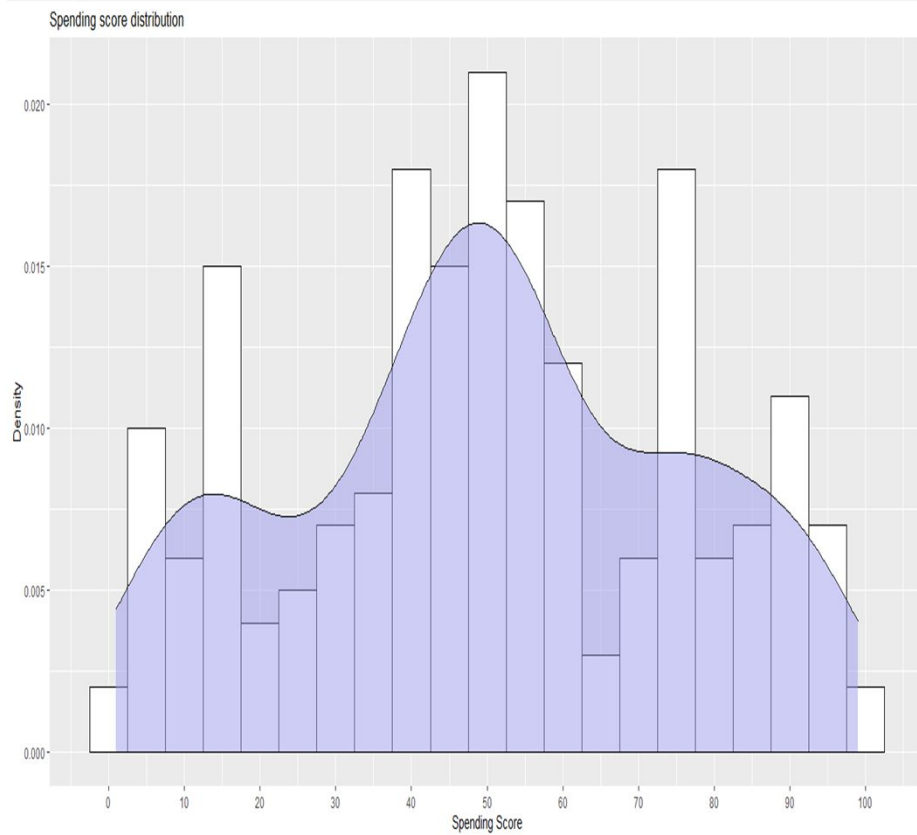
**Annual\_Income:** reddito annuale percepito dal cliente in migliaia di dollari \$

**Spending\_Score:** punteggio attribuito ad ogni cliente in base ai suoi comportamenti di acquisto(0-100)

```
C:/Users/ridol/OneDrive/Desktop/Project.R/ ↗
> str(customers) #check the structure of dataset;
spec_tbl_df [200 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ CustomerID      : num [1:200] 1 2 3 4 5 6 7 8 9 10 ...
 $ Gender          : Factor w/ 2 levels "0","1": 2 2 1 1 1 1 1 1 2 1 ...
 $ Age            : num [1:200] 19 21 20 23 31 22 35 23 64 30 ...
 $ Annual Income (k$) : num [1:200] 15 15 16 16 17 17 18 18 19 19 ...
 $ Spending Score (1-100): num [1:200] 39 81 6 77 40 76 6 94 3 72 ...
 - attr(*, "spec")=
 .. cols(
 ..   CustomerID = col_double(),
 ..   Gender = col_character(),
 ..   Age = col_double(),
 ..   `Annual Income (k$)` = col_double(),
 ..   `Spending Score (1-100)` = col_double()
 .. )
> |
```

# Analisi esplorativa





# Cluster analysis

- **Obiettivo:** raggruppare i clienti per individuare quali tra essi possono essere potenzialmente influenzabili da politiche di marketing

- Variabili considerate:

(clustering gerarchico)

*Age*

*Annual\_Income*

*Spending\_Score*

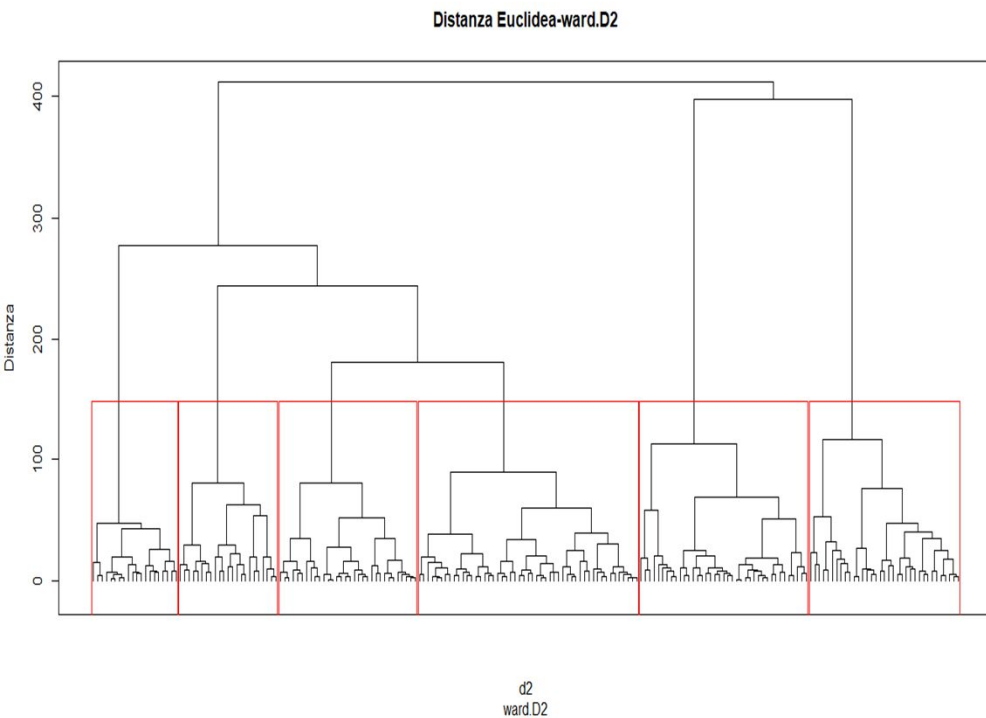
(k-means algorithm)

*Age*

*Annual\_Income*

*Spending\_Score*

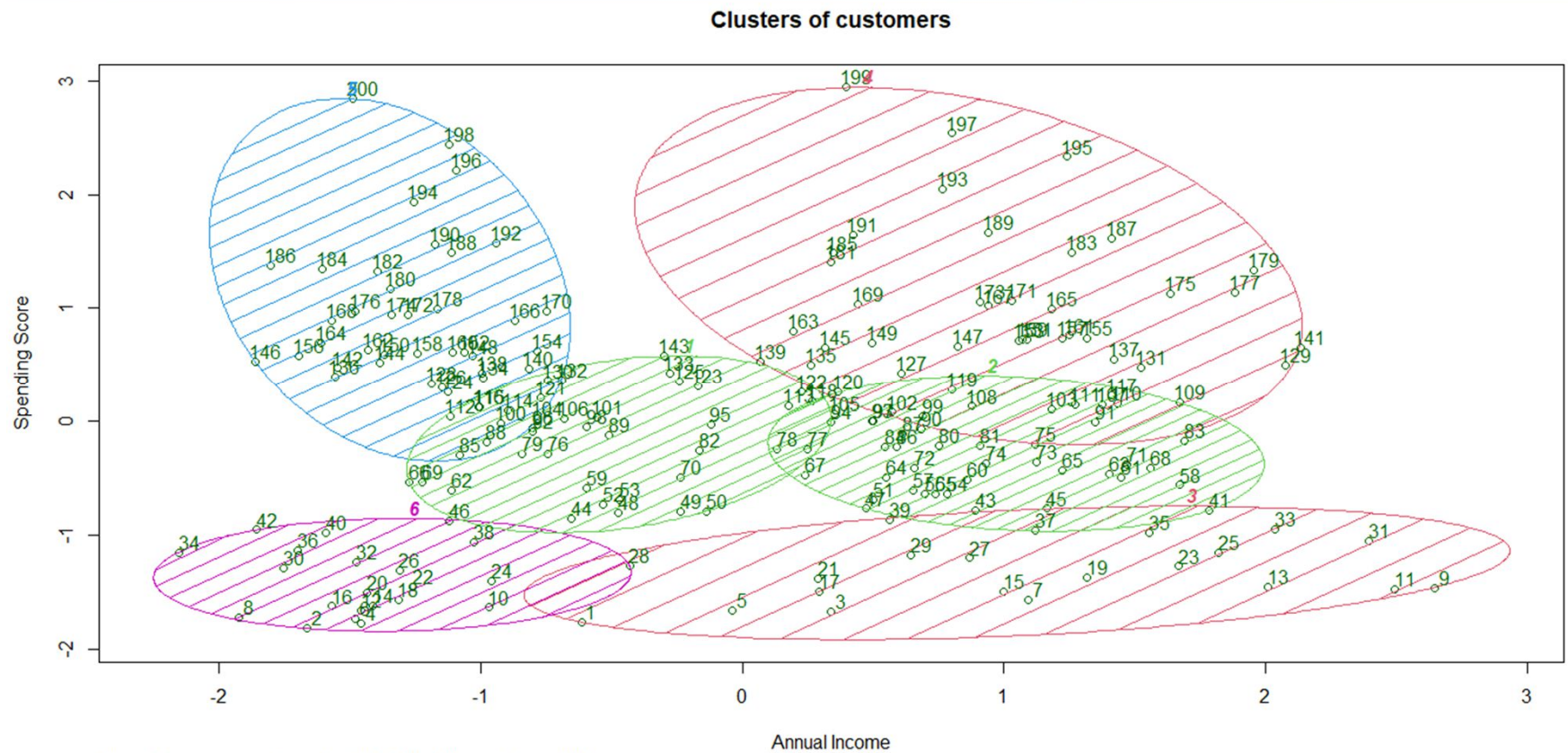
# Risultati Cluster Analysis



	classi	count	mean_Age	mean_AnnualIncome	mean_SpendingScore
1	1	23	45.21739	26.30435	20.91304
2	2	20	24.85000	24.95000	81.00000
3	3	32	24.53125	54.18750	50.25000
4	4	51	53.21569	55.11765	49.47059
5	5	39	32.69231	86.53846	82.12821
6	6	35	41.68571	88.22857	17.28571



# ‘K-Means algorithm’

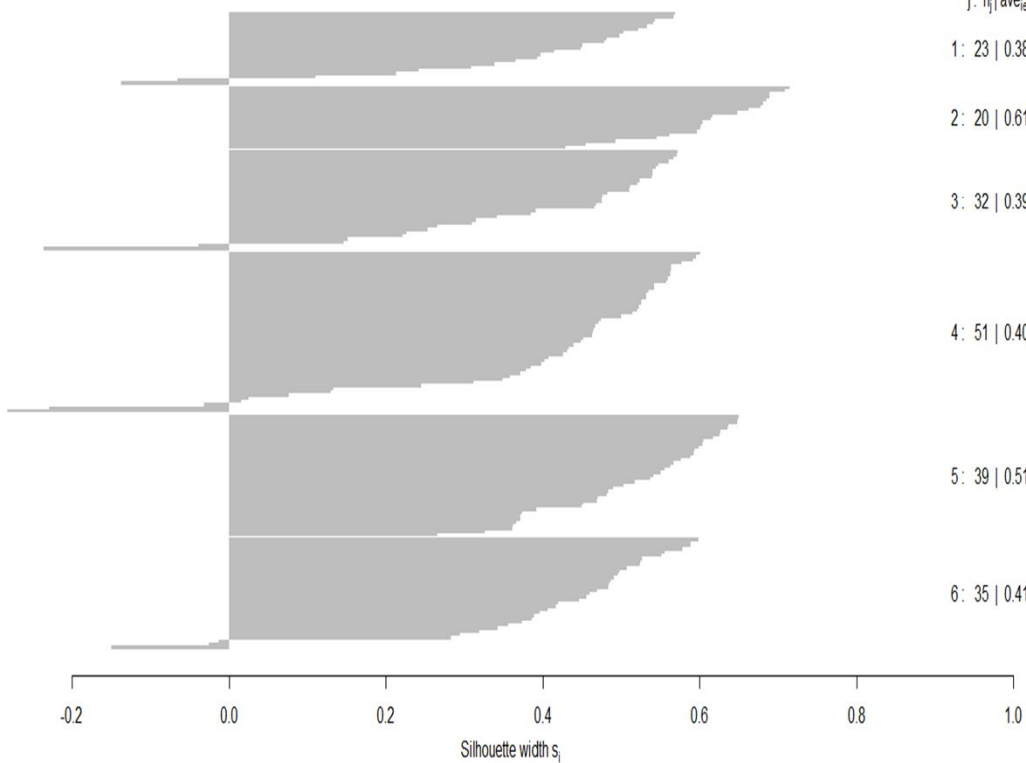


These two components explain 77.57 % of the point variability.

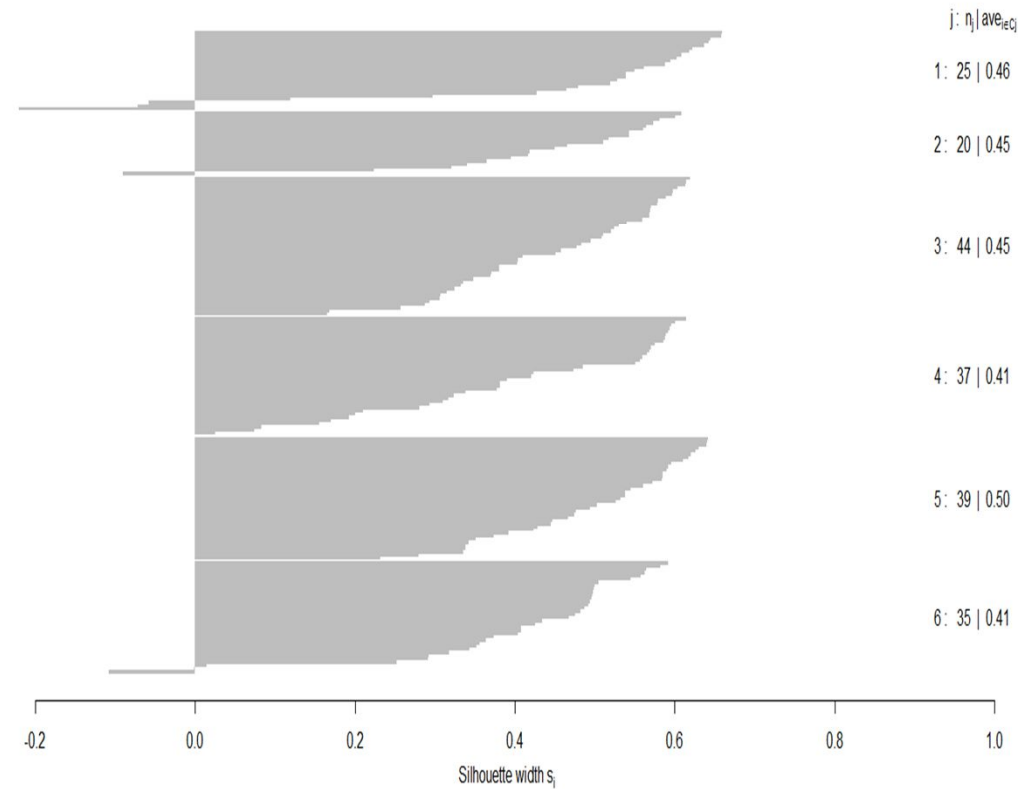
# Silhouette\_score: gerarchico e k-means

Silhouette plot of (x = cutree(gruppi, k = 6), dist = daisy(my\_customers))

n = 200

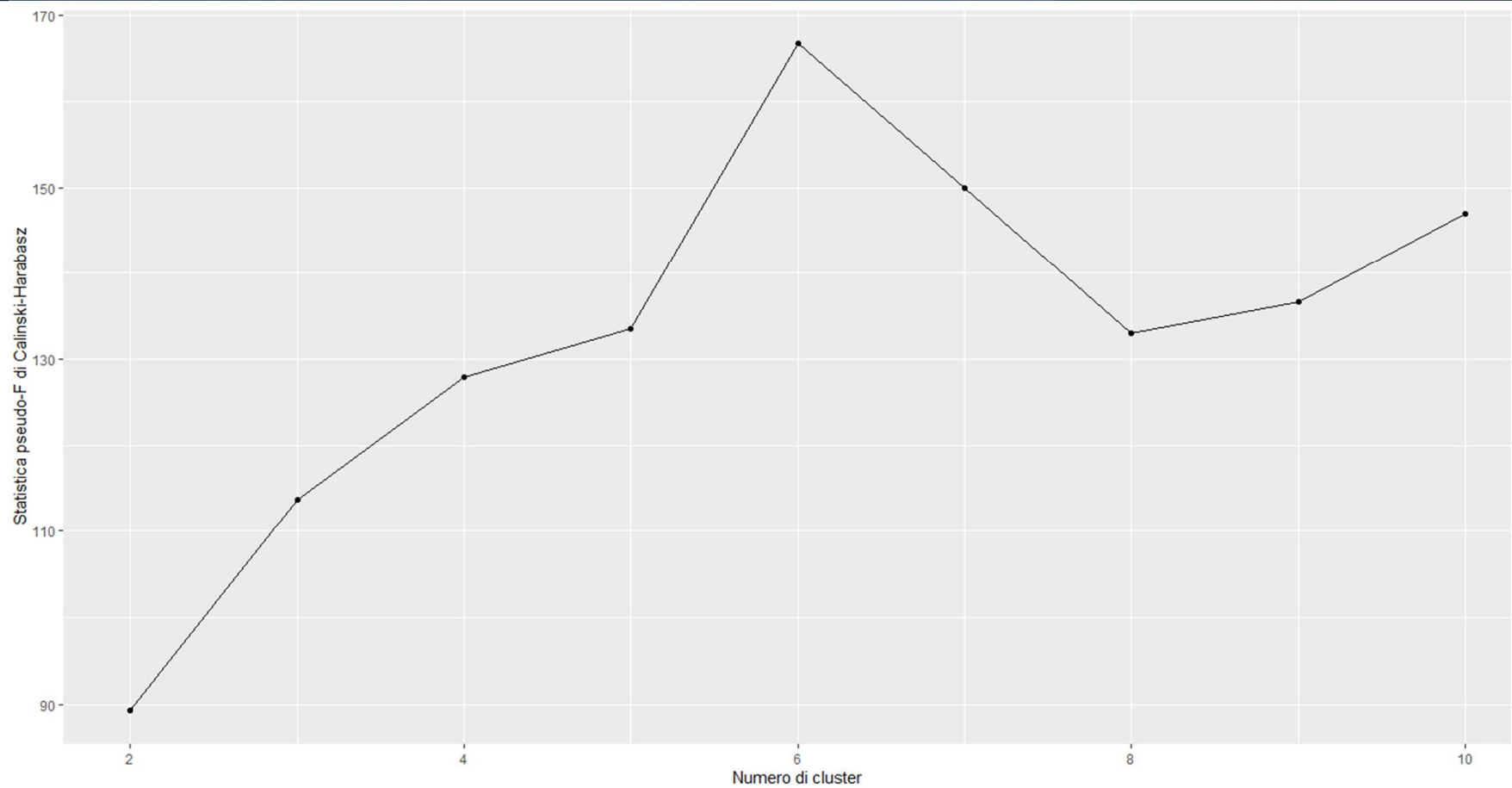


k = 6





# Pseudo F-index



# Risultati raggiunti

- Cluster analysis gerarchica: 6 gruppi
- **Gruppo 1:** annual\_Income basso, Spending\_Score basso
- **Gruppo 2:** Annual\_Income basso, Spending\_Score alto
- **Gruppo 3 e 4:** Annual\_Income e Spending\_Score nella media
- **Gruppo 5:** Annual\_Income alto, Spending\_Score alto
- **Gruppo 6:** Annual\_Income alto, Spending\_Score basso



# Risultati raggiunti

- Cluster Analysis con 'K-Means algorithm': 6 gruppi
- **Gruppo 1:** Annual\_Income basso, Spending\_Score alto, età media bassa
- **Gruppo 2:** Annual\_Income alto, Spending\_Score basso, età media 40 anni
- **Gruppo 3:** Annual\_Income basso, Spending\_Score basso, età media 45 anni
- **Gruppo 4:** Annual\_Income e Spending\_Score nella media, età media bassa
- Gruppo 5: Annual\_Income e Spending\_Score alto, età media 33 anni
- Gruppo 6: Annual\_Income e Spending\_Score nella media, età media 55 anni

# Grazie per l'attenzione

- Lavoro realizzato da *Ridolfi Lorenzo*