

CLUSTER ANALYSIS

On Open Management Dataset

Federico Rausa
Gianluca Cerea
Pablo Torres



14/10/2024



Descrizione del dataset Open Manager

Dataset costituito da:

- 50 unità statistiche, persone (manager) che compilano il questionario
- ogni domanda presenta 4 possibili opzioni :
 - 1) fortemente in disaccordo
 - 2) parzialmente in disaccordo
 - 3) parzialmente in accordo
 - 4) fortemente in accordo
- 30 domande, di cui 4 “reverse”

FONTE : **OPEN MANAGEMENT IN OPEN ORGANIZATION / Paolo Bruttini , 2022 Jun 20. 34. ciclo, Anno Accademico 2020/2021.**



Obiettivi

Si svolge un'analisi delle componenti principali (PCA) al fine di identificare e comprendere le dimensioni latenti sulle quali c'è una maggiore differenziazione fra i manager (gli individui che hanno risposto al questionario).

Ogni individuo verrà associato a un gruppo per il quale la componente principale corrispondente avrà valore massimo.

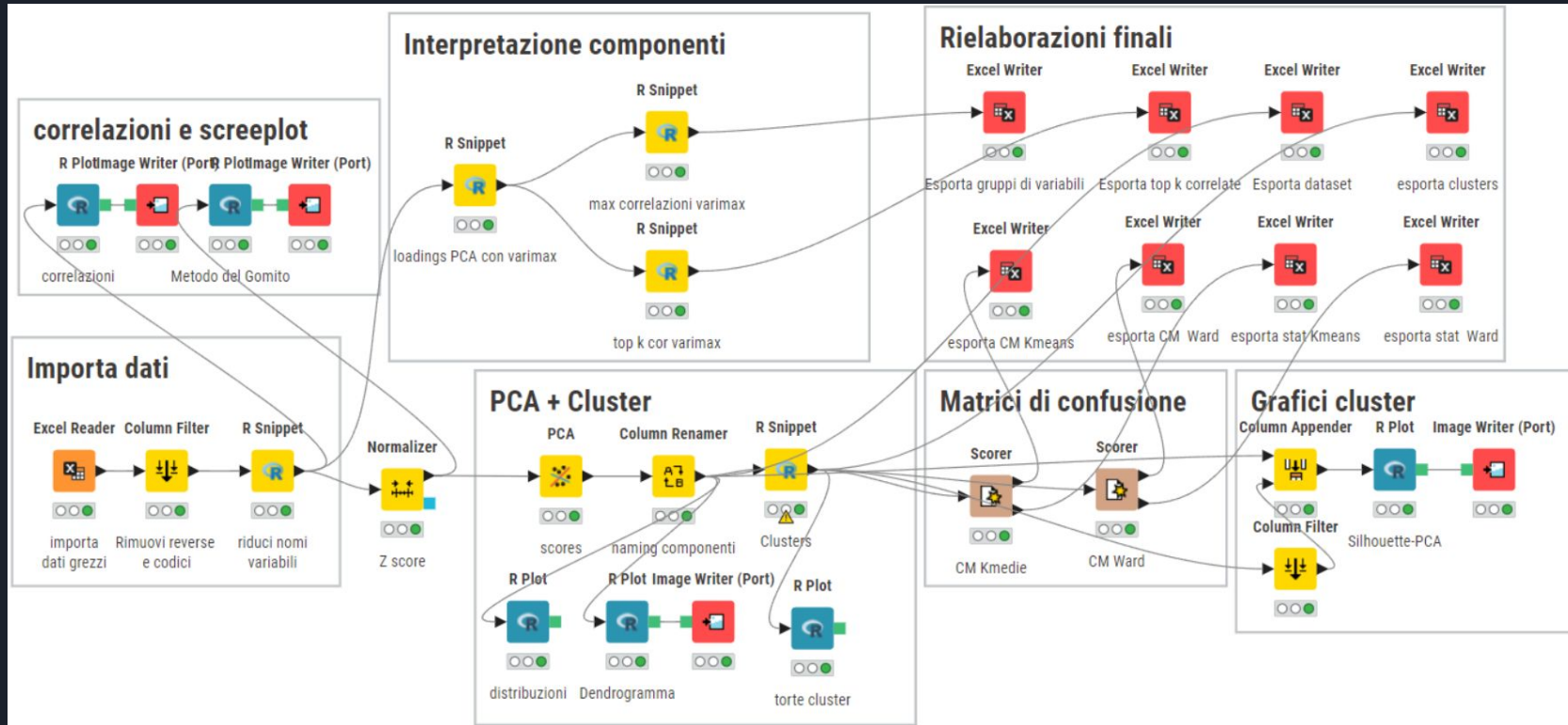
Per valutare l'efficacia della PCA, si confrontano i suoi risultati con quelli di due diversi algoritmi di clustering.



Fasi del lavoro

- 1) Pulizia del dataset
- 2) standardizzazione del dataset con Z-score
- 3) Verifica della presenza di correlazioni tra le variabili originarie
- 4) analisi delle componenti principali sul dataset
- 5) Selezione e interpretazione componenti (Scree Plot e varimax)
- 6) assegnazione componenti a istanze in base a PCA score massimo
- 7) K-means clustering
- 8) Ward clustering
- 9) Matrici di confusione

Workflow su KNIME, R e Excel

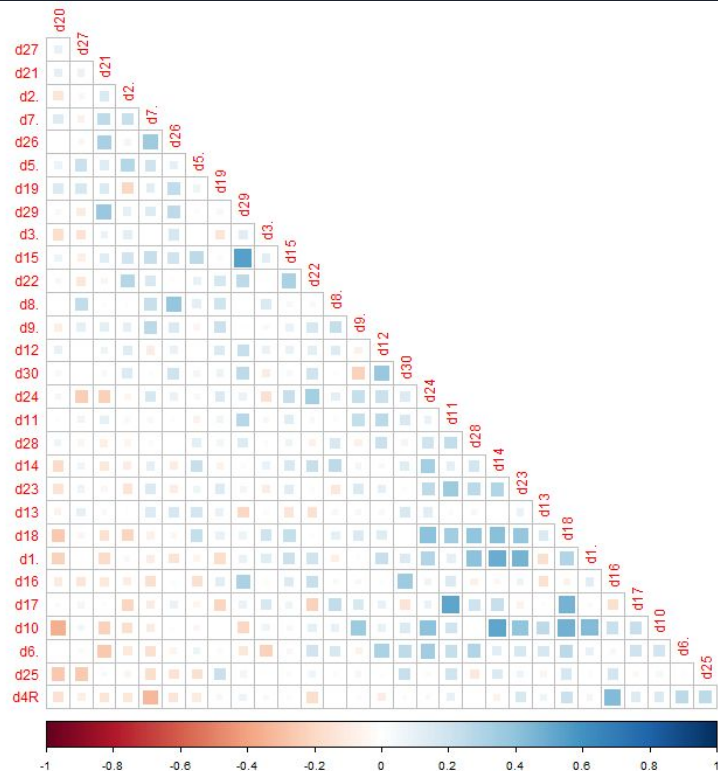




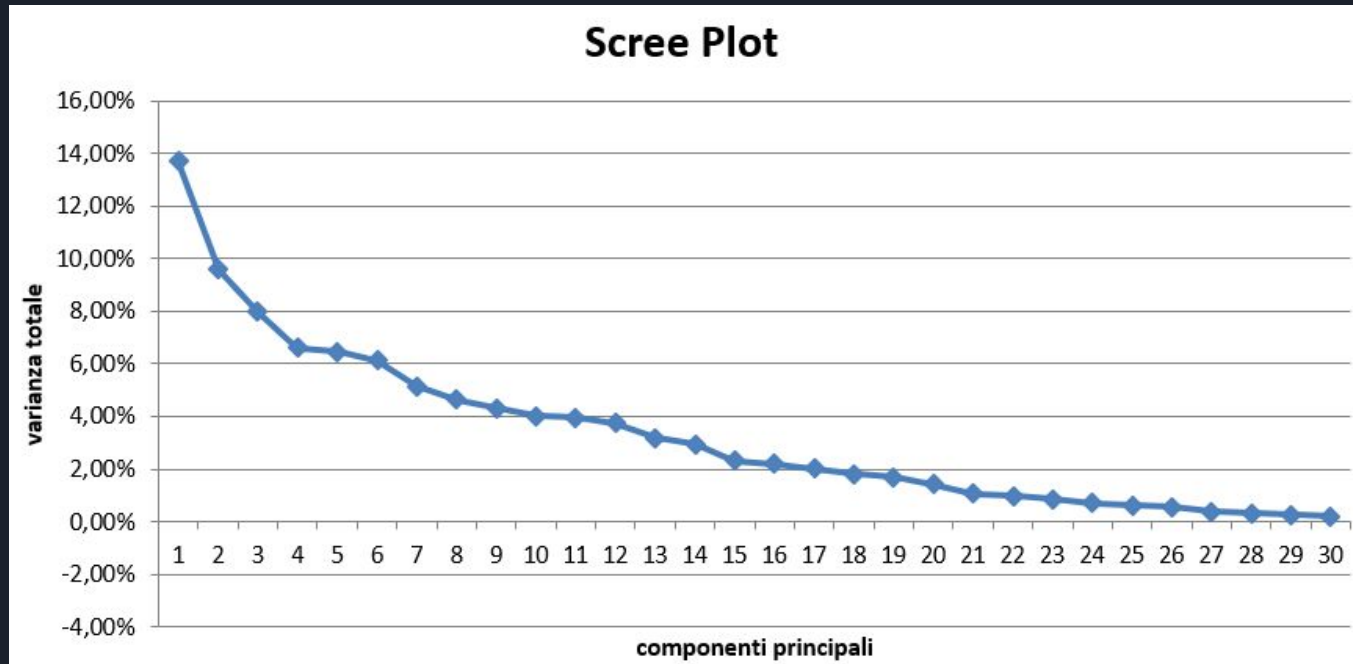
Pulizia del dataset

- rimuovere le 4 domande che non sono “reverse”
- rimuovere la colonna “codici”

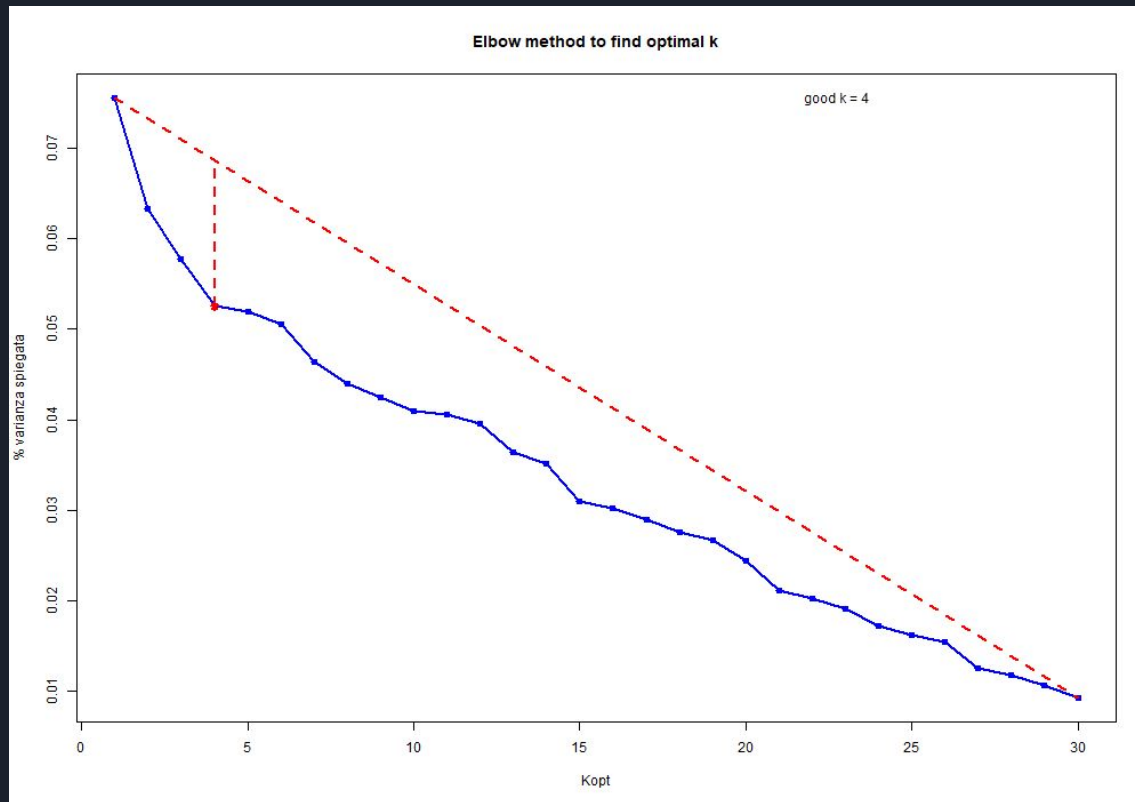
Correlazioni fra le variabili originarie



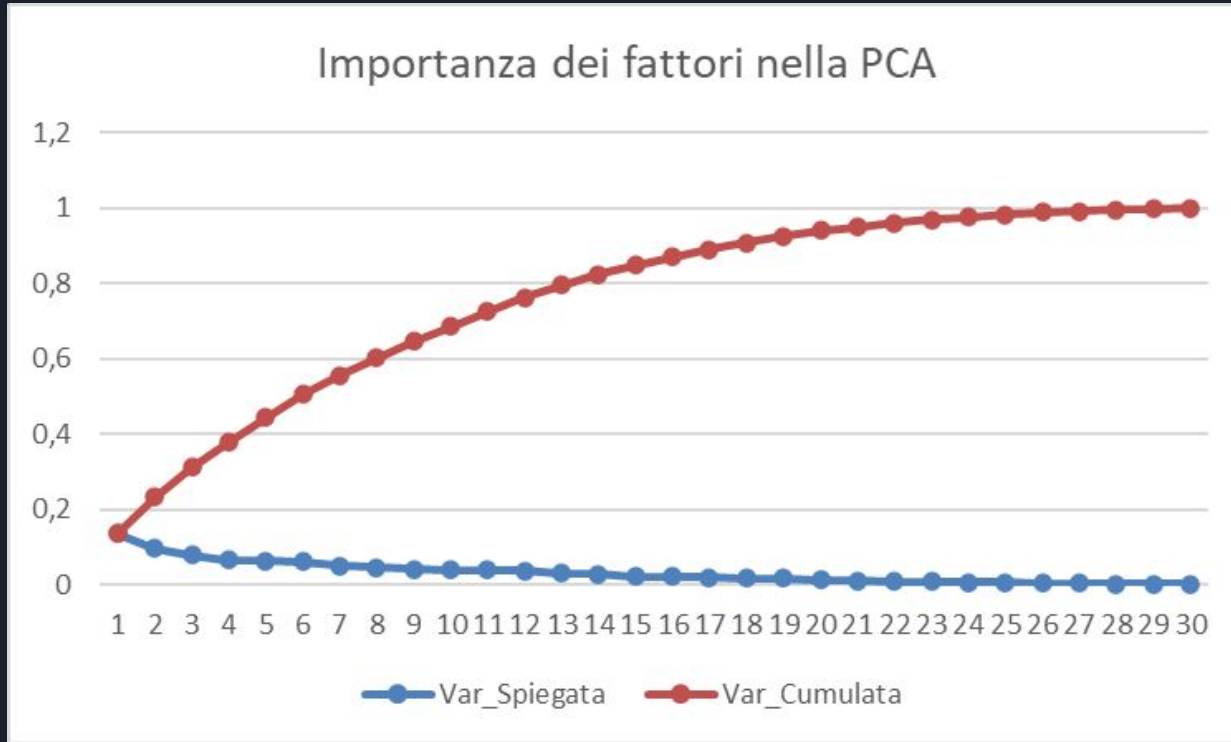
Selezione del numero di componenti: scree plot



Selezione del numero di componenti: metodo del gomito sullo scree plot



Selezione del numero di componenti: criterio della varianza% spiegata (C=4 => ExpIVar=38%)



Denominazione componenti PCA con varimax: top 5 relazioni per componente

Fiducia	Collaborazione	Indipendenza	Umiltà
PC1	PC2	PC3	PC4
d10 (0.747)	d26 (0.639)	d16 (0.69)	d22 (0.602)
d18 (0.712)	d29 (0.605)	d29 (0.541)	d2. (0.569)
d14 (0.697)	d21 (0.561)	d25 (0.476)	d17 (-0.563)
d1. (0.665)	d11 (0.504)	d30 (0.437)	d7. (0.531)
d24 (0.631)	d8. (0.444)	d27 (-0.411)	d4R (-0.503)

Denominazione componenti PCA con varimax: relazione massima per variabile

Fiducia	Collaborazione	Indipendenza	Umiltà
PC1	PC2	PC3	PC4
d10 (0.747)	d26 (0.639)	d16 (0.69)	d22 (0.602)
d18 (0.712)	d29 (0.605)	d25 (0.476)	d2. (0.569)
d14 (0.697)	d21 (0.561)	d30 (0.437)	d17 (-0.563)
d1. (0.665)	d11 (0.504)	d27 (-0.411)	d7. (0.531)
d24 (0.631)	d8. (0.444)	d13 (-0.355)	d4R (-0.503)
d23 (0.586)	d15 (0.421)		d5. (0.381)
d6. (0.473)	d19 (0.358)		
d28 (0.405)	d12 (0.339)		
d20 (-0.285)	d9. (0.319)		
	d3. (0.189)		

Interpretazione componenti:

Sono state identificate 4 dimensioni del ruolo manageriale, ovvero 4 tipologie di manager:



- **FIDUCIA:** manager ottimisti, si relazionano con tutti e vogliono mettersi alla prova



- **COLLABORAZIONE:** manager che vogliono coordinarsi con gli altri, che sanno comunicare e conoscono gli impatti delle dinamiche di gruppo.

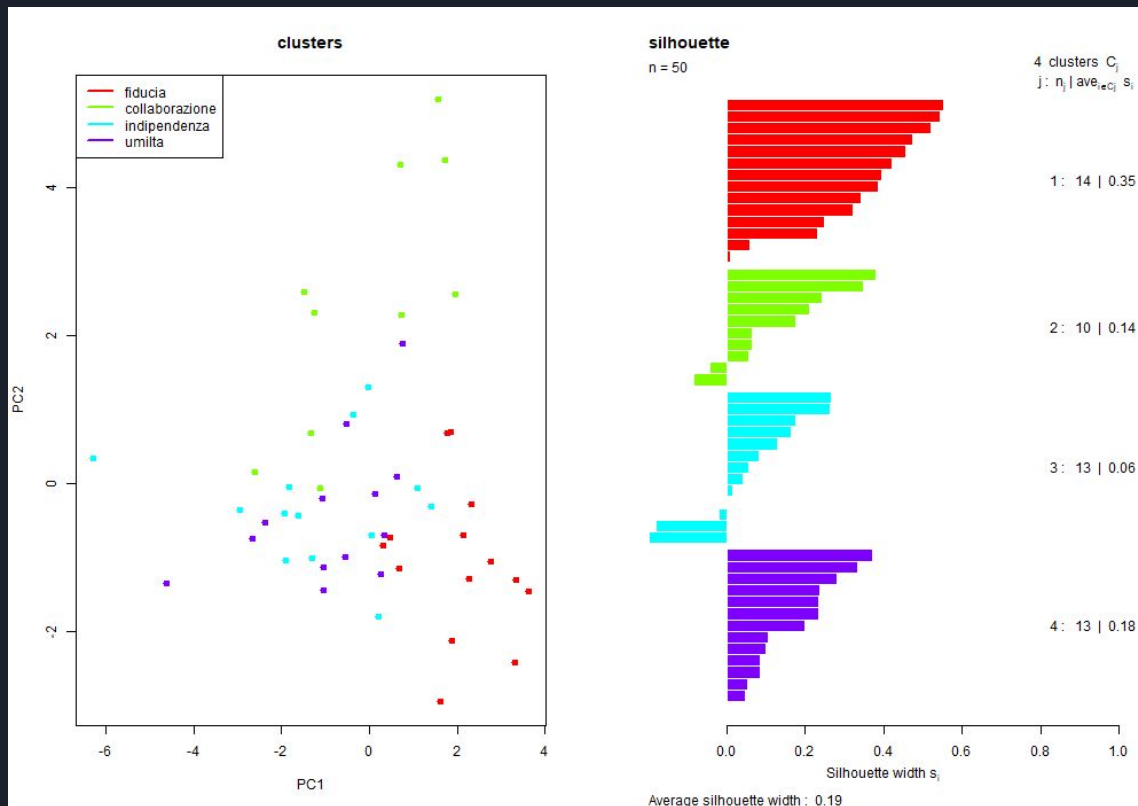


- **INDIPENDENZA:** manager solitari e ambiziosi, che presumono che ciascuno sia in grado di gestire autonomamente le proprie responsabilità. Non vogliono comunicare più del necessario.



- **UMILTA':** manager delegatori e collaborativi, pronti a mettersi in discussione e che ripongono grande fiducia nelle capacità altrui.

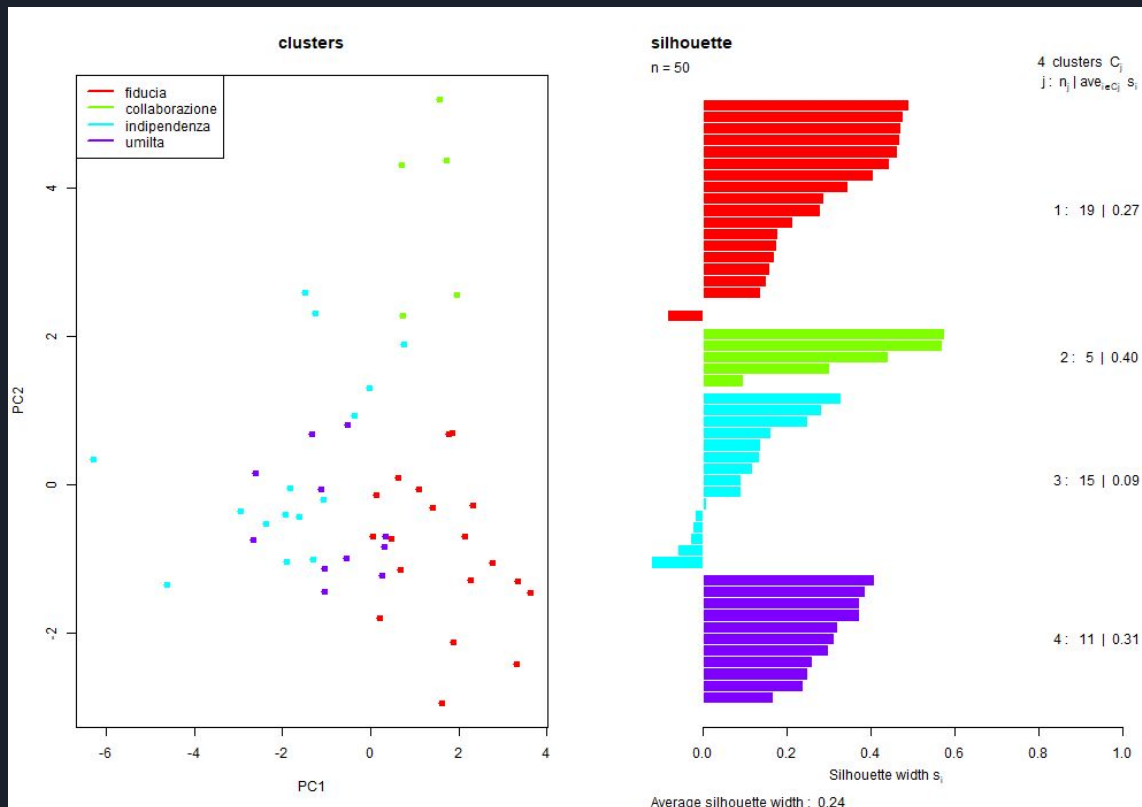
Cluster analysis: massimi score PCA



Cluster Analysis on Open Manager dataset (source from Paolo Bruttini-2022)
Federico Rausa, Gianluca Cerea, Pablo Torres

14/20/2024

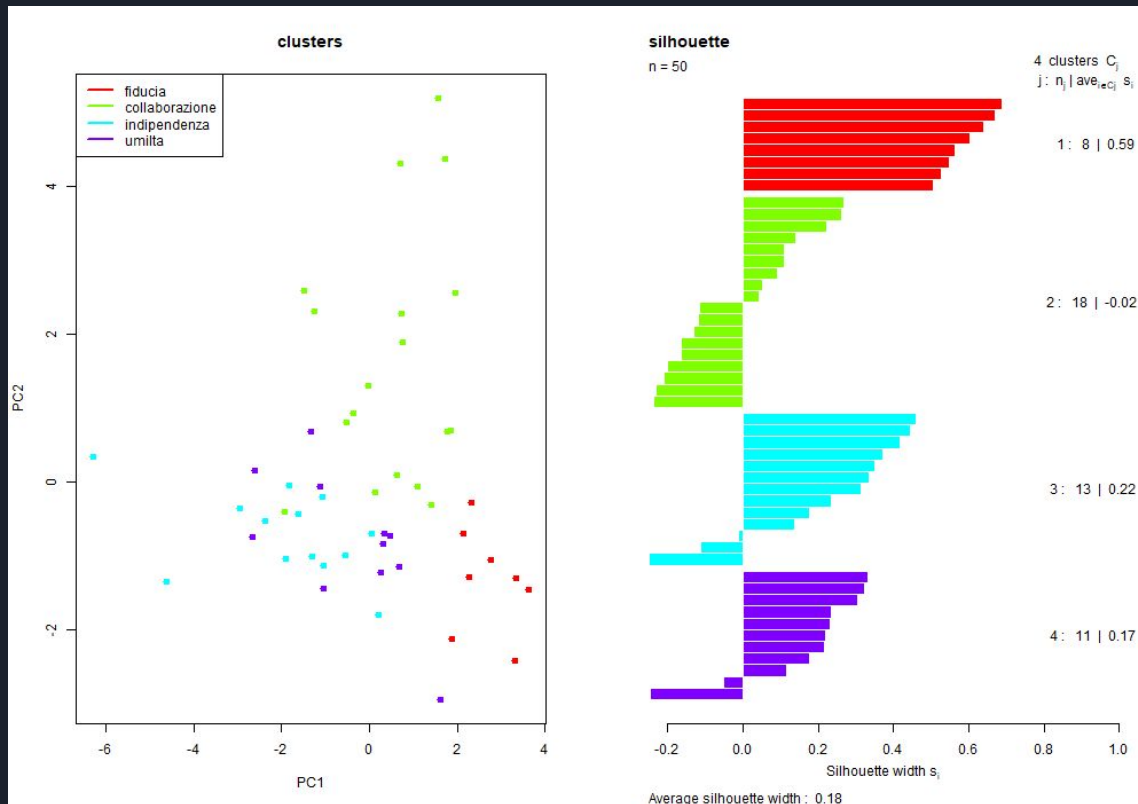
Cluster analysis: metodo delle K medie (centroidi inizializzati con PCA, num iter=1)



Cluster Analysis on Open Manager dataset (source from Paolo Bruttini-2022)
Federico Rausa, Gianluca Cerea, Pablo Torres

14/20/2024

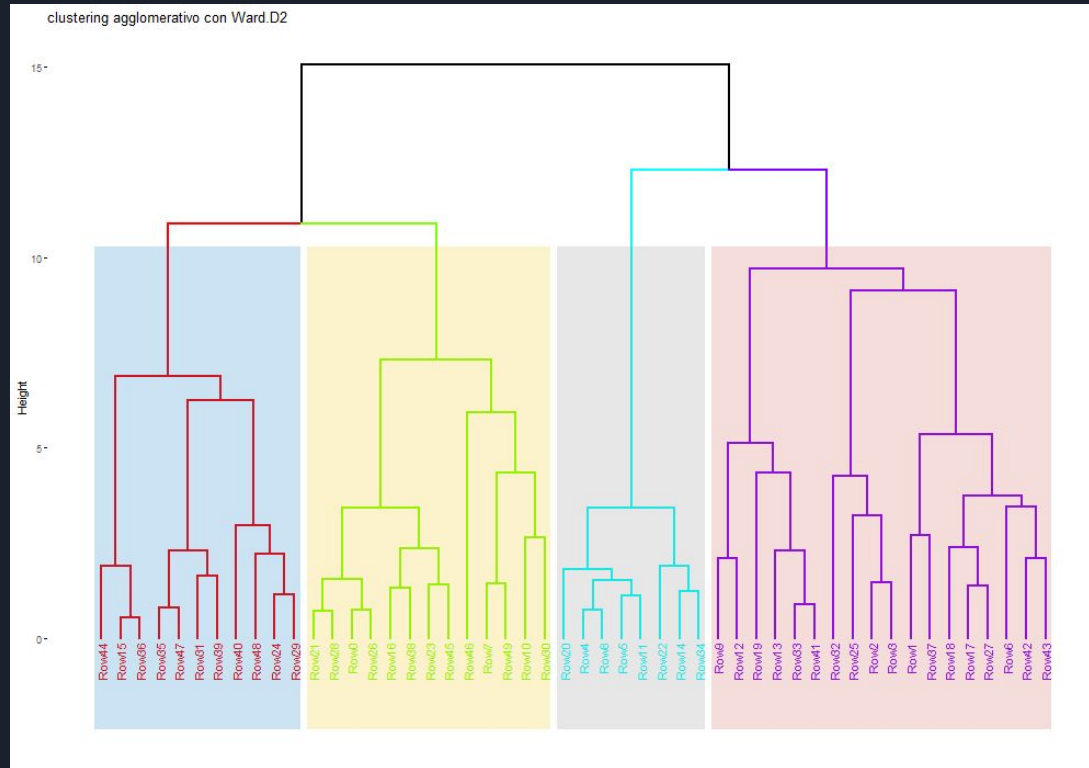
Cluster analysis: metodo di Ward



Cluster Analysis on Open Manager dataset (source from Paolo Bruttini-2022)
Federico Rausa, Gianluca Cerea, Pablo Torres

14/20/2024

Cluster analysis: metodo di Ward

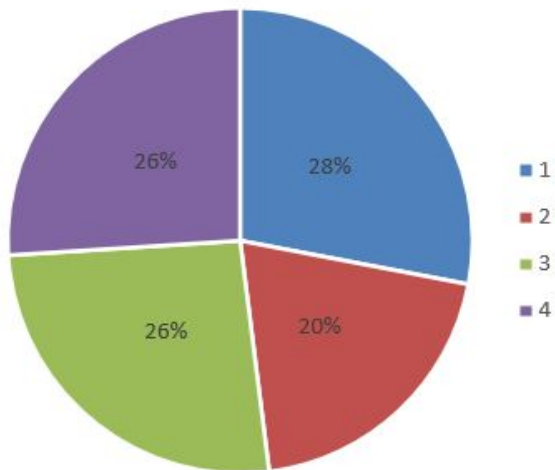


Cluster Analysis on Open Manager dataset (source from Paolo Bruttini-2022)
Federico Rausa, Gianluca Cerea, Pablo Torres

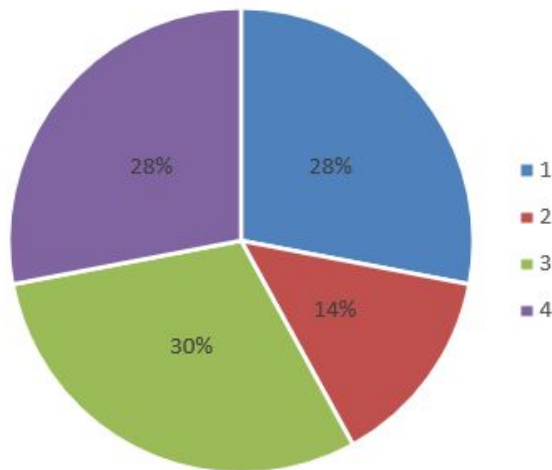
14/20/2024

Cluster analysis: numerosità dei cluster

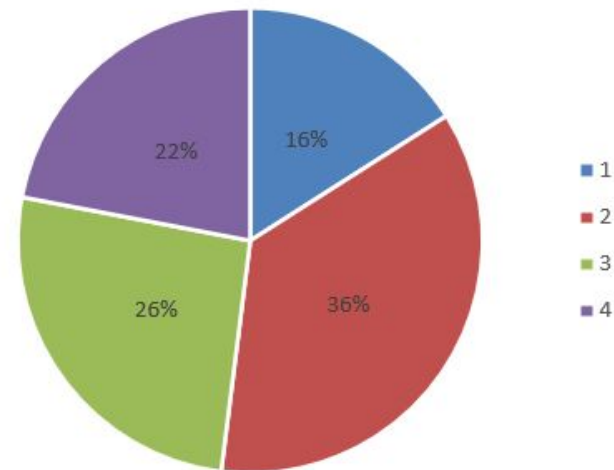
PCA score cluster



K-Medie cluster



Hierarchical-Ward cluster



Cluster analysis: matrici di confusione

		PCA score						PCA score			
		1	2	3	4			1	2	3	4
K-medie	1	13	0	4	2	H-Ward	1	8	0	0	0
	2	0	5	0	0		2	2	7	5	4
	3	0	2	9	4		3	0	0	8	5
	4	1	3	0	7		4	4	3	0	4
	Class	Precision	Sensitivity	Specificity	F-measure		Class	Precision	Sensitivity	Specificity	F-measure
	1	93%	68%	97%	79%		1	57%	100%	86%	73%
	2	50%	100%	89%	67%		2	70%	39%	91%	50%
	3	69%	60%	89%	64%		3	62%	62%	86%	62%
	4	54%	64%	85%	58%		4	31%	36%	77%	33%
	Accuracy	68%					Accuracy	54%			



Question time



Cluster Analysis on Open Manager dataset (source from Paolo Bruttini-2022)
Federico Rausa, Gianluca Cerea, Pablo Torres

14/20/2024



Thank You