



ADVISE ONLY

A DATA SCIENCE LAB PROJECT

DATA SCIENCE LAB

A.A. 2019 - 2020

Crotti Paolo
Filosa Alberto
Garavaldi Davide

TEAM



ALBERTO FILOSA

Si è occupato di:

Introduzione
Pre – Processing
Analisi Esplorativa



PAOLO CROTTI

Si è occupato di:

Cluster Analysis
Cluster Validation



DAVIDE GARAVALDI

Si è occupato di:

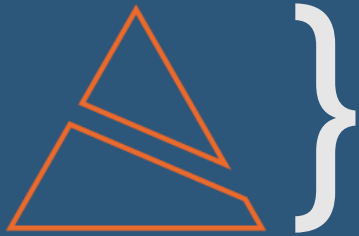
Caratterizzazione Cluster e
Confronto
Valutazioni Finali



LIBRERIE

ggplot2
Tidyverse
Cluster
ClusterR
clValid
fpc

CONTESTO



Azienda di Data Analytics specializzata in
Machine Learning, AI ed Investment Strategy

Wealth Management



Gestione degli investimenti delle persone



5000 persone di diversa Età

OBIETTIVI



Caratteristiche d'Investimento

- Statistiche Descrittive



Segmentazione dei Clienti

- Cluster Analysis
- Similarità - Dissimilarità



Abbinamento

- Cliente ↔ Investimenti

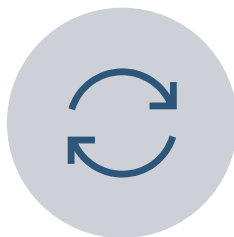


DATASET



SOCIO-DEMOGRAFICI

Provincia
Età
Reddito (Categoriale)
Sesso (Categoriale)
Data avvio



DATI COMPORTAMENTALI

Tendenza a non fidarsi
della banca
Tendenza ad andare in
panico



BISOGNI/OBIETTIVI

Bisogno di reddito
Bisogno di cure a lungo
termine da anziano
Bisogno di protezione
(capitale o persona)
Bisogno di previdenza
integrativa
Bisogno di
ottimizzazione
successoria/fiscale



DNA FINANZIARIO

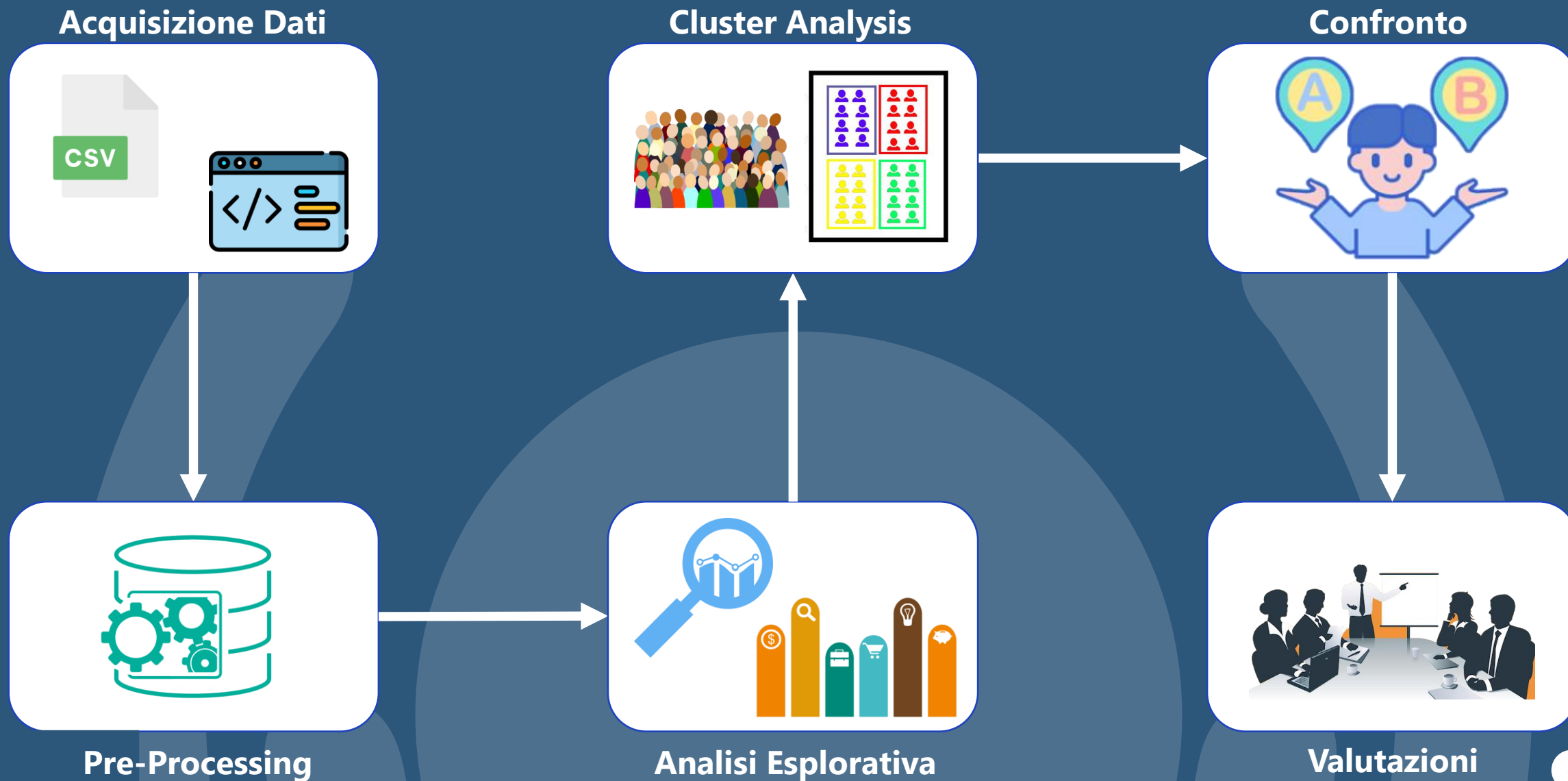
Propensione al rischio
Rischio del portafoglio
Orizzonte
d'investimento
Orizzonte
d'investimento
implicito o ideale
Livello di cultura
finanziaria
Potenziale di crescita



INVESTIMENTI EFFETTUATI

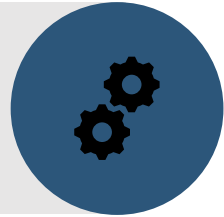
Denaro investito
% di investimento in
obbligazioni
% di investimento in
azioni
% di investimento in
strumenti monetari
% di investimento in
altri investimenti
% di investimento in
cassa

PROCEDIMENTO



PRE-PROCESSING

DATA MANIPULATION



Ricodifica Variabili Categoricali

Provincia NA (Napoli)

Normalizzazione Variabili

One Hot Encoding

Eliminate

Provincia

Sesso

Inizio Rapporto con il Cliente

Fiducia nelle Banche



VARIABLE SELECTION

ANALISI ESPLORATIVA

🔍 Analisi delle Variabili e Divisione in 2 sfere di interesse:

- Investment
- Profile

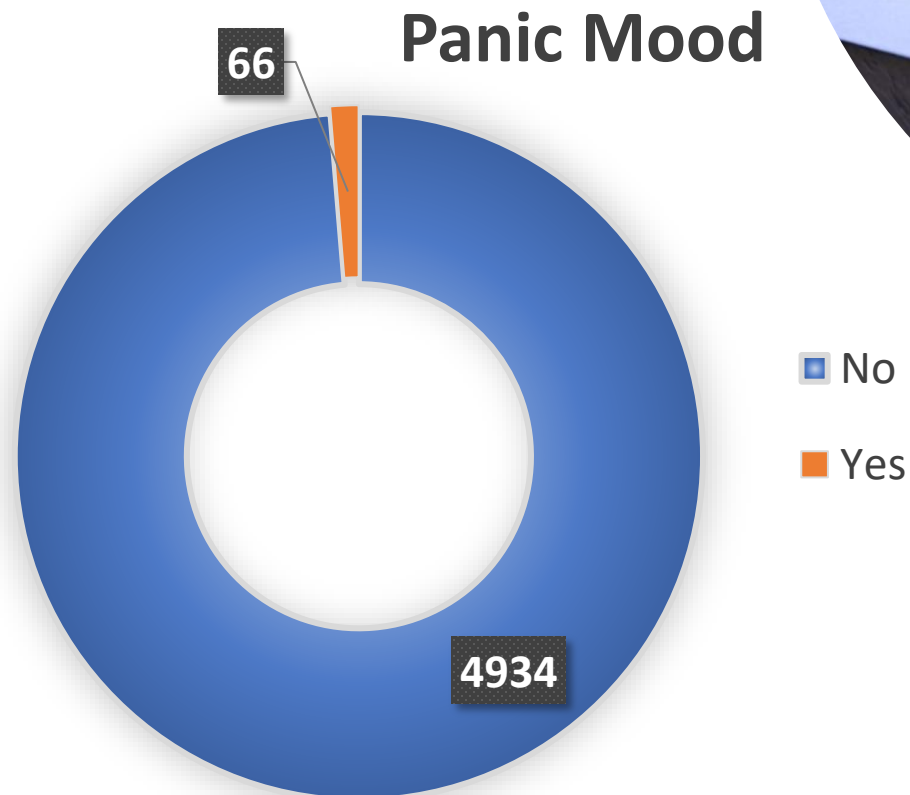
🔍 Near Zero Variance:

- Panic Mood

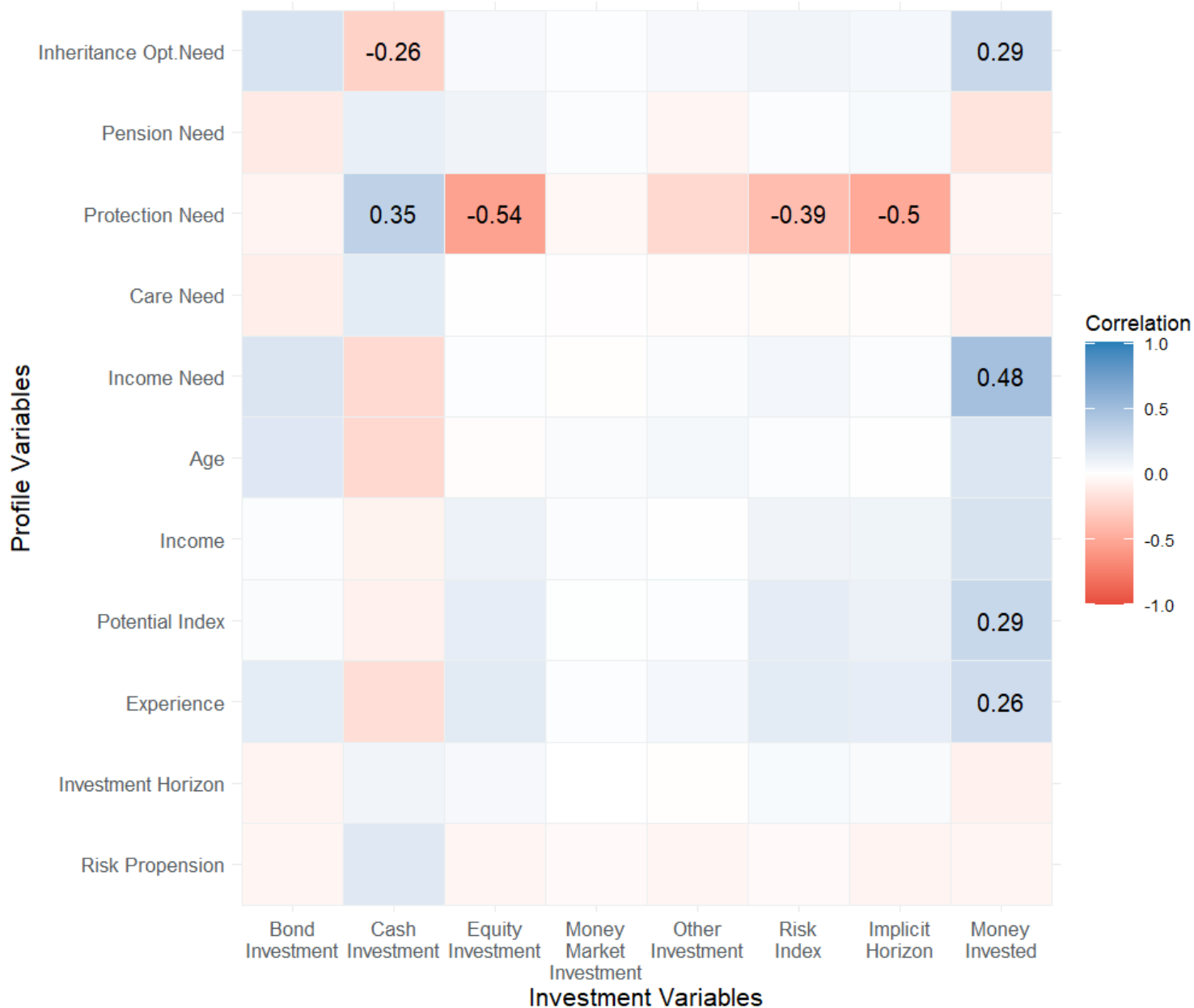
🔍 No Valori Mancanti

🔍 Correlazione

🔍 Outliers

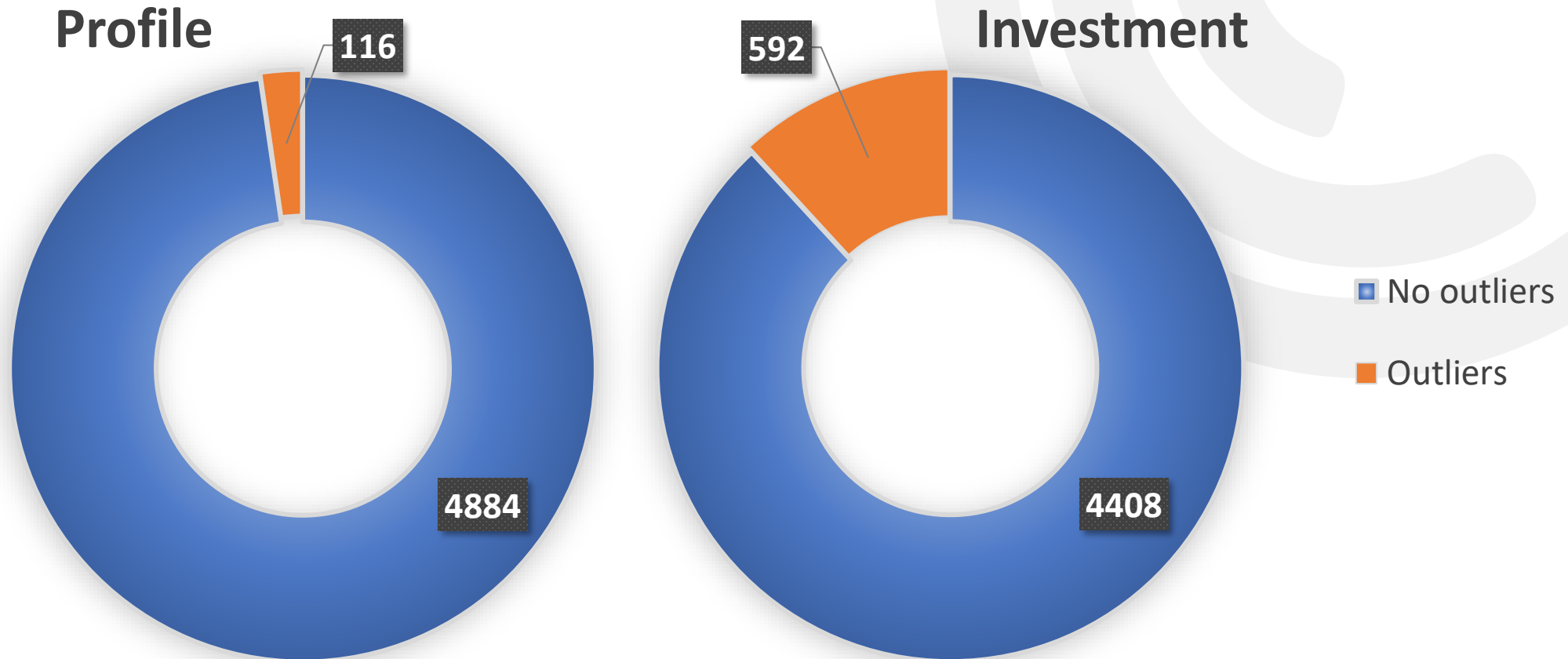


CORRELAZIONE



Matrice di correlazione incrociata tra variabili relative agli investimenti e variabili relative ai profili dei clienti

OUTLIERS



- Identificazione effettuata tramite l'algoritmo *DBSCAN*
- I 116 outliers delle variabili legate ai profili dei clienti sono in comune agli outliers delle variabili legate agli investimenti

CLUSTER ANALYSIS - OVERVIEW

ALGORITMI



Hierarchical	Complete
	Median
	Average
	Diana
	Ward
Model Based	Mixture Model using EM
Prototype	PAM
	SOM
	K-Means
	Clara
	SOTA
Fuzzy	Fanny

Internal

Silhouette

Calinski – Harabasz

Connectivity

PearsonGamma

External

Rand

Jaccard

Fowlkes & Mallows

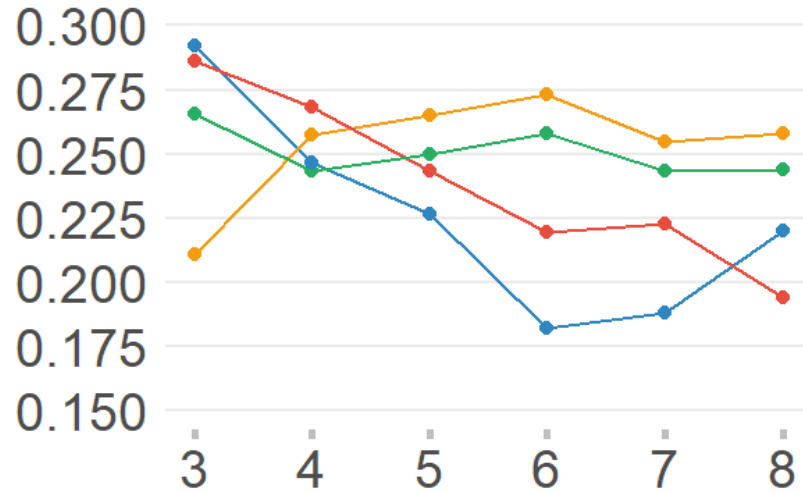
Purity

METRICHE

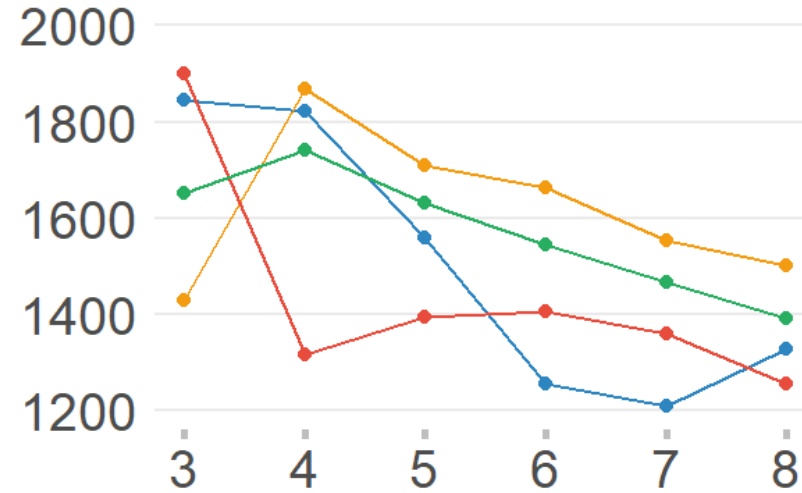


CLUSTER ANALYSIS - PROFILE

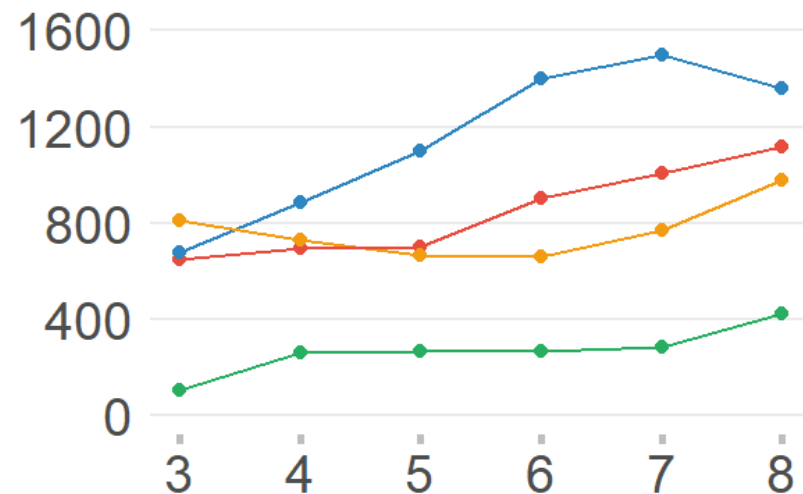
Silhouette



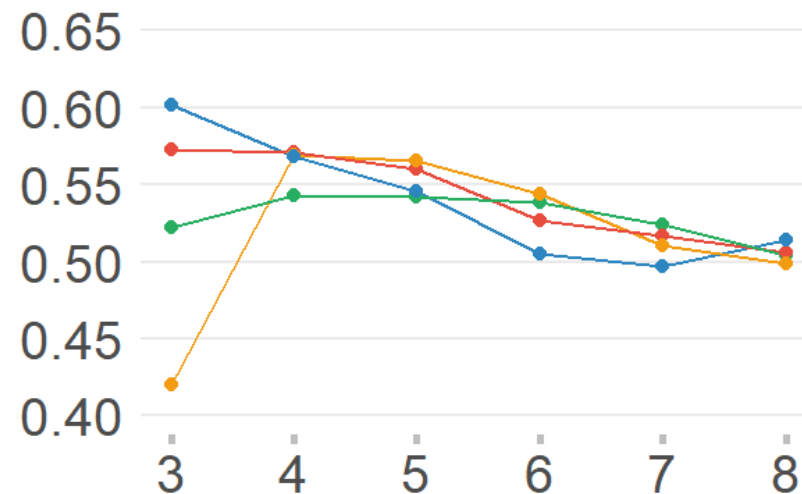
Calinski-Harabasz



Connectivity



Pearsongamma

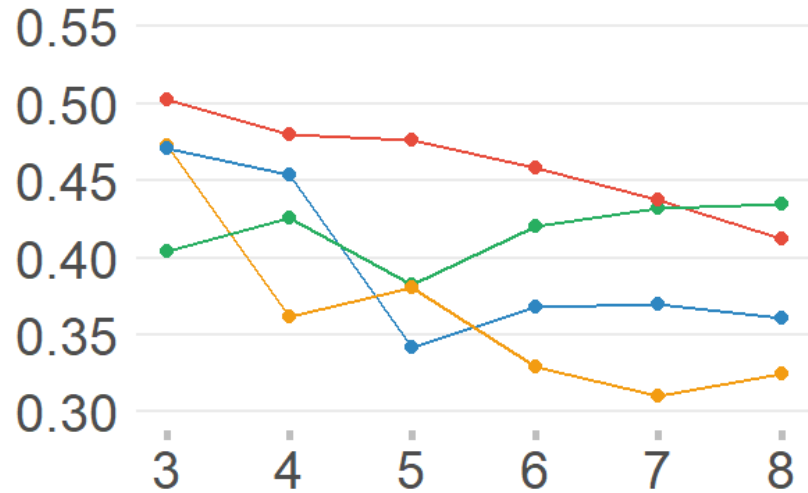


Algorithm

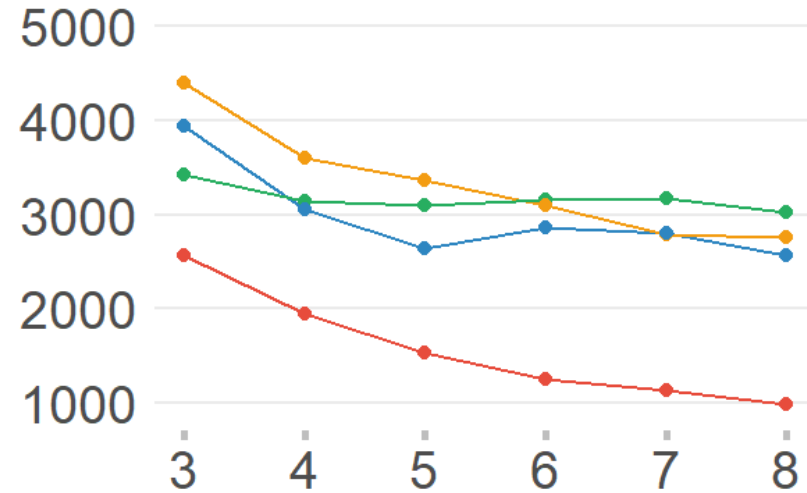
- kmean
- pam
- sota
- ward

CLUSTER ANALYSIS - INVESTMENT

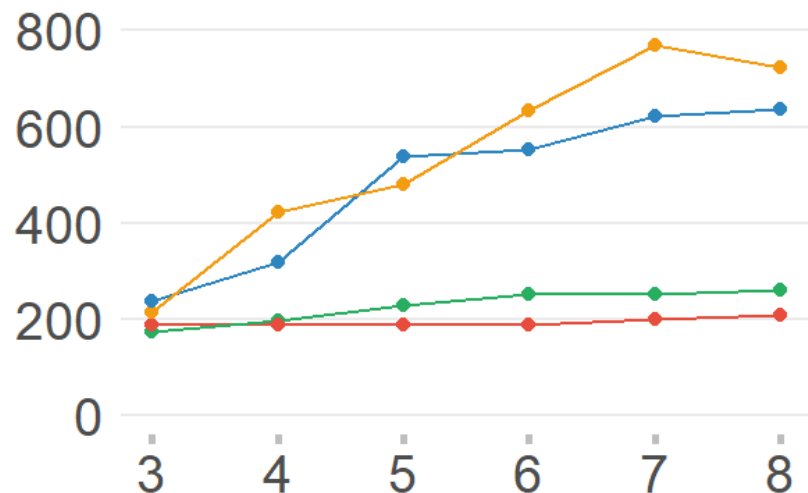
Silhouette



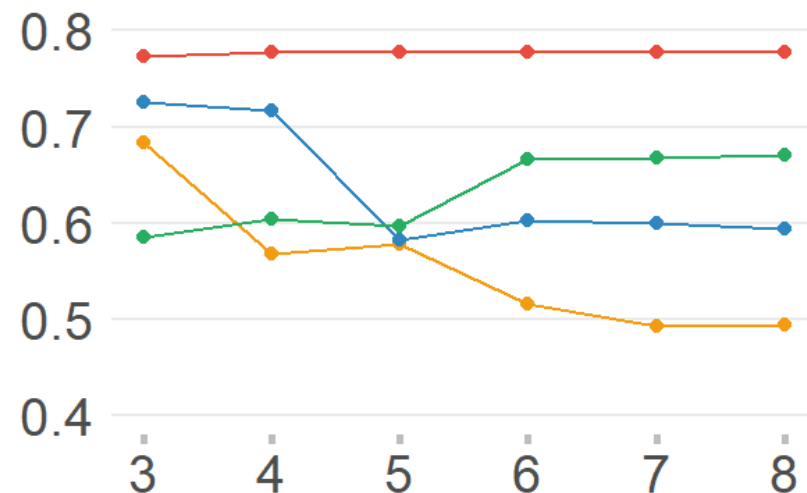
Calinski-Harabasz



Connectivity



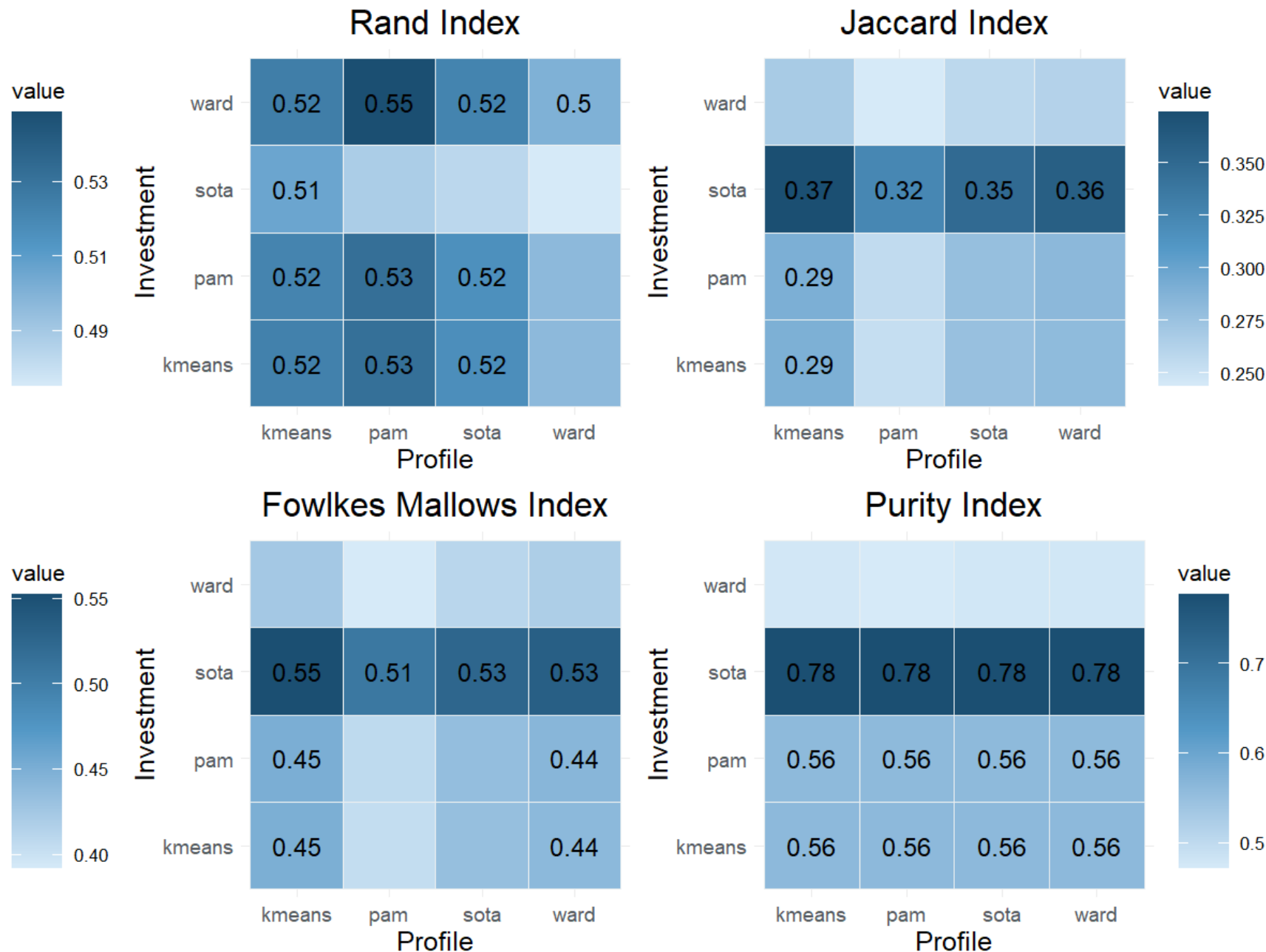
Pearsongamma



Algorithm

- kmean
- pam
- sota
- ward

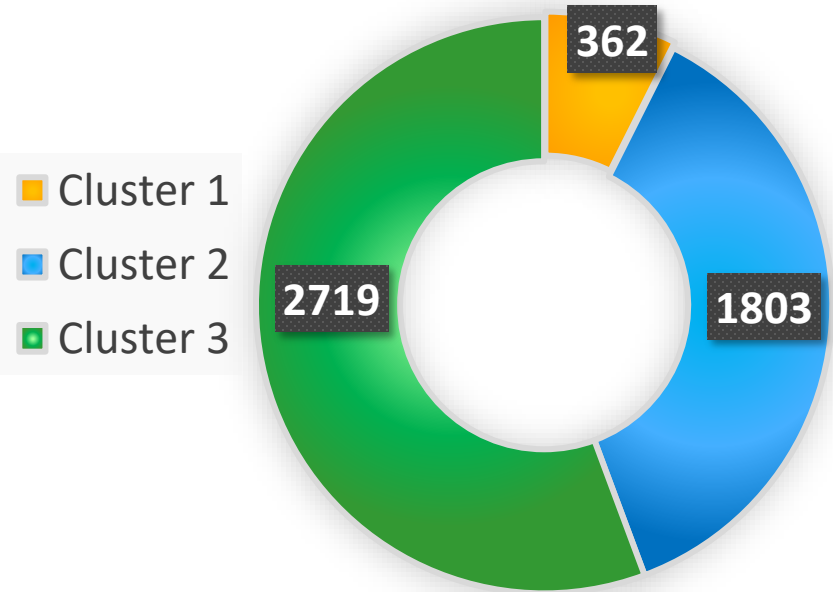
MISURE DI VALIDAZIONI ESTERNE



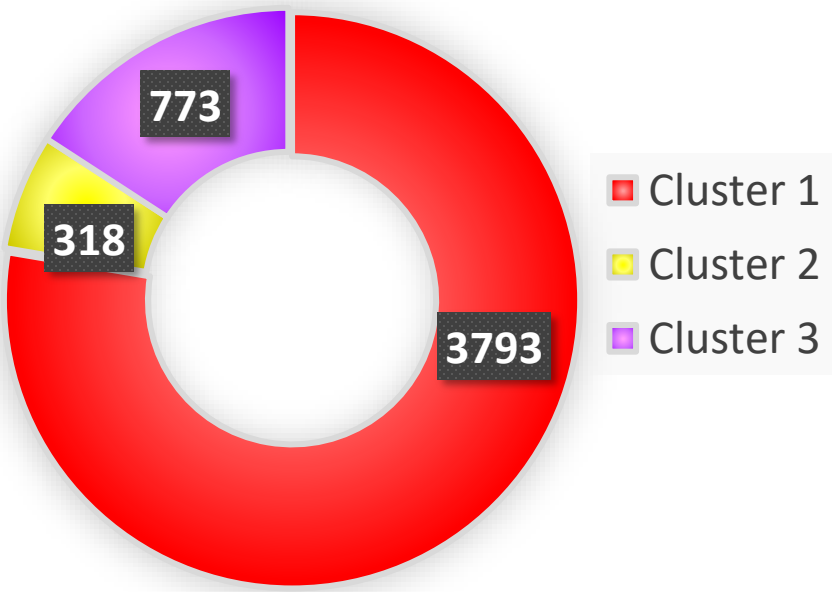
FREQUENZE DEI CLUSTER

		Investment		
		1	2	3
Profile	1	310	32	20
	2	1220	132	451
	3	2263	154	302

Profile

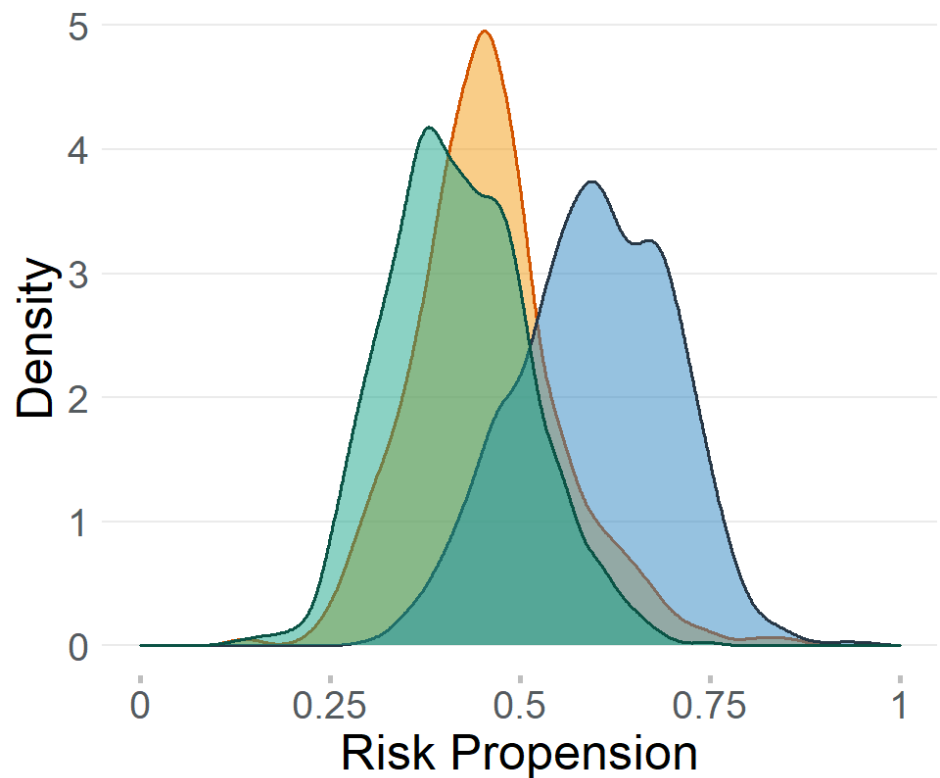


Investment

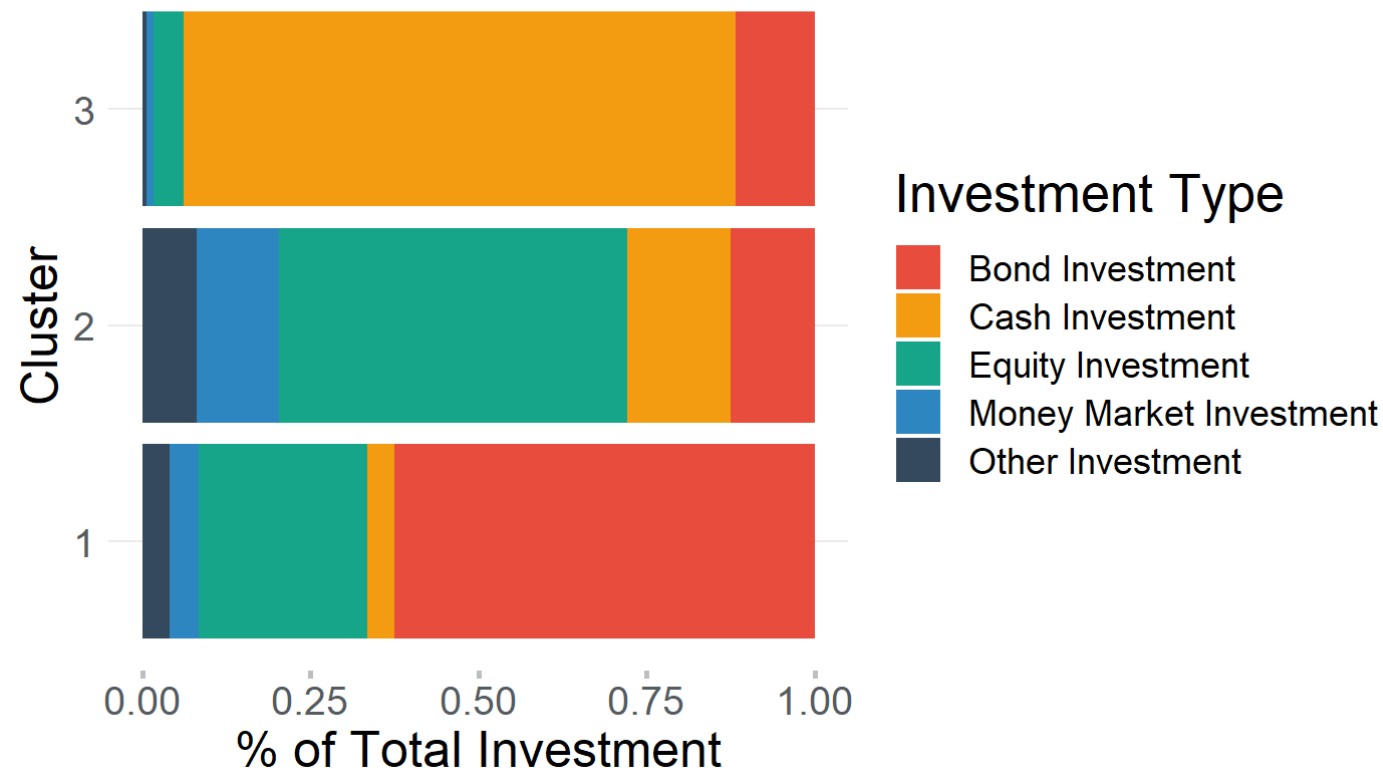


CONFRONTO

Profile - Risk Propension

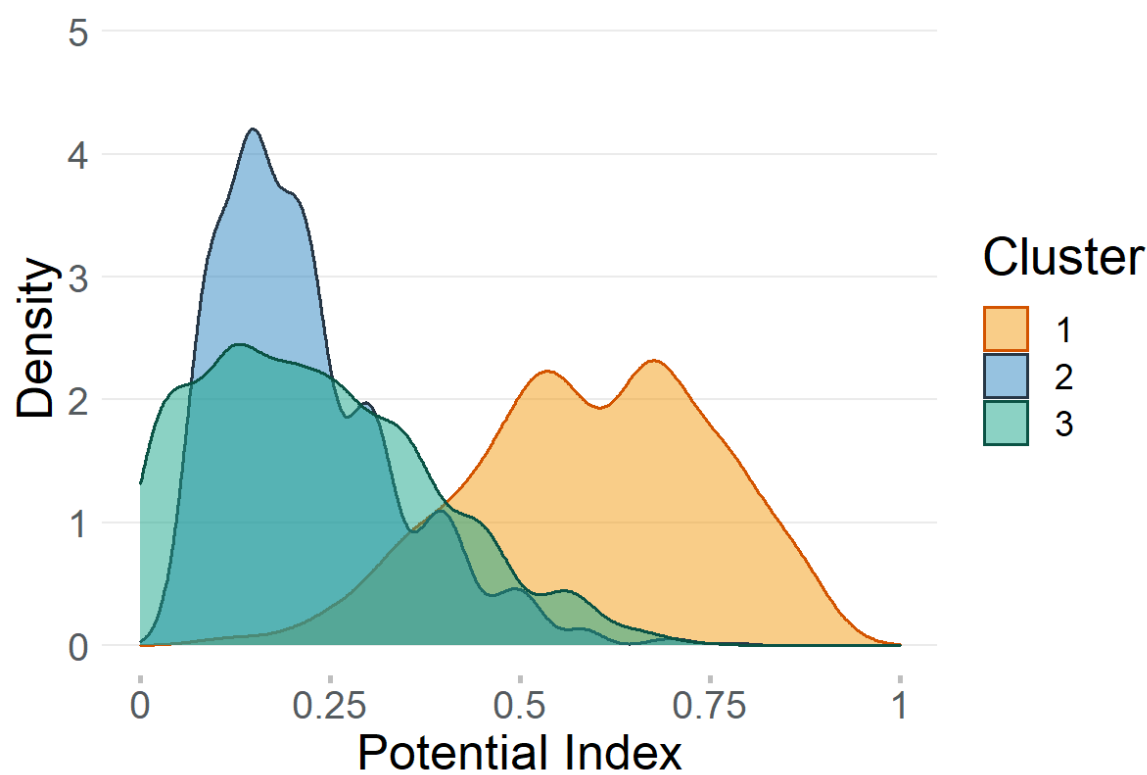


Investment

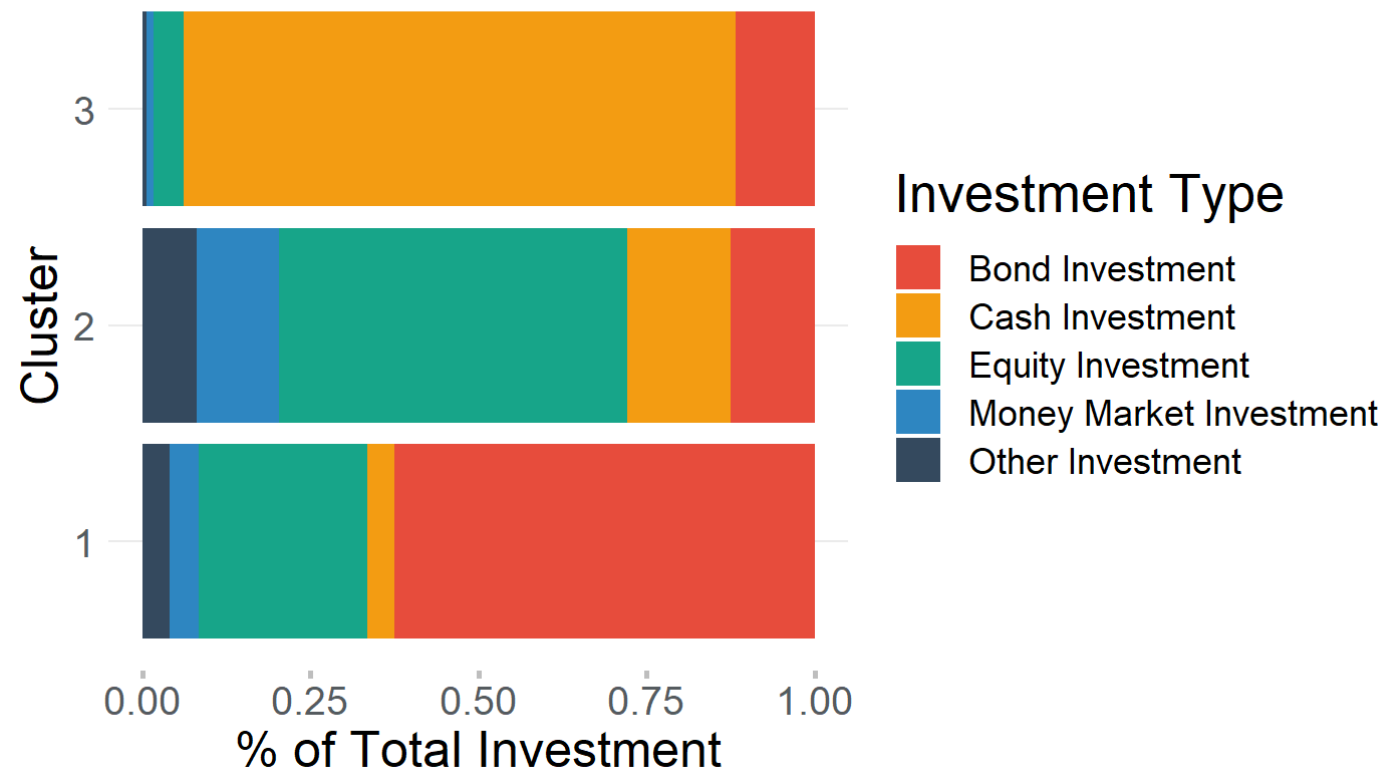


CONFRONTO

Profile - Potential Index

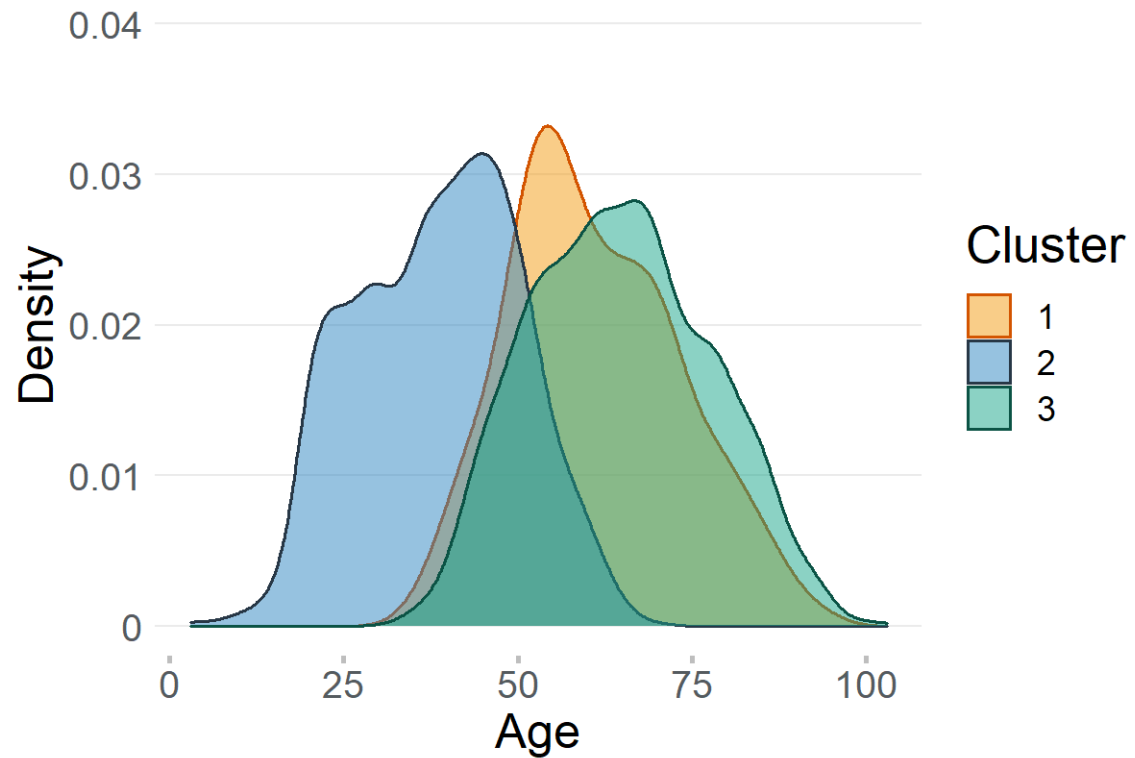


Investment

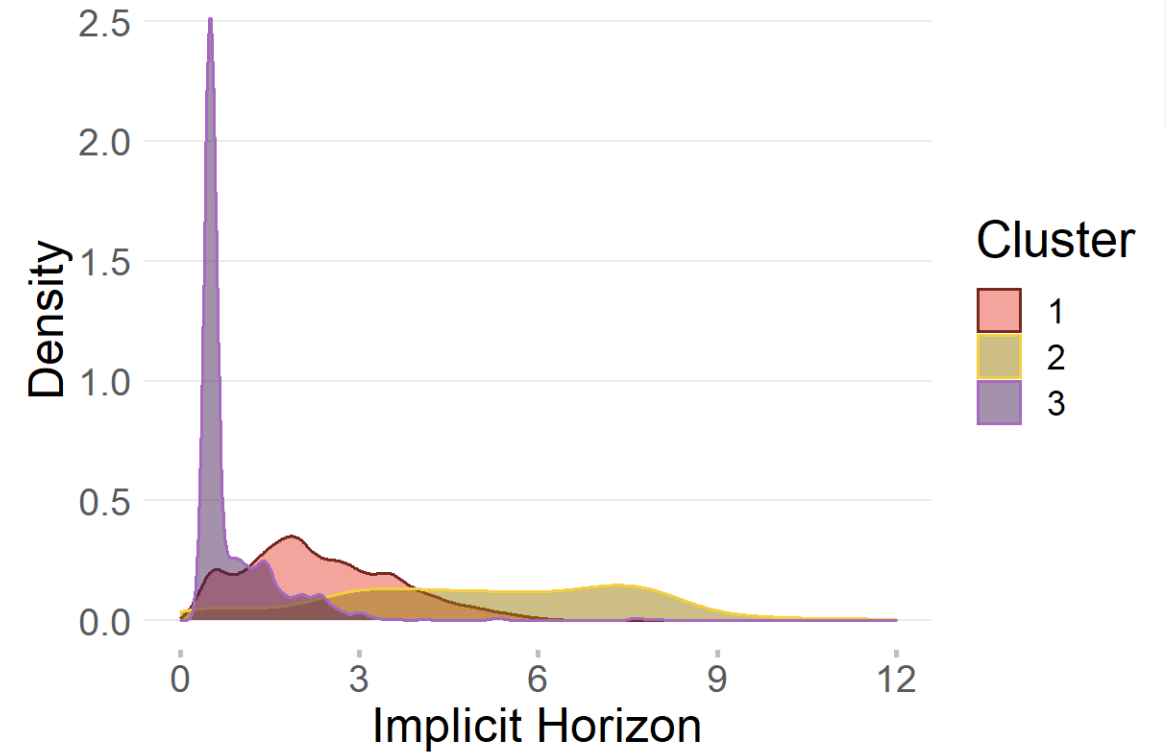


CONFRONTO

Profile - Age

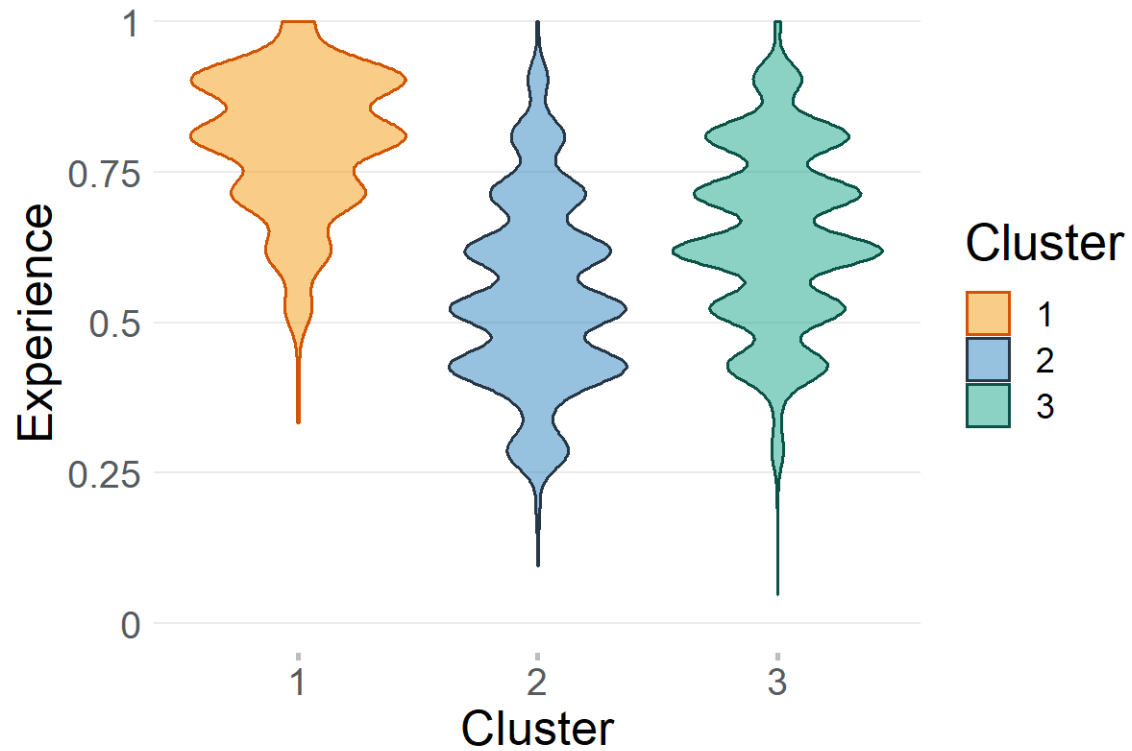


Investment - Implicit Horizon

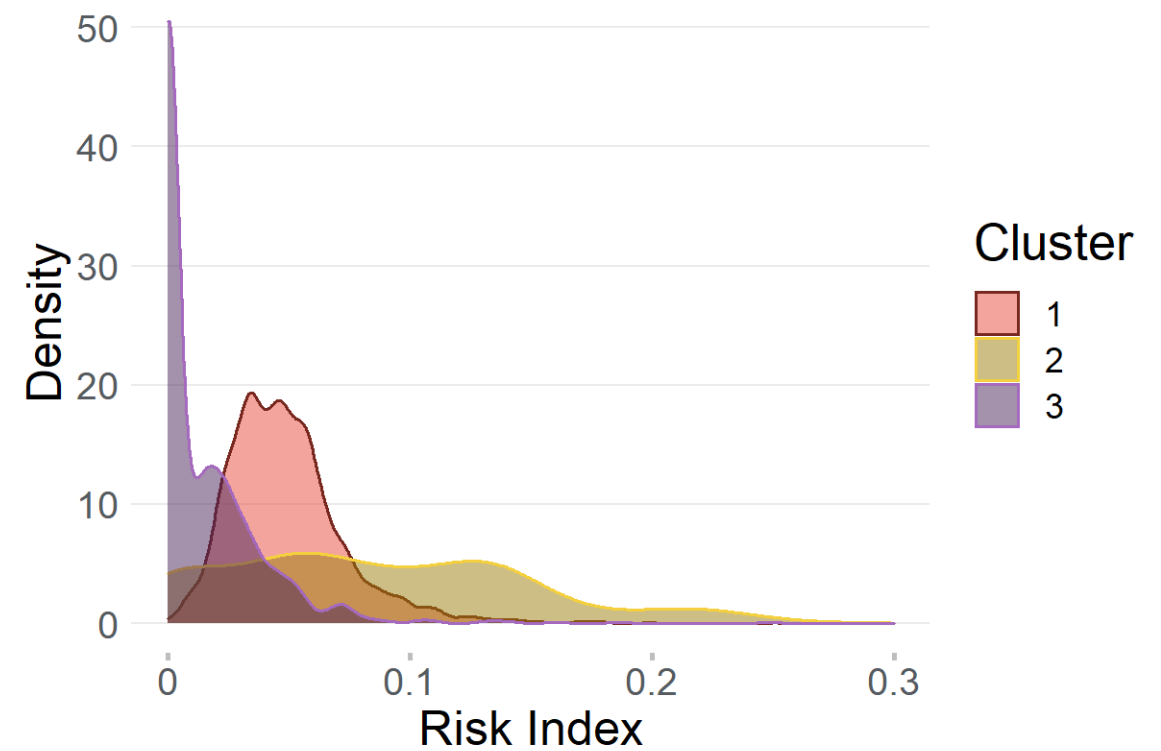


CONFRONTO

Profile - Experience



Investment - Risk Index





WHAT'S
NEXT?

SVILUPPI FUTURI

- PCA: complicano le interpretazioni, ma migliorano le performance
- Analisi Esplorativa Outliers
- Strumento di Supervisione per l'Intervento sui Clienti



**GRAZIE PER
L'ATTENZIONE**
