

# Analisi delle vendite dell'alluminio: il caso dell'azienda All.Co Group

*Strumenti Statistici per l'Analisi di Dati Aziendali*

Fornasier Luca, Parvu Marius Viorel, Piergallini Federica

Università degli Studi di Padova

A.A 2023/2024



- 1 Introduzione
- 2 Contesto
- 3 Prima domanda di ricerca
- 4 Seconda domanda di ricerca
- 5 Terza domanda di ricerca

- 1 Introduzione
  - Presentazione dell'azienda
  - Andamento della serie dei prezzi dell'alluminio
- 2 Contesto
  - Descrizione del dataset
  - Analisi esplorative e Preprocessing
  - Domande di ricerca
- 3 Prima domanda di ricerca
  - Modelli
  - Modellazione in due stadi
  - Modellazione in classi
- 4 Seconda domanda di ricerca
  - Segmentazione della clientela
- 5 Terza domanda di ricerca
  - Regole associative

# Presentazione dell'azienda

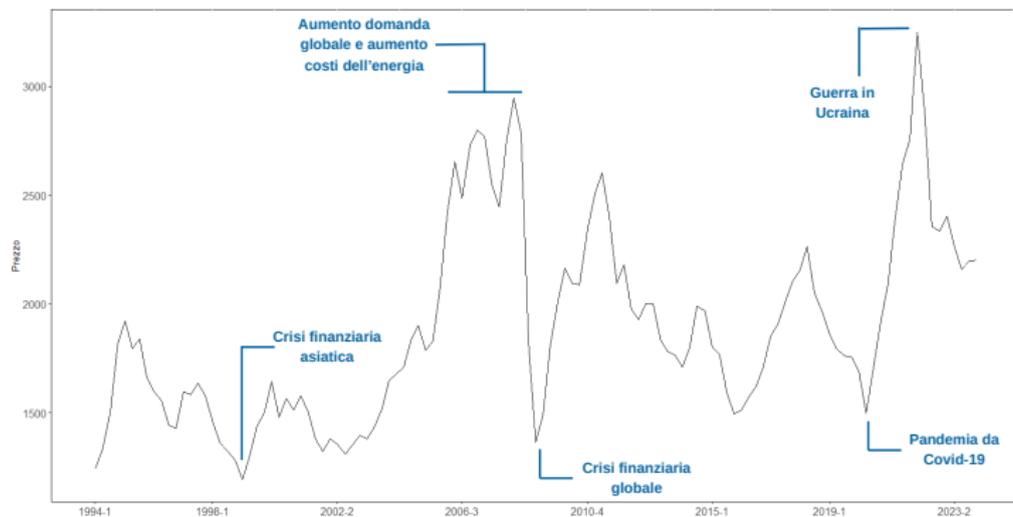
All.Co Group è un'azienda italiana leader da 50 anni nella progettazione ed estrusione di profili in alluminio per l'industria e l'architettura.

Con varie sedi in Italia, opera attivamente sia nel nostro Paese che all'Estero.



# Andamento della serie dei prezzi dell'alluminio

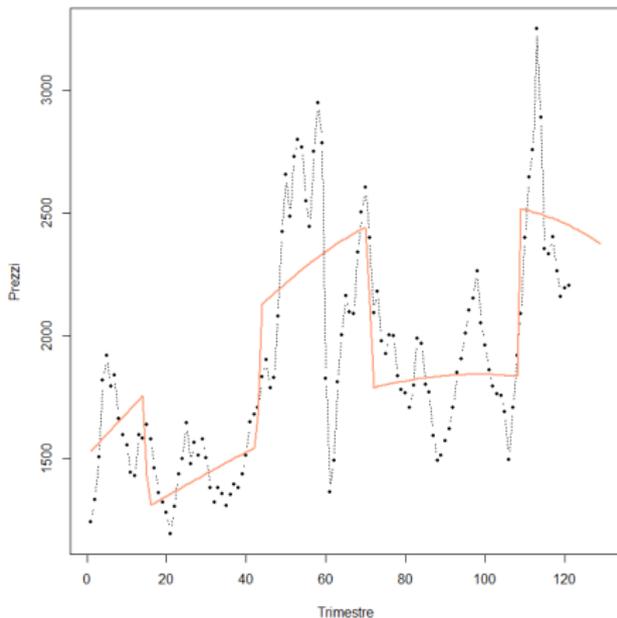
- Fonte: Federal Reserve Bank of St.Louis; <sup>1</sup>
- Periodo di riferimento: dal 1994 al 2023;
- Frequenza rilevazione: trimestrale;
- Applicazione di modelli di Bass generalizzati e modello Prophet.



<sup>1</sup><https://fred.stlouisfed.org/series/PALUMUSDQ#0>

# Andamento della serie dei prezzi dell'alluminio

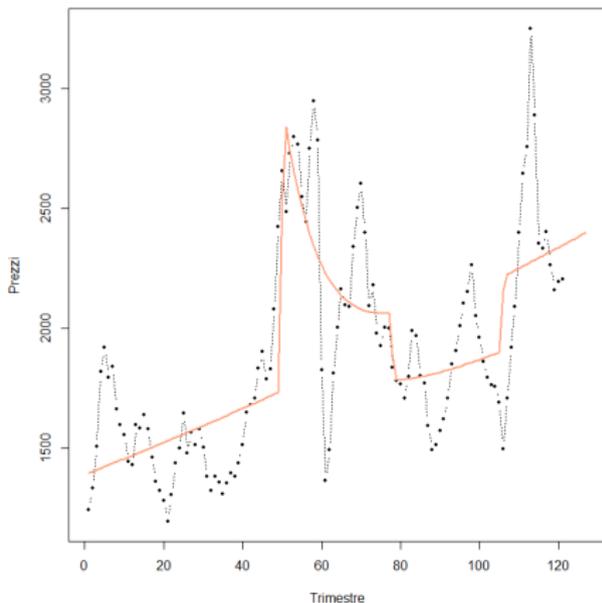
## Modelli di Bass generalizzato - 2 shock rettangolari



	<b>Estimate</b>	<b>Std.Error</b>	<b>p-value</b>
m	459407.71	22497.40	<0.0001
p	0.00331	0.00015	<0.0001
q	0.01465	0.00069	<0.0001
a1	14.27571	0.96150	<0.0001
b1	42.78742	0.73666	<0.0001
c1	-0.26573	0.01095	<0.0001
a2	70.51718	0.60812	<0.0001
b2	108.00000	0.73149	<0.0001
c2	-0.27119	0.00880	<0.0001

# Andamento della serie dei prezzi dell'alluminio

## Modelli di Bass generalizzato - shock misto

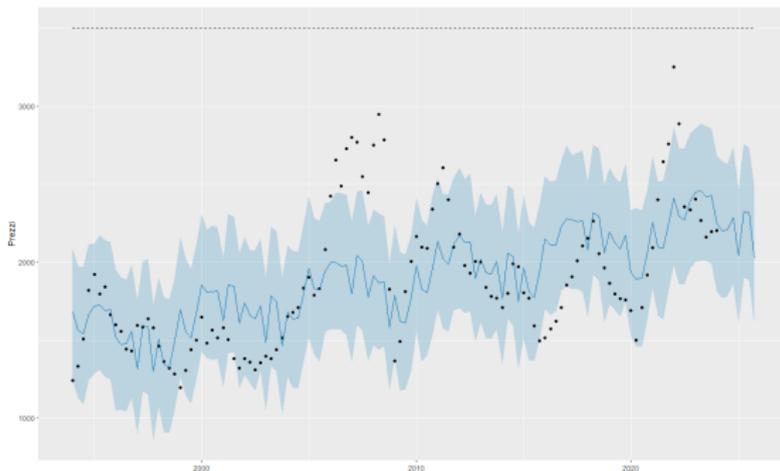


	Estimate	Std.Error	p-value
m	2769632.96	5813779.67	0.635
p	0.001	0.001	0.632
q	0.005	0.002	<b>0.013</b>
a1	49.320	0.819	<b>&lt;0.0001</b>
b1	-0.113	0.028	<b>&lt;0.0001</b>
c1	0.706	0.119	<b>&lt;0.0001</b>
a2	77.380	2.269	<b>&lt;0.0001</b>
b2	105.195	2.265	<b>&lt;0.0001</b>
c2	-0.141	0.019	<b>&lt;0.0001</b>

# Andamento della serie dei prezzi dell'alluminio

## Modello Prophet

Il trend crescente dei prezzi viene confermato anche dalla stima del modello Prophet.



**Figura:** Modello con trend non lineare logistico, stagionalità additiva sia annuale che settimanale.

- 1 Introduzione
  - Presentazione dell'azienda
  - Andamento della serie dei prezzi dell'alluminio
- 2 Contesto
  - Descrizione del dataset
  - Analisi esplorative e Preprocessing
  - Domande di ricerca
- 3 Prima domanda di ricerca
  - Modelli
  - Modellazione in due stadi
  - Modellazione in classi
- 4 Seconda domanda di ricerca
  - Segmentazione della clientela
- 5 Terza domanda di ricerca
  - Regole associative

# Descrizione del dataset

Il dataset originale era composto da 1666 osservazioni e 8 variabili:

- *nazione*: nazione di provenienza del cliente;
- *regione*: regione di provenienza del cliente (se italiano);
- *categ. cliente*: tipologia di cliente;
- *totale di quantità*: kg venduti di alluminio per cliente in un anno;
- *barrame*: kg venduti di barrame per cliente in un anno;
- *industriale*: kg venduti di industriale per cliente in un anno;
- *serramento*: kg venduti di serramento per cliente in un anno;
- *anno*: anno di riferimento, dal 2016 al 2023.

Abbiamo considerato anche variabili Open che possono essere d'aiuto nella definizione delle variabili di vendita. Le variabili aggiunte sono:

- *PIL p.c.*: PIL pro capite per ogni Paese;<sup>2</sup>
- *covid*: variabile categoriale della presenza del Covid (0:pre, 1:durante, 2:post);
- *alluminio price*: prezzo medio annuale dell'alluminio.<sup>3</sup>

---

<sup>2</sup>[https://www.imf.org/external/datamapper/NGDPDPC@WEO/OEMDC/ADVEC/WEO\\_WORLD](https://www.imf.org/external/datamapper/NGDPDPC@WEO/OEMDC/ADVEC/WEO_WORLD)

<sup>3</sup><https://fred.stlouisfed.org/series/PALUMUSDQ#0>

# Analisi esplorative e Preprocessing

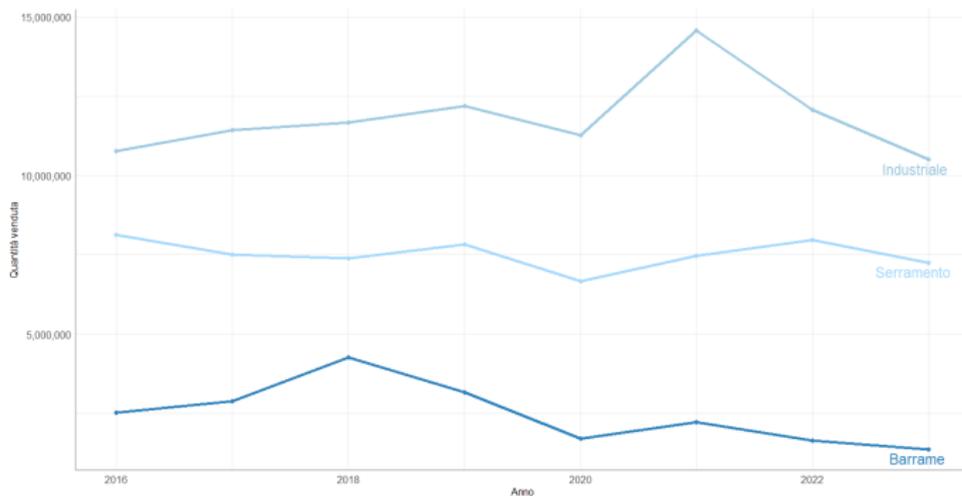


Figura: Andamento annuale della quantità venduta dei 3 prodotti.

# Analisi esplorative e Preprocessing



Figura: Quantità venduta per ripartizione geografica per categoria di prodotto.

# Analisi esplorative e Preprocessing

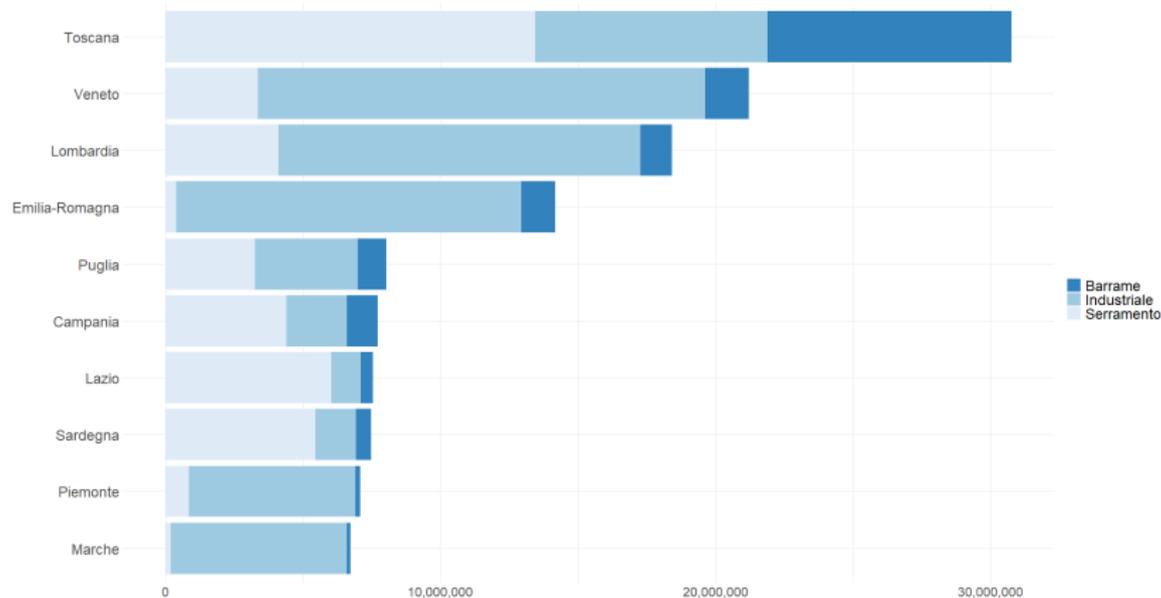


Figura: Quantità vendute negli ultimi 8 anni per le top 10 regioni italiane.

# Analisi esplorative e Preprocessing

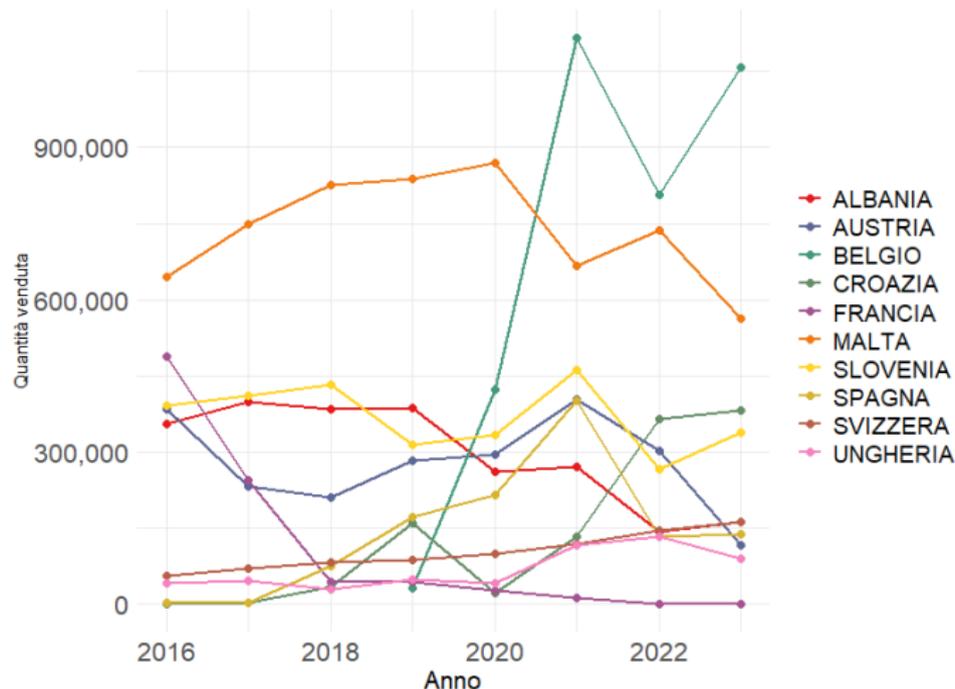


Figura: Andamento annuale delle vendite totali per le top 10 nazioni acquirenti.

# Analisi esplorative e Preprocessing

Le variabili di interesse, quando non utilizzate come variabili di risposta, vengono impiegate come variabili esplicative, categorizzate in classi.

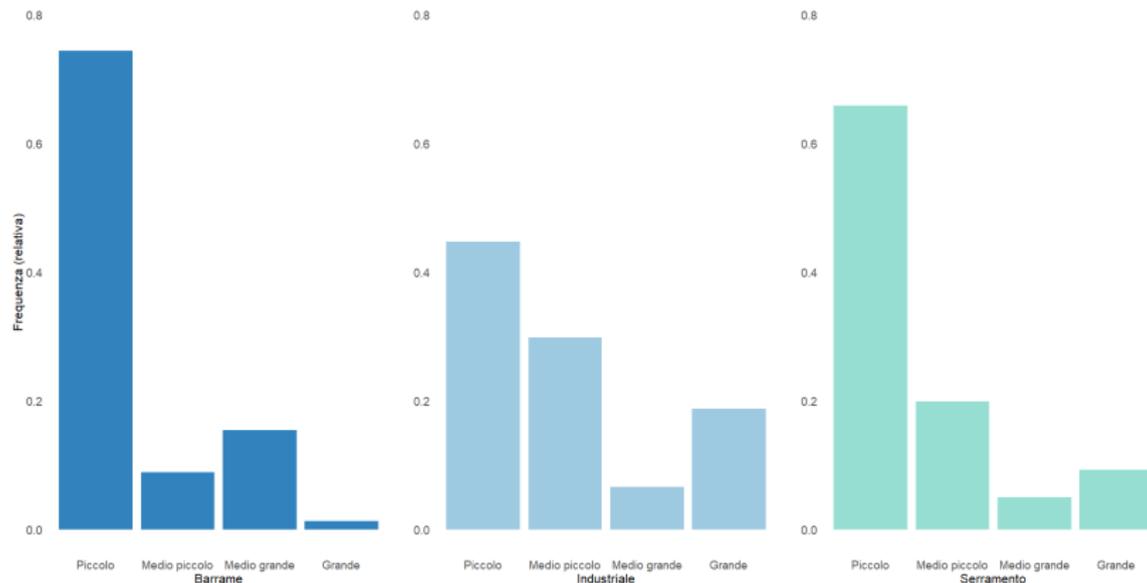
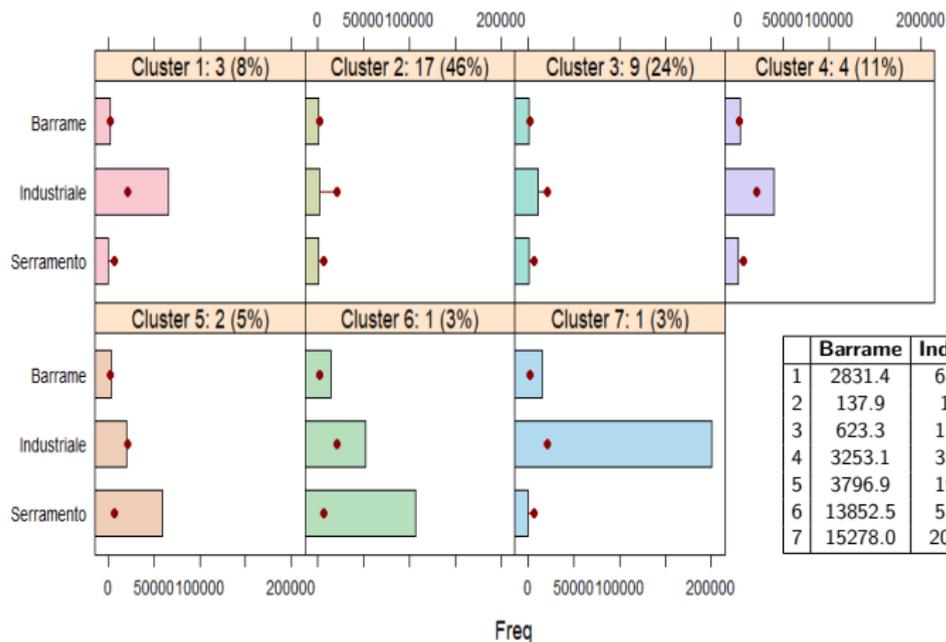


Figura: Frequenza relativa in classi dei tre prodotti.

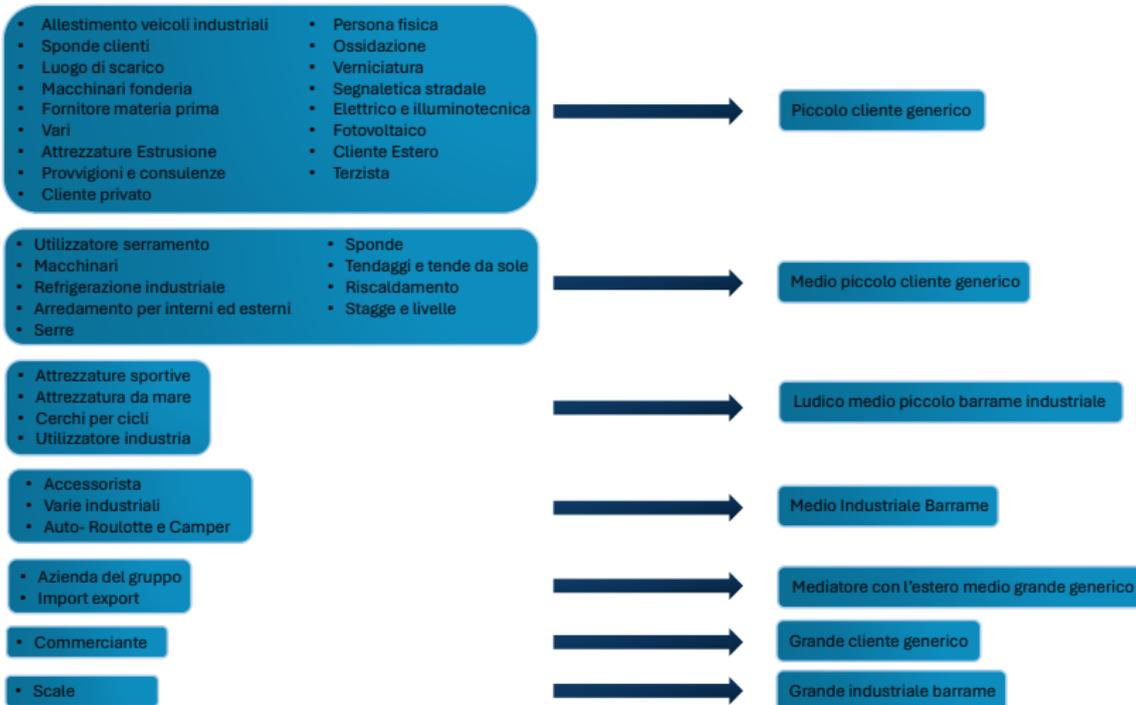
# Analisi esplorative e Preprocessing

Riduzione delle modalità della variabile categoria cliente: da 37 a 7.  
Metodo di Ward con distanza euclidea.



	Barrame	Industriale	Serramento
1	2831.4	65655.9	0.0
2	137.9	1240.8	85.7
3	623.3	11609.6	1128.6
4	3253.1	39796.6	0.0
5	3796.9	19619.1	59420.8
6	13852.5	51797.7	106037.8
7	15278.0	200973.0	0.0

# Analisi esplorative e Preprocessing



Aree di interesse dell'azienda:

- 1 Previsione delle vendite dei prodotti;
- 2 Segmentazione della clientela;
- 3 Associazioni tra tipologie di prodotto vendute.

- 1 Introduzione
  - Presentazione dell'azienda
  - Andamento della serie dei prezzi dell'alluminio
- 2 Contesto
  - Descrizione del dataset
  - Analisi esplorative e Preprocessing
  - Domande di ricerca
- 3 **Prima domanda di ricerca**
  - **Modelli**
  - **Modellazione in due stadi**
  - **Modellazione in classi**
- 4 Seconda domanda di ricerca
  - Segmentazione della clientela
- 5 Terza domanda di ricerca
  - Regole associative

## **Previsione delle vendite dei prodotti**

La quantità venduta dei prodotti ha un ruolo fondamentale per l'azienda. La previsione di questa variabile è importante per i seguenti motivi:

- pianificazione e gestione finanziaria;
- pianificazione della produzione e gestione dell'inventario (approvvigionamento, magazzino, etc.).

Punto di partenza:

Modello Lineare

Modello Lineare Stepwise

Regressione Ridge

Regressione Lasso

GAM

Random Forest

Gradient Boosting

Hurdle Negative Binomial Model

Modello Poisson

Albero di Regressione

Zero-Inflated Poisson Model

Hurdle Poisson Model

Zero-Inflated Negative Binomial Model

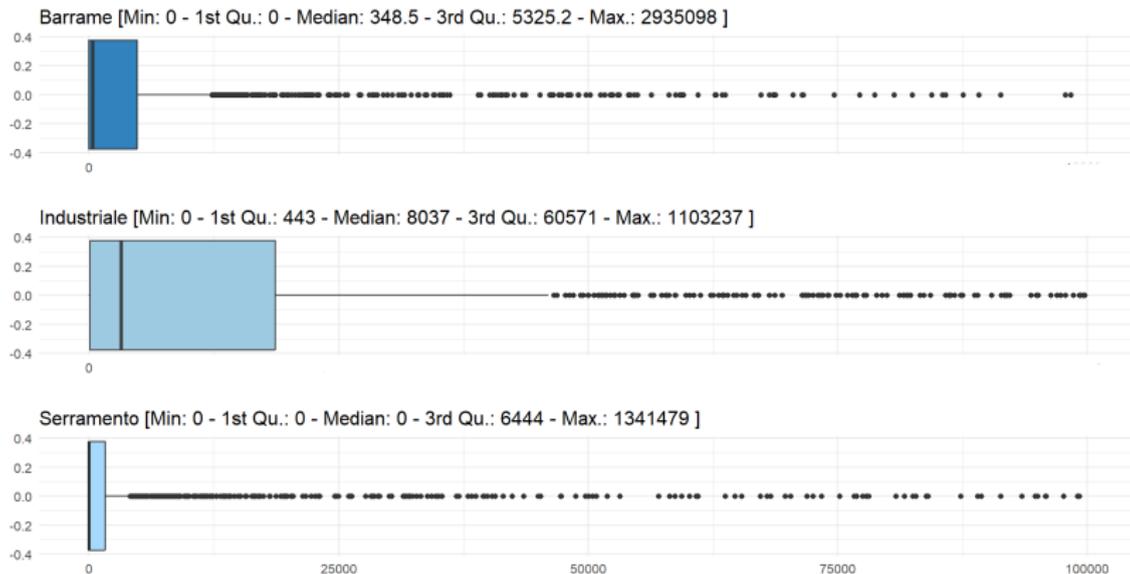
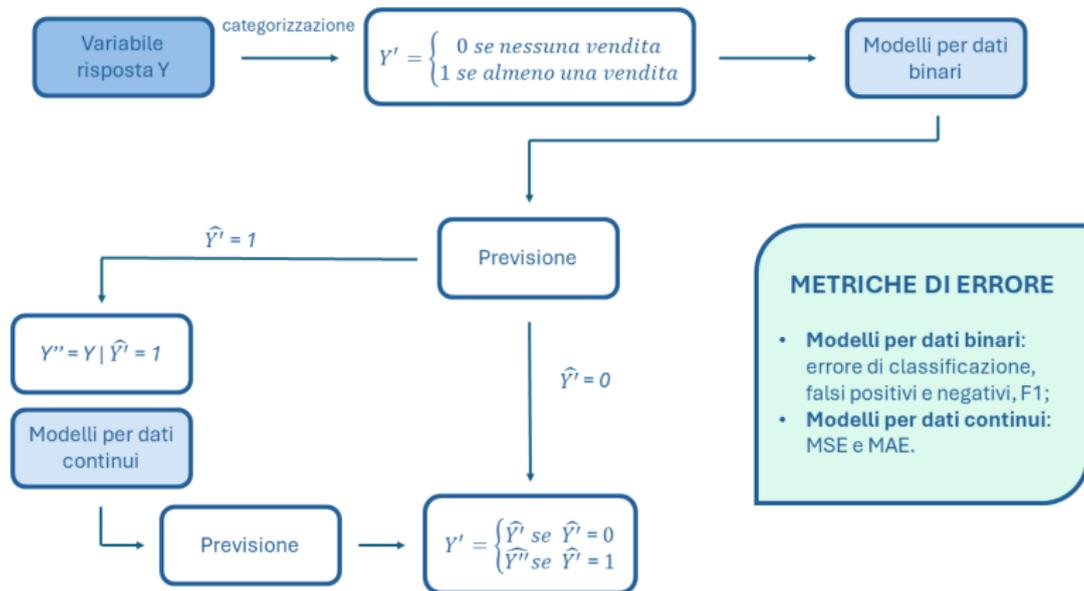


Figura: Boxplot della distribuzione delle tre variabili risposta.

# Modellazione in due stadi



# Modellazione in due stadi: Barrame

<b>Modello</b>	<b>CE</b>	<b>FP</b>	<b>FN</b>	<b>F1</b>
Lineare	0.327	1	0	0.804
Logistico	0.291	0.611	0.135	0.8
Albero	0.314	0.486	0.230	0.768
LDA	0.263	0.472	0.162	0.81
Boosting	0.282	0.389	0.230	0.786

Tabella: Metriche di errori di modelli per dati binari.

<b>Modello</b>	<b>MSE</b>	<b>MAE</b>
Lineare	2.940	1.320
Lineare Stepwise	2.942	1.323
Lasso	2.963	1.334
Tree	3.392	1.388
Random Forest	2.589	1.242
Gradient boosting stump	2.891	1.311
Gradient boosting depth4	2.562	1.230

Tabella: Metriche di errori di modelli per dati continui.

# Modellazione in due stadi: Barrame

## Modellazione primo stadio

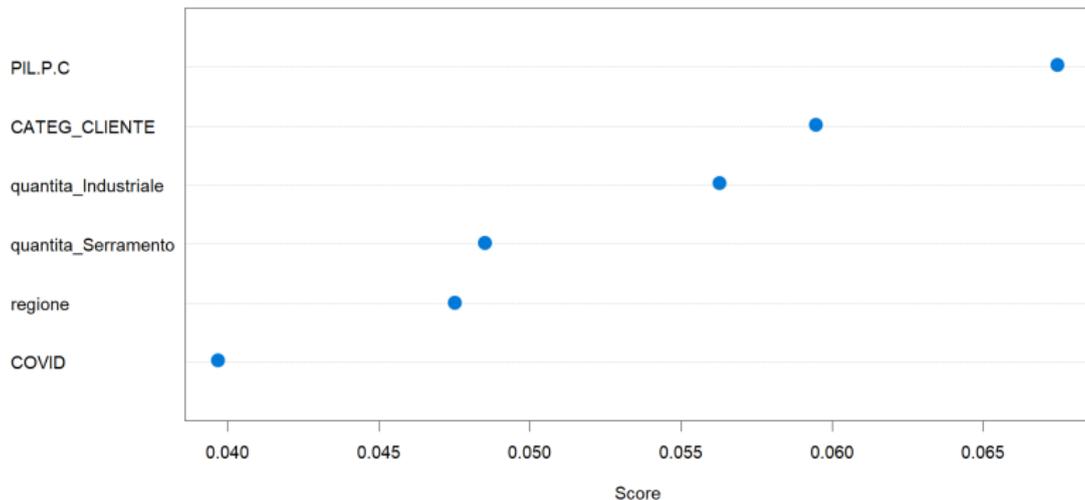


Figura: Grafico di importanza delle variabili nel Boosting.

# Modellazione in due stadi: Barrame

## Modellazione secondo stadio

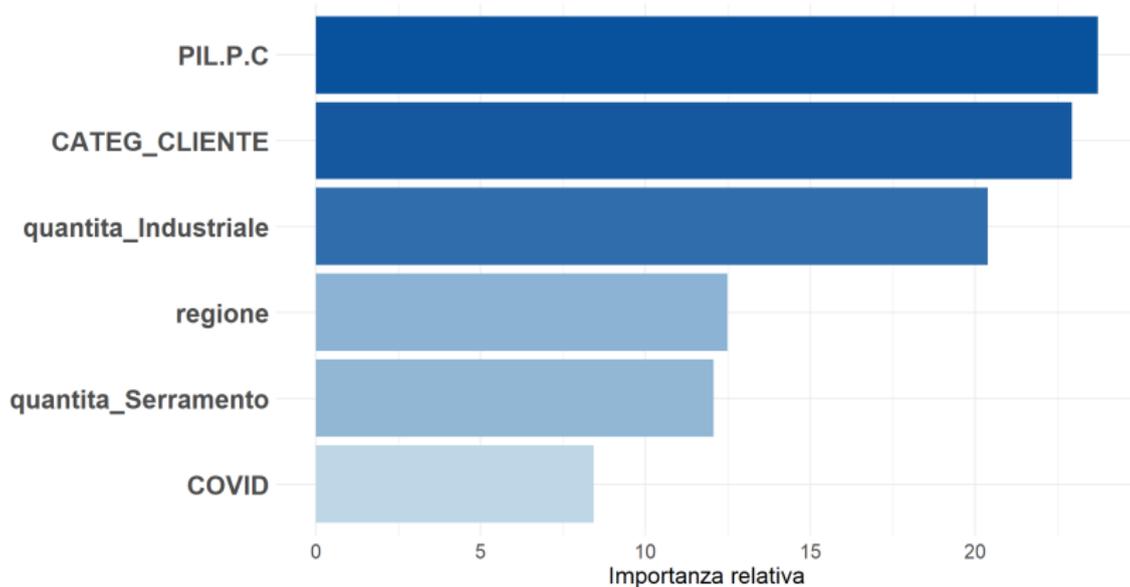


Figura: Grafico di importanza delle variabili nel GBM.

# Modellazione in due stadi: Serramento

<b>Modello</b>	<b>CE</b>	<b>FP</b>	<b>FN</b>	<b>F1</b>
Lineare	0.573	1.000	0.000	0.599
Logistico	0.309	0.095	0.596	0.528
Albero	0.295	0.254	0.351	0.652
LDA	0.318	0.135	0.564	0.539
Boosting	0.323	0.040	0.702	0.441

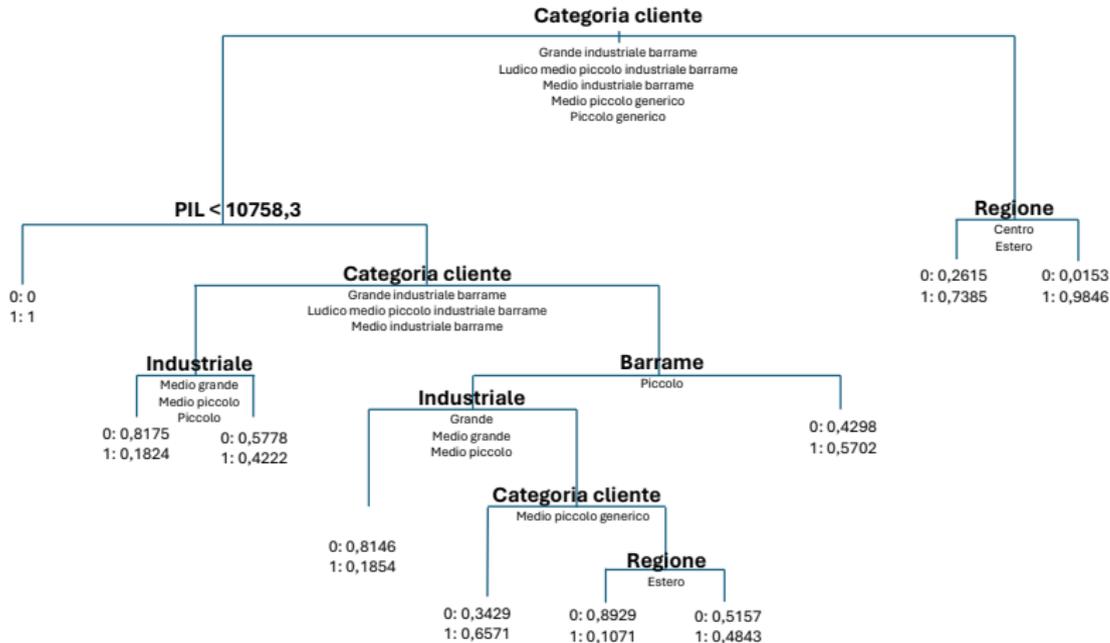
Tabella: Metriche di errori di modelli per dati binari.

<b>Modello</b>	<b>MSE</b>	<b>MAE</b>
Lineare	4.717	1.701
Lineare Stepwise	5.108	1.759
Lasso	5.243	1.673
Random Forest	4.162	1.509
Gradient boosting stump	4.656	1.667
Gradient boosting depth4	3.896	1.388
Gradient boosting depth6	3.786	1.391

Tabella: Metriche di errori di modelli per dati continui.

# Modellazione in due stadi: Serramento

## Modellazione primo stadio



# Modellazione in due stadi: Serramento

*Modellazione secondo stadio*

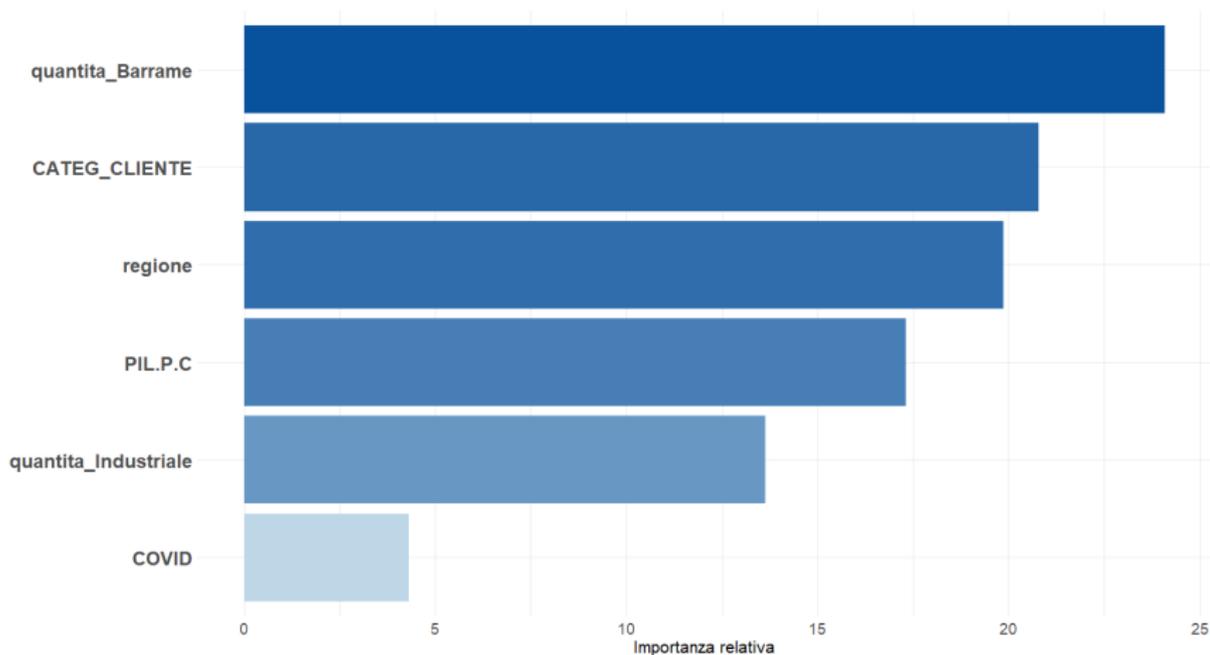


Figura: Grafico di importanza delle variabili nel GBM.

Dato che la distribuzione della variabile industriale presenta pochi zeri, la modellazione in due stadi non risulta adeguata.

Si è proceduto, quindi, a categorizzare la risposta in classi come segue:

- piccolo:  $[0,5000)$ ;
- mediopiccolo:  $[5000,56723)$ ;
- grande:  $[56723,1103237]$ .

# Modellazione in classi: Industriale

Si considerano come metriche di errori per i modelli:

- tasso di errata classificazione;
- percentuale dei falsi positivi come le osservazioni erratamente classificate in "grande";
- percentuale dei falsi negativi come le osservazioni erratamente classificate in "mediopiccolo" e "piccolo".

<b>Modello</b>	<b>CE</b>	<b>FP</b>	<b>FN</b>
LogitCumulati	0.413	0.193	0.407
MultiLineare	0.354	0.15	0.370
Multinomiale	0.391	0.192	0.444
Albero	0.386	0.128	0.518
LDA	0.35	0.139	0.407
MARS	0.354	0.113	0.426
Gradient Boosting	0.354	0.101	0.407

**Tabella:** Metriche di errori per i modelli.

# Modellazione in classi: Industriale

Coefficienti	Piccolo	MedioPiccolo	Grande
regioneVeneto	<b>-0.239</b>	0.094	<b>1.145</b>
regioneAbruzzo	<b>0.328</b>	-0.082	<b>0.754</b>
regioneBasilicata	<b>0.373</b>	0.08	<b>0.547</b>
regioneCalabria	<b>0.347</b>	-0.003	<b>0.656</b>
regioneCampania	<b>0.348</b>	-0.056	<b>0.709</b>
regioneEmilia-Romagna	-0.118	0.078	<b>1.04</b>
regioneEstero	0.14	0.039	<b>0.821</b>
regioneFriuli	<b>0.248</b>	0.022	<b>0.73</b>
regioneLazio	<b>0.237</b>	0.121	<b>0.642</b>
regioneLiguria	<b>0.506</b>	-0.196	<b>0.69</b>
regioneLombardia	-0.089	0.045	<b>1.044</b>
regioneMarche	-0.127	-0.004	<b>1.131</b>
regioneMolise	<b>0.633</b>	<b>-0.37</b>	<b>0.738</b>
regionePiemonte	-0.104	-0.028	<b>1.132</b>
regionePuglia	0.122	-0.031	<b>0.909</b>
regioneSardegna	<b>0.403</b>	-0.138	<b>0.736</b>
regioneSicilia	<b>0.324</b>	0.025	<b>0.651</b>
regioneToscana	0.154	-0.034	<b>0.881</b>
regioneTrentino	<b>0.364</b>	-0.119	<b>0.755</b>
regioneUmbria	<b>0.259</b>	-0.221	<b>0.962</b>
clustergrande cliente industriale barrame	-0.031	-0.221	<b>0.252</b>
clusterludico medio piccolo barrame industriale	-0.06	-0.101	<b>0.161</b>
clustermediatore con estero medio grande generico	-0.095	<b>0.315</b>	<b>-0.22</b>
clustermedio industriale barrame	-0.072	-0.086	<b>0.158</b>
clustermedio piccolo cliente generico	-0.015	0.091	<b>-0.076</b>
clusterpiccolo cliente generico	<b>0.205</b>	<b>-0.12</b>	<b>-0.085</b>
alluminio price	0.000	0.000	0.000
quantita Barramemedio grande	-0.092	<b>0.299</b>	<b>-0.207</b>
quantita Barramemedio piccolo	-0.148	<b>0.507</b>	<b>-0.359</b>
quantita Barramepiccolo	0.118	<b>0.384</b>	<b>-0.502</b>
quantita Serramentomedio grande	<b>0.144</b>	0.052	<b>-0.196</b>
quantita Serramentomedio piccolo	<b>0.441</b>	<b>-0.216</b>	<b>-0.225</b>
quantita Serramentopiccolo	<b>0.375</b>	-0.08	<b>-0.295</b>

Tabella: Coefficienti di regressione per il modello lineare multivariato.  
 In **grassetto** sono presentati i coefficienti statisticamente significativi al 5%.

# Sommario

- 1 Introduzione
  - Presentazione dell'azienda
  - Andamento della serie dei prezzi dell'alluminio
- 2 Contesto
  - Descrizione del dataset
  - Analisi esplorative e Preprocessing
  - Domande di ricerca
- 3 Prima domanda di ricerca
  - Modelli
  - Modellazione in due stadi
  - Modellazione in classi
- 4 Seconda domanda di ricerca
  - Segmentazione della clientela
- 5 Terza domanda di ricerca
  - Regole associative

## Segmentazione della clientela

Obiettivi:

- sviluppo di strategie mirate per ciascun segmento della clientela;
- caratterizzazione e descrizione della clientela.

# Segmentazione della clientela

Metodo	Clustering	N° cluster	Metriche
	Legame completo		
	Legame singolo	3	
	Legame medio	4	Distanza euclidea
Gerarchico	Legame mediano	5	Distanza Manhattan
	Metodo Ward	6	Indice Gower
	Centroide		
Partizionale	K-means	da 2 a 12	Distanza euclidea

**Tabella:** Riassunto delle operazioni svolte nell'ambito dell'analisi dei gruppi

# Segmentazione della clientela

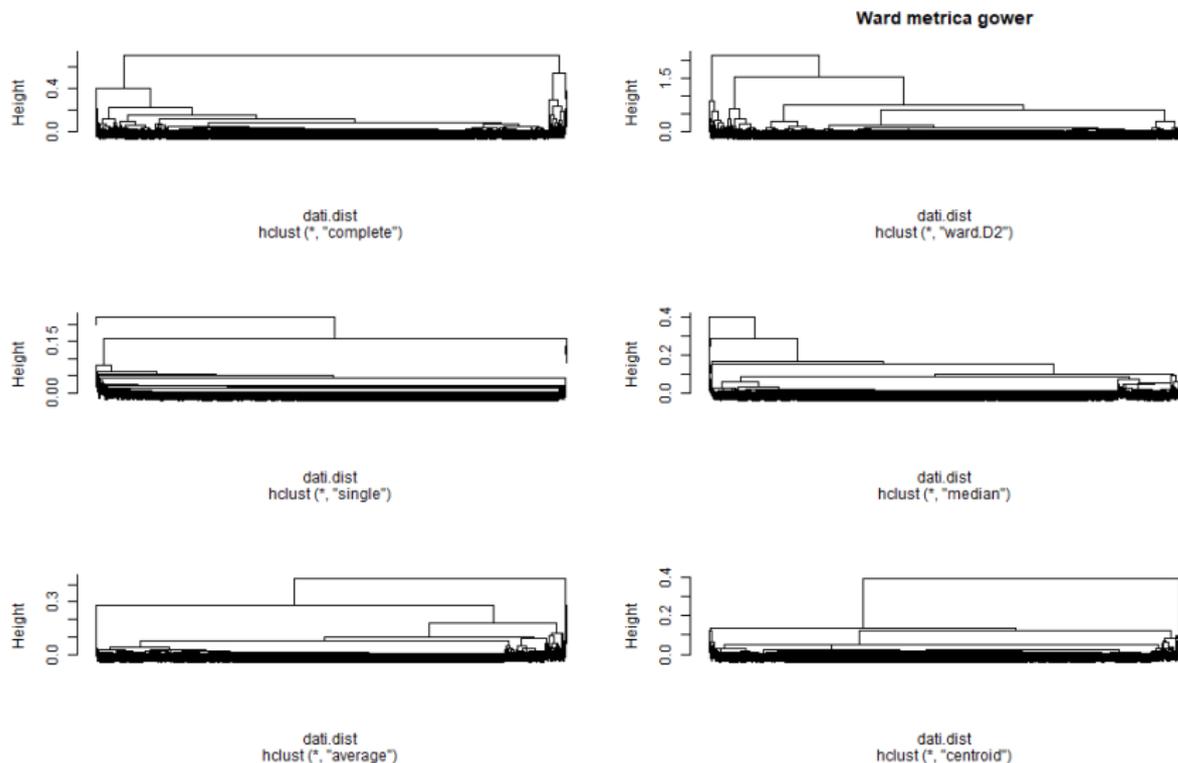
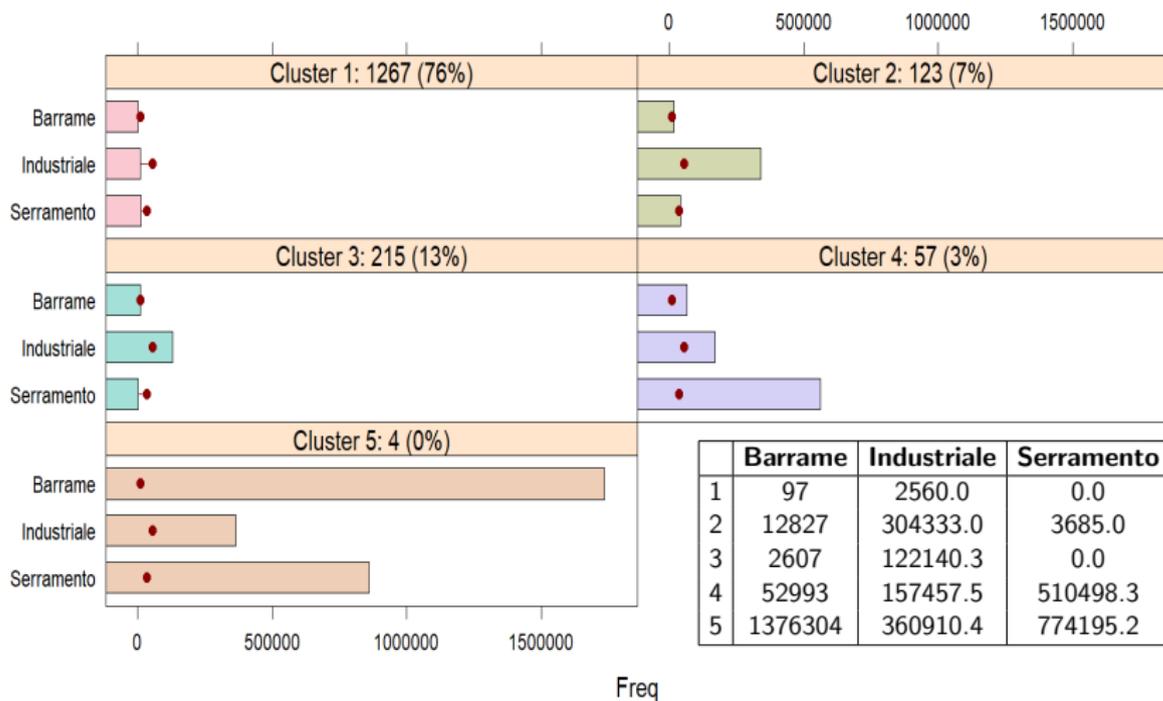


Figura: Dendrogrammi con metrica Gower.

# Segmentazione della clientela



# Segmentazione della clientela

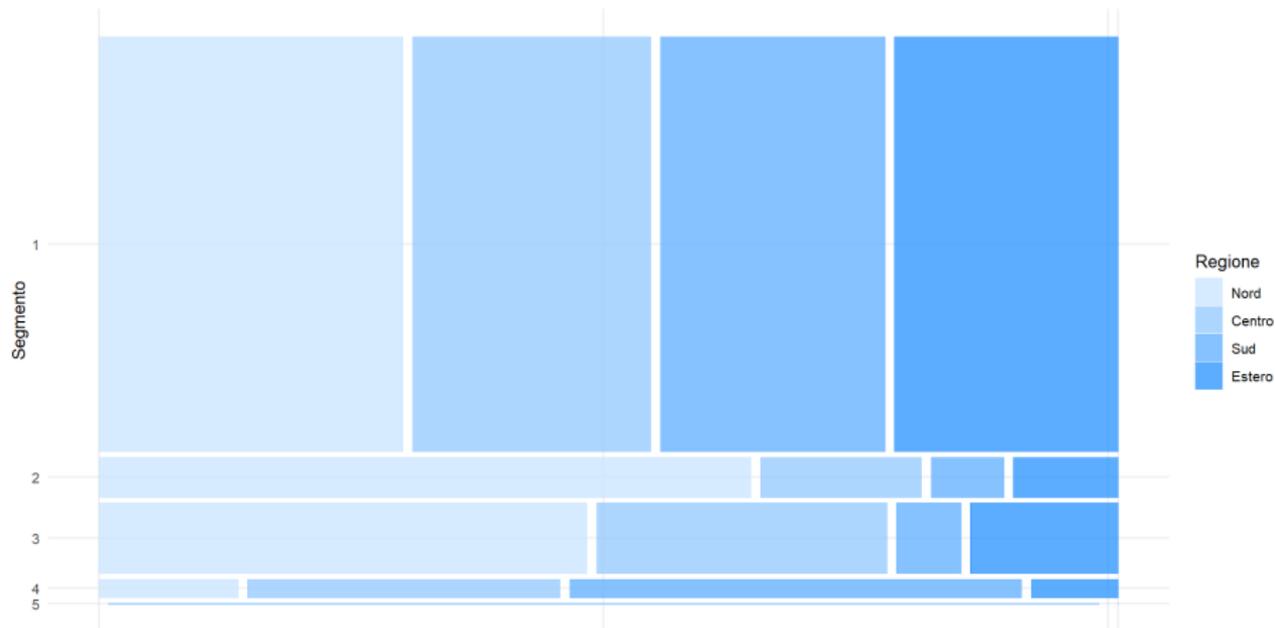


Figura: Suddivisione della clientela sulla base della regione di provenienza.

# Segmentazione della clientela

- Cluster 1 → **Club**: clienti che acquistano poco di tutto, distribuiti in modo uniforme in Italia ed estero;
- Cluster 3 → **Bronze**: clienti nel centro-nord Italia, acquistano principalmente beni industriali;
- Cluster 2 → **Silver**: acquirenti italiani del nord e paesi europei, concentrati su beni industriali;
- Cluster 4 → **Gold**: commercianti o utilizzatori di serramenti, principalmente italiani con alcuni maltesi, acquistano molto di tutto;
- Cluster 5 → **Platinum**: grandi clienti generici toscani, acquistano molto di tutto, ma non hanno acquistato durante e dopo il covid.

- 1 Introduzione
  - Presentazione dell'azienda
  - Andamento della serie dei prezzi dell'alluminio
- 2 Contesto
  - Descrizione del dataset
  - Analisi esplorative e Preprocessing
  - Domande di ricerca
- 3 Prima domanda di ricerca
  - Modelli
  - Modellazione in due stadi
  - Modellazione in classi
- 4 Seconda domanda di ricerca
  - Segmentazione della clientela
- 5 Terza domanda di ricerca
  - Regole associative

# Terza domanda di ricerca

## *Associazioni tra tipologie di prodotto vendute*

### Obiettivi:

- ottimizzazione delle strategie di marketing;
- identificazione di nuove opportunità di business;
- analisi dei comportamenti di acquisto.

# Regole associative

Antecedente	Consequente	Supporto	Fiducia	Copertura	Lift	Frequenza
Barrame medio piccolo, Serramento piccolo	Industriale medio grande	0.351	0.929	0.378	1.808	13
Barrame piccolo, Industriale piccolo	Serramento piccolo	0.378	0.875	0.432	1.704	14
Barrame medio piccolo, Serramento piccolo	Industriale piccolo	0.351	0.929	0.378	1.636	13
Barrame medio grande, Serramento nulla	Industriale grande	0.378	0.778	0.486	1.599	14
Barrame medio piccolo, Industriale piccolo	Serramento piccolo	0.351	0.813	0.432	1.582	13
Barrame piccolo, Industriale medio grande	Serramento piccolo	0.324	0.800	0.405	1.558	12
Barrame nulla, Serramento piccolo	Industriale piccolo	0.405	0.882	0.459	1.555	15
Barrame medio piccolo, Industriale piccolo	Serramento medio piccolo	0.324	0.750	0.432	1.542	12
Barrame medio piccolo, Serramento nulla	Industriale grande	0.459	0.739	0.622	1.519	17
Industriale grande, Serramento nulla	Barrame medio grande	0.378	0.778	0.486	1.515	14
Barrame medio grande	Industriale grande	0.378	0.737	0.514	1.515	14

Tabella: Regole associative ordinate secondo il lift decrescente.

# Regole associative

## Industriale

Antecedente	Consequente	Supporto	Fiducia	Copertura	Lift	Frequenza
Barrame medio piccolo, Serramento piccolo	Industriale medio grande	0.351	0.929	0.378	1.808	13
Barrame piccolo, Industriale piccolo	Serramento piccolo	0.378	0.875	0.432	1.704	14
Barrame medio piccolo, Serramento piccolo	Industriale piccolo	0.351	0.929	0.378	1.636	13
Barrame medio grande, Serramento nulla	Industriale grande	0.378	0.778	0.486	1.599	14
Barrame medio piccolo, Industriale piccolo	Serramento piccolo	0.351	0.813	0.432	1.582	13
Barrame piccolo, Industriale medio grande	Serramento piccolo	0.324	0.800	0.405	1.558	12
Barrame nulla, Serramento piccolo	Industriale piccolo	0.405	0.882	0.459	1.555	15
Barrame medio piccolo, Industriale piccolo	Serramento medio piccolo	0.324	0.750	0.432	1.542	12
Barrame medio piccolo, Serramento nulla	Industriale grande	0.459	0.739	0.622	1.519	17
Industriale grande, Serramento nulla	Barrame medio grande	0.378	0.778	0.486	1.515	14
Barrame medio grande	Industriale grande	0.378	0.737	0.514	1.515	14

Tabella: Regole associative ordinate secondo il lift decrescente.

# Regole associative

## Serramento

Antecedente	Conseguente	Supporto	Fiducia	Copertura	Lift	Frequenza
Barrame medio piccolo, <b>Serramento piccolo</b>	Industriale medio grande	0.351	0.929	0.378	1.808	13
Barrame piccolo, Industriale piccolo	<b>Serramento piccolo</b>	0.378	0.875	0.432	1.704	14
Barrame medio piccolo, <b>Serramento piccolo</b>	Industriale piccolo	0.351	0.929	0.378	1.636	13
Barrame medio grande, <b>Serramento nulla</b>	Industriale grande	0.378	0.778	0.486	1.599	14
Barrame medio piccolo, Industriale piccolo	<b>Serramento piccolo</b>	0.351	0.813	0.432	1.582	13
Barrame piccolo, Industriale medio grande	<b>Serramento piccolo</b>	0.324	0.800	0.405	1.558	12
Barrame nulla, <b>Serramento piccolo</b>	Industriale piccolo	0.405	0.882	0.459	1.555	15
Barrame medio piccolo, Industriale piccolo	<b>Serramento medio piccolo</b>	0.324	0.750	0.432	1.542	12
Barrame medio piccolo, <b>Serramento nulla</b>	Industriale grande	0.459	0.739	0.622	1.519	17
Industriale grande, <b>Serramento nulla</b>	Barrame medio grande	0.378	0.778	0.486	1.515	14
Barrame medio grande	Industriale grande	0.378	0.737	0.514	1.515	14

Tabella: Regole associative ordinate secondo il lift decrescente.

## Problematiche:

- prodotti raggruppati in tre macro-categorie;
- categorizzando le variabili si ottengono implicazioni insensate;
- regole non troppo significative.

Grazie per l'attenzione!

