STRUMENTI STATISTICI PER L'ANALISI DI DATI AZIENDALI

Andrea Nozza
Marta Polato
Luigi Venuto

INDICE

01 Introduzione

02 Preprocessing

03 Analisi Esplorativa

04 Modellistica

05 Conclusioni

01 - INTRODUZIONE

Chi è Otticanet?

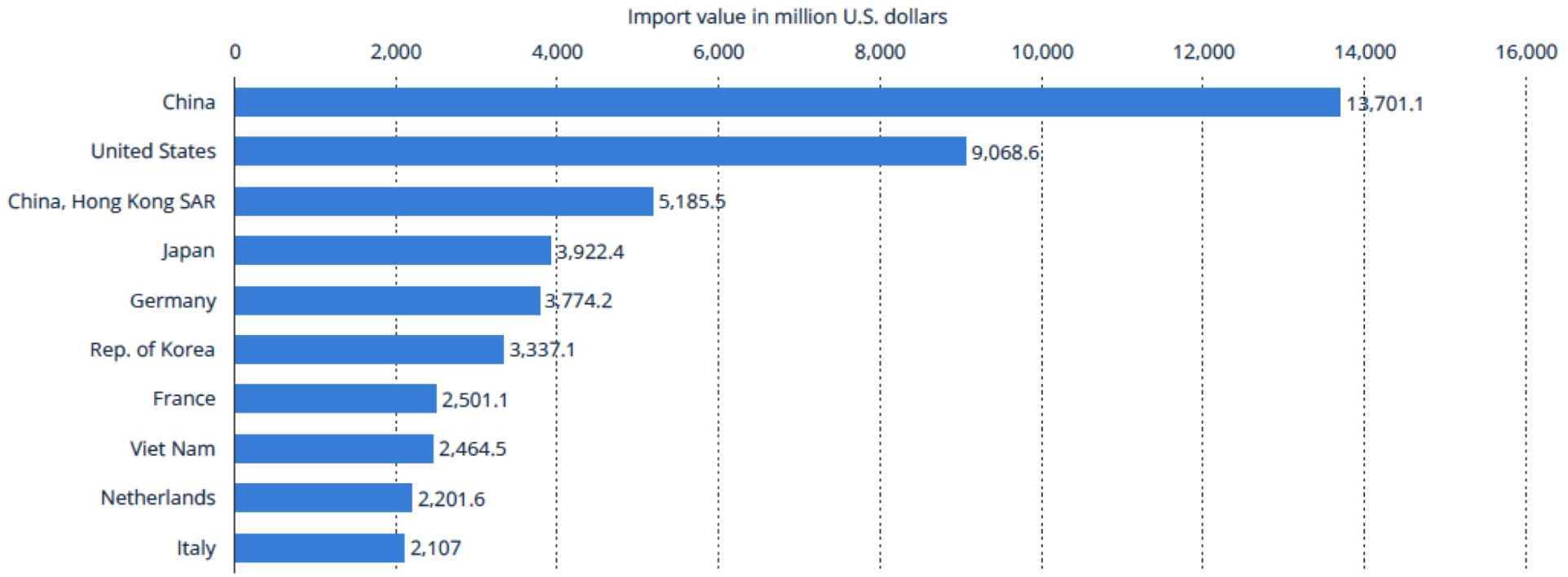
- Azienda specializzata nella vendita online di occhiali da sole, occhiali da vista, lenti a contatto e altri prodotti ottici.
- Nasce nel 2003 e dopo una lunga esperienza maturata sul campo in importanti showroom, approda nell'*e-commerce*.
- Opera principalmente attraverso il proprio sito web, offrendo una vasta selezione di marchi di occhiali di lusso e design.
- Uno dei suoi punti di forza è la possibilità di spedire i prodotti facendo affidamento a magazzini dislocati in tutto il mondo, sia di proprietà che di terzi.

Il mercato ottico

- Il mercato ottico rappresenta un settore dinamico e in continua evoluzione, dove l'eleganza e la funzionalità si fondono per offrire soluzioni ottiche di alta qualità. Gli occhiali non sono solo uno strumento per correggere la vista, ma anche un accessorio di stile che riflette la personalità e il gusto.
- I dati forniti da *statista.com* forniscono un'*overview* generale del mercato mondiale degli occhiali.

Leading importing countries of optical goods worldwide in 2021 (in million U.S. dollars)

Leading importers of optical goods worldwide 2021



8

Description: In 2021, China was the leading importer of optical goods from the rest of the world, with an import value of about 13.7 billion U.S. dollars. The United States, ranking second, imported around nine billion U.S. dollars worth of optical goods that year. Read more

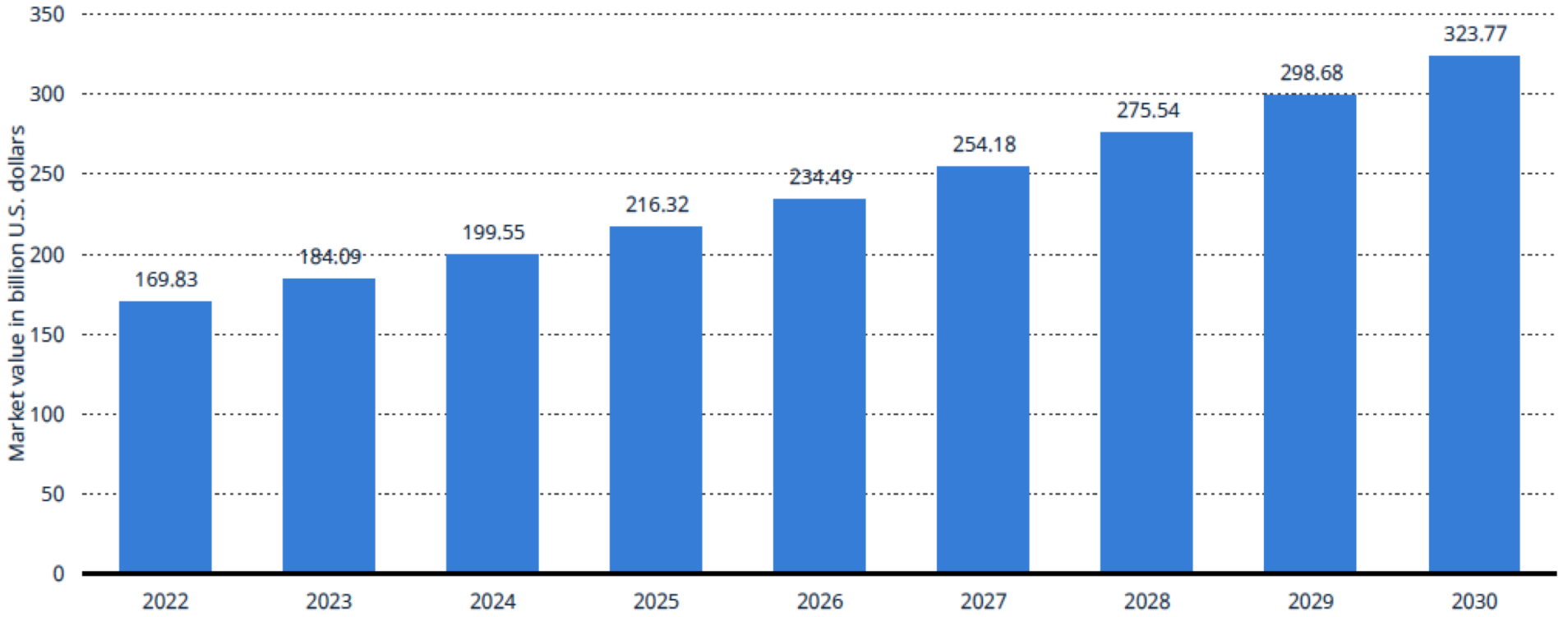
Note(s): Worldwide; 2021

Source(s): UN Comtrade

statista

Estimated value of the eyewear market worldwide from 2022 to 2030 (in billion U.S. dollars)

Value of the global eyewear market from 2022 to 2030

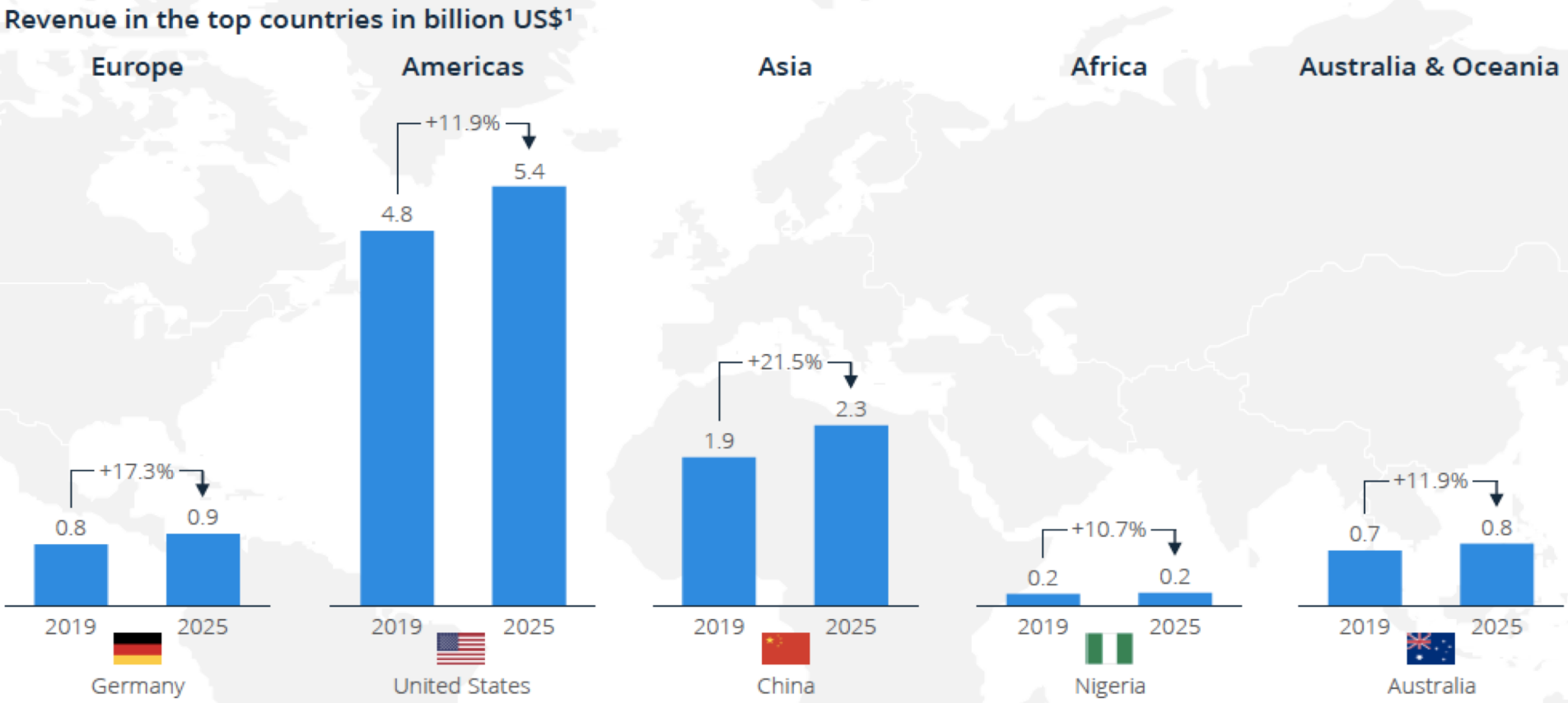


3 **Description:** The global eyewear market, which is made up of spectacles, contact lenses, sunglasses, and other eyewear products, was estimated to be worth around 170 billion U.S. dollars in 2022 and was forecast to reach a value of roughly 323.8 billion U.S. dollars by 2030. Read more.
Note(s): Worldwide; 2022
Source(s): Business Wire; Statista

statista

In China, the Sunglasses segment will grow by 21.5% from 2019 to 2025

Top countries by region



21

1: Bar heights may vary due to rounded figures
Sources: Statista Consumer Market Outlook 2020

02 - PREPROCESSING

Dataset iniziale

- Il dataset iniziale è stato fornito direttamente dall'azienda, e riguarda le vendite di prodotti tramite il loro store online relative al periodo 2019-2022.
- Inizialmente il dataset presenta circa 211.000 osservazioni, ognuna delle quali relativa ad un acquisto, e 23 variabili.
- Di seguito vengono presentate le variabili con la relativa codifica.

VARIABILI:	*CODIFICA:*
• codice_ordine	numerica
• data	data
• stato_ingresso_ordine	fattore, 3 livelli
• data_spedizione_prevista	data
• data_evasione_ordine	data
• stato_ordine	fattore, 6 livelli
• motivazione_annullamento	fattore, 9 livelli
• data_annullamento	data
• nazione	fattore, 155 livelli
• marchio	fattore, 174 livelli
• produttore	fattore, 50 livelli
• tipo_prodotto	fattore, 3 livelli
• modello	fattore, 18.646 livelli
• tipo_colore	fattore, 22.492 livelli
• descrizione_colore	fattore, 11.925 livelli
• genere_occhiale	fattore, 8 livelli
• barcode	numerica
• importo_articolo_euro_netto	numerica
• valore_promo	--
• codice_promo	fattore, 19.211 livelli
• quantita	numerica

Pulizia del dataset

Viene effettuata una pulizia preliminare del dataset per renderlo idoneo alle successive fasi dell'analisi. In particolare:

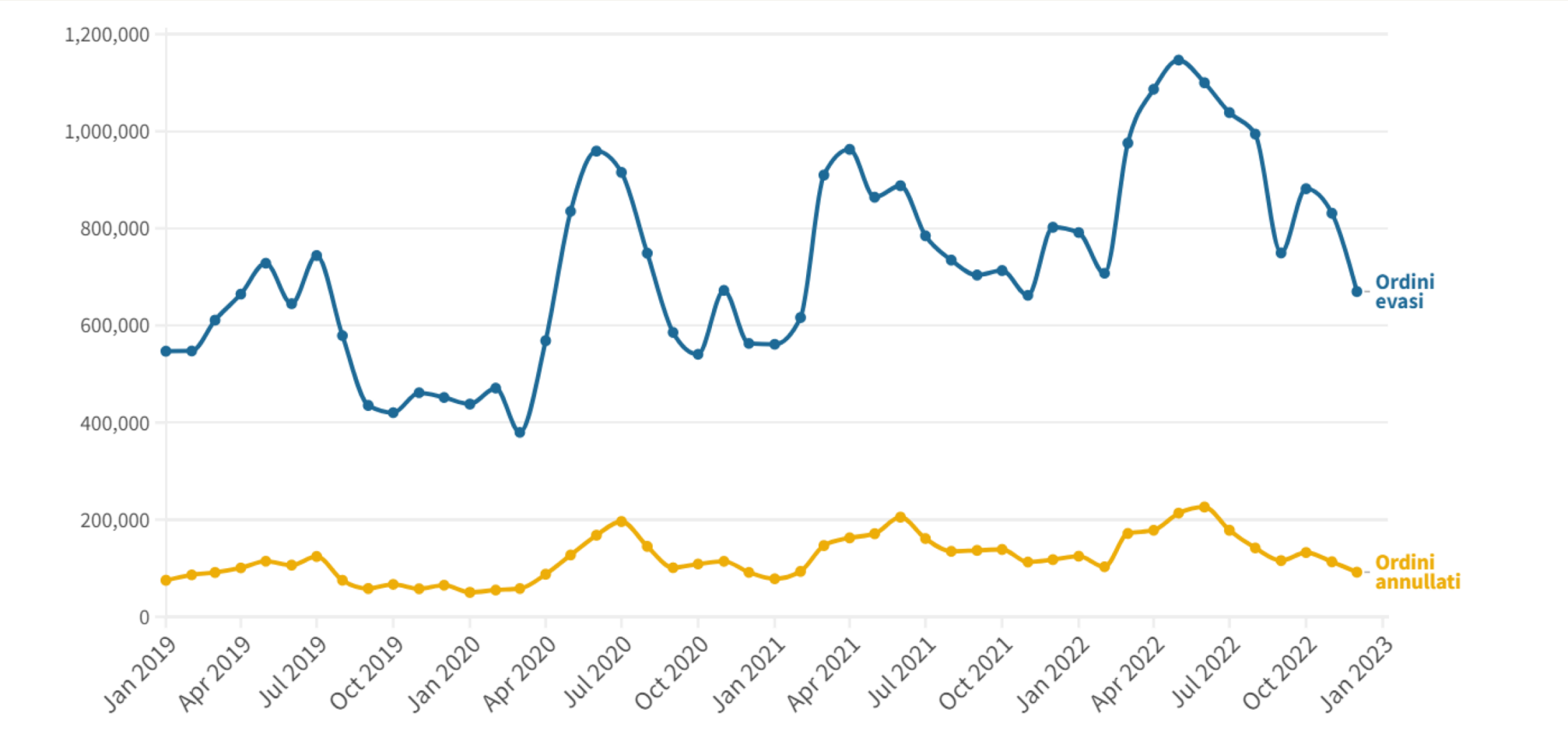
1. Eliminate le seguenti variabili:
   - `codice_ordine`, identificativa di riga;
   - `tipo_colore`, `descrizione_colore`, `modello` e `barcode`, poco informative;
   - `quantita`, degenere;
   - `valore_promo`, codificata in parte come valore percentuale e in parte come valore assoluto, e di conseguenza anche `codice_promo`.
2. Presi in considerazione solamente gli ordini evasi o annullati, eliminando quindi le osservazioni relative alle rimanenti modalità della variabile `stato_ordine`.

3. Rimosse le osservazioni che non presentano valore nella variabile `data_spedizione_prevista`, variabile fondamentale ai fini dell'analisi.
4. Eliminate le righe del dataset che presentano valori mancanti.
5. Raggruppato i livelli delle seguenti variabili fattoriali:
   - `nazione`, raggruppando le nazioni con meno di 50 ordini;
   - `produttore`, raggruppando nella categoria 'Altro' i produttori con meno di 100 ordini;
   - `motivazione_annullamento`, raggruppando nella categoria 'Altro' tutto ciò che non fosse 'Fuori Produzione' o 'Tempi Lunghi'.
6. Creato le seguenti variabili:
   - `attesa_prevista`, come differenza tra `data_spedizione_prevista` e `data`;
   - `attesa_effettiva`, come differenza tra `data_evasione_ordine` e `data`;

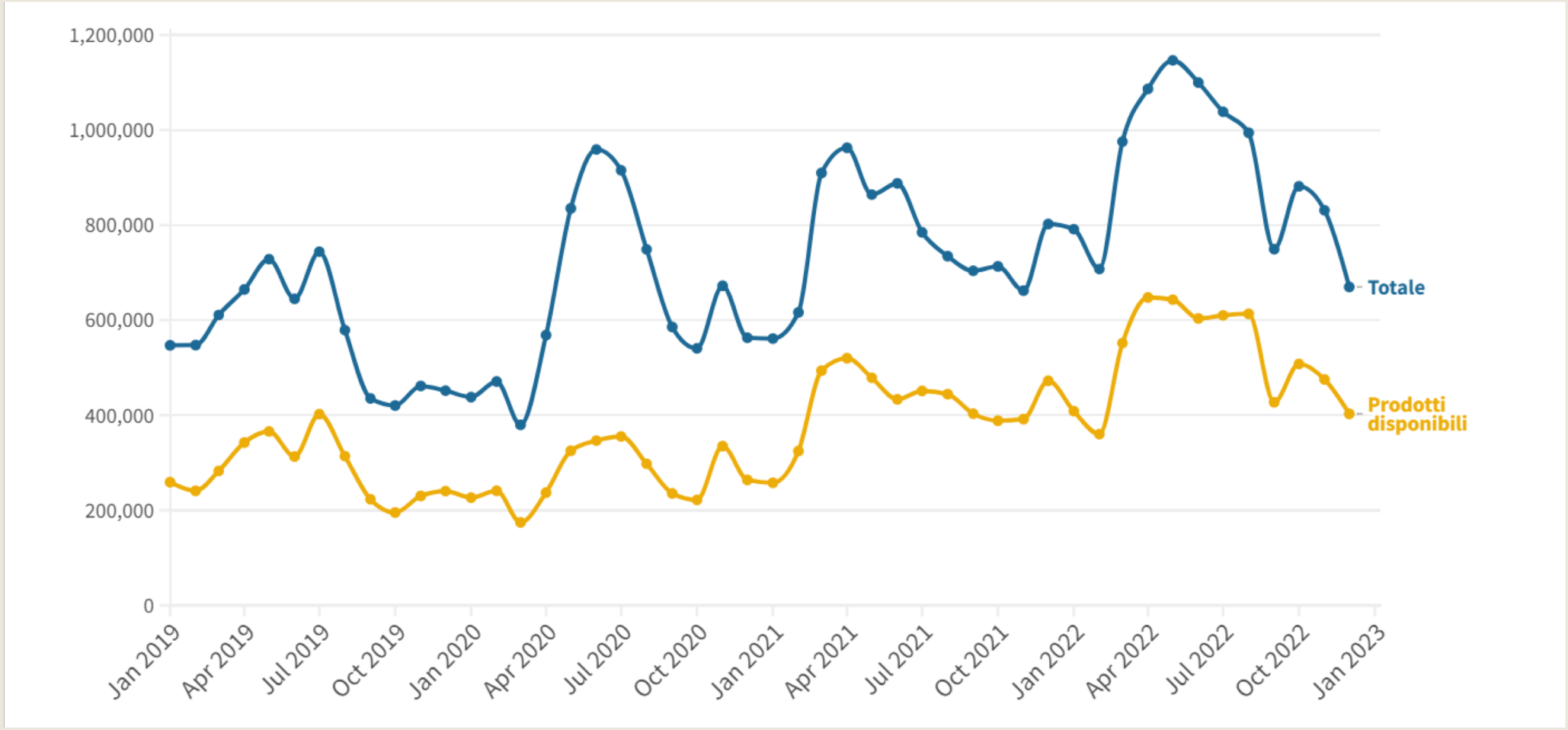
Il dataset finale presenta quindi 199.164 osservazioni e 14 variabili.

03 – ANALISI ESPLORATIVA

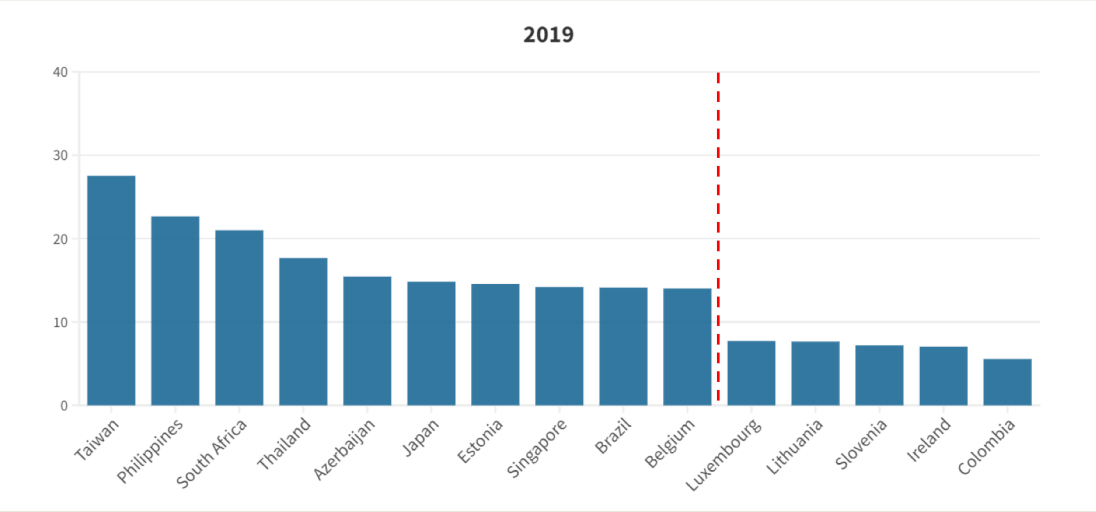
1. Confronto tra fatturato di ordini evasi e di ordini annullati

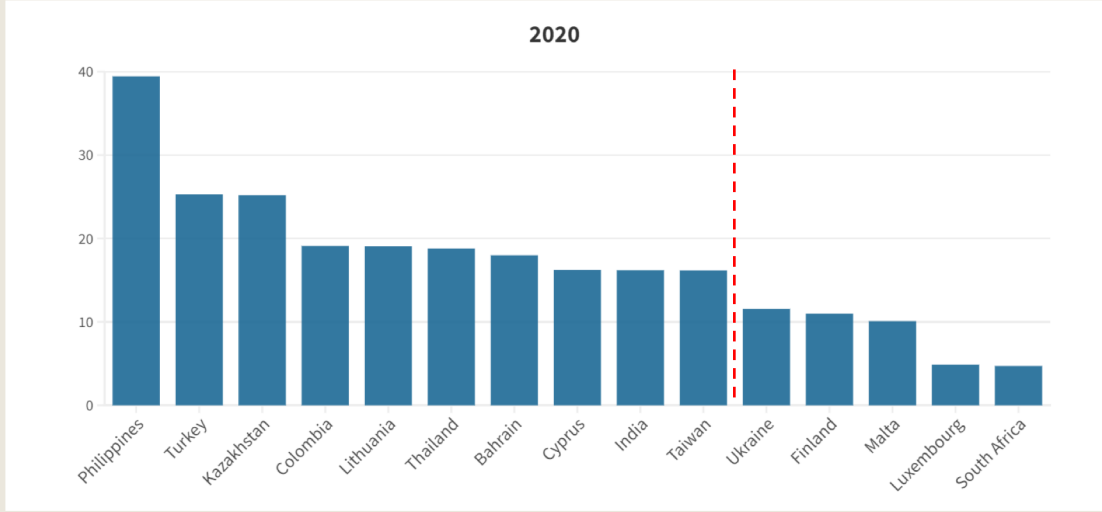


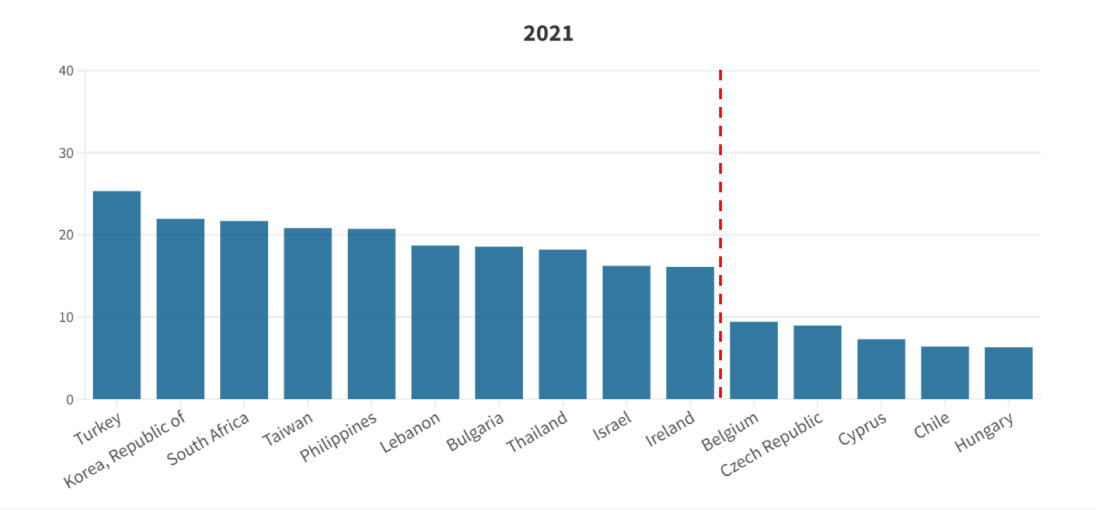
2. Confronto tra fatturato totale e fatturato dei prodotti disponibili

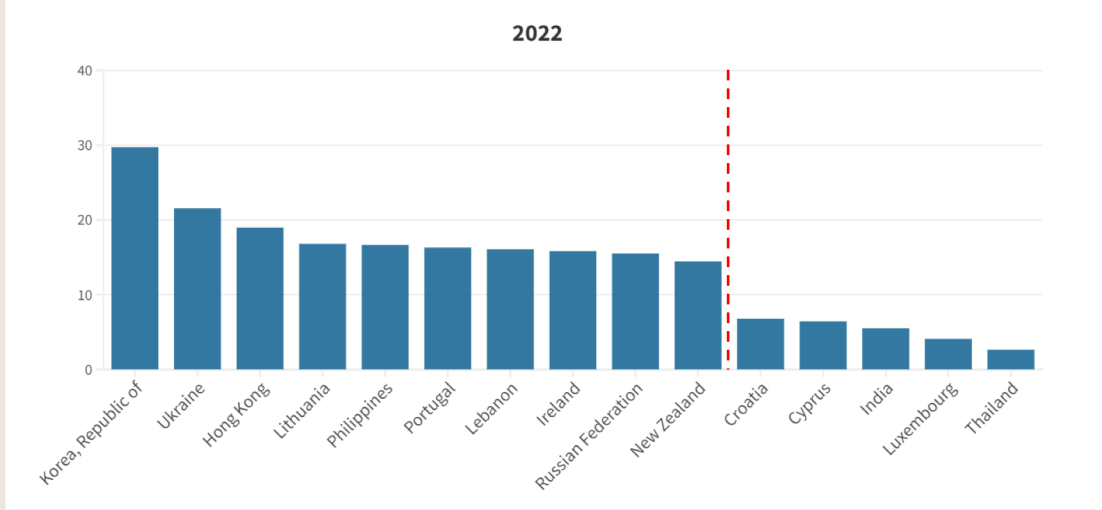


3. Confronto tra i 10 Paesi in cui si aspetta in media più giorni e i 5 in cui si aspetta meno

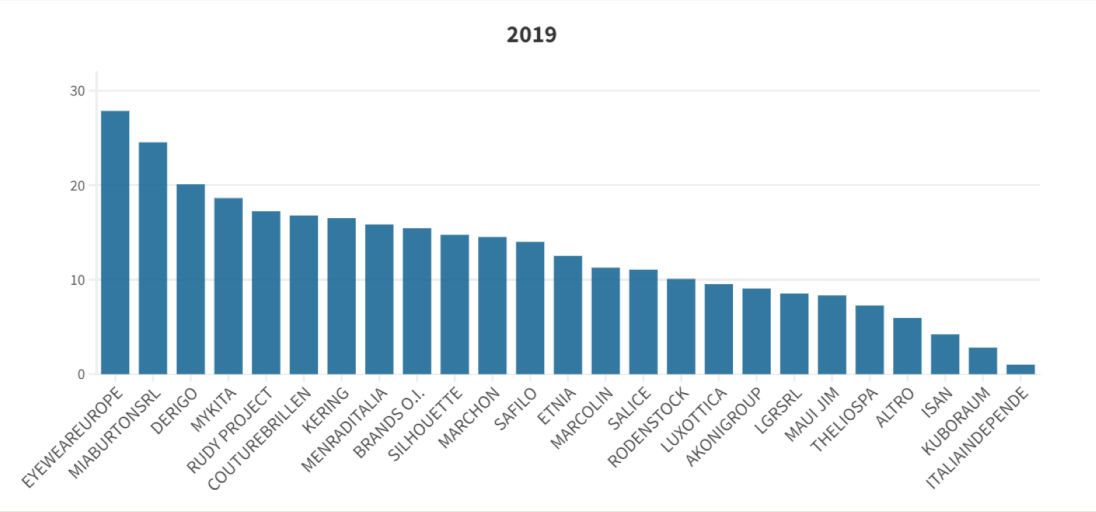


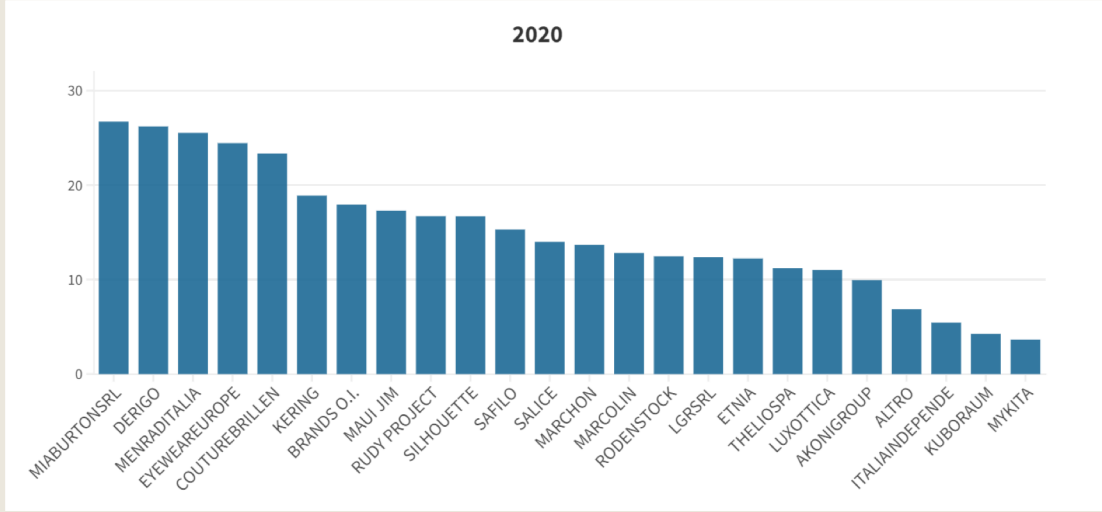


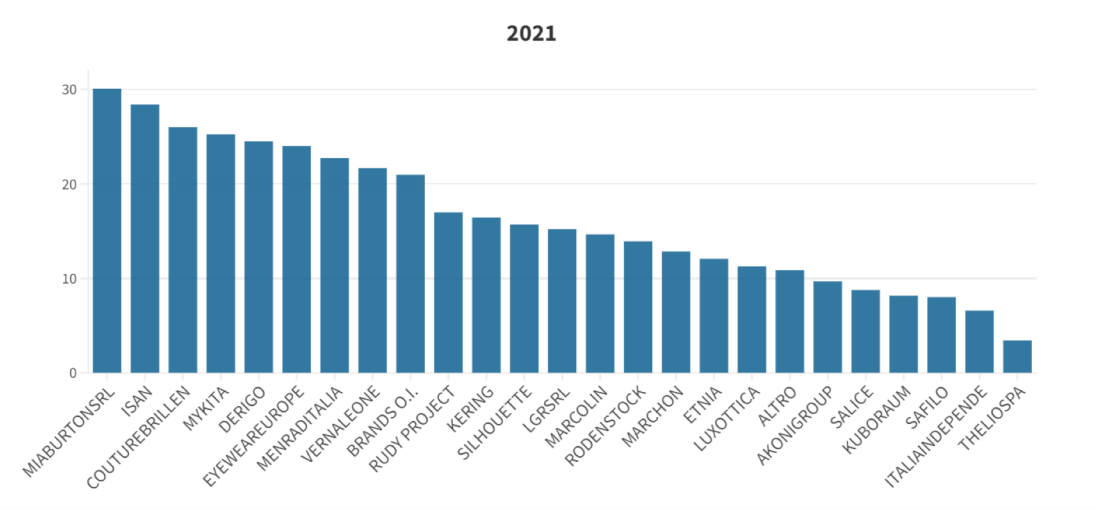


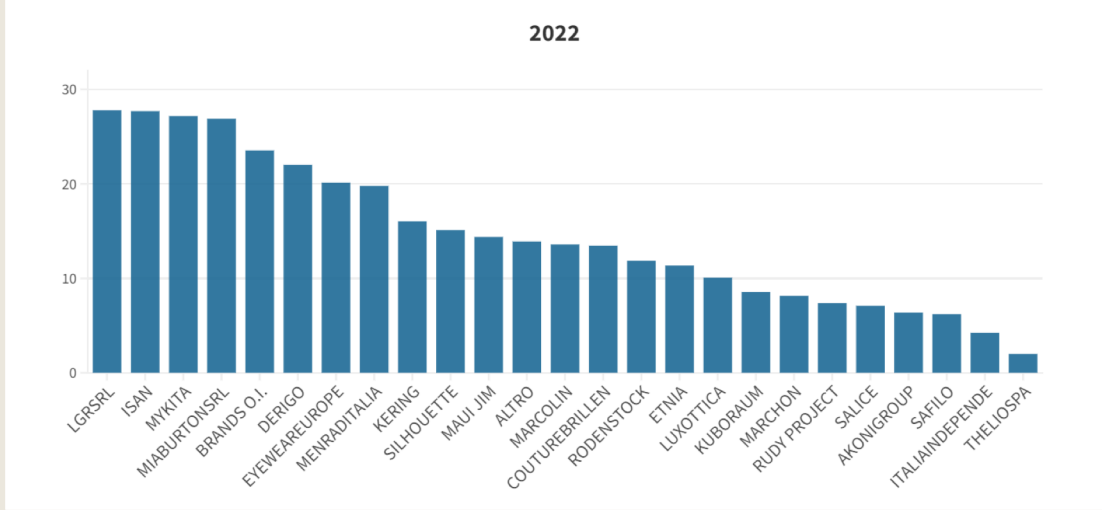


4. Confronto tra i giorni medi di attesa per ogni produttore

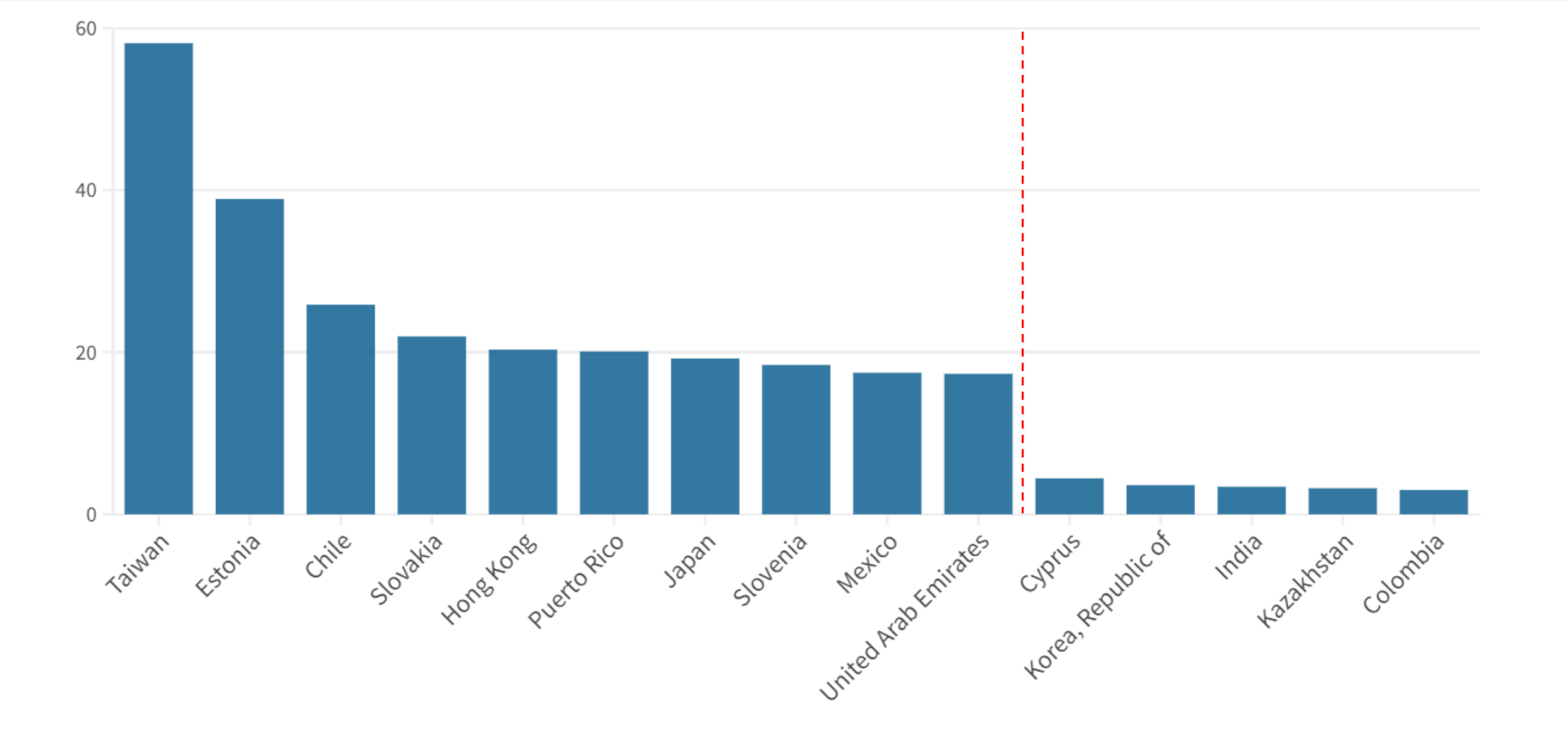




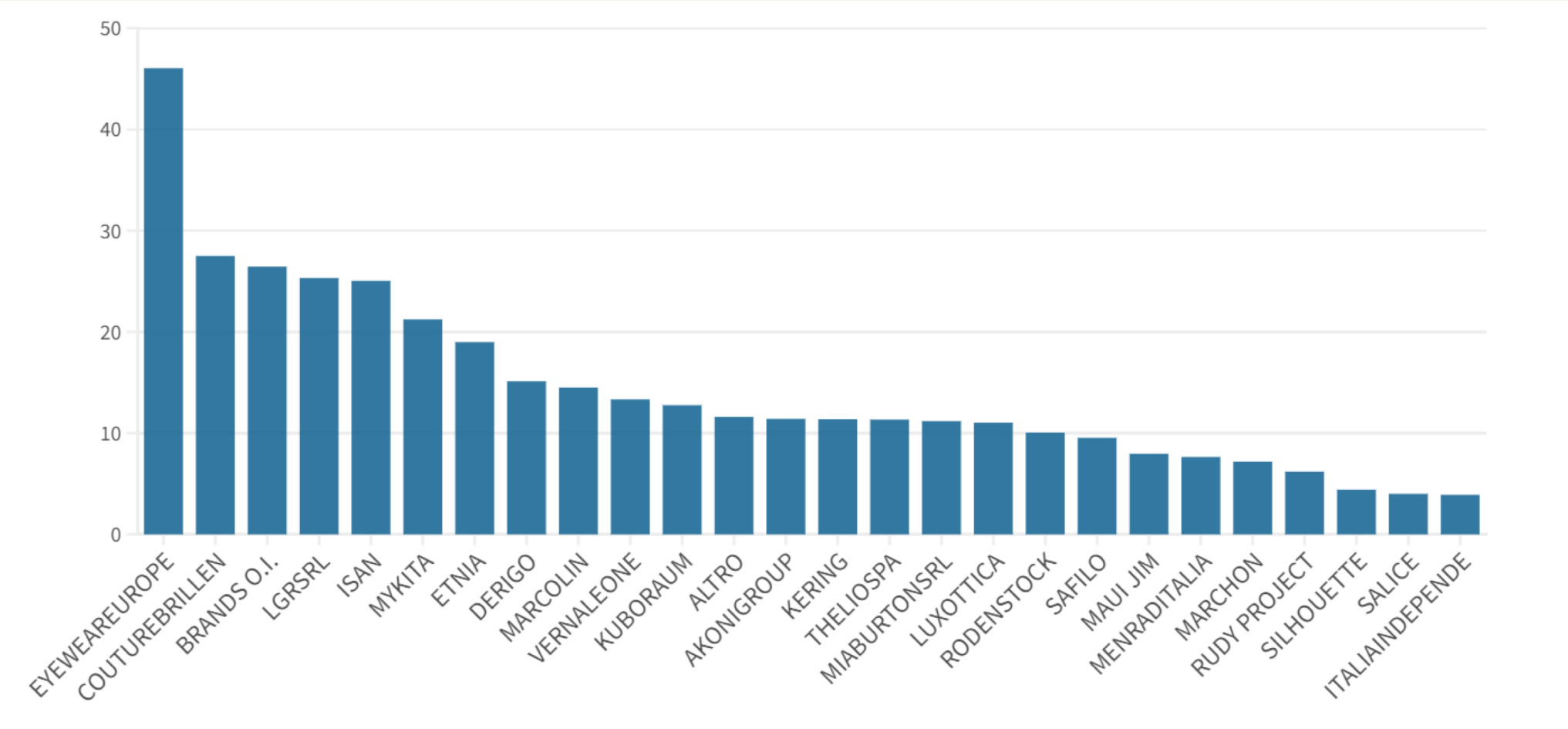




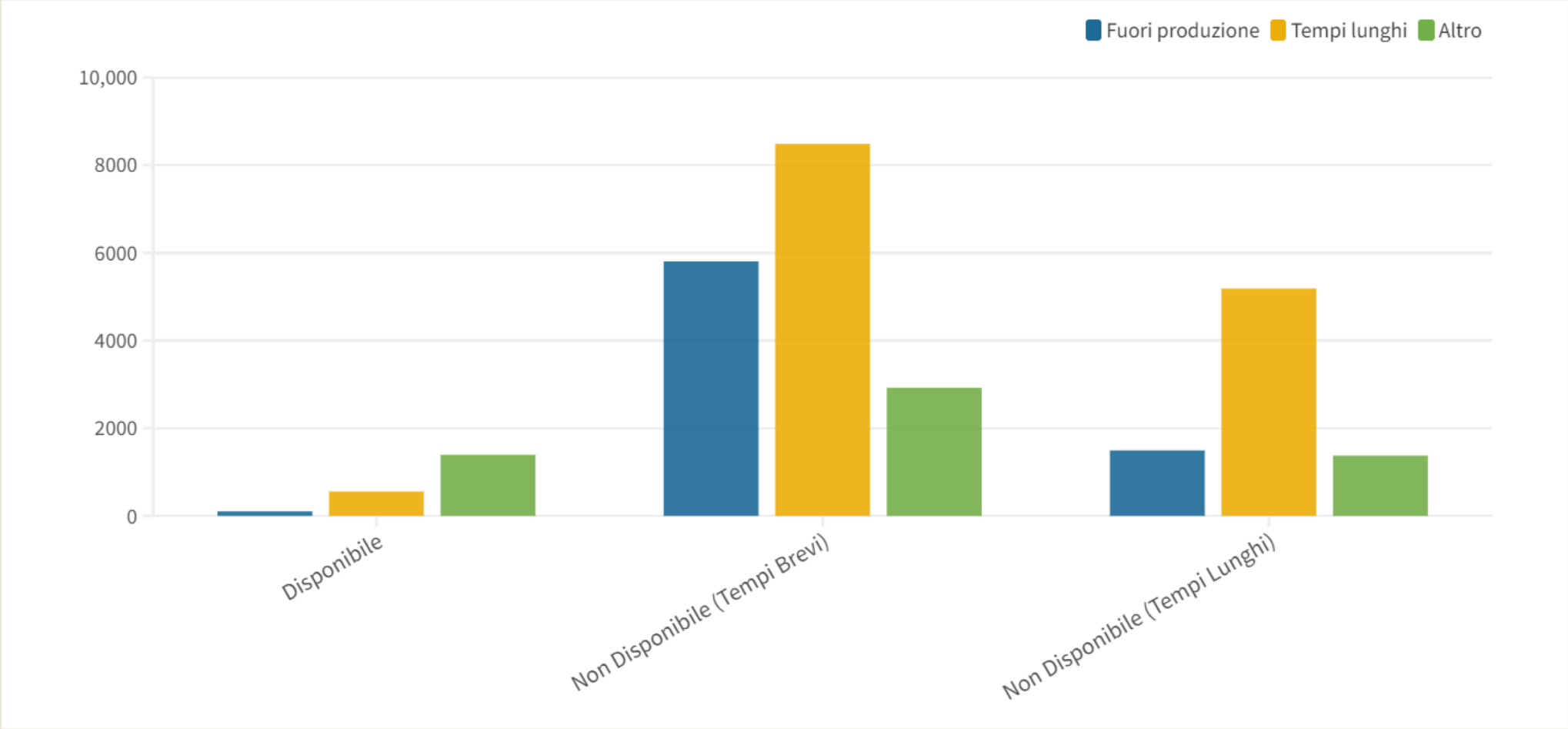
5. Confronto tra i 10 Paesi in cui la spedizione ritarda in media più giorni e i 5 in cui ritarda meno



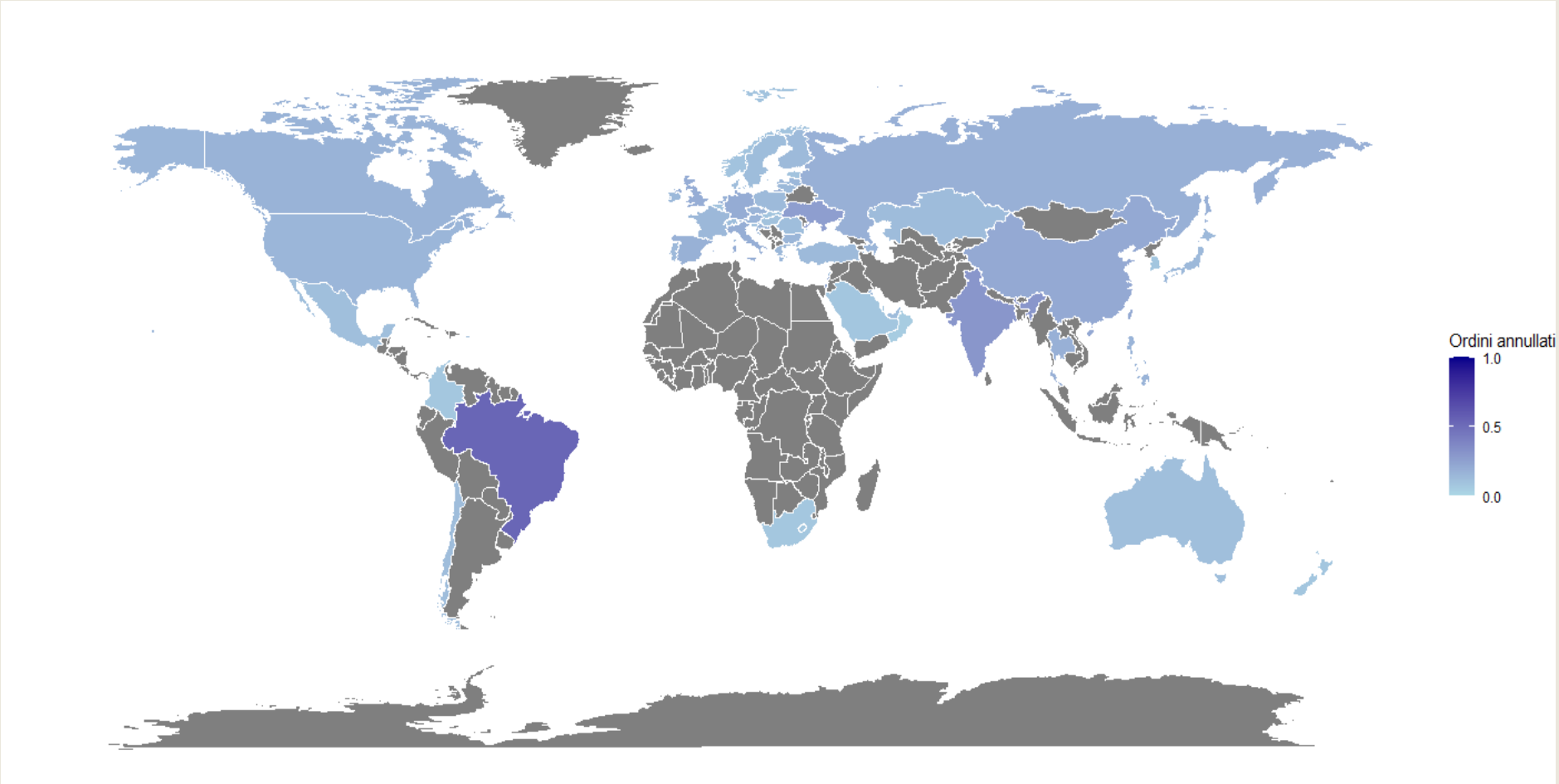
6. Confronto tra i giorni medi di ritardo nella spedizione per ogni produttore



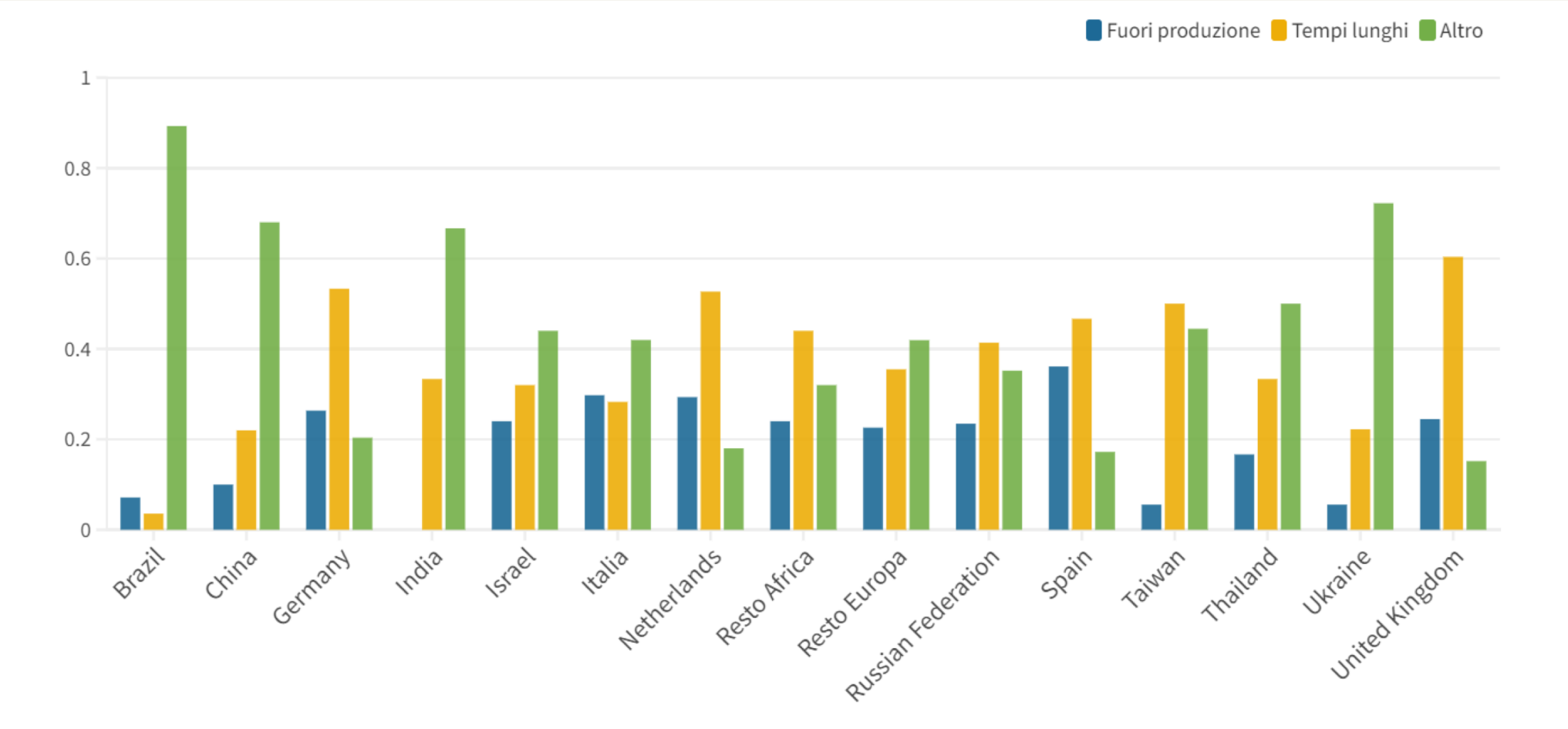
7. Motivazione di annullamento ordine rispetto alla disponibilità del prodotto



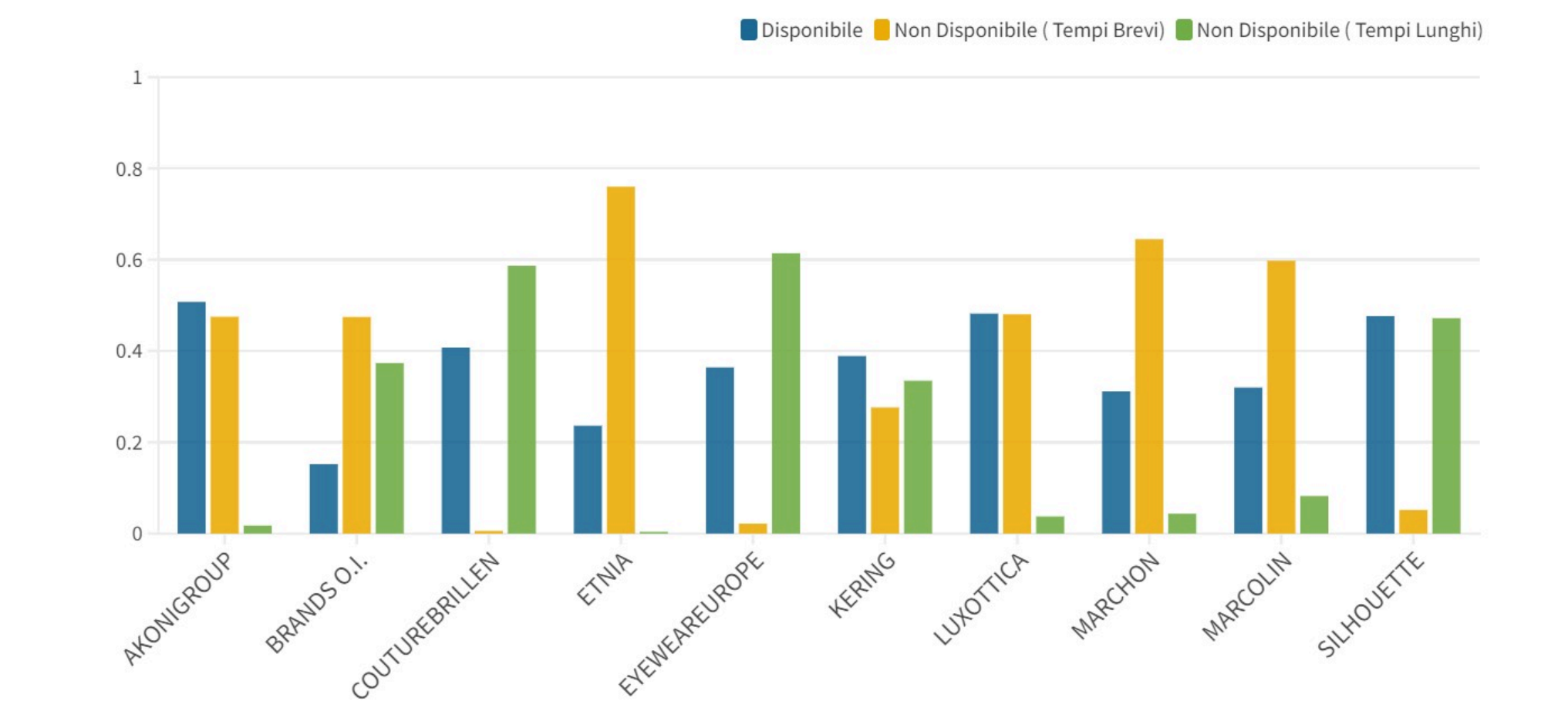
8. Mappa mondiale dei Paesi che annullano più ordini



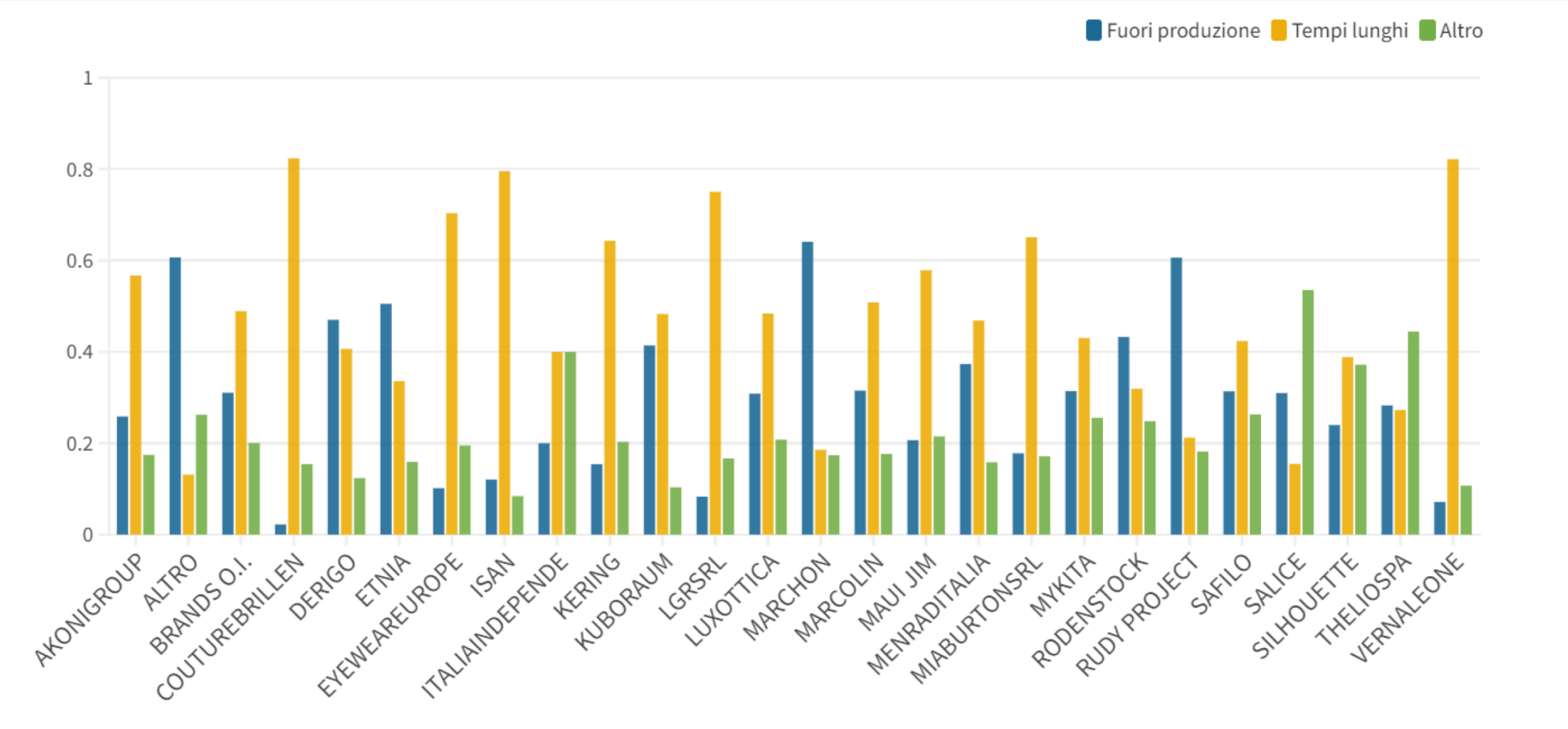
9. Motivazione di annullamento ordine dei 15 Paesi con più ordini



10. Confronto tra le disponibilità del prodotto dei 10 produttori con più vendite



11. Motivazione di annullamento ordine per produttore



04 – MODELLISTICA

Prima domanda di business – 1

- L'azienda è interessata ad analizzare quali variabili influiscono su un possibile ritardo nella spedizione dell'ordine.
- Sono stati quindi adattati modelli di classificazione volti a discriminare la variabile risposta.
- È stata creata la variabile `ritardo`, come differenza tra la data di spedizione effettiva e la data di spedizione prevista, codificata come segue:

$$y = \begin{cases} 0 \text{ se l'ordine è puntuale} \\ 1 \text{ se l'ordine è in ritardo} \end{cases}$$

- Il numero di ordini puntuali è circa 3 volte quello degli ordini in ritardo: si è deciso di non bilanciare le due classi, tenendo però conto di questo sbilanciamento nei modelli.
- In questo contesto, si è dovuto filtrare il dataset solamente per gli ordini effettivamente evasi.
- Sono state ignorate alcune variabili, ovvero `stato_ordine` e `motivazione_annullamento`.

Confronto modelli

- Il dataset finale contiene quindi 9 variabili, tra le quali la risposta, e 171.498 osservazioni.
- Viene creato un insieme di stima con 137.198 osservazioni (80% del dataset) e un insieme di verifica con 34.300 osservazioni (20% del dataset).
- Come metrica di confronto tra i vari modelli si è deciso di considerare il tasso di errata classificazione, del quale viene riportata una tabella riassuntiva.

MODELLO	ERRORE
Logistico Stepwise	0.29
LDA	0.30
MARS	0.31
Albero di Classificazione	0.27
Random Forest	0.28
Naive Bayes	0.23

Modello Logistico

```
Coefficients:
                                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                                         -1.52117    0.11414 -13.327  < 2e-16 ***
stato_ingresso_ordineNon Disponibile (Tempi Brevi)  -0.15043    0.01537  -9.787  < 2e-16 ***
stato_ingresso_ordineNon Disponibile (Tempi Lunghi) -0.80765    0.02375 -34.013  < 2e-16 ***
nazioneAltro                                         0.19827    0.05736   3.456 0.000547 ***
nazioneAustralia                                    -0.44302    0.04081 -10.855  < 2e-16 ***
nazioneAustria                                      -0.07278    0.09108  -0.799 0.424240
nazioneBahrain                                       0.63786    0.14350   4.445 8.78e-06 ***
nazioneBelgium                                      -0.39622    0.19318  -2.051 0.040263 *
nazioneCanada                                       -0.11018    0.04557  -2.418 0.015626 *
nazioneChina                                         0.59931    0.12250   4.892 9.96e-07 ***
nazioneDenmark                                      -0.15579    0.10954  -1.422 0.154957
nazioneFrance                                       -0.27887    0.05497  -5.073 3.91e-07 ***
nazioneGermany                                      -0.28254    0.04583  -6.164 7.08e-10 ***
nazioneGreece                                       -0.45204    0.19286  -2.344 0.019088 *
nazioneHong Kong                                     0.02879    0.15838   0.182 0.855747
nazioneIsrael                                        0.91831    0.14552   6.310 2.78e-10 ***
nazioneJapan                                        -0.52423    0.12804  -4.094 4.24e-05 ***
nazioneKuwait                                        0.84540    0.10066   8.399  < 2e-16 ***
nazioneNetherlands                                  -0.23617    0.10545  -2.240 0.025119 *
nazioneNew Zealand                                  -0.54370    0.12171  -4.467 7.93e-06 ***
nazioneNorway                                       -0.17063    0.05218  -3.270 0.001075 **
nazionePoland                                       -0.16687    0.14842  -1.124 0.260858
nazionePuerto Rico                                   0.36726    0.15820   2.321 0.020262 *
nazioneQatar                                         0.05726    0.10915   0.525 0.599838
nazioneRomania                                      -0.48036    0.19417  -2.474 0.013364 *
```

nazioneRussian Federation	0.84589	0.10138	8.344	< 2e-16 ***
nazioneSaudi Arabia	-0.04348	0.03925	-1.108	0.267964
nazioneSpain	-0.21549	0.09781	-2.203	0.027585 *
nazioneSweden	-0.10066	0.07348	-1.370	0.170710
nazioneSwitzerland	1.05009	0.04034	26.034	< 2e-16 ***
nazioneU.S.A.	-0.26018	0.03724	-6.986	2.82e-12 ***
nazioneUnited Arab Emirates	-0.12763	0.07538	-1.693	0.090422 .
nazioneUnited Kingdom	-0.12572	0.04662	-2.697	0.007005 **
produttoreAKONIGROUP	0.35064	0.05201	6.742	1.56e-11 ***
produttoreALTRO	0.35781	0.14385	2.487	0.012866 *
produttoreBRANDS O.I.	0.02314	0.06908	0.335	0.737681
produttoreCOUTUREBRILLEN	0.50574	0.05958	8.489	< 2e-16 ***
produttoreDERIGO	-0.20590	0.11028	-1.867	0.061889 .
produttoreETNIA	-0.05671	0.07251	-0.782	0.434160
produttoreEYEWEAREUROPE	0.57650	0.06709	8.594	< 2e-16 ***
produttoreISAN	1.00329	0.18045	5.560	2.70e-08 ***
produttoreITALIAINDEPENDE	-0.08963	0.24947	-0.359	0.719377
produttoreKERING	-0.08493	0.01926	-4.411	1.03e-05 ***
produttoreKUBORAUM	0.29865	0.21724	1.375	0.169221
produttoreLGRSRL	0.90209	0.16727	5.393	6.93e-08 ***
produttoreMARCHON	-0.24362	0.06544	-3.723	0.000197 ***
produttoreMARCOLIN	0.13190	0.02435	5.417	6.06e-08 ***
produttoreMAUI JIM	0.52017	0.07999	6.503	7.87e-11 ***
produttoreMENRADITALIA	-0.58177	0.13862	-4.197	2.71e-05 ***
produttoreMIABURTONSRL	-0.24482	0.06222	-3.935	8.33e-05 ***
produttoreMYKITA	0.10284	0.16793	0.612	0.540249
produttoreRODENSTOCK	-0.59068	0.11085	-5.329	9.89e-08 ***
produttoreRUDY PROJECT	-0.59008	0.34013	-1.735	0.082766 .
produttoreSAFILO	-0.11045	0.01989	-5.552	2.83e-08 ***
produttoreSALICE	-0.99278	0.08848	-11.221	< 2e-16 ***
produttoreSILHOUETTE	-0.65350	0.08350	-7.826	5.03e-15 ***
produttoreTHELIOSPA	-0.04852	0.10412	-0.466	0.641195
produttoreVERNALEONE	0.50890	0.15602	3.262	0.001107 **

```
tipo_prodottoOcchiali da sole          0.05320    0.10529   0.505 0.613399
tipo_prodottoOcchiali da vista         0.45466    0.10641   4.272 1.93e-05 ***
anno2020                               0.35372    0.02271  15.575  < 2e-16 ***
anno2021                               0.20977    0.02285   9.181  < 2e-16 ***
anno2022                               1.59774    0.02073  77.088  < 2e-16 ***
meseAug                               -0.05773    0.03030  -1.906 0.056710 .
meseDec                               -0.19871    0.03350  -5.931 3.01e-09 ***
meseFeb                               -0.42010    0.03430 -12.246  < 2e-16 ***
meseJan                               -0.34718    0.03355 -10.347  < 2e-16 ***
meseJul                               -0.12231    0.02954  -4.140 3.47e-05 ***
meseJun                               -0.10352    0.02944  -3.516 0.000437 ***
meseMar                               -0.22879    0.03141  -7.284 3.24e-13 ***
meseMay                               -0.14214    0.02959  -4.803 1.56e-06 ***
meseNov                               -0.38067    0.03314 -11.488  < 2e-16 ***
meseOct                               -0.16065    0.03299  -4.869 1.12e-06 ***
meseSep                               -0.20662    0.03334  -6.197 5.74e-10 ***
```

Il precedente modello logistico è stato ottenuto seguendo una procedura passo a passo ibrida, la quale ha selezionato tutte le covariate, eccetto `genere_occhiale` e `importo_articolo_euro_netto`.
Per le variabili fattoriali `nazione` e `produttore` sono stati presi come riferimenti le modalità 'Italia' e 'LUXOTTICA'.

Classificatore naïve di Bayes

Probabilità condizionate che un ordine sia spedito in ritardo:

Probabilità a priori

0	1
0.7546	0.2454

ritardo	stato_ingresso_ordine		
	Disponibile	Tempi Brevi	Tempi Lunghi
0	0.49	0.34	0.17
1	0.58	0.32	0.09

ritardo	anno			
	2019	2020	2021	2022
0	0.24	0.25	0.29	0.22
1	0.13	0.19	0.18	0.50

ritardo	mese											
	Gen	Feb	Mar	Apr	Mag	Giu	Lug	Ago	Set	Ott	Nov	Dic
0	0.074	0.073	0.083	0.089	0.101	0.102	0.102	0.086	0.070	0.069	0.080	0.072
1	0.064	0.058	0.083	0.109	0.105	0.109	0.105	0.099	0.067	0.070	0.066	0.065

Esempio:

- $\hat{P}(ritardo \mid Disponibile, Aprile, 2022) = (0.2454)(0.58)(0.109)(0.50) = 0.0078$
- $\hat{P}(orario \mid Disponibile, Aprile, 2022) = (0.7546)(0.49)(0.089)(0.22) = 0.0072$
- Vera probabilità di ritardo: $0.0078/(0.0078 + 0.0072) = \mathbf{0.52}$

- $\hat{P}(ritardo \mid Tempi\ Lunghi, Febbraio, 2019) = (0.2454)(0.09)(0.058)(0.13) = 0.00016$
- $\hat{P}(orario \mid Tempi\ Lunghi, Febbraio, 2019) = (0.7546)(0.17)(0.073)(0.24) = 0.00224$
- Vera probabilità di ritardo: $0.00016/(0.00016 + 0.00224) = \mathbf{0.069}$

Prima domanda di business – 2

- In presenza di ritardo nella spedizione dell'ordine risulta utile analizzare e quantificare il numero di giorni di ritardo e le variabili che influenzano tale fenomeno.
- Si adattano dunque dei modelli di regressione sulla variabile numerica `ritardo`, calcolata come differenza positiva tra la data di spedizione effettiva e la data di spedizione prevista, con l'obiettivo di prevedere e spiegare la variabile risposta in funzione delle covariate presenti nel dataset.
- Come visto precedentemente, si è dovuto filtrare il dataset solamente per gli ordini effettivamente evasi. Sono state quindi ignorate alcune variabili, ovvero le variabili `stato_ordine` e `motivazione_annullamento`.

Confronto modelli

- Il dataset finale contiene quindi 9 variabili, tra le quali la risposta, e 41.957 osservazioni.
- Viene creato un insieme di stima con 33.565 osservazioni (80% del dataset) e un insieme di verifica con 8.392 osservazioni (20% del dataset).
- Come metrica di confronto tra i vari modelli si è deciso di considerare l'errore quadratico medio, del quale viene riportata una tabella riassuntiva.

MODELLO	MSE
Lineare Stepwise	354.21
Lasso	353.80
MARS	333.22
Albero di Regressione	343.33
Random Forest	375.41
Gradient Boosting	325.78

Modello Lineare

```
Coefficients:
                                                  Estimate Std. Error t value Pr(>|t|)
(Intercept)                                       -2.36211    2.79383  -0.845 0.397853
stato_ingresso_ordineNon Disponibile (Tempi Brevi) 15.53302    0.25986  59.776  < 2e-16 ***
stato_ingresso_ordineNon Disponibile (Tempi Lunghi) 21.60912    0.42478  50.872  < 2e-16 ***
nazioneAltro                                       2.92067    0.90547   3.226 0.001258 **
nazioneAustralia                                   2.92951    0.66694   4.392 1.12e-05 ***
nazioneAustria                                     5.48341    1.50501   3.643 0.000269 ***
nazioneBahrain                                     1.83642    2.07053   0.887 0.375122
nazioneBelgium                                     3.79993    3.22976   1.177 0.239389
nazioneCanada                                      1.31230    0.73538   1.785 0.074348 .
nazioneChina                                      -0.61998    1.89851  -0.327 0.744002
nazioneDenmark                                    -0.85327    1.92415  -0.443 0.657439
nazioneFrance                                      1.31133    0.89553   1.464 0.143118
nazioneGermany                                     1.32263    0.74554   1.774 0.076065 .
nazioneGreece                                      1.31031    3.36165   0.390 0.696700
nazioneHong Kong                                   4.91575    2.46502   1.994 0.046138 *
nazioneIsrael                                      0.51716    2.14726   0.241 0.809676
nazioneJapan                                       6.24799    2.15635   2.897 0.003764 **
nazioneKuwait                                      1.90441    1.47878   1.288 0.197814
nazioneNetherlands                                -0.53256    1.65919  -0.321 0.748229
nazioneNew Zealand                                 3.17113    2.02467   1.566 0.117301
nazioneNorway                                      1.20626    0.85419   1.412 0.157907
nazionePoland                                     -0.14254    2.57415  -0.055 0.955842
nazionePuerto Rico                                10.86254    2.47855   4.383 1.18e-05 ***
nazioneQatar                                       3.14702    1.75725   1.791 0.073323 .
nazioneRomania                                     0.69108    3.68464   0.188 0.851224
```

nazioneRussian Federation	-0.87694	1.54686	-0.567	0.570776	
nazioneSaudi Arabia	0.76410	0.63670	1.200	0.230109	
nazioneSpain	0.93146	1.50267	0.620	0.535346	
nazioneSweden	3.07861	1.14420	2.691	0.007135	**
nazioneSwitzerland	1.05509	0.62756	1.681	0.092723	.
nazioneU.S.A.	2.23072	0.60420	3.692	0.000223	***
nazioneUnited Arab Emirates	5.88989	1.25910	4.678	2.91e-06	***
nazioneUnited Kingdom	1.90906	0.74523	2.562	0.010420	*
produttoreAKONIGROUP	0.04342	0.82290	0.053	0.957920	
produttoreALTRO	0.34272	2.35732	0.145	0.884407	
produttoreBRANDS O.I.	8.21577	1.16554	7.049	1.84e-12	***
produttoreCOUTUREBRILLEN	10.01849	0.96637	10.367	< 2e-16	***
produttoreDERIGO	-4.02782	1.85405	-2.172	0.029829	*
produttoreETNIA	1.76056	1.13935	1.545	0.122298	
produttoreEYEWEAREUROPE	25.70393	1.11490	23.055	< 2e-16	***
produttoreISAN	6.51019	2.22264	2.929	0.003402	**
produttoreITALIAINDEPENDE	-2.69710	3.65167	-0.739	0.460160	
produttoreKERING	-0.87982	0.32551	-2.703	0.006876	**
produttoreKUBORAUM	0.85477	2.73425	0.313	0.754575	
produttoreLGRSRL	4.44926	2.37396	1.874	0.060913	.
produttoreMARCHON	-1.39154	1.04343	-1.334	0.182336	
produttoreMARCOLIN	-0.14432	0.38984	-0.370	0.711221	
produttoreMAUI JIM	-4.32407	1.17665	-3.675	0.000238	***
produttoreMENRADITALIA	-5.95469	2.39106	-2.490	0.012765	*
produttoreMIABURTONSRL	-2.65952	1.03913	-2.559	0.010490	*
produttoreMYKITA	2.44581	2.64346	0.925	0.354853	
produttoreRODENSTOCK	-3.55145	1.96377	-1.808	0.070539	.
produttoreRUDY PROJECT	-4.88427	5.24633	-0.931	0.351867	
produttoreSAFILO	0.26214	0.32258	0.813	0.416431	
produttoreSALICE	-5.93248	1.54798	-3.832	0.000127	***
produttoreSILHOUETTE	-2.81945	1.43081	-1.971	0.048787	*
produttoreTHELIOSPA	-0.66668	1.85481	-0.359	0.719274	
produttoreVERNALEONE	-9.32172	2.58711	-3.603	0.000315	***

genere_occhialeDONNA	4.13309	2.70658	1.527	0.126758
genere_occhialeJUNIOR	1.06483	2.90078	0.367	0.713558
genere_occhialeUNISEX	5.77845	2.70828	2.134	0.032881 *
genere_occhialeUNISEX ADULTO	8.85041	3.32009	2.666	0.007686 **
genere_occhialeUNISEX BIMBO	4.31438	6.07078	0.711	0.477288
genere_occhialeUNISEX GIOVANE	3.31903	14.11486	0.235	0.814098
genere_occhialeUOMO	4.28763	2.70608	1.584	0.113103
anno2020	1.78860	0.39835	4.490	7.14e-06 ***
anno2021	3.68250	0.40419	9.111	< 2e-16 ***
anno2022	-2.16818	0.35997	-6.023	1.73e-09 ***
meseAug	0.55079	0.47232	1.166	0.243568
meseDec	-0.15021	0.52842	-0.284	0.776215
meseFeb	1.36758	0.55000	2.486	0.012906 *
meseJan	0.96100	0.53201	1.806	0.070872 .
meseJul	3.03938	0.46361	6.556	5.61e-11 ***
meseJun	1.21637	0.46138	2.636	0.008383 **
meseMar	0.45655	0.49159	0.929	0.353039
meseMay	-0.28848	0.46562	-0.620	0.535549
meseNov	1.79097	0.52899	3.386	0.000711 ***
meseOct	0.49106	0.51649	0.951	0.341732
meseSep	1.31491	0.52556	2.502	0.012358 *

Il precedente modello lineare è stato ottenuto seguendo una procedura passo a passo ibrida, la quale ha selezionato tutte le covariate, meno `tipo_prodotto` e `importo_articolo_euro_netto`.
Per le variabili fattoriali nazione e produttore sono stati presi come riferimenti le modalità 'Italia' e 'LUXOTTICA'.

Gradient Boosting

Importanza delle variabili:

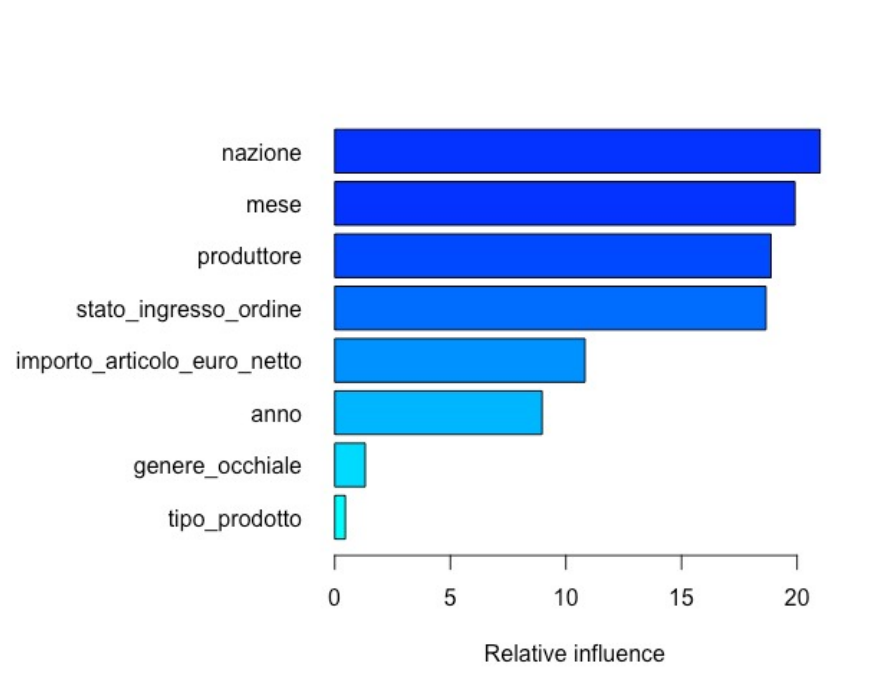
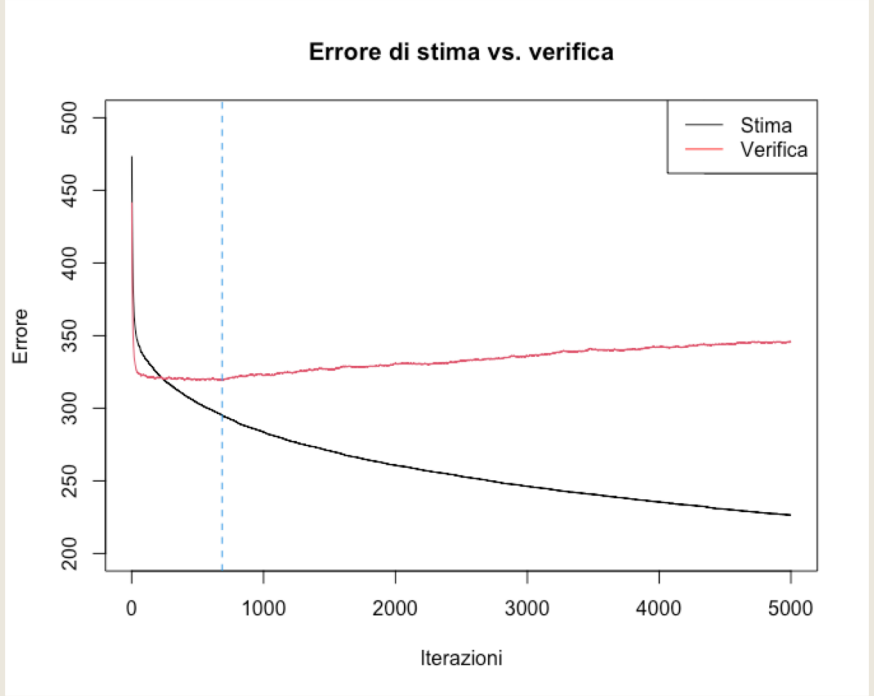


Grafico dell'errore:



Seconda domanda di business

- In secondo luogo l'azienda è interessata ad analizzare quali sono le condizioni che portano un cliente ad annullare un ordine. L'obiettivo è quindi quello di prevedere se l'ordine verrà effettivamente evaso oppure annullato in base alle sue caratteristiche.
- Si adattano dei modelli di classificazione sulla variabile `stato_ordine`, già presente del dataset.
- Il numero di ordini evasi è circa 6 volte quello degli ordini annullati: si è deciso di non bilanciare le due classi, tenendo però conto di questo sbilanciamento nei modelli.
- Anche in questo contesto, sono state ignorate alcune variabili: `motivazione_annullamento e attesa_effettiva`.

Confronto modelli

- Il dataset finale contiene quindi 10 variabili, tra le quali la risposta, e 199.164 osservazioni.
- Viene creato un insieme di stima con 159.331 osservazioni (80% del dataset) e un insieme di verifica con 39.833 osservazioni (20% del dataset).
- Come metrica di confronto tra i vari modelli si è deciso di considerare il tasso di errata classificazione, del quale viene riportata una tabella riassuntiva.

MODELLO	ERRORE
Logistico Stepwise	0.37
LDA	0.35
MARS	0.38
Albero di Classificazione	0.37
Random Forest	0.40
Naive Bayes	0.16

Modello Logistico

```
Coefficients:
                                                  Estimate Std. Error z value Pr(>|z|)
(Intercept)                                     -1.9900081  0.2094622  -9.501  < 2e-16 ***
stato_ingresso_ordineNon Disponibile (Tempi Brevi)  2.4092308  0.0295382  81.563  < 2e-16 ***
stato_ingresso_ordineNon Disponibile (Tempi Lunghi) 2.0750263  0.0433330  47.886  < 2e-16 ***
nazioneAltro                                    -0.6135378  0.0646164  -9.495  < 2e-16 ***
nazioneAustralia                                -0.8861321  0.0428502 -20.680  < 2e-16 ***
nazioneAustria                                  -0.4823708  0.1015619  -4.750 2.04e-06 ***
nazioneBahrain                                  -1.5698254  0.2853592  -5.501 3.77e-08 ***
nazioneBelgium                                  -1.0633918  0.2456760  -4.328 1.50e-05 ***
nazioneCanada                                   -0.3822212  0.0475348  -8.041 8.92e-16 ***
nazioneChina                                     0.1567753  0.1261128   1.243 0.213818
nazioneDenmark                                  -0.7562078  0.1255748  -6.022 1.72e-09 ***
nazioneFrance                                   -0.3245564  0.0602204  -5.389 7.07e-08 ***
nazioneGermany                                  -0.2152930  0.0454645  -4.735 2.19e-06 ***
nazioneGreece                                   -0.3887478  0.1930423  -2.014 0.044031 *
nazioneHong Kong                                -0.8287112  0.1925458  -4.304 1.68e-05 ***
nazioneIsrael                                   -0.0474015  0.1563846  -0.303 0.761807
nazioneJapan                                    -0.6117062  0.1276067  -4.794 1.64e-06 ***
nazioneKuwait                                   -0.5535445  0.1308226  -4.231 2.32e-05 ***
nazioneNetherlands                              -0.3071396  0.1173169  -2.618 0.008844 **
nazioneNew Zealand                              -0.9220224  0.1371174  -6.724 1.76e-11 ***
nazioneNorway                                   -0.9692927  0.0612397 -15.828  < 2e-16 ***
nazionePoland                                   -0.5528570  0.1775963  -3.113 0.001852 **
nazionePuerto Rico                              -0.7843503  0.2093942  -3.746 0.000180 ***
nazioneQatar                                    -1.1153276  0.1508807  -7.392 1.45e-13 ***
nazioneRomania                                  -0.6898650  0.2143985  -3.218 0.001292 **
```

nazioneRussian Federation	-0.2077146	0.1152176	-1.803	0.071419	.
nazioneSaudi Arabia	-1.1626326	0.0439277	-26.467	< 2e-16	***
nazioneSpain	-0.2191072	0.1063394	-2.060	0.039355	*
nazioneSweden	-0.7484324	0.0883583	-8.470	< 2e-16	***
nazioneSwitzerland	-0.7774457	0.0455226	-17.078	< 2e-16	***
nazioneU.S.A.	-0.5459201	0.0376725	-14.491	< 2e-16	***
nazioneUnited Arab Emirates	-0.5253354	0.0862143	-6.093	1.11e-09	***
nazioneUnited Kingdom	-0.1054620	0.0476163	-2.215	0.026772	*
produttoreAKONIGROUP	0.1845835	0.0634901	2.907	0.003646	**
produttoreALTRO	0.6783353	0.1850387	3.666	0.000246	***
produttoreBRANDS O.I.	-0.0805335	0.0672656	-1.197	0.231210	
produttoreCOUTUREBRILLEN	0.3895761	0.0623564	6.248	4.17e-10	***
produttoreDERIGO	0.2441435	0.0891423	2.739	0.006166	**
produttoreETNIA	0.0451461	0.0683643	0.660	0.509013	
produttoreEYEWEAREUROPE	0.6168396	0.0640993	9.623	< 2e-16	***
produttoreISAN	0.6785647	0.1626407	4.172	3.02e-05	***
produttoreITALIAINDEPENDE	-0.1434325	0.4432937	-0.324	0.746271	
produttoreKERING	-0.1899514	0.0234990	-8.083	6.30e-16	***
produttoreKUBORAUM	0.1616297	0.2780732	0.581	0.561073	
produttoreLGRSRL	0.1422059	0.1700303	0.836	0.402955	
produttoreMARCHON	0.2545685	0.0639409	3.981	6.85e-05	***
produttoreMARCOLIN	0.0264928	0.0265948	0.996	0.319169	
produttoreMAUI JIM	0.1825949	0.1243941	1.468	0.142138	
produttoreMENRADITALIA	-0.0666895	0.1046952	-0.637	0.524133	
produttoreMIABURTONSRL	0.1534535	0.0618457	2.481	0.013093	*
produttoreMYKITA	0.4773345	0.1461474	3.266	0.001090	**
produttoreRODENSTOCK	-0.2509659	0.1075105	-2.334	0.019578	*
produttoreRUDY PROJECT	0.4530325	0.2363325	1.917	0.055247	.
produttoreSAFILO	0.1615610	0.0250204	6.457	1.07e-10	***
produttoreSALICE	-2.2182585	0.1476561	-15.023	< 2e-16	***
produttoreSILHOUETTE	-0.8392035	0.1086565	-7.723	1.13e-14	***
produttoreTHELIOSPA	-0.0887553	0.1302570	-0.681	0.495627	
produttoreVERNALEONE	0.2257996	0.1854986	1.217	0.223506	

```
tipo_prodottoOcchiali da sole          -0.7267889  0.1149820  -6.321 2.60e-10 ***
tipo_prodottoOcchiali da vista         -0.3751559  0.1159143  -3.236 0.001210 **
genere_occhialeDONNA                   -0.4772608  0.1685492  -2.832 0.004632 **
genere_occhialeJUNIOR                  -0.6756082  0.1829659  -3.693 0.000222 ***
genere_occhialeUNISEX                  -0.6185911  0.1687137  -3.667 0.000246 ***
genere_occhialeUNISEX ADULTO           -0.2609531  0.2115314  -1.234 0.217338
genere_occhialeUNISEX BIMBO            -0.5397423  0.3576912  -1.509 0.131309
genere_occhialeUNISEX GIOVANE          -7.9800407 39.0849455  -0.204 0.838219
genere_occhialeUOMO                    -0.5567861  0.1685455  -3.303 0.000955 ***
anno2020                               -0.0002633  0.0240268  -0.011 0.991256
anno2021                                0.2486896  0.0234461  10.607  < 2e-16 ***
anno2022                                0.2691025  0.0234389  11.481  < 2e-16 ***
attesa_prevista                         0.2284305  0.0133631  17.094  < 2e-16 ***
meseAug                                 0.0658062  0.0368265   1.787 0.073950 .
meseDec                                -0.0311072  0.0400878  -0.776 0.437764
meseFeb                                -0.1468737  0.0398983  -3.681 0.000232 ***
meseJan                                -0.2293797  0.0406199  -5.647 1.63e-08 ***
meseJul                                 0.1839014  0.0344330   5.341 9.25e-08 ***
meseJun                                 0.1006179  0.0340841   2.952 0.003157 **
meseMar                                -0.0420537  0.0372423  -1.129 0.258817
meseMay                                 0.0631627  0.0346411   1.823 0.068252 .
meseNov                                -0.0149357  0.0393232  -0.380 0.704078
meseOct                                 0.0435143  0.0386642   1.125 0.260402
meseSep                                 0.0733822  0.0385547   1.903 0.056998 .
```

Il precedente modello logistico è stato ottenuto seguendo una procedura passo a passo ibrida, la quale ha selezionato tutte le covariate, meno `importo_articolo_euro_netto`.
Per le variabili fattoriali nazione e produttore sono stati presi come riferimenti le modalità 'Italia' e 'LUXOTTICA'.

Classificatore naïve di Bayes

Probabilità condizionata che un ordine sia annullato:

Probabilità a priori:

0	1
0.8613	0.1387

Annullato	stato_ingresso_ordine		
	Disponibile	Tempi Brevi	Tempi Lunghi
0	0.51	0.34	0.15
1	0.08	0.63	0.29

Annullato	anno			
	2019	2020	2021	2022
0	0.21	0.24	0.26	0.29
1	0.19	0.24	0.28	0.29

Annullato	mese											
	Gen	Feb	Mar	Apr	Mag	Giu	Lug	Ago	Set	Ott	Nov	Dic
0	0.071	0.069	0.083	0.095	0.100	0.101	0.103	0.089	0.070	0.069	0.076	0.070
1	0.058	0.062	0.080	0.093	0.110	0.120	0.116	0.085	0.073	0.072	0.077	0.063

Esempio:

- $\hat{P}(annullato \mid Tempi\ Brevi, Giugno, 2022) = (0.1387)(0.63)(0.120)(0.29) = 0.0031$
- $\hat{P}(evaso \mid Tempi\ Brevi, Giugno, 2022) = (0.8613)(0.34)(0.101)(0.29) = 0.0086$
- Vera probabilità di annullamento: $0.0031/(0.0031 + 0.0086) = \mathbf{0.26}$

- $\hat{P}(annullato \mid Disponibile, Gennaio, 2019) = (0.1387)(0.08)(0.058)(0.19) = 0.00012$
- $\hat{P}(evaso \mid Disponibile, Gennaio, 2019) = (0.8613)(0.51)(0.071)(0.21) = 0.01833$
- Vera probabilità di annullamento: $0.00012/(0.00012 + 0.01833) = \mathbf{0.018}$

Altre Analisi

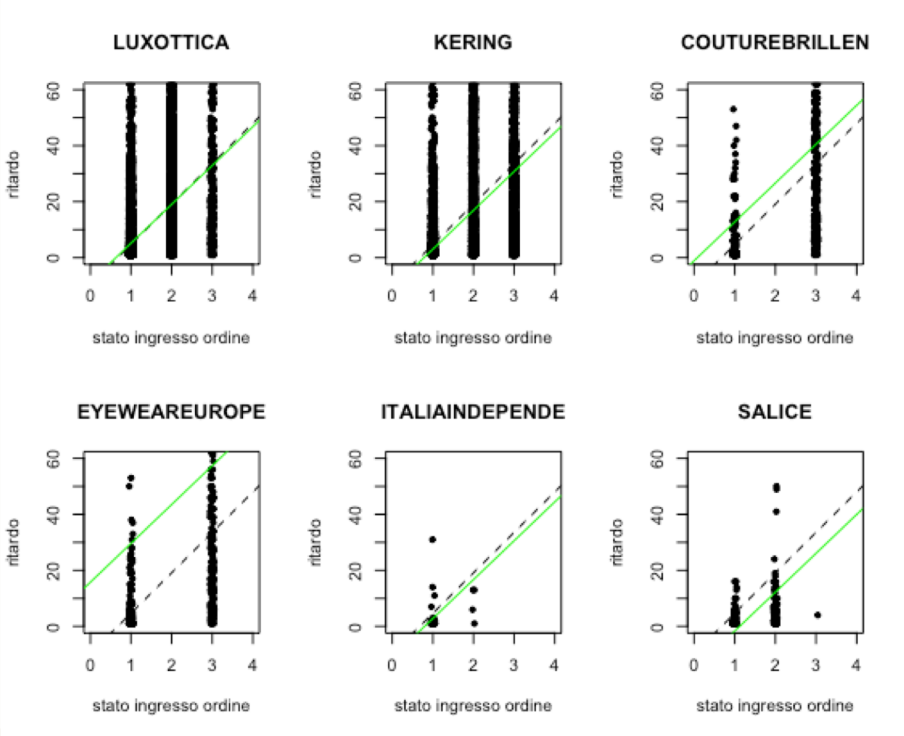
- Oltre alle domande di business richieste dall'azienda, si è ritenuto opportuno effettuare ulteriori analisi con il fine di approfondire altri aspetti potenzialmente rilevanti.

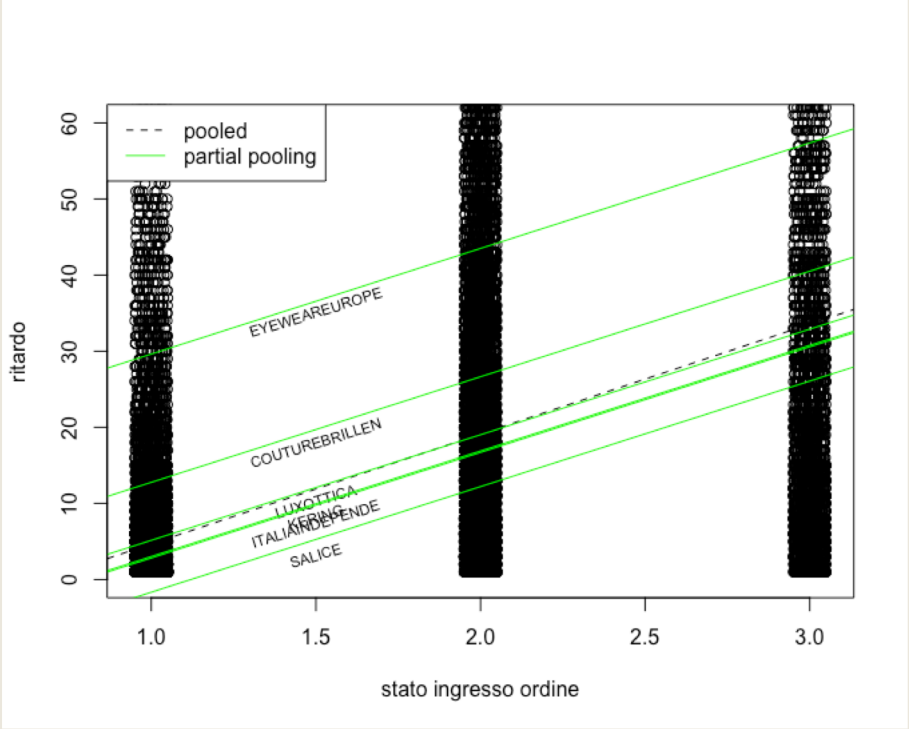
- In particolare, sono stati adattati modelli lineari gerarchici e sono state individuate delle regole associative, con lo scopo di identificare e descrivere le relazioni tra alcune variabili del dataset.

Modelli gerarchici

- Si adatta un modello lineare gerarchico utilizzando come variabile risposta la variabile numerica `ritardo`, poiché si ritiene che i dati contengano informazioni di gruppo, nel caso in esame rappresentato dalla variabile `produttore`.
- Si considera, inoltre, un secondo predittore come pendenza, rappresentato dalla variabile `stato_ingresso_ordine`.

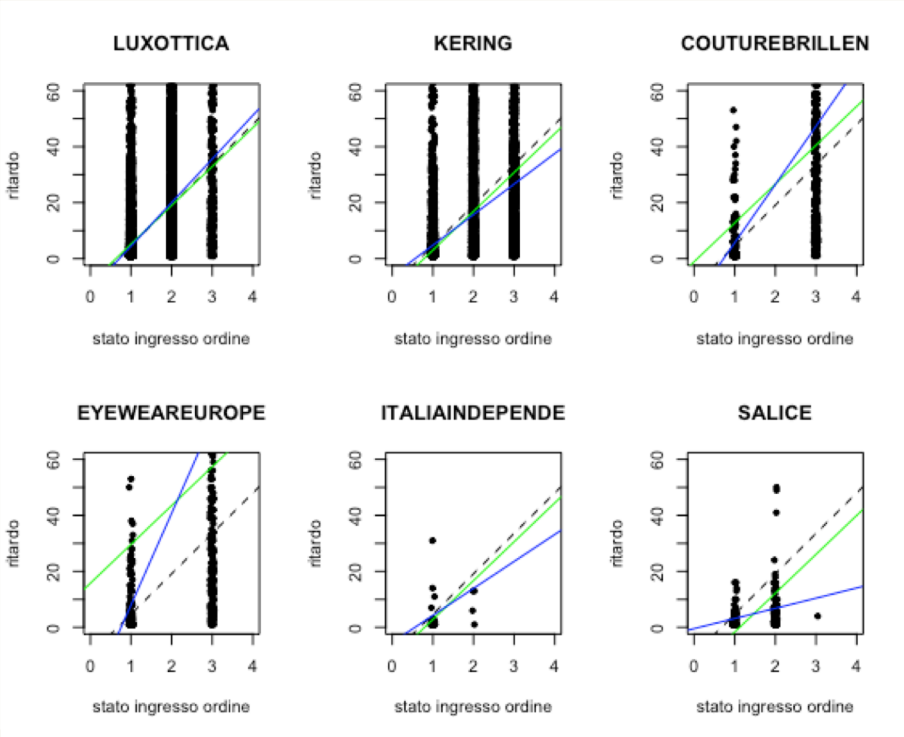
1. Modello a intercetta variabile: modello di regressione con intercetta casuale su `produttore` ed effetti fissi su `stato_ingresso_ordine`

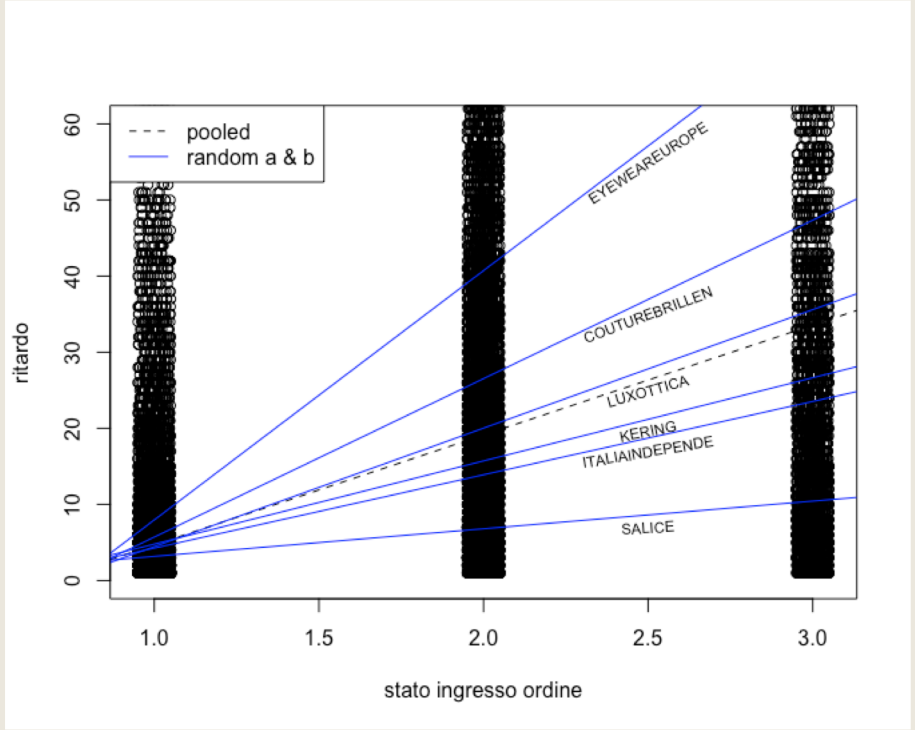




AIC: 379259.5

2. Modello a intercetta e pendenza variabile: modello di regressione con intercetta casuale su `produttore` ed effetti sia fissi che casuali su `stato_ingresso_ordine`





AIC: 368426.6

Regole associative

- Si cercano regole associative per identificare regolarità di acquisto da parte dei Paesi acquirenti.
- In particolare vengono identificati i marchi di occhiali più venduti in uno stesso Paese, come evidenziato dalla seguente tabella.

Regole associative

LHS (marchio)		RHS (nazione)	SUPPORT	CONFIDENCE	COVERAGE	LIFT	COUNT
Salice	→	Italia	0.005	0.53	0.009	12.88	1069
Mont Blanc	→	Saudi Arabia	0.003	0.70	0.004	5.26	580
Lacoste	→	Saudi Arabia	0.002	0.69	0.004	5.20	520
Tommy Hilfiger	→	Saudi Arabia	0.005	0.64	0.008	4.78	1078
Calvin Klein	→	Saudi Arabia	0.001	0.54	0.002	4.06	219
Cazal	→	USA	0.008	0.53	0.014	1.66	1632
Smith Optics	→	USA	0.003	0.53	0.007	1.65	800
Tom Ford	→	USA	0.039	0.53	0.073	1.64	8171
Alexander McQueen	→	USA	0.002	0.52	0.003	1.62	325
Oliver Peoples	→	USA	0.032	0.52	0.062	1.59	6723
Gucci	→	USA	0.044	0.50	0.089	1.56	9374

05 – CONCLUSIONI

Problemi principali

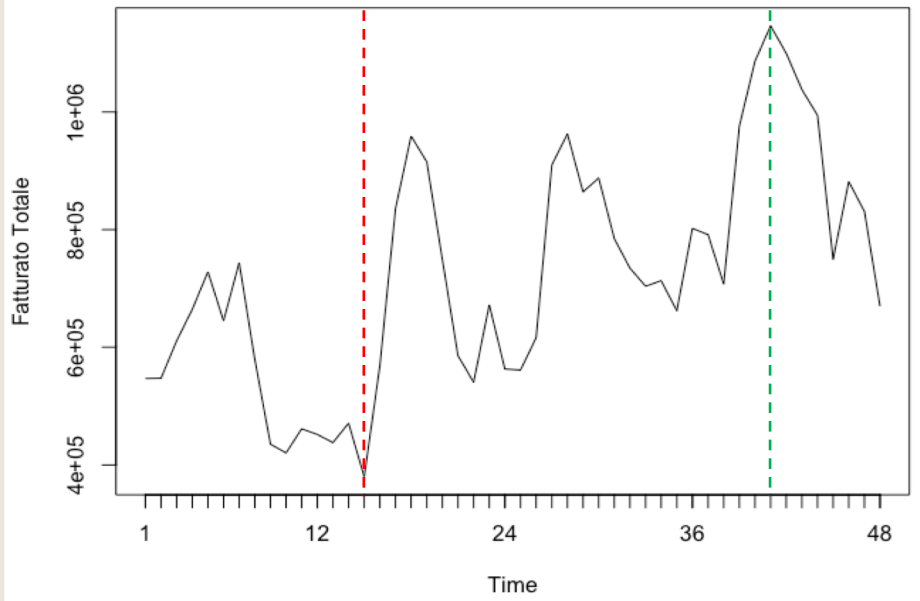
PRODOTTI NON DISPONIBILI

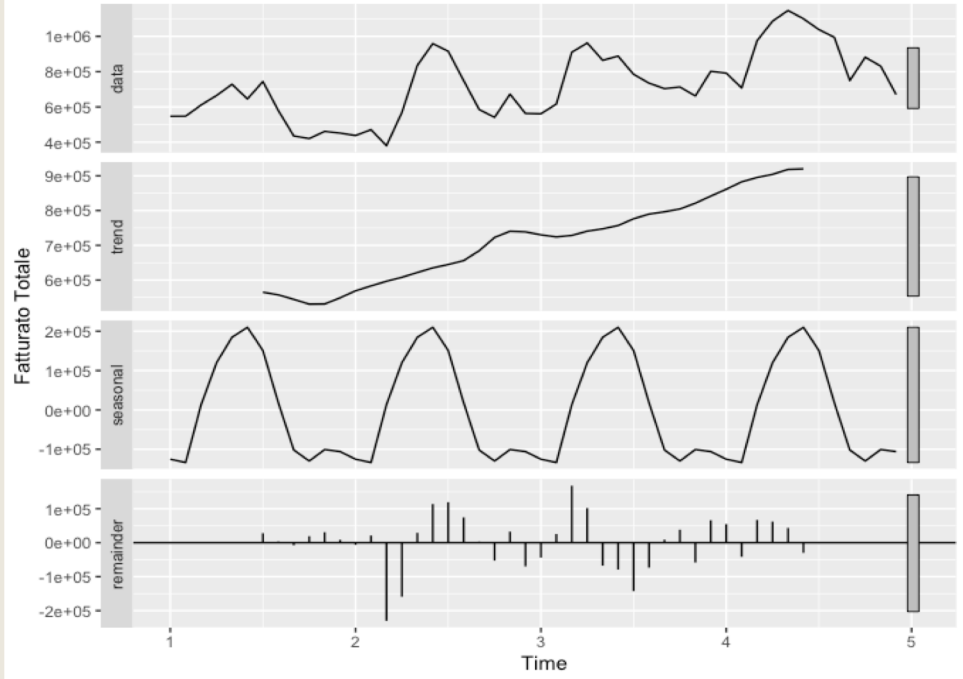
Ritardo nella spedizione

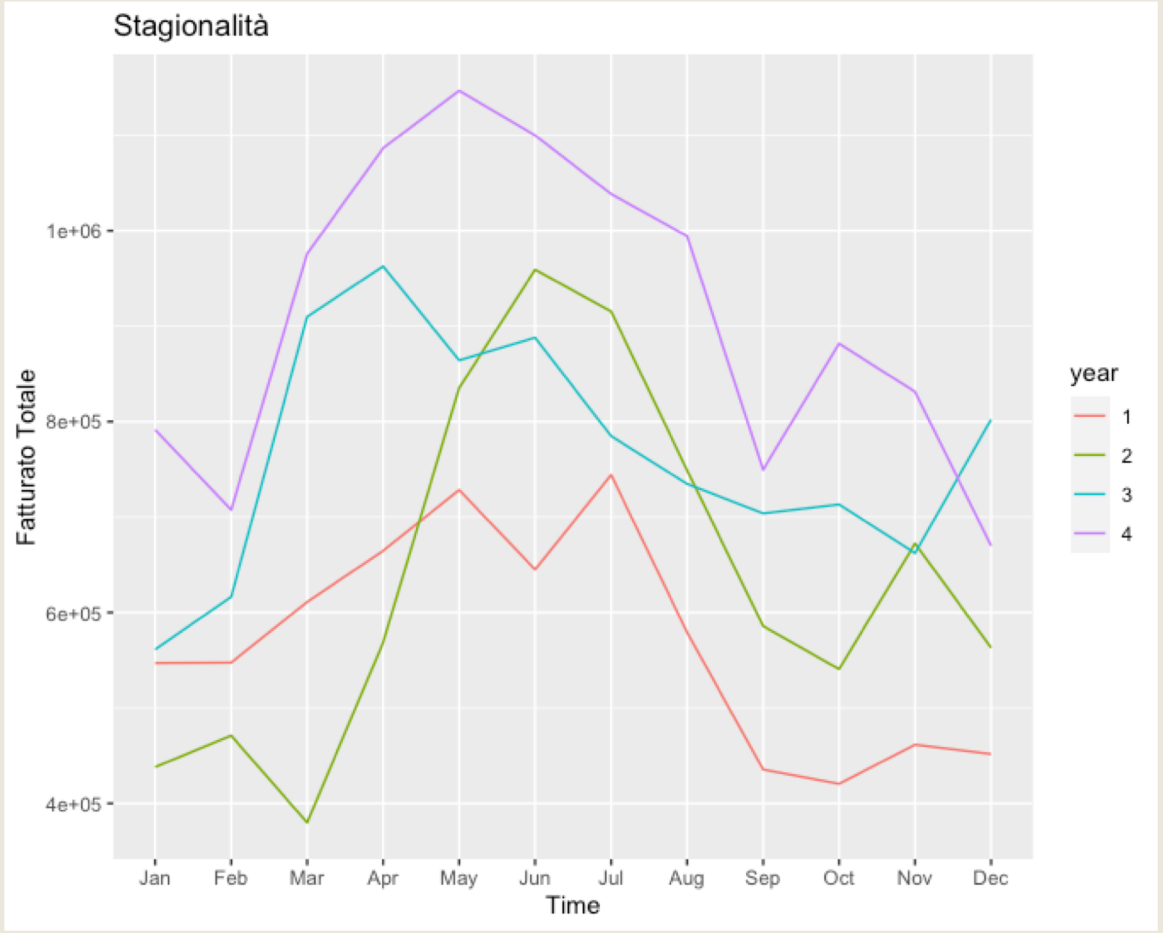
Annullamento dell’ordine

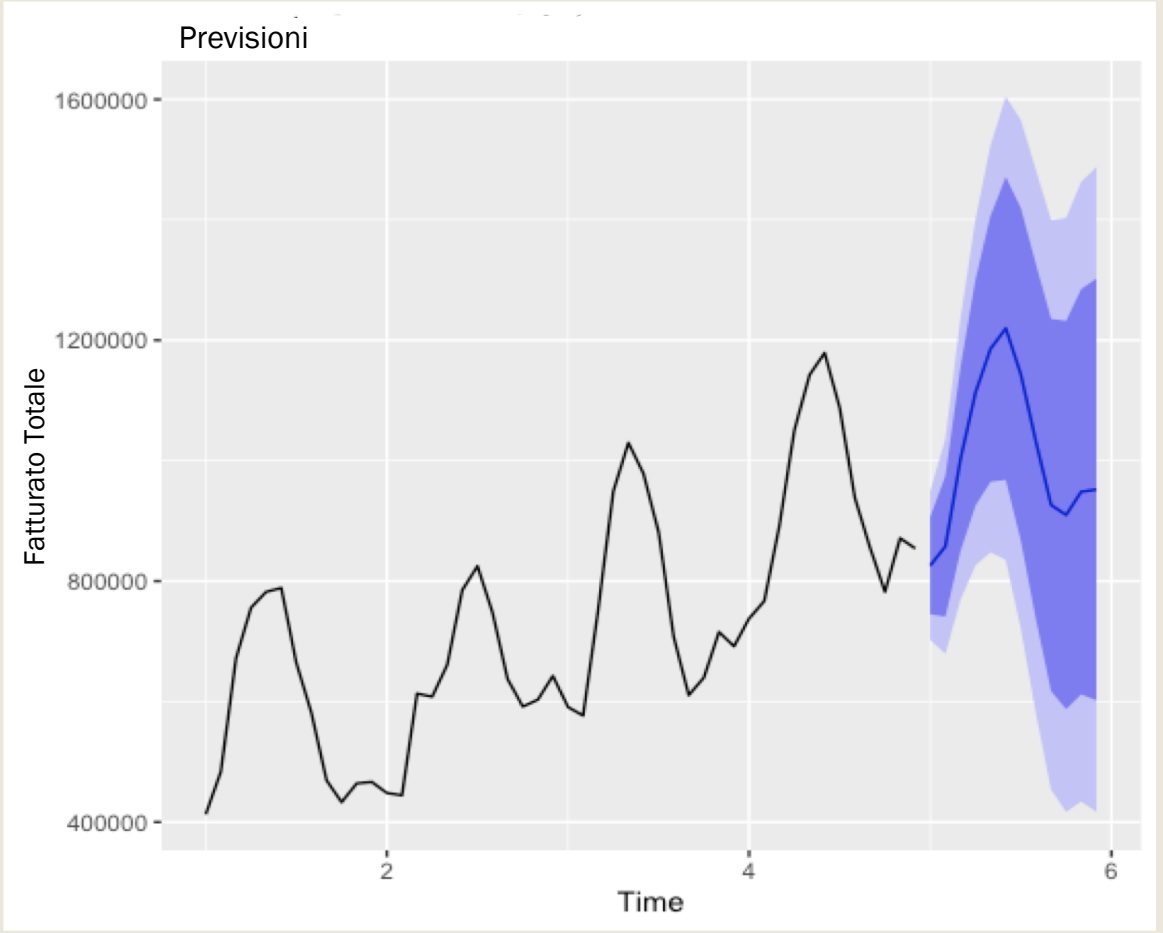
Serie mensili fatturato

- Trend crescente delle vendite
- Presenza di stagionalità
- Picchi di massimo in corrispondenza dei mesi estivi e dei giorni del Black Friday
- Picco di minimo in corrispondenza del Covid19






Stagionalità
Fatturato Totale
1e+06
8e+05
6e+05
4e+05
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Time
year
1
2
3
4


Previsioni
Fatturato Totale
1600000
1200000
800000
400000
2
4
6
Time

Ritardo nella spedizione

- Paesi con probabilità alta di ritardo nella spedizione: Svizzera, Israele, Russia e Paesi asiatici
- Mercato cinese: tra i più crescenti nei prossimi anni
- Produttori con probabilità alta di ritardo nella spedizione: Isan, L.G.R., Eyewear Europe, Couture Brillen
- Soluzioni:
  - *Rifornire magazzini nelle aree limitrofe ai Paesi con più ritardo, o istituirne di nuovi*
  - *Aumentare le scorte dei magazzini con occhiali dei produttori con più ritardo*
  - *Aumentare le scorte dei magazzini a ridosso dei mesi estivi*

Annullamento degli ordini

- Potenziali guadagni tra i 100.000€ e i 200.000€ mensili, non fatturati
- Due motivazioni principali: prodotti fuori produzione e tempi di attesa lunghi
- Fuori produzione:
  - *Rimuovere gli articoli man mano che vengono ordinati*
- Tempi di attesa lunghi:
  - *Ridefinire rifornimento magazzini*
- Paesi che sono meno propensi ad annullare: USA, Australia, Svizzera e Arabia Saudita

GRAZIE PER L'ATTENZIONE