Analisi degli effetti del Covid-19 sull'economia mondiale

Conte Enrico Matricola 852679 Doci David Matricola 799647 Filip Sara Matricola 852864

Università degli studi di Milano Bicocca

Abstract

In questo studio ci proponiamo di investigare circa la relazione esistente tra le variazioni giornaliere di persone infette dal virus Sars-Covid19 e le variazioni giornaliere degli indicatori di borsa dei relativi paesi. Non potendo analizzare tutti i paesi del mondo, abbiamo deciso di restringere il nostro studio alle principali economie mondiali. In particolare, a: Italia, Francia, Spagna, Germania, Regno Unito, Russia, USA, Brasile, Argentina, Canada, India, Turchia, Arabia Saudita, Giappone, Cina e Corea del Sud. I dati sono stati raccolti per un periodo complessivo di sei mesi, da gennaio a giugno. Per realizzare questo studio abbiamo sfruttato l'architettura Lambda resa possibile grazie agli applicativi Kafka (architettura producer-consumer), e, tramite la libreria python PyMongo, siamo riusciti ad automatizzare l'operazione di storage dei dati all'interno di MongoDB. Infine, per realizzare le visualizzazioni abbiamo sfruttato il software Tableau Desktop.

Tutti i membri hanno contribuito in egual misura allo sviluppo del progetto.

Link per le visualizzazioni interattive:

https://public.tableau.com/profile/enrico.conte#!/vizhome/Progetto DataViz 1/Indici Covid

1. Acquisizione e streaming dei dati

Avendo scelto di concentrarci su Velocity e Variety, nella parte inziale del lavoro ci siamo concentrati sulla scelta delle fonti dei dati e sulla loro acquisizione in tempo reale.

Abbiamo raccolto dati relativi al numero di persone positive al virus Sars-Covid19; i dati utilizzati, resi disponibili e aggiornati giornalmente sulla pagina Github della John Hopkins University¹, sono stati raccolti in Real Time attraverso un'architettura Lambda (Kafka). Nel dettaglio, innanzitutto abbiamo inizializzato un'istanza di Kafka Zookeeper per facilitare lo scambio di informazioni tra i diversi nodi dell'architettura (Fig.A Appendice). Successivamente, abbiamo creato un server locale di Kafka (Fig.B Appendice). Una volta connesso il server locale a Zookeeper abbiamo creato i topic atti ad ospitare i dati in streaming. Più in dettaglio, abbiamo deciso di mantenere i dati separati in questa prima fase di importazione, da ciò la creazione di due diversi topic: uno per i dati del contagio e l'altro per i dati di natura finanziaria. Creati i topic, abbiamo inizializzato anche il Producer e il Consumer, connettendoli entrambi alla porta 9092 del localhost, così da garantire che ciò che arriva al Producer venga correttamente inviato al Consumer (Fig. C e D Appendice). In Figura 1 si óuq vedere schematizzazione una dell'architettura utilizzata.

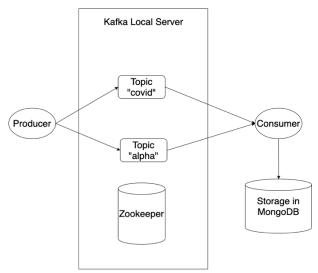


Figura.1 – Schematizzazione dell'architettura Lambda utilizzata

Le informazioni relative all'andamento degli indici borsa sono state acquisite quotidianamente in tempo reale dal sito AlphaVantage². Tale sito, infatti, mette a disposizione degli utenti una API che permette di ottenere i dati d'interesse nel tempo desiderato. Anche in questo caso, l'architettura producer-consumer ci permesso di sottoscrivere il topic "alpha" e di ricevere le nuove informazioni non appena il producer scrive nuovi dati sulla coda Kafka per il topic dichiarato.

2

https://github.com/CSSEGISandData/COVID-19/blob/master/csse covid 19 data/csse covid 19 t ime_series/time_series_covid19_confirmed_global.csv

2. Storage dei dati

Successivamente, abbiamo realizzato collegamento tra il Consumer e MongoDB tramite uno script in python, che ci consente trasferire i dati inviati al Consumer direttamente all'interno di una Collection di MongoDB in formato BSON (Figura.E Appendice). Per automatizzare il processo di raccolta ed immagazzinamento dei dati abbiamo avviato un ciclo infinito (While=True), per controllare sempre la disponibilità di nuovi dati, concedendo al ciclo, dopo ogni esecuzione, di andare in modalità sleep per un'ora; in tal modo ogni nuovo dato disponibile viene letto e stampato dal consumer.

In questo modo, abbiamo inviato i dati epidemiologici e quelli economici a due Collezioni all'interno di un unico Database locale in MongoDB.

I dati si riferiscono al periodo Gennaio - Giugno 2020, e, poiché la raccolta real time è stata effettuata nel periodo **, i restanti dati sono stati ottenuti dall'archivio storico messo a disposizione dai relativi siti. Questa valutazione è stata fatta per consentire una visione complessiva più ampia del fenomeno, poiché sarebbe stato riduttivo concentrarsi su un periodo di osservazione troppo ristretto, dal momento che il fenomeno non è nato e non si è esaurito in breve tempo.

Una volta svolte tutte le operazioni descritte, l'aspetto dei dati all'interno delle collezioni di MongoDB è quello illustrato in Figura.F in Appendice.

3. Pulizia e filtraggio dei dati

I dati collezionati sono stati in seguito filtrati e ripuliti.

Per il dataset covid, sono state scelte solo alcune delle nazioni disponibili. La scelta è stata guidata da due fattori: sono state prese in considerazione le nazioni colpite in modo significativo dal virus e sono stati tralasciati invece i Paesi in cui si sono registrati pochi casi di positività al virus nel periodo considerato; inoltre abbiamo deciso di considerare nazioni con un importante peso dell'indice di borsa nello scenario mondiale. Queste decisioni ci hanno portato a considerare le seguenti nazioni: Italia, Francia, Spagna, Germania, Regno Unito, Russia, USA, Brasile, Argentina, Canada, India, Giappone, Cina, Corea **.

Per alcune nazioni non si sono considerate le colonie (per esempio le Antille per la Francia), perché di poco impatto. Per i paesi per cui invece il conteggio dei positivi è stato effettuato e riportato per singola regione/contea (per esempio la Cina), abbiamo considerato il totale dei positivi dei come somma positivi per ogni regione/contea.

Per quanto riguardo gli indici di borsa le informazioni a disposizione erano molteplici, ma abbiamo scelto di considerare il valore di chiusura di ogni indice; a partire da tale valore abbiamo aggiunto una nuova colonna, riportante le variazioni percentuali, utili per un eventuale confronto tra indici diversi e preferibili nella realizzazione di visualizzazioni. Una volta terminate le operazioni di pulizia dei dati, i due dataset finali sono stati fusi attraverso la funzione *merge* sul campo

riguardante la data, così da ottenere un unico dataset contenente tutte le informazioni di nostro interesse; l'operazione di fusione tra i due dataset è stata necessaria per proseguire verso il nostro obiettivo di ricerca di una qualche correlazione esistente tra i due fenomeni studiati.

4. Data Visualitation

A partire dai dati utilizzati per la parte di data management, abbiamo scelto di sviluppare tre visualizzazioni interattive.

L'obiettivo delle visualizzazioni proposte è rappresentare le relazioni che intercorrono tra i fenomeni studiati e osservare come queste si modifichino nel tempo.

Disponendo di serie storiche, abbiamo privilegiato l'utilizzo di grafici che mostrassero l'andamento nel tempo dei fenomeni osservati, ovvero le linee temporali.

visualizzazione Nella prima (Figura.G, Appendice) abbiamo rappresentato l'evoluzione dei casi di positività al virus Covid-19, per le Nazioni studiate³, in relazione al tasso di variazione degli indici di borsa di ciascun paese. La scala utilizzata per gli indici è stata così costruita: per ogni Paese, gli indici sono stati dapprima normalizzati rispetto al valore del primo indice disponibile (in data 22 Gennaio 2020), e successivamente è stata calcolata la variazione percentuale di ogni valore giornaliero rispetto al 22 Gennaio. Tale scelta è stata dovuta al fatto che, in assenza di normalizzazione, il tasso di variazione comportava dei valori falsati e fuorvianti, poiché non rispecchiavano il reale andamento dell'indice di borsa. Per i dati relativi al virus abbiamo scelto di rappresentare i valori cumulati nel tempo, in linea con la nostra fonte di dati⁴.

Ad un primo sguardo è possibile avere una visione generale, per ogni continente, dell'andamento delle variabili nel tempo e poi, su richiesta e per mezzo della finestra di scelta, si può filtrare e fare uno zoom su un Paese tra quelli studiati. La rappresentazione è arricchita da una breve storia, inserita con lo scopo di accompagnare il lettore verso una comprensione completa del grafico.

Come si può notare, emerge una certa relazione negativa tra i fenomeni.

La seconda rappresentazione (Figura.H Appendice) consiste in un'animazione che riproduce l'andamento della variazione percentuale giornaliera di oro e argento in confronto alla serie storica dei casi Covid. Anche in questo caso i prezzi di oro e argento sono stati normalizzati rispetto al 23 Gennaio 2020. Si osserva una relazione inversa: l'aumento dei casi Covid corrisponde in genere ad una diminuzione del prezzo di tali beni di rifugio, specialmente per l'argento che mostra un crollo consistente. Nella stessa infografica, spostandosi lungo le storie, è possibile osservare l'andamento del prezzo del petrolio Brent e WTI sempre a confronto con i dati del contagio. È evidente che la performance del WTI risulti la peggiore tra i due prezzi, mentre il Brent, dopo un calo drastico, presenta un trend crescente.

L'ultima visualizzazione (Figura.I. Appendice) proposta consiste in una mappa coropletica: l'area dei Paesi viene caratterizzata da un colore, la cui intensità varia in funzione della forza della correlazione tra numero di casi Covid e valore di chiusura dell'indice di borsa. Il colore viene usato in modo tale che un Paese venga associato ad una categoria di

correlazione piuttosto che ad un'altra. A lato è visibile la scala di riferimento, utile a capire a quale categoria appartengono gli stati. La scala assume valori da -1 a 1, in coerenza con l'intervallo di valori dell'indice di Pearson. Inoltre, sotto alla legenda, è visibile un tasto d'informazione in cui è presente una breve spiegazione per una corretta comprensione della visualizzazione. In generale emerge che la maggior parte dei Paesi, ad eccezione dell'Argentina, mostra una correlazione negativa più o meno forte: non è un risultato inaspettato, in quanto le misure adottate dai diversi stati hanno rallentato l'economia dei Paesi. Spostandosi inoltre sui diversi Paesi è possibile visualizzare il nome del Paese e il valore della correlazione, che è stato anche stampato su ogni Stato.

Le infografiche sono state inserite in un questionario psicometrico, somministrato ad un campione di 25 utenti, per valutare alcune dimensioni della qualità dell'interazione degli utenti con la visualizzazione. Infine, abbiamo raccolti i dati delle risposte e analizzati sia per estrarre informazioni di sintesi per la valutazione della qualità, sia per capire quali fossero opportunità future di miglioramento. Abbiamo riassunto i risultati dell'analisi in tre diverse rappresentazioni (Figure L, M, N.Appendice): violin plot e correlogramma. Abbiamo scelto di utilizzare il violin plot per mostrare le risposte alle singole dimensioni di qualità e la distribuzione dei dati e il correlogramma per mettere in evidenza la correlazione tra le risposte date.

Conclusioni

In conclusione, il nostro studio ci ha permesso di comprendere più a fondo le dinamiche economiche che hanno caratterizzato i primi sei mesi di diffusione del Covid-19. In particolare, abbiamo visto quali dei paesi più rilevanti in termini economici sono riusciti a contenere meglio i danni finanziari derivanti dal blocco guasi totale della rispettiva economia domestica. Abbiamo potuto apprezzare l'entità della crisi della domanda nel campo dei carburanti fossili, che hanno fatto registrare perdite di portata storica. Abbiamo visto come l'oro si confermi ancora una volta il bene-rifugio per eccellenza, essendo l'unica commodity a guadagnare in valore rispetto a gennaio. Infine, la mappa coropletica dell'intensità delle correlazioni tra indici di borsa e numero di contagi per paese, ci ha permesso di verificare che nella maggior parte dei casi i due fenomeni sono negativamente correlati, unica eccezione è risultata l'Argentina, per la quale la correlazione risulta debolmente positiva (0,15).

Il feedback ricevuto attraverso il questionario psicometrico, inoltre, ci ha permesso di comprendere le criticità individuate da chi ha usufruito delle nostre visualizzazioni. In particolare, abbiamo visto che quella maggiormente apprezzata è stata la prima, ovvero quella relativa alle performance degli indici di borsa. Tuttavia, anche le altre due sembrano essere state apprezzate. Per migliorare il nostro studio, ci proponiamo di aumentare la quantità di informazioni mostrate e rendere più fruibile le visualizzazioni anche per il pubblico meno esperto in materia economica attraverso maggiore interattività nei grafici.

Appendice

ain /usr/local/etc/kafka/zookeeper.properties

```
er.ZooKeeperServer)
[2020-08-18 16:28:06,545] INFO Server environment:java.library.path=/Users/enricoconte/Library/Java/Extensions:/Library/Java/Extensions:/Network/Library/Java/Extensions:/System/Library/Java/Extensions:/usr/lib/java:. (org.apache.zookeepe
er.ZookeeperServer)
[2020-88-13 il-28:86,565] INFO Server environment;java.library.path=/Users/enricoconte/Library/Java/Extensions:/Library/Java/Extensions:/Network/Library/Java/Extensions:/System/Library/Java/Extensions:/usr/lib/java:. (org.apache.zookeepe perServer)
[2020-88-13 il-28:86,565] INFO Server environment;java.compiler=valx-(org.apache.zookeeper.server.ZookeeperServer)
[2020-88-13 il-28:86,565] INFO Server environment;java.compiler=valx-(org.apache.zookeeperServer)
[2020-88-13 il-28:86,565] INFO Server environment:sar.ans=makeo CS X (org.apache.zookeeper.server.ZookeeperServer)
[2020-88-13 il-28:86,565] INFO Server environment:sor.arch=v86.64 (org.apache.zookeeper.server.ZookeeperServer)
[2020-88-13 il-28:86,565] INFO Server environment:sor.arch=v86.65 (org.apache.zookeeper.server.ZookeeperServer)
[2020-88-13 il-28:86,565] INFO Server environment:sor.arch=v86.65 (org.apache.zookeeper.server.ZookeeperServer)
[2020-88-13 il-28:86,565] INFO Server environment:sor.arch=v86.65 (org.apache.zookeeper.server.ZookeeperS
```

Figura. A- Inizializzazione di Kafka Zookeeper

```
Last Login: Tow Ang 18 1118197 on trysp80.

Last Login: Last Login: Last Login: Last Login: Log
                                                                                                                                                                                   ocal/etc/kafka/zookeeper.properties ...a.Kafka /usr/local/etc/kafka/server.properties
broker.id = 0
broker.id.generation.enable = true
                                                                                                                                                                                                                                                                                                                               Figura.B - Inizializzazione Kafka Local Server
```

...a.Kafka /usr/local/etc/kafka/server.properties

.oker-list localhost:9092 --topic

.ain /usr/local/etc/kafka/zookeeper.properties

- jupyter-notebook ► python

```
Last login: Tue Aug 18 16:31:29 on ttys003
(base) enricoconte@MBP-di-Enrico ~ % kafka-console-producer --broker-list localhost:9092 --topic covid
>^C%
(base) enricoconte@MBP-di-Enrico ~ % kafka-console-producer --broker-list localhost:9092 --topic alphavantage
>test
                                                                Figura.C - Inizializzazione Producer
```

~ — jupyter-notebook ▶ python ...ain /usr/local/etc/kafka/zookeeper.properties ...a.Kafka /usr/local/etc/kafka/server.properties Last login: Tue Aug 18 16:34:52 on ttys004 (base) enricoconte@MBP-di-Enrico ~ % kafka-console-consumer --bootstrap-server localhost:9092 --topic alphavantage --from-beginning

Figura.D - Inizializzazione Consumer

```
URL_alpha_ibex = "https://www.alphavantage.co/query?function=TIME_SERIES_DAILY&symbol="IBEX&apikey=FAESZBR0KJHGLIVB"
URL_alpha_cac = "https://www.alphavantage.co/query?function=TIME_SERIES_DAILY&symbol=^FCHI&apikey=FAESZBR0KJHGLIVB"
URL_alpha_merv = "https://www.alphavantage.co/query?function=TIME_SERIES_DAILY&symbol=^MERV&apikey=FAESZBROKJHGLIVB"
URL_covid = "https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_seri
ultima_esecuzione = datetime.datetime.now()-datetime.timedelta(days=1)
while True:
    adesso = datetime.datetime.now()
    if ultima_esecuzione.date()!= adesso.date() and adesso.time() > datetime.time(12):
         #TMPORT DATT
         r_alpha_ftse = requests.get(url = URL_alpha_ftse)
r_alpha_dax = requests.get(url = URL_alpha_dax)
         r_alpha_nikkei = requests.get(url = URL_alpha_nikkei)
         r_alpha_nya = requests.get(url = URL_alpha_nya)
         r_alpha_bsesn = requests.get(url = URL_alpha_bsesn)
         r_alpha_axjo = requests.get(url = URL_alpha_axjo)
         r_alpha_gsptse = requests.get(url = URL_alpha_gsptse)
         r_alpha_ibex = requests.get(url = URL_alpha_ibex)
r_alpha_cac = requests.get(url = URL_alpha_cac)
r_alpha_merv = requests.get(url = URL_alpha_merv)
         r_covid = requests.get(url = URL_covid)
         data alpha ftse = r alpha ftse.content
         data_alpha_dax = r_alpha_dax.content
data_alpha_nikkei = r_alpha_nikkei.content
         data_alpha_nya = r_alpha_nya.content
         data_alpha_bsesn = r_alpha_bsesn.content
         data_alpha_axjo = r_alpha_axjo.content
         data_alpha_gsptse = r_alpha_gsptse.content
         data_alpha_ibex = r_alpha_ibex.content
data_alpha_cac = r_alpha_cac.content
data_alpha_merv = r_alpha_merv.content
         data_covid = r_covid.content
         #PROCESSING ALPHAVANTAGE
         now = datetime.datetime.now()-datetime.timedelta(days=1)
         nome_colonna = now.strftime("%y-%m-%d")
         dati_grezzi_alpha_ftse = data_alpha_ftse.decode('utf-8')
         dati_grezzi_alpha_ftse = pd.DataFrame(data = dati_grezzi_alpha_ftse)
         #dati_alpha_dict_ftse = ast.literal_eval(dati_grezzi_alpha_ftse)
#dati_alpha_dict_ftse = str(dati_alpha_dict_ftse)
         print(dati_alpha_dict_ftse)
         dati grezzi alpha dax = data alpha dax.decode('utf-8')
```

Figura.E – Ciclo While per raccolta e Storage dei dati

```
| Column | C
```

Figura.F – Dati all'interno della Collection Alphavantage di MongoDB, pt.1

Figura.F – Dati all'interno della Collection Alphavantage di MongoDB, pt.2

```
| C. | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1.0 | | 1
```

Figura.G – Dati all'interno della Collection Covid di MongoDB.

Indici_Covid Mappa_Cor_Info Oil_Gold

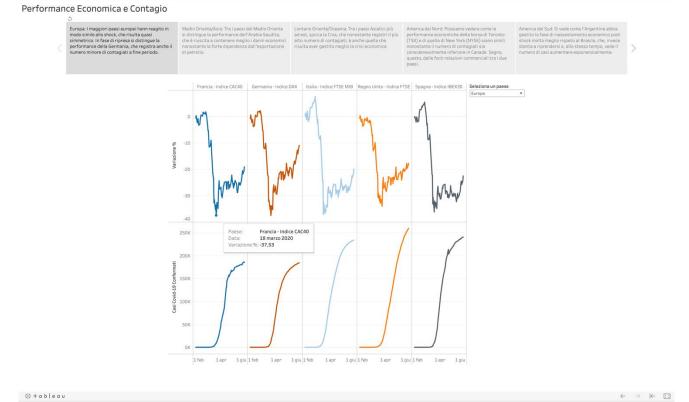


Figura G – Prima Visualizzazione

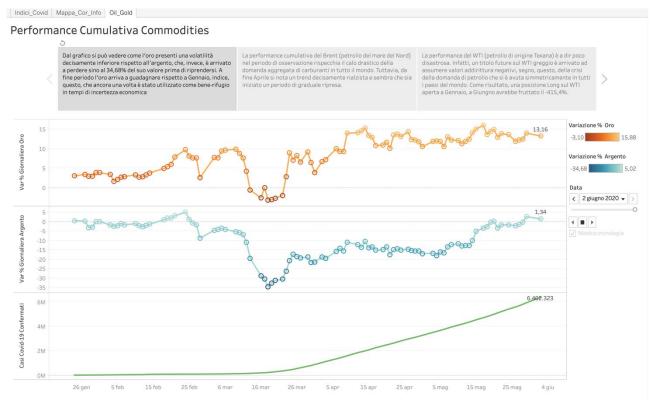


Figura H – Seconda Visualizzzione

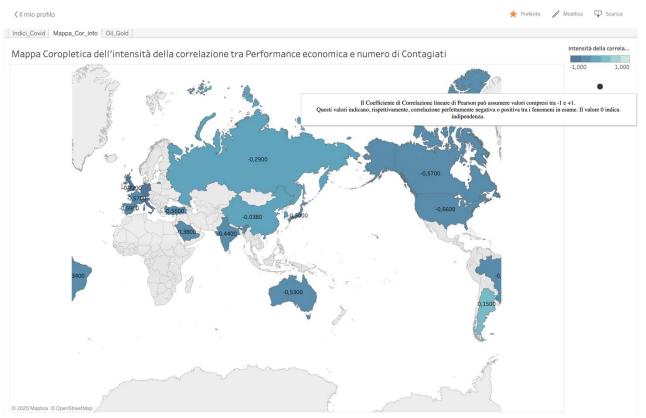


Figura I – Terza Visualizzazione

Risultati Diagnostica: Prima Visualizzazione

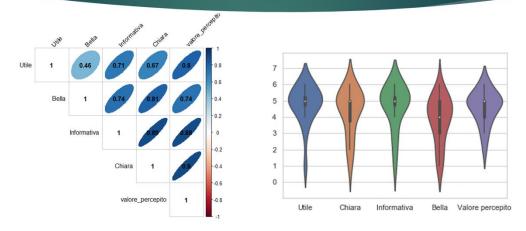


Figura.L - Diagnostica Prima Visualizzazione

sualizzazione

Risultati Diagnostica: Seconda Visualizzazione

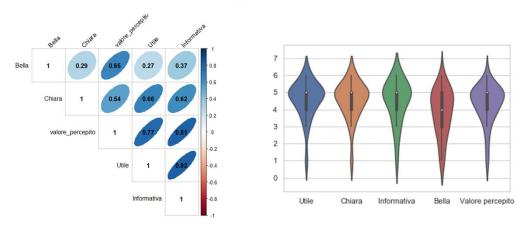


Figura. M – Diagnostica Seconda Visualizzazione

Risultati Diagnostica: Terza Visualizzazione

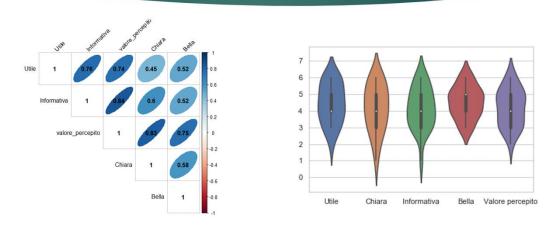


Figura.N -Diagnostica Terza Visualizzazione