

Airline On-Time Performance

- CdLM Data Science
- Riccardo Confalonieri 830404
- Rebecca Picarelli 834286
- Silvia Ranieri 878067



INDICE



INTRODUZIONE



DATA
MANAGEMENT



DATA
VISUALIZATION



CONCLUSIONI



INTRODUZIONE

- Negli Stati Uniti, a causa delle dimensioni del territorio, è noto che gli spostamenti aerei siano molto frequenti e talvolta proprio necessari.



- Analizzeremo ritardi registrati per i singoli voli, determinando i periodi e/o le compagnie migliori per viaggiare.

DATA MANAGEMENT

Ricerca dei dati

Gestione della varietà

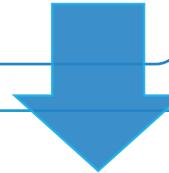
Conversione in json e salvataggio
su Mongo DB

Gestione del volume (5GB)

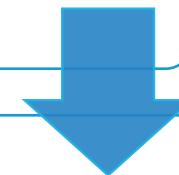


VARIETÁ

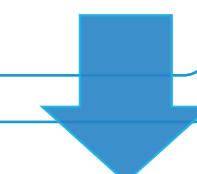
Fonte 1: BTS Dataset sui voli USA



Fonte 2: Dati sulla compagnia area Kaggle



Fonte 3 :Dati sugli aeroporti Kaggle



Fonte 4: Dati sui passeggeri ‘Airports Council International’



VOLUME



CSV



Gestione
valori null e
missing



Creazione dei
file json



Caricamento
su MongoDB



FILE JSON

```
{  
    "year": 2018,  
    "month": 1,  
    "day_of_week": 1,  
    "origin": "AVP",  
    "origin_info": {  
        "ICAO": "KAVP",  
        "name": "Wilkes Barre Scranton International Airport",  
        "type": "medium_airport",  
        "elevation_ft": 962,  
        "latitude": 41.33850097659995,  
        "longitude": -75.7233963013001,  
        "city_name": "Scranton/Wilkes-Barre",  
        "state_name": "Pennsylvania",  
        "state_code": "PA"  
    },  
    "destination": "EWR",  
    "destination_info": {  
        "ICAO": "KEWR",  
        "name": "Newark Liberty International Airport",  
        "type": "large_airport",  
        "elevation_ft": 18,  
        "latitude": 40.692501068115234,  
        "longitude": -74.168701171875,  
        "city_name": "Newark",  
        "state_name": "New Jersey",  
        "state_code": "NJ",  
        "num_passengers": [  
            {  
                "year": 2018,  
                "world_ranking": 40,  
                "total_passengers": 45851434,  
                "growth_percent": 6.082538564366158  
            }  
        ]  
    },  
    "distance": 93,  
    "distance_group": 1,  
    "flights": [ ]  
}
```

```
"flights": [  
    {  
        "day_of_month": 1,  
        "airline": "EV",  
        "airline_name": "Atlantic Southeast Airlines",  
        "crs_dep_time": 1825,  
        "dep_time": 2011,  
        "dep_delay": 106,  
        "dep_delay_group": 7,  
        "crs_arr_time": 1937,  
        "arr_time": 2203,  
        "arr_delay": 146,  
        "arr_delay_group": 9,  
        "cancelled": false,  
        "diverted": false,  
        "crs_elapsed_time": 72,  
        "actual_elapsed_time": 112,  
        "air_time": 35,  
        "carrier_delay": 39,  
        "weather_delay": 0,  
        "nas_delay": 40,  
        "security_delay": 0,  
        "late_aircraft_delay": 67  
    },  
    {  
        "day_of_month": 8,  
        "airline": "EV",  
        "airline_name": "Atlantic Southeast Airlines",  
        "crs_dep_time": 1810,  
        "crs_arr_time": 1920,  
        "cancelled": true,  
        "cancellation_code": "C",  
        "diverted": false,  
        "crs_elapsed_time": 70  
    }  
]
```

Figura 1: Esempio di documento json generato dallo script

ARCHITETTURA MONGO DB

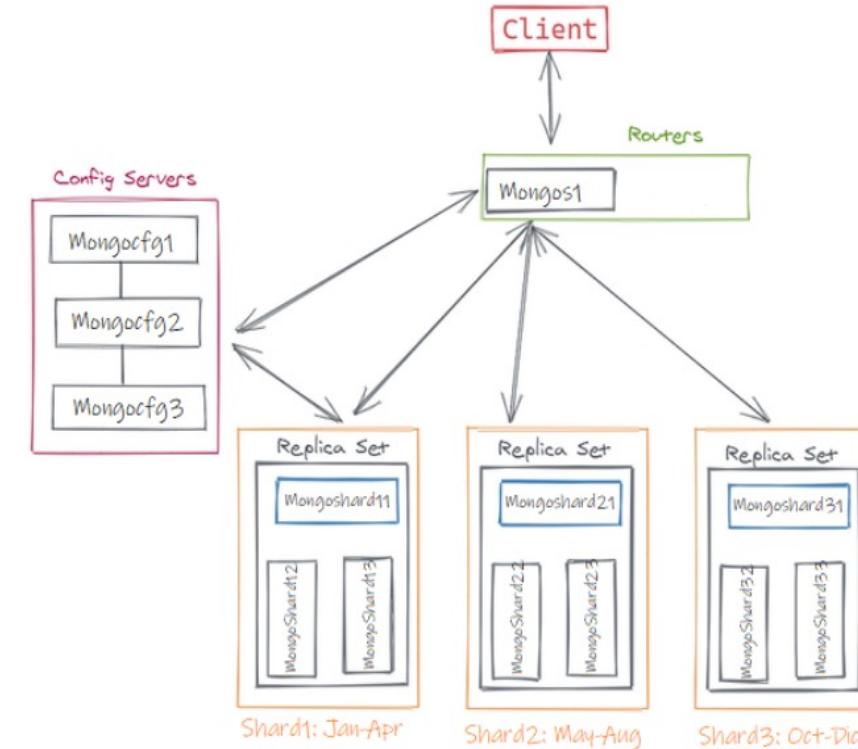
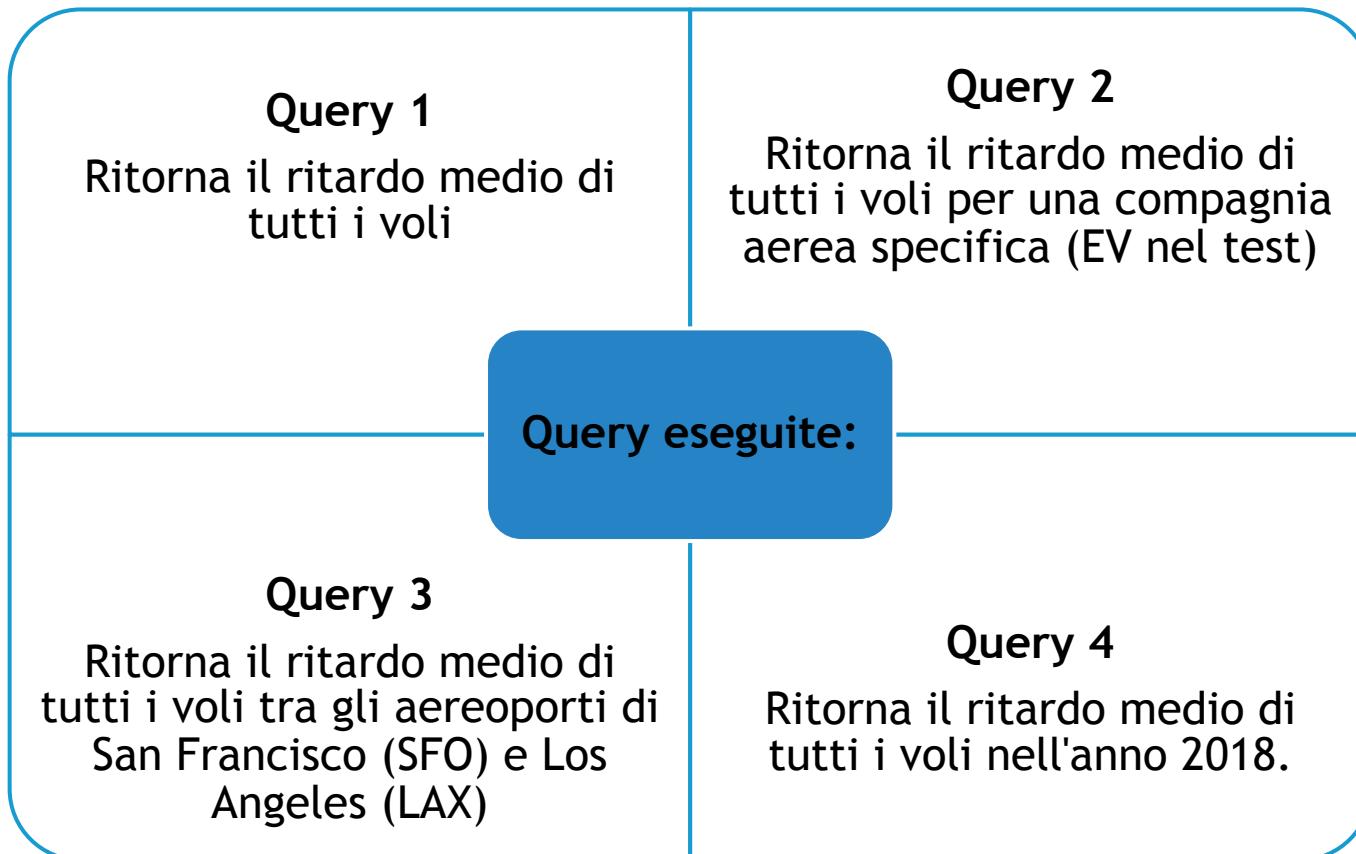
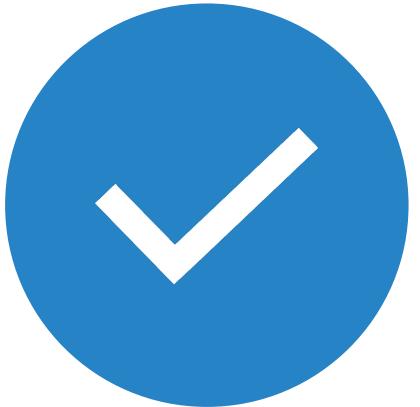


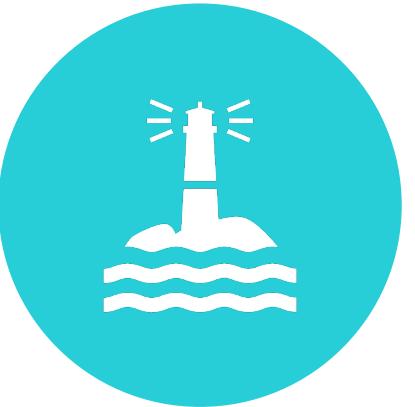
Figura 2: MongoDB architettura sharding

TESTING ARCHITETTURA





I RISULTATI DELLA STRATEGIA DI SHARDING
RISULTA EFFICIENTE E IL DATABASE RIESCE A
RISPONDERE BENE ALL'AUMENTO DI VOLUME
RICHIEDUTO.



TUTTAVIA ESTENDENDO GLI ANNI D'INTERESSE
POTREBBE ESSERE TROPPO STRINGENTE LIMITARE
LA SUDDIVISIONE A SOLI TRE SHARD ORGANIZZATI
NEL MODO PROPOSTO IN QUANTO OGNI SINGOLO
ANNO CORRISPONDE A CIRCA 2.5 GB.

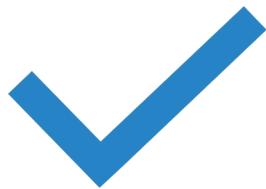


SOLUZIONE PROPOSTA AGGIUNGERE ULTERIORI
SHARD E MODIFICARE LA STRATEGIA DI
SHARDING.

OSSERVAZIONI



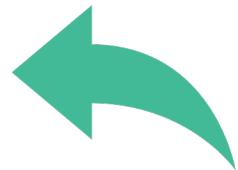
DATA VISUALIZATION



Le compagnie aeree sono
tutte uguali?



Esiste un periodo o una
destinazione migliore per
viaggiare?



L'andamento dei ritardi
rimane costante nei diversi
anni?

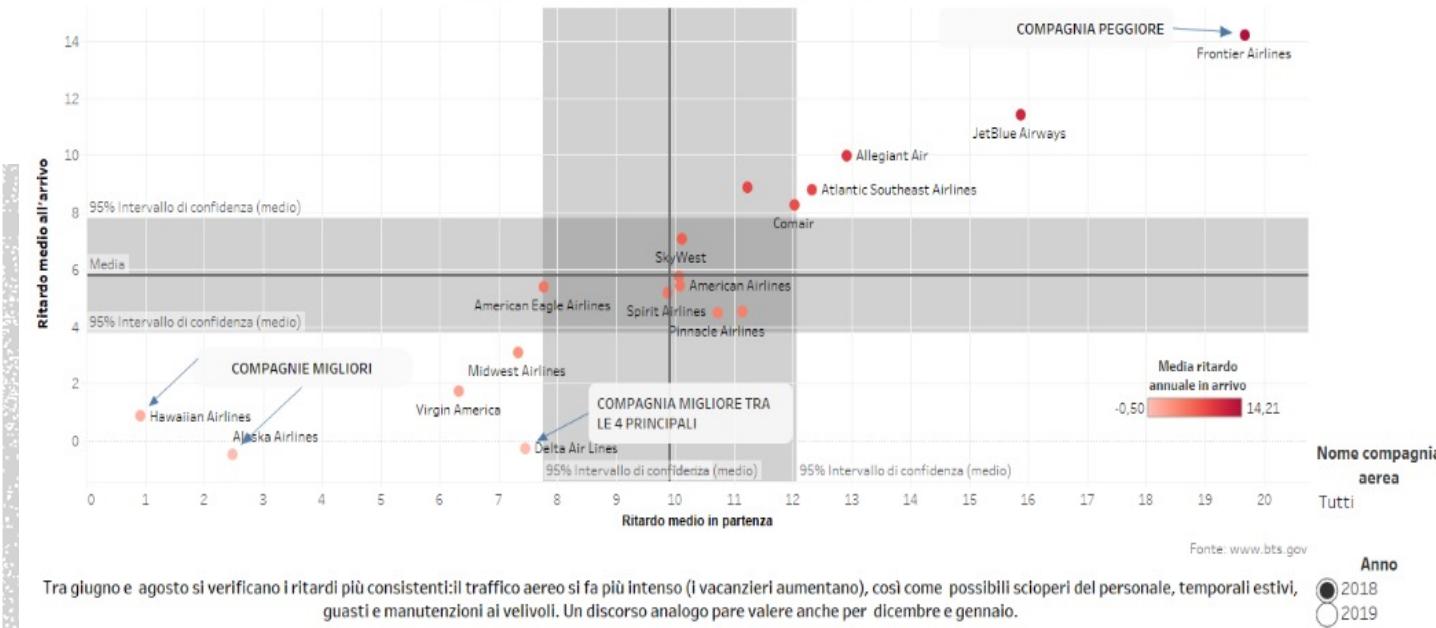


INFOGRAFICA

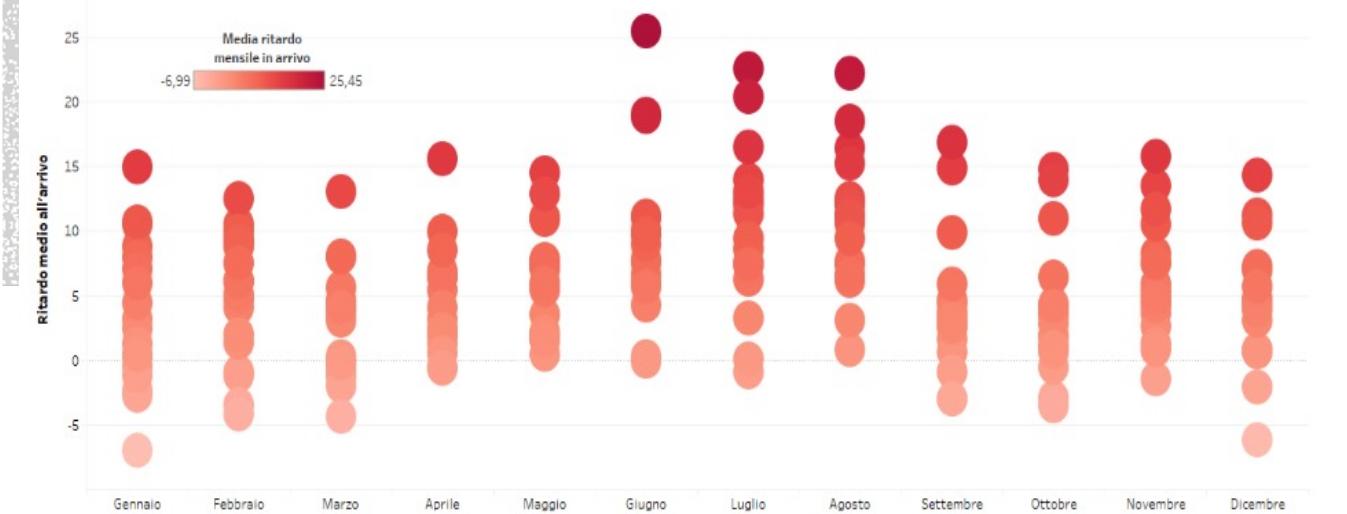
1

RITARDO MEDIO PER COMPAGNIA AEREA A LIVELLO ANNUALE E MENSILE

I ritardi medi in arrivo sono inferiori rispetto a quelli in partenza. Inoltre, i ritardi in partenza entro i 10 minuti non comportano ritardi in arrivo sopra alla media. Tra le quattro principali compagnie USA (Delta Airlines, American Airlines, United Airlines e Southwest Airlines) la Delta Airlines risulta la migliore, anche se in partenza fa registrare un ritardo medio pressoché equivalente alle altre.



Tra giugno e agosto si verificano i ritardi più consistenti: il traffico aereo si fa più intenso (i vacanzieri aumentano), così come possibili scioperi del personale, temporali estivi, guasti e manutenzioni ai velivoli. Un discorso analogo pare valere anche per dicembre e gennaio.

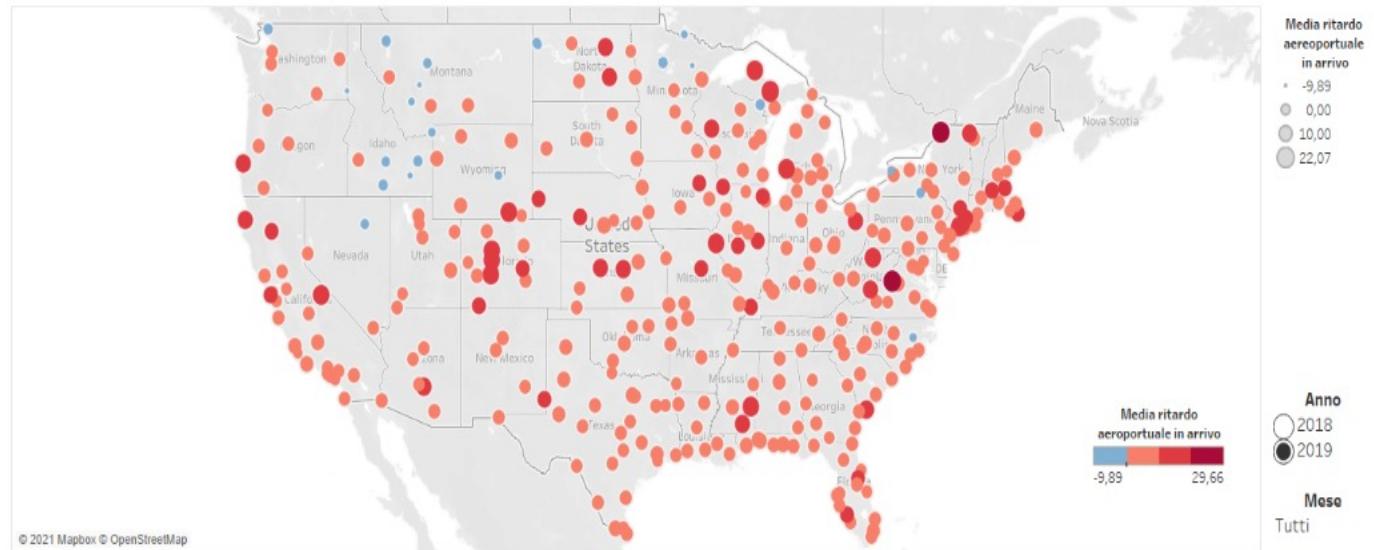


INFOGRAFICA

2

ANALISI DEI RITARDI A LIVELLO GEOGRAFICO E TEMPORALE

Le aree più densamente abitate (ovvero coste e parte orientale) tendono a mostrare un'elevata concentrazione di ritardi.



Fonte: www.bts.gov

In entrambi gli anni, marzo e settembre sono tra i mesi migliori per viaggiare, a differenza di quelli estivi chiaraente poco consigliabili, come anche il lunedì, giovedì e venerdì.
Mentre il sabato pare il giorno con le performance più soddisfacenti in assoluto.

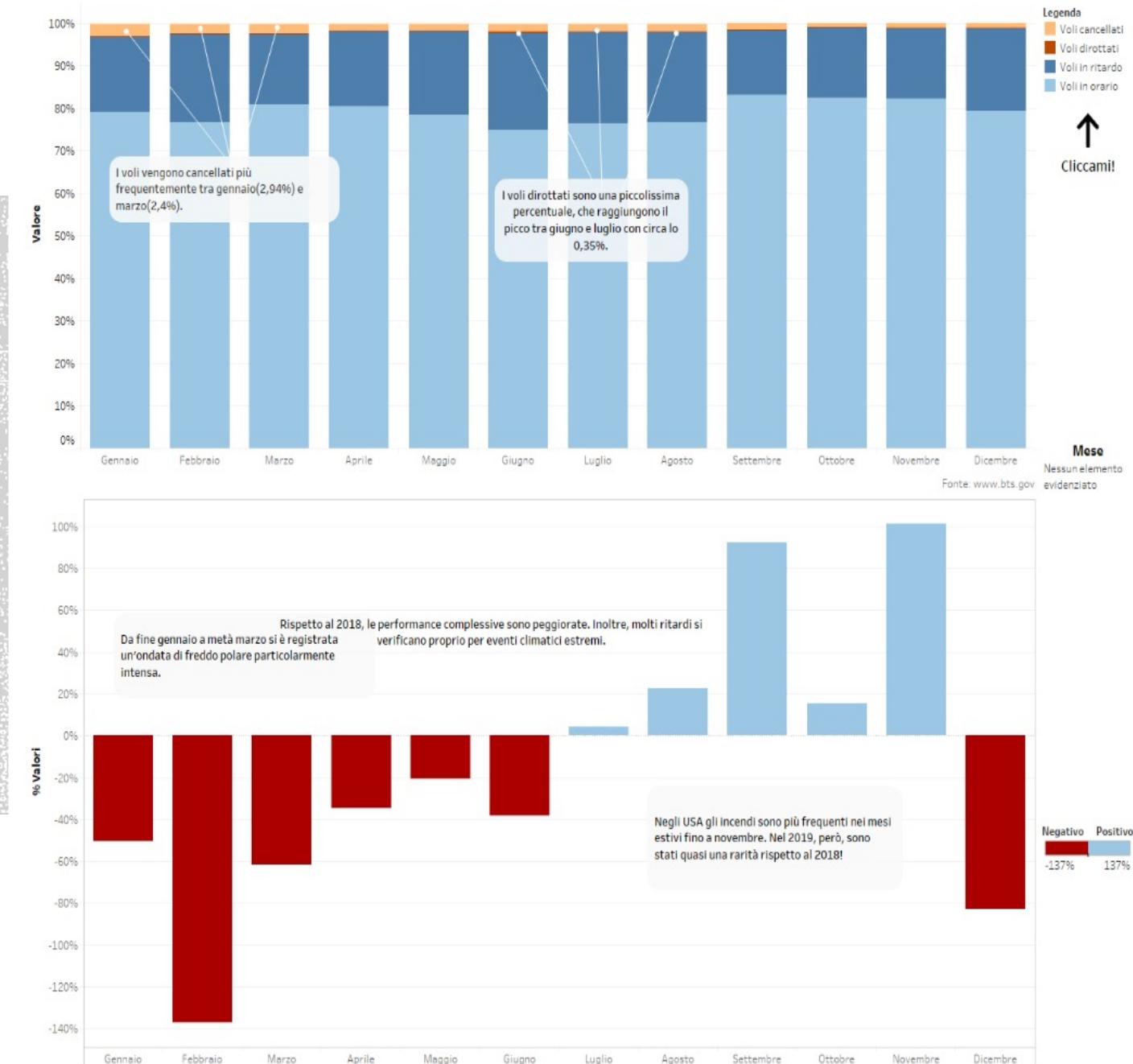
Nella tabella è evidenziato il valore massimo e minimo rilevato.



INFOGRAFICA

3

ANALISI PERFORMANCE VOLI 2019



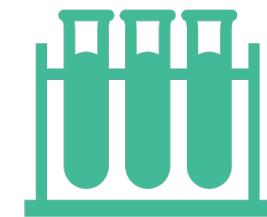
QUESTIONARI E TEST



VALUTAZIONE
EURISTICA



QUESTIONARIO
PSICOMETRICO



TEST
DELL'UTENTE



ESEMPIO

INFOGRAFICA 2

VALUTAZIONE EURISTICA

- La prima critica riguarda la scelta dei colori, infatti gli utenti avrebbero preferito la coppia Rosso-Verde per identificare peggioramenti e miglioramenti nelle performance, purtroppo quest'accoppiata di colori non è stato possibile utilizzarla per permettere anche ai daltonici di consultare la visualizzazione senza problemi
- L'altra critica riguarda la matrice colori che, secondo alcuni utenti, è risultata troppo `piena" e soprattutto senza margini tra le varie celle, quindi abbiamo modificato ottenendo il seguente risultato

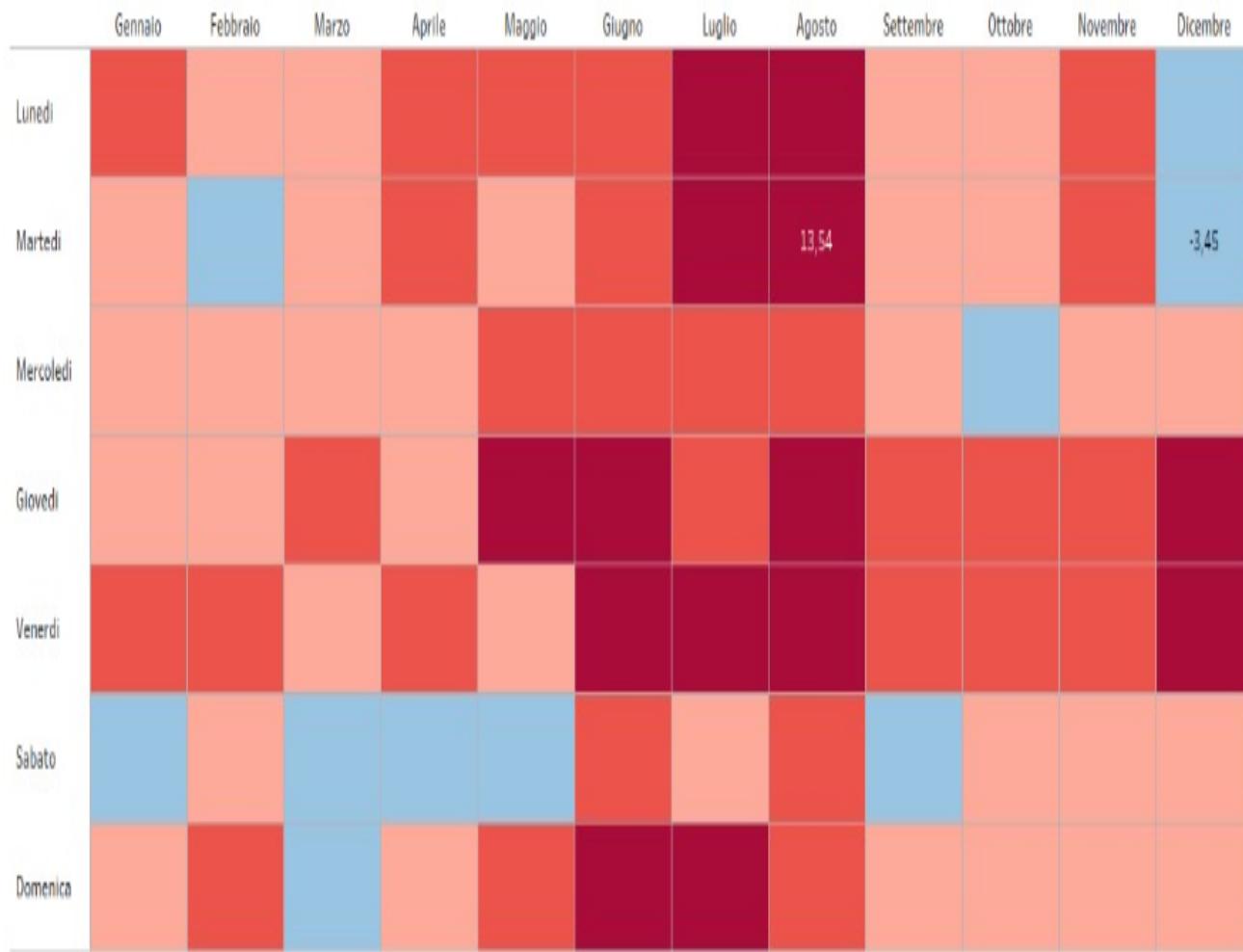


Figura 7: Tabella colori dashboard2 revisionata



ESEMPIO INFOGRAFICA 2

QUESTIONARIO PSICOMETRICO

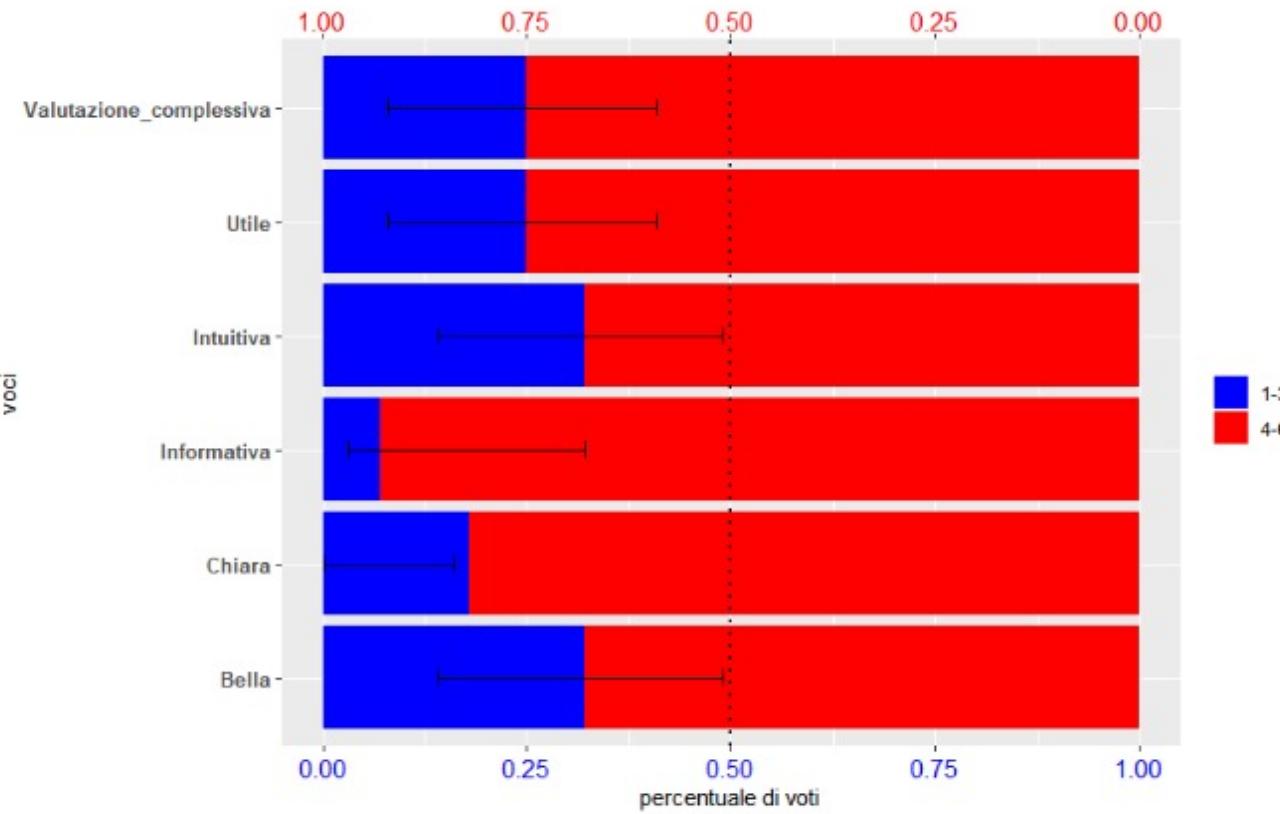


Figura 10: Stacked bar chart Dashboard 2

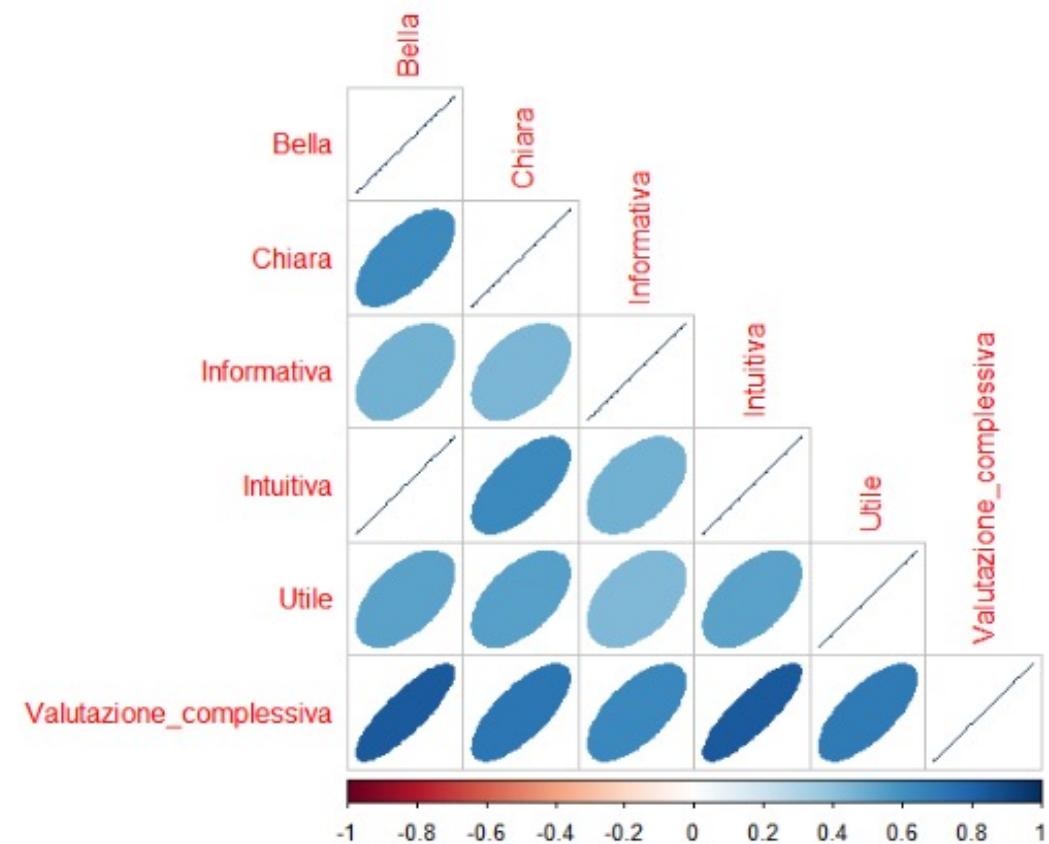


Figura 11: Correlogramma Dashboard 2



ESEMPIO INFOGRAFICA 2

- Per questa dashboard sono stati proposti due task:
- Quale ritieni che sia, per il mese di Marzo 2019, il giorno migliore per viaggiare?
- Secondo te, escludendo lo stato della California, in quale giorno è stato registrato il minor ritardo nell'anno 2019?

TEST UTENTE

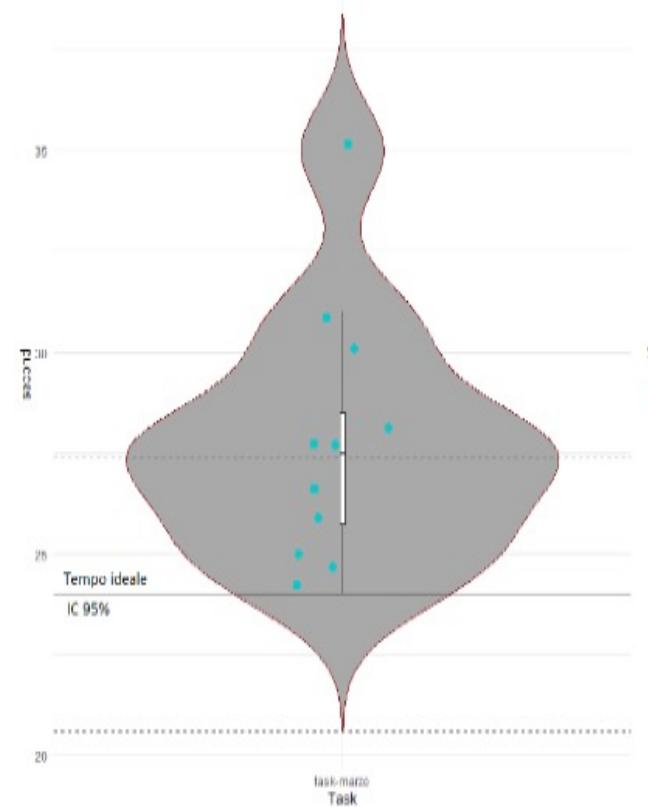


Figura 16: Task1 seconda dashboard

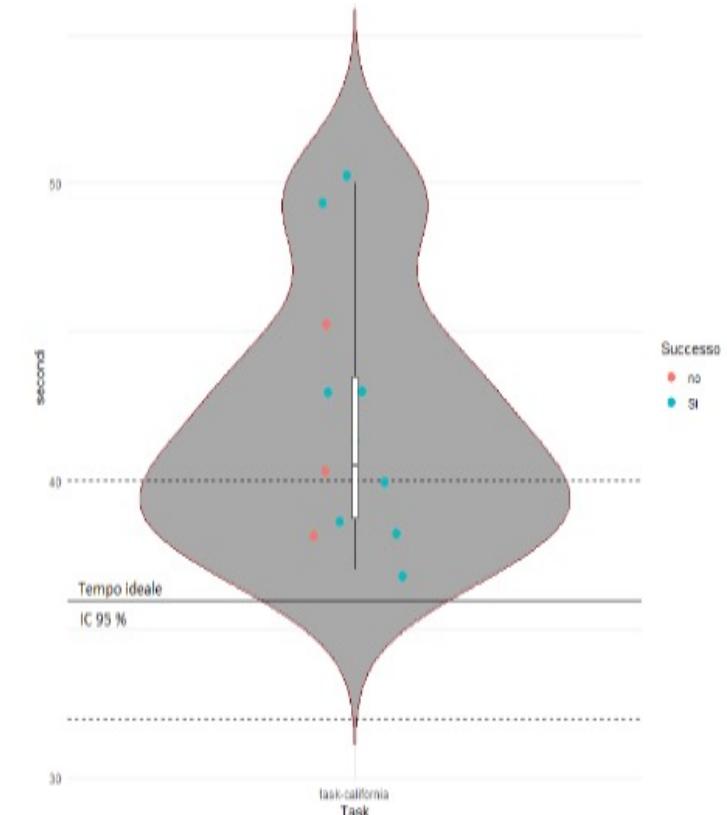


Figura 17: Task2 seconda dashboard





CONCLUSIONI

- **Infografica 1:** alcune compagnie aeree più performanti di altre in termini di ritardo negli anni, tra queste, però, non figurano le 4 compagnie statunitensi
- **Infografica 2:** si evince che i mesi estivi, ma anche per i periodi festivi di dicembre e gennaio o giorni come il lunedì, sono caratterizzati generalmente dai ritardi più consistenti
- **Infografica 3:** I voli cancellati, sono sempre molto meno rari di quelli dirottati, si concentrano tendenzialmente nei mesi invernali climaticamente sfavorevoli
- **Valutazioni e task svolti dagli utenti,** è emerso che le infografiche sono:
 - I. informative
 - II. chiare
 - III. non particolarmente belle
 - IV. Poco intuitive
 - V. la valutazione complessiva discreta

**GRAZIE
DELL'ATTENZIONE**

