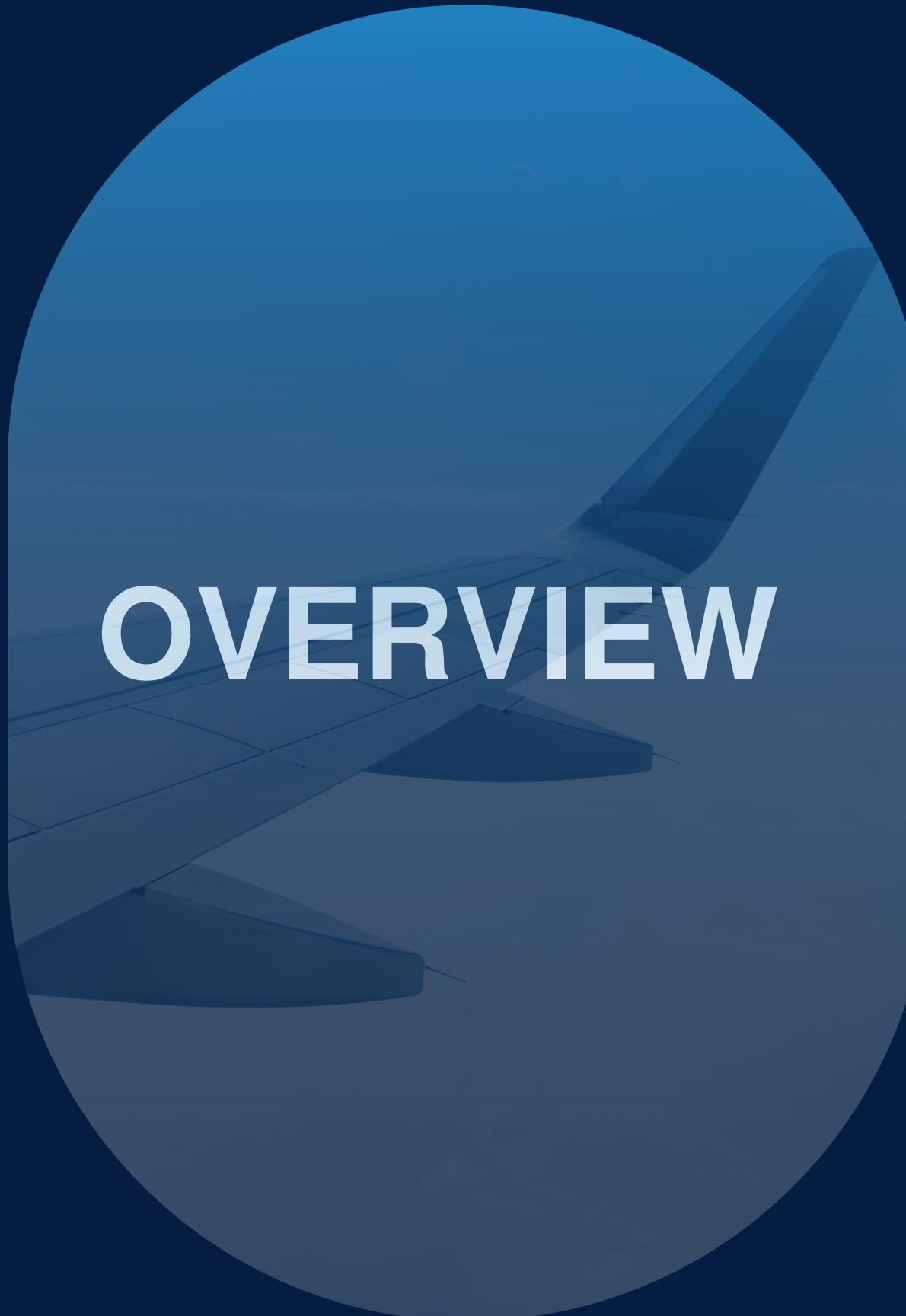


CARRIER ON-TIME PERFORMANCE PROJECT

Presentato da:
Macrì Domenico
Zappia Giuseppe





OVERVIEW

01

Il Dataset

02

Letteratura

03

Tecnologie
Utilizzate

04

Obiettivi

05

L'Applicazione

06

Sfide Principali

01 IL DATASET



DATASET UTILIZZATO

Informazioni dettagliate su oltre 6 milioni di voli effettuati negli Stati Uniti nel 2013

- 6.369.482 righe
- 110 colonne (dati voli USA, 2013)
- Informazioni su:
 - Ritardi, cancellazioni, dirottamenti
 - Origini, destinazioni, stati/città coinvolti



PRE-PROCESSING

Dati completi e pronti per l'elaborazione, minimo pre-processing necessario. Analisi preliminare basata verifiche di consistenza:

- Controllo righe duplicate
- Identificazione delle Anomalie nei Dati:
 - Distanze e tempo di volo non positivi
 - Disallineamenti tra colonne correlate

02 LETTERATURA

01

Ricerca di paper scientifici in merito allo stesso dataset o simili

02

Analisi dei paper selezionati per ricavare informazioni su dati e query principali

03

Adattamento e sviluppo di nuove interrogazioni per l'analisi dei dati

04

Indagini principali:

- Classificazione dei ritardi in base alle cause
- Relazione tra coppie di aeroporti per capire come i ritardi si propagano attraverso la rete
- Modelli stagionali e pattern giornalieri nell'occorrenza dei ritardi



03 TECNOLOGIE UTILIZZATE

Back-end

Apache Spark tramite PySpark: gestione di grandi volumi di dati

- Manipolazione efficiente dei dati (pyspark.sql)
- Operazioni di machine learning (pyspark.ml)

Front-end

Implementazione tramite Streamlit, framework Python open-source.

Vantaggi principali:

- Interfaccia intuitiva e interattiva.
- Visualizzazione dei dati semplificata con widgets.
- Caricamento rapido grazie al caching.



04

OBIETTIVI



ANALISI AVANZATA DEI DATI

Sfruttare il dataset attraverso interrogazioni mirate e significative nell'ambito della data analysis e supporto ad analisi predittive e applicazioni di machine learning



PERFORMANCE E OTTIMIZZAZIONE

Sviluppare interrogazioni e interfaccia web-based efficienti, garantendo tempi di risposta ridotti



INTERAZIONE PERSONALIZZATA

Progettare un'interfaccia utente che consenta un'interazione flessibile e personalizzabile con le interrogazioni tramite modulazione dei parametri di ricerca.



ATTENZIONE ALLA VISUALIZZAZIONE

Promuovere un approccio visivo ai dati tramite rappresentazioni grafiche avanzate, incluse mappe geografiche interattive e progress bar

05 L'APPLICAZIONE

H
O
M
E
P
A
G
E

Dashboard Voli - Panoramica

Anteprima Dataset

Voli totali

6369482

Aereo con più miglia percorse

N751AT

Giorno della settimana con più voli:

Venerdì - 952033

Percentuali

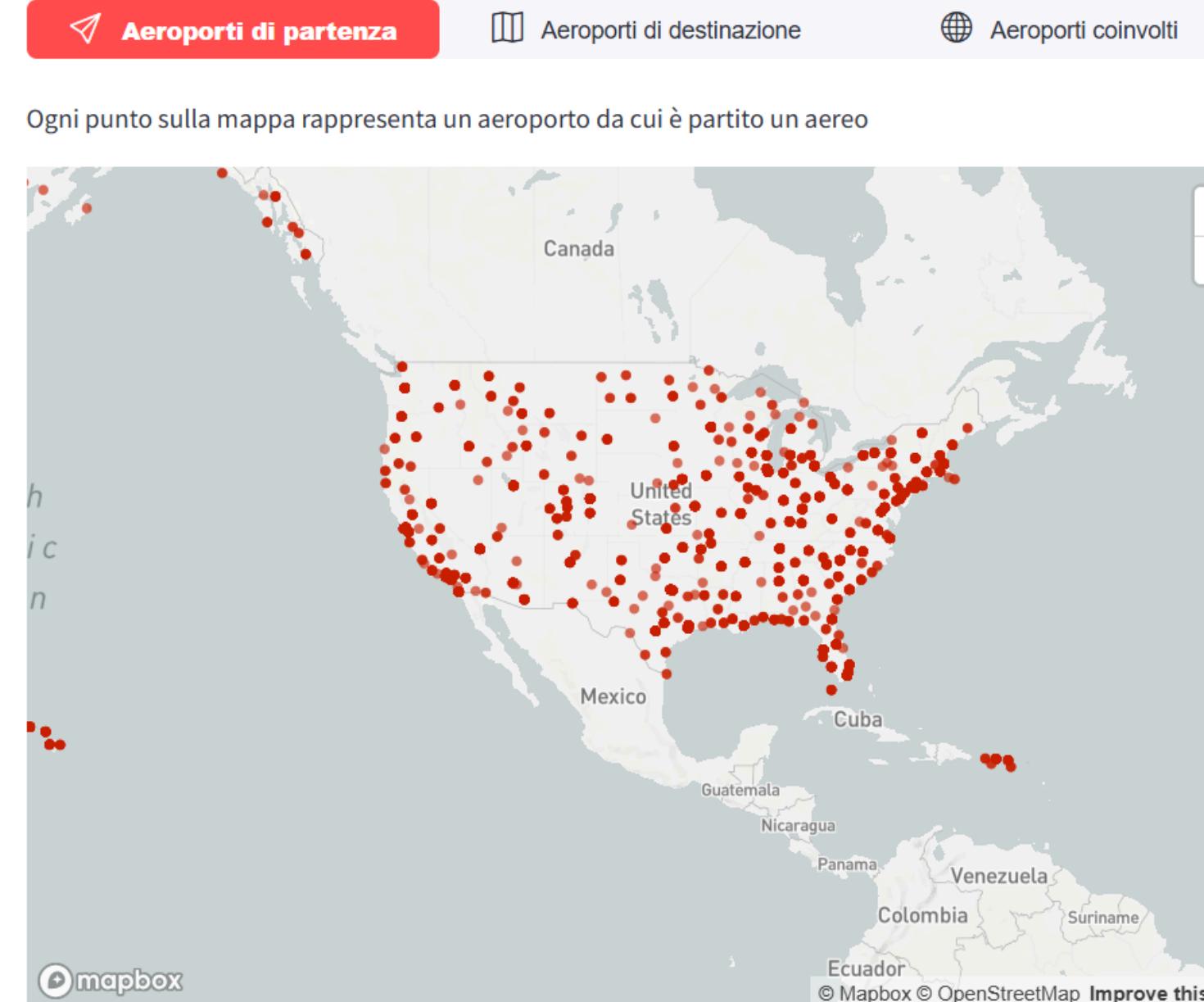
In orario



In ritardo



Mappa



Grafici

Principali compagnie per voli fatti

Stati più puntuali

Stati più visitati

Stato	Voli atterrati
California	755926
Texas	745515
Florida	443169
Georgia	416251
Illinois	414827
New York	289544
Colorado	250555
North Carolina	221079
Arizona	201509
Michigan	199349

Top Rotte Aeree

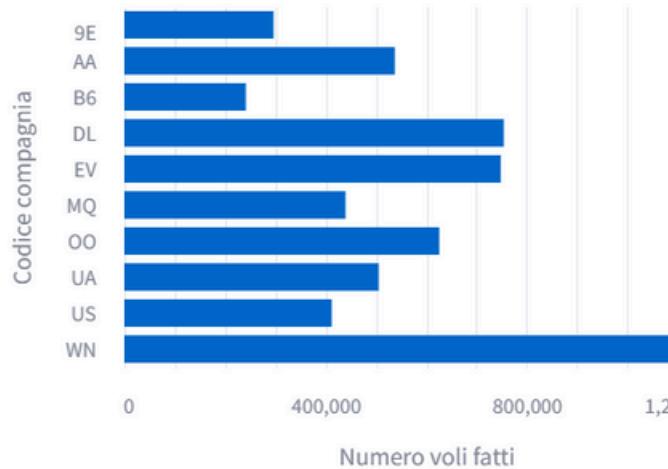
Partenza	Arrivo	NumeroVoli
SFO	LAX	1583
LAX	SFO	1579
OGG	HNL	1180
HNL	OGG	1174
LAX	LAS	1156
LAS	LAX	1152
LAX	JFK	1126

05 L'APPLICAZIONE

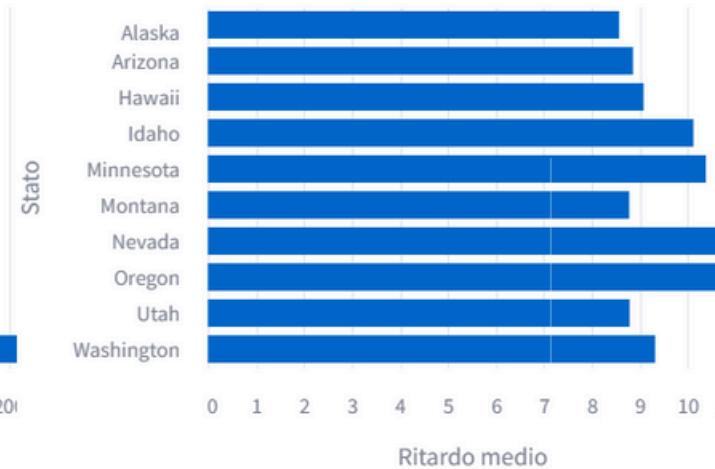
40.04 %

Grafici

Principali compagnie per voli fatti



Stati più puntuali

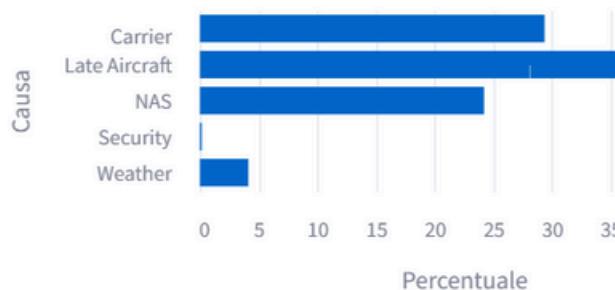


HNL	OGG	1174
LAX	LAS	1156
LAS	LAX	1152
LAX	JFK	1126
JFK	LAX	1126
LGA	ATL	1026
ATL	LGA	1025

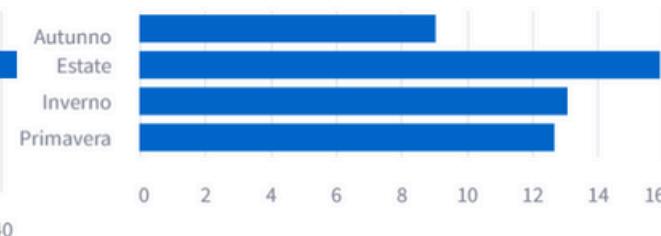
Compagnie con più miglia percorse

Compagnia	Miglia percorse
WN	782392393mi
DL	66344872mi
UA	656997085mi
AA	568996515mi
US	361588400mi
EV	357814587mi
OO	288530971mi
B6	258125810mi
MQ	218186501mi
AS	182564626mi

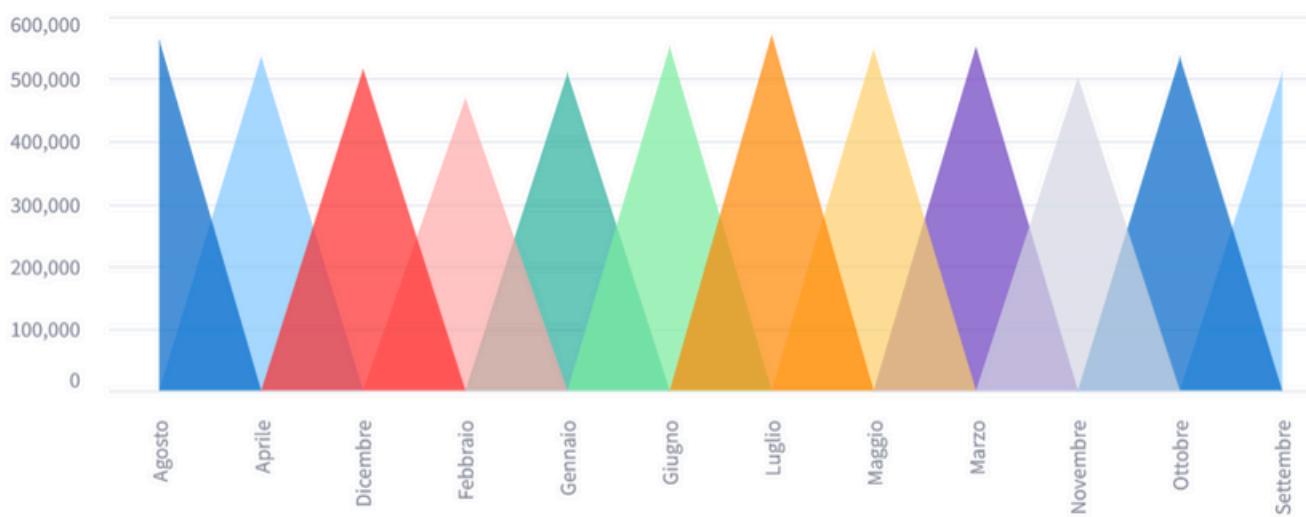
Principali cause ritardo



Ritardi medi per stagione



Numero di voli per mese



● Agosto ● Aprile ● Dicembre ● Febbraio ● Gennaio ● Giugno ● Luglio ● Maggio ● Marzo ● Nove

H
O
M
E
P
A
G
E

05 L'APPLICAZIONE

A
E
R
O
P
O
R
T
I

Analisi dei dati relativi ai singoli aeroporti

Selezione un aeroporto

Info Aeroporto

Città dell'aeroporto
New York, NY

Voli totali passati per l'aeroporto
222628

Giorno della settimana con più voli:
Venerdì - 32375

Percentuali

In orario
 **59.35 %**

In ritardo
 **40.65 %**

Numero di voli per mese

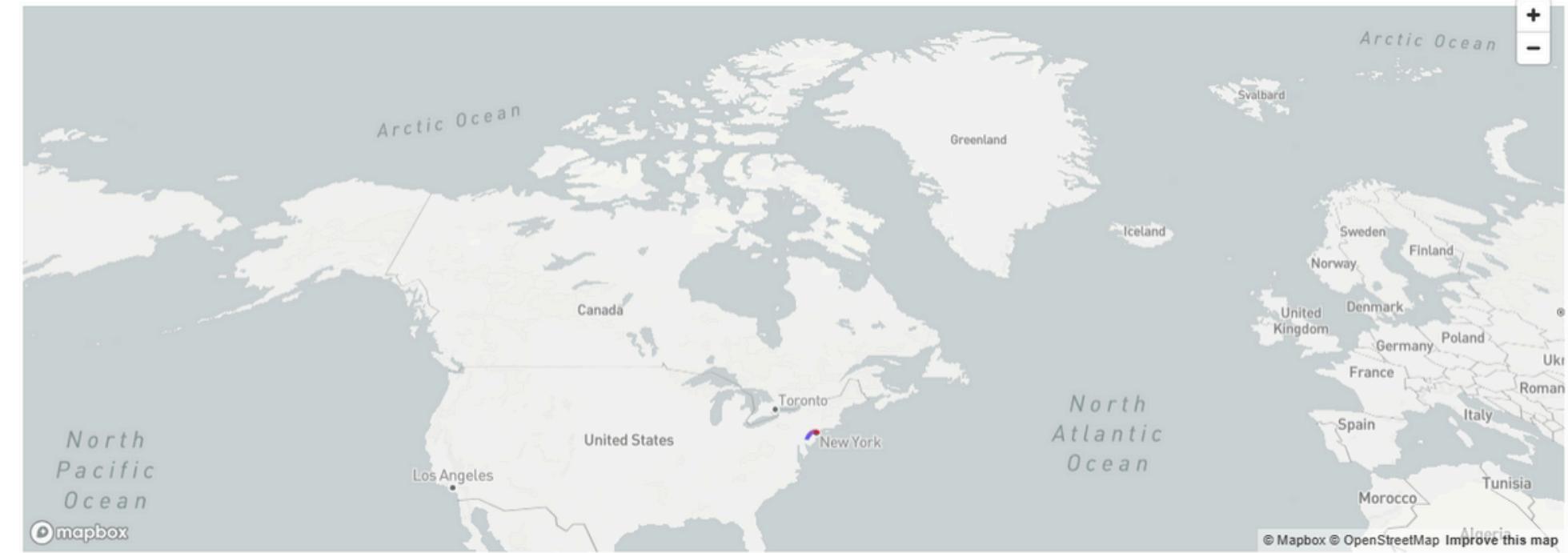
Mese	Voli translati
Gennaio	18333
Febbraio	16834
Marzo	19394
Aprile	18445
Maggio	18790
Giugno	18934
Luglio	20046

Analisi Aeroporto

Direzione
Da **Verso**

 Volo più distante  Volo più vicino  Tratta più percorsa

Tratta con distanza min partita da **New York, NY (JFK)** e arrivata a **Philadelphia, PA (PHL)** percorrendo 94.0 miglia

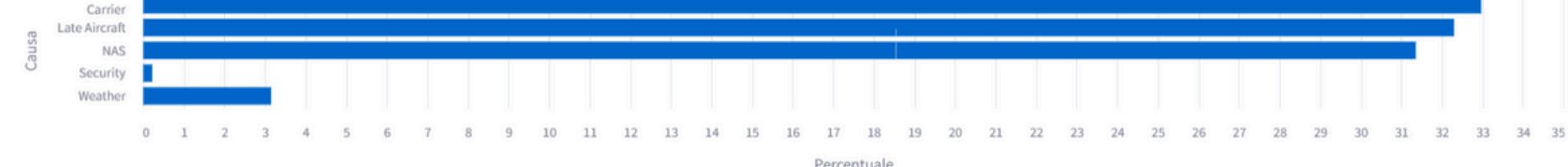


Grafici

Principali cause ritardo

Selezione il periodo
01-01-2013 **31-12-2013**

Causa
Carrier
Late Aircraft
NAS
Security
Weather



Ritardi al decollo aeroporto

Il grafico mostra la distribuzione dei ritardi al decollo per l'aeroporto. È possibile scegliere se visualizzare i ritardi mensili o giornalieri ed eventualmente specificare il periodo di interesse. Inoltre a sinistra è possibile selezionare uno o più aeroporti differenti con cui confrontare i ritardi.



Operatività nei giorni più importanti dell'anno

Voli partiti

Voli atterrati

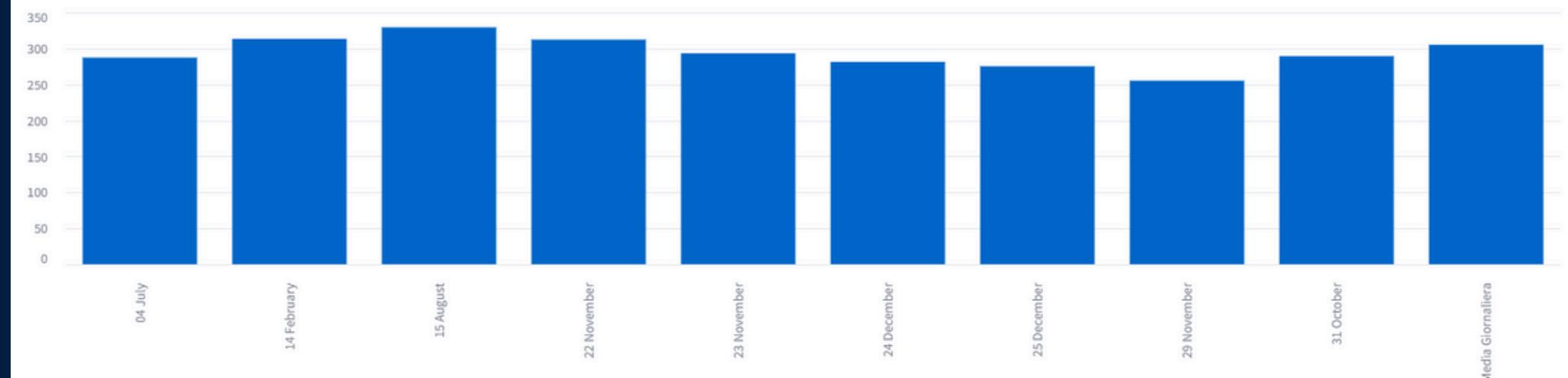
Voli totali

Percentuale Cancellazioni

Stato per confronto

Choose an option

Il grafico mostra il numero di voli partiti dall'aeroporto nelle date principali dell'anno. L'ultima barra rappresenta la media giornaliera dei voli per ogni aeroporto, permettendo un confronto immediato tra l'operatività normale e quella nei giorni speciali.



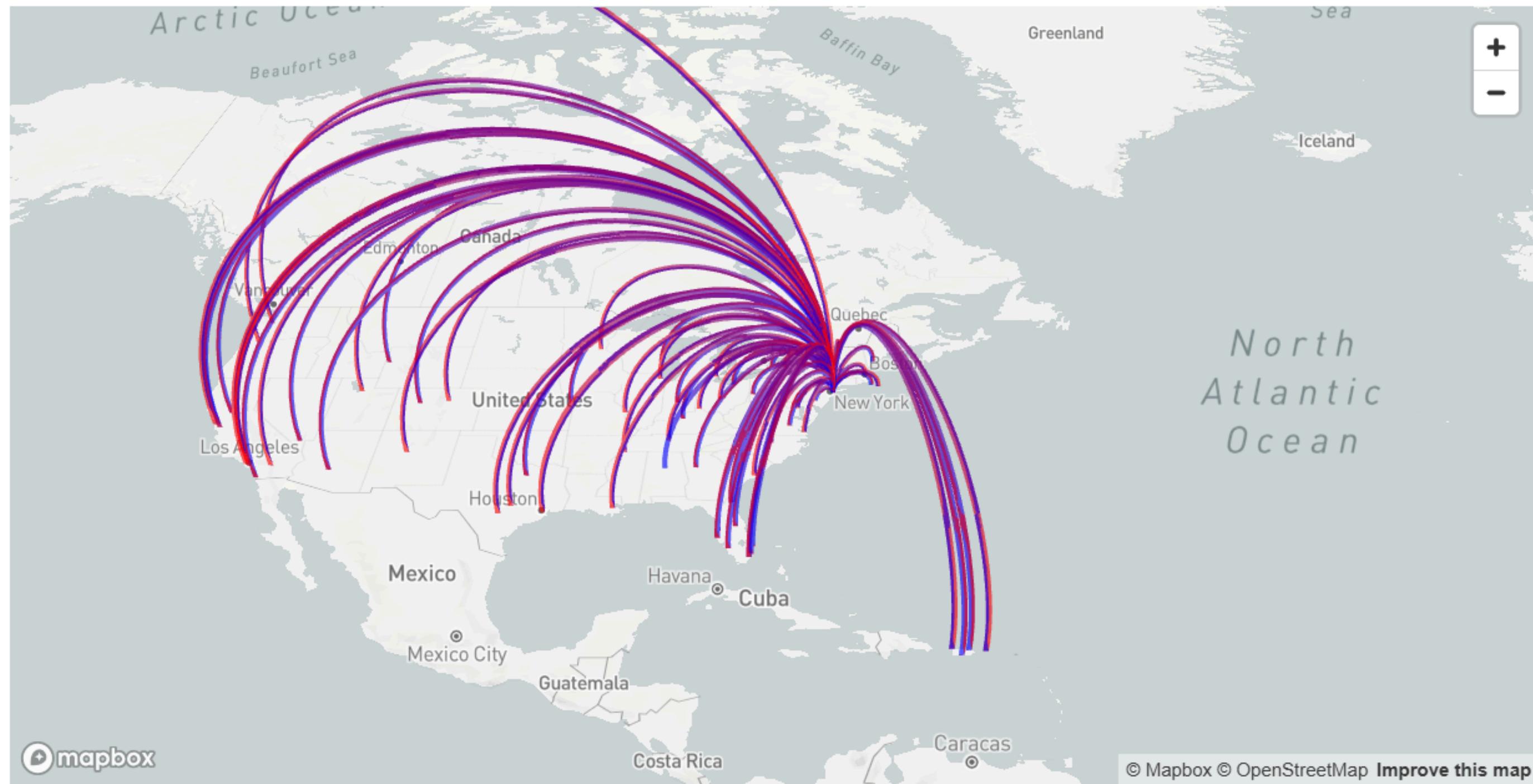
A
E
R
O
P
O
R
T

05 L'APPLICAZIONE

A
E
R
O
P
O
R
T
I

Aeroporti collegati

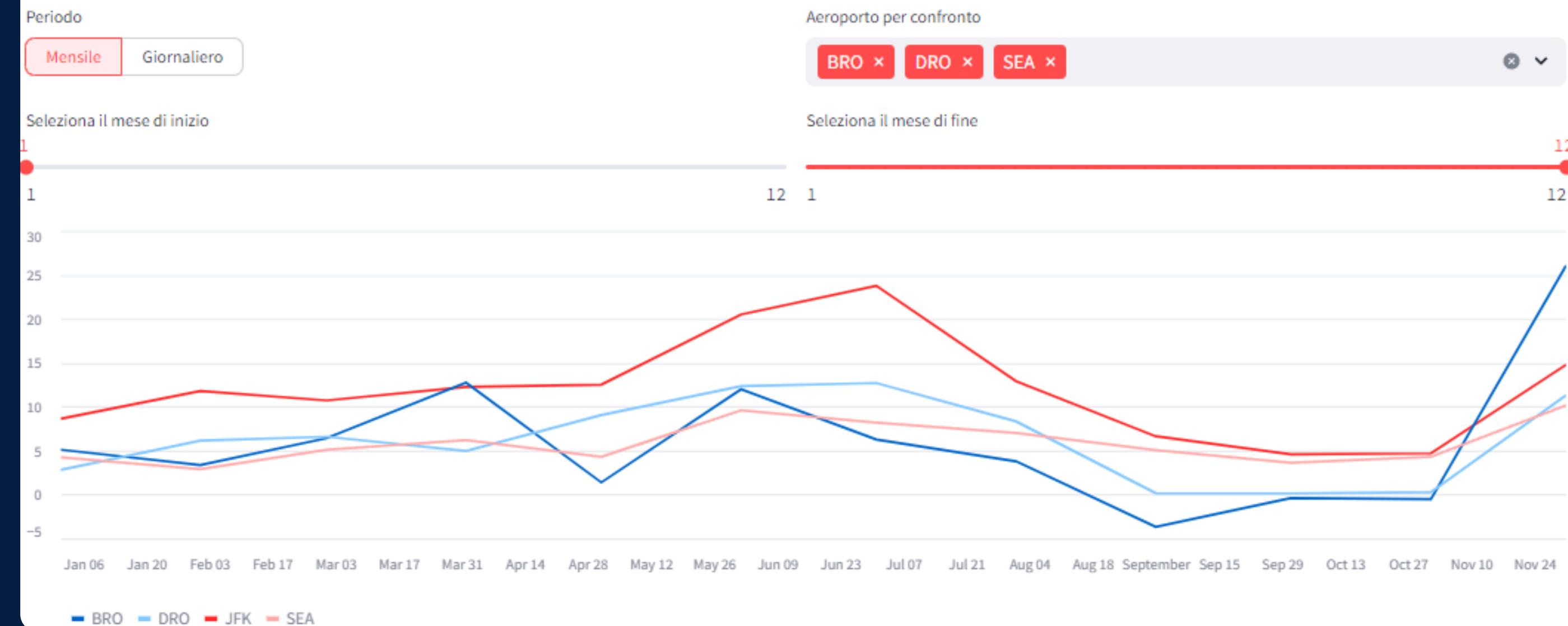
Il grafico mostra le rotte aeree che collegano l'aeroporto selezionato ad altri aeroporti



05 L'APPLICAZIONE

Ritardi al decollo aeroporto

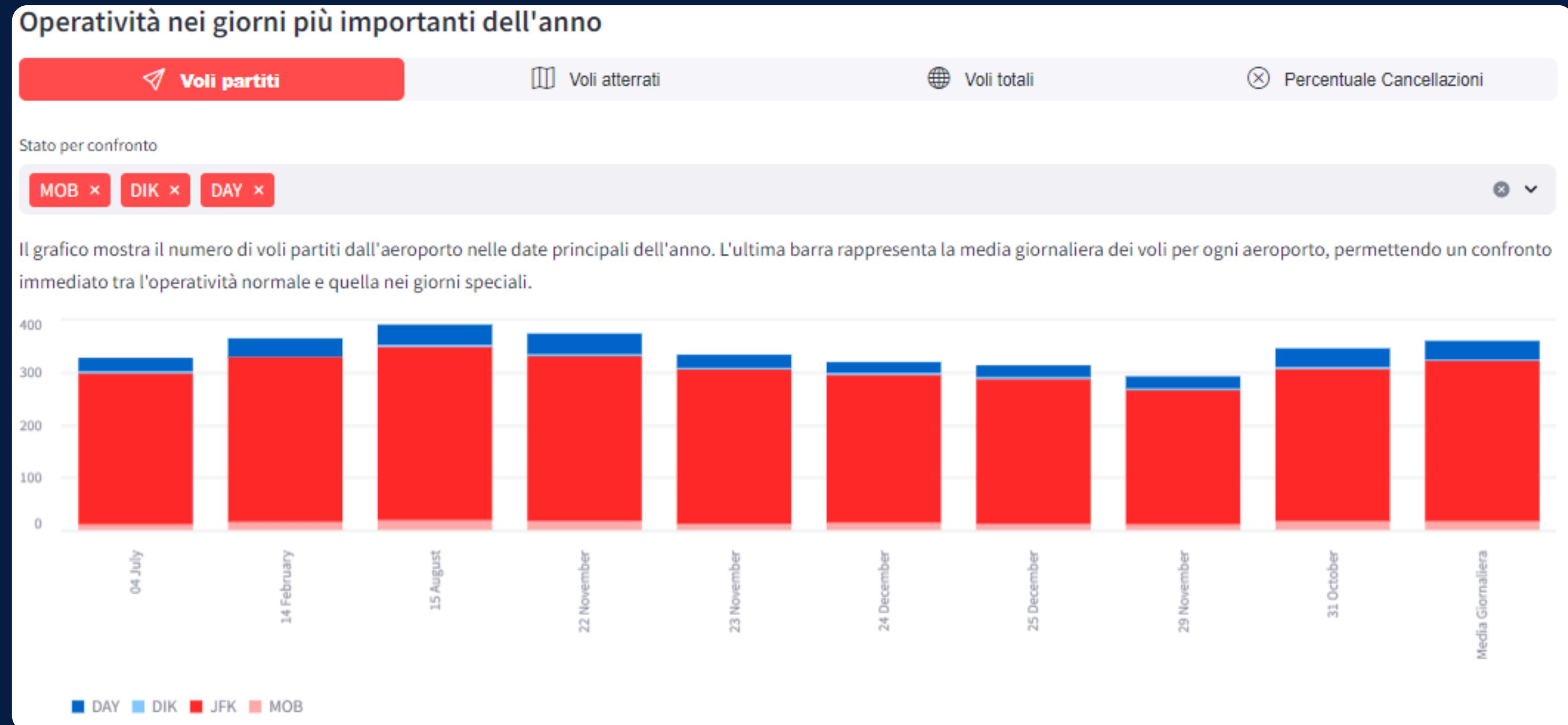
Il grafico mostra la distribuzione dei ritardi al decollo per l'aeroporto. È possibile scegliere se visualizzare i ritardi mensili o giornalieri ed eventualmente specificare il periodo di interesse. Inoltre a sinistra è possibile selezionare uno o più aeroporti differenti con cui confrontare i ritardi.



INTERATTIVITÀ

05 L'APPLICAZIONE

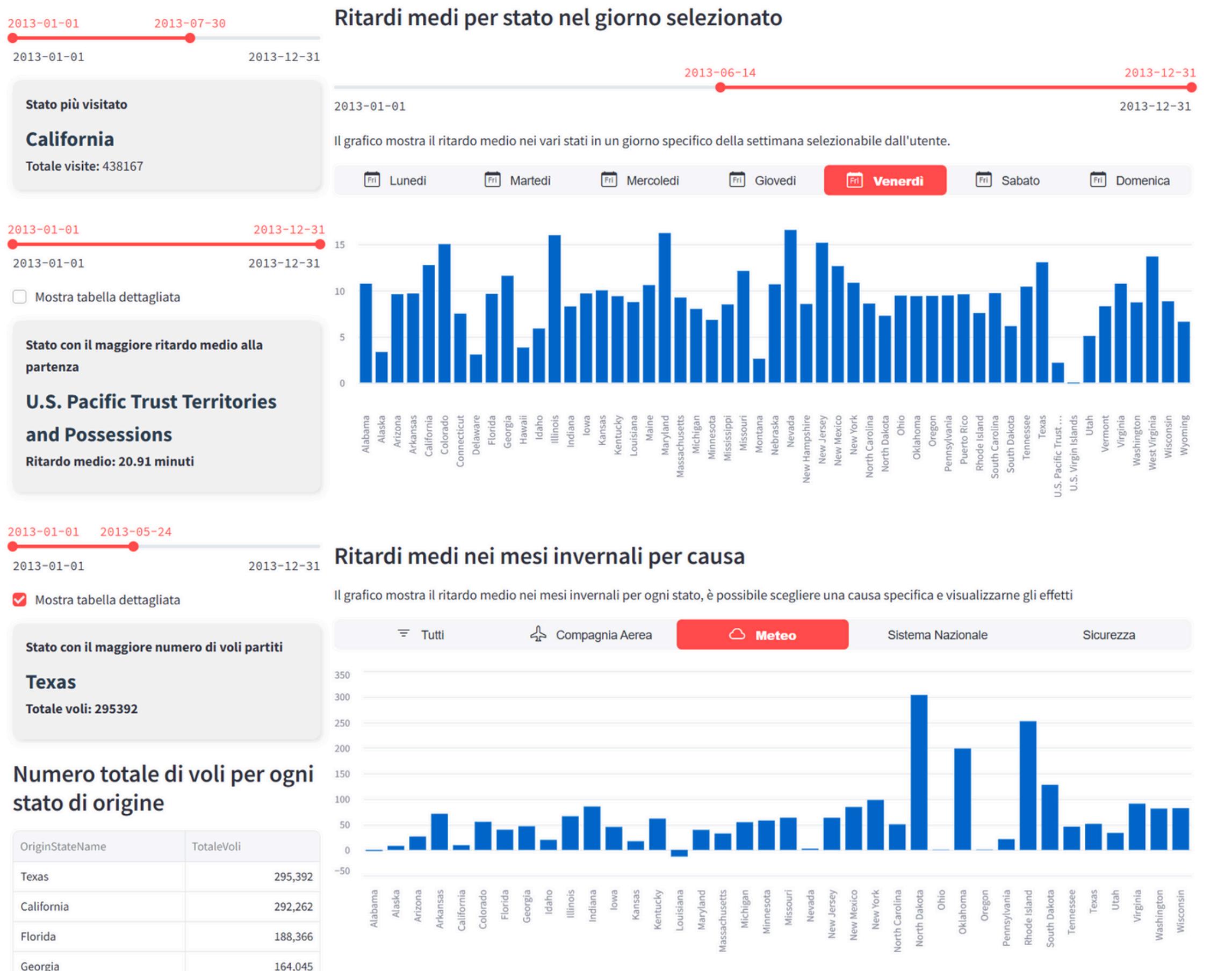
A
E
R
O
P
O
R
T
I



INTERATTIVITÀ

05 L'APPLICAZIONE

Analisi dei dati relativi ai paesi



P
A
E
S
I

05 L'APPLICAZIONE

P
A
E
S
I

North Carolina	83,644
Arizona	81,235
Michigan	76,738

2013-01-01 2013-08-05

2013-01-01 2013-12-31

Selezione uno stato per visualizzare il rapporto specifico (opzionale)

Tutti

Rapporto cancellazioni per tutti gli stati

Tabella sotto

Rapporto tra voli cancellati e voli totali per ogni stato

OriginStateName	TotalVol	VoliCancellati
North Dakota	11,130	410
South Dakota	7,070	244
West Virginia	2,198	74
New Jersey	73,648	2,410
Iowa	15,142	482
Vermont	3,507	111
Illinois	248,902	7,711
Wyoming	6,371	192
Kansas	8,070	241
New York	172,029	4,858

2013-01-01 2013-12-31

2013-01-01 2013-12-31

Selezione uno stato per visualizzare i tempi di volo (opzionale)

Tutti

Grafici

- Minor Ritardo Medio
- Stati più Visitati
- Maggiore Traffico Aereo
- Incremento Ritardi invernali rispetto estate in minuti

OriginStateName	RitardoMedio
Hawaii	1.6902
Alaska	1.8156
Montana	3.2677
U.S. Virgin Islands	3.8559
Delaware	4.1066
Utah	4.532
Idaho	4.8949
Wyoming	5.6938
Minnesota	6.5509
Washington	6.6316

DestStateName	TotalVol
California	439,211
Texas	425,653
Illinois	241,253
Florida	240,562
Georgia	238,521
New York	167,488
Colorado	143,988
North Carolina	130,232
Michigan	116,204
Washington	113,723
Arizona	113,723

OriginStateName	ChilometriTotali
California	402,973,778
Texas	297,927,080
Florida	209,486,914
Illinois	173,552,787
New York	154,201,211
Georgia	145,852,030
Colorado	118,863,359
Arizona	96,507,646
Washington	87,204,373
Virginia	82,580,343

OriginStateName	IncrementoRitardi
North Dakota	25.618
Delaware	19.6861
South Dakota	13.3346
Wyoming	11.2694
Montana	10.728
Minnesota	10.6627
Kansas	10.4746
Utah	9.5637
Wisconsin	9.5558
Iowa	9.4136

Analisi delle percentuali di cause di cancellazioni

Il grafico mostra la percentuale delle cause di cancellazione, è possibile scegliere un periodo specifico e visualizzarne le percentuali

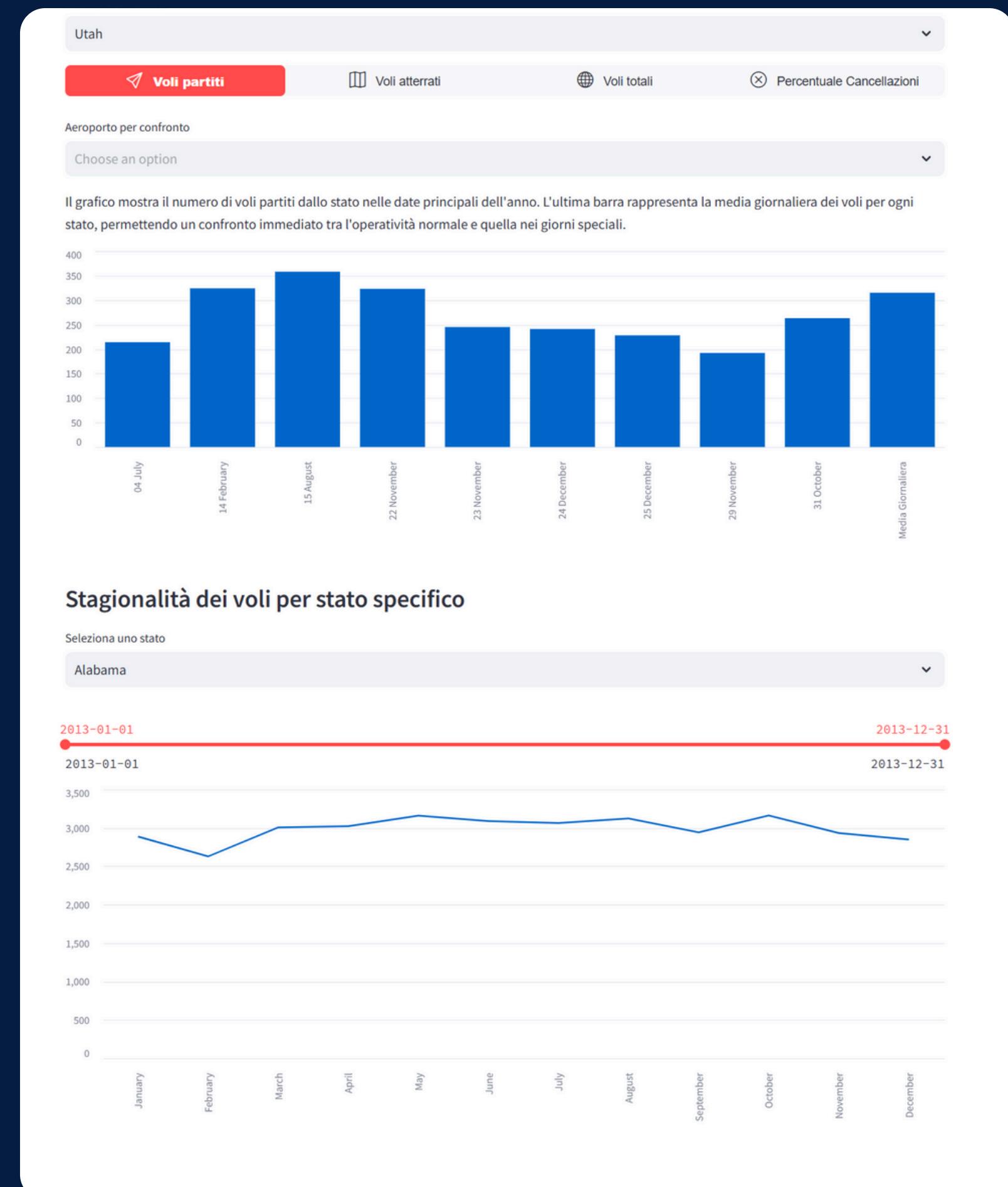
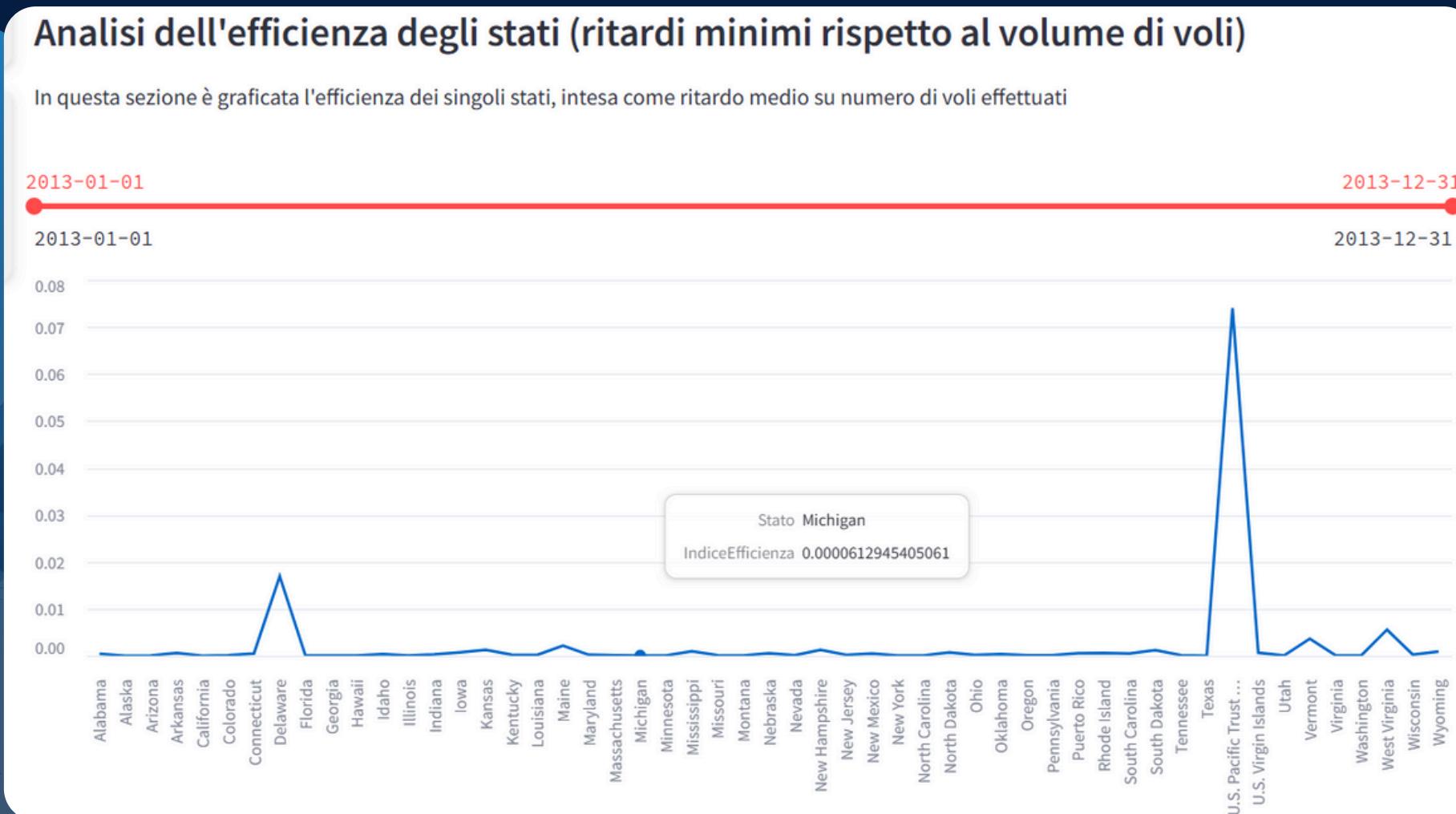
2013-01-01 2013-07-22

2013-01-01 2013-12-31

Percentuali delle cause di cancellazioni

Meteo	42.8%
Compagnia Aerea	33.6%
Sistema Nazionale	23.4%
Sicurezza	0.145%

05 L'APPLICAZIONE



05 L'APPLICAZIONE

Algoritmi di Machine Learning applicati ai dati

I
N
T
E
L
L
I
G
E
N
Z
A
A
R
T
I
F
I
C
I
A
L
E

Seleziona il tipo di analisi

Clustering degli aeroporti Clustering delle rotte

Analisi Cluster degli Aeroporti

Questa analisi raggruppa gli aeroporti in cluster basati su:

- Volume totale di voli
- Ritardo medio
- Distanza media dei voli
- Tasso di cancellazione

Seleziona il numero di cluster

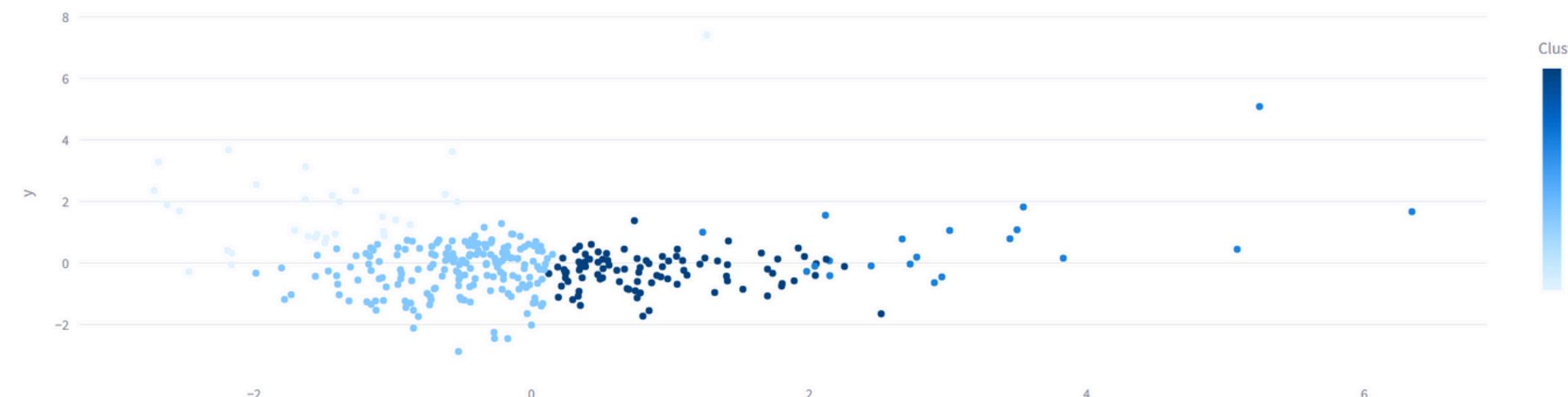
2 4 4 5

Statistiche dei Cluster

cluster	Media Voli Totali	Media Ritardo (min)	Distanza Media (miglia)	Tasso Cancellazione	Numero Aeroporti
0	2,415.19	12.54	279.25	0.05	31
1	7,770.22	5.67	343.45	0.02	178
2	323,254.1	9.92	1,122.13	0.01	20
3	52,922.1	6.78	747.08	0.01	91

[Panoramica Generale](#) [Analisi Dettagliata](#)

Panoramica Generale dei Cluster



05 L'APPLICAZIONE

I
N
T
E
L
L
I
G
E
N
Z
A
A
R
T
I
F
I
C
I
A
L
E

Questo grafico mostra una visione d'insieme dei cluster, dove:

- Ogni punto rappresenta un aeroporto
- La posizione è determinata da tutte le caratteristiche considerate
- Aeroporti vicini hanno caratteristiche simili
- I colori indicano i diversi cluster

Aeroporti per Cluster

Seleziona un cluster da visualizzare:

0

airport_code	voli_totali	ritardo_medio	distanza_media	rate_cancellazione	cluster
ASE	9,328	11.7197	403.2335	0.0444	0
MFR	6,869	10.9414	485.4541	0.0505	0
CWA	6,679	11.2251	232.7178	0.037	0
ACV	6,563	14.9041	223.2109	0.0666	0
CMI	4,658	15.8636	222.21	0.0509	0
SPI	3,712	6.8773	337.3836	0.0506	0
SUN	3,467	2.1981	226.1315	0.0617	0
SPS	2,803	9.7259	113	0.0428	0
RDD	2,346	15.4153	198.9582	0.0473	0
TXK	2,324	8.3195	181	0.0456	0

SFIDE PRINCIPALI



- ▶ VOLUME DEI DATI
- ▶ GEOLOCALIZZAZIONE
- ▶ OTTIMIZZAZIONE DELLE PRESTAZIONI
- ▶ USER EXPERIENCE (UX)
- ▶ TEMPISTICHE ED ORGANIZZAZIONE

**GRAZIE PER
L'ATTENZIONE**

