

THYNK UNLIMITED

Pacchetti
disponibili

Paesi

Info

Others

Turismo vr.

www.reallygreatsite.com

-

Alessandro Suntaxi

viaggi nel metaverso

Stiamo analizzando un dataset di un'agenzia di viaggi che offre esperienze immersive nel metaverso, permettendo di visitare luoghi esotici senza viaggiare fisicamente e con un impatto ambientale minimo.

Informazioni contenute nel dataset



Il dataset raccoglie informazioni su:

- Le **preferenze** e i **comportamenti degli utenti**
 - Le **destinazioni virtuali visitate**
- La **partecipazione a pacchetti promozionali**, come i tour guidati da 3-4 giorni
 - Metriche relative a **soddisfazione, frequenza e durata delle esperienze**

Descrizione statistica del dataset

Osservazioni Statistiche sul Dataset

Dalla descrizione statistica emergono alcune evidenze interessanti (e altre... discutibili):

- **user_id**: ID utente da 2 a 245, distribuiti in modo abbastanza uniforme. Tutti presenti, bene: possiamo tracciare ogni utente.
- **country_id**: Valori da 100 a 122, circa 20 paesi. Distribuzione compatta, ma **senza una legenda rimane solo un codice numerico sterile.**
- **pacchetto_id**: Valori tra 0 e 3, con media ≈ 1.46 . Pochi pacchetti, forse troppi pochi. Il sistema promozionale sembra ridotto
- **hotel_id**: Numeri da 21.019 a oltre 612.000. Ma parliamoci chiaro: ****questo campo non dice niente da solo****. È solo un codice — senza contesto, è rumore.
- **stars**: Tra 1 e 5, media ≈ 3.05 .

Aggiungi corpo del testo



Considerazioni

- **Zero valori nulli.**
- **Distribuzioni accettabili.**
- **Alcuni campi sembrano buttati** (es. hotel_id): serve contesto.
- **Manca profondità:** tanti numeri, poca informazione vera.



Conclusione: il dataset è un buon punto di partenza, ma ha bisogno di una interpretazione che forzata, i numeri da soli non servono per il nostro lavoro :

Noi dobbiamo **analizzare la nostra attività** partendo da ciò che abbiamo che sia scadente o meno estrapoliamo info utili



Analisi dei Cluster \ Risultati Utili per l'Azienda

Abbiamo identificato 4 gruppi in base al **comportamento** nei viaggi virtuali nel metaverso: preferenze esperienziali, lingua, Paese, piattaforma e pacchetto scelto.

Questa segmentazione non è teorica: offre **spunti operativi concreti** per ottimizzare comunicazione, prodotto e marketing.



Descrizione dei cluster

Cluster 0 – Clienti alto profilo UK/FR 25.6%

- Media hotel: 4.3 stelle virtuali → scelte orientate al **comfort e alla qualità dell'esperienza immersiva**.
- Parlano **inglese e francese**, provengono da UK e Francia.
- Preferenza per **TV e PC** (esperienza visiva più immersiva).
- Attratti da **categorie come relax e spa**.

Cluster 1 – Basic UK-only 21.7%

- Hotel virtuali scelti: 1.9 stelle → **attenzione al risparmio**.
- Solo inglese, solo utenti **britannici**.
- Dispositivi: **mobile e TV**.
- Scelgono pacchetti entry-level (0–1).

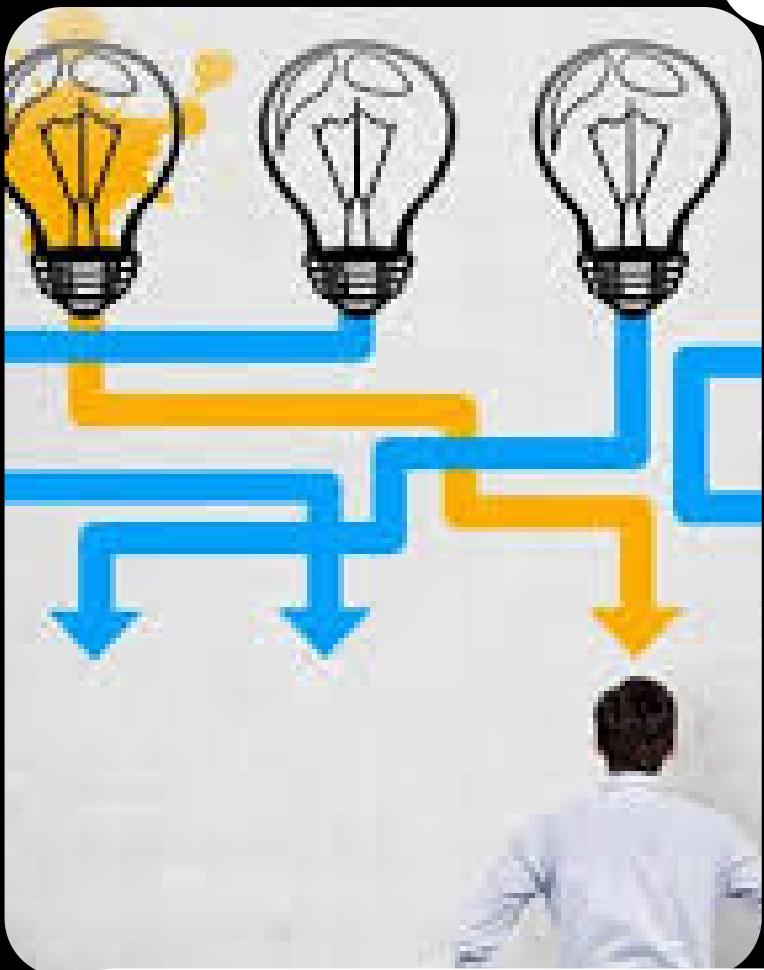
Cluster 2 – Italo-francesi low-cost 23.4%

- Hotel scelti: **1.5 stelle** → pubblico con approccio economico al turismo virtuale.
- Lingua italiana e francese, **forte presenza dall'Italia**.
- Preferenza per **PC e TV**.
- Pacchetto più usato: **3** (quello più itinerante).

Cluster 3 – Italiani benestanti 29.2%

- Hotel scelti: **4 stelle** → utenti con gusto per esperienze più curate.
- **Solo italiani**.
- Dispositivi preferiti: **TV e PC**.
- Esplorano varie categorie, bilanciando ****relax, avventura e wellness****.

Strategia



Come può agire l'azienda

Cluster | Strategia Operativa

0 | Email marketing di lusso, contenuti premium

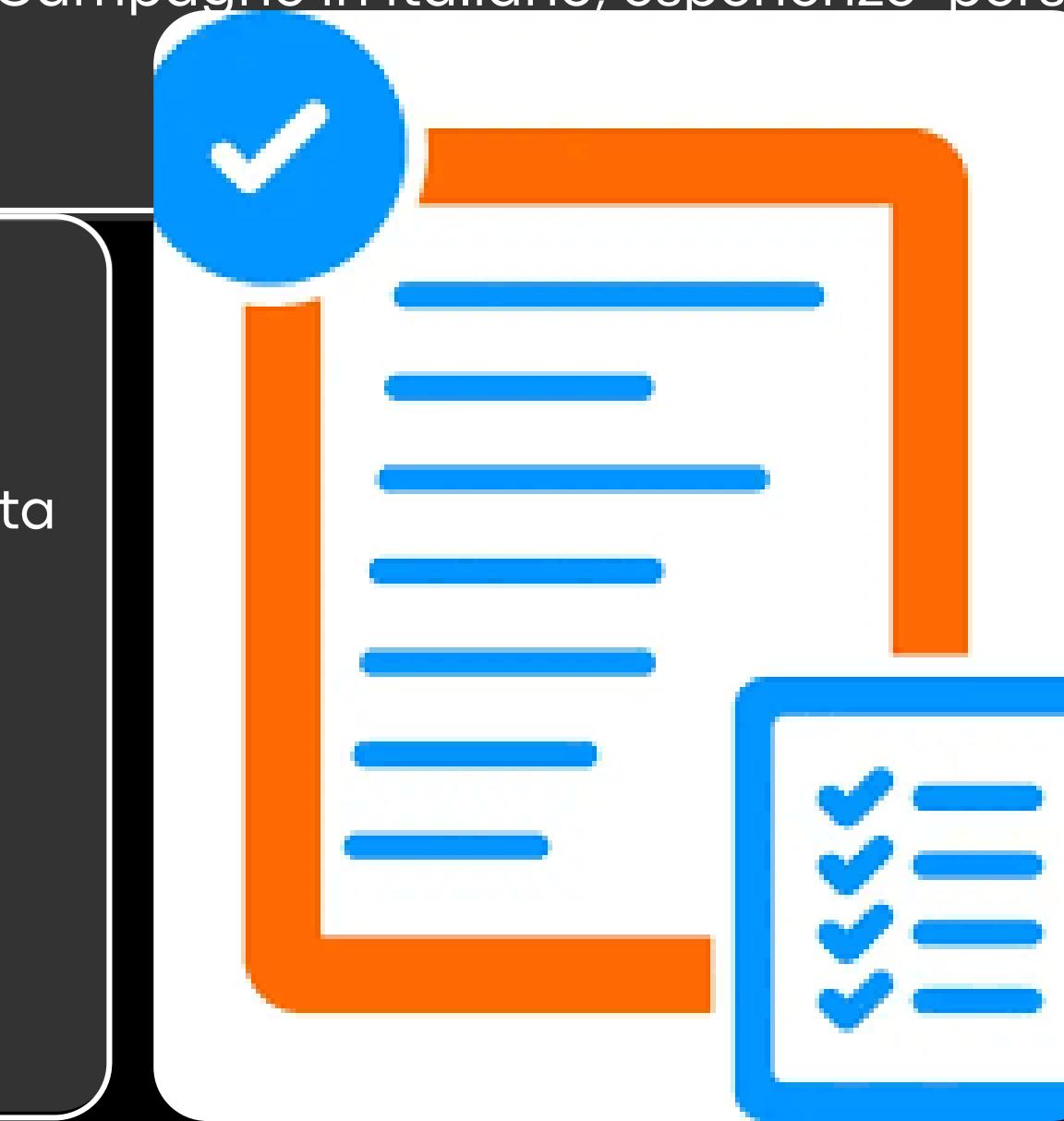
1 | Offerte last-minute, pacchetti sconto

2 | Comunicazione localizzata (IT/FR), promo smart

3 | Campagne in italiano, esperienze personalizzate

Conclusione

- I cluster non sono rivoluzionari, ma forniscono strategie
- Permettono una selezione comportamentale reale, basata su dati di utilizzo nel metaverso turistico.
- Se usati bene, possono:
 - Ottimizzare l'efficacia delle campagne
 - avere più chiaro il target per gruppo
 - Ridurre le produzioni pubblicitarie e costi



Disclaimer

Ma nel contesto dei **viaggi immersivi nel metaverso** è comunque plausibile dedurre una correlazione comportamentale:

- Chi sceglie **hotel a 4-5 stelle virtuali** cerca ambienti raffinati, ambientazioni esclusive, esperienze ad alto impatto visivo → **probabile maggiore disponibilità a spendere** o a valorizzare il servizio.
- Rispetto a chi visita solo strutture 1-2 stelle indicando per quei tipi di utenti **bassa propensione al premium**.



Attenzione però: si tratta segnali probabilistici, non di certezze. Utili per:

- Personalizzare le interazioni e offerte date le nostre supposizioni
- Testare differenti strategie di in base ai tipi di utenti

Modello di Classificazione

Predizione Scelta Pacchetto Premium all Feature

Predire se un utente sceglierà un pacchetto premium dato dall'unione di pack 1 e 2.

cercando di **scoprire** quale **feature** influenza la scelta di questi pacchetti premium

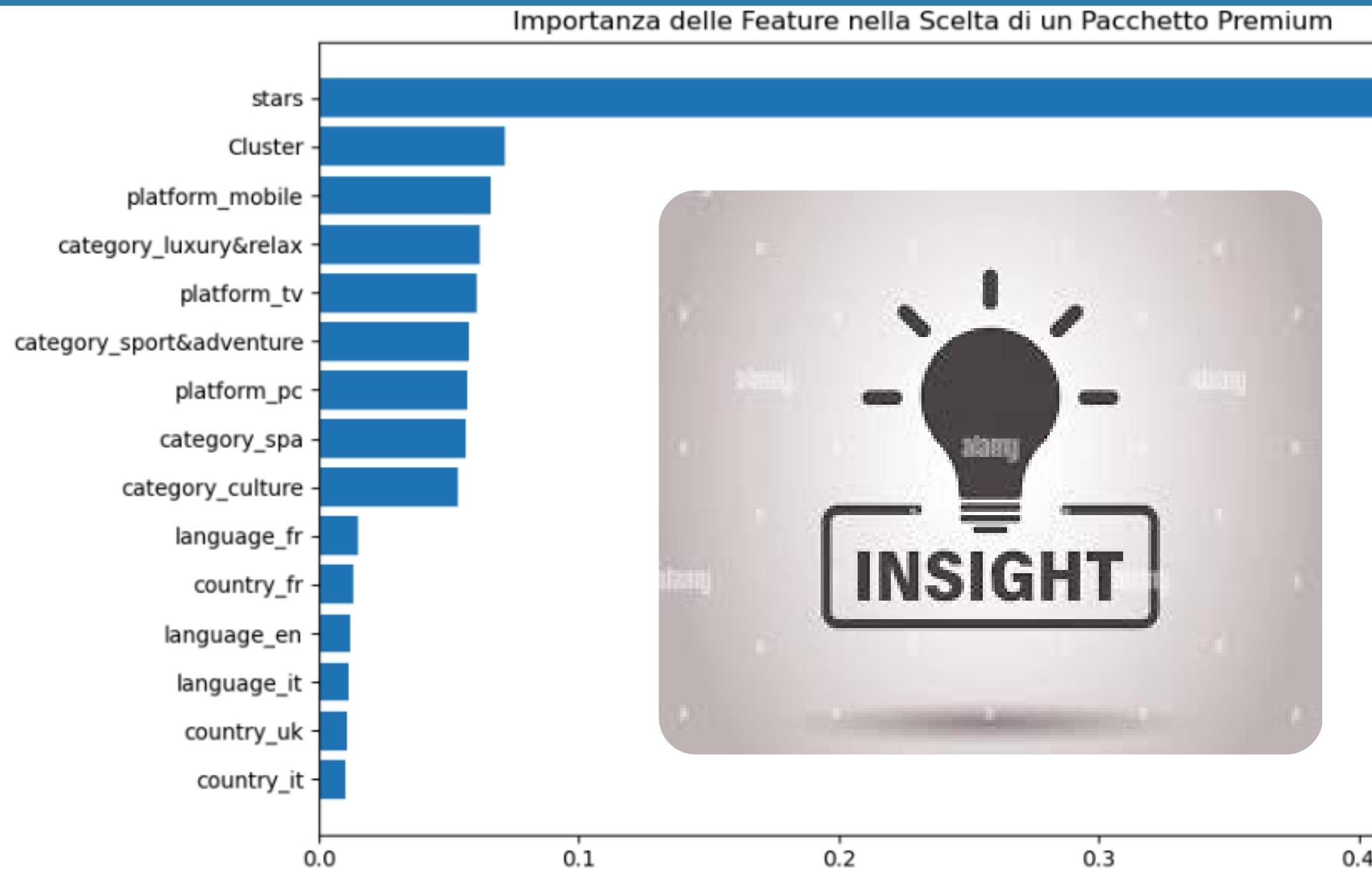
Feature Importance	
Feature	Importanza (%)
`stars`	44.3
`Cluster`	7.2
`platform_mobile`	6.6
`category_luxury&relax`	6.2
`platform_tv`	6.0
`category_sport&adv.`	5.7
`platform_pc`	5.7
`category_spa`	5.7
`category_culture`	5.3
Lingua / Paese	< 2% ciascuno

Il modello ha una accuratezza complessiva del 49%:

è in grado di distinguere in parte gruppi Base vs Premium, l'accuratezza resta limitata, suggerendo margini di miglioramento



Analisi delle percentuali feature



- La feature più rilevante è stars, che incide per oltre il 44%.
- Cluster, una feature creata, è la seconda più influente, indicando l'efficacia della segmentazione utenti.
- Le piattaforme e categorie di contenuto giocano sono rilevanti.
- Le variabili legate alle nazione sono le meno rilevanti per la scelta di un pacchetto.
- Mentre la lingua anche se segnata come feature meno rilevante ha la sua importanza per l'utente.

Analisi delle Visite e Previsioni

Disclaimer:

il modello è solido per una prima previsione, ma l'efficacia reale dipende dai dati storici se ampliato, può diventare un valido supporto alla pianificazione strategica.

Abbiamo costruito un modello di serie temporale semplice ma efficace basato sulle visite mensili presenti nel dataset. Gli obiettivo sono stati :

- Una **previsione** per i prossimi N mesi.
- L'identificazione **pattern media** per **mese**.
- Una **visual** della serie storica pulita delle visite mensili a partire dal 2023

Interpretazione del Grafico

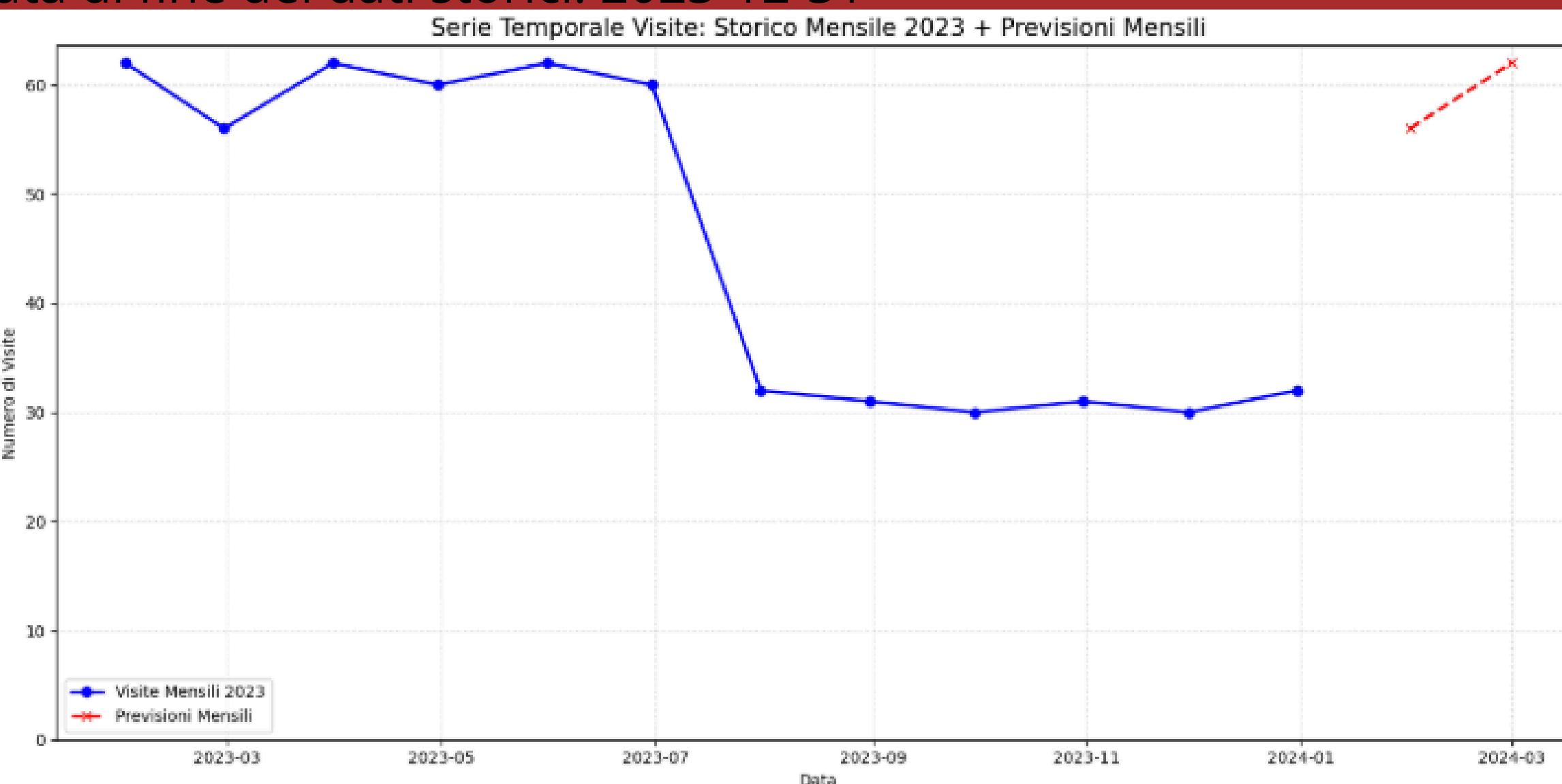
- Il 2023 mostra un calo netto nelle visite a partire da agosto.

- Le previsioni 2024 si basano su medie storiche + fattori mensili.

- Il modello prevede una ripresa moderata nei primi mesi del 2024, coerente con la stagionalità storica ma senza tener conto di tendenze recenti.

Data di inizio dei dati storici: 2022-01-31

Data di fine dei dati storici: 2023-12-31



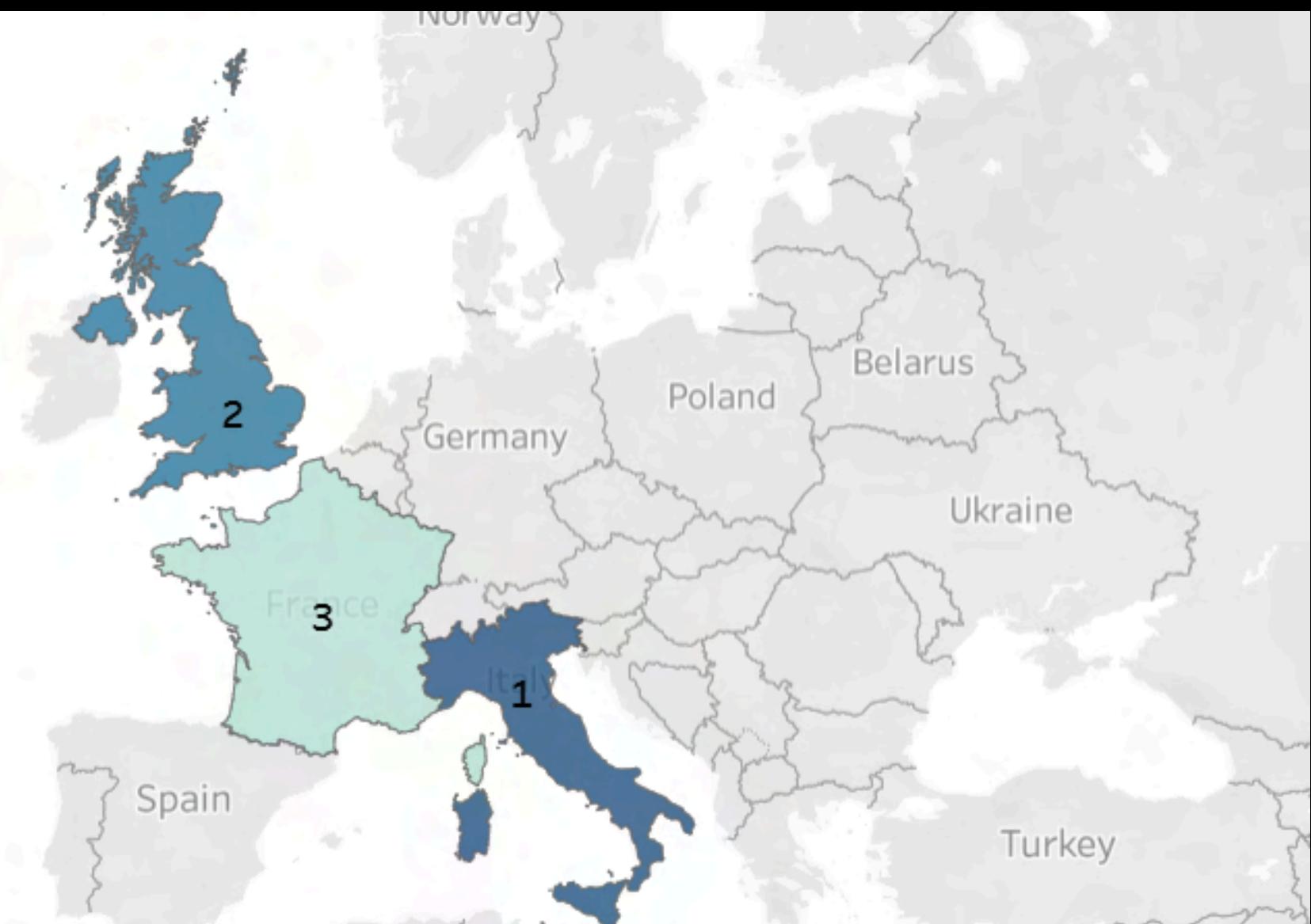
MEDIA PER MESE:

1	1.357664
2	1.226277
3	1.357664
4	1.313869
5	1.357664
6	1.313869
7	0.700730
8	0.678832
9	0.656934
10	0.678832
11	0.656934
12	0.700730

Utile come base, ma migliorabile con componenti di trend .

INFO UTILI

PAESI CON MAGGIORE VISITE OFFERTE DAI NOSTRI TOUR



Dai dati emerge che il numero di utenti attivi è così distribuito:

- 205 in Italia
- 185 nel Regno Unito
- 122 in Francia

Su Tableau ho realizzato una dashboard che rappresenta l'andamento della nostra attività in modo più sintetico e intuitivo, facilitando l'interpretazione dei dati e il monitoraggio delle performance.

Thank You

codice python

-

chart in tableau