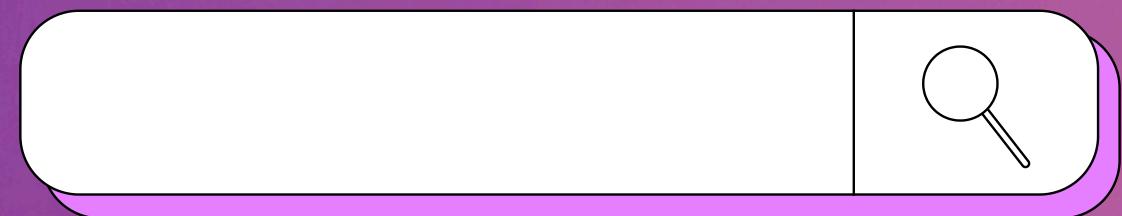




VIAGGI NEL METAVERSO: ANALISI DEL COMPORTAMENTO DEGLI UTENTI



Di Veronica Crisafio

OBIETTIVI DEL PROGETTO

1

EDA – Analisi Esplorativa con Python:

- Pulire il Dataset
- Capire i trend di utilizzo del servizio
- Identificare le categorie più richieste
- Analizzare comportamento utenti nel tempo

Clicca qui



2

VISUALIZATION - Realizzare una **dashboard interattiva in Tableau** che includa la creazione di personas di utenti tipo (o di pacchetto tipo).

Clicca qui



3

CONCLUSIONE - Proposta di strategia futura per viaggiare nel metaverso



DATASET

Column

visit_date

user_uuid

category

country_id

language

pacchetto_id

country

subscription_date

platform

hotel_id

stars

Descrizione

La data in cui l'utente ha visitato un hotel

Identificativo dell'utente

Categoria di vacanza

Identificativo della nazione

Lingua dell'utente

Identificativo del pacchetto

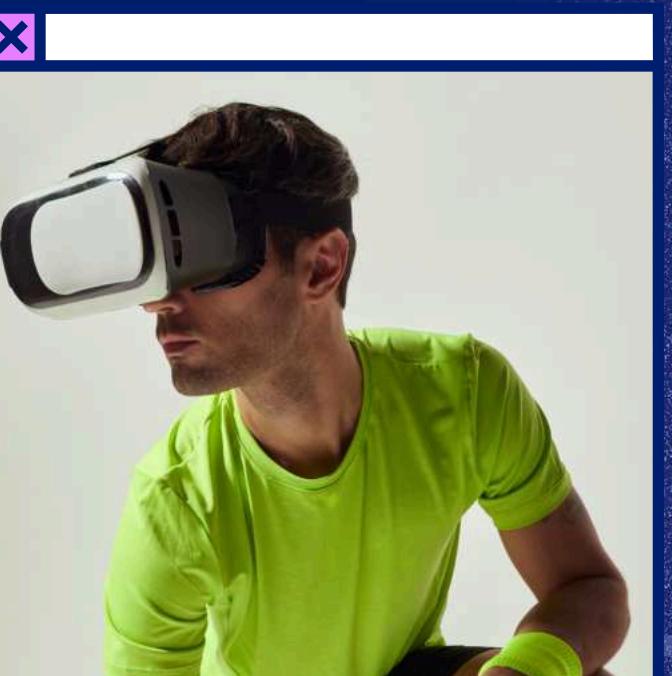
Nazionalità dell'utente

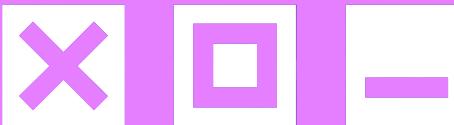
Giorno in cui l'utente si è iscritto

Piattaforma da cui l'utente sta usando l'app

Identificativo dell'hotel

Stelle medie assegnate all'hotel (da 1 a 5)





```
[ ] # Convert date columns to datetime
dftravel['visit_date'] = pd.to_datetime(dftravel['visit_date'], dayfirst=True)
dftravel['subscription_date'] = pd.to_datetime(dftravel['subscription_date'], dayfirst=True)

# Check for duplicates
duplicates = dftravel.duplicated().sum()

# Check value counts for key categorical features
category_counts = dftravel['category'].value_counts()
platform_counts = dftravel['platform'].value_counts()
language_counts = dftravel['language'].value_counts()

duplicates, category_counts, platform_counts, language_counts
```

→ (np.int64(0),
category
luxury&relax 419
sport&adventure 262
spa 194
culture 124
Name: count, dtype: int64,
platform
tv 447
pc 334
mobile 218
Name: count, dtype: int64,

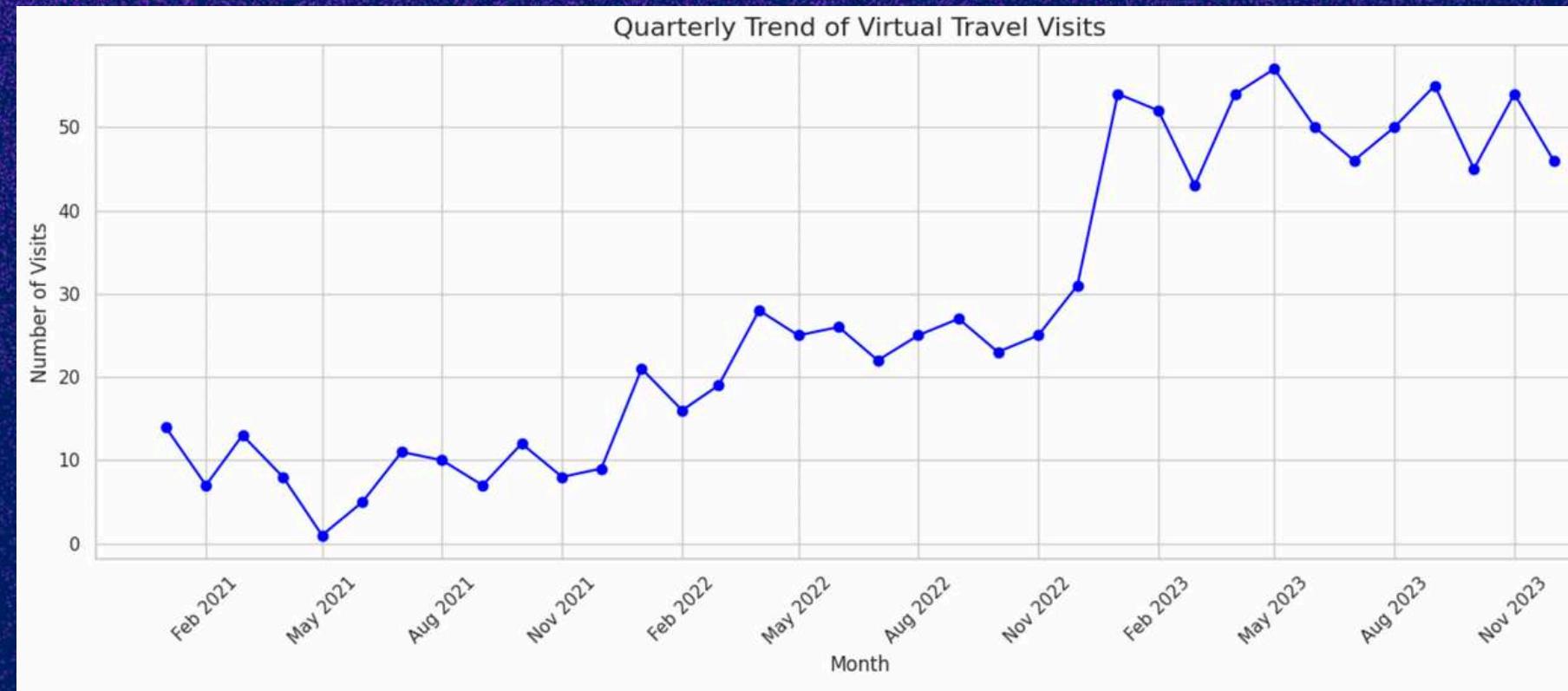
SINTESI DELL'ANALISI PRELIMINARE

Dopo aver **convertito** le colonne objects in formato dataset e aver effettuato la **pulizia iniziale**, l'analisi esplorativa ha evidenziato tre fattori chiave:

- **Categorie di viaggio:** quattro principali, con Luxury & Relax come la più scelta, seguita da Sport & Adventure, Spa e Culture.
- **Piattaforme di utilizzo:** il servizio è fruito principalmente da TV, poi PC e cellulare.
- **Lingue degli utenti:** prevale l'italiano, seguito da inglese e francese.

Inoltre, il dataset risulta pulito e privo di duplicati.

Andamento trimestrale delle visite virtuali



```
#Visualize travel trends overtime with a timeline
import matplotlib.dates as mdates

# Set style
sns.set(style="whitegrid")

# Create a new column for month and year
dftravel['visit_month'] = dftravel['visit_date'].dt.to_period('M')
monthly_visits = dftravel['visit_month'].value_counts().sort_index()
# Convert visit_month to datetime to plot nicely
dftravel['visit_month'] = dftravel['visit_date'].dt.to_period('M').dt.to_timestamp()

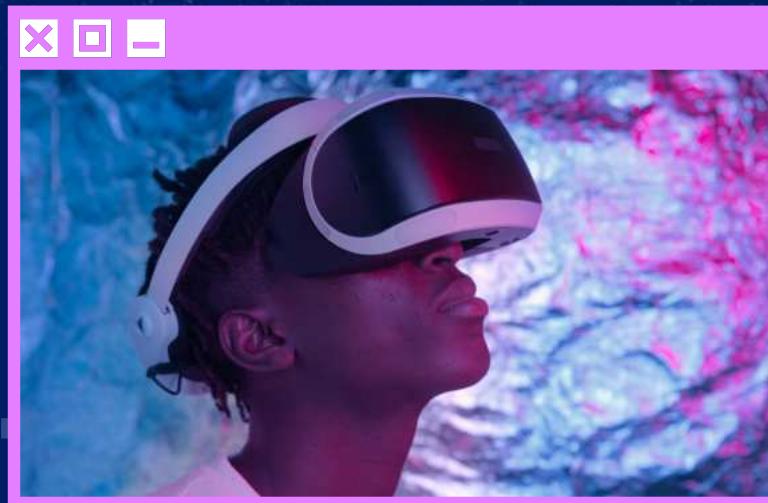
# Group by month
monthly_visits = dftravel.groupby('visit_month').size()

# Plot
plt.figure(figsize=(14, 6))
plt.plot(monthly_visits.index, monthly_visits.values, marker='o', color='blue')

# Format: show one label every 3 months
plt.gca().xaxis.set_major_locator(mdates.MonthLocator(interval=3)) # Every 3 months
plt.gca().xaxis.set_major_formatter(mdates.DateFormatter('%b %Y')) # Format: Jan 2023

# Labels
plt.title("Quarterly Trend of Virtual Travel Visits", fontsize=16)
plt.xlabel("Month")
plt.ylabel("Number of Visits")
plt.xticks(rotation=45)
plt.grid(True)
plt.tight_layout()
plt.show()
```

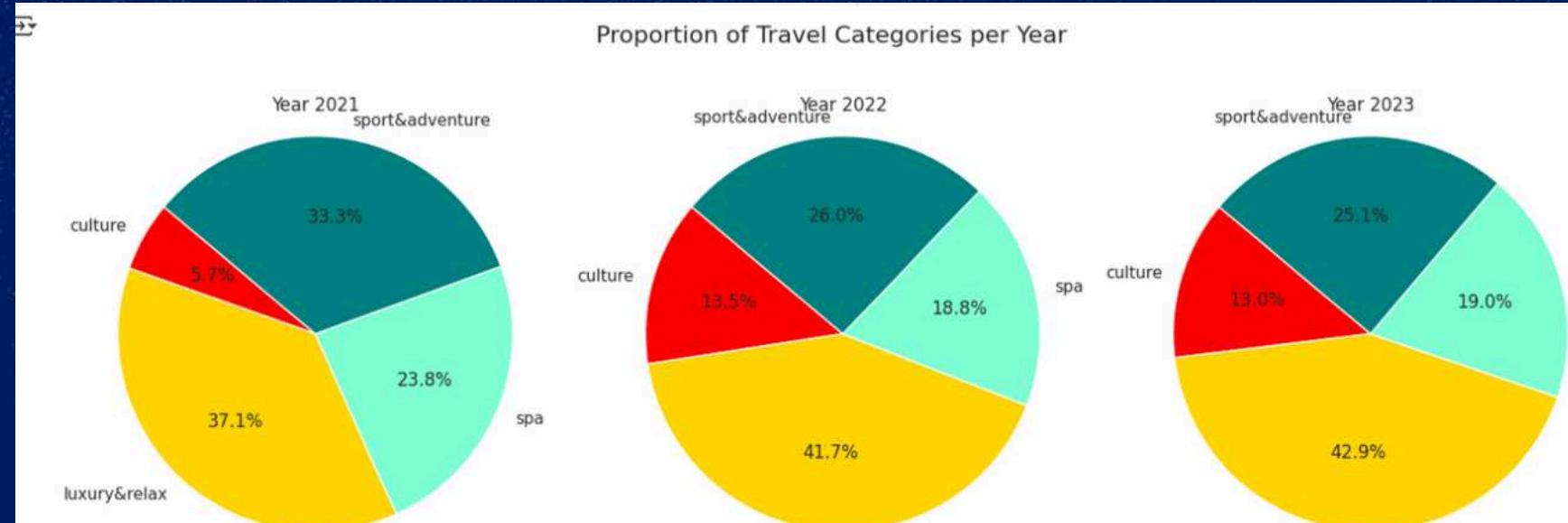
Dal grafico si evince un **costante aumento delle visite dal 2021 al 2023**. Nonostante qualche lieve calo in determinati mesi, si osserva un picco significativo all'inizio del 2023, segno di un **crescente interesse verso le esperienze di viaggio nel metaverso**.



Proporzione delle Categorie di Viaggio per Anno

L'evoluzione annuale delle preferenze mostra come **Luxury & Relax sia costantemente tra le categorie più richieste e con un trend in aumento**. Al contrario, si osserva una **diminuzione** relativa dell'**interesse per i viaggi Culturali** nel corso degli anni, indicando un possibile spostamento nelle preferenze degli utenti.

Queste informazioni sono fondamentali per ottimizzare le strategie di marketing e l'offerta in base alle tendenze stagionali e alle evoluzioni delle preferenze degli utenti.



```
# Extract the year from the 'visit_date'
dftravel['visit_year'] = dftravel['visit_date'].dt.year

# Group by year and category and count the occurrences
yearly_category_counts = dftravel.groupby(['visit_year', 'category']).size().unstack(fill_value=0)

# Get the unique years from the data
years = yearly_category_counts.index.sort_values()

# Create a figure and a set of subplots, one for each year
num_years = len(years)
fig, axes = plt.subplots(1, num_years, figsize=(5 * num_years, 5)) # Adjust figure size as needed

# Ensure axes is iterable even if there's only one year
if num_years == 1:
    axes = [axes]

# Iterate through each year and create a pie chart
for i, year in enumerate(years):
    ax = axes[i]
    year_data = yearly_category_counts.loc[year]
    categories = year_data.index
    counts = year_data.values
    colors = [color_map.get(cat, 'gray') for cat in categories] # Use color map, default to gray

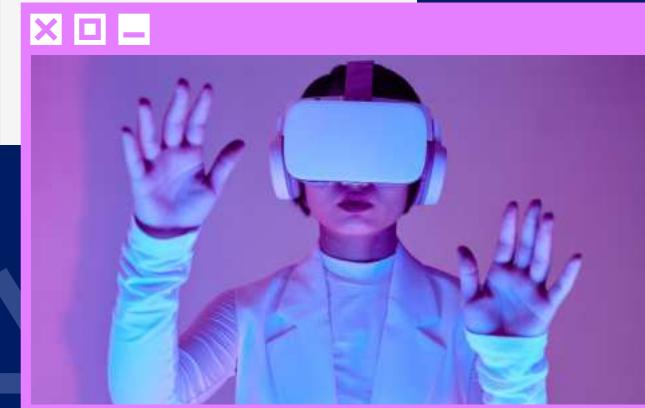
    # Create the pie chart
    wedges, texts, autotexts = ax.pie(counts, labels=categories, autopct='%.1f%%', startangle=140, colors=colors)

    # Ensure percentages are readable
    if autotexts is not None:
        for autotext in autotexts:
            if autotext is not None:
                try:
                    facecolor = autotext.get_bbox_patch().get_facecolor()
                    if isinstance(facecolor, tuple) and len(facecolor) >= 3:
                        luminosity = 0.299 * facecolor[0] + 0.587 * facecolor[1] + 0.114 * facecolor[2]
                        autotext.set_color('white') if luminosity < 0.5 else autotext.set_color('black')
                except AttributeError:
                    # Handle cases where autotext might not have get_bbox_patch
                    pass

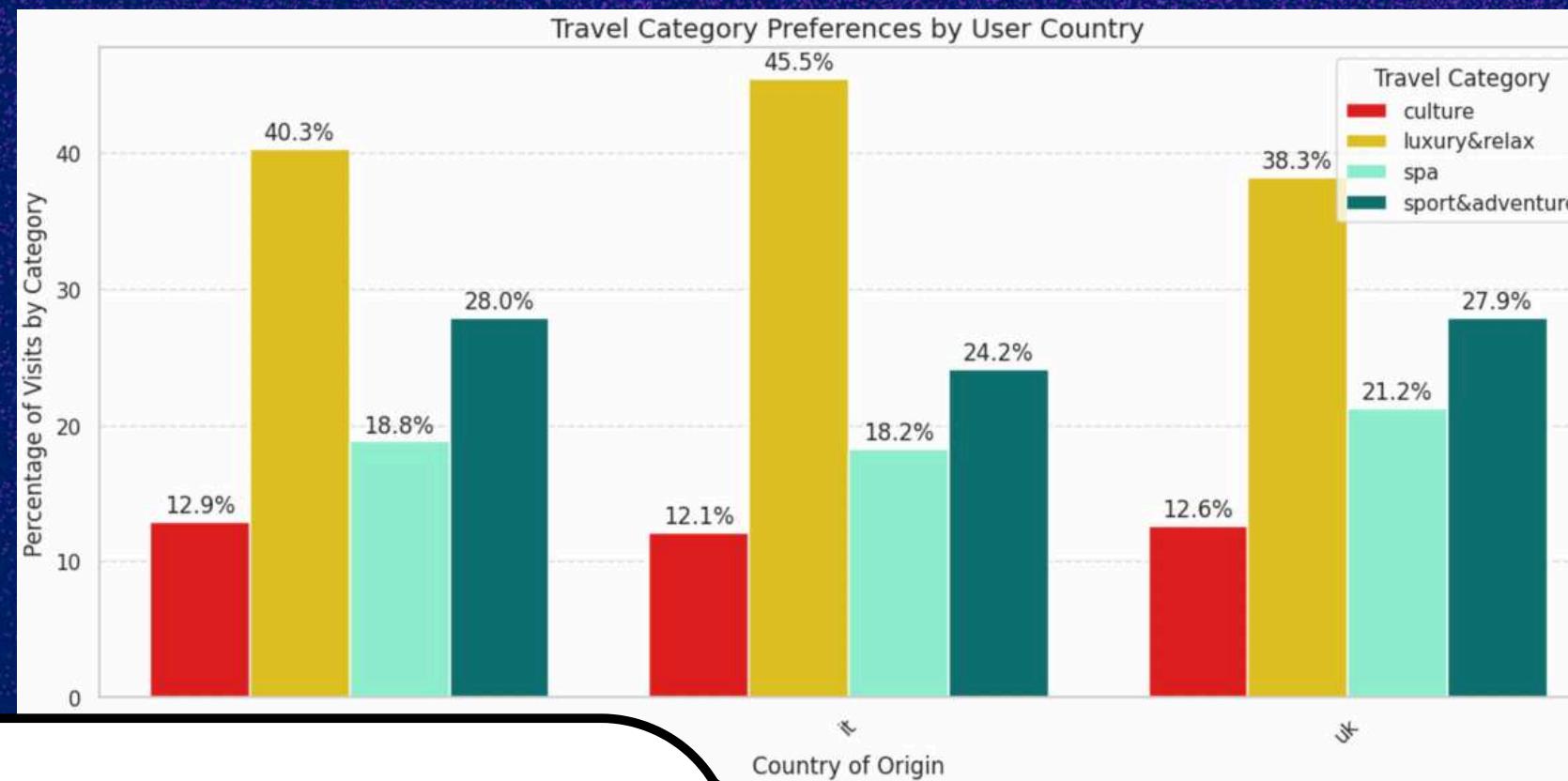
    ax.set_title(f'Year {year}')
    ax.axis('equal') # Equal aspect ratio ensures that pie is drawn as a circle.

# Add a main title for the entire figure
fig.suptitle('Proportion of Travel Categories per Year', fontsize=16, y=1.02)

plt.tight_layout()
plt.show()
```



Preferenze di viaggio per paese utente



45,5%

Degli utenti italiani sceglie la categoria Luxury & Relax, rendendola la più popolare nel Paese.



```
# Group total visits per country
total_per_country = dftrip.groupby('country').size().reset_index(name='total_visits')

# Count visits per country and category
trip_by_country_cat = dftrip.groupby(['country', 'category']).size().reset_index(name='count')

# Merge and compute percentage
trip_by_country_cat = trip_by_country_cat.merge(total_per_country, on='country')
trip_by_country_cat['percentage'] = (trip_by_country_cat['count'] / trip_by_country_cat['total_visits']) * 100

# Optional: define a custom color palette
custom_palette = {
    'luxury&relax': 'gold',
    'sport&adventure': 'teal',
    'spa': 'aquamarine',
    'culture': 'red'
}

# Create the barplot
plt.figure(figsize=(12, 6))
barplot = sns.barplot(
    data=trip_by_country_cat,
    x='country',
    y='percentage',
    hue='category',
    palette=custom_palette
)

# Add percentage labels on top of each bar
for container in barplot.containers:
    barplot.bar_label(container, fmt='%.1f%%', label_type='edge', padding=2)

# Titles and axis labels
plt.title("Travel Category Preferences by User Country", fontsize=14)
plt.xlabel("Country of Origin", fontsize=12)
plt.ylabel("Percentage of Visits by Category", fontsize=12)

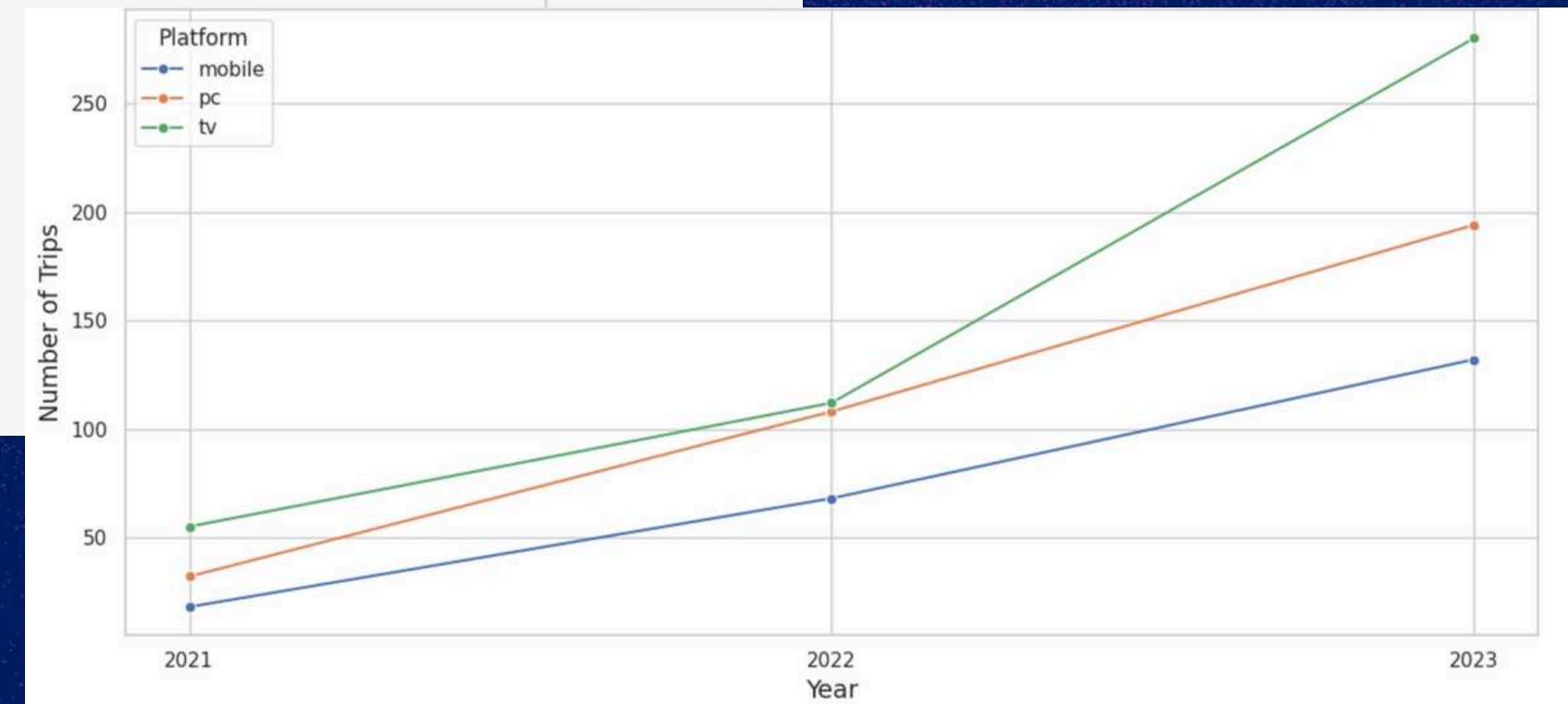
# Grid and legend
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.legend(title="Travel Category")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

Utilizzo delle piattaforme nel tempo

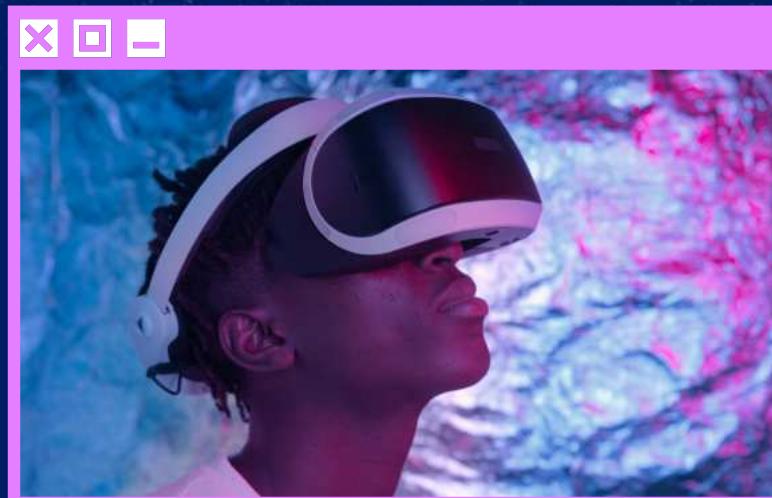
```
# Group by year and Platform
platform_usage = dftrip.groupby(['visit_year', 'platform']).size().reset_index(name='count')

# Create a lineplot
plt.figure(figsize=(12, 6))
sns.lineplot(data=platform_usage, x='visit_year', y='count', hue='platform', marker='o')

# Adjust details
plt.title("Platform Usage Over Time", fontsize=16)
plt.xlabel("Year", fontsize=14)
plt.ylabel("Number of Trips", fontsize=14)
plt.legend(title="Platform")
plt.xticks([2021, 2022, 2023])
plt.grid(True)
plt.tight_layout()
plt.show()
```



Il grafico mostra un **chiaro aumento dell'utilizzo di tutte le piattaforme tra il 2021 e il 2023**. Si può notare come nel **2022 PC e TV** avevano **quasi lo stesso utilizzo per poi differenziarsi nel 2023** dove la TV registra una crescita particolarmente marcata. Anche PC e mobile registrano una crescita costante, ma con volumi inferiori rispetto alla TV.

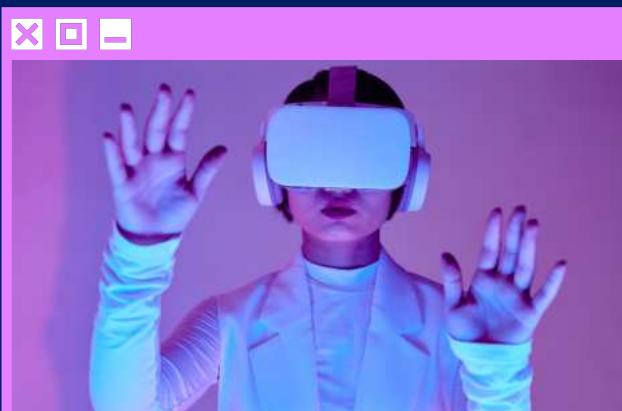


Comportamento tra iscrizione e visita

L'analisi della differenza in giorni tra la data di sottoscrizione e la data di visita (days_from_subscription) mostra quanto segue:

- Conteggio totale: 999 visite analizzate.
- **Media:** In media, gli utenti visitano 538 giorni dopo, ovvero **1 anno e 5 mesi dopo essersi iscritti**.
- Deviazione standard: 277 giorni, indica un'alta variabilità tra gli utenti.
- **Visite più rapide:** Il minimo è di **13 giorni dopo la sottoscrizione**.
- Distribuzione:
 - Il 25% degli utenti visita entro 336 giorni, ovvero circa 11 mesi.
 - Il 50% (mediana) entro 496 giorni (~1 anno e 4 mesi).
 - Il 75% entro 718 giorni, ovvero quasi 2 anni
- **Visite più tardive:** Un utente ha visitato **dopo 1395 giorni (~3 anni e 10 mesi)**.

Conclusione: **tutti gli utenti hanno effettuato la visita dopo l'iscrizione**, ma il tempo varia molto. Sarebbe interessante capire quali altri fattori possano influenzare la scelta di una visita dopo l'iscrizione.

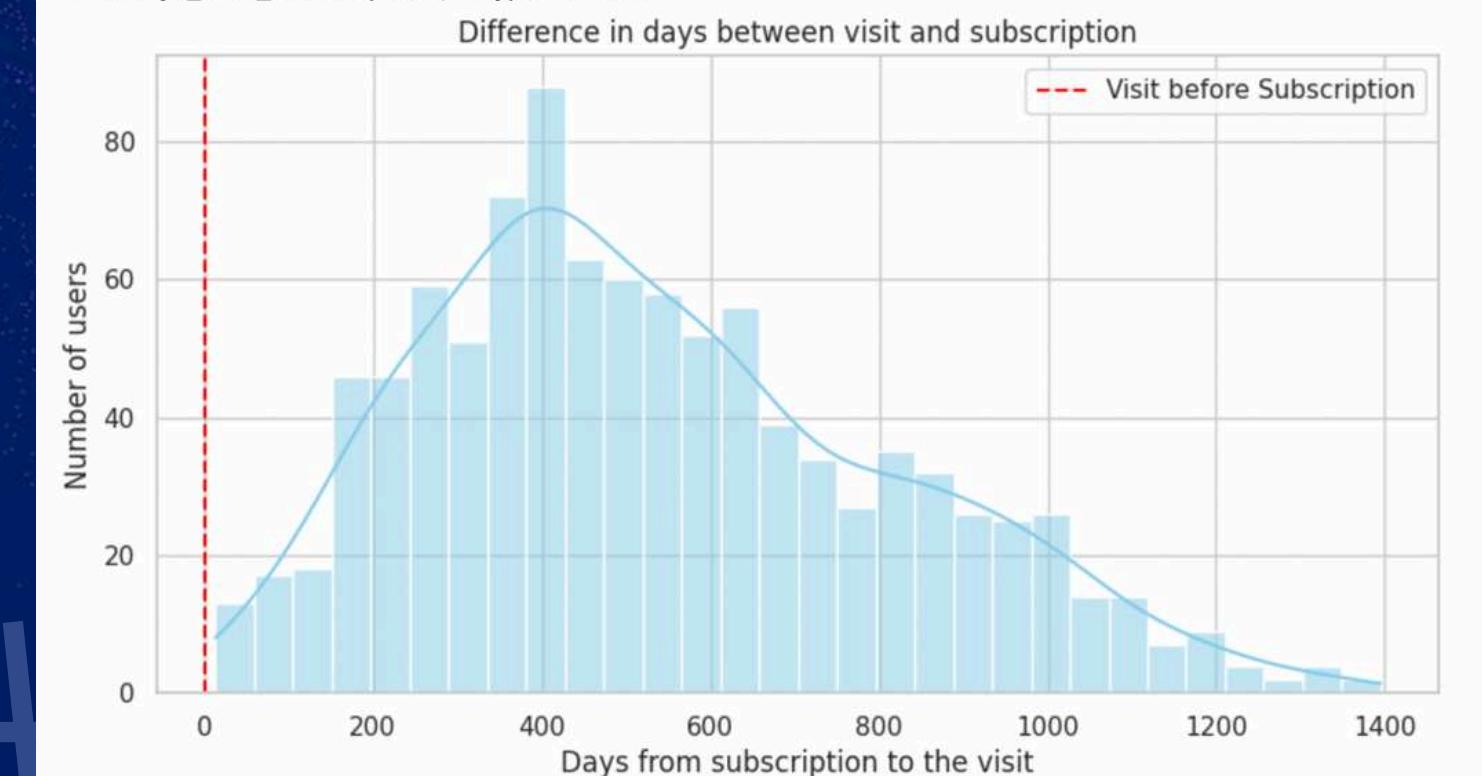


```
# Difference in days between visit and subscription
dftrip['days_from_subscription'] = (dftrip['visit_date'] - dftrip['subscription_date']).dt.days

# Statistics
print(dftrip['days_from_subscription'].describe())

# Visualization
plt.figure(figsize=(10, 5))
sns.histplot(dftrip['days_from_subscription'], bins=30, kde=True, color='skyblue')
plt.axvline(0, color='red', linestyle='--', label='Visit before Subscription')
plt.title('Difference in days between visit and subscription')
plt.xlabel('Days from subscription to the visit')
plt.ylabel('Number of users')
plt.legend()
plt.show()
```

```
count    999.000000
mean     538.356356
std      277.781129
min      13.000000
25%     336.000000
50%     496.000000
75%     718.000000
max    1395.000000
Name: days_from_subscription, dtype: float64
```



Analisi delle visite per singolo utente

Abbiamo analizzato il numero di visite effettuate per ciascun utente unico (user_uuid) all'interno del dataset.

📌 Statistiche principali:

- Utenti unici analizzati: 239
- **Numero medio di visite** per utente: 4.18
- Visite minime per utente: 1
- **Visite massime** per utente: 10
- Mediana (50%): 4 visite
- 75° percentile: 6 visite

🧠 Interpretazione:

- Gli utenti tendono a visitare in media 4 volte, con alcuni molto più attivi.
- La distribuzione è abbastanza bilanciata, con una leggera tendenza verso **utenti "fedeli"**.
- Un 25% degli utenti più attivi effettua almeno 6 visite, segno di una **buona retention**.

Il dato suggerisce che una parte degli utenti interagisce frequentemente con il servizio.

Questa informazione può rivelarsi utile per segmentare l'utenza in base alla frequenza di utilizzo, così da ideare strategie personalizzate, come sconti o promozioni mirate, per incentivare ulteriormente l'engagement.

```
visits_per_user = dftrip['user_uuid'].value_counts()  
print(visits_per_user.describe())
```

	count	mean	std	min	25%	50%	75%	max
	239.00000	4.179916	2.007615	1.000000	3.000000	4.000000	6.000000	10.000000
Name:	count							
		dtype:						
		float64						

Un **25%**

**degli utenti più attivi
effettua almeno 6 visite,
segno di una buona
retention.**



Classificazione degli utenti in base alla frequenza delle visite

```
Step 1: Count visits per user
sits_per_user = dftrip['user_uuid'].value_counts().reset_index()
sits_per_user.columns = ['user_uuid', 'visit_count']

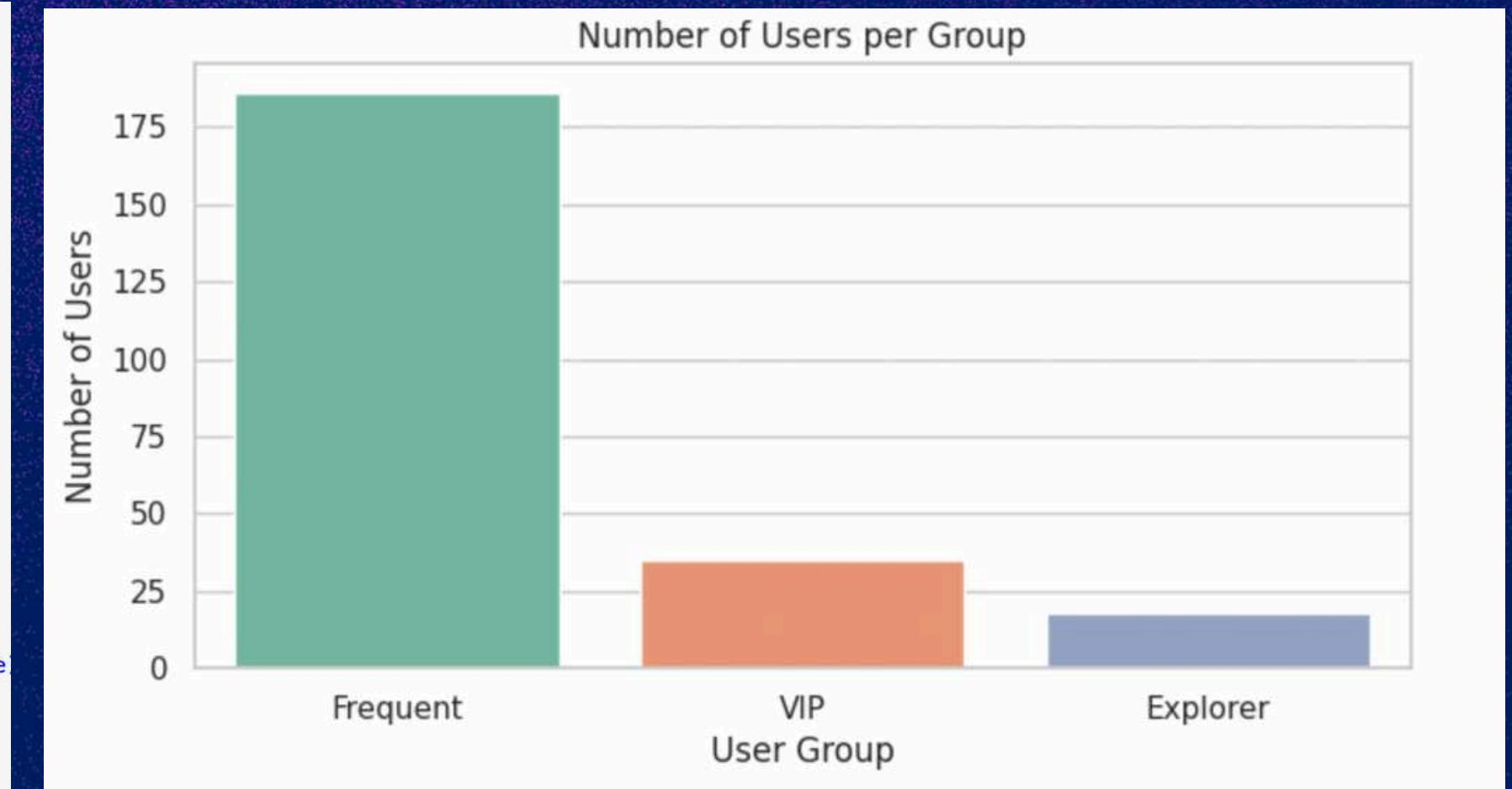
Step 2: Categorize users
def label_user_segment(visits):
    if visits == 1:
        return 'Explorer'
    elif 2 <= visits <= 6:
        return 'Frequent'
    else:
        return 'VIP'

sits_per_user['user_segment'] = visits_per_user['visit_count'].apply(label_user_segment)

Step 3: Merge back to main df to analyze behavior by group
trip_with_groups = dftrip.merge(visits_per_user[['user_uuid', 'user_segment']], on='user_uuid')

Step 4: Count users in each group
user_group_counts = visits_per_user['user_segment'].value_counts()

Step 5: Plot – number of users per group
t.figure(figsize=(8, 4))
s.barplot(x=user_group_counts.index, y=user_group_counts.values, hue=user_group_counts.index, palette='Set2', legend=False)
t.title('Number of Users per Group')
t.xlabel('User Group')
t.ylabel('Number of Users')
t.show()
```

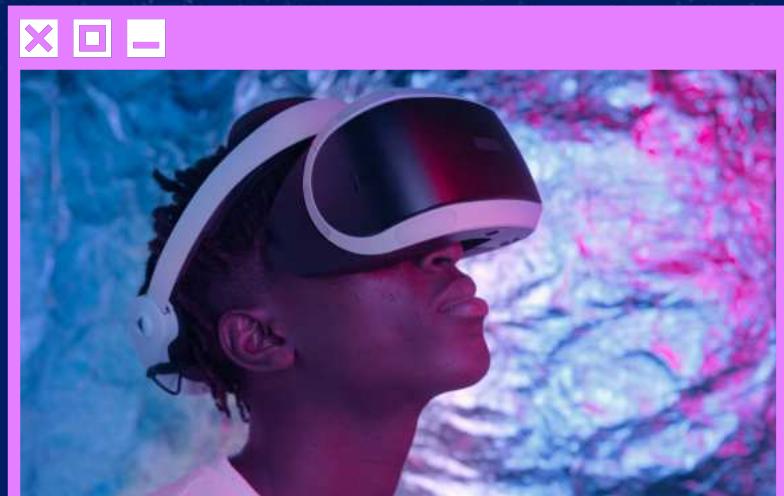


Questa classificazione è stata realizzata per identificare tre gruppi principali di utenti:

- **Explorer**: utenti occasionali con una sola visita
- **Frequent Flyer**: utenti abituali con 2-6 viaggi
- **VIP**: utenti altamente attivi con oltre 6 esperienze nel metaverso

Successivamente, questi gruppi sono stati analizzati in relazione alle diverse categorie di viaggio per evidenziare comportamenti e preferenze.

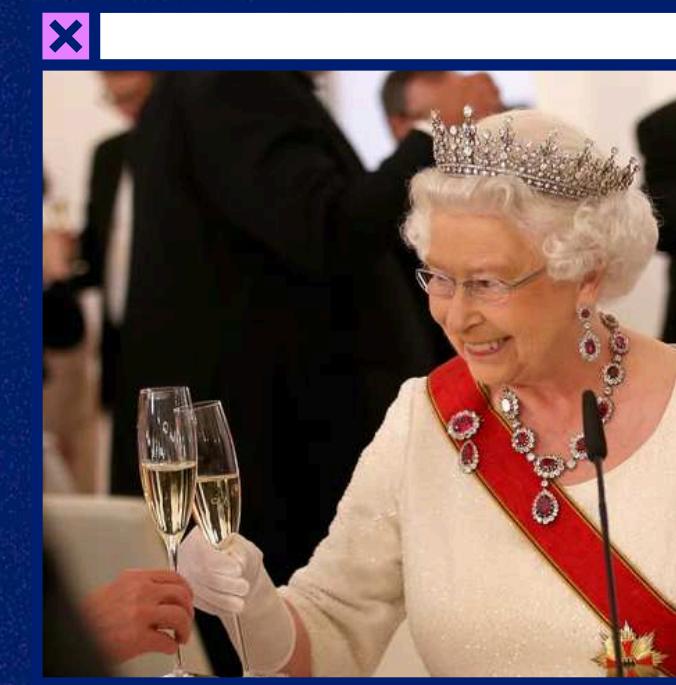
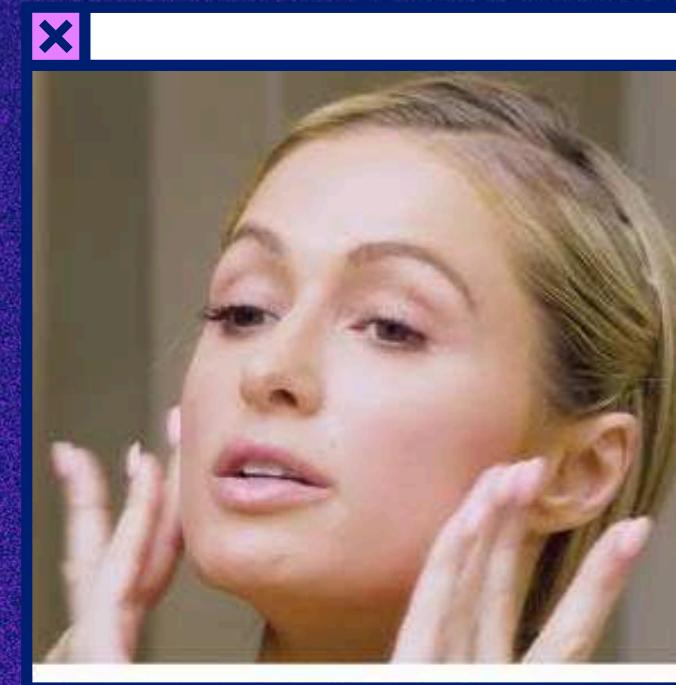
👉 Per un'analisi approfondita dell'intera EDA, è possibile consultare il seguente [LINK](#).



INTRODUZIONE AI PROFILI UTENTE

Abbiamo individuato 4 profili tipo, uno per ogni categoria di viaggio.

[CLICCA QUI PER VEDERE LA DASHBOARD DI TABLEAU](#)



Lara Croft



Paris Hilton



Queen Elizabeth



Rita Levi



Sport & Adventure

Spa

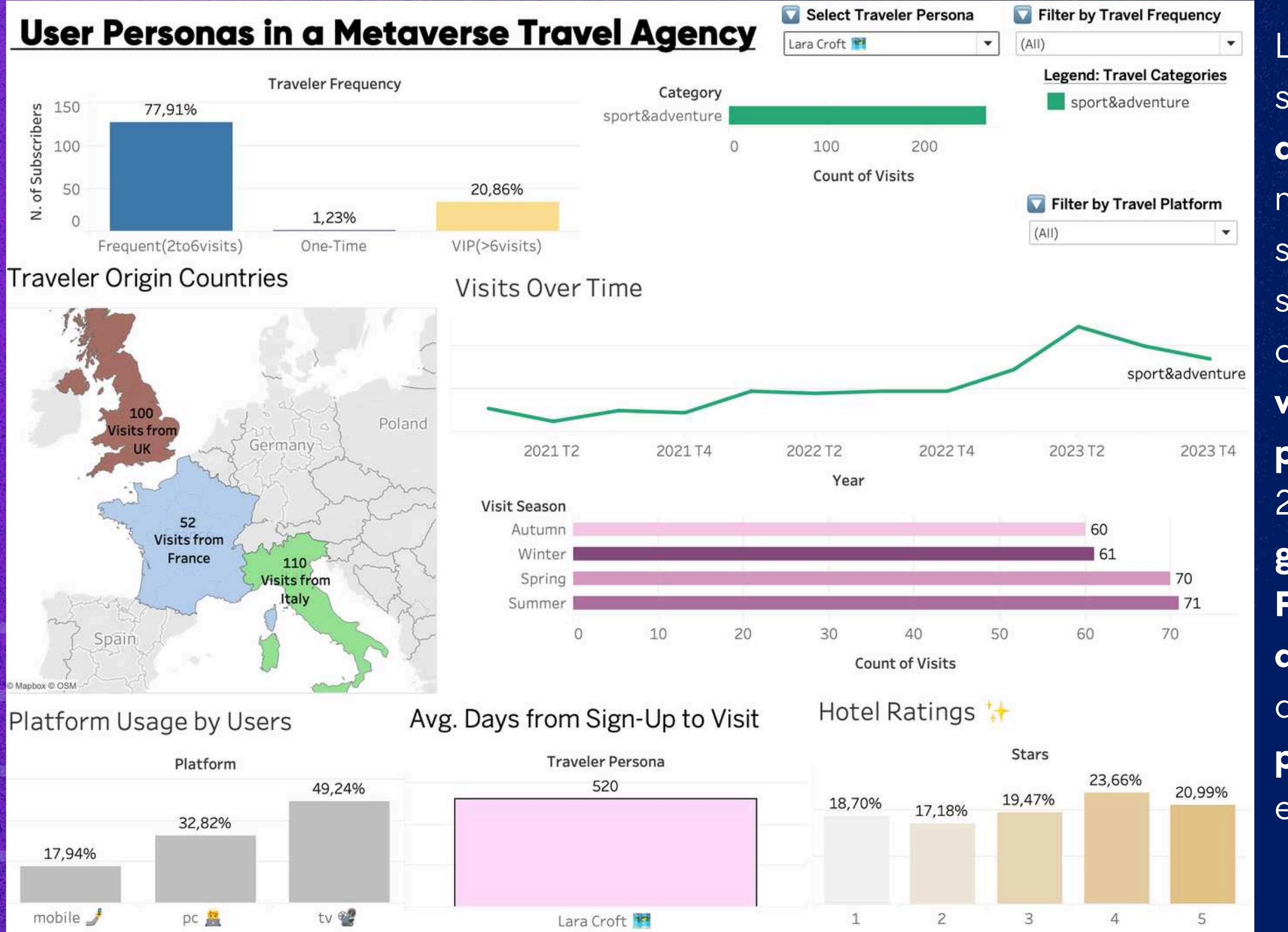
Luxury & Relax

Culture

LARA CROFT

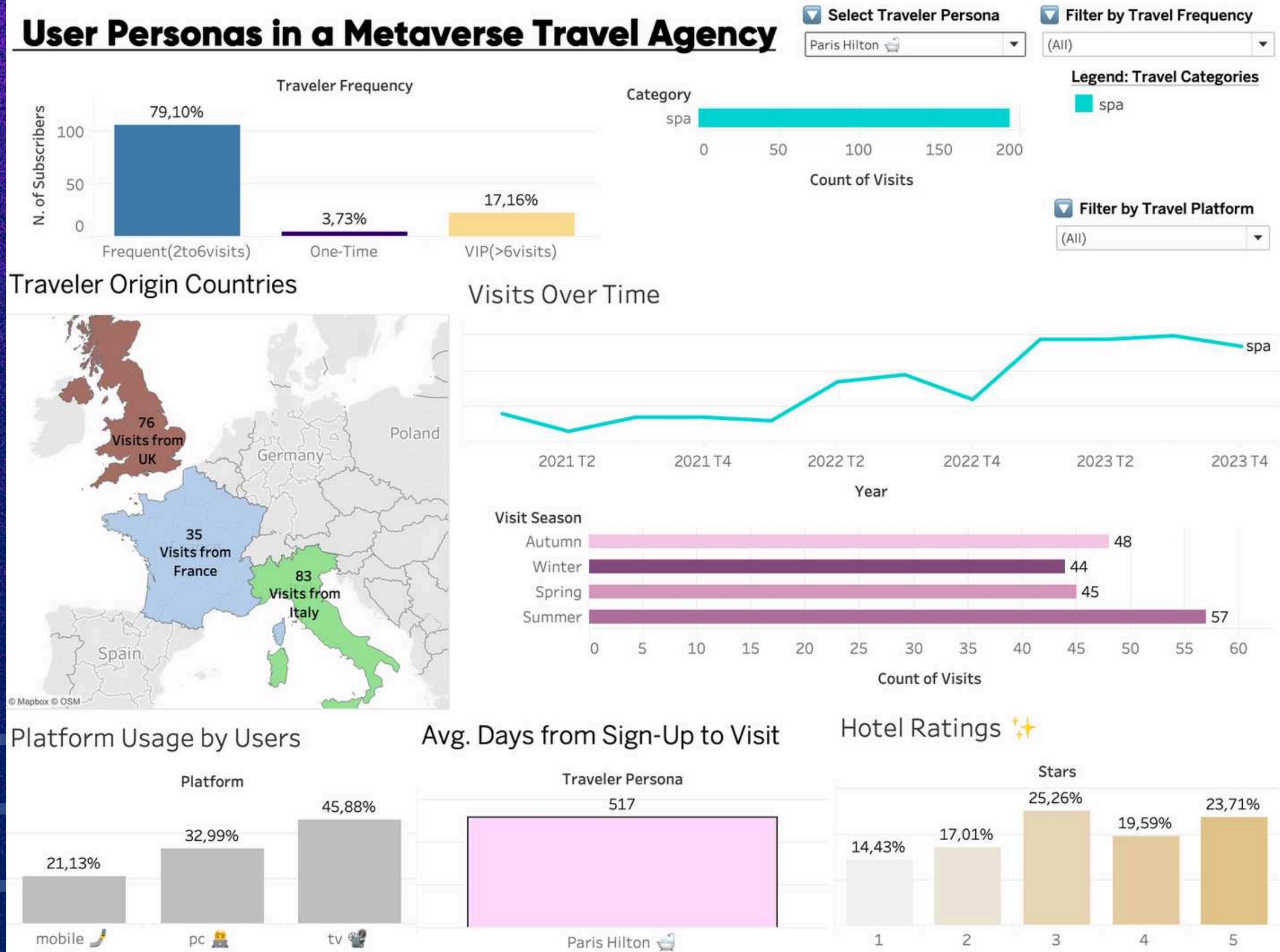


User Personas in a Metaverse Travel Agency



La persona Lara Croft incarna il viaggiatore sportivo e avventuroso nel metaverso, **attratto da esperienze dinamiche e adrenaliniche**. La maggior parte è frequent flyer (77,91%), cioè sono viaggiatori che hanno svolto dai 2 ai 6 viaggi, seguita dai VIP (20%) e da una piccola parte di one-time traveler (1,23%), per **un totale di 262 visite**. Le **stagioni preferite sono estate e primavera**, con un **andamento in crescita** dal 2021 e un lieve calo solo a fine 2023. **L'Italia guida con 110 visite, seguita da UK (100) e Francia (52)**. Il rating più comune degli hotel è di **4 stelle (23,66%)**. In media servono 520 giorni dall'iscrizione alla prima visita. **La TV è la piattaforma più usata (49,24%)**, seguita da PC e mobile.

PARIS HILTON

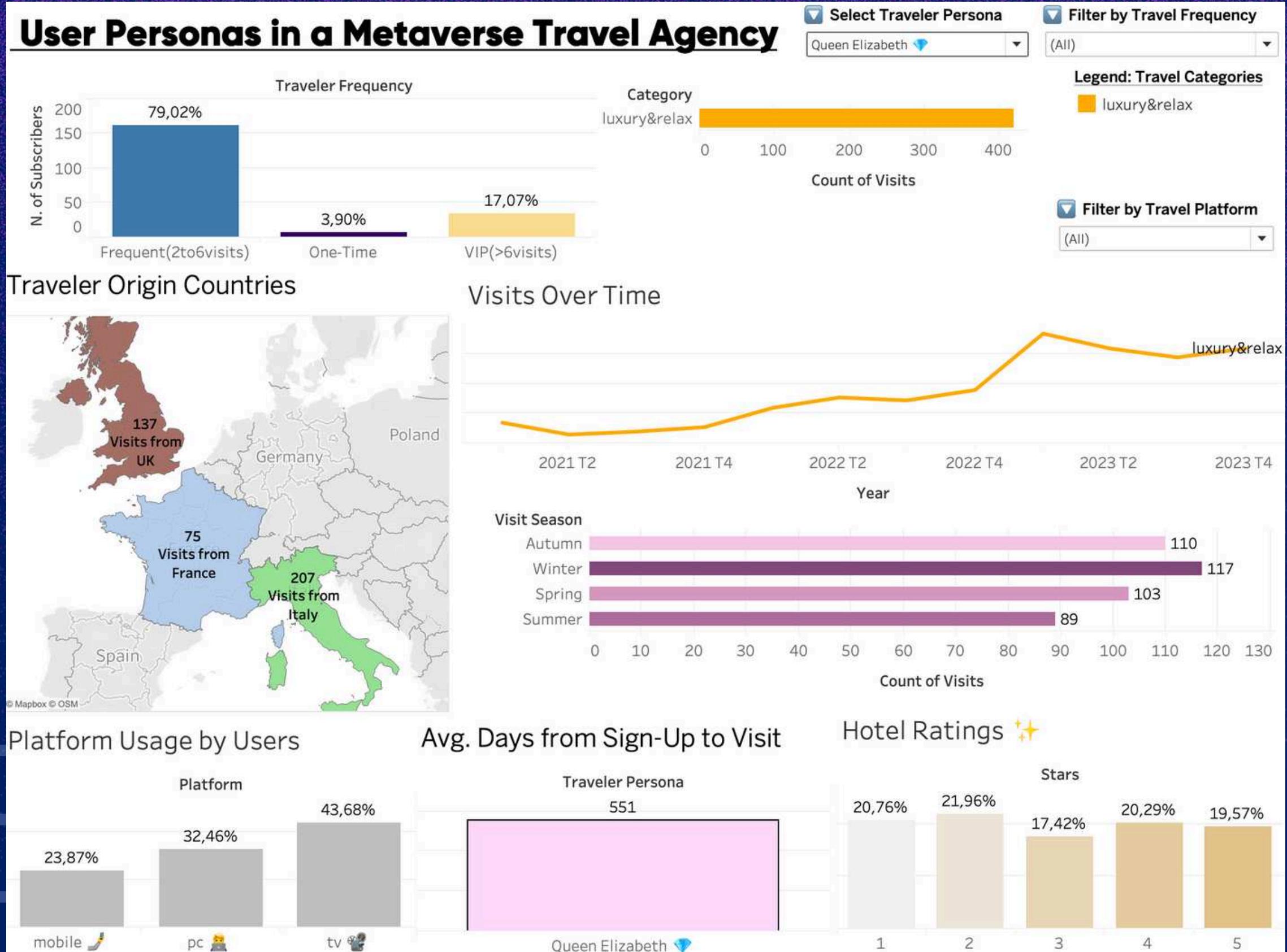


Paris Hilton è il profilo dedicato a **chi cerca relax, comfort e benessere nel metaverso, con esperienze immersive in spa e ambienti rilassanti**. Il 79,10% sono frequent flyer, il 17,16% VIP, mentre i one-time traveler rappresentano il 3,73%. Si **registrano 83 visite dall'Italia , 76 dal Regno Unito e 35 dalla Francia per un totale di 194 visite. L'andamento è in crescita** costante, con solo un lieve rallentamento a fine 2022. La **stagione più amata è l'estate**, e la piattaforma preferita rimane la TV (45,88%), **seguita da PC (32,99%) e mobile (21,13%) che hanno un maggiore utilizzo rispetto alla categoria precedente**. Servono in media 517 giorni dall'iscrizione alla prima visita. Il rating degli **hotel** è piuttosto distribuito, con **un picco sulla valutazione a 3 stelle (25,26%)**, suggerendo spazio per migliorare l'esperienza.

QUEEN ELIZABETH

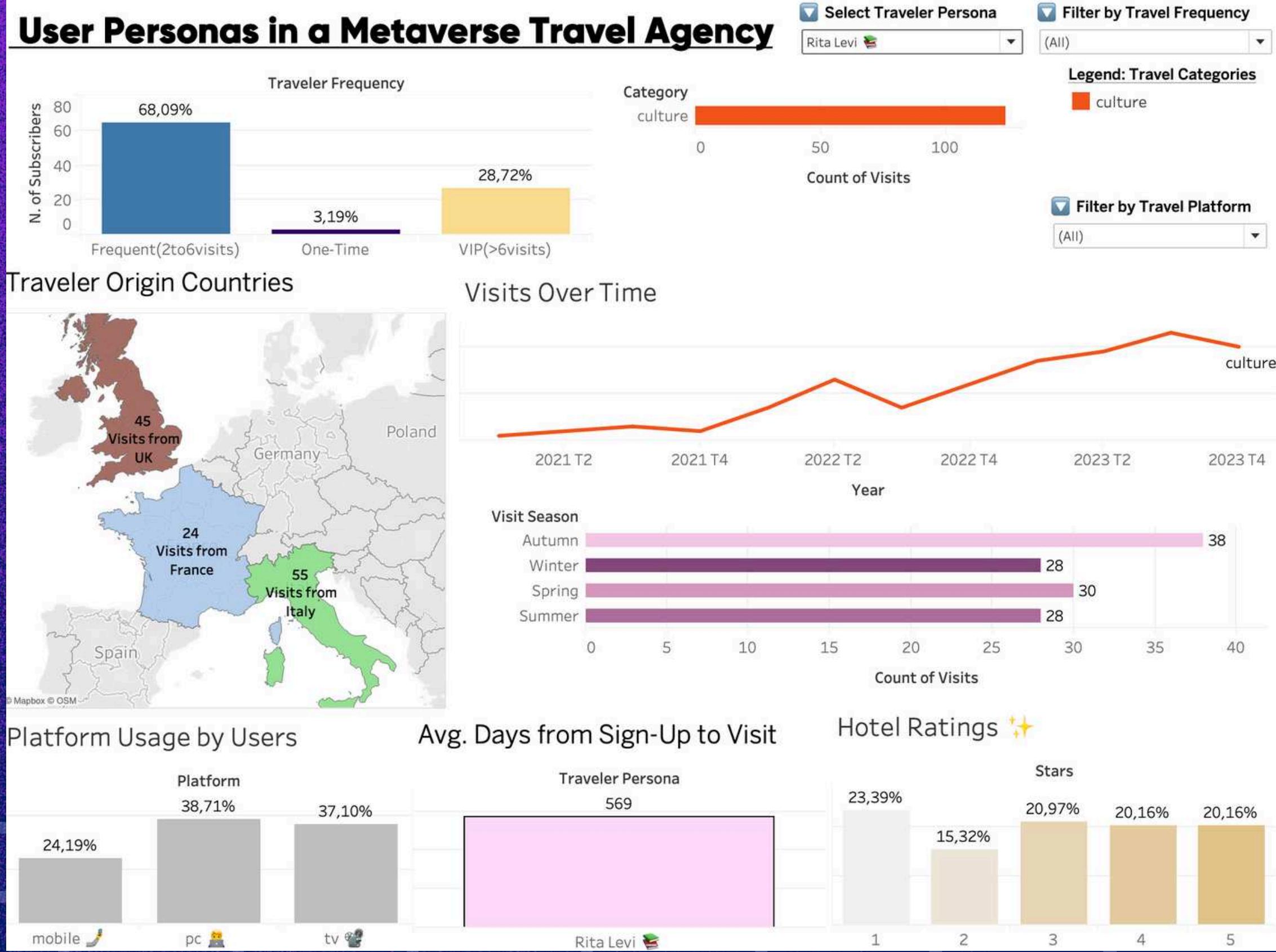


User Personas in a Metaverse Travel Agency



Queen Elizabeth rappresenta il viaggiatore in cerca di esperienze di lusso e relax nel metaverso. La maggior parte degli utenti è frequent flyer (79,02%), con un 17,7% di VIP e un 3,90% di one-time traveler. Le visite sono in costante crescita, con un picco significativo all'inizio del 2023. L'Italia domina con 207 visite, seguita da UK (137) e Francia (75) per un totale di 419 visite, facendo di questa categoria la più richiesta. A differenza degli altri profili, qui l'inverno è la stagione preferita, mentre l'estate è la meno richiesta. L'hotel rating è critico: le valutazioni più alte si concentrano su 1 e 2 stelle (20% e 21,96%), suggerendo la necessità di migliorare la qualità percepita. In media servono 551 giorni dall'iscrizione alla visita, il dato più alto tra i profili analizzati. Anche in questo caso la TV è la piattaforma più usata, ma cresce l'uso di PC e mobile.

RITA LEV-MONTALCINI



Rita Levi-Montalcini rappresenta **il viaggiatore interessato alla cultura, ispirato all'arte e al sapere nel metaverso**. È la categoria con la percentuale più alta (28,72%) di VIP, segno che, pur essendo meno richiesta in generale, **attira utenti più coinvolti e altospendenti**. Le **visite calano verso la fine del 2023, ma resta una tipologia da valorizzare**. La stagione preferita è l'autunno (38 visite), mentre l'estate è la meno scelta. Si registrano 55 visite dall'Italia, 45 dal Regno Unito e 24 dalla Francia per un piccolo **totale di 124 visite**. Il dato più critico riguarda l'hotel rating: **prevalgono le valutazioni a una stella**, segnalando l'urgenza di migliorare l'offerta. Il tempo medio dalla sottoscrizione alla visita è il più alto: 569 giorni. **Un aspetto distintivo è la fruizione**: qui PC (38,71%) supera la TV (37,10%), unico caso tra le personas analizzate e il mobile ha la percentuale più alta tra tutte le categorie (24,19%).



PROPOSTA DI STRATEGIA FUTURA PER IL VIAGGIO NEL METAVERSO

Le analisi delle personas evidenziano quattro profili distinti di viaggiatori, ognuno con esigenze, abitudini e aspettative diverse. Per **ottimizzare l'offerta nel metaverso**, è fondamentale adottare una strategia differenziata e mirata:

- **Personalizzazione dell'esperienza:** valorizzare i Frequent travelers e i VIP, soprattutto nelle categorie culturali e di lusso, con contenuti esclusivi, premi fedeltà e ambienti personalizzati.
- **Ottimizzazione della qualità alberghiera:** intervenire in modo deciso sul rating degli hotel, soprattutto per Queen Elizabeth e Rita Levi-Montalcini, migliorando la percezione e l'effettiva qualità delle strutture.
- **Stagionalità intelligente:** promuovere esperienze legate alle stagioni preferite per ciascun target, sfruttando campagne di marketing stagionale mirate.
- **Sviluppo multipiattaforma:** mantenere la centralità della TV immersiva, ma potenziare l'esperienza su PC e mobile, in particolare per i viaggiatori culturali, che dimostrano una fruizione più autonoma e dinamica.
- **Focus sui clienti altospendenti:** investire sulla categoria Rita Levi-Montalcini, che, seppur più di nicchia, coinvolge utenti fidelizzati e di valore. È una leva strategica da potenziare con contenuti curati e partnership culturali.