In [101…
```
import pandas as pd
df = pd.read_csv("C:\\Users\\jeeva\\Downloads\\archive (12)\\traffic.csv")
```

In [102…
```
df.head()
```

Out[102…

	event	date	country	city	artist	album	track	isrc	linkid
0	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
1	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
2	click	2021-08-21	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	23199824-9cf5-4b98-942a-34965c3b0cc2
3	click	2021-08-21	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	35573248-4e49-47c7-af80-08a960fa74cd
4	click	2021-08-21	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

In [103…
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 226278 entries, 0 to 226277
Data columns (total 9 columns):
 #   Column   Non-Null Count   Dtype 
---  ------   --------------   ----- 
 0   event    226278 non-null  object
 1   date     226278 non-null  object
 2   country  226267 non-null  object
 3   city     226267 non-null  object
 4   artist   226241 non-null  object
 5   album    226273 non-null  object
 6   track    226273 non-null  object
 7   isrc     219157 non-null  object
 8   linkid   226278 non-null  object
dtypes: object(9)
memory usage: 15.5+ MB
```

In [104…
```
df.shape
```

Out[104… (226278, 9)

In [105…
```
df['date'] = pd.to_datetime(df['date'], errors='coerce')
```

In [106…
```
df = df.dropna(subset=['country', 'city', 'artist', 'album', 'track'])
```

In [107…
```
df = df.drop_duplicates()
```

In [108…
```
print(df.isnull().sum())
```

```
event         0
date          0
country       0
city          0
artist        0
album         0
track         0
isrc       6277
linkid        0
dtype: int64
```

In [109…
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 122534 entries, 0 to 226274
Data columns (total 9 columns):
 #   Column   Non-Null Count   Dtype         
---  ------   --------------   -----         
 0   event    122534 non-null  object        
 1   date     122534 non-null  datetime64[ns]
 2   country  122534 non-null  object        
 3   city     122534 non-null  object        
 4   artist   122534 non-null  object        
 5   album    122534 non-null  object        
 6   track    122534 non-null  object        
 7   isrc     116257 non-null  object        
 8   linkid   122534 non-null  object        
dtypes: datetime64[ns](1), object(8)
memory usage: 9.3+ MB
```

In [110…
```
event_counts = df['event'].value_counts()
print("Event Counts:\n", event_counts)
```

```
Event Counts:
 event
pageview    73338
click       32489
preview     16707
Name: count, dtype: int64
```

In [111…
```
top_countries = df['country'].value_counts().head(10)
print("Top Countries:\n", top_countries)
```

```
Top Countries:
 country
United States     28640
India             18689
France            10565
Saudi Arabia       7682
United Kingdom     5095
Germany            4015
Canada             2784
Pakistan           2633
Iraq               2444
Turkey             2399
Name: count, dtype: int64
```

In [112…
```
top_artists = df['artist'].value_counts().head(10)
print("Top Artists:\n", top_artists)
```

```
Top Artists:
 artist
Tesher                                                    8288
Anne-Marie                                                4029
Tundra Beats                                              3951
Roddy Ricch                                               3107
Olivia Rodrigo                                            3037
Surf Mesa, Emilee                                         2956
DMNDS, Strange Fruits Music, Fallen Roses, Lujavo, Nito-Onna    2865
Reyanna Maria                                             2672
PinkPantheress                                            2446
50 Cent, Olivia                                           2390
Name: count, dtype: int64
```

In [113…
```
top_tracks = df['track'].value_counts().head(10)
print("Top Tracks:\n", top_tracks)
```

```
Top Tracks:
 track
Jalebi Baby                              8288
Beautiful                                4037
Beautiful Day                            3951
Late At Night                            3059
ily (i love you baby) (feat. Emilee)     2956
Calabria (feat. Lujavo & Nito-Onna)      2865
So Pretty                                2827
Candy Shop                               2397
Summer of Love (Shawn Mendes & Tainy)    2108
Build a Bitch                            2072
Name: count, dtype: int64
```

In [114…
```
daily_activity = df['date'].value_counts().sort_index()
print("Daily Activity:\n", daily_activity.head())
```

```
Daily Activity:
 date
2021-08-19    21143
2021-08-20    18522
2021-08-21    16701
2021-08-22    16927
2021-08-23    16412
Name: count, dtype: int64
```

In [115…
```
df.groupby('linkid')['event'].count().describe()

# Number of unique event types per session (complexity of user flow)
df.groupby('linkid')['event'].nunique().value_counts()
```

Out[115…
```
event
2    1811
1    1547
3     464
Name: count, dtype: int64
```

In [116…
```
df['country'].value_counts().head(10)
```

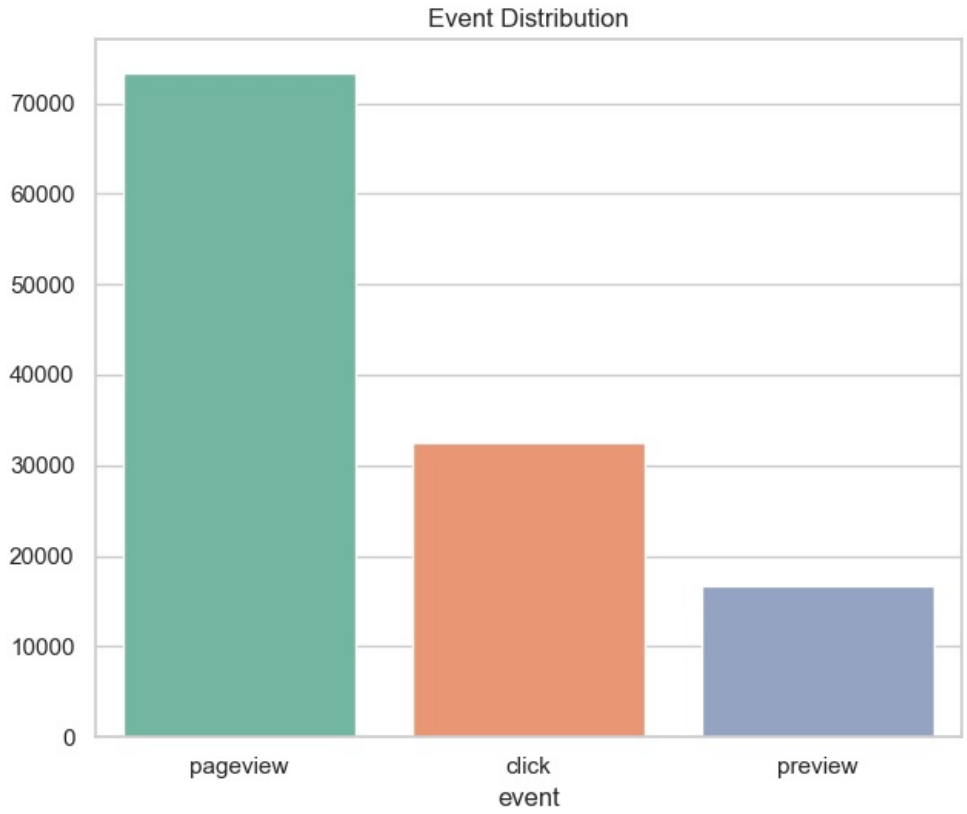
```
df['city'].value_counts().head(10)
```

Out[116…
```
city
Unknown        8790
Jeddah         2497
Riyadh         2232
Hyderabad      1088
Dammam         1002
Delhi           884
Jaipur          849
Lucknow         837
Kuwait City     816
Ahmedabad       808
Name: count, dtype: int64
```

In [117…
```
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style="whitegrid")
plt.figure(figsize=(16, 20))
plt.subplot(3, 2, 1)
sns.barplot(x=event_counts.index, y=event_counts.values, palette="Set2")
plt.title("Event Distribution")
```

```
C:\Users\jeeva\AppData\Local\Temp\ipykernel_6404\3369721957.py:6: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable
to `hue` and set `legend=False` for the same effect.
```

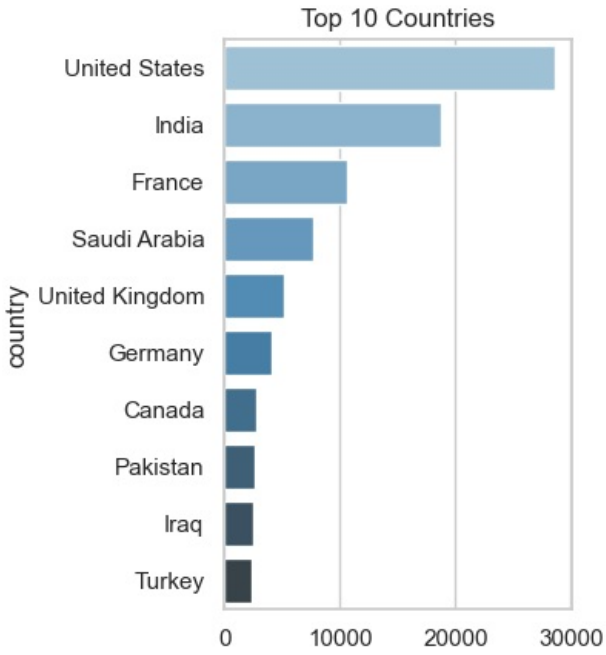
Out[117… Text(0.5, 1.0, 'Event Distribution')



In [118…
```
plt.subplot(1, 2, 2)
sns.barplot(x=top_countries.values, y=top_countries.index, palette="Blues_d")
plt.title("Top 10 Countries")
```

```
C:\Users\jeeva\AppData\Local\Temp\ipykernel_6404\173644326.py:2: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable
to `hue` and set `legend=False` for the same effect.
```

Out[118… Text(0.5, 1.0, 'Top 10 Countries')



In [119…
```
# Top Artists
plt.subplot(2, 2, 3)
sns.barplot(x=top_artists.values, y=top_artists.index, palette="Purples_d")
plt.title("Top 10 Artists")
```

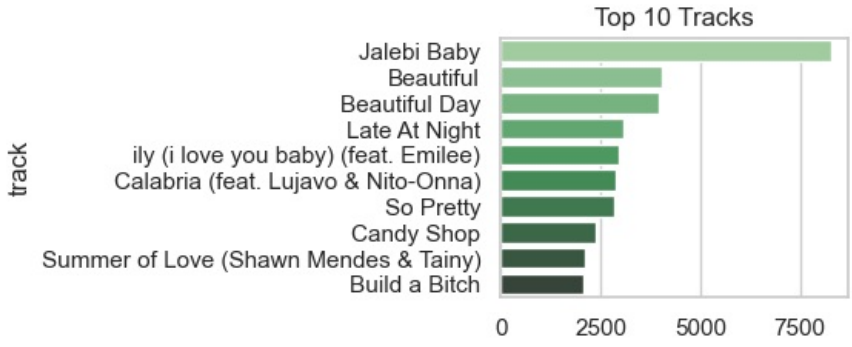
```
C:\Users\jeeva\AppData\Local\Temp\ipykernel_6404\106171222.py:3: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable
to `hue` and set `legend=False` for the same effect.
```

Out[119… Text(0.5, 1.0, 'Top 10 Artists')

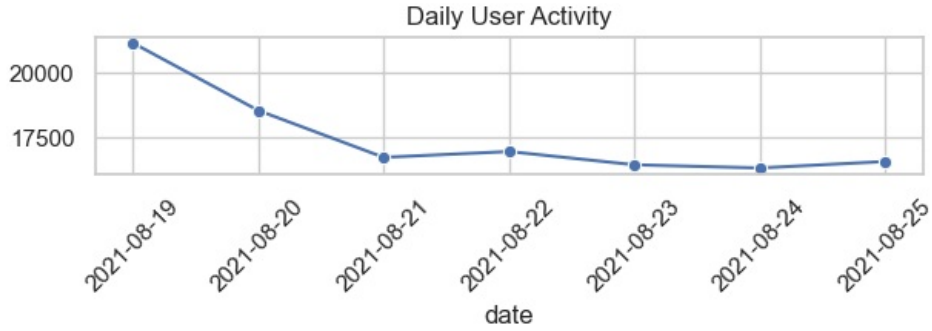
In [120…
```
# Top Tracks
plt.subplot(2, 2, 4)
sns.barplot(x=top_tracks.values, y=top_tracks.index, palette="Greens_d")
plt.title("Top 10 Tracks")
```

```
C:\Users\jeeva\AppData\Local\Temp\ipykernel_6404\131125717.py:3: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable
to `hue` and set `legend=False` for the same effect.
```

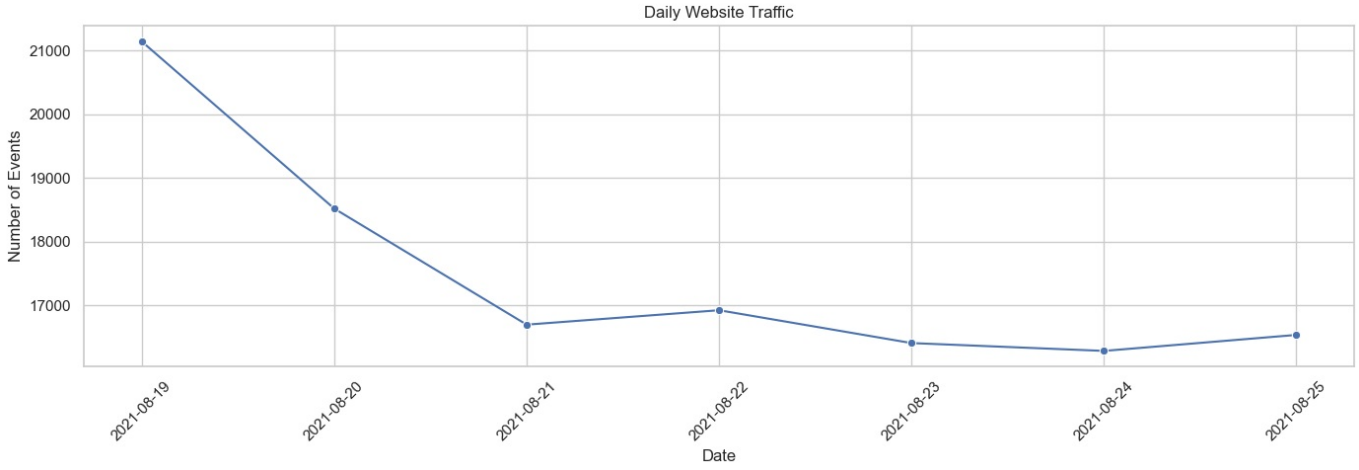
Out[120… Text(0.5, 1.0, 'Top 10 Tracks')



In [121…
```
# Daily Activity Line Plot
plt.subplot(3, 1, 3)
sns.lineplot(x=daily_activity.index, y=daily_activity.values, marker="o")
plt.title("Daily User Activity")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



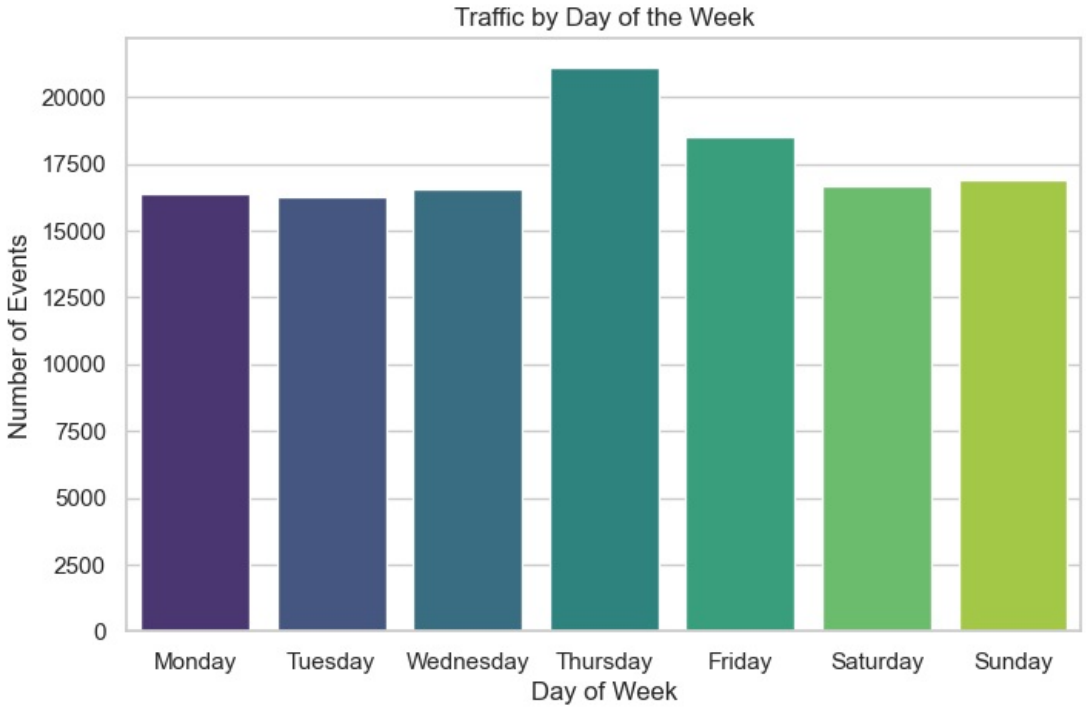
In [122…
```
# Daily event counts
daily_traffic = df['date'].dt.date.value_counts().sort_index()
plt.figure(figsize=(14, 5))
sns.lineplot(x=daily_traffic.index, y=daily_traffic.values, marker='o')
plt.title("Daily Website Traffic")
plt.xlabel("Date")
plt.ylabel("Number of Events")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```



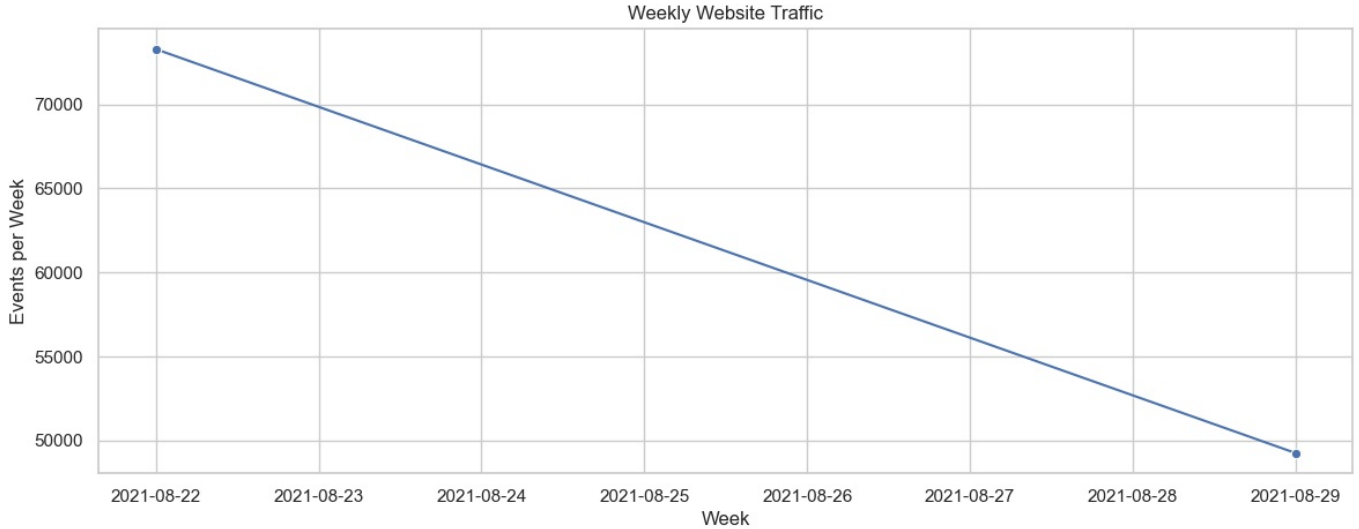
In [123…
```
df['day_of_week'] = df['date'].dt.day_name()

# Count by day of week
dow_traffic = df['day_of_week'].value_counts().reindex([
    'Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday'
])

# Plot
```

```
plt.figure(figsize=(8, 5))
sns.barplot(x=dow_traffic.index, y=dow_traffic.values, palette="viridis")
plt.title("Traffic by Day of the Week")
plt.ylabel("Number of Events")
plt.xlabel("Day of Week")
plt.show()
```

```
C:\Users\jeeva\AppData\Local\Temp\ipykernel_6404\3060480665.py:10: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable
to `hue` and set `legend=False` for the same effect.
```



In [124…
```
df.set_index('date', inplace=True)

# Weekly trend
weekly_traffic = df.resample('W').size()

plt.figure(figsize=(14, 5))
sns.lineplot(x=weekly_traffic.index, y=weekly_traffic.values, marker="o")
plt.title("Weekly Website Traffic")
plt.ylabel("Events per Week")
plt.xlabel("Week")
plt.show()

# Optional: Monthly
monthly_traffic = df.resample('M').size()
```



```
C:\Users\jeeva\AppData\Local\Temp\ipykernel_6404\1357657267.py:14: FutureWarning:

'M' is deprecated and will be removed in a future version, please use 'ME' instead.
```

In [125…
```
print(df.columns.tolist())
```

```
['event', 'country', 'city', 'artist', 'album', 'track', 'isrc', 'linkid', 'day_of_week']
```

In [126…
```
import plotly.graph_objects as go
df['event'] = df['event'].str.lower()
session_events = df.groupby('linkid')['event'].apply(set)
stage1 = session_events.apply(lambda x: 'pageview' in x).sum()
stage2 = session_events.apply(lambda x: {'pageview', 'preview'}.issubset(x)).sum()
stage3 = session_events.apply(lambda x: {'pageview', 'preview', 'click'}.issubset(x)).sum()

# Visualize funnel
funnel_values = [stage1, stage2, stage3]
funnel_labels = ['Pageview', 'Preview', 'Click']

fig = go.Figure(go.Funnel(
    y=funnel_labels,
    x=funnel_values,
    textinfo="value+percent initial+percent previous",
    marker=dict(color=["#636EFA", "#00CC96", "#EF553B"])
))
fig.update_layout(title="Conversion Funnel")
fig.show()
```

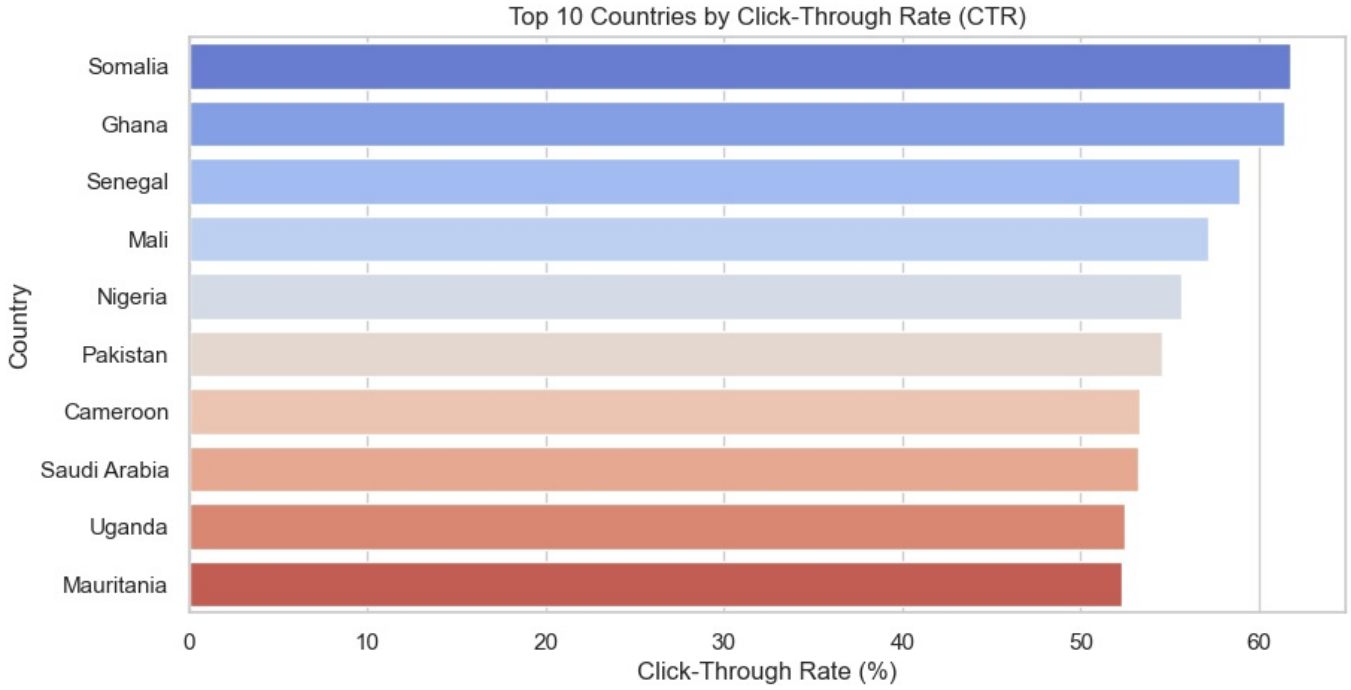
In [127…
```
conversion_rate = stage3 / stage1 * 100 if stage1 else 0
dropoff_preview = (stage1 - stage2) / stage1 * 100 if stage1 else 0
dropoff_click = (stage2 - stage3) / stage2 * 100 if stage2 else 0

print(f"Drop-off after Preview: {dropoff_preview:.2f}%")
print(f"Drop-off before Click: {dropoff_click:.2f}%")
print(f"Final Conversion Rate: {conversion_rate:.2f}% (Pageview → Click)")
```

```
Drop-off after Preview: 87.04%
Drop-off before Click: 6.26%
Final Conversion Rate: 12.15% (Pageview → Click)
```

In [128…
```
# Normalize event text
df['event'] = df['event'].str.lower()

# Create a country vs event count pivot
country_event_pivot = df.pivot_table(index='country', columns='event', values='linkid', aggfunc='count', fill_va

# Calculate click-through rate (CTR)
country_event_pivot['click_rate'] = (country_event_pivot['click'] / country_event_pivot['pageview']) * 100

# Top 10 countries by click rate (with enough data)
```

```
top_ctr = country_event_pivot[country_event_pivot['pageview'] > 50].sort_values('click_rate', ascending=False).

plt.figure(figsize=(10, 5))
sns.barplot(x=top_ctr['click_rate'], y=top_ctr.index, palette="coolwarm")
plt.title("Top 10 Countries by Click-Through Rate (CTR)")
plt.xlabel("Click-Through Rate (%)")
plt.ylabel("Country")
plt.show()
```

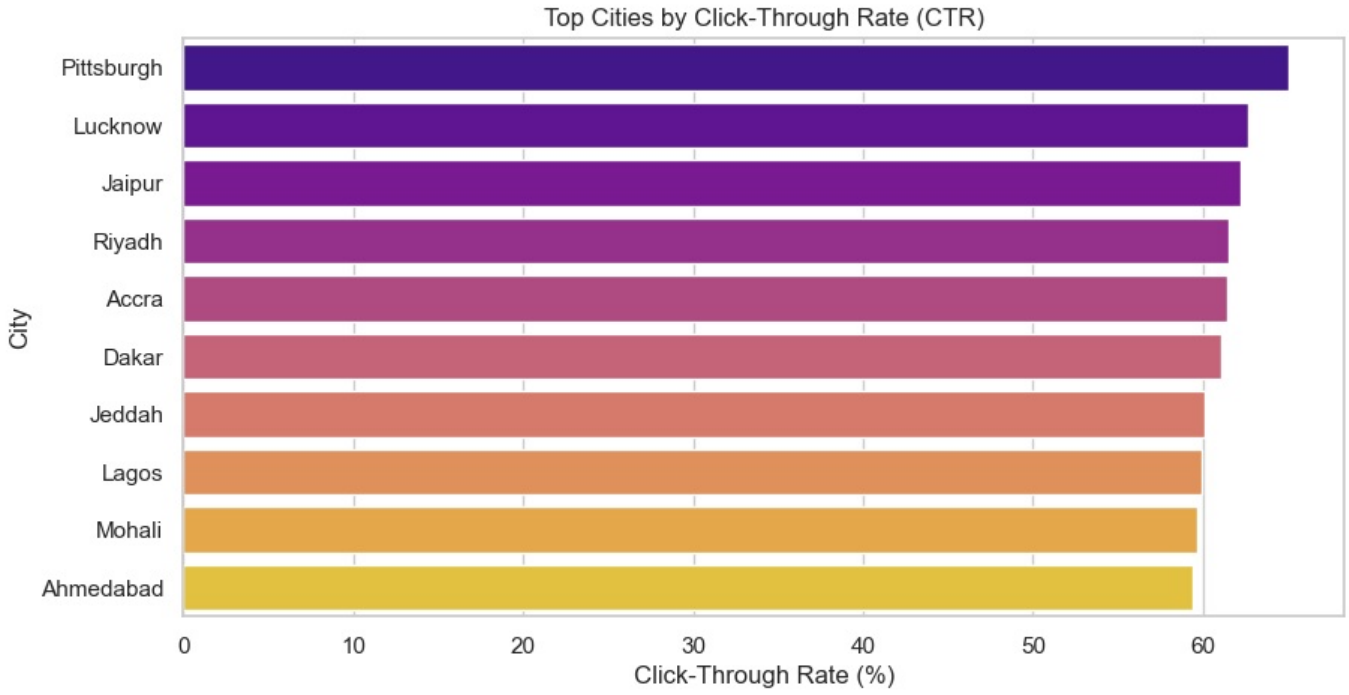
```
C:\Users\jeeva\AppData\Local\Temp\ipykernel_6404\4241633583.py:14: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable
to `hue` and set `legend=False` for the same effect.
```



In [129…

```
city_event_pivot = df.pivot_table(index='city', columns='event', values='linkid', aggfunc='count', fill_value=0
city_event_pivot['click_rate'] = (city_event_pivot['click'] / city_event_pivot['pageview']) * 100
top_city_ctr = city_event_pivot[city_event_pivot['pageview'] > 50].sort_values('click_rate', ascending=False).h

plt.figure(figsize=(10, 5))
sns.barplot(x=top_city_ctr['click_rate'], y=top_city_ctr.index, palette="plasma")
plt.title("Top Cities by Click-Through Rate (CTR)")
plt.xlabel("Click-Through Rate (%)")
plt.ylabel("City")
plt.show()
```

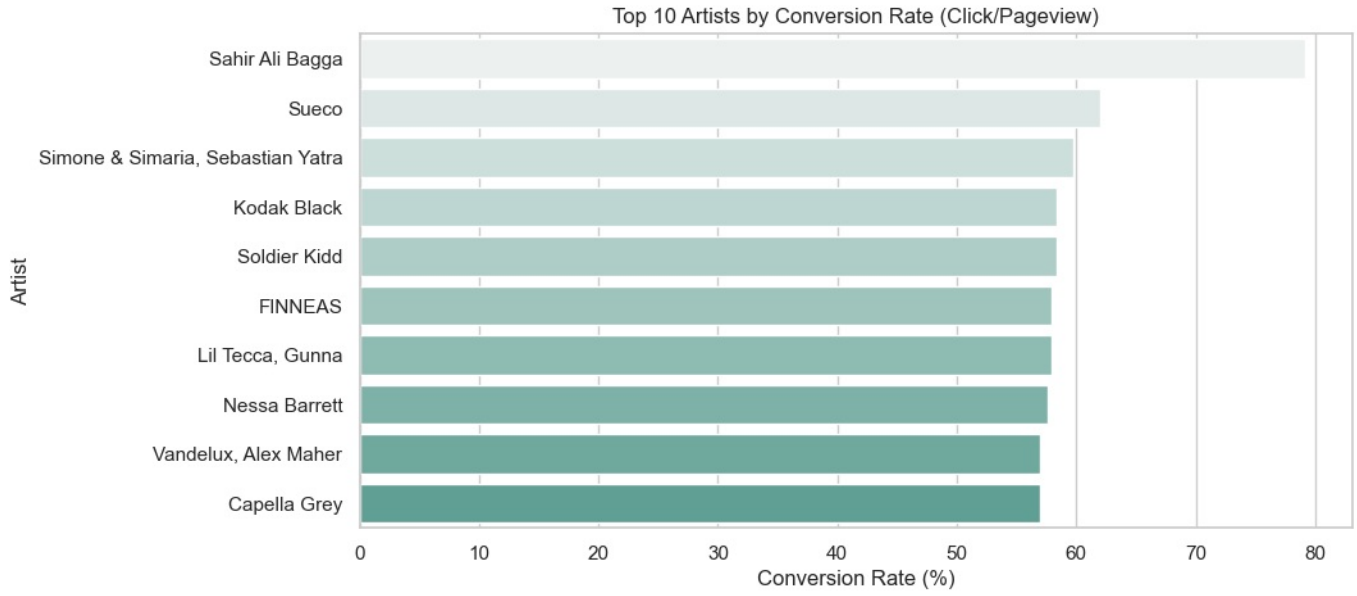
```
C:\Users\jeeva\AppData\Local\Temp\ipykernel_6404\3677565223.py:6: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable
to `hue` and set `legend=False` for the same effect.
```



In [130…

```
artist_event_pivot = df.pivot_table(index='artist', columns='event', values='linkid', aggfunc='count', fill_valu
artist_event_pivot['conversion_rate'] = (artist_event_pivot['click'] / artist_event_pivot['pageview']) * 100

top_converting_artists = artist_event_pivot[artist_event_pivot['pageview'] > 50].sort_values('conversion_rate',

plt.figure(figsize=(10, 5))
sns.barplot(x=top_converting_artists['conversion_rate'], y=top_converting_artists.index, palette="light:#5A9")
plt.title("Top 10 Artists by Conversion Rate (Click/Pageview)")
plt.xlabel("Conversion Rate (%)")
plt.ylabel("Artist")
plt.show()
```

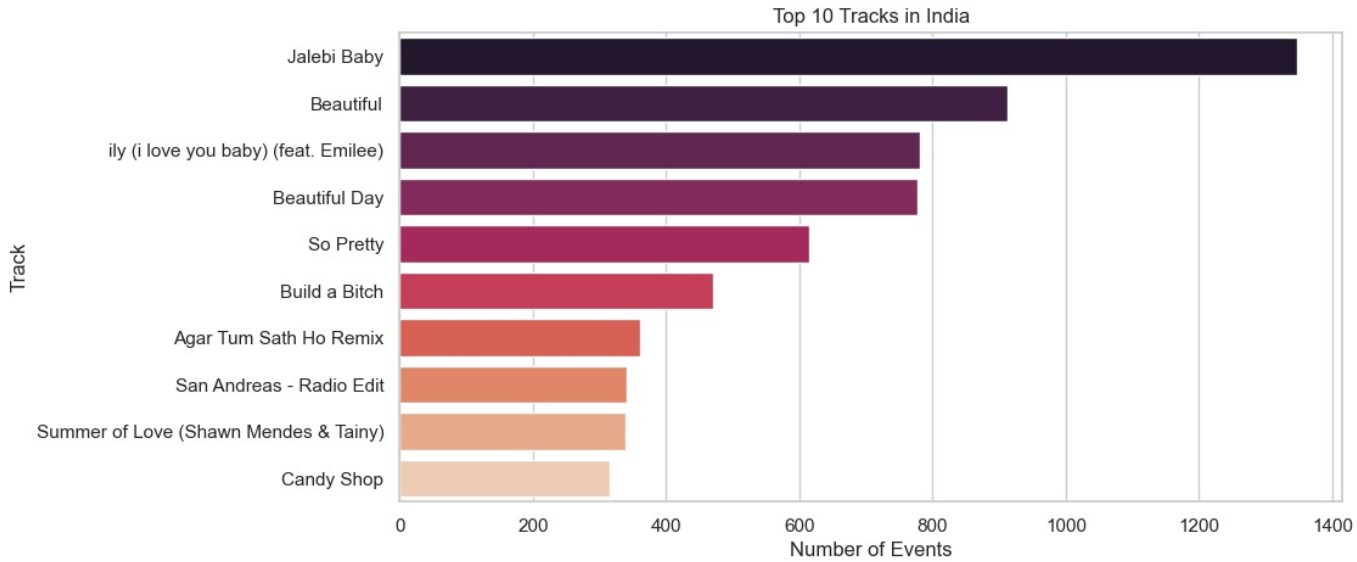
```
C:\Users\jeeva\AppData\Local\Temp\ipykernel_6404\3592157901.py:7: FutureWarning:



Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable
to `hue` and set `legend=False` for the same effect.
```



In [131…

```
country_name = 'India'
top_tracks_in_country = df[df['country'] == country_name]['track'].value_counts().head(10)

plt.figure(figsize=(10, 5))
sns.barplot(x=top_tracks_in_country.values, y=top_tracks_in_country.index, palette='rocket')
plt.title(f"Top 10 Tracks in {country_name}")
plt.xlabel("Number of Events")
plt.ylabel("Track")
plt.show()
```

```
C:\Users\jeeva\AppData\Local\Temp\ipykernel_6404\2226528957.py:5: FutureWarning:


Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable
to `hue` and set `legend=False` for the same effect.
```



In [ ]: