

```
from google.colab import files
uploaded = files.upload()

Choose Files traffic.csv
• traffic.csv(text/csv) - 30630222 bytes, last modified: 7/7/2025 - 100% done
Saving traffic.csv to traffic (1).csv
```

```
import pandas as pd
```

```
# Load the uploaded file
df = pd.read_csv('traffic.csv')
```

```
# Display structure
print(df.info())
print(df.head())
```

```
<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
None
event      date      country  city \
0 click  2021-08-21  Saudi Arabia  Jeddah
1 click  2021-08-21  Saudi Arabia  Jeddah
2 click  2021-08-21      India  Ludhiana
3 click  2021-08-21      France  Unknown
4 click  2021-08-21    Maldives    Malé

          artist      album      track \
0      Teshher  Jalebi Baby  Jalebi Baby
1      Teshher  Jalebi Baby  Jalebi Baby
2      Reyanna Maria  So Pretty  So Pretty
3 Simone & Simaria, Sebastian Yatra  No Llores Más  No Llores Más
4      Teshher  Jalebi Baby  Jalebi Baby

          isrc      linkid
0  QZNWQ2070741  2d896d31-97b6-4869-967b-1c5fb9cd4bb8
1  QZNWQ2070741  2d896d31-97b6-4869-967b-1c5fb9cd4bb8
2  USUM72100871  23199824-9cf5-4b98-942a-34965c3b0cc2
3  BRUM72003904  35573248-4e49-47c7-af80-08a960fa74cd
4  QZNWQ2070741  2d896d31-97b6-4869-967b-1c5fb9cd4bb8
```

```
# Check for missing values
print(df.isnull().sum())

# Drop rows missing key geographic or artist info
df = df.dropna(subset=['country', 'city', 'artist'])

# Convert 'date' to datetime
df['date'] = pd.to_datetime(df['date'])

# Fill other missing values with 'Unknown'
df = df.fillna('Unknown')
```

```
event      0
date       0
country    11
city       11
```

What can I help you build?

traffic.csv ×

1 to 10 of 226278 entries Filter

event	date	country	city	artist	album
click	2021-08-21	Saudi Arabia	Jeddah	Teshher	Jalebi Baby
click	2021-08-21	Saudi Arabia	Jeddah	Teshher	Jalebi Baby
click	2021-08-21	India	Ludhiana	Reyanna Maria	So Pretty
click	2021-08-21	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más
click	2021-08-21	Maldives	Malé	Teshher	Jalebi Baby
click	2021-08-21	United States	Los Angeles	KenTheMan	I'm Perfect
click	2021-08-21	Sierra Leone	Unknown	Coi Leray	No More Parties
click	2021-08-21	United Kingdom	Stockport	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Uzi Vert)
click	2021-08-21	United States	Pflugerville	Billie Eilish	Happier Than Ever
click	2021-08-21	France	Unknown	DMNDS, Strange Fruits Music, Fallen Roses, Lujavo, Nito-Onna	Calab (feat. Lujavo & Nito-Onna)

Show 10 per page

121010010002000022000226002262022628

```

artist      37
album       5
track       5
isrc        7121
linkid      0
dtype: int64

# Top 10 countries
print("Top 10 Countries:\n", df['country'].value_counts().head(10))

# Top 10 cities
print("Top 10 Cities:\n", df['city'].value_counts().head(10))

# Top 10 artists
print("Top 10 Artists:\n", df['artist'].value_counts().head(10))

# Time period covered
print("Date range:", df['date'].min(), "to", df['date'].max())

```

```

↗ Top 10 Countries:
  country
Saudi Arabia      47334
India              42992
United States     32525
France            15661
Iraq              8260
United Kingdom    5970
Pakistan          5644
Germany           4792
United Arab Emirates 3702
Turkey            3514
Name: count, dtype: int64
Top 10 Cities:
  city
Jeddah          22791
Riyadh           17454
Unknown         16653
Hyderabad        3682
Dammam           3276
Erbil            3105
Kuwait City      3010
Lucknow          2808
Muscat           2717
Delhi            2567
Name: count, dtype: int64
Top 10 Artists:
  artist
Teshar                    40841
Anne-Marie                10650
Tundra Beats              9751
Surf Mesa, Emilee         7533
DMNDS, Strange Fruits Music, Fallen Roses, Lujavo, Nito-Onna 5512
Reyanna Maria             5434
Shawn Mendes, Tainy       5409
50 Cent, Olivia           5367
Roddy Ricch               5093
Olivia Rodrigo            4115
Name: count, dtype: int64
Date range: 2021-08-19 00:00:00 to 2021-08-25 00:00:00

```

```

import matplotlib.pyplot as plt

daily_traffic = df.groupby('date').size()

plt.figure(figsize=(12,6))
daily_traffic.plot()
plt.title("Daily Website Clicks")
plt.xlabel("Date")
plt.ylabel("Number of Clicks")
plt.grid(True)
plt.show()

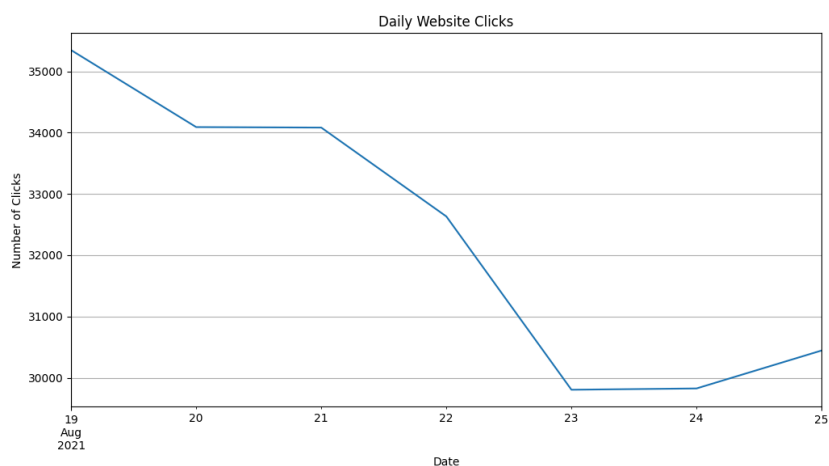
```



McAfee | WebAdvisor



Your download's being scanned.
We'll let you know if there's an issue.



```
import seaborn as sns
```

```
top_countries = df['country'].value_counts().head(10)
```

```
plt.figure(figsize=(10,6))
```

```
sns.barplot(x=top_countries.values, y=top_countries.index, palette='viridis')
```

```
plt.title("Top 10 Countries by Website Clicks")
```

```
plt.xlabel("Clicks")
```

```
plt.ylabel("Country")
```

```
plt.show()
```

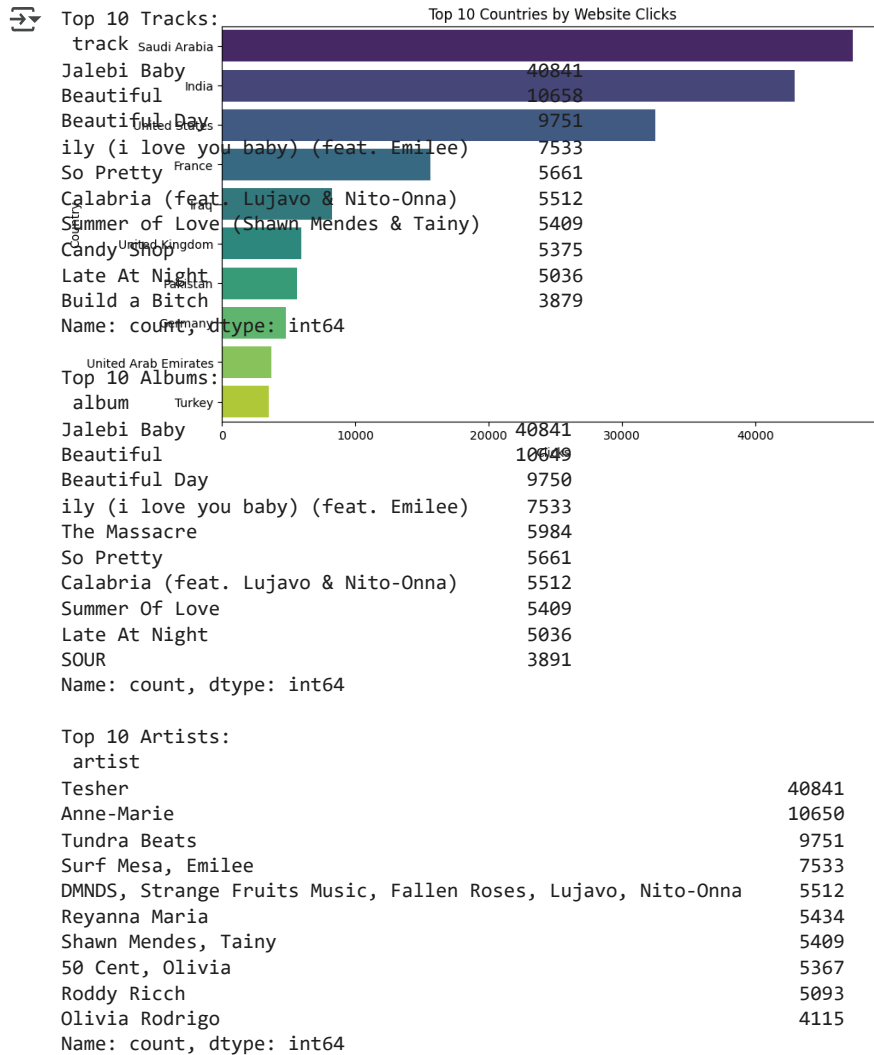


McAfee | WebAdvisor



Your download's being scanned.
We'll let you know if there's an issue.

```
print("Top 10 Tracks:\n", df['track'].value_counts().head(10))
print("\nTop 10 Albums:\n", df['album'].value_counts().head(10))
print("\nTop 10 Artists:\n", df['artist'].value_counts().head(10))
```



```
df.to_csv('cleaned_traffic.csv', index=False)
from google.colab import files
files.download('cleaned_traffic.csv')
```



McAfee | WebAdvisor



Your download's being scanned.
We'll let you know if there's an issue.