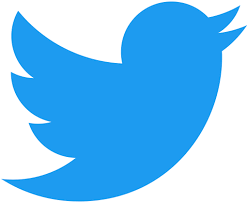
# **Social Media Sentiments Analysis**

U,{8541b970-1bf3-463b-a16d-2288c647dd48}{45},10.520833333333334,10.520833333333334



linkcode

## **Goal of the Project:** In this project, we aim to develop a Social Media Sentiment Analysis system to assist organizations, including platforms like Facebook and Twitter. The objective is to understand and analyze user sentiments, enabling the organization to better serve users by providing the information or objects they are searching for.

## Import Libraries

*# For analysis and cleaning*  
import pandas as pd  
import numpy as np  
*# for visaluatztion*  
import matplotlib.pyplot as plt  
import seaborn as sns

df=pd.read\_csv('/kaggle/input/social-media-sentiments-analysis-dataset/sentimentdataset.csv')  
df.head()

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Unnamed: 0.1 | Unnamed: 0 | Text | Sentiment | Timestamp | User | Platform | Hashtags | Retweets | Likes | Country | Year | Month | Day | Hour |
| 0 | 0 | 0 | Enjoying a beautiful day at the park! ... | Positive | 2023-01-15 12:30:00 | User123 | Twitter | #Nature #Park | 15.0 | 30.0 | USA | 2023 | 1 | 15 | 12 |
| 1 | 1 | 1 | Traffic was terrible this morning. ... | Negative | 2023-01-15 08:45:00 | CommuterX | Twitter | #Traffic #Morning | 5.0 | 10.0 | Canada | 2023 | 1 | 15 | 8 |
| 2 | 2 | 2 | Just finished an amazing workout! 💪 ... | Positive | 2023-01-15 15:45:00 | FitnessFan | Instagram | #Fitness #Workout | 20.0 | 40.0 | USA | 2023 | 1 | 15 | 15 |
| 3 | 3 | 3 | Excited about the upcoming weekend getaway! ... | Positive | 2023-01-15 18:20:00 | AdventureX | Facebook | #Travel #Adventure | 8.0 | 15.0 | UK | 2023 | 1 | 15 | 18 |
| 4 | 4 | 4 | Trying out a new recipe for dinner tonight. ... | Neutral | 2023-01-15 19:55:00 | ChefCook | Instagram | #Cooking #Food | 12.0 | 25.0 | Australia | 2023 | 1 | 15 | 19 |

df.shape

(732, 15)

*# drop column Unnamed: 0.1*  
df.drop(columns='Unnamed: 0.1',inplace=True)

*# Rename column Unnamed: 0 to id*  
df.rename(columns={'Unnamed: 0':'Id'},inplace=True)

df.isnull().sum()

Id 0  
Text 0  
Sentiment 0  
Timestamp 0  
User 0  
Platform 0  
Hashtags 0  
Retweets 0  
Likes 0  
Country 0  
Year 0  
Month 0  
Day 0  
Hour 0  
dtype: int64

df.dtypes

Id int64  
Text object  
Sentiment object  
Timestamp object  
User object  
Platform object  
Hashtags object  
Retweets float64  
Likes float64  
Country object  
Year int64  
Month int64  
Day int64  
Hour int64  
dtype: object

df['Timestamp'] = pd.to\_datetime(df['Timestamp'])  
df['Day'] = df['Timestamp'].dt.day  
df['Month'] = df['Timestamp'].dt.month  
df['Year'] = df['Timestamp'].dt.year

df['Text']= df['Text'].str.strip()  
df['Sentiment']= df['Sentiment'].str.strip()  
df['User']= df['User'].str.strip()  
df['Platform']= df['Platform'].str.strip()  
df['Hashtags']= df['Hashtags'].str.strip()  
df['Country']= df['Country'].str.strip()

## EDA

df.head(1)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Id | Text | Sentiment | Timestamp | User | Platform | Hashtags | Retweets | Likes | Country | Year | Month | Day | Hour |
| 0 | 0 | Enjoying a beautiful day at the park! | Positive | 2023-01-15 12:30:00 | User123 | Twitter | #Nature #Park | 15.0 | 30.0 | USA | 2023 | 1 | 15 | 12 |

df['Sentiment'].value\_counts().nlargest(10).plot(kind='bar')  
plt.title('Top 10 Sentiments based on Text')  
plt.xlabel('Sentiment')  
plt.ylabel('Count')  
plt.show()

U,{8900b0b5-01d7-4264-ae3b-b72cb8abb9e3}{198},11.708333333333334,11.083333333333334

df['Platform'].value\_counts()

Platform  
Instagram 258  
Twitter 243  
Facebook 231  
Name: count, dtype: int64

df['Platform'].value\_counts().plot(kind='pie', autopct='**%1.1f%%**')  
plt.title('Percentages of Platforms')  
plt.legend()  
plt.show()

U,{ab0d067d-c664-499a-81f6-632a8f0336b9}{4},8.604166666666666,8.5625

df['Country'].value\_counts().nlargest(10).plot(kind='bar')  
plt.title('Top 10 Country')  
plt.legend()  
plt.show()

U,{ab0d067d-c664-499a-81f6-632a8f0336b9}{35},11.5,10.520833333333334

df['Hashtags'].value\_counts().nlargest(10).plot(kind='bar')  
plt.title('Top 10 Hashtags')  
plt.xlabel('Hashtags')  
plt.ylabel('Count')  
plt.legend()  
plt.show()

U,{ab0d067d-c664-499a-81f6-632a8f0336b9}{66},11.8125,14.895833333333334

*# Descrip data nurimucal*  
df.describe()

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Id | Timestamp | Retweets | Likes | Year | Month | Day | Hour |
| count | 732.000000 | 732 | 732.000000 | 732.000000 | 732.000000 | 732.000000 | 732.000000 | 732.000000 |
| mean | 369.740437 | 2020-12-08 21:37:38.196721408 | 21.508197 | 42.901639 | 2020.471311 | 6.122951 | 15.497268 | 15.521858 |
| min | 0.000000 | 2010-05-15 15:30:00 | 5.000000 | 10.000000 | 2010.000000 | 1.000000 | 1.000000 | 0.000000 |
| 25% | 185.750000 | 2019-03-08 17:27:30 | 17.750000 | 34.750000 | 2019.000000 | 3.000000 | 9.000000 | 13.000000 |
| 50% | 370.500000 | 2021-09-20 14:30:00 | 22.000000 | 43.000000 | 2021.000000 | 6.000000 | 15.000000 | 16.000000 |
| 75% | 553.250000 | 2023-02-26 11:48:45 | 25.000000 | 50.000000 | 2023.000000 | 9.000000 | 22.000000 | 19.000000 |
| max | 736.000000 | 2023-10-22 20:45:00 | 40.000000 | 80.000000 | 2023.000000 | 12.000000 | 31.000000 | 23.000000 |
| std | 212.428936 | NaN | 7.061286 | 14.089848 | 2.802285 | 3.411763 | 8.474553 | 4.113414 |

numerical\_columns = df[['Day', 'Month', 'Year', 'Likes', 'Retweets']]  
  
for col **in** numerical\_columns.columns:  
 print(f"Minimum **{**col**}**: **{**df[col].min()**}** | Maximum **{**col**}**: **{**df[col].max()**}**")

Minimum Day: 1 | Maximum Day: 31  
Minimum Month: 1 | Maximum Month: 12  
Minimum Year: 2010 | Maximum Year: 2023  
Minimum Likes: 10.0 | Maximum Likes: 80.0  
Minimum Retweets: 5.0 | Maximum Retweets: 40.0

# Relationships between columns

### Q1:Top 10 hashtags retweeted

# **Answer**

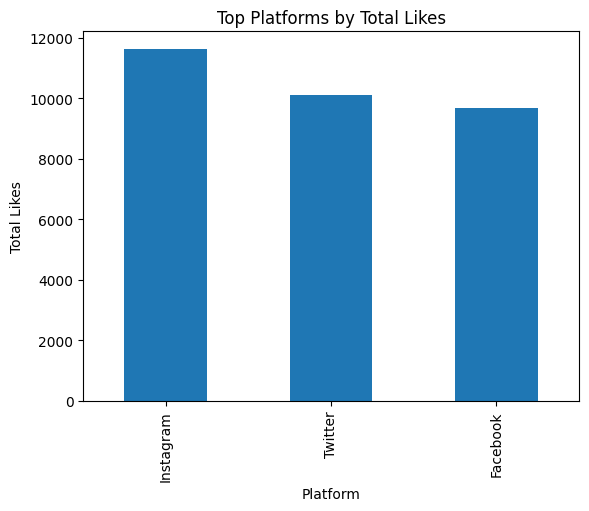
H\_R=df.groupby('Hashtags')['Retweets'].max().nlargest(10).sort\_values(ascending=False).plot(kind='bar')  
plt.title('Top 10 hashtags retweeted')  
plt.xlabel('Hashtags')  
plt.ylabel('count')  
plt.show()

U,{bee76fcc-2885-409c-8426-a446e1498e3b}{40},11.708333333333334,14.416666666666666

### Q2: What is platform top liked by users?

## **Answer**

top\_likes\_platform = df.groupby('Platform')['Likes'].sum().nlargest(10)  
top\_likes\_platform.plot(kind='bar')  
plt.title('Top Platforms by Total Likes')  
plt.xlabel('Platform')  
plt.ylabel('Total Likes')  
plt.show()



### Q3: What is the most country liked posts

## **Answer**

top\_country\_likes=df.groupby('Country')['Likes'].sum().nlargest(10)  
top\_country\_likes.plot(kind='bar')  
plt.title('Top country likes')  
plt.xlabel('Country')  
plt.ylabel('count')  
plt.show()

U,{bee76fcc-2885-409c-8426-a446e1498e3b}{170},12.083333333333334,10.520833333333334

## I will segment users for each platform

Facebook=df[df['Platform']=='Facebook']  
Twitter=df[df['Platform']=='Twitter']  
Instagram=df[df['Platform']=='Instagram']

# **Facebook**

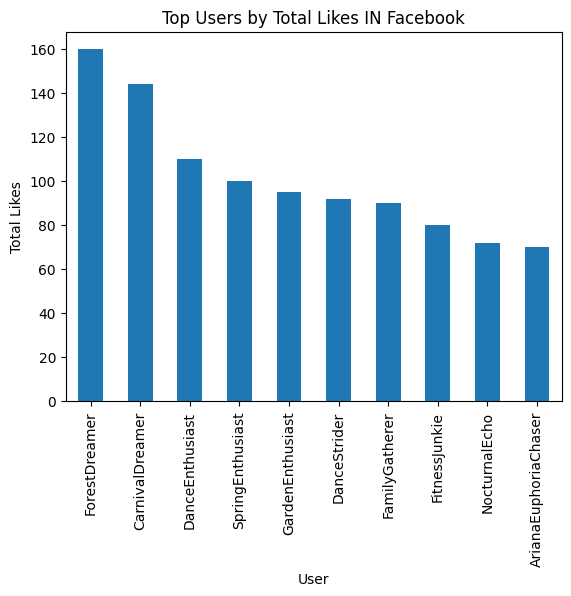
### Top 10 hashtags retweeted

H\_R\_f=Facebook.groupby('Hashtags')['Retweets'].max().nlargest(10).sort\_values(ascending=False)  
H\_R\_f.plot(kind='bar')  
plt.title('Top 10 hashtags retweeted in $/ Facebook $/')  
plt.xlabel('Hashtags')  
plt.ylabel('count')  
plt.show()

U,{adb7a784-d279-4b11-bf2f-5269ee84b2b9}{14},11.708333333333334,14.75

### Q2: Who User liked mostly ?

top\_likes\_platform\_F = Facebook.groupby('User')['Likes'].sum().nlargest(10)  
top\_likes\_platform\_F.plot(kind='bar')  
plt.title('Top Users by Total Likes IN Facebook')  
plt.xlabel('User')  
plt.ylabel('Total Likes')  
plt.show()



f = Facebook.groupby('Year')['Likes'].sum().reset\_index()  
plt.figure(figsize=(10, 6))  
sns.lineplot(data=f, x='Year', y='Likes', marker='o')  
for index, value **in** f.iterrows():  
 plt.text(value['Year'], value['Likes'], str(value['Likes']), ha='left', va='bottom')  
  
plt.title('Cumulative Likes Over Years on Facebook')  
plt.xlabel('Year')  
plt.ylabel('Cumulative Likes')  
  
plt.show()

U,{adb7a784-d279-4b11-bf2f-5269ee84b2b9}{93},12.270833333333334,7.6875

f = Facebook.groupby('Year')['Retweets'].sum().reset\_index()  
plt.figure(figsize=(10, 6))  
sns.lineplot(data=f, x='Year', y='Retweets', marker='o')  
for index, value **in** f.iterrows():  
 plt.text(value['Year'], value['Retweets'], str(value['Retweets']), ha='left', va='bottom')  
  
plt.title('Cumulative Retweets Over Years on Facebook')  
plt.xlabel('Year')  
plt.ylabel('Cumulative Retweets')  
  
plt.show()

U,{adb7a784-d279-4b11-bf2f-5269ee84b2b9}{124},12.270833333333334,7.6875

# **Twitter**

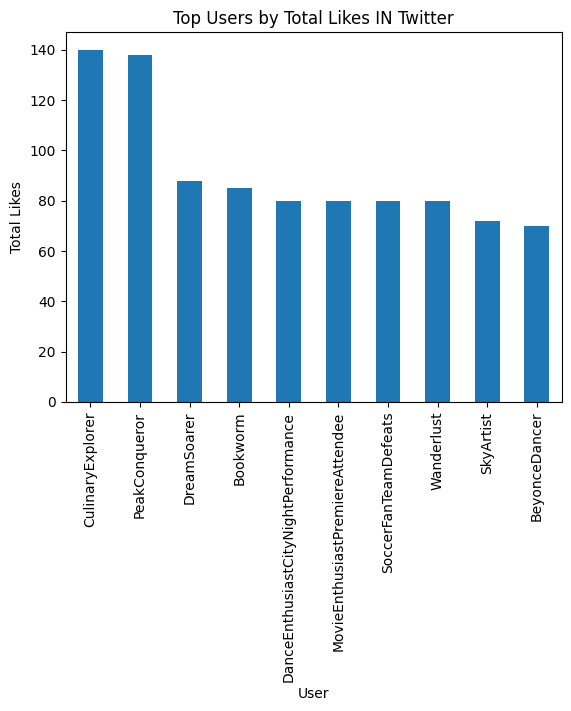
### Top 10 hashtags retweeted

H\_R\_t=Twitter.groupby('Hashtags')['Retweets'].max().nlargest(10).sort\_values(ascending=False)  
H\_R\_t.plot(kind='bar')  
plt.title('Top 10 hashtags retweeted in $/ Twitter $/')  
plt.xlabel('Hashtags')  
plt.ylabel('count')  
plt.show()

U,{adb7a784-d279-4b11-bf2f-5269ee84b2b9}{189},11.708333333333334,15.6875

### Q2: Who User liked mostly ?

top\_likes\_platform\_t = Twitter.groupby('User')['Likes'].sum().nlargest(10)  
top\_likes\_platform\_t.plot(kind='bar')  
plt.title('Top Users by Total Likes IN Twitter')  
plt.xlabel('User')  
plt.ylabel('Total Likes')  
plt.show()



f = Twitter.groupby('Year')['Likes'].sum().reset\_index()  
plt.figure(figsize=(10, 6))  
sns.lineplot(data=f, x='Year', y='Likes', marker='o')  
for index, value **in** f.iterrows():  
 plt.text(value['Year'], value['Likes'], str(value['Likes']), ha='left', va='bottom')  
  
plt.title('Cumulative Likes Over Years on Twitter')  
plt.xlabel('Year')  
plt.ylabel('Cumulative Likes')  
  
plt.show()

U,{926f7d88-b06c-494b-a13e-76f8dacaabe8}{13},12.270833333333334,7.6875

f = Twitter.groupby('Year')['Retweets'].sum().reset\_index()  
plt.figure(figsize=(10, 6))  
sns.lineplot(data=f, x='Year', y='Retweets', marker='o')  
for index, value **in** f.iterrows():  
 plt.text(value['Year'], value['Retweets'], str(value['Retweets']), ha='left', va='bottom')  
  
plt.title('Cumulative Retweets Over Years on Twitter')  
plt.xlabel('Year')  
plt.ylabel('Cumulative Retweets')  
  
plt.show()

U,{926f7d88-b06c-494b-a13e-76f8dacaabe8}{44},12.270833333333334,7.6875

# **Instagram**

### Top 10 hashtags retweeted

H\_R\_i=Instagram.groupby('Hashtags')['Retweets'].max().nlargest(15).sort\_values(ascending=False)  
H\_R\_i.plot(kind='bar')  
plt.title('Top 15 hashtags retweeted in $/ Instagram $/')  
plt.xlabel('Hashtags')  
plt.ylabel('count')  
plt.show()

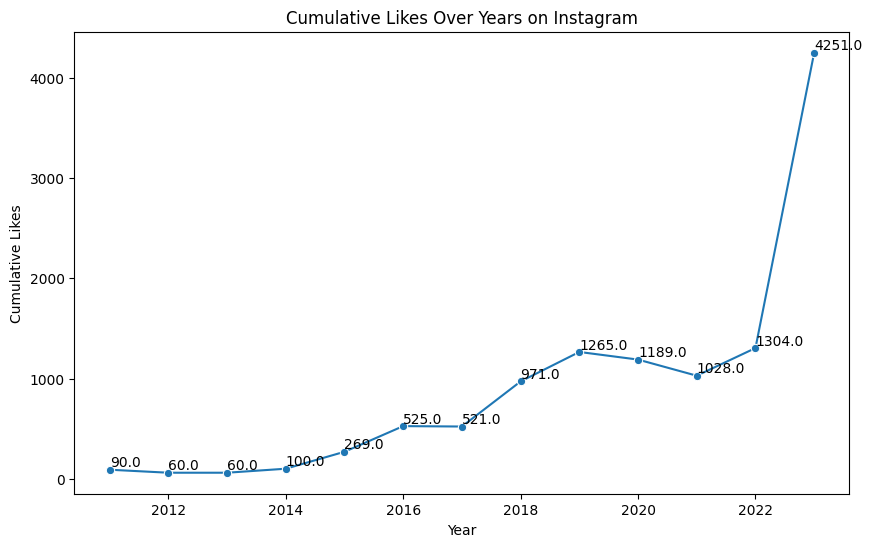
U,{926f7d88-b06c-494b-a13e-76f8dacaabe8}{109},11.708333333333334,15.270833333333334

### Q2: Who User liked mostly ?

top\_likes\_platform\_i = Instagram.groupby('User')['Likes'].sum().nlargest(10)  
top\_likes\_platform\_i.plot(kind='bar')  
plt.title('Top Users by Total Likes IN Instagram')  
plt.xlabel('User')  
plt.ylabel('Total Likes')  
plt.show()

U,{926f7d88-b06c-494b-a13e-76f8dacaabe8}{157},11.895833333333334,11.833333333333334

f = Instagram.groupby('Year')['Likes'].sum().reset\_index()  
plt.figure(figsize=(10, 6))  
sns.lineplot(data=f, x='Year', y='Likes', marker='o')  
for index, value **in** f.iterrows():  
 plt.text(value['Year'], value['Likes'], str(value['Likes']), ha='left', va='bottom')  
  
plt.title('Cumulative Likes Over Years on Instagram')  
plt.xlabel('Year')  
plt.ylabel('Cumulative Likes')  
  
plt.show()



f = Instagram.groupby('Year')['Retweets'].sum().reset\_index()  
plt.figure(figsize=(10, 6))  
sns.lineplot(data=f, x='Year', y='Retweets', marker='o')  
for index, value **in** f.iterrows():  
 plt.text(value['Year'], value['Retweets'], str(value['Retweets']), ha='left', va='bottom')  
  
plt.title('Cumulative Retweets Over Years on Instagram')  
plt.xlabel('Year')  
plt.ylabel('Cumulative Retweets')  
  
plt.show()

