**SENTIMENT ANALYSIS REPORT**

**OBJECTIVES:** This project aims to explore sentiment across various platform and countries . The main objectives are to analyze:

* Sentiments are classified in three ways i.e Positive,Neutral,Negative.
* Sentiment over time . How it sentiment varies across different Time period.
* To analyze how people use different social media Platforms(i.e Instagram,Twitter,Facebook) to share their emotions .
* Top social media users.
* To analyze sentiment over time in different country.
* Trending hashtags.
* Ranking Platform by average Likes and Retweets.

**DATASET OVERVIEW:** The dataset includes various features associated with social media posts. These are:

* Text: The main content of the post.
* Timestamp: The time of post creation
* User: The username or user id.
* Platform: The Platform where the post generated (ie Instagram).
* Hashtags: Hashtags used in each post.
* Retweets: Posts that are shared or Retweets.
* Likes: Total likes on each post.
* Country: Geographical location of the post.

**Libraries Used:**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from wordcloud import WordCloud

import nltk

from textblob import TextBlob

import plotly.express as px

from collection import Counter

**Data Cleaning:**

* Two unnecessary columns dropped from the dataset using python.
* Timestamp column is converted into datetime format.
* Checked the missing values.
* Inconsistent valueswere cleaned using str.strip()in the country and platform columns

**Univariate Distribution:**

It visualize the distribution of sentiment,country and platform using matplotlib and seaborn.

**Sentiment analysis(VADER):**

* We used VADER(Valence Aware dictionary & Sentiment Reasoner) from nltk for sentiment analysis.

from nltk.sentiment.vader import SentimentIntensityAnalyzer

nltk.download('vader\_lexicon')

data['VADER\_COMPOUND']=data['Text'].astype(str).apply(lambda x:sid.polarity\_scores(x)['compound'])

* Sentiment Classification:

def sentiment\_score(score):

if score<=0.05:

return 'Negative'

if score>=0.05:

return 'Positive'

else:

return 'Neutral'

data['VADER\_SENTIMENT']=data['VADER\_COMPOUND'].apply(sentiment\_score)

**Sentiment Distribution over Platform:**

* Positive and Negative sentiment are distributed over social media platform like Instagram,Twitter and Facebook.

plt.figure(figsize=(10,5))

sns.countplot(data=data,x='Platform',hue='VADER\_SENTIMENT')

plt.title('Overall sentiment over platform')

plt.show()

**Sentiment Distribution over Time:**

* It shows how semtiment is distributed over time , here Months are extracted from the Timestamp column to represent Time.

data["Month"]=data['Timestamp'].dt.strftime('%b')

Time=data.groupby(['Month','VADER\_SENTIMENT']).size().reset\_index(name='count')

plt.figure(figsize=(8,6))

sns.lineplot(data=Time,x="Month",hue="VADER\_SENTIMENT",y="count")

plt.title("Sentiment Over Time")

plt.show()

**Engagement Analysis:**

* To understand platform specific behavior, the average of likes and retweets have been analyzed.

engagement=data.groupby('Platform')[['Likes','Retweets']].mean().sort\_values(by='Likes',ascending=False)

print(engagement)

**Sentiment Distribution overCountry:**

* It visualize how sentiment is distributed across countries.

c1=data['Country'].value\_counts().index[:10] #Top 10 frequent Countries

new\_data=data[data['Country'].isin(c1) ]

#Plotting Heatmap

heatmap\_data=new\_data.pivot\_table(index='Country',columns='VADER\_SENTIMENT',aggfunc='size',fill\_value=0)

plt.figure(figsize=(10,5))

sns.heatmap(heatmap\_data,annot=True,fmt='d',cmap='YlGnBu')

plt.title('Distribution of Sentiment over Country')

plt.tight\_layout()

plt.show()

**Geographical Analysis:**

* The Top 3 countries were analyzed based on the number of likes and retweets on posts.

top\_Country=data['Country'].value\_counts().index[:3]

filtered=data[data['Country'].isin(top\_Country)]

filtered['Month']=filtered['Timestamp'].dt.strftime('%b')

trend=filtered.groupby(['Country','Month','VADER\_SENTIMENT']).size().reset\_index(name='count')

#Plotting

plt.figure(figsize=(10,5))

sns.lineplot(data=trend,x='Month',y='count',hue='VADER\_SENTIMENT',style='Country')

plt.title('Sentiment trends over time by Country')

plt.show()

**Top 10 User:**

* The user Ids of top 10 user who posted the most.

plt.figure(figsize=(8,6))

sns.barplot(y=Top\_User.values,x=Top\_User.index)

plt.title("Top Users")

plt.tight\_layout()

plt.xticks(rotation=45)

plt.show()

**Hashtag Trends:**

* Extracted top 20 used hashtags and visualize it.

from collections import Counter

print(data['Hashtags'].dropna().head(10))

hashtags=data['Hashtags'].dropna().str.lower().str.split()

flat\_hashtags=[]

for sublist in hashtags:

for tag in sublist:

flat\_hashtags.append(tag)

top\_hashtags=Counter(flat\_hashtags).most\_common(20)

tag,counts=zip(\*top\_hashtags)

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

sns.barplot(x=list(tag),y=list(counts))

plt.title("Top 20 Hashtags")

plt.xticks(rotation=45)

plt.show()

**Conclusion:**

In conclusion, this project successfully explored and analyzed user-generated sentiments across multiple social media platforms. Through a combination of NLP techniques and the VADER sentiment analyzer, we identified key emotional expressions such as admiration, excitement, and thrill. Visualizations provided clear insights into sentiment distribution across countries and platforms, revealing both regional and platform-specific trends. While the current model offered a strong baseline, future enhancements could include the use of advanced transformer-based models like RoBERTa for more accurate and context-aware sentiment classification. This analysis lays the groundwork for more refined social media monitoring tools and data-driven decision-making in digital marketing.