Om Yadav
BI-V8

CODE :-

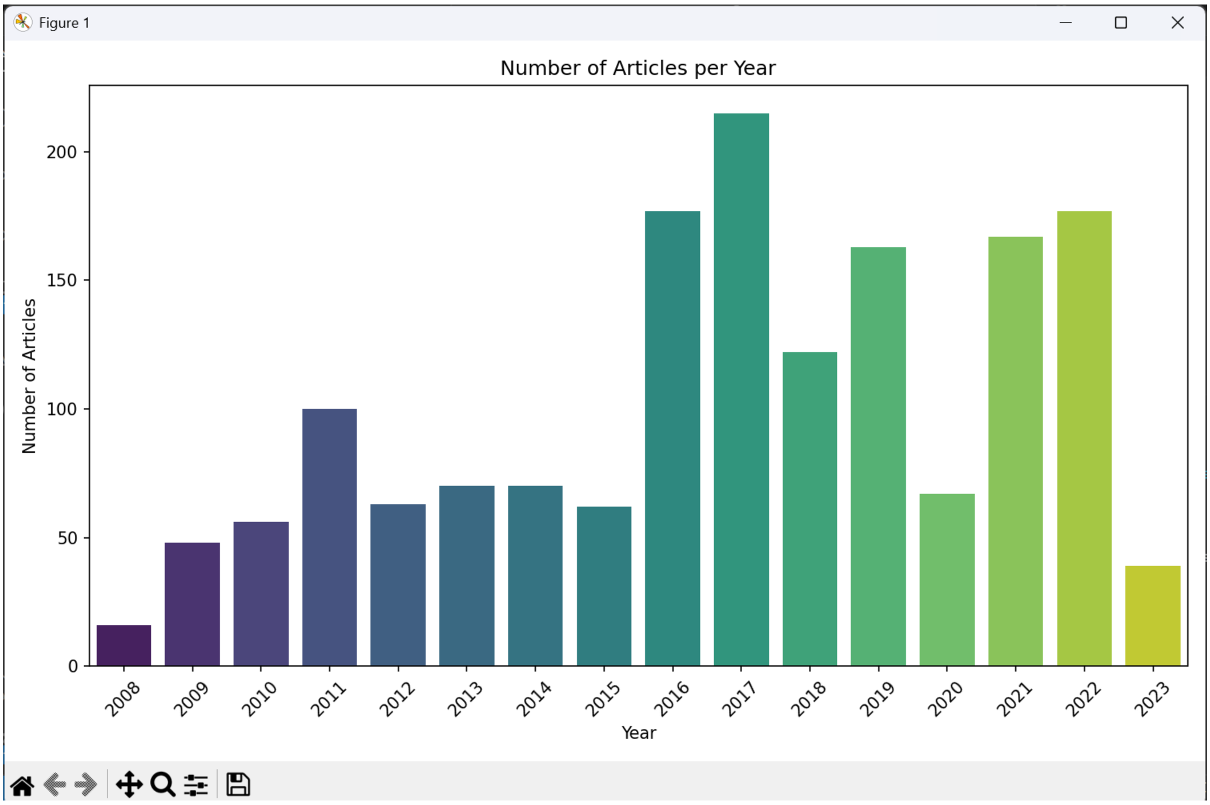
```
import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from wordcloud import WordCloud

from textblob import TextBlob


# Load the dataset

df = pd.read_csv("db/income_tax.csv")


# Clean column names

df.columns = df.columns.str.strip()


# Convert 'date' column to datetime, coerce errors to NaT

df['date'] = pd.to_datetime(df['date'], errors='coerce')


# Drop rows with invalid or missing dates

df_clean = df.dropna(subset=['date'])


# Extract year from date

df_clean['year'] = df_clean['date'].dt.year

# Insight 1: Number of articles per year

articles_per_year = df_clean['year'].value_counts().sort_index()


# Plot articles per year
```

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

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

plt.title("Number of Articles per Year")

plt.xlabel("Year")

plt.ylabel("Number of Articles")

plt.xticks(rotation=45)

plt.tight_layout()

plt.show()


# Insight 2: Most frequent keywords in titles using WordCloud

titles = ' '.join(df_clean['title'].dropna().astype(str))

wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(titles)


# Plot WordCloud

plt.figure(figsize=(12, 6))

plt.imshow(wordcloud, interpolation='bilinear')

plt.axis('off')

plt.title("Most Frequent Keywords in Titles")

plt.show()


# Insight 3: Sentiment analysis of article content

def get_sentiment(text):

    return TextBlob(str(text)).sentiment.polarity


df_clean['sentiment'] = df_clean['content'].apply(get_sentiment)


# Plot sentiment distribution
```

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

sns.histplot(df_clean['sentiment'], bins=30, kde=True, color='skyblue')

plt.title("Sentiment Distribution of Article Content")

plt.xlabel("Sentiment Polarity")

plt.ylabel("Frequency")

plt.tight_layout()

plt.show()
```

OUTPUT:-

1.



2.

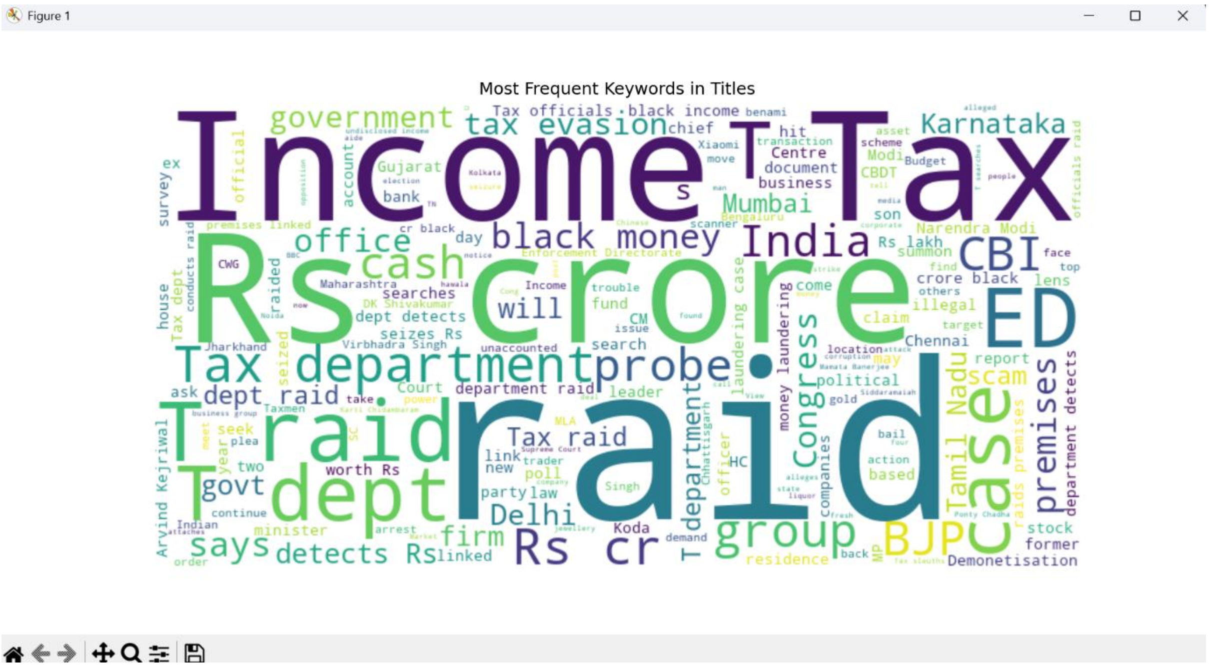


Figure 1
Most Frequent Keywords in Titles
Income Tax
Rs crore
raid
Tax department
probe
dept
T raid
case
ED
CBI
India
black money
cash
office
government
tax evasion
Karnataka
Mumbai
Congress
group
BJP
premises
Tamil Nadu
scam
Delhi
Rs cr
says
detects Rs
firm
linked
worth Rs
govt
dept raid
will
fund
money laundering case
Chennai
political
Demonetisation
former
stock
Arvind Kejriwal
Jharkhand
Maharashtra
Gujarat
bank
survey
house
official
account
Narendra Modi
Centre
document
business
CBDT
Modi
Budget
Xiaomi
chief
hit
transaction
scheme
asset
crore black
illegal
claim
report
search
seizes Rs
Virbhadra Singh
DK Shivakumar
Enforcement Directorate
searches
unaccounted
leader
Court
department raid
poll
party
law
link
new
trader
Koda
residence
officer
companies
HC
bail
action
based
Singh
minister
Indian
continue
order
arrest
two
year
seek
plea
Taxmen
ask
seized
raided
CWG
lakh
summon
son
lens
target
others
find
face
top
location
gold
may
come
Chhattisgarh
T department

3.

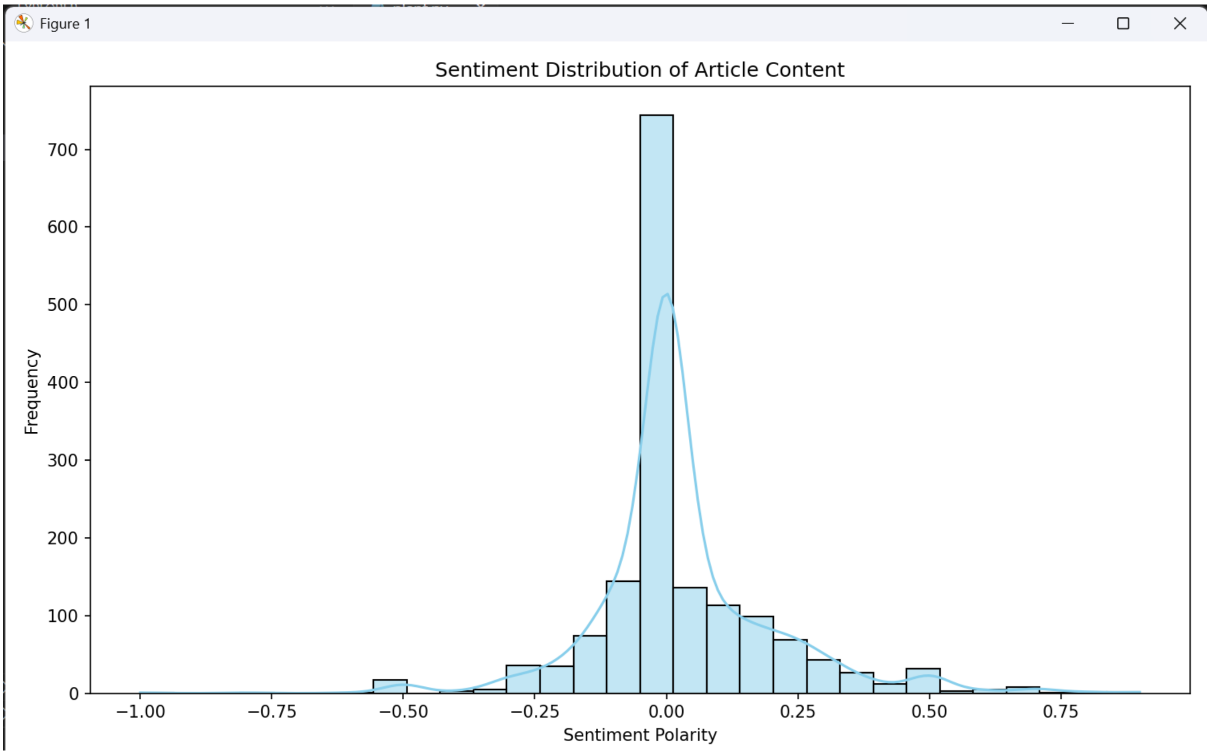


Figure 1
Sentiment Distribution of Article Content
Frequency
Sentiment Polarity
0
100
200
300
400
500
600
700
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75