

# Python Project



**FOR DATA ANALYST**

# Sentiment Analysis

[linkedin.com/in/ileonjose](https://www.linkedin.com/in/ileonjose)



**Overview:** Analyze Twitter sentiment regarding climate change to gauge public perception. This project combines data collection, text preprocessing, advanced sentiment analysis, and data visualization.

Dataset:

[kaggle.com/datasets/kazanova/sentiment140](https://kaggle.com/datasets/kazanova/sentiment140)

[linkedin.com/in/ileonjose](https://linkedin.com/in/ileonjose)



# Step 1: Gather Data

Utilize the Tweepy library to collect tweets related to climate change over a specific period.

## 1. Set Up Twitter API Access:

- Create a Twitter Developer account and obtain API keys.

## 2. Fetch Tweets:

- Collect tweets using specific keywords or hashtags like #ClimateChange or #GlobalWarming.



```
import tweepy

# Authenticate to Twitter
auth = tweepy.OAuth1UserHandler('API_KEY', 'API_SECRET_KEY', 'ACCESS_TOKEN',
                                'ACCESS_TOKEN_SECRET')
api = tweepy.API(auth)

# Fetch tweets
tweets = api.search_tweets(q="#ClimateChange OR #GlobalWarming", count=500,
                            lang="en", tweet_mode='extended')
tweet_data = [tweet.full_text for tweet in tweets]
```

[linkedin.com/in/ileonjose](https://www.linkedin.com/in/ileonjose)



# Step 2: Preprocess Text Data

Clean and prepare the text for analysis, focusing on meaningful content.

## 1. Text Cleaning Functions:

- Remove URLs, special characters, and stopwords.

## 2. Tokenization:

- Break the text into words for further analysis.



Python

```
import re
import nltk
from nltk.corpus import stopwords

nltk.download('stopwords')
stop_words = set(stopwords.words('english'))

def clean_text(text):
    text = re.sub(r'http\S+|www\S+|https\S+', '', text) # Remove URLs
    text = re.sub(r'\@w+|\#','', text) # Remove mentions and hashtags
    text = text.lower() # Convert to lowercase
    return ' '.join([word for word in text.split() if word not in
stop_words]) # Remove stopwords

cleaned_tweets = [clean_text(tweet) for tweet in tweet_data]
```

[linkedin.com/in/ileonjose](https://www.linkedin.com/in/ileonjose)



# Step 3: Advanced Sentiment Analysis

Use VADER (Valence Aware Dictionary and sentiment Reasoner) for nuanced sentiment analysis, particularly suitable for social media text.

## 1. Apply VADER:

- Get sentiment scores, which include positive, negative, and neutral values.



Python

```
from nltk.sentiment import SentimentIntensityAnalyzer

nltk.download('vader_lexicon')
sia = SentimentIntensityAnalyzer()

def get_sentiment_scores(text):
    return sia.polarity_scores(text)

sentiment_scores = [get_sentiment_scores(tweet) for tweet in
cleaned_tweets]
```

[linkedin.com/in/ileonjose](https://www.linkedin.com/in/ileonjose)





# Step 4: Aggregate and Visualize Results

Create visualizations to present sentiment distributions and trends over time.

## 1. Aggregate Scores:

- Classify tweets based on their compound scores into positive, negative, and neutral categories.

## 2. Bar and Line Graphs:

- Use Matplotlib to visualize the overall sentiment and temporal trends.



Python

```
import pandas as pd
import matplotlib.pyplot as plt

# Create DataFrame
df = pd.DataFrame(sentiment_scores)

# Classify sentiments
df['sentiment'] = df['compound'].apply(lambda x: 'positive'
if x > 0.05 else ('negative' if x < -0.05 else 'neutral'))

# Plotting
plt.figure(figsize=(10, 5))
df['sentiment'].value_counts().plot(kind='bar', color=
['green', 'red', 'gray'])
plt.title('Sentiment Distribution of Climate Change Tweets')
plt.xlabel('Sentiment')
plt.ylabel('Number of Tweets')
plt.show()
```

[linkedin.com/in/ileonjose](https://www.linkedin.com/in/ileonjose)



# Step 5: Provide Insights

Write a comprehensive report detailing your findings:

- **Overall Sentiment:** Discuss the balance of positive, negative, and neutral tweets.
- **Trends Over Time:** If data spans several days, analyze how sentiment fluctuated, especially during significant climate events.
- **Public Opinion:** Reflect on what the sentiment suggests about public awareness and concern regarding climate change.



# Recruiters POV:

- This project highlights your ability to collect and analyze social media data, apply NLP techniques, and present your findings visually.
- It demonstrates critical thinking and data storytelling, valuable skills for any data analyst role.

