```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings


train = pd.read_csv('/content/train_tweet.csv')
test = pd.read_csv('/content/train_tweet.csv')
print(train.shape)
print(test.shape)


train.head()


test.head()


train[train['label'] == 0].head(10)
train[train['label'] == 1].head(10)
train['label'].value_counts().plot.bar(color = 'pink', figsize = (6, 4))


length_train = train['tweet'].str.len().plot.hist(color = 'pink', figsize = (6, 4))


length_test = test['tweet'].str.len().plot.hist(color = 'orange', figsize = (6, 4))


train['len'] = train['tweet'].str.len()
test['len'] = test['tweet'].str.len()
train.head(10)


train['len'] = train['tweet'].str.len()
test['len'] = test['tweet'].str.len()
train.head(10)


from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(stop_words = 'english')
words = cv.fit_transform(train.tweet)


sum_words = words.sum(axis=0)
words_freq = [(word, sum_words[0, i]) for word, i in cv.vocabulary_.items()]
words_freq = sorted(words_freq, key = lambda x: x[1], reverse = True)
frequency = pd.DataFrame(words_freq, columns=['word', 'freq'])
frequency.head(30).plot(x='word', y='freq', kind='bar', figsize=(15, 7), color = 'blue')
plt.title("Most Frequently Occuring Words - Top 30")


Start coding or generate with AI.
```