EXPERIMENT NO.5

AIM: To Develop Content based social media analytics model for business.

RESOURCES REQUIRED: Windows/MAC/Linux O.S, Compatible version of Python.

THEORY:

What Is Topic Analysis?

Topic analysis (also called topic detection, topic modeling, or topic extraction) is a machine learning technique that organizes and understands large collections of text data, by assigning "tags" or categories according to each individual text's topic or theme.Topic analysis uses natural language processing (NLP) to break down human language so that you can find patterns and unlock semantic structures within texts to extract insights and help make data-driven decisions. The two most common approaches for topic analysis with machine learning are NLP topic modeling and NLP topic classification.

What Is Trend Analysis?

Trend analysis is a technique used in technical analysis that attempts to predict future stock price movements based on recently observed trend data. Trend analysis uses historical data, such as price movements and trade volume, to forecast the long-term direction of market sentiment.

What is sentiment analysis (opinion mining)?

Sentiment analysis, also referred to as opinion mining, is an approach to natural language processing (NLP) that identifies the emotional tone behind a body of text. This is a popular way for organizations to determine and categorize opinions about a product, service, or idea. It involves the use of data mining, machine learning (ML) and artificial intelligence (AI) to mine text for sentiment and subjective information.

Audio, video, image analytics:

Image analysis or imagery analysis is the extraction of meaningful information from images; mainly from digital images by means of digital image processing techniques. Image analysis tasks can be as simple as reading bar coded tags or as sophisticated as identifying a person from their face.Video content analysis or video content analytics (VCA), also known as video analysis or video analytics (VA), is the capability of automatically analyzing video to detect and determine temporal and spatial events. This technical capability is used in a wide range of domains including entertainment, video retrieval and video browsing, health-care, retail, automotive, transport, home automation, flame and smoke detection, safety, and security. The algorithms can be implemented as software on general-purpose machines, or as hardware in specialized video processing units.

CONCLUSION: Hence, we have successfully studied to Develop Content based social media analytics model for business

In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
post_df = pd.read_csv('post.csv',index_col=['id'])
comment_df = pd.read_csv('comment.csv',index_col=['comment_id'])
```

In [3]:
```
post_df.head()
```

Out[3]:

id	author	author_id	total_comments	upvote	post_type	ti
12glkw4	TheBodyPolitic1	von3w6y2	319	591	top week	Why did Python becor popular ur long a
12hj9oc	MetonymyQT	shnqm	80	527	top week	Free cours Build a mode API w FastAPI a
12egsoz	2broke2code	rs4dqilj	105	467	top week	I trainec RoastBot >120,000 fac and >0
12ffsif	midnitte	3gad9	64	378	top week	EP 684: A P Interpreter G Accept
12fzdu2	aeluro1	88efmodhn	12	370	top week	Comprehensi Reddit Sav Pos Downloade

In [4]:
```
comment_df.head()
```

Out[4]:

comment_id	post_id	parent_id	body	created_on	upvotes	author_id	author_
jfkvwx0	12glkw4	t3_12glkw4	Hardware wasn't ready for Python in that time	2023-04-09 15:27:18	5	22vat21u	Du
jfnha98	12glkw4	t3_12glkw4	Because Python was developed with the conceit ...	2023-04-10 02:55:16	2	4wtjvsh6	FredVIII
jflnf2e	12glkw4	t3_12glkw4	Perl was *the* scripting language in the early...	2023-04-09 18:37:04	2	4i9hp	tom
jflbch7	12glkw4	t3_12glkw4	I was a web developer between 2000 and 2010, a...	2023-04-09 17:13:45	2	8hi6986p	As Dress
jfmk972	12glkw4	t1_jflbch7	Adding on to this from my POV, early web in th...	2023-04-09 22:33:38	3	380he	snap

Analysing 1st post comments

In [5]:
```
fp_comments = comment_df.loc[comment_df['post_id']=='12glkw4']
fp_comments.head()
```

Out[5]:

comment_id	post_id	parent_id	body	created_on	upvotes	author_id	author_
jfkvwx0	12glkw4	t3_12glkw4	Hardware wasn't ready for Python in that time	2023-04-09 15:27:18	5	22vat21u	Du
jfnha98	12glkw4	t3_12glkw4	Because Python was developed with the conceit ...	2023-04-10 02:55:16	2	4wtjvsh6	FredVIII
jflnf2e	12glkw4	t3_12glkw4	Perl was *the* scripting language in the early...	2023-04-09 18:37:04	2	4i9hp	tom
jflbch7	12glkw4	t3_12glkw4	I was a web developer between 2000 and 2010, a...	2023-04-09 17:13:45	2	8hi6986p	As Dress
jfmk972	12glkw4	t1_jflbch7	Adding on to this from my POV, early web in th...	2023-04-09 22:33:38	3	380he	snap

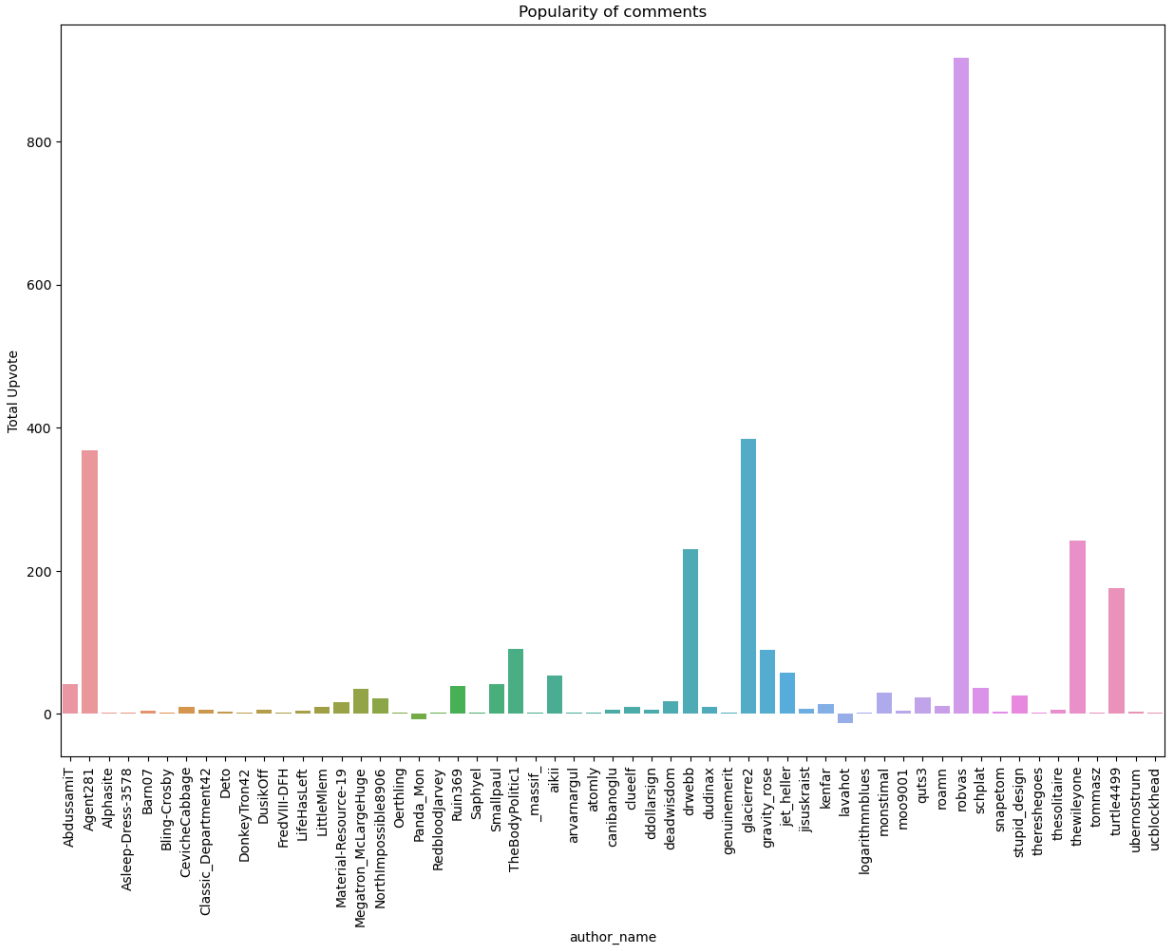
In [6]:
```
import matplotlib.pyplot as plt
import seaborn as sns
```

In [7]:
```
upvote_rec = fp_comments.groupby(fp_comments['author_name']).apply(lambda
    "total_upvote":x['upvotes'].sum(),
    'total_comments':len(x)
})
upvote_rec
```

Out[7]:
```
author_name
AbdussamiT                {'total_upvote': 42, 'total_comments': 2}
Agent281                 {'total_upvote': 368, 'total_comments': 1}
Alphasite                  {'total_upvote': 2, 'total_comments': 1}
Asleep-Dress-3578          {'total_upvote': 2, 'total_comments': 1}
Barn07                     {'total_upvote': 4, 'total_comments': 1}
Bling-Crosby               {'total_upvote': 2, 'total_comments': 1}
CevicheCabbage             {'total_upvote': 9, 'total_comments': 1}
Classic_Department42       {'total_upvote': 6, 'total_comments': 1}
Deto                       {'total_upvote': 3, 'total_comments': 1}
DonkeyTron42               {'total_upvote': 2, 'total_comments': 1}
DusikOff                   {'total_upvote': 5, 'total_comments': 1}
FredVIII-DFH               {'total_upvote': 2, 'total_comments': 1}
LifeHasLeft                {'total_upvote': 4, 'total_comments': 2}
LittleMlem                 {'total_upvote': 9, 'total_comments': 1}
Material-Resource-19      {'total_upvote': 16, 'total_comments': 1}
Megatron_McLargeHuge      {'total_upvote': 35, 'total_comments': 1}
NorthImpossible8906       {'total_upvote': 21, 'total_comments': 1}
Oerthling                  {'total_upvote': 1, 'total_comments': 1}
Panda_Mon                 {'total_upvote': -8, 'total_comments': 1}
RedbloodJarvey             {'total_upvote': 1, 'total_comments': 1}
Ruin369                   {'total_upvote': 39, 'total_comments': 1}
Saphyel                    {'total_upvote': 2, 'total_comments': 1}
Smallpaul                 {'total_upvote': 41, 'total_comments': 1}
TheBodyPolitic1           {'total_upvote': 90, 'total_comments': 3}
_massif_                   {'total_upvote': 1, 'total_comments': 1}
aikii                     {'total_upvote': 53, 'total_comments': 1}
arvarnargul                {'total_upvote': 1, 'total_comments': 1}
atomly                     {'total_upvote': 1, 'total_comments': 1}
canibanoglu                {'total_upvote': 5, 'total_comments': 1}
clueelf                   {'total_upvote': 10, 'total_comments': 1}
ddollarsign                {'total_upvote': 5, 'total_comments': 1}
deadwisdom                {'total_upvote': 18, 'total_comments': 1}
drwebb                   {'total_upvote': 230, 'total_comments': 1}
dudinax                    {'total_upvote': 9, 'total_comments': 1}
genuinemerit               {'total_upvote': 2, 'total_comments': 1}
glacierre2               {'total_upvote': 384, 'total_comments': 1}
gravity_rose              {'total_upvote': 89, 'total_comments': 1}
jet_heller                {'total_upvote': 57, 'total_comments': 1}
jisuskraist                {'total_upvote': 7, 'total_comments': 1}
kenfar                    {'total_upvote': 13, 'total_comments': 1}
lavahot                  {'total_upvote': -13, 'total_comments': 1}
logarithmnblues            {'total_upvote': 1, 'total_comments': 1}
monstimal                 {'total_upvote': 29, 'total_comments': 1}
moo9001                    {'total_upvote': 4, 'total_comments': 1}
quts3                     {'total_upvote': 23, 'total_comments': 2}
roamn                     {'total_upvote': 11, 'total_comments': 1}
robvas                   {'total_upvote': 917, 'total_comments': 1}
schplat                   {'total_upvote': 36, 'total_comments': 1}
snapetom                   {'total_upvote': 3, 'total_comments': 1}
stupid_design             {'total_upvote': 26, 'total_comments': 1}
thereshegoes               {'total_upvote': 2, 'total_comments': 1}
thesolitaire               {'total_upvote': 6, 'total_comments': 1}
thewileyone              {'total_upvote': 242, 'total_comments': 1}
tommasz                    {'total_upvote': 2, 'total_comments': 1}
turtle4499               {'total_upvote': 175, 'total_comments': 1}
ubernostrum                {'total_upvote': 3, 'total_comments': 1}
ucblockhead                {'total_upvote': 1, 'total_comments': 1}
dtype: object
```

Most upvotes users in comments

In [8]:
```
fig,ax = plt.subplots(1,1,figsize=(15,10))
plt.xticks(rotation=90)
ax.set_xlabel('author_name')
ax.set_ylabel('Total Upvote')
ax.set_title('Popularity of comments')
sns.barplot(x=upvote_rec.index,y=[data['total_upvote'] for data in upvote
plt.show()
```



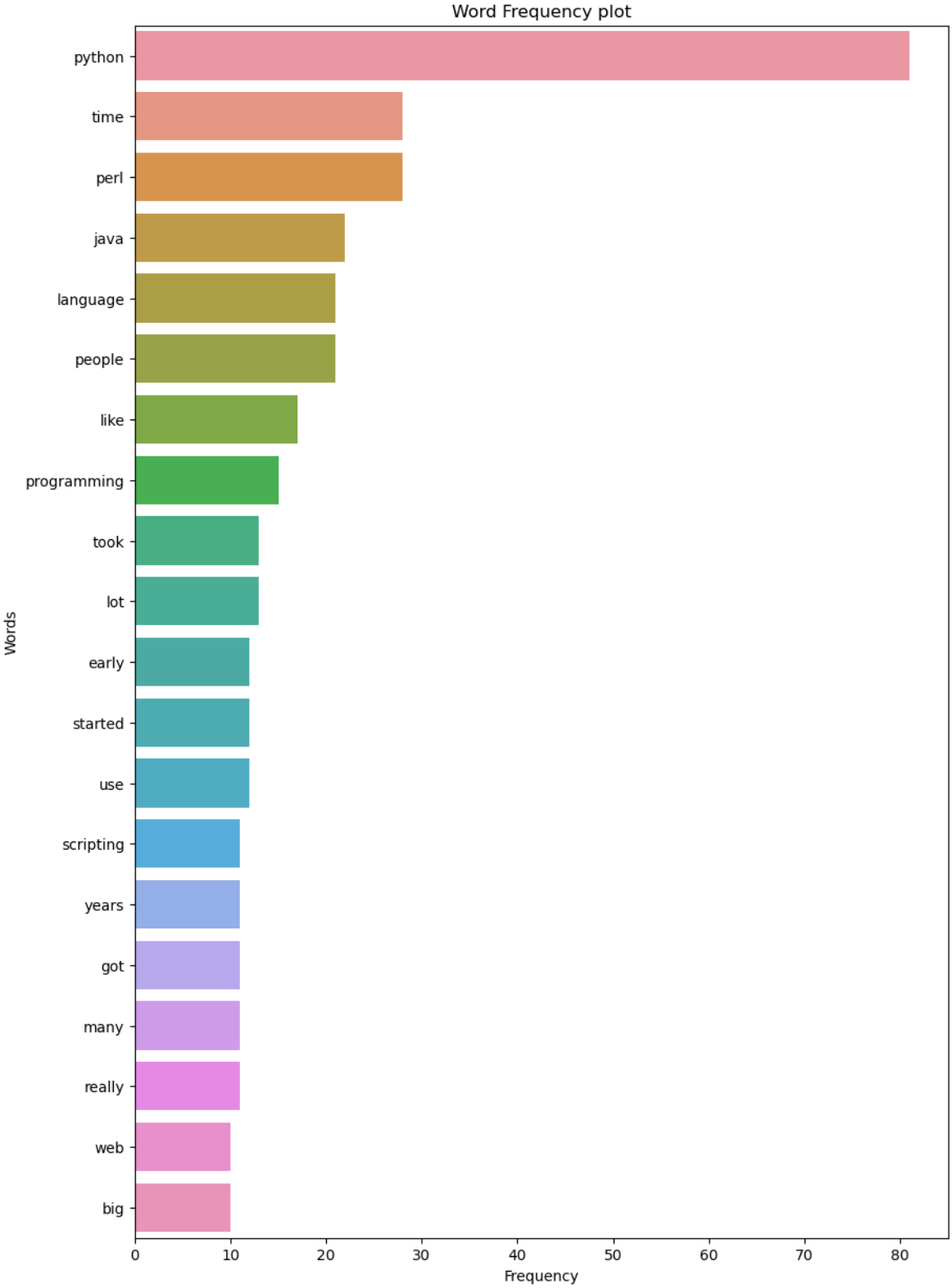
Most repeated words in comments

In [9]:
```
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize

sw = set(stopwords.words('english'))

comments = fp_comments['body'].values.tolist()

corpus = [word.lower() for comment in comments for word in word_tokenize(
corpus[:5]
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     /home/slowgamer/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Out[9]: ['hardware', 'ready', 'python', 'time', 'python']

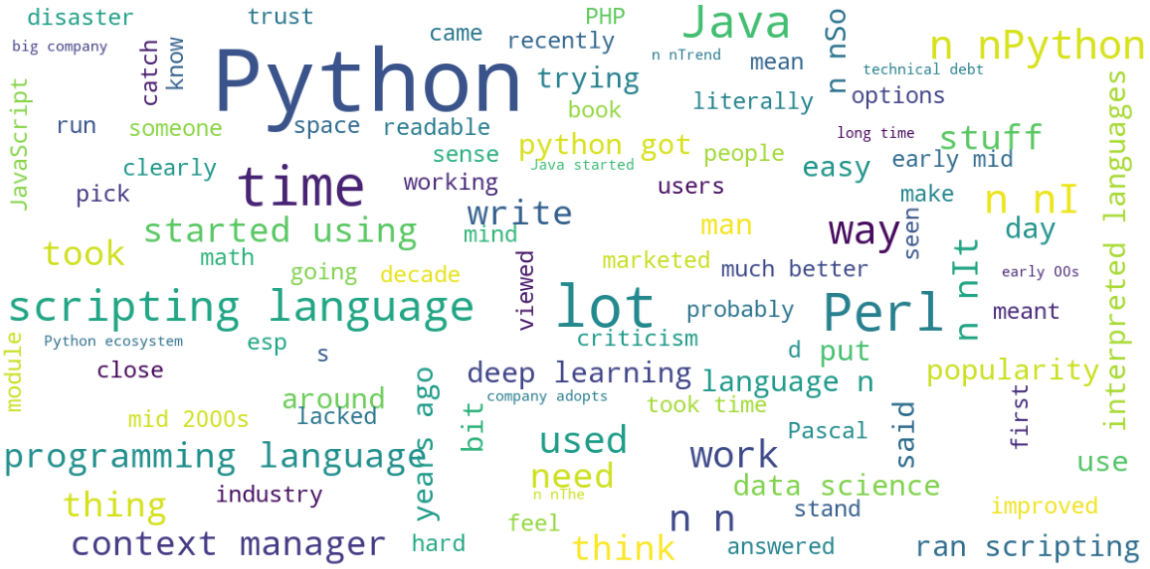
In [10]:
```
from collections import Counter

fig,ax = plt.subplots(1,1,figsize=(10,15))
word_counts = dict(Counter(corpus).most_common(20))
ax.set_xlabel('Frequency')
ax.set_ylabel('Words')
ax.set_title('Word Frequency plot')
sns.barplot(x=list(word_counts.values()),y=list(word_counts.keys()),ax=ax
plt.show()
```



In [52]:
```
from wordcloud import WordCloud, STOPWORDS

sw = set(STOPWORDS)

def generate_cloud(data,n_grams=2):
    fig, ax = plt.subplots(1,1,figsize=(15,15))
    wordcloud = WordCloud(
        background_color='white',
        stopwords=sw,
        max_words=100,
        max_font_size=30,
        scale=3,
        random_state=1,
        collocation_threshold=n_grams
    ).generate(data)

    ax.imshow(wordcloud)
    ax.axis('off')
    plt.show()
```

In [12]:
```
generate_cloud(str(fp_comments['body'].values).replace('\n',' '))
```

In [13]:
```
generate_cloud(str(fp_comments['body'].values).replace('\n',' '),n_grams=
```



Sentiment Analysis

Using Text Blob Library

In [16]:
```
from textblob import TextBlob

def polarity_score(text):
    return TextBlob(text).sentiment.polarity
```

In [30]:
```
fp_comments['comment_sent_score'] =  fp_comments['body'].apply(lambda tex
```

```
/tmp/ipykernel_30060/1826135865.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-doc
s/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  fp_comments['comment_sent_score'] =  fp_comments['body'].apply(lambda te
xt: polarity_score(text))
```

In [31]:
```
fp_comments[['body','comment_sent_score']]
```

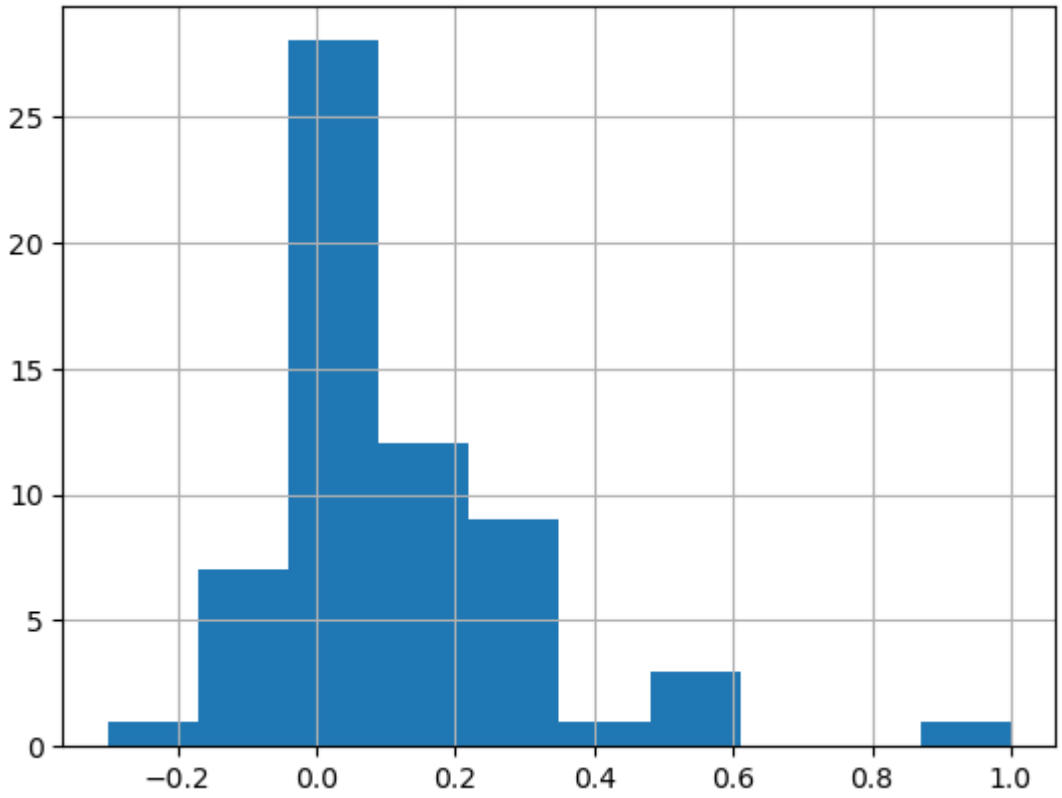
Out[31]:

comment_id	body	comment_sent_score
jfkvwx0	Hardware wasn't ready for Python in that time	0.200000
jfnha98	Because Python was developed with the conceit ...	0.212500
jflnf2e	Perl was *the* scripting language in the early...	-0.037500
jflbch7	I was a web developer between 2000 and 2010, a...	0.292308
jfmk972	Adding on to this from my POV, early web in th...	0.080903
...	...	...
jfnxolg	We have to start calling Python as JavaPython,...	0.000000
jfnvotv	Java was incredibly popular because it filled ...	0.304048
jfktbt2	Fair point, but Python became popular and I ca...	0.041667
jflq4yt	I'm wondering if Java started to loose popular...	-0.076923
jfl01if	To further this comment, Sun had a $500 millio...	0.000000

62 rows × 2 columns

In [32]:
```
fp_comments['comment_sent_score'].hist()
```

Out[32]: <AxesSubplot: >



In [29]:
```
def sentiment(score):
    return 'positive' if score>0 else 'negative' if score<0 else 'neutral
```

In [35]:
```
fp_comments['comment_sentiment'] = fp_comments['comment_sent_score'].appl
```

```
/tmp/ipykernel_30060/3742194281.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  fp_comments['comment_sentiment'] = fp_comments['comment_sent_score'].apply(lambda score:sentiment(score))
```

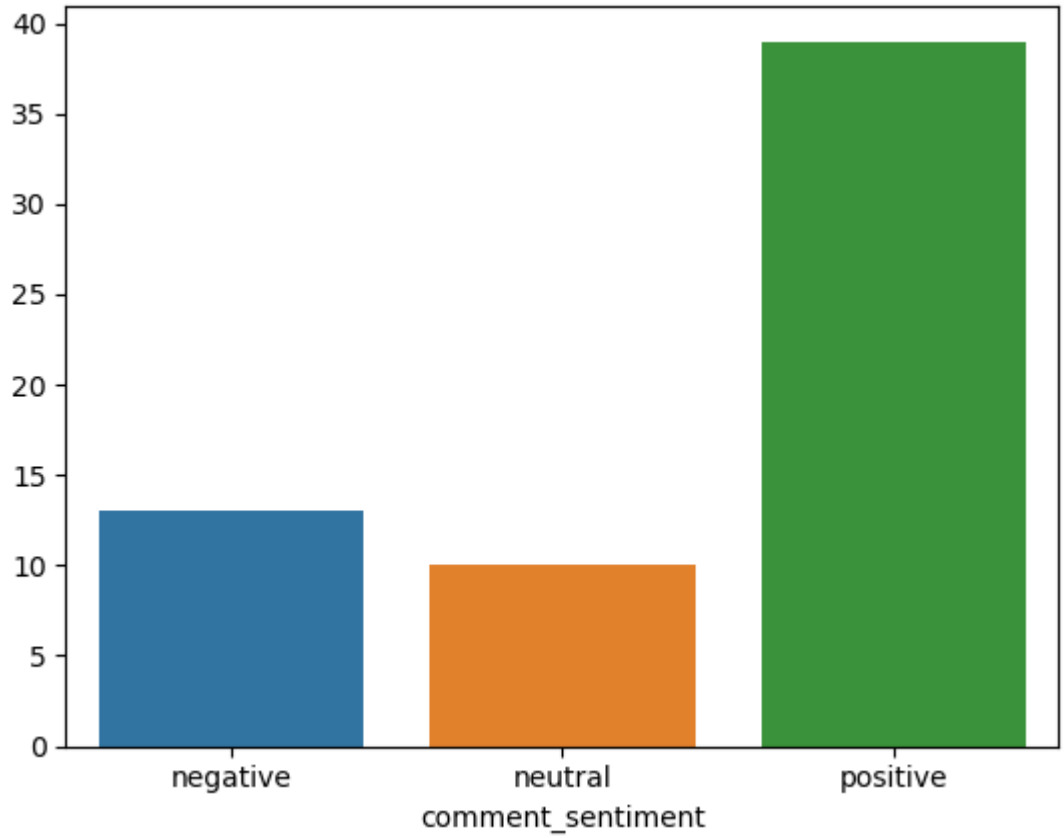
In [36]:
```
fp_comments[['body','comment_sentiment']]
```

Out[36]:

comment_id	body	comment_sentiment
jfkvwx0	Hardware wasn't ready for Python in that time	positive
jfnha98	Because Python was developed with the conceit ...	positive
jflnf2e	Perl was *the* scripting language in the early...	negative
jflbch7	I was a web developer between 2000 and 2010, a...	positive
jfmk972	Adding on to this from my POV, early web in th...	positive
...	...	...
jfnxolg	We have to start calling Python as JavaPython,...	neutral
jfnvotv	Java was incredibly popular because it filled ...	positive
jfktbt2	Fair point, but Python became popular and I ca...	positive
jflq4yt	I'm wondering if Java started to loose popular...	negative
jfl01if	To further this comment, Sun had a $500 millio...	neutral

62 rows × 2 columns

In [47]:
```
cnt = fp_comments.groupby(fp_comments['comment_sentiment'])['body'].count

sns.barplot(x=cnt.index,y=cnt.values)
plt.show()
```



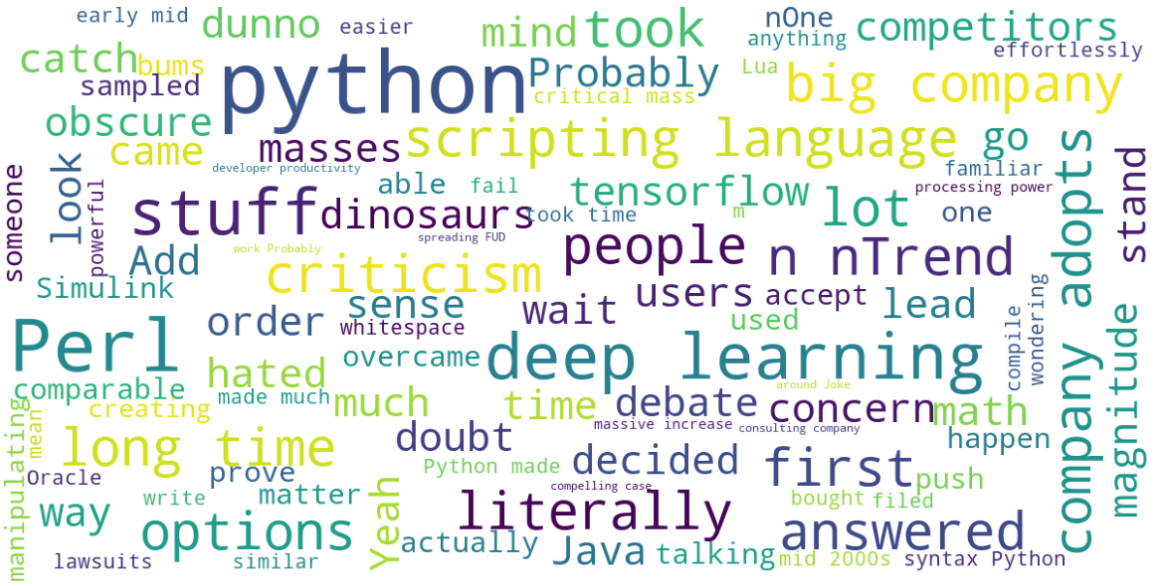
Segmenting comments by sentiment and plotting word cloud

Positive Sentiment

In [48]:
```
generate_cloud(str(fp_comments.loc[fp_comments['comment_sentiment']=='pos
```

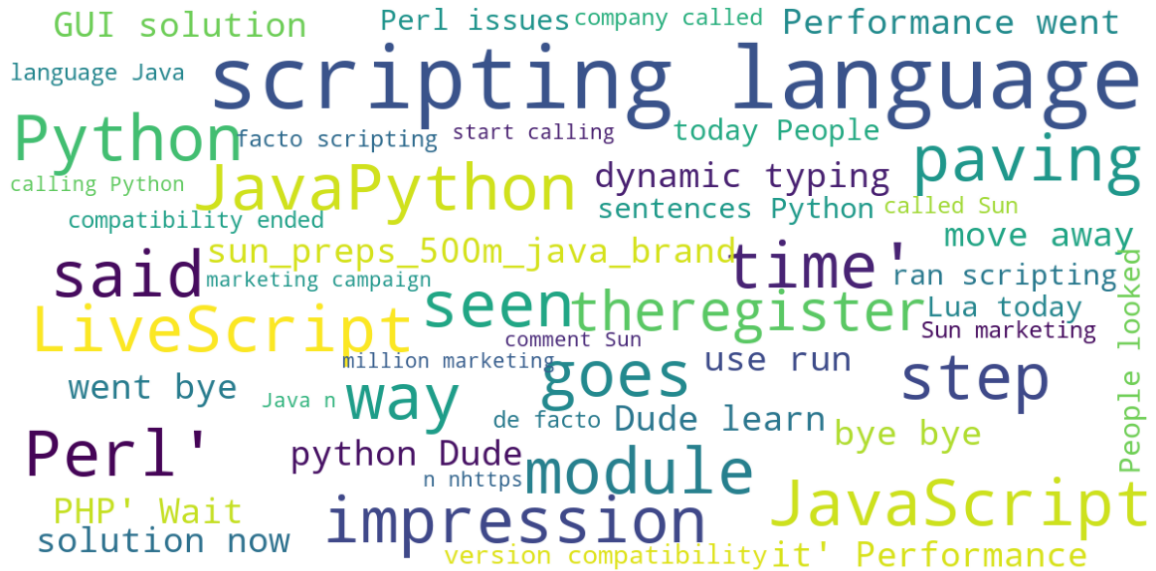
Negative Sentiment

In [54]:
```
generate_cloud(str(fp_comments.loc[fp_comments['comment_sentiment']=='neg
```



Neutral Sentiment

In [53]:
```
generate_cloud(str(fp_comments.loc[fp_comments['comment_sentiment']=='neu
```



In [ ]:
```
```