```
In [1]: import pandas as pd
         import numpy as np
         post df = pd.read csv('post.csv',index col=['id'])
In [2]:
         comment df = pd.read csv('comment.csv',index col=['comment id'])
In [3]:
         post_df.head()
Out[3]:
                           author
                                    author_id total_comments upvote post_type
                                                                                            ti
               id
                                                                                       Why did
                                                                                  Python becor
          12glkw4 TheBodyPolitic1
                                    von3w6y2
                                                          319
                                                                  591
                                                                        top week
                                                                                     popular ur
                                                                                        long a
                                                                                     Free cours
                                                                                   Build a mode
          12hj9oc
                     MetonymyQT
                                                           80
                                                                  527
                                       shnqm
                                                                        top week
                                                                                        API w
                                                                                      FastAPI &
                                                                                       I trained
                                                                                     RoastBot
          12egsoz
                      2broke2code
                                      rs4dqilj
                                                          105
                                                                  467
                                                                        top week
                                                                                  >120,000 fac
                                                                                       and >0
                                                                                   EP 684: A Po
            12ffsif
                          midnitte
                                                           64
                                                                  378
                                                                                   Interpreter G
                                       3gad9
                                                                        top week
                                                                                        Accept
                                                                                  Comprehensi
                                                                                    Reddit Sav
          12fzdu2
                          aeluro1 88efmodhn
                                                           12
                                                                  370
                                                                        top week
                                                                                           Pos
                                                                                    Downloade
In [4]:
         comment_df.head()
```

Out[4]:		post_id	parent_id	body	created_on	upvotes	author_id	author_I
	comment_id							
	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	snar
4								>

Analysing 1st post comments

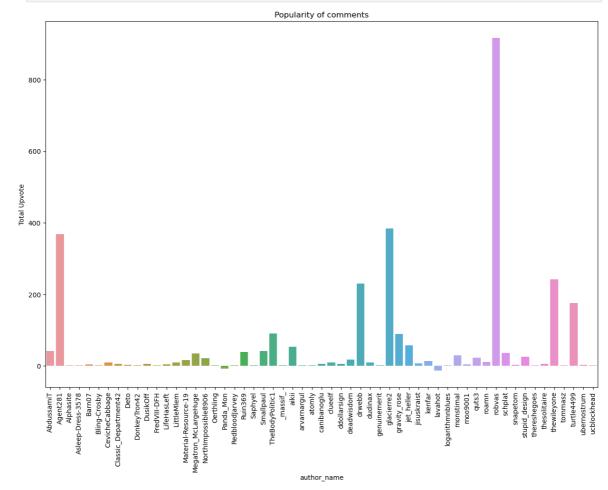
```
In [5]: fp_comments = comment_df.loc[comment_df['post_id']=='12glkw4']
fp_comments.head()
```

Out[5]: post_id parent_id body created_on upvotes author_id author_i comment_id Hardware wasn't 2023-04-09 jfkvwx0 12glkw4 t3_12glkw4 ready for 22vat21u Du 15:27:18 Python in that time Because Python was 2023-04-10 jfnha98 12glkw4 t3_12glkw4 2 4wtjvsh6 FredVIII developed 02:55:16 with the conceit ... Perl was *the* scripting 2023-04-09 2 jflnf2e 12glkw4 t3 12glkw4 4i9hp tom language 18:37:04 in the early... I was a web 2023-04-09 developer A٤ jflbch7 12glkw4 t3 12glkw4 8hi6986p between 17:13:45 Dress-2000 and 2010, a... Adding on to this from my 2023-04-09 3 jfmk972 12glkw4 t1 jflbch7 380he snar POV, 22:33:38 early web in th... 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}
                                   {'total upvote': 2, 'total comments': 1}
        Alphasite
                                   {'total upvote': 2, 'total_comments': 1}
        Asleep-Dress-3578
                                   {'total upvote': 4, 'total comments': 1}
        Barn07
                                   {'total upvote': 2, 'total comments': 1}
        Bling-Crosby
                                   {'total_upvote': 9, 'total_comments': 1}
        CevicheCabbage
                                   {'total upvote': 6, 'total comments': 1}
        Classic Department42
                                   {'total upvote': 3, 'total comments': 1}
        Deto
                                   {'total upvote': 2, 'total_comments': 1}
        DonkeyTron42
        DusikOff
                                   {'total upvote': 5, 'total comments': 1}
        FredVIII-DFH
                                   {'total upvote': 2, 'total comments': 1}
                                   {'total upvote': 4, 'total comments': 2}
        LifeHasLeft
        LittleMlem
                                   {'total upvote': 9, 'total comments': 1}
        Material-Resource-19
                                  {'total upvote': 16, 'total comments': 1}
                                  {'total_upvote': 35, 'total_comments': 1}
        Megatron_McLargeHuge
                                  {'total upvote': 21, 'total_comments': 1}
        NorthImpossible8906
        0erthling
                                   {'total_upvote': 1, 'total_comments': 1}
        Panda Mon
                                  {'total upvote': -8, 'total comments': 1}
                                   {'total upvote': 1, 'total comments': 1}
        RedbloodJarvey
                                  {'total_upvote': 39, 'total_comments': 1}
        Ruin369
        Saphyel
                                   {'total upvote': 2, 'total comments': 1}
                                  {'total upvote': 41, 'total comments': 1}
        Smallpaul
                                  {'total upvote': 90, 'total_comments': 3}
        TheBodyPolitic1
                                   {'total_upvote': 1, 'total_comments': 1}
        _massif_
        aikii
                                  {'total upvote': 53, 'total comments': 1}
                                   {'total upvote': 1, 'total comments': 1}
        arvarnargul
                                   {'total_upvote': 1, 'total_comments': 1}
        atomly
        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}
                                   {'total_upvote': 2, 'total_comments': 1}
        genuinemerit
        glacierre2
                                 {'total_upvote': 384, 'total_comments': 1}
                                  {'total_upvote': 89, 'total_comments': 1}
        gravity_rose
                                  {'total upvote': 57, 'total comments': 1}
        jet heller
                                   {'total upvote': 7, 'total comments': 1}
        iisuskraist
        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}
        auts3
                                  {'total upvote': 23, 'total comments': 2}
                                  {'total upvote': 11, 'total comments': 1}
        roamn
                                 {'total_upvote': 917, 'total_comments': 1}
        robvas
        schplat
                                  {'total_upvote': 36, 'total_comments': 1}
        snapetom
                                   {'total_upvote': 3, 'total comments': 1}
                                  {'total upvote': 26, 'total comments': 1}
        stupid design
        thereshegoes
                                   {'total upvote': 2, 'total comments': 1}
                                   {'total upvote': 6, 'total comments': 1}
        thesolitaire
                                 {'total upvote': 242, 'total comments': 1}
        thewileyone
                                   {'total upvote': 2, 'total_comments': 1}
        tommasz
        turtle4499
                                 {'total upvote': 175, 'total comments': 1}
        ubernostrum
                                   {'total upvote': 3, 'total comments': 1}
                                   {'total upvote': 1, 'total comments': 1}
        ucblockhead
        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()

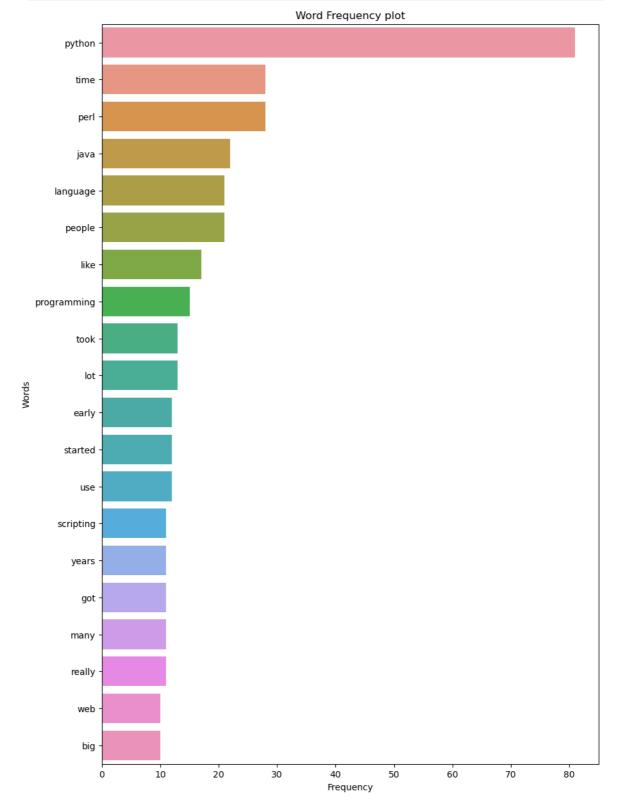
corpus = [word.lower() for comment in comments for word in word_tokenize(
   corpus[:5]

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

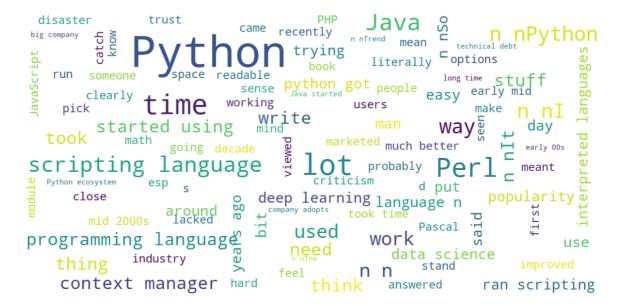
Out[9]: ['hardware', 'ready', 'python', 'time', 'python']

```
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()
         generate_cloud(str(fp_comments['body'].values).replace('\n',' '))
In [12]:
          Languages
is stand
criminal stand
criminal stand
criminal stand
criminal stand
                                                  scripting
                                                                         anguage
                                    readable
                                                 ran scripting n nThe O
                                                                               answered
                                                                marketed C
                                                industry
                                                                    Python ecosystem going
                              language n
                                               started
                                                           using
                         time
                                                                      data science
                                                  viewed
                                        <sup>clear</sup>lyway
          eted
                                                            d
                           took
                                                                   n nTrend
                                                           write
              recently context manager
                                                                          PHP
                                                                            much better
                                                     python got
                                feel
              sense easy
                                                                               used
                               Java
                                                       close know technical debt popularity
                day mid 2000s
                                           mean
                               lacked
                                                     language
             <sub>early mid</sub>think
                                    programming
                                                                          vears ago
                           nliterally math
                                                         nPython
                               hard
                                          disaster
          put
                        esp
                                         book trying
                                                                             working
        long time o
                                                                              options
                                                                    around
                                                   someone
                             space
                                                    modulething criticism Java nmind
         deep learning
                                    ν<sub>early 00s</sub>use
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-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
    fp_comments['comment_sent_score'] = fp_comments['body'].apply(lambda text: polarity_score(text))

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

Out[31]: body comment_sent_score

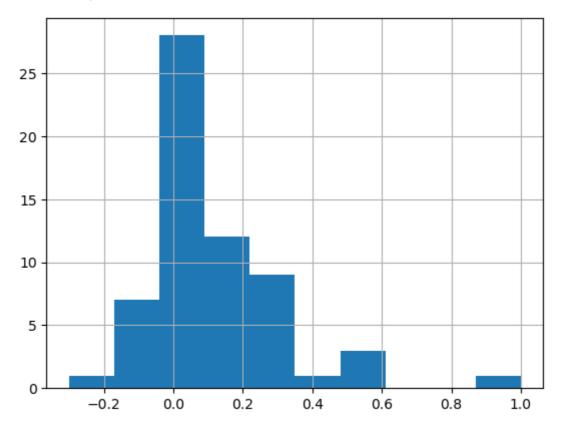
	ner	

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</pre>
```

/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'].app
ly(lambda score:sentiment(score))

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

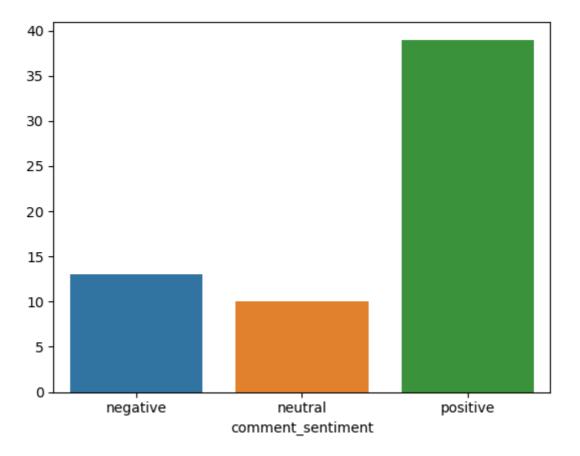
Out[36]:

body comment sentiment

comment_id		
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()
```



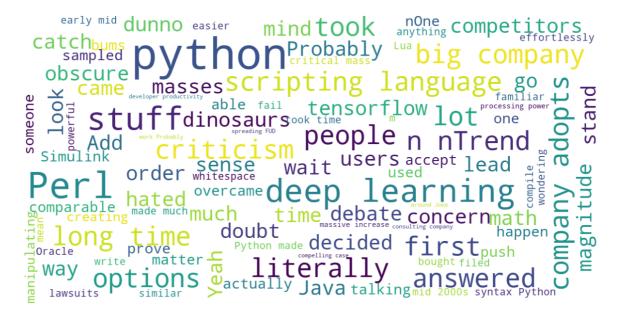
Segmenting comments by sentiment and plotting word cloud

Positive Sentiment

```
generate_cloud(str(fp_comments.loc[fp_comments['comment_sentiment']=='pos
n nThe david beazlev
putmid 90s
                                          ern based ones
  going
                                                     improved
 programming language
                               around
       various web
    think
                       people started
        none web
                               work
                                 network resourced
                     feeln latePHP
      man
                                   popularity
                           solved n
      preted languages
                                                                meant
                        bit book many purposes context manager
            technical debt
 late 90s
```

Negative Sentiment

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



Neutral Sentiment

```
GUI solution Perl issuescompany called Performance went language Java SCripting language Pythonfacto scripting start calling today People Pythonfacto scripting start calling today People Python JavaPython dynamic typing paving compatibility ended sentences Python called Sun move away said sun preps 500m java brand time ran scripting liveScript seentheregister Lua today comment Sun went bye Javan Way de facto Dude learn bye bye Perl' python Dude module JavaScript solution now impression compatibility it' Performance
```