

```
import praw
import pandas as pd
import numpy as np
import re
from collections import Counter
import matplotlib.pyplot as plt
import datetime as dt
from pprint import pprint
import seaborn as sns
```

```

        user_agent = "reddit_scraper",
        username = "EarthMetals",
        password = "FeR87fgt**")

In [112]: # hot posts, rising posts
          headlines = set()

```

```
1  
  
from praw.models import MoreComments  
times = []  
comments = []
```

```

        comments.append(top_level_comment.body)
        times.append(top_level_comment.created_utc)
    print(len(comments))
    print(len(times))

436
436

df = pd.DataFrame(comments, times)
df.head()
df.to_csv(r'C:\Syracuse DataScience\IST 736 - Text

import nltk

```

```
[nltk_data] C:\Users\deane\AppData\Roaming\nltk_data...
[nltk_data] Package vader_lexicon is already up-to-date!
Out[87]: True

In [88]: from nltk.sentiment.vader import SentimentIntensityAnalyzer as SIA
```

```
for line in
```

```
results.append(poi_score)

pprint(results[:3], width=100)
```

```
'headline': '#Ban Bet Lost\\n'
      '\\n'
      '/u/Rodrigo87 (0/1) made a bet that BABA would go to 175.0 when it was 165.02 and it
      'did not, so they were banned for a week.',
'neg': 0.182,
'neu': 0.818,
'pos': 0.0],
{'compound': -0.7996,
'headline': 'You gotta feel for CRSR gang in 2021.\\n'
      '\\n'
      "During one of the biggest bull run ever, they've been absolutely fucked non stop.",
'neg': 0.264,
'neu': 0.736,
'pos': 0.0],
{'compound': 0.2263,
'headline': 'You guys ever do something like say,buy SPY calls for $0.40, sell them later for '
      '$0.85, think you're a genius selling the top, and then watch them shoot up to about
      '$3.15?',
'neg': 0.069,
'neu': 0.807,
'pos': 0.124]]
```

```
df = df.rename(columns = {'headline':'comments'})
df.head()
```

Out[89]:

timestamp	neg	neu	pos	compound	comments
-----------	-----	-----	-----	----------	----------

	1.634294e+09	0.264	0.736	0.000	-0.9396	rod gotta feel for CSK gang in 2021. why? (reddit.com)
2	1.634298e+09	0.069	0.807	0.124	0.2263	You guys ever do something like say...buy SPY ca...
3	1.634300e+09	0.333	0.535	0.132	-0.6855	My spy puts will most likely expire worthless ...
4	1.634305e+09	0.189	0.811	0.000	-0.1027	Is it gay to pay federal taxes?

```
#creating thresholds
df['label'] = 0
df.loc[df['compound'] > 0.2, 'label'] = 1
df.loc[df['compound'] < -0.2, 'label'] = -1
df.head()
```

1	1.634294e+09	0.264	0.736	0.000	-0.7996	You g
2	1.634298e+09	0.069	0.807	0.124	0.2263	You gu
3	1.634300e+09	0.333	0.535	0.132	-0.6855	Ms

```
In [91]: df1 = df[['comments', 'label']]
df1.label.value_counts()
#normalized to percent
df1.label.value_counts(normalized=True)
```

```

0      43.577982
-1     28.440367
1     27.981651
Name: label, dtype: float64

print("Positive Comments: \n")
pprint(list(df[df['label'] ==1].comments)[:5])
```

Positive Comments:

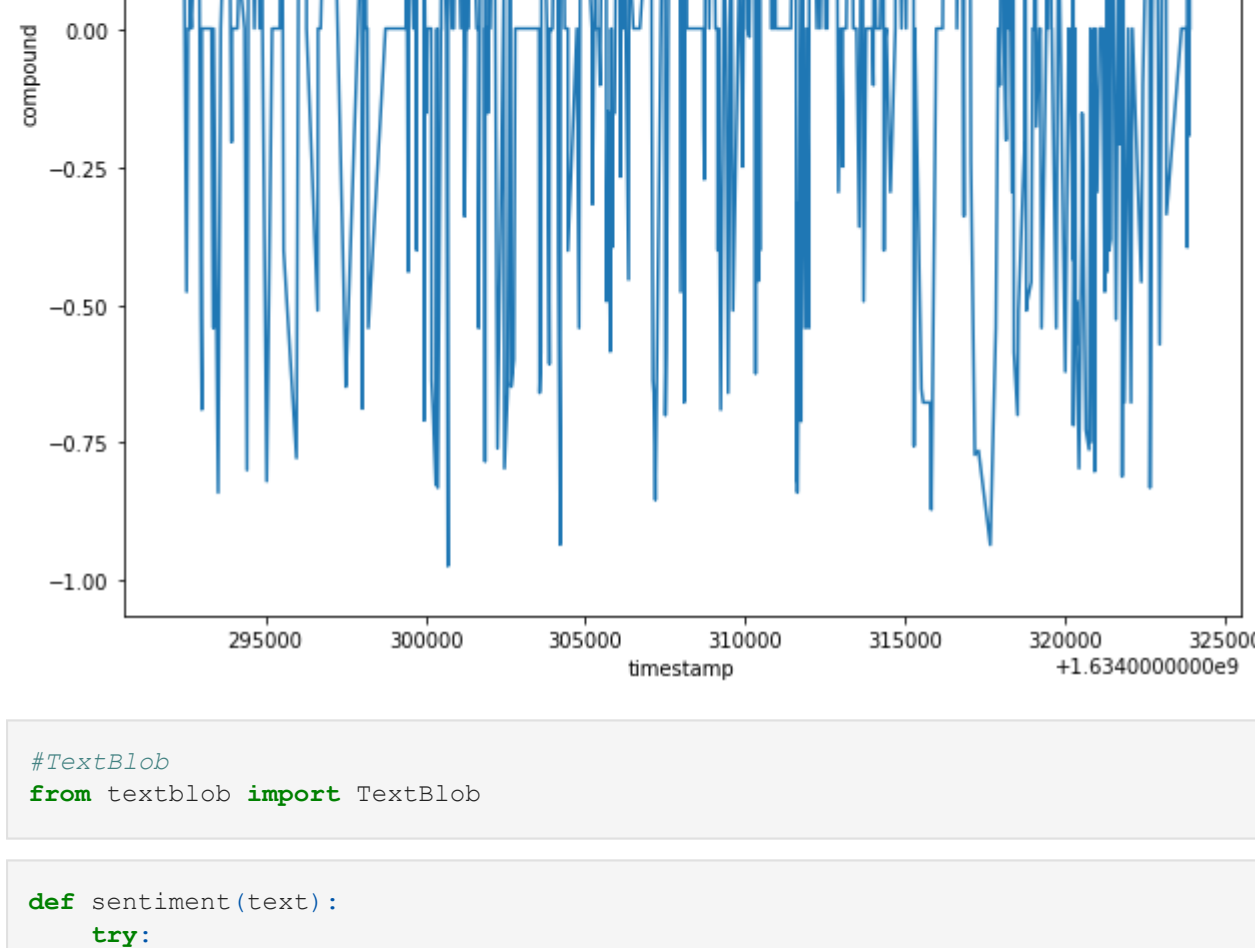
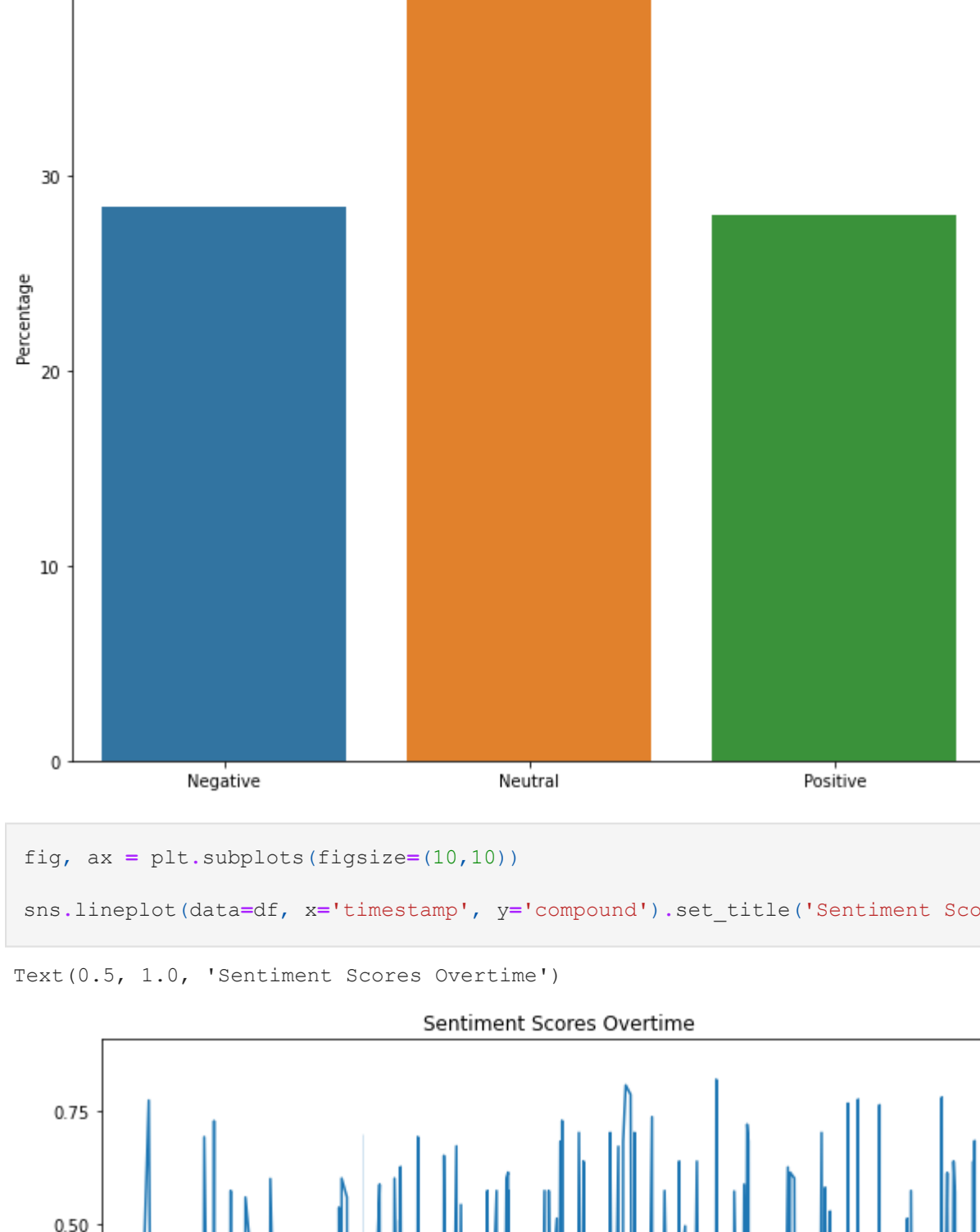
"Every time I worry about my portfolio I scroll through the daily thread and get a good laugh at all the flaming dogshit people are bag holding.",
 "I've been watching the market get absolutely wrecked. Bull market back in action confirmed. I expected him to need another week of confirmation to be honest.",
 "Watching UPST is like watching a yacht go by and all my friends are on it, waving and saying we told you to get out on".
 "I could have been a very rich man swimming in pussy juice if I did the opposite of everything I've done since last year."]
 Negative Comments:
 "[!\$Ban Bet Lost\n\n/u/Rodriguez87 (0/1) made a bet that BABA would go to 175.0 when it was 165.02 and it did not, so they were banned for a week.",
 "You gotta feel for CRSR gang in 2021.\n\nDuring one of the biggest bull run ever, they've been absolutely fuc

'My buddy was loaded to the tits on SPCE with a low cost basis. He held even during the run up to the 60s-70s saying \$100 was his exit target.\n'
'\n'
'Today he will wake up and his position will be red for the first time (not 100% of his cost basis but it should be in high teens). Hundreds of thousands in gains gone.\n'
'\n'
'Should I:\n'
'\n'
'1. Call him to invite him to lunch, check up on him.\n'
'2. Text him pic of Greta smiling with his stupid ass cowboy hat on.,
'3. TSLA really is unstoppable isn't it lol \n\nIt's completely broken the shell of a man once known as Burry and just. keeps. going.']

```
sns.barplot(x=counts.index, y=counts, ax=ax).set_title('WSB Daily Thread Sentiment-VADER')

ax.set_xticklabels(['Negative', 'Neutral', 'Positive'])
ax.set_ylabel('Percentage')

plt.show()
```



```

return None

```

```
<ipython-input-97-1a138c8115ef>:1: Setting
A value is trying to be set on a copy of
Try using .loc[row_indexer,col_indexer]
See the caveats in the documentation: ht
urning-a-view-versus-a-copy
df[['sentiment']] = df[['comments']].appr
```

```
columns = ['polarity', 'subjectivity']
tb = pd.DataFrame(sentiment_series, columns=columns, index=df1.index)
```

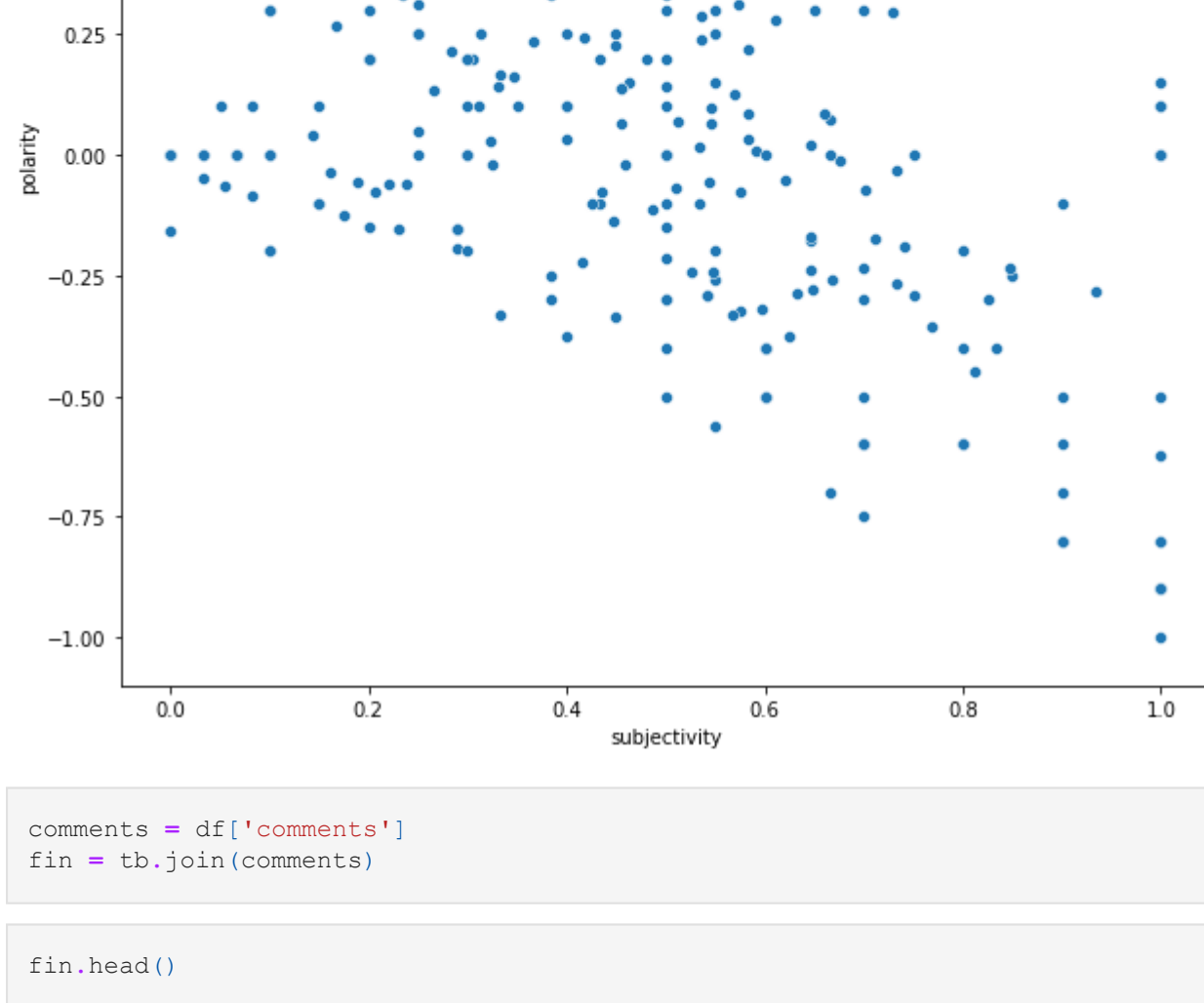
	polarity	subjectivity
0	0.000000	0.000000
1	-0.600000	0.700000

3	-0.191667	0.741667
4	0.416667	0.583333

```
fig, ax = plt.subplots(figsize=(10,10))
sns.scatterplot(data=tb, x='subjectivity', y='polarity').set_title('')
Text(0.5, 1.0, 'Polarity by Subjectivity')
```

Year	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Share of GDP	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0	2.1

Number of children	Number of children who are not in the household
10	0.5



0	0.000000	0.000000	#Ban
1	0.600000	0.700000	Yes

3 -0.19166

```
#creating thresholds
fin['label'] = 0
fin.loc[fin['polarity'] > 0.2, 'label'] = 1
fin.loc[fin['polarity'] < -0.2, 'label'] = -1
fin.head()
```

0	0.000000	0.000000	#Ban Bet Lost\n\nu/Rodrigo87 (0/
1	-0.600000	0.700000	You gotta feel for CRSR gang in 2

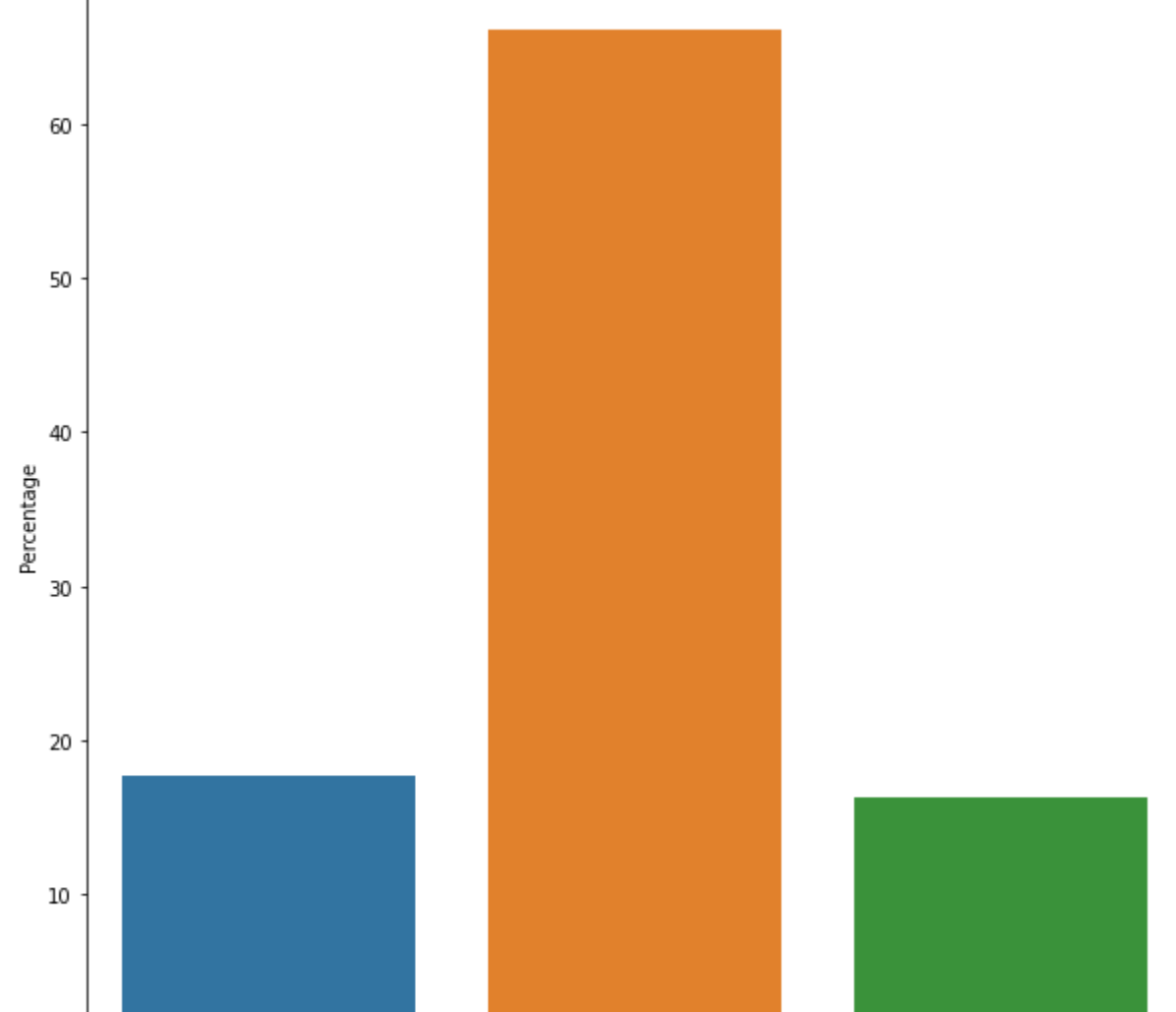
	df1	df2	df3	df4
3	-0.191667	0.741667	My spy puts will most likely expire worthless ...	0
4	0.416667	0.583333	Is it gay to pay federal taxes?	1

```
df2 = fin[['comments', 'label']]
```

```
Out[104]: 0 66.055046
```

```
fig, ax = plt.subplots(figsize=(10,10))
counts = df2.label.value_counts(normalize=True)
sns.barplot(x=counts.index, y=counts, ax=ax)
```

```
plt.show()
```



```
print("Negative Comments:\n")
pprint(list(df2[df2['label'] == -1].comments)[:5], width=200)

Positive Comments:

['You guys ever do something like say...buy SPY calls for $0.40
selling the top, and then watch them shoot up to about $3.15?
```

dogshit people are bag holding.',
 'Jim Cramer the type of dude to order unsalted fries and add salt to them',
 'Bought \$30 worth of spy calls for fun, sold for \$40. Paid a \$10 commission to vanguard. No wonder banks are making so much',
 Negative Comments:
 'You gotta feel for CRSR gang in 2021.\n\nDuring one of the biggest bull run ever, they've been absolutely fucked non stop.',
 'Bnat Bet won't\n\n/n\n/nAlmightysnoo made a bet that TSLA would go to \$250. when it was \$11.1 and it did, comfats retard.',
 'The market reliance relies entirely on the bulls decision to take profit\n\nBottom feeders and retards',
 'I'm so fucked. I accidentally hit confirm on ldtc SPY PUTS with my last \$1000. There I go',
 'I remember the good old days you buy some OTM options they expire worthless a week later and you shutup and m

"Why can't can't you retards just buy worthless options like the rest of us?"]