In [1]:
```
#Import PRAW API to access reddit posts

import pandas as pd
import pprint
import praw

#Import packages for numpy, regex, nltk and matplotlib

import numpy as np
import re
import nltk
import matplotlib.pyplot as plt
from datetime import datetime

#Read txt file with saved password info

with open('pw.txt', 'r') as f:
    pw = f.read()

#Install PRAW if necessary
#!pip install praw
```

In [2]:
```
#Create reddit instance with log-in information

reddit = praw.Reddit(
    client_id='SuiQzO3Say1QCQ',
    client_secret='pkkqLIy2F8oYQHvdZ9iJwp86Cz_8MQ',
    user_agent="my user agent",
    username='leechmilyfe4',
    password= pw,
)

print(reddit.read_only)
```

```
False
```

Parsing data

In [3]:
```
#Create subreddit object naming particular subreddit of study

subreddit = reddit.subreddit("wallstreetbets")
```

In [4]:
```
#Create loop to output post information based on top rated posts in subreddit

for submission in subreddit.top(limit=20):
    print(submission.title)
    # Output: the submission's title
    print(submission.score)
    # Output: the submission's score
    print(submission.id)
    # Output: the submission's ID
    print(submission.url)
    # Output: the URL the submission points to or the submission's URL if it's a
```

```
Times Square right now
411403
l8rf4k
https://v.redd.it/x64z70f7eie61 (https://v.redd.it/x64z70f7eie61)
UPVOTE so everyone sees we got SUPPORT
314908
l6wu59
https://i.redd.it/sgoqy8nyt2e61.png (https://i.redd.it/sgoqy8nyt2e61.png)
GME YOLO update — Jan 28 2021
280491
l78uct
https://i.redd.it/opzucppb15e61.png (https://i.redd.it/opzucppb15e61.png)
GME YOLO month-end update — Jan 2021
251696
l846a1
https://i.redd.it/r557em3t5ce61.png (https://i.redd.it/r557em3t5ce61.png)
It’s treason then
225041
l881ia
https://i.redd.it/d3t66lv1yce61.jpg (https://i.redd.it/d3t66lv1yce61.jpg)
CLASS ACTION AGAINST ROBINHOOD. Allowing people to only sell is the definition
 of market manipulation. A class action must be started, Robinhood has made ple
nty of money off selling info about our trades to the hedge funds to be able to
pay out a little for causing people to loose money now
221967
l6x130
https://www.reddit.com/r/wallstreetbets/comments/l6x130/class_action_against_ro
binhood_allowing_people_to/ (https://www.reddit.com/r/wallstreetbets/comments/l
6x130/class_action_against_robinhood_allowing_people_to/)
GME YOLO update — Feb 19 2021
215992
lnqgz8
https://i.redd.it/2xswz0h11ii61.png (https://i.redd.it/2xswz0h11ii61.png)
GME YOLO update — Jan 27 2021 ------------------------------------ guess i n
eed 102 characters in title now
210968
l6ekdz
https://i.redd.it/a309gkm5yxd61.png (https://i.redd.it/a309gkm5yxd61.png)
Used some of my GME tendies to buy Nintendo Switches from Gamestop, then donate
d them to a Children's Hospital. Got featured on the local news and brought glo
ry to WSB.
210608
l8c0u4
https://www.nbcdfw.com/news/local/north-texas-investor-uses-gamestop-gains-help
-sick-children/2537134/ (https://www.nbcdfw.com/news/local/north-texas-investor
```

-uses-gamestop-gains-help-sick-children/2537134/)
GME YOLO update — Feb 1 2021
210586
lae6j0
https://i.redd.it/og8ca1xskxe61.png (https://i.redd.it/og8ca1xskxe61.png)
IT'S POWER TO THE TRADERS NOW
209380
l7feld
https://v.redd.it/lu8aekujd6e61 (https://v.redd.it/lu8aekujd6e61)
Wow super bowl commercial for us
195699
lexy8t
https://i.redd.it/tko5nh1qy4g61.jpg (https://i.redd.it/tko5nh1qy4g61.jpg)
Can we all take a moment and appreciate the Mods who have grinded through this and held our community together. Thank you.
195008
l6jobf
https://www.reddit.com/r/wallstreetbets/comments/l6jobf/can_we_all_take_a_moment_and_appreciate_the_mods/ (https://www.reddit.com/r/wallstreetbets/comments/l6jobf/can_we_all_take_a_moment_and_appreciate_the_mods/)
That’s what I thought
192973
l890i7
https://i.redd.it/f2yhugwt5de61.jpg (https://i.redd.it/f2yhugwt5de61.jpg)
It runs very deep, my friends.
185920
l79x17
https://i.redd.it/b8zy92sy85e61.jpg (https://i.redd.it/b8zy92sy85e61.jpg)
I am proud to do my part in paying forward our good fortune with a donation of 6 Nintendo Switches and games to go with them to the Children’s Minnesota Hospital. Cant Stop. Won’t Stop. GameStop. (Still long 50 shares I WILL NOT SELL)
185701
l90oq6
https://i.redd.it/sgrbt9vthke61.jpg (https://i.redd.it/sgrbt9vthke61.jpg)
Crazy mannnnnnn. We can't let this slide at all
182075
l74tr1
https://i.redd.it/8bels7zob4e61.jpg (https://i.redd.it/8bels7zob4e61.jpg)
Drew this for you my fellow retards 💎🙌
181325
l9kn3z
https://i.redd.it/h7sap8vb5qe61.jpg (https://i.redd.it/h7sap8vb5qe61.jpg)
Mark Cuban spent nearly 2 hours answering questions in his AMA. And then immediately called into CNBC to defend WSB!
178524
lazmky
https://i.redd.it/5gwwn8rxb3f61.jpg (https://i.redd.it/5gwwn8rxb3f61.jpg)
AOC got our backs
178241
l6fg8q
https://i.redd.it/3v5c5quc5yd61.png (https://i.redd.it/3v5c5quc5yd61.png)

ID 1

In [5]:
```
#Verify comment output of ID 1
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l8rf4k")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)

#Import MoreComments to remove "Load More Comments" and "Continue this Thread" li

from praw.models import MoreComments


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Get on the rocket before its to late #SafeGalaxy.

"To infinity and beyond 🚀"

Not financial advice so do your own research.
this is second on r/all if u sort by top posts of all time
+1 (904) 401-6195
.
.
.
.
.
.
.
.
.
.
.
.
.
```

In [6]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score', '
comment_df
```

Out[6]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gx...	gxhz01m	1	1.620598e+09	Get on the rocket before its to late #SafeGala...
1	<bound method Comment.parent of Comment(id='gx...	gxe5jsw	2	1.620515e+09	this is second on r/all if u sort by top posts...
2	<bound method Comment.parent of Comment(id='gx...	gxdn0qu	1	1.620503e+09	+1 (904) 401-6195
3	<bound method Comment.parent of Comment(id='gx...	gxbsg0e	1	1.620456e+09	.
4	<bound method Comment.parent of Comment(id='gx...	gxbsfn2	1	1.620456e+09	.
...	...	...	...	...	...
274	<bound method Comment.parent of Comment(id='gn...	gnkjub1	1	1.613450e+09	[removed]

ID 2

In [7]:
```
#Verify comment output of ID 2
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l6wu59")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Get on the rocket before its to late #SafeGalaxy.

"To infinity and beyond 🚀"

Not financial advice so do your own research.
But is anyone here talking about the flips WSB can do? Never forget who runs
this show now.
Do you think they had Chick-fil-A sauce in the USSR? 'MERICA!
I mean...didn't really blow up that hard on the day this happened...oh well
I am not a American citizen.. But can't u all sue this Robinhood for illegal
practices??. Or at least switch all away from the platfor?. They seem so obvi
ous poisonous it looks like a Disney movie. 👀
Don't mind if I do! Screw utilities.  I'll but more gme
I'm pretty sure they did it to PLTR and BB too.
I am so mad that I missed out on this while it was still affordable. Even a s
mall investment would have made my life leages easier.
W
Since this post will go down in the history books I wanna say that Noita is a
really good game and it is really fun and I really want multiplayer
```

In [8]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df2 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df2
```

Out[8]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gx...	gxhz0sk	1	1.620598e+09	Get on the rocket before its to late #SafeGala...
1	<bound method Comment.parent of Comment(id='gx...	gxeivew	1	1.620521e+09	But is anyone here talking about the flips WSB...
2	<bound method Comment.parent of Comment(id='gw...	gwv004w	1	1.620124e+09	Do you think they had Chick-fil-A sauce in the...
3	<bound method Comment.parent of Comment(id='gw...	gwemvp9	1	1.619804e+09	I mean...didn't really blow up that hard on th...
4	<bound method Comment.parent of Comment(id='gw...	gw11p9u	2	1.619541e+09	I am not a American citizen.. But can't u all ...
...	...	...	...	...	...
87	<bound method Comment.parent of Comment(id='gl...	glnegcc	2	1.612242e+09	How is no one seeing $XL lol that stock is sho...
88	<bound method Comment.parent of Comment(id='gl...	gln072i	1	1.612237e+09	It seem that those are also blocked in the deg...
89	<bound method Comment.parent of Comment(id='gl...	glmwkaf	2	1.612236e+09	But why isn't AMC going up ?
90	<bound method Comment.parent of Comment(id='gl...	glmuu4n	6	1.612235e+09	I would love you guys to spread the word that ...
91	<bound method Comment.parent of Comment(id='gl...	glmtp7j	4	1.612234e+09	AMC 💎 □□

92 rows × 5 columns

ID 3

In [9]:
```
#Verify comment output of ID 3
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l78uct")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Get on the rocket before its to late #SafeGalaxy.

"To infinity and beyond 🚀"

Not financial advice so do your own research.
then double it
You fuck
Love looking hack at this and being like double it,then double it again!
Long live the king □.
Hi
I don’t understand stocks, what does this mean?
u/award-cost-bot
Please do a video dfv for president. Apes2024
Uh, he sold 50,000 shares. His last post he has 100,000 shares.

Diamond hands don't trade.
□□□□
big bruh
monke hape
```

In [10]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df3 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df3
```

Out[10]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gx...	gxhz1mu	1	1.620599e+09	Get on the rocket before its to late #SafeGala...
1	<bound method Comment.parent of Comment(id='gw...	gw676a9	1	1.619647e+09	then double it
2	<bound method Comment.parent of Comment(id='gw...	gw66web	1	1.619647e+09	You fuck
3	<bound method Comment.parent of Comment(id='gv...	gvtdo14	1	1.619386e+09	Love looking hack at this and being like doubl...
4	<bound method Comment.parent of Comment(id='gv...	gvp72vb	1	1.619301e+09	Long live the king □.
...	...	...	...	...	...
239	<bound method Comment.parent of Comment(id='gl...	glbiu2x	1	1.612000e+09	What do you guys think about MARPS, can we ear...
240	<bound method Comment.parent of Comment(id='gl...	glbhy04	2	1.612000e+09	Can someone please explain to me this second l...
241	<bound method Comment.parent of Comment(id='gl...	glbhpo8	1	1.611999e+09	Are we still buying NOK? Im $600 deep in GME
242	<bound method Comment.parent of Comment(id='gl...	glbh7mz	1	1.611999e+09	GME come over to OPK. Top 20 most shorted comp...
243	<bound method Comment.parent of Comment(id='gl...	glbfj0j	4	1.611998e+09	IF HE'S STILL IN, I'M STILL IN

244 rows × 5 columns

ID 4

In [11]:
```
#Verify comment output of ID 4
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l846a1")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Hi Guys,
What are your thought about GME atm?


will we see a spike again?
ALL GREEN BABY
I stay
u/award-cost-bot
If he’s still in I’m still in
Update with Congress testimony. To help illustrate GME GameStop current short
positions.

https://www.reddit.com/r/wallstreetbets/comments/lsbv5w/repost_but_you_might_
want_to_watch_this_again/?utm_source=share&utm_medium=ios_app&utm_name=iossmf
(https://www.reddit.com/r/wallstreetbets/comments/lsbv5w/repost_but_you_might
_want_to_watch_this_again/?utm_source=share&utm_medium=ios_app&utm_name=iossm
f)
sheesh let me get just 100000 😁
It's starting to drop slowly. Get those limit orders in boys and girls! Not f
```

In [12]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df4 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df4
```

Out[12]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gx...	gxivvzb	1	1.620615e+09	Hi Guys, \nWhat are your thought about GME at...
1	<bound method Comment.parent of Comment(id='gw...	gwaaxv0	0	1.619726e+09	ALL GREEN BABY
2	<bound method Comment.parent of Comment(id='gs...	gs3qybu	1	1.616654e+09	I stay
3	<bound method Comment.parent of Comment(id='gp...	gp43f0d	1	1.614510e+09	u/award-cost-bot
4	<bound method Comment.parent of Comment(id='go...	govoyxf	1	1.614400e+09	If he’s still in I’m still in
...	...	...	...	...	...
156	<bound method Comment.parent of Comment(id='gl...	gli76nj	1	1.612147e+09	I am hungry for some more GME! Better than fru...
157	<bound method Comment.parent of Comment(id='gl...	gli6s57	1	1.612147e+09	Next gme is General Electric buy buy
158	<bound method Comment.parent of Comment(id='gl...	gli5q3y	1	1.612146e+09	What are the other FD s we are all investing?
159	<bound method Comment.parent of Comment(id='gl...	gli4upq	1	1.612146e+09	I’m in on the #biowar BCRX squeeze Monday morn...
160	<bound method Comment.parent of Comment(id='gl...	gli46t0	3	1.612146e+09	Just spend whatever you can to be a part of th...

161 rows × 5 columns

ID 5

In [13]:
```
#Verify comment output of ID 5
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l881ia")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
A guy opens his account and sells his share
You thinks that of me?
NO I AM THE ONE WHO BUYS
I voted yesterday. When we vote is just by person or by share? Just wondering
if the more shares the more the vote counts?
[CBAT stock is my next bet!](https://www.google.com/amp/s/www.cnbc.com/amp/20
21/04/08/the-us-is-facing-a-lithium-ion-battery-shortage-with-ev-growth.html)
I don’t know shit about stocks, investing, or literally anything else.  But I
read, I see something unique happening in the news/here, I wish I understood,
I support everyone here.
*It's free stonks*
Its fun being part of a community even if you lose money
First majestic all the way
I wish my bowel movements worked just as good.
Zomedica about to blow up!
Hey i have 1.01 shares, I AM THE MOVEMENT TOO :')
WERE GOING BEAR HUNTING TODAY

AMC AMC AMC AMC AMC AMC AMC
```

In [14]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df5 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df5
```

Out[14]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gx...	gxhp458	1	1.620592e+09	A guy opens his account and sells his share\nY...
1	<bound method Comment.parent of Comment(id='gx...	gxawrd2	1	1.620441e+09	I voted yesterday. When we vote is just by per...
2	<bound method Comment.parent of Comment(id='gw...	gw9zavq	1	1.619717e+09	[CBAT stock is my next bet!](https://www.googl...
3	<bound method Comment.parent of Comment(id='gu...	guoezl2	1	1.618567e+09	I don’t know shit about stocks, investing, or ...
4	<bound method Comment.parent of Comment(id='gs...	gss5y1j	1	1.617124e+09	*It's free stonks*
...	...	...	...	...	...
367	<bound method Comment.parent of Comment(id='gl...	glcd6ep	1	1.612017e+09	I work with homeless people. It'd be real neat...
368	<bound method Comment.parent of Comment(id='gl...	glcd4jk	1	1.612017e+09	I curently have 6 shares in AMC but robinhoods...
369	<bound method Comment.parent of Comment(id='gl...	glcd4do	1	1.612017e+09	😂😂
370	<bound method Comment.parent of Comment(id='gl...	glcd0zr	1	1.612017e+09	I feel attacked
371	<bound method Comment.parent of Comment(id='gl...	glccviw	1	1.612016e+09	This is too accurate

372 rows × 5 columns

ID 6

In [15]:
```
#Verify comment output of ID 6
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l6x130")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Still waiting on Decision from First Lawsuit. Got 256k riding on fhe line.
Let Round 2 Begin!
Only way to make Robinhood bleed is to drain it dry. Cash out and close your
accounts, open them with another service. It’s the only thing people like tha
t understand.
So should I not use RH?
Let’s roll!
I canceled several limit buys just now but my own money won’t go back into my
buying power!!!!! WHY?!!emote:free\_emotes\_pack:dizzy\_face
File the in referenced  class action before the IPO Robin hood plans if my in
formation Is Correct.  If media do releases of public articles of a pending l
aw suit, that has been filed in the courts. The Civil  class action needs to
seek financial  damages of persons and stock holders. An outstanding claim by
class action suit would have to factored into the Value IPO of contingent lia
bilities that the   "Robin Hoods" have exposure to in the class suit!!!!

Thus .. IPO buyers  will not pay as high a value for the offering price.

DISCLAIMER: this is not legal advice nor is it advise to invest in
```

In [16]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df6 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df6
```

Out[16]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gx...	gx41o07	1	1.620302e+09	Still waiting on Decision from First Lawsuit. ...
1	<bound method Comment.parent of Comment(id='gw...	gwtgnua	1	1.620098e+09	Only way to make Robinhood bleed is to drain i...
2	<bound method Comment.parent of Comment(id='gw...	gw22643	1	1.619564e+09	So should I not use RH?
3	<bound method Comment.parent of Comment(id='gv...	gvqx9dn	1	1.619326e+09	Let’s roll!
4	<bound method Comment.parent of Comment(id='gv...	gvj20xn	1	1.619179e+09	I canceled several limit buys just now but my ...
...	...	...	...	...	...
190	<bound method Comment.parent of Comment(id='gl...	glaoqch	1	1.611986e+09	I wasn’t allowed to buy shares today! Received...
191	<bound method Comment.parent of Comment(id='gl...	glamh51	1	1.611985e+09	If an exchange restricts buying, does that mea...
192	<bound method Comment.parent of Comment(id='gl...	glakxyz	1	1.611985e+09	Just sold all my shares and they will not let ...
193	<bound method Comment.parent of Comment(id='gl...	glakxjt	1	1.611984e+09	Where is the best alternative? RobinHood cost...
194	<bound method Comment.parent of Comment(id='gl...	glab8sg	2	1.611981e+09	How do I fully delete my account? I'm sure the...

195 rows × 5 columns

ID 7

In [17]:
```
#Verify comment output of ID 7
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="lnqgz8")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Holy shit, buy me a happy meal please
Hero. Legend. GOAT.
How did you only pay 26$ when they were going for 40$?? Teach me da wayyy
Buy more today!!!
My body is yours.
I’m gonna leave this here to show how Melvin and Citadel  in Congress. Gabe
of Melvin Capital said not short covering and price went up. Which means Cita
del CEO  said it was short covering. At that time it was 140% short interes
t. And CEO of Interactive Broker which also stopped trading, the day before s
aid only 50 million shares available. Which CEO of Citadel and Melvin said ki
nda of maybe cover 30 to 40 million, which Gabe said it wasn’t shorts coverin
g. Would have need 140% of all available stock to cover, which only 50 millio
n were available and from that maybe covered 30 to 40 million. So with hedges
own statements they  to Congress. https://www.reddit.com/r/wallstreetbets/co
mments/lsbv5w/repost_but_you_might_want_to_watch_this_again/?utm_source=share
&utm_medium=ios_app&utm_name=iossmf (https://www.reddit.com/r/wallstreetbets/
comments/lsbv5w/repost_but_you_might_want_to_watch_this_again/?utm_source=sha
re&utm_medium=ios_app&utm_name=iossmf)
Just bought 5k worth, this is my YOLO / Balls Deep I believe in this stock, g
amers for life!
```

In [18]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df7 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df7
```

Out[18]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gv...	gvif50u	1	1.619167e+09	Keith. Keith. Keith. Keith. Keith!
1	<bound method Comment.parent of Comment(id='gv...	gvgt0sv	1	1.619141e+09	31 Mil volume mvis □ strong 🦡🐻 naked
2	<bound method Comment.parent of Comment(id='gu...	guuuu8r	1	1.618700e+09	Coming back here to comment again. Until yeste...
3	<bound method Comment.parent of Comment(id='gu...	gusfl65	1	1.618649e+09	u/award-cost-bot
4	<bound method Comment.parent of Comment(id='gu...	guak7qx	1	1.618287e+09	Triggered redditors trying to raise the award ...
...	...	...	...	...	...
210	<bound method Comment.parent of Comment(id='go...	gon0w05	2	1.614237e+09	YOLO update?
211	<bound method Comment.parent of Comment(id='go...	gon037r	4	1.614236e+09	SENSEI WHERE IS THE UPDATE?
212	<bound method Comment.parent of Comment(id='go...	gon002k	3	1.614236e+09	Are you watching now? Woop woop
213	<bound method Comment.parent of Comment(id='go...	gomzh3m	3	1.614236e+09	what a champ
214	<bound method Comment.parent of Comment(id='go...	gomz6lv	10	1.614236e+09	How?HOW? You keep predicting the future. You g...

215 rows × 5 columns

ID 8

In [19]:
```
#Verify comment output of ID 8
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l6ekdz")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
ABSOLUTE FUCKING LEGEND
I WAS HERE
LEAVING A COMMENT FOR HISTORY
The biggest bet in WSB history.
How the fuck does this shit work somebody please explain it to me
This is the highest upvoted post I've ever seen on reddit what the fuck.  Jus
t searched this guys post history and my mind is blown.  I need to learn how
to play this game, but I just cant figure out where to start.
Good stuff
This is the way
Commenting here to be part of history. Hi son!
Zomedica about to blow up!🚀🚀🚀
Glad I can still comment here to be a part of it when it’s archived.
Figure of speech
Should've sold..
BOSS
Bro should've sold right here.. ![gif](emote|free_emotes_pack|grimacing)
I think his position must have been worth over $70m when GME was close to $50
0/share
```

In [20]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df8 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df8
```

Out[20]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gu...	gux918t	1	1.618738e+09	ABSOLUTE FUCKING LEGEND \nI WAS HERE \nLEAVI...
1	<bound method Comment.parent of Comment(id='gt...	gtiulub	1	1.617703e+09	The biggest bet in WSB history.
2	<bound method Comment.parent of Comment(id='gs...	gs480lm	1	1.616662e+09	How the fuck does this shit work somebody plea...
3	<bound method Comment.parent of Comment(id='gs...	gs47fvr	1	1.616662e+09	This is the highest upvoted post I've ever see...
4	<bound method Comment.parent of Comment(id='gs...	gs3qwh7	1	1.616654e+09	Good stuff
...	...	...	...	...	...
204	<bound method Comment.parent of Comment(id='gl...	gl5bdqw	2	1.611897e+09	does any1 know which brokerage DFV is using??
205	<bound method Comment.parent of Comment(id='gl...	gl5adt8	0	1.611897e+09	I'm so confused. Is DFV rich now? Or does he h...
206	<bound method Comment.parent of Comment(id='gl...	gl59zub	2	1.611897e+09	respect holy shit
207	<bound method Comment.parent of Comment(id='gl...	gl59tkl	9	1.611897e+09	This link shows top shorted stocks. I think it...
208	<bound method Comment.parent of Comment(id='gl...	gl59qgy	3	1.611897e+09	Give us some hope you are the only person that...

209 rows × 5 columns

ID 9

In [21]:
```
#Verify comment output of ID 9
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l8c0u4")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
This makes me so fucking happy bro!!! Need more like you OP
“This is the way.”
Hopefully the controllers don’t have drift problems!
First majestic (AG) all the way
Some say he's a true Tendyman
Nicely done!
21 shares @ 170 let's gooo 🚀
That’s awesome you did that, major props. Looks like you’re not doing so bad
yourself cause you got the Elon Mobile!
This is what the fair market is for.
Take note Karl Marx! Lol.
How good does it feel to jack off with those diamond hands? Serious Question
That's what it's All about!!!🚀🚀🚀🚀🚀🚀🚀🚀
Who else got in on zenabis for. 16 and .17 cents big announcement coming from
them out here in Canada c🚀 10k diamond hands til 12 dollars
DNN
You brought shame not glory this not what wsb was about
**DISCLAIMER:  DONATING TO CHARITY IS OBVIOUSLY A GOOD THING.  I AM GLAD THAT
THESE KIDS GET THEIR NINTENDO SWITCHES!!!**
```

In [22]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df9 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df9
```

Out[22]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gx...	gx5h9k4	1	1.620339e+09	This makes me so fucking happy bro!!! Need mor...
1	<bound method Comment.parent of Comment(id='gt...	gtbvq7x	1	1.617554e+09	“This is the way.”
2	<bound method Comment.parent of Comment(id='gr...	grcm4vg	1	1.616100e+09	Hopefully the controllers don’t have drift pro...
3	<bound method Comment.parent of Comment(id='gp...	gpo9kzn	1	1.614907e+09	First majestic (AG) all the way
4	<bound method Comment.parent of Comment(id='gp...	gpbw7tx	1	1.614655e+09	Some say he's a true Tendyman
...	...	...	...	...	...
343	<bound method Comment.parent of Comment(id='gl...	gle4ho7	1	1.612060e+09	LoL — HOLD !!
344	<bound method Comment.parent of Comment(id='gl...	gle42k8	1	1.612060e+09	It’s a media ploy so others will follow!!!!
345	<bound method Comment.parent of Comment(id='gl...	gle3tza	1	1.612060e+09	Absolutely beautiful gesture. I fucking love O...
346	<bound method Comment.parent of Comment(id='gl...	gle3n91	1	1.612060e+09	That’s what up
347	<bound method Comment.parent of Comment(id='gl...	gle3d7c	1	1.612059e+09	May you all prosper for many years! It’s remar...

348 rows × 5 columns

ID 10

In [23]:
```
#Verify comment output of ID 10
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="lae6j0")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Im from the future.

You made it.
Impeccable timing! This deserves to be the all-time top upvoted post of WSB
GME and AMC on discount prices right now.
Projected to hit $385 tomorow on short squeeze
Imbecil
In short, I like the stock!
AKA “roaringking kitty”
DEEPFUCKINGVALUE THE MAN
Fuck off
[removed]
And I made $11

hi my name is mr. proud
I really hope you sold at the peak and been forging numbers ever since.
Did you really sell u/DeepFuckingValue ?!Or am I seeing fake news everywhere?
Who is this guy and where did he work before,  or still works, before this?
✨????
```

In [24]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df10 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df10
```

Out[24]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gu...	gurp81z	1	1.618636e+09	Im from the future.\n\nYou made it.
1	<bound method Comment.parent of Comment(id='gs...	gs1bgai	1	1.616614e+09	Impeccable timing! This deserves to be the all...
2	<bound method Comment.parent of Comment(id='go...	goqwofi	1	1.614313e+09	GME and AMC on discount prices right now.
3	<bound method Comment.parent of Comment(id='go...	gonecw8	1	1.614243e+09	Projected to hit $385 tomorow on short squeeze
4	<bound method Comment.parent of Comment(id='go...	go74vfc	1	1.613913e+09	Imbecil
...	...	...	...	...	...
68	<bound method Comment.parent of Comment(id='gl...	glrhb78	19	1.612319e+09	20@285 shitting my pants but I'll hold. The sh...
69	<bound method Comment.parent of Comment(id='gl...	glrh4w5	4	1.612319e+09	YOU NEED SUPPORT? SOLDIER REPORTING FOR DUTY!
70	<bound method Comment.parent of Comment(id='gl...	glrgg1j	42	1.612319e+09	PLEASE DFV THE PEOPLE NEED A WORD OF ENCOURAGE...
71	<bound method Comment.parent of Comment(id='gl...	glrfosm	1	1.612319e+09	[removed]
72	<bound method Comment.parent of Comment(id='gl...	glrf04e	9	1.612319e+09	AMCGANG □□□□□□□□□□

73 rows × 5 columns

ID 11

In [25]:
```
#Verify comment output of ID 11
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l7feld")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
[deleted]
[deleted]
First majestic (AG) all the way
Neo understands stonks
What is the name of the movie? He he
I only know how to buy, I don't know how to sell
🐰
I love a well placed “fuck you melvin!”
This didn’t age well
I don’t really like these dramas. I prefer the chic flicks. Like that sarah c
onnor terminator meme from last week. I laughed. I cried. I ate my own poop □
□□
U/savevideo
Finally! Thank you! This gif is helping me understand WTF is going on.
This aged beautifully
When this makes more sense than the movie.
Happy cake day
AMC ???
Robinhood won’t approve my account.  I will buy and hold when they do!
```

In [26]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df11 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df11
```

Out[26]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gw...	gwpp942	1	1.620021e+09	[deleted]
1	<bound method Comment.parent of Comment(id='gs...	gsyght8	1	1.617253e+09	[deleted]
2	<bound method Comment.parent of Comment(id='gp...	gpo9lsz	1	1.614907e+09	First majestic (AG) all the way
3	<bound method Comment.parent of Comment(id='go...	gox88a7	1	1.614418e+09	Neo understands stonks
4	<bound method Comment.parent of Comment(id='go...	goq433h	1	1.614302e+09	What is the name of the movie? He he
...	...	...	...	...	...
208	<bound method Comment.parent of Comment(id='gl...	gl94qda	2	1.611966e+09	$STNG Baby
209	<bound method Comment.parent of Comment(id='gl...	gl938ru	1	1.611966e+09	We are NOT selling
210	<bound method Comment.parent of Comment(id='gl...	gl92457	2	1.611965e+09	AMC 25 today let's go
211	<bound method Comment.parent of Comment(id='gl...	gl8zz9a	1	1.611964e+09	+13 more share soldiers 💎🤲
212	<bound method Comment.parent of Comment(id='gl...	gl8waxa	4	1.611963e+09	AMC

213 rows × 5 columns

ID 12

In [27]:
```
#Verify comment output of ID 12
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="lexy8t")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Lmao
APE WAY
This is most definately the way!! ;);)![gif](emote|free_emotes_pack|give_upvo
te)![gif](emote|free_emotes_pack|give_upvote)![gif](emote|free_emotes_pack|gi
ve_upvote)![gif](emote|free_emotes_pack|give_upvote)
Way to get in to the SB that quickly!
I will Propose 💎💍 to cathie Woods if ARKF hits 75 buy July 2021
We need to go after Tesla most shorted stock!!!!
reddit moment
Social media gone too far. Like a star war movie, hedge funds strikes back
Fuck Reddit

Fuck Anti Evil Operations

Fuck China

Fuck Censorship
Redditors united will be the name of the new world government organization
CCCRINGGEEEE
```

In [28]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df12 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df12
```

Out[28]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gw...	gwskjn3	1	1.620085e+09	Lmao
1	<bound method Comment.parent of Comment(id='gw...	gw304gt	1	1.619579e+09	APE WAY
2	<bound method Comment.parent of Comment(id='gw...	gw0q020	1	1.619531e+09	This is most definately the way!! ;);)![gif](e...
3	<bound method Comment.parent of Comment(id='gr...	grrag5i	1	1.616401e+09	Way to get in to the SB that quickly!
4	<bound method Comment.parent of Comment(id='gp...	gppqfdi	1	1.614930e+09	I will Propose 💎💍 to cathie Woods if ARKF hits...
...	...	...	...	...	...
183	<bound method Comment.parent of Comment(id='gm...	gmjsblt	2	1.612788e+09	Ps: it's owned by China. Vote me to the top!
184	<bound method Comment.parent of Comment(id='gm...	gmjs5el	1	1.612787e+09	We... are Reddit.
185	<bound method Comment.parent of Comment(id='gm...	gmjs20i	1	1.612787e+09	If they can rig the Stock Market, what else ar...
186	<bound method Comment.parent of Comment(id='gm...	gmjrvas	1	1.612787e+09	I thought this was important and I was right e...
187	<bound method Comment.parent of Comment(id='gm...	gmjroxw	2	1.612787e+09	Look ma, I'm on TV

188 rows × 5 columns

ID 13

In [29]:
```
#Verify comment output of ID 13
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l6jobf")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Lmao
Thank you
Not so sure where we're going anymore
First majestic (AG) all the way
r/agedlikemilk
They blocking a lot of shit rn
WE LIKE THE MODS
This did not age well...
Thank you!!
Unbelievable resilience
 Mods have given a purpose to a slot of folks
Happy to see everyday folks able to see big gains
Buy General Electric now
Bravo 👏👏👏
□□□□□□
This thread is awesome, far too often the mods work behind the scenes goes un
noticed and unappreciated on subs. The solidarity, gratitude and appreciation
expressed here warms my heart and I too join in this wonderful shout out to t
hose who have likely been put through the gristmill in these past weeks.
```

In [30]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df13 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df13
```

Out[30]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gw...	gwskjzs	1	1.620085e+09	Lmao
1	<bound method Comment.parent of Comment(id='gw...	gwkn16t	1	1.619924e+09	Thank you
2	<bound method Comment.parent of Comment(id='gu...	gurmtjw	1	1.618635e+09	Not so sure where we're going anymore
3	<bound method Comment.parent of Comment(id='gp...	gpo9mrx	1	1.614907e+09	First majestic (AG) all the way
4	<bound method Comment.parent of Comment(id='gp...	gpmoud2	2	1.614871e+09	r/agedlikemilk
...	...	...	...	...	...
95	<bound method Comment.parent of Comment(id='gl...	gl3l54z	2	1.611875e+09	I'm willing to donate a portion of my winnings...
96	<bound method Comment.parent of Comment(id='gl...	gl3ku4d	1	1.611875e+09	Thank you!
97	<bound method Comment.parent of Comment(id='gl...	gl3jq7p	1	1.611875e+09	Thank you mods □□□
98	<bound method Comment.parent of Comment(id='gl...	gl3i1a5	1	1.611875e+09	F (but in a good way)
99	<bound method Comment.parent of Comment(id='gl...	gl3hpwn	1	1.611875e+09	Salute 🚀

100 rows × 5 columns

ID 14

In [31]:
```
#Verify comment output of ID 14
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l890i7")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Lmao
This is still relevant and I love it
First majestic (AG) all the way
I just wish I had DFV superpower to memefy my personal stocks
By now they are probably making money off this..
They are still loosing money
Lol
Omg gggg
💎🙌
Soy SOy Soy 🚀😵nok😵🚀🚀🚀
We Like The Stock 💎🖐
Closer to MOON = Index funds buy more GME.

To the Moon 🚀
Love it, I happened to have bought 90 AMC calls a week or so before all this
happened (I’ve never heard of this subreddit either) because I just really mi
ss going to the movies lol. I was making a bet that if they opened up the sto
ck could recover. Thanks wsb for the preliminary boost! Lol @ the hedge funds
finally getting what they deserved 😂
```

In [32]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df14 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df14
```

Out[32]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gw...	gwskkfn	2	1.620085e+09	Lmao
1	<bound method Comment.parent of Comment(id='gp...	gpztljw	1	1.615085e+09	This is still relevant and I love it
2	<bound method Comment.parent of Comment(id='gp...	gpodw5f	1	1.614909e+09	First majestic (AG) all the way
3	<bound method Comment.parent of Comment(id='go...	gop3zn1	1	1.614284e+09	I just wish I had DFV superpower to memefy my ...
4	<bound method Comment.parent of Comment(id='gn...	gnsviyi	1	1.613621e+09	By now they are probably making money off this..
...	...	...	...	...	...
343	<bound method Comment.parent of Comment(id='gl...	glcf74u	1	1.612018e+09	”We can stay retarded longer than you can rema...
344	<bound method Comment.parent of Comment(id='gl...	glcf4cy	1	1.612018e+09	What’s the chance of gme actually going over $...
345	<bound method Comment.parent of Comment(id='gl...	glcf212	3	1.612018e+09	The malicious short-selling regimes have made ...
346	<bound method Comment.parent of Comment(id='gl...	glcf015	1	1.612018e+09	\n\nOne Snowflake does no damage, ENDLESS SN...
347	<bound method Comment.parent of Comment(id='gl...	glcezgb	1	1.612018e+09	💎□💎□💎□💎□💎□\n\nWe were going to wait for boomer...

348 rows × 5 columns

ID 15

In [33]:
```
#Verify comment output of ID 15
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l79x17")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Remember....A Bugs Life!?
"There's 100,000 Ants for every 1 Cricket! If We All stick together as 1, we
can over run them! We just have to stick Together like a FAMILY"
You tell em. Fk. Even though the senate is specifically excluded from insider
trading laws
First majestic (AG) all the way
[removed]
No longer about profits but how much money we can fuck these cock suckers out
of.
democracy ?
🚀🚀💎🤡
Everyone is loving it!
I like memes and I like this gamestonk
100100100100100100100
I am sure the managers are in a precarious spot. Their high paying jobs on th
e line. They are sitting on a mine field of unpredictability. Only to know an
y stock in their portfolio that has short interest is venerable. So who is ha
ndling the money for PERS, STRS, TSP and the many pensions? I am sure there i
s a silent panic this weekend among them all. Tactical and strategic planning
```

In [34]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df15 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df15
```

Out[34]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gx...	gx42gt7	1	1.620302e+09	Remember....A Bugs Life!? \n"There's 100,000 An...
1	<bound method Comment.parent of Comment(id='gp...	gpv0sz6	1	1.615031e+09	You tell em. Fk. Even though the senate is spe...
2	<bound method Comment.parent of Comment(id='gp...	gpody01	1	1.614909e+09	First majestic (AG) all the way
3	<bound method Comment.parent of Comment(id='gn...	gn006uc	1	1.613116e+09	[removed]
4	<bound method Comment.parent of Comment(id='gm...	gma0umc	1	1.612676e+09	No longer about profits but how much money we ...
...	...	...	...	...	...
189	<bound method Comment.parent of Comment(id='gl...	gl643o9	1	1.611909e+09	If no one paid then, I guarantee you that no o...
190	<bound method Comment.parent of Comment(id='gl...	gl642ia	4	1.611909e+09	I will never forget this. Lost over 300k not s...
191	<bound method Comment.parent of Comment(id='gl...	gl641s1	1	1.611909e+09	CLASS WARFARE BAYBEEE
192	<bound method Comment.parent of Comment(id='gl...	gl640ww	1	1.611909e+09	Is anybody has any doubt that laws exist to ke...
193	<bound method Comment.parent of Comment(id='gl...	gl63tbn	2	1.611909e+09	This generation's finest creation was not a co...

194 rows × 5 columns

ID 16

In [35]:
```
#Verify comment output of ID 16
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l90oq6")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
Rrvergrbgg by wwfcecec
I'm just noticing this now what's your shoe size and why is it 15?
Ya know what?  Gonna do the same thing with Children's Hospital Los Angeles.

brb...shopping.

This is good stuff OP
Help him get into para Olympics!!!
Let's join hands and fund him guys!!
He still needs $11,000 dollars.
Google pay no. Solai raj: 9894890503
Other details all included in this video and in description  of this video.
https://youtu.be/325Nt0FBfJY (https://youtu.be/325Nt0FBfJY)
😂
Best medicine is gaming.
Fuck yeah now you can pay mommy and daddy back the money you owe them for you
r shares lol #wallstreetchildren #itzababydick
Legendary
How much you pocket tho?
```

In [36]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df16 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df16
```

Out[36]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gv...	gv8chfn	1	1.618972e+09	Rrvergrbgg by wwfcecec
1	<bound method Comment.parent of Comment(id='gt...	gt4bjgs	1	1.617377e+09	I'm just noticing this now what's your shoe si...
2	<bound method Comment.parent of Comment(id='gp...	gpcxo0i	1	1.614672e+09	Ya know what? Gonna do the same thing with Ch...
3	<bound method Comment.parent of Comment(id='go...	go7015t	1	1.613910e+09	Help him get into para Olympics!!!\nLet's join...
4	<bound method Comment.parent of Comment(id='go...	go2zyel	1	1.613821e+09	😂
...	...	...	...	...	...
365	<bound method Comment.parent of Comment(id='gl...	glgx861	1	1.612119e+09	Nice work bro\n\nGME HOLD\n\nNot Financial adv...
366	<bound method Comment.parent of Comment(id='gl...	glgx3of	1	1.612119e+09	Billionaires hate this one simple trick.
367	<bound method Comment.parent of Comment(id='gl...	glgwidw	1	1.612119e+09	That's so awesome, we should all take note and...
368	<bound method Comment.parent of Comment(id='gl...	glgwhwu	1	1.612119e+09	Would never see those rich scum fucks do this....
369	<bound method Comment.parent of Comment(id='gl...	glgwcbc	1	1.612118e+09	Legend

370 rows × 5 columns

ID 17

In [37]:
```
#Verify comment output of ID 17
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l74tr1")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
i thought this was interesting:
Nokia wins multi-year 5G radio and core contract with A1 Austria


Press Release


Nokia wins multi-year 5G radio and core contract with A1 Austria
   \-- Will deliver nationwide 5G coverage across Austria; continues Nokia's
long-standing partnership with A1 Austria       -- Commercial implementation
of 4G and 5G network slicing following        successful pilot
18 February 2021
Just buy penny stocks I don't understand
Stupid ass fucking tweet. Melvin got fucked. good get fucked. fuck hedge fund
s. but this isn't "rich people vs poor people". in reality 99.99% of rich peo
ple are not short gamestop and dont give a fuck. most poor people who bought
gamestop have been financially devastated. stop fucking politicizing this
In times of tyranny and injustice when law appresses the people, the outlaw t
akes his place in history
```

In [38]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df17 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df17
```

Out[38]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='go...	gozk4fc	1	1.614458e+09	i thought this was interesting: \nNokia win...
1	<bound method Comment.parent of Comment(id='gn...	gn8h50a	1	1.613261e+09	Just buy penny stocks I don't understand
2	<bound method Comment.parent of Comment(id='gl...	gltetm7	2	1.612350e+09	Stupid ass fucking tweet. Melvin got fucked. g...
3	<bound method Comment.parent of Comment(id='gl...	glh73mi	1	1.612128e+09	In times of tyranny and injustice when law app...
4	<bound method Comment.parent of Comment(id='gl...	glg0u6y	1	1.612094e+09	@StockitesBreakingNews\n\n🚀🚀 Don't worry... We...
...	...	...	...	...	...
156	<bound method Comment.parent of Comment(id='gl...	gl529in	2	1.611894e+09	AH Will Be Rough\n\nAutistic Soldiers, we know...
157	<bound method Comment.parent of Comment(id='gl...	gl526mo	1	1.611894e+09	In the UK you can buy GME on Revolut.
158	<bound method Comment.parent of Comment(id='gl...	gl524r0	2	1.611894e+09	As someone who worries about the next bills a ...
159	<bound method Comment.parent of Comment(id='gl...	gl524fe	1	1.611894e+09	[deleted]
160	<bound method Comment.parent of Comment(id='gl...	gl51xwi	1	1.611894e+09	# TRADEUP\n\n# FIRST TRADE\n\n# CHASE\n\n# IBK...

161 rows × 5 columns

ID 18

In [39]:
```
#Verify comment output of ID 18
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="lazmky")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
What a boomer
Cuban went ballsdeep in Barbara, facts!

Another fact, I'm noticing alot greedy fucktards on here.

Ask for a lil help and the internet trolls pull out their 1 inch pricks to ha
ve a sword fight.  I'm hung like ya grandma, I'm gonna win that fight.


I like the stonk DNAX.  If you do, Get with it or go eat a bluewaffle.  I'm r
iding regardless.
If Mark Cuban ran for president I would 100% vote for him!!!!!
too bad cuban hates america and bows to china
Does Cuban even have position on gme/amc?
Get long and get loud! 
I’m glad he is on our side as well, an ape
Never count your money when you’re sitting at the table...no when you hold em
https://youtu.be/BEOximWoFd0 (https://youtu.be/BEOximWoFd0) 
You guys think he'd say this if it didn't help his self-image? He's not a man
```

In [40]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df18 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df18
```

Out[40]:

	Parent ID	Comment ID	Score	Created	B
0	<bound method Comment.parent of Comment(id='gp...	gp8o07h	1	1.614583e+09	What a boc
1	<bound method Comment.parent of Comment(id='go...	goq6hfc	1	1.614303e+09	Cuban went ballsdeep in Barbara, facts!\n\nA
2	<bound method Comment.parent of Comment(id='go...	gonhm0w	1	1.614245e+09	If Mark Cuban ran for president I would 100%
3	<bound method Comment.parent of Comment(id='gm...	gmvbyod	1	1.613025e+09	too bad cuban hates america and bows to c
4	<bound method Comment.parent of Comment(id='gm...	gma1601	2	1.612676e+09	Does Cuban even have position on gme/a
...	...	...	...	...	
100	<bound method Comment.parent of Comment(id='gl...	gls5nne	1	1.612329e+09	The level of respect I have this man is bana
101	<bound method Comment.parent of Comment(id='gl...	gls5l0s	2	1.612329e+09	True MVP. I am gonna buy a Luka jersey wher
102	<bound method Comment.parent of Comment(id='gl...	gls5akb	-1	1.612329e+09	https://story.snapchat.com/o/W7_EDlXWTBiXAEE
103	<bound method Comment.parent of Comment(id='gl...	gls57zl	1	1.612329e+09	#lettheirbillions
104	<bound method Comment.parent of Comment(id='gl...	gls4fty	1	1.612329e+09	This man is awesome!! So much respect for

105 rows × 5 columns

ID 19

In [41]:
```
#Verify comment output of ID 19
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l7feld")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
[deleted]
[deleted]
First majestic (AG) all the way
Neo understands stonks
What is the name of the movie? He he
I only know how to buy, I don't know how to sell
🤡
I love a well placed “fuck you melvin!”
This didn’t age well
I don’t really like these dramas. I prefer the chic flicks. Like that sarah c
onnor terminator meme from last week. I laughed. I cried. I ate my own poop 
U/savevideo
Finally! Thank you! This gif is helping me understand WTF is going on.
This aged beautifully
When this makes more sense than the movie.
Happy cake day
AMC ???
Robinhood won’t approve my account.  I will buy and hold when they do!
```

In [42]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df19 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df19
```

Out[42]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gw...	gwpp942	1	1.620021e+09	[deleted]
1	<bound method Comment.parent of Comment(id='gs...	gsyght8	1	1.617253e+09	[deleted]
2	<bound method Comment.parent of Comment(id='gp...	gpo9lsz	1	1.614907e+09	First majestic (AG) all the way
3	<bound method Comment.parent of Comment(id='go...	gox88a7	1	1.614418e+09	Neo understands stonks
4	<bound method Comment.parent of Comment(id='go...	goq433h	1	1.614302e+09	What is the name of the movie? He he
...	...	...	...	...	...
208	<bound method Comment.parent of Comment(id='gl...	gl94qda	2	1.611966e+09	$STNG Baby
209	<bound method Comment.parent of Comment(id='gl...	gl938ru	1	1.611966e+09	We are NOT selling
210	<bound method Comment.parent of Comment(id='gl...	gl92457	2	1.611965e+09	AMC 25 today let's go
211	<bound method Comment.parent of Comment(id='gl...	gl8zz9a	1	1.611964e+09	+13 more share soldiers 💎🤲
212	<bound method Comment.parent of Comment(id='gl...	gl8waxa	3	1.611963e+09	AMC

213 rows × 5 columns

ID 20

In [43]:
```
#Verify comment output of ID 20
#Create submission object to read into comments on specific post, passing post ID

submission = reddit.submission(id="l6fg8q")
submission.comment_sort = "new"
top_level_comments = list(submission.comments)


#Create output of contents from submission

submission.comments.replace_more(limit=0)
for top_level_comment in submission.comments:
    print(top_level_comment.body)
```

```
God I hope she's married to one of you.
You realize once you finally get your break you'll be the one she's taxing to
death. And you won't be able to afford it as a small time millionaire unlike
a mega corporate millionaire.
Ew
Tax the rich... □ she has an income that put her in the top %20 of income ear
ners with a pension for life should she ever decide to retire after four year
s plus the best health benefits in the world... some eat her and I don’t mean
her environmentally friendly douched crotch.
Wait if we did it correctly we are now the rich. Damn
The fudge? Based Alexandria Occasional Cortex? Thats a first.
She doesn’t get much right but every now and again she hits the nail on the h
ead
yes! those tweets will teach them!!!
□
You guys are gonna be really disappointed when you realize you're gonna half
to pay taxes on your capital gains.
She gets it.
This aged poorly
```

In [44]:
```
#Append comments to overall comment df

df_rows = [[comment.parent, comment.id, comment.score, comment.created, comment.b
comment_df20 = pd.DataFrame(df_rows, columns=['Parent ID', 'Comment ID', 'Score',
comment_df20
```

Out[44]:

	Parent ID	Comment ID	Score	Created	Body
0	<bound method Comment.parent of Comment(id='gl...	gl0ktzf	1	1.611819e+09	God I hope she's married to one of you.
1	<bound method Comment.parent of Comment(id='gw...	gw5skhz	1	1.619639e+09	You realize once you finally get your break yo...
2	<bound method Comment.parent of Comment(id='gw...	gw4qwyw	1	1.619608e+09	Ew
3	<bound method Comment.parent of Comment(id='gw...	gw4d7uu	1	1.619601e+09	Tax the rich... □ she has an income that put h...
4	<bound method Comment.parent of Comment(id='gv...	gvz669o	1	1.619500e+09	Wait if we did it correctly we are now the ric...
...	...	...	...	...	...
219	<bound method Comment.parent of Comment(id='gl...	gl1vy41	1	1.611839e+09	Hummmm.... I'm getting rich off this. She want...
220	<bound method Comment.parent of Comment(id='gl...	gl1vv9f	2	1.611839e+09	So she is telling me that if I put my recent p...
221	<bound method Comment.parent of Comment(id='gl...	gl1vttn	1	1.611839e+09	Bravoooo
222	<bound method Comment.parent of Comment(id='gl...	gl1vt1i	1	1.611839e+09	The funny thing to me is- assuming all you guy...
223	<bound method Comment.parent of Comment(id='gl...	gl1vpqi	1	1.611839e+09	Should I pack for mars or Pluto?

224 rows × 5 columns

Creating the corpus

In [45]:
```
#Initialize corpus_df and set it to first ID values

corpus_df = comment_df


#Create list of dataframes to append to corpus_df

additional_docs = [comment_df2,comment_df3,comment_df4,comment_df5,comment_df6, c


#Append dataframes

corpus_df = corpus_df.append(additional_docs)
corpus_df = corpus_df[['Comment ID','Body']]
corpus_df
```

Out[45]:

	Comment ID	Body
0	gxhz01m	Get on the rocket before its to late #SafeGala...
1	gxe5jsw	this is second on r/all if u sort by top posts...
2	gxdn0qu	+1 (904) 401-6195
3	gxbsg0e	.
4	gxbsfn2	.
...	...	...
219	gl1vy41	Hummmm.... I'm getting rich off this. She want...
220	gl1vv9f	So she is telling me that if I put my recent p...
221	gl1vttn	Bravoooo
222	gl1vt1i	The funny thing to me is- assuming all you guy...
223	gl1vpqi	Should I pack for mars or Pluto?

4304 rows × 2 columns

In [46]:
```
#Import Beautiful Soup to clean HTML tags

import bs4
from bs4 import BeautifulSoup
```

In [47]:
```
#Use lambda function to strip HTML tags

corpus_df['Body'] = corpus_df['Body'].apply(lambda x: bs4.BeautifulSoup(x, 'lxml'

#Verify corpus_df output

corpus_df['Body']
```

```
D:\Program Files\lib\site-packages\bs4\__init__.py:329: MarkupResemblesLocato
rWarning: "." looks like a filename, not markup. You should probably open thi
s file and pass the filehandle into Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://gofund.me/53244d1d" looks like a URL. Beautiful Soup is no
t an HTTP client. You should probably use an HTTP client like requests to get
the document behind the URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://youtu.be/G2Trn1_KPvA" looks like a URL. Beautiful Soup is
not an HTTP client. You should probably use an HTTP client like requests to g
et the document behind the URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://www.google.com/url?sa=i&url=https%3A%2F%2Ftenor.com%2Fvie
w%2Fgme-gamestop-stocks-stonk-trading-gif-20128040&psig=AOvVaw31yE4BZHNSXt4gh
4ERLQdV&ust=1612200243623000&source=images&cd=vfe&ved=0CAIQjRxqFwoTCLia_o_Yxu
4CFQAAAAAdAAAAABAD" looks like a URL. Beautiful Soup is not an HTTP client. Y
ou should probably use an HTTP client like requests to get the document behin
d the URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://instagram.com/wallstreetbetsredditofficial?igshid=1q041z0b
3kqnq" looks like a URL. Beautiful Soup is not an HTTP client. You should pro
bably use an HTTP client like requests to get the document behind the URL, an
d feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://t.me/trading_tipps_germany" looks like a URL. Beautiful So
up is not an HTTP client. You should probably use an HTTP client like request
s to get the document behind the URL, and feed that document to Beautiful Sou
p.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://www.usatoday.com/story/opinion/2021/01/31/gamestop-drama-r
obinhood-had-play-rules-wall-street-column/4332726001/" looks like a URL. Bea
utiful Soup is not an HTTP client. You should probably use an HTTP client lik
e requests to get the document behind the URL, and feed that document to Beau
tiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://maiar.com/?ref=ad8187f7" looks like a URL. Beautiful Soup
is not an HTTP client. You should probably use an HTTP client like requests t
o get the document behind the URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://youtu.be/ikJfa_HuXkM" looks like a URL. Beautiful Soup is
not an HTTP client. You should probably use an HTTP client like requests to g
```

```
et the document behind the URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://www.thesun.co.uk/wp-content/uploads/2017/09/nintchdbpict00
0310247157.jpg[DDP](https://www.thesun.co.uk/wp-content/uploads/2017/09/nintc
hdbpict000310247157.jpg)" looks like a URL. Beautiful Soup is not an HTTP cli
ent. You should probably use an HTTP client like requests to get the document
behind the URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://open.spotify.com/track/46z25iDIOFSb3G1DOWGiys?si=p91MUt9yS
oWbZvDQVBQPFg" looks like a URL. Beautiful Soup is not an HTTP client. You sh
ould probably use an HTTP client like requests to get the document behind the
URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://youtu.be/whp26NU1X2g" looks like a URL. Beautiful Soup is
not an HTTP client. You should probably use an HTTP client like requests to g
et the document behind the URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://twitter.com/paraszopen/status/1355469500695666688?s=19" lo
oks like a URL. Beautiful Soup is not an HTTP client. You should probably use
an HTTP client like requests to get the document behind the URL, and feed tha
t document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://www.reddit.com/r/CoronavirusMemes/comments/l8e69h/celltrio
ns_regkirona_wins_conditional_nod_to/glc5e6p/?utm_source=share&utm_medium=ios
_app&utm_name=iossmf&context=3" looks like a URL. Beautiful Soup is not an HT
TP client. You should probably use an HTTP client like requests to get the do
cument behind the URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://www.reddit.com/user/Boring_Cranberry_816/comments/l8dg47/s
ave_the_world_from_covid_19_and_save_celltrion/" looks like a URL. Beautiful
Soup is not an HTTP client. You should probably use an HTTP client like reque
sts to get the document behind the URL, and feed that document to Beautiful S
oup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:329: MarkupResemblesLocato
rWarning: "  ." looks like a filename, not markup. You should probably open t
his file and pass the filehandle into Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://www.cnbc.com/video/2015/02/06/santellis-tea-party-rant-feb
ruary-19-2009.html" looks like a URL. Beautiful Soup is not an HTTP client. Y
ou should probably use an HTTP client like requests to get the document behin
d the URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://i.imgur.com/a1UB56F.jpg" looks like a URL. Beautiful Soup
is not an HTTP client. You should probably use an HTTP client like requests t
o get the document behind the URL, and feed that document to Beautiful Soup.
  warnings.warn(
D:\Program Files\lib\site-packages\bs4\__init__.py:414: MarkupResemblesLocato
rWarning: "https://story.snapchat.com/o/W7_EDlXWTBiXAEEniNoMPwAAY-op5chBLlfxL
```

```
AXdkPk_RAXdkPk9BAO1OAA" looks like a URL. Beautiful Soup is not an HTTP clien
t. You should probably use an HTTP client like requests to get the document b
ehind the URL, and feed that document to Beautiful Soup.
  warnings.warn(
```

Out[47]:
```
0      Get on the rocket before its to late #SafeGala...
1      this is second on r/all if u sort by top posts...
2                                    +1 (904) 401-6195
3                                                    .
4                                                    .
                           ...
219    Hummmm.... I'm getting rich off this. She want...
220    So she is telling me that if I put my recent p...
221                                            Bravoooo
222    The funny thing to me is- assuming all you guy...
223                     Should I pack for mars or Pluto?
Name: Body, Length: 4304, dtype: object
```

In [48]:
```
#Place text into specific corpus object for cleaning

cleaned_corpus = corpus_df['Body']
cleaned_corpus
```

Out[48]:
```
0      Get on the rocket before its to late #SafeGala...
1      this is second on r/all if u sort by top posts...
2                                    +1 (904) 401-6195
3                                                    .
4                                                    .
                           ...
219    Hummmm.... I'm getting rich off this. She want...
220    So she is telling me that if I put my recent p...
221                                            Bravoooo
222    The funny thing to me is- assuming all you guy...
223                     Should I pack for mars or Pluto?
Name: Body, Length: 4304, dtype: object
```

In [49]:
```
#define WordPunctTokenizer and stopwords objects

wpt=nltk.WordPunctTokenizer()
stop_words=nltk.corpus.stopwords.words('english')
```

In [50]:
```
#Define normalize_corpus method to apply preprocessing, including regex formattir

def normalize_doc (doc):
    doc=re.sub('[^a-zA-Z\s0-9]+',"", doc, re.I|re.A)
    doc=doc.lower()
    doc=doc.strip()
    tokens=wpt.tokenize(doc)
    filtered_tokens=[token for token in tokens if token not in stop_words]
    doc=' '.join(filtered_tokens)
    return doc
```

In [51]:
```
#Create pipeline to normalize corpus

normalize_corpus=np.vectorize(normalize_doc)
```

In [52]:
```
#Create normalized corpus dataframe and apply normalization pipeline to corpus

normalized_corpus = normalize_corpus(cleaned_corpus)

#Verify corpus output

normalized_corpus
```

Out[52]:
```
array(['get rocket late safegalaxy infinity beyond financial advice research',
       'second rall u sort top posts time', '1 904 4016195', ...,
       'bravoooo',
       'funny thing assuming guys making obscene profits fucking hedge funds ac
tually pay taxes likely alone allow bunch money taxed likely would sheltered st
olen american people',
       'pack mars pluto'], dtype='<U4594')
```

Document Similarity

In [53]:
```
#Import count vectorizer from sklearn

from sklearn.feature_extraction.text import CountVectorizer

#Create cv object to vectorize corpus

cv = CountVectorizer(min_df=0., max_df=1.)

#Create cv_matrix object to create matrix from corpus

cv_matrix=cv.fit_transform(normalized_corpus)
cv_matrix
```

Out[53]:
```
<4304x7183 sparse matrix of type '<class 'numpy.int64'>'
        with 35409 stored elements in Compressed Sparse Row format>
```

In [54]:
```
#Create array representation to verify output

cv_matrix=cv_matrix.toarray()
cv_matrix
```

Out[54]:
```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

In [55]:
```
#Create vocab object to hold important feature names, and create dataframe to dis

vocab=cv.get_feature_names()
pd.DataFrame(cv_matrix,columns=vocab)
```

Out[55]:

	000	0000	000s	00434	005	01	011	020	050	056	...	zillionaires	zim	ziop	zjz	zom
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4299	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4300	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4301	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4302	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4303	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

4304 rows × 7183 columns

In [56]:
```
#Import TfidfTransformer from sklearn

from sklearn.feature_extraction.text import TfidfTransformer
```

In [57]:
```
#Apply TFidf to matrix

tft=TfidfTransformer(norm='l2', use_idf=True)
tft_matrix=tft.fit_transform(cv_matrix)
tft_matrix=tft_matrix.toarray()
vocab=cv.get_feature_names()
pd.DataFrame(np.round(tft_matrix,2), columns=vocab)
```

Out[57]:

	000	0000	000s	00434	005	01	011	020	050	056	...	zillionaires	zim	ziop	zjz	zom
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4299	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
4300	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
4301	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
4302	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
4303	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0

4304 rows × 7183 columns

In [58]:
```
#Import cosine_similarity for document similarity

from sklearn.metrics.pairwise import cosine_similarity
```

In [59]:
```
#Create similarity_matrix object to calculate cosine similarity, passing tf-idf m

similarity_matrix = cosine_similarity(tft_matrix)


#Create dataframe to store similarity matrix data

similarity_df = pd.DataFrame(similarity_matrix)


#Confirm dataframe output

similarity_df
```

Out[59]:

	0	1	2	3	4	5	6	7	8	9	...	4294	4295	4296	4297	4298
0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.000000	0.0	0.000000
1	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.000000	0.0	0.000000
2	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.000000	0.0	0.000000
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.000000	0.0	0.000000
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.000000	0.0	0.000000
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
4299	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.159538	0.0	0.089570
4300	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.106209	0.0	0.059629
4301	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.000000	0.0	0.000000
4302	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.095639	0.0	0.000000	0.0	0.079002
4303	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.000000	0.0	0.000000	0.0	0.000000

4304 rows × 4304 columns

In [60]:
```
#Import dendogram and linkage packages from scipy

from scipy.cluster.hierarchy import dendrogram, linkage
```

In [61]:
```
#Create object to create linkage array

linkage_array = linkage (similarity_matrix, 'ward')
linkage_array
```

Out[61]:
```
array([[3.00000000e+00, 4.00000000e+00, 0.00000000e+00, 2.00000000e+00],
       [5.00000000e+00, 4.30400000e+03, 0.00000000e+00, 3.00000000e+00],
       [6.00000000e+00, 4.30500000e+03, 0.00000000e+00, 4.00000000e+00],
       ...,
       [4.91300000e+03, 8.60300000e+03, 3.70429947e+01, 4.22000000e+03],
       [8.20300000e+03, 8.60400000e+03, 3.85757585e+01, 4.24800000e+03],
       [8.55700000e+03, 8.60500000e+03, 4.16668750e+01, 4.30400000e+03]])
```

In [62]:
```
#Create document cluster dataframe

pd.DataFrame(linkage_array, columns=['Documents\Cluster 1','Documents\Cluster 2',
                                     'Distance', 'Cluster Size'], dtype='object')
```

Out[62]:

	Documents\Cluster 1	Documents\Cluster 2	Distance	Cluster Size
0	3	4	0	2
1	5	4304	0	3
2	6	4305	0	4
3	7	4306	0	5
4	8	4307	0	6
...	...	...	...	...
4298	8600	8601	32.3596	4173
4299	5020	8602	34.0479	4197
4300	4913	8603	37.043	4220
4301	8203	8604	38.5758	4248
4302	8557	8605	41.6669	4304

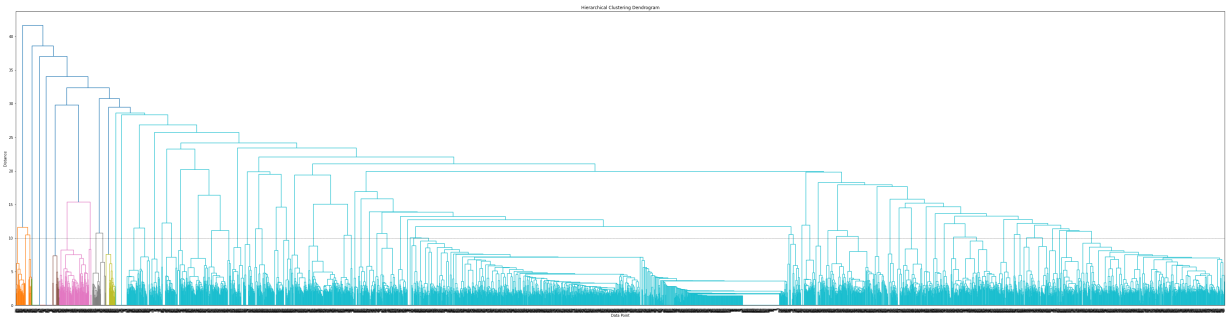
4303 rows × 4 columns

In [93]:
```
#Define figsize, title, and labels for dendogram

plt.figure (figsize = (60,15))
plt.title ('Hierarchical Clustering Dendrogram')
plt.xlabel('Data Point')
plt.ylabel('Distance')

#print dendogram with similarity matrix as the argument

dendrogram(linkage_array)
plt.axhline(y=10.0,c='k',ls='--',lw=0.5);
```



In [ ]:
```
#Our Dendogram shows 5 major clusters, with a majority of clusters falling within
#a majority of the elements of our corpus are highly related and similar, with a

#We can therefore say that much of the data is highly similar, which given the fe
#the time period, makes sense. Thus, the discussion in these posts is centric arc
#or dispersed amongst several other topics.
```

Topic Modelling

In [64]:
```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.decomposition import TruncatedSVD
```

In [65]:
```
#Create an instance of vectorizer and fit corpus data into object X

vectorizer = TfidfVectorizer ()
X = vectorizer.fit_transform (normalized_corpus)
```

In [66]:
```
#Create lsa object to apply SVD to vectorized corpus

lsa = TruncatedSVD (n_components=20, n_iter=750)
lsa.fit(X)
```

Out[66]:
```
TruncatedSVD(n_components=20, n_iter=750)
```

In [67]:
```
#check first row of lsa object to confirm output

row1 = lsa.components_[0]
row1
```

Out[67]:
```
array([0.00036783, 0.00080853, 0.00042262, ..., 0.00060562, 0.00010875,
       0.00012272])
```

In [68]:
```
#Create terms object to browse feature names

terms = vectorizer.get_feature_names()
terms
```

Out[68]:
```
['000',
 '0000',
 '000s',
 '00434',
 '005',
 '01',
 '011',
 '020',
 '050',
 '056',
 '08',
 '09',
 '10',
 '100',
 '1000',
 '10000',
 '100000',
 '100000000',
 '1000s',
```

In [69]:
```
#Use a loop to find concepts and terms associated with them

#enumerate returns index and row tuples in lsa object

for i,comp in enumerate(lsa.components_):
    componentTerms = zip(terms,comp)
    #sort component terms, by concept value,lambda x (x correpsnding tuples) and
    sortedTerms = sorted(componentTerms, key=lambda x:x[1], reverse=True) #decend
    sortedTerms = sortedTerms[:15] #select 15 most important terms within a conce
    print ("\nConcept",i,":")
    for term in sortedTerms:
        print (term)
```

```
Concept 0 :
('hold', 0.43510414253244395)
('buy', 0.37392585749638446)
('amc', 0.33192886040951)
('like', 0.2743055354020431)
('gme', 0.22593329877891755)
('stock', 0.22369621502508572)
('im', 0.16809839963999812)
('shares', 0.11068210763306778)
('still', 0.10299286971625957)
('fucking', 0.09844930013898842)
('way', 0.09634296156502901)
('good', 0.09533472631576663)
('get', 0.09515044388589106)
('dont', 0.09170378906081238)
('one', 0.08675303295363473)

Concept 1 :
```

In [70]:
```
#create an empty dictionary to hold concepts and associated words

concept_words = {}
```

In [71]:
```
#Loop through again to store concepts and associated words in concept_words dicti

for i,comp in enumerate(lsa.components_):
    componentTerms = zip(terms,comp)

    sortedTerms = sorted(componentTerms, key=lambda x:x[1], reverse=True)
    sortedTerms = sortedTerms[:15]
    concept_words["Concept " + str(i)] = sortedTerms
```

In [72]:
```
#View concepts

concept_words
```

Out[72]:
```
{'Concept 0': [('hold', 0.43510414253244395),
  ('buy', 0.37392585749638446),
  ('amc', 0.33192886040951),
  ('like', 0.2743055354020431),
  ('gme', 0.22593329877891755),
  ('stock', 0.22369621502508572),
  ('im', 0.16809839963999812),
  ('shares', 0.11068210763306778),
  ('still', 0.10299286971625957),
  ('fucking', 0.09844930013898842),
  ('way', 0.09634296156502901),
  ('good', 0.09533472631576663),
  ('get', 0.09515044388589106),
  ('dont', 0.09170378906081238),
  ('one', 0.08675303295363473)],
 'Concept 1': [('hold', 0.7406515158743162),
  ('buy', 0.13296339520636685),
  ('amc', 0.09017008965822881),
  ('lol', 0.019630206055666375),
```

In [73]:
```
#Loop through concepts to find if a particular sentence is featured in that conce

for key in concept_words.keys():
    sentence_scores = [] #a list to store all scores for sentences and concepts
    for sentence in normalized_corpus:
        words = nltk.word_tokenize(sentence)
        score = 0
        for word  in words:
            for word_with_score in concept_words[key]:
                if word  == word_with_score[0]:
                    score +=word_with_score[1]
        sentence_scores.append(score)
    print("\n"+key+":")
    for sentence_scores in sentence_scores:
        print (sentence_scores)
```

```
0.13891379791572628
0
0
0
0
0
0.06830162399242724
0.17739866582768124
0
0
0
0
0.07061639387518417
0.19739242143114072
0.11174923644417986
0.20490487197728174
0
0.06945689895786314
0
0
```

Word2Vec

In [74]:
```
#!pip install gensim
#!pip install python-Levenshtein
```

In [75]:
```
#Import packages necessary for Word2Vec

import urllib
from gensim.models import Word2Vec
from nltk.corpus import stopwords
```

```
D:\Program Files\lib\site-packages\gensim\similarities\__init__.py:15: UserWarn
ing: The gensim.similarities.levenshtein submodule is disabled, because the opt
ional Levenshtein package <https://pypi.org/project/python-Levenshtein/> is una
vailable. Install Levenhstein (e.g. `pip install python-Levenshtein`) to suppre
ss this warning.
  warnings.warn(msg)
```

In [76]:
```
#Create a list to hold elements of corpus for tokenization

normalized_corpusList = normalized_corpus.tolist()
```

In [77]:
```
#Verify list output

normalized_corpusList
```

```
 ',
 'f yeah apes together stronk',
 'e',
 'got raw files ad im thinking running uk',
 'bruh one upvoted posts popular posted one month ago',
 'woah wtf almost top post time',
 'cant believe still rising month wanna history even though comment buried',
 'commenting show grandchildren',
 'might get caught bs nuisance lawsuit fantastic legitimate dd provided say p
arty go fund legal fees might need',

 'upvoted post',
 '',
 'legendary',
 'group affiliated wall street silver reddit',
 'almost top reddit post forever know apes',
 'uawardcostbot',
 'literally upvoted post reddit',
 'uawardcostbot',
 'uawardcostbot',
 '400k upvotes fucking kidding',
```

In [78]:
```
#Import tokenizer from nltk

from nltk.tokenize import word_tokenize


#Loop through corpus list and tokenize words

tokenized_sents = [word_tokenize(i) for i in normalized_corpusList]
for i in tokenized_sents:
        print (i)
```

```
['get', 'rocket', 'late', 'safegalaxy', 'infinity', 'beyond', 'financial', 'a
dvice', 'research']
['second', 'rall', 'u', 'sort', 'top', 'posts', 'time']
['1', '904', '4016195']
[]
[]
[]
[]
[]
[]
[]
[]
[]
[]
[]
[]
[]
[]
['maybe', 'smoothbrained', 'question', 'chart', 'historically', 'accurate',
'reason', 'ask', 'second', 'last', 'candlestick', 'looks', 'seriously', 'vola
```

In [79]:
```
#Apply Word2Vec to tokenized sentence corpus

model = Word2Vec(tokenized_sents)
```

In [80]:
```
#Create vocabulary of words stored in "words"

words = model.wv.key_to_index
```

In [89]:
```
#Create object to find words with similar vectors to "robinhood"

robinhoodSimilar = model.wv.most_similar('robinhood', topn = 100)

robinhoodSimilar
```

Out[89]:
```
[('im', 0.999774158000946),
 ('even', 0.9997698664665222),
 ('market', 0.9997601509094238),
 ('short', 0.9997507333755493),
 ('need', 0.9997431635856628),
 ('put', 0.9997417330741882),
 ('day', 0.9997402429580688),
 ('would', 0.9997368454933167),
 ('make', 0.9997299909591675),
 ('like', 0.9997278451919556),
 ('time', 0.9997276663780212),
 ('trading', 0.9997259974479675),
 ('shit', 0.9997239708900452),
 ('money', 0.9997235536575317),
 ('account', 0.9997186064720154),
 ('people', 0.9997185468673706),
 ('one', 0.9997178912162781),
 ('last', 0.9997174143791199),
 ('help', 0.9997166991233826),
 ('gamestop', 0.9997127652168274),
 ('million', 0.9997091293334961),
 ('us', 0.9997031092643738),
 ('think', 0.9997026920318604),
 ('shares', 0.9996996521949768),
 ('get', 0.9996933341026306),
 ('days', 0.9996925592422485),
 ('going', 0.9996894001960754),
 ('right', 0.9996888637542725),
 ('real', 0.9996885657310486),
 ('price', 0.9996873140335083),
 ('take', 0.9996838569641113),
 ('share', 0.9996826648712158),
 ('every', 0.99968022108078),
 ('ive', 0.9996798634529114),
 ('back', 0.9996767044067383),
 ('see', 0.9996760487556458),
 ('well', 0.9996755719184875),
 ('bought', 0.9996718168258667),
 ('really', 0.9996713399887085),
 ('thing', 0.999670684337616),
 ('stocks', 0.9996663331985474),
 ('dont', 0.9996654987335205),
 ('didnt', 0.9996635913848877),
 ('way', 0.9996617436408997),
 ('holding', 0.9996612668037415),
 ('stock', 0.9996547102928162),
 ('many', 0.9996545910835266),
 ('go', 0.9996544122695923),
 ('keep', 0.9996532797813416),
 ('gme', 0.9996518492698669),
 ('everyone', 0.9996516108512878),
```

```
('please', 0.999651312828064),
('could', 0.9996507167816162),
('also', 0.9996488690376282),
('much', 0.9996487498283386),
('company', 0.9996482729911804),
('love', 0.9996480345726013),
('new', 0.9996466636657715),
('first', 0.9996459484100342),
('man', 0.9996457695960999),
('know', 0.999644935131073),
('already', 0.9996439814567566),
('made', 0.9996412992477417),
('got', 0.9996371865272522),
('5', 0.9996360540390015),
('pay', 0.9996340870857239),
('anyone', 0.9996309876441956),
('lost', 0.9996258616447449),
('let', 0.9996243715286255),
('fucking', 0.9996212720870972),
('say', 0.9996193051338196),
('guys', 0.9996192455291748),
('cash', 0.9996181726455688),
('great', 0.9996164441108704),
('doesnt', 0.9996115565299988),
('world', 0.9996110796928406),
('sure', 0.9996103048324585),
('good', 0.9996084570884705),
('fuck', 0.9996018409729004),
('buy', 0.9996010661125183),
('year', 0.9995991587638855),
('give', 0.9995983839035034),
('youre', 0.9995976686477661),
('big', 0.9995976686477661),
('shorts', 0.9995961785316467),
('bank', 0.9995946884155273),
('kids', 0.9995883107185364),
('cant', 0.9995877742767334),
('start', 0.9995861053466797),
('together', 0.9995858669281006),
('best', 0.9995834231376648),
('part', 0.9995757341384888),
('want', 0.9995748996734619),
('buying', 0.9995728731155396),
('someone', 0.9995701909065247),
('never', 0.9995696544647217),
('lol', 0.9995670318603516),
('actually', 0.9995665550231934),
('reddit', 0.9995651841163635),
('financial', 0.9995639324188232)]
```

In [88]:
```
#Create object to find words with similar vectors to "restrict"
restrictSimilar = model.wv.most_similar('restrict', topn = 100)

restrictSimilar
```

Out[88]:
```
[('isnt', 0.9942945837974548),
 ('wrong', 0.9941893219947815),
 ('remember', 0.9939219355583191),
 ('profit', 0.9938562512397766),
 ('power', 0.9938443303108215),
 ('think', 0.993805468082428),
 ('making', 0.9938042759895325),
 ('something', 0.9937946200370789),
 ('years', 0.9937798380851746),
 ('keep', 0.9937587380409241),
 ('movement', 0.9937451481819153),
 ('kind', 0.9937443733215332),
 ('little', 0.9937378168106079),
 ('open', 0.9937349557876587),
 ('trying', 0.9937239289283752),
 ('year', 0.9937204718589783),
 ('new', 0.9937132596969604),
 ('still', 0.9937111735343933),
 ('end', 0.9937000870704651),
 ('holding', 0.9936881065368652),
 ('market', 0.9936777353286743),
 ('give', 0.9936673641204834),
 ('help', 0.9936638474464417),
 ('get', 0.9936620593070984),
 ('right', 0.993661642074585),
 ('enough', 0.9936577081680298),
 ('day', 0.9936543107032776),
 ('invest', 0.993649959564209),
 ('law', 0.9936480522155762),
 ('broker', 0.9936478137969971),
 ('company', 0.9936447739601135),
 ('buying', 0.9936394691467285),
 ('made', 0.9936311841011047),
 ('long', 0.993629515171051),
 ('game', 0.9936179518699646),
 ('next', 0.9936164617538452),
 ('us', 0.9936093091964722),
 ('take', 0.9936007261276245),
 ('money', 0.993600070476532),
 ('pay', 0.9935915470123291),
 ('second', 0.9935882687568665),
 ('dont', 0.9935842752456665),
 ('know', 0.9935794472694397),
 ('shes', 0.9935791492462158),
 ('go', 0.9935786128044128),
 ('didnt', 0.9935764670372009),
 ('way', 0.9935649037361145),
 ('anyone', 0.9935609698295593),
 ('lol', 0.9935529828071594),
 ('back', 0.9935526847839355),
 ('app', 0.9935522079467773),
 ('cuban', 0.9935508370399475),
```

```
('shares', 0.9935483932495117),
('position', 0.9935482144355774),
('need', 0.9935376048088074),
('seen', 0.9935367703437805),
('instead', 0.9935356974601746),
('trading', 0.9935287833213806),
('short', 0.9935256838798523),
('please', 0.9935216903686523),
('positions', 0.9935207962989807),
('hedge', 0.9935142397880554),
('nothing', 0.9935044646263123),
('thing', 0.9934982657432556),
('fucking', 0.9934961795806885),
('price', 0.99349445104599),
('wish', 0.9934866428375244),
('better', 0.9934835433959961),
('could', 0.9934802651405334),
('line', 0.9934787154197693),
('account', 0.9934744238853455),
('afford', 0.9934739470481873),
('im', 0.9934723973274231),
('doesnt', 0.9934718608856201),
('shit', 0.9934707283973694),
('good', 0.9934680461883545),
('stuff', 0.9934651255607605),
('buy', 0.9934577345848083),
('people', 0.9934577345848083),
('stock', 0.9934549927711487),
('like', 0.9934542775154114),
('point', 0.9934541583061218),
('selling', 0.9934488534927368),
('someone', 0.993445634841919),
('able', 0.9934412240982056),
('funds', 0.9934352040290833),
('stop', 0.9934300780296326),
('investment', 0.9934217929840088),
('gamestop', 0.9934115409851074),
('put', 0.993411123752594),
('getting', 0.9934089183807373),
('id', 0.9934085011482239),
('crack', 0.9934056401252747),
('gme', 0.9934036135673523),
('never', 0.9934007525444031),
('trade', 0.9933971166610718),
('ive', 0.9933950901031494),
('lets', 0.9933922290802002),
('would', 0.9933835864067078),
('false', 0.9933809041976929)]
```

In [87]:
```
#Create object to find words with similar vectors to "legal"
legalSimilar = model.wv.most_similar('legal', topn = 100)

legalSimilar
```

Out[87]:
```
[('power', 0.9969563484191895),
 ('things', 0.9968568682670593),
 ('even', 0.9968422651290894),
 ('shes', 0.9968288540840149),
 ('stocks', 0.9968149065971375),
 ('already', 0.9968130588531494),
 ('one', 0.9968092441558838),
 ('said', 0.9968035817146301),
 ('want', 0.9967972636222839),
 ('people', 0.9967834949493408),
 ('shorts', 0.9967797994613647),
 ('know', 0.9967741966247559),
 ('new', 0.9967671036720276),
 ('time', 0.9967551231384277),
 ('whats', 0.996739387512207),
 ('days', 0.9967294931411743),
 ('im', 0.9967196583747864),
 ('positions', 0.9967164993286133),
 ('made', 0.996714174747467),
 ('never', 0.996709942817688),
 ('need', 0.9967049360275269),
 ('attack', 0.9967031478881836),
 ('wish', 0.9966993927955627),
 ('cover', 0.996698796749115),
 ('funds', 0.9966952204704285),
 ('really', 0.9966941475868225),
 ('got', 0.9966934323310852),
 ('news', 0.9966679811477661),
 ('today', 0.9966628551483154),
 ('year', 0.9966601133346558),
 ('bought', 0.996660053730011),
 ('every', 0.9966553449630737),
 ('share', 0.9966536164283752),
 ('someone', 0.9966502785682678),
 ('hands', 0.9966408014297485),
 ('start', 0.9966381192207336),
 ('well', 0.9966297745704651),
 ('help', 0.9966220259666443),
 ('dont', 0.9966135025024414),
 ('robinhood', 0.996613085269928),
 ('2', 0.9966121315956116),
 ('gamestop', 0.9966118931770325),
 ('retail', 0.9966115355491638),
 ('reddit', 0.9966113567352295),
 ('account', 0.9966071248054504),
 ('able', 0.9966051578521729),
 ('guys', 0.9966012239456177),
 ('ever', 0.9965975880622864),
 ('isnt', 0.9965945482254028),
 ('could', 0.996589183807373),
 ('worth', 0.9965800046920776),
 ('street', 0.9965795874595642),
```

```
('something', 0.9965769052505493),
('small', 0.9965732097625732),
('seen', 0.9965713620185852),
('still', 0.9965705871582031),
('lol', 0.9965686202049255),
('market', 0.9965678453445435),
('get', 0.9965614676475525),
('bad', 0.9965588450431824),
('go', 0.9965581893920898),
('didnt', 0.9965572357177734),
('trade', 0.9965550899505615),
('shares', 0.9965509176254272),
('going', 0.9965500235557556),
('everyone', 0.9965486526489258),
('work', 0.9965469837188721),
('also', 0.9965466856956482),
('company', 0.99654221534729),
('doesnt', 0.996541440486908),
('money', 0.9965406656265259),
('looking', 0.9965400695800781),
('nothing', 0.9965350031852722),
('keep', 0.996532678604126),
('wsb', 0.9965313076972961),
('gme', 0.9965303540229797),
('app', 0.9965251684188843),
('place', 0.9965212941169739),
('many', 0.9965208172798157),
('together', 0.9965149164199829),
('na', 0.9965087175369263),
('hope', 0.9965060949325562),
('short', 0.9965059161186218),
('amc', 0.9965007305145264),
('right', 0.9964993596076965),
('moon', 0.996496319770813),
('next', 0.9964946508407593),
('like', 0.9964918494224548),
('price', 0.9964912533760071),
('man', 0.9964908957481384),
('end', 0.9964827299118042),
('buy', 0.9964750409126282),
('bank', 0.9964748620986938),
('anyone', 0.9964739084243774),
('sub', 0.9964730143547058),
('much', 0.9964721202850342),
('thats', 0.9964708685874939),
('companies', 0.9964699149131775),
('day', 0.9964666962623596),
('first', 0.9964637160301208)]
```

In [91]:
```
#Create object to find words with similar vectors to "attack"
attackSimilar = model.wv.most_similar('attack', topn = 100)

attackSimilar
```

Out[91]:
```
[('company', 0.9993638396263123),
 ('got', 0.9993467926979065),
 ('stocks', 0.9993463158607483),
 ('right', 0.9993352890014648),
 ('im', 0.9993324279785156),
 ('know', 0.9993231296539307),
 ('one', 0.9993216395378113),
 ('help', 0.9993123412132263),
 ('gamestop', 0.9993104934692383),
 ('money', 0.9993051886558533),
 ('time', 0.9992981553077698),
 ('let', 0.9992930889129639),
 ('think', 0.9992923736572266),
 ('ill', 0.9992877244949341),
 ('last', 0.9992861747741699),
 ('bought', 0.999272882938385),
 ('keep', 0.9992691874504089),
 ('already', 0.9992679953575134),
 ('would', 0.9992664456367493),
 ('much', 0.9992652535438538),
 ('could', 0.999261736869812),
 ('stock', 0.9992591738700867),
 ('fucking', 0.9992578625679016),
 ('next', 0.9992573261260986),
 ('need', 0.9992517232894897),
 ('market', 0.9992426037788391),
 ('trading', 0.99924236536026),
 ('also', 0.9992420673370361),
 ('back', 0.9992367029190063),
 ('even', 0.9992331266403198),
 ('made', 0.9992319345474243),
 ('everyone', 0.999231219291687),
 ('amc', 0.9992309212684631),
 ('us', 0.9992297887802124),
 ('short', 0.9992291331291199),
 ('get', 0.9992285966873169),
 ('gme', 0.9992268681526184),
 ('someone', 0.9992256760597229),
 ('days', 0.9992212057113647),
 ('shares', 0.9992192387580872),
 ('really', 0.9992164373397827),
 ('thats', 0.9992144703865051),
 ('million', 0.9992142915725708),
 ('thing', 0.9992142915725708),
 ('robinhood', 0.999214231967926),
 ('day', 0.9992092847824097),
 ('well', 0.9992036819458008),
 ('share', 0.999203085899353),
 ('good', 0.9992011189460754),
 ('account', 0.9992005825042725),
 ('said', 0.999199628829956),
 ('guys', 0.9991977214813232),
```

```
('shes', 0.9991965293884277),
('man', 0.9991956353187561),
('street', 0.999194860458374),
('holding', 0.9991921186447144),
('shorts', 0.9991897940635681),
('going', 0.9991896152496338),
('buy', 0.9991878867149353),
('reddit', 0.9991876482963562),
('every', 0.9991862773895264),
('way', 0.9991855025291443),
('like', 0.9991852641105652),
('people', 0.9991846680641174),
('take', 0.9991830587387085),
('still', 0.9991827607154846),
('see', 0.9991762042045593),
('make', 0.9991731643676758),
('world', 0.9991731643676758),
('want', 0.9991668462753296),
('lot', 0.9991661906242371),
('didnt', 0.9991620779037476),
('first', 0.9991598725318909),
('new', 0.9991582632064819),
('price', 0.9991515874862671),
('doesnt', 0.9991462826728821),
('real', 0.9991384148597717),
('sure', 0.9991325736045837),
('shit', 0.9991307258605957),
('put', 0.9991299510002136),
('together', 0.9991297721862793),
('year', 0.9991289377212524),
('rich', 0.9991244077682495),
('high', 0.9991240501403809),
('line', 0.9991229176521301),
('dont', 0.9991228580474854),
('lets', 0.9991223216056824),
('lost', 0.9991205334663391),
('youre', 0.9991175532341003),
('love', 0.9991168975830078),
('years', 0.999113917350769),
('something', 0.9991097450256348),
('silver', 0.9991087913513184),
('part', 0.9991070628166199),
('end', 0.9990999698638916),
('cover', 0.9990963935852051),
('fuck', 0.9990904331207275),
('go', 0.9990898370742798),
('video', 0.9990895986557007),
('actually', 0.9990891814231873)]
```

In [92]:
```
#Create object to find words with similar vectors to "attack"
brokerSimilar = model.wv.most_similar('broker', topn = 100)

brokerSimilar
```

Out[92]:
```
[('please', 0.9992424249649048),
 ('need', 0.9992393851280212),
 ('get', 0.9992284774780273),
 ('robinhood', 0.9992026686668396),
 ('put', 0.9991934895515442),
 ('made', 0.9991903305053711),
 ('im', 0.9991863369941711),
 ('gme', 0.999182939529419),
 ('gamestop', 0.999174177646637),
 ('shit', 0.9991686940193176),
 ('even', 0.9991675615310669),
 ('people', 0.9991664886474609),
 ('bought', 0.9991611242294312),
 ('fucking', 0.9991608262062073),
 ('go', 0.9991589188575745),
 ('shares', 0.9991548657417297),
 ('shorts', 0.9991539120674133),
 ('first', 0.9991514086723328),
 ('anyone', 0.9991469383239746),
 ('buy', 0.9991467595100403),
 ('dont', 0.9991413354873657),
 ('short', 0.9991379380226135),
 ('buying', 0.9991365671157837),
 ('trading', 0.9991353154182434),
 ('got', 0.9991347789764404),
 ('stock', 0.9991321563720703),
 ('market', 0.9991303086280823),
 ('one', 0.9991294145584106),
 ('didnt', 0.9991270899772644),
 ('really', 0.9991231560707092),
 ('help', 0.9991217851638794),
 ('million', 0.9991216659545898),
 ('amc', 0.9991205334663391),
 ('money', 0.9991123676300049),
 ('trade', 0.9991099238395691),
 ('make', 0.9991084933280945),
 ('think', 0.9991048574447632),
 ('said', 0.9991021752357483),
 ('many', 0.9990992546081543),
 ('like', 0.9990977644920349),
 ('take', 0.9990925192832947),
 ('man', 0.9990900754928589),
 ('also', 0.9990881085395813),
 ('start', 0.9990876913070679),
 ('company', 0.9990866780281067),
 ('stocks', 0.9990846514701843),
 ('pay', 0.9990842938423157),
 ('day', 0.9990836381912231),
 ('could', 0.9990829825401306),
 ('cash', 0.9990795850753784),
 ('able', 0.9990777969360352),
 ('new', 0.9990739226341248),
```

```
('holding', 0.9990724921226501),
('way', 0.999072253704071),
('well', 0.9990696907043457),
('days', 0.9990677833557129),
('right', 0.9990642070770264),
('account', 0.999060869216919),
('still', 0.9990608096122742),
('going', 0.9990602135658264),
('know', 0.9990599155426025),
('selling', 0.999059796333313),
('never', 0.999057412147522),
('actually', 0.9990571141242981),
('already', 0.9990512132644653),
('end', 0.9990440011024475),
('share', 0.9990437626838684),
('bank', 0.9990378618240356),
('everyone', 0.9990373849868774),
('always', 0.9990372061729431),
('back', 0.9990362524986267),
('ive', 0.9990347623825073),
('us', 0.999027669429779),
('good', 0.9990262389183044),
('world', 0.999025285243988),
('lost', 0.9990221261978149),
('position', 0.9990200400352478),
('5', 0.9990196228027344),
('someone', 0.9990190267562866),
('much', 0.9990183115005493),
('real', 0.9990168213844299),
('every', 0.9990135431289673),
('financial', 0.9990109205245972),
('together', 0.9990051984786987),
('shes', 0.9990043640136719),
('something', 0.9990009665489197),
('last', 0.998999834060669),
('price', 0.9989989399909973),
('would', 0.9989964365959167),
('guys', 0.9989924430847168),
('sold', 0.9989914894104004),
('cant', 0.9989852905273438),
('love', 0.9989851713180542),
('wish', 0.9989827275276184),
('power', 0.9989814758300781),
('lot', 0.9989813566207886),
('video', 0.9989781975746155),
('say', 0.998977780342102),
('open', 0.9989755749702454),
('better', 0.9989708065986633)]
```

In [94]:
```
#Create object to find words with similar vectors to "manipulation"
manipulationSimilar = model.wv.most_similar('manipulation', topn = 100)

manipulationSimilar
```

Out[94]:
```
[('cash', 0.9986650347709656),
 ('robinhood', 0.9985054731369019),
 ('real', 0.998491644859314),
 ('gamestop', 0.9984837770462036),
 ('fucking', 0.9984831809997559),
 ('thing', 0.9984828233718872),
 ('right', 0.9984692335128784),
 ('market', 0.9984663724899292),
 ('account', 0.99846351146698),
 ('start', 0.9984602928161621),
 ('good', 0.9984580278396606),
 ('stock', 0.9984495639801025),
 ('great', 0.9984457492828369),
 ('history', 0.9984398484230042),
 ('im', 0.9984334111213684),
 ('help', 0.9984323978424072),
 ('last', 0.9984303712844849),
 ('take', 0.9984294176101685),
 ('gme', 0.9984269142150879),
 ('shit', 0.9984261989593506),
 ('like', 0.9984150528907776),
 ('new', 0.9984103441238403),
 ('app', 0.9984081983566284),
 ('5', 0.9984078407287598),
 ('trading', 0.9984069466590881),
 ('financial', 0.9984048008918762),
 ('keep', 0.9983968734741211),
 ('see', 0.9983898401260376),
 ('one', 0.998389720916748),
 ('shorts', 0.9983887672424316),
 ('anyone', 0.9983779191970825),
 ('know', 0.9983749389648438),
 ('didnt', 0.9983742237091064),
 ('rich', 0.9983737468719482),
 ('shares', 0.9983727931976318),
 ('u', 0.9983662962913513),
 ('sure', 0.998363733291626),
 ('short', 0.99836266040802),
 ('actually', 0.9983618259429932),
 ('much', 0.9983608722686768),
 ('even', 0.9983562231063843),
 ('think', 0.9983521103858948),
 ('money', 0.9983518719673157),
 ('got', 0.9983515739440918),
 ('holding', 0.9983511567115784),
 ('would', 0.9983507990837097),
 ('youre', 0.998350203037262),
 ('world', 0.9983491897583008),
 ('time', 0.9983468055725098),
 ('made', 0.9983463883399963),
 ('dont', 0.9983342885971069),
 ('buying', 0.9983338713645935),
```

```
('whole', 0.9983330368995667),
('man', 0.9983291029930115),
('cant', 0.9983271956443787),
('say', 0.9983218908309937),
('days', 0.9983199834823608),
('said', 0.998314619064331),
('best', 0.9983137249946594),
('anything', 0.9983134865760803),
('people', 0.9983099102973938),
('end', 0.9983087778091431),
('going', 0.9983076453208923),
('whats', 0.9983057975769043),
('go', 0.9983043074607849),
('please', 0.9982939958572388),
('order', 0.9982932806015015),
('never', 0.9982921481132507),
('buy', 0.9982906579971313),
('want', 0.9982902407646179),
('million', 0.9982867240905762),
('us', 0.998285710811615),
('way', 0.9982792735099792),
('lost', 0.9982783794403076),
('enough', 0.9982749819755554),
('every', 0.9982662796974182),
('robin', 0.9982653260231018),
('day', 0.9982649683952332),
('many', 0.998260498046875),
('let', 0.9982594847679138),
('believe', 0.9982593655586243),
('na', 0.9982550144195557),
('open', 0.998253583908081),
('everyone', 0.9982489943504333),
('get', 0.9982439875602722),
('bank', 0.9982417821884155),
('guys', 0.9982408285140991),
('company', 0.9982405304908752),
('getting', 0.9982404708862305),
('could', 0.9982398152351379),
('friday', 0.998236894607544),
('celltrion', 0.9982351064682007),
('join', 0.9982338547706604),
('since', 0.9982278347015381),
('hands', 0.998223602771759),
('cover', 0.9982174634933472),
('really', 0.9982168674468994),
('doesnt', 0.9982151985168457),
('bought', 0.9982139468193054),
('reddit', 0.9982070326805115)]
```

In [ ]: