wallstreetbets - investing

By: Sean Frey

Overview

- Problem Statement

- Data Collection

- EDA

- Data Visualization

- Modeling

- Results & Conclusion

Problem

I was tasked with creating a categorizer to label unmarked posts as either r/wallstreetbets or r/investing. This can be useful for both the reddit moderators and/or users trying to find out where information is coming from.

r/investing · Posted by u/_____ 2 hours ago

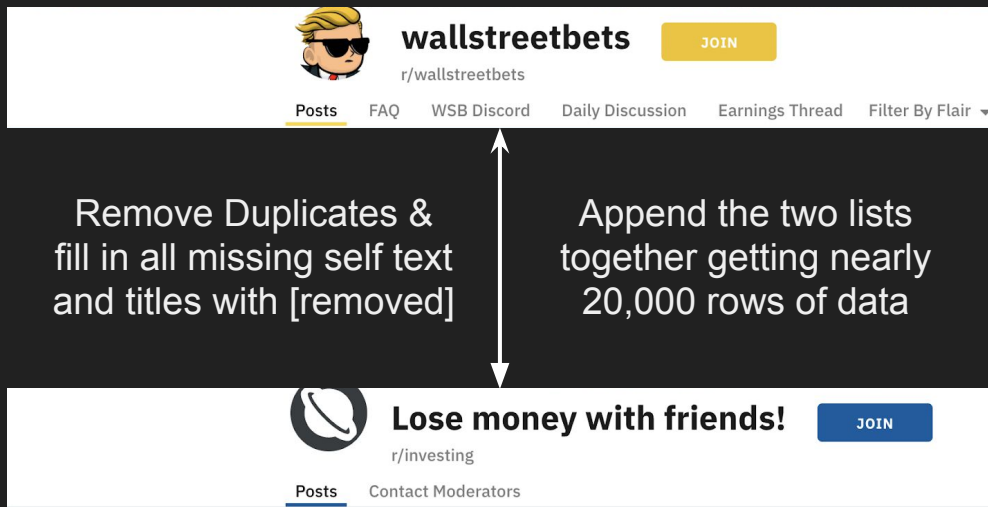### Compound Interest in Europe?

I am fairly new to investing and have been reading a lot around compound interest. As an European investor what are the best ways to invest my money so I can use the advantages of compounding?

And, is there any way that investing in stocks can return compounding?

r/wallstreetbets · Posted by u/_____ 1 hour ago

The real reason for market pullback. It's not the flu.

Technicals

https://www.reddit.com/r/wallstreetbets/comments/ekmd7o/how_to_predict_the_next_market_pullback_its_not/

Earlier this year u/Kroniic told us how to predict market pullbacks. It's not the flu. It's Lil Yachty's drops.

January 23rd, Wiz Khalifa, **Lil Yachty**, Sueco the Child, and Ty Dolla Sign - Speed Me Up

January 26th, **Lil Yachty** & Bryson Tiller - For Nothing

The day after both releases SPY fell. Prepare for the recession when he drops his new album.

Data Collection

Pushshift api

- type: submission

- size: 500

- before: date of the last submission pulled

wallstreetbets
r/wallstreetbets
JOIN

Posts   FAQ   WSB Discord   Daily Discussion   Earnings Thread   Filter By Flair ▼

Remove Duplicates &
fill in all missing self text
and titles with [removed]

Append the two lists
together getting nearly
20,000 rows of data

Lose money with friends!
r/investing
JOIN

Posts   Contact Moderators

EDA Let's look at the web pages themselves!

Posted by u/[redacted] 29 minutes ago

$SPCE is going to the moon 16 times this year! `Discussion`

"This year may see a major milestone for the Spaceport: the first commercial launch for Virgin Galactic, the facility's anchor tenant that reportedly pays the state government about $5 million per year in rent and usage fees. Previously filed documents with the U.S. Securities and Exchange Commission say the company is planning for its first launch to take place in June.

Virgin Galactic estimates 16 commercial flights will take place this year, 115 next year and 270 by 2023, according to an analyst presentation filed with the SEC."

Someone filed it with the SEC so it must be true! Easy 2/21 calls.

https://www.bizjournals.com/albuquerque/news/2020/01/30/two-new-companies-headed-to-spaceport-america.html?ana=yahoo&yptr=yahoo

💬 5 Comments   ⭐ Give Award   ➤ Share   🔖 Save   •••

EDA Let's look at the web pages themselves!

Online Valuation Calculators

What are the best online intrinsic valuation calculators? I've come across DiscoverCI which seems to offer much more than just simply an input calculator, are sites such as these worth the cost? Are there better ones out there that I should look at before deciding?

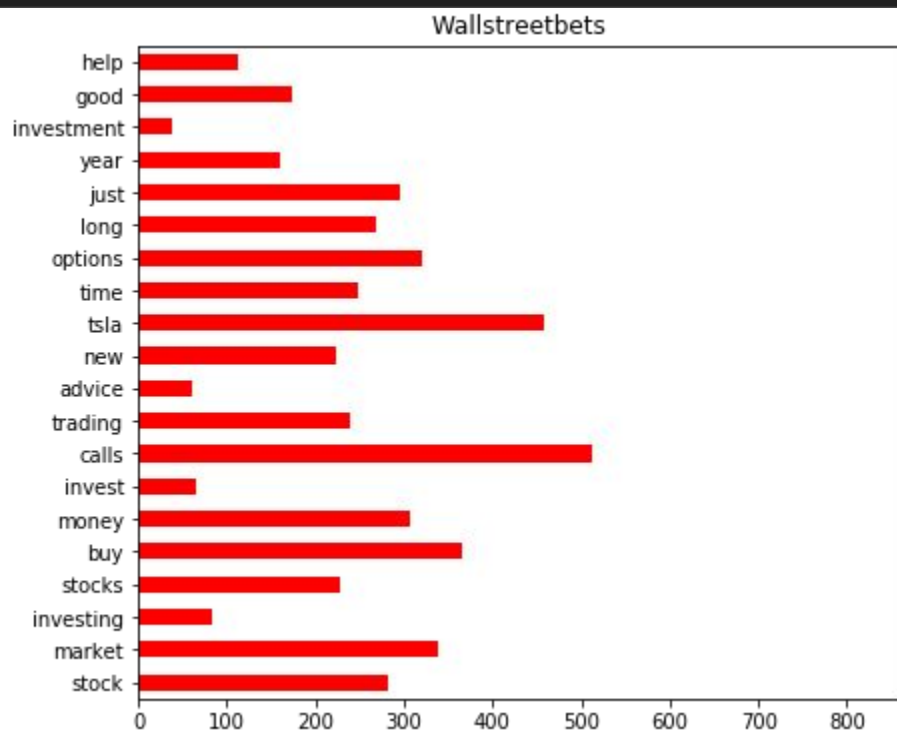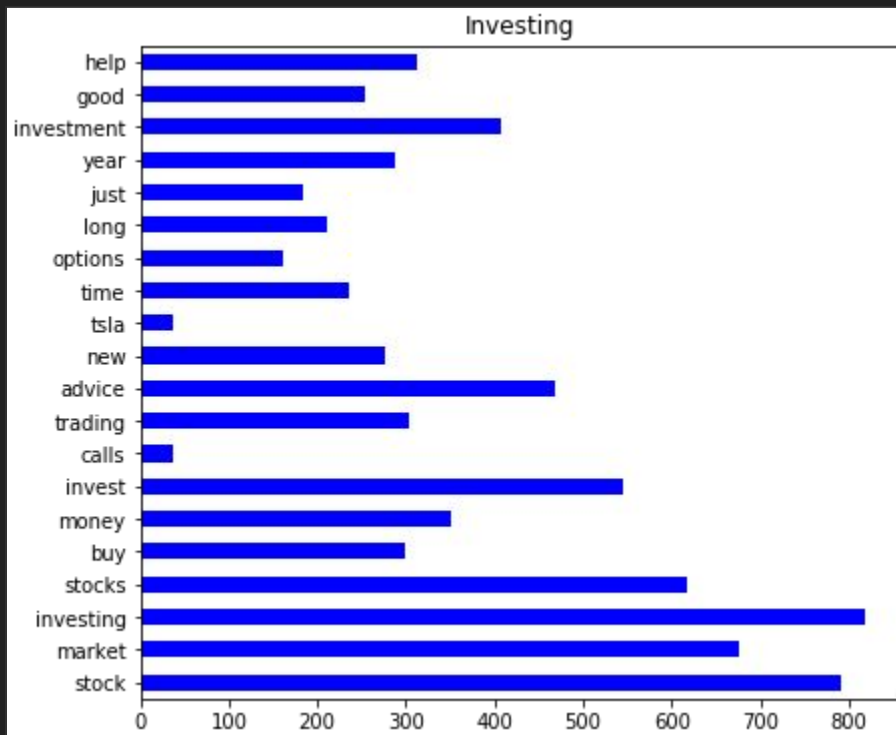Also, DiscoverCI seems to only work for American trades. Is there a Canadian equivalent out there"?
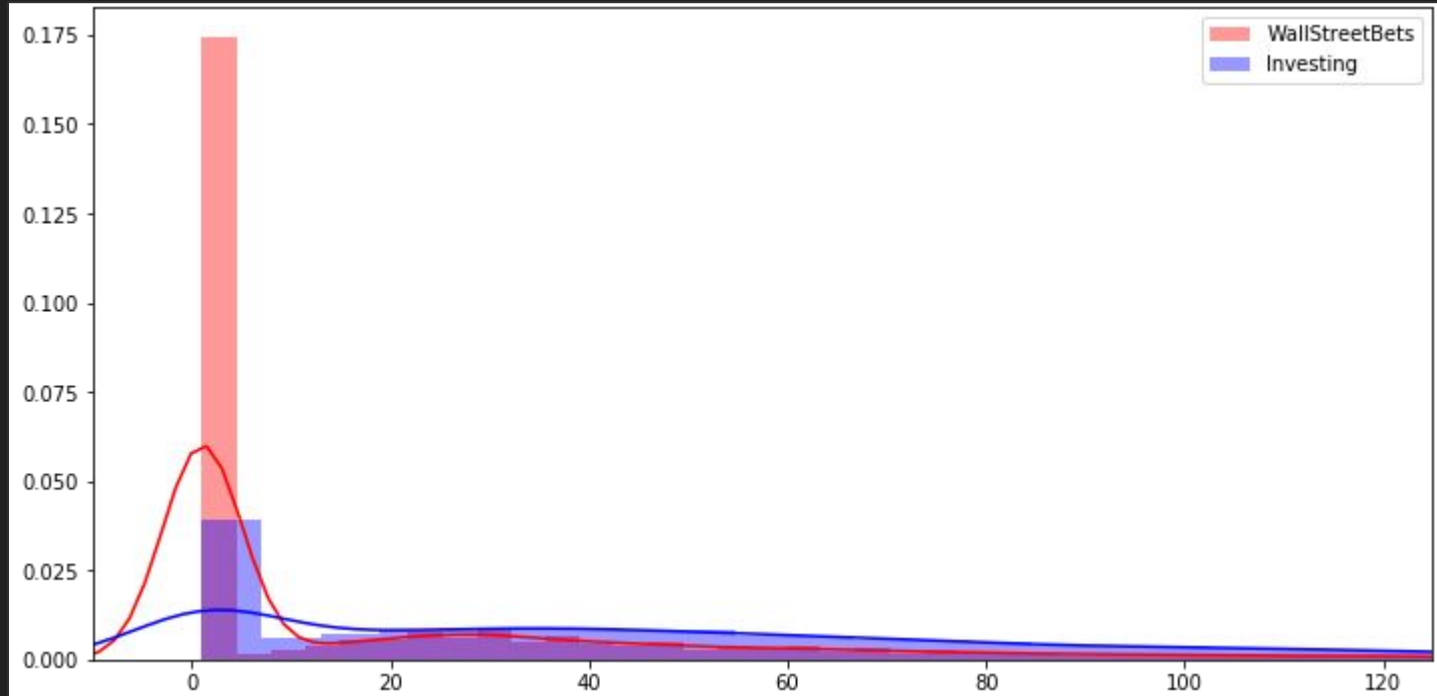
Data Visualization - Self Text
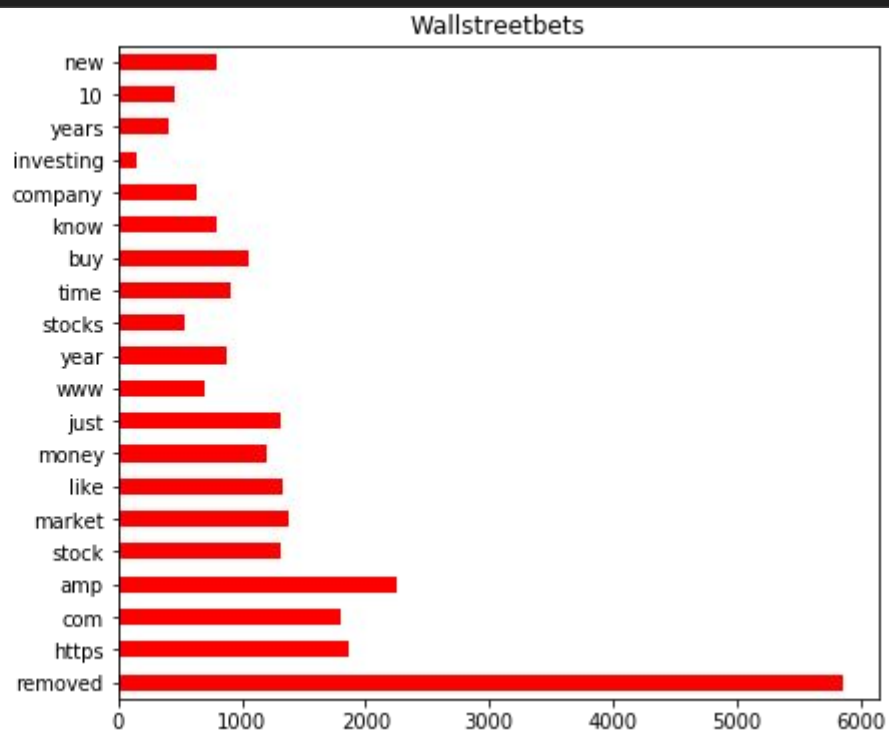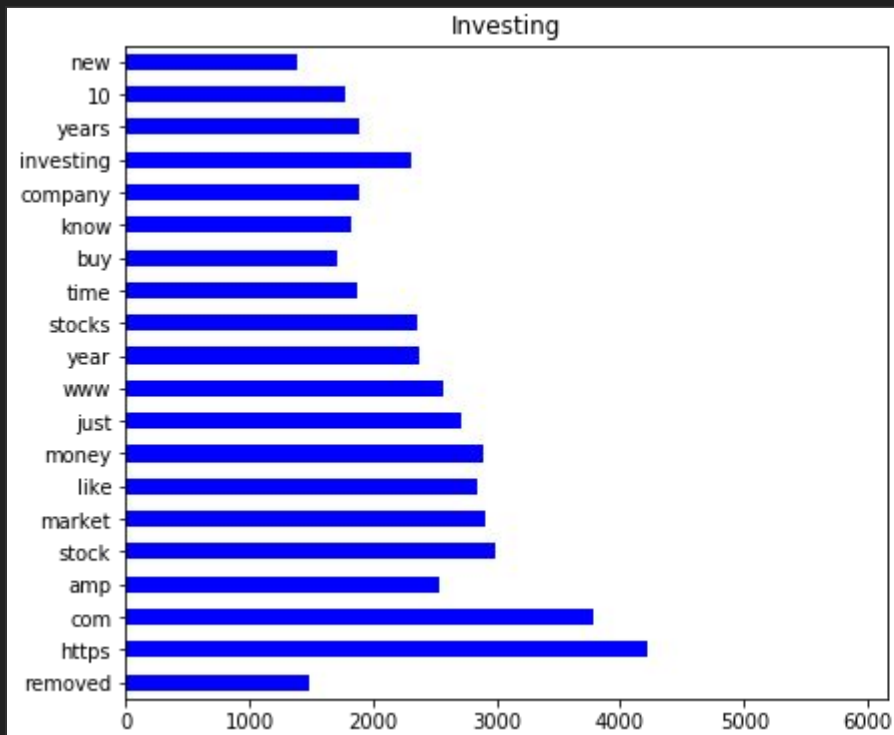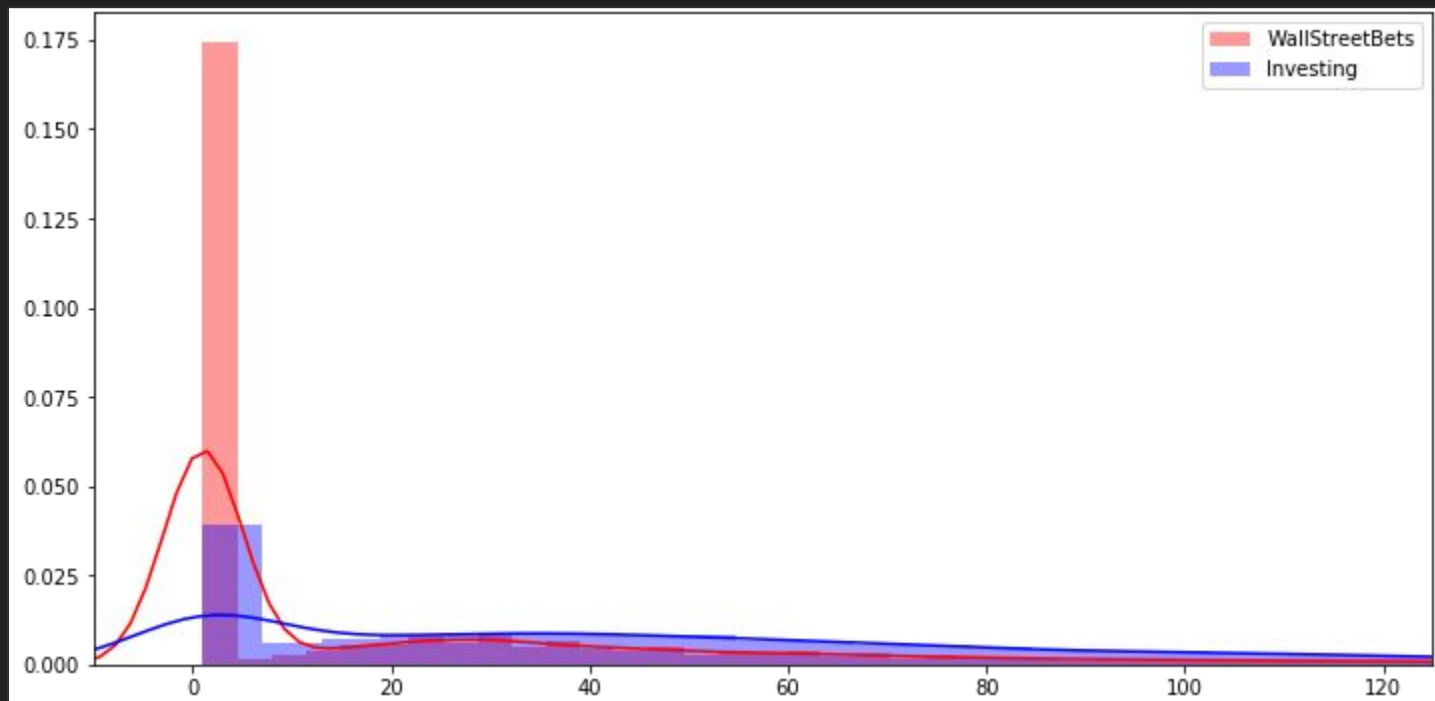
Data Visualization - Self Text
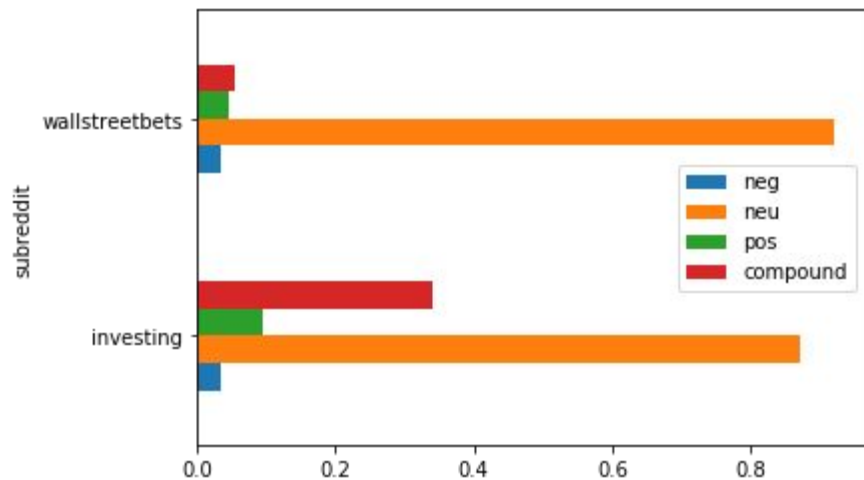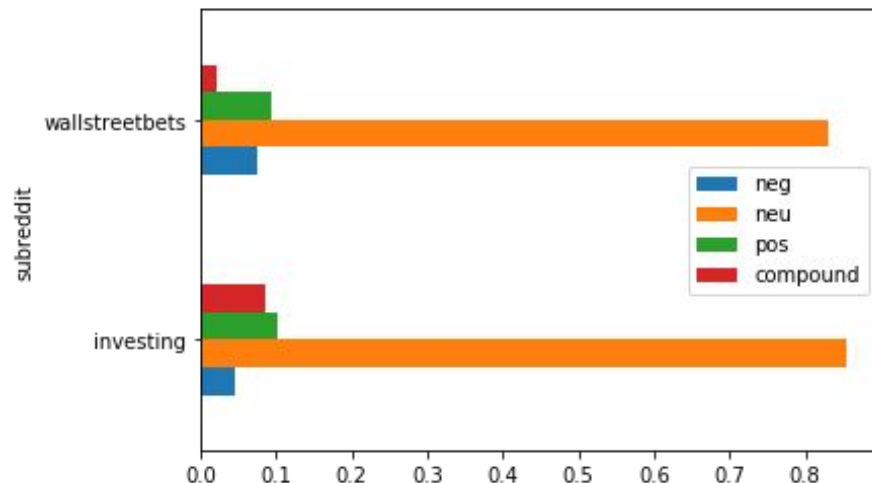
Data Visualization - Title

Data Visualization - Title

Data Visualization - Sentiment
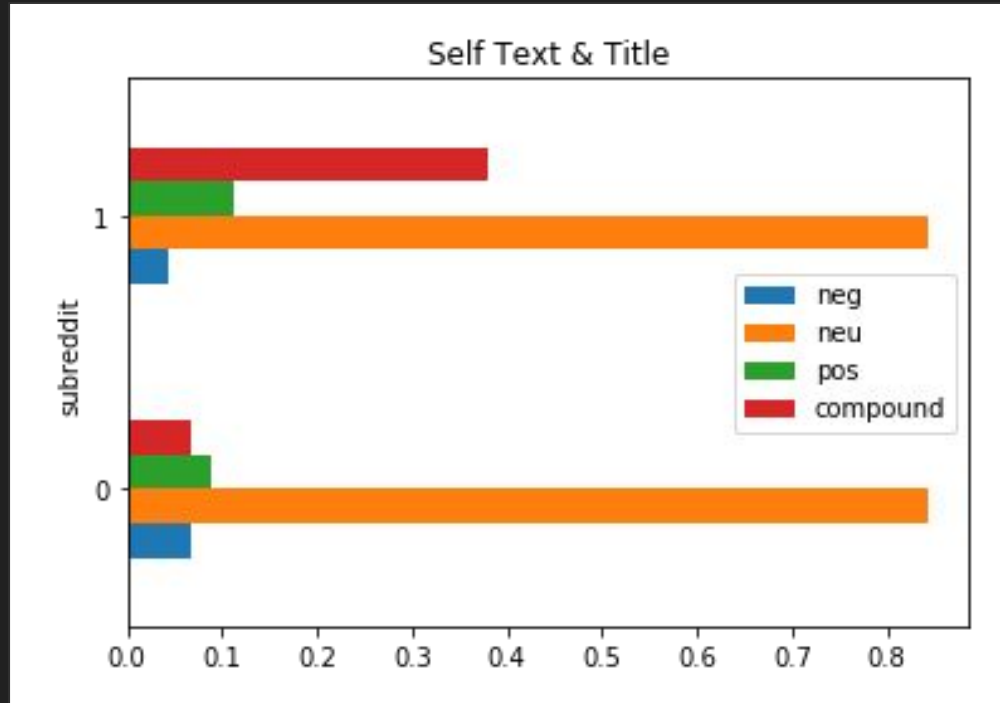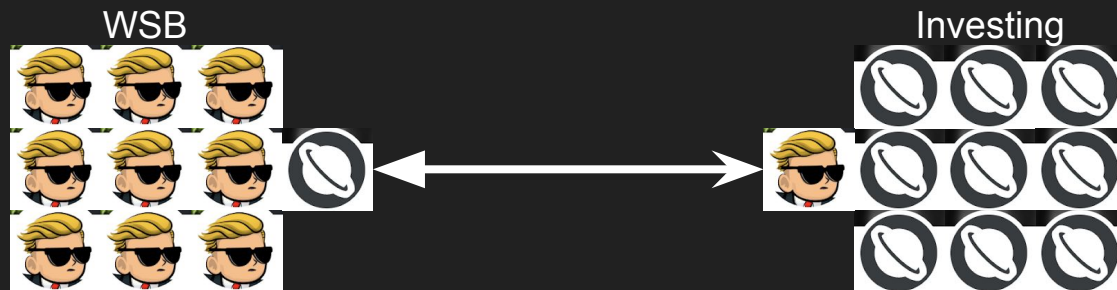
Data Visualization - Sentiment

Modeling

- Count Vectorize - Log Reg: Performed better for the self-text

- Tfidf Vectorize - Log Reg: Performed better for the self-text

- Tfidf Vecotize - Gaussian: Overfit

- Random Forest - Run duration and tuning was demanded

- Tfidf Vectorize - Log Reg: Performed better with the addition of sentiment values on the self text

Results

TfidfVectorizer	Logistic Regression
Fit_Transform a combined column with the self text and title included	Solver: Liblinear - Penalty: Ridge
	CV = 5
Accuracy Score: Training - 92.4%, Testing - 87.4%	
Baseline Accuracy: 50.1% Investing	

Questions?

Thank you.