# ESG Reddit Scraper Report

Amir Hallak

February 20, 2020

## 1 Overview

Environmental, social, and governance standards have often been overlooked when it comes to investing. Lately, however, these standards have become more prominent within the new generation of investors. Just recently, Goldman Sachs announced that they will no longer take companies public if they had an all-male board members. This could be contributed to the social and governance aspect of ESG.

The goal of this scraper is to get an idea of how ESG topics and keywords are trending over Reddit. Reddit is famously known as the "front page of the internet" because it has all the important latest stories.

## 2 Methodology

### 2.1 The Scrape

The Reddit scraper is a python script that goes through a predetermined list of subreddits chosen according to their relevance. These subreddits are financial, environmental, news, and technological subreddits. The code will go over submission information and their first level comments. It does not go into replies since they would be less relevant submissions than first level comments.

The output of the scrape consists of two data frames. One for submission information and the second for first level comments. Table below explains the data variables.

Table 1: Data for **Submissions Data Frame**

| Variable | Definition | Example |
|---|---|---|
| submission_id | Unique submission ID. | e418h |
| title | Title of the submission. | America's possibilities ... |
| author | Author of the submission. | jingowatt |
| num_comments | Number of comments on submission. | 23 |
| upvotes | The number of upvotes on a submission. | 13 |
| upvote_ratio | Ratio of upvotes to downvotes. | 0.91 |
| date | Date of submission - includes time. | 2010-11-10 15:28 |
| subreddit | The name of the subreddit. | Politics |
| keyword | The keyword used in the search. | greenwash |

Table 2: Data for **Comments Data Frame**

| Variable | Definition | Example |
|---|---|---|
| submission_id | Unique submission ID. | e418h |
| comment_id | Unique comment ID. | f8875b7 |
| comment_author | Author of the comment. | jingowatt |
| num_comments | Number of comments on submission. | 23 |
| comment | The body of the comment. | ESG is largely a gimmick... |
| keyword | The keyword used in the search. | greenwash |

With further analysis, more columns are built. However, these columns will be visually represented.

## 2.2   Sentiment Analysis & Word Count

The second part of the script is to analyze the above data. Using NLTK module, we import a sentiment analyzer tool. This tool will determine whether the overall sentiment towards these keyword submissions were either positive, negative, or neutral. The sentiment analyzer can be tuned, and is currently tuned such that any score higher than 0.1 will be labeled "positive", scores between 0.1 and -0.1 will be labeled "neutral", and scores lower than -0.1 will be labeled "negative".
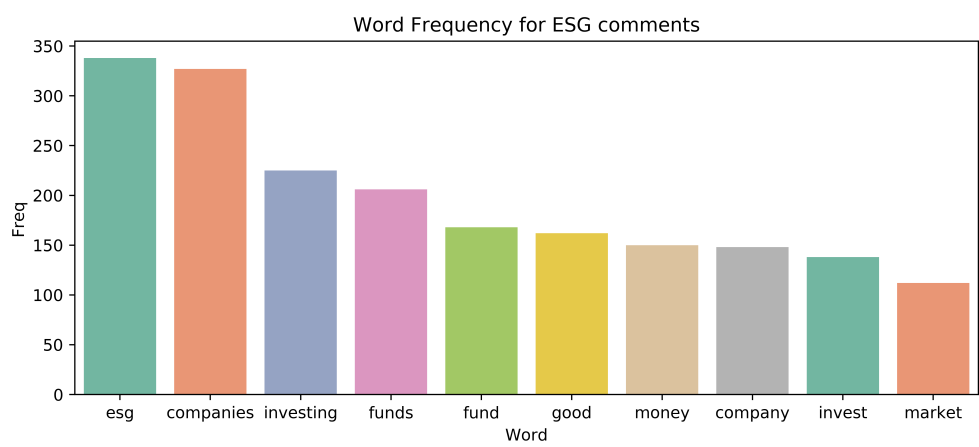
Lastly, also from the NLTK module, we import a word count processor. The aim of this processor is to see which words are being mentioned most in the comments. It will then graph the distribution and show the top 10 words.

# 3 Results

## 3.1 Word Counts

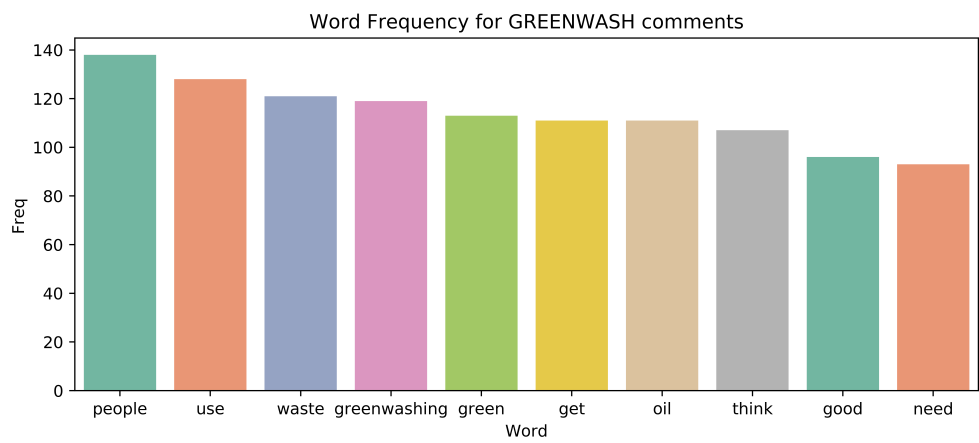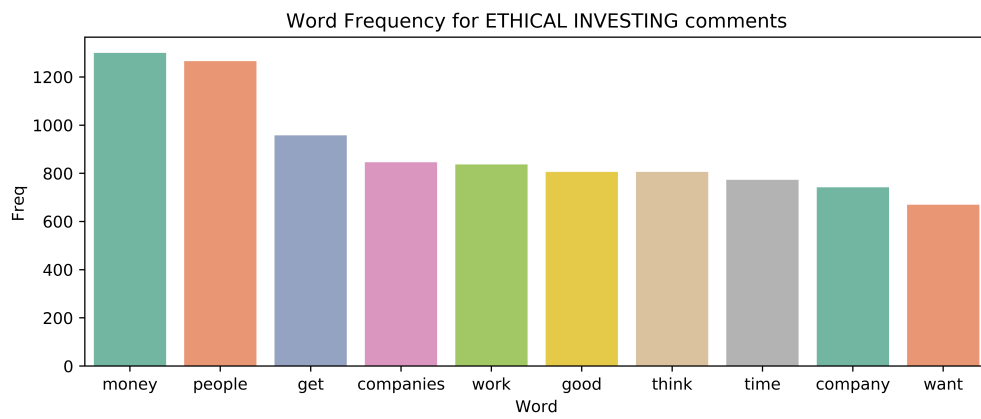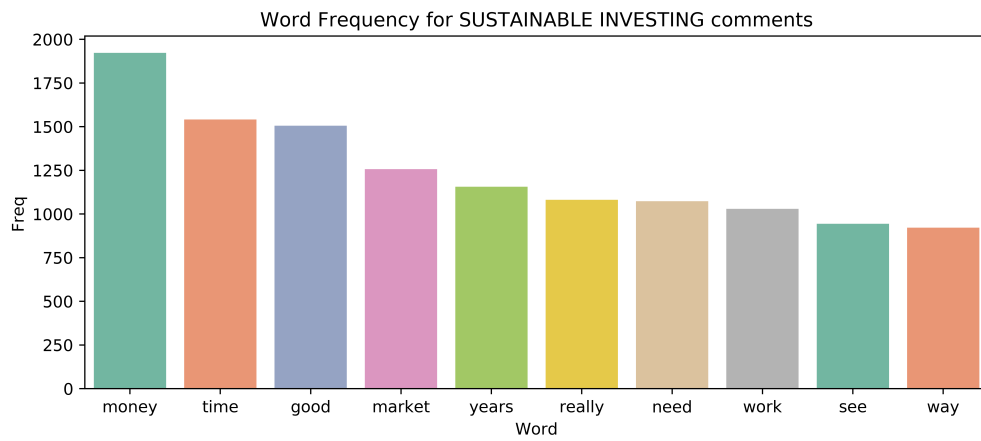Figure 1: Word Count Plots

(a) Keyword: "ESG"



(b) Keyword: "Greenwash"
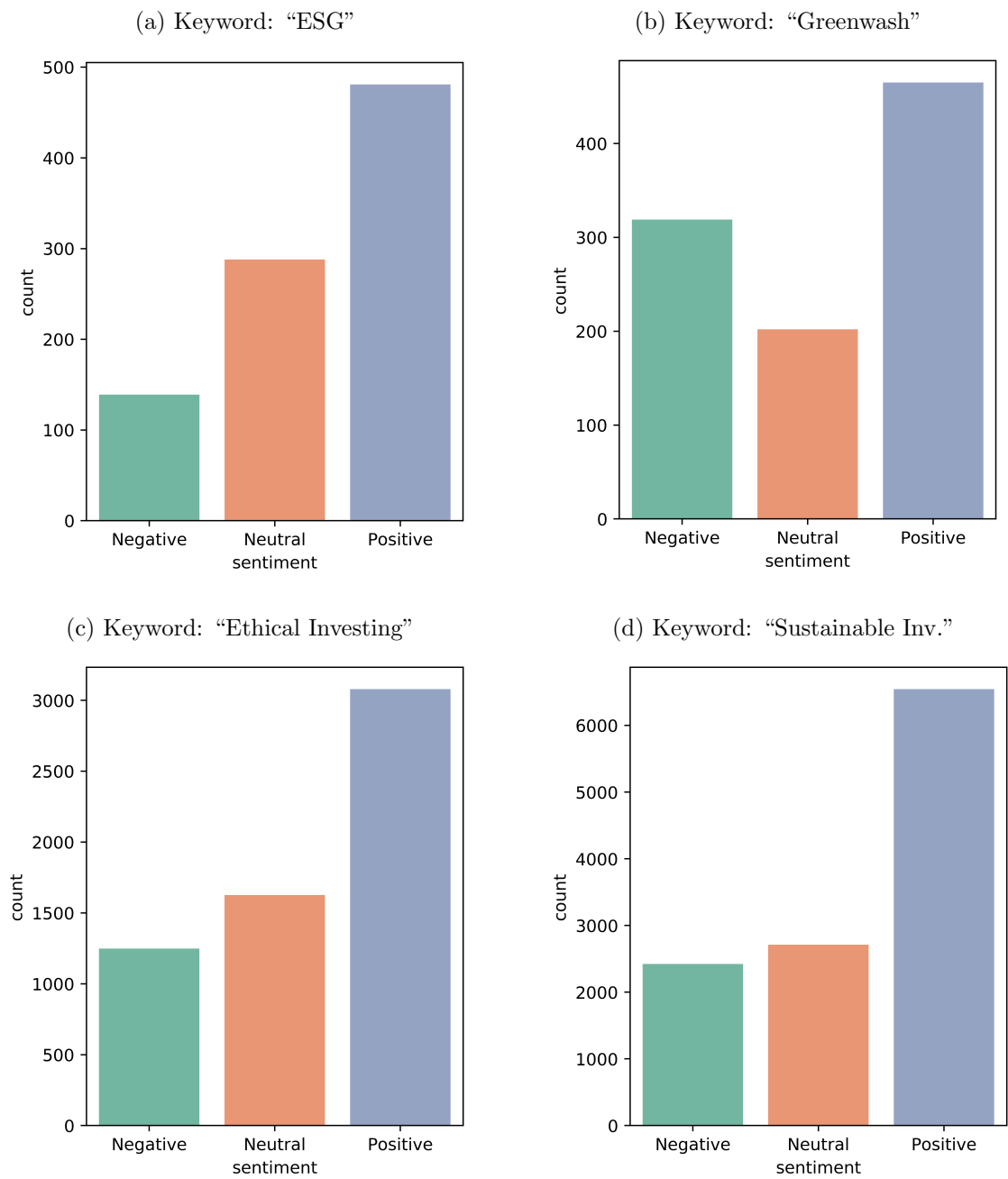
## Figure 1

### (c) Keyword: "Ethical Investing"

**Word Frequency for ETHICAL INVESTING comments**



### (d) Keyword: "Sustainable Investing"

**Word Frequency for SUSTAINABLE INVESTING comments**

## 3.2 Sentiment Analytics

Figure 2: Sentiment Analytics

(a) Keyword: "ESG"

(b) Keyword: "Greenwash"

(c) Keyword: "Ethical Investing"

(d) Keyword: "Sustainable Inv."

## 3.3 Yearly Trend

Figure 3: Yearly Trend Distribution

(a) Keyword: "ESG"



Year Distribution for ESG submissions.

(b) Keyword: "Greenwash"



Year Distribution for GREENWASH submissions.
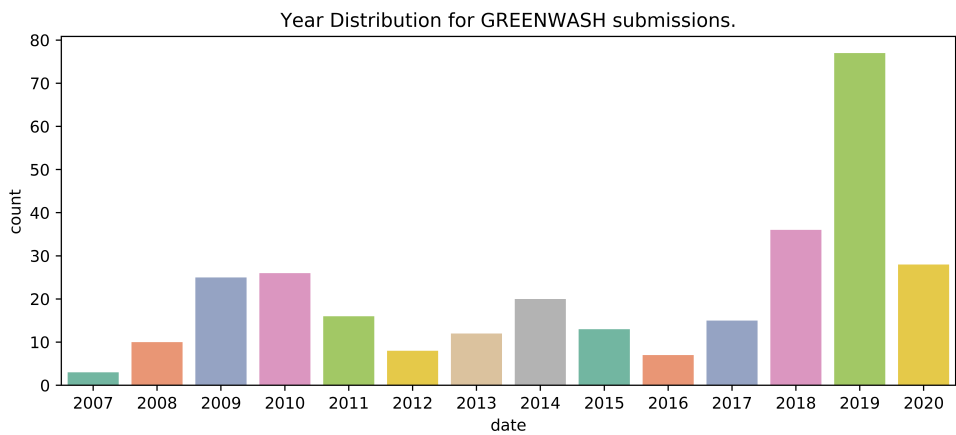
Figure 3

(c) Keyword: "Ethical Investing"

Year Distribution for ETHICAL INVESTING submissions.



(d) Keyword: "Sustainable Investing"

Year Distribution for SUSTAINABLE INVESTING submissions.
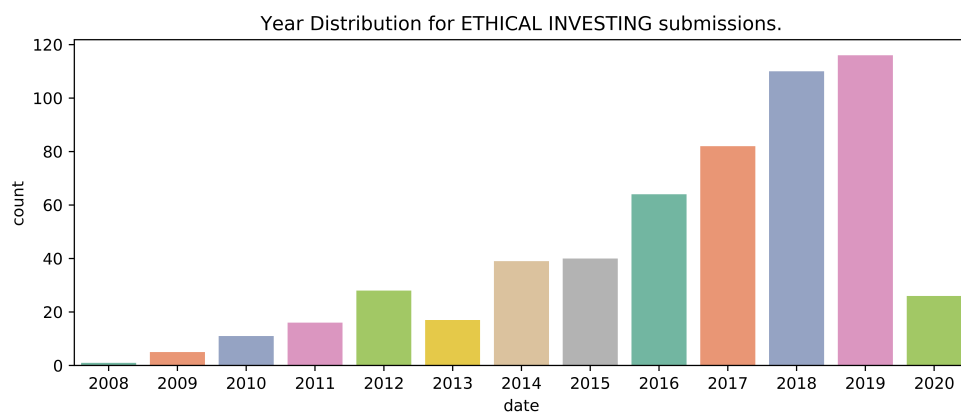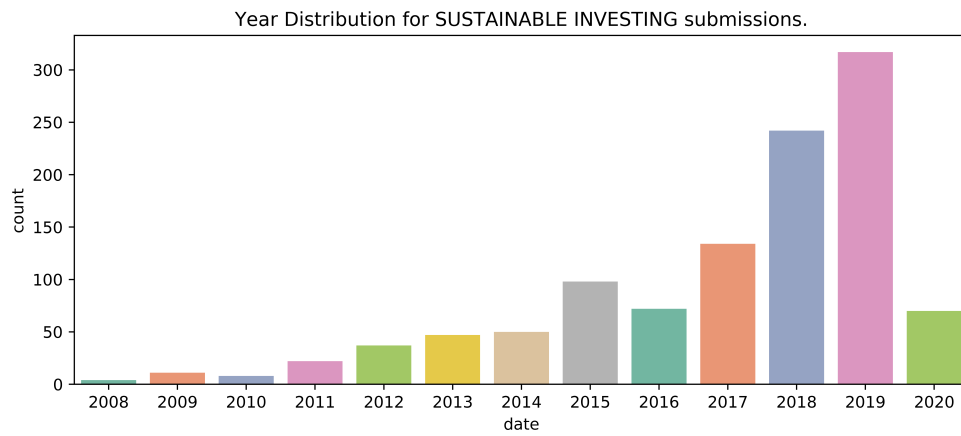
## 3.4 Analysis & Conclusion

The above graphs show that ESG topics are becoming more relevant. Sustainable and ethical investing are trending upwards with no signs of slowing down. Greenwashing is mentioned more over the years. Meaning, corporate greenwashing is getting more attention in the media which could potentially lead to increased accountability. Latest example regrading greenwashing was Jeff Bezos pledging $10 billion to fight climate change. Many Reddit submissions mentioned corporate greenwashing being the driving motivation.

Figure 1(a) shows that the bulk of the words mentioned are words that relate to investing. Investors are looking for companies that would be profitable in the long run. Companies with good ESG records will most likely carry the lowest risk of being sued as opposed to more profitable companies in the short run. Looking at Figure 1(b) we can clearly see the word "oil" as one of the top 10 words when searched for "greenwash". This goes to show that oil companies are mostly talked about when discussing corporate greenwashing.

Figures 2(a),(c) show an overwhelming positive reaction or sentiment towards ESG, ethical investing, and sustainable investing. Figure 2(b) talks about greenwashing and the negative sentiment is clearly shown.

Lastly, the yearly distribution is the biggest indicator that these ESG topics are being talked about more. There is an increasing trend that is peaking almost every year with 2020 already being at a quarter or a third of 2019.