

Assignment 2

Saujanya Acharya

Eric Ohemeng

Bozhou(Peter) Tan

2024-03-24

Introduction

In this project, the research topic we have chosen is whether the current economic situation in the United States is improving or deteriorating. As we enter 2024, some surveys and statistical data indicate that the U.S. economy is moving in a positive direction. For instance, the Consumer Sentiment Index has risen by over 20 points since last December, and inflation expectations have also decreased, falling into the range seen in the past few years. However, many individuals also express that they haven't felt the improvement in the economic situation themselves; instead, they feel it's getting worse — with inflation still high and income growth sluggish. We aim to understand the perspectives of different individuals in society regarding the current economic situation through discussions among users on Reddit.

Methods

To extract the threads and comments from Reddit, we first need to choose the keywords or subreddits. There are two methods to extract the posts through Reddit API: one approach is searching the keywords to get a list of threads and extracting comments from those threads; the other way is choosing a specific subreddit and extracting the comments from the subreddit. The former method provides the threads from a wide range of subreddits, which means there are lots of irrelevant threads in the results; the latter method can make sure all threads are under a specific topic or keyword. After several experiments, we adopted the latter method, choosing several subreddits for listening.

Economy situation is a large topic, including inflation rate, salary and unemployment rate. First of all, we chose the subreddit “economy”, because this subreddit contains threads and comments covering almost all topics about economy, which can help us understand people's feelings about economy situation from many perspectives. There are also hundreds of comments in this subreddit, making the sample size large enough. We also chose some subreddits focusing on a specific economic topic: “inflation”, “business” and “salary”. These subreddits are also helpful but there are some disadvantages compared to the subreddit “economy”. Firstly, threads from these subreddits mainly focus on one economic topic. Secondly, lots of threads and comments are not related to our

research topic. Therefore, while we also listened to these subreddits, we leaned towards utilizing comments from “economy”. If the data from “economy” is sufficient, we do not intend to utilize comments from other sources. If not, we will also reach records from them.

We listened to those four subreddits for 8 days, from March 16th to March 23rd. Every evening, one member in the group extracted comments from the subreddits posted in the last 24 hours, stored data as .csv files and uploaded them on **GitHub**¹. The details of listening results are shown in Table 1. There are at least 400 posts in “economy” subreddit every day and 4520 comments in total.

After collection, we divided the comments we have in “economy” subreddit into three parts and each member coded the comments into four categories: favor, oppose, neutral and irrelevant. Since the number of relevant posts from “economy” subreddit meets the requirement, we did not code the posts from other subreddits.

Table 1: The number of comments from different subreddits

Date	Economy	Inflation	Business	Salary
2024-03-16	723	640	49	297
2024-03-17	513	185	293	73
2024-03-18	438	311	162	87
2024-03-19	551	339	94	128
2024-03-20	515	278	180	0
2024-03-21	795	1233	85	0
2024-03-22	571	328	116	0
2024-03-23	414	338	75	161

Analysis

Building upon the comprehensive data collection framework outlined in the Methods section, we started the analysis of the posts by first cleaning the hence collected data. The posts with “Irrelevant” posts were first removed. Finally, we ended up with a list of 838 posts/comments that had some relevance to the topic of our interest.

The distribution of the sentiment of the posts is illustrated in Figure 1. The initial review of the hand-coded labels for each post showed that the majority expressed opposing sentiment to “economy is improving”. More than half of the posts indicated that the reddit users felt that the economy was not faring well, and has been worse than the past times. On the contrary, analysis indicated that less than 20% of the posts had positive/favorable sentiments associated to them.

Furthermore, the distribution of time of the posts over the day was also analyzed. Figure 2 illustrated how most of the posts were made on the 2nd half of the day through to near mid-night. The discussions were less frequent over the span of the night. Moreover, it was observed that the most frequent posts were made from 3 PM to 6 PM.

¹<https://github.com/petertbz/622HW2>

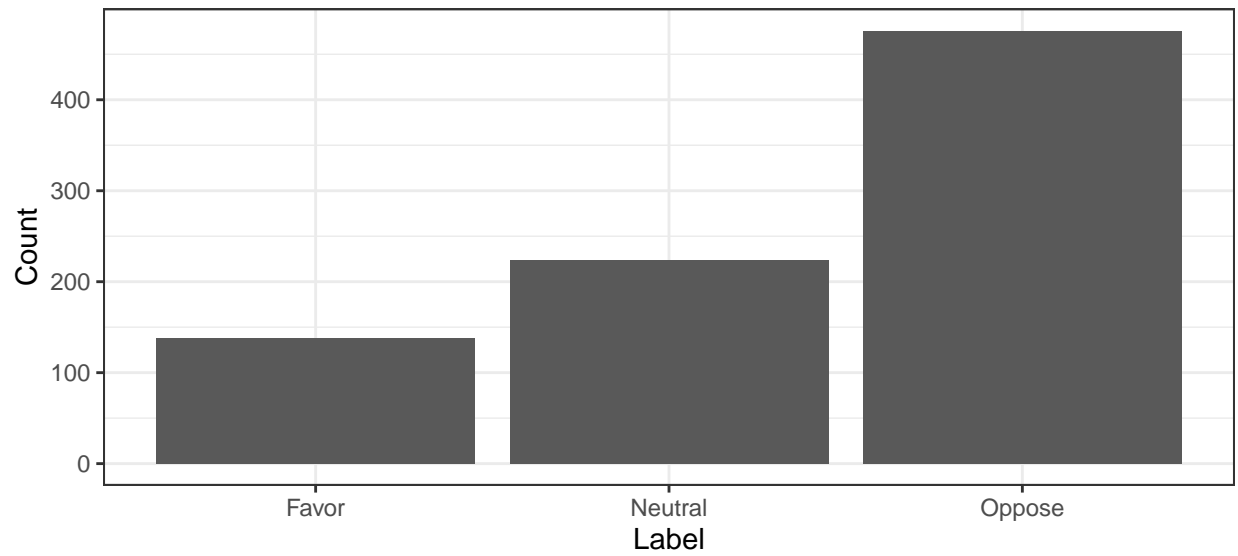


Figure 1: Distribution of in Favor (F), Neutral (N) and Opposing (O) sentiment of cleaned posts

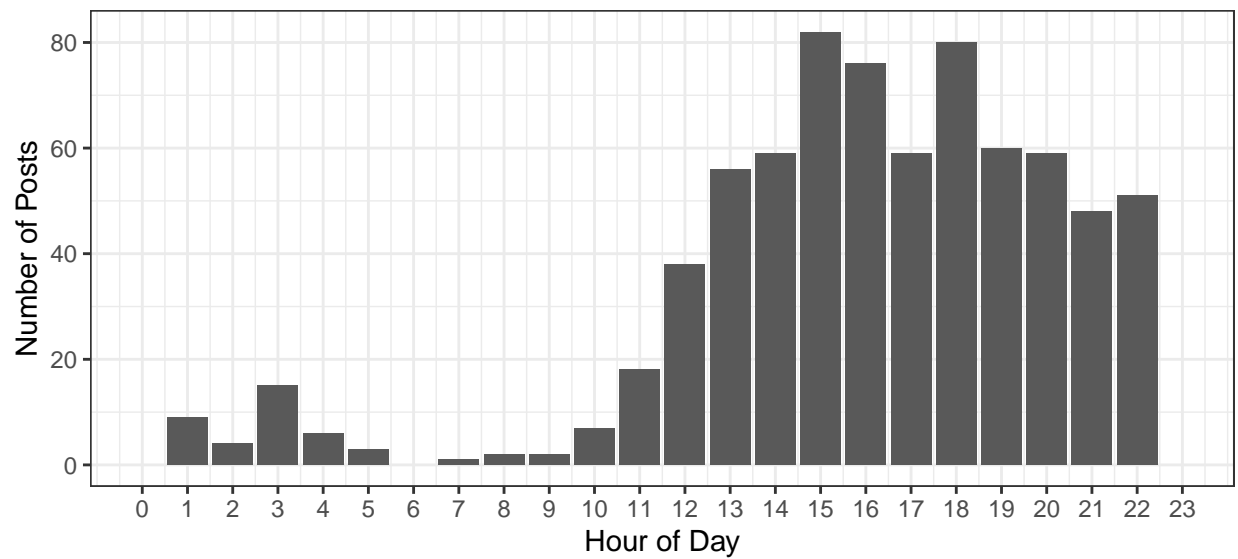


Figure 2: Distribution of posts made over the time of the day

After the initial meta-analysis, the posts were then reviewed to analyze the frequency of words to gather insights on specific topics that were being discussed in these forums. To do this, first the data was tokenized, and cleaned. Moreover, the common stop words were also removed with the use of tidytext library in R. It was observed that the most common words that appeared in the discussions were as expected from our initial assumption of the kind of discussions happening in the online forums. The frequency of the most common/frequent words occurring in the comments has been visualized as Figure 3. From the figure, we can see that the most common words were “people”, “inflation”, “money”, “economy”, “prices”, “housing”, “home”, “market”, “rates” and “tax”. Most of the words above reflect the topics that people are most concerned about in the current economic situation.

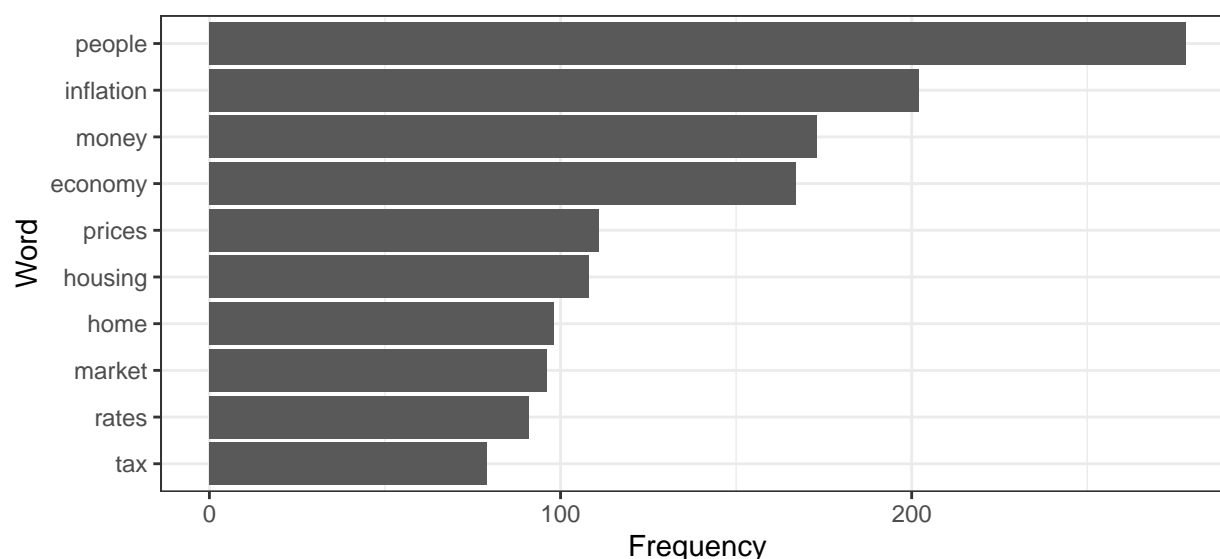


Figure 3: Frequency of most commonly used words in the discussions

We also focused on the dynamics of discussions revolving around key socioeconomic events of the week: the release of inflation figures, the Biden administration’s revision on electronic vehicles adoption target and the federal government effort to address housing challenges.

Figure 4 shows the number of comments by day with key events. The graph starts with a moderate number of comments, and we see a small peak coinciding with the “Inflation” event. This is due to the release of several inflation reports (e.g., the Consumer Price Index), which on spurred a discussion among users, reflected in a mild increase in commenting activity for some days. A significant peak occurs on March 20, 2024, the day Biden administration slashed its target for U.S. electric vehicle adoption from 67% by 2032 to 35%. This also resulted in a high level of user engagement. The conversation centered around the effect of electric cars on the U.S economy. Following the “Electric Vehicle” event, there’s a sharp decline in commenting activity before another spike on the “Housing” event day. The release of the annual Economic report which indicated a proactive stance by economists within the Biden administration, advocating for robust federal intervention to alleviate the financial burdens shouldered by prospective homeowners and renters especially Family Size Houses. Such news, given its potential impact on a wide range of issues — from homeownership to the broader economy-naturally serves as a catalyst for debate and dialogue among concerned

citizens.

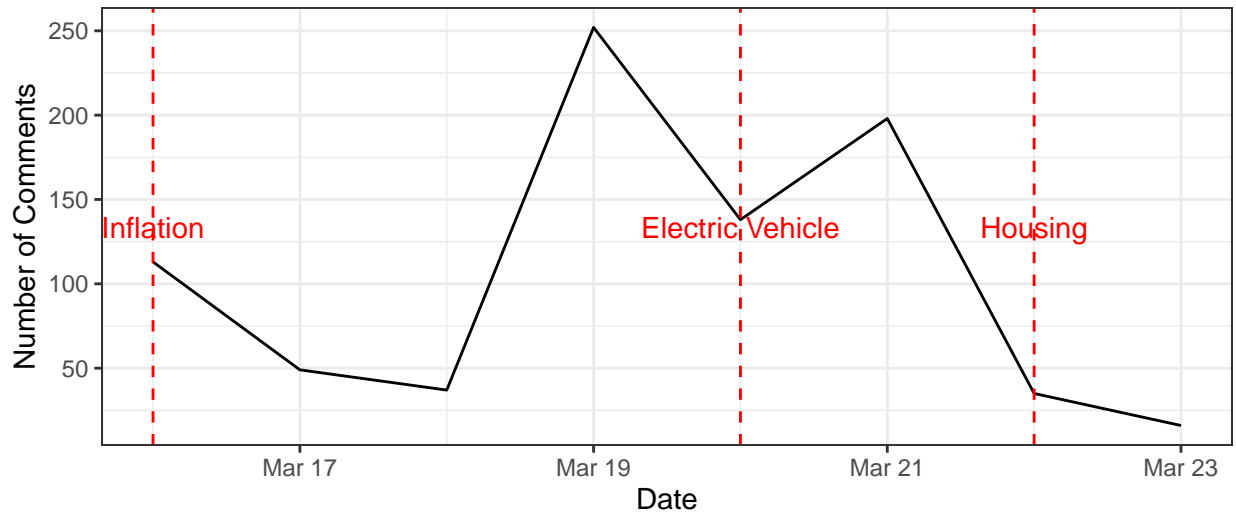


Figure 4: Number of Comments by Day with Key Events

Figure 5 shows the number of each category by day. It is evident that the “Oppose” category has the highest number of comments on all three peak days, followed by the “Neutral” category. The “Favor” category has the lowest number of comments on all three days, indicating that the majority of the comments are negative or opposing the events. However, this phenomenon does not necessarily mean that the majority of the public is against these events. It could be that people who are against these events are more vocal and active in online forums like Reddit.

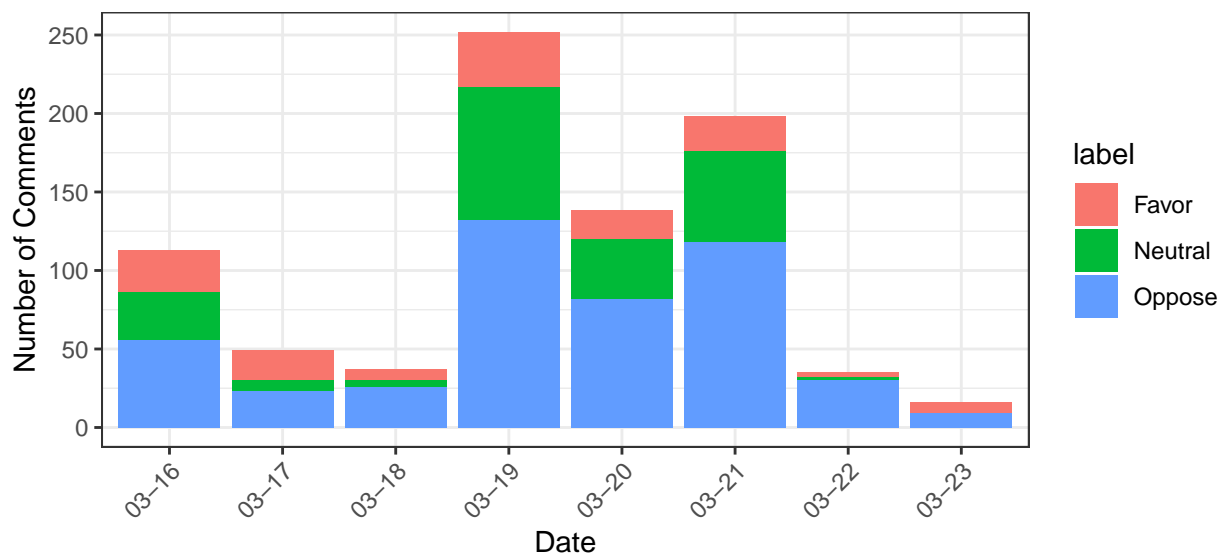


Figure 5: Composition of Comments by Day

Conclusion

Our project delves into Reddit discussions on the economy, revealing widespread concern over issues like inflation and housing. We found that people often post in the late afternoon, and significant news events tend to spark even more discussion. The data shows that despite some optimistic economic indicators, many people remain worried about the cost of living and their financial futures. However, the larger number of negative posts may also be “selection bias” in the social media. In general, this snapshot of public opinion shows us that economic issues are a major concern for many, and what happens in the economy is closely followed by the online community.