Reddit Sentiment Analysis and Topic Modeling
Rye International Data Team presents for Onomy

---

1. *Key Takeaways*

   a. Exploratory discussions on sentiment trends and compositions of different topics.
   b. Summaries on multi-sentiment analysis for adulting and finance datasets.
   c. On the keyword "social", people display anxious and struggling emotions.
   d. On keyword "advisor", people show the emotion of trust regarding bank accounts, financial advisors, credit card, insurance.
   e. On the keyword "emergency", people care about emergency funds and monthly expenses, mentioning them with a sentiment of trust.
   f. On the keyword "loan", people need advice on student loans and having lower interest rates, in which various loans are related to mixed sentiments.
   g. On the keyword "credit", people discuss their credit card usage and ways to build up a good credit, while when mentioning credit card debts, people have emotions of fear & disgust & sadness.
   h. On the keyword"college", mental health issues, getting first job and moving out is the focal point of discussion
   i. On the keyword "time", people feel positive about planning ahead about their personal finance and feel negative about being behind in life.
   j. On the keyword "parent", 'living with parents' leads to most of the negative feelings, including anger, disgust, and sadness.

2. *Motivation*

   a. **Motivation:** We try to discern the sentiment trends of social discussion about different topics relevant to adulting and personal finance by grouping and analyzing text data from Reddit. We hope to gain a better understanding of users' attitudes for different topics, which potentially provide insights for what popular concerns which Onomy could address in courses.

   b. **Questions we try to answer**

      i. What are the general sentiment trends by keywords and by subreddits?
      ii. What topics are relevant to adulting and personal finance keywords and what are their sentiment trends?
      iii. What kind of emotions are attached to the keywords?
      iv. What are the topics under each emotion?

3. *How do we conduct EDA? Where is the data from? analysis methods?*

   a. **Data source: Posts associated with 7 relevant subreddits on Reddit**
   b. **Analysis process:**
   Sentiment Score:

   We use VADAR to get a quantitative evaluation on Reddit posts in three dimensions---positive, negative, and neutral. Then, we generate a feature by computing (positive-negative)/neutral as

our metrics to measure sentiment. This score is generally between -1 and 1, while a positive figure implies positive sentiment and vice versa.

Multiple Emotion Detection:

We employed the NRC Emotion Intensity Lexicon (NRC-EIL) which is a dictionary of 9922 lemmas associated with scores reflecting the emotions they convey and their intensity.

4. *What are the results and implications?*

I. Sentiment Score by Subreddit:

Sentiment Score by Subreddits
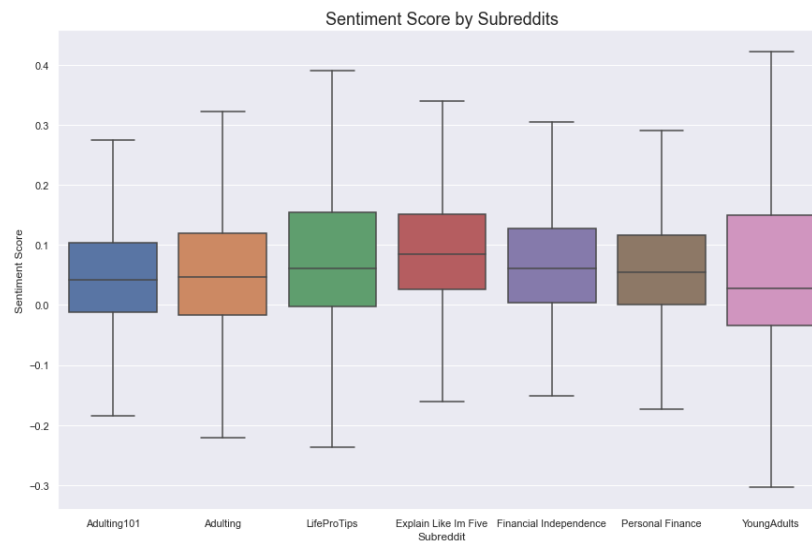
A. Difference in Subjectivity:
YoungAdults has larger variance in sentiment, which indicates they are **more emotional** in their reddit posts.
Financial related subreddit generally has smaller variance in sentiment and higher sentiment score than those related to adulting. This might imply that **users who concern financial issues are mature adults.**
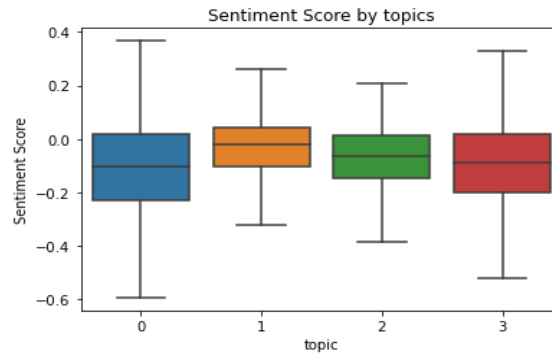Overall, the median sentiment scores in the subreddits are all above zero, indicating that people tend to express in a positive way.

II. Sentiment Score by Keywords:

In the following part, we will interpret the sentiment scores of each keyword and the sub-topics under them.

Sentiment Score by Keywords

A. Emergency: 966 posts
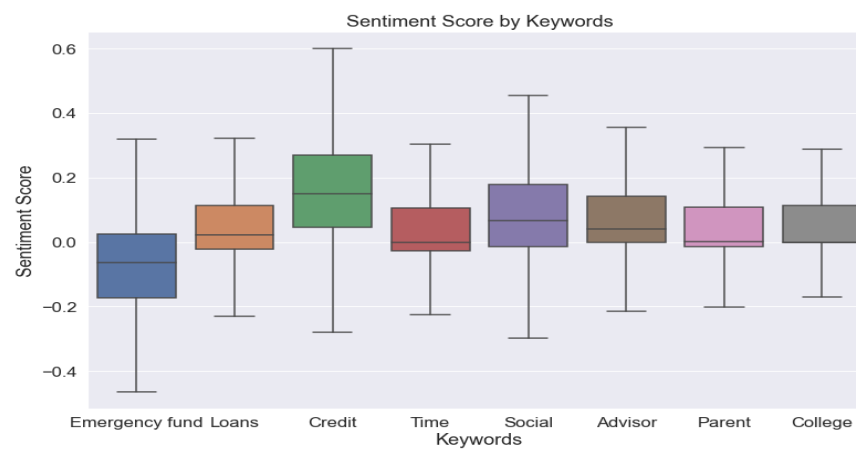
Sentiment Score by topics

Topic 0: "emergency fund", "savings", "earnings"
Topic 1: "roth ira", "loan", "debt"
Topic 2: "accounts", "yield"
Topic 3: "monthly expense"

Insights:

1. Topic 0 and 3 have the lowest sentiment scores, topic 2 has a slightly higher score, and topic 1 has the highest. Overall, **Emergency has the lowest sentiment score among all keywords, mainly because the word 'emergency' itself carries negative sentiment** and a sense of fear, worry, and sadness, as emergency probably refers to some unexpected negative events.
2. The most popular discussions were on the topic of **advice on savings accounts and emergency funds**, but generally the posts were **getting very few likes**.
3. Topic 0 is about **Cases of Using** emergency funds with negative feelings and situations. Associated words include: "ill", "worry", "pandemic", "hope"
   a. Example of topic 0: "My parents also have an emergency fund, so incase I lose my job I can ask them for money (not a lot, but a small amount)"
4. Topic 1 is on Loan, Debt and Roth IRA, which brings in some sense of anticipation of being **able to pay back**. Associated words include: "crisis", "planning", "important"
   a. Example of topic 1: "I have had my car loan for 36 months now and when I first financed the car I had no credit history whatsoever and I ended up having my dad be the primary signer on the loan (his credit score 800ish at the time) and I was the co-signer and I now have my credit score sitting anywhere between 734 - 752. I figured now is a great time given that interest rates are at an all time low, my credit score is at an all time high and I'm three years into paying it off."
5. Topic 2 is on Accounts and Yields on bank accounts that are for emergency purposes. This is more associated with "fire", "injury", and "guilt".
   a. Example of topic 2: "Currently I have one savings account and one checking account (outside of investment accounts) and I pay bills from checking, then my savings includes money for emergencies."

6. Topic 3 is on Monthly Expense, which includes people's worries about unwise money spending. Sentiment words include: "isolated", "insane", "security"
    a. Example of topic 3: "Advice on my plan of eliminating monthly expenses incase of future job loss"

B. <u>Loan: 5229 posts</u>
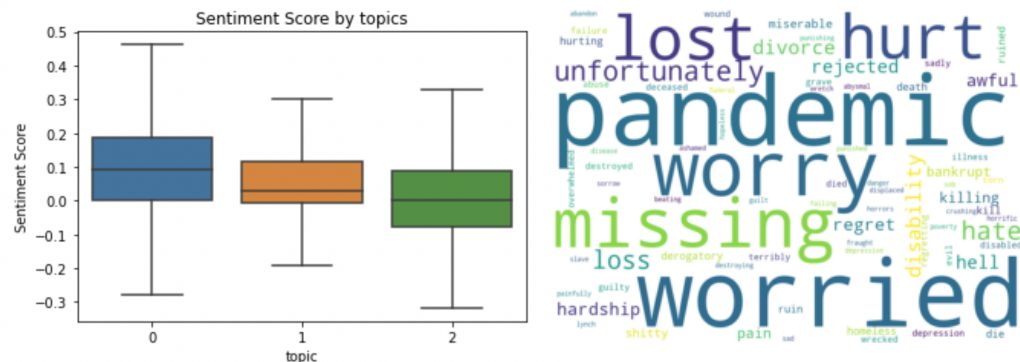
Topic 0: "credit card", "credit score", "credit debt"
Topic 1: "car loan", "interest rate", "monthly payment"
Topic 2: "student loans", "loan debt","federal student"

Insights*:*
1. The three topic groups include **different kinds of loans** with different sentiment score performance. Topic 0 has the highest sentiment score, and it's probably because "credit" carries positive meaning. Topic 1 is mixed, and Topic 2 is the most negative. We can infer that discussions around **student loans are mostly negative**.
2. Topic 0: Credit cards related. More associated with "good" and "start" compared to the other two topics, and related to "worried" and "emergency" to some extent.
    a. Example of Topic 0: "After years of being very conscious about it I finally got my score into the "excellent" section, applied for my loan consolidation and got approved for a loan with 4% interest without needing a co-signer" (credit score, positive) ; "I took out loans and opened credit cards to make ends meet, most of which I was unable to maintain payments on."
3. Topic 1: Getting advice on having **lower interest rates** on loans and on **approval of large loans**. More associated with "approved", "officer", "advice" compared to the other two topics.
4. Topic 2: Student loans. More associated with "emergency", "gross", "hurt", "worried".

      a.   Example: "I owe a student loan and $200 in a doctor bill. Once I start my job, I don't know if I should pay back my loan first or move out first, and when I should apply for the assistance."

C. <u>Credit: 5994 posts</u>

Sentiment Score by topics
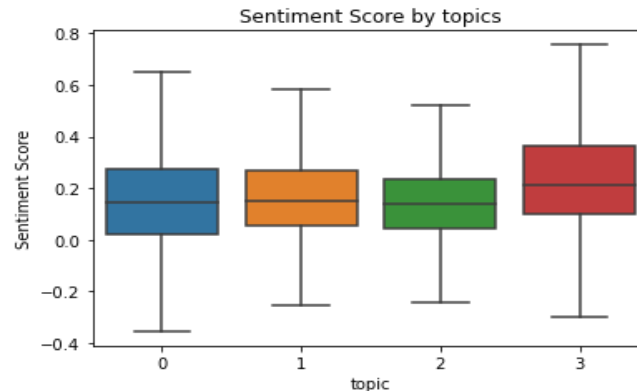
Topic 0: "Student debt", "Credit card", "payments"
Topic 1: "bank account", "loan"
Topic 2: "paying back debt", "checking credit reports"
Topic 3: "credit scores", "getting good credit", "things affecting credit"

Insights*:*
1. Credit has the **highest sentiment score** among all keywords and the **largest IQR**, mainly because the word credit itself can be used to demonstrate either a success in paying back debt and getting good credit or having to struggle to pay back loans. As a result, the comments were often polarized.
2. The most popular discussions were about **advice on the usage of credit cards** and people sharing their experience **improving their credit scores.**
3. Topic 0 is about paying back student debt and credit cards which could **include mixed feelings**. It is more associated with "struggle", "worry", and "approved".
   a. Example: "So I got the idea to fix my credit when my Dad called about a cosigner. My credit is garbage due to bad choices during and after college. Mainly my major...So I decided to use a bit of my savings to get a new credit card through my bank so I can repair my credit."
4. Topic 1 is about bank accounts and loans. This is more associated with "excellent", "friend", "ally", and "pandemic".
   a. Example: "my plan is to open a bank account, apply for credit or whatever and save the rest of the money for a rainy day."
5. Topic 2 is about paying back debt and checking credit reports. This is more associated with "derogatory", "verification", "crisis", and "freaked"
   a. Example: "Only debt i have at the moment are student loans , i paid off my credit bills once COVID became a thing, tips on budgeting/paying off student loans would be helpful"
6. Topic 3 is about credit scores and ways to get a good credit score. More associated with "nightmares", "targeted", "affirm", and "appreciation".

      a. Example: "So I'm 19 and in the Army and I want to start building my credit score but my parents are anti-credit cards so I have nowhere else to go for help."

D. <u>Time: 1320 posts</u>

Sentiment Score by topics
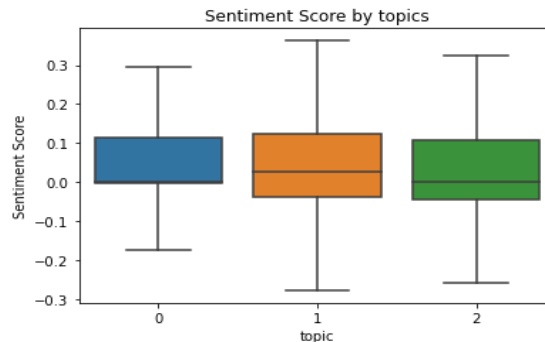
Topic 0: "tax" "month" "money" "company" "started" "earlier year"
Topic 1: "retirement" "age" "feel like" "far"
Topic 2: "don't time" "know" "work" "day"

Insights*:*

1. Topic 0 is about **planning ahead of time about one's personal finance**, which achieves the **highest** sentiment score among all topics. Sentiment words include "hope", "planning", "team", "holiday", and 'struggle' with expressions related to 'earlier year'.
   a. Example: "Like I said earlier, when I wa working my first fast food job making minimum wage at 16, the paycheck would go to the bank to save up for a car, and I'd keep about 40-50 for spending money which wa often too much cause I didn't use if on anything but a stop at 7-11 or fast food."
2. Topic 1 is about **planning about life and retirement.** Sentiment words include "professional", "pleasure", "secure", "worried", 'regret', and "failure" with many expressions related to '**leaving behind**'.
   a. Example: "I'm 28 now, just about the graduate engineering school (Yes, getting a BS at 28 is NOT good at all, there is a story for why I am like 5 year behind in my life, won't go into that here)."
3. Topic 2 is about random posts that mention the keywords 'earlier', 'behind', and 'time'.

E. <u>Social: 1082 posts</u>

Topic 0: "social security", "security medicare", " security benefit", "bank account", " pensional social"

Topic 1: "social life", "social skill", "social circle", "social groups", " social anxiety"

Topic 2: "medium platform", "time spend", "medium account", "lot people", "post social", "e commerce"

Insights:

1. Topic 0 is associated with banking and social security. The word 'security' might contribute to the positive sentiment of the topic social security/pension. Its common negative words are **typically "worried", "lost", "disability", "paranoid**". The word "**pay**" also appears as a top **negative** word, which might indicate that people have trouble dealing with payments perhaps in going out with friends and need suggestions with that:
   a. Post with top upvotes: "And I'm not even comparing myself to US colleagues that are 10x ahead with some NW of 1M USD by 27 or having 2 triplexes or 4 cars.I didn't take into account any social securities because I don't believe *I will receive* any."

2. Topic 1 is relevant to social life and social group, its negative words are more **emotional** such as "isolated", "struggle", "terrifies", "lonely", "lost", "depression"; Other negative words related to social relationships appear frequently such as **"teasing" and "awkward"**. People may be concerned on how to deal with non-satisfactory occasions in their relationships:
   a. An interesting post with the top 2 number of comments: "If you know someone who passed away recently please wait at least a day before posting on social media. This meant that a lot of people found out through social media before family members could even call them."

3. Topic 2 is more related to media. About the topic of social life and friends, the word **anxiety** frequently appears. The trend of anxiety relating to social life might indicate that people may be in need of suggestions or guidance on how to effectively navigate social life/manage their social anxiety. There are also many more **emotional words** relating to relationship status, such as "ill", "abuse", "destroying", "victim", "harmful", "injury".

F.   Advisor: 535 posts
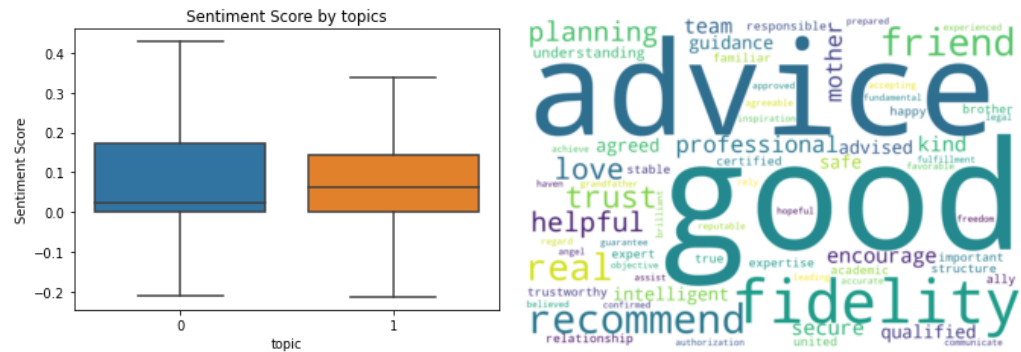
Topic 0: "money financial", "bank", "investment advice", "advisor help"
Topic 1: "roth ira", "saving account", "account robo", " schwab intelligent",
"intelligent portfolio"

Insights:

1.   Surprisingly, the topic of Robo-advisor has a **higher median** sentiment score.
     The majority of people mentioning Robo-advisor tend to have a positive
     attitude. Another interesting thing is though the topic of financial
     advisor/bank/investment has a lower median score, its **greater variance shows
     the better services that banks could offer relating to investment**.

2.   Topic 0 is related to investments advice, its positive words are **"recommend",
     "guidance", "fidelity", "understanding", "trustworthy"**; its negative
     connotations are "fear", "insane", "paranoid", "worried", "pandemic".

     a.   Example post with the most comments in Advisor set, which show that
          people convey the negative feelings towards the **expensive fees of
          Human Advisor but still the trust** for its helpfulness: "I know we like
          to say that financial advisors are too expensive, not worth it, etc, but
          someone who had simply said "Don't be a muppet and stay the course"
          would have earned 19 years' worth of fees."

     b.   Example posts with most upvotes in Advisor set: both indicate people's
          confusion on financial advising regarding the investing process and the
          overmuch options available to them.

          i.   "So I'm really confused by this whole process, and I do have
               advisors assisting me."

          ii. "We have spoken with advisors before, and honestly, they give
               too many options."

3.   Topic 1 is related to robo advisor. Common positive words are "good",
     "qualified", "planning", "helpful", "safe", "stable". Negative words in
     sentiments such as "sadness" and "fear" are "disability", "aggressive",
     "harmful", "scared", "disease", "dismay", "overwhelmed". The word "sure"
     appears as a top negative word in the robo advisor topic part, which might
     indicate that people may **feel uncertain about the accuracy of the suggestions**
     provided by robo advisor.

G. <u>Parent: 3882 posts</u>

Topic 0: "money", "house", "time"

Topic 1: "bank", "company", "credit", "account"

Topic 2: "currently living with parents"

Insights:

1. Topic 0 has the highest sentiment score, Topic 1 is mixed with large sentiment variance, and Topic 2 is the most negative. We can infer that **living with parents is problematic for people.**

2. Topic 0 is about the worries in **money and house** with parents, which is more associated with "love", "planning", "hate", and "divorce" compared to the other two topics.
   a. Example: "My parents support me financially and I'm safe and comfortable. I want my parents to enjoy their money and stop spending on a 25 year old baby.I've been looking for a job for a month and a half."

3. Topic 1 is about the **confusion in banking issues** after leaving parents. This is more associated with "fidelity", "friend" , "trust", "frustrated", "pain", and "lost" compared to the other two topics.
   a. Example: "So for years I've had issues with my parents taking everything I do and everywhere I go, and I think it's about time I get out of this. Since I've never been taught how to do thing myself, and cannot ask my parent for help, I have so many questions.How do I set up my own back account and be able to transfer the money from my checking and saving over?How do I choose a bank?Are there different kind of account and if so which do I need?Where can I learn how to properly budget all my expenses?)How do I prepare for paying for property?How do I do taxes?How do I maintain a good credit score?None of these things can be done with my parents as they will obviously refuse."

4. Topic 2 is about **living with parents**. It is more associated with "advice", "relationship" , "battle", "guilty", and "depression".
   a. Example: "I'm 22 and I still live with my parents. Back last year, my parents showed some clear signs of tension and it climaxed in

November of 2019. And now, I'm frequently feeling torn between each parent and it's pretty mentally fatiguing."

H. <u>College: 3269 posts</u>

Topic 0: "currently", "20 year old", "college student"
Topic 1: " graduated college", "recently graduated", "year ago"
Topic 2: "college degree", "job", "pay college"

Insights:
1. We can see that topic modeling successfully split the keyword college to three topic groups with three age groups. The three topic groups have different sentiment score performance. With the keyword "college", all three have a medium around zero, and most of the bar is around zero or above, indicating that the keyword "college" is generally positive. Discussions of current college students and those around jobs and paying college have more variance in their sentiment score.
2. Topic 0 is more associated with joyful words like "good" , "start", "advice" and also associated with negative words like "hate", "shit". This group is **concerned about aspects of college life and confused about the future**.
   a. Examples of topic 0: "My parents did not go to college and don't know how to advise me. My question is, would it be a good idea to switch majors to one of the business majors being offered at the art school?" (college life, positive); "How to be a motherfucking successful ass adult"(confusion about future, negative)
3. Topic 1: associated with "start", "real", "good", "pandemic". Discussion in this group includes **getting a first job and dealing with moving out**.
   a. Examples of topic 1: "Low-income housing for a single mom. Thoughts and advice?" Besides, this group might be hit by pandemic harder than the other two because they mention pandemic more frequently.
4. Topic 2: A different word that appears in the word cloud is "retirement". In this group, we notice a **higher mentioning of mental health issues**.

a. Examples of topic 2: "I 've only been out on my own for 2 and a half years, since graduating college.I have no reason to feel stressed and irritated. I simply don't feel like "lifeing" anymore."
b. "I feel like I've ruined my life in ny early 20's due to student debt. How do I dig myself out of this hole?"

III.   Sentiment Composition:

A.   General Sentiment Word Cloud

Word Cloud for Anger

Word Cloud for Sadness

Word Cloud for Fear

Word Cloud for Trust

Word Cloud for Joy

Word Cloud for Anticipation

Word Cloud for Surprise   Word Cloud for Disgust

B. <u>Sentiment by Keywords:</u>

Insights:
- We switched to using radar graph for this section (Instead of bar plot, now in the Appendix)
- The general trend is the sentiment of **"trust" appearing the most frequently,** followed by fear, then joy, and with surprise being the least frequent sentiment.
- With radar plot, we can also easily find the keyword with the highest composition of each sentiment

Sentiment	Keyword with highest sentiment composition	Keyword with lowest sentiment composition
Trust	emergency	parent
Surprise	loan	credit
Sadness	parent	time
Joy	college	loan
Fear	parent	time
Disgust	parent	emergency
Anticipation	emergency	parent
Anger	parent	emergency

- Also, if we sort the sentiment by keywords, there is the **most sentiment associated with the "parent"** keyword and followed by the "time" and "loan" keywords.

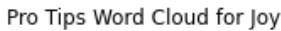**Composition Proportion of Sentiments Radar Plot (Dividing each sentiment by the total sentiment)**
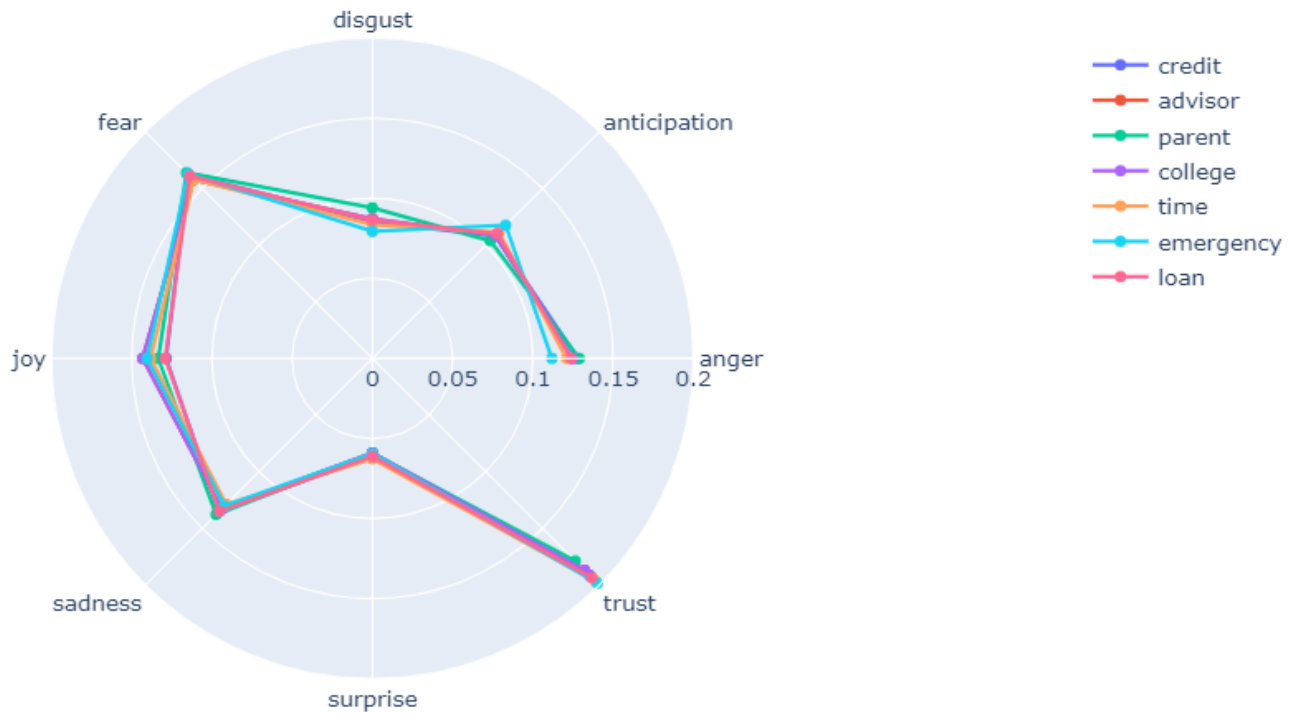
Rescaled Sentiment Radar Plot (Using same sample size)

C. Sentiment by Subreddit:

Explain Like I am Five

Financial Independence

Personal Finance

Adulting 101

Adulting

Pro Tips

Young Adults

Financial Independence Word Cloud for Joy

Pro Tips Word Cloud for Joy

Young Adult Word Cloud for Fear
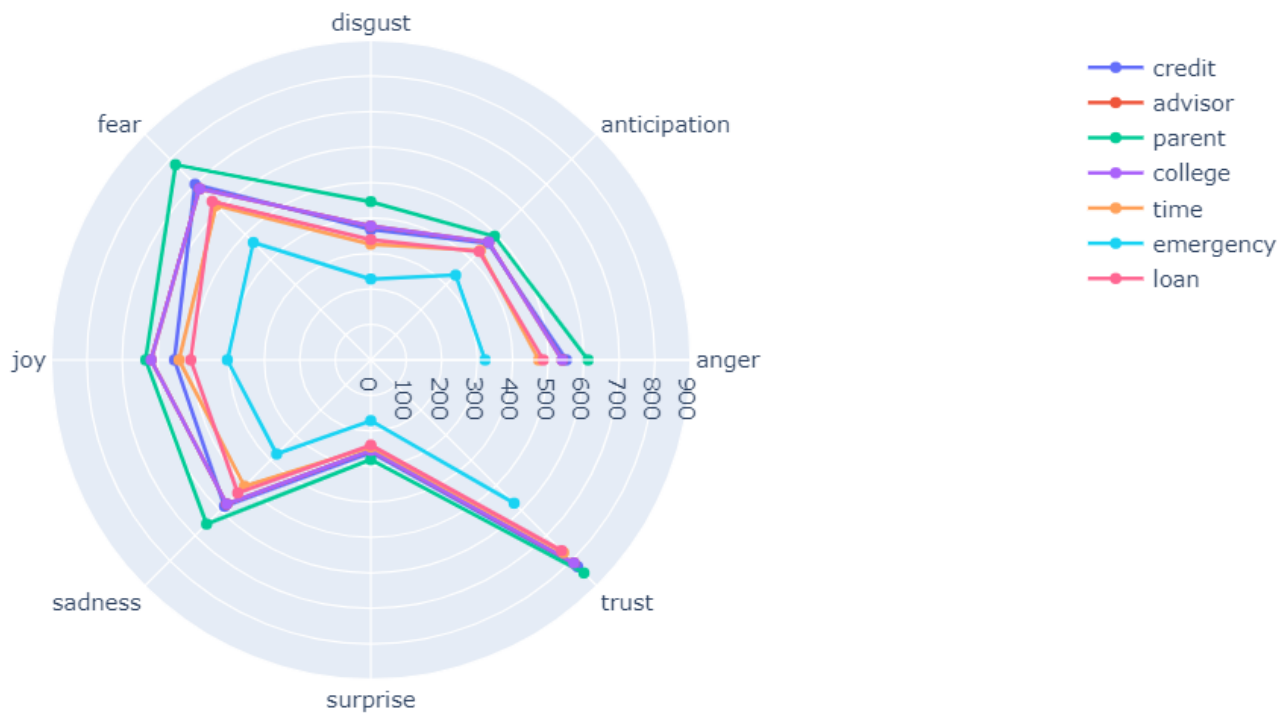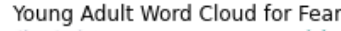
Insights:
- The common trend of sentiment composition is trust > fear > sadness
- A few outliers:
    - **Financial independence** did not follow the general trend and instead had a **more frequent sentiment of joy**, which makes sense because the term financial independence itself suggests a sense of happiness and success.
    - **Pro tips** have a slightly lower sadness than joy which means that the conversations here are likely to be **more joyful** than sad. People here talk a lot about relationships and love on the positive side and about fear of illness and struggles in life on the negative side
    - **Young adults** stand out from the rest of the datasets because they had **fear as the most frequent emotion** rather than trust, which is really rare. It might make sense because many people see adulthood as something that involves a lot of unknown and they become scared of their future. People talked about illness, fight, hanging, crisis here; which contribute to a lot negative sentiment

   IV.   **Topic modeling By Emotion:**
1. **Adulting:** We can see that in different emotions, there are different unique words that are surprising. For instance, in the topics under surprise, most people mention **couples and adulting**. Relating to disgust, people mention **throwing away old stuff**. In topics that are shared by different emotions, we see the **"year old" and " time job" (part/full time job) share both positive and negative feelings**.Topics which are commonly negative are "mental health", "high school","year old". Positive topic is "**save money**".
    a.   Unique topics in each emotion:
        i.   Anger: view poll
        ii.   Trust: bank account, insurance, credit score, credit card
        iii. Surprise: leave house, turned 18, task list, couple days, past couple, talking girl
        iv.   Anticipation: ask questions, workt time, new job
        v.   Joy: friends, family
   vi.   Disgust: throw away, car insurance

b. Shared topics in different emotions:
   i. Sadness & Joy & Fear: Time job, year old
   ii. Fear & Joy: social media
   iii. Sadness & Fear: mental health, high school
   iv. Disgust & Fear: year old
   v. Trust & Anticipation: Save money

2. **Finance:** Finance subreddits include: r/personalfinance, r/financialindependence, r/explainlikeimfive. We used topic modeling under each of the sentiments, and used n-grams to confirm the result.
   a. Unique topics in each emotion:
      i. Anger: Taxable **Brokerage Account** ; **Insurance company** and credit card account service ; **Covid hit** and "past year"(mentioned with sentiment words: hit, aggressively, terrible, frustrated)
      ii. Trust: **Health Insurance; emergency and index funds** (mentioned with sentiment words: advice, Good, real, planning)
      iii. Joy: **Buy houses** and real estate; **Company match**; Debts free and mutual fund (mentioned with sentiment words: positive, super, appreciated, reward)
      iv. Fear: **stock market** and new investment (mentioned with sentiment words : fire, ill, scared, insane and aggressive)
      v. Sadness: losing jobs, losing money and **social security** ; collection agencies and **credit reports** (mentioned with sentiment words : loss, lost, pandemic, worried, missing)
   b. Shared topics in different emotions:
      i. Anticipation & Trust & Joy & Fear : Retirement accounts and **Roth IRA**
      ii. Fear & Disgust & Sadness: Credit card debts and Car loans
      iii. Sadness & Fear & Trust: Student Loans
      iv. Disgust & Surprise: mortgage companies and **mortgage rate** (mentioned with sentiment word: gross, shit, waste)
      v. Trust & Anticipation & Joy: Save money
   vi. Sadness & Trust: Collection agencies and credit reports

Appendix

Takeaway:
- When we look at all of the multiple sentiment graphs, the general trend is the sentiment of "trust" appearing the most frequently, followed by fear, then joy, and with surprise being the least frequent sentiment.
- Also, if we sort the sentiment by keywords, there is the most sentiment associated with the "parent" keyword and followed by the "time" and "loan" keywords.
- There are some interesting outliers that does not follow the general trend:
    - The "loan" and "parent" and "credit" keywords had "sadness" as the third most frequent sentiment (the general trend is that "joy" is the third most frequent sentiment instead)
    - The "Personal Finance" dataset also had "sadness" as the third most frequent sentiment
    - The "Financial independence" dataset had "joy" as the second most frequent sentiment
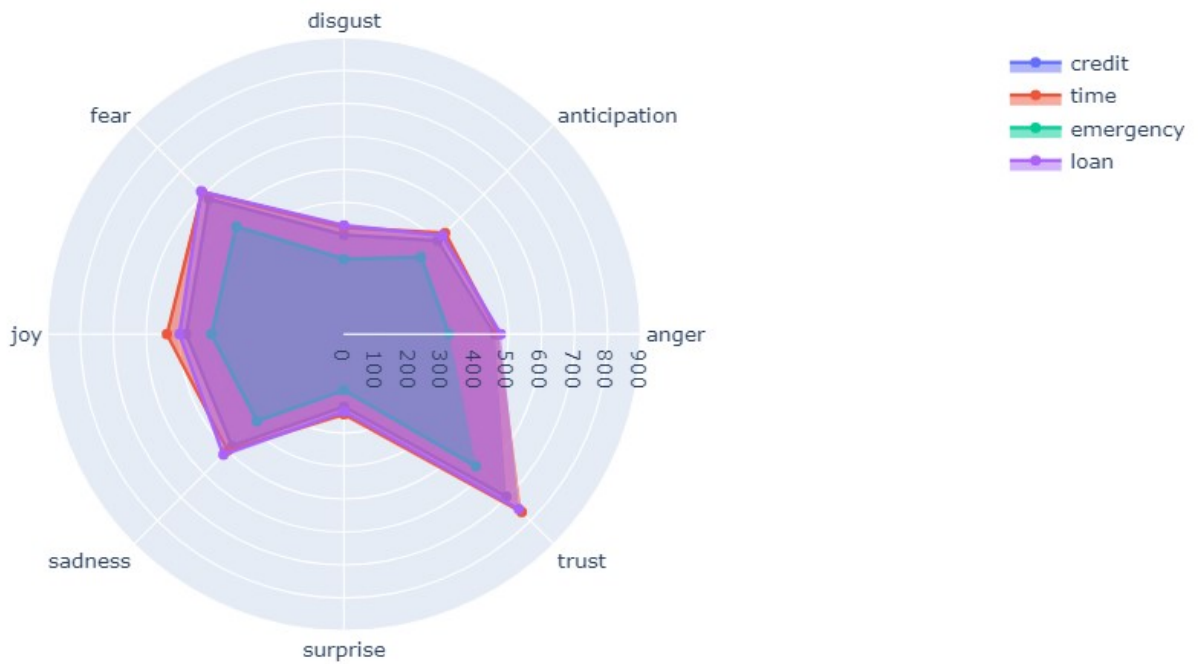
Sentiment by keyword bar plots:

keyword: advisor

keyword: credit

keyword: credit
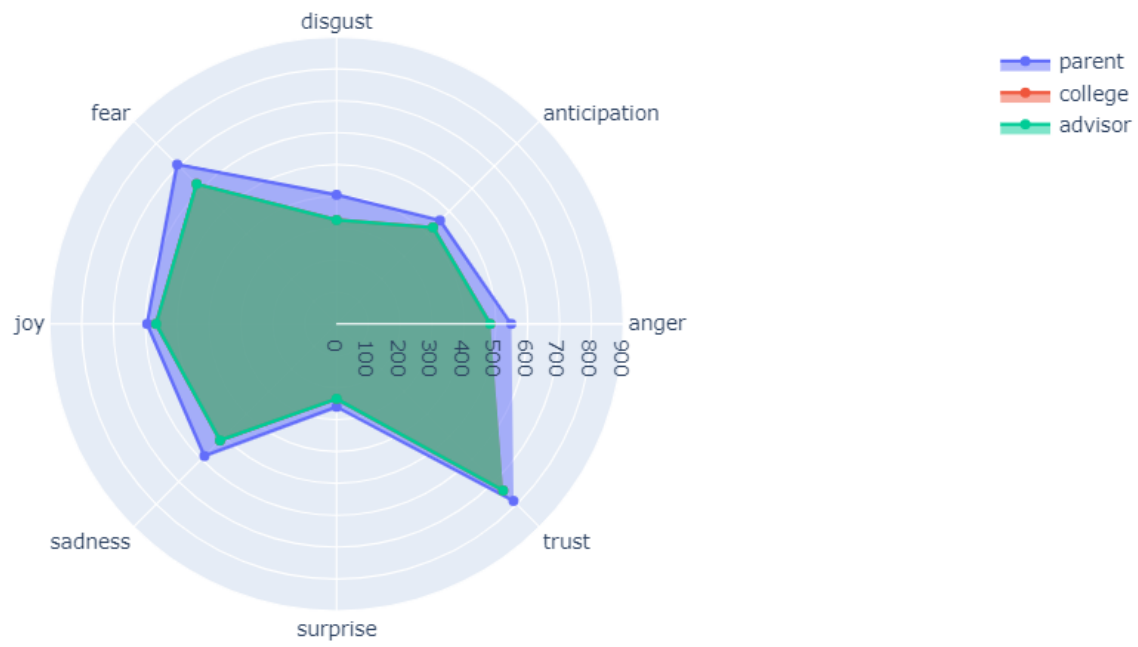
keyword: social

keyword: time
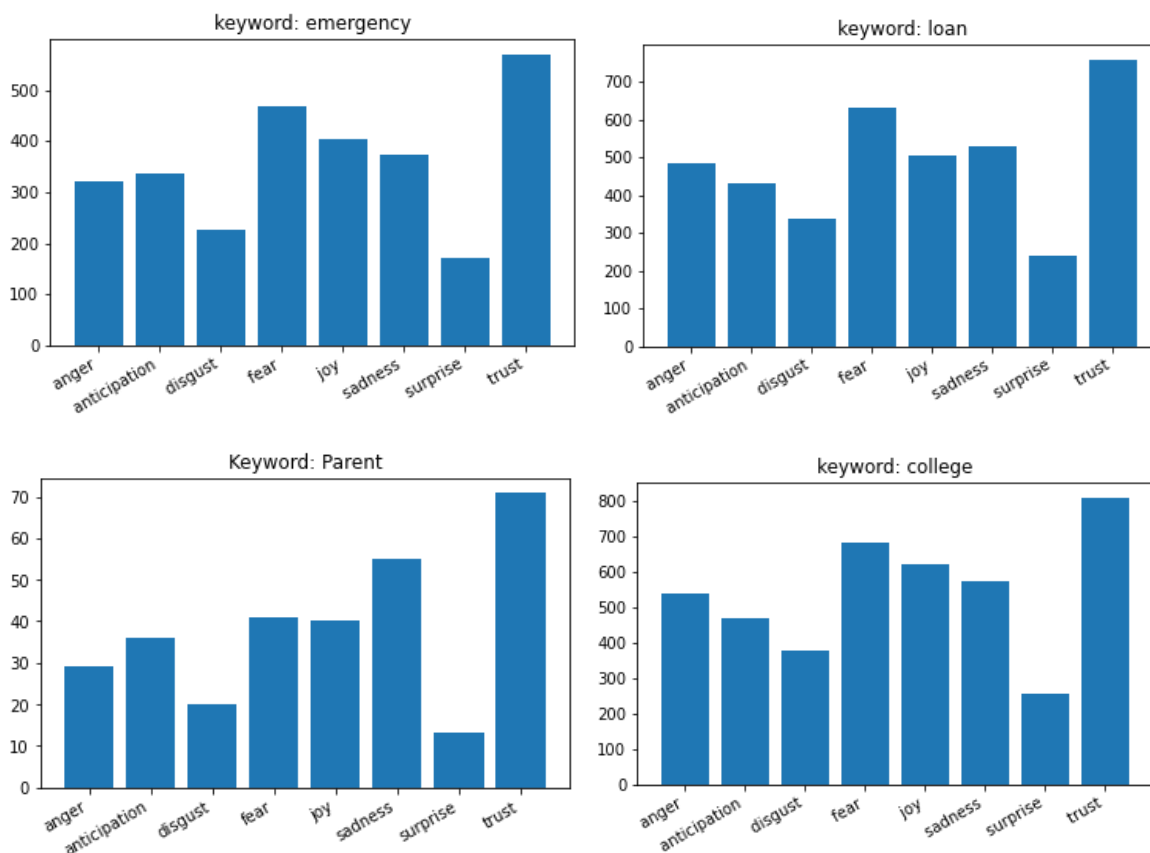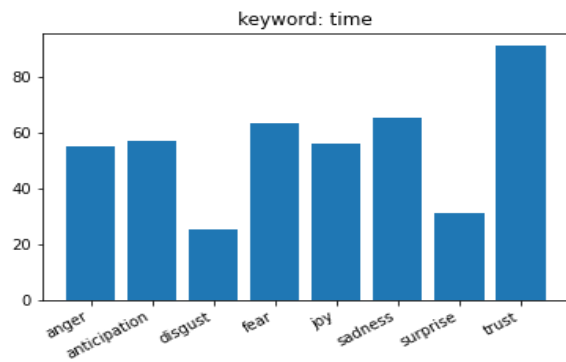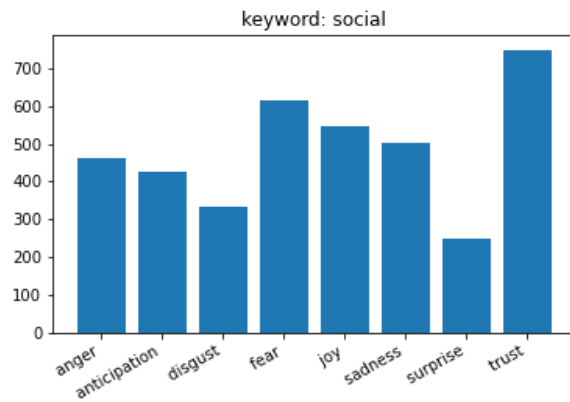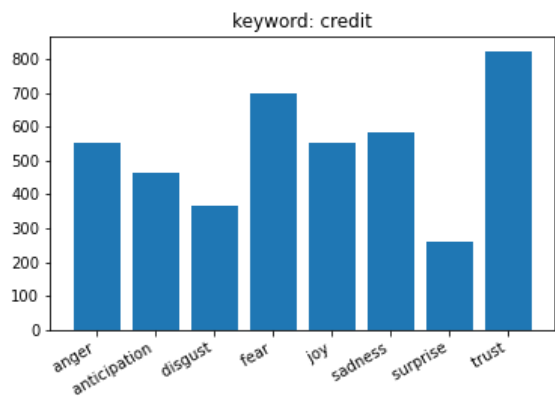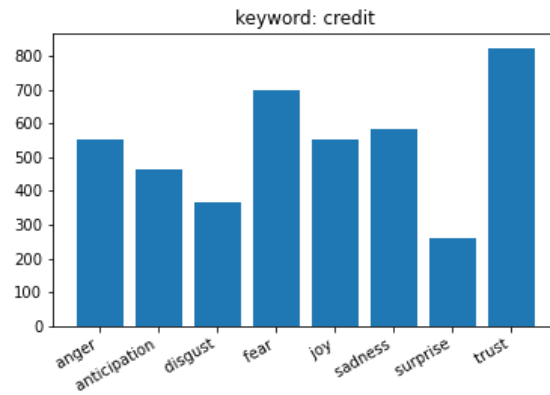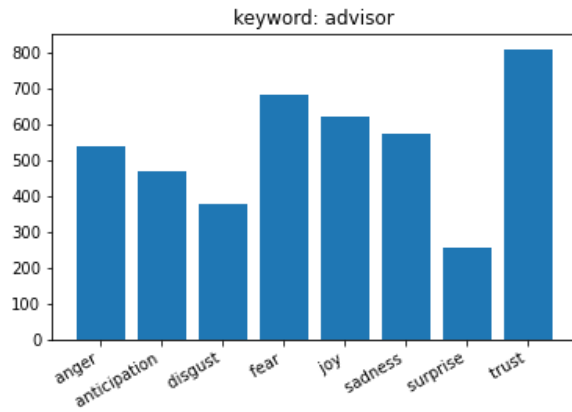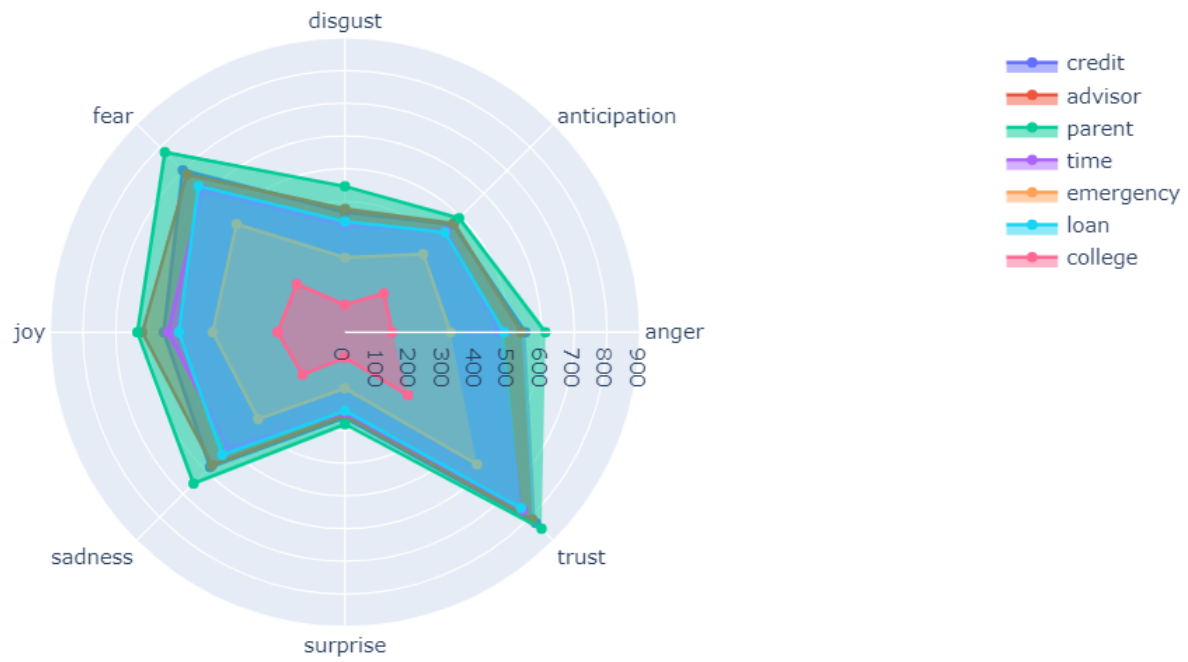
Rescaled graphs: (same sample size)

Sentiment barplot by subreddit: (old method where we counted every sentiment, we now only count sentiment with decent intensity (greater than 0.6))