Key Factors Driving Negative Sentiment Among Students in U.S. College Subreddits and Variations Across Colleges

Doron Czarny, Yella Diekmann, Claudia Schmidt

TABLE OF CONTENTS
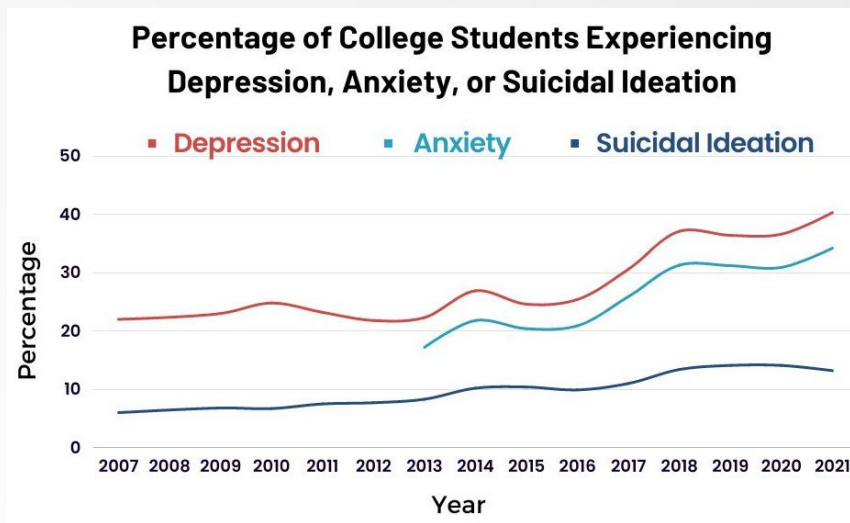
Motivation

- Increasing mental health concerns among U.S. college students due to factors like academic stress, financial struggles, and isolation.

- Existing sentiment analysis methods fail to identify specific drivers of negativity, limiting universities' ability to address root causes effectively.

- Sentiment analysis and actionable insights can enhance student support systems and mental health interventions.

- Reddit serves as an open and real-time platform for students to share their experiences, providing valuable, unfiltered data on student sentiment.
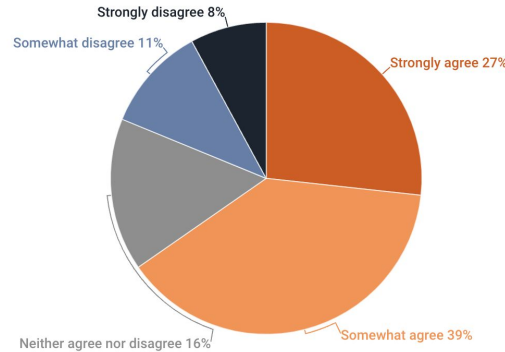

Source: Mindlyte, Data: Healthy Minds survey (2007-2021) +850,000 Survey participants

Importance

- Helps universities understand key drivers of student negativity, enabling them to design targeted strategies for student well-being.

- Provides insights into variations in stress drivers across colleges, highlighting disparities in student experiences based on factors like ranking, climate, or funding.

- Supports broader initiatives in improving mental health outcomes in higher education, contributing to a healthier academic environment.
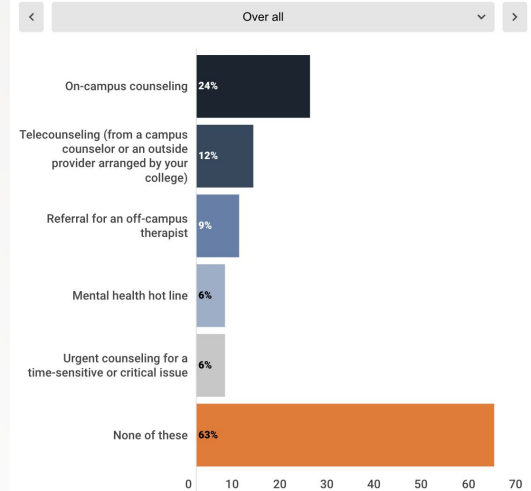
Crisis Resource Awareness

Students say how much they agree or disagree with the following (by overall share): "I know where I can seek help on campus if I or a friend is experiencing a mental health crisis."

Strongly disagree 8%
Somewhat disagree 11%
Strongly agree 27%
Somewhat agree 39%
Neither agree nor disagree 16%

Source: Inside Higher Ed/College Pulse survey of 3,000 college students.

Student Use of Campus Mental Health Resources

Share of students who say they've used the following campus-based services, over all and by self-rating of mental health

Over all

Service	Percentage
On-campus counseling	24%
Telecounseling (from a campus counselor or an outside provider arranged by your college)	12%
Referral for an off-campus therapist	9%
Mental health hot line	6%
Urgent counseling for a time-sensitive or critical issue	6%
None of these	63%

0 10 20 30 40 50 60 70

Data

Composition

- Kaggle
- Reddit posts from 128 selected universities
- Only posts from 2019

Diversity

- Private and public institutions
- Diversity of geographic locations
- Wide range of student demographics (accounts for academic, financial, social, regional, and weather related factors)

Initial Methodology Exploration

1

Attention Weights

Extracted attention weights (calculated in last layer of RoBERTa model) to identify words or phrases contributing most to the model's prediction of negative sentiment.

Tried to cluster all words. However, this was ineffective as many words were falsely added to clusters. Additionally, words that were not adjectives, nouns, or verbs did not provide meaningful semantic content.

2

POS Tagging

Using POS tagging to filter for nouns and verbs as those were most helpful in capturing the themes we were looking for.

Only top 5% of nouns and verbs were retained for clustering. This allowed us to focus on the most relevant terms.

3

Grouping

Grouped key words and phrases by applying a clustering algorithm.

Each cluster represented a theme such as academic stress or financial concerns.

Used a ThemeAnalyzer class that created matrices from the raw data and performed clustering on the attention patterns.

Initial Methodology Exploration

```
Cluster Analysis Results:

Cluster 9:
Size: 39138 posts (98.4%)
Top terms: financial, family, expensive, panic, path, choice, price, direction, pressure, payment
Sample posts: ['I have a similar problem', 'They already lost the referendum', 'Your housing and dining costs will be free unless you quit during the training week or you violate policy/break rules and lose you
-----------------------------------------------------
Cluster 0:
Size: 153 posts (0.4%)
Top terms: class, friend, course, major, food, research, home, community, handle, safety
Sample posts: ['This^ :(. In a lot of my classes attendance is mandatory or missing the lecture will set me significantly behind (Classes where the professor doesn't have a topic schedule or upload notes himsel
-----------------------------------------------------
Cluster 1:
Size: 323 posts (0.8%)
Top terms: time, campus, work, aid, test, social, fail, plan, stress, cost
Sample posts: ['I always wondered if it was a student or a full-time job for some "mascot professional"?  \n  \nMy current student job I\'ve been at for 2-ish years now, and I know 2 people out of the 15+ that
-----------------------------------------------------
Cluster 5:
Size: 20 posts (0.1%)
Top terms: professor, final, worry, expensive, depression, dining, direction, dissertation, dollar, dorm
Sample posts: ['Professor boomer, what?', 'Protestors and Journalists Arrested and Charged With Felony Rioting: [One](https://www.theguardian.com/media/2017/jan/24/journalists-charged-felonies-trump-inaugurati
-----------------------------------------------------
Cluster 2:
Size: 82 posts (0.2%)
Top terms: job, money, party, housing, graduate, future, grade, library, budget, financial
Sample posts: ['All those tuition and you still have "budgetary issues"\n\nMight want to look at your payroll, UW.\nThere might be a few fat cats around.', 'All that fucking hard work for nothing. The overall e
-----------------------------------------------------
Cluster 6:
Size: 11 posts (0.0%)
Top terms: study, interview, worry, experience, depression, dining, direction, dissertation, dollar, dorm
Sample posts: ['&gt; Turns out he doesn't actually attend any church, and his preaching stems from self-study\n\nAsk him next time if he's running a compound out in the woods somewhere for his followers.", "Thi
-----------------------------------------------------
Cluster 4:
Size: 7 posts (0.0%)
Top terms: company, worry, decision, dining, direction, dissertation, dollar, dorm, employer, exam
Sample posts: ['Bruh, gender and race are not equal to sexuality. I\'m aware that being female/a URM makes you a more appealing candidate to some companies, but most people don\'t give a fuck if you\'re queer o
-----------------------------------------------------
Cluster 7:
Size: 6 posts (0.0%)
Top terms: mental, worry, gym, depression, dining, direction, dissertation, dollar, dorm, employer
Sample posts: ['Would like to see the source. Heartbreaking – especially including the non-mental health related deaths, it seems excessive.', 'And I gave you an option to chill and reword you question without
-----------------------------------------------------
Cluster 3:
Size: 35 posts (0.1%)
Top terms: cost, plan, rent, safety, sleep, exhausted, decision, depression, dining, direction
Sample posts: ["Lol, I just checked MyPlan and found out that I got W credit for my paper even though I never asked. Literally turned in the draft a day late and my final paper was trash. I'll take it tho", 'Do
-----------------------------------------------------
Cluster 8:
Size: 4 posts (0.0%)
Top terms: stress, worry, experience, depression, dining, direction, dissertation, dollar, dorm, employer
Sample posts: ["I had this happening to me last year. The kicker there is that I got into their mental health program because I was in such bad condition, they put me in triage and tried to inpatient me during
-----------------------------------------------------
```
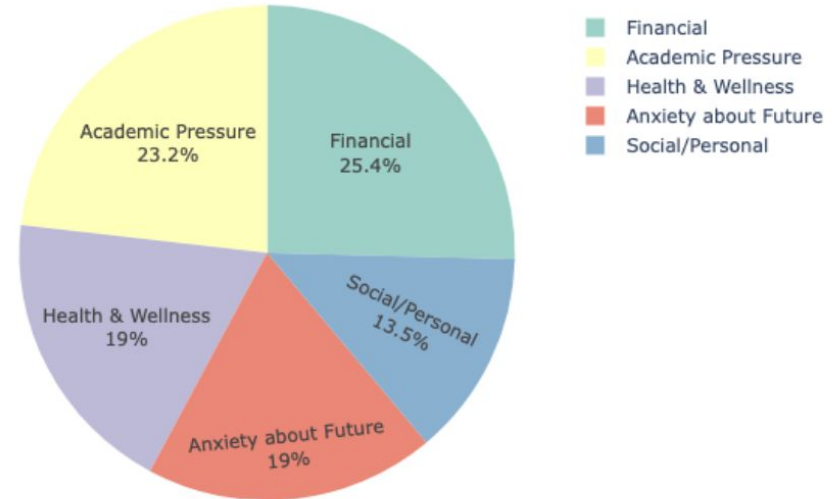
Experimental Design

- Data Collection: Analyzed 128 university subreddits with sentiment-labeled posts (neg. threshold >0.6)

- Theme Classification: Developed 5 predefined categories with 20-25 validated seed words each
  - Academic Pressure, Future Anxiety, Social/Personal, Financial, Health & Wellness

- Text Processing: Used spaCy for NLP, focusing on verb/noun phrases and pattern matching

- Theme Assignment: Applied TF-IDF vectorization with cosine similarity matching

- Quality Control: Implemented filtering criteria

Experimental Results

- Examined 147,580 posts from 128 university subreddits
- After filtering and strict quality control successfully classified 638 posts (0.43% of total)
- Strict filtering ensures high-quality theme assignments, though at the cost of reduced data volume.
- Classified posts came from 76 different universities, with varying levels of representation.

Distribution of Themes Across All Schools

Legend:
- Financial
- Academic Pressure
- Health & Wellness
- Anxiety about Future
- Social/Personal

Financial 25.4%
Academic Pressure 23.2%
Health & Wellness 19%
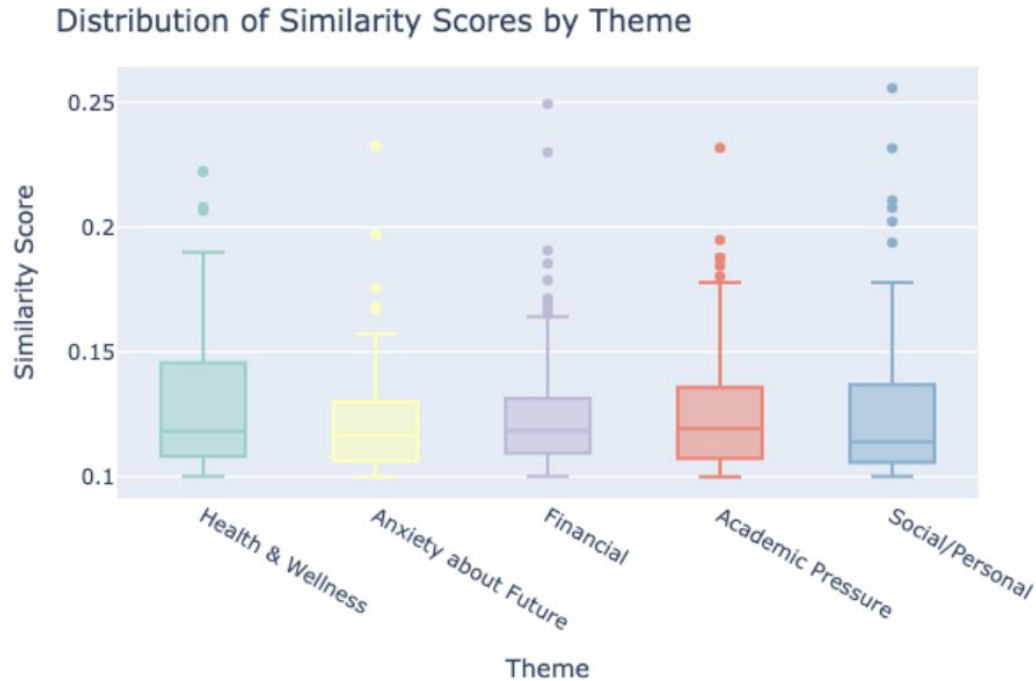Anxiety about Future 19%
Social/Personal 13.5%

Examples

Financials

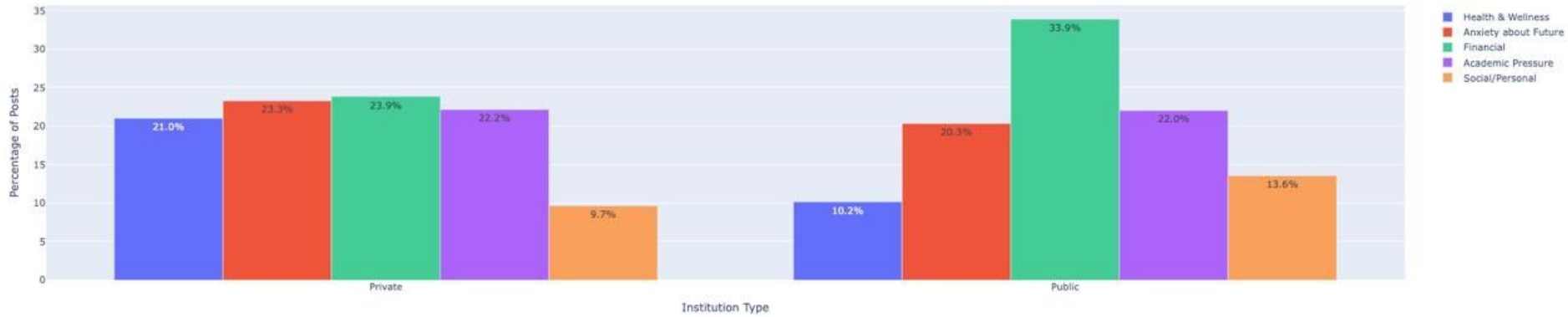"Just checked and I messed up. The $18k was tuition AND housing..."

Academics

"I retook 103, just felt I didn't really learn anything from the class the first time and ended up failing out the first time. Retook it again and still got [a low grade]..."

Experimental Results

Distribution of Similarity Scores by Theme

Experimental Results

Theme Distribution: Private vs Public Institutions

Legend:
- Health & Wellness
- Anxiety about Future
- Financial
- Academic Pressure
- Social/Personal

Private: 21.0%, 23.3%, 23.9%, 22.2%, 9.7%

Public: 10.2%, 20.3%, 33.9%, 22.0%, 13.6%

Y-axis: Percentage of Posts

X-axis: Institution Type

Limitations

- Data Constraints: Only 0.43% of posts analyzed due to strict filtering. Imbalanced university representation.
- Methodology Limitations: Predefined themes may not be exhaustive. TF-IDF threshold potentially excludes nuanced posts. Posts limited to a single category.
- Platform Bias: Self-selection bias in negative posts.
- Relative Frequency Analysis: Higher frequency may not indicate proportionally higher sentiment. No baseline comparison.

Possible Extensions

Top 20 Colleges vs. Bottom 20

- Compare the top 20 ranked schools and the bottom 20 ranked schools as determined by U.S. News rankings.
- Uncover meaningful trends in stress causes between different ranked institutions

Warm Climate vs. Cold Climate

- Geographical segmentation to investigate whether climate and regional differences influence the causes of negativity
- How environmental and regional factors might shape student experiences and emotional well-being

Small vs. Large Student Body

- Expect causes of negativity to be different between differently sized student bodies
- Enable institutions to tailor interventions to their specific contexts more effectively.

Conclusion

- Rising stress levels in college highlight the importance of providing qualitative counseling services

- Understanding causes of stress allows institutions to take a tailored approach to combating stress levels

- Our approach might allow for colleges to better understand main reasons for student stress for their individual use cases

- Findings from analyzing reddit datasets support thesis that different institutions experience different themes of stress more commonly → need for further exploration