Fannie Mae Climate Change Sentiment & Topic Analysis

Amanda Lee, Fahmina Salma, Wendy Deng
Technica 2023

Fannie Mae Challenge:

Given a climate change dataset/reddit post and comment history over the past several years.

- Calculate monthly average sentiment and chart these values
- Generate topic clusters from the post and comment text data
- Visualize the data through a comparison temporal chart or other means. Were the sentiment values more diverse or extreme in the subpopulation?

Kaggle Reddit Climate Change Dataset

```
# take a look at data
comments.head()
```

	type	id	subreddit.id	subreddit.name	subreddit.nsfw	created_utc	permalink	body	sentiment	score	time	month
0	comment	imlddn9	2qh3l	news	False	1661990368	https://old.reddit.com/r/news/comments/x2cszk/...	Yeah but what the above commenter is saying is...	0.5719	2	2022-08-31 23:59:28	8
1	comment	imldbeh	2qn7b	ohio	False	1661990340	https://old.reddit.com/r/Ohio/comments/x2awnp/...	Any comparison of efficiency between solar and...	-0.9877	2	2022-08-31 23:59:00	8
2	comment	imldado	2qhma	newzealand	False	1661990327	https://old.reddit.com/r/newzealand/comments/x...	I'm honestly waiting for climate change and th...	-0.1143	1	2022-08-31 23:58:47	8
3	comment	imld6cb	2qi09	sacramento	False	1661990278	https://old.reddit.com/r/Sacramento/comments/x...	Not just Sacramento. It's actually happening a...	0.0000	4	2022-08-31 23:57:58	8
4	comment	imld0kj	2qh1i	askreddit	False	1661990206	https://old.reddit.com/r/AskReddit/comments/x2...	I think climate change tends to get some peopl...	0.6634	1	2022-08-31 23:56:46	8

```
posts.head()
```

	type	id	subreddit.id	subreddit.name	subreddit.nsfw	created_utc	permalink	domain	url	selftext	title	score
0	post	x2smmw	2vh0v	democraticsocialism	False	1661990237	https://old.reddit.com/r/DemocraticSocialism/c...	twitter.com	https://twitter.com/SusanofTexas/status/156480...	NaN	"Cities without water" is our future. The lead...	15
1	post	x2slxy	2w844	nostupidquestions	False	1661990182	https://old.reddit.com/r/NoStupidQuestions/com...	self.nostupidquestions	NaN	Ok so I was having a conversation with my neig...	Sharks and climate change	2
2	post	x2slk5	6wzx9b	eduwriters	False	1661990150	https://old.reddit.com/r/EduWriters/comments/x...	eduwriters.pro	https://eduwriters.pro/?cid=2860/?utm_source=r...	NaN	How to start a climate change essay	1
3	post	x2sbz6	46adh7	alternatitvereality	False	1661989389	https://old.reddit.com/r/AlternatitveReality/c...	i.imgur.com	https://i.imgur.com/wJbchkQ.jpg	NaN	Ever notice that when it's hot a few days in J...	1
4	post	x2sa2a	2qh1i	askreddit	False	1661989240	https://old.reddit.com/r/AskReddit/comments/x2...	self.askreddit	NaN	[deleted]	How would you react if the next American presi...	1

Steps:

1. **Extracting date and time** from the dataset. The information was encoded in UTC format, and required conversion to Python datetime format.

2. Understanding how to extract entities from **texts** and find out their **topics** using *spaCy* and *n-grams frequency distributions*.

3. **Data visualizations**: When we tried to visualize the climate change data, it is showing in a very close plane. We used plotly to put it in scatter plot.
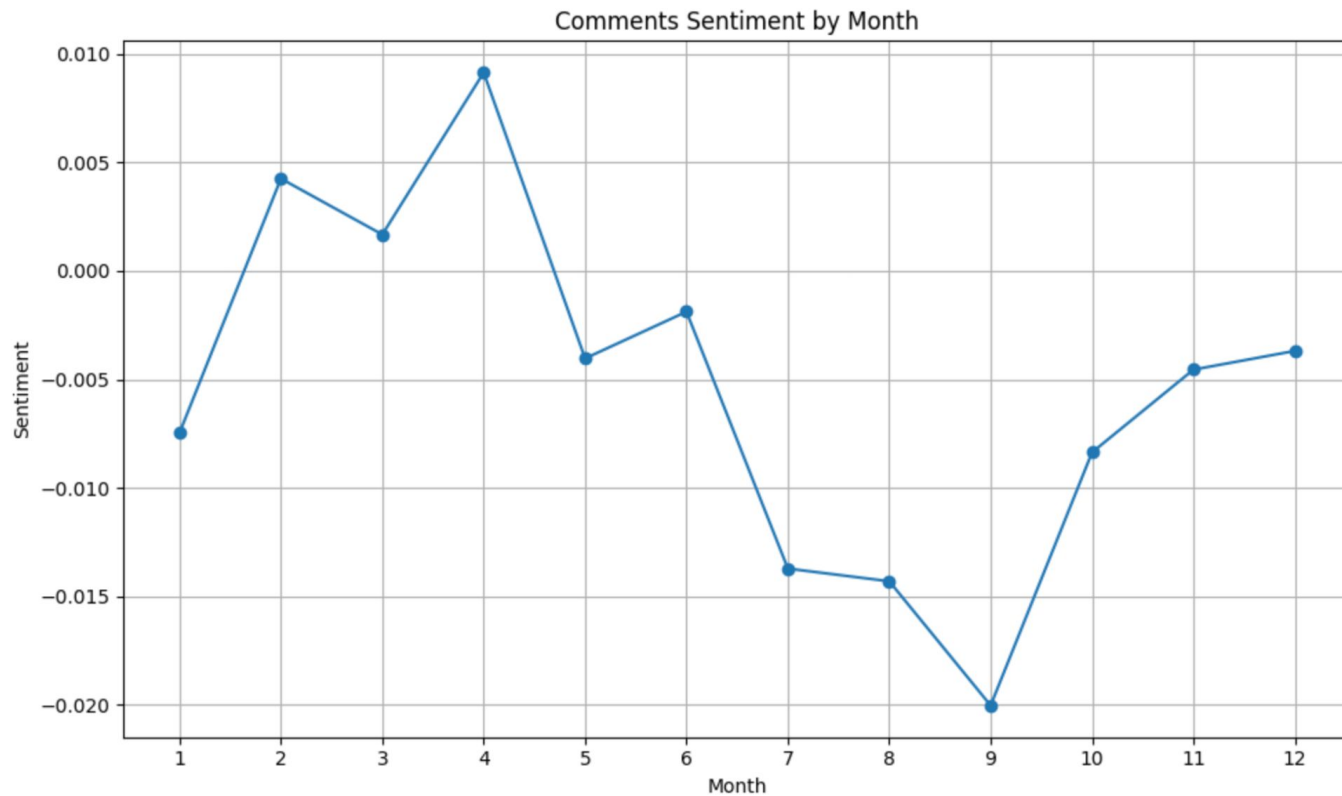
Month Extraction & Sentiment Scores

```python
# take a look at data
comments.head()
```

	type	id	subreddit.id	subreddit.name	subreddit.nsfw	created_utc	permalink	body	sentiment	score	time	month
0	comment	imlddn9	2qh3l	news	False	1661990368	https://old.reddit.com/r/news/comments/x2cszk/...	Yeah but what the above commenter is saying is...	0.5719	2	2022-08-31 23:59:28	8
1	comment	imldbeh	2qn7b	ohio	False	1661990340	https://old.reddit.com/r/Ohio/comments/x2awnp/...	Any comparison of efficiency between solar and...	-0.9877	2	2022-08-31 23:59:00	8
2	comment	imldado	2qhma	newzealand	False	1661990327	https://old.reddit.com/r/newzealand/comments/x...	I'm honestly waiting for climate change and th...	-0.1143	1	2022-08-31 23:58:47	8
3	comment	imld6cb	2qi09	sacramento	False	1661990278	https://old.reddit.com/r/Sacramento/comments/x...	Not just Sacramento. It's actually happening a...	0.0000	4	2022-08-31 23:57:58	8
4	comment	imld0kj	2qh1i	askreddit	False	1661990206	https://old.reddit.com/r/AskReddit/comments/x2...	I think climate change tends to get some peopl...	0.6634	1	2022-08-31 23:56:46	8

```python
posts.head()
```

	type	id	subreddit.id	subreddit.name	subreddit.nsfw	created_utc	permalink	domain	url	selftext	title	score
0	post	x2smmw	2vh0v	democraticsocialism	False	1661990237	https://old.reddit.com/r/DemocraticSocialism/c...	twitter.com	https://twitter.com/SusanofTexas/status/156480...	NaN	"Cities without water" is our future. The lead...	15
1	post	x2slxy	2w844	nostupidquestions	False	1661990182	https://old.reddit.com/r/NoStupidQuestions/com...	self.nostupidquestions	NaN	Ok so I was having a conversation with my neig...	Sharks and climate change	2
2	post	x2slk5	6wzx9b	eduwriters	False	1661990150	https://old.reddit.com/r/EduWriters/comments/x...	eduwriters.pro	https://eduwriters.pro/?cid=2860/?utm_source=r...	NaN	How to start a climate change essay	1
3	post	x2sbz6	46adh7	alternatitvereality	False	1661989389	https://old.reddit.com/r/AlternatitveReality/c...	i.imgur.com	https://i.imgur.com/wJbchkQ.jpg	NaN	Ever notice that when it's hot a few days in J...	1
4	post	x2sa2a	2qh1i	askreddit	False	1661989240	https://old.reddit.com/r/AskReddit/comments/x2...	self.askreddit	NaN	[deleted]	How would you react if the next American presi...	1

Monthly Average Sentiment on Comments

Comments Sentiment by Month

Topic Clustering

```
# take a look at data
comments.head()
```

	type	id	subreddit.id	subreddit.name	subreddit.nsfw	created_utc	permalink	body	sentiment	score	time	month
0	comment	imlddn9	2qh3l	news	False	1661990368	https://old.reddit.com/r/news/comments/x2cszk/...	Yeah but what the above commenter is saying is...	0.5719	2	2022-08-31 23:59:28	8
1	comment	imldbeh	2qn7b	ohio	False	1661990340	https://old.reddit.com/r/Ohio/comments/x2awnp/...	Any comparison of efficiency between solar and...	-0.9877	2	2022-08-31 23:59:00	8
2	comment	imldado	2qhma	newzealand	False	1661990327	https://old.reddit.com/r/newzealand/comments/x...	I'm honestly waiting for climate change and th...	-0.1143	1	2022-08-31 23:58:47	8
3	comment	imld6cb	2qi09	sacramento	False	1661990278	https://old.reddit.com/r/Sacramento/comments/x...	Not just Sacramento. It's actually happening a...	0.0000	4	2022-08-31 23:57:58	8
4	comment	imld0kj	2qh1i	askreddit	False	1661990206	https://old.reddit.com/r/AskReddit/comments/x2...	I think climate change tends to get some peopl...	0.6634	1	2022-08-31 23:56:46	8

```
posts.head()
```

	type	id	subreddit.id	subreddit.name	subreddit.nsfw	created_utc	permalink	domain	url	selftext	title	score
0	post	x2smmw	2vh0v	democraticsocialism	False	1661990237	https://old.reddit.com/r/DemocraticSocialism/c...	twitter.com	https://twitter.com/SusanofTexas/status/156480...	NaN	"Cities without water" is our future. The lead...	15
1	post	x2slxy	2w844	nostupidquestions	False	1661990182	https://old.reddit.com/r/NoStupidQuestions/com...	self.nostupidquestions	NaN	Ok so I was having a conversation with my neig...	Sharks and climate change	2
2	post	x2slk5	6wzx9b	eduwriters	False	1661990150	https://old.reddit.com/r/EduWriters/comments/x...	eduwriters.pro	https://eduwriters.pro/?cid=2860/?utm_source=r...	NaN	How to start a climate change essay	1
3	post	x2sbz6	46adh7	alternatitvereality	False	1661989389	https://old.reddit.com/r/AlternatitveReality/c...	i.imgur.com	https://i.imgur.com/wJbchkQ.jpg	NaN	Ever notice that when it's hot a few days in J...	1
4	post	x2sa2a	2qh1i	askreddit	False	1661989240	https://old.reddit.com/r/AskReddit/comments/x2...	self.askreddit	NaN	[deleted]	How would you react if the next American presi...	1

Deeper Look

```
newzealand_subset['body'].tolist()[0]
```

[96… "I'm honestly waiting for climate change and the impacts of that to kick some fucking sense into people. But who am I kidding itll still just be more of the poor suffering while the rich claim victim hood for handouts while letting us all starve. Its honestly hard some days to not just give up, and I truly wonder if and when anything will ever actually be done."
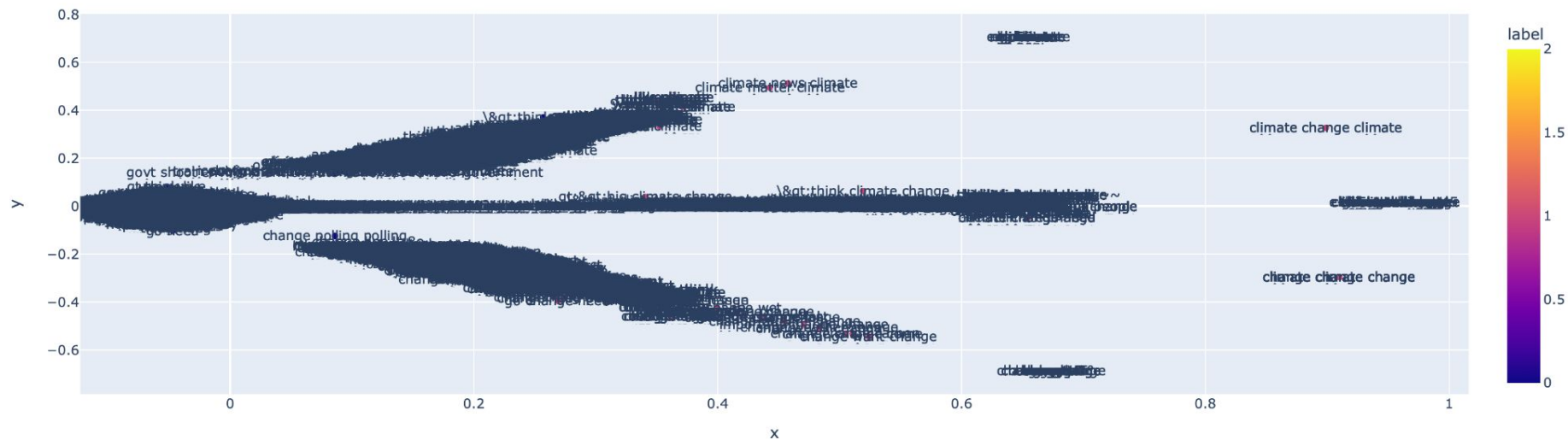
```
newzealand_posts_subset['title'].tolist()[0]
```
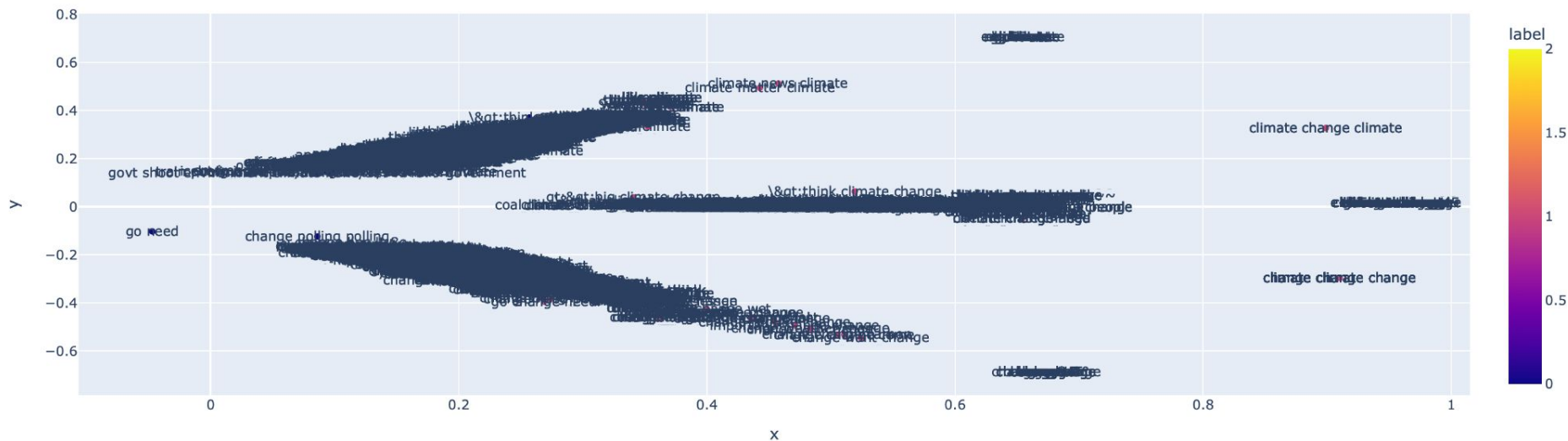
[10… 'Climate change and insurance: Weighing up the risk'

Frequent Topics by Clusters

Clustered Topics Visualization

Cleaned Frequent Topics by Clusters

Clustered Topics Visualization

Conclusion:

Overall, we were happy to have worked together and delved deep into the dataset. Thank you to Fannie Mae and Technica for this challenge!

Within less than 24 hours, we were able to:

- Leverage NLP and data cleaning to gain more insight into the data
- Experience first-hand the hurdles that come with solving real world problem and how to blend our skills to solve them

Future Considerations

- Graph data — Neo4j:
  - Data can be converted to graph data to create clusters

- Improved sentiment analysis:
  - Compute our own sentiment values on comments and perform the same analysis

- Build pipeline & scale up analysis:
  - Right now, we subsetted the original 20GB data to achieve the results
  - Moving on, we would build a pipeline to perform our analysis on the entire dataset with GPU