TEAM REDWARN

(Reddit Data for Early Warning and Response to Pandemics)

Our project is a culmination of multiple Analysis Tests on data gathered from the subreddit r/Austin. With these tests we have been able to formulate results that help community leaders predict how this community might react to similar situations in the future. Which in turn will hopefully provide a more peaceful environment to the community.

team REDWARN

Xinyi Miao

Evans Etrue Howard

Coreen Mullen

Mentors

- Emily Javan
- Oluwasegun Ibrahim
- Lydia Fletcher

Ahmad Samyono

Qimora Mason

Tasks

1. Topic Modeling
2. Time Series Analysis
3. Sentiment Analysis
4. Distribution of Posts over Time
5. Temporal Analysis

With time, our team was able to to come up with our Minimum Lovable Product, and we have been able to achieve that.

During this project our team has completed the code for the aforementioned tasks excluding Temporal Analysis, which we were close to finishing but was not needed to reach our Minimum Lovable Product.

Time Series Analysis (Task #2)

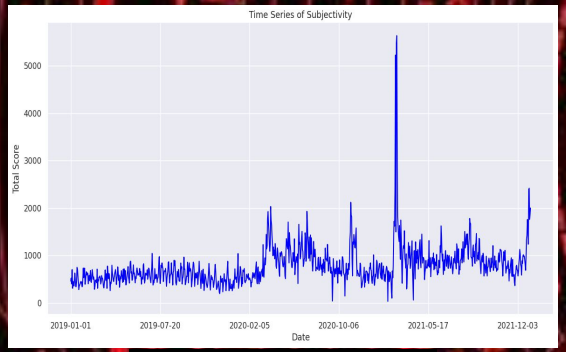


Fig.1 Time Series Analysis

From this plot we can see that the sentiment score of posts increased as time went on.

Post Distribution

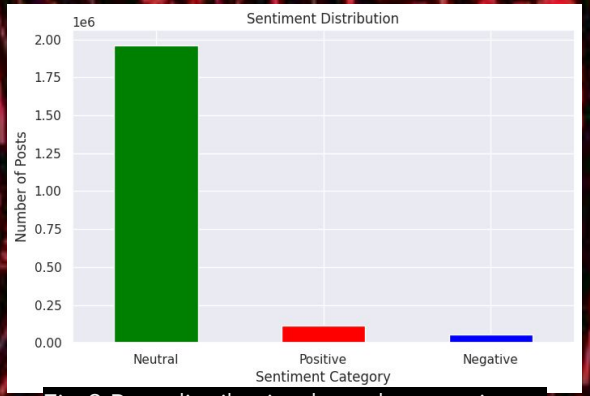


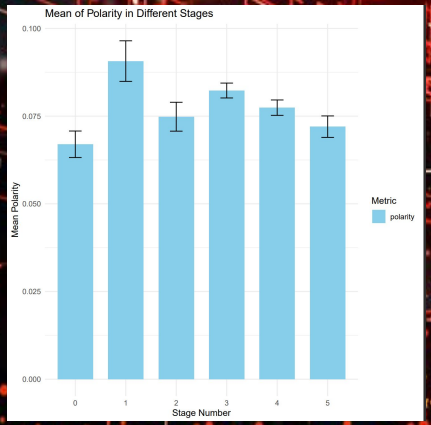
Fig.2 Post distribution based on sentiment

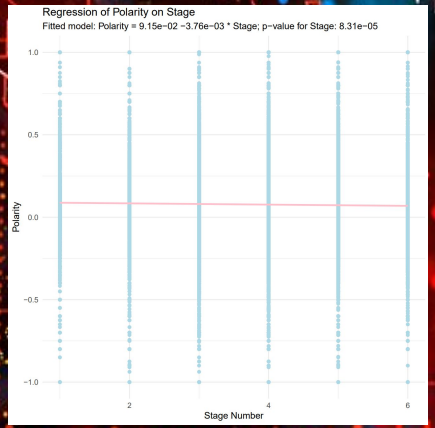
Our findings show that the vast majority of posts have a neutral sentiment.

Semantic Analysis

Example of our Data.

	X	Date	author	author_fla	author_fla	body	id	parent_id	score	subreddit	Segment	Analytic	Tone	pronoun	ppron	i	we	you
767429	767429	4/21/2020	808adw			I just did t	fo483tt	t1_fo3hgk	1	Austin	1	13.76	57.46	16.67	14.29	9.52	0	0





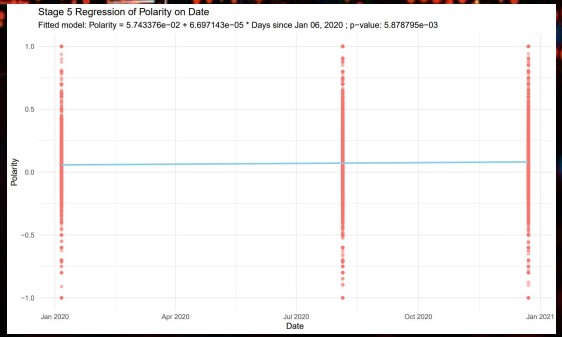


Fig 1. Positive sentiment word cloud.

Fig 2. Negative sentiment word cloud.

Topic Modeling

- Topic Modeling is a technique that extracts the theme/topics from data.

Technology Used

- Google Colab, to be able to work on our projects code as a team and view the results of each others work.

- TACC, we used TACC to handle the processing for our dataset use to the size of the dataset.

- Python, Our script for doing all of our analysis tests was done using the Python programming language.

- R, we used R to do principal component analysis and linear regression on our data.

REDWARN

(Reddit Data for Early Warning and Response to Pandemics)

Project Impact on The Community

The Impact our project intends to bring for the community of Austin, TX is an opportunity for leaders in the community to have an example of how people will react to these big events. It will not only be important information to the leaders but the community will have a better experience in these situations because the leaders know their sentiment reaction.