

Haas and Farmer School of Business Part 2

Chris Lee^{1,*}

¹Market Games, Berkeley California

*chris@marketgames.io

ABSTRACT

I scraped the comments of posts for farmer and haas. I found that comments are used often to show support for movements and people, students strongly support equity and inclusions, and there are students with larger than average influence.

Introduction

This is a follow up in response to the last report. The goal of this project was to better understand our audience. It was brought up that we may be able to do this by scraping the comments of each post. From this we can see how communities are interacting with each other.

Methods

Comment Scraper

The comment scraper code collects all comments from a post. It was adapted from user AgiMaulana under the open MIT license. It features functionality to get usernames and the associated comments. The dependencies for this are the same as the previous program. Here we also rely on css selectors instead of xpath as I could not figure out the xpath. This seemed to be the work around I was looking for.

```
def get_comments(driver):
    try:
        load_more_comment = \
            driver.find_element_by_css_selector('.MGdpg >
            button:nth-child(1)')
        print("Found {}".format(str(load_more_comment)))
        i = 0
        while load_more_comment.is_displayed() and i < int(sys.argv[2]):
            load_more_comment.click()
            time.sleep(1.5)
            load_more_comment = \
                driver.find_element_by_css_selector('.MGdpg >
                button:nth-child(1)')
            print("Found {}".format(str(load_more_comment)))
            i += 1
    except Exception as e:
        print(e)
        pass

user_names = []
user_comments = []
comment = driver.find_elements_by_class_name('gElp9 ')
for c in comment:
    container = c.find_element_by_class_name('C4VMK')
    name = container.find_element_by_class_name('_6lAjh').text
    content = container.find_element_by_tag_name('span').text
    content = content.replace('\n', ' ').strip().rstrip()
    user_names.append(name)
```

```

        user_comments.append(content)
        # debug output
        print(name)
        print(content)
        print()

# not using names ...
# deal with no comment case
try:
    user_names.pop(0)
    user_comments.pop(0)
except Exception as e:
    print(e)
    pass

return(user_comments)

```

Definitions

- Outlier: In this case we only have upper outliers. We use the common formula for computing outliers of a normally distributed population.

$$\text{Upper Outlier} = Q3 + (1.5 \cdot \text{IQR})$$

- Note: Any characters that are unrecognizable are emojis. This is due to difficulty with encoding when moving from HTML to the csv file. Unfortunately I can not do anything about this at the moment. After spending some time trying to fix this I came to the conclusion that this would take more time than I am willing to spend to fix this.

Results

Because of the issues with encoding I have excluding emojis in the results.

Common Topics

Here we remove stop words (common words such as grammar) and aggregate across all posts. We then divide by total posts to get the average times the word occurs per post.

	word	n	avg_per_post
1	love	64	0.106666667
2	congrats	56	0.093333333
3	proud	47	0.078333333
4	congratulations	38	0.063333333
5	haas	36	0.060000000
6	bears	35	0.058333333
7	amazing	32	0.053333333
8	berkeleyhaas	22	0.036666667
9	students	22	0.036666667
10	yay	21	0.035000000
11	queen	19	0.031666667
12	awesome	16	0.026666667
13	class	16	0.026666667
14	wow	16	0.026666667
15	business	15	0.025000000

Figure 1. With emojis removed this is the most commonly used words by follows of both haas and farmer

	word	n	avg_per_post
1	ðŷ	1155	1.925000000
2	ĩ	173	0.288333333
3	â	167	0.278333333
4	œðŷ	83	0.138333333
5	love	64	0.106666667
6	congrats	56	0.093333333
7	proud	47	0.078333333
8	congratulations	38	0.063333333
9	haas	36	0.060000000
10	bears	35	0.058333333
11	amazing	32	0.053333333
12	thatâ	31	0.051666667
13	œ	25	0.041666667
14	berkeleyhaas	22	0.036666667
15	students	22	0.036666667

Figure 2. This is the same chart with the emojis kept

Relationship to Likes

Using the formula for finding outliers that is listed above I found the posts that have a large amount of likes. Then I used the same procedure to see that followers were commenting.

	word	n	avg_per_post
1	percent	14	0.31818182
2	luke_sassano	8	0.18181818
3	police	8	0.18181818
4	bears	6	0.13636364
5	blacks	6	0.13636364
6	black	5	0.11363636
7	congrats	5	0.11363636
8	haas	5	0.11363636
9	macdonald	5	0.11363636
10	shootings	5	0.11363636
11	account	4	0.09090909
12	alexmdavenport	4	0.09090909
13	cinnamon	4	0.09090909
14	congratulations	4	0.09090909
15	dean	4	0.09090909
16	population	4	0.09090909

Figure 3. Without emojis common words in outlier posts

	word	n	avg_per_post
1	ðŸ	59	1.34090909
2	â	15	0.34090909
3	ĩ	15	0.34090909
4	percent	14	0.31818182
5	luke_sassano	8	0.18181818
6	police	8	0.18181818
7	bears	6	0.13636364
8	blacks	6	0.13636364
9	âœ	5	0.11363636
10	black	5	0.11363636
11	congrats	5	0.11363636
12	haas	5	0.11363636
13	macdonald	5	0.11363636
14	shootings	5	0.11363636
15	account	4	0.09090909
16	alexmdavenport	4	0.09090909

Figure 4. With emojis common words in outlier posts

Discussion

The first thing to notice is the prevalence of congrats and congratulation in the overall list. Combining their counts (because they mean the same thing) we can see that it is the most common. This is evidence of what can be inferred by skimming the posts. This is that a large chunk of engagement in the form of comments is congratulating other members of the community (most of whom the commenter seems to personally know). Abstracting a little further and looking at the list as a whole it seems that comments are most often used to show appreciation for a person and/or announcement (both insta frequently use the platform to provide information to students about various topics). For the outliers I have two thoughts. The first is the prevalence of police, black, and shootings. This mirrors the strong outpouring of support and engagement of the students with social justice initiatives. Because of this we might consider some way to show support for equity and inclusion in the classroom. Perhaps a blog post could share how our program supports equity and inclusion in the classroom. The support for this movement may provide opportunity from a marketing standpoint (especially in wake of the recent election, regardless of outcome). Secondly in the outliers we see multiple instances of specific people. While some are expected figures such as the dean (who passed away), others are apparently just popular students. What this seems to suggest is that there are community members, some of them students, that have a disproportional involvement and standing in the community. If we can identify these individuals they would likely make great people for the spotlights. Finally, this is just a fun bit, but farmer seems to have a strange love of cinnamon rolls.