SOCI 40133 - Final Project

Social Media, Misogyny, and Labor Market Outcomes

Qiuyu (Jora) Li

March 18, 2022

GitHub: `https://github.com/Qiuyu-Li/soci40133-final-media-bias`[1]

1 Introduction

My project focuses on the relationship between misogyny displayed on social media and gender-based discrimination in the labor market. Specifically, I plan to use the Twitter data to construct a county-level measure of prevailing misogyny, and examine its implications on the gaps in women's earnings or career opportunities compared to otherwise identical men. The model of Becker (1957) established that employers with a "taste for discrimination" against women will hire fewer than the profit-maximizing number of women, employing more men who are equally skilled yet more highly paid. In my project, I intend to proxy the misogyny discourse displayed on social media by the frequency of misogyny posts on Twitter and examine whether it correlates with the heterogeneity in women's outcomes across the US labor markets. To be specific, I will construct a county-level index (hereinafter "Twitter misogyny index") by taking advantage of Twitter users' self-revealed location information and calculating the proportion of misogyny tweets in a particular period. I expect to observe that counties with a higher level of prevailing misogyny on Twitter have a larger gender wage gap.

The remainder of the paper proceeds as follows. Section 2 describes the approach of identifying misogynistic tweets, which highlights on lexicon construction. In Section 3, I describe the details on tweets scrapping process and location identification. Section 4 displays some primary visualization results. Section 5 introduces the labor outcome data and presents results of the quantitative analyses. Section 6 draws the conclusions and evaluates the data collection and analyzing strategies of this project.

[1] I know the repository name does not make sense...because I actually changed the final project topic in the middle.

2 Lexicon Construction

I intend to construct an index that describes the frequency of the occurrences of misogyny tweets in a specific period and use it to make cross-country comparisons. I choose to construct the index at the county level to be able to match it with the available income and employment data, which is at the county level.

The ideal way of identifying misogynistic tweets is to train a classifier with an annotated dataset of a decent size, by employing AMT workers, for example. However, due to time and budget constraints, I wasn't able to generate such a dataset by myself. Instead, I found an annotated dataset published by Guest et al. (2021) on GitHub[2], which is an expert labeled dataset of Reddit posts and comments to enable automatic classification of misogynistic content , which contains 5,868 nonmisogynistic posts and 699 misogynistic posts. Though the annotations of this dataset also include subcategories of misogyny, this project only considers the binary indicator of misogyny versus non-misogyny. With this dataset, I trained a classifier based on the Bidirectional Encoder Representations from Transformers (BERT) from Hugging Face. Specifically, I used the "bert-base-uncased" model. I randomly selected 90% of the posts in the Guest et al. (2021) dataset to be the training set, and the remaining 10% to be the test set, and feed the former to the BERT transformer. The further trained model returns a 90% accuracy rate on the test set.

However, the size of the annotated dataset is small. I expect the BERT classifier to be only able to recognize a few patterns of misogynistic text. Therefore, instead of directly use the trained BERT classifier to label each of the tweets I scrapped, I used it to construct a lexicon.

Farrell et al. (2019) constructed a large lexicon with 9 categories and 1,300 terms to collect as many misogyny Reddit posts as possible for future researchers to train misogyny content detectors. This is not necessarily efficient for this project, which also requires identifying misogyny tweets accurately. To refine the lexicon, I first deleted all the terms in the "Racism" category, since collecting racism tweets would interrupt my future examination of gender-based discrimination. This leaves a dataset of 8 categories and 630 terms[3].

Next, I collected all the tweets containing at least one of the terms with the python package "snscrape" in the lexicon on a random date, February 26, 2021. This returned me a dataset of 2,763,637. I deleted all the terms with occurrences less than 1,000 times, and feed the remaining set containing 118 terms to the

[2]Their misogyny dataset is publicly available at `https://github.com/ellamguest/online-misogyny-eacl2021/tree/main/data`

[3]Specifically, the 8 categories are (i) Physical violence towards women, (ii) Sexual violence towards women, (iii) Hostility towards women, (iv) Belittling of women (v) Exclusion of women, (vi) Promotion of Patriarchy or Male Privilege (vii) Stoicism (This is adopted from Donna Zuckerberg's book Not All Dead White Men, where she documented that the words of the Stoics deployed to support an ideal vision of masculine life.) and (viii) Flipping the Narrative (Flipping the Narrative encapsulate terms and expressions that refer to men being oppressed by women or by other men).

Table 1: First round result of lexicon refining

NO.	Term	(1) Total number of tweets	(2) Number of misogynistic tweets	(2)/(1)
1	bitch	16028	446	0.027826
2	bitches	15393	414	0.026895
3	pussies	2892	54	0.018672
4	pussy	3061	53	0.017315
5	pussys	2791	46	0.016482
6	fucker	1016	14	0.01378
7	hoe	2602	33	0.012683
8	hoes	2655	29	0.010923
9	boob	1308	14	0.010703
10	bang	2386	25	0.010478
11	ram	1060	11	0.010377
12	blast	1260	13	0.010317
13	harm	1577	16	0.010146
14	queers	1008	10	0.009921
15	punch	1642	15	0.009135

trained BERT transformer. The classification results from BERT show that 105 of the terms identified at least one misogyny tweet. In Table 1 I showed the terms that returned the highest proportion of misogyny tweets. We can see all of them make sense to some extent, but words like “bang”, “harm” and “queers” are clearly used more frequently in other nonmisogynistic situations. Also, we see terms tend to be posited close to their plural form. This is partly because Twitter’s search engine does not match the whole word, but also part of the word (for example, searching for “bitch” would also return tweets containing “bitches”). Also, words and their plural form are often used interchangeably. Therefore, for future analysis, I only used the stemmed version of each word in the lexicon.

Since tweets in one day are not enough for me to identify more rarely occurred misogyny terms, I further collected the tweets with each of the terms in the lexicon on May 10, July 22, October 3, and December 15 (with a time interval of $365/5 = 73$ days), excluding the 105 words already occurred in the first round. This time I got a dataset of 1,131,143 tweets. Since I want to identify as many misogyny terms as possible, I feed all of them to the BERT transformer. Table 2 shows the results of this round. We can see that except for the first three terms, the proportion of misogyny tweets labeled by BERT under the set of tweets returned by each of the terms is generally less than that in the first round. This might be a result of sampling and the smaller size of the data.

Based on the results of two rounds of tweets scrapping and BERT classifications, I used the first 9 words from Table 1 and the first 11 words from Table 2. Dropping words that are plural forms of other words in this set, I got a lexicon of 16 words, shown in Table 3.

Considering that different terms in the lexicon are not equally “efficient” in identifying misogyny

Table 2: Second round result of lexicon refining

No	Term	(1) Total Number of Tweets	(2) Number of Misogynistic Tweets	(1)/(2)
1	skank fuck	4	1	0.25
2	cunt rag	5	1	0.2
3	bitch ass	5677	160	0.028184
4	cumslut	308	3	0.00974
5	mothafucka	144	1	0.006944
6	assfuck	293	2	0.006826
7	slut	5381	32	0.005947
8	whore	8075	32	0.003963
9	dumb fuck	1617	5	0.003092
10	cunt	8140	20	0.002457
11	cock suck	501	1	0.001996
12	whitey	581	1	0.001721
13	faggot	581	1	0.001721
14	lolita	599	1	0.001669
15	douchebag	610	1	0.001639

Table 3: The refined lexicon

Term	weight
bitch	1
pussy	0.394086
fucker	0.190318
hoe	0.127087
boob	0.013003
skank fuck	1
cunt rag	1
bitch ass	0.935779
cumslut	0.283572
mothafucka	0.184706
assfuck	0.180516
slut	0.149429
whore	0.079271
dumb fuck	0.048481
cunt	0.026021
cock suck	0.009719

tweets, I also calculated the weights of each term by using the below equation:

$$W_t = \frac{prop_t - prop_{r^{-1}}}{max(prop_t) - prop_{r^{-1}}} \tag{1}$$

where r stands for the round. $prop_t$ is the proportion of misogyny tweets returned by the term t; $prop_{r^{-1}}$ is the proportion of misogyny tweets returned by the term one position below the cut-off of selected term in the round r (for the first round, this refers to the term "bang"; and for the second round, this refers to the term "whitey"). $max(prop_t)$ is the largest $prop_t$ in round r. For the two outliers "skank fuck" and "cunt rag", I took a proportion of 0.03 for both words, instead of 0.25 and 0.2. The results are shown in the second column of Table 3.

It's true that by adopting a lexicon I'm unlikely to perfectly identify each misogyny tweet in the focused period. Some misogyny tweets may not contain words in the lexicon, and some tweets containing words in the lexicon may not necessarily convey misogyny meanings. However, I argue that the false identifications here won't undermine the results of later causal inference as long as their occurrences were random, and not associated with local labor market outcomes. This condition can be derived formally as:

$$P(FI_c = f|TMI_c, X_c) = P(FI_c = f), \forall c, f \tag{2}$$

where FI_c are the occurrences of false identifications in a county, TMI_c is the magnitude of the Twitter Misogyny Index of this county, and X_c is a vector of county-level controls. $P(FI_c = f|TMI_c, X_c)$ refers to the conditional distribution of FI_c, which is independent of TMI_c and X_c when it's identical to the unconditional distribution.

3 Tweets Scrapping and Location Identification

Besides the tweets I already collected in five days (February 26, May 10, July 22, October 3, and December 15), which I deleted tweets returned by terms not occurring in the final lexicon, I also appended a new set of tweets in another five days: January 20, April 3, June 15, August 27, November 8. The size of the final tweets dataset is 1,516,565.

I retrieved the location of each user from the location string in the user's profile. This task is complicated for three reasons. First, the users can literally write whatever they want, and even if they wrote a valid location tag, it can take many forms (e.g. "New York", "New York City", "New York City, NY"). Third, there are many counties and cities of the same name in the US. Therefore, I only identified locations

Table 4: Distribution of collected tweets

Terms	All tweets	Tweets with state tags	Tweets with county tags	Tweets with city tags
bitch	804568	173843	34570	37713
pussy	170992	34519	5901	7858
hoe	133952	34271	7204	7218
cunt	76943	9519	1511	4545
whore	75585	11639	2033	2817
boob	66049	7925	1223	1999
bitch ass	60387	15190	3197	3111
slut	54184	8728	1480	2070
fucker	45852	8249	1121	2197
dumb fuck	16138	3263	559	751
cock suck	4653	612	75	192
cumslut	3419	390	66	74
assfuck	2552	572	109	112
mothafucka	1208	345	68	67
skank fuck	44	11	4	5
cunt rag	39	7	0	5

that occurred in one of the below forms:

1. Have a comma in the string, the substring before the comma is a county or city in the US, and the substring after the comma is a state name or state abbreviation (e.g. "Los Angeles, CA")

2. Have a comma in the string and the substring before the comma is a state name or state abbreviation (e.g. "CA, US")

3. Is an unambiguous county or city in the US (e.g. "Los Angeles") Eventually, I got 315,151 tweets with valid state tags, 90,598 tweets with valid county tags, and 131,504 tweets with valid city tags. The distribution of the tweets by terms is shown in Table 4.

The Kolmogorov–Smirnov test shows that there's no significant difference in distributions by terms between all tweets and tweets with state tags, tweets with county tags, or tweets with city tags.

Ideally, the Twitter Misogyny Index (TMI) should be the proportion of the occurrences of misogyny tweets in a particular county to the total number of tweets generated in this county in the random dates selected. However, it's computationally costly to collect all tweets in a county on several particular days. Instead, I collected 1,702,247 tweets by searching for the keyword "the" on March 10. The word "the" should be general enough to avoid region- or user-related biases. I feed the tweets returned by this round of scrapping to the location identification function described previously and get the estimated "Twitter population". Specifically, the TMI is derived by

$$TMI_c = \frac{\sum_{c,k} NumMisoTweets_{c,k} \times W_k}{NumTheTweets_c} \tag{3}$$

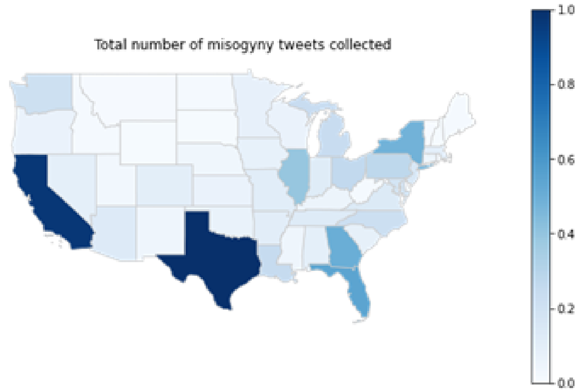

Figure 1: Total Number of Misogynistic Tweets collected

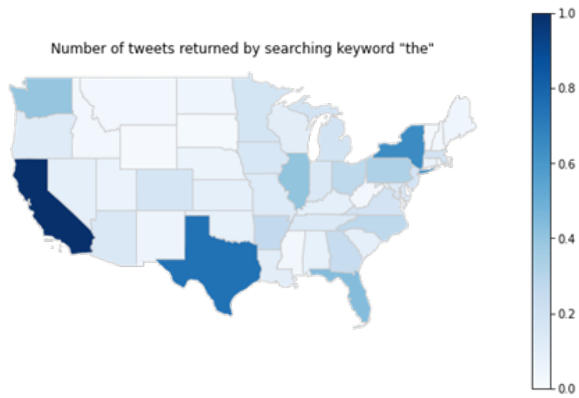

Figure 2: Number of Tweets returned by searching keyword "the"

where c refers to county, k refers to terms in the refined lexicon. W_k s the weights of the term k calculated by equation (1). Therefore, the TMI_c can be regarded as a weighted sum of the number of misogyny tweets in county c, divided by the "population" of county c estimated by searching for the term "the".

4 Primary Visualization Results

4.1 Twitter Misogyny Index

Figure 1 is the total number of tweets returned by the refined misogyny lexicon in each state, and Figure 2 is the number of tweets returned by searching the keyword "the". We can see that compared to the variation of "tweets population", the variation of the number of misogyny tweets in each state seems to be larger, with most tweets coming from California and Texas, while a lot of states in the central area are under the risk of underrepresented.

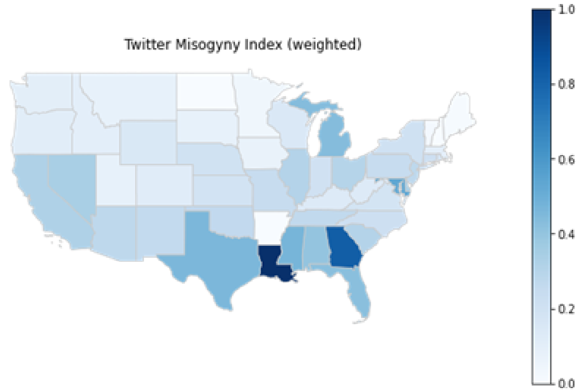

Figure 3: TMI at the state level (weighted)

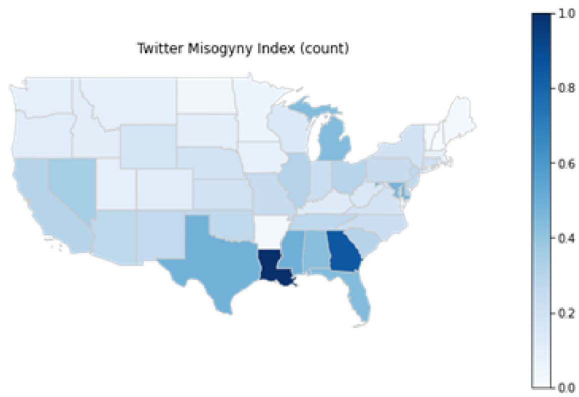

Figure 4: TMI at the state level (count)

Figure 3 is the Twitter Misogyny Index calculated at state level by equation (2). The Louisiana State ranks the highest, followed by Georgia. The states at the south tend to be more misogynistic.

Figure 4 shows the relative magnitude of misogyny by simply counting the number of tweets at each state returned by the lexicon. This has little difference compared to Figure 3. This is very likely to be due to that the word "bitch" occurs most frequently, and was also assigned the highest weights. Given this high similarity, in further analysis of this draft, I will use only the weighted TMI.

Figure 5 is the TMI calculated at the county level. The counties with white color had no misogyny tweets are tweets with the word "the" collected. This figure shows that even though more than 1,500,000 tweets are collected, they only cover a few counties in the US. As a result, we should be cautious of the state-level results, since they are driven by only part of the counties within. Also, I didn't include tweets with only city location tags in the graph, because users choosing to reveal their city and those who reveal their counties may be systematically different, and there are more duplicated names among cities than among counties. However, the observations with city tags should be able to provide important

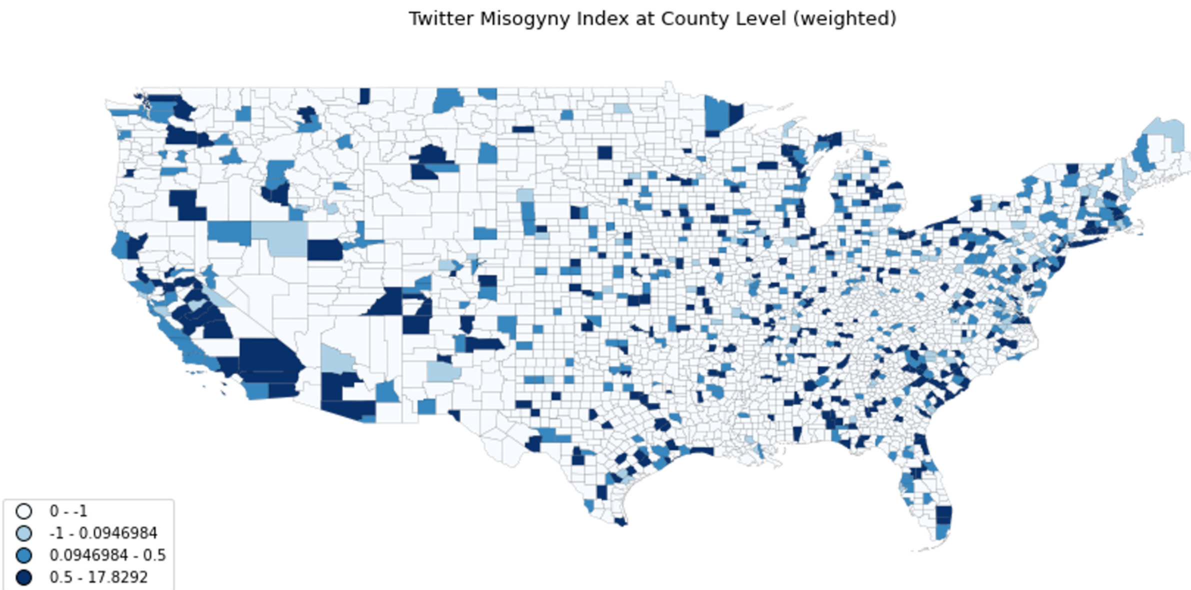

Figure 5: TMI at the county level

implications, especially that there are significantly more users choosing to reveal their cities than to reveal their counties (131,504 versus 90,598)

4.2 The text

In this section, I tried to use topic modeling as well as word embedding method to get some intuitions from the text.

4.2.1 Topic modeling

I used K-means and LDA to explore if there are some salient topics in the misogynistic tweets I collected. I first used Silhouette analysis to decide an appropriate number of cluster. The results of the Silhouette analysis are shown in Figure 6-8. For a cluster number of 2,3,4, the Silhouette scores are 0.023, 0.015, and 0.013, respectively. Therefore, I chose the cluster number to be 2. A visualization of the K-means on a PCA-reduced 2-dimensional graph is shown in Figure 9.

From Figure 9, it seems that most of the observations were classified to be part of a large cluster, and there's only a few exceptions. I printed some tweets in both clusters to observe if there's any salient differences. However, it seems that the tweets in the small cluster are only unique for being particularly short.

From Table 5, we can see that the text is comprised with dirty languages and slangs, interspersed with

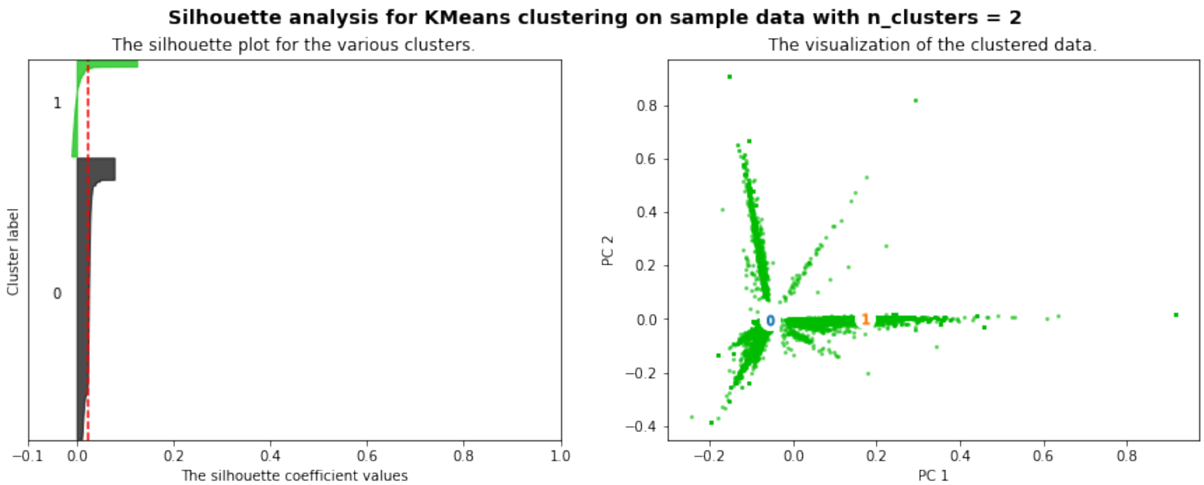

Figure 6: Silhouette analysis with a cluster number of 2

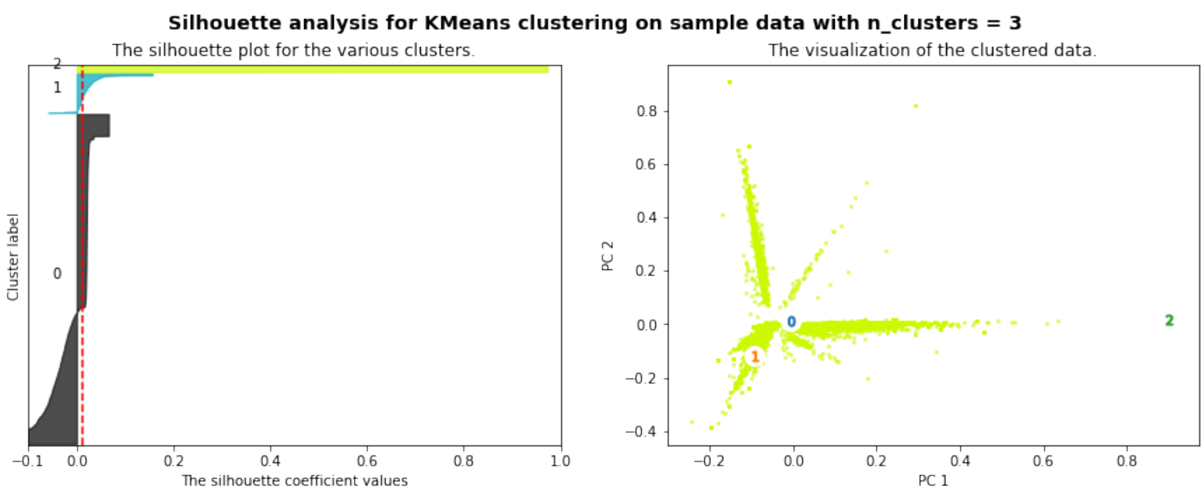

Figure 7: Silhouette analysis with a cluster number of 3

Table 5: Most salient words in each topic by LDA

	Topic_0	**Topic_1**	**Topic_2**	**Topic_3**	**Topic_4**	**Topic_5**	**Topic_6**	**Topic_7**	**Topic_8**	**Topic_9**
0	don	pussy	bitch	hat	know	shit	hoes	hoe	ou	like
1	nt	fuck	bitches	slut	want	amp	bad	love	whore	time
2	wanna	boobs	ass	need	fucking	little	good	let	work	hate
3	tell	going	itch	lol	fucker	people	nigga	big	gone	ain
4	dumb	cunt	got	right	called	man	cause	day	ike	nd
5	dick	said	ll	eat	fuckers	new	ve	hy	money	stop
6	boob	thing	think	face	damn	cunts	tryna	house	away	itches
7	ow	getting	look	wtf	ur	yo	job	girls	niggas	mad
8	start	ot	girl	kids	woman	way	come	listen	probably	play
9	boy	better	baby	went	et	oobs	looking	uck	calling	head

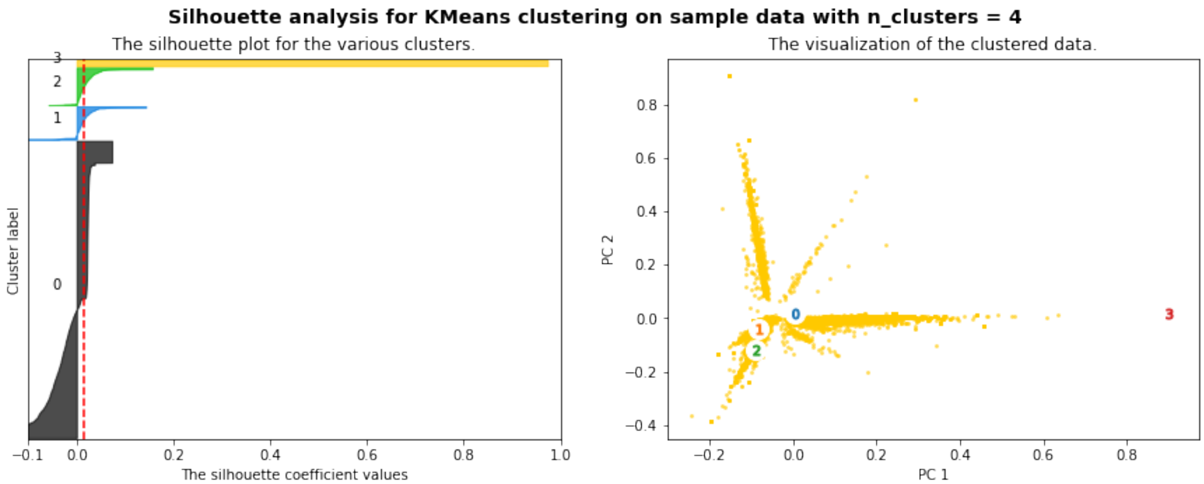

Figure 8: Silhouette analysis with a cluster number of 4

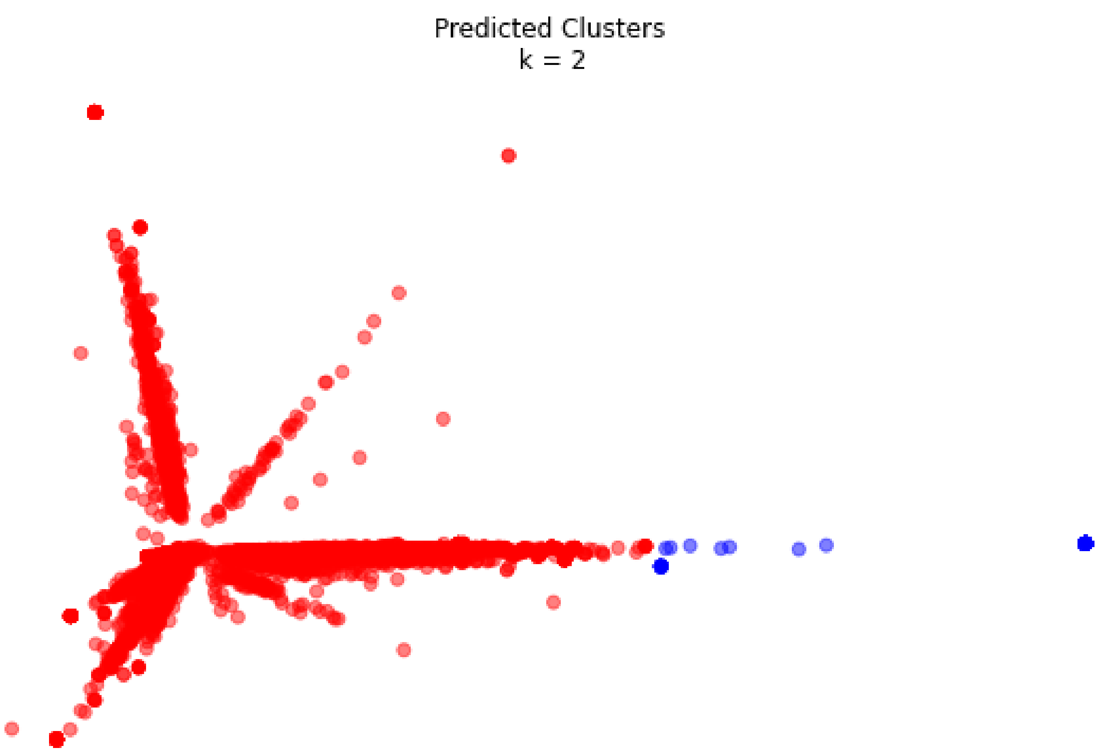

Figure 9: Visualization of the result of K-means on a 2D graph

sex-related terms. Note that racist terms like nigga occurred in Topic 5 and 8. This could be implying a potential cultural association between racism and sexism. Or alternatively, my BERT classifier was detecting hate speeches instead of misogyny.

4.2.2 Word Embedding

In this section, I used word embedding tools to explore the language space of my tweets sample. Figure 10 shows the most similar words to "woman" and "man" in my sample reported by the Gensim word embedding model. The results show that "man" is related "fine", "cool" and "friend", while "woman" is related with "ucking", "shame". These results support my strategy of identifying misogynistic tweets.

And from the visualization in Figure 11, the take away is that except for the clearly misogynistic terms, the corpora are filled with informal terms.

5 Quantitative analyses

In this section, I intend to use regression method to examine whether there is a significant relationship between misogynistic culture and women's labor outcomes, as implied by the theory of Becker (1957). I first introduced my empirical model and the labor outcomes data I used for this project. Then, I show the results.

5.1 Empirical strategy

The empirical model for examining the correlation between TMI and labor outcome is:

$$Y_c = \alpha + \beta TMI_c + \gamma X_c + \epsilon_c \tag{4}$$

The subscription c in equation (4) refers to county and the Y_c refers to county-level gender differences in earnings and labor market participation. X_c is a vector of controls, including the distributions of ages, ethnicity, education attainment, and the total population. As described above, the data of the control variables will all be collected from U.S. Census Bureau, and they are all county-level measurements.

5.2 Earnings, Education and Demographics data

The earnings and demographics data used in this project are from the American Community Survey (ACS), which is a demographics survey program conducted by the U.S. Census Bureau. The ACS is a

```
misW2V.wv.most_similar('woman')
```

```
[('ucking', 0.998742938041687),
 ('ick', 0.9986869692802429),
 ('pussies', 0.9985202550888062),
 ('takes', 0.998462438583374),
 ('ave', 0.9984195828437805),
 ('enjoy', 0.9983705878257751),
 ('girls', 0.9983689785003662),
 ('feels', 0.9983376264572144),
 ('shame', 0.9982975721359253),
 ('balls', 0.9982428550720215)]
```

```
misW2V.wv.most_similar('man')
```

```
[('fine', 0.997022807598114),
 ('place', 0.9969719052314758),
 ('cool', 0.9969708919525146),
 ('somebody', 0.9968761205673218),
 ('turn', 0.9967746138572693),
 ('find', 0.9967357516288757),
 ('ike', 0.99672532081604),
 ('ok', 0.9965589642524719),
 ('friend', 0.9965091347694397),
 ('screen', 0.9964550137519836)]
```

Figure 10: Most similar words of ”woman” and ”man”

Figure 11: Visualization of most frequent words

large, monthly, national survey of the U.S. population that is sent to about a quarter-million households each month in order to provide nationally-representative data on the equivalent of the full long-form content on a yearly basis. This project used the newest cross-sectional data in 2019.

I employed several variables related to the respondents' earnings in the past 12 months and labor market participation as the dependent variables. The earnings-related variables include the sex ratio of workers with earnings divided by the sex ratio of the whole population (hereinafter "Worker Ratio"), the sex ratio of full-time workers divided by the sex ratio of the whole population (hereinafter "Full-time Worker Ratio"), sex ratio of workers' median earnings (hereinafter "Worker Median Earnings Ratio"), sex ratio of full-time workers' median earnings (hereinafter "Full-time Worker Median Earnings Ratio"), sex ratio of full-time workers' average earnings (hereinafter "Full-time Worker Average Earnings Ratio"), sex ratio of workers earning less than 10 thousand dollars, 10-25 thousand dollars, 25-50 thousand dollars, 50-75 thousand dollars, 75-100 thousand dollars, and above 100 thousand dollars, divided by the sex ratio of all full-time workers (hereinafter "Full-time Worker Ratio Blow 10k, 10-25k, 25-50k, 50-75k, 75-100k").

Besides, a set of county-level demographic variables including population, ages, ethnicity, and education are considered as the control variables. The full list of variables selected from the 2019 ACS and their summary statistics are shown in Table 6.

Table 6: Summary statistics

Variable	Obs	Mean	Obs	Mean	p-values for t-stat
	County-level TMI non-missed		County-level TMI missed		
TMI(county)	211	0.67548	.	.	.
Worker Ratio	211	0.01136	629	0.011318	0.5565
Full-time Worker Ratio	202	0.01373	589	0.01362	0.3584
Worker Median Earnings Ratio	211	1.41104	629	1.43221	0.1601
Full-time Worker Median Earnings Ratio	211	1.24235	629	1.27313	0.0057***
Full-time Worker Average Earnings Ratio	202	1.33037	589	1.36138	0.009***
Full-time Worker Ratio -10k	202	1.4494	582	1.15397	0.2012
Full-time Worker Ratio 10-25k	202	0.69983	589	0.678264	0.243
Full-time Worker Ratio 25-50k	202	0.82454	589	0.801968	0.0421**
Full-time Worker Ratio 50-75k	202	1.11418	589	1.16586	0.0435**
Full-time Worker Ratio 75-100k	202	1.53342	589	1.64055	0.1953
Full-time Worker Ratio 100k-	202	2.19321	589	2.77127	0.0222**
Proportion $\geq$ 25 with $\geq$ high school degree	211	0.893212	629	0.896662	0.3517
Proportion $\geq$ 25 with $\geq$ Bachelor's degree	211	0.324319	629	302232	0.0107**
Total population	211	528903	629	269088	0.00***
Percentage of population under 18 years old	211	21.835	629	22.0733	0.3299
Percentage of population over 65 years old	211	17.337	629	17.1868	0.6631
Percentage of white population	198	74.816	583	78.4012	0.0037***
Percentage of black population	198	10.358	583	11.2115	0.3904

Table 6 implies some significant systematic differences between counties with non-missing Twitter Misogyny Index and those without (i.e. either there are no misogynistic tweets collected, or there are no tweets collected by searching for the word "the"). Counties with the TMI missed have higher gender differences in median earnings and average earning. Besides, these counties also have significantly fewer young people, fewer people holding Bachelor's degrees or higher among the 25-year-old-and-above population, and a larger proportion of the white population. One thing to notice is that the average population among these counties with missing TMI is around half of that for counties with non-missing TMI. Therefore, one needs to be cautious in interpreting the later regression results, since there's strong evidence that they are biased towards more populated counties.

5.3 Correlation matrix

Figure 13-14 show the heat map drawn from the calculation results of the correlation matrix between occurrences of misogynistic terms and labor outcomes at the state level and at the county level, respectively. Two major conclusions can be drawn from these results: First, the correlation between the occurrences of the misogynistic terms are impressively high, implying that it might be a better strategy to calculate one index out of all the measures instead of using all of them. Second,

5.4 Regression results

Table 7 shows the estimation results of the β in Equation (3), with the Y_c being Worker Ratio and Full-time Worker Ratio. In Columns (1) and (4), only the independent variable, TMI is included in the

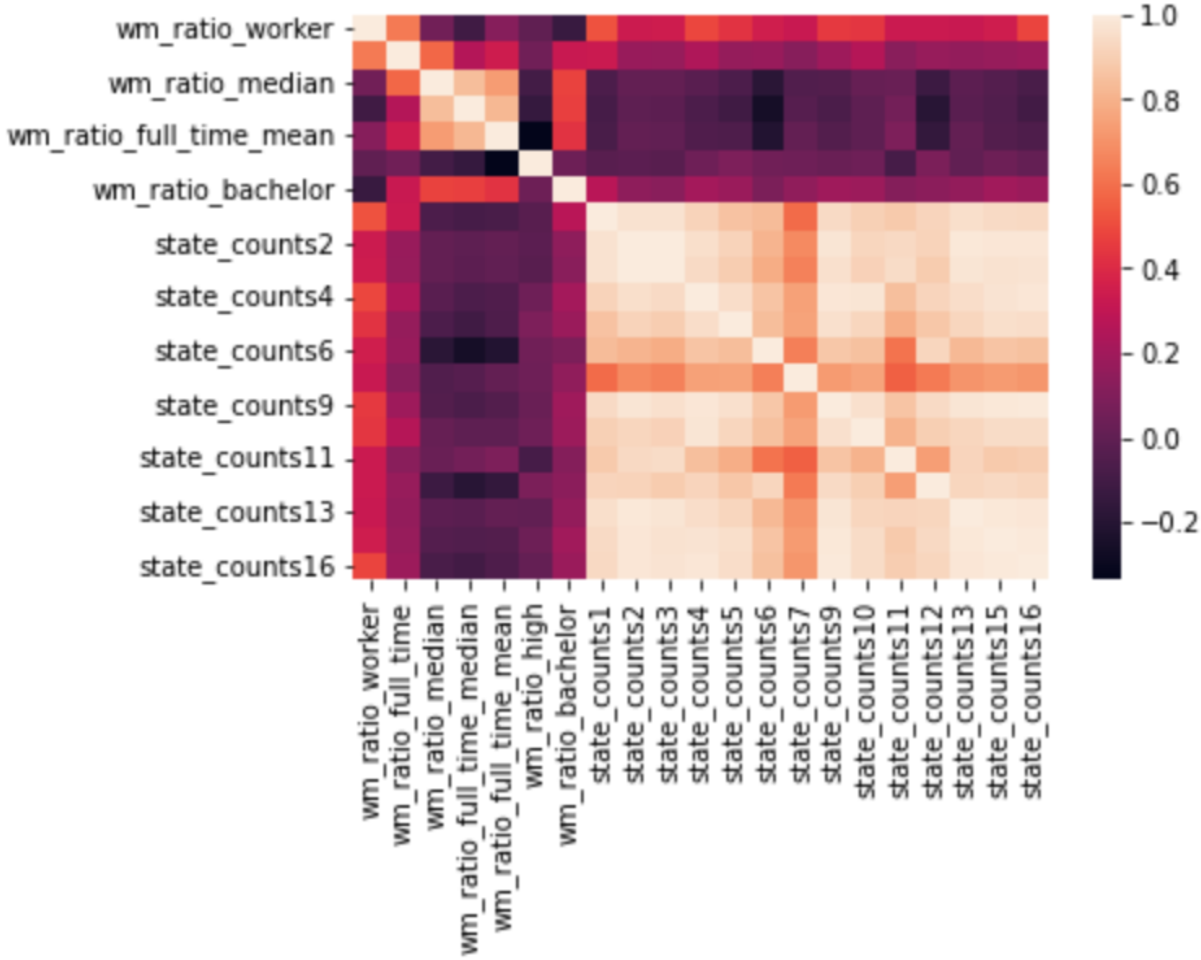

Figure 12: Heatmap of the correlation matrix between occurrences of misogynistic terms and labor outcomes at the state level

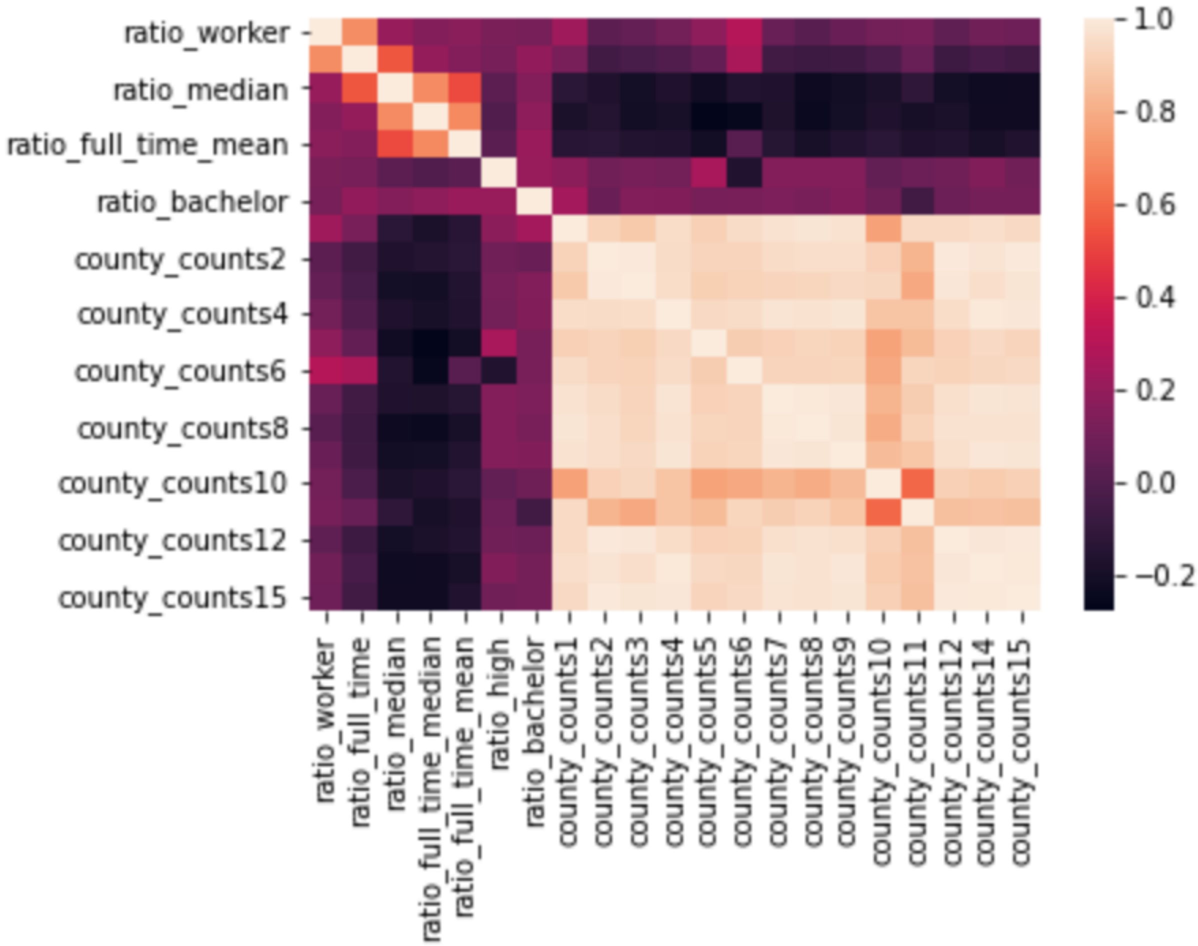

Figure 13: Heatmap of the correlation matrix between occurrences of misogynistic terms and labor outcomes at the county level

Table 7: County-level correlations between the TMI and the labor market participation

Panel A: Outcome Var: Worker Ratio			
	(1)	(2)	(3)
TMI	6.09E-05	5.30E-05	4.71E-05
	-6.12E-05	-6.24E-05	-5.63E-05
Constant	0.0113***	0.0106***	0.00891***
	-6.30E-05	-0.00112	-0.00168
County Controls	N	Y	Y
State FE	N	N	Y
Observations	211	198	198
Panel B: Outcome Var: Full-time Worker Ratio			
	(4)	(5)	(6)
TMI	0.000144	0.000171	0.00013
	-0.000124	-0.000157	-0.00015
Constant	0.0136***	0.0128***	0.00835**
	-0.00012	-0.00294	-0.00394
County Controls	N	Y	Y
State FE	N	N	Y
Observations	202	202	193

equation. Columns (2) and (5) added the county-level control variables shown in Table 6. And Columns (3) and (6) further added the state indicators. It's almost safe to say that there's no correlation between the TMI and the gender differences in labor market participation at the county level.

Table 8 shows no evidence of a statistically significant correlation between TMI and three earnings-related variables: men versus women's median earnings ratio among all workers, among full-time workers, and men versus women's average earnings ratio among full-time workers. In fact, the two are even negatively correlated except in Column (9).

From Table 9, there's a little evidence that counties with a higher intensity of misogyny tweets have slightly higher gender gaps in the population of the 25-50k yearly earnings group.

6 Discussions

In this project, I constructed a lexicon of misogyny terms based on the lexicon in Farrell et al. (2019) using two rounds of pre-collections of tweets and the BERT classifier I trained with the annotated Reddit misogyny dataset established by Guest et al. (2021). Then, I collected 1,516,565 tweets containing misogynistic terms, identified 315,151 tweets with valid state tags, 90,598 tweets with valid county tags, and 131,504 tweets with valid city tags. I first used topic modeling and word embedding method to qualitatively analyze the data. Then I used the tweets with county tags and state tags to estimate how the prevailing misogynistic discourses on Twitter are correlated with gender gaps in multiple dimensions

Table 8: County-level correlations between the TMI and the earnings

Panel A: Worker Median Earnings Ratio			
	(1)	(2)	(3)
TMI	-0.00335	-0.0123	-0.00344
	-0.0122	-0.0123	-0.014
Constant	1.413***	0.846***	0.919**
	-0.013	-0.315	-0.377
County Controls	N	Y	Y
State FE	N	N	Y
Observations	211	198	198
Panel B: Full-time Worker Median Earnings Ratio			
	(4)	(5)	(6)
TMI	-0.0137	-0.0182	-0.012
	-0.0125	-0.0111	-0.0113
Constant	1.252***	0.255	0.457
	-0.0106	-0.257	-0.316
County Controls	N	Y	Y
State FE	N	N	Y
Observations	211	198	198
Panel C: Full-time Worker Average Earnings Ratio			
	(7)	(8)	(9)
TMI	-0.00849	-0.00149	0.00539
	-0.00907	-0.00801	-0.00845
Constant	1.336***	0.187	0.0689
	-0.0111	-0.212	-0.351
County Controls	N	Y	Y
State FE	N	N	Y
Observations	202	193	193

Table 9: County-level correlations between the TMI and sex ratio of population in each earnings intervals

	(1)	(2)	(3)	(4)	(5)	(6)
Earnings interval	-10k	10-25k	25-50k	50-75k	75-100k	100k-
TMI	-0.319	0.0124	0.0165*	0.00183	0.00048	-0.0193
	-0.239	-0.0229	-0.00993	-0.0207	-0.0678	-0.0635
Constant	-4.455	0.851	2.310***	2.712***	0.639	-0.0973
	-9.815	-0.665	-0.381	-0.515	-1.978	-2.756
County-level Controls	Y	Y	Y	Y	Y	Y
STATE FE	Y	Y	Y	Y	Y	Y
Observations	193	193	193	193	193	193

in the local labor market at the county-level and state-level, respectively. The earnings, education, and demographics data were collected from the American Community Survey (ACS) Data. There's little evidence that the level of the Twitter Misogyny Index is correlated with gender gaps in earnings and labor market participation.

However, I also argue that there are important deficiencies in my current strategy of constructing lexicons, collecting tweets, and identifying locations, which may have had a significant impact on the estimations that veiled the actual correlation between the misogyny discourses on Twitter and women's labor market outcomes.

Though I collected more than three million tweets in total in the two-round procedure of pre-collection for lexicon constructing, this still leads to an overrepresentation of common words, like "bitch". This overrepresentation further results in a miscalculation of the weights of each word in the lexicon. Therefore, the Twitter Misogyny Index implied this lexicon may not reflect the true misogyny level, but is highly correlated with other variables such as the population of Twitter users. Therefore, in the next draft, I will adopt a different strategy of pre-collection for lexicon refining. First, instead of scraping the tweets by searching for each of the terms in the lexicon of Farrell et al. (2019) on just one particular day, I will collect 10,000 tweets for each of the terms in the lexicon. In addition, I will generate a random series of date-hour pairs in 2021, and loop through this time series with at most 100 tweets being collected for each term at each time point until 10,000 tweets are collected for each term in the lexicon. Therefore, I will get enough samples to calculate the proportion of misogyny tweets for the less frequently occurred words.

Second, I found it difficult to identify correct location tags in practice, not only because users may write the location string in every imaginative way, but also due to the existence of a variety of duplicated names among US cities and counties. Therefore, despite that I detected more than 130,000 observations in my current dataset that contains city tags, I wasn't able to confirm what specific cities they are. In my next draft, I will employ three strategies of location identification.

First, I intend to make a priority list of counties and cities in the US according to their population and mark the ambiguous location strings as the most populated county or city with the same name. This is for recognizing as many places as I can to draw a detailed picture of misogyny level among every possible corner in the US.

Second, I will loop through every county included in the ACS data, as well as the most populated cities in each county, to collect the same amount of tweets at each location. This is for getting accurate estimations of the correlations between TMI and local labor outcomes.

And for both strategies, I will generate a series of random date-hour pairs as in the previously planned step of lexicon refining, to avoid random factors on any particular date.

Also, this draft wasn't able to use the GSS data for cross-validations of the credibility of the TMIs (though they are clearly not credible). I believe that adding this cross-validation part will make the results of this project more convincing, and I'll continue to work on it.

There's also a weak point in the initial step, which is the dataset I used to train the classifier. The sample is constructed by Guest et al. (2021) from Reddit, with only around 7,000 pieces of posts, and around 700 misogynistic posts. It's not likely that the 700 posts will cover all the patterns or even just all the terms in Farrell's lexicon. A better solution would be to construct an annotated lexicon with tweets, but I may not have the time and money (for hiring AMT workers) for it. The suboptimal approach is to find more annotated datasets, especially those constructed with social media text, to construct a larger dataset for training the BERT classifier.