

# Topic- NLP to detect mental illnesses among social media users

ISM 6905 –Independent Study Project-  
Dr. Balaji Padmanabhan

Mentor: Negar Maleki

Presented By: Adhithyan Rangarajan



# Overview

We performed a literature review of current research in this space, understood the problem significance and methodology on detecting mental illness using social media data.

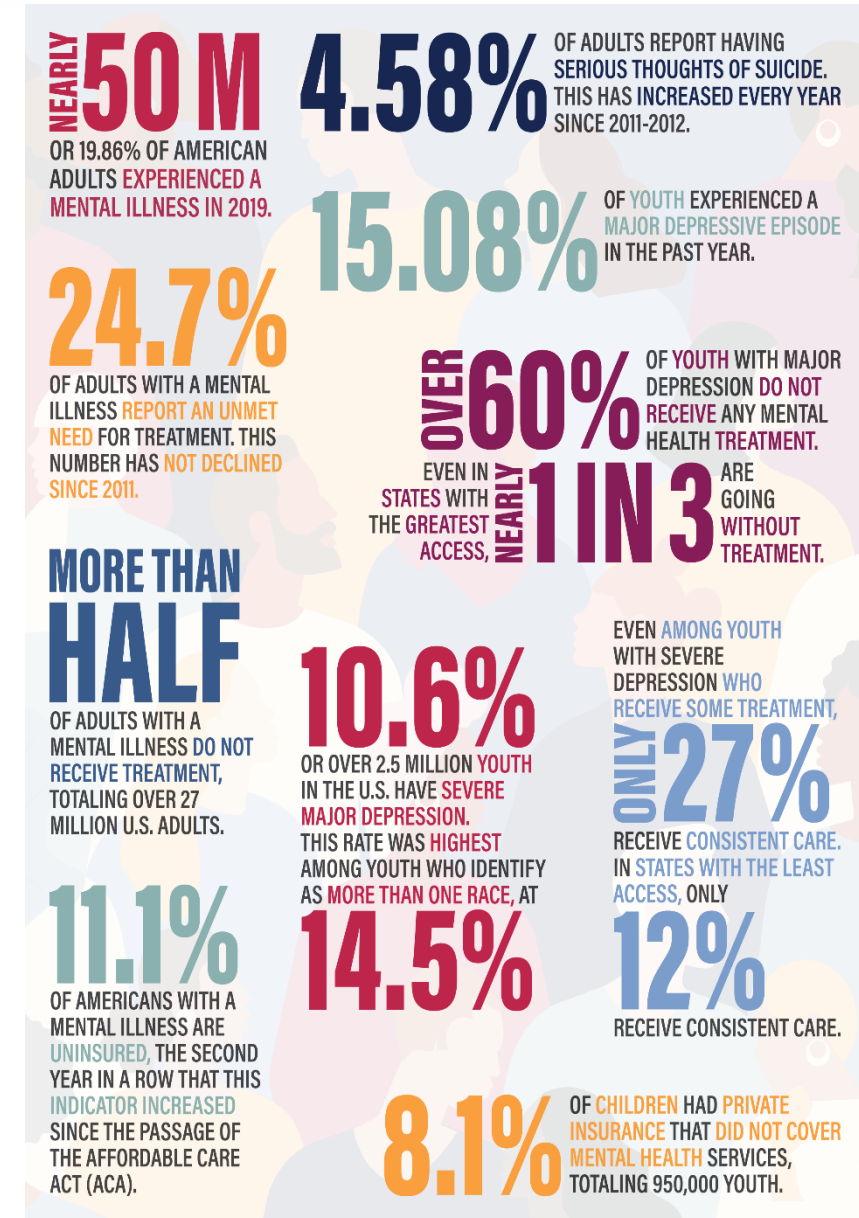
A reddit mental health dataset (**344000 posts**) from the Journal of Medical Internet Research was curated and used for training our ML models & scraped twitter data (**408000 tweets**) for users in Florida based on population density clusters was used for testing.

Analysis and recommendations were built based on model output metrics for the State of Florida & Florida Blue.

# Problem Scale and Significance

- In 2019, just prior to the COVID-19 pandemic, **19.86%** of adults experienced a mental illness, equivalent to nearly **50 million Americans**.
- Over half of adults with a mental illness do not receive treatment, totaling over **27 million adults in the U.S. who are going untreated**.
- **Over 60%** of youth with major depression **do not receive any mental health treatment**.
- **Suicidal ideation** continues to increase among adults in the U.S. at **4.58% this year** i.e., an increase of **664,000 people** from last year.
- 1 in 25 Americans lives with a serious mental illness, such as schizophrenia, bipolar disorder, or major depression.
- **1 in 5** Americans experience a mental illness in a given year. (52.9 million in 2020).
- **1 in 5 children**, either currently or at some point during their life, have had a seriously debilitating mental illness.
- **11.1%** of Americans with a mental illness **are uninsured**.

Sources – CDC, MHA National, National Alliance on Mental Illness



# Tweets and Mental Illness



#ifightBPD  
@me\_bpd

Follow

The shame I felt today by crying alone in my car was overwhelming. I admit I had enough of trying to do the right thing. Of course rejection reared it's ugly head again! I miss the old me so much 😞

#ifightbpd #SickNotWeak #mentalillness  
#Menshealth #depression #BPD #broken

5:17 PM - 23 Jan 2019

1 Retweet 6 Likes



Jenny  
@Jennycolak

Follow

Depression is the worst thing I myself have ever experienced. Constantly feeling so low, to the stage u don't want to get out bed in the morning because u dread the day ahead and no pill a doctor chucks at u take seems to fix it. I can't wait for the day I'm truly happy ❤️

4:42 PM - 22 Jan 2019

38 Retweets 282 Likes



6 38 282



Tobin  
@Haha728lol

Follow

I hate how I can never get genuinely excited anymore and I feel like when some good news is brought to me and that person is so happy and I'm just there, I feel like I bring them down because of that. This is what depression can do to people and then I just spirals down afterthat

1:06 AM - 22 Jan 2019

5 Likes



5



Audree  
@DarlinSincerest

Follow

Depression doesn't just hit you when everything is going wrong. It hits you when everything seems to be going right, too. It hits you in the middle of a laugh like a punch in the gut reminding you that you're not okay. It creeps in and swallows you piece by piece until you're

2:22 AM - 27 Dec 2018

210 Retweets 521 Likes



1 210 521



J. Neyhart (N👁️❤️🏳️🌈) @JenniferNeyhart · Feb 25, 2017



It's a heavy blanket of fog that weighs you down and sometimes all you want to do is sleep or die. [#ThisIsWhatDepressionFeelsLike](#)



J. Neyhart (N👁️❤️🏳️🌈) @JenniferNeyhart

The sun may be shining for other people, but all you see is darkness. [#ThisIsWhatDepressionFeelsLike](#)

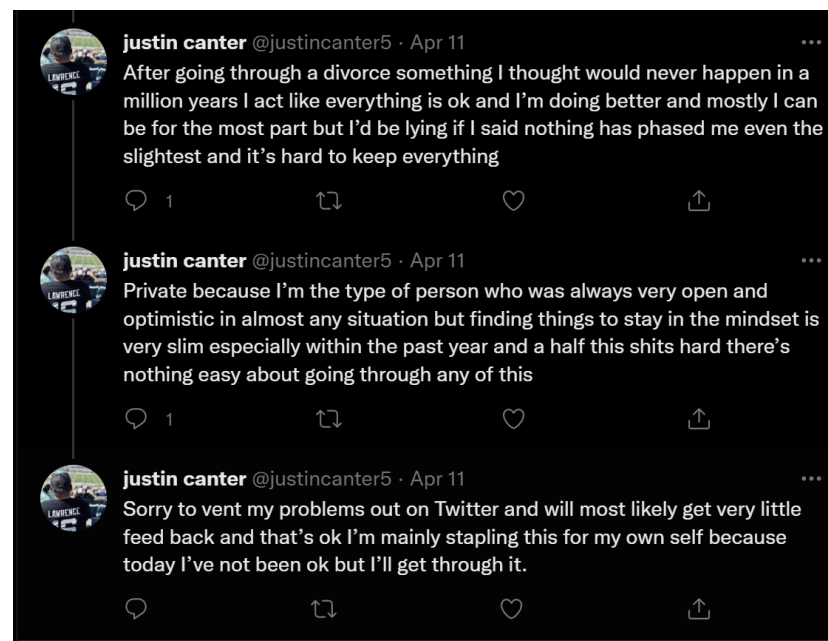
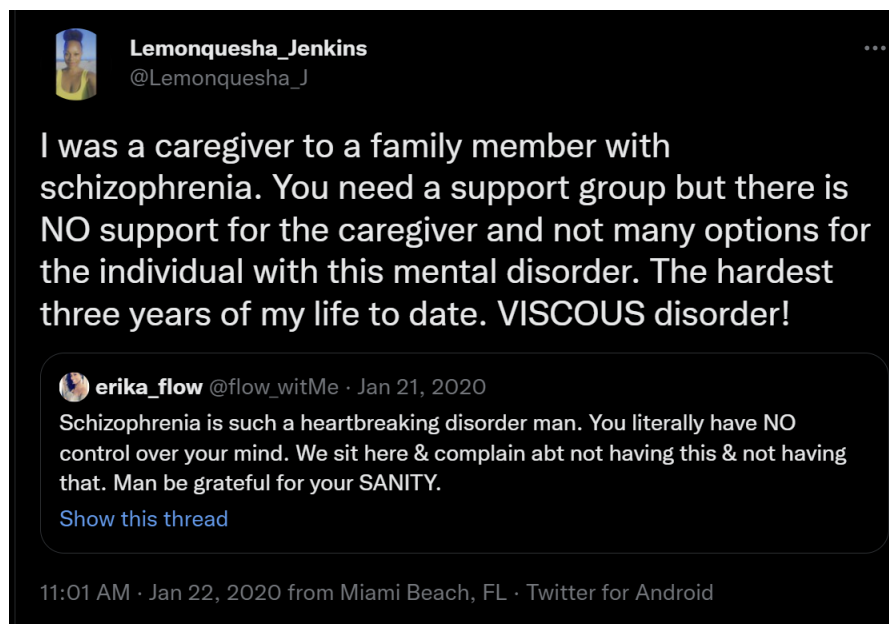
11:10 AM · Feb 25, 2017



9 Reply Share



# Some tweets from our dataset



# Methodology for Data Retrieval

Four clusters in Florida were used to pull the tweets for our dataset. These include Tampa, Miami, Jacksonville and the Panhandle.

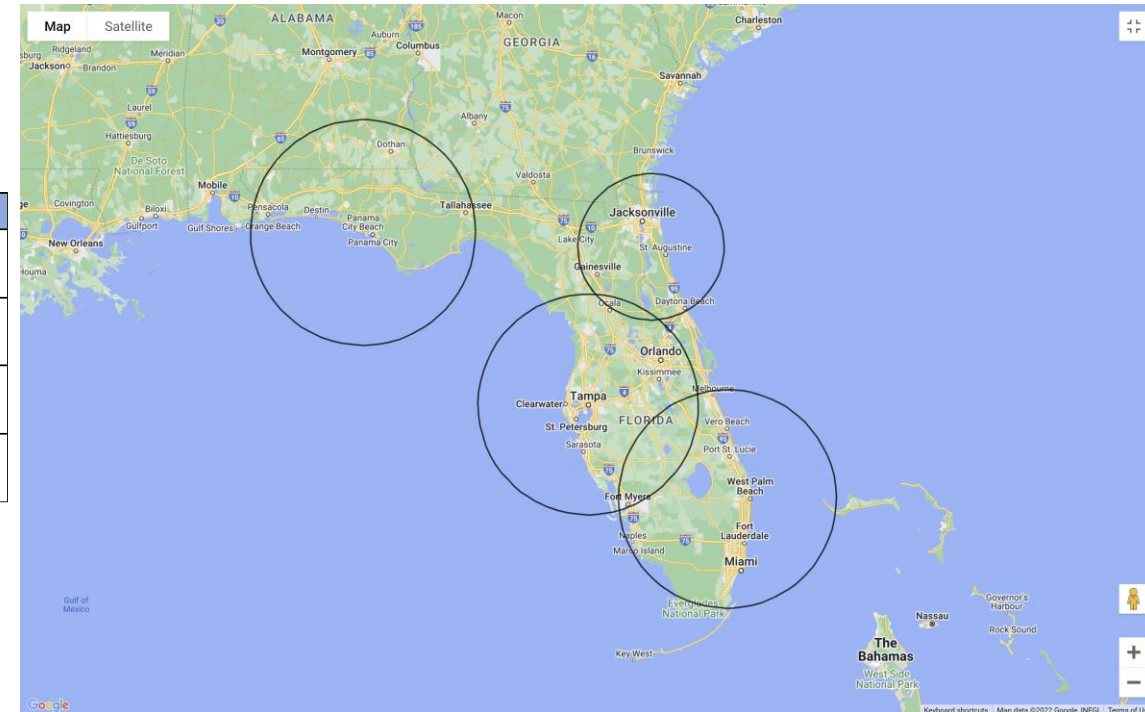
## Radius and Geolocation View of dense populations in Florida-

75% of Florida population lives in three clusters – Tampa-Orlando, Miami-Dade County and Jacksonville. (Source- US Census Bureau). Our selection to extract tweets is from the below 4 clusters to cover the maximum possible population in the state.

CITY	GEOLOCATION	RADIUS
Tampa	Position: 27.9506,-82.4572	Radius: 160934.00 Meters (100 miles)
Jacksonville	Position: 29.999398,-81.528424	Radius: 104607.10 Meters (65 miles)
West Palm Beach	Position: 26.704825,-80.392211	Radius: 160934.00 Meters (100 miles)
Panhandle - Panama City Beach/Tallahassee	Position: 30.176591,-85.805386	Radius: 160934.00 Meters (100 miles)

[www.census.gov/geographies/reference-files/2010/geo/state-local-geo-guides-2010/florida.html](https://www.census.gov/geographies/reference-files/2010/geo/state-local-geo-guides-2010/florida.html)

**Challenges and Tradeoffs-** We need to ensure coverage of location of major cities with high population and also have minimal overlap between the circles.



# Data Sample Volume

Total of 408K tweets were pulled- 213K using Geo-Location & 194K using City Name to form our final dataset.

Geo-Location Search								
	City (100 mile radius) -> (10K Records searched unless specified)	Tampa-Orlando Belt (100)	Miami(100)	Panhandle/ Panama City/ Tallahassee(100)	Jacksonville(65)			
Keyword Searched	PTSD	5963	3336	104	157			
	Trauma	4060	1303	183	335			
	Abuse	6752	6294	399	767			
	Drugs	10000	10000	887	3066			
	Alcohol	10000	5415	302	659			
	Stress(25000)	25000	14840	778	870			
	Anxiety(20000)	19392	9237	690	1628			
	Bipolar	8388	8616	38	105			
	Disorder	3576	265	87	185			
	Schizophrenia	557	351	6	18			
	Depress	6086	2750	278	170			
	Covid	10000	10000	10000	10000	FINAL TOTAL		
	TOTAL	109774	72407	13752	17960	213893	A	
City Name Search								
	City (100 mile radius) -> (10K Records searched unless specified)	Tampa-Orlando Belt (100)	Miami(100)	Panhandle/ Panama City/ Tallahassee(100)	Jacksonville(65)			
Keyword Searched	PTSD	1120	1175	302	157			TOTAL NUMBER OF TWEETS (A +B)
	Trauma	2402	1992	322	162			408359
	Abuse	4340	5992	1247	877			
	Drugs	10000	10000	5068	5521			
	Alcohol	6020	10000	2247	644			
	Stress(25000)	20767	25000	6799	4704			
	Anxiety(20000)	7671	10771	2337	1136			
	Bipolar	1464	4013	465	456			
	Disorder	668	1300	194	151			
	Schizophrenia	81	168	26	4			
	Depress	3157	5925	1271	993			
	Covid	10000	10000	3068	2289	FINAL TOTAL		
	TOTAL	67690	86336	23346	17094	194466	B	

# Twitter Dataset Sample

Post data extraction, this is a preview of the final dataset.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	url	date	content	username	keyword	date	time	month	tweethour	day	year	weekend	region	
38	https://twitter.com	2022-04-11 22:00	@CapaKAN I can see that causing you stress. At the end of th	GyrexGG	stress	4/11/2022	10:00:20 PM	Apr		22	Monday	2022 Weekday	Tampa	
39	https://twitter.com	2022-04-11 21:54	I got fired from my job and I still cant really figure out why ar	momcrotch	stress	4/11/2022	9:54:53 PM	Apr		21	Monday	2022 Weekday	Tampa	
40	https://twitter.com	2022-04-11 21:47	This project stressed me out so bad ðŸ˜~,	queforshort_	stress	4/11/2022	9:47:26 PM	Apr		21	Monday	2022 Weekday	Tampa	
41	https://twitter.com	2022-04-11 21:42	Six Stress Relieving Tips for Moving to a New City https://t.cc	Gabe_StuartFL	stress	4/11/2022	9:42:06 PM	Apr		21	Monday	2022 Weekday	Tampa	
42	https://twitter.com	2022-04-11 21:33	@dovidfeldman @justinclewis86 I was for sure. I stressed ou	drum_ear	stress	4/11/2022	9:33:35 PM	Apr		21	Monday	2022 Weekday	Tampa	
43	https://twitter.com	2022-04-11 21:14	@MsAriesMoon I know you have stomach issues so do I so h	Grateful_Gabb	stress	4/11/2022	9:14:25 PM	Apr		21	Monday	2022 Weekday	Tampa	
44	https://twitter.com	2022-04-11 21:10	"I think itâ€™s important to stress that we had 60 days wher	SunSentinel	stress	4/11/2022	9:10:01 PM	Apr		21	Monday	2022 Weekday	Tampa	
45	https://twitter.com	2022-04-11 21:00	Exhale the stress!	ThaBarKidReason	stress	4/11/2022	9:00:58 PM	Apr		21	Monday	2022 Weekday	Tampa	
46	https://twitter.com	2022-04-11 20:54	Sculpting a bat cat while I'm trying to not stress about a very	Pots N Pens	stress	4/11/2022	8:54:28 PM	Apr		20	Monday	2022 Weekdav	Tampa	



# Reddit Content used for training the models

- Reference - Low, D. M., Rumker, L., Torous, J., Cecchi, G., Ghosh, S. S., & Talkar, T. (2020). Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study. *Journal of medical Internet research*, 22(10), e22635
- Reddit data preview –

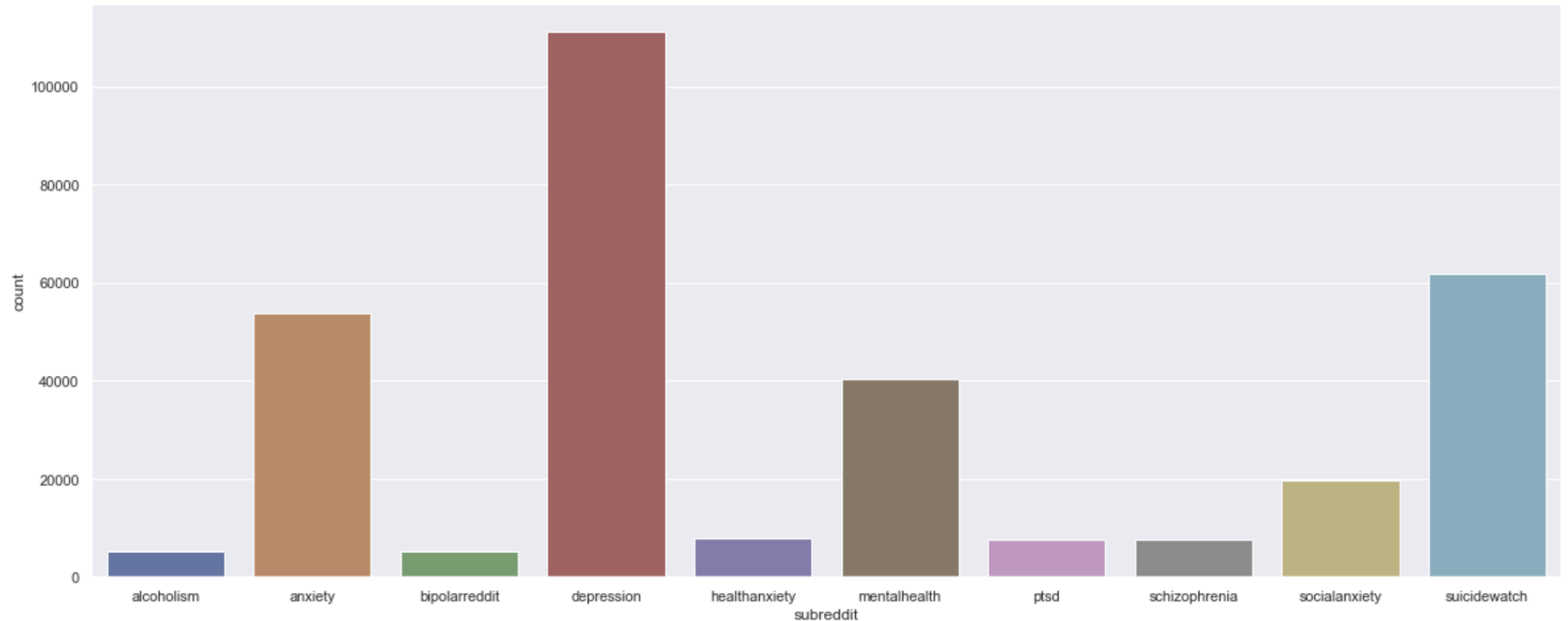
A	B	C
subreddit	post	
alcoholism	I need to quit while im young (22m) Im going through some hardcore withdrawals. I have anxiety as it is, but i just feel like im dying. Should i go to the hospital? I heard you can die	
alcoholism	It's 1:30am. So glad to be sober. I was at a New Years party tonight. All the adults got wasted and had a good time. I did not get wasted and I had a good time. Tomorrow I	
alcoholism	Might quit tomorrow but don't want to hurt myself going cold turkey. Usually drink 5-8 drinks per day from 11:00am I have been drinking heavily every day since age 21 besides a	
alcoholism	Treating alcoholism with naltrexone using Sinclair Method cures 80% of all alcoholics <a href="https://www.psychologytoday.com/blog/overcoming-addiction/201307/drink-your-way-sober">https://www.psychologytoday.com/blog/overcoming-addiction/201307/drink-your-way-sober</a>	
alcoholism	It took me being 6,000 in debt to realize I needed help. After battling alcohol for a year I finally had a mental breakdown while drunk and I am getting kicked out of my dorm. Truth	
alcoholism	I feel like I want to crawl out of my skin. I've got a bit more than 9 months sober right now. My sobriety isn't at risk; this sub and the rooms of AA have helped me learn how to ask	
alcoholism	Boyfriend has a Problematic Relationship with Alcohol My boyfriend decided to have a dry January to assess his relationship with alcohol and hopefully come up with a plan to drin	
alcoholism	Alcoholic- Wernicke? Hi All,First time poster here. I am hoping someone can help me.I have had concerns about my partners's drinking for quite some time. For some background,	
alcoholism	First time posting, I could use some advice...(trigger warning possibly) -long Hello all,First time poster but I've been reading this sub for a while. Here's a short backstory: I ha	
alcoholism	Should I go to the doctor? For the past two years I drink a 12 pack or half a fifth daily. I want to quit but should I go to the doctor and get some type of medication to help with wit	
alcoholism	My SOs alcoholic brother is being pushed away by his family My SOs brother is an alcoholic and suffers from depression. He has been hospitalized in the past, has almost gone throu	
alcoholism	Year and few days sober I started uploading videos because when i got sober it was the time i got myself into mental hospital so i will upload videos of that and also alcoholism and	
alcoholism	Irritable Am wanting to quit drinking, but I get so bloody irritable when I do. My wife and kids suffer as a result. Anyone know how long this lasts or when I can hope for it to pass?	
alcoholism	1 year sober today I can't believe I made it this far. As of today, January 3, 2018, I have been sober for 1 year. I still struggle every day. I think about drinking multiple times a day	

# Subreddits

Data from these subreddits were used-

- Alcoholism
- Anxiety
- Bipolar
- Depression
- Health Anxiety
- PTSD
- Schizophrenia
- Social Anxiety
- Suicide watch
- Mental Health

# Reddit Data Distribution



# Process and ML Models

- Social media Data from Reddit's top subreddit mental health channels were used to form a text corpus.
- Reddit data was cleaned- Converted to lowercase, Punctuations removed, Tokenized, Lemmatized , Stop words removed and text was processed.
- The processed text was trained for modelling using TF-IDF Vectorizer and n-grams.
- This text corpus was used to train a Naïve Bayes model & a Multinomial Logistic Regression model.
- Used the better performing trained multinomial logistic regression model with 74% accuracy to classify our twitter data of 408K tweets.

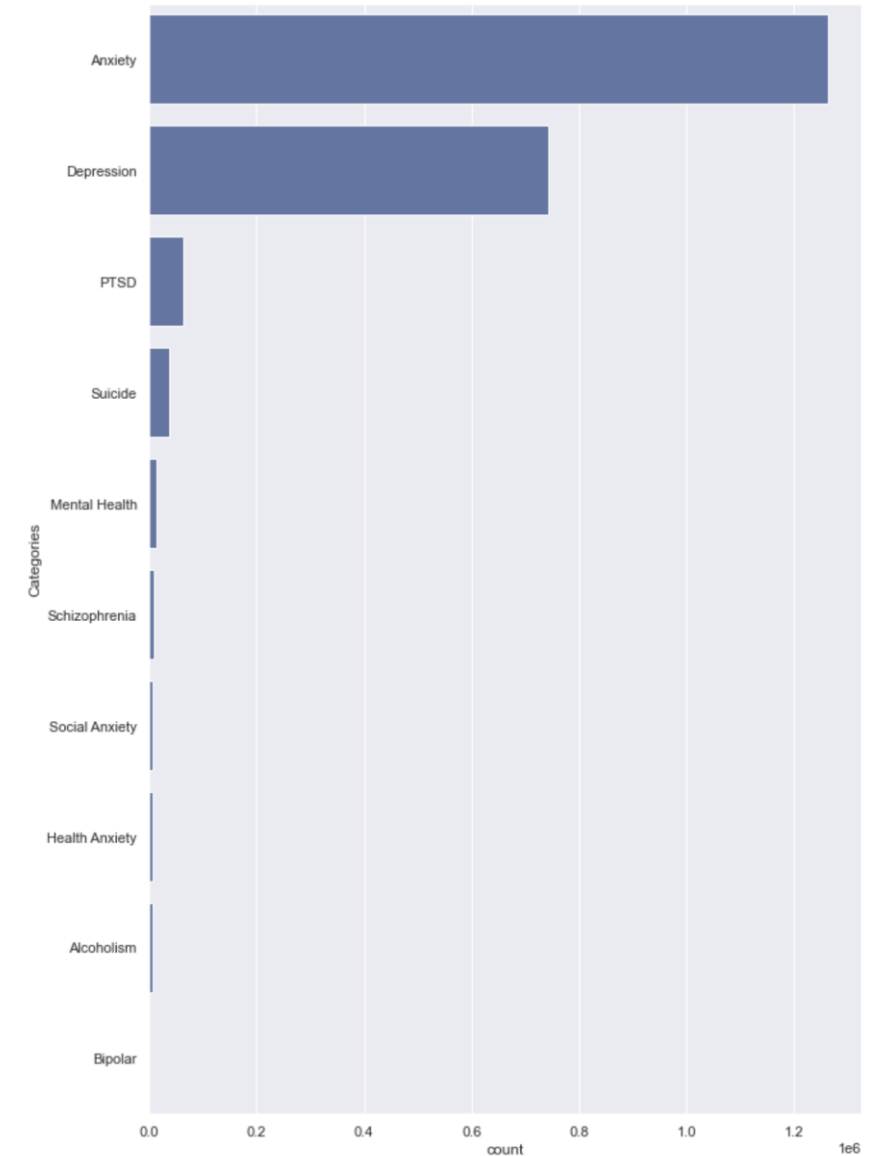
MODEL	ACCURACY
Naïve Bayes	60%
Multinomial Logistic Regression	74%

- Tweets were cleaned and processed using same approach using NLP and then the trained model was used to classify the tweets into different mental illness/issue categories. Analysis in the next slide.



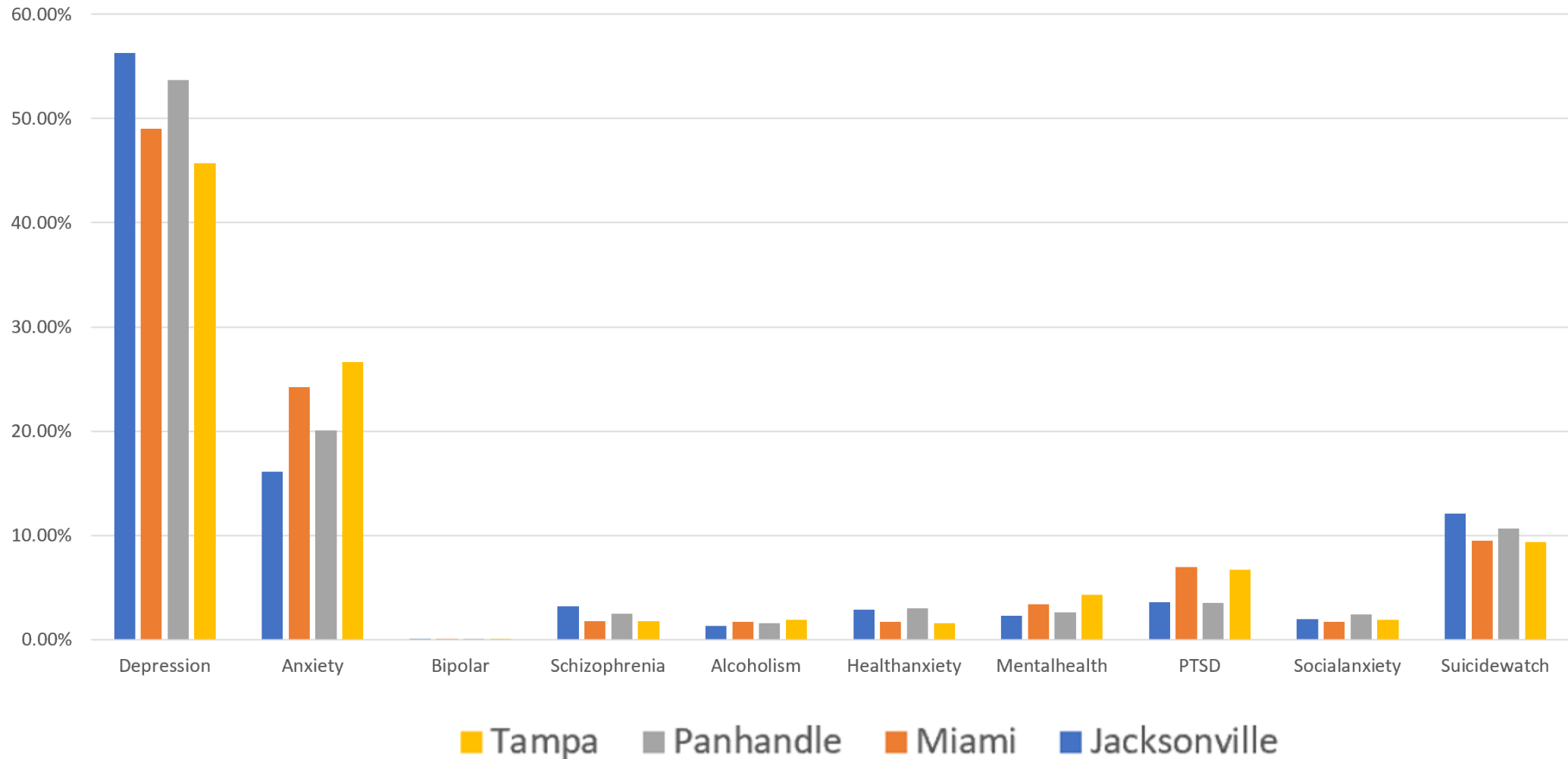
# Overall Prediction Distribution

TYPE	PERCENTAGE
Anxiety	58.78%
Depression	34.56%
PTSD	2.97%
Suicide	1.70%
Mental Health	0.64%
Schizophrenia	0.41%
Social Anxiety	0.31%
Health Anxiety	0.31%
Alcoholism	0.30%
Bipolar	0.02%

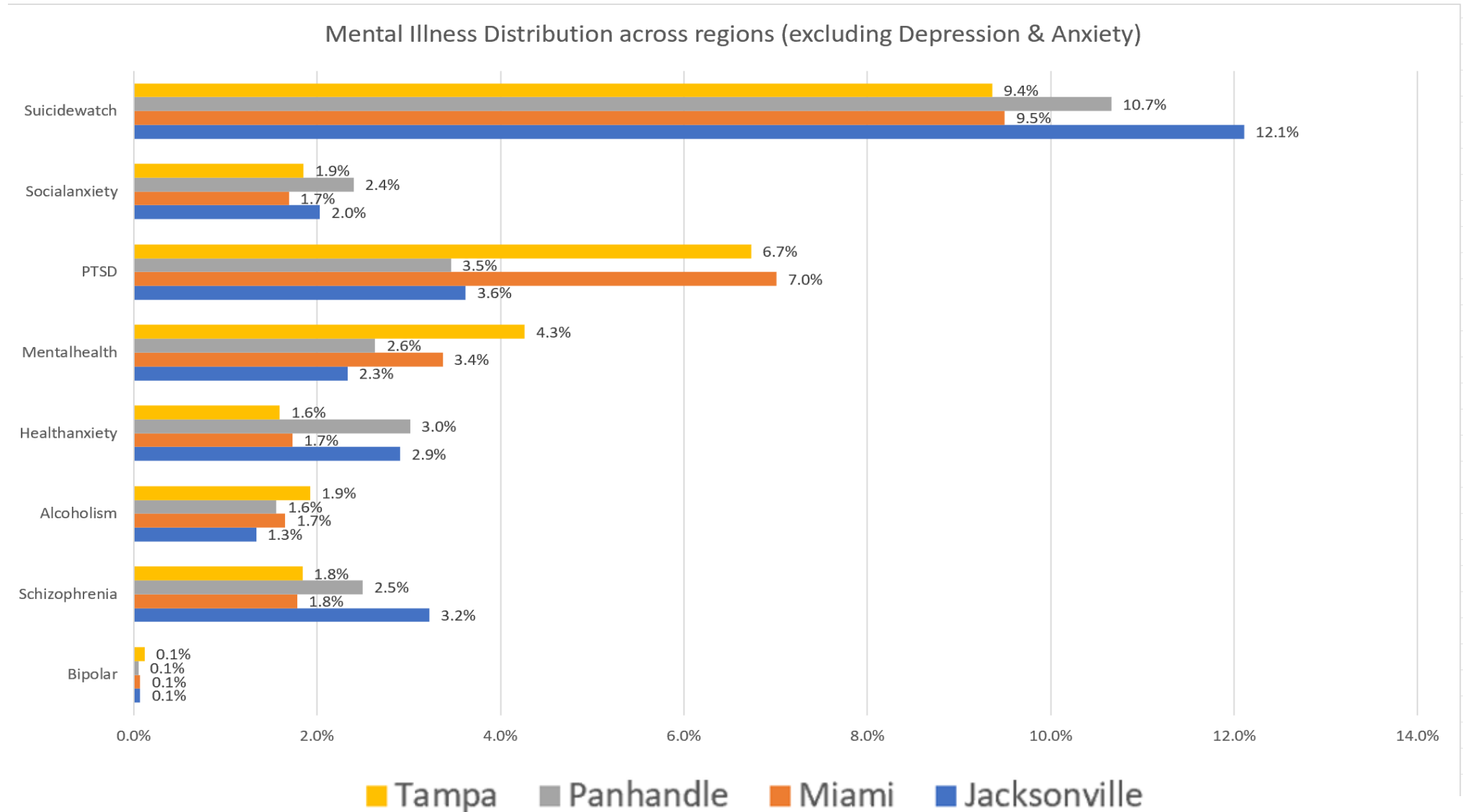


# Mental Illness – Across Regions

Mental Illness Distribution by Region



# Mental Illness – Across Regions (excluding Depression & Anxiety)



# Mental Illness – Key Takeaways

- High scale detection of Depression & Anxiety in all regions, as this is the most common detected and diagnosed.
- Jacksonville cluster has highest suicide prediction – 12.1%, followed by the Panhandle, at 10.7% in our samples.
- PTSD has highest presence in the Miami cluster - 7%, followed by Tampa, at 6.7%.
- Highest Schizophrenia was predicted for the Jacksonville cluster over others.
- Panhandle cluster has the Health Anxiety, possibly owing to poor healthcare infrastructure availability.
- Tampa region has the highest Mental Health & Alcoholism concerns over the other clusters, possibly a social lifestyle effect and correlated to mental illness, not causation.



# Recommendations

- **Increase Suicide Prevention lines and awareness**, mainly in the local **Jacksonville and Panhandle** clusters.
- Conduct periodic **surveys** for mental well-being and behavioral health.
- **Collect this user survey data over a time period**, **understand** mental health **patterns** and **recommend** specific health insurance plans to support them based on the highest ones in each cluster.
- More health insurance options with **specific** behavioral therapy, group options and **initial free sample sessions** to encourage patient visit and sign up.
- Conduct Behavior **therapy sessions** for those who suffer from depression & PTSD, mainly in **Miami and Tampa clusters**.
- Conduct free/reduced cost counselling sessions to fight Anxiety **with support groups**.
- Mandating few **PTO days off for mental health** and avoiding burnout/stress.
- Investment in treatment of mental illnesses has 4 times the ROI. (<https://www.who.int/news/item/13-04-2016-investing-in-treatment-for-depression-and-anxiety-leads-to-fourfold-return> )

# APPENDIX

- References of Papers
  - Data Visualizations
- Model Prediction Output Numbers



UNIVERSITY of  
**SOUTH FLORIDA**

# Thank you!



UNIVERSITY of  
**SOUTH FLORIDA**

# Literature Review & Papers

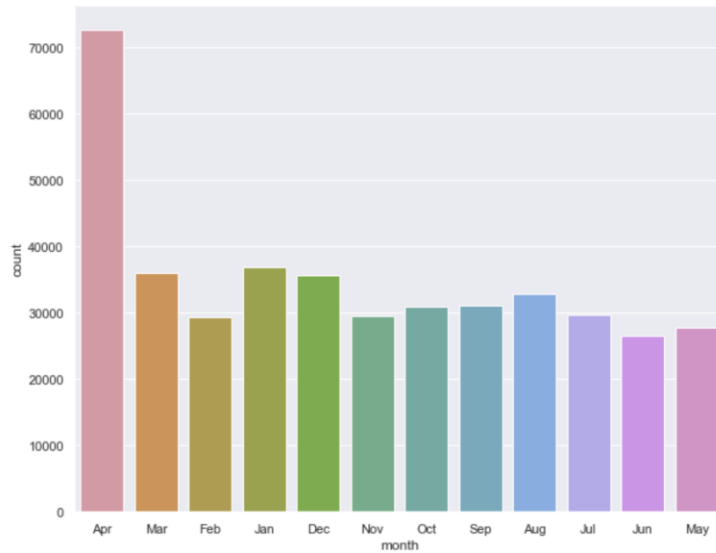
Here are the papers that were used as reference to understand the domain, problem and methodology.

PAPER	LINK
Ananthakrishnan, Uttara M and Tucker, Catherine E., <b>The Drivers and Virality of Hate Speech Online</b> (January 28,2022)	<a href="https://ssrn.com/abstract=3793801">https://ssrn.com/abstract=3793801</a> or <a href="http://dx.doi.org/10.2139/ssrn.3793801">http://dx.doi.org/10.2139/ssrn.3793801</a>
Sharath Chandra Guntuku, David B Yaden, Margaret L Kern, Lyle H Ungar, Johannes C Eichstaedt, <b>Detecting depression and mental illness on social media: an integrative review</b> , Current Opinion in Behavioral Sciences, Volume 18, 2017, Pages 43-49, ISSN 2352-1546	<a href="https://doi.org/10.1016/j.cobeha.2017.07.005">https://doi.org/10.1016/j.cobeha.2017.07.005</a> or <a href="https://www.sciencedirect.com/science/article/pii/S2352154617300384">https://www.sciencedirect.com/science/article/pii/S2352154617300384</a>
<b>NLP Model for Effect of Social Media on Health</b> . Duncan, Brianni	<a href="http://hdl.handle.net/1920/12187">http://hdl.handle.net/1920/12187</a>
Conway M, O'Connor D. <b>Social Media, Big Data, and Mental Health: Current Advances and Ethical Implications</b> . Curr Opin Psychol. 2016 Jun;9:77-82	doi: 10.1016/j.copsyc.2016.01.004. PMID: 27042689; PMCID: PMC4815031.
Babu, N.V., Kanaga, E.G.M. <b>Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review</b> . SN COMPUT. SCI. 3, 74 (2022).	<a href="https://doi.org/10.1007/s42979-021-00958-1">https://doi.org/10.1007/s42979-021-00958-1</a>
Coppersmith G, Leary R, Crutchley P, Fine A. <b>Natural Language Processing of Social Media as Screening for Suicide Risk</b> . Biomedical Informatics Insights. January 2018	doi:10.1177/1178222618792860
Kim, J., Lee, J., Park, E. et al. <b>A deep learning model for detecting mental illness from user content on social media</b> . Sci Rep 10, 11846 (2020).	<a href="https://doi.org/10.1038/s41598-020-68764-y">https://doi.org/10.1038/s41598-020-68764-y</a>

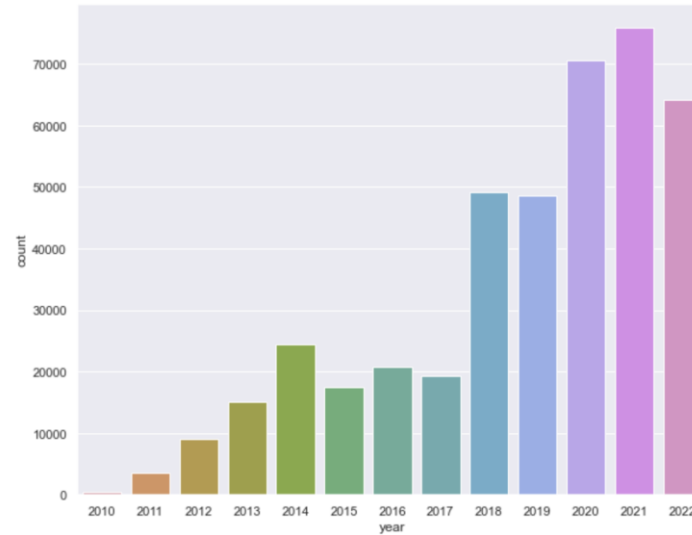


# Data Visualizations – Tweet Distribution

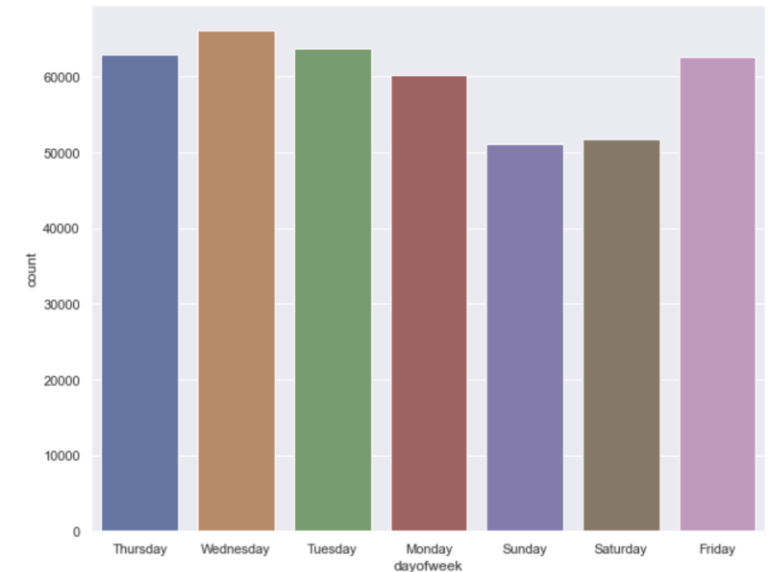
## By Month



## By Year

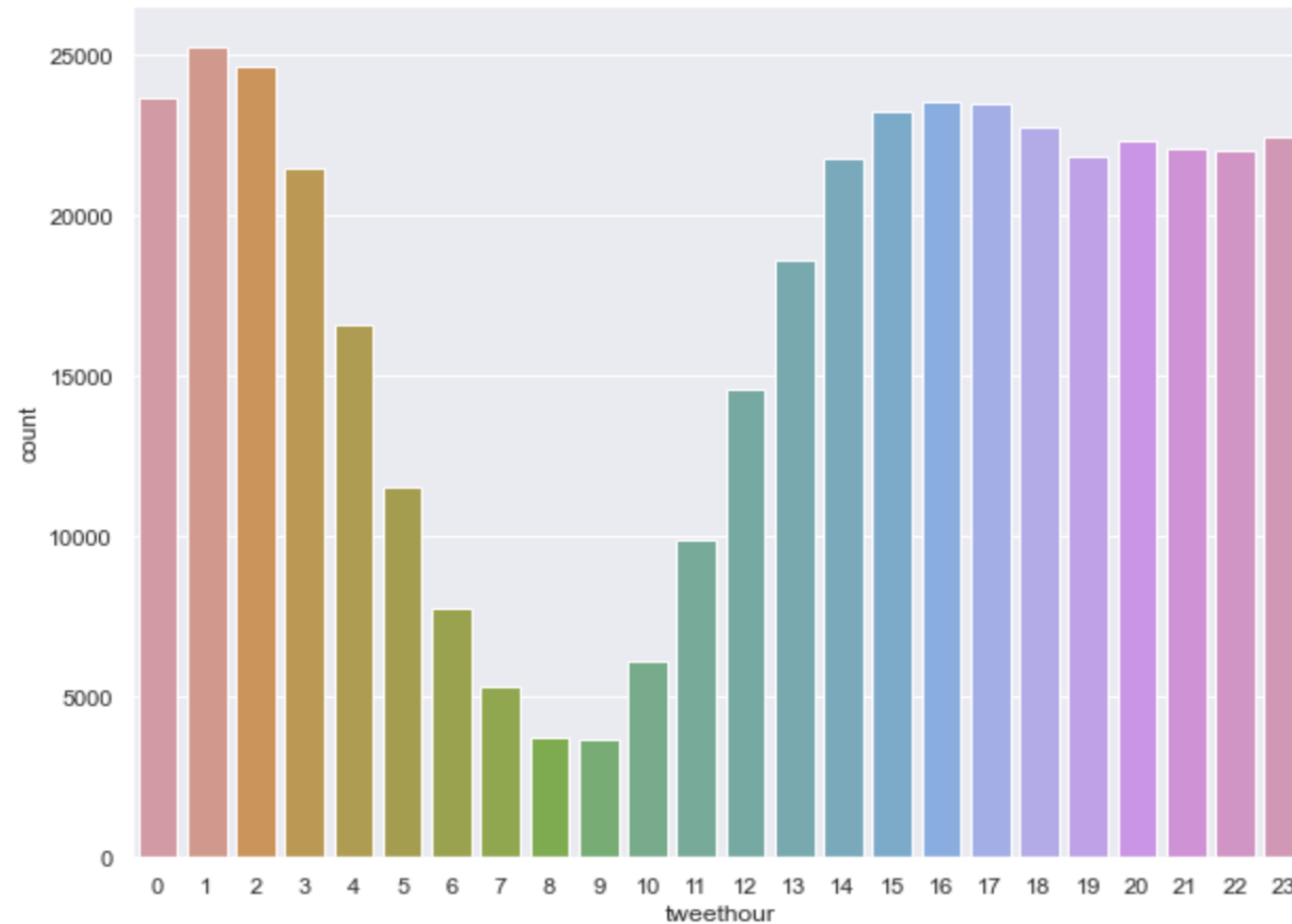


## By Day of the Week



# Data Visualizations – Tweet Distribution(contd.)

By Hour of the day



# Mental Illness - Numbers vs Percentage

					PREDICTION IN NUMBERS						
REGION	Depression	Anxiety	Bipolar	Schizophrenia	Alcoholism	Healthanxiety	Mentalhealth	PTSD	Socialanxiety	Suicidewatch	TOTAL TWEETS
Jacksonville	18029	5140	23	1033	428	930	746	1159	650	3876	32014
Miami	63388	31220	94	2307	2138	2238	4358	9061	2189	12279	129272
Panhandle	17611	6581	17	820	511	989	864	1135	787	3498	32813
Tampa	63847	37228	165	2574	2690	2229	5963	9413	2591	13087	139787
				PREDICTION IN PERCENTAGES							
REGION	Depression	Anxiety	Bipolar	Schizophrenia	Alcoholism	Healthanxiety	Mentalhealth	PTSD	Socialanxiety	Suicidewatch	TOTAL TWEETS
Jacksonville	56.3%	16.1%	0.1%	3.2%	1.3%	2.9%	2.3%	3.6%	2.0%	12.1%	100.0%
Miami	49.0%	24.2%	0.1%	1.8%	1.7%	1.7%	3.4%	7.0%	1.7%	9.5%	100.0%
Panhandle	53.7%	20.1%	0.1%	2.5%	1.6%	3.0%	2.6%	3.5%	2.4%	10.7%	100.0%
Tampa	45.7%	26.6%	0.1%	1.8%	1.9%	1.6%	4.3%	6.7%	1.9%	9.4%	100.0%

- High scale detection of Depression & Anxiety in all regions, as this is the most common detected and diagnosed.
- Jacksonville cluster has highest suicide prediction – 12.1%, followed by the Panhandle, at 10.7% in our samples.
- PTSD has highest presence in the Miami cluster - 7%, followed by Tampa, at 6.7%.
- Highest Schizophrenia was predicted for the Jacksonville cluster over others.
- Panhandle cluster has the Health Anxiety, possibly owing to poor healthcare infrastructure availability.
- Tampa region has the highest Mental Health & Alcoholism concerns over the other clusters, possibly a social lifestyle effect. – correlation , not causation.