

Topic- NLP to detect mental illnesses among social media users

ISM 6905 –Independent Study

Dr. Balaji Padmanabhan

MENTOR: Negar Maleki

Student: Adhithyan Rangarajan



Overview

- **Step 1 – Literature Review-** Performed a literature review ,understood problem significance & methodology used by researchers on predicting mental illness using social media data.
- **Step 2 – Methodology for Data Retrieval** – Understood the population density weightage in Florida and pulled data from twitter based on it.
- **Step 3 – Twitter Test Data Sample Volume** – 408000 tweets in our sample pulled using Geo-location and City Names.
- **Step 4 – Reddit Mental Health Data for training** – 344000 reddit mental health posts were used to train our ML models.
- **Step 5 – ML Models** - Ran the trained better ML model on twitter data for prediction for testing.
- **Step 6 - Analysis & Recommendations** for the State of Florida & Florida Blue.

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

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.

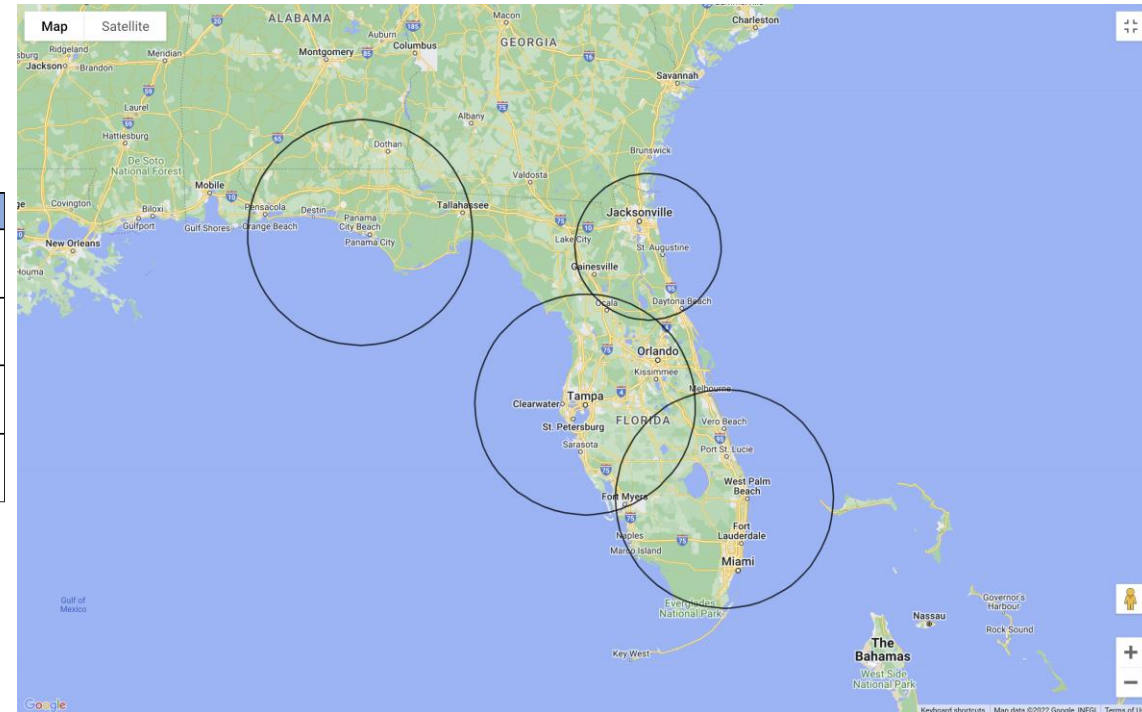
Radius and Geolocation View of dense populations in Florida-

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

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



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 Disorder	8388	8616	38	105
	Schizophrenia	3576	265	87	185
	Depress	557	351	6	18
	Covid	6086	2750	278	170
		10000	10000	10000	10000
TOTAL		109774	72407	13752	17960
		FINAL TOTAL		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
	Trauma	2402	1992	322	162
	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 Disorder	1464	4013	465	456
	Schizophrenia	668	1300	194	151
	Depress	81	168	26	4
	Covid	3157	5925	1271	993
		10000	10000	3068	2289
TOTAL		67690	86336	23346	17094
		FINAL TOTAL		194466	B

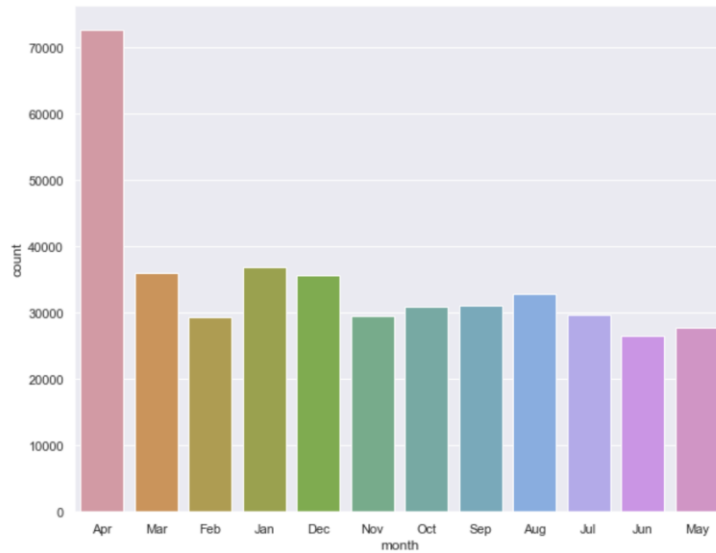
Twitter Dataset Sample

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

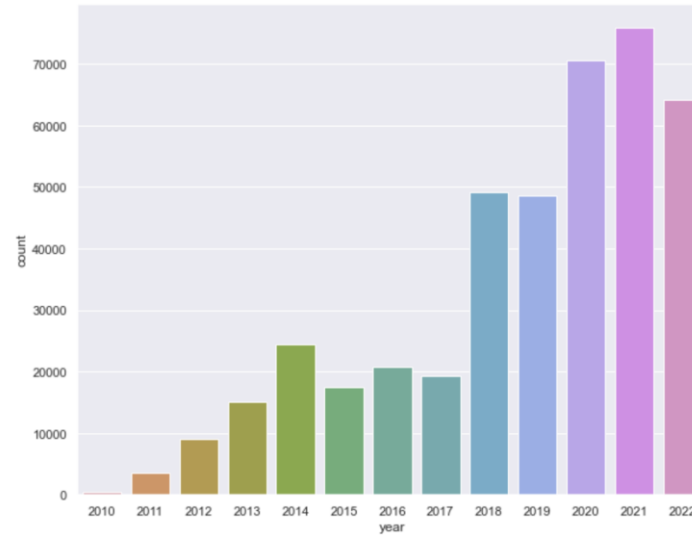
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	url	date	content	username	keyword	date	time	month	tweethour	day	year	weekend	region	
38	https://twitter.com	2022-04-11 22:00	@CapaKAN I can see that causing you stress. At the end of th	GyrexGG	stress	4/11/2022	10:00:20 PM	Apr		22 Monday	2022 Weekday		Tampa	
39	https://twitter.com	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	

Basic Visualizations – Tweet Distribution

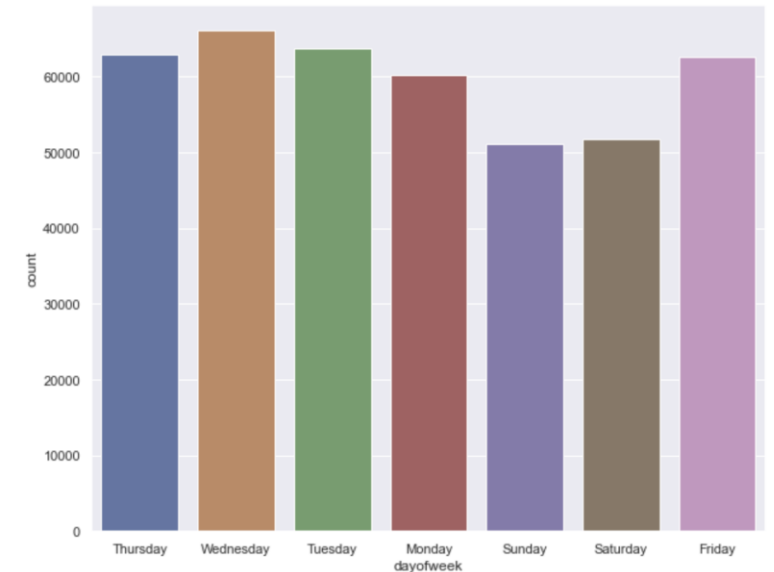
By Month



By Year

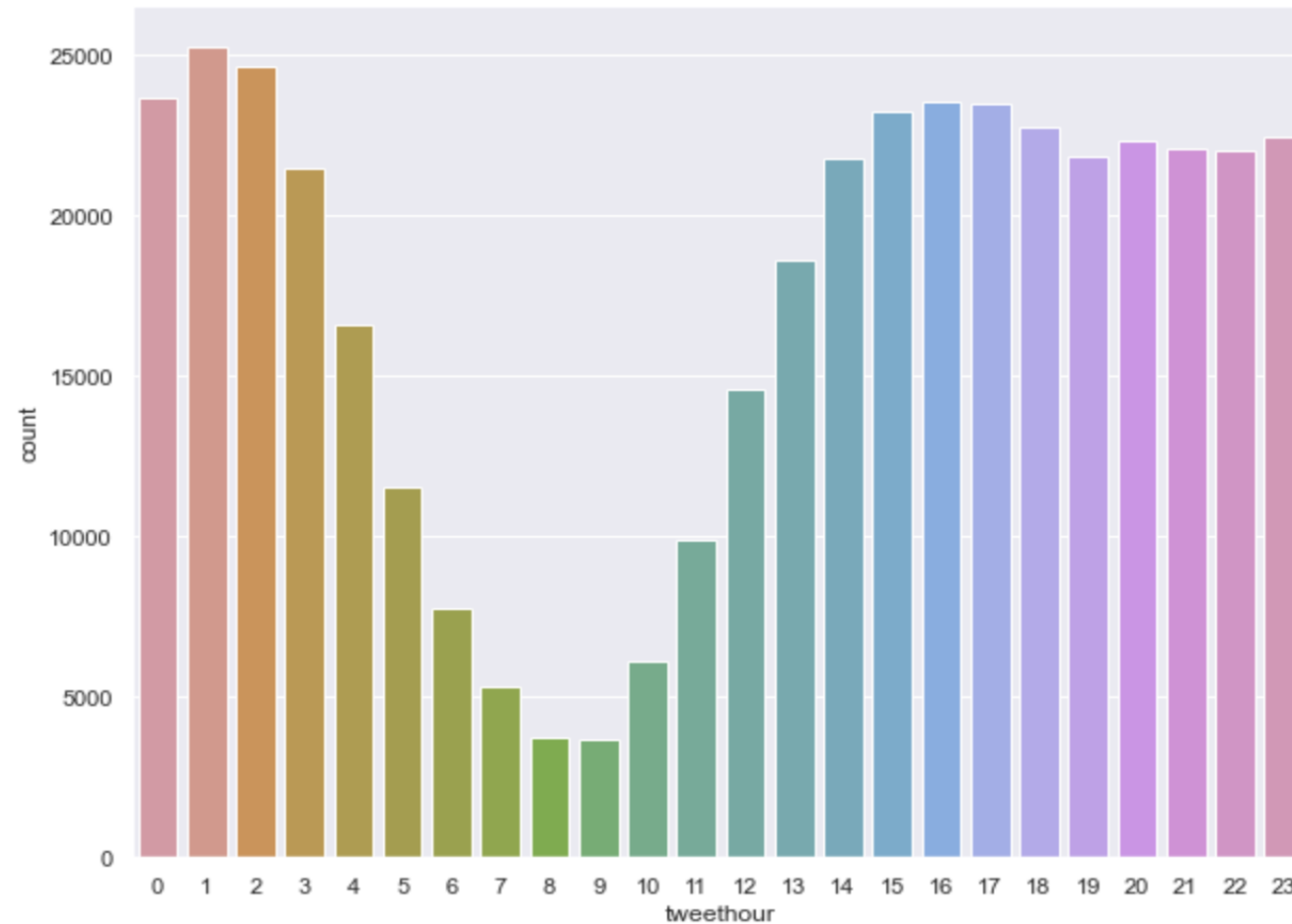


By Day of the Week



Basic Visualizations – Tweet Distribution(contd.)

By Hour of the day



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 –

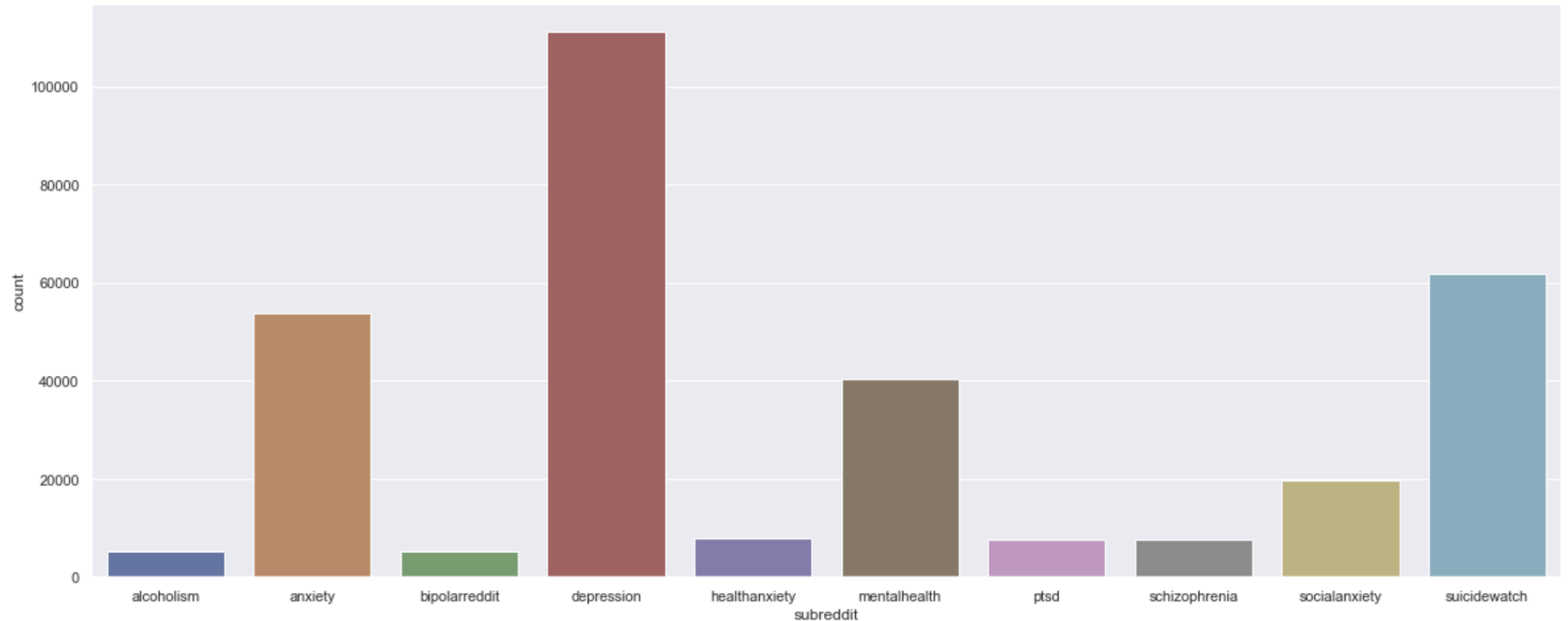
A	B	C
subreddit	post	
alcoholism	I need to quit while im young (22m) Im going through some hardcore withdrawals. I have anxiety as it is, but i just feel like im dying. Should i go to the hospital? I heard you can die	
alcoholism	It's 1:30am. So glad to be sober. I was at a New Years party tonight. All 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 https://www.psychologytoday.com/blog/overcoming-addiction/201307/drink-your-way-sober	
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

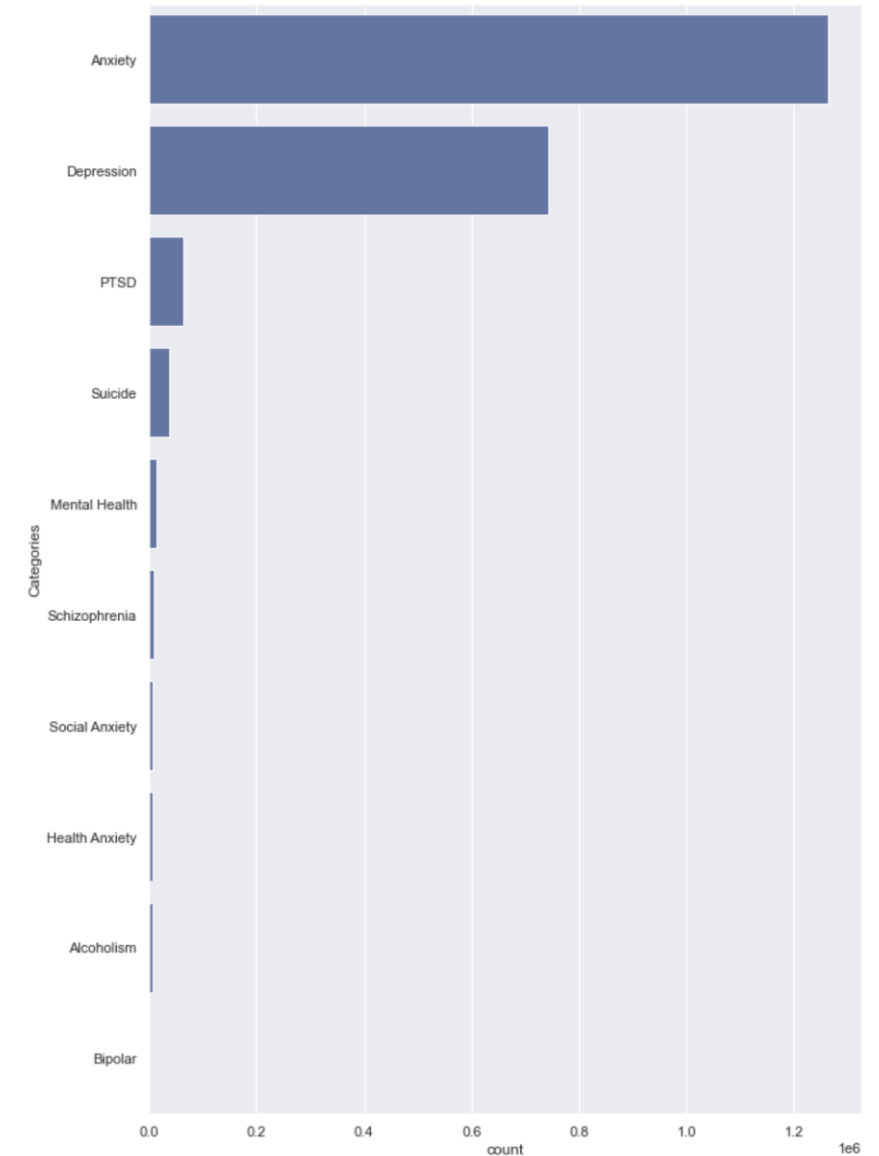
- Social media Data from Reddit's top subreddit mental health channels were used to form a text corpus.
- Reddit data was cleaned- Converted to lowercase, Punctuations removed, Tokenized, Lemmatized , Stop words removed and text was processed.
- The processed text was trained for modelling using TF-IDF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

MODEL	ACCURACY
Naïve Bayes	60%
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.

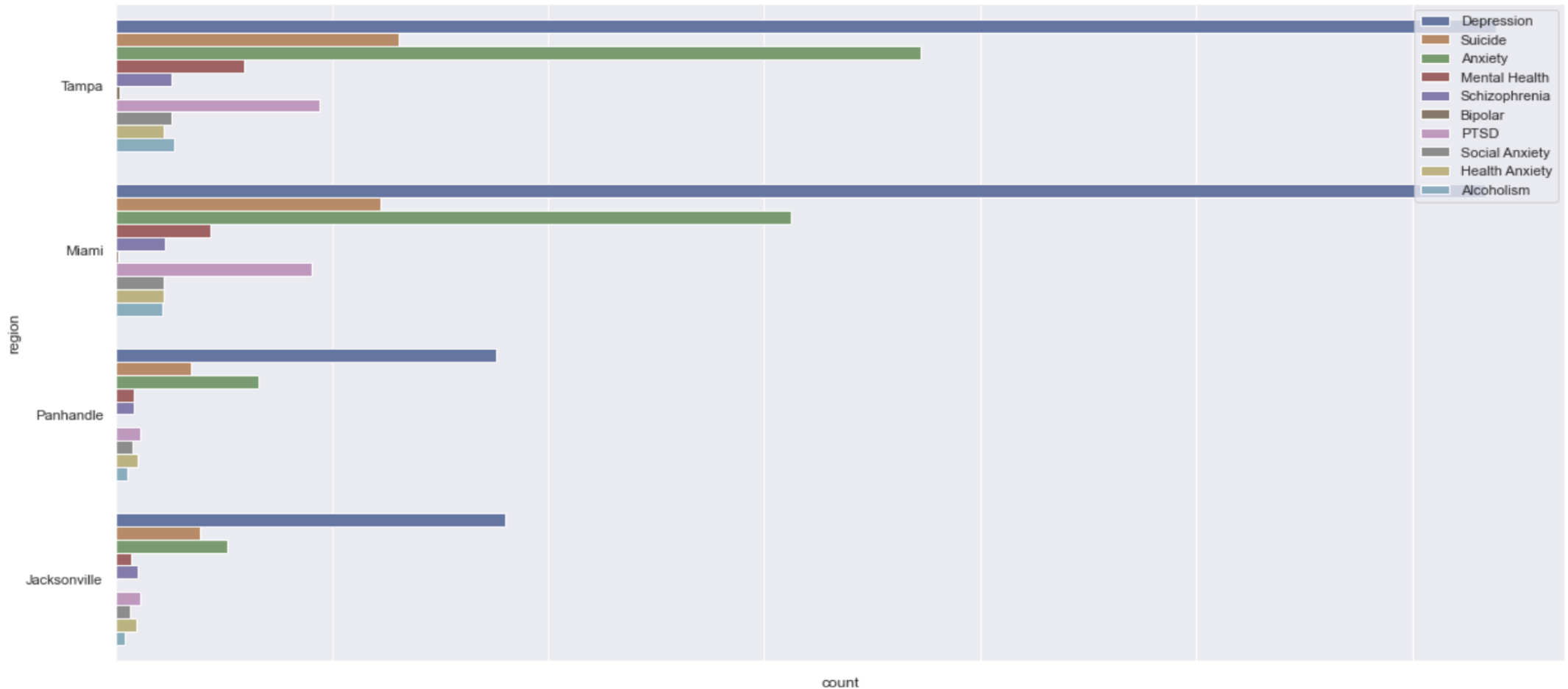
Overall Prediction Distribution

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%

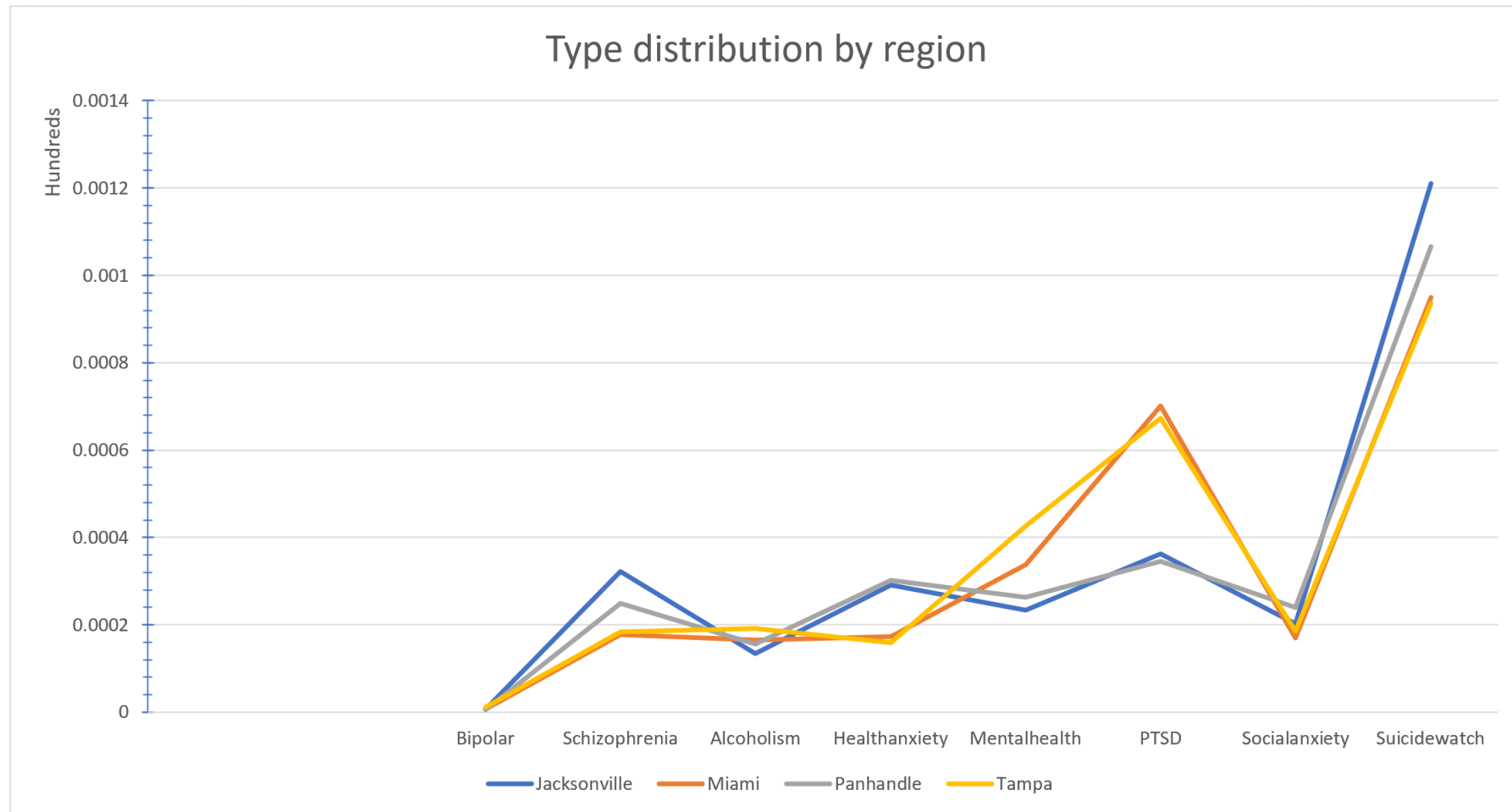


Prediction by region

- Depression and Anxiety had the highest percentages in each region, way above others. Graphical representation below-



Distribution of Mental Illness type by region



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.
- 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.
- Tampa region has the highest Mental Health & Alcoholism concerns over the other clusters.

Recommendations

- Increase Suicide Prevention lines, mainly in the local **Jacksonville and Panhandle** areas.
- 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.
- 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.
- 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>)

Thank you!



UNIVERSITY of
SOUTH FLORIDA