

Social Media Analysis

Mukunda Aithal, Cindy Huang, Luyu Huang,
Victor Wang, Kelly Wang

Agenda

1. Question and Hypothesis
2. Exploratory Data Analysis
3. Consumer Analysis
4. General CGM Analysis
5. Product-Specific CGM Analysis
6. Recommendations

Problem & Hypothesis

What do people think of CGM?

- Hypothesis: People agree CGM is an easy solution to manage diabetes

What features should CMG companies focused on improving?

- Hypothesis: Accuracy, Price, Usability

Exploratory Data Analysis

The dataset contains post-level data about post, repost, author and LexisNexis features

- **37,844** post-level data encompassing 63 features that fall into the following 7 categories
 - **Post content** (e.g. Sound Bite Text)
 - **Post sentiment** (e.g. 'Sentiment', 'Positive Objects', 'Negative Objects',
 - **Post popularity** (e.g. 'Followers/Daily Unique Visitors/Subscribers', 'Total Engagements')
 - **Post metadata:** (e.g. 'Source Type', 'Post Type', 'Is Paid', 'Media Type', 'URL', 'Media Link', 'Domain', 'Published Date (GMT-04:00) New York')
 - **Repost information** (e.g. 'Quoted Post', 'Quoted Author Name')
 - **Author account and demographics information:** (e.g. 'Author Gender', 'Author URL')
 - **LexisNexis information***: (e.g. 'LexisNexis Source Publisher')
- Among the 63 features, 24 features have only missing values, 18 features have 95%-100% values missing

*Columns contain only missing values

Low data availability features tend to be source specific and can be informative

Source	Author	Popularity Metrics	Repost
Blogs	Gender, Name, Handle, Location (Country level)	N/A	N/A
Comments	ID	Followers/Daily Unique Visitors/Subscribers	
Facebook	Location	Engagement, Comment, Like	N/A
Forums	Gender, Name, Handle, Reddit Karma	Followers/Daily Unique Visitors/Subscribers	N/A
Instagram	Location (Country level)	Engagement, Comment, Like	N/A
Twitter	Gender, Name, Handle, Location, Interests	Followers/Daily Unique Visitors/Subscribers	Available

Among the categorical features, we observe some data imbalance from multiple dimensions

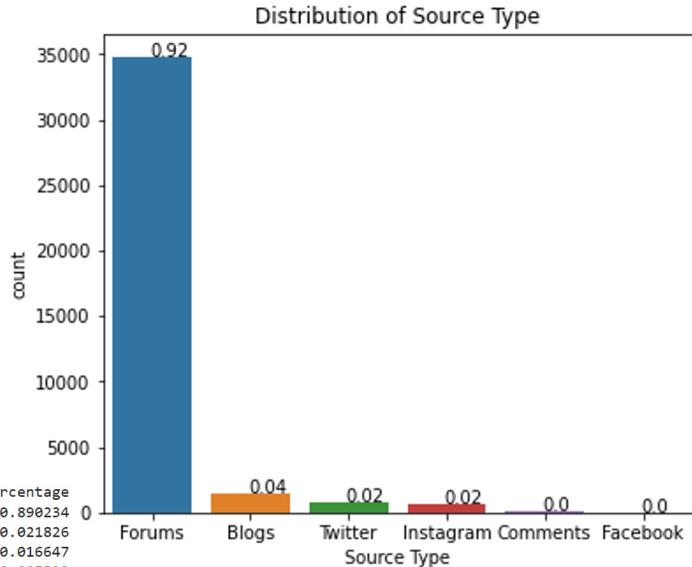
- **The majority of the posts come from forums, specifically reddit.** The discussions on forums occur the most in diabetes-related forums.
- **The majority of the posts are replies and comments,** which may be because most forum posts are replies and comments.

	Source Name_count	Source Name_percentage	
r/diabetes_t1	13284	0.394301	reddit.com
r/diabetes	9191	0.272811	twitter.com
r/Type1Diabetes	5887	0.174740	instagram.com
r/dexcom	663	0.019679	forum.tudiabetes.org
r/diabetes_t2	487	0.014455	community.whattoexpect.com
...
r/Niacin	1	0.000030	trentonryflf.canariblogs.com
r/HamptonRoads	1	0.000030	pebforum.com
r/czech	1	0.000030	kj1cbd.amoblog.com
r/MillennialBets	1	0.000030	hsn.com
r/hyderabad	1	0.000030	systematicreviewsjournal.biomedcentral.com

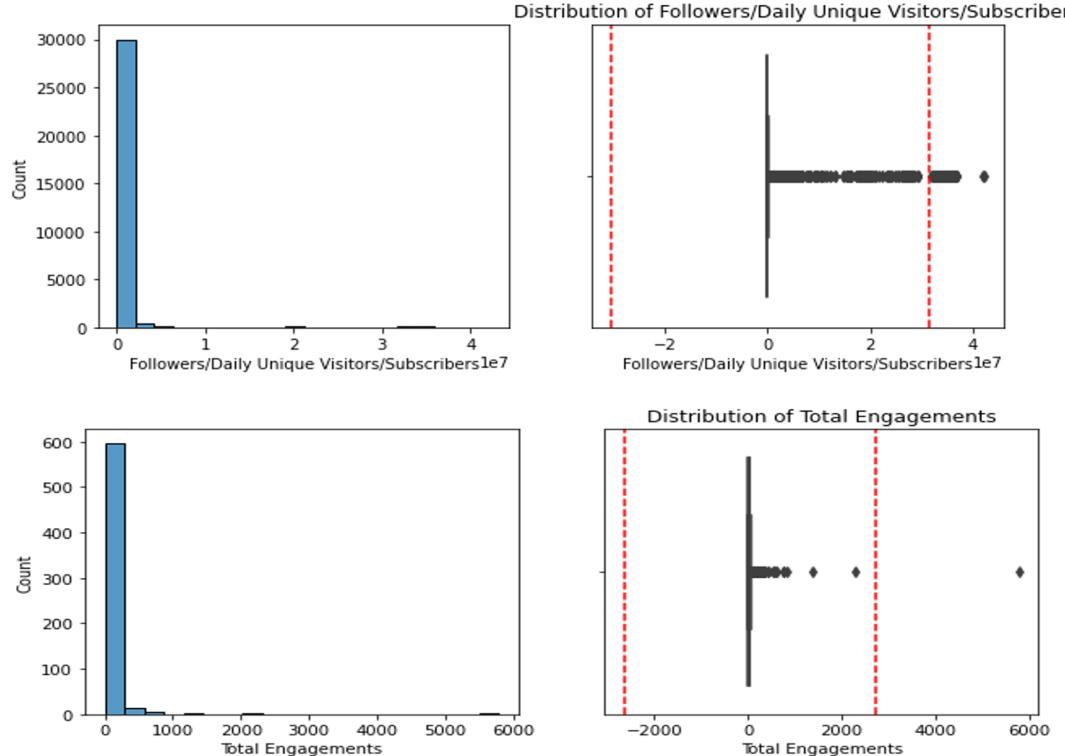
[924 rows x 2 columns]

	Domain_count	Domain_percentage	
Forums	33690	0.890234	
Blogs	826	0.021826	
Twitter	630	0.016647	
Instagram	209	0.005523	
Comments	146	0.003858	
Facebook	
	1	0.000026	
	1	0.000026	
	1	0.000026	
	1	0.000026	
	1	0.000026	

[1049 rows x 2 columns]



All of the numerical columns (mostly related to post popularity) observe long-tailed distributions

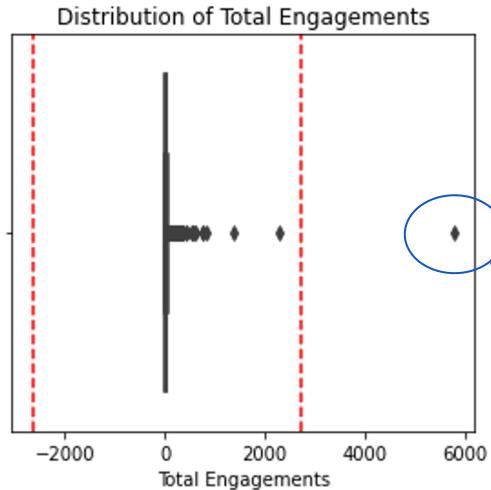
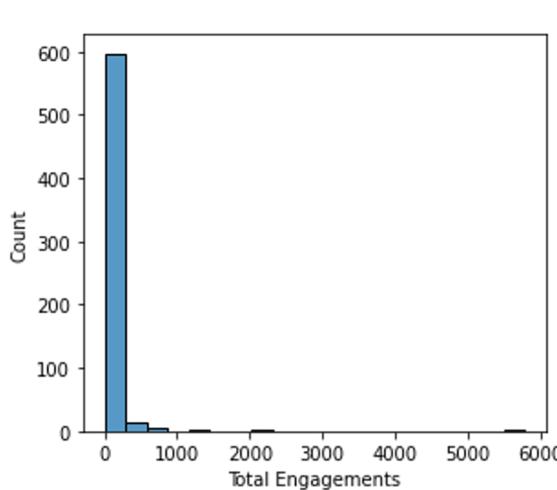


As expected, most of the posts have lower engagement, and only very few are “super posts” with a huge number of engagements, followers, likes, comments, etc.

The red cut-offs are 10 standard deviations away from the mean.

(Not?) Dealing with outliers

- In our context the outlier posts may require the most attention of the business stakeholders, as these are posts that may be representative or have bigger reach
- We do not drop these outliers and instead seek to understand the themes of the sample texts



“...today as a part of the Pro team, I am excited to bring to you a product that has been tested & proven to make change happen sustainably????”
- 5,782 total engagement

Sampled texts show that some focuses of the trending posts include insurance and affordability, lifestyle influencing, and app product issues

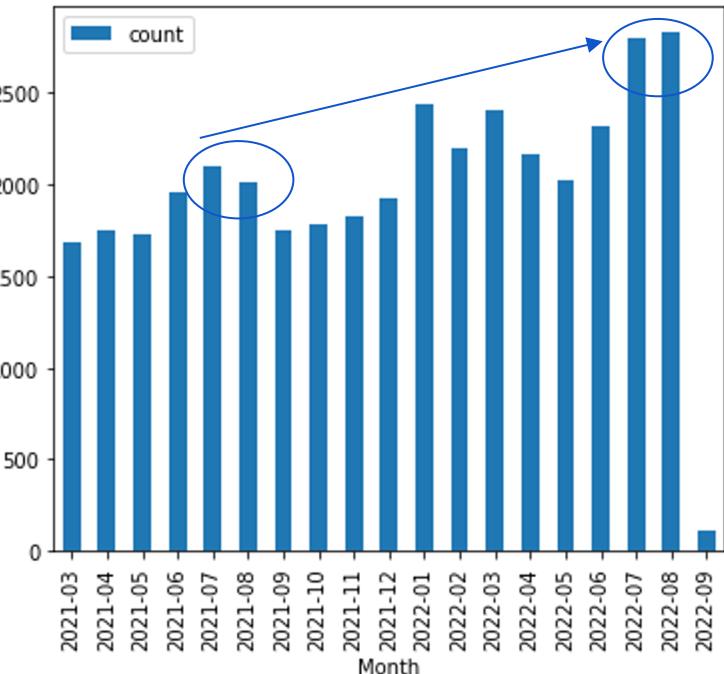
"...It costs me about 650\$ a month to be diabetic and run the system that can maintain my sugar levels. It's a joke really" - 33,463,935 Followers/Daily Unique Visitors/Subscribers

"... I went back to the g66 app for omnipot 5 and I couldn't get it to restart... I don't have a fancy phone so I just simply it was a complaint for me for somebody who has those kind of phones obviously is not a problem but for someone like me who can't afford a phone like that it is an issue" - Richness = 15
(Max value)

"...today as a part of the Pro team, I am excited to bring to you a product that has been tested & proven to make change happen sustainably????" - 5,720 likes

"...Stay tuned for me to share all of my thoughts and findings after wearing this continuous glucose monitor for 2 weeks. #sugar #dietitian #healthtips #nutritiontips #insulin #womenshealth #bloodsugar #health #nutrition" - 308 comments

Social media discussion volumes see an upward trend over time and spikes in the summer months



- There is a general upward trend of the number of posts in each month. This shows that the topic is gaining traction and discussions.
- We see a particularly increased number of posts during the summer months (July and August). This may be potentially due to people's attention to lifestyle and diet, as they may go out and consume ice cream more.

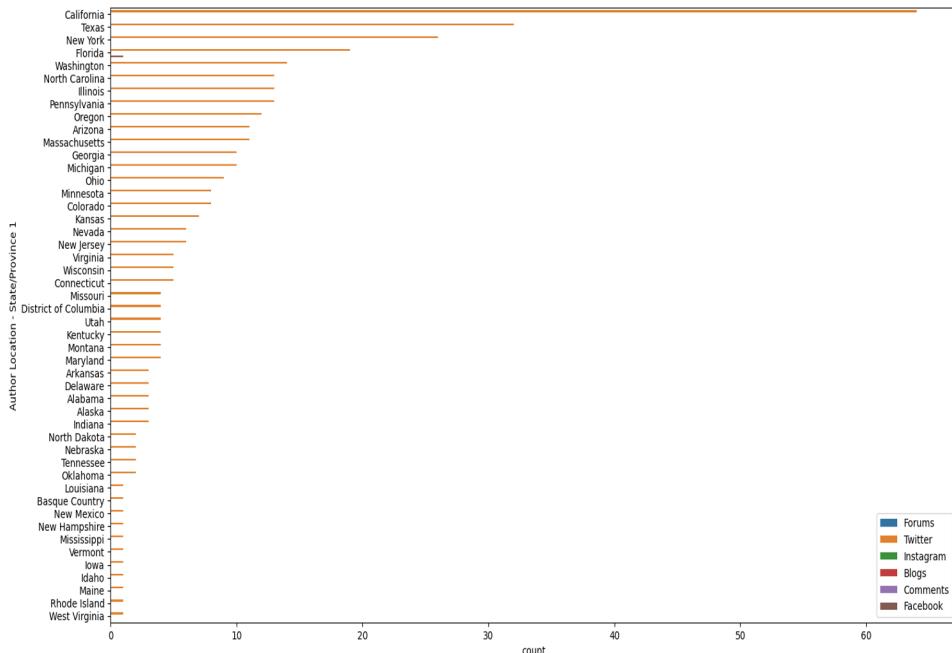
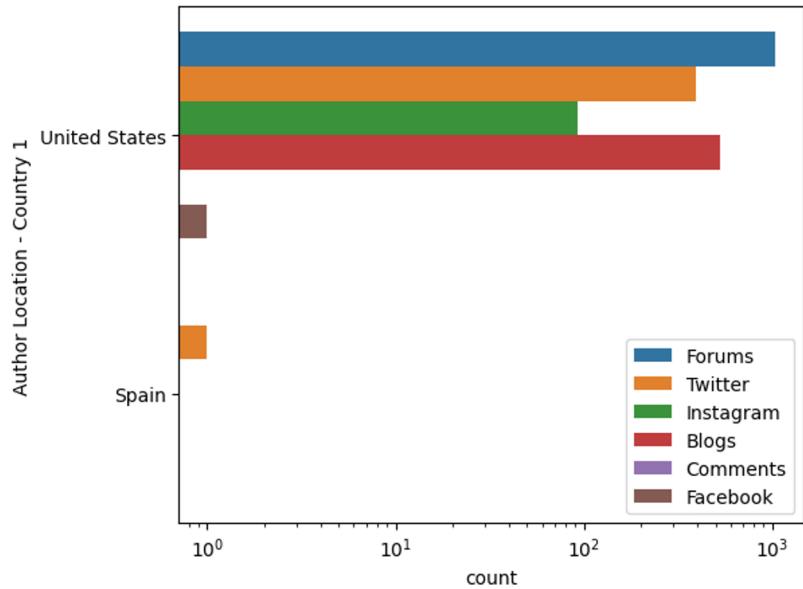
EDA Summary

1. We observe **low to no data availability** across the majority of columns. Among the 63 features, 24 features have only missing values, 18 features have 95%-100% values missing.
2. A main reason why many columns display low data availability is that **the dataset contains posts from multiple sources** with different metrics and information collected. About **92%** of the posts were sourced from **forums**. About **89%** were from **Reddit**.
3. All of the **numerical** columns (mostly related to post popularity) observe **long-tailed distributions**. In our context the outlier posts may require the most attention of the business stakeholders, as these are posts that may be representative or have bigger reach. We do not drop these outliers and instead seek to understand the themes of the sample texts. **Sampled texts show that some themes of the trending posts include insurance and affordability, lifestyle influencing, and app product issues.**
4. Social media **discussion volumes** have an **upward trend over time** and observe **spikes in the summer months**.

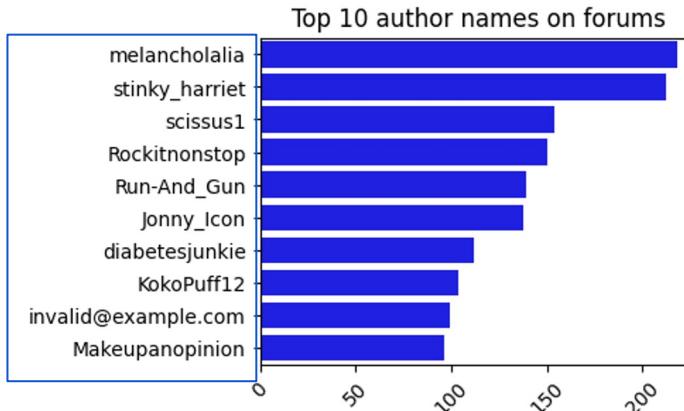
Consumer Analysis

Given what we can observe from the dataset, people who post about CGM are mostly in the US and specifically CA, TX, NY, FL, and WA

- Among Twitter post authors, the top 5 source states are: California, Texas, New York, Florida, and Washington. Those states may be where most of the CGM consumers come from.



The super authors are mostly patients or patient communities sharing device usage experience, monitoring and lifestyle tips, pricing



*"i had the guardian sensor for a bit before switching to tslim and dexcom. that thing f**king HURT to insert. felt so violent every time."*

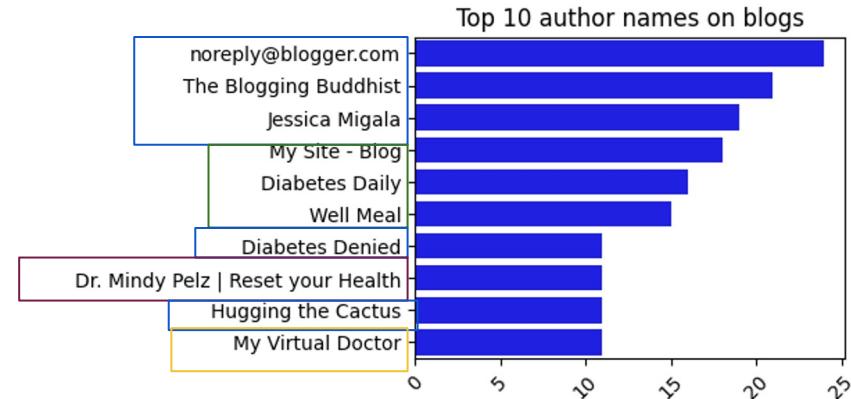
"I rely on my CGM to wake me for lows because I don't feel them."

Patients

Patient
Community

Medical
Scientist

Noise

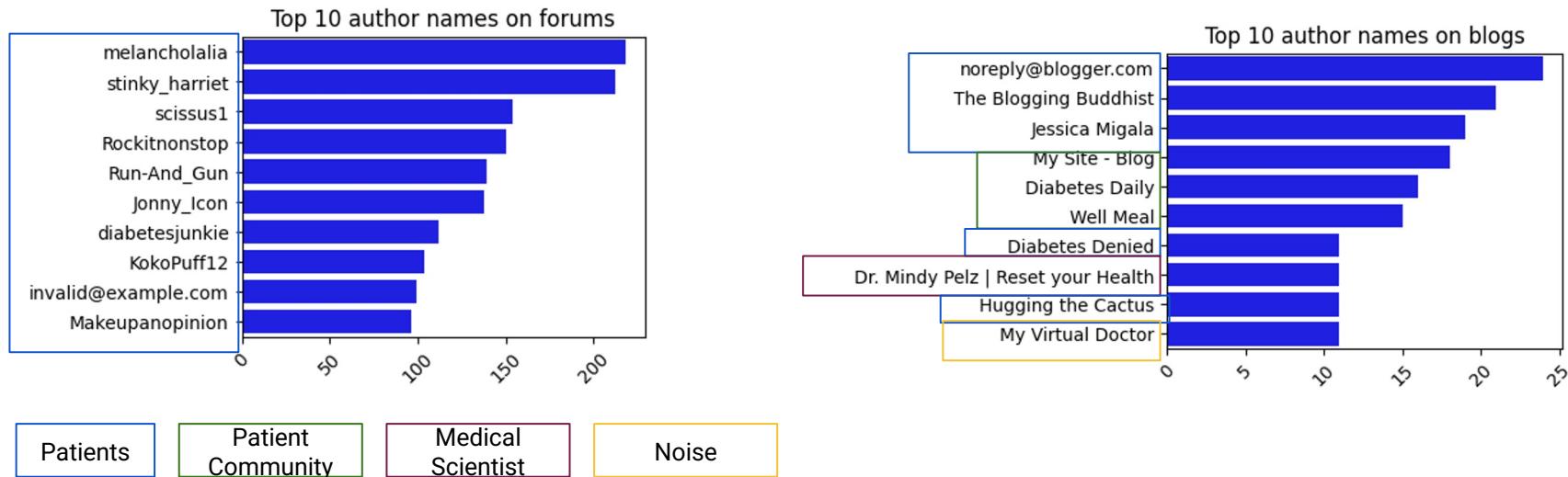


"What's worse for me right now is that I currently can't afford CGM and am using Freestyle Libre..."

"My mom has diabetes. She's tried to get a CGM, but insurance has denied her multiple times. My mom saw the data that I was getting from Levels and decided to pay full price for a CGM..."

"Take this quiz to better identify if you're as engaged with your diabetes management as you think you are"

The super authors are mostly patients or patient communities sharing device usage experience, monitoring and lifestyle tips, pricing



General CGM Analysis

range type diabetes
meal need diabetes help
kind t1d normal person
etc testing
testing
use cgm alarm hospital
fine great
great hospital
blood sugar
low type
night
point working
times day care
sugar level monitoring
good sensor glucose monitoring
time week year going
change bring data helped
cgm pump spike
small stay taking
watch
think pump
new bo minute different got cgm
feel try use don took
cost glucose level
low carb
run mean
diet definitely sure test especially
plan ask getting
check pretty day work thin
experience
a1c told instead actually control tell
let bad started
told instead
glucose monitor month
high glucose monitor
find finger stick
endo
looking bolu phone wanted
insurance makes able way
based app case
doctor
insulin want know
issue happen blood glucose eat diabetic
continuous glucose
found diagnosed pump arm drop
cgm lot right people
look start long problem eating best set hard tried
exercise medtronic
dose
cause wake information idea
gets easier effect
understand home
life

Topic modeling and clustering is utilized to see if there are CGM specific topics that are discussed

Clustering (K-Means)

Unsupervised machine learning method

K-means clustering algorithm works in three steps.

1. Select the k values.
2. Initialize the centroids.
3. Select the group and find the average.

Does not consider overlap between clusters.

Topic Modeling (LDA)

Discover the abstract “topics” that occur in a collection of documents.

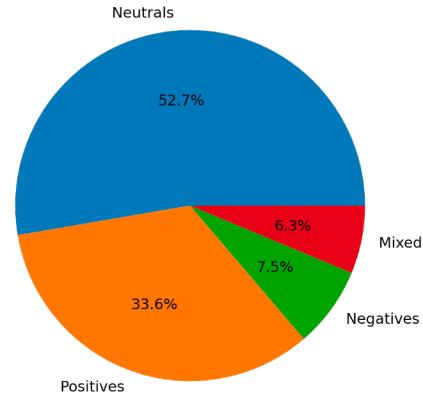
It builds a topic per document model and words per topic model, modeled as Dirichlet distributions.

Allows for topic overlap to occur.

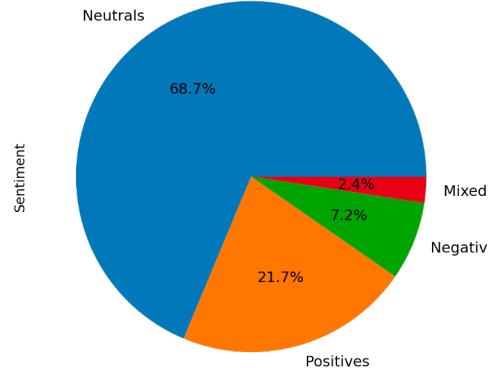
K-Means clustering shows that there are some distinct terms that pop up in certain groups

Cluster Number	Count	Top 10 Terms	General Topic
1	4581	<u>libre</u> , freestyle, sensor, dexcom, use, day, app, cgm, like, phone	CGM monitors (mainly libre)
2	5197	<u>pump</u> , medtronic, dexcom, <u>tandem</u> , cgm, year, insulin, <u>omnipod</u> , <u>tslim</u> , <u>slim</u>	Insulin pumps
3	21337	dexcom, cgm, low, like, time, day, sensor, insulin, year, use,	Unsure so far
4	6729	<u>blood</u> , <u>sugar</u> , <u>glucose</u> , monitor, continuous, cgm, low, insulin, <u>eat</u> , <u>high</u>	Personal/Life style habits?

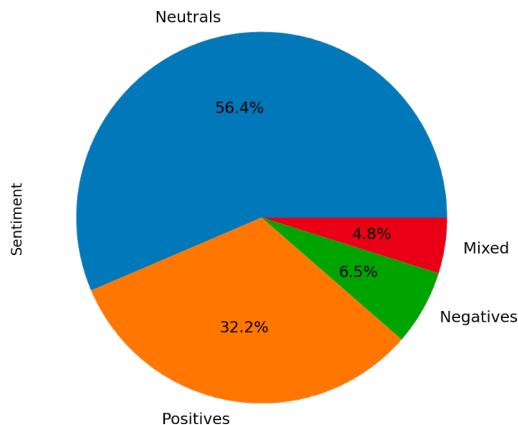
Cluster 1



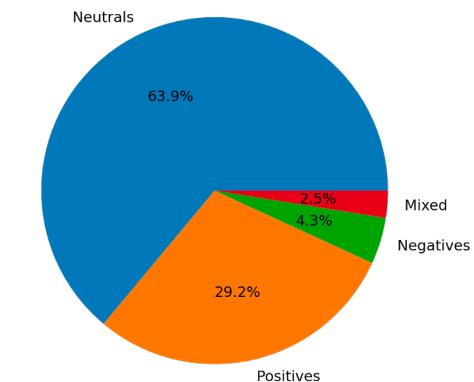
Cluster 3



Cluster 2

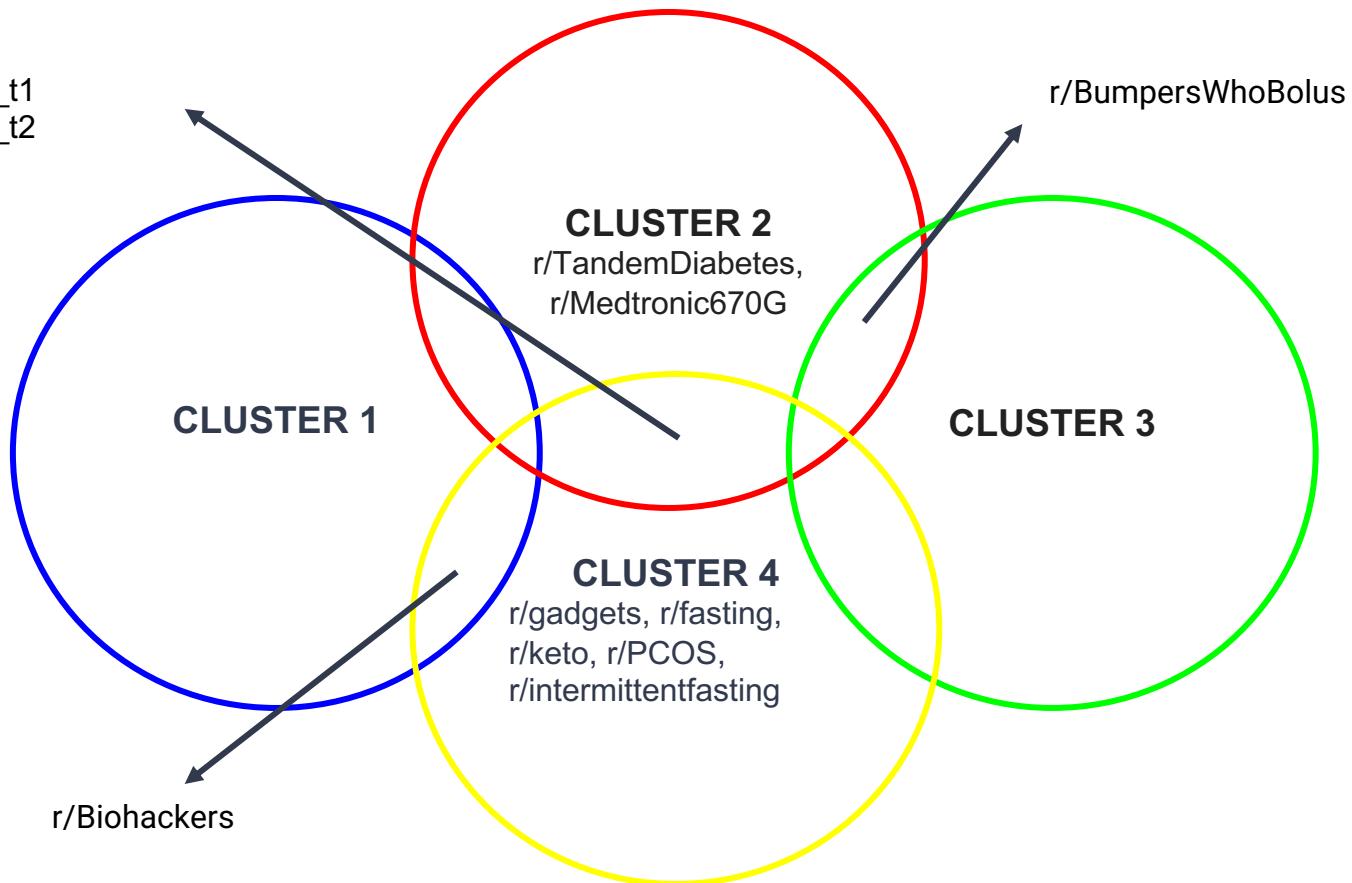


Cluster 4



Although there is not much difference in sentiment scores between clusters, clusters 1 and 2 have slightly less positives compared to clusters 3 and 4.

r/diabetes
r/diabetes_t1
r/diabetes_t2
r/dexcom

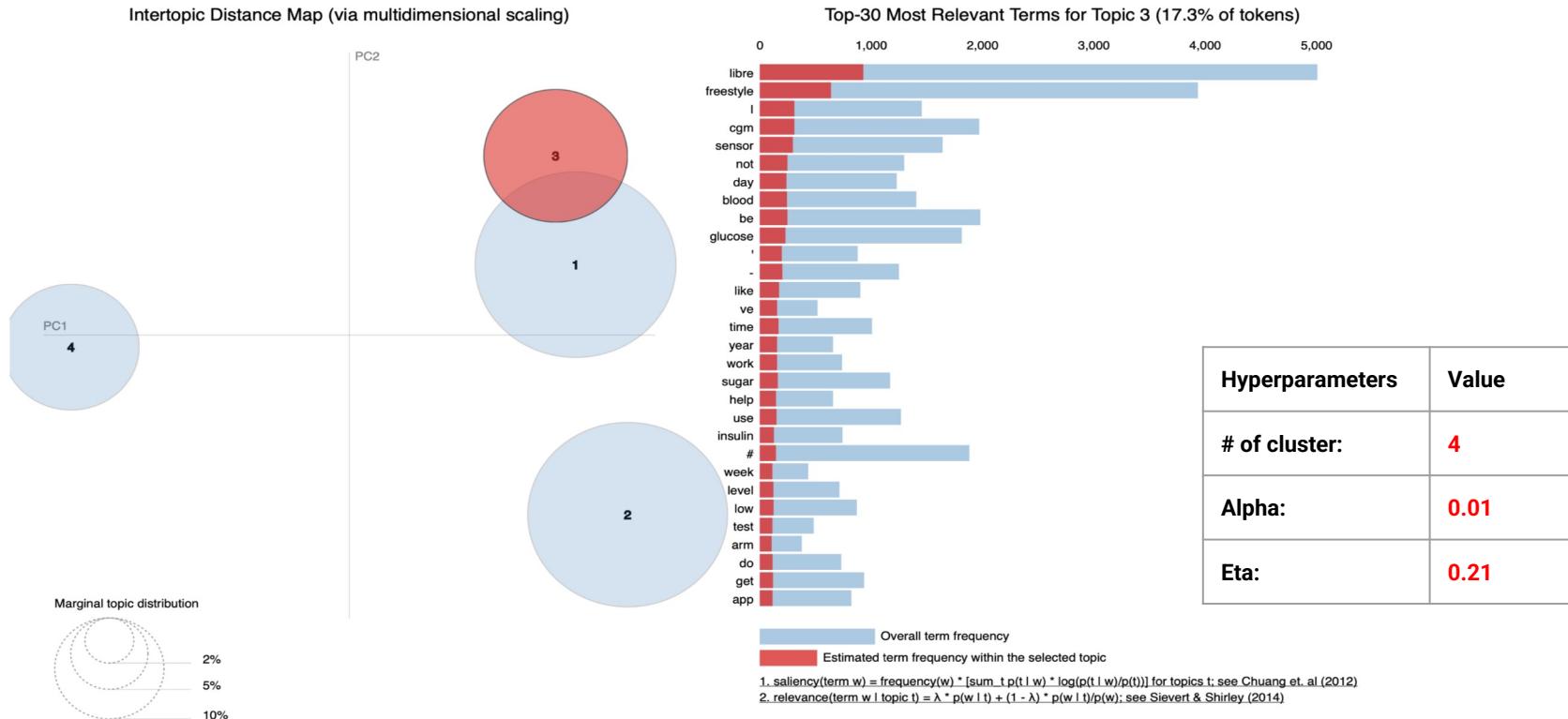


Unique subreddits that appear frequently for each cluster

Clusters seem to have a decent amount of overlap.

Latent Dirichlet allocation (LDA) allows documents to “overlap” each other in terms of content, rather than being separated into discrete groups.

Results generated from the LDA model show overlap of topics and relevant terms



Topic defining quotes for each topic provide insight into user issues and experiences with CGMs

- For each topic, rank all corresponding texts based on the contribution scores
- Distill topics from the top 10 texts within each cluster

Topic 1: The daily experience of people have diabetes

“...I was in school and my Dexcom CGM was beeping like crazy...”

“...good, my Blood Glucose responds more to bread-ish things...”

Topic 2: Features built into the product and associated hardware devices

“The ability to track and map the ups, downs and patterns...”

“Ever since a late Google update to Android 10 on my old Pixel 2...”

Topic 3: Users Questions and knowledge Gap

“And what pump do you suggest?...”

“Mine reads low sometimes, and my doctor says to trust my symptoms...”

Topic 4: Users' preference on different brands of CGMs

“I loved our MM2/Libre combo and if we hadn't had to switch to....”

“Dexcom G6 works great, far better than Medtronic CGM...”

Key Takeaways - CGMs Conclusion

Some people are concerned with the accuracy of certain CGMs.

- Can certain foods cause over fluctuations in reading levels.
- Should users put more importance into the CGM readings or their own symptoms?

There are groups who are interested in how CGMs can interact with their insulin pumps.

- Is there a certain brand of CGM that integrates with a specific insulin pump?

There are mixed opinions about alert/notification systems built into CGMs - some find it annoying while others find it helpful

People considering switching from their current CGM brand are usually comparing Dexcom and Freestyle Libre.

Dexcom vs Freestyle Product Analysis

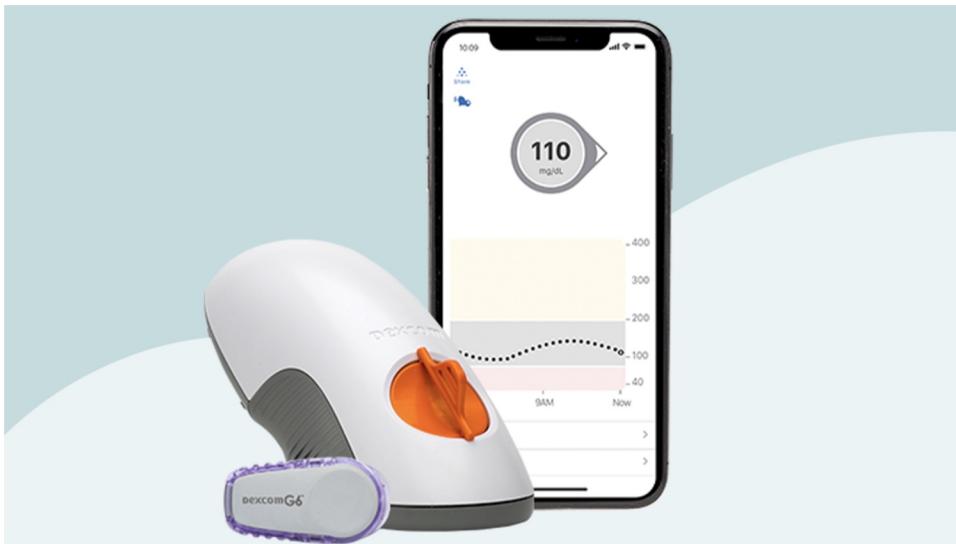
Differences at a Glance

Product	Details	Length of wear	Approved for ages
Dexcom G6	<ul style="list-style-type: none">can insert easily with one handallows you to monitor glucose levels in real time on your smartphonecan share real time data with 10 other peoplerelies on bluetooth signalmore expensive	10 days	2 years and older
FreeStyle Libre	<ul style="list-style-type: none">more affordableoptional alerts for low and high glucose levelshave to manually scan the sensor each time you want a readingmay not be as accurate traditional methods	14 days	4 years and older

<https://www.healthline.com/health/diabetes/dexcom-g6-cgm-product-review#pros-and-cons>

<https://www.healthline.com/health/diabetes/freestyle-libre>

Dexcom G6

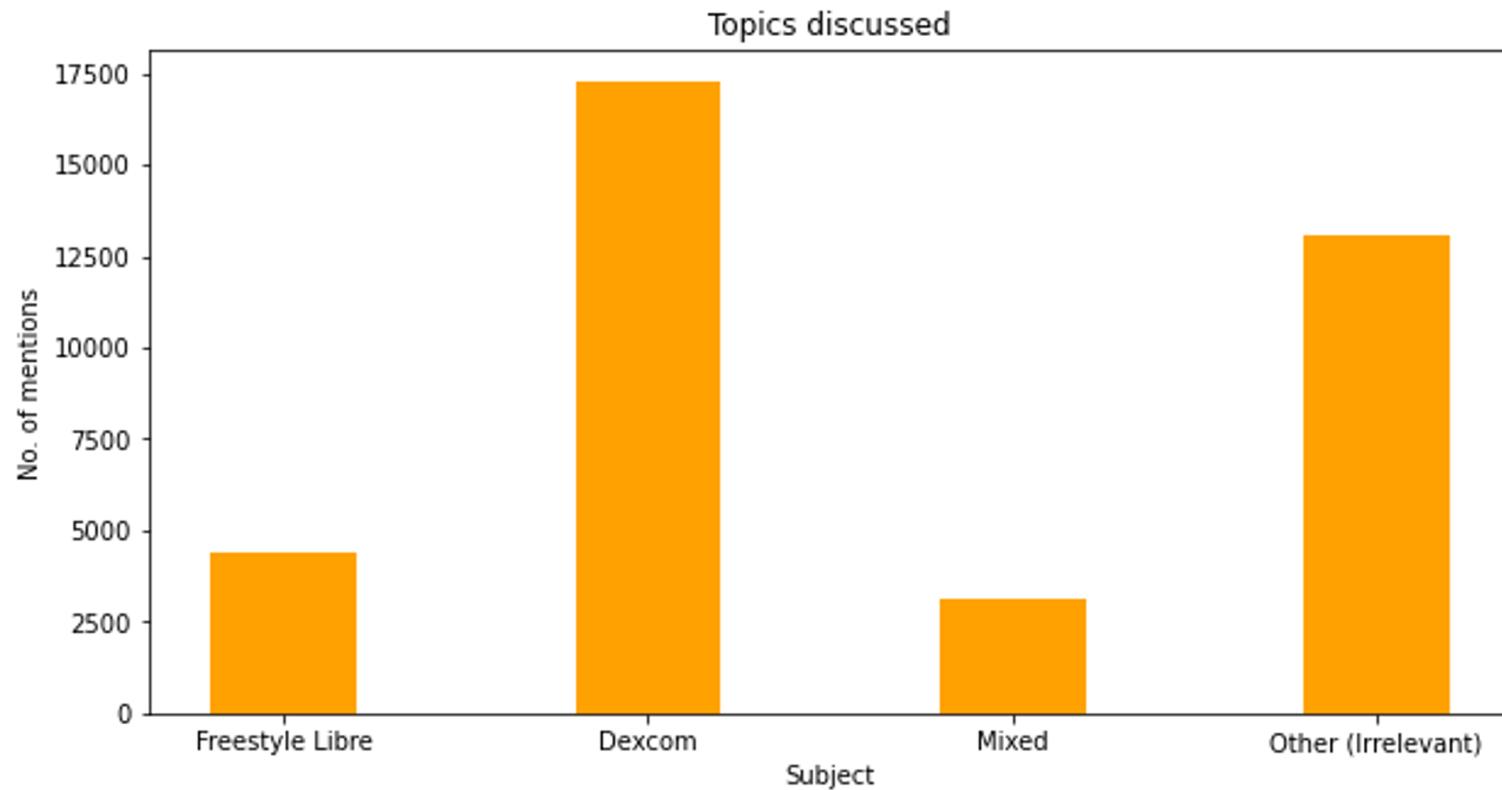


Freestyle Libre 3

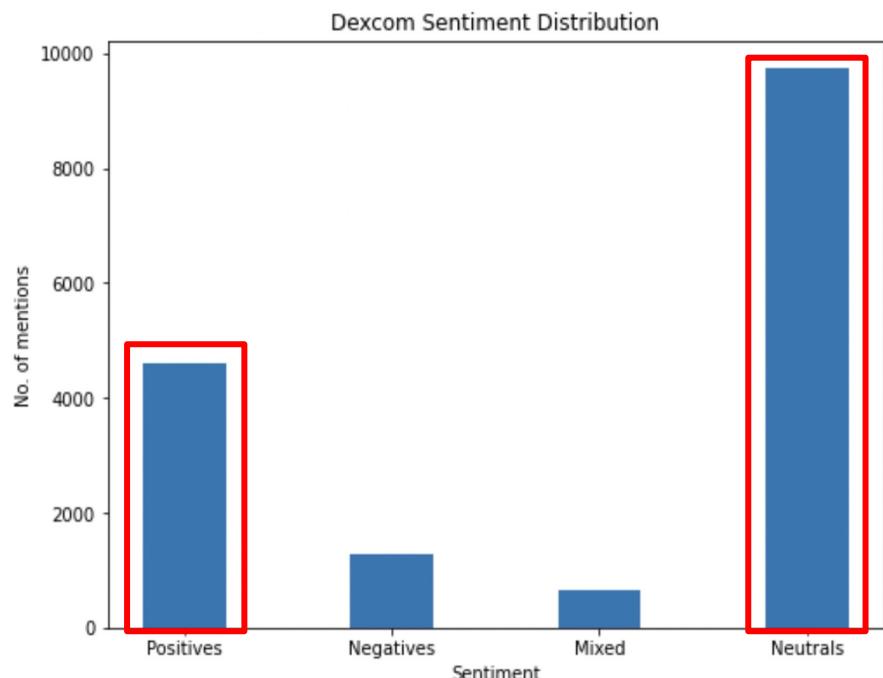
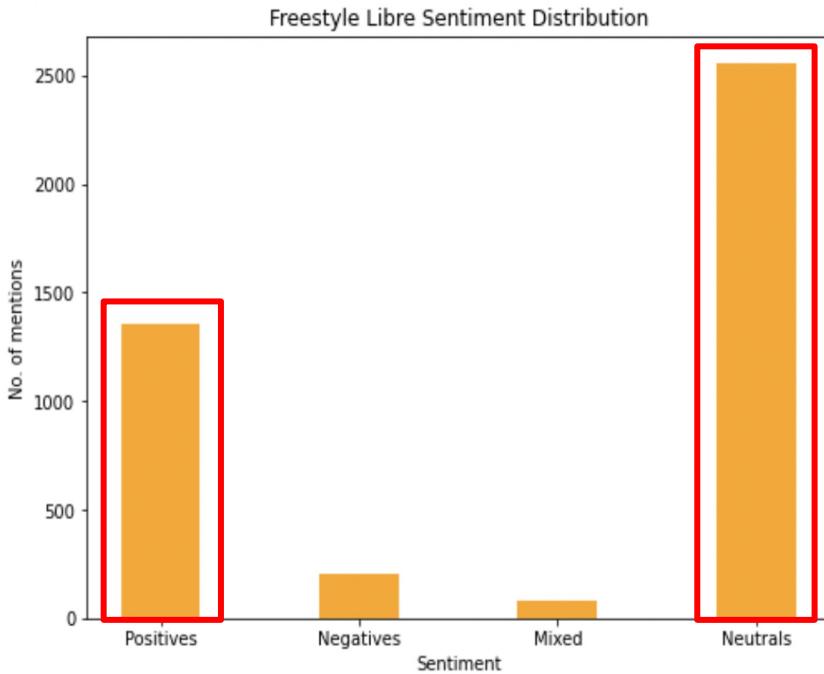


Latest Products

Imbalanced data: Dexcom is mentioned more frequently compared to Freestyle



Neutral and *Positive* are the main sentiments expressed about either product



Word Clouds by Product Brand

Dexcom Word cloud:



Freestyle Libre Word cloud:



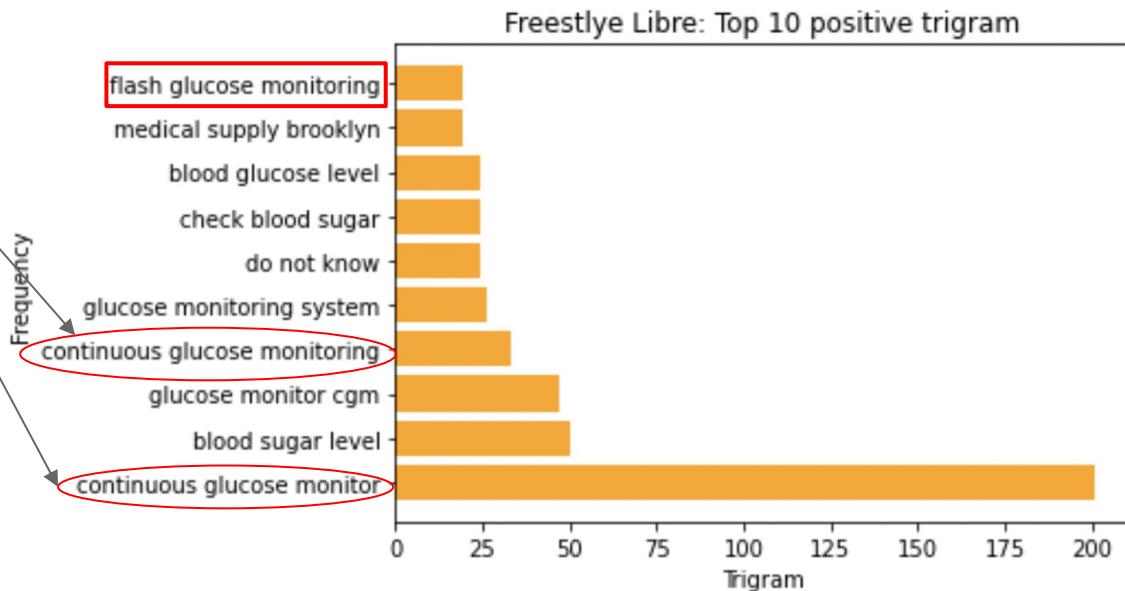
Freestyle Libre positive: Many discussions about continuous glucose monitoring

CGM = continuous glucose monitoring

No more routine fingersticks*

Flash glucose monitoring allows you to conveniently check your glucose with a painless¹, 1-second scan² of the FreeStyle Libre sensor using your phone instead of fingersticks*!

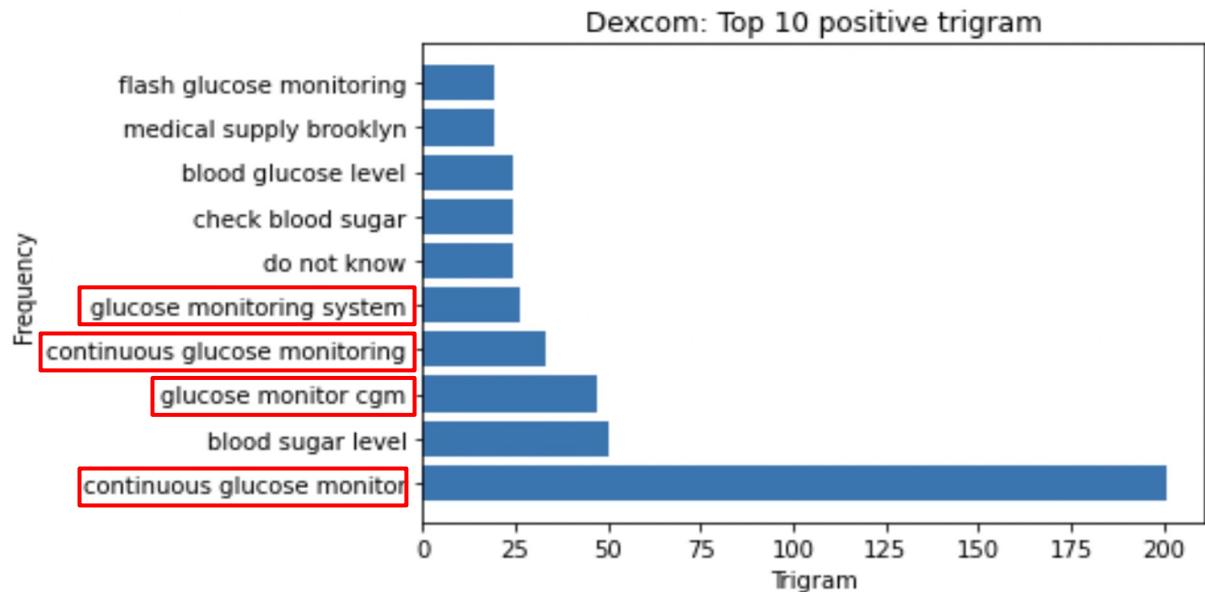
The FreeStyle Libre flash glucose monitoring system is designed to provide healthcare professionals with deeper insights and make informed treatment decisions.



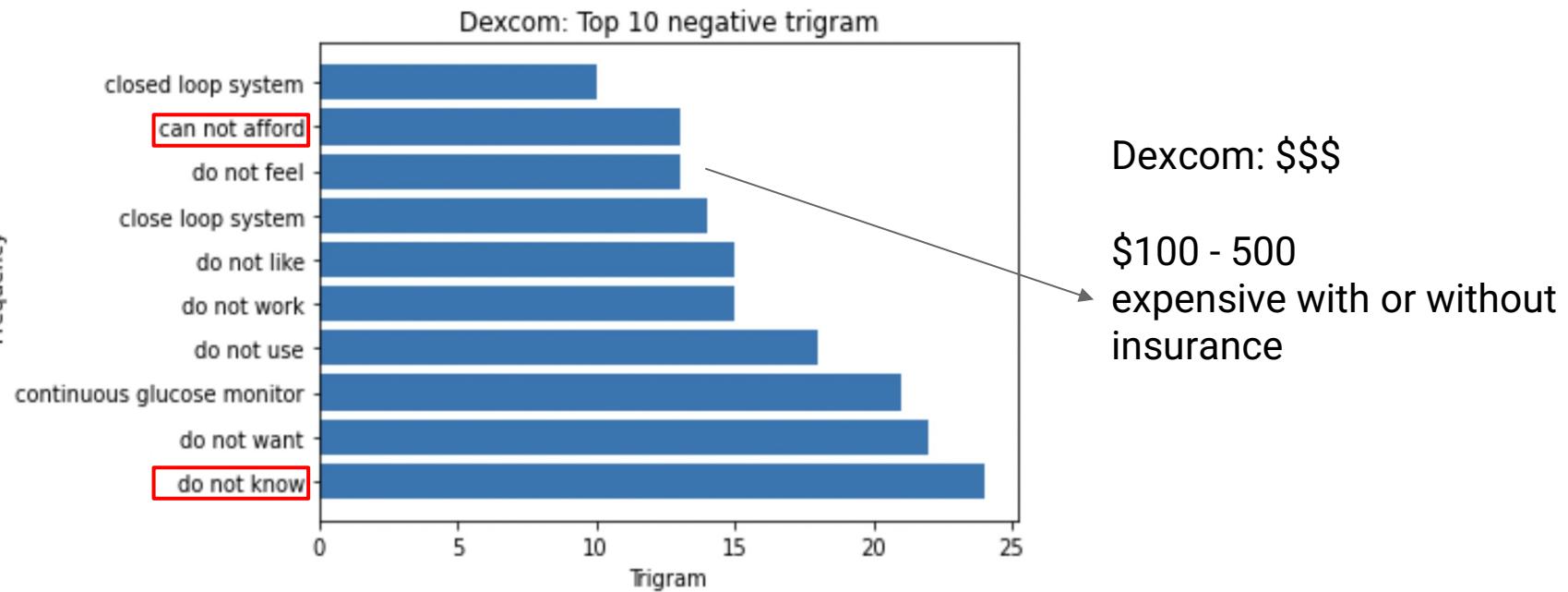
Dexcom positive: Many discussions about continuous glucose monitoring

Dexcom feature:

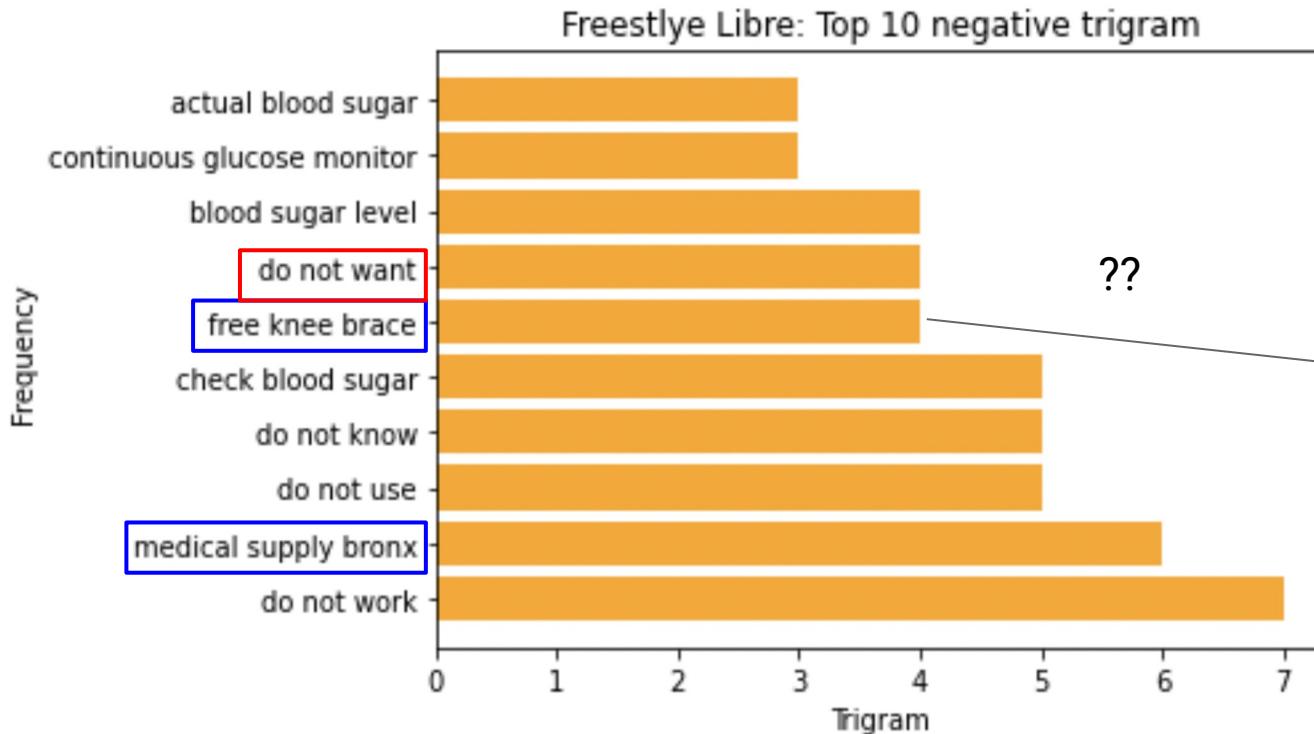
- continuously tracks glucose levels in real-time and update result on smartphone app/receiver device



What are the complaints?

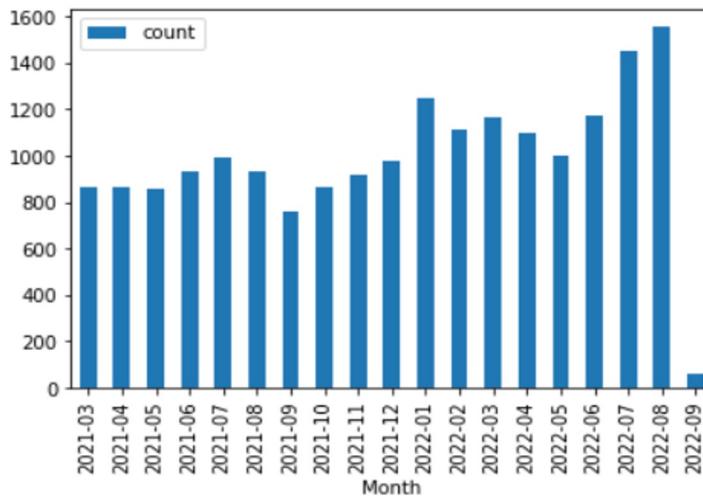


What are the complaints?

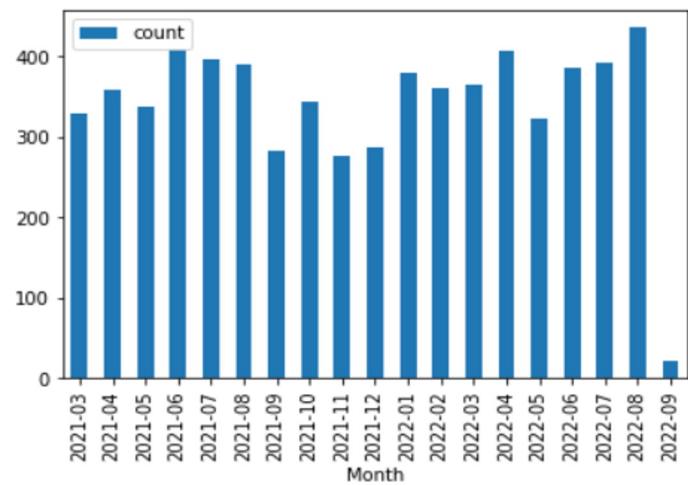


Product Popularity Trend

Dexcom related posts over time

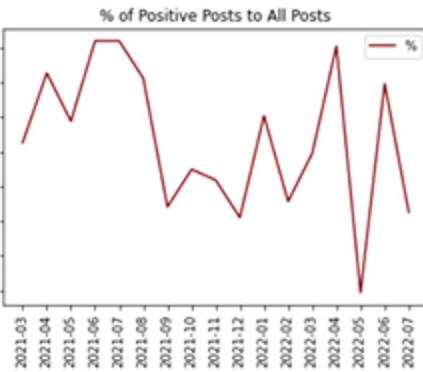
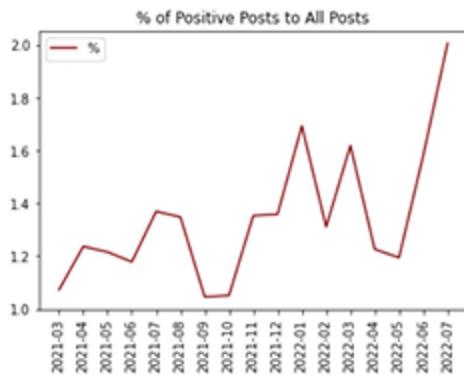


Freestyle Libre related posts over time

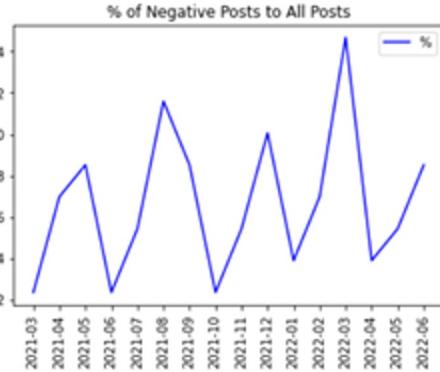
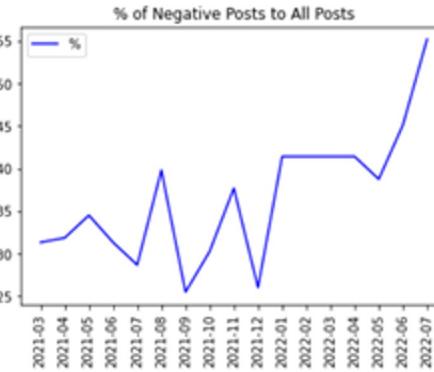


Opinions Trend

Positive Posts



Negative Posts



Top CGM Products Discussed

1. Dexcom
 2. Freestyle Libre
 3. Medtronic
- Dexcom is talked about 10x more compared with the other two brands
 - Insulin pump matters to customers

“Dexcom. It gets along well with my Tandem pump”

Key Takeaways by Brands

Dexcom

- Product price is a main concern of customers
- Insurance coverage across globe
- App and Device Interoperability

Freestyle Libre

- Integration with insulin pump
- Social media engagement
- Alert system does not need to be a development priority

Conclusion

What do people think of CGM?

- Hypothesis: People agree CGM is an easier solution to manage diabetes
- Conclusion: Yes. People agree CGM is a better alternative than fingerstick tests

What features should CGM companies focused on improving?

- Hypothesis: Accuracy, Price, Usability
- Conclusion: People are not too concerned about the accuracy of today's CGM. Insurance coverage is a key to expand patient access to CGM

Questions

