Arm & Hammer Baking Soda
Social Content Analysis

Wednesday, June 30th, 2021

Agenda

- Business Objective
- Methodology
- Data Collection
- Use Case Classification
- Key Insights
- Live Dashboard Demonstration
- Next Steps & User Guide

Business Objectives

Project Goals Defined by VIA

- Identify the **most popular use cases and topics** for Baking Soda (ex. DIY makeup)
- Rank **use cases/topics** by engagement
- Identify **emerging use cases** and topics for Baking Soda
- Describe the **platforms/channels** where emerging topics are happening, to guide where the company communicates in the future
- Visualize the above analyses in a dashboard that could be **updated continually with refreshed data**

Project Approach

What data do we require?

How do we classify the various use cases?

How do we define emerging?

Where are conversations happening?

Data Collection

Required Data

Customer Engagements

Posts and Replies from Customers that mention variations of "Baking Soda".

Time-Series Data

"Emerging" implies that something is increasing from a baseline.

Platform Information

Where are conversations happening?

Data Collection

Social Listening Tools

- Allow companies to compile user posts mentioning specific keywords.

Synthesio

- Customer Engagements
- Timestamps
- Source Information

- **For this project, we took advantage of VIA's Synthesio tool to compile a dataset containing mentions of 'baking soda'.**

Final Query Language

Synthesio Queries: ARM & HAMMER Baking Soda

General 2.0 ***[Updated 06/03]***

("BAKING SODA" OR "BAKINGSODA" OR "SODIUM BICARBONATE")

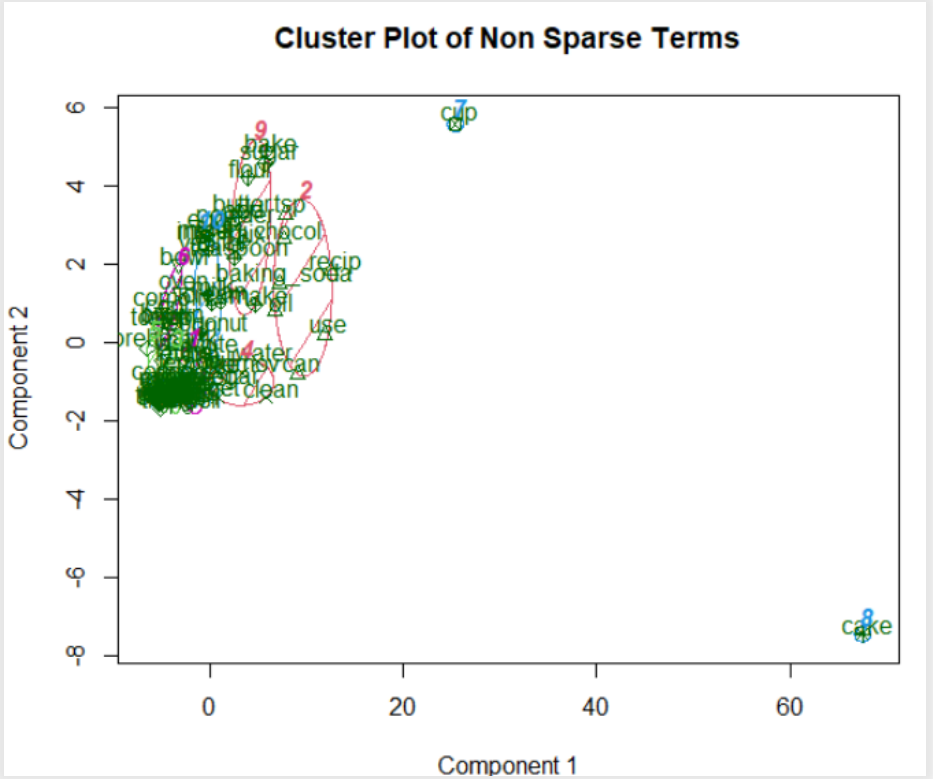
NOT

("COUPON" OR "SALE" OR "AD" OR "DEAL" OR "DEALS" OR "% OFF" OR "TRUMP" OR "ARMIE" OR "REPUBLICAN" OR "DEMOCRAT" OR "CRUZ" OR "BESTSELLER" OR "COCAINE" OR "PORN" OR "POLITICIAN" OR "METH" OR "ILLUMINATI" OR "GOVERNMENT" OR "ILLEGAL" OR "BIDEN" OR "CLINTON" OR "OBAMA" OR "NO BAKING SODA" OR "NOT BAKING SODA")

Use Case Classification

Classification: Lessons Learned

UNSUPERVISED MACHINE LEARNING: CLUSTERING



SUPERVISED MACHINE LEARNING: RANDOM FOREST, SVM, NAÏVE BAYES

```
       ACC PRECISION1 PRECISION2 PRECISION3 PRECISION4 PRECISION5 PRECISION6 PRECISION7        CE
  56.68449   58.82353   80.00000  100.00000   73.91304  100.00000   40.00000   30.90909  43.31551

$conf
          pred
target     beauty clean deodorize eat kids other spam
  beauty       10     1         0   3    0     3    6
  clean         3    12         0   2    0     7    7
  deodorize     0     2         5   1    0     0    0
  eat           1     0         0  51    0     2    7
  kids          0     0         0   4    1     1    8
  other         2     0         0   5    0    10   10
  spam          1     0         0   3    0     2   17
```

- Unsupervised clusters did not group usage words in distinct enough clusters to be useful
- Supervised methods could not reduce classification error under 30% and were demanding on system resources

Classification: Keyword-based model rapid automatic keyword extraction

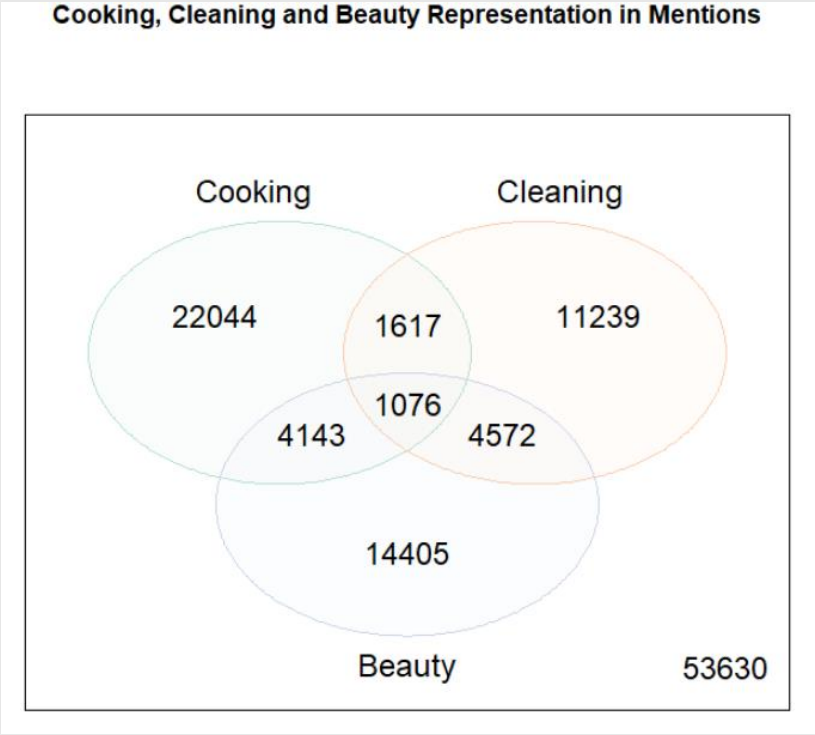
- RAKE extracts commonly occurring words and phrases.
- Commonly occurring phrases can be used to classify mentions into use cases.
- Example: 204 mentions contain the phrase “large mixing bowl”. I can tell R to classify these as “cooking/baking”.

keyword	ngram	freq
tsp baking soda	3	1956
teaspoon baking soda	3	1123
tsp baking powder	3	819
tsp vanilla extract	3	544
teaspoon baking powder	3	441
teaspoon vanilla extract	3	260
cup brown sugar	3	230
large mixing bowl	3	204
cup white sugar	3	201
apple cider vinegar	3	196
cup coconut sugar	3	189
cup baking soda	3	178
cup cocoa powder	3	169
cup coconut oil	3	166
cup vegetable oil	3	163
chocolate chip cookies	3	158
cup maple syrup	3	157
teaspoons baking powder	3	148
tier wedding cake	3	147
tbsp baking soda	3	138
cup chocolate chips	3	131

Classification: Single use vs multi use

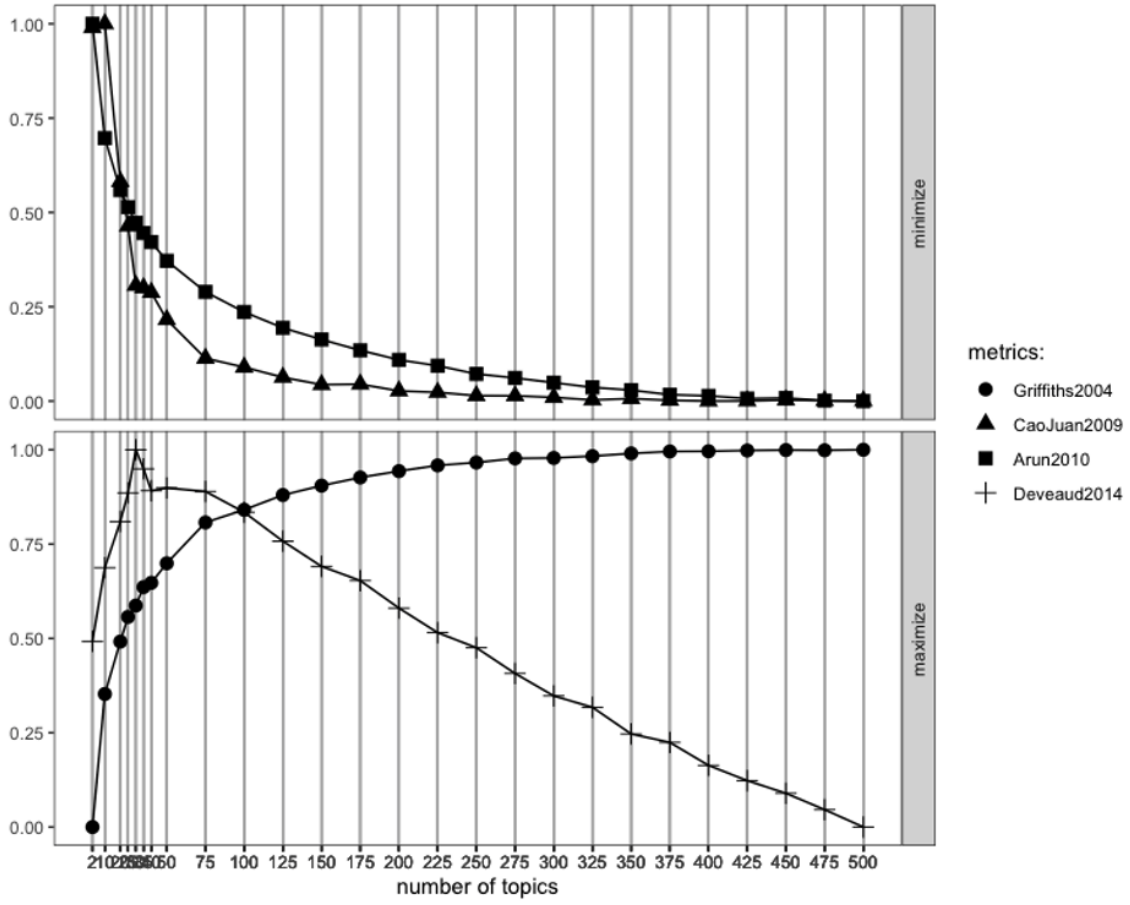
- Mentions can indicate multiple uses.
- Instead of classifying mentions into exclusive categories, we allow them to be classified into multiple categories.

low-tox cleaning so today i got inspired to do some spring cleaning in winter i felt inspired to start by cleaning off the top of the fridge. it usually has a vase with dried native flowers and some other dust collecting ornaments as well as a a few cake tins. well somehow we ended up with rust stains from the tins and i tried a few different things to remove it! then i remember lemon essential oil it just took a few drops and a bit of rubbing and the rust stain was gone! i m telling you it s really good stuff! check out the before and after photos! other uses for lemon essential oil: + d i s h s p a r k l e r 3 drops in a small glass with white vinegar in the top rack of your dishwasher during your normal washing cycle. shiny and bright! + l a u n d r y b r i g h t e n e r a few drops of lemon essential oil in the laundry basin with your whites will help to brighten them! + s m e l l y l a u n d r y in case you leave your laundry sitting in the washer too long, just add a few drop of lemon eo and your clothes won t get that nasty smell. + t e e t h w h i t e n e r mix lemon essential oil, baking soda and coconut oil and rub on your teeth for 2 minutes and then rinse. ideas/content from botanicatribe #lowtox #lowtoxliving #lowtoxcleaning #naturalliving #naturallivingfamily #healthyhome #chemicalfree #chemicalfreehome #family #cleaning #rustremoval #essentialoils #younglivingessentialoils #lemon #naturalcleaning



Classification: LDA topic modeling

- Latent Dirichlet Allocation (LDA)
- Unsupervised machine learning method often used in natural language processing that divides words into topics
- Advantage: Words can belong to more than one topic
- Disadvantage: Works best when documents are long
- Unable to determine a distinct number of topics within the baking soda data

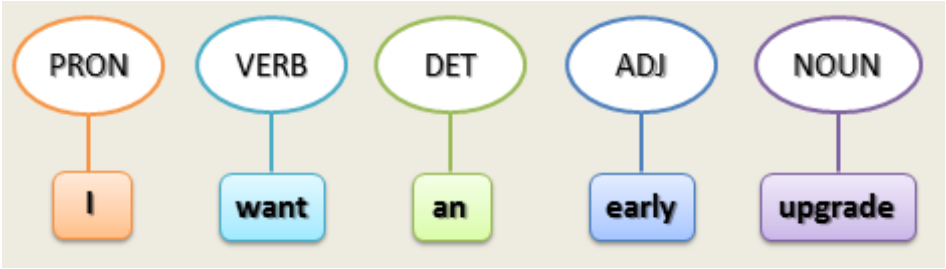



```
> terms(ldaOut,6)
     Topic 1    Topic 2  Topic 3       Topic 4   Topic 5   Topic 6
[1,] "vinegar"  "salt"   "powder"      "pan"     "cookies" "hair"
[2,] "clean"    "red"    "baking"      "bowl"    "dough"   "try"
[3,] "cleaning" "sauce"  "ingredients" "minutes" "cookie"  "tried"
[4,] "apple"    "cheese" "dry"         "batter"  "minutes" "used"
[5,] "remove"   "oil"    "wet"         "cool"    "baking"  "work"
[6,] "cider"    "cook"   "orange"      "oven"    "oven"    "shampoo"
```

Classification: Hashtag Model

- Hashtags are users telling you what they're talking about
- Limitations: Not all platforms or users use them
- Challenges/Methodology of text mining for hashtags
  - Hashtag in text mining: "(#+[a-zA-Z0-9(_)]{1,})"

Classification: Part-of-Speech Model



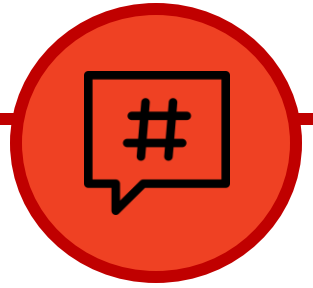
- Similar methodology to Hashtags – but **applicable to all platforms**.
- Used a custom stop-words dictionary to filter out 'irrelevant' words.
- Like hashtags, this topic model is not limited to baking soda, and is **applicable to other products**.
- Requires a subject matter expert to distinguish relevant results.

	wordComparisons.item
1	wool
2	volcano
3	chemistry
4	dissolve
5	react
6	scraper
7	swish
8	alcohol
9	steel
10	brew
11	wine
12	flare
13	beverage
14	blanket
15	boot
16	acid
17	eyebrow
18	mat
19	bitter
20	kitty
21	belly
22	starch
23	buttery
24	patrick
25	pudding

Classification: Summary

Keyword-Based Classification

Cooking
Cleaning
Health
HomeImprovement
Fun
Beauty
OdorControl



Hashtag Classification

#bakedoats
#naturalsoap
#healthylifestyle
#veganrecipes
#diy
#slime
#stressrelief

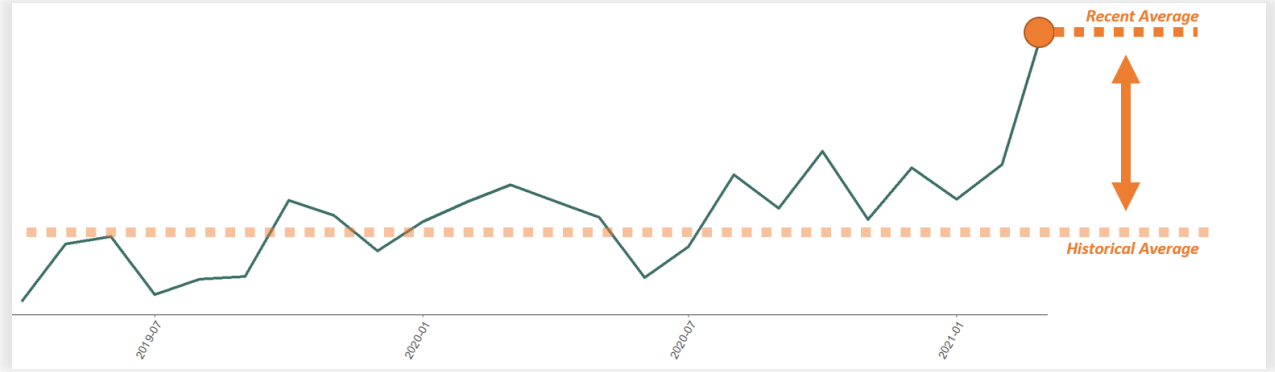
Part-of-Speech Classification

Volcano
Chemistry
Alcohol
Brew
React
Aquarium
Buttermilk

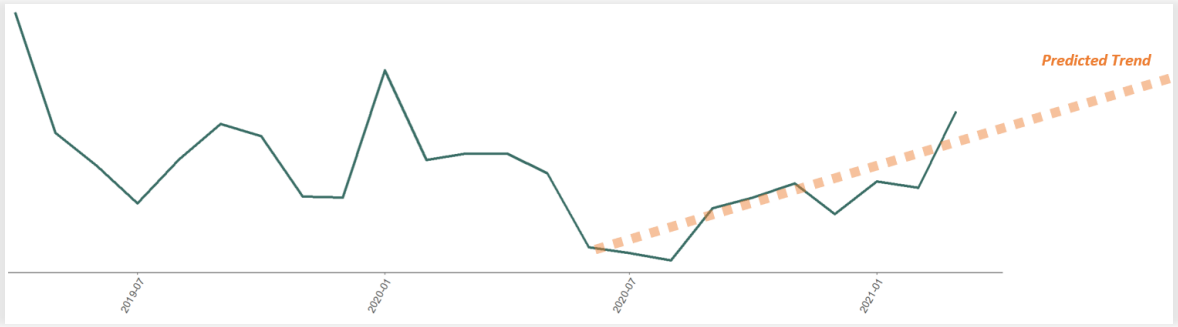
"Emerging" Use-Cases

"Emerging" vs. "Trending"

"Emerging": This past month, this use-case was **discussed more than usual.**



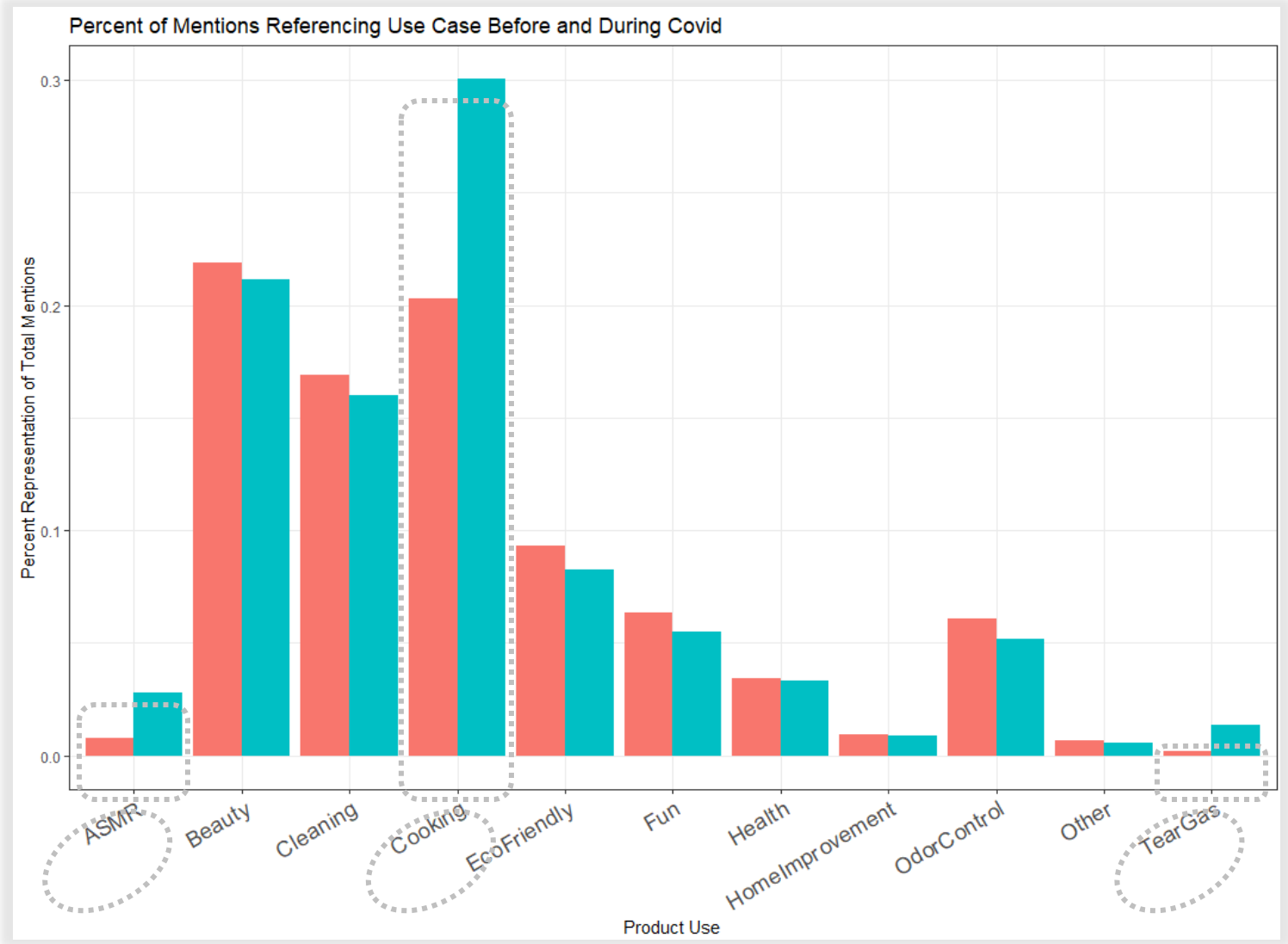
"Trending": Based on historical data, this use-case is **predicted to keep increasing in popularity.**



Dashboard

Key Insights

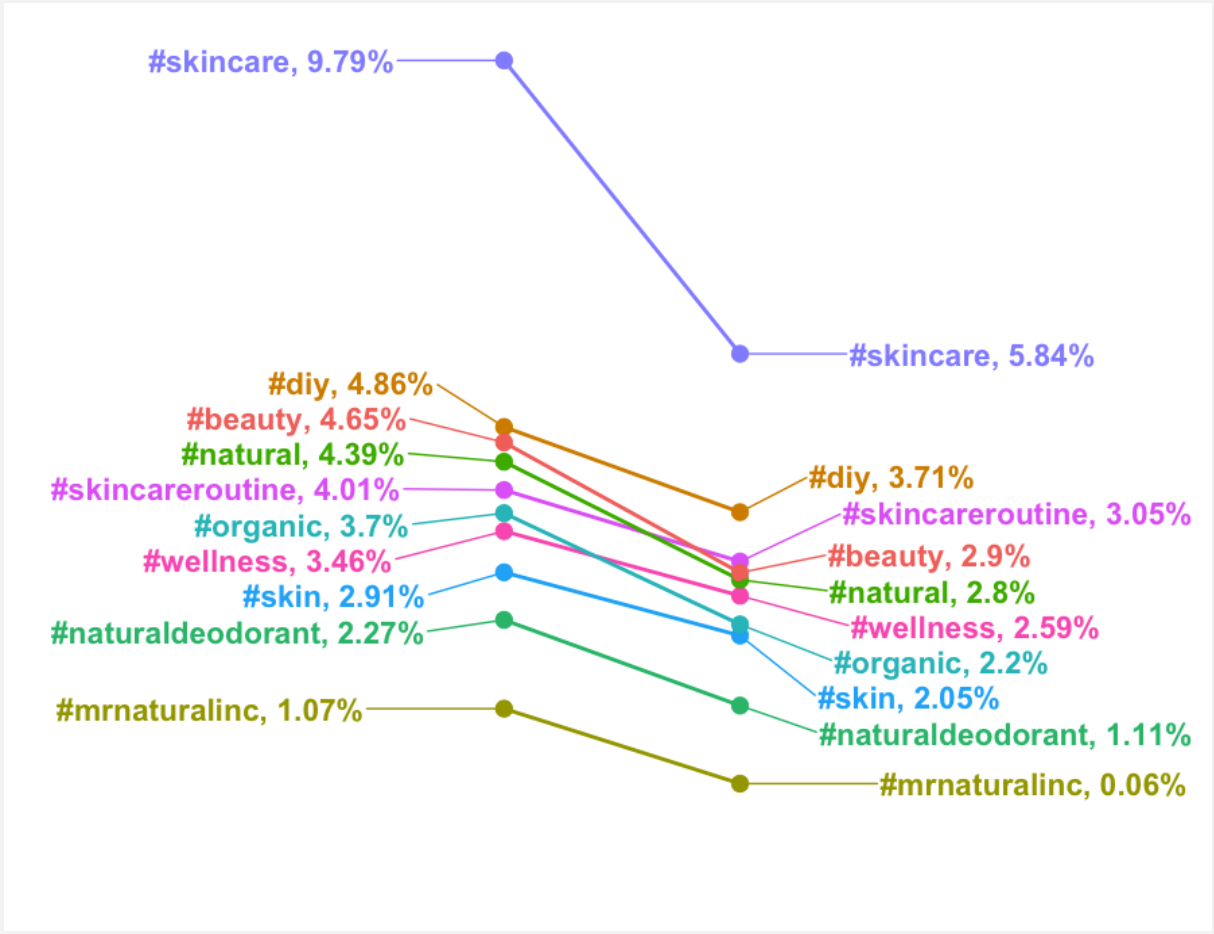
How did the pandemic influence the way people use baking soda?



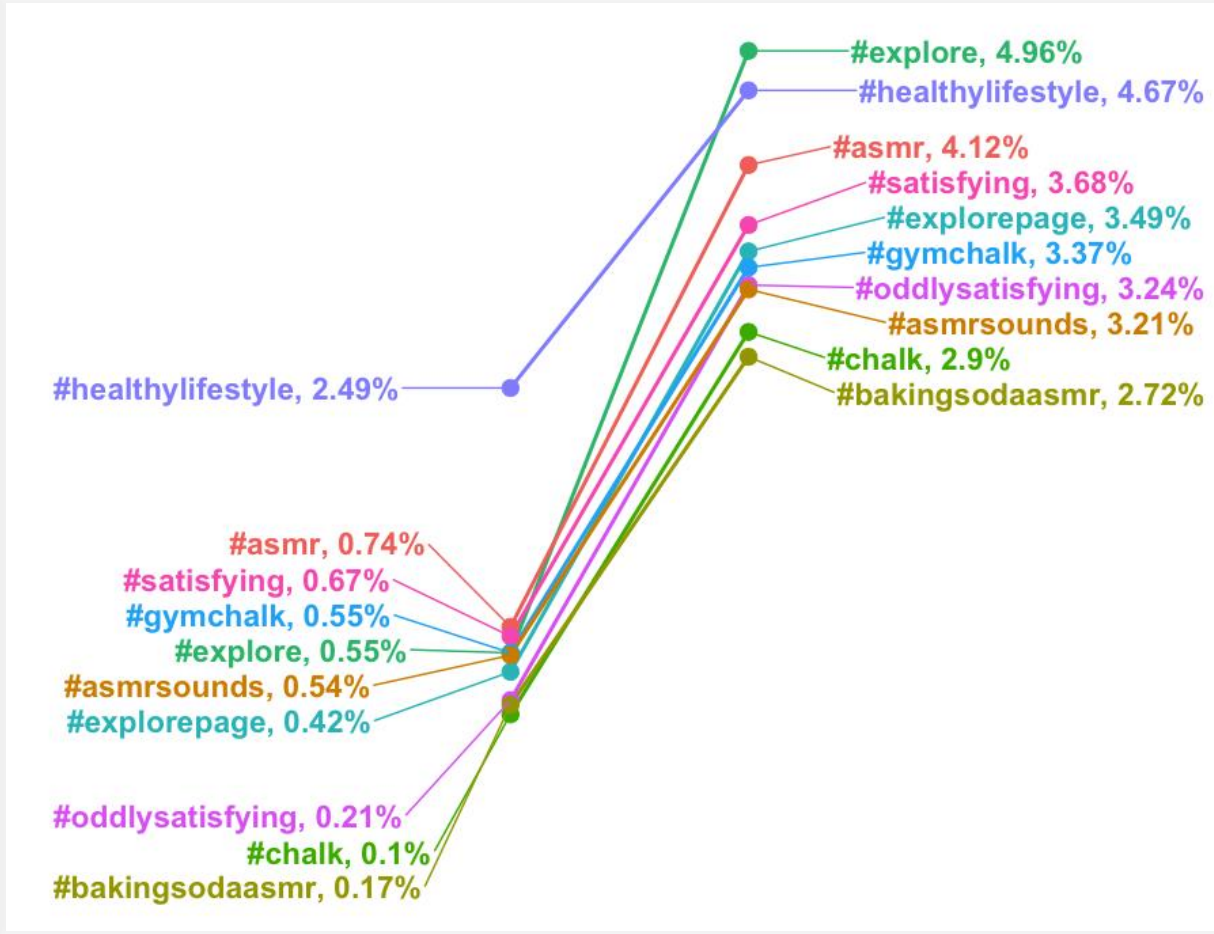
Pre-Covid
During Covid

**Pre-Covid dates are dates prior to March 1, 2020*

Hashtags with Largest DECREASE from Before to During Pandemic



Hashtags with Largest INCREASE from Before to During Pandemic



Percentages indicate frequency of hashtag within all posts containing hashtags.

ASMR

- Autonomous Sensory Meridian Response
- Sensation people get from watching certain stimulating videos
- Tingles in the spine or soothing/relaxing

Video Source: https://youtu.be/lKvjCZpXHik

Tear Gas / Pepper Spray Neutralizer

Peri
@ThePeriQuill

Stay safe out there, everybody.

> **JakeWhyman** @Jake_Whyman · May 30, 2020
> Teargass canisters can be kicked away or covered by traffic cones. Bring spray bottles with a baking soda and water solution to spray in your face to neutralize the effects of tear gas particles.
> Wear a mask (obviously).
>
> Stay safe, fellow Seattleites!
> (Damn. I hate that word...) twitter.com/seatimesbiz/st...

@___gooseman · Jun 2, 2020

I'm sure you might have seen other countermeasures, I haven't... so PSA:

Homemade tear gas neutralizer- 3 tablespoons of **baking soda** for every 8.5 oz of water.

Put mix in spray bottle and apply to eyes/mouth/skin if you're hit. Stay safe and stay strong.

#BlackLivesMatter

Stoovy · 1y

2 table spoons baking soda and 2 cups water in spray bottle. Spray into eyes if affected by teargas apply as needed.

Used this and it works amazingly. Helped lots of people with it on Saturday.

21 Share Report Save

throwawayyyy1311 · 1y

Yeah very surprised at how well baking soda and water worked

5 Share Report Save

¡ matt ! @y0uforias · May 29, 2020

baking soda and water to combat tear gas !! spray in mouth and eyes !!
#BlackLivesMatter

1 8 14

Uwagbale Edward-Ekpu @uwagbale_ · Oct 9, 2020

When going for #EndSARS protest ensure to carry water containing **baking soda** (80 ml of water/teaspoons of powder) which you can spray on your face to neutralize tear gas. Tear gas is unstable & breakdown in the presence of soda water in a process called hydrolyzation.
@SavvyRinu

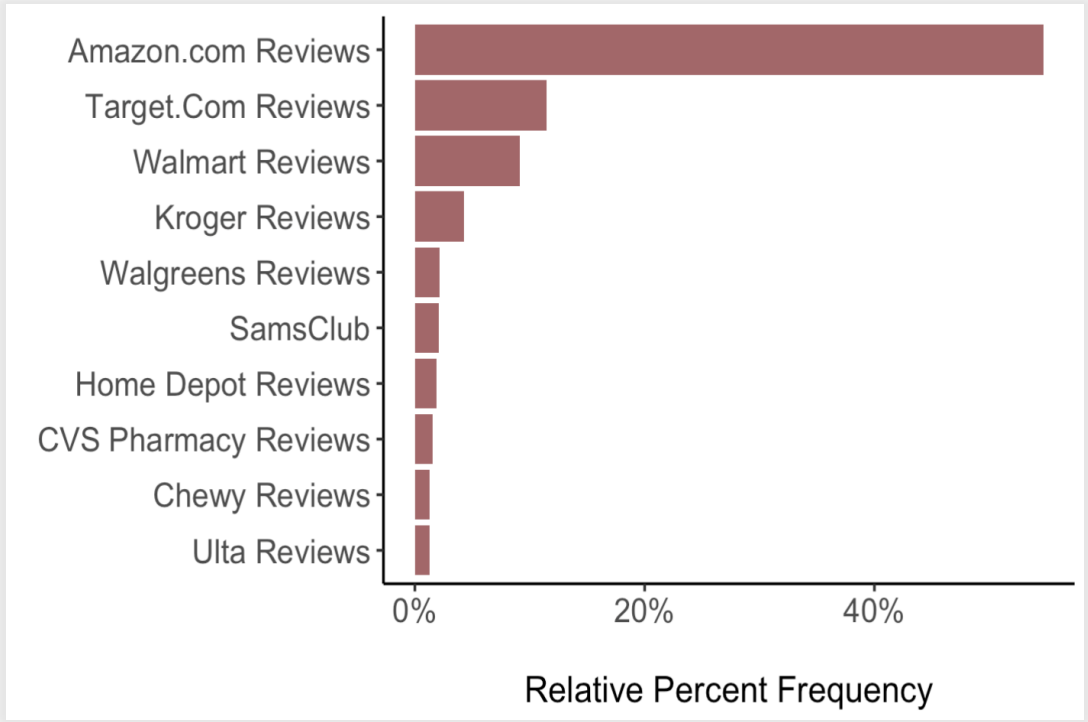
6 1K 679

jamie HTTYD BRAINROT
@peachiejam

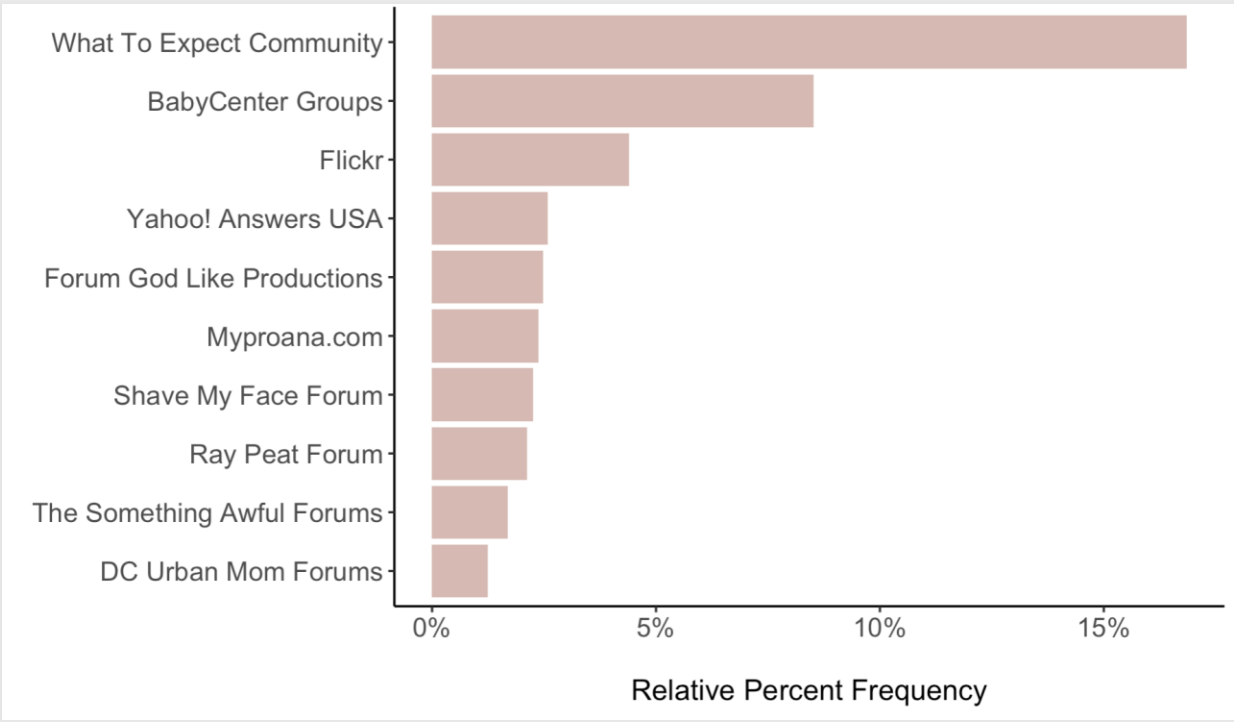
Replying to @peachiejam

Remember to use three spoons of baking soda per cup when you're using water to get tear gas out of eyes!!

Amazon is by far the most frequent **Retail Review** site



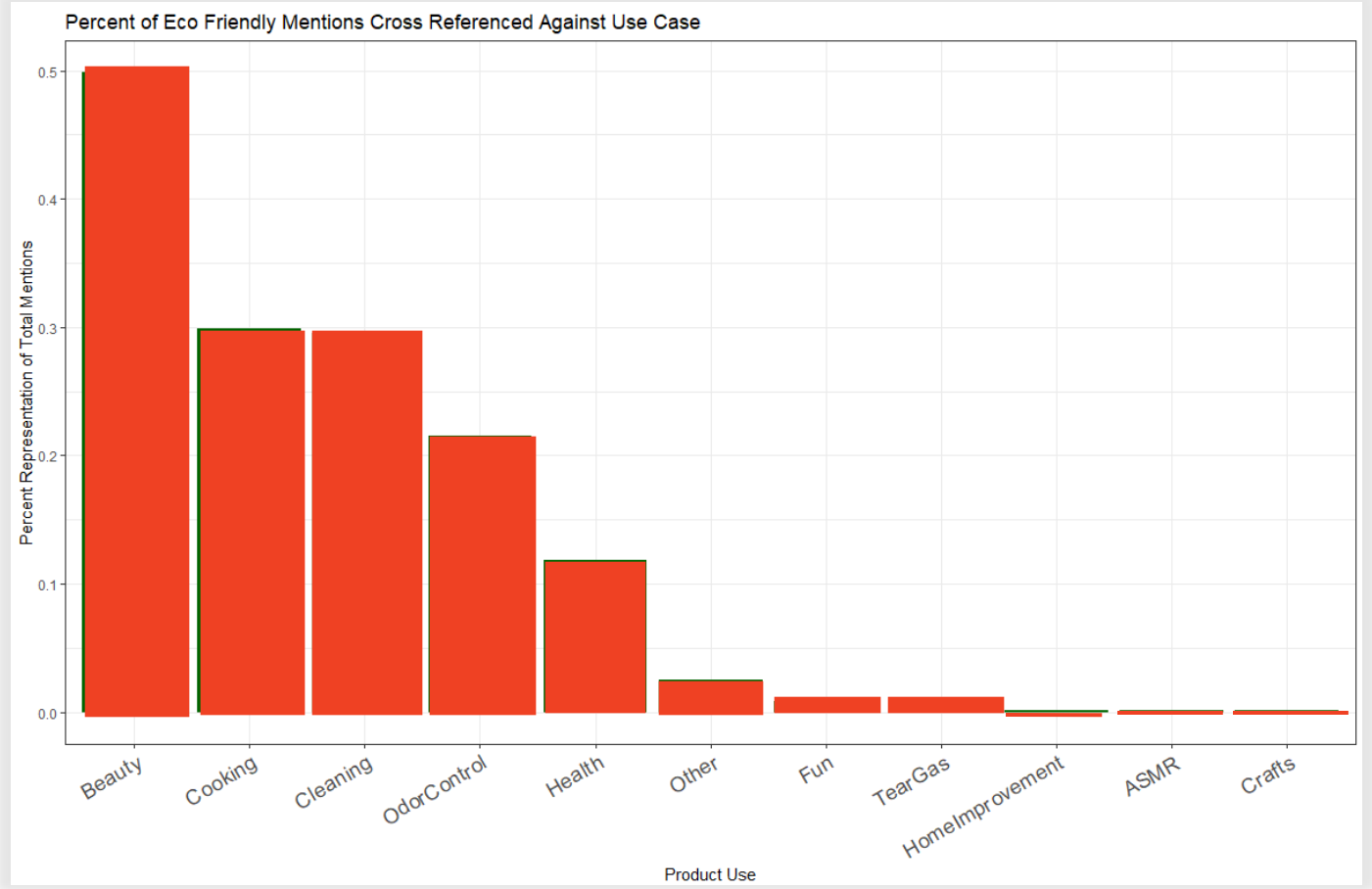
Baby groups make up over 25% of all **Blog & Forum** mentions



Next Steps & User Guide

Extending Use Case Classification

- Mentions can be classified based on other features than use case.
- This graph shows the frequency of use cases among mentions classified as Eco Friendly.



What we are Delivering

- All the code that transforms the data and generates the dashboard.
- A user guide that walks through software installation, running the tool, and information on how to customize certain features.

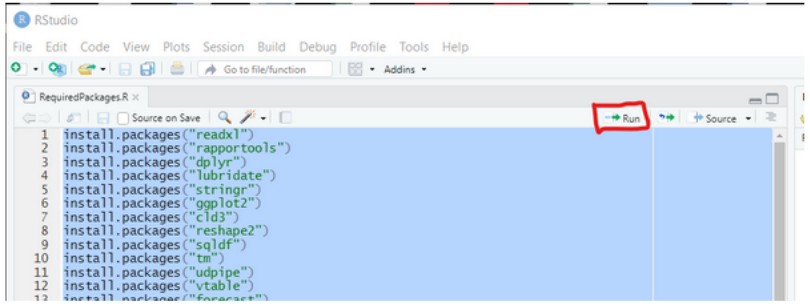
Getting Started

Downloading and Installing R and R Studio

The first thing that will need to be done is downloading and installing R and R Studio. R should be installed first, and R Studio second. First, you will want to go here and install the latest version following the steps on the website. Then, go to this website and download R Studio (the desktop version is sufficient) and install.

Installing Required Packages

Once R and R Studio are installed, you will want to open the RequiredPackages.R file included in the zipped file with R Studio. Highlight the code with CTRL+A, then select the run button:



This file will only ever need to be run one time. The files we have provided require additional libraries that are not a part of the base R functions, and this file will install all those packages onto your machine.

Loading and Running Files

Make sure that the zipped file has been downloaded and saved somewhere on your machine, and that the files have been extracted, but are still contained in the same root folder. Select the Run First file (it should open automatically in R Studio, or you can do a File > Open within R Studio). To run the file,

Limitations

The code uses the column names from the Synthesio file. If names are changed or columns omitted, it could cause the code to fail.

Use Case Classification was created specifically with baking soda in mind, so if this is run for other products, that aspect may not work.

Trend analysis is based on monthly trends, so ideally the dataset contains 6 months of data in order to make somewhat meaningful predictions.

Questions

Backup Slides: Dashboard Images

VIA x ARM AND HAMMER

- Home
- Overview
- Social
- Use Case
- Trends

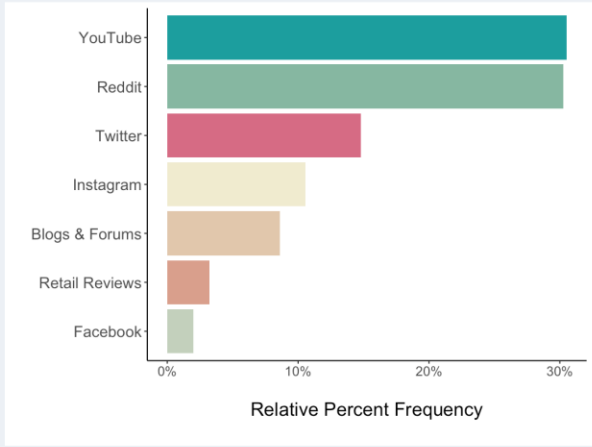
Welcome

VIA x ARM AND HAMMER

- Home
- Overview
- Social
- Use Case
- Trends

Overview

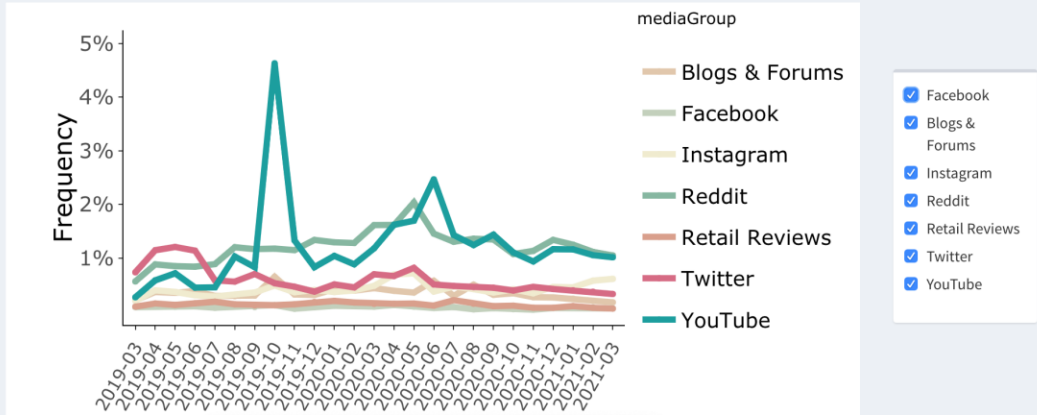
Media Groups



The media groups were regrouped from the original media types. These groups were then looked at to gauge overall prevalence (left) and change in usage over time (below).

Media Groups Over Time

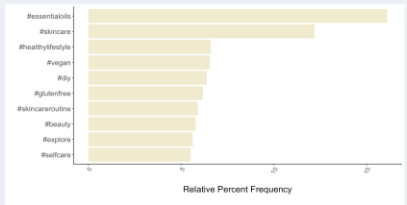
The time series below shows the movement in the media groups over time and allows the user to see when certain platforms become more popular. The filter allows for the selection of certain media groups not all need to be compared at once. The graph itself also has a scroll over (which shows the percent frequency of each month) and additionally has a zoom function which allows the user to focus on specific months and see more minor movements across time.



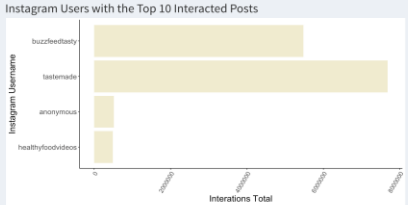
VIA x ARM AND HAMMER

Instagram

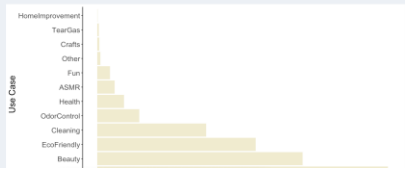
Top Hashtags for Instagram



Instagram Users with the Top 10 Interacted Posts



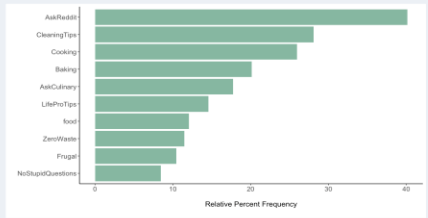
Percent of Instagram Mentions Containing Use Case Keywords



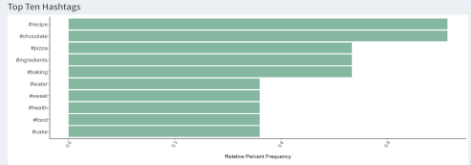
VIA x ARM AND HAMMER

Reddit

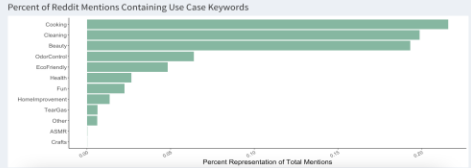
Top Ten Subreddits



Top Ten Hashtags

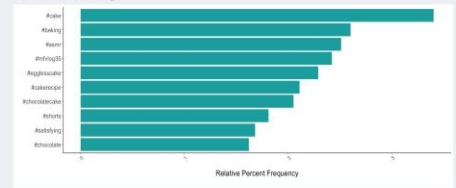


Percent of Reddit Mentions Containing Use Case Keywords

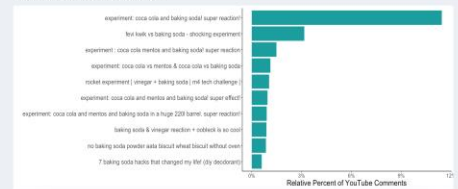


YouTube

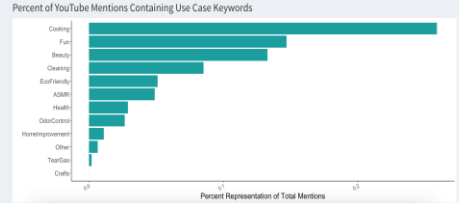
Top Ten YouTube Hashtags



Most commented YouTube Video Titles

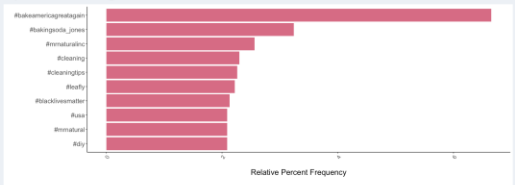


Percent of YouTube Mentions Containing Use Case Keywords

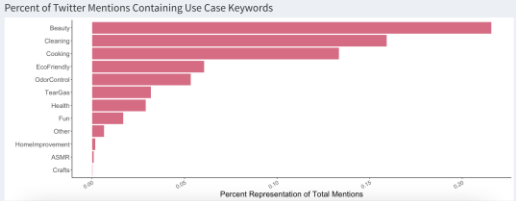


Twitter

Top Ten Twitter Hashtags



Percent of Twitter Mentions Containing Use Case Keywords



VIA x ARM AND HAMMER

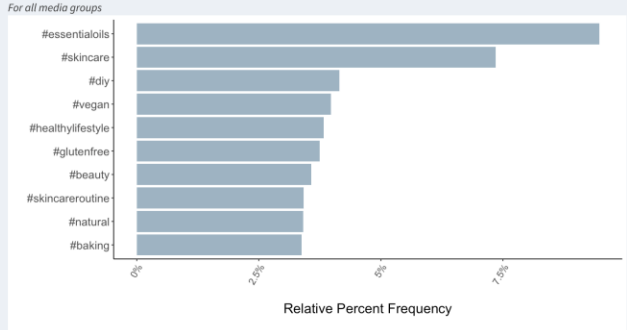
- Home
- Overview
- Social
  - Instagram
  - Reddit
  - Youtube
  - Twitter
- Use Case
- Trends

Use Case

Use case was discerned using two different techniques, both displayed on this page. The first was using hashtags, as users add hashtags to online posts to emphasize the purpose and topic of the post. The second was using manual type of classification where both words and hashtags were used to identify key groupings.

Top Hashtags

For all media groups



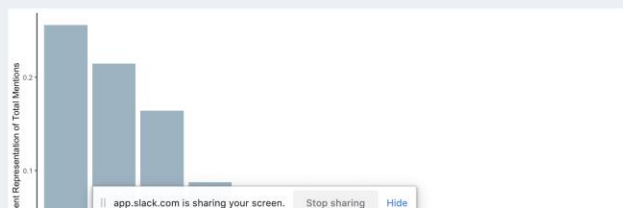
The following interaction below allows the user to select one of the top hashtags and shows hashtags that are associated. The greater the association level number, the stronger the hashtag is associated with the original selected hashtag.

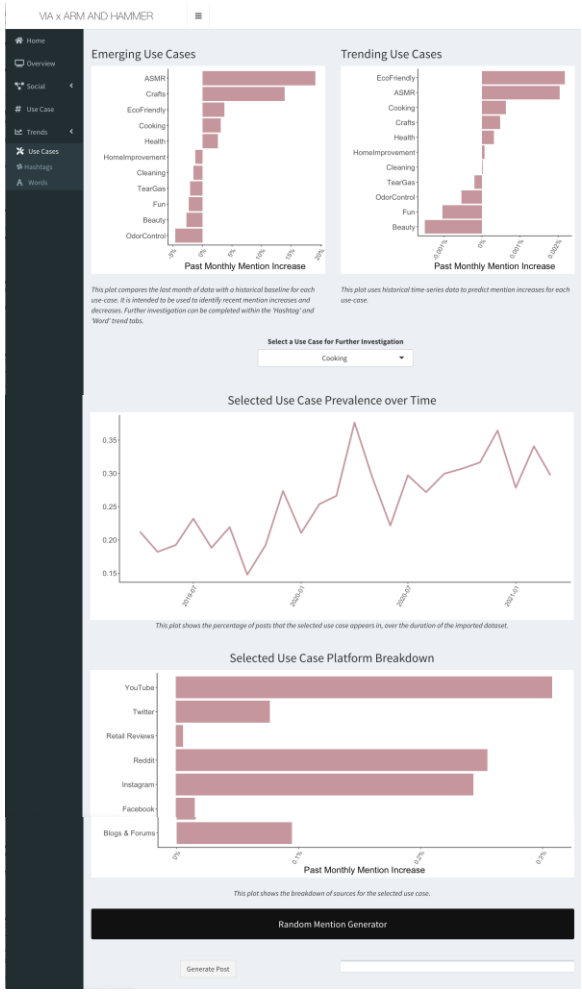
Select One

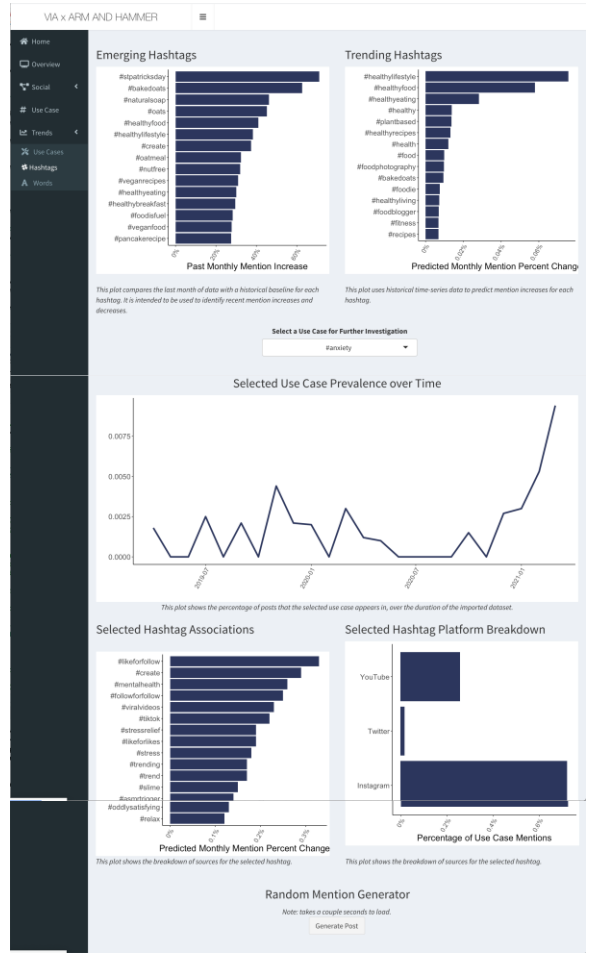
Essential Oils

Percent of Mentions Referencing Use Case

Each mention in the dataset is assigned to or excluded from a use case based on the presence of key words and hashtags that are indicative of product use. The categories themselves are not exclusive; some mentions will be identified with more than one use. This bar chart shows the percentage of mentions in the dataset that include keywords belonging to the different use cases.




VIA x ARM AND HAMMER
Home
Overview
Social
Use Case
Trends
Use Cases
Hashtags
Words
Emerging Use Cases
ASMR
Crafts
EcoFriendly
Cooking
Health
HomeImprovement
Cleaning
TearGas
Fun
Beauty
OdorControl
Past Monthly Mention Increase
Trending Use Cases
EcoFriendly
ASMR
Cooking
Crafts
Health
HomeImprovement
Cleaning
TearGas
OdorControl
Fun
Beauty
Past Monthly Mention Increase
This plot compares the last month of data with a historical baseline for each use-case. It is intended to be used to identify recent mention increases and decreases. Further investigation can be completed within the 'Hashtag' and 'Word' trend tabs.
This plot uses historical time-series data to predict mention increases for each use-case.
Select a Use Case for Further Investigation
Cooking
Selected Use Case Prevalence over Time
0.35
0.30
0.25
0.20
0.15
2019-07
2020-01
2020-07
2021-01
This plot shows the percentage of posts that the selected use case appears in, over the duration of the imported dataset.
Selected Use Case Platform Breakdown
YouTube
Twitter
Retail Reviews
Reddit
Instagram
Facebook
Blogs & Forums
Past Monthly Mention Increase
This plot shows the breakdown of sources for the selected use case.
Random Mention Generator
Generate Post


VIA x ARM AND HAMMER
Home
Overview
Social
Use Case
Trends
Use Cases
Hashtags
Words
Emerging Hashtags
#stpatricksday
#bakedoats
#naturalsoap
#oats
#healthyfood
#healthylifestyle
#create
#oatmeal
#nutfree
#veganrecipes
#healthyeating
#healthybreakfast
#foodisfuel
#veganfood
#pancakerecipe
Past Monthly Mention Increase
Trending Hashtags
#healthylifestyle
#healthyfood
#healthyeating
#healthy
#plantbased
#healthyrecipes
#health
#food
#foodphotography
#bakedoats
#foodie
#healthyliving
#foodblogger
#fitness
#recipes
Predicted Monthly Mention Percent Change
This plot compares the last month of data with a historical baseline for each hashtag. It is intended to be used to identify recent mention increases and decreases.
This plot uses historical time-series data to predict mention increases for each hashtag.
Select a Use Case for Further Investigation
#anxiety
Selected Use Case Prevalence over Time
0.0075
0.0050
0.0025
0.0000
2019-07
2020-01
2020-07
2021-01
This plot shows the percentage of posts that the selected use case appears in, over the duration of the imported dataset.
Selected Hashtag Associations
#likeforfollow
#create
#mentalhealth
#followforfollow
#viralvideos
#tiktok
#stressrelief
#likeforlikes
#stress
#trending
#trend
#slime
#asmrtriggers
#oddlysatisfying
#relax
Predicted Monthly Mention Percent Change
This plot shows the breakdown of sources for the selected hashtag.
Selected Hashtag Platform Breakdown
YouTube
Twitter
Instagram
Percentage of Use Case Mentions
This plot shows the breakdown of sources for the selected hashtag.
Random Mention Generator
Note: takes a couple seconds to load.
Generate Post


VIA x ARM AND HAMMER
Home
Overview
Social
Use Case
Trends
Use Cases
Hashtags
Words
Emerging Words
wool
volcano
chemistry
dissolve
react
scraper
swish
alcohol
steel
brew
mum
wine
flare
beverage
blanket
boot
acid
bicarb
eyebrow
mail
bitter
kitty
fluid
sprain
belly
Past Monthly Mention Increase
Trending Words
face
treat
dish
rub
dissolve
water
vinegar
whisk
texture
paste
sun
starch
spot
feel
kitchen
volcano
sand
grain
alcohol
beer
sprinkle
fragrance
fudge
hair
react
Predicted Monthly Mention Percent Change
This plot compares the last month of data with a historical baseline for each use-case. It is intended to be used to identify recent mention increases and decreases.
This plot uses historical time-series data to predict mention increases for each word.
Select a Use Case for Further Investigation
alcohol
Selected Word Prevalence over Time
0.020
0.015
0.010
0.005
2019-07
2020-01
2020-07
2021-01
This plot shows the percentage of posts that the selected use case appears in, over the duration of the imported dataset.
Selected Word Associations
water
oil
ingredient
mix
salt
rub
clean
need
sugar
skin
remove
powder
cup
product
vinegar
recipe
tsp
food
recipes
tsp
food
dry
butter
Predicted Monthly Mention Percent Change
This plot shows the breakdown of sources for the selected word.
Selected Word Platform Breakdown
YouTube
Twitter
Retail Reviews
Reddit
Instagram
Facebook
Blogs & Forums
Blogs & Forums
Predicted Monthly Mention Percent Change
This plot shows the breakdown of sources for the selected word.
Random Mention Generator
Note: takes a couple seconds to load.
Generate Post