VEGAN

TWITTER ANALYSIS

GROUP 23

BIG DATA

13/12/22

WHO ARE WE?

Viktor

Konstantin

Lennert

Artur

Wouter

DISCUSS THESE POINTS FOR AN EFFICIENT PRESENTATION

PRESENTATION OF 20 MINUTES

DISCUSS INSIGHTS IN MANAGERIAL PRESENTATION

DISCUSS TECHNICALITIES IN TECHNICAL PRESENTATION

Q&A OF 20 MINUTES

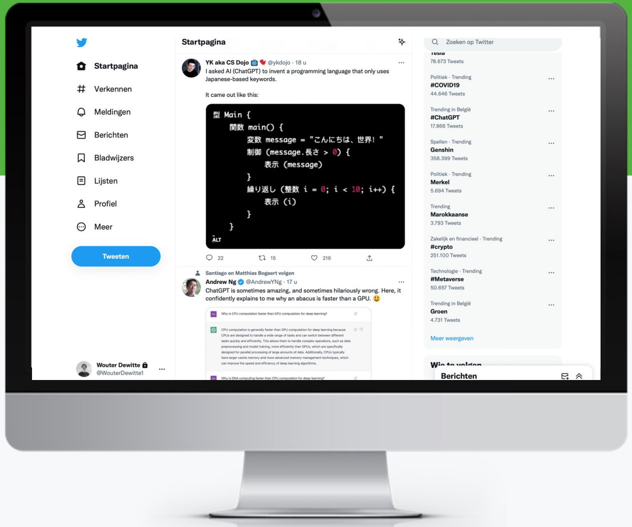
INDIVIDUAL QUESTIONS FOR 7 MINUTES

MANAGERIAL PRESENTATION

A presentation about insights and conclusions

LEARN THE WAYS OF TWITTER FOR OPTIMAL INSIGHTS

These **properties** make Twitter **different** from other platforms

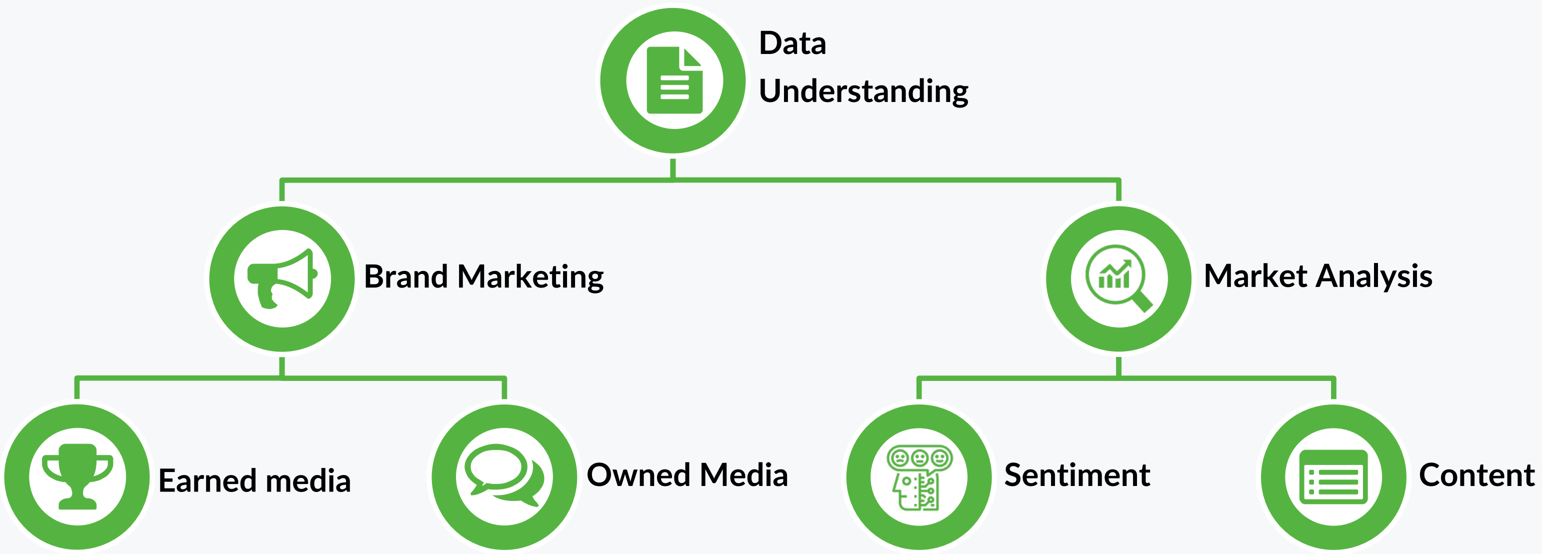


HOW IS TWITTER DEFINED?

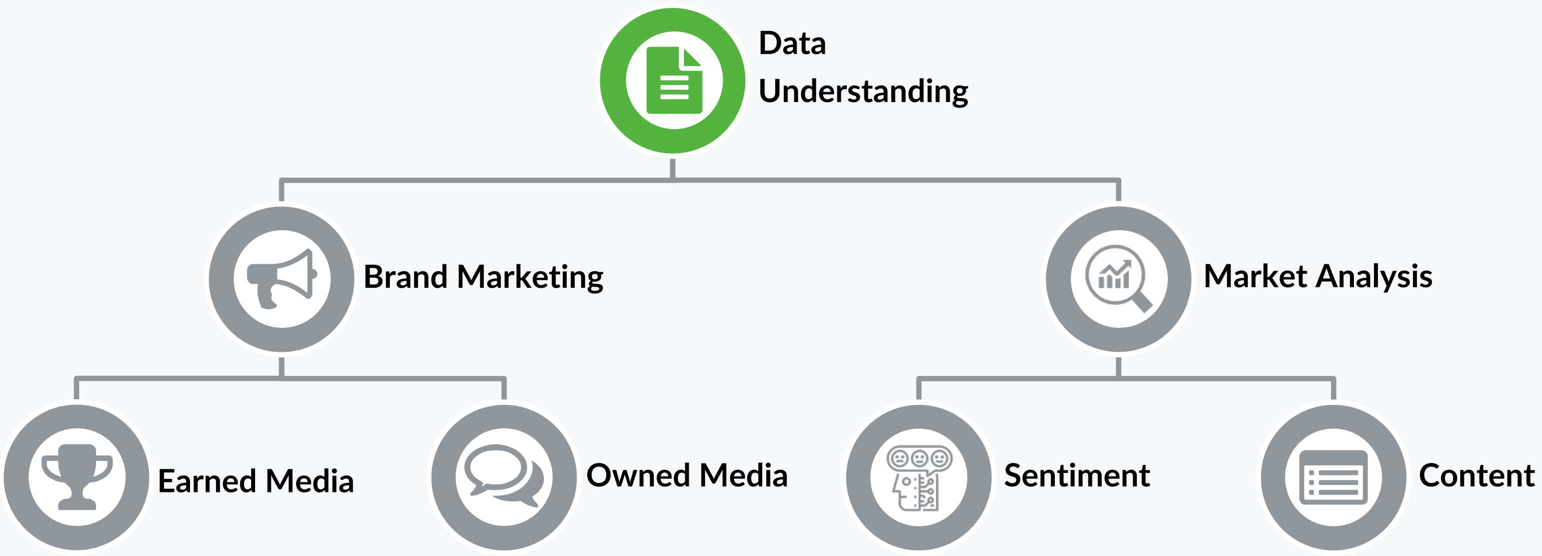
"Twitter is a **microblogging**, **social** networking service owned by American company Twitter, Inc., on which users **post** and **interact** with messages known as "**tweets**". "

- **Short** messages with a **concise** message
- **Source** of **information**
- **Large network** of interconnected users
- **Interactive** platform

VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



UNDERSTAND YOUR DATA TO CREATE VALUABLE INSIGHTS

Understand the data and then **check** the **generalizability**

VEGAN, VEGANISM ...

PERIOD 25/10/21 - 11/10/22

VOLUME OF 1.3M TWEETS

CONCLUSION FROM THE MODEL

PREDICTIVE POWER

The features extracted from the data have **predictive power** for the ETF performance

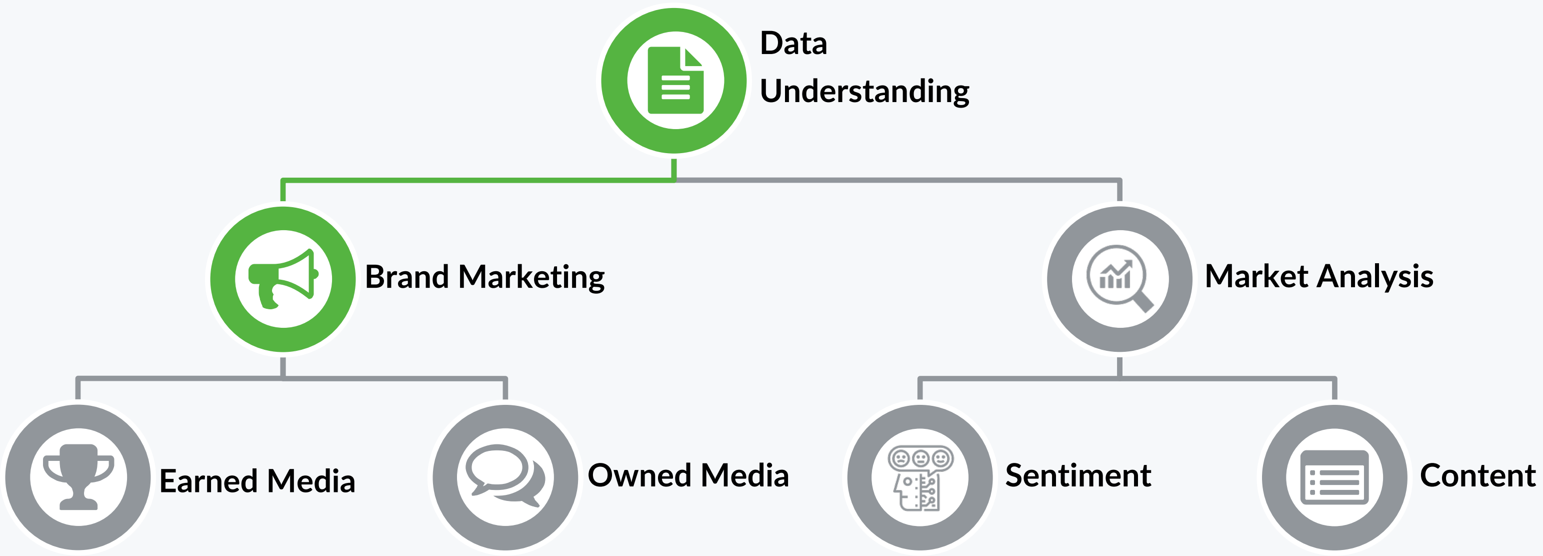
REPRESENTATIVE

The data is **representative** of the **global vegan market**

OTHER INSIGHTS

The data is a **valid source** for the extraction of **other useful insights**

VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



USE INSIGHTS TO MAXIMIZE BRAND MARKETING IMPACT TO DEVELOP YOUR BRAND

6 **reasons** why **branding** and **thus brand marketing** is so **important** for your business

RECOGNITION

Your brand will be more recognizable. People will **take note of it.**

TRUST

Branding can help **build trust** with your customers.

EMPLOYEES

Your employees will be **more motivated** to work for a company with a good brand

LOYALTY

Great branding creates loyal customers. They will continue to **come back.**

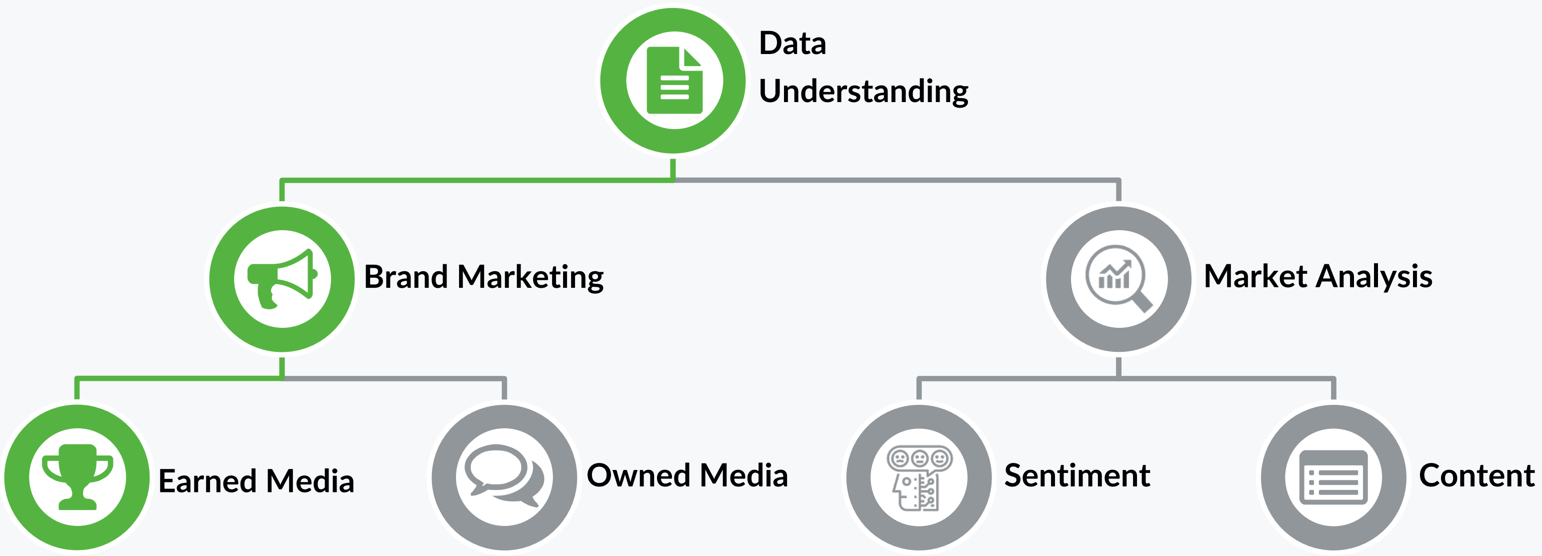
COMPETITIVE ADVANTAGE

Branding can help you **differentiate** and give you a competitive advantage.

BRAND EQUITY

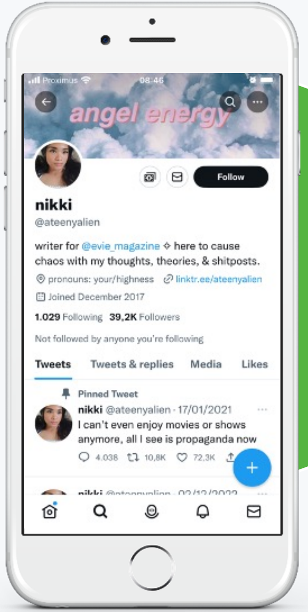
Additional value created for the business because of the brand.

VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



OPTIMIZE YOUR EARNED MEDIA BY USING INFLUENCERS DEFINED BY THRESHOLDS

Make use of **3 thresholds** and **one tip** to **find influencers**



AMOUNT OF FOLLOWERS

High amount of followers shows **high popularity**

LEVEL OF ENGAGEMENT

High level of engagement shows **high influence**

FREQUENCY

High frequency shows **active account**

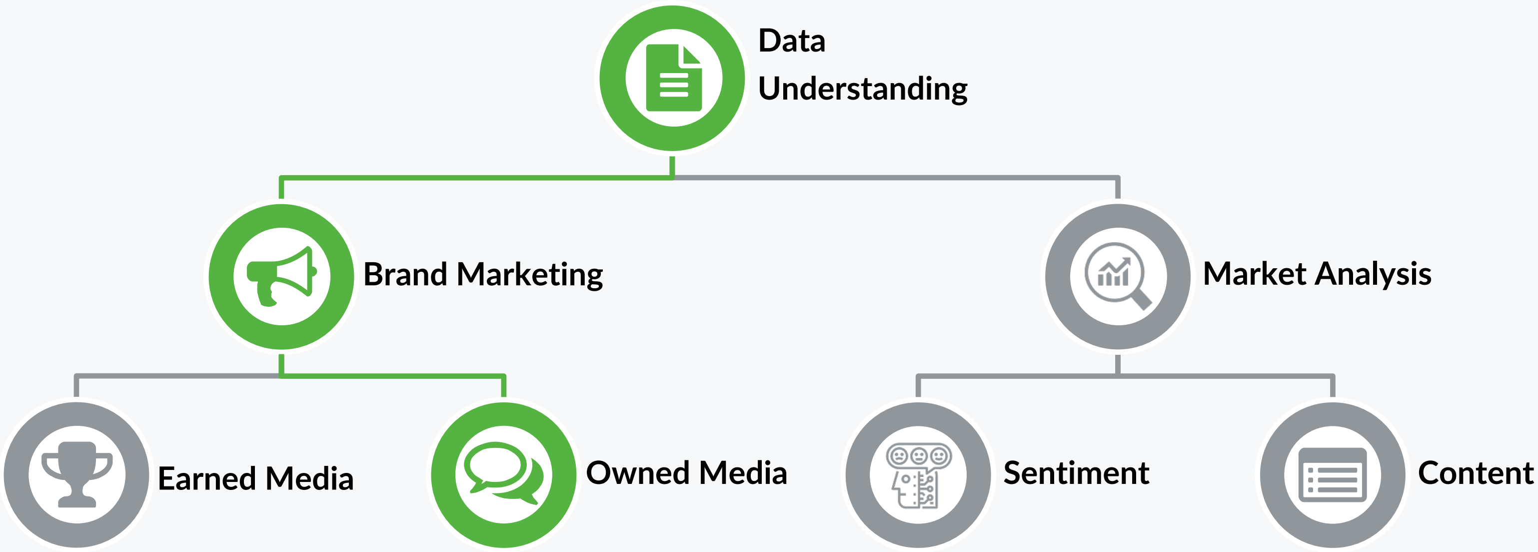
NEWER ACCOUNTS

Newer account means **higher engagement**

Top Influencers

1. amaneku_organic
2. ateenyalien
3. itsTFC
4. MissAudreyBoo
5. teonawrites

VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



OPTIMIZING OWNED MEDIA BY POSTING IN SUMMER IN USA AND UK AND TARGETING NEW ACCOUNTS

Optimize the use of your **owned** media based on **3** different **categories**

	Timing	Location	Age of Account
Engagement rate	Spring, Summer Weekend Morning	UK India USA	**Newer** means **more engagement**
Volume	December, Spring, Summer Weekday Evening	UK USA	**Very old** or **new**
Conclusion	**Spring and Summer Weekday**	**USA and UK**	**Younger accounts**

Engagement rate · Volume · Conclusion · Only on tweets with location

OPTIMIZING OWNED MEDIA BY MAXIMIZING ENGAGEMENT RATE

Use these **do's** and **don'ts** to create **better-performing tweets**

TARGET **NEWER** ACCOUNTS

POST OFTEN IN **FEBRUARY, JUNE, JULY AND NOVEMBER**

USE **TEXT** AND AVAILABLE **SPACE**

VS

DON'T FOLLOW **TOO MANY ACCOUNTS**

DON'T POST FREQUENTLY FROM **MARCH TO MAY**

DON'T BE **OBJECTIVE** OR **NEGATIVE**

GOOGLE TRENDS

SEASONAL EFFECTS

Summer and especially **winter** seasons peak

INCREASE UNTIL 2020

General popularity on Google from 2010 to 2020 increased exponentially

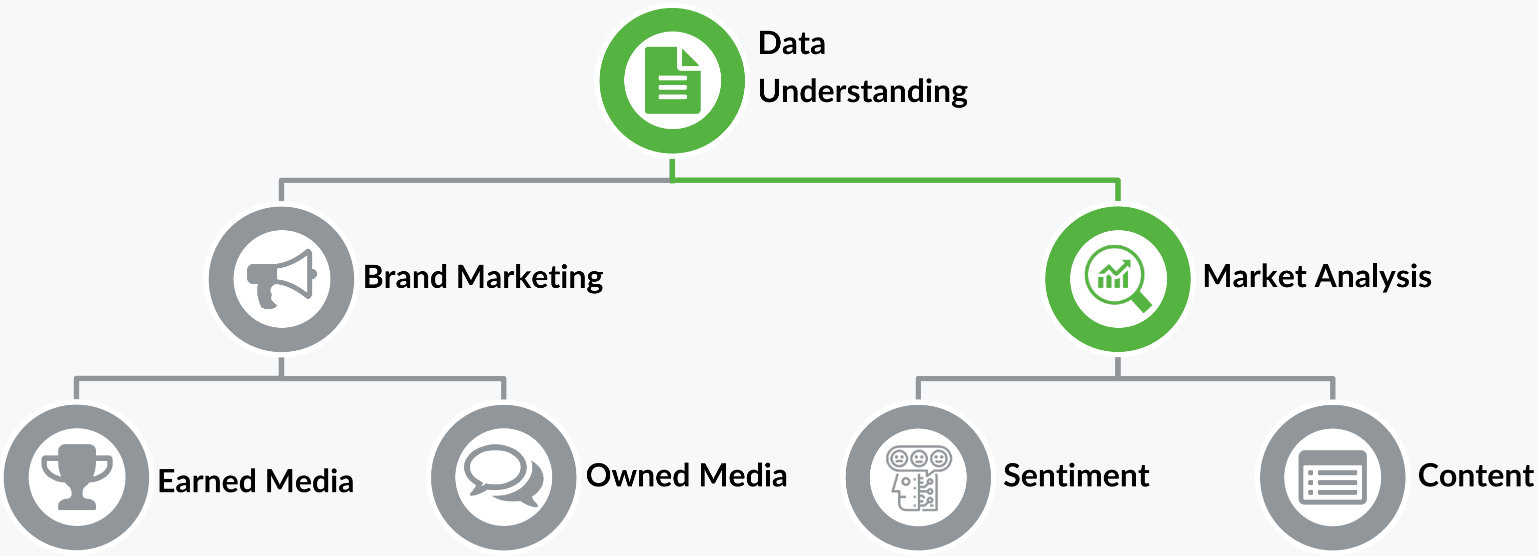
STAGNATING TREND

In 2020 popularity peaked and stagnated the following years

FORECAST

We expect the seasonality effects to be still present and about the same popularity for the following 3 years

VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



USE INSIGHTS TO KNOW MORE ABOUT YOUR MARKET AND HAVE THE ADVANTAGE

RISK REDUCTION

Market analysis will **help your decision-making** so there will be less risk.

TARGETING

With market knowledge, you can **target** the **right customers.**

EMERGING TRENDS

Staying ahead means **spotting new opportunities** and **trends**.

GROWTH OPPORTUNITIES

Knowing the market gives you the opportunity to **grow** and **expand**.

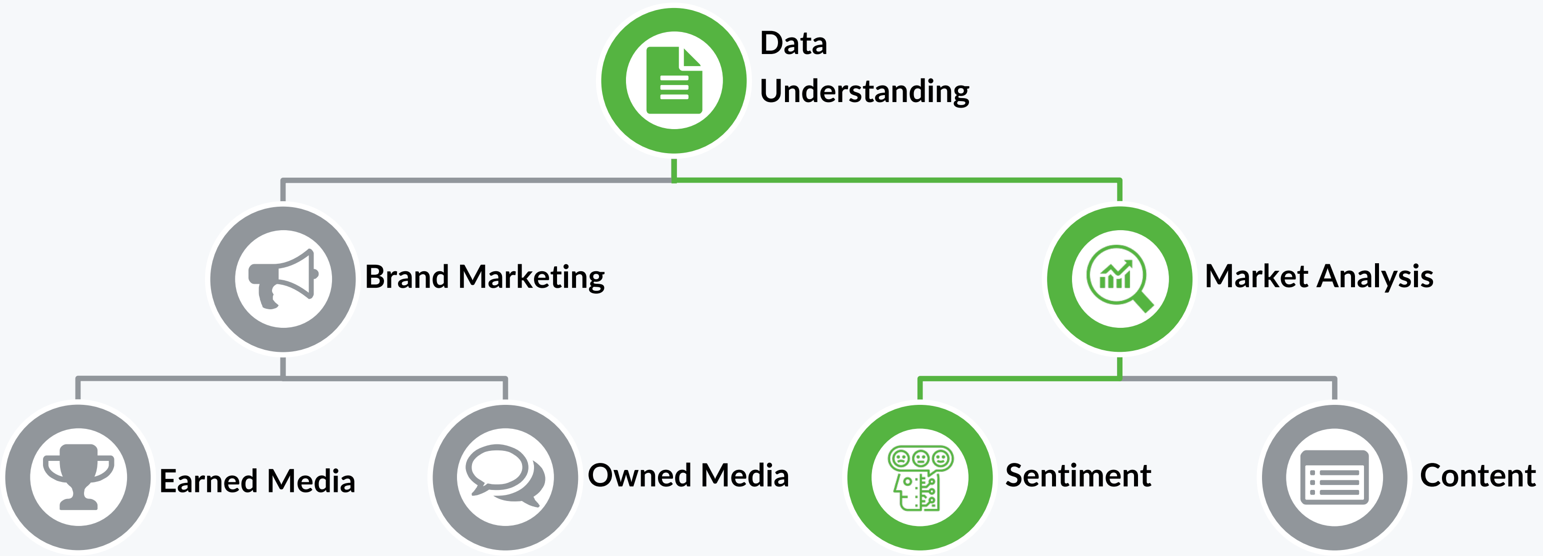
LEARN FROM MISTAKES

Marketing analytics can **explain mistakes** from the past.

REVENUE PROJECTION

Forecast your **future revenues** by using your market analysis.

VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



LEARN ABOUT THE EMOTIONS AROUND YOUR BRAND TO KNOW WHEN THE REACTION IS MOST POSITIVE

2 periods of the **year** that have **better sentiment**

01 **WINTER SEASON**

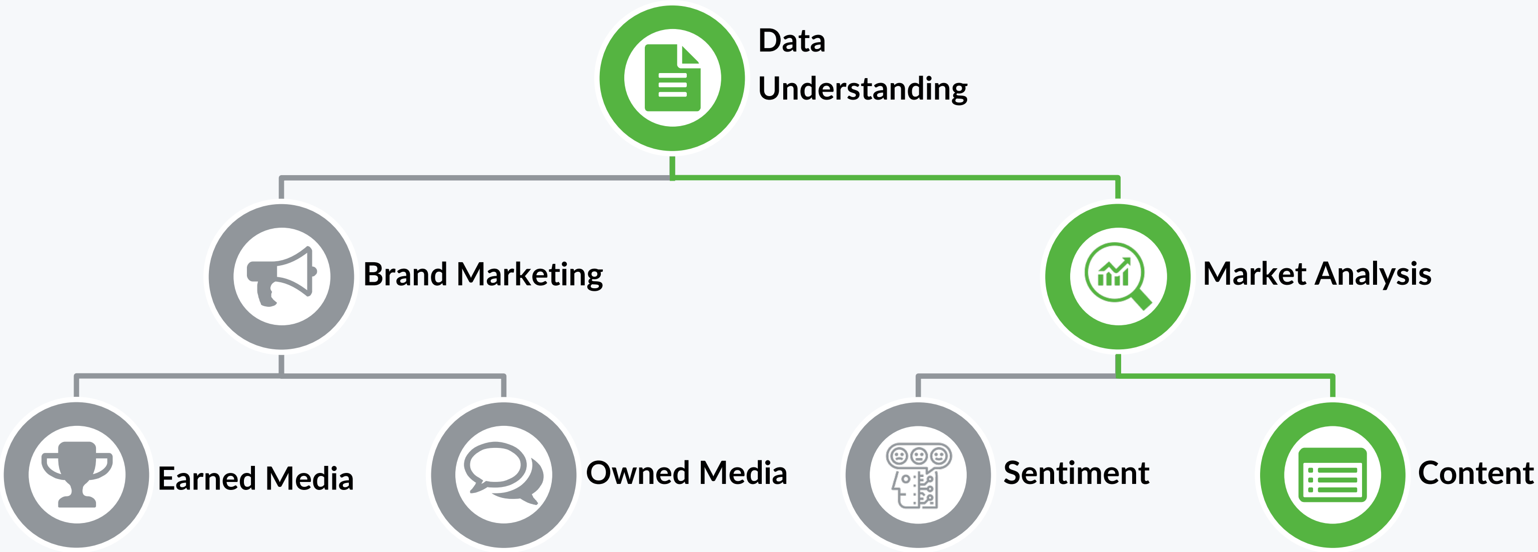
Because of good intentions for the new year

02 **JUST BEFORE SUMMER**

Because of the healthy summer goals

Sentiment means the **emotion**, **expressions**, and **feelings** that people **express** in tweets.

VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



LEARN ABOUT THE CONTENT OF YOUR BRAND TO KNOW WHAT TO POST ABOUT

These **three topics** define our **vegan** brand on **Twitter**

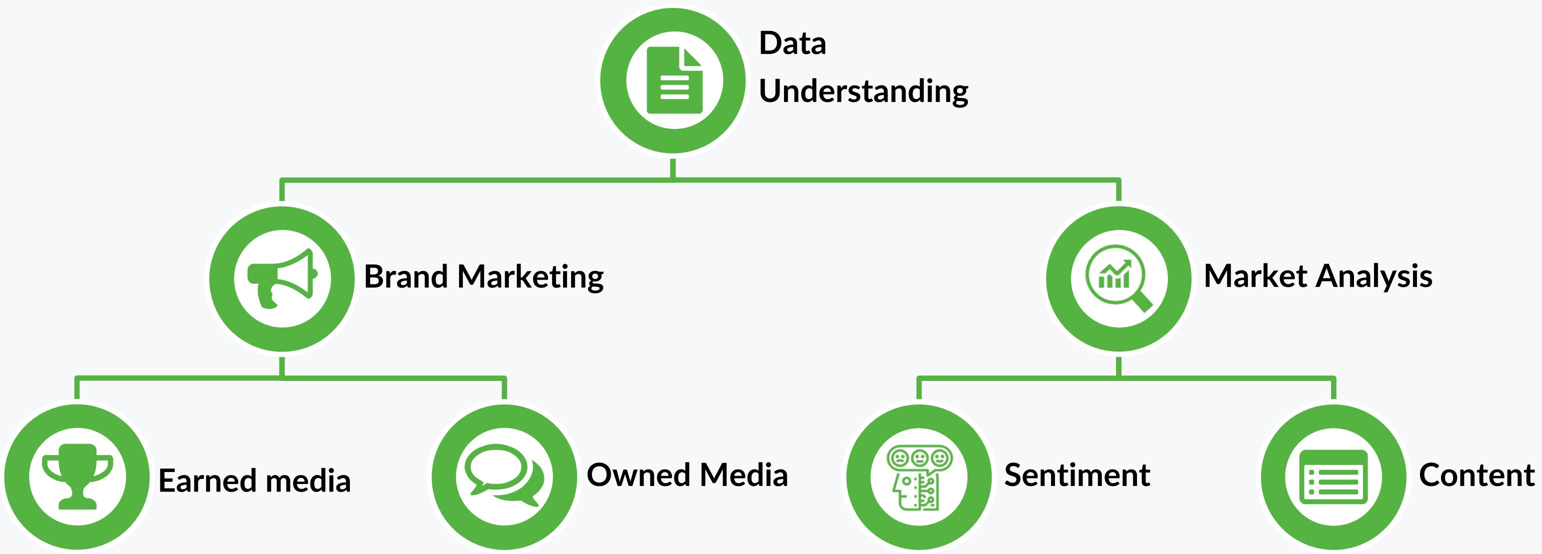
Branding	Reasons	Cooking
Talking about **beauty**, **gifts**, and so on…	**Reasons** to **become vegan**, f.e. **animals, diet**, and so on…	Focus on **recipes, plant-based, organic, and so on…**
Make the **market** and the **branding** around **vegan attractive**	Know what **people** find **important** about **being vegan**	Use **cooking** to **increase** vegan **popularity** and **awareness**

What is the topic about?

Derived objectives

Winner for engagement and volume compared to other topics

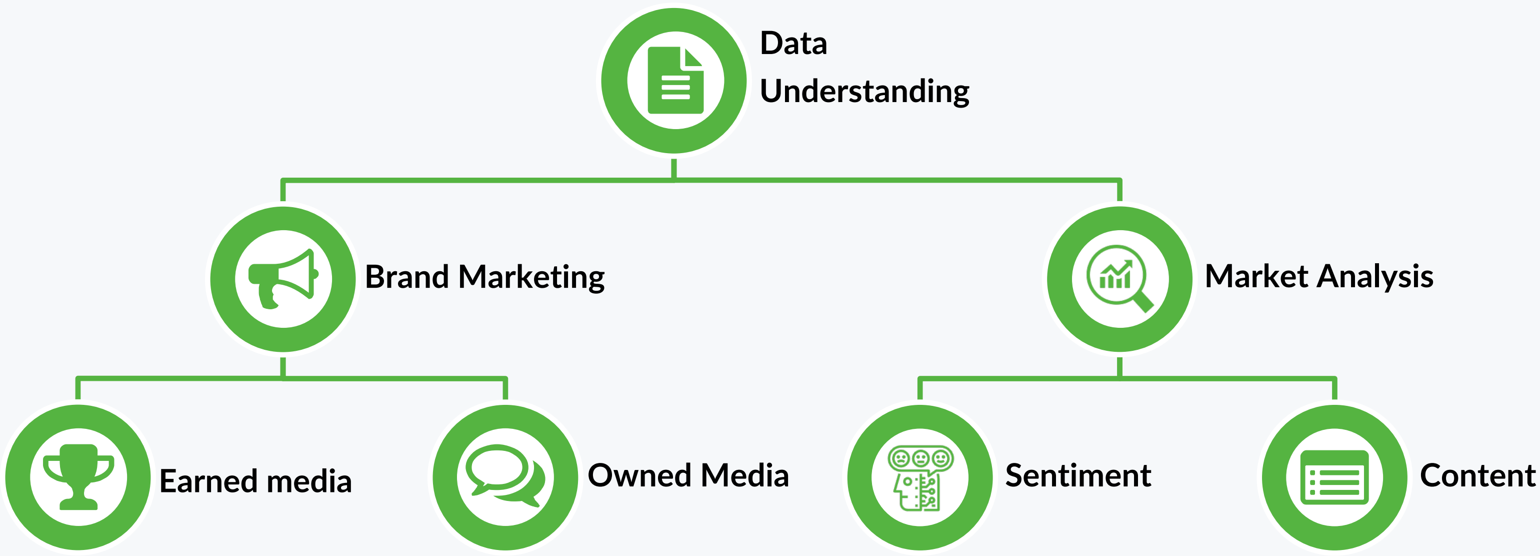
VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



TECHNICAL PRESENTATION

A presentation about our models and technical implications

VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



LIMITATIONS OF THE DATA AND SOFTWARE

Know the **limitations** so that you can create **meaningful results.**

Out of Memory Error

Not enough memory to compile model 2 on whole dataset, only able to use 5% of the tweets

Data limitations

Not able to obtain furhter information on the user: limiting network analysis

Not able to translate tweets

Not able to translate tweets due to PC memory constraints

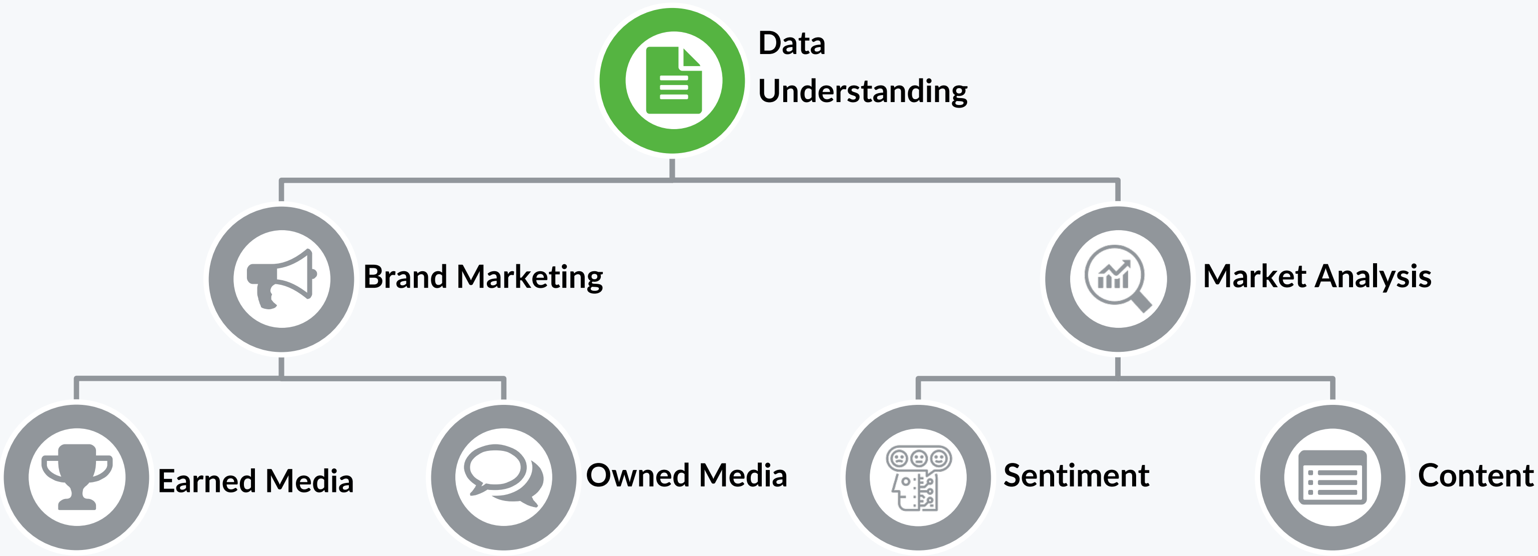
No replies

No information on the replies was available

Missing data

Very little information on the geographic location of the tweets

VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



GENERAL ML PIPELINE

Load in the data	Data Cleaning	Feature engineering	Model Building	Model Testing
Tweets in JSON structure	**Converting variables**	**Model specific features**	**Setting up PySpark Pipeline**	**Deploying model on test set**
Unnesting the data	Cleaning tweet text	Scaling	Random Forest	Cross validation
	Necessary groupings	Tokenization	Logistic regression	Hyperparameter tuning
	Converting to datetime	Feature selection		
	Removing duplicates, spam accounts and retweets			

MODEL 1 : PREDICTING VEGAN ETF TRACKER

BASETABLE CREATION FOR ETF PREDICTION

Independent

Favorites
Retweets
Followers
→ Sum

Engagement
Emojis
Number of words
Number of hashtags
Number of mentions
Num Exclamation marks
Polarity
Subjectivity
→ Average

Amount of tweets → Count

Dependent

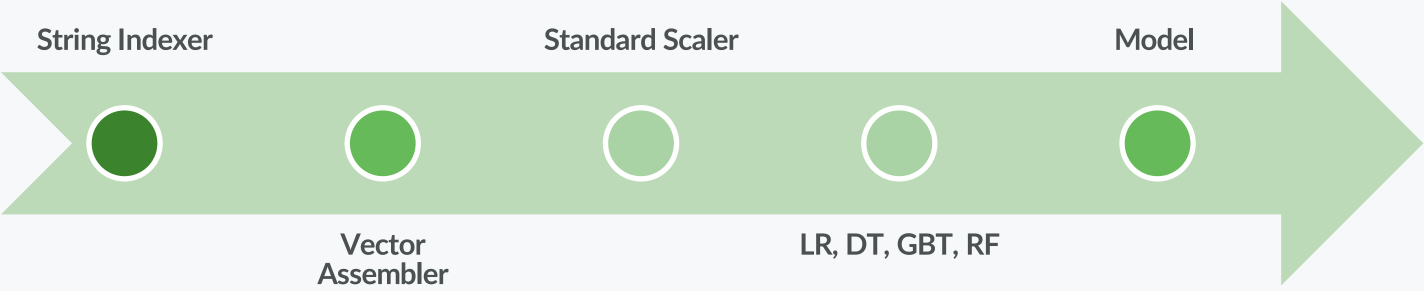
Binary Price Indicator

Join On

Date

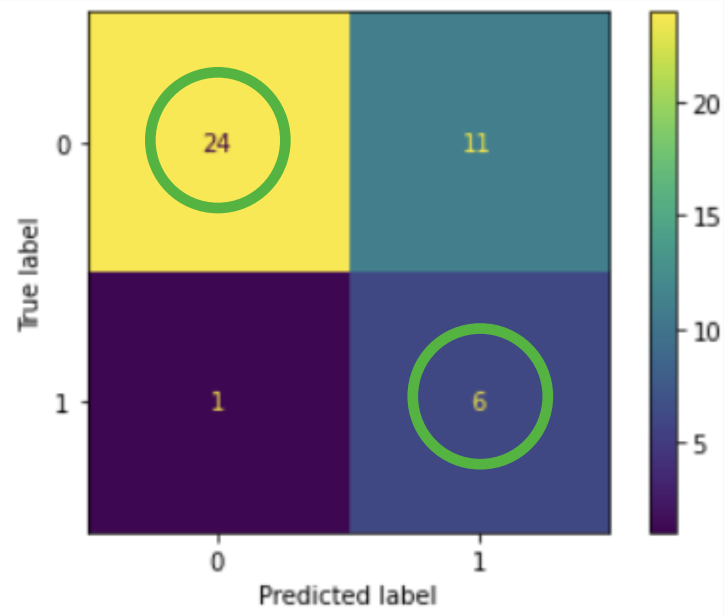
ML **PIPELINE** VEGAN TO PREDICT **BINARY ETF TRACKER**

Showing the **different steps** in the **binary predictive ETF** model

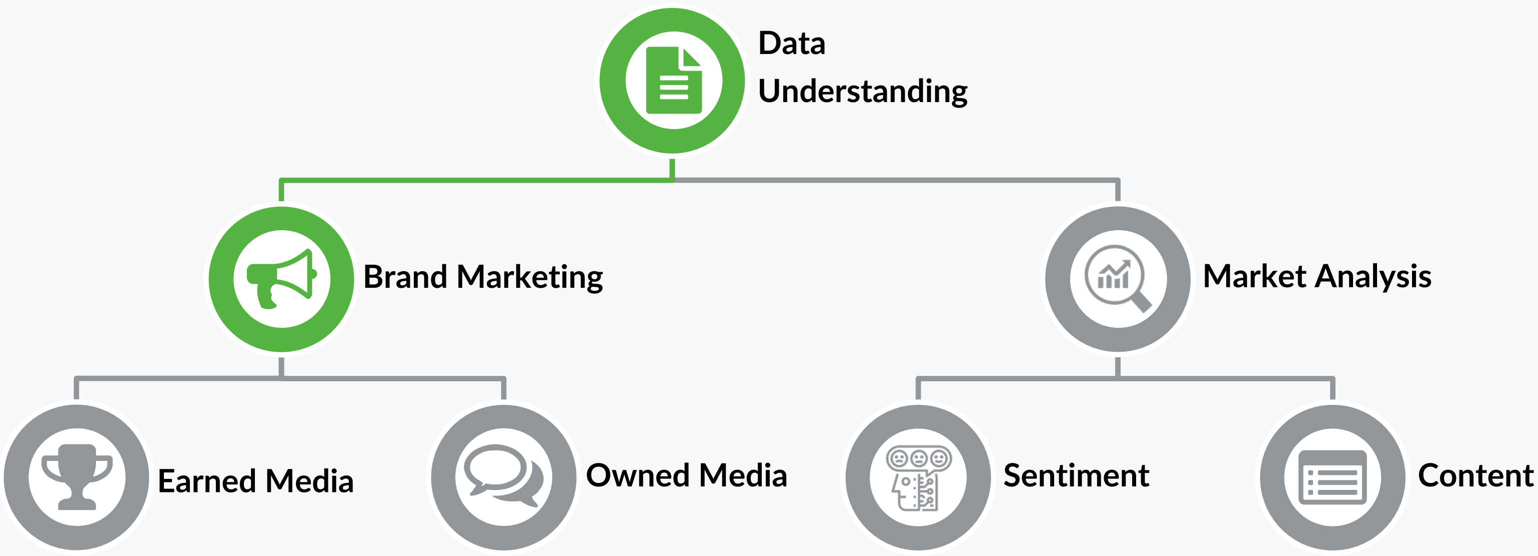


VISUALIZATION OF MODEL EVALUATION BINARY ETF TRACKER

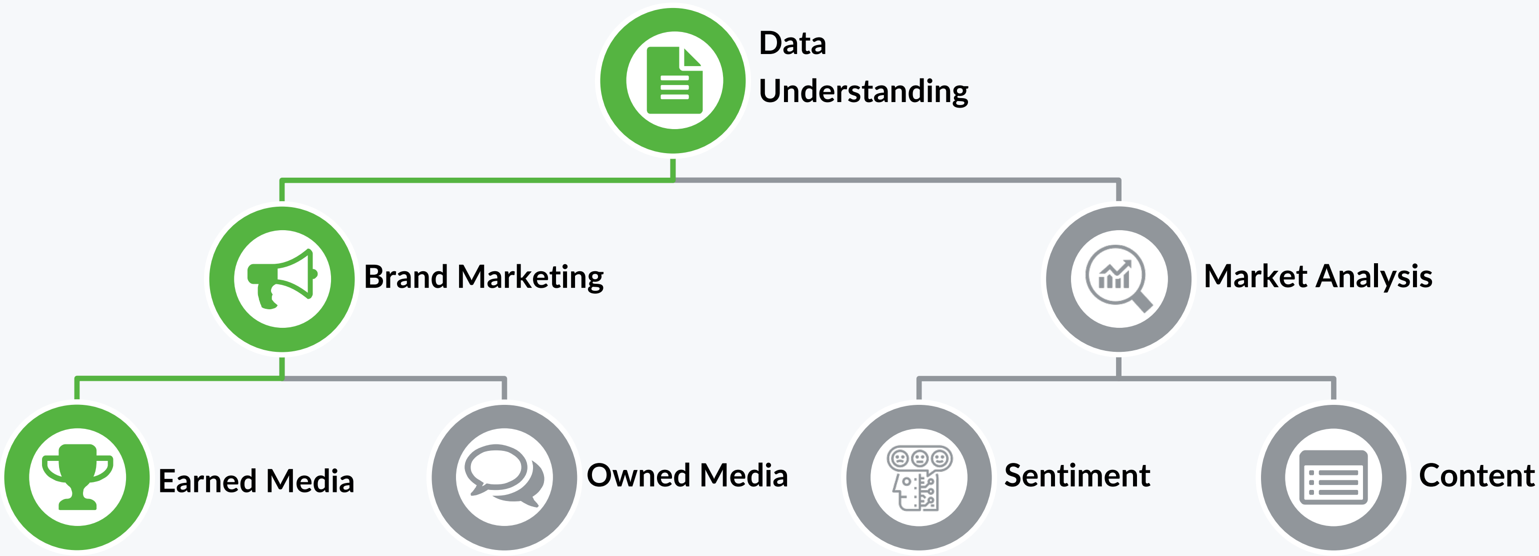
Sensitivity = 85% , Specificity = 69%



VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



OPTIMIZE YOUR **EARNED** MEDIA BY **USING INFLUENCERS**

Explanation of **thresholds** and the **function** that **determines** the **influencer accounts**

10.000 FOLLOWERS

High amount of followers shows **high popularity**

4% LEVEL OF ENGAGEMENT

High level of engagement shows **high influence**

2 TWEETS PER WEEK

High frequency shows **active account**

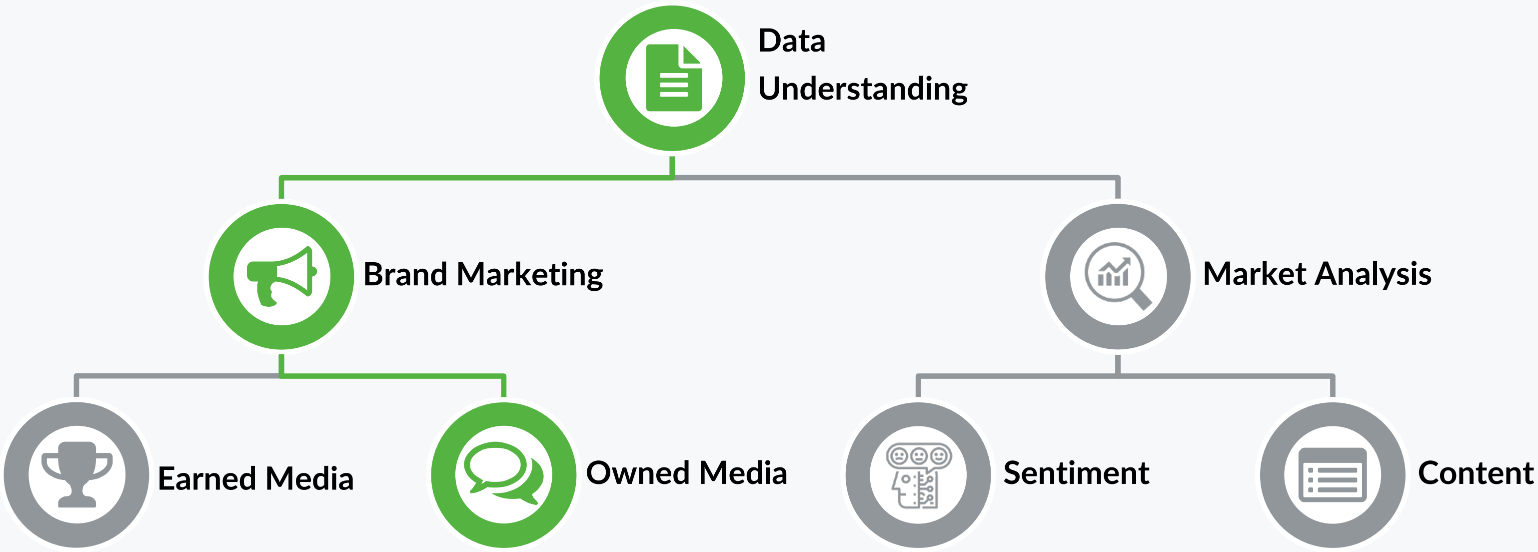
GROUP BY → **JOIN DATAFRAMES** → **FILTER ON THRESHOLDS**

Group the information of followers, engagement and frequency **per user**

Join all the **information** in **one table**

Filter the information on the **thresholds** so that we only have **relevant information**

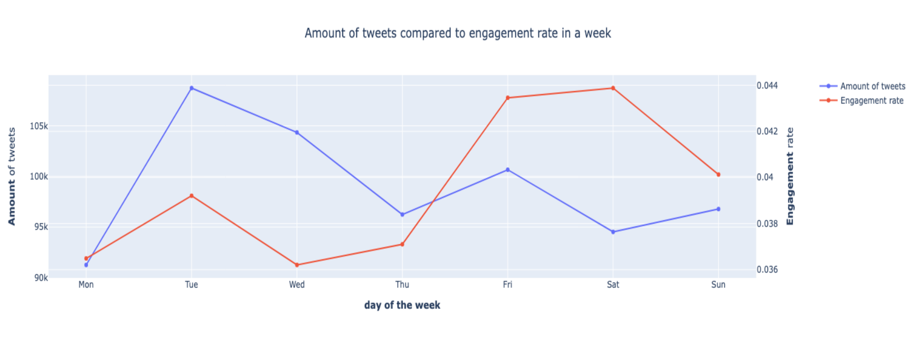
VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



OPTIMIZING OWNED MEDIA BY USING DATA

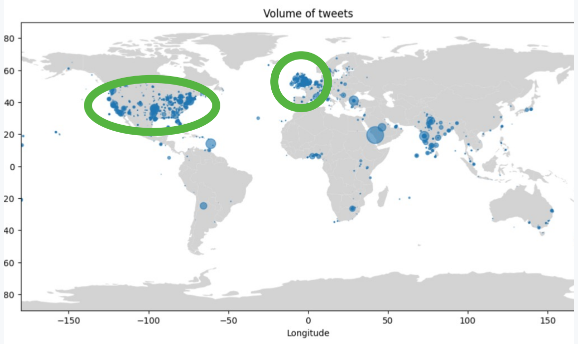
Post in Spring/Summer and Weekdays

These periodes have high volume and engagement



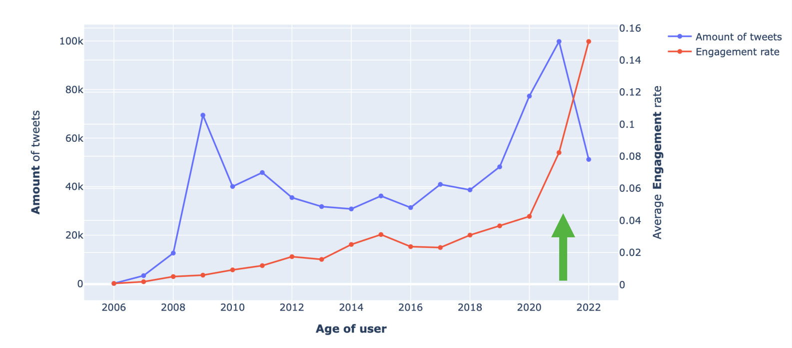
USA & UK

These regions show much larger volume



Target younger accounts

These show much higher engagement



MODEL 2: PREDICTING ENGAGEMENT

BASETABLE CREATION FOR EGAGEMENT PREDICTION

Independent

- User followers
- Age of user's account
- Number of emojis in the tweet
- Number of uppercase words
- Number of words
- Number of hashtags
- Number of mentions
- Number of exclamation marks
- Number of text characters
- Engagement
- Polarity
- Subjectivity

Numerical variables

- Type of Media
- Hour of the day
- Day of the week
- Month
- Day of the month
- Indicator of quoted
- Indicator of symbols
- Indicator of verification

Categorical variables

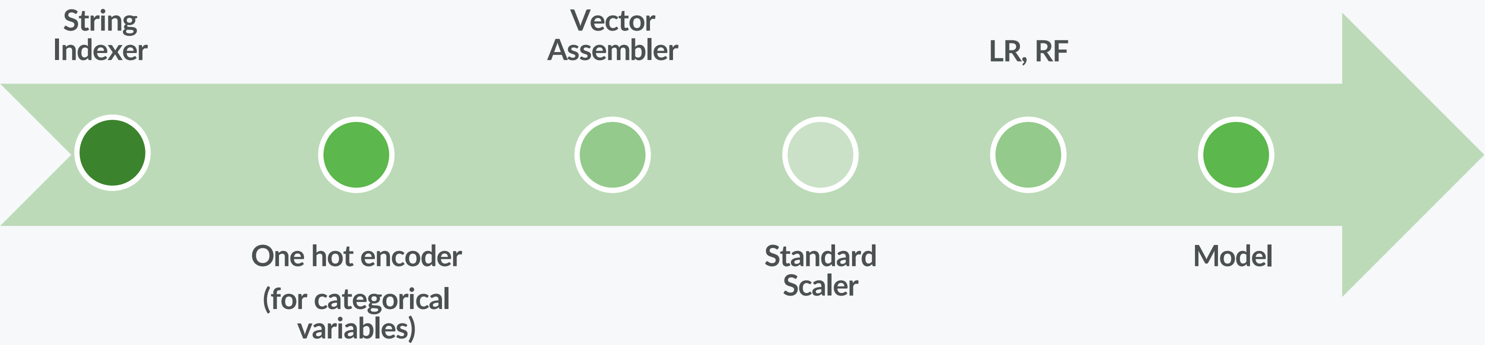
Dependent

Engagement rate

$$= \frac{\text{(count of favorites)} + \text{(count of retweets)}}{\text{(number of followers)}}$$

This variable is converted into a categorical variable with 3 classes

ML PIPELINE VEGAN TO PREDICT CATEGORICAL ENGAGEMENT SCORE



MODEL PERFORMANCE

Random Forest with weights:

F1 : 0.8139

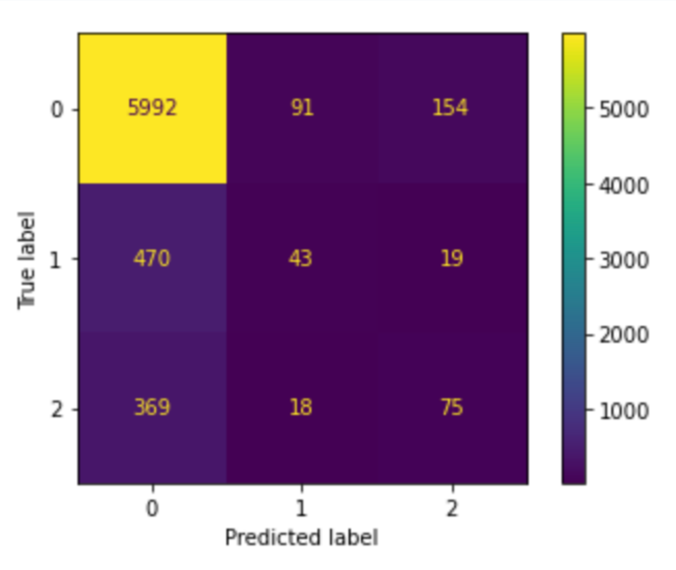
RECALL : 0.8446

PRECISION : 0.7748

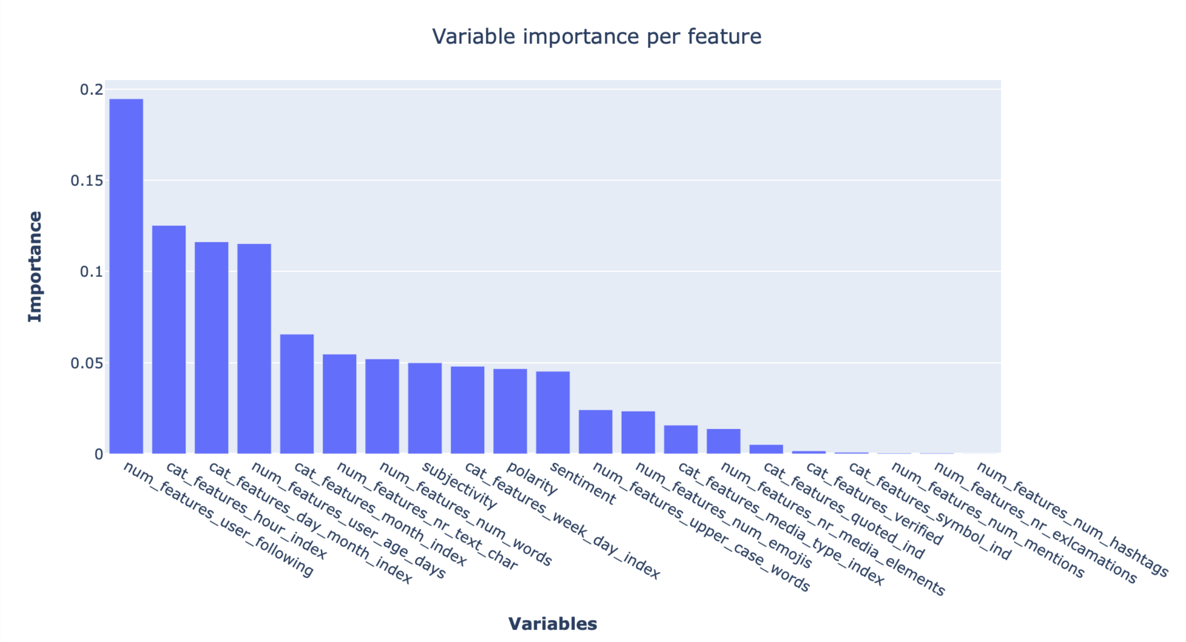
WEIGHTED_TPR : 0.8450

WEIGHTED_FPR : 0.7430

Confusion matrix Random Forest model with Ajudsted weights for class imabalance



VARIABLE IMPORTANCE

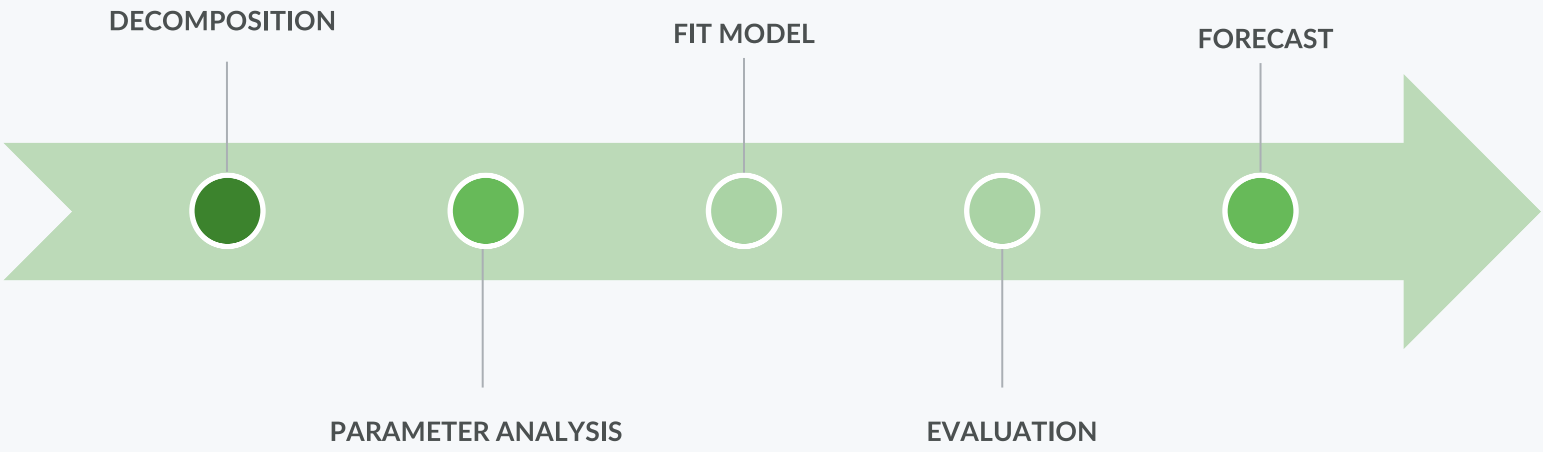


MODEL 3 : FORECASTING BASED ON HISTORICAL DATA

USING SARIMA

GOOGLE TRENDS SARIMA

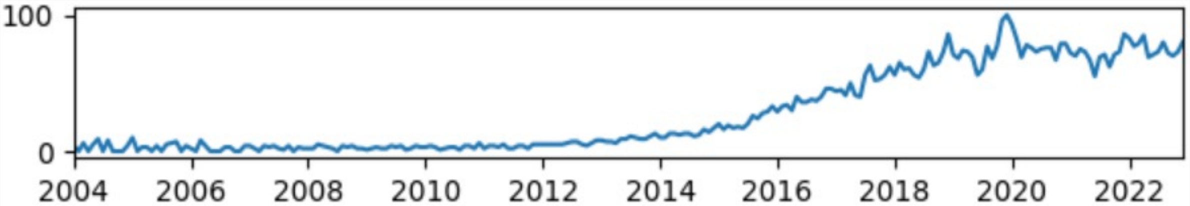
Seasonal Autoregressive Intgrated Moving Averages



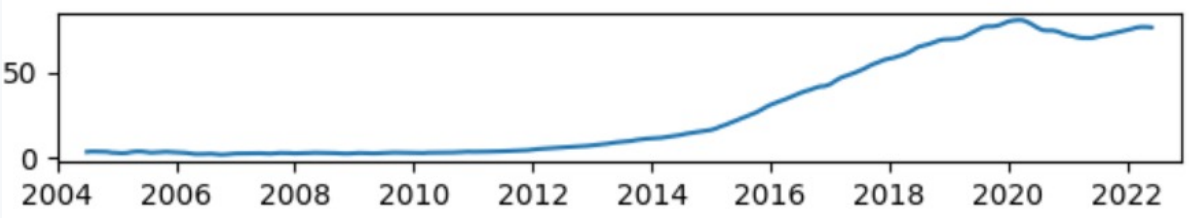
GOOGLE TRENDS SARIMA

DECOMPOSITION

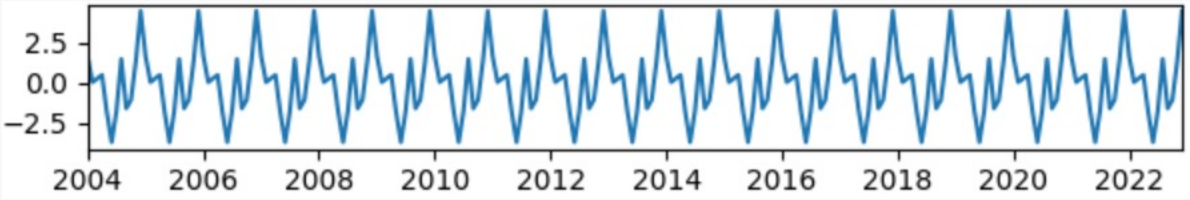
Google Trends: Vegan
from 1st of january 2004 till
11th of december 2022



Trend
component isolated

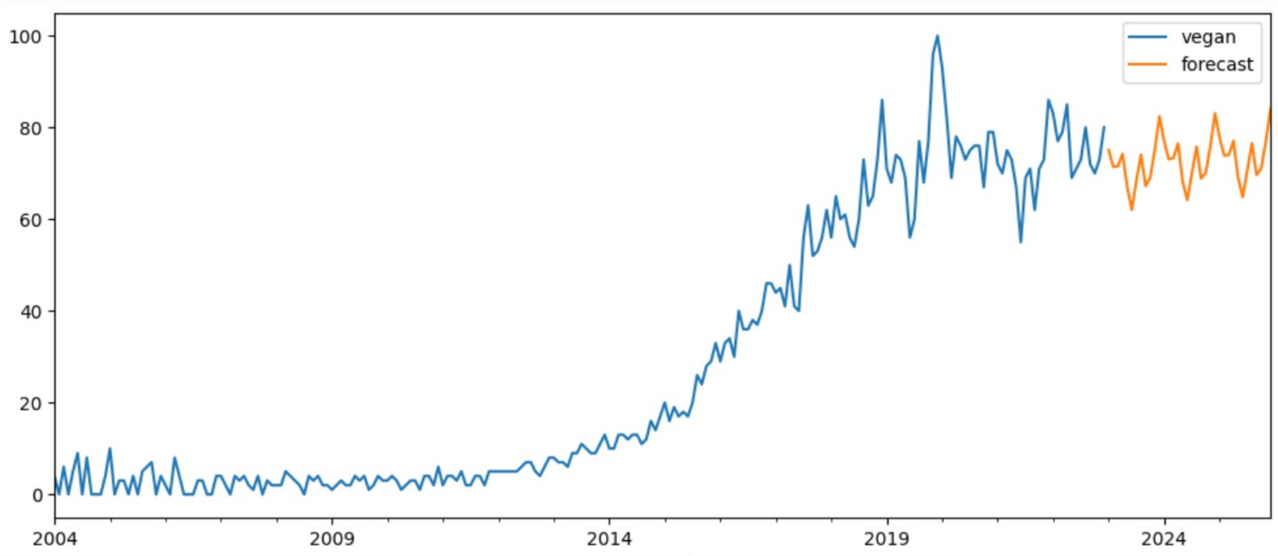


Seasonality
component isolated



GOOGLE TRENDS SARIMA

FORECASTING



Forecasts

for 3 consecutive years

Trending

same trending levels next 3 years

Seasonality

Seasonality plays a significant role

MODEL 4 : PREDICTING CHANGES IN GOOGLE TRENDS

Based on Twitter data

BASETABLE CREATION FOR GOOGLE TRENDS PREDICTION

Independent

Favorites
Retweets
Followers
Engagement
Emojis
Number of words
Number of hashtags
Number of mentions
Num Exclamation marks
Polarity
Subjectivity → Average

Amount of tweets → Count

Dependent

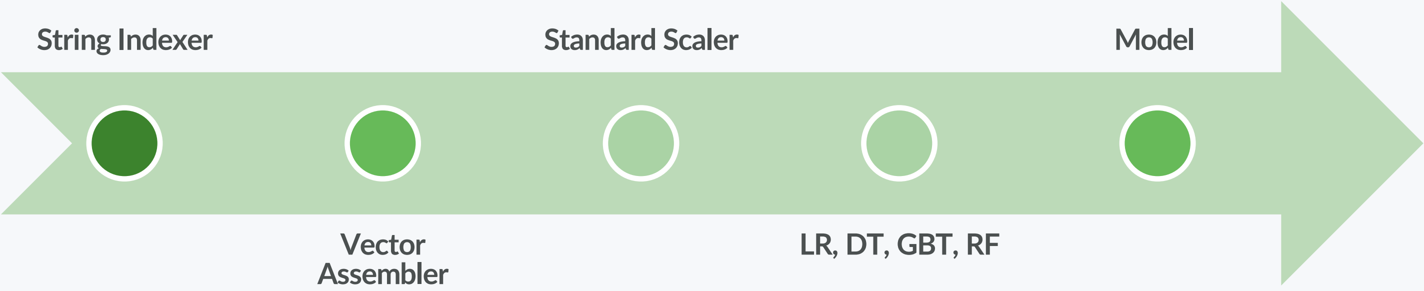
Binary Trend Indicator

Join On

Date

ML **PIPELINE** VEGAN TO PREDICT **BINARY GOOGLE TRENDS**

Showing the **different steps** in the **binary predictive Google trend** model



CONCLUSION GOOGLE TREND PREDICTION

NO PREDICTIVE POWER

Based on our **evaluation metrics** we found there is no predictive power in this model

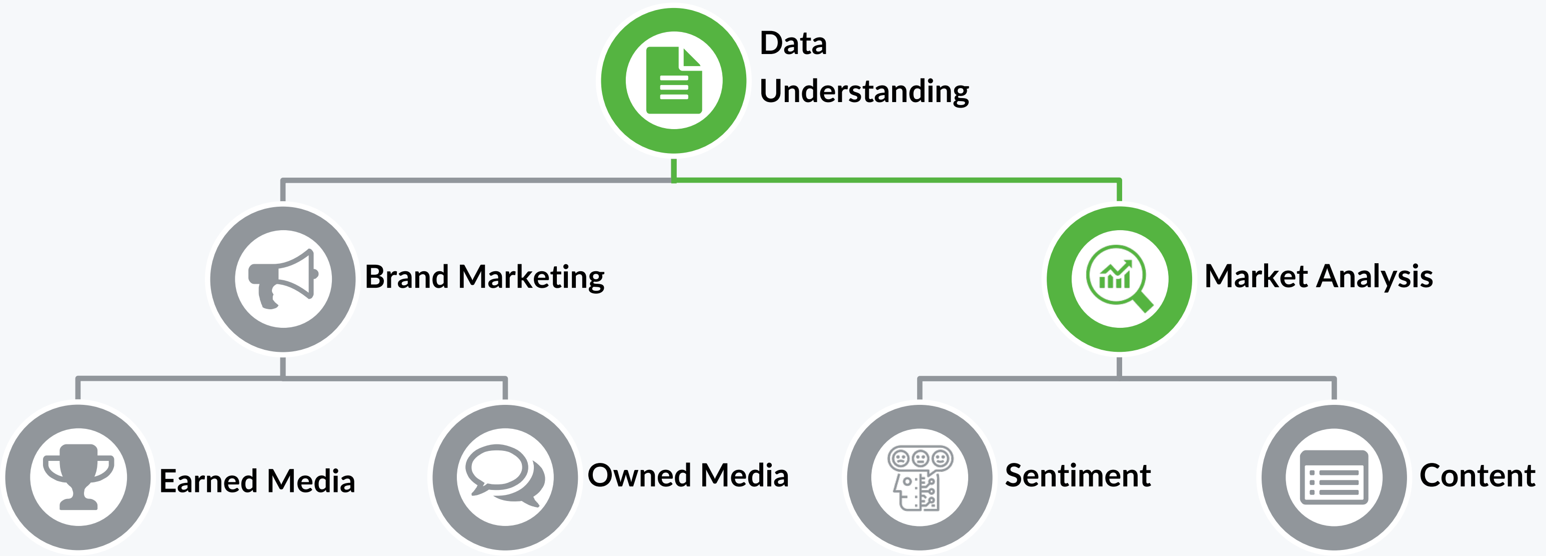
SEO

We found **no evidence** to use Twitter data for **Search Engine Optimization** in this context

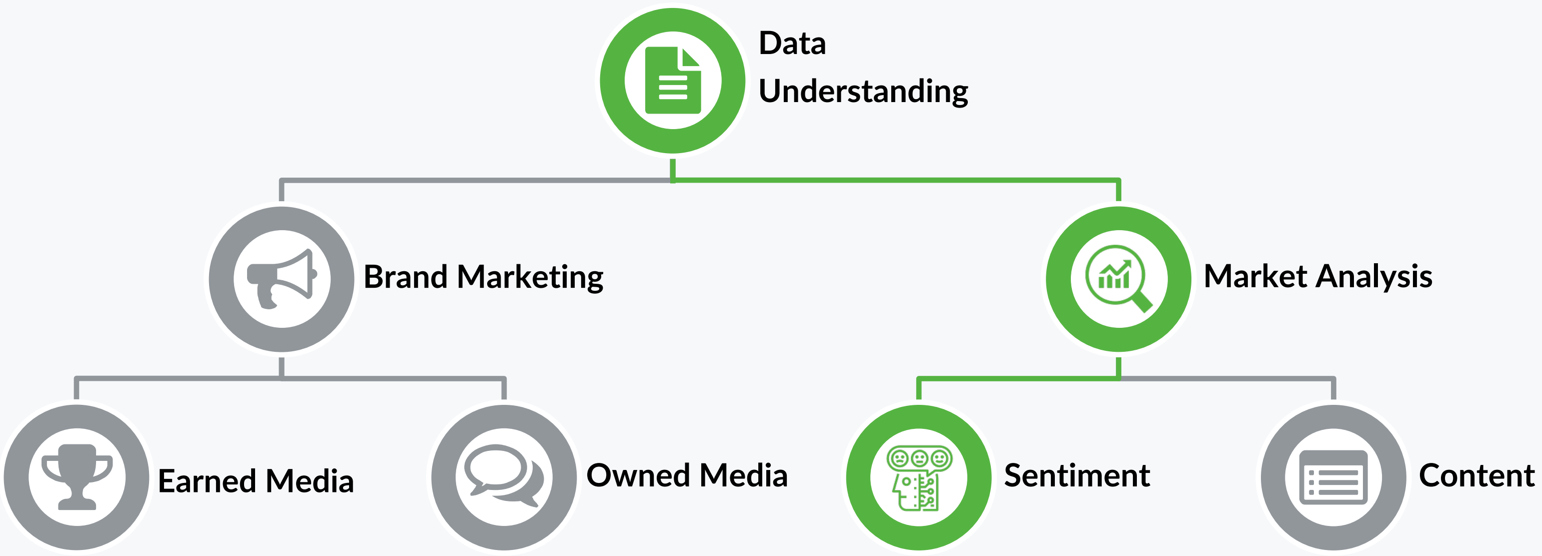
FURTHER RESEARCH

This outcome is **against our expectations** so further research is advisable

VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



LEARN ABOUT THE EMOTIONS AROUND YOUR BRAND TO HAVE THE ADVANTAGE

Using **Vader** and **TextBlob** packages to perform sentiment analysis

01

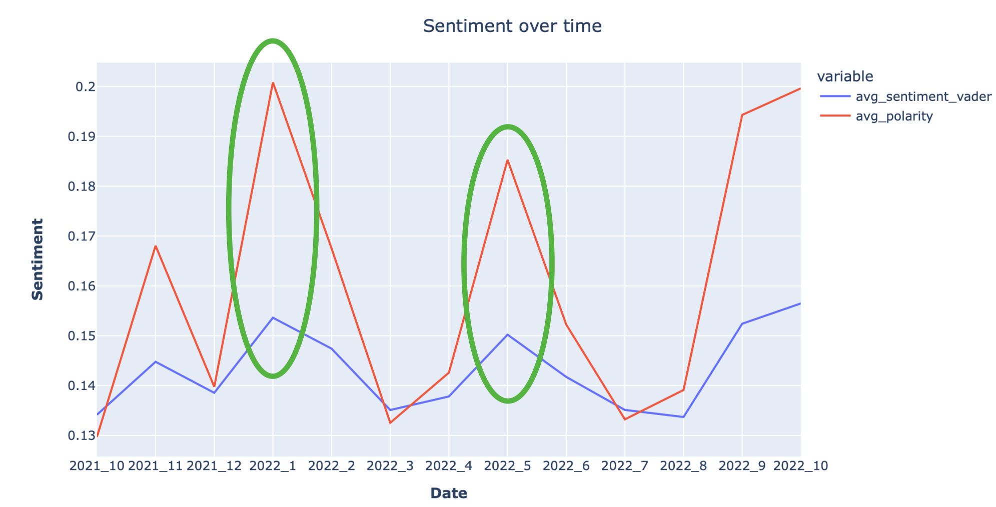
Vader Package

Lexicon and **rule-based sentiment** analysis tool that is specifically attuned to **sentiments** expressed in **social media**

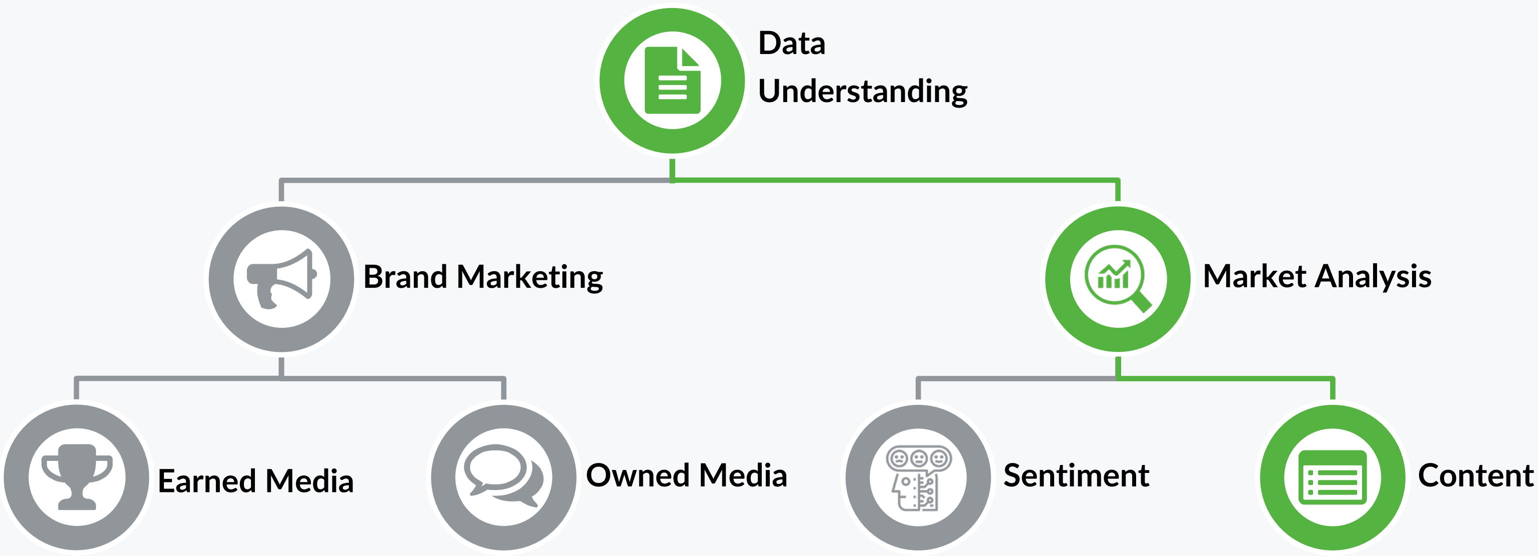
02

TextBlob Package

Using **pre-defined dictionary** and **averaging** over all words in text



VALUE TREE THAT SHOWS THE VALUE IN DIFFERENT LAYERS



USE SPARK NLP TO PERFORM TOPIC MODELING AND LEARN ABOUT THE TOPICS

Showing the **different steps** in the topic **modeling** Spark NLP process

Basic **NLP Processing** Steps

N-Gram Generator

PerceptronModel for **POS-tagging**

Filtering useless combinations of POS

TF-IDF matrix

Latent Dirichlet Allocation

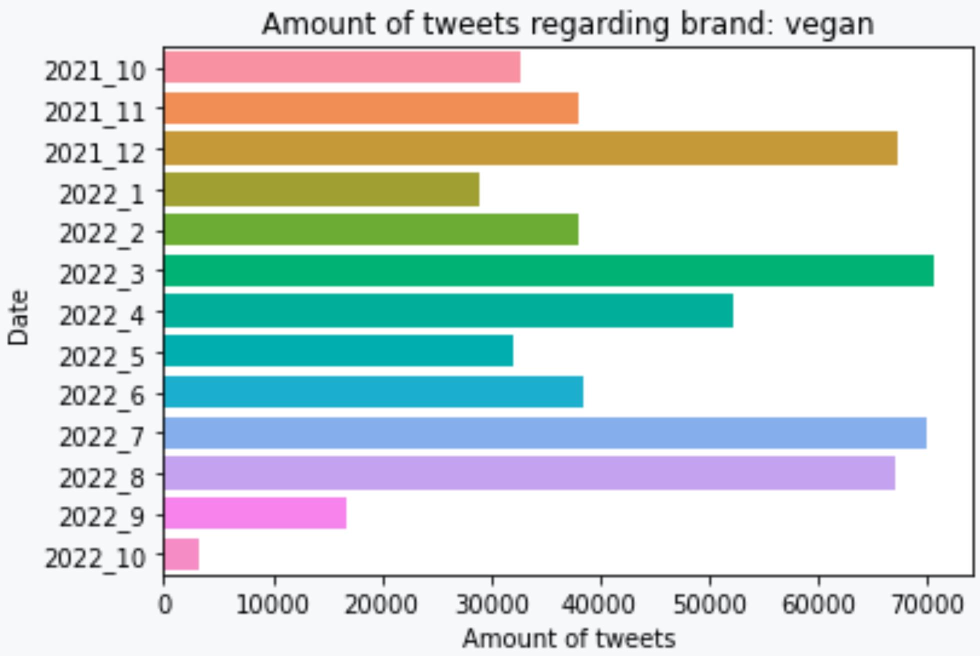
Get necessary **information**

CREATE A WORD CLOUD TO HAVE AN OVERALL IDEA ABOUT THE TOPICS

TIME FOR Q&A!

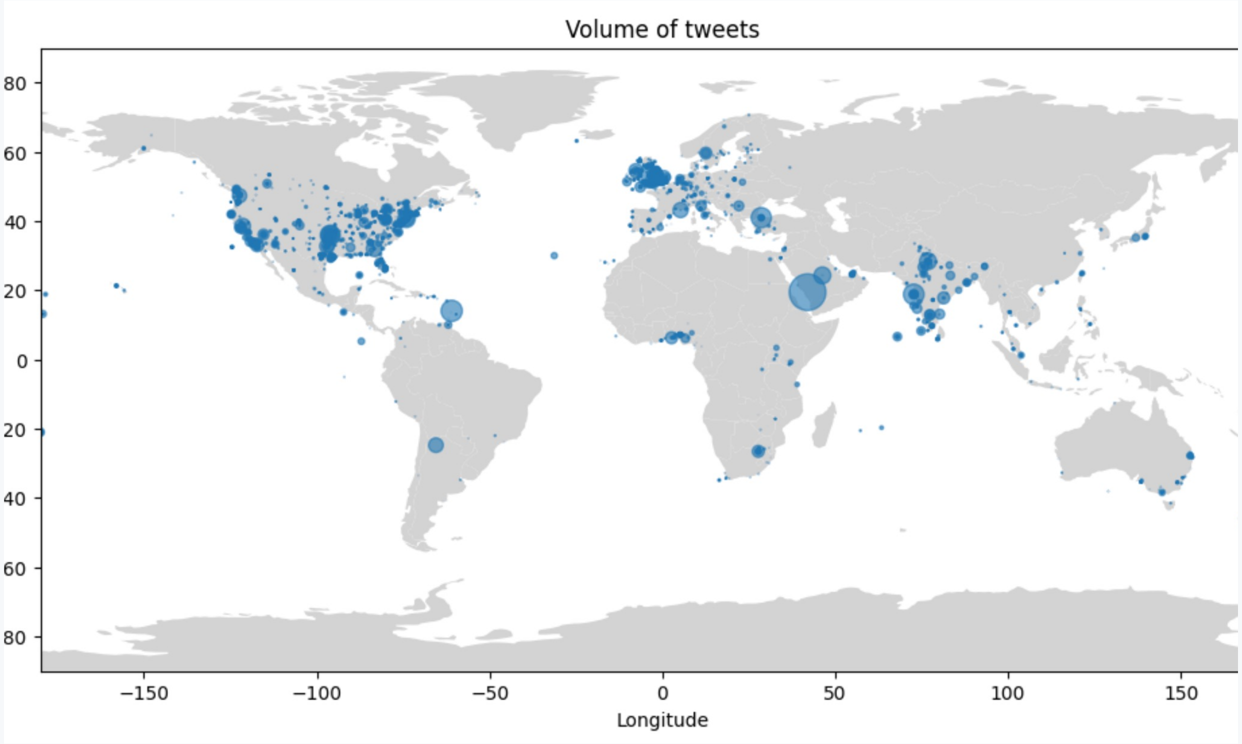
APPENDIX

VOLUME



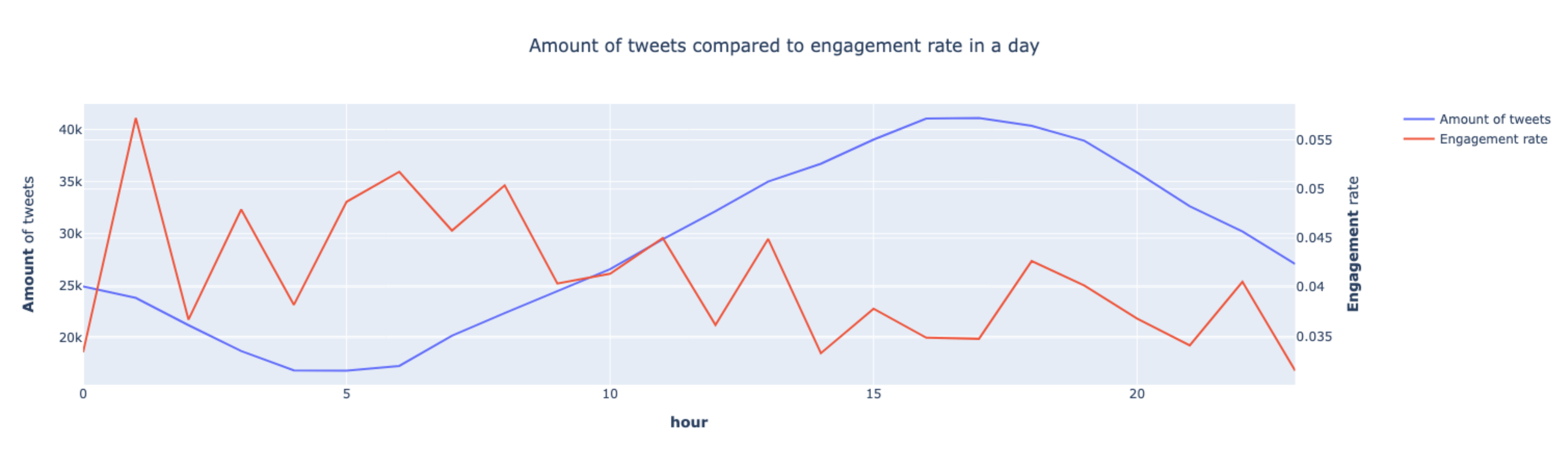
Amount of tweets regarding brand: vegan
Date
2021_10
2021_11
2021_12
2022_1
2022_2
2022_3
2022_4
2022_5
2022_6
2022_7
2022_8
2022_9
2022_10
0
10000
20000
30000
40000
50000
60000
70000
Amount of tweets

LOCATION



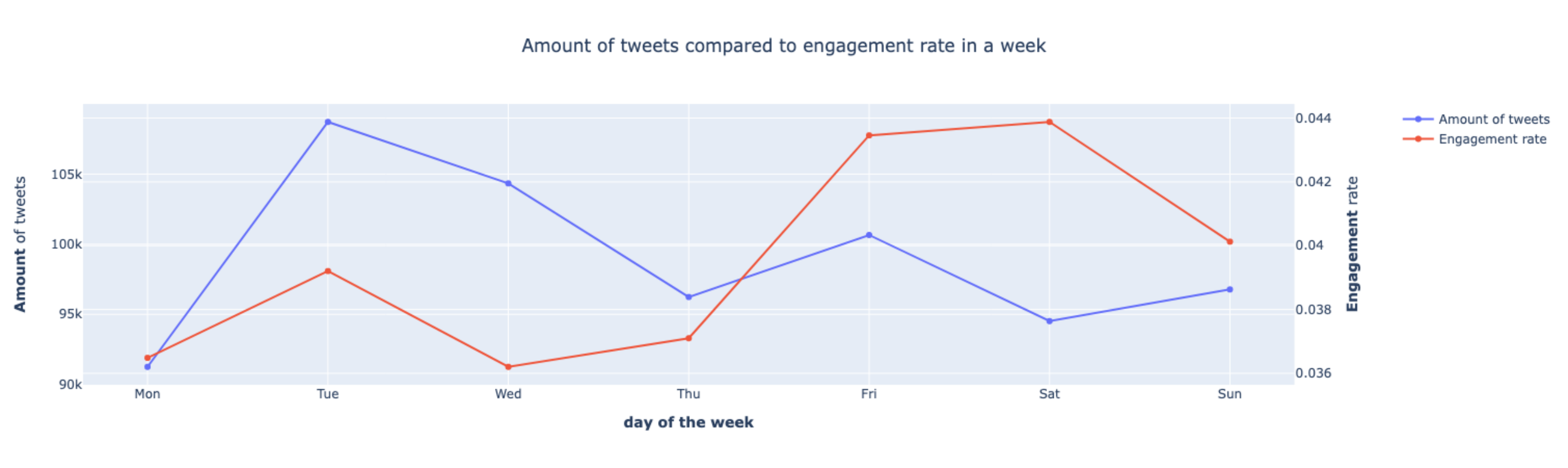
Volume of tweets
80
60
40
20
0
20
40
60
80
−150
−100
−50
0
50
100
150
Longitude

TIMING

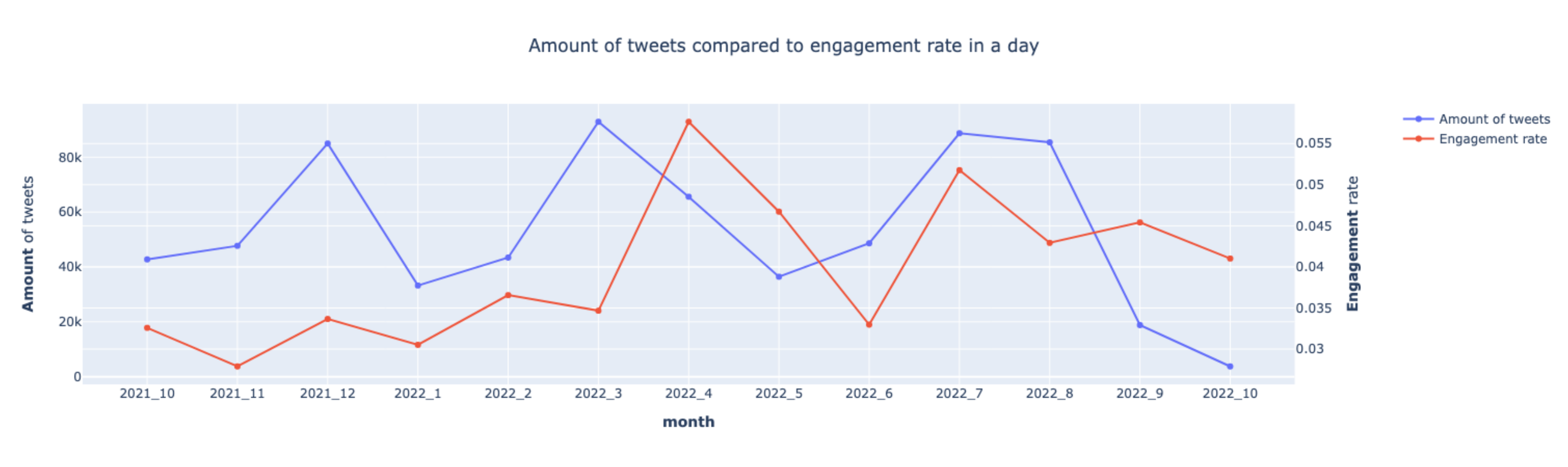


Amount of tweets compared to engagement rate in a day
40k
35k
30k
25k
20k
Amount of tweets
0.055
0.05
0.045
0.04
0.035
Engagement rate
0
5
10
15
20
hour
Amount of tweets
Engagement rate

TIMING

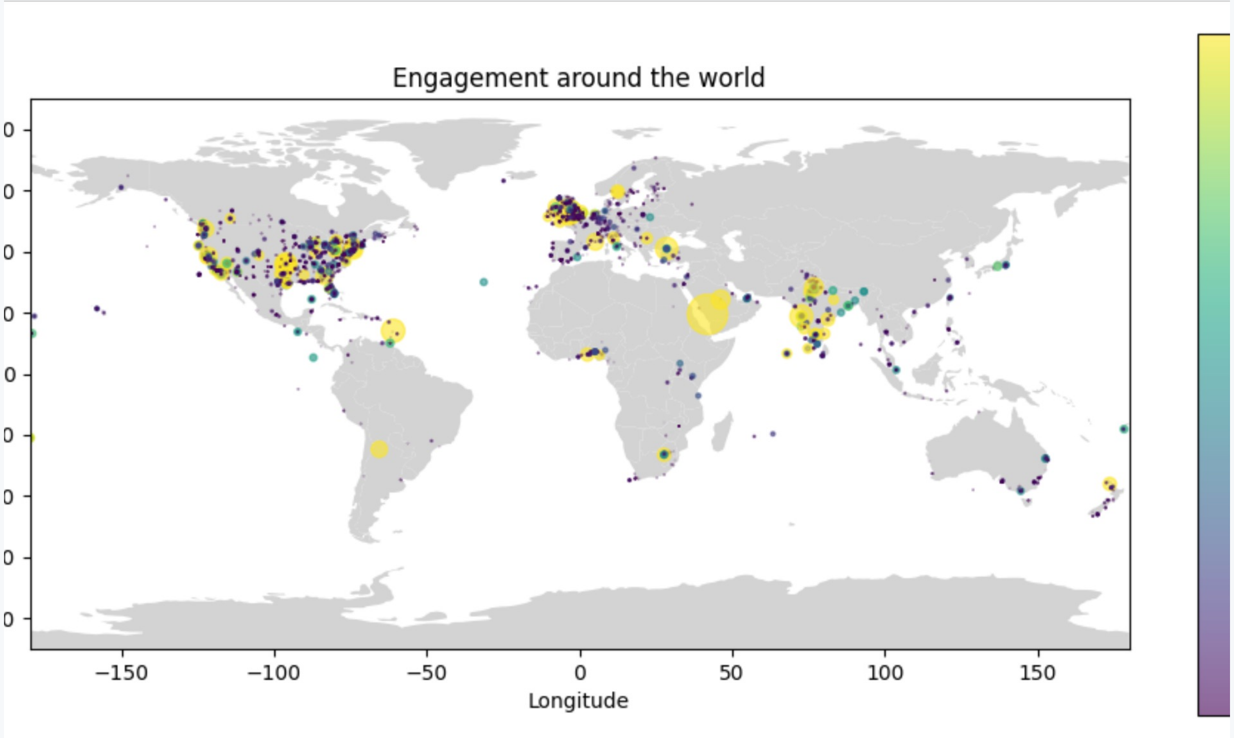

Amount of tweets compared to engagement rate in a week
Amount of tweets
Engagement rate
105k
100k
95k
90k
0.044
0.042
0.04
0.038
0.036
Mon
Tue
Wed
Thu
Fri
Sat
Sun
day of the week

TIMING

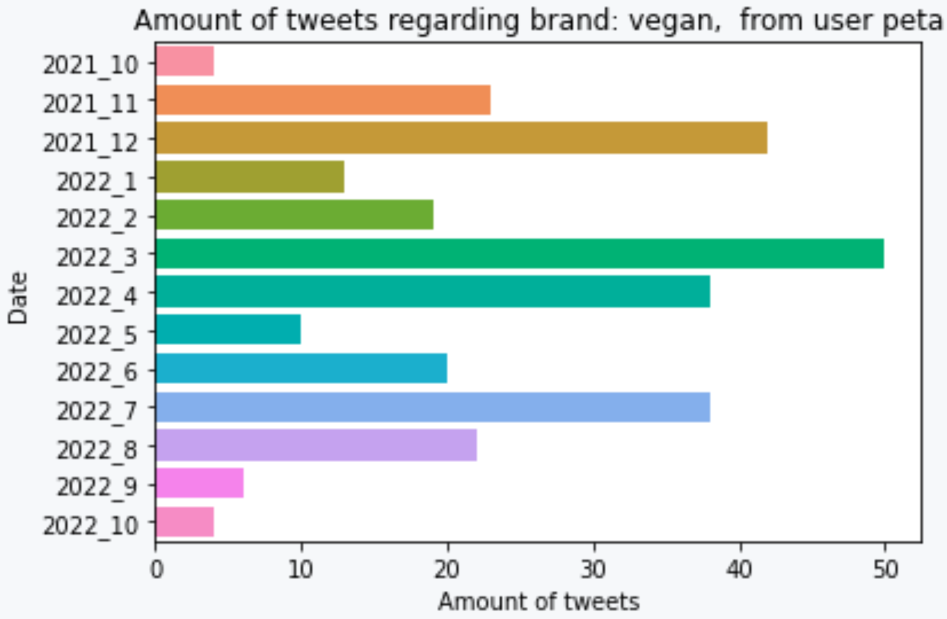


Amount of tweets compared to engagement rate in a day
Amount of tweets
Engagement rate
0
20k
40k
60k
80k
0.03
0.035
0.04
0.045
0.05
0.055
2021_10
2021_11
2021_12
2022_1
2022_2
2022_3
2022_4
2022_5
2022_6
2022_7
2022_8
2022_9
2022_10
month
Amount of tweets
Engagement rate

ENGAGEMENT AROUND THE WORLD



BRAND FOLLOWER'S IMPORTANCE



BRAND FOLLOWER'S IMPORTANCE

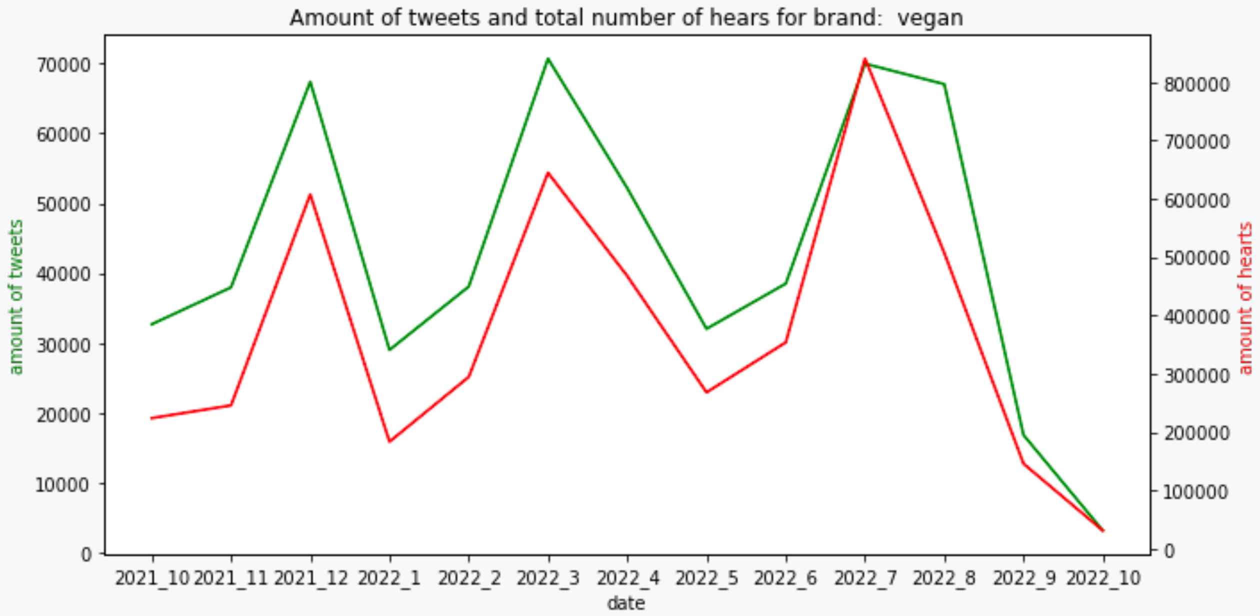
Account with the highest weight

name	screen_name	amount_of_followers	count_of_tweets	weight
PETA	peta	1082775	289	312921975
Lady Gaga	ladygaga	84876971	2	169753942
The New York Times	nytimes	53818201	2	107636402
Reuters	Reuters	24739056	4	98956224
Kompas.com	kompascom	8296037	9	74664333

Account with the highest weight with more than 50 tweets

name	screen_name	amount_of_followers	count_of_tweets	weight
PETA	peta	1082775	289	312921975
Vegan Future	veganfuture	173086	339	58676154
vegnews	VegNews	176977	231	40881687
Vegan Posters	veganposters	9645	3831	36949995
Mystic Mist Healing	TheMysticMist	4629	6622	30653238

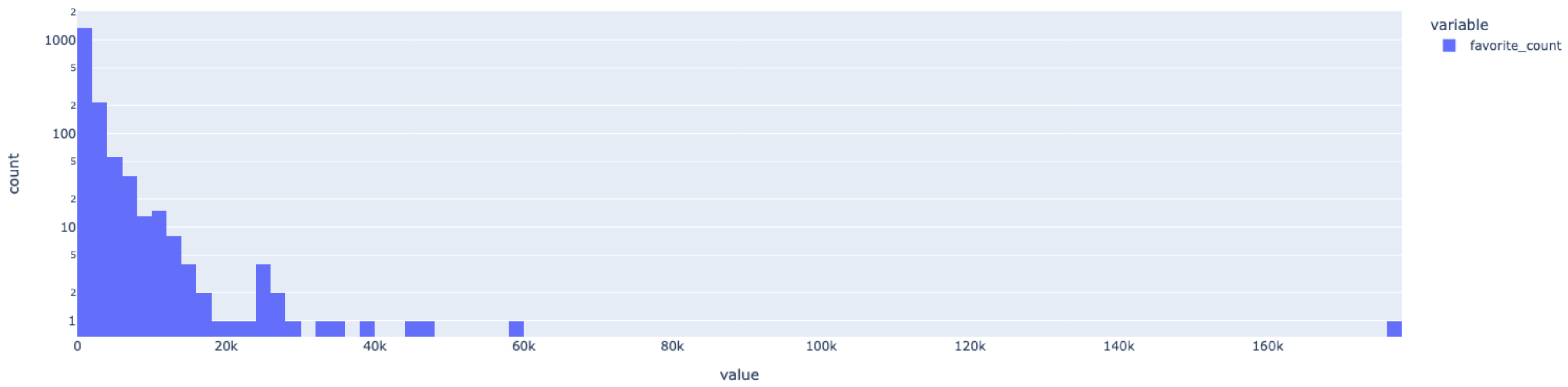
NUMBER OF LIKES



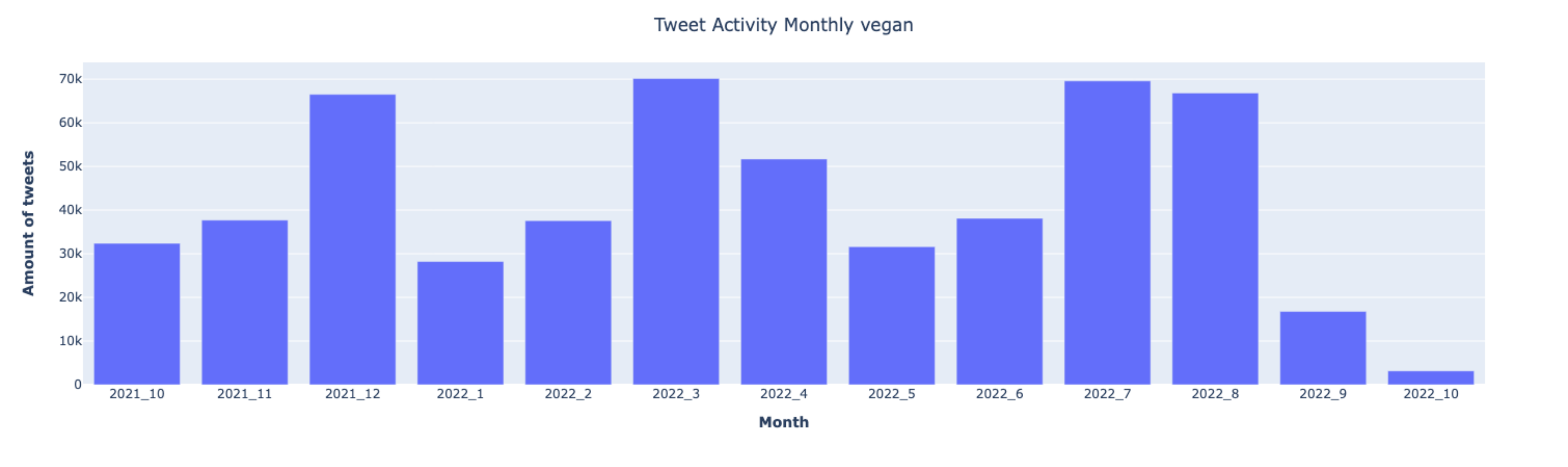
NUMBER OF LIKES

name	MaxFollowersCount
Lady Gaga	84876971
CNN	58738132
The New York Times	53833653
BBC News (World)	37368190
The Economist	26981263
Kourtney Kardashian	26324023
Reuters	25387151
A.R.Rahman	24192397
Nicki Minaj	23311089
Fox News	22083552

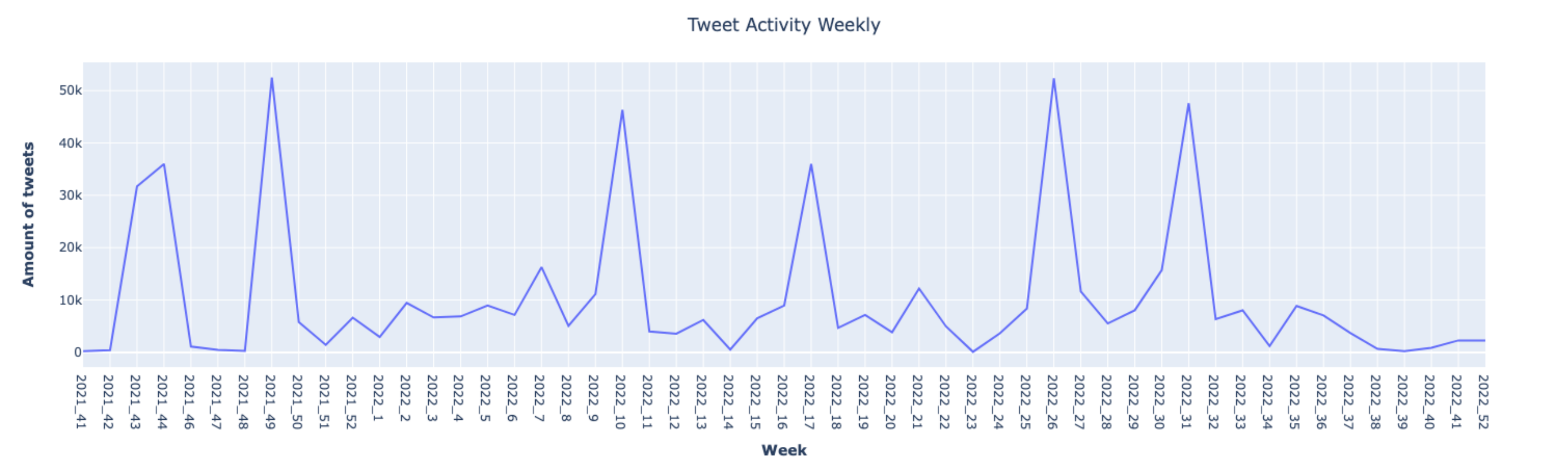
FAVOURITES PER TWEET



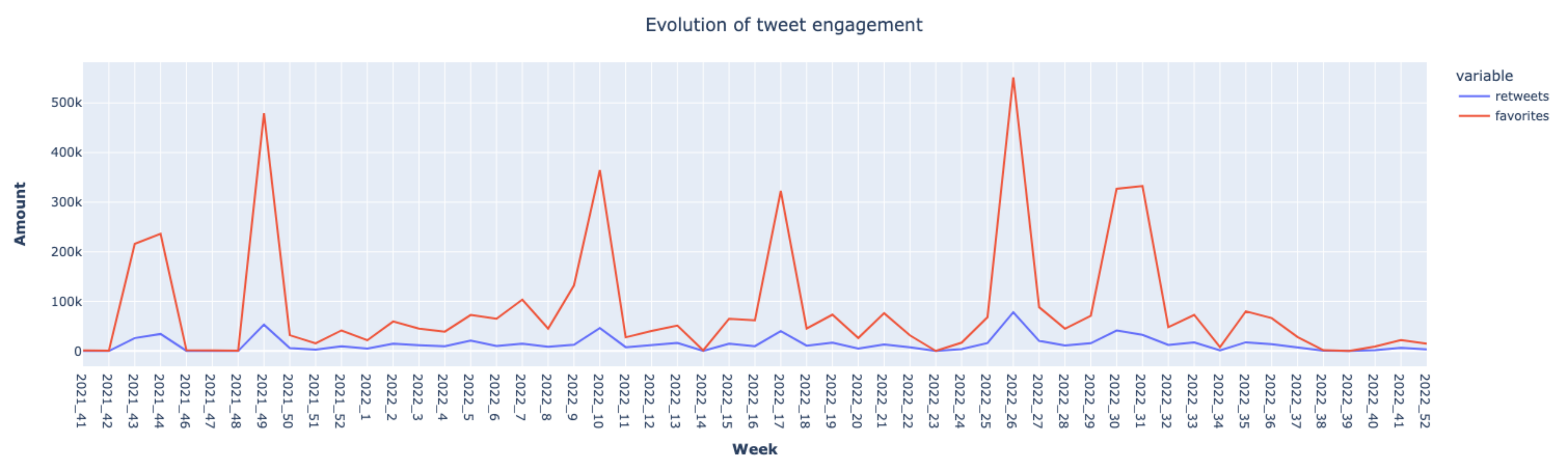
EVOLUTION OF TWEET ACTIVITY MONTHLY



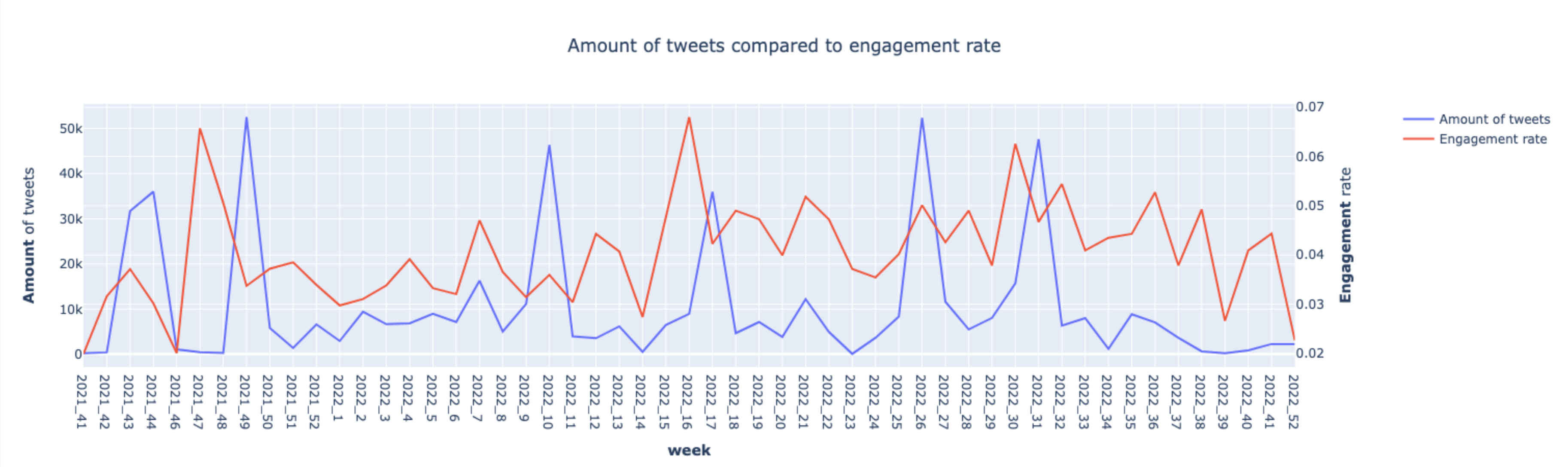
EVOLUTION OF TWEET ACTIVITY WEEKLY



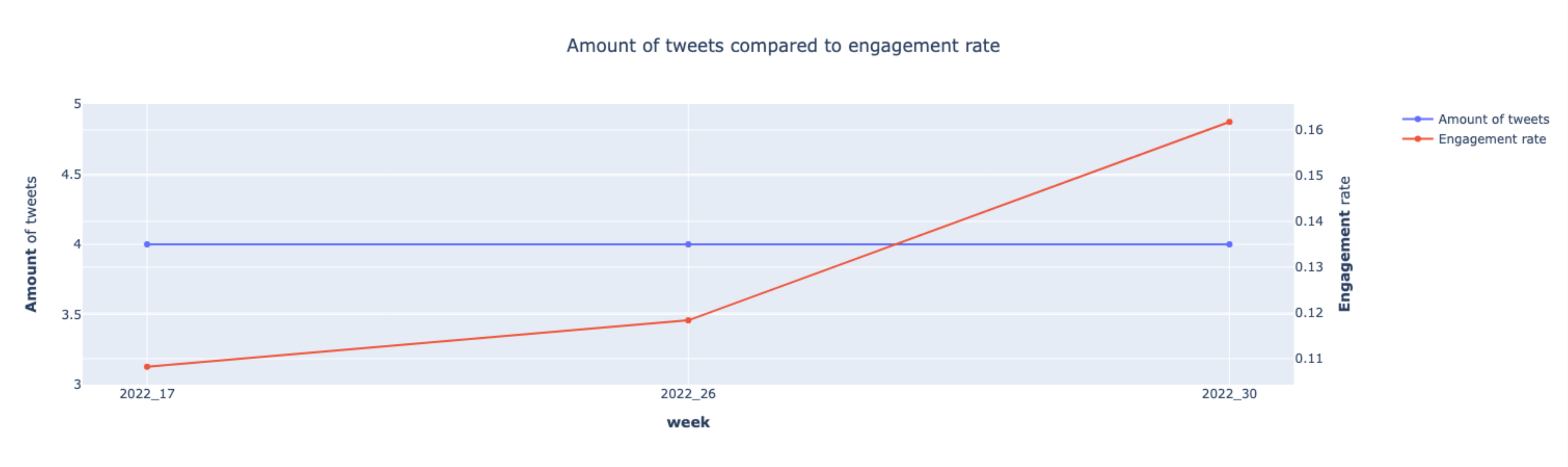
EVOLUTION OF TWEET ENGAGEMENT



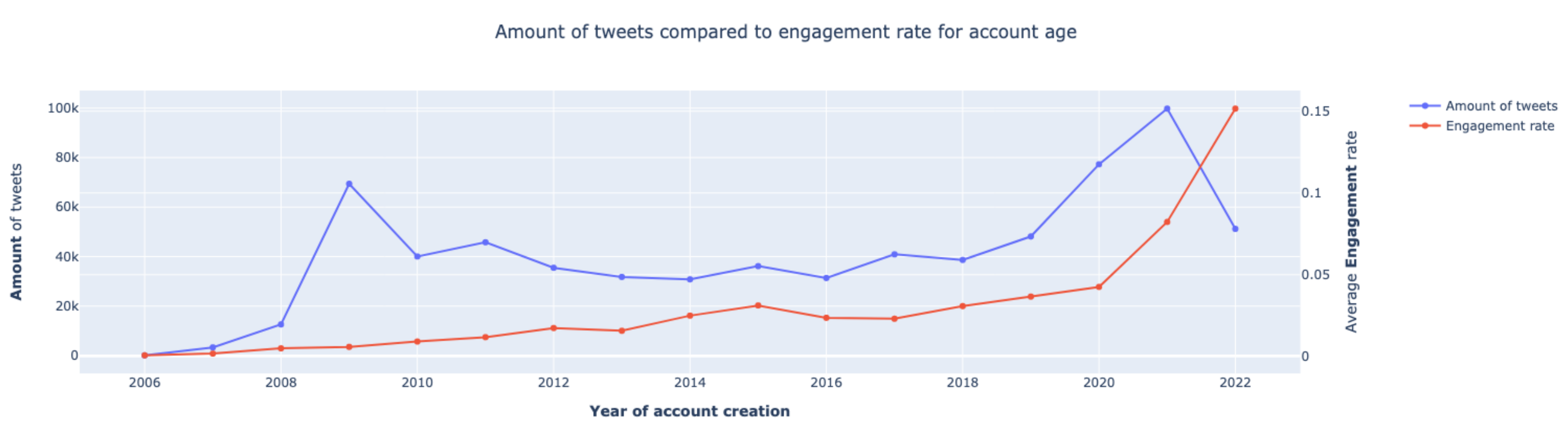
LEVEL OF SOCIAL MEDIA ACTIVITY VS ENGAGEMENT



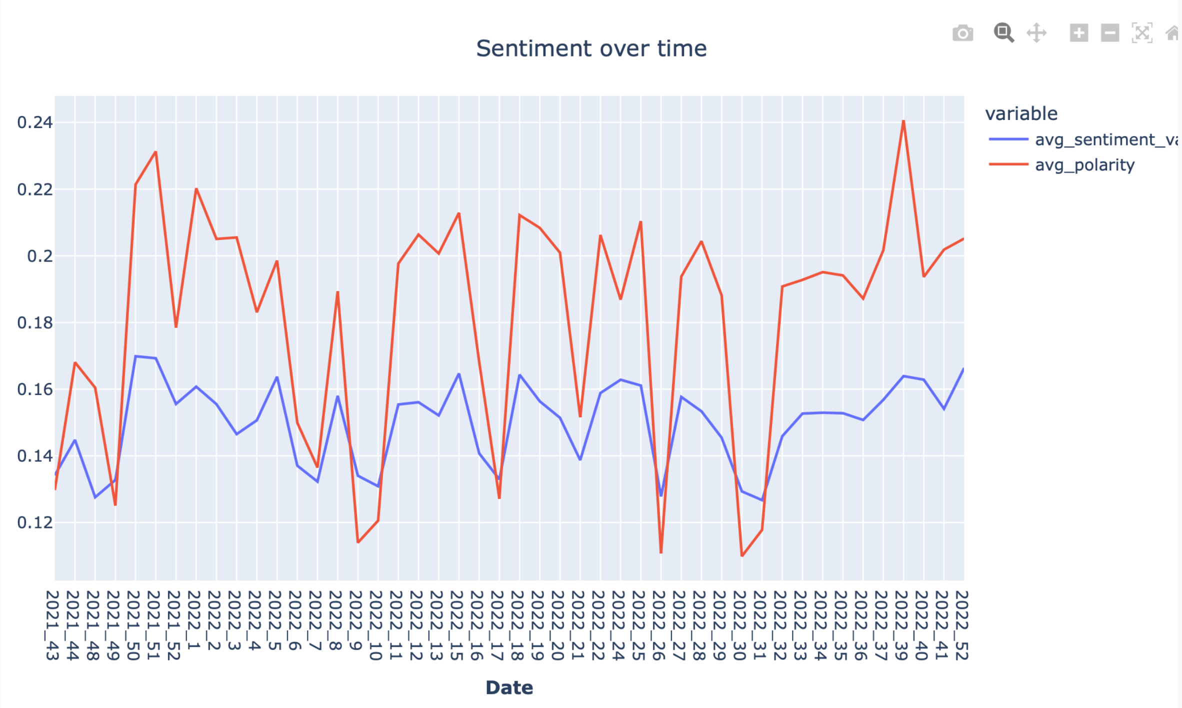
VOLUME OF INFLUENCER ACTIVITY VS LEVEL OF ENGAGEMENT



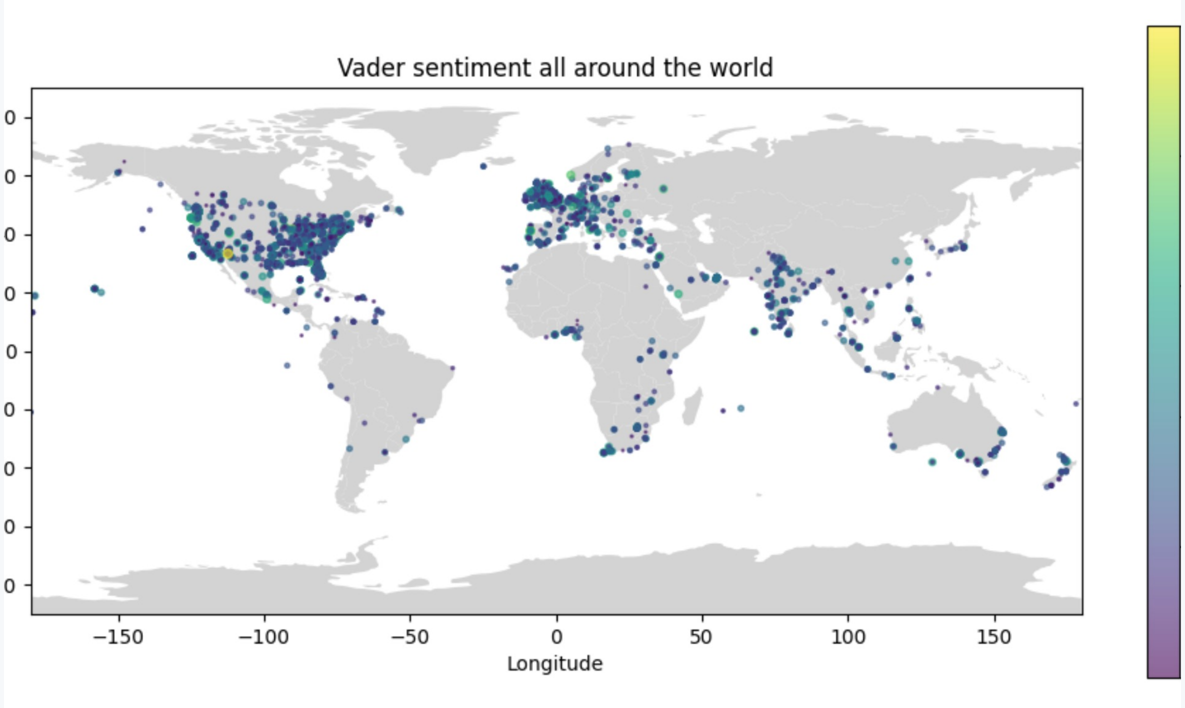
AGE OF ACCOUNTS WITH ENGAGEMENT AND VOLUME



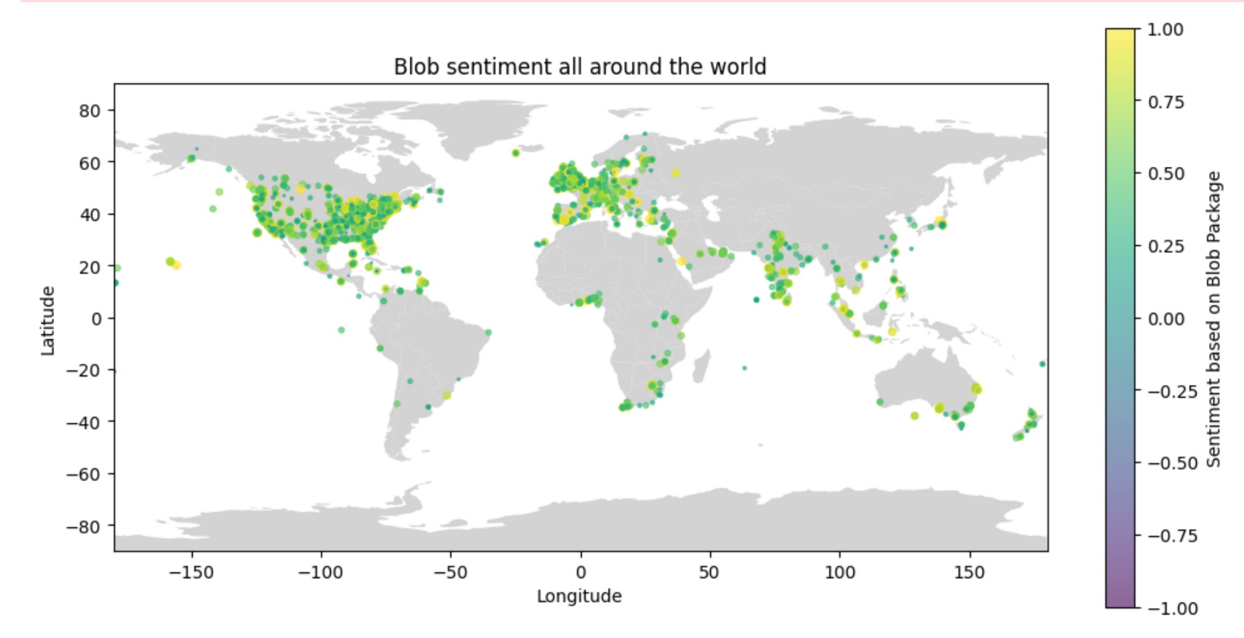
Sentiment per week



Vader sentiment around the world



Blob sentiment around the world



Variable importance Random Forest

Feature	Importance
user_following	0.194689
hour of the day	0.125309
day of the month	0.116302
account age	0.115249
month	0.065632
number of text characters	0.054684
num ber of words	0.052102
subjectivity	0.049946
weekday	0.048056
polarity	0.046693
sentiment	0.045300
number of upper case words	0.024228
number of emojis	0.023503
media type index	0.015749
number of media elements	0.013839
quoted tweet	0.005118
verified account	0.001654
symbol indicator	0.000929
number of mentions	0.000500
number of exclamations	0.000473
number of hashtags	0.000043

Coefficients logistic regression:

For the users the account is **following** and **account age**

```
# 1) Look at the number of users the accounts is following
coefs_df.filter(F.col('Feature') == 'num_features_scaled_0').show()
```

```
+--------------------+-------------------+--------------------+--------------------+
|             Feature|  Coefvalue_class_0|   Coefvalue_class_1|   Coefvalue_class_2|
+--------------------+-------------------+--------------------+--------------------+
|num_features_scal...|0.21893131811186922|-0.07389076709476614|-0.14504055101710314|
+--------------------+-------------------+--------------------+--------------------+
```

```
# 3) Age of the account in days
coefs_df.filter(F.col('Feature') == 'num_features_scaled_7').show()
```

```
+--------------------+------------------+--------------------+--------------------+
|             Feature| Coefvalue_class_0|   Coefvalue_class_1|   Coefvalue_class_2|
+--------------------+------------------+--------------------+--------------------+
|num_features_scal...|0.1422967220047017|-0.02227923167135...|-0.12001749033334655|
+--------------------+------------------+--------------------+--------------------+
```

Coefficients logistic regression:

For the number of **text characters** and **the number of words** in the tweet

```
# 6) Look at the text characters of the tweet
coefs_df.filter(F.col('Feature') == 'num_features_scaled_9').show()
```

```
+--------------------+--------------------+--------------------+--------------------+
|             Feature|   Coefvalue_class_0|   Coefvalue_class_1|   Coefvalue_class_2|
+--------------------+--------------------+--------------------+--------------------+
|num_features_scal...|-0.00644509745984...|-0.00938912471622577|0.015834222176070754|
+--------------------+--------------------+--------------------+--------------------+
```

```
# 7) Look at the number of words of the tweet
coefs_df.filter(F.col('Feature') == 'num_features_scaled_3').show(truncate = False)
```

```
+---------------------+--------------------+---------------------+--------------------+
|Feature              |Coefvalue_class_0   |Coefvalue_class_1    |Coefvalue_class_2   |
+---------------------+--------------------+---------------------+--------------------+
|num_features_scaled_3|-0.00997985303923987|0.0012823808209239502|0.008697472218315938|
+---------------------+--------------------+---------------------+--------------------+
```

Coefficients logistic regression:

For the **subjectivity**, **polarity** and **sentiment** of the tweet

```
# 8 + 9) Look at the subjectivity, polarity and sentiment of the tweet
coefs_df.filter((F.col('Feature') == 'subjectivity')|
                (F.col('Feature') == 'polarity')|
                (F.col('Feature') == 'sentiment')).show(truncate = False)
```

```
+------------+--------------------+-------------------+--------------------+
|Feature     |Coefvalue_class_0   |Coefvalue_class_1  |Coefvalue_class_2   |
+------------+--------------------+-------------------+--------------------+
|sentiment   |-0.05737200701348562 |0.01000923729822799|0.0473627697152576 5|
|polarity    |-0.007370786667381791|-0.0170725622750061|0.024443348942387838|
|subjectivity|-0.02215712103128041 |-0.0215743758010151|0.043731496832295706|
+------------+--------------------+-------------------+--------------------+
```

Coefficients logistic regression:

For the number **emojis** and **the number of upper case words** in the tweet

```
# 11 + 12) Look at extra characters: number of emojis and number of upper case words
coefs_df.filter((F.col('Feature') == 'num_features_scaled_1')|
                (F.col('Feature') == 'num_features_scaled_2')).show(truncate = False)
```

```
+---------------------+----------------------+---------------------+---------------------+
|Feature              |Coefvalue_class_0     |Coefvalue_class_1    |Coefvalue_class_2    |
+---------------------+----------------------+---------------------+---------------------+
|num_features_scaled_1|-0.0030855051461838724|0.008212959466056984|-0.005127454319873103|
|num_features_scaled_2|-0.0024891964088818714|6.398308953780299E-4|0.001849365513503825 |
+---------------------+----------------------+---------------------+---------------------+
```

Coefficients logistic regression:

For the **hour** of the day

Feature	Coefvalue_class_0	Coefvalue_class_1	Coefvalue_class_2
cat_features_hour_index_OHE_1	0.007768756020498764	0.05309877031656413	-0.06086752633706288
cat_features_hour_index_OHE_10	0.028058668600232733	0.020143892489595513	-0.04820256108982826
cat_features_hour_index_OHE_12	-0.007419555674138468	0.053780576862951394	-0.04636102118881294
cat_features_hour_index_OHE_0	-0.0216374315914389	0.054329718299690036	-0.03269228670825114
cat_features_hour_index_OHE_17	-0.009045650739182473	0.03734296658004055	-0.028297315840858057
cat_features_hour_index_OHE_3	5.06836835806825E-4	0.023520230478031652	-0.02402706731383848
cat_features_hour_index_OHE_2	0.04759111549946655	-0.02828521840595907	-0.01930589709350748
cat_features_hour_index_OHE_13	0.004627108100162384	0.00965453107742652	-0.014281639177588899
cat_features_hour_index_OHE_20	-0.009789667909068133	0.023705827998245756	-0.013916160089177608
cat_features_hour_index_OHE_21	0.0038846739245690094	0.0032827424337155764	-0.007167416358284537
cat_features_hour_index_OHE_16	-0.02470949625550453	0.030957223407659062	-0.006247727152154522
cat_features_hour_index_OHE_9	0.020411954678321707	-0.02134574348760665	9.337888092849805E-4
cat_features_hour_index_OHE_22	0.006199848253685338	-0.013398022556215238	0.007198174302529899
cat_features_hour_index_OHE_4	-0.026938348306833778	0.017870254429369493	0.009068093877464318
cat_features_hour_index_OHE_23	0.009072946141219482	-0.020433415046513786	0.011360468905294322
cat_features_hour_index_OHE_19	0.00786142568349386	-0.02263508665826263	0.014773660974768746
cat_features_hour_index_OHE_5	0.010759507854698763	-0.02860075225935315	0.017841244404654386
cat_features_hour_index_OHE_18	0.003858814451203442	-0.02199034619716972	0.018131531745966278
cat_features_hour_index_OHE_11	0.013759032139171339	-0.032551362140502214	0.01879233000133092
cat_features_hour_index_OHE_8	-0.003240295953822237	-0.028783364754317094	0.032023660708139334
cat_features_hour_index_OHE_14	0.017103250756872967	-0.05005516091700063	0.032951910160127676
cat_features_hour_index_OHE_15	0.0038284700656728734	-0.04253996489718813	0.03871149483151527
cat_features_hour_index_OHE_7	-0.028786001317671693	-0.0176545815381141	0.04644058285578583

Coefficients logistic regression:

For the **day of the month**

Feature	Coefvalue_class_0	Coefvalue_class_1	Coefvalue_class_2
cat_features_day_month_index_OHE_6	0.036024875871627456	0.05661534155052849	-0.09264021742215592
cat_features_day_month_index_OHE_4	-0.044698943898271254	0.11045424772999039	-0.06575530383171913
cat_features_day_month_index_OHE_9	-0.01236441337162952	0.06771125078322807	-0.055346837411598515
cat_features_day_month_index_OHE_5	0.0760187260578621	-0.031800651847148406	-0.044218074210713644
cat_features_day_month_index_OHE_14	0.020894747755976224	0.022077200760737863	-0.04297194851671406
cat_features_day_month_index_OHE_10	-0.020873748996749913	0.05336955777755777	-0.0324958087808079
cat_features_day_month_index_OHE_8	0.014988753759103881	0.00954814347963806	-0.02453689723874197
cat_features_day_month_index_OHE_30	-0.015191053906148406	0.0359674836598773	-0.020776429753728867
cat_features_day_month_index_OHE_2	0.010026178333876439	0.010354790220189914	-0.020380968554066334
cat_features_day_month_index_OHE_27	-0.0034050049957552644	0.013624323189912623	-0.010219318194157343
cat_features_day_month_index_OHE_11	0.00617665064391543	0.003112425702156344	-0.009289076346071762
cat_features_day_month_index_OHE_25	-0.02264304154453164	0.03022794329567801	-0.00758490175114631
cat_features_day_month_index_OHE_19	-0.005491867184441076	0.012363793082463062	-0.006871925898022
cat_features_day_month_index_OHE_31	0.013623606752705419	-0.013320669705766841	-3.0293704693857773E-4
cat_features_day_month_index_OHE_3	-0.0010449587077753417	-0.0034958293535440345	0.004540788061319361
cat_features_day_month_index_OHE_28	0.001707324790631941	-0.007082892163440611	0.005375567372808645
cat_features_day_month_index_OHE_29	0.008360406217704524	-0.02147715389547158	0.01311674767776708
cat_features_day_month_index_OHE_20	0.015167325771216943	-0.029775529090042954	0.014608203318826046
cat_features_day_month_index_OHE_22	0.011764580506433387	-0.0377917196321009	0.026027139125667523
cat_features_day_month_index_OHE_7	-0.053453081425344946	0.02414337329321806	0.02930970813212685
cat_features_day_month_index_OHE_24	0.01942654259005904	-0.04976387442376958	0.03033733183371053
cat_features_day_month_index_OHE_16	0.06279931623820924	-0.09552427511164592	0.032724958873436676
cat_features_day_month_index_OHE_12	0.01138945820205512	-0.050037523685671784	0.038648065483616656
cat_features_day_month_index_OHE_26	-0.05149314377416043	0.010859459266100314	0.04063368450806013
cat_features_day_month_index_OHE_21	0.00598204022824045	-0.049790763163285456	0.04380872293504502
cat_features_day_month_index_OHE_1	0.006146450655558484	-0.05855713647109144	0.052410685815532966
cat_features_day_month_index_OHE_23	0.004337456335799106	-0.061071298686690444	0.05673384235089134
cat_features_day_month_index_OHE_13	-0.008216816490115563	-0.049038149044345815	0.057254965534461404

Coefficients logistic regression:

For the **month** of the year

Feature	Coefvalue_class_0	Coefvalue_class_1	Coefvalue_class_2
cat_features_month_index_OHE_3	0.0032435586499102967	0.022605645590216494	-0.025849204240126773
cat_features_month_index_OHE_5	0.00669860420495559	0.017137325771302635	-0.023835929976258215
cat_features_month_index_OHE_1	0.0556420912390855l	-0.040886129618752765	-0.014755961620332702
cat_features_month_index_OHE_4	0.0034387008779550323	0.009160783641865532	-0.012599484519820577
cat_features_month_index_OHE_10	1.3467152387964192E-4	0.0015697896518880466	-0.00170446117576768
cat_features_month_index_OHE_12	-0.010846801154359351	0.012024195574231884	-0.0011773944198725252
cat_features_month_index_OHE_6	-0.011170943698131859	8.561980753634834E-4	0.010314745622768406
cat_features_month_index_OHE_11	-0.005118515705167424	-0.009672580201042169	0.014791095906209601
cat_features_month_index_OHE_2	-0.00649771240947244	-0.01300426747945528	0.01950197988892772
cat_features_month_index_OHE_8	0.004234064206301132	-0.024396587609881167	0.020162523403580056
cat_features_month_index_OHE_7	-0.003971900196553123	-0.016892237280445588	0.020864137476998737

Coefficients logistic regression:

For the **day of the week**

Feature	Coefvalue_class_0	Coefvalue_class_1	Coefvalue_class_2
cat_features_week_day_index_OHE_Thu	4.186560202989838E-4	0.0387257898325645	-0.0391444585286348
cat_features_week_day_index_OHE_Sun	0.007387957539618533	0.003065247079718391	-0.010453204619336917
cat_features_week_day_index_OHE_Fri	0.01663920597789385	-0.007710391210414071	-0.0089288147674797 6
cat_features_week_day_index_OHE_Tue	0.0034353399217008536	0.0027737812638256 63	-0.006209121185526465
cat_features_week_day_index_OHE_Sat	-0.004136082947720673	0.0095103246678568 6	-0.0053742417201361 8
cat_features_week_day_index_OHE_Wed	-0.006137857568483364	-0.0203248346839418 5	0.0264626922524252 3