
SOFTWEAVE®
NEW
SECRETLAB TITAN
SOFTWEAVE® BLACK³
NEW
SECRETLAB OMEGA
SOFTWEAVE® BLACK³
SOFTWEAVE®
BLACK³
NOW AVAILABLE IN
OMEGA, TITAN AND TITAN XL
INSPIRED BY STREETWEAR
TRIPLE-BLACK
//HIGH-PERFORMANCE FABRIC
BLACK³
BLACK³
BLACK³
BLACK³
BLACK³
BLACK³

Brand insights for

SECRET LAB

Prakash N
Data Analyst

What we’ll be solving

Gather brand insights from video analytics and comments from Secretlab review videos on Youtube.

Build a classifier that is able to sort incoming comments into key topics that deliver insights for the product team.

01

Data Collection

SEARCH

Can’t use raw results

YouTube API V3

Public API key
(No need for OAUTH creds)

CommentThreads

Able to extract both
comments and replies

Videos

Extract all public
statistics of each video

The DATA

9480 — Unique comments

1642 — Duration of all videos (mins)

184 — Videos relevant to Secretlab

02

IN DEPTH

Time to dive into EDA

Interesting facts at a glance

Days

Most **videos** were published on **Mon** and **Fri**, whereas most **comments** were published on **Tue** and **Wed**

Months

The top 2 dominant months for videos and comments published were **July** and **May** respectively

Years

More than half of the videos and comments published came from **2019** and **2020**

ESSAY

The longest comment had **721** english words in it.

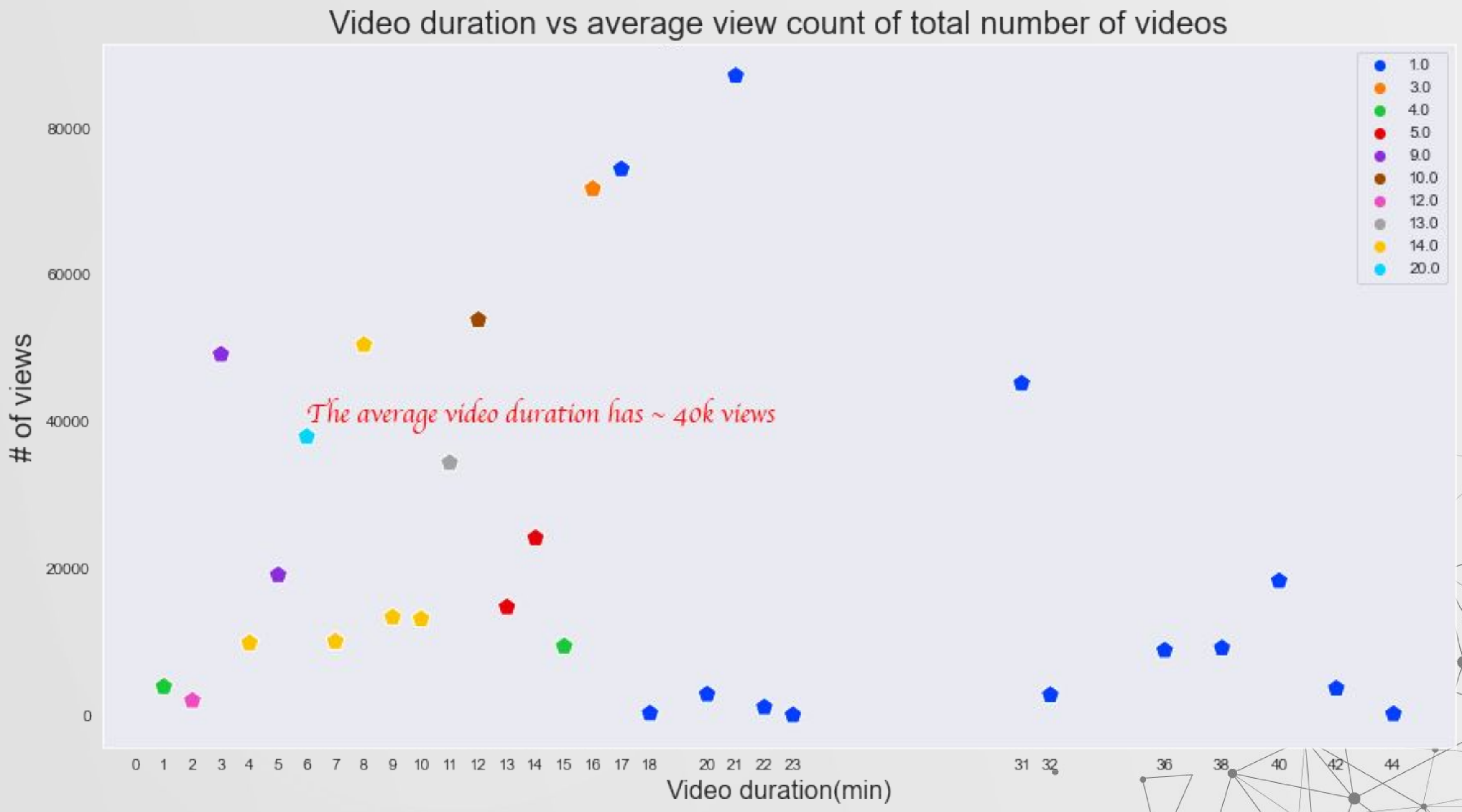
LIKE

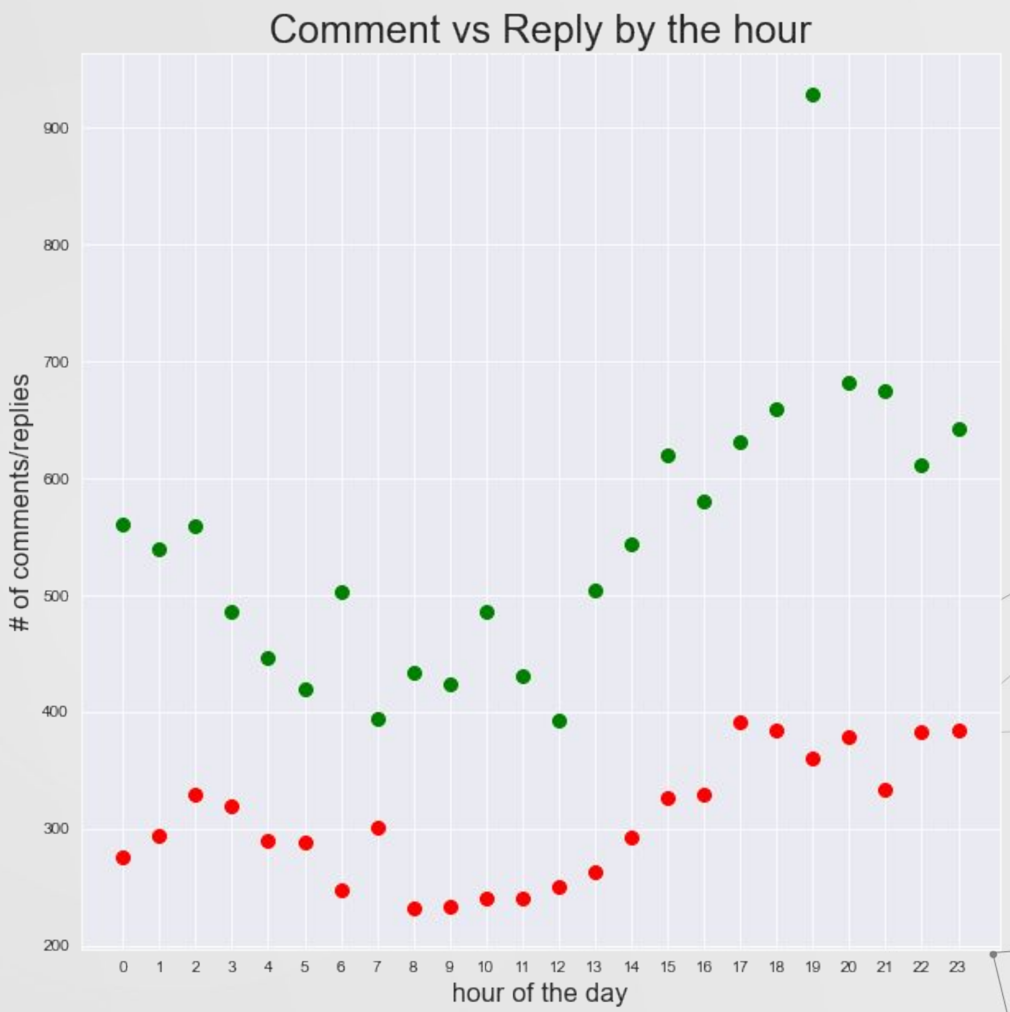
The comment with the most number of likes **(1239):**

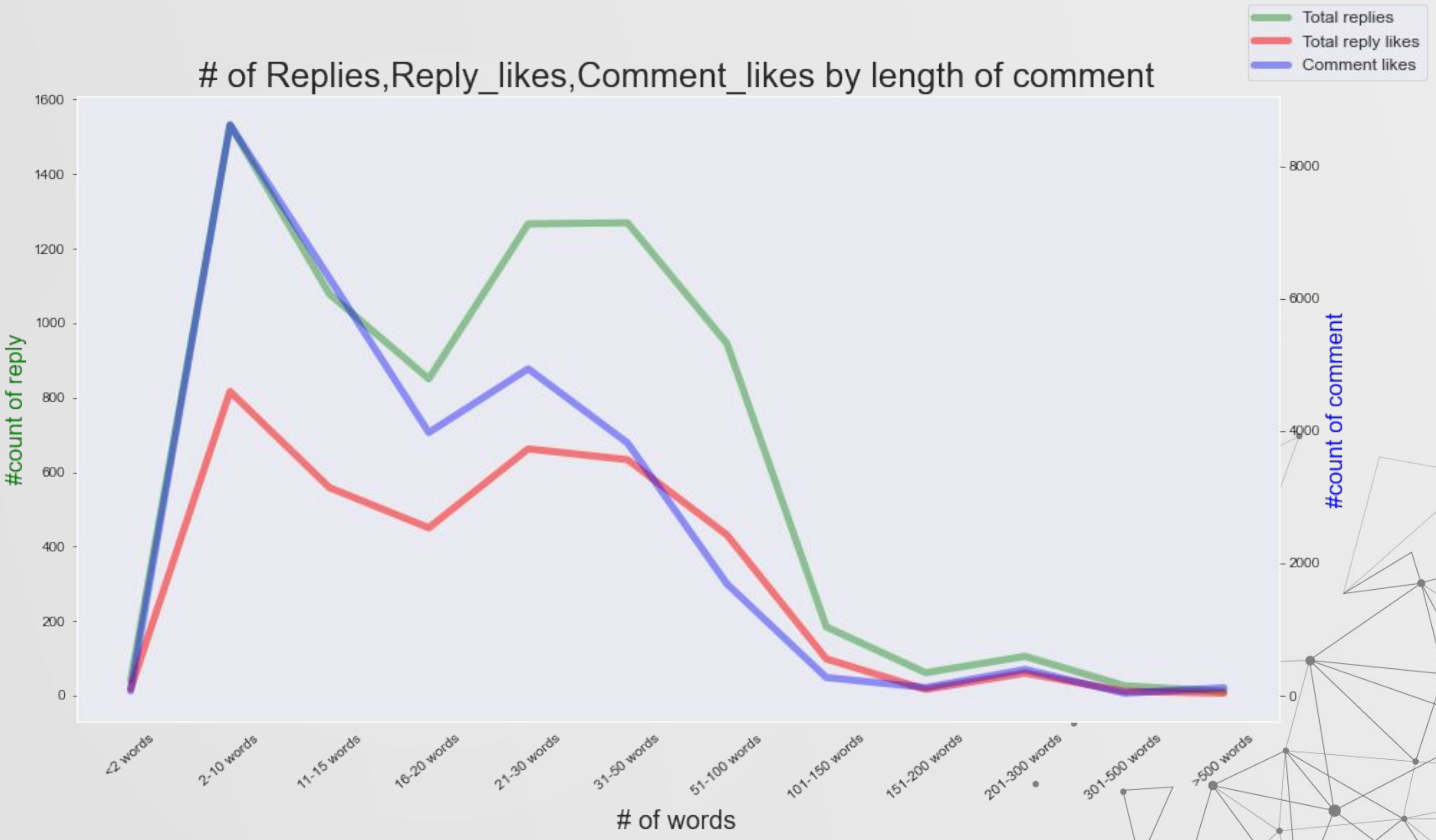
“ one of the best ads i've ever seen “

RESPONSE

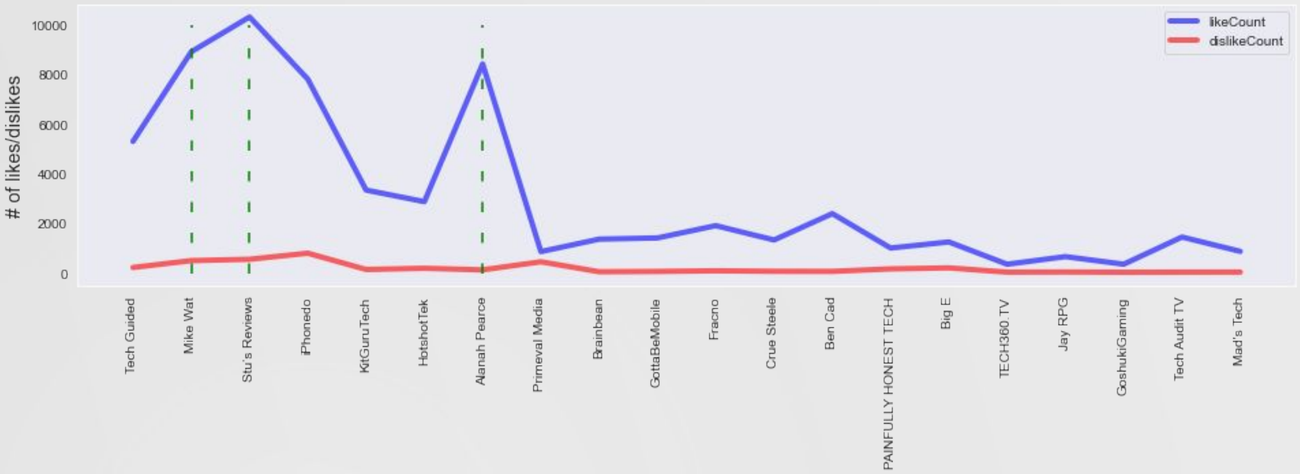
The most frequent reply response to a comment was **1 day**. The latest response to a comment was 1119 days ~ **3 years later.**


Video duration vs average view count of total number of videos
# of views
80000
60000
40000
20000
0
Video duration(min)
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 20 21 22 23 31 32 36 38 40 42 44
1.0
3.0
4.0
5.0
9.0
10.0
12.0
13.0
14.0
20.0
The average video duration has ~ 40k views

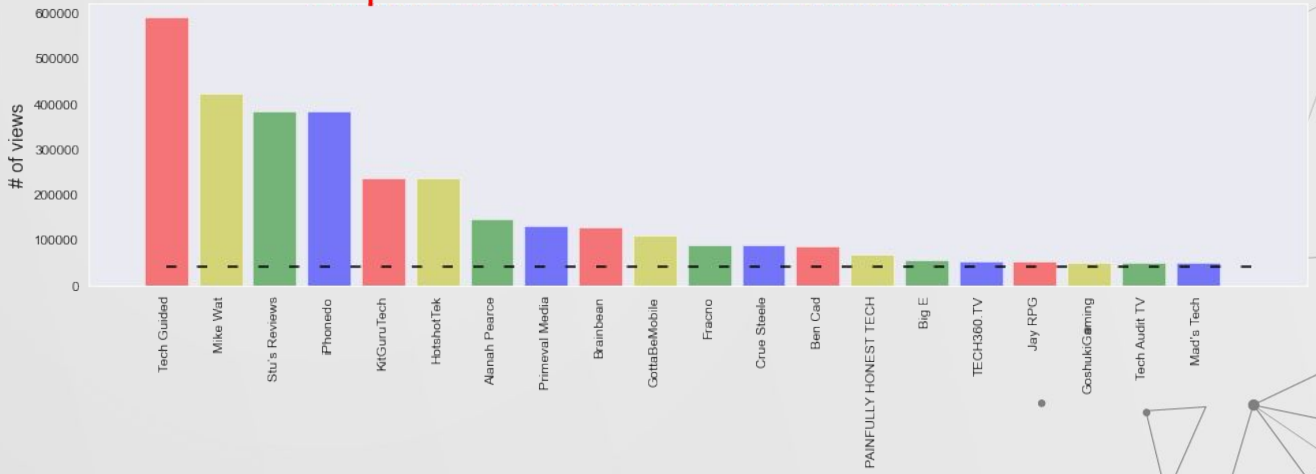

Comment vs Reply by the hour
# of comments/replies
900
800
700
600
500
400
300
200
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
hour of the day

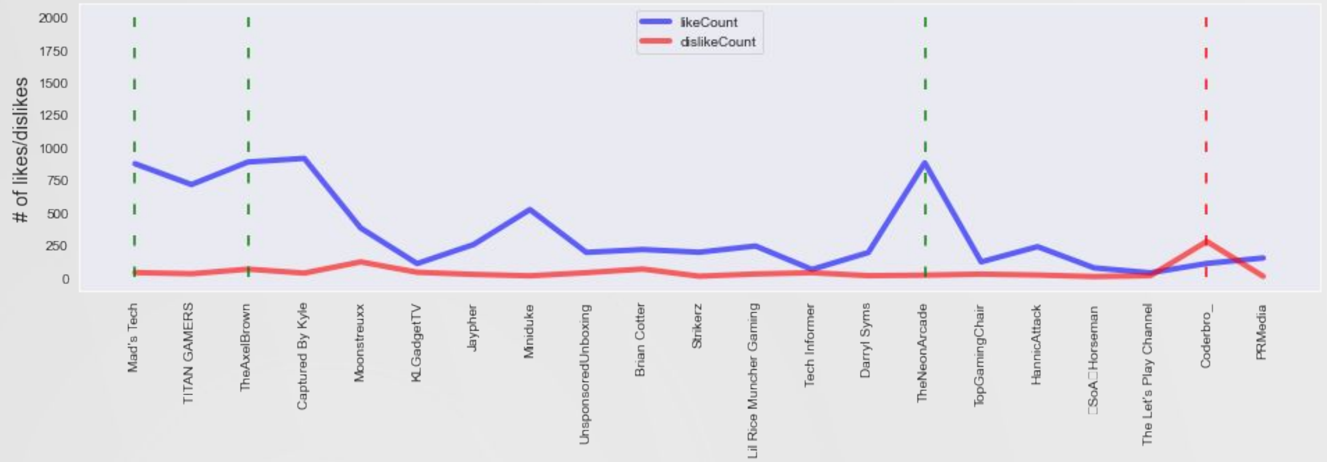


# of Replies,Reply_likes,Comment_likes by length of comment
Total replies
Total reply likes
Comment likes
#count of reply
#count of comment
# of words
0
200
400
600
800
1000
1200
1400
1600
0
2000
4000
6000
8000
<2 words
2-10 words
11-15 words
16-20 words
21-30 words
31-50 words
51-100 words
101-150 words
151-200 words
201-300 words
301-500 words
>500 words

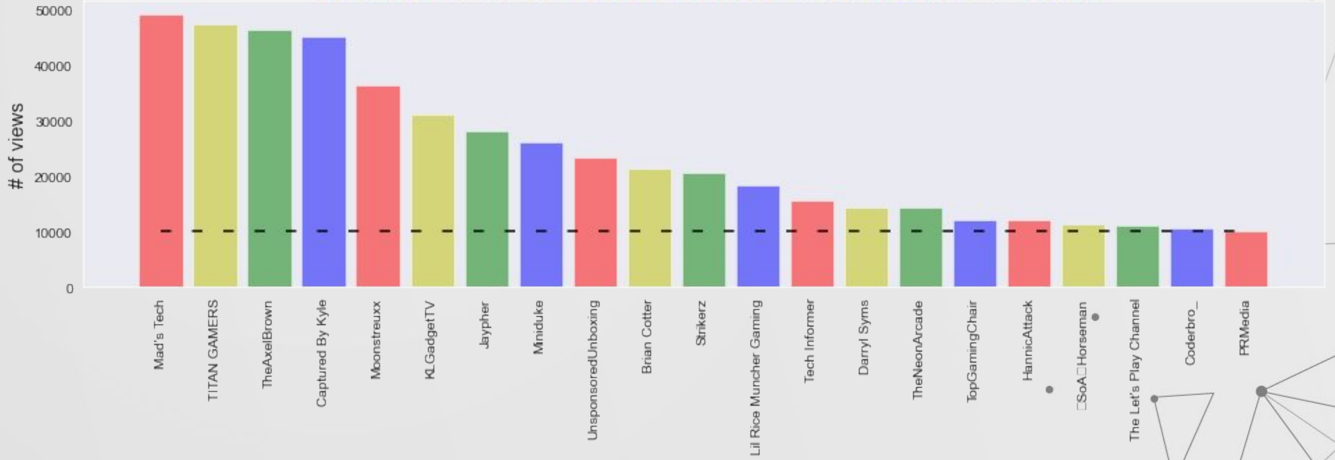


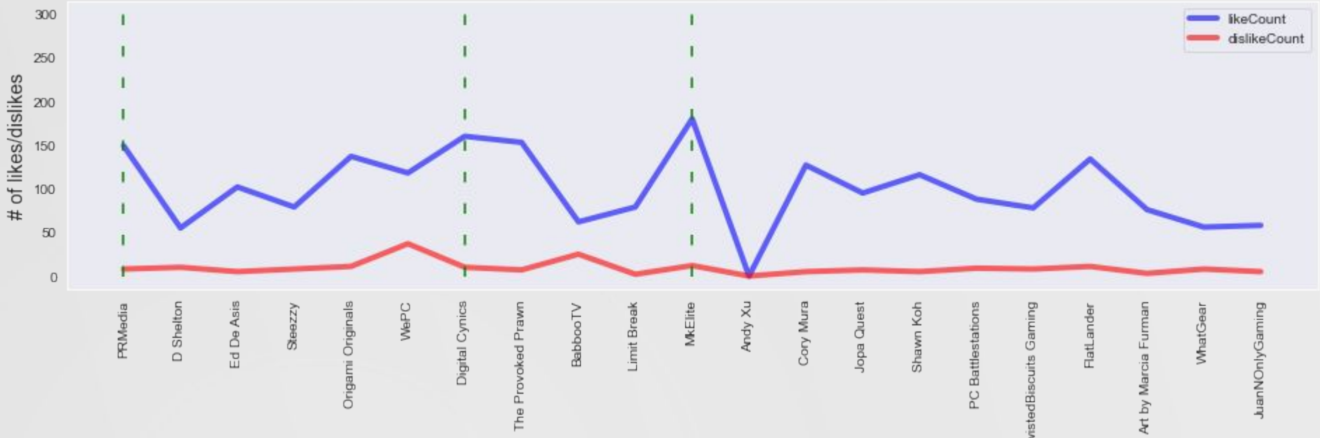
Super influencers with views abv 40k



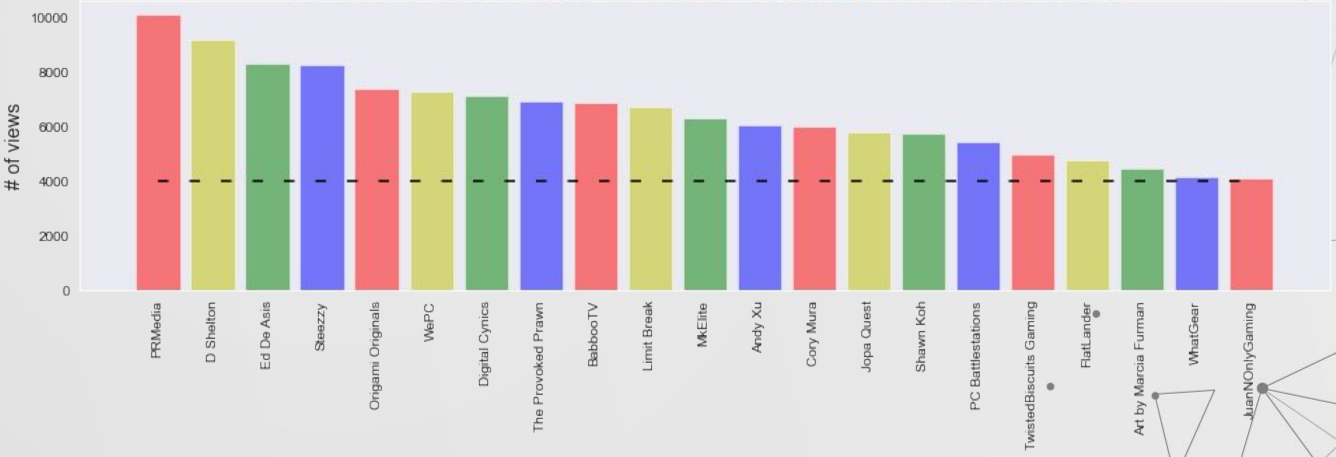


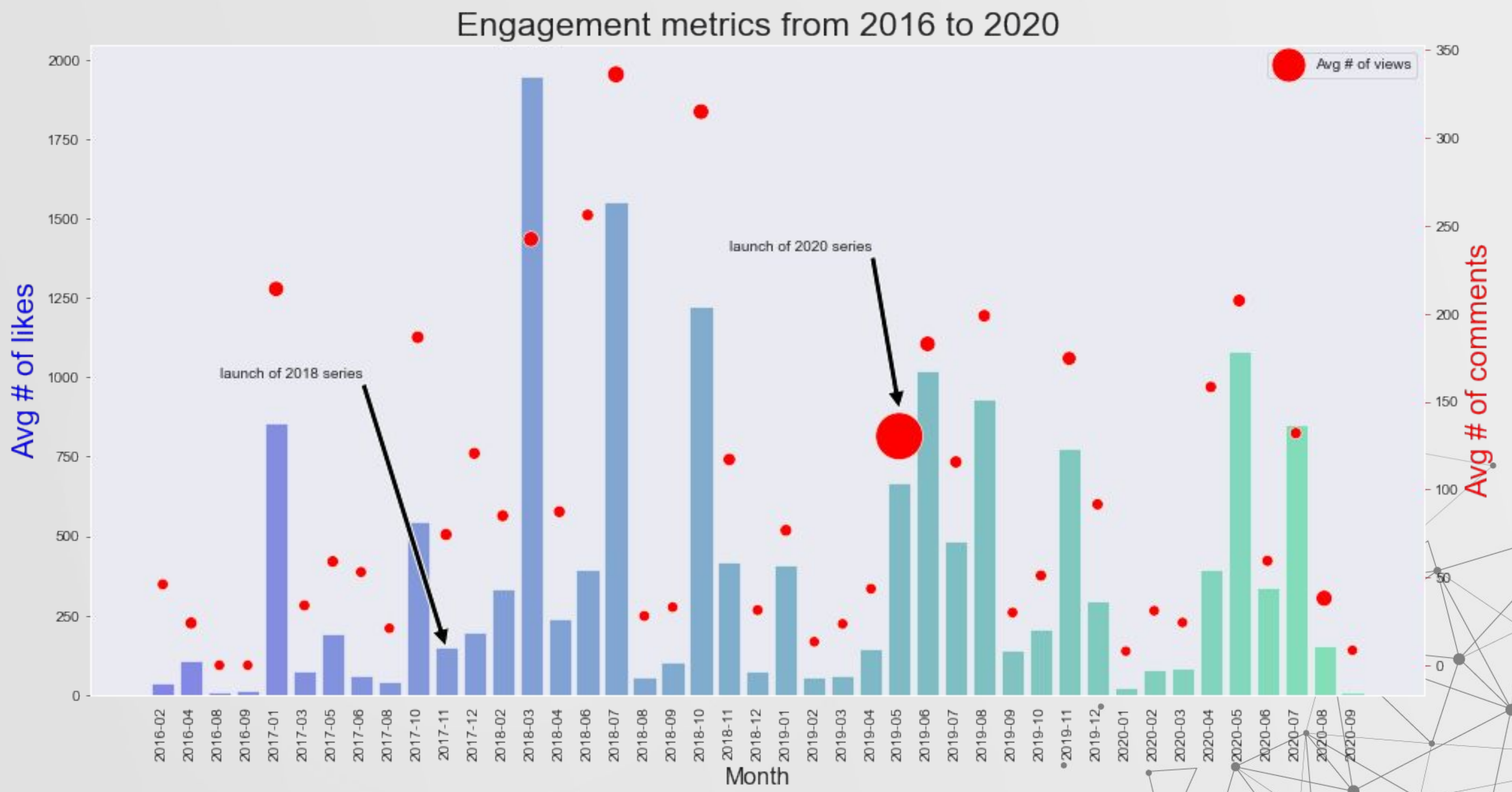
Micro influencers with views abv 10k

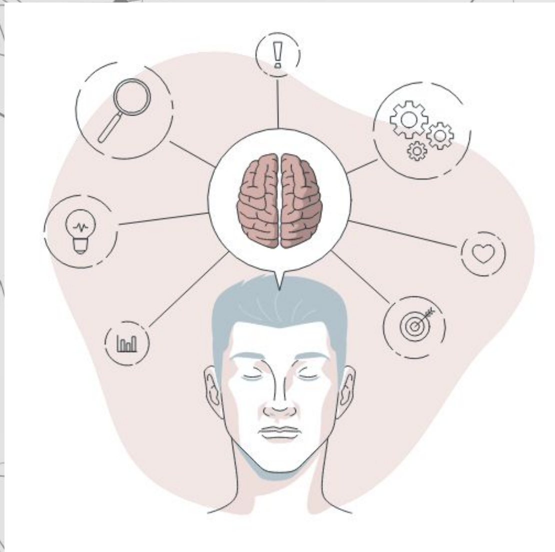




Nano influencers with views below 10k




Engagement metrics from 2016 to 2020
Avg # of views
launch of 2018 series
launch of 2020 series
Avg # of likes
Avg # of comments
Month
2000
1750
1500
1250
1000
750
500
250
0
350
300
250
200
150
100
50
0
2016-02
2016-04
2016-08
2016-09
2017-01
2017-03
2017-05
2017-06
2017-08
2017-10
2017-11
2017-12
2018-02
2018-03
2018-04
2018-06
2018-07
2018-08
2018-09
2018-10
2018-11
2018-12
2019-01
2019-02
2019-03
2019-04
2019-05
2019-06
2019-07
2019-08
2019-09
2019-10
2019-11
2019-12
2020-01
2020-02
2020-03
2020-04
2020-05
2020-06
2020-07
2020-08
2020-09



03

Ready for more EDA?

Let’s take a look at insights from SPACY and Word2Vec

SPACY Named Entity Recognition

‘TIME’

8 hours
A few hours
More than an hour
2 hours
12 hours
30mins
Long hours
Last night
This morning

‘DATE’

2020
yesterday
2018
today
tomorrow
A few weeks
A year
2 Years

SPACY Named Entity Recognition

‘ORG

XL

Aeron

Amazon

Secret labs

DXRacer

SecretLab

Last night

IKEA

‘PERSON’

Omega

Herman

Miller

2018

Steelcase

Mike

Ben

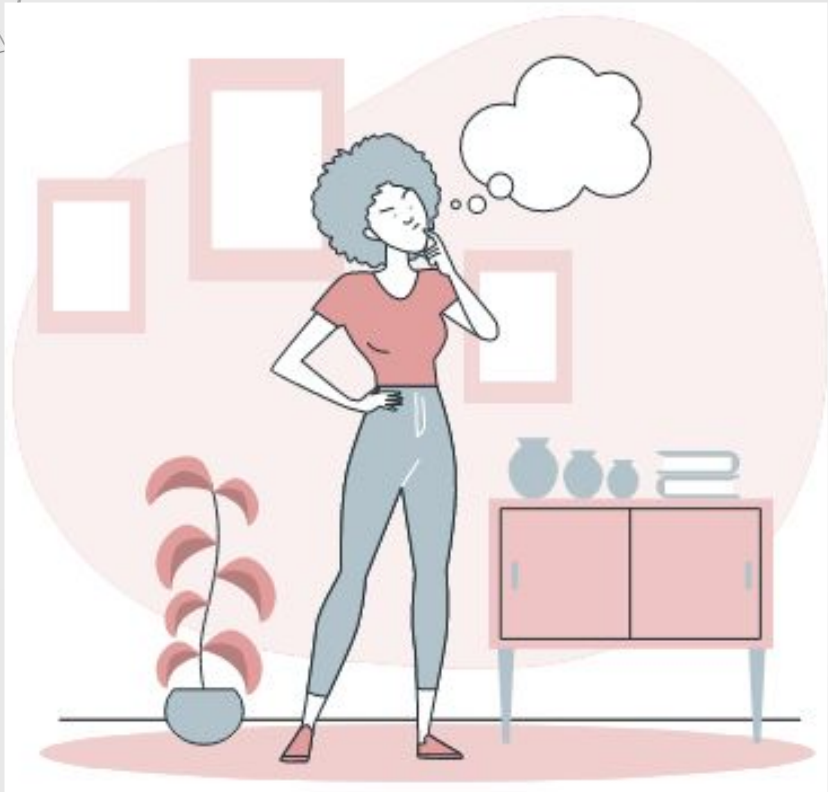
Herman Miller

Titan

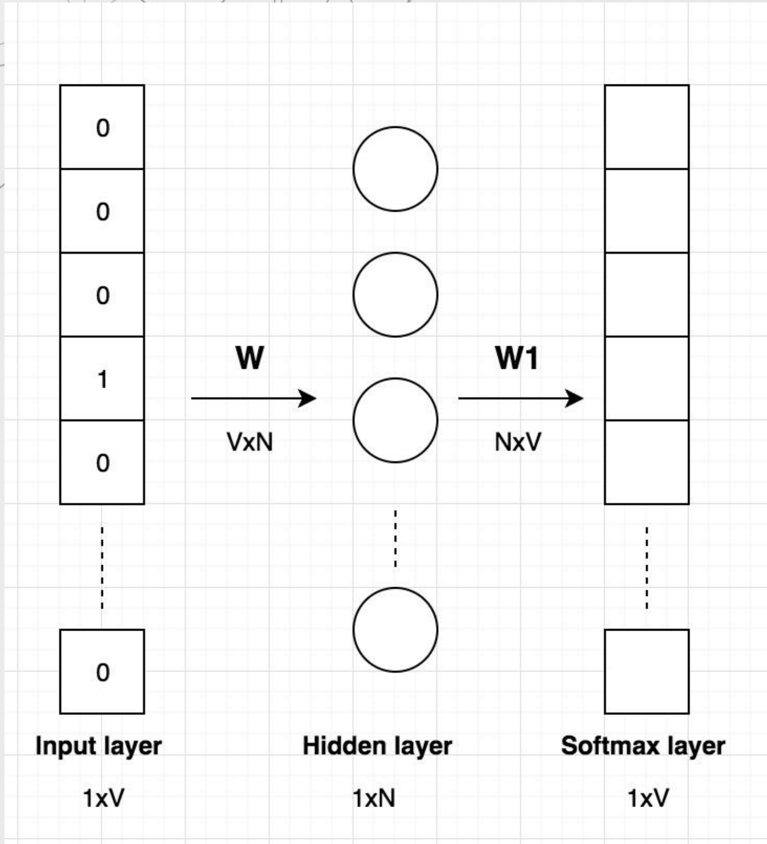
XL

Aeron

Word2Vec



Word2Vec



Skip-Gram

Given an input word in a sentence, the NN will predict how likely it is for each word in the vocab being that input word's nearby word (cosine similarity)

Works well with a small amount of the training data.

C-BOW

Given a context of words (surrounding a word) in a sentence, the NN will predict how likely it is for each word in the vocab being the word

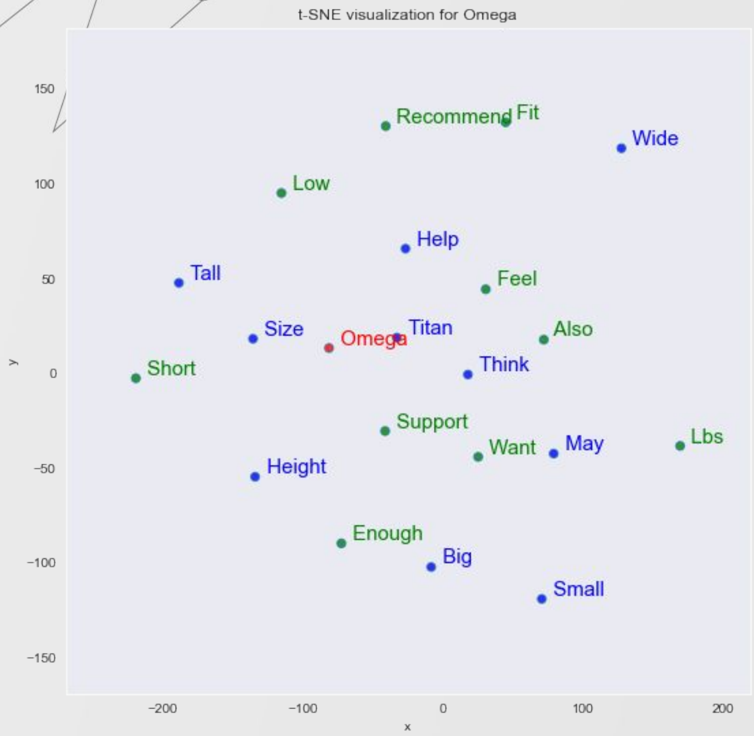
Several times faster to train than the skip-gram, slightly better accuracy for the frequent words.

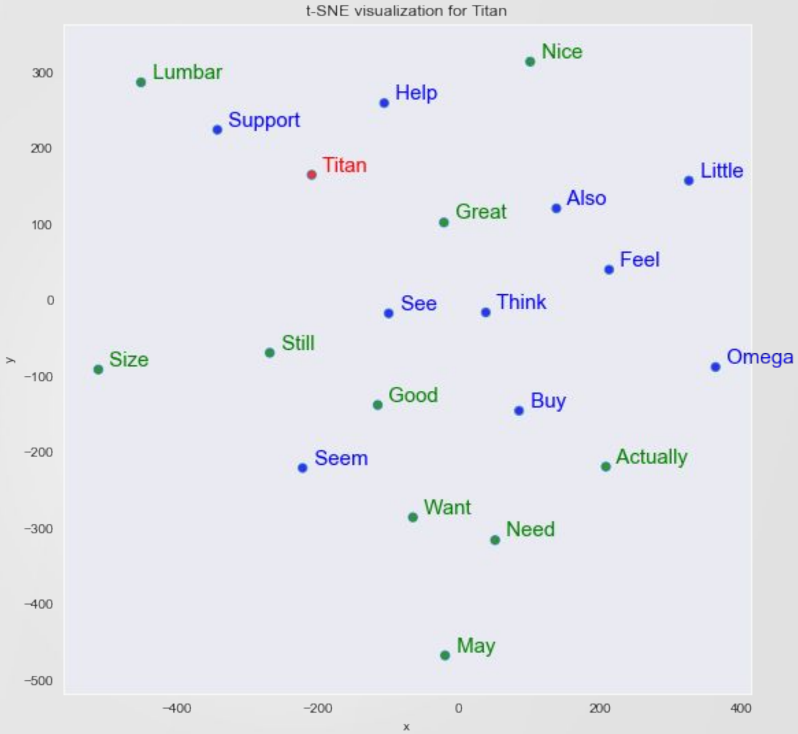
```
w2v_model.wv.relative_cosine_similarity('buy','omega')
```

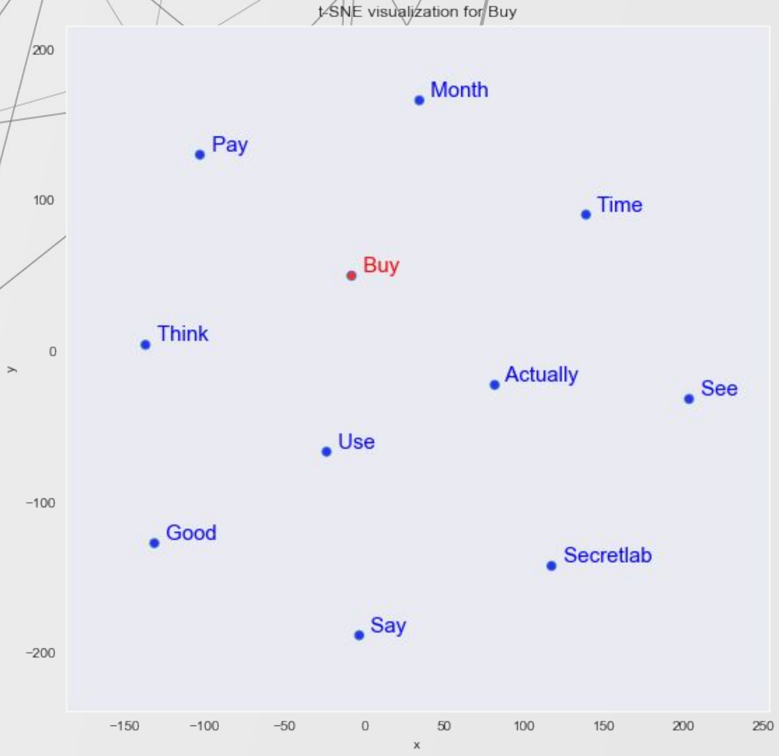
```
0.06996956645724846
```

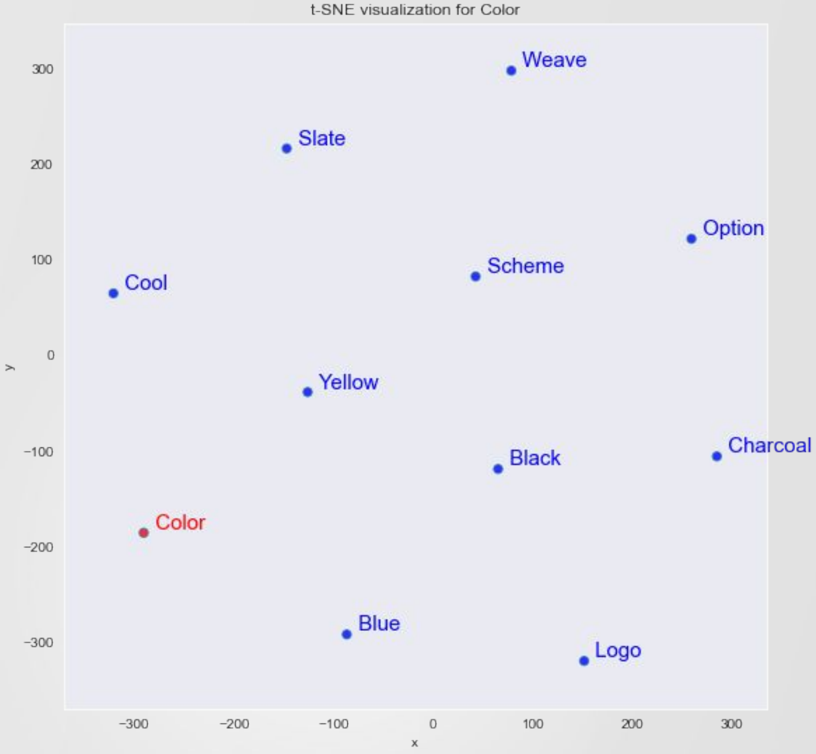
```
w2v_model.wv.relative_cosine_similarity('buy','titan')
```

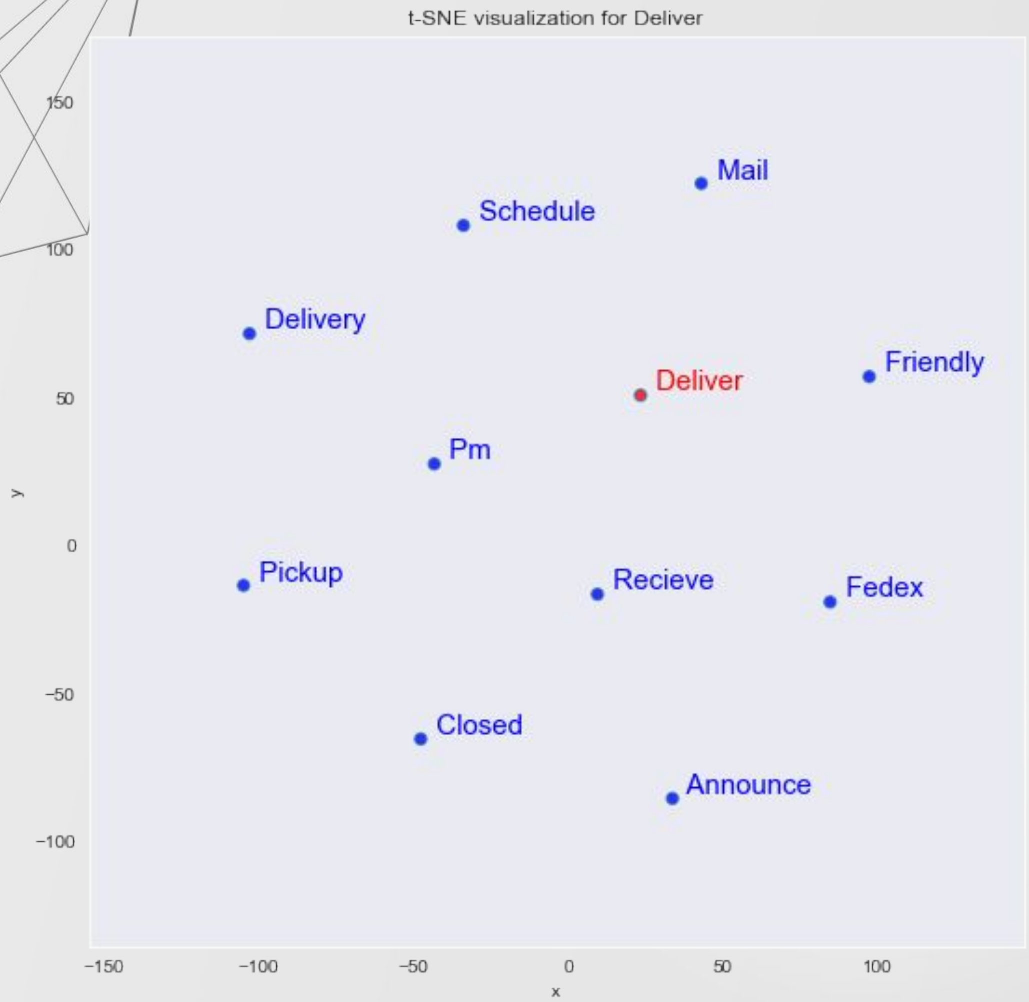
```
0.09292699185735538
```






t-SNE visualization for Buy
Month
Pay
Time
Buy
Think
Actually
See
Use
Good
Secretlab
Say
x
y


t-SNE visualization for Color
Weave
Slate
Option
Scheme
Cool
Yellow
Charcoal
Black
Color
Blue
Logo
x
y


t-SNE visualization for Deliver
Mail
Schedule
Delivery
Friendly
Deliver
Pm
Pickup
Recieve
Fedex
Closed
Announce
150
100
50
0
−50
−100
−150
−100
−50
0
50
100
x
y

Word2Vec

```
w2v_model.wv.most_similar(positive=["fabric", "leather"], negative=["cool"], topn=5)
```

```
[('softweave', 0.6305164694786072),
 ('hot', 0.6189433336257935),
 ('flake', 0.617984414100647),
 ('sweat', 0.6054732203483582),
 ('material', 0.6045715808868408)]
```

Word2Vec

```
w2v_model.predict_output_word(['my', 'next', 'chair', 'will', 'be'], topn=10)
```

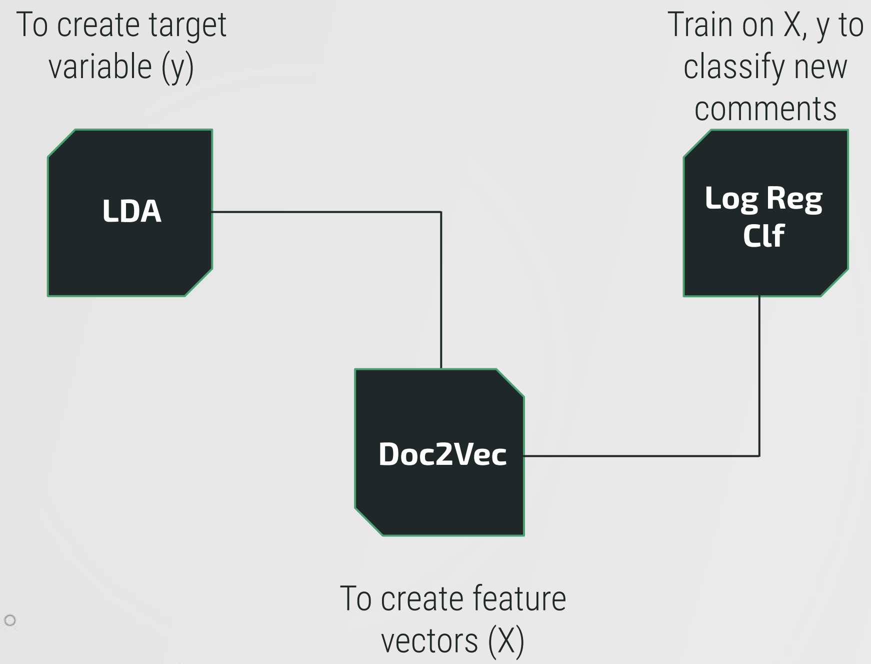
```
[('worn', 0.01168655),
 ('scream', 0.007911612),
 ('backup', 0.007831355),
 ('symptom', 0.0064958106),
 ('wave', 0.006039662),
 ('circle', 0.005978039),
 ('program', 0.0058969483),
 ('news', 0.005699329),
 ('triangle', 0.0056111636),
 ('dealer', 0.005006789)]
```

04

MODEL TIME!!

Here we go...

Modelling


To create target variable (y)
LDA
Train on X, y to classify new comments
Log Reg Clf
Doc2Vec
To create feature vectors (X)

Dominant Topics From LDA

01 Firmness of the seat causing back issues

02 Discussion on how long the chair lasts for

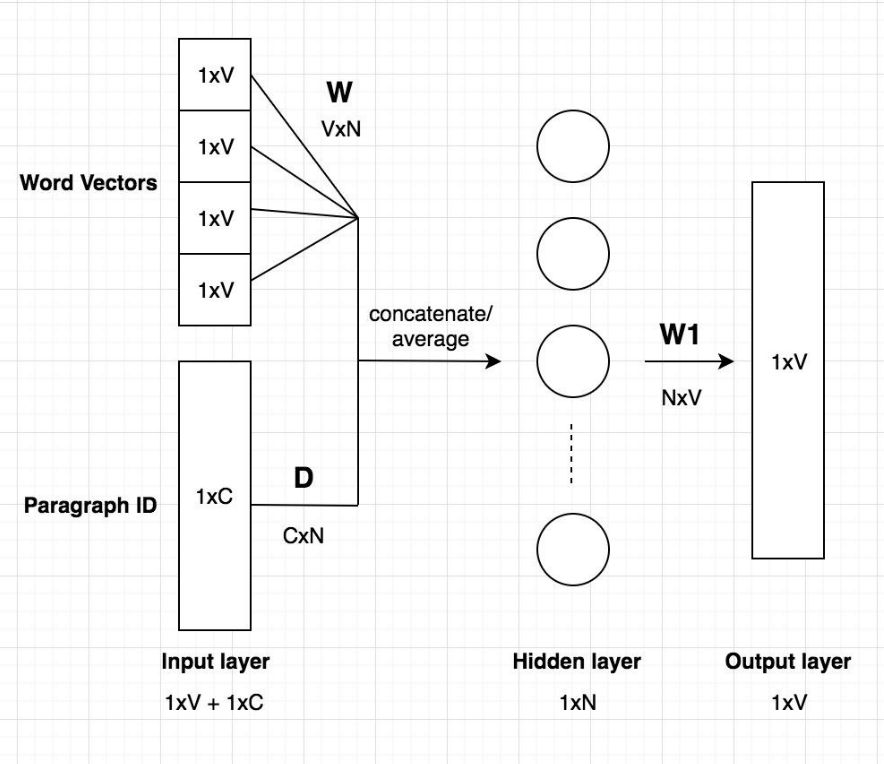
03 Considering between Fabric and Leather

04 How great the Omega chair is!

05 Lumbar Support

06 Compliments on how the video help them with their purchase decision

Doc2Vec



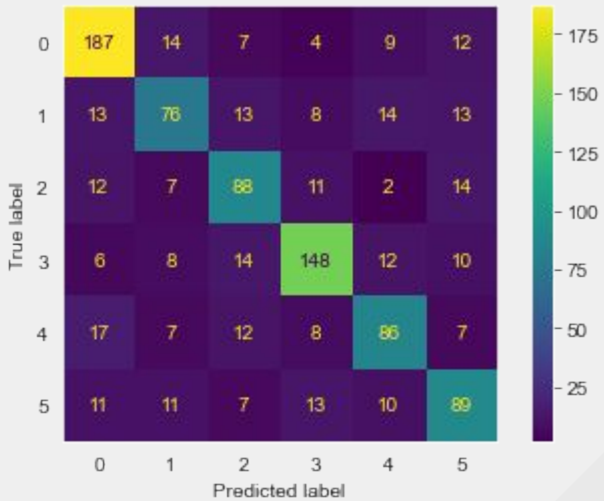
DM

Distributed Memory (DM) acts as a memory that remembers what is missing from the current context — or as the topic of the paragraph. While the word vectors represent the concept of a word, the document vector intends to represent the concept of a document.

D-BOW

DBOW is the doc2vec model analogous to Skip-gram model in word2vec. The paragraph vectors are obtained by training a neural network on the task of predicting a probability distribution of words in a paragraph given a randomly-sampled word from the paragraph.

LR Classifier



Training accuracy - 0.860
Testing accuracy - 0.687
MCC - 0.620

05

Model Evaluation

We’re not done just yet!

Topic 4: How great the Omega chair is

Words determinant to Log Reg Classifier

('comfortably', 0.25152626633644104),
('bro', 0.23484185338020325),
('snug', 0.21710629761219025),
('originally', 0.20406052470207214),
('size', 0.2028352916240692),
('male', 0.19776876270771027),
('husband', 0.18971474468708038),
('tall', 0.187089204788208),
('chill', 0.18451052904129028),
('velcro', 0.1792258620262146)

Words determinant to Doc2Vec

('symbol', 0.36029061675071716),
('ft', 0.31007319688796997),
('bro', 0.30809274315834045),
('incredible', 0.2996496260166168),
('comfortably', 0.29147326946258545),
('snug', 0.2860320508480072),
('size', 0.2838527262210846),
('beginning', 0.27516692876815796),
('male', 0.26859986782073975),
('suggestion', 0.2587391138076782)

Spearmans correlation coefficient: 0.854
Determinant words are correlated (reject H0) p=0.000

FALL IN LOVE WITH YOUR DATA AND REALLY UNDERSTAND YOUR MODEL

THAT’S A WRAP!

Does anyone have any questions?