

BLACK³

BLACK³

BLACK³

BLACK³

BLACK³

SOFTWEAVE®

NEW
SECRETLAB TITAN
SOFTWEAVE® BLACK®

NEW
SECRETLAB OMEGA
SOFTWEAVE® BLACK®

SOFTWEAVE®
BLACK³

NOW AVAILABLE IN
OMEGA, TITAN AND TITAN XL



INSPIRED BY STREETWEAR
TRIPLE-BLACK

// HIGH-PERFORMANCE FABRIC

BLACK³

BLACK³

BLACK³

BLACK³

BLACK³

BLACK³

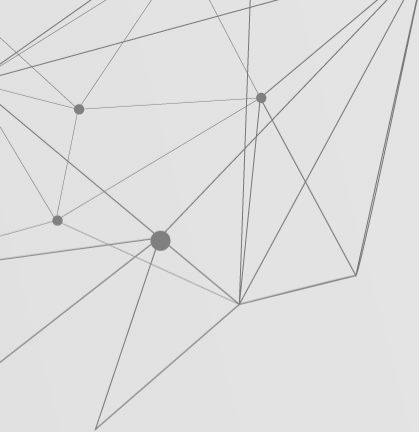
BLACK³

Brand insights for



**SECRET
LAB**

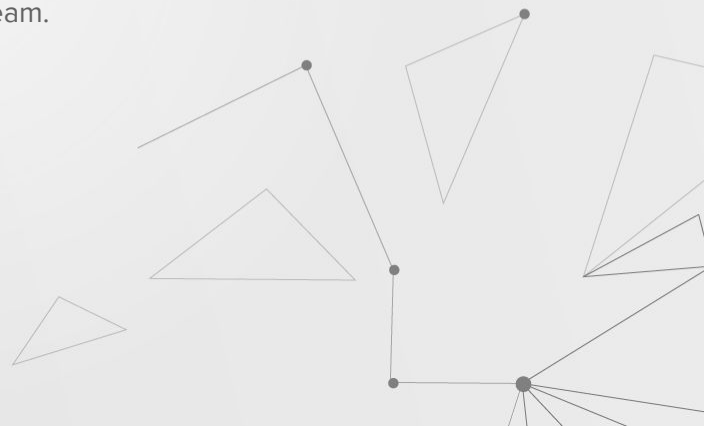
Prakash N
Data Analyst



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

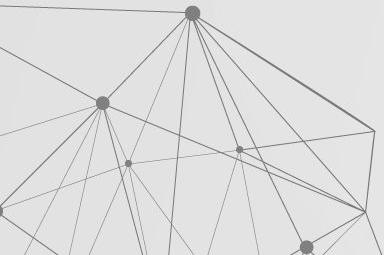
Unique comments

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**):

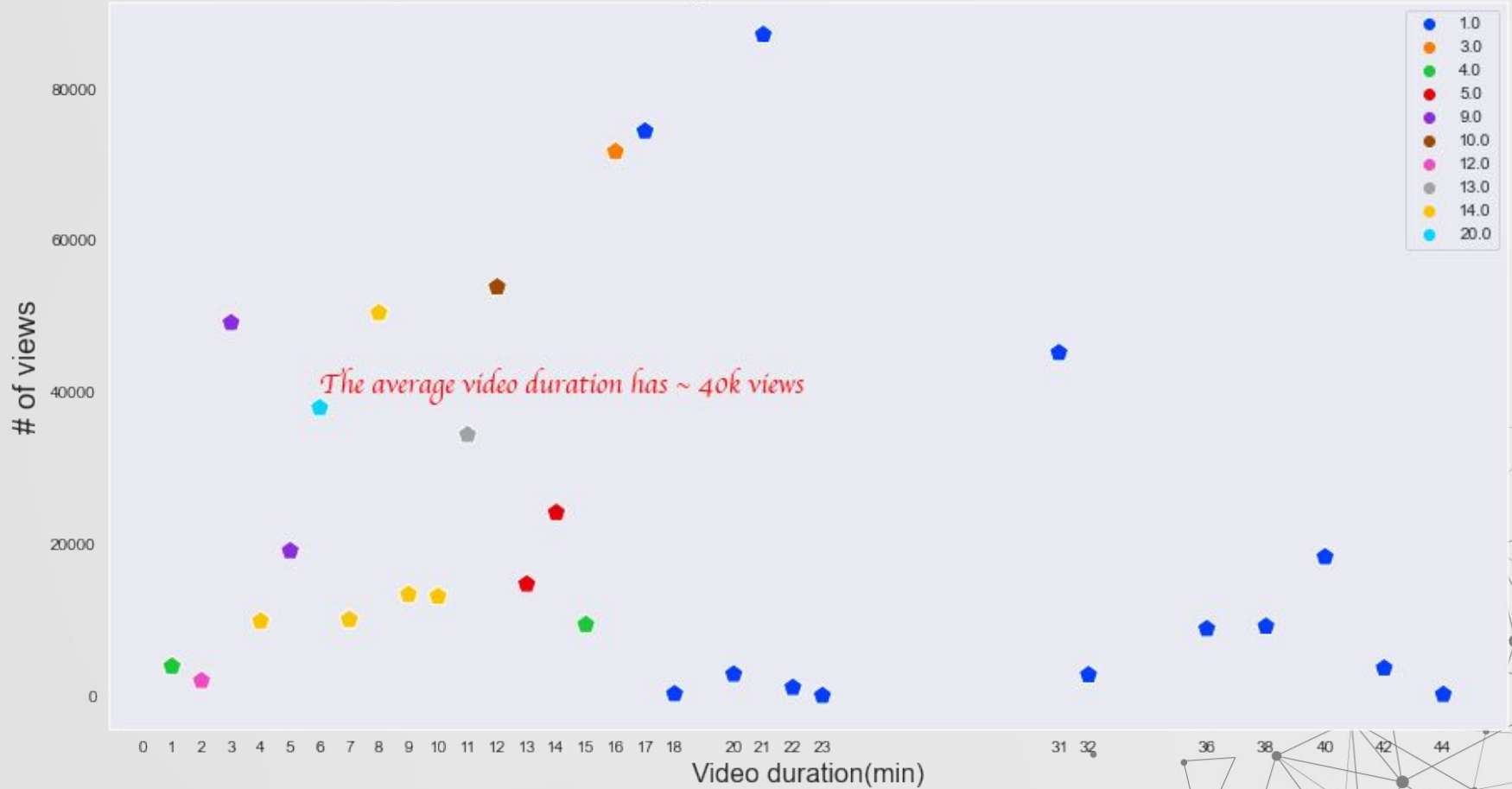
“one of the best ads i've ever seen “



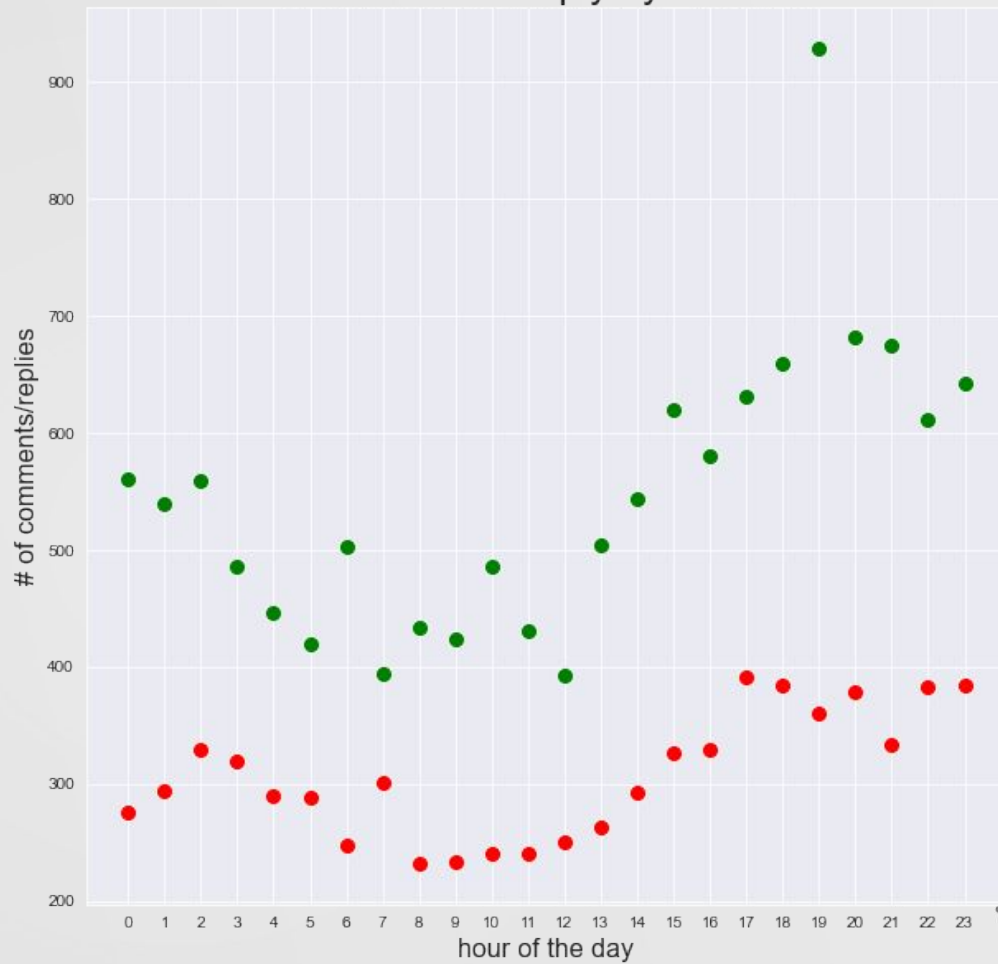
RESPONSE

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

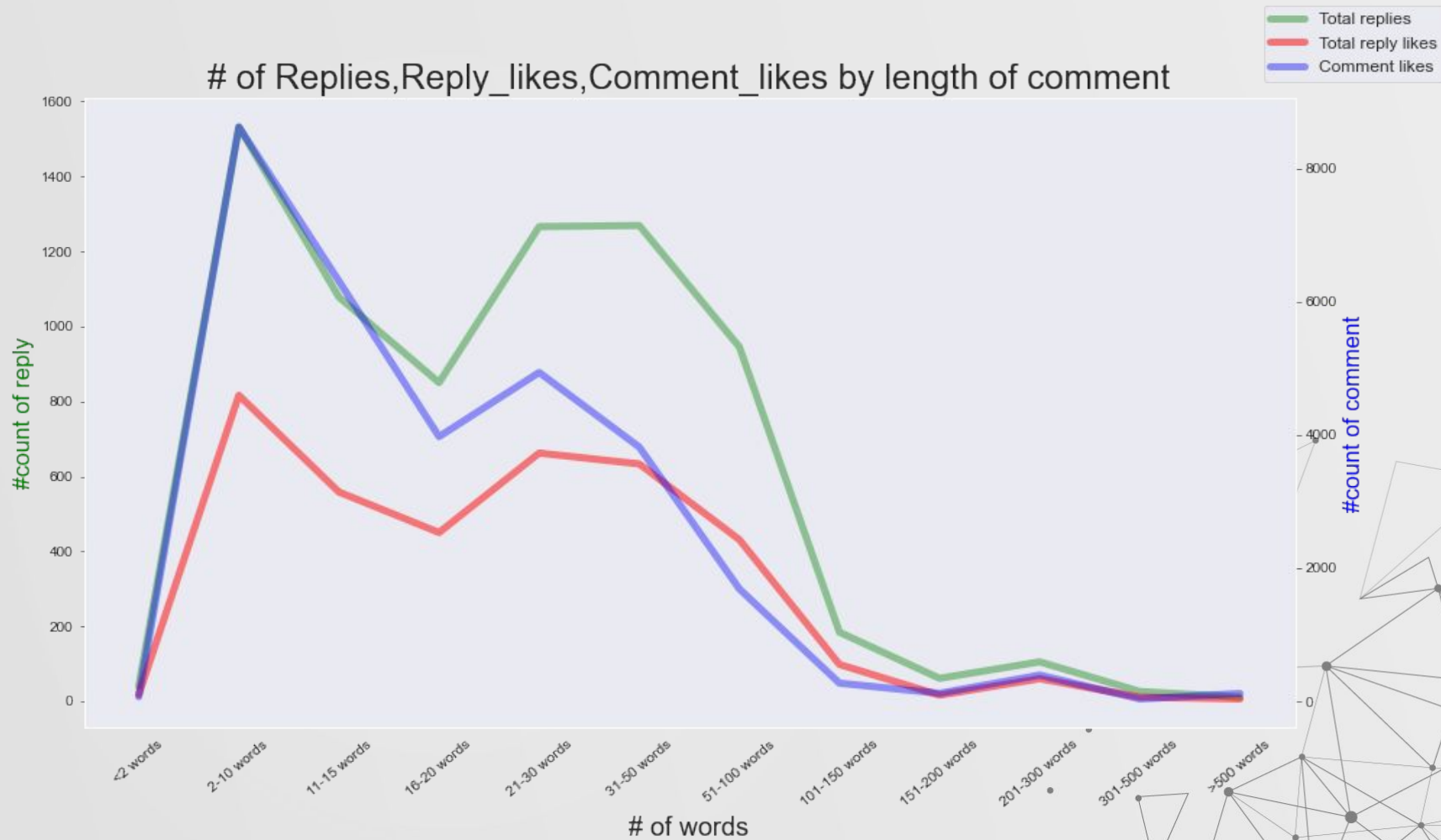
Video duration vs average view count of total number of videos

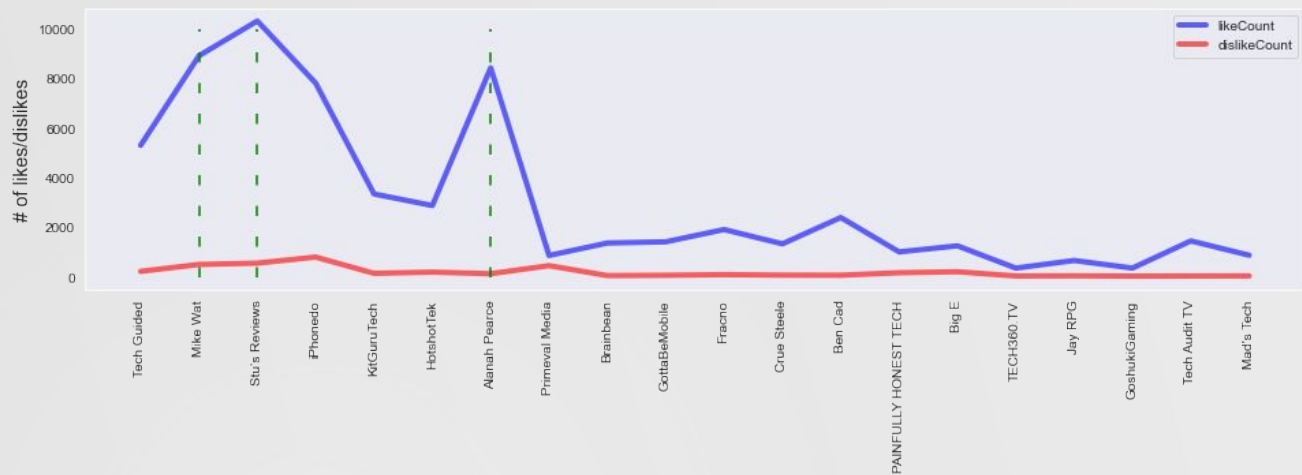


Comment vs Reply by the hour

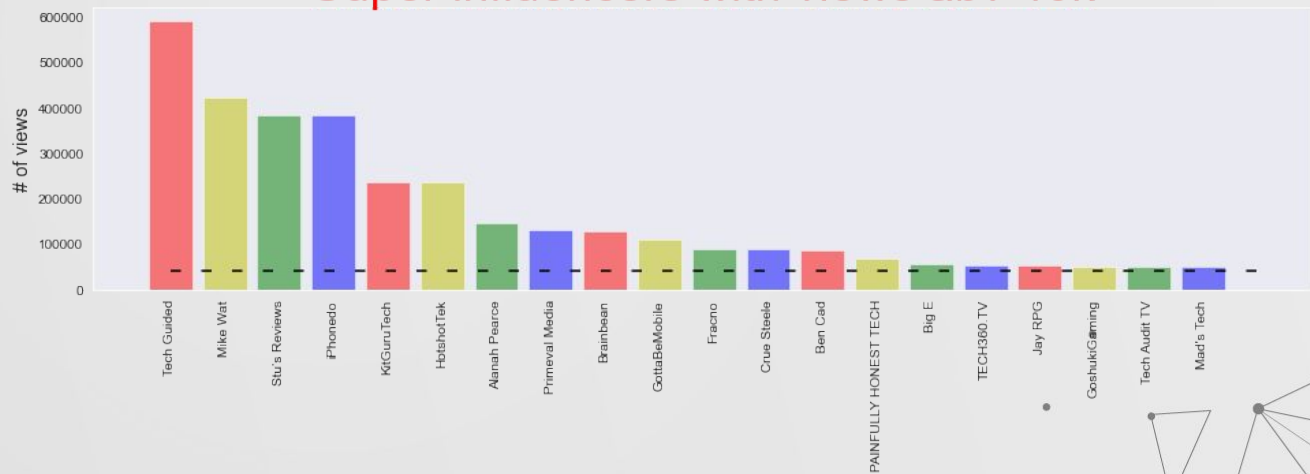


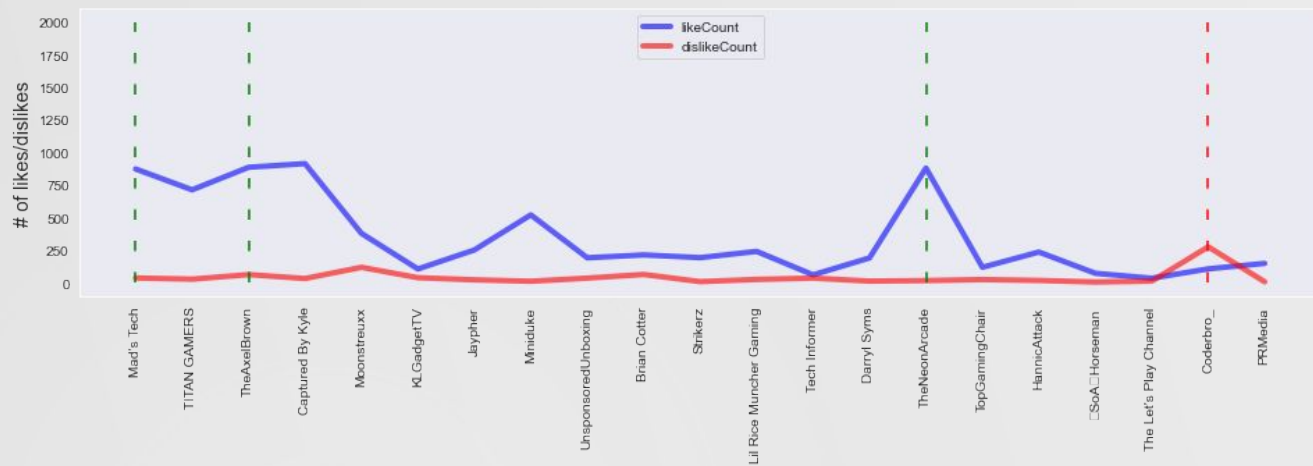
of Replies, Reply_likes, Comment_likes by length of comment



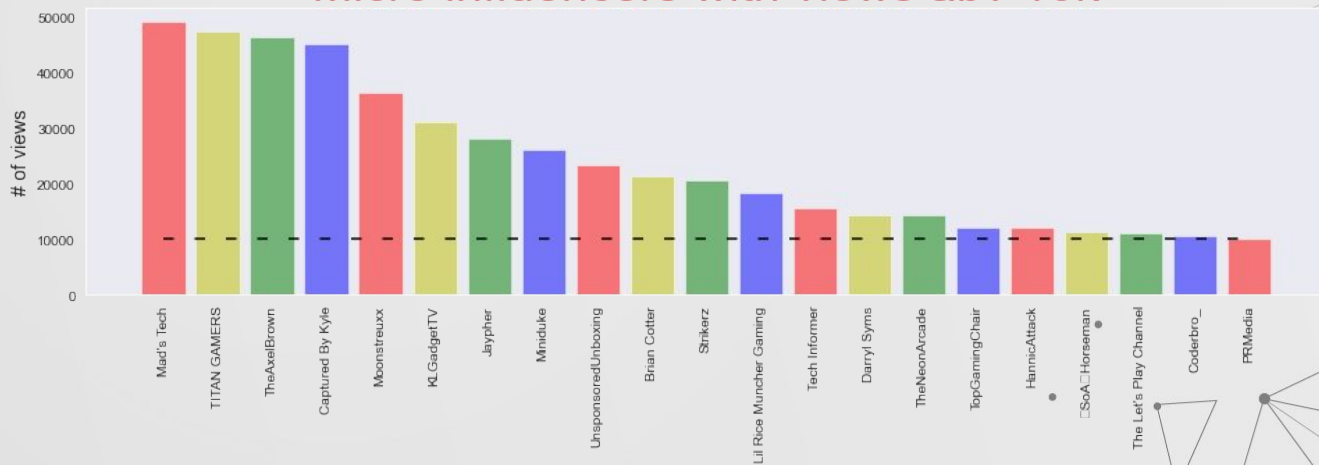


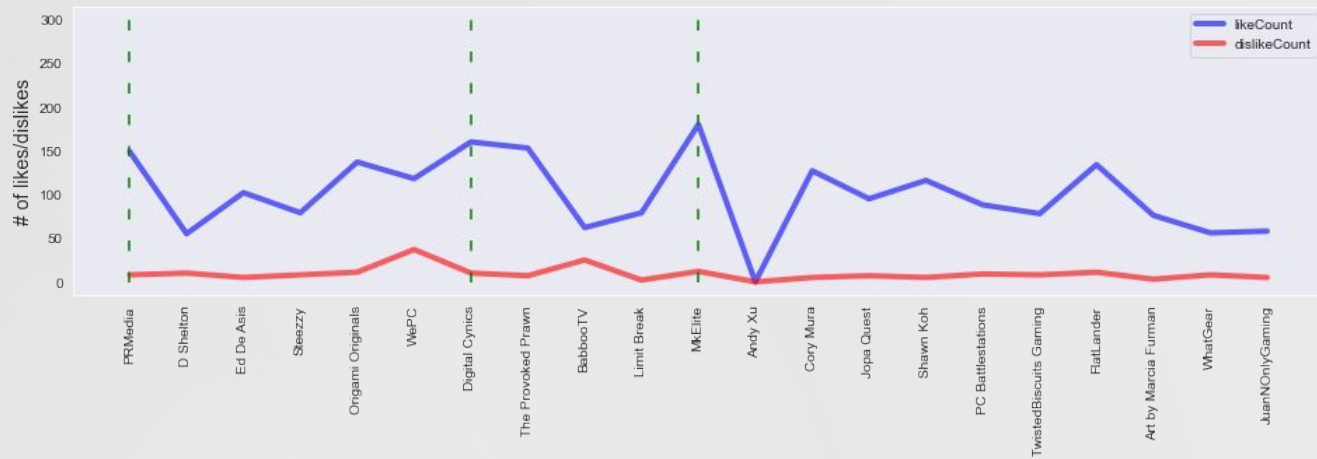
Super influencers with views abv 40k



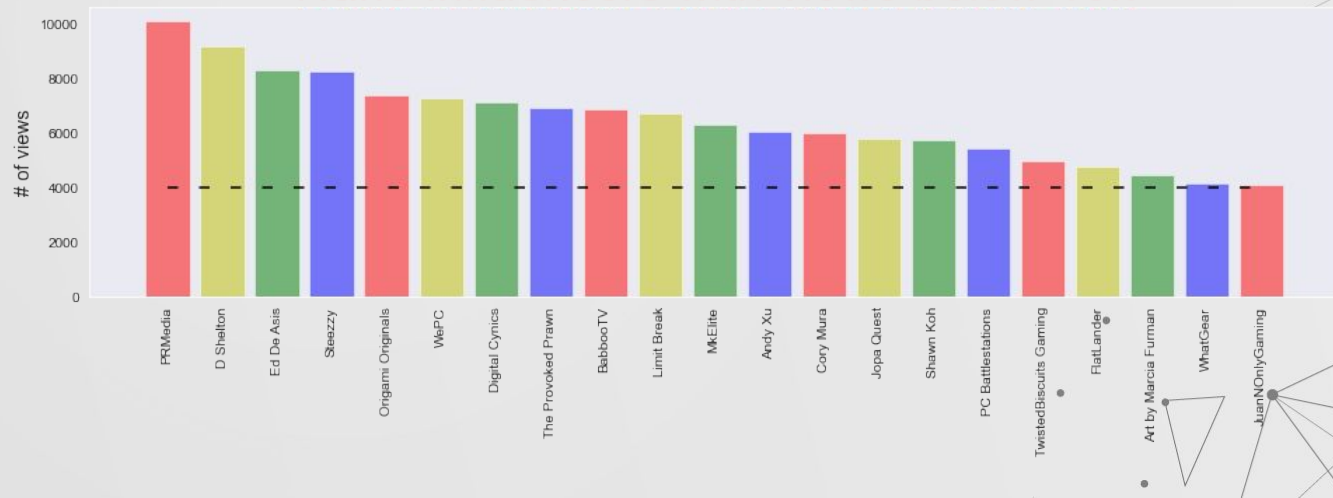


Micro influencers with views abv 10k

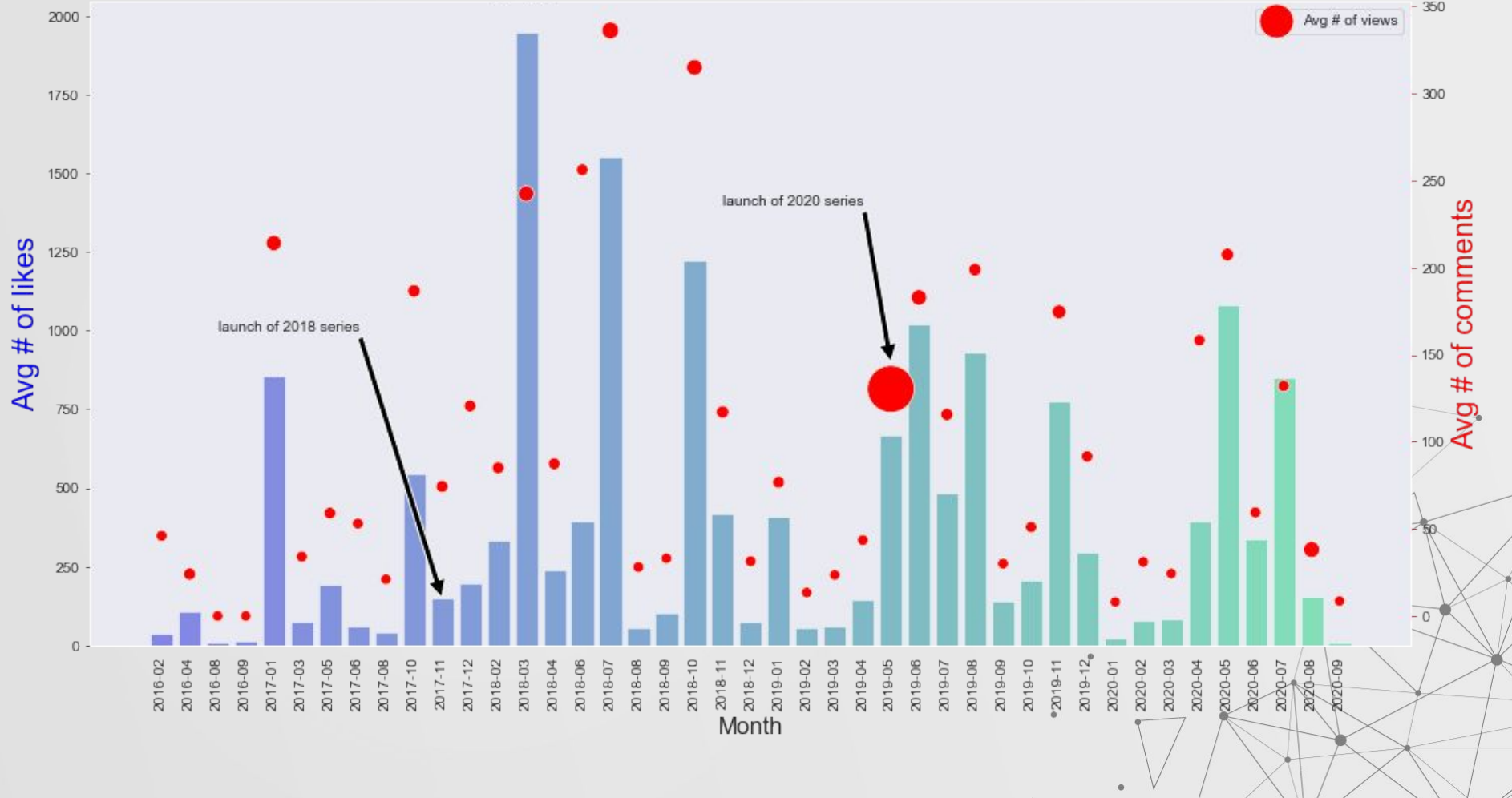


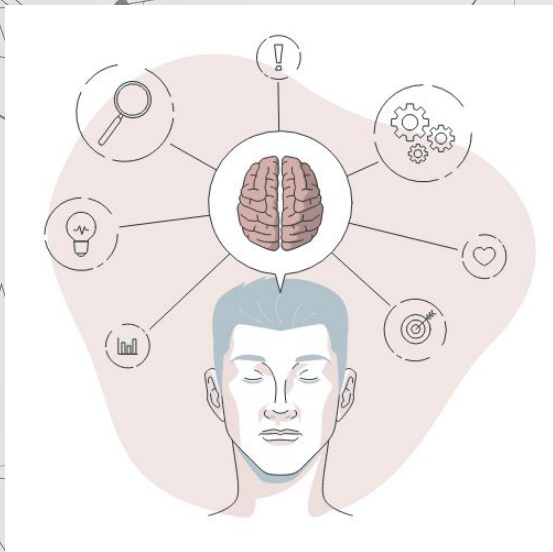


Nano influencers with views below 10k



Engagement metrics from 2016 to 2020





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

‘DATE’

2020
2018
yesterday
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
2018 Miller
Mike Ben Steelcase
Titan Herman Miller
XL Aeron

Word2Vec



Word2Vec

Skip-Gram

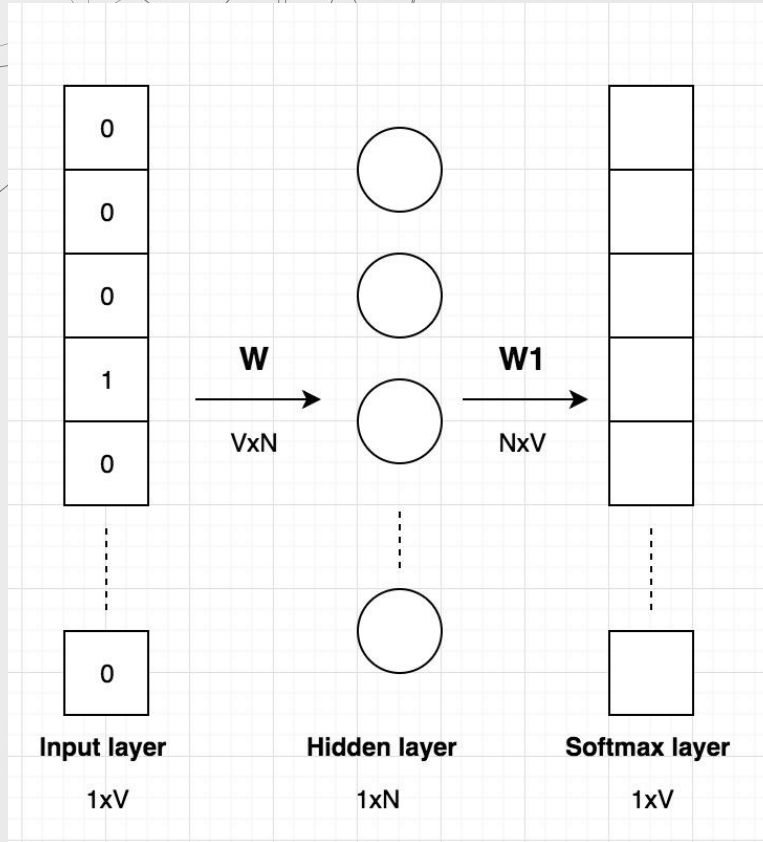
Given an input word in a sentence, the NN will predict how likely it is for each word in the vocab being that input word's nearby word (cosine similarity)

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.





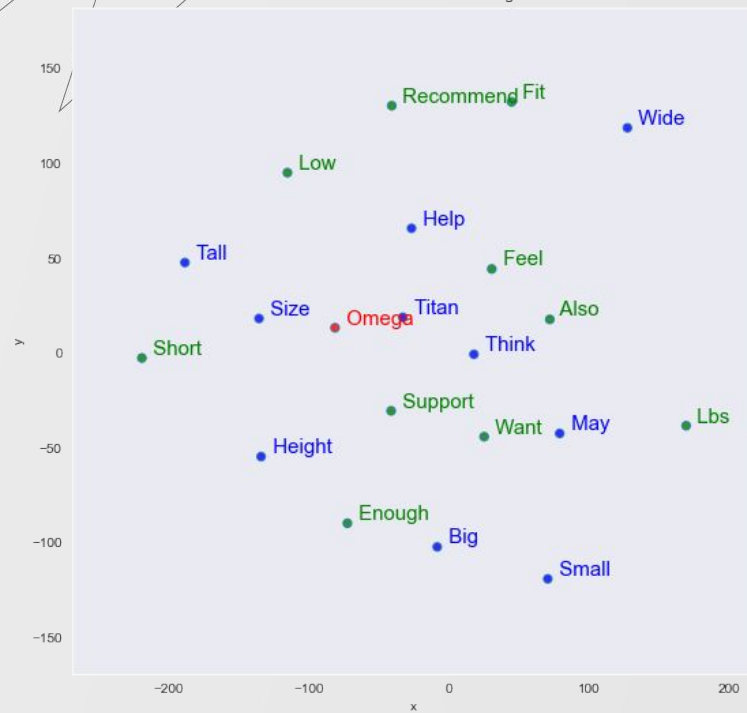
```
w2v_model.wv.relative_cosine_similarity('buy', 'omega')
```

```
0.06996956645724846
```

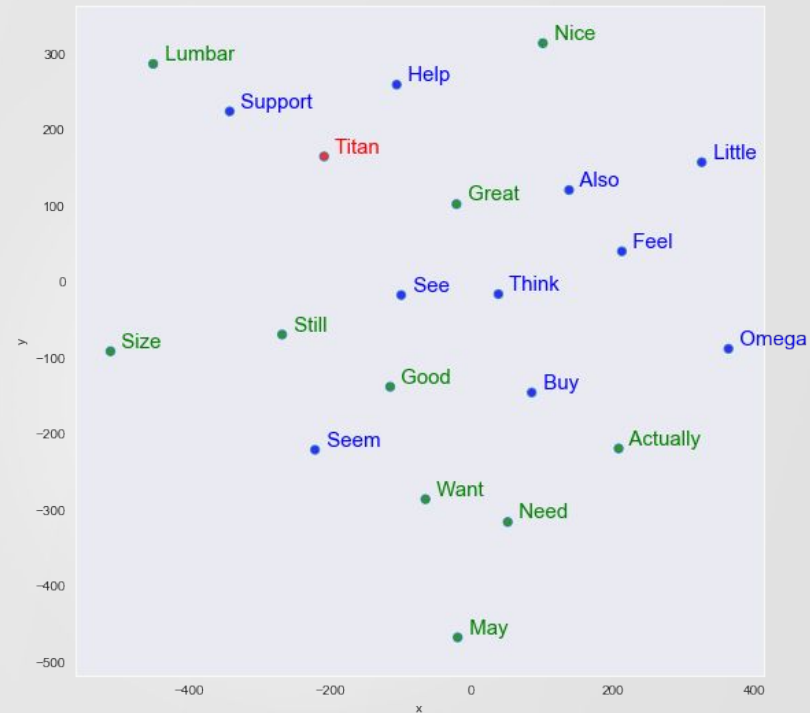
```
w2v_model.wv.relative_cosine_similarity('buy', 'titan')
```

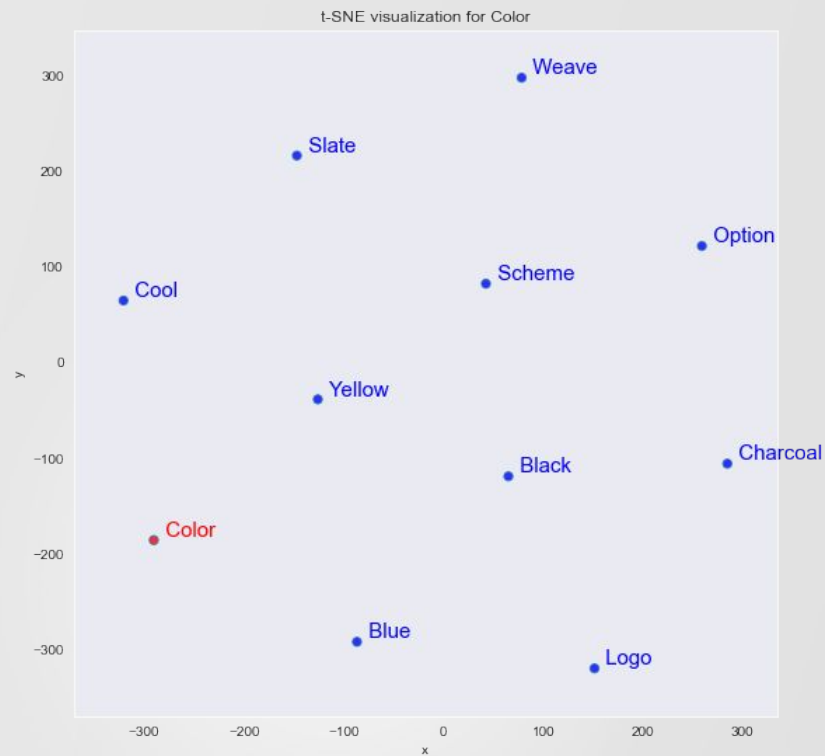
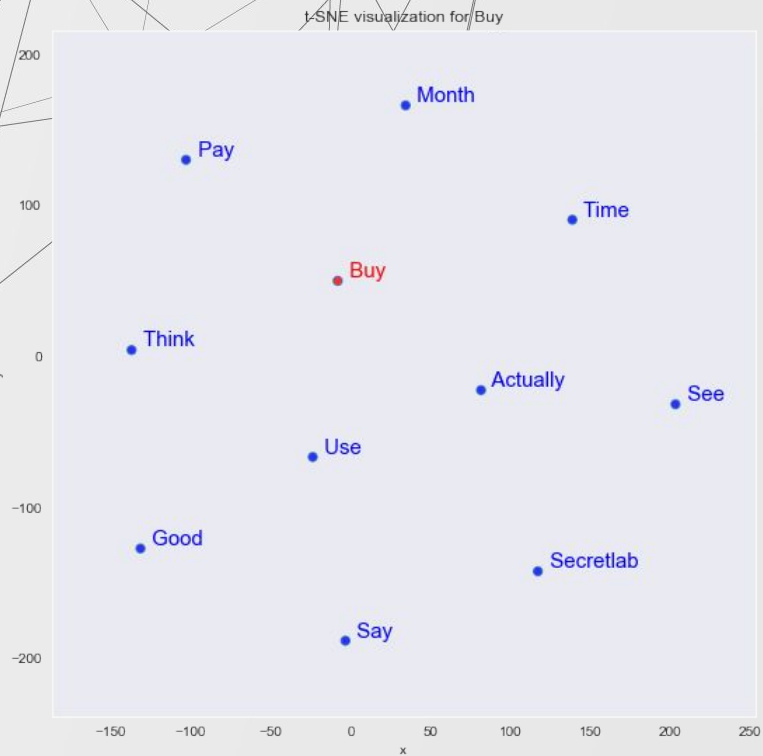
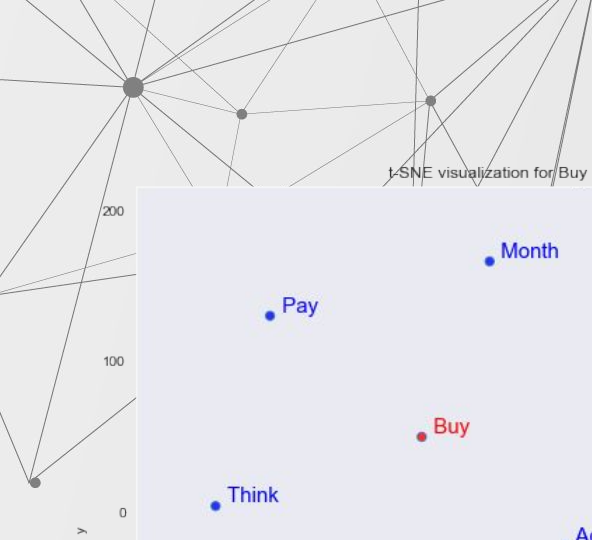
```
0.09292699185735538
```

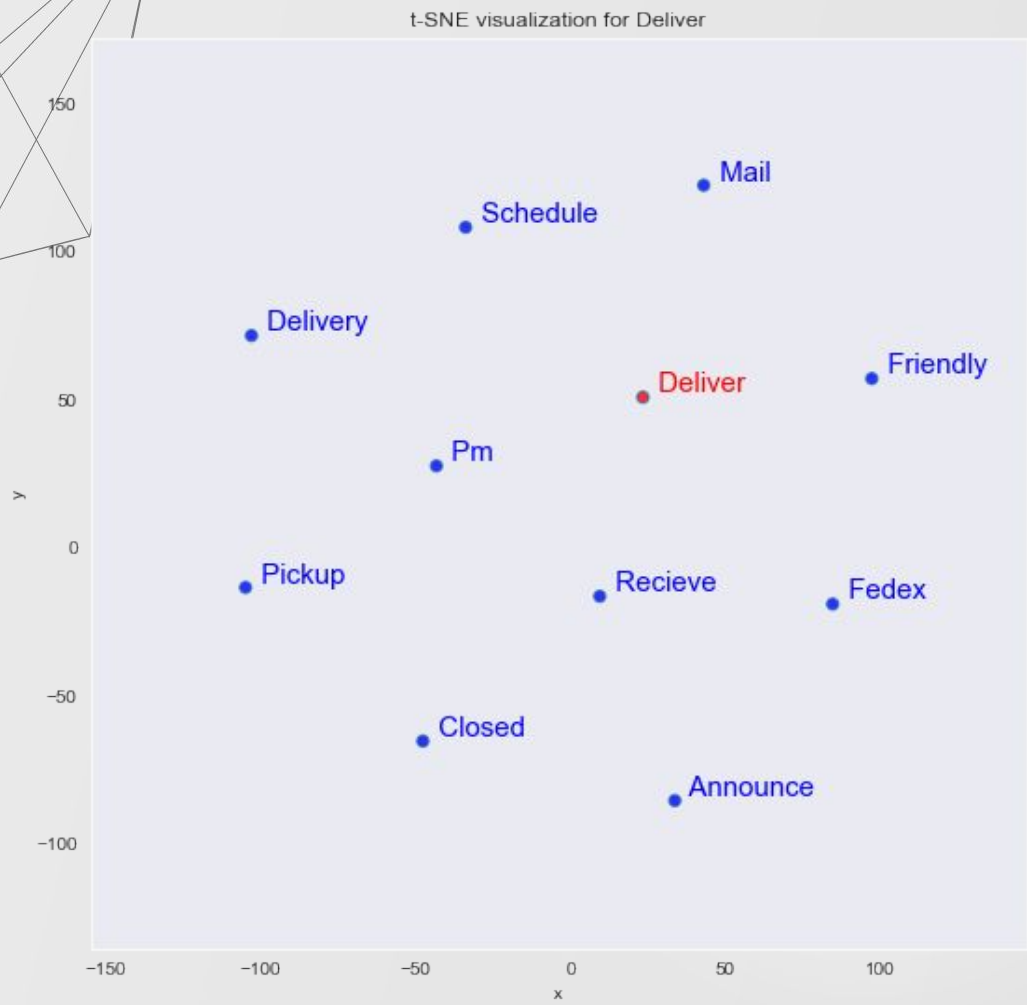
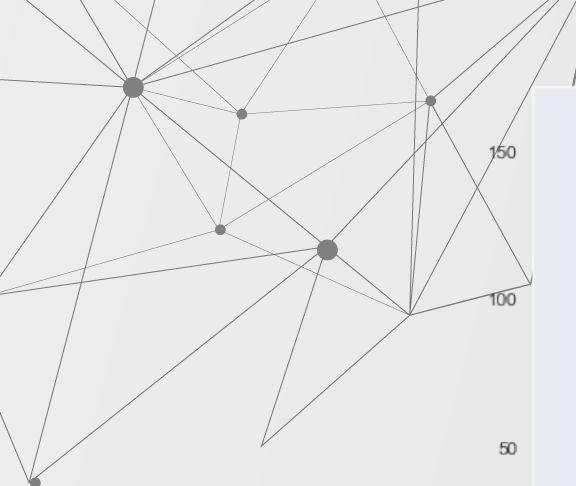
t-SNE visualization for Omega



t-SNE visualization for Titan









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)

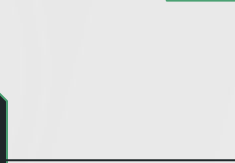
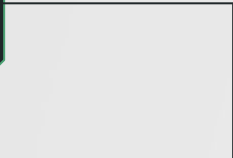
LDA

Train on X, y to
classify new
comments

**Log Reg
Clf**

Doc2Vec

To create feature
vectors (X)



Firmness of the seat
causing back issues

01

Discussion on how
long the chair lasts for

02

Considering between
Fabric and Leather

03

Dominant Topics From LDA

04

How great the Omega
chair is!

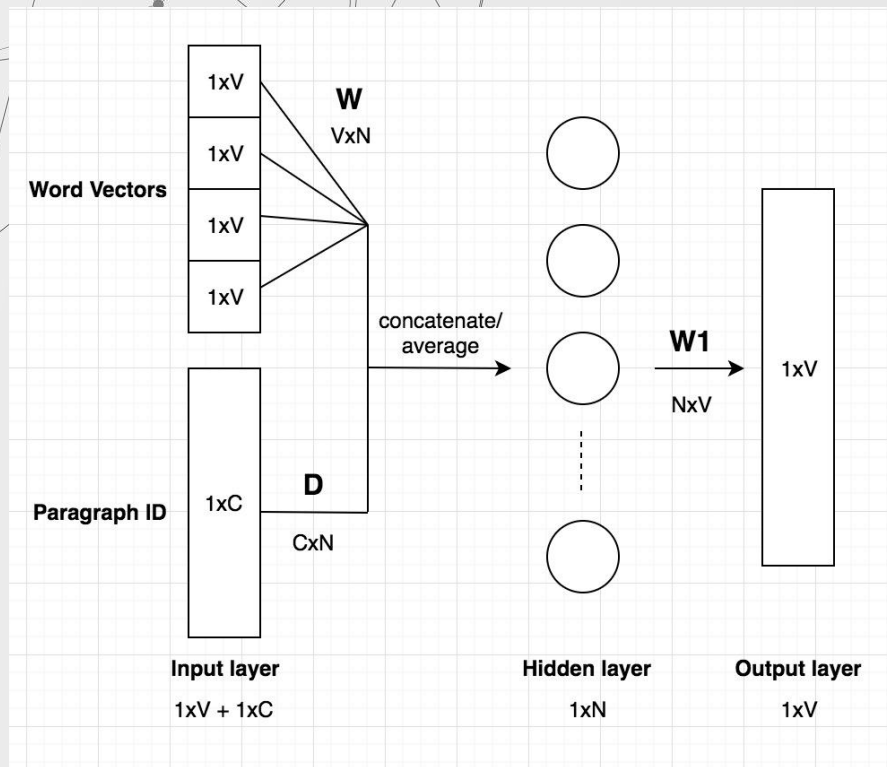
05

Lumbar Support

06

Compliments on how the
video help them with their
purchase decision

Doc2Vec



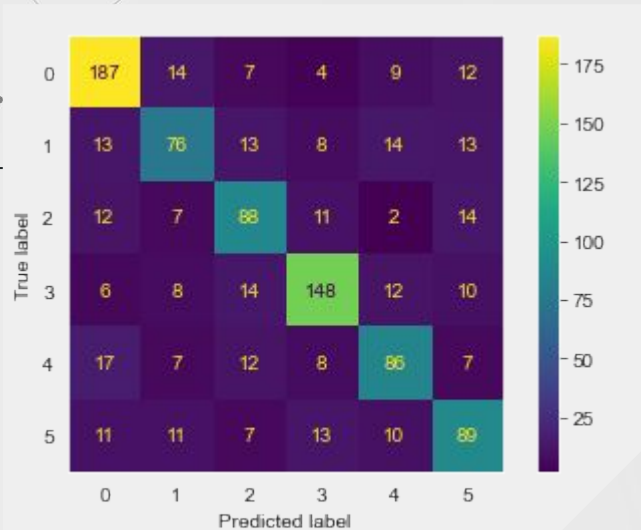
DM

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

The background features a complex network of dark gray nodes connected by thin lines, creating a web-like structure. A vertical green line is positioned on the left side of the slide.

05

Model Evaluation

We're not done just yet!

Topic 4:

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)

Spearman's 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?

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.



06

SNEAK PEEK

You can enter here the subtitle if you need it