Instanalysis

Of Instagram

not sponsored by instagram

Table of Contents

1

Background

Context, Problem, Goals, and Workflow

Context

What's the big deal about Instagram?

- Important Marketing Strategy
- CPM, CPC, CPRresult
- Well established IG page comes with benefits
- The first step to achieving that is getting likes

The Problem

justinbieber • Follow

justinbieber Above it all
5 d

alfredoflores wow
4 d 773 likes Reply
View replies (15)

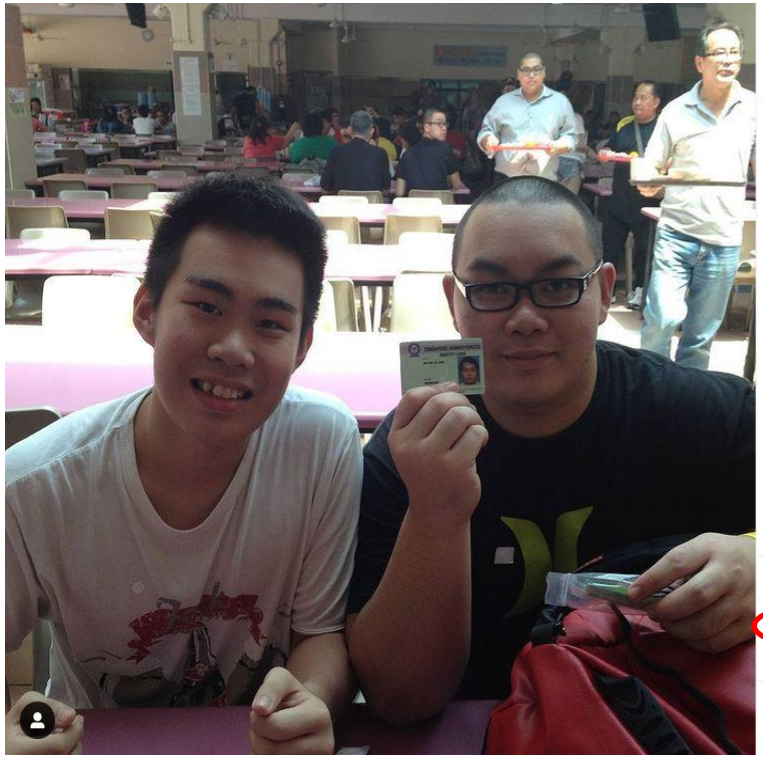
izahrabz Hi
5 d 67 likes Reply
View replies (7)

rohitt_09_nita_fans_official
5 d 60 likes Reply
View replies (2)

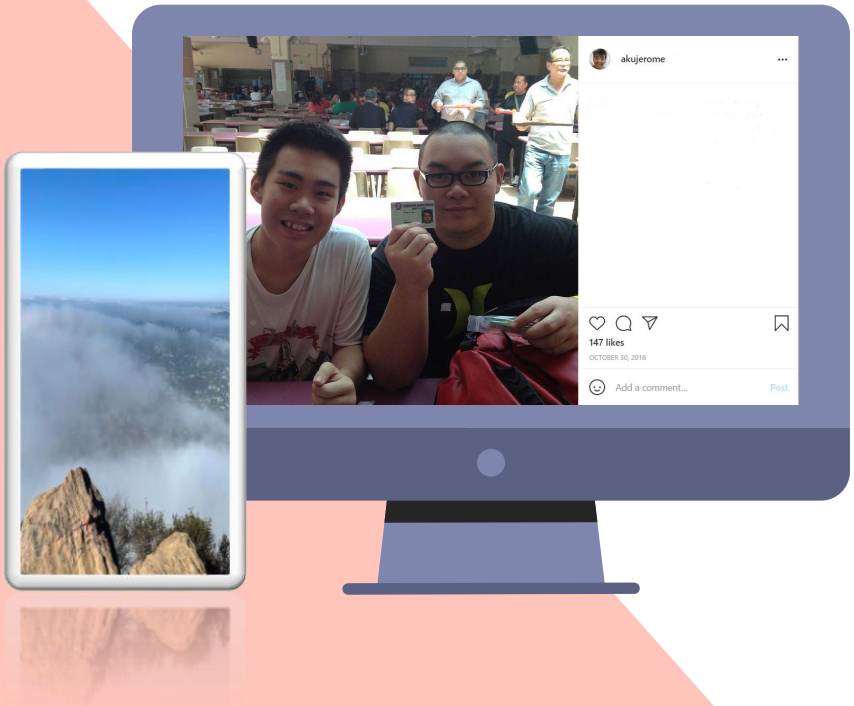
802,180 likes
5 DAYS AGO

Add a comment... Post

The Problem



The Problem



Companies post Instagram photos all the time but many just can't find success

Goals

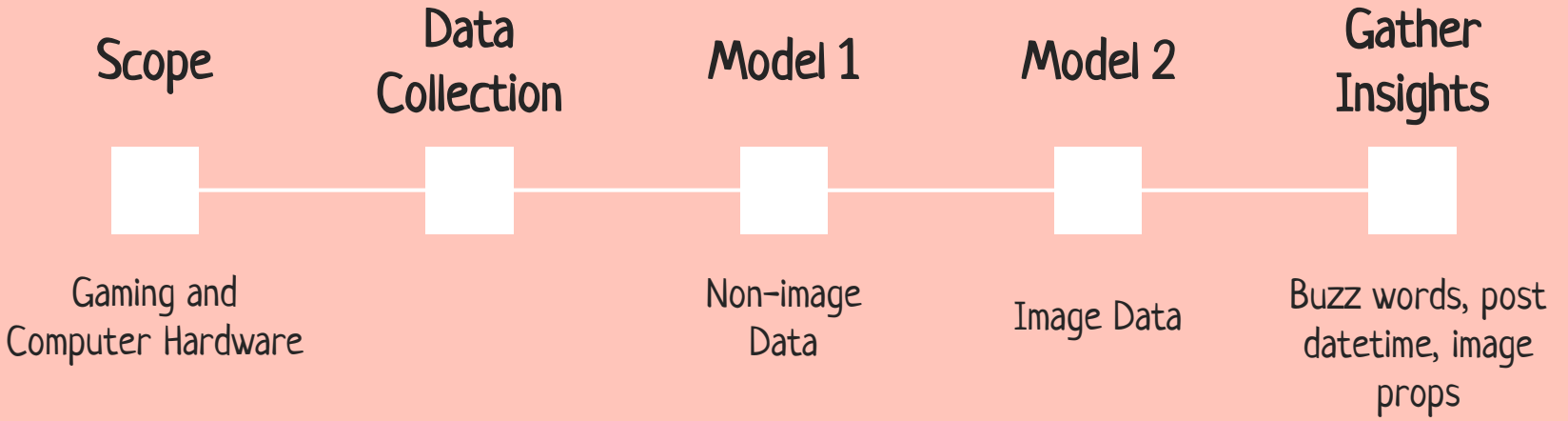
Image Factors

Find the optimal image ‘ingredients’ (props) to maximise post performance

Non-Image Factors

Find the optimal (1) caption and (2) post date/time to maximise post performance

Workflow


Scope
Data Collection
Model 1
Model 2
Gather Insights
Gaming and Computer Hardware
Non-image Data
Image Data
Buzz words, post datetime, image props

2 Non-Image Data

2.1

The Data

What & Where

Posts since Sep 2020

From Aftershock, ErgoTune, Logitech, MSI, Prism+, Razer, Secretlab, Omnidesk, and SteelSeries

From

Dun tell u

Features

Num_likes

Number of likes; int

Caption

Post's caption; str

DateTime

Date and time of post; obj

Followers

Posting account's number of followers at the time of post launch; int

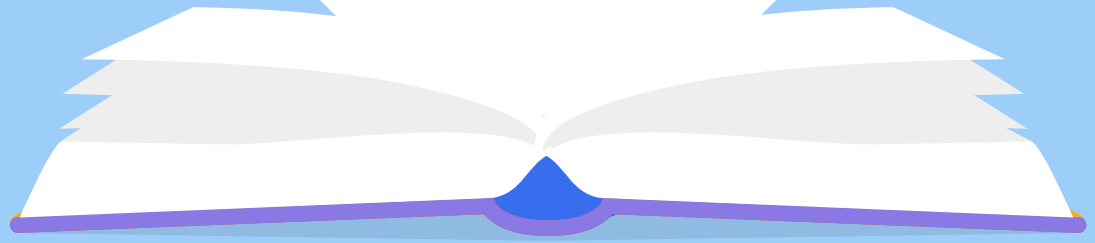
Account

Username of posting account; str

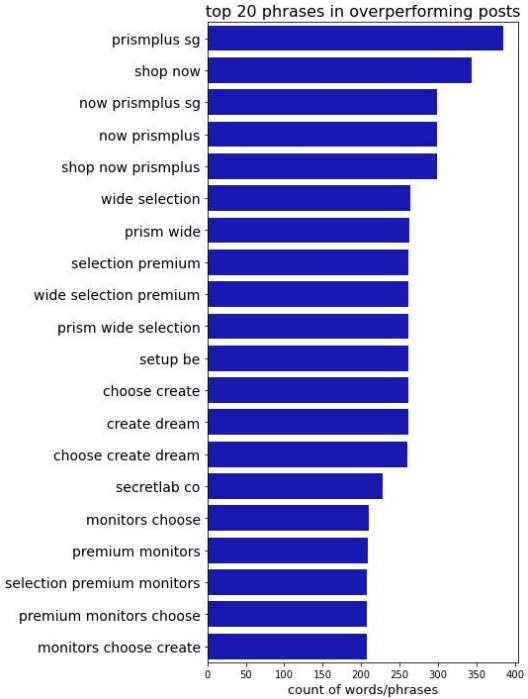
Overperforming

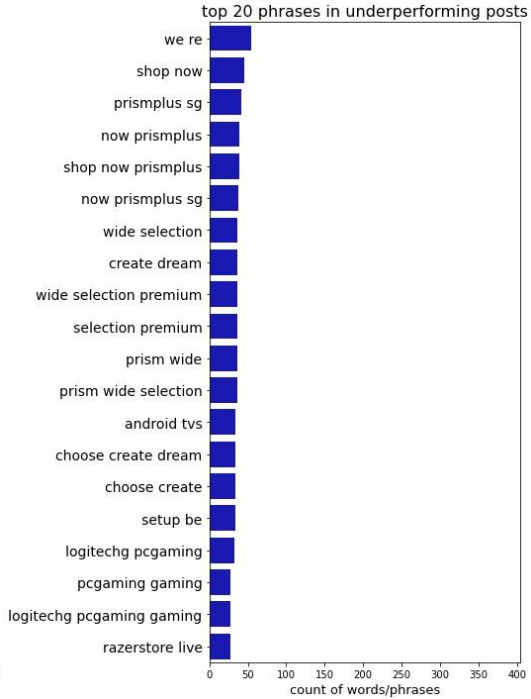
Binary feature based on weighted score between likes and followers

2.2 EDA



Captions

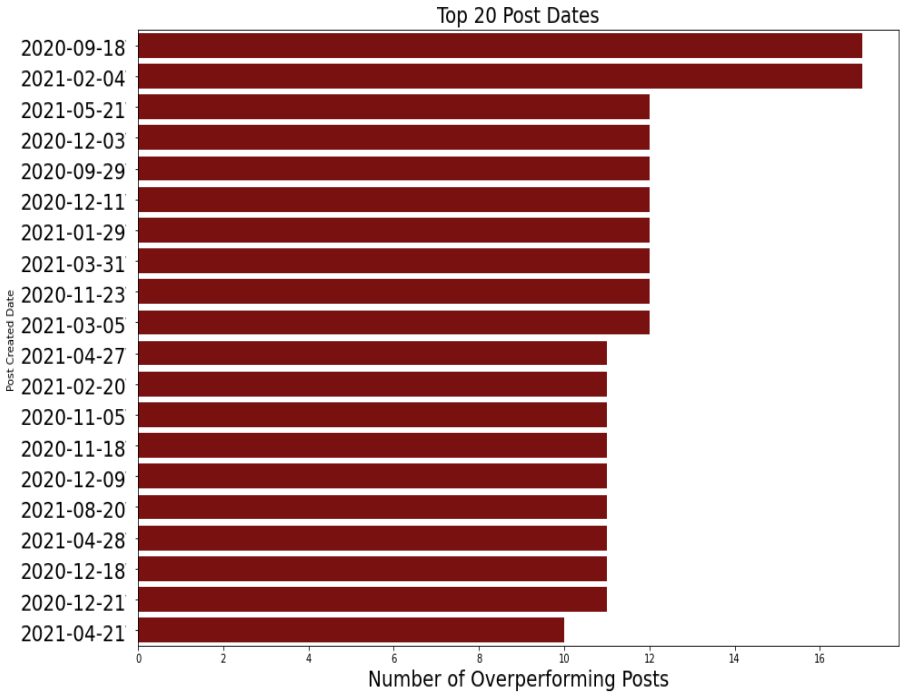


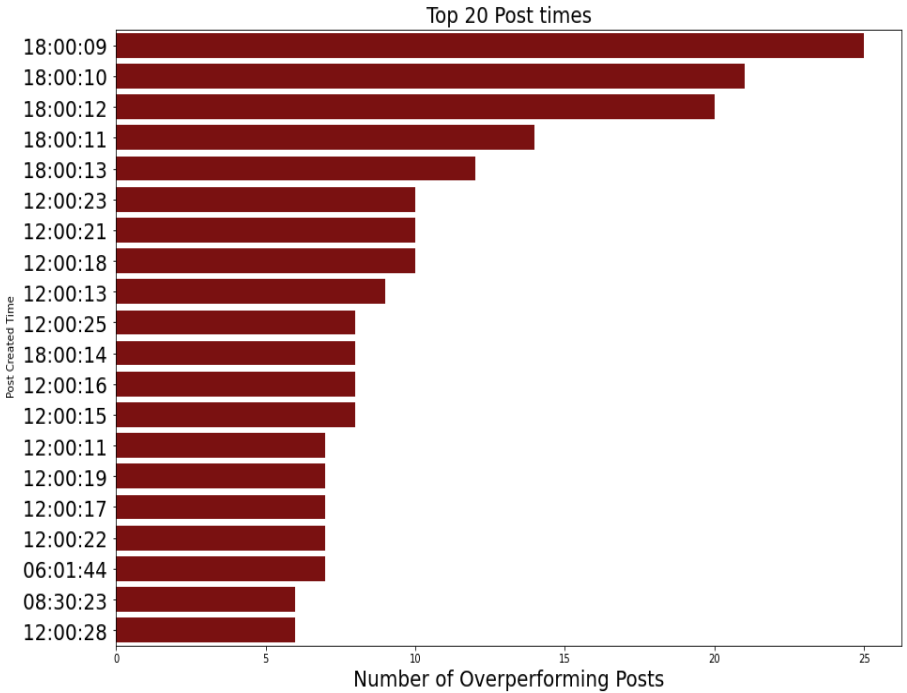


No clear distinction between phrases in posts that overperform or otherwise

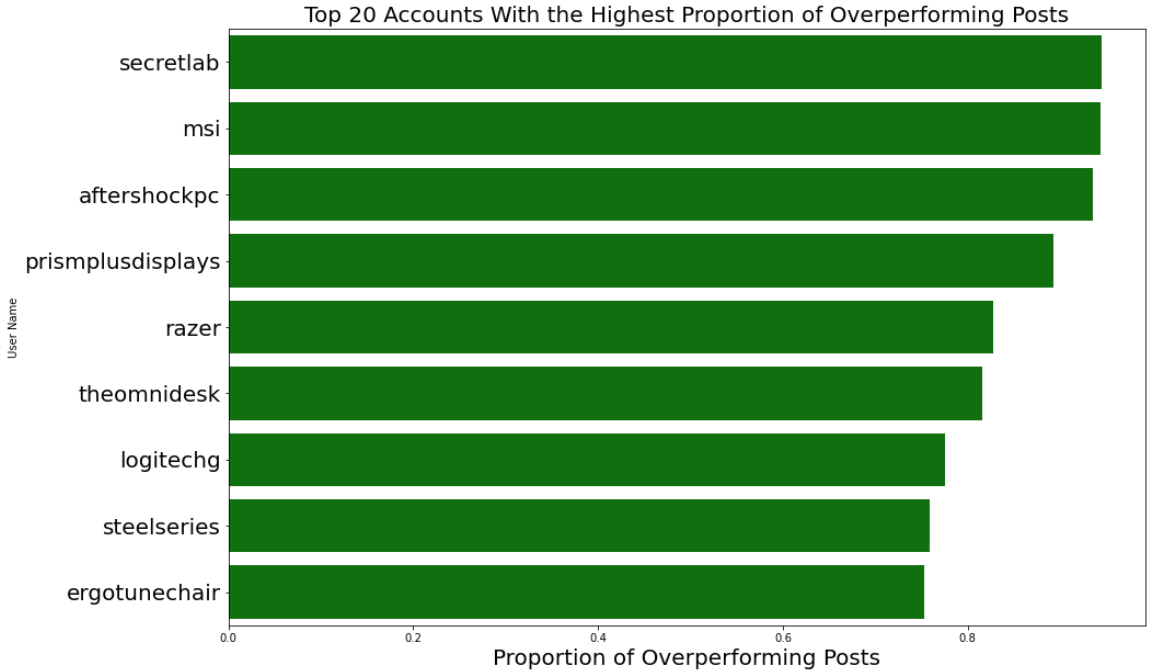
Likely due to distribution of dataset

DateTime


Top 20 Post Dates
2020-09-18
2021-02-04
2021-05-21
2020-12-03
2020-09-29
2020-12-11
2021-01-29
2021-03-31
2020-11-23
2021-03-05
2021-04-27
2021-02-20
2020-11-05
2020-11-18
2020-12-09
2021-08-20
2021-04-28
2020-12-18
2020-12-21
2021-04-21
Post Created Date
0
2
4
6
8
10
12
14
16
Number of Overperforming Posts

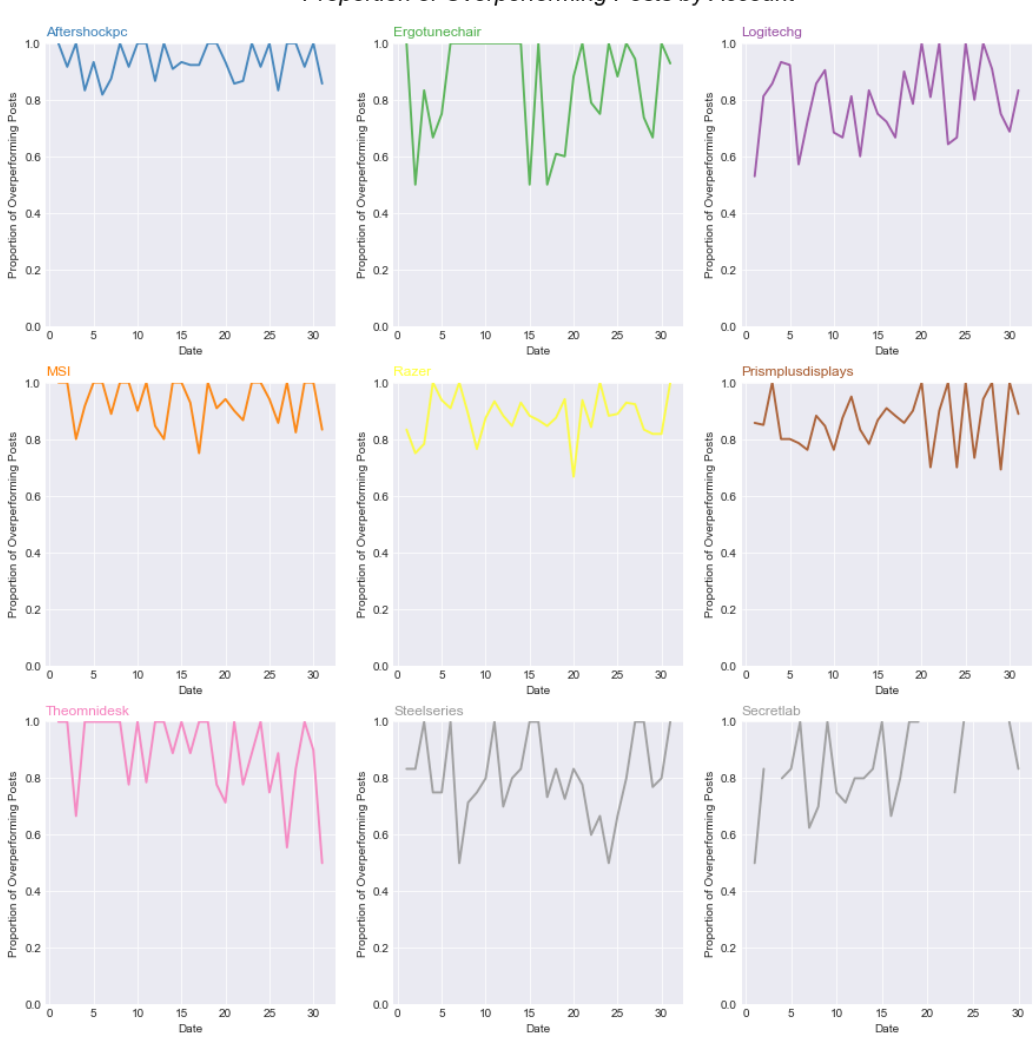

Top 20 Post times
18:00:09
18:00:10
18:00:12
18:00:11
18:00:13
12:00:23
12:00:21
12:00:18
12:00:13
12:00:25
18:00:14
12:00:16
12:00:15
12:00:11
12:00:19
12:00:17
12:00:22
06:01:44
08:30:23
12:00:28
Post Created Time
0
5
10
15
20
25
Number of Overperforming Posts

Accounts



Ergotune uses the most paid Instagram advertisements in this list

No discernible trends, highest fluctuations in ErgoTune, Steelseries, and Omnidesk



2.3

Modelling

Feature Engineering

Caption

Perform NLP

DateTime

Split into date and time

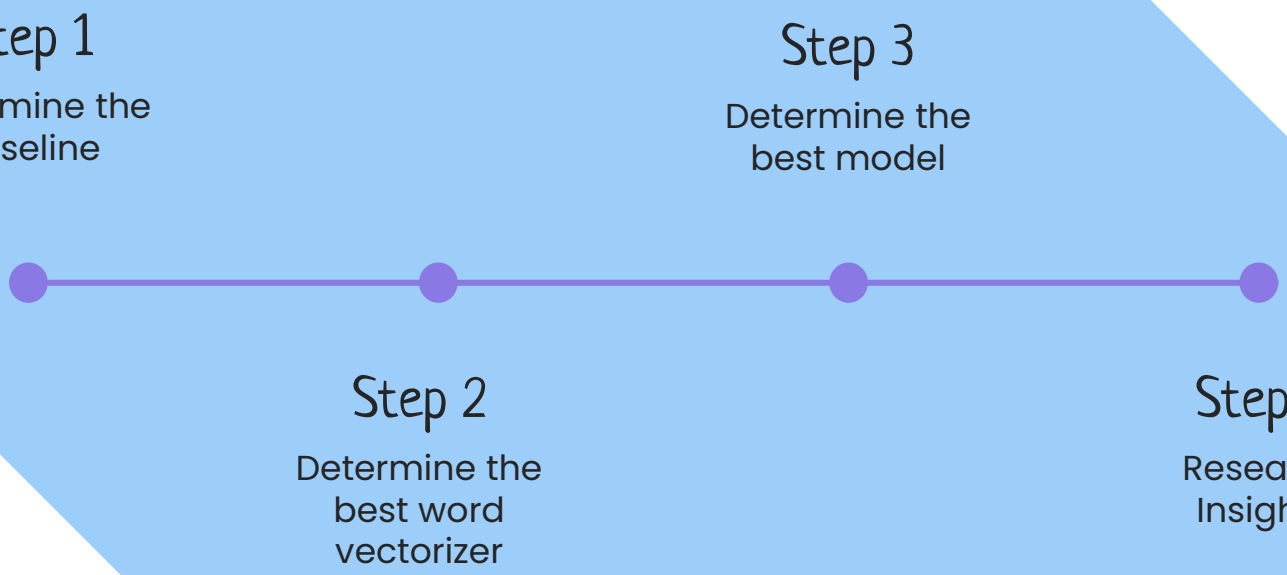
Account

Dummify

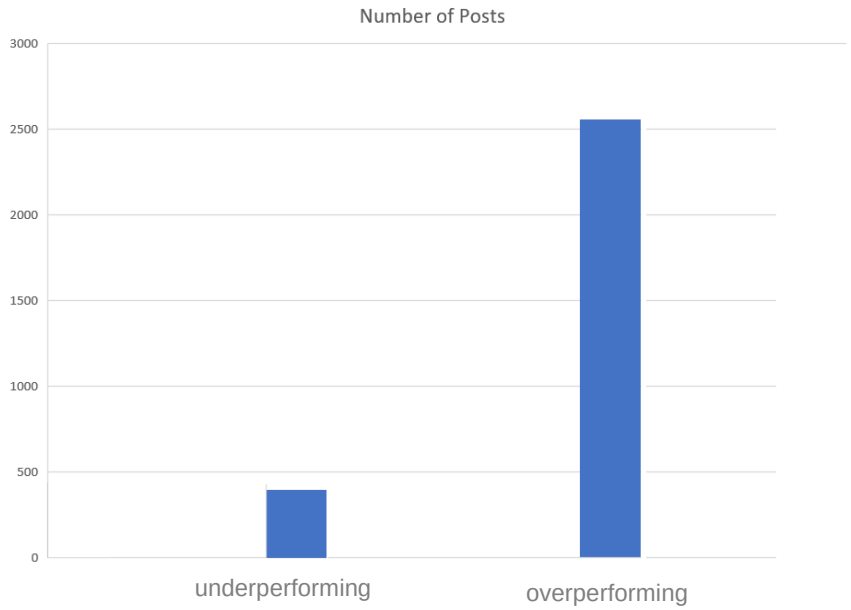
Overperforming

Our binary target variable

Modelling Workflow



Baseline



Baseline: 84.3%

2630

Overperforming posts

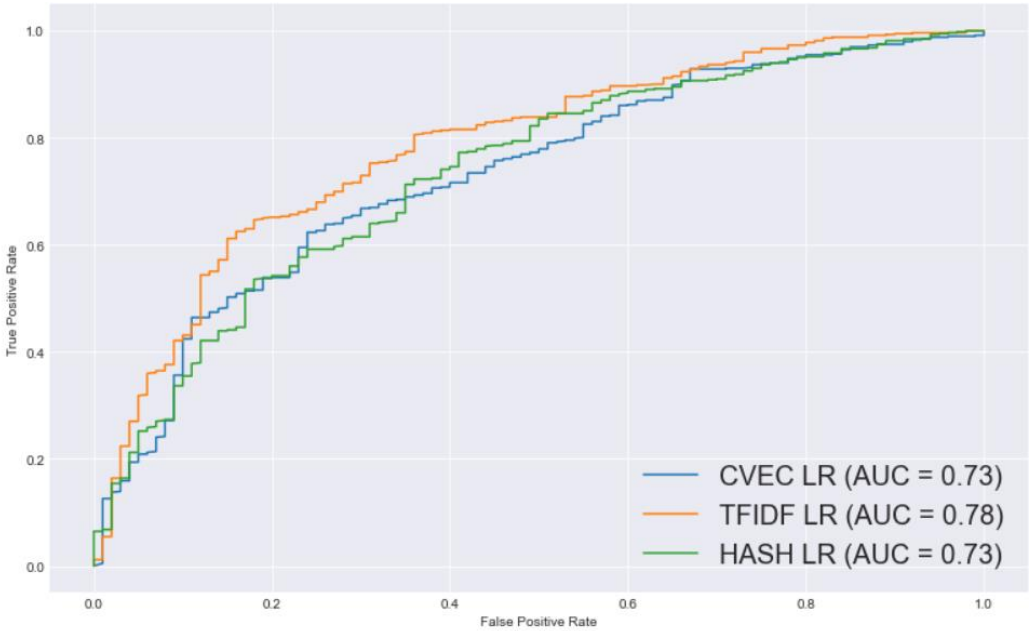
412

Underperforming posts

Determining the Best Vectorizer

	Train	Test	Spec.	Sens.
CV-LR	0.845	0.740	0.5	0.780
TFIDF-LR	0.859	0.782	0.63	0.808
Hash-LR	0.952	0.817	0.39	0.887

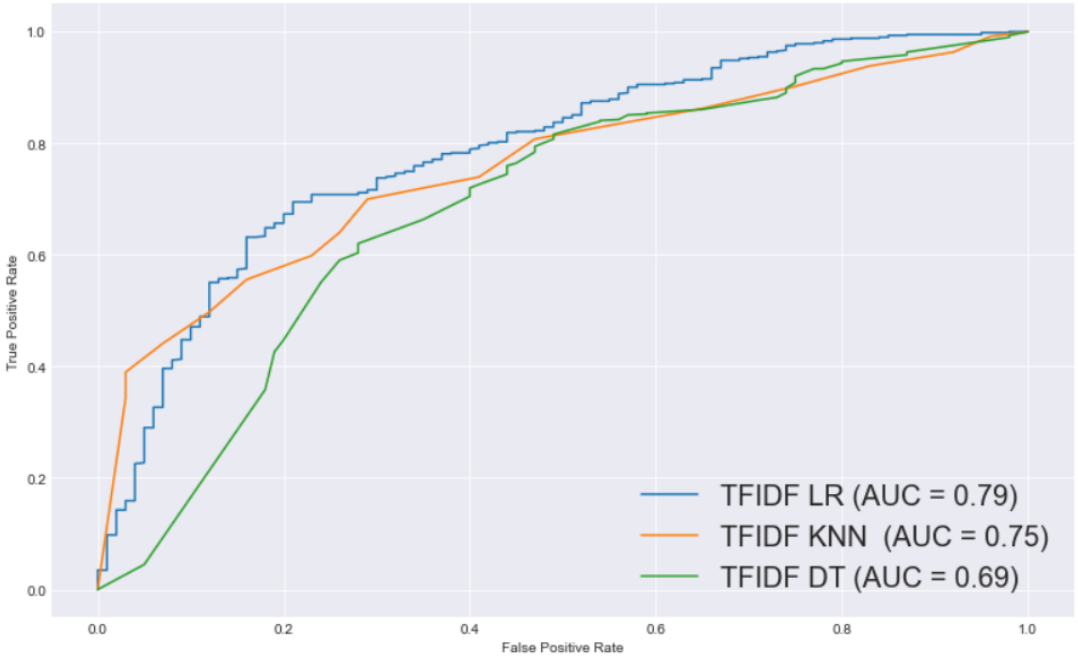
Determining the Best Vectorizer



Determining the Best Model

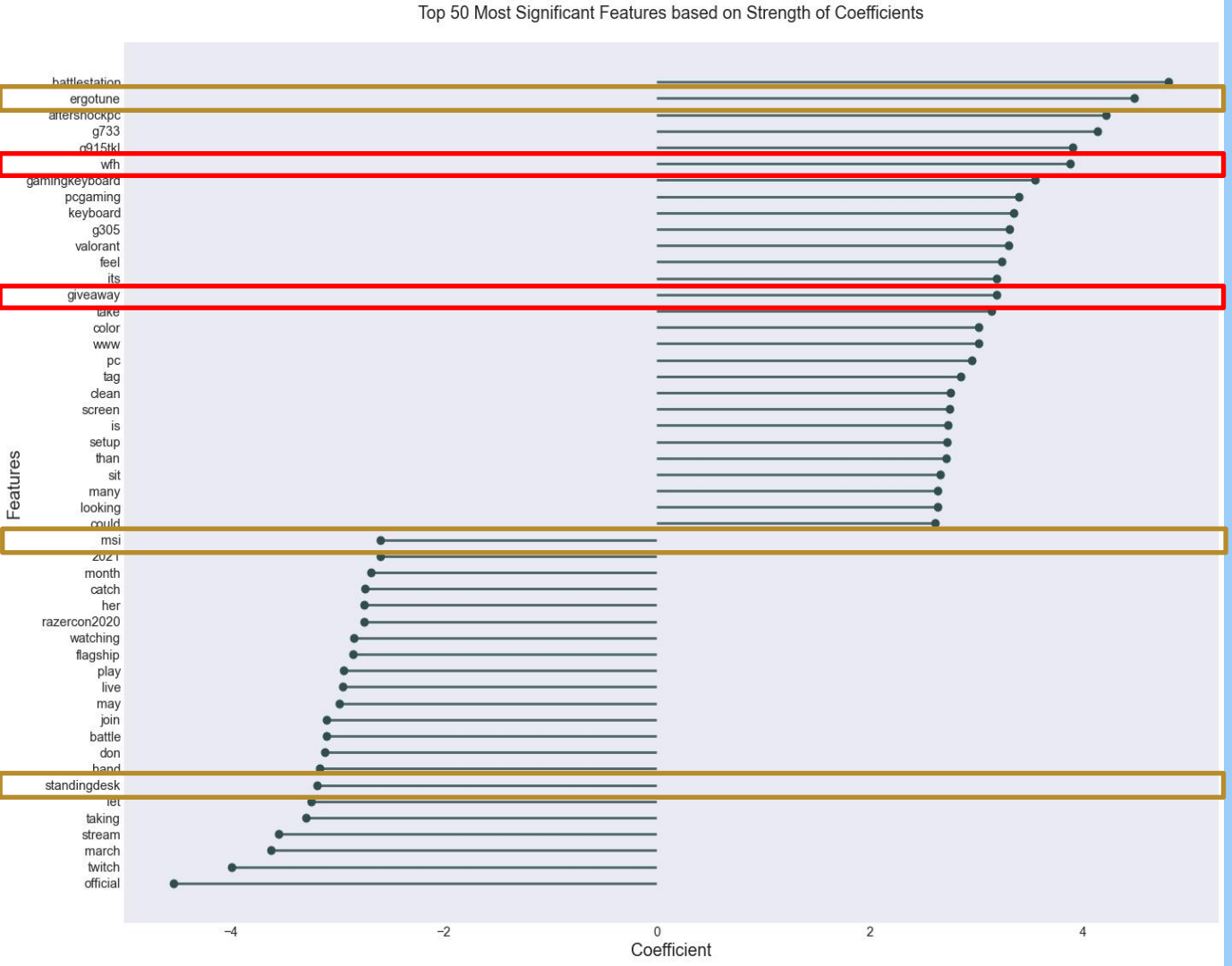
	Train	Test	Spec.	Sens.
TFIDF-LR	0.878	0.768	0.56	0.802
TFIDF-KNN	0.769	0.701	0.71	0.699
TFIDF-DT	0.859	0.785	0.46	0.839

Determining the Best Model



2.4

Findings



Accounts

Known Buzzwords

3

Image Data

3.1

The Data

What & Where

Posts since July 2021

From Aftershock, ErgoTune, Logitech, MSI, Prism+, Razer, Secretlab, Omnidesk, and SteelSeries

Selenium + Resnet50

Step 1

Seleniu
pretends t
me

Step 5

s of 9 predictions
e matched to like
number

Features

Prop_1, ... , Prop_9

Props within picture, lower number = higher prominence; str

Account

Username of posting account; str

Followers

Posting account's number of followers; int

Num_like

Number of likes; int

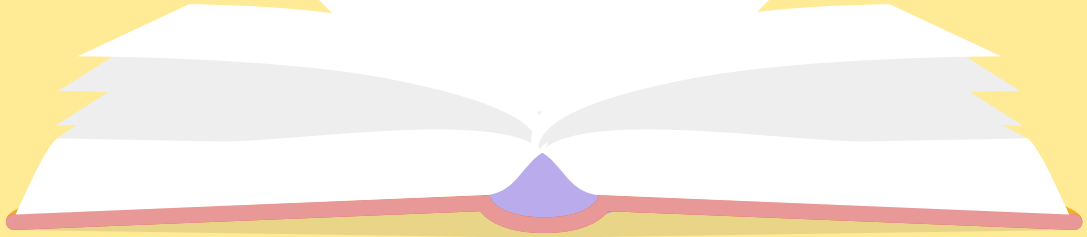
Performance

Num_like/Followers; int

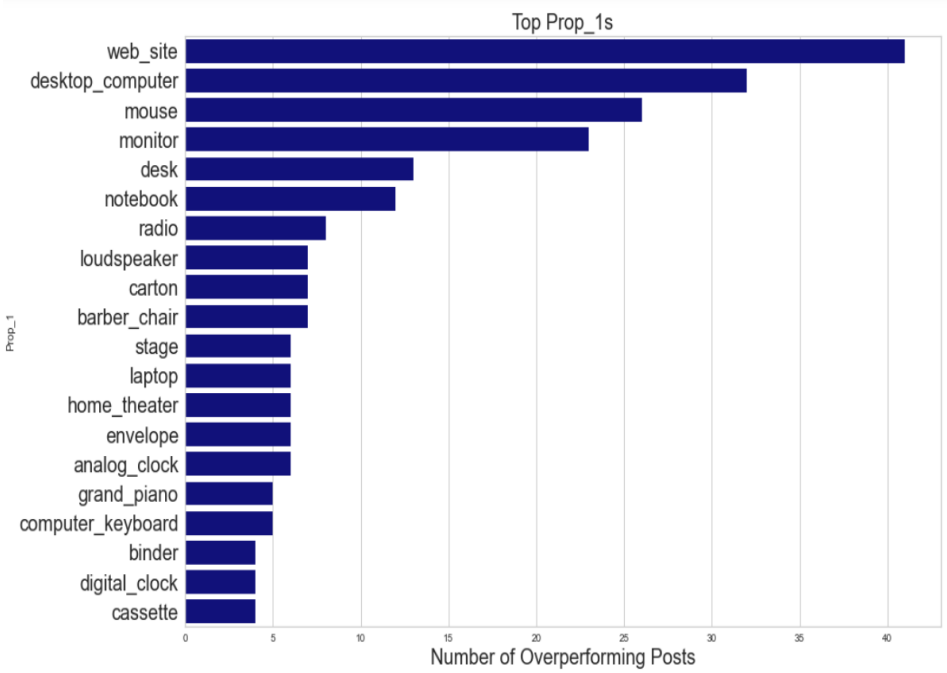
Overperforming

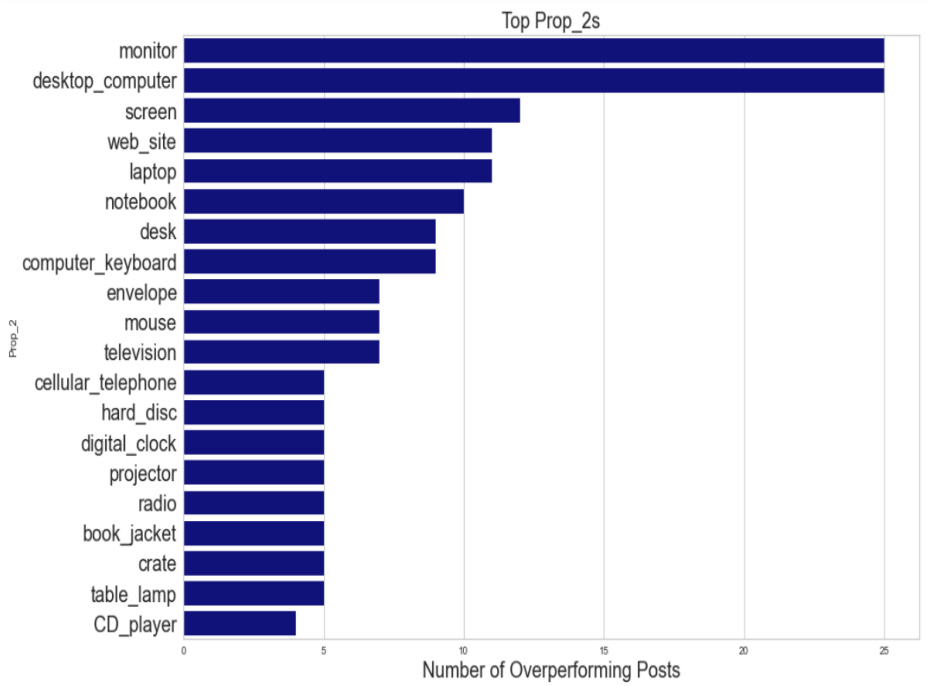
Binary feature based on cut-off performance score

3.2 EDA



EDA


Top Prop_1s
web_site
desktop_computer
mouse
monitor
desk
notebook
radio
loudspeaker
carton
barber_chair
stage
laptop
home_theater
envelope
analog_clock
grand_piano
computer_keyboard
binder
digital_clock
cassette
Prop_1
0
5
10
15
20
25
30
35
40
Number of Overperforming Posts


Top Prop_2s
monitor
desktop_computer
screen
web_site
laptop
notebook
desk
computer_keyboard
envelope
mouse
television
cellular_telephone
hard_disc
digital_clock
projector
radio
book_jacket
crate
table_lamp
CD_player
Prop_2
0
5
10
15
20
25
Number of Overperforming Posts

3.3

Modelling

Modelling Workflow

Step 1

Determine the Baseline

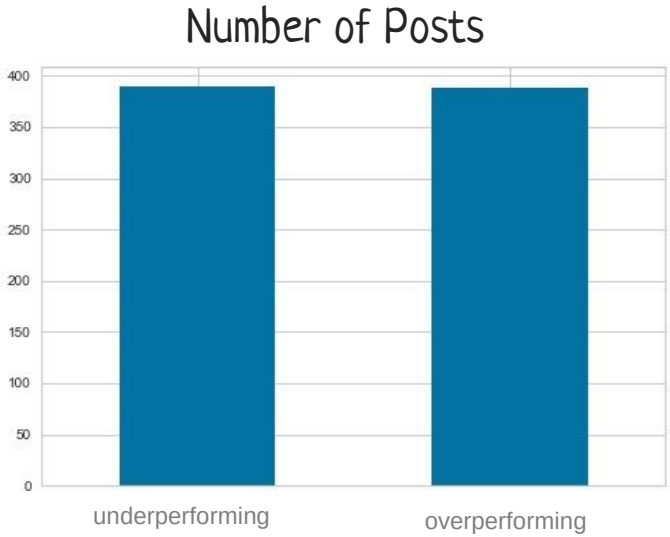
Step 2

!!!!Pycaret!!!!

Step 3

Research Insights

KPI Overview



Baseline: 50%

338

Overperforming Posts

339

Underperforming Posts

XgBoost

0.529

Accuracy

0.532

AUC

0.613

Recall

0.526

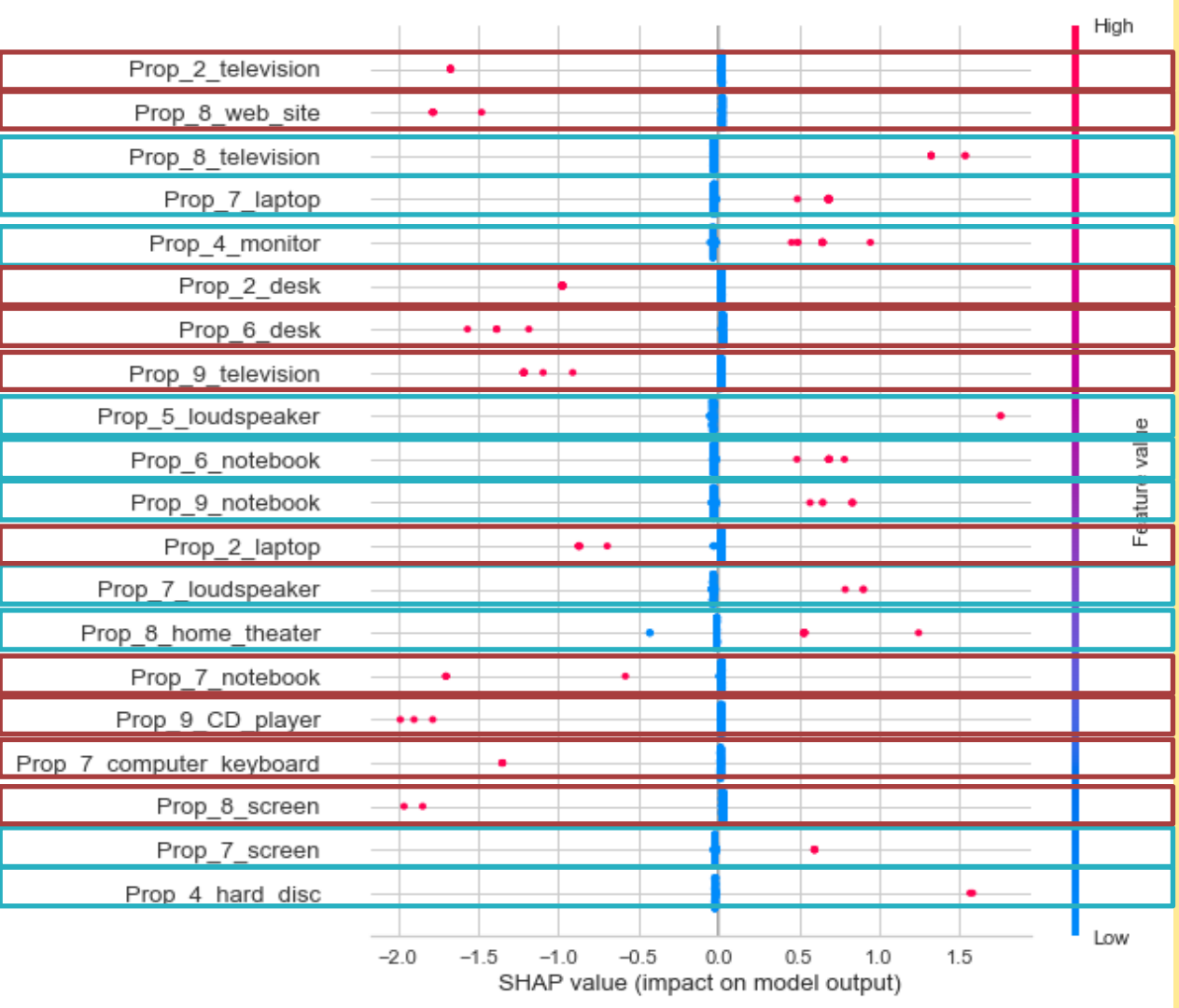
Precision

0.563

F1

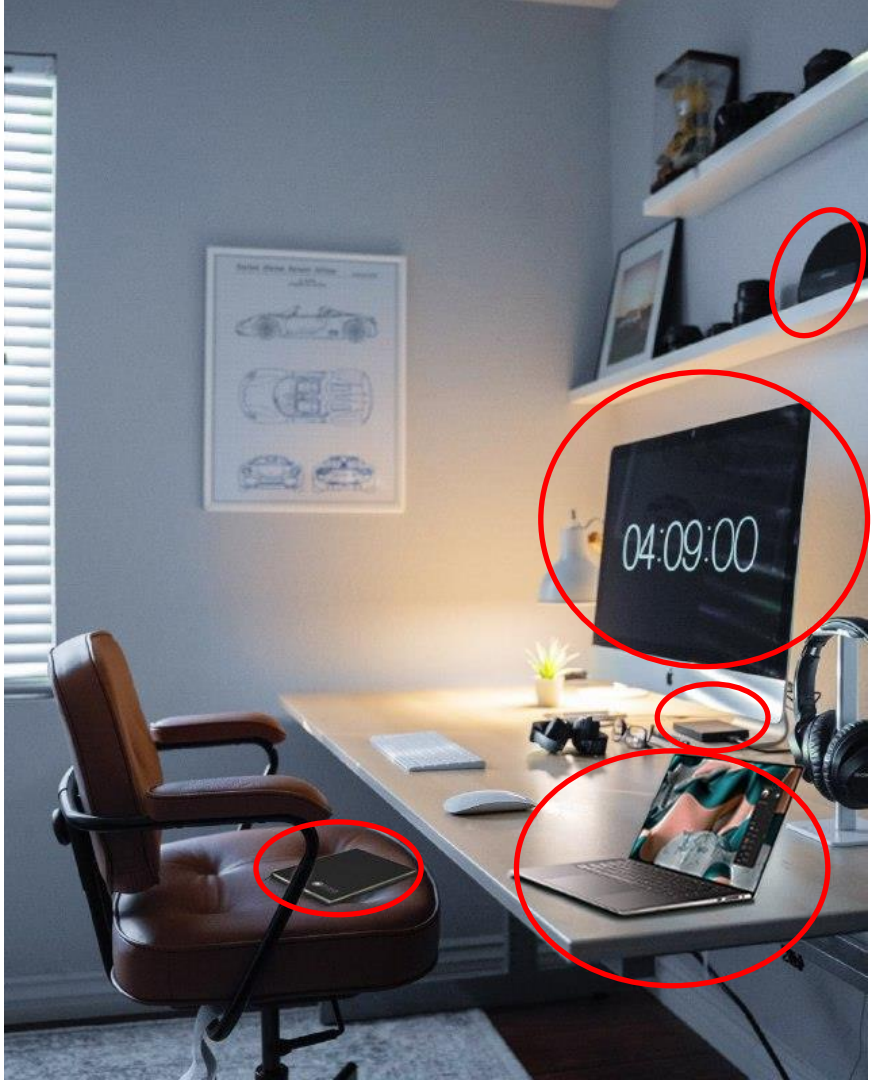
3.4

Findings

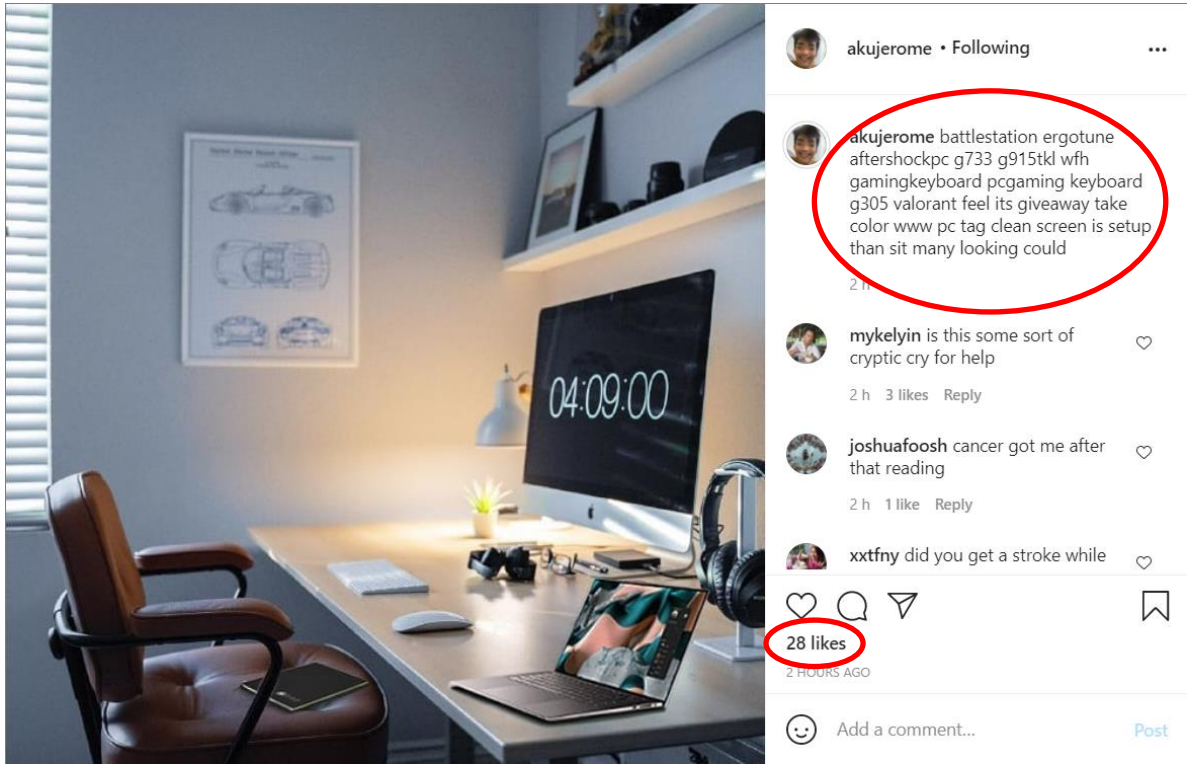


Do Include

Do Not Include


04:09:00

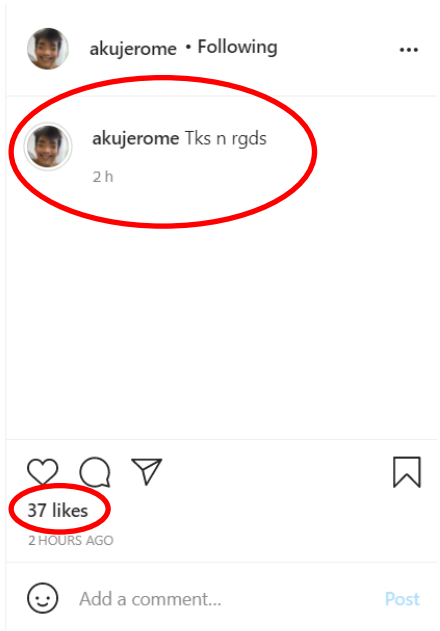
Combining non-image data and image data



Combining non-image data and image data

Hey you

- Can help me like this post? It's for some project
- Thanks



Limitations

1. Non-Image Data: The Dataset being extremely skewed toward overperforming posts forced us to resort to oversampling

2. Image: By accounting for prop prominence, our findings ended up convoluted and, at times, contradictory

Recommendations

1. Non-Image Data: Search for and include companies that have a balanced over-underperforming ratio

2. Image: ignore prop prominence

3. After making the above 2 changes, move toward a recommendation engine

Industry Use

Not a replacement for Creatives, but a tool to aid their creative process

Thanks!

Do you have any questions?
- Ask Kishan