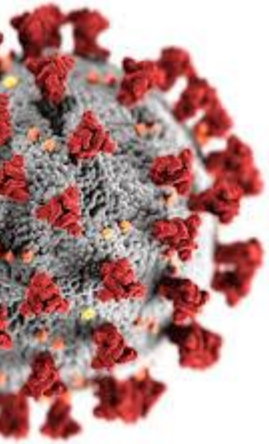
Tourists at Home

Exploring the contours of Domestic Tourism and to provide insights into how STB can market and position themselves to build an active and robust Domestic Tourism market.



S$ 365, 000, 000

As representatives of STB, we want to identify focus areas of high potential for the next phase of the #SingapoRediscovers campaign, and in turn propose targeted initiatives that can build up a resilient Domestic Tourism economy.

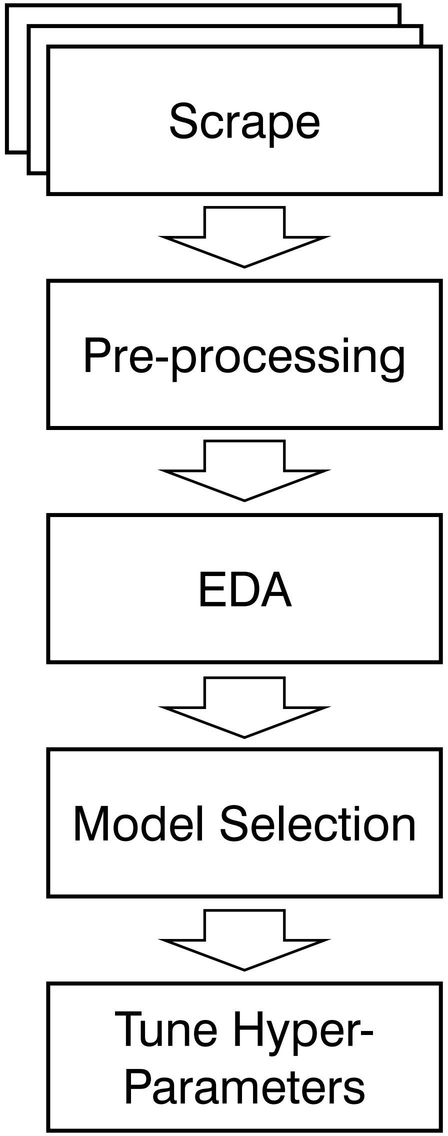
It is also this project's belief that investing in Domestic Tourism today can only add value to Singapore Tourism when travel finally resumes.

Problem Statement

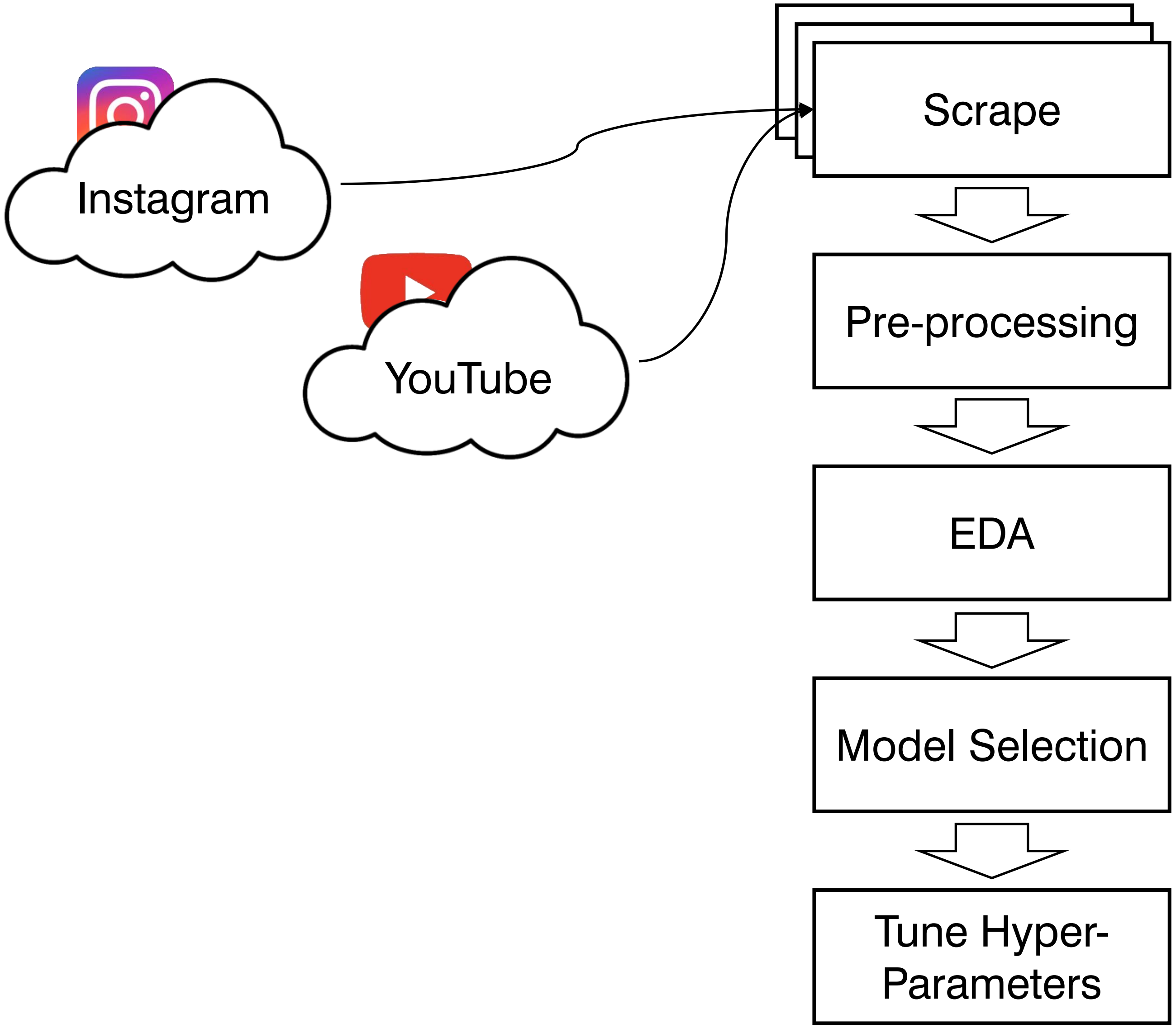
Singapore TOURISM BOARD

Process

Overview of Data Science process.

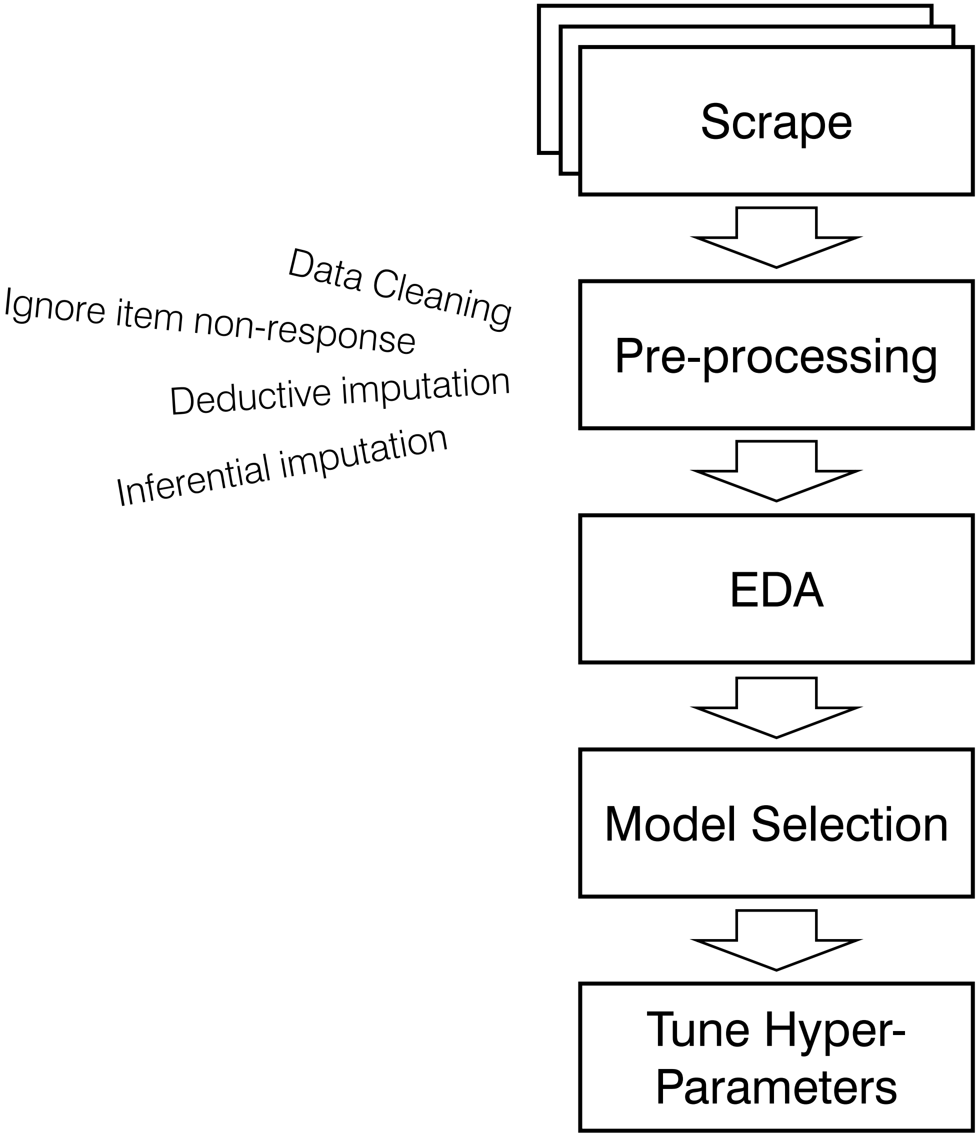


Methodology

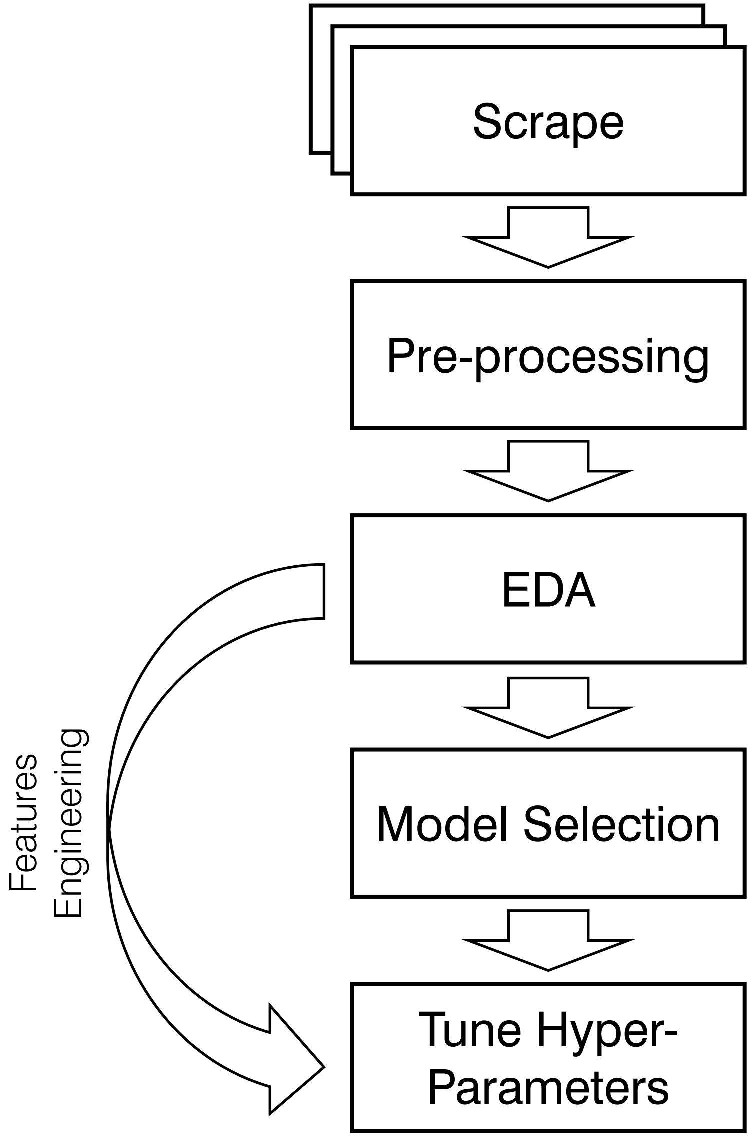


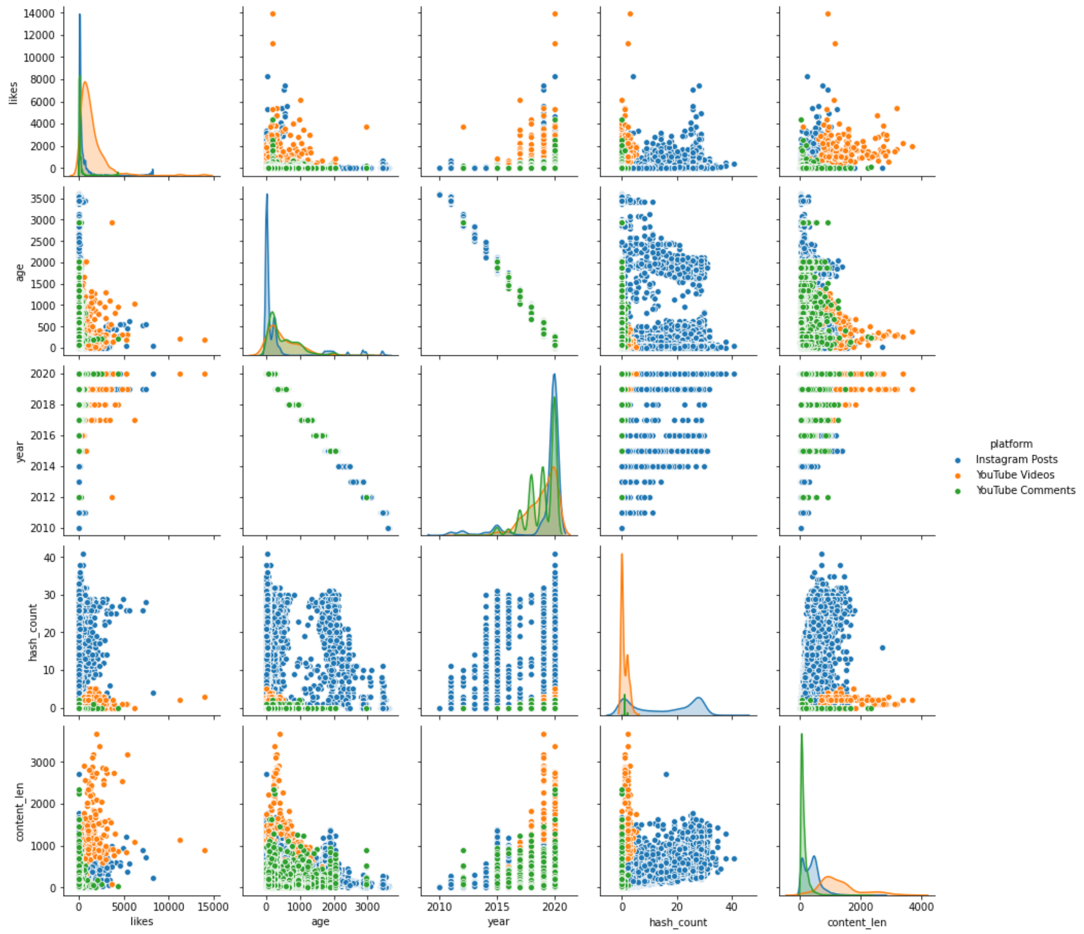
Methodology

Singapore
TOURISM
BOARD

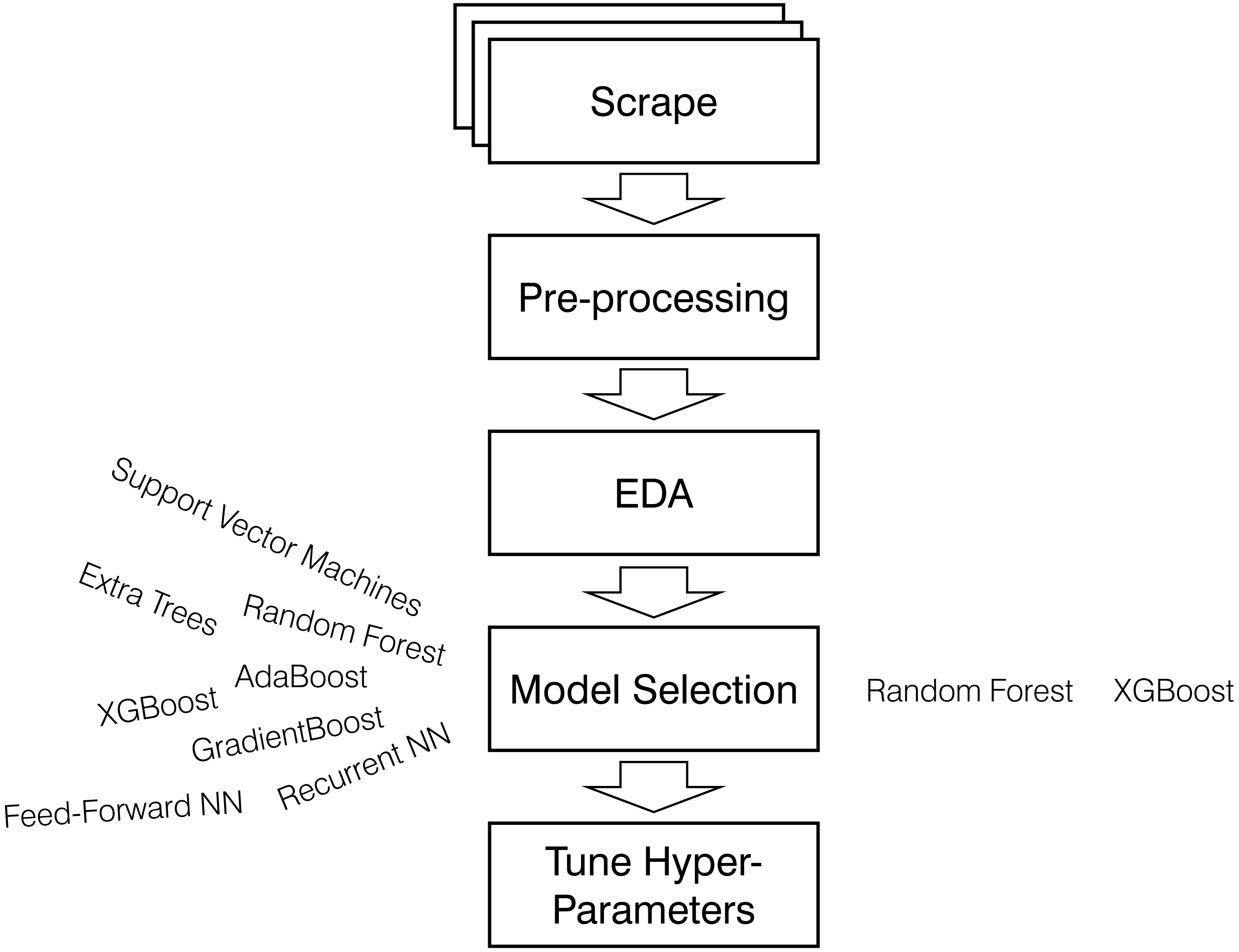


Methodology

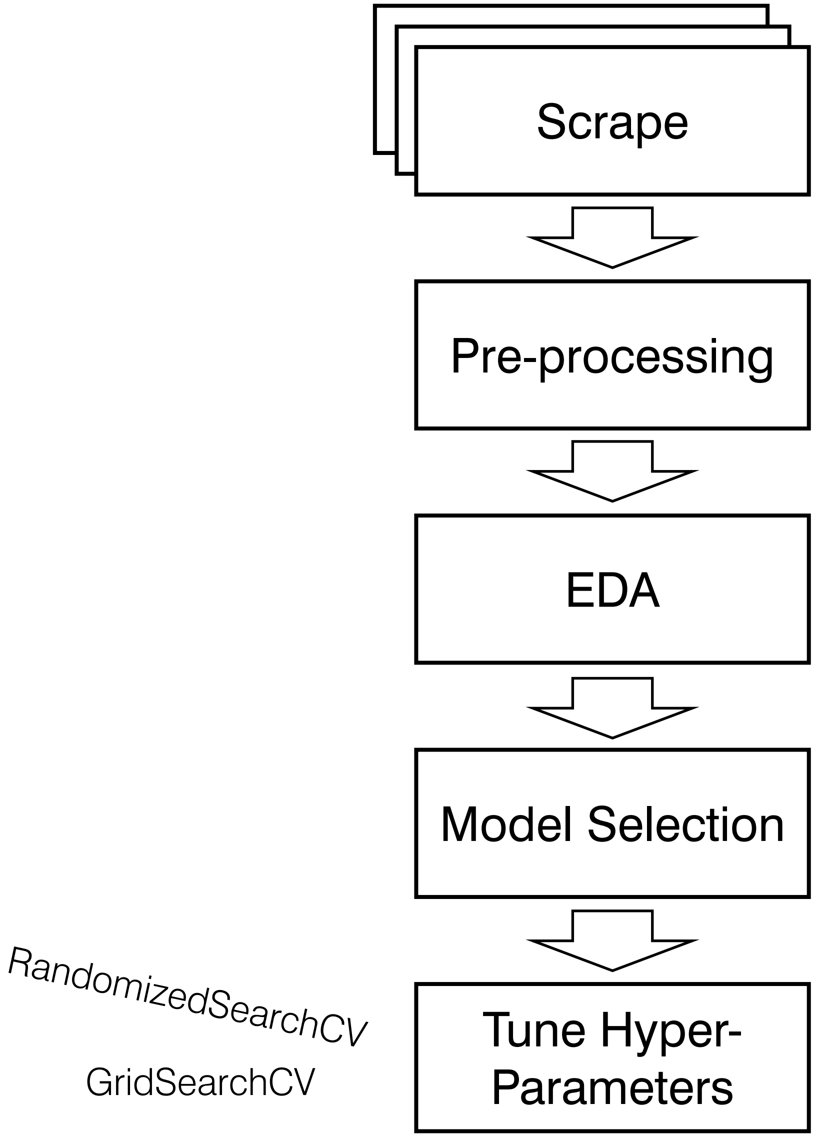


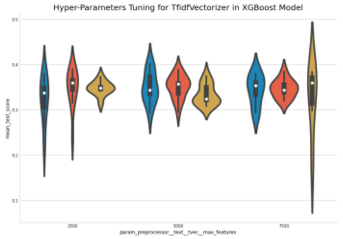


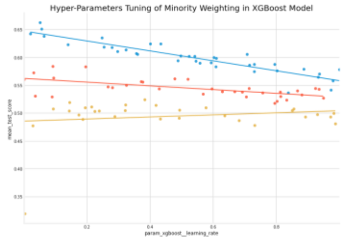
Methodology

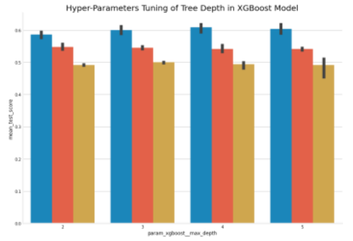


Methodology

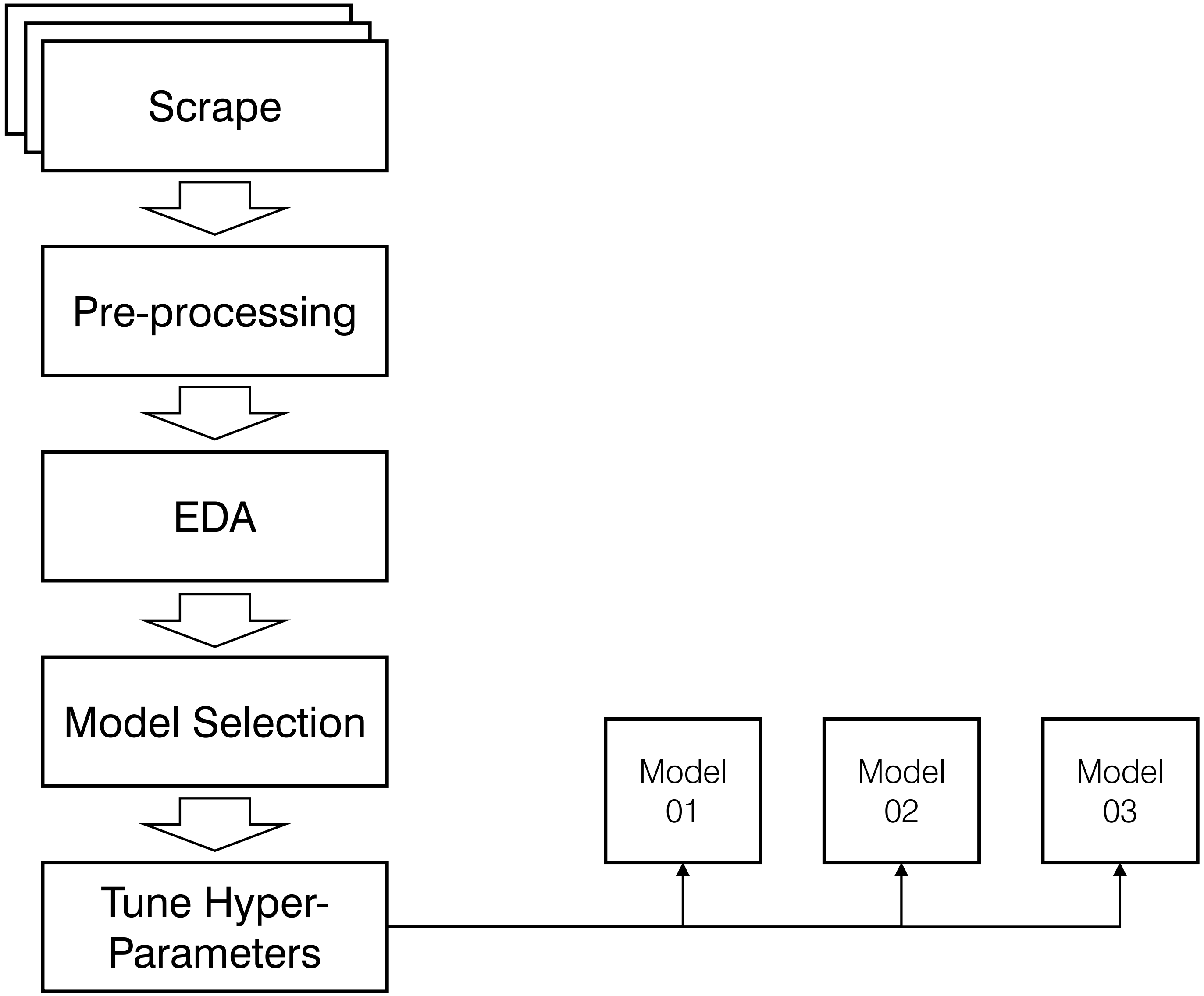




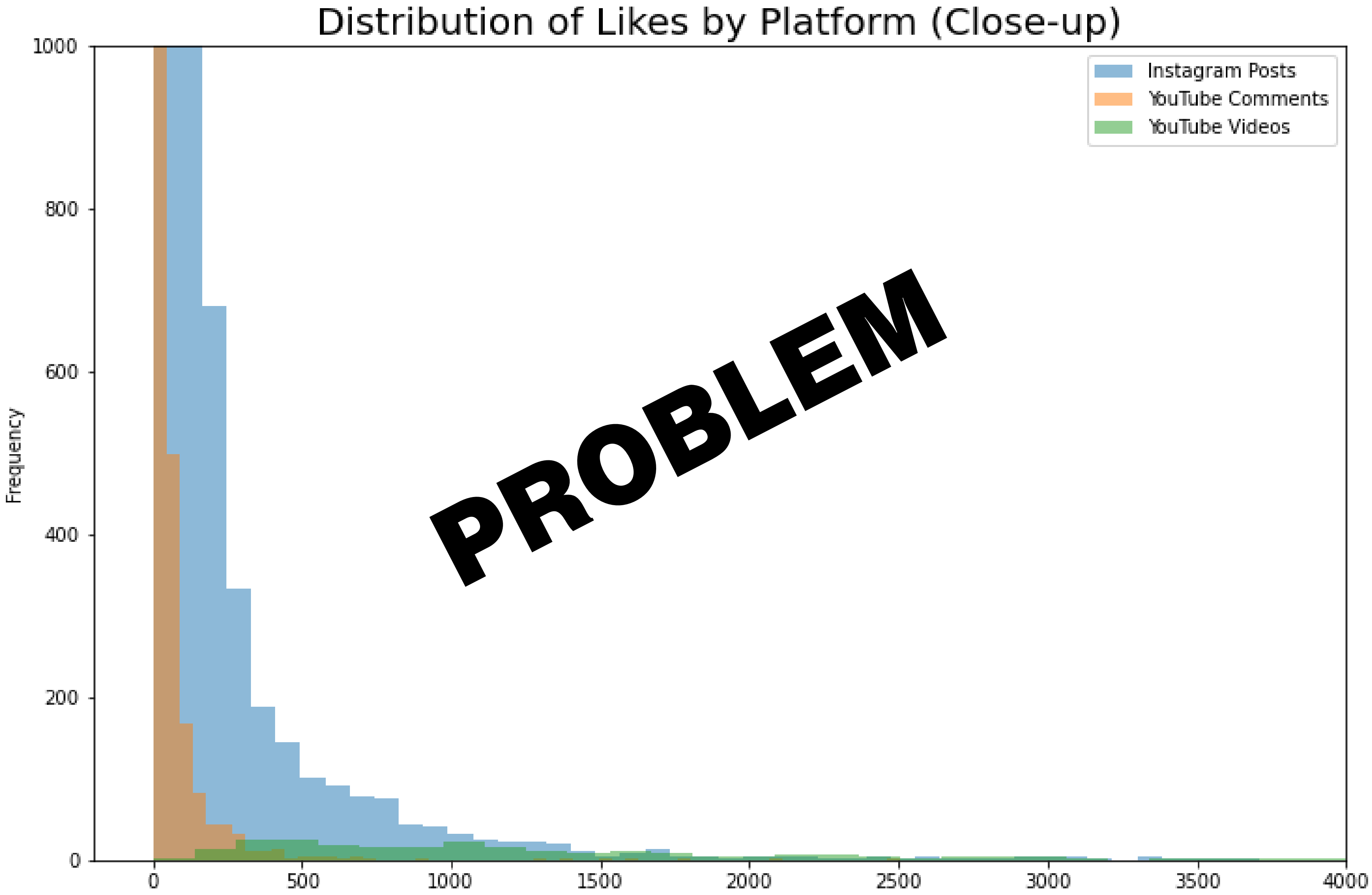




Methodology



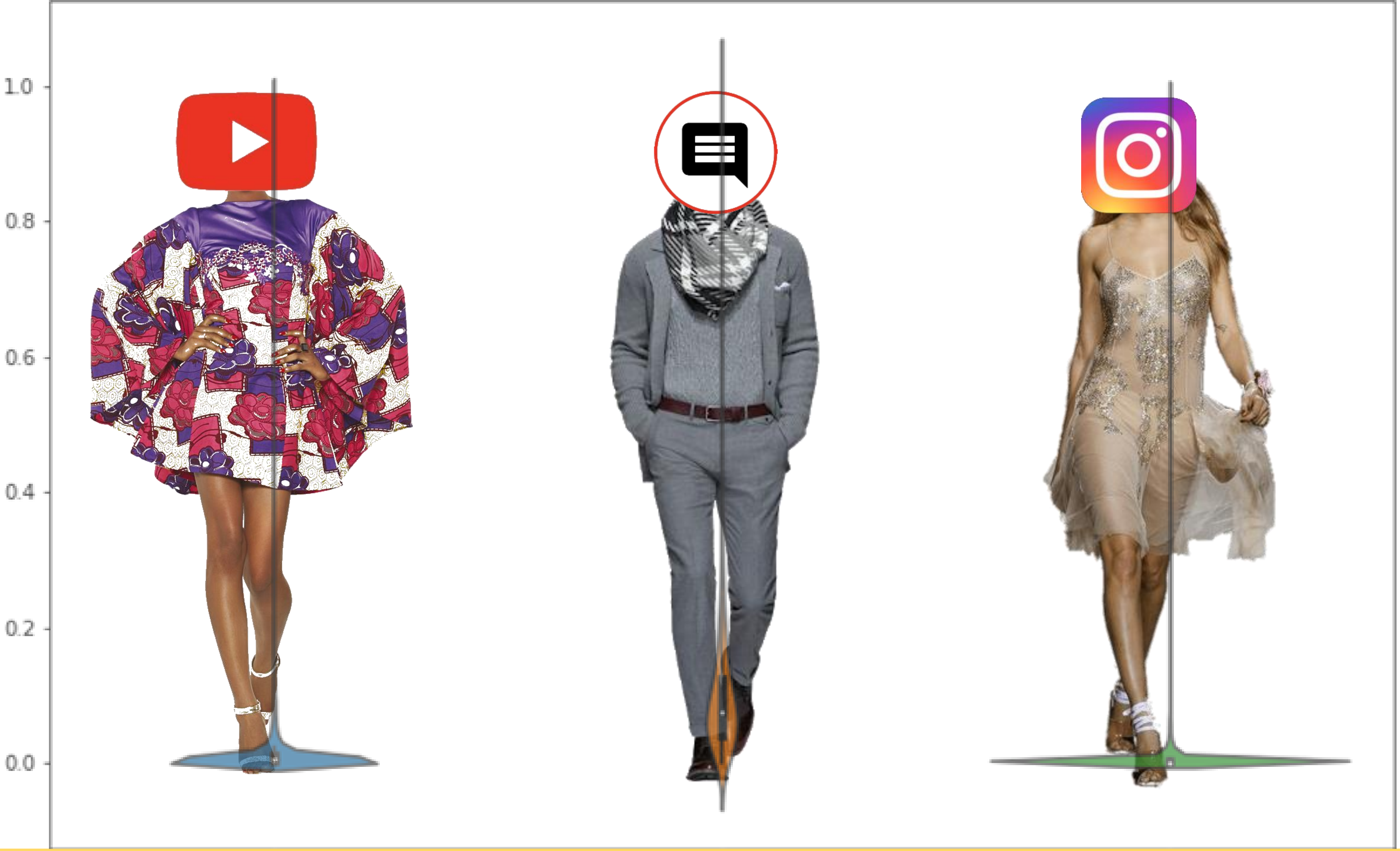
Methodology



I Like, You Like?

Platform Dummies

Singapore
TOURISM
BOARD



Rescale Target Variables



Separate Models

CLASSIFICATION

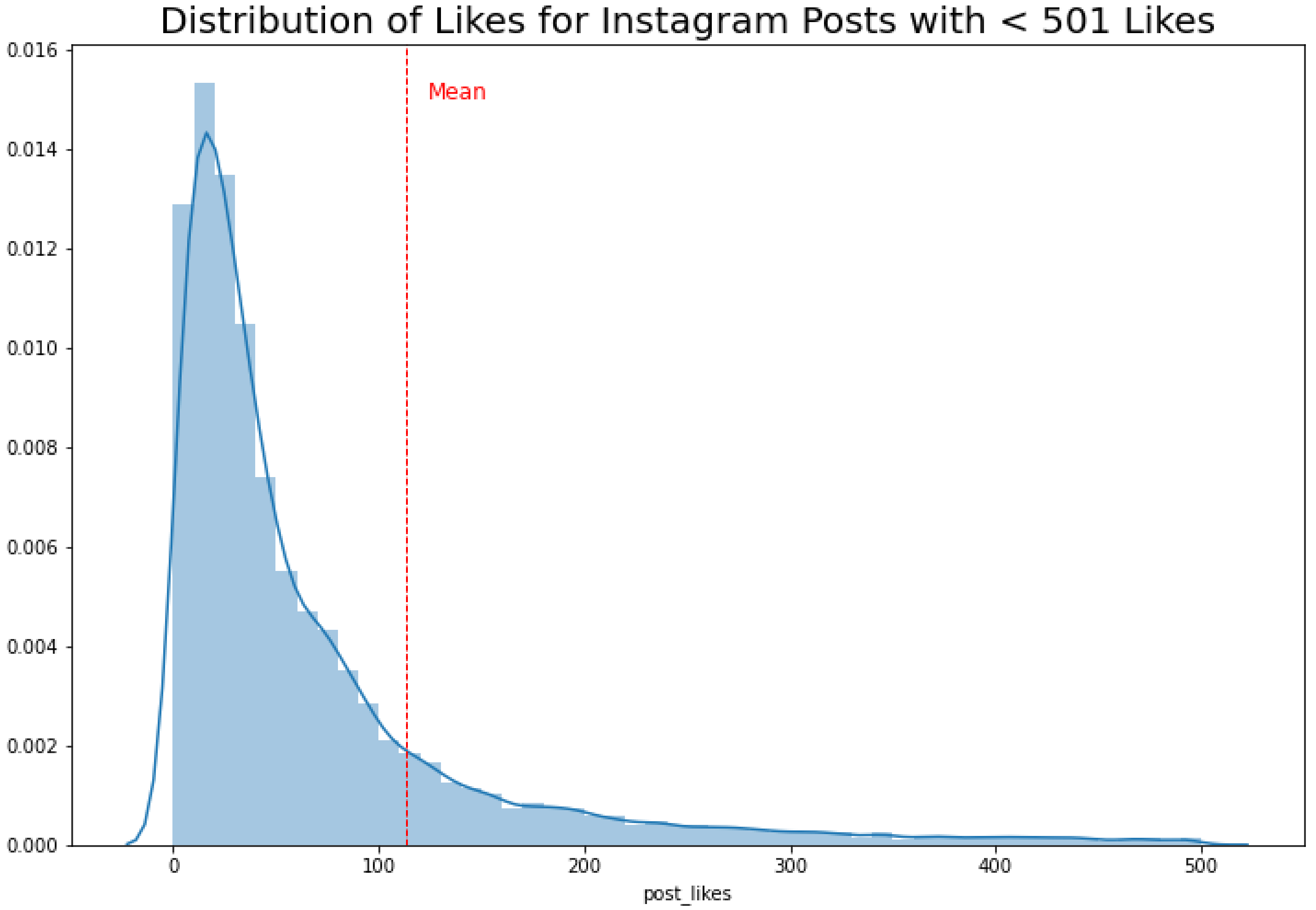
Create Target Labels

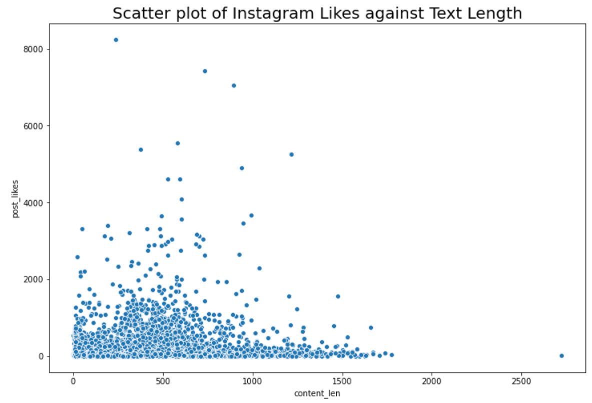
Singapore TOURISM BOARD

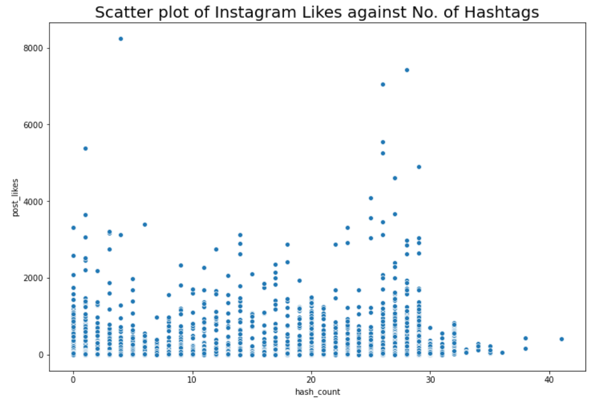
Exploratory Data Analysis

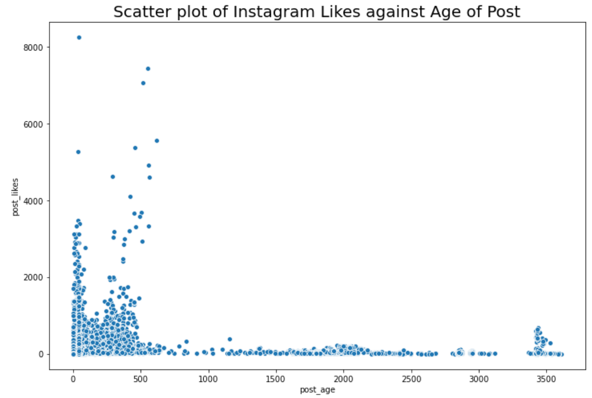
Preliminary analysis of data sets.

Singapore TOURISM BOARD

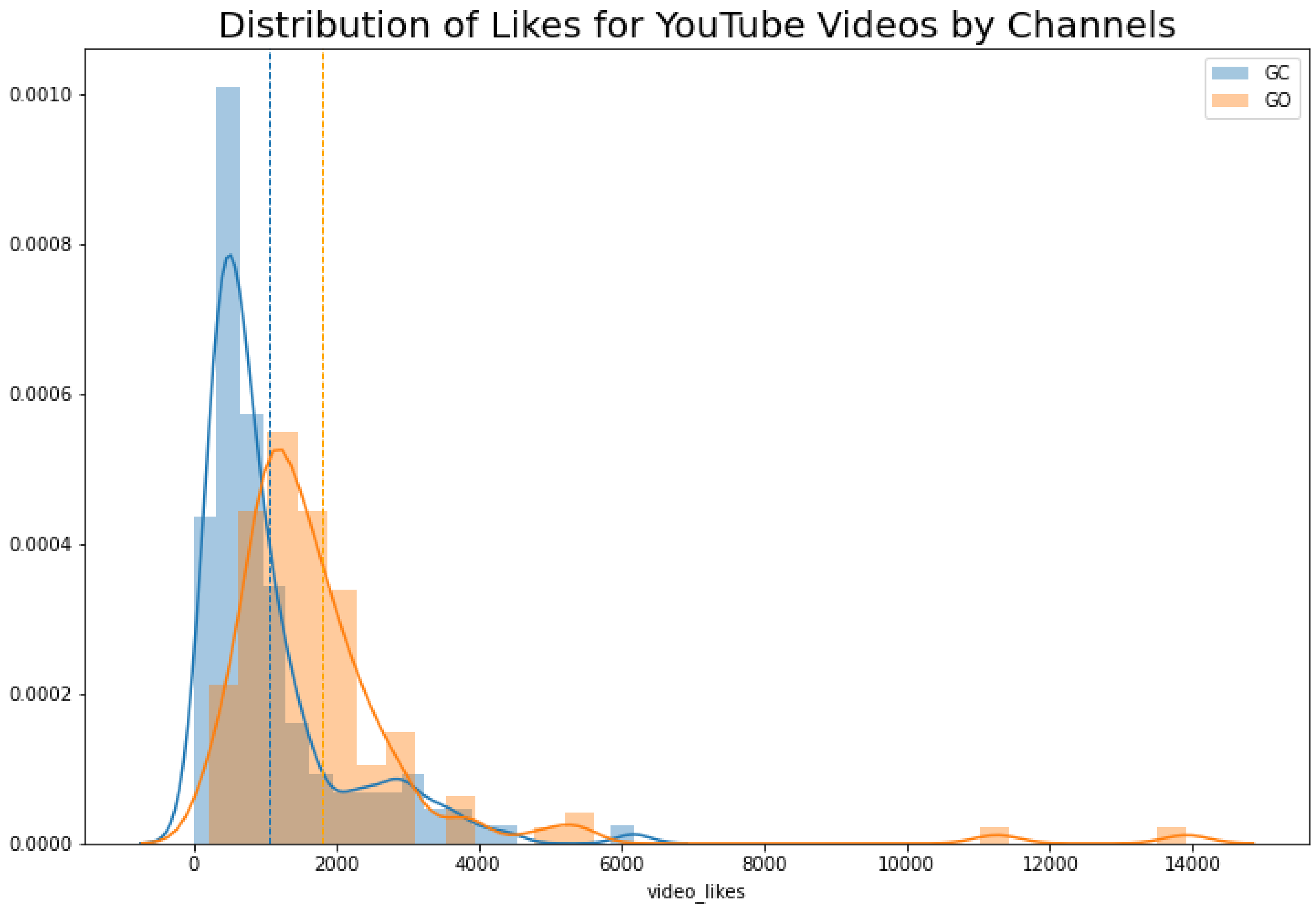


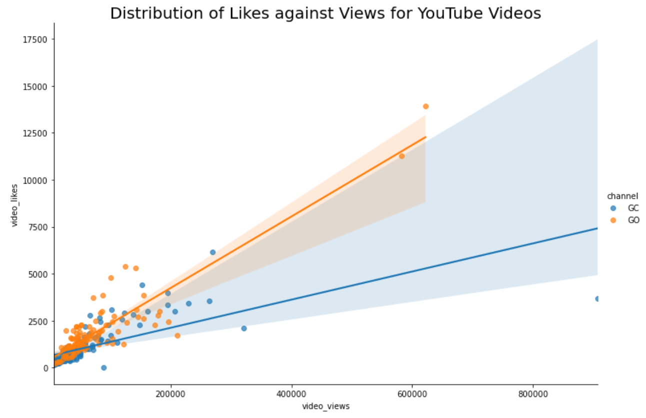


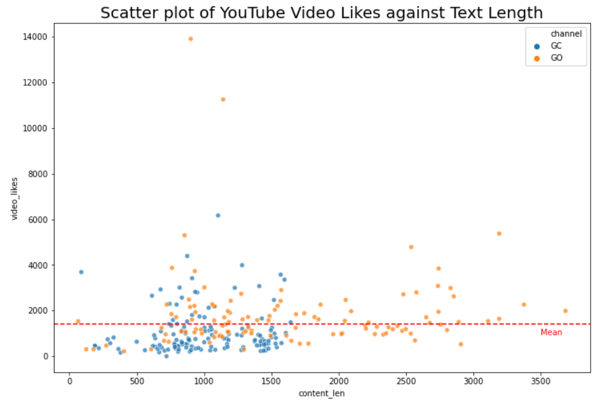




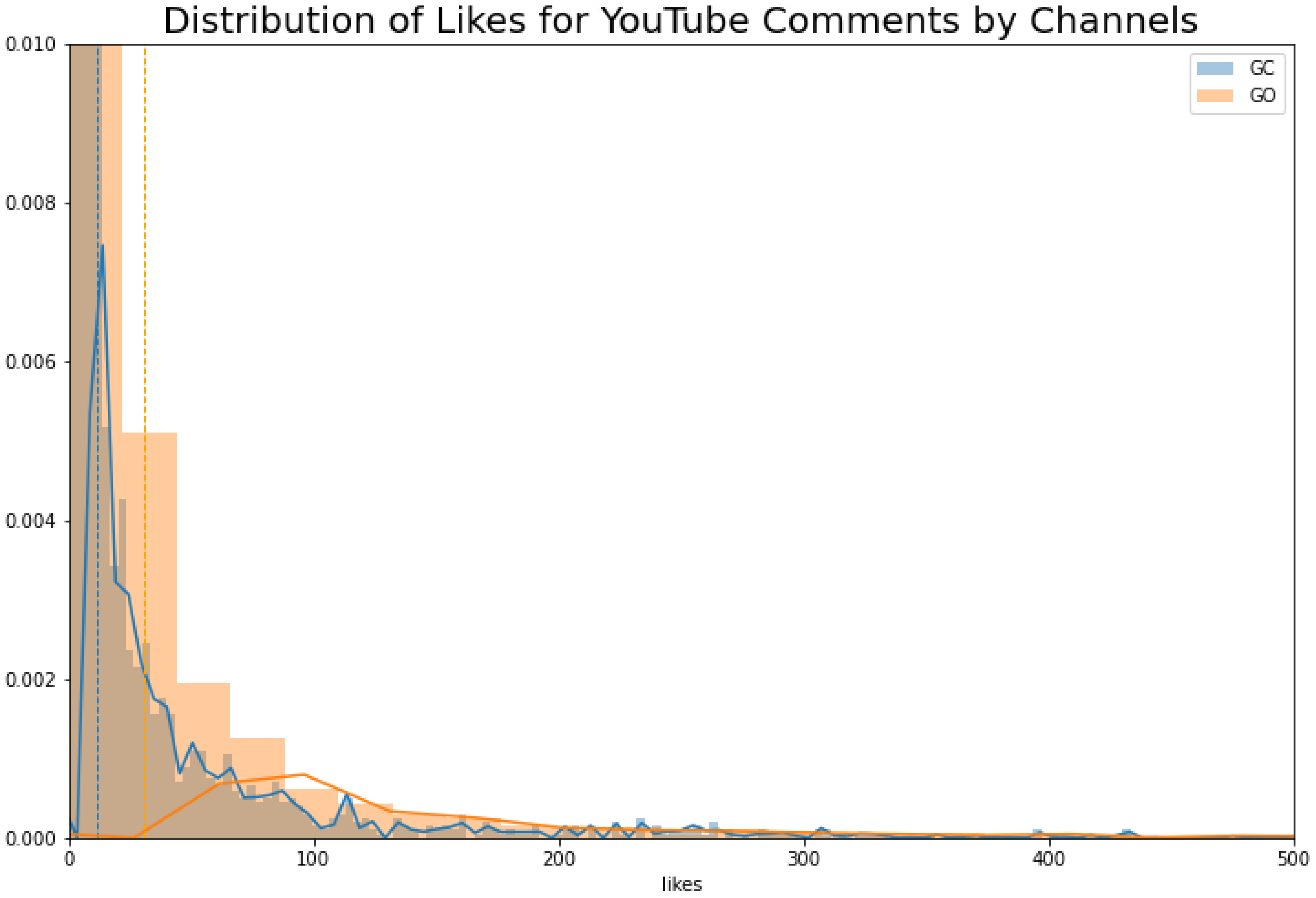
EDA – Instagram Posts

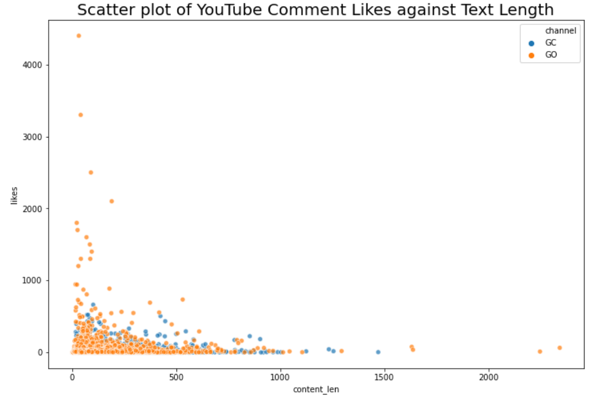


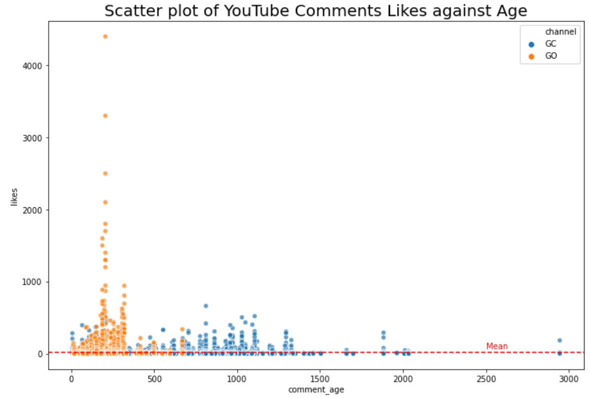




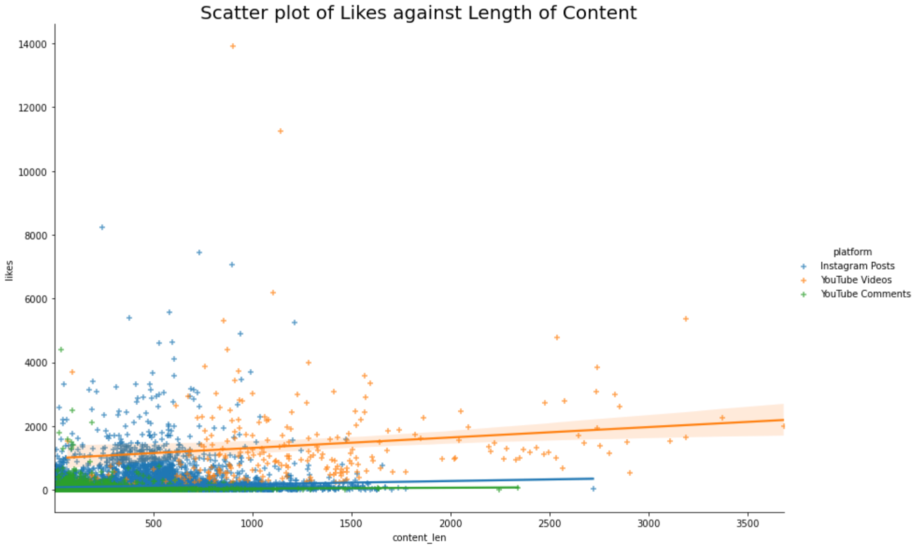
EDA – YouTube Videos

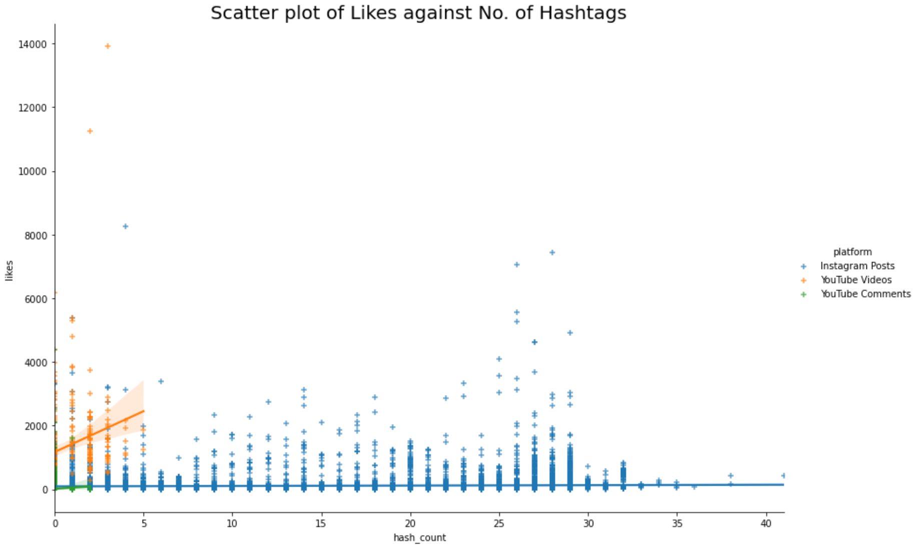


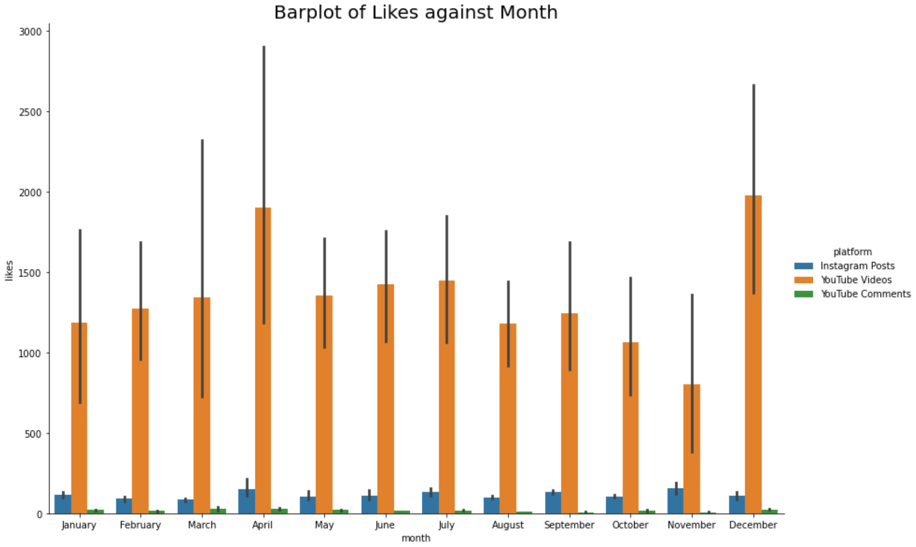


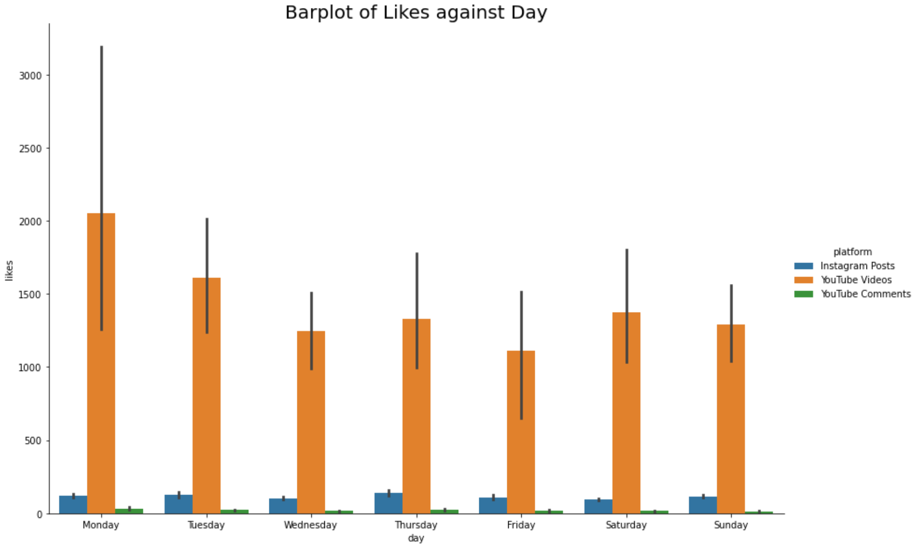


EDA – YouTube Comments


Scatter plot of Likes against Length of Content
likes
content_len
platform
Instagram Posts
YouTube Videos
YouTube Comments
0
2000
4000
6000
8000
10000
12000
14000
500
1000
1500
2000
2500
3000
3500


Scatter plot of Likes against No. of Hashtags
likes
hash_count
platform
Instagram Posts
YouTube Videos
YouTube Comments
0
2000
4000
6000
8000
10000
12000
14000
0
5
10
15
20
25
30
35
40


Barplot of Likes against Month
likes
month
platform
Instagram Posts
YouTube Videos
YouTube Comments
0
500
1000
1500
2000
2500
3000
January
February
March
April
May
June
July
August
September
October
November
December


Barplot of Likes against Day
likes
day
platform
Instagram Posts
YouTube Videos
YouTube Comments
0
500
1000
1500
2000
2500
3000
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday

EDA – Overall

Modelling

Building and evaluating models.

Model	Cross-Validation Score (R^2)	Mean Squared Error	Validation Score Gap	Training Time (s)	Training Time (hh-mm-ss)
Ridge Regressor*	0.387	58557.4	-	70	1min 10s
Support Vector Regressor	-0.035	-	-	3022	50min 22s
Random Forest Regressor	0.401	-	-	5305	1h 28min 25s
Extra Trees Regressor	0.367	-	-	10306	2h 51min 46s
Ada-Boost Regressor	-7.124	-	-	2215	36min 55s
Gradient Boosting Regressor	0.363	-	-	587	9min 47s
XGBoost (gbtree, squared error) Regressor	0.303	-	-	158	2min 38s
XGBoost (gbtree, tweedie) Regressor	0.381	-	-	180	3min
XGBoost (gblinear, squared error) Regressor	0.398	-	-	41	41s
XGBoost (gblinear, tweedie) Regressor	0.317	-	-	36	36s
Feed-Forward NN	-	61421.0	18.1	28	28s
Feed-Forward NN (Ridge)	-	61419.7	596.5	27	27s
Feed-Forward NN (Lasso)	-	61259.4	541.3	27	27s
Recurrent NN (GRU)	-	101920.1	64157.1	1362	22min 42s
Recurrent NN (LSTM)	-	90009.7	33145.1	1270	21min 10s

Model Selection

Model	Optimised on	ROC-AUC	Accuracy	Recall	Specificity	Precision	F-1 Score
baseline*	-	-	0.709	-	-	-	-
Hybrid_01	ROC-AUC	0.727	0.734	0.711	0.743	0.532	0.609
Hybrid_02	F-1	0.731	0.735	0.724	0.739	0.533	0.614
Hybrid_03	SMOTE, F-1	0.719	0.699	0.765	0.672	0.489	0.597

Model	Optimised on	ROC-AUC	Accuracy	Recall	Specificity	Precision	F-1 Score
baseline*	-	-	0.709	-	-	-	-
NLP_01	ROC-AUC	0.713	0.716	0.706	0.721	0.509	0.592
nNLP_01	ROC-AUC	0.657	0.672	0.622	0.692	0.454	0.524
NLP_02	F-1	0.719	0.748	0.649	0.788	0.557	0.600
nNLP_02	F-1	0.669	0.648	0.717	0.620	0.437	0.543
NLP_03	SMOTE, F-1	0.688	0.674	0.721	0.655	0.462	0.563
nNLP_03	SMOTE, F-1	0.647	0.604	0.750	0.544	0.403	0.525

Built Models

Singapore TOURISM BOARD

Model	Optimised on	ROC-AUC	Accuracy	Recall	Specificity	Precision	F-1 Score
baseline*	-	-	0.709	-	-	-	-
Hybrid_01	ROC-AUC	0.727	0.734	0.711	0.743	0.532	0.609
Hybrid_02	F-1	0.731	0.735	0.724	0.739	0.533	0.614
Hybrid_03	SMOTE, F-1	0.719	0.699	0.765	0.672	0.489	0.597

	Predicted Negatives	Predicted Positives
Actual Negatives	3168	1097
Actual Positives	506	1246

Model 01

Model	Optimised on	ROC-AUC	Accuracy	Recall	Specificity	Precision	F-1 Score
baseline*	-	-	0.709	-	-	-	-
Hybrid_01	ROC-AUC	0.727	0.734	0.711	0.743	0.532	0.609
Hybrid_02	F-1	0.731	0.735	0.724	0.739	0.533	0.614
Hybrid_03	SMOTE, F-1	0.719	0.699	0.765	0.672	0.489	0.597

	Predicted Negatives	Predicted Positives
Actual Negatives	3168	1097
Actual Positives	506	1246

False Positives, Risk Resources

Recall

False Negatives, Opportunity Cost

Model 01

Model	Optimised on	ROC-AUC	Accuracy	Recall	Specificity	Precision	F-1 score
baseline*	-	-	0.709	-	-		-
Hybrid_01	ROC-AUC	0.727	0.734	0.711	0.743	0.532	0.609
Hybrid_02	F-1	0.731	0.735	0.724	0.739	0.533	0.614
Hybrid_03	SMOTE, F-1	0.719	0.699	0.765	0.672	0.489	0.597

	Predicted Negatives	Predicted Positives
Actual Negatives	3168	1097
Actual Positives	506	1246

False Positives, Risk Resources

False Negatives, Opportunity Cost

Precision

Model 01

Singapore TOURISM BOARD

Model	Optimised on	ROC-AUC	Accuracy	Recall	Specificity	Precision	F-1 score
baseline*	-	-	0.709	-	-		-
Hybrid_01	ROC-AUC	0.727	0.734	0.711	0.743	0.532	0.609
Hybrid_02	F-1	0.731	0.735	0.724	0.739	0.533	0.614
Hybrid_03	SMOTE, F-1	0.719	0.699	0.765	0.672	0.489	0.597

	Predicted Negatives	Predicted Positives
Actual Negatives	3152 -16	1113 +16
Actual Positives	484 -22	1268 +22

Model 02

Model	Optimised on	ROC-AUC	Accuracy	Recall	Specificity	Precision	F-1 score
baseline*	-	-	0.709	-	-		-
Hybrid_01	ROC-AUC	0.727	0.734	0.711	0.743	0.532	0.609
Hybrid_02	F-1	0.731	0.735	0.724	0.739	0.533	0.614
Hybrid_03	SMOTE, F-1	0.719	0.699	0.765	0.672	0.489	0.597

	Predicted Negatives	Predicted Positives
Actual Negatives	3152 -16	1113 +16
Actual Positives	484 -22	1268 +22

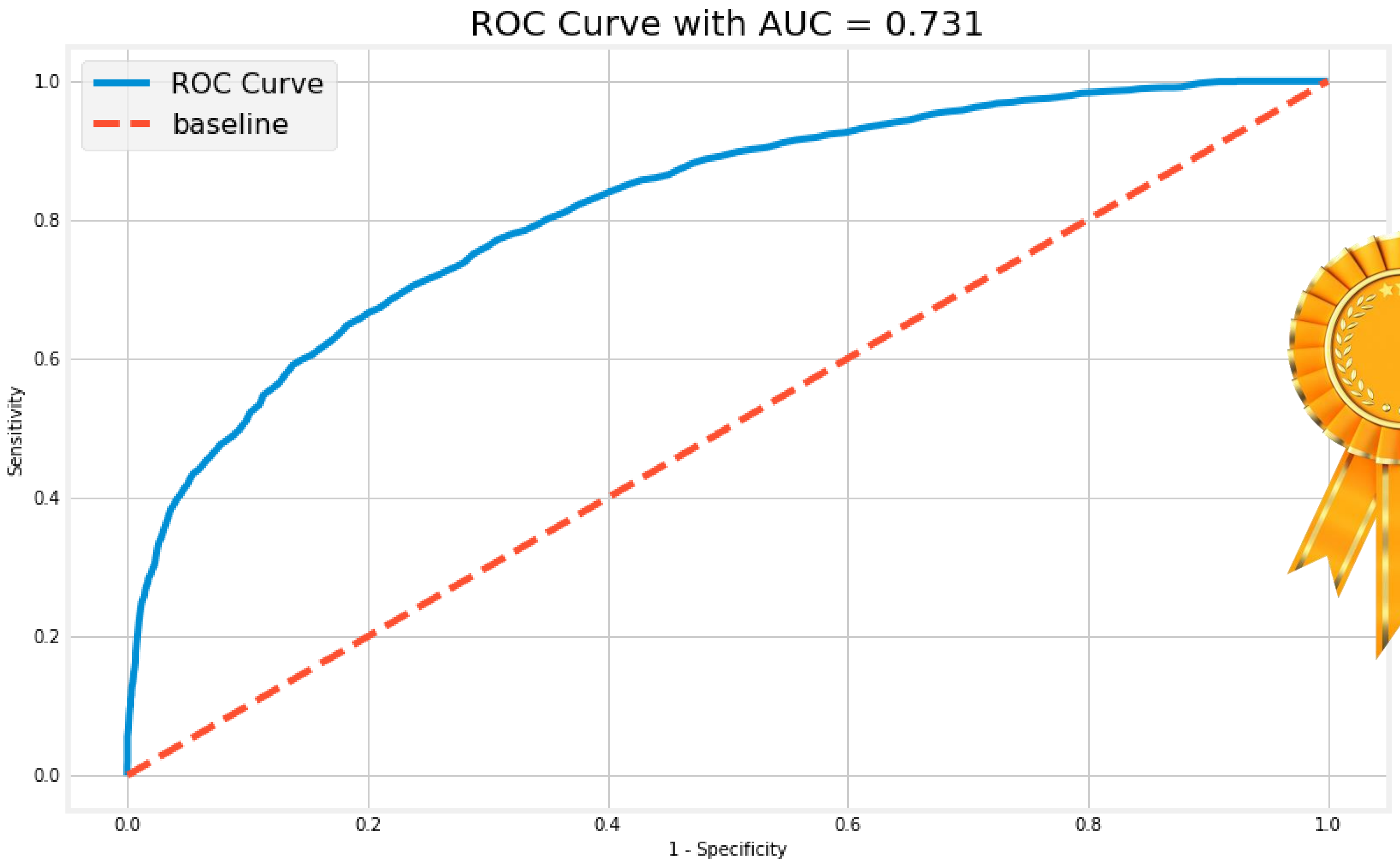
False Positives, Risk Resources

False Negatives, Opportunity Cost

Precision

Model 02

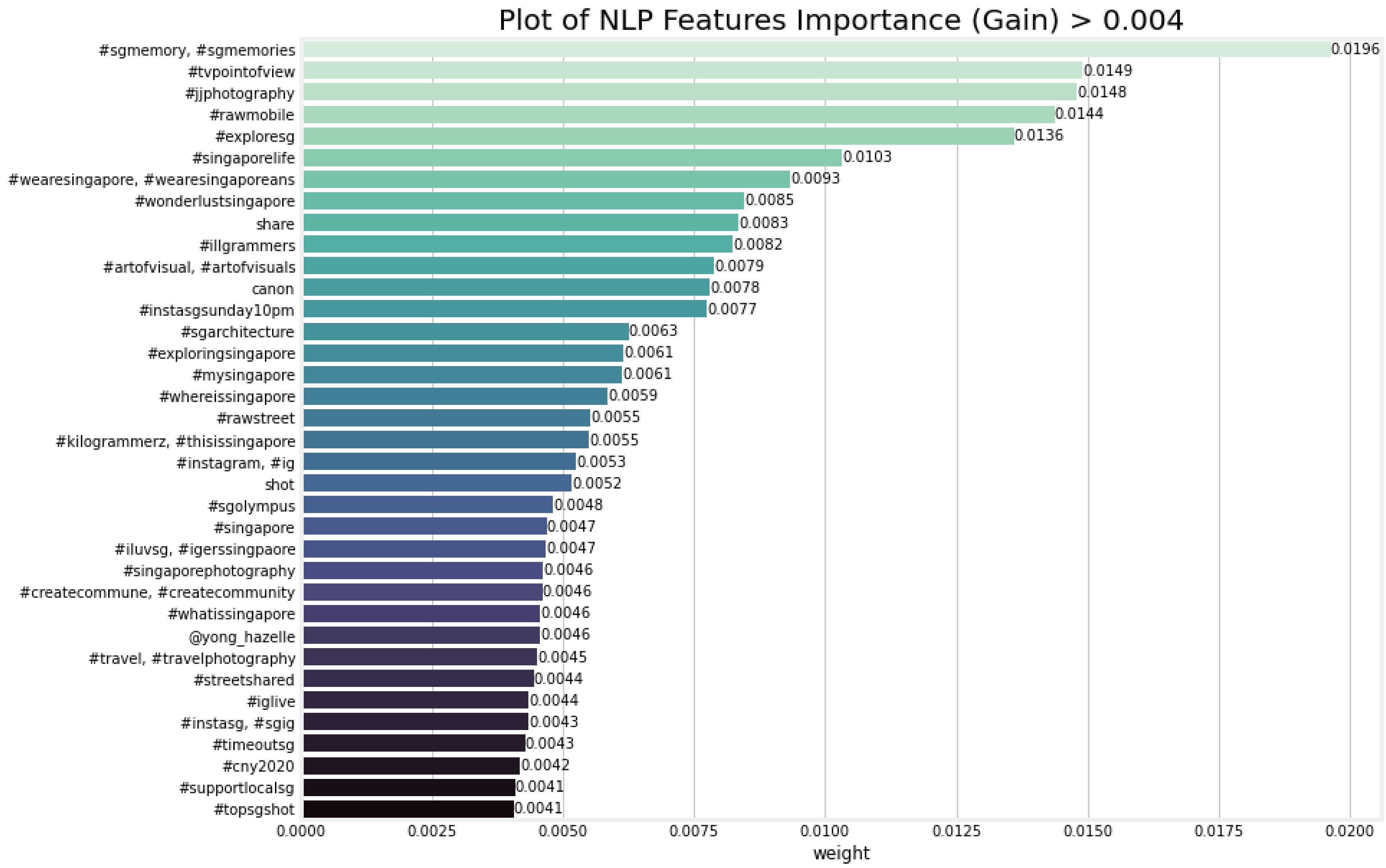
Singapore TOURISM BOARD


ROC Curve with AUC = 0.731
ROC Curve
baseline
1.0
0.8
0.6
0.4
0.2
0.0
Sensitivity
0.0
0.2
0.4
0.6
0.8
1.0
1 - Specificity

Model 02

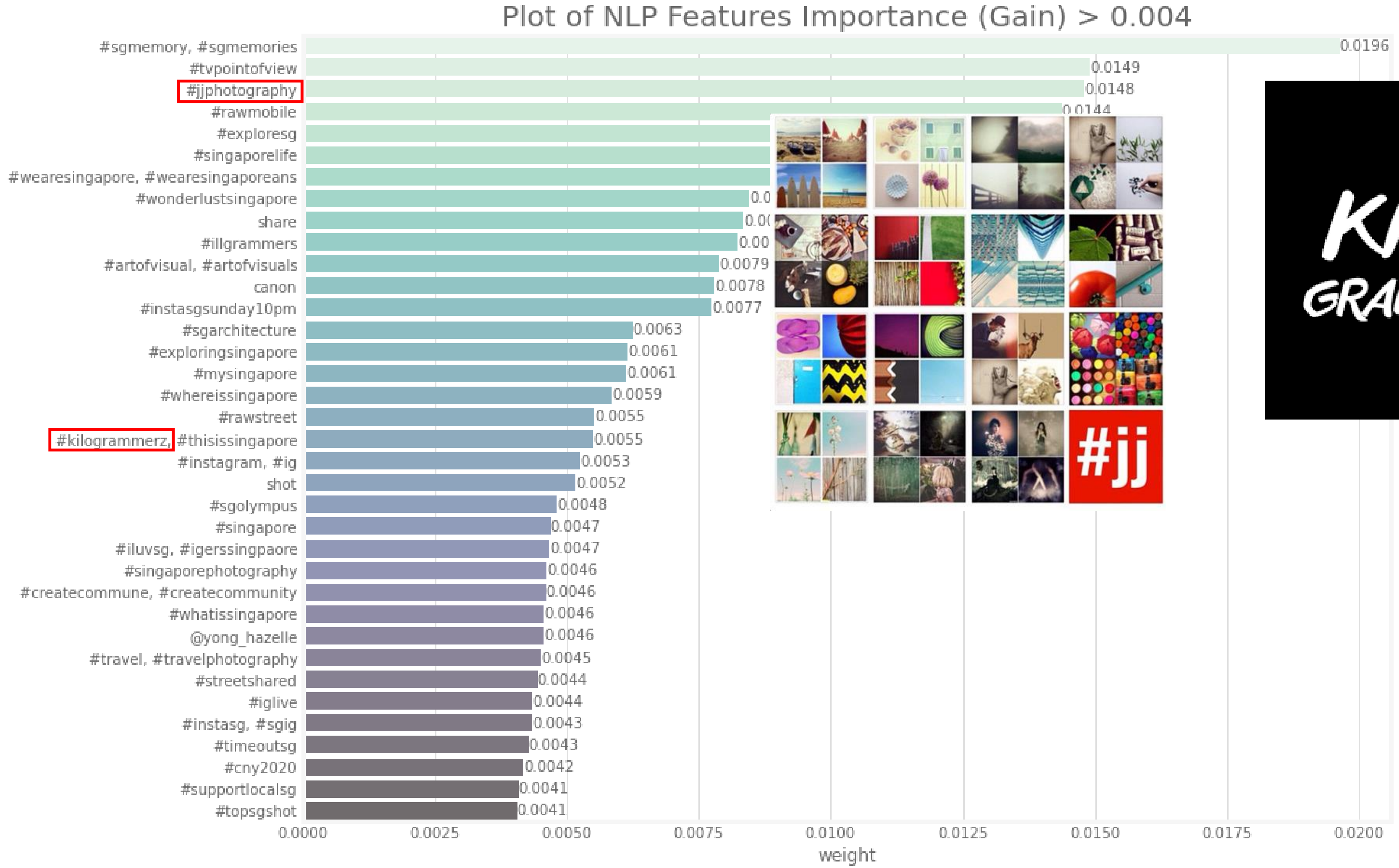
Inference

Inferring and extracting insights from model.



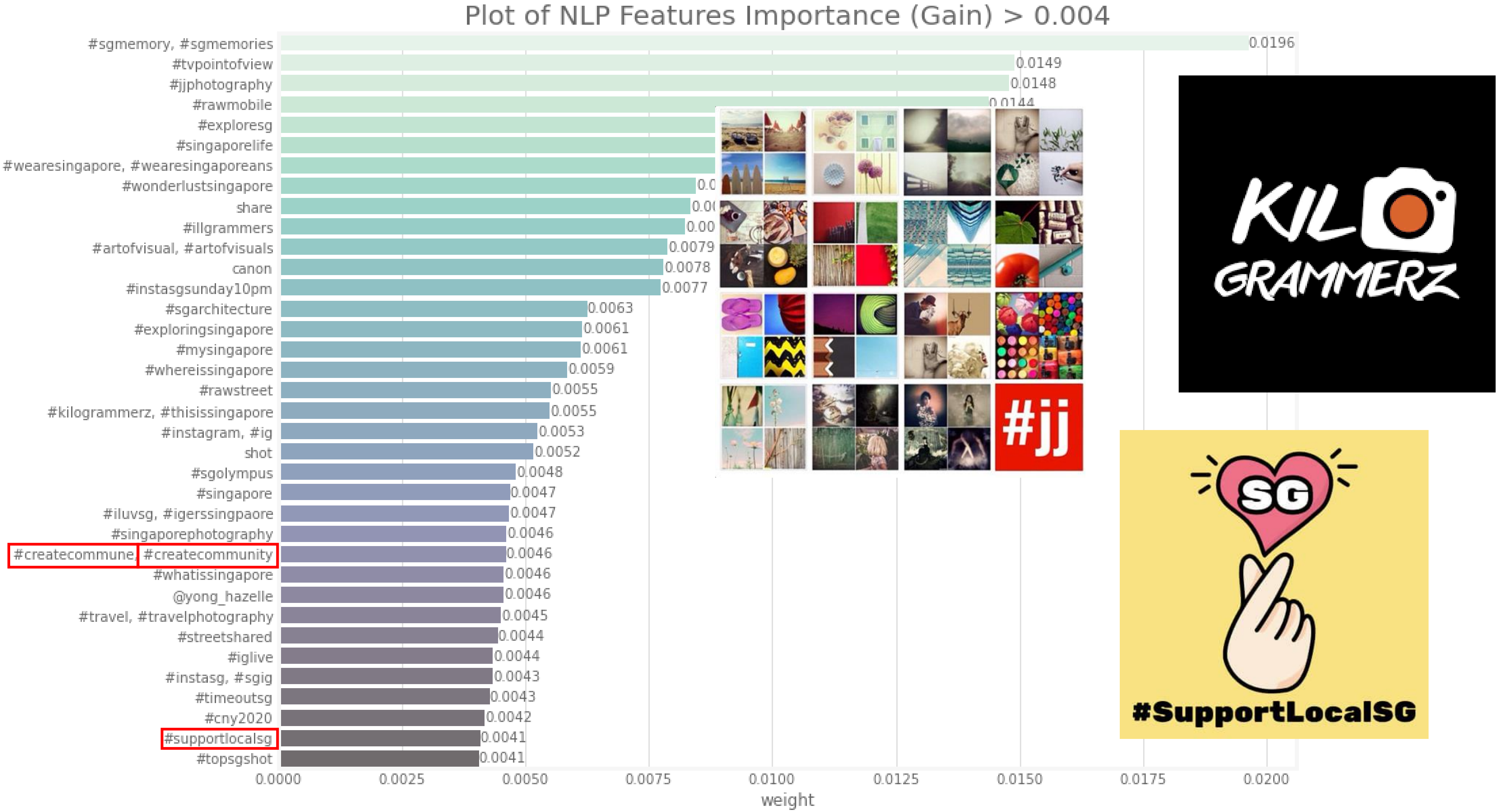
Model 02 – NLP

Singapore TOURISM BOARD



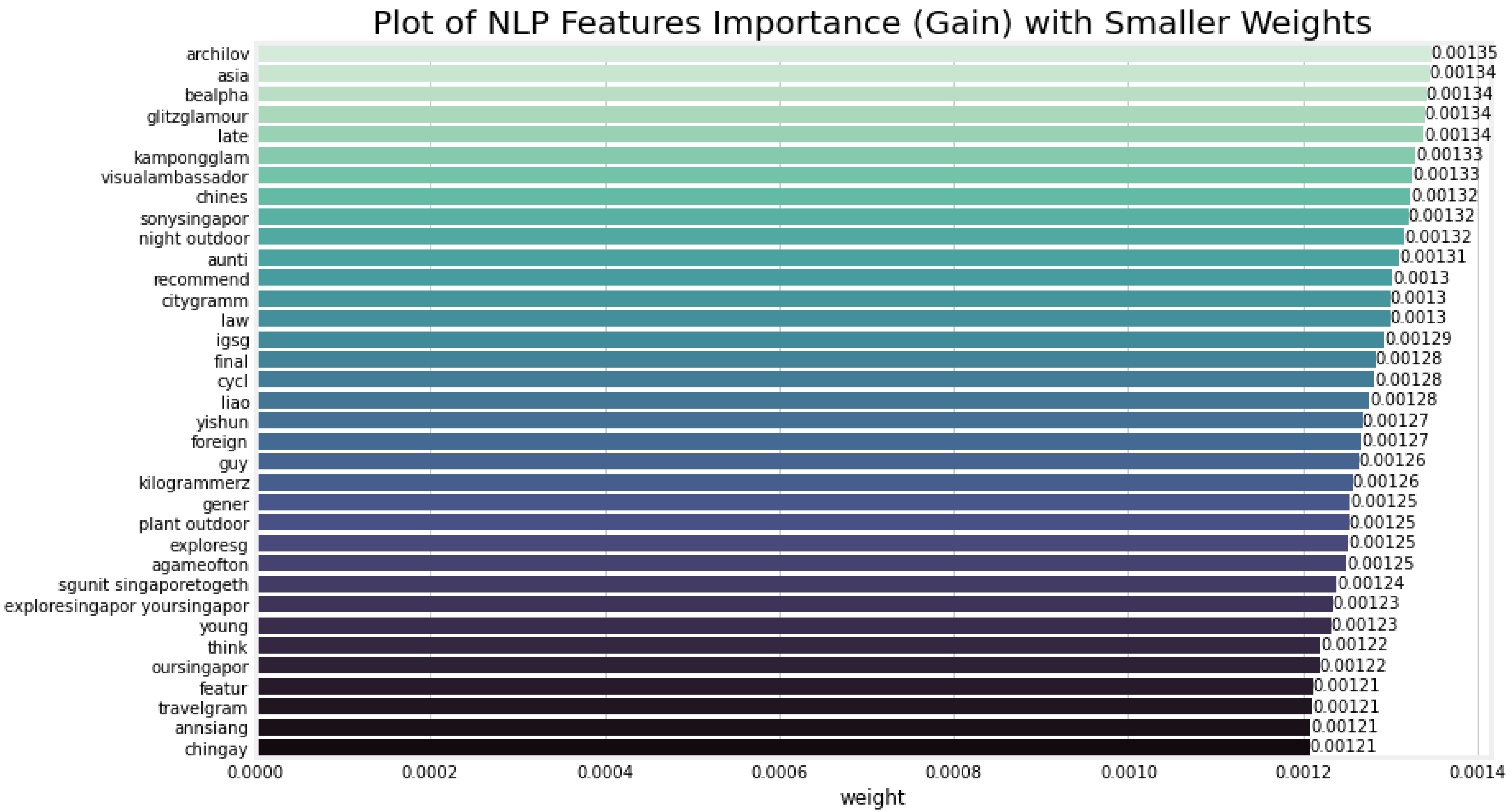
Plot of NLP Features Importance (Gain) > 0.004
#sgmemory, #sgmemories 0.0196
#tvpointofview 0.0149
#jjphotography 0.0148
#rawmobile 0.0144
#exploresg
#singaporelife
#wearesingapore, #wearesingaporeans
#wonderlustsingapore 0.0
share 0.00
#illgrammers 0.00
#artofvisual, #artofvisuals 0.0079
canon 0.0078
#instasgsunday10pm 0.0077
#sgarchitecture 0.0063
#exploringsingapore 0.0061
#mysingapore 0.0061
#whereissingapore 0.0059
#rawstreet 0.0055
#kilogrammerz, #thisissingapore 0.0055
#instagram, #ig 0.0053
shot 0.0052
#sgolympus 0.0048
#singapore 0.0047
#iluvsg, #igerssingpaore 0.0047
#singaporephotography 0.0046
#createcommune, #createcommunity 0.0046
#whatissingapore 0.0046
@yong_hazelle 0.0046
#travel, #travelphotography 0.0045
#streetshared 0.0044
#iglive 0.0044
#instasg, #sgig 0.0043
#timeoutsg 0.0043
#cny2020 0.0042
#supportlocalsg 0.0041
#topsgshot 0.0041
0.0000
0.0025
0.0050
0.0075
0.0100
0.0125
0.0150
0.0175
0.0200
weight
#jj

Model 02 – NLP



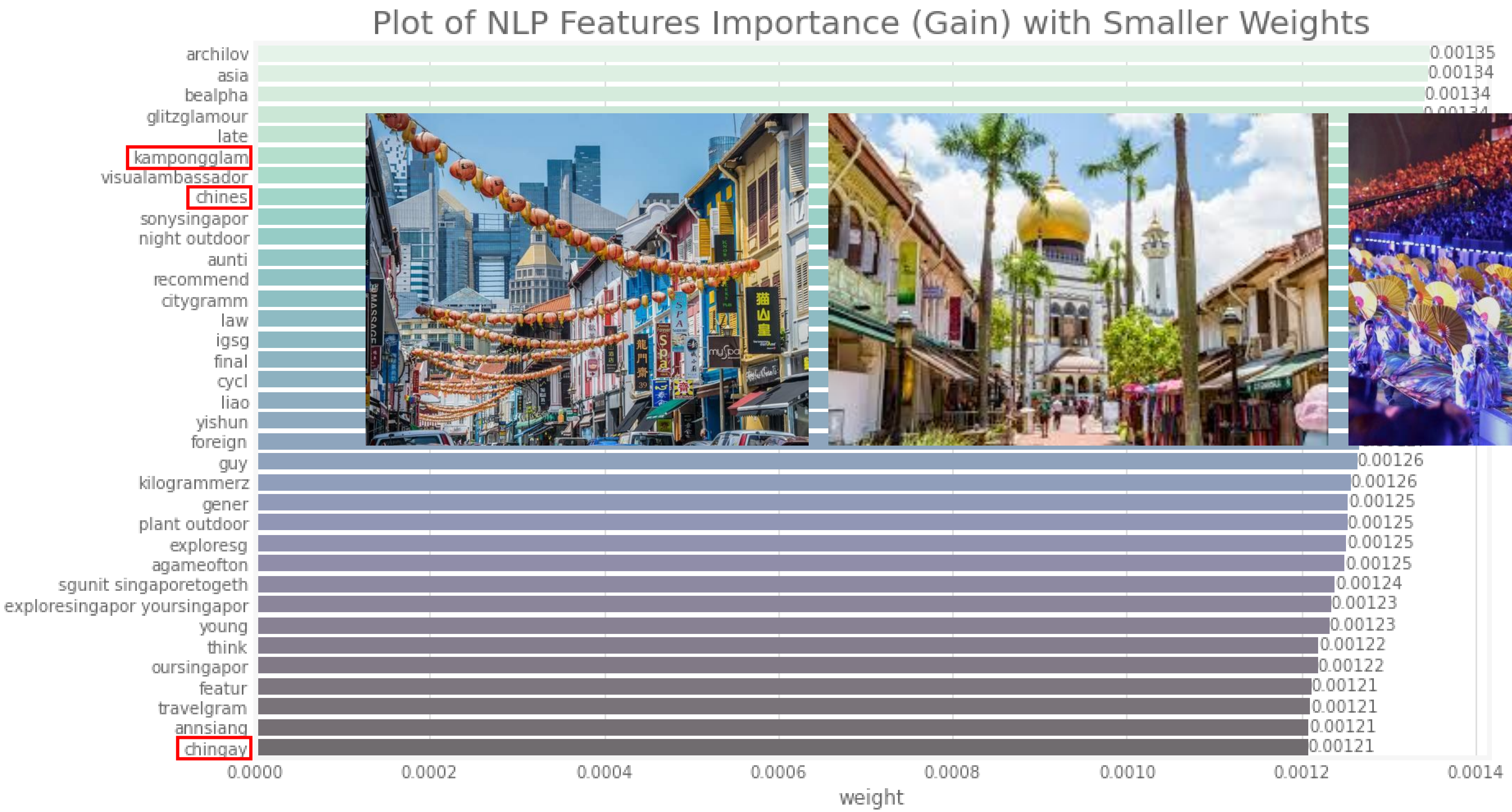
Model 02 – NLP

Singapore TOURISM BOARD



Model 02 – NLP

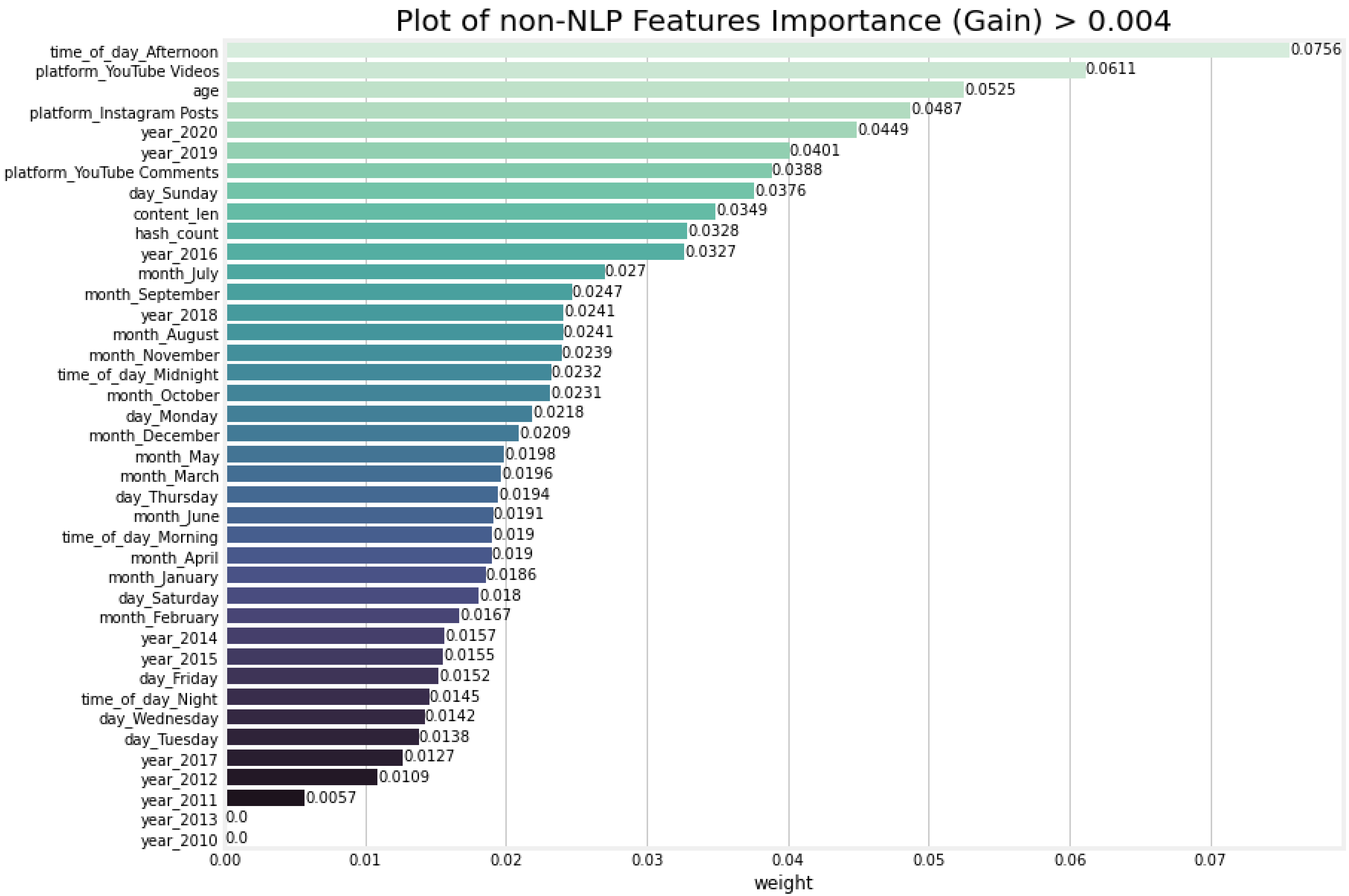
Singapore TOURISM BOARD


Plot of NLP Features Importance (Gain) with Smaller Weights
archilov 0.00135
asia 0.00134
bealpha 0.00134
glitzglamour
late
kampongglam
visualambassador
chines
sonysingapor
night outdoor
aunti
recommend
citygramm
law
igsg
final
cycl
liao
yishun
foreign
guy 0.00126
kilogrammerz 0.00126
gener 0.00125
plant outdoor 0.00125
exploresg 0.00125
agameofton 0.00125
sgunit singaporetogeth 0.00124
exploresingapor yoursingapor 0.00123
young 0.00123
think 0.00122
oursingapor 0.00122
featur 0.00121
travelgram 0.00121
annsiang 0.00121
chingay 0.00121
0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
0.0012
0.0014
weight

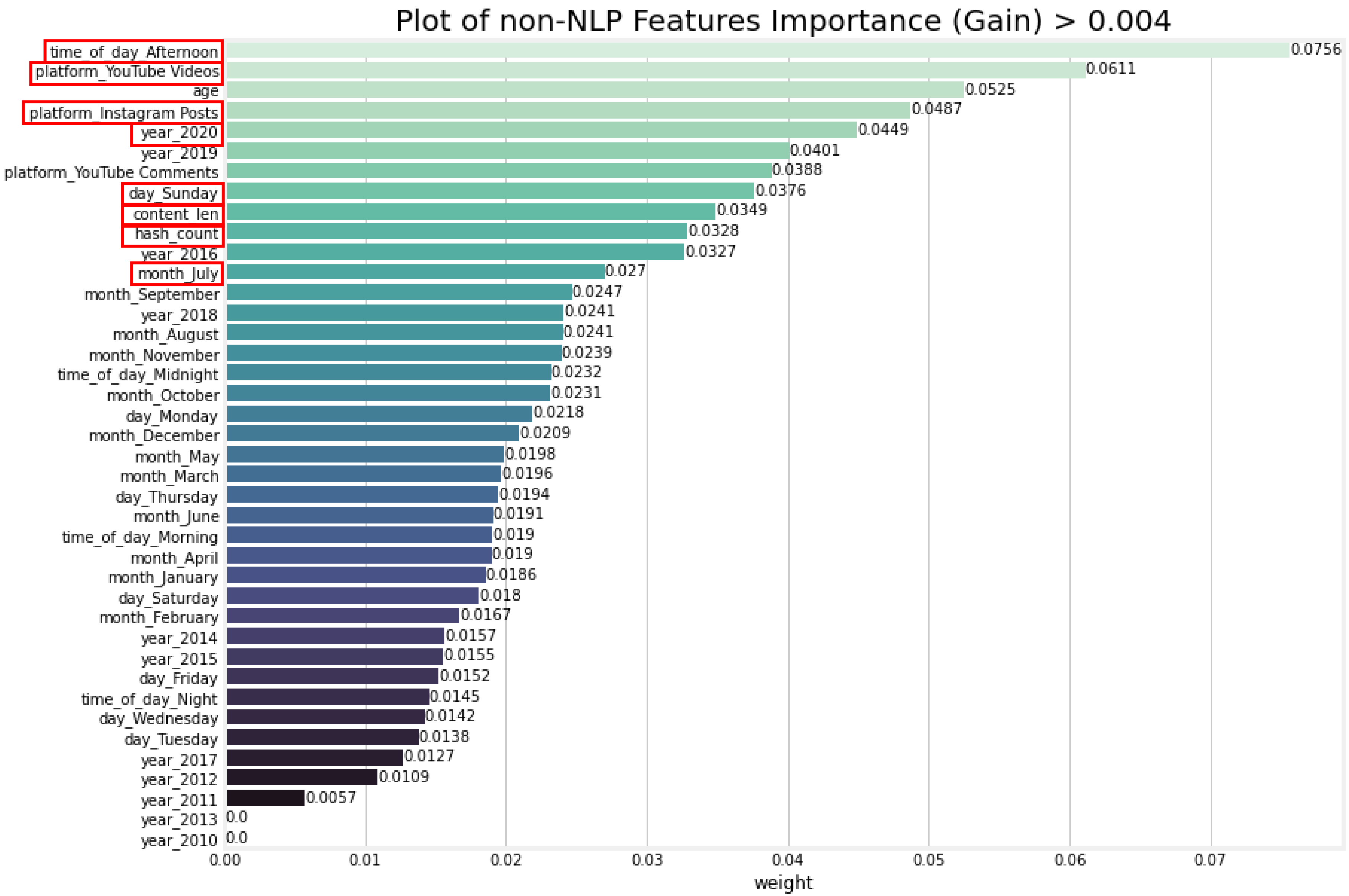

Model 02 – NLP


Plot of NLP Features Importance (Gain) with Smaller Weights
archilov
asia
bealpha
glitzglamour
late
kampongglam
visualambassador
chines
sonysingapor
night outdoor
aunti
recommend
citygramm
law
igsg
final
cycl
liao
yishun
foreign
guy
kilogrammerz
gener
plant outdoor
exploresg
agameofton
sgunit singaporetogeth
exploresingapor yoursingapor
young
think
oursingapor
featur
travelgram
annsiang
chingay
0.00135
0.00134
0.00128
0.00128
0.00121
0.00121
0.00121
0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
0.0012
0.0014
weight
littledayout.com
CHONG PANG CITY

Model 02 – NLP



Model 02 – non-NLP



Model 02 – non-NLP

Singapore TOURISM BOARD

Actionables

Recommendations based on insights for the next phase of #SingapoRediscovers campaign.

Instagram Photo Cycle

Proposal 01

Instagram Photo Cycle

Proposal 01

Singapore TOURISM BOARD

Instagram Photo Cycle

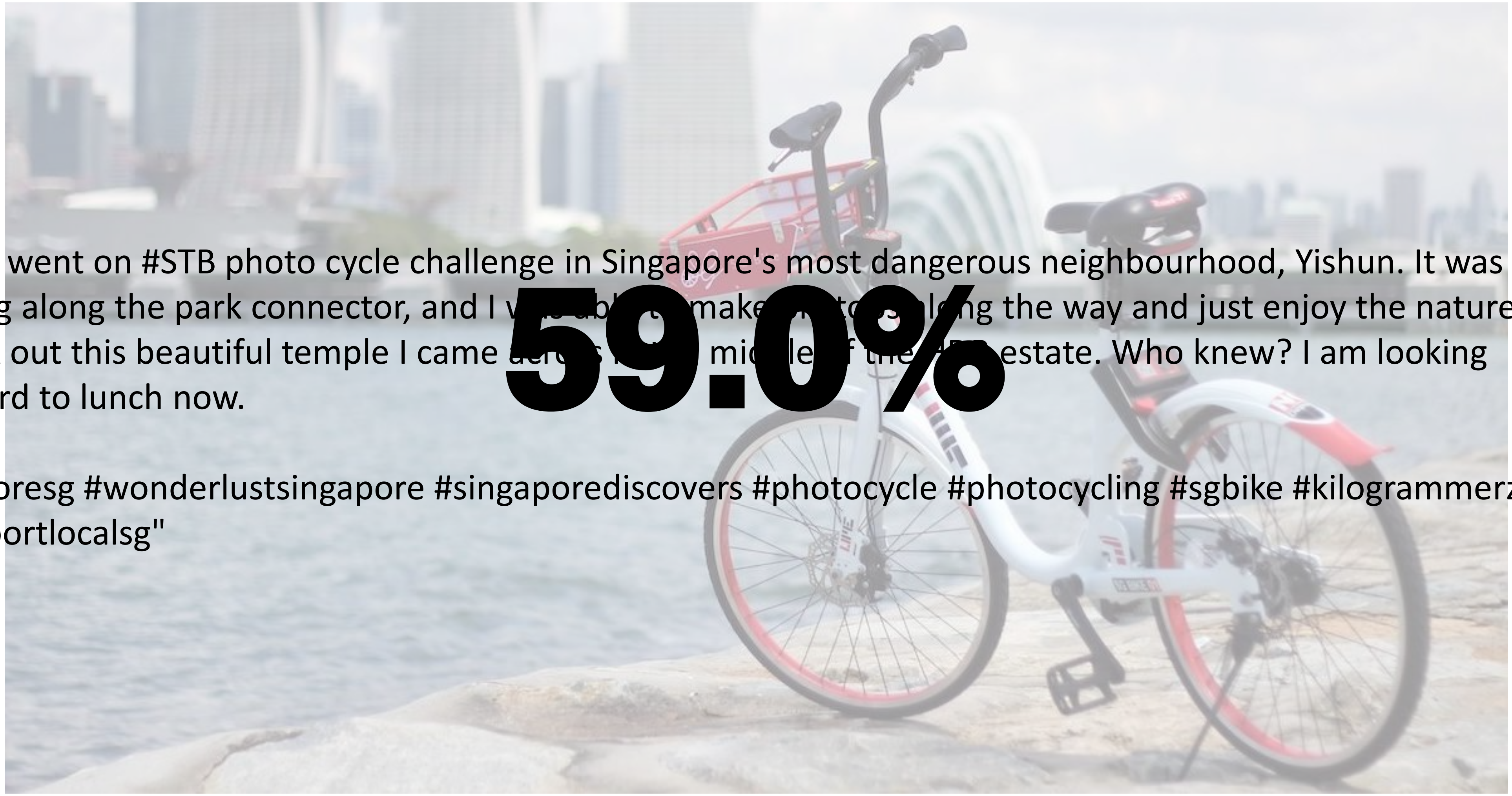
"I just went on #STB photo cycle challenge in Singapore's most dangerous neighbourhood, Yishun. It was easy cycling along the park connector, and I was able to make pitstops along the way and just enjoy the nature. Check out this beautiful temple I came across in the middle of the HDB estate. Who knew? I am looking forward to lunch now.

#exploresg #wonderlustsingapore #singaporediscovers #photocycle #photocycling #sgbike #kilogrammerz #supportlocalsg"

Proposal 01

Singapore TOURISM BOARD

Instagram Photo Cycle



"I just went on #STB photo cycle challenge in Singapore's most dangerous neighbourhood, Yishun. It was easy cycling along the park connector, and I was able to make photos along the way and just enjoy the nature. Check out this beautiful temple I came across in the middle of the HDB estate. Who knew? I am looking forward to lunch now.

#exploresg #wonderlustsingapore #singaporediscovers #photocycle #photocycling #sgbike #kilogrammerz #supportlocalsg"

59.0%

Proposal 01

Pseudo-Travel in Singapore

Proposal 02

Singapore TOURISM BOARD

Pseudo-Travel in Singapore

Proposal 02

Singapore TOURISM BOARD

Pseudo-Travel in Singapore

"Visiting the Thailand of Singapore - Golden Mile Complex

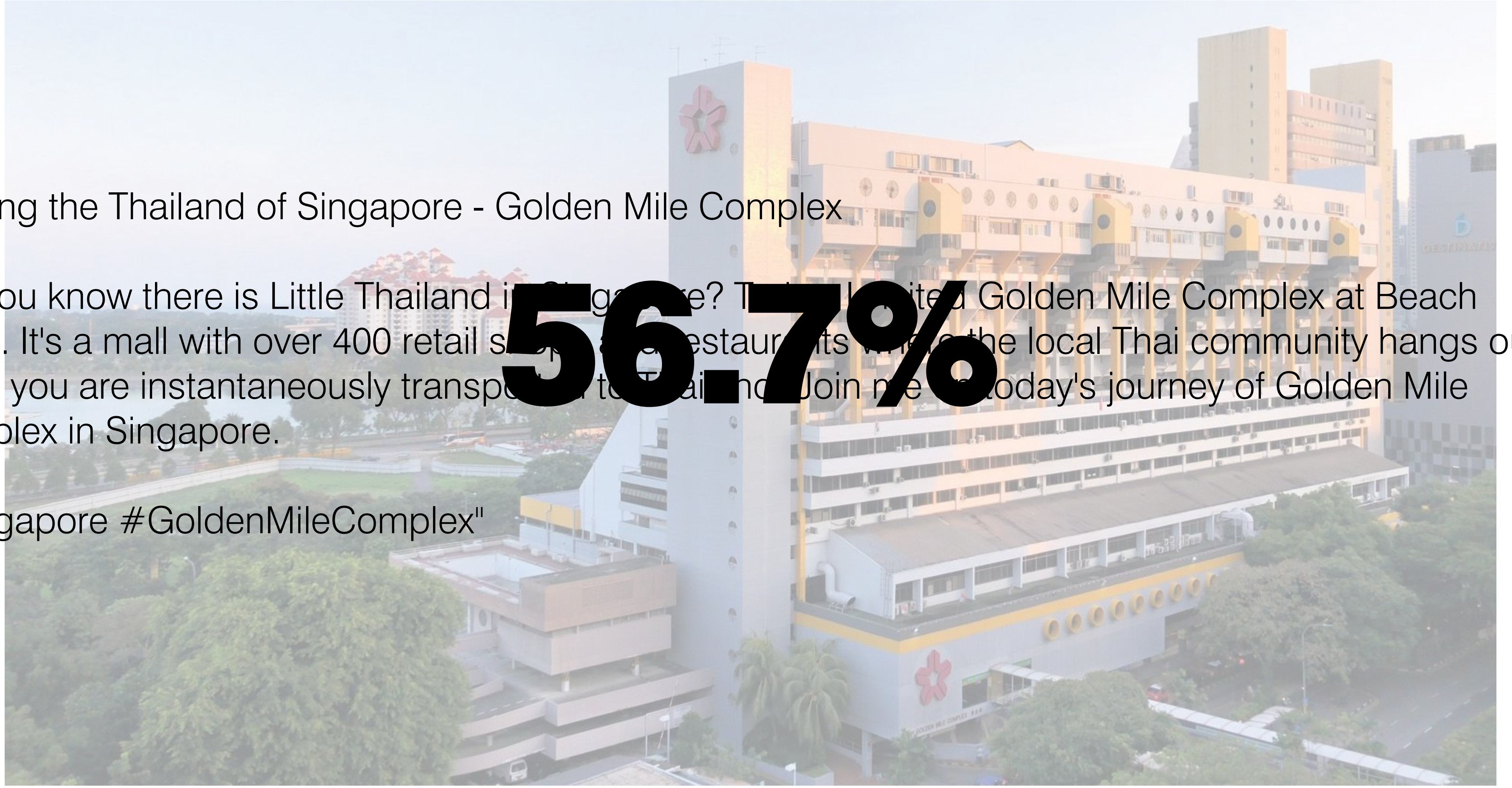
Did you know there is Little Thailand in Singapore? Today, I visited Golden Mile Complex at Beach Road. It's a mall with over 400 retail shops and restaurants where the local Thai community hangs out. Here, you are instantaneously transported to Thailand. Join me on today's journey of Golden Mile Complex in Singapore.

#Singapore #GoldenMileComplex"

Proposal 02

Singapore TOURISM BOARD

Pseudo-Travel in Singapore



"Visiting the Thailand of Singapore - Golden Mile Complex

Did you know there is Little Thailand in Singapore? Today I visited Golden Mile Complex at Beach Road. It's a mall with over 400 retail shops and restaurants where the local Thai community hangs out. Here, you are instantaneously transported to Thailand. Join me on today's journey of Golden Mile Complex in Singapore.

#Singapore #GoldenMileComplex"

56.7%

Proposal 02

Singapore TOURISM BOARD

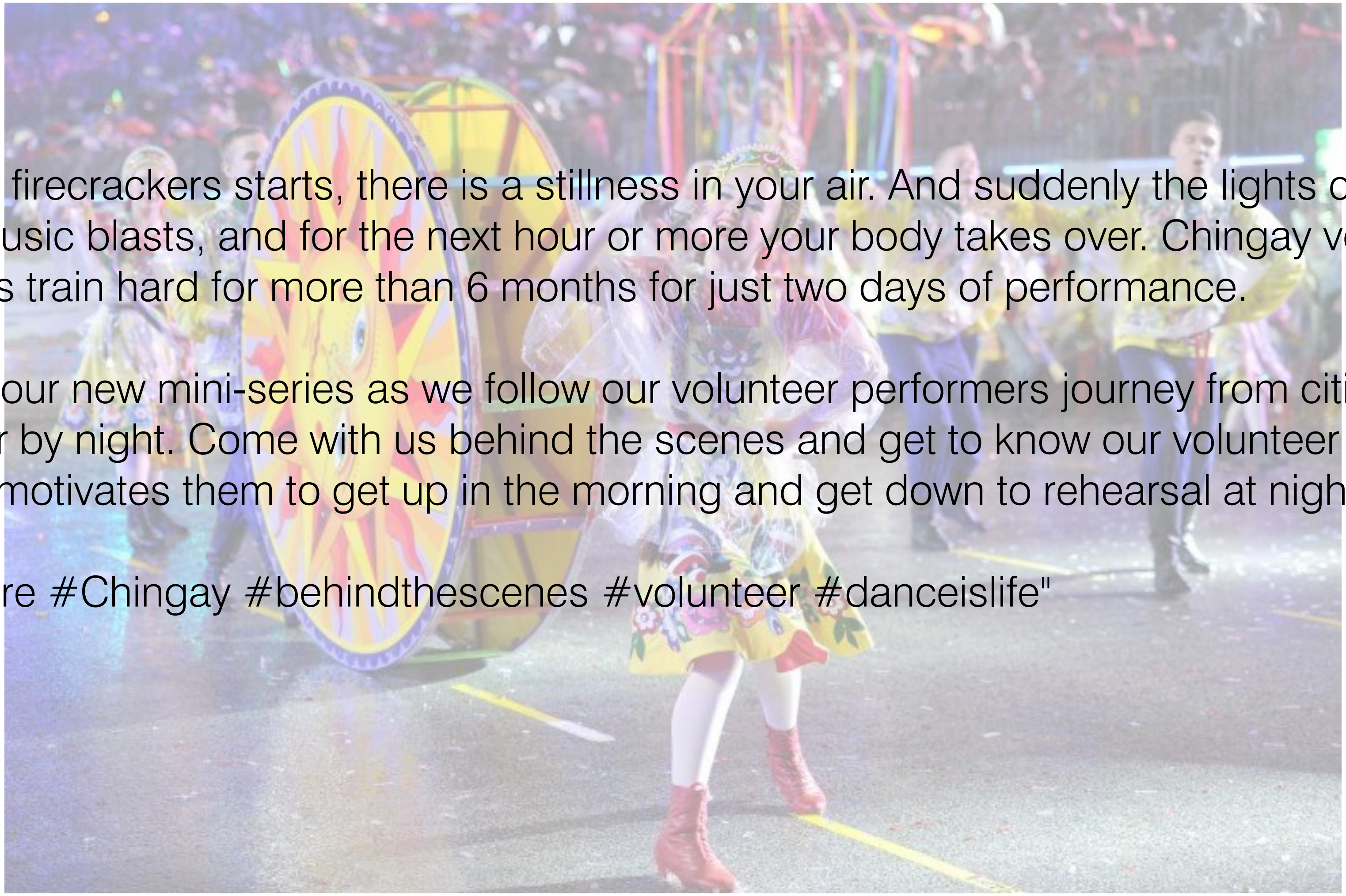
Chingay behind-the-scenes YT mini-series

Proposal 03

Chingay behind-the-scenes YT mini-series

Proposal 03

Chingay behind-the-scenes YT mini-series

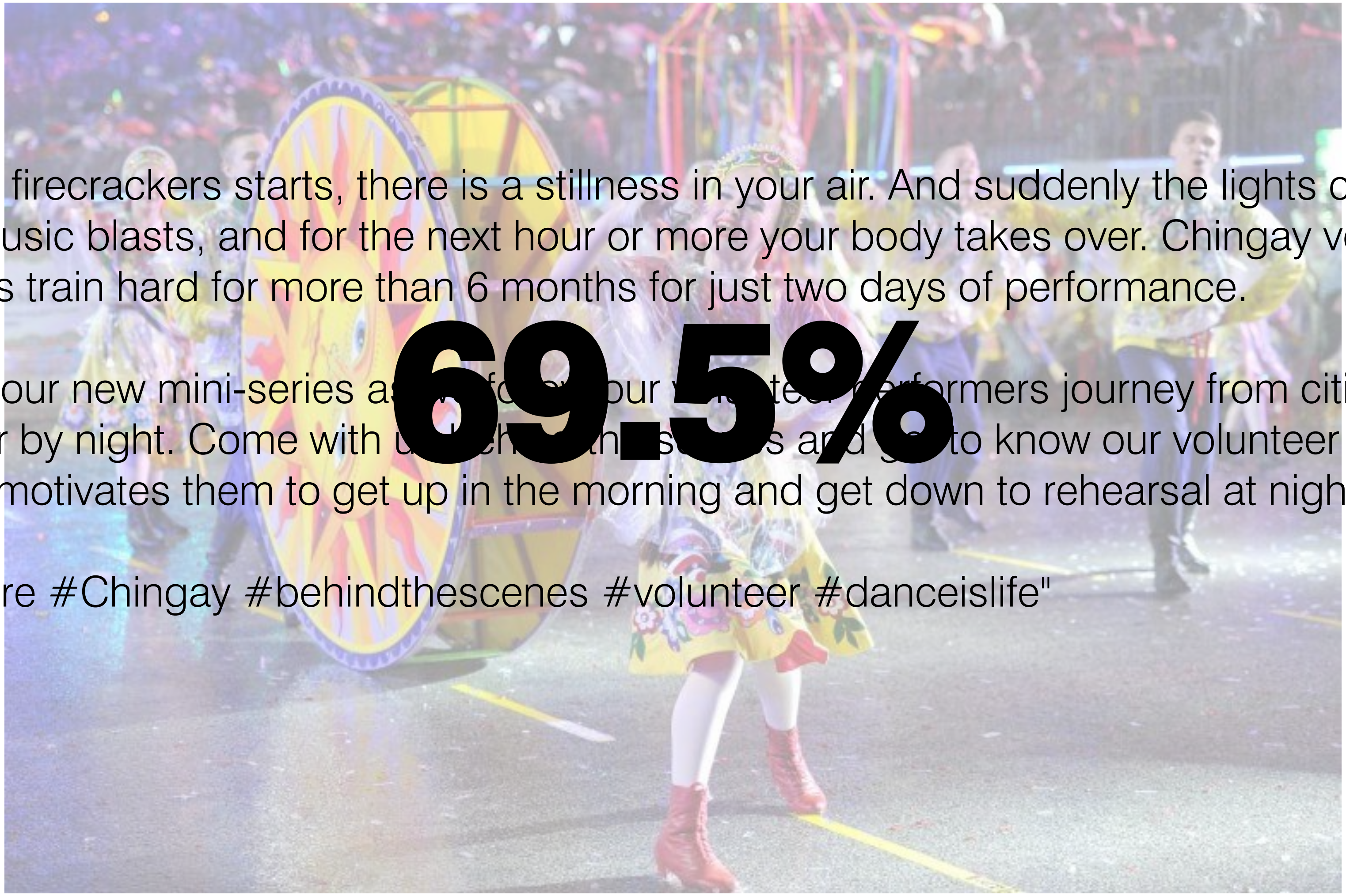


"When the firecrackers starts, there is a stillness in your air. And suddenly the lights come on, and the music blasts, and for the next hour or more your body takes over. Chingay volunteer performers train hard for more than 6 months for just two days of performance.

Join us in our new mini-series as we follow our volunteer performers journey from citizen in the day to star by night. Come with us behind the scenes and get to know our volunteer performers, and what motivates them to get up in the morning and get down to rehearsal at night.

#Singapore #Chingay #behindthescenes #volunteer #danceislife"

Chingay behind-the-scenes YT mini-series



"When the firecrackers starts, there is a stillness in your air. And suddenly the lights come on, and the music blasts, and for the next hour or more your body takes over. Chingay volunteer performers train hard for more than 6 months for just two days of performance.

69.5%

Join us in our new mini-series as we follow our volunteer performers journey from citizen in the day to star by night. Come with us behind the scenes and get to know our volunteer performers, and what motivates them to get up in the morning and get down to rehearsal at night.

#Singapore #Chingay #behindthescenes #volunteer #danceislife"

Proposal 03

Singapore TOURISM BOARD

Moving Forward

What's next?

What more can be done?

Our model has some limits which we can try to overcome:

- By gathering more data
- Trying a weighted form of aggregation
- Dropping all hashtags if enough data is acquired to offset information loss
- Building a Flask app to deploy the model
- Image modelling or transcribing videos to directly access the content

What's Next?

Singapore TOURISM BOARD

Any Questions?

Thank you for your attention!