Predicting the Popularity of YouTube Videos concerning Sephora

Presented by Shu Jiang

WHERE DOES THE DATA COME FROM?

Answer: Through API

Keyword: Sephora

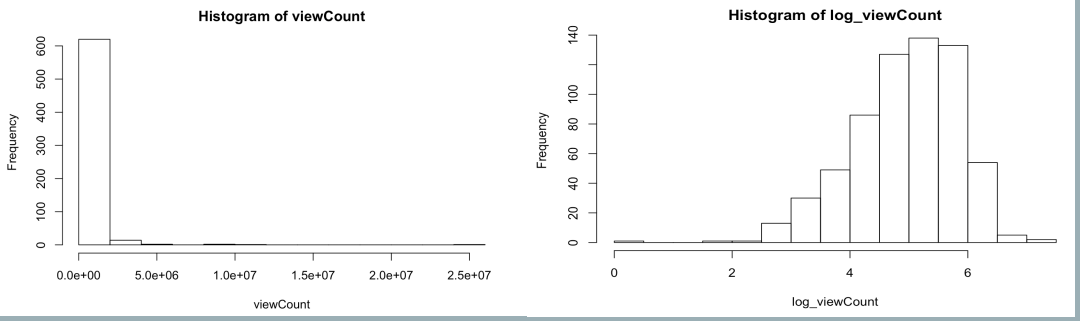
↓

Video ID and Channel ID

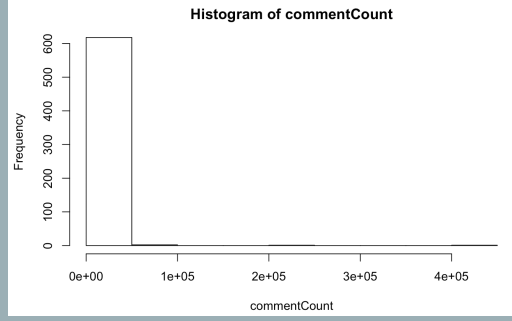
↓

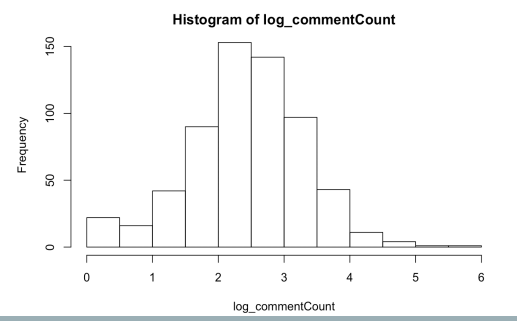
Subscribers count, View count
Comment count and so on

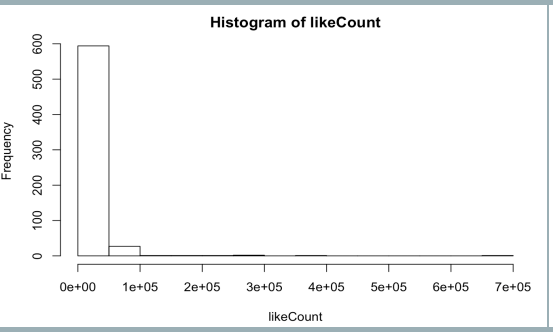
	viewCount <int>	likeCount <int>	dislikeCount <int>	favoriteCount <int>	commentCount <int>	subscriberCount <int>	hiddenSubscriberCount <fctr>	videoCount <int>
1	11475	1076	13	0	233	298713	False	1016
2	432306	17137	220	0	1010	4082602	False	923
3	5214	535	4	0	110	198671	False	261
4	73542	1752	66	0	1273	79838	False	161
5	40821	1967	68	0	320	1192973	False	2680
6	71723	3995	89	0	556	873450	False	269

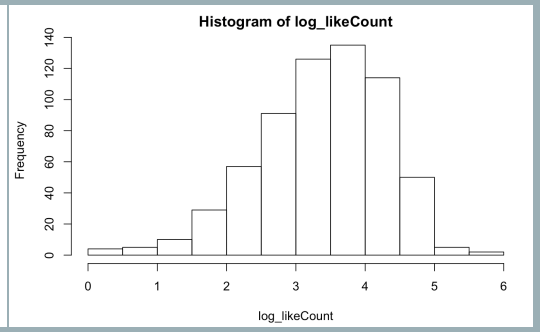

Histogram of viewCount
Frequency
0
100
200
300
400
500
600
0.0e+00
5.0e+06
1.0e+07
1.5e+07
2.0e+07
2.5e+07
viewCount

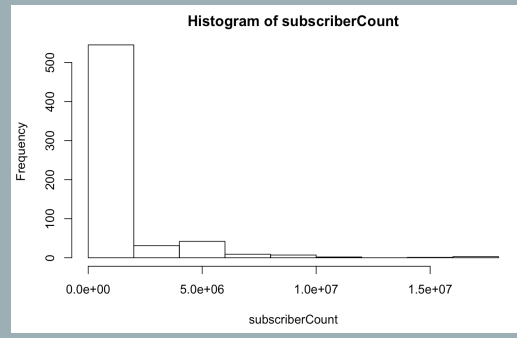
Histogram of log_viewCount
Frequency
0
20
40
60
80
100
140
0
2
4
6
log_viewCount

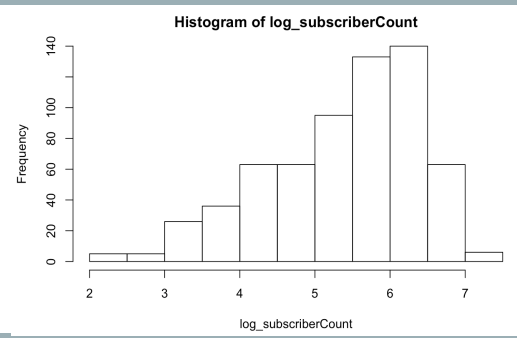

Histogram of commentCount
Frequency
0
100
200
300
400
500
600
0e+00
1e+05
2e+05
3e+05
4e+05
commentCount

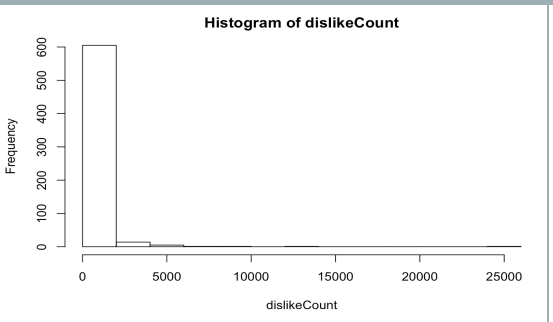

Histogram of log_commentCount
Frequency
0
50
100
150
0
1
2
3
4
5
6
log_commentCount

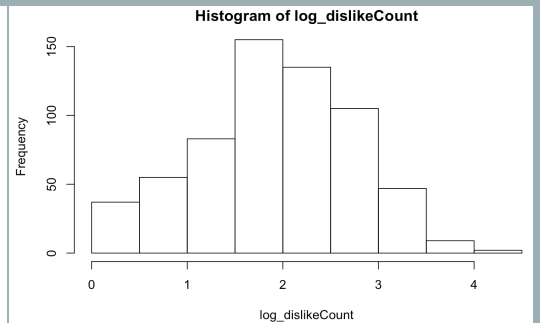

Histogram of likeCount
Frequency
0
100
200
300
400
500
600
0e+00
1e+05
2e+05
3e+05
4e+05
5e+05
6e+05
7e+05
likeCount

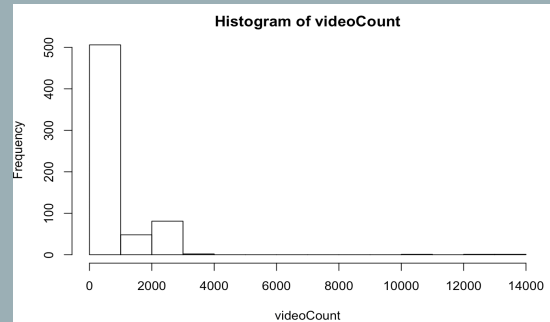

Histogram of log_likeCount
Frequency
0
20
40
60
80
100
120
140
0
1
2
3
4
5
6
log_likeCount

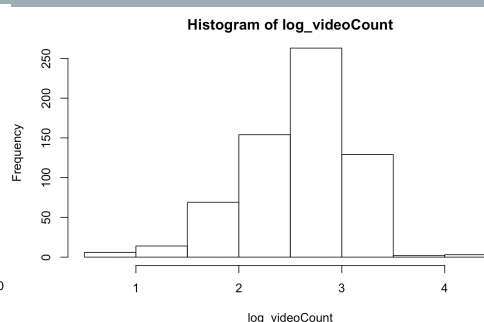

Histogram of subscriberCount
Frequency
0
100
200
300
400
500
0.0e+00
5.0e+06
1.0e+07
1.5e+07
subscriberCount


Histogram of log_subscriberCount
Frequency
0
20
40
60
80
100
140
2
3
4
5
6
7
log_subscriberCount


Histogram of dislikeCount
Frequency
0
100
200
300
400
500
600
0
5000
10000
15000
20000
25000
dislikeCount


Histogram of log_dislikeCount
Frequency
0
50
100
150
0
1
2
3
4
log_dislikeCount


Histogram of videoCount
Frequency
0
100
200
300
400
500
0
2000
4000
6000
8000
10000
12000
14000
videoCount


Histogram of log_videoCount
Frequency
0
50
100
150
200
250
1
2
3
4
log_videoCount

Proposal of intend to do

Predictor Variable: ViewCount, DislikeCount, CommentCount, SubscriberCount, VideoCount HiddensubscriberCount

Outcome Variable: likeCount

Model selection

1. Multiple Linear Regression
2. Random Forest