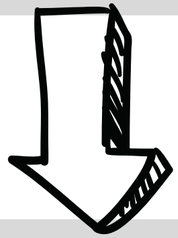
Youtube頻道分析

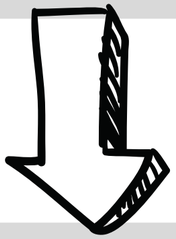
Youtube Data API

Youtube Data API

該頻道的基本資料



頻道內所有影片的ID



頻道內所有影片資料

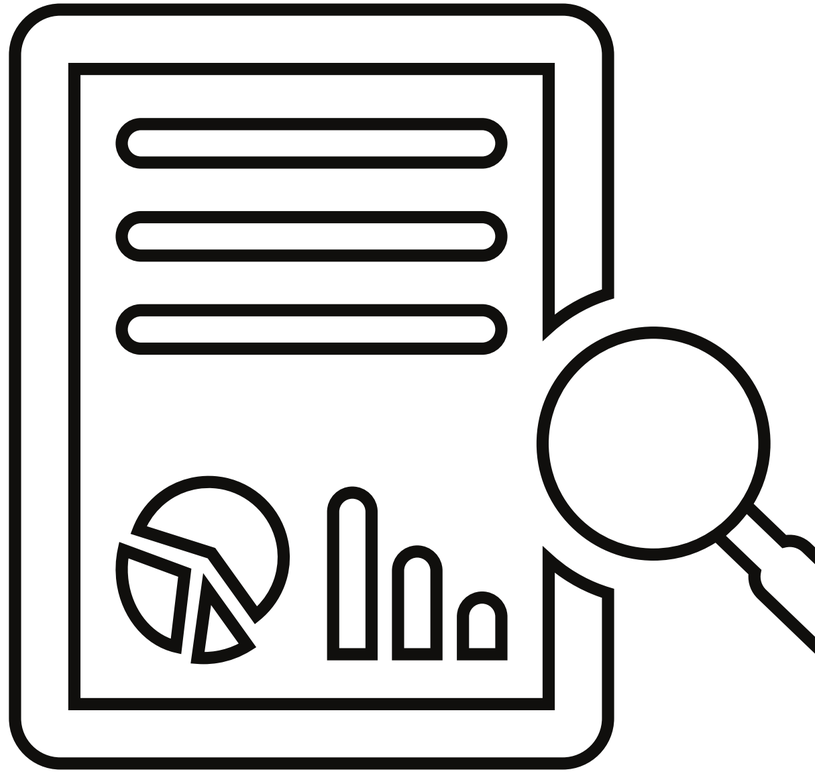
df_channel_info

id
video_length
title
description
published
tag_count
view_count
like_count
dislike_count
comment_count

歷史數據分析

歷史數據分析

文本分析



數字分析

歷史數據分析

文本分析

留言動詞

歷史數據分析

文本分析

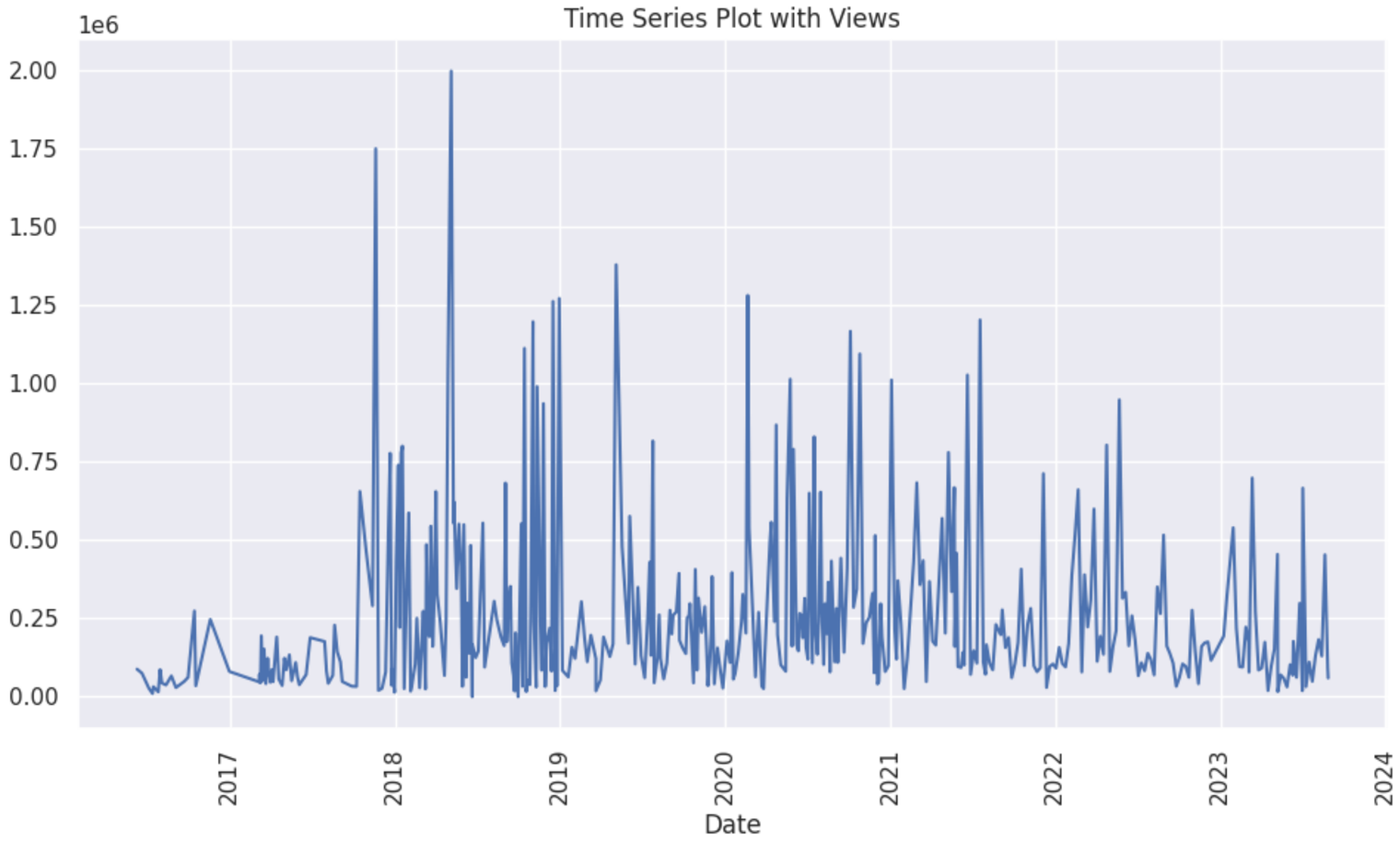
留言名詞

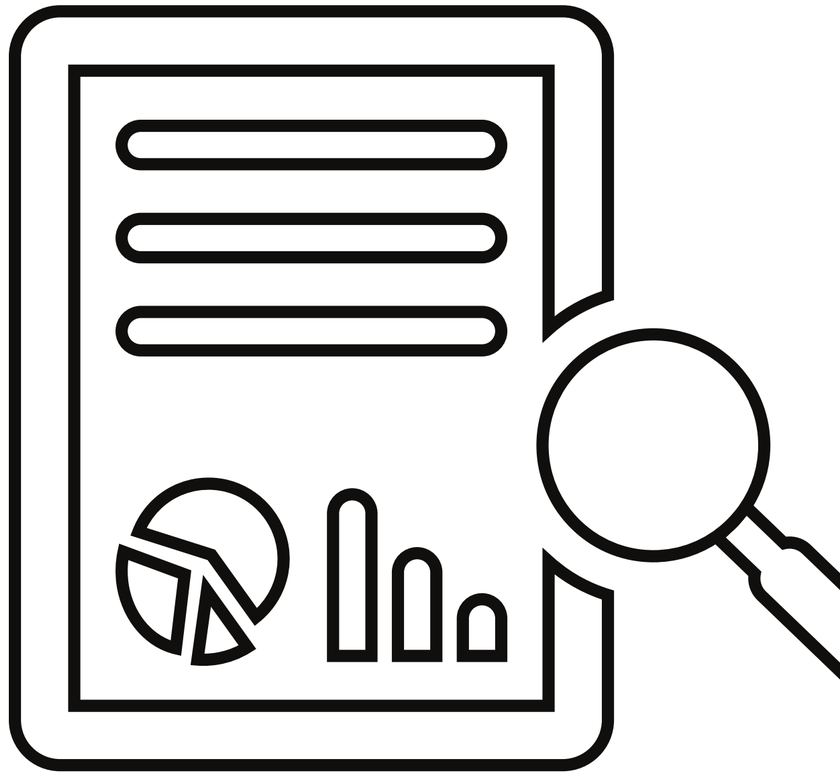
歷史數據分析

文本分析

標題名詞

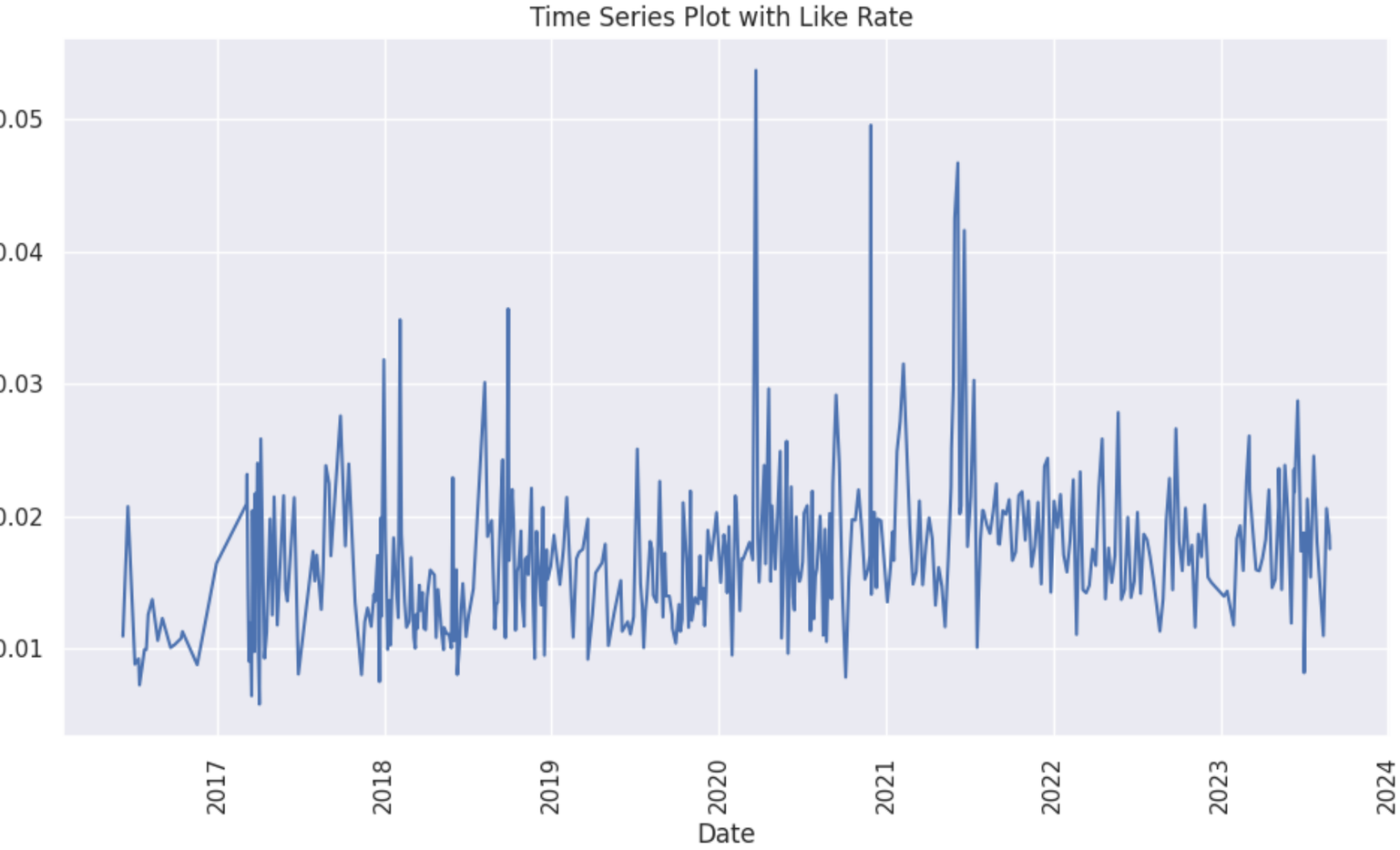
歷史數據分析

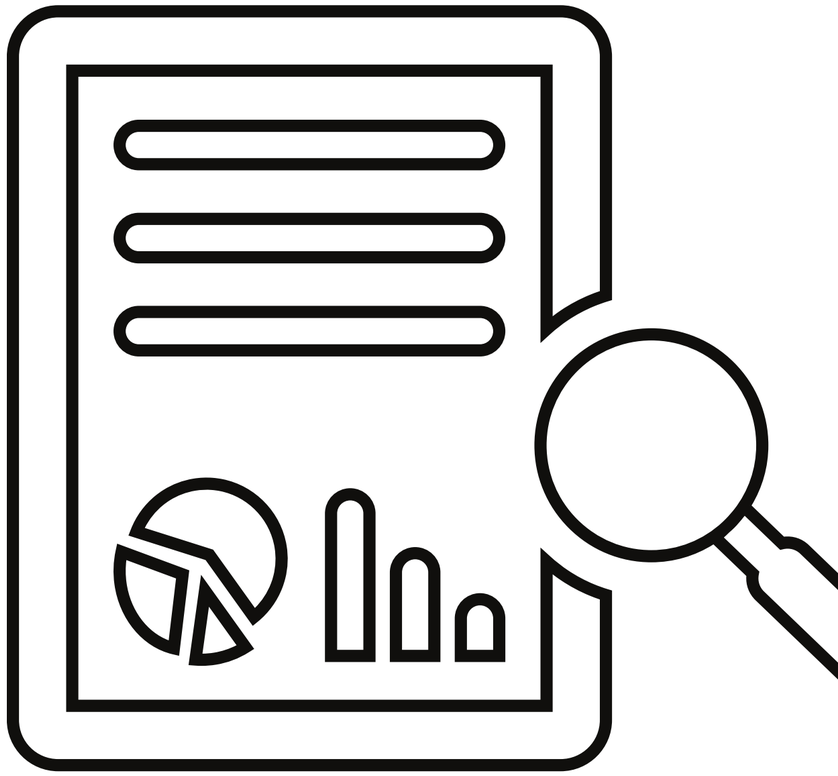

Time Series Plot with Views
1e6
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
2017
2018
2019
2020
2021
2022
2023
2024
Date



數字分析

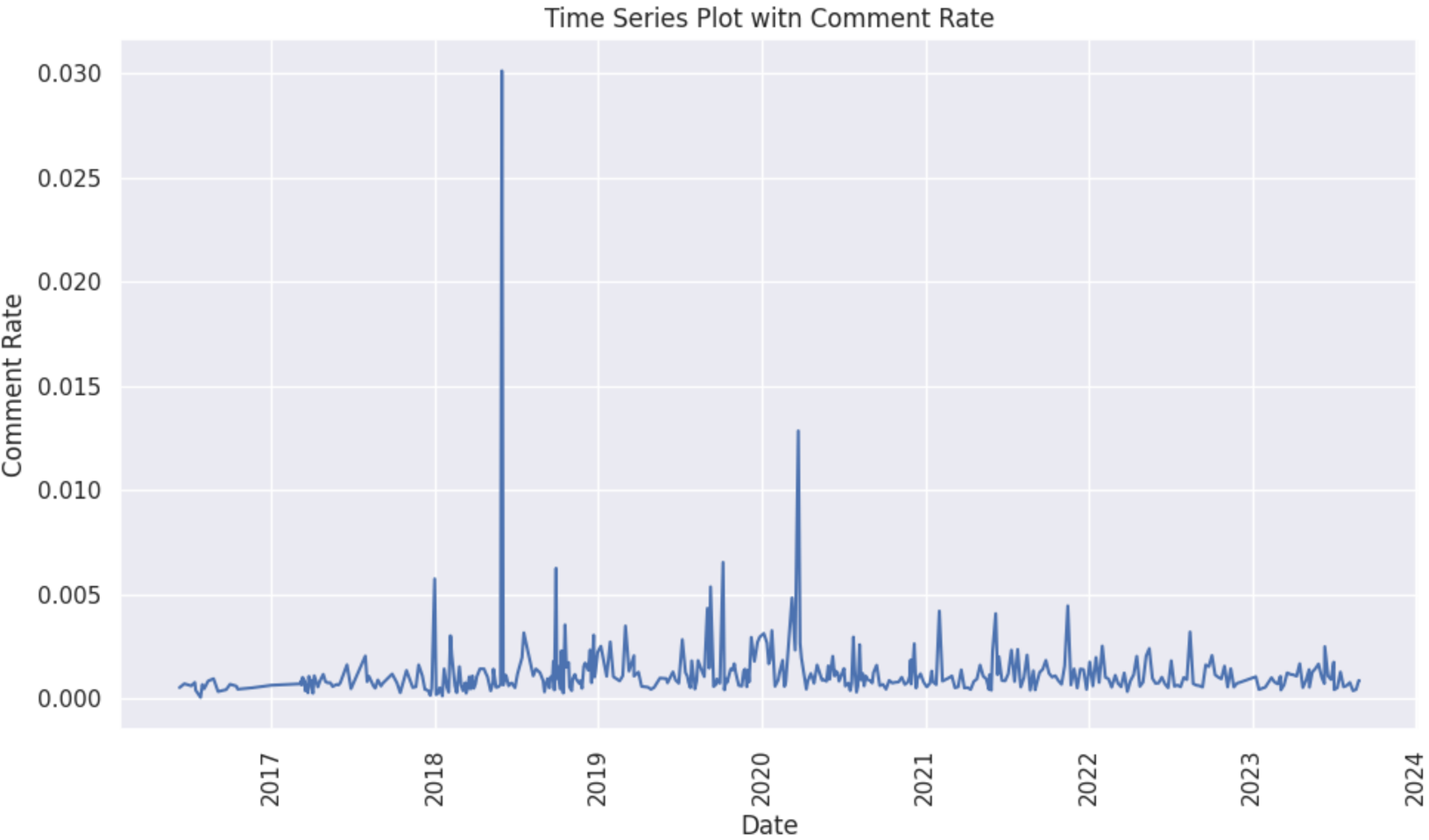
歷史數據分析

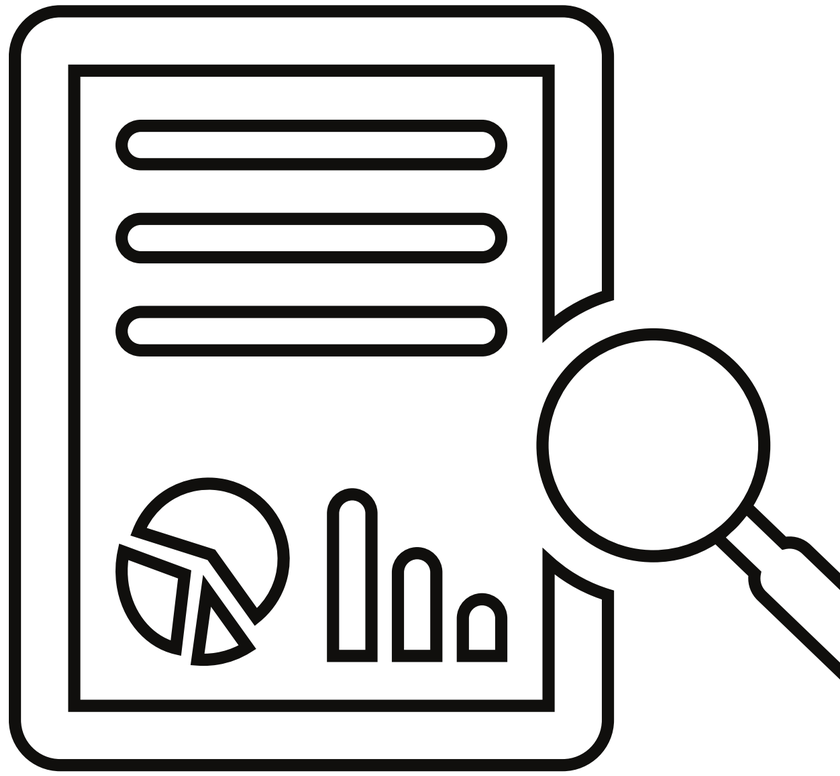

Time Series Plot with Like Rate
0.05
0.04
0.03
0.02
0.01
2017
2018
2019
2020
2021
2022
2023
2024
Date



數字分析

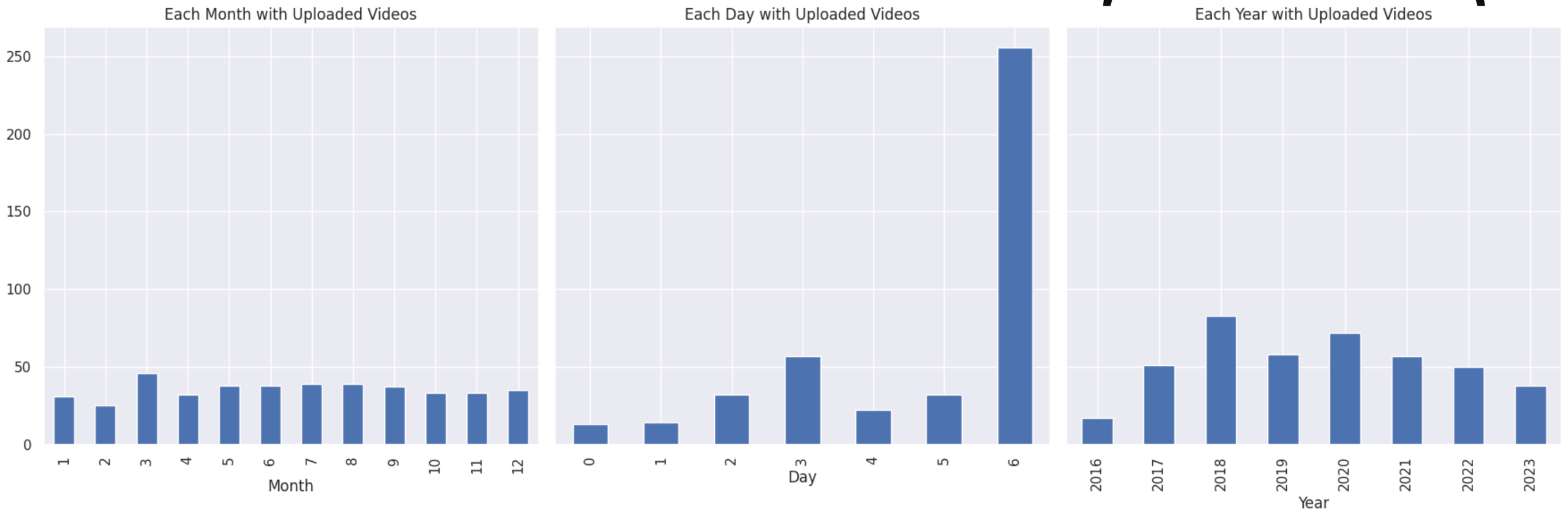
歷史數據分析


Time Series Plot witn Comment Rate
Comment Rate
0.030
0.025
0.020
0.015
0.010
0.005
0.000
2017
2018
2019
2020
2021
2022
2023
2024
Date



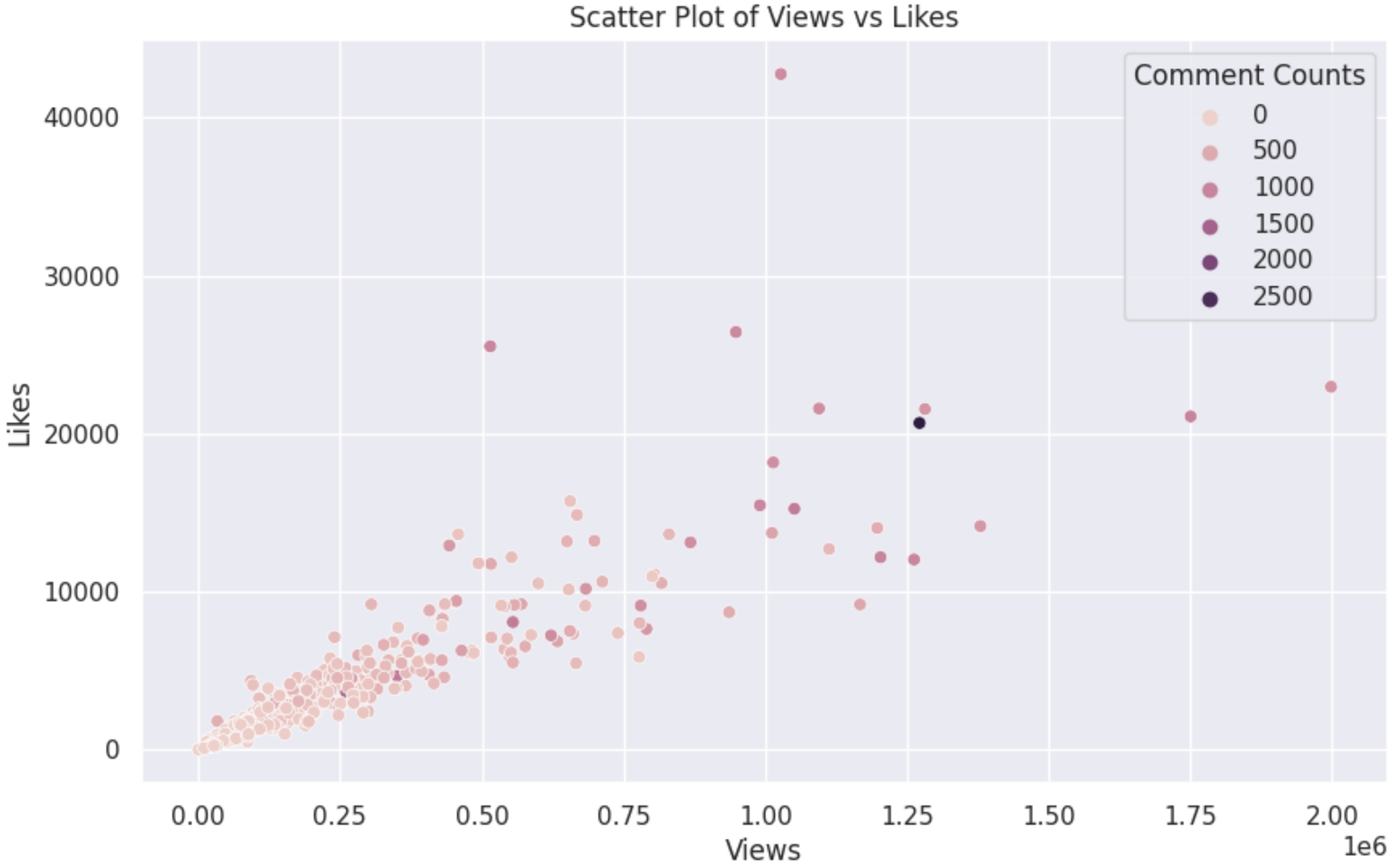
數字分析

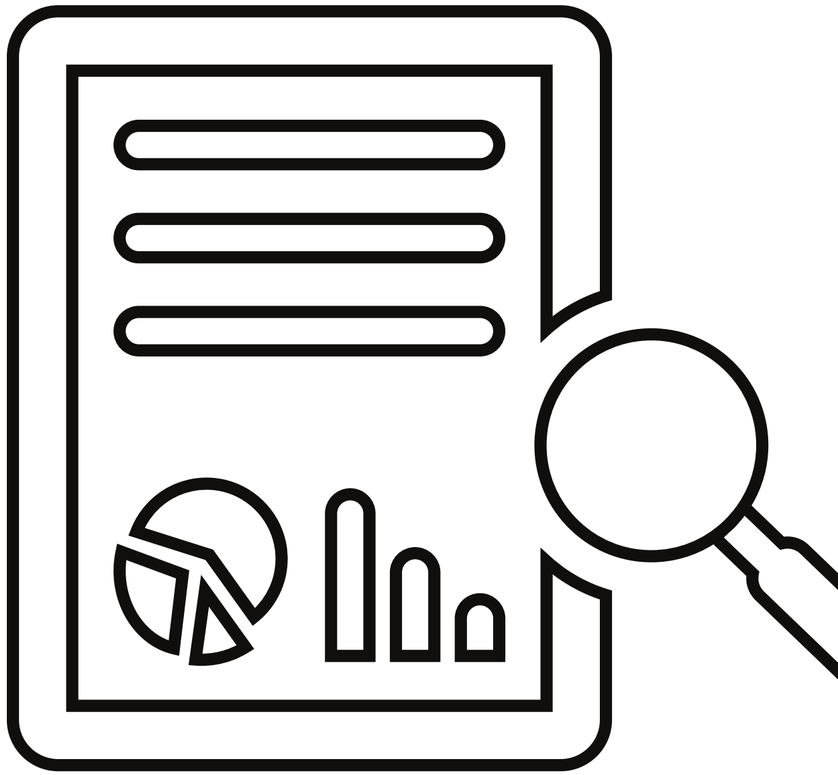
歷史數據分析


Each Month with Uploaded Videos
250
200
150
100
50
0
1
2
3
4
5
6
7
8
9
10
11
12
Month
Each Day with Uploaded Videos
0
1
2
3
4
5
6
Day
Each Year with Uploaded Videos
2016
2017
2018
2019
2020
2021
2022
2023
Year

數字分析

歷史數據分析


Scatter Plot of Views vs Likes
Comment Counts
0
500
1000
1500
2000
2500
Likes
40000
30000
20000
10000
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Views
1e6



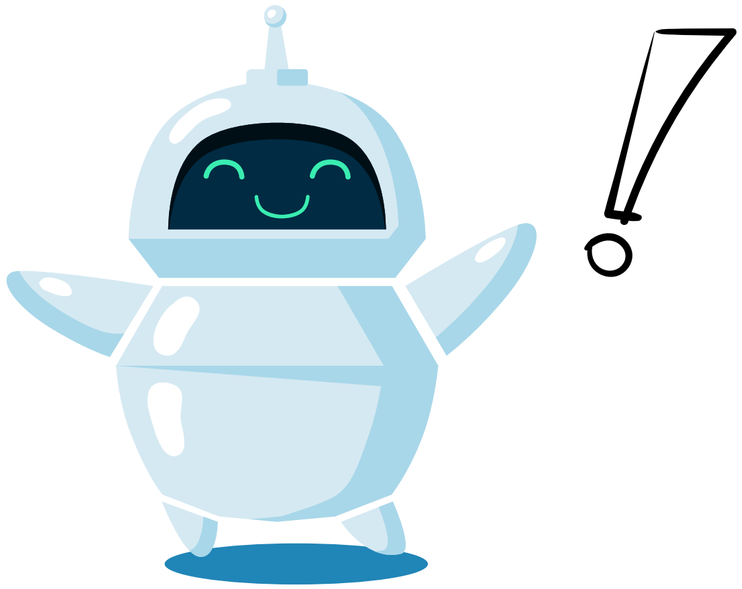
數字分析

機器學習模型
(預測平均每日觀看數)

建立模型步驟

- Feature Engineering
- Data Cleaning
- Data Preprocessing
- Model Selection
- Improve the Model

Testing Data



建立模型步驟

Feature Engineering

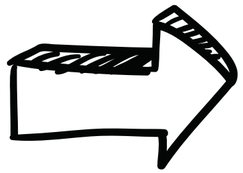
pandas

'month'
'day_of_week'
'link_counts'
'feat_or_not'
'video_length_s'
'title_text_lenght'
'avg_day_of_views'

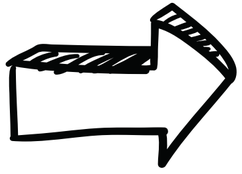
建立模型步驟

Data Cleaning

1		15
	Male	
9	Female	30



data_type_convert



1	Female	15
(1+9)/2	Male	(15+30)/2
9	Female	30

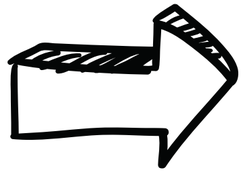
建立模型步驟

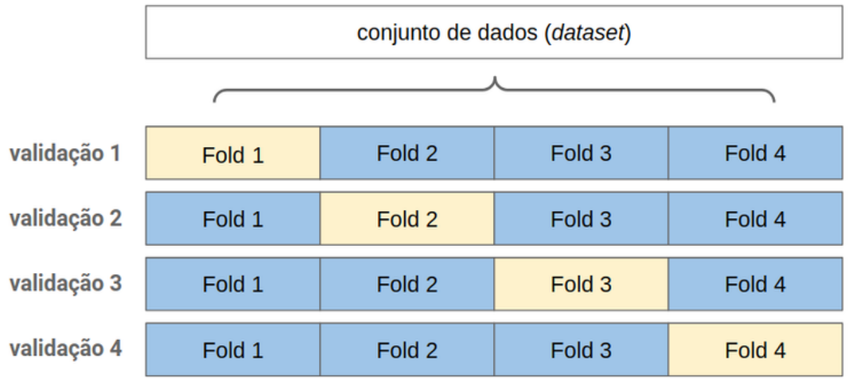
Data Preprocessing

StandardScaler
OneHotEncoder

Model Selection

LinearRegression
SVM
RandomForest



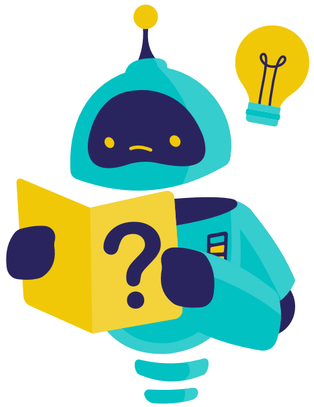


LR RMSE 1487.895065 STD 4758612.920399
SVM RMSE 1533.414071 STD 5249152.672435
RandomForest RMSE 1606.268654 STD 4787254.790045

建立模型步驟

LinearRegression

pipeline.score(X_train,(y_train))



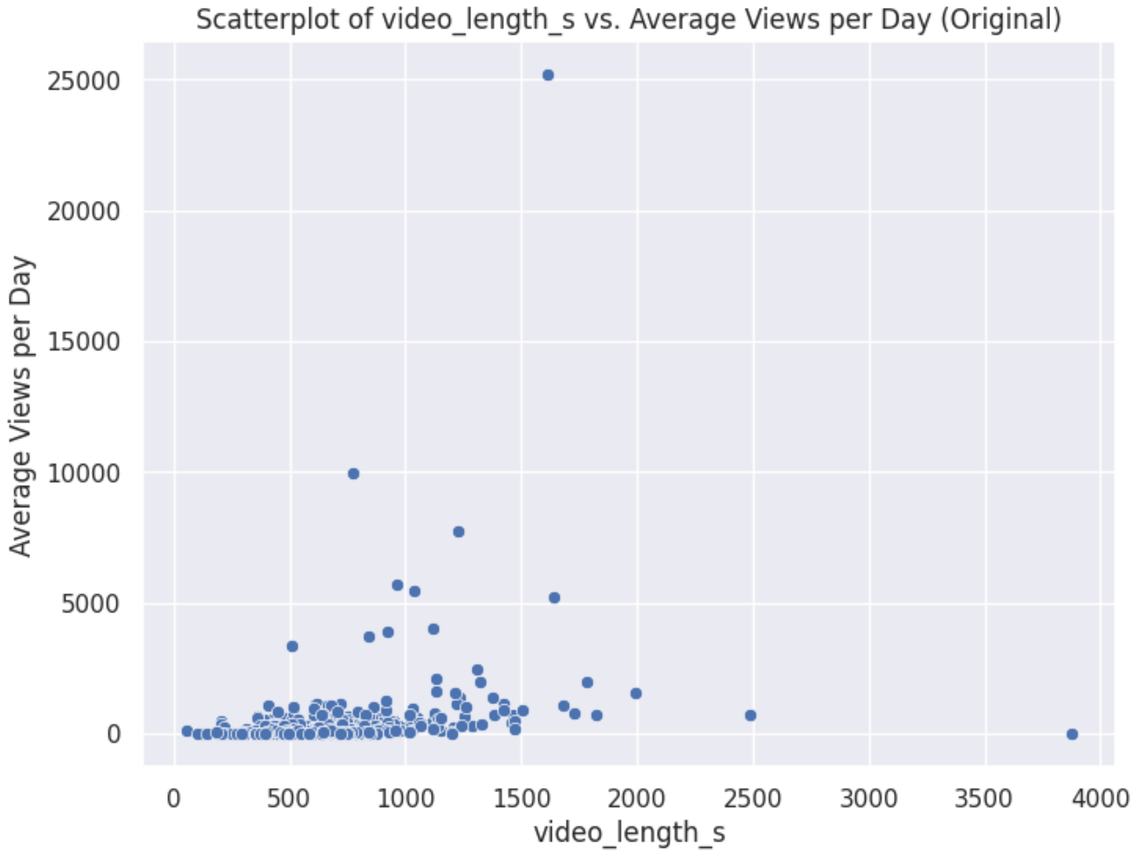
0.15

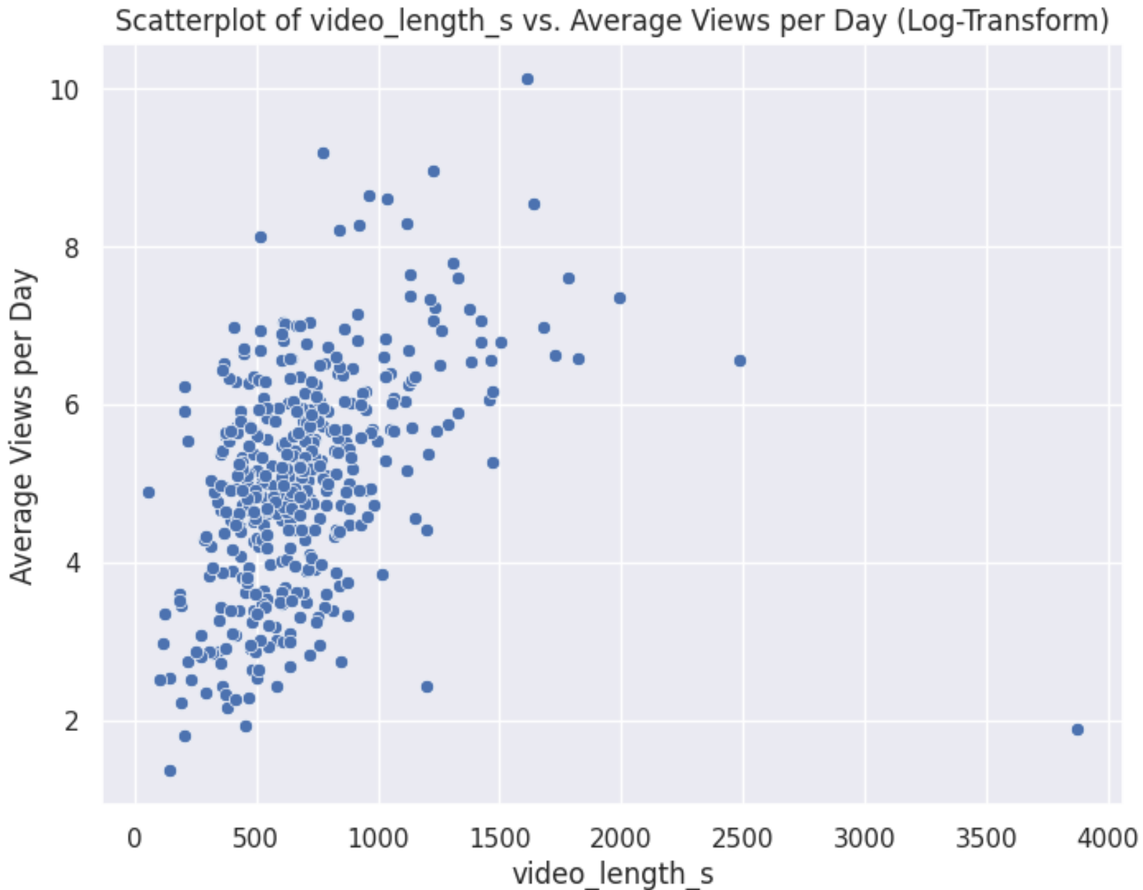
```
                            OLS Regression Results
==============================================================================
Dep. Variable:       avg_day_of_views   R-squared:                       0.150
Model:                            OLS   Adj. R-squared:                  0.127
Method:                 Least Squares   F-statistic:                     6.485
Date:                Thu, 07 Sep 2023   Prob (F-statistic):           1.65e-08
Time:                        18:28:50   Log-Likelihood:                -2959.6
No. Observations:                 341   AIC:                             5939.
Df Residuals:                     331   BIC:                             5978.
Df Model:                           9
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          -1067.6314    491.425     -2.173      0.031   -2034.342    -100.921
title_text_lenght    28.9092      6.627      4.362      0.000      15.872      41.946
video_length_s       0.7729      0.263      2.943      0.003       0.256       1.290
day_of_week_1        8.5913    607.515      0.014      0.989   -1186.486    1203.668
day_of_week_2      -16.6581    543.611     -0.031      0.976   -1086.025    1052.709
day_of_week_3      -130.1261    504.179     -0.258      0.796   -1121.925     861.673
day_of_week_4       22.2582    575.722      0.039      0.969   -1110.278    1154.794
day_of_week_5      323.1581    540.527      0.598      0.550    -740.143    1386.460
day_of_week_6      -136.2266    476.931     -0.286      0.775   -1074.425     801.971
feat_or_not_1      -129.4355    255.762     -0.506      0.613    -632.559     373.688
==============================================================================
Omnibus:                      654.073   Durbin-Watson:                   2.044
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           481115.810
Skew:                          11.985   Prob(JB):                         0.00
Kurtosis:                     185.447   Cond. No.                     1.22e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.22e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

建立模型步驟

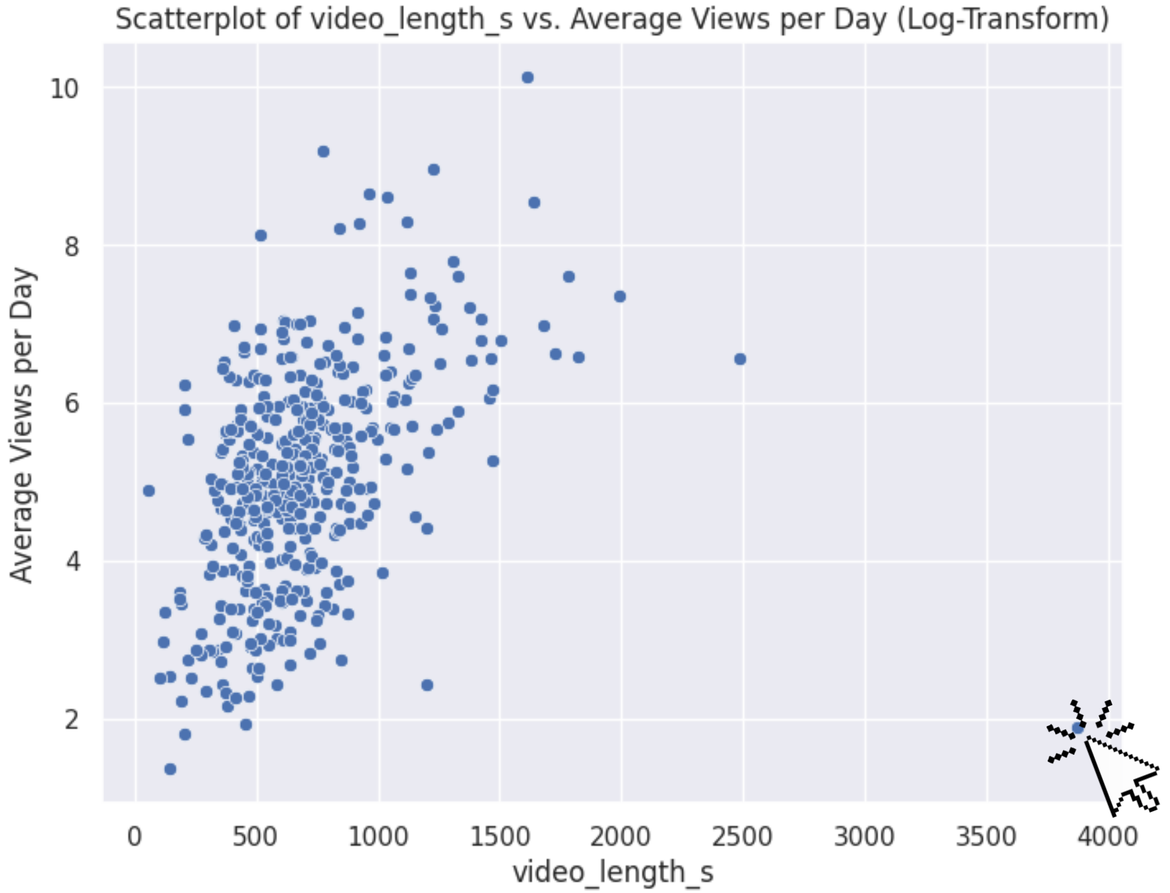
Improve the Model





建立模型步驟

Improve the Model



建立模型步驟

LinearRegression

pipeline.score(X_train,(y_train))

0.43

建立模型步驟

Testing Data

Final Test Error (RMSE):
1134.59947948093

The End