Youtube数据挖掘

1.引包

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import matplotlib as mpl
import os
from subprocess import check_output
from wordcloud import WordCloud, STOPWORDS
```

2. 数据引入

```
df_youtube_us = pd.read_csv("archive/USvideos.csv")
```

3.数据摘要

3.1数据示例

```
df_youtube_us.head()
```

	video_id	trending_date	title	channel_title	category_id	publish_time	tags	views
0	2kyS6SvSYSE	17.14.11	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	2017-11-13T17:13:01.000Z	SHANtell martin	748374
1	1ZAPwfrtAFY	17.14.11	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	2017-11-13T07:30:00.000Z	last week tonight trump presidency|"last week ...	2418783
2	5qpjK5DgCt4	17.14.11	Racist Superman | Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	2017-11-12T19:05:24.000Z	racist superman|"rudy"|"mancuso"|"king"|"bach"...	3191434
3	puqaWrEC7tY	17.14.11	Nickelback Lyrics: Real or Fake?	Good Mythical Morning	24	2017-11-13T11:00:04.000Z	rhett and link|"gmm"|"good mythical morning"|"...	343168
4	d380meD0W0M	17.14.11	I Dare You: GOING BALD!?	nigahiga	24	2017-11-12T18:01:41.000Z	ryan|"higa"|"higatv"|"nigahiga"|"i dare you"|"...	2095731

3.2 数据总个数

```
df_youtube_us.shape
```

```
(40949, 16)
```

3.3 数据频数

3.3.1 各属性的独立数据个数

```
for column_name in df_youtube_us:
    print("-------------分割线-------------")
    print("%s的总描述:\n"%(column_name),df_youtube_us[column_name].describe())
    print("%s的每个值的频次:\n"%(column_name),df_youtube_us[column_name].value_counts())
```

```
-------------分割线-------------
video_id的总描述:
 count           40949
unique           6351
top       j4KvrAUjn6c
freq               30
Name: video_id, dtype: object
video_id的每个值的频次:
 j4KvrAUjn6c    30
MAjY8mCTXWk    29
r-3iathMo7o    29
NBSAQenU2Bk    29
QBL8IRJ5yHU    29
               ..
2X5fG1A4kuc     1
A8QY1IRPGTY     1
zYWt2mnalP8     1
yW6ORWYn3g0     1
ufOnBRXy3Pc     1
Name: video_id, Length: 6351, dtype: int64
-------------分割线-------------
trending_date的总描述:
 count        40949
unique         205
top       18.05.06
freq           200
Name: trending_date, dtype: object
trending_date的每个值的频次:
 18.05.06    200
18.03.06    200
18.11.06    200
17.13.12    200
18.07.05    200
           ...
18.01.02    197
18.31.01    197
18.02.02    196
18.03.02    196
18.04.02    196
Name: trending_date, Length: 205, dtype: int64
-------------分割线-------------
title的总描述:
 count                                          40949
unique                                          6455
top       WE MADE OUR MOM CRY...HER DREAM CAME TRUE!
freq                                              30
Name: title, dtype: object
title的每个值的频次:
 WE MADE OUR MOM CRY...HER DREAM CAME TRUE!                                      30
Why I'm So Scared (being myself and crying too much)                            29
Rooster Teeth Animated Adventures - Millie So Serious                           29
Charlie Puth - BOY [Official Audio]                                             29
Mission: Impossible - Fallout (2018) - Official Trailer - Paramount Pictures    29
                                                                                ..
Black Mirror - U.S.S. Callister | Official Trailer [HD] | Netflix                1
2018 Winter Olympics Recap Day 11 I Part 1 I NBC Sports                          1
Fusion 360 to 3D Print Parts to Build a Desk                                     1
Behind The Curtain - OFFICIAL TRAILER                                            1
Natalie's 2nd Rap - SNL                                                          1
Name: title, Length: 6455, dtype: int64
-------------分割线-------------
channel_title的总描述:
 count     40949
unique     2207
top        ESPN
freq        203
Name: channel_title, dtype: object
channel_title的每个值的频次:
 ESPN                                      203
The Tonight Show Starring Jimmy Fallon    197
Vox                                       193
Netflix                                   193
TheEllenShow                              193
                                         ...
Dean Anderson                               1
Caterham Cars                               1
Tu Fan                                      1
Super Netvid                                1
CNLohr                                      1
Name: channel_title, Length: 2207, dtype: int64
```

```
-------------分割线-------------
category_id的总描述:
 count    40949.000000
mean        19.972429
std          7.568327
min          1.000000
25%         17.000000
50%         24.000000
75%         25.000000
max         43.000000
Name: category_id, dtype: float64
category_id的每个值的频次:
 24    9964
10    6472
26    4146
23    3457
22    3210
25    2487
28    2401
1     2345
17    2174
27    1656
15     920
20     817
19     402
2      384
29      57
43      57
Name: category_id, dtype: int64
-------------分割线-------------
publish_time的总描述:
 count                        40949
unique                        6269
top       2018-05-18T14:00:04.000Z
freq                            50
Name: publish_time, dtype: object
publish_time的每个值的频次:
 2018-05-18T14:00:04.000Z    50
2018-05-06T13:00:05.000Z    32
2018-05-13T18:03:56.000Z    30
2018-05-14T13:00:01.000Z    29
2018-05-14T14:00:03.000Z    29
                            ..
2017-11-15T02:17:29.000Z     1
2018-01-08T16:00:02.000Z     1
2018-02-14T11:00:00.000Z     1
2017-12-28T16:03:03.000Z     1
2017-11-24T13:00:06.000Z     1
Name: publish_time, Length: 6269, dtype: int64
-------------分割线-------------
tags的总描述:
 count      40949
unique      6055
top       [none]
freq        1535
Name: tags, dtype: object
tags的每个值的频次:
 [none]


                                                                                                           1535
ABC|"americanidol"|"idol"|"american idol"|"ryan"|"seacrest"|"ryan seacrest"|"katy"|"perry"|"katy perry"|"luke"|"bryan"|"luke
bryan"|"lionel"|"richie"|"lionel richie"|"season 16"|"american idol
XVI"|"television"|"ad"|"spring"|"2018"|"music"|"reality"|"competition"|"song"|"sing"|"audition"|"auditions"|"performance"|"live"|"fox"|"AI"
|"hollywood"|"contestant"|"official"|"american"|"official american idol"|"hollywood week"|"hometown audition"
                                                          87
Jacksfilms|"Jack Douglass"|"YGS"|"YGS 100"|"YGS 50"|"The Best of Your Grammar Sucks"|"Your Grammar Sucks"|"YIAY"|"Yesterday I Asked
You"|"Fidget Spinners"|"Emoji Movie"|"Kermit Sings"|"JackAsk"|"Jack Ask"|"Dubstep Solves Everything"|"Frozen 2"|"iPhone Parody"|"Apple
Parody"
                                                                                                                         80
James Corden|"The Late Late Show"|"Colbert"|"late night"|"late night show"|"Stephen
Colbert"|"Comedy"|"monologue"|"comedian"|"impressions"|"celebrities"|"carpool"|"karaoke"|"CBS"|"Late Late
Show"|"Corden"|"joke"|"jokes"|"funny"|"funny video"|"funny videos"|"humor"|"celebrity"|"celeb"|"hollywood"|"famous"

                                                            71
The Late Show|"Stephen Colbert"|"Colbert"|"Late Show"|"celebrities"|"late night"|"talk show"|"skits"|"bit"|"monologue"|"The Late Late
Show"|"Late Late Show"|"letterman"|"david letterman"|"comedian"|"impressions"|"CBS"|"joke"|"jokes"|"funny"|"funny video"|"funny
videos"|"humor"|"celebrity"|"celeb"|"hollywood"|"famous"|"James Corden"|"Corden"|"Comedy"
                                                                                                                              66



                                                                                                           ...
First we feast|"fwf"|"firstwefeast"|"food"|"food porn"|"cook"|"cooking"|"chef"|"kitchen"|"recipe"|"cocktail"|"bartender"|"craft
beer"|"complex"|"complex media"|"Cook (Profession)sean evans"|"Food Skills"|"Mark Iacono"|"Lucali"|"margherita pie"|"margherita"|"Dom
DeMarco"|"beyonce and jay z lucali"|"pizza"|"brooklyn"|"ugly delicious mark Iacono"|"how to make pizza"|"how to make tomato sauce"
                                                                                                                                1
Georgia Dome|"Implosion"|"2017"|"November 20"|"The Weather Channel"|"Guy Losing It Live on
Air"|"Swearing"|"MARTA"|"bus"|"ironic"|"monologue"


                      1
```

```
buzzfeed|"buzzfeedvideo"|"buzzfeed video"|"meghan markle"|"megan markle"|"prince harry"|"prince william"|"princess kate"|"kate
middleton"|"royal engagement"|"royal wedding"|"princess diana"|"british"|"americans"|"london"|"princess charlotte"|"prince george"|"prince
charles"|"funny"|"royal family"|"kate"|"queen elizabeth"|"uk"|"uk news"|"royals"|"wedding"|"engagement"|"harry"|"harry and
meghan"|"markle"|"royal"|"prince harry and meghan markle"|"meghan"|"the royals"|"kensington palace"|"the queen"|"prince harry engagement"
    1
introverts vs extroverts|"hannah"|"stocking"|"introverts"|"vs"|"extroverts"|"high school rivalry"|"how to control your boyfriend"|"dating a
pathological liar"|"High School Rivalry | Hannah Stocking & Supreme
Patty"|"lelepons"|"hannahstocking"|"rudymancuso"|"inanna"|"anwar"|"sarkis"|"shots"|"shotsstudios"|"alesso"|"anitta"|"brazil"

                                              1
tamar|"tamar braxton"|"tamar braxton vince herbert"|"tamar and vince"|"vince herbert"|"tamar braxton talk show"|"tamar vince
interview"|"tamar vince season 5"|"tamar braxton marriage"|"tamar vince divorce"|"tamar chronicles"|"the braxtons"|"braxton family
values"|"tamar and vince trailer"|"tamar braxton interview"|"tamar and vince premiere"
                                                                                                                                        1
Name: tags, Length: 6055, dtype: int64
-------------分割线-------------
views的总描述:
 count    4.094900e+04
mean     2.360785e+06
std      7.394114e+06
min      5.490000e+02
25%      2.423290e+05
50%      6.818610e+05
75%      1.823157e+06
max      2.252119e+08
Name: views, dtype: float64
views的每个值的频次:
 235077     3
427242     3
8493       3
168135     3
241387     3
          ..
82493      1
2050620    1
487995     1
500734     1
268291     1
Name: views, Length: 40478, dtype: int64
-------------分割线-------------
likes的总描述:
 count    4.094900e+04
mean     7.426670e+04
std      2.288853e+05
min      0.000000e+00
25%      5.424000e+03
50%      1.809100e+04
75%      5.541700e+04
max      5.613827e+06
Name: likes, dtype: float64
likes的每个值的频次:
 0         172
2          34
39         22
6          21
8          20
         ...
32868       1
104551      1
8304        1
67014       1
26813       1
Name: likes, Length: 29850, dtype: int64
-------------分割线-------------
dislikes的总描述:
 count    4.094900e+04
mean     3.711401e+03
std      2.902971e+04
min      0.000000e+00
25%      2.020000e+02
50%      6.310000e+02
75%      1.938000e+03
max      1.674420e+06
Name: dislikes, dtype: float64
dislikes的每个值的频次:
 0         383
1         159
4         120
5         101
2          97
         ...
3278        1
25648       1
9451        1
189472      1
2047        1
Name: dislikes, Length: 8516, dtype: int64
-------------分割线-------------
comment_count的总描述:
 count    4.094900e+04
mean     8.446804e+03
std      3.743049e+04
min      0.000000e+00
```

```
25%      6.140000e+02
50%      1.856000e+03
75%      5.755000e+03
max      1.361580e+06
Name: comment_count, dtype: float64
comment_count的每个值的频次:
 0        760
1         74
4         71
3         61
8         52
        ...
10120      1
12169      1
10525      1
8079       1
26597      1
Name: comment_count, Length: 13773, dtype: int64
-------------分割线-------------
thumbnail_link的总描述:
 count                                              40949
unique                                              6352
top       https://i.ytimg.com/vi/j4KvrAUjn6c/default.jpg
freq                                                  30
Name: thumbnail_link, dtype: object
thumbnail_link的每个值的频次:
 https://i.ytimg.com/vi/j4KvrAUjn6c/default.jpg    30
https://i.ytimg.com/vi/t4pRQ0jn23Q/default.jpg    29
https://i.ytimg.com/vi/iILJvqrAQ_w/default.jpg    29
https://i.ytimg.com/vi/r-3iathMo7o/default.jpg    29
https://i.ytimg.com/vi/NBSAQenU2Bk/default.jpg    29
                                                  ..
https://i.ytimg.com/vi/RK_B4Ez4_5Q/default.jpg     1
https://i.ytimg.com/vi/rovAxg5A48Q/default.jpg     1
https://i.ytimg.com/vi/eDelIZDzmwQ/default.jpg     1
https://i.ytimg.com/vi/yc0kcGgg3o0/default.jpg     1
https://i.ytimg.com/vi/c_KdAO6MqiM/default.jpg     1
Name: thumbnail_link, Length: 6352, dtype: int64
-------------分割线-------------
comments_disabled的总描述:
 count     40949
unique        2
top       False
freq      40316
Name: comments_disabled, dtype: object
comments_disabled的每个值的频次:
 False    40316
True       633
Name: comments_disabled, dtype: int64
-------------分割线-------------
ratings_disabled的总描述:
 count     40949
unique        2
top       False
freq      40780
Name: ratings_disabled, dtype: object
ratings_disabled的每个值的频次:
 False    40780
True       169
Name: ratings_disabled, dtype: int64
-------------分割线-------------
video_error_or_removed的总描述:
 count     40949
unique        2
top       False
freq      40926
Name: video_error_or_removed, dtype: object
video_error_or_removed的每个值的频次:
 False    40926
True        23
Name: video_error_or_removed, dtype: int64
-------------分割线-------------
description的总描述:
 count                                                 40379
unique                                                 6901
top       ► Listen LIVE: http://power1051fm.com/\n► Face...
freq                                                     58
Name: description, dtype: object
description的每个值的频次:
```

► Listen LIVE: http://power1051fm.com/\n► Facebook: https://www.facebook.com/Power1051NY/\n► Twitter: https://twitter.com/power1051/\n► Instagram: https://www.instagram.com/power1051/

58

Jukin Media Verified (Original) * For licensing / permission to use: Contact - licensing(at)jukinmediadotcom\nSubmit your videos here: http://bit.ly/2iFnUya

46

My Twitter: https://twitter.com/prozdkp\nMy Let's Play channel, Press Buttons n
Talk:\nhttps://www.youtube.com/channel/UCSHsNH4FZXFeSQMJ56AdrBA\nMy Merch/T-Shirt Store: http://www.theyetee.com/prozd\nMy Tumblr:
http://prozdvoices.tumblr.com/\nMy Twitch: https://www.twitch.tv/prozd\nMy Instagram: https://instagram.com/prozd\nMy Patreon:
http://www.patreon.com/prozd\nUse the link below and the coupon code PROZDSNACKS to get $3 off your first Japan Crate Premium or
Original:\nhttp://japancrate.com/?tap_a=13976-19476b&tap_s=76467-12d24b\nUse the link below and the coupon code PROZDRAMEN to get $3 off
your first Umai Crate:\nhttp://japancrate.com/umai?tap_a=18655-b8af8b&tap_s=76467-12d24b\nUse the link below to get a free 14-day trial of
Funimation anime streaming:\nhttps://www.funimation.com/prozd\nUse the link below and coupon code PROZD10 to get $10 off any Classic Bokksu
subscription:\nhttp://www.bokksu.com?rfsn=498614.9d328&utm_source=refersion&utm_medium=influencers&utm_campaign=498614.9d328\nUse the link
below and the coupon code ProZDCrate to get 10% off any Loot Crate:\nhttps://lootcrate.com/ProZD

38

Fortnite, PUBG, Far Cry 5? Which game would you play on this gaming PC setup?Visit SteelSeries.com and use discount code “Unbox15”(letters
in discount code ARE case sensitive) to get an Unbox 24hr exclusive of 15% off Arctis Pro + GameDAC: http://steelseries.com/arctisproThe
Chair - https://amzn.to/2Km7gC6The Monitor - https://amzn.to/2jWuQdkThe Gaming PC - https://www.xidax.com/(More info on gaming PC specs
etc. in this video - https://youtu.be/Pvakr7s7qc0)Is this the ultimate gaming PC setup?____WATCH SOME
MORE VIDEOS...Get The OnePlus 6 EARLY!https://youtu.be/yCxwmH3psxg?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34Should You Buy The Samsung Galaxy
S9?https://youtu.be/SIR67et5tcs?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The True All-Screen Smartphone is
Here...https://youtu.be/sYvH7Y16iUM?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The TRUTH About Smartphones in 2018https://youtu.be/1kllbOrLfoo?
list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34World's Biggest Fortnite Gaming Setup!https://youtu.be/8x7UtZKwfHA?list=PL7u4lWXQ3wfI_7PgX0C-
VTiwLeu0S4v34The Weirdest Phones In The World...https://youtu.be/o6T9mUq9Vgo?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The Coolest Smartphone
You'll Never Touch...https://youtu.be/5M3mKgLTn3Q?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34I'm Switching To The Samsung Galaxy
S9https://youtu.be/8g-VjqONplA?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The Most Expensive iPhone I've Ever
Seen...https://youtu.be/JUi3psxB3QA?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The Limited Smartphone You Never Knew
Existed...https://youtu.be/SMLgNZYW3XE?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The Almost All-Screen
Smartphone...https://youtu.be/jAq9RV3k9Qc?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34TOP SECRET SMARTPHONE
DELIVERYhttps://youtu.be/BNnFgT_CAEE?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The iPhone X Home Button... Is This Real Life?
https://youtu.be/Vz_EE5Ta9ZA?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34Fortnite on an INSANE $20,000 Gaming PChttps://youtu.be/Pvakr7s7qc0?
list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The $200 Smartphone You NEED To Know About...https://youtu.be/uxLOfjaWRvw?list=PL7u4lWXQ3wfI_7PgX0C-
VTiwLeu0S4v34This New Smartphone Is NOT What It Looks Like...https://youtu.be/r8vFZ0HAaz0?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34Is The
Samsung Galaxy S9 Worth The Hype?https://youtu.be/g30Rhk82rmg?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v343 Unique Gadgets You Wouldn't Expect
To Existhttps://youtu.be/z5ydE6qQqZU?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The Worst Gadget EVER On Unbox
Therapy...https://youtu.be/ZOFoPTAqZlQ?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The Worst Text You Could Ever
Receive...https://youtu.be/HUE9mCN7sek?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The Essential Phone Is Back!https://youtu.be/ZxOmJfCEgoc?
list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34What If You Could Get AirPods For Only $40? https://youtu.be/6N5V_7_n1uI?list=PL7u4lWXQ3wfI_7PgX0C-
VTiwLeu0S4v34I Bought The Cheapest Smartphone on Amazon...https://youtu.be/YkGAg9WmYBs?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v343 Unique
Gadgets You Can Buy Right Nowhttps://youtu.be/Yzsf9SECcEo?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34DON'T Buy The Google Pixel
Budshttps://youtu.be/lGkrhR2mfl8?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34How To Turn Any Android Phone Into An
iPhone...https://youtu.be/14pYNywLqDs?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34Is The LG V30 The Most Underrated Smartphone?
https://youtu.be/YsWIHhKmmvY?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The Best Wireless Headphones You Can Buy Right
Nowhttps://youtu.be/SXyObZahu-o?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34Unboxing The Samsung Galaxy S9 Clonehttps://youtu.be/1xgbmrsgrq4?
list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34It Has Double The Battery of iPhone Xhttps://youtu.be/8Np9Kk82-zA?list=PL7u4lWXQ3wfI_7PgX0C-
VTiwLeu0S4v34The Mind Blowing 33 Million Pixel Display...https://youtu.be/OKAU1Xx59ho?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v345 Cool Gadgets
Under $10https://youtu.be/hNrSNrEVpkQ?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34Which Smartphone Do They ACTUALLY Use? --- MKBHD, Austin
Evans, Linus + Morehttps://youtu.be/Hi2tjMLVpdQ?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34Unboxing The World's Smallest
Phonehttps://youtu.be/SSzyGCjH88o?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34The Most RIDICULOUS MacBook Prohttps://youtu.be/46qTg3swoEo?
list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34A Message from Apple...https://youtu.be/UiaqBdzCcBA?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v344 Unique
iPhone Accessorieshttps://youtu.be/uZgnXJz_9DM?list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34DON'T Buy The iPhone Xhttps://youtu.be/2fGXDFiFBhg?
list=PL7u4lWXQ3wfI_7PgX0C-VTiwLeu0S4v34FOLLOW ME IN THESE PLACES FOR UPDATESTwitter - http://twitter.com/unboxtherapyFacebook -
http://facebook.com/lewis.hilsentegerInstagram - http://instagram.com/unboxtherapyGoogle Plus - http://bit.ly/1auEeak 29

I will never be able to say Thank You enough... Thank you for being my family.➡ CLICK HERE - http://bit.ly/GiveAGatorItsWings➡ SUBSCRIBE TO MY 2ND CHANNEL!: http://bit.ly/2hsXpQd➡ ADD ME ON SNAPCHAT: BM885_ _♡ GRAV3YARD CURL COLLECTION: http://bit.ly/Grav3yardCurl♡ Entire Set - http://bit.ly/Grav3yardCurlSet♡ Gator Hairdryer: http://bit.ly/GatorHairdryer♡ Gator Flat Iron: http://bit.ly/GatorFlatIron♡ Gator Clipless Curler: http://bit.ly/GatorCliplessCurler♡ INSTAGRAM: http://bit.ly/1wdGBwS ♡ TWITTER: http://twitter.com/grav3yardgirl♡ FACEBOOK: http://bit.ly/2ktztLnyou might enjoy these other videos?EDIBLE JELLO GLASSES: http://bit.ly/EdibleJelloGlasses$500 Designer Mystery Box: http://bit.ly/LuxuryMysteryBox$900 Ebay 90s Mystery Box: http://bit.ly/90sMysteryBoxLucky Bag 2018: http://bit.ly/LuckyBag2018Grav3yardgirlMaking a MINIATURE Happy Meal: http://bit.ly/MiniatureHappyMealWUBBLE BUBBLE BALL: http://bit.ly/WubbleBubbleGrav3yardgirlSNO CONE SLIME DIY: http://bit.ly/SnoConeSlimeFishbowl Slime DIY: http://bit.ly/FishbowlSlime♡ EVERYTIME YOU SUBSCRIBE, A GATOR GETS HIS WINGS! ♡FTC- I am not being paid by any of the mentioned companies or designers to make this video. The views in this video are strictly my own and I am not affiliated with any of these companies.

29

..

Did you catch our ad during the BIG Game? See the full, uncut version and go to gasfireplaces.com to save 80% off your next gas fireplace.\n\nSchedule your showroom visit: gasfireplaces.com

1

Khloe and Corey hash out there issues during an intense lunch. Plus, she and Tristan Thompson break the news that they're expecting.\n\nSUBSCRIBE: http://bit.ly/Eentsub\n\nAbout Keeping Up With the Kardashians:\nThings change, but this famous family stays the same. Can you keep up with the drama?\n\nConnect with the Kardashians:\nVisit the KUWTK WEBSITE: http://bit.ly/KUWTKweb\nLike KUWTK on FACEBOOK: http://bit.ly/KUWTKfb \nFollow KUWTK on TWITTER: http://bit.ly/KUWTKtwtr \n\nAbout E! Entertainment:\nE! is on the Pulse of Pop Culture, bringing fans the very best original content including reality series, scripted programming, exclusive specials, breaking entertainment news, streaming events and more. Passionate viewers can’t get enough of our Pop Culture hits including Keeping Up with the Kardashians,” “Fashion Police,” “The Royals,” Total Divas” and “Botched.” And with new original programming on the way, fans have even more to love.\n\nConnect with E! Entertainment:\nVisit the E! WEBSITE: http://eonli.ne/1iX6d8n\nLike E! on FACEBOOK: http://on.fb.me/1fzeamg\nCheck out E! on INSTAGRAM: http://bit.ly/EInsta\nFollow E! on TWITTER: http://bit.ly/EEntTwitter\n\nKeeping Up With the Kardashians Katch-Up: S14, EP.14 | E!\nhttp://www.youtube.com/user/Eentertainment

1

화려한 골든 에이지의 시작!\nWanna One 2nd Mini Album 0+1=1 (I PROMISE YOU)\n\n'BOOMERANG(부메랑)'은 자신감 강한 남자의 거부할 수 없는 짝사랑이\n사랑하는 그녀와 만나 다시 그에게 돌아오길 바라는 마음과\n영원히 그녀만을 바라보겠다는 약속을 담은 곡이다.\n\nMore About Wanna One @ \nOfficial Facebook: https://www.facebook.com/WannaOne.official \nOfficial Twitter: https://twitter.com/WannaOne_twt \nOfficial Instagram: https://www.instagram.com/wannaone.official \nOfficial Post: http://post.naver.com/wannaone_official \nOfficial fan cafe: http://cafe.daum.net/WannaOneOfficial\n\n♬ Available on iTunes, Apple Music : https://apple.co/2FWt5Gz\n\nCJ E&M Music은 아시아 No.1 엔터테인먼트 기업인 CJ E&M의 음악사업 브랜드로 음원/음반의 투자/제작/유통부터 콘서트/페스티벌 개최까지 포함하고 있습니다. CJ E&M MUSIC과 함께 하는 K-POP 아티스트들의 신곡과 뮤직비디오, 미공개 독점 영상 등을 이곳 YOUTUBE 채널에서 가장 먼저 만나보세요.\n\nCJ E&M Music is a music business brand of CJ E&M, Asia's No.1 entertainment company. CJ E&M Music covers investment, production and distribution of album and also provides the best music festival and concerts. Meet the K-POP artists' brand new music videos and exclusive video clips on the official YouTube of CJ E&M Music.

1

Download or stream now at: https://atlanti.cr/thesedays\n\nThis is the official video for our new single These Days featuring Jess Glynne, Macklemore and Dan Caplen – out now everywhere!\n\nClick here to subscribe: http://bit.ly/SubscribeToRudimental\n\nFollow Rudimental\nhttp://www.rudimental.co.uk \nhttp://www.facebook.com/rudimentaluk \nhttp://www.twitter.com/rudimentaluk \nhttp://www.instagram.com/rudimentaluk\nhttp://www.soundcloud.com/rudimentaluk\n\nFollow Jess Glynne\nhttp://jessglynne.co.uk\nhttps://www.facebook.com/JessGlynne\nhttps://twitter.com/jessglynne\nhttps://www.instagram.com/jessglynne\n\nFollow Dan Caplen\nhttp://dancaplen.com\nhttps://www.facebook.com/dancaplen\nhttps://twitter.com/dancaplen\nhttps://www.instagram.com/dancaplen\n\nFollow Macklemore\nhttp://macklemore.com\nhttps://www.facebook.com/macklemore\nhttps://twitter.com/macklemore\nhttps://www.instagram.com/macklemore\n\nDirector | Johnny Valencia\nExecutive Producer | Honna Kimmerer\nProducer | Shabana Mansuri\nProduction Designer | John Lavin\nCinematographer | Justin Henning\nEditor | Jason Koenig & Johnny Valencia\nVideo Commissioner | Dan Curwin

1

```
Singer Patti LaBelle tells Andy Cohen what advice she would and has given to goddaughter Mariah Carey and says what she thinks about some
negative people in Mariah’s life.\n►► Subscribe To WWHL: http://bravo.ly/WWHLSub\n\nWatch WWHL Sun–Thu 11/10c:\nWWHL Website:
http://www.bravotv.com/watch-what-happens-live\nFollow WWHL: https://twitter.com/BravoWWHL\nLike WWHL:
https://www.facebook.com/WatchWhatHappensLive\nWWHL Tumblr: http://bravowwhl.tumblr.com/\n\n'Watch What Happens: Live' is Bravo's late-
night, interactive talk show that features guests from the world of entertainment, politics, and pop culture. Hosted by Andy Cohen, the
series includes lively debates on everything from fashion, the latest on everyone's favorite Bravolebrities, and what celebrity is making
headlines that week. Past guests who have joined Cohen in the Bravo Clubhouse include Sarah Jessica Parker, Tina Fey, Khloe Kardashian,
Jennifer Lopez, Liam Neeson, Kelly Ripa, Jimmy Fallon, Anderson Cooper, Jennifer Lawrence, and Lance Bass.\n\nWatch More Bravo:\nBravo
Website: http://www.bravotv.com/\nBravo Youtube: http://www.youtube.com/videobybravo\nFollow Bravo: http://www.twitter.com/bravotv\nLike
Bravo: https://www.facebook.com/BRAVO\nPin Bravo: http://www.pinterest.com/bravobybravo\nBravo Instagram:
http://www.instagram.com/bravotv\nBravo Tumblr: http://bravotv.tumblr.com/\n\nAfter Show: Patti LaBelle’s Advice For Mariah Carey | WWHL
                                              1
Name: description, Length: 6901, dtype: int64
-------------分割线-------------
likes_log的总描述:
 count    40949.000000
mean         9.599392
std          2.115725
min          0.000000
25%          8.598773
50%          9.803225
75%         10.922660
max         15.540743
Name: likes_log, dtype: float64
likes_log的每个值的频次:
 0.000000     172
1.098612      34
3.688879      22
1.945910      21
2.197225      20
            ...
11.157621      1
10.869159      1
9.624699       1
12.643895      1
11.721410      1
Name: likes_log, Length: 29850, dtype: int64
-------------分割线-------------
views_log的总描述:
 count    40949.000000
mean        13.337995
std          1.709989
min          6.309918
25%         12.398056
50%         13.432583
75%         14.416081
max         19.232552
Name: views_log, dtype: float64
views_log的每个值的频次:
 12.394161    3
10.909034    3
12.367673    3
10.121056    3
9.047115     3
            ..
13.694479    1
10.859749    1
15.130452    1
16.684928    1
13.421313    1
Name: views_log, Length: 40478, dtype: int64
-------------分割线-------------
dislikes_log的总描述:
 count    40949.000000
mean         6.387610
std          1.915583
min          0.000000
25%          5.313206
50%          6.448889
```

```
75%           7.569928
max          14.330978
Name: dislikes_log, dtype: float64
dislikes_log的每个值的频次:
 0.000000     383
0.693147     159
1.609438     120
1.791759     101
1.098612      97
            ...
9.731156       1
9.946882       1
9.595535       1
8.099554       1
10.614892      1
Name: dislikes_log, Length: 8516, dtype: int64
-------------分割线-------------
comment_log的总描述:
 count    40949.000000
mean         7.387712
std          2.057100
min          0.000000
25%          6.421622
50%          7.526718
75%          8.657998
max         14.124157
Name: comment_log, dtype: float64
comment_log的每个值的频次:
 0.000000     760
0.693147      74
1.609438      71
1.386294      61
2.197225      52
            ...
9.212638       1
9.783577       1
9.863082       1
10.135948      1
9.498222       1
Name: comment_log, Length: 13773, dtype: int64
```

3.4 五数概括和缺失值个数

```python
# 纯数值的属性
number_list = ["category_id","views","likes","dislikes","comment_count"]
```

```python
# 输出指定数据、指定属性的5数概括及缺失值的个数
# column_name: 属性名
# data: 数据来源
def FiveNumberandNull(data,column_name):
    print("%s的情况如下: "%(column_name))
    # 缺失值的情况
    data_column = data[column_name]
    null_number = data_column.isnull().sum()
    print("null的个数:%d"%(null_number))
    # 五数概括 Minimum (最小值) 、Q1 (25%) 、Median (中位数、) 、Q3 (75%) 、Maximum (最大值)
    data_column = data_column.dropna(axis = 0) #删除NaN值
    Minimum = min(data_column)
    Maximum = max(data_column)
    Q1 = np.percentile(data_column, 25)
    Median = np.median(data_column)
    Q3 = np.percentile(data_column, 75)
    print("五数概括: Minimum:%d; Q1:%d; Median:%d; Q3:%d; Maximum:%d;\n"%(Minimum , Q1 , Median , Q3 , Maximum))
```

由数据分析可以发现，数据集中这些主要的数字数据不包含NULL

```python
for column_name in number_list:
    FiveNumberandNull(df_youtube_us,column_name)
```

```
category_id的情况如下:
null的个数:0
五数概括: Minimum:1; Q1:17; Median:24; Q3:25; Maximum:43;

views的情况如下:
null的个数:0
五数概括: Minimum:549; Q1:242329; Median:681861; Q3:1823157; Maximum:225211923;

likes的情况如下:
null的个数:0
五数概括: Minimum:0; Q1:5424; Median:18091; Q3:55417; Maximum:5613827;

dislikes的情况如下:
null的个数:0
五数概括: Minimum:0; Q1:202; Median:631; Q3:1938; Maximum:1674420;

comment_count的情况如下:
```

```
null的个数:0
五数概括: Minimum:0; Q1:614; Median:1856; Q3:5755; Maximum:1361580;
```

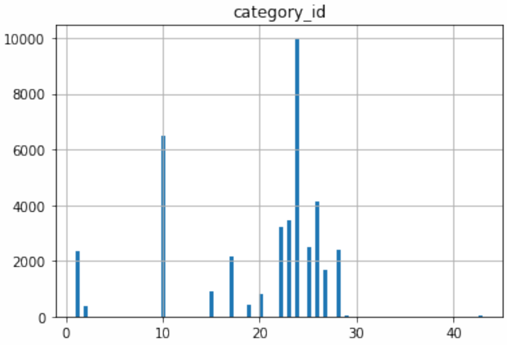
3.5 绘图

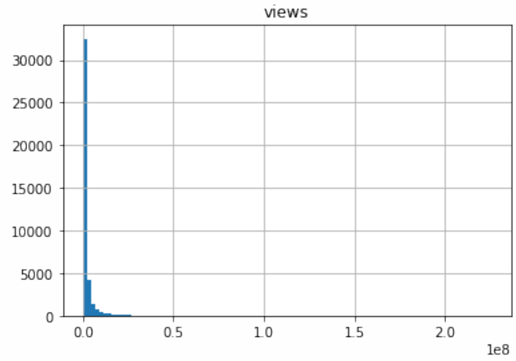
以直方图、盒图、分布图为例

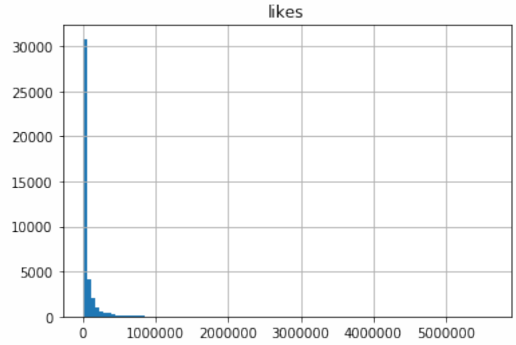
```
#绘制直方图:100bins
def histogram(column_name,data):
    histogram = data[column_name].hist(bins= 100)
    plt.title(column_name)
    plt.show()

#绘制盒图
def box(column_name,data):
    data[column_name].plot.box()
    plt.title(column_name)
    plt.show()
    df_yout['likes_log'] = np.log(df_yout['likes'] + 1)
# 绘制正态分布图
def logDistribution(data):
    data['likes_log'] = np.log(data['likes'] + 1)
    data['views_log'] = np.log(data['views'] + 1)
    data['dislikes_log'] = np.log(data['dislikes'] + 1)
    data['comment_log'] = np.log(data['comment_count'] + 1)

    plt.figure(figsize = (12,6))

    plt.subplot(221)
    g1 = sns.distplot(data['views_log'])
    g1.set_title("VIEWS LOG DISTRIBUITION", fontsize=16)

    plt.subplot(224)
    g2 = sns.distplot(data['likes_log'],color='green')
    g2.set_title('LIKES LOG DISTRIBUITION', fontsize=16)

    plt.subplot(223)
    g3 = sns.distplot(data['dislikes_log'], color='r')
    g3.set_title("DISLIKES LOG DISTRIBUITION", fontsize=16)

    plt.subplot(222)
    g4 = sns.distplot(data['comment_log'])
    g4.set_title("COMMENTS LOG DISTRIBUITION", fontsize=16)

    plt.subplots_adjust(wspace = 0.2, hspace = 0.4,top = 0.9)

    plt.show()
```

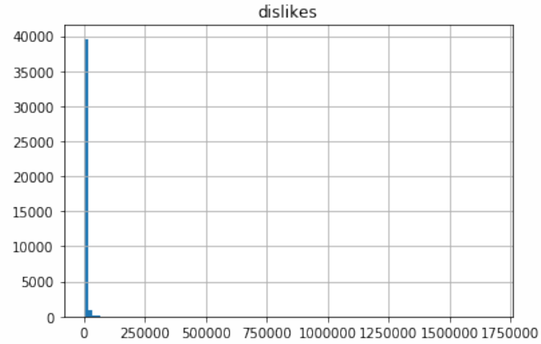
3.5.1 直方图

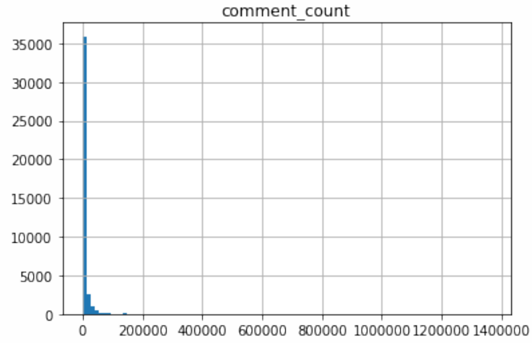
```
for column_name in number_list:
    histogram(column_name,df_youtube_us)
```







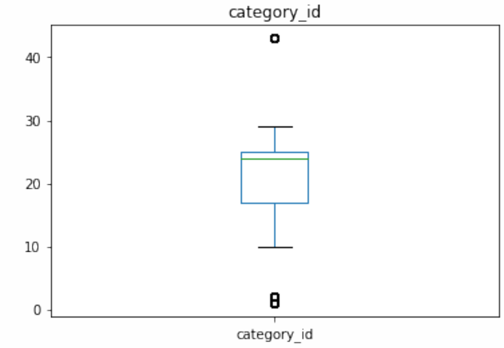


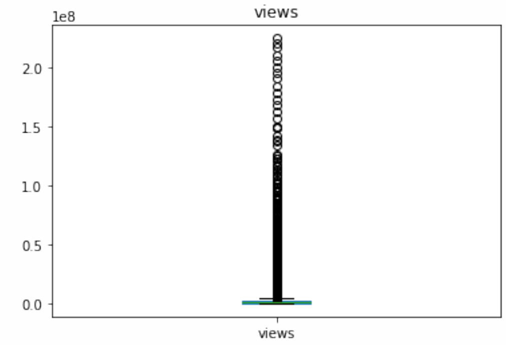


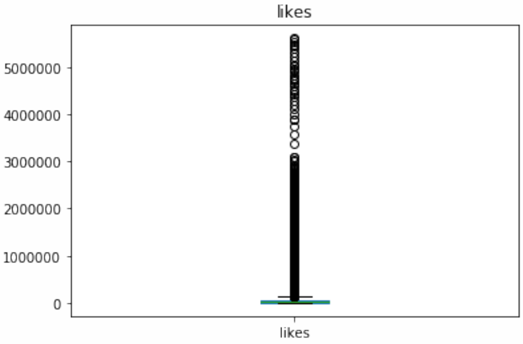
3.5.2 盒图

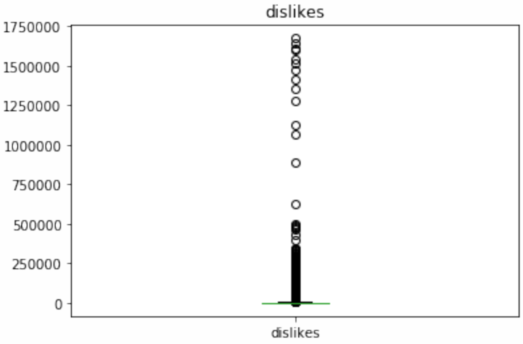
绝大多数的类别id处在10到30之间
观看次数的数量随着数值增大而减小，绝大多数位于1.5le8以下

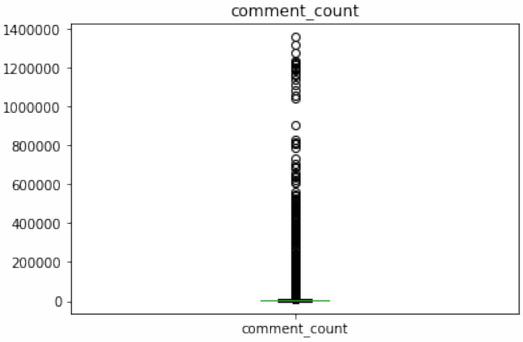
```
for column_name in number_list:
    box(column_name,df_youtube_us)
```


category_id
40
30
20
10
0
category_id


views
1e8
2.0
1.5
1.0
0.5
0.0
views


likes
5000000
4000000
3000000
2000000
1000000
0
likes


dislikes
1750000
1500000
1250000
1000000
750000
500000
250000
0
dislikes


comment_count
1400000
1200000
1000000
800000
600000
400000
200000
0
comment_count

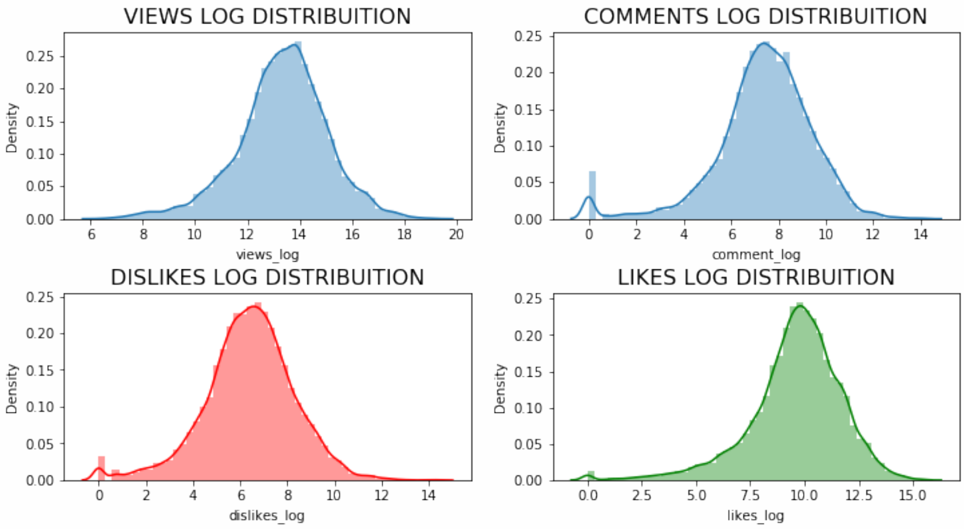
```
df_youtube_us.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 40949 entries, 0 to 40948
Data columns (total 17 columns):
video_id                  40949 non-null object
trending_date             40949 non-null object
title                     40949 non-null object
channel_title             40949 non-null object
category_id               40949 non-null int64
publish_time              40949 non-null object
tags                      40949 non-null object
views                     40949 non-null int64
likes                     40949 non-null int64
dislikes                  40949 non-null int64
comment_count             40949 non-null int64
thumbnail_link            40949 non-null object
comments_disabled         40949 non-null bool
ratings_disabled          40949 non-null bool
video_error_or_removed    40949 non-null bool
description               40379 non-null object
likes_log                 40949 non-null float64
dtypes: bool(3), float64(1), int64(5), object(8)
memory usage: 4.5+ MB
```

3.5.3 分布图

```
logDistribution(df_youtube_us)
```

```
/Users/guopeiqi/opt/anaconda3/envs/deeplearning/lib/python3.7/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/Users/guopeiqi/opt/anaconda3/envs/deeplearning/lib/python3.7/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/Users/guopeiqi/opt/anaconda3/envs/deeplearning/lib/python3.7/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
/Users/guopeiqi/opt/anaconda3/envs/deeplearning/lib/python3.7/site-packages/seaborn/distributions.py:2619: FutureWarning: `distplot` is a
deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```



3.6 缺失值处理

由于数据中的主要数值型数据不包含null值，所以不需要进行缺失值处理，若存在，可以使用以下方法进行处理，并进行可视化对比

```
# 剔除null，并可视化对比
def null_drop(data):
    data_nonull = data.dropna(axis = 0)
    # 绘制直方图
    hist = Data1.hist(bins=100)
    # 绘制盒图
    data_nonull.plot.box()
    plt.show()

# 用最高频率值来填补缺失值
def null_mode(data):
    data_nonull = data.fillna(data.mode()) #使用众数
    # 绘制直方图
    hist = data_nonull.hist(bins=100)
    # 绘制盒图
    data_nonull.plot.box()
    plt.show()
```