2023

PYTHON PROGRAMMING

BY
KIRAN AIGALI

Table of Contents

1. Import required libraries and read the provided dataset (youtube_dislike_dataset.csv) and retrieve top 5 and bottom 5 records.

```
# Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Read the Dataset
df = pd.read_csv('youtube_dislike_dataset.csv')
df
```

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	
0	--0bCF-iK2E	Jadon Sancho Magical Skills & Goals	UC6UL29enLNe4mqwTfAyeNuw	Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	footbal alemn
1	--14w5SOEUs	Migos - Avalanche (Official Video)	UCGIelM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migo Qu Music
2	--40TEbZ9Is	Supporting Actress in a Comedy: 73rd Emmys	UCIBKH8yZRcM4AsRjDVEdjMg	Television Academy	2021-09-20 01:03:32	925281	11212	401	831	
3	--4tfbSyYDE	JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO	UCsmXiDP8S40uBeJYxvyuImA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUC JO1 The
4	--DKkzWVh-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQIdVoKwjlQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retaini Jer Dir
...	...	...	...	...	...	...	...	...	...	
37417	zzd4ydafGR0	Lil Tjay - Calling My Phone (feat. 6LACK) [Off...	UCEB4a5o_6KfjxHwNMnmj54Q	Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	Lil Callin Call
37418	zziBybeSAtw	PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...	UCWJ2lWNubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NB game-
37419	zzk09ESX7e0	[MV] (MAMAMOO) - Where Are We Now	UCuhAUMLzJxIP1W7mEk0_6IA	MAMAMOO	2021-06-02 09:00:10	13346678	720854	4426	90616	MAMA WAW WAW W
37420	zzmQEb0Em5I	FELLIPE ESCUDERO- Master Podcast #12	UC8NjnNWMsRqq11NYvHAQb1g	Master Podcast	2020-10-20 20:59:30	252057	19198	1234	1471	master lord vinh
37421	zzxPZwaA-8w	Gareth Bale brace secures dramatic comeback on...	UCEg25rdRZXg32iwai6N6l0w	Tottenham Hotspur	2021-05-23 21:00:31	2252090	34063	868	2004	Spur Hotspu

37422 rows × 12 columns

```
# Display top 5 records

df.head()
```

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	
0	--0bCF-iK2E	Jadon Sancho Magical Skills & Goals	UC6UL29enLNe4mqwTfAyeNuw	Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football soccer l alemn Bundes seasc
1	--14w5SOEUs	Migos - Avalanche (Official Video)	UCGIelM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Avalar Quality Cor Music/Motown
2	--40TEbZ9Is	Supporting Actress in a Comedy: 73rd Emmys	UCIBKH8yZRdM4AsRjDVEdjMg	Television Academy	2021-09-20 01:03:32	925281	11212	401	831	
3	--4tfbSyYDE	JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO	UCsmXiDP8S40uBeJYxvyuImA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE101JA JO1 TheSTAR S
4	--DKkzWVh-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQIdVoKwjlQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retaining wall Jersey high Direct Con

```
# Display bottom 5 records

df.tail()
```

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	
37417	zzd4ydafGR0	Lil Tjay - Calling My Phone (feat. 6LACK) [Off...	UCEB4a5o_6KfjxHwNMnmj54Q	Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	Lil Tjay Calling My Calling M
37418	zziBybeSAtw	PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...	UCWJ2lWNubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NBA G L Bas game-00220 L
37419	zzk09ESX7e0	[MV] (MAMAMOO) - Where Are We Now	UCuhAUMLzJxlP1W7mEk0_6IA	MAMAMOO	2021-06-02 09:00:10	13346678	720854	4426	90616	MAMAMOO WAW MAM WAW Whe We
37420	zzmQEb0Em5I	FELLIPE ESCUDERO- Master Podcast #12	UC8NjnNWMsRqq11NYvHAQb1g	Master Podcast	2020-10-20 20:59:30	252057	19198	1234	1471	masterpodca lord vinhete
37421	zzxPZwaA-8w	Gareth Bale brace secures dramatic comeback on...	UCEg25rdRZXg32iwai6N6l0w	Tottenham Hotspur	2021-05-23 21:00:31	2252090	34063	868	2004	Spurs Tott Hotspur Tott Leice

This code imports the panda's library, reads the dataset, and then displays the top5 and the bottom5 records from the dataset.

2. Check the info of the data frame and write your inference on data types and shape of the dataset.

```
# Display information about the dataset

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37422 entries, 0 to 37421
Data columns (total 12 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   video_id       37422 non-null  object
 1   title          37422 non-null  object
 2   channel_id     37422 non-null  object
 3   channel_title  37422 non-null  object
 4   published_at   37422 non-null  object
 5   view_count     37422 non-null  int64 
 6   likes          37422 non-null  int64 
 7   dislikes       37422 non-null  int64 
 8   comment_count  37422 non-null  int64 
 9   tags           37422 non-null  object
 10  description    37422 non-null  object
 11  comments       37264 non-null  object
dtypes: int64(4), object(8)
memory usage: 3.4+ MB
```

```
df.dtypes
```

```
video_id         object
title            object
channel_id       object
channel_title    object
published_at     object
view_count        int64
likes             int64
dislikes          int64
comment_count     int64
tags             object
description      object
comments         object
dtype: object
```

```
df.shape
```

```
(37422, 12)
```

The shape of the dataset is (37422, 12), which it has 37,422 rows and 12 columns. This information confirms the number of records and attributes present in your dataset.

3. Check for the percentage of the missing values and drop or impute them.

```
# Precentage of missing values & drop or impute them

missing_percent = (df.isnull().sum() / len(df)) * 100
print("Percentage of missing values:")
print(missing_percent)
```

```
Percentage of missing values:
video_id          0.000000
title             0.000000
channel_id        0.000000
channel_title     0.000000
published_at      0.000000
view_count        0.000000
likes             0.000000
dislikes          0.000000
comment_count     0.000000
tags              0.000000
description       0.000000
comments          0.422212
dtype: float64
```

```
# Drop or impute missing values

df['comments'].fillna('', inplace = True)
threshold = 30
df.dropna(axis = 1, thresh = len(df) * (1 - threshold / 100), inplace = True)
df.dropna(axis = 0, inplace = True)
print((df.isnull().sum() / len(df)) * 100)
```

```
video_id          0.0
title             0.0
channel_id        0.0
channel_title     0.0
published_at      0.0
view_count        0.0
likes             0.0
dislikes          0.0
comment_count     0.0
tags              0.0
description       0.0
comments          0.0
dtype: float64
```

In the code above, the missing values are calculated as a percentage for each column. Then we can decide whether to drop or impute them based on our analysis. To impute the missing values in the ‘Comments’ column with an empty string and to drop the columns and rows based on a threshold percentage.

4. Check whether the statistical summary of both numerical and categorical columns and write your inferences.

```
# Numeric summary
numerical_summary = df.describe()
print("Statistical summary of numerical columns:")
print(numerical_summary)
```

```
Statistical summary of numerical columns:
         view_count         likes      dislikes  comment_count
count  3.742200e+04  3.742200e+04  3.742200e+04   3.742200e+04
mean   5.697838e+06  1.668147e+05  4.989862e+03   9.924930e+03
std    2.426622e+07  5.375670e+05  3.070824e+04   1.171003e+05
min    2.036800e+04  0.000000e+00  0.000000e+00   0.000000e+00
25%    5.122970e+05  1.323350e+04  2.810000e+02   9.000000e+02
50%    1.319078e+06  4.233050e+04  7.960000e+02   2.328000e+03
75%    3.670231e+06  1.304698e+05  2.461750e+03   6.184000e+03
max    1.322797e+09  3.183768e+07  2.397733e+06   1.607103e+07
```

- View count, likes, dislikes, and comment count are numerical columns with non-negative values, as indicated by the minimum values being greater then or equal to 0.
- You can see the counts, mean, standard deviation, minimum, 25th percentile, median 50th percentile, and 75th percentile, and maximum values for these columns.

```
# Categorical summary
categorical_summary = df.describe(include = ['object']).T
print("\nstatistical summary of categorical columns:")
print(categorical_summary)
```

```
statistical summary of categorical columns:
              count unique                         top  freq
video_id      37422  37422                --0bCF-iK2E     1
title         37422  37113                         www    21
channel_id    37422  10961  UCNAf1k0yIjyGu3k9BwAg3lg   533
channel_title 37422  10883        Sky Sports Football   533
published_at  37422  36772        2020-10-16 04:00:10     6
tags          37422  28799                               3817
description   37422  35630                                589
comments      37422  37265                                158
```

- The statistical summary for categorical column includes the count, unique values, topmost frequent values, and the frequency of the top value.
- Columns like video id, title, channel id, channel title, published at, tags, description, comments have these categorical statistics.
- The categorical summary provides the information about the most common values and their frequency, but it doesn't give measures like mean and standard deviation as it does for numerical column.

5. Convert datatype of column published at from object to pandas datetime.

```
# 5. Convert datatype of column published_at from object to pandas datetime.

df.columns
```

```
Index(['video_id', 'title', 'channel_id', 'channel_title', 'published_at',
       'view_count', 'likes', 'dislikes', 'comment_count', 'tags',
       'description', 'comments'],
      dtype='object')
```

```
# Convert published_at columns to datetime

df['published_at'] = pd.to_datetime(df['published_at'])
print(df['published_at'])
```

```
0       2021-07-01 10:00:00
1       2021-06-10 16:00:00
2       2021-09-20 01:03:32
3       2021-03-03 10:00:17
4       2021-12-07 13:00:00
               ...
37417   2021-02-12 05:03:49
37418   2021-01-16 05:39:05
37419   2021-06-02 09:00:10
37420   2020-10-20 20:59:30
37421   2021-05-23 21:00:31
Name: published_at, Length: 37422, dtype: datetime64[ns]
```

```
# Check the data type
print(df.dtypes)
```

```
video_id                 object
title                    object
channel_id               object
channel_title            object
published_at     datetime64[ns]
view_count                int64
likes                     int64
dislikes                  int64
comment_count             int64
tags                     object
description              object
comments                 object
dtype: object
```

6. Create a new column as 'published month' using the column published at (display the months only).

```
# 6. Creat new column published_month using published_at column(display only the months)

df['published_month'] = df['published_at'].dt.month
print(df['published_month'])
```

```
0         7
1         6
2         9
3         3
4        12
         ..
37417     2
37418     1
37419     6
37420    10
37421     5
Name: published_month, Length: 37422, dtype: int32
```

7. Replace the numbers in the column published month as names of the months that is 1 as 'Jan', 2 as 'Feb' and so on.

```
# 7. Replace the numbers in the column published_month as names of the months i,e., 1 as 'Jan', 2 as 'Feb' and so on.

# Replace numerical values with month names

month_mapping = {
    1: 'Jan', 2: 'Feb', 3: 'Mar', 4: 'Apr', 5: 'May', 6: 'Jun',
    7: 'July', 8: 'Aug', 9: 'Sep', 10: 'Oct', 11: 'Nov', 12: 'Dec'
}
```

```
df['published_month'] = df['published_month'].apply(lambda x: month_mapping.get(x))
print(df['published_month'])
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
File C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\indexes\base.py:3653, in Index.get_loc(self, key)
   3652 try:
-> 3653     return self._engine.get_loc(casted_key)
   3654 except KeyError as err:

File C:\ProgramData\anaconda3\Lib\site-packages\pandas\_libs\index.pyx:147, in pandas._libs.index.IndexEngine.get_loc()

File C:\ProgramData\anaconda3\Lib\site-packages\pandas\_libs\index.pyx:176, in pandas._libs.index.IndexEngine.get_loc()

File pandas\_libs\hashtable_class_helper.pxi:7080, in pandas._libs.hashtable.PyObjectHashTable.get_item()

File pandas\_libs\hashtable_class_helper.pxi:7088, in pandas._libs.hashtable.PyObjectHashTable.get_item()

KeyError: 'published_month'

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
Cell In[82], line 1
----> 1 df['published_month'] = df['published_month'].apply(lambda x: month_mapping.get(x))
      2 print(df['published month'])

File C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\frame.py:3761, in DataFrame.__getitem__(self, key)
   3759 if self.columns.nlevels > 1:
   3760     return self._getitem_multilevel(key)
-> 3761 indexer = self.columns.get_loc(key)
   3762 if is_integer(indexer):
   3763     indexer = [indexer]

File C:\ProgramData\anaconda3\Lib\site-packages\pandas\core\indexes\base.py:3655, in Index.get_loc(self, key)
   3653     return self._engine.get_loc(casted_key)
   3654 except KeyError as err:
-> 3655     raise KeyError(key) from err
   3656 except TypeError:
   3657     # If we have a listlike key, _check_indexing_error will raise
   3658     #  InvalidIndexError. Otherwise we fall through and re-raise
   3659     #  the TypeError.
   3660     self._check_indexing_error(key)

KeyError: 'published_month'
```

8. Find the number of videos published each month and arrange the months in a decreasing order based on the video count.

```
# 8. Find the number of videos published each month and arrange the months in a decreasing order based on the video count.

video_count_per_month = df.groupby(['published_month'])['video_id'].count().sort_values(ascending = False)
print(video_count_per_month)
```

```
Series([], Name: video_id, dtype: int64)
```

9. Find the count of unique video id, channel id, and channel title.

```
# 9. Find the count of unique video_id, channel_id and channel_title.

unique_video_id_count = df['video_id'].nunique()
unique_channel_id_count = df['channel_id'].nunique()
unique_channel_title_count = df['channel_title'].nunique()

print("count of unique video_id:", unique_video_id_count)
print("count of unique channel_id:", unique_channel_id_count)
print("count of unique channel_title:", unique_channel_title_count)
```

```
count of unique video_id: 37422
count of unique channel_id: 10961
count of unique channel_title: 10883
```

10.Find the top10 channel names having the highest number of videos in the dataset and the bottom10 having lowest number of videos.

```
# 10.  Find the top10 channel names having the highest number of videos in the dataset and the bottom10 having lowest numbe

# Top 10 channels

channel_video_count = df['channel_title'].value_counts()

top_10_channels = channel_video_count.head(10)
print("Top 10 channels with the highest number of videos")
print(top_10_channels)
```

```
Top 10 channels with the highest number of videos
channel_title
Sky Sports Football    533
The United Stand       301
BT Sport               246
NBA                    209
NFL                    162
WWE                    122
SSSniperWolf            99
SSundee                 98
FORMULA 1               87
NHL                     86
Name: count, dtype: int64
```

```
# Bottom 10 channels

bottom_10_channels = channel_video_count.tail(10)
print("Bottom 10 channels with the highest number of videos")
print(bottom_10_channels)
```

```
Bottom 10 channels with the highest number of videos
channel_title
Eiaz                    1
adidas                  1
Universitetet i Agder   1
Team Tapia              1
TheVincentMottola       1
Blocktrainer            1
Movie Addicts           1
1 1 minute mom          1
FLORIN CERCEL           1
Master Podcast          1
Name: count, dtype: int64
```

11. Find the title of the video which has the maximum number of likes and the title of the video having minimum likes and write your inferences.

```
# 11.  Find the title of the video which has the maximum number of likes and the title of the video having minimum likes an

max_likes_title = df.loc[df['likes'].idxmax(), 'title']
min_likes_title = df.loc[df['likes'].idxmin(), 'title']

print("Title of the video with the maximum number of liks:", max_likes_title)
print("Title of the video with the minimum number of liks:", min_likes_title)
```

```
Title of the video with the maximum number of liks: BTS () 'Dynamite' Official MV
Title of the video with the minimum number of liks: Kim Kardashian's Must-See Moments on "Saturday Night Live" | E! News
```

- The video with maximum number of likes is titled BTS () 'Dynamite' Official MV, and the video with minimum number of likes Kim Kardashian's Must-See Moments on "Saturday Night Live" | E! News
- It’s important to consider other factors like view, comments and context of the videos to have a comprehensive understanding of their popularity or reception.

12. Find the title of the video which has the maximum number of dislikes and the title of the video having minimum dislikes and write your inferences.

```
# 12. Find the title of the video which has the maximum number of dislikes and the title of the video having minimum dislik

max_dislikes_title = df.loc[df['dislikes'].idxmax(), 'title']
min_dislikes_title = df.loc[df['dislikes'].idxmin(), 'title']

print("Title of the video with the maximum number of disliks:", max_dislikes_title)
print("Title of the video with the minimum number of disliks:", min_dislikes_title)
```

```
Title of the video with the maximum number of disliks: Cuties | Official Trailer | Netflix
Title of the video with the minimum number of disliks: Kim Kardashian's Must-See Moments on "Saturday Night Live" | E! New
s
```

- Video with the maximum number of dislikes is titled Cuties | Official Trailer | Netflix, While the video with the minimum number of dislikes is titled Kim Kardashian's Must-See Moments on "Saturday Night Live" | E! News
- As with likes and other metrics, its important to consider the context, content, audience perspective when interacting the dislike counts.

13. Does the number of views have any effect on how many people disliked the video? Support your answer with a metric and a plot.

```
# 13. Does the number of views have any effect on how many people disliked the video? Support youranswer with a metric and

# Calculete the correlation between 'view_count' and th 'dislikes'

correlation = df['view_count'].corr(df['dislikes'])

plt.figure(figsize = (10, 6))
sns.scatterplot(df = df, x = 'veiw_count', y = 'dislikes', alpha = 0.5)
plt.title(f"scatter plot of view count vs Dislikes\ncorrelation:{correlation: .2f}")
plt.xlable('view count')
plt.ylable('Dislikes')
plt.grid(True)
plt.show()

print("correlation between vied_count and dislikes:", correlaton)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Cell In[83], line 7
      4 correlation = df['view_count'].corr(df['dislikes'])
      6 plt.figure(figsize = (10, 6))
----> 7 sns.scatterplot(df = df, x = 'veiw_count', y = 'dislikes', alpha = 0.5)
      8 plt.title(f"scatter plot of view count vs Dislikes\ncorrelation:{correlation: .2f}")
      9 plt.xlable('view count')

File C:\ProgramData\anaconda3\Lib\site-packages\seaborn\relational.py:742, in scatterplot(data, x, y, hue, size, style, pa
lette, hue_order, hue_norm, sizes, size_order, size_norm, markers, style_order, legend, ax, **kwargs)
    732 def scatterplot(
    733     data=None, *,
    734     x=None, y=None, hue=None, size=None, style=None,
   (...)
    738     **kwargs
    739 ):
    741     variables = _ScatterPlotter.get_semantics(locals())
--> 742     p = _ScatterPlotter(data=data, variables=variables, legend=legend)
    744     p.map_hue(palette=palette, order=hue_order, norm=hue_norm)
    745     p.map_size(sizes=sizes, order=size_order, norm=size_norm)

File C:\ProgramData\anaconda3\Lib\site-packages\seaborn\relational.py:538, in _ScatterPlotter.__init__(self, data, variabl
es, legend)
    529 def __init__(self, *, data=None, variables={}, legend=None):
    530
    531     # TODO this is messy, we want the mapping to be agnostic about
    532     # the kind of plot to draw, but for the time being we need to set
    533     # this information so the SizeMapping can use it
    534     self._default_size_range = (
    535         np.r_[.5, 2] * np.square(mpl.rcParams["lines.markersize"])
    536     )
--> 538     super().__init__(data=data, variables=variables)
    540     self.legend = legend

File C:\ProgramData\anaconda3\Lib\site-packages\seaborn\_oldcore.py:640, in VectorPlotter.__init__(self, data, variables)
    635 # var_ordered is relevant only for categorical axis variables, and may
    636 # be better handled by an internal axis information object that tracks
    637 # such information and is set up by the scale_* methods. The analogous
    933 elif isinstance(val, (str, bytes)):
    934
    935     # This looks like a column name but we don't know what it means!
    937     err = f"Could not interpret value `{val}` for parameter `{key}`"
--> 938     raise ValueError(err)
    940 else:
    941
    942     # Otherwise, assume the value is itself data
    943
    944     # Raise when data object is present and a vector can't matched
    945     if isinstance(data, pd.DataFrame) and not isinstance(val, pd.Series):

ValueError: Could not interpret value `veiw_count` for parameter `x`
```

14. Display all the information about the videos that were published in January and mention the count of videos that were published in January.

```
# 14. Display all the information about the videos that were published in January, and mention the count of videos that were

videos = df[df['published_at'].dt.month == 1]

videos
```

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	ta
27	-2Gwm7QtBnE	Q&A With Naisha	UCYwNMbogQFzMocPSuypPWg	MianTwins	2021-01-21 00:05:47	872372	38626	239	621	
48	-4sfKSHSxzA	SURPRISING BRENT WITH HIS TIKTOK CRUSH!!	UCPpATKqmMV-CNRNWYaDUwiA	Alexa Rivera	2021-01-16 21:40:04	6504784	262477	5779	7907	
95	-AJD1Fc5rpQ	WE ARE HAVING A BABY! | finding out i'm pregna...	UCVsTboAhpnuL8j-tDePWwQ	Tess Christine	2021-01-03 21:53:48	533084	38965	119	1650	
103	-AuJiwjsmWk	Do UglyFoods Taste Worse? Taste Test	UCzpCc5n9hqiVC7HhPwcIKEg	Good Mythical MORE	2021-01-19 11:00:01	1057077	22526	531	773	gmm g mythical morn rhettandlinkrl
182	-JhqO2KWr5U	Schlatt gets fit	UCWZp4yfjqBuvLfiykSs_ZBw	Big guy	2021-01-24 22:50:57	1724965	119431	325	1578	jschlatt big jschlatt highlig schlatt
...	...	...	...	...	...	...	...	...	...	
37300	zmzFL5bG-jc	DEVINE MON PERSONNAGE AVANT AKINATOR ! (c'est ...	UCIr3byh8wmXgcPx_TmSOcw	Piwerre	2021-01-16 16:12:19	670357	54462	832	1249	Piwerre frere michou crou among us devi
37329	zpzjex7qwrA	Lampard Sacked Within Days Rorys Misery| Chel...	UCkD-ZOid0a9FjIExDsHsbg	The Kick Off	2021-01-03 20:13:49	428646	12060	296	1505	Premier leag Chelsea chels 1-3 Man CityC
37345	zq yw-B6mnBM	Lil Wayne - Ain't Got Time (Audio)	UCO9zJyi7HWrIS3ojB4Lr7Yqw	Lil Wayne	2021-01-21 05:00:10	2238244	58925	2365	5539	lil wayne we weezywednes wayne carter
37383	zwfu1-24T7Q	PRADA Cup Day1 | Full Race Replay| PRADA Cup...	UCo15ZYO_XDRU9LI30OPbxAg	America's Cup	2021-01-15 04:07:55	317382	2008	83	192	America's C Americas C AC36 AC Presente
37418	zziBybeSAtw	PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...	UCWJ2lWNubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NBA G Leag Basketl game-0022000 Laker

2108 rows × 13 columns

```
print(videos)
```

```
         video_id                                              title  \
27     -2Gwm7QfBnE                                     Q&A With Naisha
48     -4sfXSHSxzA          SURPRISING BRENT WITH HIS TIKTOK CRUSH!!
95     -AJD1Fc5rpQ  WE ARE HAVING A BABY! | finding out i'm pregna...
103    -AuJiwjsmWk              Do Ugly Foods Taste Worse? Taste Test
182    -JhqO2KWr5U                                Schlatt gets fit
...            ...                                               ...
37300  zmzFL5bG-jc  DEVINE MON PERSONNAGE AVANT AKINATOR ! (c'est ...
37329  zprjex7qwrA  Lampard Sacked Within Days Rorys Misery | Chel...
37345  zqyv-B6mnBM                 Lil Wayne - Ain't Got Time (Audio)
37383  zwfu1-24T7Q  PRADA Cup Day 1 | Full Race Replay | PRADA Cup...
37418  zziBybeSAtw  PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...

                    channel_id       channel_title        published_at  \
27     UCYwNMbogQFzMccPSuy-pPWg          MianTwins 2021-01-21 00:05:47
48     UCPpATKqmMV-CNRNWYaDUwiA       Alexa Rivera 2021-01-16 21:40:04
95     UCVsTboAhpnuL6j-tDePvNwQ      Tess Christine 2021-01-03 21:53:48
103    UCzpCc5n9hqiVC7HhPwcIKEg  Good Mythical MORE 2021-01-19 11:00:01
182    UCWZp4y1jqBuvLtiyxSs_ZBw             Big guy 2021-01-24 22:50:57
...                         ...                 ...                 ...
37300  UCIlr3byh6wmXgcPx_Tm9Ocw            Piwerre 2021-01-16 16:12:19
37329  UCkD-IOixI0a9FjIExDsHsbg       The Kick Off 2021-01-03 20:13:49
37345  UCO9zJy7HWrIS3ojB4Lr7Yqw          Lil Wayne 2021-01-21 05:00:10
37383  UCo15IYO_XDRU9LI30OPtxAg       America's Cup 2021-01-15 04:07:55
37418  UCWJ21WNubArHWmf3FIHbfcQ                NBA 2021-01-16 05:39:05

       view_count   likes  dislikes  comment_count  \
27         872372   38626       239            621
48        6504784  262477      5779           7907
95         533084   38965       119           1650
103       1057077   22526       531            773
182       1724965  119431       325           1578
...           ...     ...       ...            ...
37300      670357   54462       832           1249
37329      428646   12060       296           1505
37345     2238244   58925      2365           5539
37383      317382    2008        83            192
37418     2841917   20759      1049           2624

                                                    tags  \
27                                                         
48                                                         
95                                                         
103    gmm good mythical morning rhettandlink rhett a...
182    jschlatt big guy jschlatt highlights schlatt j...
...                                                  ...
37300  Piwerre frere de michou crouton among us devin...
37329  Premier league Chelsea chelsea 1-3 Man City Ch...
37345  lil wayne weezy weezy wednesday wayne carter y...
37383  America's Cup Americas Cup AC36 AC75 Presented...
37418  NBA G League Basketball game-0022000187 Lakers...

                                             description  \
27     Hey Guys!!! this has been the most requested v...
48     He had no idea! Thank you guys so much for wat...
95     I am so happy to tell you that I am pregnant!!...
103    Today, we're doing a blind taste test to deter...
182                  #jschlatt #schlatt #bigguy #short
...                                                  ...
37300   Discord Piwerre : https://discord.gg/QBduPgAA...
37329  The Kick Off watched Manchester City destroy C...
37345  Official audio for Lil Wayne "Ain't Got Time",...
37383  The opening day of the PRADA Cup in Auckland, ...
37418  PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...

                                            comments published_month
27     I feel like Nate and Aishas personality match ...            None
48     Thank you guys for watching and don't forget t...            None
95     Okay I needed a moment to collect my thoughts ...            None
103    "there's nothing wrong with it being bent"\nI ...            None
182    Schlatt is single handedly wiping out all the ...            None


                                            comments published_month
27     I feel like Nate and Aishas personality match ...            None
48     Thank you guys for watching and don't forget t...            None
95     Okay I needed a moment to collect my thoughts ...            None
103    "there's nothing wrong with it being bent"\nI ...            None
182    Schlatt is single handedly wiping out all the ...            None
...                                                  ...             ...
37300  Mdr michou quand c'est pas ses tournage il fou...            None
37329  True Its like a fighter who Geordie But I thou...            None
37345  RIP Juice Wrld, wrote that on two cups, pour o...            None
37383  Incredible how these boats evolve in a short t...            None
37418  Montrezl Harrell is going crazy with the rebou...            None

[2108 rows x 13 columns]
```

```
df['published_at'] = pd.to_datetime(df['published_at'])
count_january_videos = videos.shape[0]
print("count of videos published in january:", count_january_videos)
```

```
count of videos published in january: 2108
```