Exploratory Data Analysis on YouTube data

Submitted by: Fayyas kp

Data Collection

Q1) Import required libraries and read the provided dataset (youtube_dislike_dataset.csv) and retrieve top 5 and bottom 5 records.¶

```python
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity='all'

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Loading data set into Dataframe
df=pd.read_csv(r"D:\Project\Python\youtube_dislike_dataset.csv")
df
```

```
         video_id                                       title  \
0       --0bCF-iK2E              Jadon Sancho  Magical Skills & Goals
1       --14w5SOEUs                Migos - Avalanche (Official Video)
2       --40TEbZ9Is        Supporting Actress in a Comedy: 73rd Emmys
3       --4tfbSyYDE            JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO
4       --DKkzWVh-E                         Why Retaining Walls Collapse
...             ...                                               ...
37417   zzd4ydafGR0  Lil Tjay - Calling My Phone (feat. 6LACK) [Off...
37418   zziBybeSAtw  PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...
37419   zzk09ESX7e0                  [MV]  (MAMAMOO) - Where Are We Now
37420   zzmQEb0Em5I                FELLIPE ESCUDERO- Master Podcast  #12
37421   zzxPZwaA-8w  Gareth Bale brace secures dramatic comeback on...
```

```
                      channel_id           channel_title
published_at  \
0        UC6UL29enLNe4mqwTfAyeNuw                Bundesliga   2021-07-01
10:00:00
1        UCGIelM2Dj3zza3xyV3pL3WQ                MigosVEVO    2021-06-10
16:00:00
2        UClBKH8yZRcM4AsRjDVEdjMg        Television Academy   2021-09-20
01:03:32
3        UCsmXiDP8S40uBeJYxvyulmA                       J01   2021-03-03
10:00:17
4        UCMOqf8ab-42UUQIdVoKwjlQ      Practical Engineering  2021-12-07
13:00:00
...                           ...                       ...
...
37417    UCEB4a5o_6KfjxHwNMnmj54Q                  Lil Tjay   2021-02-12
05:03:49
37418    UCWJ2lWNubArHWmf3FIHbfcQ                       NBA   2021-01-16
05:39:05
37419    UCuhAUMLzJxlP1W7mEk0_6lA                   MAMAMOO   2021-06-02
09:00:10
37420    UC8NjnNWMsRqq11NYvHAQb1g             Master Podcast  2020-10-20
20:59:30
37421    UCEg25rdRZXg32iwai6N6l0w          Tottenham Hotspur  2021-05-23
21:00:31

        view_count      likes   dislikes   comment_count  \
0          1048888      19515        226            1319
1         15352638     359277       7479           18729
2           925281      11212        401             831
3          2641597      39131        441            3745
4           715724      32887        367            1067
...            ...        ...        ...             ...
37417    120408275    2180780      35871           81360
37418      2841917      20759       1049            2624
37419     13346678     720854       4426           90616
37420       252057      19198       1234            1471
37421      2252090      34063        868            2004

                                                      tags  \
0        football soccer ftbol alemn Bundesliga season ...
1        Migos Avalanche Quality Control Music/Motown R...
2
3        PRODUCE101JAPAN             J01     TheSTAR STA...
4        retaining wall New Jersey highway Direct Conne...
...                                                    ...
37417    Lil Tjay Steady Calling My Phone Calling My Ph...
37418    NBA G League Basketball game-0022000187 Lakers...
37419    MAMAMOO  WAW  WAW MAMAMOO WAW Where Are We Now...
37420    master masterpodcast lord lord vinheteiro z z ...
```

```
37421  Spurs Tottenham Hotspur   Tottenham Leicester ...

                                        description  \
0      Enjoy the best skills and goals from Jadon San...
1      Watch the the official video for Migos - "Aval...
2      Hannah Waddingham wins the Emmy for Supporting...
3      JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO\n\n--...
4      One of the most important (and innocuous) part...
...                                                 ...
37417  Official video for "Calling My Phone" by Lil T...
37418  PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...
37419  [MV]  (MAMAMOO) - Where Are We Now\n\nInstagra...
37420  DOCTOR HAIR\nhttps://www.thedoctorhair.com/?fb...
37421  Two minute highlights from Tottenham Hotspur's...

                                           comments
0      Respect to Dortmund fans,must be sad losing hi...
1      Migos just makes me want to live my live to th...
2      Hannah's energy bursts through any screen. Wel...
3      youngVer><  REN is really PERFECT. It's not ju...
4       Keep up with all my projects here: https://pr...
...                                             ...
37417  'DESTINED 2 WIN' OUT NOW !! https://liltjay.ln...
37418  Montrezl Harrell is going crazy with the rebou...
37419  I honestly do not know why this song hit so ha...
37420  Foi um prazer passar esta tarde com vocs debat...
37421  Thank you Kane for everything you have given t...

[37422 rows x 12 columns]
```

#Printing top 5 rows
df.head(5)

```
      video_id                                        title  \
0  --0bCF-iK2E         Jadon Sancho  Magical Skills & Goals
1  --14w5SOEUs          Migos - Avalanche (Official Video)
2  --40TEbZ9Is  Supporting Actress in a Comedy: 73rd Emmys
3  --4tfbSyYDE      JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO
4  --DKkzWVh-E                  Why Retaining Walls Collapse

                 channel_id       channel_title
published_at  \
0  UC6UL29enLNe4mqwTfAyeNuw            Bundesliga  2021-07-01
10:00:00
1  UCGIelM2Dj3zza3xyV3pL3WQ             MigosVEVO  2021-06-10
16:00:00
2  UClBKH8yZRcM4AsRjDVEdjMg   Television Academy  2021-09-20
01:03:32
3  UCsmXiDP8S40uBeJYxvyulmA                   JO1  2021-03-03
10:00:17
```

```
4  UCMOqf8ab-42UUQIdVoKwjlQ  Practical Engineering  2021-12-07
13:00:00

   view_count    likes   dislikes   comment_count  \
0     1048888    19515        226            1319
1    15352638   359277       7479           18729
2      925281    11212        401             831
3     2641597    39131        441            3745
4      715724    32887        367            1067

                                                  tags  \
0  football soccer ftbol alemn Bundesliga season ...
1  Migos Avalanche Quality Control Music/Motown R...
2
3  PRODUCE101JAPAN              JO1    TheSTAR STA...
4  retaining wall New Jersey highway Direct Conne...

                                           description  \
0  Enjoy the best skills and goals from Jadon San...
1  Watch the the official video for Migos - "Aval...
2  Hannah Waddingham wins the Emmy for Supporting...
3  JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO\n\n--...
4  One of the most important (and innocuous) part...

                                              comments
0  Respect to Dortmund fans,must be sad losing hi...
1  Migos just makes me want to live my live to th...
2  Hannah's energy bursts through any screen. Wel...
3  youngVer><  REN is really PERFECT. It's not ju...
4   Keep up with all my projects here: https://pr...
```

```python
# Printiong bottom 5 Rows.
df.tail(5)
```

```
          video_id                                              title
\
37417  zzd4ydafGR0  Lil Tjay - Calling My Phone (feat. 6LACK) [Off...

37418  zziBybeSAtw  PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...

37419  zzk09ESX7e0                      [MV]  (MAMAMOO) - Where Are We Now

37420  zzmQEb0Em5I              FELLIPE ESCUDERO- Master Podcast  #12

37421  zzxPZwaA-8w  Gareth Bale brace secures dramatic comeback on...


                        channel_id      channel_title
published_at  \
37417  UCEB4a5o_6KfjxHwNMnmj54Q             Lil Tjay  2021-02-12
05:03:49
```

```
37418   UCWJ2lWNubArHWmf3FIHbfcQ                   NBA   2021-01-16
05:39:05
37419   UCuhAUMLzJxlP1W7mEk0_6lA              MAMAMOO   2021-06-02
09:00:10
37420   UC8NjnNWMsRqq11NYvHAQb1g      Master Podcast   2020-10-20
20:59:30
37421   UCEg25rdRZXg32iwai6N6l0w  Tottenham Hotspur   2021-05-23
21:00:31

        view_count    likes   dislikes   comment_count  \
37417   120408275   2180780      35871           81360
37418     2841917     20759       1049            2624
37419    13346678    720854       4426           90616
37420      252057     19198       1234            1471
37421     2252090     34063        868            2004

                                                       tags  \
37417   Lil Tjay Steady Calling My Phone Calling My Ph...
37418   NBA G League Basketball game-0022000187 Lakers...
37419   MAMAMOO  WAW   WAW MAMAMOO WAW Where Are We Now...
37420   master masterpodcast lord lord vinheteiro z z ...
37421   Spurs Tottenham Hotspur   Tottenham Leicester ...

                                                description  \
37417   Official video for "Calling My Phone" by Lil T...
37418   PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...
37419   [MV]  (MAMAMOO) - Where Are We Now\n\nInstagra...
37420   DOCTOR HAIR\nhttps://www.thedoctorhair.com/?fb...
37421   Two minute highlights from Tottenham Hotspur's...

                                                   comments
37417   'DESTINED 2 WIN' OUT NOW !! https://liltjay.ln...
37418   Montrezl Harrell is going crazy with the rebou...
37419   I honestly do not know why this song hit so ha...
37420   Foi um prazer passar esta tarde com vocs debat...
37421   Thank you Kane for everything you have given t...
```

Q2) Check the info of the dataframe and write your inferences on data types and shape of the dataset.

```
df.info() #checking the info of the dataframe.
df.dtypes #checking the Data types of the dataset.
df.shape #checking the shape of the dataframe.

<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

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

(37422, 12)
```

Inference:

```
""""The Following dataframe consists of 37422 rows and 12 columns.
Each row represents the data about a video published on youtube. The
dataframe hosts 12 variables and is having Null Values only in the
comments column. The whole dataframe is of Object datatype and on
individual level the dataframe hosts 8 columns of object datatypes and
4 of Int64 datatypes."""

'"The Following dataframe consists of 37422 rows and 12 columns. Each
row represents the data about a video published on youtube. The
dataframe hosts 12 variables and is having Null Values only in the
comments column. The whole dataframe is of Object datatype and on
individual level the dataframe hosts 8 columns of object datatypes and
4 of Int64 datatypes.'
```

Q3) Check for the Percentage of the missing values and drop or impute them.

```python
#Analyzing the percentages of missing values
df.isnull().mean()*100
```

```
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
```

0.42% of the values is missing

```python
#Dropping Comments column.
df.drop(columns=['comments'], inplace=True)

"""Referring to the Analysis only Comments feature is having null
values. Viewing all the Questions metioned in the problem
statement, there is no need for the comments column. With respect to
the observation the comments column is dropped."""
```

```
'Referring to the Analysis only Comments feature is having null
values. Viewing all the Questions metioned in the problem\nstatement,
there is no need for the comments column. With respect to the
observation the comments column is dropped.'
```

Q4) Check the statistical summary of both numerical and categorical columns and write your inferences.

```python
# getting the Statistical Summary of the Numerical columns.
df.describe()
```

	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
```

```python
# getting the Statistical Summary of the categorical columns.
df.describe(include =['object'])
```

```
           video_id  title                   channel_id
channel_title  \
count         37422  37422                        37422
37422
unique        37422  37113                        10961
10883
top      --0bCF-iK2E    www  UCNAf1k0yIjyGu3k9BwAg3lg  Sky Sports
Football
freq              1     21                          533
533


              published_at    tags description
count                37422  37422       37422
unique               36772  28799       35630
top     2020-10-16 04:00:10
freq                     6   3817         589
```

```python
"""The Numerical Columns consists of
view_counts,likes,dislikes,comment_count. The Statistical summary for
the numerical datatypes provides us with the descriptive details about
our dataset. We can get a good idea from viewing the mean of the
numberical columns of the dataframe. and getting the quartiles helps
us to undersatand the data points. Whereas the Categorical column
provides us with a good understanding of the dataset as awhole."""
```

```
'The Numerical Columns consists of
view_counts,likes,dislikes,comment_count. The Statistical summary for
the numerical datatypes provides us with the descriptive details about
our dataset. We can get a good idea from viewing the mean of the
numberical columns of the dataframe. and getting the quartiles helps
us to undersatand the data points. Whereas the Categorical column
provides us with a good understanding of the dataset as awhole.'
```

Q5) Convert datatype of column published_at from object to pandas datetime.

```python
#Conversion of column published_at from object to pandas datetime.
df["published_at"]= pd.to_datetime(df["published_at"])
#df
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 37422 entries, 0 to 37421
Data columns (total 11 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  datetime64[ns]
 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
dtypes: datetime64[ns](1), int64(4), object(6)
memory usage: 3.1+ MB
```

```
"""Here Using to_datetime function given column is converted from
object to datetime type"""
```

```
'Here Using to_datetime function given column is converted from object
to datetime type'
```

6. Create a new column as 'published_month' using the column published_at (display the months only)

```python
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]
```

```python
# creating a column to present the published month(in no.) of the
video
df['published_month']=df['published_at'].dt.month
df.head()
```

```
      video_id                                        title  \
0   --0bCF-iK2E         Jadon Sancho  Magical Skills & Goals
1   --14w5SOEUs            Migos - Avalanche (Official Video)
2   --40TEbZ9Is   Supporting Actress in a Comedy: 73rd Emmys
```

```
3   --4tfbSyYDE       JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO
4   --DKkzWVh-E                      Why Retaining Walls Collapse

                  channel_id          channel_title        published_at
\
0  UC6UL29enLNe4mqwTfAyeNuw             Bundesliga 2021-07-01 10:00:00

1  UCGIelM2Dj3zza3xyV3pL3WQ             MigosVEVO 2021-06-10 16:00:00

2  UClBKH8yZRcM4AsRjDVEdjMg    Television Academy 2021-09-20 01:03:32

3  UCsmXiDP8S40uBeJYxvyulmA                   JO1 2021-03-03 10:00:17

4  UCMOqf8ab-42UUQIdVoKwjlQ  Practical Engineering 2021-12-07 13:00:00


   view_count    likes  dislikes  comment_count  \
0    1048888    19515       226           1319
1   15352638   359277      7479          18729
2     925281    11212       401            831
3    2641597    39131       441           3745
4     715724    32887       367           1067

                                              tags  \
0  football soccer ftbol alemn Bundesliga season ...
1  Migos Avalanche Quality Control Music/Motown R...
2
3  PRODUCE101JAPAN          JO1    TheSTAR STA...
4  retaining wall New Jersey highway Direct Conne...

                                      description  published_month

0  Enjoy the best skills and goals from Jadon San...                7

1  Watch the the official video for Migos - "Aval...                6

2  Hannah Waddingham wins the Emmy for Supporting...                9

3  JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO\n\n---...                3

4  One of the most important (and innocuous) part...               12
```

Q7) Replace the numbers in the column published_month as names of the months i,e., 1 as 'Jan', 2 as 'Feb' and so on.....

```
# Editing the column published_month to present the months for the
videos published
```

```python
df['published_month'] = df['published_at'].dt.strftime('%B')
df.head()
```

```
      video_id                                        title  \
0  --0bCF-iK2E         Jadon Sancho  Magical Skills & Goals
1  --14w5SOEUs           Migos - Avalanche (Official Video)
2  --40TEbZ9Is  Supporting Actress in a Comedy: 73rd Emmys
3  --4tfbSyYDE       JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO
4  --DKkzWVh-E                   Why Retaining Walls Collapse

                channel_id        channel_title         published_at
\
0  UC6UL29enLNe4mqwTfAyeNuw          Bundesliga 2021-07-01 10:00:00

1  UCGIelM2Dj3zza3xyV3pL3WQ           MigosVEVO 2021-06-10 16:00:00

2  UClBKH8yZRcM4AsRjDVEdjMg    Television Academy 2021-09-20 01:03:32

3  UCsmXiDP8S40uBeJYxvyulmA                 JO1 2021-03-03 10:00:17

4  UCMOqf8ab-42UUQIdVoKwjlQ  Practical Engineering 2021-12-07 13:00:00


   view_count    likes  dislikes  comment_count  \
0     1048888    19515       226           1319
1    15352638   359277      7479          18729
2      925281    11212       401            831
3     2641597    39131       441           3745
4      715724    32887       367           1067

                                              tags  \
0  football soccer ftbol alemn Bundesliga season ...
1  Migos Avalanche Quality Control Music/Motown R...
2
3  PRODUCE101JAPAN              JO1    TheSTAR STA...
4  retaining wall New Jersey highway Direct Conne...

                                        description published_month
0  Enjoy the best skills and goals from Jadon San...            July
1  Watch the the official video for Migos - "Aval...            June
2  Hannah Waddingham wins the Emmy for Supporting...       September
3  JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO\n\n---...           March
4  One of the most important (and innocuous) part...        December
```

Q8) Find the number of videos published each month and arrange the months in a decreasing order based on the video count.

```python
#sorting the Count of videos published per month in descending order.
monthly_published = df.groupby('published_month')
['video_id'].count().sort_values(ascending=False)
monthly_published

published_month
October      4991
September    4880
November     4851
August       4262
December     3072
July         2340
June         2316
March        2258
February     2137
April        2126
January      2108
May          2081
Name: video_id, dtype: int64
```

Q9) Find the count of unique video_id, channel_id and channel_titl

```python
#getting count of unique video_id
unique_video_id_Count = df["video_id"].nunique()
#getting count of unique channel_id
unique_channel_id_count =  df["channel_id"].nunique()
# getting count of unique channel_title
unique_channel_title_count = df["channel_title"].nunique()

# Printing the result
print("The Unique video id count is",unique_video_id_Count)

print("The Unique channel id count is",unique_channel_id_count)

print("The Unique channel title count is",unique_channel_title_count)

The Unique video id count is 37422
The Unique channel id count is 10961
The Unique channel title count is 10883
```

Q10) Find the top10 channel names having the highest number of videos in the dataset and the bottom10 having lowest number of videos.

```python
# Using groupby creating a variable containing top 10 Channels having
maximum number of videos.
top_10 = df.groupby(["channel_title"])
["title"].count().sort_values(ascending= False).head(10)

# Using groupby creating a variable containing bottom 10 Channels
having least number of videos.
bottom_10 = df.groupby(["channel_title"])
["title"].count().sort_values(ascending= True).head(10)

print("top_10 Channel Names:")
top_10
print("Bottom_10 Channel Names:")
bottom_10
```

```
top_10 Channel Names:

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: title, dtype: int64


Bottom_10 Channel Names:

channel_title
   SilverName              1
Mini Muka                  1
Mini Ladd                  1
MindYourLogic              1
Mind Body Tonic With Dr Sita    1
Mimi Ar                    1
Millyz                     1
Milkair                    1
Milissa Grande             1
MikuruSong                 1
Name: title, dtype: int64
```

Q11) 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.

```python
# Finding the title of the video having maximim number of likes.
max_likes = df["likes"].max()
max_likes_title = df[df["likes"] == max_likes]['title'].values[0]

print("The title of the video with maximum likes is:")
print(max_likes_title)

The title of the video with maximum likes is:
BTS () 'Dynamite' Official MV

#Finding the title of the video having least number of likes.
min_likes = df["likes"].min()
min_likes_title = df[df["likes"] == min_likes]["title"].values[0]

print("The title of the video with minimum likes is:")
print(min_likes_title)

The title of the video with minimum likes is:
Kim Kardashian's Must-See Moments on "Saturday Night Live" | E! News

"""The most liked video is the BTS () \'Dynamite\' Official MV having
26143 likes and the minimum is of Kim Kardashian\'s Must-See Moments
on "Saturday Night Live" haing 18654 likes. It is also observed that
the videos having higher number of likes are having higher number of
views too."""

'The most liked video is the BTS () \'Dynamite\' Official MV having
26143 likes and the minimum is of Kim Kardashian\'s Must-See Moments
on "Saturday Night Live" haing 18654 likes. It is also observed that
the videos having higher number of likes are having higher number of
views too.'
```

Q12) 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.

```python
# Finding the title of the video having maximim number of dislikes.
max_dislikes = df["dislikes"].max()
max_dislikes_title = df[df["dislikes"] == max_dislikes]
['title'].values[0]

print("The title of the video with maximum dislikes is:")
max_dislikes_title
```

```
The title of the video with maximum dislikes is:

'Cuties | Official Trailer | Netflix'

#Finding the title of the video having least number of dislikes.
min_dislikes = df["dislikes"].min()
min_dislikes_title = df[df["dislikes"] == min_dislikes]
["title"].values[0]

print("The title of the video with minimum dislikes is:")
min_dislikes_title

The title of the video with minimum dislikes is:

'Kim Kardashian\'s Must-See Moments on "Saturday Night Live" | E!
News'

"""The Max dislikes video is: \'Cuties | Official Trailer | Netflix\'.
and the least disliked video is \'Kim Kardashian\'s Must-See Moments
on "Saturday Night Live" | E! News"""

'The Max dislikes video is: \'Cuties | Official Trailer | Netflix\'.
and the least disliked video is \'Kim Kardashian\'s Must-See Moments
on "Saturday Night Live" | E! News'
```

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.

```
# To check the relation we use correlation
corr = df["view_count"].corr(df["dislikes"])
corr

0.6844687753905536

import matplotlib.pyplot as plot

#creating a scatter plot to find relation.
plot.scatter(df["view_count"], df["dislikes"])
plot.xlabel("Views")
plot.ylabel("Dislikes")
plot.title("Correlation between Views and Dislikes")
plot.show()

<matplotlib.collections.PathCollection at 0x29230d79450>

Text(0.5, 0, 'Views')

Text(0, 0.5, 'Dislikes')

Text(0.5, 1.0, 'Correlation between Views and Dislikes')
```
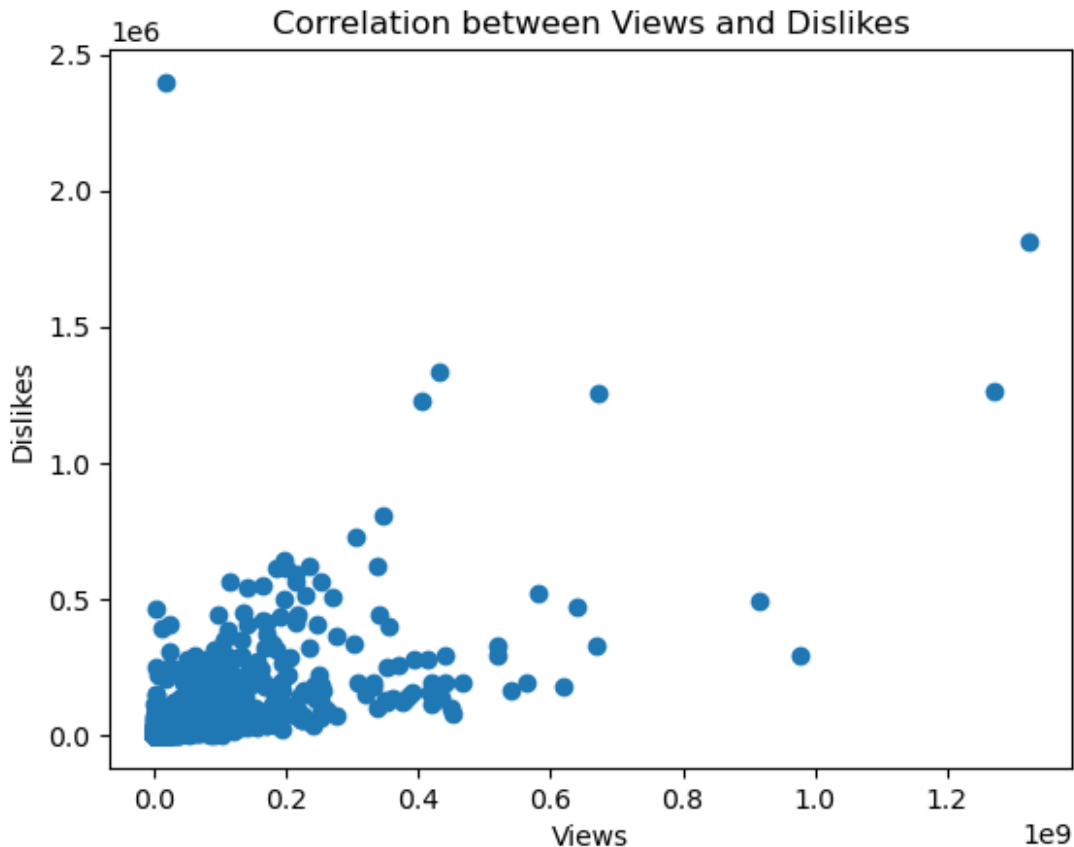
Correlation between Views and Dislikes

```
"""The correlation between views and dislikes is 0.684. It means it's
having a direct relation and increase in one leads to increase in
another and vice-versa."""
```

```
"The correlation between views and dislikes is 0.684. It means it's
having a direct relation and increase in one leads to increase in
another and vice-versa."
```

Q14) Display all the information about the videos that were published in January, and mention the count of videos that were published in January.

```python
# Filter the dataframe to only include videos published in January
jan_videos = df[df['published_month'] == 'January']
print("the Videos published in the month of January are: ")
jan_videos
print("The descriptive summary of the videos published in January
is :")
jan_videos.describe()

print("Number of videos published in January:")
jan_videos.shape[0]
```

the Videos published in the month of January are:

```
          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  zpzjex7qwrA  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-ZOixI0a9FjIExDsHsbg      The Kick Off 2021-01-03
20:13:49
37345  UCO9zJy7HWrIS3ojB4Lr7Yqw         Lil Wayne 2021-01-21
05:00:10
37383  UCo15ZYO_XDRU9LI30OPtxAg    America's Cup 2021-01-15
04:07:55
37418  UCWJ2lWNubArHWmf3FIHbfcQ               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
published_month
27     Hey Guys!!! this has been the most requested v...
January
48     He had no idea! Thank you guys so much for wat...
January
95     I am so happy to tell you that I am pregnant!!...
January
103    Today, we're doing a blind taste test to deter...
January
182                   #jschlatt #schlatt #bigguy #short
January
...                                                  ...    .
..
37300   Discord Piwerre : https://discord.gg/QBduPgAA...
January
37329  The Kick Off watched Manchester City destroy C...
January
37345  Official audio for Lil Wayne "Ain't Got Time",...
January
37383  The opening day of the PRADA Cup in Auckland, ...
January
37418  PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...
```

```
January

[2108 rows x 12 columns]

The descriptive summary of the videos published in January is :

         view_count          likes        dislikes   comment_count
count  2.108000e+03   2.108000e+03     2108.000000    2.108000e+03
mean   4.976449e+06   1.315030e+05     3614.568311    8.923385e+03
std    2.055331e+07   3.886887e+05    13946.711333    3.664662e+04
min    2.845000e+04   2.800000e+01        3.000000    0.000000e+00
25%    4.322335e+05   1.032400e+04      249.000000    8.747500e+02
50%    1.147500e+06   3.173900e+04      729.000000    2.328500e+03
75%    3.227483e+06   1.076512e+05     2243.250000    6.302000e+03
max    6.203274e+08   7.828036e+06   409144.000000    1.095119e+06

Number of videos published in January:

2108

df[df['published_month']=='Jan']['video_id'].count()

0

import pandas as pd
import sweetviz

# Using Sweetviz library creating a dashboard for easier understanding
of the Dataset.
report=sweetviz.analyze(df)
report
```

{"model_id":"3e036cf756c64e4ab4bef4838580cabe","version_major":2,"version_minor":0}

```
<sweetviz.dataframe_report.DataframeReport at 0x23443248810>

# Printing the report
report.show_html("Youtube_report.html")

Report Youtube_report.html was generated! NOTEBOOK/COLAB USERS: the
web browser MAY not pop up, regardless, the report IS saved in your
notebook/colab files.
```