REPORT ON YOUTUBE DISLIKES DATASET

Name : Venkatesh Kulkarni

Batch : GLCA-DA Sept-23

Branch : Bangalore

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Domain: Social media

Executive Summary

In a fairly recent move by Youtube, it announced the decision to hide the number of dislikes from users around November 2021. However, the official YouTube Data API allowed you to get information about dislikes until December 13, 2021. Doing an EDA-exercise can help to draw some unseen insights from this dataset.

Introduction

The purpose of this whole exercise is to explore the dataset. Do the exploratory data analysis. Explore the dataset using central tendency and other parameters. The data consists of 37422 different entries with 8 columns. The analysis should help the learner to explore Data Analysis using Pandas.

Objective

To do data analysis and explore the youtube dislikes dataset using numpy and pandas libraries and drive meaningful insights by performing Exploratory data analysis.

Data Description

YouTube Dislikes Dataset:

- This dataset contains information about trending YouTube videos from August 2020 to December 2021 for the USA, Canada, and Great Britain.

- This dataset contains the latest possible information about dislikes,likes,views and more which was collected just before December 13. The information was collected by videos that had been trending in the USA, Canada, and Great Britain for a year prior.

- Dataset link: https://www.kaggle.com/datasets/dmitrynikolaev/youtube-dislikes-dataset

Sl.No	Coulumn Name	Description
1	Video id	Unique video id
2	Title	Video title
3	Channel id	Id of the channel
4	Channel title	Title of the Channel
5	Published at	Video publication date
6	View count	Number of views
7	Likes	Number of likes
8	Dislikes	Number of dislikes
9	Comment count	Number of counts
10	Tags	Tags(in one string)
11	Description	Video description
12	Comments	20 video comments(in one string)

Table 1 : Data Description

Sample of the Dataset

	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
1	--14w5SOEUs	Migos - Avalanche (Official Video)	UCGIelM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729
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	UCsmXiDP8S40uBeJYxvyulmA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745
4	--DKkzWVh-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQIdVoKwjlQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067

Fig 1 : Sample of the Dataset

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

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

df=pd.read_csv('youtube_dislike_dataset(1).csv')

In [55]:
```
#reading the dataset
df=pd.read_csv('youtube_dislike_dataset.csv')
df
```

Out[55]:

	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	fo
1	--14w5SOEUs	Migos - Avalanche (Official Video)	UCGIelM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	
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	PRO JO1
4	--DKkzWVh-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQIdVoKwjlQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	Jers
...	...	...	...	...	...	...	...	...	...	
37417	zzd4ydafGR0	Lil Tjay - Calling My Phone (feat. 6LACK) [Off...	UCEB4a5o_6KfjxHwNMnmj54Q	Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	

Fig 2 : Importing the dataset(csv file)

Retrieving the top and bottom 5 records

```
In [5]: #reading top 5 rows
        df.head()
```

Out[5]:

	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 s alemn E
1	-14w5SOEUs	Migos - Avalanche (Official Video)	UCGIelM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Qual Music/M
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	UCsmXiDP8S40uBeJYxvyulmA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE JO1 TheST
4	--DKkzWVh-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQIdVoKwjlQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retainin Jersey high

Fig 3 : Printing the top 5 rows

```
In [6]: #reading the bottom 5 rows
        df.tail()
```

Out[6]:

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	tag
37421	zzxPZwaA-8w	Gareth Bale brace secures dramatic comeback on...	UCEg25rdRZXg32iwai6N6l0w	Tottenham Hotspur	2021-05-23 21:00:31	2252090	34063	868	2004	S Totten Hot Totten Leicest
37418	zziBybeSAtw	PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...	UCWJ2lWNubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NE Le Baske ga 002200 Lak
37419	zzk09ESX7e0	[MV] (MAMAMOO) - Where Are We Now	UCuhAUMLzJxlP1W7mEk0_6lA	MAMAMOO	2021-06-02 09:00:10	13346678	720854	4426	90616	MAMAM WAW V MAMAM WAW W Are We N
37420	zzmQEb0Em5I	FELLIPE ESCUDERO- Master Podcast #12	UC8NjnNWMsRqq11NYvHAQb1g	Master Podcast	2020-10-20 20:59:30	252057	19198	1234	1471	m masterpoc lorc vinheteir
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 St Callin Phone Ca My

Fig 4 : Printing the bottom 5 rows

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

```
In [11]: #checking the info of 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
```

```
In [ ]: #there are 4 numeric columns and 8 categorical columns
```

```
In [10]: df.shape
```

```
Out[10]: (37422, 12)
```

```
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
```

fig 5 : Getting the info of the dataset

- From the info, the dataset has a total 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
- 4 columns are integer type and the other 8 are of object datatype.

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

```
In [13]: #checking for null values
         df.isnull().mean()*100

Out[13]: 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

In [ ]: #there are 42% missing values in the comments column
```

Fig 6 : Checking for null values

- Referring to the Analysis only Comments feature is having null values.
- With respect to the observation the comments column is dropped.

In [17]:
```
#dropping the null values
df.dropna()
```

Out[17]:

	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	fo
1	--14w5SOEUs	Migos - Avalanche (Official Video)	UCGIelM2Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	
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	UCsmXiDP8S40uBeJYxvyulmA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRO JO1
4	--DKkzWVh-E	Why Retaining Walls Collapse	UCMOqf8ab-42UUQIdVoKwjlQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	Jers
...	...	...	...	...	...	...	...	...	...	
37417	zzd4ydafGR0	Lil Tjay - Calling My Phone (feat. 6LACK) [Off...	UCEB4a5o_6KfjxHwNMnmj54Q	Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	
37418	zziBybeSAtw	PELICANS at LAKERS | FULL GAME HIGHLIGHTS |	UCWJ2lWNubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	0022

In [17]:
```
#dropping the null values
df.dropna()
```

Out[17]:

	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	tags	description
	LNe4mqwTfAyeNuw	Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football soccer ftbol alemn Bundesliga season ...	Enjoy the best skills and goals from Jadon San...
	Dj3zza3xyV3pL3WQ	MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Avalanche Quality Control Music/Motown R...	Watch the the official video for Migos - "Aval...
	ZRcM4AsRjDVEdjMg	Television Academy	2021-09-20 01:03:32	925281	11212	401	831		Hannah Waddingham wins the Emmy for Supporting...
	8S40uBeJYxvyulmA	JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE101JAPAN JO1 TheSTAR STA...	JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO\n\n---...
	ab-42UUQIdVoKwjlQ	Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retaining wall New Jersey highway Direct Conne...	One of the most important (and innocuous) part...
	...	...	...	...	...	...	...	...	...
	_6KfjxHwNMnmj54Q	Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	Lil Tjay Steady Calling My Phone Calling My Ph...	Official video for "Calling My Phone" by Lil T...
	lubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NBA G League Basketball game-0022000187 Lakers...	PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...

Fig 7 : Dropping the null values

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

```
In [18]: #statistical summary
         #numeric summary
         df.describe()
```

Out[18]:

	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

Fig 8 : Statistical Summary

```
In [19]: #categorical summary
         df.describe(include=['object'])
```

Out[19]:

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

Fig 9 : Categorical Summary

- The Numerical Columns consists of view_counts,likes,dislikes,comment_count.
- The Statistical summary for the numerical datatypes provides us with the descriptive deta ils about our dataset.
- We can get a good idea from viewing the mean of the numberical columns of the datafra me.
- 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

```
In [23]: #Converting datatype of column published_at from object to pandas datetime.
         df["published_at"]= pd.to_datetime(df["published_at"])
         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
```

Fig 10 : Converting datatype of published_at to pandas datetime

- Conversion of column published at from object to pandas datetime. Loading the info after conversion
- Observing the data types we have another data type added to the dataset that is ‘datetime’.

Q6. Create a new column as 'published_month' using the column published_at (display the months only).

```
In [26]: #Creating a new column as 'published_month' using the column published_at
         df['published_month']=df['published_at'].dt.month
         df
```

Out[26]:

:hannel_title	published_at	view_count	likes	dislikes	comment_count	tags	description	published_month
Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football soccer ftbol alemn Bundesliga season ...	Enjoy the best skills and goals from Jadon San...	7
MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Avalanche Quality Control Music/Motown R...	Watch the the official video for Migos - "Aval...	6
Television Academy	2021-09-20 01:03:32	925281	11212	401	831		Hannah Waddingham wins the Emmy for Supporting...	9
JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE101JAPAN JO1 TheSTAR STA...	JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO\n\n---...	3
Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retaining wall New Jersey highway Direct Conne...	One of the most important (and innocuous) part...	12
...	...	...	...	...	...	...	...	...
Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	Lil Tjay Steady Calling My Phone Calling My Ph...	Official video for "Calling My Phone" by Lil T...	2
NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NBA G League Basketball game-0022000187 Lakers...	PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...	1

Fig 11 : Creating a new column

- Creating a new column ‘published month’ to the given dataset and displaying it, as shown in Fig11.

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.....

```
In [28]: df['published_month'] = df['published_at'].dt.strftime('%B')
         df
```

Out[28]:

channel_title	published_at	view_count	likes	dislikes	comment_count	tags	description	published_month
Bundesliga	2021-07-01 10:00:00	1048888	19515	226	1319	football soccer ftbol alemn Bundesliga season ...	Enjoy the best skills and goals from Jadon San...	July
MigosVEVO	2021-06-10 16:00:00	15352638	359277	7479	18729	Migos Avalanche Quality Control Music/Motown R...	Watch the the official video for Migos - "Aval...	June
Television Academy	2021-09-20 01:03:32	925281	11212	401	831		Hannah Waddingham wins the Emmy for Supporting...	September
JO1	2021-03-03 10:00:17	2641597	39131	441	3745	PRODUCE101JAPAN JO1 TheSTAR STA...	JO1'YOUNG (JO1 ver.)' PERFORMANCE VIDEO\n\n---...	March
Practical Engineering	2021-12-07 13:00:00	715724	32887	367	1067	retaining wall New Jersey highway Direct Conne...	One of the most important (and innocuous) part...	December
...	...	...	...	...	...	...	...	...
Lil Tjay	2021-02-12 05:03:49	120408275	2180780	35871	81360	Lil Tjay Steady Calling My Phone Calling My Ph...	Official video for "Calling My Phone" by Lil T...	February
NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NBA G League Basketball game-0022000187 Lakers...	PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...	January

Fig 12 : Replacing the month numbers with month names

- Editing the column to present the months for the videos published and Loading Dataframe after editing Column. As shown in fig12.

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

```
In [29]: df.groupby('published_month')['video_id'].count().sort_values(ascending=False)

Out[29]: 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
```

Fig 13 : Printing the number of videos publishes each month

- sorting the Count of videos published per month in descending order.
- Highest videos published month is October
- Lowest videos published month is May

Q9. Find the count of unique video_id, channel_id and channel_title.

```
In [30]: #count of unique video_id
         df["video_id"].nunique()

Out[30]: 37422

In [32]: #count of unique channel_id
         df['channel_id'].nunique()

Out[32]: 10961

In [33]: #count of unique channel_title
         df['channel_title'].nunique()

Out[33]: 10883
```

Fig 14 : Count of video_id, channel_id, channel_title

- The Unique video_id_ count is 37422
- The Unique channel_id count is 10961
- The Unique channel_title count is 10883

Q10. Find the top 10 channel names having the highest number of videos in the dataset and the bottom10 having lowest number of videos.

```
In [34]: # channels having highest number of videos
         df.groupby(["channel_title"])["title"].count().sort_values(ascending= False).head(10)

Out[34]: 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

In [35]: # channels having lowest number of videos
         df.groupby(["channel_title"])["title"].count().sort_values(ascending= True).head(10)

Out[35]: 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
```

Fig 15 : Finding the channel names having highest number and lowest number of videos

- Using groupby creating a variable containing top and bottom 10 Channels having maximum number of videos.

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.

```
In [38]: #maximum likes
         df[df["likes"] == df["likes"].max()]['title'].values[0]

Out[38]: "BTS () 'Dynamite' Official MV"

In [40]: #minimum likes
         df[df["likes"] == df['likes'].min()]["title"].values[0]

Out[40]: 'Kim Kardashian\'s Must-See Moments on "Saturday Night Live" | E! News'
```

Fig 16 : Finding the title of the video having maximum and minimum number of likes

- The most liked video is the BTS () 'Dynamite' Official MV having 26143 likes
- the minimum is of Kim Kardashian's Must-See Moments on "Saturday Night Live" having 18654 likes.

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.

```
In [42]: #Maximum dislikes
         df[df["dislikes"] == df["dislikes"].max()]["title"].values[0]

Out[42]: 'Cuties | Official Trailer | Netflix'

In [43]: #minimum dislikes
         df[df['dislikes']==df['dislikes'].min()]['title'].values[0]

Out[43]: 'Kim Kardashian\'s Must-See Moments on "Saturday Night Live" | E! News'
```

Fig 17 : Finding the title of the video having maximum and minimum number of dislikes

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

Q13. Does the number of views have any effect on how many people disliked the video? Support your answer with a metric and a plot.

```
In [31]: #finding the correlation between number of views and dislikes
         corr = df["view_count"].corr(df["dislikes"])
         corr

Out[31]: 0.6844687753905547

In [32]: import matplotlib.pyplot as plot
         plot.scatter(df["view_count"], df["dislikes"])
         plot.xlabel("Views")
         plot.ylabel("Dislikes")
         plot.title("Correlation between Views and Dislikes")
         plot.show()
```

Fig 18 : Finding the correlation between view_count and dislikes

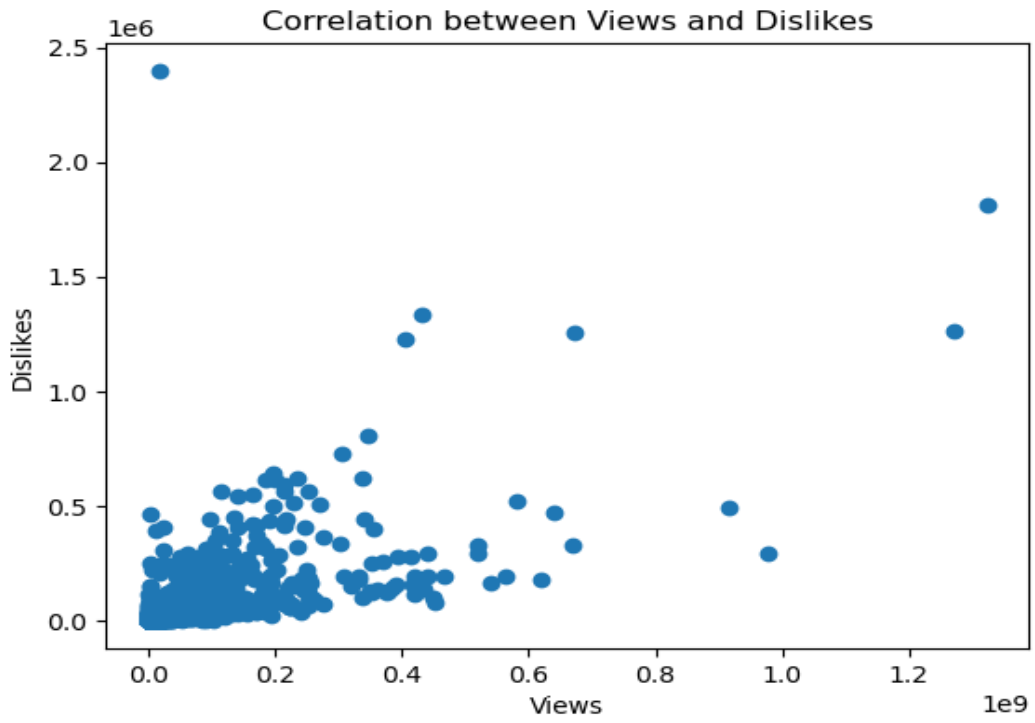

Fig 19 : Correlation plot between view_count and dislikes

- The number of views and dislikes are positively correlated which means the increase in one variable results in increase in another variable.
- The correlation between views and dislikes is around 0.684.

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

```
In [49]: jan=df[df['published_month'] == 'January']
         jan
```

Out[49]:

	video_id	title	channel_id	channel_title	published_at	view_count	likes	dislikes	comment_count	tag
27	-2Gwm7QfBnE	Q&A With Naisha	UCYwNMbogQFzMccPSuy-pPWg	MianTwins	2021-01-21 00:05:47	872372	38626	239	621	
48	-4sfXSHSxzA	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...	UCVsTboAhpnuL6j-tDePvNwQ	Tess Christine	2021-01-03 21:53:48	533084	38965	119	1650	
103	-AuJiwjsmWk	Do Ugly Foods Taste Worse? Taste Test	UCzpCc5n9hqiVC7HhPwcIKEg	Good Mythical MORE	2021-01-19 11:00:01	1057077	22526	531	773	gmm g myth mor rhettanc rhett
37329	zpzjex7qwrA	Lampard Sacked Within Days Rorys Misery | Chel...	UCkD-ZOixI0a9FjIExDsHsbg	The Kick Off	2021-01-03 20:13:49	428646	12060	296	1505	Pre lea Che chelsea Man
37345	zqyv-B6mnBM	Lil Wayne - Ain't Got Time (Audio)	UCO9zJy7HWrIS3ojB4Lr7Yqw	Lil Wayne	2021-01-21 05:00:10	2238244	58925	2365	5539	lil wa we we wednes wa carte
37383	zwfu1-24T7Q	PRADA Cup Day 1 | Full Race Replay | PRADA Cup...	UCo15ZYO_XDRU9LI30OPtxAg	America's Cup	2021-01-15 04:07:55	317382	2008	83	192	Ameri Amer Cup A A Present
37418	zziBybeSAtw	PELICANS at LAKERS | FULL GAME HIGHLIGHTS | Ja...	UCWJ2lWNubArHWmf3FIHbfcQ	NBA	2021-01-16 05:39:05	2841917	20759	1049	2624	NB Lea Baske ga 0022000 Lake

2108 rows × 12 columns

```
In [50]: jan.shape
Out[50]: (2108, 12)
```

Fig 20 : Information of all the videos that were publishes in January

- the Videos published in the month of January are shown in above Fig.20.
- Number of videos pblished in January.

Conclusion

The above analysis concludes that there are 37422 rows and 12 columns in the dataset and there are 4 integer and 8 object datatype columns. The variables ‘number of views’ and ‘dislikes’ are positively correlated.