





[illegible]



[illegible]



[illegible]



[illegible]



[illegible]



[illegible]



[illegible]



[illegible]

```
In [14]: #let's save the collected data in csv format using this -
data_df.to_csv('zeemusiccompany.csv',index=False)

# By running the above code, latest dataset will be downloaded and saved in 'zeemusic'

In [15]: # Importing operating system library to views files and interacting with system
import os
os.listdir() # Shows all the files available in current directory

Out[15]: ['.ipynb_checkpoints', 'YOUTUBE Data Analysis.ipynb', 'zeemusiccompany.csv']
```

## Data Preparation and cleaning

We have our raw dataset. Now, we'll remove the unwanted data, and will make the dates readable and will extract the information from it (date, time, day, month/year) and will store them in separate columns.

```
In [16]: # Storing information from csv file to Pandas dataframe
zeemusic_raw_df = pd.read_csv('zeemusiccompany.csv')

In [17]: zeemusic_raw_df

Out[17]:
```

	video_id	channel_id	published_date	video_title	video_description	likes	dislikes	
0	a6CJAFFqnJ	UCFbwnve3yfe2-tVXi7yHqg	2020-07-29T06:30:19Z	Jaani Ban Gaye - Khuda Haafiz [Viduyat Jammwal ...	👤 SUBSCRIBE to Zee Music Company - https://bi...	1432584	30897	10896

1	7gUAGZ30F9o	UCFFbwbnve3yf62-tVXkTyHqg	2018-09-12T10:32:14Z	Raftaar x Nawazuddin Siddiqui - MANTOVIAT   Manto	👉 SUBSCRIBE to Zee Music Company - <a href="https://bi...">https://bi...</a>	1383962	27498	5502
2	7gUAGZ30F9o	UCFFbwbnve3yf62-tVXkTyHqg	2018-09-12T10:32:14Z	Raftaar x Nawazuddin Siddiqui - MANTOVIAT   Manto	👉 SUBSCRIBE to Zee Music Company - <a href="https://bi...">https://bi...</a>	1383962	27498	5502
3	votbP_3o73qj	UCFFbwbnve3yf62-tVXkTyHqg	2020-06-25T07:31:08Z	Jaan Nisaar - Lyrical   Kedamath   Arjit Singh...	👉 SUBSCRIBE to Zee Music Company - <a href="https://bi...">https://bi...</a>	1183101	33166	11325
4	BT0z0d0kmTxM	UCFFbwbnve3yf62-tVXkTyHqg	2021-01-23T12:00:13Z	Jersey - Official Trailer   Shahid Kapoor   Mr...	👉 SUBSCRIBE to Zee Music Company - <a href="https://bi...">https://bi...</a>	862768	39444	4735
...	...	...	...	...	...	...	...	...
95	INURbRt7ESo	UCFFbwbnve3yf62-tVXkTyHqg	2019-03-06T14:00:58Z	Neend Na Aave Mainu   Bandit Vajje   Jatinder Singh	👉 SUBSCRIBE to Zee Music Company - <a href="https://bi...">https://bi...</a>	68855	3575	1111

	96	-Lw8kdhYGVK	UCFFbwvnr3y3f62-tVxkTyHqg	2019-09-21T06:30:12Z	Dil Hi Tuh Hai - The Sky Is Pink   Priyanka Ch...	SUBSCRIBE to Zee Music Company - https://bi...	112804	4192	1346
	97	Ew_cZGIGzGMt	UCFFbwvnr3y3f62-tVxkTyHqg	2015-07-14T13:05:15Z	Das Dae Full Video   Ishqedariyaan   Mahaaksh...	SUBSCRIBE to Zee Music Company - https://bi...	26136	473	134
	98	okv_erK7QHUx	UCFFbwvnr3y3f62-tVxkTyHqg	2020-09-21T14:37:20Z	Mann Ki Dori - Full Video  Gurjan Saxena   Jan...	SUBSCRIBE to Zee Music Company - https://bi...	54818	1695	313
	99	6Ksq7k-E2Gc	UCFFbwvnr3y3f62-tVxkTyHqg	2020-08-03T06:30:27Z	Khuda Haafiz Title Track - Vidyut Jammwal Shiv...	SUBSCRIBE to Zee Music Company - https://bi...	247316	4984	1215
	100 rows × 9 columns								
In [18]:	<pre># Removing unwanted columns - channel_id and video_id zeemusic_df=zeemusic_raw_df.drop(['channel_id','video_id'], inplace=False,axis=1)</pre>								
In [19]:	<pre># Our new dataframe with required information zeemusic_df</pre>								
Out[19]:	published_date	video_title	video_description	likes	dislikes	views	comment_count		
	0	2020-07-29T06:30:19Z	Jaan Bar Gaye - Khuda Haafiz  Vidyt Jammwal...	SUBSCRIBE to Zee Music Company - https://bi...	142584	30897	108963761	31024	

1	2019-09-12T10:32:14Z	Raftaar x Nawazuddin Siddiqui   MANTOYAT   Manto	🔄 SUBSCRIBE to Zee Music Company - <a href="https://bit.ly/3839662">https://bit.ly/3839662</a>	1383962	27498	55026141	70978
2	2018-10-12T10:32:14Z	Raftaar x Nawazuddin Siddiqui   MANTOYAT   Manto	🔄 SUBSCRIBE to Zee Music Company - <a href="https://bit.ly/3839662">https://bit.ly/3839662</a>	1383962	27498	55026141	70978
3	2020-06-25T07:31:08Z	Jaan Nisaar - Lyrical   Kedarnath Arijit Sing...	🔄 SUBSCRIBE to Zee Music Company - <a href="https://bit.ly/3839662">https://bit.ly/3839662</a>	1183101	33166	113258341	33414
4	2021-11-23T11:00:13Z	Jersey - Official Trailer   Shahid Kapoor   Mr...	🔄 SUBSCRIBE to Zee Music Company - <a href="https://bit.ly/3839662">https://bit.ly/3839662</a>	862768	39444	47304061	53233
...			...	...	...	...	...
95	2019-03-06T14:00:58Z	Neend Na Aave Mannu   Band Waaje   Jatinendra Sh...	🔄 SUBSCRIBE to Zee Music Company - <a href="https://bit.ly/3839662">https://bit.ly/3839662</a>	68855	3575	11115516	1348
96	2019-09-	Dil Hi Toh Hai - The Story of a Love	🔄 SUBSCRIBE to Zee Music Company - <a href="https://bit.ly/3839662">https://bit.ly/3839662</a>	112804	4192	13454898	3084

	2170630.12Z	Priyanka Ch...		https://bi...					
97	2015-07-13T05:15Z	Das Dee Full Video lshqedarriyaa Mahaaksh...	👤	SUBSCRIBE to Zee Music Company - https://bi...	26136	473	1348880	760	
98	2020-09-21T14:37:20Z	Mann Ki Dori - Full Video  Gurjan Saxena   Jan...	👤	SUBSCRIBE to Zee Music Company - https://bi...	54818	1695	3137722	705	
99	2020-08-03T06:30:27Z	Khuda Haafiz Title Track - Vidvut Jammwal Shiv...	👤	SUBSCRIBE to Zee Music Company - https://bi...	247316	4984	12190929	6842	

100 rows x 7 columns

### Converting published date and time into more readable format

```
In [20]: # Importing datetime library which provides great functions to handle date and time in python
import datetime
i=0
for i in range(zee music raw df.shape[0]):
    date_time_obj = datetime.datetime.strptime(zee music df['published_date'].at[i], '%Y-%m-%d %H:%M:%S')
    zee music df['published_date'].at[i] = date_time_obj
    i = i+1
```

In [21]:		zeemusic_df						
Out[21]:	published_date	video_title	video_description	likes	dislikes	views	comment_count	
0	2020-07-29 06:30:19	Jaan Ban Gaye - Khushi Haazil [vidyut Jammwal ...	👉 SUBSCRIBE to Zee Music Company - https://bi...	1432584	30897	108963761	31024	
1	2018-09-12 10:32:14	Raftaar x Nawazuddin Siddiqui - MANTO[YAT Manto	👉 SUBSCRIBE to Zee Music Company - https://bi...	1383962	27498	55026141	70978	
2	2018-09-12 10:32:14	Raftaar x Nawazuddin Siddiqui - MANTO[YAT Manto	👉 SUBSCRIBE to Zee Music Company - https://bi...	1383962	27498	55026141	70978	
3	2020-06-25 07:31:08	Jaan Nisaar - Lyrical [Kedarnathi] Anjli Sing.	👉 SUBSCRIBE to Zee Music Company - https://bi...	1183101	33166	113258341	33414	
4	2021-11-23	Jersey - Official Trailer   Iqbal Sing.	👉 SUBSCRIBE to Zee Music Company - https://bi...	862768	39444	47304061	53223	

	12:00:13	Kapoor   Mr.	https://bi...	68855	3575	1115516	1348
...	...	...	...	...	...	...	...
95	2019-03-06 14:00:58	Neend Na Aave Mam   Band Naje   Jatinder Sh...	🔗 SUBSCRIBE to Zee Music Company - https://bi...	68855	3575	1115516	1348
96	2019-09-21 06:30:12	Dil Hi Toh Hai - The Sky is Pink   Priyanka Ch...	🔗 SUBSCRIBE to Zee Music Company - https://bi...	112804	4192	13484898	3084
97	2015-07-14 13:05:15	Das Dae Full Video   Ishqedarriyaan   Mahaaksh...	🔗 SUBSCRIBE to Zee Music Company - https://bi...	26136	473	1348880	760
98	2020-09-21 14:37:20	Mann Ki Dori - Full Video   Gunjan Saxena   Jan...	🔗 SUBSCRIBE to Zee Music Company - https://bi...	54818	1695	3137722	705
99	2020-08-03 06:30:27	Khuda Haafiz Title Track - Vidut Jammwal   Shiv...	🔗 SUBSCRIBE to Zee Music Company - https://bi...	247316	4984	12190929	6842

100 rows × 7 columns

```
In [22]: i=0
         date=[]
         time=[]
         year=[]
         month=[]
         day=[]
         for i in range(zeemusic_df.shape[0]):
             d = zeemusic_df['published_date'][i].date();
             t = zeemusic_df['published_date'][i].time();
             y = zeemusic_df['published_date'][i].date().year;
             m = zeemusic_df['published_date'][i].date().month;
             da = zeemusic_df['published_date'][i].date().day;
             date.append(d) # Storing dates
             time.append(t) # Storing time
             year.append(y) # Storing years
             month.append(m) # Storing months
             day.append(da) # Storing days
             i = i+1
         zeemusic_df.drop(['published_date'], inplace=True,axis=1)
         zeemusic_df['published_date']=date
         zeemusic_df['year']=year
         zeemusic_df['month']=month
         zeemusic_df['day']= day
```

Cleaned data set								
<pre>In [23]: # this is our cleaned dataset, we'll use this for data analysis zeemusic_df</pre>								
Out[23]:	video_title	video_description	likes	dislikes	views	comment_count	published_date	published
0	Jaan Ban Gaye - Khuda Haafiz   Idyut Jammwal ...	👉 SUBSCRIBE to Zee Music Company - https://bi...	1432584	30897	108963761	31024	2020-07-29	06
1	Raftaar x Nawazuddin Siddiqui   MANTOIYAT   Manto	👉 SUBSCRIBE to Zee Music Company - https://bi...	1383962	27498	55026141	70978	2018-09-12	10
2	Raftaar x Nawazuddin Siddiqui   MANTOIYAT   Manto	👉 SUBSCRIBE to Zee Music Company - https://bi...	1383962	27498	55026141	70978	2018-09-12	10
	Jaan Nisaar -	👉 SUBSCRIBE to						

3	 <b>Lyrical</b> [Kedamath] Anjith Sing...	 <b>Zee Music</b> Company - https://bi...	1183101	33166	113258341	33414	2020-06-25	07
4	 <b>Jersey - Official</b> Trailer   Shahid Kapoor   Mr...	 <b>SUBSCRIBE to</b> Zee Music Company - https://bi...	862768	39444	47304061	53233	2021-11-23	12
...	...	...	...	...	...	...	...	...
95	 <b>Neend Na</b> Aave Maine   Band Vajje   Jatinder Sh...	 <b>SUBSCRIBE to</b> Zee Music Company - https://bi...	68855	3575	11115516	1348	2019-03-06	14
96	 <b>DI HI Toh Hai -</b> The Sky Is Pink   Priyanka Ch...	 <b>SUBSCRIBE to</b> Zee Music Company - https://bi...	112804	4192	13484898	3084	2019-09-21	06
97	 <b>Das Dae Full</b> Video   Ishgeddaryyaan   Mahaaksh...	 <b>SUBSCRIBE to</b> Zee Music Company - https://bi...	26136	473	1348880	760	2015-07-14	13
98	 <b>Mann Ki Dori -</b> Full Video   Gurleen Sengra	 <b>SUBSCRIBE to</b> Zee Music Company - https://bi...	54818	1695	3137722	705	2020-09-21	14

	Jan	https://bi...							
	Khuda Haafiz Title Track - Vidyt Jammwal Shiv...	➔ SUBSCRIBE to Zee Music Company - https://bi...	247316	4984	12190929	6842	2020-08-03	06	

100 rows × 11 columns

### Size of the dataset

```
In [24]: print('Number of rows = ',zeemusic_df.shape[0],'\nNumber of columns = ',zeemusic_df.shape[1])
Number of rows = 100
Number of columns = 11
Size of the dataset = 1100 elements.
```

## EXPLORATORY DATA ANALYSIS AND VISUALIZATION

Let us view various statistics and relationships in the extracted dataset.

```
In [25]: zeemusic_df.describe()
```

```

count    1.000000e+02    100.000000    1.000000e+02    100.000000    100.000000    100.000000
mean     4.475560e+05    17076.220000    4.284732e+07    13747.140000    2018.560000    7.300000    16.440000
std      6.617886e+05    33084.952528    6.970944e+07    21228.353097    1.960829    2.886751    8.135168
min      1.820000e+04    319.000000    7.617560e+05    353.000000    2014.000000    1.000000    1.000000
25%      6.642300e+04    1927.000000    6.185257e+06    1113.000000    2017.000000    6.000000    11.000000
50%      1.769070e+05    4725.000000    1.873568e+07    4416.500000    2021.000000    7.000000    15.000000
75%      5.867200e+05    14781.000000    4.515084e+07    14107.000000    2020.000000    9.000000    24.000000
max      3.580386e+06    181157.000000    3.301177e+08    86083.000000    2021.000000    12.000000    30.000000

```

Total likes, dislikes, views and comments of all videos

```

In [26]: zeemuslc_df[['views', 'likes', 'dislikes', 'comment_count']].sum()

Out[26]: views          4284731922
         likes          44755605
         dislikes      17076222
         comment_count    1374714
         dtype: int64

```

```

Average number of views, likes, dislikes and comments on videos

In [27]: AvgLikes = zeeusmic_df.describe()['likes']['mean']
AvgDislikes = zeeusmic_df.describe()['dislikes']['mean']
AvgViews = zeeusmic_df.describe()['views']['mean']
AvgComments = zeeusmic_df.describe()['comment_count']['mean']
print('Average number of views on video = ',AvgViews,'\nAverage number of likes on video = ',AvgLikes)

```



