



```
In [2]: import pandas as pd
data=pd.read_csv(r"C:\Users\NITYA PRIYA\Downloads\USvideos.csv")
print(data.head())
print(data.info())
```

```

video_id trending_date \
0 2kyS6SvSYSE 17.14.11
1 1ZAPwfrtAFY 17.14.11
2 5qpjK5DgCt4 17.14.11
3 puqaWrEC7tY 17.14.11
4 d380meD0W0M 17.14.11

```

```

title channel_title
\
0 WE WANT TO TALK ABOUT OUR MARRIAGE CaseyNeistat
1 The Trump Presidency: Last Week Tonight with J... LastWeekTonight
2 Racist Superman | Rudy Mancuso, King Bach & Le... Rudy Mancuso
3 Nickelback Lyrics: Real or Fake? Good Mythical Morning
4 I Dare You: GOING BALD!? nigahiga

```

```

category_id publish_time \
0 22 2017-11-13T17:13:01.000Z
1 24 2017-11-13T07:30:00.000Z
2 23 2017-11-12T19:05:24.000Z
3 24 2017-11-13T11:00:04.000Z
4 24 2017-11-12T18:01:41.000Z

```

```

tags views likes \
0 SHANTell martin 748374 57527
1 last week tonight trump presidency|"last week ... 2418783 97185
2 racist superman|"rudy"|"mancuso"|"king"|"bach"... 3191434 146033
3 rhett and link|"gmm"|"good mythical morning"|"... 343168 10172
4 ryan|"higa"|"higatv"|"nigahiga"|"i dare you"|"... 2095731 132235

```

```

dislikes comment_count thumbnail_link \
0 2966 15954 https://i.ytimg.com/vi/2kyS6SvSYSE/default.jpg (h
https://i.ytimg.com/vi/2kyS6SvSYSE/default.jpg)
1 6146 12703 https://i.ytimg.com/vi/1ZAPwfrtAFY/default.jpg (h
https://i.ytimg.com/vi/1ZAPwfrtAFY/default.jpg)
2 5339 8181 https://i.ytimg.com/vi/5qpjK5DgCt4/default.jpg (h
https://i.ytimg.com/vi/5qpjK5DgCt4/default.jpg)
3 666 2146 https://i.ytimg.com/vi/puqaWrEC7tY/default.jpg (h
https://i.ytimg.com/vi/puqaWrEC7tY/default.jpg)
4 1989 17518 https://i.ytimg.com/vi/d380meD0W0M/default.jpg (h
https://i.ytimg.com/vi/d380meD0W0M/default.jpg)

```

```

comments_disabled ratings_disabled video_error_or_removed \
0 False False False
1 False False False
2 False False False
3 False False False
4 False False False

```

```

description
0 SHANTELL'S CHANNEL - https://www.youtube.com/s... (https://www.youtube.co
m/s...)
1 One year after the presidential election, John...
2 WATCH MY PREVIOUS VIDEO ► \n\nSUBSCRIBE ► http...
3 Today we find out if Link is a Nickelback amat...
4 I know it's been a while since we did this sho...

```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 40949 entries, 0 to 40948
```

Data columns (total 16 columns):

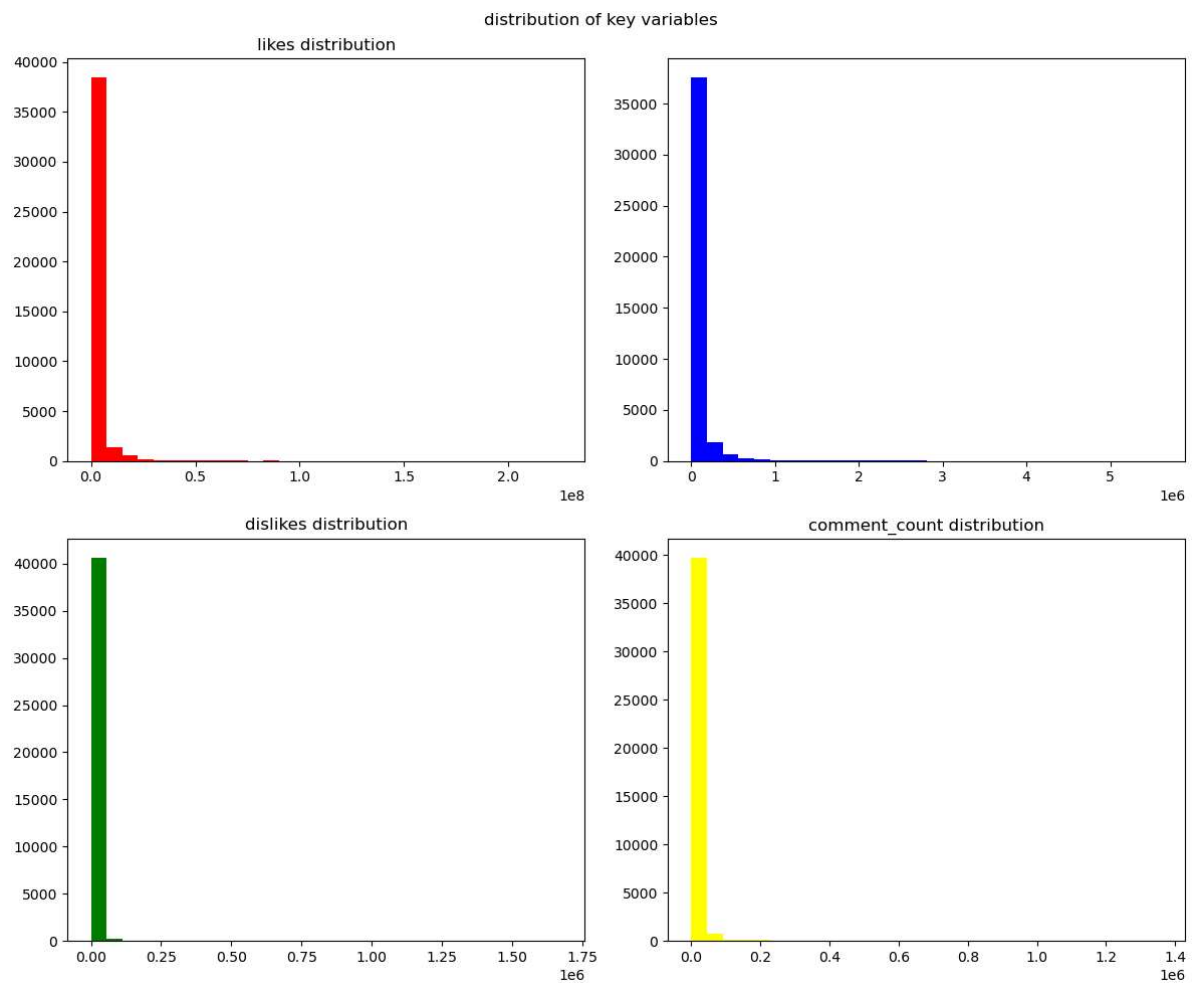
#	Column	Non-Null Count	Dtype
0	video_id	40949 non-null	object
1	trending_date	40949 non-null	object
2	title	40949 non-null	object
3	channel_title	40949 non-null	object
4	category_id	40949 non-null	int64
5	publish_time	40949 non-null	object
6	tags	40949 non-null	object
7	views	40949 non-null	int64
8	likes	40949 non-null	int64
9	dislikes	40949 non-null	int64
10	comment_count	40949 non-null	int64
11	thumbnail_link	40949 non-null	object
12	comments_disabled	40949 non-null	bool
13	ratings_disabled	40949 non-null	bool
14	video_error_or_removed	40949 non-null	bool
15	description	40379 non-null	object

dtypes: bool(3), int64(5), object(8)

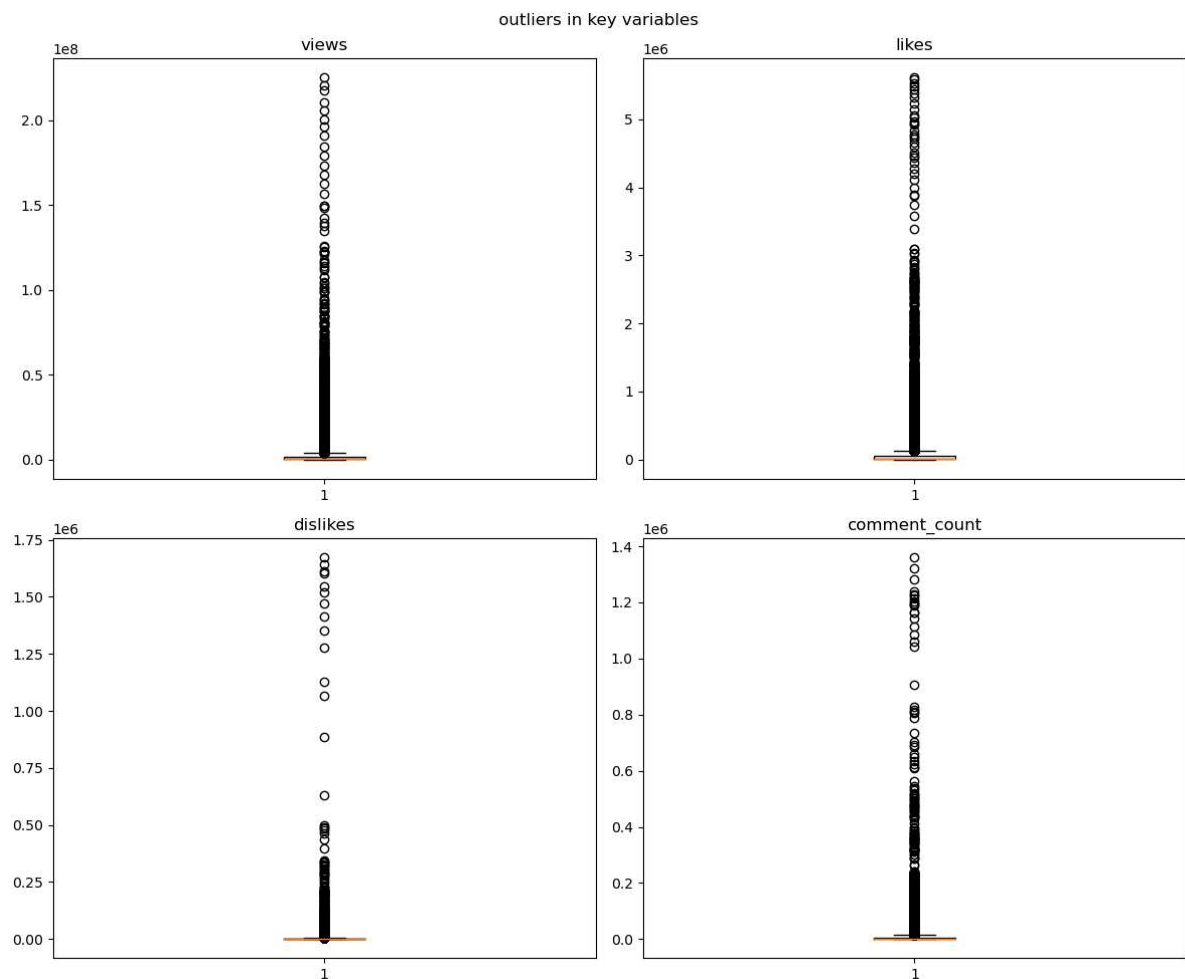
memory usage: 4.2+ MB

None

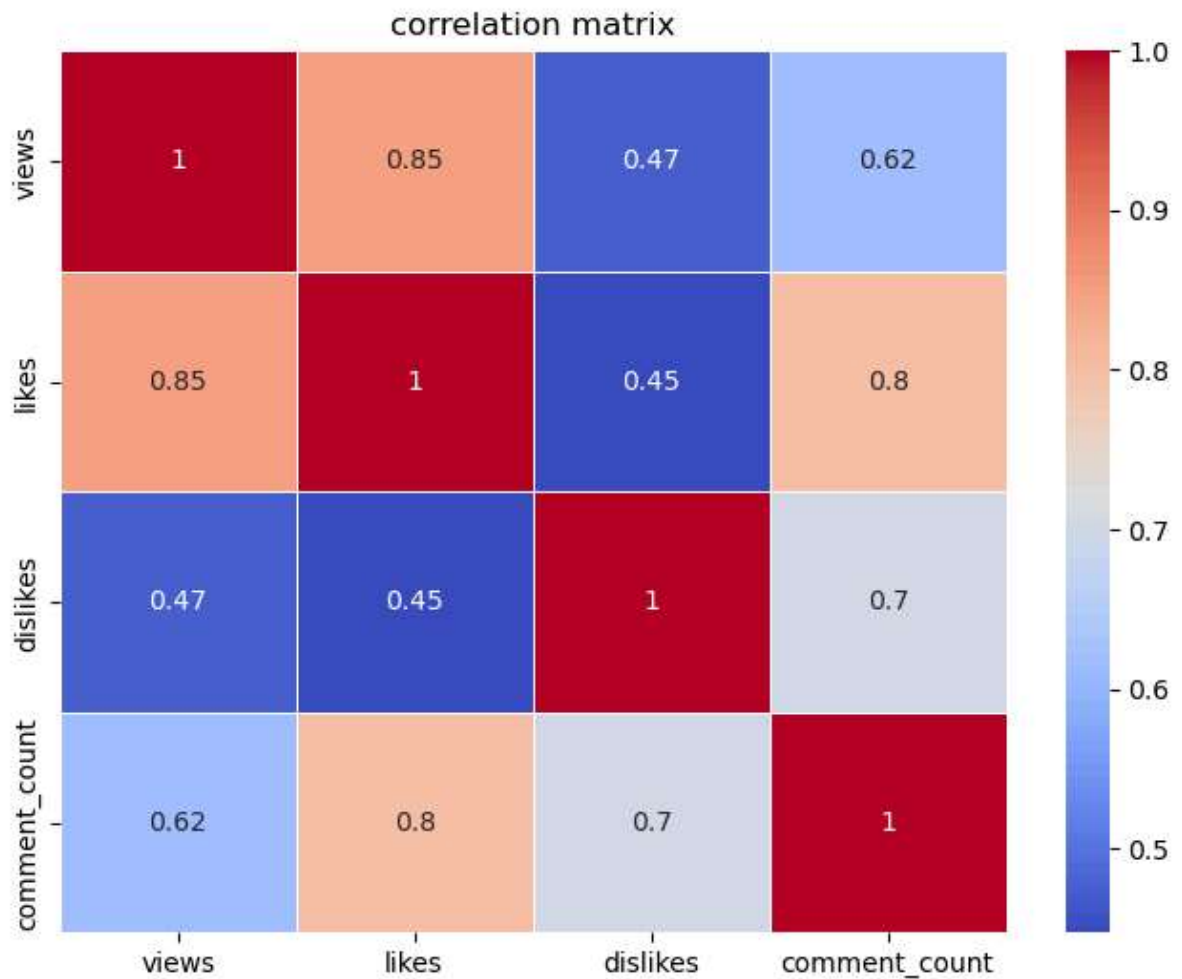
```
In [7]: import matplotlib.pyplot as plt
#plot histogram for distribution
fig,axes=plt.subplots(2,2,figsize=(12,10))
fig.suptitle('distribution of key variables')
axes[0,0].hist(data['views'],bins=30,color='red')
axes[0,0].set_title('views distribution')
axes[0,1].hist(data['likes'],bins=30,color='blue')
axes[0,1].set_title('likes distribution')
axes[1,0].hist(data['dislikes'],bins=30,color='green')
axes[1,0].set_title('dislikes distribution')
axes[1,1].hist(data['comment_count'],bins=30,color='yellow')
axes[1,1].set_title('comment_count distribution')
plt.tight_layout()
plt.show()
```



```
In [11]: #plot boxplots to identify outliers
fig, axes = plt.subplots(2, 2, figsize=(12, 10))
fig.suptitle('outliers in key variables')
axes[0, 0].boxplot(data['views'].dropna())
axes[0, 0].set_title('views')
axes[0, 1].boxplot(data['likes'].dropna())
axes[0, 1].set_title('likes')
axes[1, 0].boxplot(data['dislikes'].dropna())
axes[1, 0].set_title('dislikes')
axes[1, 1].boxplot(data['comment_count'].dropna())
axes[1, 1].set_title('comment_count')
plt.tight_layout()
plt.show()
```



```
In [13]: import seaborn as sns
#select relevant columns and calculate correlation
correlation_matrix=data[['views','likes','dislikes','comment_count']].corr()
#plot the heatmap
plt.figure(figsize=(8,6))
sns.heatmap(correlation_matrix,annot=True,cmap='coolwarm',linewidth=0.5)
plt.title('correlation matrix')
plt.show()
```



In [ ]:

In [ ]:

In [ ]:

In [ ]:

