

```
import pandas as pd

df = pd.read_csv('/train.csvm')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3198 entries, 0 to 3197
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Video_id                             3198 non-null   object
1   category_id                           3192 non-null   object
2   channel_title                         3195 non-null   object
3   subscriber                           3175 non-null   float64
4   title                                3195 non-null   object
5   tags                                  3046 non-null   object
6   description                           3133 non-null   object
7   Trend_day_count                       3197 non-null   float64
8   Tag_count                             3197 non-null   object
9   Trend_tag_count                       3197 non-null   object
10  comment_count                         3198 non-null   object
11  comment_disabled                      3198 non-null   object
12  like dislike disabled                 3198 non-null   object
13  likes                                3198 non-null   object
14  dislike                              3198 non-null   object
15  tag appered in title                 3197 non-null   object
16  views                                3198 non-null   object
17  Unnamed: 17                          1 non-null     float64
18  Unnamed: 18                          1 non-null     object
dtypes: float64(3), object(16)
memory usage: 474.8+ KB
```

```
df.head()
```

ation	Trend_day_count	Tag_count	Trend_tag_count	comment_count	comment_disabled	like dislike disabled	likes	dis
NEL - m/s...	6.0	21	6		FALSE	FALSE	13342	
er the ohn...	1.0	23	1	116266	TRUE	FALSE	5761	
IOUS RIBE http...	10.0	22	3	257850	TRUE	TRUE	0	
ik is a mat...	12.0	17	5	263939	TRUE	TRUE	0	
while sho...	11.0	15	7	268085	TRUE	TRUE	0	

```
df.head(10)
```

on	Trend_day_count	Tag_count	Trend_tag_count	comment_count	comment_disabled	like dislike disabled	likes	disli
L - s...	6.0	21	6		FALSE	FALSE	13342	60
the r...	1.0	23	1	116266	TRUE	FALSE	5761	30
JS BE p...	10.0	22	3	257850	TRUE	TRUE	0	
s a it...	12.0	17	5	263939	TRUE	TRUE	0	
ile c...	11.0	15	7	268085	TRUE	TRUE	0	
on't e...	10.0	9	1	198248	TRUE	FALSE	8619	36
ate (...	4.0	17	5		FALSE	FALSE	9951	9
ion l...	9.0	14	7	103280	TRUE	FALSE	18010	110
ion r...	8.0	20	1	214583	TRUE	FALSE	12126	78
ive c...	11.0	8	2	223434	TRUE	TRUE	0	

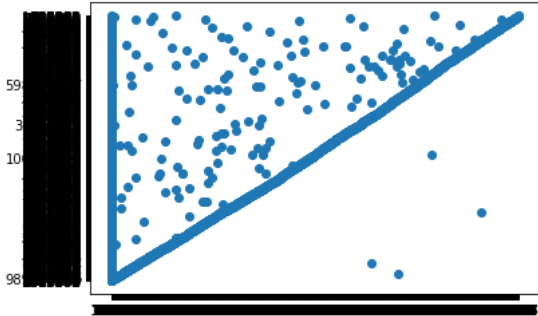
```
df.describe()
```

	subscriber	Trend_day_count	Unnamed: 17
count	3.175000e+03	3197.000000	1.0
mean	3.823981e+06	7.964342	2544.0
std	2.865771e+07	78.556055	NaN
min	0.000000e+00	0.000000	2544.0
25%	2.428800e+05	4.000000	2544.0
50%	1.241220e+06	7.000000	2544.0
75%	3.812622e+06	10.000000	2544.0
max	1.576229e+09	4444.000000	2544.0

```
import matplotlib.pyplot as plt
plt.hist(df['views'])
plt.show()
```



```
plt.scatter(df['likes'], df['views'])
plt.show()
```



```
import seaborn as sns
sns.heatmap(df.corr(), annot=True)
plt.show()
```

