```
import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

import seaborn as sns


df = pd.read_csv(r"path to csv\train.csv")

df.shape

df.dtypes

df.isnull().sum()*100 / df.shape[0]

df.drop(df.columns[[-1, -2]], axis=1, inplace=True)


df['category_id'].fillna(0, inplace=True)

df['channel_title'].fillna('Anonymous', inplace=True)

df['subscriber'].fillna(0, inplace=True)

df['title'].fillna('No title', inplace=True)

df['tags'].fillna('No tags', inplace=True)

df['description'].fillna('No description', inplace=True)

df['Trend_day_count'].fillna(0, inplace=True)

df['Tag_count'].fillna(0, inplace=True)

df['Trend_tag_count'].fillna(0, inplace=True)

df['tag appered in title'].fillna('No title tag', inplace=True)


df['Tag_count'] = df['Tag_count'].replace(regex=[r'^[A-Za-z)(<>+-/@_.&#\s]+', '', '#VALUE!', True, False], value=0)  #Repeat this step for columns 'Trend_tag_count', 'comment_count', 'likes', 'dislike', 'views'

df[Tag_count] = df[Tag_count].astype(float)   #Repeat this step for columns 'Trend_tag_count', 'comment_count', 'likes', 'dislike', 'views'


plt.hist(x= df['like dislike disabled'])

plt.title("Distribution of Likes, Dislikes Disabled")

plt.show()
```

```
plt.hist(x= df['comment_disabled'])
plt.title("Distribution of Comment Disabled")
plt.show()

fig = plt.figure(figsize=(8,8))
ax = fig.gca()
youtube_df.hist(ax=ax)
plt.show()

plt.boxplot([df['likes'], df['dislike'], df['comment_count']])
plt.title("Distribution of Likes, Dislikes and Comments")
plt.show()

plt.boxplot([df['Trend_day_count'], df['Tag_count'], df['Trend_tag_count']])
plt.title("Distribution of Trend_day_count, Tag_count and Trend_tag_count")
plt.show()

plt.boxplot([df['subscriber'], df['views']])
plt.title("Distribution of Subscribers, Views")
plt.show()

df.corr()

sns.pairplot(data= df)
plt.show()

sns.heatmap(data= df.corr(), annot=True, linewidth=0.5, cmap="Blues", cbar=True, vmin=-1, vmax=1)
plt.show()
```