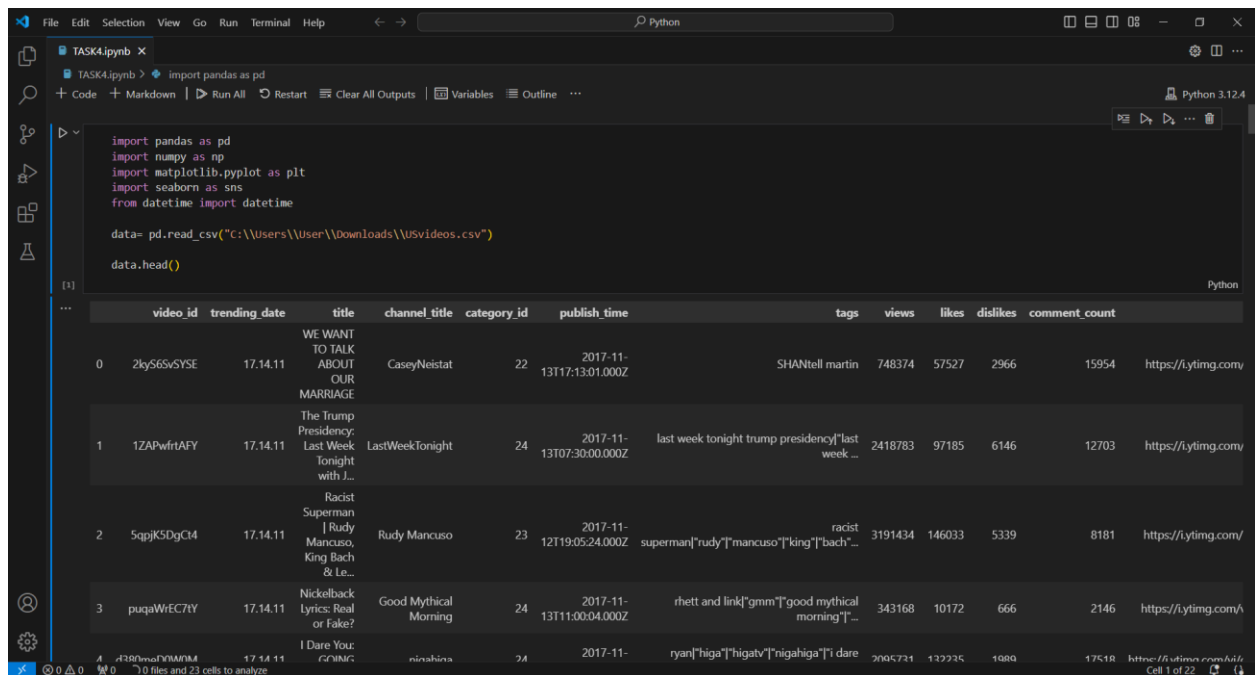


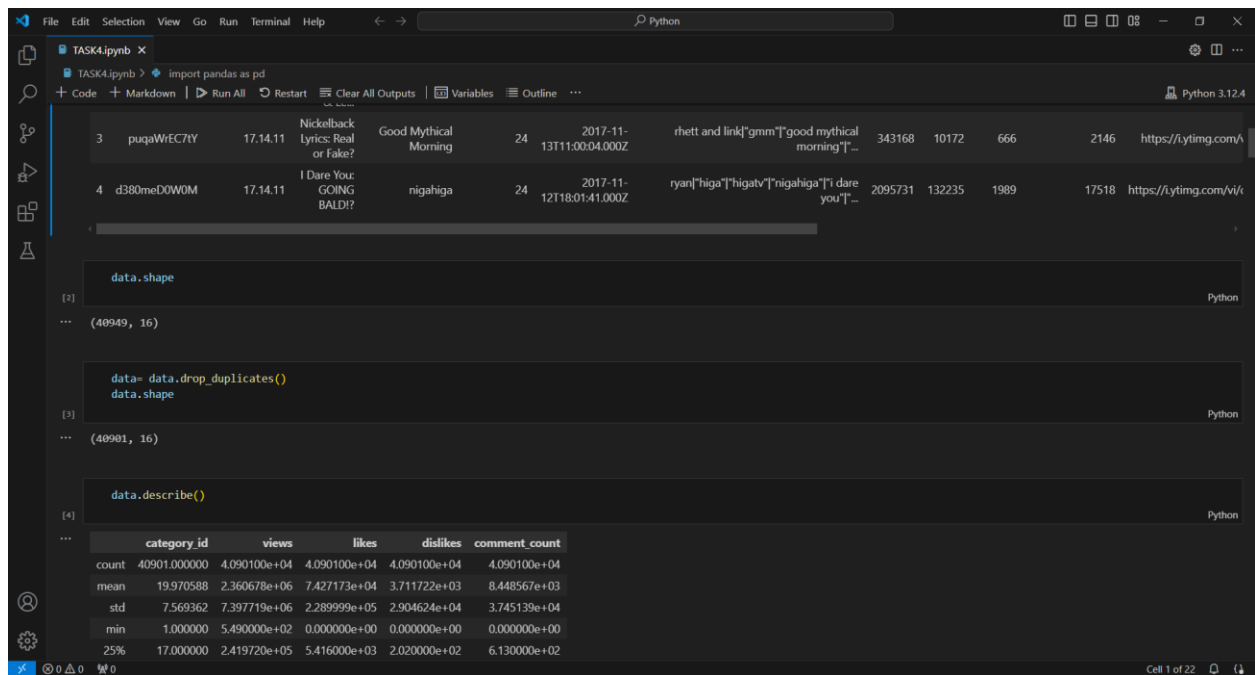
# TASK 4



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime

data= pd.read_csv("c:\\Users\\User\\Downloads\\USVideos.csv")
data.head()
```

	video_id	trending_date	title	channel_title	category_id	publish_time	tags	views	likes	dislikes	comment_count	
0	2kyS6svSYSE	17.14.11	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	2017-11-13T17:13:01.000Z	SHANTell martin	748374	57527	2966	15954	<a href="https://i.ytimg.com/v/2kyS6svSYSE">https://i.ytimg.com/v/2kyS6svSYSE</a>
1	1ZAPwrtAFY	17.14.11	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	2017-11-13T07:30:00.000Z	last week tonight trump presidency("last week ...	2418783	97185	6146	12703	<a href="https://i.ytimg.com/v/1ZAPwrtAFY">https://i.ytimg.com/v/1ZAPwrtAFY</a>
2	5qpjKSDgCt4	17.14.11	Racist Superman   Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	2017-11-12T19:05:24.000Z	superman "rudy" mancuso "king" bach*...	3191434	146033	5339	8181	<a href="https://i.ytimg.com/v/5qpjKSDgCt4">https://i.ytimg.com/v/5qpjKSDgCt4</a>
3	puqaWfEC7tY	17.14.11	Nickelback Lyrics: Real or Fake?	Good Mythical Morning	24	2017-11-13T11:00:04.000Z	rhett and link "gmm" good mythical morning" ...	343168	10172	666	2146	<a href="https://i.ytimg.com/v/puqaWfEC7tY">https://i.ytimg.com/v/puqaWfEC7tY</a>
4	d380meDOW0M	17.14.11	I Dare You: GOING BALD!?	nigahiga	24	2017-11-12T18:01:41.000Z	ryan "higa" higatv "nigahiga" i dare you" ...	2095731	132235	1989	17518	<a href="https://i.ytimg.com/v/d380meDOW0M">https://i.ytimg.com/v/d380meDOW0M</a>



```
data.shape
data.drop_duplicates()
data.shape
data.describe()
```

	category_id	views	likes	dislikes	comment_count
count	40901.000000	4.090100e+04	4.090100e+04	4.090100e+04	4.090100e+04
mean	19.970588	2.360678e+06	7.427173e+04	3.711722e+03	8.448567e+03
std	7.569362	7.397719e+06	2.289999e+05	2.904624e+04	3.745139e+04
min	1.000000	5.490000e+02	0.000000e+00	0.000000e+00	0.000000e+00
25%	17.000000	2.419720e+05	5.416000e+03	2.020000e+02	6.130000e+02

```
File Edit Selection View Go Run Terminal Help Python
TASK4.ipynb X
TASK4.ipynb > import pandas as pd

+ Code + Markdown Run All Restart Clear All Outputs Variables Outline Python 3.12.4

25% 17.000000 2.419720e+05 5.416000e+03 2.020000e+02 6.130000e+02
50% 24.000000 6.810640e+05 1.806900e+04 6.300000e+02 1.855000e+03
75% 25.000000 1.821926e+06 5.533800e+04 1.936000e+03 5.752000e+03
max 43.000000 2.252119e+08 5.613827e+06 1.674420e+06 1.361580e+06

data.info()

[5] Python

<class 'pandas.core.frame.DataFrame'>
Index: 48981 entries, 0 to 48948
Data columns (total 16 columns):
#   column                Non-Null Count  Dtype
---  -
0   video_id              48981 non-null  object
1   trending_date         48981 non-null  object
2   title                 48981 non-null  object
3   channel_title         48981 non-null  object
4   category_id           48981 non-null  int64
5   publish_time          48981 non-null  object
6   tags                  48981 non-null  object
7   views                 48981 non-null  int64
8   likes                 48981 non-null  int64
9   dislikes              48981 non-null  int64
10  comment_count         48981 non-null  int64
11  thumbnail_link        48981 non-null  object
12  comments_disabled     48981 non-null  bool
13  ratings_disabled      48981 non-null  bool
14  video_error_or_removed 48981 non-null  bool
15  description            48332 non-null  object
dtypes: bool(3), int64(5), object(8)
memory usage: 4.5+ MB
```

```
File Edit Selection View Go Run Terminal Help Python
TASK4.ipynb X
TASK4.ipynb > import pandas as pd

+ Code + Markdown Run All Restart Clear All Outputs Variables Outline Python 3.12.4

columns_to_remove=['thumbnail_link', 'description']
data=data.drop(columns=columns_to_remove)
data.info()

[6] Python

<class 'pandas.core.frame.DataFrame'>
Index: 48981 entries, 0 to 48948
Data columns (total 14 columns):
#   column                Non-Null Count  Dtype
---  -
0   video_id              48981 non-null  object
1   trending_date         48981 non-null  object
2   title                 48981 non-null  object
3   channel_title         48981 non-null  object
4   category_id           48981 non-null  int64
5   publish_time          48981 non-null  object
6   tags                  48981 non-null  object
7   views                 48981 non-null  int64
8   likes                 48981 non-null  int64
9   dislikes              48981 non-null  int64
10  comment_count         48981 non-null  int64
11  comments_disabled     48981 non-null  bool
12  ratings_disabled      48981 non-null  bool
13  video_error_or_removed 48981 non-null  bool
dtypes: bool(3), int64(5), object(6)
memory usage: 3.9+ MB

from datetime import datetime

[7] Python

import datetime

[8] Python
```

```
File Edit Selection View Go Run Terminal Help Python
TASK4.ipynb X
TASK4.ipynb > import pandas as pd

+ Code + Markdown Run All Restart Clear All Outputs Variables Outline Python 3.12.4

import datetime

[9] Python

data['trending_date'] = data['trending_date'].apply(lambda x: datetime.datetime.strptime(x, '%y.%d.%m'))
data.head(3)

[10] Python

video_id  trending_date  title  channel_title  category_id  publish_time  tags  views  likes  dislikes  comment_count  comments_disabled  ra
0  2kyS6svSYSE  2017-11-14  WE WANT TO TALK ABOUT OUR MARRIAGE  CaseyNeistat  22  2017-11-13T17:13:01.000Z  SHANTell martin  748374  57527  2966  15954  False
1  1ZAPwfrtAFY  2017-11-14  The Trump Presidency: Last Week Tonight with J...  LastWeekTonight  24  2017-11-13T07:30:00.000Z  last week tonight trump presidency|"last week ...  2418783  97185  6146  12703  False
2  5apjKSDgCt4  2017-11-14  Racist Superman | Rudy Mancuso, King Bach & L...  Rudy Mancuso  23  2017-11-12T19:05:24.000Z  superman|"rudy|"mancuso|"king|"bach"...  3191434  146033  5339  8181  False

data['publish_time']=pd.to_datetime(data['publish_time'])
data.head(2)

[11] Python
```

Python

TASK4.ipynb X

TASK4.ipynb > data[publish\_month]=data[publish\_time].dt.month

+ Code + Markdown Run All Restart Clear All Outputs Variables Outline ... Python 3.12.4

```
data['publish_time']=pd.to_datetime(data['publish_time'])
data.head(2)
```

video_id	trending_date	title	channel_title	category_id	publish_time	tags	views	likes	dislikes	comment_count	comments_disabled	ratings_disabled	video_error_r
0	2kyS6svSYSE	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	2017-11-13 17:13:01+00:00	SHANTell martin	748374	57527	2966	15954	False	False	
1	1ZAPwfrtAFY	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	2017-11-13 07:30:00+00:00	last week tonight trump presidency last week ...	2418783	97185	6146	12703	False	False	

```
data['publish_month']=data['publish_time'].dt.month
data['publish_day']=data['publish_time'].dt.day
data['publish_hour']=data['publish_time'].dt.hour
data.head(2)
```

video_id	trending_date	title	channel_title	category_id	publish_time	tags	views	likes	dislikes	comment_count	comments_disabled	ratings_disabled	video_error_r
0	2kyS6svSYSE	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	2017-11-13 17:13:01+00:00	SHANTell martin	748374	57527	2966	15954	False	False	
1	1ZAPwfrtAFY	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	2017-11-13 07:30:00+00:00	last week tonight trump presidency last week ...	2418783	97185	6146	12703	False	False	

Spaces: 4 Cell 11 of 22

Python

TASK4.ipynb X

TASK4.ipynb > data[publish\_month]=data[publish\_time].dt.month

+ Code + Markdown Run All Restart Clear All Outputs Variables Outline ... Python 3.12.4

video_id	trending_date	title	channel_title	category_id	publish_time	tags	views	likes	dislikes	comment_count	comments_disabled	ratings_disabled	video_error_r
1	1ZAPwfrtAFY	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	2017-11-13 07:30:00+00:00	last week tonight trump presidency last week ...	2418783	97185	6146	12703	False	False	

```
print(sorted(data['category_id'].unique()))
[1, 2, 10, 15, 17, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 43]
```

```
[np.int64(1), np.int64(2), np.int64(10), np.int64(15), np.int64(17), np.int64(19), np.int64(20), np.int64(22), np.int64(23), np.int64(24), np.int64(25), np.int64(26), np.int64(27), np.int64(28), np.int64(29), np.int64(30), np.int64(43)]
```

```
data['category_name']=np.nan
data.loc[(data['category_id']==1), 'category_name'] = "Film and Animation"
data.loc[(data['category_id']==2), 'category_name'] = "Autos and Vehicles"
data.loc[(data['category_id']==10), 'category_name'] = "Music"
data.loc[(data['category_id']==15), 'category_name'] = "Pets and Animals"
data.loc[(data['category_id']==17), 'category_name'] = "Sports"
data.loc[(data['category_id']==19), 'category_name'] = "Travel and Events"
data.loc[(data['category_id']==20), 'category_name'] = "Gaming"
data.loc[(data['category_id']==22), 'category_name'] = "People and Blogs"
data.loc[(data['category_id']==23), 'category_name'] = "Comedy"
data.loc[(data['category_id']==24), 'category_name'] = "Entertainment"
data.loc[(data['category_id']==25), 'category_name'] = "News and Politics"
data.loc[(data['category_id']==26), 'category_name'] = "How to and Style"
data.loc[(data['category_id']==27), 'category_name'] = "Education"
data.loc[(data['category_id']==28), 'category_name'] = "Science and Technology"
data.loc[(data['category_id']==29), 'category_name'] = "Non Profits and Activism"
data.loc[(data['category_id']==30), 'category_name'] = "Movies"
```

Spaces: 4 Cell 11 of 22

Python

TASK4.ipynb X

TASK4.ipynb > data[publish\_month]=data[publish\_time].dt.month

+ Code + Markdown Run All Restart Clear All Outputs Variables Outline ... Python 3.12.4

```
data.loc[(data['category_id']==30), 'category_name'] = "Movies"
data.loc[(data['category_id']==43), 'category_name'] = "Shows"

data.head()
```

```
C:\Users\User\AppData\Local\Temp\ipykernel_8600\2331478760.py:2: FutureWarning: Setting an item of incompatible dtype is deprecated and will raise an error in a future version of pandas
data.loc[(data['category_id']==1), 'category_name'] = "Film and Animation"
```

video_id	trending_date	title	channel_title	category_id	publish_time	tags	views	likes	dislikes	comment_count	comments_disabled	ratings_disabled	video_error_r
0	2kyS6svSYSE	WE WANT TO TALK ABOUT OUR MARRIAGE	CaseyNeistat	22	2017-11-13 17:13:01+00:00	SHANTell martin	748374	57527	2966	15954	False	False	
1	1ZAPwfrtAFY	The Trump Presidency: Last Week Tonight with J...	LastWeekTonight	24	2017-11-13 07:30:00+00:00	last week tonight trump presidency last week ...	2418783	97185	6146	12703	False	False	
2	5qpiKSDgCt4	Racist Superman   Rudy Mancuso, King Bach & Le...	Rudy Mancuso	23	2017-11-12 19:05:24+00:00	superman rudy mancuso king bach ...	3191434	146033	5339	8181	False	False	
3	puqaWfEC7iY	Nickelback Lyrics: Real or Fake?	Good Mythical Morning	24	2017-11-13 11:00:04+00:00	rhett and link gmm good mythical morning ...	343168	10172	666	2146	False	False	
4	d380meDOW0M	I Dare You: GOING BALDI?	nigahiga	24	2017-11-12 18:01:41+00:00	ryan higa higatv nigahiga i dare you ...	2095731	132235	1989	17518	False	False	

Spaces: 4 Cell 11 of 22

