Home Page - Select or create a n | Untitled19 - Jupyter Notebook | grossglockner high alpine road h

localhost:8888/notebooks/Untitled19.ipynb?kernel_name=python3

jupyter Untitled19 Last Checkpoint: 2 hours ago (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [1]:
```
import math
import warnings
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import seaborn as sns
import plotly.offline as py
import plotly.graph_objs as go
import matplotlib.pyplot as plt
```

In [10]:
```
# Let's import to our data and check the basics.
yt= pd.read_csv('Task 1 YouTube Streamer Analysis-20231109T103241Z-001.zip')
yt
```

Out[10]:

	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments	Links
0	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy...
1	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://youtube.com/channel/UCX6OQ3DkcsbYNE6H8u...
2	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://youtube.com/channel/UCbCmjCuTUZos6Inko4...
3	4	SETIndia	NaN	162600000.0	India	15600.0	166.0	9.0	http://youtube.com/channel/UCpEhnqL0y41EpW2TvW...
4	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http://youtube.com/channel/UCk8GzjMOrta8yxDcKf...
...	...	...	...	...	...	...	...	...	...
995	996	hamzymukbang	NaN	11700000.0	Estados Unidos	397400.0	14000.0	124.0	http://youtube.com/channel/UCPKNKldggioffXPkSm...
996	997	Adaahqueen	NaN	11700000.0	India	1100000.0	92500.0	164.0	http://youtube.com/channel/UCk3fEpqI5kDMf... mLIP

Type here to search | 25°C | ENG | 18:58 09-11-2023

Home Page - Select or create a n | Untitled19 - Jupyter Notebook | grossglockner high alpine road h

localhost:8888/notebooks/Untitled19.ipynb?kernel_name=python3

jupyter Untitled19 Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [11]:
```
yt.head()
```

Out[11]:

	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments	Links
0	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy...
1	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://youtube.com/channel/UCX6OQ3DkcsbYNE6H8u...
2	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://youtube.com/channel/UCbCmjCuTUZos6Inko4...
3	4	SETIndia	NaN	162600000.0	India	15600.0	166.0	9.0	http://youtube.com/channel/UCpEhnqL0y41EpW2TvW...
4	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http://youtube.com/channel/UCk8GzjMOrta8yxDcKf...

In [12]:
```
yt.columns
```

Out[12]:
```
Index(['Rank', 'Username', 'Categories', 'Suscribers', 'Country', 'Visits',
       'Likes', 'Comments', 'Links'],
      dtype='object')
```

In [13]:
```
yt.isnull().sum()
```

Out[13]:
```
Rank            0
Username        0
Categories    306
Suscribers      0
Country         0
Visits          0
Likes           0
Comments        0
Links           0
dtype: int64
```

Type here to search

Earn...

ENG 18:59 09-11-2023

Home Page - Select or create a n | Untitled19 - Jupyter Notebook | grossglockner high alpine road h

localhost:8888/notebooks/Untitled19.ipynb?kernel_name=python3

jupyter Untitled19 Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [14]:
```
yt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   Rank        1000 non-null   int64  
 1   Username    1000 non-null   object 
 2   Categories  694 non-null    object 
 3   Suscribers  1000 non-null   float64
 4   Country     1000 non-null   object 
 5   Visits      1000 non-null   float64
 6   Likes       1000 non-null   float64
 7   Comments    1000 non-null   float64
 8   Links       1000 non-null   object 
dtypes: float64(4), int64(1), object(4)
memory usage: 70.4+ KB
```

In [16]:
```
yt["Categories"].info()
```

```
<class 'pandas.core.series.Series'>
RangeIndex: 1000 entries, 0 to 999
Series name: Categories
Non-Null Count  Dtype 
--------------  ----- 
694 non-null    object
dtypes: object(1)
memory usage: 7.9+ KB
```

In [17]:
```
yt["Country"].info()
```

Type here to search | 25°C | ENG | 19:01 09-11-2023

Home Page - Select or create a n | youtube streamers analysis - Jup

localhost:8888/notebooks/youtube%20streamers%20analysis.ipynb

jupyter youtube streamers analysis Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [14]:
```
yt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Rank        1000 non-null   int64
 1   Username    1000 non-null   object
 2   Categories  694 non-null    object
 3   Suscribers  1000 non-null   float64
 4   Country     1000 non-null   object
 5   Visits      1000 non-null   float64
 6   Likes       1000 non-null   float64
 7   Comments    1000 non-null   float64
 8   Links       1000 non-null   object
dtypes: float64(4), int64(1), object(4)
memory usage: 70.4+ KB
```

In [16]:
```
yt["Categories"].info()
```

```
<class 'pandas.core.series.Series'>
RangeIndex: 1000 entries, 0 to 999
Series name: Categories
Non-Null Count  Dtype
--------------  -----
694 non-null    object
dtypes: object(1)
memory usage: 7.9+ KB
```

In [17]:
```
yt["Country"].info()
```

Type here to search | 25°C | ENG | 19:07 09-11-2023

Home Page - Select or create a n | youtube streamers analysis - Jup

localhost:8888/notebooks/youtube%20streamers%20analysis.ipynb

jupyter youtube streamers analysis Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [17]:
```
yt["Country"].info()
```

```
<class 'pandas.core.series.Series'>
RangeIndex: 1000 entries, 0 to 999
Series name: Country
Non-Null Count  Dtype
--------------  -----
1000 non-null   object
dtypes: object(1)
memory usage: 7.9+ KB
```

In [18]:
```
yt["Country"].mode()
```

Out[18]:
```
0    Estados Unidos
Name: Country, dtype: object
```

In [20]:
```
country_youtubers_count = yt["Country"].value_counts().head(10)
country_youtubers_count
```

Out[20]:
```
Estados Unidos    293
India             241
Unknown           171
Brasil             64
México             58
Indonesia          38
Rusia              25
Tailandia          18
Colombia           16
Filipinas          13
Name: Country, dtype: int64
```

In [21]:
```
import matplotlib.pyplot as plt
```

Type here to search

25°C ENG 19:07 09-11-2023

Home Page - Select or create a n | youtube streamers analysis - Jup

localhost:8888/notebooks/youtube%20streamers%20analysis.ipynb

jupyter youtube streamers analysis Last Checkpoint: 2 hours ago (autosaved)

Logout

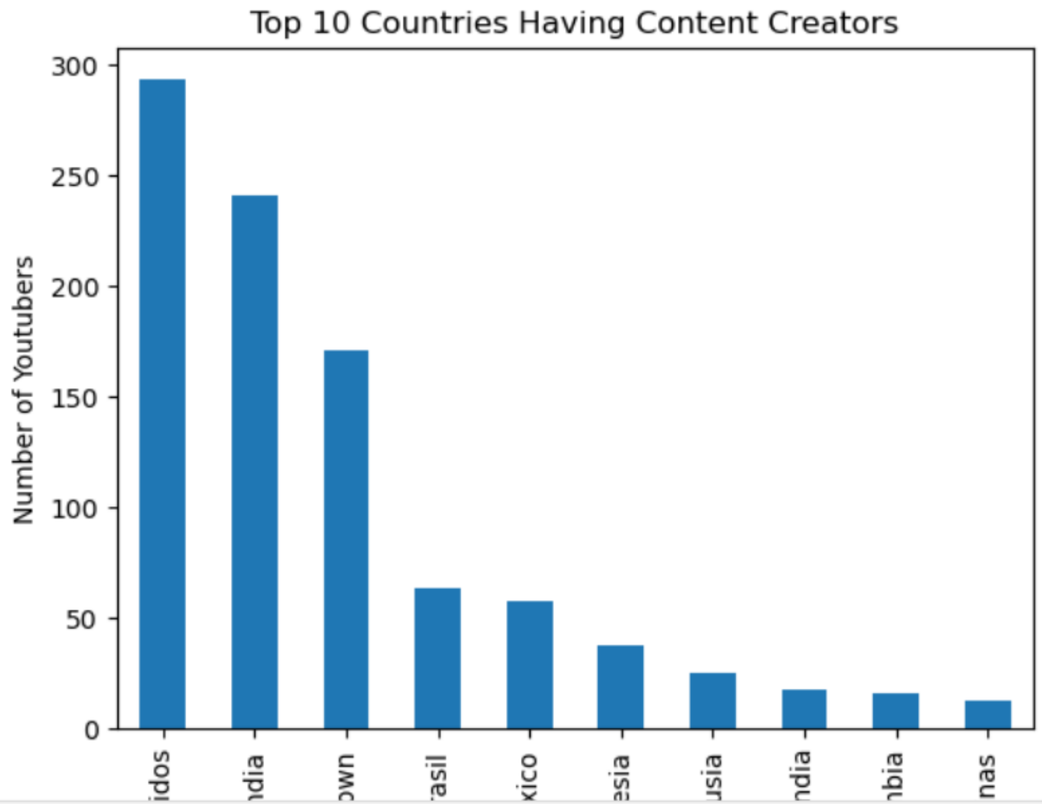
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [21]:

```
import matplotlib.pyplot as plt
country_youtubers_count.plot(kind = "bar")
plt.xlabel("Country")
plt.ylabel("Number of Youtubers")
plt.title("Top 10 Countries Having Content Creators")
plt.show()
```



Type here to search

25°C ENG 19:07 09-11-2023

In [23]:
```
total_categories = yt["Categories"].value_counts()
total_categories.head(15)
```

Out[23]:
```
Música y baile               160
Películas, Animación          61
Música y baile, Películas     41
Vlogs diarios                 37
Noticias y Política           36
Películas, Humor              34
Animación, Videojuegos        34
Animación, Juguetes           29
Animación, Humor              27
Películas                     24
Educación                     24
Animación                     22
Videojuegos                   19
Videojuegos, Humor            17
Música y baile, Animación     16
Name: Categories, dtype: int64
```

In [24]:
```
yt.dropna(subset = 'Categories',inplace = True)
yt.describe()
```

Home Page - Select or create a n | youtube streamers analysis - Jup

localhost:8888/notebooks/youtube%20streamers%20analysis.ipynb

jupyter youtube streamers analysis Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

In [24]:
```
yt.dropna(subset = 'Categories',inplace = True)
yt.describe()
```

Out[24]:

	Rank	Suscribers	Visits	Likes	Comments
count	694.000000	6.940000e+02	6.940000e+02	6.940000e+02	694.000000
mean	495.298271	2.241556e+07	1.210730e+06	5.347360e+04	1558.793948
std	289.222212	1.824123e+07	6.038274e+06	2.979711e+05	7967.470234
min	1.000000	1.170000e+07	0.000000e+00	0.000000e+00	0.000000
25%	244.250000	1.380000e+07	3.692500e+04	5.685000e+02	2.000000
50%	492.500000	1.680000e+07	1.587000e+05	3.550000e+03	78.000000
75%	746.750000	2.390000e+07	8.339000e+05	2.377500e+04	499.750000
max	1000.000000	2.495000e+08	1.174000e+08	5.300000e+06	154000.000000

In [25]:
```
yt.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 694 entries, 0 to 999
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Rank        694 non-null    int64
 1   Username    694 non-null    object
 2   Categories  694 non-null    object
 3   Suscribers  694 non-null    float64
 4   Country     694 non-null    object
 5   Visits      694 non-null    float64
 6   Likes       694 non-null    float64
```

Type here to search | 25°C | ENG | 19:08 09-11-2023

Home Page - Select or create a n | youtube streamers analysis - Jup

localhost:8888/notebooks/youtube%20streamers%20analysis.ipynb

jupyter youtube streamers analysis Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [26]:

```
categories = yt['Categories'].unique()
print("Categories in the dataset:")
for categorie in categories:
    print(categorie)
```

```
Categories in the dataset:
Música y baile
Videojuegos, Humor
Educación
Animación, Juguetes
Películas, Videojuegos
Juguetes
Videojuegos
Películas, Animación
Películas
Noticias y Política
Animación, Humor
Música y baile, Animación
Música y baile, Películas
Películas, Juguetes
Películas, Humor
Vlogs diarios
Videojuegos, Juguetes
Animación, Videojuegos
Animación
Música y baile, Humor
Diseño/arte, DIY y Life Hacks
Ciencia y tecnología
Fitness, Salud y autoayuda
Belleza, Moda
Humor
Comida y bebida
Deportes
```

Type here to search

25°C ENG 19:08 09-11-2023

Home Page - Select or create a n | youtube streamers analysis - Jup

localhost:8888/notebooks/youtube%20streamers%20analysis.ipynb

jupyter youtube streamers analysis Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

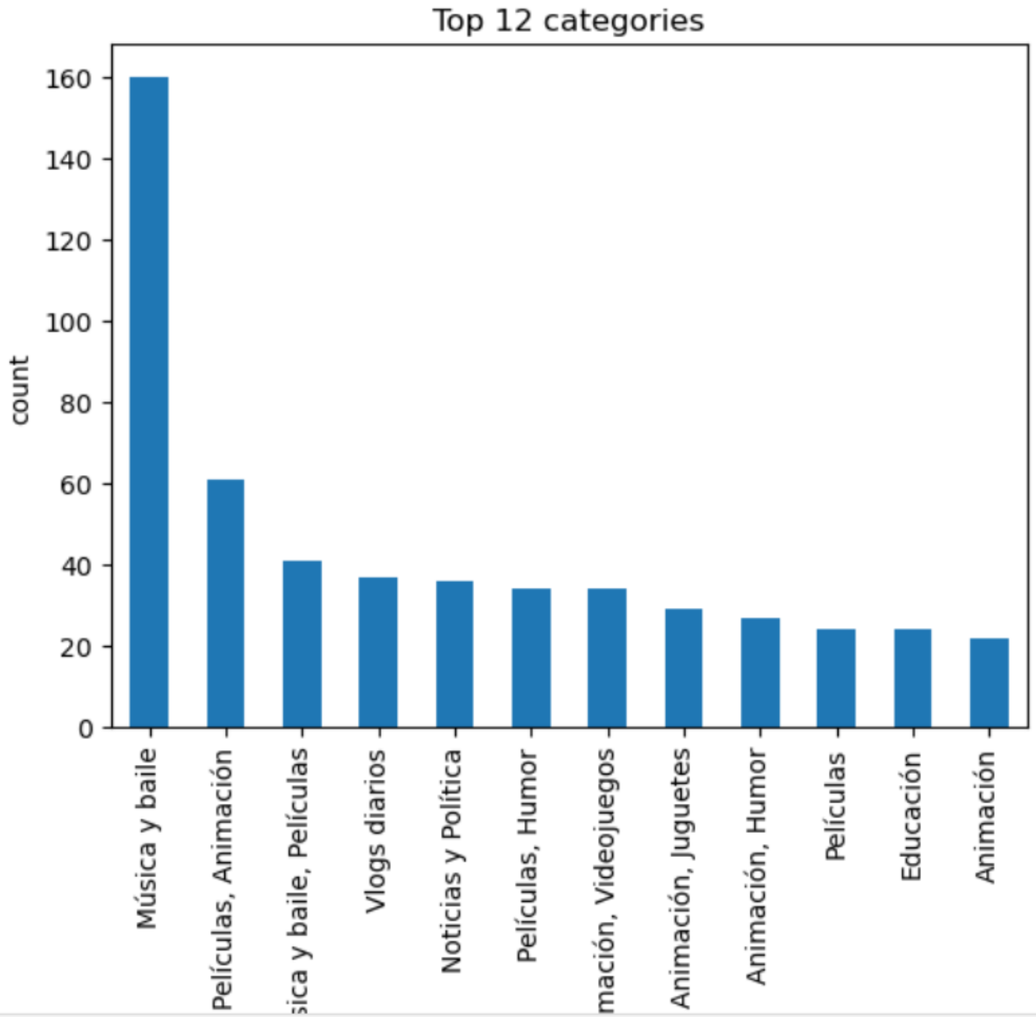
```
Ciencia y tecnología
Fitness, Salud y autoayuda
Belleza, Moda
Humor
Comida y bebida
Deportes
Fitness
Viajes, Espectáculos
Comida y bebida, Salud y autoayuda
Diseño/arte
DIY y Life Hacks, Juguetes
Educación, Juguetes
Juguetes, Coches y vehículos
Música y baile, Juguetes
Animales y mascotas
ASMR
Moda
DIY y Life Hacks
Diseño/arte, Belleza
Coches y vehículos
Animación, Humor, Juguetes
ASMR, Comida y bebida
Comida y bebida, Juguetes
Juguetes, DIY y Life Hacks
```

In [30]:

```
categories = yt["Categories"].value_counts().head(12)
categories.plot(kind = "bar")
plt.xlabel("category")
plt.ylabel("count")
plt.title("Top 12 categories")
plt.show()
```

Type here to search

25°C ENG 19:08 09-11-2023


Top 12 categories
count
0
20
40
60
80
100
120
140
160
Música y baile
Películas, Animación
sica y baile, Películas
Vlogs diarios
Noticias y Política
Películas, Humor
mación, Videojuegos
Animación, Juguetes
Animación, Humor
Películas
Educación
Animación

Home Page - Select or create a n | youtube streamers analysis - Jup

localhost:8888/notebooks/youtube%20streamers%20analysis.ipynb

jupyter youtube streamers analysis Last Checkpoint: 2 hours ago (autosaved)

Logout

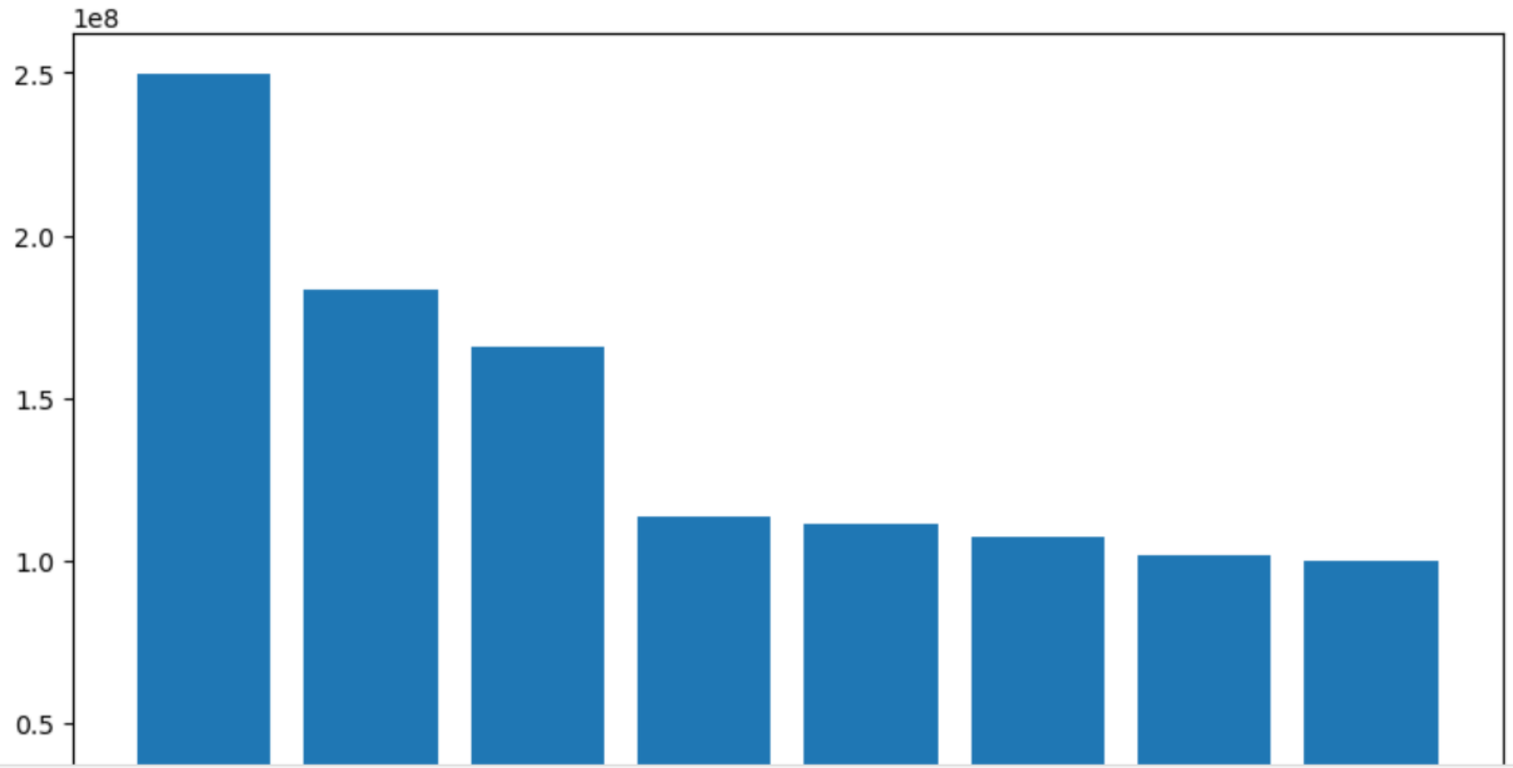
File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run | Code

In [31]:

```
suscribers = yt.head(8)
plt.figure(figsize=(10,6))
plt.bar(suscribers['Username'],suscribers['Suscribers'])
plt.xlabel='suscribers'
plt.ylabel='Youtuber'
plt.title='Top 8 Youtube Suscribers'
plt.show()
```



Type here to search

25°C ENG 19:08 09-11-2023

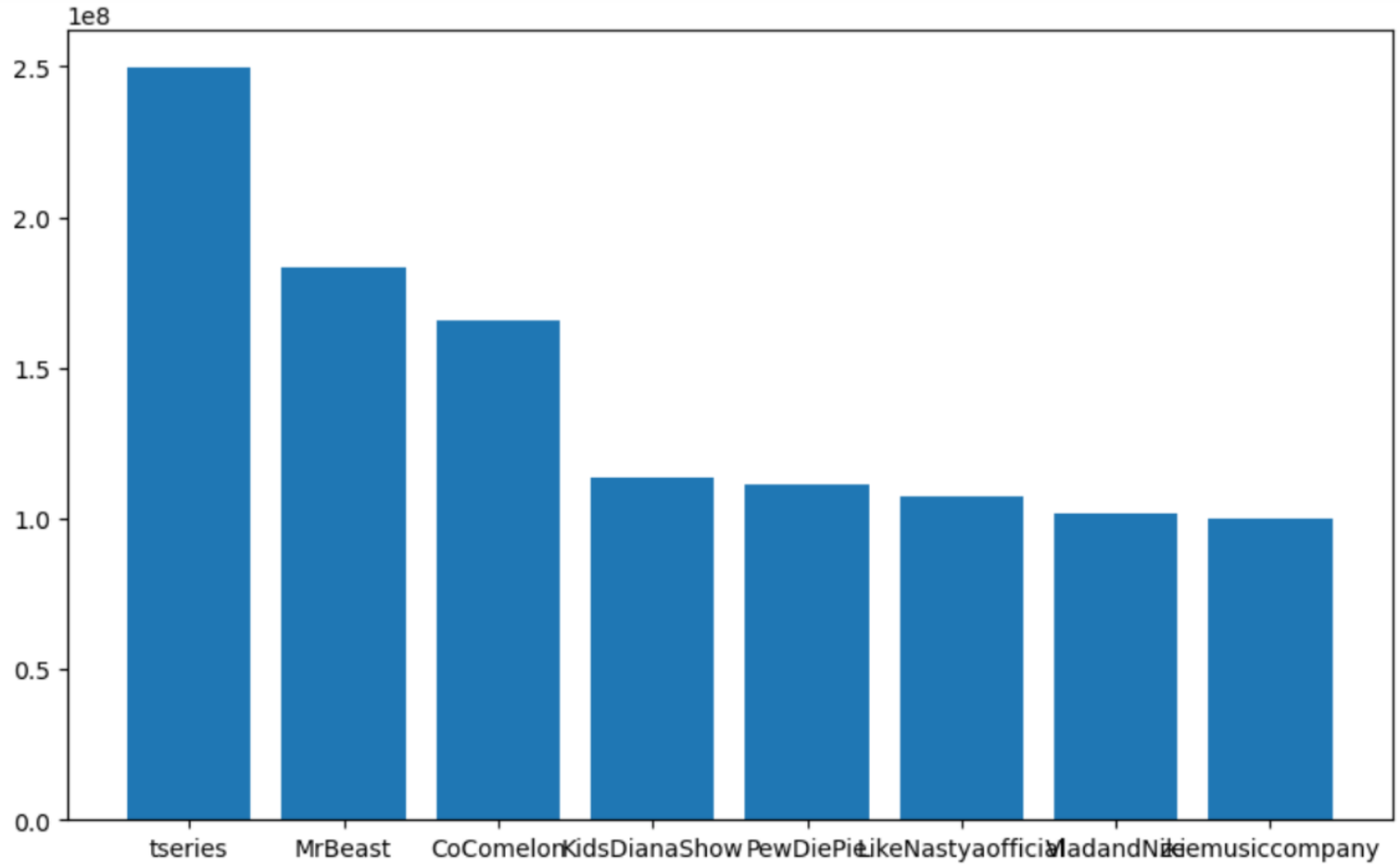
jupyter youtube streamers analysis Last Checkpoint: 2 hours ago (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted Python 3 (ipykernel)

Run Code



In [ ]: