



## youtube analysis project

File Edit View Insert Runtime Tools Help All changes saved

+ Code

+ Text



## DATA ANALYSIS ON TOP YOUTUBERS



```
▶ import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
df=pd.read_csv("youtubers_df.csv")
df
```



	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments	Links
0	1	tseries	Música y baile	249500000.0	India	862000.0	2700.0	78.0	<a href="http://youtube.com/channel/UCq-FJ5jknLsUf-MWSy...">http://youtube.com/channel/UCq-FJ5jknLsUf-MWSy...</a>
1	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	<a href="http://youtube.com/channel/UCX6OQ3DkcsbYNE6H8u...">http://youtube.com/channel/UCX6OQ3DkcsbYNE6H8u...</a>
2	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	<a href="http://youtube.com/channel/UCbCmjCuTUZos6lnko4...">http://youtube.com/channel/UCbCmjCuTUZos6lnko4...</a>
3	4	SETIndia		NaN	162600000.0	India	15600.0	166.0	<a href="http://youtube.com/channel/UCpEhnqL0y41EpW2TvW...">http://youtube.com/channel/UCpEhnqL0y41EpW2TvW...</a>
4	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	<a href="http://youtube.com/channel/UCk8GzjMOrta8yxDcKf...">http://youtube.com/channel/UCk8GzjMOrta8yxDcKf...</a>
...	...	...	...	...	...	...	...	...	...
995	996	hamzymukbang		NaN	11700000.0	Estados Unidos	397400.0	14000.0	<a href="http://youtube.com/channel/UCPKNKLdggioffXPkSm...">http://youtube.com/channel/UCPKNKLdggioffXPkSm...</a>
996	997	Adaahqueen		NaN	11700000.0	India	1100000.0	92500.0	<a href="http://youtube.com/channel/UCK3fPql5kDMf_mUP...">http://youtube.com/channel/UCK3fPql5kDMf_mUP...</a>
997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	<a href="http://youtube.com/channel/UCdrHrQf0o0TO8YDntX...">http://youtube.com/channel/UCdrHrQf0o0TO8YDntX...</a>
998	999	PenMultiplex		NaN	11700000.0	India	14000.0	81.0	<a href="http://youtube.com/channel/UCOByBrdrtQ20BU9PxH...">http://youtube.com/channel/UCOByBrdrtQ20BU9PxH...</a>
999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	<a href="http://youtube.com/channel/UCOJgc1p2hJ4GZi6pQQ...">http://youtube.com/channel/UCOJgc1p2hJ4GZi6pQQ...</a>

1000 rows x 9 columns



Search



colab.research.google.com/drive/1iIBurZrqb0qOQuwW1pq3aDoDoxVu8uq#scrollTo=F850uP4\_ZVA4

Comment Reconnect

youtube analysis project

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text 1000 rows x 9 columns

Double-click (or enter) to edit

{x}

df.shape  
(1000, 9)

DATA CLEANING AND DATA EXPLORATION

```
[ ] df.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   Subscribers 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
```

```
[ ] df.columns
```

22°C sunny

Search

CO youtube analysis project ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

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

CSV df.rename(columns= {'Suscribers': 'Subscribers'}, inplace=True)

[ ] df.dtypes

Rank	int64
Username	object
Categories	object
Subscribers	float64
Country	object
Visits	float64
Likes	float64
Comments	float64
Links	object
dtype:	object

[ ] df.isnull().sum()

Rank	0
Username	0
Categories	306
Subscribers	0
Country	0
Visits	0
Likes	0
Comments	0
Links	0
dtype:	int64

20°C 8:45 AM

Q Search

File Edit View Insert Runtime Tools Help

All changes saved

+ Code + Text

[ ]

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

CSV df.rename(columns= {'Suscribers': 'Subscribers'}, inplace=True)

[ ] df.dtypes

Rank	int64
Username	object
Categories	object
Subscribers	float64
Country	object
Visits	float64
Likes	float64
Comments	float64
Links	object
dtype:	object

[ ] df.isnull().sum()

Rank	0
Username	0
Categories	306
Subscribers	0
Country	0
Visits	0
Likes	0
Comments	0
Links	0
dtype:	int64

20°C 8:45 AM

Q Search

File Edit View Insert Runtime Tools Help

All changes saved

+ Code + Text

[ ]

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

CSV df.rename(columns= {'Suscribers': 'Subscribers'}, inplace=True)

[ ] df.dtypes

Rank	int64
Username	object
Categories	object
Subscribers	float64
Country	object
Visits	float64
Likes	float64
Comments	float64
Links	object
dtype:	object

[ ] df.isnull().sum()

Rank	0
Username	0
Categories	306
Subscribers	0
Country	0
Visits	0
Likes	0
Comments	0
Links	0
dtype:	int64

## CO youtube analysis project ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

```
[ ] #df['categories'] = 'unknown'  
#df.fillna({'categories': 'unknown'}, inplace=True)  
df['Categories'].fillna('Unknown', inplace=True)
```

```
{x} df.isnull().sum()
```

```
Rank      0  
Username  0  
Categories 0  
Subscribers 0  
Country   0  
Visits    0  
Likes     0  
Comments  0  
Links     0  
dtype: int64
```

```
[ ] df.duplicated()
```

```
0      False  
1      False  
2      False  
3      False  
4      False  
...  
995    False  
996    False  
997    False  
998    False  
999    False  
Length: 1000, dtype: bool
```

CO youtube analysis project ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

df['Subscribers'].describe()

	count	mean	std	min	25%	50%	75%	max	Name	Dtype
{x}	1.000000e+03	2.189440e+07	1.682775e+07	1.170000e+07	1.380000e+07	1.675000e+07	2.370000e+07	2.495000e+08	Subscribers	float64

[ ] df['z\_scores'] = (df['Subscribers'] - df['Subscribers'].mean())/df['Subscribers'].std()

df

Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Link
0	tseries	Música y baile	249500000.0	India	862000.0	2700.0	78.0	<a href="http://youtube.com/channel/UCv24jy">http://youtube.com/channel/UCv24jy</a>
1	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	<a href="http://youtube.com/channel/UCXVJGJ">http://youtube.com/channel/UCXVJGJ</a>
2	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	<a href="http://youtube.com/channel/UCXSVmP">http://youtube.com/channel/UCXSVmP</a>
3	SETIndia	Unknown	162600000.0	India	15600.0	166.0	9.0	<a href="http://youtube.com/channel/UCSEnqL">http://youtube.com/channel/UCSEnqL</a>
4	KidsDianaShow	Animacion, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	<a href="http://youtube.com/channel/UCkbrsgj">http://youtube.com/channel/UCkbrsgj</a>
...	...	...	...	...	...	...	...	...
995	hamzymukbang	Unknown	11700000.0	Estados Unidos	397400.0	14000.0	124.0	<a href="http://youtube.com/channel/UCPKKMK">http://youtube.com/channel/UCPKKMK</a>
996	Acaahqueen	Unknown	11700000.0	India	1100000.0	92500.0	134.0	<a href="http://youtube.com/channel/UCX9S31">http://youtube.com/channel/UCX9S31</a>

Q Search

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

df['Subscribers'].describe()

	count	mean	std	min	25%	50%	75%	max	Name	Dtype
{x}	1.000000e+03	2.189440e+07	1.682775e+07	1.170000e+07	1.380000e+07	1.675000e+07	2.370000e+07	2.495000e+08	Subscribers	float64

[ ] df['z\_scores'] = (df['Subscribers'] - df['Subscribers'].mean())/df['Subscribers'].std()

df

Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Link
0	tseries	Música y baile	249500000.0	India	862000.0	2700.0	78.0	<a href="http://youtube.com/channel/UCv24jy">http://youtube.com/channel/UCv24jy</a>
1	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	<a href="http://youtube.com/channel/UCXVJGJ">http://youtube.com/channel/UCXVJGJ</a>
2	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	<a href="http://youtube.com/channel/UCXSVmP">http://youtube.com/channel/UCXSVmP</a>
3	SETIndia	Unknown	162600000.0	India	15600.0	166.0	9.0	<a href="http://youtube.com/channel/UCSEnqL">http://youtube.com/channel/UCSEnqL</a>
4	KidsDianaShow	Animacion, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	<a href="http://youtube.com/channel/UCkbrsgj">http://youtube.com/channel/UCkbrsgj</a>
...	...	...	...	...	...	...	...	...
995	hamzymukbang	Unknown	11700000.0	Estados Unidos	397400.0	14000.0	124.0	<a href="http://youtube.com/channel/UCPKKMK">http://youtube.com/channel/UCPKKMK</a>
996	Acaahqueen	Unknown	11700000.0	India	1100000.0	92500.0	134.0	<a href="http://youtube.com/channel/UCX9S31">http://youtube.com/channel/UCX9S31</a>

Q Search

Colab research.google.com/drive/1iiYBuzZqz0qOQuwVWipz5Pd0caYu8uq#scrollTo=3000000000000000000

# youtube analysis project

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

	996	997	Adaahqueen	Unknown	11700000.0	India	1100000.0	92500.0	164.0	http://youtube.com/channel/UCK3fFpqI5kD
{x}	997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	http://youtube.com/channel/UCdrHrQf0o0
Key	998	999	PenMultiplex	Unknown	11700000.0	India	14000.0	81.0	1.0	http://youtube.com/channel/UCObyBrdrtQ2
1000 rows x 10 columns	999	1000	OneindiaHindi	Noticias y Politica	11700000.0	India	2200.0	31.0	1.0	http://youtube.com/channel/UCOjgc1p2hJ

Checking outliers based on like

```
[ ] df['z_likes'] =(df['Likes'] -df['Likes'].mean()) / df['Likes'].std()
df
```

Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links	
0	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy.	
1	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://youtube.com/channel/UCX6OQ3DkcsbYNEdH8u.	
2	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://youtube.com/channel/UCbCmjCuTUZos6lnko4	
3	SETIndia	Unknown	162600000.0	India	15600.0	166.0	9.0	http://youtube.com/channel/UCpEhnqL0y41EpW2TvW.	
4	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http://youtube.com/channel/UCk8GzjMOrta8yxDcKf.	
...	...	...	...	...	...	...	...	...	
995	996	hamzymukbang	Unknown	11700000.0	Estados Unidos	397400.0	14000.0	124.0	http://youtube.com/channel/UCPKNKLdggi0ffXPkSm.

22°C

Search PRO L M A T S N X

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

	Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links	z_scores
996	997	Adaahqueen	Unknown	11700000.0	India	1100000.0	92500.0	164.0	<a href="http://youtube.com/channel/UCk3fFpqj5kDMf_mUP...">http://youtube.com/channel/UCk3fFpqj5kDMf_mUP...</a>	-0.60580
997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	<a href="http://youtube.com/channel/UCdrHrQf0o0TO8YDntX...">http://youtube.com/channel/UCdrHrQf0o0TO8YDntX...</a>	-0.60580
998	999	PenMultiplex	Unknown	11700000.0	India	14000.0	81.0	1.0	<a href="http://youtube.com/channel/UCObbyBrdrtQ20BU9PxH...">http://youtube.com/channel/UCObbyBrdrtQ20BU9PxH...</a>	-0.60580
999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	<a href="http://youtube.com/channel/UCOjgc1p2hJ4GZi6pQQ...">http://youtube.com/channel/UCOjgc1p2hJ4GZi6pQQ...</a>	-0.60580

1000 rows × 11 columns

### CHECKING OUTLIERS ON COMMENTS

```
[ ] df['Comments'].describe()
```

count	mean	std	min	25%	50%	75%	max
1000.000000	1288.768000	6778.188308	0.000000	2.000000	67.000000	472.000000	154000.000000

Name: Comments, dtype: float64

```
[ ] df['z_comments'] = (df.Comments - df.Comments.mean())/df.Comments.std()
df
```

Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links	z_scores
0	1	Música y baile	310500000.0	India	86000.0	2700.0	70.0	<a href="https://youtube.com/channel/UCk3fFpqj5kDMf_mUP...">https://youtube.com/channel/UCk3fFpqj5kDMf_mUP...</a>	-0.60580

22°C Sunny

youtube analysis project											
File Edit View Insert Runtime Tools Help All changes saved											
+ Code + Text											
Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links	z_sc		
0	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy...	13.52	
1	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://youtube.com/channel/UCX6OQ3DkcsbYNE6H8u...	9.60	
2	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://youtube.com/channel/UCbCmjCuTUZos6Inko4...	8.53	
3	4	SETIndia	Unknown	162600000.0	India	15600.0	166.0	9.0	http://youtube.com/channel/UCpEhnqL0y41EpW2TvW...	8.36	
4	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http://youtube.com/channel/Uck8GzjMOrta8yxDcKf...	5.44	
...	...	...	...	...	...	...	...	...	...	...	
995	996	hamzymukbang	Unknown	11700000.0	Estados Unidos	397400.0	14000.0	124.0	http://youtube.com/channel/UCPKNKldggioffXPkSm...	-0.60	
996	997	Adaahqueen	Unknown	11700000.0	India	1100000.0	92500.0	164.0	http://youtube.com/channel/Uck3fFpqI5kDMf__mUP...	-0.60	
997	998	LittleAngellIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	http://youtube.com/channel/UCdhrQf0o0TO8YDntX...	-0.60	
998	999	PenMultiplex	Unknown	11700000.0	India	14000.0	81.0	1.0	http://youtube.com/channel/UCObYBrdtQ20BU9PxH...	-0.60	
999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	http://youtube.com/channel/UCOjgc1p2hJ4GZi6pQQ...	-0.60	
1000 rows × 12 columns											
OUTLIERS ON VISITS											
22°C Sunny	Search	PRE	L	O	I	A	S	N	Spotify	Notepad	File Explorer

## youtube analysis project ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

df['visits'].describe()

```
{x} count    1.000000e+03
mean     1.209446e+06
std      5.229942e+06
min      0.000000e+00
25%      3.197500e+04
50%      1.744500e+05
75%      8.654750e+05
max      1.174000e+08
Name: visits, dtype: float64
```

[ ] df['z\_visits'] = (df.Visits -df.Visits.mean())/(df.Visits.std())

		Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links
0	1	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	<a href="http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy...">http://youtube.com/channel/UCq-Fj5jknLsUf-MWSy...</a>
1	2	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	<a href="http://youtube.com/channel/UCX6OQ3DkcsbYNEdH8u...">http://youtube.com/channel/UCX6OQ3DkcsbYNEdH8u...</a>
2	3	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	<a href="http://youtube.com/channel/UCbCmjCuTUZos6Inko4...">http://youtube.com/channel/UCbCmjCuTUZos6Inko4...</a>
3	4	4	SETIndia	Unknown	162600000.0	India	15600.0	166.0	9.0	<a href="http://youtube.com/channel/UCpEhnqL0y41EpW2TvW...">http://youtube.com/channel/UCpEhnqL0y41EpW2TvW...</a>
4	5	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	<a href="http://youtube.com/channel/UCk8GzjMOrta8yxDcKf...">http://youtube.com/channel/UCk8GzjMOrta8yxDcKf...</a>
...	...	...	...	...	...	...	...	...	...	...
995	996	996	hamzymukbang	Unknown	11700000.0	Estados Unidos	397400.0	14000.0	124.0	<a href="http://youtube.com/channel/UCPKNKldggioffXPkSm...">http://youtube.com/channel/UCPKNKldggioffXPkSm...</a>



youtube analysis project																			
File Edit View Insert Runtime Tools Help All changes saved																			
+ Code	+ Text																		
	[ ] df.rename(columns = {'z_scores' : 'z_subscribers'},inplace=True)																		
	df[(df.z_subscribers > 3)   (df.z_subscribers < -3)]																		
Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links	z										
0	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http://youtube.com/channel/UCq-Fj5jnLsUf-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/UCbCmjcTuZos6Inko4...										
3	4	SETIndia	Unknown	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/UCk8GzMOrta8yxDcKf...										
5	6	PewDiePie	Películas, Videojuegos	111500000.0	Estados Unidos	2400000.0	197300.0	4900.0	http://youtube.com/channel/UC-IHJZR3Gqxm24_Vd_...										
6	7	LikeNastyofficial	Juguetes	107500000.0	Unknown	2600000.0	28000.0	0.0	http://youtube.com/channel/UCJplp5SjeGSdVdwsfb...										
7	8	VladandNiki	Animación, Juguetes	101400000.0	Unknown	4100000.0	22100.0	0.0	http://youtube.com/channel/UCvIE5gTbOvjiolFIEm...										
8	9	zeemusiccompany	Música y baile	99700000.0	India	74300.0	2600.0	32.0	http://youtube.com/channel/UCFFbwnve3yF62-tVXk...										
9	10	WWE	Videojuegos	97200000.0	Estados Unidos	184500.0	6300.0	214.0	http://youtube.com/channel/UCJ5v_MCY6GNUBTO8-D...										
10	11	BLACKPINK	Música y baile	91300000.0	Estados Unidos	863200.0	146900.0	3400.0	http://youtube.com/channel/UComHUn--16B90oW2L6...										
22°C Sunny			Search																



+ Code + Text

[ ]	16	17	zeetv	Unknown	72500000.0	India	7300.0	144.0	1.0	http://youtube.com/channel/UCppHT7SZKKvar4Oc9J...
New Data Filtered										
	Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links	z_s
17	18	Pinkfong	Animación, Juguetes	69600000.0	Unknown	506400.0	397.0	0.0	http://youtube.com/channel/UCcdwLMPsaU2ezNSJU1...	
18	19	ChuChuTV	Juguetes	67500000.0	Unknown	652100.0	17800.0	0.0	http://youtube.com/channel/UCBnZ16ahKA2DZ_T5W0...	
19	20	KondZilla	Música y baile	66600000.0	Brasil	30400.0	1100.0	4.0	http://youtube.com/channel/UCffDXn7ycAzwL2LDlb...	
20	21	colorstv	Unknown	66600000.0	India	14700.0	413.0	3.0	http://youtube.com/channel/UC55IWqFLDH1Xp7iu1...	
21	22	filmigaane	Música y baile	66400000.0	India	33000.0	392.0	18.0	http://youtube.com/channel/UCP6uH_XlsxrXwZQ4DI...	
...	...	...	...	...	...	...	...	...	...	...
995	996	hamzymukbang	Unknown	11700000.0	Estados Unidos	397400.0	14000.0	124.0	http://youtube.com/channel/UCPKNKldggioffXPkSm...	
996	997	Adaahqueen	Unknown	11700000.0	India	1100000.0	92500.0	164.0	http://youtube.com/channel/UCk3fFpqI5kDMf__mUP...	
997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	http://youtube.com/channel/UCdrHrQf0o0TO8YDntX...	
998	999	PenMultiplex	Unknown	11700000.0	India	14000.0	81.0	1.0	http://youtube.com/channel/UCObyBrdrtQ20BU9PxH...	
999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	http://youtube.com/channel/UCOjgc1p2hJ4GZi6pQQ...	





## youtube analysis project ⭐

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

983 rows x 13 columns



```
❶ df = new_df
df.shape
```

```
❷ (983, 13)
```

```
[ ] df[(df.z_likes > 3) | (df.z_likes < -3)]
```

	Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links	z_
136	137	MrBeast2	Vlogs diarios	31300000.0	Estados Unidos	83100000.0	5000000.0	11600.0	http://youtube.com/channel/UC4-79UOIP48-QNGgCk...	
153	154	DaFugBoom	Animación, Humor	29800000.0	Estados Unidos	52700000.0	1700000.0	82800.0	http://youtube.com/channel/UCsSsgPaZ2GSmO6i8C...	
341	342	triggeredinsaan	Humor	20400000.0	India	11100000.0	1400000.0	38000.0	http://youtube.com/channel/UCfLuT3JwLx8rvHjHT...	
488	489	BeastPhilanthropy	Comida y bebida	16900000.0	Estados Unidos	21500000.0	952100.0	24000.0	http://youtube.com/channel/UCAiLfjNXkNv24uhpzU...	
734	735	Mythpat	Unknown	13900000.0	India	11700000.0	1200000.0	15800.0	http://youtube.com/channel/UCx6F-rETGiz7xf_vkM...	

```
[ ] new_df = df[(df.z_likes < 3) & (df.z_likes > -3)]
new_df
```

Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links	z_
		Animación							



22°C

Sunny



```
new_df = df[(df.z_likes < 3) & (df.z_likes > -3)]  
new_df
```

	Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links
	17	18	Pinkfong	Animación, Juguetes	69600000.0	Unknown	506400.0	397.0	0.0 <a href="http://youtube.com/channel/UCcdwLMPsaU2ezNSJU1...">http://youtube.com/channel/UCcdwLMPsaU2ezNSJU1...</a>
	18	19	ChuChuTV	Juguetes	67500000.0	Unknown	652100.0	17800.0	0.0 <a href="http://youtube.com/channel/UCBnZ16ahKA2DZ_T5W0...">http://youtube.com/channel/UCBnZ16ahKA2DZ_T5W0...</a>
	19	20	KondZilla	Música y baile	66600000.0	Brasil	30400.0	1100.0	4.0 <a href="http://youtube.com/channel/UCffDXn7ycAzwL2LDlb...">http://youtube.com/channel/UCffDXn7ycAzwL2LDlb...</a>
	20	21	colorstv	Unknown	66600000.0	India	14700.0	413.0	3.0 <a href="http://youtube.com/channel/UC55IWqFLDH1Xp7iu1...">http://youtube.com/channel/UC55IWqFLDH1Xp7iu1...</a>
	21	22	filmigaane	Música y baile	66400000.0	India	33000.0	392.0	18.0 <a href="http://youtube.com/channel/UCP6uH_XlsxrXwZQ4Dl...">http://youtube.com/channel/UCP6uH_XlsxrXwZQ4Dl...</a>
	...	...	...	...	...	...	...	...	...
	995	996	hamzymukbang	Unknown	11700000.0	Estados Unidos	397400.0	14000.0	124.0 <a href="http://youtube.com/channel/UCPKNKldggioffXPkSm...">http://youtube.com/channel/UCPKNKldggioffXPkSm...</a>
	996	997	Adaahqueen	Unknown	11700000.0	India	1100000.0	92500.0	164.0 <a href="http://youtube.com/channel/UCk3fFpqI5kDMf__mUP...">http://youtube.com/channel/UCk3fFpqI5kDMf__mUP...</a>
	997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0 <a href="http://youtube.com/channel/UCdrHrQf0o0TO8YDntX...">http://youtube.com/channel/UCdrHrQf0o0TO8YDntX...</a>
	998	999	PenMultiplex	Unknown	11700000.0	India	14000.0	81.0	1.0 <a href="http://youtube.com/channel/UCObyBrdrtQ20BU9PxH...">http://youtube.com/channel/UCObyBrdrtQ20BU9PxH...</a>
	999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0 <a href="http://youtube.com/channel/UCOjgc1p2hJ4GZi6pQQ...">http://youtube.com/channel/UCOjgc1p2hJ4GZi6pQQ...</a>

978 rows × 13 columns



22°C  
Sunny



Search



+ Code + Text

▶ df = new\_df  
df.shape

▶ (978, 13)

+ Code + Text

Add text cell

[ ] df[(df.z\_comments > 3) | (df.z\_comments < -3)]

Rank		Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links
43	44	A4a4a4a4	Animación, Humor	47300000.0	Rusia	9700000.0	330400.0	22000.0	<a href="http://youtube.com/channel/UC2tsySbe9TNrl-xh2l..">http://youtube.com/channel/UC2tsySbe9TNrl-xh2l..</a>
123	124	MRINDIANHACKER	Unknown	32600000.0	India	6500000.0	617400.0	26000.0	<a href="http://youtube.com/channel/UCSiDGb0MnHFGjs4E2W..">http://youtube.com/channel/UCSiDGb0MnHFGjs4E2W..</a>
177	178	DanTDM	Animación, Videojuegos	27800000.0	Estados Unidos	3500000.0	285000.0	52500.0	<a href="http://youtube.com/channel/UCS5Oz6CHmeoF7vSad0..">http://youtube.com/channel/UCS5Oz6CHmeoF7vSad0..</a>
436	437	BispoBrunoLeonardo	Música y baile	18000000.0	Brasil	762100.0	276400.0	154000.0	<a href="http://youtube.com/channel/UCVNouUw3dI5JYVCxh..">http://youtube.com/channel/UCVNouUw3dI5JYVCxh..</a>
958	959	dojacat	Música y baile	11900000.0	Estados Unidos	13600000.0	395300.0	73000.0	<a href="http://youtube.com/channel/UCzpl23pGTHVYqvKsgY..">http://youtube.com/channel/UCzpl23pGTHVYqvKsgY..</a>

◀▶ new\_df = df[(df.z\_comments < 3) & (df.z\_comments > -3)]  
new\_df

Rank		Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links
17	18	Pinkfong	Animación, Juguetes	69600000.0	Unknown	506400.0	397.0	0.0	<a href="http://youtube.com/channel/UCcdwLMPsaU2ezNSJU1...">http://youtube.com/channel/UCcdwLMPsaU2ezNSJU1...</a>

**CO** youtube analysis project ☆

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

	998	999	PenMultiplex	Unknown	11700000.0	India	14000.0	81.0	1.0	http://youtube.com/channel/UCObbyBrdrtQ20BU9PxH...
▶	999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	http://youtube.com/channel/UCOjgc1p2hJ4GZi6pQQ...

973 rows × 13 columns

```
[ ] df = new_df
df.shape

(973, 13)

[ ] df[(df.z_visits > 3) | (df.z_visits < -3)]
```

Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links	
288	289	VillageCookingChannel	Unknown	22500000.0	India	21500000.0	321500.0	5900.0	http://youtube.com/channel/UCK3JZr7eS3pg5AGE...
390	391	jaanvipatel	Unknown	19100000.0	India	19100000.0	0.0	2200.0	http://youtube.com/channel/UCRqs0TZQqe5nnwP-B...

Generate Using ... 10 random numbers using numpy

```
[ ] new_df = df[(df.z_visits < 3) & (df.z_visits > -3)]
new_df
```

Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	Links	
17	18	Pinkfong	Animación, Juguetes	69600000.0	Unknown	506400.0	397.0	0.0	http://youtube.com/channel/UCcdwLMPsaU2ezNSJU1...

22°C Sunny

Search

PRE

**CO** youtube analysis project ★

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

▶	998	999	PenMultiplex	Unknown	11700000.0	India	14000.0	81.0	1.0	http://youtube.com/channel/UCObyBrdtQ20BU9PxH..
➡	999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	http://youtube.com/channel/UCOjgc1p2hJ4GZi6pQQ..

971 rows × 13 columns

```
[ ] df = new_df
df.shape

(971, 13)
```

## CONCLUSION

```
[ ] !pip install googletrans
import googletrans

Requirement already satisfied: googletrans in /usr/local/lib/python3.10/dist-packages (3.0.0)
Requirement already satisfied: httpx==0.13.3 in /usr/local/lib/python3.10/dist-packages (from googletrans) (0.13.3)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-packages (from httpx==0.13.3->googletrans) (2023.11.17)
Requirement already satisfied: hstspreload in /usr/local/lib/python3.10/dist-packages (from httpx==0.13.3->googletrans) (2023.1.1)
Requirement already satisfied: sniffio in /usr/local/lib/python3.10/dist-packages (from httpx==0.13.3->googletrans) (1.3.0)
Requirement already satisfied: chardet==3.* in /usr/local/lib/python3.10/dist-packages (from httpx==0.13.3->googletrans) (3.0.4)
Requirement already satisfied: idna==2.* in /usr/local/lib/python3.10/dist-packages (from httpx==0.13.3->googletrans) (2.10)
Requirement already satisfied: rfc3986<2,>=1.3 in /usr/local/lib/python3.10/dist-packages (from httpx==0.13.3->googletrans) (1.5.0)
Requirement already satisfied: httpcore==0.9.* in /usr/local/lib/python3.10/dist-packages (from httpx==0.13.3->googletrans) (0.9.1)
Requirement already satisfied: h11<0.10,>=0.8 in /usr/local/lib/python3.10/dist-packages (from httpcore==0.9.*->httpx==0.13.3->googletrans)
Requirement already satisfied: h2==3.* in /usr/local/lib/python3.10/dist-packages (from httpcore==0.9.*->httpx==0.13.3->googletrans) (3.2.0)
Requirement already satisfied: hyperframe<6,>=5.2.0 in /usr/local/lib/python3.10/dist-packages (from h2==3.*->httpcore==0.9.*->httpx==0.13.3->googletrans)
Requirement already satisfied: hpack<4,>=3.0 in /usr/local/lib/python3.10/dist-packages (from h2==3.*->httpcore==0.9.*->httpx==0.13.3->googletrans)
```



+ Code + Text

Requirement already satisfied: hpack&lt;4,&gt;=3.0 in /usr/local/lib/python3.10/dist-packages (from h2==3.\*-&gt;httpcore==0.9.\*-&gt;httpx==0.13.3-)

```
● from googletrans import Translator  
import pandas as pd  
t = googletrans.client.Translator()
```

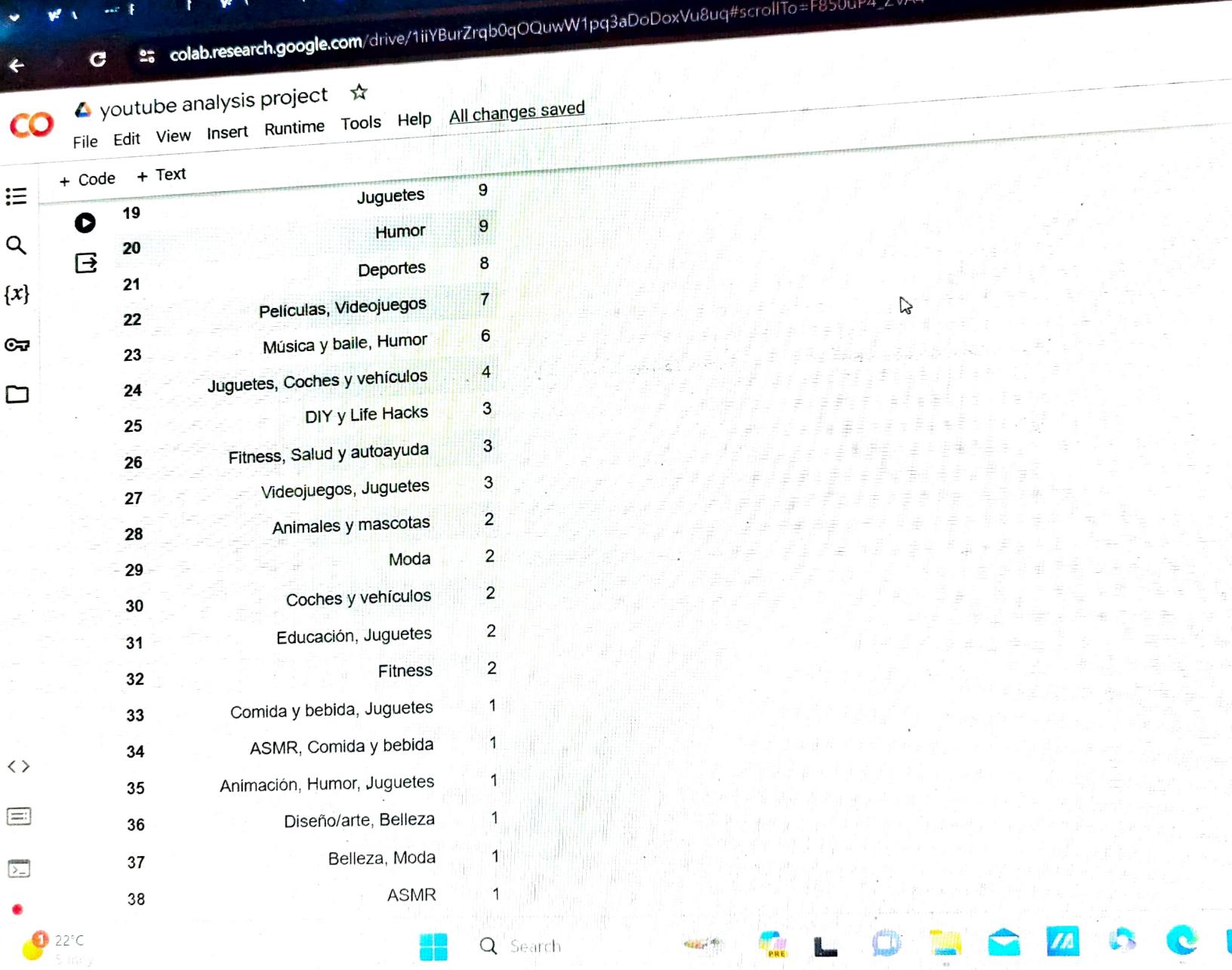
+ Code

+ Text

```
[ ] most_pop = pd.DataFrame(df['Categories'].value_counts())  
pop_df = most_pop.reset_index()  
pop_df.columns = ['Categories', 'Count']  
pop_df
```

Add text cell

	Categories	Count	grid icon
0	Unknown	299	bar chart icon
1	Música y baile	153	edit icon
2	Películas, Animación	60	
3	Música y baile, Películas	41	
4	Noticias y Política	36	
5	Vlogs diarios	36	
6	Películas, Humor	34	
7	Animación, Videojuegos	33	
8	Animación, Juguetes	27	
9	Animación, Humor	25	
10	Películas	23	



```
In [38]: pop_df['Categories'] = temp  
pop_df.head(11)
```

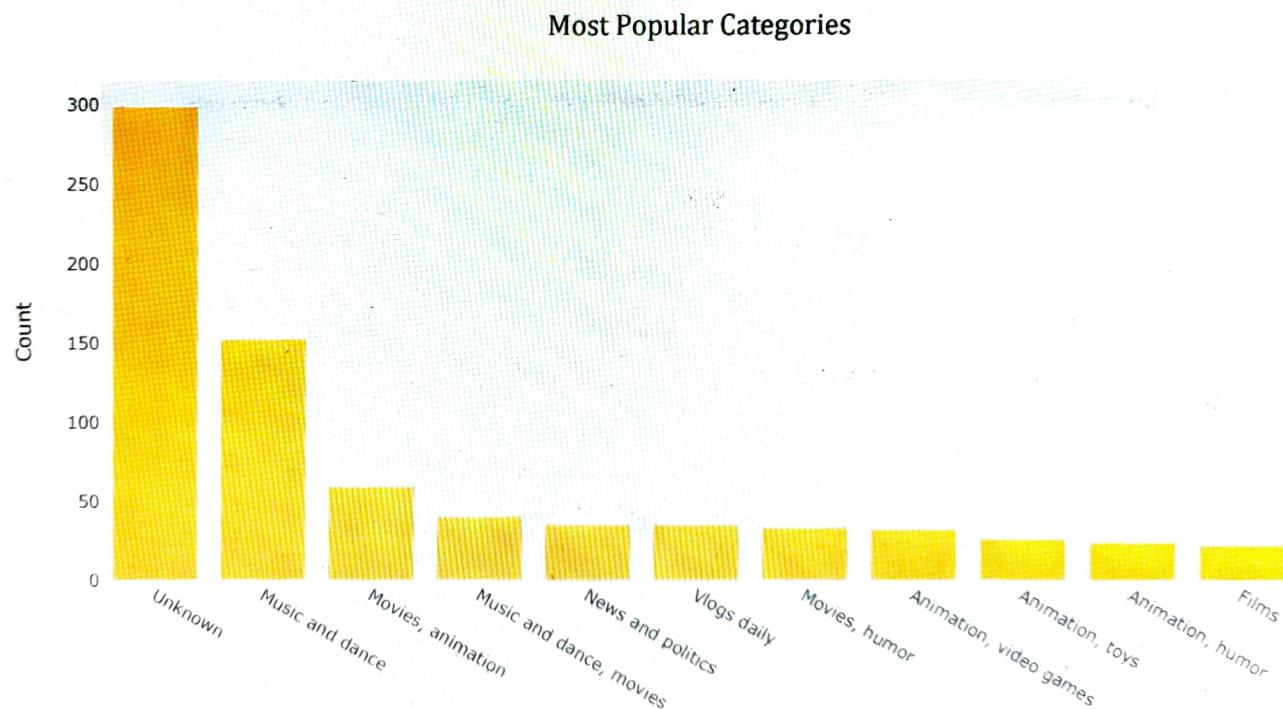
Out[38]:

	Categories	Count
0	Unknown	299
1	Music and dance	153
2	Movies, animation	60
3	Music and dance, movies	41
4	News and politics	36
5	Vlogs daily	36
6	Movies, humor	34
7	Animation, video games	33
8	Animation, toys	27
9	Animation, humor	25
10	Films	23

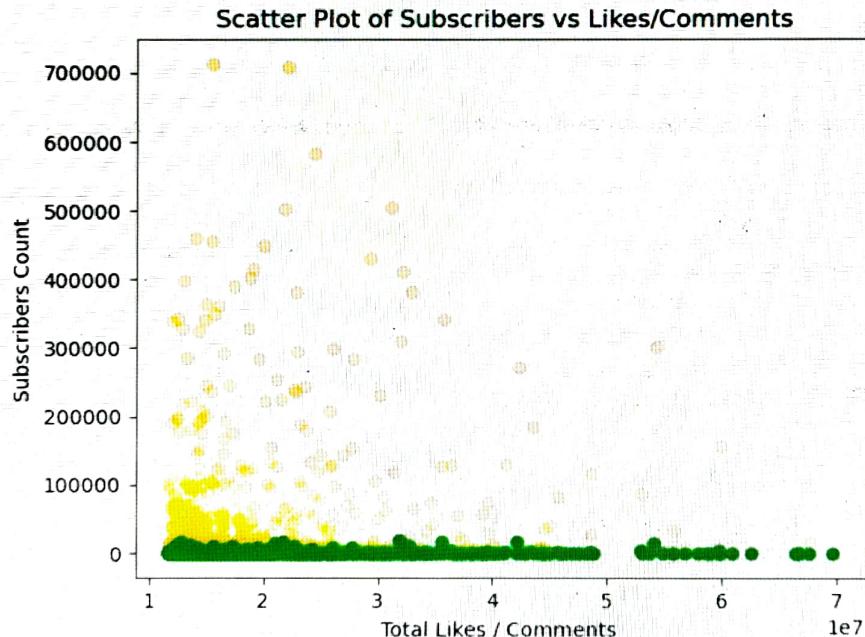
```
In [39]: import plotly.express as px
```

```
In [40]: fig = px.bar(data_frame = pop_df.head(11), x = 'Categories', y = 'Count', color_discrete_sequence = ['#ffb346'])  
fig.update_layout(  
    title = 'Most Popular Categories',  
    title_x = .5,  
    title_font = dict(color = 'black', size = 20, family = 'Cambria'))
```

```
In [40]: fig = px.bar(data_frame = pop_df.head(11), x = 'Categories', y = 'Count', color_discrete_sequence = ['#ffb346'])
fig.update_layout(
    title = 'Most Popular Categories',
    title_x = .5,
    title_font = dict(color = 'black', size = 20, family = 'Cambria'))
fig.show()
```



```
In [42]: plt.scatter(df['Subscribers'], df['Likes'], alpha = .4, label = 'Likes', color = 'orange')
plt.scatter(df['Subscribers'], df['Comments'], alpha = 1, label = 'Comments', color = 'green')
plt.xlabel('Total Likes / Comments')
plt.ylabel('Subscribers Count')
plt.title('Scatter Plot of Subscribers vs Likes/Comments')
plt.show()
```



From above plot it is clear that relation between subscribers count and likes/comments is arbitrary. But, lets consolidate this fact using correlation matrix as well.

```
In [43]: numeric_df = df[['Subscribers', 'Visits', 'Likes', 'Comments']]
```

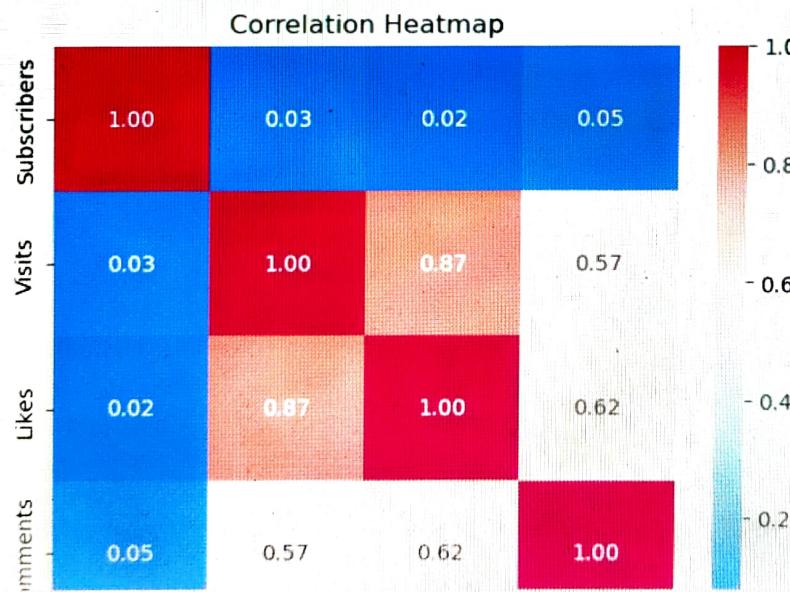


```
In [43]: numeric_df = df[['Subscribers','Visits','Likes', 'Comments']]
correlation_matrix = numeric_df.corr()
correlation_matrix
```

Out[43]:

	Subscribers	Visits	Likes	Comments
Subscribers	1.000000	0.030769	0.018870	0.047266
Visits	0.030769	1.000000	0.871883	0.568362
Likes	0.018870	0.871883	1.000000	0.615077
Comments	0.047266	0.568362	0.615077	1.000000

```
In [44]: sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
plt.title('Correlation Heatmap')
plt.show()
```



Search



```
In [45]: country_distribution = df.groupby('Country')['Visits'].sum().sort_values(ascending = False).reset_index()
```

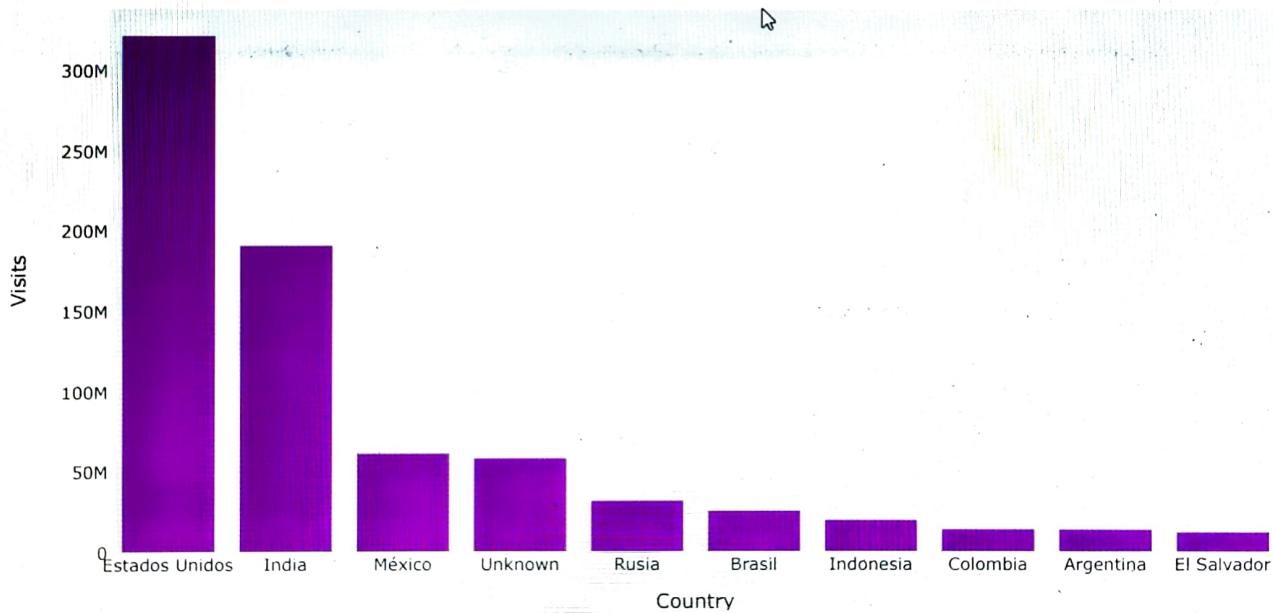
Out[45]:

	Country	Visits
0	Estados Unidos	322860860.0
1	India	191995055.5
2	México	61966100.0
3	Unknown	58949900.0
4	Rusia	32340300.0
5	Brasil	26058300.0
6	Indonesia	20344100.0
7	Colombia	14566800.0
8	Argentina	14474200.0
9	El Salvador	12900000.0
10	Filipinas	9323200.0
11	Reino Unido	9250700.0
12	Iraq	6326700.0
13	Francia	6059100.0
14	Tailandia	4688600.0
15	Arabia Saudita	4574500.0
16	Japón	4195000.0
17	Turquía	3904900.0
18	Perú	2219400.0
19	Pakistán	2179200.0
20	España	2039900.0
21	Somalia	1900000.0
22	Argelia	924100.0



```
In [46]: fig = px.bar(data_frame = country_distribution.head(10), x = 'Country', y = 'Visits', color_discrete_sequence = ['purple', 'blue', 'teal', 'orange', 'red', 'pink', 'yellow', 'green', 'brown', 'purple'],  
fig.update_layout(  
    title = 'Traffic From Different Countries',  
    title_x = .5,  
    title_font = dict(color = 'black', size = 20, family = 'Cambria'))  
fig.show()
```

Traffic From Different Countries



Search



```
In [47]: content_category_distribution = df.groupby(['Country', 'Categories'])['Visits'].sum().unstack().fillna(0)
content_category_distribution
```

Out[47]:

Categories	ASMR, Comida y bebida	Animación	Animación, Humor	Animación, Humor, Juguetes	Animación, Juguetes	Animación, Videojuegos	Animales y mascotas	Belleza, Moda	Ciencia y tecnología	... Películas, Animación	Per
Country											
<b>Arabia Saudita</b>	0.0	0.0	0.0	0.0	0.0	0.0	1200000.0	0.0	0.0	0.0	0.0
<b>Argelia</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>Argentina</b>	0.0	0.0	5700000.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	594800.0 56
<b>Bangladesh</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>Brasil</b>	0.0	0.0	8839000.0	5331000.0	0.0	0.0	1865800.0	0.0	0.0	404100.0	3859900.0 10
<b>Colombia</b>	0.0	0.0	0.0	4508300.0	0.0	0.0	16900.0	0.0	0.0	0.0	0.0
<b>Ecuador</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>Egipto</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>El Salvador</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>España</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	238400.0
<b>Estados Unidos</b>	368500.0	557500.0	2388700.0	23002600.0	0.0	28200.0	23545700.0	4462900.0	964500.0	10796500.0	4492500.0 124
<b>Filipinas</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0 48
<b>Francia</b>	0.0	0.0	0.0	0.0	8000.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>India</b>	0.0	0.0	0.0	0.0	0.0	0.0	839300.0	0.0	0.0	0.0	7519900.0 38
<b>Indonesia</b>	0.0	0.0	0.0	0.0	0.0	0.0	80900.0	0.0	0.0	56400.0	3097200.0 3
<b>Iraq</b>	0.0	0.0	0.0	0.0	0.0	0.0	3300000.0	0.0	0.0	0.0	0.0
<b>Japón</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>Jordania</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>Marruecos</b>	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>México</b>	0.0	0.0	1999500.0	10900000.0	0.0	0.0	2461500.0	0.0	0.0	1100000.0	8561400.0 5



Search

