

```
1 from google.colab import files
2 import pandas as pd
3
4 # Upload the CSV file
5 uploaded = files.upload()
```



Choose Files youtubers\_df.csv

- youtubers\_df.csv(text/csv) - 125368 bytes, last modified: 8/24/2024 - 100% done
- Saving youtubers\_df.csv to youtubers\_df.csv

```
1 df = pd.read_csv('youtubers_df.csv')
2 df
```



	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments	
0	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http://youtube.com/channel/UCq-Fj5J
1	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://youtube.com/channel/UCX6OQ3DI
2	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://youtube.com/channel/UCbCmjCu
3	4	SETIndia	NaN	162600000.0	India	15600.0	166.0	9.0	http://youtube.com/channel/UCpEhnqL0
4	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http://youtube.com/channel/UCk8GzjH
...	...	...	...	...	...	...	...	...	...
995	996	hamzymukbang	NaN	11700000.0	Estados Unidos	397400.0	14000.0	124.0	http://youtube.com/channel/UCPKNKl
996	997	Adaahqueen	NaN	11700000.0	India	1100000.0	92500.0	164.0	http://youtube.com/channel/UCk3fFpc
997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	http://youtube.com/channel/UCdrHrQf

Next steps:

[Generate code with df](#)[View recommended plots](#)[New interactive sheet](#)

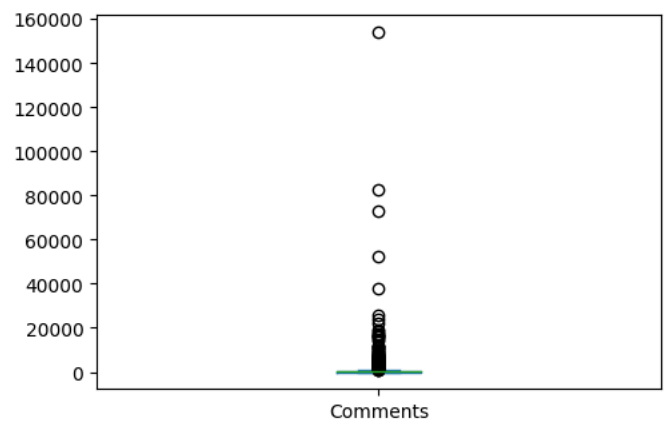
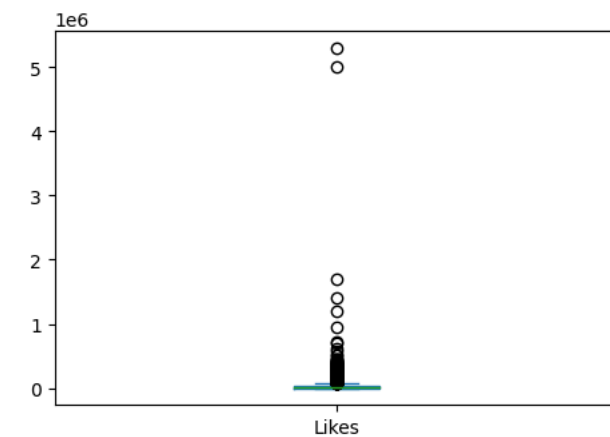
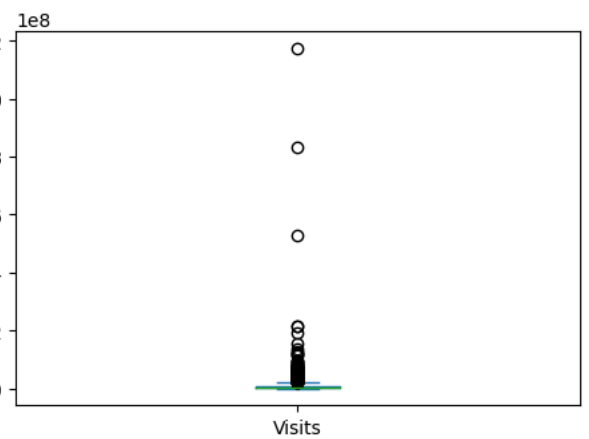
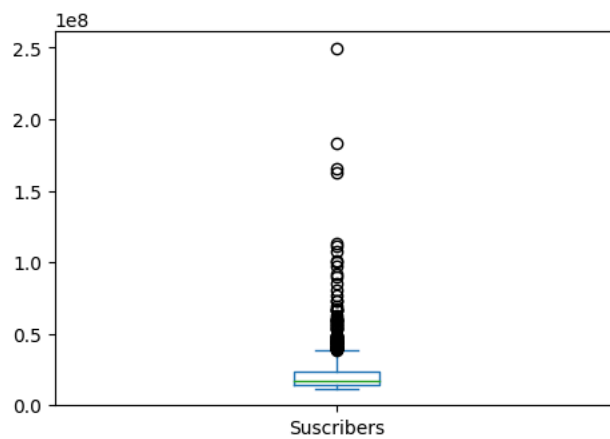
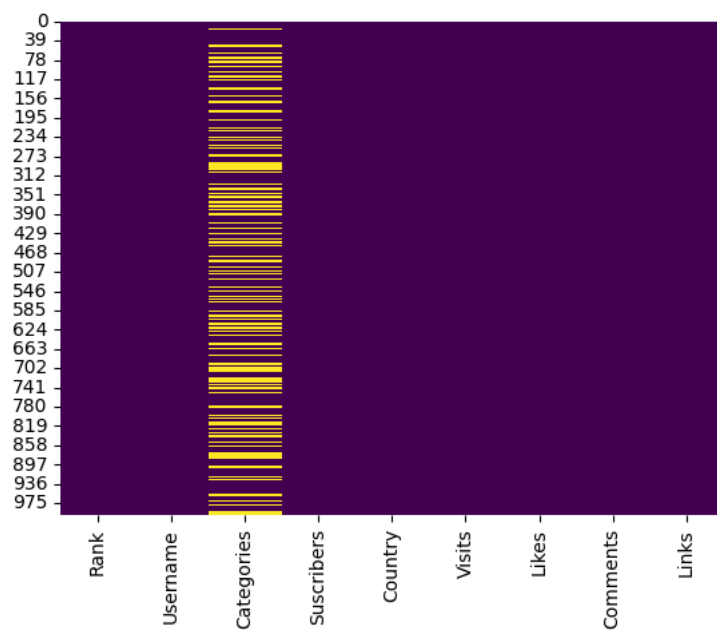
```
1 df.info()
```



```
<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   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
```

```
1 import matplotlib.pyplot as plt
2
3 # Check for missing data
4 print(df.isnull().sum())
5
6 # Visualize missing data
7 import seaborn as sns
8
9 sns.heatmap(df.isnull(), cbar=False, cmap='viridis')
10 plt.show()
11
12 # Check for outliers using boxplots
13 df[['Suscribers', 'Visits', 'Likes', 'Comments']].plot(kind='box', subplots=True, layout=(2, 2), figsize=(12, 8))
14 plt.show()
```

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



```

1 # Popular categories
2 popular_categories = df['Categories'].value_counts()
3 print(popular_categories)
4
5 # Correlation between subscribers, likes, and comments
6 correlation_matrix = df[['Subscribers', 'Likes', 'Comments']].corr()
7 print(correlation_matrix)
8
9 # Visualize correlation using a heatmap
10 sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
11 plt.show()

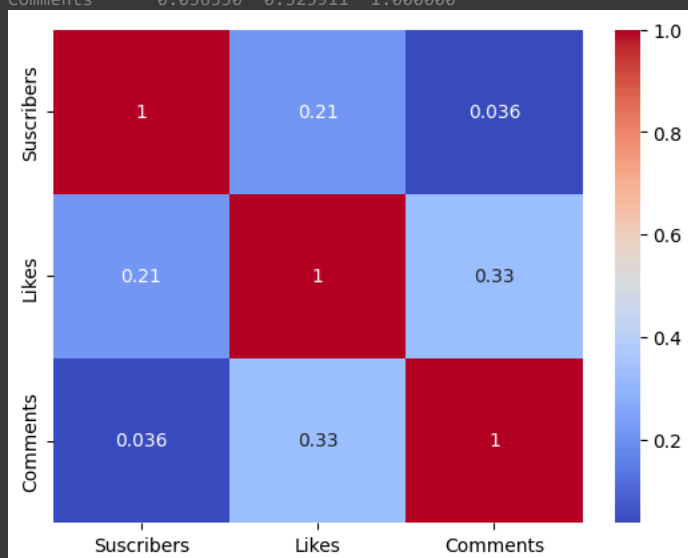
```

```

Categories
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
Ciencia y tecnología          14
Comida y bebida               12
Humor                         10
Juguetes                      10
Películas, Juguetes           9
Películas, Videojuegos        8
Deportes                      8
Música y baile, Humor         6
Juguetes, Coches y vehículos  4
DIY y Life Hacks              3
Fitness, Salud y autoayuda    3
Videojuegos, Juguetes         3
Animales y mascotas           2
Moda                          2
Coches y vehículos            2
Educación, Juguetes           2
Fitness                       2
Comida y bebida, Juguetes     1
ASMR, Comida y bebida         1
Animación, Humor, Juguetes    1
Diseño/arte, Belleza          1
Belleza, Moda                 1
ASMR                          1
Música y baile, Juguetes      1
Diseño/arte, DIY y Life Hacks 1
DIY y Life Hacks, Juguetes    1
Diseño/arte                   1
Comida y bebida, Salud y autoayuda 1
Viajes, Espectáculos          1
Juguetes, DIY y Life Hacks    1
Name: count, dtype: int64

```

	Subscribers	Likes	Comments
Subscribers	1.000000	0.211639	0.036350
Likes	0.211639	1.000000	0.325911
Comments	0.036350	0.325911	1.000000

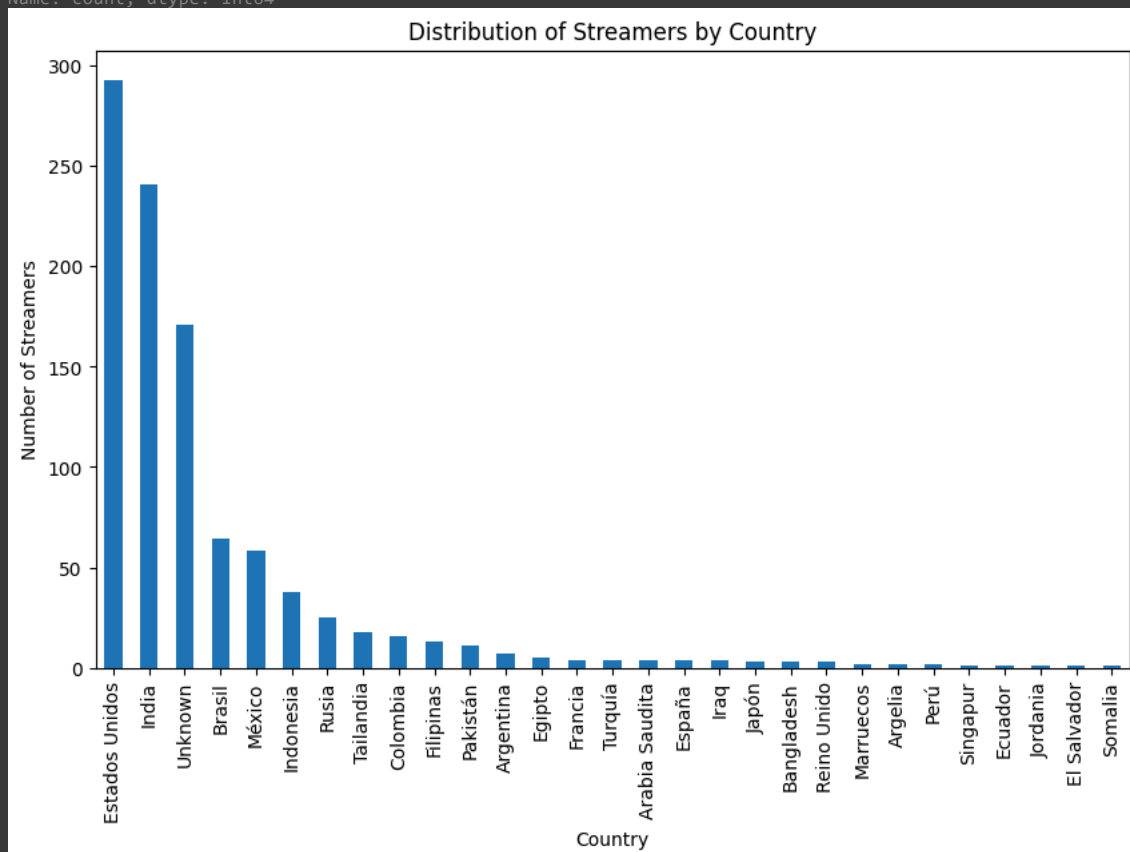


```
1 # Distribution by Country
2 country_distribution = df['Country'].value_counts()
3 print(country_distribution)
4
5 # Plot the distribution of streamers by country
6 country_distribution.plot(kind='bar', figsize=(10, 6))
7 plt.title('Distribution of Streamers by Country')
8 plt.xlabel('Country')
9 plt.ylabel('Number of Streamers')
10 plt.show()
11
12 # Country and Category Preferences
13 country_category_preferences = df.groupby(['Country', 'Categories']).size().unstack().fillna(0)
14 print(country_category_preferences)
15
16 # Visualize country and category preferences
17 country_category_preferences.plot(kind='bar', stacked=True, figsize=(12, 8))
18 plt.title('Country Preferences for Content Categories')
19 plt.xlabel('Country')
20 plt.ylabel('Number of Streamers')
21 plt.show()
```

```

Country
Estados Unidos    293
India              241
Unknown           171
Brasil            64
México            58
Indonesia         38
Rusia             25
Tailandia         18
Colombia          16
Filipinas         13
Pakistán          11
Argentina          7
Egipto            5
Francia           4
Turquía           4
Arabia Saudita    4
España            4
Iraq              4
Japón             3
Bangladesh        3
Reino Unido       3
Marruecos         2
Argelia           2
Perú              2
Singapur          1
Ecuador           1
Jordania           1
El Salvador        1
Somalia           1
Name: count, dtype: int64

```



Categories	ASMR	ASMR, Comida y bebida	Animación	Animación, Humor
Country				
Arabia Saudita	0.0	0.0	0.0	0.0
Argelia	0.0	0.0	0.0	0.0
Argentina	0.0	0.0	1.0	0.0
Bangladesh	0.0	0.0	0.0	0.0
Brasil	0.0	0.0	3.0	2.0
Colombia	0.0	0.0	0.0	2.0
Egipto	0.0	0.0	0.0	0.0
España	0.0	0.0	0.0	0.0
Estados Unidos	1.0	1.0	4.0	17.0
Filipinas	0.0	0.0	0.0	0.0
Francia	0.0	0.0	0.0	0.0
India	0.0	0.0	0.0	0.0
Indonesia	0.0	0.0	0.0	0.0
Iraq	0.0	0.0	0.0	0.0
Japón	0.0	0.0	0.0	0.0
Jordania	0.0	0.0	0.0	0.0
Marruecos	0.0	0.0	0.0	0.0
México	0.0	0.0	4.0	2.0
Pakistán	0.0	0.0	0.0	0.0
Perú	0.0	0.0	0.0	0.0

Reino Unido	0.0	0.0	0.0	0.0
Rusia	0.0	0.0	1.0	1.0
Singapur	0.0	0.0	1.0	0.0
Somalia	0.0	0.0	0.0	0.0
Tailandia	0.0	0.0	0.0	0.0
Turquía	0.0	0.0	0.0	0.0
Unknown	0.0	0.0	8.0	3.0

Categories	Animación, Humor, Juguetes	Animación, Juguetes	\
------------	----------------------------	---------------------	---

Country			
Arabia Saudita	0.0	0.0	
Argelia	0.0	0.0	
Argentina	0.0	0.0	
Bangladesh	0.0	0.0	
Brasil	0.0	0.0	
Colombia	0.0	0.0	
Egipto	0.0	0.0	
España	0.0	0.0	
Estados Unidos	0.0	1.0	
Filipinas	0.0	0.0	
Francia	1.0	0.0	
India	0.0	0.0	
Indonesia	0.0	0.0	
Iraq	0.0	0.0	
Japón	0.0	0.0	
Jordania	0.0	0.0	
Marruecos	0.0	0.0	
México	0.0	0.0	
Pakistán	0.0	0.0	
Perú	0.0	0.0	
Reino Unido	0.0	0.0	
Rusia	0.0	0.0	
Singapur	0.0	0.0	
Somalia	0.0	0.0	
Tailandia	0.0	0.0	
Turquía	0.0	0.0	
Unknown	0.0	28.0	

Categories	Animación, Videojuegos	Animales y mascotas	Belleza, Moda	\
------------	------------------------	---------------------	---------------	---

Country				
Arabia Saudita	1.0	0.0	0.0	
Argelia	0.0	0.0	0.0	
Argentina	0.0	0.0	0.0	
Bangladesh	0.0	0.0	0.0	
Brasil	2.0	0.0	0.0	
Colombia	1.0	0.0	0.0	
Egipto	0.0	0.0	0.0	
España	0.0	0.0	0.0	
Estados Unidos	19.0	2.0	1.0	
Filipinas	0.0	0.0	0.0	
Francia	0.0	0.0	0.0	
India	1.0	0.0	0.0	
Indonesia	1.0	0.0	0.0	
Iraq	1.0	0.0	0.0	
Japón	0.0	0.0	0.0	
Jordania	0.0	0.0	0.0	
Marruecos	0.0	0.0	0.0	
México	6.0	0.0	0.0	
Pakistán	0.0	0.0	0.0	
Perú	0.0	0.0	0.0	
Reino Unido	0.0	0.0	0.0	
Rusia	2.0	0.0	0.0	
Singapur	0.0	0.0	0.0	
Somalia	0.0	0.0	0.0	
Tailandia	0.0	0.0	0.0	
Turquía	0.0	0.0	0.0	
Unknown	0.0	0.0	0.0	

Categories	Ciencia y tecnología	...	Películas	Películas, Animación	\
------------	----------------------	-----	-----------	----------------------	---

Country	...			
Arabia Saudita	0.0	...	0.0	0.0
Argelia	0.0	...	0.0	0.0
Argentina	0.0	...	1.0	2.0
Bangladesh	0.0	...	0.0	0.0
Brasil	2.0	...	2.0	12.0
Colombia	0.0	...	0.0	0.0
Egipto	0.0	...	1.0	0.0
España	0.0	...	0.0	1.0
Estados Unidos	8.0	...	6.0	10.0
Filipinas	0.0	...	0.0	0.0
Francia	0.0	...	0.0	0.0
India	0.0	...	5.0	5.0
Indonesia	1.0	...	2.0	5.0
Iraq	0.0	...	0.0	0.0
Japón	0.0	...	0.0	0.0
Jordania	0.0	...	0.0	0.0
Marruecos	0.0	...	0.0	0.0
México	1.0	...	3.0	12.0
Pakistán	0.0	...	0.0	1.0
Perú	0.0	...	0.0	0.0
Reino Unido	0.0	...	0.0	1.0

Rusia	0.0	...	2.0	2.0
Singapur	0.0	...	0.0	0.0
Somalia	0.0	...	1.0	0.0
Tailandia	0.0	...	0.0	0.0
Turquía	0.0	...	0.0	0.0
Unknown	2.0	...	1.0	10.0

Categories	Películas, Humor	Películas, Juguetes	Películas, Videojuegos	\
Country				
Arabia Saudita	0.0	0.0	0.0	
Argelia	0.0	0.0	0.0	
Argentina	1.0	0.0	0.0	
Bangladesh	0.0	0.0	0.0	
Brasil	4.0	0.0	1.0	
Colombia	0.0	0.0	0.0	
Egipto	0.0	0.0	0.0	
España	0.0	0.0	1.0	
Estados Unidos	9.0	0.0	2.0	
Filipinas	2.0	0.0	0.0	
Francia	0.0	0.0	0.0	
India	8.0	0.0	1.0	
Indonesia	2.0	0.0	0.0	
Iraq	0.0	0.0	0.0	
Japón	0.0	0.0	0.0	
Jordania	0.0	0.0	0.0	
Marruecos	0.0	0.0	0.0	
México	1.0	0.0	1.0	
Pakistán	0.0	0.0	0.0	
Perú	0.0	0.0	0.0	
Reino Unido	0.0	0.0	0.0	
Rusia	0.0	0.0	1.0	
Singapur	0.0	0.0	0.0	
Somalia	0.0	0.0	0.0	
Tailandia	1.0	0.0	0.0	
Turquía	0.0	0.0	0.0	
Unknown	6.0	9.0	1.0	

Categories	Viajes, Espectáculos	Videojuegos	Videojuegos, Humor	\
Country				
Arabia Saudita	0.0	0.0	0.0	
Argelia	0.0	0.0	0.0	
Argentina	0.0	0.0	0.0	
Bangladesh	1.0	0.0	0.0	
Brasil	0.0	0.0	0.0	
Colombia	0.0	0.0	0.0	
Egipto	0.0	0.0	0.0	
España	0.0	1.0	0.0	
Estados Unidos	0.0	9.0	14.0	
Filipinas	0.0	0.0	0.0	
Francia	0.0	0.0	0.0	
India	0.0	0.0	1.0	
Indonesia	0.0	1.0	0.0	
Iraq	0.0	0.0	0.0	
Japón	0.0	0.0	0.0	
Jordania	0.0	0.0	0.0	
Marruecos	0.0	0.0	0.0	
México	0.0	0.0	0.0	
Pakistán	0.0	0.0	0.0	
Perú	0.0	0.0	0.0	
Reino Unido	0.0	0.0	1.0	
Rusia	0.0	8.0	1.0	
Singapur	0.0	0.0	0.0	
Somalia	0.0	0.0	0.0	
Tailandia	0.0	0.0	0.0	
Turquía	0.0	0.0	0.0	
Unknown	0.0	0.0	0.0	

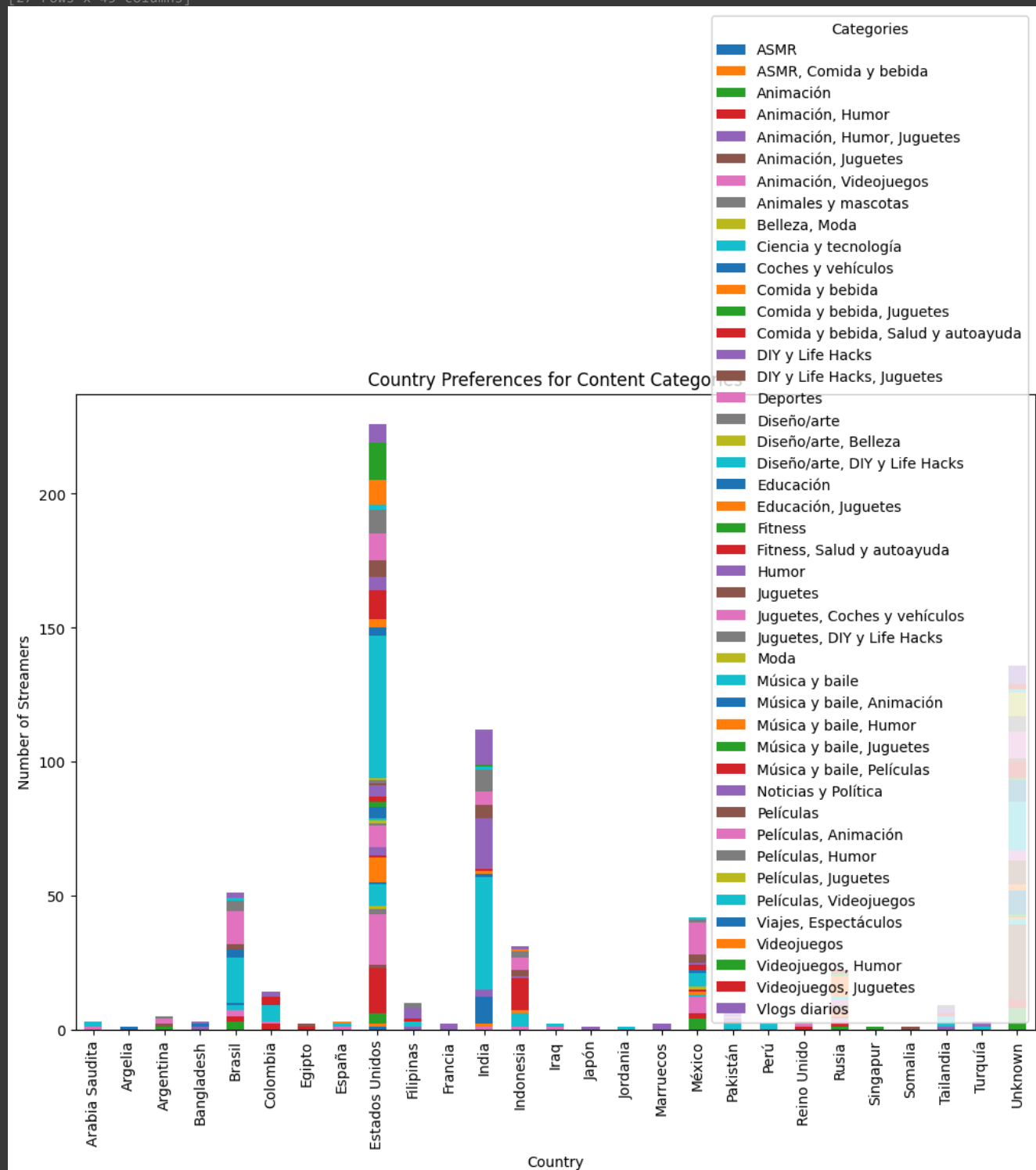
Categories	Videojuegos, Juguetes	Vlogs diarios
Country		
Arabia Saudita	0.0	0.0
Argelia	0.0	0.0
Argentina	0.0	0.0
Bangladesh	0.0	1.0
Brasil	0.0	2.0
Colombia	0.0	2.0
Egipto	0.0	0.0
España	0.0	0.0
Estados Unidos	0.0	7.0
Filipinas	0.0	0.0
Francia	0.0	1.0
India	0.0	13.0
Indonesia	0.0	1.0
Iraq	0.0	0.0
Japón	0.0	0.0
Jordania	0.0	0.0
Marruecos	0.0	0.0
México	0.0	0.0
Pakistán	0.0	1.0
Perú	0.0	0.0
Reino Unido	0.0	0.0

```

Rusia          1.0      0.0
Singapur       0.0      0.0
Somalia        0.0      0.0
Tailandia       0.0      0.0
Turquía        0.0      2.0
Unknown        2.0      7.0

```

[27 rows x 45 columns]



```

1 # Calculate average metrics
2 avg_metrics = df[['Suscribers', 'Visits', 'Likes', 'Comments']].mean()
3 print(avg_metrics)
4
5 # Visualize average metrics
6 avg_metrics.plot(kind='bar', figsize=(8, 5))
7 plt.title('Average Performance Metrics')
8 plt.ylabel('Average Count')
9 plt.show()
10
11 # Check for patterns or anomalies
12 df[['Suscribers', 'Visits', 'Likes', 'Comments']].plot(kind='box', subplots=True, layout=(2, 2), figsize=(12, 8))
13 plt.show()

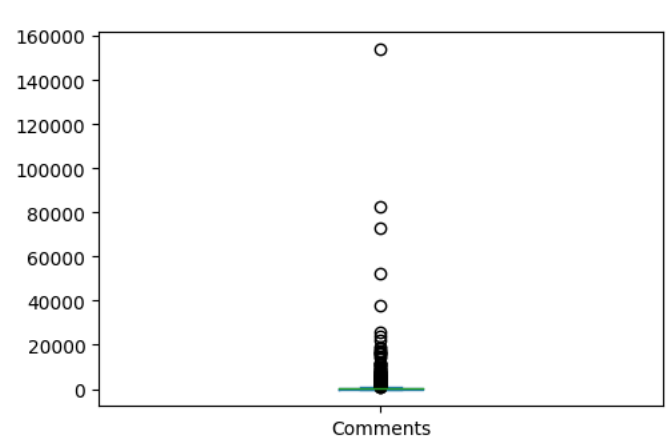
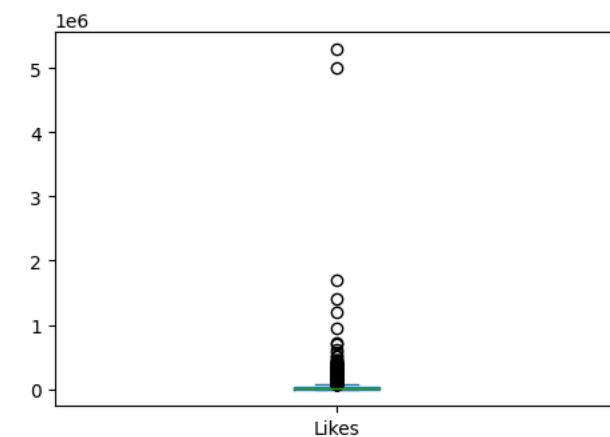
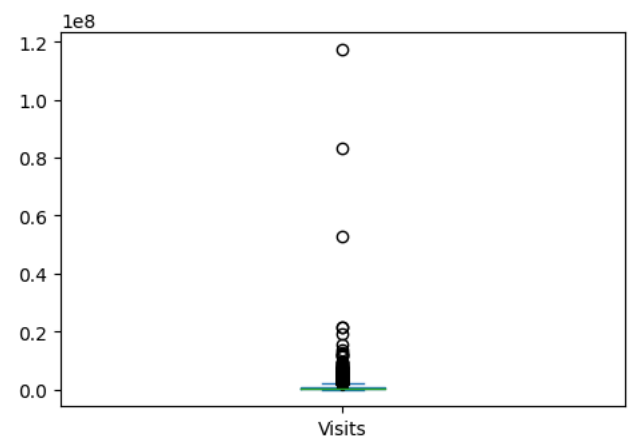
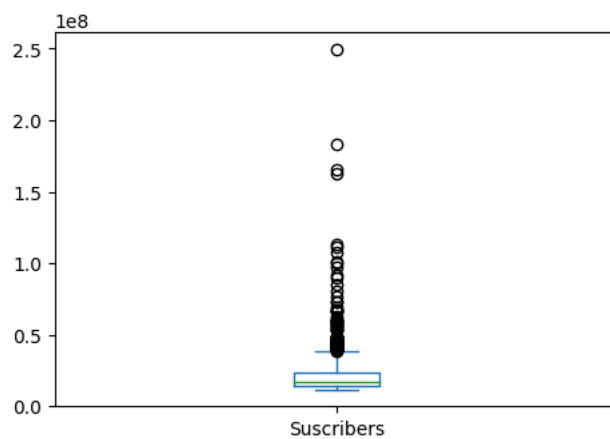
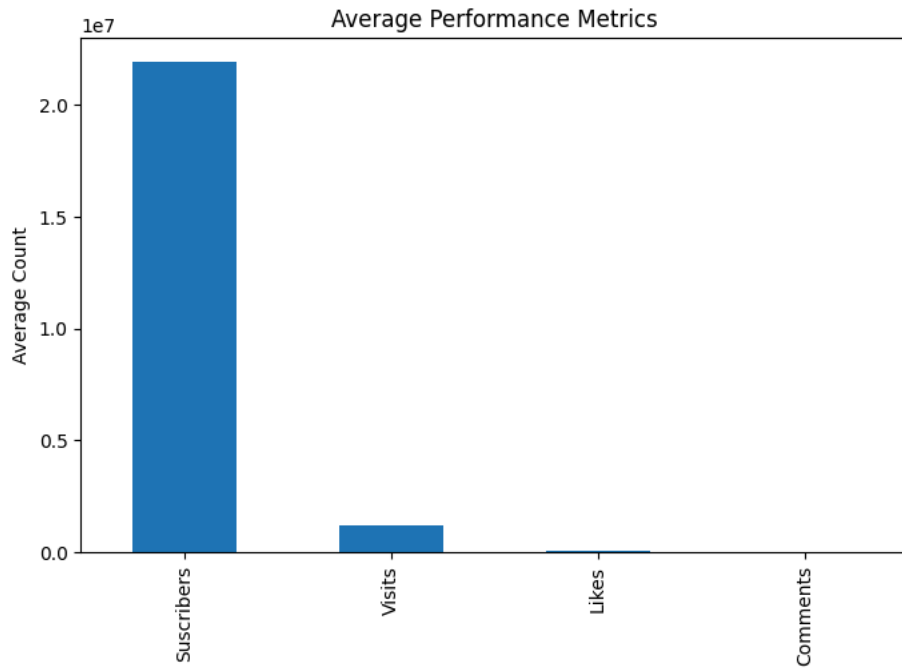
```



```

Suscribers    2.189440e+07
Visits        1.209446e+06
Likes         5.363259e+04
Comments      1.288768e+03
dtype: float64

```



```

1 # Distribution of Categories
2 category_distribution = df['Categories'].value_counts()
3 print(category_distribution)
4
5 # Visualize category distribution
6 category_distribution.plot(kind='bar', figsize=(10, 6))
7 plt.title('Distribution of Content Categories')
8 plt.xlabel('Categories')
9 plt.ylabel('Number of Streamers')
10 plt.show()

```

```
Categories
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
Ciencia y tecnología 14
Comida y bebida 12
Humor 10
Juguetes 10
Películas, Juguetes 9
Películas, Videojuegos 8
Deportes 8
Música y baile, Humor 6
Juguetes, Coches y vehículos 4
DIY y Life Hacks 3
Fitness, Salud y autoayuda 3
Videojuegos, Juguetes 3
Animales y mascotas 2
Moda 2
Coches y vehículos 2
Educación, Juguetes 2
Fitness 2
Comida y bebida, Juguetes 1
ASMR, Comida y bebida 1
Animación, Humor, Juguetes 1
Diseño/arte, Belleza 1
Belleza, Moda 1
ASMR 1
Música y baile, Juguetes 1
Diseño/arte, DIY y Life Hacks 1
DIY y Life Hacks, Juguetes 1
Diseño/arte 1
Comida y bebida, Salud y autoayuda 1
Viajes, Espectáculos 1
Juguetes, DIY y Life Hacks 1
Name: count, dtype: int64
```

Distribution of Content Categories

