

```
In [19]: import pandas as pd
import numpy as np
```

```
In [25]: df=pd.read_csv("C:/Users/manju/Desktop/youtubers_df.csv")
```

```
In [26]: df.isnull()
```

```
Out[26]:
```

	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments	Links
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	True	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...
995	False	False	True	False	False	False	False	False	False
996	False	False	True	False	False	False	False	False	False
997	False	False	False	False	False	False	False	False	False
998	False	False	True	False	False	False	False	False	False
999	False	False	False	False	False	False	False	False	False

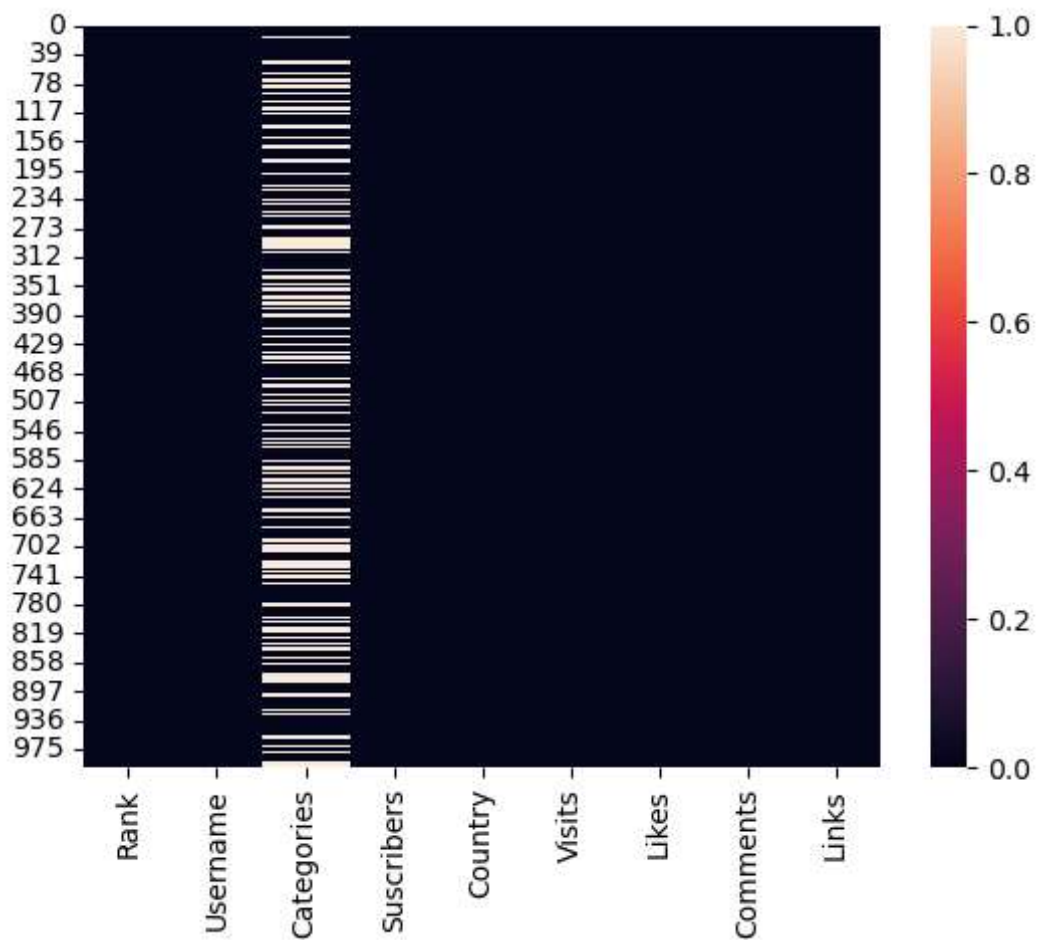
1000 rows × 9 columns

```
In [27]: df.isnull().sum().sum()
```

```
Out[27]: 306
```

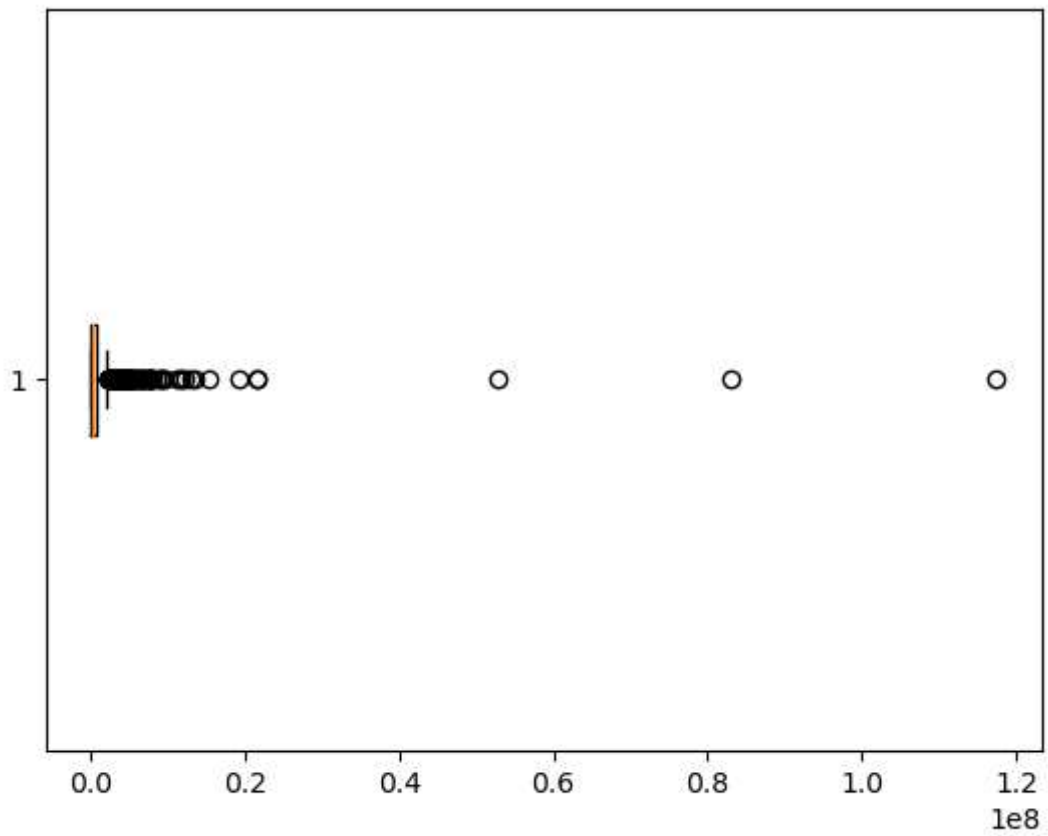
```
In [28]: import seaborn as sns
sns.heatmap(df.isnull())
```

```
Out[28]: <Axes: >
```



```
In [29]: import matplotlib.pyplot as plt
cc=['Visits']

plt.boxplot(df[cc],vert=False)
plt.show()
```



```
In [30]: Q1= df[cc].quantile(0.25)
Q3= df[cc].quantile(0.75)
IQR=Q3 - Q1

l_b= Q1 - 1.5*IQR
u_b= Q3 + 1.5*IQR
```

```
In [32]: outliers=df[(df[cc]<l_b) | (df[cc] > u_b)]
print("Outliers:")
print(outliers)
```

Outliers:

	Rank	Username	Categories	Suscribers	Country	Visits	Likes	\
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	NaN	117400000.0	NaN	
2	NaN	NaN	NaN	NaN	NaN	7000000.0	NaN	
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	3900000.0	NaN	
..	
995	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
996	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
997	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
998	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
999	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

	Comments	Links
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	NaN
..
995	NaN	NaN
996	NaN	NaN
997	NaN	NaN
998	NaN	NaN
999	NaN	NaN

[1000 rows x 9 columns]

In []:

In []:

In []:

In []:

In []:

In []: