Fhatuwani Makharamedzha.

Aspiring Data Analyst/Software Developer.

YOUTUBERS_ANALYSIS(intern career).

Importing Libraries.

In [18]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

Data Importing and exploration.

In [19]:
```
df = pd.read_csv('youtubers_df.csv')
```

In [20]:
```
df
```

Out[20]:

	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments	
0	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http:
1	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://yo
2	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://y
3	4	SETIndia	NaN	162600000.0	India	15600.0	166.0	9.0	http://yo
4	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http:/
...	...	...	...	...	...	...	...	...	
995	996	hamzymukbang	NaN	11700000.0	Estados Unidos	397400.0	14000.0	124.0	http://
996	997	Adaahqueen	NaN	11700000.0	India	1100000.0	92500.0	164.0	http://y
997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	http://y
998	999	PenMultiplex	NaN	11700000.0	India	14000.0	81.0	1.0	http://y
999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	http://y

1000 rows × 9 columns

In [21]:
```
df.describe()
```

Out[21]:

	Rank	Suscribers	Visits	Likes	Comments
count	1000.000000	1.000000e+03	1.000000e+03	1.000000e+03	1000.000000
mean	500.500000	2.189440e+07	1.209446e+06	5.363259e+04	1288.768000
std	288.819436	1.682775e+07	5.229942e+06	2.580457e+05	6778.188308
min	1.000000	1.170000e+07	0.000000e+00	0.000000e+00	0.000000
25%	250.750000	1.380000e+07	3.197500e+04	4.717500e+02	2.000000
50%	500.500000	1.675000e+07	1.744500e+05	3.500000e+03	67.000000
75%	750.250000	2.370000e+07	8.654750e+05	2.865000e+04	472.000000
max	1000.000000	2.495000e+08	1.174000e+08	5.300000e+06	154000.000000

In [22]:
```
df.head()
```

Out[22]:

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

In [23]:
```
df.tail()
```

Out[23]:

	Rank	Username	Categories	Suscribers	Country	Visits	Likes	Comments	
995	996	hamzymukbang	NaN	11700000.0	Estados Unidos	397400.0	14000.0	124.0	http://youtube.c
996	997	Adaahqueen	NaN	11700000.0	India	1100000.0	92500.0	164.0	http://youtube.c
997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	http://youtube.c
998	999	PenMultiplex	NaN	11700000.0	India	14000.0	81.0	1.0	http://youtube.cc
999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	http://youtube.cc

In [24]:
```
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
```

In [25]:
```
df.shape
```

Out[25]:
```
(1000, 9)
```

In [30]:
```
df.size
```

Out[30]:
```
9000
```

In [31]:
```
df.columns
```

Out[31]:
```
Index(['Rank', 'Username', 'Categories', 'Suscribers', 'Country', 'Visits',
       'Likes', 'Comments', 'Links'],
      dtype='object')
```

In [32]:
```
#FROM THE DATA WE NOTICE SOME COLUMNS[Suscribers] ARE NOT SPELLED WELL.
df.rename(columns = {'Suscribers': 'Subscribers'}, inplace=True)
```

In [33]:
```
df
```

Out[33]:

	Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	
0	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http:
1	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://yo
2	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://y
3	4	SETIndia	NaN	162600000.0	India	15600.0	166.0	9.0	http://yo
4	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http:/
...	...	...	...	...	...	...	...	...	
995	996	hamzymukbang	NaN	11700000.0	Estados Unidos	397400.0	14000.0	124.0	http://
996	997	Adaahqueen	NaN	11700000.0	India	1100000.0	92500.0	164.0	http://y
997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	http://y
998	999	PenMultiplex	NaN	11700000.0	India	14000.0	81.0	1.0	http://y
999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	http://y

1000 rows × 9 columns

In [34]:
```
df.isnull().sum()
```

Out[34]:
```
Rank            0
Username        0
Categories    306
Subscribers     0
Country         0
Visits          0
Likes           0
Comments        0
Links           0
dtype: int64
```

In [38]:
```
df.duplicated().sum()
```

Out[38]:
```
0
```

In [41]:
```
df.dropna(inplace=True)
```

In [42]:
```
df
```

Out[42]:

	Rank	Username	Categories	Subscribers	Country	Visits	Likes	Comments	
0	1	tseries	Música y baile	249500000.0	India	86200.0	2700.0	78.0	http:
1	2	MrBeast	Videojuegos, Humor	183500000.0	Estados Unidos	117400000.0	5300000.0	18500.0	http://yo
2	3	CoComelon	Educación	165500000.0	Unknown	7000000.0	24700.0	0.0	http://y
4	5	KidsDianaShow	Animación, Juguetes	113500000.0	Unknown	3900000.0	12400.0	0.0	http:/
5	6	PewDiePie	Películas, Videojuegos	111500000.0	Estados Unidos	2400000.0	197300.0	4900.0	http://y
...	...	...	...	...	...	...	...	...	
989	990	cut	Humor	11700000.0	Estados Unidos	359000.0	8800.0	342.0	http://yc
990	991	JoeHattab	Películas	11700000.0	Somalia	1900000.0	98500.0	2900.0	http:
991	992	BeAmazed	Educación	11700000.0	Estados Unidos	477800.0	9900.0	556.0	http://yc
997	998	LittleAngelIndonesia	Música y baile	11700000.0	Unknown	211400.0	745.0	0.0	http://y
999	1000	OneindiaHindi	Noticias y Política	11700000.0	India	2200.0	31.0	1.0	http://y

694 rows × 9 columns

In [43]:
```
df.isnull().sum()
```

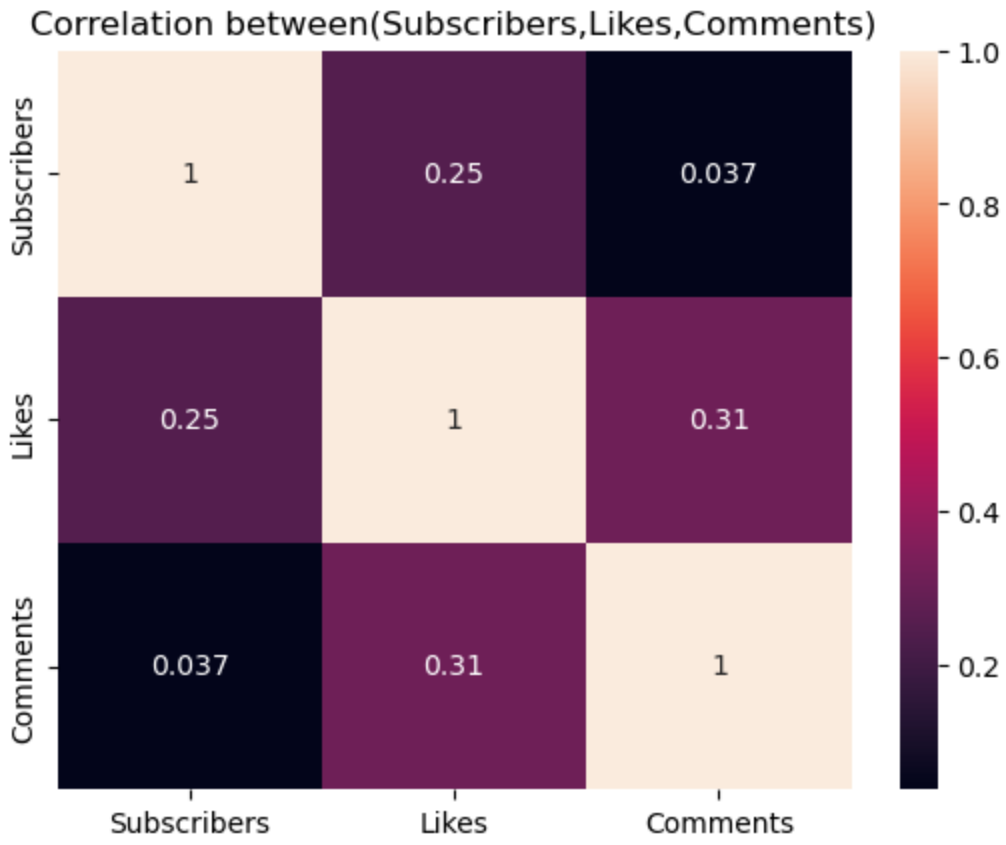
Out[43]:
```
Rank          0
Username      0
Categories    0
Subscribers   0
Country       0
Visits        0
Likes         0
Comments      0
Links         0
dtype: int64
```

Trend Analysis and plotting.

In [53]:
```
#This is the most popular trend in the "Categories" column
df['Categories'].max()
```

Out[53]:
```
'Vlogs diarios'
```

In [112…
```
#This is the "correlation" between number of Subscribers,Likes and Commnets
correlation = df[['Subscribers','Likes','Comments']].corr()
sns.heatmap(correlation, annot=True)
plt.title('Correlation between(Subscribers,Likes,Comments)')
plt.show()
```

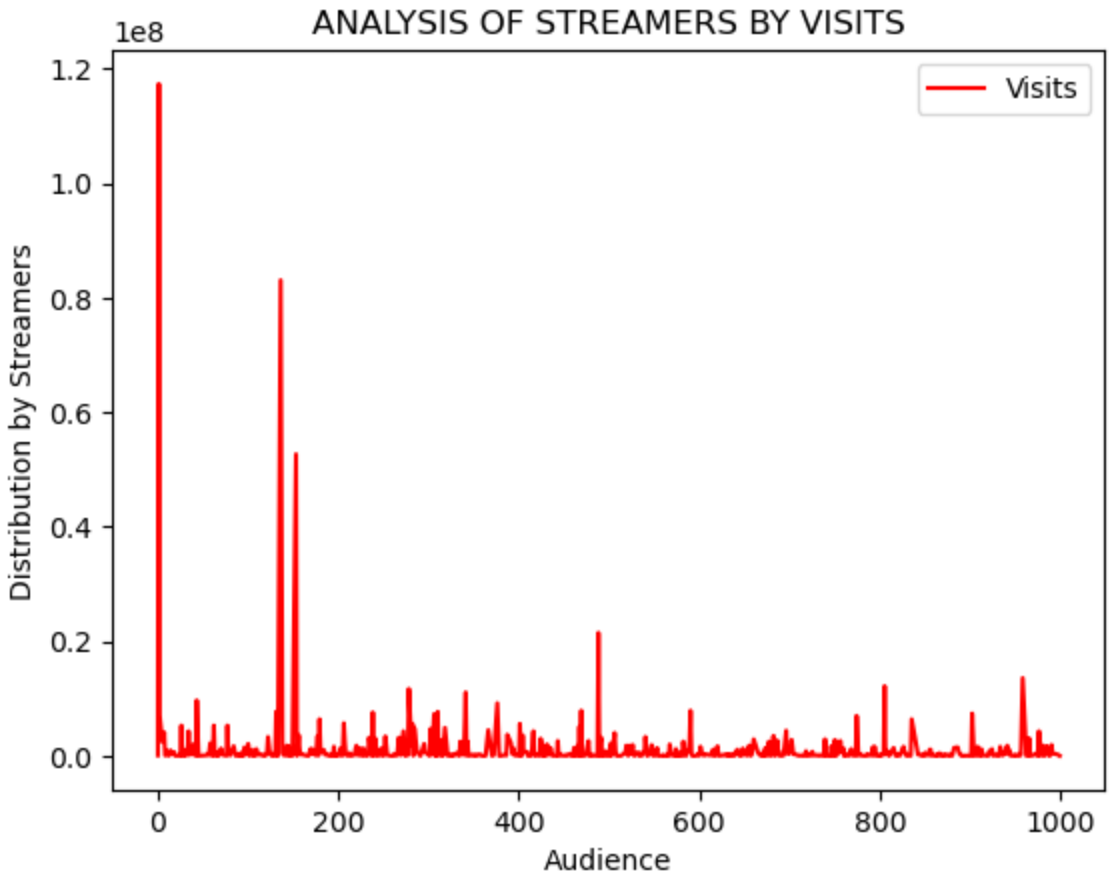


Audience Study and Analysis.

In [127…
```
#This is the distribution of streamers"Visits" in graph form
df['Country'].max()
```

Out[127]:
```
'Unknown'
```

In [133…
```
#This graph is the analysis by visits"streamers
Streamers = df[['Visits']].plot(kind='line', color='red')
plt.title('ANALYSIS OF STREAMERS BY VISITS')
plt.xlabel('Audience')
plt.ylabel('Distribution by Streamers')
plt.show()
```



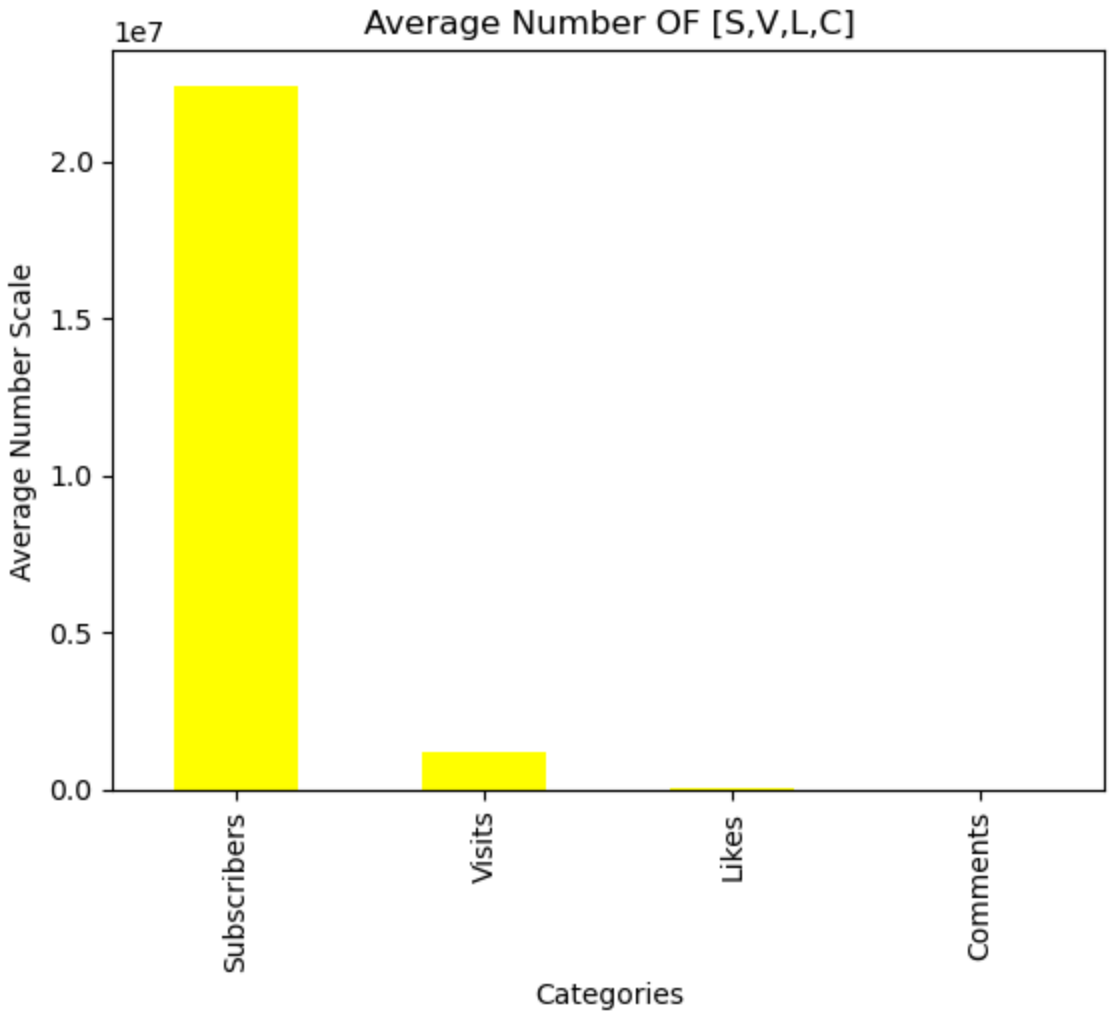
Perfomance Metrics.

In [190…
```
#Calculating and visualising the average number of subscribers, visits, likes, and comme
AverageNumber = df[['Subscribers','Visits','Likes','Comments']].mean()
```

In [191…
```
AverageNumber.sum()
```

Out[191]:
```
23681324.03458213
```

In [193…
```
#Visualisation of the Average Number of subscribers, visits, likes, and comments.
AverageNumber.plot(kind ='bar', color='yellow')
plt.ylabel('Average Number Scale')
plt.xlabel('Categories')
plt.title('Average Number OF [S,V,L,C]')
plt.show()
```



Content Categories.

In [171…

```
#Distribution of Content Categories[With the HIGHEST NUMBER OF STREAMERS].
df['Categories'].value_counts()
```

Out[171]:
```
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
```

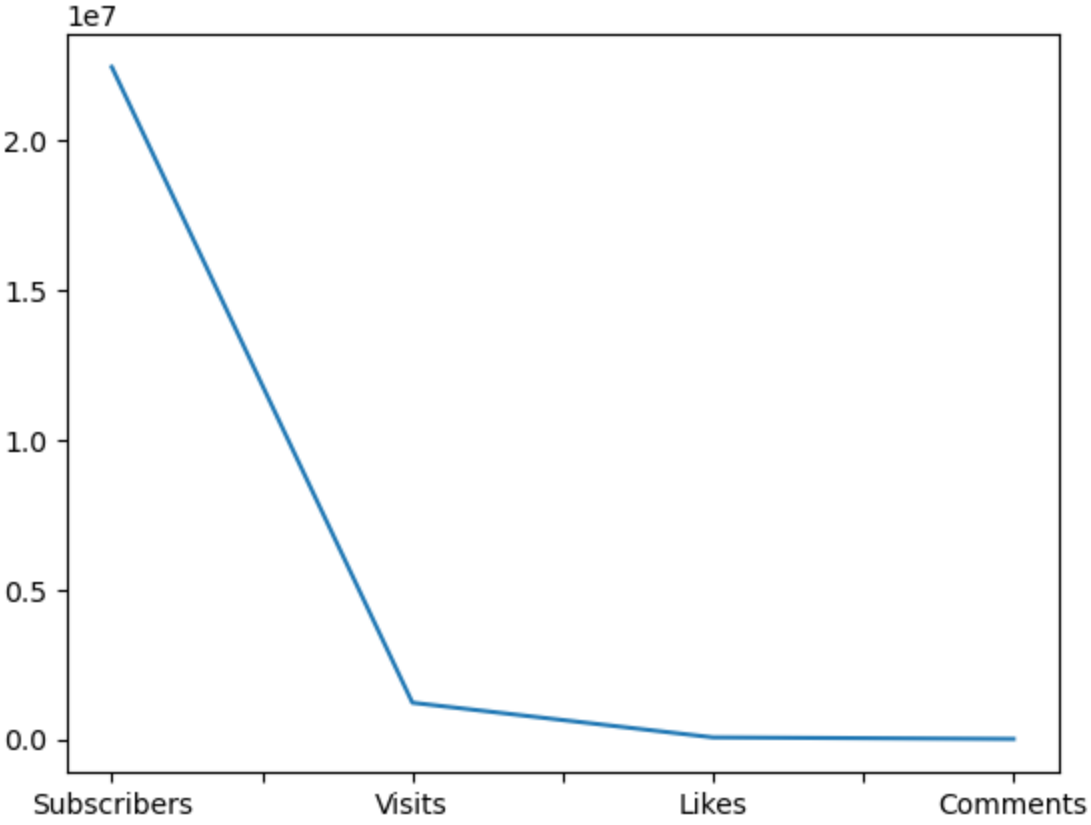
BenchMarking.

In [215…
```
#- Identify streamers with above-average performance in terms of subscribers, visits, li
AverageNumber >  df[['Subscribers','Visits','Likes','Comments']].mean()
```

Out[215]:
```
Subscribers    False
Visits         False
Likes          False
Comments       False
dtype: bool
```

In [216…
```
AverageNumber.plot()
```

Out[216]:
```
<Axes: >
```



:)

In [221…

```
df.to_csv('youtubers_df.csv')
```