instagram Analysis Using Python

The instagram accounts dataset has the information about the Followers, Profession & Country.

The Data set available from Flexible which is a Third Party instagram accounts which engine, and available on Kaggle dataset for free.

Import Library

In [1]:
```
import pandas as pd
```

In [2]:
```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import seaborn as sns
```

```
C:\Users\Syed Arif\anaconda3\lib\site-packages\scipy\__init__.py:146: UserWar
ning: A NumPy version >=1.16.5 and <1.23.0 is required for this version of Sc
iPy (detected version 1.25.1
  warnings.warn(f"A NumPy version >={np_minversion} and <{np_maxversion}"
```

Uploading Csv fle

In [3]:
```
df = pd.read_csv(r"C:\Users\Syed Arif\Desktop\instagram.csv")
```

Data Preprocessing

.head()

head is used show to the By default = 5 rows in the dataset

In [4]:
```
df.head()
```

Out[4]:

	Rank	Username	Owner	Followers(millions)	Profession/Activity	Country
0	1	@instagram	Instagram	645.0	Social media platform	United States
1	2	@cristiano	Cristiano Ronaldo	594.0	Footballer	Portugal
2	3	@leomessi	Lionel Messi	476.0	Footballer	Argentina
3	4	@selenagomez	Selena Gomez	423.0	Musician, actress, and businesswoman	United States
4	5	@kyliejenner	Kylie Jenner	395.0	Television personality and businesswoman	United States

.tail()

tail is used to show rows by Descending order

In [5]:
```
df.tail()
```

Out[5]:

	Rank	Username	Owner	Followers(millions)	Profession/Activity	Country
45	46	@snoopdogg	Snoop Dogg	80.0	Musician	United States
46	47	@davidbeckham	David Beckham	79.3	Former footballer, president of MLS club Inter...	United Kingdom
47	48	@jennierubyjane	Jennie	79.1	Musician	South Korea
48	49	@khaby00	Khaby Lame	79.1	Social media personality	Italy Senegal
49	50	@gigihadid	Gigi Hadid	78.5	Model	United States

.shape

It show the total no of rows & Column in the dataset

In [6]:
```
df.shape
```

Out[6]:
```
(50, 6)
```

.Columns

It show the no of each Column

In [7]:
```
df.columns
```

Out[7]:
```
Index(['Rank', 'Username', 'Owner', 'Followers(millions)',
       'Profession/Activity', 'Country'],
      dtype='object')
```

.dtypes

This Attribute show the data type of each column

In [8]:
```
df.dtypes
```

Out[8]:
```
Rank                   int64
Username              object
Owner                 object
Followers(millions)  float64
Profession/Activity    object
Country                object
dtype: object
```

.unique()

In a column, It show the unique value of specific column.

In [9]:
```
df["Country"].unique()
```

Out[9]:
```
array(['United States', 'Portugal', 'Argentina', 'Canada', 'India',
       'Trinidad and Tobago\xa0United States', 'Brazil', 'Barbados',
       'Spain', 'Europe', 'France', 'Israel', 'Thailand',
       'United Kingdom\xa0Albania', 'Colombia', 'United States\xa0Canada',
       'United Kingdom', 'South Korea', 'Italy\xa0Senegal'], dtype=object)
```

.nuique()

It will show the total no of unque value from whole data frame

In [10]:
```
df.nunique()
```

Out[10]:
```
Rank                   50
Username               50
Owner                  50
Followers(millions)    47
Profession/Activity    30
Country                19
dtype: int64
```

.describe()

It show the Count, mean , median etc

In [11]:
```
df.describe()
```

Out[11]:

	Rank	Followers(millions)
count	50.00000	50.000000
mean	25.50000	208.976000
std	14.57738	136.262332
min	1.00000	78.500000
25%	13.25000	96.850000
50%	25.50000	161.000000
75%	37.75000	288.000000
max	50.00000	645.000000

.value_counts

It Shows all the unique values with their count

In [12]:
```
df["Country"].value_counts()
```

Out[12]:
```
United States                          28
India                                   3
Canada                                  2
Spain                                   2
Israel                                  1
South Korea                             1
United Kingdom                          1
United States Canada                    1
Colombia                                1
United Kingdom Albania                  1
Thailand                                1
Europe                                  1
France                                  1
Portugal                                1
Barbados                                1
Brazil                                  1
Trinidad and Tobago United States       1
Argentina                               1
Italy Senegal                           1
Name: Country, dtype: int64
```

.isnull()

It shows the how many null values

In [13]:
```
df.isnull()
```

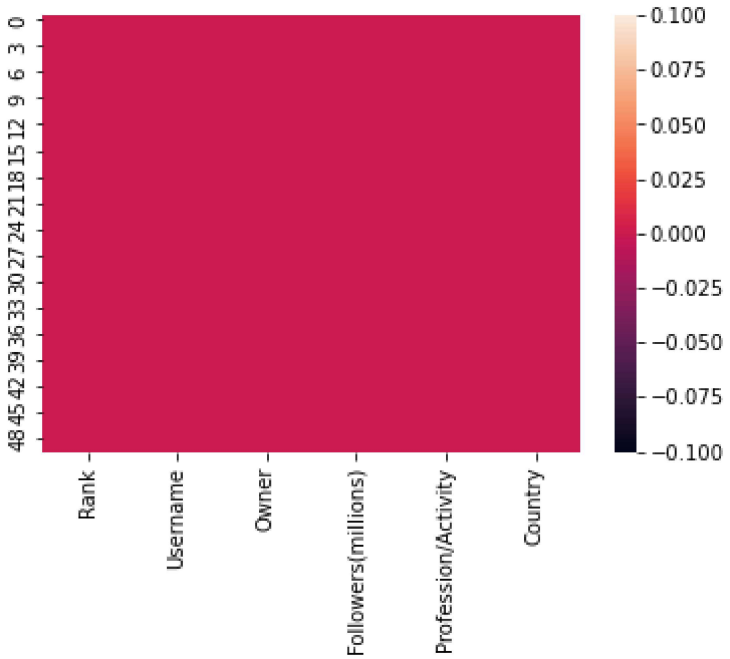
Out[13]:

	Rank	Username	Owner	Followers(millions)	Profession/Activity	Country
0	False	False	False	False	False	False
1	False	False	False	False	False	False
2	False	False	False	False	False	False
3	False	False	False	False	False	False
4	False	False	False	False	False	False
5	False	False	False	False	False	False
6	False	False	False	False	False	False
7	False	False	False	False	False	False
8	False	False	False	False	False	False
9	False	False	False	False	False	False
10	False	False	False	False	False	False
11	False	False	False	False	False	False
12	False	False	False	False	False	False
13	False	False	False	False	False	False
14	False	False	False	False	False	False
15	False	False	False	False	False	False
16	False	False	False	False	False	False
17	False	False	False	False	False	False
18	False	False	False	False	False	False
19	False	False	False	False	False	False
20	False	False	False	False	False	False
21	False	False	False	False	False	False
22	False	False	False	False	False	False
23	False	False	False	False	False	False
24	False	False	False	False	False	False
25	False	False	False	False	False	False
26	False	False	False	False	False	False
27	False	False	False	False	False	False
28	False	False	False	False	False	False
29	False	False	False	False	False	False
30	False	False	False	False	False	False
31	False	False	False	False	False	False
32	False	False	False	False	False	False
33	False	False	False	False	False	False
34	False	False	False	False	False	False

	Rank	Username	Owner	Followers(millions)	Profession/Activity	Country
35	False	False	False	False	False	False
36	False	False	False	False	False	False
37	False	False	False	False	False	False
38	False	False	False	False	False	False
39	False	False	False	False	False	False
40	False	False	False	False	False	False
41	False	False	False	False	False	False
42	False	False	False	False	False	False
43	False	False	False	False	False	False
44	False	False	False	False	False	False
45	False	False	False	False	False	False
46	False	False	False	False	False	False
47	False	False	False	False	False	False
48	False	False	False	False	False	False
49	False	False	False	False	False	False

In [14]:
```
sns.heatmap(df.isnull())
```

Out[14]: <AxesSubplot:>



Top 10 Performers with the Highest Number of Followers

In [15]:
```
top_10_followers= df.sort_values(by = "Followers(millions)" , ascending = Fals
top_10_followers
```

Out[15]:

	Rank	Username	Owner	Followers(millions)	Profession/Activity	Country
0	1	@instagram	Instagram	645.0	Social media platform	United States
1	2	@cristiano	Cristiano Ronaldo	594.0	Footballer	Portugal
2	3	@leomessi	Lionel Messi	476.0	Footballer	Argentina
3	4	@selenagomez	Selena Gomez	423.0	Musician, actress, and businesswoman	United States
4	5	@kyliejenner	Kylie Jenner	395.0	Television personality and businesswoman	United States
5	6	@therock	Dwayne Johnson	385.0	Actor and professional wrestler	United States
6	7	@arianagrande	Ariana Grande	375.0	Musician, actress and businesswoman	United States
7	8	@kimkardashian	Kim Kardashian	360.0	Television personality, model and businesswoman	United States
8	9	@beyonce	Beyoncé	312.0	Musician and businesswoman	United States
9	10	@khloekardashian	Khloé Kardashian	309.0	Television personality and model	United States

Correlation between Followers and Rank

In [16]:
```
correlation = df['Followers(millions)'].corr(df['Rank'])
correlation
```

Out[16]:
```
-0.9121488888123581
```

Average Number of Followers

In [17]:
```
average_followers = df['Followers(millions)'].mean()
average_followers
```

Out[17]:
```
208.976
```

Countries with the Most Performers

In [18]:
```
country_counts = df['Country'].value_counts()
country_counts
```

Out[18]:
```
United States                          28
India                                   3
Canada                                  2
Spain                                   2
Israel                                  1
South Korea                             1
United Kingdom                          1
United States Canada                    1
Colombia                                1
United Kingdom Albania                  1
Thailand                                1
Europe                                  1
France                                  1
Portugal                                1
Barbados                                1
Brazil                                  1
Trinidad and Tobago United States       1
Argentina                               1
Italy Senegal                           1
Name: Country, dtype: int64
```

Most Common Professions

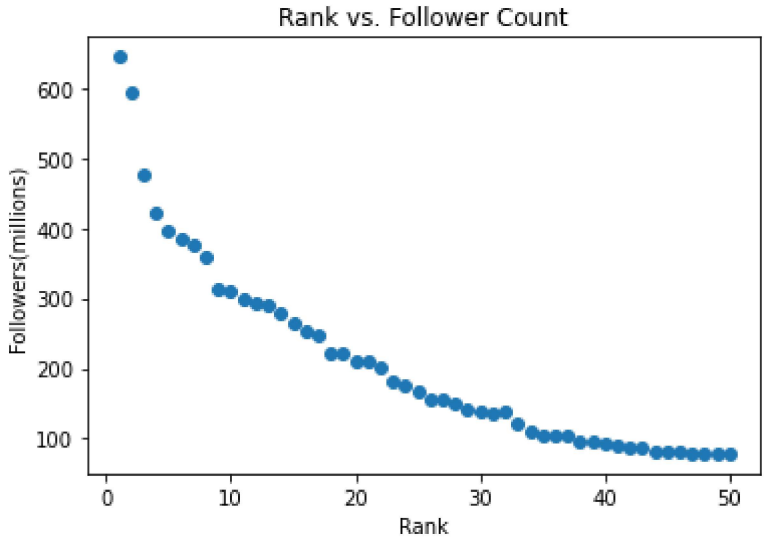
In [19]:
```
common_professions = df['Profession/Activity'].value_counts()
common_professions
```

Out[19]:
```
Musician                                                  12
Musician and actress                                       4
Footballer                                                 3
Television personality and model                           2
Actress                                                    2
Musician and businesswoman                                 2
Football club                                              2
Club football competition                                  1
Footballer at Paris Saint-Germain                          1
Social media platform                                      1
Actor                                                      1
Space agency                                               1
Actress and musician                                       1
Professional basketball league                             1
Former footballer, president of MLS club Inter Miami CF    1
Social media personality                                   1
Comedian and television personality                        1
Actress and singer                                         1
Basketball player                                          1
Comedian and actor                                         1
Cricketer                                                  1
Magazine                                                   1
Model and television personality                           1
Sportswear multinational                                   1
Television personality, model and businesswoman            1
Musician, actress and businesswoman                        1
Actor and professional wrestler                            1
Television personality and businesswoman                   1
Musician, actress, and businesswoman                       1
Model                                                      1
Name: Profession/Activity, dtype: int64
```

Relationship Between Rank and Follower Count

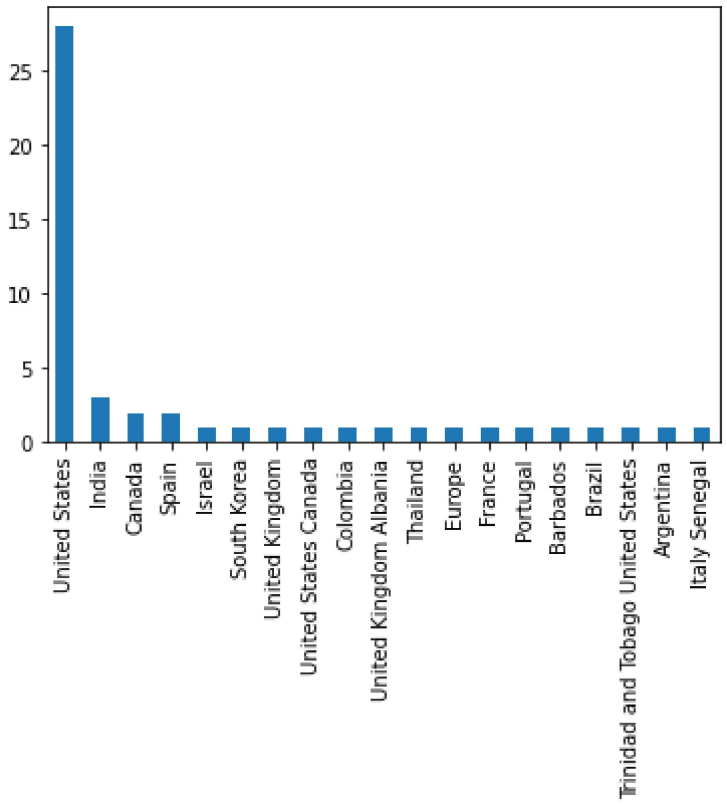
In [20]:
```
import matplotlib.pyplot as plt

plt.scatter(df['Rank'], df['Followers(millions)'])
plt.xlabel('Rank')
plt.ylabel('Followers(millions)')
plt.title('Rank vs. Follower Count')
plt.show()
```



In [21]:
```
df.Country.value_counts().plot(kind = "bar")
```

Out[21]: <AxesSubplot:>



Show all the records where Owner == Cristiano Ronaldo

In [27]:
```
df[df['Owner'] == "Cristiano Ronaldo"]
```

Out[27]:

	Rank	Username	Owner	Followers(millions)	Profession/Activity	Country
1	2	@cristiano	Cristiano Ronaldo	594.0	Footballer	Portugal

In [ ]:
```

```