# PROJECT  REPORT

A project report submitted in fulfillment of

the requirements for the subject

**CAS702 – FOSS LAB**

MSC - Semester II

By

**SHRESTHA BHATTACHARJEE**

**(205322019)**

**DEPARTMENT OF COMPUTER APPLICATIONS**

**NATIONAL INSTITUTE OF TECHNOLOGY**

**TIRUCHIRAPPALLI-620015**

**February 2023**

# PROJECT TITLE :

# INSTAGRAM DATA ANALYSIS USING PYTHON

# Introduction :

In recent years, Instagram has become a powerhouse for influencer marketing, with brands leveraging the power of social media to reach their target audiences through partnerships with influential individuals. In this essay, we will explore the basics of being an Instagram influencer, including how much they can earn per post, what factors determine their rates, and the importance of engagement over follower numbers. We will also discuss the future of influencer marketing, including the push for transparency and the challenges posed by changing regulations. Whether you are an aspiring influencer or simply curious about this rapidly evolving industry, this essay will provide valuable insights into the world of Instagram influencer marketing.

# AIM OF MY PROJECT :

The aim of my project is to find out the majority of the participation of the audiences across various countries. After getting to know about the country, having the highest audience count for the Instagram influencers, the aim lies in analyzing and interpreting the top Instagram influencer in that particular country, in which category does they fall and to have the actual count of the total no. of followers each of these influencers hold.

The DATA SET that has been used throughout the project is given in the Following link :

https://drive.google.com/drive/folders/1RtGinbrBRfETEZOUtP989s0EaZhcAQtw

# DATA MANIPULATION FOR ANALYSIS:

- **Python** language is used for the data analysis.
- The packages that has been used :
    **Numpy, Pandas, Matplotlib**;
- The Pandas library was used to **import the csv file** and a data frame was created.
- We use **head(),tail() , .info and .describe and the pivot tables** to get the relationships among the various aspects of the data set.
- A data frame was created for the influencers with the '**United States**' as their '**Audience Country'.**
- More **pivot tables** were used to get information about the categories' followers and audience engagement in the United States.
- **Matplotlib** was used to have the graphical representations over the **numerical** and the **categorical data**.

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

```
In [2]: data=pd.read_csv('Instagram_influencers.csv')
        data
```

# OUTPUT:

Out[2]:

| | Country | Rank | Account | Title | Link | Category | Followers | Audience Country | Authentic engagement | Engagement avg |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | All | 1 | cristiano | Cristiano Ronaldo | https://www.instagram.com/cristiano/ | Sports with a ball | 400100000 | India | 7800000 | 9500000 |
| 1 | All | 2 | kyliejenner | Kylie ♡ | https://www.instagram.com/kyliejenner/ | Fashion\|Modeling\|Beauty | 308800000 | United States | 6200000 | 10100000 |
| 2 | All | 3 | leomessi | Leo Messi | https://www.instagram.com/leomessi/ | Sports with a ball\|Family | 306300000 | Argentina | 4800000 | 6500000 |
| 3 | All | 4 | kendalljenner | Kendall | https://www.instagram.com/kendalljenner/ | Modeling\|Fashion | 217800000 | United States | 3400000 | 5400000 |
| 4 | All | 5 | selenagomez | Selena Gomez | https://www.instagram.com/selenagomez/ | Music\|Lifestyle | 295800000 | United States | 2700000 | 3600000 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 95 | All | 996 | senoritasaeva | Dina Saeva | https://www.instagram.com/senoritasaeva/ | Lifestyle\|Music\|Modeling | 7700000 | Russia | 246600 | 318200 |
| 96 | All | 997 | manuelneuer | Manuel Neuer | https://www.instagram.com/manuelneuer/ | Sports with a ball | 11500000 | Germany | 146500 | 210200 |
| 97 | All | 998 | sahilkhan | India's Youth & Fitness ICON® | https://www.instagram.com/sahilkhan/ | Fitness\|Gym | 10100000 | India | 176500 | 239800 |
| 98 | All | 999 | mohanshakti | Shakti Mohan | https://www.instagram.com/mohanshakti/ | Art\|Artists\|Cinema\|Actors/actresses | 13700000 | India | 146400 | 175500 |
| 99 | All | 1000 | eduincaz | Eduin Caz | https://www.instagram.com/eduincaz/ | Lifestyle | 6200000 | Mexico | 305600 | 391900 |

00 rows × 11 columns

## Using head() to fetch the top 5 datas:

In [8]: `data.head()`

Out[8]:

| | Country | Rank | Account | Title | Link | Category | Followers | Audience Country | Authentic engagement | Enga |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | All | 1 | cristiano | Cristiano Ronaldo | https://www.instagram.com/cristiano/ | Sports with a ball | 400100000 | India | 7800000 | |
| 1 | All | 2 | kyliejenner | Kylie ♡ | https://www.instagram.com/kyliejenner/ | Fashion\|Modeling\|Beauty | 308800000 | United States | 6200000 | 1( |
| 2 | All | 3 | leomessi | Leo Messi | https://www.instagram.com/leomessi/ | Sports with a ball\|Family | 306300000 | Argentina | 4800000 | |
| 3 | All | 4 | kendalljenner | Kendall | https://www.instagram.com/kendalljenner/ | Modeling\|Fashion | 217800000 | United States | 3400000 | |
| 4 | All | 5 | selenagomez | Selena Gomez | https://www.instagram.com/selenagomez/ | Music\|Lifestyle | 295800000 | United States | 2700000 | |

## Using info to get the information about the data set :

In [9]: `data.info`

Out[9]:
```
<bound method DataFrame.info of        Country  Rank         Account                         Account
0          All     1        cristiano              Cristiano Ronaldo
1          All     2      kyliejenner                     Kylie ♡
2          All     3         leomessi                      Leo Messi
3          All     4    kendalljenner                        Kendall
4          All     5      selenagomez                   Selena Gomez
..         ...   ...              ...                            ...
995        All   996    senoritasaeva                     Dina Saeva
996        All   997      manuelneuer                   Manuel Neuer
997        All   998        sahilkhan  India's Youth & Fitness ICON®
998        All   999      mohanshakti                   Shakti Mohan
999        All  1000          eduincaz                      Eduin Caz

                                        Link  \
0            https://www.instagram.com/cristiano/
1          https://www.instagram.com/kyliejenner/
2             https://www.instagram.com/leomessi/
3         https://www.instagram.com/kendalljenner/
4          https://www.instagram.com/selenagomez/
..                                           ...
995       https://www.instagram.com/senoritasaeva/
996        https://www.instagram.com/manuelneuer/
997          https://www.instagram.com/sahilkhan/
998         https://www.instagram.com/mohanshakti/
999           https://www.instagram.com/eduincaz/

                         Category  Followers Audience Country  \
0              Sports with a ball  400100000            India
1         Fashion|Modeling|Beauty  308800000    United States
2         Sports with a ball|Family  306300000        Argentina
3               Modeling|Fashion  217800000    United States
4                 Music|Lifestyle  295800000    United States
..                           ...        ...              ...
995      Lifestyle|Music|Modeling    7700000           Russia
```

```
                            Category  Followers Audience Country  \
0                  Sports with a ball  400100000            India
1             Fashion|Modeling|Beauty  308800000    United States
2           Sports with a ball|Family  306300000        Argentina
3                    Modeling|Fashion  217800000    United States
4                     Music|Lifestyle  295800000    United States
..                                ...        ...              ...
995           Lifestyle|Music|Modeling    7700000           Russia
996                 Sports with a ball   11500000          Germany
997                       Fitness|Gym   10100000            India
998   Art|Artists|Cinema|Actors/actresses  13700000            India
999                          Lifestyle    6200000           Mexico

     Authentic engagement  Engagement avg  Scraped
0                 7800000         9500000  50:24.8
1                 6200000        10100000  50:24.8
2                 4800000         6500000  50:24.8
3                 3400000         5400000  50:24.8
4                 2700000         3600000  50:24.8
..                    ...             ...      ...
995                246600          318200  50:24.8
996                146500          210200  50:24.8
997                176500          239800  50:24.8
998                146400          175500  50:24.8
999                305600          391900  50:24.8

[1000 rows x 11 columns]>
```

## Using the describe() to get the overall statistical summary about the data.

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

Out[11]:

|       | Rank        | Followers    | Authentic engagement | Engagement avg |
|-------|-------------|--------------|----------------------|----------------|
| count | 1000.000000 | 1.000000e+03 | 1.000000e+03         | 1.000000e+03   |
| mean  | 500.500000  | 2.483610e+07 | 6.491228e+05         | 8.653886e+05   |
| std   | 288.819436  | 3.872143e+07 | 1.098804e+06         | 1.267515e+06   |
| min   | 1.000000    | 2.800000e+06 | 0.000000e+00         | 5.200000e+04   |
| 25%   | 250.750000  | 9.000000e+06 | 2.091000e+05         | 3.086000e+05   |
| 50%   | 500.500000  | 1.440000e+07 | 3.637000e+05         | 5.239000e+05   |
| 75%   | 750.250000  | 2.565000e+07 | 6.832500e+05         | 9.508000e+05   |
| max   | 1000.000000 | 4.696000e+08 | 1.340000e+07         | 1.340000e+07   |

```
In [13]: data['Audience Country'].value_counts()

Out[13]: United States           283
         Brazil                  161
         India                   143
         Indonesia               130
         Mexico                   50
         Spain                    34
         Russia                   30
         Argentina                24
         Iran                     17
         United Kingdom           17
         Turkey                   16
         Italy                    15
         South Korea              13
         Colombia                  9
         Philippines               7
         France                    6
         Egypt                     6
         Germany                   5
         Nigeria                   4
         Thailand                  4
         Iraq                      4
         Morocco                   3
         Japan                     2
         Saudi Arabia              2
         Kazakhstan                2
         Syria                     1
         China                     1
         Algeria                   1
         United Arab Emirates      1
         Poland                    1
         Chile                     1
         Name: Audience Country, dtype: int64
```

**Through the Above Data Series we could see that the UNITED STATES serves as the greatest "audience country" for the top 1000 instagram influencers.**

```
In [18]: data.pivot_table(index="Category",values="Followers", aggfunc=['mean', 'sum', 'min', 'max'], sort=False)
```
Out[18]:

:

|  | mean | sum | min | max |
|---|---|---|---|---|
|  | Followers | Followers | Followers | Followers |
| **Category** |  |  |  |  |
| **Sports with a ball** | 3.225825e+07 | 3322600000 | 5400000 | 400100000 |
| **Fashion\|Modeling\|Beauty** | 3.088000e+08 | 308800000 | 308800000 | 308800000 |
| **Sports with a ball\|Family** | 1.730500e+08 | 346100000 | 39800000 | 306300000 |
| **Modeling\|Fashion** | 6.622500e+07 | 264900000 | 11400000 | 217800000 |
| **Music\|Lifestyle** | 4.848000e+07 | 484800000 | 7600000 | 295800000 |
| ... | ... | ... | ... | ... |
| **Photography\|Lifestyle** | 7.900000e+06 | 7900000 | 7900000 | 7900000 |
| **Modeling\|Music** | 6.800000e+06 | 6800000 | 6800000 | 6800000 |
| **Beauty\|Art\|Artists\|Modeling** | 1.360000e+07 | 13600000 | 13600000 | 13600000 |
| **Humor\|Fun\|Happiness\|Shows\|Art\|Artists** | 1.570000e+07 | 15700000 | 15700000 | 15700000 |
| **Lifestyle\|Music\|Modeling** | 7.700000e+06 | 7700000 | 7700000 | 7700000 |

165 rows × 4 columns

**Through this pivot table we have obtained the summary for all the categories of the Instagram influencers.**

```
In [21]: data['Category'].value_counts()
```

```
Out[21]: Music                                                 170
         Cinema|Actors/actresses                               165
         Sports with a ball                                    103
         Lifestyle                                              58
         Music|Cinema|Actors/actresses                         20
                                                               ...
         Modeling|Fashion|Lifestyle                             1
         Fashion|Modeling|Beauty                                1
         Cinema|Actors/actresses|Comics  sketches|Shows         1
         Lifestyle|Cinema|Actors/actresses|Music                1
         Lifestyle|Music|Modeling                               1
         Name: Category, Length: 165, dtype: int64
```

**Now we get the value of each of the category counts and also :**

```
In [22]: data['Category'].value_counts().shape
Out[22]: (165,)
```

**Where 165 shows that there are total 165 categories of the different Instagram influencers.**

**We now plot a bar graph of the categories with respect to their value counts :**

```
In [24]: import matplotlib.pyplot as plyt
         data['Category'].value_counts().head().plot(kind='barh')
Out[24]: <AxesSubplot:>
```
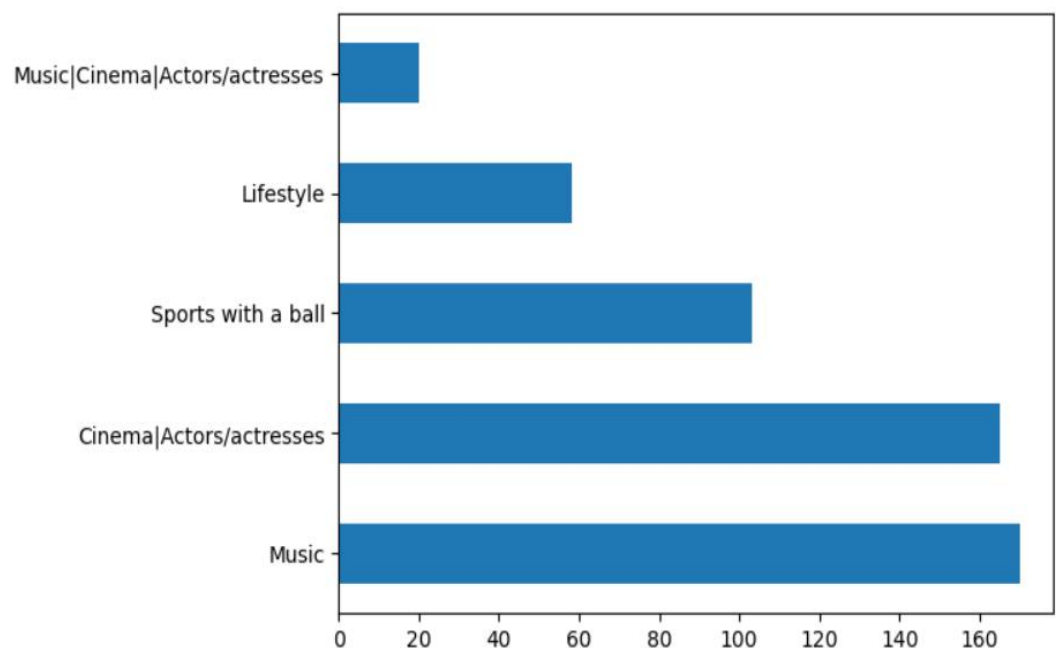
**So, among the top influencers across the world , the category "Music" has more popularity.**

## Now we are here to create a new data frame for the UNITED STATES

```
In [26]: mask=data['Audience Country']=='United States'
         df_USA=data[mask]
         df_USA.head(6)
```

## OUTPUT:

Out[26]:

| | Country | Rank | Account | Title | Link | Category | Followers | Audience Country | Authentic engagement | Engageme a |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | All | 2 | kyliejenner | Kylie ♡ | https://www.instagram.com/kyliejenner/ | Fashion\|Modeling\|Beauty | 308800000 | United States | 6200000 | 101000 |
| 3 | All | 4 | kendalljenner | Kendall | https://www.instagram.com/kendalljenner/ | Modeling\|Fashion | 217800000 | United States | 3400000 | 54000 |
| 4 | All | 5 | selenagomez | Selena Gomez | https://www.instagram.com/selenagomez/ | Music\|Lifestyle | 295800000 | United States | 2700000 | 36000 |
| 5 | All | 6 | zendaya | Zendaya | https://www.instagram.com/zendaya/ | Cinema\|Actors/actresses\|Fashion | 127800000 | United States | 5800000 | 78000 |
| 6 | All | 7 | kimkardashian | Kim Kardashian West | https://www.instagram.com/kimkardashian/ | Fashion\|Beauty | 284900000 | United States | 2200000 | 33000 |
| 7 | All | 8 | beyonce | Beyoncé | https://www.instagram.com/beyonce/ | Music\|Fashion | 237200000 | United States | 2500000 | 36000 |

## From the above Data frame we could see that "Kylie Jennier" serves as the top most influencer in USA.

```
In [27]: topfollowers = df_USA.nlargest(10,'Followers')

         topfollowers.pivot_table(index="Category", values="Followers", aggfunc=['sum'], sort=False)
```

Out[27]:

| | sum |
|---|---|
| | Followers |
| **Category** | |
| Fashion\|Modeling\|Beauty | 308800000 |
| Music\|Lifestyle | 295800000 |
| Music | 294100000 |
| Fashion\|Beauty | 284900000 |
| Music\|Fashion | 237200000 |
| Clothing\|Outfits\|Lifestyle | 219400000 |
| Modeling\|Fashion | 217800000 |
| Nature landscapes\|Science\|Photography | 206600000 |
| Fitness\|Gym\|Shopping\|Retail\|Clothing\|Outfits | 199000000 |
| Music\|Cinema\|Actors/actresses | 194900000 |

**Through the Output above we could see that in USA the most preferred Category is Fashion/Modelling/Beauty**

**Whereas world wide for the top 1000 instagram influencers it was the MUSIC.**

**Next we plot a histogram to evaluate the Follower Counts of Top Instagram Accounts in the US:**

```
In [34]: plt.figure(figsize=(8,5))
         plt.hist(df_USA.Followers, color='red', range=(0,60000000))
         plt.xlabel('Followers', fontsize = 16)
         plt.ylabel('Number of Instagram Accounts', fontsize = 16)
         plt.title('Histogram: Follower Counts of Top Instagram Accounts in the US', fontsize = 20)
         plt.tight_layout()
         # remove scientific notation
         plt.ticklabel_format(style='plain')
```

**OUTPUT:**



Histogram: Follower Counts of Top Instagram Accounts in the US
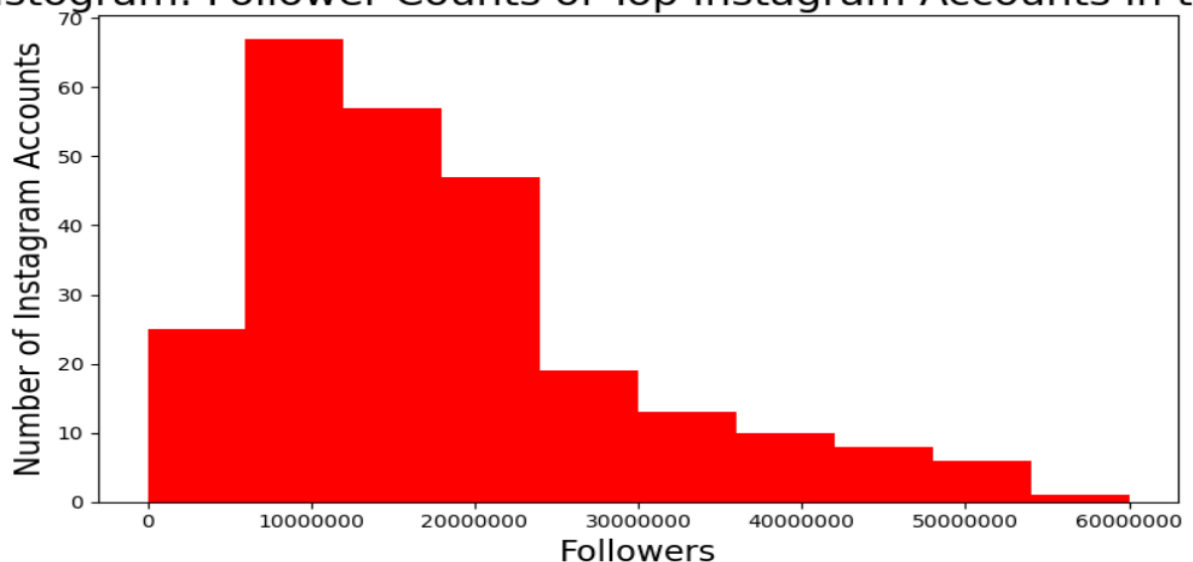
# CONCLUSIONS:

The realm of Instagram influence is a rapidly growing and competitive field, with various categories vying for prominence. A cursory analysis of the top Instagram influencers in the United States reveals that the most popular categories include Fashion, Music, Cinema, Acting, Modeling, Beauty, and Lifestyle. Of these, the Fashion category stands out as the most prominent, with influencers in this category enjoying the highest number of followers and engagement.

It is evident that Instagram has become an indispensable platform for fashion influencers, who leverage their content to showcase the latest trends, styles, and products to their followers. The ability to connect with a vast audience through visually stunning images and videos has made Instagram an ideal platform for fashion influencers to build their brand and establish a loyal following. As a result, many of the top Instagram influencers in the United States are fashion bloggers, designers, stylists, and models.

The Fashion category's dominance in the world of Instagram influence is a reflection of the industry's global appeal, which generates billions of dollars in revenue each year. With the rise of social media and e-commerce platforms, fashion influencers have become a crucial component of the industry's marketing and advertising strategies. Their ability to reach a large and diverse audience has made them valuable partners for brands looking to promote their products and services to a wider market.

In conclusion, the analysis of the top Instagram influencers in the United States reveals that the Fashion category is the most prominent and influential category. Fashion influencers leverage Instagram's visual platform to showcase the latest trends, styles, and products to their followers, resulting in high engagement and a vast following. As such, the Fashion category should be considered the top category for Instagram influencers in the United States.