

DMDD Final Project: Social Media Influencer Data Repository

[Project Objective](#)

[Team Members](#)

[ER Diagram](#)

[Real-Time Data Collection: Web Scrapping YouTube & Twitter](#)

- 1) [The Twitter Scrapper: \(twitterscrapper_dmdd.py\)](#)
- 2) [The YouTube Scrapper: \(youtubescrapper_dmdd.py\)](#)

[Data Cleaning](#)

[Null checks, Duplicates check \(DataCleaning.ipynb \)](#)

[Data Standardization](#)

[Data Audit](#)

[Data Audit validity/ accuracy](#)

[Audit completeness](#)

[Audit consistency/uniformity](#)

[Table Creation](#)

[Normalisation](#)

[1st Normal Form](#)

[Table Name: youtube_profile_data](#)

[Table Name: profile_data_youtube_twitter](#)

[Table Name: tweets](#)

[Table Name: profile_data_twitter](#)

[Tables Added in 1NF: Category, Profile platform](#)

[2nd Normal Form](#)

[Table Name: Youtube_Profile_Data](#)

[Table Name: profile_data_youtube_twitter](#)

[Table Name: tweets](#)

[Table Name: twitter profile](#)

[3rd Normal Form](#)

[Table Name: Youtube_Profile_Data](#)

[Table Name: profile_data_youtube_twitter](#)

[Table Name: tweets](#)

[Table Name: twitter profile](#)

[Use Cases](#)

[Data Insights with Visualisations](#)

- 1) [Table Name: Twitter profile data](#)
- 2) [Table Name: profile_data](#)
- 3) [Table Name: youtube_profile_data](#)
- 4) [Table Name: Tweets](#)

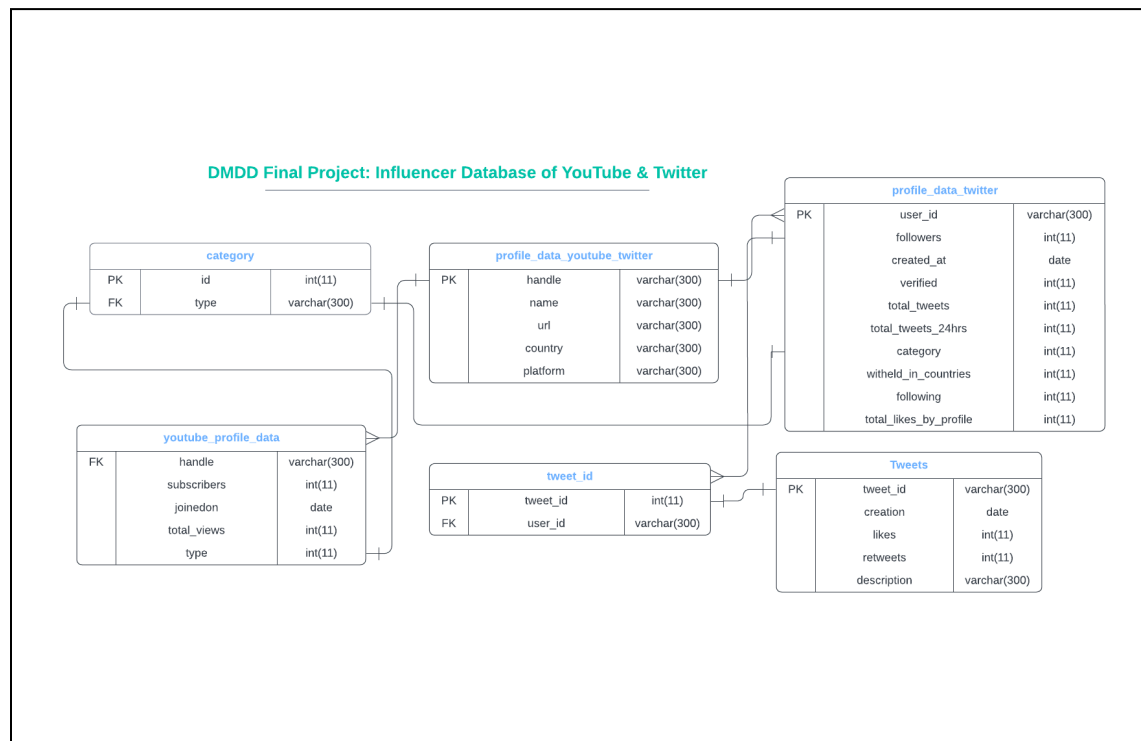
Project Objective

The objective of our project is to design a database of YouTube and Twitter influencers catering to business marketing needs. The scope of our project includes records on the influencers reach, type of content, and origin of the content, amongst other things. This sophisticated system can further be used to deduce and match which social media influencer is a good fit for advertising businesses based on many criteria.

Team Members

1. Hamzah Mukadam (NUID: 002741426)
2. Shreya Maher (NUID: 002770227)
3. Veenadharini Shukla (NUID: 002704948)

ER Diagram



Real-Time Data Collection: Web Scrapping YouTube & Twitter

We are designing a database for influencers on YouTube and Twitter and hence our primary sources of data would be their respective websites

- 1) For YouTube Data: <https://www.youtube.com/>
- 2) For Twitter Data: <https://twitter.com/>

We will be scraping profiles of Influencers from both platforms across categories as:

News, Programming, Gaming, Fashion, Makeup, Fitness, Movie Reviews, and Technology

1) The Twitter Scraper: (twitterscrapper_dmdd.py)

- Fetch top users under a particular category with their complete profile information and tweets made in past 24 hours.
- Can be used to perform analysis of reach and engagement of a particular users profile and to check fit of a particular influencer for certain marketing needs
- Input: Keyword
- Output: 2 CSV files with users information and information of related tweets from profiles
- (Parameters in 2 files: User ID, Tweet Created At, Tweet Likes, Retweets, Tweet Description, Name, Followers, Location, Created At, Verified, URL, Total Tweets, Total tweets in last 24hrs, Latest Status updated by User, Withheld In Countries, Following, Total Likes By Profile)

2) The YouTube Scraper: (youtubescrapper_dmdd.py)

- Fetch top users under a particular category with their complete profile information and latest videos uploaded
- Can be used to perform analysis of reach and engagement of a particular users profile and to check fit of a particular influencer for certain marketing needs
- Input: Keyword
- Output: 2 CSV files with users information and information of related videos from profiles
- (Parameters in 2 files: User ID, Profile Views, Profile Subscribers, Name, Location, Created At, URL)

Data Cleaning

Null checks, Duplicates check (DataCleaning.ipynb)

File DataCleaning.ipynb was used to clean data. Cleaning operations that were performed are as shown below:

1) Null check done on files:

```
import pandas as pd

pd.set_option('display.max_rows',None)

df = pd.read_csv(r'C:\Users\maher\Desktop\YouTube Profile Data.csv')

df.head()
```

	handle	type	subscribers	total	on
0	@CamodoGaming	Gaming	2.99M	1,89,58,38,340	Jan 5, 2016
1	@MagicMusicGroup	Gaming	4.62M	93,25,71,647	May 6, 2016
2	@UnspeakablePlays	Gaming	7.33M	2,25,16,66,389	Jun 26, 2017
3	@GameToonsGaming	Gaming	1.53M	48,56,38,386	Feb 22, 2021
4	@JKrewGaming	Gaming	460K	21,61,17,057	Sep 29, 2020

```
df.isna().sum()
#since there are no null values, data is in required format
```

```
handle      0
type        0
subscribers 0
total      29
on          0
dtype: int64
```

Since Total views can be null on profiles these values are relevant but should be zero instead of null

2) Further replacing null values

```
df.isna().sum()
#since there are no null values, data is in required format. |
df.fillna(0, inplace=True)
df.isna().sum()
```

✓ 0.3s

```
handle      0
type        0
subscribers 0
total       0
on          0
dtype: int64
```

3) Dropping entries with null values:

```

df2.isna().sum()

[4] ✓ 0.7s

...
name      0
handle    3
url       3
desc     19
country  96
platform  0
dtype: int64

df2new = df2.dropna()
df2 = df2new
df2.isna().sum()

[5] ✓ 0.3s

...
name      0
handle    0
url       0
desc     0
country   0
platform  0
dtype: int64

```

4) Checking Duplicates

```

df2new = df2.dropna()
df2 = df2new
df2.isna().sum()
df2.duplicated().sum()
#df2.head()

✓ 0.5s

6

```

5) After removing duplicates from file

```

df2new = df2.dropna()
df2 = df2new
df2.isna().sum()
df2.duplicated().sum()
df2.drop_duplicates(inplace = True)
df2.duplicated().sum()
#df2.head()

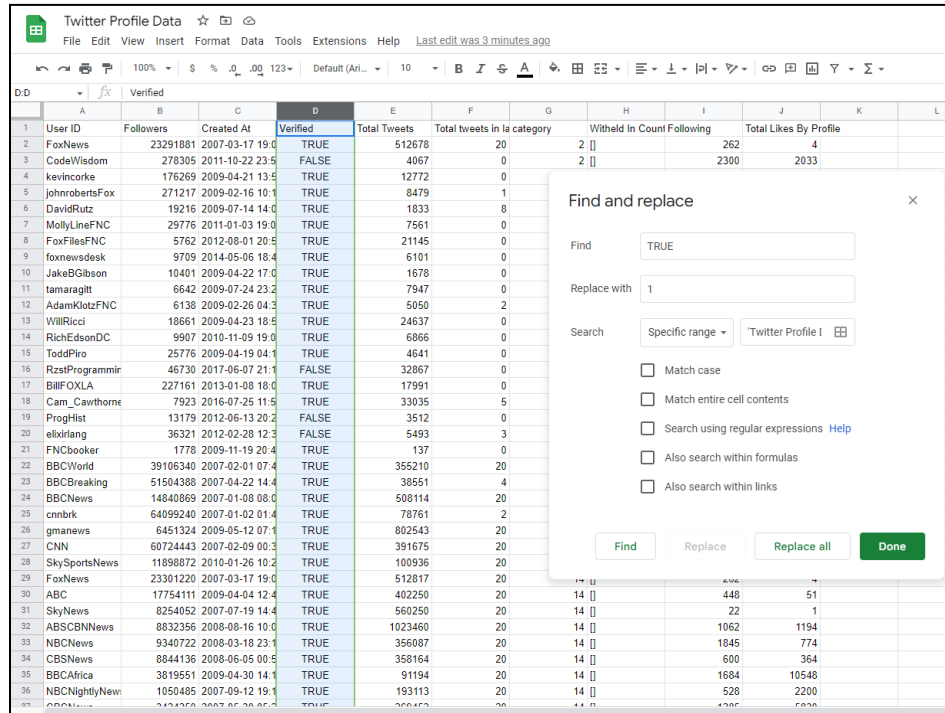
✓ 0.7s

0

```

Data Standardization

Setting verified status from TRUE and FALSE to 1,0 respectively



Find and replace

Find: TRUE

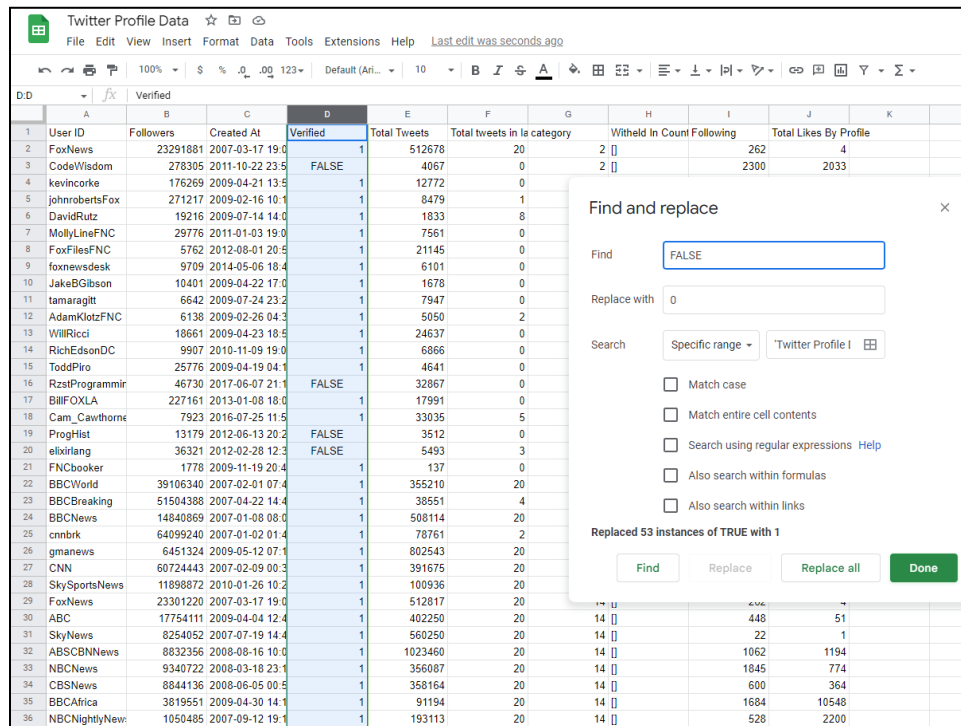
Replace with: 1

Search: Specific range | Twitter Profile I

☐ Match case
☐ Match entire cell contents
☐ Search using regular expressions [Help](#)
☐ Also search within formulas
☐ Also search within links

Find Replace Replace all Done

	A	B	C	D	E	F	G	H	I	J	K	L
	User ID	Followers	Created At	Verified	Total Tweets	Total tweets in la category	Withheld In Count Following	Total Likes By Profile				
1	FoxNews	23291881	2007-03-17 19:0	TRUE	512678	20	2		262	4		
2	CodeWisdom	278305	2011-10-22 23:5	FALSE	4067	0	2		2300	2033		
3	kevincorke	176269	2009-04-21 13:5	TRUE	12772	0						
4	johnrobertsFox	271217	2009-02-16 10:1	TRUE	8479	1						
5	DavidRutz	19216	2009-07-14 14:0	TRUE	1833	8						
6	MollyLineFNC	29776	2011-01-03 19:0	TRUE	7561	0						
7	FoxFilesFNC	5762	2012-08-01 20:5	TRUE	21145	0						
8	foxnewsdesk	9709	2014-05-06 18:4	TRUE	6101	0						
9	JakeBGibson	10401	2009-04-22 17:0	TRUE	1678	0						
10	tamaragit	6642	2009-07-24 23:2	TRUE	7947	0						
11	AdamKlotzFNC	6138	2009-02-26 04:3	TRUE	5050	2						
12	WillRicci	18661	2009-04-23 18:5	TRUE	24637	0						
13	RichEdsonDC	9907	2010-11-09 19:0	TRUE	6866	0						
14	ToddPiro	25776	2009-04-19 04:1	TRUE	4641	0						
15	RzstProgrammir	46730	2017-06-07 21:1	FALSE	32867	0						
16	BIIFOXLA	227161	2013-01-08 18:0	TRUE	17991	0						
17	Cam_Cawthorne	7923	2016-07-25 11:5	TRUE	33035	5						
18	ProgHist	13179	2012-06-13 20:2	FALSE	3512	0						
19	elixirfang	36321	2012-02-28 12:3	FALSE	5493	3						
20	FNCbooker	1778	2009-11-19 20:4	TRUE	137	0						
21	BBCWorld	39106340	2007-02-01 07:4	TRUE	355210	20						
22	BBCBreaking	51504388	2007-04-22 14:4	TRUE	38551	4						
23	BBCNews	14840869	2007-01-08 08:0	TRUE	508114	20						
24	cnbrk	64099240	2007-01-02 01:4	TRUE	78761	2						
25	gmanews	6451324	2009-05-12 07:1	TRUE	802543	20						
26	CNN	60724443	2007-02-09 00:3	TRUE	391675	20						
27	SkySportsNews	11898872	2010-01-26 10:2	TRUE	100936	20						
28	FoxNews	23301220	2007-03-17 19:0	TRUE	512817	20						
29	ABC	17754111	2009-04-04 12:4	TRUE	402250	20	14		448	51		
30	SkyNews	8254052	2007-07-19 14:4	TRUE	560250	20	14		22	1		
31	ABSCBNNews	8832356	2008-08-16 10:0	TRUE	1023460	20	14		1062	1194		
32	NBCNews	9340722	2008-03-18 23:1	TRUE	356087	20	14		1845	774		
33	CBSNews	8844136	2008-06-05 00:5	TRUE	358164	20	14		600	364		
34	BBCAfrica	3819551	2009-04-30 14:1	TRUE	91194	20	14		1684	10548		
35	NBCNightlyNew	1050485	2007-09-12 19:1	TRUE	193113	20	14		528	2200		



Find and replace

Find: FALSE

Replace with: 0

Search: Specific range | Twitter Profile I

☐ Match case
☐ Match entire cell contents
☐ Search using regular expressions [Help](#)
☐ Also search within formulas
☐ Also search within links

Replaced 53 instances of TRUE with 1

Find Replace Replace all Done

	A	B	C	D	E	F	G	H	I	J	K	L
	User ID	Followers	Created At	Verified	Total Tweets	Total tweets in la category	Withheld In Count Following	Total Likes By Profile				
1	FoxNews	23291881	2007-03-17 19:0	FALSE	512678	20	2		262	4		
2	CodeWisdom	278305	2011-10-22 23:5	FALSE	4067	0	2		2300	2033		
3	kevincorke	176269	2009-04-21 13:5	1	12772	0						
4	johnrobertsFox	271217	2009-02-16 10:1	1	8479	1						
5	DavidRutz	19216	2009-07-14 14:0	1	1833	8						
6	MollyLineFNC	29776	2011-01-03 19:0	1	7561	0						
7	FoxFilesFNC	5762	2012-08-01 20:5	1	21145	0						
8	foxnewsdesk	9709	2014-05-06 18:4	1	6101	0						
9	JakeBGibson	10401	2009-04-22 17:0	1	1678	0						
10	tamaragit	6642	2009-07-24 23:2	1	7947	0						
11	AdamKlotzFNC	6138	2009-02-26 04:3	1	5050	2						
12	WillRicci	18661	2009-04-23 18:5	1	24637	0						
13	RichEdsonDC	9907	2010-11-09 19:0	1	6866	0						
14	ToddPiro	25776	2009-04-19 04:1	1	4641	0						
15	RzstProgrammir	46730	2017-06-07 21:1	FALSE	32867	0						
16	BIIFOXLA	227161	2013-01-08 18:0	1	17991	0						
17	Cam_Cawthorne	7923	2016-07-25 11:5	1	33035	5						
18	ProgHist	13179	2012-06-13 20:2	FALSE	3512	0						
19	elixirfang	36321	2012-02-28 12:3	FALSE	5493	3						
20	FNCbooker	1778	2009-11-19 20:4	1	137	0						
21	BBCWorld	39106340	2007-02-01 07:4	1	355210	20						
22	BBCBreaking	51504388	2007-04-22 14:4	1	38551	4						
23	BBCNews	14840869	2007-01-08 08:0	1	508114	20						
24	cnbrk	64099240	2007-01-02 01:4	1	78761	2						
25	gmanews	6451324	2009-05-12 07:1	1	802543	20						
26	CNN	60724443	2007-02-09 00:3	1	391675	20						
27	SkySportsNews	11898872	2010-01-26 10:2	1	100936	20						
28	FoxNews	23301220	2007-03-17 19:0	1	512817	20						
29	ABC	17754111	2009-04-04 12:4	1	402250	20	14		448	51		
30	SkyNews	8254052	2007-07-19 14:4	1	560250	20	14		22	1		
31	ABSCBNNews	8832356	2008-08-16 10:0	1	1023460	20	14		1062	1194		
32	NBCNews	9340722	2008-03-18 23:1	1	356087	20	14		1845	774		
33	CBSNews	8844136	2008-06-05 00:5	1	358164	20	14		600	364		
34	BBCAfrica	3819551	2009-04-30 14:1	1	91194	20	14		1684	10548		
35	NBCNightlyNew	1050485	2007-09-12 19:1	1	193113	20	14		528	2200		

Data Audit

Data Audit validity/ accuracy

Null Checks are performed on Tweets.csv

```
print(f'User ID: {df.User_ID.isnull().sum()}')
print(f'Tweet Created at: {df.Tweet_Created_At.isnull().sum()}')
print(f'Tweet Likes: {df.Tweet_Likes.isnull().sum()}')
print(f'Retweets: {df.Retweets.isnull().sum()}')
print(f'Tweet Description: {df.Tweet_Description.isnull().sum()}')
```

[28] ✓ 0.1s Python

```
... User ID: 0
    Tweet Created at: 0
    Tweet Likes: 0
    Retweets: 0
    Tweet Description: 0
```

Audit completeness

Performed univariate analysis for Tweets.csv

```
import numpy as np

print(df.Tweet_Likes.min())
print(df.Tweet_Likes.quantile(.25))
print(df.Tweet_Likes.quantile(.50))
print(df.Tweet_Likes.quantile(.75))
print(df.Tweet_Likes.max())
print(df.Tweet_Likes.mean())
print(df.Tweet_Likes.median())
print(df.Tweet_Likes.mode().values[0])
```

[33] ✓ 0.4s Python

```
... 0
    0.0
    14.0
    50.0
    1883
    94.69166666666666
    14.0
    0
```

Checked uniqueness of the data:

```
df.shape
```

[20] ✓ 0.3s

```
... (120, 6)
```



```

print(f'User ID: {df.User_ID.nunique()}')
print(f'Tweet Created at: {df.Tweet_Created_At.nunique()}')
print(f'Tweet Likes: {df.Tweet_Likes.nunique()}')
print(f'Retweets: {df.Retweets.nunique()}')
print(f'Tweet Description: {df.Tweet_Description.nunique()}')

```

[23] ✓ 0.2s Python

... User ID: 13
 Tweet Created at: 120
 Tweet Likes: 57
 Retweets: 30
 Tweet Description: 120

Checked the datatype of each column:

```

print(f'User ID: {df.User_ID.dtype}')
print(f'Tweet Created at: {df.Tweet_Created_At.dtype}')
print(f'Tweet Likes: {df.Tweet_Likes.dtype}')
print(f'Retweets: {df.Retweets.dtype}')
print(f'Tweet Description: {df.Tweet_Description.dtype}')

```

[25] ✓ 0.4s Python

... User ID: object
 Tweet Created at: object
 Tweet Likes: int64
 Retweets: int64
 Tweet Description: object

```

print(f'User ID: {pd.api.types.is_numeric_dtype(df.User_ID.dtype)}')
print(f'Tweet Created at: {pd.api.types.is_numeric_dtype(df.Tweet_Created_At.dtype)}')
print(f'Tweet Likes: {pd.api.types.is_numeric_dtype(df.Tweet_Likes.dtype)}')
print(f'Retweets: {pd.api.types.is_numeric_dtype(df.Retweets.dtype)}')
print(f'Tweet Description: {pd.api.types.is_numeric_dtype(df.Tweet_Description.dtype)}')

```

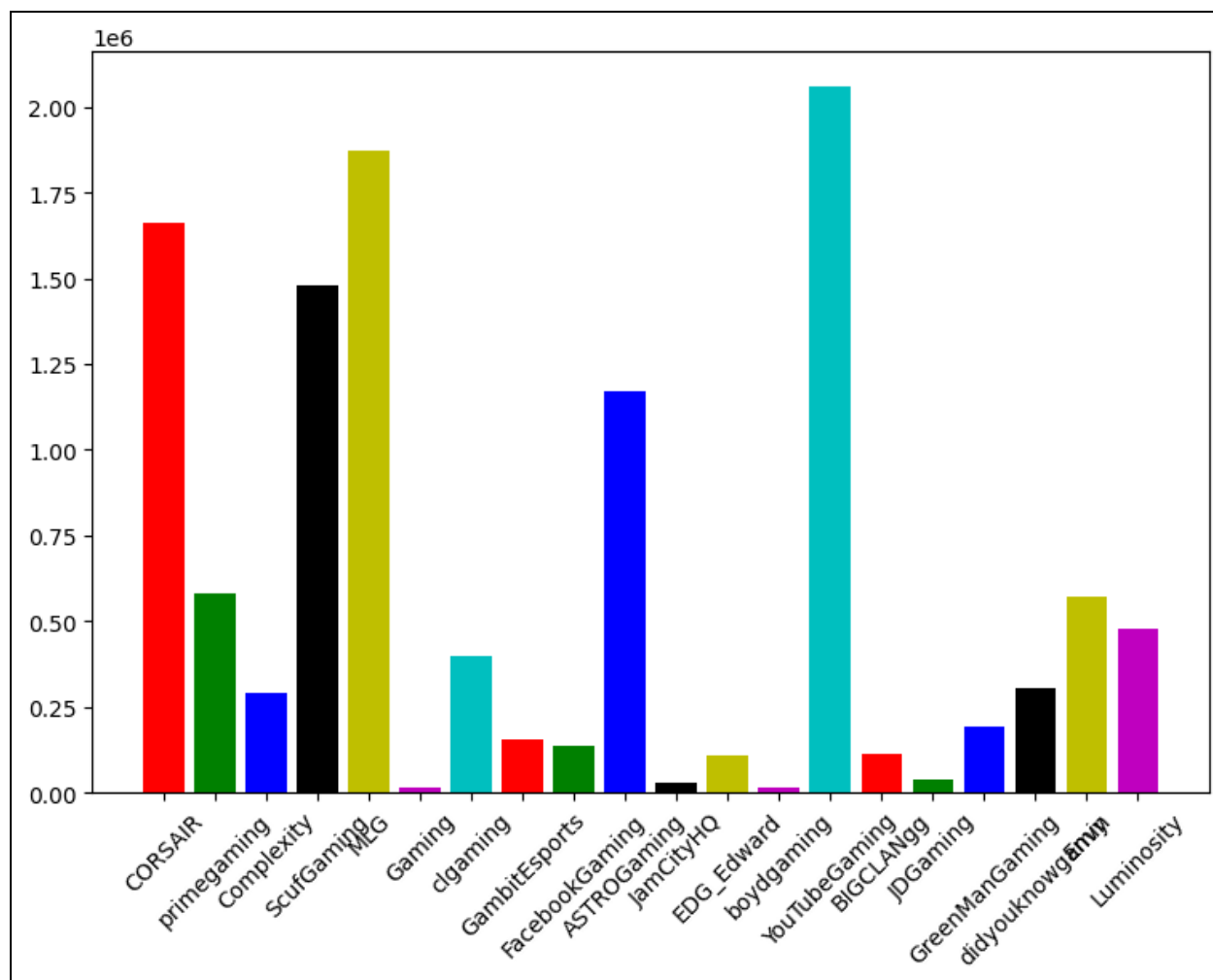
[27] ✓ 0.3s Python

... User ID: False
 Tweet Created at: False
 Tweet Likes: True
 Retweets: True
 Tweet Description: False

Audit consistency/uniformity

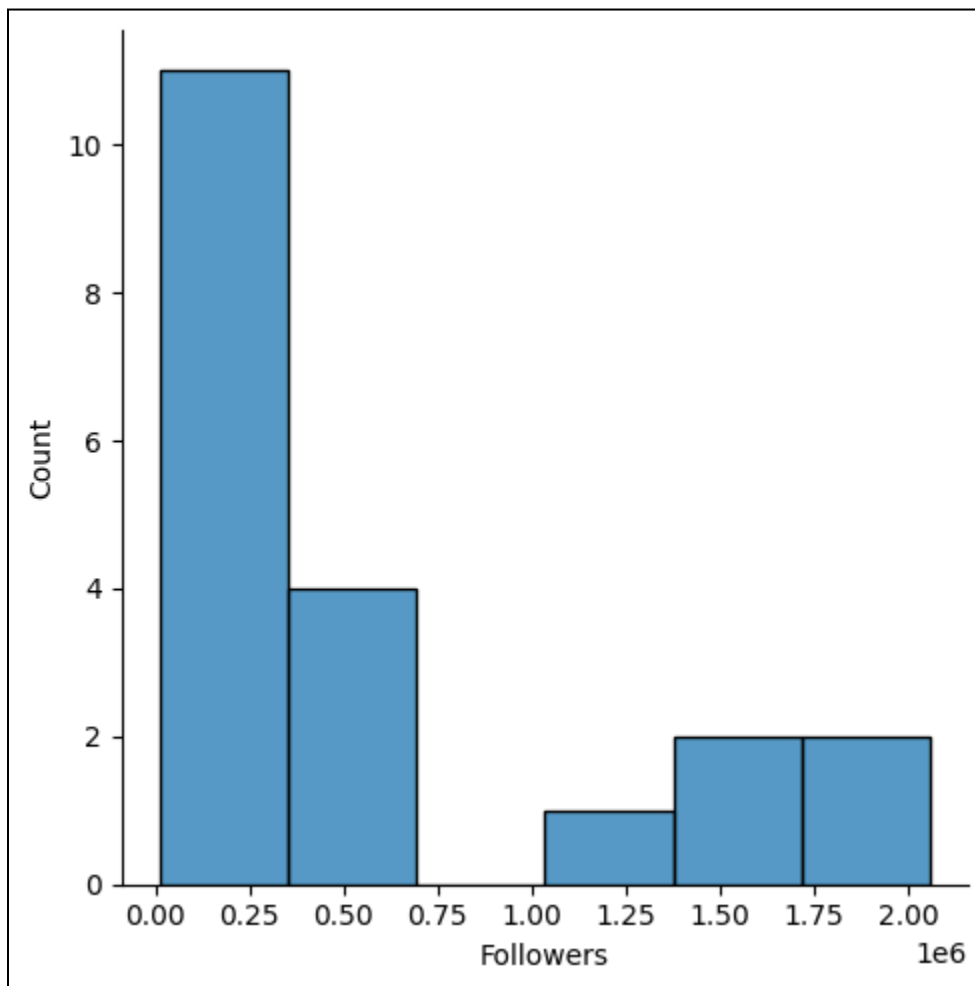
Data Visualizations created for the Twitter_Users.csv

Twitter Followers count of each user using Bar graph:



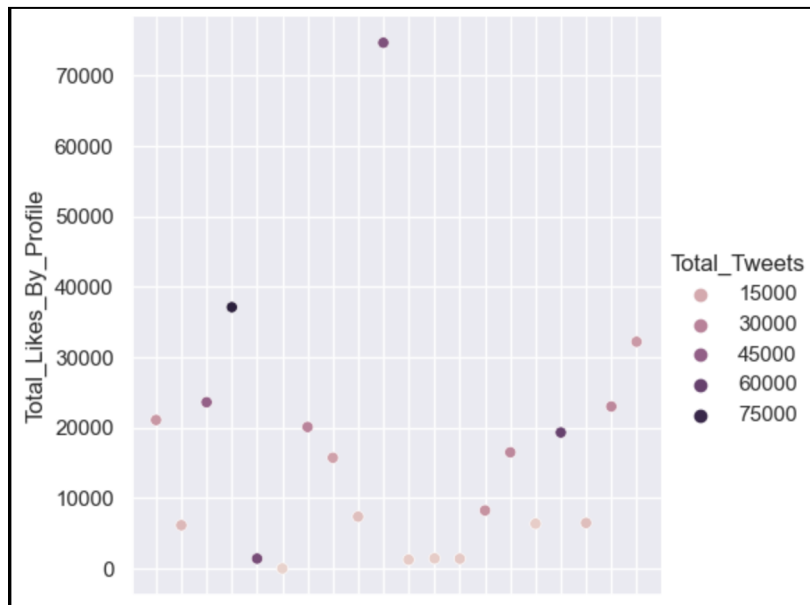
The above graph visualizes the data in the form of the follower count against the handle of the twitter accounts. From this, we can analyze which twitter user would be the most suited to campaign for a specific organization's business needs. The above twitter user's are from the gaming category and could be suitable for marketing in the same domain. This bivariate analysis of handle on follower count is extremely insightful and gets the relevant information that would assist in choosing the right influencer.

Displot for Followers to the count of tweets.



The above displot maps the number of followers with the tweets, which gives insight on how the interaction on the tweets increases with the increase in the number of followers. This visualized data can be used to decide which is the sweet spot between the follower count and engagement on posts

PairPlot for Twitter_User:



The above pairplot gives the relation between the total number of tweets made by a profile to the total likes received by a profile on all their tweets combined. This gives us insightful information on how the total likes increase based on the sheer number of tweets made. This will help the organization decide if they are partnering with an influencer that has more engagement on their tweets or an influencer that has a higher number of tweets but lower overall engagement. This could help the organizations decide whether they want to go for quality or quantity.

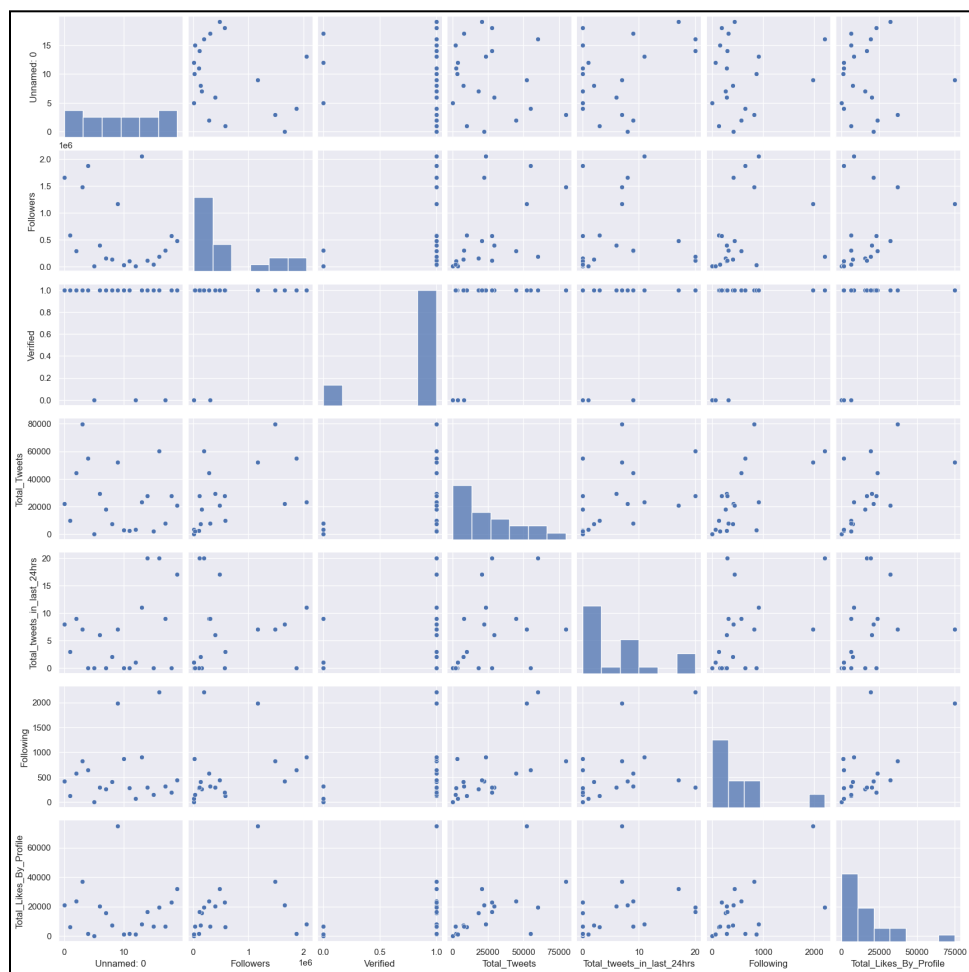


Table Creation

```
CREATE TABLE profile_data_twitter (
  User_ID varchar(300),
  Followers int(11),
  Created_At date,
  Verified int(11),
  Total_Tweets int(11),
  Total_tweets_24hrs int(11) ,
  category int(11),
  Withheld_In_Countries int(11),
  Following int(11),
  Total_Likes_By_Profile int(11),
  PRIMARY KEY ( User_ID )
);
```

```
CREATE TABLE profile_data_youtube_twitter (
  name varchar(300),
  handle varchar(300),
  url varchar(300),
  country varchar(300),
  platform int(11),
  PRIMARY KEY ( handle )
);
```

```
CREATE TABLE `tweets` (
  `Creation` varchar(300),
  `Likes` int,
  `Retweets` int,
  `Description` varchar(300),
```

```
`tweet_id` int(11) NOT NULL AUTO_INCREMENT,
PRIMARY KEY (`tweet_id`)
);
```

```
CREATE TABLE `youtube_profile_data` (
  handle varchar(300),
  subscribers int(11),
  joinedon date,
  total_views int(11),
  type int(11)
PRIMARY KEY (`handle`)
);
```

Normalisation

1st Normal Form

Table Name: youtube_profile_data

1. Eliminate duplicative columns from the same table.

- No duplicative columns found

Query 1 | youtube_profile_data | Limit to 1000 rows

1 • SELECT * FROM influencer_demo.youtube_profile_data;

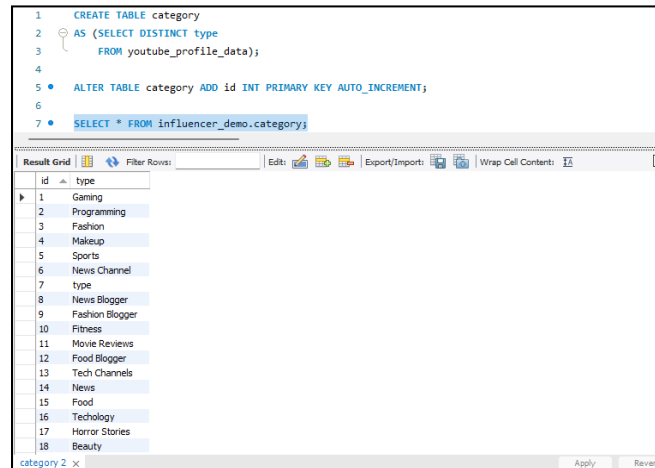
Result Grid | Filter Rows: | Export: | Wrap Cell Content: I A

handle	subscribers	otherLinks	joinedon	total views	type
CamodoGaming	2.99M	[https://www.youtube.com/redirect?event=ch...	2016-5-1	1,895,838,340	Gaming
MagicMusicGroup	4.62M	[https://www.youtube.com/redirect?event=ch...	2016-6-5	932,571,647	Gaming
UnspeakablePlays	7.33M	[https://www.youtube.com/redirect?event=ch...	2017-26-6	2,251,666,389	Gaming
GameToonsGaming	1.53M	[]	2021-22-2	485,638,386	Gaming
JKrewGaming	460K	[https://www.youtube.com/c/JKrew/featured,...	2020-29-9	216,117,057	Gaming
KCityGaming	1.32M	[https://www.youtube.com/redirect?event=ch...	2016-6-4	849,365,921	Gaming
UnderdogsGaming	186K	[]	2020-28-4	55,573,982	Gaming
RoyaltyGaming1	4.21M	[https://www.youtube.com/channel/UCja7QU...	2020-4-9	564,672,311	Gaming
MoreJStu	7.38M	[https://www.youtube.com/redirect?event=ch...	2011-25-12	2,084,680,075	Gaming
NotWethan	128K	[https://www.youtube.com/redirect?event=ch...	2016-21-9	0	Gaming
JivanTV	229K	[https://www.youtube.com/redirect?event=ch...	2013-28-4	24,114,252	Gaming

2. Create separate tables for each group of related data

Table Name: Youtube_Profile_Data

- The column type could be separated in a different table called category with primary key as id:



3. Identify each row with a unique column(primary key)

Each row is uniquely identified by column: handle

```
ALTER TABLE youtube_profile_data ADD PRIMARY KEY(handle);
```


Table Name: profile_data_youtube_twitter

1. Eliminate duplicative columns from the same table.

-No duplicative columns found, all columns are of the same type

1 • `SELECT * FROM influencer_demo.profile_data_youtube_twitter;`

Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

name	handle	url	country	platform
Assia Beauty	@AssiaBeauty	https://www.youtube.com/@AssiaBeauty	United Kingdom	YouTube
Ali-A	@alia	https://www.youtube.com/@alia	United Kingdom	YouTube
Camodo Gaming	@CamodoGaming	https://www.youtube.com/@CamodoGaming	United States	YouTube
Magic Music	@MagicMusicGroup	https://www.youtube.com/@MagicMusicGroup	United States	YouTube
UnspeakablePlays	@UnspeakablePlays	https://www.youtube.com/@UnspeakablePlays	United States	YouTube
GameToons Gaming	@GameToonsGaming	https://www.youtube.com/@GameToonsGaming	United States	YouTube
JKrew Gaming	@JKrewGaming	https://www.youtube.com/@JKrewGaming	United States	YouTube
K-City Gaming	@KCityGaming	https://www.youtube.com/@KCityGaming	United States	YouTube
JStu	@MoreJStu	https://www.youtube.com/@MoreJStu	United States	YouTube
Toasty Bros	@ToastyBros	https://www.youtube.com/@ToastyBros	United States	YouTube
MrBeast Gaming	@MrBeastGaming	https://www.youtube.com/@MrBeastGaming	United States	YouTube
GamingWithKev	@GamingWithKev	https://www.youtube.com/@GamingWithKev	United States	YouTube
Gaming With K	@GamingWithK	https://www.youtube.com/@GamingWithK	United States	YouTube
4K Gaming Nepal	@4KGamingNepal	https://www.youtube.com/@4KGamingNepal	United States	YouTube
freeCodeCamp.org	@freecodecamp	https://www.youtube.com/@freecodecamp	United States	YouTube
Aaron Jack	@AaronJack	https://www.youtube.com/@AaronJack	United States	YouTube
Chill Music Lab	@MusicLabChill	https://www.youtube.com/@MusicLabChill	United States	YouTube
CrashCourse	@crashcourse	https://www.youtube.com/@crashcourse	United States	YouTube

2. Create separate tables for each group of related data

For type Platform since can be grouped we have created a new table and inserted grouped data

```
CREATE TABLE `profile_platform` (  
  `platform_key` int(11) DEFAULT NULL,  
  `platform_type` varchar(30) DEFAULT NULL  
)
```

```
1 • INSERT INTO `influencer_demo`.`profile_platform`  
2   (`platform_key`,  
3    `platform_type`)  
4   VALUES  
5   (2,  
6    'Twitter');  
7 • SELECT * FROM influencer_demo.profile_platform;
```

< | Result Grid | Filter Rows: | Export: | Wrap Cell Content: |

	platform_key	platform_type
▶	1	YouTube
	2	Twitter

Table after changes:

1 • `SELECT * FROM influencer_demo.profile_data_youtube_twitter;`

name	handle	url	country	platform
Camodo Gaming	@CamodoGaming	https://www.youtube.com/@CamodoGaming	United States	1
Magic Music	@MagicMusicGroup	https://www.youtube.com/@MagicMusicGroup	United States	1
UnspeakablePlays	@UnspeakablePlays	https://www.youtube.com/@UnspeakablePlays	United States	1
GameToons Gaming	@GameToonsGaming	https://www.youtube.com/@GameToonsGaming	United States	1
JKrew Gaming	@JKrewGaming	https://www.youtube.com/@JKrewGaming	United States	1
K-City Gaming	@KCityGaming	https://www.youtube.com/@KCityGaming	United States	1
JStu	@MoreJStu	https://www.youtube.com/@MoreJStu	United States	1
Toasty Bros	@ToastyBros	https://www.youtube.com/@ToastyBros	United States	1
MrBeast Gaming	@MrBeastGaming	https://www.youtube.com/@MrBeastGaming	United States	1
GamingWithKev	@GamingWithKev	https://www.youtube.com/@GamingWithKev	United States	1
Gaming With K	@GamingWithK	https://www.youtube.com/@GamingWithK	United States	1
4K Gaming Nepal	@4KGamingNepal	https://www.youtube.com/@4KGamingNepal	United States	1
freeCodeCamp.org	@freecodecamp	https://www.youtube.com/@freecodecamp	United States	1

- Identify each row with a unique column(primary key) ALTER TABLE
`influencer_demo`.`profile_platform`
CHANGE COLUMN `platform_key` `platform_key` INT(11) NOT NULL ,
ADD PRIMARY KEY (`platform_key`);

Table Name: tweets

- Eliminate duplicative columns from the same table.

The table does not consist duplicate columns

1 • `SELECT * FROM influencer_demo.tweets;`

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [↗](#)


	User	Description	Likes	Retweets	Creation
▶	FoxNews	Connecticut police officer wounded in ambush honored by WWE superstars h...	44	6	2022-12-05 0
	FoxNews	"Cowboys embarrass Colts, tally eight touchdowns in blowout win https://t.c...	44	9	2022-12-05 0
	FoxNews	"Bengals' Joe Burrow continues success against Chiefs, snaps Kansas City's w...	32	7	2022-12-05 0
	FoxNews	"Mississippi man charged with burning cross to intimidate Black neighbors, ple...	90	24	2022-12-05 0
	FoxNews	SIC' EM: These are the four teams that will battle for the college football nati...	44	8	2022-12-05 0
	FoxNews	Ravens lose Lamar Jackson for game after knee injury https://t.co/M0w4Cck...	39	6	2022-12-05 0
	FoxNews	"Pro-life group christens, holds day of mourning for baby whose abortion was...	200	62	2022-12-05 0
	FoxNews	FOX Bet Super 6: Terry Bradshaw's 100K jackpot at stake in NFL Week 13 htt...	27	5	2022-12-05 0
	FoxNews	Netflix's "Crime Scene: The Texas Killing Fields": Could multiple mur...	78	17	2022-12-05 0
	FoxNews	Packers' Aaron Rodgers screams at head coach Matt LaFleur for not throwing...	52	10	2022-12-05 0
	FoxNews	George Clooney helps wife Amal steal the show at Kennedy Center where he'...	77	13	2022-12-05 0
	FoxNews	Netanyahu insists he won't step on 'landmine' question about Trump: 'Keep tr...	216	33	2022-12-05 0
	FoxNews	George Clooney says twins 'don't really care' he's a star at Kennedy Center H...	78	11	2022-12-05 0
	FoxNews	"Geno Smith orchestrates Seahawks' game-winning drive, hands Rams sixth s...	86	10	2022-12-05 0
	FoxNews	Raheem Sterling leaves World Cup due to armed burglary at London home htt...	107	18	2022-12-05 0
	FoxNews	Georgia child sex offender apprehended in Mexico after 5 years on the loose ...	508	79	2022-12-05 0
	FoxNews	"Fred McGriff elected to Baseball Hall of Fame: Barry Bonds, Roger Clemens. ...	280	42	2022-12-05 0

2. Create separate tables for each group of related data

No separate tables needed as the table does not comprise of group data

3. Identify each row with a unique column(primary key)

Since Tweets are multiple we cannot have user as primary key, hence we created a new primary key called tweet_id


 Table Name: Schema: **influencer_demo**
 Charset/Collation: Engine:
 Comments:

Column Name	Datatype	PK	NN	UQ	B	UN	ZF	AI	G	Default/Expression
◇ User	VARCHAR(300)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
◇ Creation	VARCHAR(300)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
◇ Likes	VARCHAR(300)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
◇ Retweets	VARCHAR(300)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
◇ Description	VARCHAR(300)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	NULL
💡 tweet_id	INT(11)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	

1 • `SELECT * FROM influencer_demo.tweets;`

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

		Likes	Retweets	Description	tweet_id
▶	05 08:10:02+0...	44	6	Connecticut police officer wounded in ambush honored by WWE superstars h...	1
	05 08:00:01+0...	44	9	"Cowboys embarrass Colts, tally eight touchdowns in blowout win https://t.c...	2
	05 07:50:03+0...	32	7	"Bengals' Joe Burrow continues success against Chiefs, snaps Kansas City's w...	3
	05 07:40:07+0...	90	24	"Mississippi man charged with burning cross to intimidate Black neighbors, ple...	4
	05 07:30:14+0...	44	8	SIC' EM: These are the four teams that will battle for the college football nati...	5
	05 07:20:03+0...	39	6	Ravens lose Lamar Jackson for game after knee injury https://t.co/M0w4Cck...	6
	05 07:10:06+0...	200	62	"Pro-life group christens, holds day of mourning for baby whose abortion was...	7
	05 06:59:02+0...	27	5	FOX Bet Super 6: Terry Bradshaw's 100K jackpot at stake in NFL Week 13 htt...	8
	05 06:49:04+0...	78	17	Netflix's "Crime Scene: The Texas Killing Fields": Could multiple mur...	9
	05 06:40:01+0...	52	10	Packers' Aaron Rodgers screams at head coach Matt LaFleur for not throwing...	10
	05 06:33:12+0...	77	13	George Clooney helps wife Amal steal the show at Kennedy Center where he'...	11
	05 06:00:14+0...	216	33	Netanyahu insists he won't step on 'landmine' question about Trump: 'Keep tr...	12
	05 05:54:14+0...	78	11	George Clooney says twins 'don't really care' he's a star at Kennedy Center H...	13

Table Name: profile_data_twitter

1. Eliminate duplicative columns from the same table.

-No duplicative columns found, all columns are of the same type

1 • `SELECT * FROM influencer_demo.profile_data_twitter;`

	User_ID	Followers	Created_At	Verified	Total_Tweets	Total_tweets_24hrs
▶	FoxNews	23291881	2007-03-17 19:01:26+00:00	1	512678	20
	CodeWisdom	278305	2011-10-22 23:50:32+00:00	0	4067	0
	kevincorke	176269	2009-04-21 13:51:33+00:00	1	12772	0
	johnrobertsFox	271217	2009-02-16 10:19:01+00:00	1	8479	1
	DavidRutz	19216	2009-07-14 14:05:52+00:00	1	1833	8
	MollyLineFNC	29776	2011-01-03 19:07:25+00:00	1	7561	0
	FoxFilesFNC	5762	2012-08-01 20:53:13+00:00	1	21145	0
	foxnewsdesk	9709	2014-05-06 18:48:29+00:00	1	6101	0
	JakeBGibson	10401	2009-04-22 17:09:33+00:00	1	1678	0
	tamaragitt	6642	2009-07-24 23:23:35+00:00	1	7947	0
	AdamKlotzFNC	6138	2009-02-26 04:36:56+00:00	1	5050	2
	WillRicci	18661	2009-04-23 18:53:44+00:00	1	24637	0
	RichEdsonDC	9907	2010-11-09 19:04:56+00:00	1	6866	0
	ToddPiro	25776	2009-04-19 04:14:17+00:00	1	4641	0
	RzstProgramming	46730	2017-06-07 21:17:59+00:00	0	32867	0
	BillFOXLA	227161	2013-01-08 18:03:50+00:00	1	17991	0

2. The column category could be separated in a different table called category with primary key as id:

before:

```
1 • SELECT * FROM Influencer_demo.twitter_profile;
```

	Total_Tweets	Total_tweets_in_last_24hrs	category	Withheld_In_Countries	Following	Total_Likes_By_Profile
512678	20	0	Programming	0	262	4
4067	0	0	Programming	0	2300	2033
12772	0	0	Programming	0	62067	290488
8479	1	0	Programming	0	504	105
1833	8	0	Programming	0	2138	4154
7561	0	0	Programming	0	1695	1492
21145	0	0	Programming	0	975	18285
6101	0	0	Programming	0	5417	115
1678	0	0	Programming	0	1655	382
7947	0	0	Programming	0	492	2035
5050	2	0	Programming	0	1058	8447
24637	0	0	Programming	0	3989	22890
6866	0	0	Programming	0	2328	3
4641	0	0	Programming	0	652	14356
32867	0	0	Programming	0	171	14942
17991	0	0	Programming	0	2504	7686
18116	5	0	Programming	0	1315	21576

Creation of category as in table 2

```
1 CREATE TABLE category
2 AS (SELECT DISTINCT type
3 FROM youtube_profile_data);
4
5 • ALTER TABLE category ADD id INT PRIMARY KEY AUTO_INCREMENT;
6
7 • SELECT * FROM influencer_demo.category;
```

id	type
1	Gaming
2	Programming
3	Fashion
4	Makeup
5	Sports
6	News Channel
7	type
8	News Blogger
9	Fashion Blogger
10	Fitness
11	Movie Reviews
12	Food Blogger
13	Tech Channels
14	News
15	Food
16	Technology
17	Horror Stories
18	Beauty

After updation based on table

```
1 • SELECT * FROM Influencer_demo.twitter_profile;
```

yr_ID	Followers	Created_At	Verified	Total_Tweets	Total_tweets_in_last_24hrs	category
News	23291881	2007-03-17 19:01:26+00:00	TRUE	512678	20	2
eWisdom	278305	2011-10-22 23:50:32+00:00	FALSE	4067	0	2
ncorke	176269	2009-04-21 13:51:33+00:00	TRUE	12772	0	2
robertsFox	271217	2009-02-16 10:19:01+00:00	TRUE	8479	1	2
idRutz	19216	2009-07-14 14:05:52+00:00	TRUE	1833	8	2
yLineFNC	29776	2011-01-03 19:07:25+00:00	TRUE	7561	0	2
FileFNC	5762	2012-08-01 20:53:13+00:00	TRUE	21145	0	2
newsdesk	9709	2014-05-06 18:48:29+00:00	TRUE	6101	0	2
BGibson	10401	2009-04-22 17:09:33+00:00	TRUE	1678	0	2
bragitt	6642	2009-07-24 23:23:35+00:00	TRUE	7947	0	2
mklotzFNC	6138	2009-02-26 04:36:56+00:00	TRUE	5050	2	2
llico	18661	2009-04-23 18:53:44+00:00	TRUE	24637	0	2
EdsonDC	9907	2010-11-09 19:04:56+00:00	TRUE	6866	0	2
Spino	25776	2009-04-19 04:14:17+00:00	TRUE	TRUE	0	2
Programming	46730	2017-06-07 21:17:59+00:00	FALSE	32867	0	2
DILA	227161	2013-01-08 18:03:50+00:00	TRUE	17991	0	2
l_Caythorne	7921	2016-07-25 11:54:09+00:00	TRUE	1315	5	?

3. Identify each row with a unique column(primary key)

```
ALTER TABLE `influencer_demo`.`twitter_profile`  
CHANGE COLUMN `User_ID` `User_ID` VARCHAR(300) NOT NULL ,  
ADD PRIMARY KEY (`User_ID`);  
;
```

User_ID is the primary key and uniquely identifies data in table

Tables Added in 1NF: Category, Profile platform

2nd Normal Form

Table Name: Youtube_Profile_Data

1. Meet all requirements of 1NF

-We have demonstrated that all requirements of 1NF are met in the above screenshots.

2. Remove subsets of data that apply to multiple rows of a table and place them in separate tables.

-We did not have duplicative columns of data to begin with, hence not required to separate the subsets of the data to a different table.

For 'type' column we identified multiple rows of duplicate data hence separated that into a table 'category' with an additional auto-increment primary key 'id'.

3. Create relationships between these new tables and their predecessors through use of foreign keys:

-Updated the type column in the Youtube_Profile_data table with the foreign key:

	i»handle	subscribers	otherLinks	joinedon	total views	type
	TechyPathshala	606K	[https://www.youtube.com/channel/UC9Cowa...	2017-16-2	165,727,011	1
	QAYZERGAMING	104K	[https://www.youtube.com/channel/UCfXoodv...	2018-27-11	50,137,826	1
	RAHULGAMING920	1.6M	[https://www.youtube.com/redirect?event=ch...	2014-25-10	166,347,105	1
	4KGamingNepal	1.04M	[https://www.youtube.com/redirect?event=ch...	2017-18-10	84,339,506	1
	MiroHQ	24.7K	[https://www.youtube.com/redirect?event=ch...	2012-12-9	0	2
	2withmosh	2.95M	[https://www.youtube.com/redirect?event=ch...	2014-6-10	142,329,938	2
	freecodecamp	6.81M	[https://www.youtube.com/redirect?event=ch...	2014-16-12	465,384,906	2
	AaronJack	338K	[https://www.youtube.com/redirect?event=ch...	2018-9-9	21,085,075	2
	MusicLabChill	782K	[https://www.youtube.com/redirect?event=ch...	2019-5-7	93,156,733	2
	TEDxTalks	37M	[https://www.youtube.com/redirect?event=ch...	2009-23-6	0	2
	crashcourse	14.2M	[https://www.youtube.com/redirect?event=ch...	2006-19-5	1,750,944,033	2
	namanhkapur	136K	[https://www.youtube.com/redirect?event=ch...	2012-14-9	7,507,520	2
	CroatCode	513K	[https://www.youtube.com/redirect?event=ch...	2012-17-6	52,967,719	2

Table Name: profile_data_youtube_twitter

1. Meet all requirements of 1NF

-We have demonstrated that all requirements of 1NF are met in the above screenshots.

2. Remove subsets of data that apply to multiple rows of a table and place them in separate tables.

-We did not have duplicative columns of data to begin with, hence not required to separate the subsets of the data to a different table.

For 'type' column we identified multiple rows of duplicate data hence separated that into a table 'category' with an additional auto-increment primary key 'id'.

3. Create relationships between these new tables and their predecessors through use of foreign keys:

```
ALTER TABLE `influencer_demo`.`profile_platform`
CHANGE COLUMN `platform_key` `platform_key` INT(11) NOT NULL ,
ADD PRIMARY KEY (`platform_key`);
;
```

Foreign key relation developed with platform

name	handle	url	country	platform
Camodo Gaming	@CamodoGaming	https://www.youtube.com/@CamodoGaming	United States	1
Magic Music	@MagicMusicGroup	https://www.youtube.com/@MagicMusicGroup	United States	1
UnspeakablePlays	@UnspeakablePlays	https://www.youtube.com/@UnspeakablePlays	United States	1
GameToons Gaming	@GameToonsGaming	https://www.youtube.com/@GameToonsGaming	United States	1
JKrew Gaming	@JKrewGaming	https://www.youtube.com/@JKrewGaming	United States	1
K-City Gaming	@KCityGaming	https://www.youtube.com/@KCityGaming	United States	1
Underdogs Gaming	@UnderdogsGaming	https://www.youtube.com/@UnderdogsGaming	Australia	1
Royalty Gaming	@RoyaltyGaming1	https://www.youtube.com/@RoyaltyGaming1		1
JStu	@MoreJStu	https://www.youtube.com/@MoreJStu	United States	1
Wethan	@NotWethan	https://www.youtube.com/@NotWethan		1
JivanTV	@JivanTV	https://www.youtube.com/@JivanTV		1
Toasty Bros	@ToastyBros	https://www.youtube.com/@ToastyBros	United States	1
Norris Nuts Gaming	@NorrisNutsGaming	https://www.youtube.com/@NorrisNutsGaming	Australia	1
MrBeast Gaming	@MrBeastGaming	https://www.youtube.com/@MrBeastGaming	United States	1
WhatCulture Gaming	@WhatCultureGaming	https://www.youtube.com/@WhatCultureGaming		1
GamingWithKev	@GamingWithKev	https://www.youtube.com/@GamingWithKev	United States	1

Table Name: tweets

1. Meet all requirements of 1NF

-We have demonstrated that all requirements of 1NF are met in the above screenshots.

2. Remove subsets of data that apply to multiple rows of a table and place them in separate tables.

Breaking tweet_id and user

Before:

User	Creation	Likes	Retweets	Description	tweet_id
FoxNews	2022-12-05 08:10:02+00:00	44	6	Connecticut police officer wounded in ambush honored by WWE superstars h...	1
FoxNews	2022-12-05 08:00:01+00:00	44	9	'Cowboys embarrass Colts, tally eight touchdowns in blowout win https://t.c...	2
FoxNews	2022-12-05 07:50:03+00:00	32	7	'Bengals' Joe Burrow continues success against Chiefs, snaps Kansas City's w...	3
FoxNews	2022-12-05 07:40:07+00:00	90	24	'Mississippi man charged with burning cross to intimidate Black neighbors, ple...	4
FoxNews	2022-12-05 07:30:14+00:00	44	8	SIC EM: These are the four teams that will battle for the college football nat...	5
FoxNews	2022-12-05 07:20:03+00:00	39	6	Ravens lose Lamar Jackson for game after knee injury https://t.co/M0w4Cck...	6
FoxNews	2022-12-05 07:10:06+00:00	200	62	'Pro-life group christens, holds day of mourning for baby whose abortion was...	7
FoxNews	2022-12-05 06:59:02+00:00	27	5	FOX Bet Super 6: Terry Bradshaw's 100K jackpot at stake in NFL Week 13 htt...	8
FoxNews	2022-12-05 06:49:04+00:00	78	17	Netflixâ€™s â€œCrime Scene: The Texas Killing Fieldsâ€œ: Could multiple mur...	9
FoxNews	2022-12-05 06:40:01+00:00	52	10	Packers' Aaron Rodgers screams at head coach Matt LaFleur for not throwing...	10
FoxNews	2022-12-05 06:33:12+00:00	77	13	George Clooney helps wife Amal steal the show at Kennedy Center where he'...	11
FoxNews	2022-12-05 06:00:14+00:00	216	33	Netanyahu insists he won't step on 'landmine' question about Trump: 'Keep tr...	12
FoxNews	2022-12-05 05:54:14+00:00	78	11	George Clooney says twins 'don't really care' he's a star at Kennedy Center H...	13
FoxNews	2022-12-05 05:26:47+00:00	86	10	'Geno Smith orchestrates Seahawks' game-winning drive, hands Rams sixth s...	14
FoxNews	2022-12-05 04:53:56+00:00	107	18	Raheem Sterling leaves World Cup due to armed burglary at London home htt...	15
FoxNews	2022-12-05 04:41:09+00:00	508	79	Georgia child sex offender apprehended in Mexico after 5 years on the loose ...	16
FoxNews	2022-12-05 04:27:26+00:00	280	42	'Fred McGriff elected to Baseball Hall of Fame; Barry Bonds, Roger Clemens, ...	17
FoxNews	2022-12-05 03:58:39+00:00	7703	779	Elon Musk says Constitution is greater than any president https://t.co/urhbn...	18




Dropping user from the table

Creation	Likes	Retweets	Description	tweet_id
2022-12-05 08:10:02+00:00	44	6	Connecticut police officer wounded in ambush honored by WWE superstars h...	1
2022-12-05 08:00:01+00:00	44	9	'Cowboys embarrass Colts, tally eight touchdowns in blowout win https://t.c...	2
2022-12-05 07:50:03+00:00	32	7	'Bengals' Joe Burrow continues success against Chiefs, snaps Kansas City's w...	3
2022-12-05 07:40:07+00:00	90	24	'Mississippi man charged with burning cross to intimidate Black neighbors, ple...	4
2022-12-05 07:30:14+00:00	44	8	SIC EM: These are the four teams that will battle for the college football nat...	5
2022-12-05 07:20:03+00:00	39	6	Ravens lose Lamar Jackson for game after knee injury https://t.co/M0w4Cck...	6
2022-12-05 07:10:06+00:00	200	62	'Pro-life group christens, holds day of mourning for baby whose abortion was...	7
2022-12-05 06:59:02+00:00	27	5	FOX Bet Super 6: Terry Bradshaw's 100K jackpot at stake in NFL Week 13 htt...	8
2022-12-05 06:49:04+00:00	78	17	Netflixâ€™s â€œCrime Scene: The Texas Killing Fieldsâ€œ: Could multiple mur...	9
2022-12-05 06:40:01+00:00	52	10	Packers' Aaron Rodgers screams at head coach Matt LaFleur for not throwing...	10
2022-12-05 06:33:12+00:00	77	13	George Clooney helps wife Amal steal the show at Kennedy Center where he'...	11
2022-12-05 06:00:14+00:00	216	33	Netanyahu insists he won't step on 'landmine' question about Trump: 'Keep tr...	12
2022-12-05 05:54:14+00:00	78	11	George Clooney says twins 'don't really care' he's a star at Kennedy Center H...	13
2022-12-05 05:26:47+00:00	86	10	'Geno Smith orchestrates Seahawks' game-winning drive, hands Rams sixth s...	14
2022-12-05 04:53:56+00:00	107	18	Raheem Sterling leaves World Cup due to armed burglary at London home htt...	15
2022-12-05 04:41:09+00:00	508	79	Georgia child sex offender apprehended in Mexico after 5 years on the loose ...	16
2022-12-05 04:27:26+00:00	280	42	'Fred McGriff elected to Baseball Hall of Fame; Barry Bonds, Roger Clemens, ...	17
2022-12-05 03:58:39+00:00	7703	779	Elon Musk says Constitution is greater than any president https://t.co/urhbn...	18

And breaking it into another to link tweet id and user

1 • `SELECT * FROM influencer_demo.tweet_id;`

<

Result Grid   Filter Rows: | Export:  | Wrap Cell Content: [I/A](#)

UserID	ID
FoxNews	1
FoxNews	2
FoxNews	3
FoxNews	4
FoxNews	5
FoxNews	6
FoxNews	7
FoxNews	8
FoxNews	9
FoxNews	10
FoxNews	11
FoxNews	12
FoxNews	13
FoxNews	14
FoxNews	15
FoxNews	16
FoxNews	17
FoxNews	18

tweet_id 1 x

Table Name: twitter profile

1. Meet all requirements of 1NF

-We have demonstrated that all requirements of 1NF are met in the above screenshots.

2. Remove subsets of data that apply to multiple rows of a table and place them in separate tables.

-We did not have duplicative columns of data, hence not required to separate the subsets of the data to a different table.

3rd Normal Form

Table Name: Youtube_Profile_Data

1. Meet all requirements of 2NF

-We have demonstrated that all requirements of 1NF are met in the above screenshots

2. Remove the columns that are not dependent upon the primary key

- We do not have any columns dependent upon the primary key

Table Name: profile_data_youtube_twitter

1. Meet all requirements of 2NF

-We have demonstrated that all requirements of 1NF are met in the above screenshots

2. Remove the columns that are not dependent upon the primary key

- We do not have any columns dependent upon the primary key

Table Name: tweets

1. Meet all requirements of 2NF

-We have demonstrated that all requirements of 1NF are met in the above screenshots

2. Remove the columns that are not dependent upon the primary key

- We do not have any columns dependent upon the primary key

Table Name: twitter profile

1. Meet all requirements of 2NF

-We have demonstrated that all requirements of 1NF are met in the above screenshots

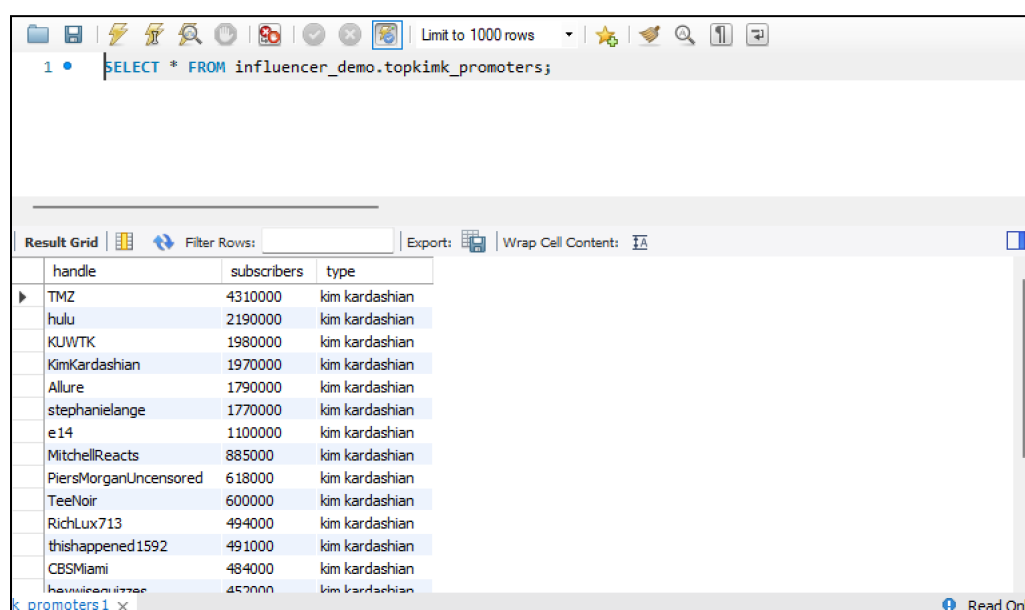
2. Remove the columns that are not dependent upon the primary key

- We do not have any columns dependent upon the primary key

Use Cases

1) Top Youtube influencers with highest number of subscribers of a particular category (Kim Kardashian)

```
CREATE VIEW TopKimK_promoters AS
SELECT a.handle, a.subscribers, b.type
FROM youtube_profile_data AS a inner join category AS b
ON a.type = b.id
WHERE b.type = "Kim Kardashian"
ORDER BY a.subscribers DESC
```



1 • SELECT * FROM influencer_demo.topkimk_promoters;

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [F1](#)

handle	subscribers	type
TMZ	4310000	kim kardashian
hulu	2190000	kim kardashian
KUWTK	1980000	kim kardashian
KimKardashian	1970000	kim kardashian
Allure	1790000	kim kardashian
stephanielange	1770000	kim kardashian
e14	1100000	kim kardashian
MitchellReacts	885000	kim kardashian
PiersMorganUncensored	618000	kim kardashian
TeeNoir	600000	kim kardashian
RichLux713	494000	kim kardashian
thishappened1592	491000	kim kardashian
CBSMiami	484000	kim kardashian
hawaiiianidm	457000	kim kardashian

k_promoters1 x Read Only

2) Top Youtube influencers with highest number of subscribers of a particular category (Gaming) for a particular country (United States)

```
CREATE VIEW TopGamers_US AS
SELECT a.handle, a.subscribers, b.type, c.country
FROM youtube_profile_data AS a inner join category AS b
ON a.type = b.id
inner join profile_data_youtube_twitter AS c
ON a.handle = c.handle
WHERE b.type = "Gaming" AND c.country = "United States"
ORDER BY a.subscribers DESC
```


1 • `SELECT * FROM influencer_demo.topgamers_us;`

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [A](#)

handle	subscribers	type	country
MoreJStu	7380000	Gaming	United States
MagicMusicGroup	4620000	Gaming	United States
RoyaltyGaming1	4210000	Gaming	United States
CamodoGaming	2990000	Gaming	United States
WhatCultureGaming	1720000	Gaming	United States
GameToonsGaming	1530000	Gaming	United States
KCityGaming	1320000	Gaming	United States
4KGamingNepal	1040000	Gaming	United States
ToastyBros	483000	Gaming	United States
GamingWithK	478000	Gaming	United States
JKrewGaming	460000	Gaming	United States
JivanTV	229000	Gaming	United States
NotWethan	128000	Gaming	United States

3) Top Twitter influencer with verified profiles based on the selected category (Gaming)

```
CREATE VIEW TopVerifiedGamers_Twitter AS
SELECT a.User_ID, a.followers, a.verified, b.type
FROM profile_data_twitter AS a inner join category AS b
ON a.category = b.id
WHERE b.type = "Gaming" AND a.Verified = "1"
ORDER BY a.followers DESC
```

1 • `SELECT * FROM influencer_demo.topverifiedgamers_twitter;`

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [A](#)

User_ID	followers	verified	type
YouTubeGaming	2058106	1	Gaming
MLG	1872999	1	Gaming
CORSAIR	1660036	1	Gaming
ScufGaming	1480792	1	Gaming
ASTROGaming	1168197	1	Gaming
primegaming	578811	1	Gaming
Envy	573184	1	Gaming
Luminosity	478278	1	Gaming
dgaming	399781	1	Gaming
Complexity	292522	1	Gaming
GreenManGaming	191853	1	Gaming
GambitEsports	155583	1	Gaming
FacebookGaming	135273	1	Gaming
RTGCLAN	115401	1	Gaming

Read Only

4) Top Twitter influencers Tweets with average retweets of more than 50 based on selected category (News)

```

SELECT a.User_ID, a.followers, d.description, d.Retweets, c.type
FROM profile_data_twitter AS a inner join tweet_id AS b
ON a.User_ID = b.UserID
inner join tweets as d
ON b.ID = d.tweet_id
inner join category as c
on a.category = c.id
WHERE d.Retweets > 50 AND c.type = "News"
ORDER BY a.followers DESC

```

Result Grid					
		Filter Rows:		Export:	Wrap Cell Content:
	User_ID	followers	description	Retweets	type
▶	FoxNews	23301220	"Pro-life group christens, holds day of mourning...	62	News
	FoxNews	23301220	Georgia child sex offender apprehended in Mexi...	79	News
	FoxNews	23301220	Elon Musk says Constitution is greater than any...	729	News
	FoxNews	23301220	"Indonesian volcano erupts, releasing river of la...	51	News

5) Name and handle of the YouTuber with the highest view count for selected category (Fashion)

```

CREATE VIEW HighestViewsFashion_Youtubers AS
SELECT a.name, a.handle, max(b.total_views) AS max_views, c.type
FROM profile_data_youtube_twitter AS a
INNER JOIN youtube_profile_data AS b
ON a.handle = b.handle
INNER JOIN category AS c
ON b.type = c.id
WHERE c.type = "Fashion"
GROUP BY a.name, a.handle
ORDER BY max(b.total_views) DESC

```

1 • `SELECT * FROM influencer_demo.highestviewsfashion_youtubers;`

	name	handle	max_views	type
▶	WOWPresents	WOWPresents	1680512430	Fashion
	Architectural Digest	Archdigest	1061127763	Fashion
	David Bowie	davidbowie	893390856	Fashion
	Unexplained Mysteries	UnexplainedMysteriesOfficial	522304869	Fashion
	WGN News	wgntv	415802016	Fashion
	Shea Whitney	SheaWhitney	163654415	Fashion
	HauteLeMode	hautelemode	145054381	Fashion
	Ankhal	Ankhal	131842680	Fashion
	Versace	Versace	115271119	Fashion
	Steve Gaskins92	stevegaskins92	44474526	Fashion

6) Name and Country of the top YouTuber with the highest amount of subscribers for selected category (Fitness)

```

CREATE VIEW TopCountryFitness_Youtuber AS
SELECT a.country, a.handle, MAX(b.subscribers) as subscribers, c.type
FROM profile_data_youtube_twitter AS a
INNER JOIN youtube_profile_data AS b
ON a.handle = b.handle
INNER JOIN category AS c
ON b.type = c.id
WHERE c.type = "Fitness"
GROUP BY a.handle,a.country
ORDER BY MAX(b.subscribers) DESC

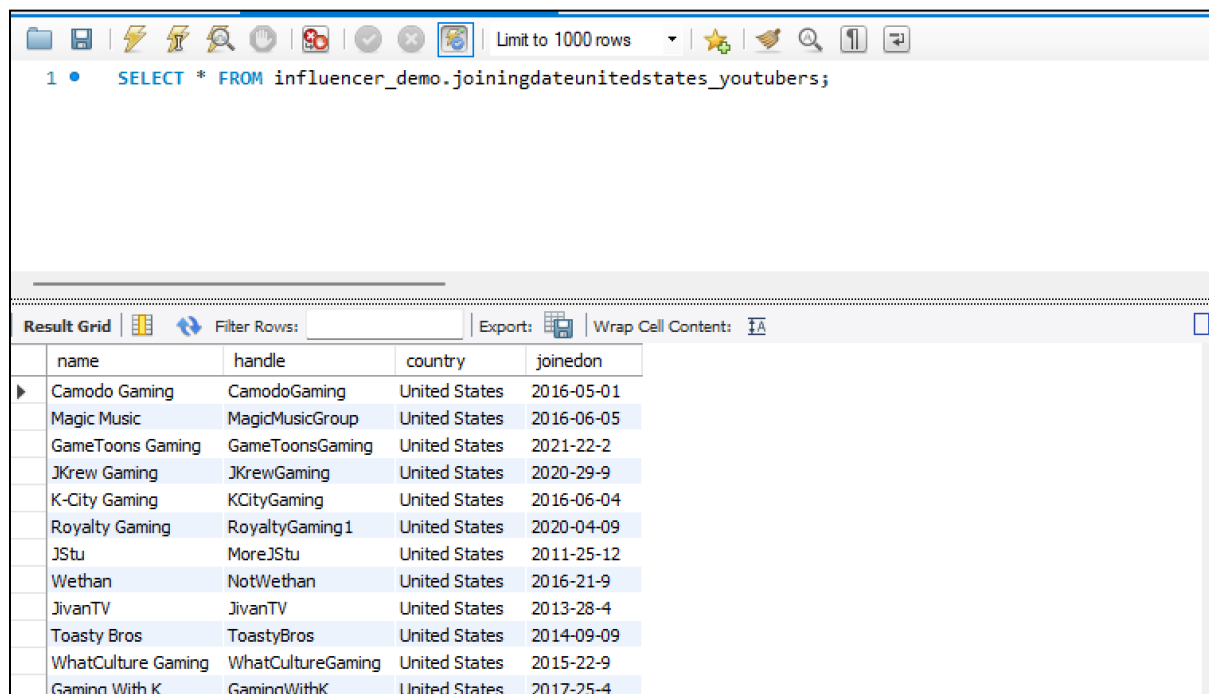
```

1 • `SELECT * FROM influencer_demo.topcountryfitness_youtuber;`

	country	handle	subscribers	type
▶	United States	TEDxTalks	37000000	Fitness
	United States	DoctorMike	10400000	Fitness
	Canada	MadFit	7520000	Fitness
	United Kingdom	RobertasGym	4980000	Fitness
	India	beerbiceps	4600000	Fitness
	United States	LeslieSansonesWalkatHome	4230000	Fitness
	United States	SunnyV2	2920000	Fitness
	United States	InsiderScience	2390000	Fitness
	United States	hubermanlab	1870000	Fitness
	United States	DingProductions	1440000	Fitness
	United States	self	1220000	Fitness
	United States	Sean_Nalewanyj	859000	Fitness
	United States	earthfedmuscle1560	760000	Fitness
	United States	thelleanbeefatty	695000	Fitness

7) Join date of the recommended YouTuber based on country (United States).

```
CREATE VIEW JoiningDateUnitedStates_Youtubers AS
SELECT a.name, a.handle, a.country, b.joinendon
FROM profile_data_youtube_twitter AS a
INNER JOIN youtube_profile_data AS b
ON a.handle = b.handle
WHERE a.country = 'United States'
GROUP BY a.name, a.handle, b.joinendon
```

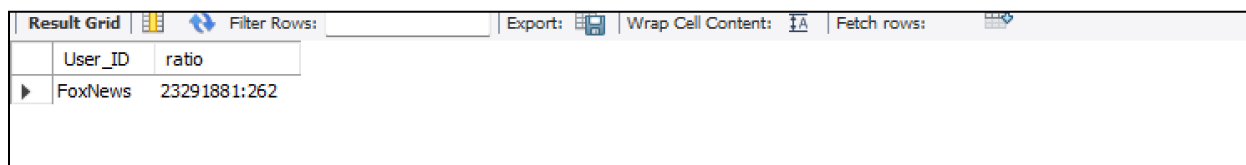


The screenshot shows a database query interface. At the top, a SQL query is entered: `SELECT * FROM influencer_demo.joiningdateunitedstates_youtubers;`. Below the query, the results are displayed in a table with the following columns: name, handle, country, and joinendon. The results list 14 YouTube influencers from the United States, including Camodo Gaming, Magic Music, GameToons Gaming, JKrew Gaming, K-City Gaming, Royalty Gaming, JStu, Wethan, JivanTV, Toasty Bros, WhatCulture Gaming, and Gaming With K, along with their join dates.

name	handle	country	joinendon
Camodo Gaming	CamodoGaming	United States	2016-05-01
Magic Music	MagicMusicGroup	United States	2016-06-05
GameToons Gaming	GameToonsGaming	United States	2021-22-2
JKrew Gaming	JKrewGaming	United States	2020-29-9
K-City Gaming	KCityGaming	United States	2016-06-04
Royalty Gaming	RoyaltyGaming1	United States	2020-04-09
JStu	MoreJStu	United States	2011-25-12
Wethan	NotWethan	United States	2016-21-9
JivanTV	JivanTV	United States	2013-28-4
Toasty Bros	ToastyBros	United States	2014-09-09
WhatCulture Gaming	WhatCultureGaming	United States	2015-22-9
Gaming With K	GamingWithK	United States	2017-25-4

8) Get the followers to following ratio of the particular Twitter influencer

```
SELECT a.User_ID, CONCAT(a.Followers,':',a.Following) AS ratio
FROM profile_data_twitter AS a
WHERE a.User_ID = "FoxNews"
ORDER BY a.User_ID
```



The screenshot shows a database query interface. At the top, a SQL query is entered: `SELECT a.User_ID, CONCAT(a.Followers,':',a.Following) AS ratio FROM profile_data_twitter AS a WHERE a.User_ID = "FoxNews" ORDER BY a.User_ID`. Below the query, the results are displayed in a table with the following columns: User_ID and ratio. The results show the ratio for FoxNews as 23291881:262.

User_ID	ratio
FoxNews	23291881:262

9) Get the website URL of Top Youtubers for a specified category (Fitness)

```
CREATE VIEW TopFitnessYoutuber_URL AS
SELECT a.handle, a.url, b.subscribers, c.type
FROM profile_data_youtube_twitter AS a
INNER JOIN youtube_profile_data AS b
on a.handle = b.handle
INNER JOIN category AS c
on b.type = c.id
WHERE c.type = "Fitness"
GROUP BY a.handle, a.url, b.subscribers
HAVING MAX(b.subscribers)
ORDER BY MAX(b.subscribers) DESC;
```

1 • `SELECT * FROM influencer_demo.topfitnessyoutuber_url;`

handle	url	subscribers	type
TEDxTalks	https://www.youtube.com/@TEDxTalks	37000000	Fitness
DoctorMike	https://www.youtube.com/@DoctorMike	10400000	Fitness
MadFit	https://www.youtube.com/@MadFit	7520000	Fitness
RobertasGym	https://www.youtube.com/@RobertasGym	4980000	Fitness
beerbiceps	https://www.youtube.com/@beerbiceps	4600000	Fitness
LeslieSansonesWalkatHome	https://www.youtube.com/@LeslieSansonesWa...	4230000	Fitness
SunnyV2	https://www.youtube.com/@SunnyV2	2920000	Fitness
InsiderScience	https://www.youtube.com/@InsiderScience	2390000	Fitness
hubermanlab	https://www.youtube.com/@hubermanlab	1870000	Fitness
DingProductions	https://www.youtube.com/@DingProductions	1440000	Fitness
self	https://www.youtube.com/@self	1220000	Fitness
Sean Nalewanvi	https://www.youtube.com/@Sean Nalewanvi	859000	Fitness

10) To check if the Twitter influencer's profile is withheld from any country

```
SELECT User_ID, Withheld_In_Countries
FROM profile_data_twitter
WHERE User_ID = "FoxNews";
```

User_ID	Withheld_In_Countries
FoxNews	0

11) Get top influencers who have had a Twitter presence for more than 5 years in a selected category

```
CREATE VIEW TwitterInfluencers_5Plusyears AS
SELECT User_ID, DATE_FORMAT(FROM_DAYS(DATEDIFF(NOW(),Created_At)), '%Y') + 0 AS 'years'
FROM profile_data_twitter
WHERE DATE_FORMAT(FROM_DAYS(DATEDIFF(NOW(),`Created_At`)), '%Y') + 0 >5;
```

1 • `SELECT * FROM influencer_demo.twitterinfluencers_5plusyears;`

User_ID	years
FoxNews	15
CodeWisdom	11
kevincorke	13
johnrobertsFox	13
DavidRutz	13
MollyLineFNC	11
FoxFilesFNC	10
foxnewsdesk	8
JakeBGibson	13
tamaragitt	13
AdamKlotzFNC	13
WillRicci	13

12) Types of Youtubers sorted by countries

```
CREATE VIEW youtuberTypes_country AS
SELECT a.name, a.handle, a.country, c.type from
profile_data_youtube_twitter AS a INNER JOIN youtube_profile_data AS b
ON a.handle = b.handle
INNER JOIN category AS c
ON b.type = c.id
```

1 • `SELECT * FROM influencer_demo.youtubetypes_country;`

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

	name	handle	country	type
▶	Camodo Gaming	CamodoGaming	United States	Gaming
	Magic Music	MagicMusicGroup	United States	Gaming
	GameToons Gaming	GameToonsGaming	United States	Gaming
	JKrew Gaming	JKrewGaming	United States	Gaming
	K-City Gaming	KCityGaming	United States	Gaming
	Underdogs Gaming	UnderdogsGaming	Australia	Gaming
	Royalty Gaming	RoyaltyGaming1	United States	Gaming
	JStu	MoreJStu	United States	Gaming
	Wethan	NotWethan	United States	Gaming
	JivanTV	JivanTV	United States	Gaming
	Toasty Bros	ToastyBros	United States	Gaming
	Norris Nuts Gaming	NorrisNutsGaming	Australia	Gaming

13) Top Twitter influencers Tweets with likes more than 1000 based on selected category (News)

Create view CategoryLikesMoreThan AS

SELECT a.User_ID, a.followers, d.description, d.likes, c.type

FROM profile_data_twitter AS a inner join tweet_id AS b

ON a.User_ID = b.UserID

inner join tweets as d

ON b.ID = d.tweet_id

inner join category as c

on a.category = c.id

WHERE d.likes > 1000 AND c.type = "News"

ORDER BY a.followers DESC

1 • `select * from CategoryLikesMoreThan`

<

Result Grid Filter Rows: Export: Wrap Cell Content:

	User_ID	followers	description	likes	type
▶	FoxNews	23301220	Elon Musk says Constitution is greater than any...	7703	News
	NBCNews	9340722	BREAKING 8YŠ": Cristiano Ronaldo is yet to agr...	2423	News
	NBCNews	9340722	"England BEAT Pakistan by 74 runs in the first T...	1407	News
	ABSCBNNews	8832356	FULL-TIME: Brazil with a statement victory over...	1572	News
	ABSCBNNews	8832356	8YŠ" Manchester United are out of the race to ...	2611	News
	ABSCBNNews	8832356	"Al-Nassr have made the best offer so far to sig...	1251	News
	ABSCBNNews	8832356	""Thank you for your compliment	3979	News
	ABSCBNNews	8832356	"Merseyside derby 8š"i,	1347	News
	gmanews	6451324	"Prince Harry speaks of ""pain and suffering"" o...	1836	News
	gmanews	6451324	England beat Senegal 3-0 setting up World Cup ...	1257	News

14) Top Twitter Influencers with more than 10 tweets count in the last 24 hours for a particular category(News)

Create view TweetsIn24hours AS

SELECT a.name, a.handle, a.country, b.Total_tweets_24hrs, c.type

FROM profile_data_youtube_twitter AS a

inner join profile_data_twitter AS b

ON a.handle = b.User_ID

inner join category as c

on b.category = c.id and c.type = "News"

WHERE b.Total_tweets_24hrs >= 20

1 • `SELECT * FROM influencer_demo.tweetsin24hours;`

<

Result Grid Filter Rows: Export: Wrap Cell Content:

	name	handle	country	Total_tweets_24hrs	type
▶	NBC News	NBCNews	United States	20	News
	Fox News	FoxNews	United States	20	News
	CNN	CNN	United States	20	News

15) Complete data for a particular Youtuber including Total views, subscribers

```
SELECT c.name, a.handle, a.subscribers, a.total_views, c.country
FROM youtube_profile_data AS a
INNER JOIN profile_data_youtube_twitter AS c
ON a.handle = c.handle
WHERE a.handle = "CamodoGaming"
```

Result Grid					
		Filter Rows:			
		Export:			
		Wrap Cell Content:			
	name	handle	subscribers	total_views	country
▶	Camodo Gaming	CamodoGaming	2990000	1895838340	United States

Data Insights with Visualisations

We have further used the above cleaned, normalized data to form data summaries of each table

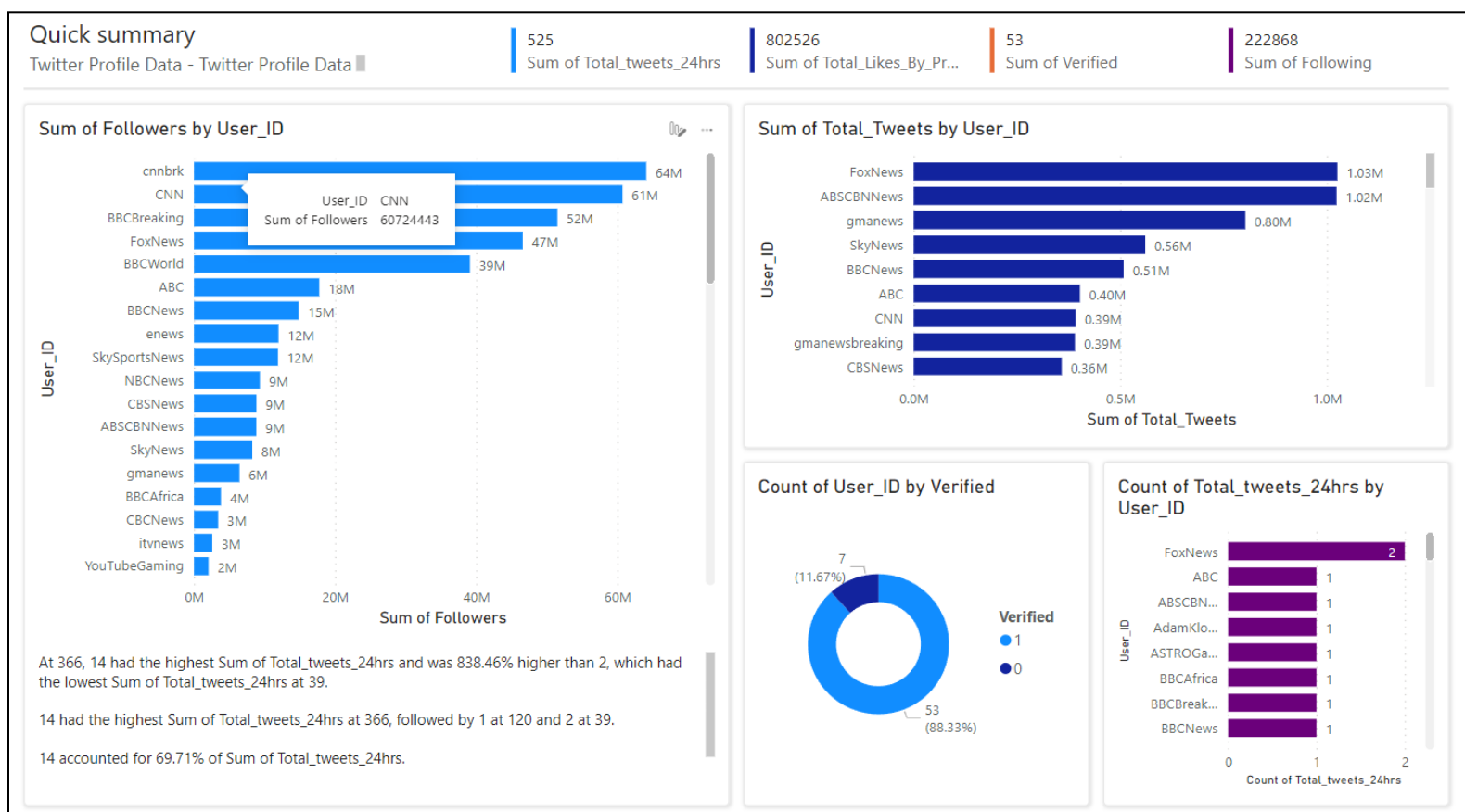
1) Table Name: twitter profile data

Link:

https://app.powerbi.com/links/mMhQB1-8SX?ctid=a8eec281-aaa3-4dae-ac9b-9a398b9215e7&pbj_source=linkShare

Insights:

The below visual depicts twitter profile data and shows various twitter handles, their data, sum of total tweets by a particular user, checks whether the users profile is verified, etc



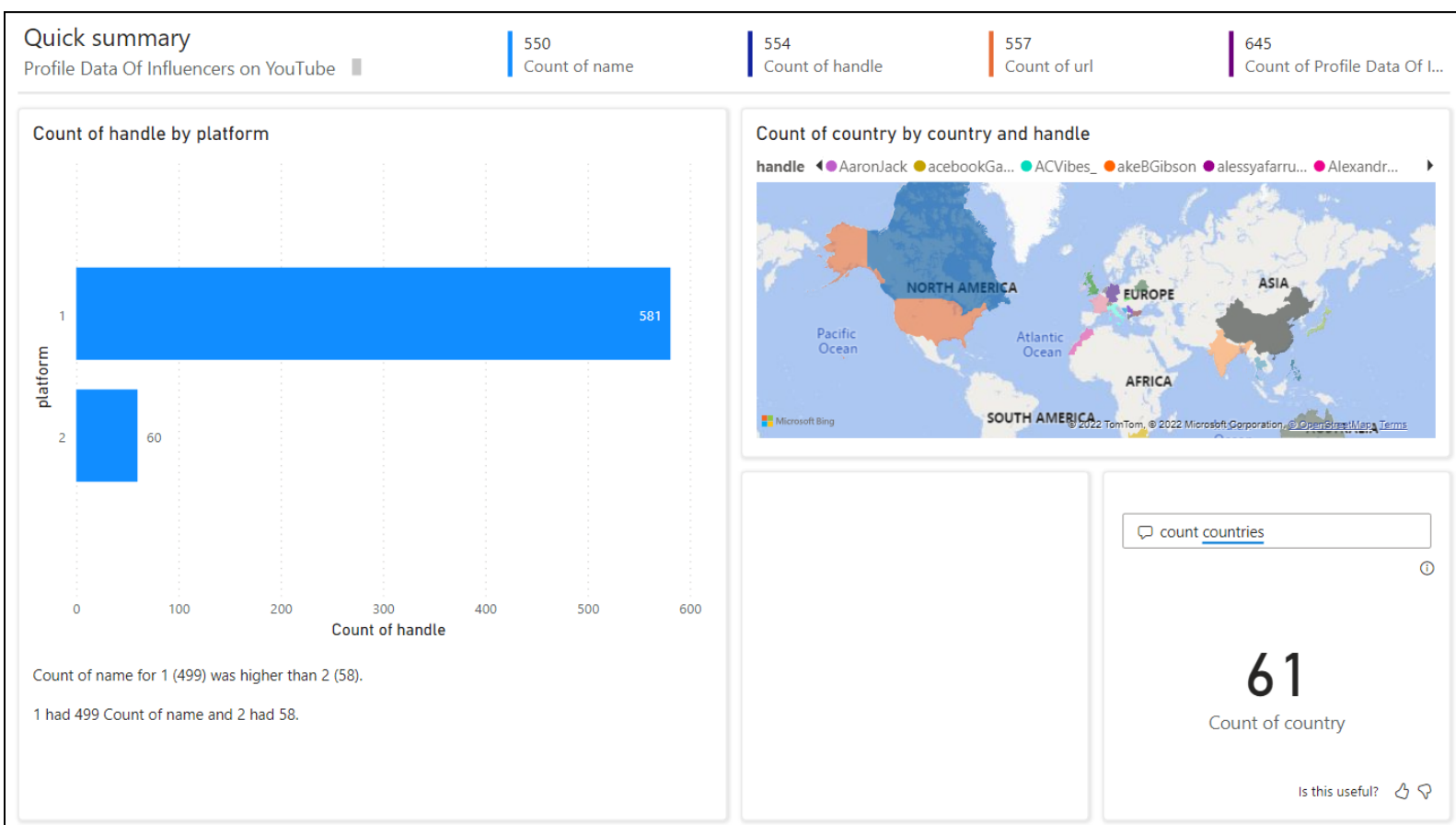
2) Table Name: profile_data

Link:

<https://app.powerbi.com/links/y7cVfh0STN?ctid=a8eec281-aaa3-4dae-ac9b-9a398b9215e7&pbisource=linkShare>

Insights:

The below shows visuals of platform and number of handles collected for the platforms. Here 1 in platform represents youtube and 2 represents twitter. The map depicts handles and countries and the total count of countries for which data is brought



3) Table Name: youtube_profile_data

Link:

<https://app.powerbi.com/links/60DQeK0mCc?ctid=a8eec281-aaa3-4dae-ac9b-9a398b9215e7&pbisource=linkShare&bookmarkGuid=cb529969-5984-4ba0-9a96-136a1d134807>

Insights:

The below visual shows the handles vs total subscribers of the account (this can further be verified online to match data on YouTube dated: 1st December, 2022)

It also compares category and different accounts in that category, average subscribers in a particular category, etc



4) Table Name: tweets

Link:

https://app.powerbi.com/links/60DQeKOmCc?ctid=a8eec281-aaa3-4dae-ac9b-9a398b9215e7&pb_source=linkShare

Insights:

The below shows visuals of Twitter tweets of various profiles, it depicts twitter user ID against the total likes and retweets on the account, it also shows the total likes the profile has got and the average retweets the account gets on posts

