Shorya Mehra
86052200035

Welcome

What?

Steps

Conclusion

YouTube

Statistical Analysis

Course: **Functional Programming**
Professor: **Dr. Kavita Jain**

Shorya Mehra
86052200035

Welcome

What?

Steps

Conclusion

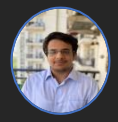
What is Data Analysis?

- Data analysis is the act of turning raw, messy data into useful insights by cleaning the data up, transforming it, manipulating it, and inspecting it.

- The insights gathered from the data are then presented visually in the form of charts, graphs, or dashboards.

- The insights discovered can help aid the company’s or organization’s growth. Decision-makers will be able to come to an actionable conclusion and make the right business decisions.

Tools for Data Analysis



Shorya Mehra
86052200035

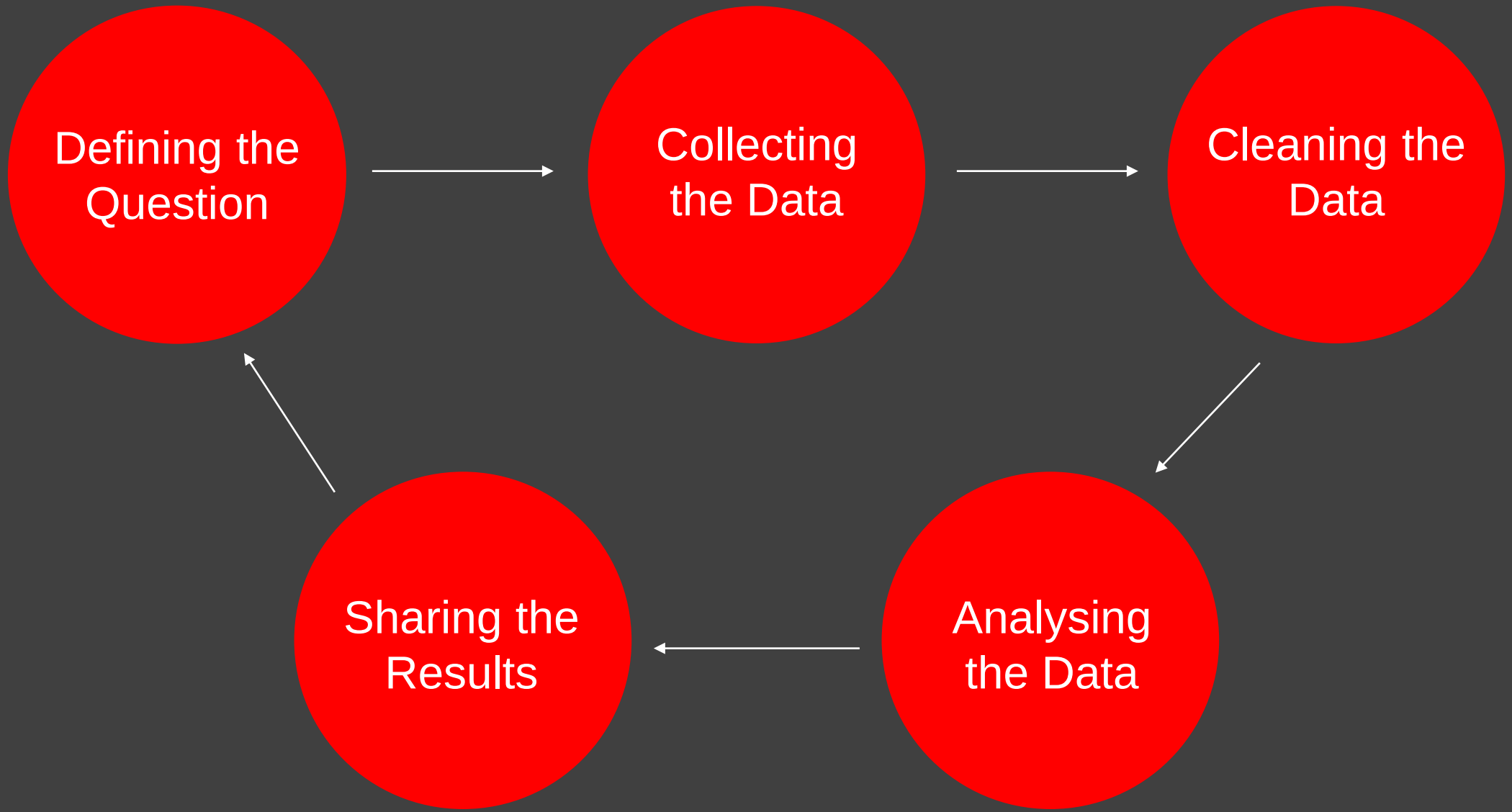
Welcome

What?

Steps

Conclusion

Steps of Data Analysis


Defining the Question
Collecting the Data
Cleaning the Data
Analysing the Data
Sharing the Results

Step 1 | Defining the Question

- The first step in any data analysis process is to define your objective. In data analytics jargon, this is sometimes called the ‘problem statement’.

- Defining your objective means coming up with a hypothesis and figuring how to test it. Start by asking: What business problem am I trying to solve? While this might sound straightforward, it can be trickier than it seems.

YouTube

youtube.com/ShoryaMehra

Views **Likes** **Comments** **Keywords**

Step 2 | Collecting the Data

- Once you’ve established your objective, you’ll need to create a strategy for collecting and aggregating the appropriate data. A key part of this is determining which data you need.

- This might be quantitative (numeric) data, e.g. sales figures, or qualitative (descriptive) data, such as customer reviews.

	Title	Video ID	Published At	Keyword	Likes	Comments	Views
0	Apple Pay	wAZZ-UW(	23-08-2022	tech	3407	672	135612
1	The most E	b3x28s61q	24-08-2022	tech	76779	4306	1758063
2	My New H	4mgePWW	23-08-2022	tech	63825	3338	1564007
3	Petrol Vs L	kXiYSI7H2b	23-08-2022	tech	71566	1426	922918
4	Best Back	ErMwWXC	08-08-2022	tech	96513	5155	1855644
5	Brewmast	18fwz9Itbv	05-11-2021	tech	33570	1643	943119
6	Tech Mon	jXf04bhcjb	13-06-2022	tech	135047	9367	5937790
7	I bought th	2TqOmtTA	07-08-2022	tech	216935	12605	4782514
8	15 Emergi	wLlL46pYc	08-12-2021	tech	45565	2882	7001236
9	Toxicologi	R7qsau3X6	14-07-2022	tech	24252	1068	667767
10	Dope Tech	MEiq0oCU	15-08-2022	tech	118001	4123	2359142
11	Cool Tech	pT_9hntW	06-08-2022	tech	20999	3091	413179
12	Cool Back	cj4lxmHQV	13-08-2022	tech	15322	547	389114
13	Best Tech/	d-BdIo8_w	15-12-2021	tech	17866	157	444953
14	My Massiv	eFhhW6fs/	09-07-2022	tech	13217	442	371563
15	Why Mille	N88OE2ZC	04-06-2021	tech	26890	4999	1633059
16	10 Coolest	PKATJiyz0i	13-08-2021	tech	9562	199	760249
17	17 Coolest	qiMnSaZW	21-07-2022	tech	11743	143	1086568
18	Almost EV	4AnyhHl3_	14-08-2022	tech	146978	11105	3186890
19	I bought th	nmY2kgW'	25-03-2022	tech	363771	13609	11422924
20	iPhone 14	HMqyLcNr	24-08-2022	tech	2117	486	63650
21	Doctor An	Q2ysbTVlP	12-05-2022	tech	42841	1357	1280387
22	Cool Tech	zI8KQq5zn	02-07-2022	tech	23689	2842	530889
23	Wolfgang	iEAlq1GPgı	18-06-2021	tech	55231	2175	2122433
24	Battle Of T	IHnhSxizV(	31-07-2022	tech	1289	183	35948
25	OnePlus 1(	Ql1-RZ2FU	24-08-2022	tech	970	207	23175

kaggle™

bit.ly/shoryadata

salesforce

sas

Step 3 | Cleaning the Data

- Once you’ve established your objective, you’ll need to create a strategy for collecting and aggregating the appropriate data. A key part of this is determining which data you need.

- This might be quantitative (numeric) data, e.g. sales figures, or qualitative (descriptive) data, such as customer reviews.

- This step is also known as **Data Preprocessing**.

pandas

```
vid_df = pd.DataFrame(pd.read_csv("videos-stats.csv"))
```

```
vid_df.head().reset_index(drop=True).style.background_gradient(cmap="Reds")
```

	Unnamed: 0	Title	Video ID	Published At	Keyword	Likes	Comments	Views
0	0	Apple Pay Is Killing the Physical Wallet After Only Eight Years | Tech News Briefing Podcast | WSJ	wAZZ-UWGVHI	2022-08-23	tech	3407.000000	672.000000	135612.000000
1	1	The most EXPENSIVE thing I own.	b3x28s61q3c	2022-08-24	tech	76779.000000	4306.000000	1758063.000000
2	2	My New House Gaming Setup is SICK!	4mgePWWCAmA	2022-08-23	tech	63825.000000	3338.000000	1564007.000000
3	3	Petrol Vs Liquid Nitrogen | Freezing Experiment | പെട്രോളിനെ ഐസ് ആകാൻ പറ്റുമോ | M4 Tech |	kXiYSI7H2b0	2022-08-23	tech	71566.000000	1426.000000	922918.000000
4	4	Best Back to School Tech 2022!	ErMwWXQxHp0	2022-08-08	tech	96513.000000	5155.000000	1855644.000000

```
vid_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1881 entries, 0 to 1880
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Unnamed: 0    1881 non-null   int64  
 1   Title         1881 non-null   object 
 2   Video ID      1881 non-null   object 
 3   Published At  1881 non-null   object 
 4   Keyword       1881 non-null   object 
 5   Likes         1879 non-null   float64
 6   Comments      1879 non-null   float64
 7   Views         1879 non-null   float64
dtypes: float64(3), int64(1), object(4)
memory usage: 117.7+ KB
```

```
vid_df.describe()[["Likes", "Comments","Views"]].style.background_gradient(cmap="Reds")
```

	Likes	Comments	Views
count	1879.000000	1879.000000	1879.000000
mean	170061.020756	7863.331559	11612916.491219
std	796229.267020	37879.964926	108444994.303802
min	-1.000000	-1.000000	25.000000
25%	2672.500000	199.000000	84515.000000
50%	14787.000000	814.000000	591721.000000
75%	60906.000000	3377.500000	2804977.500000
max	16445558.000000	732818.000000	4034122271.000000

```
vid_df.isnull().sum()
```

```
Unnamed: 0      0
Title           0
Video ID        0
Published At    0
Keyword         0
Likes           2
Comments        2
Views           2
dtype: int64
```

```
vid_df.dropna(axis=0,inplace=True )
```

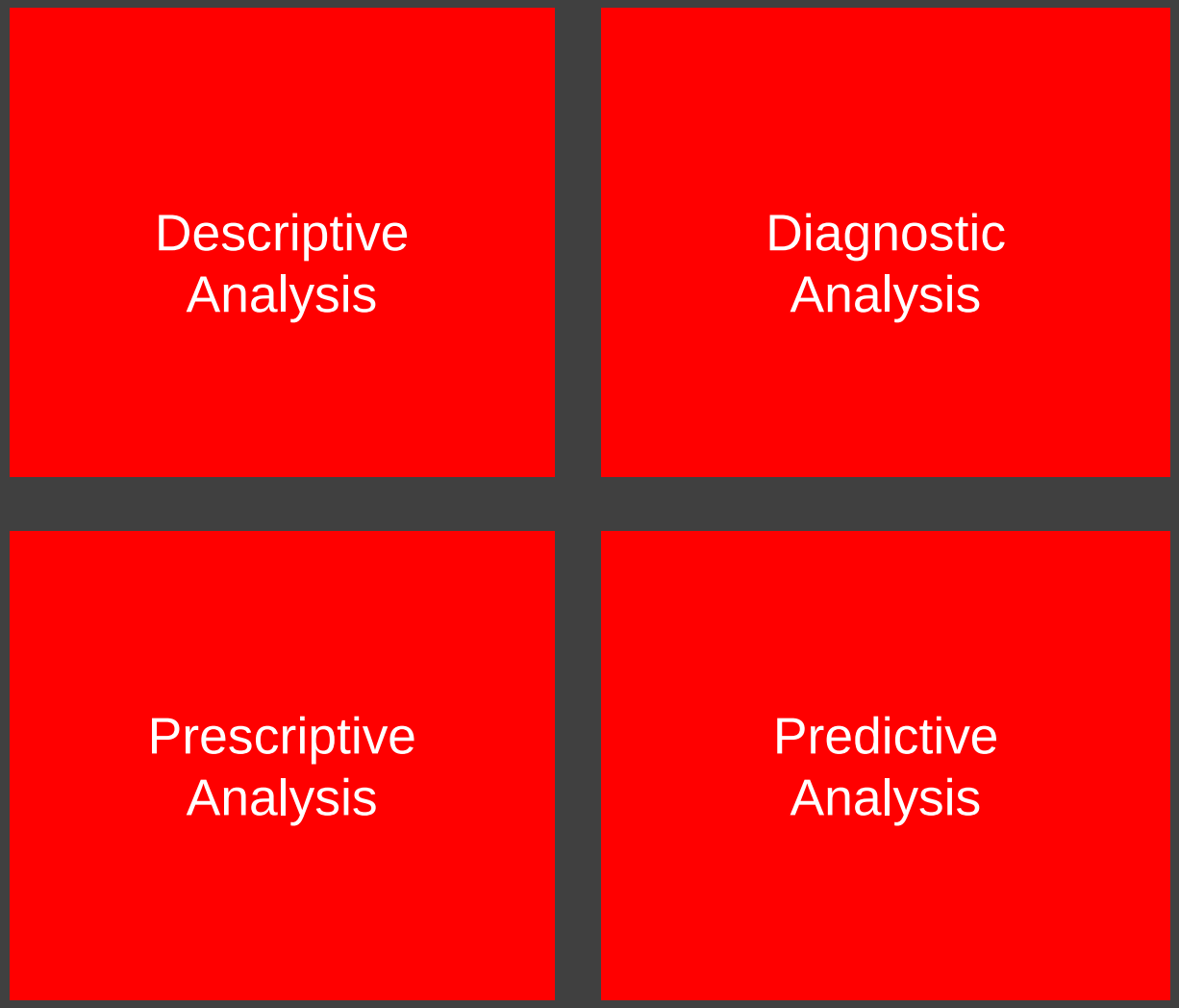
```
vid_df.isnull().sum()
```

```
Unnamed: 0      0
Title           0
Video ID        0
Published At    0
Keyword         0
Likes           0
Comments        0
Views           0
dtype: int64
```

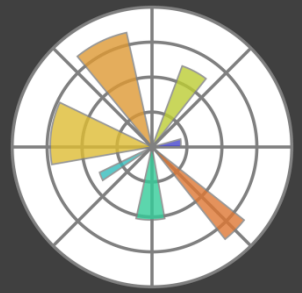
```
vid_df.shape
```

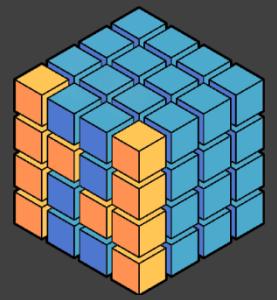
```
(1879, 8)
```

Step 4 | Analysing the Data


Descriptive Analysis
Diagnostic Analysis
Prescriptive Analysis
Predictive Analysis

Exploratory Data Analysis (EDA)





```
vid_df = pd.DataFrame(pd.read_csv("videos-stats.csv"))
```

```
vid_df.head().reset_index(drop=True).style.background_gradient(cmap="Reds")
```

	Unnamed: 0	Title	Video ID	Published At	Keyword	Likes	Comments	Views
0	0	Apple Pay Is Killing the Physical Wallet After Only Eight Years | Tech News Briefing Podcast | WSJ	wAZZ-UWGVHI	2022-08-23	tech	3407.000000	672.000000	135612.000000
1	1	The most EXPENSIVE thing I own.	b3x28s61q3c	2022-08-24	tech	76779.000000	4306.000000	1758063.000000
2	2	My New House Gaming Setup is SICK!	4mgePWWCAmA	2022-08-23	tech	63825.000000	3338.000000	1564007.000000
3	3	Petrol Vs Liquid Nitrogen | Freezing Experiment | പെട്രോളിനെ ഐസ് ആകാൻ പറ്റുമോ | M4 Tech |	kXiYSI7H2b0	2022-08-23	tech	71566.000000	1426.000000	922918.000000
4	4	Best Back to School Tech 2022!	ErMwWXQxHp0	2022-08-08	tech	96513.000000	5155.000000	1855644.000000

Most Liked Videos

```
max_like_vids = vid_df.nlargest(10,"Likes")
max_like_vids[["Title", "Likes"]].reset_index(drop=True).set_index([pd.Index([1,2,3,4,5,6,7,8,9,10])]).style.background_gradient(
```

	Title	Likes
1	El Chombo - Dame Tu Cosita feat. Cutty Ranks (Official Video) [Ultra Music]	16445558.000000
2	$456,000 Squid Game In Real Life!	14259033.000000
3	Martin Garrix - Animals (Official Video)	11025176.000000
4	Powfu - death bed (coffee for your head) (Official Video) ft. beabadoobee	7786057.000000
5	The Weeknd - Save Your Tears (Official Music Video)	6823113.000000
6	Glass Animals - Heat Waves (Official Video)	6177588.000000
7	I Spent 50 Hours Buried Alive	6055504.000000
8	Maroon 5 - Animals (Official Music Video)	5743875.000000
9	history of the entire world, i guess	5471653.000000
10	One Direction - History (Official Video)	5400589.000000

Keyword-Like Graph

```
stats_likes_df = vid_df[['Keyword', 'Likes']].groupby(['Keyword']).agg(['max', 'min', 'mean', 'sum'])
stats_likes_df.sort_values(by=('Likes', 'sum'), ascending=False).style.background_gradient(cmap='Reds')
```

	Likes			
	max	min	mean	sum
Keyword				
mrbeast	14259033.000000	31521.000000	2105914.360000	105295718.000000
animals	11025176.000000	37.000000	760775.789474	28909480.000000
google	16445558.000000	-1.000000	471167.488889	21202537.000000
bed	7786057.000000	25.000000	473682.181818	20842016.000000
cubes	2307773.000000	150.000000	303061.200000	15153060.000000
music	6823113.000000	94.000000	314188.108696	14452653.000000
history	5471653.000000	160.000000	273410.660000	13670533.000000
marvel	1790536.000000	-1.000000	210379.200000	10518960.000000
tutorial	981522.000000	1341.000000	167720.880000	8386044.000000
how-to	1165120.000000	63.000000	158898.145833	7627111.000000
mukbang	1221507.000000	-1.000000	141406.644444	6363299.000000
apple	4144389.000000	184.000000	118078.904762	4959314.000000
physics	1194852.000000	962.000000	98700.900000	4935045.000000
food	672749.000000	774.000000	95014.645833	4560703.000000
sports	867074.000000	15.000000	90518.877551	4435425.000000
reaction	3001265.000000	462.000000	73959.780000	3697989.000000

```
vid_df["Published At"] = pd.to_datetime(vid_df["Published At"])
stats_time_df = vid_df.sort_values(by='Published At').reset_index(drop=True).sort_values(by="Published At", ascending=False)
stats_time_df.drop("Video ID", axis=1,inplace=True)
```

Ten Latest Videos

```
stats_time_df.head(10)[["Title",'Published At', 'Views', 'Likes', 'Comments']].reset_index(drop=True).set_index([pd.Index([1,2,3
```

	Title	Published At	Views	Likes	Comments
1	ASI serán las VOTACIONES MC LIVE 2022!	2022-08-24 00:00:00	1044608.000000	98631.000000	3188.000000
2	Xbox REVEALS Big Updates For Xbox Series X Exclusives & Xbox Game Pass For PC | Xbox News	2022-08-24 00:00:00	16396.000000	926.000000	258.000000
3	The REAL REASON Bitcoin Will Hit $500,000 Per Coin By 2030 (NOT CLICKBAIT)	2022-08-24 00:00:00	77109.000000	5595.000000	641.000000
4	Youtuber Gets Sued By Crypto Millionaire	2022-08-24 00:00:00	474368.000000	27738.000000	2355.000000
5	Biden Announces Plan For Federal Student Loan Debt Relief	2022-08-24 00:00:00	54596.000000	1200.000000	2841.000000
6	Apple September 7 iPhone 14 Event ANNOUNCED! 🥳 EVERYTHING We're Getting!	2022-08-24 00:00:00	83807.000000	4187.000000	723.000000
7	Andrew Tate Announces New "Business" After Ban	2022-08-24 00:00:00	26499.000000	441.000000	496.000000
8	Apple's September 7th iPhone 14 Event Everything Leaked!	2022-08-24 00:00:00	42191.000000	1294.000000	163.000000
9	Apple Announces Fall iPhone Event (iPhone 14, Apple Watch Pro & More)	2022-08-24 00:00:00	47811.000000	915.000000	107.000000
10	Biden plans to forgive $300 billion in student loan debt: Report	2022-08-24 00:00:00	60528.000000	1573.000000	2670.000000

Ten Oldest Videos ¶

```
stats_time_df.tail(10)[["Title",'Published At', 'Views', 'Likes', 'Comments']].reset_index(drop=True).set_index([pd.Index([1,2,3,
```

	Title	Published At	Views	Likes	Comments
10	J. Holiday - Bed (Official Music Video)	2007-07-16 00:00:00	78137822.000000	515049.000000	16568.000000
9	J Holiday - Bed	2007-12-11 00:00:00	26826274.000000	119272.000000	3743.000000
8	Pink Floyd - We Don't Need No Education Lyrics in Description!	2008-07-25 00:00:00	7178954.000000	50074.000000	3095.000000
7	Biology: Digestive System Video	2009-02-16 00:00:00	17716356.000000	132214.000000	5981.000000
6	Ice Cube - It Was A Good Day	2009-02-24 00:00:00	168546247.000000	2307773.000000	71897.000000
5	Bon Jovi - Bed Of Roses (Official Music Video)	2009-06-17 00:00:00	302125099.000000	1072689.000000	37571.000000
4	Girls Aloud - Biology	2009-06-26 00:00:00	7073338.000000	18353.000000	2560.000000
3	The Animals - House of the Rising Sun (1964) HQ/Widescreen ♫ 58 YEARS AGO	2009-07-03 00:00:00	153478497.000000	1357197.000000	39624.000000
2	Lec 1 | MIT 6.00 Introduction to Computer Science and Programming, Fall 2008	2009-08-20 00:00:00	6091315.000000	61813.000000	3416.000000
1	Yung Berg - The Business ft. Casha	2009-10-25 00:00:00	28325416.000000	229141.000000	7461.000000

Most Viewed Videos

```
most_views = vid_df.nlargest(10,"Views").reset_index(drop=True)[["Title","Views"]].set_index([pd.Index([1,2,3,4,5,6,7,8,9,10])])
most_views.style.background_gradient(cmap="Reds")
```

	Title	Views
1	El Chombo - Dame Tu Cosita feat. Cutty Ranks (Official Video) [Ultra Music]	4034122271.000000
2	Martin Garrix - Animals (Official Video)	1582262997.000000
3	The Weeknd - Save Your Tears (Official Music Video)	915457091.000000
4	Maroon 5 - Animals (Official Music Video)	826423766.000000
5	Powfu - death bed (coffee for your head) (Official Video) ft. beabadoobee	524709805.000000
6	One Direction - History (Official Video)	434352213.000000
7	PIKOTARO - PPAP (Pen Pineapple Apple Pen) (Long Version) [Official Video]	425478119.000000
8	Glass Animals - Heat Waves (Official Video)	378164492.000000
9	Young Money - Bed Rock (Official Music Video)	321977550.000000
10	Rihanna - California King Bed	308501014.000000

Most Commented/Discussed Videos

```
most_comments = vid_df.nlargest(10,"Comments").reset_index(drop=True)[["Title","Comments"]].set_index([pd.Index([1,2,3,4,5,6,7,8,
most_comments.style.background_gradient(cmap="Reds")
```

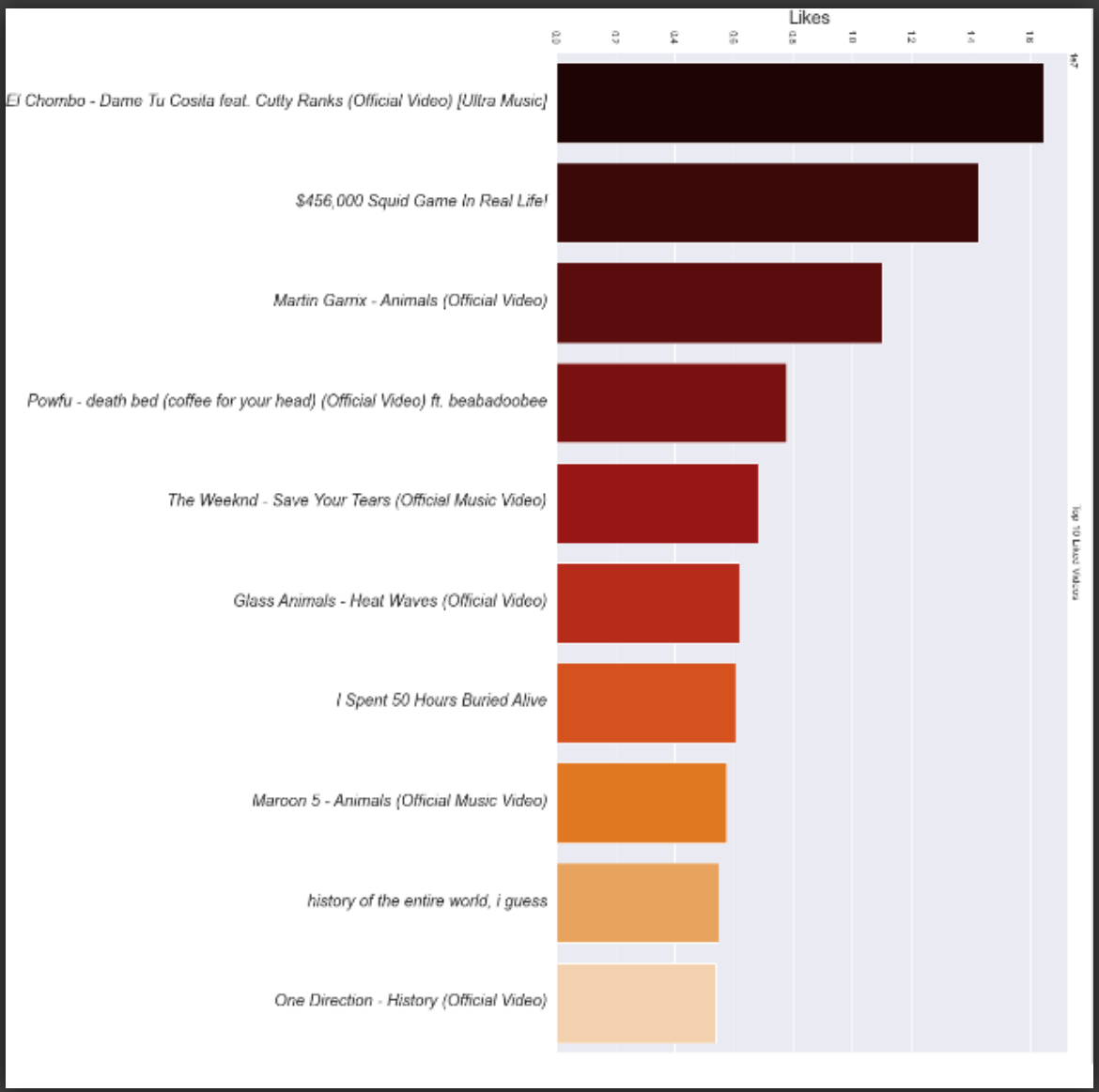
	Title	Comments
1	El Chombo - Dame Tu Cosita feat. Cutty Ranks (Official Video) [Ultra Music]	732818.000000
2	I Opened A Restaurant That Pays You To Eat At It	728774.000000
3	$456,000 Squid Game In Real Life!	611757.000000
4	One Direction - History (Official Video)	434688.000000
5	history of the entire world, i guess	395657.000000
6	Martin Garrix - Animals (Official Video)	335455.000000
7	PIKOTARO - PPAP (Pen Pineapple Apple Pen) (Long Version) [Official Video]	294428.000000
8	The Weeknd - Save Your Tears (Official Music Video)	270948.000000
9	I Spent 50 Hours Buried Alive	239172.000000
10	I Gave My 100,000,000th Subscriber An Island	224423.000000

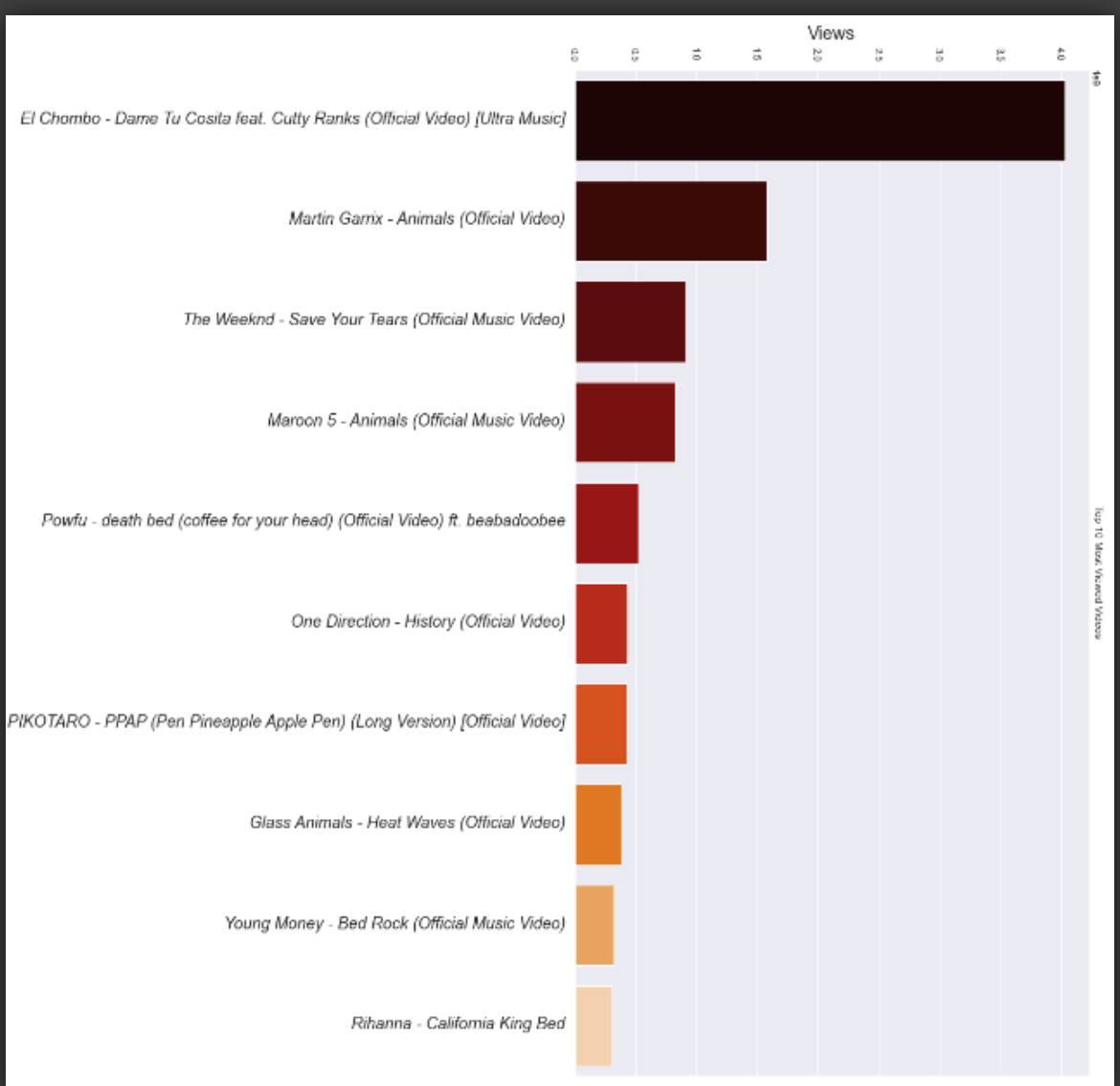
Step 5 | Sharing the Results

- The final step of the data analytics process is to share these insights with the wider world.

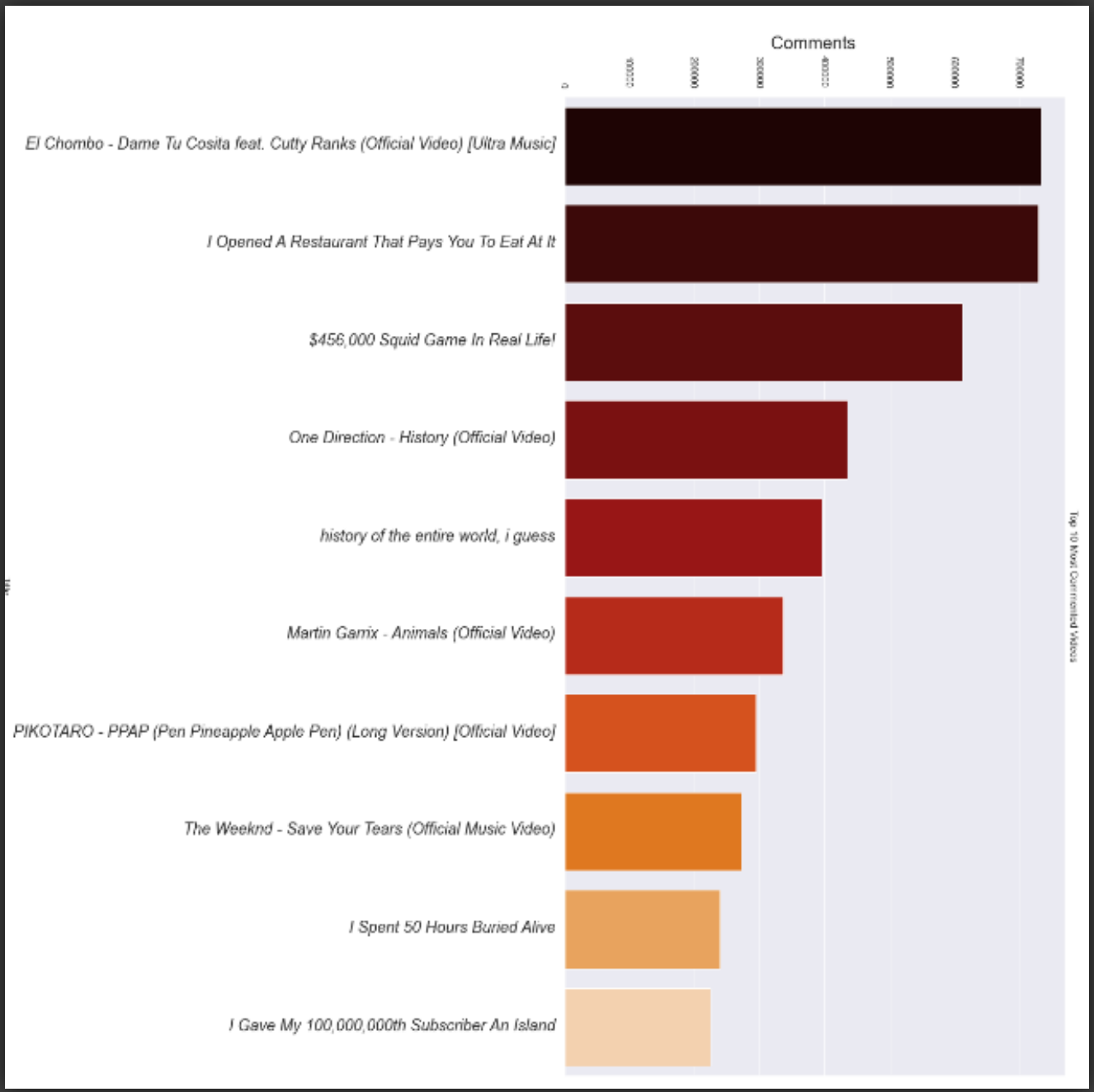
- This is more complex than simply sharing the raw results of your work—it involves interpreting the outcomes, and presenting them in a manner that's digestible for all types of audiences.

- Since you'll often present information to decision-makers, it's very important that the insights you present are 100% clear and unambiguous. For this reason, data analysts commonly use reports, dashboards, and interactive visualizations to support their findings.


Likes
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1e7
El Chombo - Dame Tu Cosita feat. Cutty Ranks (Official Video) [Ultra Music]
$456,000 Squid Game In Real Life!
Martin Garrix - Animals (Official Video)
Powfu - death bed (coffee for your head) (Official Video) ft. beabadoobee
The Weeknd - Save Your Tears (Official Music Video)
Glass Animals - Heat Waves (Official Video)
I Spent 50 Hours Buried Alive
Maroon 5 - Animals (Official Music Video)
history of the entire world, i guess
One Direction - History (Official Video)


Views
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
1e9
El Chombo - Dame Tu Cosita feat. Cutty Ranks (Official Video) [Ultra Music]
Martin Garrix - Animals (Official Video)
The Weeknd - Save Your Tears (Official Music Video)
Maroon 5 - Animals (Official Music Video)
Powfu - death bed (coffee for your head) (Official Video) ft. beabadoobee
One Direction - History (Official Video)
PIKOTARO - PPAP (Pen Pineapple Apple Pen) (Long Version) [Official Video]
Glass Animals - Heat Waves (Official Video)
Young Money - Bed Rock (Official Music Video)
Rihanna - California King Bed

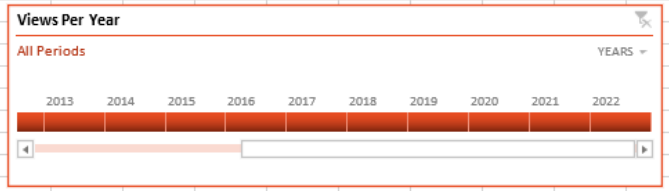
sum of y
16M
14M
12M
10M
8M
6M
4M
2M
0
x
google
mrbeast
animals
bed
music
history
apple
reaction
cubes
marvel
business
lofi
minecraft
mukbang
physics
how-to
tutorial
sports
sat
food
asmr
nintendo
mathchemistry
interview
xbox
machine learning
trolling
education
biology
tech
movies
gaming
chess
finance
computer science
game development
crypto
data science
literature
cnn
news

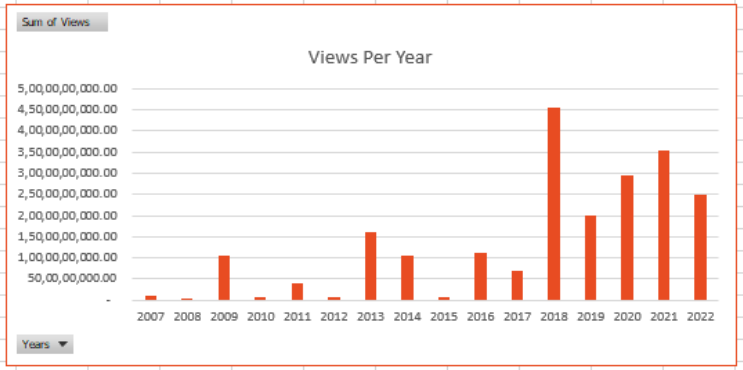

Comments
El Chombo - Dame Tu Cosita feat. Cutty Ranks (Official Video) [Ultra Music]
I Opened A Restaurant That Pays You To Eat At It
$456,000 Squid Game In Real Life!
One Direction - History (Official Video)
history of the entire world, i guess
Martin Garrix - Animals (Official Video)
PIKOTARO - PPAP (Pen Pineapple Apple Pen) (Long Version) [Official Video]
The Weeknd - Save Your Tears (Official Music Video)
I Spent 50 Hours Buried Alive
I Gave My 100,000,000th Subscriber An Island
Top 10 Most Commented Videos

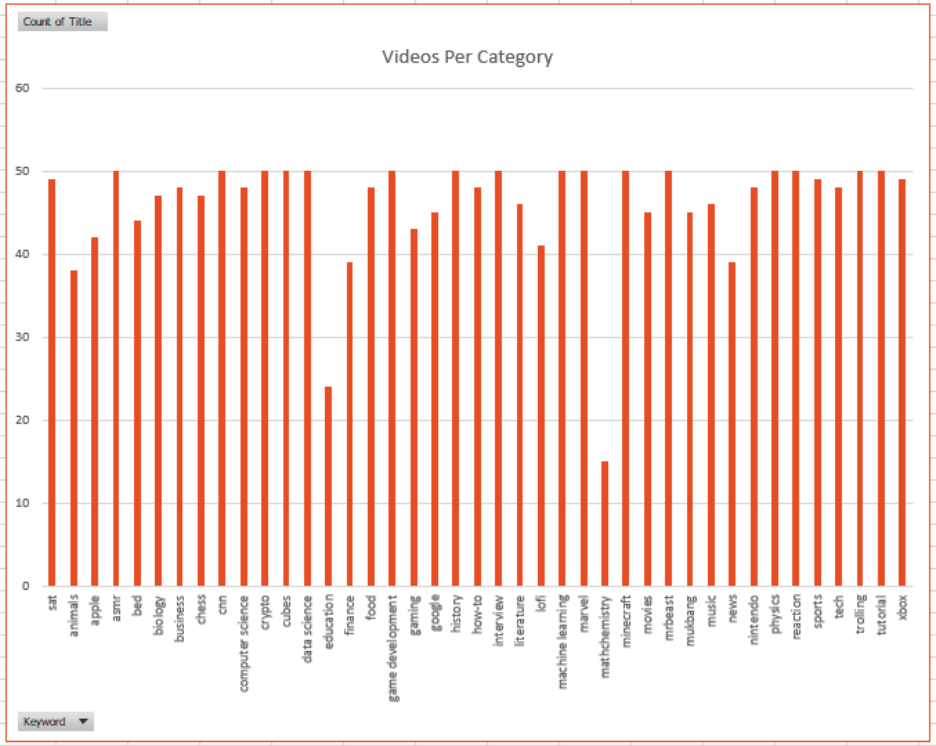
Name: **SMDash**

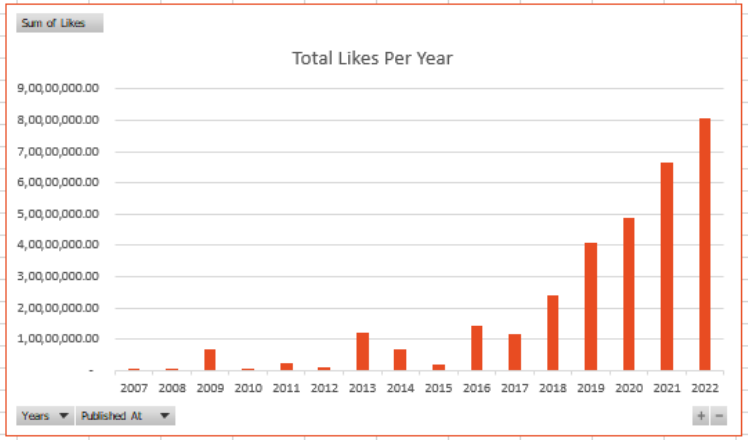
Password: **shoryadata**

YouTube Statistics









Shorya Mehra
86052200035

Welcome

What?

Steps

Conclusion

Thank you!