Website Scraper

Objectives

- Extract the Website Data like Twitter tweets , YouTube Comments, Facebook Comments and Other website data.

Benefits

- Extract tweets from specific User.
- Extract tweets from using Twitter Advance Search Query.
- Extract data from Flipkart Mobile Data .

Data Sharing Agreement(Username)

- Data File Name: **TwitterData _<USERNAME>_.csv**
- Number of Columns: 7
- Number of Rows: <User Defined>
- Columns Name: Index, Username, Date, Tweets, Like Count, Reply Count, Retweet Count.
- Columns Datatype: **Number, Date, String**

Data Sharing Agreement(Query)

- Data File Name: **TwitterDataQuery.csv**

- Number of Columns: 7

- Number of Rows: <User Defined>

- Columns Name: Index, Username, Date, Tweets, Like Count, Reply Count, Retweet Count.

- Columns Datatype: **Number, Date, String**

Data Sharing Agreement(Flipkart)

- Data File Name: **Flipkart Mobile Data_<Page Number>csv**
- Number of Columns: 7
- Number of Rows: <User Defined>
- Columns Name: Index, Model Name, Price, Image
- Columns Datatype: **Number, String**

Architecture
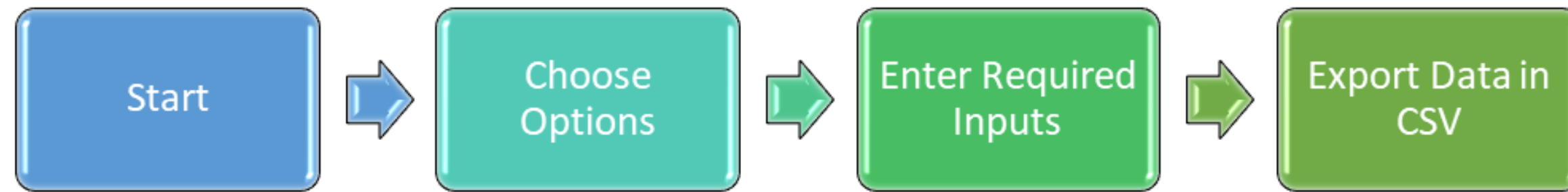
For Twitter by Username

Step 1: Choose the Choice

Step 2: Enter the Username and Number of Tweets

Step 3: Output Data Saved in .csv file

	Username	Date	Tweets	Like Count	Reply Cou	Retweet Count
0	elonmusk	2022-05-09 13:35:52+00:00	Chocolate milk is insanely good. Just had some.	1034579	67553	71787
1	elonmusk	2022-05-09 00:51:26+00:00	If I die under mysterious circumstances, itâ€™s been nice knowin ya	1810252	142621	175147
2	elonmusk	2022-05-09 00:40:57+00:00	.@Rogozin sent this to Russian media https://t.co/eMI08NnSby	104407	8547	9058
3	elonmusk	2022-05-06 15:33:10+00:00	Another Starlink mission completed https://t.co/8ZUeQtTegE	148355	8131	10748
4	elonmusk	2022-05-06 15:14:54+00:00	If Twitter acquisition completes, company will be super focused on hardcore softw	392982	21246	36340
5	elonmusk	2022-05-04 11:03:33+00:00	I dunno â€¦ seems kinda fungible	479452	35813	32007
6	elonmusk	2022-05-03 20:00:53+00:00	Ultimately, the downfall of the Freemasons was giving away their stonecutting serv	438747	26080	33746
7	elonmusk	2022-05-03 19:34:55+00:00	Who funds these organizations that want to control your access to information?	401490	23484	58388
8	elonmusk	2022-05-02 19:00:20+00:00	As I was saying â€¦ https://t.co/tsGz6fCWuW	1366785	55556	109505
9	elonmusk	2022-05-02 18:12:19+00:00	the elusive beauty of imperfection	425460	24221	30349
10	elonmusk	2022-05-01 06:23:15+00:00	Since Iâ€™ve been asked a lot:	1134855	45342	115677
11	elonmusk	2022-04-29 12:28:10+00:00	The far left hates everyone, themselves included!	1230066	65541	131540
12	elonmusk	2022-04-28 20:53:49+00:00	.@CocaCola Oh hi lol https://t.co/VhhVM9GzLJ	647022	35690	45472
13	elonmusk	2022-04-28 17:47:22+00:00	https://t.co/Q9OjlJhi7f	1612187	95504	222530
14	elonmusk	2022-04-28 01:57:27+00:00	Listen, I canâ€™t do miracles ok https://t.co/z7dvLMUXy8	2652047	80205	223046
15	elonmusk	2022-04-28 01:53:50+00:00	Letâ€™s make Twitter maximum fun!	2741015	119961	207345
16	elonmusk	2022-04-28 00:56:58+00:00	Next Iâ€™m buying Coca-Cola to put the cocaine back in	4884624	195800	711226
17	elonmusk	2022-04-28 00:13:47+00:00	Twitter DMs should have end to end encryption like Signal, so no one can spy on or	1523042	44338	115584
18	elonmusk	2022-04-27 20:38:20+00:00	For Twitter to deserve public trust, it must be politically neutral, which effectively m	1609391	84329	154359
19	elonmusk	2022-04-27 18:06:21+00:00	https://t.co/1CE7rjBrNH	498317	20978	51339
20	elonmusk	2022-04-27 05:01:00+00:00	Truth Social is currently beating Twitter & TikTok on the Apple Store https://t.c	635097	48016	54501
21	elonmusk	2022-04-26 20:15:32+00:00	Per aspera ad astra!	539213	38091	44344
22	elonmusk	2022-04-26 20:09:46+00:00	Raptor 2 rocket engines at Starbase, each producing over half a million pounds (230	420293	20205	23478
23	elonmusk	2022-04-26 18:26:46+00:00	The extreme antibody reaction from those who fear free speech says it all	1722803	84246	208580
24	elonmusk	2022-04-25 19:43:22+00:00	ðŸš€ðŸ'«â™¡¸ Yesss!!! â™¡¸ðŸ'«ðŸš€ https://t.co/0T9HzUHuh6	2694565	157938	370576
25	elonmusk	2022-04-25 16:12:30+00:00	I hope that even my worst critics remain on Twitter, because that is what free spee	3335266	188916	390233
26	elonmusk	2022-04-25 15:35:41+00:00	And be my love in the rain	829064	41545	56713
27	elonmusk	2022-04-24 11:46:48+00:00	Moving on â€¦	506039	34539	33412

For Twitter by Query

Step 1: Choose the Choice

Step 1: Create Twitter Advance Search Query
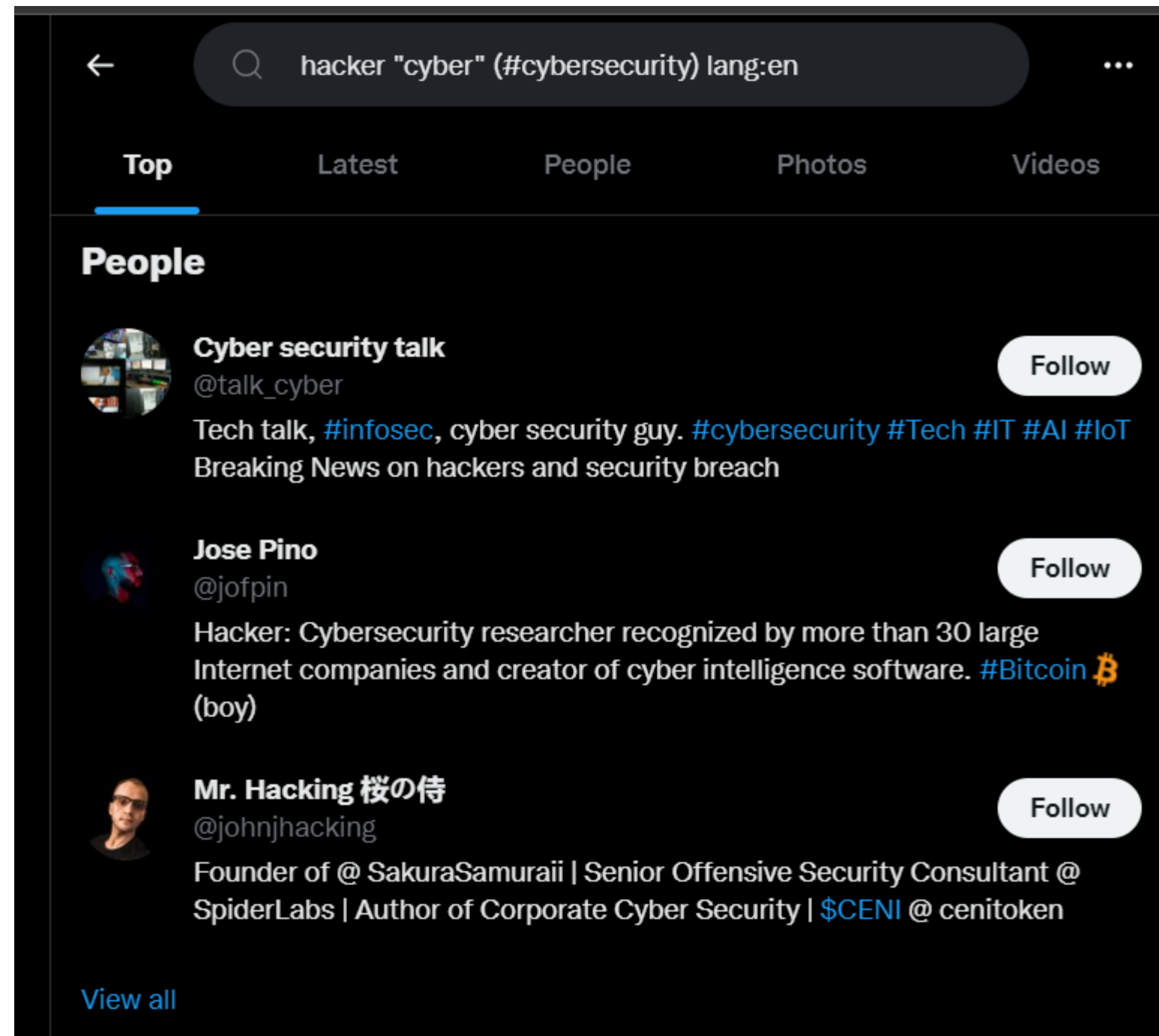
Step 1: Enter the Query and Number of Tweets

Step 1: Output of Query Result

	Username	Date	Tweets	Like Count	Reply Cou	Retweet Count
0	Securityblog	2022-05-12 02:50:	Illinois college, hit by ransomware attack, to shut down | #m	1	0	3
1	Securityblog	2022-05-12 02:48:(	Customers Sue Horizon Actuarial Over Ransomware Attack |	0	0	0
2	reach2ratan	2022-05-12 02:31:(	The Importance of Defining Secure Code	5	0	7
3	DakotaWrench	2022-05-12 01:40:	Russian hackers increasing efforts after cyberattack shut	0	0	0
4	reach2ratan	2022-05-12 00:31:(	15.3 Million Request-Per-Second #DDoS Attack	3	0	6
5	bronwyn_reid	2022-05-11 23:00:	Hundreds of thousands of business owners have lost their	0	0	0
6	reach2ratan	2022-05-11 22:30:	Itâ€™s business as usual for REvil ransomware	2	0	7
7	reach2ratan	2022-05-11 20:30:	Decade-old bugs discovered in Avast, AVG antivirus	9	0	7
8	Cybersecinsider	2022-05-11 20:10:(	Hackers try to cyber scam by posing Chief Executive of Linco	0	0	0
9	timsandle	2022-05-11 19:12:	Cyber-aware: It doesn't matter if you're a small business or €	0	0	0
10	reach2ratan	2022-05-11 18:31:(	CANs Reinvent LANs for an All-Local World	4	0	9
11	reach2ratan	2022-05-11 16:30:	Thereâ€™s no sugarcoating it: That online sugar daddy may	4	1	4
12	ingliguori	2022-05-11 14:59:	#infographic: Prioritizing Cyber Threats	18	0	15
13	recruitersuper7	2022-05-11 14:38:	Looking for a Sr Cyber Security consultant, if interested	0	0	0
14	reach2ratan	2022-05-11 14:30:	F5 Warns of Critical Bug Allowing Remote Code Execution in	1	0	5
15	three_cube	2022-05-11 14:26:	Check out my latest Interview with David Bombal	27	0	7
16	NSEIndia	2022-05-11 13:00:	Amature hackers attack systems. Professionals attack	18	0	2
17	o_alejaldre	2022-05-11 12:32:	Hackers pose as UK cyber-security chief for online scam | #p	0	0	0
18	reach2ratan	2022-05-11 12:30:	#Google to Add Passwordless Authentication Support to	2	0	3
19	LanceSchukies	2022-05-11 11:50:	What your cyber insurance application form can tell you	0	0	0
20	LawSocietyNL	2022-05-11 11:24:	With increased cyber-attacks and increased sophistication of	1	0	0
21	reach2ratan	2022-05-11 10:30:	NIST Releases Updated Guidance for Managing Software	0	0	1
22	UKCyberWeek	2022-05-11 09:22:	Hackers working for the UK's National Cyber Force (NCF)	1	0	0
23	reach2ratan	2022-05-11 08:30:	Google fixes two critical Pixel vulnerabilities: Get your	3	0	3
24	Team_EDENIAS	2022-05-11 07:28:	THE STATUS OF INDIAâ€™S NATIONAL CYBER SECURITY	1	0	0
25	shahriarsabet	2022-05-11 06:49:	The hackers being â€œroyallyâ€ hacked is a pretty cool	0	0	0
26	wreckingcrew44	2022-05-11 06:49:(	With Cyber Security threats happening continuously, itâ€™s t	0	0	0
27	reach2ratan	2022-05-11 06:21:	The $43 billion Business Email Compromise threat	0	0	3

Steps to Deploy

- git clone https://github.com/sahsisunny/WebsiteScrapperProjectVIEH

- cd WebsiteScrapperProjectVIEH

- pip install -r requirements.txt

- python WebScrapper.py