

✓ TikTok imports

```
import pandas as pd
import time
from tqdm.auto import tqdm
```

```
import requests
from datetime import datetime
```

```
!pip install ensembledata
from ensembledata.api import EDClient
```

```
Requirement already satisfied: ensembledata in /usr/local/lib/python3.10/dist-packages (0.2.3)
Requirement already satisfied: httpx in /usr/local/lib/python3.10/dist-packages (from ensembledata) (0.27.2)
Requirement already satisfied: anyio in /usr/local/lib/python3.10/dist-packages (from httpx->ensembledata) (3.7.1)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-packages (from httpx->ensembledata) (2024.8.30)
Requirement already satisfied: httpcore==1.* in /usr/local/lib/python3.10/dist-packages (from httpx->ensembledata) (1.0.7)
Requirement already satisfied: idna in /usr/local/lib/python3.10/dist-packages (from httpx->ensembledata) (3.10)
Requirement already satisfied: sniffio in /usr/local/lib/python3.10/dist-packages (from httpx->ensembledata) (1.3.1)
Requirement already satisfied: h11<0.15,>=0.13 in /usr/local/lib/python3.10/dist-packages (from httpcore==1.*->httpx->ensembledata) (0.14.0)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from anyio->httpx->ensembledata) (1.2.2)
```

```
# initialize the ensembledata client
client = EDClient("pu0xGAobo0WV6jG2")
```

```
# get a free token at https://dashboard.ensembledata.com/register
# free trial of 50 units per day for 7 days, starting 23/11/24
#
# costs:
# - 1 unit = 200 posts in full_hashtag_search :)
# - 1 unit = 20 posts in hashtag_search
# - 1 unit = 20 posts in keyword_search or full_keyword_search :(
# - 1 unit = 30 comments to a post
# - 1 unit = 30 replies to a comment
#
# documentation available at https://ensembledata.com/apis/docs#tag/Tiktok
# and https://github.com/EnsembleData/tiktok-scraper
# source code available at
# https://github.com/EnsembleData/ensembledata-python/tree/main/ensembledata/api
```

✓ Read posts from TikTok

Hashtag search (200 posts at the cost of 1 unit)

```
# read posts related to an hashtag
result = client.tiktok.full_hashtag_search(
    hashtag = "rome", # hashtag name
    days = 360, # filter out posts created more than N days ago
    remap_output = True, # False for mobile app, True for web
    max_cursor = 200) # max number of posts fetched,
                      # cost is 1 unit per 200 posts

# convert json to dataframe
df = pd.json_normalize(result.data["posts"])

# extract relevant info
df = df[['itemInfos.id', 'itemInfos.createTime', 'itemInfos.authorId',
        'itemInfos.text', 'itemInfos.video.urls', 'itemInfos.shareCount',
        'itemInfos.playCount', 'itemInfos.commentCount']]

# rename columns
df.columns = ['id', 'createTime', 'authorId', 'text', 'video.urls',
              'shareCount', 'playCount', 'commentCount']

# format time to date in string format
df["createTime"] = [str(datetime.fromtimestamp(int(i)).date())
                    for i in df["createTime"]]

# save
df_hashtag = df
df_hashtag.to_csv('rome_tiktok_hashtag_posts.csv', index=False)

display(df_hashtag)
```



	id	createTime	authorId	text	video.urls	shareCount	playCount	commentCount
0	7430530232064412961	2024-10-27	24722567	Pov: visiti Roma nel perido sbagliato #giubile...	[https://v16m.tiktokcdn-eu.com/376fc3098634709...	16583	2139172	1022
1	7327390400950390048	2024-01-23	6993424758629123077	Roma città eterna 🍷❤️ #romacittàeterna #roma...	[https://v16m.tiktokcdn-eu.com/88d81d39c53857c...	2738	1165146	413
2	7318787567552318752	2023-12-31	7045849422799275013	STOP ROMANTICIZING ITALY. #italy #rome #italian	[https://v16m.tiktokcdn-eu.com/566dee66c818fcb...	10655	5112326	6379
3	7410868793632771334	2024-09-04	6806842005659010054	Rome 🍷❤️ #rome #romeitaly #rometravel #italy #i...	[https://v16m.tiktokcdn-eu.com/d721642e0359126...	3433	401507	143
4	7397000549079452960	2024-07-29	7231946113031406619	Total fail #viral#trend#rome#Summer	[https://v16m.tiktokcdn-eu.com/ee2093d6c3ef97e...	18564	6333479	1406
...	...	...	...	...	...	...	...	...
115	7443910610724621600	2024-12-02	6873737272078124038	ROMA 🍷🇮🇹 #rome #roma #italia #italy #trip #nov...	[https://v77.tiktokcdn-eu.com/c817dd6771cf29f9...	0	320	0
116	7433816227442674976	2024-11-05	6842439146486776838	Roma 🍷🇮🇹 #roma #rome #europa #italia #italy #fy...	[https://v77.tiktokcdn-eu.com/889a7b7c510e17b1...	9	3569	3
117	7373359666325146912	2024-05-26	6739164005518967814	They didn't have galvanized steel at the time ...	[https://v77.tiktokcdn-eu.com/7b8b71a353eec3cd...	7	2503	0
118	7438950870999387424	2024-11-19	7252341322874553370	POV: You're visiting Rome for the first time 🇮🇹 ...	[https://v77.tiktokcdn-eu.com/e8ead9aeab98d383...	48	10500	5
119	7438364642252016952	2024-11-17	23155521	🤔🤔 #roma #construction #rometoday #romahoy #ro...	[https://v77.tiktokcdn-eu.com/9bf327098c0076cb...	16	4205	2

120 rows x 8 columns

Keyword search (100 posts at the cost of 5 units) - limited to 180 days!

```
# read posts related to a keyword
keyword = "visit rome" # keyword name
period = 180 # must be one of 0, 1, 7, 30, 90, 180 days
            # returns posts more recent than the given period
country = "us" # leave blank "" for any country
max_cursor = 100 # max number of posts fetched, cost is 1 unit per 20 posts

# loop through keyword_search to control for max_cursor
posts = list()
cursor = None
for i in tqdm(range(int(max_cursor/20))):
    result = client.tiktok.keyword_search(keyword = keyword,
```

```
        period = period,
        country = country,
        cursor = cursor)

posts.extend(result.data["data"])
cursor = result.data.get("nextCursor")

# If there is no next cursor, we've fetched all the available posts
if cursor is None:
    break
```



100%

5/5 [00:35<00:00, 7.68s/it]

```
# convert json to dataframe
df = pd.json_normalize(posts)

# extract relevant info
df = df[['aweme_info.aweme_id', 'aweme_info.create_time',
        'aweme_info.author.uid', 'aweme_info.desc',
        'aweme_info.video.play_addr.url_list',
        'aweme_info.statistics.share_count',
        'aweme_info.statistics.play_count',
        'aweme_info.statistics.comment_count']]

# rename columns
df.columns = ['id', 'createTime', 'authorId', 'text', 'video.urls',
              'shareCount', 'playCount', 'commentCount']

# format time to date in string format
df["createTime"] = [str(datetime.fromtimestamp(int(i)).date())
                    for i in df["createTime"]]

# save
df_keyword = df
df_keyword.to_csv('rome_tiktok_keyword_posts.csv', index=False)

display(df_keyword)
```



	id	createTime	authorId	text	video.urls	shareCount	playCount	commentCount
0	7443243111238602002	2024-12-01	7004818490176750598	The eternal city's glow ✨ #italy #rome #visiti...	[https://v77.tiktokcdn-eu.com/9d49151a42f5e478...	5649	473469	134
1	7443157624809360662	2024-11-30	6993424758629123077	La Città Eterna, dove i ricordi durano per sem...	[https://v77.tiktokcdn-eu.com/1450ca3a67e0bd5a...	10566	1829475	518
2	7443847971235466504	2024-12-02	6770556526920303622	Everything you need to know for Christmas in R...	[https://v77.tiktokcdn-eu.com/d7fce877b020e04f...	1590	96623	66
3	7441298945273384214	2024-11-25	15534388	#rome #romeitaly #italy🇮🇹	[https://v77.tiktokcdn-eu.com/74454842d6110292...	13	9096	21
4	7443510987539402006	2024-12-01	6993424758629123077	Christmas in Rome 🎄❤️✨ #roma #italia #italy #i...	[https://v77.tiktokcdn-eu.com/6d57b54652889d22...	550	58354	33
...	...	...	...	...	...	...	...	...
92	7385610241104252193	2024-06-28	6792304655574828037	Rome is always a good idea ☐☐ 🇮🇹 ✨ ...	[https://v77.tiktokcdn-eu.com/b3f10aa8b00a1686...	1197	175922	56
93	7427179185480764704	2024-10-18	6786288333636420614	Rome 2024 was a vibe 🏠 #rome #2024 #construct...	[https://v77.tiktokcdn-eu.com/7531c56c0e021752...	4219	1134854	127
94	7444224296395738400	2024-12-03	7273433970195661857	The TRUTH about visiting the Colosseum in Rome...	[https://v77.tiktokcdn-eu.com/3125289492edd1e3...	0	286	1
95	7437098354825284872	2024-11-14	6770556526920303622	November in rome outfit inspo! Also you guys a...	[https://v77.tiktokcdn-eu.com/2f1d16921854e754...	102	29302	42
96	7433939815092997381	2024-11-05	6775748372335756293	Un dia manifestas que estaras en Italia y ese...	[https://v77.tiktokcdn-eu.com/aee64d29c6b4200e...	0	735	3

97 rows x 8 columns

✓ Read comments from TikTok

Comments search

```
df_hashtag = pd.read_csv('rome_tiktok_hashtag_posts.csv')
df_keyword = pd.read_csv('rome_tiktok_keyword_posts.csv')

# order by comment count
df_hashtag = df_hashtag.sort_values(by=['commentCount'], ascending=False).reset_index(drop=True)
df_keyword = df_keyword.sort_values(by=['commentCount'], ascending=False).reset_index(drop=True)
# delete posts with less than 30 comments
```

```
df_hashtag = df_hashtag[df_hashtag['commentCount']>=30]
df_keyword = df_keyword[df_keyword['commentCount']>=30]
# state how many
print(f'{len(df_hashtag)} posts for hastags with more than 30 comments')
print(f'{len(df_keyword)} posts for keywords with more than 30 comments')
```

84 posts for hastags with more than 30 comments
57 posts for keywords with more than 30 comments

display(df_hashtag)

🔍

		id	createTime	authorId	text	video.urls	shareCount	playCount	commentCount
0	7349672610927529249	2024-03-23	6801498920364377093	Rome 🇮🇹 #rome #romeitaly #italytravel #euros...	[https://v77.tiktokcdn-eu.com/ed58c7d2b5145d9...	162392	9318909	7096	
1	7379659891151047968	2024-06-12	6993424758629123077	La capitale d'Italia 🇮🇹 #roma #rome #italy #it...	[https://v45.tiktokcdn-eu.com/d29b867e13337b0...	40895	8542455	6716	
2	7352900587245014305	2024-04-01	54711960121925632	Rome 🍷 #fyp #tiktoktravel #rome #romeitaly #ro...	[https://v77.tiktokcdn-eu.com/cee3c4794acc121...	217835	12227149	6626	
3	7318787567552318752	2023-12-31	7045849422799275013	STOP ROMANTICIZING ITALY. #italy #rome #italian	[https://v16m.tiktokcdn-eu.com/566dee66c818fc...	10655	5112326	6379	
4	7381345760148786465	2024-06-17	6612593138022760454	#valentino #valentinoperfume #borninroma #rome...	[https://v58.tiktokcdn-eu.com/video/tos/useas...	46638	12770138	5044	
...	...	...	...	...	...	...	...	...	
79	7375173485728992544	2024-05-31	6810102378110649350	oh how much I miss this beautiful city 🇮🇹 #ro...	[https://v77.tiktokcdn-eu.com/697bc3f34eb0f2d...	387	93014	48	
80	7421961779326045473	2024-10-04	7186645069767033861	October in Rome, Italy 🇮🇹🍷🍊 Are you visiting ...	[https://v77.tiktokcdn-eu.com/603139637ac42bc...	480	85183	35	
81	7402265058476166432	2024-08-12	7166665197695943686	No wonder Rome is called the Eternal City 🏛️ ...	[https://v16m.tiktokcdn-eu.com/5fb6265e7815c2...	893	163032	34	
82	7427917993281178912	2024-10-20	6846800960322061317	The construction in rome is no joke, and there...	[https://v77.tiktokcdn-eu.com/f3dd8570b418387...	455	51163	31	
83	7411045340872183072	2024-09-05	7357657783099786272	something i'll never forget 🏛️🔒 #colosseum #col...	[https://v77.tiktokcdn-eu.com/d969ececf0f94d9e...	2128	160823	30	

84 rows x 8 columns

```
# set search criteria
comments_to_retrieve = 60 # 30 is a safe number = 1 credit!
```

```
comments = list()
for i in range(2): # change the range to get more!
    aweme_id = df_hashtag["id"][i]
    cursor = None
    for _ in range(int(comments_to_retrieve/30)):
        result = client.tiktok.post_comments(
            aweme_id = aweme_id,
            cursor = cursor
        )
        comments.extend(result.data["comments"])
        cursor = result.data.get("nextCursor")
    if cursor is None:
        break

# convert json to dataframe
df = pd.json_normalize(comments)

# extract relevant info
df = df[['aweme_id', 'cid', 'create_time', 'text', 'reply_comment_total']]

# format time to date in string format
df["create_time"] = [str(datetime.fromtimestamp(int(i)).date())
                     for i in df["create_time"]]

# order by number of replies
df = df.sort_values(by=['reply_comment_total'], ascending=False).reset_index(drop=True)

# save
df_comments = df
df_comments.to_csv('rome_tiktok_hashtag_comments.csv', index=False)

display(df_comments)
```



	aweme_id		cid	create_time	text	reply_comment_total
0	7379659891151047968	7383697395861701381		2024-06-23	what's language 🤔	20
1	7379659891151047968	7413769745060070149		2024-09-12	I need friends	13
2	7379659891151047968	7382366231100293906		2024-06-19	Yes to whatever she said. ROMA	12
3	7379659891151047968	7387257871748694789		2024-07-03	I just realized how close Italian is to French.	12
4	7379659891151047968	7392388066424718112		2024-07-17	I love 🇮🇹 !	11
...	...	...		...	...	...
95	7349672610927529249	7354511623425409797		2024-04-05	Algún día!!!!!!!	0
96	7349672610927529249	7376212990095377185		2024-06-03	Там просто шик,була там тиждень тому. Хочу ще раз	0
97	7349672610927529249	7380780692793787141		2024-06-15	Where was the second clip from?	0
98	7349672610927529249	7352578697973515041		2024-03-31	KLASSENFAHRT IM MAI !!! ♡ hab so bock	0
99	7379659891151047968	7388853852877931296		2024-07-07	rome 🍕🤔	0

100 rows × 5 columns

Replies search

```
# set search criteria
replies_to_retrieve = 30 # 30 is a safe number = 1 credit!

replies = list()
for i in range(4): # change the range to get more!
    aweme_id = df_comments["aweme_id"][i]
    cid = df_comments["cid"][i]
    cursor = None
    for _ in range(int(replies_to_retrieve/30)):
        result = client.tiktok.post_comment_replies(
            aweme_id = aweme_id,
            comment_id = cid,
            cursor = cursor
        )
        replies.extend(result.data["comments"])
        cursor = result.data.get("nextCursor")
    if cursor is None:
        break
```

```
# convert json to dataframe
df = pd.json_normalize(replies)

# extract relevant info
df = df[['aweme_id', 'cid', 'reply_id', 'create_time', 'text']]

# format time to date in string format
df["create_time"] = [str(datetime.fromtimestamp(int(i)).date())
                     for i in df["create_time"]]

# save
df_replies = df
df_replies.to_csv('rome_tiktok_hashtag_replies.csv', index=False)

display(df_replies)
```



	aweme_id	cid	reply_id	create_time	text
0	7379659891151047968	7383779338464625413	7383697395861701381	2024-06-23	italian
1	7379659891151047968	7411179305041953556	7383697395861701381	2024-09-05	can you teach Lil bit Italian language
2	7379659891151047968	7420488232829600530	7383697395861701381	2024-09-30	@Asad mobile:I am Asad Mughal. All country hel...
3	7379659891151047968	7438124683721573175	7383697395861701381	2024-11-17	Hello 🙋👋
4	7379659891151047968	7407909671468745504	7383697395861701381	2024-08-27	melonian
5	7379659891151047968	7401013238711059218	7383697395861701381	2024-08-09	Seriously??
6	7379659891151047968	7407745019991491333	7383697395861701381	2024-08-27	hi 🙋👋👋👋👋
7	7379659891151047968	7386382000457155334	7383697395861701381	2024-06-30	italian
8	7379659891151047968	7394948636206383905	7383697395861701381	2024-07-23	Dacian
9	7379659891151047968	7387543360343507720	7383697395861701381	2024-07-03	Norwegian
10	7379659891151047968	7406408205059064608	7383697395861701381	2024-08-23	deutsch
11	7379659891151047968	7405688029191291654	7383697395861701381	2024-08-21	🙄
12	7379659891151047968	7403505256221115169	7383697395861701381	2024-08-15	south italian
13	7379659891151047968	7399705602970796805	7383697395861701381	2024-08-05	Mandarin Chinese 👍
14	7379659891151047968	7387852185041634080	7383697395861701381	2024-07-04	german
15	7379659891151047968	7385695787517788934	7383697395861701381	2024-06-28	Albanian
16	7379659891151047968	7384206504302035745	7383697395861701381	2024-06-24	Arabic
17	7379659891151047968	7385273246100439829	7383697395861701381	2024-06-27	No
18	7379659891151047968	7416429078486336257	7413769745060070149	2024-09-19	I will be?
19	7379659891151047968	7416444744501084933	7413769745060070149	2024-09-19	Alright, how are you doing?
20	7379659891151047968	7416449377361740565	7413769745060070149	2024-09-19	Il fine hbu?
21	7379659891151047968	7416451806802281222	7413769745060070149	2024-09-19	I'm doing pretty good
22	7379659891151047968	7436515498051420983	7413769745060070149	2024-11-12	I will be
23	7379659891151047968	7436567276129846034	7413769745060070149	2024-11-13	I need best friend
24	7379659891151047968	7429290114381955845	7413769745060070149	2024-10-24	same here can we ge