



ANJUMAN-I-ISLAM'S KALSEKAR TECHNICAL CAMPUS  
**School of Engineering & Technology**

Affiliated to : University of Mumbai, Recognised by : DTE (Maharashtra) & Approved by : AICTE (New Delhi)

<b>Name: Farhan Shikalgar</b>	<b>Course Name: SMA Lab</b>
<b>Class: BECO</b>	<b>Roll no: 19CO56</b>

## **Experiment No: 2**

**Aim:** Data Collection-Select the social media platforms of your choice(Twitter, Facebook, LinkedIn, YouTube, Web blogs etc) ,connect to and capture social media data for business (scraping, crawling, parsing).

### **Theory:**

#### **Social Media Scraping of Myntra using Python:**

##### **1. Instagram:**

```
2. from instagramy import InstagramUser
3. import pandas as pd
4.
5. user = InstagramUser("Myntra")
6. print(f"Username: {user.fullname}")
7. print(f"Biography: {user.biography}")
8. print(f"Verified User: {user.is_verified}")
9. print(f"Website: {user.website}")
10. print(f"Followers: {user.number_of_followers}")
11. print(f"Following: {user.number_of_followings}")
12. print(f"No. Of Posts: {user.number_of_posts}")
13.
14. posts = user.posts
15. instaPosts = []
16. for i in range(10):
17.     post = {}
18.     post["Post"] = i+1
19.     post["Likes"] = posts[i].likes
20.     post["Comments"] = posts[i].comments
```

```

21.     post["Caption"] = posts[i].caption
22.     instaPosts.append(post)
23. print(f"Latest posts:\n {pd.DataFrame(instaPosts)}")
24.

```

## Output

```

PS E:\SEM-8\SMA\FK> & C:/Users/farha/AppData/Local/Programs/Python/Python310/python.exe e:/SEM-8/SMA/FK/Exp2-SMA.py
Username: MYNTRA
Biography: If you never try, you'll never know ❤️
📸 #MyMyntraLook get featured on Myntra Studio
Visit Myntra Studio for more fashion & style advice.
Verified User: True
Website: https://myntra.onelink.me/eWcy/gnozbjm9
Followers: 3167991
Following: 262
No. Of Posts: 10799
Latest posts:

```

	Post	Likes	Comments	Caption
0	1	3006	62	None
1	2	2626	43	None
2	3	14290	46	None
3	4	577	14	None
4	5	793	8	None
5	6	742	20	None
6	7	1242	55	None
7	8	1020	23	None
8	9	2043	16	None
9	10	349	27	None

## 2)Facebook:

```

25. # Facebook
26. from facebook_scraper import get_posts
27.
28. fbPosts = []
29. for post in get_posts('Myntra', pages=10):
30.     fbPosts.append({x: post[x] for x in ['post_id', 'text', 'likes']})
31. print(print(f"Latest posts:\n {pd.DataFrame(fbPosts)}"))

```

## Output

```
Latest posts:
      post_id      text      likes
0  10161168767013221  YAAASSS, we can't believe our eyes either 🤔🤔🤔\n...  6.0
1  10161168690158221  YAAASSS, we can't believe our eyes either 🤔🤔🤔\n...  10.0
2  10161167893718221  This Pink monochrome look is ❤️❤️\n\n👉 @simran...  24.0
3  10161166649103221  Whether you're going on a date, or giving your...  79.0
4  10161165440628221  Mark your calendar & set your reminder 'cause ...  23.0
5  10161162761178221  Jump into an unparalleled listening experience...  28.0
6  10161164921723221  Shhhh...something extraordinary coming your way,...  69.0
7  10161164580113221  Shhhh...something extraordinary coming your way,...  24.0
8  10161164466943221  Shhhh...something extraordinary coming your way,...  20.0
9  10161164422793221  Shhhh...something extraordinary coming your way,...  27.0
10 10161163521578221  Denim game strong with pantaloons!\nCheck out ...  34.0
11 10161163439843221  A race before the actual race!\nThe Early Bird...  28.0
12 10161163209568221  Keep swiping 😊  121.0
13 10161161257373221  Vibe check 🖤\n👉@phoolfilledvish\n#Reels #Fash...  61.0
14 10161161191548221  ❤️\n👉@thankgod_itsfashion\n#Reels #FashionReel...  19.0
15 10161161063803221  The clock is ticking! 🕒\nYou don't wanna miss ...  45.0
16 10161160787473221  Pastel magic!\nCheck out 🌟Myntra Studio 🌟for fa...  59.0
17 10161158857943221  Cute outfit inspo🥰\nCheck out 🌟Myntra Studio 🌟f...  NaN
18 10161158772288221  Tea party ready! 🍰\nCheck out our stunning rang...  81.0
19 10161158759248221  Nothing like a good fitted suit! Check out 🌟Myn...  113.0
20 10161158642653221  This ain't no place for basic fashion 😊\n\nChe...  64.0
21 10161158356338221  Saree charm doesn't fade away💖 Check out 🌟Myntr...  72.0
22 10161157987323221  It's always the right time for Insiders to mak...  65.0
23 10161157988108221  Wishlist karne ke bhi alag maje hai 😊  60.0
24 10161157980433221  @houseofpataudi is an integral part of the Myn...  63.0
25 10161157593608221  RUNWAY ICONS - a one-stop destination for cura...  57.0
26 10161156389858221  Latest range of new launches feel your favouri...  64.0
27 10161156376458221  Run towards your fitness goals as fast as you ...  30.0
28 10161156110263221  'Dear Troopers,\nOur D-Day was full of celebra...  43.0
29 10161155738693221  What about yours? 😊  64.0
30 10161155699848221  Anarkali vibes ❤️❤️\nCheck out 🌟Myntra Studio 🌟fo...  24.0
31 10161155593708221  The #Maxessorize Sale is now LIVE 🥳\nGet 50-80...  44.0
32 10161154540828221  Cozy & Comfy 🧡\nCheck out 🌟Myntra Studio 🌟for f...  86.0
33 10161154535858221  Denim on Denim is a treat 🦋\nCheck out 🌟Myntra ...  59.0
34 10161154512788221  Bandhani Print is a must in your wardrobe 🥰\n\n...  110.0
35 10161154187103221  Rocking red ❤️🥰\nCheck out 🌟Myntra Studio 🌟for ...  74.0
```

### 3)Twitter:

```
from twitter_scraper_selenium import get_profile_details

twitter_username = "Myntra"
filename = "twitter_api_data"
get_profile_details(twitter_username=twitter_username, filename=filename)

import json

with open('twitter_api_data.json', 'r') as json_file:
    json_object = json.load(json_file)

print(json.dumps(json_object, indent=1))
```

## Output

2023-02-06 14:15:46,817 - twitter\_scraper\_selenium.profile\_details - INFO - Data Successfully Saved to E:\SEM-8\SMA\FK\twitter\_api\_data.json

```
{
  "id": 89443197,
  "id_str": "89443197",
  "name": "Myntra",
  "screen_name": "myntra",
  "location": "Bengaluru, India",
  "profile_location": null,
  "description": "If you never try, you\u2019ll never know \ud83e\udd0d Shop NOW\ud83d\udc47\ud83c\udfffb",
  "url": "https://t.co/wIOLMicLBd",
  "entities": {
    "url": {
      "urls": [
        {
          "url": "https://t.co/wIOLMicLBd",
          "expanded_url": "https://myntra.onelink.me/ewcy/gnozbjm9",
          "display_url": "myntra.onelink.me/ewcy/gnozbjm9",
          "indices": [
            0,
            23
          ]
        }
      ]
    }
  },
  "description": {
    "urls": []
  },
  "protected": false,
  "followers_count": 692685,
  "fast_followers_count": 0,
  "normal_followers_count": 692685,
  "friends_count": 37,
  "listed_count": 497,
  "created_at": "Thu Nov 12 13:16:01 +0000 2009",
  "favourites_count": 5803,
  "utc_offset": null,
  "time_zone": null,
  "geo_enabled": true,
  "verified": true,
  "statuses_count": 497237,
  "media_count": 8500,
  "lang": null,
  "status": {
```

```
"in_reply_to_screen_name": "JeetPrajapat20",
  "geo": null,
  "coordinates": null,
  "place": null,
  "contributors": null,
  "is_quote_status": false,
  "retweet_count": 0,
  "favorite_count": 0,
  "favorited": false,
  "retweeted": false,
  "lang": "en",
  "supplemental_language": null
},
  "contributors_enabled": false,
  "is_translator": false,
  "is_translation_enabled": false,
  "profile_background_color": "000000",
  "profile_background_image_url": "http://abs.twimg.com/images/themes/theme11/bg.gif",
  "profile_background_image_url_https": "https://abs.twimg.com/images/themes/theme11/bg.gif",
  "profile_background_tile": false,
  "profile_image_url": "http://pbs.twimg.com/profile_images/1617407797632831488/xE29nKpB_normal.jpg",
  "profile_image_url_https": "https://pbs.twimg.com/profile_images/1617407797632831488/xE29nKpB_normal.jpg",
  "profile_banner_url": "https://pbs.twimg.com/profile_banners/89443197/1672990539",
  "profile_link_color": "000000",
  "profile_sidebar_border_color": "000000",
  "profile_sidebar_fill_color": "000000",
  "profile_text_color": "000000",
  "profile_use_background_image": false,
  "has_extended_profile": false,
  "default_profile": false,
  "default_profile_image": false,
  "pinned_tweet_ids": [],
  "pinned_tweet_ids_str": [],
  "has_custom_timelines": true,
  "following": null,
  "follow_request_sent": null,
  "notifications": null,
  "advertiser_account_type": "promotable_user",
  "advertiser_account_service_levels": [
    "dso"
  ],
  "business_profile_state": "none",
  "translator_type": "none",
  "withheld_in_countries": [],
  "require_some_consent": false
}
```



```

"media_count": 8500,
"lang": null,
"status": {
  "created_at": "Mon Feb 06 08:43:35 +0000 2023",
  "id": 1622516340518162433,
  "id_str": "1622516340518162433",
  "text": "@JeetPrajapat20 I hear you & I am here for you, Jeet! Rest assured, I have got this highlighted, and one of our cas\u2026 https://t.c
o/069kk07w11",
  "truncated": true,
  "entities": {
    "hashtags": [],
    "symbols": [],
    "user_mentions": [
      {
        "screen_name": "JeetPrajapat20",
        "name": "Jeet Prajapat",
        "id": 1600448314482118656,
        "id_str": "1600448314482118656",
        "indices": [
          0,
          15
        ]
      }
    ],
    "urls": [
      {
        "url": "https://t.co/069kk07w11",
        "expanded_url": "https://twitter.com/i/web/status/1622516340518162433",
        "display_url": "twitter.com/i/web/status/1\u2026",
        "indices": [
          121,
          144
        ]
      }
    ]
  },
  "source": "<a href=\"https://www.sprinklr.com\" rel=\"nofollow\">Sprinklr</a>",
  "in_reply_to_status_id": 1622515741483495424,
  "in_reply_to_status_id_str": "1622515741483495424",
  "in_reply_to_user_id": 1600448314482118656,
  "in_reply_to_user_id_str": "1600448314482118656",

```

## Conclusion:

We have successfully collected Myntra social media data from websites like Instagram, Facebook, and Twitter. In order to connect and capture social media data for businesses (scraping, crawling, parsing).