

| **Name: Khan shoab akhat vakil ahamad** | **Course Name: SMA-LAB** |
| --- | --- |
| **Class: BE-CO** | **Batch: 01** |
| **Roll no: 19CO33** | **Experiment No: 03** |

**Aim:** Data Cleaning and Storage- Preprocess, filter and store social media data for business (Using Python, MongoDB, R, etc).

**Theory:** Social Media Scraping of Unilever using Python.

**1. Instagram:**

2. # Instagram

3. from instagramy import InstagramUser

4. user = InstagramUser("unilever")

5. print(f"Username: {user.fullname}")

6. print(f"Biography: {user.biography}")

7. print(f"Verified User: {user.is\_verified}")

8. print(f"Website: {user.website}")

9. print(f"Followers: {user.number\_of\_followers}")

10.print(f"Following: {user.number\_of\_followings}")

11.print(f"No. Of Posts: {user.number\_of\_posts}")

12.posts = user.posts

13.print(posts[0])

14.instaPosts = []

15.for i in range(10):

16. post = {}

17. post["Likes"] = posts[i].likes

18. post["Comments"] = posts[i].comments

19. post["post\_source"] = posts[i].post\_source

20. post["post\_url"] = posts[i].post\_url

21. post["time"] = posts[i].taken\_at\_timestamp

22. instaPosts.append(post)

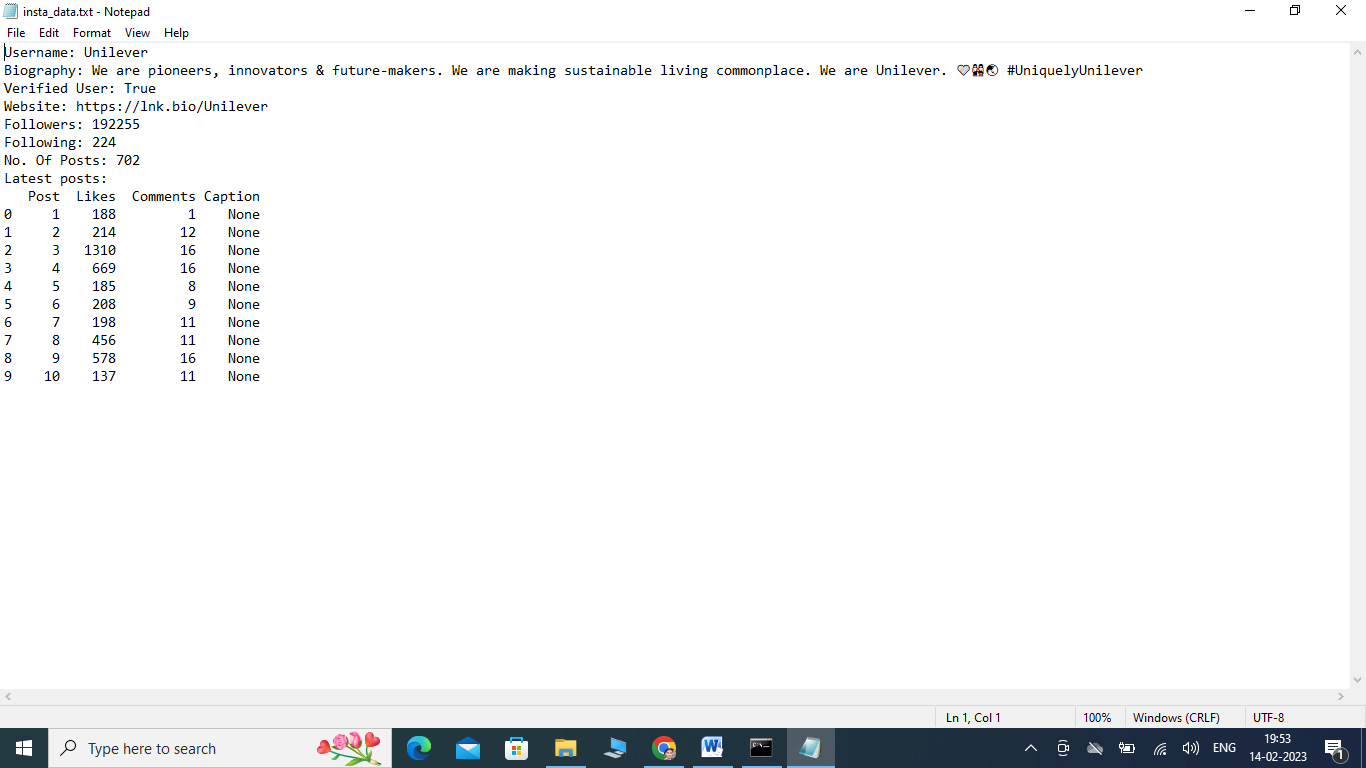
23.insta\_df = pd.DataFrame(instaPosts)

24.insta\_df.info()

25.print(insta\_df.isna())

26.print(f"Latest posts:\n {insta\_df}")

**OUTPUT :**

****

**2.Facebook:**

28.# Facebook

29.from facebook\_scraper import get\_posts

30.fbPosts = []

31.for post in get\_posts('Unilever', pages=10):

32. fbPosts.append(post)

33.print(fbPosts)

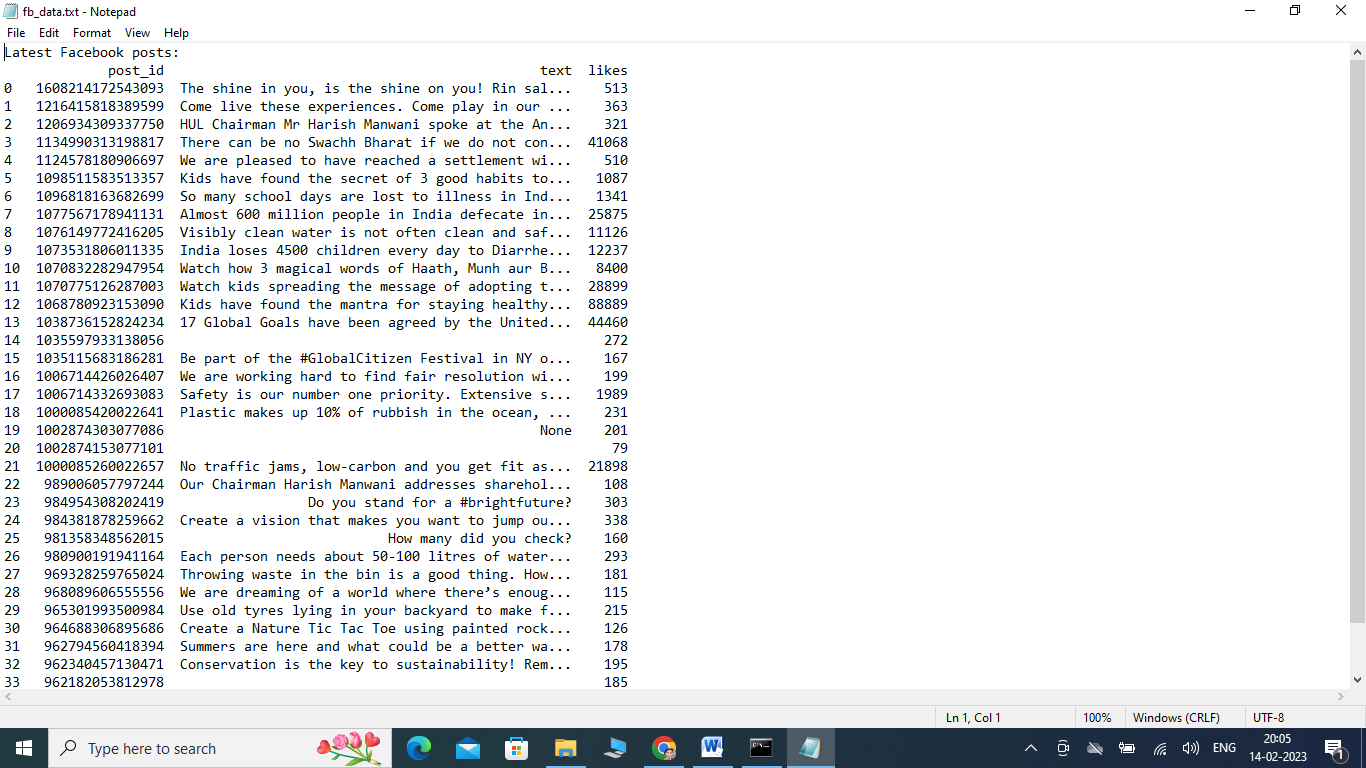
34.df = pd.DataFrame(fbPosts)

35.df.info()

36.df = df[['post\_id', 'text', 'post\_url', 'page\_id', 'comments', 'likes', 'time']]

37.print(df.isna())

38.print(print(f"Latest posts:\n {df}"))

**OUTPUT : **

**3.Twitter:**

40.# Twitter

41.import twint

42.c = twint.Config()

43.c.Lang = "en"

44.c.Username = "Unilever"

45.c.Pandas = True

46.# Run

47.twint.run.Search(c)

48.Tweets\_df = twint.storage.panda.Tweets\_df

49.Tweets\_df.info()

50.print(Tweets\_df.isna())

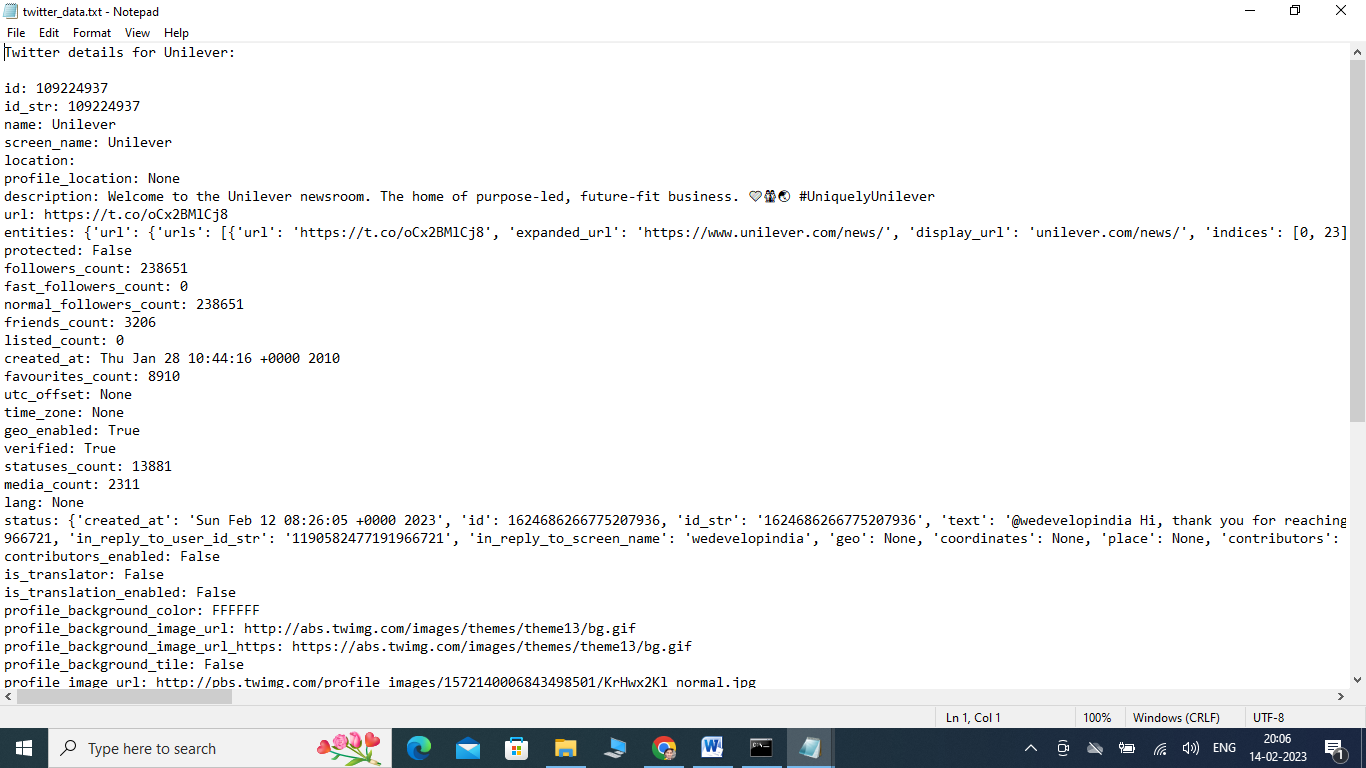
51.Tweets\_df = Tweets\_df[Tweets\_df.language == "en"]

52.Tweets\_df = Tweets\_df[['id', 'date', 'nlikes', 'language',

'nreplies','tweet', 'username']]

53.print(Tweets\_df)

**Output :**

****

**Conclusion:** We have successfully cleaned Unilever’s social media data from websites like Instagram, Facebook, and Twitter using various data preprocessing techniques.