

|  |  |
| --- | --- |
| **Name: Sayyed Sohail Rashid** | **Course Name: SMA-LAB** |
| **Class: BE-CO** | **Batch: 01** |
| **Roll no: 18CO48** | **Experiment No: 03** |

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

# Theory:

**Social Media Scraping of Apple using Python:**

**Instagram:**

1. # Instagram
2. from instagramy import InstagramUser
3. user = InstagramUser("Apple")
4. print(f"Username: {user.fullname}")
5. print(f"Biography: {user.biography}")
6. print(f"Verified User: {user.is\_verified}")
7. print(f"Website: {user.website}")
8. print(f"Followers: {user.number\_of\_followers}")
9. print(f"Following: {user.number\_of\_followings}")
10. print(f"No. Of Posts: {user.number\_of\_posts}")

12.posts = user.posts

13.print(posts[0])

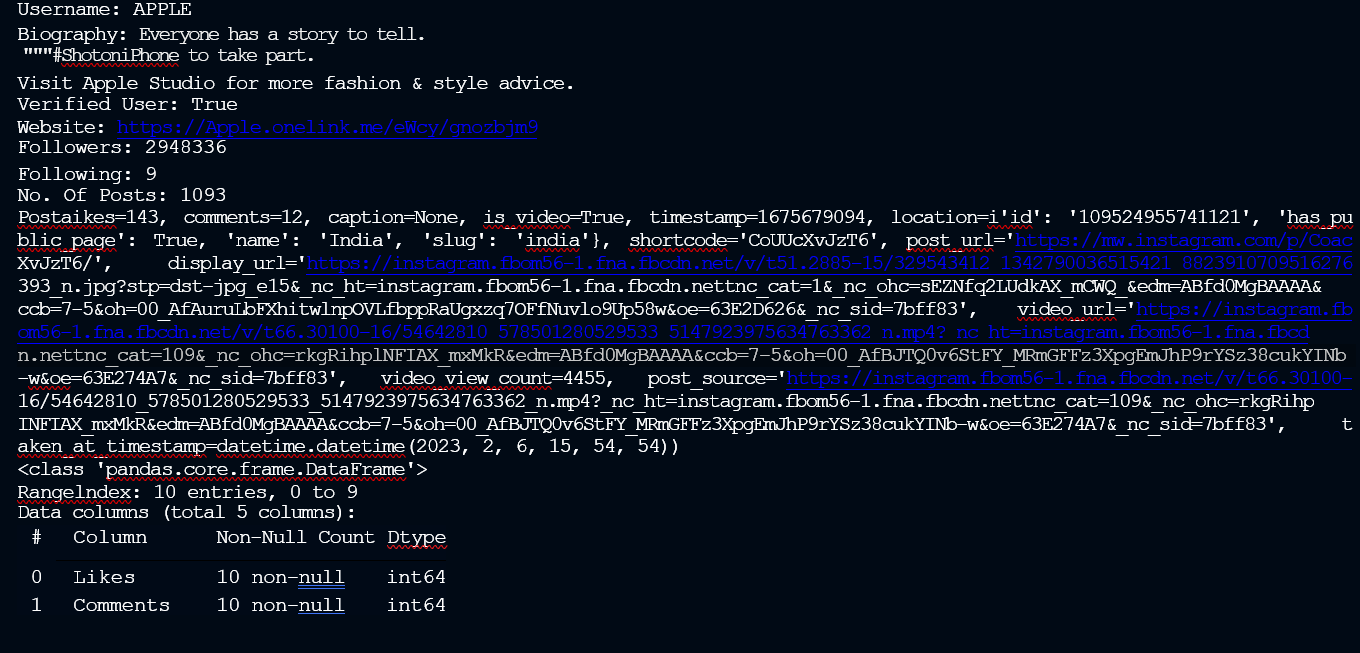
14.instaPosts = []

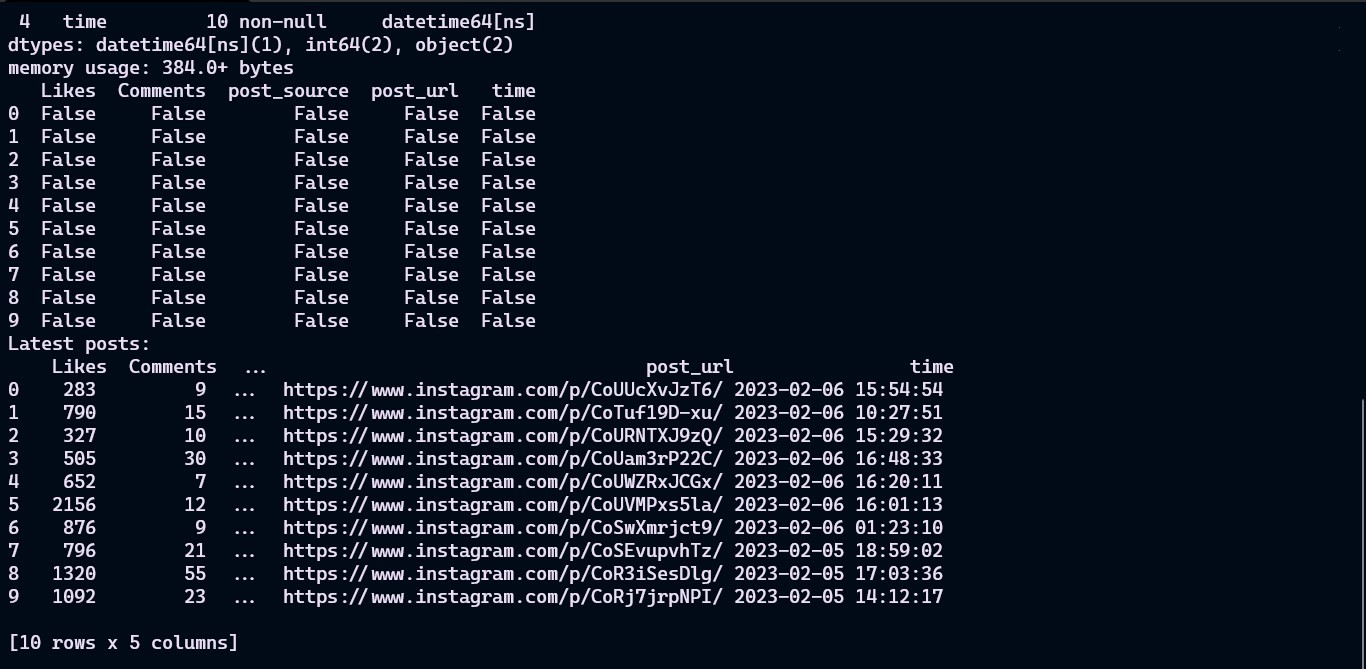
1. for i in range(10):
2. post = {}
3. post["Likes"] = posts[i].likes
4. post["Comments"] = posts[i].comments
5. post["post\_source"] = posts[i].post\_source
6. post["post\_url"] = posts[i].post\_url
7. post["time"] = posts[i].taken\_at\_timestamp
8. instaPosts.append(post)

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

24.insta\_df.info()

1. print(insta\_df.isna())
2. print(f"Latest posts:\n {insta\_df}")





**Facebook:**

1. # Facebook
2. from facebook\_scraper import get\_posts

30.fbPosts = []

1. for post in get\_posts('Apple', pages=10):
2. fbPosts.append(post)
3. print(fbPosts)

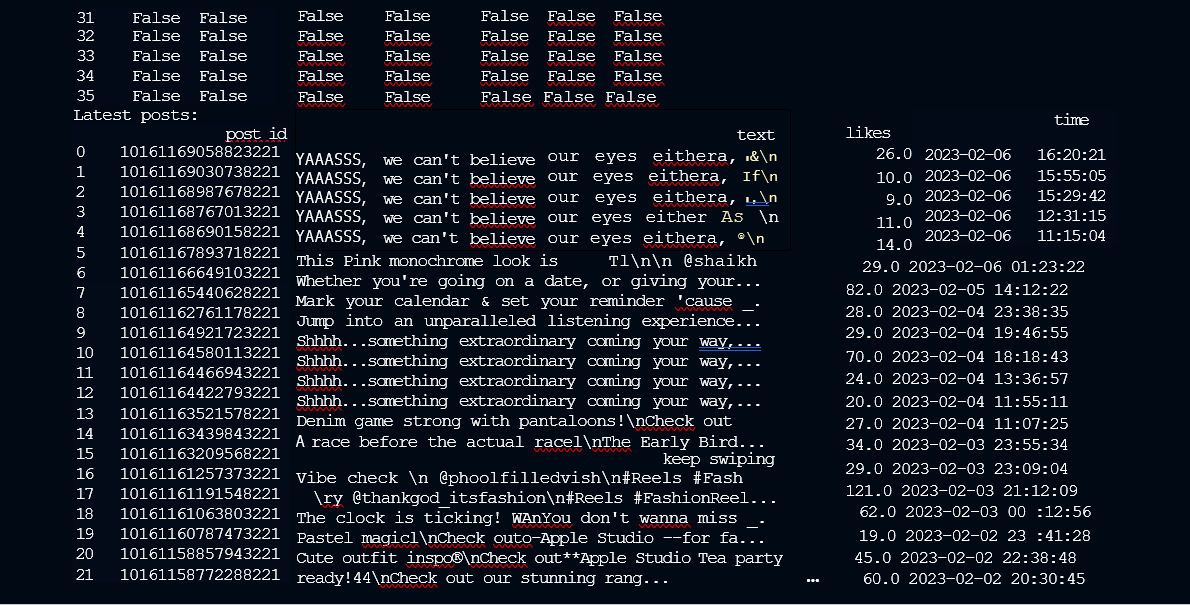
34.df = pd.DataFrame(fbPosts)

35.df.info()

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

1. print(df.isna())
2. print(print(f"Latest posts:\n {df}"))

****

****

**Twitter:**

1. # Twitter
2. import twint

42.c = twint.Config()

43.c.Lang = "en"

44.c.Username = "Apple"

45.c.Pandas = True

1. # Run
2. twint.run.Search(c)

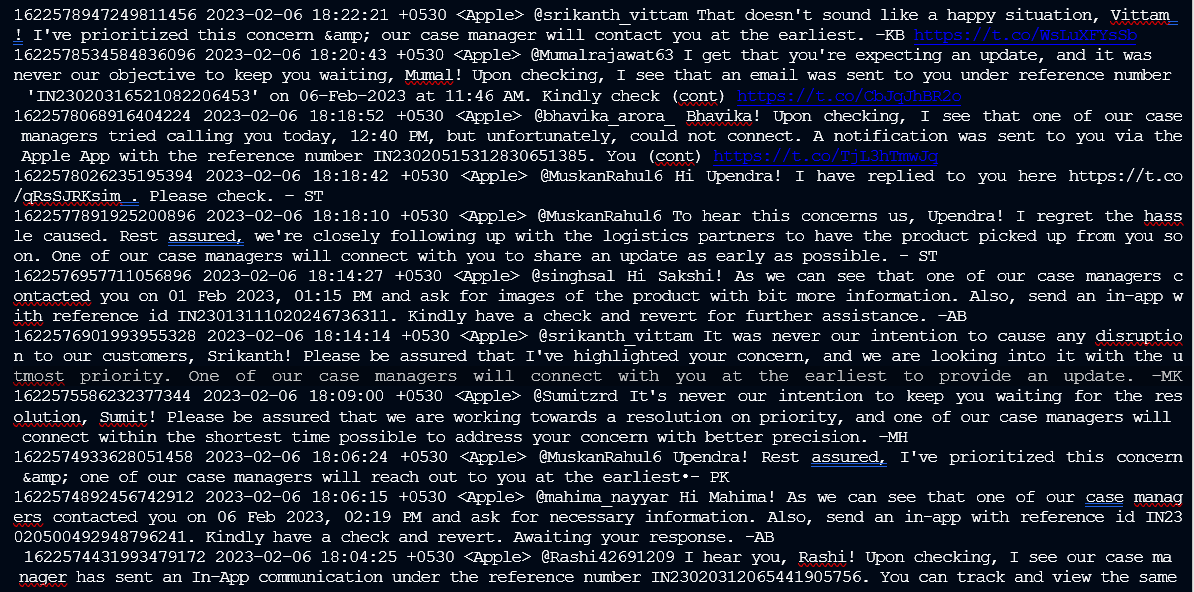
48.Tweets\_df = twint.storage.panda.Tweets\_df

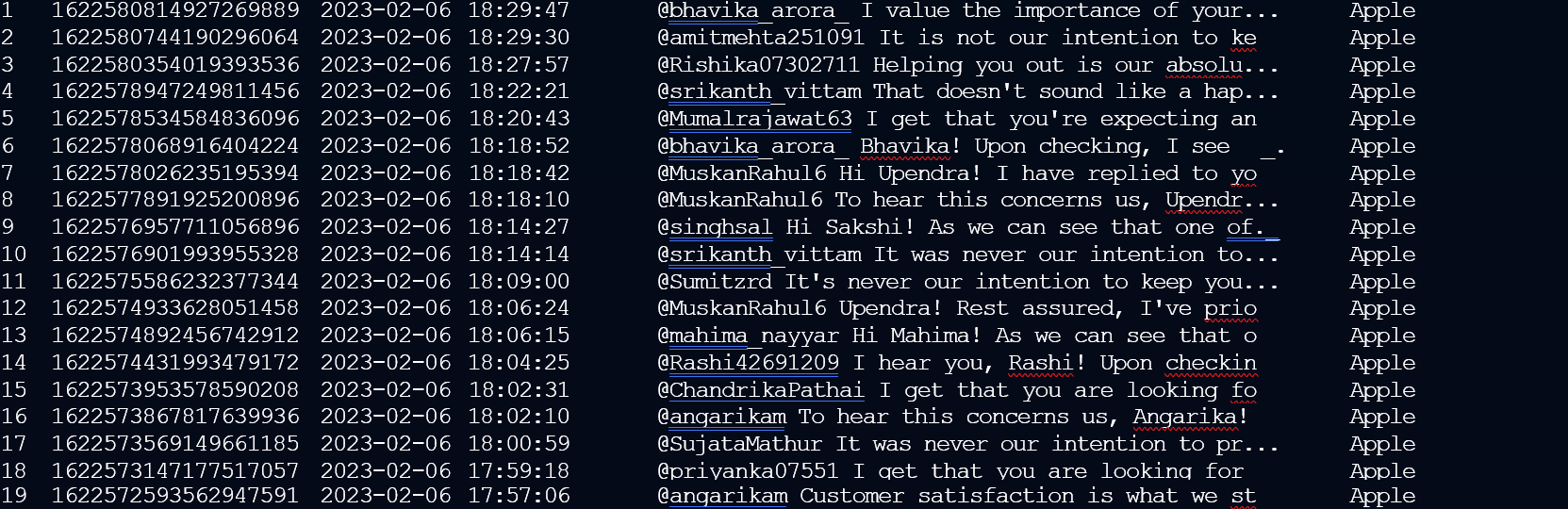
1. Tweets\_df.info()
2. print(Tweets\_df.isna())

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)

****

****

**Conclusion:**

We have Performed cleaning for Apple’s social media data from websites like Instagram, Facebook, and Twitter using various data preprocessing techniques.