Project Midterm Report

1. Team: 551_pg 37

Zihan Liu zliu2193@usc.edu USC ID: 1534981630
Shuchan Zhou zhoushuc@usc.edu USC ID: 9442515557
Zeyu Li zli86605@usc.edu USC ID: 8454499854

2. Current Progress

1) Web-scraping: First, the corresponding Instagram account was searched according to the name of the USC sports club. Second, we extracted the URLs of each club's posts from its account homepage. Finally, we extracted the text, images, likes, and comments of the posts.

2) Data cleaning: Delete posts without any images (these posts only include videos).

3) Transform the data format to JSON data.

Sample of post data including post ID, creation account name number of likes, and post article:

```
1  {"post_id": "B8Z3Zu0JzBa", "account_name": "uscfigureskating", "like_number": "36", "article": "#meetusmond
2  {"post_id": "B8PyHtPJYeu", "account_name": "uscfigureskating", "like_number": "64", "article": "Lastly\u201
3  {"post_id": "B8Mk-WBp6QG", "account_name": "uscfigureskating", "like_number": "71", "article": "Huge thank
4  {"post_id": "B8MiU4-pE25", "account_name": "uscfigureskating", "like_number": "68", "article": "So proud of
5  {"post_id": "B8IYpC5JNow", "account_name": "uscfigureskating", "like_number": "37", "article": "Today\u2019
```

Sample of image URLs of posts including post ID and corresponding image URLs:

```
1  {"post_id": "B8Z3Zu0JzBa", "image_urls": "https://scontent-iad3-2.cdninstagram.com/v/t51.2885-15/84527676_7
2  {"post_id": "B8PyHtPJYeu", "image_urls": "https://scontent-iad3-2.cdninstagram.com/v/t51.2885-15/83886519_3
3  {"post_id": "B8PyHtPJYeu", "image_urls": "https://scontent-iad3-2.cdninstagram.com/v/t51.2885-15/83401413_7
4  {"post_id": "B8PyHtPJYeu", "image_urls": "https://scontent-iad3-2.cdninstagram.com/v/t51.2885-15/84156258_1
5  {"post_id": "B8PyHtPJYeu", "image_urls": "https://scontent-iad3-2.cdninstagram.com/v/t51.2885-15/83434093_1
```

Sample of comment data including post ID and corresponding comments:

```
1  {"post_id": "B8Z3Zu0JzBa", "comments": "Oh hiiiii! \ud83d\ude0d\ud83d\ude18"}
2  {"post_id": "B8Z3Zu0JzBa", "comments": "Lovely.  Beautiful.  If your legs get cold try these fleece side zi
3  {"post_id": "B8PyHtPJYeu", "comments": "We rocking\ud83d\ude4c\ud83d\ude4c"}
4  {"post_id": "B8PyHtPJYeu", "comments": "FIGHT ON FOREVER"}
5  {"post_id": "B8PyHtPJYeu", "comments": "my family\u2764\ufe0f"}
```

4) Uploaded the dataset to MongoDB Database, and connected EC2 MongoDB from the local machine.

5) Started to implement the WebSocket server, and almost finished writing the corresponding python script. However, there is some error when running corresponding commands on the terminal.

6) Finished building the main part of the front-end UI interface

Screenshots of the current page of our UI:
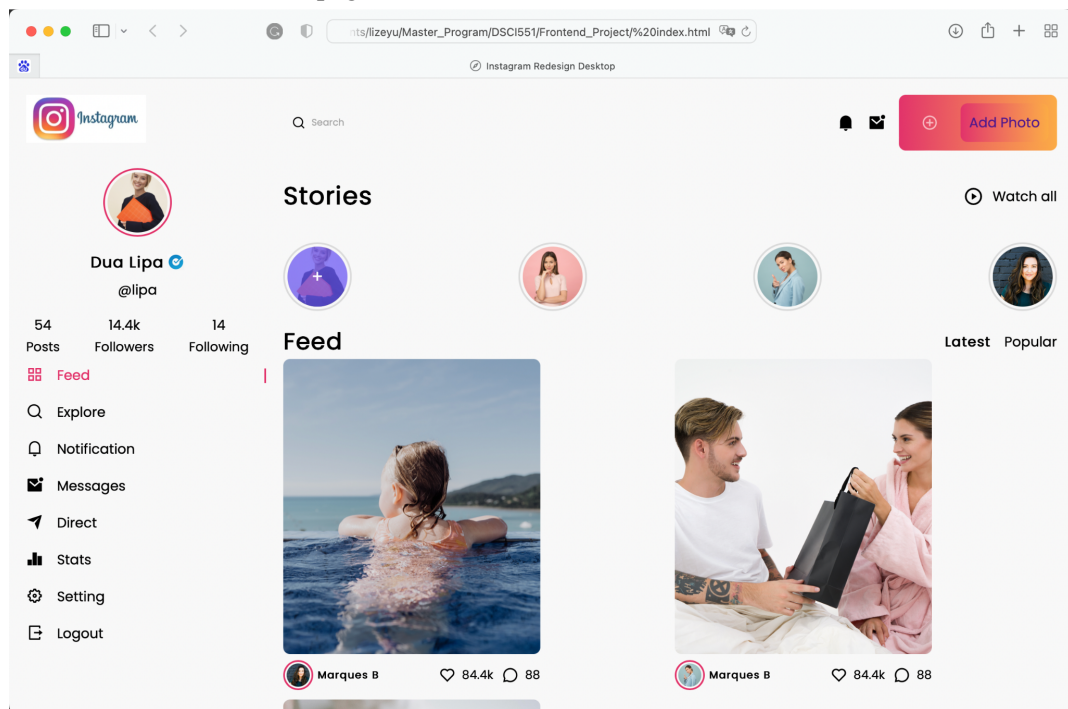
The current screenshot shows the completed front-end framework. Because we are not familiar with the front-end interface and are currently in the state of learning, many functions currently appearing in the interface will be deleted or modified in the final actual results. Therefore, the current UI interface belongs to the preliminary framework version.

3. Challenges Encountered

1) Instagram's API has a limit on the number of requests, thus we could only access to the data in parts.
2) When connecting MongoDB from the local machine, we met some problems with configuration settings.
3) When uploading data to MongoDB, it was necessary to transform structured data (tables) to semi-structured data (JSON), which led to some modifications in the schema.
4) Struggling in building web socket server. When we try to run the "curl -X GET " command in the terminal, there is some connection error.
5) In the design of the front-end interface, it is difficult to set the size of many buttons or functions or refer to different interfaces.

4. Future Plans

1) Finish the implementation of the WebSocket server
2) Revise the frontend UI
3) Test the backend database and functions and frontend UI interface, and make sure the frontend and backend interfaces correspond
4) Connect frontend UI to the backend database
5) Organize demo codes, record presenting video, and write the final report