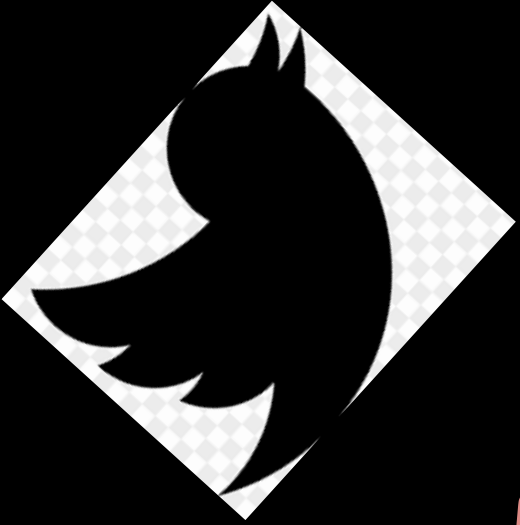


Social Media Data Mining



List of Contents

Introduction

Tools List

Web User Interface

Backend

Demo

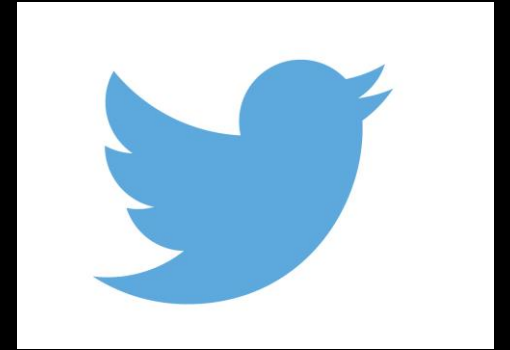


- Twitter currently has **396.5 million** users
- **206 million** users access Twitter daily. **75%** of them are not based in the US
- 68.1% of Twitter users are male. 31.9% are female.
- US adults spend an average of 6 minutes on Twitter per day in 2021.



- Important Twitter Statistics
- Reddit Has One Billion Native Video Views Per Month
- Reddit Saw a 43% Increase for Health & Fitness Content
- 38% of Users are Technology Enthusiasts

Tools List



Twitter	Reddit
Username	Username
Tweet	ID
Screen Name	Title
Tweet ID	Description
Posted Date	Score
User's Profile Picture URL	Timestamp



Data Models

Create a new post



What's happening ?

Tweet



New Post

Title

Text

Post

Web User Interface

Data Storage - Atlas Mongo DB

The screenshot displays the MongoDB Atlas web interface. At the top, the header includes the 'Georgian College' organization name, 'Access Manager', and 'Billing' links. On the right, there are links for 'All Clusters', 'Get Help', and a user profile for 'Sujeevan'.

The main navigation bar shows 'Project 0', 'Atlas', 'Realm', and 'Charts' tabs. The left sidebar is divided into 'DEPLOYMENT' and 'SECURITY' sections. Under 'DEPLOYMENT', 'Databases' is selected, showing a tree view with 'Web_Scraping' expanded and 'Twitter' selected. Other options include 'Triggers', 'Data Lake', 'Database Access', 'Network Access', and 'Advanced'.

The central panel is titled 'Web_Scraping.Twitter' and shows collection statistics: 'COLLECTION SIZE: 10.53KB', 'TOTAL DOCUMENTS: 40', and 'INDEXES TOTAL SIZE: 36KB'. Below this are tabs for 'Find', 'Indexes', 'Schema Anti-Patterns', 'Aggregation', and 'Search Indexes'. An 'INSERT DOCUMENT' button is in the top right.

A filter bar contains the text '{ field: 'value' }' and buttons for 'Apply' and 'Reset'. Below the filter, the section 'QUERY RESULTS 1-20 OF MANY' displays two document snippets. Each snippet shows fields like '_id', 'User Name', 'Tweet', 'Screen Name', 'Tweet ID', 'Posted Date', and 'User Profile Pi...'. The first document's 'Tweet' value is 'Saaaaaaaaaaaaa' and the second is 'Hi There'.

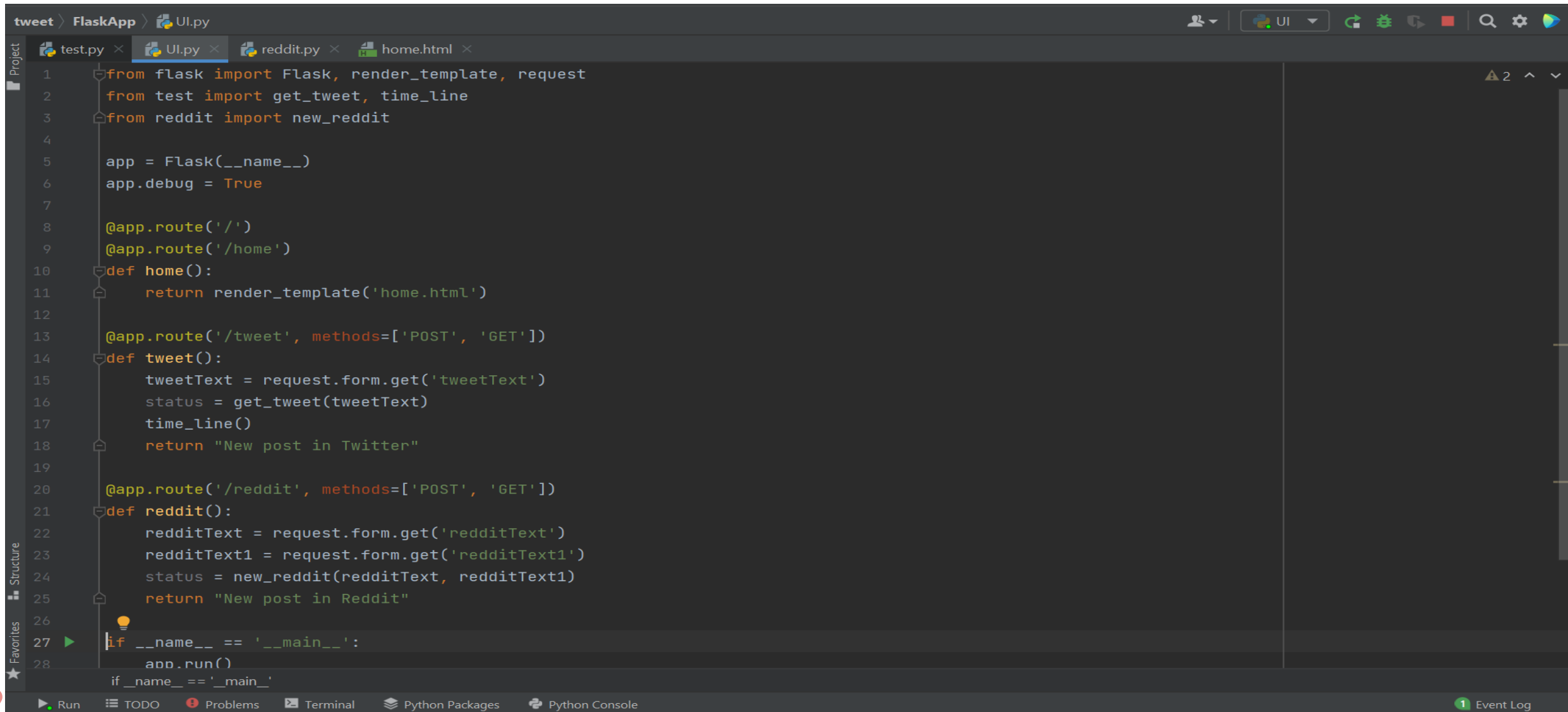
At the bottom right, there is a green circular chat icon.

Web UI – UI.html

```
tweet | FlaskApp | templates | home.html
test.py x UI.py x reddit.py x home.html x
2 <html lang="en">
3 <head>
4 <style> h1 {text-align: center;}</style>
5 <meta charset="utf-8">
6 <meta name="viewport" content="width=device-width, initial-scale=1, shrink-to-fit=no">
7 <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/bootstrap@4.5.3/dist/css/bootstrap.min.css" integrity="sha384-TX8t27EcR56" crossorigin="anonymous">
8 <title>Home Page</title>
9 </head>
10 <body>
11 <nav class="navbar navbar-expand-md navbar-dark bg-dark">
12 <a class="navbar-brand" href="#">Social Media Bot</a>
13 <button class="navbar-toggler" type="button" data-toggle="collapse" data-target="#navbarNav">
14 <span class="navbar-toggler-icon"></span>
15 </button>
16 <div class="collapse navbar-collapse" id="navbarNav">
17 <ul class="navbar-nav mr-auto">
18 <li class="nav-item active">
19 <a class="nav-link" href="#">About <span class="sr-only">(current)</span></a>
20 </li>
21 </ul>
22 <ul class="navbar-nav"></ul>
23 </div>
24 </nav>
25
26 <h1>Create a new post</h1><br><br>
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
html > body > nav.navbar.navbar-expand-md.navbar-dark.bg-dark > div#navbarNav.collapse.navbar-collapse > ul.navbar-nav.mr-auto
```

```
tweet | FlaskApp | templates | home.html
test.py x UI.py x reddit.py x home.html x
28 <div class="container" style="background-color:black; width:1150px; margin:0 auto;">
29 <div class="form-group">
30 <form method="POST" action="{url_for('tweet')}}">
31 <a href="/"></a>
32 <h1 class="p-3">What's happening ?</h1>
33 <div class="input-group input-group-lg">
34 <input type="text" class="form-control" name="tweetText" id="tweetText" aria-label="Large" aria-describedby="inputGroup-sizing-lg">
35 <input class="btn btn-primary" type="submit" value="Tweet">
36 </div><br><br>
37 </form>
38 </div>
39 </div>
40
41 <div class="container" style="background-color:black; width:1150px; margin:0 auto;">
42 <div class="form-group">
43 <form method="POST" action="{url_for('reddit')}}">
44 <a href="/"></a>
45 <h1 class="p-3">New Post</h1>
46 <div class="input-group input-group-lg">
47 <h2>Title</h2><input type="text" class="form-control" name="redditText" id="redditText" aria-label="Large" aria-describedby="inputGroup-sizing-lg">
48 <h3>Text</h3><input type="text" class="form-control" name="redditText1" id="redditText1" aria-label="Large" aria-describedby="inputGroup-sizing-lg">
49 <input class="btn btn-primary" type="submit" value="Post">
50 </div><br><br>
51 </form>
52 </div>
53 </div>
html > body > div.container > div.form-group > form > image > div.input-group.input-group-lg
```


Backend – User Interface

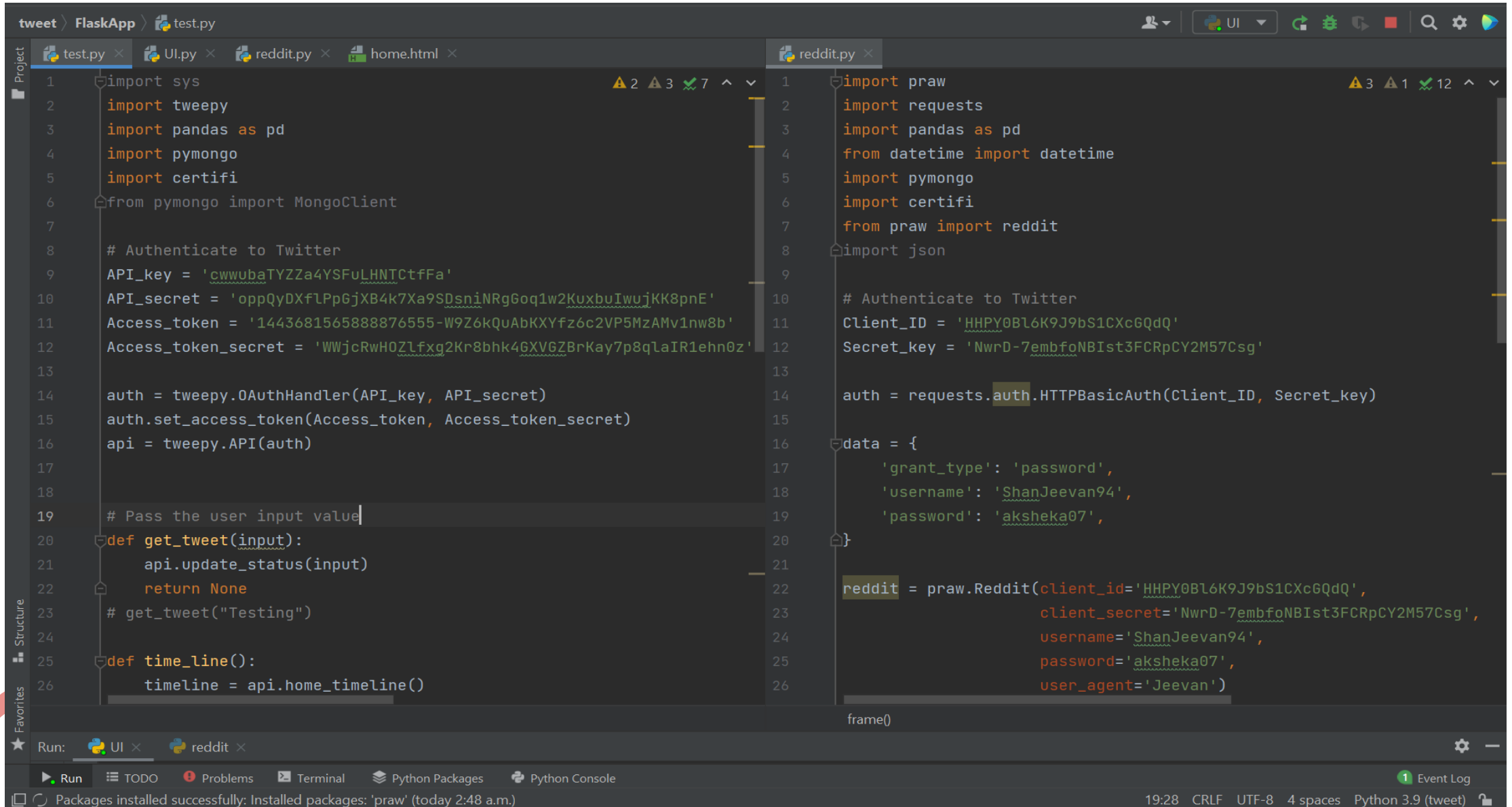


The screenshot shows a code editor with a dark theme. The top bar indicates the project is 'tweet' and the current file is 'FlaskApp/UI.py'. The editor has four tabs: 'test.py', 'UI.py' (active), 'reddit.py', and 'home.html'. The code in 'UI.py' is as follows:

```
1 from flask import Flask, render_template, request
2 from test import get_tweet, time_line
3 from reddit import new_reddit
4
5 app = Flask(__name__)
6 app.debug = True
7
8 @app.route('/')
9 @app.route('/home')
10 def home():
11     return render_template('home.html')
12
13 @app.route('/tweet', methods=['POST', 'GET'])
14 def tweet():
15     tweetText = request.form.get('tweetText')
16     status = get_tweet(tweetText)
17     time_line()
18     return "New post in Twitter"
19
20 @app.route('/reddit', methods=['POST', 'GET'])
21 def reddit():
22     redditText = request.form.get('redditText')
23     redditText1 = request.form.get('redditText1')
24     status = new_reddit(redditText, redditText1)
25     return "New post in Reddit"
26
27 if __name__ == '__main__':
28     app.run()
    if __name__ == '__main__'
```

The left sidebar shows the 'Project' view with a tree structure of the files. The bottom status bar includes 'Run', 'TODO', 'Problems', 'Terminal', 'Python Packages', 'Python Console', and 'Event Log'.

Backend – Web Scraping & Live Status Update



```
tweet > FlaskApp > test.py
test.py x UI.py x reddit.py x home.html x
1 import sys
2 import tweepy
3 import pandas as pd
4 import pymongo
5 import certifi
6 from pymongo import MongoClient
7
8 # Authenticate to Twitter
9 API_key = 'cwwubaTYZZa4YSFuLHNTCtfFa'
10 API_secret = 'oppQyDXfLPp6jXB4k7Xa9SDsniNRg6oq1w2KuxbuIwujKK8pnE'
11 Access_token = '144368156588876555-W9Z6kQuAbKXYfz6c2VP5MzAMv1nw8b'
12 Access_token_secret = 'WWjcRwH0Zlfxg2Kr8bhk46XV6ZBrKay7p8qlaIR1ehn0z'
13
14 auth = tweepy.OAuthHandler(API_key, API_secret)
15 auth.set_access_token(Access_token, Access_token_secret)
16 api = tweepy.API(auth)
17
18 # Pass the user input value
19
20 def get_tweet(input):
21     api.update_status(input)
22     return None
23 # get_tweet("Testing")
24
25 def time_line():
26     timeline = api.home_timeline()
27
28 reddit.py x
29
30 1 import praw
31 2 import requests
32 3 import pandas as pd
33 4 from datetime import datetime
34 5 import pymongo
35 6 import certifi
36 7 from praw import reddit
37 8 import json
39
40 # Authenticate to Twitter
41 Client_ID = 'HHPY0B16K9J9bS1CXcGQdQ'
42 Secret_key = 'NwrD-7embfoNB1st3FCRpCY2M57Csg'
43
44 auth = requests.auth.HTTPBasicAuth(Client_ID, Secret_key)
45
46 data = {
47     'grant_type': 'password',
48     'username': 'ShanJeevan94',
49     'password': 'aksheka07',
50 }
51
52 reddit = praw.Reddit(client_id='HHPY0B16K9J9bS1CXcGQdQ',
53                      client_secret='NwrD-7embfoNB1st3FCRpCY2M57Csg',
54                      username='ShanJeevan94',
55                      password='aksheka07',
56                      user_agent='Jeevan')
57
58 frame()
```

Run: UI x reddit x

Run | TODO | Problems | Terminal | Python Packages | Python Console

Packages installed successfully: Installed packages: 'praw' (today 2:48 a.m.)

19:28 CRLF UTF-8 4 spaces Python 3.9 (tweet) 1 Event Log

tweet > FlaskApp > reddit.py

Project

test.py × UI.py × reddit.py × home.html ×

reddit.py ×

```
28 filename = 'tweet.csv'
29 f = open(filename, "w")
30 headers = "User Name,Tweet,Screen Name, Tweet ID, Posted Date, User
31 f.write(headers)
32
33 for tweet in timeline:
34     print(f"{tweet.user.name} said {tweet.text}")
35     print(f"{tweet.user.name}'s nick name is' {tweet.user.screen_n
36     print(f"{tweet.user.name}'s ID:' {tweet.id}")
37     print(f"{tweet.user.name} created the post {tweet.created_at}")
38     print(f"{tweet.user.name}post image URL is {tweet.user.profile
39     f.write(tweet.user.name + "," + tweet.text + "," + tweet.user.
40         tweet.created_at) + "," + tweet.user.profile_image_url + "
41 f.close()
42
43 # Import CSV Files into Pandas Dataframes
44 df = pd.read_csv("tweet.csv")
45 # print(df)
46
47 # connecting to mongo DB and storing the data
48 ca = certifi.where()
49 client = pymongo.MongoClient("mongodb+srv://Jeevan:aksheka07@clust
50
51 db = client["Web_Scraping"]
52 collection = db["Twitter"]
53 collection.insert_many(df.to_dict('records'))
54
```

Structure

time_line()

Favorites

Run: UI × reddit ×

Run

TODO

Problems

Terminal

Python Packages

Python Console

56:50

CRLF

UTF-8

4 spaces

Python 3.9 (tweet)

Event Log

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

```
subr = 'pythonsandlot' # Choose your subreddit
subreddit = reddit.subreddit(subr) # Initialize the subreddit to a
title = heading
selftext = body
subreddit.submit(title, selftext=selftext)
return None

def frame():
    df = pd.DataFrame()
    Posts = []
    for post in reddit.redditor(RedditUsername).submissions.top():
        df = df.append({"ID": post.id,
                        "Title": post.title,
                        "Description": post.selftext,
                        "Score": post.score,
                        "Timestamp": datetime.fromtimestamp(post.create
        print(df)

    ca = certifi.where()
    client = pymongo.MongoClient(
        "mongodb+srv://Jeevan:aksheka07@cluster0.vqsfw.mongodb.net/myF
        tlsCAFile=ca)

    db = client["Web_Scraping"]
    collection = db["Reddit"]
    collection.insert_many(df.to_dict('records'))
```

frame()

The background features several abstract geometric shapes. In the top left, there is a purple circle and a blue triangle. Below the triangle is a large purple semi-circle and a smaller purple circle. To the left of the semi-circle are two vertical red dashed lines. In the bottom left, there is a blue square and a large purple circle. Above the large purple circle are three red dashed lines of varying lengths and orientations.

Demo



Thank you



<https://youtu.be/xMAynt02igw>