

Client: Dr. Daniel Golberg

Project Members: Sapir Dooley, Chibuike Imo

Project Title: Instagram Web Scraper

I. Introduction

- A. The purpose of this project is to create a Python script that will read an Instagram feed and collect information based on user-inputted criteria. This script is intended to solve the challenge of collecting customer-related information through Instagram at a large scale. With this script, the user should be able to simply run the program; the program will log in to a specified account and collect information, in this case: location data of users that fit a specific set of criteria. In this proposal we shall lay out the specific functions of the script as well as the criteria that will drive data collection. With this script, the user will have their stresses alleviated in regards to the marketing aspect of their company.

II. Motivation

- A. For this specific instance of this code, we are interested in assisting the marketing team for the company TK Paddles, LLC. This company produces custom wooden wall decorations, or “paddles”, for college students in the Fraternity and Sorority community. There are thousands of Greek organizations in the United States and each organization may have tens to hundreds of chapters. In many of these organizations, it is common to give a paddle to either big/little brother/sister or the organization's new member educator each semester. This will not only be helping this specific company's marketing team out, but it will also be helping us figure out a newer way to tackle marketing in a newer age of technology. As a team we will not only be solving this for our current client, but also solving this to meet the needs of our future clients to come.

III. Functions

- A. The first function of this script will be to open a browser and log in to an instagram account. This will be achieved using a simple web driver that will allow us to simulate a web browser. Next the code will load the feed by scrolling down the page a set number of times. After this the code will read the caption of each individual post, looking for specific keywords. If a post is found, the script will like that post and search it for a location tag. If no location tag is found the script will open the users profile and search for terms that the user can infer location from. After this is completed, the script will save the URL for the post.

IV. Criteria

- A. While each organization is different, generally the terms used are the same. When a class of new members is presented to the public, the terms “crossing”, “presentation”, “probate”, “rollout”, and “Neophyte” are used just to name a few. When most of the followers of the account are Greek, it is unlikely that these terms would be used elsewhere. This criteria will help the script determine what posts are most important as they will likely relate to a class that has just joined and therefore have not purchased a paddle yet. The terms that will help us determine location will be university names or acronyms as it is highly likely that is where a paddle would be shipped were it purchased.

V. Management

- A. Roles and responsibilities

1. Sapir Dooley (Lead Developer and Co-Project Manager)

- a) As Lead Developer, Sapir will be directly responsible for engineering the code for the project with the assistance of his Assistant Developer. All code will be coded using the coding platform Visual Studio Code and the front-end and back-end coding language being utilized is fully Python. As Co-Project Manager Sapir will make sure that both him and his counterparts working on the project are working diligently and communicating well in order to make this project successful.
2. Chibuike Imo (Assistant Developer & Co-Project Manager)
 - a) As Assistant Developer, Chibuike will be responsible for debugging and beta testing the code to make sure the code executes successfully and meets the needs of the desired project functionality requested by the client. His role is important as he will be making sure that the project also aligns with a set timeline that is agreed upon with the client. As Co-Project Manager, Chibuike will also be making sure that all communications run smoothly and all operations are executed to make the project a success.

VI. Budget

- A. The budget of this project can be explained to be \$0. Visual Studio Code is a free platform that has zero costs to be able to code and meet all the project requirements/deliverables. The python extension within Visual Studio Code is also free so that contributes to the low project budget. Utilizing the instagram platform and running an instagram account is also free of charge, so we've avoided high expenses in regards to the platform integration as well. Our team's means of communication is fully through direct iMessage via Apple iPhones and since both project members are local residents of the U.S. they do not have to worry about paying any extra fees to communicate as a team at all so that is a saved expense. With expenses kept to a minimum we really cut several costs.

VII. Completed Project Data and Methods

- A. Work in progress:
 1. In this project, data will be collected directly from the Instagram website. The completed project will collect the following data from each and every post that matches the user criteria, if available: the username of the original poster, a link to their profile, a link to the original post, the caption of the post, the date of the post, any hashtags, the location of where the post was tagged, the location of the poster, the university of the poster, and the greek organization of the poster. Data will be stored in an excel spreadsheet.
 2. As of November 13th, there is no more work in progress. If there was a continuation or addition to this project it'd be an ongoing implementation of more code to make sure the software runs seamlessly. Another potential for additive work could be if the client desires to have the software have more functions to better meet the company's expansion or cumulated work load of marketing.
- B. Work Completed
 - a) Sapir

- b) As lead developer, Sapir has implemented all the coding that is required to get everything to function well. The Instagram web scraper now can fully function and it runs smoothly. The functionality of the code is as follows: When running the code, the client will have the instagram web scraper program log into instagram using their login data embedded in the code. Their login information will be secure and specific to only the client while being very confidential/secure. The client's code then proceeds to scroll down a designated quantity amount of times that the client desires. So if the client wants the program to scroll until it gets through every single one of his followers in his database it surely can. The bot being used in the program obtains every link to every single post that it scrolls through in the process leaving no room for misplaced data. The program and bot saves the username and profile information of the author, the date it was published, the caption the author titled the post, any hashtags the author decided to add within the post, and any locational data the author added as well while using his geolocation. The program makes sure to filter the entire caption for specific indicated keywords chosen by the client to decipher, it filters likes and comments to see if there are any keywords indicated by other customers of the client or within the users of the social media database. The program distinguishes if the post has keywords and a location tied to it and automatically instantaneously searches google for the latitude, longitude, and specific coordinates of the post by the user . If it matches any of the keywords that the client has designated, then the bot proceeds to like the post, saves the post in the database, and then finally stores information in a JSON file.

2. Chibby

- a) As assistant developer, Chibby has focused on beta testing and making sure that all functionalities of the program/bots are ironed out and ready for the client to receive. Chibby has tested out the code twice and has given his feedback to lead developer and co-project manager Sapir. Chibby has tracked how many times the team has met together to gameplan and update each other on deliverables and work in scope, while making sure that deadlines are met so that the client has the best possible program in his hands.
- b) Chibby assisted Sapir with compiling a presentational video to present to the client Dr. Daniel Goldberg.
 - (1) The presentation consisted of presenting the entire project to Dr. Goldberg and the rest of his team at GEOG 476. The presentation was seamless and was a duration of ten to twelve minutes in length. There were some inconsistencies with our listed functionality and we were completely transparent with the client in telling him that we have completed the majority of the heavy lifting and all that was left were some final touches. Sapir

and myself came to the conclusion that our client would completely understand that we were working diligently and we desired to deliver a product that the client would not have any errors or see any flaws in.

VIII. Conclusion

- A. The purpose of this project is to scrape an instagram feed for specific keywords and save those posts so that someone can go back and easily identify potential customers and their general locations. This information is valuable to the company so that the marketing team can know who needs to be reached out to, what regions are most likely to benefit from advertising, and what regions are not being touched. The company will have their marketing pains alleviated and they will be able to have a system that works for them to not only maximize their profits but also grow their social media presence as large while leveraging their Instagram profile. In this technology age this is a huge business move being made in order to guarantee the success of their company in a competitive age of online companies like theirs.
- B. The company received notification of the completed application of our code on November 14th. This code is fully functional and meets all the needs of the client. Upon our completion of this code it is a common desire to send a version of the application/code to the client so that they will have the code to start practicing the functionalities and capabilities with their company to better affect their marketing plans for the future. Sapir and Chibby beta tested all of the app functionality as if putting ourselves in the Client's shoes so that every single aspect the client would possibly try to utilize in the code would work for them and not bug out at all. Of course it is expected that there will definitely be comments from the client on the functionality of the code and if the software is actually working properly and executing all the functionalities the client desires, but this is normal. Sapir and Chibby are convinced that if the client is not happy with the product they will assist in any way possible to meet the needs of the client's commentary on the execution of the code. This is very important when working with a client that is not coding the application himself, because sometimes the desires may not be fully transferable to the actual code functionality itself.
- C. Overall this project was a success and the client will love the ease it will bring to the marketing scheme overall. In the future Sapir and Chibby definitely have learned from past mistakes and hiccups down the road how to make the job get done more efficiently and diligently to meet the end goal at a quicker pace. Our team will be readily accessible at any time to not only our client but any future clients that may want anytype of assistance to their companies as well.