Client: Dr. Daniel Golberg

Project Members: Sapir Dooley, Chibuike Imo

Project Title: Instagram Web Scraper

1. Introduction
   1. 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.
2. Motivation
   1. 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.
3. Functions
   1. 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.
4. Criteria
   1. 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.
5. Management
   1. Roles and responsibilities
      1. Sapir Dooley (Lead Developer and Co-Project Manager)
         1. 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 and Co-Project Manager )
         1. 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 any operations that need to be executed to make the project a success he catalyzes directly.
6. Budget
   1. 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.
7. Conclusion
   1. 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.