COMP4801 Final Year Project 2023-2024
Interim Report

Name: Mak Ho Chit Terry

UID: 3035785764

Supervisor: Professor Zhiyi Huang

Second Examiner: Professor Giulio Chiribella

Abstract

This progress report presents a comprehensive overview of the ongoing efforts and development stages of the E-Commerce Analytics and Product Monitoring System project. This project was designed to address the inefficiencies in the market of business solutions. The main focuses of the project are real-time data extraction, automated product tracking, and integrated analytical tools for third-party sellers operating their businesses on Amazon. The project aims to enhance operational efficiency and optimize business decisions for sellers.

Currently, the project is undergoing the development phase, where a website is being hosted for the system. Due to unforeseen challenges, the design and implementation of the system have been amended to ensure that all expectations can be met. Further details will be provided and discussed in this report.

In the coming weeks, the design of the website will be finalized. Tools and functions will be set up and connected to the website and its respective database. Data obtained will be injected as demo data to the first draft of the platform for testing purposes.

Table of Contents

List of Figures

Project Objective

The objective of the project is to develop a one-stop E-commerce solution that specializes in real-time data extraction, automated product tracking and integrated analytical tools. The system will be implemented as a website that is accessible to anyone. These features are currently lacking in the marketplace, as most of the tools are designed to be either for capturing or analyzing data. To be accurate, below are a list of objectives this project will cover:

1) <u>To develop a One-stop E-commerce Solution</u>
   The website will contain all essential tools for daily operations, so that users do not have to switch between different software for specific services.

2) <u>To integrate Analytical Tools</u>
   Analytical tools will be integrated for users to analyze data obtained from other functions of the system, or any data will the same structure as the tool.

3) <u>To extract Real-time Data</u>
   Real-time Data will be extracted to guarantee the authenticity and accuracy of extracted data. This is to ensure that users get the most updated and accurate data instead of approximated or inaccurate data.

4) <u>To automate Product Tracking</u>
   Product Tracking will be automatically executed on user-defined time frame for users to monitor specific products. A list of a large number of products can be tracked, data obtained will be stored in either the website's database or as an output file.

5) <u>To ensure Scalability and Evolution</u>
   The system will be designed with the ability to handle more users in the future and edit the parameters for attribute searching. This allows users to have more precise and tailor-made results, and future improvements and updates can be made without erasing the current structure.

6) <u>To design with a User-Centric approach</u>
   The website will be designed to be user-friendly so that users can extract the full potential of the system. Instructions will be clear and actions will be designed to be as simple as possible.

Background

Third-party sellers are independent sellers who offer a variety of new and used products to customers of other businesses [1]. These businesses are usually small in scale when compared to traditional companies as the demand for manpower of E-commerce is less than traditional businesses due to the digitalized business structure. A report from Amazon reveals that more than 60% of sales in Amazon's store are from these independent sellers [2], proving the dominance and importance of third-party sellers. To maximize profit margins, third-party sellers require large amount of accurate and updated data to keep track of product performances, as well as to gain insights into other competitor products. Since third-party sellers are often small businesses, it is unlikely for them to design, create, and maintain software and programs to suit their needs. Currently, independent sellers monitor product performances either by manually tracing and collecting data, or subscribing to paid services to collect data from their platforms. However, these subscription services all contain flaws. Some provide inaccurate data due to the use of approximation and prediction of data trends instead of directly extracting data, some not user-friendly, and some others lack crucial functionalities for tracking and monitoring purposes [3]. Thus, it is believed that by creating a project that 1) provides accurate, real-time data, 2) automates the product tracking process, and 3) provides useful analyzing tools, users can increase efficiency and have a competitive advantage over other competitors in the market. This will ultimately improve business efficiency and bring upon higher revenue streams to users.

Methodology

The methodology of this project has remained hugely the same as the project plan. It is proposed to develop three principal functionalities, namely the Search Page, the Automated Worker, and the Analytical Page. The Search Page is a dedicated function that enables users to search specific product attributes instantly. The Automated Worker is a system designed to autonomously monitor selected products and capture relevant data. The Analytical Page is a dedicated space equipped with tools to visualize and interpret the collected data. These three functions will be the backbone of this project as users will be able to speed up daily operation and increase efficiency by incorporating the use of these three functions into their operations.

However, some changes were made to various parts as the original plan had some fatal problems. This section will only mention the changes, while the findings will be documented in the next section.

1) ScraperAPI will be used in the Scrapy codes for rotating user agents in the current version. However, if time allows, a rotating agent will be developed and replace ScraperAPI since the free subscription cannot afford the daily requests from the website.
2) Github Actions will not be used for automation. Instead, a server will be hosted to handle requests from the website.
3) Retool will not be used for the development of the website. A website has been created instead as replacement.
4) SQL Database will be used to store data entries instead of PostgreSQL.

Accomplishments (First Semester)

In October, the research and planning stage was completed. All tools and software were selected after thorough research. The project timeline was updated to allocate more work to November and December so that more time can be allocated for the development and testing stages.

In November, the proposed tools and software were set up and tested to see if the required features function as expected. Two web scraping functions were created, one for tracking specific product details, the other for tracking product rankings. They were then finetuned to meet the needs of the specific users of this website (third-party sellers on Amazon). It was discovered that the Scrapy functions cannot crawl data successfully without rotating the user agents as Amazon would identify the Scrapy Spiders as robots and block them. Therefore, ScraperAPI, an additional user-agent rotating software was used and implemented into the functions. However, the free plan only contains 1000 credits per month, while every webpage being crawled by Scrapy Spiders would consume around 25 credits. This hugely limits the accessibility of the webpage as the functions would not work after the credits were all consumed. Therefore, if time allows next semester, a rotation agent will be attempted to be created to eliminate the use of ScraperAPI.

Usage

0/ 5 concurrent requests

965/ 1,000 API credits

Resets on Sun, February 4, 2024

Figure 1: ScraperAPI provides very limited credits with a free plan subscription.

In December, the UI design for all three functions was drafted. The PostgreSQL database, along with the Retool account, was created and set up. The web scraping functions were updated so that outputs were stored in the database instead of showing in the terminal. They were then uploaded to Github Actions for testing the automation process. It was then discovered that Github Actions can only be scheduled to execute once every 5 minutes, instead of executing whenever it is being called. Also, after being scheduled, an Action will be executed around 20 minutes later due to the fairness of resources distributed by Github. This implies that users might have to wait for more than 20 minutes after they press the "Search" button to get the results, which is unreasonable. Thus, a server will be created to execute the Scrapy functions instantly upon request. Github Actions will be abandoned. In addition, Retool was discovered to provide limited UI options. The pages developed are crude and cannot achieve aesthetic purposes. Furthermore, it was discovered that users are unable to access a Retool page unless they login with the same credentials as my Retool account, which is not what is expected. Thus, Retool, along with the PostgreSQL account, were abolished. A website has been hosted as replacement using InfinityFree. An SQL database was instantiated and connected to the website. The website is called AmaSight, with the meaning of bringing Amazon Insights for users. The SSL certificate has been acquired, making the site more secure and safe from cyberattacks.

```
1    name: Run Scrapy Spiders
2
3    on:
4      schedule:
5        # * is a special character in YAML so you have to quote this string
6        - cron: "16 15 * * *"
7                        Runs at 15:16.
8    jobs:              Actions schedules run at most every 5 minutes using UTC time. Learn more
9      build:
10       runs-on: ubuntu-latest
```

Figure 2: Github Actions cannot run instantly on request.
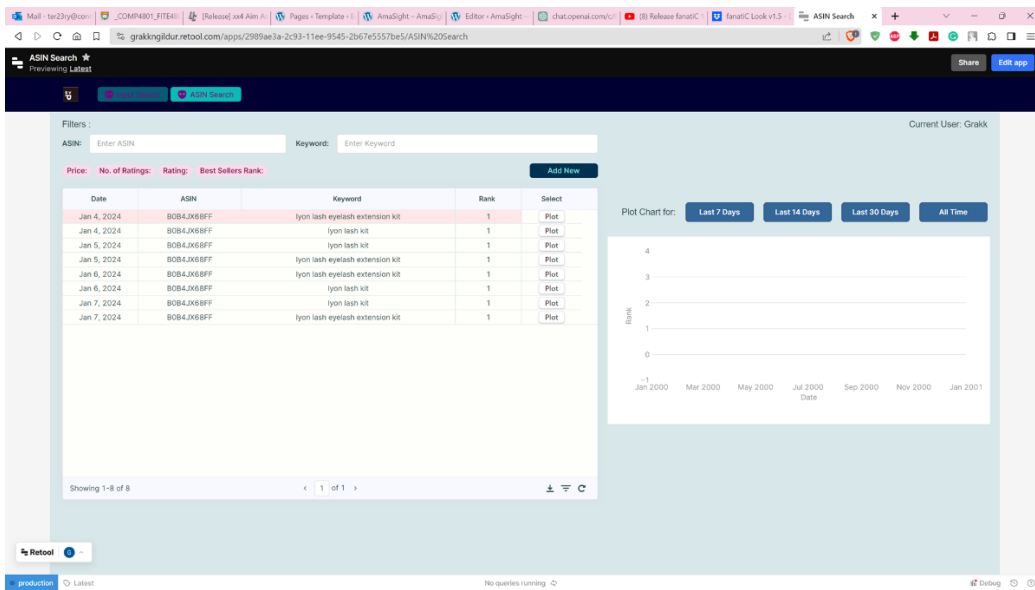
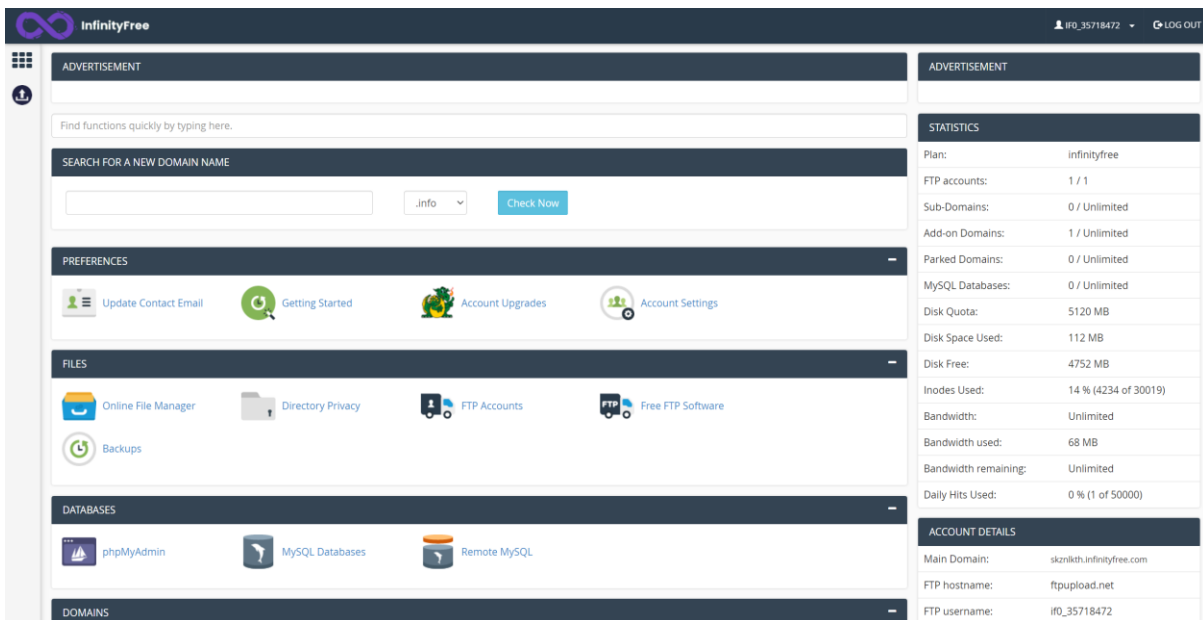*Figure 3: UI options in Retool are very limited without the option to build a website instead of a portal.*

Figure 4: The website is currently being built step by step.

Future Work (Second Semester)

In the coming weeks, the structure of the website will be constructed. Both frontend and backend will be set up to make the website functional. After this is done, the proposed project schedule will be adopted: It is hoped that all three pages can be developed by the end of March, leaving April for testing and making improvements. By the end of April, it is hoped that the public version of this website will be constructed and published.

Conclusion

The E-Commerce Analytics and Product Monitoring System represents a pioneering initiative for third-party sellers on Amazon. It aims to provide all-round support for their daily operations, and also to enhance their user experiences. In this Interim Report, the background and methodology have been stated. They are mostly unchanged when comparing with the project plan, except for multiple areas where the expected methods did not work out during testing stage. In the past month, progress have been made in improving the usability and reliability of the website by abandoning existing selection and replacing them with better options. With the changes made, it is hoped that the website will be able to assist users and bring great value to them.

References

1) Aftership. [Online]. Available: https://www.aftership.com/glossary/third-party-seller#:~:text=A%20third%2Dparty%20seller%20is,products%20on%20the%20Amazon%20marketplace. [Accessed: 25-Oct 2023].

2) Aboutamazon. [Online]. Available: https://press.aboutamazon.com/2023/5/amazons-u-s-small-business-empowerment- report-shows-continued-sales-growth-for-independent-sellers. [Accessed: 25-Oct- 2023].

3) Akashdeep, "Helium 10 vs Jungle Scout: Unbiased comparison of 2023!," Demandsage.com, 08-Aug-2023. [Online]. Available: https://www.demandsage.com/helium-10-vs-jungle-scout/. [Accessed: 28-Nov-2023].