

Midterm Progress Report

Please report your progress: by midterm, you should have API fully working, data set collected. Please report how you meet your milestone (setting up a database, UI development, etc.)

API Access

For API Access, we were required to create a Walmart account and retrieve the consumer ID. We are also required to create private keys and hash the keys.

Challenges

This API is not like other APIs that are easy to use. It actually needed us to hash the keys ourselves and write lots of code in order to access it.

Dataset Collection

We have collected two types of data. First is taxonomy data. The API returns a nested JSON output. We also found out that the depth of the data is 7, which means that the data's largest nested depth is 7. Having this depth is hard to store into a tabular data structure and we are trying to store it into a relational database like MySQL, so we also need to flatten the nested JSON file to only one layer.

Another type of data is paginated data. Each API call will return a page of Walmart products and the next page's URL. We iterated the process and scraped 5181 pages of products and each page has 200 items. So we have over a million products.

Challenges

We actually didn't have any problems here, but the code did take a long time to run. Also, it's hard to temporarily store the data, so we stored the data on Google Drive for now.

How To Meet Our Milestone

Setting Up Database

We decided to use MySQL to store our data instead of previously said Cassandra since we are more familiar with relational databases and SQL.

New updates: We decided to switch to NoSQL databases like MongoDB

We decided to finish importing JSON files into MongoDB by the end of March.

Challenges

We don't have experience importing large amounts of data into databases before, so it could be a challenge for us.

New updates: We have tried MySQL, but MySQL was very slow to import and very slow to query. So we decided to use MongoDB. MongoDB only offers 512MB for the free tier, which is much smaller than what we need, so we might want to upgrade it or do something else accordingly.

UI Development

For now, we are thinking of using the Vue.js framework in python and flask to build our UI. We are thinking of building a searching UI for better searching and filtering options.

Challenges

We both don't have experience with UI development, so Vue.js and flask are new things for us to learn.