Final Project Proposal

**Data Source**

In the final project, I am planning to analyze data to help user decide which cleaning product is the most suitable choice for him/her. I plan to crawl data from Amazon on Zep products (multiple pages), and I will collect the product name, price, rate, rate count, product URL link as my raw material to analyze.

**Data Process**

After crawling data from amazon and writing to a csv file, I will group data into a tree structure. The parent level is clean products usage, and the next level will be product type. Each product will have its own attributes such as price, rate, rate count, etc.

**Data Presentation**

I will use command line interface to interact with users. Specifically, I will give users the best choice by asking for their willing about the usage of the product (hand cleaning, general disinfection, heavy-duty cleaning, etc.), and the price range, rate requirement and rate count. They can also specify the type of cleaner they want, like spray or gel. Then I will give users a set of suitable products name and their URLs. I will also try other flasks if time is enough.

**Progress**

By now, I have written a python script to crawl data form Amazon. And I have already collected enough data, however, the script does not work every time. Sometimes the website blocks my request.