**DRUG RECOMMENDATION SYSTEM PROJECT**

**Motivation:** to further solidify the fundamentals of Python through web scraping and EDA as well as develop skills in analytics/story telling by building dashboards. Through the process, I hope to learn how to efficiently write Python code, understand the process of unit testing and development in general.

**Why a Drug recommendation system**

When you go to the pharmacy and ask for medicines that your doctor has prescribed for simple health problems like allergies, cold sore, or body ache it’s inconvenient if the pharmacy don’t have the same medication you’ve been prescribed. It’s often a hassle to then reach out to your doctor for alternatives, so it would be useful to have a system that can recommend similar OTC or prescribed drugs to solve this issue.

WebMD’s database has authentic content for the drug components of several medicines as well as medical news. Each medicine has reviews which can be extracted and analysed, with the possibility of using it for alternative recommendations.

WebMD has a lot of info about many kinds of conditions and medicines to use, but it would be good to have a place where you can just see:

* Alternative medicines for simple conditions
* Similarities between medicines for each condition
* What the medicine looks like
* Reviews of the medicines

**Plan**

Project subtasks:

* Web scraping
* EDA
* Data visualization

Do research on each subtask

Build subtasks, one step at a time

Put subtasks together

**Duration:** 3 weeks – 4 weeks (checkpoint: 1 day per week for the first 2-3 weeks and 2 days per week for the last week)

**Documentation/Communication**

GitHub Repo containing code, documentation, and dashboards

**Action – first week**

* Search WebMD for drugs recommended for 3 different health problems:
  + Backache
  + Cold sore
  + Allergy
* Extract **key info**:
  + Drug description
  + Reviews for each drug
* Use cosine similarity between drugs or similar drug components to build recommendations?
* Use **streamlit** to build the system?

**Friday – 21/10/2022**

* Created a local repository and pushed it to GitHub:
  + In terminal:
    - Initialized Git repo = **git init**
    - Staged all changes made in local repo = **git add** . + **git commit -m “First commit”**
    - Pushed existing repo into empty GitHub repo that was made = **git remote add origin + url** + **git push origin master**
    - Link *Local* master branch to *remote* master branch = **git push –set-upstream origin master**
    - To push changes to GitHub:
      * git add .
      * git commit -m “Commit message”
      * git push
* Read on Selenium, its tools and how to use it
* Make note on the type of drugs for each health condition

**Medicines**

* **Allergy**
  + Antihistamines & decongestions:
    - OTC
      * Cetirizine
      * Benadryl
      * Fexofenadine
      * Levocetirizine
      * Loratadine
      * Diphenhydramine
      * Chlorpheniramine
      * Xyzal
      * Zyrtec
      * Zyrtec-d
* **Backache (NSAIDs)**
  + Ibuprofen
  + Naproxen + Aleve
  + Acetaminophen (paracetamol) + Tyleon
* **Cold sore**
  + Docosanol (OTC)
  + Acyclovir
  + Famciclovir
  + Valacyclovir

**Monday (24/10/2022)**

* Used selenium to extract names and links of drugs recommended for allergies (both OTC and prescribed)
* Plan:
  + Extract all relevant URLs into a csv file
  + Find the reviews section, filter the reviews to show only simple health problems, extract the review data and iterate this over every link

**Tuesday (25/10/2022)**

* Scrape over links that have been collected
* Extract reviews of drugs and store into a dataframe/file

**To produce dataframe of drug + visualization**

**Tuesday (2:45pm) next week**