

TEAM SNAKE:

Tiffany Huang, Jack Jacobs, Sara Khoshhal, Andrew Nunn

Matching Engine for Cancer Clinical Trials

ABSTRACT

One of the most important decisions pharmaceutical companies must make concerning a clinical trial is the geography in which to locate it. Several factors influence this decision, including the prevalence of the given disease in the region, the demographics, the hospital's capacity, and the saturation of ongoing trials already in the area. With these factors in mind, we created an engine to match cancer clinical trials to potential hospital partners. More specifically, this application allows the user to choose from a list of the top ten deadliest cancers in the United States, with the option to indicate a target sex, age, or race group, and returns a list of the top ten hospitals in the United States that match the chosen criterion. To construct this program, we used basic iterations and callable functions as well as advanced Python features and packages (like Request and JSON) to sort through API data from the US Census and CDC Data in the form of CSV to rank counties that matched the desired demographics and disease prevalence rates. Furthermore, we used web scraping methods (Selenium) to retrieve data from the US Clinical Trials website, ranking hospitals based on hospital experience and trial saturation. The county rankings and hospital rankings are then combined to output a list of the top ten hospitals which best match the user criterion. With this application, we hope to create a more efficient method to match clinical trials with hospitals and promote better distribution of resources and research to areas that may need it most.

USER INSTRUCTION

Before Running the Script

Download the zip file called “FINAL FOLDER” containing all the required scripts and files. Extract all the files into a designated folder and ensure the working directory in Python is set to that same folder. Before running the program, check that you have Anaconda Navigator 2.0.3 downloaded, as well as the required package Selenium. You can install Selenium by typing “conda install selenium” into your console. After installing Selenium, restart your kernel.

Additionally, the program requires the ChromeDriver program compatible with the user's version of Google Chrome. You can download ChromeDriver at:

<https://chromedriver.chromium.org/downloads>'. To check your version of Google Chrome, click on the 3 vertical dots at the far right side of your Chrome toolbar, hover mouse pointer over the ‘Help’ menu, click ‘About Google Chrome’, and it will display your Google Chrome version. Ensure that the ChromeDriver is saved into the same folder as the rest of the script and data.

Check that the following scripts are included in the folder:

- clinical_trial_matching_engine.py
- census.py
- census_score.py
- cdc.py
- hospitals.py
- hospital_score.py
- clinical_trial_search.py
- chromedriver.exe

Refer to the Documentation to check that the required CSV files are also downloaded. Open the script called _clinical_trial_matching_engine.py and run the script.

Running the Script

First, indicate whether you would like to continue to the main menu by typing “1” or if you would like to quit by typing “0”.

Next, from the list of cancers provided:

- 1) Female Breast Cancer
- 2) Corpus and Uterus, NOS
- 3) Ovary
- 4) Prostate

- 5) Lung and Bronchus Cancer
- 6) Colon and Rectum
- 7) Non-Hodgkin Lymphoma
- 8) Leukemias
- 9) Pancreas
- 10) Liver

Choose the cancer of interest by typing its corresponding number (“1-10”) into the console. Type “0” if you would like to quit.

Given the following age ranges:

- 1) Under 18 years
- 2) 18 to 64 years
- 3) 65+ years
- 4) All ages

Choose a desired age range by typing its corresponding number (“1-4”) into the console. Type “0” if you would like to quit.

From the following race groups:

- 1) All Races and Ethnicities
- 2) White
- 3) Black
- 4) American Indian and Alaskan Native
- 5) Asian and Pacific Islander
- 6) Hispanic/Latino

Choose a desired race group by typing its corresponding number (“1-6”) into the console. Type “0” if you would like to quit.

Given the following sex groups:

- 1) Male and Female
- 2) Male Only
- 3) Female Only

Choose a desired sex group by typing its corresponding number (“1-3”) into the console. Type “0” to quit.

Finally, allow time for the script to run. Once the search is complete, the top 10 hospitals will be printed in the console with the hospital’s corresponding phone numbers. Additionally, the

program will save an excel file of the top 20 hospitals into your folder, providing more information regarding the score breakdown and a website link. To retrieve the excel file, go to your folder and find the excel file named “ranked_hospital_partner_prospects_[date]_[time]”.

If you would like to run another search, go back to the console and enter “1”. If you would like to quit, enter “0”.