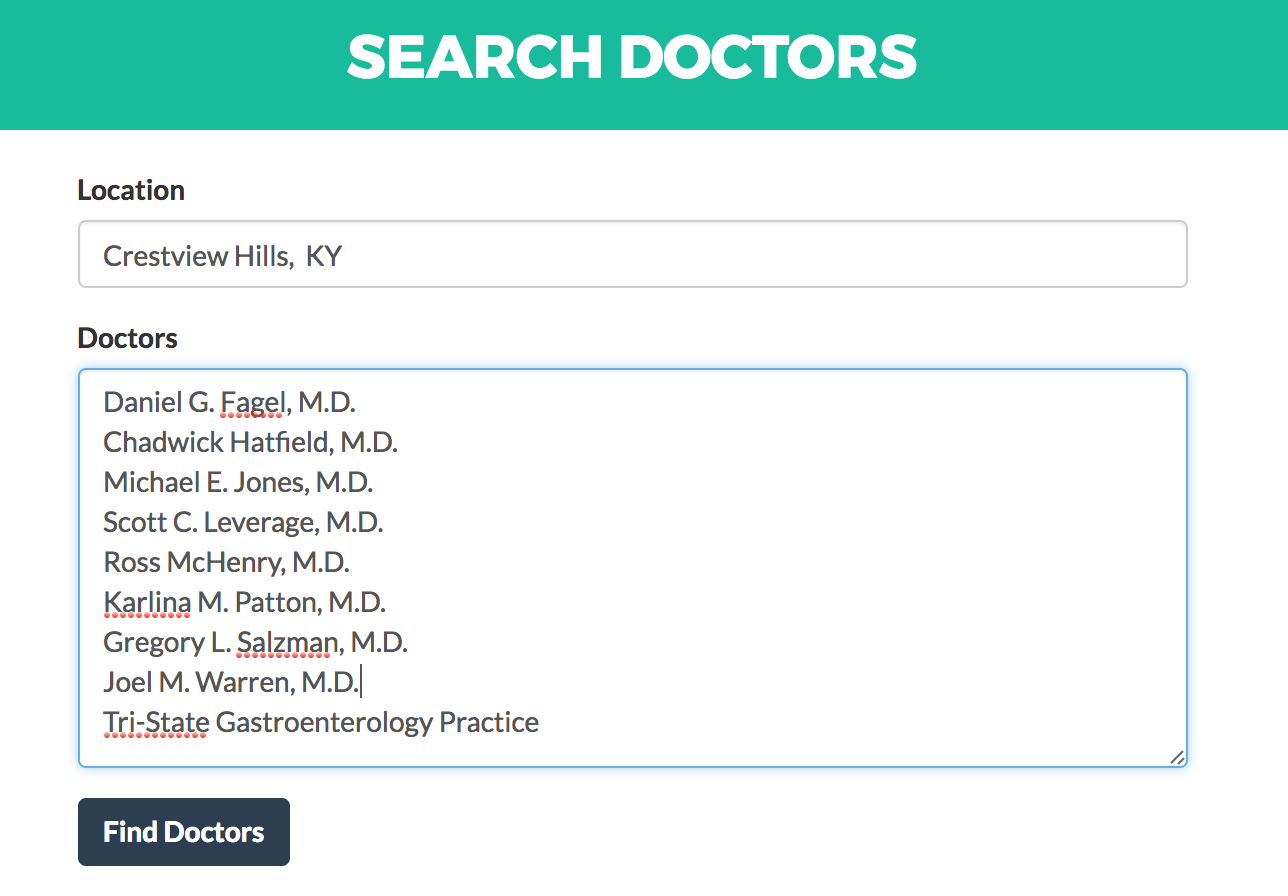
# Suggested Scrape Process

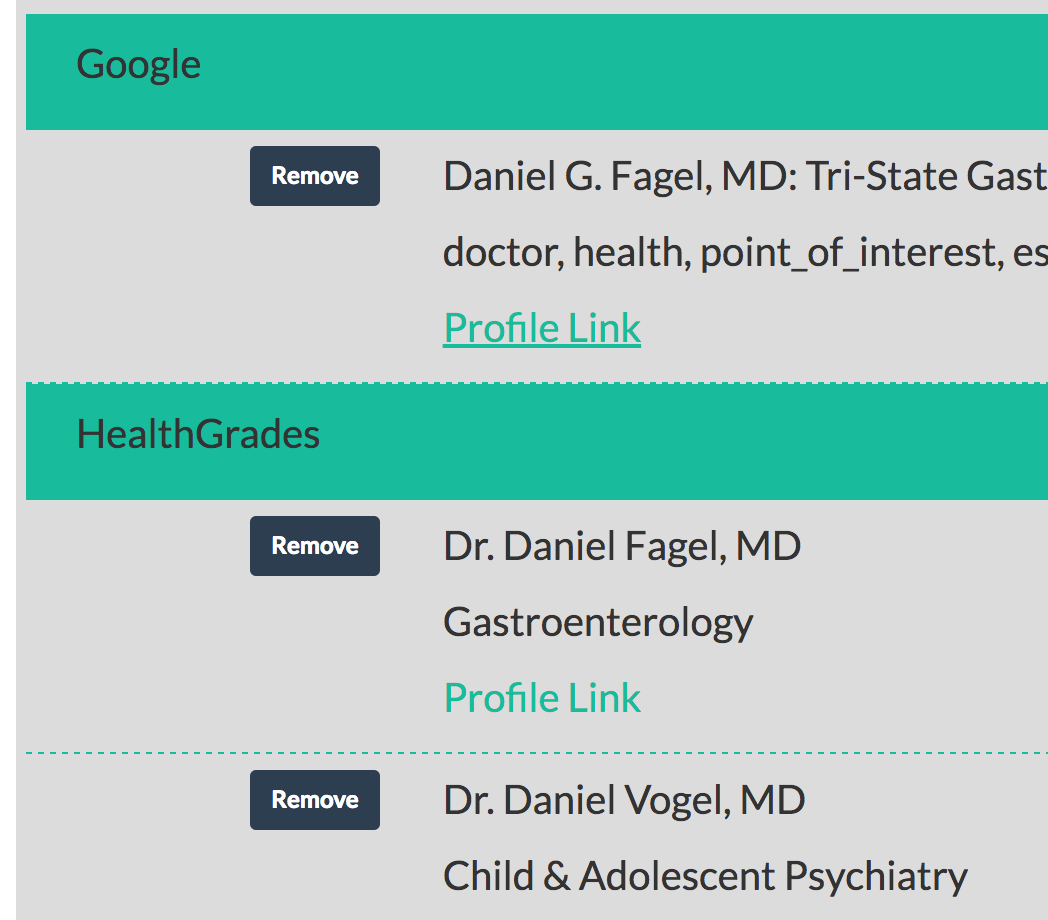
This document details the “suggested” process to follow to get the most relevant data out of the scrape functionality in the ConnectedMD application. The goal of the application is to provide enough search flexibility and features that this document will become unnecessary.

1. **Do an initial search.** Use the application to do an initial search for doctors in a specific city/state. <http://connectedmd.quantumconnect.com/doctors/bulkadd>

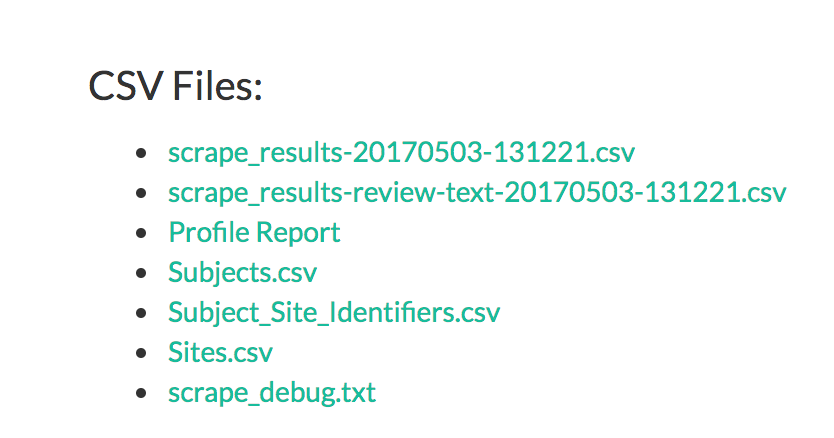


NOTE: to get more profiles returned from this initial search try adding the doctor names without “M.D.” or middle initials

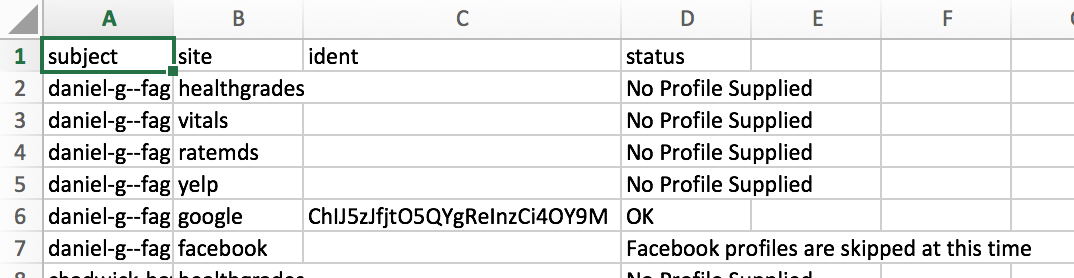
1. **Review the search results.** Take the time to “Remove” irrelevant profiles, leaving only the profiles you want to include in the initial scrape.

****

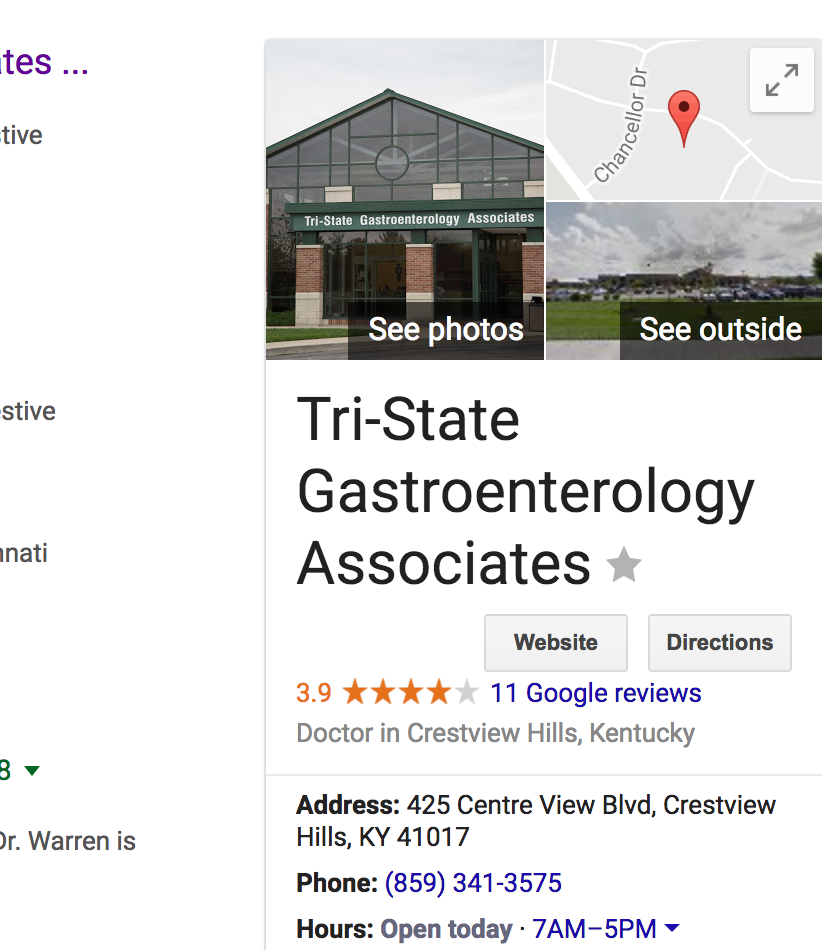
1. **Run the scrape on the chosen profiles.** When you have reviewed the profiles you want to include in the initial scrape, click on the “Scrape Profiles” button.
2. **Review the files created during the scrape.**



* **Profile Report**: this csv file will look similar to the Subject\_Site\_Identifier.csv except it will contain a “status” column containing success/failure information. You can use this file to determine the success/failure of your scrape. It can also be used to identify what profiles you will need to manually find.



1. **Download the Subjects.csv and Subject\_Site\_Identifiers.csv.** When you run a scrape from the search functionality, it will automatically create a correctly formatted Subjects.csv and Subject\_Site\_Identifier.csv file. The Subject\_site\_identifier.csv file is guaranteed to have at least 1 line for every subject/site combination. You will easily be able to tell what is missing and what you need to do search for manually.
2. **Manually add profiles if necessary.** If you need to add profiles manually, you will need to add a line to the Subject\_Site\_Identifiers.csv file. You will need to make sure you add the following on each line:
   1. A valid “subject\_key” that matches a “subject\_key” in the Subjects.csv file.
   2. A valid “site\_key” from the list (ratemds, healthgrades, vitals, google, yelp, and facebook
   3. A properly formatted “site\_subject\_ident” value for the site profile you are adding. In all cases except for Google, the “site\_subject\_ident” value is the portion of the URL after the main site url. For example, the site\_subject\_ident value for the RateMds profile “<https://www.ratemds.com/doctor-ratings/3640635/Dr-Tejas-Patel-Ahmedabad-GJ.html>” is “/doctor-ratings/3640635/Dr-Tejas-Patel-Ahmedabad-GJ.html”. Examples for each site are listed below:
      1. Ratemds**:** /doctor-ratings/3640635/Dr-Tejas-Patel-Ahmedabad-GJ.html
      2. Healthgrades: /physician/dr-aaron-jarrett-y9pvx4z?cc=MTPSR-1
      3. Vitals: /doctors/Dr\_Joe\_Howell.html
      4. Yelp: /biz/pinball-jones-fort-collins
      5. Google: SEE BELOW
      6. Facebook: /focopinball/
3. **Adding Google Profiles.**  Adding google profiles manually requires a little more effort. When you are going a search for something in google, you will notice sometimes on the right side that a “profile” of sorts shows up pertaining to the search you are conducting.

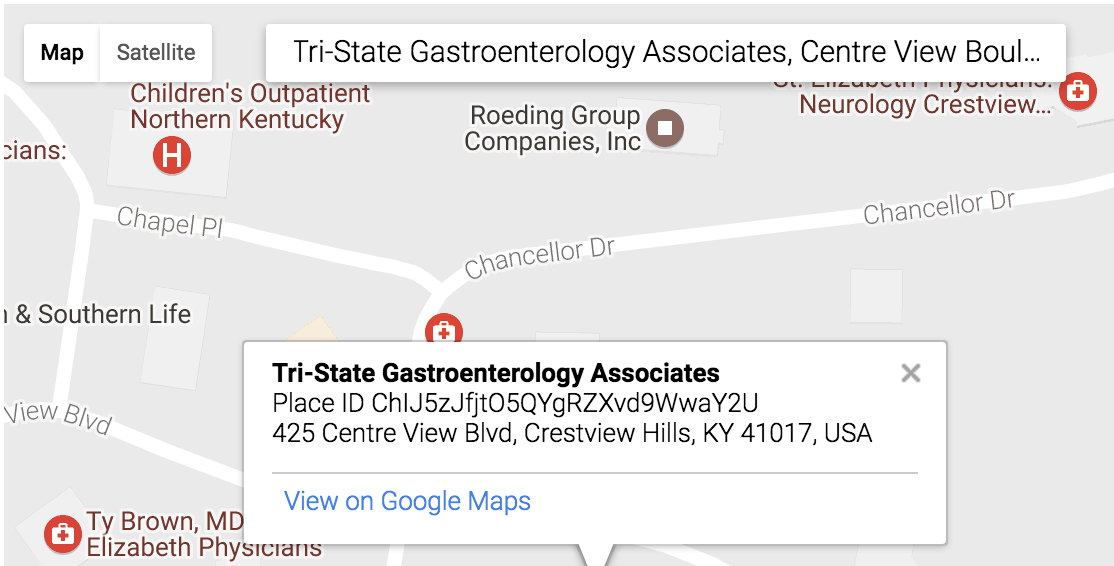


In google terms, this is a “place”. It is a location (business, person, etc) that has been added to the google maps database. When a place has been added, google users are then able to add comments and review the “place”. These are the comments/reviews that we are going after.

In order to scrape these google “places”, we need to add the “place\_id” to the Subject\_Site\_Identifiers.csv file.

How to find the google Place ID:

* 1. Do a normal google search for the doctor. Modify the search terms until you find the correct google place on the google map or the correct side-bar profile information shows up on the search results page.
  2. Use the Google Place ID finder: <https://developers.google.com/places/place-id>
  3. You will want to type the name of the place you found in the step above the exactly. You will then see your place on the map with an info box containing the Place ID.



* 1. Put ONLY the value after “Place ID” in the above image into the Subject\_site\_Identifier.csv file.

1. **Use the “Upload CSV Files” to re-scrape the updated files**
2. **Review the ProfileReport.csv**