Technical Report

**The sources of data that you will extract from.**

Clinicaltrials.gov

Bmsstudyconnect.com

**\*The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).**

**Scraping CTGOV**

Using requests and beautiful soup we pulled data using ct.gov URL by nctID

Appended site name, recruiting status, city, and zip to dictionary

Loaded records into a json

Using a for loop records were inserted row by row into MongoDB

**Scraping BMS**

Using chrome driver and selenium the html was scraped from the page source

Time delay was needed because the page load was slow

Class definition was set and then span was used to find the trial id

Collected trial id, location, & recruiting status- had to replace “\n” characters and “ ”. and appended into a location dictionary was inserted into MongoDB

**The type of final production database to load the data into (relational or non-relational).**

MongoDB

**The final tables or collections that will be used in the production database.**



