Here is the link to a spreadsheet that contains a current "SELECT \* FROM" four of the five tables in my DB. They are each represented in separate tabs at the bottom:  
  
[https://docs.google.com/spreadsheets/d/1phDQpdFoHmAIiXD8BbrLZ0F2Uxs6g6Ye\_Db3SHx4VBA/edit?usp=sharing](https://nam12.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdocs.google.com%2Fspreadsheets%2Fd%2F1phDQpdFoHmAIiXD8BbrLZ0F2Uxs6g6Ye_Db3SHx4VBA%2Fedit%3Fusp%3Dsharing&data=04%7C01%7Crbswift%40seattleschools.org%7Cc65b7db6c2f24df588a208d8b0ffbfd5%7Cd431d15860744832878351ea6f6dd227%7C0%7C0%7C637453959784486237%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=HCAYY1E6ma6YLdAS4PDopIWCHUsvutHW5jfbD1GOB40%3D&reserved=0)

The data you would be collecting would be added to the following two. There is already some data in these tables, but if it doesn't take too long, and you grab the data from these too, it might be a good way to check the success of the scraping:

**races**

* This table lists all the separate races, team places, and scores. It references the meets table, since there are multiple races at every meet.

Columns break down thusly:

|  |  |
| --- | --- |
| id | the unique id for each row |
| date | date |
| location | this references the "courses" table with a unique code |
| season | season - this is based on date. (unless we have the abbreviated spring XC season this year. for record keeping purposes, i will probably still refer to it as 2020, so as not to conflict with the fall season) |
| bg | this is an indicator of whether this was a boys or girls race |
| race\_name | this is the name listed in [athletic.net](https://nam12.safelinks.protection.outlook.com/?url=http%3A%2F%2Fathletic.net%2F&data=04%7C01%7Crbswift%40seattleschools.org%7Cc65b7db6c2f24df588a208d8b0ffbfd5%7Cd431d15860744832878351ea6f6dd227%7C0%7C0%7C637453959784496230%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=VHB2yC9fWaopekSBzDPNdD9739%2FVGImT%2Ft%2BbQ%2B6UioY%3D&reserved=0) for the specific race. this is usually something banal like "5000 meters Varsity" but isn't always, so it is good to have. |
| distance | distance in miles |
| type | varsity, jv, or mix. the first two are self explanitory, the last is races with both classifications in one race. I have considered a freshman one, but those are just a specific subset of jv races, technically, so for now, I am not using this label. |
| score | the score for the team in that race |
| place | the team's place in that race |
| teams | the number of teams competing in that race. using the numbers in the last two columns, we can later on determine a relative metric for success when compared to other races |

**performances**

* This is just what it says. The list of all performances. It references the rams table, since I use an id number for each athlete, I explain it more below. If you are mapping your scraping results to fit this, you can just put the athlete's name into that column for now.

Column break down:

|  |  |
| --- | --- |
| id | the unique id for each row |
| ram | this references the "rams" table and is the unique code for each runner. |
| race | this references "races" for the time being, just add the name for "race\_name" here. Once i import the races into the db, I replace this with the unique id given in that table. I import without that column, and the db assigns numbers sequentially. |
| race\_time | race times. it is formatted minutes:seconds:hundreths of a second (precision of 2 decimal places) |
| depth | this is the place amongst rams in the specific race. this can be obtained manually when you eliminate other teams in the selector on the site. I am not sure exactly how your code works, but if it is a simple automation of my video steps, then it should do this fine. |
| place | this is their place within all runners in the race |
| grade | grade |
| season | season |
| meet\_team\_rank | this is their place amongst all rams of the same gender at the same meet. if there is only one race, this is equal to "depth". when more than one race, we rank all rams performances at that meet with the same gender. |

Those two work in concert with:

**meets**

* This is the list of all meets. It references a courses table, which is just a list of the places we race, since some courses host multiple meets. I didn't include that, here, because it is even more straightforward than these two here.
* **the urls for all available athletic.net meets are listed here, but also the id column just uses the athletic.net code for each meet without the rest of the url**. The exception is in the case of a few meets in 2006/7 which don't have results on athletic.net. I have to enter this data manually.

**rams**

* This is the list of all athletes. Each is represented by a unique ID for each season they participated. I suppose I could have done that a little bit differently, but at the time I made this DB, I wasn't too savvy with foreign keys that combined multiple columns. The last bit of "data massage" I do is to reference a copy of this in a spreadsheet and do a simple VLOOKUP to grab the id numbers for each athlete before i import them into the DB. if there is some fancy way you can do the replacement in "performances" accurately, using this info, then let me know. if not, it really isn't a huge deal if it is the only massage I have to do.

Let me know if this helps. Aside from sharing the spreadsheet above, I have shared this folder for you to drop results: [https://drive.google.com/drive/folders/1h\_qwzfsd15bzYsoBxhMXPTKJkLyPGnJo?usp=sharing](https://nam12.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdrive.google.com%2Fdrive%2Ffolders%2F1h_qwzfsd15bzYsoBxhMXPTKJkLyPGnJo%3Fusp%3Dsharing&data=04%7C01%7Crbswift%40seattleschools.org%7Cc65b7db6c2f24df588a208d8b0ffbfd5%7Cd431d15860744832878351ea6f6dd227%7C0%7C0%7C637453959784496230%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C1000&sdata=zvEl2wvVZXEyi2Y4Iofbkq6oUUtf70PBcEilehh7jiA%3D&reserved=0)