

Status Report

So far, the majority of my project work has been focused on scraping the data. Although the project requirements state that the data does not need to be real, I want this project to be something that I can look back on and continue to use myself as a personal application.

Currently the problem stems from acquiring this data. With 100 races, each with 50+ cars, each with 3+ drivers, there is a lot of data. My current solution is by downloading all the Wikipedia html files, then parsing these files, exporting the data contained in the results field. From there I can generate the cars, drivers, teams, and results, which should be enough to start making an application.

I don't have any worries about the front end as I'll be using streamlit, a program I am familiar with from using it in CPSC 325.