Sarah Schubel

Professor Willamson

DS 2002

October 20th, 2024

While I do not think this project was incredibly high in complexity, the vague outline for what it should accomplish made it more difficult. Since I am not the very best or efficient at coding, it took me a while to write and debug what I had written to make sure it worked. For instance, I noticed that every time I wrote a new file, whether it was CSV or JSON, it would need to be deleted from my files in Colab before rerunning to code for further debugging. Though I am sure there was a much more eloquent fix for this problem, I decided that the new file would be named a random number from 0 to 100,000,000 and I no longer had another issue with it, just be sure not to run it a hundred million times. Again, I am not sure why you would anyways. Sure the code is a little janky, but it gets the job done and I'm pretty proud of it too. The biggest difficulty I had with writing this code was taking a JSON input, because the dataset I initially used from Kaggle was a little too complex for me, so I instead created a new dataset based on the CSV one I already had. Overall, this project was helpful in learning how to convert data from one data format to another especially considering that in the data science field not everything will be homogenous.