

Submit a 1 pager on your experience reflecting on the challenges, what was easier than you thought and what was harder. How would a utility like this be useful for other data projects you may encounter?

The greatest challenges I encountered in this project were recalling all the methods we learned in class to interconvert and modify data sources. I did not have very much experience programming outside of this Data Science minor's classes, so I had to repeatedly look up documentation for handling the different formats. It was also challenging to work with user input, and ensure that my script could perform well or produce informative error messages for any combination of the two user inputs. For this I tested the processor a large number of times.

What was easier than I thought about this project was finding and using open datasets. I had thought it would require more research to find two data sources to use in this project in the proper format and that I could understand without specialized domain knowledge. However, I was able to find many open datasets in CSV and JSON formats which also had very readable column names and data organization. The Charlottesville Transit dataset which I used, in fact, had an informative description on its website and multiple formats to choose from. I feel as though I am fortunate to live in a time where there is already a wealth of knowledge and open data resources readily available to me. A few years ago I do not think I would have had the same ease.

As mentioned previously, what was harder than I thought about this project was recalling specific methods and syntax for each data file format and integrating them into one script. I often ran into the problem where I applied a method meant for one format on an operation for another format and received error messages. This was also my first time creating a processing tool to be used by other people (rather than a solution to a problem just intended for me to solve), so it was a challenge to organize my work clearly and think about what documentation and instructions would be most helpful to the user.

This utility may be helpful to me in future projects I encounter by helping me to interconvert between data file formats. If a format I am using in the future is not included in this project, I anticipate it will still be a useful summary of methods and a guide for generating ideas. The operations to find the

number of records and columns in the data file were also a good base for summary statistics, and could be expanded on in the future to be an exploratory tool. For instance I could also count null values, find maximum and minimum values, and perform operations like average and standard deviation.