

Lesly Veizaga

Professor Williamson

DS 2002

Oct 20, 2024

Data Project Reflection

Over the past weeks, I have been working on this Data ingestion project where I chose my first disease csv data from Kaggle, and my second news data JSON from Kaggle. Concluding my first data science project, I have learned how to manipulate data from different sources like JSON, CSV, SQL. What I found the most challenging at first was deciding when to modify my dataset, either before or after converting the data. After pondering, and consulting with the Professor, we found it best to modify the data before converting it. For my first Dataset (CSV to JSON into SQL), handling a huge file, I decided to drop all columns but 3, where then I added 1 new column. The column I added was the “Health risk score” which I had to calculate based on the “Age” and “Diagnosis” data. I found this helpful overall because I created a new column that can potentially help in the medical field by displaying who is at most risk. Another challenge I kept running into was handling missing data values and getting lost with all the code. I found what helped with this the most was to label each step with comments to help me keep organized of what I am doing and easier for the Professor to read. A utility like this can be useful for other data projects I may encounter in the future because I learned how to manipulate, and pull data from APIs and convert them into a different file type like JSON or CSV. This is important because it helps me sharpen my coding skills and modifying.