

Consider the work you just completed in the Data Engineer Exercise. Please write no more than 2 paragraphs explaining to a client how you approached this task. Remember that clients do not need to know every process detail, but do want to understand how and why your choices contribute to our overall strategy and any benefits of the end product.

In order to create the people csv file, I first merged the three files into one file. The people schema called for 5 variables: email, code, is_unsub, created_dt, and updated_dt. A few of the files had information about several of these variables; however, I decided to use the constituent information file for created_dt and updated_dt since the schema called for insight about the constituent. The end product includes all five variables with appropriate data types for easy accessibility, reproducibility and readability.

For the acquisition facts file, I used the constituent information file since the schema called for information about the constituents. This file is simple, with two columns, 1) the calendar date of a constituent acquisition and 2) the number of constituents acquired on that date. I cleaned the values in the column that gives insight about when constituents were acquired - specifically to follow a calendar date format. Then, I determined the number of constituents that were acquired on each of the date that the data contains. The acquisition facts file includes these two variables with appropriate data types, similarly to people.csv. All of the steps that was executed to produce the end product allow for easy accessibility and reproducibility of the data.