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.

Hello [Client Name],

The analytics portion required me to make a spreadsheet with the specified fields, and make calculations with the information given to answer questions. The first thing I did was put the given data into the spreadsheet by copying, pasting/transposing. After inserting the given data, I fill the other columns with formulas to find more insightful figures to analyze. Now that I have the needed information, I move onto the questions. I use formulas and basic algebra to find and explain changes in percentage and performance specified.

The data engineering portion of the exercise was a simple ETL(Extract, transform, load). The first step was extracting. To do so, I downloaded the files with the specified links, and loaded them into our program as dataframes. Next is transformation. The first is a new dataframe that contains bits of the three I’ve just extracted. To make this data frame, I join the 3 separate tables into one based on common fields, in this case on different ids. I then drop all the unnecessary tables, and am left with a table that matches the schema listed in the exercise. Secondly, I need to create a new table that lists dates, as well as how many acquisitions were made on that date. To do this, I drop more unnecessary tables and group by the “create by” dates. I then save these data frames as .csv files locally. This is the load step of this ETL. That concludes my submission for the mission wired data engineer exercises. Please feel free to respond with any questions you may have, and let me know if there is anything I can clear up.

Best,

Ronald Tun