**Client Communications**

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.

Dear Client,

Using the data I have turned the 3 data frames ('cons.csv', 'cons\_email.csv', and 'cons\_email\_chapter\_subscription.csv') into one large data frame inner combining of the data on column names: cons\_id and cons\_email\_id because they are primary keys meaning they are unique for each entry. Next, I only included email users from 'cons\_email.csv', where the variable ‘is\_primary’ is equal to true because we want to ensure we are only including primary email users. Next, I only included email users with subscriptions from ‘cons\_email\_chapter\_subscription.csv', where ‘chapter\_id’ is equal to 1 because we only care about those subscription statuses. Using the one large data frame, I have the data to complete the ‘people’ data frame that includes: the ‘email’ column from ‘cons\_email.csv', the code column from ‘cons.csv', the ‘is\_unsub’ column from ‘cons\_email\_chapter\_subscription.csv', and lastly the created\_dt and updated\_dt columns from ‘cons.csv'. Although I was unsure on the ‘code’ column for the ‘people’ data frame, I used the ‘source’ column from ‘cons.csv’ because the description says source code.

Using the ‘people’ data frame I was able to create an ‘acquisition\_facts.csv’ in order to know when people in the dataset were acquired. First I had to create the acquisition\_date column by using the create\_dt variable of type, datetime and changing it to type, date. Lastly, using the acquisition\_date column, I am able to use a ‘groupby’ function in order to group by the dates and get the acquisition count of each date appearing.

Best Regards,

Data Engineer