

Hello [client name],

Thank you for sending the 3 datasets for Sprocket Central. I have taken a look at all 3 and highlighted some data quality issues for your data analytics/science team to resolve; as well as recommendations to prevent these issues from arising again. Please let us know if you or anyone else at the company has further questions regarding the data cleaning portion of the analysis.

| Dataset      | Number of Records | Unique Customer IDs | Date Data Received |
|--------------|-------------------|---------------------|--------------------|
| Transactions | 20000             | 3494                | 2017               |
| Demographic  | 4000              | 4000                | unknown            |
| Address      | 3999              | 3999                | unknown            |

#### Data Types

Within all 3 datasets, the data types listed are objects/strings. It would be best if you include other data types. For example datetime, float, etc.

#### Data Completeness

It seems that there are missing values in each dataset. Within the transactions dataframe, online order, brand, product line, and 4 other columns are missing data - some more than others. In the demographic dataframe, last name, DOB, job title, and 3 more columns are missing values. And in the address dataframe, at least one customer id is missing along with the values coinciding with it.

#### Data Currency

I was able to identify that the year of transaction evaluation was 2017, however, I cannot identify what year is being evaluated from the other 2 datasets.

#### Data Uniqueness

After comparing the uniqueness of customer ids within each table there are varying differences. As outlined in the table above each table has a different number of customer ids present.

*Recommendations: Incorporate the date of data accumulation for each table so that we can ensure that all tables are evaluated accurately against one another. And fill in missing information, especially regarding customer ids. It is imperative that we effectively evaluate the tables by the same criteria in order to ensure that the model and analysis is as accurate as possible.*

Thank you for your cooperation and trust in KPMG to accurately present your data for your company's growth.

Best,  
Chanah Dunbar