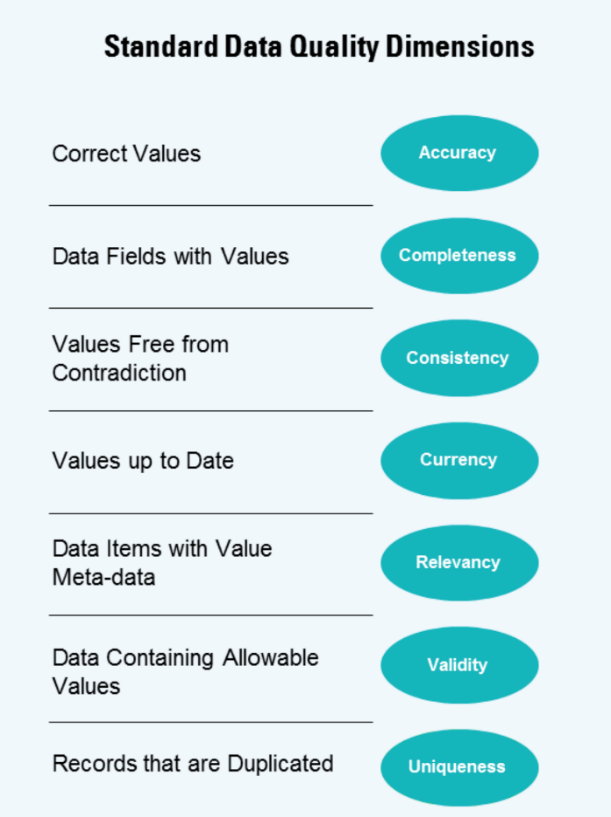
Greetings from KPMG,

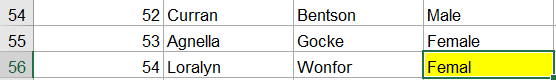
As a starting point, below are the seven main dimensions our company had used to assess Sprocket Central.



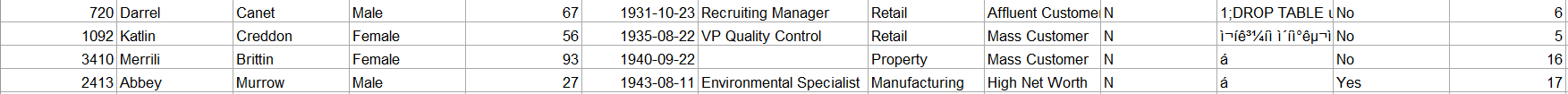
The data provided include the transactions and customers data. There are few things I realized while I was analysing the data set

* Customers’ ID links CustomerDemographic and CustomerAddress
* Customers in NewCustomerList have not yet been assigned customer\_id
* NewCustomerList is not yet related to CustomerDemographic and CustomerAddress

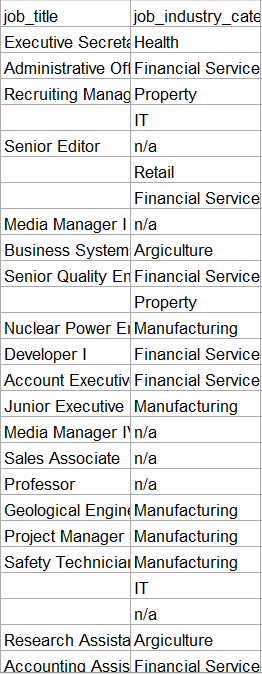
Moving forward to the analyses for the quality dimensions. The gender column in data set of CustomerDemographic with customer\_id of 54 has a value of “femal” instead of “female” or “male” or “u” which represents “unknown.” This indicates that the respective row prevents the data set from being accurate.



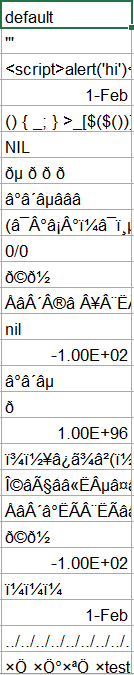
Once CustomerDemographic is filtered with ascending norm, there are few data fields which are blank or incomplete making it not meeting the completeness factor. Below are the data fields example from CustomerDemographic. In order to go into the next phase of analysis, these fields ought to be filled with relevant value.



The data relevancy is another dimension which could not be covered by the data set. In data set CustomerDemographic, the attribute job\_title should define job\_industry\_category and vice versa. However, there are some blank fields and fields with “n/a” value. Hence, it could not be considered as a relevant data set.



The default column on CustomerDemographic does not considered as valid data. It is shown on the second row where it is filed with script while others are filled with numbers and characters.



Data consistency, on the other hand, could not be measured since all four data sets are completely independent one another. Currency and relevancy of the dimension are both not as important as the others are skipped since majority of the dimensions were failed to be met. Meanwhile, all the rows from four data sets are all meeting the dimension of uniqueness.

As a result, the data set is not yet ready to be progressed further into the second step since more than half of the dimensions are not satisfied. It is justified by the previous analysis where four(Accuracy, Completeness, Relevancy and Validity) out of seven dimensions was not fulfilled.

It would be better if all the fields are to be placed with its desired inputs without leaving any blank and with correct grammar to prevent any misunderstanding.

Regards,

KPMG Team