**Data quality issues concerning the data-set of Sprocket Central Pty Ltd.**

To whom it may concern

My name is Oluwatosin Alebiosu and I work with the KPMG Data Analytics and Modelling. I am writing you this email as per the data quality issues discovered in the three data-sets sent to our company for analysis. Thank you for providing us with the three data-sets.

Below are the **discoveries** from the data-sets received:

* Various columns, such as the brand of a purchase, or job title, have empty values in certain records.
* Additional customer ids in the ‘Transactions table’ and ‘Customer Address table’ but not in ‘Customer Master (Customer Demographic)’
* Inconsistent values for the same attribute (e.g., Victoria being represented as “V”, “Vic” and “Victoria”)
* A certain customer with ID 34 has an incorrect date of birth value (1843-12-21). This data is not accurate as the approximate age would be 180 as of today.
* Values in the ‘default’ column of the Customer Demographic table is not allowable and not understandable.
* Numerous incomplete/blank data values, which are relevant to the task.

I am also listing out recommendations you can go about **mitigating** these issues respectively:

* If only a small number of rows are empty, filter out the record entirely from the training set for prediction. Else, if it is a core field, impute based on distribution in the training dataset.
* : Please ensure that all tables are from the same period. Only customers in the Customer Master list will be used as a training set for our model.
* Use regular expression to replaced extended values into abbreviations to ensure consistency across addresses.
* Try confirming the accurate date of birth from the data source.
* Kindly shed more light into what the items in the ‘default’ column entails.
* Kindly check from other source and provide the complete data if available.

Also, I want to confirm if the following acronyms “NSW, QLD, VIC” stands for “New South Wales, Queensland, Victoria” respectively. If not, kindly state the full meaning of those acronyms.

Moving forward, the team will continue with the data cleaning, standardisation and transformation process for the purpose of model analysis. I am looking forward to your response. Thank you.

Kind Regards

Oluwatosin Alebiosu

Junior Data Analyst.