KPMG

Dear Client,

Thank you for providing us with three datasets to identify ways in which we can improve the quality of Sprocket Central Pty Ltd's data. The three tables analysed are Customer Demographic, Customer Address and Transactions. Inconsistencies and errors have been pointed out as well as solutions and recommendations to avoid any reoccurrences of such data.

- **There are 4000 customer_ids in the Customer Demographic table, however, additional customer_ids can be found in the Customer Address table and Transaction Table**
  Solution: Only use customers that are found in the Customer Demographic table to model the data. Otherwise, the outliers will create errors and skewness in the results.

- **Columns have empty fields such as DOB, job_title and last_name in the Customer Demographic Table (empty fields can be found in all tables)**
  Solution: Enforce a must fill policy for the user entering data on important fields. For the current existing data set we can remove any rows with empty fields if the ratio of non-empty fields to empty fields is fairly low.

- **Values of the same meaning have different representations**
  Solution: Inconsistent representations such as male being referred to as 'M' or 'Male'. To create meaningful variables, enforce a radio button with specific data values. The gender column also contains values of 'U' which can be replaced by either 'M' or 'F' whilst still maintaining the same ratio of male to female.

- **Values have inconsistent formats**
  Solution: Columns such as list_price in the Transaction table does not contain a specified currency before each value. Values need to be up to date to avoid creating any confusion.

The team will continue to analyse the datasets for solutions to any outliers. Please contact me regarding any questions regarding the email.

Kind regards,
Ahmad Sultan.