Dear [to the person in contact],

Firstly, we would like to thank you for choosing KPMG as your trustworthy partner and entrusting us with providing an access to your database. We are pleased to work with you on this journey of yours to make a data driven decisions and to improve your business from all dimensions.

Secondly, our team analyzed all the three datasets provided with us by the **Sprocket Central Pty Ltd** and we found out that there are some inconsistencies lies within the dataset which could possibly reduce the quality of the data i.e. data quality and the accuracy of the result, if it would have occurred in a large scale. But, since the ratio between the total number of rows and the number of rows with inconsistencies is low, we just filtered and normalized them based on our industry standard methods. We have attached some notable inconsistencies and the methods which we have used to mitigate them, along with some recommendations to prevent those same occurrences in the future.

Presence of empty rows in a column (Incompleteness):

- **Recommendation:** Implement an option to enter or select "***N/A***" or "***Not Available***" or "***Others***" based on an application, whenever or wherever data is not available.
- **Mitigation:** This issue can be dealt by filtering out all rows from the table which has an empty values (Null values) in it, if the number of rows containing null values are low. But at some cases we can also resort to filling out all those null values with the mean value of a respective column if possible.
    1. Rows containing null values are filtered out, since the number of rows with null value is low.

Different words for same purpose (Inconsistencies): This is one of the most common data quality issues where different words are used to refer same meaning. Example: "New York" and "NY" are used for same purpose.

- **Recommendation:** Implement scroll down menu with pre-fixed options in it or implement check boxes with options - at places where data's are being typed manually.
- **Mitigation:** Replacing all those words with a single unique word.
    1. ***Customer Demographic dataset***: Replaced "Female, Femal, and F" to "F".
    2. ***Customer Address dataset***: "New South Wales, NSW" to "NSW" and "Victoria, VIC" To "VIC".

Presence of entries from different time period (Inconsistency): ***Customer Demographic*** dataset has a lot of outliers present in it, which will ultimately affect the quality of the data driven outcome.

- **Recommendation:** Make sure that all the entries in the dataset are from the same time period, location and same group.
- **Mitigation:** Removed all those rows which had outlier values present in it.

These are the data quality issues which our team has faced while analyzing the quality of the dataset, and we have also attached some of the recommendations to prevent those reoccurrence of quality issues in future. At present our team is working on further processes and we will update you regarding the progress updates periodically.

Thank you for taking your quality time.

With kind regards,
Dineshkumar.