To:
Sprocket Central Pty Ltd

Subject: Data Quality Assessment

Dear Sir,

Thank you for allowing [**my company**] to handle your data analysis project. We received three datasets from you, tabulated below for your perusal.

Dataset	Key column header	No. of rows (before cleaning)	No. of rows (after cleaning)
CustomerDemographic	customer_id	4000	3911
CustomerAddress	customer_id	3999	3999
Transactions	customer_id	20000	19804

We noticed some data quality issue with the data. I will highlight these issues, provide recommendation on how the dataset will be cleaned, and ways to mitigate the issues.

1. **Incomplete records**: Some of the data fields have no data inputted and this can lead to error in analysis.

Mitigation: Ensure to input all the records carefully

Recommendation: Records with blanks that can affect analysis will be erased for purpose of analysis.

2. **Inconsistency**: Some fields have two different strings represent the same thing. E.g.,"vic" and "victoria" in the "state" field.

Mitigation: Data should be represented with the same value

Recommendation: Abbreviations will be used to represent all data in the State field.

3. **Different data types**: A field containing numeric, string, date datatype cannot be used for analysis.

Mitigation: One datatype should be used throughout any field.

Recommendation: Field will be deleted or changed to one datatype where possible.

4. **Additional customer ids**: The customer_id is the primary key for all three datasets, but some datasets have customer_id that others don't have. This won't allow the table to match and can lead to error in analysis.

Mitigation: Provide the records for additional customer_id.

Recommendation: The extra dataset that have no records in all three datasets will either be deleted or added for analysis. These are the data quality issues with the three datasets you provided us with.

Thank you.