The Managing Director,

Sprocket Central Pty Ltd.

Dear Sir/MA,

Thank you for providing us with four datasets from your company, we have reviewed the dataset and during the data quality analysis, these were our finding:

There are four tables provided in the datasets, namely;

* Transactions Table
* New Customer List
* Customer Demographic
* Customer Address

The Transaction Table consist of 20,000 rows and 13 columns, the New Customer List has 1000 rows and 23 columns, the Customer Demographic has 4000 rows and 13 columns, and the Customer Address has 3999 rows and 6 columns.

Here are some of the data quality issues we observed and the recommendations we provided:

1. The Customer\_ID provided for Customer Demographic and Customer Address is totally different from the one provided in the Transaction table and the New Customer List does not have any customer\_id.
2. The data type for Product First Sold Date column in the Transaction table is appearing as float instead of data time data type. We converted it to date time data type so that it can appear as date.
3. There are 555 rows having missing values in the Transaction Table ( 360 missing value from the Online Order, 197 missing value each from Brand, Product Class, Product Size, Product Line, Standard Cost and Product First Sold Date). In the New Customer List Table, there are 285 rows of missing values ( 29 missing values from Last Name column, 17 from Date of Birth, 106 from Job title and 165 from Job Industry Category). Also, in the Customer Demographic Table, there 1220 rows of missing values ( 125 missing values from Last name, 87 from Date of Birth, 506 from Job title and 656 from Job Industry Category).
4. There are some columns in the New Customer List Table ( Unnamed 16-20) and a column in the Customer Demographic Table ( Default) that we found irrelevant to our analysis. We dropped them from both Tables in the dataset.
5. The gender column in the New Customer List Table has SEX- Male, Female and U. while that of Customer Demographic Table has SEX- Male, Female, F, M, Femal and U. We replaced the SEX with Male, Female and Unspecified ( where F was replaced with Female, M was replaced with Male, Femal was replaced with Female and U was replaced with Unspecified).

Regards,

Stella ( for KPMG)