**E-Mail**

Subject: Data Quality Issues

From: Akinlabi Otebolaku (KPMG Data Analyst)

To: Sprocket Central Pty Ltd

Good day sir/ma,

As regards the data set given to us, the following observations were made:

**The Transactions data**

• The values in the columns: “product\_size” and “product\_class” are in lower case and so had to be replaced with upper cases starting as this might present a problem when trying to find a character.

• The “list\_price” column was not in currency format and had to be changed to currency format as it denotes the currency in which it is being sold.

• The “product\_first\_sold\_date” column was not in date format and had to be changed to date format as it is a date column

• Blank cells were found in the “online\_order, brand, product\_line, product\_class, product\_size, standard\_cost and product\_first\_sold\_date” columns which will greatly affect our analysis results.

**The New Customer List**

• There was an empty row at the top of the data set which had to be removed so as not to distort the dataset.

• There were blank values in the “last\_name, DOB, job\_title” columns which will defect our quantitative and qualitative analysis.

• The “ past\_3\_years\_bike\_related\_purchases, post\_code and property\_valuation” columns were not in number formats but in text format which would affect the way the data is viewed.

• Five (5) terrestrial columns without column headers were found in the data set with strange values.

• The DOB column was not in date format.

• There is a strange character “U” in the gender column other than “Male or Female”.

• There is a strange character named “n/a” in the “job industry” category.

**Customer Demographic**

• There were blank cells in all of the columns present in the dataset except the customer id column.

• In the “gender” column, there were errors like “U, M, F”, e.t.c instead of “Male or Female”.

• There were errors such as “n/a” in the “job\_industry\_category” column.

• There is a strange column named “default” containing characters such as “ï½ï½¨(Â´âï½â©”.

**Customer Address**

• Some characters were in abbreviation while some were in full form under the “state” column.

To mitigate these data quality issues**,** a proper check should be made on the data set before analysisso as to ensure effective analysis of data for the making of better informed decisions and for the overall growth of the company.

With regards to these data quality issues, we hope you improve on the data gathering process so as to incite a more accurate analysis. Thank you.