**Email to Sprocket Central Pty Ltd**

**Subject Line:** Information on Data Quality Assessment

Dear Sprocket Central Pty Ltd,

I'm pleased to share the following details regarding the data quality assessment on the four dataset you have provided for us.

Below I have listed the datasets along with the issues I found and their respective solutions.

**Transactions data**

Problems:

1. Columns order\_status, brand, product\_line, product\_class, and product\_size are of type object

2. product\_first\_sold\_date is not in datetime format

3. Missing values in online\_order, brand, product\_line, product\_class, product\_size, standard\_cost, and product\_first\_sold\_date

Solutions:

1. Convert columns to type category or convert to dummy variables wherever possible.

2. Change date to datetime format

3. Either filter out missing values completely or impute them

**Customer Demographic data**

Problems:

1. Columns with category values labeled as type `object`

- gender

- job\_industry\_category

- wealth\_segment

- owns\_car

2. Unnecessary default column

3. Two deceased customers in data

4. Missing values in DOB, job\_title, job\_industry\_category, and tenure

- There are missing values in last\_name as well however I'm assuming those are customers without a last name to begin with.

Solutions:

1. Convert columns to type category or convert to dummy variables

2. Drop column

3. Filter out deceased customers

4. Filter out missing values or impute them

**Customer Address data**

Problems:

1. state column is of type object

2. Similar values in state column

3. Unnecessary country column

Solutions:

1. Convert to type category

2. Match values into one

3. Drop column

If you have any questions, please don't hesitate to ask.

Best regards,

Saul Chirinos

Junior Consultant

KPMG Data Analytics Team