**Title: Data Assessment**

To: Sprocket central pty Ltd

Dear Addressee,

This is Sohaib Muhammad Khan from KMPG Analytics,

I have received your data and as a Data Analyst after looking and analysing the dataset there are some data quality issues focused on accuracy, completeness, uniqueness. I would like to brief these issues towards you and suggest some strategies to mitigate them.

Transaction Sheet:

* 360 rows in the "online order" column are blank. You can consider either removing these rows if they do not provide essential information or inputting missing values.
* 197 rows in the "brand," "product\_line," "product\_class," and "product\_size" columns are blank. You can also consider either removing these rows if they do not provide essential information or inputting missing values.
* There is a customer\_id of 5034 which exceeds the limit of ids because of the demographic sheet which was (1-4000).
* There are 197 blank rows in the "standard\_cost" column. Similar to the previous point, you can either remove these rows or input missing values.
* The "product\_first\_sold\_date" column is not in proper date format and contains 197 blank rows. To resolve this, you can use Excel's Text to Columns feature once again, selecting the "Date" option.
* It is important to check for the data type of each column for more clarity and visibility.
* Finally, to ensure there are no duplicates in the data. You can perform this feature by utilizing Excel's Remove Duplicates feature.

Customer Demographic Sheet:

* There are 125 blank rows in the "last\_name" column. Similar to previous cases, you can either remove these rows or input missing values using suitable techniques.
* The "gender" column contains entries such as "F," "M," "Femal," and ''U." You can standardize the gender entries by replacing them with consistent values, such as "Female," "Male," and "Unknown".
* Customer\_id with id’s 753 and 3790 are having their deceased status as Yes. Since we’re targeting only living customers they will be removed.
* There are 87 blank rows in the "DOB" column, with an odd year "1843" as a date of birth. You may consider removing these rows or imputing missing values using appropriate techniques.
* There are 506 blank rows in the "job\_title" column. Similar to previous cases, you can either remove these rows or input missing values using suitable techniques.
* The job\_industry" column has 656 rows with 'n/a' entries. You can either remove these rows or replace 'n/a' with a suitable category that represents missing data.
* The "default" column is irrelevant and can be removed from the dataset to improve data cleanliness.
* There are 87 blank rows in the "tenure" column. Similar to previous cases, you can either remove these rows or impute missing values using suitable techniques.
* It is important to check for the data type of each column for more clarity and visibility.
* Also remove duplicated if there is a need by utilizing Excel’s Remove Duplicate feature.

CustomerAddress Sheet:

* In the "state" column, there are entries, such as "VIC" or "Victoria," "NSW," or "New South Wales." To address this issue, you can use Excel's Find and Replace feature to replace the state names.

I hope you have got the insights of the data quality if you have any queries please reach out by replying to this mail.

Thanks, and regards.