Hello Sir,

This email is in reply to your previous mail about ensuring the data quality of the 3 datasets given by Sprocket Central Pty. Ltd.

* **Customer Demographic**
* **Customer Addresses**
* **Transaction data in the past three months**

We have reviewed the data quality of the datasets and made sure that it is ready for analysis in phase 2. We have also referred to the data quality framework that was provided by you as a sample. Following are the changes that need to be made in the three datasets.

* **Transaction data in the past three months:**

1. In the online\_order column, there are certain blank and false entries that need to be removed to ensure accuracy and completeness of the data. False entries mean that the order is given offline which does not meet our requirements and neither can we keep blank entries which would mean that neither offline nor online orders are considered. This ensures consistency of the data.

2. In the order\_status column, the cancelled orders need to be removed as they don’t affect the revenue in any way.

3. From the brand column, we need to remove the blank brands as they are not needed and to ensure completeness of the data.

4. An additional profit column should also be calculated by using the formula profit=list\_price-standard\_cost.

5. Also, we need to change the date format into short date to make it a standardised and a readable date format.

* **Customer demographics:**

1. In the gender column, there are certain incompetencies. The gender female is written as duplicate parameters and in order to make it standardised, we need to replace all the F values with Female and similarly for male replace all the M values with Male. This ensures uniqueness of the data.

2. An additional age column after DOB column should also be included, which calculates the age of a person by using NOW()-DOB/365. The filters for ages 100+ need to be removed, because that cant be an age of a person practically. This ensures validity of the data.

3. In the job\_title column,the blank fields need to be removed to ensure accuracy and completeness if the data.

4. From the deceased\_indicator column, the Y part can be removed, because if the customer is deceased then he wont be able to do online shopping. This ensures validity of the data.

5. The script column can be deleted as we cannot comprehend it.

* Customer address:

1. In the state column, New south wales and Victoria have repeated parameters, so they can be replaced by only one parameter NSW and VIC respectively. This ensures uniqueness of the parameters as well as data.

We hope that all your requirements have been satisfied and the quality of data has been ensured. Kindly revert back to the same.

Thanks and regards,

Tony Smith

Partner in KPMG Lighthouse & Innovation Team

A picture containing screenshot

Description automatically generated

A picture containing screenshot

Description automatically generated

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA picture containing screenshot

Description automatically generatedA picture containing screenshot

Description automatically generatedA picture containing screenshot

Description automatically generated