Dear Manager,

Thank you for the Sprocket Central Pty Ltd datasets. Below section addresses the quality issues in the given datasets.

# **CustomerDemographic dataset:**

Total number of records are 4000 in the CustomerDemographic dataset.

#### **Accuracy:**

***Mitigation*:** 33rd record in the DOB column has age of 177 and it is removed.

***Recommendation:***Create the new column for age and calculate the age from the entered DOB. Also, create a threshold value for DOB to avoid the irrelevant DOB value.

#### **Completeness:**

**Column name Null value**

last\_name 125

DOB 87

job\_title 506

job\_industry\_category 656

tenure 87

***Mitigation:*** From above mentioned incomplete records, the DOB records with the null values are removed.

***Recommendation*:** even though last\_name, job\_title, job\_industry\_category, and tenure columns got null value, it won’t affects our model. Hence removing DOB is enough.

#### **Consistency:**

Gender: There are inconsistency in the gender entry such as F = 1; Femal = 1; Female = 2037; M = 1; Male = 1872; U = 88. 87

***Mitigation:*** changed the F and Femal to Female and M to Male.

***Recommendation*:** create a dropdown to choose the gender so the inconsistency in the records will be avoided.

#### **Currency:**

deceased\_indicator: 2 customers are deceased

***Mitigation:*** deceased customer are filtered out.

#### **Relevancy:**

Default attribute: records has no meaning

***Mitigation:*** filtered out the default column.

**Validity:** No issues found

# **Transaction dataset**

**Accuracy:** No issues found

**Completeness:**

|  |  |
| --- | --- |
| **Column name** | **Null value** |
| online\_order | 360 |
| Brand | 197 |
| product\_line | 197 |
| product\_class | 197 |
| product\_size | 197 |
| standard\_cost | 197 |

The above table shows the missing values in the transaction dataset.

***Mitigation:*** change the columns list\_price and standard\_cost name to list\_price\_in\_USD and standard\_cost\_USD

***Recommendation:***In the given dataset all the customers are from Australia (currency is Australian dollars) and its fine to mention only USD at the moment. But, creating separate column or drop down for currency type might help in future for understanding the customer demographic better.

#### **Consistency:**

No issues found

#### **Relevancy:**

***Mitigation*:** There are 179 orders are cancelled and are removed at the moment.

***Recommendation*:** Cancelled order might help in understanding the customer’s interest and creating the reason column for cancelled order give more insight.

#### **Validity:**

Product\_first\_ sold\_ date column has different date format

***Mitigation:*** changed the date format of Product\_first\_ sold\_ date column.

***Recommendation*:** make sure the data format is similar to transaction\_date format

# **CustomerAddress**

Accuracy: No issue found

Completeness: No issue found

#### **Consistency:**

NSW 2054

New South Wales 86

QLD 838

VIC 939

Victoria 82

***Mitigation:*** there is inconsistency in the state name and all are replaced to ensure the consistency ***Recommendation*:** Create a drop down for state

**Relevancy:** No issue found

**Validity:** No issue found

Please let me know if additional information is needed.