Dear John Doe,

Thank you for providing us with the datasets from Sprocket Central Pty Ltd. Please let us know if you have any queries surrounding the key quality issues that we discovered within the datasets, presented in the table below.

**Customer Demographics**

|  |  |
| --- | --- |
| ***Field name*** | ***Errors*** |
| DOB | 1 record is of year 1843  87 records as blank |
| Last name | 125 records as blank |
| Job\_title | 506 records as blank |
| Job\_industry | 656 records mentioned ‘N/A’ |
| Tenure | 87 records as blank |
| Gender | 88 records are termed as ‘U’  Inconsistency in values |
| Default | 3317 record values are ‘special characters’ including null and blanks |

**Transactions**

|  |  |
| --- | --- |
| ***Field name*** | ***Errors*** |
| Online\_order | 94 records Blanks |
| brand | 48 records Blanks |
| product\_line | 48 records Blanks |
| product\_class | 48 records Blanks |
| product\_size | 48 records Blanks |
| standard\_cost | 48 records Blanks |
| product\_first\_sold\_date | 48 records Blanks |

**Customer Address**

|  |  |
| --- | --- |
| ***Field name*** | ***Errors*** |
| state | Inconsistency in name of states |

Below are description of data quality issues discovered and methods of mitigation used.

**Mitigations:**

* Filter out outlier in DOB, blanks for job\_title, filter all inconsistency in gender in “Customer Demographic” dataset.
* Filter all customer\_ids from 1 to 3500 in “Customer Demographic” dataset.
* Filter all inconsistency in gender in “Customer Address”
* Filter out blanks for online\_order and brand\_column in “Transactions” dataset.
* Filter out customers marked “Y” in deceased\_indicator in “Customer Demographic” as these are not the current customers.
* Deleted metadata in default\_column.
* Filter out “Cancelled” in order\_status.
* Formatted product\_sale\_date to short date format, list\_price to currency in “Transaction” dataset.

The mitigation strategies suggested are simple and effective ways of improving data quality for future analysis. These will eventually help us to improve the analysis output for your company.

Please let us know if you have questions regarding mitigation or any data quality issues identified.

Regards,

Kartikeya Gupta

KPMG (Data Analytics Team)