Hello,

This is Jyoti Sarpal, from KPMG Data Analytics team. We have reviewed the data sets which were provided by your company and during the data quality analysis, and found some errors in the data sets as mentioned below.

The data quality analysis is the core phase and due to errors in the data set, we suggest the following mitigates in order to improve the data quality, which will eventually help in getting better results.

The following are the details of error encountered in the data set.

**Data Quality Framework Table**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Datasets** | **Accuracy** | **Completeness** | **Consistency** | **Currency** | **Relevancy** | **Validity** |
| **Customer**  **Demographic** | Inaccurate D.O.B  Age  Missing | Job title: blanks | Gender: inconsistency | Filter out deceased  customers | Delete: default  column |  |
| **Customer**  **Address** |  |  | States: inconsistency |  |  |  |
| **Transactions** | Profit: missing | Online order: blanks  Brand: blanks |  |  | Cancelled order status:  filter out | List price: format  Product first sold date: format |

The following are the details of errors:

**Accuracy Issues**

* DOB was inaccurate for the “Customer Demographic” dataset.
* Profit column and age column are missing in the “Transactions” dataset.
* **Mitigate:** Filter out DOB.
* **Recommendations:** Add age column in the dataset. Add profit or profit margin column in the “Transactions” dataset for better insight of sales data.

**Completeness**

* Blanks in job title for the “Customer Demographic” dataset.
* Blanks in online order and brand column in the “Transactions” dataset.
* **Mitigate:** Filter out job title; brand; and online order columns.
* **Recommendations:**  Blank values in columns affect data analysis in a negative way, so avoid null values.

**Consistency**

* Inconsistency in gender column for “Customer Demographic” dataset.
* Inconsistency in states column for “Customer Address” dataset
* **Mitigate:** Filter out ‘M’ as ‘Male’ and ‘F’ as ‘Female’ in gender column. Filter out ‘New South Wales’ as ‘NSW’ and ‘Victoria’ as ‘VIC’ in the states column.

**Currency**

* Deceased indicator column having value ‘Y’ in “Customer Demographic” dataset is not a current customer.
* **Mitigate:** Filter out deceased indicator column having value ‘Y’ in dataset.
* **Recommendations:** Up to date data of deceased customer will increase efficiency of data.

**Relevancy**

* Lack of relevance in default column for “Customer Demographic” dataset and order status column in “Transactions” dataset.
* **Mitigate:** Delete meta-data in default column in “Customer Demographic”. Remove the cancelled value from order status column in “Transactions”.

**Validity**

* Format of list price; and product first sold date column in “Transactions” dataset.
* **Mitigate:** Change format of list price column to currency and product first sold date to short date.
* **Recommendations:** Assign suitable format for any column to make data more readable and understandable.

Please contact us for any queries and issues regarding the data quality framework table.

Best regards,

Jyoti Sarpal