Subject: Data Quality (DQ) Assessment Result

Dear Manager,

Thank you for providing us with the datasets from Sprocket Central Pty Ltd. It consists of three raw datasets named Transactions, Customer Demographics, and Customer Addresses. We have reviewed the data to ensure that it’s ready for analysis in phase two. We have found some data quality issues in the datasets provided to us.

Below is a summary table to highlight key data quality issues observed with DQ dimensions.

|  |  |  |
| --- | --- | --- |
| **Dataset No** | **Dataset Name** | **Data Quality Issue** |
| 1 | Transaction | Completeness- Missing Values found in data. |
| Validity- “product\_first\_sold\_date” is not a date. |
| Accuracy- Profit data could be included for analysis. |
|  |  |  |
| 2 | Customer Demography | Completeness- Missing Values found in data. |
| Consistency- “gender” column has inconsistent values |
| Validity- “default” column has invalid data. |
| Timeliness- “DOB” has one outlier for |
|  |  |  |
| 3 | Customer Address | Consistency- “state” column has inconsistent data. |

Recommendations:

We would like to recommend you some ideas to mitigate data issues in the future,

* Make sure while entering DOB data as well as missing values can be filled with a central tendency “mode “.
* Make sure we are entering the correct data.
* If you are using Excel, use a filter option to check and do replace data for consistency in data. Or if you are using Pandas, check the ‘value\_counts()’ then do replace the data with the appropriate one.
* All dataset can be merged for data analysis as it is relevant to each other on “customer\_id”
* Null values can be dropped or filled by the appropriate value.
* Inconsistency in data could avoid giving correct values as “Female”, “F”, “Femal”, and “Fem” can be replaced with “Female” which is more relevant and readable.
* Null Values can be dropped but it’s advised to use the fill method to fill null values as data means value.

If you have any feedback or inquiries on the aforementioned, please do not hesitate to ask.

Thanking you.

Regards,

Nikhil Neve

Data Analytics Consultant, KPMG\*

\*Note:

This draft email is generated for the purpose of the KPMG Data Analytics Virtual Internship.