Dear KPMG Data Analytics Consulting Team,

I appreciate the opportunity given to us by the Skyrocket Central Pty Ltd by lending us these datasets for assessment. This is my statistical assessment regarding these datasets.

|  |  |  |
| --- | --- | --- |
| **Table Name** | **No of Rows** | **Date Acquired** |
| Transactions | 20000 | 15/04/2023 |
| Customer Demographic | 4000 | 15/04/2023 |
| Customer Address | 3999 | 15/04/2023 |

Issues with the dataset's quality are listed below, along with solutions.

**Transactions**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Problem** | **Solution** | **Data Quality Assessment Parameter** |
| online\_order | 360 blank spaces | Should be removed/ filled with the accurate data | Completeness |
| brand | 197 blank spaces | Blank spaces should be removed | Completeness |
| product\_line | 197 blank spaces | Blank spaces should be removed | Completeness |
| product\_class | 197 blank spaces | Blank spaces should be removed | Completeness |
| product\_size | 197 blank spaces | Blank spaces should be removed | Completeness |
| standard\_cost | 197 blank spaces | Blank spaces should be removed | Completeness |
| product\_first\_sold\_date | 197 blank spaces | Blank spaces should be removed | Completeness |

**Customer Demographic**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Problem** | **Solution** | **Data Quality Assessment Parameter** |
| Date of Birth | 87 blank spaces | Should be removed/ filled with the accurate data | Completeness |
| Job Title | 506 blank spaces | Should be removed/ filled with the accurate data | Completeness |
| Default | Bad Data | Should be removed/ filled with the accurate data | Completeness/Accuracy |
| Tenure | 87 blank spaces | Should be removed/ filled with the accurate data | Completeness/Relevancy |
| Deceased Indicator | Data items with meta data value | Changing the “N” and “Y” to Yes and No | Relevancy/Validity |
| DOB | Inaccurate Data | Correcting the DOB (1843-12-21) | Accuracy |
| Gender | Spelling Errors | Changing “U” to Unidentified, changing “F” and “Femal” to Female and changing “M” to Male | Consistency/Accuracy |

**Customer Address**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Problem** | **Solution** | **Data Quality Assessment Parameter** |
| State | Inconsistency of Data | Changing “NSW” to New South Wales, changing “VIC” to “Victoria” to Female and changing “QLD” to Queensland | Consistency/Accuracy |

**N.B;** To reduce data errors, data validation strategies including using complete numbers, text, and dates should be used in the subsequent data collection process.

**Warm Regards,**

**ORUNTA EMMANUEL UCHENNA**