**Email**

Dear manager,

Please find the data quality issues and strategies to mitigate these issues. We need

clarification from the client on some data points.

Best Regards,

Samvel Kocharyan

**\_\_\_\_\_\_**

**Data Quality Framework**

**“CustomerDemographics” Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field** | **Issue** | **Dimension** | **Comment** | **Fix Strategy** |
| last\_name | Missed Values | Completeness | 125 empty cells | Prepare list of future possible issues when last-name is unknown |
| gender | Unknown Categories | Accuracy | 88 entries with “U” labels | Ask customer about “U” category label mean. |
| gender | Misspelled Category Names | Accuracy, Consistency | “Femal”, “M” | Fix misspelling |
| past\_3\_years | Probability of Outliers | Validity | Entries with “99”, “98”, etc values | Check that orders are real |
| DOB | Missed Values | Completeness | 87 empty cells | Prepare list of future possible issues when DOB is unknown |
| job\_title | Missed Values | Completeness | 506 empty cells | Prepare list of future possible issues when job\_title is unknown |
| job\_title | Probability of redundancy | Consistency | “\_I”, “\_II”, “\_III” There are a bunch of similar job titles | Prepare a list of joins. Approve labels join. |
| job\_industry | Missed Values | Completeness | “n/a” | Prepare list of future possible issues when job\_indusrty is unknown |
| default | Unkown field | Relevancy | Probability of “junk” data  Probability of wrong encoding | Clarify field source. In case of useful data provide field fix strategy |

**“CustomerAddress” Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field** | **Issue** | **Dimension** | **Comment** | **Fix Strategy** |
| state | Format | Consistency | “Victoria”, “New South Wales” | Fix long names to abbreviations |

**“Transactions” Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field** | **Issue** | **Dimension** | **Comment** | **Fix Strategy** |
| transaction\_date | Probability of redundancy | Validity | Date range = 1 year | Fix long names to abbreviations |
| online\_order | Missed Values | Completeness | 360 empty cells |  |
| brand | Missed Values | Completeness | 197 empty cells |  |
| product\_line | Missed Values | Completeness | 197 empty cells |  |
| product\_size | Missed Values | Completeness | 197 empty cells |  |
| product\_first\_sold\_date | Formar problem | Validity |  | Change column format |

**“NewCustomerList” Table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Field** | **Issue** | **Dimension** | **Comment** | **Fix Strategy** |
| last\_name | Missed Values | Completeness | 29 empty cells | Prepare list of future possible issues when last-name is unknown |
| gender | Unknown Categories | Accuracy | 17 entries with “U” labels | Ask customer about “U” category label mean. |
| DOB | Missed Values | Completeness | 106 empty cells |  |
| job\_title | Missed Values | Completeness | 106 empty cells |  |
| product\_size | Missed Values | Completeness | 197 empty cells |  |
| product\_first\_sold\_date | Formar problem | Validity |  | Change column format |