9th November, 2023

**SUBJECT: QUALITY ISSUE IDENTIFIED IN DATASET**.

Hi,  
Our team at KPMG has gone through your dataset and has identified some quality issues that will affect the outcome of our analysis. Here are the identified issues and the steps taken to mitigate the problems.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **DATASET TITLE** | **COLUMN TITLE** | **DATA QUALITY ISSUE** | **REMARKS** | **RESOLUTION** |
| Transactions | Customer ID | Validity | There is a record with 5034, while the dataset has a record of 4000 customers | The record is changed to "Unknown". |
| Transactions | Online Order | Completeness | There are 360 records with empty cells | The record is changed to "Unknown". |
| Transactions | Brand, Product line, Product size, Product class, Standard cost | Completeness | There are 197 records with empty cells, which is a small percentage of the whole dataset. | The records are deleted as it is viewed to be a very small percentage of the whole set which will not affect the outcome of our analysis. |
| Transactions | Product first date sold | Consistency | The date format is wrong | The format is changed to the right date format. |
| Customer Demographic | Last Name | Completeness | 125 customers with no last name. | These records will not affect our analysis so will be left unchanged. |
| Customer Demographic | Gender | Accuracy/Validity | 17 records with "U" as gender. | Replace "U" with "Unknown". |
| Customer Demographic | Gender | Consistency | There are records where the gender is wrongly spelled and some are written out in full. | Replace all categories with their abbreviations, F & M |
| Customer Demographic | DOB | Completeness | There are 87 customers with no date of birth. | They are filled with "Unknown" |
| Customer Demographic | DOB | Validity | There is a date of birth that is out of range | Changed from 1843 to 1943 |
| Customer Demographic | Job Title | Completeness | There were 506 without job titles. | The records are filled with "Unknown" |
| Customer Demographic | Default | Relevancy | These are metadata which is irrelevant to our analysis | Delete the column |
| Customer Demographic | Tenure | Relevancy | The column might not be needed in the analysis | The column is left unchanged |
| Customer Address | State | Consistency | Some records have states written in full and some are denoted with “F” or “M”. | The names are replaced with either “F” or “M”. |

I hope this mail sheds more light on how we hope to help your company in achieving its goal, we look forward to your cooperation.

Thank you.