Data Quality Report

Hello,

My name is Keerthana, and I'm with the KPMG Data Analyst (Virtual Internship) team. We analyzed the data sets that your organization submitted, and during our data quality analysis, we discovered certain errors in the data sets.

The data quality analysis is the most important phase, and due to flaws in the data set, we recommend the following mitigations to improve data quality, which will allow us to deliver better analytics and results for your firm.

- We can take a mode year value for the missing records of customer's DOB.

- We can assign a uniform last name of customers, which values are missing.

- Replaced gender 'U' with reverence to the customer's name and make a consistency.

- For tenure values, we can take a mean of rest of the values and assign the mean value to the missing fields in order to maintain the consistence of data.

- Eliminate the blank orders considering fake orders.

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

Transactions (Total records 20,000-past 3months):

Field Name	Missing Values
Online_order	360
brand	197
product_line	197
product_class	197
standard_cost	197
product_first_sold_date	197

New Customer List (Total records 1,000):

Field Name	Missing Values
last_name	29
DOB	17
job_title	106

job_industry_category	165

Customer Demographic (Total records 4,000):

Field Name	Missing Values
last_name	125
DOB	87
job_title	506
job_industry_category	656
default	302
tenure	87

Customer Address (Total records 3,999):

Missing values are not included in the Customer Address.