Hello,

This is Navaneeth joshi from KPMG Data Analytics (Virtual Internship) team. We have reviewed the data sets which were provided by your company and during the data quality analysis, we have found some errors in the data sets. The summary table below highlights key quality issues we discovered within the three data sets

|  | Accuracy | Completeness | Consistency | Currency | Relevancy | Validity |
| --- | --- | --- | --- | --- | --- | --- |
| Customer  Demographic | **DOB**: Inaccurate | **DOB**: Missing  **Last name**: Missing  **Job title, job industry**: Missing  **Tenure**: Missing | Gender:  Inconsistency | **Deceased customers**:  Filter out | **Default**:  DROP column |  |
| Customer  Address |  |  | **States:**  Inconsistency |  |  |  |
| Transactions |  | **Online order:** Missing  **Brand:** Missing |  |  | **Canceled orders:**  Filter out | **Product first sold date**: Format invalid |

The data quality analysis is the core phase and due to errors in the data set, we suggest the following mitigates in order to improve the data quality, which will eventually help us to drive better analytics, results for your company.

* We can take a mode year value for the missing records of customers’ DOB.
* We can assign a uniform last name of customers, which values are missing.
* Replace gender ‘U’ with reference to the customer name and make a consistency.
* For tenure values, we can take a mean of the rest of the values and assign the mean value to the missing fields in order to maintain the consistency of data.
* Eliminate the blank orders considering fake orders.