Dear Sir,   
  
Thank you for reaching our company for the assessment of the data quality.   
Let me share some of the key issues identified concerning your data:

**Dataset - Customer Demographic**

**Completeness**

* Last name column in 97% complete.
* DOB is 98% complete.
* Job title is 87% complete.
* Job\_industry\_category is 84% complete.
* Tenure is 98% complete.

**Accuracy**

* The DOB column consists  a incorrect value that implies a customer named Jephthah Bachmann is 175 years old which is out of range.

**Consistency**

* Gender column is inconsistent as it contains categories such as F, U, M, Annott, Femal  along with Male and Female which is not expected.
* Gender Column is 96% consistence.

**Validity**

* Last name column has 97% validity.
* DOB has 98% validity.
* Job title has 98% validity.
* Tenure has 98% validity.
* Default consists invalid data that is not relevant according to dataset.

**Uniqueness**

* no duplicate data.

**Relevancy**

* Default column is irrelevant.
* column number 13, 14 and 15 is irrelevant.
* The column owns\_car consists of irrelevant data.
* The column tenure also consists of irrelevant data.

**Dataset - Customer address**

**Completeness**

* All the columns are 100% complete.

**Consistency**

* The State column is inconsistent, VIC and Victoria denote the same, NSW and New South Wales denote the same.

**Uniqueness**

* no duplicate data.

**Dataset - Transactions**

**Completeness**

* online\_order  column in 98.2% complete.
* brand column  is 99% complete.
* product\_line is 99% complete.
* product\_class is 99% complete.
* product\_size is 99% complete.
* standard\_cost is 99% complete.
* product\_first\_sold\_date is 99% complete.

**Accuracy**

* The standard\_cost  column consists of  values such as $13.44,  $7.21 which are very low compared to other values.

**Consistency**

* The standard\_cost column is inconsistent as it contains two very low values compared to others.

**Validity**

* The product\_first\_sold\_date column is in the wrong format.
* The List price column is not represented by the currency.

**Dataset - new customer list**

**Completeness**

* Last name column in 97% complete.
* DOB is 98% complete.
* Job title is 89% complete.
* Job\_industry\_category is 84% complete.

**Consistency**

* Gender column is inconsistent as it contains U along with Male and Female which is not expected.

**Validity**

* Last name column has 97% validity
* DOB has 98% validity
* Job title has 89% validity

**Relevancy**

* The Value column is irrelevant.
* column numbers 16 to 20 is irrelevant.

**Modifications to be done by us in order to the clean the data and improve the quality standard-**

* Optimize the dataset by removing the blanks
* Remove the columns that are irrelevant which in turn reduces the dimensionality and helps maintain the relevancy and better analysis of the data.
* Inconsistent values can be replaced by standard values of similar kind.
* Using Machine learning tools, incorrect format data and data types can be handled properly.
* Removing the inaccurate data values.

**Resource provided by You:**

* You can provide the missing data if it is available.
* You can give information about the DOB of Jephthah Bachmann and the inaccurate data.

Thank you, We look forward to hearing from you soon.

Kind Regards,

Saba Firdaus Ansaria