Hi there,

As per voicemail, I found the datasets attached from Sprocket Central Pty Ltd:

- Customer Demographic
- Customer Addresses
- Transaction data in the past three months
- NewCustomerList

I have reviewed the data quality to ensure that it is ready for our analysis in phase two. I have taken note of all assumptions or issues we need to fix. As well as recommendations going forward to mitigate current data quality concerns.

I've also attached a data quality assessment report as a guideline. Let me know if you have any questions.

Thanks for your help.

Kind Regards
Analyst

KPMG Virtual Internship

Sprocket Central Pty Ltd Data Set

Data Quality Assessment Report

Transactions Sheet

Completeness

- online_order contains 360 blanks.
- brand, product_line, product_class, product_size, standard_cost, product_first_sold_date contains 196 blanks.

Validity

- list_price should be in currency format.
- product_first_sold_date not in date format.

NewCustomerList Sheet

Accuracy

- job_industry_category contains 164 n/a as input.

Completeness

- last_name contains 28 blanks.
- DOB contains 16 blanks.
- job_title contains 105 blanks.

Consistency

- gender contains mixed values U, Male, Female.
- 16 values contain U as gender.
- property_valuation contains 4 float values.
- Value contains inconsistent float values.

Relevancy

- Column Q, R, S, T, V does not contain headers.

Customer Demographic Sheet

Accuracy

- job_industry_category contains 665 n/a as input.
- Default contains irregular values.

Completeness

- last_name contains 124 blanks.
- DOB contains 86 blanks.
- job_title contains 505 blanks.
- Tenure contains 86 blanks.

Consistency

- gender contains mixed values F, M, U, Male, Female, Femal.
- 1 value contain F as gender.
- 1 value contain M as gender.
- 1 value contain Femal as gender.
- 87 values contain U as gender.
- DOB contains 1 value as 1843-12-21.

CustomerAddress Sheet

Consistency

- State mixed values NSW, QLD, VIC, New South Wales, Victoria.
- 85 values contain New South Wales instead of NSW.
- 81 values contain Victoria instead of VIC.

Recommendations

These recommendations can be used going forward to mitigate current data quality concerns.

- We need to fix the blank rows in the data and reinstate completeness.
- We need to fix the data formats for some columns.
- We need to replace the "n/a" fields in the job category with the correct ones.
- We need to provide the headers for some columns.
- We need to fix irregular data in the rows to reinstate consistency.

By using these recommendations, we can fix the issues with the data, and we can also establish a data pipeline for data preprocessing and cleaning to fix these automatically using python programming language.