

To: SprocketCentralPtyLtd@mail.com
Subject: Underlying data quality issues
Body:

Hi XYZ,

I am Preeti from the Lighthouse & Innovation team of KPMG. I went through all the data files provided by you and below recommendations will help with the mitigation of data quality issues.

Data Quality Issues and Recommendations:

1. **Customer data:** The customer data has irrelevant column 'default' that should be removed from the file set for better analysis. The 'DOB' column has some of the dates above 100 of years which needs to be checked and entered again. The 'gender' column has 5 attributes like 'Femal', 'F' which needs to be corrected to only three attributes: 'Female', 'Male', 'U'. The 'tenure' column has float datatype which is not required and must be converted to int datatype. There are some missing values in the data but they won't interfere much with the analysis and can be ignored.
2. **New Customer data:** New customer data also has 'DOB' column having values in 1950s and 1960s which needs to be corrected. The 'property_valuation' column has mixed datatype values: int64 and float64, this can cause problems with analysis and needs correction.
3. **Customer Address data:** The 'state' column has two values 'North south wales' and 'NSW' for same state.
4. **Transaction data:** Transaction data doesn't have any issues.

These issues can easily be fixed using Pandas and NumPy. After mitigating these problems, we can move forward to next task.

Regards,

PREETI
Data Analyst
he Lighthouse & Innovation team KPMG