



Respected *Sprocket Central Pty Ltd*,

This is Shubha Hegde from the *Data Analytics Team at KPMG*. After analyzing the datasets provided to us, I have found erroneous data like missing records, inconsistent entries, and contradictory values which will have an adverse effect on the decisions and predictions made.

The mitigations I suggest for data quality improvement involve the following:

- Filling in missing entries with the mode for categorical attributes like brand, product line, product size, and so on.
- Dropping records with contradictory values like an incorrect date and dropping columns (Eg.: default column in CustomerDemographic) with noisy data and > 50% missing values.
- Normalization of inconsistent data (columns representing the same information but are represented in an alternate format) like Gender: Male, Female, F, and State: NSW, New South Whales and VIC, Victoria.

CustomerDemographic :

Contradictory Data	i)Jephtah Bachmann's DOB is recorded as 1843 which is impossible (179 years old). ii) Stevena Allcock and Darnell Wistance's DOB is stated in the years 2002 and 2001 respectively but their job titles have the title 'Senior' and average age for such posts is around 40.
Missing Values	Last Name: 125 records, Job Title: 506 records. DOB: 87 records. Entire column for attribute 'tenure' is missing.
Incorrect Entries	i)Gender has been stated as 'U' for 87 records. ii) Loralyn Wonfar's gender has been incorrectly entered as 'Femal'.
Unnecessary Data	i)656 records in Job Industry have the value n/a. ii) Entire column 'default' has junk entries which provide no significant value and corrupt the data.
Inconsistent Data	i)Gender: Male, Female, F

CustomerAddress:

Inconsistent Data	State: NSW, New South Whales and VIC, Victoria.
-------------------	---



New Customer List:

Contradictory Data	Burk Wortley born in 2001 has job title of Senior Sales Associate which cannot be acquired by someone of that age.
Missing Values	Job Title: 506 records, Last Surname: 29 records.
Incorrect Entries	Gender has value 'U' for 17 records
Inconsistent Data	i) Past_3_Years and Postcode have Numbers formatted as text or preceded by an apostrophe. ii) 5 Customers have values 6,7,8, and 11 in whole numbers as well as floating-point numbers up to two decimal places.

Transactions:

Missing Values	i) Online_order: 360 records, Brand, Product_line, Product_class, Product_size, Product_first_sold_date, Standard Cost: 197 records.
Inconsistent Data	All three records with Customer_id 5034 have incorrect Standard Cost format: No \$ symbol and 7 decimal places instead of 2.

We hope to assess you regarding the above findings and help build the best decisions based on quality data and improve the growth of your company. Please feel free to reach out to the *KPMG Data Analytics Team* for any further clarifications or suggestions.

Regards,
KPMG Data Analytics Team.