

Hello,

This is Satyam Tiwari from KPMG Data Analytics (Virtual Internship) team. After reviewing the data sets supplied by your firm, we performed a data quality check and discovered certain inaccuracies in the data sets.

The study of the data quality is the key step, and because the data set contains flaws, we recommend the following mitigations to aid with data quality. This will help us produce better analytics results for your business.

- 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.
- For tenure values, we can take a mean of 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.
- Covert product_first_sold_date colum to date format.
- Insert currency sign to list price (Transactions).
- Delete column – Default (Customer Demographic).
- Replace VIC with Victoria and NSW with New South Wales in state column (customer address).

The following are the details of error encountered in the data set.

Customer Demographic (Total records 4000)

FIELD NAME	ERRORS
<u>DOB</u>	01 record = invalid value (year=1843) 88 records show 122 years of age
<u>last_name</u>	126 records Blanks
<u>Gender</u>	88 records gender 'U' Values are not consistence M, Male, F, Female, Femal, U
<u>job_title</u>	506 records Blanks
<u>job_industry</u>	656 records mention 'N/A'
<u>Default</u>	'special characters' includes null and Blanks
<u>Tenure</u>	88 records Blanks

Transactions (Total records 20000 -past 3months)

FIELD NAME	ERRORS
<u>Online_order</u>	361 records Blanks
<u>brand</u>	198 records Blanks
<u>product_line</u>	198 records Blanks
<u>product_class</u>	198 records Blanks
<u>product_size</u>	198 records Blanks
<u>standard_cost</u>	198 records Blanks
<u>product_first_sold_date</u>	198 records Blanks

Customer Address (Total records 3999)

FIELD NAME	ERRORS
state	2054 records represents NSW (short form of New South Wales) 939 records represents VIC (short form of Victoria)

Regards,
KPMG (Data Analytics Team)