Hello,

I am Chaitra Hegde from KPMG Virtual Internship Data Analytics team. We would like to thank you for giving Datasets, and we have reviewed the datasets provided by your company. During Data quanlity analysis we have found some errors, some of them are:

* In Transactions Dataset the production\_first\_date column does not have understandable date format
* Many null values are present in various columns (eg. Online\_order, brand, product\_line etc). Eliminate blank orders
* No proper naming in Gender, state column. Same attribute are given different name (eg. Female is given in 3 forms they are Female,Femal,F and male is given as Male,M etc)
* Assign a uniform last name of customers, which values are missing.
* For null values present in Dataset we can take the mean of rest of the values and assign mean values to missing values

Details of error present in Dataset

Transactions

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| Online\_order | 360 records Blank |
| brand | 197 records Blanks |
| Product\_line | 197 records Blanks |
| Product\_class | 197 records Blanks |
| Product\_size | 197 records Blanks |
| Standard\_cost | 197 records Blanks |
| Product\_first\_sold\_date | 197 records Blanks |

Customer Demographics

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| DOB | 01 record 1843/12/21  87 records Blanks |
| last\_name | 125 records Blanks |
| Gender | 88 records gender ‘U’  Values are inconsistent M, Male, F, Female, Femal, U |
| job\_title | 506 records Blanks |
| job\_industry | 656 records mention ‘N/A’ |
| Default | 3317 records value ‘special characters’ includes null and Blanks |
| Tenure | 87 records Blanks |

Customer Address

|  |  |
| --- | --- |
| **FIELD NAME** | **ERRORS** |
| State | Values are inconsistent NSW,New South Wales,VIC,Victoria |

Regards,

KPMG Data Analytics Team