Hello,

This is Viki P. Yeole from KPMG Data Analytics (Virtual Internship) team. I have reviewd the dataset provided for Data Quality Assessment that was provide by your company. During the analysis, we found some errors in the datasets

The data quality phase is the most important phase and because of this you can predict the future analytics but due to some errors you cant proceed further. so Please clear the following errors that We found before working with the dataset.

|  |  |
| --- | --- |
| **Name of Dataset sheet** | **Errors in dataset.** |
| Transactions | 1) There are some unnamed columns that should not be there in the dataset because it creates unnecessary NA values that are not required for us  2) There should be proper datatype for numeric columns Like standard cost, list\_price as they are decimal we should explain it as float  3) There are some Null values as mention  online\_order 360  brand 197  product\_line 197  product\_class 197  product\_size 197  standard\_cost 197  product\_first\_sold\_date 197 |
| NewCustomerList | 1) Clearly here it is not mentioned the names of columns no 16 to 20 so we have to drop these columns  2) There are some missing values in columns  last\_name 29  gender 0  past\_3\_years\_bike\_related\_purchases 0  DOB 17  job\_title 106  job\_industry\_category 165  3) There are unknown and unspecified values of U that we are not able to find |
| CustomerDemographic | 1) There are around 14 columns which have no name.  2) There are many values of Gender while there are only 2 available in the real world  3) There is one columns of default who have na values and we dont know much about values and also data is not known  4) here are some na values in Dataset of Customerdemographic.  last\_name 125  DOB 87  job\_title 506  job\_industry\_category 656  default 302  tenure 87 |
| CustomerAddress | 1) in the above sheet there are some unnamed columns around 20 and there are no NA values in the sheet so |

Moving forward, the team will continue with the data cleaning, standardisation and transformation process for the purpose of model analysis. After we have completed this, it would be great to spend some time with your data SME to ensure that all

assumptions are aligned with Sprocket Central’s understanding.

Kindly Regards,

[Data Analyst]