Data Quality Issues in KPMG New\_raw\_data\_update\_final : -

In CustomerDemographic Table:

* In Last Name 125 records found as Blanks from total of 4000 records. We can remove those values or we can write N/A
* In Gender Column there is some inconsistency in values like it has F, Femal, Female, M, Male, U. So we can convert F and Femal data to Female(2037 from 4000) as it has more records and M to Male(1872 from 4000) and gender U can be categorized as others.
* In D0B, there is error date 1843-12-21(175 years old) which can be removed from data and 87 records as Blanks which can be removed or we can write Not Available.
* In job\_title column total of 506 records as Blanks.
* In job\_industry\_category column total of 656 records as n/a.
* In default column it has total inconsistency which cannot be used in analysis.

In Transactions Table:

* In customer\_id column it has one outlier as 5034(3 columns)
* In online\_order, 360 records from 20000 found to be as Blanks which all can be removed.
* In brand column, product\_line, product\_class, product\_size, standard\_cost, product\_first\_sold\_date columns have 197 Blank records these records can be removed.

[SUBJECT LINE:] Information You Requested About Data Quality for the dataset from Sprocket Central.

Hi Sir/Madam,

Thanks for your request for information about Data Quality Check for Sprocket Central dataset.

I’m happy to share these details with you.

In Customer Demographic Table:

* In Last Name 125 records found as Blanks from total of 4000 records. We can remove those values or we can write N/A
* In Gender Column there is some inconsistency in values like it has F, Femal, Female, M, Male, U. So we can convert F and Femal data to Female(2037 from 4000) as it has more records and M to Male(1872 from 4000) and gender U can be categorized as others.
* In D0B, there is error date 1843-12-21(175 years old) which can be removed from data and 87 records as Blanks which can be removed or we can write Not Available.
* In job\_title column total of 506 records as Blanks.
* In job\_industry\_category column total of 656 records as n/a.
* In default column it has total inconsistency which cannot be used in analysis.

In Transactions Table:

* In customer\_id column it has one outlier as 5034(3 columns)
* In online\_order, 360 records from 20000 found to be as Blanks which all can be removed.
* In brand column, product\_line, product\_class, product\_size, standard\_cost, product\_first\_sold\_date columns have 197 Blank records these records can be removed.

Please let me know if I can be of more help.

Thank you,