Hey there,

I am Subhash, a virtual intern at KPMG Data Analytics. After reviewing the data sets supplied by your firm, we performed a data quality check and discovered certain inaccuracies in the data sets.

The following are the details of analysis done on the dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** |  | **Table Records** | |
|  | **Blank Columns** | **Before Data Cleaning** | **After Data Cleaning** |
| Transaction Data | 1542 | 20000 rows & 13 columns | 19445 rows & 14 columns |
| New Customer List | 152 | 1000 rows & 18 columns | 878 rows & 18 columns |
| Customer Demographic | 806 | 4000 rows & 13 columns | 3413 rows & 13 columns |
| Customer Address | 0 | 3999 rows & 6 columns | 3999 rows & 6 columns |

# More customer ids exist in the "Transactions table" and "Customer Address table," but not in the "Customer Master (Customer Demographic) table."

# In certain records, several fields, like the brand of purchase, an online order, or a job title, have empty values.

# Incorrect DOB information

# Values for the same property that are inconsistent

# Different data types are used for the same property (e.g., numeric values for some fields and strings for others)

# Sincerely,

# Subhash Nayaka