Manager,
KPMG.

SUBJECT:Data Quality issues in the KPMG Dataset.

The following are the data quality issues observed in the dataset:

- Accuracy:

In the customer demographic table, one date was inaccurate,stating that the individual was born In 1843.

- Completeness:
  Missing values in the last name in customer demographics which is a key field as it is included in contact information.
  The list price in the transactions table does not show the currency.
- Consistency:

In the Transactions table product_first_sold_date is written in an integer data type and has been expressed as a whole number instead of the date data type.

- Relevancy:In the customer list table columns 17,18,19 are unlabeled.In the customer demographic table the default column is irrelevant as the values do not show any relevant information.
- Uniqueness: In the customer demographic table,some values for male and female were represented with m and f or femal respectively.

To address these data quality issues and improve the overall reliability of our analyses, I propose the following strategies:

- Data Cleaning and Standardization:

Implement rigorous data cleaning processes to identify and correct inaccuracies, inconsistencies, and missing values in the dataset.

- Data Validation and Verification:

Develop validation checks and procedures to verify the accuracy and completeness of incoming data, including automated checks for duplicates and outliers.

- Regular Data Audits and Maintenance:

Conduct periodic data audits to identify and address emerging data quality issues proactively.

- Training and Awareness:

Provide training and resources to staff involved in data entry, management, and analysis to raise awareness of data quality best practices and protocols

Regards;
Melinda Muthoni.