DATA QUALITY FRAMEWORK

Hi there,

According to the task assigned, I reviewed the potato chips dataset in order to ensure that the data quality is standard and conforms to the data quality framework and below are results.

1. **Transactions Table**

This table contains inaccurate records in the first column “DATE”

Mitigation: Convert all the rows in the column into date format

It contains outliers in the column “Product quantity” i.e., 200 is unrelatable number to the dataset.

Mitigation: Delete all rows with outlier values

No other quality issues as all the standards are optimum

* Timeliness
* Completeness
* Validity
* Uniqueness
* Consistency

1. **Customer Demographics Table**

No major quality issues all the checks are optimum

* Timeliness
* Completeness
* Validity
* Uniqueness
* Consistency
* Accuracy

Examine transaction data – look for inconsistencies, missing data across the data set, outliers, correctly identified category items, numeric data across all tables. If you determine any anomalies make the necessary changes in the dataset and save it. Having clean data will help when it comes to your analysis.

Examine customer data – check for similar issues in the customer data, look for nulls and when you are happy merge the transaction and customer data together so it’s ready for the analysis ensuring you save your files along the way.

Data analysis and customer segments – in your analysis make sure you define the metrics – look at total sales, drivers of sales, where the highest sales are coming from etc. Explore the data, create charts and graphs as well as noting any interesting trends and/or insights you find. These will all form part of our report to Julia.

Deep dive into customer segments – define your recommendation from your insights, determine which segments we should be targeting, if packet sizes are relative and form an overall conclusion based on your analysis.