

To,
The Manager,
KPMG.

Subject – Data Quality Assessment of Sprocket’s Data.

The dataset provided by Sprocket Central Pty Ltd had several Data Quality issues. The problem and the measures taken to solve them are mentioned below.

Transactions:

Dimension	Features	Action
Accuracy	- - -	- -
Completeness	1. <i>blanks</i> in online_order 2. <i>blanks</i> in brand	1. Filter out blanks 2. Filter out blanks
Consistency	- - -	- -
Currency	- - -	- -
Relevancy	1. <i>canceled</i> order_status 2. gibberish product_first_sold_date	1. Remove canceled orders 2. Delete column
Validity	1. <i>plaintext</i> list_price	1. Change to currency
Uniqueness	- - -	- -

Customer Demographics:

Dimension	Features	Actions
Accuracy	1. DOB	1. Create Age column and remove ages like 174. Recommended to add an age column.
Completeness	1. <i>blanks</i> in job_title 2. <i>n/a</i> in job_industry	1. Keep it as it is. 2. Changed nothing.
Consistency	1. <i>gender</i>	1. convert male to M and convert female to F. Recommended to add a select button.
Currency	1. Y in deceased_indicator	1. Remove data of deceased customers.
Relevancy	1. <i>gibberish</i> in ‘default’ column	1. Delete column
Validity	- - -	- - -
Uniqueness	- - -	- - -

Customer Address:

Dimension	Features	Actions
Accuracy	- - -	- - -
Completeness	- - -	- - -
Consistency	1. state	1. convert New South Wales to NSW and Victoria to VIC
Currency	- - -	- - -
Relevancy	- - -	- - -
Validity	- - -	- - -
Uniqueness	- - -	- - -

Solving these data quality issues will result in KPMG providing better analysis and Sprocket getting best use of their data.

Sincerely,
Yusuf Nadim