Manager,
KPMG.

<u>Subject: Sprocket's Data Quality Assessment</u>

The dataset I received earlier had several Data Quality issues. The problems and steps taken to address them are mentioned below.

Transactions:

Dimension	Columns	Actions
Accuracy	- - -	- - -
Completeness	1. *blanks* in online_order 2. *blanks* in brand	1. Filter out blanks 2. Filter out blanks
Consistency	- - -	- - -
Currency	- - -	- - -
Relevancy	1. *canceled* order_status	1. Remove canceled orders
Validity	1. *plaintext* list_price 2. *plaintext* product_first_sold_date	1. Change to currency 2. Change to date
Uniqueness	- - -	- - -

Customer Demographics:

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

Customer Address:

Dimension	Columns	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,
Shitab Tanzim