TASK 1: KPMG Data Quality Assessment Email.

Hi,

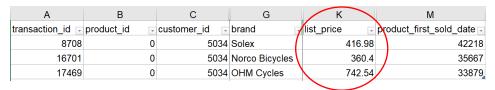
As per the voicemail, please find below all the data inconsistencies as per the data quality framework table. I have mentioned the rows in the database that require attention. Before we move to the next stage, I recommend our clients to fix these tables. It could lead to problems in future analysis. The recommendations have been highlighted.

TRANSACTIONS TABLE

Product_first_sold_date attribute: Not in date format.

Date format should be in DD-MM-YYYY for consistency with transaction date attribute

2. Customer ID 5034: Standard_cost not in dollars



Change to currency type from numbers.

- 3. 6 columns are BLANK for 196 rows.
 - product_id = 0
 - order_status = Approved
 - **Brand, product_line, product_class, standard_cost** and **product_first_sold_date** are also **BLANK** for all blanks in **product_size**.
- 4. The same product_id, brand, product_line, product_class and product_size have different list_price and standard_cost. This inconsistency is present throughout the transaction table across the various products and brands. Sample case:

A	В	С	D	E	F	G	н	1 1	1		К		М
transaction (*	_		transaction_da =	_			product line J	product_clas =	product siz	e w li		standard co ▼	product first sold da =
40		0 2448			Approved			medium	medium		544.05	\$376.84	38647
898		0 1710	03-04-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		543.39	\$407.54	42696
1129		0 161	1 27-09-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		543.39	\$407.54	34527
1665		0 1024	4 08-07-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		543.39	\$407.54	33888
1698		0 1707	7 13-10-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		543.39	\$407.54	42696
1865		0 2307	7 09-07-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		544.05	\$376.84	42105
1911		0 1109	01-04-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		543.39	\$407.54	42696
2312		0 3325	15-11-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		543.39	\$407.54	42696
2478		0 676	27-07-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		544.05	\$376.84	42710
2953		0 1410	31-10-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		543.39	\$407.54	42458
3096		0 3252	22-05-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		544.05	\$376.84	38859
3117		0 2494	18-02-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		544.05	\$376.84	38647
3737	•	0 3126	22-05-2017	TRUE	Cancelled	Norco Bicycles	Road	medium	medium		543.39	\$407.54	42696
4409		0 544	4 23-01-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		544.05	\$376.84	36668
4505		0 3487	7 04-07-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		543.39	\$407.54	42696
4590		0 640	01-12-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		544.05	\$376.84	34996
4671		0 1039	30-05-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		544.05	\$376.84	36668
5187		0 1859	21-05-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		543.39	\$407.54	35052
5326		0 1304	30-09-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium		544.05	\$376.84	38647
5500		0 220	07.02.2047	TDUE	Approved	Marco Digueleo	Dood	mondium	modium.		EAADE	F276 04	20647

5. The same product_id have different products listed with different standard_costs. This inconsistency



6. 360 tuples show BLANK for online_order attribute.

sample case:



NEW CUSTOMER LIST TABLE:

1. past_3_years_bike_related_purchases column, property_valuation column, postcode column: some tuples data has been entered as text and not as numerical data.

This should be changed into a field of numerical data as suitable

2. Column 1-5 present as hidden fields between property_valuation and rank attributes.

These do not follow proper nomenclature and hence are junk values. These columns should be named accurately with proper formatting.

- 3. BLANKS:
 - Job_title is blank for 106 tuples.
 - Last_name is blank for 29 tuples.
 - DOB is blank for 17 tuples
- 4. DOB is not in date format. It has been entered as text.

CUSTOMER DEMOGRAPHIC TABLE

- 1. BLANKS:
 - Job title is blank for 506 tuples.
 - Tenure and DOB is blank for 87 tuples.
 - Last_name is blank for 125 tuples.

2. Gender attribute in table:



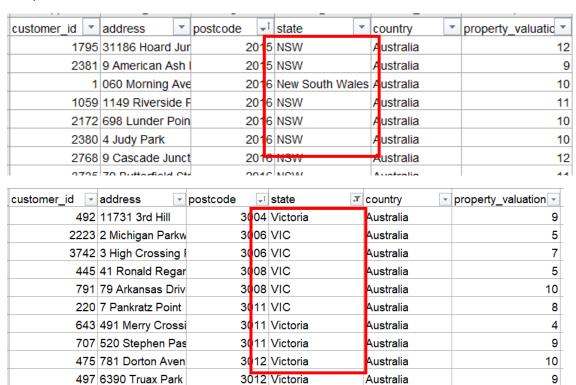
Gender should be specified either as F, M and U OR as Female, Male and U. 'Femal' is also an inconsistency found for custID 54.

- 3. Default column has junk values and should be formatted correctly in human readable form.
- 4. Customer_ID 34 has incorrect DOB (1843).

CUSTOMER ADDRESS:

1. Inconsistency – New South Wales also mentioned as NSW. Victoria also mentioned as VIC.

Sample cases:



A consistent reference should be used.

Happy to be part of the team.

Regards,

Roshni Balasubramanian