

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

1. 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

A	B	C	G	K	M
transaction_id	product_id	customer_id	brand	list_price	product_first_sold_date
8708	0	5034	Solex	416.98	42218
16701	0	5034	Norco Bicycles	360.4	35667
17469	0	5034	OHM Cycles	742.54	33879

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	B	C	D	E	F	G	H	I	J	K	L	M
transaction_id	product_id	customer_id	transaction_date	online_order	order_status	brand	product_line	product_class	product_size	list_price	standard_cost	product_first_sold_date
40	0	2448	28-11-2017	TRUE	Approved	Norco Bicycles	Road	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	1611	27-09-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium	543.39	\$407.54	34527
1665	0	1024	08-07-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium	543.39	\$407.54	33888
1698	0	1707	13-10-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium	543.39	\$407.54	42696
1865	0	2307	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	23-01-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium	544.05	\$376.84	36668
4505	0	3487	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
5592	0	2284	07-02-2017	TRUE	Approved	Norco Bicycles	Road	medium	medium	544.05	\$376.84	28647

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

transaction_id	product_id	brand	product_line	product_class	product_size	list_price	standard_cost	product_first_sold_date
64	0	Trek Bicycles	Standard	medium	medium	499.53	\$388.72	39031
83	0	OHM Cycles	Standard	medium	medium	235.63	\$125.07	37626
100	0	Solex	Standard	medium	medium	478.16	\$298.72	36146
192	0	Giant Bicycles	Standard	medium	medium	230.91	\$173.18	39031
201	0	Trek Bicycles	Road	medium	medium	533.51	\$400.13	41064
376	0	Solex	Standard	medium	medium	478.16	\$298.72	41245
400	0	Solex	Standard	medium	medium	71.49	\$53.62	37659

6. 360 tuples show BLANK for online_order attribute.

sample case:

A	B	C	D	E	F	G	H	I	J	K
transaction_id	product_id	customer_id	transaction_date	online_order	order_status	brand	product_line	product_class	product_size	list_price
98	49	333	23-06-2017		Approved	Trek Bicycles	Road	medium	medium	533.51
167	90	3177	26-04-2017		Approved	Norco Bicycles	Standard	low	medium	363.01
170	6	404	16-10-2017		Approved	OHM Cycles	Standard	high	medium	227.88
251	63	1967	11-04-2017		Approved	Solex	Standard	medium	medium	1483.2
301	78	2530	24-03-2017		Approved	Giant Bicycles	Standard	medium	large	1765.3
337	82	1615	30-10-2017		Approved	Norco Bicycles	Standard	high	medium	1148.64
343	61	1478	11-03-2017		Approved	Norco Bicycles	Standard	medium	small	586.45
399	83	1306	18-12-2017		Approved	Solex	Touring	medium	large	2083.94
477	73	367	12-02-2017		Approved	Solex	Standard	medium	medium	1945.43
530	62	1202	31-03-2017		Approved	Solex	Standard	medium	medium	478.16
559	20	3104	21-03-2017		Approved	Trek Bicycles	Standard	medium	small	1775.81
577	1	149	04-05-2017		Approved	Giant Bicycles	Standard	medium	medium	1403.5
761	4	2428	20-12-2017		Approved	Giant Bicycles	Standard	high	medium	1129.13
770	4	2414	10-09-2017		Approved	Giant Bicycles	Standard	high	medium	1129.13
844	40	3412	09-07-2017		Approved	OHM Cycles	Standard	high	medium	1458.17
915	39	2756	16-03-2017		Approved	Giant Bicycles	Standard	medium	large	1812.75
918	35	2281	30-04-2017		Approved	Giant Bicycles	Standard	medium	medium	1403.5

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  

F

Femal

Female

M

Male

U

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.

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

CUSTOMER ADDRESS:

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

Sample cases:

customer_id	address	postcode	state	country	property_valuatic
1795	31186 Hoard Jur	2015	NSW	Australia	12
2381	9 American Ash I	2015	NSW	Australia	9
1	060 Morning Ave	2016	New South Wales	Australia	10
1059	1149 Riverside F	2016	NSW	Australia	11
2172	698 Lunder Poin	2016	NSW	Australia	10
2380	4 Judy Park	2016	NSW	Australia	10
2768	9 Cascade Junct	2016	NSW	Australia	12
2795	70 Butterfield St	2016	NSW	Australia	11

customer_id	address	postcode	state	country	property_valuation
492	11731 3rd Hill	3004	Victoria	Australia	9
2223	2 Michigan Parkw	3006	VIC	Australia	5
3742	3 High Crossing f	3006	VIC	Australia	7
445	41 Ronald Regar	3008	VIC	Australia	5
791	79 Arkansas Driv	3008	VIC	Australia	10
220	7 Pankratz Point	3011	VIC	Australia	8
643	491 Merry Crossi	3011	Victoria	Australia	4
707	520 Stephen Pas	3011	Victoria	Australia	9
475	781 Dorton Aven	3012	Victoria	Australia	10
497	6390 Truax Park	3012	Victoria	Australia	9

A consistent reference should be used.

Happy to be part of the team.

Regards,

Roshni Balasubramanian