

Respected Sir/Madam,

In the **Customer Demographics dataset**:

The Customer Demographic dataset has the following missing values,

- last_name = 125
- DOB = 87
- job_title = 506
- job_industry_category = 656
- default = 302
- tenure = 87

#	Column	Non-Null	Count	Dtype
0	customer_id	4000	non-null	int64
1	first_name	4000	non-null	object
2	last_name	3875	non-null	object
3	gender	4000	non-null	object
4	past_3_years_bike_related_purchases	4000	non-null	int64
5	DOB	3913	non-null	datetime64[ns]
6	job_title	3494	non-null	object
7	job_industry_category	3344	non-null	object
8	wealth_segment	4000	non-null	object
9	deceased_indicator	4000	non-null	object
10	default	3698	non-null	object
11	owns_car	4000	non-null	object
12	tenure	3913	non-null	float64

‘last_name’, ‘DOB’, ‘default’ and ‘tenure’ have information irrelevant in making analysis of our customers. So, the rows containing these missing values can be kept as such and common/average values can be inserted here.

However, the missing values in ‘job_title’ and ‘job_industry_category’ cannot be ignored since they can be useful for understanding the financial aspect of the customer. So, it's better to get the information of these customers and update the values in the dataset.

In the **Transactions dataset**:

The Transactions dataset has the following missing values:

- online_order = 360
- brand = 197
- product_line = 197
- product_class = 197
- product_size = 197
- standard_cost = 197
- product_first_sold_data = 197

#	Column	Non-Null	Count	Dtype
0	transaction_id	20000	non-null	int64
1	product_id	20000	non-null	int64
2	customer_id	20000	non-null	int64
3	transaction_date	20000	non-null	datetime64[ns]
4	online_order	19640	non-null	float64
5	order_status	20000	non-null	object
6	brand	19803	non-null	object
7	product_line	19803	non-null	object
8	product_class	19803	non-null	object
9	product_size	19803	non-null	object
10	list_price	20000	non-null	float64
11	standard_cost	19803	non-null	float64
12	product_first_sold_date	19803	non-null	float64

All these values are important in making the analysis. Ignoring the rows containing the missing values results in loss of data, so it is better to fill these missing values into the dataset.

The ‘customer_ids’ are from 1 to 4000. However, in the Transactions dataset there is a ‘customer_id’ with 5034 – which is outside the range of ‘customer_ids’ values. Therefore, check with these values.

Also, 'product_first_sold_data' is not in date format (is in numerical format). Kindly provide it in the appropriate format.

In the **Customer Address dataset**:

The 'customer_ids' – 3, 10, 22, 23 which are in Customer Demographics Dataset are not in the Customer Address dataset. The values corresponding to these customer ids must be retrieved.

Thank You