

Hello,

Based on available dataset, there are 3 (three) sheet of data available, namely:

1. Transactions
2. NewCustomerList
3. CustomerDemographic

There are a few things to note:

1. There are some columns that don't have the correct data type.

transaction_id	int64
product_id	int64
customer_id	int64
transaction_date	datetime64[ns]
online_order	float64
order_status	object
brand	object
product_line	object
product_class	object
product_size	object
list_price	float64
standard_cost	float64
product_first_sold_date	object
dtype:	object

I think that column "product_first_sold_date" should have a datetime data type.

2. There are quite a lot of missing values in each column.

transaction_id	0
product_id	0
customer_id	0
transaction_date	0
online_order	360
order_status	0
brand	197
product_line	197
product_class	197
product_size	197
list_price	0
standard_cost	197
product_first_sold_date	197
dtype:	int64

There are 360 missing values in online_order columns, 197 missing values in product_line, product_class, product_size, standard_cost, and product_first_sold_date

In the NewCustomerList sheet, the total percentage of missing values in column job_title reaches 10.6% which shows quite a lot of missing values.

- ```
array(['F', 'Male', 'Female', 'U', 'Femal', 'M'], dtype=object)
```

5. There is one column that I don't know the meaning of the values.

Sheet CustomDemographic, column 'default'