

Dear Sprocket Central Pty Ltd officer

This is Nour elHouda from KPMG virtual internship data-analytics team, Thanks for providing us with your organizations' data. Our team have done some data exploration to assess its quality and identify ways to improve its accuracy, which will eventually help us in a better analytics results.

Errors and issues encountered are summarized below:

Starting with **Transactions dataset**

<i>Column</i>	<i>Issue</i>	<i>Suggestion</i>
online_order	360 missing values	By eliminating missing orders Brand, product_line and product size blank fields issue will be fixed.
brand	197 missing values	-
product_line	197 missing values	-
product_class	197 missing values	-
product_size	197 missing values	-
product_first_sold_date	1. Wrong data type 2. Dates refer to the same year 1970-01-01 with different time. 3. 197 Missing values	1. Change datatype to date 2. Check dates 3. Assign the earliest date available

As for the **NewCustomerList dataset:**

<i>Column</i>	<i>Issue</i>	<i>Suggestion</i>
last_name	29 Missing values	assign a uniform last name
DOB	18 Missing values	Assign a mode year value
job_title	106 Missing values	-
job_industry_category	165 Missing values	-
tenure	87 Missing values	Assign the mean value to the missing fields.

finally for the CustomerDemographic dataset:

<i>Column</i>	<i>Issue</i>	<i>Suggestion</i>
default	The column contains nonsense values	Delete the entire column
gender	Contains inconsistent values such as F or femal instead of female and M instead of male.	Correct them to female, male, and changing U to unspecified for more consistency.

The team also noticed differences in customer_id counts between datasets some are present in transactions dataset and customer address while missing on customer demographic, we would like you to confirm whether all data is collected from the same period.

Kind regards,