Hello,

This is Sarthak Chitre from KPMG Data Analytics (Virtual Internship) team. We have reviewed the 3 data sets which were provided by your company and during the data quality analysis, we have found some errors in the data sets.

We have added some recommendations and mitigated the errors which we found in the data; this will eventually help you out to drive business decisions accordingly.

**DataSets:**

**Transaction**

1. Additional customer\_id in the transaction dataset, the data is not in sync with other tables there may be some record loss.
2. We have found 360 blank items in the **online\_order** column. We can also neglect the false orders present in it, if needed.
3. We don’t need to consider cancelled (**order\_status**).
4. There were 197 records found which didn’t had **brand** of the cycles mentioned in the dataset, we cannot come up with a common brand name here for the blanks since it depends upon the (**product\_class, product\_line, product\_size, standard \_cost**) which is not mentioned as well.
5. The data mentioned in (**product\_first\_sold\_out\_date)** is not in proper format. To display it in correct way, change the Text format into Short date format. This will easily help us to analyze the dates.
6. We can come up with the profits and some other insights of each transaction in dataset.

**Customer Demographic**

1. The gender column needs a proper classification the (M & F) should be replaced with (Male & Female) respectively. We can replace the “U” in gender by assigning it to be Male or Female depending upon the customer’s name or we can keep the first name of the customer in that column.
2. In DOB column, one of the customers has a birth dated back in 1893 (Jephthah Bachmann). It should be either re-checked or we can consider it 1993 instead of 1893.

There are 87 records found to be blank, these records can be replaced with the mean age of the dataset.

1. There are 125 records found to be blank in the (**last\_name)** column. We can either come up with a common last name or kept it as it is.
2. There are 506 records whose (**job\_title)** has been kept blank.
3. 2 individuals with customer id-753 & 3790 where found to be deceased (**deceased indicator**).
4. No clear insights can be drawn from the **default** column. We need proper verification regarding what it was meant for, or we can simply eliminate.
5. 87 records where left blank in the **tenure** column. Probably that won’t affect the data.

**Customer Address**

1. Need to replace the records containing NSW and VIC with New South Wales and Victoria respectively in the state’s column.

Regards,

KPMG (Data Analytics Team)