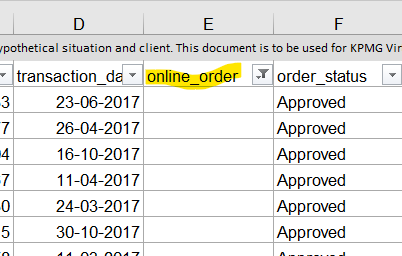
Hello,

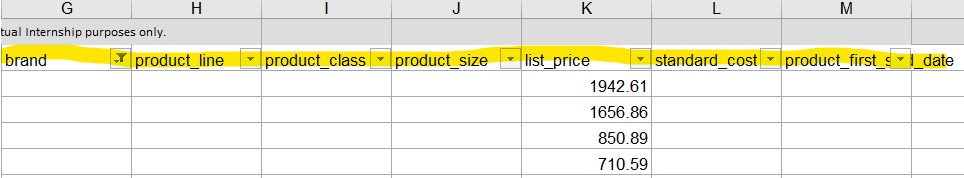
I have run some preliminary analyses to analyze data quality. Here are the few concerns that I have found:

**1.** **File: Transactions**

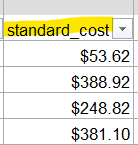
**a.** **Online order field** has blank values. About **1.8%** of data in the field has blank values. So if any analysis needs to be performed on the online order field then the analysis on 1.8% of the data can't be performed. Here is a snippet of the data for your reference.



**b.** **Brand field** has blank values. Also, other fields **like product line, Product Class, Product Size, Standard cost, and Product First Sold Date** are dependent on the brand field. About **1%** of the data in the Brand field has blank values so if any analysis needs to be performed on the brand field then analysis on 1% of the data can’t be performed. Here is a snippet of the data for reference.

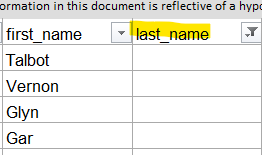


**c.** **Standard Cost** is in string format it should be in the numeric format so that operations can be performed on that data. (This Task can be easily done from our end). Here is the data snippet for reference.

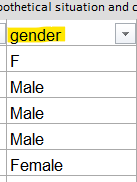


**2.** **File: Customer Demographics.**

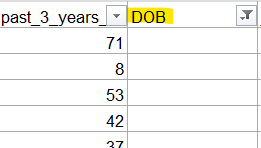
**a. Last Name field** has blank values. About **3.5%** of the data has blank values for the last name. So if any analysis needs to be performed on the last name then analysis can’t be performed on 3.5% of the data. Here is the data snippet for your reference.



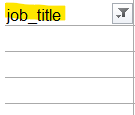
**b. Gender Field** is not standardized. (It can be easily done from our end) Also “U” value in the field is not known needs clarification on the value. We also suggest while collecting data for the gender field. Autofill values can be used or checkboxes to record the data can also be used which will solve the standardization problem. Here is the data snippet for your reference.



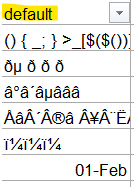
**c.** In the **Date of Birth field** about **2%** of the values have blanks. So if any analysis needs to be performed on the date of birth then analysis on 2% of the data can’t be performed. Here is the snippet of data for your reference.



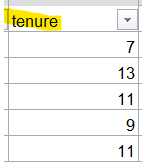
**d. Job Title** has about **12.65%** blank values. So if any analysis needs to be performed on Job title then analysis can’t be performed on 12.65% of the data. Here is the data snippet for your reference.



**e. Default field** has invalid characters. Need clarity on the field. Here is the data snippet for the reference.

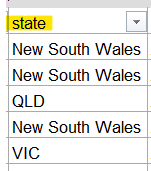


**f.** Whether the **Tenure Field** value is in months or years that needs to be specified. Here is the data snippet for reference.



**3**. **File: Customer Address**

**a. State Field** has abbreviations as well as full state names so standardization is necessary( can be done at our end). We also suggest that while collecting state data autofill values can be used to collect data. Here is the data snippet for your reference.



**b. Property** Valuation field’s unit and demonization are not specified.(Description can be mentioned to improve the quality) Here is the data snippet for reference.

