Below is a chart to displaying the issues which each column. There is no duplicated data in this sheet.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Online order | brand | Product line | Product Class | Product size | Standard cost | Product first sold date |
| 360 Missing values | 197 Missing values | 197 Missing values | 197 Missing values | 197 Missing values | 197 Missing values | 197 Missing values  Format of date is  incorrect |

Recommendations:

To filling numerical missing values with sum of each column value. For categorical, using backfill or forward fill to fill missing values.

To correct the format of time, use pd.to\_datetime function to fix the format of the time into a object.

For the Customer Demographic excel sheet there are a few issues. Below is a chart displaying the issues with each column.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Last Name | DOB | Job Title | Job Industry | Tenure | Gender | Default |
| 125 Missing values | 87 missing values | 506 Missing values | 656 Missing values | 87 Missing values | Attributes are misspelled. There should be only two (male, female) | Values are non- distinguishable, not numerical or categorical |

Recommendations:

The default has no distinguishable values, this column should be dropped.

For the numerical values will with sum or mode of each column value. For categorical, use the backfill or forward fill function to fill missing values.

To correct the misspelled gender variables, a replace function for each variable is done.

For the Customer Address Sheet only one issue is displayed. For the feature State, New south Wales is not formatted correctly. To fix this issue, values should be changed to correspond and match the rest of the data set.