To

Data Science Team Leader,

Task Completed:

1. After importing the data set to Python notebook, basic data cleaning was performed like removing null and duplicate values.
2. Some of the data types of the column were changed for example changing timestamp column to datetime.
3. Histograms was plotted to check the skewness.
4. Performed Chi-Square test to find the association between category and quantity sold.
5. Correlation analysis was also done to check the relationship between continuous variables

Findings:

1. ‘Unit price’, and ‘total’ column is skewed towards right.
2. fruits, vegetables, packaged foods, baked goods, canned foods are the top selling categories where stock re-fulling must be done quickly.
3. From the chi-square test we conclude that there is no relationship between category of good and quantity sold.
4. There is a positive relationship of 0.79 between unit price and total amount payable by the customer.

Recommendations:

1. More feature need to included in the data sets for comprehensive analysis.
2. This dataset only information of one week, more data is needed.

Best regards,

Madhusudhanan.s