Dear Data Science Team Leader,

I was assigned the task to go through sample\_sales\_data.csv of Gala Groceries to explore and understand the data deeply. I was successfully able to complete the Exploration of the data.

Following are some of my findings about the data.

* The Unnecessary column “Unnamed: 0” is dropped.

“Timestamp” column has been converted to Date-Time from Object type.

* Features are divided into Date-Time, Categorical and Numerical Features and different explorations are performed on the basis of the data.
  1. Date-Time feature is decomposed into Year, Month, Day and Hour.
     1. Year and Month found to be same for all records.
     2. Day and hour were considered for further analysis.
  2. Count plots were plotted for Categorical Features and found out that,
     1. Transaction ID is unique Transaction identifier with 7829 unique values.
     2. *Product ID is unique product identifier with 300 different products.*
     3. There are 22 different categories of Products.
     4. There are 5 Customer Types which includes, non-member, basic, standard, premium, and gold.
     5. Customers have made payments through 4 different types includes, Cash, Debit Card, Credit Card, and E-Wallet.
  3. Descriptive Statistics have been calculated and Histograms, Boxplots were plotted for Numerical Features and found out that,
     1. There is high correlation between Unit price and Total since, total is product of Unit price and quantity.
     2. Unit price hasn’t been Normally distributed and also an outlier has been founded.
     3. Quantity is a discrete variable with 1 being lowest to 4 being highest.

Recommendations:

* This dataset is a sample from a store across a week. We have to check this sample is representing the population at what extent. So, we need data from multiple stores on different days.
* Client wants to know how to better stock the items that they sell. This data is not enough to answer that question. We need to clarify the problem statement first, and then we would be able to frame the proper questions and the data we required to achieve that.
* We will need more features if we want to build a predictive model for sales, from which Client can be able to stock the items in hand and maintain the inventory efficiently.

Best Regards,

Varun Bhat