

Dear Data Science Team Leader,

I have completed the data exploration and analysis of the 'sales' data for a week provided by Gala Groceries.

Here is what I found:

- Highest sales have been of items with unit price under '10'
- Purchase quantity is mostly between 1 to 20 units, gradually dropping beyond that point.
- Transactions are quite evenly distributed among the four categories 'debit', 'credit', 'cash' and 'e-wallet', with 'cash' being slightly higher.
- 'Fruits' and 'Vegetables' are their most purchased products while 'spices and herbs' and 'pets' are the purchased categories.
- The busiest hours are 11th, 16th and 18th hours of the days.

From this dataset, it is impossible to answer that question. In order to make the next step on this project with the client, it is clear that:

- We need more rows of data. The current sample is only from 1 store and 1 week worth of data
- We need to frame the specific problem statement that we want to solve. The current business problem is too broad, we should narrow down the focus in order to deliver a valuable end product
- We need more features. Based on the problem statement that we move forward with, we need more columns (features) that may help us to understand the outcome that we're solving for.

Best regards,

Somthirtha Bhowmik