Dear Data Science Team Leader,

As per the task given to me to explore the data given by Gala Groceries to find out answer to the question – “How to better stock the items that they sell?”, I have performed following steps:

* A complete data analysis where first I checked for any null or duplicates that are present in the data.
* Transformed columns to get more deep insights into data
* Data visualization Univariate as well as Bivariate to gain a deeper insight from data.

Below are my findings:

* Top 3 product category sold are - fruit, vegetables, packaged foods.
* Bottom 3 product category sold are - spices and herbs, pets, personal care.
* Maximum number of transactions are observed on Tuesday and Wednesday.
* Lowest number of transactions are observed on Thursday and Saturday.
* The maximum number transactions are done by customer type - non-member followed by standard, premium, basic, gold.
* The maximum number of transactions are carried out by payment type - cash followed by credit card, e-wallet, debit card.
* Unit Price is positively skewed that means products with lower unit price are sold more as compared to products with a higher unit price.
* Total price is also positively skewed as it is the product of unit price and quantity which means customers are making more transactions of products with a lower unit price than the higher unit price.
* There is no significant difference between the quantity of products that customers are buying as it shows nearly same number of transactions.
* The maximum number of transactions are observed at 11 a.m., 4 p.m. and 6 p.m. and lowest number of transactions are observed at 3 p.m., 1 p.m., 5 p.m.

Although initial analysis provides some insight into data, however I cannot provide a suitable answer to the question that Gala Groceries is asking. Below are some of my recommendations about the same:

* We need more rows of data. The current sample is only from 1 store and 1 week worth of data.
* We need data for at-least 2-3 years so that we can also compare transactions during holidays and in different weather type, location etc.
* 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.
* We need features like age, gender so that we can classify transactions accordingly to find out what products and their type should we stock and how much.

Best regards,

Akshay Pandurang Paunikar