Respected Data Science team leader,

Exploratory Data Analysis (EDA) of the ‘Gala Groceries’ technology-led grocery store chain dataset and gained the following insights for solving the required business problem:

* In visualization of the dataset using seaborn library, found that the distribution of unit price as positively skewed, which shows that the grocery store to sell more products that are cheap, and few products that are really expensive.
* There are total 4 unique values exists in the distribution of quantities as 1, 2, 3 &4 and they are evenly distributed.
* Category feature shows, there are total 22 unique values with ‘fruit’ and ‘vegetables’ being the 2 most frequently purchased product categories while ‘spices’ and ‘herbs’ being the least.
* In customer type, according to the bar-plot, ‘non-members’ appear to be the most frequent type of customers, closely followed by ‘standard’ and ‘premium’ customers with least as ‘gold’.
* Payment type feature contains 4 unique values, where ‘cash’ seems to be the most frequently used method of payment and ‘debit cards’ being the least.
* With timestamp feature, we got 11th, 16th and 18th hour of the day are the top 3 hours of the day for transactions being processed.

Recommendations for the given business problem:

* The given dataset is from one store and contains only a week of data, we require more rows of data for solving the required business problem.
* We need to frame the specific problem statement that we want to solve as the current business problem is too broad and we should narrow down the focus in order to deliver a valuable end product like a mobile application for searching available products using k-means clustering.
* We require more columns (features) to understand and evaluate the business problem that we are solving for.

Best regards,

Kalyan Budharam