Dear Team Leader,

I received the sample dataset from the Data Engineering team and I’ve been analyzing the sample on behalf of the Data Science team.

Insights drawn from the data:

* The data is of 1 outlet and 1 week from 1 September 2023 to 7 September 2023. 11am is the busiest hour with regards to number of transactions
* There were 22 categories of products among which highest sold ones are from fruits and vegetables category. The least sold ones are from spices and herbs, pets, personal care & condiments and sauces.
* Unit Price and total are strongly positively correlated that means more unit price is related to more total. Quantity and total are moderately correlated.
* The total purchase made by most customers lies between 6.57 to 21.9. The maximum unit price is 23.99 belonging from medicine category and minimum is 0.19 belonging from fruit, spices and herbs & snacks categories. The average unit price ranges from 3.99 to 11.19.
* The majority of customers are non-members. The most common method of payment is cash.

As a reminder, the client indicated that they wanted to know the following: “How to better stock the items that they sell.”

With respect to this business question, my recommendations are as follows:

* The current business statement is too broad. We need to be more specific and narrow down our objectives in order to deliver a better accuracy and valuable end product. For example, can we predict the demand of products on an hourly basis in order to procure products more intelligently?
* We need more data of outlets, as well as data from the last few months, to analyse. As the current sample is of 1 store and 1 week.
* We need more datasets to help describe the outcome that we’re trying to model. For example, if we’re modeling demand for products, we may want to include information about stock levels or weather conditions.

Best regards,

Ayush Verma