# Requested Quantity Forecasting

I have proposed following two solutions:
1. Using LightGBM
2. Using the Top Down, Bottom Up and Ordinary Least Square method of Hierarchical Time Series (HTS) forecasting using the Prophet model.

The lightGBM model outperforms the second approach of HTS using Prophet, therefore we proceed with LightGBM.

Steps of Preprocessing:
1. The **Business Unit** only has one unit '2'. It is dropped since it wouldn't add any information to our model.
2. There are two **Business Groups:** '9050' and '9052'. Business Group '9050' is dropped since it has only 3 samples.
3. Each unique **lineage** is mapped to a unique **product_id.** Keeping both wouldn't add any information to the model and therefore lineage is dropped and product_id is kept.
4. The hierarchy now boils down to:
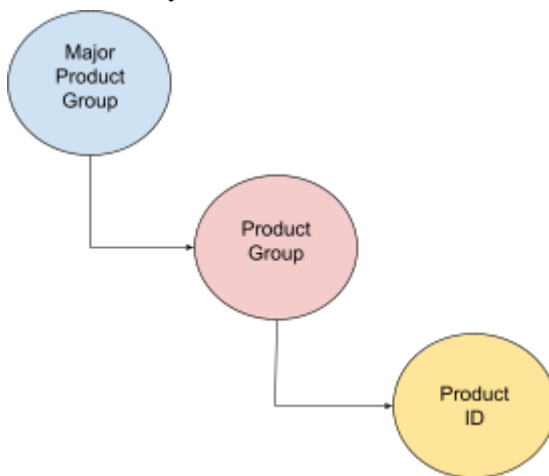
5. In the next step the data is pivoted for every product's requested_qty for every month (32 months).
6. All the other features are merged with this data and then melted to get a dataframe with 335*32=10720 rows of data.
7. The past 2 month's sell_in_qty is mapped to the present month to add an additional column and the data for the first two months is deleted because they have no past 2 month's sell_in_qty available.
8. The test and train data is randomly divided from the data after removing the last two months data.
9. Finally the forecast is done for the last two months June 2020 and July 2020.
10. If the forecasted values are negative they are considered zero.

The features taken for prediction are with decreasing order of importance are:

1. sell_shifted: Past 2 months 'sell_in_qty' named 'sell_shifted'
2. Price
3. Categorical form of Product_id
4. month_number , which consists the information of month and year(categorical)
5. Month: which contains information about the month (categorical)
6. Product Group in categorical form
7. Followed by all the discounts and Major Product Group.

This is shown in the following diagram:
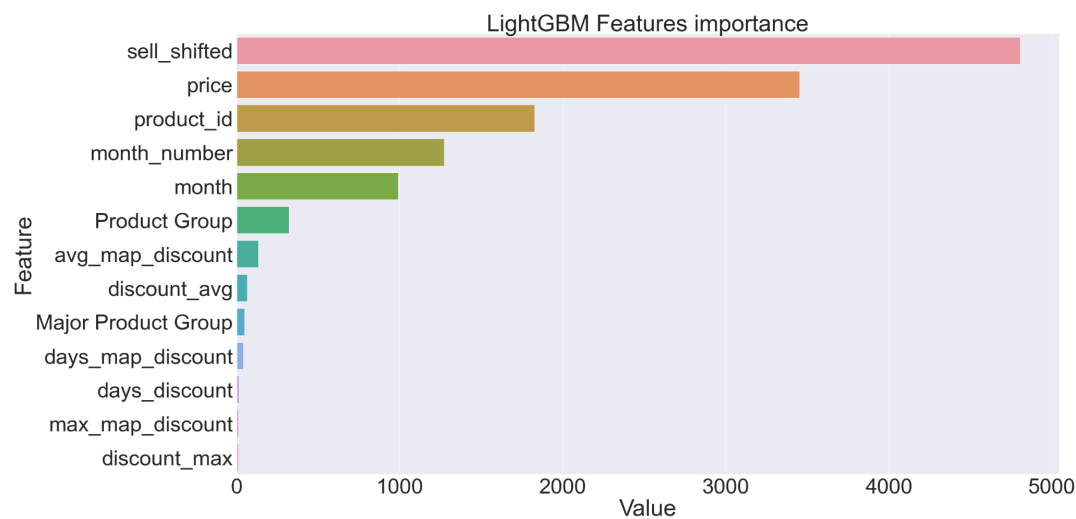
Following are the values of metrics obtained after selecting best model and forecasting for test set:

Mean Absolute Error is 780.8717030948758
Root Mean Squared Error is 2078.557098711053
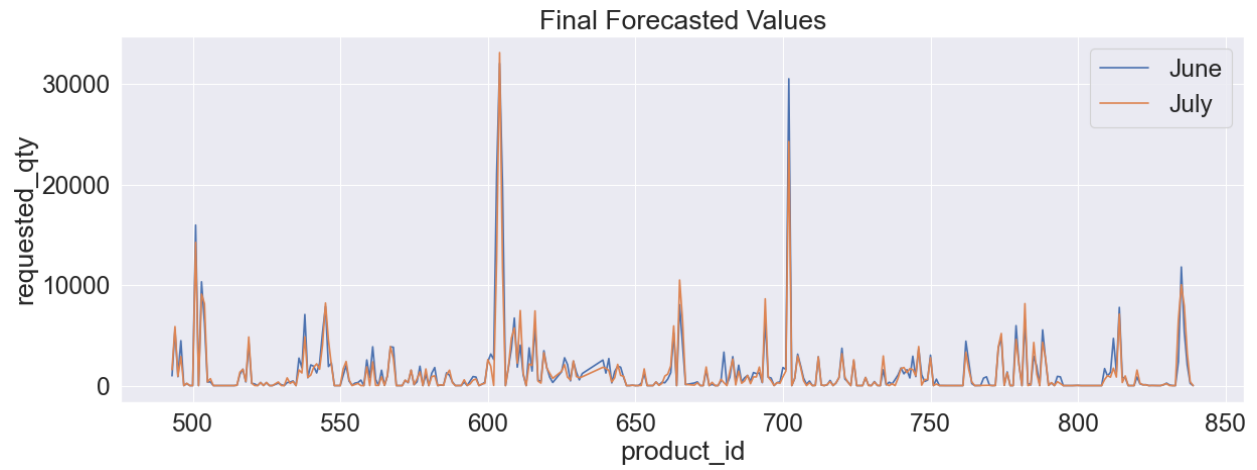**R2 score is 0.6545057152881609**
Aggregated Measure is 628.851509538425

Here, Aggregated Measure(AGM) is calculated using the following formula:
AGM = (mae+mse)/2 * (1-r2)

Here, MAPE is not the right error metric since many forecasted values are 0.

The final forecasted values are:

This result as compared to the previous may month values can be seen in the following graph: