Business Problem: (Academic Project)

"Global Mart" is an online store super giant having worldwide operations. It takes orders and delivers across the globe and deals with all the major product categories - consumer, corporate & home office.

Now as a sales/operations manager, you want to finalise the plan for the next 6 months. So, you want to forecast the sales and the demand for the next 6 months, that would help you manage the revenue and inventory accordingly.

The store caters to 7 different market segments and in 3 major categories. You want to forecast at this granular level, so you subset your data into 21 (7*3) buckets before analysing these data.

But not all of these 21 market buckets are important from the store's point of view. So you need to find out 2 most profitable (and consistent) segment from these 21 and forecast the sales and demand for these segments.

Data Understanding:

The data currently has the transaction level data, where each row represents a particular order made on the online store. There are 24 attributes related to each such transaction. The "Market" attribute has 7-factor levels representing the geographical market sector that the customer belongs to. The "Segment" attribute tells which of the 3 segments that customer belongs to. You will find the complete data dictionary for the dataset from the link below.

Note: There is a data dictionary file also added in the repo which we can refer.

Data preparation:

You would need to first segment the whole dataset into the 21 subsets based on the market and the customer segment level. Next, comes the most important data preparation step. That is to convert the transaction-level data into a time series. Thus, you would need to aggregate the 3 attributes - Sales, Quantity & Profit, over the Order Date to arrive at monthly values for these attributes. Once, you arrive at these 3 time series for each of the 21 segments, we need to find the 2 most profitable and consistently profitable segments. For this, the metric that you can use is the coefficient of variation of the Profit for all 21 market segments.

Model building:

Once you arrive at the 2 most profitable segments, the next challenge is to forecast the sales and quantity for the next 6 months. You are supposed to use classical decomposition and auto ARIMA for forecasting. Also, it is advised that you smoothen the data before you perform classical decomposition.

Model evaluation:

Once you come up with a satisfactory model, the next step would be to forecast the sales/demand for next 6 months using this model. To test the accuracy of your forecast, you must initially separate out the last 6 months values from your dataset, after aggregating the transaction level data into the monthly data. Then check your 6 months forecast using the out-of-sample figures. You can use MAPE for this.