# Business Problem

**Business understanding**:

"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 5 most profitable (and consistent) segment from these 21 and forecast the sales and demand for these segments.

You can download the transaction level data from the link below.

**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 market sector that the customer belongs to. The "Segment" attribute represents the 3 segment that the product belongs to. You will find the complete data dictionary for the dataset from the link below.

**Data preparation:**

You would need to first segment the whole dataset into the 21 subsets based on the market and the product 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 5 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 5 most profitable segment, the next challenge is to forecast the sales and quantity for the next 6 months. You have already learnt about the different time series models that you can use for forecasting. The step-by-step process that you can follow are:-

1. Plot the time series for sales/quantity.
2. Smoothen the series using any of the smoothing techniques. This would help you identify the trend/seasonality component.
3. Now, run a regression model using feature engineering, by identifying the features that can best model the series.

4. Now predict the values of the series using this model and build a new series from the difference between the original and the predicted series.

5. This residual series should be close to pure noise. Carry out auto.arima fit to confirm this.

6. If the auto.arima fit shows still some significant p,d,q parameters, then try the regression with some different features.

7. Even after repeated attempt, if you are unable to get rid of the p,d,q parameters in auto.arima, it means that the series has significant autoregressive component and cannot be captured by regression alone.

8. In such a case, proceed with the ARMA/ARIMA technique on the original series itself.

9. Come up with the optimal values of p,d,q to improve the model.

**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. The parameter that you can use is the MAPE.

# Checkpoints

**Checkpoint 1: (Data Understanding & Data Preparation) (10% marks)**

- Make the subsets of the complete dataset using the 7 factor levels of "Market" and 3 factor levels of "Segment". (3%)

- Aggregate the transaction level data to the month level time series using the "Sales", "Quantity" and the "Profit". (3%)

- For each of the 21 segments, find the profitability index. Find the top 5 profitable segment. (you can use Coefficient of Variation as the measure for this). (3%)
- For each time series that you are forecasting, separate out last 6 months data for out of sample testing. (1%)

## Checkpoint 2: (Time series modelling) (60%) (6% per model x 10 time series)

- Plot and smoothen the time series. (1.5%)
- Use feature engineering to come up with the best regression fit. (1.5%)
- Check the residual series for White noise. (1.5%)
- Find the optimal value of p,d,q for ARIMA modelling (1.5%)

## Checkpoint 3: (Model Evaluation) (20%) (2% per model x 10 series)

- Calculate the MAPE value for the regression - forecasted values using out-of-sample data (1%)
- Calculate the MAPE value for the ARIMA - forecasted values using out-of-sample data (1%)

## Checkpoint 4: (Result Interpretation) (10%)

- Based on the 10 forecasts that you have got, comment on the major source of revenue for the store over the next 6 months. (5%)
- Also, comment on the resource allocation that the store should do, based on demand forecast. (5%)