Introduction:

Demand forecasting is a critical requirement to manage the Supply Chain of any FMCG company. Typically forecasts are generated weeks or month in advance and fulfillment is planned.

Forecasters encounter three main issues when generating the forecasts -

1. FMCG companies have any many Stock Keeping Units (SKUs) and the SKUs change pretty frequently. For example even the change of the packing could change the SKU number.
2. The second problem is the that the forecasts have to be generated at different product and geographical hierarchies.
3. The final issue involves the fact that demand is a function of external factors including seasonal effects, festivals and even weather. Incorporating all these effects into forecast is challenging.

Problem Statement:

Given the data set as described later, there are 2 specific goal of this exercise

Goal 1:

The first goal is to design a product that will solve the above problem for FMCG companies. The product must have the following features -

1. All technology stack components should come out of the open source stack.
2. Which database will you use that will store historical sales value (example for 22 product types is given) for 100K SKUs at a daily aggregation level.
   a. What type of database will you use and why?
   b. How will you include external data such as weather data here?
3. What other components will be part of this architecture
4. How will the user interact with this system - please provide some scenarios and functionalities

Please provide a write-up using your preferred tool.

Goal 2:

Give the data, generate the forecast for the next 6 months.
The forecast should be given at the following levels -

● National Brand sales
● Regional Warehouse Brand sales
● Dealer Level Brand sales

- National Sub-Brand sales
- Regional Warehouse Sub-Brand sales
- Dealer Level Sub-Brand sales

In addition to the given forecast, you are encourage to check online sources for weather and festivals and if required incorporate them in the forecast.

Please provide *replicable* source code in either Python or R. Code should be submitted as a Python Notebook or as R Markdown. If you giving raw code files please include a detailed writeup explaining your analysis.

Data:

The data headers have the following description
- Year - Year of Sales
- Month - Month of Sales
- Sub-Brand - The SKU
- Brand - Type of product
- Regional Warehouse - Regional Warehouse
- Dealer Code - Dealer Code
- Retail Quantity - Quantity sold to the dealer