Case Study

Data – Case\_Study\_ML.xlsx (Base Sheet) –

The following problem can be solved using R or Python. The code and the output file should be in a zipped folder.

It contains 4 columns:

1. SKU
2. ISO\_Week
3. Sales
4. Season

Promotion is going on for the following SKU on these weeks:

|  |  |
| --- | --- |
| FU | Weeks |
| SKU\_101 | 2018-11 |
| SKU\_101 | 2018-12 |
| SKU\_101 | 2018-13 |
| SKU\_101 | 2018-14 |
| SKU\_101 | 2018-15 |
| SKU\_101 | 2018-16 |
| SKU\_101 | 2018-17 |
| SKU\_101 | 2018-19 |
| SKU\_101 | 2018-20 |
| SKU\_101 | 2018-24 |
| SKU\_101 | 2018-26 |
| SKU\_101 | 2018-45 |
| SKU\_101 | 2018-48 |
| SKU\_101 | 2018-49 |
| SKU\_101 | 2018-50 |
| SKU\_101 | 2018-51 |
| SKU\_101 | 2018-52 |
| SKU\_103 | 2018-21 |
| SKU\_103 | 2018-22 |
| SKU\_103 | 2018-23 |
| SKU\_103 | 2018-24 |
| SKU\_103 | 2018-25 |
| SKU\_103 | 2018-26 |
| SKU\_103 | 2018-27 |
| SKU\_103 | 2018-28 |
| SKU\_103 | 2018-29 |
| SKU\_103 | 2018-31 |
| SKU\_103 | 2018-32 |
| SKU\_103 | 2018-33 |
| SKU\_103 | 2018-34 |
| SKU\_103 | 2018-35 |
| SKU\_103 | 2018-36 |
| SKU\_103 | 2018-37 |
| SKU\_103 | 2018-46 |
| SKU\_103 | 2018-47 |
| SKU\_103 | 2018-48 |
| SKU\_103 | 2018-49 |
| SKU\_103 | 2018-50 |
| SKU\_103 | 2018-51 |
| SKU\_103 | 2018-52 |

The objective of this exercise is to predict sales of 2018-42 to 2018-52 of each SKU. The data present in the mentioned week are only for computation of accuracy and bias only. So please use the sales data till 2018-41.

**The following pre-processing treatment needs to be done if needed:**

1. **Initial zero removal:**

There are some SKU’s for which initial week sales are 0.it means sales was not started after that period. Those weeks needs to be removed before fitting the data into the model.

1. **Duplicate treatment**
2. **Outlier treatment**

**Expected Model output 1:**

The expected output should be in a excel in the below Format:

|  |  |  |
| --- | --- | --- |
| SKU | ISO\_Week | Forecast |
| SKU\_101 | 2018-42 |  |
| SKU\_101 | 2018-43 |  |
| SKU\_101 | 2018-44 |  |
| SKU\_101 | 2018-45 |  |
| SKU\_101 | 2018-46 |  |
| SKU\_101 | 2018-47 |  |
| SKU\_101 | 2018-48 |  |
| SKU\_101 | 2018-49 |  |
| SKU\_101 | 2018-50 |  |
| SKU\_101 | 2018-51 |  |
| SKU\_101 | 2018-52 |  |
| SKU\_102 | 2018-42 |  |
| SKU\_102 | 2018-43 |  |
| SKU\_102 | 2018-44 |  |
| SKU\_102 | 2018-45 |  |
| SKU\_102 | 2018-46 |  |
| SKU\_102 | 2018-47 |  |
| SKU\_102 | 2018-48 |  |
| SKU\_102 | 2018-49 |  |
| SKU\_102 | 2018-50 |  |
| SKU\_102 | 2018-51 |  |
| SKU\_102 | 2018-52 |  |
| SKU\_103 | 2018-42 |  |
| SKU\_103 | 2018-43 |  |
| SKU\_103 | 2018-44 |  |
| SKU\_103 | 2018-45 |  |
| SKU\_103 | 2018-46 |  |
| SKU\_103 | 2018-47 |  |
| SKU\_103 | 2018-48 |  |
| SKU\_103 | 2018-49 |  |
| SKU\_103 | 2018-50 |  |
| SKU\_103 | 2018-51 |  |
| SKU\_103 | 2018-52 |  |

**Expected output 2:**

Another excel sheet named “**Results**” showing the accuracy and the bias of the model. It should be prepared in such a way that accuracy and bias for each SKU can be seen easily when filtered.