Help The Ag-Industry Solution

Aim: To understand trends in APMC (Agricultural produce market committee)/mandi price & quantity arrival data for different commodities in Maharashtra.

Objective:

1. Test and filter outliers.
2. Understand price fluctuations accounting the seasonal effect
    1. Detect seasonality type (multiplicative or additive) for each cluster of APMC and commodities
    2. De-seasonalise prices for each commodity and APMC according to the detected seasonality type
3. Compare prices in APMC/Mandi with MSP(Minimum Support Price)- raw and deseasonalised
4. Flag set of APMC/mandis and commodities with highest price fluctuation across different commodities in each relevant season, and year.

Solution :

1. The two datasets (monthly_cmo and cmo_msp) had 11 and 6 columns resp. The 'Monthly_cmo' dataset had max,min,modal prices for different commodities along diffrent APMC for three years (2014-16). Whereas the 'Cmo_msp' had MSP prices of different commodities for years 2012-16.

2. The first task done was to load and explore the dataset . The commodity column in monthly_cmo dataset had lower and upper case values for same dataset along with their whitespaces . These were removed and all commodities were made into lower case .

3. Datatypes were converted after creating two new column "Year and Month" by dropping their previous version in the given dataset . Further exploration of unique commodities,APMC, no. of clusters(4720),datatypes etc. of each columns was done

4. The next thing was to detect outliers. Outliers are observation points that are distant from other observations. There are many ways to detect and filter outliers . In this solution outliers are first detected using a simple approach and then filtered out from the working dataset . The approach to detect outliers is as follows :

    a. Divide the total dataset for 3 years (2014-16)

b. Draw boxplots of modal prices of commodities separately for all three years to get the rough idea for no. of outliers
c. Now, a function is written to detect the outliers , i.e the datapoints that are below lower bound and above the upper bound of modal prices .Lower_bound = q1 - (1.5 * (q3-q1)), Upper bound = q3 + (1.5 * (q3-q1)) where q3 and q1 are 75%tile and 25%tile respectively
d. The outliers were then filtered .

5. The total data of three years was then combined

Detecting Seasonality

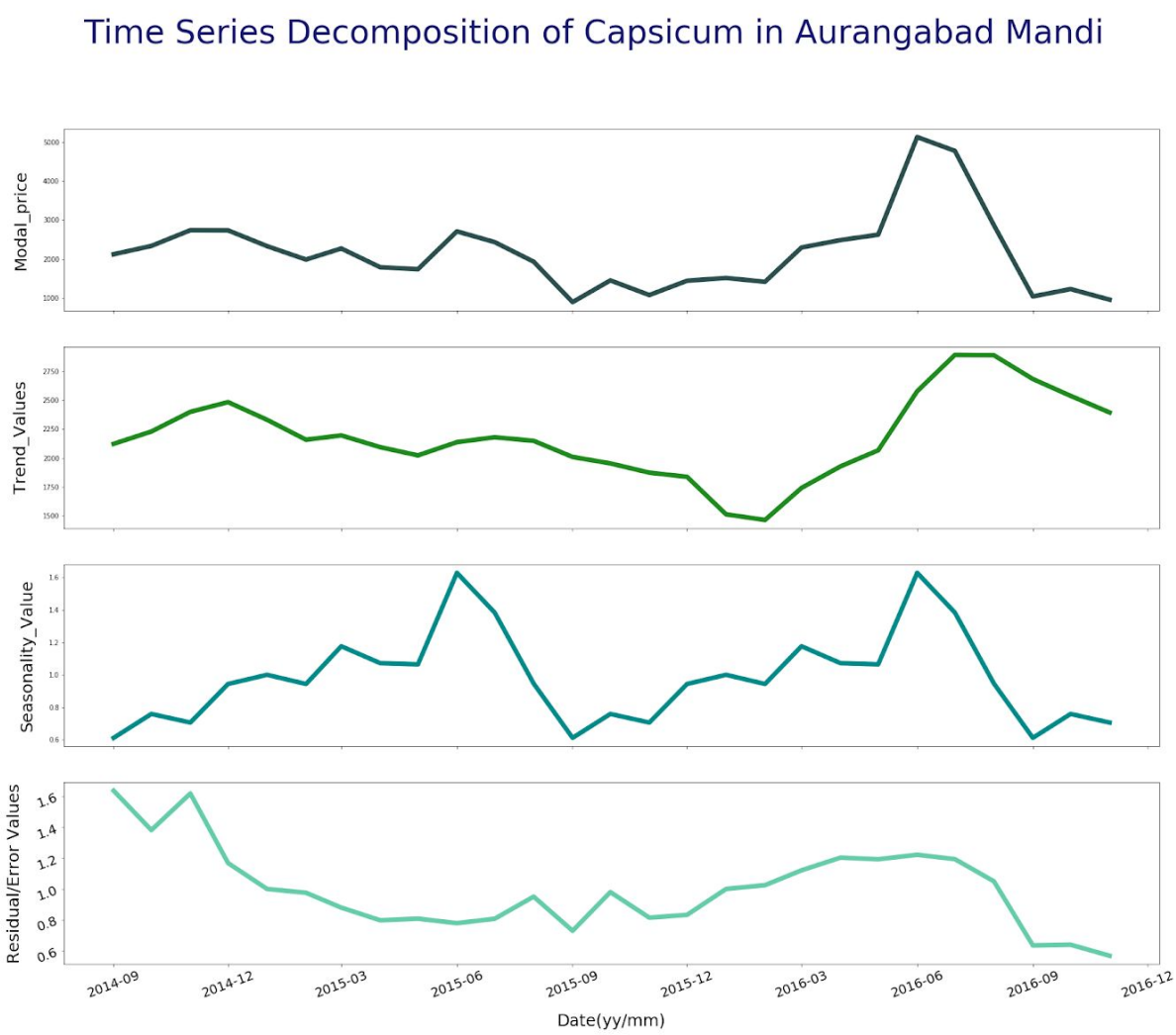
6. The first step was to calculate trend . Trend in simple terms is tendency of data over long period of time (up/ down / constant) . A method of moving average has been used to detect the trend . The moving average is calculated for a cluster of APMC, Commodity for months present for each year (2014-16) .

7. The data is then detrended (both additive/multiplicative).

8. After that seasonality of for each cluster is calculated . Seasonality ,simply put, is a repetitive pattern that occurs over a period of time . Two types of seasonality has been calculated - additive and multiplicative . These have been calculated using mean of detrended values ,additive/ multiplicative , respectively for each cluster . Also residual(Noise) of each cluster was calculated

9. Then using the comparison of the sum of auto-correlation function for additive and multiplicative seasonality , it was decided what type of seasonality the clusters have . It came out to be Multiplicative .

10 . An example of this is Time Series Decomposition for each cluster of APMC, Commodity is of Capsicum in Aurangabad APMC/Mandi for three years . It is as follows :



Deaseasonalising the prices :

11. After this modal_prices for each cluster of APMC,Commodity for three years was deasonalised . This was done dividing modal_prices by seasonality values(multiplicative)

Compare prices in APMC/Mandi with MSP(Minimum Support Price)- raw and deseasonalised

12. First, the cmo_prices dataset was cleaned and sorted according to the need .

13 .The deasonalised and cmo_msp dataset was divided into 3 datafarmes based on their years

14 . Deasonalised and cmo_msp datasets were joined for each year .

15 . The ‘msp’ were then compared with modal prices and deasonalised prices . Accordingly two new column were created and values of “Below_MSP”/”Above_MSP” were assigned based on modal prices and deasonalised prices relative to ‘msp’ prices of commodity .

16. The dataset of all three years was then combined

Clusters with Highest Price Fluctuations

17 . For this task for each row the fluctuation is calculated as difference between Max and Min price .

18. Then the monthly_cmo is divided into three years(2014-16) as three separate dataframes

19. For each dataframe maximum fluctuation Cluster is identified and stored

20. The final data for all three years is combined and stored