Introduction **2**

Methodology to attempt these tasks: **3**

Outliers - modified z score 3

Seasonality Detection 3

Seasonality Exploratory Analysis 4

APMC Cluster Seasonality Analysis 5

Trend and Seasonality detection 6

Trend detection: Rolling Mean 6

Seasonality detection: Differencing 7

b.) Deseasonalised Prices for commodity 7

3. Minimum support price analysis 9

Introduction

The assignment has three parts:

1. Test and filter outliers
2. Understanding the price fluctuations accounting the seasonal effect
    a. Detecting seasonality in
        i. Commodities
        ii. APMC clusters
    b. Deseasonalise prices for commodity and APMC
3. Compare prices in APMC with the minimum support price
    a. Raw
    b. deseasonalised

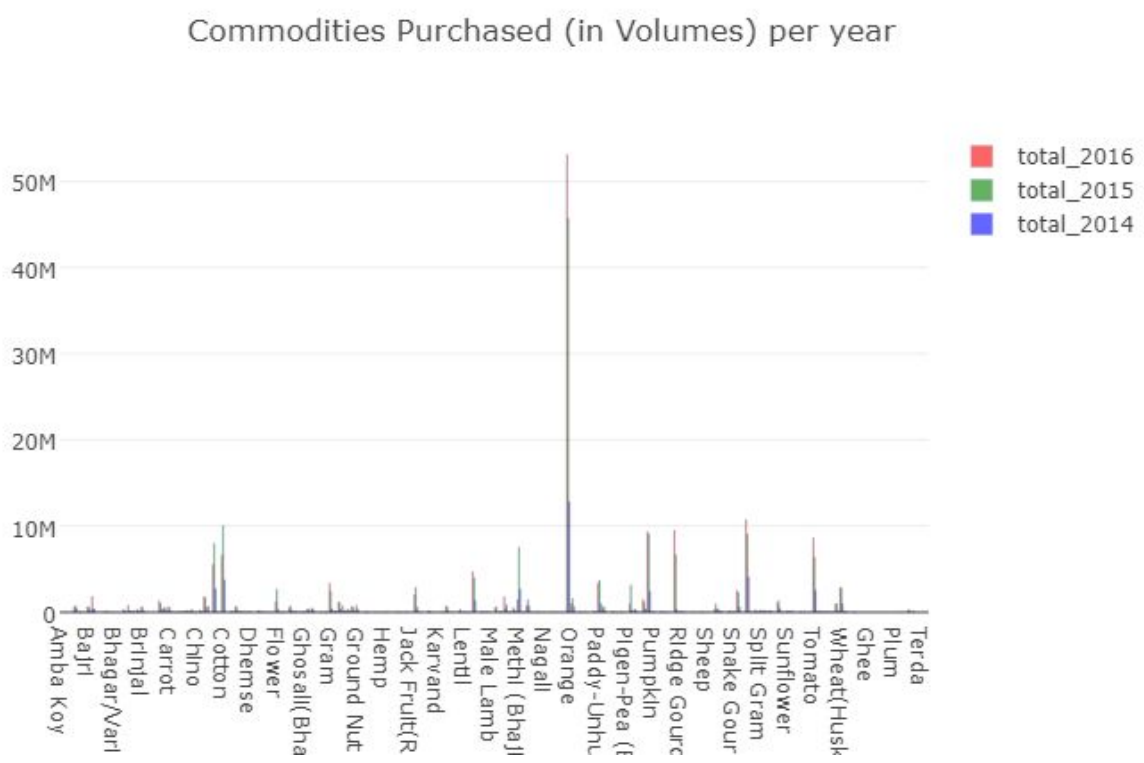
Methodologies (for above-assigned tasks):

Outliers - modified z score

Data has four numeric columns, arrivals_in_qtl, modal_price, min_price, max_price. By plotting the histogram for each of these columns, It can be seen that our data is highly skewed and hence simple parametric approach will not work. I first used z score to detect and eliminate outliers but then while researching on some optimal technique (as the dataset has only 4 columns) I come across **modified z score**.
Modified z score method was first applied on all four columns but there was a huge drop in rows which led to a decrease in the number of commodities and other important features of data. So, i applied it over modal_price, min_price, max_price. This resolved the problem of unwanted feature reduction.
Thing is it is possible to have a large amount of order or arrival of a commodity in Mandi.

Seasonality Detection

From the commodity top, 10 commodities in market share are selected for further analysis of Seasonality and fluctuations in price and demand.



Seasonality Exploratory Analysis

The demand of these commodities is plotted with modal_price to analyse the demand price fluctuations.

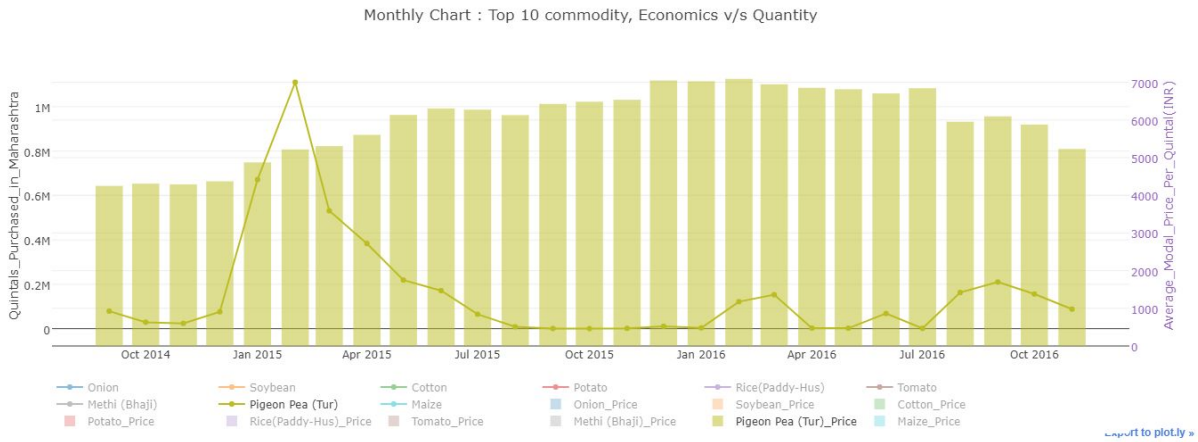


fig. Pigeon Pea (Tur) Demand v/s Avg. Modal Price

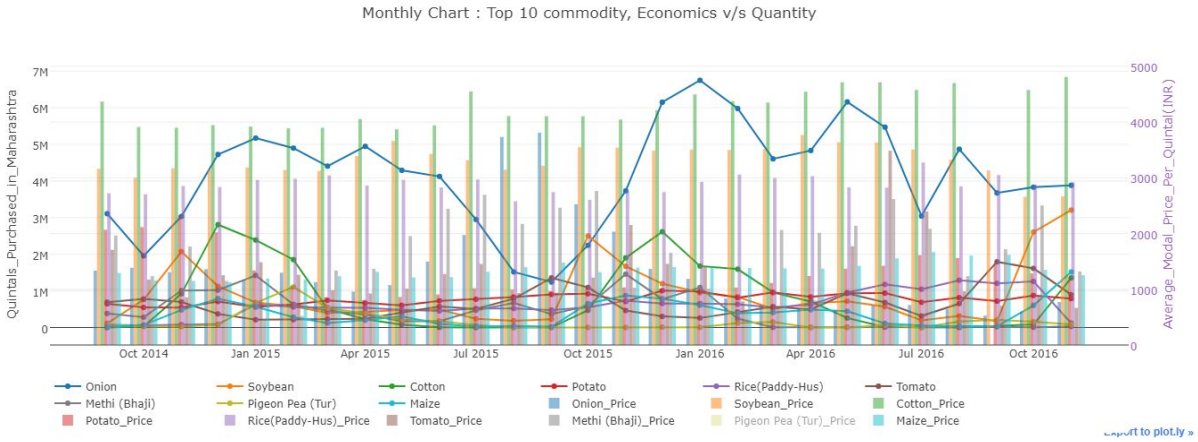


Fig. Top 10 commodities Demand v/s Average Modal Price

The following results were obtained from this analysis:

1. Maize, Cotton, Soybean, Pigeon Pea are winter crops so their demands are higher in the winter months (NOV to FEB)
   There is no particular co-relation between the price and the amount purchased for both crops

2. Methi (Bhaji) has interestingly dropped its demands in the year 2016
   There is a negative co-relation between Price and Demand (As price Increases, Demand Decreases)

3. Onion has seen a steep drop in demand in the month of Aug - Nov 2015 , owing to the high price increase
   There is a negative co-relation between Price and Demand (As price Increases, Demand Decreases)

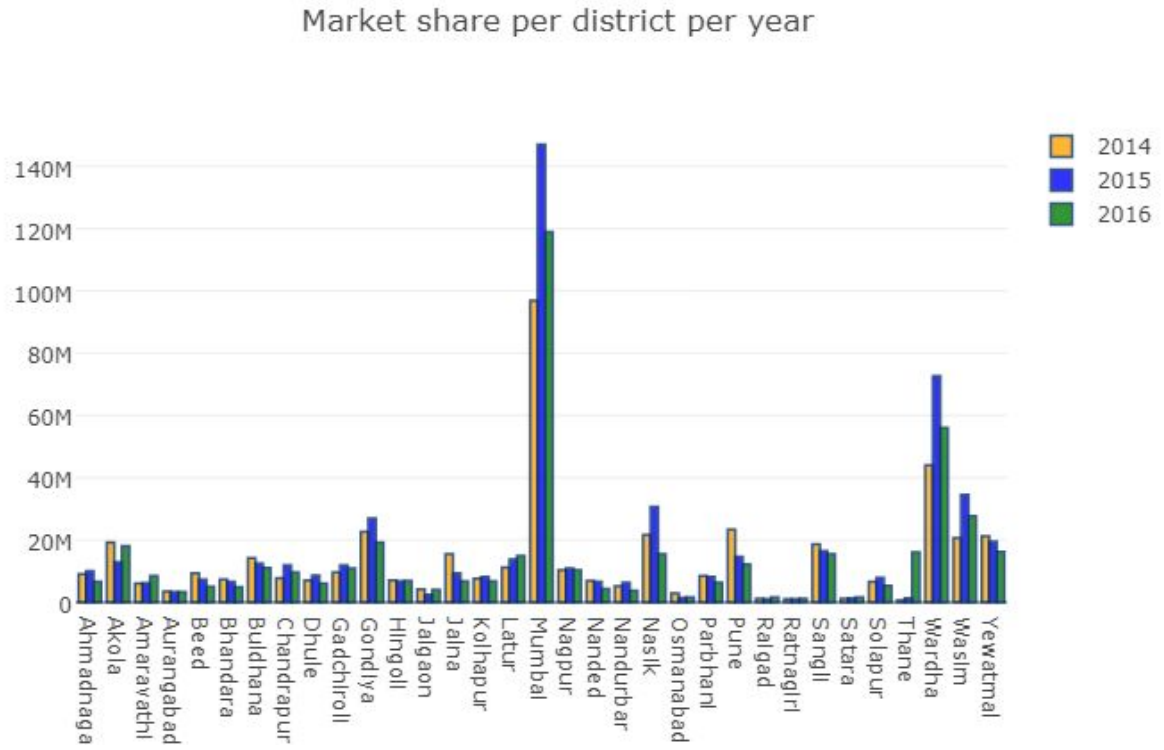
4. Potato does not show any particular seasonal trend in its demand. However, over the time the demand has increased.
   There is a negative co-relation between Price and Demand (As price Increases, Demand Decreases)

5. Rice has no particular seasonal trend. However, over the time, the demand has significantly grown.
   There is no co-relation between the price and demand for this crop.

6. Tomato has no seasonal Trend.
   There is a negative co-relation between Price and Demand (As price Increases, Demand Decreases)

APMC Cluster Seasonality Analysis

For apmc cluster i selected a district and those which comes under that district were considered as an apmc cluster.



For analysis, a district is selected and the demand v/s average modal price graph is plotted to see the fluctuations in price in that particular APMC cluster/district

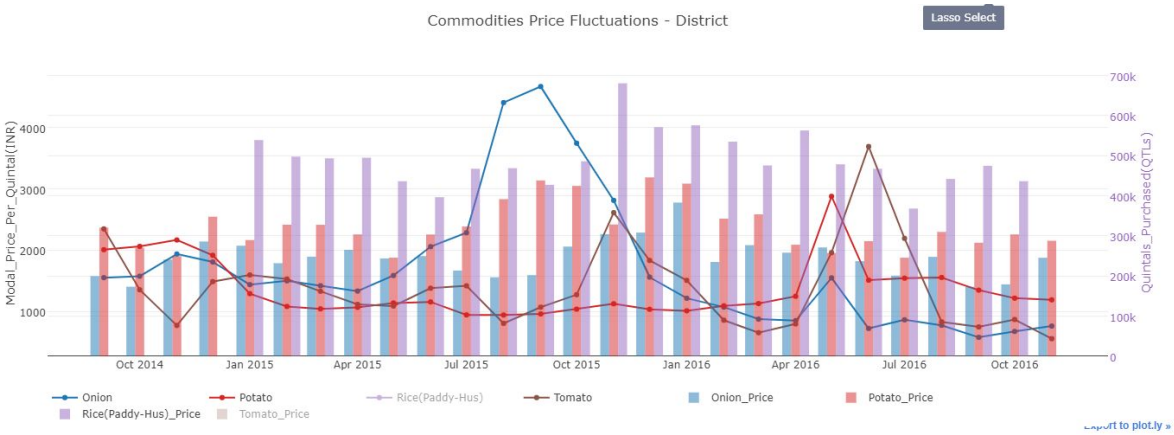


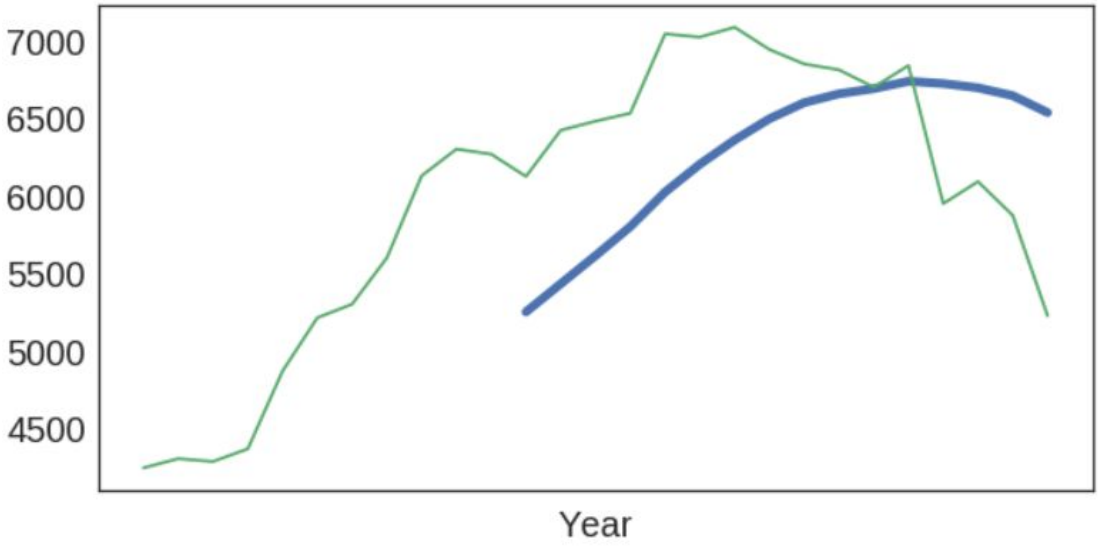
Fig. Mumbai Commodity demand v/s Avg. Modal Price

We can see that there lie components of trend and seasonality in both demand and price data. These trends and seasonalities are detected by using rolling mean and differencing respectively.

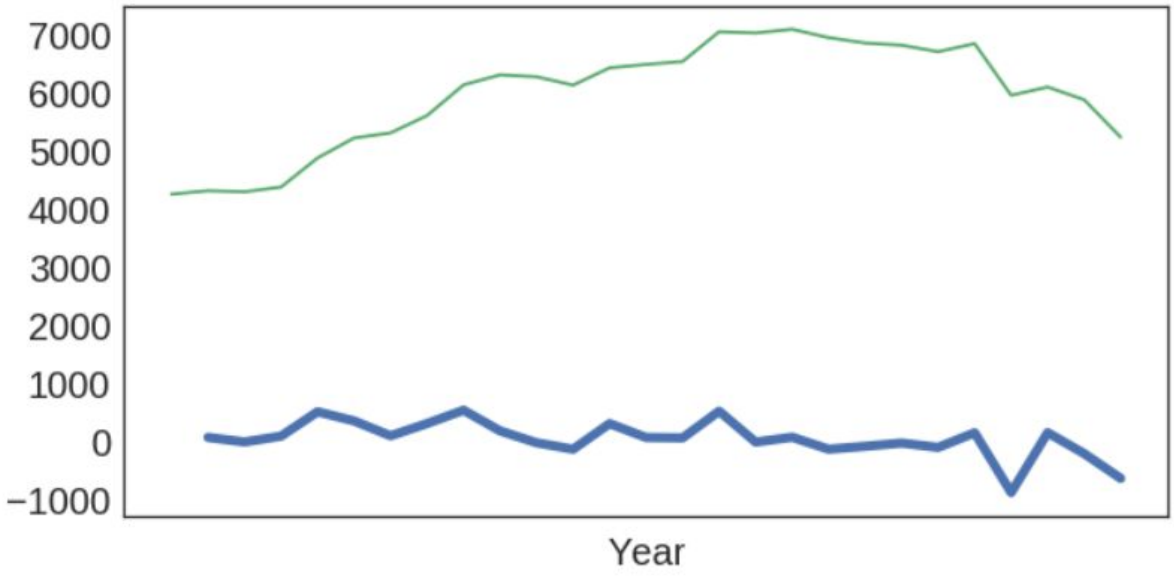
Trend and Seasonality detection

First I ,used Rolling means and differencing to obtain the underlying trend and seasonality in the commodity. But this method

Trend detection: Rolling Mean



Seasonality detection: Differencing



But these methods were not quite accurate.

b.) Deseasonalised Prices for commodity

I switched to method from time series analysis section of statsmodels library- (**statsmodels.tsa.seasonal.seasonal_decompose**)
seasonal_decompose which basically decomposed the time series into respective components and returns a object which contains these 3 components along with observed values.

- Observed
- Trend
- Seasonality
- Residual

If the seasonality component is removed from the observed prices (the raw price), the obtained prices will be deseasonalised.

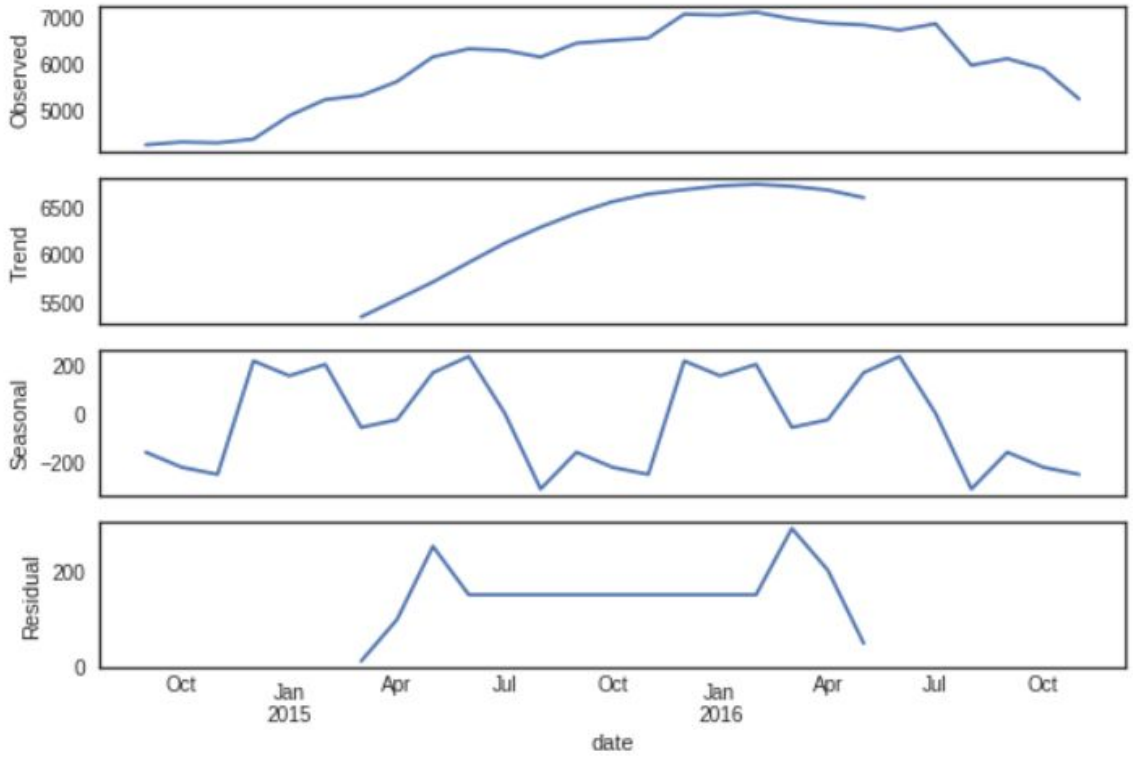


Fig. Seasonal Decomposition

<matplotlib.axes._subplots.AxesSubplot at 0x7fc2764ac240>

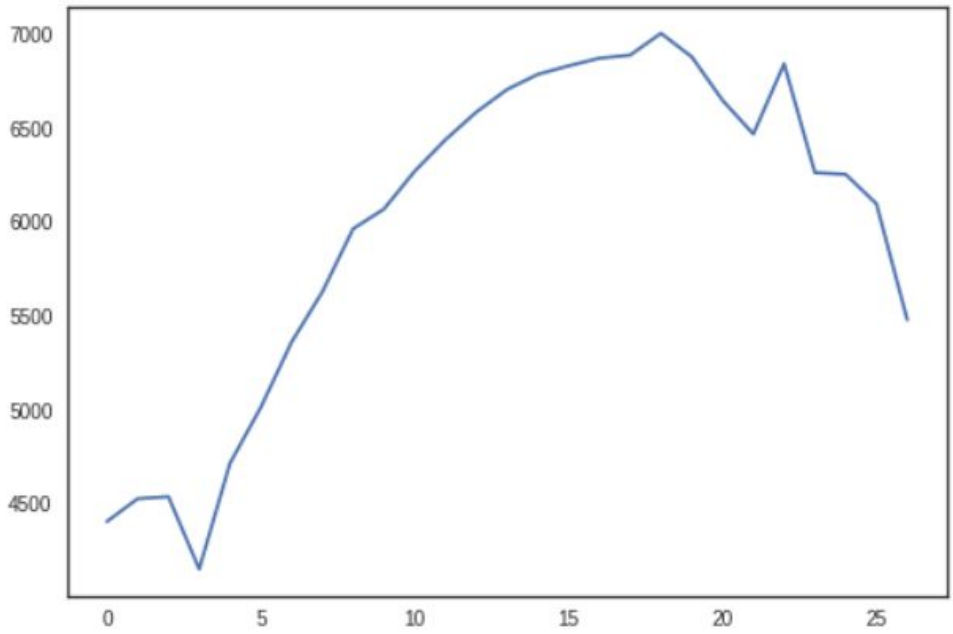


Fig. Deseasonalised Prices

3. Minimum support price analysis

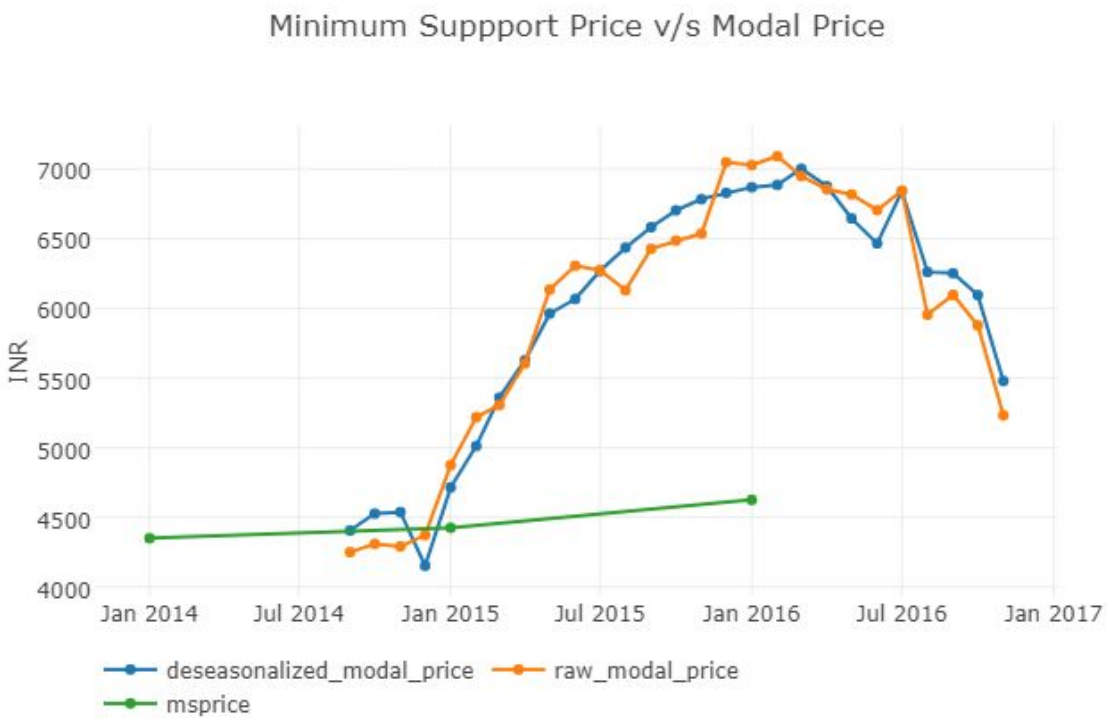


Fig. For commodity = ‘Pigeon Pea (Tur)’

Minimum support price analysis gives an overall idea about the market. It gives a basis which can be used to observe how high the prices are.

MSP data contains ~30 commodities. These prices plotted against the raw modal price of the commodity and deseasonalised price of the commodity.

Conclusion:
I’ve completed 3 out of 4 tasks mentioned in the assignment however, I am not able to figure out an automated way to flag set commodities and mandies with highest price fluctuations as there are a lot of commodities. Manually drawing all the commodities will be time consuming but doable.