

a) Link to Github for Python script for webscrapping: -

<https://github.com/shubhamJain1203/Agritech>

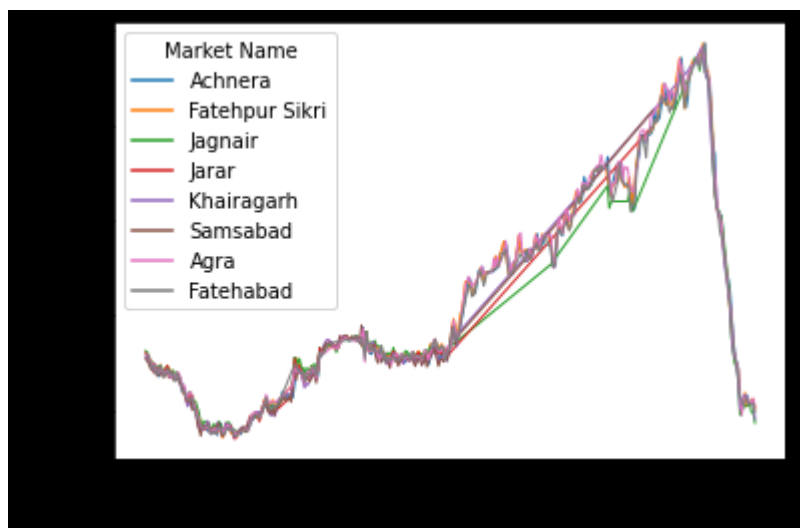
b) Major markets in Agra district are: -

Achnera, Fatehpur Sikri, Agra, Fatehabad

```
df['Market Name'].value_counts()

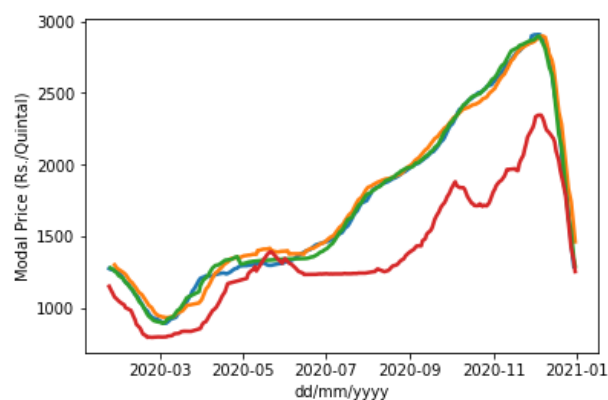
Achnera      306
Fatehpur Sikri 290
Agra         285
Fatehabad    258
Jagnair      187
Khairagarh   187
Samsabad     168
Jarar        146
Name: Market Name, dtype: int64
```

Pattern of Modal prices is as follows: -



We observe that in the beginning of the year the prices for potato were at lower end and the prices rose steadily till the end of november. In december 2020 there was sudden drop in prices of potato.

If we try to observe top four markets (i.e Achnera, Fatehpur Sikri, Agra, Fatehabad). We see that prices went up to 3000 rupees per quintal in these markets near to end of november.



c) Answers:

I. Data pre-processing required: -

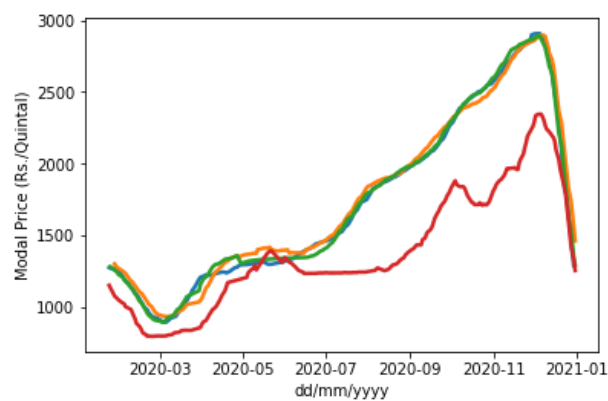
- 1) Remove NaN type rows, columns.
- 2) Convert dates to datetime objects.
- 3) Integer encoding for Market name and Variety.
- 4) Normalizing All the 3 types of prices.

II. Features to be used can be: -

- 1) Market Name
- 2) Variety
- 3) Grade
- 4) Min price
- 5) Max price
- 6) Price Date (after pre-processing)

III. This problem can be framed as regression problem.

The target Variable will be Modal prices.



- IV. The algorithm that can be used is LSTM as the data is given in form of time series.
- V. Loss function that can be used is Mean square error loss.