Gramoday Assignment

- Bindu Priyatha Y

Part-B

Below are the price patterns of each Market

Market:Khairagarh

Market:Jarar

Market:Jagnair
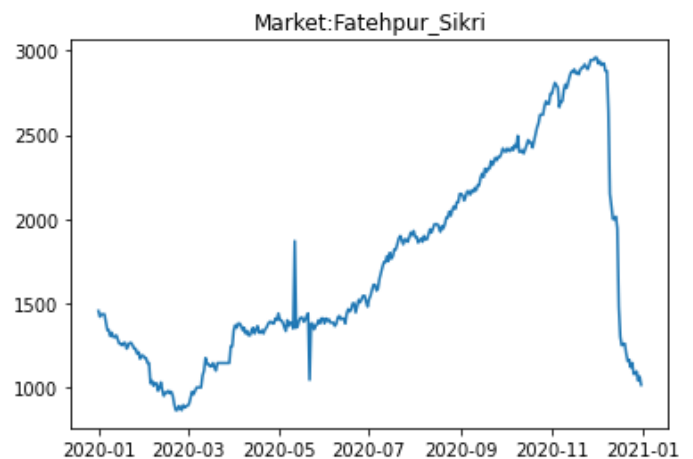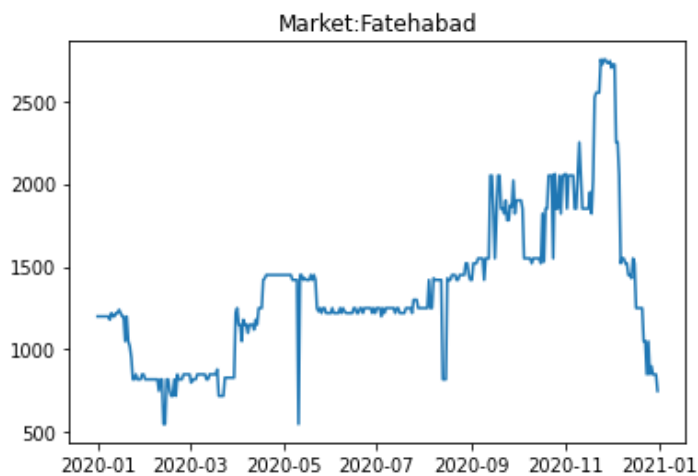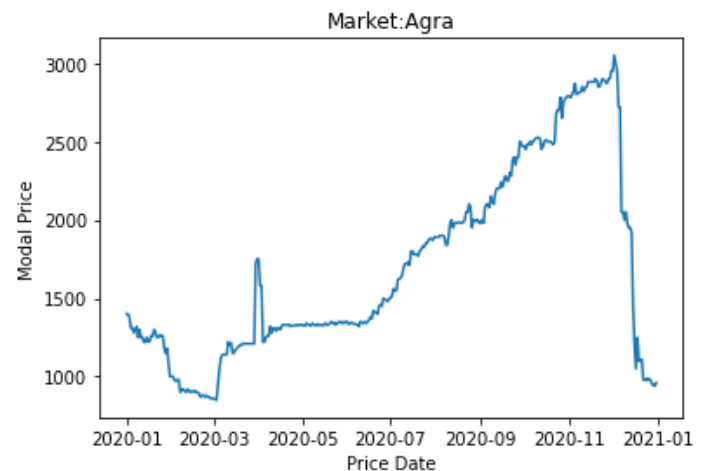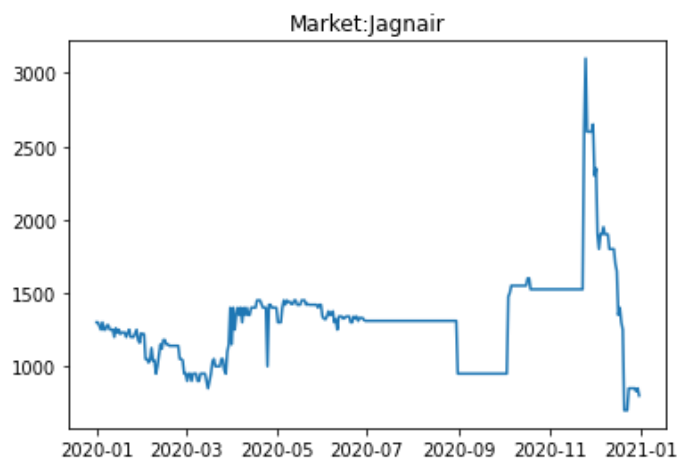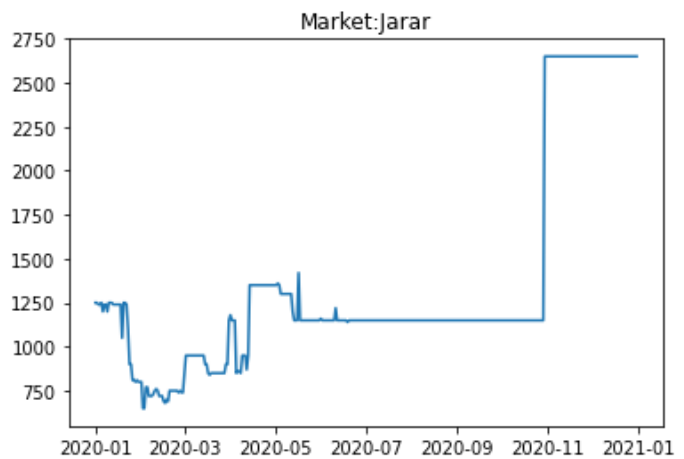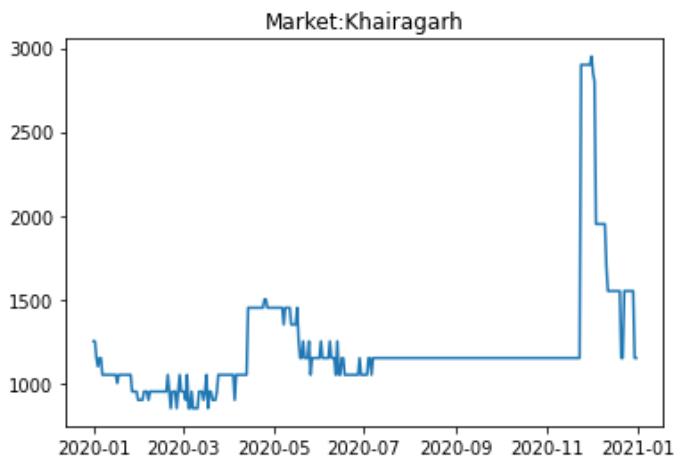
The first three major Markets for Agra as observed from the plots are :
1. Achnera
2. Agra
3. Fatehpur Sikri

Part-C

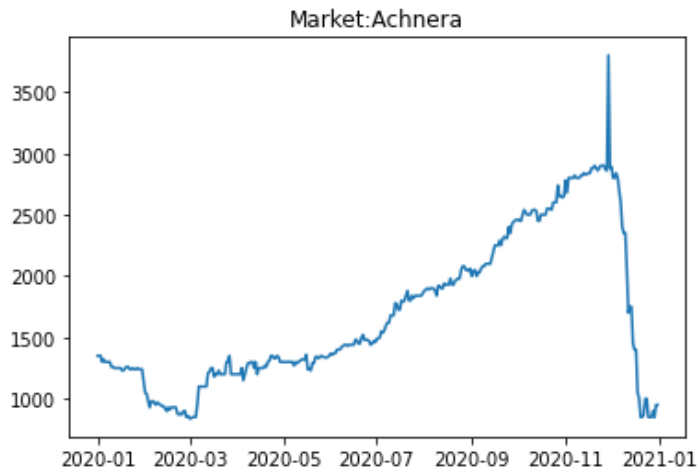i) Data Preprocessing Techniques used are:
1. We group the entire dataset into groups of different Markets.

1. Sorting the data of each market with respect to Price Date keeping the most recent to the last.
2. Since we have missing data on some dates, we fill the missing values such that the prices are equal to that of the previous sample's (i.e. Padding) and if there is no previous sample then give the values of the next sample.(i.e., backward fill). Instead, we can as well use the mean previous and next sample to fill the present sample.
3. Outliers can be removed by using a median filter if needed.

ii) Features of a given sample are Min, Max and Modal Prices of the previous samples upto certain specified time steps.
Or the difference between Min, Max can be taken as one feature and Modal Price as the other and further this data can be normalised for better model training.

Additional features that affect the crop yield like Daily weather and the marketing factors like available regional consumers ,etc can be considered if we have the data for early forecasting of the prices.

iii) By using the previous dates' prices we can forecast the prices for the coming few days, so input is a sequence of previously observed prices and the output will be the predicted prices of the future and this makes it an ML problem. The target variable is a vector containing 'Modal price', 'Min price', 'Max Price' of the current input sample.

iv) I've used LSTM(Long Short Term Memory layer), a modified RNN that performs best in capturing time-series dependent features. 'Relu' Activation function is used after each LSTM layer.

In general various other variants of LSTM like ConvLSTM and Bidirectional LSTM can also be used.

```
Model: "sequential_12"
_____
Layer (type)                 Output Shape              Param #
=================================================================
lstm_24 (LSTM)               (None, 4, 100)            41600

lstm_25 (LSTM)               (None, 100)               80400

reshape_10 (Reshape)         (None, 2, 50)             0

dense_11 (Dense)             (None, 2, 3)              153
=================================================================
Total params: 122,153
Trainable params: 122,153
Non-trainable params: 0
_____
```

v) Loss function used is RMSE(Root Mean Squared Error). The optimizer used is 'Adam(Adaptive Moment Estimation).

Implemented Code flow is as follows:
1. Data is separation based on Markets
2. Filling Data of Missing Dates using methods like Padding and Backwardfill
3. Splitting the Data: Out of 366 days of the year 2020, Train-test-split ratio is 300 days:66 days
4. Model defining,training and optimising.
5. Prediction of test data
6. Visualizing the test data.

Results:

Epochs	400
Optimizer	Adam
Loss fn	RMSE
Input sequence size	4
Output sequence size	2

Prediction of Modal Price one day ahead

Predicted Price Range in Nov and Dec

True Price Range in Nov and Dec