ASSIGNMENT

a. Write a python script to fetch data of prices for the year 2020 (date wise from 1st Jan'2020 to 31st Dec'2020) for district "Agra" of Uttar Pradesh from the data sources mentioned in the data section (can take point b as a reference).

This is the output for this question,

```
Data of prices for the year 2020 (date wise from 1st Jan'2020 to 31st Dec'2020) for district "Agra" of Uttar Pradesh
```

Out[9]:

	Sl no.	District Name	Market Name	Commodity	Variety	Grade	Min Price (Rs./Quintal)	Max Price (Rs./Quintal)	Modal Price (Rs./Quintal)	Price Date
1826	1827	Agra	Samsabad	Potato	Local	FAQ	1100	1300	1200	2020-01-01
305	306	Agra	Achnera	Potato	Desi	FAQ	1300	1400	1350	2020-01-01
1325	1326	Agra	Jagnair	Potato	Desi	FAQ	1250	1350	1300	2020-01-01
1658	1659	Agra	Khairagarh	Potato	Desi	FAQ	1200	1300	1250	2020-01-01
1138	1139	Agra	Fatehpur Sikri	Potato	Local	FAQ	1400	1520	1455	2020-01-01
...	...	...	...	...	...	...	...	...	...	...
1472	1473	Agra	Khairagarh	Potato	Desi	FAQ	1100	1200	1150	2020-12-31
591	592	Agra	Fatehabad	Potato	Desi	FAQ	700	800	750	2020-12-31
306	307	Agra	Agra	Potato	Desi	FAQ	800	1100	960	2020-12-31
849	850	Agra	Fatehpur Sikri	Potato	Local	FAQ	900	1100	1015	2020-12-31
1139	1140	Agra	Jagnair	Potato	Desi	FAQ	750	850	800	2020-12-31

1827 rows × 10 columns

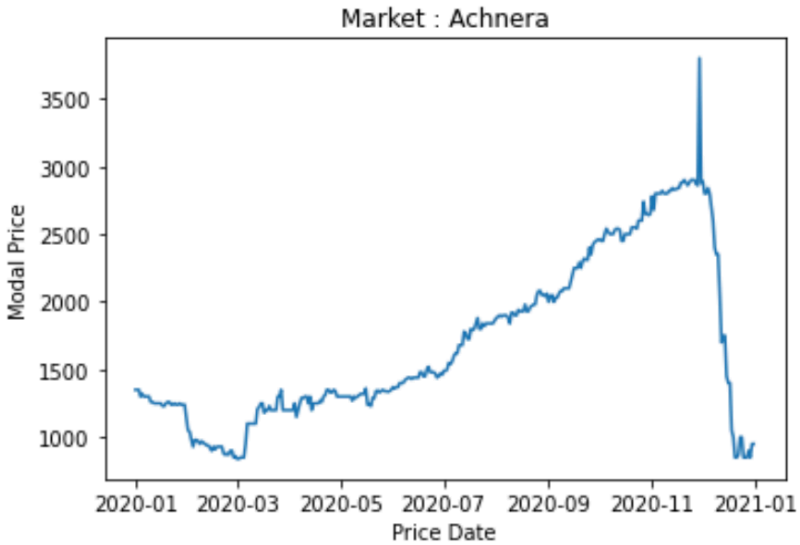
b. Identify major markets for the district "Agra" and plot price patterns for each of them. What patterns do you identify?

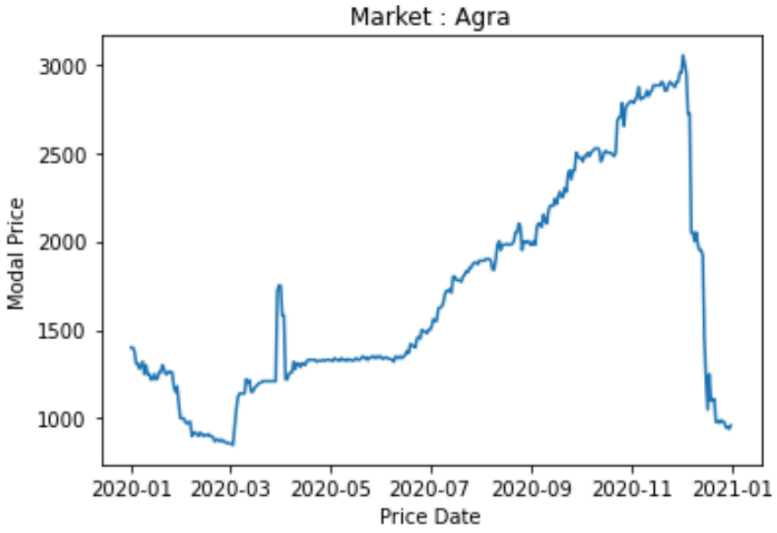
After observing plots, for the Agra district, major markets are,

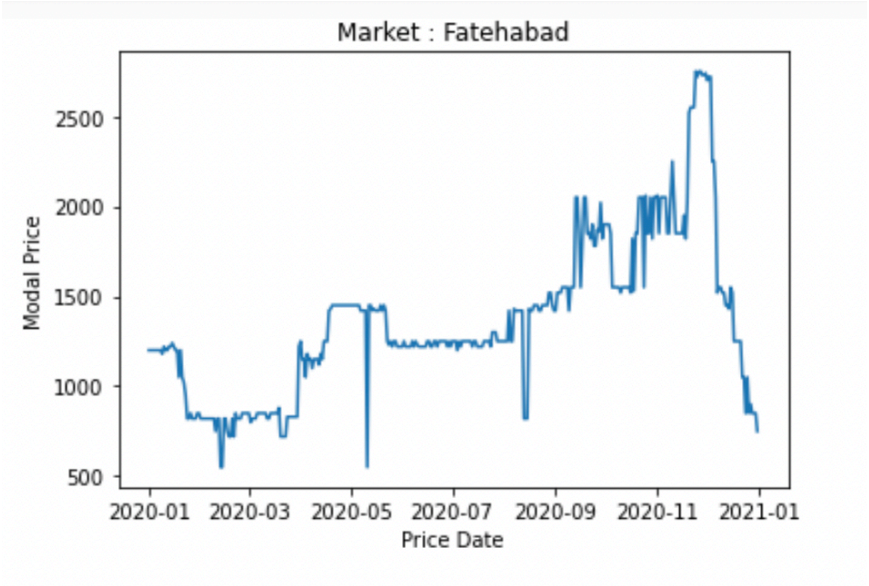
1. Achnera
2. Agra
3. Fatehpur_Sikri

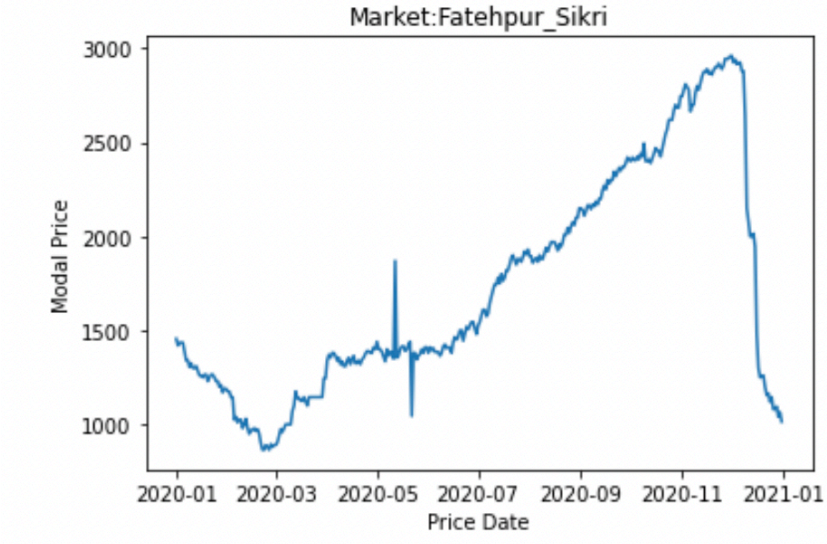
These three markets have high modal price when compared to other.

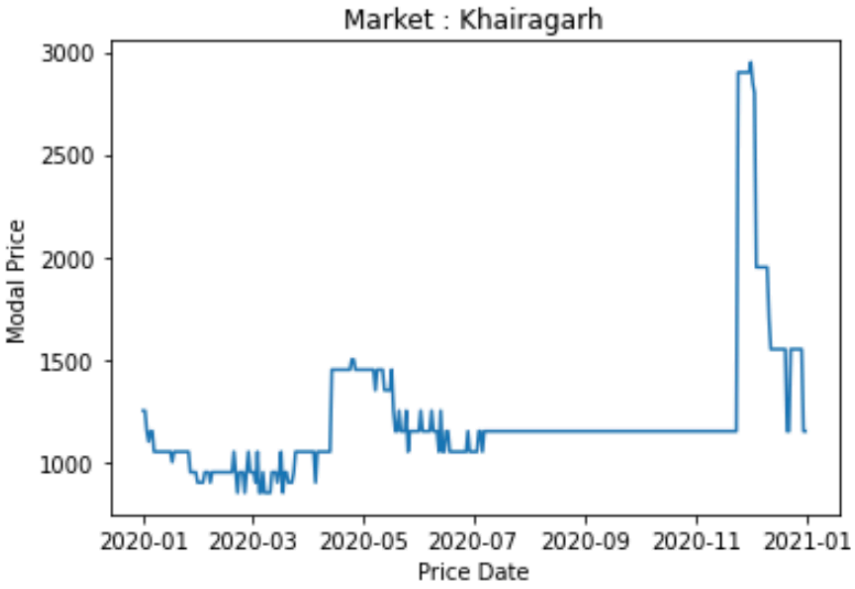
Here are the screen shots of the marketwise price patterns,

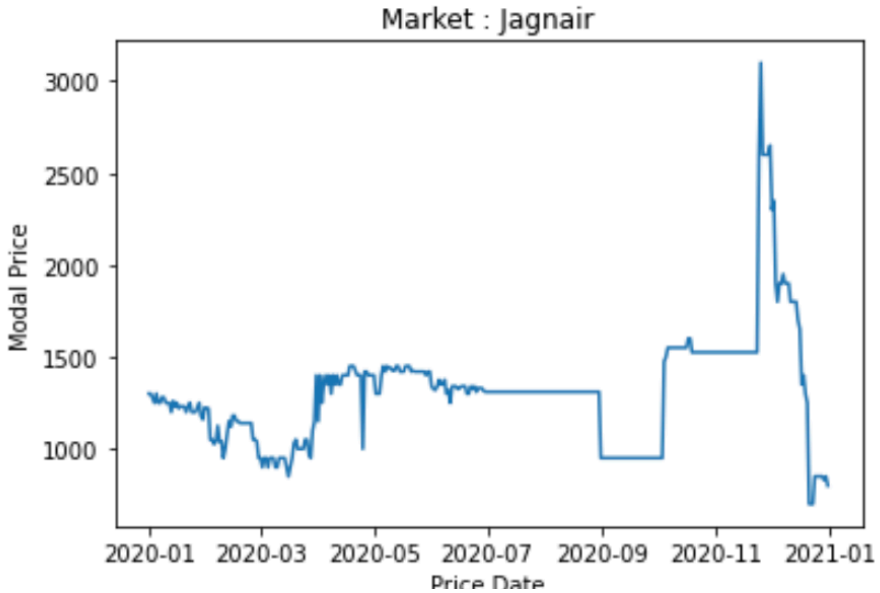


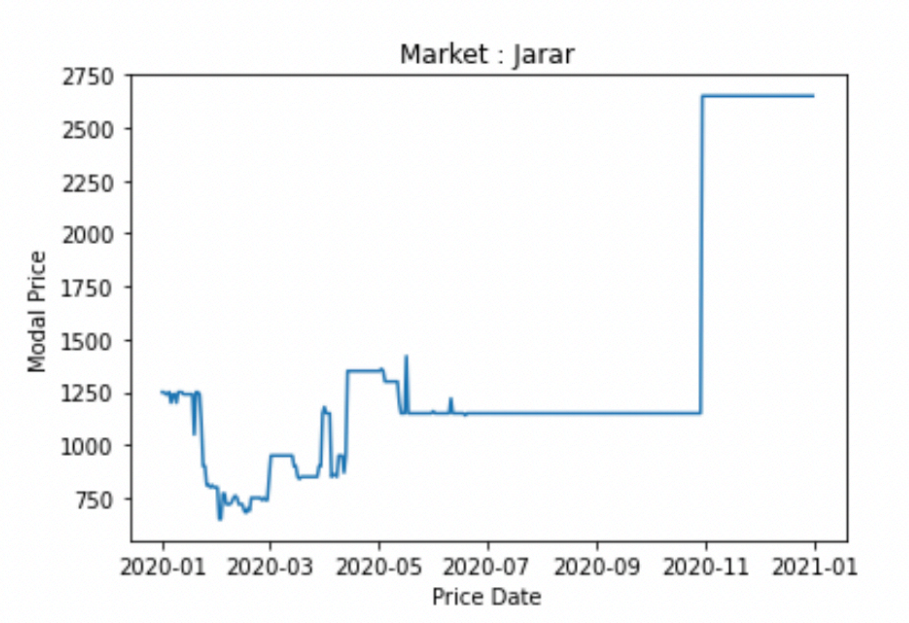

Market : Agra
Modal Price
3000
2500
2000
1500
1000
2020-01
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
Price Date


Market : Fatehabad
Modal Price
2500
2000
1500
1000
500
2020-01
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
Price Date


Market:Fatehpur_Sikri
Modal Price
3000
2500
2000
1500
1000
2020-01
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
Price Date


Market : Khairagarh
Modal Price
3000
2500
2000
1500
1000
2020-01
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
Price Date


Market : Jagnair
Modal Price
3000
2500
2000
1500
1000
2020-01
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
Price Date


Market : Jarar
Modal Price
2750
2500
2250
2000
1750
1500
1250
1000
750
2020-01
2020-03
2020-05
2020-07
2020-09
2020-11
2021-01
Price Date

c. Comment on how you can leverage machine learning to predict prices for a given market in Agra for the crop "Potato".

i. What are the data pre-processing / cleaning techniques you would apply?

- Firstly, data cleaning is more important for getting accurate results during prediction. So, missing values in the dataset should be identify. All the null values should be removed or replaced by the mean, mode values() whatever is suitable, and etc.
- I grouped the dataset into groups of different markets and also dropped the unnecessary columns that are not required.
- Outliers can also be removed using median filter. If there will be a vast difference in the features, then feature scale the dataset using standardization or normalization.

ii. What are the features you would use to create the model?

Selecting an appropriate dependent variables is a key feature and reduces computational cost. When preprocessing and cleaning the data step will be taken then that will give accurate results and also increases the performance. Visualization for each data would be necessary to get an accurate model.
In this project, the features are Max, Min and Modal prices. The difference between min, max can be assign as one feature and model price as other. This data can be used for normalising for better model training.

iii. How would you frame this problem as a machine learning problem? What would be the target variable?

This problem comes under predictive modeling because we are predicting the price for the potato on the basis of previous data in Agra District.
This means that, input will be the past observed prices and output will be the predicted prices of the future.
According to the data given, the target variable should be the Model Price, Min Price, Max price.

iv. Which algorithm would you use for price prediction? v. What would be the loss function you would use?

I would use Long Short Term Memory layer(LSTM), a modified type of RNN that would perform best.

vi. What would be the loss function you would use?

Loss function would be Root Mean Squared Error(RMSE).

vi. Any other comments you want to add?

There are many algorithms and model that we can use but we always need to find the best fit.