**Problem statement**:

**Overview**:

There are over 4,000 agriculture markets (commonly known as mandis) in the country. Everyday prices fluctuate in the markets basis supply and demand of the crop.

Prediction of crop prices is one of the most important task to ensure efficient crop planning and food safety in the country.

The problem statement revolves around prediction of prices for the crop **Potato** in District “Agra” in the state of **Uttar Pradesh** across year 2020.

**Data**:

i. The historical data for prices in district “Agra” of state “Uttar Pradesh” are reported daily on [Agmarknet](https://agmarknet.gov.in/).

ii. Prices for a particular date (say 20 Mar’2021) can be extracted from a URL on Agmarknet:

<https://agmarknet.gov.in/SearchCmmMkt.aspx?Tx_Commodity=24&Tx_State=UP&Tx_District=1&Tx_Market=0&DateFrom=20-Mar-2021&DateTo=20-Mar-2021&Fr_Date=20-Mar-2021&To_Date=20-Mar-2021&Tx_Trend=0&Tx_CommodityHead=Potato&Tx_StateHead=Uttar+Pradesh&Tx_DistrictHead=Agra&Tx_MarketHead=--Select-->

**Description**:

Following are the tasks which need to be done:

1. 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). Following is the output schema expected:

#### https://github.com/paritosh-007/Gramoday

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

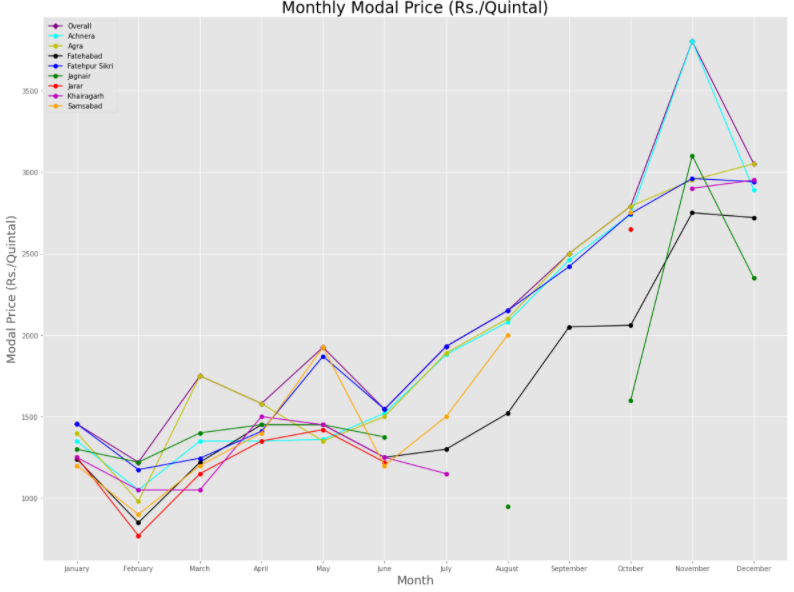


Potato (Solanum tuberosum L.) popularly known as ‘The king of vegetables’, has emerged as fourth most important food crop in India after rice, wheat and maize. Indian vegetable basket is incomplete without Potato.

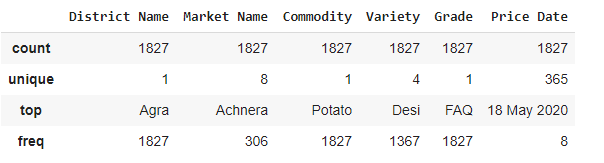
Rabi crops are sown during winter in India, same pattern can be observed for Potato Prices which are demand and supply driven.

Typical crop cycle follows 2.5 to 3.5 Months . It matures in 75-100 days in plains and 140 days in hills.

We can observe the same trend of Price fluctuating in this 2 month interval as UP is a plain region.



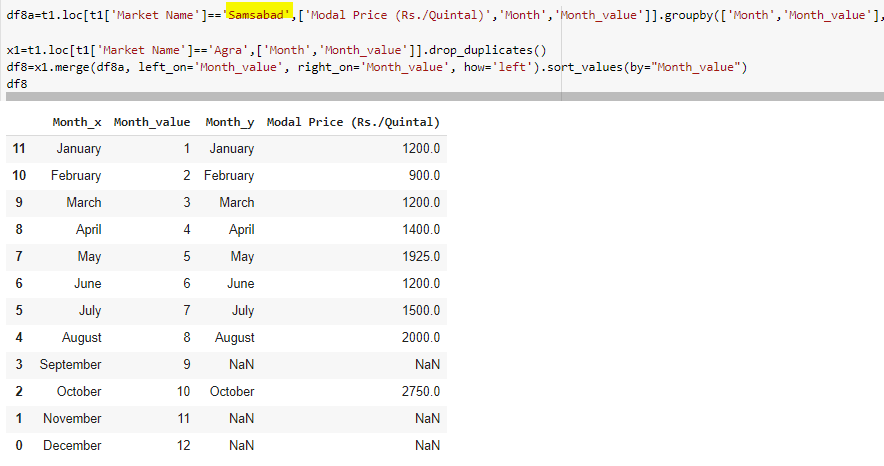
Further analysis can also be done of the Weekdays that observes higher price, but the initial hypothesis is that this factor is minimal in nature.





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

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



Few markets have missing values , for instance in this example

plt.plot(x,z5, color='g',marker='o',label='Jagnair')

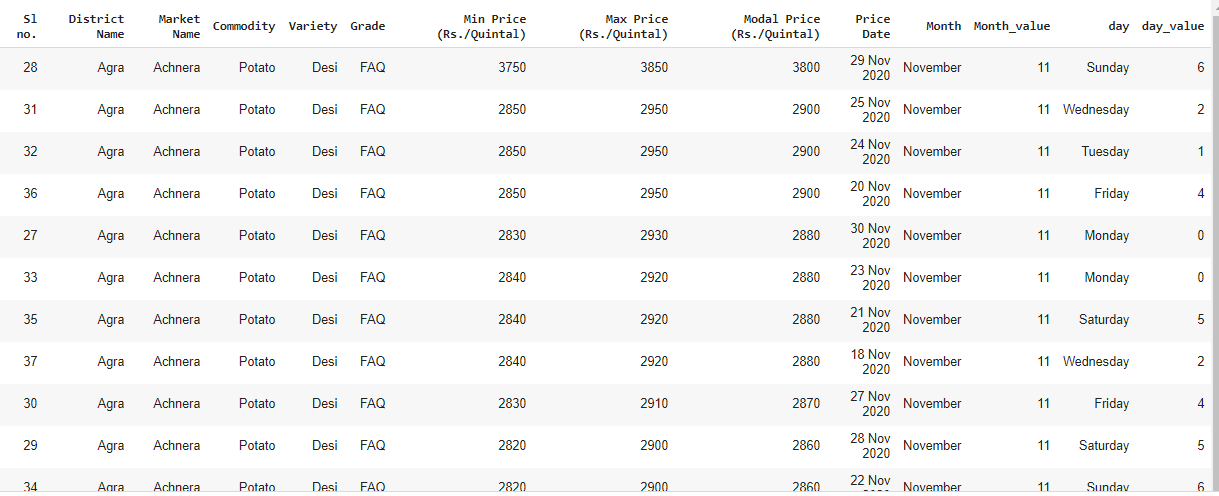
plt.plot(x,z6, color='r',marker='o',label='Jarar')

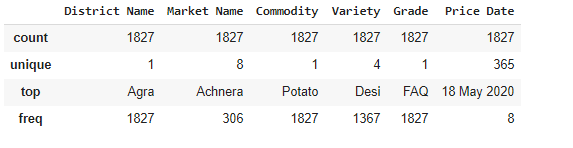
plt.plot(x,z7, color='m',marker='o',label='Khairagarh')

plt.plot(x,z8, color='orange',marker='o',label='Samsabad')

These 4 markets had missing value so we made a left join with Complete Agra District records to make them dimensionally equivalent with others and to be plotable.

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





No point in taking Grade, Commodity, District Name as they are predefined and have just 1 unique value.

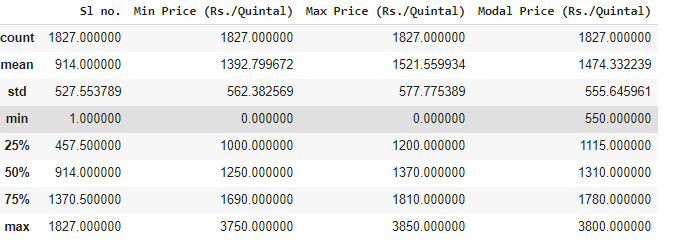
Same goes for SI. no. which is irrelevant here. So , we are left with

**Market Name, Variety, Price date, Price Day, Month, and Prices which is our target variable here.**

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

Price prediction/forecasting is our aim so Prices will be our target variable.

Since we can observe variation in different price parameters, hence for predicting one we can use the other two . We might also fall into the problem of overfitting, so hyperparameters tuning is the key here.

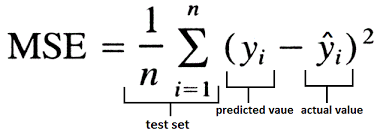


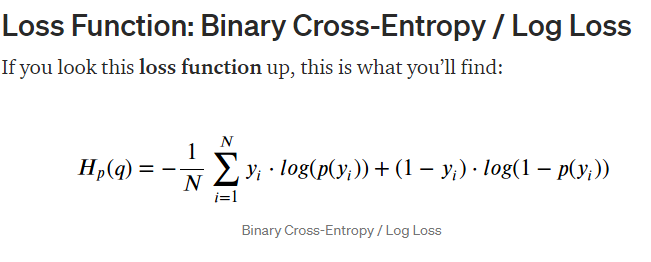
1. Which algorithm would you use for price prediction?

We can implement a mix of machine learning algorithms to predict the future Potato prices, starting with simple algorithms like averaging and linear regression, and then move on to advanced techniques like Auto ARIMA and LSTM.

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

For regression :



For LSTM technique : 

1. Any other comments you want to add?

The work done through this project has got immense potential because it can bring subtle nuances or reveal hidden behaviour .

We can leverage the inferences to balance the supply demand curvature and make the competition fair and democratic for the producers.

**Output**:

1. Please share python script to extract prices as mentioned in the description. Provide instructions to run the script (in a README file) and also share dependencies if any.
2. Collate the output of points (b) and (c), into a word document.

Please collate all these files into a github repository and mail the access link to [admin@agrilinks.in](mailto:admin@agrilinks.in).