**AGRILINLKS DATA ASSIGNMENT**

b)

The major markets, according to the data, for the potato sales are Achnera, Agra, Fatehabad, Fatehpur Sikri, Jagnair, Jarar, Khairagarh and Samsabad, for the Agra District in Uttar Pradesh.

As can be seen from the graph, the prices in Markets of Jagnair and Fatehabad are lowest while the Prices in the markets of Agra, Achnera and Fatehpur Sikri continue to be high.

Another pattern that can be seen is the increment in the prices after the 7th month of 2020. Apart from Jagnair, all other cities’ markets have shown a sudden rise in the prices of Potatoes of the July.

c)

i. Firstly, we get rid of the columns/features that aren’t important for our job. Then we normalize the data using the MinMaxScaler available in scikit-learn library. Scaling will increase the accuracy of any model we build.

Then, we split the data into training and testing data, following the general rule of 80:20 ratio. We can either use train\_test\_split() or do it manually.

For training, we divide the data into chunks of size 30-60. This is basically done to predict the price of the new data point. To predict the price of any given point, the machine will go through the past 30/60 intervals and accordingly predict the new output.

ii. The time or date is one of the most prominent features for out model, then we could also use the Markets Names, as the individual markets also show difference in prices also the patterns.

iii. The target variable for the problem will be the price of the commodity potato for the next day, for this we may use the past data of 30-60 days to predict the new quantity. Using the time series, and the market data, we can easily formulate a model.

iv. LSTM (or Long Short-Term Memory), a special case of Recurrent Neural Network can be used here, as we are required to use past data. LSTM is capable of storing relevant data for extended period of time, so that it can be used further in the future for delivering results, in my opinion LSTM would be the most appropriate choice. We could also use simple Regression or Logistic Regression.

v. Loss function will be the difference between the price predicted and the actual price of the commodity ‘Potato’ for a given day and a given market. We can use Mean Squared Error as well.