Prediction of prices for the crop Potato in District "Agra" in the state of Uttar Pradesh across year 2020.

-Sanjay kazi

A. Data Mining:
Data has been extracted from Agmarknet website by automated web scraping using selenium and beautifulsoup python extensions.

B. Identifying the major markets in Agra:
After obtaining data, I applied basic Exploratory Data Analysis [EDA], using describe and summary inbuilt functions. The **Sweetvz** library provides an extensive description of data. I used Sweetvz to make the conclusion for major markets in Agra. The plots of market and its price trends are as bellow.
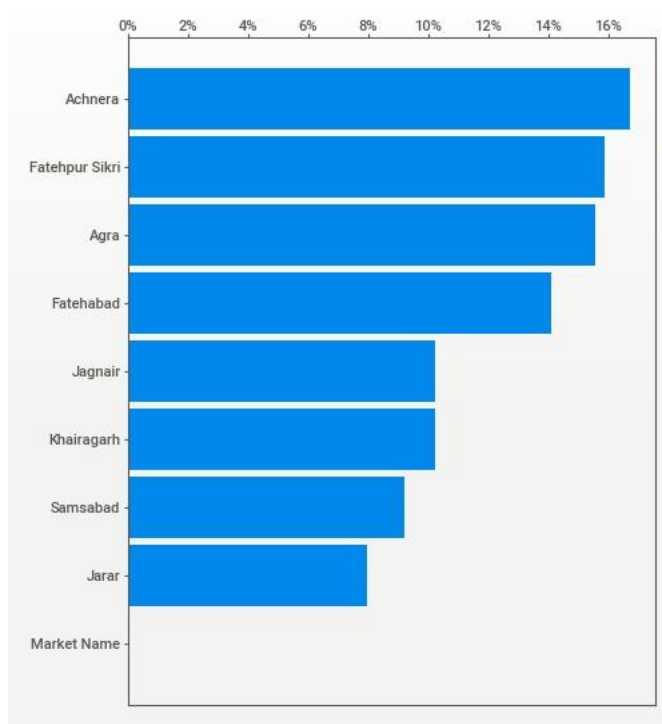
Here it is evident that **Achnera** tops the list; where market price sum is highest.

Fatehpur Sikri, Agra, Fatehabad also shows an above average trend.

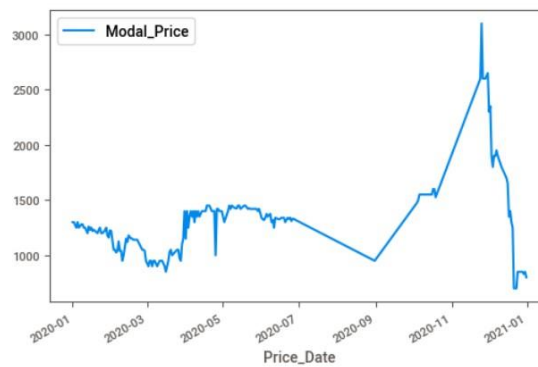The Market vs Price plots are as follows:
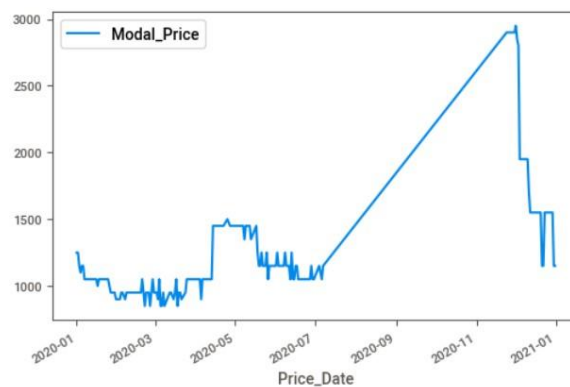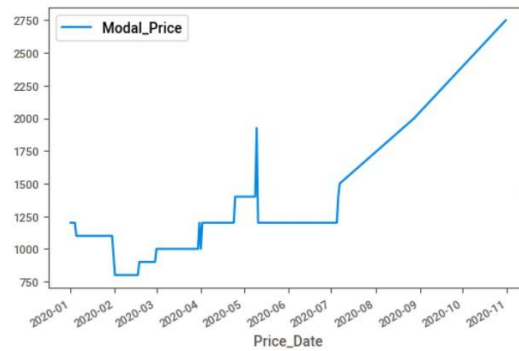
1. Fatehabad

2.Fatehpur Sikri

3. Jagnair

4. Jarar

5. Khairagarh

6. Samsabad

7. Achnera

8. Agra

C. Leveraging machine learning to predict prices for a given market in Agra for the crop "Potato".

    i.) Firstly, I load the data in Jupyter Lab with help of pandas library; the necessary data cleansing techniques I applied are as follows.

        a.) Looked out for any Nan values; luckily there were no Nan value in the extracted data.

        b.) Renamed the dataframe with relevant names for the simplicity's sake.

        c.) Applied reset_index to regularize the indexing after renaming feature names in column.

d.) Dropped the unnecessary features which were not contributing to the enhancement of the model.

e.) This is a time series Analysis problem so I converted the date feature into normal date64 format using datetime library of python.

f.) Finally, data got ready for ML model development.

ii.) Features used to create ML model are below:

a.) Price_date, Modal_Price and Market Name.
This is useful to consider single dependent feature as there are Min. and Max. prices too. Beside that variety could also be taken into consideration.

iii.) Problem framing:
This is a time series analysis problem which includes deep learning RNN techniques to built a robust model. This problem is similar to stock market price prediction. I would pick Modal price as target variable as it is average of the min and max price.

iv.) I have used LSTM (Long Short-term Memory) an artificial RNN architecture. This model is widely used in Time series analysis problem because this technique uses feedback connections.

v.) Mean squared error is used as a loss function; it is the most commonly used loss function for regression. The loss is the mean overseen data of the squared differences between true and predicted values.

vi.) This is really an awesome task and I would love to explore deeper into this.