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.

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:

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). Following is the output schema expected:

S_Number	District Name	Market Name	Commodity	Variety	Grade	Min Price (Rs./Quintal)	Max Price (Rs./Quintal)	Modal Price (Rs./Quintal)	Price Date
1	Agra	Achnera	Potato	Desi	FAQ	600	680	640	20-Mar-21
2	Agra	Agra	Potato	Desi	FAQ	550	750	660	20-Mar-21
3	Agra	Fatehpur Sikri	Potato	Local	FAQ	650	740	690	20-Mar-21
4	Agra	Jagnair	Potato	Desi	FAQ	600	700	650	20-Mar-21
5	Agra	Khairagarh	Potato	Desi	FAQ	600	700	650	20-Mar-21

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

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?
    ii. What are the features you would use to create the model?
    iii. How would you frame this problem as a machine learning problem? What would be the target variable?
    iv. Which algorithm would you use for price prediction?
    v. What would be the loss function you would use?
    vi. Any other comments you want to add?

Output:

i. 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.
ii. Collate the output of points (b) and (c), into a word document.

Please collate all these files into a folder and mail it to admin@agrilinks.in.