**Assignment Questions**

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

There is an interesting pattern in almost all the Markets as in first 2months of 2020 the price daily reduced and from 3rd month it started increasing exponentially. Till 7th month the increase was mild numbers but after 7th it was drastic eventually the price was at peak in the last two months of 2020

This can be observed by the plots in the notebook

The cause of the Price Hike is due to lockdown heavy usage and low transportation reported by some traders and by November cultivation wasn’t matching the supply that’s why it was at peak

Some of the Major Market are Khairagarh, Samsabad, Jarar and Jagnair as this market have low modal price compare to others which makes them more effective to people to use. I have plotted this in the Notebook you can visually see it

**B. 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?**

* I would first check for any missing values and null values if any will treat those by imputation and deletion depends on the variable how much it is important for the model building if there is more than 70% of data is missing then its not of much use so can be deleted other than those can be treated accordingly. As far as this dataset there is no such issues
* For pre-processing I could group the dates(months) into particular seasons as it can give some useful insight for the analysis and model building
* Remaining variables all are good to go no need any processing

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

I would use Min Price, Max Price, Seasons (price dates) as it will have definitely an impact on price. This three are the major other than this we can also use markets and varieties but as seen in analysis variety is not having a much input to analysis

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

Supervised Learning Is an algorithm where we have a bunch of labeled data using those we tarin the data to predict an outcome Here we have same scenario we have labeled dataset we need to predict the price so it’s the best option for as opt for machine learning algorithms to solve this problem

We can make use max min prices to tarin data and then we will predict the average price

**Modal Price (Average Price) will be the Target Variable** as that’s what gives the idea how the price is behaving over the time period

1. **Which algorithm would you use for price prediction?**

I would use simple linear regression to predict the price as its best suited for this problem

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

I would use Mean Squared Error (MSE) as we want the predictions to be as close as possible to the actual data points

1. **Any other comments you want to add?**

I was having no idea on how this agricultural market functions but somehow managed to draw some insights with little research. If there is something wrong feedback is much appreciated