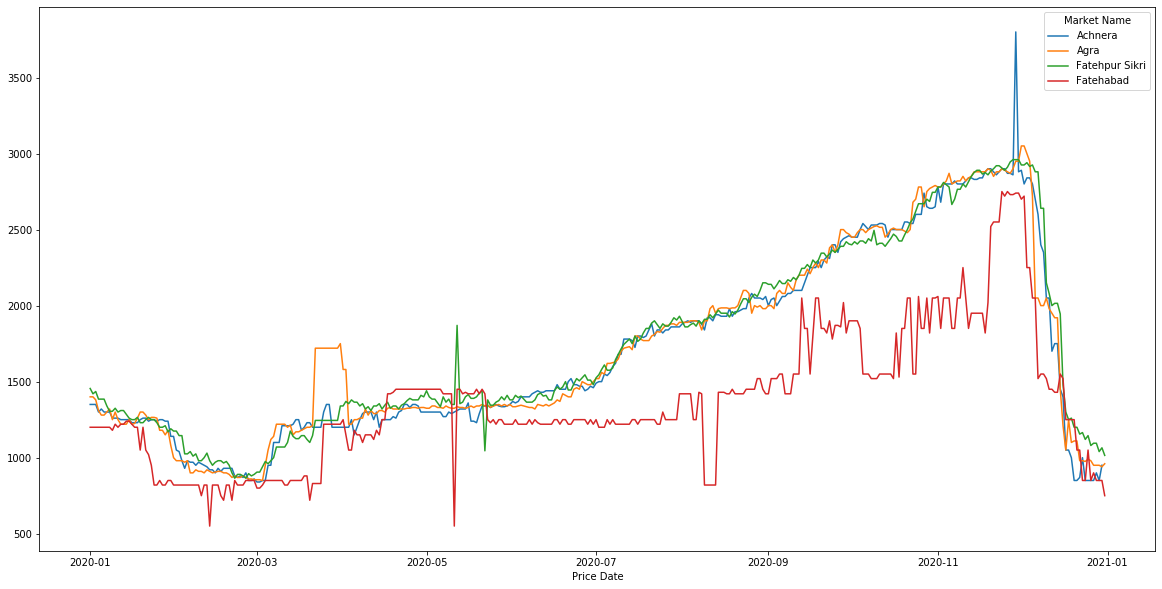
Ques 2: 

Major Markets identified were Achnera, Agra and Fatehpur Sikri because the mean modal price of these markets were the highest as you can see in the below picture.



Also, there were many missing data points in the modal prices, so I have used back propagation method to fill those missing values.

After plotting the prices of these three markets, we can clearly see that the price pattern of potato has seen an increasing trend since March and the prices of potato saw a drastic drop in the month of December.

One of the reason of the increasing trend from March can be the pandemic as pandemic hit in our country almost during march and since potato is a necessity vegetable, it’s prices increased due to high unnatural demand and low supply.

Ques 3:

1. Data Pre-processing techniques:
2. First of all I will look at the correlation between variables and check if one variable is being explained by the another or not because keeping two variables that give same information is of no use.
3. After that I’ll look at the missing values and see different ways to fill them up. If some variables follow a certain pattern then the missing values will be filled up accordingly.
4. Next I’ll look at the outliers using different kinds of plots like boxplot, scatter plot etc and if there are any outliers available then I’ll find out ways to either remove them or replace them.
5. I can create month, quarterly and weekly based features for the model as it will give me more clear picture about the problem. I can also give ranks to various districts according to their high prices, which will further help me to predict the prices.
6. I’ll frame this as a regression model and predict the modal prices of the commodity across various markets.
7. There are various regression models like Linear regression, Lasso regression, Decision tree regression, Random forest regression and many more. But I’ll start with Linear regression.
8. If I take Linear regression as the model then my error function will be my loss function. Error function is the sum of squares of different between actual and predicted values and since the main objective of the loss function is to minimise it, so the error function is our best choice.
9. The data given here is very straight forward, so I’ll need to work more on feature engineering part to get more insights and fit our model with more efficiency.