b) Patterns observed:

* In Achnera Market the Median Price is around Rs. 1500/Quintal for the potatoes and from the boxplot we can see that one day the price was around Rs. 3800/Quintal which looks like an outlier.
* In Fatehpur Sikri Market the Median Price is around Rs. 1500 with no abnormalities seen in Fatehpur Market.
* In Agra Market the Median Price is around Rs. 1500/Quintal for the potatoes with no abnormalities seen in the market
* In Fatehabad Market the median Price is around Rs. 1200/Quintal for the potatoes and from the boxplot we can see that some days the price reached in the interval [2200,3000].
* In Jagnair Market the Median Price is around Rs. 1300/Quintal for the potatoes and from the boxplot we can see that some days the price even reached the interval [2000,3000].
* In khairagarh Market the Median Price is around Rs. 1000/Quintal for the potatoes and from the graph we can again see some outliers.
* In Samsabad Market the Median Price is around Rs. 1250/Quintal for the potatoes and we can see some outliers also from the boxplot.
* In Jarar Market the Median Price is around Rs. 1150/Quintal for the potatoes and we can see an outlier in the boxplot

C)

1) Data Processing steps:

* Market Name and Variety are the Categorical feature which have texts in it and to let model understand the whole of the data we have to first Map the text features to Number for which we can use One hot encoding.
* Although the Min Price, Max Price and Modal Price features show some outliers, we won’t drop them and not change them as it is possible for a particular vegetable to become expensive over a night, example of this situation is the situation which happened recently where the price of onions sky rocketed.
* We can drop District Name, Commodity, Grade, SI no and Price Date before fitting the model.
* We can also standardise the columns to bring them in the similar range.

2) Market Name and Variety (after one hot encoding), Min Price (Rs./Quintal), Max Price (Rs./Quintal), Modal Price (Rs./Quintal)

3). XGBoost

4). reg : linear

5). This is a regression problem as we have to predict the price which is a continuous quantity.

6). We could have used the LinearRegressionClassifier to predict the prices but we did not use it because our data has some outliers which we did not remove as they were providing us the valuable information that the prices of the potato could sky rocket over the night, We also did not have any missing values present in our data and hence there was no imputation done.