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

I have observed that most of the Market name that is Agra and achnera are having highest count values. Thus all other are lesser than that. The Modal Price are arranged between 500 to 1000 for the most of the variety of market. The most selled variety is Desi rather than the local. The max price and modal price has been same for mean years. But all the prices such as modal, max and min decreased in later years as compared to the starting i.e. 2017.The most market is stated in samsabad for the year 2020(from 1st Jan 2020 to 31st Dec 2020) as observed and mostly the variety was desi, then local. Potato and other didn’t included much.

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 Clean the dataset first to get accurate prediction, Identifying and handling the missing values in the dataset, identify and removing the outliers encoding all the categorical data, removing null values or replacing it, segregating the dataset into training and test data, splitting the data, if theirs a vast different in the features will feature scale the dataset using normalization or standardization.

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

Selecting a appropriate dependent variables is a key feature and reduces computational cost. When I would preprocess and clean the data that would led to accurate results and increase the performance. Visualization for each data would be necessary to get a accurate model.

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

I would frame this problem under predictive modelling, as we are predicting the data/price for the potato on the basis of past data in Agra. The target variable according to the given data in assignment I would select the modal price and places where the potato is been trending (Maximum trade) and the max and min prices comparison.

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

I would use Navie bayes algorithm to predict the price and also K means clustering.

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

Mean squared Error or Confusion matrix

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

Prediction of crop depends on various aspects and there are many algorithms and model that we can use but we always need to find the best fit.