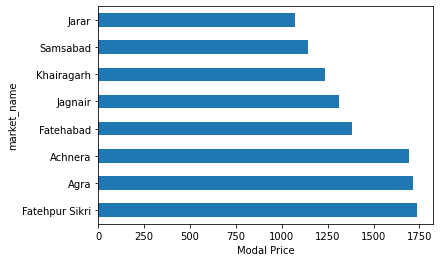
**Mukund Sojitra**

**Data Science Problem Statement Assignment**

**ANSWER B.**

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

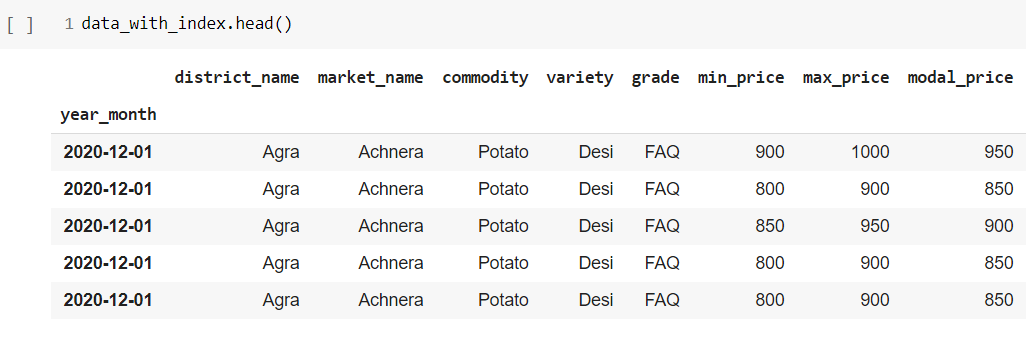
Ans. I have identified top 3 major markets which name is **Fatehpur Sikri, Agra and, Achnera**.

These top 3 markets have Modal Price values greater than average value **1411.7686** of the total market. 

**ANSWER C.**

1. What are the data pre-processing / cleaning techniques you would apply?

Ans. Here in data pre-processing, I have converted the date column in data time object and set that date column in index, which helps in time series analysis.



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

Ans. For predicting Modal Price, I have taken Market name, Commodity, Min Price, Max Price columns as features.

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

Ans. Here in this dataset, machine learning problem can be viewed as to predict the Modal Price for given Market name, Commodity, Min Price, Max Price.

4. Which algorithm would you use for price prediction?

Ans. I have use Support Vector Machine and accuracy of train data is 90.0555% accuracy of test data is 92.2152%.

5. What would be the loss function you would use?

Ans. I have use Mean Absolute Error and Mean Squared Error.

6. Any other comments you want to add?

Ans. Here I have applied Normality test on column of ‘**modal\_price**’.

So, I conclude that the modal\_price doesn’t follow normal distribution.

