Connected new commodity prices to current SQL-database (crop-project2) (in the broken prices\_data.ipynb

Models Walk Through:

We first attempted to use a Long Short-Term Memory (LSTM) neural network for our price predictions. We had several issues with attempting to use this model, including constant package installation error. After fixing our package installation issue, we did further analysis the type of data that was being used to run this model. The data that would have been pulled to create the predictions was the Date column, which would have not been a helpful correlation for prices. We decided that this model set up would not give us helpful price predictions.

Our next step was to find the correct data to use for correlation. We use the panda’s .shift function in our Jupyter notebook to find the previous day price, as current and future prices would be more related to previous prices than dates. We used the previous price data in two different ways; comparing the previous crop price to the current crop price and the previous corn price on the current live cattle price. We thought that the corn price would influence the cattle price, since majority of the corn is used as livestock feed. However, in our initial analysis, the correlation between the two was poor. There was a strong correlation between previous crop prices and current crop prices. Since the previous price correlated so well, we wanted to see if there was trend in prices buy weekday and/or monthly. Using pandas we calculated the average prices by day and month and found there was no daily or monthly seasonally.

Our next steps were to use the previous prices and use the linear regression model. We used the model to predict prices and graphed the actual prices vs the predicted prices. When then graphed this data in a scatter plot. In the plots you can see the strong correlations between the actual prices and the predicted prices. We also found the ran the current days prices against a price lag of 5, 10, and 20 days.

Our final model was a Logistic Regression. We took the previous day price and compared it to the current day price to see if prices went up or down. We then converted the Boolean data into integers, with 1 being up and 0 being down and put the data in a heat map. With an accuracy score of 0.50 and less, this tells us you cannot use the current day’s price to predict if tomorrow’s price would go up or down.

Corn Charts-

There is a strong correlation between the previous day price and the current’s day price.

When you try to predict the current price based on price from 5,10 or 20 days ago, there is a good correlation, but not as strong as the previous day comparison, since the r-squared value does slightly drop (0.996 to .0978). Above the $500 mark the predicted price and the actual price is not as tightly correlated since the dots are sparser. With an accuracy score of 0.489 you will not be able to predict if tomorrow’s prices will be higher or lower than today’s prices.

Further Price Analysis: See what prices are consider high and look more in the different of actual prices verses predicted prices.

Soybean Charts-

There is a strong correlation between the previous day price and the current’s day price.

When you try to predict the current price based on price from 5,10 or 20 days ago, there is a good correlation, but not as strong as the previous day comparison, since the r-squared value does slightly drop (0.995 to 0.975). Above the $1100 mark the predicted price and the actual price is not as tightly correlated since the dots are sparser. With an accuracy score of 0.4784 you will not be able to predict if tomorrow’s prices will be higher or lower than today’s prices.

Further Price Analysis: See what prices are consider high and look more in the different of actual prices verses predicted prices. Consider if $1600-$1800 should be classified as outliers.

Wheat Charts-

There is a strong correlation between the previous day price and the current’s day price.

When you try to predict the current price based on price from 5,10 or 20 days ago, there is a good correlation, but not as strong as the previous day comparison, since the r-squared value does slightly drop (0.991 to 0.958). Above the $1100 mark the predicted price and the actual price is not as tightly correlated since the dots are sparser. With an accuracy score of 0.4784 you will not be able to predict if tomorrow’s prices will be higher or lower than today’s prices.

Further Price Analysis: See what prices are consider high and look more in the different of actual prices verses predicted prices.

Next steps:

Correlated soybean prices and see if we can predict future oil and meal prices based on the soybean price or vice versa.