Project 2 Outline

- Determine best source of data:
  - CME DataMine
  - Nasdaq Data Link (Quandl)
  -
- Other resources
  - USDA - National Agricultural Statistics Services
  - USDA Economic Research Service
  - CME Institute
  - Hog Feeding Spreads
  - Trading Opportunities in Lean Hogs
  - Niche Pork Production
  - Soybean Crush Synthetic Futures
- Consider other variables
  - Weather/rainfall
  - Oil/natural gas prices
  - Corn prices rising because of demand abroad for ethanol
  - ???
- What are our goals
  - Write a brief summary of our interests and intent
  - Maximize for profit
  - Farmers are looking to hedge
  - Farmers need to determine quantity to purchase
- What environment do we want to work in?
  - Jupyter, Google Colab, or AWS SageMaker
- Data load in & clean up
  - HE - lean hogs
  - ZC - corn
  - ZM - soybean meal
  - LE - live cattle
  - GF - feeder cattle
  - Need to look at continuous contracts
  - All have different expirations
  - Visual data historical data to show our starting point
- Time-series analysis
  - Filter out noise
  - Forecast returns using ARMA - output & plot
  - Forecast settle price using ARIMA - output & plot
  - Forecast volatility using GARCH - output & plot

- Classification
    - Split data into train and test
    - Scale the training and testing data
    - Resample the data with the balanced random forest classifier
        - Display confusion matrix and print reports
    - Train the classifier and calculate the balanced accuracy score
        - Display confusion matrix and print reports

- Sentiment analysis
    - Research best sources for in-depth industry knowledge - Twitter?
- LSTM Stock Predictor

Technical Requirements

- Create our slideshow presentation
    - PowerPoint, Keynote or Google Slides
- Summarize our conclusions and predictions
- **Optional**, apply a dimensionality reduction technique to reduce the input features, or perform feature engineering to generate new features to train the model
- Create one or more machine learning models.
- Fit the model(s) to the training data.
- Evaluate the trained model(s) using testing data. Include any calculations, metrics, or visualizations needed to evaluate the performance
- Show the predictions using a sample of new data. Compare the predictions if more than one model is used.
- Save PNG images of your visualizations to distribute to the class and instructional team for inclusion in our presentation and our repo's README.md
- Use one new machine learning library, machine learning model, or evaluation metric that hasn't been covered in class
- Create a README.md in your repo with a write-up summarizing your project. Be sure to include any usage instructions to set up and use the model