# MinneAnalytics

Student Data Set for MinneMUDAC

# Backstory

Growing the food that feeds the expanding world population is the job of farmers. The one position in the world where every day you can say you are solving world hunger. 98% of farmers are family owned operations, and small businesses. The most challenging part of any small business is the number of skill sets needed to grow and maintain a profitable business.

A farmer's core competency is often in growing the commodities he/she sells. Doing so profitability requires a farmer to decide at what price they can sell at in order to break even.

Farms are a very cash flow intensive business. For those that are not familiar with agriculture some key things to understand this small business are:

- Avg price of an acre of land in MN is >$10k per acre
- Avg price of combine (new) to harvest commodity crops > $500k
- Avg cost of a planter is >$250k
- Avg cost of other necessary equipment >$400k
- Avg cost of seed for an acre of soybeans is $60-$80
- Avg cost of agronomic applications (fertilizer, spray etc.) for an acre of soybeans is >$100/acre

# Novice Challenge

- Create a presentation including visualizations and insights using the provided dataset of predictive features
  - Perform descriptive statistics on the provided dataset to explore the variables provided
  - Eliminate variables that are not correlated as needed
  - Perform feature engineering as necessary

# Novice Challenge: Analytic Thought Starters

- What trends do you see in U.S. soybean exports?
- What seasonality do you see in the data?
- Do you see any relationships between commodity exports (substitutable products: corn/soybeans)?
- Are we trade neutral for agricultural products in a given month/year (consider columns past BE)?
- Does competitor exports of soybeans (column AK) impact US exports of soybeans?

Farm Femmes

# Novice Challenge: Evaluation

Insights will be measured against each deliverable distinctly:

1. Strength of evidence provided in suggesting key relationships or trends to monitor
2. Applicability in guiding decision making for sell recommendations
3. Creativity in informing a farmer's understanding of the market trends

# Undergrad Challenge

In this challenge, you will build a predictive model that will return 3 key insights to the farmer:

1. Predicted commodity closing (EOD) price 5 consecutive trading days in September & October (10 total days). Dates will not be disclosed prior to November 9.
   a. *Target variable Contract dates (March, May, July)*
   b. *Total measurements will be 30 (10 days x 3 contract dates)*
2. Is today's price a peak / sell recommendation (binary)
3. Predicted commodity EOD price for December 6th
   a. Most accurate prediction will be announced at FASTCon (Dec 9th)

Model development will occur against July & August timeframes

Predictive accuracy will be measured against September and October actuals

Farm Femmes

# Undergraduate Challenge: Analytic Thought Starters

- What is the price difference by contract date?
- Is there seasonality in the target variable for a given contract date?
- Is canola price predictive of soybean price?
- Is there an effect of the delayed 2019 seeding on soybean prices (see time lag in datasets on Oil Crops Outlook tab: oil crops chart gallery figure 1)?

# Undergraduate: Evaluation

Accuracy will be measured against each deliverable distinctly:

1. Predicted commodity closing (EOD) price 5 consecutive trading days in September & October (dates will not be disclosed prior to November 9 - 10 total days)
   a. *You will provide a file with predicted EOD price for September 3-October 31 it will include:*
      i. *Predicted price for each of the distinct contract dates (March, May, July)*
   b. *Total measurements will be 30 (10 days x 3 contract dates)*
2. Is today's price a peak / sell recommendation (binary)
   a. *You will provide a file with SELL or HOLD as your recommendation for September 3-October 31 it will include:*
      i. *A recommendation of SELL or HOLD for each of the distinct contract dates (March, May, July)*
   b. *Total measurements will be 30 (10 days x 3 contract dates)*
      i. SELL recommendation w/in $0.16 of the max price = 10 pts
      ii. HOLD recommendation w/in $0.16 of the min price = 10 pts
      iii. HOLD recommendation w/in $0.16 of the max price = -5 pts
      iv. SELL recommendation w/in $0.32 of the min price = -5 pts
3. Predicted commodity EOD price on December 6th for each of the three contract dates
   a. Most accurate prediction will be announced at FASTCon (Dec 9th)

# Graduate Challenge

Create a predictive model to address the 3 questions outlined in the undergraduate challenge

Differentiate yourselves by considering how the below information could be utilized to increase your predictive accuracy

- Incorporation of additional open source datasets as predictive features
    - Macro-economic variables
        - Weather conditions in key competitive markets for soybeans (examples below are from Brazil)
            - https://weather-and-climate.com/average-monthly-Rainfall-Temperature-Sunshine,Cuiaba,Brazil
            - https://www.weather-atlas.com/en/brazil/cuiaba-climate
    - Geo-political events
        - Tweets
        - Tariffs
            - https://www.bloomberg.com/news/articles/2019-08-23/soybeans-sink-to-two-week-low-as-china-hikes-u-s-tariffs
            - https://www.rt.com/business/468096-us-china-trade-war-soybeans/
        - Brazilian fires and possible expanded area available for crop production
            - https://www.reuters.com/article/us-usa-trade-china-soybeans/on-the-front-lines-trade-war-sinks-north-dakota-soybean-farmers-idUSKCN1VC0ZX
    - Feature engineering of additional predictive variables

# Graduate Challenge: Analytic Thought Starters

● Is there a lag between tweets related to agriculture and/or trade and the markets?
● What are the most impactful indicators to watch from the next top three soybean producing countries (Brazil, Argentina and China)?
● What other data sources could you use as macroeconomic indicators, and what is their relative importance?

# Graduate: Evaluation

Accuracy will be measured against each deliverable distinctly:

1. Predicted commodity closing (EOD) price 5 consecutive trading days in September & October (dates will not be disclosed prior to November 9 - 10 total days)
   a. *You will provide a file with predicted EOD price for September 3-October 31 it will include:*
      i. *Predicted price for each of the distinct contract dates (March, May, July)*
   b. *Total measurements will be 30 (10 days x 3 contract dates)*
2. Is today's price a peak / sell recommendation (binary)
   a. *You will provide a file with SELL or HOLD as your recommendation for September 3-October 31 it will include:*
      i. *A recommendation of SELL or HOLD for each of the distinct contract dates (March, May, July)*
   b. *Total measurements will be 30 (10 days x 3 contract dates)*
      i. SELL recommendation w/in $0.16 of the max price = 10 pts
      ii. HOLD recommendation w/in $0.16 of the min price = 10 pts
      iii. HOLD recommendation w/in $0.16 of the max price = -5 pts
      iv. SELL recommendation w/in $0.32 of the min price = -5 pts
3. Predicted commodity EOD price on December 6th for each of the three contract dates
   a. Most accurate prediction will be announced at FASTCon (Dec 9th)
4. Incorporation of additional data sources
5. Communication of results in a farmer friendly format