Farming and Market Insights 2025

Joshua Wacks

The High Level Requirement

Bring innovative and fresh insights on the farming and market data, in the context of the following questions:

1. How can farming practices be optimised while promoting sustainability and reducing environmental impact?
2. How can agriculture be transformed into a resource-efficient and profitable endeavour?
3. What trends within the data can be identified with reasoning and solutioning where possible?
4. Design a separate out of the box question and answer based off the data

The Farming Data

	Farm_ID	Soil_pH	Soil_Moisture	Temperature_C	Rainfall_mm	Crop_Type	Fertilizer_Usage_kg	Pesticide_Usage_kg	Crop_Yield_ton	Sustainability_Score
0	1	7.073643	49.145359	26.668157	227.890912	Wheat	131.692844	2.958215	1.576920	51.913649
1	2	6.236931	21.496115	29.325342	244.017493	Soybean	136.370492	19.204770	3.824686	47.159077
2	3	5.922335	19.469042	17.666414	141.110521	Corn	99.725210	11.041066	1.133198	50.148418
3	4	6.845120	27.974234	17.188722	156.785663	Wheat	194.832396	8.806271	8.870540	89.764557
4	5	6.934171	33.637679	23.603899	77.859362	Corn	57.271267	3.747553	8.779317	51.033941

Very Clean Data overall

- No missing values
- No outliers
- No duplicate records
- Considered changing the ‘Crop_Yield_ton’ column to ‘Crop_Yield_Kg’ to have standard units for this analysis, but was not necessary.
- Only one categorical variable that needed to be handled: ‘Crop_Type’

The Market Data

	Market_ID	Product	Market_Price_per_ton	Demand_Index	Supply_Index	Competitor_Price_per_ton	Economic_Indicator	Weather_Impact_Score	Seasonal_Factor	Consumer_Trend_Index
0	1	Rice	180.251212	196.085900	199.509124	300.549219	1.093636	28.474810	Medium	148.472131
1	2	Rice	420.527970	188.452400	150.789483	492.097798	0.526307	70.978063	High	97.298888
2	3	Wheat	457.260398	171.179384	78.989326	323.003342	1.292393	80.853592	Low	131.113236
3	4	Soybean	237.179113	196.970677	50.464363	232.978384	0.627663	60.676069	Low	95.169417
4	5	Wheat	324.032925	113.165416	145.878647	312.428652	1.491255	45.379516	Low	130.297138

Very Clean Data overall

- No missing values
- No outliers
- No duplicate records
- Considered changing the price per ton to price per kg to have standard units for this analysis, but was not necessary.
- Two categorical variable that needed to be handled: ‘Product’ and ‘Seasonal_Factor’

Question 1

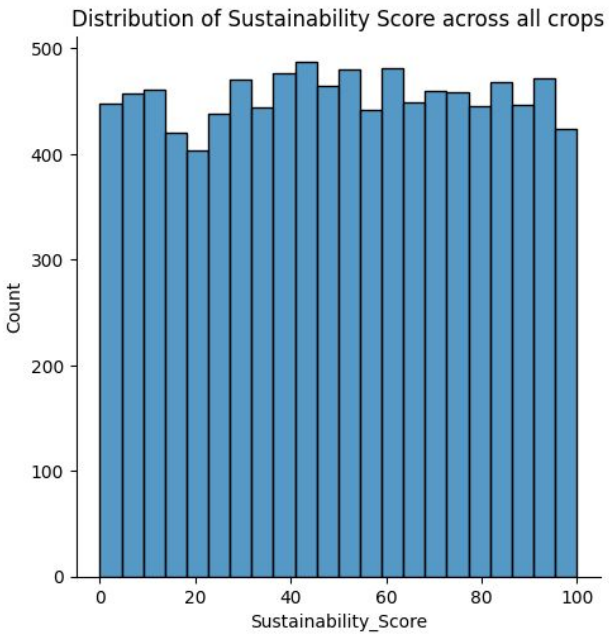
How can farming practices be optimised while promoting sustainability and reducing environmental impact?

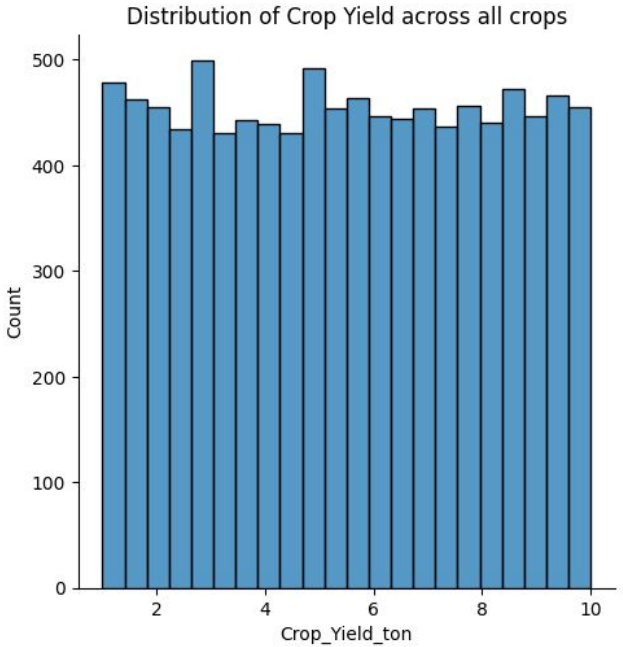
Question 1

Key Insights

- No linear relationships between any of the farming techniques and crop yield.
- No crop significantly yields more than the other crops.
- No linear relationships between any one of the farming techniques and sustainable farming.
- No crop is significantly more sustainable than the others.
- To improve sustainable farming and increase crop yield, it must be done on a per crop level.

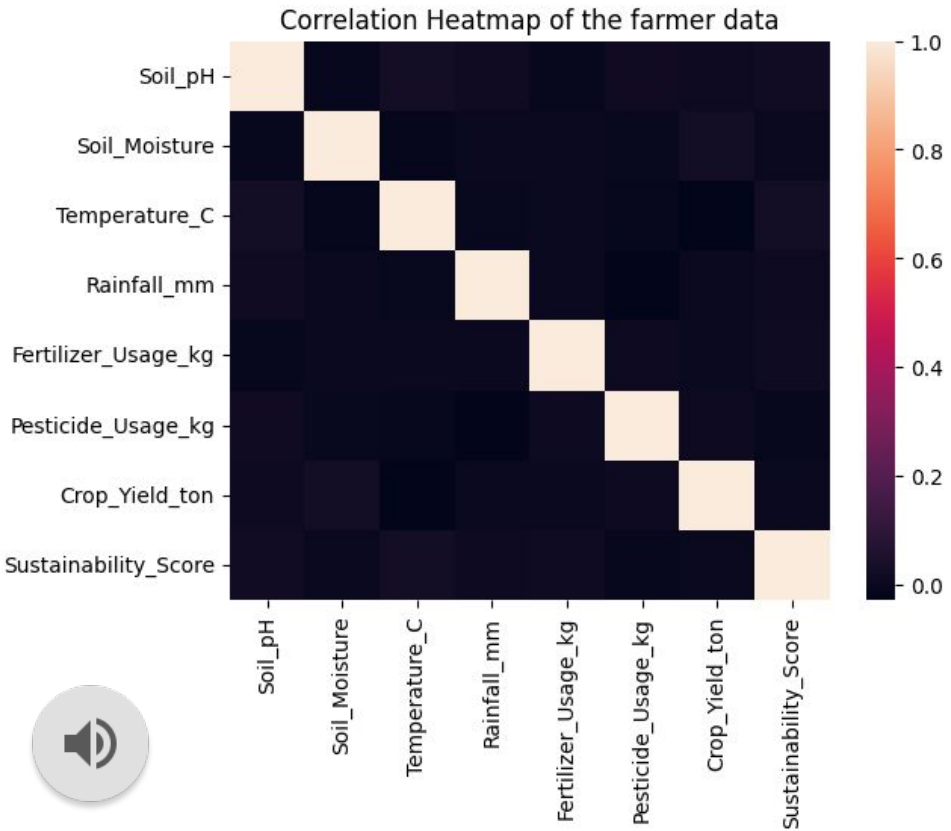
Question 1: Overall Data Analysis





No large skew in the distribution of either the sustainability score or the crop yield.

Question 1: Overall Data Analysis

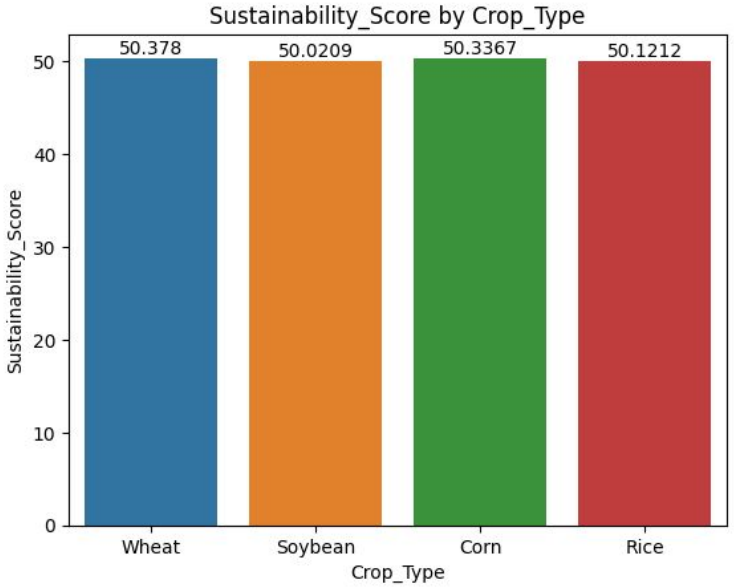


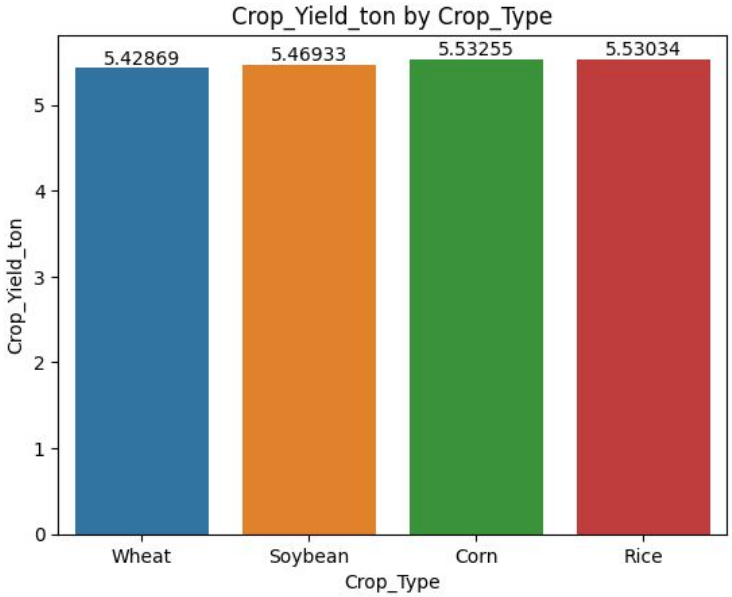
Correlation heatmap clearly shows only weak/ no correlation between the two metrics we are investigating; ‘Sustainability_Score’ and ‘Crop_Yield_ton’.

To gain further insight we need to do a per crop analysis which makes sense as crops generally have different ideal conditions required to grow, such as different levels of pesticide and fertilizer.

Question 1: Per Crop Analysis

Is there a crop that has a statically significant difference in their Sustainability_Score or Crop_Yield_ton?





Using a one-way ANOVA test, with the independent variable being crop type, the differences above were found to be statistically insignificant as the p-value for the test for the Sustainability_Score was 0.966 and the p-value for the Crop_Yield_Ton was 0.424. As both p-values were greater than 0.05 the null hypothesis is not rejected.

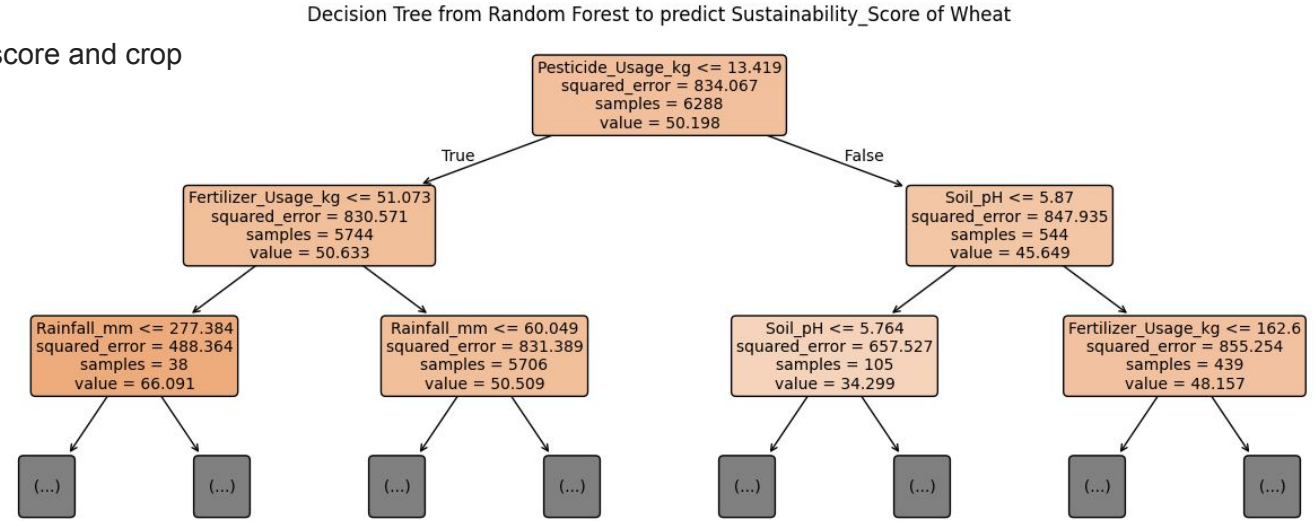
Question 1: Random Forest Analysis

Built a random forest regressor to predict the sustainability score and crop yield on a per crop basis.
Each model achieved at least 80% on each crop type, suggesting it was able to capture non-linear relationships in the data.

Each crop had different features the model found to be most influential in predicting the sustainability score and crop yield.

Features that affect sustainability score and crop yield across all crops:

- Fertilizer Usage
- Rainfall
- Soil pH
- Pesticide Usage
- Soil Moisture



Question 2

How can agriculture be transformed into a resource-efficient and profitable endeavour?

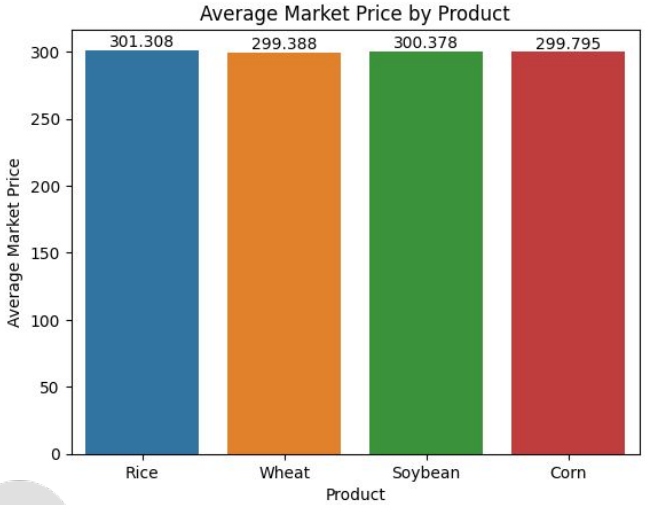
Question 2

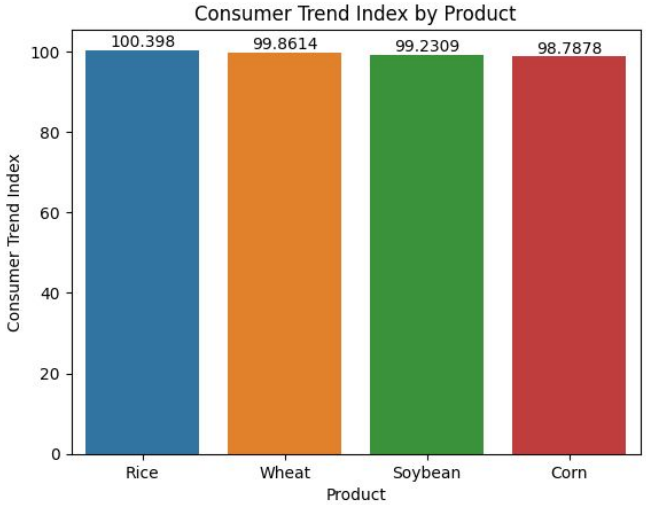
Key Insights

- Use findings from Question 1 to optimise resource usage on farms on a per-crop level.
- Increase rice production as it is the only product where demand is currently less than supply and performs well on all other metrics.

Question 2: Comparing By Crop

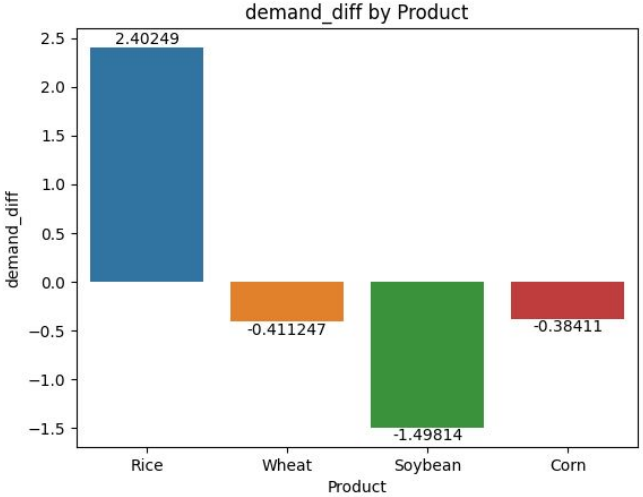
Performing an analysis on each of the factors such as economic indicator and consumer trend index in the data, showed no significant difference in these values across crops at each of the markets.
Furthermore, creating an average price feature([Market_Price_per_ton + Competitor_Price_per_ton]/2) for each of the markers showed no significant difference between the products.

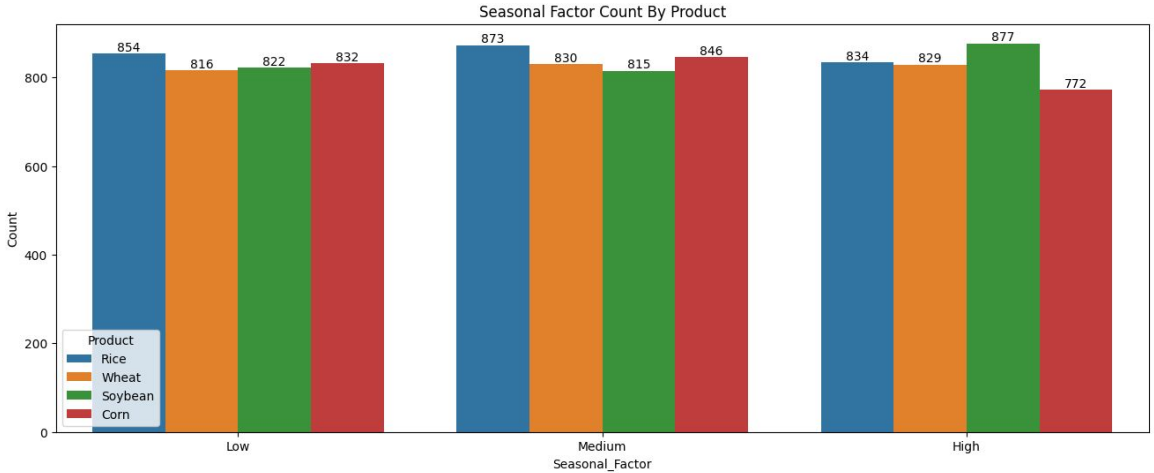




Question 2: Comparing By Crop Cont

Creating a feature which captured the difference between the demand and supply did show there was a product in short supply at the markets, rice which was the only product with a demand greater than supply. As well as this, rice has a low seasonal factor at markets making it viable product to invest in.





Question 3

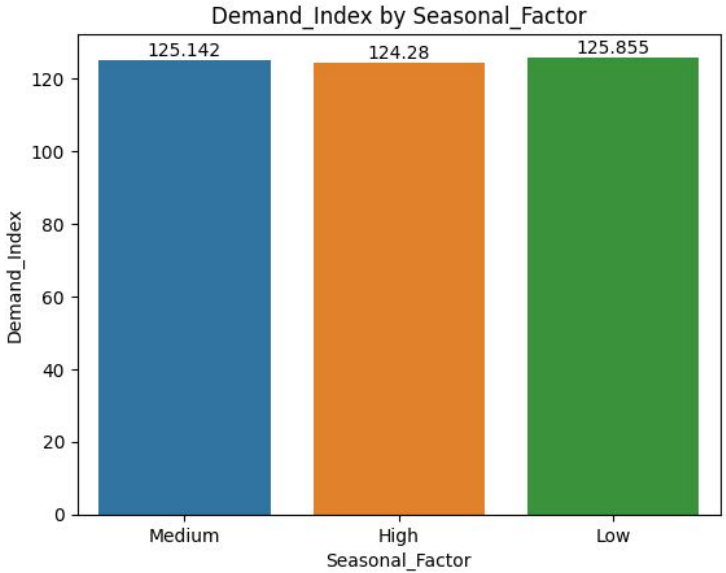
What trends within the data can be identified with reasoning and solutioning where possible?

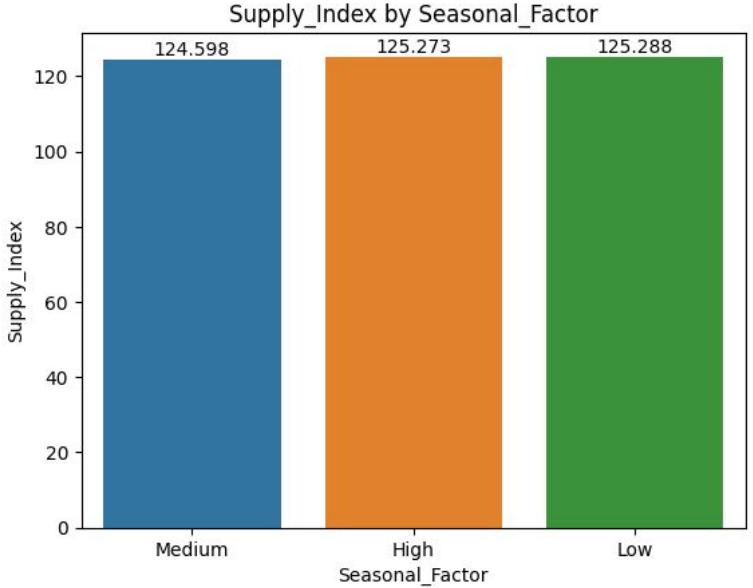
Question 3

Key Insights

- Slight increase in demand when seasonal factor is low vs supply which is not influenced by seasonal factor.
- Specifically difficult to identify trends in non time-scale based data.

Question 3





Slight increase in demand when seasonal factor is low vs supply which is not influenced by seasonal factor.

Question 4

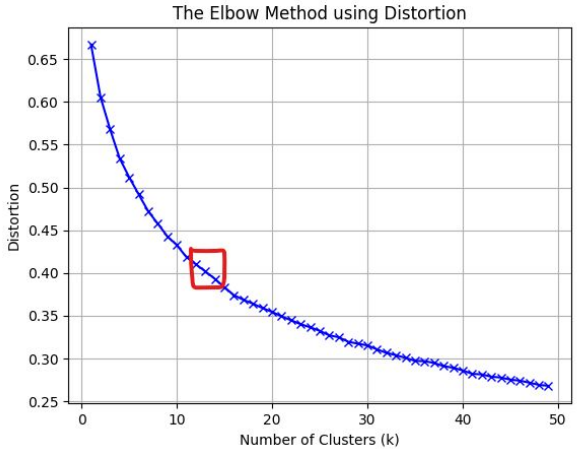
Question Chosen: Can we group farms beyond crop type, on other characteristics, to optimise crop yield and sustainability?

Question 4: Approach

Grouped the farms using an unsupervised machine learning method to find similar farms, separate from the crop type identified earlier.

Two different approaches were investigated:

- The DBSCAN algorithm. Iterated through many different parameter values but was not able to find sufficient clusters.
- K-Means algorithm. Used the ‘elbow method’ on Distortion and Inertia distance metrics and an optimal number of clusters was found to be 15.



Question 4: Findings

Using this clustering we could communicate the information to each of the 15 groups of farms that there are other farms that share similar characteristics in terms of the following features:

- Soil pH
- Soil Moisture
- Temperature
- Rainfall
- Fertilizer Usage
- Pesticide Usage

Allowing the farmers to learn and share information with one another, specifically knowing that these are farmers that grow crops under the same conditions they do.

Farming and Market Insights

Joshua Wacks