Opportunities for Innovations from Data Science.

## Food Waste

optimizes marketing campaign / improving operation

- Predictive analytics – future outcome prediction attempts.
- The use of Big Data – Predicting lifespan of products
- Sentiment Analysis – feelings / enjoyment of customers.
  - Are customers happy with the freshness + quality. Change if they're not.

## Demand of Farmer's.

- Marketing – demographics. Learn more about particular areas, supply to these areas.
- Promotions to direct consumer purchases.

## Environment

- Marketing
- Help build transparency. What's in the food, where it came from, how it was produced.
- 'The Yield' – Monitor planting beds + agricultural ecosystem

Implications

- Data Quality needs to be improved. The degree of which data can be used is largely determined by their quality.

Consumer trends constantly changing

# Predictive Analytics

Predictive models use known results to develop or train a model that can be used to predict values for different or new data.
Modeling provides result in the form of predictions that represent a probability of the target variable based on estimated significance from a set of input variables.

Two Models:

1) Classification Models → Qualitative dat (eg. likely to buy a particular product.)
Regression Models → Predicting a number. (When they are likely to buy this product.)

Modeling Techniques:

1) Decision Trees -
(Classification Models)
- → partition data into subsets based on categories of input variables.
- → Each branch represents a choice between alternatives.
- → The model looks at the data and tries to the one variable that splits the data into logical groups that are the most different.
- → Handles missing values well.

2) Regression - Quantitative Data.
- estimates relationships among variables.
- Intended for continuous variables.
- Finds key patterns in large data sets
- Used to determine how much specific factors influence the movement. Eg price influence sale.

3) Neural Networks -
- Capable of modeling extremely complex relationships
- Powerful + flexible
- Ability to handle non-linear relationships. Common as more data is being collected.
- Often used to confirm findings from regression / decision trees.
- Based on pattern recognition.
- Work well when no mathematical formula is known that relates to inputs / outputs, when prediction is more important than explanation or when there's a lot of training data.

training data → extremely large data set used to ~~train~~ teach a machine learning model.

Other Techniques:
- Bayesian Analysis: "degrees of belief"
- Gradient Boosting: Can be used to improve accuracy of decision tree.
- Incremental Response → Model the change in probability caused by action → the effects of different marketing programs.
- Support Vector Machine → analyse data and recognise patterns.

- Time Series Data Mining → time-stamped + collected over particular time at a particular interval. Combines traditional data mining techniques + forecasting techniques. Sampling, Clustering, + decision trees.