Problem Description: Sustainability of the human race in different parts of the world is challenged by the shortage of food. The world population has grown six hundred percentage - from one billion to about six billion - in the last two hundred years. According to the Population Institute, roughly, 230 thousand more babies are born every day. The World Food Programme estimates that about 795 million people do not have adequate food to lead a healthy life. About 3.1 million children die every year because of poor nutrition. On the other hand, land used for farming has been decreasing which makes the burden of food shortage acute. Regardless, simply attempting to increase the land available for farming is unlikely to sustain the needed food supply. To address this great problem, this project expects you to develop an analytics framework to aid soybean farmers select up to a given number of varieties of soybeans from a large set of available varieties to maximize the yield at a target farm.

Every year soybean farmers make decisions about the varieties to be grown at their farm. While making this decision, they consider uncertainty due to weather, soil conditions, and yield studies of different varieties. They could choose just one variety or a mix of few varieties to hedge against uncertainties. You are expected to utilize the dataset provided to propose a framework which integrates descriptive, predictive, and prescriptive analytics to optimally select up to five varieties of soybeans.

Deliverables:

1. Perform exploratory data analytics to unearth patterns in the given data and utilize those patterns in making predictions and prescriptions.
2. Construct one or more prediction models to predict yield of different experimental varieties.
3. Optimize the portfolio of (experimental) varieties to be grown at the target farm. The optimal portfolio can have at most 5 varieties of soybean. It is not necessary but you are welcome to use the methods you learn in prescriptive analytics class to construct the optimal portfolio.

Data Sets:

1. Training Data for Ag Project
2. Evaluation Dataset for Ag Project

Key:

GrowingSeason	Year	Date
Location	trial location code	Id number
Genetics	breeding group	Group ID
Experiment	Experiment number	Experiment ID
Latitude	Latitude	Decimal degrees
Longitude	Longitude	Decimal degrees
Variety	Variety code	Variety ID
Variety_Yield	Variety yield	Bushels per acre adjusted by moisture
Commercial_Yield	Commercial yield for the trial	Bushels per acre adjusted by moisture
Yield_Difference	yield difference between experiment and commercial varieties in a trial	Bushels per acre adjusted by moisture
Location_Yield	Average site yield (approximately, checks across experiments)	Bushels per acre adjusted by moisture
RelativeMaturity	Relative Maturity Interval	Relative maturity interval (region) based on the location
Weather1	Climate type based on temperature, precipitation and solar radiation	Climate class
Weather2	Season type	Season class
Probability	Probability of growing soybean	Probability of growing soybeans in the nearby area of the site

RelativeMaturity25	Probability of growing soybean of RM 2.5 to 3	Probability of growing soybeans in the nearby area of the site
Prob_IRR	Probability of irrigation	Probability of field irrgation nearby the area of the site
Soil_Type	Soil type based on texture, available water holding capacity, and soil drainage	Soil Class
TEMP_03	Sum of the temperatures for the season 2003	Daily degree Celsius sum between April 1st and October 31st
TEMP_04	Sum of the temperatures for the season 2004	Daily degree Celsius sum between April 1st and October 31st
TEMP_05	Sum of the temperatures for the season 2005	Daily degree Celsius sum between April 1st and October 31st
TEMP_06	Sum of the temperatures for the season 2006	Daily degree Celsius sum between April 1st and October 31st
TEMP_07	Sum of the temperatures for the season 2007	Daily degree Celsius sum between April 1st and October 31st
TEMP_08	Sum of the temperatures for the season 2008	Daily degree Celsius sum between April 1st and October 31st
TEMP_09	Sum of the temperatures for the season 2009	Daily degree Celsius sum between April 1st and October 31st

Median_Temp	Median Sum of temperatures for season between 1994 and 2007	Daily degree Celsius sum between April 1st and October 31st
PREC_03	Sum of the precipitation for the season 2003	Daily degree Celsius sum between April 1st and October 31st
PREC_04	Sum of the precipitation for the season 2004	Precipitation sum between April 1st and October 31st
PREC_05	Sum of the precipitation for the season 2005	Precipitation sum between April 1st and October 31st
PREC_06	Sum of the precipitation for the season 2006	Precipitation sum between April 1st and October 31st
PREC_07	Sum of the precipitation for the season 2007	Precipitation sum between April 1st and October 31st
PREC_08	Sum of the precipitation for the season 2008	Precipitation sum between April 1st and October 31st
PREC_09	Sum of the precipitation for the season 2009	Precipitation sum between April 1st and October 31st
Median_Prec	Median Sum of precipitation for season between 1994 and 2007	Precipitation sum between April 1st and October 31st
RAD_03	Sum of the solar radiation for the season 2003	Daily Watts per sq. meter solar radiation sum between April 1st and October 31st
RAD_04	Sum of the solar radiation for the season 2004	Daily Watts per sq. meter solar radiation sum between April 1st and October 31st
RAD_05	Sum of the solar radiation for the season 2005	Daily Watts per sq. meter solar radiation sum between April 1st and October 31st

RAD_06	Sum of the solar radiation for the season 2006	Daily Watts per sq. meter solar radiation sum between April 1st and October 31st
RAD_07	Sum of the solar radiation for the season 2007	Daily Watts per sq. meter solar radiation sum between April 1st and October 31st
RAD_08	Sum of the solar radiation for the season 2008	Daily Watts per sq. meter solar radiation sum between April 1st and October 31st
RAD_09	Sum of the solar radiation for the season 2009	Daily Watts per sq. meter solar radiation sum between April 1st and October 31st
RAD_MED	Median Sum of solar radiation for season between 1994 and 2007	Daily Watts per sq. meter solar radiation sum between April 1st and October 31st
PH1	Topsoil (10 to 20 cm depth) pH	pH units
AWC1	Topsoil (10 to 20 cm depth) Available water capacity in 150 cm soil profile	cm
Clay1	Topsoil clay content (10 to 20 cm depth)	Percentage
Silt1	Topsoil silt content (10 to 20 cm depth)	Percentage
Sand1	Topsoil sand content (10 to 20 cm depth)	Percentage
Sand2	Soil sand content from another soil source	Percentage (5-30 cm)
Silt2	Soil silt content from another soil source	Percentage (5-30 cm)

Clay2	Soil clay content from another soil source	Percentage (5-30 cm)
PH2	Soil ph from another soil source	pH (5-30 cm)
CEC	Soil cation exchange from another soil source	cmol per kilo (5-30 cm)
CE	Soil cation exchange from another soil source	cmol per kilo (5-30 cm)