FORECASTING TRENDS IN NATIONAL

FOOD SUPPLIES

GOAL

To better understand how the steady growth in population effects a country's supply of fresh fruits and vegetables.

HYPOTHESIS

As countries grow in population, their capacity to produce agricultural commodities decreases and their dependency on importing agricultural commodities increases.

FOOD & AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

The FAO provides Food Balance Sheets which presents the pattern of a country's food supply during a specified reference period.

Data is available for the years 1961 to 2013. While the time period for this dataset is a bit over 50 years, recent innovations in food production may not be accounted for.

The data provides insight as to how much of a certain commodity a country imported, exported, how much was available for consumption, how much was used as feed for livestock, and how much was lost through wastage.

The data also includes population growth and growth predictions for countries from 1961 to the year 2050.

Using these data points, we should hopefully get a fairly accurate prediction on the amounts of fruits and vegetables the US will be importing and producing in the near future and in the next generation (by the year 2050).

DATA CLEAN UP AND PREPROCESSING

The original dataset was not formatted in a way that would best fit the training models I wanted to use. I therefore had to do the following data pre-processing:

- Transposed row data into column data
- Created polynomial features based off the existing feature set
- Created Interaction Terms based off the findings of the existing feature set and their relationships

MODELING AND PREDICTING THE DATA

PREDICTED FEATURE VALUES:

The only predicted value that was not calculated by the models was population. These predicted values came from the FAO.

TIME SERIES CONSIDERATION:

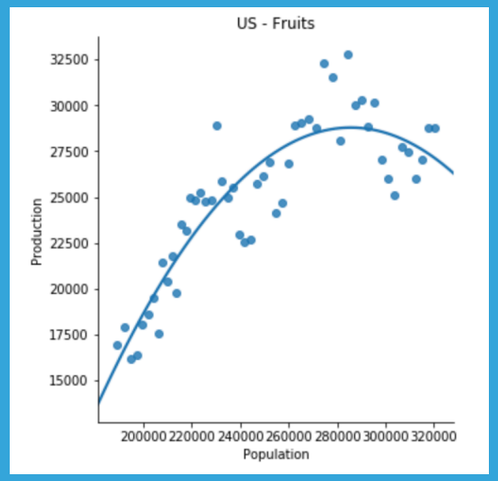
While the data is presented in a Time Series (year-by-year) I believe the model(s) needed are not necessarily time sensitive, but dependent on population.

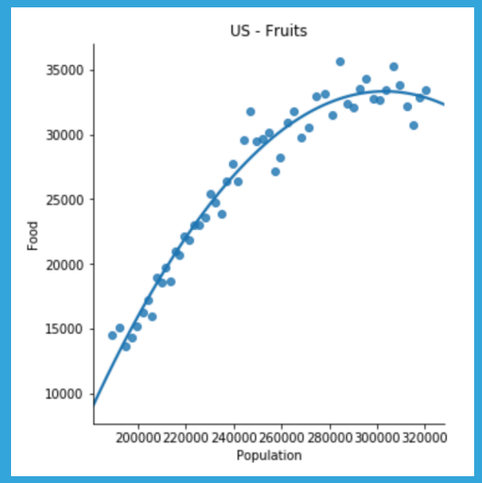
RUNNING MULTIPLE MODELS:

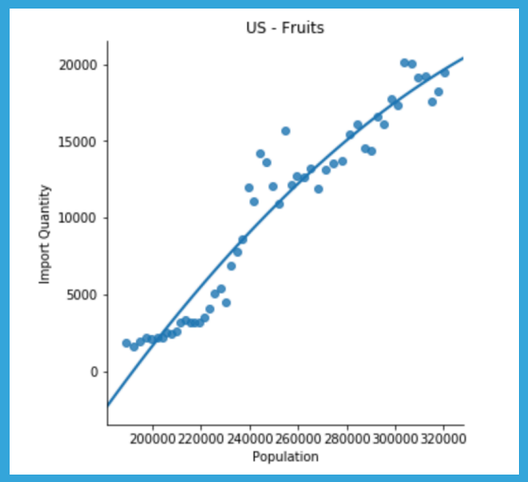
To get the full feature set required to properly train a model that targets import quantity, I created a set of models, each targeting the individual features needed in the final feature set. Each of these models were then put into functions and ran synchronously.

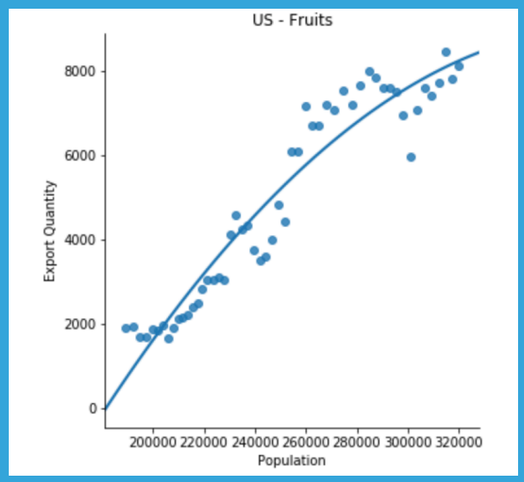
MODEL TYPES:

I chose to use Linear Regression models with K-Folds cross-validation.



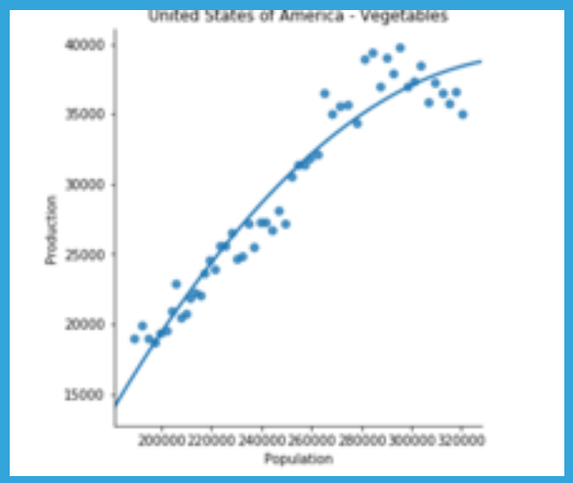


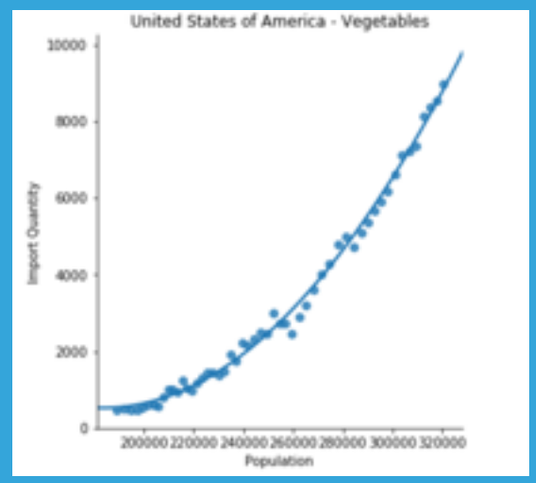




	Country Code	Country	Item Code	Item	Domestic supply quantity	Food	Export Quantity	Import Quantity	Losses	Production	Feed	Stock Variation	Year	Population	Land Use
2768	231.0	United States of America	2919.0	Fruits - Excluding Wine	39208.0	33813.0	7400.0	19161.0	1423.0	27448.0	0.0	0.0	2009.0	309492.0	NaN
2780	231.0	United States of America	2919.0	Fruits - Excluding Wine	37538.0	32212.0	7719.0	19256.0	1352.0	26001.0	0.0	0.0	2010.0	312247.0	NaN
2792	231.0	United States of America	2919.0	Fruits - Excluding Wine	36176.0	30768.0	8468.0	17616.0	1371.0	27028.0	0.0	0.0	2011.0	314912.0	NaN
2804	231.0	United States of America	2919.0	Fruits - Excluding Wine	39122.0	32820.0	7817.0	18239.0	2051.0	28744.0	0.0	-45.0	2012.0	317505.0	NaN
2816	231.0	United States of America	2919.0	Fruits - Excluding Wine	40169.0	33454.0	8135.0	19496.0	2148.0	28757.0	0.0	51.0	2013.0	320051.0	NaN

	Country Code	Country	Item Code	Item	Domestic supply quantity	Food	Export Quantity	Import Quantity	Losses	Production	Feed	Stock Variation	Year	Population
2774	231.0	United States of America	2918.0	Vegetables	40195.0	37686.0	4265.0	7362.0	2316.0	37289.0	195.0	-192.0	2009.0	309492.0
2786	231.0	United States of America	2918.0	Vegetables	40015.0	37496.0	4498.0	8137.0	2320.0	36535.0	200.0	-158.0	2010.0	312247.0
2798	231.0	United States of America	2918.0	Vegetables	39119.0	36650.0	5057.0	8371.0	2281.0	35762.0	192.0	43.0	2011.0	314912.0
2810	231.0	United States of America	2918.0	Vegetables	40010.0	37515.0	5094.0	8525.0	2318.0	36557.0	181.0	22.0	2012.0	317505.0
2822	231.0	United States of America	2918.0	Vegetables	38938.0	36472.0	5538.0	8967.0	2278.0	35058.0	193.0	452.0	2013.0	320051.0





FRUITS

VEGETABLES

PREDICTED POPULATION

↓

Test Set Average: 1336
Test Set - MAE: 155

MODEL TARGETING LOSSES

Test Set Average: 1724
Test Set - MAE: 81

↓

Test Set Average: 24071
Test Set MAE: 1611

MODEL TARGETING PRODUCTION

Test Set Average: 30520
Test Set MAE: 210

↓

Test Set Average: 27435
Test Set MAE: 1087

MODEL TARGETING FOOD

Test Set Average: 28932
Test Set MAE: 352

↓

Test Set Average: NA
Test Set MAE: NA

MODEL TARGETING FEED

Test Set Average: 149
Test Set MAE: 4

↓

Test Set Average: 10119
Test Set MAE: 447

MODEL TARGETING IMPORTS

Test Set Average: 2979
Test Set MAE: 69

↓

Test Set Average: 5378
Test Set MAE: 539

MODEL TARGETING EXPORTS

Test Set Average: 2812
Test Set MAE: 174

US PREDICTIONS

	BY 2020	BY 2030	BY 2040	BY 2050
POPULATION	333783*	354712*	374069*	389592*
VEGETABLE IMPORTS	22%	27%	28%	32%
FRUIT IMPORTS	52%	60%	67%	70%

Population = In millions

Import Percentage = total imports / total production + total imports - total exports

Export Percentage = total exports / total production

CANADA PREDICTIONS

	BY 2020	BY 2030	BY 2040	BY 2050
POPULATION	37603*	40617*	43004*	44948*
VEGETABLE IMPORTS	71%	83%	100%	106%
FRUIT IMPORTS	111%	115%	121%	124%

Population = In millions

Import Percentage = total imports / total production + total imports - total exports

Export Percentage = total exports / total production

FRANCE PREDICTIONS

	BY 2020	BY 2030	BY 2040	BY 2050
POPULATION	65721*	67894*	69648*	70609*
VEGETABLE IMPORTS	64%	73%	78%	83%
FRUIT IMPORTS	66%	72%	83%	89%

Population = In millions

Import Percentage = total imports / total production + total imports - total exports

Export Percentage = total exports / total production

UK PREDICTIONS

	BY 2020	BY 2030	BY 2040	BY 2050
POPULATION	66798*	68158*	68595*	68258*
VEGETABLE IMPORTS	90%	103%	106%	104%
FRUIT IMPORTS	94%	93%	89%	88%

Population = In millions

Import Percentage = total imports / total production + total imports - total exports

Export Percentage = total exports / total production

GERMANY PREDICTIONS

	BY 2020	BY 2030	BY 2040	BY 2050
POPULATION	82540*	82186*	81099*	79238*
VEGETABLE IMPORTS	62%	62%	57%	51%
FRUIT IMPORTS	82%	76%	68%	55%

Population = In millions

Import Percentage = total imports / total production + total imports - total exports

Export Percentage = total exports / total production

JAPAN PREDICTIONS

	BY 2020	BY 2030	BY 2040	BY 2050
POPULATION	127363*	125228*	121750*	118774*
VEGETABLE IMPORTS	19%	16%	10%	6%
FRUIT IMPORTS	57%	46%	31%	24%

Population = In millions

Import Percentage = total imports / total production + total imports - total exports

Export Percentage = total exports / total production

S. KOREA PREDICTIONS

	BY 2020	BY 2030	BY 2040	BY 2050
POPULATION	51506*	52701*	52409*	50456*
VEGETABLE IMPORTS	13%	16%	15%	11%
FRUIT IMPORTS	13%	25%	18%	14%

Population = In millions

Import Percentage = total imports / total production + total imports - total exports

Export Percentage = total exports / total production

MEXICO PREDICTIONS

	BY 2020	BY 2030	BY 2040	BY 2050
POPULATION	133870*	147540*	157689*	164279*
VEGETABLE EXPORTS	51%	72%	85%	99%
FRUIT EXPORTS	25%	35%	40%	51%

Population = In millions

Import Percentage = total imports / total production + total imports - total exports

Export Percentage = total exports / total production

PREDICTING IMPORTS AND EXPORTS

SUMMARY OF MODELS

SUMMARY

TARGET VALUES SEEM TO ALIGN WITH OTHER SOURCE MATERIALS

- Crop lands are decreasing in western countries
- Instead of relying on domestic growth, most developed countries are relying on imports for their nutritional needs
- Values from the models use are similar to other research projects

MODEL DATA DOES NOT TAKE INTO CONSIDERATION OUTSIDE FORCES

- Environmental issues such as drought and flooding are not being evaluated
- Socio/political conflicts are not considered
- Changes in countries overall diet and lifestyle are not being considered

OVERALL FINDINGS

Based off the MAEs and the predicted target values for developed countries; I feel the models reflect possible outcomes for countries with stable economies, higher GDP and consistent patterns of population growth and decline.