Better ATE than never: Reducing Food Waste

M3 Challenge 2018

Team 10388

1 Executive Summary

It has become more clear that as times goes on, food wastage relentlessly increases in exponential numbers. It is said that with great power comes great responsibility, but unfortunately even with the relatively lavish and secure lifestyle that citizens of first world countries such as the United States have, said citizens fail to optimize this great advantage and continue to be world leaders in the food wastage category. While many organizations such as charities exist to distribute food to the needy, we must first focus on the redistribution of potentially wasted pre-existing food to food-insecure people or at least to farmers for compost and regrowth.

We first are required to see if the total food wasted in a state can be used to feed the food-insecure population of that state. To do that, we find the total calories wasted in a state, based on different research and caloric density conversions, and then find out how many children and adults can be fed by those calories. We then find the percentage of the total food-insecure people in the state that can be fed by the wasted calories. For part two, we needed to model the amount of food wasted per household, based on the habits and traits of those households. We boiled down the traits and habits to three main factors: age, with five different categories, income, and number of people in the household. The five categories of ages are toddlers, children, teens, adults, and the elderly. Each of these categories have their specific amount of caloric intake needed for a healthy diet. We then used logistic regression to determine the amount of money used for food based on income, and from that we subtracted the calories necessary in a healthy diet for each person. The leftover money theoretically is wasted, and we found out how much food could be bought with that money, giving us the wasted food.

Our target community that we chose to analyze was the state of Illinois. We first identified major locations of food banks around Illinois to where leftover food from all over the state would be transported to. We then created a Transportation Cost Efficiency score (TCE) computed via leftover food quantity, food value, the real distance transformed from the haversine distance from anywhere in Illinois to the food banks, the cost of transportation per kilometer, and the cost per pound and quantity of compost. This score, when less than 1 would indicate to us that it is more efficient to use the food as compost and vice versa when greater than 1 in that it is more efficient in terms of transportation costs to donate the food. At a score of 1, the break off point, we can choose to donate the food as a sign of moral values.

2 Introduction

Due to the inherent wasteful human consumption tendencies that are prominent in modern society - an unjust paradox exists. Because of modern technology, developed nations such as the United States are able to mass produce food at reduced costs but still are faced with a food-insecure population of over 42 million Americans.[**2**] Thus, as a nation we need to look forward in reducing food waste at both the consumption stage as well as the production and processing stages. This will take a conscious effort but also can be supplemented with the use of mathematical models to present various options to aid in the present issue as seen below. In the ever growing world of computing, big data is helping to lead the way. Now it is time for us as a society to utilize it and the paint the way for a better tomorrow.

3 Just Eat it!

3.1 Restatement of the Problem

The problem asks us to create a robust model that can take relevant state data and output whether that state can feed its food-insecure population. Finally, we will demonstrate our model on Texas in a case-study format.

3.2 Assumptions and Justifications

Assumption: All caloric intake data in our model is using the amount of calories for the average sedentary male.
Justification: Accounting for race, age, and gender will create too much complexity in our model. If a state is actually going to use this model, we need to simplify the caloric needs of people to give a general idea of whether we can actually feed the food-insecure population. Thus, we used 2000 calories as our average caloric intake for an adult and created a small model for children under the age of 18 below [**1**].

Assumption: The data and proportions of categories of food wasted are from North America and Oceania.
Justification: The proportions of the different categories of food wasted will be quite similar from North America to that of Oceania.

Assumption: Population is directly correlated to food wasted.
Justification: We took sample points from 6 points of varying populations and found a linear regression line with $r^2 = 0.9684$.

3.3 Summary of Model

In order to determine the total amount of food-insecure people that can be fed based on food that's wasted, we decided that our model needed to consider 4 criteria. Quantity of food wasted per state, population of food-insecure people per state, types of food being wasted, general nutritional needs of the average human.

3.4 Caloric Intake

We already have the caloric intake of the average adult sedentary human from above. Now, we need to find the expected caloric intake of the average person under the age of 18. To do this, we found data from USDA that describes the caloric needs of children. Then we applied, we applied a linear regression model ($C(A)$) relating age (A) and caloric intake (C), for A from 0 to 18.

$$C(A) = 76.5A + 827 \tag{1}$$

$$\frac{\int_0^{18} C(A)\partial A}{\int_0^{18}(1)\partial A} \tag{2}$$

The integral from above gives us an average caloric intake of children to be 1515.5.
Taking into account both the average adult caloric intake and the value that was obtained from

the average child caloric intake, we wanted to find one specific caloric intake score. We weighted the child's caloric intake based on the proportion of food-insecure children in the state to the total amount of food-insecure people in the state (P_c).
We then weight the adult's caloric intake against the total amount of food-insecure people in the state and added the two values (P_a). P_t is the total amount of food insecure people from a state. Now, we can construct a very simple weighted equation. To find P_t, P_a, and P_c, we look at the overall food insecurity for entire states to get P_t. To find P_c insecurity for an entire state, we use the Child Food Insecurity tables. To find P_a we say $P_a = P_t - P_c$ **[2]**.

$$K = (\frac{P_c}{P_t} * (1515.5)) * (\frac{P_a}{Pt} * (2000)) \quad (3)$$

3.5 Food Wasted

Finding the total amount of food wasted in a state is imperative to deciding how many of the food-insecure people can be fed. In order to do so, we decided to create a linear model to infer the amount of food wasted for a given population (which we can do with assumption 3). Below, we use seven states with varying populations where we could find data as base points for **[3][4][5][6][7][8][9]**. From that data we found a linear model to be defined as

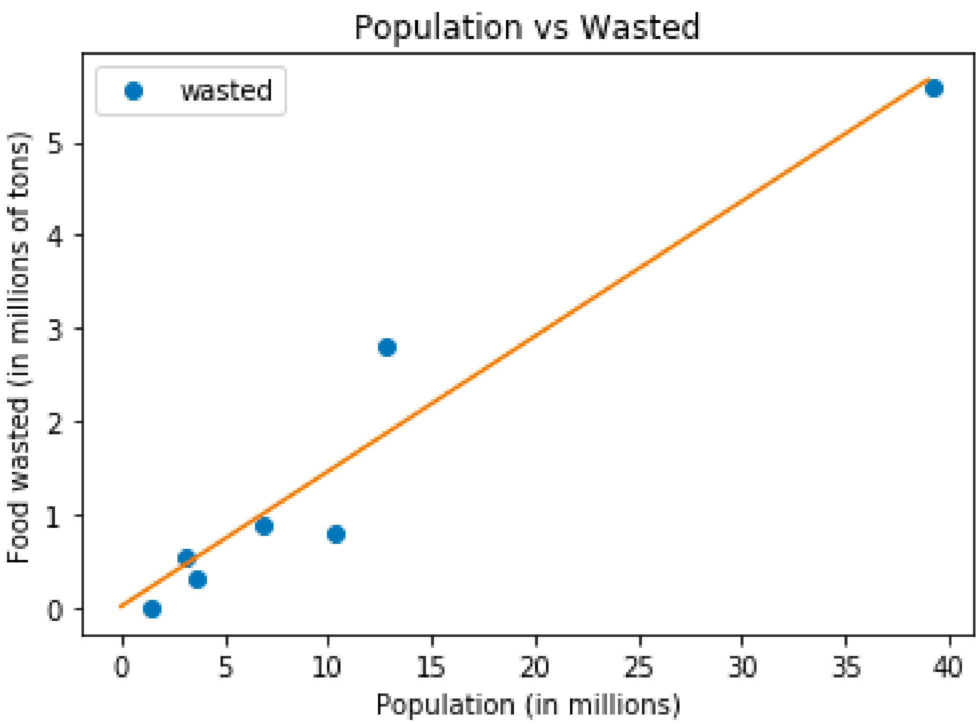
$$W\hat{(}P) = P_0 + b_1 P \quad (4)$$

State	Population (millions)	Food Wasted (millions of tons per year)
California	39.25	5.6
Illinois	12.8	2.8
Georgia	10.31	0.8
Massachusetts	6.81	0.9
Connecticut	3.6	0.32
Iowa	3.1	0.55
Hawaii	1.4	0.26

Looking at the scatterplot and associated line of best fit shows us that a linear model is most applicable. Also, we made sure to calculate the that the coefficient of correlation, $r^2 = 0.937$.

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2(y_i - \overline{y})^2}} \quad (5)$$

This formula is the Pearson Correlation formula used to calculate the r^2 value.



Using a python script, we can input a population size in millions to receive a predicted amount of food wasted. This is solved by our linear model $W(P) = P_0 + b_1P$, where $P_0 = 0.013227968135046009$ and $b_1P = 0.14502829$*x which describes linear regression model. Our script, interactively calculates the average caloric intake from a given population using our model $K = (\frac{P_c}{P_t} * (1515.5)) * (\frac{P_a}{Pt} * (2000))$, where K is equal to the average calories.

Reference to Appendix A: Part 1 Code

If the state itself has data on the amount of food wasted, the linear model is not needed, and as such, no additional calculations are necessary.

3.6 Converting the Food to Calories

It's understood that different types of foods have different caloric densities. Because of this fact, we convert the weight of our categories of food to calories. Given the amount of food wasted within each state, we separate the food wasted into different categories. Each category of food contains foods with similar caloric densities. The different types were: Cereals, Roots and Tubers, Oilseeds and pulses, Fruits and vegetables, Meat, Fish and seafood, and finally Milk.

In order to figure out how much food to put into each category, we used the proportion of type of food to total food wasted **[9]**. Then, we convert the weight of food in each category to calories by using a specific caloric conversion factor for each food, shown in the table below.

Food Category	Calories per Pound
Cereals	1358
Roots and Tubers	181
Oilseeds and Pulses	2024
Fruits and Vegetables	151
Meat	651
Fish and Seafood	953
Milk	191

$$P_{f_c} = \begin{bmatrix} .236498 & .199442 & .094427 & .358824 & .025953 & .053309 \end{bmatrix}$$

$$C_{f_c} = \begin{bmatrix} 1358 \\ 181 \\ 2024 \\ 151 \\ 651 \\ 953 \\ 191 \end{bmatrix}$$

Adding all of the calories, we find the total calories of wasted food in an entire state.
The general form of this would be:

$$T_{w_c} = (W_t * 2000) * (P_{f_c} \cdot C_{f_c}) \tag{6}$$

where T_{w_c} is the total amount of wasted calories in a given state, P_{f_c} is the proportion of food categories, and C_{f_c} is the calorie amount of the respective food groups. The order of P_{f_c} from left to right is the same order of the Food Category table from above meaning it goes from Cereals to Milk. The same order applies for C_{f_c}. W_t is the total amount of food wasted determined from a given state's population. This is what our $W\hat{(P)}$ is for. Also, we must multiply W_t (total amount of food wasted by a given state in tons per year) by 2000 so that our units are compatible with our proportion and calories matrices.

3.7 The Model

Putting it all together, we can find now find the total amount of people that the wasted food can feed. This will be:

$$\frac{T_{w_c}}{K * 365} \tag{7}$$

We divide the total calories of food wasted per year by the average caloric intake of a person per year.In order to find the percentage of people fed with the food, we simply divide this value by the population of food insecure people in the state. Doing so, gives us an percentage of people fed by the wasted food in one year.

$$FSP = \frac{\frac{T_{w_c}}{K*365}}{P_{state}} \tag{8}$$

FSP stands for food security percentage. This is essentially, the percentage of food insecure people in a state that can be fed by the wasted food.

In our python script, the FSP is calculated by the variable peoplefed = (totalcal/(averagecalories * 365))/((childpop + adultpop)*(10**6)). Where T_{w_c} is totalcal, K is averagecalories, and P_{state} is the total population of adults and children in a given population. FSP is also based on a annual model. **Reference to Appendix A: Part 1 Code**

3.8 Case Study: Texas

We now apply our model to Texas. We put all of our data in the table below.

State Population	P_t	P_c	P_a	T_{c_w}	K	FSP
27,469,114	4,320,050	1,713,430	2,606,620	$2.63 * 10^9$	1807.8363	184.52066

After inputting these table values into equation **(8)**, we get an **FSP** of 184.52066% of food insecure population of Texas can be fed yearly. Our Average caloric intake value would then be $K = 1807.8363$. K is obtained from equation **(3)**. We get these values from interactively inputting the state population of Texas into our python script. **Reference to Appendix A: Part 1 Code**

3.9 Validity of Model

The model that we have is very accurate if the data is specific, but lacks when the data is inferred. The linear regression line $W\hat{(P)} = P_0 + b_1 P$.

will not be able to pin-point the amount of food wasted completely accurately, as the r^2 value is not 1. However, it is not entirely viable to measure every single pound of wasted food anyways, so it is not a major factor.

Our model also assumes that every adult needs the same caloric intake, which is not entirely accurate. We also assume that the children that are in food-insecure are evenly distributed over the ages 0-18. As this might not be entirely true, it is an assumption that must be made.

Although we convert the weight of the food to the calories mathematically, there are different factors that could affect the accuracy of the outputs. Every food has a unique caloric density. This makes it difficult to measure the calories of the food completely accurately, which led us to group foods based on category.

4 Food Foolish?

4.1 Restatement of the Problem

This part asks us to create a general model that can be used to determine the amount of food waste a household generates in a year based on their traits and habits.

4.2 Summary of Model

Our model essentially finds the amount of money that was used on wasted food, and converting that value into pounds of food. To do this, we find the total amount of money spent on food based on income and subtract from that the money needed to sustain a healthy diet.
The value that was left over from the subtraction is the amount of money that was used on wasted food. Finally, we convert that value into pounds of food based on a healthy daily diet.

4.3 Assumptions and Justifications

Assumption: A healthy diet for the average human of an age group consists of the average amount an individual in that group would intake on a daily basis.
Justification: Given a large sample size, individuals that eat more than the average human and those that eat less than the norm will average out and yield a healthy caloric intake for the average human of that age group.

Assumption: With left over money not spent on essential food, individuals would spend money on food the same as they would normally.
Justification: To get a logical sense of the amount of food waste that will result from an excess in income in households, we can assume that that income would be spent towards food just like they would spend on it if it were supplying their normal healthy diet.

Assumption: Each category of food can be represented by one main product of it in pricing models.
Justification: Pricing towards specific food groups usually will hover around one of the main products of the whole category (i.e. the price of milk to represent the pricing of dairy products as a whole)

4.4 Setting up the Data

Dietary needs for minimal living (healthy diet) per day[**11**][**12**]

Toddlers (2-8)	Children (8-13)	Teens (13-18)	Adults (18-50)	Elderly (50+)
4 oz grains	5.5 oz grains	6.46 oz grains	6.47 oz grains	5.5 oz grains
1.25 cups vegetables	2.25 cups vegetables	0.92 cups vegetables	1.59 cups vegetables	2 cups vegetables
1.25 cups fruits	1.5 cups fruits	1.08 cups fruits	1.05 cups fruits	1.5 cups fruits
2.25 cups dairy	3 cups dairy	2.16 cups dairy	1.64 cups dairy	3 cups dairy
3 oz meat	5 oz meat	4.33 oz meat	6.13 oz meat	5 oz meat

Data below represents the cost per unit of measurement and is from NationMaster [**13**]

Grains (1 oz)	Vegetables (1 cup)	Fruits (1 cup)	Dairy (1 cup)	Meat (1 oz)
\$0.134376659	\$0.29375	\$0.46125	\$0.221875	\$0.945

$$C_{p_x} = ((W_{p_g} * 0.134376659) + (W_{p_v} * 0.29375) + (W_{p_f} * 0.46125) + (W_{p_d} * 0.221875) + (W_{p_m} * 0.945)) * (365) \quad (9)$$

Here we multiply the amount of food in each category by its cost per unit conversion factor for all 5 subgroups of food. Then, we are able to obtain the total cost of essential food for a specific member of a household. Every p subscript on the RHS of equation **(9)** denotes the category of that person (ie.toddler, children, teen, adult, elderly). In the Appendix Part 2 Code, we have the real values from the tables above inputted into the general equation **(9)**. We multiply by 365 for a yearly calculation. In our python script the variable C_{p_x} is the cost of a given person in a house hold. **Reference to Appendix B: Part 2 Code**

To find the minimum cost of food for a household in a given year, we use the equation below. This accounts for each age group within the household and will add up each individuals total cost for essential food.

$$T_m = (C_{p_t} * T) + (C_{p_c} * C) + (C_{p_te} * T_e) + (C_{p_a} * A) + (C_{p_e} * E) \tag{10}$$

In our code T_m is the "total" of the cost in given household. **Reference to Appendix B: Part 2 Code**.

T_m is going to give us the total minimum cost of food for a household in a given year, we use the equation below. T(Toddlers), C(Children), Te(Teenagers), A(Adults), and E(Elderly) indicates the number of people living in that household of those respective categories.

4.5 Weight of Food Needed (tons per year)

The total amount of weight in tons of the essential food consumed by each specific age group of a given household are represented by these matrix calculations below.

We take the trace of each respective category of that person to get their specific weight calculations.

$$TT = 365 * \begin{bmatrix} 0.000125 & 0 & 0 & 0 & 0 \\ 0 & 0.000110358 & 0 & 0 & 0 \\ 0 & 0 & 0.000110358 & 0 & 0 \\ 0 & 0 & 0 & 0.000198645 & 0 \\ 0 & 0 & 0 & 0 & 0.00009375 \end{bmatrix} \tag{11}$$

$$T_{t_w} = tr(TT) \tag{12}$$

$$TC = 365 * \begin{bmatrix} 0.000171875 & 0 & 0 & 0 & 0 \\ 0 & 0.000198645 & 0 & 0 & 0 \\ 0 & 0 & 0.00013243 & 0 & 0 \\ 0 & 0 & 0 & 0.00026486 & 0 \\ 0 & 0 & 0 & 0 & 0.00015625 \end{bmatrix} \tag{13}$$

$$T_{c_w} = tr(TC) \tag{14}$$

$$TTE = 365 * \begin{bmatrix} 0.000201875 & 0 & 0 & 0 & 0 \\ 0 & 0.000081224 & 0 & 0 & 0 \\ 0 & 0 & 0.00013243 & 0 & 0 \\ 0 & 0 & 0 & 0.000190699 & 0 \\ 0 & 0 & 0 & 0 & 0.0001353125 \end{bmatrix} \tag{15}$$

$$T_{te_w} = tr(TTE) \tag{16}$$

$$TA = 365 * \begin{bmatrix} 0.0002021875 & 0 & 0 & 0 & 0 \\ 0 & 0.000140376 & 0 & 0 & 0 \\ 0 & 0 & 0.000092701 & 0 & 0 \\ 0 & 0 & 0 & 0.00014479 & 0 \\ 0 & 0 & 0 & 0 & 0.0001915625 \end{bmatrix} \tag{17}$$

$$T_{a_w} = tr(TA) \tag{18}$$

$$TE = 365 * \begin{bmatrix} 0.000171875 & 0 & 0 & 0 & 0 \\ 0 & 0.000176573 & 0 & 0 & 0 \\ 0 & 0 & 0.00013243 & 0 & 0 \\ 0 & 0 & 0 & 0.00026486 & 0 \\ 0 & 0 & 0 & 0 & 0.00015625 \end{bmatrix} \tag{19}$$

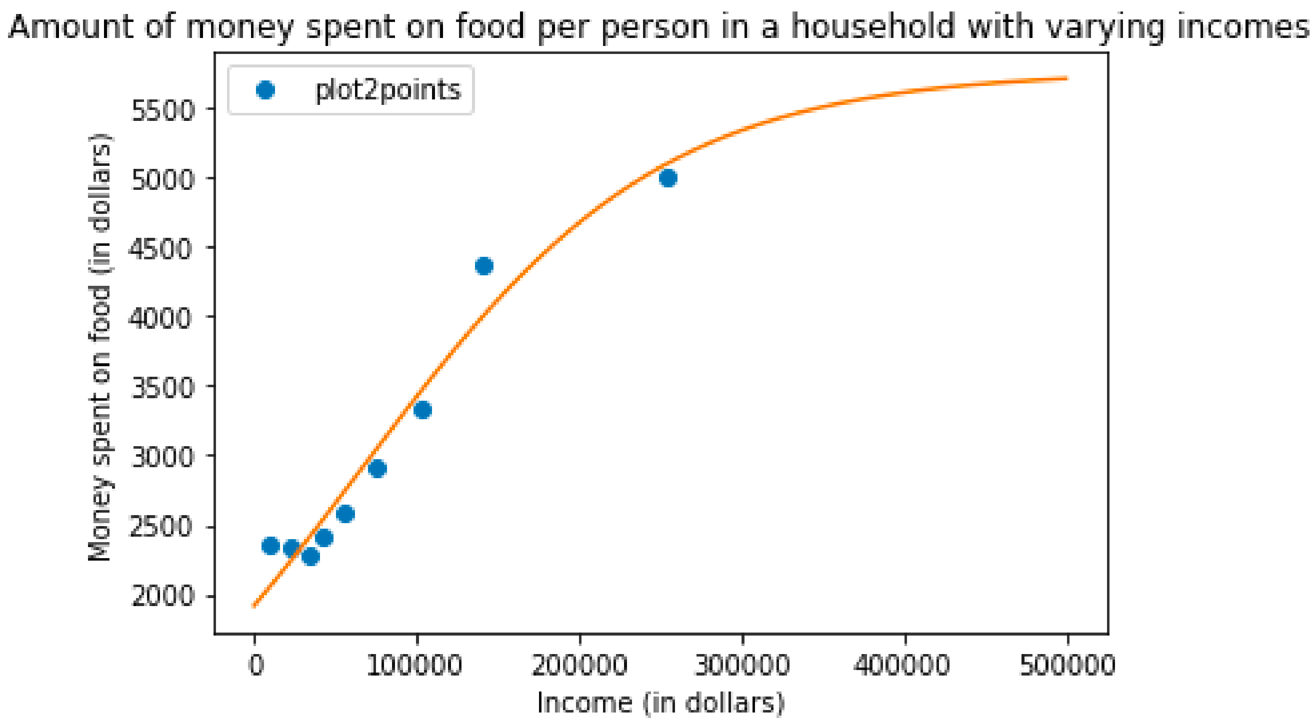
$$T_{e_w} = tr(TE) \tag{20}$$

$$T_w = (T_{t_w} * T) + (T_{c_w} * C) + (T_{te_w} * T_e) + (T_{a_w} * A) + (T_{e_w} * E) \tag{21}$$

T_w is the total weight of food wasted for a given household in tons per year - adding up each specific age group's total weights.

4.6 Money Spent on Food Based on Income

Using the ConsumerBehaviorBasedonIncome data set, we discovered that there is a positive relationship between income and money spent on food. Another note from the data is that there is a minimum amount of money spent on food per person per year and there is also a maximum amount of money spent on food per person per day. One must buy a minimum amount of food to barely survive, and while many wealthy individuals can buy the most expensive foods everyday, there will still be a maximum capacity. Since there is a minimum and maximum, we used a logistic curve to model the data shown below.



The logistic regression curve is described by the equation

$$M(I) = \frac{5766.728826}{1 + 2.009830121e^{0.0000107580841}} \tag{22}$$

The value of I must be greater than 0, as it is unlikely to have a negative income.

Our python script, uses the matplotlib to calculate the graph of the logistic regression curve. We use the formula $\frac{q}{1+we^{-rx}}$ where q = 5766.728826, w = 2.009830121, r = 1.0758084 * (10**-5). This is the amount of money spent on food per person in a household with varying incomes. **Reference to Appendix B: Part 2 Code**

4.7 Final Model

H is the total amount of people in a household.

$$H = T + C + TE + A + E \tag{23}$$

$$LF = \frac{(H * M(I)) - T_m}{T_m} * T_w \tag{24}$$

Our final equation above, shows the left over food in tons per year. This is the amount of food wasted by a household per year based on their traits and habits.

In our python script, LF represent the leftover food in tons per year. We then use our logistic regression model which is $M(I)$ which is multiplied by the number of people in a given household respectively, H. We subtract this by the the total T_m and divide all of that by T_m. We take this number and multiply it by T_w,which is the total weight per year in tons of each category of person in household. **Reference to Appendix B: Part 2 Code**

4.8 Case Study 1

We now need to evaluate our model with a single parent and one toddler with an annual income of $20,500. All data is in tons per year. Since the amount of leftover money in our model evaluates to $-253.96 which is less than 0, we can proceed with the assumption that the household is not going to buy any excess food that they will not consume entirely.

Toddlers	Children	Teens	Adults	Elderly
1	1	0	0	0

Income	Total Minimum Cost of Household	Leftover Money	Total Amount of Food Wasted
$20,500	$4669.44	$-253.96	0

4.9 Case Study 2

We now need to evaluate our model with 2 parents and 2 teenage children with an annual income of $135,000. All data is in tons per year. With an excess income of $5332.39 for this household according to our model, we can move with the assumption that they are to spend that on food similarly as they did before to create 0.57 tons of wasted food in that fiscal year.

Toddlers	Children	Teens	Adults	Elderly
0	0	2	2	0

Income	Total Minimum Cost of Household	Leftover Money	Total Amount of Food Wasted
$135,000	$10355.12	$5332.99	0.57

4.10 Case Study 3

We now need to evaluate our model with an elderly couple with an annual income of $55,000. All data is in tons per year. With only a slight excess income of $51.73 in this household, this couple will buy and end up wasting .0063 tons of food with our model.

Toddlers	Children	Teens	Adults	Elderly
0	0	0	0	2

Income	Total Minimum Cost of Household	Leftover Money	Total Amount of Food Wasted
$55,000	$5408.62	$51.73	0.0063

4.11 Case Study 4

We now need to evaluate our model with a single 23 year old with an annual income of $45,000. All data is in tons per year.

Toddlers	Children	Teens	Adults	Elderly
0	0	0	1	0

Income	Total Minimum Cost of Household	Leftover Money	Total Amount of Food Wasted
$45,000	$2911.79	$-335.69	0

4.12 Validation

For case study 1, we see a single parent with a toddler making $20,500 annually. Using our model, the results show that the total minimum amount of money spent on food of this household in a year is $4669.44. The leftover money is $-253.96 because they aren't enough money to waste food. This is all based on the daily nutritional needs of the average adult and toddler. So, this is why our final amount of food wasted annually in tons is 0.

For case study 2, we see 2 parents and 2 teenage children making $135,000 annually. Using our model, the results show that the total minimum amount of money spent on food of this household in a year is $10,355.12. The leftover money is $5332.99 because they are making and spending enough money to exceed their daily nutritional needs. Thus, they are wasting the excess food that their higher income affords them. This is all based on the daily nutritional needs of the average adult and teenager. So, this is why our final amount of food wasted annually in tons is 0.57.

For case study 3, we see an elderly couple making $55,000 annually. Using our model, the results show that the total minimum amount of money spent on food of this household in a year is $5,408.62. The leftover money is $51.73 because they are making and spending enough money to exceed their daily nutritional needs. Thus, they are wasting the excess food that their higher income affords them. This is all based on the daily nutritional needs of the average elderly person. So, this is why our final amount of food wasted annually in tons is 0.0063.

For case study 4, we see a single adult making $45,000 annually. Using our model, the results show that the total minimum amount of money spent on food of this household in a year is $2911.79. The leftover money is $-335.68 because he/she is spending more of his/her income on housing and other confounding factors that appear while living alone. So, that's why our final amount of food wasted annually in tons is 0. This is all based on the daily nutritional needs of the average adult.

5 Hunger Game Plan?

5.1 Restatement of the Problem

In this part, we were tasked with optimizing a community that we are a part of by creating a model to provide insight on which strategies we could adopt to re purpose the maximal amount of food at the minimum cost.

5.2 Overview

We want to create a model that decides what to do with leftover food without wasting any of it. We want to create an efficiency score, the Transportation Cost Efficiency (TCE), for each food bank, which tells us how to handle the food in the most cost efficient manner. The two options for leftover food are either to serve it to food-insecure people in Illinois, or to compost it into nutrients for the growth of new crops. We create a model that compares the cost efficiency of either option, based on distance of food to the food bank, where food is stored for the food-insecure, the cost per pound of food, and the amount of food being sent.

All of these calculations are from our python model.

Reference to Appendix B: Part 2 Code

5.3 Assumptions and Justifications

Assumption: In our model we did not factor in the expiration of food because we are only within Illinois.
Justification: Since we are in Illinois, the transportation from across the state will always be less than a day, so expiration should not be a major factor in our model.
Assumption: There are food-insecure people near all Illinois Food Banks.
Justification: According to the food map[**15**], there are food-insecure people in every food-bank region in Illinois. Therefore, for every food bank, there are hungry people.
Assumption: Compost created by food will be the same weight as the food.
Justification: We assume this by the conservation of mass.

5.4 Cost for Transportation of Food

Are first order of business is to determine the cost of delivering food to the closest food bank. We do this by using the latitude and longitude of both the food bank and the location of the food, and finding the estimated drive distance between the two points, and finding the cost it takes for a truck to drive between the points.

To do so, we use the Haversine formula, which gives us the great-circle distance between two latitude-longitude points. The location of the food-banks is found by the [**15**]. We then have a function

$$d = 2r\arcsin\left(\sqrt{sin^2\left(\frac{\phi_2-\phi_1}{2}\right) + cos(\phi_1)cos(\phi_2)sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)}\right) \tag{25}$$

(ϕ_1, λ_1) are points in latitude and longitude and ($\phi 2$, $\lambda 2$) are points in latitude and longitude. Using the Haversine equation above and our points, we can easily find the great-circle distance with d.

The great-circle distance on Earth is not accurate for the actual distance traveled by a truck, as the roads and highways are not straight from one location to another. However, we note that there is a relationship between the great-circle distance and the actual distance via roads and highway in the United States of America.

Using one reference point in Illinois, we compare it to 15 points in the United States and test to see if there is a relationship between the great-circle distance and the road-highway distance. These 15 points vary in distance and direction from the original point, in order to take in account and variables that may be affecting the relationship.

Plotting these points and using linear regression, we find a curve that fits the data points with a r^2 coefficient 0.999, which gives us an extremely accurate line to model the relationship between the great-circle distance and the road-highway distance. The line is below:

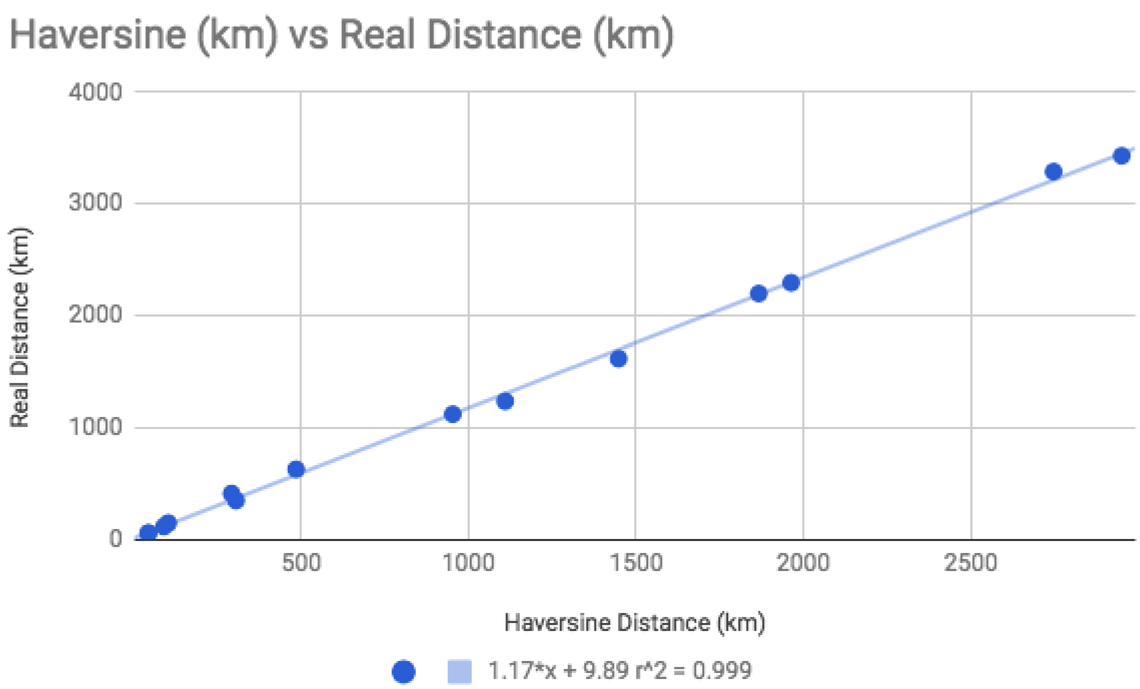
$$R(H) = 1.17H + 9.89 \tag{26}$$

where R is the actual road-highway distance, and H is the great-circle distance.

Locations	Haversine Distance	Real Distance	Real:Haversine (Ratio)
CL to Chicago	42.708	51.98181	1.217144563
CL to Waukesha County Airport	102.649	144.5191	1.407895839
CL to Denver	1447.677	1617.391	1.117231952
CL to Helena	1963.301	2293.315	1.168091393
CL to Seattle	2747.86	3289.499	1.197113026
CL to San Francisco	2951.543	3431.121	1.162483826
CL to Lansing	291.996	408.773	1.399926711
CL to Jackson	1109.859	1235.98	1.113636957
CL to Indianappolis	305.647	346.009	1.132054298
CL to Naperville	43.694	55.6833	1.274392365
CL to Rockford	89.912	113.2978	1.260096539
CL to Naples (FL)	1866.495	2196.755	1.176941272
CL to Harrisburg	953.231	1120.1	1.175056204
CL to Columbus	485.107	626.035	1.290509104

CL is central location which is defined as the location where we are doing the challenge! The latitude of **CL** is 42.130832 and the longitude is -88.

Reference to Appendix C: Part 3 Code



From our linear model, we compute that $r^2 = 0.999$ meaning that there is a strong positive correlation between Haversine Distance (km) and Real Distance (km) meaning that we can create a conversion factor between the two.

Using this, we can find the efficiency of driving a truck between two latitude longitude points. According to Truck Driver Report[**14**], it takes 0.85795 dollars per kilometer for a truck. The total cost of driving between two points would then be:

$$(LB_f * R_f) - (0.85795)(R(d)) \tag{27}$$

Where LB_f is the weight of the food, and R_f is the pound to dollar rate of the type of food, given by the table below.

Because the food that is being transported will have value, we subtract the cost to transport a truck from the cost of the food. This allows us to see the net cost it takes to transport the food, which we want to find.

5.5 Cost of Compost

We need to compare the value of compost for a given value of food, and the net cost to transport food found in Section 6.4.

By The National Compost Costs[**16**], we find that the average value of compost is 1.875 dollars per pound. To find the value of compost, we take the weight of food and multiply by 1.875. This gives us the value of the compost.

5.6 The Model

We now find the TCE score for each food bank, we need to compare what the net cost of transporting food would be and what the cost of the compost created by the food being transported would be.

We decide to find TCE by finding the ratio between the net cost of transporting food and the supposed cost of composted created by the food. This allows us to see if it is more cost efficient to transport the food to the food bank, or if it more cost efficient to compost the food. The final model is:

$$M(LB_f, R_f, d) = \frac{(LB_f * R_f) - (0.85795)(R(d))}{(1.9375 * LB_f)} \tag{28}$$

Where LB_f is the weight of the food, R_f is the pound to dollar rate of the type of food, given by the table below, and R(d) is the road-highway distance between the two locations.

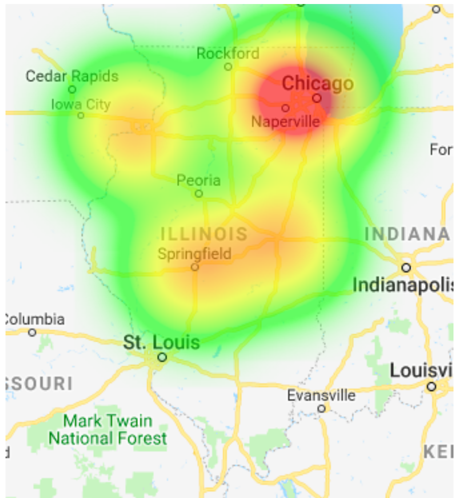
0.85795 is the rate for the truck found previously, and 1.9375 is the value of compost per pound.

If the model gives us a TCE score of below one, we know that it is more cost efficient to compost the food. If the TCE score is exactly one, or above one, we know it is more cost efficient to transport the food to the food bank. The reason we transport the food even if the TCE score equals one is because if both the net cost of transporting food and the supposed cost of composting the food are equal (making the ratio one), we decide that it is better to give food immediately to people who need it. That allows us to give food to people who need it immediately, instead of composting it.

In order to apply our model, we find the TCE score for the location of food to each of the food banks in Illinois.

If all of the TCE scores for the food banks are less than one, we compost the food. However, if the TCE score for one of the food banks is equal to or greater than one, we transport the food to that specific food bank. If the TCE score is equal to or more than one for multiple food banks, we want to transport the food to the nearest food bank. The nearest food bank would be the location with the highest TCE score, as the compost value would be the same for all Food Banks, which means that the closest food bank would have the highest TCE score.

By doing this, we find the most cost efficient option for all unused food. During the next section, we test the model.



We assembled a general heat map surrounding the food banks in Illinois. Each one of these banks create a point of interest on our heat map, and areas in the colored area represent the general places where donating food to charities is going to be cost efficient. These points will vary depending on the type of food and how much there is, but this is a general map that helps you conceptually visualize what locations will benefit from donating the food to charity. The red area is very cost efficient to donate food, while the green is less cost efficient to donate food, and all the un-colored area represents locations that will benefit more from turning the food into compost.

5.7 Case Study 1

For this case study, we will take a theoretical trip from the town of Palatine to the food bank in Chicago. The truckload will consist of 550 pounds of grain. To calculate the score we input the longitudes and latitudes for both cities, food rate, and the weight of the food.

Palatine:

- Latitude: 42.1103
- Longitude: -88.0342

Chicago Food Bank:

- Latitude: 41.818101
- Longitude: -87.726805

Food Rate: 1.47
Food Pounds: 550
Distance: 41.251 km

This yields an efficiency score of 0.7119138796246334, which is less than 1, which means that this food is more efficiently used by converting it to compost.

5.8 Case Study 2

For this case study, we will take a theoretical trip from the town of Bloomington to the food bank in Champaign. The truckload will consist of 50 pounds of fruit. To calculate the score we input the longitudes and latitudes for both cities, food rate, and the weight of the food.

Bloomington:

- Latitude: 40.4842
- Longitude: -88.9937

Champaign Food Bank:

- Latitude: 40.137927
- Longitude: -88.222839

Food Rate: 1.57
Food Pounds: 50
Distance: 75.86 km

This yields an efficiency score of -0.06285539612903229, which is less than 1, which means that this food is more efficiently used by converting it to compost.

5.9 Case Study 3

For this case study, we will take a theoretical trip from the town of Oak Park to the food bank in Chicago. The truckload will consist of 1000 pounds of vegetables. To calculate the score we input the longitudes and latitudes for both cities, food rate, and the weight of the food.

Oak Park:

- Latitude: 41.8850
- Longitude: -87.7845

Chicago Food Bank:

- Latitude: 41.818101
- Longitude: -87.726805

Food Rate: 1.87
Food Pounds: 1000
Distance: 8.841 km

This yields an efficiency score of 0.9562061268258064, which is less than 1, which means that this food is more efficiently used by converting it to compost.

5.10 Case Study 4

For this case study, we will take a theoretical trip from the town of Decatur to the food bank in Springfield. The truckload will consist of 300 pounds of meat. To calculate the score we input the longitudes and latitudes for both cities, food rate, and the weight of the food.

Decatur:

- Latitude: 39.8403
- Longitude: -88.9548

Springfield Food Bank:

- Latitude: 39.795006
- Longitude: -89.628637

Food Rate: 3.87
Food Pounds: 300
Distance: 57.77 km

This yields an efficiency score of 1.8831141991397848, which is more than 1, which means that this food is more efficiently used by donating the food to charity.

5.11 Case Study 5

For this case study, we will take a theoretical trip from the town of Rock Island to the food bank in Davenport. The truckload will consist of 800 pounds of meat. To calculate the score we input the longitudes and latitudes for both cities, food rate, and the weight of the food.

Rock Island:

- Latitude: 41.5095
- Longitude: -90.5787

Davenport Food Bank:

- Latitude: 41.499524
- Longitude: -90.636543

Food Rate: 3.87
Food Pounds: 800
Distance: 8.841 km

This yields an efficiency score of 1.9887484752741937, which is more than 1, which means that this food is more efficiently used by donating the food to charity.

6 Citations

[**1**]https://www.cnpp.usda.gov/sites/default/files/usda_food_patterns/EstimatedCalorieNeedsPerDayTable.pdf

[**2**]http://www.feedingamerica.org/research/map-the-meal-gap/data-by-county-in-each-state.html

[**3**]http://www.epa.illinois.gov/topics/waste-management/food-waste/index

[**4**]https://www.kcet.org/food-living/how-californians-are-fighting-food-waste-on-the-farm-at-the-store-and-at-home

[**5**]http://www.hawaiinewsnow.com/story/31883111/hawaii-wastes-1b-worth-of-food-a-year-and-consumers-are-the-worst-culprits

[**6**]http://www.ct.gov/deep/cwp/view.asp?a=2718&q=552676&deepNav_GID=1645

[**7**]https://www.desmoinesregister.com/story/entertainment/dining/2017/12/15/iowans-wasted-over-556000-tons-food-2017/952888001/

[**8**]https://epd.georgia.gov/food-residuals-diversion

[**9**]https://recyclingworksma.com/how-to/materials-guidance/food-waste-2/

[**10**]http://www.fao.org/docrep/014/mb060e/mb060e00.pdf

[**11**]http://www.nia.nih.gov/health/serving-and-portion-sizes-how-much-should-i-eat

[**12**]https://pediasure.com/child-development-nutrition/5-food-groups-kids

[**13**]www.nationmaster.com/country-info/stats/Cost-of-living/Prices-at-markets

[**14**]https://www.thetruckersreport.com/infographics/cost-of-trucking/

[**15**]http://map.feedingamerica.org/

[**16**]http://www.recycle.cc/compostprices.pdf

7 Software Used

- Python
- Jupyter Notebook
- Google Maps API
- Matplotlib

- Pandas
- Numpy
- https://github.com/eckels/m3challenge2018

Appendix A Part 1 Code

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
dataset = pd.read_csv('food_waste.csv')

X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_
test_split(X, y, test_size=0.2, random_state=0)

from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
print(regressor.coef_)
print(regressor.intercept_)

def graph(formula, x_range):
    x = np.array(x_range)
    y = eval(formula)
    dataset.plot(x='population', y='wasted', style='o')
    plt.title('Population vs Wasted')
    plt.xlabel('Population (in millions)')
    plt.ylabel('Food wasted (in millions of tons)')
    plt.plot(x, y)
    plt.show()

graph('0.013227968135046009 + 0.14502829*x', range(0, 40))
y_pred = regressor.predict(X_test)
df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})

userx = float(input('Population in millions here: '))
prediction[0] = regressor.predict(userx)

print('\nFood wasted in millions of tons: ', prediction)

poundprediction = prediction * (10**6) * 2000

print('Food wasted in millions of pounds: ', poundprediction, '\n')
```

```
graincal = poundprediction * 1358
print('Cereal/Grain Caloric Intake: ', graincal)

rootcal = poundprediction * 181
print('Roots/Tubers Caloric Intake: ', rootcal)

oilcal = poundprediction * 2024
print('Oilseeds and Pulses Caloric Intake: ', oilcal)

fruitcal = poundprediction * 151
print('Fruits and Vegetables Caloric Intake: ', fruitcal)

meatcal = poundprediction * 651
print('Meat Caloric Intake: ', meatcal)

fishcal = poundprediction * 933
print('Fish and Seafood Caloric Intake: ', fishcal)

milkcal = poundprediction * 191
print('Milk Caloric Intake: ', milkcal)

totalcal = graincal + rootcal + oilcal + fruitcal + meatcal + fishcal + milkcal
print('\nTotal Caloric Intake: ', totalcal)

childpop = float(input('Children population in millions: '))
adultpop = float(input('Adult population in millions: '))


averagecalories = (childpop/(childpop + adultpop))*(1515.25) +
(adultpop/(childpop + adultpop))*(2000)
peoplefed = (totalcal/(averagecalories * 365) /((childpop + adultpop)*(10**6))

print('\nAverage caloric intake of this population: ', averagecalories)
print('Population fed annually in millions: ', peoplefed)
```

Appendix B Part 2 Code

```
#Determine amount of food waste a
household generates in a year  based on habits and traits
import math
#Logistic curve
z = 5766.728826
y = 2.009830121
b = 1.0758084 * (10**-5)
e = math.exp(1)
x = float(input("Enter Income: "))
```

```
lcurve = z/(1 + y*e**(-b * x))
#COSTS OF FOODS
# 1 oz of grains
gr = .13437659
# 1 oz of meat
me = .945
# 1 cup of fruit
fr = .46125
# 1 cup veg
veg = .29375
# 1 cup of dairy
da = .221875


adult =  ((6.47 * gr) + (1.59 * veg) + (1.05 * fr)
+ (1.64 * da) + (6.13 * me)) * (365)
toddlers = ((4 * gr) + (1.25 * veg) + (1.25 * fr)
+ (2.25 * da) + (3 * me)) * (365)
children = ((5.5 * gr) + (2.25 * veg) + (1.5 * fr)
+ (3 * da) + (5 * me)) * (365)
teens = ((6.46 * gr) + (0.92 * veg) + (1.08 * fr)
+ (2.16 * da) + (4.33 * me)) * (365)
elderly = ((5.5 * gr) + (2 * veg) + (1.5 * fr)
+ (3 * da) + (5 * me)) * (365)
a = float(input("Number of adults: "))
t = float(input("Number of toddlers: "))
c = float(input("Number of children: "))
te = float(input("Number of teens: "))
e = float(input("Number of elderly: "))
h = a + t + c + te + e
total = ((adult * a) + (toddlers * t)
+ (children * c) + (teens * te) + (elderly * e))
leftoverMoney = (h * lcurve) - total
print("Total minimum cost of a given household in a year: ",total)
print("Leftover money: ", leftoverMoney)
######## ADULTS  WEIGHT PER YR IN TONS #########
gr1 = 0.0002021875 * 365
veg1 = 0.000140376 * 365
fr1 = 0.000092701 * 365
da1 = 0.00014479 * 365
me1 = 0.0001915625 * 365
totalAdW = gr1 + veg1 + fr1 + da1 + me1
######## TODDLERS  WEIGHT YR IN TONS  #########
gr2 = 0.000125 * 365
veg2 = 0.000110358 * 365
fr2 = 0.000110358 * 365
da2 = 0.000198645 * 365
me2 =  .00009375 * 365
```

```
totalTodW = gr2 + veg2 + fr2 + da2 + me2
######## CHILDREN  WEIGHT YR IN TONS #########
gr3 = 0.000171875 * 365
veg3 = 0.000198645 * 365
fr3 = 0.00013243 * 365
da3 = 0.00026486 * 365
me3 =  0.00015625 * 365
totalChiW = gr3 + veg3 + fr3 + da3 + me3
######## TEENS  WEIGHT YR IN TONS #########
gr4 = 0.000201875 * 365
veg4 = 0.000081224 * 365
fr4 = 0.00013243 * 365
da4 = 0.000190699 * 365
me4 =  0.0001353125 * 365
totalTeenW = gr4 + veg4 + fr4 + da4 + me4
######## ELDERLY  WEIGHT YR IN TONS #########
gr5 = 0.000171875 * 365
veg5 = 0.000176573 * 365
fr5 = 0.00013243 * 365
da5 = 0.00026486 * 365
me5 =  0.00015625 * 365
totalEldW =  gr5 + veg5 + fr5 + da5 + me5
totalWeight = (a * totalAdW) +
(t * totalTodW) + (c * totalChiW) + (te * totalTeenW) + (e * totalEldW)
leftoverFood = (leftoverMoney / (total)) * totalWeight
if (leftoverMoney < 0):
  print("total amount of food wasted in tons: 0")
else:
  print("This is the total weight of food
  wasted for given household per year in tons: ", leftoverFood)
```

Appendix C Part 3 Code

```
def haversine(coord1: object, coord2: object):
    import math
    #Coordinates in decimal degrees (e.g. 2.89078, 12.79797)
    lat1,lon1 = coord1
    lat2,lon2 = coord2
    R = 6371000  # radius of Earth in meters
    phi_1 = math.radians(lat1)
    phi_2 = math.radians(lat2)
    delta_phi = math.radians(lat2 - lat1)
    delta_lambda = math.radians(lon2 - lon1)
    a = math.sin(delta_phi / 2.0) ** 2
    + math.cos(phi_1) * math.cos(phi_2) * math.sin(delta_lambda / 2.0) ** 2

    c = 2 * math.atan2(math.sqrt(a), math.sqrt(1 - a))
    meters = R * c  # output distance in meters
```

```
    km = meters / 1000.0  # output distance in kilometers
    meters = round(meters, 3)
    km = round(km, 3)
    print(f"Distance: {km} km")
    return km

def score(lat1, long1, lat2, long2, foodrate, foodlbs):
    topeqn = (foodlbs * foodrate) -
    (0.8575 * (( 1.17 * haversine([lat1, long1], [lat2, long2]) ) + 9.89))
    boteqn = (1.9375) * foodlbs
    fineqn = topeqn / boteqn
    print('\nEfficiency Score: ', fineqn)
    if (fineqn > 1):
        print('More cost efficient to donate the food.')
    elif (fineqn == 1):
        print('Equal cost efficiency for donating or composting food.')
    else:
        print('More cost efficient to compost the food.')



ulat1 = float(input('First Lat: '))
ulong1 = float(input('First Long: '))
ulat2 = float(input('Second Lat: '))
ulong2 = float(input('Second Long: '))
ufoodrate = float(input('Food Rate: '))
ufoodlbs = float(input('Food Pounds: '))

score(ulat1, ulong1, ulat2, ulong2, ufoodrate, ufoodlbs)
```