Urban Food Access Analysis

Andrew Baer

Tanin Nazmul

Amanpreet Kaur

Aditya Thakur

Varinder Singh

July 19th, 2017

MIS 6324.5UI

Dr. Russell R. Torres

Table of Content

Executive Summary

This report contains a thorough review of the data analysis of Urban Census tracts in the state of Texas to determine the data objects that are connected to Food Deserts. Food deserts are defined as parts of the country without accessible and affordable healthy and nutritious foods, essentially low income and low access areas.

To identify food deserts in urban areas the variable "lalowi1share" within the dataset was utilized. lalowi1share is defined as a percentage of the urban tract population that are living within a food desert.

The analysis identified various expected and unexpected results. Some of the expected results show that increasing the poverty rate and decreasing the median family income of a tract will increase the chance that the tract is a food desert.

However, additional unexpected findings within the decision trees modeling showed that the predictor TractAsian, total count of Asian population in tract, was the most important predictor in determining where to branch off from. Also the clusters generated within the clustering model show great initial separation at Family Income, but then separate by the ethnicities of the food deserts. For example, the clustering model gave food deserts that were predominately white, predominately black, and a few that were predominately Asian.

These findings could be utilized to identify current deserts and could lead to legislation that helps these disenfranchised areas. Additionally, these findings could be utilized to reduce future urban tracts of ever becoming a food desert. By analyzing the data and trends, potential future food deserts can be identified before they are ever established.

Furthermore, by identifying which areas of urban tracts are food deserts, city and county officials would have the ability to incentivize grocery stores in certain key areas to reduce the overall risks of food deserts.

Motivation and Background

From the start of this analysis project, our group unanimously found ourselves centered around one desire. We wanted to obtain a data set that would be impactful and memorable to both ourselves as future business leaders, but also our peers.

After putting several options on the table, food access stood out to us as a large opportunity for social impact. Our team believes that our analysis can help improve food access for more people. Food access aligns with our mutual desires to enact positive social change and this keeps us motivated.

For far too many people in our country, healthy food is not an option, financially or spatially. People are subjected to driving long distances or paying inflated prices to find nourishment. Proper nutrition is at the corner stone of an active, happy, and healthy lifestyle. Without access, these communities face hardships for both youth and adult populations. Research studies show the nutrition in a child's early years is linked to their health and academic performance in later years.[1] Poor nutrition in adults has been tied to nearly half of deaths from heart disease, diabetes, and strokes.[2] Food impacts all stages of life.

Population growth in the United States is being driven by these very communities in need.[3] The future of our country will depend on these communities. Our report is not only an opportunity to utilize and discuss our business analytics skill, but a call to action for our peers. Food is a human right. What will our future look like if we don't fix this problem?

Data Description

We analyzed Food Access Research Atlas, secondhand data collected by the United States Department of Agriculture, to arrive at our findings. The data is a composite of 2010 Census and 2010-14 American Community Survey.

To help understand the Atlas, we added some contextual information and definitions about food deserts. Additionally, a table of all variables used in our analysis is provided at the end of this section.

- The Atlas was compiled to identify food deserts. Food deserts are defined as parts of the country without accessible and affordable healthy and nutritious foods. The data was constructed to allow visibility into many factors that contribute to deserts, income, distance to grocery stores, and vehicle access.
- States are composed of counties. For the Census, counties are further subdivided into census tracts or just tracts. Tracts are small and relatively permanent statistical subdivisions.
- Group Quarters are residential engagements where a business provides housing as a service to individuals on their property. These living arrangements are flagged, because they frequently provide food services to their residents. Group Quarters maybe far from grocery stores, but not truly have limited access to healthy and affordable foods. (Examples: College Dorm, Homeless Shelter, Correctional Facility, Medical Treatment Facility, etc.)
- The Supplemental Nutrition Assistance Program (SNAP) is more commonly known as food stamps. It is a federal entitlement program that helps low-income families buy healthy foods.
- Poverty rate is the ratio of the number of people by age group, whose income falls below half the median income of the total populate.

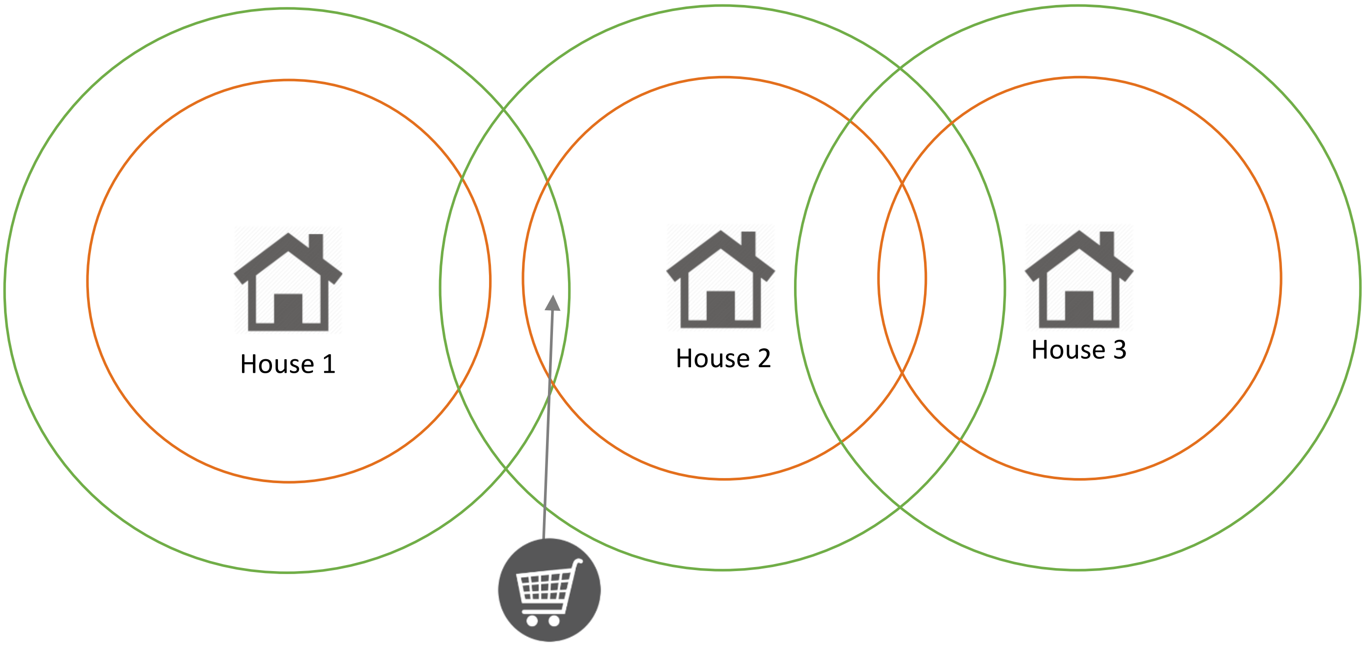
Field	LongName	Description
CensusTract	Census tract	Census tract number
County	County	County name
GroupQuartersFlag	Group quarters, tract with high share	Flag for tract where >=67%
lalowi1share	Low access, low-income population at 1 mile, share	Share of tract population that are low income individuals beyond 1 mile from supermarket
MedianFamilyIncome	Tract median family income	Tract median family income
OHU2010	Housing units, total	Occupied housing unit count from 2010 census
PCTGQTRS	Group quarters, tract population residing in, share	Percent of tract population residing in group quarters
POP2010	Population, tract total	Population count from 2010 census
PovertyRate	Tract poverty rate	Share of the tract population living with income at or below the Federal poverty thresholds for family size
TractAIAN	Tract American Indian and Alaska Native population, number	Total count of American Indian and Alaska Native population in tract
TractAsian	Tract Asian population, number	Total count of Asian population in tract
TractBlack	Tract Black or African American population, number	Total count of Black or African American population in tract
TractHispanic	Tract Hispanic or Latino population, number	Total count of Hispanic or Latino population in tract
TractHUNV	Tract housing units without a vehicle, number	Total count of housing units without a vehicle in tract
TractKids	Tract children age 0-17, number	Total count of children age 0-17 in tract
TractLOWI	Tract low-income population, number	Total count of low-income population in tract
TractNHOPI	Tract Native Hawaiian and Other Pacific Islander population, number	Total count of Native Hawaiian and Other Pacific Islander population in tract
TractOMultir	Tract Other/Multiple race population, number	Total count of Other/Multiple race population in tract
TractSeniors	Tract seniors age 65+, number	Total count of seniors age 65+ in tract
TractSNAP	Tract housing units receiving SNAP benefits, number	Total count of housing units receiving SNAP benefits in tract
TractWhite	Tract White population, number	Total count of White population in tract

Data Preparation

After a cursory look of our data, our team identified that there were both tract and share variables for some indicators. A tract variable is a population count of a certain descriptor, while share is a percentage of the population, simply the tract divided by the tract population. In more simple terms, the share is a derived variable calculated from population and the tract variable. We removed all share variables, because of their redundancy.

Next our group, decided to focus on only urban food deserts. In the data, food deserts were divided into urban and rural and then further subdivided by distance. By focusing on urban, we were able to remove any food desert distances associated with rural deserts, namely predictors focusing on 10 or 20 miles. Furthermore, we were able to strip out some composite flag predictors for urban and rural. An example of a composite flag predictors would be LA1and20, a flag for low access tract at 1 mile for urban areas or 20 miles for rural areas. We already have a flag for low access at 1 mile for urban, LATracts1. Therefore, we can remove LA1and20. We applied this same logic to remove several more predictors and removed them.

Now our data set was left with urban data. During model construction, we determined that our half mile indicators and our mile indicators are not independent variables. For sake of explanation, consider the example illustrated below.



House	Half Mile Food Desert	One Mile Food Desert
1	Y	N
2	N	N
3	Y	Y

‘House 3’ would be counted as both a half mile and one mile food dessert. The deserts are not independent and neither are their predictors given this relationship. Because of their lack of independence, we choose to further reduce our data to only focus on food deserts at a distance of a mile or more in urban areas. We purposefully choose one mile, because it is inclusive of half mile food deserts.

Lastly, we removed one more grouping of variables for lack of independence. As mentioned previously, our target variable is the percentage of the population in a census tract that is in a food dessert. The data additionally contained this percentage broken down by race. These percentages are not independent of each other and lead to autocorrelation, so we removed them.

With these three major reductions, we were able to reduce our data set from one-hundred and forty-three predictors to twenty predictors.

Models: Linear Regression

Regression analysis is the analysis of the relationship between a response or outcome variable and another set of variables. The relationship is expressed through a statistical model equation that predicts a response variable (also called a dependent variable or criterion) from a function of regression variables (also called independent variables, predictors, explanatory variables, factors, or carriers) and parameters.

In a linear regression model the predictor function is linear in the parameters (but not necessarily linear in the regression variables). The parameters are estimated so that a measure of fit is optimized. For example, the equation for the observation might be

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Target variable – (Continuous variable)	Low access, low-income population at 1 mile, share

Since the target variable is a continuous one we went ahead with the Linear regression to get the accuracy of the model. The following observations were obtained upon processing the data in SAS miner. 21% of the variation in the dependent variable has been explained by the model. However, the overall significance of the model is very high since the value of ‘F’ is very significant

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	16	23.283184	1.455199	70.42	<.0001
Error	4078	84.270117	0.020665		
Corrected Total	4094	107.553301			

Model Fit Statistics

R-Square	0.2165	Adj R-Sq	0.2134
AIC	-15868.9107	BIC	-15866.7690
SBC	-15761.5128	C(p)	17.0000

The variables which are highly significant are as follows

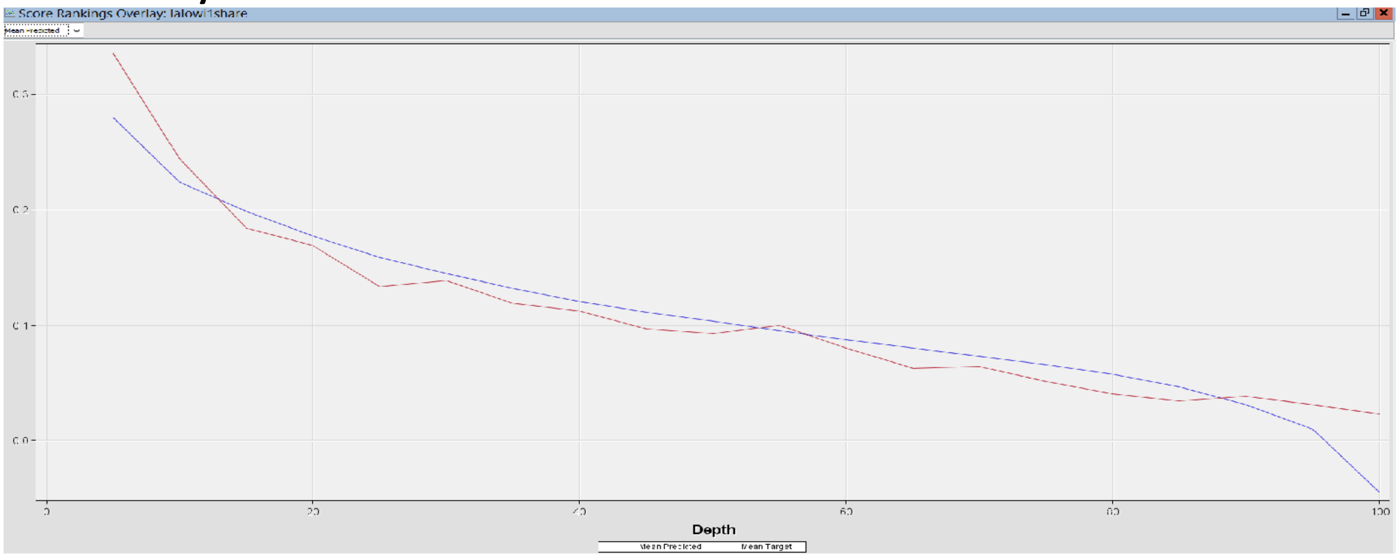
HUNVFlag, MedianFamilyIncome, POP2010, PovertyRate, pOHU2010, ptractAIAN, ptractAsian, ptractBlack,pTractNHOPI, pTractSNAP, pTractSeniors and pTractWhite

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.2031	0.0483	4.20	<.0001
HUNVFlag	1	0.0993	0.00637	15.58	<.0001
MedianFamilyIncome	1	-5.24E-7	1.035E-7	-5.06	<.0001
PCTGQTRS	1	0.0636	0.0471	1.35	0.1769
POP2010	1	-2.43E-6	1.049E-6	-2.32	0.0205
PovertyRate	1	0.00194	0.000290	6.70	<.0001
pOHU2010	1	-0.1588	0.0666	-2.39	0.0171
pTractAIAN	1	1.7305	0.5293	3.27	0.0011
pTractAsian	1	-0.2603	0.0418	-6.23	<.0001
pTractBlack	1	0.0476	0.0180	2.64	0.0082
pTractHUNV	1	-1.7440	0.1415	-12.32	<.0001
pTractHispanic	1	-0.0155	0.0154	-1.01	0.3145
pTractKids	1	0.1090	0.0726	1.50	0.1332
pTractNHOPI	1	2.1459	0.7624	2.81	0.0049
pTractOMultir	1	-0.3852	0.0360	-10.69	<.0001
pTractSNAP	1	0.5056	0.1198	4.22	<.0001
pTractSeniors	1	-0.1317	0.0483	-2.73	0.0064
pTractWhite	0	0	.	.	.

The T test of this model clearly shows that there are 13 significant variables from the input variables.

Score overlay

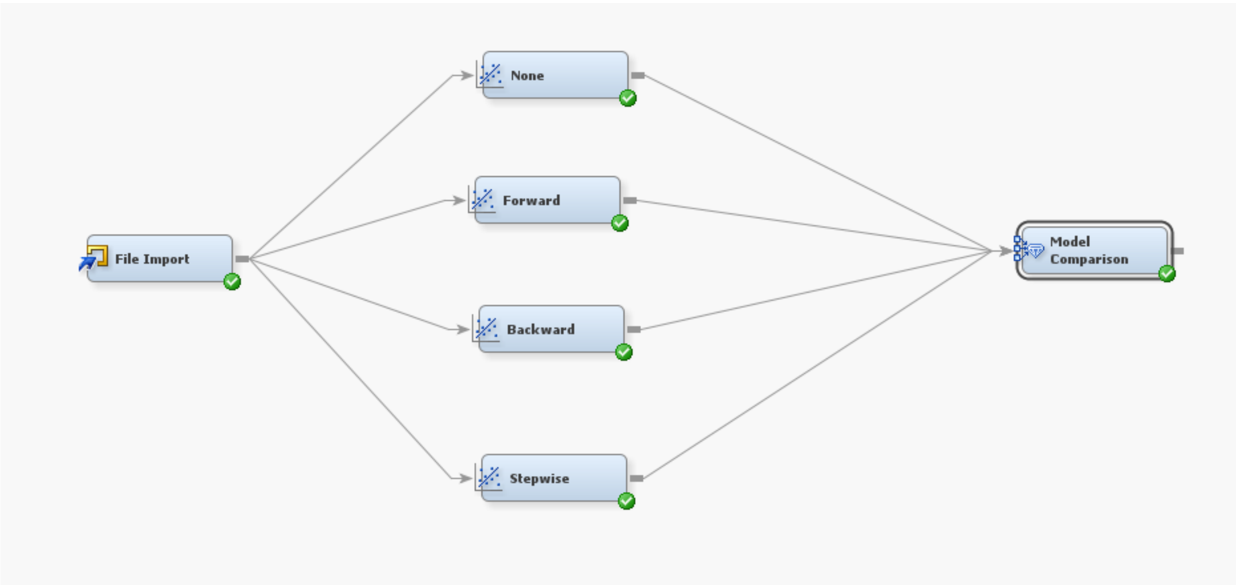


Fit statistics

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
lalowi1share	lalowi1share	AIC	Akaike's Information Criterion	-15868.9	.	
lalowi1share	lalowi1share	ASE	Average Squared Error	0.020579	.	
lalowi1share	lalowi1share	AVERR	Average Error Function	0.020579	.	
lalowi1share	lalowi1share	DFE	Degrees of Freedom for Error	4078	.	
lalowi1share	lalowi1share	DFM	Model Degrees of Freedom	17	.	
lalowi1share	lalowi1share	DFT	Total Degrees of Freedom	4095	.	
lalowi1share	lalowi1share	DIV	Divisor for ASE	4095	.	
lalowi1share	lalowi1share	ERR	Error Function	84.27012	.	
lalowi1share	lalowi1share	FPE	Final Prediction Error	0.02075	.	
lalowi1share	lalowi1share	MAX	Maximum Absolute Error	0.73498	.	
lalowi1share	lalowi1share	MSE	Mean Square Error	0.020665	.	
lalowi1share	lalowi1share	NOBS	Sum of Frequencies	4095	.	
lalowi1share	lalowi1share	NW	Number of Estimate Weights	17	.	
lalowi1share	lalowi1share	RASE	Root Average Sum of Squares	0.143453	.	
lalowi1share	lalowi1share	RFPE	Root Final Prediction Error	0.14405	.	
lalowi1share	lalowi1share	RMSE	Root Mean Squared Error	0.143752	.	
lalowi1share	lalowi1share	SBC	Schwarz's Bayesian Criterion	-15761.5	.	
lalowi1share	lalowi1share	SSE	Sum of Squared Errors	84.27012	.	
lalowi1share	lalowi1share	SUMW	Sum of Case Weights Times Freq	4095	.	

The Average square error of the Regular model is 0.020579 and the mean square is 0.020665

On comparing the various variants of Linear regression



We observe the following output on comparison of the various regression models

Fit Statistics
Model Selection based on Train: Average Squared Error (_ASE_)

Selected Model	Model Node	Model Description	Train: Average Squared Error
Y	Reg	None	0.020579
	Reg3	Backward	0.020598
	Reg2	Forward	0.020600
	Reg4	Stepwise	0.020600

The default regular model of Linear regression has the least Average Squared Error.

Models: Cluster Analysis

Data Preparation Details

- *Only the data for Texas and urban environments are utilized to reduce the sample size.*
- *Most binary indicators were removed as they are derived from other interval indicators within the dataset.*
- *There is not a target variable identified within clustering models. The target which was utilized within the decision tree and regression models, "Share of tract population that are low income individuals beyond 1 mile from supermarket" was utilized as an input variable. This was identified as an interval variable.*
- *The indicators used for input represent income, housing status, and ethnic makeup of the tracts. These are farther divided into more specific indicators.*
- *NORMALIZATION: The second set of the Cluster Analysis was run after making a significant change in the input data. The absolute counts of ethnic groups in tracts does not tell the whole story of the racial makeup. It is the proportion of ethnic groups that really gives an idea of how each group's presence affects food access. While there are indicators available that provide proportion, these variables only account for those people who also suffer from low food access; these indicators represent compound attributes. Therefore, each ethnicity count variable, count of children variable, and the variable for "Occupied Housing Units" was divided by the total tract population. While not perfect, this results in a more normalized input data-set which is adjusted for varying population sizes per tract.*

Initial Cluster Analysis

The clustering data model has the ability to determine the similarity between a collection of data objects. In this dataset, clustering would be able to identify which variables are typically clustered together when speaking about food deserts. Initially, the normalized data set was run through the clustering model with an automatic specified number of clusters. However, it was realized that the clustering model within SAS automatically normalizes the variables, therefore the dataset containing the actual population numbers, instead of percentages was utilized.

With automatic selection of clusters enabled, SAS identified 20 clusters, which also happens to be the default maximum, through its initial run within the dataset. However, many of these clusters contained a “frequency of cluster” of 1, meaning there was only 1 census tract contained within that cluster. After identifying the maximum number of clusters was reached, the model was modified to see SAS Enterprise Miner would identify more clusters if the maximum was raised. The model was changed to have a “final maximum” set to 50.

After another automatic selection of clusters, SAS identified 50 clusters, the new maximum amount allowed. This shows that the dataset is very dissimilar, and that many urban census tracts in Texas may not be like the others. The further SAS was allowed to identify similarities between data objects (by increasing allowed maximum of clusters), the more dissimilar the clusters became.

Through many trial and errors runs in the clustering model, 8 clusters were identified to best represent the population. Initially the process started with the minimum number of clusters, 2, and the automatically selected amount of 20. Over 20 clusters were never tried manually since it was likely that creating more than 20 clusters would create many clusters with 1 urban tract in them. With 2 clusters, the main dissimilarity between the segments were the median family income. More and more models were tried until 8 clusters seemed to be the appropriate amount.

The 8 segments that SAS enterprise miner returned all had varying frequencies. The frequencies ranged from 5-1220. The 8 clusters contained 4 large clusters, and 4 small clusters. To better summarize the data, a chart was created below. Names for each cluster were also generated. Additionally, please see the appendix for the segment profile.

Analysis of the Initial Dataset

Segment ID (Pop Size)	**Frequency of Cluster**	**Median Income**	**PCTG QTRS**	**Poverty Rate**	**Tract Asian**	**Tract Black**	**Tract White**	**Food Desert %**	**Name**
1 (6571)	23	Medium	Low	Medium	Medium High	High	Middle	25%	Average Income Food Desert
2 (2821)	1178	Medium	Low	Low	Medium Low	Medium Low	Middle	11%	Small Tract with Plentiful Food
3 (5627)	5	Very Low	Low	Very High	Low	Medium Low	Middle	31%	Very Poor Tract
4 (7686)	542	Medium	Low	Medium	Medium	Medium	Middle	15%	Average Urban Texas Tract
5 (4661)	49	Medium	Very High	Medium	Medium	Medium	Middle	20%	Group Quarters Segment
6 (14246)	81	High	Low	Very Low	Very High	High	High	8%	Large Wealthy Tracts
7 (5157)	1220	High	Low	Very Low	Medium High	Low	Middle	5%	Small Wealthy Tracts
8 (4709)	997	Low	Low	Very High	Medium Low	Medium	Middle	12%	Poor Tracts

Cluster Analysis using the Normalized Dataset

SAS Enterprise Miner returned a total of 6 clusters, all with varying frequencies. The frequencies ranged from 4-1986. There were 2 very large clusters (Segment 2 &3), 1 averaged sized cluster (Segment 5), and 3 small clusters (Segment 1, 4, & 6). Segment 1's main deviation from the rest of the clusters was that it contained a high "PCTGQTRS" rating. This is the percentage of people living in group quarters in that census tract. Segment 1 scored a percentage of 67% of group quarters while all the other segments did not even reach 2%.

Segment 2's main deviation from the rest of the clusters is the median family income. Segment 2 averaged a median family income of over $90,000, while the next highest was $48,000. Segment 2 additionally also contained the highest average white census tract at 76% white. Segment 3 contains very similar basic demographics as segment 2, except with a significantly lower median family income.

Segment 4 only contained a segment of 4 census tracts, and it's main deviation was that it was very high in the Asian tract, 13%, with the next highest being 6%. Segment 5 contained the same income and poverty rate as segment 3, except its main demographic was black. Over 50% of this segment's population was black compared to the next highest of 20%. Segment 6 only contained 6 census tracts, with the main similarity between those tracts being that it is a poor urban tract.

To better summarize the data, a chart was created below. Names for each cluster were also generated. Additionally, please see the appendix for the segment profile.

Analysis of the Normalized Dataset

Segment ID	Frequency of Cluster	Median Income	PCTG QTRS	Poverty Rate	Tract Asian	Tract Black	Tract White	Food Desert %	Name
1	55	Medium	High	Middle	Middle	Middle	Middle	20%	Group Quarters Segment Desert
2	1986	High	Low	Low	Middle	Low	Very High	7%	Wealthy White Tract
3	1531	Medium	Low	High	Low	Low	High	14%	White Food Desert
4	4	Medium Low	Low	Middle	High	Low	Middle	16%	Asian Food Deserts
5	513	Medium Low	Low	High	Low	Very High	Low	16%	Black Food Deserts
6	6	Low	Low	High	Low	Low	Middle	25%	Poor Food Desert

The normalized dataset showed more of a dissimilarity within the ethnicities of the regions than what the initial cluster analysis showed. This could be since the values were normalized, it was easier to tell if the proportions were different.

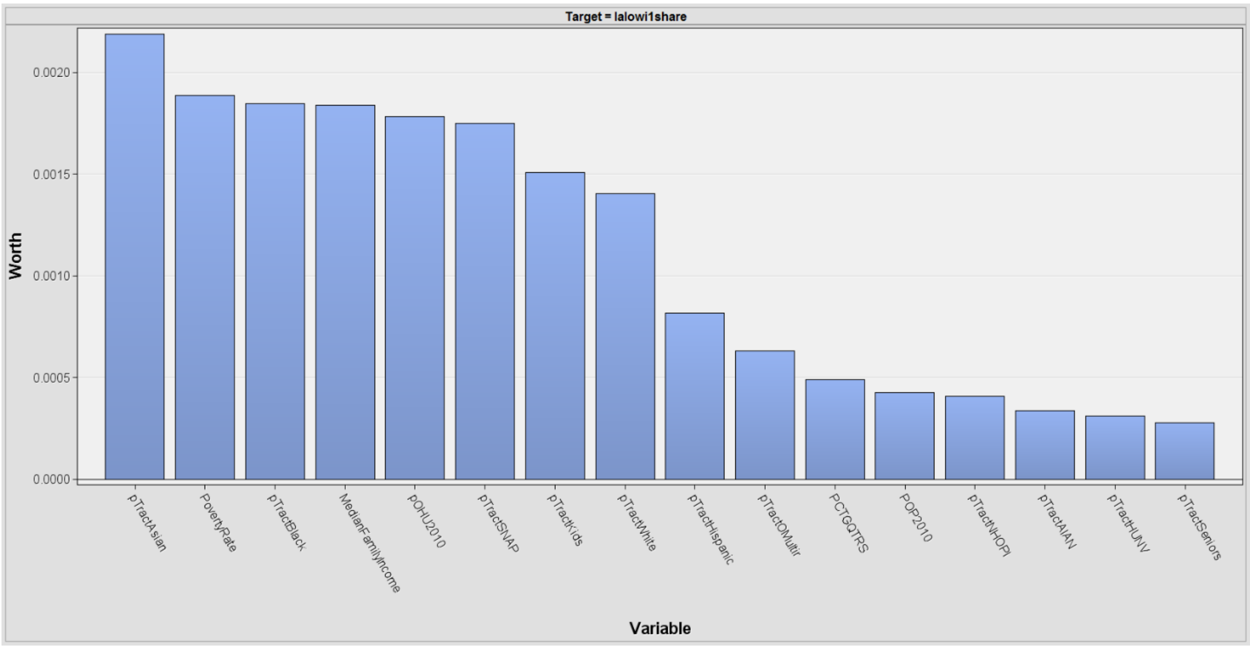
Models: Decision Trees

Data Preparation Details

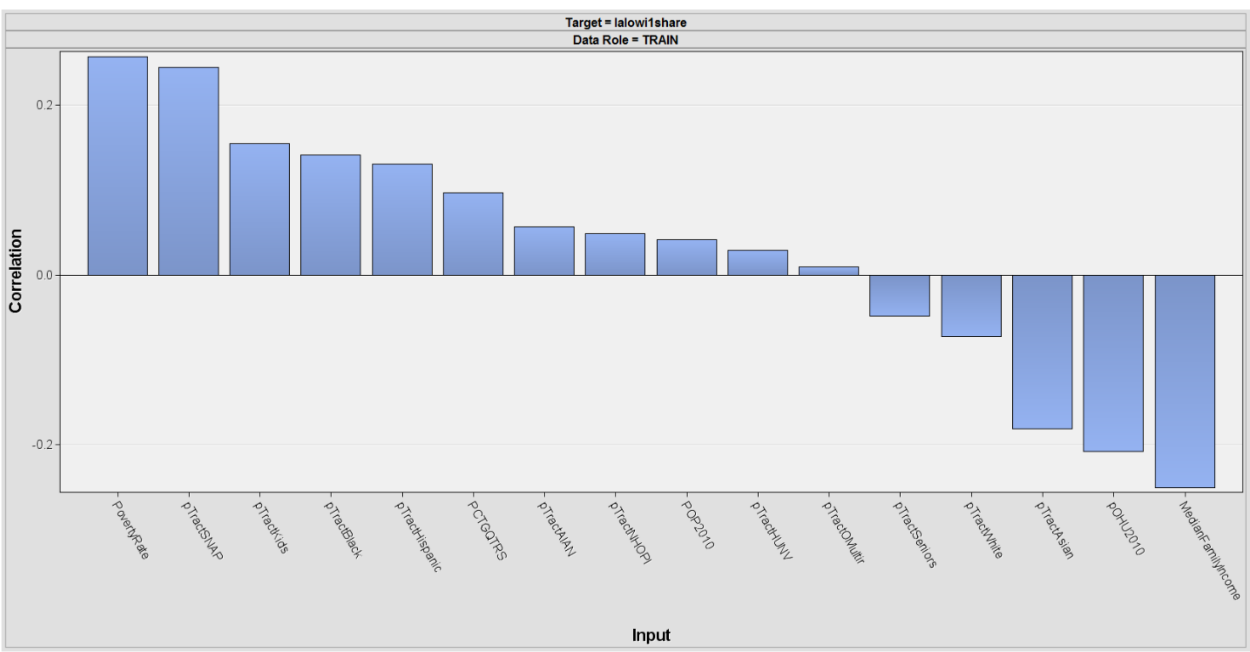
- *Only the data for Texas and urban environments are utilized to reduce the sample size.*
- *Most binary indicators were removed as they are derived from other interval indicators within the dataset.*
- *The target is set to the variable representing ”Share of tract population that are low income individuals beyond 1 mile from supermarket” – this is an interval variable; we did not want to utilize the food desert flag variables because whether or not an area is a food desert is based on the interpretation of the researchers for those variables.*
- *The indicators used for input represent income, housing status, and ethnic makeup of the tracts. These are farther divided into more specific indicators.*
- *NORMALIZATION: The second set of Decision Tree was run after making a significant change in the input data. The absolute counts of ethnic groups in tracts does not tell the whole story of the racial makeup. It is the proportion of ethnic groups that really gives an idea of how each group's presence affects food access. While there are indicators available that provide proportion, these variables only account for those people who also suffer from low food access; these indicators represent compound attributes. Therefore, each ethnicity count variable, count of children variable, and the variable for "Occupied Housing Units" was divided by the total tract population. While not perfect, this results in a more normalized input data-set which is adjusted for varying population sizes per tract.*

Initial Indicator Analysis

Before modeling the data using decision trees, the indicators were analyzed for importance and correlation with the Target data. In the below plots, the Normalized-population count data (second dataset) was utilized.



The variables representing Asian and black ethnicity, poverty rate, and median family income appeared to be the most important indicators of food access as expressed in the above chart. Housing, population, and age-related variables appeared to be less impactful.

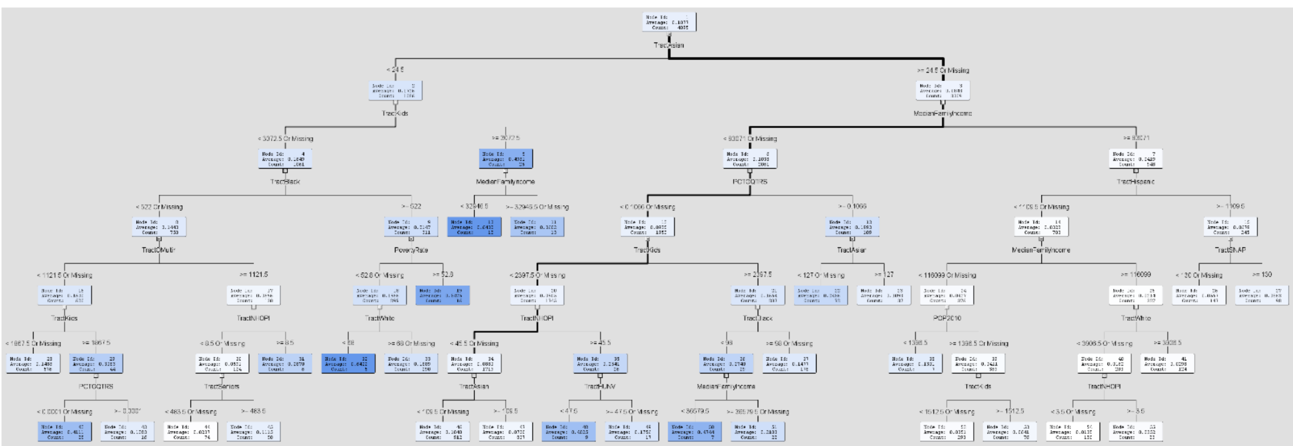


The chart immediately above displays the strength of correlation *and* direction of each variable as it influences the target. Greater poverty rates, larger numbers of

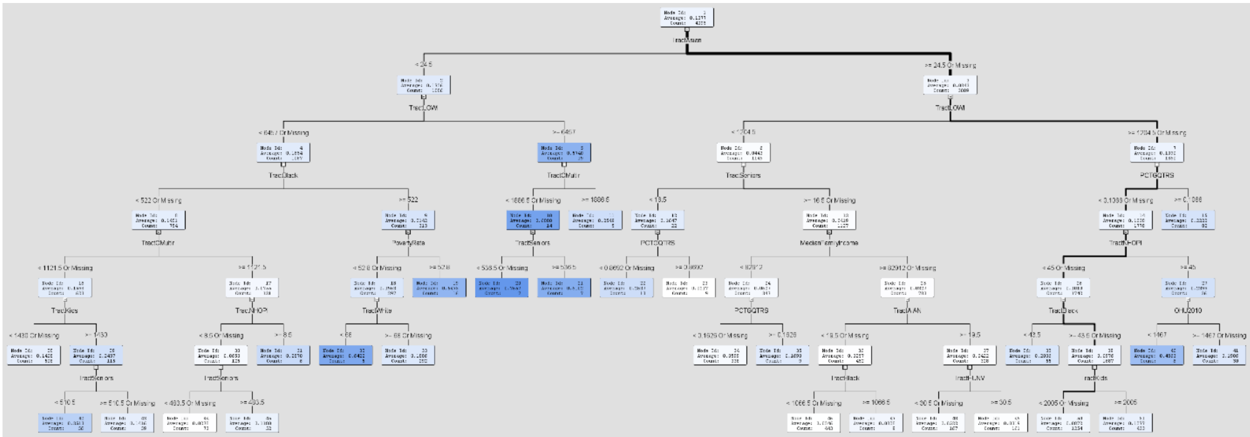
SNAP beneficiaries, and greater child population proportions all increase the likelihood of poor food access, as evidenced by the large positive magnitude of the corresponding chart bars. In contrast, greater median family income, housing unit occupancy, and Asian population makeup all correlate with greater food access, which is seen in the strong negative magnitude for these values.

Decision Trees Using Population Segment Counts

The first set of decision trees was run utilizing the unmodified dataset containing absolutely population counts. As seen immediately below, this resulted in a very complex tree containing 54 leaves. While certain conclusions could be drawn from this diagram, a simpler tree was desired.



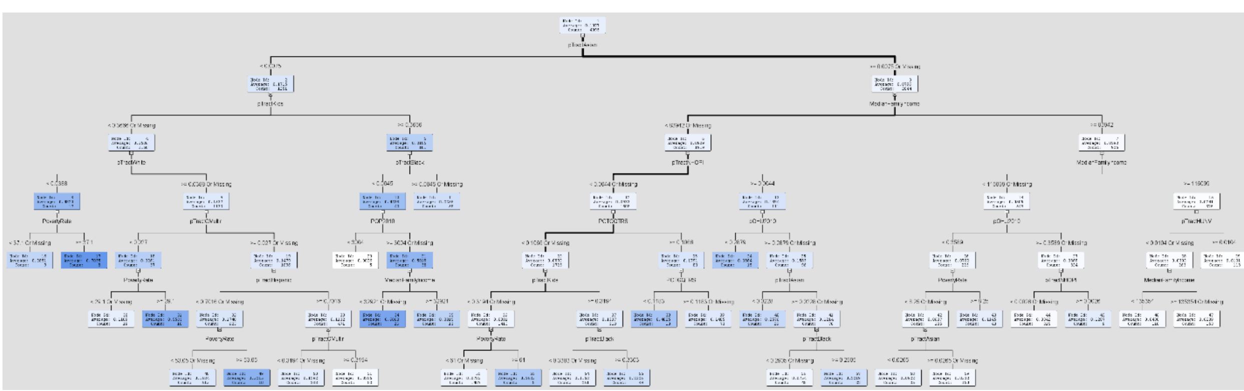
The decision tree below was developed using the same dataset as the tree above but with one key change. Each input was only permitted to be utilized once; thus, a certain ethnic variable, for example, could not appear multiple times down the same branch of the tree. The resulting simplified tree, shown below, contains a slightly more reasonable 50 leaves.



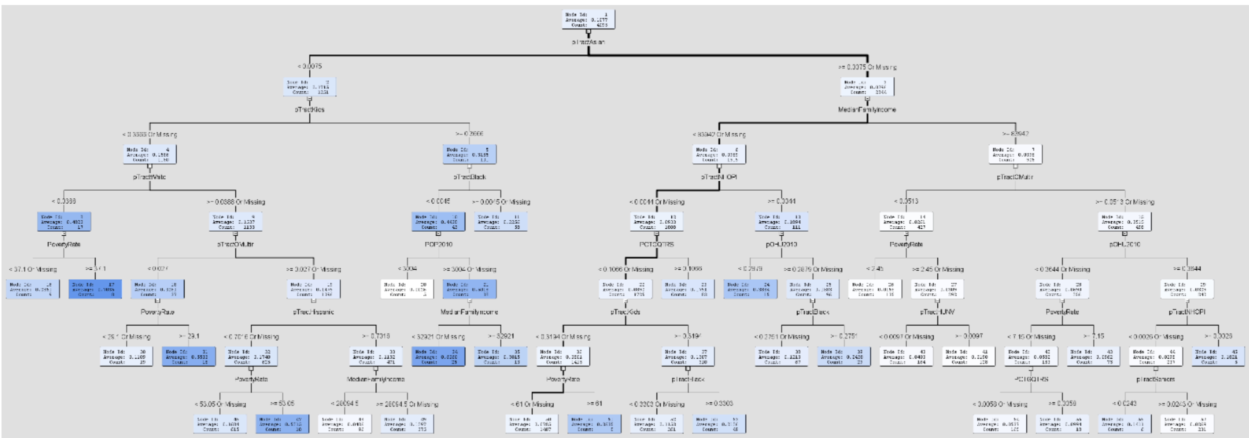
With either approach, the most important variable splitting the top of the decision trees appears to be the count of people of “Asian” ethnicity. Tracts with greater Asian populations contained far less people with poor access to food. The two preceding models had significant SSE values of 82.5541 and 83.137. The model with fewer leaves thus exhibited a slightly greater error.

Decision Trees Using Normalized Population Segment Proportions

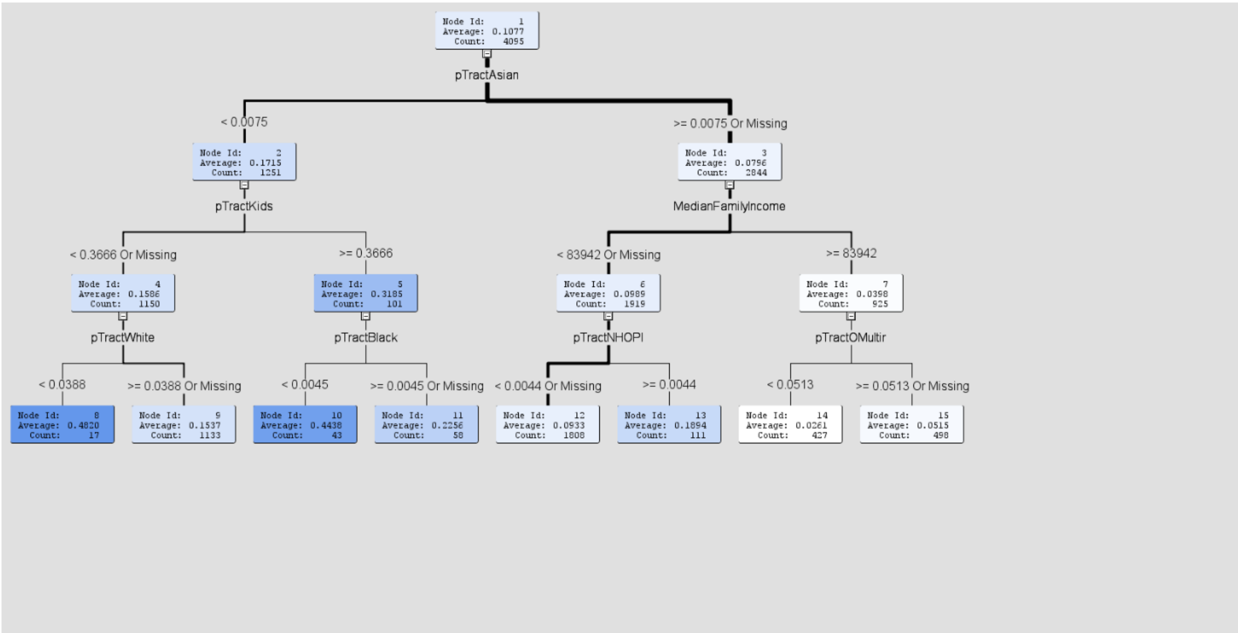
As described above, normalized data theoretically makes different tracts of different sizes more comparable. The same kind of modeling was run with ethnic population, SNAP, and Housing unit count variables normalized over the total tract populations. The first iteration is immediately below.



Once again, the most important variable appears to be Asian population in the tract. However, the next more important variables relate both to income and number of children present in the tracts. Regardless, this tree contains 58 leaves, larger than the trees generated in the first round of analysis with non-normalized data. At the same time, the SSE is improved down to 78.522. Generating a subsequent tree with the restriction of utilizing each indicator once per branch results in a slightly simpler tree with 56 leaves but with a bit more error at an SSE value of 79.440 (seen right below). This tree is still too complicated and almost certainly overfitting of the training data.



In a final attempt to farther prune the decision tree and reduce overfitting, the maximum depth of the tree was restricted to 3 levels. As displayed below, the resulting tree mainly favors Asian ethnic makeup, median income, child population, and black population. It is far simpler and highlights the most important factors determining which areas have more and less access to food. Despite the simplicity, this decision tree suffers the greatest error by far at an SSE of 91.579.



While the latest tree harbors the most error, the team believes that a simpler model is far more interpretable and far more practical to utilize for projections. It also highlights the striking finding that Asian ethnic makeup, or lack thereof, is the biggest indicator for poor food access. Tracts with less than 0.75% Asian population seem to suffer the most. Beyond that, the model also shows that a large number of children per capita is also a strong indicator. Tracts with children making up at least 36.7% of the population are home to twice as many low-access individuals. Low median income is also another driver of poor food access based on the tree results. Farther more, lower white population makeup concurrent with a smaller child population makeup and lower black population makeup in the presence of larger child population makeup also favor poorer food access

Conclusion

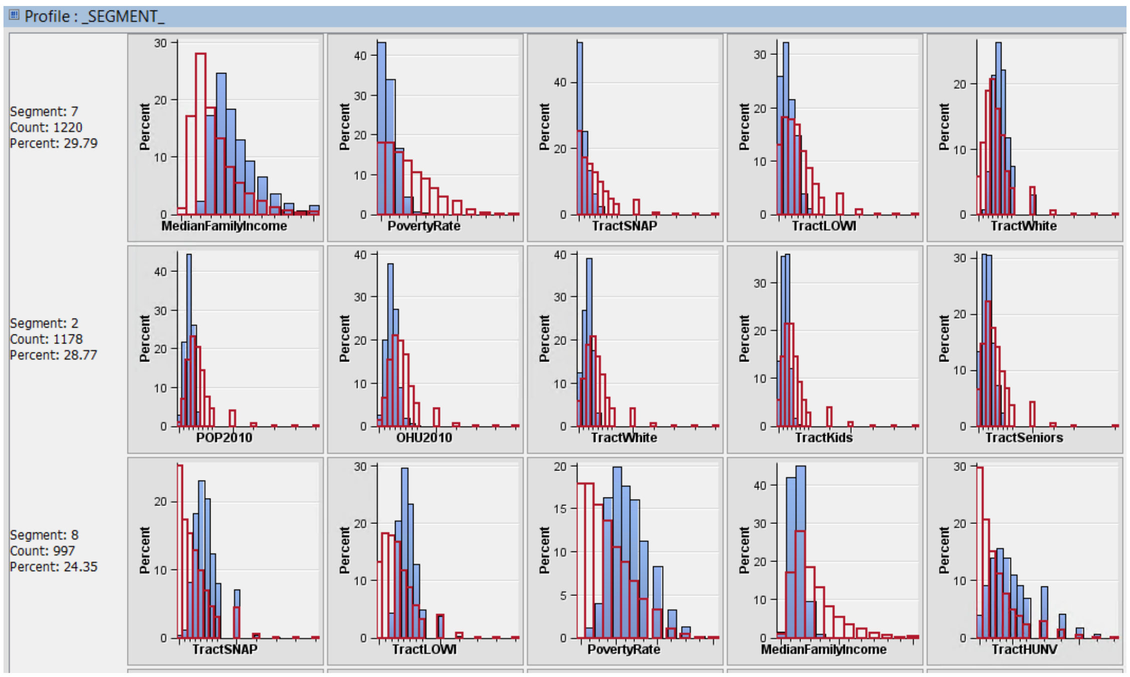
Approximately 17% of the Urban Texas population reside in a food desert. These food deserts can contain many varying factors, some predominately white, some predominately black, but the effects that exist are detrimental throughout. It has been proven that the relationship between health, learning, and nutrition is very strongly correlated. Nutrition is one of the major factors that can affect a child's health. Having poor nutrition or under nutrition can lead to poor brain development which can lead to irreversible effects.

The findings within the analysis could be utilized to identify current deserts and could lead to legislation that helps these disenfranchised areas. Also, they could be utilized to target urban tracts which are at risk of becoming a food desert. By analyzing the data and trends, these future food deserts can be recognized before they are ever established.

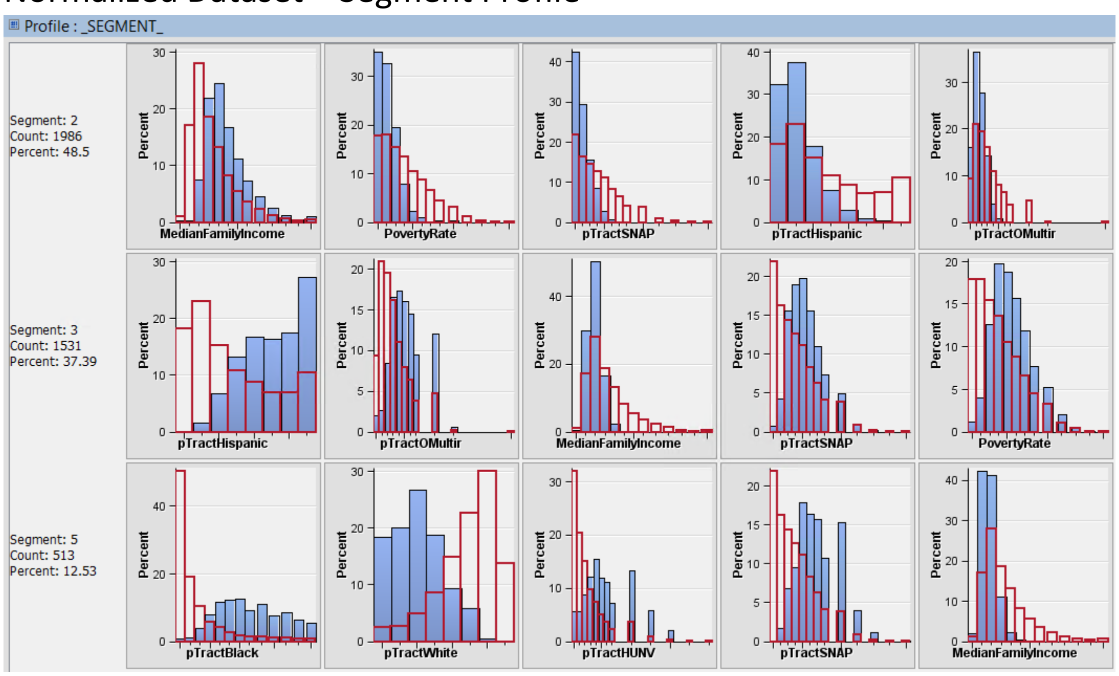
Additionally, by identifying which areas of urban tracts are food deserts, city and county officials would have the ability to incentivize grocery stores in certain key areas to reduce the overall risks of food deserts.

Appendix

Initial Cluster Analysis – Segment Profile



Normalized Dataset – Segment Profile



References

1. https://novakdjokovicfoundation.org/importance-nutrition-early-childhood-development/
2. https://news.aetna.com/2017/04/poor-nutrition-tied-to-nearly-half-of-deaths-from-heart-disease-diabetes-stroke/
3. http://thefoodtrust.org/uploads/media_items/access-to-healthy-food.original.pdf