The Pennsylvania State University

The Graduate School

Department of Industrial and Manufacturing Engineering

SPATIAL ACCESSIBILITY IN FOOD DESERTS

A Thesis in

Industrial Engineering

by

Varun Sudhurshun Chellappa

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

December 2016

The thesis of Varun Sudhurshun Chellappa was reviewed and approved* by the following:

Soundar R.T. Kumara

Allen E. Pearce and Allen M. Pearce Professor of Industrial and Manufacturing Engineering

Thesis Advisor

Conrad S. Tucker

Assistant Professor of Industrial and Manufacturing Engineering

Thesis Reader

Janis P. Terpenny

Peter and Angela Dal Pezzo Professor

Head of the Department of the Industrial and Manufacturing Engineering

*Signatures are on file in the Graduate School

ABSTRACT

In recent times, food insecurity has become a pervasive topic in the United States of America. Previous research has indicated the prevalence of food insecurity is high amongst low-income neighborhoods that have poor access or cannot afford enough food in order to avoid hunger, also known as food deserts, and is currently rising amongst children below the age of 18. This is a major concern for non-profit and charitable organizations like food banks, whose main objective is to identify and provide food to people in need. This thesis considers the case study of the Central Pennsylvania Food Bank for the purpose of identifying the food deserts in need of the food bank's resources by using different data analysis and visualization techniques on multiple datasets. Currently, the Central Pennsylvania Food Bank operates based on the demand identified by the food agencies present in the different counties, but do not consider the possible factors contributing to the demand. The key factors contributing to the increase in food insecurity are identified and the relationship between these factors are analyzed in this thesis. Furthermore, an implementation plan consisting of the recommendations and future work to be done are also detailed to help the Central Pennsylvania Food Bank target their customers in amore effective way.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would take this opportunity to thank everyone who has helped me in successfully completing this thesis.

I would like to thank Dr. Soundar R.T. Kumara, Professor at The Department of Industrial and Manufacturing Engineering, for his constant encouragement, valuable guidance and great insights during the course of my Master's Thesis.

I would like to express my gratitude to Dr. Conrad S. Tucker, Assistant Professor of Industrial and Manufacturing Engineering, for agreeing to be the reader for this thesis and providing his valuable insights.

I also thank Ms. Susan Steadman and her team at the Central Pennsylvania Food Bank for providing me with the data and information required to make this thesis possible.

Finally, I express my profound gratitude to my parents Chellappa Venkatachari and Jayalakshmi Chellappa for their continuous encouragement and unfailing support throughout my graduate studies.

Chapter 1

Introduction

The concept of Food Insecurity has been defined various ways. In general, we define food insecurity as the state of being without reliable access to sufficient quantity of affordable and nutritious food [1]. U.S. Department of Agriculture (USDA) defines food insecurity as “the household-level economic and social condition of limited or uncertain access to adequate food” [2]. Food Insecurity has been a growing concern over the past few years, due to the increased dependency on fast food. This tendency has led to people having low access to nutritious and fresh food over the past decade. The implications of this change is becoming clearer in the current generation in the form of increased obesity, diabetes and other heart related issues in much younger adults. In 2014, nearly 48.1 million Americans that lived in households with poor access to affordable and nutritous food, including 32.8 million adults and 15.3 million children. Households with children reported higher food insecurity (19%) than those without children (12%) [3].

Food insecurity has become a concept that is not limited to rural areas. This is largely due to the decrease in accessibility to fresh and healthy foods in urban areas, which increases the risk of the urban population having poor nutrition, which in-turn puts them at risk of chronic diseases like obesity, heart-related disorders and type 2 diabetes [4,5]. The data provided by USDA and Feeding America suggests a strong relationship between poverty, hunger and food habits. Poverty leads to people tending towards cheaper and unhealthy food, which may lead to them becoming overweight, obese or both. Overweight and obesity are both risk factors for type-2 diabetes, heart disease and high blood pressure [38]. There are a number of food banks and food pantries across the country that cater to the needs of the food insecure population. Food banks coordinate with the local agencies to identify these people and help them [3].

Feeding America defines Meal Gap as “a conversion of the total annual food budget shortfall in a specified area divided by the weighted cost per meal in that area”. Meal Gap is used a standard measure to determine the food shortfall in terms of number of meals at the county level. The Additional number of meals required is defined as “the remaining number of meals needed to satisfy the meal gap after the completed efforts of the food bank” [6]. The main objectives of these food banks are aligned towards reducing the meal gap that is prevalent in the country.

The main goal of this thesis is to understand how the Central Pennsylvania Food Bank helps the people living in various counties in Central PA bridge the meal gaps and suggest possible ways to decrease the meal gap. The case study of Central Pennsylvania is used to explore the concept food insecurity in depth and how this problem could be approached using a more efficient Food Supply Chain System. The thesis is only limited to the Central Pennsylvania Food Bank due to the availability of demand data within their respective counties.

1.1 Motivation

The growth of the food insecurity in the country has led to Food Banks to design new ways and coordinate food distribution under Feeding America [3] and other independent agencies. Previous research work related to evolution of food deserts have indicated that, over a period of time, neighborhoods with lower socio-economic status have the poorest access to fresh food supply. This indicated the existence of spatial inequalities in major cities. Studies have shown that the fresh food markets are more likely to be set up in localities, which are more profitable [7].

Currently, the Central Pennsylvania Food Bank makes an effort to satisfy hunger in 27 counties within Pennsylvania (Figure 1-1). The Food Bank has two warehouses, one in Harrisburg and the other one in Williamsport. These warehouses provide the food required to the people through local agencies and food pantries within the independent counties [26]. Since, the number of agencies in some counties are not sufficient to cater to the needs of the people in hunger; there exists an inequality in distribution of food all across the 27 counties.

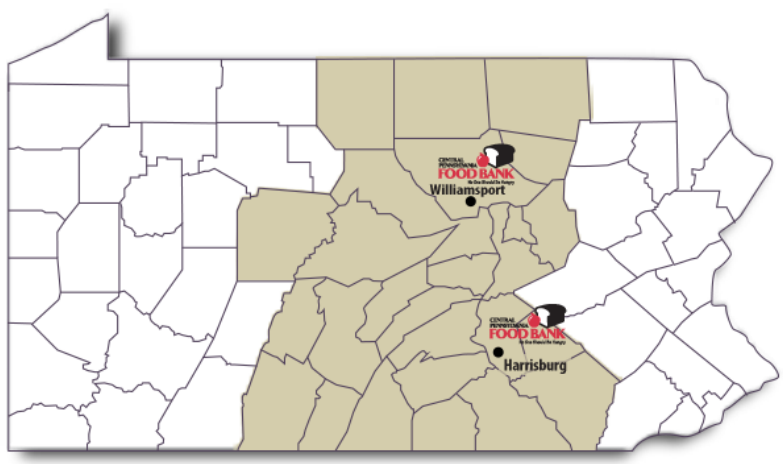

Figure 1-1. Central Pennsylvania Census Tracts with Locations of Food Banks

Davis [9] has shed some light on this problem by identifying satellite locations to plant Food Delivery Points (FDPs), based on the frequency of the agencies in each county.

Food Delivery Points are where the partner agencies can receive deliveries. In this case, a set covering model was used in order to assign agencies to respective FDPs.

Several similar research efforts addressed the problem the same way as Davis (2014) [9], but most of them did not address the granularity (level of detail present in data) of the problem. The past research work has only accounted for the distance of the population from food agencies and demand at the food agencies, but never accounted for the population demographics that considers the number of low-access and low income individuals at the Census tract level. This helped us in establishing a correlation between the food insecurity and the socio-economic status of the people in these regions. In this thesis, we address the issue at a more granular level. This research makes an effort to understand the effect of socio-economic and demographic factors on food insecurity at the Census Tract level. U.S. Census Bureau defines Census Tract as “small, relatively permanent statistical subdivisions of a county or equivalent entity that are updated by local participants prior to each decennial census as part of the Census Bureau's Participant Statistical Areas Program” [8]. Census tracts are further divided into Block groups [8].

1.2 Problem Definition

The food banks are primarily non-profit organizations that aim to provide relief in the form of food to people affected by food insecurity. Food banks collect, store and distribute food to charitable agencies. These agencies operate under different food distribution programs (Eldershare, SNAP, USDA etc.) [26] and cater to different strata of the society.

The Central Pennsylvania Food Bank [26] collected the demand data from the food agencies and aggregated the Meal Gap data at the county level, based on where the food agency is located and have also provided supply data at the agency level. This led us to question the method used to distribute the available resources. The data was not successful in justifying the prioritization of one county over the other, which leads us to question the equity of distribution.

The Central Pennsylvania Food Bank [26] currently employs a strategy targeting the population that is adversely affected by food insecurity, based on the definition of the USDA's research. USDA defines these sections of the population as "low access and low income" [2]. This is measured at 1 mile for urban areas and at 10 miles for rural areas. Since the concept of food insecurity is no longer applicable to rural areas alone and now includes the accessibility to fresh food supply, several other factors need to be considered to measure accessibility in addition to just the distance.

We aim to identify the factors affecting the Census Tracts [8] classified as food insecure based on the demand in each region, its geolocation (distance to the agency) and their respective demographics using different approaches such as clustering, descriptive analytics, visual analytics and spatial analytics aimed at minimizing the overall Meal Gap. We use these variables mentioned above due to its availability in the same frequencies.

1.3 Brief Description of Methodology

The background literature on Food Deserts, Food Banks and Spatial Inequality were explored to understand these research efforts. We identified that most of the studies only considered the spatial accessibility in terms of distance alone and are very nascent.

The past research on food deserts [7,9,14,18] has not considered the impact of demographic features at a more granular level i.e. Census Tract and Block Group level. The method used in the previous works neither accounts for capacity at the nearest food agency nor does it account for the median household income in each of the regions. In this study, we use different approaches (clustering, descriptive, visual and spatial analytics) to study the data provided by the Central Pennsylvania Food Bank separately and in relation to the data obtained from the Census Bureau on Census Tracts' demographics. These approaches combine all the available data in order to draw more accurate conclusions about demand data at the different counties and understand their importance.

The analyses in this research will help us in prioritizing the counties, the tracts and block groups within them for the purpose of equitable distribution of resources by the Central Pennsylvania Food Bank. The various datasets used for the study, their sources and description are listed in Table 1-1. These were the only datasets used mainly due to their availability.

Table 1-1. Datasets used in the study

Dataset Name	**Definition**
Meal Gap data	The demand in each of the counties are aggregated
Census Bureau data	The demographics data on the County, Census Tract and Block Group level
Tiger Line Shapefile data	The geographical data spatial polygons for the purpose of mapping are provided
Food Desert data	Regions affected by Food Insecurity and classification of population

Exploratory data analyses were performed on each of the datasets through descriptive analytics and data visualization on maps. Preliminary analysis on the Meal Gap data suggested that four counties (Dauphin, York, Lancaster and Lebanon) were the most affected. These four counties were used to measure accessibility and study the use of k-means clustering (of the block groups) in identifying consolidation points (centroid of the cluster) through which food would be distributed to the specific regions within these counties that are in need.

k-means clustering is used as it is one of the more simple algorithms which uses unsupervised learning. For the purpose of k-means cluster analysis, demand (aggregated at the county level) and the distance of a block group to the nearest agency were used as primary attributes. Further analysis was performed to determine the ideal number of clusters.

The data from USDA Food Access research [2] was explored to understand its relationship to the Meal Gap data. Demand-Supply analysis of the Food Bank identified the counties with the highest requirement of resource from the Food Bank (based on additional amount of food (in lbs.) required per year). The additional number of meals per year is defined as the amount of food that is required per year in addition to what resources has already been provided by the food bank per year. The Supply data provided by the food bank was further used to identify the various types of food, distribution programs and the delivery patterns followed by the food bank. The purpose of this analysis is to determine the amount of fresh food being distributed to the food agencies by the food bank. The data on demographics, demand of food, and accessibility to the food agencies were aggregated at the Census tract level to perform some spatial analysis on the county and tract level in order to identify actual driving distances from the Food Bank and its respective agencies used to distribute food.

The agencies under the Food Bank are classified on the basis of the county in which they are present and accessibility to food is measured only on this basis. The data sourced from the Census bureau [29] is further used to develop a methodology to prioritize each of the counties based on more significant factors. These factors include the number of agencies in the Census tract, the median household income and the number of people with low access to the food resources, as defined by USDA [2].

1.4 Contributions of the Thesis

Different methodologies were used to approach the problem and provide a solution more applicable. The k-means clustering methodology provided a solution obtained by brute force of using the meal gap, population within a census block group and the distance of a block-group from the nearest food agency. The descriptive analysis of demographics and the distribution of resources by the Central Pennsylvania Food Bank details the type of food distributed to majority of the region affected and the amount of food (in lbs.) supplied to each of these counties. The visual and spatial analysis helped understand the impact of each possible contributing factor i.e. household income, accessibility etc. on the demand at the Census tract level. This work lays the foundation for integrating socio economic, meal gap and health data to build integrated systems which can help less fortunate in the USA.

1.5 Organization of Thesis

This thesis contains seven chapters. Chapter 2 describes the previous work and research gaps. Data collection and Data description is detailed in chapter 3. Chapter 4 includes the

Methodology used. Chapter 5 is devoted to Data Analysis. Chapter 6 contains results from the research work. Chapter 7 describes the Implementation Plan and recommendations for the food bank.

Chapter 2

Background Literature

Following the financial crisis in 2008, there has been a dramatic increase in the price of food all across the USA, based on estimates from the Food and Agriculture Organization (FAO) (Figure 2-1) [10]. This led to a large number of people being dependent on the services of the Food Bank for a daily meal.

Increased dependency on fast-food has played an important part in the change of eating habits leading to health issues. This change started to have more health implications, that has been indicated by the increasing statistics of cardiac issues and diabetes by the Centre for Disease Control and Prevention (CDC) [11]. The recent trend in many children turning towards more fast food, due to the availability of more options at a much lower cost has increased the number of obese and overweight children in the country [12]. The same idea also applies to the low-income section of the population, primarily due to the financial constraints. Both these sections of the population have jointly increased the number of people who are at risk of having food-related medical conditions such as coronary heart disease, type 2 diabetes or a stroke [38].

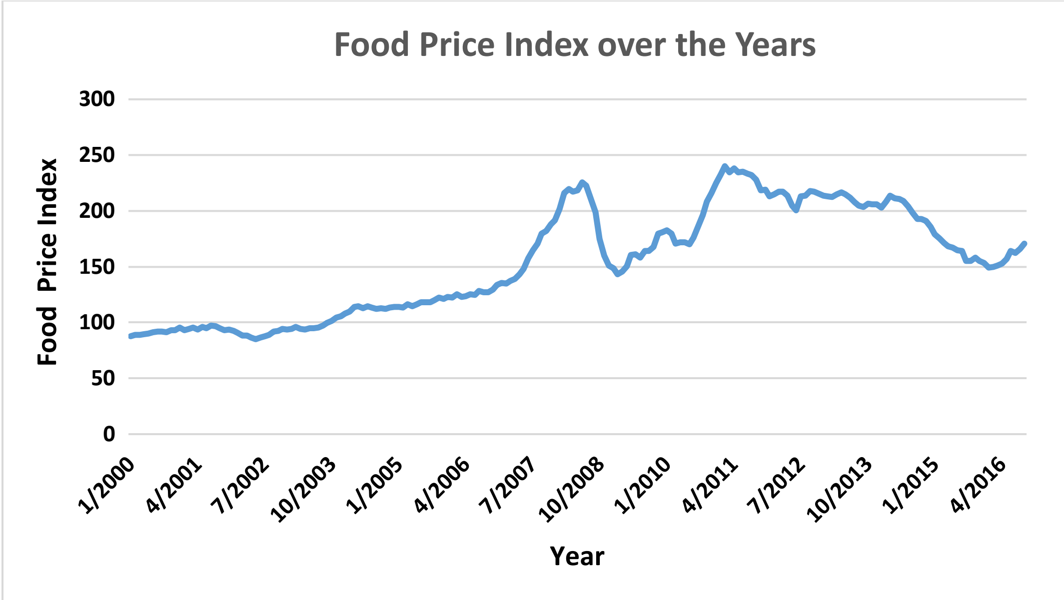

Figure 2-1: Change in Price of food over the years

Facing this problem, food banks are making an effort to increase the availability of fresh food to the people with low access to food (food deserts) and have a low income. In this chapter, we explore past research related to Food Deserts, Food Banks and Spatial Accessibility.

2.1 Food Deserts

The technical brief provided by Feeding America [13] discusses the methodology used to determine the food insecurity rate at the state level. This research takes into consideration the demographics of the state to estimate a model for food insecurity. It fails to take the geographic distribution of the population demographics at the county level into account, which makes a significant difference based on works that were more recent. Larsen et al. [7], describe the evolution of 'food deserts' in the city of London, Ontario over the years 1961-2005 by plotting them on a map. The research used geographic information system (GIS) based techniques to identify the precise location of supermarkets for computing access to them. Based on the longitudinal data Larsen et al., [7] determine that neighborhoods with the lowest socioeconomic status have the poorest access to fresh food supply from super markets. This research [7] also describes the access to the public transportation system in relation to the supermarket access. Larsen et al. [7] proves the existence of spatial inequalities in a major city but fails to provide a solution to the problem.

Waity [14] studies 24 sample counties within the state of Indiana and states that there are two types of assistance programs: Government assistance programs and Community assistance programs. Usage of community assistance programs was found to be lower than that of government programs. The main government assistance program is the Federal Supplemental Nutrition Assistance Program (SNAP). As for community programs, they were classified into two main types, namely food pantries and soup kitchens. The two types differ in terms of the assistance they provide, with food pantries providing mainly supplemental groceries that usually last for several days to a week, and soup kitchens providing free meals that could be consumed either inside the facility or could be taken away. Although usage of community assistance programs is lower than government assistance programs, significant amount of people use both [14]. The location of SNAP offices does not have the same degree of importance as the location of food pantries and soup kitchens. Depending on the state, people can have SNAP benefits through a government issued card that they can use in a network of shops and stores that accept them. SNAP card is widely accepted, and more importantly, discrete [3].

Waity [14] analyzes the spatial inequality that exists between rural and urban areas in access to food assistance agencies. This project maps the location of food agencies on the 24 sample counties in Indiana using geographic information system (GIS) analysis. It introduced the concept of food assistance desert (instead of the conventional concept of food desert) to investigate access to food assistance by food agencies (not access to food in general as it is the case in food desert). Using the population center of the census block group, the distance from the population center to the nearest food assistance agency was measured. If the closest agency was more than a mile away, the census block group was considered a food assistance desert.

According to the Food Access Research Atlas [2] provided by the Economic Research Service (ERS) of USDA, there are many ways to measure food store access for individuals and for neighborhoods, and many ways to define which areas are food deserts—neighborhoods that lack healthy food sources. However, most measures and definitions take into account at least some of the following indicators of access:

- Accessibility to sources of healthy food, as measured by distance to a store or by the number of stores in an area.
- Individual-level resources that may affect accessibility, such as family income or vehicle availability.
- Neighborhood-level indicators of resources, such as the average income of the neighborhood and the availability of public transportation.

Agrawal et al. [18] redefine the concept of food desert based on simulating the eating and purchasing behavior of low-income households. Agrawal et al. [18] makes assumptions about the shopping frequency, household location, fresh food consumption and the shopping quantity assumption. They also account for the changes in the shopping quantity by adopting a shopping type changing assumption. Additionally, they have analyzed the impact of having and not having a farmer's market in the neighborhood, which is open only from June to October in any given year. Based on the results of the simulation, Agrawal et al. [18] concluded that households near fast-food stores shop less fresh food and shop more fresh-food when the farmer's market is open (mostly from early April to end of November).

Food deserts have been covered extensively in the literature. Although the existence of food deserts cannot be considered as direct cause for food insecurity, they are good indicators of areas that are more likely to have food insecurity (Morton et al. [19]). As far as factors contributing to an area being a food desert, Dutko et al. [20] list a number of them. These factors include: high numbers of residents with low income or education, low levels of vehicle ownership, high rates of poverty, high rates of unemployment, high rates of minority population, and high rates of vacant housing.

2.2 Research on Food Banks

Davis, et al. [9] describe the problem at hand more accurately. Their research focuses on developing transportation schedules that enable the food bank to collect donations from local sources and deliver them to the charitable agencies. They identify satellite locations to plant food delivery points (FDPs), where agencies can receive deliveries. A set covering model was used to solve the problem of assigning agencies to their respective FDPs. Using the optimal assignment of agencies to FDPs, they identify a weekly transportation schedule that addresses collection and distribution of donated food and incorporates constraints related to food safety, operator workday, collection frequency, and fleet capacity.

Lien, et.al. [21] shed some light through their research on Supply Chain in Non-profit Organizations. When the country's GDP reached a whopping 15.8 trillion dollars a year in 2011 [36], almost 15% of the population was below the poverty line. Approximately 37 million people rely on Feeding America and their network of pantries, shelters and soup kitchens for food. Lien's case study deals with the Greater Chicago Food Depository (GCFD) and explains the change in objective function for the problem while considering non-profit operations (as opposed to commercial operations). While commercial operations focus on minimizing operational cost or maximizing profits, the same objective may lead to inequitable distribution of resources in non-profit operations. The study emphasizes the necessity of having an objective function that is aimed at equitable and effective service. Lien et al. [21] also implement a two-node decomposition (TND) heuristic with discrete demand distribution to mimic near-optimal conditions. Their research, motivated by the goal of GCFD to provide equitable service while maintaining a high level of distribution to its agencies, concluded by developing an objective function to maximize the expected minimum fill rate. Lien et al. [21] however does not account for the distance traveled for each pickup and delivery route. Lien et al. [21] have also not considered a model to prioritize the requirement of the agency based on the primary determinants of demand in that region.

2.3 Research on Spatial Accessibility

Guagliardo [22] discusses a very relevant aspect of spatial accessibility in the context of health geographics. His research work necessitates the need of a measure of accessibility far more complex than just distances and accounts for multiple parameters relative to providing the most fundamental source of healthcare to people to facilitate overall population health. One of the major drawbacks pointed out by this work is that distance to the nearest source of healthcare loses validity in heavily congested urban areas. Guagliardo [22] discusses the use of grouping barriers (availability, accessibility, affordability, acceptability and accommodation) that can impede progression from potential to realization of success. The main contribution of Guagliardo's work is the application of the combination of a gravity-based accessibility developed by Reilly [23] by modifying Newton's laws of gravitation, used to help with land use planning, further strengthened by Bantock et al. [24] using population demand to accessibility in healthcare services.

Apparicio et al. [25] describes a number of concepts in the field of spatial accessibility using various measures of distances while determining access to hospitals and other medical facilities. This research work compared the measures of distances such as Cartesian distances (Euclidean and Manhattan) and network distances (shortest network and shortest network time distances). The comparison was performed at the Census tract and Census block level to test the accuracy of the methods used. The results indicated that the Cartesian distances were less precise that network distances in the suburban level.

Waity [14] claims that the situation is quite different for community assistance programs. In order to be able to receive assistance, people must typically be present in person in specific locations of food agencies. Thus, the spatial structure of this network of food agencies plays an important role in order to improve food access. Waity [14] discusses that spatial inequality is a dimension of food insecurity and accessibility that is often overlooked. The concept of spatial inequality deals with the importance "who gets what, where" instead of the traditional conceptions of "who gets what, when". Allard [15], Kissane [16] and Sherman [17] found that spatial inequality affects the accessibility significantly. Waity (2016) [14] studied 24 sample counties in the state Indiana and showed that rural areas have less access, especially rural areas with high poverty rates.

Waity [14] makes three major statements on spatial accessibility based on the observations from her research. The first statement is that rural areas are more likely to be food assistance deserts. Waity claims that this may hold true as most of the smaller community-based agencies like food pantries and soup kitchens are less likely to be situated in rural areas due to availability of resources in urban areas, based on interviewing directors of different food agencies [14]. The second statement is with regard to the poverty level and food assistance deserts and was called “responsive community”. “Responsive community” states that counties with the highest poverty levels has fewer food assistance deserts and local churches and community service programs would to respond to the need by providing more aid [14]. The third statement is an alternative to the second one with regard to poverty and food assistance deserts and is called “spatial mismatch”. “Spatial mismatch”, states that the counties with higher poverty rates are more likely to be food assistance deserts as the agencies do not have the resources to assist everyone in need while the agencies in counties with lower poverty rates are more in number simply due to the availability of supporting resources [14]. Waity [14] concludes that the results of the analysis were supportive of the “responsive community” hypothesis. However, it is also noted that the results may differ if the sample space was much larger i.e. the project only accounts for the counties in Indiana.

2.4 Research Gaps

Previous research shows an extensive network of work done with respect to food deserts, food insecurity and spatial accessibility. However, the depth of the research work done is limited by the approach used by different researchers and have considered attributes that are more relevant to their approaches. For example, Davis et al. [9] focused on cost-effectivity, Lien et al. [21] focused on equity of distribution and Waity [14] used a more sociological approach. These research works are significant in their own way but seem disjointed while approaching the subject of food insecurity. Firstly, in the case of food deserts, Larsen et al. [7] and Agrawal et al. [18] describe the change in the behavior of shopping patterns in relation to the locations of the households. They do not describe the impact of the demographics of these household locations on the existence of food deserts accurately due to the unavailability of data on these households. Secondly, in the case of food banks, Davis et al. [9] explains the identification of satellite locations to plant FDPs, where agencies can receive scheduled deliveries, which poses a severe financial constraint on most of the non-profit operations. Lien et al. [21] overcome the drawback of Davis' [9] methodology by accounting for the financial constraint and equity of distribution, but do not account for the distance traveled for each pickup and delivery route. Lastly, in the case of the spatial accessibility, Guagliardo [22] and Apparacio et al. [25] complement each other's work and help develop measures of accessibility that can be used in both, the presence and absence of sufficient data. Their research together could contribute to the advancement of spatial accessibility determined for multiple purposes.

This thesis aims to bridge the gaps in the research works discussed in this chapter. The thesis focuses on studying all the existing attributes mentioned in the past research and use more recent data by analyzing the relationships between the multiple attributes in order to advance the research done in this field. It also focuses on contributing effectively to the field of “Operations in Non-profit Organizations”.

Chapter 3

Data Collection and Description

This chapter discusses the process of data collection and the description of the various datasets used in the thesis. Most of the research on Logistics and Supply Chain Management is in the industrial/commercial environment where the main objective is to maximize profit or minimize operational cost. Supply chain operations in non-profit organizations, where the main objective is the equity of distribution while maintaining operational effectivity Is not studied exhaustively. The objectives used in industrial/commercial settings may lead to affecting the equity in distribution of resources amongst the customers of the non-profit organization. Since these objectives are difficult to quantify in non-profit settings, we need to understand the parameters that affect to equity and effectiveness of distribution more carefully. In this thesis, we try to discern the differences between different possible contributing factors and understand impact of each factor on the primary objective.

To study and explore contributing factors we consider the case study of the Central Pennsylvania Food Bank, which caters to 27 counties within Central Pennsylvania. The data obtained from the Central Pennsylvania Food Bank website [26] provides the food distribution data for all the 27 counties in Central Pennsylvania. The radial distances of each of the 27 counties from the assigned food bank warehouses are provided below in

Table 3-1. The data on the counties and its specifics were collected from the Census Bureau website [27,28].

Table 3-1. Radial Distances of Counties from the Food Bank warehouses in Central Pennsylvania

Warehouse Location	**County**	**Radial distance (in miles)**
Harrisburg	Dauphin	9.71
	Snyder	35.80
	Perry	24.92
	Fulton	73.24
	Huntingdon	61.28
	Bedford	90.46
	York	24.87
	Juniata	35.24
	Mifflin	49.20
	Adams	34.94
	Cumberland	24.50
	Franklin	53.43
	Lebanon	20.03
	Lancaster	34.24
	Blair	79.92
Williamsport	Tioga	38.10
	Clearfield	74.65
	Bradford	48.70
	Clinton	29.87
	Columbia	37.27
	Sullivan	32.54
	Northumberland	32.55
	Montour	25.60
	Potter	55.45
	Lycoming	7.70
	Centre	46.11
	Union	18.77

3.1 Data Collection

In addition, the geographic description of Pennsylvania and its counties using the shapefile data provided by the United States Census Bureau [28] were obtained. The American Fact Finder website [29] provided the population data at the block group level and both datasets were integrated to provide us a comprehensive view with regard to the demographics in each county. The data collection process is shown in Figure 3-1.

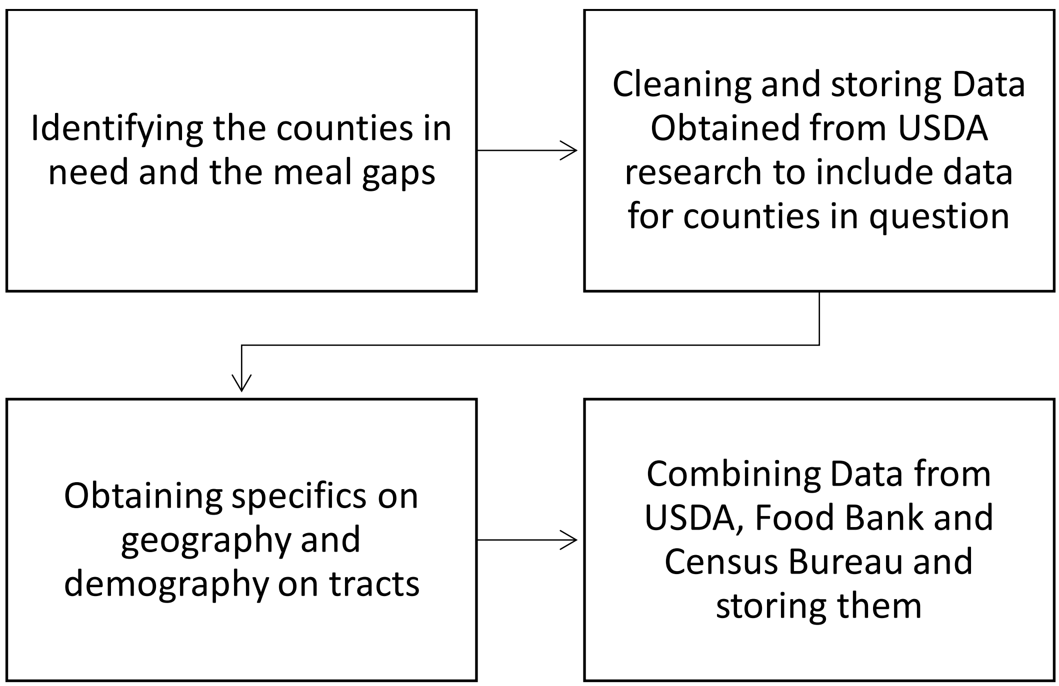

Figure 3-1: Data Collection process

3.1.1 Meal Gap Data

The Meal Gap (as shown in Fig 3-2) provided data about the distribution of food through the Food Bank, its agencies and other Federal Government programs, the Meal Gap and Additional Meals Required at the County level on a yearly basis. The meal gap is defined as the total number of meals required per year in order to eliminate hunger. The total amount of food (in lbs.) distributed to the food agencies in each of the counties by the Food Bank through its warehouses for 2014 are listed in Figure 3-2. The number of meals provided through the different programs such as Supplemental Nutrition Assistance Program (SNAP), Eldershare and Kid’s Cafe® [26] gives us an idea as to what section of the population the program is aimed at. The amount of food possible through other programs and the number of people eligible to participate in each of these programs, within each county, are also described in Figure 3-2. The fundamental idea behind the operation of the food bank is to fight hunger and improve lives of the individuals and families in need. Hence, the assumption made about the data is that it targets households and certain individuals who are below the poverty line (Poverty Line is defined as estimated minimum level of income needed to secure the necessities of life). Healthcare.gov [31] dictates that the federal poverty line as $ 24,300 for a family of 4 in the United States of America. The state of Pennsylvania determines that a county is below the poverty level if the county has a median household income that is below 80% of the median household income for the state of Pennsylvania. The median household income for Pennsylvania is $53,115 [30].

Central Pennsylvania Food Bank												
	FY 2014	CY 2014										Additional
County	Total Pounds Distributed	Total Pounds Distributed	Total Pounds Less non/food and beverages (10%)	Total Meals from food	Sep-14 Potential Snap participants	Total eligible applicants	*Total possible meals from snap	*Meals provided by agencies from other sources(25%)	Potential school breakfast participants	Total possible meals from school breakfast	Meal Gap	Total Meals Needed
Adams	1,091,184	1,058,495	1,041,622	868,018	8,354	418	440,674	217,005	6,088	54,792	1,830,200	249,712
Bedford	1,225,810	1,326,634	1,313,049	1,094,208	6,304	0	0	273,552	2,988	0	1,117,400	0
Blair	1,086,448	1,234,962	1,202,257	1,001,881	11,221	0	0	250,470	7,630	0	2,824,900	1,572,549
Bradford	249,671	306,039	304,072	253,393	8,429	0	0	63,348	3,760	0	1,320,500	1,003,758
Centre	698,365	708,776	692,391	576,993	29,176	0	0	144,248	9,245	0	3,910,800	3,189,559
Clearfield	274,082	304,601	301,585	251,321	7,390	0	0	62,830	4,902	0	1,884,500	1,570,349
Clinton	306,761	330,561	325,867	271,556	4,643	0	0	67,889	2,356	0	972,000	632,555
Columbia	383,728	417,817	416,971	347,476	6,534	0	0	86,869	4,040	0	1,583,200	1,148,855
Cumberland	1,685,907	2,056,534	2,028,175	1,690,146	22,258	1,113	1,174,110	422,536	14,134	127,206	4,579,800	1,165,802
Dauphin	6,624,094	7,406,981	7,184,330	5,986,942	13,776	689	726,684	1,496,735	16,096	144,864	6,823,300	0
Franklin	913,687	938,115	921,236	767,697	16,409	0	0	191,924	8,983	0	3,021,100	2,061,479
Fulton	148,834	168,921	166,038	138,365	1,298	0	0	34,591	891	0	327,500	154,544
Hungtingdon	271,448	305,087	297,817	248,181	3,284	0	0	62,045	2,757	0	1,077,100	766,874
Juniata	293,493	298,013	287,373	239,478	3,731	0	0	59,869	1,479	0	502,600	203,253
Lancaster	3,395,203	4,057,498	3,935,632	3,279,693	50,739	2,537	2,676,482	819,923	31,187	280,683	10,410,100	3,353,318
Lebanon	672,005	728,492	713,944	594,953	10,965	548	578,404	148,738	8,019	72,171	2,534,200	1,139,934
Lycoming	1,520,613	1,464,119	1,408,028	1,173,357	7,495	375	395,361	293,339	6,971	62,739	2,795,200	870,404
Mifflin	480,717	537,494	527,703	439,753	6,772	0	0	109,938	2,803	0	1,096,400	546,709
Montour	77,098	91,697	90,668	75,557	1,699	0	0	18,889	1,097	0	350,300	255,854
Northumberland	595,425	710,263	689,932	574,943	9,457	473	498,857	143,736	5,675	51,075	2,276,800	1,008,189
Perry	1,039,053	1,156,240	1,138,302	948,585	4,115	0	0	237,146	2,760	0	900,200	0
Potter	173,650	163,954	159,049	132,541	2,067	0	0	33,135	1,048	0	406,300	240,624
Snyder	290,661	361,341	351,350	292,792	4,689	0	0	73,198	2,384	0	840,700	474,710
Sullivan	114,732	120,088	119,833	99,861	1,123	0	0	24,965	386	0	134,900	10,074
Tioga	520,448	519,994	508,276	423,563	5,855	0	0	105,891	2,521	0	956,200	426,746
Union	275,215	294,706	285,247	237,706	5,007	0	0	59,426	2,699	0	1,001,800	704,668
York	3,237,039	3,821,312	3,688,492	3,073,743	25,201	1,260	1,329,353	768,436	26,115	235,035	8,870,700	3,464,133
Other FBs	4,984,466	7,059,026	N/A	N/A							N/A	N/A
Total	32,629,837	37,947,760	30,099,239	25,082,699	277,991	7,412	7,819,924	6,270,675	179,014	1,028,565	64,348,700	26,214,653

Figure 3-2. Meal Gap Data

3.1.2 U.S. Department of Agriculture Food Desert Data

The Food Access Research Data [2] provides demographic information about the County at the tract level. This data helps us identify Food Desert locations in a county and people with low access and low income within rural and urban census tracts. Low Access is defined as 1 mile for urban census tracts and 10 miles for rural census tracts. Low Income tract is defined as the tract that has a median household income that is below 80% of the median household income of the county [2].

3.1.3 Census Bureau Demographics Data

The demographics data obtained from the Census Bureau [27] provides us with the details of number of people within a census tract, the median household income and the geographic location of their centroids. This data is used to define the important attributes that contribute to identifying the counties and census tracts that were affected by food insecurity and prioritizing them in terms of satisfying resource requirements.

3.1.4 Tiger Line Shapefile Data

The shapefile data provides us with the geographic boundaries determined by the Census Bureau [27] to identify the divisions within the state and their respective counties. The data pertaining to our study i.e. the 27 counties under the Central Pennsylvania Food Bank were extracted from the shapefile data for the purpose of map analytics.

3.2 Data Description

The data on census tracts from the Central Pennsylvania Food Bank [26], USDA [2] and Tiger Line shapefiles [27] is cleaned and stored as different datasets. The various datasets were consolidated and their description are explained in this section.

3.2.1 Meal Gap and Data at the County Level

The meal gap data, the list of partner agencies provided by the food bank and the Demographic data at the Block group level (discussed in Chapter 1) for all the counties within Pennsylvania are cleaned and stored separately. The data description of the datasets are given in tables 3-2, 3-3 and 3-4.

Table 3-2. Data description of meal gap data

Attribute	**Description**	**Variable type**
County	Name of the County within Pennsylvania	Character
Meal Gap	The conversion of the total annual food budget shortfall in a specified area divided by the weighted cost per meal in that area	Numeric

Table 3-3. Data description of Partner agencies

Attribute	**Description**	**Variable type**
Name	Name of the partner agency	Character
Address	Physical address of the agency	Character
Quantity	Amount of food provided to the agency (in lbs)	Numeric
Type	Type of food supplied to the agency	Categorical
Longitude	Geocoded longitude of the agency	Numeric
Latitude	Geocoded latitude of the agency	Numeric

Table 3-4. Data description of Demographic data at block group level

Attribute	**Description**	**Variable type**
GEOID	Block group Identification number assigned by Census Bureau	Character
Population	Population of the block group	Numeric
Longitude	The longitude of the block group	Numeric
Latitude	The geocoded latitude of the block group	Numeric

3.2.2 USDA Food accessibility research data

U.S. Department of Agriculture [2] collected data on the food accessibility in different states across the USA at the Census Tract level. They included statistics on people with accessibility to food who had less access in terms of distance, without vehicular access and existence of low-income tracts. This data was cleaned for the purpose of analysis and stored separately. The data description is given in Table 3-5.

Table 3-5. Data description of USDA's Food Access Research

Attribute	**Description**	**Variable type**
County	Name of the County within Pennsylvania	Character
Additional Meals	Additional Meals Required in each county per day	Numeric
Litracts	Low-income tract or not	Binary
Lapop1	Low access population at 1 mile	Numeric
Lalowi1	Low access and Low income population at 1 mile	Numeric
Lakids1	Low access kids at 1 mile	Numeric
Laseniors1	Low access seniors (ages 65+) at 1 mile	Numeric
Lapop10	Low access population at 10 miles	Numeric
Lalowi10	Low access and Low income population at 10 miles	Numeric

Lakids10	Low access kids at 10 miles	Numeric
Laseniors10	Low access seniors (ages 65+) at 10 miles	Numeric
Population	2010 Census population in counties	Numeric

3.2.3 Census Bureau Data on Demographics at the Tract level

The Census Bureau provides a comprehensive dataset on the demographic features of each tract. This includes the location of the centroid of each tract, population and household income in each Census tract [29]. The data description for this dataset is given in Table 3-6.

Table 3-6. Data description of demographics at tract level

Attribute	**Description**	**Variable type**
GEOID	Block group Identification number assigned by Census Bureau	Character
Longitude	The longitude of tract centroid	Numeric
Latitude	The latitude of tract centroid	Numeric
Household Income	Median Household Income of the respective Census Tract	Numeric
Population	Population of the Census Tract	Numeric

All of the datasets described above have been stored as comma-separate value (.csv) files on Excel and are extracted when needed for the purpose of analyses. Each of these datasets are analyzed either individually or in relation to the another dataset to determine the relation between multiple attributes extracted from different sources.

Chapter 4

Methodology

The objective of this thesis as mentioned in Chapter 1 is to study the use of different methodologies in the field of food insecurity and food accessibility and identify the major factors that hinder spatial accessibility and equity of distribution of food in food deserts. The different methodologies are studied using different datasets consisting of possible attributes that could contribute to minimizing the meal gap and the results obtained are discussed. A trial and error methodology is used that is described in Figure 4-1 below.

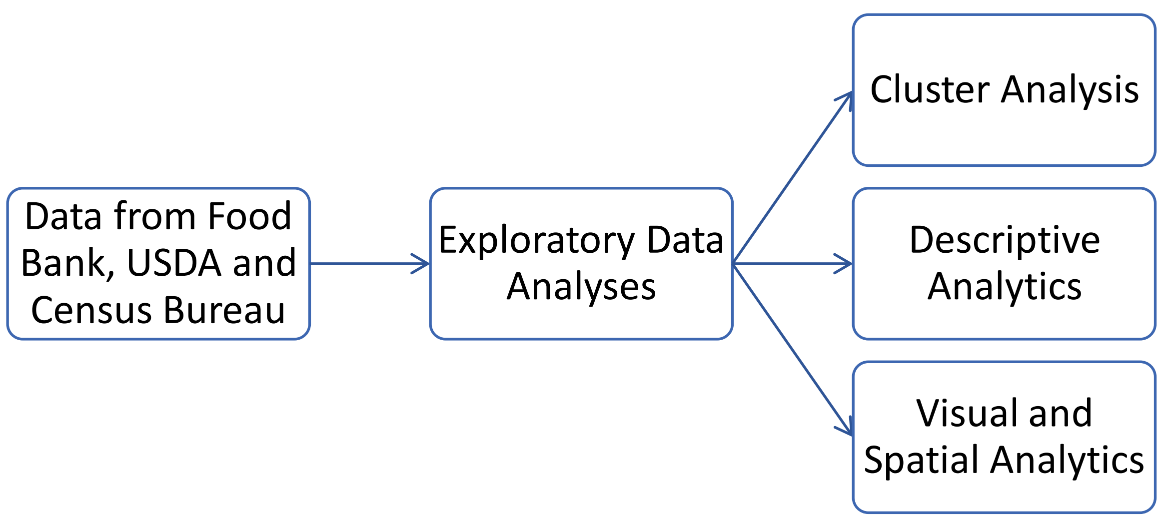

Figure 4-1. Flowchart of Methodology

All the data obtained are in the same frequency, except for the demand data from the Food Bank. The data obtained was cleaned for any missing data and then used for the purpose of analyses.

4.1 Exploratory Data Analysis

The datasets obtained from the Food Bank, USDA’s research on food deserts and the US Census Bureau are cleaned and the attributes for the purpose of each methodology are obtained at this stage. The different attributes are analyzed individually in order to observe patterns.

4.2 Different Methodologies Used

At this stage, k-means clustering is performed initially in order to identify consolidation points for the purpose of improving accessibility. K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining [32]. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. K-means clustering is applied to the data set containing the replicated data tuples.

Further descriptive analysis is performed using visualizations on the data provided by the food bank on the meal gap at the county level to understand distribution of resources in detail and compare them with the demographics of the county. This was done to understand the distributions of possible contributing attributes from the multiple datasets used (discussed in section 3.2).

Lastly, a visual and spatial analysis approach was used to identify the distributions of different attributes at the Census tract level. Visual analytics [33] is an outgrowth of the fields of information visualization and scientific visualization that focuses on analytical reasoning facilitated by interactive visual interfaces. Spatial analysis [34] or spatial statistics includes any of the formal techniques, which study entities using their topological, geometric, or geographic properties. These techniques helped provide a visual reasoning on the possible contributing factors to food insecurity.

The application of these methodologies to the problem have been discussed in detail in Chapter 5.

Chapter 5

Analysis

Literature review (see Chapter 2) indicated that there are research gaps in the past work, primarily due to the unavailability of data while studying non-profit operations. In addition, the access to healthy food, which is a growing concern, has not been explored very effectively by past research. This section aims to 1) explore the data provided on the Central Pennsylvania Food Bank and its operations, 2) study the possible factors contributing to food insecurity from different sources in detail and 3) establish a relationship between relevant factors that would help the Central Pennsylvania Food Bank prioritize their customers' demand more efficiently.

The Central Pennsylvania Food Bank is responsible for 27 counties located in Central Pennsylvania and has two warehouses situated in Harrisburg and Williamsport respectively. The warehouse at Harrisburg supplies to 15 of the 27 counties while the warehouse at Williamsport supplies to the remaining 12 counties as shown in Figure 5-1. The following analyses were performed keeping in mind the hierarchy the food bank mentioned above.

It was important to study the operational policies and delivery patterns of the Food Bank to understand how they function. This step was crucial to identify policies and develop a framework in order to perform the analyses successfully. Also exploring the other datasets would help identify the right set of attributes that would be useful to understand the data patterns and their relation to the problem defined in Chapter 1.2.

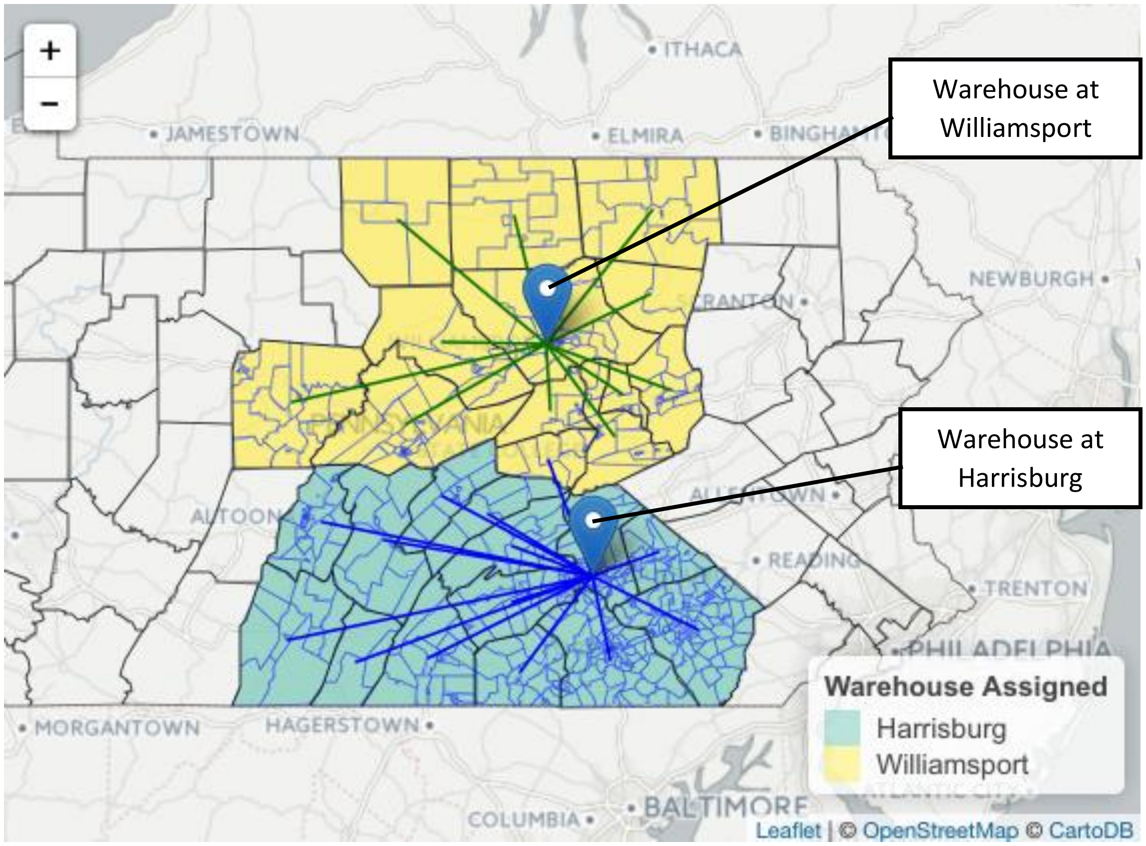

Figure 5-1. Counties served by the respective Warehouses (Harrisburg and Williamsport)

Exploratory data analysis, along with visual and descriptive analytics are used to answer the following questions:

1. What are the current conditions of the food bank in terms of distribution of food?
2. How does this distribution affect the healthcare aspects of the low-income population within the counties?
3. What are the major factors affecting the additional number of meals required per year, even though there is huge network of agencies in counties extremely affected by food insecurity?

4. How does spatial analytics relate to the problem? What attributes can be used to understand the results of spatial analytics in a better way?

We try to answer all of the above questions through the following analyses.

5.1 Identification of Consolidation points based on Meal Gap

This section discusses the analyses done by prioritizing just the meal gap similar to Davis [9]. The data collected at the Census block group level on its geography and demographics [8] are used to understand the impact of demographics and possibility of developing consolidation points based on higher population and low accessibility. A measure of accessibility is used to determine the number of block groups that have access to food agencies versus those that do not have access to food agencies. This measure in-turn is used to identify how each block group identifies itself in terms of population and distance from a food agency.

5.1.1 Supply Vs Demand

The following bar chart (Figure 5-2) shows the total supply and demand for the two food banks. The food bank at Harrisburg covers 15 out the 27 counties in central Pennsylvania, while the food bank at Williamsport covers the remaining 12 counties. It is apparent that the demand is more than the supply for both the food banks. The gap (or additional meals required or unsatisfied demand; all these terms are used interchangeably) for Harrisburg is estimated to be 18.2 million pounds per year, and the gap for Williamsport is estimated to be 13.3 million pounds per year (based on CY 2014 data).

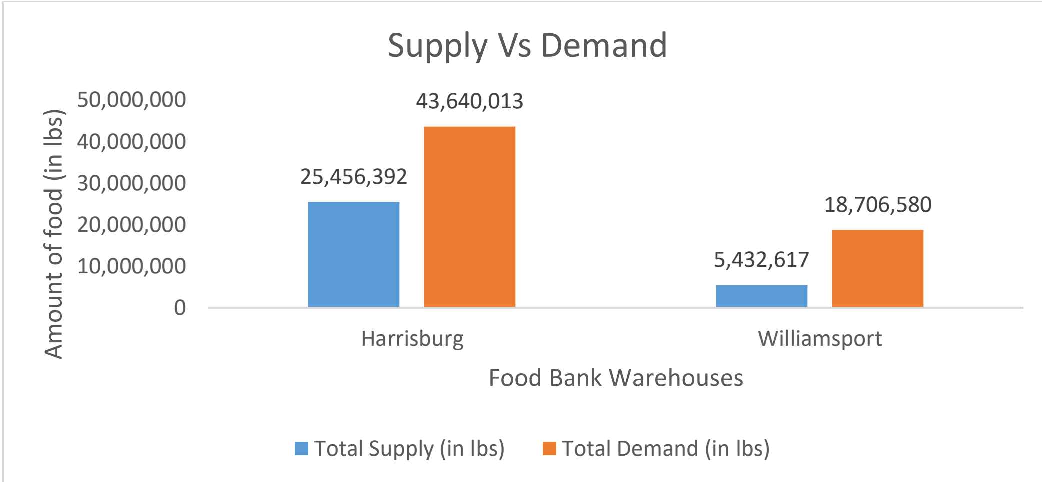

Figure 5-2. Supply Vs Demand for Food Bank Warehouses

5.1.2 Meal Gap Distribution across Counties

The following pie chart (Figure 5-3) shows the distribution of the Meal Gap across the counties within Central Pennsylvania.

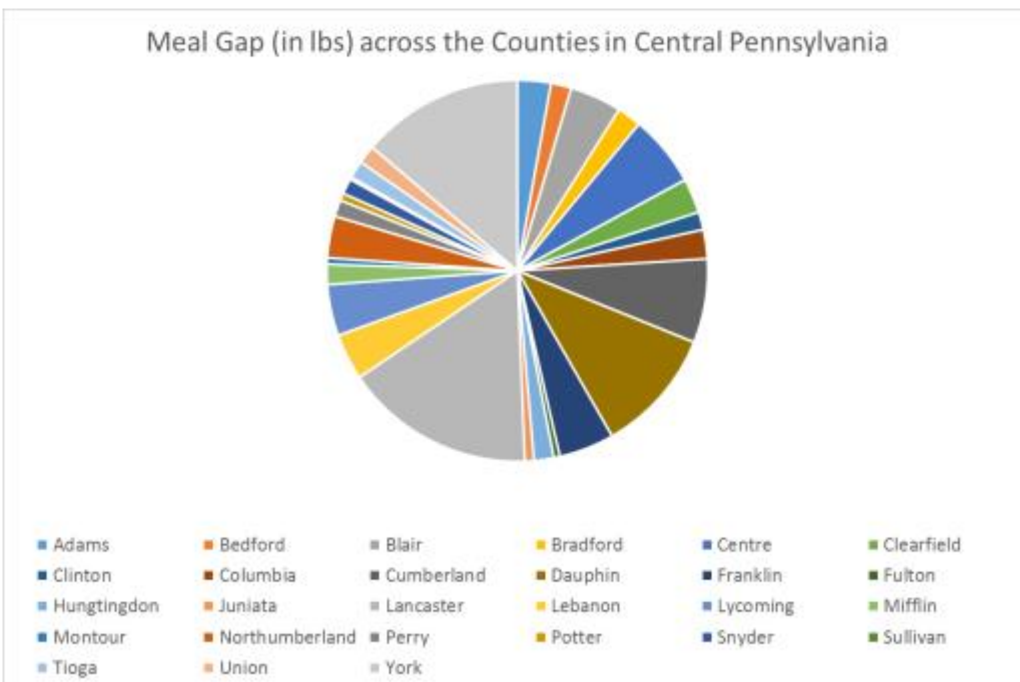

Figure 5-3. Meal Gap across the Counties in Central Pennsylvania

The meal gap in Central Pennsylvania alone amounts to 77,218,440 lbs. per year. For the purpose of our analyses, we consider the four counties that are adjacent to each other and have a relatively high meal gap. The four counties are Dauphin (10.6%), York (13.8%), Lancaster (16.2%) and Lebanon (3.9%). These four counties alone contribute to almost 45% of the overall meal gap.

5.1.3 Computing Accessibility Index for Block Groups

The addresses of the agencies within the four chosen counties were geocoded and stored in the form of geographic coordinates. The coordinates of the centroid and the population of the block groups were obtained from the shapefile. The geographic coordinates for the agencies and the centroid of the block groups were used to find the distance between a block group and an agency. A spatial accessibility parameter, similar to the one discussed by Guagliardo [22] was used.

The minimum radial distance (d_{ij}) between the centroid of the block group 'i' and the agency 'j' and the population of the block group (p_i) were normalized and used in developing a spatial accessibility parameter (A_i) which is defined as follows:

$$A_i = \frac{p_i \times d_i}{\sum_{j=1}^{n}(p_j \times d_j)} \times 100 \text{ (Equation 5-1)}$$

5.1.4 Implementation of k-means Cluster Methodology

This section discusses the segmentation of block groups without access (measured in 5.1.3) to food agencies in the four counties. The methodology and the assumptions made are also described in detail.

5.1.4.1 Assumptions

The assumptions made about different attributes in the analyses are listed below:

- Population: The population of the block group are used as a proxy for the expected number of people in need of food from food banks due to the unavailability of data on meal gap at the block group level. In other words, as the population density of the block group increases, the expected number of people in need of food from the food banks increases proportionately.
- Block Groups: In order to improve accessibility of food agencies, block groups, which are at a distance greater than a defined threshold from the nearest agency, were identified.

5.1.4.2 Classification of Block Group based on Accessibility

The analysis assumed a distance of two kilometers (2 km) as the threshold for determining if the block group had access to a food agency or not. The following figures (Figure 5-4 and Figure 5-5) shows the spread of food agencies and block groups in Dauphin, York, Lancaster and Lebanon counties.

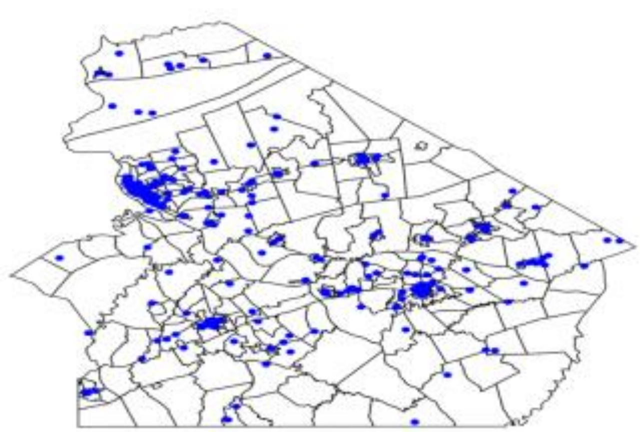

Figure 5-4. Distribution of Food Agencies (blue) in Dauphin, Lancaster, Lebanon and York

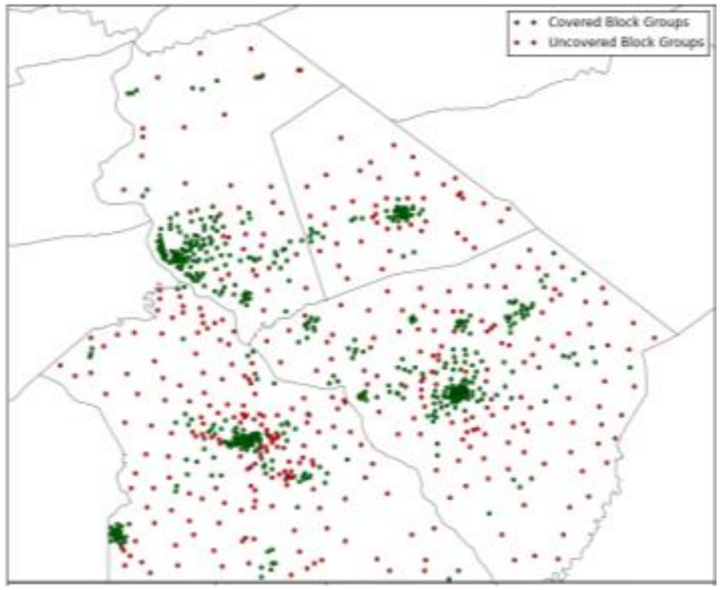

Figure 5-5. Distribution of Covered (in green) and Uncovered (in red) Block Groups

The following figure (Figure 5-6) shows an overlay of the food agencies and the block groups.

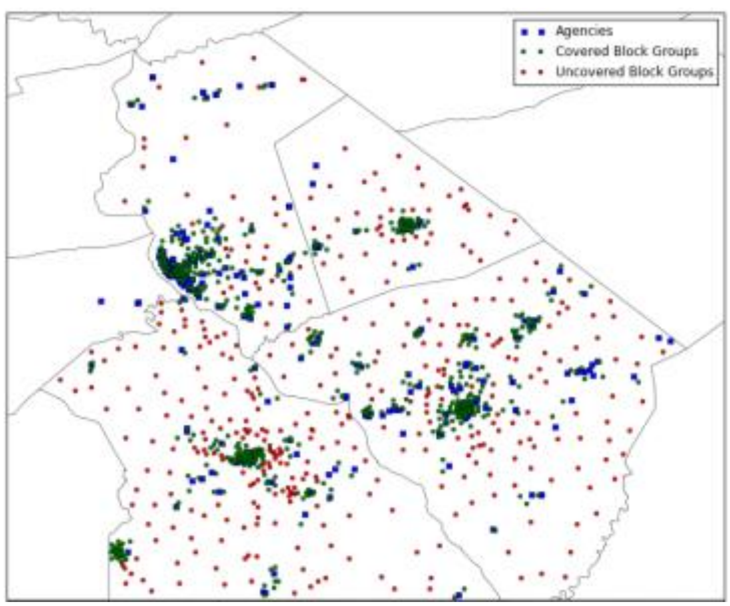

Figure 5-6. Overlay of Food Bank Agencies (in blue) at Block Groups

5.1.4.3 Weighted Clustering of Consolidation Centers

This section describes the weighted clustering used on the uncovered block groups based on the population and the Euclidean distance of the uncovered block group from the nearest agency. The Euclidean distance is used for the purpose of quick and easy analysis. The weighted clustering approach was used so that centroid of the cluster would be biased to the block group that had a comparatively higher population and/or was far from the centroid of the cluster.

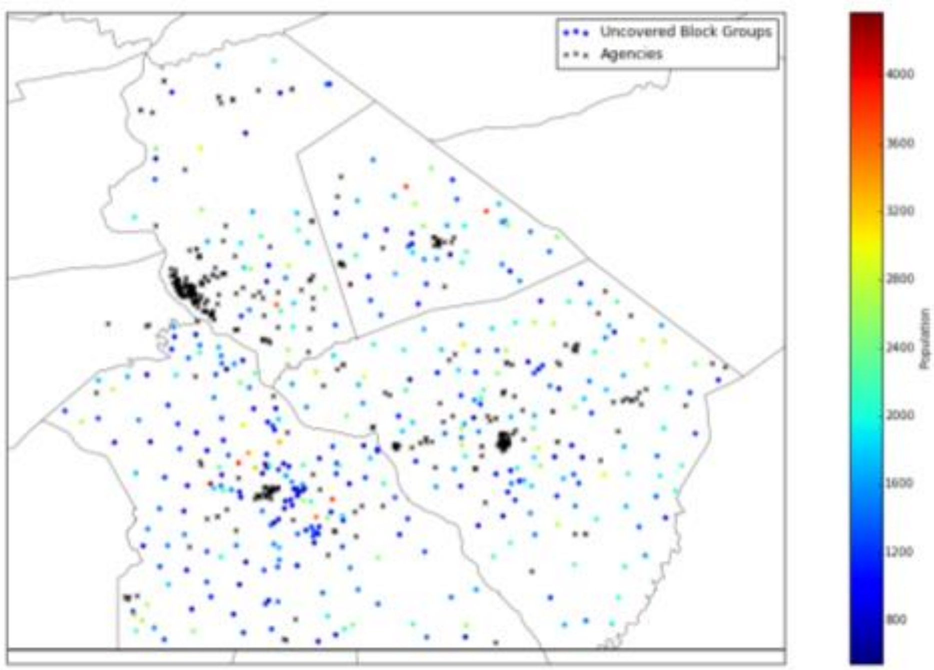

Figure 5-7. Heat Map of population within uncovered block groups and the food agencies

The following steps were used to perform the k-means clustering on the uncovered block groups (shown in Figure 5-7):

- Normalization of distance and population: In the previous section (5.1.4.2), we identified a subset S of uncovered block groups.

For each block group $x \in S$, the following attributes were calculated:

k_x: Euclidean distance between the block group x and the nearest agency.

P_x: Population of the block group

Ideal value scaling is performed in order to scale the range of both k_x and P_x to [0,1].

- Calculation of cluster weights: The following formula is used to calculate the

cluster weights for each $x \in S$,

$$w_x = \alpha\, k_x + (1-\alpha)\, P_x, \text{ where } \alpha \in [0,1] \quad \text{(Equation 5-2)}$$

- Replication of data tuples: Each data tuple $x \in S$ were replicated based on the

corresponding value of 100w (100 is multiplied in order to improve the resolution between the cluster weights). The objective of this step is to direct the clustering algorithm to give more importance to those values of $x \in S$ which have higher values of w. This consequently directed the algorithm to identify cluster centers towards block groups with higher populations and at a greater distance from the nearest agency (adversely affected).

- Clustering: K-means clustering is applied on the data set containing the

replicated data tuples. The normalization of the latitude and longitude has been applied prior to K means clustering. This ensures that the latitude and longitude is normalized to have unit variance. The clustering algorithm was tested for different values of α and for different number of clusters.

- Location of consolidation centers: The consolidation centers are then located at the identified cluster centroids. Additionally, a measure called distortion has been defined. Distortion is defined as the sum of the squared differences (Radial distance) between each block group and its corresponding centroid. This measure is used mainly to compare the different test results, mainly to determine the appropriate number of clusters.

5.2 Distribution of Resources by Food Bank

This section describes distribution of food by the Food Bank and the distribution patterns of the deliveries made by the Food Bank from the warehouses at both Harrisburg and Williamsport.

5.2.1 Additional Meals required by each County

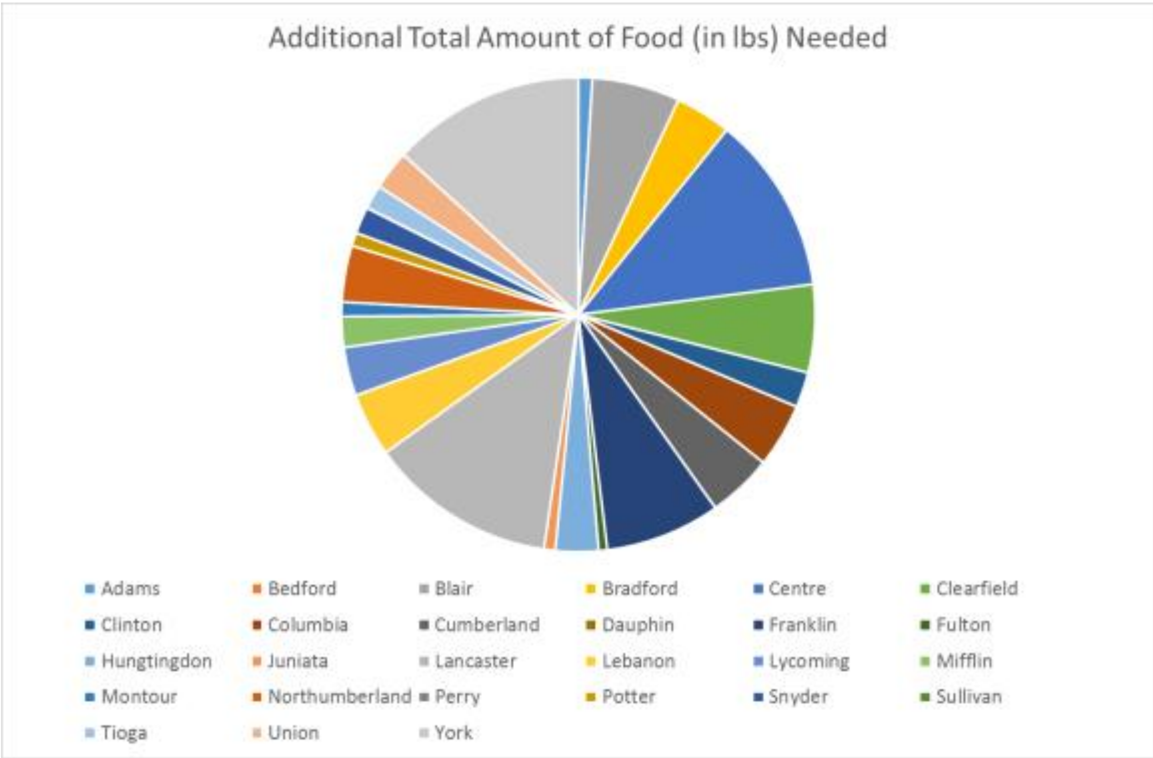

Figure 5-8. Additional Amount of Food needed (per year) in the Individual Counties

For the purpose of this analysis, we take into account the additional amount of food needed in order to satisfy the meal gap in each county. The pie chart (Figure 5-8) illustrates the distribution of additional meals required by each county. The additional meals required is the unsatisfied demand per year, which amounted to 31,457,584 lbs. in total. The top four counties in this case are Lancaster (13%), York (13%), Centre (12%), and Franklin (8%).

5.2.2 Demand-Supply data for the Individual Warehouses

The Demand-Supply analyses performed shows the amount of food supplied and the additional food needed in each county with respect to the warehouses in-charge of distribution in the county.

The following bar chart (Figure 5-9) shows the Supply Vs Demand for each county covered by the food bank at Harrisburg.

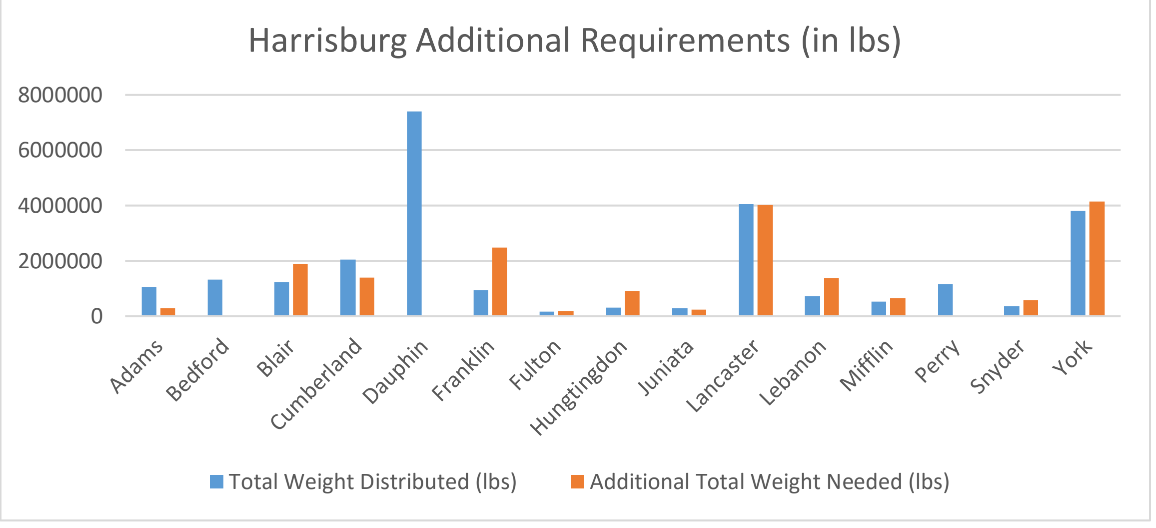

Figure 5-9. Additional Requirements at Harrisburg Warehouse

The following bar chart (Fig 5-10) shows the Supply Vs Demand for each county covered by the food bank at Williamsport.

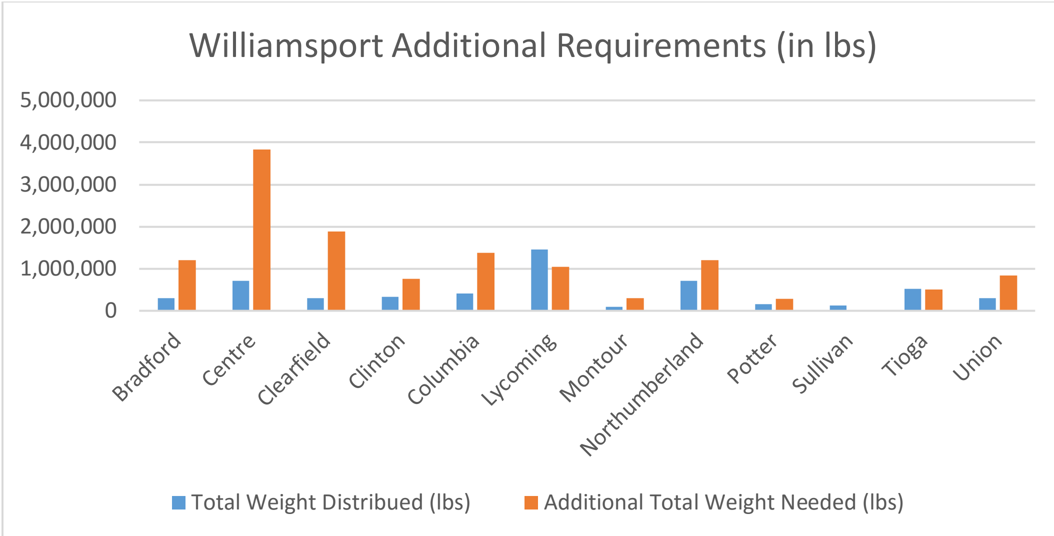

Figure 5-10. Additional Requirements at Williamsport Warehouse

From figures 5-9 and 5-10, it is observed that the counties under Williamsport are the most affected in terms of equity of distribution, as they have higher requirement, given the existing supply. This analysis also indicated that the additional amount of food required is a far better measure than meal gap (as opposed to what was discussed in 5.1) to prioritize the county in need of more resources.

5.2.3 Comparison of Demographics from USDA with Food Bank Data

The following table (Table 5-1) shows the attributes for each Census tract (aggregated at the county level), as discussed in 3.2.2. Some of the data is collected from USDA Food desert analysis [2]. Assuming that the demographics would not have changed considerably since 2010, we used both the data provided by Food Bank [3] and collected from USDA Food desert [2].

Table 5-1. Comparison of Meal gap data with USDA data on food access research

County	Number of additional meals needed per day (2014)	Number of people in LaLowi1 (1 mile)	Number of people in LaLowi10 (10 mile)	2010 census population	% of people with LaLowi
Adams	694	15,605	0	101,407	15%
Bedford	0	14,680	1,829	49,762	33%
Blair	4,368	17,293	0	127,089	14%
Bradford	2,788	16,314	1,354	62,622	28%
Centre	8,860	17,542	332	153,990	12%
Clearfield	4,362	19,534	897	81,642	25%
Clinton	1,757	7,580	660	39,238	21%
Columbia	3,191	9,379	0	67,295	14%
Cumberland	3,238	18,082	8	235,406	8%
Dauphin	0	41,276	0	268,100	15%
Franklin	5,726	23,754	942	149,618	17%
Fulton	429	4,103	669	14,845	32%
Huntingdon	2,130	9,967	302	45,913	22%
Juniata	565	7,498	551	24,636	33%
Lancaster	9,315	59,050	0	519,445	11%
Lebanon	3,166	13,552	0	133,568	10%
Lycoming	2,418	18,887	1,478	116,111	18%
Mifflin	1,519	11,056	70	46,682	24%
Montour	711	3,710	0	18,267	20%
Northumberland	2,801	14,234	0	94,528	15%
Perry	0	9,893	1,131	45,969	24%
Potter	668	4,833	1,490	17,457	36%
Snyder	1,319	8,931	655	39,702	24%
Sullivan	28	2,273	872	6,428	49%
Tioga	1,185	10,373	374	41,981	26%
Union	1,957	12,526	211	44,947	28%
York	9,623	42,030	0	434,972	10%

One interesting finding is that there is a negative correlation (r = -0.6) between the percentage of people with low income/low access and the number of meals required daily. This may be because food banks favoring counties with higher percentages of low income/low access people at the expense of other counties. Based on the results discussed above, investigation of equal distribution of resources might lead to an ideal solution.

5.2.4 Types of Food Distributed across different Programs

Exploratory analysis of delivery data revealed the following stacked chart that illustrates the amount of food for each type and for each program (Figure 5-11).

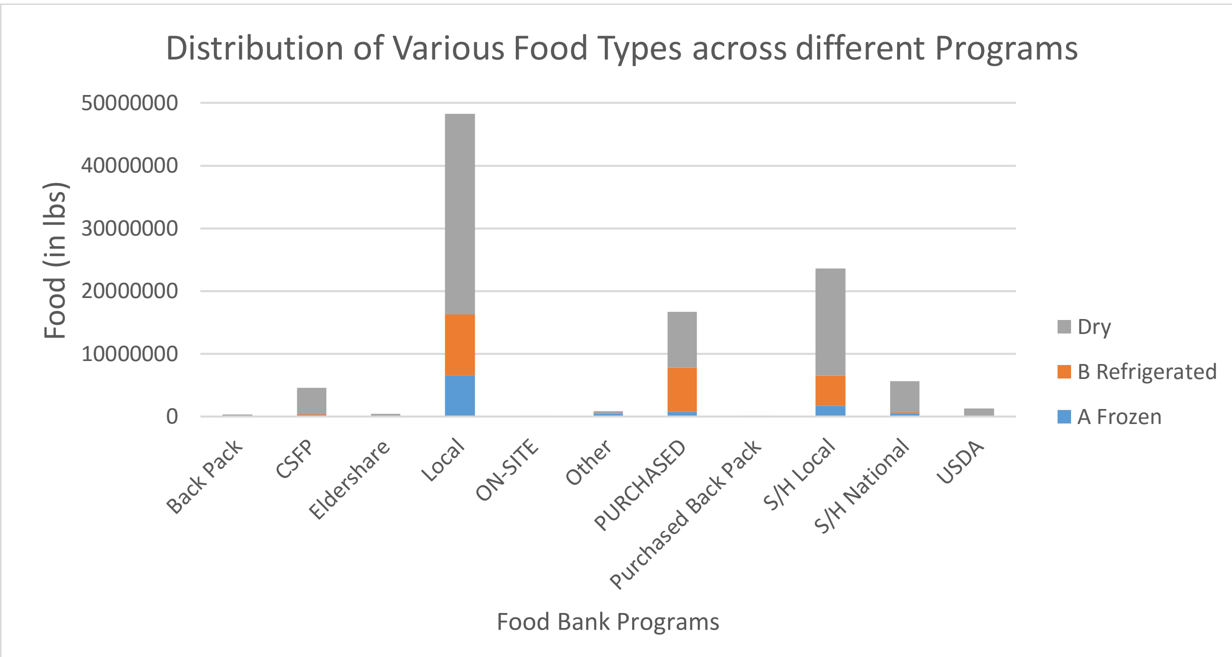

Figure 5-11. Types of Food across different Programs

It seemed that ‘Dry food’ is the dominant type of food distributed to the people in need. There is a lower tendency to distribute refrigerated and frozen food. This may be due to lack of storage facilities. Another factor that could explain this is perishability factor that is associated with these type of foods. The perishability factor might be higher for these types of foods, especially the refrigerated food. This analysis also proves that the amount of fresh food (Refrigerated Food) is much lesser compared to the 'Dry food', which is striking health concern.

5.3 Impact of Demographics on Food Insecurity using Census Data

The earlier analysis performed were based upon just the meal gap aggregated at the county level. The block group analyses were done by identifying the nearest agencies irrespective of the county boundary. One of the major realizations was that the people in need within each county could only have access to the food from the agencies present within each county.

To accommodate the recent changes, the level of data granularity was brought up to the tract level in order to identify the boundaries and it now even accounted for the data provided by the census bureau. This analysis is important as the counties in question have tracts ranging from 2 to almost 100 in number and the population per tract ranges from 1000 to 10000. The more recent data accumulated now consists of the Census Tract identification details, the household income, population, demand, and its coverage by the nearest agency.

5.3.1 Census Tract Data Analysis

The data collected at the tract level was analyzed based on the county they belonged to and a visual analysis was conducted. This determined the tracts within the county that were both covered and uncovered by the agencies within the county.

The shortest distance between the centroid of the tract and the nearest agency was computed and used as measure of accessibility. An example of a plot of Potter County is shown below (Figure 5-12) highlighting the difference between the centroid of the tract and the agencies within Potter County.

Figure 5-12. Distribution of Food Agencies and Census tracts in Potter County

This analysis was performed to determine the existence of a food agency in any given tract. Further exploratory analysis of the different new attributes was performed to understand their distribution at the tract level and are described below.

5.3.2 Descriptive analytics of the Census Tract

A descriptive analysis was conducted by testing the correlation matrix between the attributes at the tract level: presence of agency within tract (Cov); distance to the nearest food agency (Distance); low-income tract or not (litracts); census population of tract (population); median household income (hhincome); and the low access and low-income population in urban and rural tracts (refer to Table 3-5). This was done to test if one attribute contributes more to the additional amount of food required based on the county which has a requirement. An example using Census Tracts in Bedford County is shown in Figure 5-14.

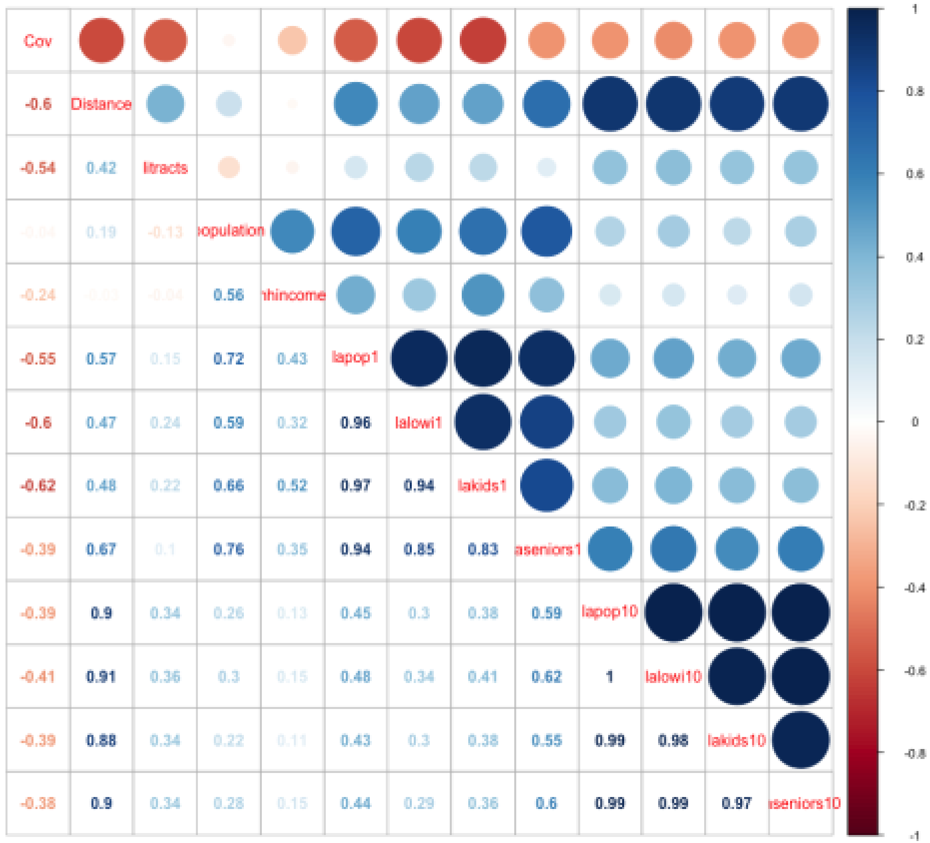

Figure 5-14. Correlation matrix of attributes in Bedford County

Further analysis was performed to understand the factors contributing to the additional meals in each county individually.

5.3.3 Visual Analysis of demographics in the Tracts

The population data obtained from the Census Bureau was used to create an interactive map of the Census tracts with population in order to visualize the density of population without access to food and low income, the median household income and the presence of a food agency in the tracts being analyzed as part of the study (Figure 5-13).

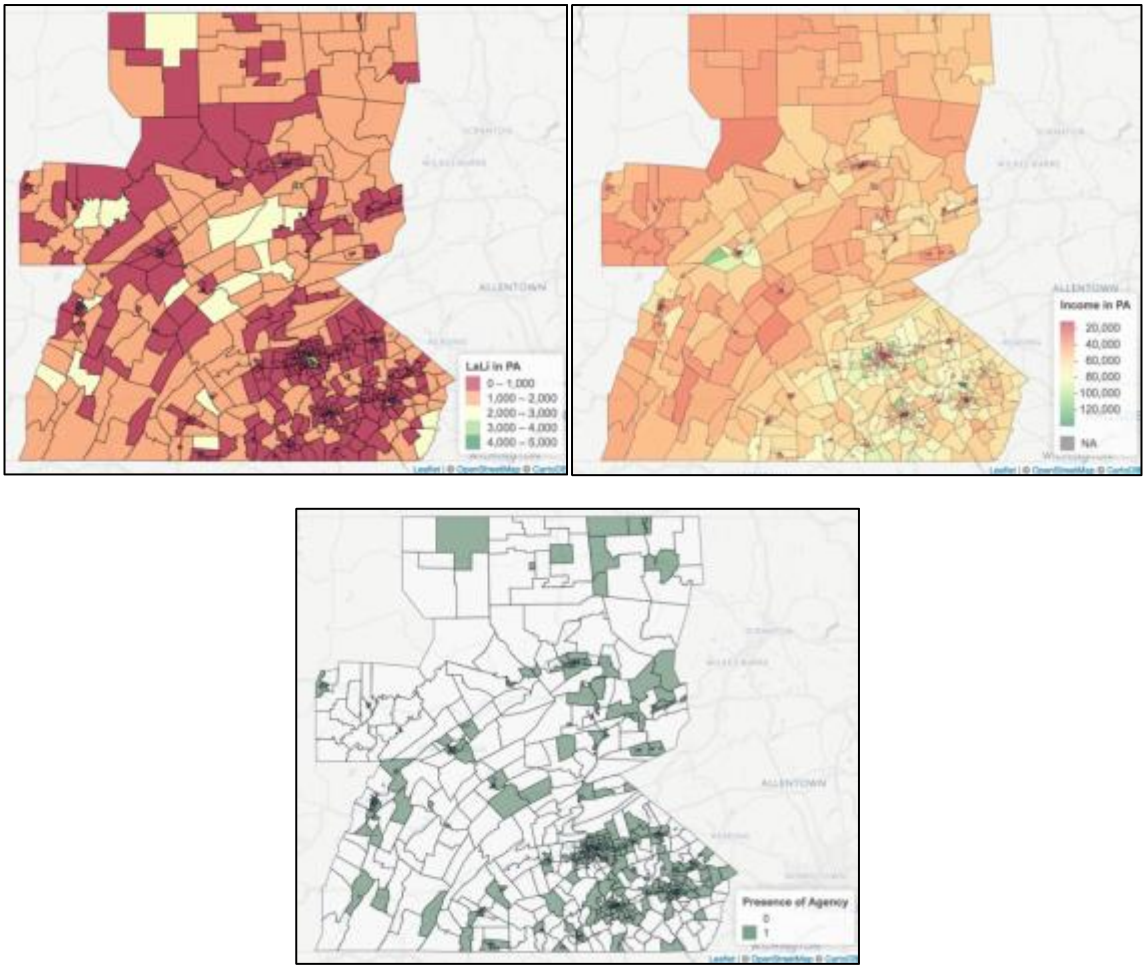

Figure 5-13: Low access low-income population (top-right), household income (top-left) and presence of a food agency (bottom) at the Census Tract level

The above figure (Figure 5-13) helps us visually understand the effect of different factors on different Census tracts.

5.4 Spatial Analysis in Central Pennsylvania

The spatial analysis for the 27 counties with the data at the Census tract level were performed and mapped in order to understand how the demographics and the county were related at the tract level. The different types of counties are discussed in this section.

In the case of Adams County, the median household income, the number of people with low access and low income within 1 mile and the distances from each Census Tract (centroid) to the nearest food agency were calculated and visualized in Figure 5-14.

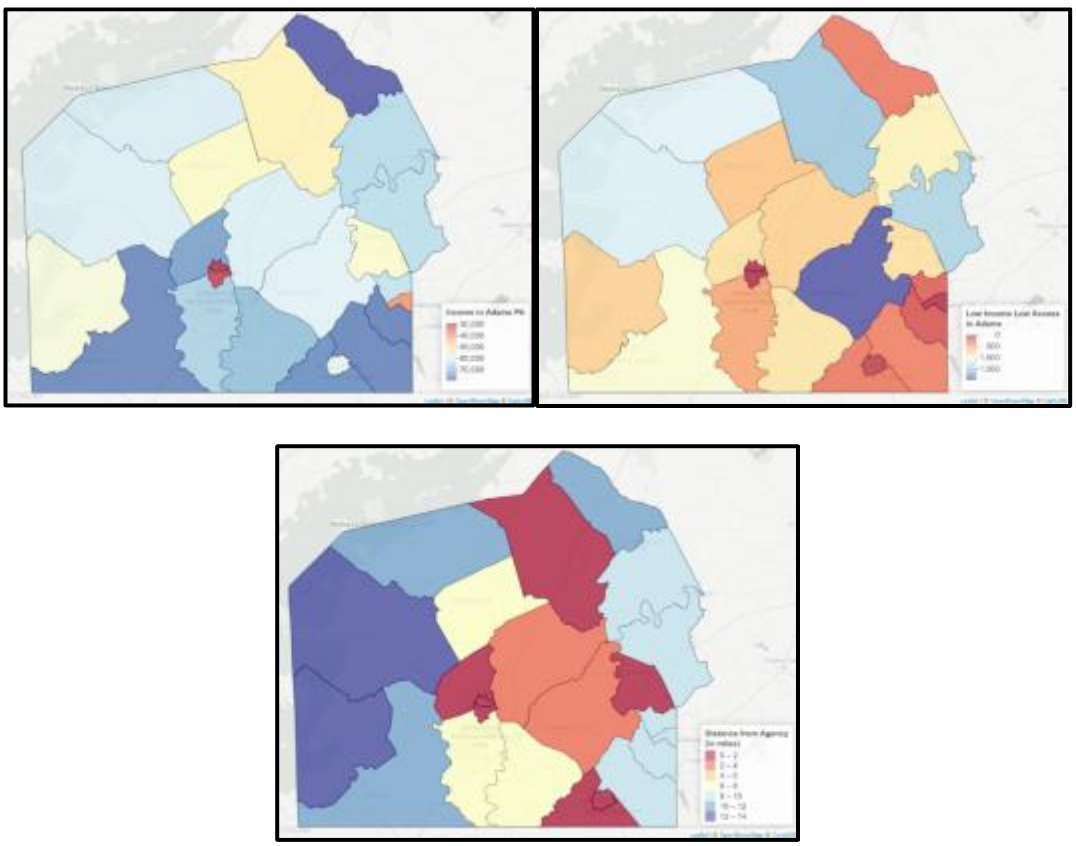

Figure 5-14. Maps of Adams County showing Median household income (top-right), low access low-income people (top-left) and distance from nearest agency (bottom)

In the case of York County, the demographics are visualized similar to Adams County in Figure 5-15.

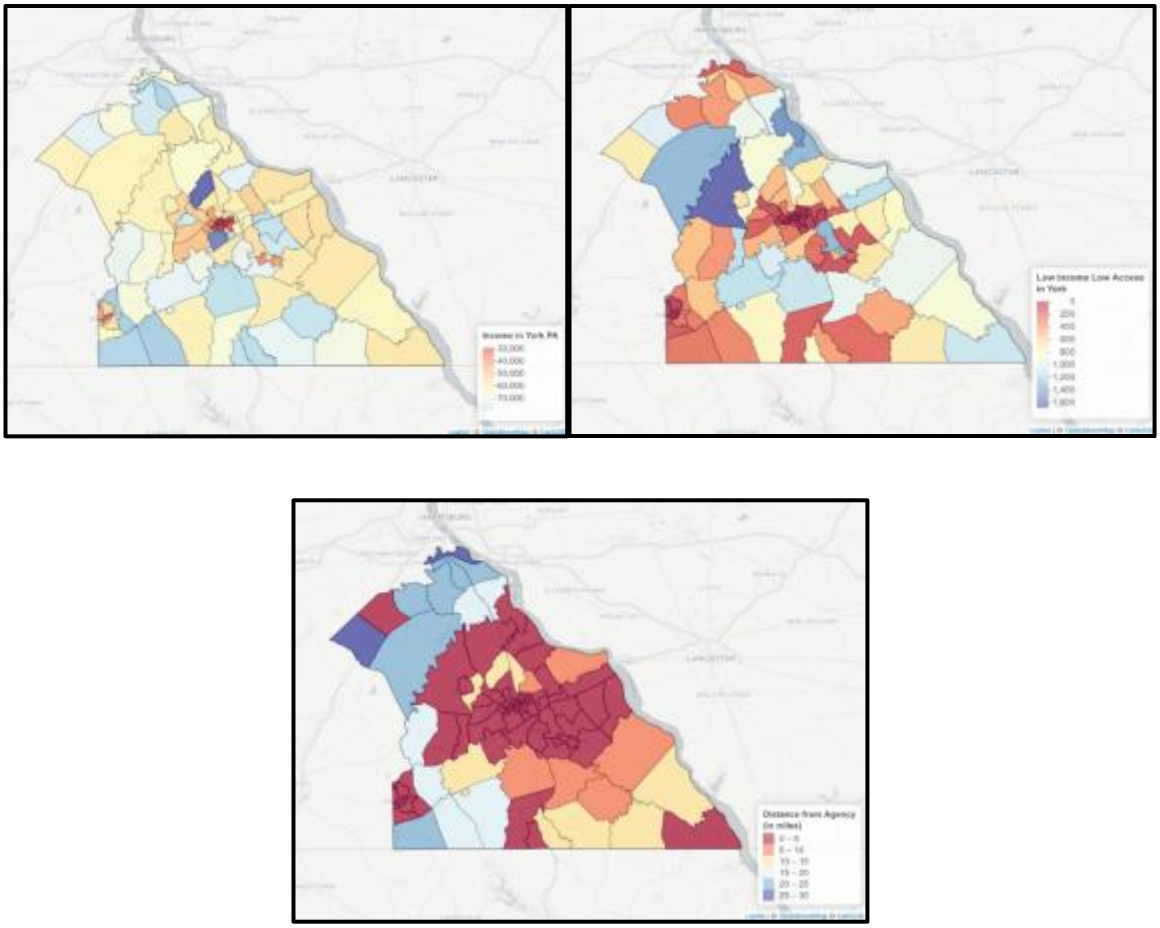

Figure 5-15. Maps of York County showing Median household income (top-right), low access low-income people (top-left) and distance from nearest agency (bottom)

5.5 Summary of Analysis

This section summarizes the analyses done in order to answer the questions that were posed at the beginning of this chapter. The identification of consolidation points was done using the brute force of descriptive analysis using the population and the distance from the nearest agency. This analysis was done in order to identify where the census block groups and food agencies were located within the worst affected counties (identified by analyzing meal gap in 5.1.2) and provide a rough solution to improve the process of distribution through the Food Bank.

The descriptive analysis of the meal gap was performed with respect to the warehouse catering to the needs of the respective county. This revealed the amount of food (in lbs) required to satisfy the meal gap per day in each county. The type of food distributed across the food agencies were also identified to observe the existence of equity in distribution and its implications to healthcare in the counties. The data was further correlated with the number of people with low access and low-income to resources in each county.

The visual analysis approach was used to understand the relation between the Census Data at the tract level and the additional amount of food (in lbs,) required in each County. The spatial analysis performed on the different counties showed the dominance of certain attributes in identifying the tracts that were in need over other attributes.

Chapter 6

Results and Conclusions

Analysis of Meal gap data alone helped in understanding the operations of the Central Pennsylvania food bank. Exploratory data analysis revealed that:

- There is a demand-supply issue;
- Low distribution of fresh food to the regions;
- The right set of factors that should be used to analyze the demand-supply problem are median household Income and the distance from the nearest agency.

The datasets used for analysis were combined interchangeably to identify the relationships between them and approach the problem with a different method in each case. Chapter 5 discussed each stage of the analysis and the repetitive cleaning of data based on the results obtained at each stage. The methods used for descriptive analytics were k-means clustering, correlations and visual analytics. This chapter discusses the results of the approaches used and their benefits.

6.1 k-means clustering in the identification of consolidation points

The accessibility index (developed in 5.1.3) helped identify the frequency of block groups in Dauphin, York, Lancaster and Lebanon that had access to a food agency. Accessibility indices range between 0 and 1. It means that lower the accessibility index of a block group; lower the population in that block group and lower the distance of the block group from the closest food agency. Higher the accessibility index; higher the population in that block group and greater the distance of the block group from the closest food agency. Figure 6-1 shows a histogram of the accessibility index of the block groups. The accessibility index is used primarily to provide a summary statistics of the block groups that have poor access to the food agencies within a county. This is not used in the clustering methodology but merely provided us with a number of block groups that should be targeted within each county.

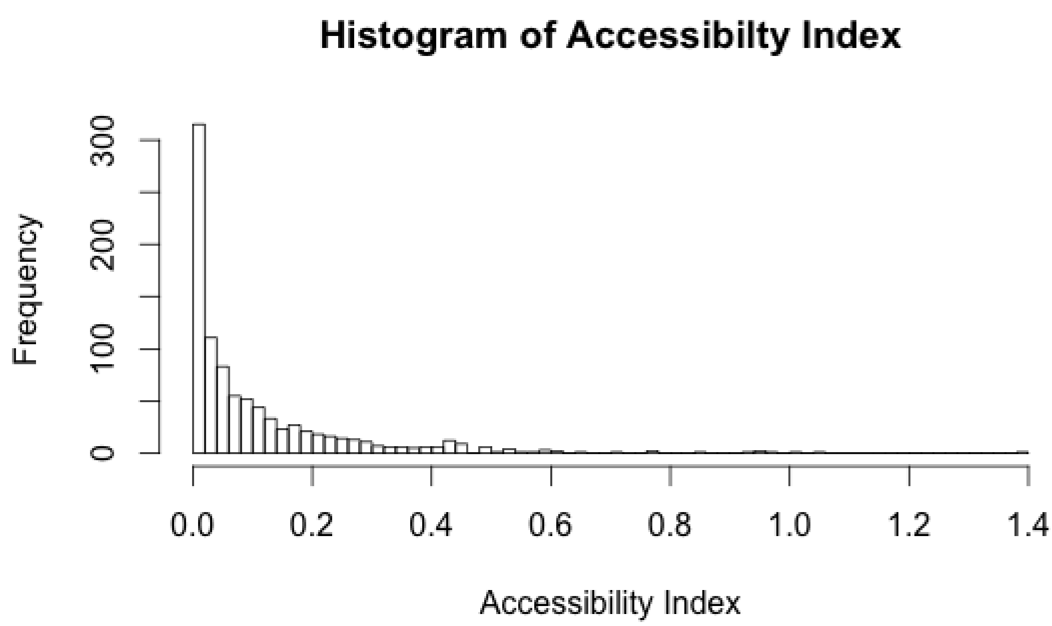

Figure 6-1. Frequency plot of Accessibility index

The plot (Figure 6-1) indicated that the majority of block groups, both covered and uncovered, have accessibility index value of 0.2 or less. Based on the 2 km threshold developed, 379 out of 923 block groups were considered to be uncovered. The accessibility index was originally developed to be gravitational factor that would account for the demand of food in the closest agency as well, similar to the one used by Guagliardo [22]. A more basic accessibility index due to the absence of demand data at the block group level.

k-means clustering was performed multiple times for varying values of $\alpha \in [0,1]$ and different number of clusters ranging from 7 to 20. The values of distortion obtained for each significant case are tabulated in Table 6-1.

Table 6-1. Distortion values for the different k-means scenarios

α value	**Number of clusters**	**Distortion**
0	7	3,634.03
	15	1,440.45
	20	1,090.63
0.25	7	3,581.34
	15	1,572.60
	20	1,092.25
0.5	7	3,668.85
	15	1,548.74
	20	1,039.77
0.75	7	3,626.41
	15	1,422.59
	20	1,069.52
1	7	3,516.04
	15	1,378.48
	20	1,150.95

Initially, a substantial reduction in distortion was noted for just a minor increase in the number of clusters. However, the reduction in distortion becomes minor after some point. In this case, it occurred when the number of clusters were 15. Hence, the number of clusters were chosen to be 15. K-means clustering was run using k=15 as the input and the output generated is color-coded by cluster as shown in Figure 6-2.

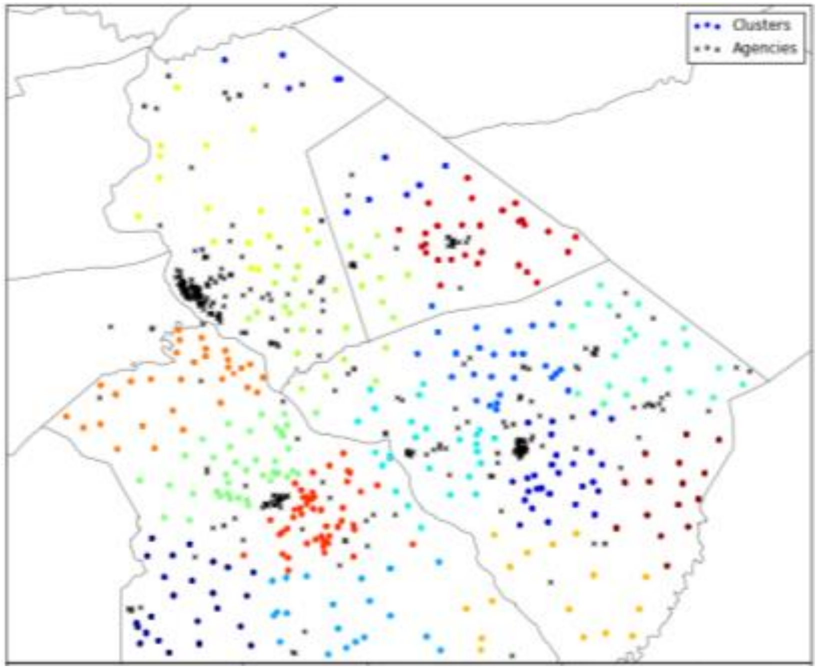

Figure 6-2. k-means clustering of uncovered block groups

Based on the cluster analysis, we determined that the consolidation points should be placed at the center of each cluster in order to effectively satisfy the demand in the uncovered block group. This analysis is very similar to the one performed by Davis et. al. [9], on identifying satellite locations to collect deliveries. However, the setting up of consolidation required considerable investment and would pose a heavy financial constraint to a non-profit organization like the Central Pennsylvania Food Bank. The results could be interpreted in an alternate way as well. The alternative to the current solution is that a food agency that is located to the center of each cluster could be used to as the consolidation point, if they have the capacity to store and distribute the resources required. The distribution of the food and its relation to USDA's research of food access are discussed in the next section.

6.2 Equity of Distribution of Food

The analysis of the distribution of food by the Central Pennsylvania Food Bank in over a year revealed the amount of food distributed to each individual county and was compared with the data obtained from USDA on food desert (tabulated in Table 5-1). This analysis resulted in the additional number of meals per day having a negative correlation (r = -0.6) with the number of people who had low access and low-income rates. The result can be interpreted as the food bank favoring counties with higher percentages of lower access and income at the expense of other counties. The analysis helped us understand that there was no equity in distribution of resources amongst the counties.

Data visualization of the food bank distribution data helped revealed that 'dry food' had the maximum distribution, compared to 'frozen food' and 'refrigerated food'. The inference could be that the food agencies did not have the capacity required to store these kind of food, since they were more likely to be local churches or soup kitchens. There is a need to increase the amount of fresh food ('refrigerated food') distributed, in order to improve the health conditions of the people with low-income and low access to resources.

The results of this analysis bore a very significant resemblance to the second hypothesis proposed by Waity [14] (discussed in 2.3) on "responsive community". The "responsive community" hypothesis suggested that the food agencies are more concentrated in regions with higher percentages of the population having low-income and poor accessibility to resources. However, the next section will discuss that all the counties in Central Pennsylvania do not essentially follow the "responsive community" hypothesis.

6.3 Impact of Visual and Spatial analysis in Central Pennsylvania

The spatial analysis shown in Figure 5-14 show that Adams County has tracts, which have considerable low access low-income people, low household income and they are very close to the agency. These tracts are in support of the “responsive community” statement, mentioned by Waity [14]. There were an equal number of Census tracts within Adams county that have relatively less number of people with low access and low-income but are extremely far from the nearest agency, which support Waity’s “spatial mismatch” statement [14] (discussed in 2.3).

The results of spatial analysis in York County (Figure 5-15) showed that majority of the tracts in York County were in support of the “responsive community” statement (discussed in 2.3) [14]. Similar analyses were done on the other counties as well. The results of the analysis are tabulated in Table 6-2. This leads to think of how much weight each of the attributes need to be given in order to prioritize the demands of each county.

Table 6-2. List of Counties in Central Pennsylvania with supported statements [14]

County	**Statements supported**	**Food bank assigned**
Adams	“responsive community”	Harrisburg
Bedford	“responsive community”	Harrisburg
Blair	“spatial mismatch”	Harrisburg
Bradford	“spatial mismatch”	Williamsport
Centre	“spatial mismatch”	Williamsport
Clearfield	“responsive community”	Williamsport
Clinton	“responsive community”	Williamsport
Columbia	“responsive community”	Williamsport
Cumberland	“spatial mismatch”	Harrisburg
Dauphin	“responsive community”	Harrisburg

Franklin	“responsive community”	Harrisburg
Fulton	“spatial mismatch”	Harrisburg
Huntingdon	“spatial mismatch”	Harrisburg
Juniata	“responsive community”	Harrisburg
Lancaster	“responsive community”	Harrisburg
Lebanon	“responsive community”	Harrisburg
Lycoming	“responsive community”	Williamsport
Mifflin	“responsive community”	Harrisburg
Montour	“spatial mismatch”	Williamsport
Northumberland	“spatial mismatch”	Williamsport
Perry	“responsive community”	Harrisburg
Potter	“responsive community”	Williamsport
Snyder	“spatial mismatch”	Harrisburg
Sullivan	“responsive community”	Williamsport
Tioga	“spatial mismatch”	Williamsport
Union	“responsive community”	Williamsport
York	“responsive community”	Harrisburg

The results of Table 6-2 indicate that the majority of the Census tracts in the counties both Harrisburg and Williamsport warehouses support the “responsive community” hypothesis (null hypothesis), but there are significant number of Census tracts in other counties that support the “spatial mismatch” (alternate hypothesis) [14]. The spatial analysis revealed that the additional food (in lbs.) in the counties that support “spatial mismatch” [14] alone contribute to approximately 39.5% of the total additional food required in Central Pennsylvania. This meant that some of the counties had food agencies situated in regions where they had resources to sustain. The results of the visual and spatial analysis also revealed that the median household income and the distance from the nearest food agency were the key indicators affecting the equity of distribution and should be used to identify the priority of the Census tract.

6.4 Conclusion

Different methodologies such as k-means clustering, descriptive, visual and spatial analysis have been used to approach the problem. The results of each analysis have been detailed in the previous sections. This section describes the summary of conclusions, limitation of each methodology used, and the future work to be done.

6.4.1 Summary of Conclusions

- The k-means clustering helped identify where the consolidation points need to be placed in Dauphin, York, Lancaster and Lebanon to answer the demand-supply issue.
- The analysis of the type of food distributed explained the relation to the probable health conditions of the people situated in each of the affected regions. Additionally, the Food Bank favors the distribution of resources to the counties that are in close proximity of the warehouses, i.e. Dauphin, Lycoming etc. (Table 3-1).
- The analysis of the distribution revealed that the Food Bank distributes its resources to the counties that have a higher need at the expense of counties with relatively lower need, thus affecting the equity of distribution.
- There key demographic factors such as the median household income and the number of low access and low-income people in each county were used in the spatial and visual analysis to determine the Census tracts within a county that were affected and did not have sufficient resources.

- The visual and spatial analysis showed that a considerable number of the food agencies were situated based on availability of resources to support the agency.

6.4.2 Limitations of Methodologies Used

- The k-means clustering performed using the population of the block group and the distance of the block group from the nearest agency posed a financial constraint i.e. setting up of consolidation points require a lot of investment from non-profit organizations. The clustering did not account for the demographics such as median household income and number of low access and low-income people within a block group.
- The descriptive analysis of the Census data and the Meal Gap data overcame the limitations of the clustering methodology and provided the demographic details of each Census tract in detail. However, the analysis did not account for the demand data at the Census tract level.
- The visual and spatial analysis accounted for the demographics of each Census tract and helped identify the major attributes that contributed to food insecurity, however, did not provide any details on the weightage of each attribute's contribution to food insecurity.

6.4.3 Future Work

- A more sophisticated multi-feature clustering could be used to determine the Census tracts that were affected by food insecurity.
- The attributes used in the analysis could be used to generate a preemptive goal programming model that would account for satisfying demand and equity of

distribution of resources more efficiently similar to what has been proposed by Lien et.al. [21].

- A resource allocation model that takes into account the population affected by both accessibility and low-income, and the demand of tract needs to be developed.

Chapter 7

Implementation Plan

The results of thesis indicated that the demand-supply issue at the Central Pennsylvania Food Bank were due to fact that the warehouses in Harrisburg and Williamsport were biased to the counties that had better access to them. The results also revealed that were plenty of other contributing factors that were not used efficiently in addressing the problem in a very comprehensive manner. This chapter aims at discussing the implementation plan for the Central Pennsylvania Food Bank using the analysis of other major contributing factors done in this thesis.

The data gathered at the Census tract level on all the 27 counties under the Central Pennsylvania Food Bank could be used in further analysis to identify a robust solution to improve distribution, while accounting for equity of distribution in all the counties within Central Pennsylvania.

7.1 Identification of Affected Tracts

A multi-feature k-means clustering can be used in order to cluster the tracts based on different number of attributes and clusters. This clustering would provide us with clusters consisting of tracts having different variations of attributes. These patterns can be used to assign priorities to the Census tracts within different counties and enhance the equity of distribution of resources by the food bank. This method could be a suitable alternative to the solution proposed by Lien et.al. [21], due to the existence of data to backup the methodology used. The clusters could also be used to classify uncovered Census tracts (tracts that do not have a food agency within 2 miles) based on the county they are situated in. The frequency of uncovered could tell us more about why the county requires more amount of food (in lbs.).

7.2 Type of Agency in the Counties

The distribution of the different types of food agencies in the different counties need to be analyzed in order to identify the possible capacity constraints that could be hindering the distribution of food. For example, Kid's Cafe® [26] is a food distribution program aimed at distributing food to children that have poor access to food and belong to low-income families. The existence of Kid's Cafe® [26] can be correlated with the tracts consisting of low access children and determine if the programs have been situated ideally across the different counties.

7.3 Prioritization of Objectives

In any organization, it is necessary to prioritize the multiple objectives of a problem while solving it, such that the existing hierarchy organization is not affected drastically by the change proposed through the solution. Goal programming [37] is a branch of multiobjective optimization, which in turn is a branch of multi-criteria decision analysis (MCDA). This is an optimization programme that can be thought of as an extension or generalisation of linear programming to handle multiple, normally conflicting objective measures.

In the case of the Central Pennsylvania Food Bank, some of the multiple objectives that need to be addressed are:

- Maximizing the amount of food distributed (in lbs.) to minimize the Meal Gap
- Minimizing the distance travelled by the food trucks and population
- Maximizing the equity of distribution of resources from both warehouses

These objectives need to be addressed using preemptive goal programming, that provides weights to the multiple objective functions. Preemptive goal programming is generally used when there are major differences in the importance of goals.

References

[1] "Food Insecurity" [Online]. Available: http://en.oxforddictionaries.com/definition/food_insecurity

[2] USDA ERS Documentation, 2013, USDA ERS Documentation "Food access research atlas" [online] (2013) Available from http://www.ers.usda.gov/data-products/food-access-research-atlas/documentation.aspx

[3] Feeding America, Hunger In America 2014, Feeding America, Chicago, IL (2014)

[4] Tarasuk VS, Beaton GH. Women's dietary intakes in the context of household food insecurity. J Nutr. 1999;129:672–9. [PubMed]

[5] Vozoris NT, Tarasuk VS. Household food insufficiency is associated with poorer health. J Nutr. 2003;133:120–6. [PubMed]

[6] "Meal Gap" [Online]. Available: http://map.feedingamerica.org/county/2012/overall

[7] K Larsen, J Gilliland. Mapping the evolution of 'food deserts' in a Canadian city: supermarket accessibility in London, Ontario, 1961–2005. Int J Health Geogr. 2008;7:16.

[8] "Census Tract" and "Block Group" [Online]. Available: https://www.census.gov/geo/reference/gtc/gtc_ct.html

[9] Davis, L. B., Sengul, I., Ivy, J. S., Brock, L. G., & Miles, L. (2014). Scheduling food bank collections and deliveries to ensure food safety and improve access. Socio-Economic Planning Sciences, 48(3), 175-188.

[10] “Food and Agriculture Organization” [Online]. Available:
https://en.wikipedia.org/wiki/Food_and_Agriculture_Organization

[11] “Heart Disease Fact Sheet” [Online]. Available:
http://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_heart_disease.htm

[12] “Childhood Obesity Facts” [Online]. Available:
https://www.cdc.gov/healthyschools/obesity/facts.htm

[13] Gundersen C., Engelhard E., Satoh A. and Waxman E.. 2014, “Map the Meal Gap 2014: Technical Brief” [Online]. Available:
http://map.feedingamerica.org/county/2014/overall

[14] Waity, J. F. (2015). Spatial inequality in access to food assistance in Indiana. Sociological Inquiry. 86(1), 103-126.

[15] Allard, S. W. (2008). Out of reach: Place, poverty, and the new American welfare state. Yale University Press.

[16] Kissane, R. J. (2010). “We call it the badlands”: How Social- Spatial Geographies Influence Social Service Use. Social Service Review, 84(1), 3-28.

[17] Sherman, J. (2009) Those who work, those who don’t: Poverty, morality, and family in rural America. Minneapolis: University of Minnesota Press.

[18] Agrawal, D., Chen, C., and Kumara, S. (2015). Improving food access using agent based simulation. IIE Annual Conference Proceedings, 761-770.

[19] Morton, L. W., Bitto, E. A., Oakland, M. J., & Sand, M. (2005). Solving the problems of Iowa food deserts: Food insecurity and civic structure. Rural Sociology, 70(1), 94-112.

[20] Dutko, P., Ver Ploeg, M., & Farrigan, T. L. (2012). Characteristics and influential factors of food deserts. US Department of Agriculture, Economic Research Service.

[21] Lien, Robert W., Seyed MR Iravani, and Karen R. Smilowitz. "Sequential resource allocation for nonprofit operations." Operations Research 62.2 (2014): 301-317.

[22] Guagliardo, Mark F. "Spatial accessibility of primary care: concepts, methods and challenges." International journal of health geographics 3.1 (2004): 1.

[23] Reilly, William John. The law of retail gravitation. WJ Reilly, 1931.

[24] Joseph, Alun E., and Peter R. Bantock. "Measuring potential physical accessibility to general practitioners in rural areas: a method and case study." Social science & medicine 16.1 (1982): 85-90.

[25] Apparicio, P., Abdelmajid, M., Riva, M., & Shearmur, R. (2008). Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. International journal of health geographics, 7(1), 1.

[26] "Central Pennsylvania Food Bank - Fighting Hunger" [Online]. Available: http://www.centralpafoodbank.org/

[27] "Tiger-Line Shape Files" [Online]. Available: https://www.census.gov/geo/maps-data/data/tiger-data.html

[28] "tigris" [Online]. Available: https:/cran.r-project.org/web/packages/tigris/tigris.pdf

[29] "American Fact Finder" [Online]. Available: http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml

[30] "Pennsylvania Income Statistics" [Online]. Available: https://www.incomebyzipcode.com/pennsylvania

[31] "Federal Poverty Level" [Online]. Available: https://www.healthcare.gov/glossary/federal-poverty-level-FPL/

[32] “K-means clustering” [Online]. Available: https://en.wikipedia.org/wiki/K-means_clustering

[33] “Visual Analytics” [Online]. Available: https://en.wikipedia.org/wiki/Visual_analytics

[34] “Spatial Analysis” [Online]. Available: https://en.wikipedia.org/wiki/Spatial_analysis

[35] “leaflet” [Online]. Available: https://cran.r-project.org/web/packages/leaflet/index.html

[36] U.S. Bureau of Economic Analysis (2013) National income and product accounts tables: Table 1.1.5. Gross domestic product [Online]. Available: http://www.bea.gov/iTable/iTable.cfm?ReqID=9&step=1.

[37] “Goal programming” [Online]. Available: https://en.wikipedia.org/wiki/Goal_programming

[38] “Overweight and Obesity Statistics” [Online]. Available: https://www.niddk.nih.gov