From a Concrete Jungle to a Concrete Farm

Nicholas Vardaro
New York University
New York, United States
njv242@nyu.edu

Jacqueline Abalo
New York University
New York, United States
jma587@nyu.edu

Simran Arora
New York University
New York, United States
simran.arora@nyu.edu

***Abstract* —**

The planet must produce more food in the next four decades than all farmers in history have harvested over the past 8,000 years [1]. Our existing agriculture and food distribution systems have accommodated global population growth and changes in consumption with varying environmental effects, and have caused unprecedented rates of land and freshwater use. Agriculture alone accounts for roughly 70% of global freshwater use. Continued expansion of areas for agriculture and forestry as a response to growing population demands and consumption is not sustainable.

New and sustainable methods for food production and distribution must be explored and prioritized, particularly for urban areas. Today, around 55% of the world's population lives in an urban area or city, and that figure is estimated to rise to roughly two-thirds or 68% by 2050 [2]. New York City, the largest city in the U.S. with a population of roughly 8.6 million people and the highest population density with over 27 thousand people per square mile, is an ideal candidate to explore alternative and sustainable solutions [3]. New York City is also home to the largest wholesale market in the world, which supplies 60% of the food for its five boroughs and feeds 9% of the U.S. population in a 50-mile radius [4].

Our proposed solution and analytics look at general food consumption patterns and New York City's distribution system. We advocate for the integration of vertical farms in the city's neighborhoods. The proximity of vertical farms to an individual neighborhood can better meet the demands of the local population and unlock access to healthy foods to those historically neglected. We see vertical farms as part of a broader sustainable solution aimed at addressing environmental, socioeconomic, and health issues.

I. INTRODUCTION

Farmland, which provides subsistence, is vulnerable to the effects of our continued expansion as well as ongoing climate change and weather extremes. Human-induced land degradation has contributed to increasing greenhouse gas emissions, loss of natural ecosystems, and decreases in biodiversity. Continued expansion may accommodate future demand, but at the expense of more permanent degradation, and ultimately exacerbate challenges for future food production. Vertical farms, positioned as a viable, sustainable solution, alleviate land degradation, and can help reshape our food systems.

Vertical farms grow produce and plants in a vertical space, and are arranged to maximize the use of a location's square footage. The entire production process, from seedling to retail, is integrated into a single system. Because most vertical farms are set up in structures such as buildings and shipping containers, a controlled artificial environment is required. Controlling the environment helps ensure plants get the required level of nutrition and incubates them from the elements, resulting in increased crop yield and higher nutritional value. Vertical farms yield high-quality produce, with no need for pesticides or preservatives, at minimal cost to the environment.

A major selling point with vertical farms is locality. Integrating vertical farms directly into urban areas allows consumers to "buy local" and reduces food miles. Food miles are the distance food is transported from production until the end consumer. Food miles vary by product, but with preservatives, transporting food from long distances is economically viable. As investments continue in vertical farms, economies of scale and innovation will drive costs down, therefore making a more local approach attractive.

Most supermarkets in New York City currently facilitate food distribution to their respective stores through the Hunts Point Terminal Market. Hunts Point Terminal Market, located in the Bronx, is one of the best-kept secrets in New York City. Hunts Point Terminal Market is the largest wholesale produce market in the world with 2,500 fruit and vegetable grocers interacting daily, 210 million packages of produce being handled annually, and $2.3 billion in sales being generated annually [4]. 15,000 trucks travel to Hunts Point every day to deliver and pick up food products. 12,000, or 80% of those trucks are heading outbound, often to a retail location [5].

Retail partnerships are critical to a farm's success. Selling in mass quantities to supermarkets reduces the volatility of demand and allows farms to focus on fewer customers. As part of day-to-day operations supermarkets cope with the volatility of consumer demand. Having a local vertical farm nearby could help a supermarket better meet the demands of its customers and reduce waste.

Vertical farms can grow almost anything, but do have specific challenges. The top products that a vertical farm can produce are leafy greens, tomatoes, peppers, and microgreens. If a product can be delivered faster, fresher, and with more nutritional value, consumers would most likely prefer that product. Faster and fresher delivery is possible with vertical farms.

II. Motivation

The world's population is expected to increase by more than 2 billion people in the next 30 years [6]. However, the amount of land and water available for agriculture will not increase [7], and is more likely to decline due to urbanization, land erosion, and pollution [8][9][10][11]. Industrial agriculture consumes large amounts of pesticides, mineral fertilizers, energy, and freshwater resources and produces large volumes of greenhouse gas emissions. The use of pesticides is a particular concern. Poisoning by pesticides and other agrochemicals is one of the leading causes of accidents for agricultural workers. Moreover, runoff and seepage of pesticides pollute water bodies.

A large share of human-induced greenhouse gas emissions results directly or indirectly from agricultural production and the subsequent processing, storage, transport, and disposal of food [12]. In the U.S., 11% of the food system's greenhouse gas emissions are due to transportation. Final delivery from producer or processor to retail only accounts for 4% of the system's emissions [13]. Therefore, reducing food miles may not have as significant an impact on emissions as one might think. 83% of emissions occur before food even leaves the farm, the largest culprits being meat and dairy. In the U.S., 42% of agricultural emissions come from animals [14]. Many experts have concluded that a shift in consumer habits towards plant foods is the best way to counter these emissions. By bringing plant foods closer to people, vertical farms make them more accessible, and facilitate the necessary shift.

New York City, being one of the densest urban areas in the world, does not have enough land for agricultural use. The lack of land means food is transported into the city. 325 million tons of cargo enter, leave, or pass through the city every year [15]. According to the 2017 Urban Mobility Report, New York City ranks second in congestion delay in the U.S. [16]. The truck congestion cost, which is the value of increased travel time, operating expenses, and new diesel consumed, was $1.8 billion. The total congestion cost, which is the value of delay and fuel cost, was $15 billion. Excess fuel consumed was 323 million gallons, and the extra time-traveled because of congestion was 811 million hours. Traffic congestion is a city-wide issue, and the constant stream of trucks to and from Hunts Point Market notably make an impact.

To manage the challenges of urban freight delivery, the Department of Transportation has introduced a congestion charge, under which trucks would need to pay about $25 to enter the congestion pricing zone (below 60th street) [17]. Business owners argue that the cost of the plan, which takes effect in 2021, will be borne by consumers. Growing more food within the city through vertical farming can limit the impact of new legislation on consumers.

Another primary motivation is consumer's quality of diet. Hunger is a lack of calories, but a healthy diet does not only consist of enough energy, but also a balanced combination of proteins, carbohydrates, fats, and a large number of essential micronutrients. Often, an imbalanced diet consists of the excessive consumption of energy-dense foods, combined with a lack of physical activity, which results in obesity. The World Health Organization (WHO) estimated in 2014 that 1.9 billion people were overweight, with a third of them being obese. In New York City, more than half of adults are overweight (34%) or obese (22%). Malnutrition can have grave or fatal consequences, especially for infants and pregnant women. Therefore, the key to a balanced and healthy diet lies in the cultivation and consumption of a range of plants and other products with different vitamins and minerals. Again, increasing access to nutritious food can help to improve healthy lifestyles and reduce the incidence of malnutrition. Vertical farms can help address existing deficiencies.

Another motivating factor is the age of U.S. farmers. The average age of all U.S. farmers in 2017 was 57.5 years, up 1.2 years from 2012. That is seven and a half years older than the average age of a farmer in the early 1980s and 17 years older than the average age of an American worker. The continued trend of aging for the U.S. farmer population was something we learned about on our visit to Square Roots. Square Roots is an urban farming company in Williamsburg, Brooklyn. With the world population expected to increase by more than 2 billion people in the next 30 years [6], there are real concerns about having enough farmers to feed the planet.

One big reason that new young farmers are not entering the field is because of costs. Younger farmers cannot afford to buy in or accumulate land, equipment, insurance, fuel, and supplies to be able to run their own farm. In order to address the shortage of farmers, we believe that the demographics and skillsets of the position must evolve. With the adoption of vertical farming, there is a huge opportunity to change the profile of a farmer and integrate big data and data science. These skills are not traditionally associated with farming but could be critical to address demand in the future. In response to the continued trend of aging, Square Roots has established their own farmers' curriculum focused on plant science, controlled environment agriculture, business and leadership.

III. Related Work

Today, people are rightfully questioning and demanding more local engagement in the production of their food. For a long time, the idea of an expanded supply chain went unchallenged; however, an expanded supply chain regularly neglects environmental impacts, which can include an increase in greenhouse gas emissions and a waste of resources. On top of environmental impacts, expanded supply chains possess other risks such as inconsistent quality of products, government intervention, local disruptions, and unpredictable delivery times. Most firms use nonlocal suppliers, production facilities, and distribution channels to secure the cheapest, most efficient, and flexible solution.

Food distribution is no exception to the outsourcing and globalization trend. More than half of the fresh fruit and almost a third of the fresh vegetables Americans buy now come from other countries. According to a recent Agriculture Department report, fresh produce imports will rise 45 percent from 2016 to

2027, meaning three-quarters of our fruits and almost half of our vegetables will be imported [18].

Vertical farms can be a solution and help facilitate local distribution [19]. As advocated by the authors of the paper, going local and growing produce and vegetables using advanced agriculture technology can address complications of a global supply system. Ideal locations within specific geographical regions can optimize the distribution of locally grown food. The challenge of finding optimal locations is similar to a study done in Flint, Michigan, an urban area which has lost five grocery store chains since 2010. In this study, which was done for small-scale healthy food retail (SSHFR) interventions [20], the author recognizes that there are more than individual factors in people's food choices. Inaccessibility of healthy food options also plays a part.

Areas that have limited access to affordable and nutritious food are defined as food deserts [21]. Adi Segal discusses inner-city food deserts by doing a case study on Harlem in New York City. This study helps uncover the causes and impacts of food deserts. The paper sheds light on the following facts and figures:

1. Poor access to supermarkets and healthful foods is greater in lower-income and minority neighborhoods, whereas residents with better access to supermarkets tend to have healthier food intake.
2. People who are not allowed to eat healthy are often the same people on Medicaid.
3. A 2000-calorie diet would cost \$3.52 a day if it consisted of junk foods, compared with \$36.32 a day for a diet of low-energy-dense foods.
4. According to the World Watch Institute, food, on average, travels 1,500 miles from farm to fork.
5. Studies show, local organic farming, within 100 miles of a city, can produce better food at a lower price.

All the above makes New York City a hotspot for further research. Considering these factors, we try to classify and distribute vertical farms and minimize the stated problems.

Food deserts serve as motivation for identifying locations for SSHFR interventions. In [20], five key factors are used as a measure: healthy food availability, socioeconomic distress, population density, proximity to bus-stops, and proximity to neighborhood centers. A particular location's suitability for SSHFR interventions is scored using a weighted sum of the measures of each factor. The challenge is how to assign weights that represent the relative importance of each factor. The author argues that expert knowledge from people based in a particular area, combined with rigorous analytical methods, are needed to determine the weights. The Flint, Michigan study therefore used the analytic hierarchy process (AHP). In this method, variables are stacked against one another and assigned a score which indicates how much more one variable matters than the other. Results are recorded in a matrix, from which the weight of each factor is mathematically derived. Using this process, the author averages inputs from eleven different experts to settle on the final weights. Optimal locations for vertical farms can also consider these five factors. Moreover, the analytic hierarchy process and a weighted sum of factors can similarly be used to define the analytic.

Understanding the evaluation process of the project's results is extremely useful. Having had experts compare the importance of the five factors for the AHP, the author asked them to mark areas on a map which they believed were most in need of SSHFR interventions. The author then identified areas where at least two of the experts' markings overlapped. Once the weights from the AHP had been used to score all locations in the Flint urban area, they checked to see if the areas with the highest scores coincided with the areas marked by multiple experts. Indeed, many of them did.

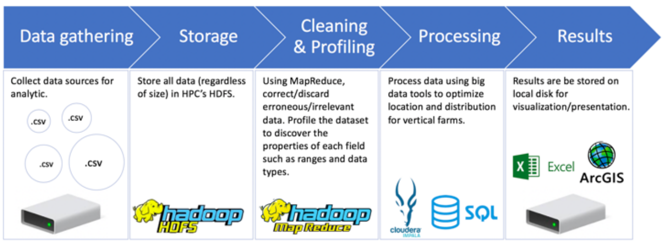
As part of establishing a vertical farm, it is important to understand the population it will be servicing. By analyzing the population distribution, allocation and area covered by retail food stores, we can understand the demand and supply dynamics of that area. The total number of retail stores and area they cover can help identify the catchment zone of a neighborhood. Identifying the catchment zone was part of a study analyzing a waterborne transportation system for Hunts Point Terminal Market [5]. The study identified specific delivery sites for waterborne vessels based on outbound distribution information and estimated weekly distribution volumes. While the study advocated for using intermediate delivery sites to alleviate congestion, we advocate for moving production to the final destination. Relocating production would further reduce food transportation miles and have a greater impact than solely leveraging a waterborne system.

Another example of understanding the optimal catchment zone is for a Health Information Exchange (HIE) organization [22]. The geographic distribution of the patients served by one HIE organization in New York City was the objective of the cluster analysis. In this study, it was confirmed that most patients who seek care at members of the New York Clinical Information Exchange (NYCLIX) live within a well-defined area. The study also confirmed a decrease in patients visiting NYCLIX sites from long distances. The author points out that understanding the geographical distribution of patients in the regional health information organization (RHIO) can inform the decision of adding new participant organizations in surrounding areas. Through our analysis, we try to identify the demand that grocery stores can accommodate which will help us identify the catchment zone for the vertical farms.

Although this project focuses on vertical farming in a single city, the concept could apply on a global scale. With the threat of climate change, and efforts to strengthen traditional farming [23], the pressure to put more land into farming and keep up with global population demands has resulted in the cutting down forests. This predictably exacerbates global warming and with climate change and unsustainable agricultural practices depleting farmland, food insecurity grows. The development of genetically modified plants to resist the effects of client change might present a short-lived victory. However, that victory would most likely be temporary because pests and pathogens eventually evolve and gain the upper hand.

Vertical farms stand up to these threats in a different way. For one, plants are much more strongly protected from pests and diseases. LED lights also recreate optimal conditions for plant growth. Hydroponic and aeroponic technologies combined with the reuse of greywater to help vertical farms use up to 70% less water than traditional ones. Using these solutions means there is no requirement for fossil fuels to harvest, transport, or refrigerate. Spoilage, from excessive handling, can also be avoided.

IV. DESIGN AND IMPLEMENTATION



The development of each analytic to support our proposal followed the same general process. Input data, either in a tab or comma-delimited file format, was put into the Hadoop Distributed File System after being copied to DUMBO, NYU's Hadoop cluster. In the data profiling stage, MapReduce jobs were submitted to determine properties such as the counts, ranges, and the number of invalid values. In the data cleaning phase, irrelevant columns and invalid data were removed. Data was also refined to only apply to the region of interest, New York City. Once data was clean, analytics were developed using Impala, and finally, visuals were created using tools such as Microsoft Excel and ArcGIS Map. Impala was chosen as a big data tool for analysis not only for its simplicity, but also because the size of the datasets made them ideal for in-memory processing. MapReduce jobs, leveraged in the cleaning and profiling phases, were written in Java and Python. The language chosen was based on the programmer's preference and language tools. The following sections describe the details of the analytics and data preparation.

Based on population and annual fresh food availability per capita, we calculated annual consumption of leafy greens in pounds, for all neighborhoods in New York City. We obtained neighborhood population information from the American Community Survey (ACS) dataset, and food availability from the United States Department of Agriculture (USDA). Food availability was used as a proxy for consumption. The leafy greens whose consumption we calculated include: spinach, kale, escarole, romaine lettuce, leaf lettuce, head lettuce, mustard greens, turnip greens, and collard greens. By summing pounds consumed of each of these leafy greens, we obtained the annual leafy green consumption for each neighborhood.

According to Artermis (formerly Agrilyst), a company that offers its analytics platform for vertical farms, the average annual yields of leafy greens are 3.75 pounds per square foot for container farms and 5.45 pounds per square foot for indoor hydroponic vertical farms. Dividing the annual leafy green consumption by these values informs the required size, in square feet, of an indoor vertical farm, and of a container farm. Using the dimensions of a standard 40 ft x 8.5 ft x 8 ft container, we also estimate the number of containers needed. Farms may leverage the available space of a shipping containers differently. Therefore, we calculated the number of containers needed assuming various set-ups – plants growing on two long walls, plants growing on two long walls and a short wall (back of the container), and the like.

The second analytic we calculated was the suitability of each neighborhood for healthy food interventions. For our research, the intervention would be the introduction of vertical farms. Our analytic uses the same methodology described in [20], except we only leveraged three factors, not five. Those three factors are socioeconomic distress, healthy food availability, and population density.

Socioeconomic distress was calculated based on rate of low education attainment, unemployment, poverty, and lone parenthood. For each neighborhood, socioeconomic distress is the sum of the z-scores of each of the rates. These rates were based on data in the social and economic ACS data tables. Socioeconomic distress is divided into five quantiles. Neighborhoods in the least distressed quintile are assigned a score of 0, 2 to the second least distressed quintile, 5 to the average distressed quintile, 8 to the second most distressed quintile, and 10 to the most distressed quintile.

Population density was calculated by combining the population statistics from the demographic ACS table with area statistics from the Neighborhood Tabulation Areas (NTAs) dataset. Using a score of 0 for the least dense neighborhoods, and 10 for the densest neighborhoods, the population density was also divided into quintiles.

We measure healthy food availability as the number of square feet of grocery stores and supermarkets per person in each neighborhood. To derive this information, we need to know which grocery stores belonged to which neighborhood. The New York City Department of City Planning (DCP) defined grocery stores or supermarkets as those in the Retail Food Stores dataset with over 5,000 square feet and whose primary purpose was the sale of food for consumption at home. During the cleaning stage, we limited the dataset to these constraints. Based on the latitude and longitude coordinates of each grocery store and the boundaries in the NTAs dataset, we determined the neighborhood for each grocery store. For stores that were missing latitude and longitude coordinates in the Retail Food Stores dataset, we used the Google Maps geocoding API to retrieve their geographic coordinates based on the address. Again, healthy food availability was divided into five quintiles, with a score of 0 for the quintile with the most grocery store square feet per person, and 10 for the least.

Using the weighted sum of the scores for socioeconomic distress, population density, and healthy food availability, the final suitability score was calculated. The weights we used were those in [20], which were 0.46, 0.22, and 0.13, respectively.

The suitability score serves as an indicator of which neighborhoods should be prioritized for the development of a vertical farm. Another simple indicator is the deficit in grocery stores in a neighborhood. The Department of City Planning stipulates a standard of 30,000 square feet of grocery stores and supermarkets for every 10,000 residents. With population statistics, it was straightforward to calculate the deficit as the square footage of grocery stores a neighborhood should have (according to the stipulation) minus the square footage of grocery stores a neighborhood contains.

If vertical farms can be integrated into individual neighborhoods, a big opportunity is big data and advanced analytics. Big data and analytics can help optimize food production and minimize food waste. According to McKinsey, 38% of food waste occurs during consumption in developed economies and 32% during production in developing economies. For New York City, having crop production close to the neighborhood it serves could lower food miles, minimize waste, and better meet the demands of its residents.

One noteworthy example of data and analytics combined with food consumption is Instacart. Instacart is an on-demand grocery delivery platform available in most major U.S. cities. The platform is robust enough to be able to deliver groceries to customers in as little as one hour after an order is submitted. Instacart, as a company, has grown substantially in the past several years as has the overall online grocery market. Online groceries currently make up a small percentage of the market, but sales continue to increase. The U.S. online grocery market grew from $12 billion in 2016 to $26 billion in 2018. Online grocery sales are predicted to hit $100 billion by 2025. Walmart currently dominates the online grocery market, but Instacart is in second ahead of Amazon and Shipt (a company recently acquired by Target).

In 2017, Instacart released a dataset with over three million orders for over two hundred thousand customers. This dataset, was normalized and split into six separate files. As part of our analysis, we calculated analytics that could help a vertical farm assess consumer demand and inform crop production.

The first analytics we calculated were orders by day of the week, orders by the hour of the day, and days since prior order. These analytics were calculated by grouping rows that have the same values and counting how many rows were in each group. Most ordered products was calculated by grouping rows with the same product ID but also required a join statement to return the product name associated with the ID. Prior to connecting the product name to the product ID, a subquery was run to return data used in the main query.

We also calculated the percentage of produce for each order. To extract this analytic, two subqueries were run to first identify the department ID of each product and then connect that information to the department name. Finally, we grouped rows based on order ID to count the number of products per order and the number of products per order from the produce department. By dividing the number of products by the number of products from the produce department, we could calculate the percentage of each order containing produce. Once that information was calculated, we needed to create groups to bucket the percentages. We bucketed the information into ten groups which provided a relatively even distribution.

V. Datasets

This section describes each dataset, including its source, size, and a brief description. A schema is provided, showing a small sample of some of the relevant fields.

A. Neighborhood Tabulation Areas (NTAs) [24]

This data contains boundaries of Neighborhood Tabulation Areas defined by the New York City Department of City Planning. The Shapefile format is about 1.8Mb in size. Each neighborhood is associated with a unique NTACode, as well as its area in square feet.

Field	Description	Type
NTACode	Unique Identifier	String
NTAName	Neighborhood name	String
Shape_Area	Area in square feet	Float

B. American Community Survey (ACS) Data Tables [25]

The ACS is an extensive nationwide survey providing estimates on demographic, socioeconomic, and housing characteristics. This survey is conducted over five years and is filtered by the Department of Planning to limit the results to New York City. The demographic and socioeconomic datasets for the 2013-2017 period are about 4.6 MB in size. The GeoId field in this data set and NTACode field in the NTAs dataset share the same domain.

Social Data		
Field	**Description**	**Type**
GeoId	Unique Identifier	String
MHnWChU18	Number of single male householder families with children	Integer
MrdFam	Number of married-couple families	Integer

Economic Data		
Field	**Description**	**Type**
GeoId	Unique Identifier	String

CvLFUEm1	Unemployed persons in the civilian labor force	Integer
PBwPv	Population below the poverty level	Integer

C. Food Availability per Capita [26]

The Food Availability per Capita dataset includes estimates for over 200 commodities, including individual fruits, vegetables, and grains, for years up to 2018. According to the Economic Research Service of the USDA which provided this information, this data serves as a proxy for actual food consumption. This project makes use of fresh vegetable food availability data, which is about 0.5MB in size. Below are some of the fields from each vegetable's table.

Field	Description	Type
Year	Year the data was collected	String
Farm	Per capita food availability at farm weight	Float
Retail	Per capita food availability at retail weight	Float

D. Retail Food Stores [27]

Retail Food Stores in New York, last updated in June 2019, is a listing of all retail food stores in the state of New York licensed by the Department of Agriculture and Markets. The size of this dataset is 7 MB. This dataset, with 29,389-rows and 15-columns, was refined to limit its scope to the five boroughs of New York City.

Field	Description	Type
Establishment Type	The purpose of the store (e.g., sale, processing, manufacturing, etc.)	String
Square Footage	Size in square feet	Integer
Location	Address of the store, sometimes accompanied by its coordinates	String

E. Instacart Online Grocery Shopping Dataset [28]

This dataset contains over 3 million grocery orders for more than 200,000 Instacart users. Six files are included in the public dataset, but only four were used for our analysis. The products and departments files are under 2 MB, the orders file is 109 MB, and the order products prior file is 577 MB.

Departments (departments.csv)		
Field	**Description**	**Type**
department_id	Unique Identifier	Integer
department	Name of the department	String

Products (products.csv)		
Field	**Description**	**Type**
product_id	Unique Identifier	Integer
product_name	Name of the product	String
department_id	Department number of the product	Integer

Grocery Orders (order_products_prior.csv)		
Field	**Description**	**Type**
order_id	Unique Identifier	Integer
product_id	Unique Identifier	Integer

Orders (orders.csv)		
Field	**Description**	**Type**
order_id	Unique Identifier	Integer
user_id	Unique Identifier	Integer
order_dow	Day of the week the order was placed (0 – 6)	Integer
order_hour_of_day	Hour of the day order was placed (0 – 23)	Integer
days_since_prior_order	Days since user's last order (0 – 30), orders greater than 30 days reflect 30 days	Integer

VI. RESULTS

As the most straightforward measure of the need for healthy food, we calculated the deficit in grocery stores per person in each neighborhood. This is shown in Figure 1. Based on the DCP's designation of 30,000 square feet of grocery stores per 10,000 people, which translates to 3 square feet per person, we divided neighborhoods into four categories: those with per person deficits less than or equal to 0, indicating no need (and a surplus of square footage), those between 0 and 1 with a low need, those between 1 and 2 with an average need, and those between 2 and 3 with a high need.

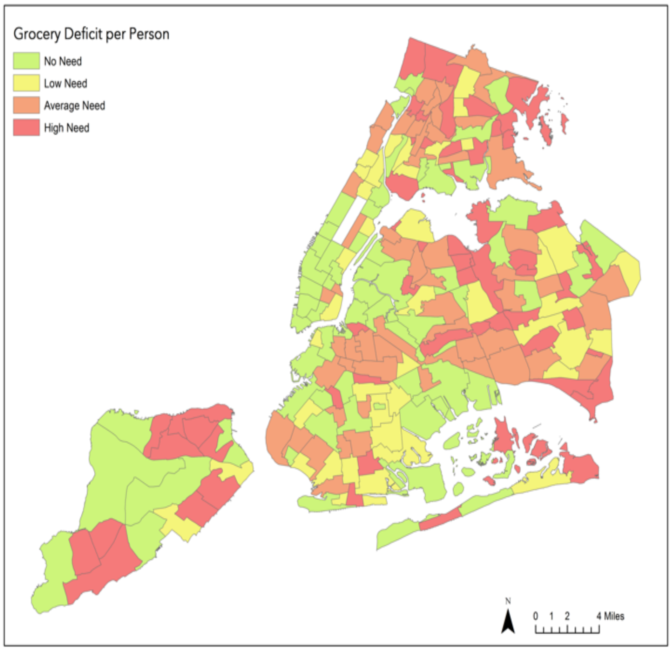


Figure 1: Grocery Stores Deficit per Person

A better measure of suitability for a vertical farm would take more factors into account. Using the previously described process from the food retail intervention study, we calculated a suitability score for each neighborhood. A higher suitability score indicates that a neighborhood is more likely to benefit from a vertical farm. Figure 2 shows each of the features that were involved in calculating neighborhood scores: socio-economic distress, population density, and square feet of grocery stores.

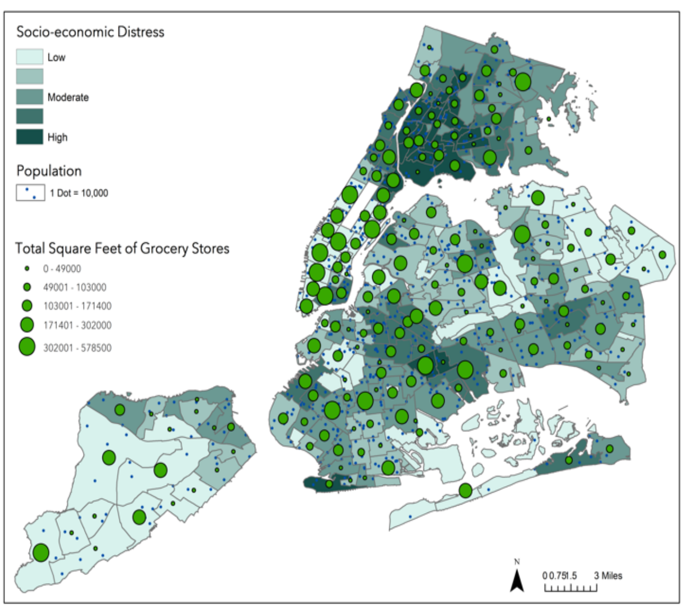
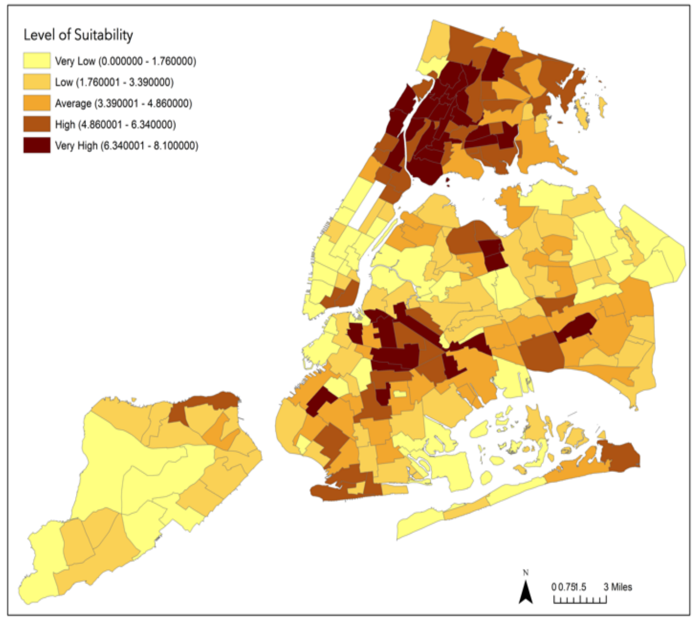


Figure 2: Suitability Score Factors

Figure 3 shows the suitability scores for each neighborhood, ranging from 0 to 8.1.



Figured 3: Suitability Scores

As was similarly done in [20], suitability zones were created to bucket neighborhoods into five different categories: very low suitability (less than 1.76), low (1.76 – 3.39), average (3.39 – 4.86), high (4.86 – 6.34), and very high (above 6.34). These buckets were chosen based on the Jenks natural breaks classification method, an available option in ArcGIS. This method aims to reduce the variance within classes and maximize the variance between classes. Six neighborhoods, four of which were in Staten Island, received the lowest possible score of 0. The other neighborhoods that received a score of 0 were in the outskirts of Queens. Areas with high and very high suitability are mostly located in the Bronx and Brooklyn. From this exhibit, it is visually apparent how higher levels of socio-economic distress, which is the highest weighted factor at 0.44, correlate with increased suitability for a vertical farm.

The goodness of this analytic can be measured in multiple ways. First, we compare our findings to those of the Supermarkets Needs Index (SNI). This was developed by the Department of City Planning and is reflected in Figure 4.

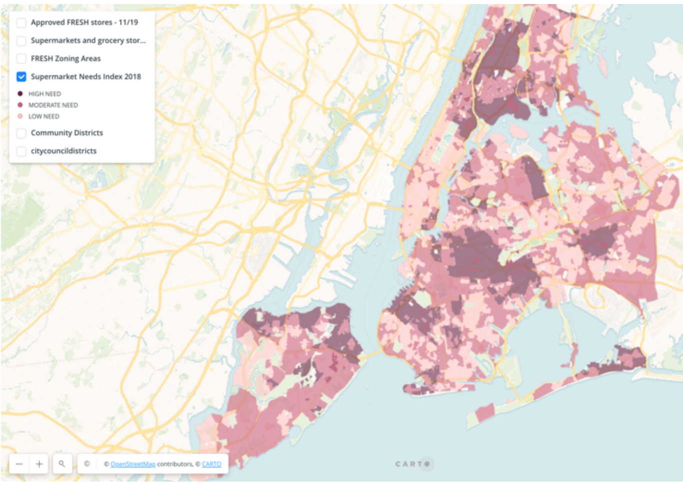


Figure 4: Supermarkets Needs Index

In analyzing both the Supermarkets Needs Index and suitability scores, many of the high or very high suitability zones determined by our analytic can be corroborated by the high need zones identified in the SNI.

Another way to confirm our analytic is the Food Retail Expansion to Support Health (FRESH) program. The FRESH program, offers zoning incentives and financial benefits as a way to introduce fresh food options in underserved areas. Figure 5 shows our suitability scores overlaid with the locations where FRESH offers zoning and/or discretionary tax incentives.

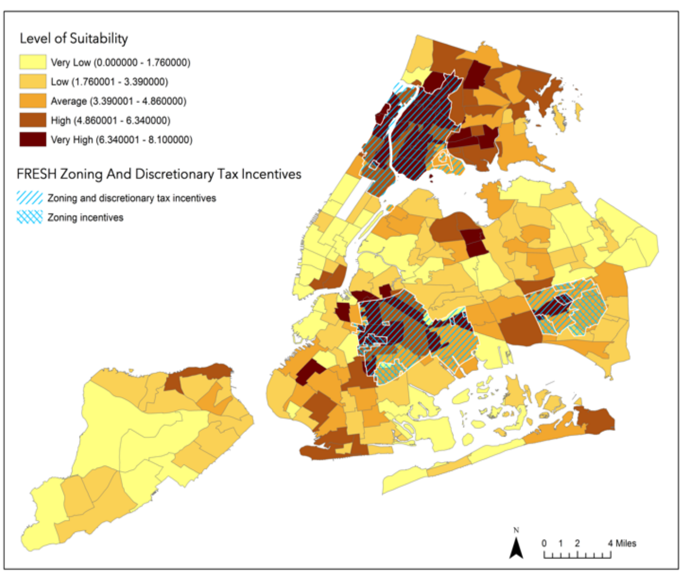


Figure 5: Suitability Scores with Zoning and/or Discretionary Tax Incentives

The FRESH zones align reasonably well with our areas of high or very high suitability, thus confirming the need to prioritize healthy food intervention in these areas. Neighborhoods with high suitability scores that fall outside of the FRESH program zones can be targeted as the city's next focus for fresh food stores.

Vertical farms should take advantage of the zoning and tax incentives and establish a presence in these areas. Doing so may also help alleviate some of the material upfront costs required to establish a vertical farm.

In order for a vertical farm to properly set up its operation it would need to know its size requirements. As part of our analytic we calculate the size, in square feet, of a vertical hydroponic farm that could support the leafy green consumption habits of a particular neighborhood. For perspective, Figure 6 shows the percentage of the neighborhood's total surface area that would be occupied by the farm, alongside the population density. The maximum proportion is 3.28% for Yorkville in Manhattan. It makes sense that areas with higher population densities would need to dedicate a greater proportion of their area to vertical farms. However, if this proves to be a burden, a tactical redistribution of vertical farms to areas with more space may relieve this burden. However, this would mean an increase in distance of food from the consumer.

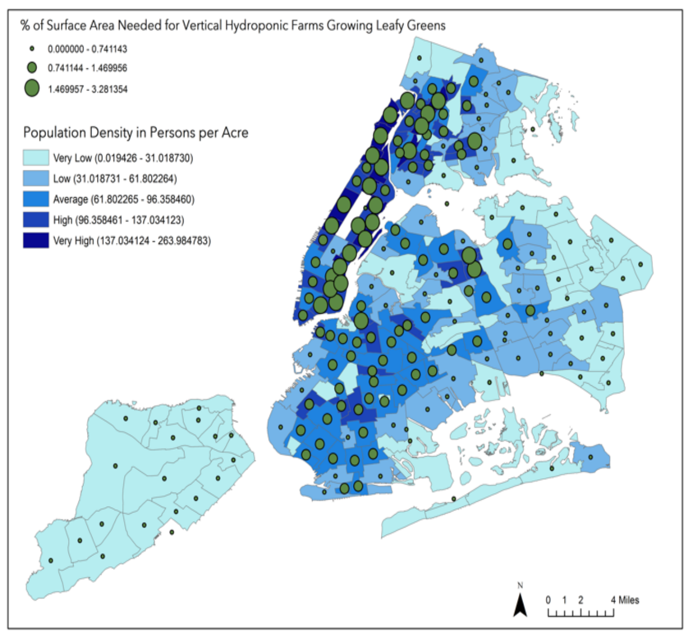


Figure 6: Percentage of Area Needed against Population Density

Container hydroponic farms are a type of indoor farming facility that does not use soil. As explained in the design section, we calculated the number of 40 ft x 8.5 ft x 8 ft containers that would be necessary to support each neighborhood's consumption. We assume five different setups of container farms. Setup 1 uses the two long walls of the container, setup 2 uses the two long walls and one of the short walls, setup 3 uses all four walls, setup 4 uses two long walls, a short wall and the floor, and Setup 5 uses all four walls and the floor. Figure 7 shows the average number of containers a neighborhood would need based on these setups. Setup 1 requires the most containers, about 518, while Setup 5 requires about 310. 518 containers would have a footprint of about 166,000 square feet. For comparison, a 290-car parking lot occupies 132,000 square feet of space.

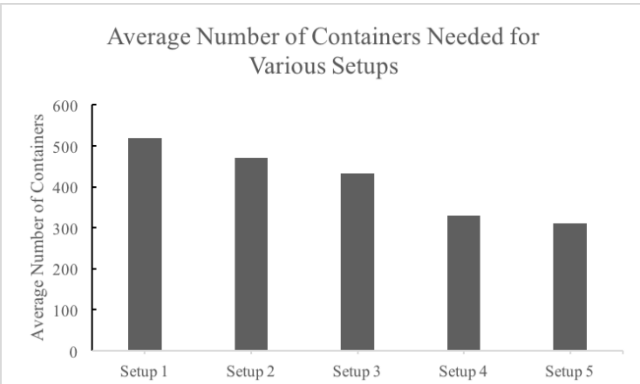


Figure 7: Average Number of Containers Needed for Various Setups

Once a vertical farm is established, it will want to understand the consumption patterns of its catchment zone. There are many interesting patterns in the Instacart analysis that provide insights into the buying habits of the company's customers. Having these insights would help inform a vertical farm's operation. The insights provide information on the products the vertical farm should grow and the frequency at which they need to grow in order to satisfy demand. Consumer habits change over time, and studying this information would help a vertical farm identify those changes.

To start, we calculated the number of products per order.

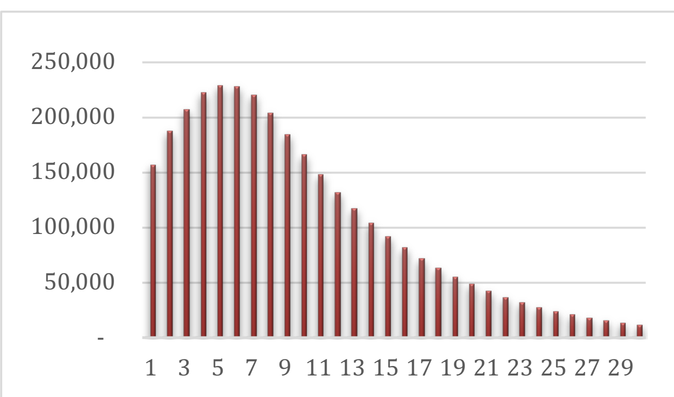


Figure 8: Products Per Order

The number of products per order varied considerably and ranged from 1, all the way up to 145. Most of the orders in the dataset had five products. Having only five products was surprising, considering Instacart's business model charges a service and delivery fee for every order. Customers usually tip the delivery person as well. All of these smaller changes add up and can negatively impact the final cost.

The next analytics were orders by day of the week and orders by the time of the day.

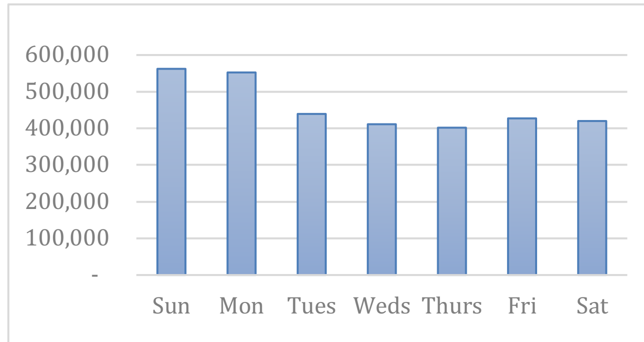


Figure 9: Orders by Day of Week

Instacart users are most likely to place their grocery orders on Sunday and Monday of each week. Roughly 35% of the orders from the dataset were submitted on those two days. We thought this pattern made sense as most customers most likely place their orders at the end of the weekend when they may not be working or at the beginning to have food for the rest of the week.

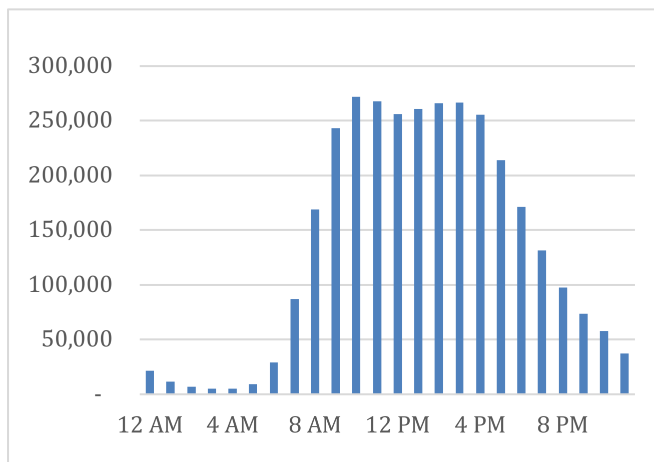


Figure 10: Orders by Time of Day

Most of the orders were placed between 10 A.M. and 3 P.M. The least amount of orders were placed between midnight and 5 A.M. Given Instacart's business model, most customers likely place their orders during the day and schedule delivery for the evening. Instacart customers appreciate the flexibility and convenience of the service. Ordering groceries during the day helps customers avoid having to go to the grocery store in the evening and gives time back for optional activities.

The next analytics was the top selling products. The three top three selling products were bananas, strawberries, and baby spinach.

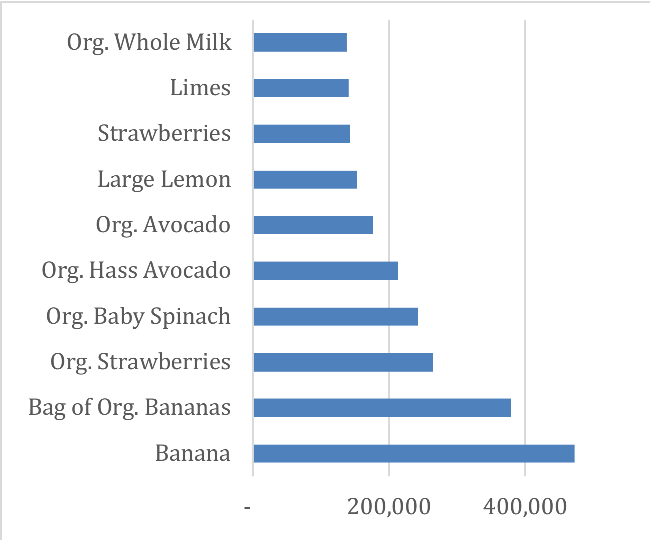


Figure 11: Top Selling Products

These three items as the top items were surprising. Many consumers like to see and touch produce. Trusting delivery personnel to select fresh produce speaks to the promise of online ordering and grocery delivery in the future. Looking at the other top-selling products, most of them are fruits and vegetables.

Interestingly, a lot of the other top items include the term organic. Organic items tend to be more expensive, but consumers are becoming wearier of the effects of pesticides. American's desire for organic foods has grown over the past few years. According to the Organic Trade Association, organic food sales in the U.S. rose 5.9% in 2018 to reach $47.9 billion. Because vertical farm production takes place in a contained environment, pesticides are not needed. However, the National Organic Standards Board advises that organic is more than just the lack of pesticides or artificial fertilizers. Certified organic means biodiversity and biological soil activity. Vertical farms often use hydroponics, which is growing plants without soil. Instead, they use mineral nutrient solutions in a water solvent. Because soil is not part of the vertical farming process, vertical farms are not always able to meet the strict definition of organic. Some agencies are allowing for organic certifications if the vertical farms can prove they only use organic inputs, but the debate over whether crops grown in a vertical farm is organic is likely to continue.

The final two analytics are days since prior order and the percentage of produce for all orders. Excluding orders placed, thirty days or higher, 57% of the orders occurred within seven days of a customer's previous order. This data point would help a vertical farm understand how often they would need to crops to meet the demand of the area they feed.

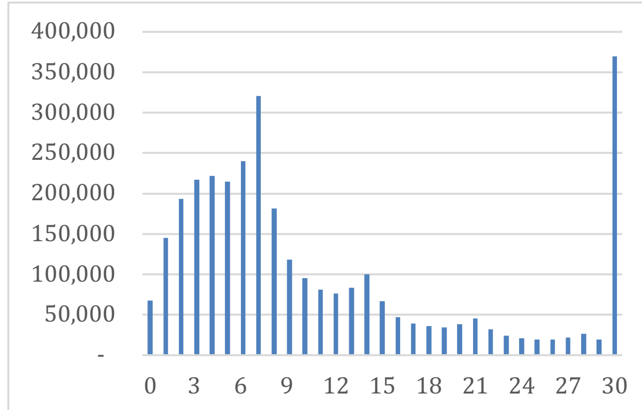


Figure 12: Days Since Prior Order

The percentage of produce for all orders highlights that 30% of all orders contain less than 10% of items from the produce department. 54% of orders contain 10% - 40% of produce.

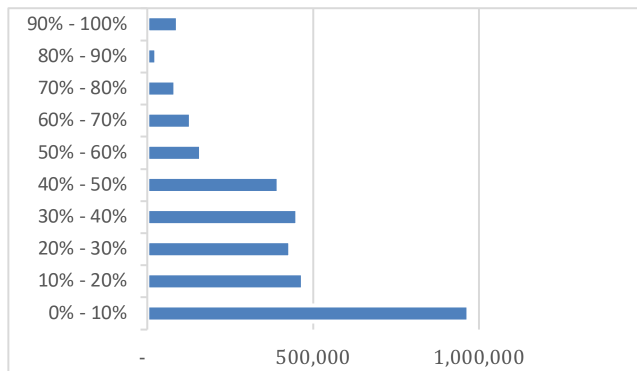


Figure 13: Percentage from Produce

The dataset does not contain the quantity of each product in an order, but based on current fruit and vegetable consumption trends, we thought this distribution made sense. In fact, according to the CDC, just 12% of American adults meet the standard for fruit consumption and 9% for vegetable consumption. Dietary guidelines by the CDC recommend that adults consume 1.5 to two cups of fruit per day, and two to three cups of vegetables per day. Americans are eating fruit once per day and vegetables 1.7 times per day.

VII. Future Work

Vertical farms are a considerable prospect to address challenges for urban areas both today and in the future. One of the biggest challenges and often the first barrier to mass implementation of vertical farms is the significant upfront costs. The set-up of a vertical farm requires substantial investments. These investments cover infrastructure expenses such as real estate, facilities, and capital equipment.

Additionally, because plants grow indoors, constant exposure to LED lights is required. Some vertical farm facilities combine natural sunlight with LED lights to limit the expense, but costs can be material. At this moment, the amount of energy required for a vertical farm and the carbon footprint it creates could outweigh the environmental benefits. One vertical farm facility may require millions of dollars in upfront investments and substantial monetary commitments to maintain the operation. Quantifying the economics of a vertical farm and its carbon footprint is what we see as the next step in this study. Climate change and other determinants will necessitate innovation in farming, but the affordability of vertical farming will eventually increase.

Vertical farms are also a massive opportunity for educational institutions. New York University (NYU) is committed to fostering and promoting sustainable business principles and should continue to lead by establishing a vertical farm. To date, the university acts on sustainability by reducing, reusing, and recycling, with a primary focus on reducing. NYU should continue to evolve past waste management and quality assurance and establish its own vertical farm. Having a vertical farm at the university could help feed the more than 11,000 undergraduate, graduate, and professional students who receive housing through the school each year. It could also serve as a vast applied research opportunity serving as the cumulation of several of the university's schools. Vertical farming involves aspects of urban planning, biology, computer science, data science, big data, and engineering.

Given that vertical farms are still fairly new, a robust set of data is not widely available. For this project, the only product for which the yield per square foot in a vertical farm was available was leafy greens. With new data on the yields for other produce, it will be possible to develop a more holistic assessment of vertical farm square footage needs for a wider range of products.

VIII. Conclusion

In summary, the diminishing availability of farmable land and the need for increased healthy food accessibility are motivators for the adoption of new and sustainable methods of food production and distribution. Vertical farms are part of a broader sustainable solution and should be further prioritized and researched for New York City. Based on factors like socio-economic distress, population density, and grocery store availability, we identify a reasonable scheme for prioritizing the expansion of vertical farms across neighborhoods. Our calculations of the square footage requirements of vertical farms for the production of leafy greens show that the solution is physically feasible, amounting to a 300-car parking lot in each neighborhood. It is also financially viable if existing financial incentives are exploited. Finally, our analysis of existing food purchase habits sheds light on the consumption pattern of the catchment area of vertical farms.

Hunts Point Terminal Market and New York City's food distribution system is unique because it caters to the largest ethnically diverse region in the world and receives deliveries by rail, truck, and air cargo from every part of the world. Dispersing distribution and production of food to different parts of the city is similar to distributing the processing of large datasets. Doing so could alleviate traffic congestion, minimize food waste, unlock access to healthy foods in neighborhoods historically neglected and decrease the reliance on Hunts Point Terminal Market as the primary food distribution channel for 9% of the U.S. population.

IX. Acknowledgment

This research and paper were part of a project for a real-time and big data analytics class. The main focus of the class was the Apache Hadoop framework and the distributed storage and computing power of the cluster system. For the class, we used the "Hadoop: The Definitive Guide" textbook by Tom White. With this comprehensive guide, we learned about large-scale data storage, data processing and data locality. Data locality, which refers to moving the computation to the data, instead of moving data to the computation, increases the throughput of the cluster system. The concept of data locality was a massive inspiration for advocating for the decentralization of Hunts Point Terminal Market. In the Hadoop cluster system, data locality avoids passing data through the Name Node and across the network. Data locality minimizes network congestion, which is a critical resource when dealing with large datasets. Our ability to store, process, and analyze large datasets was possible because of NYU's High-Performance Computing (HPC). HPC supports the sharing of high-performance clusters for students and researchers.

We would also like to thank ArcGIS and Google. We would like to thank ArcGIS for making their software available to NYU students. We used ArcGIS to map grocery deficit per person zones, suitability zones, tax incentive zones and map vertical farm size requirements. We would like to thank Google for their Google Maps API. Using their API, we were able to find the coordinates for grocery stores throughout New York City.

Alison Kopf, founder and CEO of Artemis was also helpful in clarifying our understanding of the company's report on the State of Indoor Farming. Finally, Square Roots, an urban farming company located in South Williamsburg, Brooklyn, also deserves acknowledgment. As part of a Meetup event, we took a tour and visited their facilities. We wanted to see a vertical farm in-person and ask questions about the operation. Seeing the farm was impressive and gave us the idea to use shipping containers. Shipping containers can be compared to nodes in a cluster. As one needs to distribute the load to increase processing power, one can add a node. With our use case, one can add a shipping container to increase crop production and satisfy demand.

X. References

1. (n.d.). How the Netherlands Feeds the World - National Geographic. Retrieved November 14, 2019, from https://www.nationalgeographic.com/magazine/2017/09/holland-agriculture-sustainable-farming/
2. (2018, May 16). 68% of the world population projected to live in urban areas... Retrieved November 14, 2019, from https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html
3. (n.d.). Planning-Population-NYC Population Facts - DCP - NYC.gov. Retrieved November 14, 2019, from https://www1.nyc.gov/site/planning/data-maps/nyc-population/population-facts.page
4. (n.d.). About The Market - Hunts Point Produce Market. Retrieved November 14, 2019, from https://huntspointproducemkt.com/about-hunts-point-market/
5. "17-10 Hunts Point Terminal Market - NYSERDA - NY.gov." https://www.nyserda.ny.gov/-/media/Files/Publications/Research/Transportation/17-10-Hunts-Point-Terminal-Market-Demand-Waterborne-Transportation.pdf. Accessed 3 Dec. 2019.
6. "2019 Urban Mobility Report - Texas A&M University." https://static.tti.tamu.edu/tti.tamu.edu/documents/mobility-report-2019.pdf. Accessed 3 Dec. 2019.
7. "Growing at a slower pace, world population is expected to ...." 17 Jun.2019, https://www.un.org/development/desa/en/news/population/world-population-prospects-2019.html. Accessed 3 Dec. 2019.
8. "Evolution not revolution of farming systems will best feed and ...." https://www.sciencedirect.com/science/article/pii/S2211912412000193. Accessed 15 Nov. 2019.
9. "Agriculture Crossroads Agriculture Crossroads." https://www.weltagrarbericht.de/fileadmin/files/weltagrarbericht/IAASTDBerichte/GlobalReport.pdf. Accessed 15 Nov. 2019.
10. "Reaping the benefits - Royal Society." 21 Oct. 2009, https://royalsociety.org/topics-policy/publications/2009/reaping-benefits/. Accessed 15 Nov. 2019.
11. "SOLAW: SOLAW Home." http://www.fao.org/nr/solaw/solaw-home/en/. Accessed 15 Nov. 2019.
12. "Building bridges: an integrated strategy for sustainable food ...." https://pubag.nal.usda.gov/catalog/613826. Accessed 15 Nov. 2019.
13. "Agriculture at a Crossroads - Global Agriculture." 9 Feb. 2016, https://www.globalagriculture.org/fileadmin/files/weltagrarbericht/EnglishBrochure/BrochureIAASTD_en_web_small.pdf. Accessed 15 Nov. 2019.
14. "Is Local Food Better? - GeoPathfinder." http://geopathfinder.com/IsLocalFoodBetter.pdf. Accessed 15 Nov. 2019.
15. "Meat and global warming - The New York Times." 25 Jan. 2018, https://www.nytimes.com/2018/01/25/climate/cows-global-warming.html. Accessed 15 Nov. 2019.
16. "Improving the efficiency of truck deliveries in NYC - NYC.gov." https://www1.nyc.gov/html/dot/downloads/pdf/truck-deliveries-ll189.pdf. Accessed 16 Nov. 2019.
17. "Congestion Pricing - The New York Times." 18 Apr. 2019, https://www.nytimes.com/2019/04/18/nyregion/newyorktoday/nyc-news-congestion-pricing.html. Accessed 15 Nov. 2019.
18. "Most of America's Fruit Is Now Imported. Is That a Bad Thing ...." 13 Mar. 2018, https://www.nytimes.com/2018/03/13/dining/fruit-vegetables-imports.html. Accessed 3 Dec. 2019.
19. "Global versus local: an exploration on how vertical ...." 18 Feb. 2014, https://www.researchgate.net/publication/260192500_Global_versus_local_an_exploration_on_how_vertical_farms_can_lead_the_way_to_more_sustainable_supply_chains. Accessed 15 Nov. 2019.
20. "Integrating expert knowledge in a GIS to optimize siting ...." 16 Jun. 2016, https://ij-healthgeographics.biomedcentral.com/articles/10.1186/s12942-016-0048-6. Accessed 15 Nov. 2019.
21. "Food Deserts: A Global Crisis in New York City" https://journals.cdrs.columbia.edu/wp-content/uploads/sites/25/2016/10/120-237-1-PB.pdf
22. "Geographical distribution of patients visiting a health ... - NCBI." https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3715355/. Accessed 17 Nov. 2019.
23. "The vertical farm: controlled environment agriculture carried …." 3 Dec. 2010, https://link.springer.com/content/pdf/10.1007%2Fs00003-010-0654-3.pdf. Accessed 3 Dec. 2019.
24. "Neighbourhood Tabulation Areas | NYC Open Data." https://data.cityofnewyork.us/City-Government/Neighborhood-Tabulation-Areas/cpf4-rkhq. Accessed 2 Dec. 2019.
25. "Population – American Community Survey (ACS) Data Tables." https://www1.nyc.gov/site/planning/planning-level/nyc-population/american-community-survey.page. Accessed 2 Dec. 2019.
26. "Food Availability (Per Capita) Data System." https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/. Accessed 2 Dec. 2019.
27. "Retail Food Stores – Open Data NY – NY.gov." https://data.ny.gov/Economic-Development/Retail-Food-Stores/9a8c-vfzj. Accessed 2 Dec. 2019.
28. "Instacart Online Grocery Shopping Dataset 2017." https://www.instacart.com/datasets/grocery-shopping-2017. Accessed 2 Dec. 2019.