**Executive Summary**

*This section provides an overview to the project. It should briefly touch on the motivation, data question, data to be used, along with any known assumptions and challenges.*

**Motivation**

*Here you will go into more detail about why you have chosen this project.*

When I think about it, my selected topic has been something I’ve always been interested in based on the many conversations my sister and I would have about it over the years. As a native New Yorker, I’ve been fortunate to experience the world-renowned culinary scene the city has to offer. No matter what culture it is, if you want to experience it, you can find it in New York. Although New York has plenty of options for food when it comes to eating out, can we say the same when it comes to buying your own food to cook? Do we have enough grocery stores around the city to meet the demand of a population of almost 9 million people? Out of the all the grocery stores located in the city, how many of them can we say sell healthy/quality food? Is the food even affordable for everyone?

This has been a topic of concern the city had for a while, but the COVID-19 pandemic really exposed how big of an issue it was. When the pandemic first hit, food was already scarce due to panic buying, inventory problems, and restaurants closing (both temporarily and permanently) due to restrictions or inability to pay rent. Now that social distancing dramatically limited the amount of physical interaction people had with one another, they were pushed into doing things digitally via computer and/or smart phone, which included buying food.

During this time my grandmother, only knowing how to use a smartphone for certain things, would call her grandchildren to order food for her through apps such as Uber Eats, Door Dash, etc. As grandchildren, we were concerned about her options when it came to choices of food based on the neighborhood she was in; and although she did live on a block where the grocery store was down the street, it was expensive for the demographic it served and didn't always have the best products. For my grandmother to buy higher quality food, she would have to travel a way (sometimes across the borough) to shop at the grocery stores that had these products. Before the pandemic it wasn't a problem to getting to those distant stores. Public transportation was reliable and there were friends and family that were able to pick her up and take her whenever needed. When the pandemic hit, there were restrictions and changes in operations for both public transportation and grocery stores. Also, with people in isolation and on different schedules than before, it was harder to find a time when someone would be able to take her to the market.

Reflecting on that, it had me questioning what other areas of the city do people have trouble getting access to healthier food. I was fortunate that in the part of city I lived in, there were at least three supermarkets that were in relative walking distance and a farmer's market at the nearby park every Saturday. Many people in the city would be ecstatic to have one supermarket that was in walking distance. It's also important remember that although technology has become very common there are still families out there that don't have access to internet or smartphone, so the apps that could have potentially given these families access to better food stores during a time of social distancing such as Instacart wouldn't have been helpful (And we haven't even factored the cost of food yet into the equation).

With the analytic skills I've acquired and the public access to New York City data, I would like to know what areas of New York City have the best and/or most options when it comes to accessing healthy foods.

**Data Question**

*Present your question. Feel free to include any research/articles that are relevant or show where others have attempted to answer this question.*

**Primary Question**: What areas of New York City have the best and/or most options when it comes to buying healthy and fresh foods?

* Are there specific boroughs and/or community districts that have significantly better options than others?
* Are the locations of the better food options (farmers/market, recognized shop healthy stores (Bodegas & Grocery Stores Receiving Recognition from Borough President's Office), etc.) benefiting certain demographics more than others, especially now with the impact COVID-19 had on the city?
  + Where does the location of SNAP Centers in NYC compare to the better food options in terms of community districts?
    - *Note*: SNAP (Supplemental Nutrition Assistance Program) is a federal program that provides food-purchasing assistance for low- and no-income people.
* For community districts that don't have access to the healthy food stores, what type of restaurants are located in the area?
  + What’s the average health inspection grade for restaurants in those districts?

Potential Bonus Question(s) (Depending on how early I finish answering the primary question for this capstone)

How does the quality of food in the schools of the community districts compare to the restaurants?

* What are the typical health inspection grades for school food in the community districts?
* Do children get better food in school depending on what community district they live in?

Note: Depending on how early I finish answering the primary question and the results of the data analysis, I may possibly add some sort of case scenario to the capstone.

**Minimum Viable Product (MVP)**

*Define your MVP. This should be a description of what your final capstone will look like, including visualizations, how the analysis will be presented, who the intended audience is, etc.*

Here’s a tentative order of the overall capstone presentation. It will be presented via Tableau story (Order and topics subject to change based off analysis of the data):

1. **Intro/Background**
   1. Intro/Background to the capstone will be presented via PowerPoint slides that will be imported into Tableau.
2. **Recognized shop healthy stores**
   1. Community District analyzation
   2. SNAP locations in relation to the healthy stores
   3. Borough analyzation
3. **Farmers Market**
   1. Community District analyzation
   2. Borough analyzation
4. **Restaurants**
   1. Community District analyzation
   2. Borough analyzation
   3. Health Inspection Grades
   4. What type of restaurants are located in the community districts that have the least access to recognized shop healthy stores and farmers markets? Are they healthier restaurants or just fast food?
5. **Conclusion/Thank you/Questions**
   1. Conclusion/Thank you/Questions slide(s) will be presented via PowerPoint that will be imported into Tableau.

*Note: Demographics will be discussed throughout the presentation and how it relates to all topics.*

Visualizations specifics

* Bar, line, scatter, and table charts will be used to compare data between community districts in regards to specific demographics, farmers/market, recognized shop healthy stores, and restaurants (Other visualizations besides the ones listed could be used to compare the data if needed).
* There will be a dashboard in Tableau story focusing on the Top & Bottom 5 community districts in regard to access to healthy food. In this dashboard we will be able to see what the similarities and differences are in the communities that made either list. This will most likely be towards the conclusion of the presentation.
* Bubble & Tree maps will be used compare data between boroughs in regards to specific demographics, farmers/market, recognized shop healthy stores, and restaurants (Other visualizations besides the ones listed could be used to compare the data if needed).
* The card visualization will be used to describe general demographic information about the community districts (ex. Brooklyn Community District 1 Population = 173,083)
* Pie/Donut charts will be used to analyze the percentage of Farmers Market throughout the city that accept EBT as well as the percent that are open year-round.
* Python will be used to create geospatial map(s) highlighting the location of farmers markets, recognized shop healthy stores, SNAP offices, and restaurants in New York City by community district. The geospatial map(s) produced in Python will be presented through the Tableau story board.

Intended audience

The intended audience for this project is those that are into health, nutrition, community/city planning, and finding solutions to dipartites. Residents and enthusiast of New York City and State would also likely find this capstone interesting.

**Schedule**

1. Get the Data (5/7/2022)
2. Clean & Explore the Data (5/22/2022)
3. Create Presentation of your Analysis (6/17/2022)

* Should be a presentation, but could include a Jupyter Notebook or dashboard in Excel, Tableau, or PowerBI

1. Internal demos (6/21/2021)
2. Demo Day!! (6/28/2022)

**Data Sources**

*Document the data you use and the source of that data*

**List of Data Sources**:

From the State of New York

CSV:

<https://data.ny.gov/Economic-Development/Retail-Food-Stores/9a8c-vfzj>

From the City of New York

CSV:

<https://communityprofiles.planning.nyc.gov/>

<https://opendata.cityofnewyork.us/>

<https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets/8vwk-6iz2>

<https://data.cityofnewyork.us/Social-Services/Directory-of-SNAP-Centers/tc6u-8rnp>

<https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j>

<https://data.cityofnewyork.us/Social-Services/Borough-Community-District-Report-SNAP-Population/jye8-w4d7>

<https://data.cityofnewyork.us/Health/DOHMH-School-Cafeteria-inspections/9hxz-c2kj>

<https://data.cityofnewyork.us/Health/Recognized-Shop-Healthy-Stores/ud4g-9x9z>

From the City of New York (GeoJSON)

<https://data.cityofnewyork.us/City-Government/Community-Districts/yfnk-k7r4>

<https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm> (possibly)

<https://data.cityofnewyork.us/Business/FRESH-Food-Stores-Zoning-Boundaries/w9uz-8epq> (possibly)

**Known Issues and Challenges**

*Explain any anticipated challenges with your project, and your plan for managing them. Be sure to include:*

* *If you need to request data or an api key*
* *Based on your data sources, known data cleaning steps*

1. When it comes to the community districts data. I will need to clean/merge the data of 59 individual datasets (CSV) that make up all the community districts of New York City.
   1. Solution: Create a CSV that has all the data I need for every community district (There are multiple columns in this dataset that I will not need for the purposes of this capstone project).
   2. Alternative Solution: Use python to merge all 59 datasets and from there drop the unnecessary columns. This process could get complicated and become very time consuming.
2. Visualizations not coming out the way I want in Tableau **(I don’t anticipate this issue, but it could happen)**.
   1. Solution: Use Power BI to create the visualizations
   2. Possible Alternate Solution: It may possible to import visualizations from Power BI into Tableau. I will need to look into this a little more.
      1. Importing visualizations from Excel is also an option as well.
3. Importing PowerPoint Slides into Tableau
   1. Solution: There are multiple videos on YouTube that walkthrough the process of importing PowerPoint Slides into Tableau
   2. Alternative Solution: Ask for help/advice from colleagues that are more experienced or experts in Tableau.