For the final project, I learned to apply many of the skills that I was exposed to for the first time in this class. It was an intimidating, exciting, and humbling experience to see not only how far I have come but also how much more opportunity there is for growth. During this project, I took on the role of extracting Whole Foods grocery store locations and coordinated by using the Google API, Pandas and Jupyter Notebook. My initial attempt was to use the knowledge I obtained in the web scraping module. Early on in this process I ran into a bit of a challenge with the data available on the Whole Foods website as this did not have the coordinated and only had the addresses. For this reason, I decided to use the Google API instead which retrieved the address and coordinates. In addition to the extraction of whole foods locations, I put together and deployed the website for our project. This piece was a collaborative effort. As a team, we started the base template for the website and then broke up into tasks who would work each section of the website. Once the team provided their sections of code, I put it all together and deployed the site on my GitHub.

Yahel and Mahire took the lead in cleaning the Food Access and Zillow data which would be used not only for data visualization but for the machine learning models as well. Through collaboration with the team, we reviewed the datasets to determine which data would be valuable for our analysis. Yahel used this data to put together the data visualization piece in Tableau and Mahire worked on creating some amazing machine learning models using the 2015 and 2019 data obtained. Ryan also took on the role of extracting fast food locations and coordinates. Liya took the role as our project team lead and data management. Liya put together the final PowerPoint presentation and was a key contributor in narrowing down our project research questions. We all worked well as a team and I feel the biggest key to our success was that we all communicated well and carved out a significant amount of time outside of class and office hours to collaborate, check in on where we were at and helped each other out when we were stuck or had questions.

There were some challenges early in the project due to figuring out who would do which tasks and avoiding the overlapping of tasks. For this reason, we ran into some merging errors in GitHub due to multiple people working on the same file and not pushing or pulling when needed. These were things we learned through trial and error, overcame, and ironed out throughout the process. Another challenge we ran into was outdated data and the lack of data. Due to the scope of the project, and data available, and time allotted we had to shift or focus and reevaluate which direction to take for the analysis of this project.

The focus of this project was to analyze housing prices and food access to see if there was a connection between the two. Although we did not find a direct correlation between the two, we did find that the median housing value with more organic grocery stores is double the median housing values next to locations with more fast-food locations. We used supervised machine learning for our project and ran Simple Regression, Polynomial Regression, Decision Tree, Random Forest, and Neural Networking models. In comparing the R2 of these models, it was determined the Random Forest model had the highest score.