**BarginGoods**

*An app that finds the store with the best price for your grocery list*

Fidelia Nawar, Soo Sung, Mariah Meehan, Sichen Zhong, Pony Ameri

#### **WHY**

#### **Problem Statement**

* Describe the problem you are solving.
  + Hundreds of thousands of New York City residents live in so-called food deserts, or low-income areas without large supermarkets and areas that lack affordable food and grocery options. For many low income families, being able to find the cheapest grocery items at stores is necessary for them to remain within their budget and they struggle to do so if they live in a food desert.
    - https://pix11.com/news/created-equal/why-food-deserts-persist-in-low-income-nyc-neighborhoods/
* Why is this a compelling and impactful problem to solve? Why is this a big opportunity?
  + Currently, there exists no apps on the market that provide the functionality for users to find the most optimized, budget-friendly recommendations for which grocery store to purchase their groceries from.
  + This is a big opportunity because for many individuals in urban areas like San Francisco/NYC there exists many different grocery stores at various distances, so it’s more difficult to identify which store has the best budget options based on the specific items a person needs for grocery shopping.
  + We also hope to implement the functionality of recommending cheaper food alternatives based on similar nutritional value (which currently does not exist in the market either)
    - For example, recommending chicken ($3/lb) if it’s cheaper than turkey ($4/lb) if it’s the same nutritional value
* What assumptions are you making about the problem / opportunity? (these assumptions would directly influence the key elements and features you would build in the minimal viable product (MVP)
  + Assuming that different grocery stores have drastically different prices that affect buyers’ decisions
    - This statement holds true, such as buying Great Value brand items from Walmart vs organic brands from Whole Foods
  + Assuming that most low-income individuals are seeking the cheapest/budget-friendly grocery options with minimal traveling necessary (to reduce transportation costs)

**Impact and market opportunity**

* How would you go about quantifying impact, market opportunity for this particular problem
  + This would impact large groups of low-income individuals in urban cities who are seeking to reduce their total grocery costs each month
  + This app would simplify the process of finding the cheapest options for grocery shoppers and providing them cheaper food alternatives based on nutrition value
  + Cost-optimization features could save people up to $50-100 per month on groceries hopefully (just an average estimate)
  + Being recommended the cheapest grocery store option will also save time for users so that they don’t have to visit multiple grocery stores and spend more money on groceries and transportation
* MarketSize (How big is this market, based on your research?)
  + According to NYC Opportunity, the citywide poverty rate in New York City was about 17.9% of the total population (about 3 million residents)
    - <https://www1.nyc.gov/site/opportunity/poverty-in-nyc/poverty-measure.page#:~:text=The%20New%20York%20City%20Government%20Poverty%20Measure%2C%202019&text=The%20citywide%20poverty%20rate%20fell,points%20that%20is%20statistically%20significant>.
  + Hundreds of thousands of the low-income residents mentioned above live in food deserts, further preventing them from being able to have easy access to the most affordable groceries near them
    - <https://pix11.com/news/created-equal/why-food-deserts-persist-in-low-income-nyc-neighborhoods/>
  + During times of crisis or disaster like COVID, many individuals became panic-stricken and cleaned out the aisles (some wealthy people even hired limo services to deliver groceries to their residences), but low-income people do not have the luxury or access to do so and would need an efficient way to find the most optimal grocery store for affordable groceries
    - <https://www.businessinsider.com/nyc-grocery-stores-struggling-wealth-flight-jobs-cut-2020-4>

**Target Customer and user/customer discovery**

Who is the primary customer/user? What is the use case? What are key assumptions you are making about the primary customer/ user and use case?

* Identify targeted user/customer segments for the MVP.
  1. Low income grocery shoppers with smartphones or computer access
* Define the primary use case validated by this target user/customer.
  1. A shopper enters a list of groceries they need to buy in the app. They also allow the app to use their location. When they hit enter/find, the app returns a list of grocery stores and alternative recommendations for items that are optimized for their location and the total cost of their grocery list.
* What might be other key assumptions that are important to validate with the target user/customer?
  1. Additional user assumptions
     + Users prefer to shop locally since travel incurs additional cost.
     + Users are looking for ways to grocery shop on budget while keeping nutritional values of their groceries intact.
  2. Additional store price assumptions
     + The grocery store prices we have access to include stores that the user wants to shop at.
     + The grocery prices are up to date.
     + Smaller grocers who might not have a website with their price info do not have lower prices compared to grocery stores we have data for.
* How would you go about validating these additional key assumptions?
  1. For the user assumptions we can do some research on local NYC shoppers to understand how and where they usually shop based on income and neighborhood. This can be done with USDA data sets on types of grocery items bought from where. We will also research studies on specifically in the NYC areas on grocery purchases.
     + <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4620064/> - study showed that these shoppers valued location and shopped almost daily at nearby bodegas for that convenience.
  2. For the store price assumptions we may have to get access to some locals to understand the types of smaller grocers that are available and how their prices compare to the large grocery chains. The following study shows comparisons between local grocers vs supermarkets in the NYC areas:
     + <https://www.nycfoodpolicy.org/comparing-nyc-supermarkets/>
* Who will you contact to conduct initial user research and feedback?
  1. We are planning to reach out to various non-profits in our respective areas to learn more about low income families. There are services that provided fresh produce for free to locals during the pandemic. We want to contact those types of organizations to see what they understand about low income families and access to food. (e.g. <https://www.nycservice.org/>)
* What is the user journey and UI/UX for this data product?
  1. New user
     + User opens app and allows location access
     + They’re prompted to create an account if they want to. This allows them to save their previous grocery list history and track how much they’ve saved by using this app.
     + They enter grocery items like bananas, celery etc. using keyboard or choose from common items from drop down
     + They hit the find button and the app returns a list of grocery stores that have all their grocery items sorted by total cost and distance from their current location.
  2. Experienced user
     + User opens the app, logs in (if account created), and pulls up previous grocery list(s)
     + They make any necessary modifications to desired items
     + They hit the find button and get a list of stores that optimize on total cost and distance.

**Market Landscape / Competitive Landscape / Existing companies solving the same / similar problem**

Who are the major players and main vendors in the space? What are the existing solutions?

| **Company Name** | **Stage (startup, enterprise)** | **Product / Solution overview** | **Who is the primary customer?** | **Key differentiation vs your proposal (based on your understanding/**  **research)** |
| --- | --- | --- | --- | --- |
| **Basket Saving Inc.** | Enterprise | Basket - a grocery shopping app having the function to find the nearby cheapest grocery store for items in the shopping bag | People who want to save on grocery shopping | Our app will have alternative suggestions in same nutrition but cheaper price |
| **mygrocerydeals.com/** | Enterprise | <https://www.mygrocerydeals.com/> - Grocery shopping webpage finding deals | People who want to shop for deals | We include live price information instead of just deals, also our app will have alternative suggestions in same nutrition but cheaper price |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

If you can’t identify existing solutions or similar solutions that solve the problem, please explain why there isn’t an existing solution.

**Relevant readings, market research, white papers, academic research (share title and link)**

* The app shows you where to find the cheapest groceries near you. <https://www.eatthis.com/app-shows-cheapest-groceries-near-you/>
* NBC News is tracking changes in grocery prices to monitor the impact on consumers’ food bills.

<https://www.nbcnews.com/business/business-news/graphic-track-grocery-price-trends-n1264037>

* USDA economic research on food price outlook

<https://www.ers.usda.gov/data-products/food-price-outlook/summary-findings/>

* American households struggle as inflation continues - new survey

<https://www.weforum.org/agenda/2022/09/inflation-causing-hardship-us-households-money-income>

* SNAP Benefit Adjustments Will Help Low-Income Households Cope With Food Inflation

<https://www.cbpp.org/blog/snap-benefit-adjustments-will-help-low-income-households-cope-with-food-inflation>

**WHAT**

**Minimal Viable Product (MVP)**

* What is the minimal viable product that you are building that will specifically test the fundamental assumptions you have about the problem and the value of your solution?
  1. What are the main features and why?
     + The main features of this product are
       - User Account: Where they can store their information, location and previous grocery lists
       - Location Services: this would be opt-in and would be used to optimize for distance. The user would also have the option to type in a location instead.
       - Grocery List: Where the user can create a grocery list for the current grocery run.
       - Comparison Page: The page which shows the total price for each store with the cheapest options selected. The top recommendation will be the one that minimizes cost and distance.
       - Store Page: When a user clicks on a grocery store from the Comparison page, they will see their list and which brand they will need to buy for each item on their list in order to get the displayed price. This page will also have recommendations for cheaper options that have the same nutritional value (ex. swapping salmon for beef).
  2. What is the value delivered to your user/customer?
     + The value is providing an easy way to get exactly what the user needs from the grocery store and know that they are getting the cheapest price (for their entire list). It also provides a way to compare discounts or promotions going on at stores to see if going to a different store in a given week is cheaper. Lastly, it provides recommendations for cheaper alternatives on their list within a given store to aid them in minimizing cost.
  3. What is the key question or questions (max 3) that your target user/customer will be able to answer using your Capstone product?
     + At which grocery store, in my area, will I pay the least amount for my grocery list?
     + Once I’m at the store, are there any items that I could swap out to get a cheaper price but still get the same nutritional value?
* What data science approach would you intend to use for the MVP? (this is NOT UI / UX but technical discussion)
  1. The data science approach we will take is to do optimization for the lowest total cost for the grocery list and minimum distance from the user's location.
  2. We will use NLP and K-fold classification to classify items into categories which will be used for our cheaper alternative recommendations
  3. The data science techniques are outlined in greater detail in the *Technical Approach* section.
* How would you potentially test the efficacy of the MVP? When would you start testing?
  1. We will start with a group of 10 people who are in New York. We would like the people to be representative of the population we are targeting, however if we are unable to gain access to these people in time, we will use our personal networks for the test.
  2. Then we will have them test the following features and fill out a Google Forms survey afterwards. The features we will want to focus on are:
     + Creating an account: we will want to see if they are willing to enable location services
     + Creating a grocery list: we want to make sure our database has enough options and covers all the items a user might think of or need
     + Picking a store and finding the items: This will be the real test. Once the user picks a store they need to actually go to that store and find the brand of the items. We want to see whether they took our recommendations for the cheaper alternative, that will be a question on our survey. We will also want to make sure our data is up to date so we aren’t recommending brands or products that are out of stock.
  3. We would like to start testing as soon as we have a prototype available aiming for the end of October.

#### What is the key **differentiation** between your MVP and the existing solutions and/or approaches?

The key differentiators for our app is the recommendation system for cheaper alternatives with the same nutritional value. Our closest competitor is the Baskets app which allows the user to create a shopping list and see prices at local grocery stores but it does not recommend cheaper alternatives. Other grocery shopping apps only allow for searching and comparing a single item or shopping at just one store.

#### **Value Proposition** (what value/utility does your project/product provide to your intended users?)

* State the value that your MVP brings to the target customer segment.
  + Knowing where to grocery shop to find the best deal can be time consuming and difficult. We assume we know the cheapest store to go just based on past experiences but prices change everyday. What if we could be spend less and still get everything we need without adding more time that we don’t have? BarginGoods provides an easy way to get exactly what you need from the grocery store and know that you are getting the cheapest price. It even tries to get you the same nutrition for less by providing recommendations for cheaper alternatives. Our goal is to make your life easier and help you spend less.

#### **Mission Statement**

#### **Helping you spend less and save more at the grocery store by making it easy to see which store has the cheapest cost for everything you need and providing recommendations for cheaper alternatives that have the same nutritional value.**

#### **HOW**

#### **Data sets**

* What datasets do you intend to use?
  + We will try to modify the data scraper from the github repo listed below for grocery store price (focused on NYC store)
    - <https://github.com/vcheeney/Pricebook>
    - Trader Joes, H mart, fairway market, sunrise mart, buonitalia, westerly natural market, etc
  + Wegmans api for product detail. (Might need to explore using Microsoft Azure)
  + Kroger api for product detail
    - <https://developer.kroger.com/reference/#operation/authorizationCode>
  + Google places api for user location
* Are the datasets public?
  + Scraping is not publicly available but we will not use the data for profit at this stage.
  + Apis may include charges.
* What are the datasets attributes / metadata that could make the exploratory data analysis easier / harder?
  + Text information and numeric values would be preferred. Image data can be used later for classification from the product.

#### **Project Management**

* What is the role of each member (who will do what specifically)?
* Who is the project manager and chief facilitator (and tie breaker)
  + Mariah
* Who is the resident SME?
  + Fidelia Nawar
* Who is the product manager?
  + Fidelia Nawar
* Who is the lead on infrastructure and data engineering?
  + Sichen Zhong
* Who is the lead on EDA?
  + Sichen Zhong
* Who is the lead on model evaluation?
  + Pony Ameri
* Who is the lead machine learning engineer?
  + Pony Ameri
* Who is the lead MVP application developer?
  + Soo Sung
* Who are the backups to key roles?

|  | Role and responsibilities (immediate) | Role and responsibilities (long term) / Alternate or additional role / pair |
| --- | --- | --- |
| Fidelia Nawar | Product Manager/Resident SME | Help identify key features/functionality of product, update project based on scope/time/resource constraints / ML Engineering |
| Pony Ameri | Model Evaluation  Machine Learning Engineer | Help with Data Engineering |
| Mariah Meehan | Project Manager/Chief Facilitator | Data Engineering/EDA |
| Soo Sung | MVP Application dev | Can help with machine learning and project management |
| Sichen Zhong | Data engineering and EDA | Can help with model validation and machine learning  Can help organize meeting and record the todo list |

\*\*Some teams have used pairing principles to assign two members of the team to main tasks.

* What are the strengths and weaknesses of each team member?

|  | Strengths | Weaknesses |
| --- | --- | --- |
| Fidelia Nawar | Coding (Python/C++/JavaScript/SQL), Project management, Front/backend development, creating presentations, UI design | Cloud infrastructure setup, ML pipeline development |
| Pony Ameri | Data Engineering, Machine Learning, Data Visualization | Devops |
| Mariah Meehan | Data analytics, Data Cleaning/manipulation, data visualization, Python | App development, cloud infrastructure setup |
| Soo Sung | Documentation, front end web dev (JS, HTML, CSS), EDA, data visualization, Cloud, Tensorflow | Data engineering, backend development, ML pipeline development |
| Sichen Zhong | Time management/Data analytics/Data engineering/Data visualization | App development/Ethical and privacy related topics |

* Submit the Team Process Agreement.

#### **Technical Approach and Planning**

* What methodologies would you use for initial data exploratory analysis to ensure your datasets are sufficient and meaningful?

1. Structure Investigation
   1. Structure of non-numerical features
      1. Check for unique values and frequency of most prominent values
   2. Structure of numerical features
      1. Check for statistical insights to get details on various statistical data like Mean, Standard Deviation, Median, Max Value, Min Value
2. Quality Investigation
   1. Checking null entries
   2. Check for missing values
   3. Check for unwanted entries and recording errors
   4. Checking for duplicates
3. Content Investigation
   1. Feature distribution
      1. Check for normal distribution or skewed distribution
      2. Outlier detection
   2. Feature patterns
      1. Multivariate analysis to check if there are existing relationships between more than two variables
   3. Feature relationships
      1. Correlation plots to check for relationships between features and randomness in data

* What data science algorithms are you intending to develop and build for the project? What challenges do you potentially foresee?
  + Content based filtering
    - Use item features to recommend other items similar to what the user likes, based on their grocery list
    - Algorithms:
      * Bayesian classifiers
      * Decision tree models
        + Every level of a decision tree can be used to filter out the various preferences of the user to make a more refined choice.
      * Deep neural networks
        + Captures the specific interests of a user and improves the relevance of recommendations.
    - Disadvantages
      * The model has limited ability to expand on the users' existing interests.
      * The model can only be as good as the hand-engineered features.
* What help do you need?
  + How to build a recommendation system?