Group Assignment #3: Project Plan and Evaluation Metrics

Grocery Assistance: Mariam Mayanja, Geovani Castro, Linxing Preston Jiang

Part I: Evaluation Metrics

We start designing the evaluation metrics by asking questions like: Will our application increase the efficiency at which a user is able to track down their desired essential food product? Can we remove the need to visit multiple stores in a row in order to track it down, as well as increase efficiency for future users?

In general, the following metrics offer reliability by setting a concrete scale for qualitative measures and by focusing on the change in the amount of time taken to perform a task between actions with and without the app. This approach creates reliability by helping standardize the data we are looking at.

As for mitigating over alignment, we are making a conscious effort to avoid over-alignment to the demographics of our user base. Try to avoid questions that involve demographics information or make assumptions about the resources a user might have. As such, we take into account things like the mode of transportation and the number of grocery stores within a reasonable distance of the user. For example, in accounting for the number of grocery stores a user has access to our data does not accidentally label a user as having more success with the app if they only visit 2 stores because there are only 2 in their area. Realistically, visiting all of the grocery stores in your area before you are able to find food or deem that there is not any is a failure, not a success. However, for someone with easy access to 5 or more stores, this might look like a success.

Consistency is achieved through rigorously sticking to the same predetermined scales and descriptions for qualitative assessment. For quantitative metrics, this takes form in the way we consistently measure the data, which is also determined in the scale. With a set scale for each measure, we are able to more consistently evaluate a user’s experience.

While the definition of success, the scale of the metric, and face validity analysis are all unique to each metric, having them listed and clear from the beginning helps us gather better data that better proves or disproves our previous hypothesis.

Goal 1:

Decrease the number of stores a user has to go before finding the available product or deciding to no longer purchase it

Quantitative of qualitative:

All of these measures are quantitative.

Scale for the metric:

For measuring transportation we are given an estimate of the amount of time it takes to arrive at the store. For measuring stores within a reasonable distance, we ask the user for a subjective view of “reasonable” as everyone has different transportation means. We ask for the average amount of stores a user visits per time they go grocery shopping.

Face validity:

This offers face validity by giving us a way to measure if we decrease the number of stores a user needs to go to.

Definition of success:

For this to be successful we want to see the number of stores traveled to per grocery outing decrease proportional to the number of stores a person has access to. We would like to see this decrease by at least 20%.

Metric (Number of stores traveled to)	Without App	With App
Mode of transportation and the amount of time to arrive at the grocery store?	Blanca Alvarez said that she uses a personal car and that each grocery store is between 5-10 minutes away from her home.	
Amount of stores within a reasonable distance?	4 stores	
Amount of stores traveled to per grocery outing?	3 stores	
Notes		

Goal 2:

Reduce the amount of time the user spends to find the essential food products

Quantitative of qualitative:

These metrics are measured qualitatively.

Scale for the metric:

This metric is measured qualitatively and is dependent on the users subjective observations of their own shopping experience. This takes shape in the form of notes, anecdotes, and estimation of how much time and strain they spend finding essential food products. Ideally measurements will be in hours if the user is able to approximate the time they spend.

Face validity:

This metric helps us determine if the app effectively decreases the amount of time a user spends looking for essential food products.

Definition of success:

We can consider success as the user cutting their time spent in subsequent stores in half. On average, a grocery store trip takes a person about 41 minutes according to Spendmont. We are not focused on lowering this metric, but rather lowering the number of subsequent stores they visit and the amount of time spent in those subsequent stores. This will be successful if the amount of time spent finding an alternative product is decreased by 30-50%.

Limitations:

One limitation of measuring this metric is that people are not always able to give an estimate on the amount of time it took to perform a task. To avoid discrepancies in reporting, we will attempt to phrase the question consistently and record notes on the user's qualitative assessments.

Metric (amount of time spent finding products)	Without App	With App
Mode of transportation and amount of time to arrive at the grocery store?	Personal car	
Amount of time to determine the product is not in stock (per item)?	About 1-3 minutes after arriving at the store (per product). She says it only takes the amount of time it takes to arrive at the isle plus looking around in case isles are messy.	
Average amount of time spent in each grocery store?	About 40 minutes to an hour	
Amount of time spent finding an alternative if the product is not in stock?	If the initial store does not have the items, she said she spends an additional **hour to two** traveling to other grocery stores and looking there.	

Notes	This metric was a little more difficult to measure as people very rarely spend time quantifying how much demand looking for individual products put on them.	

Goal 3:

Decrease how frequently a user goes to the store to find the desired or "essential" item is out of stock

Quantitative of qualitative:

These measures are qualitative because we cannot concretely measure them.

Scale for the metric:

For measuring things being out of stock we have a scale from 1-5.

1 - never out of stock

2 - sometimes out of stock

3 - Out of stock for about half of your recent grocery trips

4 - Out of stock of most of your recent grocery trips

5 - Always out of stock

Face validity:

This metric offers face validity by providing us with an understanding of how how frequently items that users expect to see are out of stock. This allows us to determine how effective our solution is by measuring the person's ability to determine what is available.

Definition of success:

For this to be successful, we would like to see the ratings fall to a 1 (never out of stock when you arrive at the grocery store).

Special limitations:

Due to the constraints on gathering enough crowdsourced data in the time we have remaining in the quarter, we will need to simulate data in the app to gain an approximation of what these metrics could look like with assistance from the app. This, however, might also be impacted by the slowly improving situation. It is possible that these questions could measure differently not because of the app, but because the situation is simply improving. As such, it would be ideal to simulate data that mirrors the panic buying and stock quantities in May 2020.

Metric (Reliability in determining item availability)	Without App	With App
What do you define as essential products?	Milk, baby formula, nursery water	Milk, baby formula, nursery water
How frequently do you find these products out of stock when you expected to see them when you arrive at the grocery store?	Alvarez said that “when this all started,” which she defines as March, each product ranked a 4. Now, in May she would rank each product’s availability at a 2.	
Notes		

Goal 4:

The ability for user to upload crowdsourced data

Quantitative of qualitative:

This metric is measured both quantitatively and qualitatively. The amount of time to navigate to uploads and the amount of time spent filling out the upload form is quantitative and to be measured in seconds. The frequency a user assists members of their community is qualitative. All metrics are to be measured as followed.

Scale for the metric:

The frequency will give an estimate of how many times in a week the user assists members of their community by sharing information on what items they saw in stock. This will be contrasted with how many times a week the user goes grocery shopping.

Both other metrics, ease of navigation and ease of upload, will be measured in seconds. This will be uniformly measured in the amount of time spent from the landing page to the upload page and from the landing page to submit a new data point.

Face validity:

This metric provides face validity by offering us a way to measure the effectiveness of collecting data to help inform users’ shopping decisions. In measuring the change in their perceived assistance of community members and their ability to share information we can effectively design a form that allows better data sharing.

Definition of success:

For this metric to be successful, we need to see users contributing information to the platform. We phrase this as "assisting members of your community" to incentivize people to post information in a way that highlights the good it does. Ideal data contribution is considered to be a minimum of once per grocery visit, this lets you contribute each time you go to the store without asking users to put themselves at any additional risk.

Metric (Contribution to stock quantity)	Without App	With App
How many times a week do you go grocery shopping?	1 shopping trip/week	
How frequently do you assist members of your community find what items are or are not in stock?	Approximately 1 time/week, She said she reaches out to her friends, especially those with infants or young children, whenever she goes to Walmart. However, she says this is only behavior she does during the pandemic.	
How easy was it to navigate to the upload tab?	This is not applicable. Ideally, we would like to see it take 10 seconds or fewer for the user to identify and navigate to the upload portion of the app.	
How easy was it to propagate the data upload form?	This is not applicable Ideally, we would like to see it take 20 seconds or less for the user to identify if an item was available.	
Notes		

Part II: Project Plan

We start the discussion on the project plan by elaborating on the goals we plan to reach. We then describe the internal architecture of our web application, including a flowchart of activities and the structure of the databases. We conclude by discussing the limitations of our project faces, and a storyboard picturing a few likely user scenarios at the end.

Project Specification

<u>Part 1: Project Goals</u>

- Initialization: enter location [1 day]
    - The user can choose to enable location services or manually enter their zip code
- Displays and searching [1.5 week]
    - The user is able to navigate through a three-level (Category → Subcategory → Product) structure to browse all WIC-eligible food products. The UI has both text labels and image labels.
    - The user can search the database using text for a list of relevant products or use UPC/PLU code for a more accurate search.
    - When the user clicks on a specific product, the application displays a list of ranked stores we recommend, shown with the estimated quantity of the product at each store and the time when this entry was last updated.
    - Google Maps will be opened for navigation when the user clicks on a specific store.
- User feedback & Crowdsourcing [1 week]
    - The user can provide feedback to the application by clicking on the feedback button
        - If the user enabled location service, we will rank the stores the user is most likely to be at based on the live location.
        - If not, the user can choose which store whose products s/he wants to provide feedback for.

<u>Part 2: Internal Architecture</u>

Please check the page at the end of the file (the landscape page).

<u>Part 3: Database Architecture</u>

The application relies on two databases for its functionalities. One is the WIC food product database, which contains relevant information about all WIC-eligible food products such as categories, subcategories, size, unit, etc. A snippet of the database is shown below at Table 1. The first column is highlighted as it is used as the index key.

UPC/PLU	Category	Subcategory	Description	Size	Unit
72220110616	Bread/Whole Grains	Bread - 100% Whole Wheat	Abiqua Farms 100% Whole Wheat Bread 24 oz	24	Oz
52200034767	Infant	Infant Cereal	Beech Nut Sensitive Oatmeal Cereal 8oz	8	Oz

Table 1: A snippet of the WIC Food Product Database

The other database is the food quantity database, which contains the UPC/PLU code, the store, and the estimated quantity, and the time when this information is last updated. A snippet of the database is shown below at Table 2. The first column is highlighted as it is used as the index key.

UPC/PLU	Store	Quantity	Last Updated
72220110616	100 NW 85th St. Seattle, WA	30	2 hours ago
52200034767	2500 SW Barton Ste B-1 Seattle, WA	10	1 day ago

Table 2: A snippet of the Food Quantity Database

Project Plan

Overall, our project aims to make shopping for WIC-eligible food products easier during public emergencies by providing product availability estimates using crowdsourced data. Our web application recommends/ranks WIC vendors to the users based on their location and the estimated amount of the product of interest in store, calculated from user feedback.

Our progress so far has focused on backend development through a few pair programming sessions done by Geovani and Preston. We have implemented a subset of the essential logic needed for various features. First, we decided to use Pandas and Python as the supporting data analysis backend. Using the Excel file containing all WIC-eligible food products acquired from our communication with the Washington State Department of Health, we loaded the information as a Pandas DataFrame object that supports fast and flexible queries. Second, we cleaned and preprocessed some data to make the category information more readable and easy to search

(For example, we grouped all “Infant” related product categories such as “Infant - Meats” and “Infant - Vegetable” to the large “Infant” category, with “Meats” and “Vegetable” being the subcategories, see Table 1). Third, based on the need to track the remaining number of products at each store, we created another database/dataframe object that contains live-updated product quantities at specific stores. We decided to use UPC/PLU code as the index key for our database for fast queries since they are unique, and have two other columns as store locations and estimated quantities (see Table 2).

The remaining backend logic to be implemented includes:

1. Scrape the information of all WIC vendors listed on King County website [Preston]
2. Search box: take the search string or UPC/PLU code entered by the user and return a list of relevant products [Preston]
3. Rank the stores based on the number of remaining products and the distance of the store from the user [Geovani]
4. Display map snippets based on the recommended store, and open Google Maps to the snippet when the user clicks a specific store recommended [Geovani]
5. Take users feedback and update the corresponding quantity in the database [Preston & Geovani]

The last piece of the development is to create a web frontend for our application. Mariam is responsible for bootstrapping an existing website and making sure it’s accessible using the following online assessment tools:

- https://webaim.org/services/evaluation/
- https://www.washington.edu/doit/web-accessibility-guidelines-administrators

Geovani is responsible for using frontend frameworks such as Jinja and React.JS to link frontend event listeners to backend logic. Lastly, all team members are responsible for simulating user activities and using the metrics defined to test the logic & performance of the website before the showcase in lecture.

Limitations

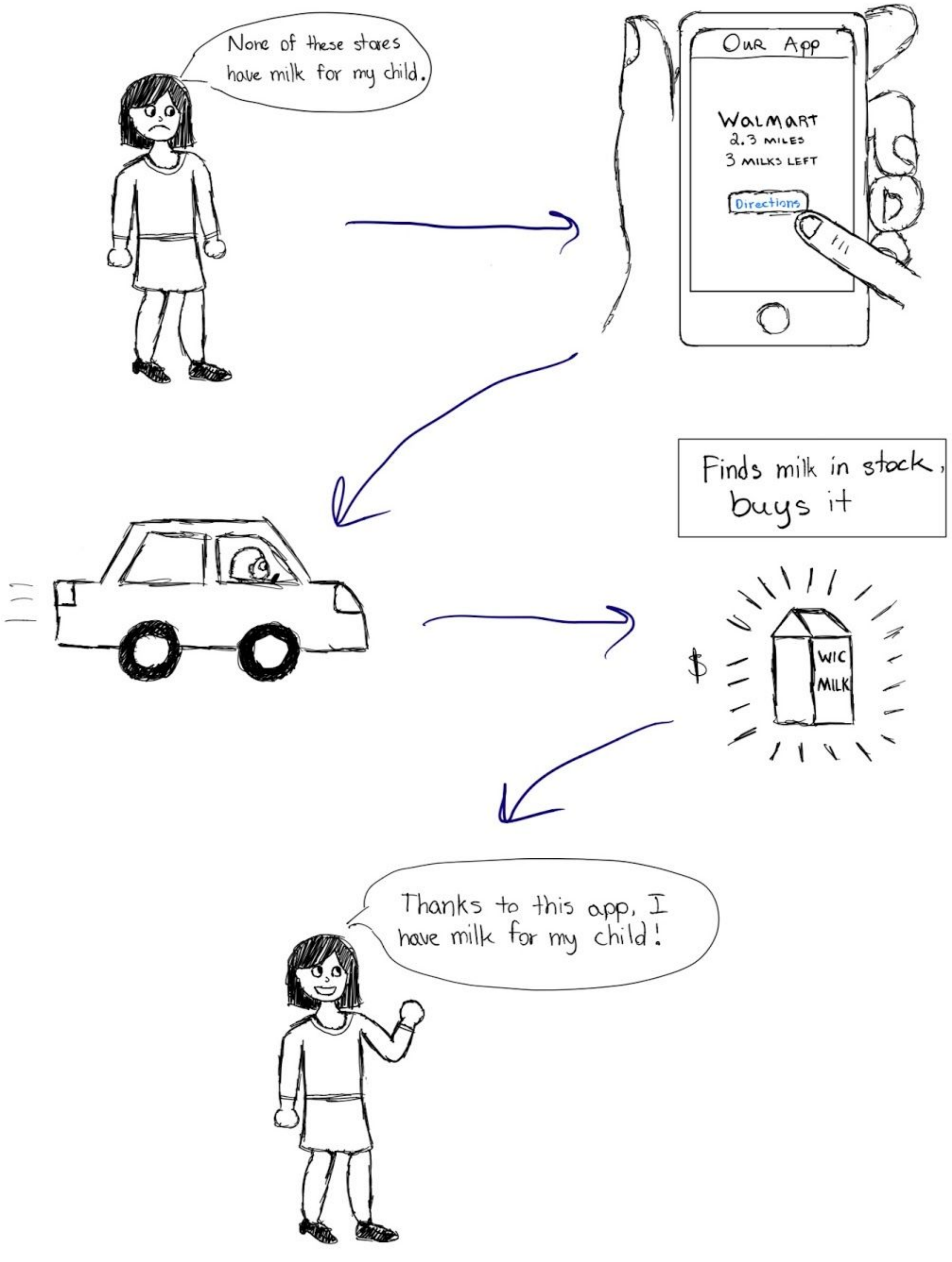
We have encountered a few limitations due to time, technical ability, and data constraints. First and foremost, there is no preexisting data that we have access to which will allow us to start pre-populating the crowdsourced data. In a meeting with the Department of Health, Mariam found that there is an API of all past WIC purchases that include location, a timestamp, and even who made the purchase. While this data would be tremendously beneficial to our database, it provides an ethical concern for the Department of Health. Since this data is so rich, it could be easily exploited by big companies like Nestle or Kroger to boost sales. Therefore, we do not have any pre-existing data to aggregate our crowdsourcing platform.

Moreover, due to the tight time constraints of this project, we do not have time to generate crowdsourced data to test our product. As such, we will be creating simulated data to run user testing. While this allows for some testing because it is simulated data we cannot measure how effective it is in getting food into a person’s hands nor how reliable it is. We can only measure how much it helps people navigate potential shortages before they arrive at the grocery store.

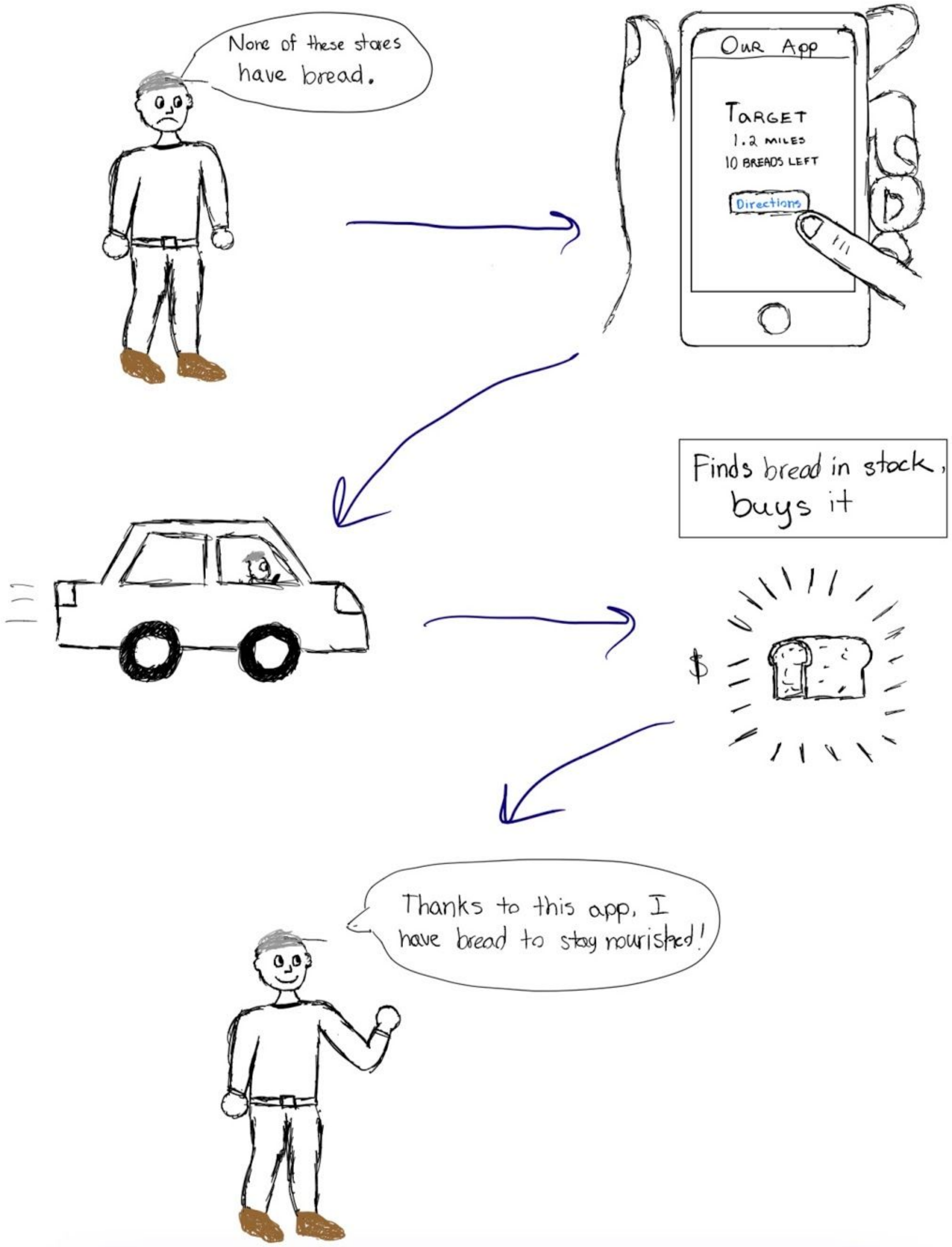
Nonetheless, we are still brainstorming how to continue to measure our outcomes and metrics effectively and in a way that will help us make initial changes as our product slowly builds a larger crowdsourcing community.

Storyboards

Scenario 1: A mother wants to buy food for her child but finds that the stores she frequents are continually out of stock. She downloads our app and searches for a store that is in stock. She finds a store 2.3 miles away with 3 milks and drives to the store to buy it.


None of these stores have milk for my child.
OUR App
WALMART
2.3 MILES
3 MILKS LEFT
Directions
Finds milk in stock, buys it
$
WIC MILK
Thanks to this app, I have milk for my child!

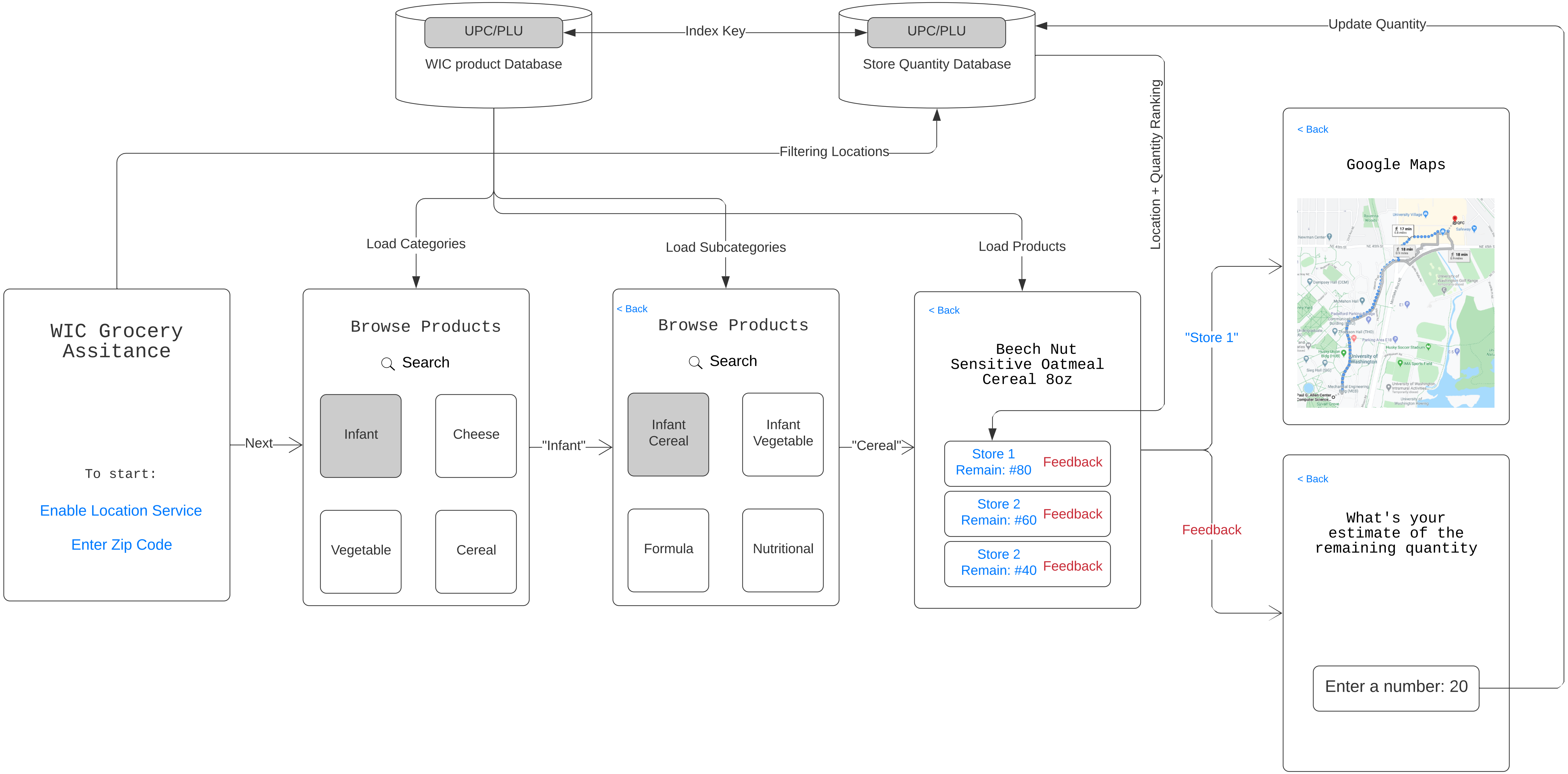
Scenario 2: A an elderly person has been affected by panic buying and is unable to find bread to purchase. He downloads our app and searches for a store that has bread in stock. He finds a store 1.2 miles away with 10 breads and drives to the store to buy it.


None of these stores have bread.
Our App
TARGET
1.2 MILES
10 BREADS LEFT
Directions
Finds bread in stock, buys it
$
Thanks to this app, I have bread to stay nourished!

Part III: Usability Test

Number	Description	WORST	-2	-1	0	1	2	BEST
OVERALL IMPRESSION								
1		TERRIBLE						WONDERFUL
2		DIFFICULT						EASY
3		FRUSTRATING						SATISFYING
4		RIGID						FLEXIBLE
EASY TO USE								
5	It is easy to use.	DISAGREE						AGREE
6	It requires the fewest steps possible to accomplish what I want to do with it.	DISAGREE						AGREE
7	It is flexible.	DISAGREE						AGREE
8	I don't notice any inconsistencies as I use it.	DISAGREE						AGREE
9	I can recover from mistakes quickly and easily.	DISAGREE						AGREE
10	I can use it successfully every time.	DISAGREE						AGREE
11	It's easy to find the product I'm interested in buying	DISAGREE						AGREE
12	It's easy to find a store to go to	DISAGREE						AGREE
13	I can provide a feedback on product quantity easily	DISAGREE						AGREE
EASY TO LEARN								
14	I learned to use it quickly.	DISAGREE						AGREE
15	I easily remember how to use it.	DISAGREE						AGREE
16	It is easy to learn to use it.	DISAGREE						AGREE

17	I quickly became skillful with it.	DISAGREE						AGREE
USER GUIDANCE								
18	I am able to quickly revert your action if I made a mistake,	DISAGREE						AGREE
19	The error messages are helpful.	DISAGREE						AGREE
20	The error messages are informative.	DISAGREE						AGREE
21	The content is well-organized.	DISAGREE						AGREE
22	The categories listed can help me quickly identify the food I want.	DISAGREE						AGREE
23	My search results are relevant to what I searched for.	DISAGREE						AGREE
USER SATISFACTION								
24	Overall, I am satisfied with the ease of completing the tasks using the app.	DISAGREE						AGREE
25	Overall, I am satisfied with the amount of time I use to find food products.	DISAGREE						AGREE
26	Overall, this app saves my time spent on WIC food shopping.	DISAGREE						AGREE
27	Overall, this app provides a good estimate of the remaining products in store.	DISAGREE						AGREE
28	I will continue to use this app and recommend it to my friends.	DISAGREE						AGREE
29	Overall, it was easy to upload information about available products	DISAGREE						AGREE


UPC/PLU
WIC product Database
Index Key
UPC/PLU
Store Quantity Database
Update Quantity
Filtering Locations
Location + Quantity Ranking
Load Categories
Load Subcategories
Load Products
WIC Grocery Assitance
To start:
Enable Location Service
Enter Zip Code
Next
Browse Products
Search
Infant
Cheese
Vegetable
Cereal
"Infant"
< Back
Browse Products
Search
Infant Cereal
Infant Vegetable
Formula
Nutritional
"Cereal"
< Back
Beech Nut Sensitive Oatmeal Cereal 8oz
Store 1 Remain: #80
Feedback
Store 2 Remain: #60
Feedback
Store 2 Remain: #40
Feedback
"Store 1"
Feedback
< Back
Google Maps
< Back
What's your estimate of the remaining quantity
Enter a number: 20