

Milestone 1

CJ Hillbrand

Research Questions

Question One

Question: How does a vegetarian diet compare, cost wise, to a meat-inclusive diet?

Justification: In a day and age where consumers think more about how their choices in the grocery store can affect their environment, consumers want to make more eco-friendly choices. It is important for consumers to be informed on the cost of each choice and how much that can change.

Question Two

Question: Is there a price difference between the meals at Kroger, Safeway, and Trader Joes for a week's worth of meals?

Justification: A consumer is always curious about which grocery stores are the cheapest. Comparing three popular grocery stores gives consumers insights on what may be the cheapest grocery store for them.

Question Three

Question: Is there a price difference between meal prepping (cooking same thing for 1, 2, ... n meals) versus cooking different things for (1, 2, ... n meals)?

Justification: Consumers can have the option to purchase a larger quantity of food to make multiple times throughout the week. This can be a way for a consumer to help save time, if they make the meals at once for the rest of the week.

The downside of this is having to eat the same meal multiple times throughout the week. It is valuable for consumers to discover if meal prepping is doable for any to all meals and understand the cost implications.

Question Four

Question: How much money can consumers save by buying in bulk?

Justification: Many people believe buying bulk saves consumers money in the long run. Although the flip side of this idea is that buying in bulk constitutes items to be in a pantry or fridge for much longer. Consumers can find value in buying in-bulk for some items but may end up wasting food if buying in bulk for quickly expiring food.

Question Five

Question: What is the economic impact of buying Organic versus Non-Organic?

Justification: Like the justification of buying for a meat-inclusive or exclusive diet, buying organic and non-organic has impacts in the consumers environment. Consumers can find value

in considering how to cut costs while still aligning with making positive choices at the grocery store.

Data Source

Source One

The first data source is the Kroger API. The Kroger API offers prices on all Kroger subsidiary grocery stores. I'll be using this data to gather product information on products sold in Kroger grocery stores. This API can be scripted against whose data can then be pulled into a database for further normalization, and labelling.

Source Two

The Safeway website has verbose information on their products and what they have to offer. The second data source is web-scraped data from the Safeway website. The Safeway website offers price, name, and price per unit which can all be scraped and stored locally in a database for further use and labelling.

Source Three

Like data source two, the Trader Joe's website offers information on product name, product size, and price. This data can be pulled from the website via a web scraper and stored locally for further normalization and labelling.

Operationalization Technique

The approach to operationalization is firstly attempting to normalize all the data. Prior to pulling data from the sources, a schema will be established so each data source can be fitted to the schema. This allows the application layer to operate on a congruent data schema.

Secondly to relate individual product items to consumer's buyer habits, several popular recipes will be gathered, broken down by their ingredients and respective quantities and stored locally. Joining the recipe information with the product information can answer the above research questions.

Manual effort will be required to create primary and foreign keys in the recipe and product tables so that they can join seamlessly. Since product names are not usually just "Apple" or "Mustard" they have additional quantifiers, like "Granny Smith Apple" or "Heinz Yellow Mustard". Non-quantified names should be produced, to allow for the joining of products and recipes.

Lastly, labels that are necessary to answer the questions above (i.e., organic, non-vegetarian, etc) Must be applied to operationalize the data.