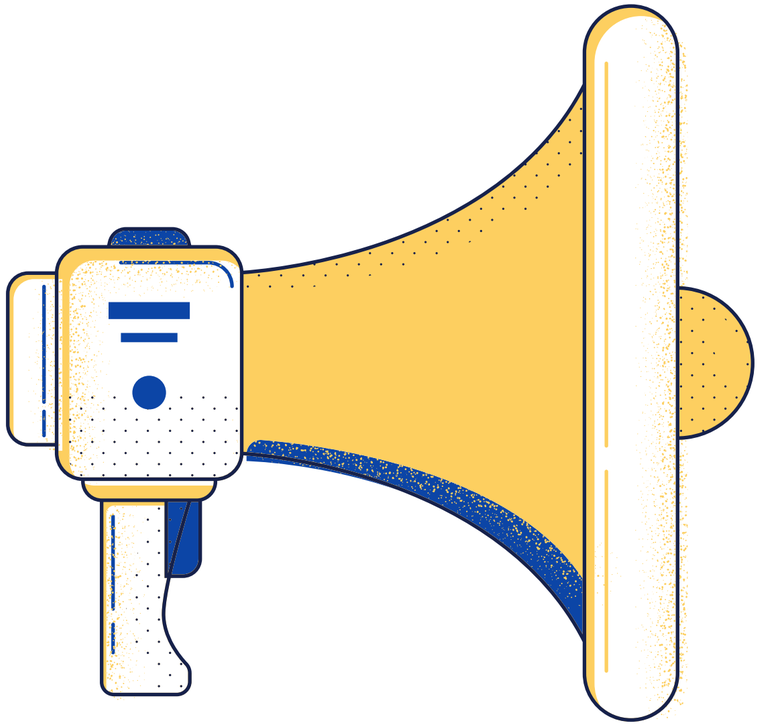
Modeling groceries price

Senior Consultant
Esther Yang

- Model: Here
- Data: Here



How might we understand the variance of grocery prices by product, country, and store brand?

DATA OVERVIEW

Example usage of data

- The base price of 1 liter of full cream milk might be 0.70 €.
- ALDI stores is only 0.9 times as expensive as the average store.
- Germany stores might be 1.4 times more expensive than average country' stores.

Conclusion: 1 liter of full cream milk in an ALDI store in Berlin =
0.70 € × 0.9 × 1.4 = 0.88 € with some random variation around that value.

10 Products

- Apple
- Bananas
- Tomatoes
- Potatoes
- Flour
- Rice
- Milk
- Butter
- Egg
- Chicken Breast

4 Countries

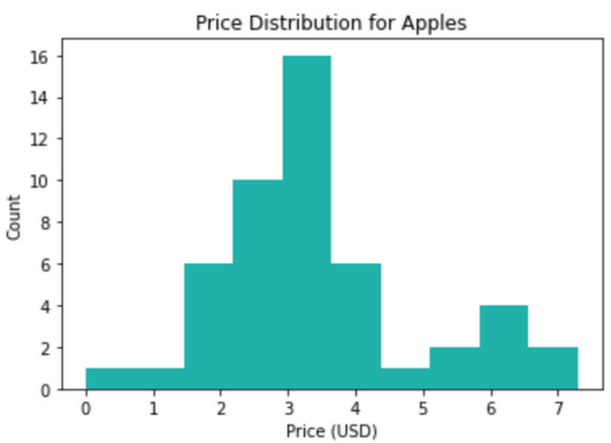
- UK
- US
- Germany
- Korea

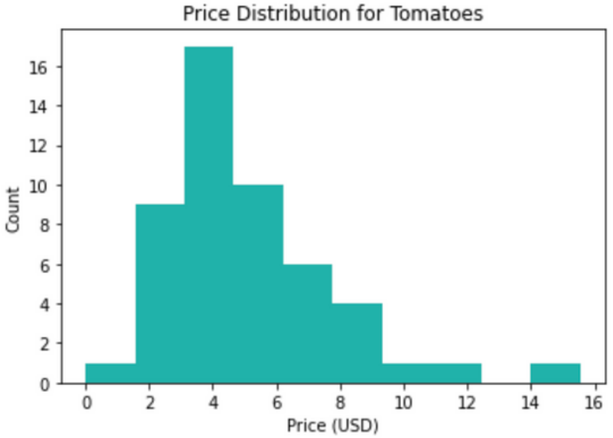
3 Store Brands

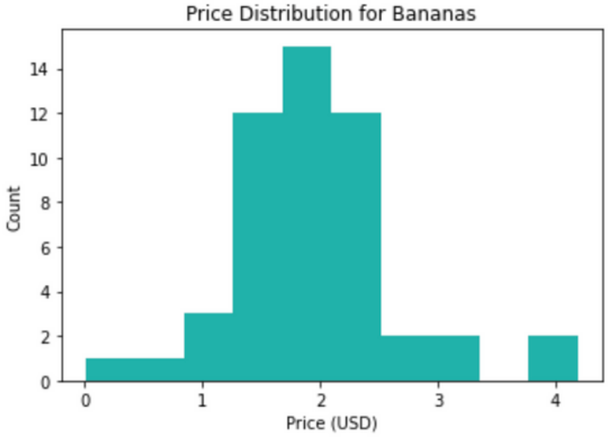
- Cheap
- Mid-range
- Luxury

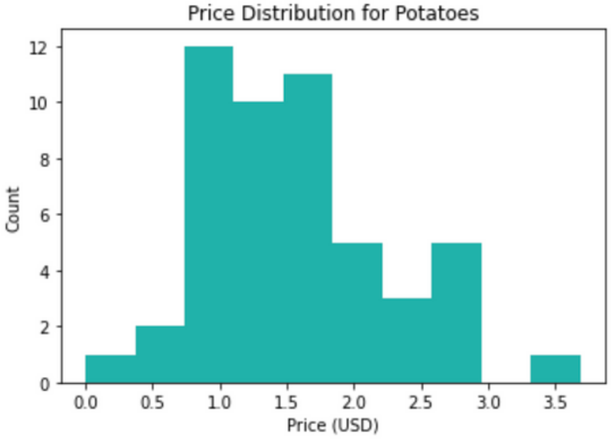
EXPLORE PRICE DATA

We explored the price data after cleaning and normalizing the data, such as deleting unreasonable prices and converting currency into USD. **Here, we show the price distribution across different products.**

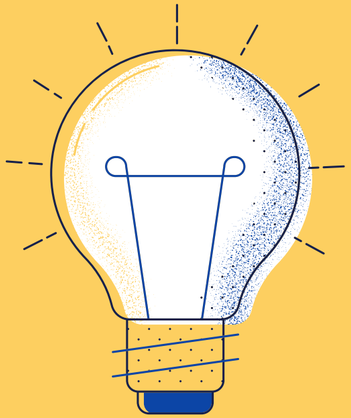








Note: We converted product, currency, brand into ids, so Stan model can read it easily.



BUILDING UP THE MODEL

Priors

We assume three priors

1. **base prices** ~ lognormal distribution
2. **Store multiplier** ~ lognormal distribution
3. **Brand multiplier** ~ lognormal distribution

We assume different vase prices for different products. From price data exploration, we know most product prices follow a lognormal or normal distribution. Since multipliers should be positive continuous vales, we assume they are lognormal distribution too.

Likelihood

We multiply base price, brand multiplier, and brand multiplier and fit in a normal distribution model. In addition, we simulate sigma square from a gamma distribution for the variance in the normal distribution.

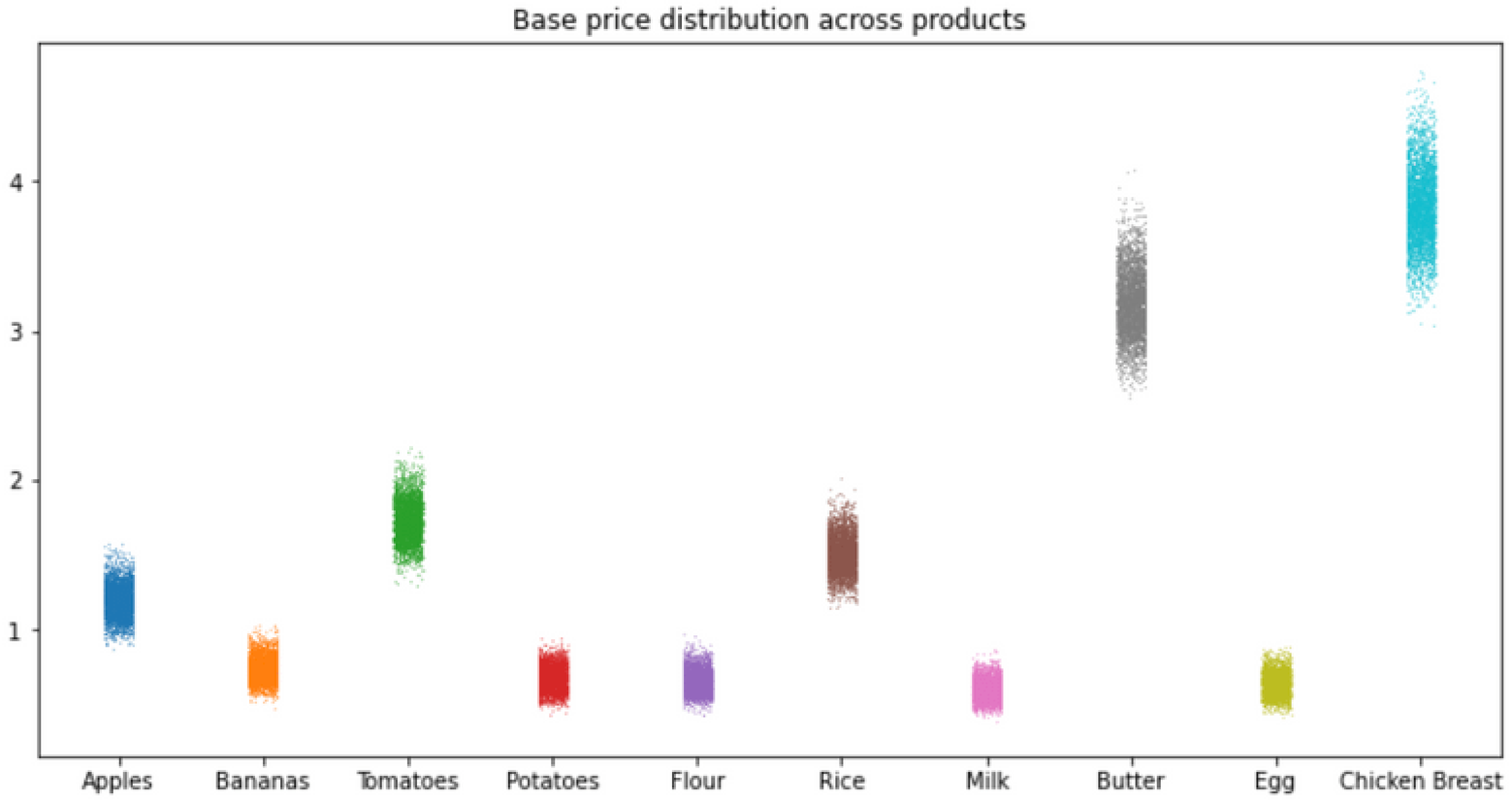
Parameters

The unknown parameter in the Stan model includes brand_multiplier, country_multiplier, base_price, sigma sqruare.

RESULTS

Base prices: Chicken breast has the highest base price.

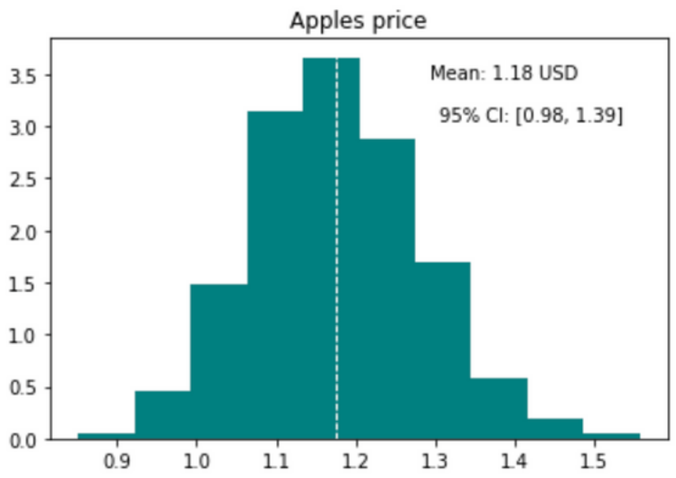
Base price is the universal price before adjusting for country and brand differences.
We can see that chicken breast has the highest base price but with the most variability. Eggs, bananas, milk, flour, potatoes have a low base price.

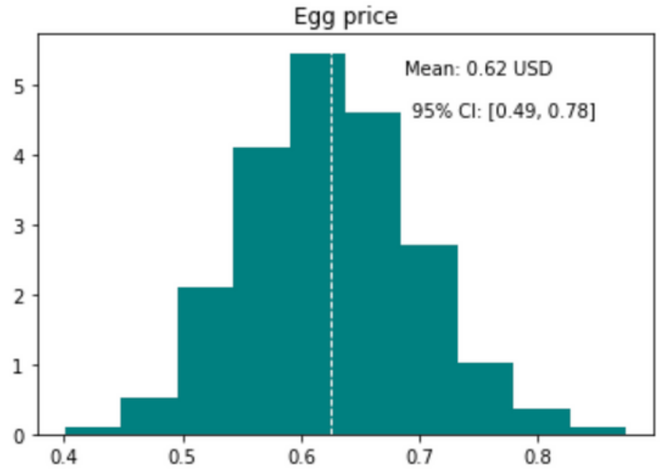


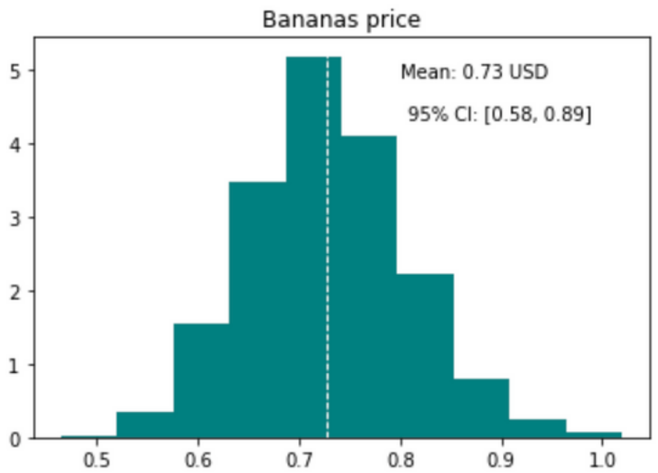
RESULTS

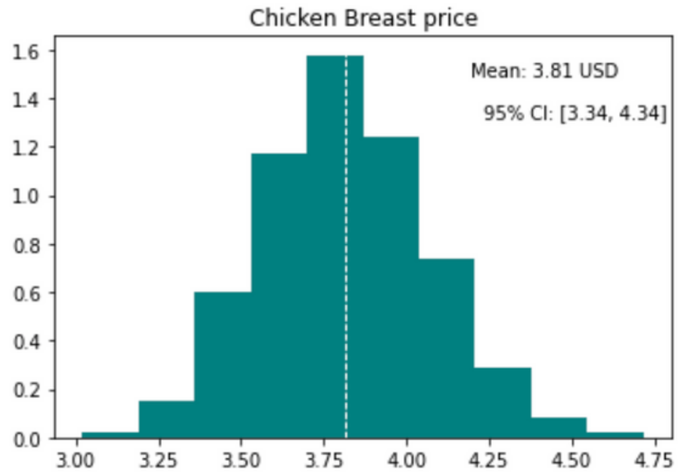
Base prices breakdown

These are the mean and confidence interval for price distribution. We take apples, eggs, bananas, and chicken breast base prices as an example.





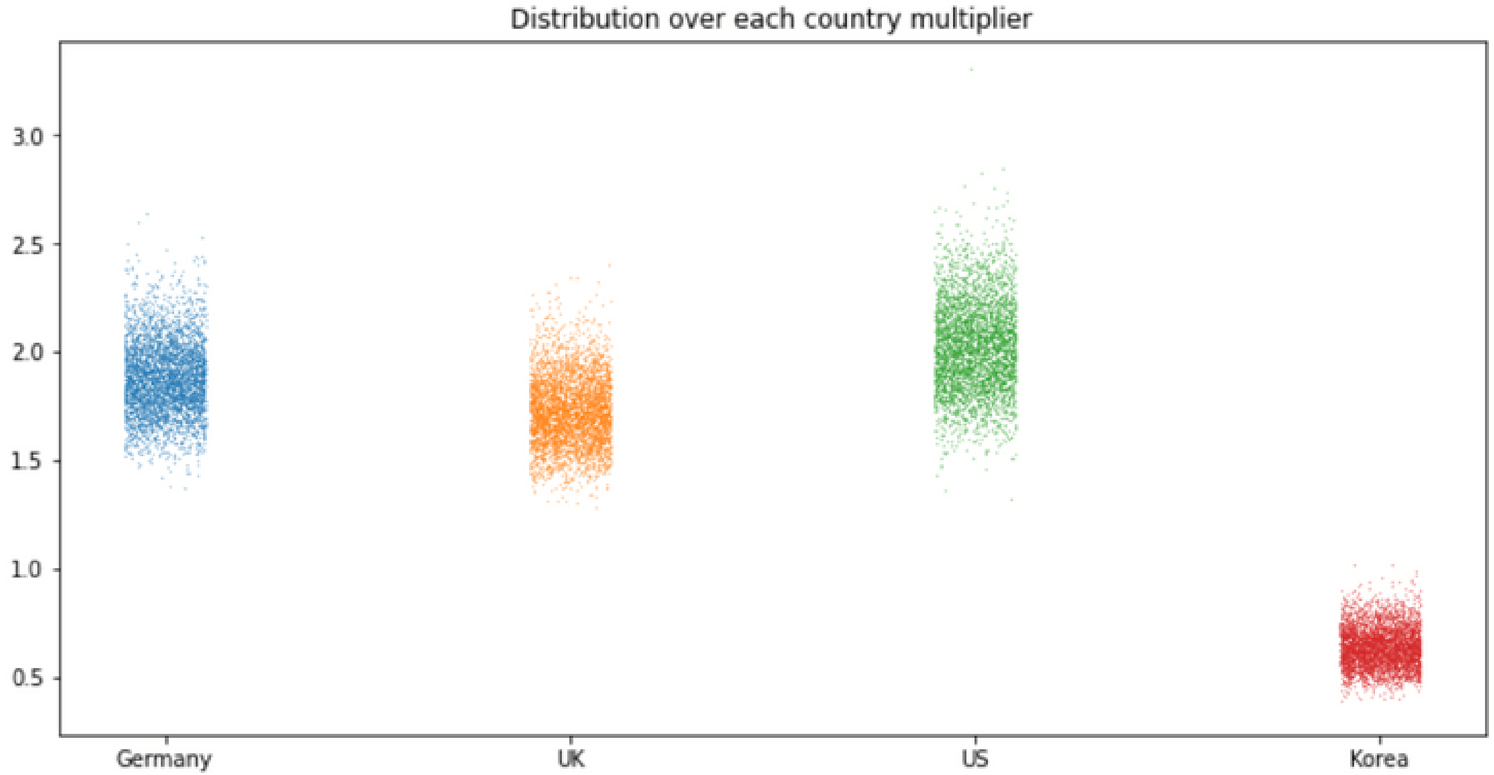




RESULTS

Country multiplier: The U.S. has the highest country multiplier, while Korea has the lowest.

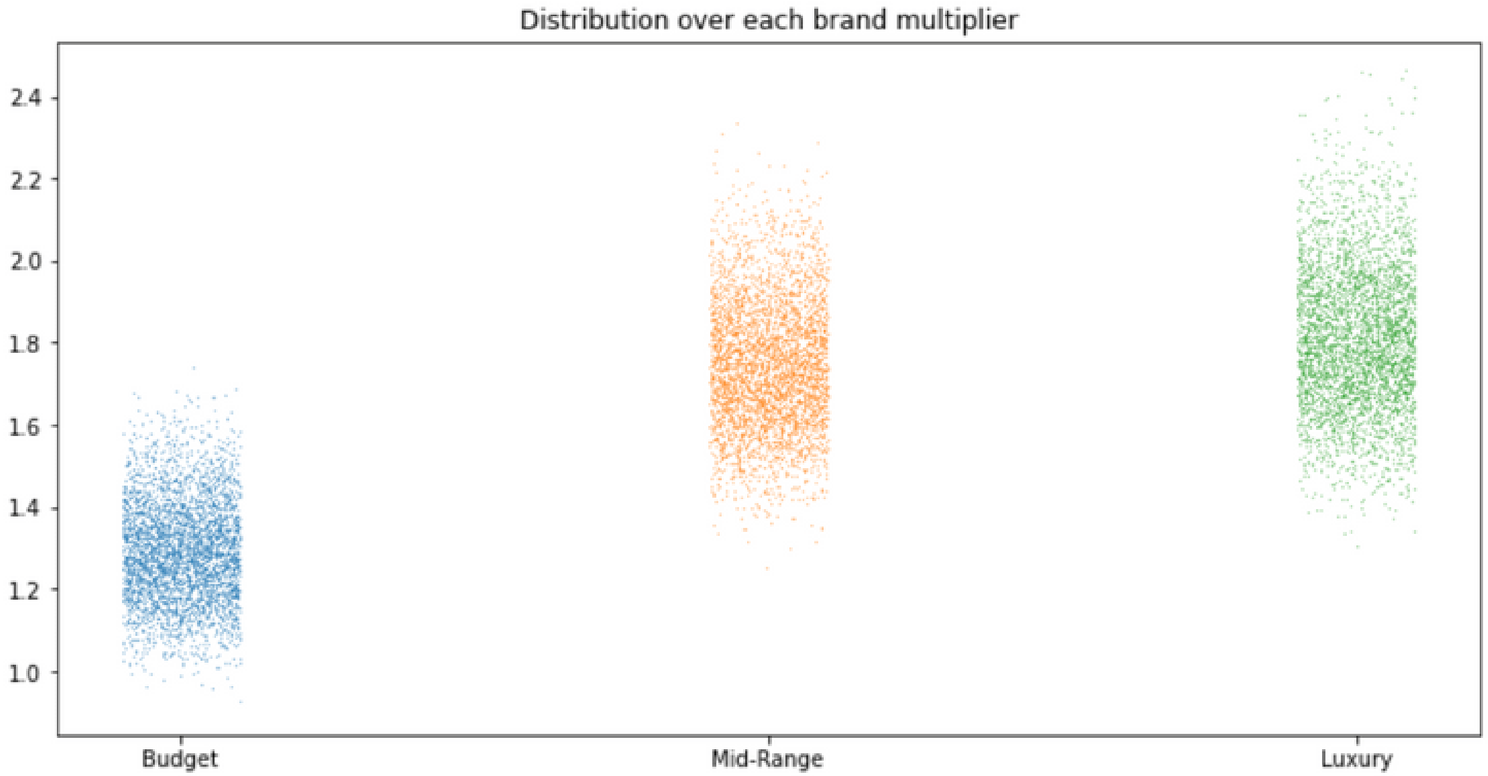
- Germany
  - Mean: 1.87(USD)
  - CI: [1.57,2.23]
- U.K.
  - Mean: 1.72(USD)
  - CI: [1.44,2.05]
- U.S.
  - Mean: 2.01(USD)
  - CI: [1.65,2.44]
- Korea
  - Mean: 0.63(USD)
  - CI: [0.47,0.82]



RESULTS

Brand multiplier: Luxury > Mid-range > Budget

- Budget
  - Mean: 1.29(USD)
  - CI: [1.07, 1.55]
- Mid-range.
  - Mean: 1.74(USD)
  - CI: [1.46, 2.07]
- Luxury
  - Mean: 1.81(USD)
  - CI: [1.5, 2.18]



RESULTS

Correlation between rent and price: There isn't a strong correlation.

We don't consider Korea and U.S. because there aren't enough data points. For Germany and U.K., there isn't a strong correlation.

Germany

	average_rent	price
average_rent	1.000000	-0.090596
price	-0.090596	1.000000

U.K.

	average_rent	price
average_rent	1.00000	0.03777
price	0.03777	1.00000