

**Tartan Data Science Cup Episode II:**  
**I do not like Kroger eggs and ham. I do not like them Sam I am.**  
*By Apoorva Havanur, Rohan Nagalkar, and Shreya Punya*

## Methods

We first approached the problem of prediction which customers will purchase eggs by considering multiple factors that would influence this decision. The factors that we decided to focus on were household size, income, cyclical buying trends, and loyalty. We also investigated the potential effects of complementary goods and overall spend, but decided to eliminate these to make predictions at the household level.

We began our analysis by calculating the proportions of baskets with eggs that each household had and mapping that to their income and household sizes. We also created a frequency table of each commodity for each household to find potential complement goods that are purchased in the same baskets as eggs. Using these values, we were able to approximate the probabilities of purchasing an egg given income size, household size, or another commodity that was purchased.

These values allowed us to calculate a Naive Bayes Classifier to find the probability a household will purchase eggs based on the previously mentioned factors. In order to account for covariance, we calculated the joint probability of multiple independent events (household income, household size, etc.) and removed this from our values. Using these results, we used the following formula to make predictions based on each household's characteristics:

$$P(\text{eggs} \mid \text{basket features}) = (P(\text{basket} \mid \text{eggs})P(\text{eggs})) / (P(\text{basket} \mid \text{eggs})P(\text{eggs}) + P(\text{basket} \mid \overline{\text{eggs}})P(\overline{\text{eggs}}))$$

## Analysis

### *Cyclical Trends in Purchasing*

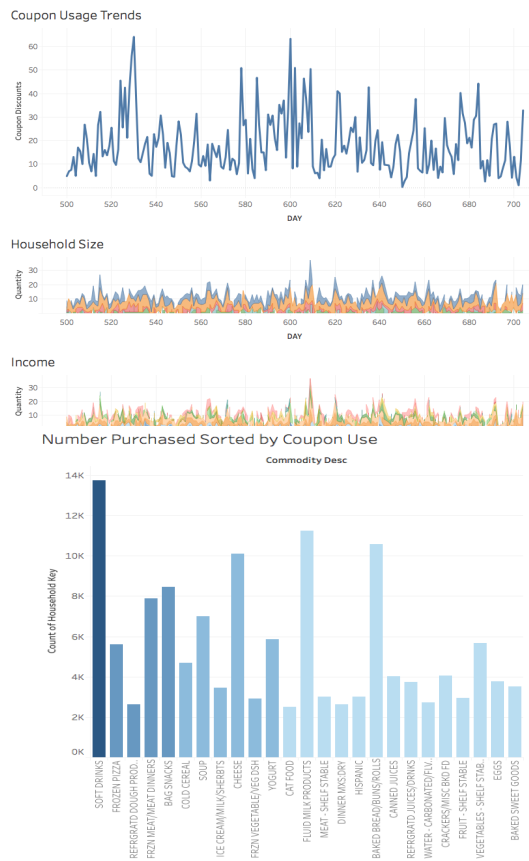
One form of analysis we attempted was to quantify the likelihood of purchase based on whether or not the upcoming week fell into the household's egg-purchasing "cycle". Since eggs are a common purchase, it made sense to assume that a household would buy eggs on a regular basis, whether that be every week, every other week, each month, etc. Therefore, we calculated the number of days in between each purchase of eggs for each household, and plotted the data to look at the distribution of the differences. We found that most of the time, a household would buy eggs with regularity, which we determined by subtracting the average number of days they spent between purchases from each purchase gap. We found that for most households, consumers tended to buy eggs within 5 days before or after they normally would (i.e, if they usually went 20 days in between trips for eggs, they would often go between 15 and 25 days for any given trip). From that, we modeled the delay as an exponential distribution to calculate the probability that a household would be buying eggs based on how many days it had been since they last purchased eggs, and how that compared to the average number of days they usually had in between eggs, and used that as a factor as to whether or not they would purchase in the upcoming week.

### *Loyalty/Promotion Insights*

Upon calculating the timing of the purchases of eggs, we took a closer look into whether promotions and discounts, such as those obtained through a coupon or loyalty membership, positively or adversely affected egg sales. The advent of saving money is surely a reasonable motive to make a trip to Kroger, but our group suspected that eggs are a necessity and would not be directly impacted by the lack of potential discounts available. We were correct: customers who

bought eggs had an average savings percent that was 4.88%, which is 5.79% less than the average savings per transaction. On average, they paid \$1.54. This led us to decide that eggs are an inelastic normal good, and validated our previous hypothesis that timing is everything in this problem. However, we found some encouraging results regarding loyalty and promotion discounts. Customers save a lot on complementary goods that they purchase with eggs:

Complementary Product	Probability of Buying with Eggs	Average Savings on Product
BACON	41.34%	18.11%
MARGARINES	37.34%	8.17%
BREAKFAST SAUSAGE/SANDWICHES	34.04%	15.76%
BUTTER	33.90%	11.32%



## Results and Conclusion

Based on our predictions on what factors predict the purchasing of eggs,, we also looked at trends in consumer activity regarding coupon use and the timeline of egg purchases. Figure 1.1 (left) displays a timeline of the 200-day period by comparing the total savings in coupons during this time. Figures 1.2 and 1.3 show the total spending on groceries by household size and household income to display changes in spending trends over the course of this same time period. By comparing the activity of both charts, we found that many times, spikes in coupon usage is unrelated to spending on eggs. Instead, from our results, as qualified by Figure 2, we found that consumers were more likely to use discounts on complementary goods when purchasing eggs, which we believe will be a useful strategy in promoting eggs.

In addition, our group discussed that if we were given more time and had external data, some important factors that could help us reach a better prediction would be customer geography and shopping habits outside of Kroger Foods. While this lies outside the scope of the data set, we believe that distance from the grocery store is a huge confounding variable that influences shopping

habits. A customer who lives further away from Kroger is relatively less likely to shop for eggs there then.

To conclude, our group emphasizes that Kroger should reevaluate its promotional offerings and market complementary goods during the off periods where buying eggs is not popular.