SpringBoard.com
Data Science Intensive
Capstone Project

**Family Food
In UK Shopping Carts**

Lucas González Santa Cruz
Oct 6th 2016

Research goal

- Food consumption in the UK can be explored through the purchases of food surveyed in a sample of households.

- What are the determinants of "good shopping carts", defined as those where "junk food" is below the average?

Good shopping carts

- Food groups of clearly "low nutritional value":

  - 'ALCOHOLIC DRINKS', 'BEVERAGES', 'BISCUITS, CAKES, BUNS, CRISPBREADS', 'CONFECTIONERY', 'SOFT DRINKS', 'SUGAR AND PRESERVES'

- Foods of "mixed or high nutritional value":

  - 'ALL BREAD', 'ALL CARCASE MEAT', 'ALL FATS', 'ALL FISH', 'ALL NON-CARCASE MEAT AND MEAT PRODUCTS', 'ALL OTHER FOODS', 'ALL PROCESSED VEGETABLES', 'BEVERAGES', 'BISCUITS, CAKES, BUNS, CRISPBREADS', 'CEREALS, EXCL. BREAD,BUNS,CAKES,BISCUITS', 'CONFECTIONERY', 'EGGS', 'FLOUR', 'FRESH FRUIT', 'FRESH GREEN VEGETABLES', 'FRUIT & FRUIT PRODS. NOT FRESH', 'LIQUID WHOLEMILK, INC SCHOOL & WELFARE', 'OTHER FRESH VEGETABLES', 'OTHER MILK & CREAM', 'POTATOES', 'SOFT DRINKS', 'SUGAR AND PRESERVES', 'TOTAL CHEESE'

- Score: number of items bought per week/total

- Classification: score > median

Dataset and processing

http://data.defra.gov.uk/Food/NationalFoodSurvey

- 6 years: 1995-2000
- 42,960 families (53 descriptive variables per family + one week of purchases)
- 1,581,381 food items bought
- 243 food-items in 24 food-groups

- Python script
- Logistic Regression

Proportion of "good food items" in family shopping carts
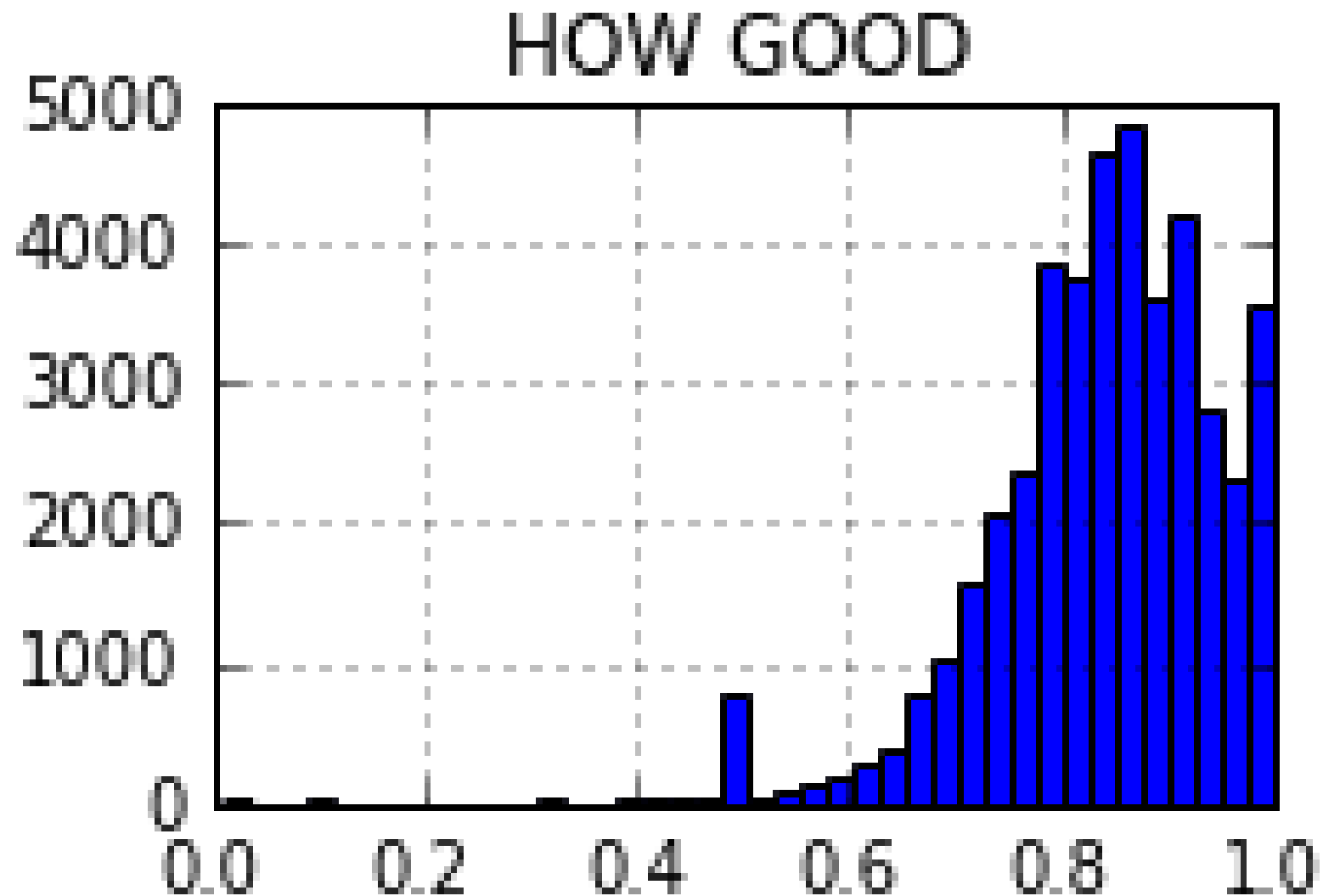
Logistic Regression

- Most important predictive variables: those related to income (including owning the home, and the job status of the head of the household), the geographical region, and the number of children.

- This smaller set of predictive variables does as good a job as the full set of variables.

- The model predicts only a bit better than throwing a dice (55% vs 50%).

Next steps

- Re-create, with experts in nutrition, the index for what a "good cart" is.

- Explore several alternative diets (supply chain disruption that forces the population to eat locally produced food only, or a sudden increase in interest in vegetarian diets).

- Thank you!