**GfK National Shoppers Lab Data**

Shopper card-based data from a grocery retailer that offers both instore and online shopping.

**Basket Files:**

These files are summaries of transactions by household ID (HHID) by date. A transaction is summarized by HHID, store, date, and time of day.

* Files:
  + Summary\_Basket\_2019.csv (14,190,773 records) – 1/1/2019-12/31/2019
  + Summary\_Basket\_2020.csv (13,068,051 records) – 1/1/2020-12/31/2020
  + Summary\_Basket\_2021.csv (24,983,103 records) – 1/1/2021-12/31/2021
* Fields:
  + HHID – household ID
  + Store – store number
  + Date – date of transaction
  + Totqty – total units bought in transaction
  + Totdol – total dollars spent in transaction
  + Online – 0=Instore purchase, 1=Online purchase

**UPC Level Data:**

These files are UPC level purchases in the All Purpose Cleaners and Ready to Eat Cereal categories.

* Files:
  + UPC\_Hack\_Cleaners2.csv (549,661 records) – 1/1/2019-12/31/2021
  + UPC\_Hack\_Cereal2.csv (6,606,866 records) – 1/1/2019-12/31/2021
* Fields:
  + HHID – household ID
  + Store – store number
  + UPC – product UPC (see Hackathon Data Maps.xlsx for product descriptions and brand groupings)
  + Date – date of transaction
  + Qty – units bought of item
  + Price – dollars paid for item
  + Online – 0=Instore purchase, 1=Online purchase

**Hackathon Data Maps.xlsx** – contains descriptions, category, and parent brand/brand grouping for the RTE Cereal and AP Cleaners categories.

**Notes:**

* This data comes from a subset of stores within a larger retailer.
* HHIDs and store numbers have been masked but are consistent across all files (meaning store 1 is the same store in all files and the same goes for HHIDs).
* Households were selected because they purchased either a Ready to Eat Cereal or All Purpose Cleaner item at some point in the 3 year data period.
* There are 447,542 unique HHIDs and 56 unique store numbers.