



34.51°

TARTAN DATA SCIENCE CUP

Data Dictionary

TARTAN DATA SCIENCE CUP: DATASET DETAILS

This dataset contains household level transactions over 205 days from a group of 967 households who are frequent shoppers at Kroger. Each row of the dataset is essentially the same as what would be found on a grocery store receipt.

The columns correspond to information about each product sold and demographic information for a portion of households. (Due to nature of the data, the demographic information is not available for all households.) The columns / variables in the dataset are described below:

transaction_data_8451

Variable	Description
household_key	Uniquely identifies each household
BASKET_ID	Uniquely identifies a purchase occasion
DAY	Day when transaction occurred
PRODUCT_ID	Uniquely identifies each product
QUANTITY	Number of products purchased during the trip
BASE_SPEND_AMT	Dollar amount of products(s) without any discounts
LOY_CARD_DISC	Discount applied due to retailer's loyalty card program
COUPON_DISC	Discount applied due to coupon
NET_SPEND_AMT	Final dollar amount household pays
DEPARTMENT	Groups similar products together
COMMODITY_DESC	Groups similar products together at a lower level
SUB_COMMODITY_DESC	Groups similar products together at the lowest level
AGE_DESC	Estimated age range
MARITAL_STATUS_CODE	Marital Status (A - Married, B - Single, U - Unknown)
INCOME_DESC	Household Income
HOMEOWNER_DESC	Homeowner, renter, etc.
HH_COMP_DESC	Household composition
HOUSEHOLD_SIZE_DESC	Size of household up to 5+
KID_CATEGORY_DESC	Number of children present up to 3+

It may be important to calculate the price per product. To do so, use this formula:

- $PRICE_PER_PRODUCT = NET_SPEND_AMT / QUANTITY$