dunnhumby | SOURCE FILES

Breakfast at the Frat: A Time Series Analysis

USER GUIDE

BREAKFAST AT THE FRAT: A TIME SERIES ANALYSIS

Breakfast at the Frat contains sales and promotion information on the top five products from each of the top three brands within four selected categories (mouthwash, pretzels, frozen pizza, and boxed cereal), gathered from a sample of stores over 156 weeks. Included in this Source File:

- Unit sales, households, visits, and spend data by product, store, and week
- Base Price and Actual Shelf Price, to determine a product's discount, if any
- Promotional support details (e.g., sale tag, in-store display), if applicable for the given product/store/week
- Store information, including size and location, as well as a price tier designation (e.g., upscale vs. value)
- Product information, including UPC, size, and description

This Source File can be successfully used in classroom projects and case studies, and is ideally suited for this use. It allows students to interact with 'real-world' data and search for their own insights. The richness of this data and the potential analyses it enables make it a valuable teaching tool.

The data set also provides students with the opportunity to understand the process required to mine data. Since this is a relational database, students will need to merge multiple data tables together and aggregate data in search of insights.

EXAMPLE QUESTIONS

The following are examples of questions that could be submitted to students from this data set:

- What is the range of prices offered on products?
- Are there specific price thresholds that, if crossed, drive significant differences in sales?
- What is the price elasticity of the products? What will happen to unit sales if the price changes by X%?
- How are sales impacted by changing price gaps between items?
- What is the impact on sales of promotions, displays, or being featured in the circular?
- What is the impact on units/visit of promotions?
- How do the above differ by products? By categories?
- How do the above differ by geographies (e.g., southwestern states vs. midwestern states)?
- How do the above differ by store price tier (e.g., upscale stores vs. value stores)?
- If you were the retailer, for which products would you be more likely to lower the price to increase sales? Why?
- If you were the retailer, for which products would you be more likely to increase the price to improve profits? Why?

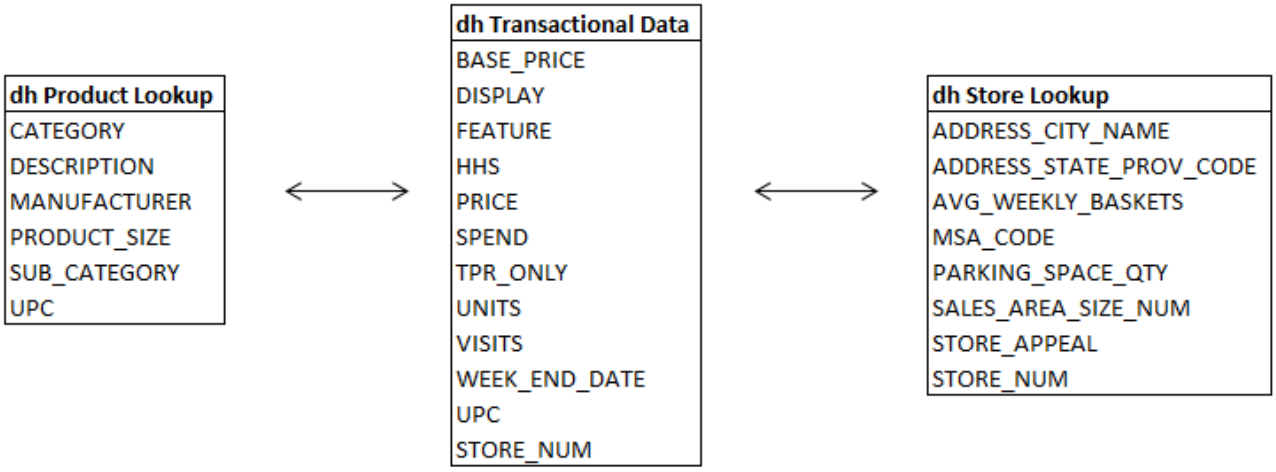
DATA OUTLIERS

Please note, there are some 'real-world' learnings regarding the nature of product/store/week-level data. Of the 525,000 rows of data in this file, we estimate there to be less than 0.5% that might be labeled as "outliers", but they do exist. To identify outliers, it is suggested that students:

- **Create and examine a Units/Visit metric.** When comparing the ratio of units vs. number of visits, students will identify some observations that appear very high (e.g., 13 units/visit on a 15oz. bag of pretzels).
- **Create and examine a Visits/Household metric.** When comparing the ratio of visits vs. number of households, students will identify some observations that appear very high (e.g., 9 visits in a single week by a household to purchase cereal).
- **Look for unusually low sales levels.** In some cases, an item may be out-of-stock or discontinued for a store.

Again, the number of outlier observations is very low, but how to address them – leave them in the dataset, adjust these observations, delete them – will be up to the student. It is also useful to note that in most cases these outliers are real. For example, there actually is a household in Texas that purchases $5,000 of mini-pretzels annually!

BREAKFAST AT THE FRAT: DATASET DETAILS



dh Product Lookup

Description: Provides detailed product information for each upc in 'dh Transaction Data'.	
# of Records: 58	
VARIABLE NAME	**DESCRIPTION**
CATEGORY	category of product
DESCRIPTION	prodcut description
MANUFACTURER	manufacturer
PRODUCT_SIZE	package size or quantity of product
SUB_CATEGORY	sub-category of product
UPC	(Universal Product Code) product specific identifier

BREAKFAST AT THE FRAT: DATASET DETAILS

dh Transaction Data

Description: This table contains a sample of 156 weeks of mouthwash, pretzels, frozen pizza, and boxed cereal transactions, at the product level by store, by week.

# of Records: 524,950

VARIABLE NAME	DESCRIPTION
BASE_PRICE	base price of item
DISPLAY	product was a part of in-store promotional display
FEATURE	product was in in-store circular
HHS	# of purchasing households
PRICE	actual amount charged for the product at shelf
SPEND	total spend (i.e., $ sales)
STORE_NUM	store number
TPR_ONLY	temporary price reduction only (i.e., shelf tag only, product was reduced in price but not on display or in an advertisment)
UNITS	units sold
VISITS	number of unique purchases (baskets) that included the product
WEEK_END_DATE	week ending date
UPC	(Universal Product Code) product specific identifier

dh Store Lookup

Description: Provides detailed store information for each store in 'dh Transaction Data'.

# of Records: 79

VARIABLE NAME	DESCRIPTION
ADDRESS_CITY_NAME	city
ADDRESS_STATE_PROV_CODE	state
AVG_WEEKLY_BASKETS	average weekly baskets sold in the store
MSA_CODE	(Metropolitan Statistical Area) geographic region with a high core population density and close economic ties throughout the surrounding areas
PARKING_SPACE_QTY	number of parking spaces in the Kroger parking lot
SALES_AREA_SIZE_NUM	square footage of Kroger store
STORE_APPEAL	Kroger's designated store appeal
STORE_NUM	store number

CONTACT INFORMATION

For general questions about dunnhumby or the Source Files programme, or for technical questions regarding the use of this dataset, please contact:

DUNNHUMBY SOURCE FILES SUPPORT
E: sourcefiles@dunnhumby.com