Breakfast at the Frat: A Time Series Analysis

User guide

dunnhumby

Breakfast at the frat: a time series analysis

Breakfast at the Frat contains a representation of sales and promotion information on five products from three brands within four categories (mouthwash, pretzels, frozen pizza, and boxed cereal), 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.**

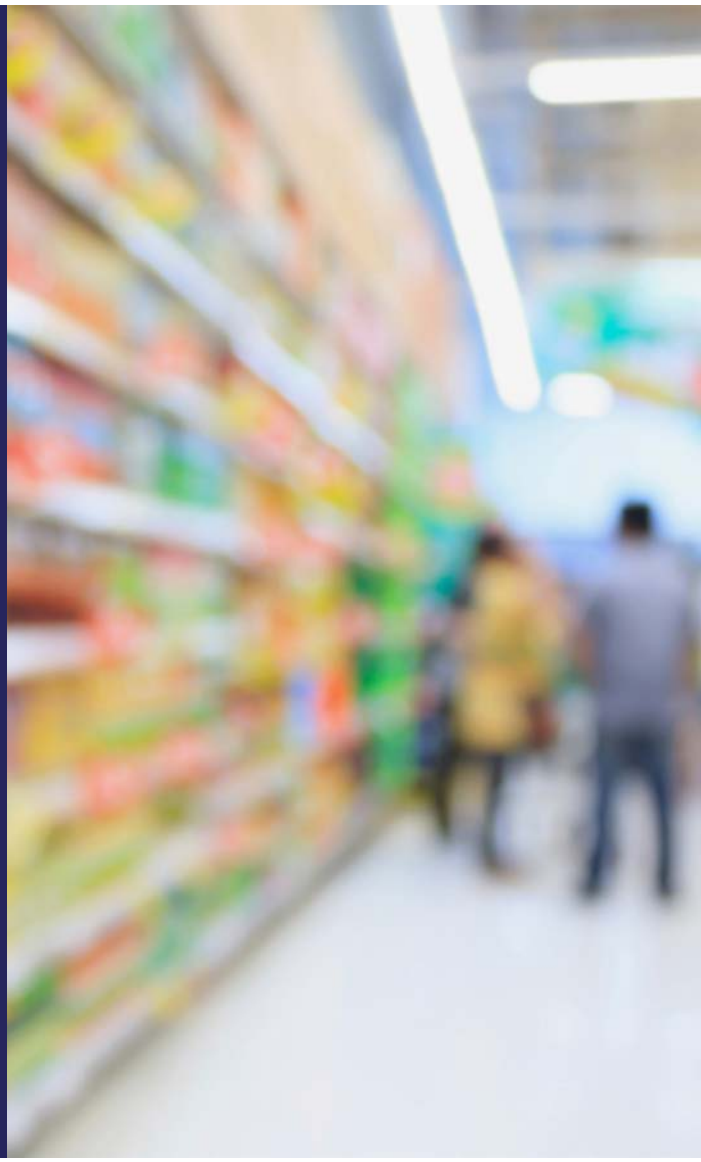
This Source File can be successfully used in classroom projects and case studies and is ideally suited for this use. It allows students to interact with 'real-world'-like data and search for their own insights. The richness of this data and the potential analyses it enables make it a valuable teaching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



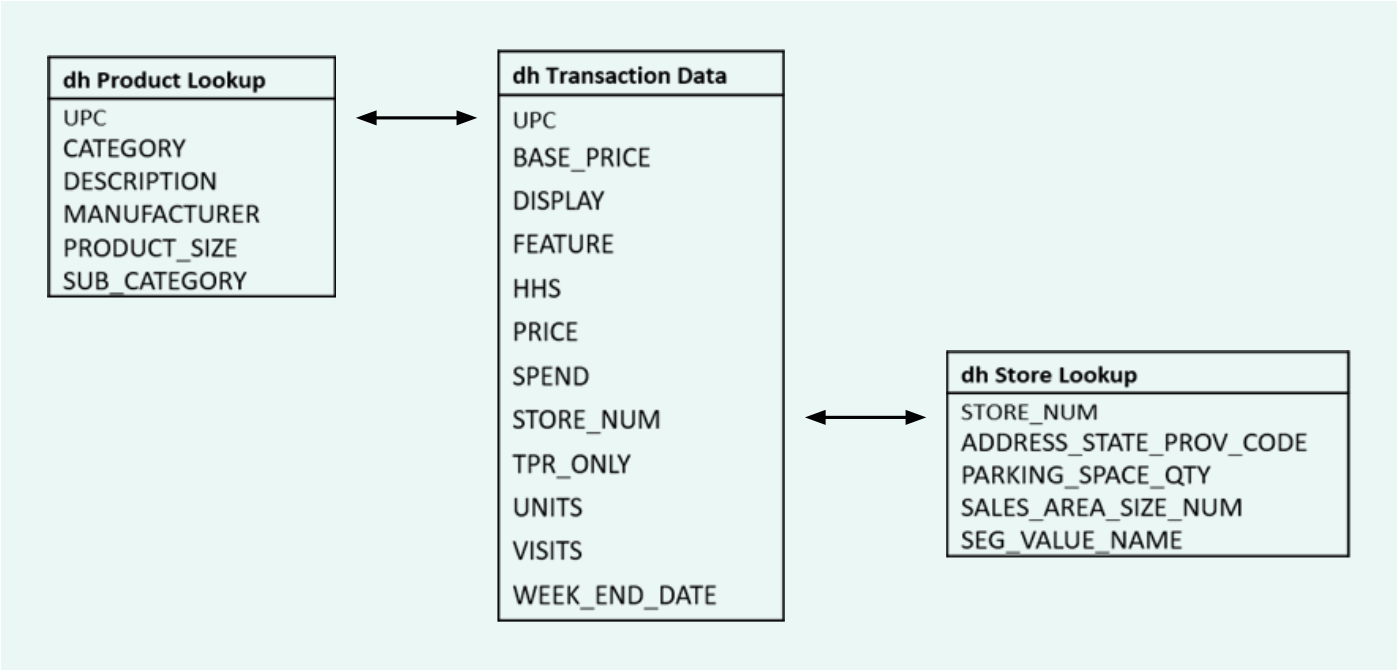
Data outliers

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 labelled 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 these outliers reflect what happens in the real world: some households do spend thousands of dollars on mini-pretzels annually!

Breakfast at the frat: dataset details



dh Transaction Data

Description: This table contains 156 weeks of Mouthwash, Pretzels, Frozen Pizza and Boxed Cereal transactions, at the product level by store, by week.
Number of records: 524,950

Variable name	Description
base_price	Base price of item
display	Indicator of whether product was part of in-store display. 1 if on display, 0 if not.
feature	Indicator of whether product was in in-store circular. 1 if in the circular, 0 if not.
hhs	Number of purchasing households
price	Actual amount charged for the product at shelf
spend	Total spend (ie. dollar sales)
store_num	Unique store identifier
tpr_only	Temporary price reduction only (ie. shelf tag only, product was reduced in price but not on display or in an advertisement). 1 if on temporary price reduction only, 0 if not.
units	Units sold
visits	Number of unique purchases (baskets) that included the product
week_end_date	Week ending date
upc	Standard unique product code

dh Product Lookup

Description: Provides detailed product information for each upc in 'dh Transaction Data'.
Number of records: 58

Variable name	Description
category	Category of the product
description	Product description
manufacturer	Manufacturer
product_size	Package size or quantity of the product
sub_category	Sub-category of the product
UPC	Standard unique product code

dh Store Lookup

Description: Provides detailed store information for each store in 'dh Transaction Data'.
Number of records: 79

Variable name	Description
address_state_prov_code	State
parking_space_qty	Number of parking spaces in the parking lot
sales_area_size_num	Square footage of store
seg_value_name	Designated store value segment. Possible values: Value, Mainstream and Upscale
store_num	unique store identifier

CONTACT INFORMATION

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

sourcefiles@dunnhumby.com

dunnhumby THE WORLD'S FIRST CUSTOMER DATA SCIENCE PLATFORM

dunnhumby is the global leader in Customer Data Science, empowering businesses everywhere to compete and thrive in the modern data-driven economy. We always put the Customer First. Our mission: to enable businesses to grow and reimagine themselves by becoming advocates and champions for their Customers.

With deep heritage and expertise in retail — one of the world's most competitive markets, with a deluge of multi-dimensional data — dunnhumby today enables businesses all over the world, across industries, to be Customer First.

The dunnhumby Customer Science Platform is our unique mix of technology, software and consulting enabling businesses to increase revenue and profits by delivering exceptional experiences for their Customers – in-store, offline and online. dunnhumby employs over 2,000 experts in offices throughout Europe, Asia, Africa, and the Americas working for transformative, iconic brands such as Tesco, Coca-Cola, Meijer, Procter & Gamble, Raley's and L'Oreal.

Connect with us to start the conversation

dunnhumby.com