

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT



# Retail Scanner Dataset Manual

---

## The Kilts Center Archive of The Nielsen Company

1/22/2016

## Contents

I.	Introduction .....	4
A.	Purpose .....	4
B.	Technical notes regarding data.....	4
C.	Data Description and Summary .....	5
	• Weekly Price/Sales Volume/Store Merchandising Conditions .....	5
	• Products .....	5
	• Stores, Retail Chains and Channels .....	6
II.	Files and File Layouts .....	7
A.	Overview of Files.....	7
B.	File Layouts .....	9
	Table 1: Stores Files .....	9
	Table 2: Products File .....	11
	Table 3: Movement Files.....	14
	Table 4: RMS Versions File .....	18
	Table 5: Product Extra Attributes Files .....	18
	Table 6: Brand Variations File .....	21
C.	Other Tables.....	22
	Table 7: Kilts Center Store Coverage - All Commodity Volume (ACV) by Channel and Scantrack Market.....	22
	Table 8: Channel Codes.....	24
D.	Data Notes .....	25
	Retailer Level Analysis Between the Retail Scanner Dataset and Consumer Panel Dataset .....	25
III.	Appendix I: Policies .....	26
IV.	Appendix II: File Names and Sizes.....	28
V.	Appendix III: Maps .....	29
	.....	30

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

VI.	Appendix IV: Version revisions.....	33
A.	Revisions 2016-1-22 .....	33
B.	Revisions 2015-05-25.....	33
C.	Revisions 2014-09-22.....	34
D.	Revisions 2014-04-02.....	34
E.	Revisions 2014-03-05.....	34

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

## **I. Introduction**

The Retail Scanner Data (also referred to as RMS data) consists of weekly pricing, volume, and store merchandising conditions generated by participating retail store point-of-sale systems in all US markets. Data are included from approximately 35,000 participating grocery, drug, mass merchandiser, and other stores. Products from all Nielsen-tracked categories are included in the data, such as food, non-food grocery items, health and beauty aids, and select general merchandise. Currently, the years 2006-2011 are covered. We expect to update the data on an annual basis each January. Updates are expected to be available in the first quarter of each calendar year, and will always lag by 2 years (e.g. 2013 Retail Scanner data is expected to be released in January 2015).

### **A. Purpose**

The purpose of this document is to provide a basic understanding of the Nielsen Retail Scanner datasets for approved users conducting academic research.

All academic use of this data must be registered and approved by the Kilts Marketing Data Center at University of Chicago Booth School of Business.

The Nielsen – Kilts datasets will be approved for academic research projects focused on important issues of wide and general interest. Narrowly focused projects on subjects of interest to only specific industries or firms may be declined.

Access to these data is designed to facilitate research projects initiated by individual academic researchers. Research projects commissioned, sponsored or funded by an industry source, government organization, industry group, nonprofit organization or private company are prohibited, unless permission is explicitly granted by Nielsen. Use of the data for consulting purposes of any kind, including litigation support, is strictly prohibited. Also prohibited are any projects with a primary purpose of evaluating Nielsen's data collection approach or methodologies.

All approved researchers are bound by policies found in Appendix I: Policies (page 26) as well as any additional limitations named in their respective contract between Chicago Booth and their individual academic institution.

## **B. Technical notes regarding data**

When importing files, users should be aware of the following.

1) For all files with UPCs in them, many software packages try to read the UPC as a single precision float. However, this data type does not have 12 digits of guaranteed precision, which will cause some UPCs to

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

be rounded to a different value on import, corrupting the data. Using a data type with more precision, such as double precision float or string, is necessary.

2) For the products and brand variations file, some fields contain literal single or double quote characters, e.g. in the brand\_descr field. Many import parsers fail to properly handle the way we have written these characters. It may be necessary to pre-process or use 3rd party importers if your software does not handle the file correctly.

Note for individuals submitting publications: Data samples may not be submitted to any outside party. If a journal indicates such a requirement, we suggest that the author(s) explain at the time of submission that the data are accessible to others who wish to replicate analyses. We advise that author(s) should make this clear and request permission of the editor to proceed with the submission.

### C. Data Description and Summary

The dimensions that describe the structure of the retail scanner dataset include the weekly price/volume/store environment information, products, geographic markets, and retail channels and stores.

- **Weekly Price/Sales Volume/Store Merchandising Conditions:** Each individual store reports weekly data for every UPC code that had any sales volume during the week. Information captured for each UPC code includes: week ending date, units sold, weighted average unit price, price multiple (e.g. 3 units for \$1.00 would have a multiple of 3 and a price of \$1.00). Some stores also report whether the product was featured or on display during each week. **Important Note:** Only a subset of the stores within each retail chain has information recorded by Nielsen about their features and displays. This “causal” sample is maintained by Nielsen and may change over time. There may be other stores within the chain that have the same displays or features, but the information is not reported for those stores. In such cases, feature and display fields have been set to “not observed”. Researchers will need to make their own assumptions about the possibility of features and displays being available in other stores within a given retail chain and market. Nielsen uses a projection methodology to report features and displays to the marketplace. This methodology is not included in this dataset.
- **Products:** One master UPC table serves both the retail scanner and consumer panel data and contains over 3 million UPC codes, of which approximately 2.6 million are utilized in the retail scanner dataset.

The products found in the data include all products in Nielsen-tracked categories that are labeled with a UPC. The Scanner Data does not contain non-UPC products such as random weight meat, fruits, and vegetables (i.e., “Magnet Data” in the Consumer Panel Data). The

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

hierarchical organization of product data is the same as in the consumer panel dataset (10 Departments, ~125 Product Groups, ~1,100 Product Modules).

- **Stores, Retail Chains and Channels:** The retail scanner dataset contains information from approximately 35,000 individual stores from approximately 90 retail chains (these numbers vary by year), representing as of year-end 2011 the following estimated percent of all commodity volume (ACV) in each channel:

<b>% of Sales Covered</b>	
<b>Food</b>	53%
<b>Drug</b>	55%
<b>Mass</b>	32%
<b>Liquor</b>	1%
<b>C-Store</b>	2%

**Table 7** contains ACV estimates by channel for the major Scantrack markets for the Kilts Center's Retail Scanner dataset.

Note that not every Nielsen retail cooperator has agreed to share their scanner data with the Kilts Center. For participating retailers, typically all stores in a retail chain within the 48 contiguous states are included. In rare instances, a retailer may consider a small number of their stores as confidential and exclude them from the dataset.

**Retail Chains:** Retail chains are identified by a parent\_code for each chain, and every store is assigned a parent\_code. Depending on how a retailer supplies data to Nielsen, the parent\_code is either:

- The Corporate Parent (if the company centralizes their data release to Nielsen across all of their retail brands)
- The Retail Banner of the store (if the company decentralizes their data release to Nielsen through multiple data centers representing each of their retail banners)

Where available, a retailer\_code inferred from the Consumer Panel Dataset is also provided, which may or may not be different than the parent\_code for a particular store. The retailer\_code can be at a more granular than the parent\_code (e.g. a retailer may have multiple banners, and the retailer\_code associated with the Consumer Panel Dataset may be at the "retail banner" level rather than the "parent" level). Researchers who are linking the Consumer Panel dataset and the Retail Scanner dataset will find cases where, over time, a retailer\_code in

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

the Consumer Panel data will map to multiple parent\_codes in the Retail Scanner data. This is legitimate, and can occur when a store changes ownership.

In order to protect retailer confidentiality, retail chain names are **not** provided. Per the third-party agreement signed by each researcher, researchers agree not to try to uncover retailer names or identify individual stores. If a researcher somehow determines a retailer's name or identifies an individual store, **the researcher agrees not to disclose retailer names or identify individual stores in any publications, working papers, presentations, or any other forums.** Participating retailers agreed to provide their data to the Kilts Center under this condition of anonymity, so it's critically important for researchers to comply with this request for confidentiality.

**Stores:** Several geographic variables are provided for each individual store, including the first three digits of the zip code, FIPS county code, FIPS county description, Designated Market Area code (DMA), and DMA description. All individual stores are assigned a store code that can be used, where available, to link the retail scanner dataset with the consumer panel data. Note that projection factors for individual stores are not provided. Researchers should not attempt to project data to the total volume of geographic markets or to the total U.S. The data available is not the complete Nielsen sample, so any attempt to project to market totals may be misleading.

Individual stores may change ownership over time if that store or the entire chain is sold to another retailer. So an individual store's parent\_code may change over time.

**Channels:** Retail channels available are food, drug, mass merchandiser, liquor, and convenience (also referred to as C-Store). Channel codes can be found in **Table 8**.

The time series of weekly data for each store should be relatively complete. There may be interruptions in reporting due to retailer POS outages, store construction, or contract lapses between the retailer and Nielsen. Stores may enter or exit the data as they are opened or closed. There may be cases where a parent company acquires or divests one or more of their banners, affecting the availability of data.

## II. Files and File Layouts

### A. Overview of Files

There are three major types of files associated with the Retail Scanner Data: *Stores*, *Products*, and *Movement* (i.e., weekly sales and pricing). The *Stores* file contains information about each individual store location. The *Products* file contains information about each UPC. The *Movement* files contain the

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

price and quantity of goods sold at specific stores on a specific week. Because of the magnitude of the *Movement* data, there is one *Movement* file for each product module code for each year.

While the *Stores* and *Products* files are relatively small files, the *Movement* files can each be tens of gigabytes in size. Since there are thousands of *Movement* files, the overall size of the dataset from 2006-2011 is approximately 3.5 TB, with the expectation it will grow approximately 600 GB each year.

The *Products* file also includes three related files:

- *RMS\_Versions* - The *Products* file tracks different versions of a UPC code over time. A new version of a UPC code is stored whenever one of the core attributes changes for the UPC. The core attributes are Product Module, Brand, Size1, and Multi. However, the weekly *Movement* files do not include the UPC version field (in order to reduce the file sizes and avoid having to reprocess the *Movement* files if we make a change to any UPC versions). So the *rms\_versions* file enables researchers to lookup the correct version of a UPC code to use with the Retail Scanner data for any given year.
- *Product\_Extra (Attributes)*: In addition to the “core” attributes, some products also have “extra” attributes, reported in the *products\_extra* files. Not every product has extra attributes.
- *Brand\_Variations*: In some cases we collapsed several Brand Descriptions into one value when that description had very minor changes. In these cases, all the variations of the Brand Description field have been retained in the *brand\_variations* file.

As part of the reference files you received with the data, we include a copy of the product module hierarchy table (an Excel file) that lists all modules and corresponding groups and departments. If you are working with the consumer panel data, this is the same product hierarchy file you receive with those data.

Note that some modules in the hierarchy file are not present in the data or data appears rarely for these modules. There are two reasons for this, and the hierarchy file indicates the modules to which these exceptions apply.

- 1) Over time, Nielsen stops coding for some module codes. These are called deferred modules. The historical panel data may include purchases made in these modules. In some cases, there may be movement files for deferred modules. **See column G in the hierarchy file for this indicator. Data for ALL deferred modules should be used with caution because the UPCs may not be fully coded, or the UPC information may be outdated.**
- 2) There are a small number of modules for which we did not receive scanner data one or more years. See column H in the hierarchy file for these notes.

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT



## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

The product hierarchy is as follows:

- Department (10 Departments)
  - Product Group (~125 Product Groups)
    - Product Module (~1,075 Product Modules)
- UPC (~3.2 million UPC Codes)\*

\*Note that ~3.2 million UPC codes are contained in the combined consumer panel and retail scanner data files. Only ~1.5 million UPC codes are present in the consumer panel files only.

## B. File Layouts

**Table 1: Stores Files**

Number of Files: One file for each year. Each record contains information for one store.

Variable Name	Variable Type	Variable Description	Notes
store_code_uc (key)	Integer	A code used to identify the individual store.	A
year (key)	Integer	The year of validity for this observation.	B
retailer_code	Integer	A code used to identify the store's retailer (this matches the retailer_code in the Consumer Panel data).	C
parent_code	Integer	A code used to identify the store's parent company.	D
channel_code	Char (1)	A code used to identify the channel type of the retailer.	
store_zip3	Char (3)	The first 3 digits of the zip code where the store is located.	E
fips_state_code	Integer	The 2 digit state FIPS code where the store is located.	
fips_state_descr	Char (2)	The name of the state.	
fips_county_code	Integer	The 3 digit county FIPS code where the store is located.	
fips_county_descr	Char (50)	The name of the county.	

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

<b>dma_code</b>	Char (5)	The 3-digit Designated Market Area code	F
<b>dma_descr</b>	Char (100)	The name of the Designated Market Area	

A) store\_code\_uc can be used to link the retail Scanner Data to the Consumer Panel Data via the Consumer Trips file. The store\_code\_uc references a physical location of a store.

B) Many store\_code\_uc-year combinations will be identical. Since the store\_code\_uc references a physical location, the owner (retailer\_code and/or parent\_code) can change over time, which may also affect the channel\_code. Further, due to changes in the assignment of Zip Codes and perhaps for other reasons, the store\_zip3 field may change, although this is unlikely since we only report the first 3 digits of the ZIP code. Except for changes to county line boundaries, the FIPS codes should not change.

C) The retailer\_code is a numeric code that identifies a store as being part of a retail chain. The retailer name is not disclosed by Nielsen in order to protect retailer confidentiality. This retailer\_code field is the same as the retailer\_code field in the Consumer Panel dataset, so these datasets can be matched using this field. If a particular store does not have any purchase transactions in the Consumer Panel dataset, then the retailer\_code field will be blank. To enable easier merging of the Consumer Panel and Retail Scanner datasets for researchers, we provide both the parent\_code and retailer\_code for a given store (when available).

D) Some retailers have multiple “banners” associated with their stores (i.e. the banner names displayed on the front of the store differ, even though the parent company is the same). In these cases, the “retailer\_code” field relates to a specific store brand, while the parent\_code field may be different. Depending on how a retailer supplies data to Nielsen, the parent\_code is either:

- The Corporate Parent (if the company centralizes their data release to Nielsen across all of their retail banners)
- The Retail Banner of the store (if the company decentralizes their data release to Nielsen through multiple data centers representing each of their retail banners)

E) Only the first 3 digits of the store’s zip code are made available to keep the store’s identities confidential.

F) See Appendix V for maps depicting DMA codes.

### Missing stores in 2012:

- In the 2012 data, there are 38 stores with incomplete geographical information, i.e. the variables store\_zip3, fips\_state\_code, fips\_state\_descr, fips\_county\_code, fips\_county\_descr, dma\_code, and dma\_descr are missing. These stores are known to be missing in 2012 only. The reason is

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

that we received from Nielsen the 2012 movement data for these stores, but we did not receive additional store information for that year. It was an anomaly related to the timing of when Nielsen sent us various files.

- We recommend users check the other years to see if any of these fields are available for these stores (e.g. 3-digit zip code). If those geographic fields are available from prior years, they clearly wouldn't have changed in 2012 (a specific store is unlikely to change its 3-digit zip code from 1 year to another).

### Table 2: Products File

**Number of Files:** One master file for all years, updated annually. Each record contains information for one unique version of a UPC code. *This is a shared file with the Consumer Panel dataset, so if you are using both datasets, the Products file delivered at any particular year should be identical across the two datasets.*

**Note:** We create a new version of the UPC for each unique combination of values for the 4 core attributes. We consider the following 4 UPC attributes as "core":

- product\_module\_code
- brand\_code\_uc
- multi
- size1\_code\_uc

A change in the value of any of the 4 core attributes results in a new UPC version. For example, if the size1\_code\_uc of a UPC changes temporarily to reflect special promoted product size and then reverts to its original value, we will create a new UPC version for the first change to a new size1\_code\_uc but not for the second change to an already existing UPC version.

For all scanner-only UPCs, we receive one set of product characteristics each year – if any of the product characteristics changed during the year, we only have the characteristics as of the date Nielsen sent us the data.

Variable Name	Variable Type	Variable Description	Notes
upc (key)	Char (12)	UPC code of a product.	A
upc_ver_uc (key)	Integer	A code used with upc to determine a unique product.	B
upc_descr	Char (50)	Description of the product.	
product_module_code	Integer	Code classifying the product into a module.	
product_module_descr	Char (100)	Description of the product module.	
product_group_code	Integer	Code classifying the product into a group.	

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

<b>product_group_descr</b>	Char (100)	Description of the product group.	
<b>department_code</b>	Integer	Code classifying the product into a department.	
<b>department_descr</b>	Char (100)	Description of the department.	C
<b>brand_code_uc</b>	Integer	Code classifying the product as a specific brand.	
<b>brand_descr</b>	Char (30)	Description of the brand.	G & See brand_variations file information.
<b>multi</b>	Integer	How many of those goods appear in a given pack (e.g. a 6 pack of soda).	
<b>size1_code_uc</b>	Integer	Code used with product module code to distinguish various sizes.	
<b>size1_amount</b>	Decimal (3)	Numeric quantity of the good in individual packaging.	
<b>size1_units</b>	Char (6)	Unit of measure for that quantity.	D
<b>datasets_found_uc</b>	Char (3)	Indicates whether this UPC appears in the Retail Scanner and Consumer Panel datasets, The Retail Scanner dataset only, or the Consumer Panel dataset only.	E
<b>size1_change_flag_uc</b>	Integer	Indicates whether the size may have changed during the 2006-2010 period in the Retail Scanner data.	F
<p>A) UPCs associated with private label goods are altered by Nielsen to protect the identity of the retail banner and proprietary sales data associated with the product and the store. Multiple Private Label UPCs across retailers are mapped to a single “masked” UPC. In these cases, the assumption is that the core attributes are identical for all of these UPCs. Prices may vary, and there may be some custom characteristics that differ based on how the different products are marketed. Nielsen codes all of the private label brands as “Control Brand” – specifically, the brand description is “CTL BR” (or a similar variation).</p> <p>B) UPCs are sometimes recycled by companies for different products, or a similar product with different packaging (e.g., a different sized bottle of soda or changed multipack value). UPC versions are generated based on 5 attributes: upc, product_module_code, brand, size1, and multipack. If any of these attributes change, a new UPC version is generated. If a UPC’s core attributes revert to a prior version of that UPC, the prior version is reassigned.</p> <p>C) See the Department codes and descriptions below.</p> <p>D) See the Size1_units descriptions below.</p>			

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

- E) This flag indicates in which datasets the UPC can be found: Retail Scanner, Consumer Panel, or both.
- F) When Nielsen provided the 2006-2010 Retail Scanner data, we only received the 2010 Size1 values for each UPC. However, Nielsen did provide information so that we could determine whether the Size1 field changed in any of the years 2006-2010, but not enough information to reliably state the historical Size1 values. Where possible, we have imported historical data from the Consumer Panel dataset and created new UPC versions when the Size1 field changed. When we did not have corresponding data from the Consumer Panel dataset, we have simply flagged that the Size1 field may have changed during the 2006-2010 period.

***Department Codes and Descriptions***

Code/ Value	Description	Approximate Number of Product Groups
0	Health and Beauty Aids	<b>21</b> (e.g. baby care, cosmetics, cough & cold remedies, deodorant, hair care, oral hygiene, pain remedies, skin care, shaving)
1	Dry Grocery	<b>41</b> (e.g. baby food, baking mixes, bottled water, candy, carbonated beverages, cereal, coffee, condiments, crackers, pet food, prepared foods, snacks, soup, canned vegetables)
2	Frozen Foods	<b>12</b> (e.g. ice cream, frozen pizza, frozen vegetables)
3	Dairy	<b>12</b> (e.g. cheese, eggs, yogurt)
4	Deli	<b>1</b>
5	Packaged Meat	<b>1</b>
6	Fresh Produce	<b>1</b>
7	Non-Food Grocery	<b>12</b> (e.g. detergent, diapers, fresheners/deodorizers, household cleaners, laundry supplies, pet care)
8	Alcohol	<b>4</b> (e.g. beer, wine, liquor, coolers)
9	General Merchandise	<b>19</b> (e.g. batteries/flashlights, candles, computer/electronic, cookware, film/cameras, insecticides, lawn & garden, motor vehicle, office supplies)
99	Magnet Data Products	<b>1</b> Magnet Data is not reported in the Retail Scanner Dataset. This department_code is assigned by University of Chicago and should not be used with the Retail Scanner data. See the Consumer Panel Data for specifics on handling magnet data.

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

**Size1\_Units Codes and Descriptions**

The size1\_code\_uc field is the code for the size of a product, usually weight.

Size1\_units descriptions are as follows:

size1_units Code	size1_units Description
CF	Cubic Foot
CT	Count
LI	Liter
FT	Foot
YD	Yard
EXP	Expired
PO	Pound
OZ	Ounce
QT	Quart
OZ	Ounce
SQ FT	Square Feet

**Table 3: Movement Files**

Number of Files: One file for each product module for each year. Each record contains information on the average weighted price and unit sales volume for 1 UPC code from 1 store during 1 specific week.

Variable Name	Variable Type	Variable Description	Notes
store_code_uc (key)	Integer	A code used to identify the individual store.	
upc (key)	Char (12)	UPC code of product.	
week_end (key)	Date (YYYYMMDD)	The Saturday ending the week for which the data is collected.	A, B, C
units	Integer	The number of units sold.	D
prmult	Integer	Price multiplier.	E
price	Decimal (2)	The volume weighted average price of the product for that week.	D, F, G, H, I, N
feature	Integer	Whether the product was a featured item that week.	J, K, M
display	Integer	Whether the product was on display that week.	J, L, M
A) The data are not available at a more granular (e.g. daily) level.			

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

- B) The “week-ending day” associated with the weekly price/volume movement records is reported as the same for each retailer – always showing the week ending day being a Saturday. In reality, retailers may use different “7-day” periods when reporting their weighted average prices and units sold. All data is for a 7-day period, but it may or may not be for a 7-day period that begins on Sunday and ends on Saturday. The 7-day period could start/end on other days of the week. This is because a retailer might begin their promotions on a certain day each week (e.g. Thursdays), and then want their weekly Scanner data reported by Nielsen to be in alignment with their promotional calendar (e.g. Thursdays through Wednesdays).

**Additional Details from Nielsen on this topic:**

*Nielsen uses a Saturday week-ending label to identify weekly data being reported for scanning store data as well as panel data. For scanning data, not all retailers provide weekly data using a Sunday to Saturday definition. Some retailers provide data based on their promotion week, which varies by retailer. Nielsen maps non-Saturday ending weeks received from retailers to the best fit Saturday. Panel data is generally a Sunday to Saturday aggregation for all panelists. Researchers who are interested in aligning panel data to scan data for the same individual week should be aware of the potential for differences in days in such comparisons. In most cases for panel data, panelists do not enter prices for retailers for which Nielsen receives scanner data. Prices are ascribed from the scanner data for the specific store a panelist shopped in the reported week. As such, prices should generally be consistent for scanning data and panel data and represents the weighted average price for the reported week. In some instances, prices may differ between scanning data and panel data. This can occur if the retailer is an exception where Nielsen does not ascribe scanning prices to panel data (panelists enter prices) or situations where panelist data transmissions are received late. In these situations a researcher may want to use the scanning prices as this will provide consistency across all items in the store for the week (including items not purchased by any panelists).*

- C) Scanning-based prices provided by select cooperating retailers are also used for Nielsen panel data. Nielsen panelists are not asked to enter price if they shop in these stores. Instead, the price is inferred from that store’s Retail Scanner data supplied to Nielsen. The inferred price is based on the specific store the panelist shopped in for the same week-ending date as the scanner data. In some cases where a panelist transmitted their data after the week-ending, the imputed price is based on the following week’s scanning data.
- D) Volume and individual prices can be imputed under certain circumstances (e.g. incomplete/missing data and outliers). Nielsen performs a battery of usability edits each week on all scanning stores received. Occurrences of data modifications or imputation are relatively rare. Volume and individual prices can be imputed under certain circumstances (e.g. incomplete/missing data and/or outliers). In less than ~1% of the cases, adjustments are made to some of a store’s weekly price/volume data sent. In less than ~3% of the cases, all of a store’s

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

weekly price/volume data is imputed.

In less than ~1% of the cases, prices are changed. Approximately 25% of price changes occur because a price outlier is found and therefore price is imputed using historical expectations from the store or if a new item, from other similar stores. Approximately 75% of the price changes occur when Nielsen aligns price to be on the same basis as their other retailers (e.g. for soft drinks, some retailers might send “price per can” and others send “price per 6-pack”, so Nielsen changes the prices to all be on the same basis). The raw data after cleansing is what Nielsen and its clients use for analyses. Nielsen does not supply an indicator for which prices/volumes were modified or imputed but uses this information to update edit rules on a regular basis.

- E) Used to indicate deals such as 3 for \$1. The unit price would be  $\text{price}/\text{prmult}$ . The week's total dollar sales would be  $(\text{price}/\text{prmult}) \times \text{units}$ . The total volume would be  $\text{units} \times \text{size1\_amount} \times \text{multi}$  in  $\text{size1\_units}$ .
- F) Retailer discounts and manufacturer coupon usage are not recorded in the Retail Scanner data. Retailer discounts and specials (e.g. discounts associated with the use of a retailer's loyalty card or coupons) are factored into the weighted average weekly price. Manufacturer coupons are not factored into the weighted average weekly price.
- G) Not all goods will be sold at a uniform price throughout the week. Nielsen uses the volume weighted price of all sales for that good in that week to determine the price.
- H) Local sales tax will typically not be included in the reported price. However, some excise taxes, e.g. federal cigarette taxes, are included in the price of the good.
- I) If a store does not report any sales for a UPC in a given week, no movement record is available for that UPC and therefore no price is reported by Nielsen. It is up to researchers to decide how to treat product-store-weeks with no reported prices.
- J) Feature and Display indicator observations: Nielsen only observes Features and Displays in a subset of the stores, and those stores can vary on a weekly basis. If one store for a retailer has a feature in a given week, it is reasonable to assume that all other stores for that retailer in the same DMA have that same item featured. However, researchers should not assume that an item on Display would necessarily also be displayed at all stores in that retail chain within the same DMA as display execution can vary by store within chain/market.

Note that only stores having **both** Features and Displays for a given week represent the subset of stores that were directly audited by Nielsen. The retail scanner dataset does contain stores that have **either** Features **or** Displays in a given week, but such stores were not directly audited by Nielsen. Therefore, when researchers are using causal information (Feature & Display

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT



## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

settings), they should use stores that have **both** Features and Displays for a given week.

**Starting in 2013**, we blank out the Feature and Display fields for all movement records in stores that didn't have at least one UPC with Feature=Y and at least one UPC with Display=Y in any given week.

- K) **Feature** – All retailer advertisements found in local newspapers, free standing inserts (FSIs), and free standing circulars, and may also include online ads from the retailer's website. The vast majority of featured items will include a price discount, but they don't have to. Features include Major Ads (which typically include an image as well as the price of an item), Line Ads (only has the name and price of the item), and retailer coupons that can be redeemed at the register.
- L) **Display** – a secondary location of an item in the store that is non-permanent and intended for merchandising purposes. Displays are located in the Store Lobby, Front of Store, End of Aisle, In Aisle, or Back of Store. Displayed items may or may not have an associated price decrease.
- M) **Temporary Price Reduction (TPR)** – there is no indicator in the Retail Scanner dataset for TPRs, but researchers can make their own assumptions. Looking at the historical average price for an item, the industry suggests that any price at least 5% lower than the average non-promoted price could be considered a TPR.
- N) **1 cent prices** - Nielsen does not allow prices to be zero. If a good is free, then it is coded/converted to 1 cent. This can happen with a conditional promotion, e.g. buy an item and get another good free. There are also other "low prices" found in the RMS data because the price in RMS is a weighted average price ... some customers may have received the product for free as part of a promotion, and others paid full price for the item.

### "Almost Duplicate" Movement Records:

- In 2006-2011, there are approximately 300,000 "almost duplicate" movement records. In these cases, the movement records contain the same store, week and UPC code, but one or more of the other fields differs. These "almost duplicate" records occur when a store changes parent companies, but the parent companies have different "week ending" dates for reporting their scanner data. Nielsen states that both records are valid. In such cases, researchers must decide how to handle these records. Going forward, Nielsen has committed to eliminating this issue when they provide future years of retail scanner data.

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

### Missing movement data for Wisconsin 2006:

- Researchers should be aware that we identified a known issue for 2006 scanner data related to sales in the state of Wisconsin only. Wisconsin RMS sales data go up 175% from 2006 to 2007. The movement records are unavailable. Refer to the spreadsheet “Wisconsin2006data.csv” located in the reference documentation folder to see increase in sales for each module 2006-2007 for Wisconsin.

### Table 4: RMS Versions File

Number of Files: One file for each year. Each record contains information for one UPC/UPC\_version for each year.

Variable Name	Variable Type	Variable Description	Notes
upc (key)	Char (12)	UPC code	
upc_ver_uc (key)	Integer	UPC version	
year (key)	Year (YYYY)	A year UPC and version is found in RMS data	

The Retail Scanner data movement files do not contain the upc\_ver\_uc variable. Retail Scanner data for the UPC characteristics are provided by Nielsen on an annual basis, thus within year changes in UPC versions are undetectable. Further, a single shipment was made for the years 2006-2010, except in some cases for the Size1 field (see details about Size1 under size1\_change\_flag\_uc in products table). To merge the UPC version with the Retail Scanner movement data, the year of the week\_end date should equal the year for a given UPC in the versions file. The resulting file can then be merged to the Products file using the upc and upc\_ver\_uc fields.

### Table 5: Product Extra Attributes Files

Number of Files: One file per panel year. Each record contains information for one unique version of a UPC code. *This is a shared file with the Consumer Panel dataset, so if you are using both datasets, the Product\_Extra\_Attributes files delivered for any particular year should be identical across the two datasets.*

The UPC codes for some products have additional attributes. There are 19 total extra attributes (each with a “code” and “description” field), however, many products do not have any value listed in these fields.

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

In the extra attributes, Nielsen uses two different values to indicate missing data: “0” and blank (“”). We have preserved these original values, but both have the same meaning.

In this file, the description associated with the integer code values for each attribute can vary by product\_module. In order to determine the description of a particular code for an attribute, not only the code itself is needed but also the product\_module. Additionally, the descriptions themselves can vary slightly by year. Ultimately, the product\_module, the panel\_year and the attribute code value are all needed to determine a code’s meaning.

**Note to researchers regarding flavor data for 2010:** It is a known issue that the flavor code and flavor description are missing from the 2010 extra attributes file. We have a file available with these missing data. Refer to the spreadsheet “Latest\_Flavor\_2010.csv” located in the reference documentation folder. This is an interim solution, as we will eventually rework the files to include the data.

Variable name	Variable Type	Variable Description	Notes
Upc	Char (12)	UPC code of product	
upc_ver_uc	Integer	UPC code version	
panel_year	Integer (4)	Year associated with this file	
flavor_code	Integer	Flavor code	
flavor_descr	Char (50)	Flavor description	
form_code	Integer	Form code	
form_descr	Char (50)	Form description	shape of product (e.g. Chocolate in form of solid bunny shape; sugar granulated or powder; nuts slivered or chopped, etc.
formula_code	Integer	Form code	
formula_descr	Char (50)	Formula description	e.g. ½ Less Fat, 0-6 Month, 110 Calorie, Acne Fighting, Cold & Cough, Day or Night, Extra Moist, etc.
container_code	Integer	Container code	

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

<b>Container_descr</b>	Char (50)	Container description	e.g. plastic, glass, or canned drinks
<b>salt_content_code</b>	Integer	Salt content code	
<b>salt_content_descr</b>	Char (50)	Salt content description	e.g. low salt, sea salt, no salt, etc.
<b>style_code</b>	Integer	Style code	
<b>style_descr</b>	Char (50)	Style description	e.g. shelf-stable product; kosher, dill, sour pickles; salted nuts; breakfast blend coffee, etc.
<b>type_code</b>	Integer	Type code	
<b>type_descr</b>	Char (50)	Type description	e.g. decaffeinated; sugar-free; extra- virgin olive oil; soft- drink; etc.
<b>product_code</b>	Integer	Product code	
<b>product_descr</b>	Char (50)	Product description	e.g. Gravy, Mint Truffle Square, Sauce, etc.
<b>variety_code</b>	Integer	Variety code	
<b>variety_descr</b>	Char (50)	Variety description	e.g. Angel Hair, Beef, Blarney, etc.
<b>organic_claim_code</b>	Integer	Organic claim code	
<b>organic_claim_descr</b>	Char (50)	Organic status claimed by product.	
<b>usda_organic_seal_code</b>	Integer	USDA organic seal code	
<b>usda_organic_seal_descr</b>	Char (50)	USDA Organic Seal on product.	
<b>common_consumer_name_code</b>	Integer	Commonly used name code	
<b>common_consumer_name_descr</b>	Char (50)	CCN description	Perch, shovel, toy, etc.
<b>strength_code</b>	Integer	Strength code	
<b>strength_descr</b>	Char (50)	Strength description	
<b>scent_code</b>	Integer	Scent code	
<b>scent_descr</b>	Char (50)	Scent description	
<b>dosage_code</b>	Integer	Dosage code	
<b>dosage_descr</b>	Char (50)	Dosage description	

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

<b>gender_code</b>	Integer	Gender code	
<b>gender_descr</b>	Char (50)	Gender description	
<b>target_skin_condition_code</b>	Integer	Skin condition targeted code	
<b>target_skin_condition_descr</b>	Char (50)	Skin condition targeted description	e.g. dry skin, cracked skin, etc.
<b>use_code</b>	Integer	Intended to be used by code	
<b>use_descr</b>	Char (50)	Used by description	Men, Teen, Regular, Child, etc.
<b>size2_code</b>	Integer	Size of product code	A different size measurement (compared to size1_code) of a product
<b>size2_amount</b>	Decimal (3)	Amount in product size	A different amount measurement (compared to size1_amount) of a product

**Table 6: Brand Variations File**

Number of Files: One master file for all years, updated annually. Each record contains information for one alternative brand descriptions associated with brand\_code\_uc. *This is a shared file with the Consumer Panel dataset, so if you are using both datasets, the Brand Variations file delivered at any particular year should be identical across the two datasets.*

This file retains minor variations in a brand\_description, where punctuation and other small variations may exist. These variations were not deemed significant enough to warrant creating a new brand\_code\_uc which would necessitate creating a new version for a UPC code. We provide this file so researchers can see other brand names that may have been used for a given brand\_code\_uc.

Variable name	Variable Type	Variable Description	Notes
<b>brand_code_uc</b>	Integer	Code classifying the product as a specific brand.	A
<b>brand_descr</b>	Char (30)	Primary Brand	

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

		Description
<b>brand_descr_alternative</b>	Char (30)	Alternative Brand Description
<b>start_date</b>	Date (YYYY-MM-DD)	Start date when alternative brand description can be found.
<b>end_date</b>	Date (YYYY-MM-DD)	End date when alternative brand description can be found.
<b>datasets_found_uc</b>	Char (3)	In which datasets the alternative brand description can be found.

A) The Kilts Center merged “similar” brand descriptions into a single brand description in order to simplify the dataset. This was done at the UPC level. In these cases, the full-text “alternative” brand descriptions have been retained in this table.

## C. Other Tables

**Table 7: Kilts Center Store Coverage - All Commodity Volume (ACV) by Channel and Scantrack Market**

	ACV by Channel		
Scantrack Markets	Food	Drug	Mass
ALBANY	43%	51%	30%
ATLANTA	54%	51%	29%
BALTIMORE	68%	56%	35%
BIRMINGHAM	26%	37%	22%
BOSTON	83%	39%	35%
BUFFALO-ROCHESTER	11%	71%	38%
CHARLOTTE	86%	50%	28%
CHICAGO	65%	77%	47%
CINCINNATI	64%	55%	34%
CLEVELAND	52%	36%	36%
COLUMBUS	67%	35%	32%
DALLAS	68%	50%	25%

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

DENVER	86%	92%	33%
DES MOINES	58%	71%	35%
DETROIT	36%	46%	50%
GRAND RAPIDS	6%	75%	37%
GREENVILLE	77%	40%	25%
HARTFORD-NEW HAVEN	62%	41%	31%
HOUSTON	50%	63%	26%
INDIANAPOLIS	38%	32%	27%
JACKSONVILLE	47%	51%	30%
KANSAS CITY	30%	44%	27%
LAS VEGAS	76%	58%	33%
LITTLE ROCK	51%	57%	16%
LOS ANGELES	52%	45%	48%
LOUISVILLE	66%	69%	25%
MEMPHIS	55%	67%	22%
MIAMI	26%	54%	37%
MILWAUKEE	72%	77%	42%
MINNEAPOLIS	41%	62%	54%
NASHVILLE	60%	66%	25%
NEW ORLEANS-MOBILE	39%	60%	22%
NEW YORK	42%	56%	51%
OKLAHOMA CITY-TULSA	1%	62%	20%
OMAHA	50%	71%	38%
ORLANDO	33%	63%	28%
PHILADELPHIA	49%	51%	39%
PHOENIX	84%	73%	30%
PITTSBURGH	60%	59%	32%
PORTLAND	67%	88%	30%
RALEIGH-DURHAM	77%	55%	27%
RICHMOND	81%	60%	33%
SACRAMENTO	51%	55%	34%
SALT LAKE CITY-BOISE	45%	91%	26%
SAN ANTONIO	6%	66%	30%
SAN DIEGO	69%	34%	37%
SAN FRANCISCO	48%	55%	49%
SEATTLE	71%	80%	38%
ST. LOUIS	53%	82%	31%

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

*Kilts Nielsen Retail Scanner Dataset Manual*

*Updated –1.22.2016*

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

SYRACUSE	15%	79%	28%
TAMPA	29%	58%	28%
WASHINGTON, D.C.	73%	28%	39%
WEST TEXAS	22%	74%	27%
Grand Total	53%	55%	32%

See Appendix [Y](#) for a map depicting ScanTrack areas.

**Table 8: Channel Codes**

Channel Code	Channel Description
<b>C</b>	Convenience
<b>D</b>	Drug
<b>F</b>	Food
<b>M</b>	Mass Merchandiser
<b>L</b>	Liquor

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**



## D. Data Notes

### Retailer Level Analysis Between the Retail Scanner Dataset and Consumer Panel Dataset

While the Stores files take care of matching a given store location in the Retail Scanner Data and the Consumer Panel Data, doing retailer level analyses proves more complicated. The complication can be exemplified in the below table. This table is generated by taking the stores files and counting the number of store\_code\_uc's associated with each unique combination of parent\_code and retailer\_code.

parent_code	retailer_code	store count
9	9	2242
9	199	82
9	801	1
9	832	2
9	873	119
9	3997	17
9		1833

While we cannot be sure, it appears that parent\_code operates 1 major banner (retailer code 9) and two minor banners (retailer codes 199 and 873). We see this by looking at the store counts. It is possible that retailer\_code 3997 is also a minor banner. However, we can be fairly certain that retailer\_codes 801 and 832 are not part of this parent\_codes banner but were assigned due to unavoidable timing alignment issues when merging the Retail Scanner and Consumer Panel datasets. Nevertheless, 42% of the store-years have no retailer code assigned to them because there is not a shopping trip associated with them in the Consumer Panel Data.

There are several strategies for rolling this data up to the parent/retailer level. Which strategy is most appropriate depends on what the researcher is trying to accomplish. It is possible that maintaining the separate banners found in the Consumer Panel Dataset is more useful. Depending on how the banners are utilized by the parent, it may be possible to recover some of the original retailer\_codes by looking at the geographic distribution of the various retailer\_codes for a given parent. If there is a clear separation, then the retailer\_code may be inferred. However, if the various banners are evenly distributed or at least highly intermixed (as is often the case in 'Super' banners versus 'Regular' banners), then it may not be possible to recover any of the retailer\_codes for the 1833 store-years.

The main lesson here is that retailer/parent level analysis may not be as straightforward as it might initially seem. Researchers should have an explicit strategy for handling the various difficulties that are present when handling the data at this level.

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

### **III. Appendix I: Policies**

Researchers who have received the data, including all research assistants and students under their supervision, must comply with several policies and legal requirements. Violating any of these conditions will result in immediate termination of data access for the researchers and all other project investigators affiliated with their institution.

**Researchers must agree that their proposal adheres to the following Nielsen criteria:**

- Unless otherwise approved by Nielsen and the Kilts Center, I will not conduct research commissioned, sponsored, or funded by an industry source, government organization, industry group, nonprofit organization, or private company.
- I will not use the data for consulting purposes of any kind, including litigation support.
- Evaluating Nielsen's data collection approach or methodologies is not a primary purpose of this project.

**Researchers must also agree to the following Nielsen legal requirements:**

- I will not share the data with any other individuals, other than registered PhD students who are individually registered and approved by the Kilts Center.
- I will not share the data with any student in an undergraduate or master's program without notifying the Kilts Center of plans to direct and accept full responsibility for the student's access in a classroom or independent study setting.
- I will ensure that data is stored only on devices and computers owned by the subscribed institution. I understand that insubstantial subsets of data required for particular analysis processes may be temporarily downloaded to such personal computers as required for analysis, but I will ensure that these data are purged as soon as possible.
- I will not disclose any personally identifiable information about any consumer panelists, or details that would permit re-identification of Nielsen's consumer panelists, without prior Nielsen approval.
- I will not disclose any information about specific retailers or individual stores, or details that would permit re-identification of Nielsen's retailers or individual stores, without prior Nielsen approval.
- I will not disclose UPC-level pricing data that may be incorporated in the Nielsen data received from retailers who have agreed to provide this data to Nielsen.
- I will include Nielsen's copyright in any disclosed information and publications.
- Within fifteen (15) business days of expiration or termination of the agreement, I will destroy or purge all data (including any copies) in my possession or control, and provide written certification of the destruction and purging to the Kilts Center.

**Finally, researchers must agree to submit working papers and final publications resulting from their use of Nielsen data:**

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

### **DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

- I will submit an annual update to the Kilts Center that includes copies of all working papers arising from use of the Nielsen data. Continued use is contingent upon submitting an annual status report and adhering to the terms and conditions of the contract.
- For final publications, I will submit these papers to Nielsen and the Kilts Center at the time the publication is accepted and at least 30 days prior to public dissemination.
- I will not publish more than limited excerpts of Nielsen information in a bona fide publication of academic research - a "limited excerpt" means such portion of the Nielsen data that would not ordinarily substitute for a purchase of access to the Nielsen data.
- In all working papers or publications of any kind that use the Nielsen data, I will acknowledge use of the data by citation of the Data Center at The University of Chicago Booth School of Business and noting that information on availability and access to the data is available at <http://research.chicagobooth.edu/nielsen>. I will also include attribution to Nielsen as follows:

*"Calculated (or Derived) based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business."*

- I will submit all working papers produced using the Kilts-Nielsen database to the [Kilts Center for Marketing Nielsen Data Research Working Paper Series at Social Science Research Network \(SSRN\)](#).

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](#), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

### **DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

## IV. Appendix II: File Names and Sizes

1. <b>Stores:</b> 1 file for each year. One new file will be created for each additional year as data is released.	
Naming convention:	Stores_YYYY.tsv
File Sizes:	~2.5 MB or less each year
2. <b>Products:</b> 1 master file (combining all panel years). An updated version of this file will be regenerated each year with annual updates.	
Naming convention:	products.tsv
File Size:	~550 MB
3. <b>Movement:</b> One file for each product module for each year, where <module_code> is the 4-digit Product Module Code. There are over 1,100 movement files in each year.	
Naming convention:	<module_code>_YYYY.tsv
File sizes:	Varies from 1MB to 10s of GBs per module code per year
4. <b>RMS Versions:</b> 1 file for each year. One new file will be created for each additional year as data is released.	
Naming convention:	rms_versions_yyyy.tsv
File sizes:	~20 MB or less each year
5. <b>Product Extra Attributes Files:</b> 1 file for each panel year. One new file will be created for each additional panel year as data is released.	
Naming convention:	products_extra_XXXX.tsv for XXXX=20XX
Files sizes:	Ranges from ~70-120 MB each year
6. <b>Brand Variations:</b> 1 master file (combining all years). An updated version of this file will be regenerated each year with annual updates.	
Naming convention:	brand_variations.tsv
File Size:	Less than 50 KB

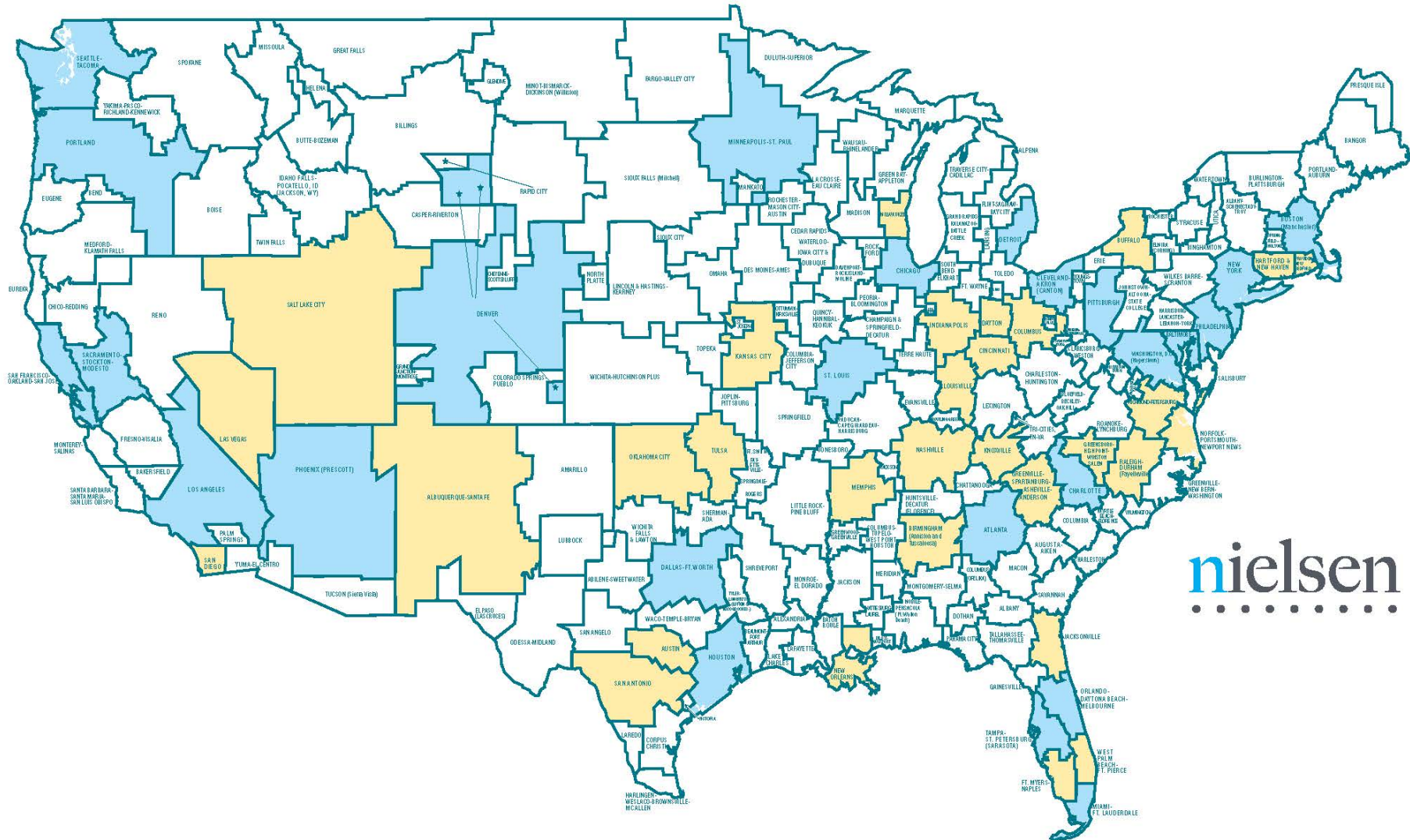
Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## **V. Appendix III: Maps**

Page intentionally left blank.

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

# Nielsen DMA—Designated Market Area



nielsen

**DMA® region:** Generally a group of counties in which the commercial TV stations in the Metro/Central area achieve the largest audience share. This is non-overlapping geography for planning, buying and evaluating television audiences across various markets. See the Local Reference Supplement for more information.

County time zone designations per Department of Transportation except in the case of split counties where the NSI® service assigns the entire county to one time zone on majority of population basis.

★ Counties assigned to a DMA region but not contiguous to other counties of that market.

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

**DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT**

*Kilts Nielsen Retail Scanner Dataset Manual*

*Updated –1.22.2016*







## VI. Appendix IV: Version revisions

### A. Revisions 2016-1-22

1. Known issue regarding 38 stores in 2012. See the note on page [10](#) for note regarding this issue.
2. Known issue regarding extra attribute “flavor” for year 2010. Data previously missing is now available in separate file. See note in section [Table 5: Product Extra Attributes Files](#) for details.
3. Pricing note added regarding products with a price of one cent (or similar amount). See note N in [Table 3: Movement Files](#) on page [14](#)
4. Known issue regarding missing movement data for the state of Wisconsin in 2006. See note on page [18](#) for more information.
5. All revisions to the hierarchy file (included in reference documentation files) are highlighted in that document. Note the following changes to modules.
  - a. Three modules added. Only module 7807 has data in 2014. Nielsen has just begun coding products with these data. Future years are expected to include data for 7808 and 7809.

7807	SEXUAL ENHANCING DEVICE	6012	MEDICATIONS/REMEDIES /HEALTH AIDS	0	HEALTH & BEAUTY CARE
7808	GRASS SEED	5508	FLORAL, GARDENING	9	GENERAL MERCHANDISE
7809	POOL AND SPA CHEMICALS AND TREATMENT	4504	HOUSEHOLD CLEANERS	7	NON-FOOD GROCERY

- a. Module 8621 – Cellular Phone – changed from product group 5507 (ELECTRONICS, RECORDS, TAPES) to group 5516 (LIGHT BULBS, ELECTRIC GOODS). **Data for years 2006-2013, this module will be categorized in product group 5507. Beginning in 2014, data for this module will be categorized in product group 5516.**
- b. Added column J to indicate modules that may not contain data ANY years in any dataset.

### B. Revisions 2015-05-25

1. Starting in 2013, we blank out the Feature and Display fields for all movement records in stores that didn’t have at least one UPC with Feature=Y **and** at least one UPC with Display=Y in any given week. Researchers using these fields for prior years should only use the Feature and

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](https://kiltsforum.chicagobooth.edu/forum/), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

## DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Display causal indicators for stores that have at least 1 UPC features and 1 UPC on display in any given week.

2. The “datasets\_found\_uc” flag in the Products file has been updated. There were cases where this flag was previously inaccurate, indicating that a UPC should be found in the Retail Scanner dataset when that was not the case.

### C. Revisions 2014-09-22

1. Updated footnote about control brands. See footnote A in section [Table 2: Products File](#), page [4](#).
2. Note regarding article submissions and data. See [Technical notes regarding data](#), page [4](#).
3. Added information about deferred modules. See description of deferred modules in [Overview of Files](#) and refer to page [8](#).
4. Updates regarding stores indicated with features and displays. See second paragraph of footnote J for the Movement Files layout on page [16](#).

### D. Revisions 2014-04-02

1. Addition of section [Technical notes regarding data](#). See page [4](#).

### E. Revisions 2014-03-05

#### **Products file update**

1. We have updated the brand description for 10 UPC codes. See list below.

007410813801	007410815295	007410825589
007410815166	007410820634	007410826631
007410815197	007410822685	
007410815211	007410823743	

In addition, brand\_id 33011 has been changed to 33012. This change affected one UPC 1963433 whose brand\_id changed to 33012.

#### **Maps**

2. Addition of maps for Nielsen Designated Marketing Areas (DMA) and Scantrack markets.

Another resource for approved Nielsen data users to refer to is the [Kilts data online users forum](#), an online discussion site for approved researchers to ask and answer questions about the data, post helpful information, and share ideas. <https://kiltsforum.chicagobooth.edu/forum/>

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT