DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

CHICAGO BOOTH | Kilts Center for Marketing

NielsenIQ

Retail Scanner Dataset Manual 2006 - 2024

The Kilts Center Archive of The NielsenIQ Company

1/26/2026

Contents

I. Important notes about this documentation 4
II. Introduction 4
A. Special notes: Batch 2 (2021 onward) 4
B. Purpose (Universal) 5
C. Technical notes regarding data (Universal) 6
D. Data Description and Summary 6
- Weekly Price/Sales Volume/Store Merchandising Conditions (Universal) 6
- Products: Batch 1 (2004-2020): 7
- Product Description: Batch 2 (2021 onward): 7
- Stores, Retail Chains and Channels (Universal): 7

II. Files and File Layouts 9
A. Overview of Files Batch 1 (2004-2020) 9
B. Overview of Files Batch 2 (2021 onward) 10
C. File Layouts: Batch 1 (2004-2020) 11
Table 1: Stores Files 11
Table 2: Products File 12
Table 3: Movement Files 15
Table 4: RMS Versions File 19
Table 5: Product Extra Attributes Files 19
Table 6: Brand Variations File 22
D. File Layouts: Batch 2 (2021 onward) 22
Table 1: Stores Files 22
Table 2: Product Description File 24
Table 3: Movement Files 25
Table 4: Product Hierarchy File 28
Table 5: Product Extra Attributes Files 29
E. Other Tables (Universal) 30
F. Data Notes (Universal) 30
III. Appendix I: Policies (Universal) 35

IV. Appendix II: File Names and Sizes Batch 1 (2004-2020) 38

V. Appendix II: File Names and Sizes Batch 2 (2021 onward) 38

VI. Appendix IV: Version Revisions (Universal) 39

VII. Appendix V: FAQs 42

I. Important notes about this documentation

This document contains information relevant to all years of consumer panel data. Please note that we group the scanner data into two batches because NielsenIQ made significant changes starting in the year 2021: Batch 1 is 2004 - 2020, and Batch 2 is 2021 onward. Below we summarize some of the key differences

1. Batch 1 (2004 – 2020)
    a. A one-to-one product hierarchy organized by 3 attributes: Product Modules, Product Groups, and Departments;
    b. One master UPC file across all years, with UPC versions;
    c. A more limited set of 19 Extra UPC attributes.
2. Batch 2 (2021 onward)
    a. The product hierarchy is now 5 attributes, and is no longer a one-to-one hierarchy;
    b. UPC files will be released annually (rather than having one master UPC file), without UPC versions for 2021 and 2022 data;
    c. The Brand Variations file was eliminated;
    d. A larger set of 65 Extra UPC attributes are included;
    e. Private Label UPCs are still "masked" in the scanner dataset. However, starting in 2021, NielsenIQ implemented a new Private Label UPC coding system. NielsenIQ created 2 cross reference files to help match "old" Private Label UPCs from Batch 1 to the "new" Private Label UPCs in Batch 2;
    f. DMA and Syndicated Major Market (SMM) definitions have been updated.

In this document, you will see text in several colors.

- **Black text** is for content that applies to all years of data, regardless of batch.
- **Blue text** is information that only applies to data from Batch 1: (2004-2020).
- **Green text** is information that only applies to data from Batch 2: (2021 onward).

II. 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. Depending on the year, data are included from approximately 30,000-50,000 participating grocery, drug, mass merchandiser, and other stores. Products from all NielsenIQ-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 2006-2021 are included. We expect to update the data on an annual basis. Updates are expected to be available in the first quarter of each calendar year, and will always lag by 2 years (e.g. 2022 Retail Scanner data is expected to be released in Q1 of 2024).

A. Special notes: Batch 2 (2021 onward)

1. **UPC versioning has been discontinued for 2021 & 2022**: We are no longer "versioning" UPCs across years, and no longer maintaining a "Master UPC file" across all years. Starting in 2021, we include annual UPC files that contain MORE UPCs than those found in the Consumer Panel and Retail Scanner data. Versioning is

back for the 2023 data. You can find the start date of the UPC in the productdesc file. Versioning is back for the 2023 data. You can find the start date of the UPC in the productdesc file.

2. **Private Label UPCs** continue to be "masked" in the Retailer Scanner data. However, starting in 2021, NielsenIQ changed the Private Label coding system in the Retail Scanner data. They also provided two cross-reference files to help map Private Label UPCs in the Retail Scanner data between pre-2021 and 2021. See details in the Product Description file layout, under Note 1. Note: Starting in 2021, Private Label UPCs are now "masked" in Consumer Panel dataset as well. Prior to 2021, Private Label UPCs were "unmasked" in the Consumer Panel data. One of the cross reference files can be used to map "unmasked" Private Label UPCs in the pre-2021 Consumer Panel data to the 2021 "masked" Private Label.

3. **UPC Attribute Code Values have all changed**: NielsenIQ changed the attribute coding systems for all UPC attributes. The code values found in 2021 are completely different from the code values found in prior years.

4. **Product Hierarchy has completely changed**: NielsenIQ completely changed the product hierarchy. The new hierarchy is no longer a 1-to-1 hierarchy of Product Modules, Product Groups, and Departments. However, for at least the near future, NielsenIQ is continuing to provide the Product Module Code in the Product Description File.

5. **UPC data is now contained in 3 UPC files**: There are now 3 UPC files: UPC Description, UPC Hierarchy, and UPC Extra Attributes. The number of UPC Extra Attributes has increased to 65 attributes.

6. **DMA and SMM definitions have changed**: NielsenIQ changed their definitions of Designated Market Areas (DMAs) and Syndicated Major Markets (SMMs). The newest DMA and SMM definitions are contained in a new FIPS-to-DMA and FIPS-to-SMM cross reference file that can be found in the documentation folder.

7. **KFSS** (Kilts File Select System) is not available due to the new hierarchy file sizes. These files will be available in Globus based in the hierarchy level Super Category. Please see the New Hierarchy Outline folder by year in the reference documentation. We organized the data by KFSS Departments, shown in the New Hierarchy SuperCategory tab.

B. Purpose (Universal)

The purpose of this document is to provide a basic understanding of the NielsenIQ 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 NielsenIQ – 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 NielsenIQ. 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 NielsenIQ's data collection approach or methodologies.

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

C. Technical notes regarding data (Universal)

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 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 Batch 1 (2004-2020) 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.

D. 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 (Universal):** 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 NielsenIQ about their features and displays. This "causal" sample is maintained by NielsenIQ 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. NielsenIQ uses a projection methodology to report features and displays to the marketplace. This methodology is not included in this dataset.

- **Products: Batch 1 (2004-2020):** One master UPC table serves both the retail scanner and consumer panel data, though not all UPCs appear in the retail scanner dataset.

  The products found in the data include all products in NielsenIQ-tracked categories that are labeled with a UPC. The hierarchical organization of product data is the same as in the consumer panel dataset (10 Departments, ~125 Product Groups, ~1,100 Product Modules).

- **Product Description: Batch 2 (2021 onward):** One UPC table serves both the retail scanner and consumer panel data, though not all UPCs appear in the retail scanner dataset.

  The products found in the data include all products in NielsenIQ-tracked categories that are labeled with a UPC. The hierarchical organization of product data is the same as in the consumer panel dataset: Department (~19), Super Category (~154), Category (~1,399), Segment (~3,385), Subcategory (~3,562).

  New for 2020 is perishable products that do not use standard UPC codes -- items such as fruits, vegetables, meats, and in-store baked-goods. Note: The magnet data from consumer panel cannot be linked with the retail scanner perishables.

- **Stores, Retail Chains and Channels (Universal):** The retail scanner dataset contains information from approximately 30,000-50,000 individual stores from approximately 90 retail chains (these numbers vary by year). Significant changes occurred again with the 2018 data, and the 2018 Kilts sample includes almost 50,000 stores.

Note that not every NielsenIQ 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 NielsenIQ, the parent_code is either:

- The Corporate Parent (if the company centralizes their data release to NielsenIQ across all of their retail brands)
- The Retail Banner of the store (if the company decentralizes their data release to NielsenIQ 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 level 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 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, journal articles 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 NielsenIQ 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 codes can be found in Table 6.

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 NielsenIQ. Stores may enter or exit the data as they are opened or closed, or as the Kilts sample changes from year to year. 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 Batch 1 (2004-2020)

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 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 2006-2019 is approximately 7.96TB unzipped.

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 Product Modules and corresponding Product 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 someProduct Modules in the hierarchy file are not present in the data, or data appears rarely for these Product Modules. There are two reasons for this, and the hierarchy file indicates the Product Modules to which these exceptions apply.

1) Over time, NielsenIQ stops coding for some Product Module codes. These are called deferred modules. The historical panel data may include purchases made in these Product Modules. In some cases, there may be movement files for deferred Product Modules. **See column G in the hierarchy file for this indicator. Data for ALL deferred Product 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 Product Modules for which we did not receive scanner data one or more years. See column H in the hierarchy file for these notes.

The product hierarchy has a one-to-one mapping, organized as follows:

- Department (10 Departments)
    - Product Group (~125 Product Groups)
        - Product Module (~1,075 Product Modules)
            - UPC

B. Overview of Files Batch 2 (2021 onward)

There are three major types of files associated with the Retail Scanner Data: *Stores*, *Product Description*, and *Movement* (i.e., weekly sales and pricing). The *Stores* file contains information about each individual store location. The *Product Description* file contains information about each UPC. The *Movement* files contain the price and quantity of goods sold at specific stores on a specific week.

While the *Stores* and *Product Description* 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 2006-2020 is approximately 8.48 TB unzipped.

The *Product Description* file has three related files:

- *Product_Extra Attributes*: In addition to the "core" attributes, some products also have "extra" attributes, reported in the *productattributes & productattributes_bp* files. **Not every product has extra attributes.**
- *Product Hierarchy:* Each record contains information for one UPC code. Most (but not all) UPC codes are assigned to 1 or more of the 5 product hierarchy attributes.

Retail Scanner product data are organized into Departments, Super Category, Category, Segment, Subcategory, and UPC Codes. Departments, Super Category, Category, Segment, and Subcategory are all NielsenIQ defined, while UPC codes are defined by manufacturers, with the exception of Magnet products, which are grouped into NielsenIQ-generated UPCs. Private Label UPC codes are also Nielsen-defined.

Note that the hierarchy file has designations of "NOT TO RELEASE." This means that NielsenIQ has stopped coding UPC characteristics for these UPCs and/or these are UPCs for which NielsenIQ does not deliver any data to their clients.

The product hierarchy is as follows (please note that there is no longer a 1-to-1 mapping within these 5 levels of NielsenIQ's product hierarchy. For example, a specific Super Category may map to multiple Departments, and a specific Category can map to multiple Super Categories, etc.):

- Department (19 Departments)
    - Super Category (~154 Super Categories)
        - Category (~1,399 Categories)

- Segment (~3,385 Segments)
    - SubCategory (~3,562 Subcategories)

C. File Layouts: Batch 1 (2004-2020)

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.	
dma_code	Char (5)	The 3-digit Designated Market Area code	F
dma_descr	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 NielsenIQ 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 NielsenIQ, the parent_code is either:**
- **The Corporate Parent (if the company centralizes their data release to NielsenIQ across all of their retail banners)**
- **The Retail Banner of the store (if the company decentralizes their data release to NielsenIQ 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 Reference Documentation 2006-2020 for DMA map and scantrack map with scantrack tables.**

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 that we received from NielsenIQ 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 NielsenIQ 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 NielsenIQ 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.	
product_group_descr	Char (100)	Description of the product group.	
department_code	Integer	Code classifying the product into a department.	
department_descr	Char (100)	Description of the department.	C
brand_code_uc	Integer	Code classifying the product as a specific brand.	
brand_descr	Char (30)	Description of the brand.	G & See brand_variations file information.
multi	Integer	How many of those goods appear in a given pack (e.g. a 6 pack of soda).	
size1_code_uc	Integer	Code used with product module code to distinguish various sizes.	
size1_amount	Decimal (3)	Numeric quantity of the good in individual packaging.	
size1_units	Char (6)	Unit of measure for that quantity.	D
datasets_found_uc	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
size1_change_flag_uc	Integer	Indicates whether the size may have changed during the 2006-2010 period in the Retail Scanner data.	F

A) **UPCs associated with private label goods are altered by NielsenIQ 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. NielsenIQ codes all of the private label brands as "Control Brand" – specifically, the brand description is "CTL BR" (or a similar variation).

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.**

C) **See the Department codes and descriptions below.**

D) **See the Size1_units descriptions below.**

E) **This flag indicates in which datasets the UPC can be found: Retail Scanner, Consumer Panel, or both. An error was found in the flag indicating the dataset in which the UPC can be found. Please do not use this flag.**

F) **When NielsenIQ provided the 2006-2010 Retail Scanner data, we only received the 2010 Size1 values for each UPC. However, NielsenIQ 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	**21** (e.g. baby care, cosmetics, cough & cold remedies, deodorant, hair care, oral hygiene, pain remedies, skin care, shaving)
1	Dry Grocery	**41** (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	**12** (e.g. ice cream, frozen pizza, frozen vegetables)
3	Dairy	**12** (e.g. cheese, eggs, yogurt)
4	Deli	**1**
5	Packaged Meat	**1**
6	Fresh Produce	**1**
7	Non-Food Grocery	**12** (e.g. detergent, diapers, fresheners/deodorizers, household cleaners, laundry supplies, pet care)

8	Alcohol	**4** (e.g. beer, wine, liquor, coolers)
9	General Merchandise	**19** (e.g. batteries/flashlights, candles, computer/electronic, cookware, film/cameras, insecticides, lawn & garden, motor vehicle, office supplies)
99	Magnet Data Products	**1** 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.

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
GRAM	Gram
KG	Kilogram
ML	Milliliter

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.**

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 NielsenIQ to be in alignment with their promotional calendar (e.g. Thursdays through Wednesdays).**

Additional Details from NielsenIQ on this topic:
NielsenIQ 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. NielsenIQ 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 NielsenIQ 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 NielsenIQ 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 NielsenIQ panel data. NielsenIQ 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 NielsenIQ. 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). NielsenIQ 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 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 NielsenIQ 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 NielsenIQ changes the prices to all be on the same basis). The raw data after cleansing is what NielsenIQ and its clients use for analyses. NielsenIQ 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 price/prmult. The week's total dollar sales would be (price/prmult)*units. The total volume would be units*size1_amount*multi in 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. NielsenIQ 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 NielsenIQ. It is up to researchers to decide how to treat product-store-weeks with no reported prices.**

J) **Feature and Display indicator observations: NielsenIQ only observes Features and Displays in a subset of stores (called "audited stores"), and those stores can vary on a weekly basis. For where Features and Displays aren't observed by NielsenIQ: 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 because display execution can vary by store within chain/market.**

Feature and Display values are as follows:
For years 2004-2012:

- **0=No**
- **1=Yes**
- **For 2004-2012, we did not blank out the Feature/Display settings in non-audited stores. Researchers will need to ignore Feature/Display values for non-audited stores.**

You will know that a store was audited by NielsenIQ in a given week if at least 1 UPC has Feature=1 and at least 1 UPC has Display=1. Therefore, when researchers are using causal information (feature & display settings), they should consider only stores that have both Features and Displays for a given week, and should ignore the Feature/Display values for non-audited stores.

Years 2013 forward: We have blanked out the Feature and Display fields for all movement records in stores that were from non-audited stores in any given week. Therefore, only audited stores have the following values (all the stores have blank values):

- **0=No**
- **1=Yes**

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 - NielsenIQ 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. NielsenIQ states that both records are valid. In such cases, researchers must decide how to handle these records. Going forward, NielsenIQ has committed to eliminating this issue when they provide future years of retail scanner data. Due to adding the missing retailers for the 2017 data processed summer of 2021, there are "almost duplicate" records.

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 NielsenIQ 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_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.

In the extra attributes, NielsenIQ 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	
Container_descr	Char (50)	Container description	e.g. plastic, glass, or canned drinks
salt_content_code	Integer	Salt content code	
salt_content_descr	Char (50)	Salt content description	e.g. low salt, sea salt, no salt, etc.
style_code	Integer	Style code	
style_descr	Char (50)	Style description	e.g. shelf-stable product; kosher, dill, sour pickles; salted nuts; breakfast blend coffee, etc.
type_code	Integer	Type code	
type_descr	Char (50)	Type description	e.g. decaffeinated; sugar-free; extra-virgin olive oil; soft-drink; etc.
product_code	Integer	Product code	

product_descr	Char (50)	Product description	e.g. Gravy, Mint Truffle Square, Sauce, etc.
variety_code	Integer	Variety code	
variety_descr	Char (50)	Variety description	e.g. Angel Hair, Beef, Blarney, etc.
organic_claim_code	Integer	Organic claim code	
organic_claim_descr	Char (50)	Organic status claimed by product.	
usda_organic_seal_code	Integer	USDA organic seal code	
usda_organic_seal_descr	Char (50)	USDA Organic Seal on product.	
common_consumer_name_code	Integer	Commonly used name code	
common_consumer_name_descr	Char (50)	CCN description	Perch, shovel, toy, etc.
strength_code	Integer	Strength code	
strength_descr	Char (50)	Strength description	
scent_code	Integer	Scent code	
scent_descr	Char (50)	Scent description	
dosage_code	Integer	Dosage code	
dosage_descr	Char (50)	Dosage description	
gender_code	Integer	Gender code	
gender_descr	Char (50)	Gender description	
target_skin_condition_code	Integer	Skin condition targeted code	
target_skin_condition_descr	Char (50)	Skin condition targeted description	e.g. dry skin, cracked skin, etc.
use_code	Integer	Intended to be used by code	
use_descr	Char (50)	Used by description	Men, Teen, Regular, Child, etc.
size2_code	Integer	Size of product code	A different size measurement (compared to size1_code) of a product
size2_amount	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
brand_code_uc	Integer	Code classifying the product as a specific brand.	A
brand_descr	Char (30)	Primary Brand Description	
brand_descr_alternative	Char (30)	Alternative Brand Description	
start_date	Date (YYYY-MM-DD)	Start date when alternative brand description can be found.	
end_date	Date (YYYY-MM-DD)	End date when alternative brand description can be found.	
datasets_found_uc	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.**

D. File Layouts: Batch 2 (2021 onward)

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.	F
fips_county_code	Integer	The 3 digit county FIPS code where the store is located.	
ACV	Integer	All Commodity Volume	G

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 NielsenIQ 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 NielsenIQ, the parent_code is either:**

- **The Corporate Parent (if the company centralizes their data release to NielsenIQ across all of their retail banners)**
- **The Retail Banner of the store (if the company decentralizes their data release to NielsenIQ 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) **In the reference folder is a map of FIPS codes to the Nielsen DMAs and Scantrack Markets. Geographic Codes: Batch 2 FIPS-to-DMA and FIPS-to-Syndicated Major Market (SMM) Mapping:**

- **SMMs: NielsenIQ no longer uses Scantrack Markets. NielsenIQ explained that the main purpose of the redesign of their SMMs was to create, wherever possible, smaller and more localized markets. Note that numeric SMM Codes no longer exist, just the SMM text descriptions.**
- **DMAs (Designated Market Areas): In 2020 the DMA's from Nielsen Media have changed.**
  We included a new Excel file in the reference documentation called FIPS-SMM-DMA-Convertion-Table-2021.xlsx.

G) **ACV (All Commodity Volume) is the annaul total sales in the store for all products. This number is an average since the actual sales dollars will vary from week to week.**

Table 2: Product Description File

Number of Files: One file per year. Each record contains information for one unique UPC code.

Variable	Format	Description	Other Notes
upc (key)	Text (12)	UPC code of product – manufacturer assigned code. (or NielsenIQ-assigned code for Private Label products).	Use year to link to product_hierarchy and product_extra_attributes files See Note 1 below for Private Label products. See Note 2 about UPCs.
product_descr	Text (100)	Product description	
product_module_code	Integer	Product module code	A list of product module descriptions is included in the attached hierarchy document from 2004-2020 (Excel file). *Note: Product Module Codes and Descriptions have been discontinued as of 2021 data, but for now, NielsenIQ continues to provide this field.
product_module_descr	Text (100)	Product module description	A list of product module descriptions is included in the attached hierarchy document from 2004-2020 (Excel file). *Note: Product Module Codes and Descriptions have been discontinued as of 2021 data, but for now, NielsenIQ continues to provide this field.

multi	Integer	Number of units in a multipack.	This field indicates whether a product is part of a multipack. Value represents number of units in multipack. If the pack is not multi, the value will be 1.
year	Integer (YYYY)		
start_date	YYYY-MM-DD	UPC start date	NielsenIQ has provided us with UPC start dates to help with versioning. Note: Start date is not available for data years 2021 & 2022.

Note 1: Private Label UPCs in the Retail Scanner dataset have always been "masked" (i.e., NielsenIQ converts the actual private label UPCs into a ***Nielsen-generated private label UPC code***, which NielsenIQ calls a "***masked***" UPC code). **Starting in 2021, NielsenIQ implemented a new Private Label UPC coding system. NielsenIQ created 2 cross reference files to help match "old" Private Label UPCs from Batch 1 to the "new" Private Label UPCs in Batch 2.** The files are located in the reference documentation called Private Label 2020 Unmasked 2021 Masked UPCs.txt and Private Label Masked UPCs 2020-2021.txt.

Homescan Consumer Panel private label UPCs from 2004 – 2020 were ***unmasked*** (i.e., private label UPCs in the Homescan data were the ***actual*** UPC codes found on a product's package). Starting in 2021, private label UPCs in the Homescan Consumer Panel data are now "***masked***" The reason NielsenIQ masks private label UPC codes is to keep retailers anonymous. NielsenIQ has provided researchers with a cross-reference file to map 2020 "unmasked" private label UPCs in the Homescan data with 2021 "masked" private label UPCs in the Homescan data. The file is located in the reference documentation and called ***Private Label 2020 Unmasked 2021 Masked UPCs.txt.***

Note 2: We have provided all UPCs, which is significantly more than the movement records, from NielsenIQ to include the new hierarchy and extra attributes.

Table 3: Movement Files

Number of Files: One file for each Super Category from the UPC Hierarchy each year. For 2021 there are 124 files that align with the Super Category Master values. Each of the 124 files is found within one of the 11 directories that group similar Super Category Master Values.

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.**

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 NielsenIQ to be in alignment with their promotional calendar (e.g. Thursdays through Wednesdays).**

Additional Details from NielsenIQ on this topic:
NielsenIQ 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. NielsenIQ 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 NielsenIQ 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 NielsenIQ 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 NielsenIQ panel data. NielsenIQ 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 NielsenIQ. 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). NielsenIQ 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 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 NielsenIQ 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 NielsenIQ changes the prices to all be on the same basis). The raw data after cleansing is what NielsenIQ and its clients use for analyses. NielsenIQ 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 price/prmult. The week's total dollar sales would be (price/prmult)*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. NielsenIQ 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 NielsenIQ. It is up to researchers to decide how to treat product-store-weeks with no reported prices.**

J) **Feature and Display indicator observations: NielsenIQ only observes Features and Displays in a subset of stores (called "audited stores"), and those stores can vary on a weekly basis. For where Features and Displays aren't observed by NielsenIQ: 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 because display execution can vary by store within chain/market.**

Feature and Display values are as follows:

- **0=No**
- **1=Yes**

IMPORANT CHANGE for Years 2021 forward: NielsenIQ has indicated that the Feature and Display fields contained in the movement files are valid for *both* audited and non-audited stores. Therefore, we are retaining all Feature and Display information sent to us by NielsenIQ in the movement files, for both audited and non-audited stores.

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 - NielsenIQ 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.**

Table 4: Product Hierarchy File

Number of Files: One file per panel year. Each record contains information for one UPC code.

Most UPC codes are assigned to 1 or more of the 5 product hierarchy attributes.

Variable	Format	Description	Other notes
upc	Text (12)	UPC code of product – manufacturer assigned code.	Use panel_year to link to product_hierarchy and product_extra_attributes files
year	Integer (YYYY)		
super_category	Text (100)		Product grouping is at the super_category level.
super_category_cd	Integer		
category	Text (100)		

category_cd	Integer		
sub_category	Text (100)		
sub_category_cd	Integer		
segment	Text (100)		
segment_cd	Integer		
department	Text (100)		
department_cd	Integer		
Note: NielsenIQ maps all of the historical UPCs "old" Product Master file to the new "super_category_cd." This mapping will allow researchers to use a common definition of a "super_category" between the "old" hierarchy "product modules" and the "new" hierarchy.			

Table 5: Product Extra Attributes Files

Number of Files: Two files 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 ProductAttributes and ProductAttributes_bp files delivered for any particular year should be identical across the two datasets.*

The UPC codes for some products have additional attributes. There are 65 total extra attributes (each with a "code" and "description" field). However, many products do not have any value listed in these fields. This file layout is NOT identical to the Batch 1 Product Extra Attributes file layout.

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

Due to the size of this table, to see the names and descriptions of all 65 attributes, please see **Product Extra Attributes 2021.xlsx** in the reference documentation. Also, due to size, we have broken the attributes into two files on Globus. The 2 files are: 1. Regular attributes, and 2. Branded Pack (BP_CONCAT) attributes. Banded Pack products are products where the manufacturer has shrink wrapped multiple products together.

E. Other Tables (Universal)

Table 6: Channel Codes

Channel Code	Channel Description
C	Convenience
D	Drug
F	Food
G	Convenience*
M	Mass Merchandiser (also includes Pet stores)
L	Liquor

*2016 – NielsenIQ created a new code "G" to indicate convenience stores that have Gas stations. However, a "G" code is assigned if the parent retailer mostly has gas stations associated with their stores. That code is not intended to indicate that a particular store has a gas station, but rather that the parent retailer mostly has gas stations at their C-stores. This is why, in the Kilts Center's scanner data archive, almost all of the Convenience stores have a "G" channel code starting in 2016, and there are not many store records with Channel Code "C" any longer.

F. Data Notes (Universal)

Retailer Level Analysis Between the Retail Scanner Dataset and Consumer Panel Dataset (Universal)
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.

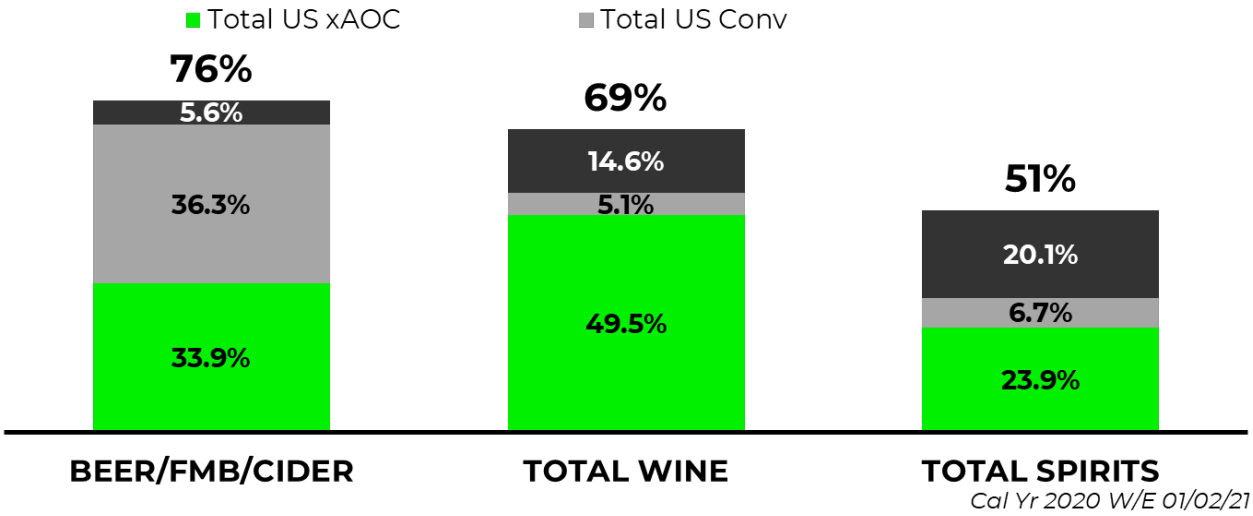
Uniqueness of Alcoholic Beverage Categories (Universal)

For researchers looking using data from the alcoholic beverage category, NielsenIQ advises several notes of caution.

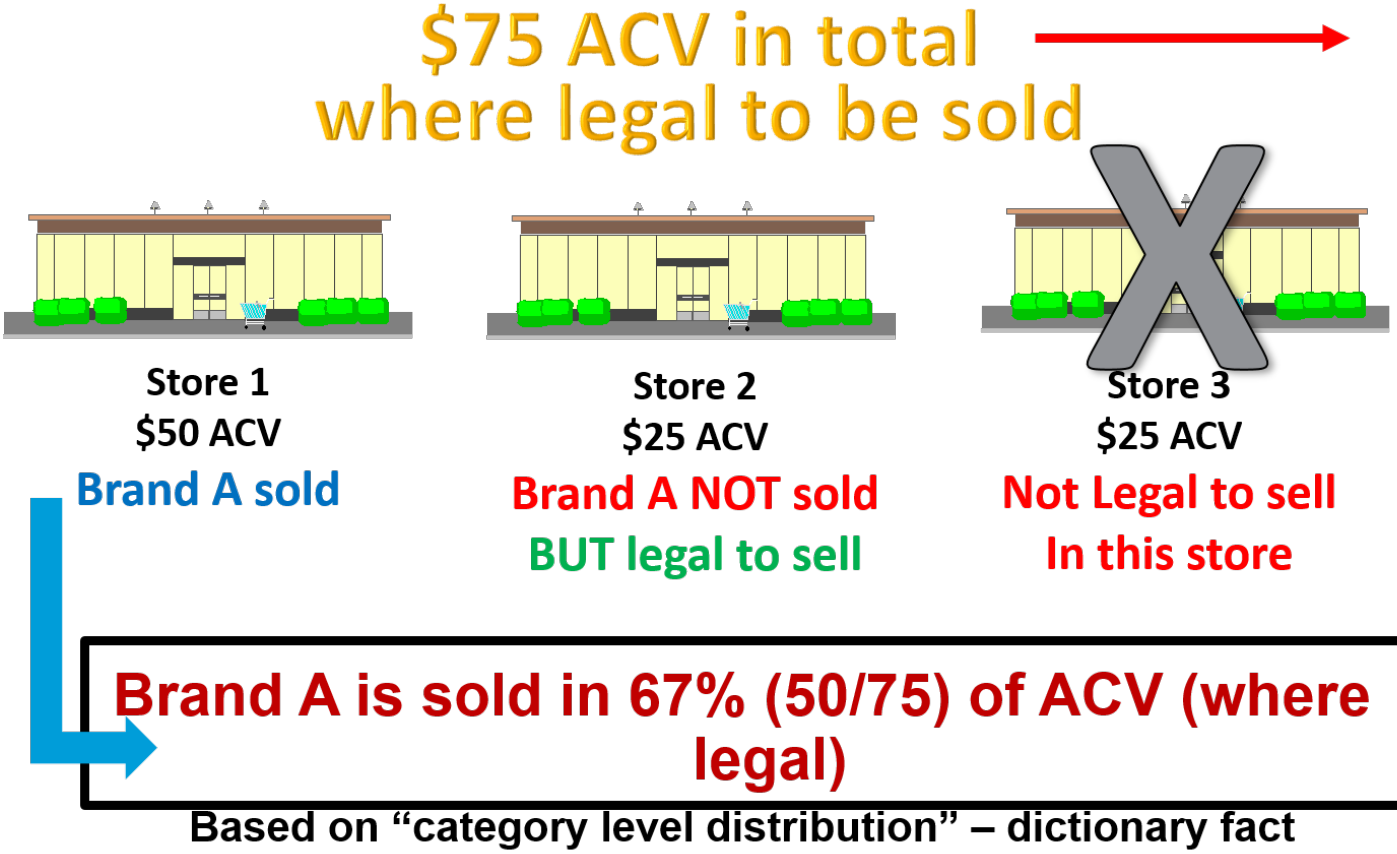
- Alcohol state laws change – sometimes slowly – which can add layers of difficulty for tracking or evaluating long-term trends.
    - Examples:
        - Washington allowed spirits to be sold in grocery stores mid-2012
        - Tennessee allowed wine to be sold in grocery stores mid-2016
        - Arkansas changes in October 2017 allowing expanded wine sales in grocery
        - Oklahoma will allow wine and full-strength beer in grocery and convenient stores October 2018
        - Other gradual changes in licensing over several years have occurred int Colorado, Pennsylvania, and Massachusetts
- Keep in mind that NielsenIQ scanner data only accounts for off-premise alcohol purchases in channels such as grocery, drug stores, mass merchandise and a small amount of the volume for convenience and liquor stores. Neither on premise sales (e.g. restaurant and bars) nor e-commerce sales are covered by scanner data.
- Collected/reported alcohol purchases data only comes from consumers 21 years and older.
- The following graph illustrates estimates of volume coverage of alcoholic beverages by category

NielsenIQ estimated volume coverage (cases)

Total US



- Alcoholic beverages have unique ACV distributions facts (i.e. where legally distributed). For example, Brand A may have $75 ACV where it is legal to sell, however, only a portion ($50/$75 or 67%) of that ACV is sold in legal off-premise locations. The other $25 ACV is legal to sell in a location but is not sold. See illustration below.


$75 ACV in total
where legal to be sold
Store 1
$50 ACV
Brand A sold
Store 2
$25 ACV
Brand A NOT sold
BUT legal to sell
Store 3
$25 ACV
Not Legal to sell
In this store
Brand A is sold in 67% (50/75) of ACV (where legal)
Based on "category level distribution" – dictionary fact

ALCOHOL LAWS VARY BY STATE AND EVEN SOMETIMES BY COUNTY

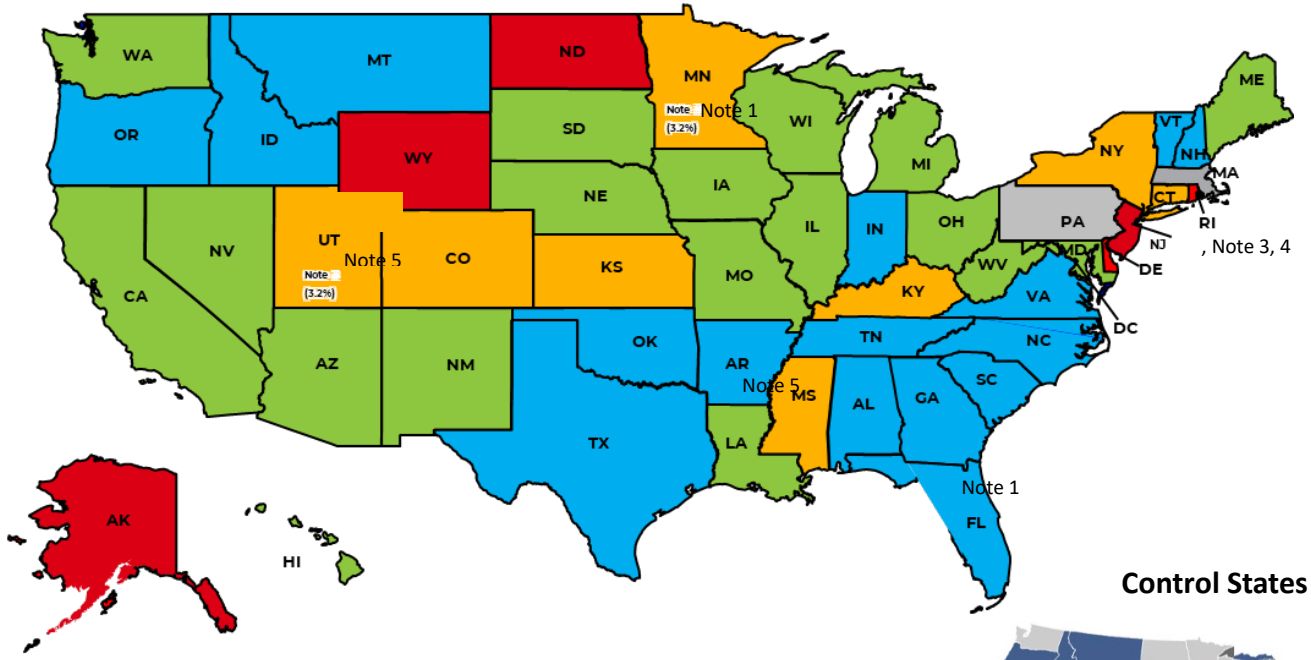
Wine, Liquor, and Beer available in grocery stores

Wine and Beer available in grocery stores. No Liquor can be sold in grocery stores

Beer available in grocery stores. No Wine or Liquor can be sold in grocery stores

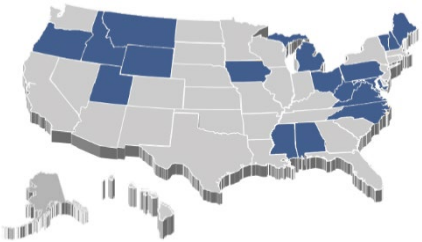
No Alcoholic Beverages can be sold in grocery stores

Limited # of licenses per chain allowed to sell one or more categories (generally Beer and Wine); increases gradually over time



Where legal to sell in **GROCERY**

Control States



	Wine, Liquor and Beer available in grocery stores.
	Wine and Beer available in grocery stores. No Liquor can be sold in grocery stores.
	Beer available in grocery stores. No Wine or Liquor can be sold in grocery stores.
	No Alcoholic Beverages can be sold in grocery stores.
	Wine and Beer less than or equal to 5% ABW available in grocery stores. No Liquor or Products greater than 5% ABW can be sold in grocery stores.

NOTE:

1. Separate entrance for alco required. (For liquor in FL, wine & beer in AK).
2. Beer sales limited to 3.2% in MN
3. Selected accounts sell wine & beer, varies by local ordinance.
4. Selected accounts sell liquor, varies by local ordinance.
5. Beer sales limited to 5.0%.

III. Appendix I: Policies (Universal)

Researchers who have received the data, including all paid 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. Please view our policies page for updates that happen after this document is prepared.

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

- Unless otherwise approved by NielsenIQ and the Kilts Center, I will not conduct research commissioned, sponsored, supported, or funded by an industry source, government organization, industry group, nonprofit organization, or private company.
- I will only use these data for academic research purposes.
- I will not use the data for consulting purposes of any kind, including litigation support.
- Evaluating NielsenIQ's data collection approach or methodologies is not a primary purpose of this project.
- All coauthors must be employed full-time by and/or enrolled as a student in an academic institution. Coauthors outside of an accredited academic institution are prohibited.
- Client shall not use the Services or Information (a) to enrich or enhance third-party retail, market or consumer measurement data sets; (b) in any Client external facing application or any third-party application; or (c) in conjunction with any machine learning, neural network, deep learning, predictive analytics or other artificial intelligence computer or software program.

Researchers must also agree to the following NielsenIQ 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 unless they are my paid research assistant.
- I will ensure that data is stored only on devices and computers owned by the subscribed institution or permissible Third Party Hosting Systems.
    - The permissible Third Party Hosting Systems must have at least one of the following information security certifications: (a) federal Family Educational Rights and Privacy Act (FERPA); (b) SSAE16 SOC2; and/or (d) ISO27001 security certifications.
- I will not disclose any personally identifiable information about any consumer panelists, or details that would permit re-identification of NielsenIQ's consumer panelists, without prior NielsenIQ approval.

- I will not disclose any information about specific retailers or individual stores, or details that would permit re-identification of NielsenIQ's retailers or individual stores, without prior NielsenIQ approval.
- I will not disclose UPC-level pricing data that may be incorporated in the NielsenIQ data received from retailers who have agreed to provide this data to NielsenIQ.
- 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.
- I will disclose any changes to their project scope, co-authors, or funding sources during the course of my research.
- If I am licensed by NielsenIQ to use any other consumer panel data, I understand that I must have written permission from NielsenIQ to link anonymized household IDs.

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

1. I will submit an annual update to the Kilts Center that includes copies of all working papers arising from use of the NielsenIQ data. Continued use is contingent upon submitting an annual status report and adhering to the terms and conditions of the contract.
2. For final publications, I will submit these papers to NielsenIQ and the Kilts Center at least 30 days prior to public dissemination.
3. I will not publish more than limited excerpts of NielsenIQ information in a bona fide publication of academic research - a "limited excerpt" means such portion of the NielsenIQ data that would not ordinarily substitute for a purchase of access to the NielsenIQ data.
4. I will include in all presentations, papers, dissertations, abstracts, manuscripts, reports, journal articles, and any other publication type the following disclaimers/citations:

NielsenIQ - Consumer Panel and Retail Scanner

"Researcher(s)' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business."

"The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein."

Nielsen - Ad Intel

"Researcher(s)' own analyses calculated (or derived) based in part on data from The Nielsen Company (US), LLC and marketing databases provided through the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business."

"The conclusions drawn from the Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein."

NielsenIQ and Nielsen

"Researcher(s)' own analyses calculated (or derived) based in part on (i) retail measurement/consumer data from Nielsen Consumer LLC ("NielsenIQ"); (ii) media data from The Nielsen Company (US), LLC ("Nielsen"); and (iii) marketing databases provided through the respective NielsenIQ and the Nielsen Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business."

"The conclusions drawn from the NielsenIQ and Nielsen data are those of the researcher(s) and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein."

5. I will submit presentations, papers, dissertations, abstracts, manuscripts, reports, journal articles, and any other publication type to Kilts Center **30-days prior to any public dissemination**.
6. After the 30-day review period, I will submit all working papers to the Kilts Center for Marketing NielsenIQ Data Research Working Paper Series at Social Science Research Network (SSRN).
7. I will notify the Kilts Center of publications resulting from my work with the data, even after I have stopped using the data.

IV. Appendix II: File Names and Sizes Batch 1 (2004-2020)

1. **Stores**: 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. **Products**: 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:	~640 MB

3. **Movement**: 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. **RMS Versions**: 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. **Product Extra Attributes Files**: 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_YYYY.tsv
Files sizes:	Ranges from ~70-120 MB each year

6. **Brand Variations**: 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:	~75 KB

V. Appendix II: File Names and Sizes Batch 2 (2021 onward)

1. **Stores**: 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. **Product Description**: 1 file for each year. An updated version of this file will be regenerated each year with annual updates.	
Naming convention:	Product_Descr_YYYY.tsv
File Size:	~ 823 MB each year

3. **Movement**: 1 file for each product hierarchy super category for each year.	
Naming convention:	<supercategory;department>.tsv
File sizes:	~ 802 GB per year

4. **Product Hierarchy:** 1 file for each year. One new file will be created for each additional panel year as data is released.	
Naming convention:	Product_hierarchy_YYYY.tsv
File sizes:	~ 1 GB per year

5. **Product Extra Attributes Files**: 1 file for each year. One new file will be created for each additional panel year as data is released.	
Naming convention:	products_extra_YYYY.tsv
Files sizes:	~ 4.22 GB per year

VI. Appendix IV: Version Revisions (Universal)

Revisions 2024-12-01

1. NIQ provided the UPC start date, which now serves as the basis for UPC versioning. This information is available in the productdesc file. Note that UPC versioning is not available for the data years 2021 and 2022.
2. ACV (All Commodity Volume) has been added to the *stores* file for the 2020 data.
3. SMM ACV (All Commodity Volume per market area) for 2023 can be found in the reference documentation. NIQ plans to provide this data annually moving forward.

Revisions 2024-03-31

1. Updates were made to the Alcohol chart (page 32) and map (page 34).
2. ACV – All Commodity Volume has been added to the stores file.
3. A new movement super_category called “OTHER” includes UPCs that are not in the hierarchy.
4. SMM ACV - All Commodity Volume per market area for 2020 – 2022, found in the reference documentation.

Revisions 2023-10-01

1. Significant changes for the 2021 data moving forward are outlined throughout the document.
2. The scanner data is from December 26, 2020 – January 1, 2022 whereas the panel year for 2021 data is January 3, 2021 – January 1, 2022. This is a correction from the 2020 data.

Revisions 2022-02-14

1. Perishables were added to the 2020 data to include bakery items, fruit, vegetables, etc.
2. The 2020 scanner data is from December 29, 2019 – December 27, 2020 whereas the panel year for data is December 29, 2019 – January 2, 2021. The missing week in retail scanner will be delivered with the 2021 data.
3. In September 2020, the 2017 missing retailers were added to the data. Note: Researchers will find "almost duplicates" in the data.

Revisions 2021-03-05

1. Nielsen has split into two separate companies on March 5, 2021. NielsenIQ covers consumer panel and retail scanner. Nielsen covers Ad Intel.

Revisions 2020-05-31

1. Several new retailers gave permission for their data to be included with the 2018 release, so the number of Stores increased to ~50,000.

Revisions 2019-01-31

1. Policy section updated to reflect newly revised or refined contract requirements. Researchers should regularly review policies starting on page 35.
2. All Commodity Volume by Channel chart was updated under section Data Description and Summary to reflect the 2017 store composition. There was a significant change in store composition that occurred starting in 2017.

Revisions 2018-02-19

1. Additional notes regarding alcoholic beverage categories. See Uniqueness of Alcoholic Beverage Categories.
2. Added map illustrating alcoholic beverage sales and laws around the U.S.

Revisions 2016-10-26

1. Additional information provided for feature and display variables. See feature and display details for note J in table Table 3: Movement Files.

Revisions 2016-1-22

1. Known issue regarding 38 stores in 2012. See the note on page 12 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
3. Table 5: Product Extra Attributes Files for details.

4. Pricing note added regarding products with a price of one cent (or similar amount). See note N in Table 3: Movement Files.
5. Known issue regarding missing movement data for the state of Wisconsin in 2006. See note on page 19 for more information.
6. 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. NielsenIQ has just begun coding products with these data. Future years are expected to included 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.

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 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.

Revisions 2014-09-22

1. Updated footnote about control brands. See footnote A in section Table 2: Products File.
2. Note regarding article submissions and data. See Technical notes regarding data.
3. Added information about deferred modules. See description of deferred modules in Overview of
.

4. Updates regarding stores indicated with features and displays. See second paragraph of footnote J for the Movement Files layout on page 15.

Revisions 2014-04-02

1. Addition of section Technical notes regarding data.

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.

VII. Appendix V: FAQs

General Questions (Universal)

1. What is the format of the files?

- The files are in the tab-delimited-format, with the variable names in the first row.

2. Why do some variables have "_uc" attached to the end?

-The suffix "_uc" denotes that these variables are generated by the Kilts-NielsenIQ Data Center and were not originally part of the NielsenIQ data. This tends to happen in cases when The NielsenIQ Company has requested the data to be masked, or it indicates that the Kilts Center generated new fields that would be helpful to researchers.

3. What standard is used for the UPC variable? Why can't I merge another UPC dataset with the UPC's found in the Retail Scanner Data?

- The UPC codes (UPC and UPC_ver_uc) in the consumer panel data do not include the check digits. The codes are made up of the EAN (International Article Number), with

the check digit dropped.

- In order to merge the UPCs of the consumer panel data with external data containing other formats and types of UPCs, researchers can generate the check digits. It is important to note that the consumer panel data promotes all UPC-Es (short form codes) to UPC-As, minus the check digit. See #8 for more details about steps for merging consumer panel UPC codes with external datasets.

4. Why do I see UPCs aligned with different products over time?

- Manufacturers reuse UPCs over time. When an old UPC is assigned to a new product, the consumer panel purchase transactions may reflect the previously assigned product characteristics for some time before NielsenIQ recognizes that the UPC is being reused. In the purchase data, a reused UPC may initially come through with the "old" product characteristics. We do not make retroactive corrections for these UPCs. Instead, we provide the data the same way we receive it from NielsenIQ.

5. Is it possible to generate the check digit to add to the retail scanner UPC codes (UPC and UPC_ver_uc)?

- Although the check digit does not identify a product but rather to validate data entry, some external resources do retain these numbers. Therefore, if looking to compare these variables to external data, researcher s may want to compute the check digit of a UPC (UPC-A, EAN-8, or EAN-13). To do so, follow this algorithm:

  1) Starting at the right most digit excluding the check digit, sum all the digits moving left, multiplying every other number by 3.
  2) Take this result module 10.
  3) If the result is not 0, subtract from 10. This is your check digit; otherwise, 0 is your check digit.

  An example for EAN-13 7895144603049 (789514460304 with no check digit):

  1) 7+8*3+9+5*3+1+4*3+4+6*3+0+3*3+0+4*3 = 3*(4+3+6+4+5+8) + (0+0+4+1+9+7) = 111
  2) 111 mod 10 ≡ 1
  3) 10 - 1 = 9, which checks.

  A final note of concern is that EAN-8 and UPC-E may collide. While UPC-E are not stored in the consumer panel data, check with your data sources to be sure whether this may occur and how to handle it.

6. My external data source shows a different type of UPC code. Why, and what can I do to compare consumer panel data to these other data?

- Researchers using UPC codes from other, external datasets may find that the UPC codes do not match those in the consumer panel data. There are in fact four widely used barcode standards: UPC-A, UPC-E, EAN-13, and EAN-8 (listed below), which can be converted for comparison or merging purposes.

    - UPC-A is the long existing 12-digit UPC codes.
    - UPC-E is the 6 digit codes found especially on smaller goods, e.g. packs of gum.
    - EAN-13, a 13 digit code, is a superset of UPC-A recently adopted to alleviate crowding of the space available to manufacturers to list the UPC code for an increasingly large range of products.
    - EAN-8 is an 8 digit code that serves an analogous purpose to UPC-E.

- It is possible to convert between these standards in certain circumstances, except for EAN-8. As mentioned, EAN-13 is a superset of UPC-A. You can always convert a UPC-A to EAN-13 by simply prepending a 0 to the UPC-A code. EAN-13 can only be converted to UPC-A when its first digit is 0 by dropping the 0. Be careful to account for whether your data contain a check digit to know whether the EAN-13 should be 12 or 13 digits. When stored as a numeric type, you should see no difference between an equivalent UPC-A and EAN-13 as their check digits are the same.

- Converting between UPC-A and UPC-E is also possible, however only a subset of UPC-A's can be converted to UPC-E's. Although it is possible to detect when a UPC-A is eligible to be converted to UPC-E, there is no way to know if the manufacturer used a UPC-A or UPC-E by just looking at the code. Thus, in an external data source, you may find that the product is a UPC-A or a UPC-E. Because all UPC-E's can be converted to UPC-A's, it is our recommendation that all UPC-E's be promoted to UPC-A's/EAN for maximum compatibility. In the consumer panel data, all UPC-E's are promoted to UPC-A's, minus the check digit.

- A further complication exists with regard to UPC-E's. UPC-E's are sometimes referred to as 6 digit and sometimes as 8 digit codes. The difference is that the 8 digit version includes a leading 0 and the UPC-A calculated check digit is appended.

- To convert a UPC-E to UPC-A, use the following table where the UPC-E is "abcdef" (alternatively "0abcdefC") and C is the UPC-A computed check digit:

    f UPC-A

0-2	0abf0000cdeC
3	0abc00000deC
4	0abcd00000eC
5-9	0abcde0000fC

7. What does RBC stand for in the product_module_descr mean?

- RBC stands for Regionally Brand Coded.
- These modules (RBC) are deferred meaning they are not “fully coded” modules and have a limited set of characteristics in both Consumer Panel and Retail scanner.
- Caution is advised for all “deferred” modules.

8. What does “Reference Card” mean in the product_module_descr?

- These are products without UPC codes, (e.g., fresh produce, bakery, floral, etc.).

9. Is there a way to decipher the UPC description?

- This is the website https://www.upcitemdb.com. Choose "Validator" not "UPC" section and paste the whole digit of a UPC code from the NielsenIQ data. Not all UPC codes come up on the page, but many have enough to identify the abbreviations of UPC descriptions.

General Questions Batch 1 (2004-2020)

10. What is the “upc_ver_uc” variable?

- The characteristics associated with a UPC code sometimes change in the data. In some cases, the same UPC code can refer to completely different products. In other cases, just a few characteristics of a product may have changed. Finally, sometimes the product has been assigned to a different product module by NielsenIQ, and none of the main characteristics have changed.

  For example, a characteristic may change due to a temporary or permanent change in a product characteristic (e.g. the size of the product changes). However, in some cases the change may simply be due to NielsenIQ-generated fields (e.g. a UPC code may have been assigned to a new Product Module Code).

  Instead of losing information or imposing our own interpretations, we assigned a UPC version to indicate the different versions. Therefore, to merge the product characteristics into the purchase file, one should use both UPC and UPC_ver_uc as the merging variables.

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
- multi
- size1_code

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 promotion 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.

11. Why do some modules show a noticeable change in the number of total purchases from year to year? Are there missing transactions?

- The definition of product modules can change over time and therefore a UPC code may change modules over time.

12. Why is some attribute information missing from the brand and size fields in the products file?

- There are a small number of cases where a brand_code_uc value exists but brand_description is missing, and/or a size1_code_uc value exists but size1_units and size1_amount are missing. These attribute descriptions are not available from NielsenIQ.

13. Why are UPCs listed in the product file but not in the movement data?

- We use a master file from NielsenIQ that **does not** confirm if the UPC is in the movement file.
- If you cannot find the UPC in the Movement files listed in the Product file, it does not exist.

14. In product file, datasets_found_uc list RMS or ALL, but the UPC cannot be found?

- An error was found in the flag indicating the dataset in which the UPC can be found. Please do not use this flag.

General Questions Batch 2 (2021-onward)

15. Why do the brand codes from 2004-2020 not match the 2021 data?

- Starting with 2021 data, NielsenIQ transitioned their UPC database to a new global platform. As a result, starting with 2021 data, the UPC coding system for all UPC attributes has changed. Therefore, all UPC attribute codes (including Brand Codes) change starting in 2021.

16. ***The UPCs from 2021 onward no longer have UPC versions. How should researchers determine what version the UPC is now?***

- For 2004-2020, we assigned a new UPC version if any of the following 4 "core" attributes changed: Product Module, Brand, Size, or Multi. Starting in 2021, we are providing annual UPC files, and are no longer "versioning" UPCs. Researchers will need to perform their own analysis to investigate any changes in a UPC's characteristics. Researchers can look up the UPC characteristics each year using the annual UPC files, compare those characteristics vs. prior years to see if there were changes, and make their own determination as to which UPC characteristics changed and whether to those changes should be considered a new "version" for any particular UPC. Starting in 2023, NIQ has provided us with the UPC start date in the product description file for versioning. Versioning does not exist for data years 2021 & 2022.

17. ***I am trying to map pre-2021 private label UPCs to Batch 2 private label UPCs, but the matching file appears to be have many duplicate matches, and it is not one-to-one. Can I still use it?***

-The additional private label matches were added in error from a different list by NielsenIQ. The document cannot be reliably used. NielsenIQ is working on sending a corrected matching file. When they do, researchers will be notified by email and provided with the file.