

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT



Consumer Panel Dataset Manual

The Kilts Center Archive of
The Nielsen Company

1/22/2016

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Table of Contents

I.	Introduction	5
A.	Purpose.....	5
B.	Special notes regarding data	5
II.	Consumer Panel Data Summary	6
A.	Years Available and Panel Size	7
B.	Panelist Demographic and Geographic Data.....	7
C.	Trips and Purchase Transactions	7
D.	Retailers.....	8
E.	Products.....	9
F.	Magnet data	10
G.	Overview of File Types.....	10
i.	Panelists	10
ii.	Trips.....	10
iii.	Purchases	11
iv.	Retailers	11
v.	Products	11
vi.	Product Extra Attributes	11
vii.	Brand Variations.....	11
III.	Data Collection Methodologies	12
A.	Sampling	12
B.	Incentives.....	14
C.	Panelist Retention	14
D.	Validation and Quality Checks to Ensure Data Accuracy.....	15
E.	Purchase Static	15
F.	Projection System.....	15

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

G.	Decomposing Sales by Consumer Segment	18
H.	Expected Sampling Variability	18
I.	Price Determination	20
IV.	Detailed File Layouts	20
A.	Panelists File	20
B.	Trips File.....	25
C.	Purchases File	27
D.	Retailers File	28
E.	Products File	28
F.	Products Extra Attributes Files	32
G.	Brand Variations File.....	36
V.	Appendix I: Policies	37
VI.	Appendix II: Codebook.....	39
A.	Panelist File Codes.....	39
i.	Demographic Codes	39
ii.	Geographic Codes	43
iii.	Ownership and Usage Codes	48
B.	Trips File Codes.....	49
C.	Retailers File Codes.....	50
i.	Retail Channel Descriptions	50
D.	Products File Codes	51
i.	Departments	51
ii.	Size1_Amount Description.....	51
VII.	Appendix III: File Names and Sizes.....	53
VIII.	Appendix IV: Maps	55
IX.	Appendix V: FAQs.....	58

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

A. General Questions	58
B. Years Available and Panel Size	59
C. Panelist Demographic and Geographic Data.....	59
D. Trips and Purchases.....	63
E. Retailers.....	69
F. Products and Product Extra Attributes.....	71
G. Other.....	75
X. Appendix VI: Version revisions.....	78
A. Revisions 2016-1-22	78
B. Revisions 2015-05-25	78
C. Revisions 2014-12-23	79
D. Revisions 2014-09-22	79
E. Revisions 2014-04-02	79
F. Revisions 2014-03-05	79
G. Revisions 2013-10-21	80
H. Revisions 2013-09-19	83
I. Revisions 2012-12-28	83
J. Revisions 2012-11-09	83
K. Revisions 2012-06-22	84
L. Revisions 2012-05-14	84
M. 2012-03-28	84

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

I. Introduction

Please note that a detailed list of revisions to the data and documentation from the prior Consumer Panel data release are listed in chronological order in

A. Purpose

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

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

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

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

All approved researchers are bound by policies found in [Appendix I: Policies](#) (page 37) as well as any additional limitations named in their respective contract between Chicago Booth and their individual academic institution.

In addition to this document, we provide an FAQ document and an excel file with product hierarchies (modules, groups, and departments). Please refer to the documentation folder and contents, provided to you with the data.

B. Special notes regarding data

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

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

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.

- b. For the products and brand variations file, some fields contain literal single or double quote characters, e.g. in the brand_descr field. Many import parsers fail to properly handle the way we have written these characters. It may be necessary to pre-process or use 3rd party importers if your software does not handle the file correctly.
2. 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.
 3. When we released the 2012 data, we made an update that affects 4 UPC codes ONLY present in the panel data in 2011. For the following four UPCs, the brand_code_uc field was not unique. In the products.tsv file released with the 2012 data, these brand codes were corrected in 2011 to reflect the new brand_code_uc listed below.
 - UPC=001200040006, Current (incorrect) brand_code_uc=669917, NEW brand_code_uc=670141, brand_desc=DIET PEPSI
 - UPC=004300084024, Current (incorrect) brand_code_uc=669914, NEW brand_code_uc=670138, brand_desc=KOOL-AID
 - UPC=004300083992, Current (incorrect) brand_code_uc=669916, NEW brand_code_uc=670140, brand_desc=KOOL-AID TWISTS
 - UPC=067609010043, Current (incorrect) brand_code_uc=669915, NEW brand_code_uc=670139, brand_desc=VICTORY LIGHT USA LLC-NBL

II. Consumer Panel Data Summary

The Consumer Panel Data represents a longitudinal panel of approximately 40,000-60,000 U.S. households who continually provide information to Nielsen about their households, what products they buy, as well as when and where they make purchases. The Nielsen Homescan panelists use in-home scanners to record all of their purchases, from any outlet, intended for personal, in-home use.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Products include all Nielsen-tracked categories of food and non-food items, across all retail outlets in all U.S. markets. Nielsen samples all states and major markets. Panelists are geographically dispersed and demographically balanced. Each panelist is assigned a projection factor, which enables purchases to be projectable to the entire United States.

A. Years Available and Panel Size

The Consumer Panel data tracks purchases of a panel of households. The data describes when, where, and what the panelists purchase, and at what price. Data is initially available for the years 2004-2010, with annual updates planned for future years. Updates are expected to be available in the first quarter of each calendar year, and will always lag by 2 years (e.g. 2013 Consumer Panel data is expected to be released in January 2015). The data contains approximately 40,000 households for 2004-2006, and 60,000 for 2007 onwards. Some panelists stay on the panel for several years, while others may join or drop off each year. (Also see Note 2 under [Panelists File](#) layouts (page [20](#)) for more information about the “Panel Year”, which usually will not exactly coincide with the calendar year).

B. Panelist Demographic and Geographic Data

Demographic and product ownership variables are recorded for the entire household and the head of household, as well as demographics for other household members. Demographic variables include household size, income, age, presence and age of children, employment, education, marital status, occupation, type of residence, and race.

Panelists complete a survey of household information and update it annually, with the exception of the WIC indicator. See Data Collection Methodologies, [Sampling](#) (page [12](#)) for additional information about demographic composition. See Appendix II: Codebook, [Demographic Codes](#) (page [39](#)) in for information about the WIC indicator variable, which researchers should use with caution.

The panelists are geographically dispersed. Geographic information available for each panelist includes their zip code, FIPS state and county codes, DMA code, Scantrack Market code (assigned by Nielsen), and region.

The panel composition is designed to be projectable to the total United States population, and to some Scantrack markets.

C. Trips and Purchase Transactions

For each shopping trip, summary information is provided about the trip (e.g. household, date, retailer code, store code, store zip code, and total dollars spent). Store code and store zip are provided only for

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

a subset of trip. Note that as part of the agreement with Nielsen, in order to protect the confidentiality of individual stores, only the first 3 digits of a store's zip code are available.

Not all scanner equipment provides an exact zip code. Therefore some zip codes are inferred from a panelist's home zip code, based on the fact that Nielsen research finds most panelist trips are made to a store closest to their home.

Within a trip, detailed transaction information is reported for each product purchased (e.g. UPC code, quantity, price, deals, and coupons). See Detailed File Layouts for sections [Trips File](#) (page 25) and [Purchases File](#) (on page 53) for further details.

Where available, individual stores have been assigned a code that will enable the Consumer Panel Data to be linked to the Retail Store Scanner Data. This code is called "store_code_uc." For researchers familiar with TDLinx codes, the "store_code_uc" code serves the same purpose as the TDLinx code for a store.

Researchers interested in knowing the sequence in which data fields are entered by the panelist, in general the sequence of data entry is:

1. Information related to the overall trip (e.g. the date and store -- note that different scanners use different methods for helping the consumer select the store)
2. Individual UPCs purchased:
 - a. Price - entered only if the product was purchased at a store that doesn't provide Nielsen with retail scanner data. Otherwise, the price is filled in by Nielsen.
 - b. Number of units purchased
 - c. If the item was perceived by the consumer to be on a "deal", then this is indicated by the deal_flag_uc variable.
 - d. If the panelist used a coupon, the total discount amount due to coupon is recorded.

D. Retailers

Data collection is performed for products purchased in many different retail channels. Nielsen assigns each retail chain a retailer code and a channel type. Channels are classified into 66 mutually exclusive categories. See Detailed File Layouts, [Retailers File](#) (page 28) for further information.

Note that as part of the agreement with Nielsen, retailer names are not revealed and individual stores are not identified. If through other means a researcher is able to determine a retailer's name or an individual store's identity, the legal agreement signed by each researcher **strictly prohibits the public disclosure in any format or forum of any retail chain names or any individual store names.**

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

E. Products

UPC codes for all 10 Nielsen Food and Non-Food Departments are included in the product data. Detailed attributes for each UPC code are also included. Changes to any core attributes of a UPC code are tracked over time by assigning different UPC code “versions.” Further, beginning in 2007 Nielsen includes additional products called “Magnet.” See [Magnet data](#) (page [10](#)) for more detail.

Consumer Panel product data are organized into Departments, Product Groups, Product Modules, and UPC Codes. Departments, Product Groups and Product Modules are all Nielsen defined codes, while UPC codes are defined by manufacturers, with the exception of Magnet products, which are grouped into Nielsen-generated UPCs. As part of the reference files you received with the data, we include a copy of the product module hierarchy table (an Excel file) that lists all modules and corresponding groups and departments.

Note that the hierarchy file indicates product modules that are “deferred.” For most modules, this means that Nielsen has stopped coding UPCs and UPC characteristics for that module. These modules are rarely present in the scanner data. Historical panel data may include purchases made in these modules. **See column G in the hierarchy file for this indicator. Data for ALL deferred modules should be used with caution because the UPCs may not be fully coded, or the UPC information may be outdated.**

The product hierarchy is as follows:

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

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

See [Detailed File Layouts](#) (page [20](#)) for further details.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

F. Magnet data

The Magnet data include products that do not use standard UPC codes -- items such as fruits, vegetables, meats, and in-store baked-goods. Only some of the panelist households report these types of purchases. Magnet data is not available prior to 2007.

Because only a subset of households report Magnet product purchases, researchers must note that we outline several cautions for how to use the projection system when analyzing data from these households. These are explained in detail in [Projection System](#) (page 15).

The Magnet product module codes were expanded starting in 2011, and now Magnet Products are no longer classified in only one Product Group (99) and only one Department (99).

Years	Product Module code	Product Group codes	Department codes
2007-present	0750	99	99
Starting in 2011, Nielsen added new magnet data Product Modules codes:	0445-0468	Varies by product module, but includes many different Product Group codes	Varies by product module, but includes many different Department codes

G. Overview of File Types

Seven types of files can be found in the Consumer Panel data: Panelists, Trips, Purchases, Retailers, Products, Product Extra Attributes, and Brand Variations. See [Detailed File Layouts](#) (page 20) in this document for detailed descriptions of each type of file.

i. Panelists

Panelist files contain demographic, geographic, and product ownership information on the panelists. Nielsen collects and updates this information annually, so new panelist files are available each year. After performing basic data validations, Nielsen reports the demographic data the way it was entered by the panelist. Note that some of the panelists are also part of Nielsen's "Magnet" panel. See [Magnet data](#) (page 10) explanation for more details.

ii. Trips

Trips files provide summary information about each of the shopping trips made by panelists (e.g. the panelist ID, date, retail chain shopped, and total amount spent for that trip).

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

iii. Purchases

Purchases files provide information about the specific products a panelist bought on a shopping trip, including the quantity, price paid, and any perceived deals.

iv. Retailers

The Retailers file contain available information about the retail chains from which products were purchased, such as the channel type.

v. Products

The Products file contains detailed product information for each UPC code, such as UPC description, brand description, multi pack and size. This files also contains Nielsen codes and descriptions for the Department, Product Group and Product Module that each UPC code is assigned to. The Products file contains only the “core” product attributes that are common for every UPC code. The file also contains what Nielsen calls Magnet products. Please see [Magnet data](#) (page [10](#)) for more detail and caveats on how to use these data.

vi. Product Extra Attributes

In addition to the “core” attributes, products also have “extra” attributes, reported in the products_extra_attributes files. Not every product has extra attributes for its UPC codes.

vii. Brand Variations

Brand Descriptions can change over time. This file captures all very minor changes that are observed in Brand Descriptions but were not deemed significant enough for us to create a new Brand Code and therefore to create a new version for a UPC code.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

III. Data Collection Methodologies

A. Sampling

The Nielsen Company uses a stratified, proportionate sample for the Homescan Consumer Panel. The design calls for the recruitment and maintenance of a panel of households that match a selected group of demographic characteristics (see below) at the total U.S., major market and remaining Census Region levels. The 100K sample is stratified into 61 geographic areas (52 major markets and nine remaining Census Divisions).

The sample is balanced on nine demographic characteristics to reflect the universe of household in the U.S.:

- Household Size (4 levels)
 - 1 member
 - 2 members
 - 3/4 members
 - 5+ members
- Household Income (6 levels)
 - <\$25k
 - \$24-\$34k
 - \$35k-\$49k
 - \$50k-\$69k
 - \$70k-\$99k
 - \$100+
- Household Head Age (3 levels)
 - <35
 - 35-54
 - 55+
- Female Head Education (4 levels)
 - HS Grad or Less
 - Some College
 - College Grad
 - No Female Head
- Male Head Education (4 levels)
 - HS Grad or Less
 - Some College

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

- College Grad
- No Male Head
- The presence of children
 - Yes
 - No
- Race (3 levels)
 - White
 - Black
 - Other
- Hispanic
 - Yes
 - No
- Household Head Occupation (3 levels)
 - White Collar
 - Blue Collar
 - Other

See Appendix II: Codebook, [Demographic Codes](#) (page 39) for additional details about each demographic category. In addition to demographic representation, the sample is also balanced on county size dispersion as well as key county population totals.

Households are randomly recruited to join the panel. Historically, the Nielsen Company has recruited panelists via direct mail, using a variety of targeted and general industry name lists. They send a brief overview letter to prospective panelists highlighting the information needed. For those returning the completed survey, Nielsen sends a detailed demographic questionnaire and a summary of what is required for membership.

In 2004, Nielsen instituted a new method of recruiting via Internet. A variety of online vendors are used to provide household email information and random site invitation recruitment, but the basic process remains the same. As Internet households do not represent total U.S. household demographics, there remains a portion of the panel population who are recruited through The Nielsen Company's traditional mail methodology in order to maintain a nationally representative sample.

On a weekly basis, Nielsen performs a sample maintenance process to evaluate overall and individual market sample demographic representation and replenishes the sample as needed.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Sampling rates for the 125,000 household panels includes:

- 52 major market sampling rate: approximately 940 households per 1MM people in the market
- Remaining U.S. Sampling rate: approximately 940 households per 1MM people in the market

B. Incentives

Nielsen offers households a variety of incentives to join and stay active. These are designed to be non-biasing; Nielsen does not provide account-specific coupons out of concern for potential impact to natural purchase selection of outlets and products. Among the incentives that Nielsen offers on a regular basis include:

- Monthly prize drawings
- Gift points awarded for weekly transmission of data
- Sweepstakes

Nielsen regularly tests different incentives for their correlation with retention rates.

Nielsen has ongoing communication with panelists to ensure cooperation, create enthusiasm, and monitor workload:

- Monthly Newsletter
- Q & A Section With Helpful Tips and Reminders
- Notice of Monthly Sweepstakes Winner
- Letter From the President
- Gift Point Statement
- Help Desk
- An 800 number for panelists to call
- Telecommunication
- After transmitting, panelists may receive personalized computerized tips and reminders
- Personalized letters for reporting problems and questions
- Exit Interviews

C. Panelist Retention

As a matter of course, Nielsen does not force drop cooperating panelists. It is necessary in many categories that have long purchase cycles (e.g. Health and Beauty) to ensure an adequate supply of long-term panelists. Additionally, having long-term panelists provides the ability to perform many analyses such as long-term consumer value or change in household composition.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Nielsen has a comprehensive program of dropping and replacing panelists that do not perform to minimum reporting standards.

The best metric of retention is to evaluate the proportion of panelists who are currently part of the sample who were also panelists at the same time last year. Currently, Nielsen retains about 80% of its active panel each year.

D. Validation and Quality Checks to Ensure Data Accuracy

To ensure quality data, The Nielsen Company has a number of systems and processes in place:

- A quarterly coverage monitor report is generated comparing projected Consumer Panel data to store-based scanning data. There is a quarterly review of the trip data associated with these categories.
- A weekly sample report is generated that tracks historical static or usable sample counts on a weekly, monthly, quarterly, and annual basis.
- A weekly analysis reviews sample demographic representation and county size dispersion for each major market and remaining U.S. sample segment.

E. Purchase Static

The goal of a purchase static is to filter out households who are poor reporters (who could skew results for certain types of analyses). The first criterion is that a household must be considered “active,” thus allowing that household to be included in report processing. Next, depending on household size, the household must transmit the minimum required spending dollars per four week period to be considered usable. All of the households in the Chicago Booth – Nielsen Data Center meet Nielsen’s 12-month static.

F. Projection System

It is extremely important not to treat all households equally. In Nielsen’s study design, some households are more likely to be selected according to their household characteristics compared to the population. The projection factors, which are sampling weights, must be used to correct for this sample bias. The sum of the weights is the total household population (i.e. number of households) in the U.S.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Projection factors for each year are based on Nielsen's demographic estimates using available data from quarter four of the prior year. (Note: please see FAQ section "Years Available and Panel Size" regarding the panel_year.) **Any researcher using the projection system for Magnet data should review the cautions below.**

There are two main objectives of this projection system:

1. The first objective is to better represent census estimates of demographic composition by re-weighting or balancing the raw (un-projected) panel. The standard demographic variables balanced are:

Household Size	Male Head Education
Household Income	Female Head Education
Head of Household Age	Head of Household Occupation
Race	Presence of Children
Hispanic Origin	Nielsen County Size

For each reporting period, the Homescan panel is projected to universe targets using the nine critical demographics used for sampling in each of the 61 U.S. geographies. In addition to demographic representation, the sample is also balanced on county size dispersion (A,B, C and D county population size) as well as key county population totals within a market (i.e. Cook county within the Chicago market). Demographics are updated for the universe and individual panelist on an annual basis. Population estimates are updated monthly.

A Linear Programming optimization routine is used to create a unique projection weight for each household. The factor solution is such that all geographic population and demographic constraints are met at all geographies with the smallest factor variance across households within each market.

2. The second objective is to project the raw panel to represent census estimates of absolute population size that enables the reporting of key household statistics such as dollar sales, unit sales, buying households, etc. for the geography reported.

Another important aspect of the projection system is to minimize the chance that too few households represent too many in the universe. The linear algebra mathematics used in the projection system can control on many population variables simultaneously. The resulting samples

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

therefore have more precise matches to true census population targets, thus producing the most accurate consumer measures in the industry.

Like any consumer panel, certain types of households are difficult to recruit and maintain. These include young, mobile singles, and certain very old and wealthy households. Nielsen's recruiting process and projection system is designed to ensure that the data provide the best possible representation of all demographic groups. With their movement to Internet-based recruiting, Nielsen has shown improvement among difficult-to-recruit young singles.

Researchers can use the projection system in the following ways, **noting how to properly use projection factors with Magnet data.** See [Departments](#) (page [51](#)) for department names.

IMPORTANT NOTE: CAUTIONS FOR USING PROJECTION SYSTEM WITH MAGNET DATA

Researchers who perform analysis on data from the Magnet product modules **should trim their sample to include only the Magnet households**, and use **only the Magnet projection factors**. Furthermore, they should use the magnet projection factors for **all** products included in their analysis (**not just for the magnet products**). It is **not** appropriate when conducting analyses with Magnet households to use the magnet projection factors for the Magnet products and regular projection factors for all other products. The magnet projection factors for Magnet households should be used for **all products** in these cases.

Similarly, if doing analyses with non-magnet households, researchers **should trim the products to remove all magnet products** from their analysis.

	When using "projection_factor"	When using "projection_factor_magnet"
Households to include	All households	Magnet households only (trim sample to exclude non-magnet households)
Products to include	Non-magnet products only (trim sample to exclude magnet products)	All products

See details below about level of projection for Magnet data and note that Magnet data should not be used to project to the local market level.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Using the field “projection_factor” in the PANELISTS file, researchers can project all data EXCEPT the Magnet data to:

- National Level
- Regional Level (using “region_code”)
- Scantrack markets that indicate they are “projectable” (by seeing which “scantrack_code” values in [Scantrack Market Codes](#) (page 44) are projectable)

Using the field “projection_factor_magnet” in the PANELISTS file, researchers can project the Magnet households and their purchases ONLY to:

- **National level:** see note below for important details.
The Nielsen Magnet sample does not use a projection system with local Scantrack markets. While it is reasonable to aggregate product purchases for Magnet households at a Scantrack market level using magnet projection factors, the variance associated with these projections will be much higher than the projections from all households of their non-magnet purchases within “reportable” Scantrack markets ([Scantrack Market Codes](#) (page 44) indicates which markets are projectable).

G. Decomposing Sales by Consumer Segment

Nielsen always recommends using projected data for sales decomposition.

When a subset of the sample is identified as a buying group based on the demographic characteristics of the sample households, the projected sales associated with this group must be viewed as a statistical estimate. Since the projection methodology incorporates only a selected set of demographic variables, many of the demographic segments will be subjected to some level of statistical variability. These sales measures represent the best possible statistical estimates that can be made for the segment given the constraints of a syndicated projection scheme. For example, the youngest head of household segment that is controlled for in the projection system is 18 to 34 years old. A client may want to look at consumer sales for the 18 to 24 versus the 25 to 34 segments. As these two segments are not controlled for explicitly, the reported sales information will have more statistical variability than the combined 18 to 34 segment would have.

H. Expected Sampling Variability

The sampling error associated with Homescan data is directly related to the number of unique households buying the reported product in the report period. Nielsen statisticians have run numerous

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

sampling error calculations across many categories and have generalized “typical” sampling error for an average brand or category to be as follows:

Homescan™ - Total U.S.		
Volume/Share Relative SE estimates - 95% confidence		
<u>Penetration (%)</u>	<u>Sales SE</u>	<u>Trend of sales SE*</u>
5	12.8%	9.9%
6	11.6%	9.1%
7	10.7%	8.5%
8	10.0%	7.9%
9	9.3%	7.5%
10	8.8%	7.1%
11	8.3%	6.7%
12	7.9%	6.4%
13	7.5%	6.1%
14	7.1%	5.8%
15	6.8%	5.6%
20	6.2%	5.2%
25	6.0%	5.0%
30	5.7%	4.8%
35	5.5%	4.7%
40	5.3%	4.5%
45	5.2%	4.4%
50	5.0%	4.2%
60	4.7%	4.0%
70	4.4%	3.8%
80	4.1%	3.6%
90	3.9%	3.4%
* period vs year ago		
Copyright © 2013, The Nielsen Company		

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

I. Price Determination

The price reported in Consumer Panel data is determined in two different ways, depending on whether the panelist shopped at a store where Nielsen receives point of sale (POS) data.

Panelists are instructed to scan all items purchased that have a bar code. After the scan, the panelist is asked to enter units. At this point, if the outlet shopped is a retailer for which Nielsen receives electronic point of sale (POS) data, the price prompt is bypassed. In this case, the price reported is an average weighted price for the item that week in that particular store. This is done to minimize the data entry burden of panelists. If the outlet shopped is not a retailer where Nielsen receives POS data, then the panelist is instructed to enter the price paid (prior to any coupons or deals).

Nielsen **does not** distinguish prices entered by the panelist from those that are imputed from Nielsen cooperating retailers, as they believe that all prices provided in the Consumer Panel data are accurate.

Finally, the panelist is prompted for Deal. If the panelist perceived a deal, then this is indicated. If the panelist used a coupon, they enter the amount discounted.

Frequent Shopper Card Discounts: If the panelist has a frequent shopper card and is aware that a particular item was discounted for card members, they can indicate this as a deal but are not required to do so.

IV. Detailed File Layouts

Note that fields ending in "uc" were generated by the University of Chicago's Kilts Center. These are not fields that were sent by Nielsen.

A. Panelists File

Number of Files: One file for each panel year. Each record contains information for one panelist's household. All demographic data is gathered from panelists during the fall of the year prior to the panel year.

<u>Variable Name</u>	<u>Format</u>	<u>Other Notes</u>
----------------------	---------------	--------------------

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

household_code (key)	Integer	Links to Purchases file
panel_year (key)	Integer (YYYY)	Indicates which panel year file to use for panelist data. See Note 2 for detailed information about the panel_year variable (which does not coincide with the calendar year).
projection_factor	Integer	Use to project all data (EXCEPT Magnet data) to a larger geographic area or to the entire U.S. See Projection System (page 15),
projection_factor_magnet	Integer	A non-null value indicates that the panelist is also part of the Magnet Panel. Use to project Magnet data to the entire U.S. See Data Collection Methodologies, Projection System (page 15) for more information. A null value indicates that the panelist is NOT part of the Magnet Panel.
Household and Head(s) Demographic Variables	All corresponding codes for these variables are available in Appendix II: Codebook, Demographic Codes (page 39).	
household_income	Integer	Range of total household income, estimated on annual basis. Year of household income is for <u>2 years prior to the panel year</u> . See Note 3 for additional details.
household_size	Integer	Number of individuals residing in home.
type_of_residence	Integer	Whether a house, apartment, condo, etc.
household_composition	Integer	Describes gender and general relationship of those living in the household. See the male_head_age and female_head_age variable notes for info on head of household.
age_and_presence_of_children	Integer	Whether there are children under the age of 18 in the household, and if so, their age range(s).
female_head_age	Integer	Value is 1-9 based on the age range selected by the panelist. "0" indicates no female head of household.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

		See note 4 for details about how head of household is identified.
male_head_age	Integer	Value is 1-9 based on the age range selected by the panelist. "0" indicates no male head of household. See note 4 for details about how head of household is identified.
male_head_employment	Integer	Approximate number of hours per week the male head is employed.
female_head_employment	Integer	Approximate number of hours per week the female head is employed.
male_head_education	Integer	Highest degree earned by male head.
female_head_education	Integer	Highest degree earned by female head.
male_head_occupation	Integer	See Note 4. Type of employment for male head. Zero indicates that there is no male head of the household.
female_head_occupation	Integer	See Note 4. If no male head, type of employment for female head. Zero indicates that there is no female head of the household.
male_head_birth	Text (7) (YYYY-MM)	
female_head_birth	Text (7) (YYYY-MM)	
marital_status	Integer	Marital status of household heads.
race	Integer	Represents racial identity of the household.
hispanic_origin	Integer	Represents whether members of the household are of Hispanic origin.
Geographic Variables	All corresponding codes for these variables can be found in Appendix II: Codebook, Geographic Codes (page 43).	
panelist_zip_code	Text (5)	Household's 5-digit zip code.
fips_state_code	Integer	State code for panelist household. These codes are assigned by the federal government for US states.
fips_state_descr	Text (2)	2-characters assigned to each state by the U.S. Postal Service.
fips_county_code	Integer	County code for panelist household. These codes are assigned by the federal government for US counties.
fips_county_descr	Text (50)	The county name.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

region_code	Integer	Code for household's regional location in the United States. See region codes in Appendix II: Codebook, Region Codes (page 43).
scantrack_market_code	Integer	See Appendix II: Codebook, Scantrack Market Codes (page 44). See map depicting scantrack areas in Appendix IV: Maps
scantrack_market_descr	Text (50)	See Appendix II: Codebook, Scantrack Market Codes (page 44). This is the name of the Scantrack Market.
dma_code	Integer	Indicates which of the Designated Market Area (DMA) of the panelist household. See map depicting DMA in Appendix IV: Maps
dma_descr	Text (100)	Indicates which of the Designated Market Area (DMA) of the panelist household.
Ownership and Usage Variables	All corresponding codes for these variables can be found in Appendix II, Ownership and Usage Codes (page 48).	
kitchen_appliances	Integer	Type of kitchen appliances reported as present in the home.
tv_items	Integer	Presence of TV and type of cable TV used by the household.
household_internet_connection	Integer	Whether Internet is active in household. 1= yes; 2= no
wic_indicator_current	Integer	Indicates that the household is receiving WIC assistance during the panel year. See Note 5 below.
wic_indicator_ever_notcurrent	Integer	Indicates that the household was receiving WIC assistance at some time in the past but not during the panel year. See Note 5 below.
Family Member Demographic Variables		
Member_1_Birth	Text (7) (YYYY-MM)	
Member_1_Relationship_Sex	Integer	
Member_1_Employment	Integer	Whether household member is employed. 1= employed; blank=unemployed/too young to work
Member_2_Birth	Text (7) (YYYY-MM)	

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Member_2_Relationship_Sex	Integer	
Member_2_Employment	Integer	Whether household member is employed. 1= employed; blank=unemployed/too young to work
Member_3_Birth	Text (7) (YYYY-MM)	
Member_3_Relationship_Sex	Integer	
Member_3_Employment	Integer	Whether household member is employed. 1= employed; blank=unemployed/too young to work
Member_4_Birth	Text (7) (YYYY-MM)	
Member_4_Relationship_Sex	Integer	
Member_4_Employment	Integer	Whether household member is employed. 1= employed; blank=unemployed/too young to work
Member_5_Birth	Text (7) (YYYY-MM)	
Member_5_Relationship_Sex	Integer	
Member_5_Employment	Integer	Whether household member is employed. 1= employed; blank=unemployed/too young to work
Member_6_Birth	Text (7) (YYYY-MM)	
Member_6_Relationship_Sex	Integer	
Member_6_Employment	Integer	Whether household member is employed. 1= employed; blank=unemployed/too young to work
Member_7_Birth	Text (7) (YYYY-MM)	
Member_7_Relationship_Sex	Integer	
Member_7_Employment	Integer	Whether household member is employed. 1= employed; blank=unemployed/too young to work
Note 1: Unless otherwise specified, all code values for demographics variables are found in Appendix II: Codebook, Demographic Codes (page 39).		
<p>Note 2: The panel_year variable denotes the Nielsen "data year" (which does not perfectly align with calendar year), and should be used to look up the appropriate Panelist record. For example, 2004 contains data for 2003-12-28 through 2004-12-25, so the panel_year variable on 2003-12-29 would equal 2004.</p> <p>The panel_year variable is not coincident with the calendar year. Nielsen starts collecting data on entering panelists on the first Sunday before the start of the new year, or if Sunday is 1 January, then that Sunday. This has several effects of which researchers should be aware.</p>		

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

- Panelists who exit the panel will have transactions/trips up until the last Saturday before the end of the year. They will have no data for the last several days of the month.
- New panelists entering the panel will have data for the last several days of the previous year, starting on that last Sunday.
- When merging demographic data to the trips and/or transactions files, panelists who remain in the data from one year to the next will have potentially conflicting demographic data in the month of December if that demographic information changes from one panel year to the next (e.g. changed income, projection factor, etc.)

Note 3: Household Income: Panelists are asked in the fall prior to the start of the panel year for their total annual income as of year-end of the **previous** calendar year. For example, panelists in the 2004 panel are surveyed in September of 2003 about their total annual income at the end of 2002.

Note 4: Identify head of household: when a panelist signs up for participation, she/he tell us whether there are 2 heads of household or 1 head of household. We then use that to code the Male Head and Female Head fields in the data. Those 2 heads of household could be married, single, related, etc. Researchers should use the Male Head and Female Head indicators/demos as the guide and use the Household Composition variable secondarily.

Occupation: Panelists select from high level occupational categories (blue collar, white collar, or service field), and then select a more specific job category within this. Nielsen aggregates responses into 12 categories as defined in Appendix II: Codebook, [Demographic Codes](#) (page 39).

Note 5: WIC Indicator variables.

Researchers should be cautious about using this field due to the following:

- The WIC Indicator field is collected in supplemental studies (not part of the annual demographic survey) starting in 2006. As a result, there is no WIC data for 2004-2005, which is indicated by a blank. For years 2006 and later, panelists are not required to answer this question and therefore missing data still exist. Therefore, 2006 forward, blank can mean either “no participant response” or that the panelist indicated the household is “not on WIC.”

For more detail on codes, see Appendix II: Codebook, [Ownership and Usage Codes](#) (page 48).

B. Trips File

Number of Files: One file for each panel year. Each record contains information for one shopping trip.

Variable	Format	Description	Other Notes
trip_code_uc (key)	Integer	Unique identifier for trip	Links to Purchases file
household_code	Integer	ID of household that made the purchase	Use with panel_year to link to Panelists file

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

purchase_date	Text (10) (YYYY-MM-DD)	Date of purchase	
retailer_code	Integer	ID code for the entire retail <u>chain</u> (not the individual store)	Links to Retailers file
store_code_uc	Integer	ID code for each <u>individual store</u>	This store_code_uc can be used to link consumer panel data with the retail scanner data. Note that a value of zero represents "missing."
panel_year	Integer (YYYY)	Panel file with panelist data	Indicates which panel year file to use for panelist data
store_zip3	Text (3)	First 3 digits of zip code of store where purchase was made	Note that some store zip codes are inferred based on panelist home address. Also, approximately half of the trips do not have a store zip code.
total_spent	Decimal (2)	Total expenditure for the trip, as entered by the consumer from the bottom of their receipt.	This field is capped at \$999.99 except for some Magnet-only trip records which exceed this cap. The sum of all purchase transactions for a trip likely will not equal this number (See FAQ document).
method_of_payment_cd	Text (2)	Method of payment used for this trip.	

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Note 1: The panel_year variable denotes the Nielsen "data year" (which does not perfectly align with calendar year), and should be used to look up the appropriate Panelist record. For more details, see Note 2 under Panelists file layout.

C. Purchases File

Number of Files: One file each panel year. Each record contains information for one UPC purchased during one Trip.

Note that, unlike other files, there are no key variables in the purchases file.

Variable	Format	Description	Other Notes
trip_code_uc	Integer	Unique identifier for trip	Links to Trips file
upc	Text (12)	UPC code of product – manufacturer assigned code.	Use with upc_ver_uc to link to Products file; Use with upc_ver_uc and panel_year to link to products_extra_attributes files
upc_ver_uc	Integer	UPC code version.	This field is generated by Univ. of Chicago to retain information related to earlier versions of a UPC code. See Note 1 below for detailed information about this field.
quantity	Integer	Number of units purchased	See Note 2 below.
total_price_paid	Decimal (2)	Total price paid for all units, <u>before</u> discounts	See Note 2 below.
coupon_value	Decimal (2)	Total discount for amount due to coupon	See Note 2 below.
deal_flag_uc	Text	Indicates presence of a deal (1=deal, 0=no deal)	

Note 1: Explanation of upc_ver_uc

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 others cases, just a few characteristics may have changed. Finally, Nielsen may have reclassified the product into a different module and none of the characteristics actually changed

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

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 Nielsen-generated fields (e.g. a UPC code may have been assigned to a new 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.

Note 2: To calculate the *final* price paid by panelist, subtract “coupon_value” from “total_price_paid.” To determine *per unit cost*, divide the calculated final price paid by quantity.

D. Retailers File

Number of Files: One master file for all years, updated annually. Each record contains information for one retail chain. **File Layout:**

Variable	Format	Description	Other Notes
retailer_code (key)	Integer	ID code for retailers	Links to Purchases file
channel_type	Text (30)	Description of retailer channel	Retail Channel Descriptions (page 50)

E. 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.

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

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

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.

See [Magnet data](#) (page [10](#)) for exceptions and notes about Magnet data product modules.

File Layout:

Variable	Format	Description	Other notes
upc (key)	Text (12)	UPC code of product	Same as in Transactions
upc_ver_uc (key)	Integer	UPC code version	See Note 1 under Purchases file layout for more details.
upc_descr	Text (50)	Product description	Abbreviated
product_module_code	Integer	Product module code	Nielsen assigned; modules represent detailed categories of products
product_module_descr	Text (100)	Product module description	A list of product module descriptions is included in the attached hierarchy document (Excel file).
product_group_code	Integer	Product group code	Nielsen assigned; Product Groups are groupings of Product Modules within the same Department; See Appendix II : Departments page 51
product_group_descr	Text (100)	Product group description	A list of product group descriptions is included in the attached hierarchy document (Excel file).

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

department_code	Integer	Department code	Nielsen assigned; see Appendix II : Departments page 51 or notes about department codes for Magnet data (page 10).
department_descr	Text (100)	Department description	See Appendix II : Departments page 51 Descriptions also listed in the attached hierarchy document (Excel file).
brand_code_uc	Integer	Brand code generated by University of Chicago.	Changes in Brand Description for a UPC will be captured in the Brand Variations file.
brand_descr	Text (30)	Brand description	
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.
size1_code_uc	Integer	Size code generated by University of Chicago.	A unique size1_code_uc is assigned for each unique combination of the “size1_amount” and “size1_units” fields.
size1_amount	Decimal (3)	Numeric size of the product	See Note 1 for further info. See Size1 Amount Description page 51 for details.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

size1_units	Text (6)	Description of the “unit of measure” for the size1_amount	See Note 1 for further info. See Size1 Amount Description page 51 for details.
dataset_found_uc	Text (3)	Indicates whether the product is found in consumer panel data, scanner data, or both.	Values are HMS (Homescan Consumer Panel dataset), RMS (Retail Scanner dataset), or ALL.
size1_change_flag_uc	Text (6)	This variable is relevant only to researchers who are linking the consumer panel and retail scanner datasets. It indicates a product size change occurred at some point in time in the scanner data. Please refer to the retail scanner data manual for more details.	

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Note 1: Every core attribute has a code (Integer) and a description (text), e.g. “brand_code_uc” and “brand_descr”. The multi field and size1 fields are the exceptions.

- Multi has a single numeric value without any description.
- There are two description fields for size1.
 - “size1_amount” is a number representing the numeric size
 - “size1_units” describes the unit of measure. For example, “size1_amount” might be “16.0”, and “size1_units” might be “OZ.”
 - For multi packs (i.e. “multi”>1), to determine **total units for a product**, multiply “multi” by “size1_amount”.

An illustration of the size1_code, size1_amount, and size1_units is below.

Six-Pack of Coca-Cola

upc: 004900000463

upc_ver_uc: 1

multi: 6

size1_code_uc: 2475

size1_amount: 12

size1_units: OZ

Single can of Coca-Cola

upc: 004900000463

upc_ver_uc: 2

multi: 1

size1_code_uc: 2475

size1_amount: 12

size1_units: OZ

F. Products Extra Attributes Files

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

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

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

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

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

Note to researchers regarding flavor data for 2010: It is a known issue that the flavor code and flavor description are missing from the 2010 extra attributes file. We have a file available with these missing data. Refer to the spreadsheet “Latest_Flavor_2010.csv” located in the reference documentation folder. This is an interim solution, as we will eventually rework the files to include the data.

(Extra attributes) Variable_Name	Format (Length)	Description	Other notes
upc	Text (12)	UPC code of product	Same as in Purchases
upc_ver_uc	Integer	UPC code version	See FAQ for more info
panel_year	Integer (4)	Panel file with panelist data	Same as in Purchases
flavor_code	Integer	Flavor code	
flavor_descr	Text (50)	Flavor description	
form_code	Integer	Form code	
form_descr	Text (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	Formula code	
formula_descr	Text (50)	Formula	e.g. ½ Less Fat, 0-6

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

(Extra attributes) Variable_Name	Format (Length)	Description	Other notes
		description	Month, 110 Calorie, Acne Fighting, Cold & Cough, Day or Night, Extra Moist, etc.
container_code	Integer	Container code	
container_descr	Text (50)	Container description	e.g. plastic, glass, or canned drinks
salt_content_code	Integer	Salt content code	
salt_content_descr	Text (50)	Salt content description	e.g. low salt, sea salt, no salt, etc.
style_code	Integer	Style code	
style_descr	Text (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	Text (50)	Type description	e.g. de-caffeinated; sugar-free; extra-virgin olive oil; soft-drink; etc.
product_code	Integer	Product code	
product_descr	Text (50)	Product description	e.g. Gravy, Mint Truffle Square, Sauce, etc.
variety_code	Integer	Variety code	
variety_descr	Text (50)	Variety description	e.g. Angel Hair, Beef, Blarney, etc.
organic_claim_code	Integer	Organic claimed by product code	
organic_claim_descr	Text (50)	Organic claimed	

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

(Extra attributes) Variable_Name	Format (Length)	Description	Other notes
		by product description	
usda_organic_seal_code	Integer	USDA Organic seal on product code	
usda_organic_seal_descr	Text (50)	USDA Organic seal on product description	
common_consumer_name_code	Integer	Commonly used name code	
common_consumer_name_descr	Text (50)	CCN description	Perch, shovel, toy, etc.
strength_code	Integer	Strength code	
strength_descr	Text (50)	Strength description	
scent_code	Integer	Scent code	
scent_descr	Text (50)	Scent description	
dosage_code	Integer	Dosage code	
dosage_descr	Text (50)	Dosage description	
gender_code	Integer	Gender code	
gender_descr	Text (50)	Gender description	
target_skin_condition_code	Integer	Skin condition targeted code	
target_skin_condition_descr	Text (50)	Skin condition targeted description	e.g. dry skin, cracked skin, etc.
use_code	Integer	Intended to be used by code	
use_descr	Text (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

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

(Extra attributes) Variable_Name	Format (Length)	Description	Other notes
			product
size2_amount	Decimal (3)	Amount in product size	A different amount measurement (compared to size1_amount) of a product
size2_units	Text (50)	Unit of amount of product	A different unit measurement (compared to size1_unit) of a product

G. Brand Variations File

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	Format	Description	Other Notes
brand_code_uc	integer		
brand_descr	Text (30)		
brand_descr_alternative	Text (30)		
start_date	Text (10) (YYYY-MM-DD)		
end_date	Text (10) (YYYY-MM-DD)		
dataset_found_uc	Text (3)	Indicates whether the product is found in panel data, scanner data, or both.	Values are HMS, RMS, or ALL.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

V. Appendix I: Policies

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

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

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

Researchers must also agree to the following Nielsen legal requirements:

- I will not share the data with any other individuals, other than registered PhD students who are individually registered and approved by the Kilts Center.
- I will not share the data with any student in an undergraduate or master's program without notifying the Kilts Center of plans to direct and accept full responsibility for the student's access in a classroom or independent study setting.
- I will ensure that data is stored only on devices and computers owned by the subscribed institution. I understand that insubstantial subsets of data required for particular analysis processes may be temporarily downloaded to such personal computers as required for analysis, but I will ensure that these data are purged as soon as possible.
- I will not disclose any personally identifiable information about any consumer panelists, or details that would permit re-identification of Nielsen's consumer panelists, without prior Nielsen approval.
- I will not disclose any information about specific retailers or individual stores, or details that would permit re-identification of Nielsen's retailers or individual stores, without prior Nielsen approval.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

- I will not disclose UPC-level pricing data that may be incorporated in the Nielsen data received from retailers who have agreed to provide this data to Nielsen.
- I will include Nielsen's copyright in any disclosed information and publications.
- Within fifteen (15) business days of expiration or termination of the agreement, I will destroy or purge all data (including any copies) in my possession or control, and provide written certification of the destruction and purging to the Kilts Center.

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

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

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

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

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

VI. Appendix II: Codebook

A. Panelist File Codes

i. Demographic Codes

	Code Value
Household Size	
Single Member	1
Two Members	2
Three Members	3
Four Members	4
Five Members	5
Six Members	6
Seven Members	7
Eight Members	8
Nine+ Members	9
Type of Residence - this variable captures all combinations of 2 questions: # of families, and Condo/Coop (yes/no).	
One Family House	1
One Family House (Condo/Coop)	2
Two Family	3
Two Family House (Condo/Coop)	4
Three+ Family House	5
Three+ Family House (Condo/Coop)	6
Mobile Home or Trailer	7
Unreported	Null
Household Income - the values represent ranges of total household income for the full year that is <u>2 years prior</u> to the Panel Year.	
Under \$5000	3
\$5000-\$7999	4
\$8000-\$9999	6
\$10,000-\$11,999	8
\$12,000-\$14,999	10
\$15,000-\$19,999	11
\$20,000-\$24,999	13
\$25,000-\$29,999	15
\$30,000-\$34,999	16
\$35,000-\$39,999	17

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

\$40,000-\$44,999	18
\$45,000-\$49,999	19
\$50,000-\$59,999	21
\$60,000-\$69,999	23
\$70,000-\$99,999	26
\$100,000 +	27 (Note: in 2004-2005, and again in 2010) "27" is the highest value and refers to anything \$100,000 and above
\$100,000 - \$124,999	27 (this value applies to this range ONLY in 2006-2009)
\$125,000 - \$149,999	28 (value only present 2006-2009)
\$150,000 - \$199,999	29 (value only present 2006-2009)
\$200,000 +	30 (value only present 2006-2009)
Household Composition	
Married	1
Female Head Living with Others Related	2
Male Head Living with Others Related	3
Female Living Alone	5
Female Living with Non-Related	6
Male Living Alone	7
Male Living with Non-Related	8
Age – Male & Female Head of Household	
Under 25 Years	1
25-29 Years	2
30-34 Years	3
35-39 Years	4
40-44 Years	5
45-49 Years	6
50-54 Years	7
55-64 Years	8
65+ Years	9
No Male/Female Head	0
Age and Presence of Children	
Under 6 only	1

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

6-12 only	2
13-17 only	3
Under 6 & 6-12	4
Under 6 & 13-17	5
6-12 & 13-17	6
Under 6 & 6-12 & 13-17	7
No Children Under 18	9
Male/Female Head Employment	
Under 30 hours	1
30-34 hours	2
35+ hours	3
Not Employed for Pay	9
No Male/Female Head	0
Male/Female Head Education	
Grade School	1
Some High School	2
Graduated High School	3
Some College	4
Graduated College	5
Post College Grad	6
No Male/Female Head or Unknown	0
Marital Status	
Married	1
Widowed	2
Divorced/Separated	3
Single	4
Unknown	Blank (currently, Marital Status is never blank)
Male/Female Head Occupation	
No male/female head of household	0
Economist/Mathematician/Scientist/Researcher Accountant/Auditor Architect/Engineer/Pilot Artist/Entertainer/Writer/Dancer/Composer/Athlete Business professional, non- managerial Computer Programmer/System Analyst/Data Processor Dentist/Doctor/Pharmacist/Physician's Assistant/Psychologist Educator/Lecturer/Librarian/Teacher/Coach	1

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Lawyer/Paralegal/Judge Medical Technician/Paramedic/RN/Therapist/Social Worker Religious/Clergy Member	
Administrator/Company/Officer/Manager/Supervisor Banker/Controller/Financial Analyst Builder/Contractor/Landscaper Owner of Business, Company, or Store Buyer/Purchasing Agent Public Official/Politician/Government Employee	2
Bank Teller/Bookkeeper/Cashier Clerk/Gas Attendant/Stock/Inventory Computer/Graphic Design Insurance Adjuster/Underwriter Mailroom/Messenger/Postal Worker Receptionist/Secretary/Typist/Data Entry	3
Sales – Industrial/Wholesale Sales – Insurance/Real Estate/Services Sales – Retail	4
Foreman Baker/Butcher/Seamstress/Tailor Carpenter/Electrician/Painter/Plumber/Exterminator Construction or Road Machine Operator Mechanic/Repairman Technician (except medical) Utility Lineman/Serviceman/Building Inspector	5
Factory Machine Operator Delivery/Route man Driver-Bus/Taxi/Truck Factory/Transportation Worker (Airline, Railroad, Cruise)	6
Member of Armed Forces	7
Barber/Beautician/Nail Technician/Salon/Makeup Artist Bartender/Chef/Food Service/Hotel Service Child Care Worker/Housekeep/Maid Dental Assistant/Practical Nurse/Dental Hygienist Fire fighter/ Police Officer/Sanitation/Security Officer Janitor/Porter	8
Farmer (Manager, Owner, Worker)	9
Students Employed <30 hours	10
Construction worker/Shipping and Receiving	11

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Fisherman/Gardener/Lumberman	
Housewife Retired Unable to work Unemployed/Laid off	12
Race	
White/Caucasian	1
Black/African American	2
Asian	3
Other	4
Hispanic Origin	
Yes	1
No	2
Member Relationship / Sex	
Son	1
Daughter	2
Other Male Relative	3
Other Female Relative	4
Male Not Related	5
Female Not Related	6
No Member or Unknown	Blank
Member Employment	
Employed	1
Not Employed or No Member	Blank

ii. **Geographic Codes**

FIPS State and County Codes

For more information on FIPS, see http://quickfacts.census.gov/qfd/meta/long_fips.htm

For a list of state FIPS codes, see <http://www.epa.gov/enviro/html/codes/state.html>

The FIPS and DMA codes map 1-to-1, using the dma_code variable in the panelist file.

See map depicting DMAs in [Appendix IV: Maps](#)

Region Codes

Codes indicate region of the country in which a household is located. These region codes map to the Census regions. See appendix IV: maps.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Region	Code Value	States included (fips code)
New England	1	Connecticut (09), Maine (23), Massachusetts (25), New Hampshire (33), Rhode Island (44), Vermont (50)
Middle Atlantic	2	New Jersey (34), New York (36), Pennsylvania (42)
East North Central	3	Indiana (18), Illinois (17), Michigan (26), Ohio (39), Wisconsin (55)
West North Central	4	Iowa (19), Kansas (20), Minnesota (27), Missouri (29), Nebraska (31), North Dakota (38), South Dakota (46)
South Atlantic	5	Delaware (10), District of Columbia (11), Florida (12), Georgia (13), Maryland (24), North Carolina (37), South Carolina (45), Virginia (51), West Virginia (54)
East South Central	6	Alabama (01), Kentucky (21), Mississippi (28), Tennessee (47)
West South Central	7	Arkansas (05), Louisiana (22), Oklahoma (40), Texas (48)
Mountain	8	Arizona (04), Colorado (08), Idaho (16), New Mexico (35), Montana (30), Utah (49), Nevada (32), Wyoming (56)
Pacific	9	Alaska (02), California (06), Hawaii (15), Oregon (41), Washington (53)

Scantrack Market Codes

The Scantrack market variable identifies the Nielsen-defined Scantrack markets and the years for which they are reportable (meaning that panelist_projection_factors can be used to estimate total spending for products for a Scantrack market). The variable is blank for the panelists who live in other markets.

See map depicting Scantrack areas in [Appendix IV: Maps](#)

IMPORTANT NOTE: Please review cautions for using [Projection System](#) (page 15) with Magnet data.

Nielsen's projection design is based on geographical and demographical marginal targets of the US household population. For UPC-based items, the geographical controls include selected major markets and as such these markets are considered by Nielsen as reportable. While every Nielsen panelist is assigned a projection weight, some households live in markets that Nielsen did not explicitly control on in the projection design for a given year due to sample size. These projection weights are based on higher geographical patricians known as remaining US markets. While it is reasonable

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

to aggregate projected household data into non-reportable markets, Nielsen does not recommend this technique as the variance associated with these estimates will be much larger than with reportable markets.

As noted in detailed notes about using the projection system with Magnet data, the Nielsen Magnet households do not use a projection system with local market targets. Again, while it is reasonable to aggregate projected Magnet households at a local market level, the variance associated with these markets will be much higher than the reportable UPC-based markets.

Scantrack Market Description		Reportable Markets								
		2004	2005	Jan-Jun 2006	Jul-06	2007	2008	2009	2010	2011-2013
001	Boston	X	X	X	X	X	X	X	X	X
002	Chicago	X	X	X	X	X	X	X	X	X
003	Houston	X	X	X	X	X	X	X	X	X
004	Indianapolis			X	X	X	X	X	X	X
005	Jacksonville				X	X				
006	Kansas City				X	X	X	X	X	X
007	Los Angeles	X	X	X	X	X	X	X	X	X
008	Suburban NY									
009	New York - used code or "urban NY"	X	X	X	X	X	X	X	X	X
010	Exurban NY									
011	Orlando				X	X	X	X	X	X
012	San Francisco	X	X	X	X	X	X	X	X	X
013	Seattle	X	X	X	X	X	X	X	X	X
014	Atlanta	X	X	X	X	X	X	X	X	X

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Scantrack Market Description		Reportable Markets								
		2004	2005	Jan-Jun 2006	Jul-06	2007	2008	2009	2010	2011-2013
015	Cincinnati			X	X	X	X	X	X	X
016	Cleveland			X	X	X	X	X	X	X
017	Dallas	X	X	X	X	X	X	X	X	X
018	Denver	X	X	X	X	X	X	X	X	X
019	Detroit	X	X	X	X	X	X	X	X	X
020	Miami	X	X	X	X	X	X	X	X	X
021	Milwaukee				X	X	X	X	X	X
022	Minneapolis	X	X	X	X	X	X	X	X	X
023	Nashville				X	X	X	X	X	X
024	Philadelphia	X	X	X	X	X	X	X	X	X
025	Pittsburgh			X	X	X	X	X	X	X
026	Portland			X	X	X	X	X	X	X
027	San Diego			X	X	X	X	X	X	X
028	St. Louis	X	X	X	X	X	X	X	X	X
029	Tampa	X	X	X	X	X	X	X	X	X
030	Baltimore*	X	X		X	X	X	X	X	X
031	Birmingham			X	X	X	X	X	X	X
032	Buffalo-Rochester	X	X	X	X	X	X	X	X	X
033	Hartford-New Haven				X	X	X	X	X	X
034	Little Rock				X	X				
035	Memphis				X	X	X	X	X	X
036	New Orleans-Mobile					X	X	X	X	X
037	Oklahoma City-Tulsa			X	X	X	X	X	X	X
038	Phoenix	X	X	X	X	X	X	X	X	X
039	Raleigh-Durham				X	X	X	X	X	X
040	Salt Lake City-Boise				X	X	X	X	X	X
041	Columbus	X	X	X	X	X	X	X	X	X

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Scantrack Market Description		Reportable Markets								
		2004	2005	Jan-Jun 2006	Jul-06	2007	2008	2009	2010	2011-2013
042	Washington, D.C.*	X	X		X	X	X	X	X	X
043	Albany				X	X				
044	Charlotte	X	X	X	X	X	X	X	X	X
045	Des Moines				X	X				
046	Grand Rapids				X	X	X	X	X	X
047	Louisville			X	X	X	X	X	X	X
048	Omaha				X	X				
049	Richmond				X	X	X	X	X	X
050	Sacramento	X	X	X	X	X	X	X	X	X
051	San Antonio	X	X	X	X	X	X	X	X	X
052	Syracuse				X	X				
053	Remaining Greenville									
054	Remaining Atlanta									
055	Remaining Boston									
056	Remaining Charlotte									
057	Remaining Denver									
058	Remaining Detroit									
059	Remaining Indianapolis									
060	Remaining Jacksonville									
061	Remaining Kansas City									
062	Remaining Memphis - Little Rock									
063	Remaining Milwaukee									
064	Remaining Minneapolis									
065	Remaining New Orleans - Mobile									
066	Remaining North California									

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Scantrack Market Description		Reportable Markets								
		2004	2005	Jan-Jun 2006	Jul-06	2007	2008	2009	2010	2011-2013
067	West Texas				X	X				
068	Remaining Omaha									
069	Remaining Philadelphia									
070	Remaining Pittsburgh									
071	Remaining Richmond - Norfolk									
072	Remaining West Texas									
073	Remaining Seattle - Portland									
074	Remaining St. Louis									
075	Remaining Las Angeles - collar									
076	Las Vegas				X	X				

*To project the total New York City market, researchers should use the aggregation of all three NY Scantrack codes: urban NY (009), exurban NY (010), and suburban NY (008)

**Note that Baltimore and Washington, D.C. were combine prior to 2007.

iii. **Ownership and Usage Codes**

Kitchen Appliances	
Microwave Only	1
Dishwasher Only	2
Garbage Disposal Only	3
Microwave & Dishwasher	4
Microwave & Garbage Disposal	5
Dishwasher & Garbage Disposal	6
Microwave, Dishwasher, & Garbage Disposal	7
None	8
Not Available	9

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

TV Items	
No Television	0 (currently, there are no occurrences of "0" in the data)
Television with No Cable	1
Regular Cable Only	2
Regular & Pay Cable	3
Unclassified	Blank
Household Internet Connection	
Yes	1
No	2
WIC Indicators (wic_indicator_current and wic_indicator_ever_notcurrent)	
NOTE: No data prior to 2005. Beginning in 2006, panelists are asked for WIC data but not required to provide it.	
Yes	1
No	Blank
N/A (missing)	Blank

B. Trips File Codes

<u>Method of Payment Codes</u>	<u>Definition</u>
00	Scanner does not collect Method of Payment
01	Cash
02	Check
03	Credit Card/American Express
04	Credit Card/Discover
05	Credit Card/Master Card
06	Credit Card/Visa
07	Credit Card/Other
08	Debit-ATM

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

09	Other Payment Type
----	--------------------

C. Retailers File Codes

i. Retail Channel Descriptions

A retail channel is the classification of the type of retail outlet from which products were purchased. Retail channels are classified differently than in the standard Consumer Panel data, which has stores classified into one or more of eight channel types. In the data provided, channels are classified into one of 66 mutually exclusive classifications.

Channel Types		
All Other Stores	Department Store	News/Book Store
Apparel Stores	Discount Store	Office Supplies Store
Athletic Footwear	Dollar Store	Online Shopping
Automotive Store	Drug Store	Optical Store
Bakery	Electronics Store	Party Supply Store
Barber/Salon	Fish Market	Pet Store
Beauty Supply Store	Free Sample/Gift	Pizzeria
Beverage Store	Fruit Stand	Pro Shop
Bodega	Garden Stores	Quick Serve Restaurants
Butcher	as Mini Mart	Music/CD Store
Camera Shop	Grocery	Restaurant
Candy Store	Hardware/Home Improvement	Service Station
Catalog Showroom	Health Food Store	Shoe Store
Cheese Stores	Home Delivery	Sporting Goods
Close Out Store	Home Furnishings	Stationery Store
Coffee Store/Gourmet Coffee	Home Inventory	Swap meet Flea Market
Computer Store	Hypermarket	TV/Home Shopping
Convenience Store	Kennel/Vet	Tobacco Store
Coop/Farm/Feed	Liquor Store	Toy Store
Craft Stores	Mail Order	Vending Machine
Dairy Store	Manufacturer Outlet	Video Store
Delicatessen	Military Store	Warehouse Club

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

D. Products File Codes

i. Departments

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 (Department 99 and Product Group 99 are for the Magnet data products associated with Product Module 0750. Note that starting in 2011, Magnet data is now also found in Product Modules 0445-0468, but these Product Modules are assigned to many different Product Groups and many different Departments).

ii. Size1_Amount Description

The size1_code field is the code for the size of a product, usually weight.
Size1_unit descriptions are as follows:

size1_units Code	size1_units Description
CF	Cubic Foot
CT	Count

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

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

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

VII. Appendix III: File Names and Sizes

1. Retailers: 1 master file (combining all panel years). If any changes occur, an additional file will be created <u>containing only changes</u> .	
Naming convention:	retailers.tsv
File Size:	Less than 100 KB

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:	~550 MB

3. Brand Variations: 1 master file (combining all panel years). An updated version of this file will be regenerated each year with annual updates.	
Naming convention:	brand_variations.tsv
File size:	Less than 50 KB

4. Panelists: 1 file for each panel year. One new file will be created for each additional panel year as data is released.	
Naming convention:	panelists_XXXX.tsv for XXXX=20XX
File sizes:	~10 MB or less each year

5. Trips: 1 file for each panel year. One new file will be created for each additional panel year as data is released.	
Naming convention:	trips_XXXX.tsv for XXXX=20XX
File sizes:	~ 600 MB or less each year

6. Purchases: 1 file for each panel year. One new file will be created for each additional panel year as data is released.	
Naming convention:	purchases_XXXX.tsv for XXXX=20XX
File sizes:	~ 3 GB or less each year

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

7. 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_XXXX.tsv for XXXX=20XX
File sizes:	~125 MB or less each year

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

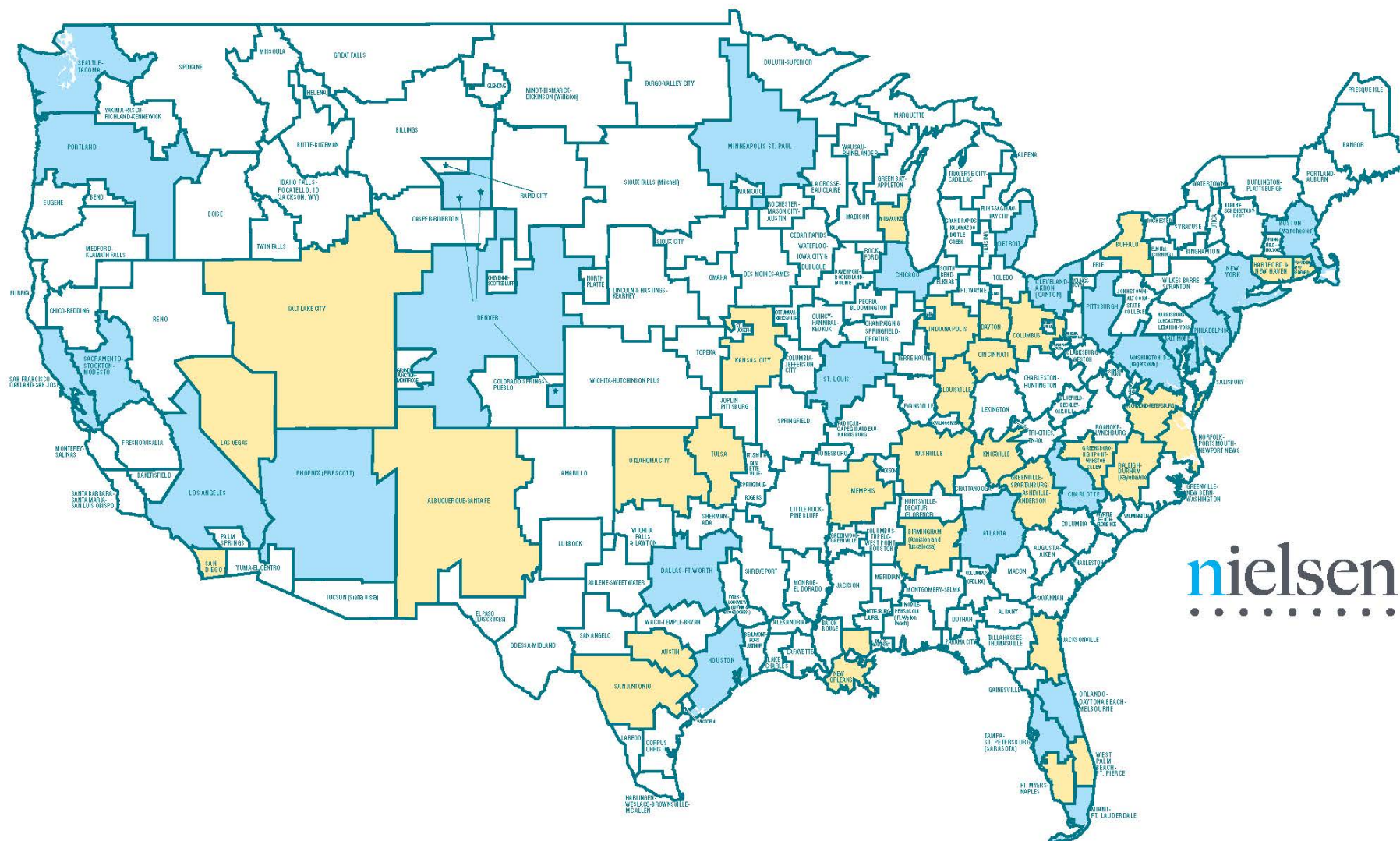
DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

VIII. Appendix IV: Maps

This page intentionally left blank.

Nielsen DMA—Designated Market Area

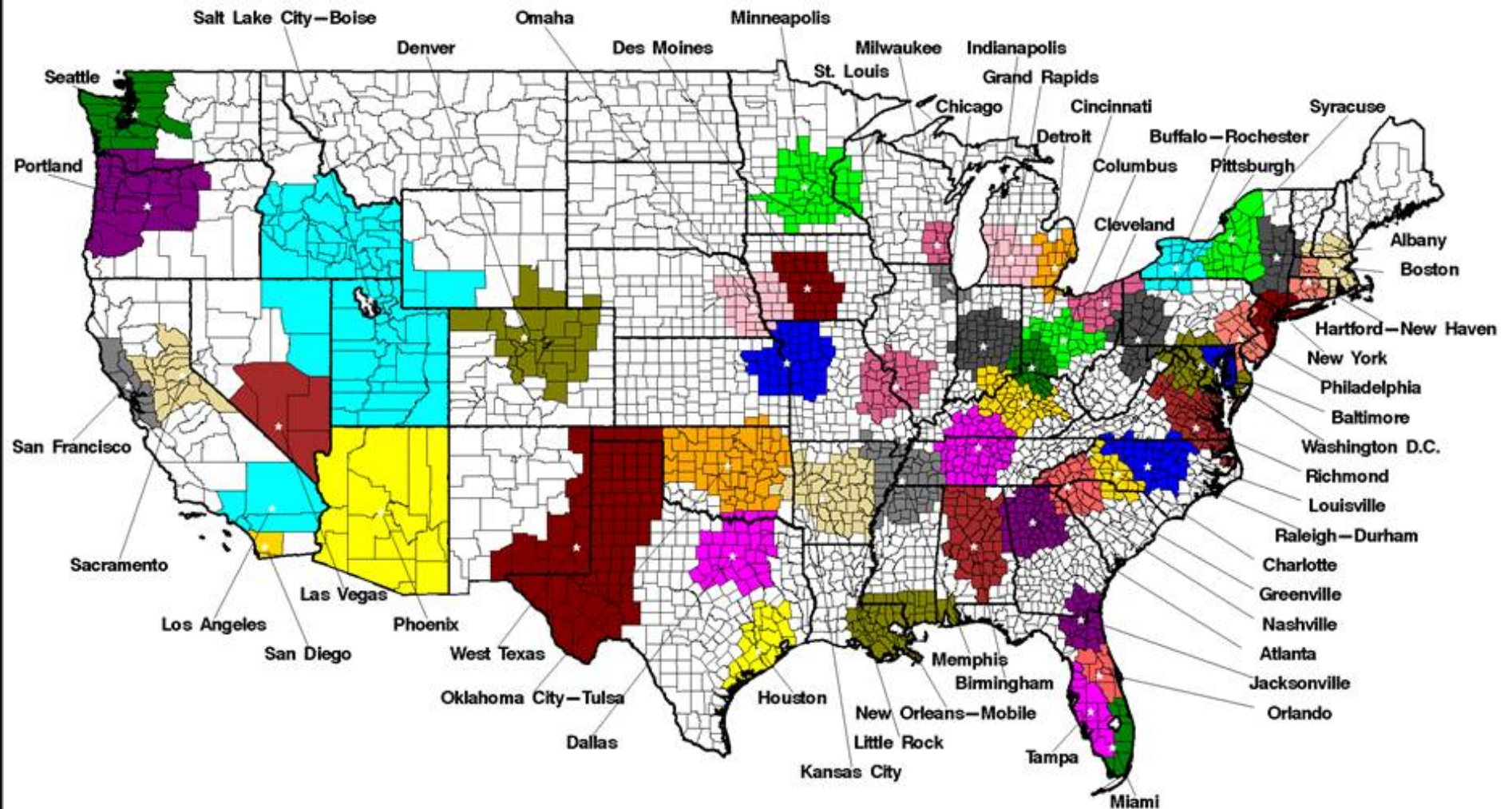


DMA® region: Generally a group of counties in which the commercial TV stations in the Metro/Central area achieve the largest audience share. This is non-overlapping geography for planning, buying and evaluating television audiences across various markets. See the Local Reference Supplement for more information.

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

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

NIELSEN TOTAL U.S. 52 SCANTRACK® MARKETS



IX. Appendix V: FAQs

A. General Questions

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. *Are there other fields that Nielsen has but have not been released as part of this dataset?*

- The consumer panel data (i.e. Homescan data) that we released includes ***all*** available variables and data fields that the Nielsen Company shares with academic researchers. There no additional fields available from Nielsen.

3. *Why do some variables have “_uc” attached to the end?*

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

4. *How many of the household members collect data?*

- Nielsen requests that all purchases made by all household members be recorded. Typically however, one member of each household scans data. This member is either the head of the household, or if the household has both a male and female head, then one primary shopper tends to collect data. The Nielsen scanner is often located in a central location in the home (such as the kitchen) to facilitate access by all household members.

5. *Does Nielsen have an estimate of what fraction of household consumption spending is represented by consumer panel data categories? Can they provide some context regarding how economically significant these data are as a percentage of total household consumption?*

- Nielsen estimates that approximately 30 percent of household consumption is accounted for by consumer panel data categories; however, they do not track other sources of consumer spending beyond the Nielsen-tracked categories. Nielsen suggests using data from the BLS looking at their NAICS codes, then comparing Nielsen-tracked categories vs. aggregate spending across all consumer categories.

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

B. Years Available and Panel Size

6. What is the “panel_year” variable, and when does the panel year begin?

- The panel_year variable denotes the Nielsen "data year" (which does not perfectly align with calendar year), and should be used to look up the appropriate Panelist record. For example, 2004 contains data for 2003-12-28 through 2004-12-25, so the panel_year variable on 2003-12-29 would equal 2004.
- The panel_year variable is not coincident with the calendar year. Nielsen starts collecting data on entering panelists on the first Sunday before the start of the new year, or if Sunday is 1 Jan, that Sunday. This has several effects of which researchers should be aware.
 - Panelists who exit the panel will have transactions/trips up until the last Saturday before the end of the year. They will have no data for the last several days of the month.
 - New panelists entering the panel will have data for the last several days of the previous year, starting on that last Sunday.
 - When merging demographic data to the trips and/or transactions files, panelists who remain in the data from one year to the next will have potentially conflicting demographic data in the month of December if that demographic information changes from one panel year to the next (e.g. changed income, projection factor, etc.)."
 - Projection factors change for most households when the panel year begins. Thus, data aggregations making use of the projection factors must take this into account.

C. Panelist Demographic and Geographic Data

7. Do the data include all households in the Nielsen panel?

- The data include all households that meet the Nielsen reporting period static. Those that do not meet the Nielsen static (i.e. the minimum reporting requirement) are not included.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

8. *What requirements does Nielsen have for households to meet the static?*

- A household must first be considered “active” by Nielsen. Next the household must transmit the minimum required spending dollars per four week period, depending on the household size, to be considered eligible for the static. All of the households in the data meet Nielsen’s 12-month static requirements for each corresponding calendar year of reporting.

9. *When are the demographics recorded?*

- Surveys are sent to existing panelists in Q4 of each year and implemented in the first week of January the following year. New panelist’s demos are based on survey responses at the time of recruitment. The demographics in a given reporting year can be thought to correspond closest to the demographics of the latter months (October-December) of the prior year.
- For household income, panelists are asked to report their combined total household annual income as of year-end of the previous calendar year. For example, panelists reporting in the 2004 data are surveyed in the Fall of 2003 about their total annual income at the end of 2002. Nielsen believes panelists are actually reporting their “annualized” estimated income as of the time of the survey and not referring to previous year tax returns. As such, in 2011 the instructions for income no longer mention previous calendar year, but rather “annualized” income.

10. *Does the household_code change over time?*

- No, this is a unique key assigned to each household. Households may drop in and out of the panel, but their household_code remains constant.

11. *Why do some of the demographic variables change over time?*

- There are some anomalies observed in the demographic reporting. For example, in some cases, the birthdates for both household heads change simultaneously. In other cases, race is reported differently from year to year. Other than some basic validations aimed at reducing data entry errors, Nielsen reports the demographic data exactly as the panelist entered it. Changes could be due to a panelist correcting previous

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

information or deciding to select a different option from a previous survey. Nielsen will use the demos as reported by panelists for the data period being measured. Due to the sensitivity surrounding personal information, Nielsen does not ask panelists to “validate” any updates they make to reported demos over time.

12. The code values for the field “Type of Residence” are not intuitively obvious. Is there more explanation for what this variable represents?

- As part of the demographic survey, Nielsen asks panelists two different questions: 1) what type of residence do you live in (one family, two family, 3+ family, mobile/trailer home), and 2) do you live in a condominium or co-op (Yes/No variable). The table represents the full range of options when combining responses to both questions.

One Family House	1
One Family House (Condo/Coop)	2
Two Family	3
Two Family House (Condo/Coop)	4
Three+ Family House	5
Three+ Family House (Condo/Coop)	6
Mobile Home or Trailer	7

13. Why don’t all of the years have data that includes panelists coded with a household income level of \$125,000 or higher?

- In 2004 and 2005, the highest household income range option on the demographic survey was \$100,000-\$124,999, which was coded as “27.” In years 2006-2009, Nielsen added three more options for household income range (values 28-30). The highest income range in the data from 2006 and later is \$200,000 or more, which is coded as “30.” In 2010, Nielsen went back to collapsing the household income ranges, so the highest level is \$125,000 and higher (value 27).

14. Is there any way to identify “Retired” vs. “Unemployed”?

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

- The occupation code value “12” contains both “retired” and “unemployed.” Nielsen suggests using the “Age” field to help distinguish retired vs. unemployed.

15. Does Nielsen capture in the panelist demographic survey information on home ownership?

- These data fields are not provided to Nielsen’s clients, including the Kilts Center.

16. Why is there no WIC data for 2004-2005?

- a. The WIC Indicator field is collected in supplemental studies (not part of the annual demographic survey) starting in 2006. As a result, there is no WIC data for 2004-2005, which is indicated by a blank. For years 2006 and later, panelists are not required to answer this question and therefore missing data still exist. In 2006 forward, the values are either “1” (yes, panelist reported WIC) or “blank” (not on WIC, or no response).

Note that panelists update their WIC information quarterly, and Nielsen provides the Kilts Center an update annually. This update contains one value for the year rather than a quarterly report.

17. What is the projection_factor_magnet variable, and how is it different from the projection_factor variable?

- Products include all Nielsen-tracked categories of food and non-food items, across all retail outlets in all U.S. markets. Nielsen samples all states and major markets. Panelists are geographically dispersed and demographically balanced. Each panelist is assigned a projection factor, which enables purchases to be projectable to the entire United States, regions within the US, and some Scantrack markets. The variable projection_factor is used for all data except when including Magnet product purchases in an analysis. Refer to projection system section of the data manual for important notes about instructions on this variable related to Magnet data.

18. Why is the Scantrack market variable empty for some panelists?

- The Scantrack market variable only contains Scantrack Major Markets, so the variable is blank for the panelists who live in the remaining US Scantrack regions.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

- Importantly, refer to projection system section of the data manual for notes regarding the projection system for Magnet data vs. non-Magnet data. The Magnet data can only be projected to the National level, not to the Scantrack market level.

D. Trips and Purchases

19. What standard is used for the UPC variable? Why can't I merge another UPC dataset with the UPC's found in the Consumer Panel 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 Section VIII: [Other](#) (page 75) for more details about steps for merging consumer panel UPC codes with external datasets.

20. 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 Nielsen, 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 Nielsen-generated fields (e.g. a UPC code may have been assigned to a new Product Module Code).

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

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.

21. How is the price field determined?

- The price reported in Homescan data is determined in two different ways, depending on whether the panelist shopped at a store where Nielsen receives point of sale (POS) data.

Panelists are instructed to scan all items purchased that have a bar code. After the scan, the panelist is asked to enter units. At this point, if the outlet shopped is a retailer for which Nielsen receives electronic point of sale (POS) data, the price prompt is bypassed.

In this case, the price reported is an average weighted price for the item that week in that particular store. This is done to minimize the data entry burden of panelists. If the outlet shopped is not a retailer where Nielsen receives POS data, then the panelist is instructed to enter the price paid (prior to any coupons or deals).

Nielsen **does not** distinguish prices entered by the panelist from those that are imputed from Nielsen cooperating retailers, as they believe that all prices provided in the Homescan data are accurate.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

22. I see an observation with total_price_paid=0. What does this mean?

- There are two reasons that the total price paid field could be zero:
 1. Zero could be the price that the consumer entered and Nielsen does not have any historical information regarding a valid price for the item.
 2. Zero is the recorded price from the retailer (e.g. a valid price based on a “free goods” promotion).

Nielsen suggests that if the transaction was recorded as a deal with the deal_flag_uc variable (and therefore the deal_type variable), researchers should use it. If there was no deal, consider the \$0 to mean that the price was “Not Available”.

23. The variables total_price_paid and coupon_value are included, but how do you determine the final price paid by the panelist for an item?

- To calculate the final price paid by panelist, subtract “coupon_value” from “total_price_paid.” To determine per unit cost, divide the calculated final price paid by quantity.

24. What is the process that panelists use to record deals?

- In general the sequence of data entry is:
 1. Information related to the overall trip (e.g. the date and store -- note that different scanners use different methods for helping the consumer select the store)
 2. Individual UPCs purchased:
 - a. Price - entered only if the product was purchased at a store that doesn’t provide Nielsen with POS data. Otherwise, the price is filled in by Nielsen.
 - b. Number of units purchased
 - c. If the items was perceived to be on “deal”, then this is indicated by the deal_flag_uc variable.
 - d. If the panelist used a coupon, the total discount amount due to coupon is recorded.

25. How is the store location determined for a shopping trip?

- Nielsen has deployed multiple scanning devices. Some scanning devices have store lists with addresses, while with others the consumer selects only the Retailer Name

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

for the store where they shopped and then Nielsen determines the most likely address for that store (which is the store nearest their home with that Retailer Name). This is to reduce panelist data entry burden.

26. Why do we see, beginning in 2007 and continuing in 2008-2009, a significant drop in percentage of transactions that have a deal?

- Kilts-Nielsen Data Center researchers found that prior to 2007, the deal percentage is about 30 percent. Starting in 2007, it drops to about 24 percent. With the expansion in the panel sample in 2007, a new scanner device was deployed that has a different interface for capturing deals. Among households with this scanner device, Nielsen also finds this same percentage decrease in deals starting in 2007. A new capture method with a new set of households is the root source of the decrease.

27. Why does the sum of total purchase transactions from a trip (computed from the PURCHASES files) differ from the total_spent amount from the TRIPS file?

- Think of total_spent as the amount at the bottom of the receipt. There are several reasons why the sum of all purchase transactions likely won't add to this number:
 - The panelist didn't scan all products purchased. Some items never make it into the home to get scanned. Consider items purchased at a hardware store that might get stored in the garage rather than being brought into the home, or a candy bar that was purchased and eaten before the consumer got home.
 - The total_spent amount, in a small proportion of cases, equals zero. These are valid trips.
 - Some items don't have a UPC code (e.g. Magnet product data is not entered by non-Magnet panelists)
 - There may have been a scanning error – e.g. the UPC code was torn.
 - Some items aren't "coded" by Nielsen – Nielsen mostly tracks fast-moving consumer goods (e.g. not most apparel, electronics or home furnishings, etc.).
 - For most products, individual purchase transactions don't include taxes, but the total_spent would include taxes.
 - Imputed price on an item (when the panelist doesn't enter price, but rather the price is imputed from the store's scanner data sent to Nielsen) may not exactly equal the price spent by the consumer.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

- Total_spent is capped at \$999.99 for all non-magnet data, however, there are some magnet-only trip records for which the calculated total does exceed this cap.

28. Why are there cases where adding up the spending for all purchase transactions in a trip exceeds the value of total_spent?

- This occurs on some trips that contain both UPC transactions and Magnet transactions. Nielsen separately edits/validates the sum of UPC transactions and the sum of Magnet transactions versus total_spent:
 - The sum of all purchase transactions on a trip with UPC codes should not exceed total_spent. If this sum does exceed total_spent, then Nielsen edits total_spent to be the sum of spending for transactions with UPC codes.
 - The sum of all purchase transactions on a trip without UPC codes (i.e. Magnet product transactions) should also not exceed total_spent.
- However, when summing up both the UPC and Magnet transactions for a trip, the sum can exceed total_spent.

29. Why are all trips that contain only Magnet product transactions missing store_code_uc, store_zip, and total_spent?

- For trips with only Magnet transactions, Nielsen doesn't store these fields. However, researchers could use the retailer_code field to make inferences for that panelist regarding which store they may have purchased from (e.g. the most frequently shopped store by that panelist that has the same retailer_code). Also, researchers can sum the purchase transaction amounts to generate total_spent.

30. Some panelists have very few transactions in a year. Is it possible this could be the case and that these panelists still made the "static"?

- Yes. It could be that they have purchased items but that these items are not coded by Nielsen. Their shopping trips are recorded, and you can look at the total_spent variable to see how much these panelists have spent even on trips where no purchase transactions with Nielsen-tracked UPC codes were recorded.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

31. Why do some trips not have a store_code_uc reported?

- This field is only filled in for those retail stores where Nielsen also receives (currently or in the past) point-of-sale scanner data from the store.

32. Why are some transactions repeated in multiple product modules?

- These duplicates are the result of how the raw data was originally pulled by Nielsen. Nielsen uses the most recent module coding in the data they provide. Over time, UPCs can be re-coded based on new information about a product's characteristics. There are some rare cases in which a UPC is assigned to 2 different modules at the same time based on its characteristics.

The Kilts-Nielsen Data Center was able to eliminate the large majority of such duplicates. However ~.02 percent of the 2004-2008 transactions are product module duplicates. The University of Chicago team determined that eliminating these remaining product module duplicates would be extremely difficult. Beginning in 2010, the product module duplicate transactions should not be present.

The approximate number of product module duplicate transactions for each year is listed below.

- 2004: 5,800
- 2005: 7,300
- 2006: 8,400
- 2007: 8,700
- 2008: 13,000

33. Why is there an increase in unique non-zero values for the store_code_uc variable, beginning in 2010 data?

- In 2010, Nielsen changed their filtering technique for delivering data on individual stores. This resulted in a higher availability in information about individual stores starting in 2010. The primary result of this is that more trips can be located to a particular store instead of just a retailer.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

Further, when looking at a given geographic unit in the trips files, a general increase in the number of stores is to be expected. This is not an economically interesting phenomenon but an artifact of this change in the data. Researchers need to account for this change when analyzing such trends. An alternative would also be to use the relevant geographic unit from the panelist files instead of the store files.

34. Why are there trips without any purchases?

- Thanks for noting this discrepancy. We will add this to the documentation as well. To explain: there is a percentage of trip record that do not have purchases. This can be true for two reasons. 1) It's possible that the panelist did not purchase any Nielsen-coded items during their trip. 2) Many of these cases are the result of non-magnet households purchasing only magnet products. In this scenario, the panelist would record the trip and total spending, but the individual purchases (all magnet products) would not be recorded because they are not a magnet household.

E. Retailers

35. Are store zip codes exact? Why are there so many missing observations in this variable?

- Depending on the type of scanner the panelist uses, the store zip code variable may be either the zip code of the exact store shopped based on the store entered by a panelist, or it may be the zip code associated with a store that is closest to the panelist's home for a retail chain entered. Nielsen finds that most trips are made to the closest store of a preferred chain.
- Zip code is only captured by Nielsen for a subset of all stores where panelists have shopped.

36. How are specific store locations determined for a shopping trip?

- Nielsen deploys multiple scanning devices. Some scanning devices have store lists with addresses. For others, the consumer selects only the Retailer Name for the store where they shopped, and Nielsen determines the most likely address for that store

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

(e.g. identify the retailer store name that is nearest panelist's home address). This is to reduce panelist data entry burden.

37. Why are the last two digits of the zip codes suppressed in the data?

- In order to maintain retailer confidentiality, Nielsen only allows the release of the first three digits of the zip code.

38. In what channel are mass merchandiser stores such as Kmart, Target, and Wal-Mart coded?

- Nielsen codes all mass merchandisers, including Kmart, Target, and Wal-Mart in the "Discount Store" channel.

39. Why do I see an increase in the number of unique stores in 2010? Why do I see an increase in the number of stores in a given zip code/dma/other geographic area in 2010?

- In 2010, Nielsen changed their filtering technique for delivering data on individual stores. This resulted in a higher availability in information about individual stores starting in 2010. The primary result of this is that more trips can be located to a particular store instead of just a retailer. Further, when looking at given geographic unit in the trips files, a general increase in the number of stores is to be expected. This is not an economically interesting phenomenon but an artifact of this change in the data. Researchers need to account for this change when analyzing such trends. An alternative would also be to use the relevant geographic unit from the panelist files instead of the store files.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

F. Products and Product Extra Attributes

40. Given that there is a 45-70% match in retailers between the scanner and panel data, why don't we see more UPCs matched across the two datasets?

- i. In the scanner data, the UPCs for private label products are masked. The UPCs are NOT masked in panel data, thus these products cannot be matched between the two datasets. Note that Nielsen codes all of the private label brands as "Control Brand" – specifically, the brand description is "CTL BR" (or a similar variation).
- ii. Panel data covers purchases across a much wider range of stores than scanner data. In the scanner data, we have ~37,000 stores. However, panelists in Homescan record purchases from a wider range of retailers.
- iii. We have two years of panel data (2004 and 2005) that do not appear in the scanner data. Those UPCs that were only available in 2004-2005 won't be found in the RMS data.
- iv. Homescan panelists may not buy all of the products available at Nielsen-participating retailers in the RMS data. Panelists are a sample of the overall U.S., so some of the products that appear in grocery, drug and mass merchandiser stores (perhaps very low volume/niche products) don't end up getting purchased by a panelist.

41. What is Magnet data, and how is it different than the other data?

- The Magnet data include products that do not use standard UPC codes -- items such as fruits, vegetables, meats, and in-store baked-goods. Only some of the panelist households report these types of purchases. Magnet data is not available prior to 2007.

Because only a subset of households report Magnet product purchases, researchers must note that we outline several cautions for how to use the projection system when analyzing data from these households. These are explained in detail in the data manual.

The way that Magnet purchases are coded (product module codes and department codes) changes beginning in 2011. See data manual for details.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

42. Why do some attribute code values in the Product Extra Attribute files appear to have multiple descriptions associated with them?

- The unique description for any attribute code in the Product Extra Attributes file is determined by using the attribute code value, the product_module_code, and the panel_year.

43. Why do some of the Product Extra Attributes file code values not have a corresponding description?

- This was due to a timing issue related to the fact that Nielsen pulled some of the older years of data at different times. When a description field is blank, there is no description available because that code value is no longer being used.

44. Why are some UPCs not found in the Transactions file?

- The data received for some panelists and transactions were not valid, however, the product characteristics from these transactions were retained. As a result, there are some UPCs found in the Products File for which there are no Transaction file records.

45. Why do some UPCs in the Products file not match to the Product Extra Attributes files?

- In the Product Extra Attributes files, only UPCs that a) have extra attributes and b) were purchased in that year, are included in the corresponding year's Product Extra Attributes file. Thus not all UPCs in the core file will have corresponding records in the Product Extra Attributes file.

46. There are a large number of observations with missing value for some variable x in the Product Extra Attributes files. Why is this?

- Not every variable is applicable to every item (e.g. flavor does not apply to plastic bags).

47. Why are there multiple Size fields in the product files?

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

- The size1_code_uc field is the code generated by University of Chicago for the size of a product.

There are two separate description fields: size1_amount and size1_units. The value of size1_amount is a number representing the numeric size, and the value of size1_units describes the unit of measure. For example, size1_amount might be "16.0", and size1_units might be "OZ."

- For multipacks (i.e. "multi">1), to determine **total units for a product**, multiply "multi" by "size1_amount".
- See the data manual for a list of all size units and corresponding codes.

48. Does the list of product modules in the data stay the same over time?

- No. Some modules drop out of the data, and some are added to reflect marketplace changes in product consumption patterns.
- In addition, a very small number of product module codes are occasionally "reused", meaning that Nielsen renames and reuses them for a totally different type of products. Finally, Nielsen occasionally changes the description for a small number of product modules while the overall type of products assigned to the module code remain the same. A list of changes in product module descriptions will be included in the Consumer Panel Data documentation manual beginning with 2014 panel data.
- Longitudinal comparisons should be done with this in mind.

49. Why do I find different product module codes for Magnet data?

- In data from years 2004-2010, all Magnet products were coded product module code 0750. However, beginning in 2011, Magnet products are coded in modules 0445-0468 in addition to 0750.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

- Importantly, product module 0750 is assigned to Product Group Code “99” and Department Code “99”. Product modules 0445-0468 are assigned to many different Product Group Codes and many different Department Codes.

50. 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.

51. 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 Nielsen.

52. Why are some purchases missing product group or department code for modules under 1000 (except for Magnet data)?

- For some modules (those under 1000, other than Magnet data), there is purchase transaction data but no product group or department value. The primary reason for the missing group and/or department is that these are combo packs with products from different modules. For example, a combo pack of toothpaste and a toothbrush is not assigned a module. In other cases, the products are simply not yet classified into product groups or departments.

53. Why do I find modules 7141, 7732, 8602 (starting in 2011) with very few transactions and no retail scanner movement data?

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

These modules are still active. There are only a few products coded in those modules (23, 17 and 68, respectively) and they are being sold in markets where retail scanner data is not collected.

54. 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 Nielsen 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 Nielsen.

55. Some product brand descriptions include abbreviations at the end of the field. What do these mean?

- These abbreviations do not have a meaning. Researchers should ignore those additional lettering.

G. Other

56. Is it possible to generate the check digit to add to the consumer panel 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, researchers 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.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

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) $111 \bmod 10 \equiv 1$
- 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.

57. 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.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

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

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

X. Appendix VI: Version revisions

A. Revisions 2016-1-22

1. Known issue regarding extra attribute “flavor” for year 2010. Data previously missing is now available in separate file located in the reference documentation folder. See note in section [Products Extra Attributes Files](#) for details.
2. All revisions to the hierarchy file (included in reference documentation files) are highlighted in that document. Note the following changes to modules.
 - a. Three modules added. Only module 7807 has data in 2014. Nielsen has just begun coding products with these data. Future years are expected to 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

- b. 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.**
 - c. Added column J to indicate modules that may not contain data ANY years in any dataset.
3. Added FAQ regarding brand_descr abbreviations. See FAQ 55 on page [75](#).
4. Added FAQ regarding trips that have no purchases. See FAQ 34 on page [69](#).

B. Revisions 2015-05-25

1. Starting in 2013, the Trips file now has a method_of_payment_cd field indicating the method of payment used for each trip.
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 Consumer Panel dataset when that was not the case.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

C. Revisions 2014-12-23

1. Note regarding how Nielsen identifies head of household, of which there can be two. See note 4 in file layouts for panelist file. See [Panelists File](#), page [20](#).

D. Revisions 2014-09-22

1. Merged FAQ document (used to be separate) into Consumer Panel Dataset Manual.
3. Added information about deferred modules. See description of deferred modules in section on [Products](#), starting on page [9](#).
4. Note regarding article submissions and data. See [Special notes regarding data](#), page [5](#).
5. Notice of change to 4 UPCs in the new product file released with 2012 data. See [Special notes regarding data](#), page [5](#).
6. Added state breakdown detail to geographic region codes. See [Geographic Codes](#), page [43](#).
7. Added note to FAQ #40 regarding how Nielsen codes private label brands. See page [71](#).
8. Added FAQ about why there are modules 7141, 7732, 8602 (starting in 2011) with very few transactions in the panel data and no retail scanner movement data? See FAQ #53, page [74](#).

E. Revisions 2014-04-02

1. Addition of section [Special notes regarding data](#). See page [5](#).
2. Added additional detail to occupation code descriptions. See [Demographic Codes](#) (page [39](#))

F. 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.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

2. Addition of DMA and Scantrack maps in Appendix IV.

G. Revisions 2013-10-21

General notes and updates

1. Kilts data users forum has a new URL. Please update bookmarks to <https://kiltsforum.chicagobooth.edu/forum/>
2. Addition of product hierarchy document, which is included in the reference files you received with the data.
3. Revised file layouts and code updates. See Detailed File Layouts and Appendix II: Codebook for details.
4. Some variable names were updated to achieve consistent naming convention across data sets. The changes are listed below.

File Type	Original variable name	New variable name
Panelist	household_id	household_code
	fips_county_code_descr	fips_county_descr
	scantrack_market_identifier_code	scantrack_market_code
	scantrack_market_identifier_descr	scantrack_market_descr
Trips	trip_id_uc	trip_code_uc
	household_id	household_code
	retailer_id	retailer_code
	store_id_uc	store_code_uc
Purchases	trip_id_uc	trip_code_uc
Retailers	retailer_id	retailer_code
Products	brand_code	brand_code_uc
	size1_code	size1_code_uc
	(did not exist in past versions)	dataset_found_uc
	(did not exist in past versions)	size1_change_flag_uc

5. Addition of the Brand Variations file.
6. Additional documentation details provided about the sampling methodology, projection system, decomposing sales by consumer segment, and expected sampling variability.

Panelist file updates

7. Updates to the WIC indicator.
 - a. WIC indicator value updated from “hyphen” to “blank” for no data in 2006 forward.
 - b. After 2006, the values are either “1” (yes, panelist reported WIC) or “blank” (not on WIC, or no response.)

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

9. In the panelist files, we added two variables to indicate designated market area (DMA): dma_code and dma_descr. Prior to this release, we had provided a separate Excel document mapping FIPS to DMA codes.
10. In the geographic codes, the Scantrack Markets chart is updated to contain the “Remaining” markets on the list (e.g. “Remaining Denver”, etc.)
11. The region_code is now reported using further sub-divisions. Find new region names and codes in the codebook section.
12. In the Panelist files, the value for fips_state_descr changed for Connecticut from CN to CT.
13. Analysts found anomalies in the Member_Birth_Date fields relative to the “Age_and_Presence_of_Children” field values. We therefore created new Panelist extract files for 2004-2009 with corrected Member Birth Date values.
14. Updated head of household variables. We found that Nielsen always codes male_head_occupation as non-zero number even in cases when there is no male head of the household (signified by male_head_age=0 and male_head_birth_month and male_head_birth_year being empty). We applied a simple fix that zeroes out male_head_occupation when no male head is present in a household.
15. In 2004-2006 Panelist files, Dade county was merged with Miami to create Miami-Dade County. The values changed from “12025” to the newer one, “12086.” The description also changed to reflect this merger.
16. All county names have changed. Previously, the source of the names was the government standards body that keeps track. Instead, we are now using Nielsen's spelling. Changes include (1) all names are now all upper case, (2) 'County' and other designations are no longer included in the name, and (3) other spelling changes.

Trips file updates

17. Store_code_uc (formerly named Store_id_uc) now has a field value of zero if it is missing. In the past, it was blank.
18. Addition of FAQ about increase in unique non-zero values for the store_code_uc, beginning in 2010 data. Refer to revisions listed in FAQ document.
19. Added FAQ about why some transactions are repeated in multiple product modules. Refer to revisions listed in FAQ document.

Purchases file updates

20. Variable deal_type was dropped from the purchases files. Nielsen no longer provides deal_type, so we eliminated the variable. Instead, researchers should use the variables deal_flag_uc and coupon_value to determine whether there was any type of deal on the purchase. If coupon_value and deal_flag_uc are both zero, there was no deal on the purchase.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

21. Variable total_uc was dropped from the purchases files. In the past, for duplicate purchase transactions we reported only 1 transaction record and incremented the “total_uc” field. We now provide each observation separately, making the total_uc variable unnecessary.

Retailers file updates

No changes or updates.

Products file updates

22. In prior versions of the products files, the product_module and the attribute code value were both needed to determine a code’s meaning. We eliminated this issue by creating a 1-to-1 mapping between brand_code_uc and brand_descr, and also between size1_code_uc and size1_amount plus size1_units.
23. Added FAQ about missing product group or department code for modules under 1000 (except for Magnet data). Refer to revisions listed in FAQ document.
24. Added to the product file a variable (dataset_found_uc) indicating which Nielsen datasets (panel, scanner, or both) in which the product is included.
25. Changes to the codes for size1_units variable. We have aggregated into a consistent abbreviation.

size 1_units field		
Consolidated New Code	Old Codes	Description
CF	CCCF, GRCF	Cubic Foot
CT	CGCT, CTCT, COUNT, SHSH, GRCT	Count
LI	CILITER	Liter
FT	CMFT, FTFT	Foot
YD	CMYD	Yard
EXP	CTEXP	Expired
PO	GRLB, GRPO, GRPOUND, POUN	Pound
OZ	GROZ, MLOZ, LIQUID OUN	Ounce
QT	GRQT, MLQU, MLQUART	Quart
ML	MLLI, MLLITER, MLM, MLML	Milliliter
SQ FT	SFSQ FT	Square Feet

26. Add FAQ about why more UPCs don’t match across panel and scanner datasets. See FAQ #[40](#) , page [71](#).

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

H. Revisions 2013-09-19

1. Add FAQ about why some transactions are repeated in multiple product modules. See FAQ #32, page 68.
2. Add FAQ about why there is an increase in unique non-zero values for the store_code_uc variable, beginning in 2010 data. See FAQ#33, page 68.
3. Add FAQ about why there is an increase in the number of unique stores in 2010. See FAQ #39, page 70.
4. Add FAQ about why there are different product module codes for Magnet data. See FAQ #49, page 73.
5. Add FAQ about why some purchases are missing product group or department code for modules under 1000 (except for Magnet data). See FAQ #52, page 74.

I. Revisions 2012-12-28

1. new fips_to_dma file. For details of the changes, see:
<https://kiltsnielsen.chicagobooth.edu/forum/viewtopic.php?f=2&t=32>

In the previous file, some counties were mapped to more than one DMA code. The explanation is that Nielsen primarily uses DMA codes for their media data and therefore defines them by OTA (over the air) TV signal strength. By this definition, some counties are split based on competing TV signals from different markets. The signal breakdowns are not relevant to the consumer panel data use, so we worked with Nielsen to clean up the mapping and eliminate the split counties. Now counties are mapped to where the majority of the county falls into the DMA. Therefore, in this file, no counties are split between DMA's.

J. Revisions 2012-11-09

1. Correction to WIC indicator.
 - a. This indicator was originally marked as integer variable but was corrected to be a text variable. The values did not change.
2. Kilts-Nielsen online users forum URL is updated. User forum moved. Please update bookmarks.

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

DO NOT DISTRIBUTE, SHARE OR POST THIS DOCUMENT

K. Revisions 2012-06-22

1. Additional clarification of how to use projection factors. See [Projection System](#) on page [15](#).
2. Clarification of the availability of magnet data - magnet data are available beginning in 2007.
3. Note that attribute information occasionally missing from size1_amount/unit and brand_descr fields.
 - b. There are a small number of cases where a brand_code value exists but brand_description is missing, and/or a size1_code value exists but size1_units and size1_amount are missing. These attribute descriptions are not available from Nielsen.
4. Added FAQ about “panel_year” variable. What is the “panel_year” variable, and when does the panel year begin? See FAQ #[6](#), page [59](#), last bullet point.
5. Added FAQ about why some modules show a noticeable change in the number of total purchases from year to year. See FAQ#[50](#), page [74](#).
6. Added FAQ about why some attribute information is missing from the brand and size fields in the products file. See FAQ #[51](#), page [74](#).

L. Revisions 2012-05-14

1. Addition of Product Extra Attributes files and documentation about the files.
2. Corrections to geography-related data in the panel files.
3. Added FAQ #[10](#) “What standard is used for the UPC variable? Why can't I merge another UPC dataset with the UPC's found in the Consumer Panel Data?” See page [81](#).

M. 2012-03-28

1. Original public release of data
2. Note: data released without Products Extra Attributes files.

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