

Comscore Web Behavior Database

Updated: November 2018

Comscore is a trusted partner for planning, transacting, and evaluating media across platforms. With transformative data science and vast audience insights across digital, linear TV, over-the-top (OTT), and theatrical viewership, we are a powerful third-party source for reliable measurement of cross-platform audiences.

The Comscore data available through WRDS represents a sample of US internet users' internet browsing behavior, purchases, and demographics. Specifically, it is a subset of the opt-in panel data used to inform Media Metrix—Comscore's premier audience measurement product.

This data is made available to WRDS subscribers for academic research only and is subject to Comscore's data usage policy, which can be found here:

<https://www.comscore.com/Insights/Data-Usage-Policy>

Data Support:

Keith Wine | kwine@comscore.com | (212) 497-1763
academicsupport@comscore.com

Sales Support:

Megan Leonhard | mleonhard@comscore.com | (650) 244-5489
Brandon Hubbard | bhubbard@comscore.com | (212) 497-1767

Demographics

Demographics are self-reported when a user enrolls a machine in Comscore's panel. It is possible for multiple impaneled machines to exist in one household. Values in this table are numeric and map to the lookup table provided at the end of this document.

| Column Name | Data Type | Description |
|--------------------|-----------|-----------------------------------------------------------|
| machine_id | bigint | unique machine identifier |
| hoh_most_education | int | highest level of education completed by head of household |
| census_region | int | household's US geographic region |
| household_size | int | number of people in household |
| hoh_oldest_age | int | age of head of household |
| household_income | int | combined income of household |
| children | int | whether children present in home (child_present in 2002) |
| racial_background | int | race of head of household |
| connection_speed | int | whether household has broadband (measured > 1 mbps) |
| country_of_origin | int | Hispanic self-identification |
| zip_code | char(5) | US postal code, stored as text to preserve leading zeros |

Transactions

Where transactions are identified within the sessions data, a page-scraping agent collects information about and categorizes the purchase. This agent is maintained for approximately 150 of the largest ecommerce retailers in the United States. As with any scraping agent, it does not work perfectly, so unusual data can often be attributed to miscollection. Product category ID lookups are provided in a separate spreadsheet.

| Column Name | Data Type | Description |
|------------------|-----------------|---------------------------------------------------------|
| machine_id | bigint | unique machine identifier |
| site_session_id | unsigned bigint | unique identifier for user's browsing session on a site |
| prod_category_id | int | unique identifier for category of product purchased |
| prod_name | varchar(500) | name of product purchased |
| prod_qty | int | number of product purchased |
| prod_totprice | numeric(12,2) | total price of product |
| basket_tot | numeric(12,2) | total price of all products purchased |
| event_date | varchar(8) | yyyymmdd |
| event_time | varchar(8) | hh:mm:ss (UTC) |
| domain_id | unsigned bigint | unique identifier for domain where product purchased |

Sessions

Sessions data measures panelists' web browsing. Browsing is segmented into browsing sessions and site sessions. There may be many site sessions within each overall browsing session.

| Column Name | Data Type | Description |
|-----------------|-----------------|---------------------------------------------------------|
| machine_id | bigint | unique machine identifier |
| site_session_id | unsigned bigint | unique identifier for user's browsing session on a site |
| user_session_id | unsigned bigint | unique identifier for user's overall browsing session |
| domain_id | unsigned bigint | unique identifier for domain |
| pages_viewed | int | number of pages viewed in browsing session |
| duration | int | number of minutes spent in browsing sessions |
| event_date | varchar(8) | yyyymmdd |
| event_time | varchar(8) | hh:mm:ss (UTC) |
| ref_domain_name | char(64) | the domain that sent the user to the current one |

Domain Lookup

| Column Name | Data Type | Description |
|-------------|-----------------|------------------------------|
| domain_id | unsigned bigint | unique identifier for domain |
| domain_name | varchar(100) | domain visited |

Demographic Lookup

| Column Values | Description | | Description |
|--------------------|---------------------------------|--------------------|-----------------------|
| hoh_most_education | | household_size | |
| 0 | Less than a high school diploma | 1 | 1 person |
| 1 | High school diploma or GED | 2 | 2 people |
| 2 | Some college but no degree | 3 | 3 people |
| 3 | Associate degree | 4 | 4 people |
| 4 | Bachelor's degree | 5 | 5 / 5+ people (02–13) |
| 5 | Graduate degree | 6 (2014 – present) | 6+ people |
| 99/3001/3002/3005 | Unknown | 99 | Unknown |
| | | | |
| census_region | | racial_background | |
| 1 | North East | 1 | Caucasian |
| 2 | North Central | 2 | African American |
| 3 | South | 3 | Asian |
| 4 | West | 5 | Other |
| 88/99 | Unknown | -88/99 | Unknown |
| | | | |
| hoh_oldest_age | | household_income | |
| 1 | 18-20 | 2002 – 2013: | |
| 2 | 21-24 | 1 | Less than \$15,000 |
| 3 | 25-29 | 2 | \$15,000 – \$24,999 |
| 4 | 30-34 | 3 | \$25,000 – \$34,999 |
| 5 | 35-39 | 4 | \$35,000 – \$49,999 |
| 6 | 40-44 | 5 | \$50,000 – \$74,999 |
| 7 | 45-49 | 6 | \$75,000 – \$99,999 |
| 8 | 50-54 | 7 | \$100,000+ |
| 9 | 55-59 | 2014 – Present: | |
| 10 | 60-64 | 11 | Less than \$25,000 |
| 11 | 65 and over | 12 | \$25,000 – \$39,999 |
| 99 | Unknown | 13 | \$40,000 – \$59,999 |
| | | 14 | \$60,000 – \$74,999 |
| children | | 15 | \$75,000 – \$99,999 |
| 0 | No | 16 | \$100,000 – \$149,999 |
| 1 | Yes | 17 | \$150,000 – \$199,999 |
| | | 18 | \$200,000+ |
| country_of_origin | | 99 | Unknown |
| 1 | Hispanic | | |
| 0 | Non-Hispanic | | |
| | | | |
| connection_speed | | | |
| 0 | Not broadband | | |
| 1 | Broadband | | |