Social Data Collaboratory at NORC

New NORC Study Reveals Social Media Users' Exposure to Food and Beverage Marketing

This survey was conducted by NORC at the University of Chicago's Social Data Collaboratory with the AmeriSpeak Omnibus team and funding from NORC.
Interviews: 08/08-12/2024
1,137 adults

Margin of sampling error: +/- 3.8 percentage points at the 95% confidence level among all adults

NOTE: All results show percentages among all respondents, unless otherwise labeled.

Study Methodology

This study was conducted by NORC's Social Data Collaboratory in collaboration with the AmeriSpeak® Omnibus team and was supported by a NORC internal investment.

Recruitment

Data were collected using the AmeriSpeak® Omnibus, a bi-monthly multi-client survey using NORC's probability-based panel designed to be representative of the U.S. household population. The survey was part of a larger study that included questions about other topics not included in this report. During the initial recruitment phase of the panel, randomly selected U.S. households were sampled with a known, non-zero probability of selection from the NORC National Sample Frame and then contacted by U.S. mail, email, telephone, and field interviewers (face-to-face). The panel provides sample coverage of approximately 97 percent of the U.S. household population. Those excluded from the sample include people with P.O. Box-only addresses, some addresses not listed in the USPS Delivery Sequence File, and some newly constructed dwellings.

Online surveys were conducted between August 8-12, 2024, with adults ages 18 and over representing the 50 states and the District of Columbia. Panel members were randomly drawn from AmeriSpeak, and 1,137 completed the survey. Respondents were offered a small monetary incentive for completing the survey. The final stage completion rate is 17.0 percent, the weighted household panel recruitment rate is 26.2 percent, and the weighted household panel retention rate is 77.0 percent, for a weighted cumulative response rate of 3.4 percent.

The overall margin of sampling error is +/-3.8 percentage points at the 95 percent confidence level, including the design effect. The margin of sampling error may be higher for subgroups.

Sampling error is only one of many potential sources of error and there may be other unmeasured error in this or any other survey.

Quality assurance checks were conducted to ensure data quality. In total, 11 interviews were removed for nonresponse to at least 50 percent of the questions asked of them, for completing the survey in less than one-third the median interview time for the full sample, or for straight-lining all grid questions asked of them. These interviews were excluded from the data file prior to weighting.

Once the sample was selected and fielded and all the study data were collected and made final, a poststratification process was used to adjust for any survey nonresponse as well as any noncoverage or under and oversampling resulting from the study-specific sample design.

Poststratification variables included age, gender, census division, race/ethnicity, and education. Weighting variables were obtained from the 2024 Current Population Survey. The weighted data reflect the U.S. population of adults ages 18 and over.

Additional information on the AmeriSpeak Panel methodology is available at: https://amerispeak.norc.org/about-amerispeak/Pages/Panel-Design.aspx.

Eligibility for Data Donation

All respondents were asked:

"Do you have a social media account for any of the following platforms?" Responses are summarized in table 1:

Table 1. Distribution of account holders across social media platforms

Platform	N	%
Facebook	797	70.1
Instagram	504	44.3
X (formerly Twitter)	221	19.4
TikTok	279	24.5
YouTube	499	43.9
Reddit	164	14.4
Snapchat	231	20.3
LinkedIn	312	27.4
Pinterest	259	22.8
Threads	62	5.5
Other[1]	15	1.3
I do not have an account	122	10.7

[1]The "Other" includes Nextdoor (3), Truth Social (3), WhatsApp (2), etc.

Respondents who reported they had either Facebook, Instagram, or both (n=869/76.4%) were considered eligible to donate data. Eligible respondents were then shown a screen that explained that NORC is interested in learning more about the effects of social media on society and asking respondents to share their own social media data with NORC for research purposes. A detailed description of the steps involved in data donation was presented. In addition, respondents were assured that "All social media data shared with NORC will be stored securely, and personally identifiable information (PII) will be removed. Any results from analysis will be reported in aggregate, with no PII disclosed."

Opt-in and Consent to Data Donation

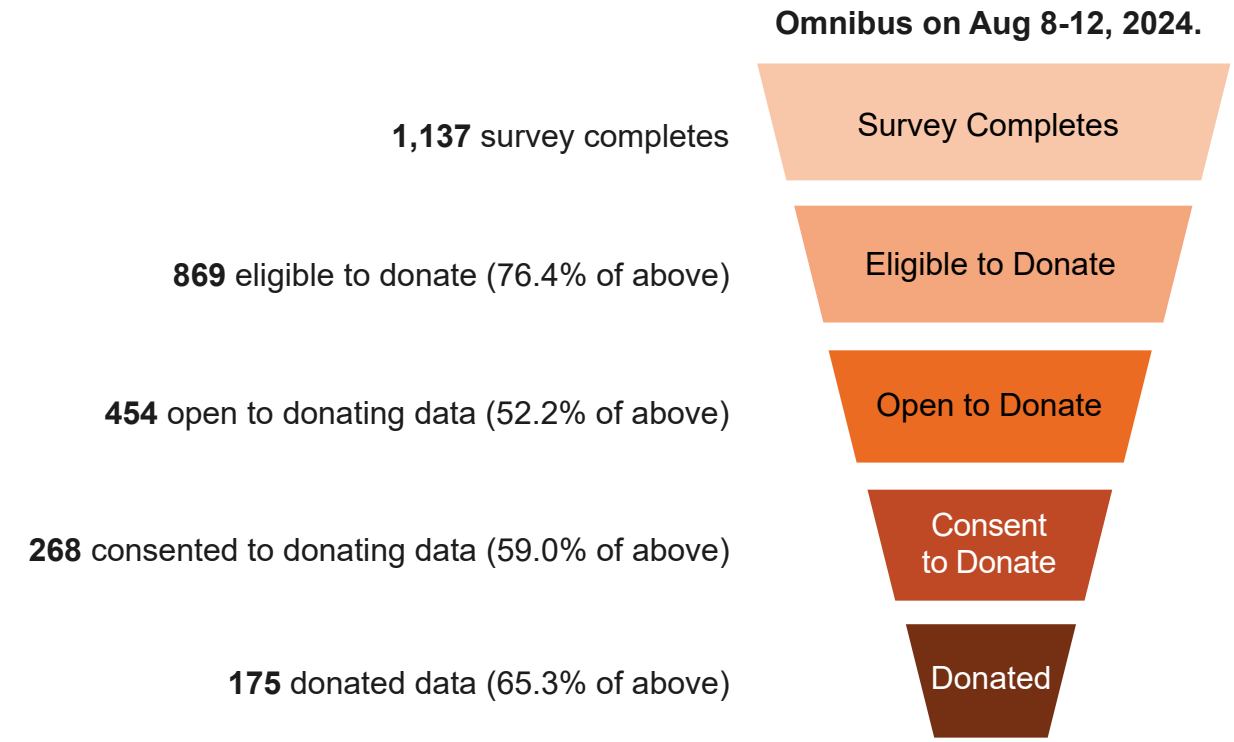
After seeing the screen with the description of data donation, all eligible respondents were asked:

"Would you consider sharing or 'donating' your social data from any platform to NORC to use for research, if provided adequate compensation?" Response categories included Yes, Maybe, and No.

Immediately following this question, all eligible respondents were informed that NORC would provide an additional incentive with a value equivalent to $25 per platform (Facebook and Instagram) as a thank you for their time and contribution of their data. They were then asked a second time, "Are you willing to donate your social media data to NORC?"

Slightly over half of eligible respondents (n=454/52.2%) stated they would be willing to donate their data. Willing respondents were next directed to an informed consent module. Of those who were willing to donate, nearly 60% (n=267/59.0%) provided informed consent. Of those who consented, nearly two-thirds retrieved their social media data from their Instagram and/or Facebook accounts and successfully uploaded those files to NORC's secure AmeriSpeak web portal. Figure 1 represents the flowchart showing eligibility, consent, and data donation.

Figure 1. Data donation process



Source: NORC at the University of Chicago

Evaluation of Selection Bias

Because the social media data were obtained within the framework of a nationally representative survey, it is possible to assess whether the respondents who contributed their social media data are statistically comparable to those who did not participate in data donation. Table 2 provides a summary of demographic characteristics of respondents who proceeded through each stage of the data donation process compared to those who did not.

Table 2. Demographic characteristics of respondents who proceeded through the data donation process

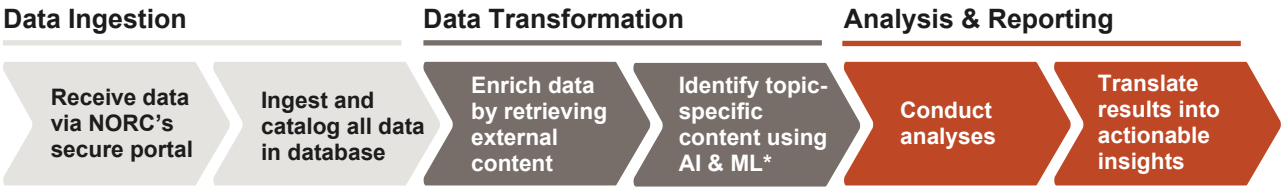
Characteristics		Eligible	Willing	Consented	Donated
Gender	Male	45.7	50.9	50.7	50.3
	Female	54.3	49.1	49.3	49.7
Age	18-29	12.7	16.7	19.4	20.0
	30-39	19.7	24.0	25.4	28.6
	40-49	19.6	21.2	20.9	24.6
	50 or older	48.1	38.1	34.3	26.9
Race/ethnicity	Non-Hispanic White	65.8	60.1	59.7	61.1
	Non-Hispanic Black	11.4	12.3	16.0	12.0
	Other	22.8	27.5	24.3	26.9
Education	High school graduate or lower	22.7	26.9	28.0	24.6
	Some college	41.2	41.9	36.9	38.3
	Bachelor's	22.0	17.6	21.3	23.4
	Postgraduate	14.2	13.7	13.8	13.7
Household income	<$30,000	20.5	26.2	25.7	24.6
	$30,000-$59,999	25.3	24.4	25.4	25.1
	$60,000-$99,999	25.7	24.9	24.3	25.1
	≥$100,000	28.5	24.4	24.6	25.1
Region	Northeast	12.5	12.6	12.3	11.4
	Midwest	25.4	27.1	25.7	25.7
	South	36.8	34.8	36.6	38.3
	West	25.2	25.6	25.4	24.6
Living in Metropolitan	Non-Metro Area	17.4	17.6	17.5	16.6
	Metro Area	82.6	82.4	82.5	83.4
Housing	Owned	66.9	61.2	57.1	54.9
	Rented/other	33.1	38.8	42.9	45.1

Social Media Data Analysis

The social media data provided by the platforms to respondents arrived in raw JSON or HTML format in a zip folder consisting of several files. The process of creating measures from this raw data involved several steps. Figure 2 summarizes our data ingestion, transformation, and analysis pipeline.

Figure 2. Data process pipeline: ingestion, transformation, analysis & reporting

Our resource-intensive workflow first catalogs interactions from complex data archives, enabling complete download and AI analysis of every associated URL.



*—AI is artificial intelligence and ML is machine learning

Source: NORC at the University of Chicago

First, the data are ingested into a database, in which every interaction, search, and app they used are sorted and cataloged. Files ingested into the database include data on time-stamped views of paid advertisements, along with videos and posts viewed, respondent-specific ad targeting, searches conducted, purchases, and other consumer product/media account logins using Meta account credentials. For the current project, we analyzed the Facebook ads viewed by each respondent.

The files we receive typically contain a mix of text data, including web addresses pointing to specific content viewed. Next, we enrich the data by retrieving the content of linked web addresses. To transform the raw data into measures of advertising exposure, our content classification process used a multi-stage pipeline combining both account-level and content-level analysis to identify food and beverage products and brands, fast- and full-service restaurants, food delivery services and meal-kits, and food stores within the corpus of the ads viewed.

Figure 3 illustrates our mixed-method approach for classifying advertising content. Ads were classified based on advertiser or account-level information and content-level information. Brand-related keywords in account profile descriptions and usernames were flagged at the account level, while a large language model (LLM) classified accounts based on metadata and profile information. At the content level, posts were flagged using extracted brand-related terms from post information (such as Optical Character Recognition and detected objects), followed by initial zero-shot classification using all available content-level data. Posts with low confidence were reclassified using refined prompts and potentially reclassified again for predicted positives and a random sample, incorporating all prior information. The final classification for each post fused these intermediate results with subject matter expertise to ensure validity and accuracy.

Figure 3. Illustration of topic-specific classification

The final classification for each post fused intermediate classifications at different levels with subject matter expertise.

Source: NORC at the University of Chicago

Measurement Creation

For each respondent, we created an average daily exposure measure to specific content. Then, we summed up the number of food and beverage ads viewed each day for one week prior to requesting Facebook data. The Omnibus surveys include demographic data for each respondent, enabling us to analyze patterns of exposure to food and beverage advertising by individual demographic characteristics.

About the Social Data Collaboratory

NORC's Social Data Collaboratory (SDC) brings together a diverse team of social science, data science, and communication experts. With decades of experience, cutting-edge technology, and a commitment to being responsible data custodians, the SDC offers unparalleled expertise in collecting, processing, and interpreting social media data. The SDC can access and analyze data from a wide range of platforms, including Facebook, Instagram, YouTube, TikTok, Reddit, and X (formerly Twitter), and we continuously add new platforms to our portfolio. We employ advanced computational methods, including natural language processing, machine learning, and AI capabilities, to provide unbiased research and actionable insights.

Our portfolio of work is agnostic to subject matter, and we are agile in our capability to collect and analyze social media data from varying platforms. We have conducted comprehensive analyses on social, health, and political topics. Our partners trust us to harness the potential of social media data for transformative research, informed decision-making, and insights into societal trends.