



Cornell
SC Johnson College of Business

Newly Available Datasets for Research: Quick Overview and Access

Jura Liukonyte



Consumer transactions, ownership and behavior



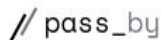
Aggregated, anonymized credit and debit card transaction data that tracks consumer spending behavior across thousands of brands, companies, and merchant categories, with breakouts by geography, channel, and demographics. Data includes cumulative spending and transaction volume, and panel data allows for understanding dynamics. Data ranges from January 2018 to September 2025, refreshes on the 8th.

Aggregated, privacy-compliant foot-traffic and mobility data derived from a panel of mobile devices, linked to points of interest (POIs) and geographic areas such as census block groups. Data includes counts of visits, visitation patterns, and movement behavior at and between locations- useful for studying consumer mobility. Data range from January 2017 to December 2025, refreshes monthly.



Transaction-level, SKU-level point-of-sale (POS) data from tens of thousands of independent U.S. convenience stores, detailing items sold, store attributes, discounts, and shopper activity. Data refreshes monthly on the 2nd.

Longitudinal market insight and consumer survey data on durable goods purchases in the United States, capturing detailed information on what products consumers buy, where they buy them, why they make those choices, and how brands perform over time. As a primary, quantitative dataset, it provides comprehensive key performance indicators that are not available through point-of-sale data, receipt data, or other traditional sources. Data spans 2012 to 2022, refreshed quarterly.



AI-enhanced, privacy-compliant foot-traffic and visitor characteristic data for millions of retail points of interest (POIs) in the U.S., including weekly, daily, and hourly visit counts and aggregated demographic/psychographic profiles of visitors. Data spans December 2018 to December 2024, refreshed quarterly.

Comprehensive U.S. consumer, household, property, and auto ownership data built on Verisk's TrueSource™ identity graph- including detailed demographic and geographic attributes, sequential address histories, property characteristics, and household vehicle information for hundreds of millions of records. Data provides insights into population characteristics, residential and automotive asset profiles, and socio-economic patterns. Data spans 30+ years and is refreshed monthly.




Detailed, privacy-safe location and foot-traffic data on millions of places in the U.S., including points of interest, visit patterns, and neighborhood-level mobility trends. Data includes customer spending, transaction volume, as well as where else consumers spend money and the breakdown of online/offline spending. Data ranges from January 2019 to October 2025, refreshes on the 21st.


Anonymized, high-quality mobile location and visit data that shows when and where devices visit millions of points of interest across the U.S., along with derived attributes like home and work locations. Data includes detailed information about location of visit, time and duration of visit. Data spans 2019 to present, refreshed monthly.




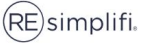
Global digital traffic and engagement data for millions of websites and mobile apps- including estimated visit counts, user behavior metrics, and organic search keywords driving traffic to domains, as well as daily app engagement statistics. Data includes visits to a given domain, daily app engagement, and desktop clicks per search term per landing page. Data spans 2019 to present, refreshed monthly.


Geolocation, rental and real estate markets

 Comprehensive global point-of-interest (POI) data with detailed location attributes (such as name, category, address, coordinates, and contact information) updated regularly and covering many countries around the world. The data consists of rich, standardized place-level data useful for spatial analysis of business distribution. Data is refreshed on an ad hoc basis.


 Comprehensive, nationwide U.S. real estate and property data-including tax assessor records, historical property assessments, recorded transactions, pre-foreclosure events, neighborhood and school boundaries, and automated valuation model (AVM) estimates for more than 155 million properties. The data provides rich, granular property-level information on ownership, valuations, sales and mortgage history, and geographic context ideal for analyzing housing market dynamics, neighborhood change, and real estate finance. Datasets span 1989 to present, refreshed monthly.


 Nationwide residential rental listing data aggregated from over 1 million public rental advertisements per week across 41,000+ U.S. ZIP codes, with detailed attributes like rent price, unit features, geolocation, amenities, and listing timelines. Data enables high-resolution, longitudinal rental market insights for analyses of housing affordability, regional rent dynamics, and urban development. Data spans 2014 to present, refreshed ad hoc.

 Comprehensive commercial real estate listings data for properties across the United States, including sale and lease attributes such as property type, location, asking price, cap rate, and status, sourced from public and private listing services and normalized for research use. Datasets cover listings, brokers, and suites. Data spans 2021 to present, refreshed quarterly.


 Rich, global IP address intelligence including geolocation mapping of billions of public IPs to cities/regions/countries, standardized Whois registry data for IP range ownership, organizational/company context for IP ranges, and privacy/traffic characterization such as VPN, proxy, and TOR detection. Data provides granular, structured insights into internet infrastructure and activity, useful for studies in cybersecurity, network analysis, and digital behavior. Data is refreshed monthly.


Construction and building


 Construction activity and equipment usage data by tracking the movement and utilization of construction site assets-such as open-top dumpsters (roll-offs) and heavy machinery rentals (e.g., lifts, booms, and other equipment)-across the U.S. and Canada. Data includes location of job, type of service, date, charge, and description of the product. Spans 2018 to present, refreshed monthly.

 Standardized, nationwide U.S. building permit data that capture details on construction activity-including permit types, project scope, costs, timelines, and contractor information- across the municipalities in the US. Data include building geolocation information, permit information, timeline information and project information. Spans 1980 to present, refreshed monthly.

Climate and the environment

 Quantitative measures of the natural environment at specific locations- such as NatureScore and Urban Heat Index values that assess the quantity and quality of natural elements and heat exposure for geographic areas. Data include indicators of heat risk and environmental health of a location to allow incorporating environmental indicators into analyses. Historical coverage is a snapshot, and refreshed annually.

 Historical and high-frequency U.S. weather and climate data- including daily and hourly observations of temperature, humidity, precipitation, wind speed, dew point, visibility, and sea-level pressure from thousands of weather stations. Data includes rich, time-series meteorological information useful for incorporating weather into analyses. Spans 2018 to present, refreshed monthly.

 Proprietary, zip-code-level climate risk ratings and projections for multiple hazards (such as extreme heat, precipitation, drought, flooding, and fire) based on climate change models through 2050, synthesized from internationally accepted climate projections and government sources. Data refreshed ad hoc.



Firms and employment



Rich, structured company and workforce data, including global company profiles with firmographic attributes (industry, size, employee counts, tenure, and growth trends) and, with institutional access, detailed historical people-level work and education histories from millions of resumes. Data provides comprehensive insights into labor market dynamics, organizational workforce changes, employment histories, and global business characteristics. Spans 2010 to present, refreshed monthly.



Comprehensive, government-sourced data on U.S. companies- including firmographics, payroll and employment, financials, loan records, private stock filings, and occupational wage statistics covering millions of active businesses. Data spans 2010 to the present, and is refreshed ad hoc.



Global, GDPR-compliant B2B company and office intelligence with detailed technographic, firmographic, and location attributes for hundreds of millions of businesses- including industry codes, employee counts, office locations, technology usage, domains, and social account data. Data covers companies and offices and ranges from basic information to more detailed financial data. Data spans 15 years and is refreshed monthly.



Massive, structured global job posting and compensation data with hundreds of millions of individual job records that include salaries (both posted and modeled), job titles, company, industry, and location metadata. Data maps roles to job family and government occupational codes for standardized analyses of roles. Data begins in 2016, refreshed monthly on the 15th.



Detailed, longitudinal U.S. company layoff records derived from public WARN (Worker Adjustment and Retraining Notification) notices- including employer name, state, notice and layoff dates, temporary status, and industry information. Data enables analysis of labor market dynamics, economic distress cycles, firm risk, and regional employment trends over time. Spans 1988 to present, refreshed ad hoc.



Proprietary, global job market data sourced directly from employer websites, including full historical and ongoing job posting records with structured fields and derived analytics aggregated at company and ticker levels. Data includes job posting information and dynamics derived from scrapes that enables analysis of hiring trends, industry and geographic job flows, and firm-level employment signals. Data spans 2007 to present, refreshed monthly.



Comprehensive global firmographic business profiles with standardized core company attributes (such as company type, industry classifications, locations, employee count, and other contextual metadata) that are refreshed regularly and built from AI-enhanced data collection processes. Data provides a stratified random sample of 20% of companies per country from the full Veridion universe of 134M+ firms. Data is refreshed monthly on the 30th.

Miscellaneous



Real-time and historical predictive pricing and liquidity metrics for fixed-income securities (e.g., bid/ask prices, spreads, and liquidity indicators across thousands of USD-denominated bonds) generated by 7 Chord's AI-driven pricing engine. Researchers can expect high-frequency, market-informed bond pricing analytics useful for studying issuer credit risk, relative value, fixed-income liquidity dynamics, and other corporate finance or financial economics topics. Data spans 2020 to present, refreshed daily.



Detailed data on professional football (soccer) player salaries across top global leagues (including the major European leagues and MLS) and financial statements of football clubs, with historical coverage over multiple seasons. Data includes club value and player salary for the biggest leagues in the world and their second divisions. Data spans 2013 to present, refreshed quarterly.



Cleaned, structured social sentiment metrics derived from large volumes of unstructured text data- most prominently normalized Twitter (X) social sentiment signals tied to equities and other financial instruments, processed using NLP and machine-learning techniques. Sentiment indicators reflect market and investor mood that can be integrated into quantitative finance, behavioral finance, risk modeling, and marketing analyses to study how social media signals relate to asset prices and volatility. Data spans 2011 to present, refreshed weekly.



Global financial market data including precise foreign exchange rates for numerous currencies, end-of-day equity prices with adjustment factors across 170+ exchanges, and readable corporate actions information for securities worldwide. Data provides reliable, standardized pricing and reference data that support empirical analysis of asset pricing, portfolio performance, and corporate events across global equity and currency markets. Spans 2014 to present, refreshed ad hoc.



Comprehensive, structured historical and ongoing federal lobbying activity, including millions of lobbying contracts, detailed information on clients, lobbyists, issues, bills, and agencies, with entities mapped to standardized identifiers such as stock tickers for easier integration with other datasets. Data includes over 1.6 million lobbying contracts across 200,000+ entities, 13,000+ lobbying firms, 78,000+ registered lobbyists. Spans 1999 to present, refreshed monthly on the 2nd.



Detailed, standardized global shipping and supply chain performance data- including metrics on container volumes, vessel arrival/departure events, and import/export port efficiency aggregated at daily and monthly levels across major ocean ports. Datasets provide insights into how efficiently ports are processing imports and exports, including how performance has changed over time. Data spans 2022 to present, refreshed monthly.



- All SC Johnson researchers currently have free access to this data. In the next billing cycle, if you continue using the data, you may be asked to contribute toward the subscription renewal.
- Please note that our license restricts access to researchers within the SC Johnson College of Business. You may collaborate with co-authors outside SC Johnson (including outside Cornell), but the license does not allow sharing or providing access to the data with those co-authors. In those cases, external co-authors would need to obtain their own license.

If you would like to request access to the data:

1. Create a free Dewey account using your Cornell email (preferably your Cornell NetID email): <https://www.deweydata.io>
2. Fill out [this internal JCB form](#).

Once you have completed steps (1) and (2), your account will be upgraded to a premium subscription, allowing you to access Dewey's datasets. Please note that it may take a couple of days for the confirmation to go through.