

## **FREE DATA SOURCES**

Kaggle Datasets: A platform for data scientists to share and find datasets for machine learning projects.

ImageNet: A dataset of over 14 million images that can be used for image classification, object detection, and image segmentation.

Open Images: A dataset of over 9 million images annotated with object detection and segmentation labels.

UCI Machine Learning Repository: A repository of various datasets for machine learning, including text, image, and time-series data.

Google Cloud Datasets: A collection of various public datasets provided by Google for machine learning, including image, text, and tabular data.

GitHub: A platform for hosting and sharing code, including datasets for machine learning projects.

## **FEDERAL GOVERNMENT**

Data.gov - This is a central repository for open data from federal agencies, including datasets related to budget, procurement, and human resources.

USA Spending - This is a government-wide database that provides information on federal spending, including budget data, contracts, grants, loans, and assistance programs.

Office of Personnel Management (OPM) - The OPM provides various datasets related to human resources, including federal employee salaries, benefits, and demographics.

Federal Procurement Data System (FPDS) - FPDS is the central repository of federal procurement data, containing information on all federal contracts over \$10,000. The dataset includes information on contract awards, vendors, agencies, and contract types. It can be used to forecast future contract opportunities by analyzing historical patterns and trends.

General Services Administration (GSA) Schedules - The GSA Schedules program is a procurement vehicle that provides federal agencies with access to commercial products and services. The GSA provides a dataset containing information on all vendors and their products and services available through

the schedules program. This dataset can be used to identify vendors and products that are likely to be in demand in the future.

Federal Funding Accountability and Transparency Act (FFATA) dataset - The FFATA requires federal agencies to report data on all subcontracts and grants over \$25,000. The dataset can be used to identify potential subcontracting opportunities and track federal funding.

Department of Defense (DOD) contracts dataset - The DOD provides a dataset containing information on all of its contracts, including contract number, award date, and vendor. This dataset can be used to identify potential contracting opportunities and track vendor performance.

Department of Energy (DOE) contracts dataset - The DOE provides a dataset containing information on all of its contracts, including contract number, award date, and vendor. This dataset can be used to identify potential contracting opportunities and track vendor performance.

Department of Defense (DOD) Acquisition Program Baselines dataset - This dataset provides information on the cost, schedule, and performance goals for major DOD acquisition programs. It can be used to identify potential opportunities for contractors to provide support and services to these programs.

Department of Homeland Security (DHS) Procurement Innovation Lab (PIL) dataset - The PIL provides a dataset containing information on DHS contracts and procurement opportunities. This dataset can be used to identify potential contracting opportunities with DHS.

Small Business Administration (SBA) dataset - The SBA provides a dataset containing information on all small businesses that are eligible for federal government contracts. This dataset can be used to identify potential partners for teaming arrangements or subcontracting opportunities.

Defense Logistics Agency (DLA) dataset - The DLA provides a dataset containing information on its supply chain, including inventory levels, suppliers, and demand forecasts. This dataset can be used to identify potential opportunities for supplying products and services to the DLA.

Federal Business Opportunities (FedBizOpps) dataset - FedBizOpps is the primary source for federal government procurement opportunities over \$25,000. The dataset includes information on upcoming solicitations, contract awards, and small business opportunities. It can be used to identify potential opportunities and track upcoming solicitations.

Federal Procurement Data System (FPDS) - This is a government-wide database that provides information on federal contracting, including contract awards, vendors, and pricing data.

Federal Financial Institutions Examination Council (FFIEC) - The FFIEC provides various datasets related to financial institutions, including data on bank salaries, loans, and assets.

Department of Labor - The Department of Labor provides various datasets related to labor market data, including wage data by occupation and geographic area.

Bureau of Labor Statistics (BLS) - The BLS provides various datasets related to labor market data, including employment, unemployment, wages, and benefits.

United States Census Bureau - The Census Bureau provides various datasets related to population, demographics, and income data, including the American Community Survey (ACS) and the Decennial Census.

Department of Education - The Department of Education provides various datasets related to education data, including information on graduation rates, student loans, and school finances.

Department of Agriculture - The Department of Agriculture provides various datasets related to agriculture and rural development, including information on farm incomes, subsidies, and food assistance programs.

Social Security Administration (SSA) - The SSA provides various datasets related to social security benefits, including information on retirement, disability, and survivor benefits.

Bureau of Economic Analysis (BEA) - The BEA provides various datasets related to economic data, including information on GDP, personal income, and international trade.

Department of Housing and Urban Development (HUD) - The HUD provides various datasets related to housing and community development, including information on housing affordability, homelessness, and community development block grants.

Congressional Budget Office (CBO) - The CBO provides various datasets related to federal budget and economic data, including information on revenues, expenditures, and economic indicators.

Government Accountability Office (GAO) - The GAO provides various datasets related to federal government performance, including audits, reports, and other assessments.

Federal Reserve Economic Data (FRED) - FRED is a database maintained by the Federal Reserve Bank of St. Louis that provides economic data on a wide range of topics, including employment, inflation, and monetary policy.

Internal Revenue Service (IRS) - The IRS provides various datasets related to tax data, including information on tax returns, deductions, and credits.

National Center for Education Statistics (NCES) - The NCES provides various datasets related to education data, including information on school budgets, student enrollment, and teacher salaries.

Environmental Protection Agency (EPA) - The EPA provides various datasets related to environmental data, including information on hazardous waste, air pollution, and water quality.

National Institutes of Health (NIH) - The NIH provides various datasets related to biomedical research, including information on grants, research funding, and clinical trials.

National Oceanic and Atmospheric Administration (NOAA) - NOAA produces a wide range of environmental datasets, including weather data, oceanographic data, and climate data. These datasets are available through the NOAA National Centers for Environmental Information (NCEI) website.

National Aeronautics and Space Administration (NASA) - NASA produces a variety of datasets related to Earth observation, climate science, and space exploration. Many of these datasets are available through the NASA Earth Science Data Systems (ESDS) Data Access System (DAS).

United States Geological Survey (USGS) - The USGS produces a range of geological, hydrological, and biological datasets. These datasets are available through the USGS Science Data Catalog.

National Institutes of Health (NIH) - NIH produces a variety of biomedical and health-related datasets, including genomic and clinical data. These datasets are available through the NIH Data Sharing Repositories.

Centers for Disease Control and Prevention (CDC) - CDC produces a range of public health datasets, including data on infectious diseases, chronic diseases, and health behaviors. These datasets are available through the CDC National Center for Health Statistics (NCHS) website.

Department of Labor - The Department of Labor produces a range of datasets related to employment, including data on unemployment rates, job growth, and wages. These datasets are available through the Department of Labor's Bureau of Labor Statistics (BLS) website.

Department of Transportation - The Department of Transportation produces a variety of transportation-related datasets, including data on traffic flow, vehicle safety, and transportation infrastructure. These datasets are available through the Department of Transportation's Bureau of Transportation Statistics (BTS) website.

Department of Energy - The Department of Energy produces a variety of energy-related datasets, including data on energy consumption, production, and prices. These datasets are available through the Department of Energy's Energy Information Administration (EIA) website.

Environmental Protection Agency (EPA) - The EPA produces a range of environmental datasets, including data on air and water quality, toxic substances, and hazardous waste. These datasets are available through the EPA's Environmental Data Gateway.

National Institute of Standards and Technology (NIST) - NIST produces a range of datasets related to measurement science, engineering, and technology. These datasets are available through the NIST Data Gateway.

Department of Homeland Security (DHS) - DHS produces a variety of datasets related to national security and public safety, including data on border security, immigration, and emergency management. These datasets are available through the DHS Data and Visualization Library.

National Science Foundation (NSF) - NSF produces a variety of datasets related to scientific research and innovation, including data on research funding, scientific publications, and research facilities. These datasets are available through the NSF's Data Management and Sharing website.

Department of Housing and Urban Development (HUD) - HUD produces a range of datasets related to housing and urban development, including data

on affordable housing, homelessness, and community development. These datasets are available through the HUD Exchange.

National Archives and Records Administration (NARA) - NARA produces a variety of datasets related to historical records and government archives, including data on presidential libraries, federal records centers, and archival collections. These datasets are available through the NARA's Catalog.

Department of Agriculture - The Department of Agriculture produces a variety of datasets related to agriculture, food safety, and nutrition, including data on crop yields, food prices, and dietary intake. These datasets are available through the USDA's National Agricultural Library.

Department of Commerce - The Department of Commerce produces a range of datasets related to trade, economic development, and innovation, including data on international trade, patents, and small business growth. These datasets are available through the Commerce Data Service.

Department of the Interior - The Department of the Interior produces a variety of datasets related to natural resources and public lands, including data on wildlife populations, geologic formations, and land use. These datasets are available through the Interior's Data.gov page.

National Endowment for the Arts (NEA) - The NEA produces a variety of datasets related to the arts and cultural heritage, including data on arts participation, cultural organizations, and creative industries. These datasets are available through the NEA's Research & Analysis page.

Department of Veterans Affairs (VA) - The VA produces a range of datasets related to veteran health and welfare, including data on healthcare utilization, disability compensation, and veteran homelessness. These datasets are available through the VA's Open Data Portal.

Consumer Financial Protection Bureau (CFPB) - The CFPB produces a range of datasets related to consumer finance, including data on mortgage lending, credit card complaints, and student loans. These datasets are available through the CFPB's Data and Research page.

United States Geological Survey (USGS) - The USGS produces a variety of datasets related to geology, hydrology, and natural hazards, including data on earthquakes, landslides, and water resources. These datasets are available through the USGS's Science Data Catalog.

## COMMERCIAL DATASETS

Bureau of Labor Statistics (BLS) - provides data on employment, wages, and benefits, as well as industry-specific trends and projections.

PayScale - offers data on salary, compensation, and benefits for various job titles and industries.

LinkedIn Workforce Report - provides insights into employment trends and skills gaps in various industries and regions.

Centers for Medicare & Medicaid Services (CMS) - offers data on healthcare costs, utilization, and outcomes for various medical conditions.

MedlinePlus - provides health information for patients, families, and healthcare professionals.

HealthData.gov - offers access to various healthcare datasets, including Medicare and Medicaid claims data.

USASpending.gov - provides data on federal government spending by agency, program, and location.

Gartner - offers research and insights on technology, customer experience, and business strategy.

Forrester Research - provides data and insights on customer behavior, digital transformation, and emerging technologies.

Zendesk Benchmark - provides data on customer service metrics and industry-specific benchmarks.

### *Human Resources:*

- LinkedIn Talent Insights: a dataset that provides talent analytics to help companies understand the job market and identify hiring trends.
- Glassdoor Economic Research: a dataset that provides salary and compensation data for different industries and job titles.
- SHRM Research: a dataset that provides research and reports on various HR topics, including employee engagement, retention, and workplace diversity.

### *Salary Trends:*

- Bureau of Labor Statistics (BLS) Occupational Employment Statistics (OES): a dataset that provides employment and wage estimates for different occupations and industries.
- PayScale: a dataset that provides salary data and compensation insights for different job titles and industries.
- LinkedIn Salary: a dataset that provides salary data for different job titles and locations based on information from LinkedIn members.
- Healthcare:
- Centers for Medicare & Medicaid Services (CMS) Medicare Claims: a dataset that includes claims data for various healthcare services provided to Medicare beneficiaries.
- National Health and Nutrition Examination Survey (NHANES): a dataset that provides information on the health and nutritional status of the US population.
- American Hospital Association (AHA) Annual Survey: a dataset that provides information on hospitals and their characteristics, such as bed size, staffing, and services offered.

### *Budget:*

- US Census Bureau Annual Survey of State and Local Government Finances: a dataset that provides information on the finances of state and local governments, including revenue, expenditures, and debt.
- National Science Foundation (NSF) Science and Engineering Indicators: a dataset that provides information on the federal budget for research and development, as well as data on research and development activities.
- Bureau of Economic Analysis (BEA) National Income and Product Accounts (NIPA): a dataset that provides data on national income, output, and expenditure.

### *Call Center:*

- **Zendesk Benchmark:** a dataset that provides metrics and insights for customer service performance, including response times, customer satisfaction, and agent productivity.
- **Cisco Contact Center Analytics:** a dataset that provides data on call center performance and customer experience, including metrics such as average handle time, abandonment rate, and first-call resolution.
- **Genesys Contact Center Data:** a dataset that provides data on contact center operations, including agent performance, customer interactions, and contact center technology.