Opportunity Insights Economic Tracker Data Dictionary

last updated on 2023-08-29

1 Overview

Each data source and level of aggregation has a separate CSV, named using the following convention: *Data source – Geographic Level of Aggregation – Temporal Level of Aggregation*

Additionally, we have three files, **GeoIDs – State** and **GeoIDs – County** and **GeoIDs – City**, that provide information on geographic crosswalks and aggregation. These can be merged to any file sharing the same geographic level of aggregation using the geographic identifier. Additionally, **GeoIDs – County** indicates the commuting zone (CZ) and state that each county belongs to. The City-level data (listed under "Metro" on the tracker site) associates the largest cities in the United States with a representative county one-to-one (except in the case of New York City which includes the 5 boroughs).

Finally, we have gathered a collection of key state-level policy dates relevant for changes in other series trends and values. These are contained in the **Policy Milestones – State** file.

A description of the columns in each file follows.

2 GeoID File Descriptions

2.1 GeoIDs - State.csv

Geographic identifier: `statefips`

- `statename`: The name of the state.
- `stateabbrev`: The 2-letter state abbreviation.
- `state_pop2019`: The population of the state in 2019, from Census Bureau estimates.

2.2 GeoIDs - County.csv

Geographic identifier: `countyfips`

- `countyname`: The name of the county.
- `cityid`: The city identifier that the county is assigned to.
- `cityname`: The name of the city that the county is assigned to.
- `cz`: The numeric identifier of the commuting zone (CZ) in which the county is contained.
- `czname`: The name of the commuting zone (CZ) in which the county is contained.
- `statename`: The name of the state in which the county is contained.
- `statefips`: The FIPS code of the state in which the county is contained.
- `stateabbrev`: The 2-letter abbreviation of the state in which the county is contained.
- `county_pop2019`: The population of the county in 2019 according to Census Bureau estimates.

2.3 GeoIDs - City.csv

Geographic identifier: `cityid`

- `cityname`: The name of the city.
- `stateabbrev`: The 2-letter abbreviation of the primary state in which the city is contained.
- `statename`: The name of the primary state in which the city is contained.
- `statefips`: The FIPS code of the primary state in which the city is contained.
- `lat`: Latitude of the city.
- `lon`: Longitude of the city.
- `city_pop2019`: The population of the city in 2019 according to Census Bureau estimates, calculated as population of the county or counties assigned to the city.

3 Data File Descriptions

3.1 Affinity

Credit/debit card spending data from Affinity Solutions.

- `spend_all`: Spending in all merchant category codes (MCCs).
  - `spend_all_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
  - `spend_all_incmiddle`: ... by consumers living in ZIP codes with median income in the middle two quartiles.
- `spend_aap`: Spending in apparel and accessories (AAP) MCCs.
  - `spend_aap_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_acf`: Spending in accomodation and food service (ACF) MCCs.
  - `spend_acf_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_aer`: Spending in arts, entertainment, and recreation (AER) MCCs.
  - `spend_aer_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_apg`: Spending in general merchandise stores (GEN) and apparel and accessories (AAP) MCCs.
  - `spend_apg_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_gen`: Spending in general merchandise stores (GEN) MCCs.
  - `spend_gen_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_grf`: Spending in grocery and food store (GRF) MCCs.
  - `spend_grf_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_hcs`: Spending in health care and social assistance (HCS) MCCs.
  - `spend_hcs_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_hic`: Spending in home improvement centers (HIS) MCCs.
  - `spend_hic_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_sgh`: Spending in sporting goods and hobby (SGH) MCCs.
  - `spend_sgh_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_tws`: Spending in transportation and warehousing (TWS) MCCs.

  - `spend_tws_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_retail_w_grocery`: Spending in retail (BLD, CLO, ELC, FBS, FUR, GEN, SPO) MCCs including grocery spending.
  - `spend_retail_w_grocery_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_retail_no_grocery`: Spending in retail (BLD, CLO, ELC, FUR, GEN, SPO) MCCs excluding grocery spending.
  - `spend_retail_no_grocery_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_durables`: Spending in durable goods (BLD, ELC, FUR, SPO, TEL, VEH) MCCs.
  - `spend_durables_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_nondurables`: Spending in non-durable goods (CLO, FBS, GAS, GEN, HPC, MSC, WHO) MCCs.
  - `spend_nondurables_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_remoteservices`: Spending in remote services (ADM, EDU, FIN, INF, NSR, PST, PUB, UCM) MCCs.
  - `spend_remoteservices_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_inperson`: Spending in in-person services (ACF, HCS, AER, TWS, REN, REP, PLS) MCCs.
  - `spend_inperson_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `spend_inpersonmisc`: Spending in other in-person services (REN, REP, PLS) MCCs.
  - `spend_inpersonmisc_q#`: ... by consumers living in ZIP codes with median income in quartile `#`.
- `provisional`: Indicator to mark that the date is within the most recent three weeks of data and is subject to non-negligible changes as new data is posted.
- `freq`: Marks whether the data represents a daily ("d") or weekly ("w") value.

All spending variables are measured relative to January 6 to February 2, 2020, seasonally adjusted, and calculated as a 7 day moving average. When we subdivide by income using the median income of the ZIP codes, `q1` is the quartile with the lowest median income and `q4` is the quartile with the highest median income. At the national level, we release a variety of breakdowns *without seasonal adjustment* in variables that begin with `spend_s_` (relative to January 2019 for 2019 data, relative to January 2020 for 2020 data onward) or `spend_19_` (relative to Janurary 7 to February 3, 2019 for all data) instead of `spend_`.

The merchant category codes (MCC) making up the grouped spending categories are:

- **Retail spending:** CLO clothing and clothing accessories; BLD building materials, garden equipment, and supplies; ELC electronics and appliances; FBS food and beverage stores; FUR furniture and home furnishings; GEN general merchandise stores; and SPO sporting goods, hobbies, musical instruments, and book stores.
- **Durable goods:** BLD building materials, gardening equipment, and supplies; ELC electronics and appliances; FUR furniture and home furnishings; SPO sporting goods, hobbies, musical instruments, and bookstores; TEL telecommunications; and VEH motor vehicles and parts.
- **Non-durable goods:** CLO clothing and clothing accessories; FBS food and beverage stores; GEN general merchandise; HPC health and personal care stores; and WHO wholesale trade.

- **Remote services:** ADM administrative and support and waste management and remediation services; EDU education; FIN finance and insurance; INF information; PST professional, scientific, and technical; PUB public administration; and UCM utilities, construction, and manufacturing.
- **In-person services:** ACF accomodation and food services; HCS healthcare and social assistance; AER arts, entertainment, and recreation; TWS transportation and warehousing; REN rental and leasing; REP repair and maintenance; and PLS personal and laundry services.

In addition, four supplemental files are included (see the documentation for more details on these files):

- *Affinity Income Shares - National - 2019.csv*
- *Affinity Income Shares - National - 2020.csv*
- *Affinity Industry Composition - National - 2020.csv*
- *Affinity Daily Total Spending - National - Daily.csv*

3.2 Job Postings

Job postings data from Lightcast (formerly known as Burning Glass Technologies).

- `bg_posts`: Average level of job postings relative to January 4 to 31, 2020.
- `bg_posts_ss30`: Average level of job postings relative to January 4 to 31, 2020 in manufacturing (NAICS supersector 30).
- `bg_posts_ss55`: Average level of job postings relative to January 4 to 31, 2020 in financial activities (NAICS supersector 55).
- `bg_posts_ss60`: Average level of job postings relative to January 4 to 31, 2020 in professional and business services (NAICS supersector 60).
- `bg_posts_ss65`: Average level of job postings relative to January 4 to 31, 2020 in education and health services (NAICS supersector 65).
- `bg_posts_ss70`: Average level of job postings relative to January 4 to 31, 2020 in leisure and hospitality (NAICS supersector 70).
- `bg_posts_jz1`: Average level of job postings relative to January 4 to 31, 2020 requiring little/no preparation (ONET jobzone level 1).
- `bg_posts_jz2`: Average level of job postings relative to January 4 to 31, 2020 requiring some preparation (ONET jobzone level 2).
- `bg_posts_jz3`: Average level of job postings relative to January 4 to 31, 2020 requiring medium preparation (ONET jobzone level 3).
- `bg_posts_jz4`: Average level of job postings relative to January 4 to 31, 2020 requiring considerable preparation (ONET jobzone level 4).
- `bg_posts_jz5`: Average level of job postings relative to January 4 to 31, 2020 requiring extensive preparation (ONET jobzone level 5).
- `bg_posts_jzgrp12`: Average level of job postings relative to January 4 to 31, 2020 requiring low preparation (ONET jobzone levels 1 and 2).
- `bg_posts_jzgrp345`: Average level of job postings relative to January 4 to 31, 2020 requiring high preparation (ONET jobzone levels 3, 4 and 5).

In addition, the following supplemental file is included (see the documentation for more details on this file):

- *Job Postings Industry Shares - National - 2020.csv*

3.3 COVID

COVID cases and deaths numbers are from the New York Times and the Centers for Disease Control and Prevention, hospitalizations numbers are from the U.S. Department of Health and Human Services, tests numbers are from Johns Hopkins University, and vaccination numbers are from the Centers for Disease Control and Prevention.

- `case_rate`: Confirmed COVID-19 cases per 100,000 people, seven day moving average.
  - `case_count`: Confirmed COVID-19 cases, seven day moving average.
- `new_case_rate`: New confirmed COVID-19 cases per 100,000 people, seven day rolling sum.
  - `new_case_count`: New confirmed COVID-19 cases, seven day rolling sum.
- `death_rate`: Confirmed COVID-19 deaths per 100,000 people, seven day moving average.
  - `death_count`: Confirmed COVID-19 deaths, seven day moving average.
- `new_death_rate`: New confirmed COVID-19 deaths per 100,000 people, seven day rolling sum.
  - `new_death_count`: New confirmed COVID-19 deaths, seven day rolling sum.
- `test_rate`: Confirmed COVID-19 tests per 100,000 people, seven day moving average.
  - `test_count`: Confirmed COVID-19 tests, seven day moving average.
- `new_test_rate`: New confirmed COVID-19 tests per 100,000 people, seven day moving average.
  - `new_test_count`: New confirmed COVID-19 tests, seven day moving average.
- `vaccine_rate`: First vaccine doses administered per 100 people.
  - `vaccine_count`: First vaccine doses administered.
- `new_vaccine_rate`: New first vaccine doses administered per 100 people, seven day moving average.
  - `new_vaccine_count`: New first vaccine doses administered, seven day moving average.
- `fullvaccine_rate`: Vaccine series completed per 100 people.
  - `fullvaccine_count`: Vaccine series completed.
- `new_fullvaccine_rate`: New vaccine series completed per 100 people, seven day moving average.
  - `new_fullvaccine_count`: New vaccine series completed, seven day moving average.
- `booster_first_rate`: First booster doses administered per 100 people.
  - `booster_first_count`: First booster doses administered.
- `new_booster_first_rate`: New first booster doses administered per 100 people, seven day moving average.
  - `new_booster_first_count`: New first booster doses administered, seven day moving average.
- `hospitalized_rate`: New patients currently hospitalized in an inpatient bed who have suspected or confirmed COVID-19 per 100,000 people, seven day moving average.
  - `hospitalized_count`: Newly patients currently hospitalized in an inpatient bed who have suspected or confirmed COVID-19, seven day moving average.

3.4 Google Mobility

GPS mobility data indexed to January 3 to February 6, 2020 from Google COVID-19 Community Mobility Reports.

- `gps_away_from_home`: Time spent outside of residential locations.
- `gps_retail_and_recreation`: Time spent at retail and recreation locations.
- `gps_grocery_and_pharmacy`: Time spent at grocery and pharmacy locations.
- `gps_parks`: Time spent at parks.
- `gps_transit_stations`: Time at inside transit stations.

- `gps_workplaces`: Time spent at work places.
- `gps_residential`: Time spent at residential locations.

3.5 Employment

Employment levels relative to January 4 to 31, 2020 from Paychex and Intuit.

- `emp`: Employment level for all workers.
- `emp_incq1`: Employment level for workers in the bottom quartile of the wage distribution (annualized wage lower than the federal poverty line).
- `emp_incq2`: Employment level for workers in the second quartile of the wage distribution (annualized wage between 1x and 1.5x the federal poverty line).
- `emp_incmiddle`: Employment level for workers in the middle two quartiles of the wage distribution (annualized wage between 1x and 2.5x the federal poverty line).
- `emp_incq3`: Employment level for workers in the third quartile of the wage distribution (annualized wage between 1.5x and 2.5x the federal poverty line).
- `emp_incq4`: Employment level for workers in the top quartile of the wage distribution (annualized wage greater than 2.5x the federal poverty line).
- `emp_incbelowmed`: Employment level for workers in the bottom half of the wage distribution (annualized wage less than 1.5x the federal poverty line).
- `emp_incabovemed`: Employment level for workers in the top half of the wage distribution (annualized wage greater than 1.5x the federal poverty line).
- `emp_ss40`: Employment level for workers in trade, transportation and utilities (NAICS supersector 40).
- `emp_ss60`: Employment level for workers in professional and business services (NAICS supersector 60).
- `emp_ss65`: Employment level for workers in education and health services (NAICS supersector 65).
- `emp_ss70`: Employment level for workers in leisure and hospitality (NAICS supersector 70).
- `emp_retail`: Employment level for workers in retail (NAICS sector 44-45).
- `emp_retail_inclow`: Employment level for workers in retail (NAICS sector 44-45) and in the bottom quartile of the wage distribution (annualized wage lower than the federal poverty line).
- `emp_retail_incmiddle`: Employment level for workers in retail (NAICS sector 44-45) and in the middle two quartiles of the wage distribution (annualized wage between 1x and 2.5x the federal poverty line).
- `emp_retail_inchigh`: Employment level for workers in retail (NAICS sector 44-45) and in the top quartile of the wage distribution (annualized wage greater than 2.5x the federal poverty line).
- `emp_s72`: Employment level for workers in accommodation and food services (NAICS sector 72).
- `emp_subset_unweighted_q1`: Employment level for workers in the bottom quartile of the wage distribution (annualized wage lower than the federal poverty line) in county x industry (2-digit NAICS code) cells with nonzero employment for all four wage quartiles.
- `emp_subset_unweighted_q2`: Employment level for workers in the second quartile of the wage distribution (annualized wage between 1x and 1.5x the federal poverty line) in county x industry (2-digit NAICS code) cells with nonzero employment for all four wage quartiles.
- `emp_subset_unweighted_q3`: Employment level for workers in the third quartile of the wage distribution (annualized wage between 1.5x and 2.5x the federal poverty line) in county x industry (2-digit NAICS code) cells with nonzero employment for all four wage quartiles.

- `emp_subset_unweighted_q4`: Employment level for workers in the top quartile of the wage distribution (annualized wage greater than 2.5x the federal poverty line) in county x industry (2-digit NAICS code) cells with nonzero employment for all four wage quartiles.
- `emp_subset_reweighted_q1`: Employment level for workers in the bottom quartile of the wage distribution (annualized wage lower than the federal poverty line) in county x industry cells with nonzero employment for all four wage quartiles, reweighting to match the county x industry (2-digit NAICS code) distribution of workers in the top quartile of the wage distribution.
- `emp_subset_reweighted_q2`: Employment level for workers in the second quartile of the wage distribution (annualized wage between 1x and 1.5x the federal poverty line) in county x industry cells with nonzero employment for all four wage quartiles, reweighting to match the county x industry (2-digit NAICS code) distribution of workers in the top quartile of the wage distribution.
- `emp_subset_reweighted_q3`: Employment level for workers in the third quartile of the wage distribution (annualized wage between 1.5x and 2.5x the federal poverty line) in county x industry cells with nonzero employment for all four wage quartiles, reweighting to match the county x industry (2-digit NAICS code) distribution of workers in the top quartile of the wage distribution.
- `emp_subset_reweighted_q4`: Employment level for workers in the top quartile of the wage distribution (annualized wage greater than 2.5x the federal poverty line) in county x industry cells with nonzero employment for all four wage quartiles, reweighting to match the county x industry (2-digit NAICS code) distribution of workers in the top quartile of the wage distribution.

In addition, the following supplemental file is included (see the documentation for more details on this file):

- *Earnin - ZCTA - 2020.csv*

3.6 UI Claims

Unemployment insurance claims data from the Department of Labor (national and state-level) and numerous individual state agencies (county-level).

- `initclaims_rate_regular`: Number of initial claims per 100 people in the 2019 labor force, Regular UI only
  - `initclaims_count_regular`: Count of initial claims, Regular UI only
- `initclaims_rate_pua`: Number of initial claims per 100 people in the 2019 labor force, PUA (Pandemic Unemployment Assistance) only
  - `initclaims_count_pua`: Count of initial claims, PUA (Pandemic Unemployment Assistance) only
- `initclaims_rate_combined`: Number of initial claims per 100 people in the 2019 labor force, combining Regular and PUA claims
  - `initclaims_count_combined`: Count of initial claims, combining Regular and PUA claims
- `contclaims_rate_regular`: Number of continued claims per 100 people in the 2019 labor force, Regular UI only
  - `contclaims_count_regular`: Count of continued claims, Regular UI only
- `contclaims_rate_pua`: Number of continued claims per 100 people in the 2019 labor force, PUA (Pandemic Unemployment Assistance) only
  - `contclaims_count_pua`: Count of continued claims, PUA (Pandemic Unemployment Assistance) only

- `contclaims_rate_peuc`: Number of continued claims per 100 people in the 2019 labor force, PEUC (Pandemic Emergency Unemployment Compensation) only
  - `contclaims_count_peuc`: Count of continued claims, PEUC (Pandemic Emergency Unemployment Compensation) only
- `contclaims_rate_combined`: Number of continued claims per 100 people in the 2019 labor force, combining Regular, PUA and PEUC claims
  - `contclaims_count_combined`: Count of continued claims, combining Regular, PUA and PEUC claims

3.7 Womply

Small business openings and revenue data from Womply.

- `merchants_all`: Percent change in number of small businesses open, calculated as a seven-day moving average, seasonally adjusted, and indexed to January 4 to 31, 2020.
  - `merchants_inchigh`: ... in high income (quartile 4 of median income) ZIP codes.
  - `merchants_incmiddle`: ... in middle income (quartiles 2 & 3 of median income) ZIP codes.
  - `merchants_inclow`: ... in low income (quartile 1 of median income) ZIP codes.
  - `merchants_retail`: ... in retail businesses (NAICS 2-digit codes 44-45).
  - `merchants_food_accommodation`: ... in food and accommodation businesses (NAICS 2-digit code 72)
  - `merchants_professional`: ... in professional services businesses (NAICS 2-digit code 54)
  - `merchants_other_services`: ... in other services businesses (NAICS 2-digit code 81)
  - `merchants_health`: ... in health & social service businesses (NAICS 2-digit code 62)
- `revenue_all`: Percent change in net revenue for small businesses, calculated as a seven-day moving average, seasonally adjusted, and indexed to January 4 to 31, 2020.
  - `revenue_inchigh`: ... in high income (quartile 4 of median income) zipcodes.
  - `revenue_incmiddle`: ... in middle income (quartiles 2 & 3 of median income) zipcodes.
  - `revenue_inclow`: ... in low income (quartile 1 of median income) zipcodes.
  - `revenue_retail`: ... in retail businesses (NAICS 2-digit codes 44-45).
  - `revenue_food_accommodation`: ... in food and accommodation businesses (NAICS 2-digit code 72).
  - `revenue_professional`: ... in professional services businesses (NAICS 2-digit code 54).
  - `revenue_other_services`: ... in other services businesses (NAICS 2-digit code 81)
  - `revenue_health`: ... in health & social service businesses (NAICS 2-digit code 62)

In addition, the following supplemental file is included (see the documentation for more details on this file):

- *Womply - ZCTA - 2020.csv*

3.8 Zearn

Online math learning data from Zearn.

- `engagement`: Average level of students using platform relative to January 6 to February 21, 2020.

- `engagement_inclow`: Average level of students using platform relative to January 6 to February 21, 2020 for schools in the 25% of ZIP codes with the lowest median income.
- `engagement_incmiddle`: Average level of students using platform relative to January 6 to February 21, 2020 for schools in ZIP codes with median income between the 25th and 75th percentiles.
- `engagement_inchigh`: Average level of students using platform relative to January 6 to February 21, 2020 for schools in the 25% of ZIP codes with the highest median income.
- `badges`: Average level of student achievements earned (badges) on platform relative to January 6 to February 21, 2020.
- `badges_inclow`: Average level of student achievements earned (badges) on platform relative to January 6 to February 21, 2020 for schools in the 25% of ZIP codes with the lowest median income.
- `badges_incmiddle`: Average level of student achievements earned (badges) on platform relative to January 6 to February 21, 2020 for schools in ZIP codes with median income between the 25th and 75th percentiles.
- `badges_inchigh`: Average level of student achievements earned (badges) on platform relative to January 6 to February 21, 2020 for schools in the 25% of ZIP codes with the highest median income.

Note that for every variable listed here, there is a corresponding variable with the prefix `break_` (for example, `break_engagement`). During the period in which schools are on summer or winter break, we record the outcomes in these `break_` variables instead of the usual variables. These numbers are not displayed on the Economic Tracker because they do not reliably measure differences in student learning across geography and income groups when many schools are on break.

To ensure privacy, the results for some counties are masked. Where possible, masked county levels values are replaced by commuting zone means, as indicated by the `imputed_from_cz` variable. The masking criteria are explained in further detail in our data documentation.

4 Policy Milestones

Key state-level policy dates relevant for changes in other series trends and values.

- `statefips`: 2-digit FIPS code of the U.S. state a milestone was recorded in
- `date`: Date the policy milestone occurred
- `policy_description`: Description of policy milestone
- `schools_first_closed`: Whether a milestone is the date at which the state first ordered all public K-12 schools statewide to physically close
- `nonessential_biz_first_closed`: Whether a milestone is the date on which the state government first ordered all nonessential businesses to close statewide
- `stayathome_first_start`: Whether a milestone is the date on which the state government first told residents to stay home, save for excepted activities statewide