Opportunity Insights Economic Tracker Data Dictionary

last updated on 2020-09-23

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).

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.
- `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, from Census Bureau estimates.

2.3 GeoIDs - **City.csv**

Geographic identifier: `cityid`

- `cityname`: The name of the city.
- `stateabbrev`: The 2-letter abbreviation of the state in which the city is contained.
- `statename`: The name of the state in which the city is contained.
- `statefips`: The fips code of the state in which the city is contained.
- `city_pop2019`: The population of the city in 2019.

3 File Descriptions

3.1 Affinity

Spending data from Affinity Solutions.

- `spend_all`: Seasonally adjusted credit/debit card spending relative to January 4-31 2020 in all merchant category codes (MCC), 7 day moving average.
- `spend_acf`: Seasonally adjusted credit/debit card spending relative to January 4-31 2020 in accomodation and food service (ACF) MCCs, 7 day moving average, 7 day moving average.
- `spend_aer`: Seasonally adjusted credit/debit card spending relative to January 4-31 2020 in arts, entertainment, and recreation (AER) MCCs, 7 day moving average.
- `spend_apg`: Seasonally adjusted credit/debit card spending relative to January 4-31 2020 in general merchandise stores (GEN) and apparel and accessories (AAP) MCCs, 7 day moving average.
- `spend_grf`: Seasonally adjusted credit/debit card spending relative to January 4-31 2020 in grocery and food store (GRF) MCCs, 7 day moving average.
- `spend_hcs`: Seasonally adjusted credit/debit card spending relative to January 4-31 2020 in health care and social assistance (HCS) MCCs, 7 day moving average.
- `spend_tws`: Seasonally adjusted credit/debit card spending relative to January 4-31 2020 in transportation and warehousing (TWS) MCCs, 7 day moving average.
- `spend_all_inchigh`: Seasonally adjusted credit/debit card spending by consumers living in ZIP codes with high (top quartile) median income, relative to January 4-31 2020 in all merchant category codes (MCC), 7 day moving average.
- `spend_all_incmiddle`: Seasonally adjusted credit/debit card spending by consumers living in ZIP codes with middle (middle two quartiles) median income, relative to January 4-31 2020 in all merchant category codes (MCC), 7 day moving average.
- `spend_all_inclow`: Seasonally adjusted credit/debit card spending by consumers living in ZIP codes with low (bottom quartiles) median income, relative to January 4-31 2020 in all merchant category codes (MCC), 7 day moving average.

3.2 Burning Glass

Job postings data from Burning Glass Technologies.

- `bg_posts`: Average level of job postings relative to January 4-31 2020.
- `bg_posts_ss30`: Average level of job postings relative to January 4-31 2020 in manufacturing (NAICS supersector 30).
- `bg_posts_ss55`: Average level of job postings relative to January 4-31 2020 in financial activities (NAICS supersector 55).

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

3.3 COVID

COVID case numbers and death numbers from the New York Times COVID-19 repository. COVID test numbers from the COVID Tracking Project.

- `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 moving average.
  - `new_case_count`: New confirmed COVID-19 cases, seven day moving average.
- `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 moving average.
  - `new_death_count`: New confirmed COVID-19 deaths, seven day moving average.
- `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.

3.4 Google Mobility

GPS mobility data indexed to Jan 3-Feb 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 Jan 4-31 from Paychex, Intuit, Earnin and Kronos.

- `emp_combined`: Employment level for all workers.
- `emp_combined_inclow`: Employment level for workers in the bottom quartile of the income distribution (incomes approximately under $27,000).
- `emp_combined_incmiddle`: Employment level for workers in the middle two quartiles of the income distribution (incomes approximately $27,000 to $60,000).
- `emp_combined_inchigh`: Employment level for workers in the top quartile of the income distribution (incomes approximately over $60,000).
- `emp_combined_ss40`: Employment level for workers in trade, transportation and utilities (NAICS supersector 40).
- `emp_combined_ss60`: Employment level for workers in professional and business services (NAICS supersector 60).
- `emp_combined_ss65`: Employment level for workers in education and health services (NAICS supersector 65).
- `emp_combined_ss70`: Employment level for workers in leisure and hospitality (NAICS supersector 70).
- `emp_combined_inclow_advance`: Indicator (0 or 1) for whether `emp_combined_inclow` is a forecasted employment level based on timecard data from Kronos.

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 Merchants

Small business openings 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 4-31 2020.
- ~~`merchants_inchigh`: Percent change in number of small businesses open calculated as a seven-day moving average seasonally adjusted and indexed to January 4-31 2020 in high income (quartile 4 of median income) ZIP codes.~~
- ~~`merchants_incmiddle`: Percent change in number of small businesses open calculated as a seven-day moving average seasonally adjusted and indexed to January 4-31 2020 in middle income (quartiles 2 & 3 of median income) ZIP codes.~~
- ~~`merchants_inclow`: Percent change in number of small businesses open calculated as a seven-day moving average seasonally adjusted and indexed to January 4-31 2020 in low income (quartile 1 of median income) ZIP codes.~~
- `merchants_ss40`: Percent change in number of small businesses open calculated as a seven-day moving average seasonally adjusted and indexed to January 4-31 2020 in transportation (NAICS supersector 40).
- `merchants_ss60`: Percent change in number of small businesses open calculated as a seven-day moving average seasonally adjusted and indexed to January 4-31 2020 in professional and business services (NAICS supersector 60).
- `merchants_ss65`: Percent change in number of small businesses open calculated as a seven-day moving average seasonally adjusted and indexed to January 4-31 2020 in education and health services (NAICS supersector 65).
- `merchants_ss70`: Percent change in number of small businesses open calculated as a seven-day moving average seasonally adjusted and indexed to January 4-31 2020 in leisure and hospitality (NAICS supersector 70).

Note: County-level and city-level data and breakdowns by High/Middle/Low income ZIP codes have been temporarily removed since the August 21st 2020 update due to revisions in the structure of the raw data we receive. We hope to add them back to the OI Economic Tracker soon.

3.8 Womply Revenue

Small business revenue data from Womply.

- `revenue_all`: Percent change in net revenue for small businesses, calculated as a seven-day moving average, seasonally adjusted, and indexed to January 4-31 2020.
- ~~`revenue_inchigh`: Percent change in net revenue for small businesses, calculated as a seven-day moving average, seasonally adjusted, and indexed to January 4-31 2020 in high income (quartile 4 of median income) zipcodes.~~
- ~~`revenue_incmiddle`: Percent change in net revenue for small businesses, calculated as a seven-day moving average, seasonally adjusted, and indexed to January 4-31 2020 in middle income (quartiles 2 & 3 of median income) zipcodes.~~
- ~~`revenue_inclow`: Percent change in net revenue for small businesses, calculated as a seven-day moving average, seasonally adjusted, and indexed to January 4-31 2020 in low income (quartile 1 of median income) zipcodes.~~
- `revenue_ss40`: Percent change in net revenue for small businesses, calculated as a seven-day moving average, seasonally adjusted, and indexed to January 4-31 2020 in transportation (NAICS supersector 40).
- `revenue_ss60`: Percent change in net revenue for small businesses, calculated as a seven-day moving average, seasonally adjusted, and indexed to January 4-31 2020 in professional and business services (NAICS supersector 60).
- `revenue_ss65`: Percent change in net revenue for small businesses, calculated as a seven-day moving average, seasonally adjusted, and indexed to January 4-31 2020 in education and health services (NAICS supersector 65).
- `revenue_ss70`: Percent change in net revenue for small businesses, calculated as a seven-day moving average, seasonally adjusted, and indexed to January 4-31 2020 in leisure and hospitality (NAICS supersector 70).

Note: County-level and city-level data and breakdowns by High/Middle/Low income ZIP codes have been temporarily removed since the August 21st 2020 update due to revisions in the structure of the raw data we receive. We hope to add them back to the OI Economic Tracker soon.

3.9 Zearn

Online math learning data from Zearn.

- `engagement`: Average level of students using platform relative to January 6-February 21 2020.
- `engagement_inclow`: Average level of students using platform relative to January 6-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-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-February 21 2020 for schools in the 25% of ZIP codes with the highest median income.

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