Opportunity Insights Economic Tracker Data Documentation

last updated on 2023-07-18

1 Overview

This document provides an overview of the sources and processing applied to each data series within the Opportunity Insights Economic Tracker. The documentation is organized sequentially by series in the tracker, then broken down into categories of information describing each series, its source data, and our processing steps.

You can refer to additional documentation published by Opportunity Insights for complementary information:

- The Economic Tracker's **Data Dictionary** lists each data file and variable available for public use, with short descriptions of the contents of each variable.
- The **accompanying paper** provides detailed information about the methodology used to construct the series.

Please note that both the data and this data documentation are updated regularly and that the following information is subject to change.

2 Data Series

2.1 Consumer Spending

Summary: Aggregated and anonymized purchase data from consumer credit and debit card spending. Spending is reported based on the ZIP code where the cardholder lives, not the ZIP code where transactions occurred.

Data Source: Affinity Solutions

Update Frequency: Weekly

Date Range: January 13, 2020 until the most recent date available.

Data Frequency: Data is daily until June 5, 2022, presented as a 7-day lookback moving average. Since June 5, 2022 we only receive weekly data on consumer spending, and the data is presented as weekly data points.

Indexing Period: January 6 to February 2, 2020

Indexing Type: Seasonally adjusted change since January 2020. We calculate the change relative to the January index period: 2019 data is indexed relative to January 2019 (January 7, 2019 to February 3, 2019), data in 2020 onward is indexed relative to January 2020 (January 6, 2020 to February 2, 2020). We then seasonally adjust by dividing by the indexed 2019 value,

which represents the difference between the change since January 2020 compared to the change since January observed since 2019. We account for differences in the dates of federal holidays between 2019 and 2020 by shifting the 2019 reference data to align the holidays before performing the year-over-year division.

Geographies: National, State, County, Metro

Breakdowns:

- *By Industry.* Industries are constructed by grouping merchant codes that are used by Affinity Solutions to identify the category of merchant and merchant activity.
  - Entertainment and Recreation
  - Grocery
  - Health Care
  - Restaurants and Hotels
  - Retail
  - Transportation
- *By Consumer Zip Code Income.* Transactions are linked to zip codes where the consumer lives and zip codes are classified into income categories based on measurements of median household income and population provided by the American Community Survey (2014 - 2018).
  - High Income (top quartile of median household income; approximately greater than $78,000 per year)
  - Middle Income (middle two quartiles of median household income; approximately between $46,000 per year and $78,000 per year)
  - Low Income (bottom quartile of median household income; approximately less than $46,000 per year)

Data masking: For the state-level breakdowns by income quartile and the county-level data, we mask locations with average daily spending $<$ $70,000 in January 2019. The raw data contains discontinuous breaks caused by entry or exit of credit card providers from the sample: counties with multiple structural breaks are dropped from the sample. Additionally, Affinity Solutions suppresses any cut of the data with fewer than five transactions. For more details refer to the accompanying paper.

Notes: We require at least 3 weeks of data in order to reliably identify and correct discontinuous breaks caused by entry or exit of credit card providers from the sample. The most recent 3 weeks of data are therefore marked 'provisional' and are subject to non-negligible changes as new data is posted. For breaks found prior to the last 3 weeks, we correct for it using a method outlined in the paper. Otherwise we substitute the national mean for more recent breaks while we gather enough data to implement the corrections outlined in the paper. We typically only allow series to have one significant break, however, in the case of areas with large populations we relax this rule.

Supplemental files:

- *Affinity Income Shares - National - 2019.csv*
  - `share_jan2019`: the share of weekly total spending from January 7 to February 3, 2019 by income quartile.
- *Affinity Income Shares - National - 2020.csv*
  - `share_jan2020`: the share of weekly total spending from January 6 to February 2, 2020 by income quartile.
- *Affinity Industry Composition - National - 2020.csv*

  - `share_jan2020`: the share of weekly total spending from January 6 to February 2, 2020 by MCC group.
  - `share_decline_covidfirstwave`: the share of the decline in weekly total spending in the first wave of the pandemic by MCC group. To measure the decline, we first take the sum of total spending in four periods: from January 7 to February 3, 2019, from March 25 to April 14, 2019, from January 6 to February 2, 2020, and from March 30 to April 20, 2020. Then, the decline is calculated as the change in spending between April 2019 and April 2020, minus the change in spending between January 2019 and January 2020. This measures the decline in total spending during the first wave of the COVID-19 pandemic, accounting for the normal decline in spending across the same months in 2019.
- *Affinity Daily Total Spending - National - Daily.csv*
  - `daily_spend_19_all` and `daily_spend_19_q#`: daily total spending indexed to total spending from January 7 to February 3, 2019 by income quartile, not seasonally adjusted.
  - The data in this file is not smoothed using a 7-day moving average, and it does not apply the intensive/extensive margin switching adjustment described in Appendix B.2 of the paper.

2.2 Small Business Revenue

Summary: Small business transactions and revenue data aggregated from several credit card processors. Transactions and revenue are reported based on the county or ZIP code where the business is located.

Data Source: Womply

Update Frequency: Weekly

Date Range: January 15, 2020 until the most recent date available.

Data Frequency: Weekly

Indexing Period: January 4 to January 31, 2020

Indexing Type: Seasonally adjusted change since January 2020. We calculate the change relative to the January index period: 2019 data is indexed relative to January 2019, data in 2020 onward is indexed relative to January 2020. We then seasonally adjust by dividing by the indexed 2019 value, which represents the difference between the change since January 2020 compared to the change since January observed since 2019.

In all series we adjust for holidays while seasonally adjusting a given series to ensure that we are comparing weeks containing a significant holiday to a corresponding week containing that holiday. This adjustment has three components: (1) weeks with the same holiday are always compared to each other; (2) weeks before the same holiday are always compared to each other; and (3) weeks after holidays are always compared to each other. Thus, even if a holiday does not fall in the same week in the year being normed and the comparison year, weeks will be aligned when compared to each other.

Geographies: National, State, County, Metro

Breakdowns:

- *Industry*, by 2-digit NAICS.
  - Health & Social Services
  - Food & Accommodation

  - Professional Services
  - Retail
  - Other Services

- *Business Zip Code Income.* Transactions are linked to ZIP codes where the business is located and ZIP codes are classified into income categories based on measurements of median household income and population provided by the American Community Survey (2014 - 2018).
  - High Income (top quartile of median household income; approximately greater than $78,000 per year)
  - Middle Income (middle two quartiles of median household income; approximately between $46,000 per year and $78,000 per year)
  - Low Income (bottom quartile of median household income; approximately less than $46,000 per year)

Data Masking:

The data we receive from Womply is restricted to businesses that have an annual revenue that is less than the SBA thresholds by industry, and have an average revenue that is within 3 standard deviations of the state average.

We omit counties that don't have a minimum of 3 businesses operating in the first week of January 2020, January 2021, and January 2022. For the county-level series, we mask any counties for which revenue is less than $250,000 during the indexing period (January 4 to 31, 2020): such counties are incorporated into state-level or national-level aggregates but are not reported isolated in the county-level data.

To reduce outliers, we manually exclude some state x industry breakdowns that present extreme variation from our state and national level calculations, as well as a small number of counties that demonstrate extreme variation.

Notes: Subnational 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.

Supplemental files:

- *Womply - ZCTA - 2020.csv*
  - `revenue_all_apr2020`: Percent change in net revenue for small businesses from January 4 to 31, 2020 to March 23 to April 12, 2020, seasonally adjusted.
  - `revenue_all_jul2020`: Percent change in net revenue for small businesses from January 4 to 31, 2020 to June 29 to July 26, 2020, seasonally adjusted.

2.3 Small Businesses Open

Summary: Number of small businesses open, as defined by having had at least one transaction in the previous 3 days.

Data Source: Womply

Update Frequency: Weekly

Date Range: January 15, 2020 until the most recent date available.

Data Frequency: Weekly

Indexing Period: January 4 to 31, 2020

Indexing Type: Seasonally adjusted change since January 2020. We calculate the change relative to the January index period: 2019 data is indexed relative to January 2019, data in 2020 onward is indexed relative to January 2020. We then seasonally adjust by dividing by the indexed 2019 value, which represents the difference between the change since January 2020 compared to the change since January observed since 2019.

In all series we adjust for holidays while seasonally adjusting a given series to ensure that we are comparing weeks containing a significant holiday to a corresponding week containing that holiday. This adjustment has three components: (1) weeks with the same holiday are always compared to each other; (2) weeks before the same holiday are always compared to each other; and (3) weeks after holidays are always compared to each other. Thus, even if a holiday does not fall in the same week in the year being normed and the comparison year, weeks will be aligned when compared to each other.

Geographies: National, State, County, Metro

Breakdowns:

- *Industry*, by 2-digit NAICS.
  - Health & Social Services
  - Food & Accommodation
  - Professional Services
  - Retail
  - Other Services
- *Business Zip Code Income.* Transactions are linked to ZIP codes where the business is located and ZIP codes are classified into income categories based on measurements of median household income and population provided by the American Community Survey (2014 - 2018).
  - High Income (top quartile of median household income; approximately greater than $78,000 per year)
  - Middle Income (middle two quartiles of median household income; approximately between $46,000 per year and $78,000 per year)
  - Low Income (bottom quartile of median household income; approximately less than $46,000 per year)

Data Masking:

The data we receive from Womply is restricted to businesses that have an annual revenue that is less than the SBA thresholds by industry, and have an average revenue that is within 3 standard deviations of the state average.

We omit counties that don't have a minimum of 3 businesses operating in the first week of January 2020, January 2021 and January 2022. For the county-level series, we mask any counties for which revenue is less than $250,000 during the indexing period (January 4 to 31 2020): such counties are incorporated into state-level or national-level aggregates but are not reported isolated in the county-level data.

To reduce outliers, we manually exclude some state x industry breakdowns that present extreme variation from our state and national level calculations, as well as a small number of counties that demonstrate extreme variation.

Notes: Subnational 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.

2.4 Job Postings

Summary: Weekly count of new job postings, sourced from over 40,000 online job boards. New job postings are defined as those that have not had a duplicate posting for at least 60 days prior.

Data Source: Lightcast (formerly known as Burning Glass Technologies)

Update Frequency: Weekly

Date Range: January 17, 2020 until the most recent date available.

Data Frequency: Weekly data points, with each week ending on Friday.

Indexing Period: January 4 to 31, 2020

Indexing Type: Change relative to the January 2020 index period, not seasonally adjusted.

Geographies: National, State, County, Metro.

Breakdowns:

- *Industry*, by NAICS supersector.
  - Educational and Health Services
  - Financial Activities and Services
  - Leisure and Hospitality
  - Manufactoring
  - Professional and Business Services
- *Education Requirement*, by ONET Jobzone's Education Requirement Classification.
  - Minimal - Jobzone 1
  - Some - Jobzone 2
  - Moderate - Jobzone 3
  - Considerable - Jobzone 4
  - Extensive - Jobzone 5

Data Masking: In order to avoid extreme outliers, we calculate a cutoff of one standard deviation above the 97th percentile of the state-level data for each variable and mask values that exceed this threshold. Additionally, at the county level, only subgroup data for the 200 largest counties is able to be disclosed for firm data privacy reasons. For the remaining counties' subgroups, all values are imputed from the share of state postings that are made up of a given subgrouping multiplied by the number of county postings in total.

Supplemental files:

- *Job Postings Industry Shares - National - 2020.csv*
  - `share_jan2020`: the share of job postings by industry (2-digit NAICS) in the period from January 4 to 31, 2020.

2.5 Employment

Summary: Number of active employees, aggregating information from multiple data providers. This series is based on firm-level payroll data from Paychex and Intuit.

Data Source: Paychex, Intuit

Update Frequency: Weekly

Date Range: January 15, 2020 until the most recent date available. The most recent date available for the full series depends on the combination of Paychex and Intuit data.

Data Frequency: Weekly

Indexing Period: January 4 to 31, 2020

Indexing Type: Change relative to the January 2020 index period, not seasonally adjusted.

Geographies: National, State, County, Metro

Breakdowns:

- *Wage.*
  - High Wage (annualized wage greater than 2.5x the federal poverty line, adjusted for CPI inflation within each calendar year)
  - Middle Wage (annualized wage between 1x and 2.5x the federal poverty line, adjusted for CPI inflation within each calendar year)
  - Low Wage (annualized wage lower than the federal poverty line, adjusted for CPI inflation within each calendar year)
  - Above Median (annualized wage greater than 1.5x the federal poverty line, adjusted for CPI inflation within each calendar year)
  - Below Median (annualized wage less than 1.5x the federal poverty line, adjusted for CPI inflation within each calendar year)
- *Industry*, by NAICS supersector.
  - Professional and Business Services
  - Education and Health Services
  - Retail and Transportation
  - Leisure and Hospitality
- *Industry*, by NAICS sector.
  - Retail

Data Masking: As the employment series is a composite series, each of its component series have their own masking standards that in aggregate determine masking for the series.

- *In the Paychex series*, we reduce the weight of cells in which we detect firm entry/exit over time. In each county x industry (two-digit NAICS code) x firm size x wage quartile cell, we compute the change in employment relative to January 4 to 31, 2020, and the change in employment relative to July 1 to 31, 2020. For county x industry x firm size x wage quartile cells between January 2020 and the end of the series, we reduce the weight we place on the series if we observe changes in employment that indicate firm entry or exit.
  - For cells with over 50 employees:
    - We reduce the weight by 2 percentage points for each percentage point of decline we observe below 50 percentage points relative to July 2020.
    - We reduce the weight by 0.5 percentage points for each percentage point of growth we observe above 600 percentage points relative to January 2020.
  - For cells with 50 employees or less:
    - We reduce the weight by 2 percentage points for each percentage point of decline we observe below 50 percentage points relative to July 2020
    - We reduce the weight by 0.1 percentage points for each percentage point of growth we observe above 4000 percentage points relative to January 2020.

  The difference in weighting between small and large cells is to account for large amounts of small firm births, particularly in the second half of 2020, which played a strong role in the economic recovery from the pandemic.

- *In the Intuit series*, we do not make any sample restrictions.

Supplemental files:

- *Earnin - ZCTA - 2020.csv*
  - `emp_incq1_apr2020`: Change in employment level for all workers in the Earnin sample (low-wage) from January 2020 to April 2020.
  - `emp_incq1_jul2020`: Change in employment level for all workers in the Earnin sample (low-wage), from January 2020 to July 2020.

2.6 Unemployment Claims

Summary: Weekly unemployment insurance claims counts and rates (as a share of the 2019 labor force) for all states, as well as initial unemployment insurance claims for select counties where the data is publicly available.

Data Source: State-level and national statistics are reported by the U.S. Department of Labor.

The county-level series is only available for states whose respective state agencies publish county level data:

- Alabama: Alabama Department of Labor
- Arizona: Arizona Commerce Authority
- California: Employment Development Department of California
- Colorado: Colorado Department of Labor and Employment
- Georgia: Georgia Department of Labor
- Hawaii: Hawaii Department of Labor
- Idaho: Idaho Department of Labor
- Illinois: Illinois Department of Employment Security
- Indiana: Indiana Department of Workforce Development
- Iowa: State of Iowa
- Kentucky: Kentucky Center for Statistics
- Maryland: Maryland Department of Labor
- Massachusetts: Massachusetts Department of Unemployment Assistance
- Missouri: State of Missouri
- Nebraska: NEworks (Government of Nebraska)
- Nevada: Nevada Department of Employment; Training and Rehabilitation
- New York: New York State Department of Labor
- Ohio: Ohio Department of Job and Family Services
- Pennsylvania: Government of Pennsylvania
- Washington: Washington State Employment Security Department
- Wisconsin: Wisconsin Department of Workforce Development
- Wyoming: Wyoming Department of Workforce Services

Update Frequency: Weekly (where available, in the case of county-level data)

Date Range: January 18, 2020 until the most recent date available.

Data Frequency: Weekly data points, with each week ending on Saturday.

Note that county-level claims in California, Georgia, Kentucky, and Illinois are reported at the monthly level and imputed to weekly data points for the county-level series. For more information about the imputation methodology, see the **accompanying paper**

Indexing Period: No indexing applied, the published numbers directly report quantities.

Indexing Type: No indexing applied, the published numbers directly report quantities.

Geographies: National, State, County, Metro.

Breakdowns:

- *Initial Claims*
  - Regular Claims
  - PUA Claims
  - Combined Claims
- *Continued Claims*
  - Regular Claims
  - PUA Claims
  - PEUC Claims
  - Combined Claims

Data Masking: No masking is performed by Opportunity Insights, but county-level data is subject to varying masking rules implemented by the state agencies that release the data. For more details, check with the relevant state agency for that state's particular masking rules.

Notes: Unemployment claims rates are calculated by dividing unemployment claims counts by the Bureau of Labor Statistics labor force estimates from 2019.

Under the CARES Act, all states provide 13 additional weeks of federally funded Pandemic Emergency Unemployment Assistance (PEUC) benefits to people who exhaust their regular state benefits. Under the Act, through the end of 2020, some people who exhaust all these benefits, and others who have lost their jobs for reasons arising from the pandemic but who are not normally eligible for UI in their state, are eligible for Pandemic Unemployment Assistance (PUA). "Combined Claims" are defined as the sum of regular, PUA and PEUC unemployment benefit claims.

National totals for all programs' unemployment benefit claims are the sum of the claims counts for all states and DC and exclude other territories such as Puerto Rico and the U.S. Virgin Islands.

2.7 Online Math Participation

Summary: Number of students using Zearn Math, a curriculum from the non-profit Zearn, among schools that already used Zearn Math in course instruction before the pandemic.

Data Source: Zearn

Update Frequency: Weekly, except during summer and winter school breaks.

Date Range: January 6, 2020 until the most recent date available. The data series is not updated during summer or winter school holidays.

Data Frequency: Weekly data points, with each week ending on Sunday.

Indexing Period: January 6 to February 7, 2020

Indexing Type: Change relative to the January 2020 index period, not seasonally adjusted.

Geographies: National, States, County, Metro

To ensure privacy, the data we obtain are masked such that any county with fewer than two districts, fewer than three schools, or fewer than 50 students on average using Zearn Math is excluded. Where possible, masked county levels values are replaced by commuting zone means.

Breakdowns:

- *School Income.* Schools are classified by income based on the share of students in the school eligible for free and reduced lunch based on data provided by Zearn.
  - High Income (35.7% students are free and reduced lunch eligible)
  - Middle Income (56.9% students are free and reduced lunch eligible)
  - Low Income (80.4% students are free and reduced lunch eligible)

Data masking: Data is masked such that any county with fewer than two districts, fewer than three schools, or fewer than 50 students on average using Zearn Math during the period from January 6 to February 7, 2020 is excluded. Masked county level data is replaced with the commuting zone average so long as there are more than two school districts in the commuting zone or at least three schools in the commuting zone. If these condition are not met the county-level data remains masked. Additionally we exclude schools who did not have at least 5 students using Zearn Math for at least one week from January 6 to February 7, 2020.

2.8 Student Progress in Math

Summary: Number of lessons completed by students each week using Zearn Math, among schools that already used Zearn Math in course instruction before the pandemic.

Data Source: Zearn

Update Frequency: Weekly, except during summer and winter school breaks.

Date Range: January 6, 2020 until the most recent date available. The data series is not updated during summer or winter school holidays.

Data Frequency: Weekly data points, with each week ending on Sunday.

Indexing Period: January 6 to February 7, 2020

Indexing Type: Change relative to the January 2020 index period, not seasonally adjusted.

Geographies: National, States, County, Metro

To ensure privacy, the data we obtain are masked such that any county with fewer than two districts, fewer than three schools, or fewer than 50 students on average using Zearn Math is excluded. Where possible, masked county levels values are replaced by commuting zone means.

Breakdowns:

- *School Income.* Schools are classified by income based on the share of students in the school eligible for free and reduced lunch based on data provided by Zearn.
  - High Income (35.7% students are free and reduced lunch eligible)
  - Middle Income (56.9% students are free and reduced lunch eligible)
  - Low Income (80.4% students are free and reduced lunch eligible)

Data Masking: Data is masked such that any county with fewer than two districts, fewer than three schools, or fewer than 50 students on average using Zearn Math during the period between January 6 to February 7, 2020 is excluded. Masked county level data is replaced with the commuting zone average so long as there are more than two school districts in the commuting zone or at least three schools in the commuting zone. If these condition are not met the county-level data remains masked. Additionally we exclude schools who did not have at least 5 students using Zearn Math for at least one week from January 6 to February 7, 2020.

2.9 COVID-19 Infections

Summary: The daily count and rate per 100,000 people of confirmed COVID-19 cases, deaths, hospitalizations, or tests performed.

Data Source: The New York Times, The Johns Hopkins Coronavirus Resource Center, U.S. Department of Health & Human Services, Centers for Disease Control and Prevention

Update Frequency: Daily

Date Range: January 22, 2020 until the most recent date available.

Data Frequency: Daily, presented as a 7-day moving average or 7-day rolling sum

Indexing Period: No indexing applied, the published numbers directly report quantities.

Indexing Type: No indexing applied, the published numbers directly report quantities.

Geographies: National, State, Country, Metro

Breakdowns:

- *New* Cases, Deaths, or Tests (presented as a 7-day moving average for Tests and a 7-day rolling sum for Cases and Deaths)
- *Total* Cases, Deaths, or Tests (presented as a 7-day moving average)
- *Other* Hospitalized (presented as a 7-day moving average)

Data Masking: No masking is performed by Opportunity Insights.

2.10 COVID-19 Vaccinations

Summary: Percentage of the population who have received one or more doses of any COVID-19 vaccine, completed a COVID-19 vaccination series, or received a COVID-19 booster or additional dose.

Data Source: The Centers for Disease Control and Prevention

Update Frequency: Daily

Date Range: February 24, 2021 until the most recent date available.

Data Frequency: Daily, presented as a 7-day moving average for new vaccinations

Indexing Period: No indexing applied, the published numbers directly report quantities.

Indexing Type: No indexing applied, the published numbers directly report quantities.

Geographies: National, State, County, Metro

Breakdowns:

- *New Vaccinations.* Percent of population newly vaccinated with at least one vaccine dose
- *Total Vaccinations.* Percent of population in total vaccinated with at least one vaccine dose
- *New Completed Vaccinations.* Percent of population newly having completed a vaccine series
- *Total Completed Vaccinations.* Percent of population in total having completed a vaccine series
- *New Boosters.* Percent of population newly vaccinated with a booster (or additional) dose
- *Total Boosters.* Percent of population in total vaccinated with a booster (or additional) dose

Data masking: No masking is performed by Opportunity Insights.

Notes: CDC data published prior to February 24, 2021 used a different methodology to assign vaccinations to the state where they were administered, producing numbers that are not directly comparable to those published after February 24.

2.11 Time Outside Home

Summary: Time spent away from home, estimated using cellphone location data from Google users who have enabled the Location History setting.

Data Source: Google COVID-19 Community Mobility Reports, American Time Use Survey

Update Frequency: When released by Google, typically every 4-7 days.

Date Range: February 24, 2020 until the most recent date available.

Data Frequency: Daily

Indexing Period: January 3 to February 5, 2020

Indexing Type: Change relative to the January 2020 index period, not seasonally adjusted.

Geographies: National, State, County, Metro

Breakdowns:

- Time Away From Home
- Retail and Restaurants
- Transit
- Parks
- Grocery
- Workplace

Data Masking: Google does not release data for geographies where their internal quality and privacy thresholds are not met. Therefore some geographic areas are omitted from the series for certain breakdowns and certain dates.

Notes: When data is missing for 1 or 2 consecutive days we linearly interpolate the missing values and construct the 7 day moving average including these interpolated values. If data is missing for 3 or more consecutive days, the corresponding 7 day moving average is also recorded as missing whenever it overlaps with the missing data.

Time Away From Home is calculated by multiplying the mean time spent inside home from the American Time Use Survey by the percent change in time spent at residential locations reported by Google. For more information about this imputation, see the **accompanying paper**.

2.12 Policy Milestones

Summary: Key state-level policy dates relevant for changes in other series trends and values. Includes start and end of stay at home order dates, public school closure dates, and non-essential business closure and re-opening dates.

Data Source(s): New York Times, MCH Strategic Data, the Institute for Health Metrics and Evaluation, and local news and government sources.

Update Frequency: This file is not being updated with data beyond June 30, 2022.

Geographies: State