

US Microbusiness Data

Dictionary

Last updated August 2024

Below is a list of the columns contained within the files that contain all the microbusiness count and density information. For ease of use, these files are split into the city center-, CBSA-, county-, and state-level files.

All US geographic levels currently use Total Population figures from Census.gov (ACS-5yr) to calculate microbusiness density. Venture Forward previously used Population 18 & Over to calculate microbusiness density with the exception of the City center-level files, which have used Total Population since of mid-2023.

*active**: A raw count of the number of active microbusinesses in a community, defined by GoDaddy as microbusiness with an active website. This is an integer derived by summing across the zip codes within a county or CBSA.

cbsa: A unique numeric identifier (up to 5 digits) for each Core-Based Statistical Area (CBSA). These are the micropolitan and metropolitan areas defined by the U.S Department of Housing and Urban Development (HUD). HUD crosswalk files are used for each update, making the geographic area defined the most up-to-date as possible.

cbsa_name: In a string, the legal description of the CBSA denoted by the row, as understood by HUD.

cfips: A unique numeric identifier (up to 5 digits) for each county. The first two digits represent the state Federal Information Processing System (FIPS) code, while the last 3 digits are assigned to each of the counties within each state, as defined by the Department of Housing and Urban Development (HUD). HUD crosswalk files are used for each update, making the geographic area defined the most up to date as possible.

city: In a string, the legal description of the municipality denoted by the row, as understood by the U.S. Postal Service (USPS).

city_name: In a string, the legal description of the CBSA denoted by the row, as understood by HUD.

city_id: A unique numeric identifier (up to 5 digits) for each city center/municipality, generated by GoDaddy for ease of data wrangling.

county: The legal description in a string of the county denoted by the row, as understood by HUD.

county_name: The legal description in a string of the county denoted by the row, as understood by HUD.

date: The month of observation for the associated row of data, one per each geographic entity

fips: The first two digits representing the state in the Federal Information Processing System (FIPS) code, as defined by the Department of Housing and Urban Development (HUD).

MAI_composite: The overall composite index score for the associated geography in the associated rowdate.

engagement: The engagement sub-index score for the associated geography in the associated rowdate. This measures how active websites are in the community.

participation: The participation sub-index score for the associated geography in the associated rowdate. This measures the number of online microbusinesses and operators in the community.

infrastructure: The infrastructure sub-index score for the associated geography in the associated rowdate. This measures how prepared a community is to use the internet, in terms of necessary physical and intellectual infrastructure.

md: The number of microbusinesses per 100 people in the geographic unit, during the month described by the row's data. Updated through the end of June 2024.

topind_shoppers: Character string denoting the industry which had the largest number of active merchants for the month denoted by the row's year and month combined.

micro_metro: A flag written in string to denote whether a CBSA is a micropolitan (under 50k population but over 10K) or a metropolitan statistical area.

pop_18over_YYYY: Adult population 18 years of age and older for the given year YYYY. This is the figure used to calculate density for all geographic units.

state_abbrev: In a string, the legal description of the state denoted by the row, as understood by the USPS.

state_name: Full state name.

total_pop_2021: Total population for a given geographic area in 2021. This is not the population figure used to calculate density anywhere.

topind_orders: Character string denoting the industry which had the largest number of completed transactions/orders for the month denoted by the row's year and month combined.

topind_sales: Character string denoting the industry which had the largest total sales in US dollars for the month denoted by the row's year and month combined.