

BDS tables on DATA.CENSUS.GOV versus BDS CSV DATASETS

This document describes some of the key differences between the BDS data tables on data.census.gov versus the BDS CSV datasets on the census.gov website. Users should note that the tables on data.census.gov are sourced from the [BDS API](#) and represent selected views of the API data. Therefore, the differences outlined below between the tables on data.census.gov and the CSV datasets apply to the BDS API as well.

1. Missing Alaska (AK) Counties

There are 6 counties in AK that appear on the CSV datasets but do not appear on the data tables on data.census.gov or in the BDS API. These 6 AK counties include:

- 02-010 (Aleutian Islands, AK)
- 02-140 (Kobuk, AK)
- 02-201 (Prince of Wales-Outer Ketchikan, AK)
- 02-231 (Skagway-Yakutat-Angoon, AK)
- 02-232 (Skagway-Hoonah-Angoon, AK)
- 02-280 (Wrangell-Petersburg, AK)

The reason these 6 counties appear on the CSV datasets and not on data.census.gov/API is because of the difference in how the county codes are validated for each data product. The CSV datasets, in general, change vintages of county (and state) codes over time, following the timing of changes in the vintages used by the source files from the Business Register and the County Business Patterns (CBP) program. These county and state codes on the CSV files are then validated against the historical vintages of county and state codes used in the CBP program, see: <https://www.census.gov/programs-surveys/cbp/technical-documentation/reference/state-county-geography-reference.html>. As long as a given county and state code in a given year on the CSV datasets was valid at some point in the historical vintages used in the CBP program, it is considered valid for publication in the CSV files. The data on data.census.gov/API, however, while sourced from the CSV files, validate county and state codes based only on the 2017 vintage of these codes, and not the historical vintages used by the CBP program. Some of the historic county codes that appear in the CSV datasets are no longer valid in 2017 and therefore do not appear on the data.census.gov tables/API. The 6 counties in AK listed above are examples of codes that are no longer valid in 2017.

2. NAICS 5250

The CSV datasets with the 4-digit industry detail include the NAICS code 5250. The 5250 code is not, however, included in the tables on data.census.gov/API. This code is not included in these tables or on the API because it is not a valid 4-digit NAICS code—as listed in the official [2012 NAICS codes](#)—and therefore cannot be validated against the NAICS coding system used for all tables on data.census.gov and the API. The 5250 code does, however, appear on the CSV datasets. This is because of the unique vintage-consistent NAICS processing used by the LBD/BDS, which relies on data from the Economic Census. The Economic Census, moreover, has different industry scope rules than the BDS (and CBP).

One of these differences is the 3-digit NAICS code 525, which is out-of-scope for the Economic Census, but in-scope for BDS and CBP. The vintage-consistent NAICS processing used by the BDS, therefore, has to have a special way of handling codes such as NAICS 525 that are out of scope to the Economic Census. Currently, the system converts any code that starts with 525 to 525000. The 4-digit version of 525000 is 5250, which is what gets published in the 4-digit NAICS datasets in the BDS. Data users should note that in the BDS CSV datasets, the data for 5250 is equivalent to the data for 525.

3. Geocoding

The BDS CSV datasets use some geographies and geographic coding that are not part of the standardized system used by the tables on data.census.gov and by the Census API. The 'non-standard' geographies/coding on the CSV datasets, therefore, had to be translated to fit the system used by the tables on data.census.gov and by the Census Bureau API. The following is a list of the non-standard geographies/coding that appear on the CSV datasets and how they are translated to appear on data.census.gov and on the API:

- Code '999' is the 'statewide' designation in the county CSV datasets and is represented by the 'statewide' designation in the state tables on data.census.gov
- Code '998' is the 'unclassified designation' in the county CSV datasets and is represented by the 'location not specified' designation in the state tables on data.census.gov
- Code 'SW' is the 'statewide' designation in the metro CSV datasets and is represented by the 'statewide' designation in the U.S. tables on data.census.gov
- Code '97' is the 'nationwide' designation in the state CSV datasets, and is represented by the 'location not specified' designation in the U.S. tables on data.census.gov.