Blended Data in the Census Bureau’s Monthly State Retail Sales Data Product

November 8, 2021

Stephanie Studds, U.S. Census Bureau

Disclaimer: Any views expressed are those of the author and not necessarily those of the United States Census Bureau.

Census Bureau has reviewed Monthly State Retail Sales product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied. (Approval ID: CBDRB-FY21-ESMD002-033)
Delivering on data users’ requests

• More timely state-level retail sales are among the most requested data by our data users.

• In September 2020, the Census Bureau released the new blended Monthly State Retail Sales (MSRS) data product.

• First version of these experimental data.

• MSRS was created using existing survey data, administrative data, and third-party/alternative data sources as its inputs. No new data was collected.
Total Retail Sales Excluding Nonstore Retailers by State

Source: July 2021 MSRS Report
S = Estimate suppressed due to quality concerns
* The 90 percent confidence interval includes zero. There is insufficient statistical evidence to conclude that the actual change is different from zero.
Note: State retail sales data not adjusted for seasonal variation, trading-day differences, moving holidays or price changes.
Blended data approach

**Admin**
- Administrative Data:
  - Gross payroll for retailers

**Survey**
- Monthly Retail Trade Survey (MRTS) Data:
  - Monthly retail sales data

**3P**
- Third-Party Point-of-Sale Retail Sales Data
  - Monthly sales by store location for a set of retailers
  - Monthly sales by state for curated groupings of retailers
Lessons Learned: Alternative Data Source Use

• Maintain realistic expectations
• Assess trade-offs between granularity and scalability
• Accept that off-the-shelf data products will rarely align with Census data
• Maintain good third-party data provider relationships
Maintain realistic expectations

• Lessons Learned
  • Third-party data providers can overpromise what data they have available, how well it aligns to Census data, and what data they are able share. They will also try to package their proprietary tools with their data.
  • Proofs of concept or sample datasets are a great way to obtaining a good understanding of the data before committing to a purchase.
  • Do not go into a project thinking a single dataset will be a solution.
  • Consider a blended data approach. Blended datasets will provide more coverage, more protection against the creation of competing data products, and lower risk of negative impacts of losing an input data source.
Assess trade-offs between granularity and scalability

• Lessons Learned
  • The most granular data can be quite useful but also comes with both cost and quality concerns.
    • Individual company data for an entire sector is not a scalable effort.
    • For satellite imagery, more frequent imagery is available but must accept lower resolution images.
  • In products where granularity of the third-party data itself is less important (Monthly State Retail Sales), we worked with data providers to obtain alternative datasets.
  • Finding data providers who will allow us to customize data aggregations has been important.
Accept that off-the-shelf data products will rarely align with Census data

• Lessons Learned
  • Precision in contractual specifications is critical, including asks for comprehensive methodology documentation, dataset formats, and timing of deliveries to be useful inputs to our data products. Get as many sample files before the contract as possible.
  • Collaborative third-party data providers will work with us to modify the data to better meet Census needs but there will always been differences that cannot be addressed.
  • Third party data does not align to the federal statistical system and standards. A heavy lift is required to fully understand methodology, quality, and fitness for use.
Build good third-party data provider relationships.

• Lessons Learned
  • Seek out data providers who are transparent and provide detailed answers to questions.
  • Find data providers that are willing to collaborate with us to improve the dataset to meet our needs.
  • Third party data products are also improved due to the relationship with Census.