

Data Analysis of US Economic and Business Data

US Census/BFS Monthly Series vs. Business Applications (US & Puerto Rico)

Prepared for: Data Science Blog
By: Churamani Paudel
August 22, 2025

This post distills an end-to-end exploration of multiple U.S. economic datasets—weekly date tables, monthly Business Formation Statistics (BFS) for the United States and Puerto Rico, and a separate Business Applications file. We focus on: (1) reproducible cleaning; (2) trend discovery in BFS; and (3) a careful, transparent comparison with Business Applications for the one month of overlap.

DATA OVERVIEW & PREPARATION — DATASETS

- `date_table.csv` — weekly date ranges with year/week indices; some years include 53 weeks.
- `bfs_monthly.csv` — monthly Business Formation Statistics (BFS) for the United States across sectors/series.
- `bfs_monthly_pr.csv` — monthly BFS for Puerto Rico across sectors/series.
- `business_applications.xlsx` — Business Applications by application type and region (required heavy cleaning).

PREPARATION & CLEANING

- Business Applications: set the correct header row, drop metadata rows, and reshape (melt) to long format.
- Create a proper datetime index for Business Applications to enable time-series work.
- BFS (US and PR): melt wide-to-long, standardize series identifiers (e.g., `BA_BA`, `BF_WAVERING`), and build a monthly datetime index.
- Build consistent keys (e.g., month-level datetime) for cross-dataset alignment.

KEY FINDINGS — DATE TABLE

- Weekly coverage extends through 2025 with ISO-like week indices.
- Years with 53 weeks are present; downstream joins should prefer explicit date boundaries over naive week numbers.

KEY FINDINGS — BFS MONTHLY (US & PUERTO RICO)

- Total Business Applications (BA_BA) trend upward from 2004 onward, with cyclical fluctuations.
- BF_WAVERING (likelihood of becoming an employer) varies over time and remains well below total BA_BA levels.
- US and Puerto Rico differ in levels and dynamics, suggesting structural/regional differences.

KEY FINDINGS — BUSINESS APPLICATIONS

- After cleaning, Business Applications only contained a single usable month (July 2025), limiting time-series analysis.

COMBINED ANALYSIS (LIMITED OVERLAP)

Due to minimal overlap, we avoided forced merges. For July 2025—the single overlapping month—descriptive statistics expose a large magnitude gap between Business Applications and BFS BA_BA. This indicates non-equivalent definitions, coverage, or counting methods.

JULY 2025 SNAPSHOT (AS PROCESSED)

- Business Applications (total) — 941,142 (July 2025).
- BFS BA_BA (US) — 4,258 (July 2025, as processed in this analysis).
- BFS BA_BA (Puerto Rico) — 4,258 (July 2025, as processed in this analysis).
- High-Propensity Applications — 336,310 (Business Applications, July 2025) vs. much smaller BFS counts (e.g., BF_WAVERING).

DISCREPANCIES & NOTABLE DIFFERENCES

- Definitions/units differ: Business Applications likely aggregates broader activity (and/or different units) than BFS BA_BA.
- Temporal granularity mismatch (weekly vs. monthly) complicates alignment.
- Limited historical depth in Business Applications blocks robust cross-series validation.

LIMITATIONS

- Only one overlapping month after cleaning Business Applications.
- Potential differences in seasonal adjustment, revisions, or unit scaling between sources.
- Lack of a published, unified data dictionary spanning both products.

CONCLUSION & NEXT STEPS

In short: BFS reveals long-run growth in business formation, while the Business Applications file—as provided—offers only a one-month snapshot. The magnitude differences in July 2025 reinforce the need to reconcile definitions and align granularity before drawing cross-source conclusions.

RECOMMENDED NEXT STEPS

- Acquire additional historical months for Business Applications to enable rolling comparisons.
- Obtain methodology notes and precise definitions for each series (BFS vs. Business Applications) to reconcile units and scope.
- Harmonize granularity: aggregate weekly series to monthly (or vice versa) before comparison; document choices.
- Test merges on richer keys (e.g., state/region × sector × month) once definitions align.

REPRODUCIBILITY NOTES

- Notebook contained ~34 hidden cells covering: header fixes, row drops, long-format melts, and date indexing.
- BFS: melted to long, standardized identifiers (BA_BA, BF_WAVERING), and created a monthly datetime index.
- Business Applications: established proper header row, removed metadata, melted to long, and built a datetime index.
- Comparison restricted to July 2025 given available scope after cleaning.