

Methodology

This dashboard uses data from various federal statistical agencies and proprietary data aggregation firms. This document explains Pew’s methodology and the data used for each chart.

Delinquency on bills; Financial stability; Balance on credit accounts

These charts are based on data from [Experian](#), which assembles data on tens of millions of commercial entities from a wide variety of sources to generate credit reports and marketing stats for sale to lenders, marketers, business analysts, researchers, and many others. Experian’s data sources include creditors, such as lenders and suppliers; third-party data providers, such as payment processors; government agencies and courts, including federal bankruptcy courts and state regulatory agencies; other aggregators and processors of proprietary data, including partners and subcontractors; AI-powered web and social media searches; and business entities. Experian says it subjects the data to continuous quality control and verification processes.

Experian refers to its financial data generally as “credit” data and updates it constantly, based on business activity. Experian receives daily, monthly, and quarterly feeds from many credit sources and generates its own analyses, credit scores, and ratings using its proprietary algorithms. It codes each entity’s data under Experian’s own business identification number, or BIN, a nine-digit unique identifier that it assigns to each business for data linking and verification purposes.

Experian refers to its descriptive company data—corporate structure, locations, executive names, industry sector, employment details, open hours, e-commerce activity, and many other details—as “marketing” data and updates it monthly. Experian and its data partners use various methods, including call centers, to verify companies’ status and data. Where fields are missing data, such as number of employees, Experian may impute figures from other sources or models. The resulting records are tagged as verified, pre-verified, suspected invalid, and verified out of business. Pew’s dashboard uses verified records from the marketing data.

Experian then matches records across the credit and marketing datasets, resulting in comprehensive portraits of firms with hundreds of possible data points, including Experian’s proprietary ratings where data was available. Experian makes the records available for purchase through self-service online portals; Pew uses the TargetIQ portal intended for marketing clients. TargetIQ is searchable by many variables, including the city where a business has a physical location. Pew downloads every available verified business record with a physical address in Philadelphia at the end of every month, effectively a snapshot of all data updated (or not) since the prior month. Pew’s initial download was on Jan. 3, 2021.

In addition to TargetIQ, Pew acquired archival data—including credit and marketing data—from Experian at four quarterly points in 2019 and 2020 in order to include in the dashboard the first year of the pandemic. The archival data files contained many more records than those obtained through TargetIQ, requiring a process for reconciling the datasets and making a uniformly defined sample across periods. First, Pew cross-matched BIN codes in each period with BIN codes found in the immediately prior and subsequent periods. Then it tagged each record as being found or not found in the adjacent time period files. In the ongoing dashboard, Pew limits all chart calculations to records tagged as found in one prior period. This creates a uniform set of records across at least two periods on a rolling basis,

including between archival and TargetIQ datasets, ensuring that any company being measured has been in existence for at least three months.

Pew's interactive dashboard allows users to filter the data by four fields:

1. Industry sector at the two-digit level in NAICS (the North American Industry Classification System, which goes to six digits).
2. Location by ZIP code of Philadelphia physical address.
3. Size of business by number of employees at a Philadelphia location.
4. Gender of the top-listed executive by given name.

Almost all records in the source dataset have data in these fields, except for gender data, which is missing from about two-thirds of the records. Pew marks all such blank fields as "unknown." When records contain credit data but no descriptive marketing data (i.e., "unknown"), Pew still uses the credit data in its total aggregated results but does not allow dashboard users to filter for "unknown."

To protect the confidentiality of businesses, and under conditions set by Experian, Pew publishes only aggregated summaries of the data and anonymizes all data in the dashboard. It also suppresses any result in which a filter produces fewer than 20 records, similar to the U.S. Census Bureau's practice of suppressing some data in its ZIP Codes Business Patterns releases.

Delinquency on bills

This chart shows the percentage of establishments, out of the citywide total, that are 31 or more days late, on average, paying their invoiced due amount. It is based on Experian's Days Beyond Terms (DBT) measurement, which is the dollar-weighted average number of days beyond the due dates that a business pays its bills, based on creditor accounts that were updated in the previous three months. DBT is calculated starting at the due dates set in contracts or invoices (i.e., "terms") across all accounts, weighted by the total amounts owed, meaning that a DBT of 5 can represent less than or more than five actual days, depending on the amount owed. In the TargetIQ dataset, the DBT field is called "DBT average" and includes many zero values, which can mean either that the company is current on its bills or that it was not evaluated by Experian. Another field, "DBT summary," indicates whether the record was evaluated; Pew uses that summary field to sift the average field, leaving true zero values that signify that a company is current on its bills. The archival data doesn't have a DBT summary field, so Pew uses the DBT summary field in the 2020-Q4 data file from TargetIQ to perform the same sifting process on the archival records, matched by BINs. (Related information on DBT is here: <http://www.bci2experian.com/credit-report/understanding-dbt-days-beyond-terms.>)

Financial stability

This chart provides a percentile figure ranging from zero to 100 representing Experian's statistical prediction of a firm's risk of default or bankruptcy within 12 months, based on Experian's analysis of firms' trajectories following instances of high delinquency and other events. The Financial Stability Risk (FSR) score is based on a firm's record over the preceding one to six years on the following factors: delinquency at more than 61 and/or more than 91 days late; accounts placed for collection; high credit use compared with previous high balances; derogatory public record filings such as liens, judgments, and prior bankruptcy filings; current bankruptcy filings; and data on the business itself, such as its age and the condition of its industry sector. Pew's chart calculates the average FSR score, where available,

for all Philadelphia establishments, excluding those with FSR values of 999 or 998, which denote insufficient data for regular scoring. (Related information on FSR is here: <https://www.experian.com/assets/business-information/brochures/financial-stability-risk-score-ps.pdf>.)

Balance on credit accounts

This chart shows the total amount that a company owes on all creditor accounts, called “tradelines” by Experian, for which the combined balance is greater than \$0. The data includes balances in accounts with activity within the past seven years (the statute of limitations for collection action) as well as accounts more than seven years old, regardless of whether the creditor was actively seeking payment. Pew’s chart shows the median of all establishments’ total account balances. It seeks to show only balances attributed to Philadelphia locations while excluding those of parent companies or headquarters, which would not reflect conditions in Philadelphia. To do so, Pew excludes from its median calculation all establishments that Experian explicitly identifies as “branches” of larger corporations. It also excludes establishments if several carry the same BIN, which in TargetIQ typically links to the parent company’s credit balance.

Hourly wage by sector; Work locations by sector

These charts are based on data from the U.S. Bureau of Labor Statistics’ Quarterly Census of Employment and Wages (QCEW), found at <https://www.bls.gov/cew/data.htm>. BLS compiles the data from state unemployment insurance (UI) systems, which are comprehensive sources of labor trends among employers with UI accounts, as required by law in all states. Detailed information on QCEW methodology is available here: <https://www.bls.gov/cew/questions-and-answers.htm>.

Pew’s “hourly wage by sector” chart uses the QCEW field “avg_wkly_wage” divided by 40 hours, which is the standard workweek for U.S. full-time workers, on average, then calculates the percent change between the current quarter and the comparable quarter in pre-pandemic 2019. The data includes industry sector designations at the two-digit level in NAICS. Detail on QCEW wage data can be found at <https://www.bls.gov/cew/publications/employment-and-wages-annual-averages/current/home.htm>.

Pew’s “work locations by sector” chart uses the QCEW field “qtrly_estabs” and calculates the percent since March 2020, because the sector designations do not change as frequently as wages do. The data includes industry sector designations at the NAICS two-digit level. Detail on QCEW’s establishment data can be found at <https://www.bls.gov/cew/publications/reporters-guide>.

Business sales by sector

This chart is based on the sales and use taxes (SUT) that businesses, categorized by their primary two-digit NAICS code, owed to the commonwealth of Pennsylvania for their transactions in Philadelphia in a given quarter. The tax liability figures reflect just Philadelphia’s 2% addition to the state’s 6% sales tax. In each quarter, we compute the percentage difference in tax liability since the fourth quarter of 2019, the last full quarter before the COVID-19 outbreak. For confidentiality reasons, the state Department of Revenue suppresses data when the number of businesses reporting a tax liability is four or fewer. To meet this suppression threshold and still show the sales trends, tax data for the arts, entertainment, and recreation (sector 71) and accommodation and food services (72) are combined into a “supersector,” leisure and hospitality. In addition, data is combined for agriculture (11), mining (21), utilities (22), and public administration (92), and all unclassified sectors are grouped into the “all others” category. The

latest full list of goods and services subject to SUT in Pennsylvania was last published in May 2022 and can be found at

<https://www.revenue.pa.gov/FormsandPublications/FormsforBusinesses/SUT/Documents/rev-717.pdf>.

More recent changes in the taxable list can be found at

<https://www.revenue.pa.gov/TaxLawPoliciesBulletinsNotices/TaxSummaries/Pages/default.aspx>.

Jobs by sector

This chart uses the U.S. Bureau of Labor Statistics' state and metro area employment data, found at <https://www.bls.gov/sae> and produced by the BLS Current Employment Statistics program (CES). BLS collects the data in ongoing surveys and analyses, and translates the results down to the county level with a two-month lag, typically revising the prior one or two periods with each new release. Details on its data and methods can be found at <https://www.bls.gov/sae/overview.htm>.

Pew's "jobs by sector" chart uses the monthly job estimates that are not seasonally adjusted, calculating the percent change from the comparable period in pre-pandemic 2019.

The CES data is available at the supersector level and only some NAICS two-digit levels, enabling Pew to manually disaggregate most of the supersectors and show them at the NAICS two-digit level. The one exception is the "Trade, Transportation, and Utilities" supersector, which cannot be fully disaggregated using the CES data. So Pew estimates the breakdown using the BLS QCEW data. Pew first calculates the percentage of "utilities" and "transportation and warehousing" portions of the total found in QCEW. Then it applies those percentages to the CES data, rounding to the nearest 100. Since QCEW's release lags CES' release by five to six months, Pew projects the latest QCEW percentages forward using the trend found in the CES supersector, labels it as an estimate, and revises all figures with each data release.

Jobs by establishment size

This chart is based on data from the U.S. Census Bureau's Quarterly Workforce Indicators (QWI), a project of the bureau's Longitudinal Employer-Household Dynamics program, run by its [Center for Economic Studies](#). Data originates from each state's labor market information agency's unemployment wage records and businesses each quarter, and is then merged with census demographic information on each area.

Pew's "jobs by establishment size" chart uses QWI filters as follows: indicator set to "Beginning of Quarter Employment Counts"; geography set to "Philadelphia County"; X-axis set to "Quarters"; group set to "Firm size." Data is calculated as the percent change since the comparable period in pre-pandemic 2019. Detail on QWI explorer data can be found at <https://lehd.ces.census.gov/doc/FirmAgeAndSizeOnePager.pdf>.

QWI's establishment size categories differ from those used by other government agencies and by Pew in other dashboard charts. In this chart only, Pew defines small as 1-49 employees and midsize as 50-499 employees, leaving large at 500 or more.

Bankruptcy filings

This chart uses data provided by the U.S. Administrative Office of the Courts and accessible through

the court's data portal PACER. First, Pew goes to the separate PACER page for each of the following court divisions, in case a Philadelphia-based company files in a neighboring jurisdiction: U.S. Bankruptcy Courts for Southeastern Pennsylvania (Philadelphia and Reading divisions), Delaware (Wilmington), and New Jersey (Camden and Trenton). Then, in PACER, Pew applies the following filters:

- Case type: bk
- Chapter: 7, 11, 12, 13
- Date field: "filed date"
- Open cases: checked
- Closed cases: checked
- Party info: checked
- Nature of debt: business
- Include dismissed cases: checked

All other fields are left blank. The resulting downloaded data includes multiple records for each case, including one for each lawyer assigned to a case. Pew removes as many of these duplicate party records as possible by cross-matching the main debtor's data with the main case filing number. Still, some

bankruptcies that may have been instigated by a single entity may cause many related entities to file at the same time. Because Pew cannot easily discern which debtor may be the main one, Pew leaves all related-entity filings in the data.

Pre-pandemic citywide and ZIP code area profiles; Pre-pandemic business-owner profiles

Data used in these charts comes from three U.S. Census Bureau products: ZIP Codes Business Patterns (ZBP); American Community Survey five-year sample Table DP05 (DP05) and Table DP03 (DP03); and Annual Business Survey (ABS).

All businesses per 1,000 residents age 16+: The total number of business establishments in the ZIP code area (ZBP), divided by the total population age 16 or older in the ZIP code area (DP03), multiplied by 1,000. To use the ZBP annual data with the five-year ACS sample, we use an average of five years of ZBP data over the same period covered by the ACS sample.

Small businesses per 1,000 residents age 16+: The total number of business establishments with 500 or fewer employees in the ZIP code area (ZBP), divided by the total population age 16 or older in the ZIP code area (DP03), multiplied by 1,000. To use the ZBP annual data with the five-year ACS sample, we use an average of five years of ZBP data over the same period covered by the ACS sample.

Average establishment size, in employees: The total number of business establishments in the ZIP code area (ZBP), divided by the total number of employees at those establishments within the ZIP code area.

Self-employed working business owners per 100 workers: The percentage of the civilian employed population age 16 or older that is self-employed in a nonincorporated business within the ZIP code area (DP03).

Labor participation rate: The percentage of the population age 16 or older in the labor force in the ZIP code area (DP03).

Median household income: The total median household income in inflation-adjusted dollars in the ZIP code area (DP03).

Establishment ownership (citywide): The number of establishments owned by each racial and ethnic group, divided into total establishments classifiable by race (ABS).

Sales revenue (citywide): The annual sales revenue at establishments owned by each racial and ethnic group, divided into total sales revenue at establishments classifiable by race (ABS).