sageworks

The Leader in the Financial Analysis of Privately-Held Companies

About Our Data

Sageworks strives to provide you with the most representative, relevant, and useful data by maintaining our Sageworks Database platform with a uniform, systematic, and statistically valid filtering system. We believe that useful data must be representative, relevant, and timely—our filtering system ensures that the industry averages used throughout our products present you with representative information that accurately reflects the economic conditions of your industry for your financial analysis, whether you segment the data by geographic region, sales range, or data source. First, this paper describes how our cooperative model provides you with timely data. Second, this paper outlines the method behind our filtering process.

Benchmark Data

The Sageworks Database allows access to real-time benchmark data that we collect directly from our ProfitCents and Sageworks Analyst users through our cooperative data model. This allows us to offer straightforward industry data to our clients who are primarily interested in trending performance metrics for industries and mining industry data through queries.

When customers run reports through our ProfitCents and Sageworks Analyst programs, we collect and store that data. We do not divulge identifiable financial information from the specific companies that our users enter into our software. Sageworks Database only displays aggregated and anonymous industry-level data. Each day, our customers run over one thousand reports and we incorporate that new data into our industry statistics every morning. Thus, the data that you receive in your report are live.

We believe the data from our cooperative model to be the best source of private-company industry data available because they are entered into the system by financial professionals who have an incentive to create the most accurate analyses possible for the business. Despite this incentive, our data are only representative of those private companies that hire financial professionals and, thus, may be limited in scope and/or biased towards more positive results. This potential limitation notwithstanding, the primary purpose of Sageworks Database is to provide insight into the private sector of the economy—for example, our data are not meant to be predictive of industry or stock market trends.

The Five Filters Method

Our five filters method omits inaccurate data from our industry averages. Inaccurate data are neither representative nor relevant—thus, they do not aid your financial analysis. We guard our data for accounting anomalies, input errors, nonsensical testing data, account sample and demo data, and other data inaccuracies that skew our reported averages and common-sized values. Before accepting industry data into our database, we screen them using the five filters outlined below:

Filter 1. In the first pass, all data are filtered to ensure the validity of the values we collect. We have based these filters on the way that users input financials into our system.

1. No duplicates are included—each set of data must be from a unique business with a unique fiscal period and period length for that business.
2. No data marked as sample data, employee data, or with an unknown data source are allowed.
3. No missing elements are permitted—industry, region, and time period for each financial statement are required.
4. Accepted data must be from audited, reviewed, compiled or tax return-based financial statements.
5. No periods that have been annualized in the program are allowed.

6. Data are not included if they were entered in the “None of the above” industry category or if no industry was indicated.
7. Data that was entered by a customer who chose to opt out of our data collection and aggregation feature are excluded.
8. No data with an invalid date or a future period end date are allowed.

Filter 2. In the second pass, we eliminate data that violate commonsense accounting rules. We based the following filters on fundamental accounting rules.

1. Net Profit before Taxes cannot be greater than Sales.
2. Adjusted Net Profit before Taxes cannot be greater than Sales.
3. Net Income cannot be greater than Sales.
4. Payroll, Rent, Advertising, Overhead, and Depreciation cannot be less than 0.
5. Cash, Accounts Receivable, and Inventory must be less than or equal to Total Current Assets.
6. Total Current Assets and Net Fixed Assets must be less than or equal to Total Assets.
7. Accounts Payable and Other Current Liabilities must be less than or equal to Total Current Liabilities.
8. Total Current Liabilities must be less than or equal to Total Liabilities.
9. Quick Ratio must be less than or equal to the Current Ratio.
10. Gross Profit cannot be greater than Sales.
11. Interest, Other Operating Income, Other Operating Expenses, Other Income, Other Expenses, Extraordinary Income, and Extraordinary Expenses cannot be less than 0.
12. Other Income / Sales and Other Expenses / Sales cannot be greater than Sales.
13. Taxes Paid / Sales cannot be greater than 1.
14. Other Current Assets must be less than or equal to Total Current Assets.
15. Accumulated Depreciation cannot be greater than Gross Fixed Assets.
16. Other Assets must be less than or equal to Total Assets.
17. Current Portion of Long-Term Debt must be less than or equal to Total Current Liabilities.
18. Long Term Liabilities must be less than or equal to Total Liabilities.
19. Ending Retained Earnings must be less than or equal to Total Equity.

Filter 3. We guard against incomplete financial entry—for example, if a customer runs a report wherein all input fields are 0, then we classify this dataset as incomplete and thus inaccurate data. A set of financials with all 0s does not make sense and we omit that report’s values as to not skew our data by lowering our averages with inaccurate 0s.

Filter 4. In the fourth pass, we omit all data that our exception rules generate. Exception rules, which account for, among others, 0s in a ratio’s denominator, ensure that we properly display relevant graphs and N/As throughout our products; however, they are not relevant data points per se. We do not believe that ratios generated from exception rules present actionable and representative data for our users—for example, a company with a current ratio of 15500/0 does not present a useful current ratio for that company’s industry. We have listed all ratios generated from exception rules in *Appendix B*.

Filter 5. In the final pass, we omit outliers from our reported averages. *Appendix A* presents the calculations behind this filter. Suppose, for example, we had dataset where an industry’s current ratios were 1.2, 1.1, 1.05, 2.5, 2.9, and 15. In this context, that current ratio of 15 would be an outlier that skews the reported current ratio mean from 1.75 to 3.96—this is a substantial leap for a current ratio. Although that 3.96 is the straight average of this industry’s current ratio, that outlier of 15 skews the reported mean so much that the mean would neither be representative nor actionable. Moreover, although 15 is a valid current ratio that one of our customers could have entered, including it in the industry average would not yield accurate financial analysis.

To filter these outliers, we use our two pass standard deviation method to create upper and lower bounds at the five-digit NAICS level—any values that we collect through our cooperative model that fall outside these upper and lower bounds we identify as outliers and we omit from the reported mean. This does not affect how we calculate the SIDA data, which calculates a new mean from the data within the

bounds each time you run a report. Finally, calculating outlier bounds at the five-digit NAICS level ensures that the outlier bounds accurately reflect each industry.

Before we calculate our bounds, we duplicate the five-digit NAICS level dataset and transform that dataset to approximate a normal distribution using a natural logarithmic transformation, which we found to yield the most consistently normal distributions. We then implement an offset function, which combines the logarithmic transformation with a linear transformation, to normalize datasets with 0s and negative values. We then run our first standard deviation script on this normalized dataset to calculate our dataset's lower and upper bounds

After we calculate bounds and omit the values that fall outside those bounds, we duplicate the dataset and repeat the process again, finally arriving at our industry-specific bounds. This is our two pass method. The second pass picks up outliers that our first pass did not. This is so that we can calculate a more representative average from the dataset that falls within the first standard deviation of any dataset's frequency distribution.

It is important to note that we only omit data—we never delete them from the database. We incorporate that omitted data the next time we calculate our bounds. This ensures that our bounds reflect the entire dataset and not just our already-filtered dataset.

Finally, our team of finance professionals adjusts a small percentage of these bounds, for example, for industries with small sample sizes, to reflect fundamental financial logic. Statistically determined bounds do not always make sense. This is because the standard statistical method +/- two standard deviations creates symmetrical bounds, e.g., a ratio cannot exceed 100 and cannot be less than -100. This does not make business sense for certain ratios, e.g., inventory days. A negative bound for inventory days is not a product of financial logic but rather the product of statistical symmetry—in order to have a negative inventory turnover you must have either negative COGS or negative inventory, both of which do not make business sense.

Thus, we adjust some of the bounds to fit the business scope of our metrics, not just the statistical scope of our metrics. Moreover, if for any reason the two pass standard deviations method produces an unreasonable bound, for example, a gross profit margin upper bound of more than 100%, then we cap those nonsensical bounds at the next most logical result, which for this gross profit margin example would be 100%.

If you have additional questions on Sageworks's industry data, please contact us at: 866.603.7029 option 1.

Appendix A: Filter 5.

$$For\ all\ financial\ datasets\ \{x_1, x_2 \dots x_i\}$$

$$Define\ Y\ s.t.\ Y = abs[\min(\{x_1, x_2 \dots x_i\})]$$

$$Define\ new\ dataset\ \{x_1', x_2' \dots x_i'\}\ s.t.\ \forall\ x_i : x' = x + Y$$

$$Define\ \mu_1 = \frac{\sum_{i=1}^{n} x_i'}{n}\ and\ define\ \sigma_1 = \sqrt{\frac{\sum_{i=1}^{n} (x_i' - u_1)^2}{n-1}}$$

$$Define\ UBound_1 = \mu_1 + 2\delta_1$$

$$Define\ LBound_1 = \mu_1 - 2\delta_1$$

$$Define\ new\ dataset\ \{x_1'', x_2'' \dots x_i''\}\ s.t.\ \forall x_i'' : LBound_1 \leq x'' \leq UBound_1$$

$$Define\ Z\ s.t.\ Z = abs[\min(\{x_1'', x_2'' \dots x_i''\})]$$

$$Define\ new\ dataset\ \{x_1''', x_2'''' \dots x_i''''\}\ s.t.\ \forall x_i'''' : x''' = x'' + Z$$

$$Define\ \mu_2 = \frac{\sum_{i=1}^{n} x_i'''}{n}\ and\ define\ \sigma_2 = \sqrt{\frac{\sum_{i=1}^{n}(x_i''' - u_2)^2}{n-1}}$$

$$Define\ UBound_2 = \mu_2 + 2\delta_2$$

$$Define\ LBound_2 = \mu_2 - 2\delta_2$$

$$Define\ new\ dataset\ \{x_1'''', x_2'''' \dots x_i''''\}\ s.t.\ \forall x_i''' : LBound_2 \leq x''' \leq UBound_2$$

$$Finally, we\ calculate\ the\ average: \frac{\sum_{i=1}^{n} x_i''''}{n}$$

Appendix B: All Exception Rules

CurrentRatio
QuickRatio
NetProfitMargin
AdjNPBTPctChange
SalesPctChange
InventoryDays
AccountsReceivableDays
AccountsPayableDays
InterestCoverage
DebtToEquity
ReturnOnEquity
ReturnOnAssets
FixedAssetTurnover
GrossProfitMargin
AdvertisingToSales
PayrollToSales
TotalPayrollToSales
RentToSales
DebtServiceRatio
EBITDAMargin
DebtToEBITDA
ProfitPerEmployee
AccountsPayable
AccountsReceivable
AccumulatedAmortization
AccumulatedDepreciation
AdditionalPaidInCapital
AdjNPBT
Advertising
Amortization
Cash
COGSDepreciationToSales
CommonStock
CostOfSales
CurrentLongTermDebt
Depreciation
DirectLaborToSales
DirectMaterialsToSales
EBITDAMargin
GrossFixedAssets
GrossIntangibleAssets
GrossProfit
Interest
Inventory
LongTermLiabilities
NetFixedAssets
NetIncome
NetIntangibleAssets
NPBT
OperatingProfit
OtherAssets
OtherCurrentAssets
OtherCurrentLiabilities
OtherExpenses
OtherIncome
OtherLongTermLiabilities
OtherOperatingExpenses
OtherOperatingIncome
OtherStock
Overhead
OwnComp
Payroll
PreferredStock
Rent
RetainedEarnings
SeniorDebt
ShortTermDebt
SubordinatedDebt
Taxes
TotalCurrentAssets
TotalCurrentLiabilities
TotalEquity
TotalLiabilities
WorkingCapital
InventoryDaysPctChange
AccountsReceivableDaysPctChange
AccountsPayableDaysPctChange
InterestCoveragePctChange
DebtToEquityPctChange
ReturnOnEquityPctChange
ReturnOnAssetsPctChange
FixedAssetTurnoverPctChange
OverheadToSales
CashFlowMargin
OperatingCycle
OperatingCashFlowMargin
LaborCostRatio
SalesPerEmployee
DebtLeverageRatio
AssetComposition
ReturnOnLabor
ProfitPerEmployee
CashFlowMarginPctChange
Debt to Tangible Net Worth Ratio
Senior Debt to Cash Flow
Debt to Cash Flow
Debt to Capitalization
Operating Profit Margin
Sales Growth Rate
Net Income Growth Rate
Notes Payable / Current Portion of Long Term Debt
Notes Payable / Senior Debt
Notes Payable / Subordinated Debt
Dividends Paid / Withdrawals
WorkingCapitaltoAssets
G&APayrolltoSales