Small Business Equity Tool™ Methodology

Small Business Equity Data Tool Methodology v1

INTRO

The *Small Business Equity Tool* displays, and allows users to compare, the performance of Black-, Asian-, Latino-, and Women-owned employer firms in the country's 100 largest metropolitan areas and cities[1].

Business ownership is a key pathway to build community wealth and close the racial, ethic, and gender wealth gaps that persist in America. The main goal of the *Small Business Equity Tool* is to capture and rank the performance of these businesses owned by female entrepreneurs and entrepreneurs of color to promote a clearer understanding of the small business ecosystem and its gaps, encourage goal setting and provide useful comparison points to economic decision-makers at all levels.

The *Small Business Equity Tool* has two predominant uses for local leaders. First, it incorporates a Business Index to rank the performance of Black-, Asian-, Latino-, and Women-owned employer businesses across metropolitan areas and cities. The Index uses six metrics under three dimensions that reveal the fundamentals of the small businesses economy prior to the coronavirus pandemic: number, size, and sector of businesses. Second, it provides groups of comparison peer cities based on city demographic and economic characteristics. With these peers, local decision makers can easily obtain meaningful comparison points for local performance of Black-, Asian-, Latino-, and Women-owned businesses.

GROUPS

The *Small Business Equity Tool* contains business data for five different population groups. For each group we defined a comparison group to contextualize the statistics and calculate metrics of relative performance. The groups and their comparisons are listed below.

GROUP	DESCRIPTION	COMPARISON GROUP
Minority-owned Business[2]	Businesses owned by population identified with any race and ethnicity combination other than non-Hispanic White (i.e. African American, Asian, American Indian and Alaska Native, Native Hawaiian and other Pacific Islander, other races, Hispanic non-white).	Non-Minority-owned Business
Black-owned Business	Businesses owned by population with race self-identified as Black or African American.	White-owned Business
Asian-owned Business	Businesses owned by population with race self-identified as Asian.	White-owned Business
Latino-owned Business	Businesses owned by population with ethnic self-identified as Latino or Hispanic.	Non-Latino-owned Business
Women-owned Business	Businesses owned by population with sex self-identified as Female.	Men-owned Business

---

1 The *Small Business Equity Tool* at the city level presents data for 105 U.S. cities, including the largest 100 cities and 5 cities selected based on ongoing partnerships. Technically speaking, the tool examines Census designated "Economic Places." Among the most populous economic places, the vast majority are cities or coterminous cities and counties.

2 "Minority-owned business" is a defined U.S. Census Bureau classification.

REFERENCE TABLE

The Table below lists all the variables displayed in the *Small Business Equity Tool*, their descriptions, and data sources. It also states whether or not the variable was considered to calculate the Business Index. The two main data sources used for the *Small Business Equity Tool* are the 2018 Annual Business Survey (2018 ABS) and the 2017 American Community Survey (2017 ACS). Further explanation on the data and data sources is provided in the following sections.

VARIABLE	DESCRIPTION	INDEX	SOURCE
Population	Total population.	-	2017 ACS
Minority Population	Total population identified as Minority.	-	2017 ACS
Black Population	Total population identified as African American or Black alone.	-	2017 ACS
Asian Population	Total population identified as Asian alone.	-	2017 ACS
Hispanic Population	Total population identified as Latino or Hispanic.	-	2017 ACS
Number of Businesses	Number of employer businesses for the selected group.	-	2018 ABS
Jobs Created	Total paid employees in employer businesses from the selected group.	-	2018 ABS
Average Salary	Average wage by employee in employer businesses from the selected group.	-	2018 ABS
Business Density	Number of employer businesses from the selected group per 1,000 residents of the same group.	Yes	2018 ABS
Business Density – Parity Ratio	Ratio of the business density of the selected group compared to the comparison group.	Yes	2017 ACS
Average Annual Sales	Average annual sales by employer business from the selected group, in $1,000s.	Yes	2018 ABS
Average Annual Sales – Parity Ratio	Ratio of the average annual sales by employer business for the selected group compared to the comparison group.	Yes	2017 ACS
High-Wage (HW) Industry Participation	Percentage of employer firms from the selected group operating in a high-paying industry.	Yes	2018 ABS
High-Wage Industry Participation – Parity Ratio	Ratio of the percentage of employer firms from the selected group operating in a high-paying industry compared to the comparison group.	Yes	2018 ABS
Business Density for Comparison Group	Number of employer businesses from the comparison group per 1,000 residents of the comparison group.	-	2018 ABS
Average Annual Sales for Comparison Group	Average annual sales by employer business from the comparison group, in $1,000s.	-	2018 ABS
HW Industry Participation for Comparison Group	Percentage of employer firms from the comparison group operating in a high-pay industry.	-	2018 ABS

A Note on Parity Ratios

Parity ratios measure the relative performance of businesses owned by the selected population group compared to the noted comparison group. For example, the Parity Ratio for Business Density for Black-owned businesses is the ratio between the Business Density of Black-owned businesses and the Business Density of White-owned businesses. A parity ratio closer to 0 represents large inequalities and high gaps between Black and White owners. A parity ratio of 1 reflects total equality and occurs when Black- and White-owned businesses have the same performance for a particular metric.

INDEX METHODOLOGY

Dimensions & Metrics

The *Small Business Equity Tool* ranks the performance of Minority-, Black-, Asian-, Latino-, and Women-owned employer businesses in addition to the existing gaps in the business landscape. The Index measures six metrics under three dimensions that reveal the fundamentals of the small businesses economy before the coronavirus pandemic hit: number, size, and sector of businesses.

Each of the three dimensions includes absolute and a relative metrics. **Absolute metrics** are the numerical values of the different variables. They allow us to assess businesses' performance in the selected group, regardless of the performance of businesses owned by other population groups. These absolute metrics partly capture the area's dynamism and competitiveness[3]. **Relative metrics** assess the performance of businesses owned by the selected population group compared to the performance of their comparison groups[4]. These relative metrics account for local conditions that could have similar effects in both groups and consider the size of the gap between businesses owned by Minority and Non-Minority populations. Both metrics are composed of the following three dimensions:

Number. Number of employer businesses per 1,000 residents of the selected group (we call this metric Business Density). A higher ratio of firms to residents relates to more employment and wealth-building opportunities for workers from the same group.

Size. Average annual sales from employer businesses. The larger the annual average sales of these businesses, the greater their economic relevance and impact.

Sector. Percentage of employer businesses operating in the high-wage industries. Healthier entrepreneurial ecosystems tend to have a higher percentage of businesses operating in markets that yield higher sales and offer better wages.

DIMENSION	METRIC	DESCRIPTION
Number	Business Density	Number of employer businesses from the selected group per 1,000 residents of the same group.
	Business Density – Parity Ratio	Ratio of the business density of the selected group compared to the comparison group.
Size	Average Annual Sales	Average annual sales for employer businesses in the selected group, in $1,000s.
	Average Annual Sales – Parity Ratio	Ratio of the average annual for employer businesses in the selected group compared to the comparison group.

---

3   For example, regional economic activity can boost or constrain the economic environment for forming and maintaining a successful business, regardless of the demographics of business owners, which would negatively affect the performance of both minority- and non-minority-owned businesses.

4   See groups and their comparison groups in the Table above.

Sector	High-Wage Industry Participation	Percentage of employer firms from the selected group operating in a high-paying industry. As a proxy for high-paying industry we consider the percentage[5] of employer businesses in the Professional and Technical Services sector[6].
	High-Wage Industry Participation – Parity Ratio	Ratio of the percentage of employer firms from the selected group operating in a high-paying industry compared to the comparison group.

Scoring & Weighting

Once each of the six metrics was created for each category, the Business Index was built following two key steps:

Metric Normalization and Ranking. To normalize the metrics, a ranking technique was used for simplicity's sake. Each of the six metrics are ranked according to where each metropolitan area or city fell in relation to the top performer (the city or metro with the highest value, having a score of 1). This method is not affected by outliers and allows the performance of cities and metropolitan areas to be tracked over time in terms of relative positions (rankings)[7].

Weighting. Once each metric has its ranking, the rankings of the absolute and relative metrics within the same dimension are averaged, and again a ranking is generated for each dimension. Finally, the rankings of each of the three dimensions are averaged and re-ranked to make up the final Index. Thus, each dimension receives a final weighting of 1/3. In cases where information is not disaggregated by sector (and we therefore cannot calculate any of the metrics for sectors), the scoring only considers the rankings in the other two dimensions. We weigh all metrics and dimensions equally (simple averages). As a result, all components have the same impact on the Index performance.

The Business Index is an ordinal ranking that takes values from 1 (top) to 100 (bottom).

DATA SOURCES

The two main sources of data for the *Small Business Equity Tool* are the 2018 Annual Business Survey (2018 ABS)[8] and the 1-year estimates of the 2017 American Community Survey (2017 ACS).

2018 Annual Business Survey (ABS)

The 2018 ABS data was collected in 2017. The data covers all nonfarm employer businesses filing the 941, 944, or 1120 tax forms with receipts of $1,000 or more in 2017. There are two key elements of this data that are important for interpreting our results. First, business ownership is defined as having more than 50% of the stock or equity in the business. Firms unclassifiable by sex, ethnicity, and race include firms that have no owners with a 10% or greater ownership of the stock or equity in the business. These are likely larger companies. In addition, it's important to note that in our classification, businesses can be tabulated in more than one racial group because the sole owner was reported to be of more than one race, the majority owner was reported to be of more than one race, or a majority combination of owners was reported to be of more than one race.

Second, the figures from the ABS are estimated from a sample. This means they will differ from the figures that would have been obtained from a complete census. The data can become less accurate on more granular levels. This is especially the case for sectoral data. The increased errors at more granular levels emerge in large part because the Census Bureau has radically slashed funding for the Annual Business Survey in recent years, causing a sharp decline in the number of businesses sampled in the survey

---

5 The denominator includes the total number of firms in the selected category. Given that not all firms are classifiable, this number may be underestimating the real participation of the selected sectors.

6 Our decision is explained by two factors. First, this industry ranks 3rd in terms of average wages in the U.S. Second, among the top high-pay industries, data availability by sector and race/ethnic of the owner is significantly higher for the Professional and Technical Services sector than for the Information sector and the Finance sector –the other 2 top high-paying industries.

7 Handbook on Constructing Composite Indicators, OECD 2008.

8 2018 Annual Business Survey Methodology, U.S. Census Bureau.

and thus decreasing its local accuracy. All survey and census results contain sampling errors and may contain non-sampling errors. Sampling errors of all metrics included in the *Small Business Equity Tool* are available for download on the Small Business Equity Toolkit website.

A Note on the Data

Because of the significant decrease in sample size, the Census did not release information on the number of minority- and women-owned businesses in a number of metropolitan areas and cities for which data was previously available. As a result, not all metropolitan areas and cities included in the SBET have usable data for all population groups. For example, only 88 of the largest 100 metropolitan areas and 90 of the largest 105 cities have usable data on Black-owned businesses from the 2018 ABS. In addition to this lack of timely data from the Census, St. Louis and Kansas City, MO, both exhibited large but difficult-to-explain changes to the data, which led us to remove them from the SBET.

PEER CITY COMPARISON

For comparison purposes, the Small Business Equity Data Tool identifies groups of cities with similar demographic and economic characteristics. This peer-selection aims to provide local policymakers with useful comparisons points.

We used a k-means clustering algorithm to generate seven groups with fifteen cities each. The clustering algorithm groups cities in such a way that cities in the same cluster are more similar to each other than to those in other clusters. The clustering algorithm considers ten variables that capture key socio-economic features of places. All variables included come from the 2017 American Community Survey (2017 ACS) and are available for download on the SBET website.

VARIABLES	DESCRIPTION
Total Population	Total population in each city.
Black Population (%)	Percentage of population self-identified as Black or African American alone in each city.
Asian Population (%)	Percentage of population self-identified as Asian alone in each city.
Latino or Hispanic Population (%)	Percentage of population self-identified as Latino or Hispanic in each city.
Poverty Rate	Percentage of total population below the poverty level.
Unemployment Rate	Percentage of workers in the labor force who do not currently have a job but are actively looking for work.
Median Household Income	Median household income.
Population with College Degree (%)	Percentage of population that holds a bachelor's degree or a graduate's degree.
Jobs in Manufacturing (%)	Percentage of total jobs in the Manufacturing sector.
Jobs in Non-tradable sectors (%)	Percentage of total jobs in the Education, Health Care and Social Assistance, Retail Trade, and Construction sector.

Before running the algorithm, we performed an exploratory data analysis on these 10 variables to understand the relation between them. The correlation plot shows how a few variables are highly correlated (e.g., the percentage of black population presents a high linear positive correlation with the poverty rate and the unemployment rate). Variables that are correlated have a stronger weight in defining the grouping in the clustering activity[9].

---

9   We attempted to use a Principal Component Analysis (PCA) to remove multicollinearity among the features, but the clustering was not optimal.

Before running the clustering activity, we used a TSNE technique to visualize the 10 variables into 2 dimensions of the feature space. We also scaled all the features to make our database suitable for the clustering method.

We used a k-means++ clustering technique, which generates different cities count in each cluster. To fulfill the special requirement of having 7 equal-size groups (15 cities each), we used the same size constrained k-means technique, a k-means clustering implementation where a minimum and a maximum size for each cluster can be specified. This k-means technique modifies the cluster assignment step by formulating it as a Minimum Cost Flow (MCF) linear network optimization problem. This is then solved using a cost-scaling push-relabel algorithm.
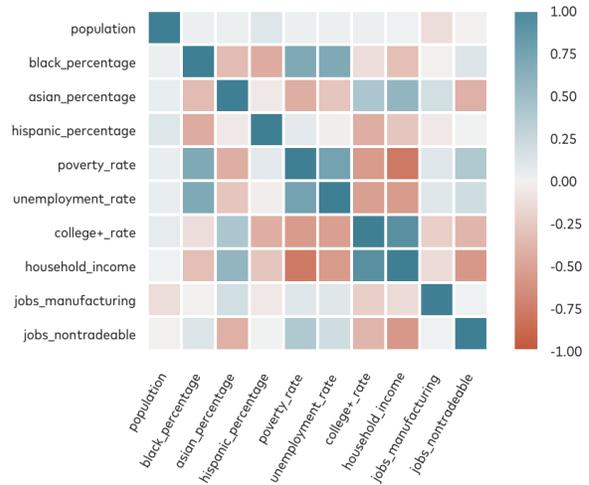
Once we obtained 7 clusters of 15 cities each, we performed a cluster profiling analysis to describe the groups of cities based on the variables used for cluster analysis. Thus, groups can be described from multiple angles. Using this profiling analysis and after carefully reviewing the final groups of cities, we decided to implement a few changes in the initial grouping to ensure better comparability between cities in the same group.

STANDARD ERRORS

Given that the figures from the ABS are estimated from a sample and will differ from the figures that would have been obtained from a complete census, the relative standard errors of all the variables in the *Small Business Equity Tool* are provided. These standard errors are relative as they are an expression of the standard error as a percent of the quantity being estimated.

The example below is illustrative for interpreting the standard errors. Suppose that the average annual sales of black-owned businesses in a metropolitan area are $900,000 and the relative standard error provided is 6 (6%). The standard error is then $54,000 (6% of $900,000). An approximate 90% confidence interval is found by first multiplying the standard error by 1.6 and then adding and subtracting that result from the estimate to obtain the upper and lower bounds. Since 1.6 x $54,000 = $86,400, the confidence interval in this example is $900,000 +/- $86,400 or the range $813,600 to $986,400.

The relative standard errors of each variable are provided by the U.S. Census Bureau. However, for the metrics included in the Business Index that are a combination of different estimates, we calculated the relative standard errors using the following equation[10] (below we present an example for the Business Density metric):

$$\frac{\partial \; Business \; Density}{Business \; Density} = \sqrt{\left(\frac{\partial \; Firms}{Firms}\right)^2 + \left(\frac{\partial \; Population}{Population}\right)^2}$$

---

10 We are assuming that the quantities have errors which are uncorrelated and random, which is not necessarily true, especially in those cases where estimates come from the same survey. If this is the case, we should expect the provided relative standard errors to be lower bounds of the relative errors.