Demographical Analysis of U.S. Businesses for 2018

2Js&aK

Jeremy Garrison, Jesse Noss, Kendra Johnson

4/24/2023

**Introduction**

Starting in 2018, the Census Bureau collected data from approximately 850,000 employer businesses pertaining to economic and demographic characteristics for businesses, owners, and their employees. By analyzing the data provided on census.gov, the variations across demographically quickly immerged.

This study began with asking:

* How do ownership demographics change across industry sectors?
* Do the ownership demographics contribute to the size of a company? The size of a firm?
* Is company revenue affected by their ownership demographics?
* Which demographics existed in the data for comparison?

**Data Sources**

Bureau, U. S. C. (2022, October 28). *Annual Business Survey (ABS) apis*. Census.gov. Retrieved April 22, 2023, from https://www.census.gov/data/developers/data-sets/abs.2019.html

*North American Industry Classification System - NAICS*. United States Census Bureau. (n.d.). Retrieved April 22, 2023, from https://www.census.gov/naics/?58967%3Fyearbck=2017

*Visualization with python*. Matplotlib. (n.d.). Retrieved April 22, 2023, from http://www.matplotlib.org/

Lynn, S. (n.d.). *Python Tutorials*. Shane Lynn. Retrieved April 22, 2023, from https://www.shanelynn.ie/category/tutorials/

**ETL**

The US Census data used in this document was imported through the Census API, using a get request code to pull in a text formatted file containing all of the necessary data. Due to its structure, this file was processed as a JSON file, allowing for the importing of its contents into a pandas dataframe within Python. In this section, the focus is on the race and ethnicity demographics of business owners, so the following steps were taken to narrow down the data to focus on that.

As an early effort to append the data in this dataframe, columns which provided less specific information or information that did not relate to the questions addressed in this document were removed. This included columns such as "RACE\_GROUP" (the code representing each "RACE\_GROUP\_LABEL") or "GEO\_ID" (geographical location identifier - for this data, these were all attributed to the United States). Once the preliminary dataframe was built, it was broken down into two focused dataframes - one on race and one on ethnicity.

To filter the data going into these, "total", or larger aggregated, categories were removed. These included "All firms" tags of EMPSZFI\_LABEL (number of employees at a firm), "Total", "Classifiable", "Unclassifiable", "Minority", "Nonminority" and "Equally minority/nonminority" tags of RACE\_GROUP\_LABEL, and "Total", "Classifiable" and "Unclassifiable" tags of ETH\_GROUP\_LABEL (ethnicity group label). These were removed as they were either: a) inclusive of data represented under more specific tags, or b) ambiguous in terminology (something tagged as a "classifiable" race group could very well also be tagged as "Black or African American", "Asian", or another more specific race group classification).

**Conclusion**

The analysis of this survey was completed by examining the responses over the years. The data quickly showed that there was a major increase in the responses to the 2020 survey as compared in the earlier years. As suspected, during the height of the pandemic, the amount of business owners that completed this online survey was almost four times higher than the previous years.

Focusing in on the 2018 data, the questions were answered by breaking the demographics into three major groups:

* Male/Female
* Veteran/Non-Veteran
* Race/Ethnicity

**Gender:**

1. Professional, scientific, and technical services lead the way for both males and females, but after that we see some significant differences. Males more often own businesses in the Manufacturing, and Transportation sectors, while females have a higher percentage of businesses in the Healthcare, and Waste Management sectors.
2. From our violin plot, female owners lean towards smaller firms with less employees, while male owners have a higher percentage of businesses with 100+ employees.
3. The highest revenue companies in this dataset are owned by males with a higher number of employees. Most companies though, male or female led, reside below 2 billion revenue.

**Veteran Status:**

1. Professional, scientific, and technical services led scored much higher for non-Veterans, however, it was the top industry for both groups. The major change is when both groups have equal interest in the company and Health care and social assistance reigns supreme.
2. Using the violin plot, we can conclude that the number of employees at each firm is relatively consistent across all categories.
3. Significant difference was noted between each of these categories; however, it did not seem to affect the breakdown by sector, nor the size of the business. (Number of employees)

**Race/Ethnicity**

1. The following treemap displays the breakdown of number of businesses in each sector, divided by the race group of the business owner. As seen in the chart, White owned businesses make up a majority of the overall businesses, with a relatively even spread across industries. Minority-owned businesses, on the other hand, can be seen to be more strongly prevalent in a few industries - Asian and Black or African American owned businesses are more present in health care, accommodation, and professional, scientific and technical services. Similarly, Native Hawaiian owned businesses see most of their presence within professional, scientific and technological services, as well as in the administrative support industry. American Indian and Alaska Native owned businesses, on the other hand, seem too have a higher prevalence in the construction and manufacturing industries, as well as in professional, scientific and technological services and health care.

In terms of ethnicity groups, businesses with Hispanic and non-Hispanic owners seem to have their presences spread across industries in similar proportions, although there are many more total non-Hispanic owned businesses.

1. The violin charts below depict the distribution of business firms with varying numbers of employees, divided between the owners' race group and ethnicity group. From the first chart, it can be seen that there are generally less firms with more employees across the board, although it is important to note that there is a strong trend of more even distributions as race groups become larger.

A similar trend can be seen when looking at businesses owned by Hispanic vs non-Hispanic ethnicities, where Hispanic and equally Hispanic/non-Hispanic owned businesss firms with less employees have a higher prevalence than those with more, while non-Hispanic owned business firms are a fair bit more evenly distributed.

1. To begin answering these questions, it is important to note the overall average revenue generated by businesses under different ownership race and ethnic groups. In the bar charts below, these comparisons are displayed.

As seen above, the disparity between average revenues of non-Hispanic or White owned businesses and minority groups is immense, with a difference of about 10x increase in average revenue between the highest and second highest average business revenues of different race-group owners, and a 20x increase between non-Hispanic and Hispanic average business revenues.

The following scatterplots are used to show the relationship between number of employees at a business and the revenue generated by that business, grouped by owner race group or ethnicity group. These also serve to highlight the disparity between the size and revenue of businesses with owners of differing groups.

As White owned businesses heavily outweigh others in terms of revenue and number of employees, multiple scatterplots were included to better visualize the trends seen in businesses owned by minority groups.

Across the board, there is a similar positive trend between increasing employees and increasing revenue. For businesses owned by larger demographic groups, this trend is larger, nearing around 200/1 ($ revenue)/(number of employees). This can be seen with the data points for White and Asian owned businesses, as well as non-Hispanic owned businesses. On the other hand, owners of smaller demographic groups see trends closer to 100/1 ($ revenue)/(number of employees). This comes to show that, while these businesses may start at similar points in terms of revenue while their number of employees is smaller, businesses owned by minority groups will only see half the revenue of similarly sized businesses under White or non-Hispanic ownership.