Exploration of Annual Business Survey (ABS) APIs

Authors: Vahram Khachikyan, Tyler Koizumi, Parker Smith

Approach

Our research looked at the United States Census' Annual Business Survey (ABS) APIs to query data relating to the Minneapolis-St. Paul metro area business demographics. We were looking for correlative relationships within the data to represent the worker demographics, economics of business, and overall fiscal health of the metro area. The data, pulled from the 2019 ABS Survey, helped to illuminate trends within the data that we were able to convert into representative graphs. The ABS Survey data is deconstructed into 4 aggregated APIs, two of which (Company Summary API & Characteristics of Business API) we called upon to access different slices of survey responses when needed.

To better access and understand the schema of each API, we spent the first portion of our time constructing different API queries, considering the aggregation, and looking for statistically-significant data points. At first we had difficulty with comprehending the construction of the API, and the organizational style. The data was not only aggregated, but also was grouped by different business characteristics (BUSCHAR) and question/responses (QDESC). This made separating the data difficult. We decided not to utilize two of the APIs, the Characteristics of Business Owners and the Technology Characteristics of Businesses, as they had limited data regarding the Minneapolis-Saint Paul metro area.

Although we struggled to sort through and properly clean some of the API responses, we were able to find important data and construct representative graphs. These graphs helped us to better understand the prior years' of economic development,

the population of business ownership in the Twin Cities, and other various points of interest.

Getting data

Before the team was able to obtain, each member had to request an API key from the United States Bureau site by registering the company and individual email information.

Figure 1: API Key URL Request Format

When using the API key to access the information, it had to be attached to the end of the API request endpoint after the '`&key=`' portion of the snippet shown in the figure above. In order to access the data through python coding, the group used the following method to get the raw api data:

Figure 2: How To Request API Data in Python

The 'url1' represents the variable to which the user is setting the api data requested to. When requesting the data, the user can input what columns of data they want by inputting the column names after the '`get=`' followed by commas if more than one is requested. The census api will also automatically aggregate the data based on the query itself so precautions by the group when doing so. The api data returns a list so the .json() function was used to convert the raw data into json. By doing so, the group members were able to convert the json data into a pandas dataframe.

```
import pandas

business_summary = pandas.DataFrame(test1, columns = ['NAME','NAICS2017_LABEL', 'SEX','EMP', 'PAYANN'])
```

Figure 3: Converting Json data into Pandas Dataframe

In order to convert the json data into pandas, the pandas module must be imported via the command ‘import pandas’. The ‘business_summary’ represents the variable taking in the pandas dataframe data and the column names are set within the dataframe in the brackets of the code shown in Figure 3.

```
business_summary_minny = business_summary[business_summary.NAME == 'Minneapolis-St. Paul-Bloomington, MN-WI Metro Area']
```

Figure 4: Finding Minneapolis Specific Data

Since the group wanted to find data pertaining specifically to the Minneapolis-St. Paul Bloomington Metro Area, the dataset was cleaned by creating a new dataframe that only contained rows that had the name value ‘`Minneapolis-St. Paul-Bloomington, MN-WI Metro Area`’. The method used to get these specific values is shown in Figure 4.

```
business_summary_minny_ethnicity =  business_summary_minny_ethnicity[business_summary_minny_ethnicity.RACE_GROUP != '00']
```

Figure 5: Dropping Rows With Certain Column Values

Team members then dropped rows that were not needed for the visualizations in the next section with the method shown in Figure 5. In Figure 5, rows that contained the value within the single quotes at the end of the code snipped were excluded from the dataset. The ‘!=’ signifies does not equal but this can be changed to ‘==’ so that only those values would be included in the new version of the dataset.

Visualizations

Characteristics of Business API:

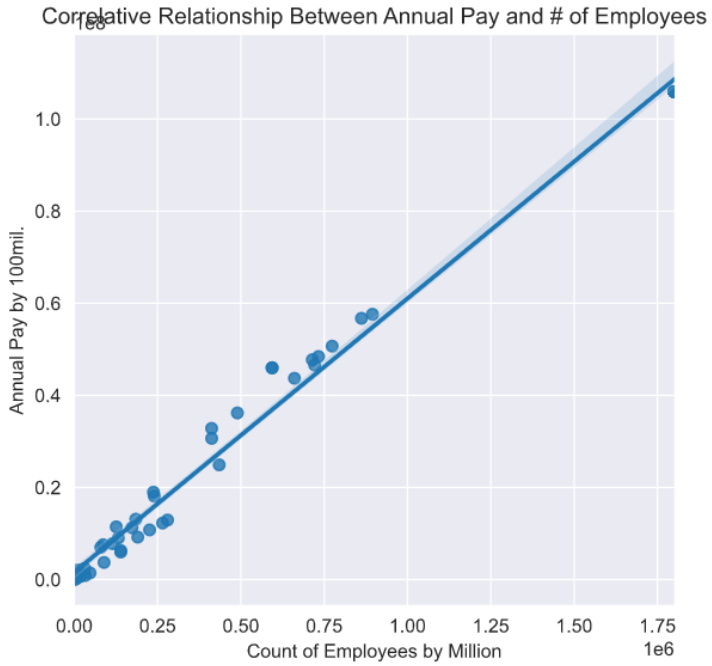

Figure 6: Correlative Pay and Employee Relationship

For the Characteristics of Business API, we were curious to see the correlation between employee size, and annual pay. We figured that there would be a strong correlative relationship, as the higher count of employees would usually factor a larger payroll. As observed above, the relationship almost entirely follows the trend line, with an extremely large outlier at the tip of the trendline. Even the outlier exists within the trend line space, suggesting that the correlative relationship is strong.

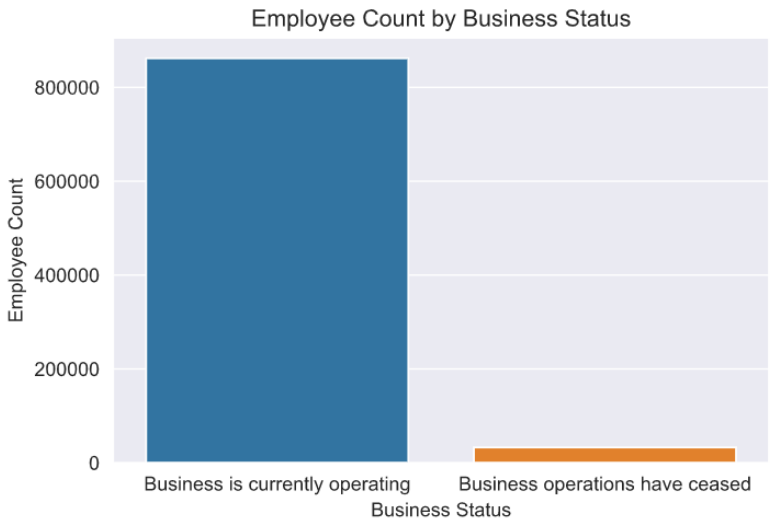

Figure 7: Business status versus employees currently working

In Fig. 2 above, we represented the proportion of employees who remained employed through the year, and compared that count to the proportion of employees whose' workplaces had ceased operations into 2019. This drastic difference is a positive indicator that a majority of businesses remained open, as well as a significantly smaller number of employees were displaced by their employer shutting down.

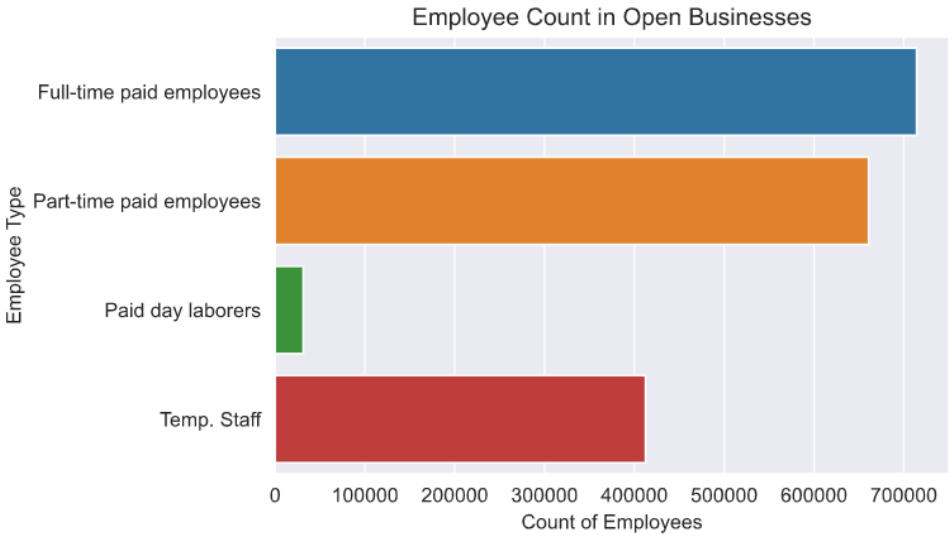

Figure 8: Types of employees at currently operating businesses

Figure 3 represents the discrepancies between types of employees working for currently active (c.a 2019) employers. This graph illustrates the different types of positions currently being filled, and is indicative of a balance between full-time and part-time employment. One interesting point of data is the high count of temporary staff, which is an aggregated value of contracted workers, temporary hires, and seasonal workers.

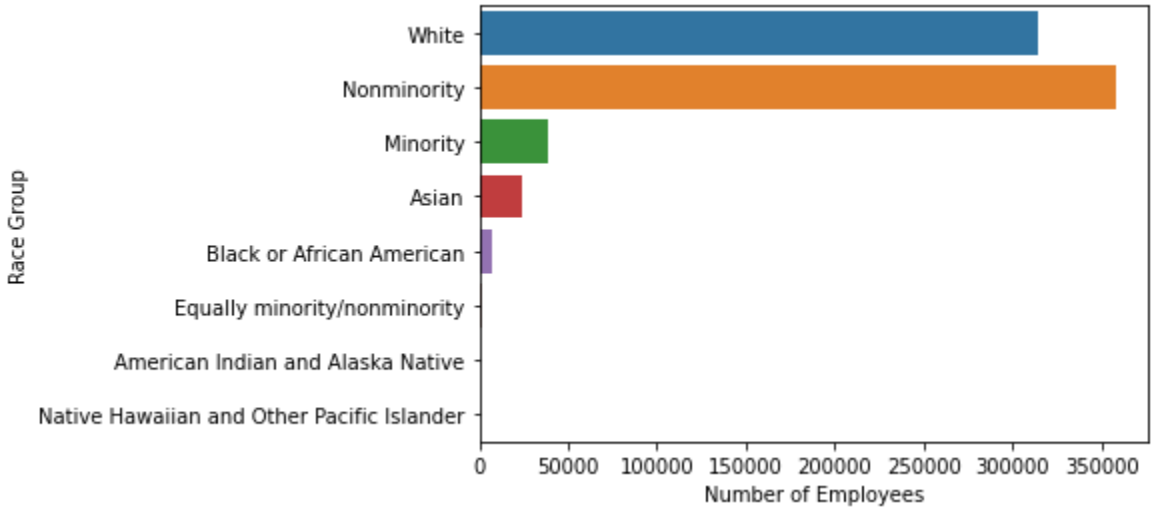

Figure 9: Number of employees by race group

In the figure above, we can see the number of employees based on race group. This barplot was done using seaborn, and the ‘Total’ value of the Race Group was filtered out, because it would not add a significant meaning to this graph. It also includes the nonminority and minority labels represented in the graph, and so some of these labels are inside others (for example, Black or African American Group is inside the Minority Group).

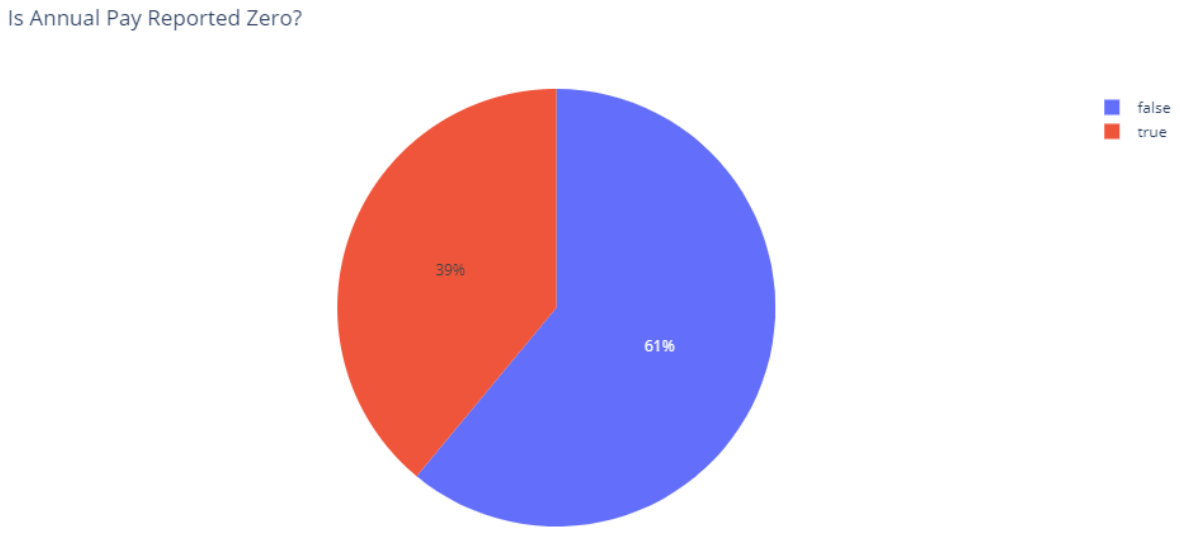

Figure 10: Percentage of businesses that reported their annual pay as zero

This simple pie chart represents the percentage of data that reported their annual pay as zero. This pie chart is made using Pyplot, and a separate column needed to be created to represent if the Annual Pay was zero or not. I created this pie chart, because it was suspicious to me to see so many zeros in the Annual Pay column.

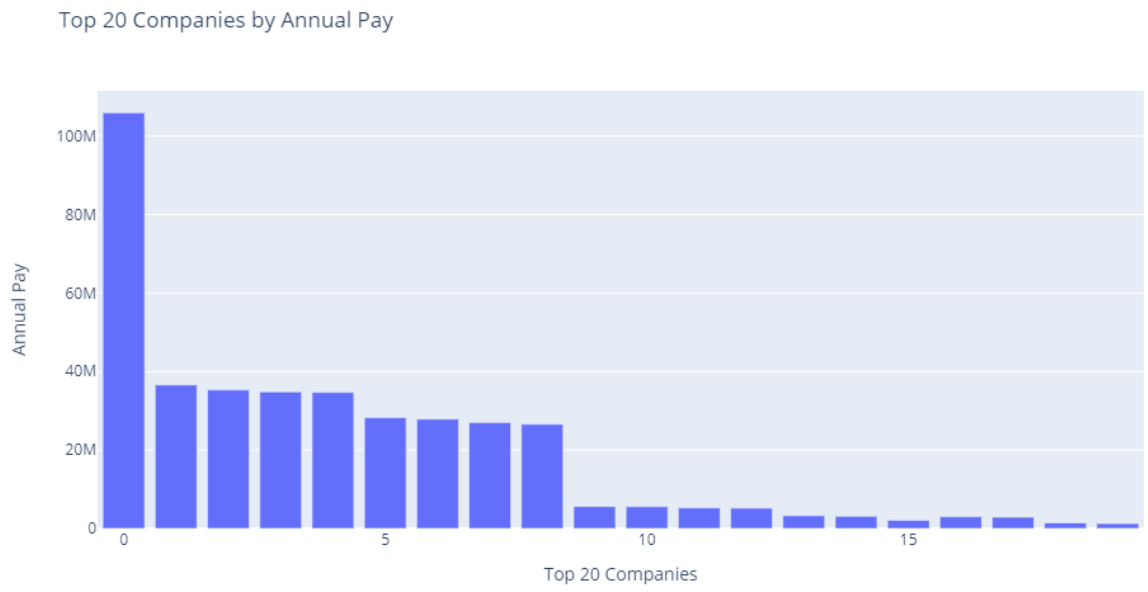

Figure 11: Top 20 companies by Annual Pay

The figure above shows the top 20 companies based on Annual Pay. This figure is made using Pyplot. I had to sort values based on Annual Pay, and then reset the index. Unfortunately, since the reports do not have a company name, I had to include their indexes as the x axis. This figure might be useful for people who want to enter the market, to see their competitors' annual pay and check to see if the market is competitive, or if there is just one or two businesses controlling the market.

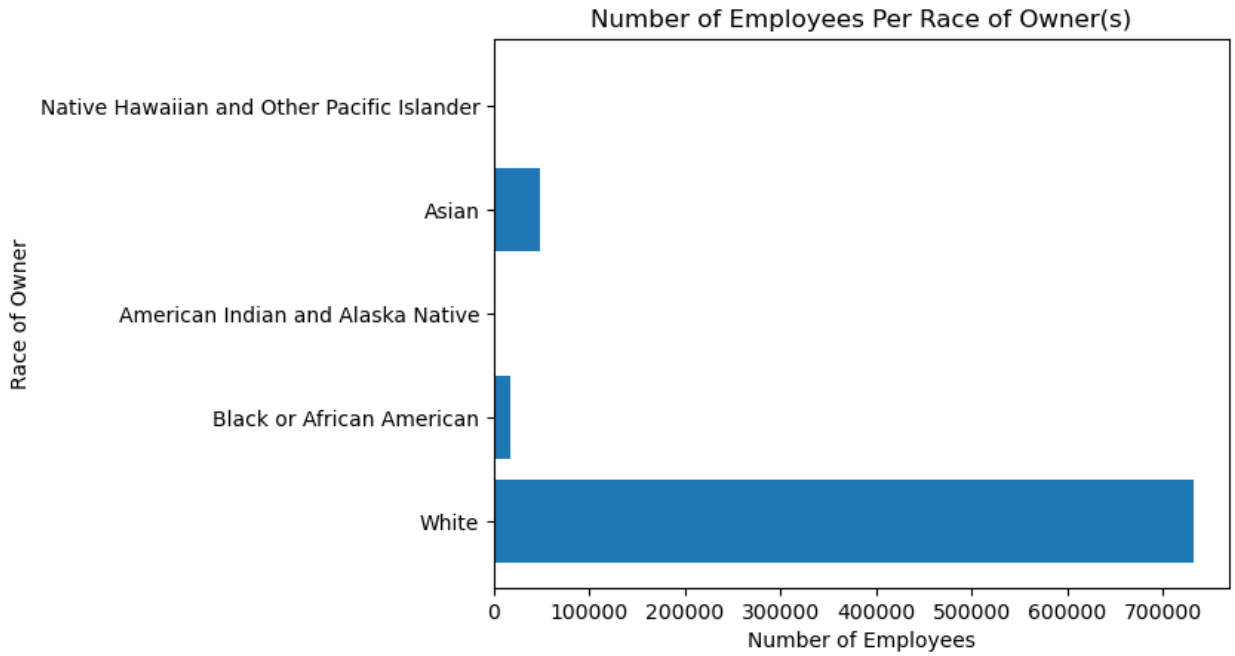

Figure 12: Company Employee Size Per Owner Race

Figure 12 aimed to look at the race demographics of the owners of companies and see how many people that each demographic employed. In the Minneapolis metro area, a large majority of the people employed were working under White business owners followed by Asian and then black or African American business owners. The visualization also seems to indicate that there are no Native Hawaiian and Other Pacific Islander or American Indian business owner or that those demographics did not participate in the census survey.This visualization seems to indicate that most business owners in the area are white as it does not show the average amount of employees employed per given demographic. More research would be required to see if there are any variations in company size on average per given demographic.

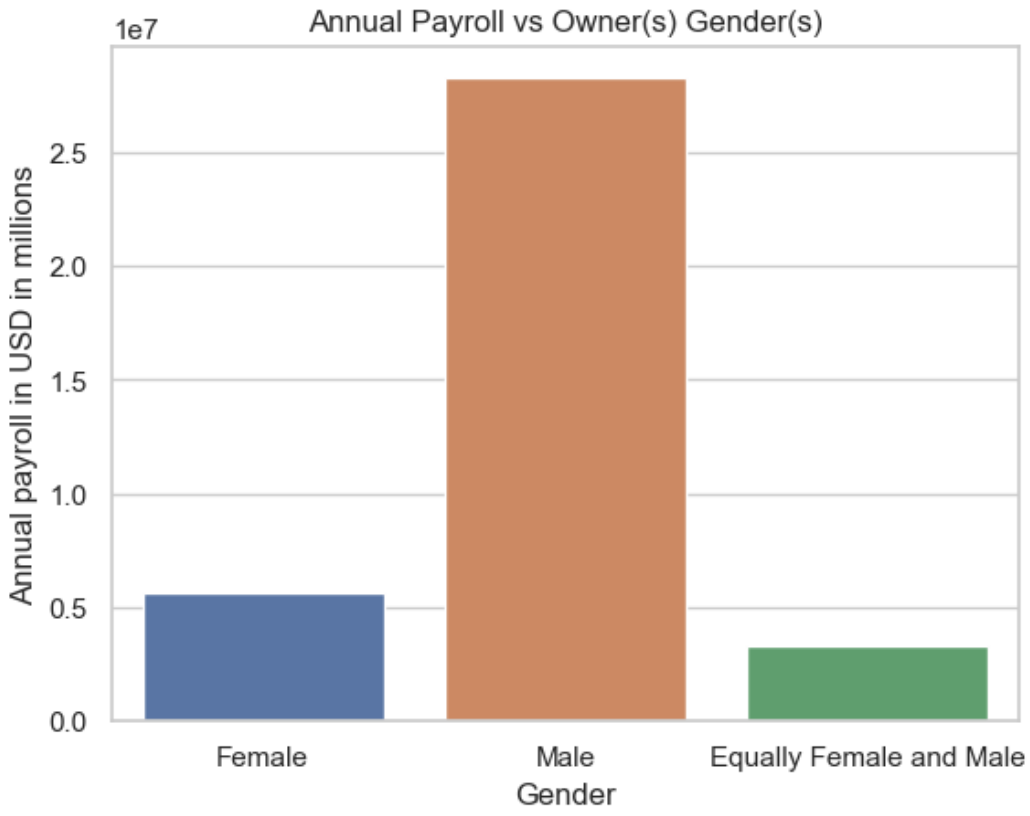

Figure 13:Annual Payroll Per Owner Gender

Figure 13 aimed to look at the gender demographics of business owners in the Minneapolis metro area and see how the total annual payroll for those given demographics. The male business owners seemed to have the largest total payroll out of the given gender demographics followed by female and then equally female and male. However it is worth noting that this does not reflect the average payroll per given demographic as this only shows total payroll. This however seems to indicate that most business owners in the metro area are comprised of males.

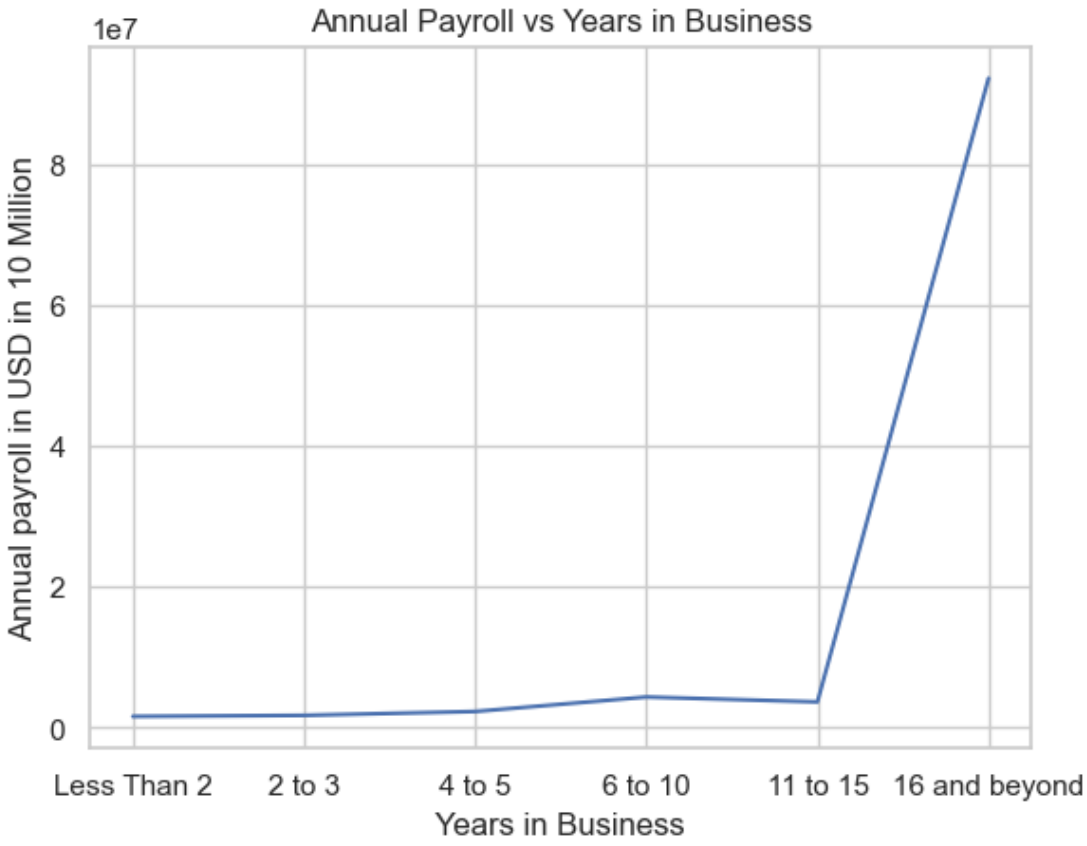

Figure 14: Annual Payroll vs Age of Company

Figure 14 aimed to find how the total annual payroll correlates with how long the company has been in business. The visualization seems to indicate that the average payroll is largely flat for companies in the first 15 years of existence with a small increase between 4 to 5 years and 6 to 10 years. There also seems to be a drop in annual payroll between the companies of age 6 to 10 years and 11 to 15 years. However there is a large jump in total annual payroll after 16 years and beyond. A solid conclusion cannot be made from the given data as the dataset does not break down the the years beyond 16 years. The only conclusion that can be drawn is that companies that have been in existence for at least 16 years make up a large amount of the payroll in the Minneapolis metro area.

Conclusions

From our findings exploring the Annual Business Survey APIs, we were able to make surface level inferences about the demographics of employees, owners, and fiscal operations of the Minneapolis-Saint Paul metro area.

One of the most important of such discoveries was the drastic economic inequity between white-male owned businesses, and minority-or women-owned businesses. According to our initial research with this data, the annual payroll of men-owned businesses was more than five times that of women-owned businesses, suggesting a higher level of commercial success. Additionally, white business owners employed over 75% of total employees within the metro area, further suggesting the economic success of said demographic.

Other observations regarding the economic index of success lied in the ratio of open versus closed business, and the count of displaced employees by such circumstance. However, we did also find that a significant number of businesses failed to, or opted not to, report their annual pay, which could lead to a skew in our representative data.

References

**U.S. Census Bureau. (2021, March 11). U.S. Census Bureau Annual Business Survey (ABS) API's (2019). Census.Gov. https://www.census.gov/data/developers/data-sets/abs.2019.html

**This website includes all documentation for querying and accessing the nested API's. It contains details regarding the data of each API, as well as variable and geographic documentation.