Identifying the Questions:

Our group was tasked with analyzing a large dataset obtained from the Census API. We decided that it would be prudent to gather information that we thought related to each other. Given the nature of the information we saw, we decided that looking at a business's financial data through different lenses made the most sense. Three primary questions emerged from this exploration:

1. What is the impact of sales and payroll based on how long a company has been in business?

2. How do the total sales and payroll of companies with different sizes compare, broken down by gender?

3. Did the Industry play a large role in determining the racial group of a company?

Problems Encountered:

In order to answer these questions, we had to clean the data first. We got rid of many columns in the CSV that were not needed for any type of questions any of us would be answering. Then we personally went through the remaining data and decided what would be acceptable to use and what wouldn't be. We ended up cleaning the data in a way that there was only the Payroll, Sales, Years in Business, and the Sex\_label left. We kept the Sex\_label because it was an idea from one of our group members that made the most sense to avoid duplicating data in our CSVs.

We encountered several issues during the analysis process. For the first question, we found that the visuals did not provide much new information beyond what was expected. The data showed that the companies that had been in business for 16 or more years dominated the companies that had been in business for less. Separating by men and women also did not show much of a pattern, perhaps due to the inconsistent number of men and women in the dataset.

For the second question, we found that the trends for each gender were not consistent as one might expect. A reason for this may be that there were different numbers of men and women employed by companies of different sizes. Further research would best entail the monetary amount of sales and payroll not only by gender, but also normalized by the amount of that specific gender working in a specific firm.

For the third question, we found that outside of the three visuals created, the data could not tell us more in regards to the question focused upon. Further research into why certain industries were dominated by certain races as opposed to others would be an area of interest.

End Result of the Data:

Our analysis yielded some interesting insights, but also raised further questions for future exploration. For the first question, we found that it makes sense that if a company has been in business a long time, it is successful. However, we wanted to see if we could find anything interesting beyond that. Maybe there is a threshold and after a certain amount of time a company is in business, it ceases to increase payroll or sales. It meets a so-called "ceiling".

For the second question, we found that the data did not show clear patterns, but it may be due to the inconsistent number of men and women in the dataset. Future research could look at the monetary amount of sales and payroll by gender normalized by the amount of that specific gender working in a specific firm.

For the third question, it was found that racially white companies dominated a majority of the industries. Research on industries that promoted minority and persons of color ran companies would be intriguing to analyze reasons to promote future companies.

Overall, our analysis highlighted the importance of thorough data cleaning and the need for continued exploration of the data to uncover new insights.