Effect of Societal and Technological Changes on Fortune 500 Companies

Franklin Glance, Matthew Heeter, Parker Schell, Esther Yi

SYS 2202 – University of Virginia

Table of Contents

[Abstract 4](#_Toc103068556)

[**Introduction** 5](#_Toc103068557)

[How have technological and societal changes affected the success of Fortune 500 businesses since 1985? 5](#_Toc103068558)

[How did the rise in technology affect successful businesses/industries? 5](#_Toc103068559)

[How does inflation of goods and services affect the success of Fortune 500 Companies? 5](#_Toc103068560)

[How does Gender in executive positions affect the success of Fortune 500 Companies? 6](#_Toc103068561)

[**Related Work** 7](#_Toc103068562)

[Technology 7](#_Toc103068563)

[Inflation 7](#_Toc103068564)

[Gender 7](#_Toc103068565)

[Covid-19 8](#_Toc103068566)

[**Analysis Design** 9](#_Toc103068567)

[Technology 9](#_Toc103068568)

[Inflation 9](#_Toc103068569)

[Gender 9](#_Toc103068570)

[Covid-19 10](#_Toc103068571)

[**Analysis** 11](#_Toc103068572)

[Technology Impact Analysis 11](#_Toc103068573)

[Inflation Impact Analysis 11](#_Toc103068574)

[Gender Impact Analysis 18](#_Toc103068575)

[Covid-19 Impact Analysis 32](#_Toc103068576)

[**Discussion of Results and Findings** 40](#_Toc103068577)

[Technology Analysis Results 40](#_Toc103068578)

[Inflation Analysis Results 40](#_Toc103068579)

[Gender Analysis Results 41](#_Toc103068580)

[Covid-19 Analysis Results 41](#_Toc103068581)

[References 43](#_Toc103068582)

Abstract

To be completed at the end. tk

**Introduction**

This report aims to clarify through a detailed analytical process, answers to the following overarching question:

# How have technological and societal changes affected the success of Fortune 500 businesses since 1985?

Due to the broad nature of this question, it was found to make more sense to break it up into four subcategories, each focusing on a change related to businesses on the Fortune 500. The sub-categories are the following:

* *Technology*
* *Inflation of Goods and Sevices*
* *Gender*
* *Covid-19*

In general, the primary metric used for analyzing success is annual revenue. However, in each section, slightly different specifics were used as appropriate for the question at hand. Categorical analysis of the effects these changes have had on Fortune 500 companies has provided a deep insight into not only the performance of successful businesses over time, but how each change specifically impacted successful Fortune 500 companies.

# How did the rise in technology affect successful businesses/industries?

The past 40 years have demonstrated a technological revolution of massive proportion. Computation power has scaled exponentially, resulting in more interconnected and globalized society than ever seen in history. This technological revolution has impacted the lives of everyone, in small and large ways. This sub question is focused on analyzing the effect that the rise in these technologies has had on successful businesses and industries, specifically Fortune 500 companies.

Many analyses currently exist on the topic of successful businesses.

# **How does inflation of goods and services affect the success of Fortune 500 Companies?**

Over the course of time, the definition of a successful company has changed significantly. This is demonstrated by the change in what allows a company to be on Forbes’s Fortune 500 company list, which displays the 500 most successful companies every fiscal year based on revenue. Before 1995, the companies on the Fortune 500 list were solely of manufacturing and industrial companies, but this changed with the addition of service-oriented companies to the list in 1995 (Catalyst 33). Also, over time, the value of products and services drastically change based on economic conditions and the overall state of the economy in addition to other external factors. Things that were deemed less valuable at the time, i.e., silicon, could massively inflate in price and vice versa. While the Fortune 500 list is a strong indication of how well a company is performing, comparing the revenues of companies on the Fortune 500 list to the inflation rates of the goods and services these companies provide could give better insight into what makes a company “successful”. In addition, it could provide a useful insight into how economic factors can affect the success of a company. For example, we can look to see how the crashes of certain companies like Blockbuster or Sears were affected by the inflation or deflation in price of what they were selling. The opposite could be observed as well. This led us to ask the following question: How does inflation of goods and services affect the success of companies between 1997 and 2017?

# **How does Gender in executive positions affect the success of Fortune 500 Companies?**

There are many signs of inequality in the workforce between men and women. Starting in the early 1900s, women formally began entering the workforce. Although laws were enacted in protection for female workers to ensure equality, it is still questionable as to whether the practice of these laws has led to actual equal opportunities. Studies have shown on average, women make significantly less than their male counterparts, and these numbers drop even more when further categorized by race. Although the workforce is approximately equal in the number of women and men working, it has been observed that companies are more reluctant to provide opportunities for professional growth. One way to observe these discrepancies is by looking at leadership in the Fortune 500, more specifically, the CEOs. Within these successful businesses of the Fortune 500, we can hope that the qualities of being a leader are not dictated by gender but rather by skill set. By looking at the raw number of women on the list, what sectors are more likely to have female CEOS, and the revenue and ranks of the companies they work for, we can test how these different characteristics affect a company’s likelihood to adopt a female CEO. After the analysis, it was observed that although the number of female CEOs has been growing linearly since 1995, the projected year for reaching an equal number is not for many years from now with different sectors having more or less females. Also, when looking at the revenue distribution, female run companies and male run companies perform differently. Consequently, company rank, and sector also determines the likelihood of adopting a female leader.

**How has Covid-19 impacted Fortune 500 Companies?**

Over the past two years, the world has been greatly affected by the Covid-19 pandemic. Every aspect of our lives has been impacted, from family, to business, to religion. At first, the economy plummeted due to the government mandating businesses to shut down. However, as more research and vaccines become available, the economy has been able to restabilize. Although this pandemic is detrimental to many businesses, it has also been beneficial to others. To study how different businesses performed through Covid-19, we will be examining the Fortune 1000 from years 2017 to 2021. The Fortune 1000 provides ample data of successful businesses and through careful analysis, trends and predictions can be made. The Fortune 1000 dataset includes the sector of each business. By comparing each sector's revenues, profits and number of employees, we can draw conclusions about how covid affected industries differently. We will also be looking at the changes in the Fortune 1000 by state. This will show us if states with harsher Covid regulations were affected differently than states with less restrictions. This can help us draw conclusions on whether the trends that we see are because of the pandemic or other factors. Throughout this analysis, we will analyze the effect of Covid-19 on successful businesses and determining how sectors were affected differently than others.

**Related Work**

As with any popular research topic, there does exists previous analyses into similar topics discussed in this report. To effectively research the existing analyses and findings for this topic, a topic-by-topic analysis was performed. Our research into related works has concluded that while many similar individual studies do exist, one combining research towards the effects of these four major changes (Technology, Inflation, Gender, Covid-19), has not been seen before.

# Technology

Insert Franklin related work here. A discussion of existing work that tries to analyze this problem and their findings

# Inflation

While there have been many analyses on inflation, especially since the beginning of the COVID pandemic, most of them focus on why inflation shot up during COVID and how companies have responded. During the COVID pandemic, inflation significantly increased, shown through increases in the price of consumer goods; in April, inflation of prices shot up 0.8%, which was the largest single-month jump in prices since 2009 (Winck, Kaplan 1). In Winck and Kaplan’s article, they addressed that this overall inflation is due to a large increase in the prices of goods; his example is that in May of 2021, the 0.6% increase in inflation was attributed to a 7.3% spike in car and truck costs, a 0.4% jump in food costs, and a 0.7% decrease in gasoline prices. While this comparison is somewhat like our hypothesis, our analysis differs by comparing the long-term effects of inflation on the revenues of companies instead of short-term inflationary effect on the economy. The other types of comparisons that we have come across are those talking about how companies have adjusted to inflation due to COVID. Will Daniel’s article lists 10 companies that have decided to increase their prices to combat a potential decrease in profit. While this article shows what companies due to combat inflation, our analysis does not address how companies combat inflation but rather how inflation affects their revenues. We see this indicator as what makes a successful company.

# Gender

Despite the fairly equal distribution of men and women in the workforce, there are still many differences in how they are treated. According to the US Bureau of Labor Statistics, as of 2012, women make up slightly more than half of the white-collar workers in America. White collar refers to work in offices or other professional environments. However, they only make up less than 5% percent of the executive titles.  The percent of women employees tends to be based on the type of occupation. For example, only 18.7 % of software developers were women in 2020 compared to the 88.9% in nursing. Also, only 27.6% of employees in chief executive positions in management, profession, and related occupations which has a 51.8% female employee rate. Even though women are already not given the same chances, a study done in 2010 showed that CEOs who were women only earned about 74.5% compared to their male counterparts with this percentage getting even smaller for people of color. Though the argument can be made that they are getting paid less and are not being promoted to these higher positions because of their lack of competency, a study looking at firms and CEO gender showed that gender did not affect the extent of real earnings in management.  Another study on female leaders on firm performance showed that there was a net increase in overall sales per person when a female CEO was introduced. When a firm had a workforce of less than 20%, sales per person went up 3.7% and this number increased to 14% when more than 20% of the staff was female. Looking at this, it can be concluded that the reason that women are not in these higher positions is because they are simply not being given the opportunity and the environment for women leaders is more hostile than that of male leaders. However, when the opportunity is given the company is also likely to benefit in a sales increase. The Fortune 500 are the top 500 revenue companies in the United States; however, they are not an exception to the patterns of other general companies. Given the higher status of these companies, they may or may not follow a similar pattern as the general population. The previous studies mentioned look at certain sectors, or just overall businesses, however this study looks at how the top performing companies in America are hiring CEOs given the fact that they are the most successful businesses. Because these companies are the ones that receive the most media and general public’s attention, ideally, they would hold a more progressive standpoint and be more open to hiring females as their CEO.

# Covid-19

There is currently extensive research about the Covid-19 pandemic and the economic impact. From employment to the stock market, to the supply chain, researchers and economists have been analyzing how these business factors have changed due to covid and what they might look like in the future. In Sharon Stang’s research on the impact of Covid on businesses and employees, the industries that faced the most government mandated closures were arts, entertainment, and recreation, with 48% of businesses experiencing a closure, and accommodation and food services, with 36% of businesses experiencing a closure (Stang). The industries that faced the least number of closures were mining, quarrying, and oil and gas extraction, and agriculture, forestry, fishing and hunting, both with only 3% of businesses experiencing a closure (Stang). Although Stang’s work is related to our question, we will be analyzing the trends based on monetary value instead of the number of employees. Stang’s findings were based on the number of employees that each industry lost. Also, the sectors that we will be looking at are different from the industries that she looked at because we will be using the Fortune 1000 data to only look at successful businesses.

According to Cory Breaux, Jessica Fernandez and Ben Griffis, there were clear winners and losers of the Covid -19 Pandemic (Breaux, Fernandez, Griffis). The writers analyzed the revenue of four industries, real estate, truck transportation, transportation and warehousing and couriers in messengers. The last three were studied to show whether people are consuming less good. All four analysts show a sharp decrease in revenue mid 2020 and then a sharp increase in revenue soon after. This shows that although people may have stopped consuming at first, they soon started buying again once they got adjusted with living in a pandemic. Our analyst differs from this because we will be looking at a wider range of data. The researchers only presented these four trends; however, we will be looking at several more. We will also not only focus on revenue but also profit and the state.

**Analysis Design**

Each of our individual analyses was aimed at one of the four changes studied in this analysis. Since the nature of each subsection varies slightly, a uniform analysis technique wouldn’t be suitable to fairly analyze each part. For example, it doesn’t make sense to evaluate the same success metrics when talking about Covid-19 versus the rise in technology since 1985. The following subsections will provide detailed insight into the analysis method used to pragmatically answer each sub question.

# Technology

# Inflation

For this hypothesis, how has the inflation of goods and services affected the success of businesses from 1997 to 2017, we thought that it was relevant because revenue is a key performance indicator for a company’s success, and inflation is strongly correlated with the annual rate of increase or decrease in revenue. While the overall health of the economy effects all companies in general, there are certain goods and services that will change in comparison with others. If the price of certain goods increase, the then revenue of a company in that related market should also increase. If the price of a certain good decreases, then the revenue of the company shouldn’t decline more than the rate of inflation.

To explore our hypothesis, we need to use two data sets. The first data set, called fortune500, contains the name rank and revenue of every company on the US Fortune 500 list from 1955 to 2021. The second data set, called price\_changes, displays the Entity (country), Code (Abbreviation of the country), Year (1997-2017), and the increase/decrease in value in various goods and services (Clothing, New cars, Toys, etc.). It reveals how these consumer goods/services categories have inflated or deflated in value. Using these two data sets, we can further explore our hypothesis.

We first decided to find the average annual revenue throughout the 20 years of Fortune 500 data and then filter out the companies that didn’t hit this benchmark value. We then found the companies who have been on the Fortune 500 list for the complete range of 20 years. After using these filters, we then created a categorical variable that displayed the sector that the companies operated in. Doing this allowed us later on to compare the inflation of goods and services with the companies that operate in that sector. We then joined the two data sets and proceeded to graph the inflation of goods and services per year and compared it to the annual revenues of the companies in each category of goods and services, which are demonstrated in the *Analysis* section.

# Gender

The sub question “How are Fortune 500 companies affected by gender,” was analyzed from a few different perspectives: number of companies on the list with female CEOs, revenue comparison between male and female run companies, how rank is different between male and female ceo companies, and if sector plays a role in the likelihood of whether or not the CEO is a female. The first round of analyzing looked just at the number of females on the list by grouping the data by years and then counting the number of companies that were listed to have a female. Then, using the data about sectors of companies from another dataset, the number of female CEOs in different sectors of the Fortune 500 was found as well for comparison. By looking at these, it can be determined if companies of certain sectors are more open to having a female run the company. For the revenue, a line graph and box plots were used to measure growth between men and women run companies.

# Covid-19

To analyze how the sectors were affected by Covid-19, we need to look at the changes in revenue and profit for each sector from the years 2017 - 2021. Examining data from three years before the pandemic is crucial because it will differentiate whether the factors are changing because of a trend or because of covid. For both revenue and profit, we will create three different graphs, one bar graph showing the revenue/profit of the different sectors for each year individually, another showing all the years in a grouped bar graph and lastly a line graph. Through these three graphs we will be able to see whether the trends during Covid (years 2020 - 2021) are different from the previous years. Following this we will be looking at the states and the revenues and profits that the companies in each produce. This can help solidify our conclusion because we will be able to see if different restrictions affected Fortune 1000 businesses in the long run. For this, we will graph a choropleth map of the United States with the variable (profit or revenue) of each state being different shades.

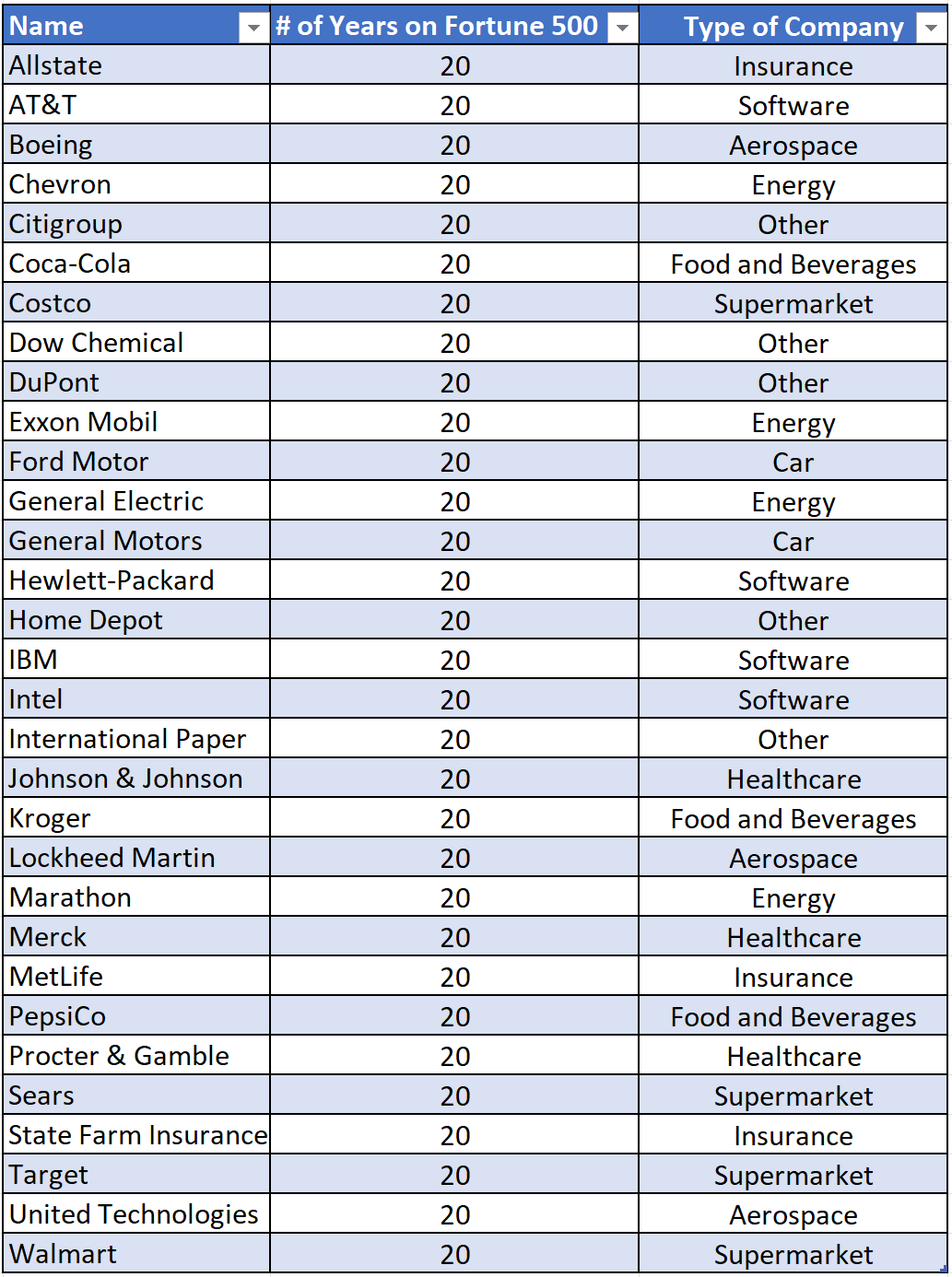
**Analysis**

As described in the previous section, our analysis was broken into four sections, parallel to the following sections.

# Technology Impact Analysis

# Inflation Impact Analysis

After consideration of the Analysis Design mentioned earlier, we decided to look at variables in the two datasets and decide which ones are relevant. The Name variable is categorical, the inflation rate of each good and service is numerical, and Revenue is categorical, however we plan on changing this to a numeric variable later on. You can check this by using the glimpse() function for the two data sets to see the data type of each variable. We think that Name is a relevant variable because we are comparing how a company's revenue changes over time in comparison to the inflation/deflation of goods and services. The Name variable also distguishes companies from one another. This variable allows us to categorize companies into their industry sectors to make a comparison with the goods and services that belong to that respective sector. Without this variable, you could not make a comparison by sector or at all. There is not any missing data for the Name variable, which we think is another reason why this variable is relevant. The Revenue variable is relevant because revenue is the sales rate of a company and is the primary and best key performance indicator of a company’s monthly, quarterly, and annual sales performance. Based on the change in revenue, one can predict the growth or decline of a company’s sales performance and, to some extent, the overall success of a company. The inflation rate, as mentioned earlier, is used to compare with each company’s revenue and is used as a core variable when comparing a sector’s inflation rate with the companies that reside in it and their revenues. There are other variables that can help with answering this hypothesis, but they are mainly used for sorting and merging the two data sets.

After filtering data, creating new variables, and merging the two datasets, we narrowed down our list of companies to 31 companies that have all been on the Fortune 500 list for the entire 20 years and have above average annual revenue. These 31 companies along with their company type are shown in the table on the right.

We then compared the companies of each sector with their respective sector's inflation (some companies fell into multiple categories so will graph them in multiple categories). We did this by filtering the by the variable we created earlier on called Company\_type. We compared by creating line graphs of each company in the specific sector, using the parameters Year and Revenue\_billions, to view the companies' trends and compare them with the inflation of the goods/services in that market, which was done by creating a column graph with the parameters being Year on the x axis and the different goods' inflation rate.

Chart

Description automatically generatedChart, line chart

Description automatically generated

We first compared the inflation percentage for new cars, shown in Figure 4, with the revenue of companies in this category. We included Insurance companies in Figure 5, which measures the revenues of car and insurance companies over time, because car insurance is affected by people buying new cars, the price of new cars being insured. Also, with declining or no new car sales, the total replacement cost of cars being insured lowers each year. If people aren’t buying new cars, then they are not buying car insurance as much, which makes the two markets dependent on each other. This comparison above has shown an interesting correlation. When the inflation in value of new cars is positive, which is after 2010 and represented by the blue line in Figure 5, the revenues of all of the companies increases. However, the opposite does not seem to be true; negative inflation/deflation does not always entail a decrease in revenue in this case. A massive decrease in revenue is shown between 2007 and 2010, a period of deflation for new cars, in all companies except for State Farm Insurance. This can be explained by the Great Recession in 2008. During periods of depression, people are much less likely to purchase cars, which explains why there is a decrease in revenue for these companies. However, it seems that the Insurance companies are not solely dependent on cars as State Farm didn’t decrease during this time and have other sectors that they belong to, such as Healthcare.

Chart

Description automatically generatedChart, line chart

Description automatically generated

We then looked at the inflation of clothing and companies that would be categorized as Supermarkets and Department stores, as these would be companies that sell clothes . Figures 6 and 7 display an interesting but different correlation from the comparison discussed in the previous paragraph . When the inflation in value of clothing is positive, the revenue of all of the companies above except Sears increase; Costco and Target steadily increase while inflation is positive, but Walmart dramatically increases. The same applies when the inflation rate of clothing is negative. Sears slightly decreases when the inflation of clothing is positive, which only happens during 1998. This comparison is different from the previous because for most of the years, the inflation rate is negative, while most of the companies increase in revenue. This leads us to think that the deflation of clothing does not seem to affect Revenue across the board. Sears overall decreases in revenue, Target increase and then starts to plateau, and Costco and Walmart increase steadily. These companies that increase during a period of deflation could indicate success, but all of these companies in this category have multiple markets, so there is more to explore in regards to those companies.

Chart, bar chart

Description automatically generatedChart, line chart

Description automatically generated

Next, we compared the inflation of software, as shown in Figure 8, with companies in that category and their revenues. We included Aerospace companies in Figure 9 because aircraft are tightly knit with software, as it helps with flight controls, etc. The comparison between these two criteria show another completely different correlation. As shown in the first graph, Software as goods drastically deflate over time as more software has become available. While this happens, the revenues of Intel steadily increases, Hewlett-Packard dramatically increases and then decreases dramatically closer towards 2017, IBM stays relatively stable, and AT&T dramatically increases over time. The Aerospace companies on the other hand, steadily increase over the course of 20 years, which shows that Aerospace companies, because they fall into multiple categories, such as public transportation and software, are affected by the other category to some extent.

Chart, line chart

Description automatically generatedChart, bar chart, histogram

Description automatically generated

Based on Figures 10 and 11, the relationship between revenue and the inflation of food and beverages seems to indicate a completely different relationship than the other graphs. There is not a point in time where the value of Food and Beverages deflates, so it is hard to see if a company is successful despite deflation of goods. In regard to these companies above, all of their revenues increase over time but at different rates; PepsiCo and Coca-Cola slightly increase, Target, Costco, and Kroger steadily increase, and Walmart drastically increases in revenue. This follows the trend of inflation causing an increase in revenue.

Chart, line chart

Description automatically generatedChart, bar chart, histogram

Description automatically generated

In regard to Figure 12, the Household Energy sector inflates drastically over time, but the Household Energy companies, drastically increase until and then start to decrease after 2008, which is shown by Figure 13. While these companies increase during a period of inflation, they also decrease, implying that the success of a business in this category cannot simply be dependent on inflation but in fact other external causes, which in this case is most likely attributed to the Great Recession in 2008. This is further indicated by the divide between the constant inflation rate of Household energy post Great Recession and the revenues of the energy companies listed.

Chart, line chart

Description automatically generatedChart, bar chart, histogram

Description automatically generatedChart, bar chart

Description automatically generated

Figures 14 and 15 show that both the childcare and medical care sectors drastically inflate over the 20 year span. As shown in Figure 16, each company increases their revenues overall, with the exception of Merck. However, the company revenues seem to fluctuate once again during the Great Recession but otherwise follow the trend of increasing inflation rates of child and medical care. While Allstate and State Farm are Insurance companies, they also operate in other industries, so you need to look at the other sectors that they are involved in, like the automotive industry.

Chart, line chart

Description automatically generatedChart

Description automatically generated

The toys market drastically deflates over time as seen in Figure 17. However, as shown in Figure 18, the three companies in this category have a steady increase in revenue over time. All though this is true, these three companies have other goods and services that create their total revenue, so while this is an indicator of a good business, there are more to these companies than just toys.

Chart

Description automatically generatedChart, line chart

Description automatically generated

The trends between the companies in Figure 20 and TVs change in inflation, shown in Figure 19 20, are very similar to the trend between toys and these companies. They both have consistent negative inflation percentages. Walmart again drastically increases over time, while Target and Costco slightly increase, and Sears steadily decreases in revenue.Chart, line chart

Description automatically generatedChart, bar chart, histogram

Description automatically generated

Based on Figures 21 and 22, the aerospace companies’ relationship with public transportation inflation seems to hold up the trends we have seen in the previous graphs. These companies are all successful based on my criteria since the companies consistently increase in revenue after periods of inflation. They also seem to follow the trend of public transportation inflation as well.

Chart, line chart

Description automatically generatedChart, histogram

Description automatically generated

While this comparison between companies in the other category and the overall change in inflation, shown in Figures 23 and 24, might not be as specific as the other comparisons, we feel that this gives a good overall comparison between these companies without specific inflation rates and general inflation percentages. Based on these trends from earlier graphs, a good company would have increased revenue during a period of positive inflation and flat or increasing revenue during a period of deflation. However, these companies don't always follow these trends. There are periods of positive inflation where all companies significantly decreased in revenue.

# Gender Impact Analysis

To observe the overall number of women on the Fortune 500 list, as well as in each sector, two datasets were used. The first one had the year, company name, revenue, and whether or not the CEO was a female. This data set spanned from 1955 to 2021 and each observation was one of the Fortune 500 companies for a given year. The second data set had the top 1000 companies information such as rank change, profit, city, state, and many more, but it only had it for one year. However, although it had only one years worth of data, it was sufficient in joining the companies by name to also have the sector information. There were a total of 48 out of 393 NA values. Using the known values, a conclusion on sectors was made.

The years covered in the first dataset began from 1955. However, the information needed to answer the question started after 1955, thus the data was filtered to only include years after 1955. This is because the qualification to be considered for the Fortune 500 changed starting 1995. Prior, companies that were “industrial'' were added to the list, but that was later changed to all company types. Then for the second dataset, because there were a lot of variables unnecessary to this analysis including the city and state the company is located in,if it was profitable, website handle, ticker, and more. The only fields left after filtering were name, and sector. The datasets were merged to create one dataset that had the company name, rank, sector, revenue, and if the CEO was a female. They were merged using a full join so that the missing value would also be recorded. Then, the type of revenue had to be changed from a character to numeric so that it could be observed over time as a value. One drawback of using two different datasets was that in some cases, the same companies were listed under different names. This is why I believe there were 48 NA values for the sector for some companies.

In the combined dataset, the observations without a female CEO were marked as NA. Those values were changed to “No” to clearly demonstrate that that company, in that year, did not have a female at the top. Outliers, companies that made an extreme amount more than the median company on the Fortune 500, were not discarded from the list of observations because they are important in demonstrating the success of a company. The more of an outlier it is the higher the annual revenue is or the more “successful” it is. Also it is important to keep those to not discount the fact those companies may be more prone or less prone to hiring a female CEO which is also a part of the analysis. When looking at the differences in revenue per year between male and female CEO companies in a line graph, the female company graph was much more unstable compared to the male graph. This is most likely due to the fact that there are more observations of male companies which allows for a general pattern to be shown. However, because there are few female observations no official pattern can be observed. For that reason, a series of box plots was also graphed to compare the median values and see the spread so that sample size did not play a role and so that outliers could be noted. The reason a line graph was used initially was because both time and revenue are both continuous. To make the boxplots continuous, the years 1996-2021 were split into three different time periods and then they were graphed for comparison. The same goes for the rank where it was initially graphed as a line graph because both are continuous variables but then it was changed to a series of bar graphs.

Groups for every 100 ranks were made so that the number of female CEOs per year of that particular subset of ranks could be counted. This way, you could see the rank subset that allowed the most number of female CEOs.

Chart, scatter chart

Description automatically generated

Because the data sets were so big it was important to narrow the data down to just what was needed for the analysis. New variables of the number of certain observations needed to be made as well as new vectors with the list of possibilities.

First, the number of women per year on the Fortune 500 list was graphed to follow the growth and see how it has changed over time. The scatter plot displayed a linear growth with a line of best fit equation of y= 1.445x+2 where y indicates the number of women on the list and the x is the number of years after 1995. The correlation coefficient is 0.9132 which indicates a fairly strong correlation between year and the number of women on the list. Although there is a positive slope on the growth of women, in comparison to the Fortune 500, given the current data available, if society stays on the same trend, 2168 will be the year when there is an equal number of male and female CEOs, which is 146 years from 2022.

Chart

Description automatically generated with medium confidence

This slow growth can be seen when the scatter plot is scaled to 500 on the y axis. In comparison to the whole of 500, the actual number is much smaller despite the positive growth which makes sense as to why the predicted year for 250 females and 250 males is 146 years away.

Then, a sector analysis was done. A sector is what a company defines its goods and services which includes, Wholesale, Transportation, Technology, Finance, Energy, and Food & Beverage & Tobacco. An outer join between the first data set with the information on CEOs and the second dataset with the sector information made it possible to analyze how the sector might influence the likelihood of the number of CEOs. To look at the count of the number of females in different sectors, a series of bar graphs were used These graphs displayed the different rates of growth to see if there was a relationship between men and female run companies. If the graphs for the lines were similar then we can assume there is no difference.

Chart

Description automatically generated

To start the analysis, we added four data sets, one that included the Fortune 1000 companies from 2017 - 2020, this dataset is called “fortune” , another that included the companies in 2021, called “fortune\_1000”, one that included the global fortune 1000, called “int\_fortune” companies from years 2019 - 2021 and one that included the fortune 500 companies in 2020, called “Fortune\_500\_match”. Once all the data was loaded into R, we needed to create one single dataset with all the important variables. In order to accomplish this, we first had to clean the datasets. For the 2020 fortune 500 dataset, the variables city and state were in the same column. We separated this into two different columns called “city” and “state”. The fortune dataset did not include the state or sector of the businesses. Using the function mutate and match, we compared the companies’ names in the fortune dataset to the names in the fortune\_1000, int\_fortune, and fortune\_500\_match datasets, and if the companies had the same name, then the state and sector will be added to the fortune dataset.

After creating a dataset with the most important variables, we have to add an additional year to it. Currently, fortune only consists of data from the years 2017 -2020, however, because we are studying the effects from Covid-19, it is important to add 2021. To add the data from 2021 to fortune we use the fortune\_1000 dataset. We started by selecting only the columns that we are using from each set. This included company name, rank, profit, sector, revenue, state, and number of employees. After this we had to rename the column to make sure both datasets had the same names. Once they had the same names and number of columns, we used the rbind function to add the 2021 fortune 1000 to the fortune dataset.

Regarding missing values, the only values that were NA were the states and sectors for some companies. This was due to the dataset we used to match the companies' names had some with slightly different spellings. We decided to remove the missing values because there were not enough to skew our data and we are examining the mean revenue and profit, therefore having one less value will have little impact on the analysis.

We decided to keep the outliers in the data because the outliers may bring in the majority of the revenue or profit of the sector. We cannot exclude outliers because then we may be missing a large portion of the revenues and profits that the sector generates.

Once we had all the data into one dataset, we started by creating a graph of the distribution of the number of companies in each sector from 2017 - 2021. Although this graph does not show much, we were able to see which sectors had the most companies and we were also able to realize that in order to get a better visualization of our data, we need to reduce the number of sectors.

To combine sectors, we used the ifelse function and tried to group them by how they are related. Retailing and Apparel became Apparel, Food and Drug Stores, and Food, Beverages and Tobacco became Food, Beverages and Drugs Stores, Household Products and Hotels, Restaurants and Leisure became Leisure and Home goods, Financials and Business Services because Business and Finance, materials, Industrials, Aerospace and Defense, Motor Vehicles and Parts and Engineering and Construction became manufacturing. Chemicals and energy became, chemicals and energy, and lastly, media and telecommunications became communication. Healthcare, technology and transportation all remained the same.

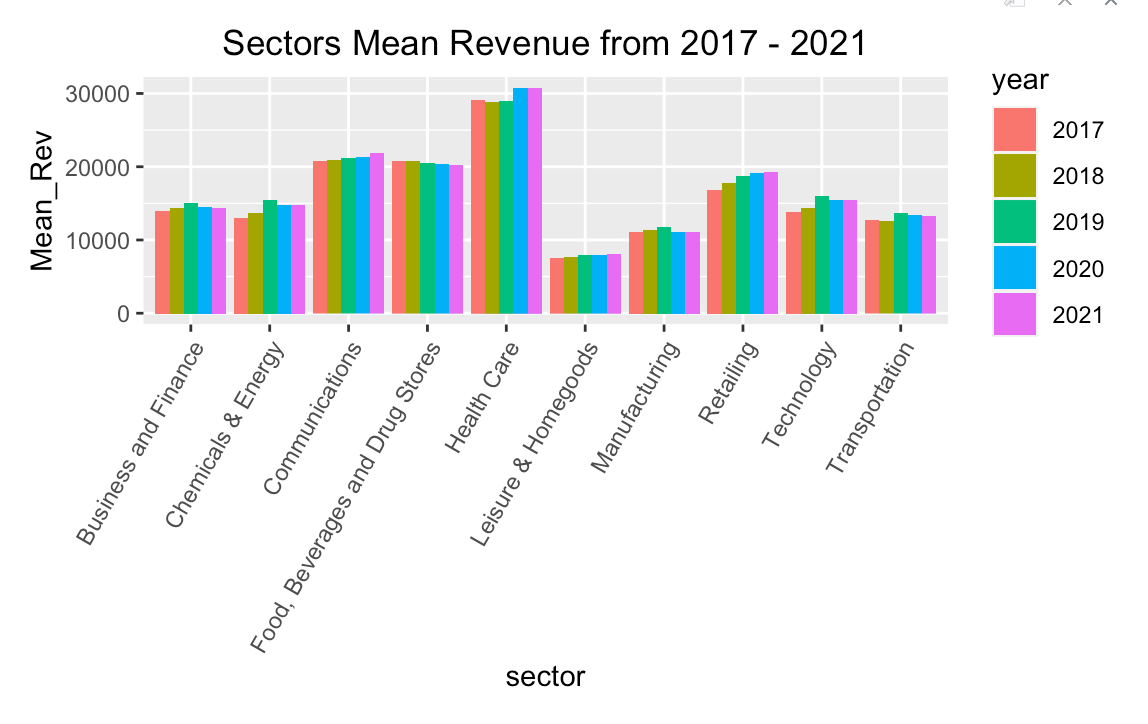
Once the data was grouped, we then graphed it again.

As seen above, it is a lot easier to read and understand the data now that there are only a few sectors.

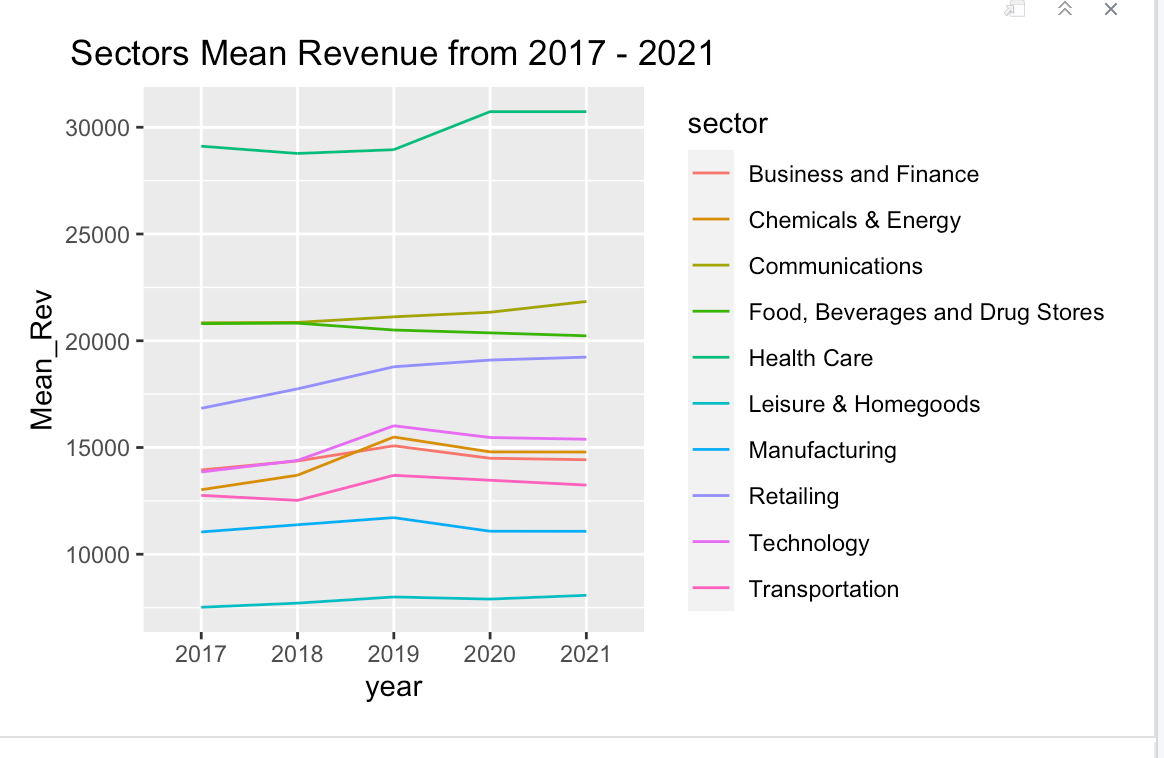
Following this we found the mean revenue and profit of each sector throughout the years and added these values to a column in the fortune dataset called Mean\_Rev and Mean\_Profit.

**Revenue of Sectors**

Using the Mean\_Rev, we created a bar graph for each year.



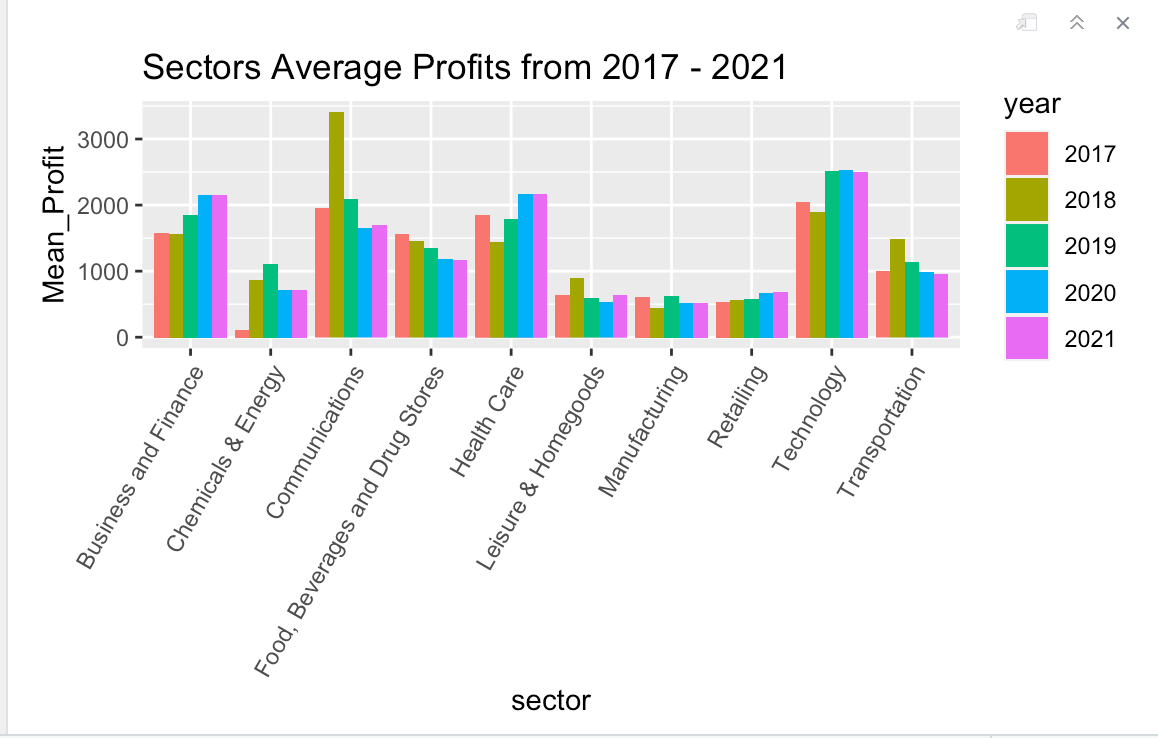
As seen in the bar graph above, the mean revenues trends of each sector from 2017 - 2021 seem to be consistent. There are no large increases or decreases in revenue that appear like an outlier. To further analyze the mean revenues through these years, we created a line graph as well.



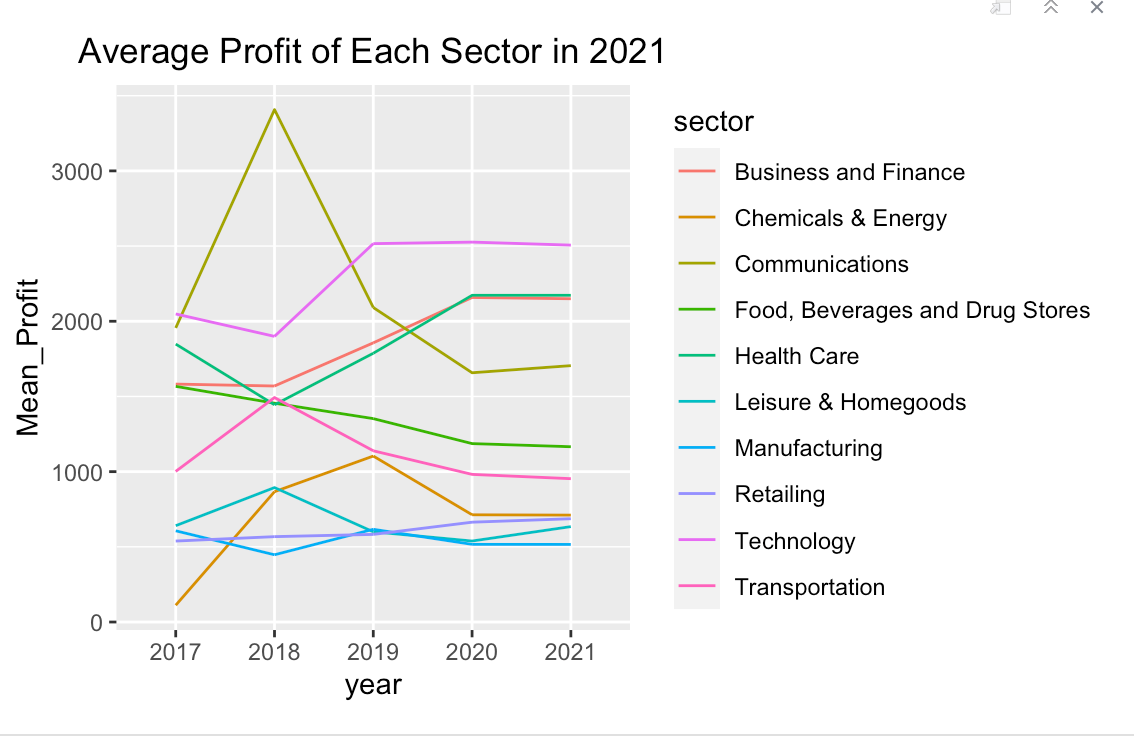
The line graph is the most helpful in this analysis. It gives us a more clear visual of the trends in revenue. This shows that during covid (2020 - 2021) there was no huge change in revenue. Surprisingly, it also shows that covid did not decrease the company's revenue (from 2019 - 2020).

**Profits of Sectors**

Next we looked at the average profit per year.



Through this graph we are able to see the changes of mean profit form 2017 - 2021. The changes here are greater than the changes in revenue. Some sectors like Business and Finance, Energy and Chemicals and Health Care show different trends during the covid years.

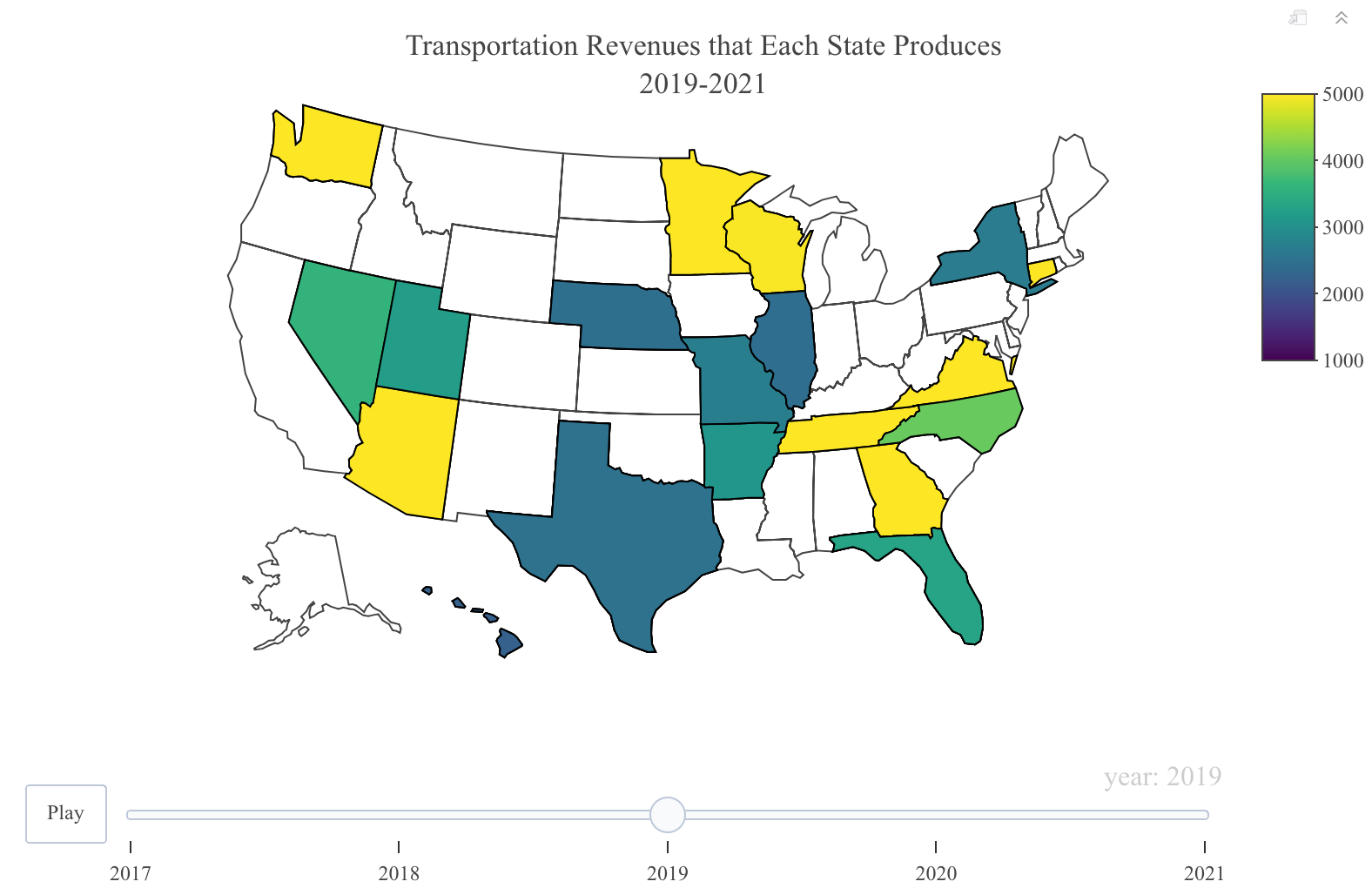


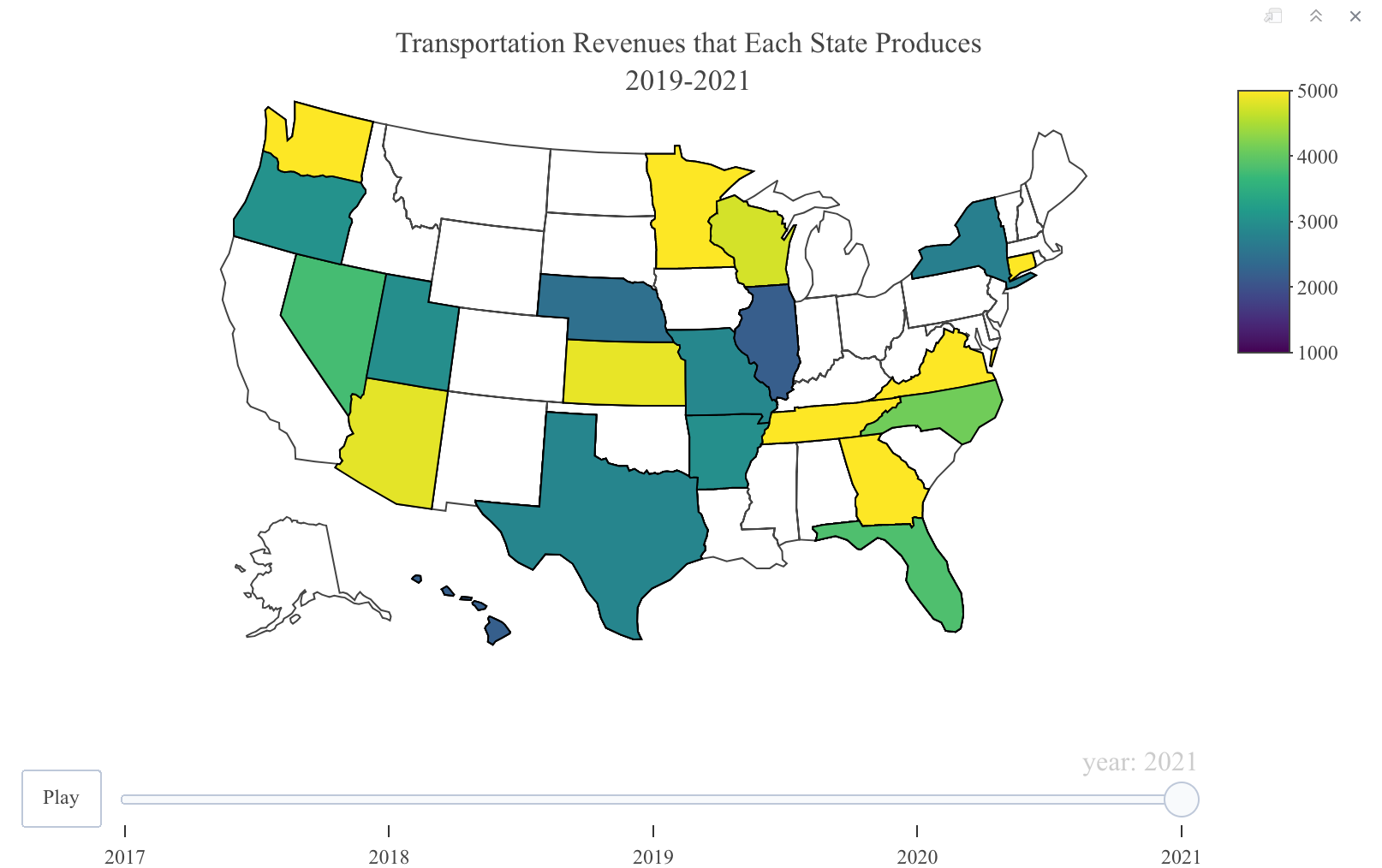
This graph also shows the changes in average mean profit. Although this graph is help[ful in pointing out trends, it can be somewhat difficult to read because of how many overlapping profits. It can be hard to differentiate between the sectors.

**States**

Following the analysis of the revenue and profit of each industry, we looked at data regarding different states. We took the sectors that were affected the most by Covid (Food Beverages and Drug Stores, Transportation, Chemicals and Energy, and Business and Finances) and we compared their revenue per state from 2019 - 2021. To do so we had to load in the plotly, dplyr and readr packages. We also had to add a new dataset called statecode. This dataset allowed us to convert the name of the state to the code. In order to do this we used the match function. We needed the code of the state because when we graph the revenue, the graph will use the state code, not the state name. We then made different datasets for each sector that will be analyzed by using the factor tool.

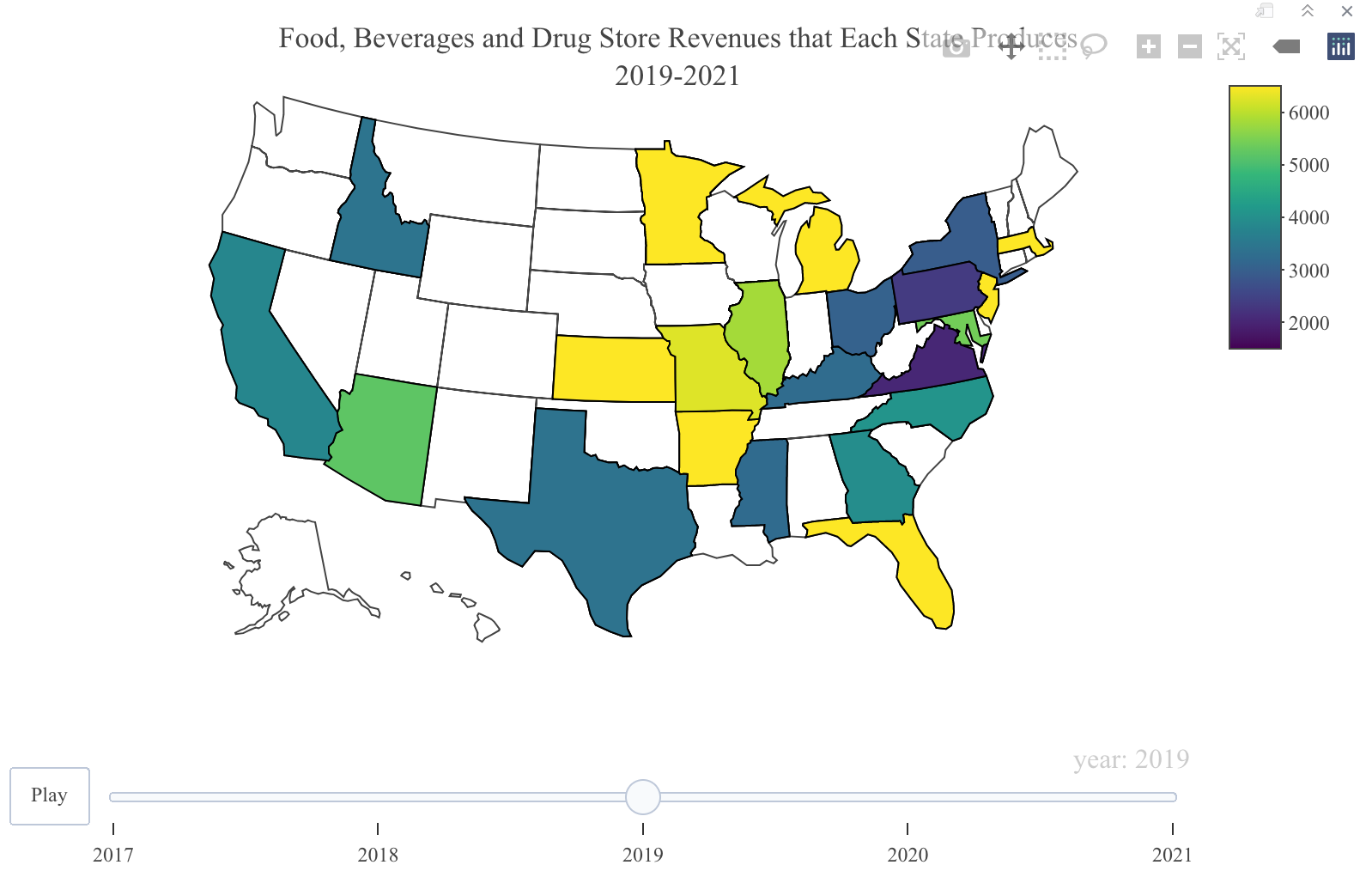
We started by analyzing the transportation revenue in different states from 2019 - 2021.

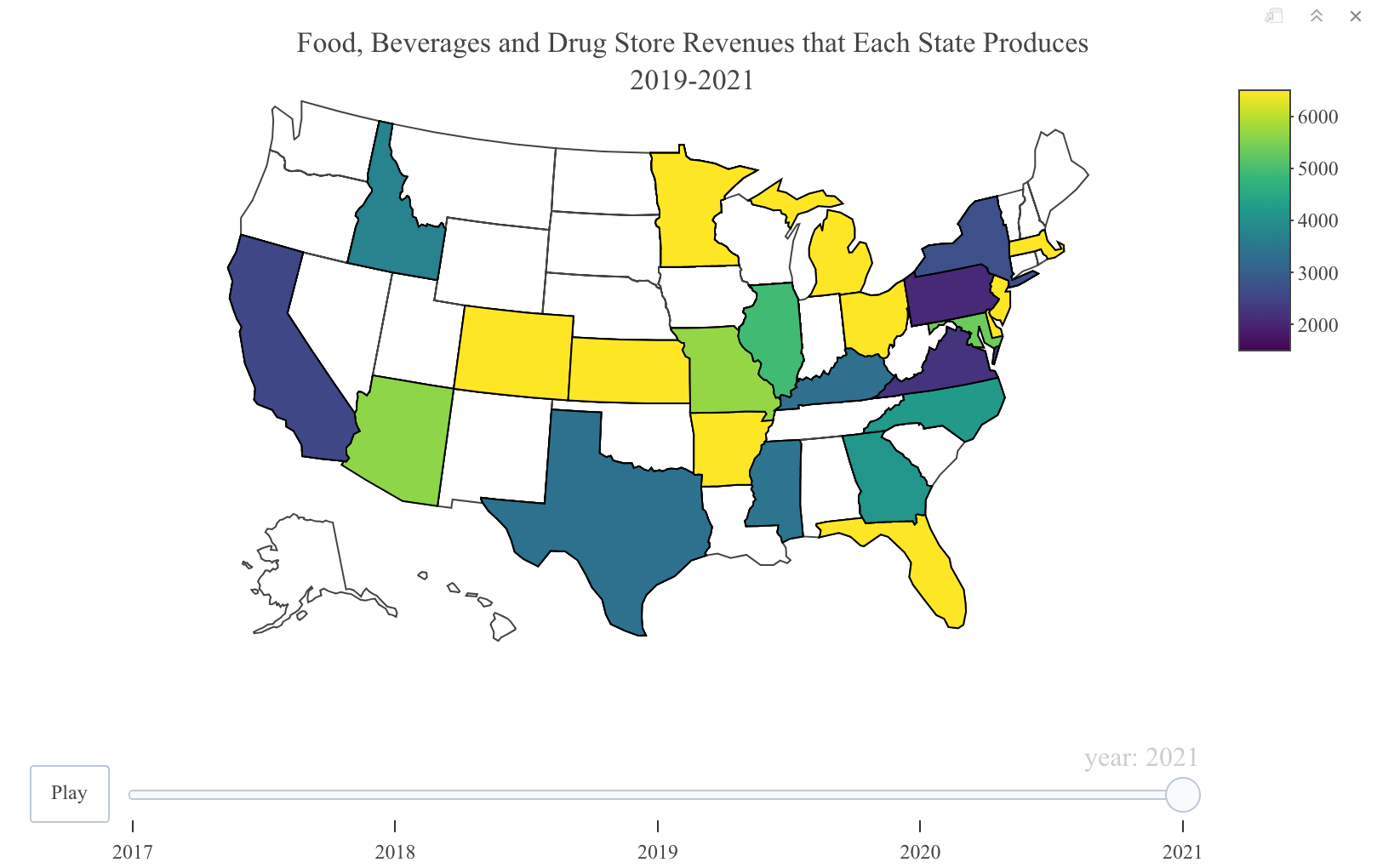




As seen in the two graphs above, the main places that transportation revenues increased are in the central United States. Kansas, which had no transportation fortune 1000 companies in 2019, now has one of the largest transportation revenues.

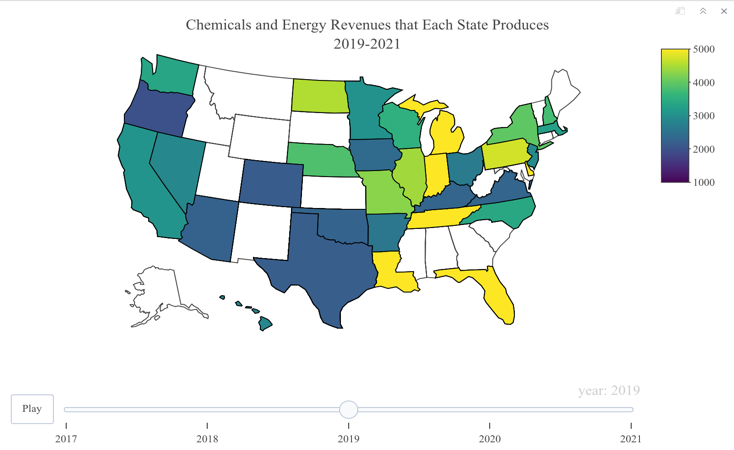
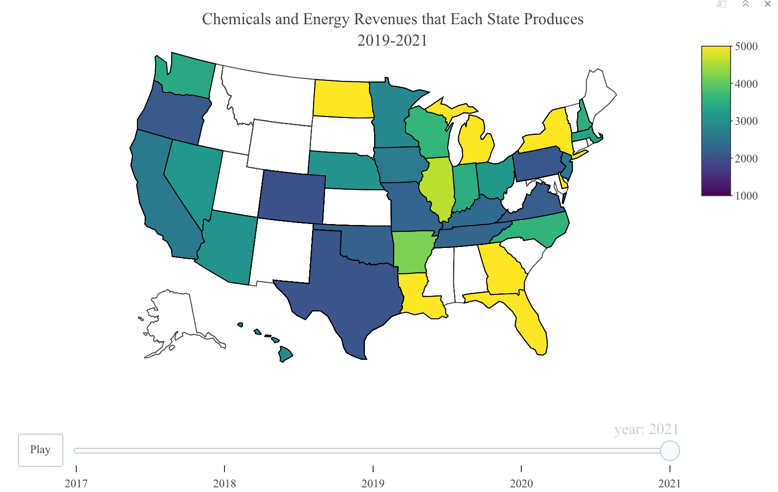
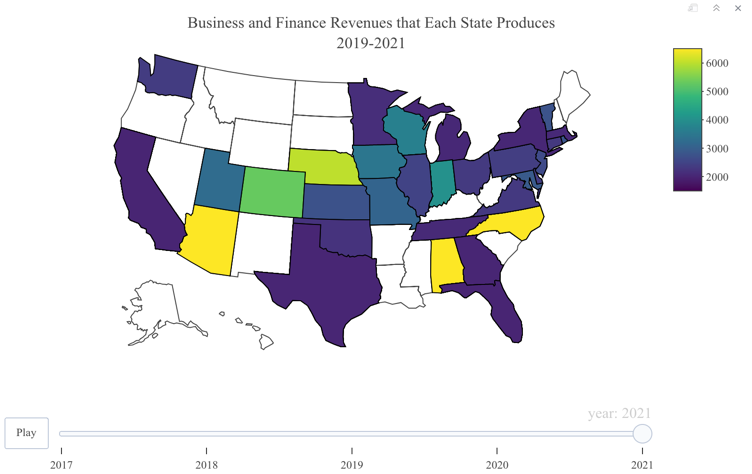
The next sector that we analyzed was Food, Beverages and Drug Stores.





The revenues per state of the Food, Beverages and Drug store sector follow a similar trend as the transportation sector. More of the revenue is coming from the central United States and less is coming from coastal regions like California and New York.

Following Food, Beverage and Drug Stores, we looked at the trends in the Chemicals and Energy Sector.

A picture containing timeline

Description automatically generated

Chemicals and Energy seem to follow a different trend than the ones seen above. Almost all of the states’ revenues decreased. The ones that increased consist of Georgia, New York and North Dakota.

Chart, bar chart

Description automatically generated

Lastly we looked at the Business and Finance sector.

This trend follows a similar pattern to the Energy and Chemical sector. It appears that almost all states’ revenues decreased from 2019 - 2021. One state that revenue increased in was New Hampshire.

However, there were many more NA values than anticipated which also makes it hard to have a real conclusion. I then removed the NA values to see how the distribution looked despite the unknown sector companies. .

In conclusion, when looking at the distribution for the known sector companies, Food & Drug stores have the most female CEOs while Chemicals, Engineering & Construction, Telecommunications, Cosmetics, and Media have the least. A follow-up question that would be interesting would be to look at the proportion in each sector that has a female CEO. When looking merely at the count, an issue that may occur is that the Fortune 500 is disproportionately much more of one sector. In that case, overall, that sector may have many more female CEOs compared to other sectors, but it may be a misrepresentation when looking at the proportion.

To observe differences in revenue between male and female run companies a line graph and a series of box plots were made. Initially a line graph was used to see the revenue difference over time. Three lines were graphed: Female, Fortune 500, and Male.

Chart, line chart

Description automatically generated

The graph has a hole in it from the year 2015 because the data set was missing the revenues for that year. The Fortune 500 line and the Male line are very similar because such a larger percent of the companies are males so the weight of the male owned companies will play a bigger role in the average. The female line is very unstable which can be attributed to the fact that there are fewer sample values, so it is hard to have a clear pattern, but the graph shows that except for a few years, on average, the female CEO companies performed better than their male counterparts. To solve the issue of having a small sample size, a box plot can be created to look at the median value. The first box plot shows the total difference in revenue form 1995- 2021. The third quartile value is much higher, and the median is slightly above that of female CEOs, but male CEO companies have more outlier values. Although the median is about the same, the interquartile range is much larger for that of females.

**Chart

Description automatically generated**

Because societal standards for business change over time, such as being more inclusive, the box plot was further divided into three groups each a third of the years from 1995-2021.

Chart, box and whisker chart

Description automatically generated

In all three-year categories, women run companies outperformed their male counterparts in interquartile range. This means that generally, female owned companies did better in revenue but because male owned companies had far more outlier values which would indicate an extremely high revenue company compared to the other Fortune 500 companies.

Chart, bar chart, histogram

Description automatically generatedChart, bar chart, histogram

Description automatically generatedWhen looking at rank, it is important to remember that a lower rank signifies a more successful company. Because companies are ranked based on revenue, the higher the revenue, the lower rank. First, male CEO companies were filtered out and then the data was grouped by every 100 ranks, for example 1-101. After looking at the graphs, you can see that ranks 201-300 have the highest volume of CEOs with the last 100 ranks and top 100 ranks having the least. This matches the information found on the boxplots about the revenues. The lower rank companies have much higher revenue and are likely to be the outlier values found on the boxplots. Seeing as to how little female CEOs there are in the top 100 ranked companies, the fact that there are so many more outliers for the male companies makes sense.

Chart, bar chart, histogram

Description automatically generatedChart, bar chart, histogram

Description automatically generatedChart, bar chart

Description automatically generated

# Covid-19 Impact Analysis

To start the analysis, we added four data sets, one that included the Fortune 1000 companies from 2017 - 2020, this dataset is called “fortune” , another that included the companies in 2021, called “fortune\_1000”, one that included the global fortune 1000, called “int\_fortune” companies from years 2019 - 2021 and one that included the fortune 500 companies in 2020, called “Fortune\_500\_match”. Once all the data was loaded into R, we needed to create one single dataset with all the important variables. In order to accomplish this, we first had to clean the datasets. For the 2020 fortune 500 dataset, the variables city and state were in the same column. We separated this into two different columns called “city” and “state”. The fortune dataset did not include the state or sector of the businesses. Using the function mutate and match, we compared the companies’ names in the fortune dataset to the names in the fortune\_1000, int\_fortune, and fortune\_500\_match datasets, and if the companies had the same name, then the state and sector will be added to the fortune dataset.

After creating a dataset with the most important variables, we have to add an additional year to it. Currently, fortune only consists of data from the years 2017 -2020, however, because we are studying the effects from Covid-19, it is important to add 2021. To add the data from 2021 to fortune we use the fortune\_1000 dataset. We started by selecting only the columns that we are using from each set. This included company name, rank, profit, sector, revenue, state and number of employees. After this we had to rename the column to make sure the both datasets had the same names. Once they had the same names and number of columns, we used the rbind function to add the 2021 fortune 1000 to the fortune dataset.

In regards to missing values, the only values that were NA were the states and sectors for some companies. This was due to the dataset we used to match the companies names had some with slightly different spellings. We decided to remove the missing values because there were not enough to skew our data and we are examining the mean revenue and profit, therefore having one less value will have very little impact on the analysis.

We decided to keep the outliers in the data because the outliers may bring in the majority of the revenue or profit of the sector. We cannot exclude outliers because then we may be missing a large portion of the revenues and profits that the sector generates.

Once we had all the data into one dataset, we started by creating a graph of the distribution of the number of companies in each sector from 2017 - 2021. Although this graph does not show much, we were able to see which sectors had the most companies and we were also able to realize that in order to get a better visualization of our data, we need to reduce the number of sectors.

Chart

Description automatically generated

To combine sectors, we used the ifelse function and tried to group them by how they are related. Retailing and Apparel became Apparel, Food and Drug Stores, and Food, Beverages and Tobacco became Food, Beverages and Drugs Stores, Household Products and Hotels, Restaurants and Leisure became Leisure and Homegoods, Financials and Business Services because Business and Finance, materials, Industrials, Aerospace and Defense, Motor Vehicles and Parts and Engineering and Construction became manufacturing. Chemicals and energy became, chemicals and energy, and lastly, media and telecommunications, became communication. Healthcare, technology and transportation all remained the same.

Once the data was grouped, we then graphed it again.

Chart, histogram

Description automatically generated

As seen above, it is a lot easier to read and understand the data now that there are only a few sectors.

Following this we found the mean revenue and profit of each sector throughout the years and added these values to a column in the fortune dataset called Mean\_Rev and Mean\_Profit.

**Revenue of Sectors**

Using the Mean\_Rev, we created a bar graph for each year.

A picture containing text, writing implement, stationary, pencil

Description automatically generated

As seen in the bar graph above, the mean revenues trends of each sector from 2017 - 2021 seem to be consistent. There are no large increases or decreases in revenue that appear like an outlier. To further analyze the mean revenues through these years, we created a line graph as well.

Chart, line chart

Description automatically generated

The line graph is the most helpful in this analysis. It gives us a more clear visual of the trends in revenue. This shows that during covid (2020 - 2021) there was no huge change in revenue. Surprisingly, it also shows that covid did not decrease the company's revenue (from 2019 - 2020).

**Profits of Sectors**

Next we looked at the average profit per year.

A screenshot of a graph

Description automatically generated with medium confidence

Through this graph we are able to see the changes of mean profit form 2017 - 2021. The changes here are greater than the changes in revenue. Some sectors like Business and Finance, Energy and Chemicals and Health Care show different trends during the covid years.

Chart, line chart

Description automatically generated

This graph also shows the changes in average mean profit. Although this graph is help[ful in pointing out trends, it can be somewhat difficult to read because of how many overlapping profits. It can be hard to differentiate between the sectors.

**States**

Following the analysis of the revenue and profit of each industry, we looked at data regarding different states. We took the sectors that were affected the most by Covid (Food Beverages and Drug Stores, Transportation, Chemicals and Energy, and Business and Finances) and we compared their revenue per state from 2019 - 2021. To do so we had to load in the plotly, dplyr and readr packages. We also had to add a new dataset called statecode. This dataset allowed us to convert the name of the state to the code. In order to do this we used the match function. We needed the code of the state because when we graph the revenue, the graph will use the state code, not the state name. We then made different datasets for each sector that will be analyzed by using the factor tool.

We started by analyzing the transportation revenue in different states from 2019 - 2021.

A picture containing graphical user interface

Description automatically generated

A picture containing graphical user interface

Description automatically generated

As seen in the two graphs above, the main places that transportation revenues increased are in the central United States. Kansas, which had no transportation fortune 1000 companies in 2019, now has one of the largest transportation revenues.

The next secor that we analyzed was Food, Beverages and Drug Stores.

A picture containing chart

Description automatically generated

A picture containing chart

Description automatically generated

The revenues per state of the Food, Beverages and Drug store sector follow a similar trend as the transportation sector. More of the revenue is coming from the central United States and less is coming from coastal regions like California and New York.

Following Food, Beverage and Drug Stores, we looked at the trends in the Chemicals and Energy Sector.

A picture containing graphical user interface

Description automatically generated

A picture containing chart

Description automatically generated

Chemicals and Energy seem to follow a different trend than the ones seen above. Almost all of the states’ revenues decreased. The ones that increased consist of Georgia, New York and North Dakota.

Lastly we looked at the Business and Finance sector.

A picture containing timeline

Description automatically generated

A picture containing graphical user interface

Description automatically generated

This trend follows a similar pattern to the Energy and Chemical sector. It appears that almost all states’ revenues decreased from 2019 - 2021. One state that revenue increased in was New Hampshire.

**Discussion of Results and Findings**

Our analysis on factors affecting success of Fortune 500 companies gleaned several interesting insights and conclusions. ….

# Technology Analysis Results

# Inflation Analysis Results

At the beginning of this analysis, we asked the question: How have changes in price/inflation of goods affected the success of businesses from 1997 to 2017? Once we asked this question, we thought that potential variables from the price\_changes and fortune500 datasets for examination would include: Year, Revenue, and the inflation rate of each market (Clothing, New\_cars, etc.). We decided to further examine the Revenue and Name variables because their data seemed to be the easiest to demonstrate the success of companies; when we think of successful companies, we think of the revenues of their companies, which is why the combination of these two variables makes the most sense to compare.

After filtering, merging, and graphing the data, we have found the following trends. As shown in Figures 4 and 5, car companies have an inconclusive correlation with the inflation rate of new cars. We concluded that there is an inconclusive correlation because before 2010, there is an inverse relationship while the relationship after 2010 is direct. Insurance companies have an inconclusive correlation. As shown in Figures 4 and 5, Insurance companies have an indirect relationship with the inflation rate of new cars, however, they also have a direct correlation to the inflation rates of medical care and childcare, which are shown in Figures 14, 15, and 16. As shown in Figures 6 and 7, there is a correlation between the inflation rate of clothing and the revenues of the companies operating in that sector, which are supermarkets and department stores. This relationship is inverse by nature because when the inflation rate of clothing rapidly deflates in value, the revenues of most of the companies increase, which makes sense as people tend to buy more clothes when they are cheaper. The relationship between the revenues of software companies and the inflation rate of software is also an inverse relationship, which is shown by Figures 8 and 9. As the inflation rate of software significantly decreases, the revenues of every company increase. The relationship between the inflation rate of food and beverages and the companies that operate in that sector, which are supermarkets and food and beverage companies, show a correlation. This correlation, as shown in Figures 10 and 11, is direct in nature because the revenues of these companies increase in value as the inflation rate of food and beverages increases. There is also a direct relationship between household energy companies and the inflation rate of household energy even with the Great Recession. As shown in Figures 12 and 13, while the inflation rate is always positive, the revenue decreases when there are dips in the inflation rate and vice versa, which is why we concluded that this is a direct correlation. Healthcare companies and the inflation rates of medical care and childcare have a direct correlation as well. As shown in Figures 14, 15, and 16, the inflation rates of medical care and childcare both increase while the revenue of healthcare companies increase as well; this makes sense because people will always need healthcare and childcare. The relationship between the inflation rate of toys and the companies that operate in this sector, which are supermarkets and department stores, have the same correlation to the relationship between clothing companies and the inflation rate of clothing. This inverse correlation is demonstrated in Figures 17 and 18, which show that the revenues of these companies increase when the inflation rate of toys decreases. The relationship between the inflation rate of TVs and the companies that operate in this sector, which are supermarkets and department stores, also have the same correlation to the relationship between clothing companies and the inflation rate of clothing. This inverse correlation is demonstrated in Figures 19 and 20, which show that the revenues of these companies increase when the inflation rate of TVs decreases. Aerospace companies show a direct correlation between their revenue and the inflation rate of public transportation, which is depicted in Figures 21 and 22. The companies in the Other category have no correlation with the total inflation rate, which is depicted in Figures 23 and 24 and is shown by the sporadic behavior of the company revenues in this category that do not follow the trend of inflation at all.

Based on these findings, we can see that seven out of ten of the company types have a correlation with the inflation rates in their categories. With most of the companies fitting our hypothesis, we have determined that we have enough information to come to a conclusion. I conclude that inflation rates do, in fact, affect the revenues of companies and thus, their success. While most companies fit the hypothesis, we also conclude that the relationship between inflation and revenue depends on the sector in consideration as some companies, like software companies for example, have an inverse relationship with the inflation rate of software, while other companies, like aerospace companies for example, have a direct relationship with the inflation rate of public transportation.

# Gender Analysis Results

When looking at the data for the number of female CEOs per year there has been a steady increase however when comparing it to the whole of the Fortune 500, the fact that the growth is actually very slow can be observed. When divided by sector again, you can see that certain sectors are more prone to having female CEOs such as food & drug stores have the most female CEOs while Chemicals, Engineering & Construction, Cosmetics, and Media have the least. In regards to revenue there does seem to be some level of correlation between male CEO companies having many more outliers which may or may not be attributed to the fact that there are more male CEO companies. The interquartile range is consistently higher for female run companies with them reporting significantly higher third quartile ranges which means the top 25% of female run companies do better than the top 25% of male run companies. There have been less female CEOs in the top 100 and bottom 100 companies on the Fortune 500 list and and more in the middle 300.

These results demonstrate some of the ways that successful businesses may be discriminatory towards women. Women are not being given the same chances for upward mobility as men. However, it does not seem to be a problem with competency seeing as to how female run companies seem to be doing better than male run companies. Also, you can see the difference in how sectors play a role in that as well. Women are more likely to be picked as CEO in more stereotypically female correlated sectors such as food & grocery and less likely in technological fields such as chemical engineering. In conclusion, certain companies are more likely to choose a female CEO over others.

We did not think there was going to be a trend of female CEO companies doing better, but I originally thought it was going to be equal. This may have happened because the sample size was too small or it may be true because women have to be more qualified for the same position which means they are more qualified than their counterparts on the list which results in higher revenues.

# Covid-19 Analysis Results

The trends over the past few years of revenues for different sectors has been a slow yet steady increase. However, for a lot of sectors during the years between 2020 -2021, it appears that the changes in revenue have become almost negligible. Sectors such as health care and retailing seem to have benefited from Covid in terms of revenue, this can be seen in the line graph and box plot.

The profit analysis trends are slightly less consistent than the revenue trends. Sectors such as Business and Finance and health care, seemed to increase in profits from 2019 - 2021, while sectors such as Chemicals and Energy, communications, Food beverages and drug stores, and transportation decreased.

Revenues of each state follow similar trends. It appears that both variables are decreasing in coastal states and increasing in states within the midwest. This could be due to a lot of the states in the midwest having less covid regulations, therefore, the companies are able to operate normally.

This analysis shows that the sector that benefited from Covid-19 the most is health care because it increased in both profits and revenues. Other businesses that also benefited are Business and Finance and retailing. The sectors that did not benefit are Chemicals and energy, communications, food beverages and drug stores and transportation.

References