Brielle Bernardy
Haddi Fadia
Karen Fan
Jessica Warner

Social Security Administration Analysis

Introduction

Social Security reaches almost every family and provides funding in the source of supplemental income. Social Security helps older Americans, workers who develop disabilities, and families where a parent or spouse dies. In June 2022, about 51 million retirees and their families received Social Security payouts, which is the category of beneficiary that received the most aid. Social Security replaces a percentage of a worker's pre-retirement income based on their lifetime earnings, and is a benefit that many Americans look forward to around the age of retirement.

However, according to the 2022 annual report of the Social Security Board of Trustees- this funding may be depleted by 2035; the system will exhaust its cash reserves and pay out only what it takes in yearly through Social Security taxes.

Project Background

We focused our analysis on exploring the Social Security Payout over the period of a decade (from 2010 to 2020). While exploring through the data, some of the question we found interesting were:

What is the trend of Social Security payout over time (2010- 2020)?
What is the average amount of payout per individual?
What is the average amount of payout by State?
What is the average amount of payout by Category (Retirement, disability, survivors)?

Finding an answer to these questions were motivated by the fact that everyone is affected by Social Security in one way or the other. Either by being in the workforce contributing to Social Security taxes on a bi-weekly basis through a paycheck or by being on the receiving end through retirement or as a surviving spouse.

Data Collection

The data for our project was collected from the Social Security Administration website. They publish data every year on individuals who are paid in December of each year and what amount was paid out to those individuals. We used data from 2010, 2015, and 2020. This is to show the payouts over time.

The data is broken down by:
- State or outlying areas,
- County,
- Type of benefit (retirement, survivors, disability), and
- Sex of beneficiaries over age 65.

The data was separated by the number of beneficiaries and amount of benefits paid. We left this data separated, but then merged the csv files for cleanup using pandas in Jupyter Notebook.

```python
# Read CSV.files
aob_2010 = pd.read_csv("data/aob_2010.csv", thousands=',')
aob_2015 = pd.read_csv("data/aob_2015.csv", thousands=',')
aob_2020 = pd.read_csv("data/aob_2020.csv", thousands=',')
nob_2010 = pd.read_csv("data/nob_2010.csv", thousands=',')
nob_2015 = pd.read_csv("data/nob_2015.csv", thousands=',')
nob_2020 = pd.read_csv("data/nob_2020.csv", thousands=',')
```

```python
# Merged Files
df_2010 = pd.merge(aob_2010, nob_2010, on=["State or area", "County"], how='outer')
df_2015 = pd.merge(aob_2015, nob_2015, on=["State or area", "County"], how='outer')
df_2020 = pd.merge(aob_2020, nob_2020, on=["State or area", "County"], how='outer')
```

Data Exploration and Cleanup

Before we could do any manipulation of our data, we needed to prepare and clean our data. After reading in our csv files, we viewed the data types and created a dictionary to convert almost all of our columns from strings into floats. We wouldn't have been able to perform any data manipulation if the data type stayed as strings. We then merged our files because some of the data from multiple csv files would be necessary to answer some of our questions. Another form of data cleaning we did was looking for any null values that might skew our data or at least not be necessary, but also keeping in mind to not change our data and how we wanted it presented. We did have an instance of a null value for an entire row, so we decided to remove it from the dataset because it would have very little effect on our analysis by leaving it in and would actually cause issues with our charts and graphs. Another minor change we made in preparing our data was to rename some of the columns.

```python
# Convert data types for nob 2020
convert_dict = {"Total": float,
                "Retirement Retired workers": float,
                "Retirement Spouses": float,
                "Retirement Children": float,
                "Survivors Widow(er)s and parents": float,
                "Survivors Children": float,
                "Disability Disabled workers": float,
                "Disability Spouses": float,
                "Disability Children": float,
                "Aged 65 or older Men": float,
                "Aged 65 or older Women" : float}
nob_2020 = nob_2020.astype(convert_dict)
```

```python
# Merged Files
df_2010 = pd.merge(aob_2010, nob_2010, on=["State or area", "County"], how='outer')
df_2015 = pd.merge(aob_2015, nob_2015, on=["State or area", "County"], how='outer')
df_2020 = pd.merge(aob_2020, nob_2020, on=["State or area", "County"], how='outer')
```

```python
# Sum of NaN values in df_2010
df_2010.isna().sum()
```

```python
# Drop outlying areas in df_2010
df_2010.drop(df_2010[df_2010['State or area'].str.contains('Outlying areas')].index, inplace=True)
```

Major Findings

What is the trend of Social Security payout over time?

       The first thing I did was calculate the correlation between the number of beneficiaries and the amount of benefits paid for each county or area on the list for each year we looked at. The correlation for each set of data for 2010, 2015, and 2020 was +1.0. This is a perfect positive correlation between the amount of benefits paid out and the number of beneficiaries. Incremental steps of one variable is exactly proportional to the value change of the other variable.
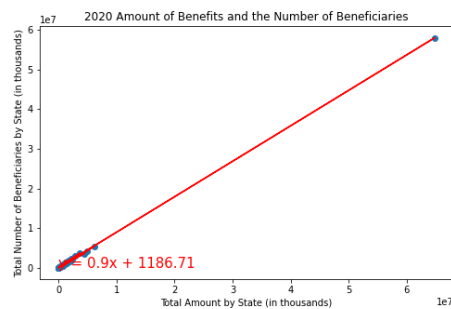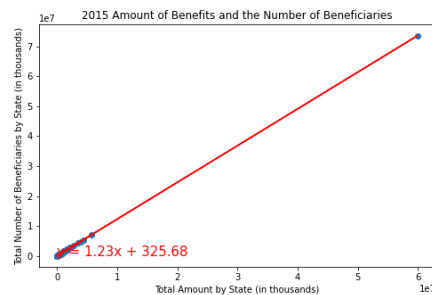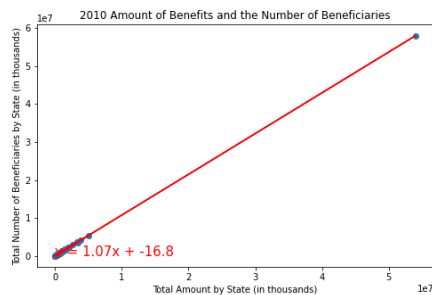
       In order to determine the trend of the payout overtime, I grouped the data by state or area and then summed all the counties for each state or area to determine the average payout in thousands per individual over a decade.

The following three bar graphs show the average payout over a decade for Minnesota, Wisconsin, and all state and outlying areas. All three figures depict the same trend. An increase in payouts from 2010 through 2015 and a decrease from 2015 through 2020.

Average Payout Over a Decade

Minnesota Average Payout Over a Decade

Wisconsin Average Payout Over a Decade
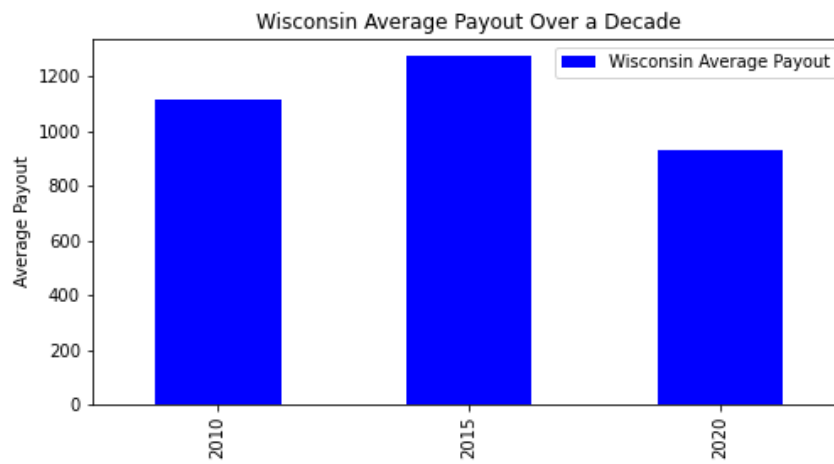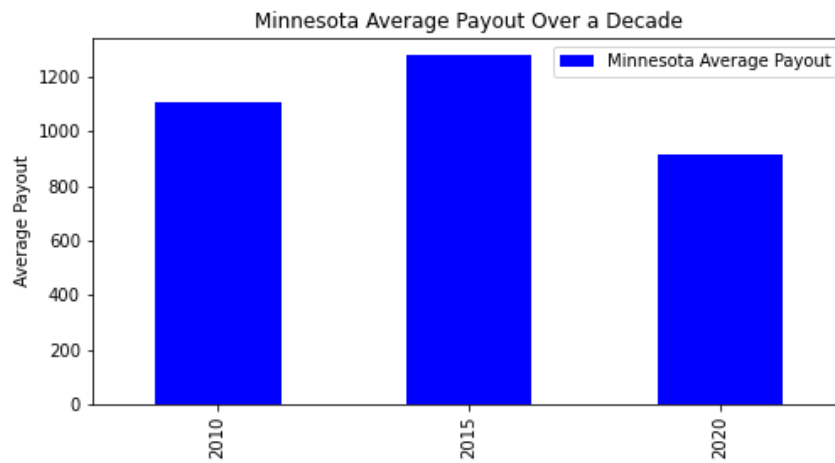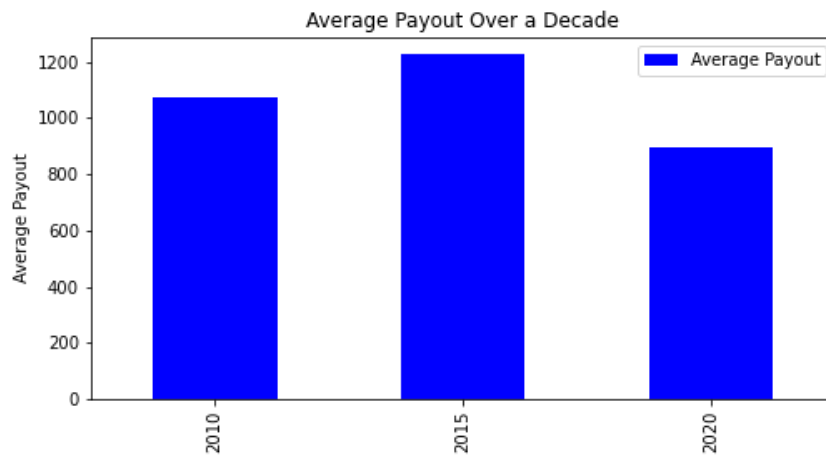
There was a 14.3% increase in the average payout from 2010 - 2015 in the month of December. There was a 27.1% decrease in the average payout from 2015 - 2020 in the month of December.

If you look at the percent change over a 10 year period, there was a 16.6% decrease from 2010-2020 in average payout for the month of December.

Some questions that I have after answering this question is why did the number of beneficiaries decrease from 2015 to 2020? Our world was experiencing a global pandemic that had severe effects on the elderly. Did the COVID-19 global pandemic affect the number of people who qualified for Social Security payouts resulting in a drastic decline in payouts? Did less people working affect the amount of money brought in versus the amount of money being paid out? This is something that if given more time I would want to look into and see if we could find an answer.
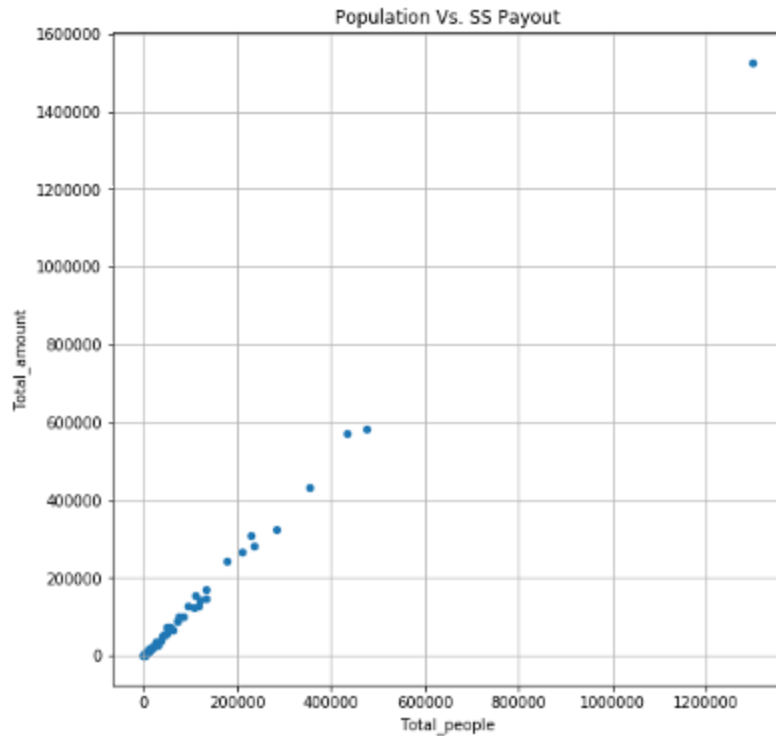
Question 2

While exploring the data, I calculated the average payout by individual for the year 2010, 2015 and 2020.

	2010 Avg Pay	2015 Avg Pay	2020 Avg Pay
0	1074.332421	1228.115789	895.104378

The average payout for 2010 was $1,074, 2015 was $1,228 and 2020 was $895. The 2020 average payout per person was lower versus 2010 and 2015. This could be due to the global covid-19 pandemic since the death rate in 2020 was significantly high. In addition, the calculation of Social Security payout is greatly affected by the economy and the economy in 2020 was in the low vs. 2015 and 2020. Also, since COLA - cost of living adjustment is factored into payout calculation, it is no surprise that the 2020 average payout is lower versus 2010 and 2015.
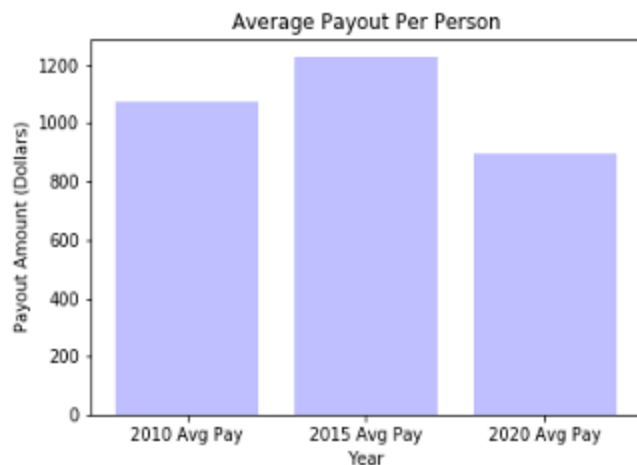
Furthermore, I explored the total Social Security payout by state for 2015 and found that California has the highest amount of Social Security payout versus all other states in the United States. With an overall total payout of 9% for the single state. That was not a shocker since California is the most populous state in the United States. Comparing the total California payout by county, we found that Los Angeles county received 22% of the overall California Social Security payout which was also not a surprise since Los Angeles County is the most populous county in California. The top five counties with the highest Social Security payouts in California were Los Angeles county, San Diego, Orange, Riverside and San Bernardino. I analyzed the relationship between population and Social Security payouts for the counties in California using a scatter plot. Based on the analysis, it was concluded that there is a positive correlation between the two. The higher the population per county, the higher the total Social Security payout. Los Angeles county showed as an outlier on the scatter plot because it is the most populous state with the most Social Security payout versus all the other counties in California.

Population Vs. SS Payout

Question 3

My analysis was to look into the big picture scale of average payouts per state and compare between 2015 and 2020 to see if there was a trend within the five years. A quick glance at the Average Payout by State graphs, first callout that I noticed was that the graphs look exactly the same between 2015 and 2020. This is because between 2015 and 2020, the average amount of payout by state all decreased within the five years. The average payout in 2015 was 31 million dollars compared to 2020's 24.3 million dollars. California, Connecticut and New Jersey have the highest number of payouts, with California having the max amount. (2015- 119 million to 2020- 92.8 million). This data shows that payout has decreased within the 5 year range.

2015 Average Payout by State

2020 Average Payout by State
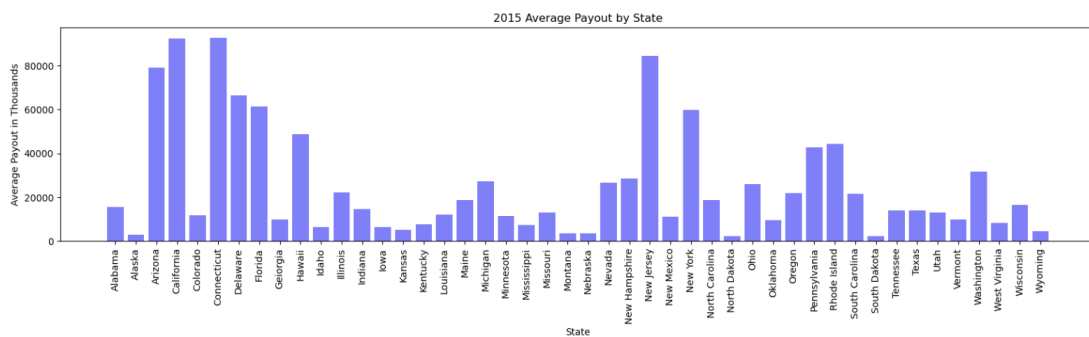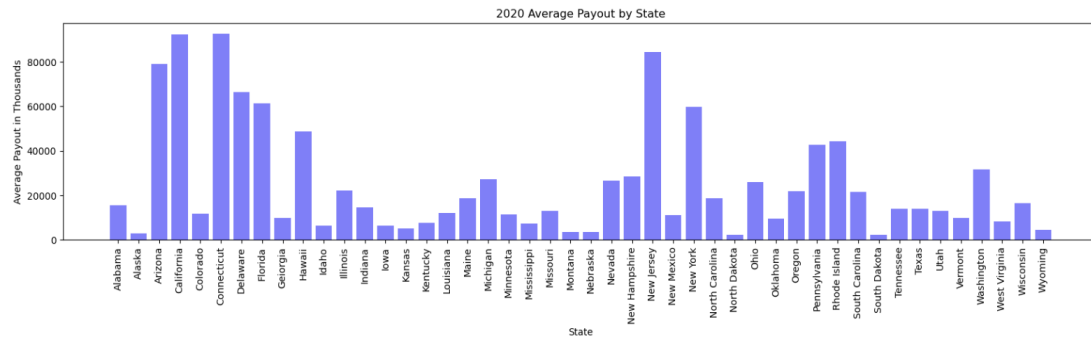
Question 4

      My goal for my part of the project included looking at the breakdown of payouts by Social Security type. For this, I decided to look at how the percentages changed for each state, rather than look at an average of all the states added together. My thought process behind this was that with percentage, I would be able to look at how the states distribute the Social Security by category and then compare them between the years 2015 and 2020. Additionally, I looked more closely at one of the category breakdowns, which was payout by gender, ages 65 and up and how those two compare across 2015 and 2020. For my data manipulation, I grouped by state and took the sum for each of the separate columns under each category. Then I looped through each state total for the broken down categories to find the total for each of the 4 main categories: retirement, survivor, disability and 65 and older. Specifically, I wanted to analyze how the percentages changed for the categories from 2015 to 2020, when the pandemic happened. My findings were that the survivor category received a greater percentage of the total payouts, while a different category like retirement decreased.

      For my second question, I looked at the difference in payouts by gender. For this I simply took the totals for each state and compared them side by side in a bar chart. Though, to get a closer look, I took 5 of the states that received most of the benefits and looked at if there was a difference in the amount of payouts. My analysis was that most of the payouts to men were higher in both years, but there were a few instances where women received a higher payout. And in comparison by year, generally it seems the distribution seemed to be about the same.

Payout by Gender 2015

Payout by Gender 2020

Retirement Percentage by State

Survivor Percentage by State
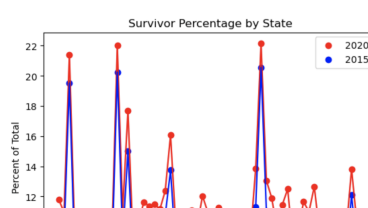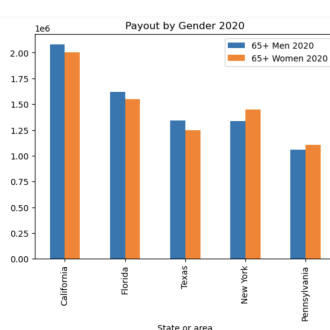
However, as we have seen in other parts of our project, there was a decrease in payouts in 2015 than in 2020.

Conclusion and Implications of Findings

In sum, our data supports that Social Security payout has decreased year over year through the questions we looked into - the overall trend from 2010- 2020, payout by individual, state and within the categories that receive Social Security benefits. We cannot confidently say through our findings that we will run out of funding reserves by 2035 but implications are that we should not have high expectations when it comes to Social Security payouts.

Social Security Administration. OASDI Beneficiaries by State and County, 2021. (n.d.). Retrieved February 2, 2023, from https://www.ssa.gov/policy/docs/statcomps/oasdi_sc/