## Investigating U.S. State Poverty Rates (1995 to 2020)

### Capstone Project Technical Report

By: Financial Services Group 1
Beth Vander Hoek, Jerad Ipsen, Joel Garcia, Shannon Bayless
October 7th, 2022

**Executive Summary**

We researched factors that contribute to poverty across the United States to predict levels of poverty by state and year. We explored the national U.S. poverty level between 1995 and 2020 to see how inflation, federal government spending, and political parties in control of the House, Senate, and Executive office can affect poverty. We then narrowed our focus to state-level for the same time period, investigating whether the following factors affected state poverty rates:

- Drug overdose rate
- Violent crime rate
- Property crime rate
- Unemployment rate
- Party control of state government
- State alcohol revenue
- State tobacco revenue
- State debt
- State property tax
- State education attainment (high school and Bachelor's degree)
- State spending on education
- State spending on social services

Based on our findings and model predictions, we hope to provide recommendations to government officials that can inform improved future policies to reduce poverty rates across the U.S.

**Introduction**

Poverty is prevalent around the world. You can spin a globe and wherever your finger lands, that country will have a percentage of poverty no matter how small. It's not biased with how well a country is developed. In 2021 alone, around 9% of the world's population, which is roughly equal to 698 million people, were found to be living in

extreme poverty.  The effects of COVID over these past 2 years had caused the world to come to a halt as countries were having to deal with containing the virus while having to juggle their own economy as life came to a stand still. At its most basic definition, poverty is a condition where one barely has access to the basic necessities of life. These basic needs are things such as food, water, and shelter.  It's a right that every human needs, but also there's more to it than that.  Things such as education, infrastructure, and access to resources are other needs that need to be met in the modern world.  We thrive off of community and cooperation, but when it comes to poverty, there are people who tend to ignore these things in favor of claiming that as long as the most basic 3 needs are covered, then everything is fine.  However, that is not the case as the world is always changing, and at the rate it's going with trying to get it back up since the start of COVID, there needs to be more work done to ensure that future generations can thrive from access to these resources and cut down on this condition.

When looking at the United States where in 2021 alone, there was a poverty percentage of around 11.6 percent, or 37.9 million US citizens.  However, poverty has always affected the country even before the virus.  Now more than ever, it has become increasingly noticeable that the federal and state governments need better policies to combat this condition that is affecting citizens. With this report, we explore the data we have collected, our data analysis, and machine learning model to answer the following questions:

Federal Level
1. How has the nationwide U.S. poverty level changed between 1995-2020?
2. Is there a relationship between the U.S. national poverty level and: party control of the Senate, House, and President, inflation, or federal spending allocated to social services categories?

State Level
3. What factors influence poverty rates on a state-level?
4. Is state tobacco or alcohol revenue related to state poverty levels?

5. Are state poverty levels and state education spending related?
6. Are state poverty levels and state social services spending related?
7. Are we able to predict poverty levels of U.S. states based on the factors explored in Question 3?

The purpose of these questions is to gain a better understanding of what factors influence poverty levels.  We started at the federal level to gain a broad understanding of poverty on a national scale.  This was to see how and what the government has done in the past to combat this issue while taking into account the result of inflation.  We also looked at party control overall to see how that can potentially influence the decisions that the federal government makes. We then moved onto the state level to see each state's response to poverty through the use of various expenses, taxes, and revenue. To broaden the scope beyond poverty at the state level, we also incorporated drug overdoses, education attainment, unemployment, violent and property crimes.  In order to make the data more consistent from the various datasets we pulled in, we focused on a year range from 1995 to 2020.  By the end of our exploration, we provide some insight on ways that both the federal and state governments can create future policies to reduce poverty levels across the U.S.

**Section I: Federal Level**

In order to see and understand poverty on a state level, we first look at the current state of poverty for the nation as a whole. Again, we are interested in whether national poverty levels have shifted from 1995 to 2020, and how, as well as if national party control, inflation, or allocations of federal spending factor into national poverty rates. To do this we investigated historical data on national party control, the U.S. Census Bureau's reports on national poverty rates, and historical data on the national CPI (Consumer Price Index) for changing inflation rates.
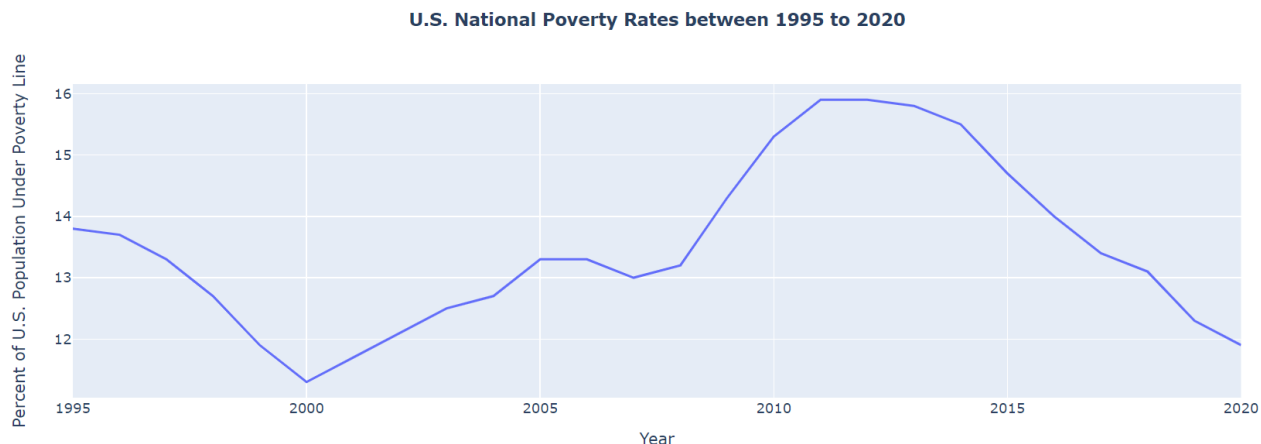
**U.S. National Poverty Rates between 1995 to 2020**



**Fig 1.** US Poverty Rate over time

When examining U.S. poverty rates over the period of 1995 to 2020 from the image above (Fig 1), we saw major dips in poverty in both 2000 and 2020 as well as high points in 2011 and 2012. We were also interested in knowing if government spending on social services or payments to individuals followed a similar trend. After accounting for inflation, we found that federal spending on social security, medicare, and additional payments to individuals steadily rose during the same period with little to no fluctuation at all. The only anomaly that occurred during that period in terms of federal payments was the drastic increase in payments towards individuals during 2020. This was likely due to the introduction of stimulus checks during the initial response to COVID-19, with the first of these payments being issued on March 27th of 2020. We do not see any clear indication that federal spending on social services follow similar trends to poverty levels over this period. Although there was a decrease in poverty rates and increased government spending on individuals in 2020, we cannot say there is any specific link between the two as poverty started to decline before 2020, more around 2014.
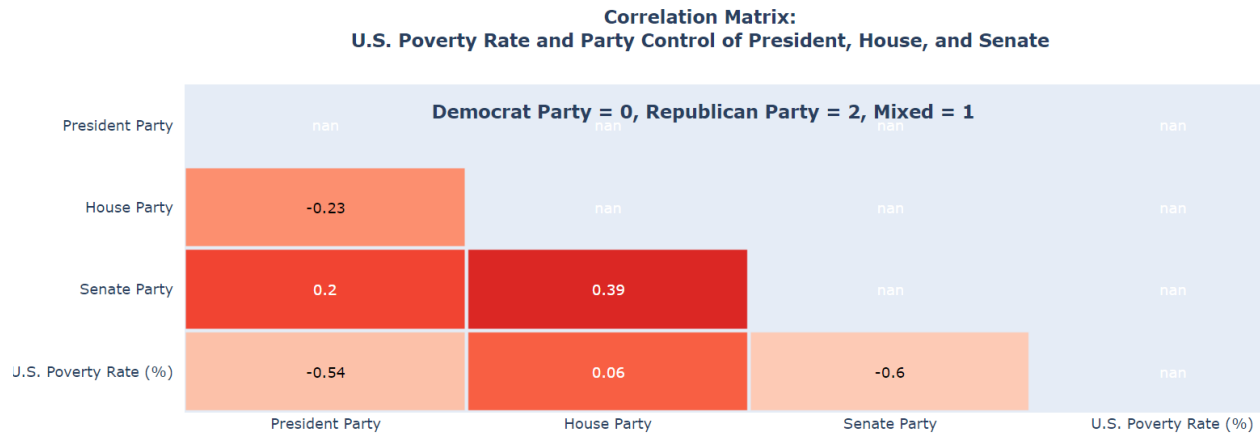
**Fig 2.** Correlation matrix demonstrating how poverty rates are related to the parties controlling the different chambers of congress and the executive branch.

We were also interested in knowing if there is any correlation between federal party control and national poverty rates. To do this we created a numeric scale for party control in which democratic power is coded as 0, republican as 2, and mixed as 1. In the above image there are correlation percentages between poverty percentage and the different branches of power based on party control. Looking at the final row of the image we see there is little to no correlation between party control of the house and poverty percentage. However, we notice there is a fairly high negative correlation between poverty percentage and both senate party control and the president's party, meaning poverty rates increase as party control of the senate and presidency leans more democratic (Fig 2). Although an interesting revelation, one thing to keep in mind is that this correlation matrix does not reflect any correlation between national poverty rates and the effects of policies enacted by prior administrations.

Looking into our potential factors related to poverty rates, we found that on a national level, federal spending on social services and payments for individuals does not follow a direct trend with national poverty rates. We did see a high correlation between national poverty levels and party control of the senate and presidency, but again it is important to note that we did not directly look at correlations between policy in place and poverty rates. The look into party power control as a factor is only an initial one and would need some follow up research on how policies implemented by certain parties affect poverty. Following the federal level, we were interested to see if other

things such as educational attainment, drug overdose percentages, and specific taxes play a role in poverty rates at the state level.

**Section II: State Level**

When examining factors that influence poverty at the state level, we noticed that education levels (both high school and bachelor's) seem to be correlated with poverty rates across all states (Fig 3). However, when narrowing down to a state level, we can see that the factors related to poverty rates change drastically.
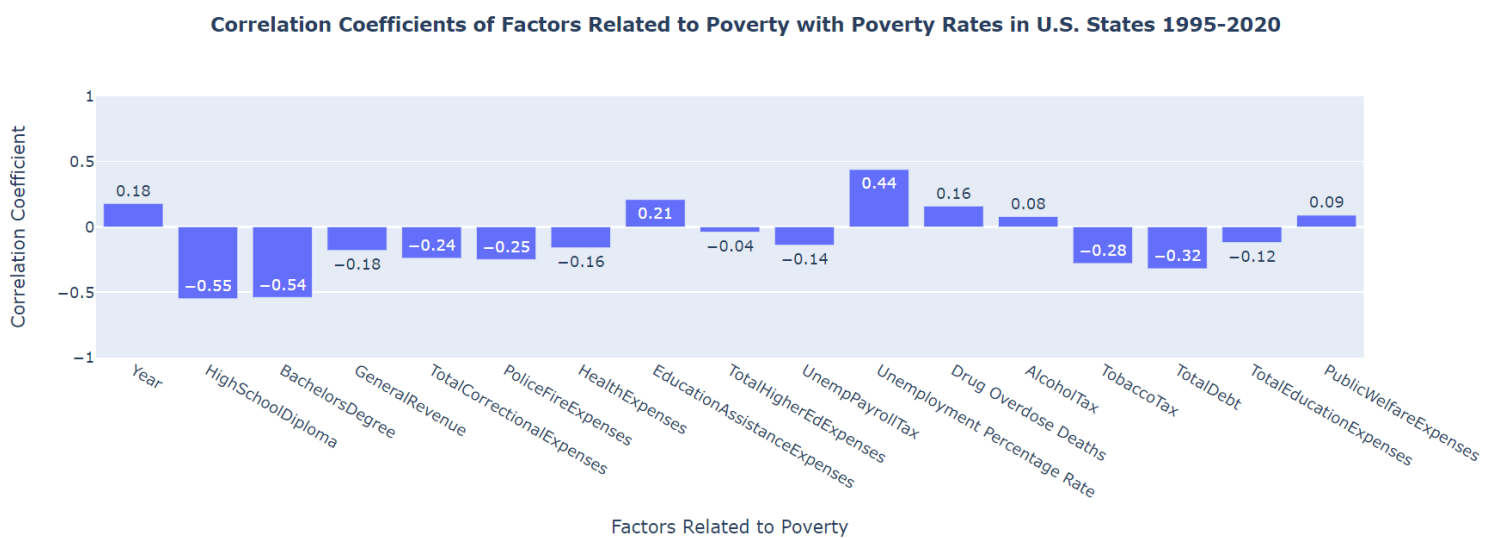
**Correlation Coefficients of Factors Related to Poverty with Poverty Rates in U.S. States 1995-2020**



**Fig 3.** Correlation coefficients for the relationship between poverty rates and other factors.

In New Hampshire, as health expenses increase, poverty rates decrease. These are highly related with a negative correlation coefficient of -0.92 (Fig 4a). On the other hand, poverty rates and health expenses in New Jersey have a positive correlation coefficient of 0.82 (Fig 4b) meaning as health expenses increase, so does poverty.

**Fig 4a (left).** Poverty Percentage vs. Health Expenses in New Hampshire. **Fig 4b (right).** Poverty Percentage vs. Health Expenses in New Jersey.

Other factors that vary from state to state are tax revenue from alcohol and tobacco sales. Continuing with the New Hampshire example, alcohol tax revenue and poverty rates have a negative relationship and a correlation coefficient of -0.68 (Fig 5a) while tobacco tax revenue has a strong positive relationship with poverty rates with correlation coefficient of 0.84 (Fig 5b). This stark difference between alcohol and tobacco tax revenue can also be seen in Colorado, Maine, Ohio, and Wisconsin.





**Fig 5a (left).** Poverty Percentage vs. Alcohol Revenue in New Hampshire. **Fig 5b (right).** Poverty Percentage vs. Tobacco in New Hampshire.

In some states, only tobacco or alcohol revenue has a strong correlation with poverty rates, while the other has almost no correlation. For example, in New Jersey, alcohol revenue and poverty rates have a strong correlation (correlation coefficient = 0.84), while tobacco revenue and poverty rates have almost no relationship (correlation coefficient = 0.03). The opposite is true in Tennessee. Tobacco revenue and poverty rates have a strong correlation (correlation coefficient = 0.79) (Fig 6c), while alcohol revenue and poverty rates have almost no relationship (correlation coefficient = 0.01) (Fig 6d).
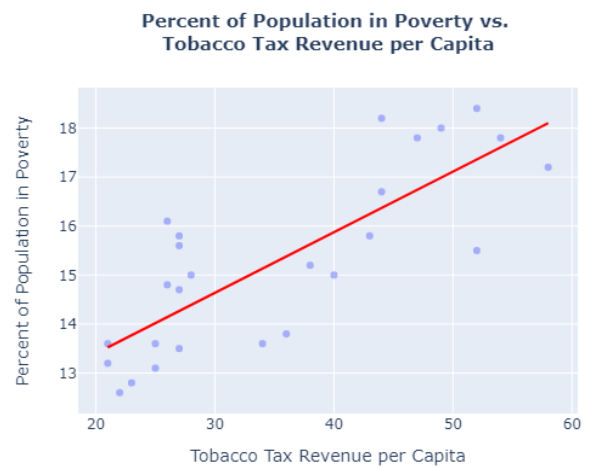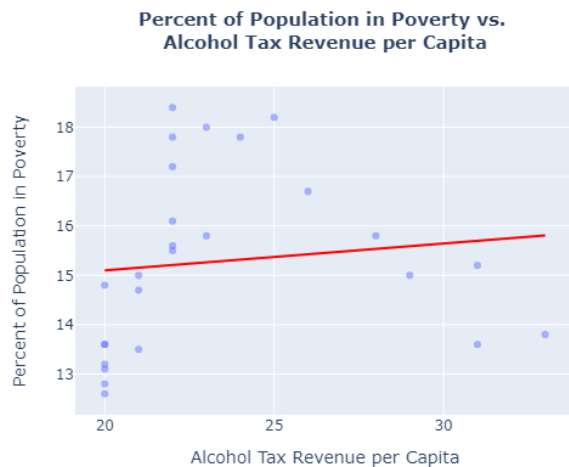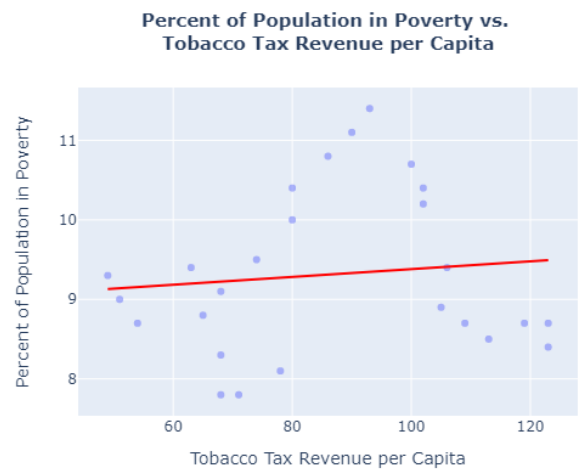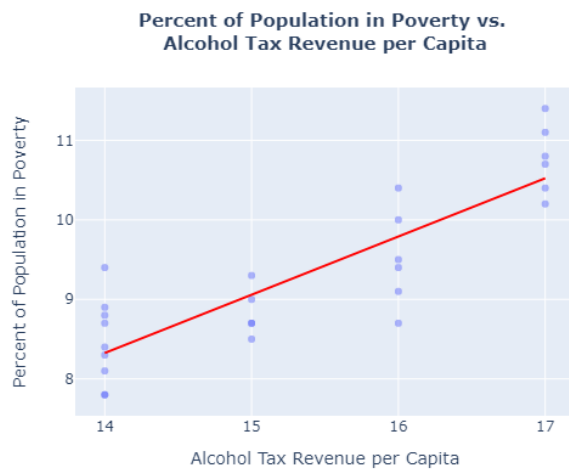
**Fig 6a (top left).** Poverty Percentage vs. Alcohol Revenue in New Jersey. **Fig 6b (to right).** Poverty Percentage vs. Tobacco Revenue in New Jersey. **Fig 6c (bottom left).** Poverty Percentage vs. Alcohol Revenue in Tennessee. **Fig 6d (bottom right).** Poverty Percentage vs. Tobacco in Tennessee.

Total education spending has a surprising relationship with poverty rates. Overall, education expenses and poverty rates have a weak, negative correlation (correlation coefficient = -0.12). However, when looking at individual states, none have a strong negative correlation. Most states have at least a weak positive relationship between poverty rates and education spending. The strongest relationship between these two variables is in Ohio with a correlation coefficient of 0.88 (Fig 7). Other states with strong correlations are Iowa and Delaware.
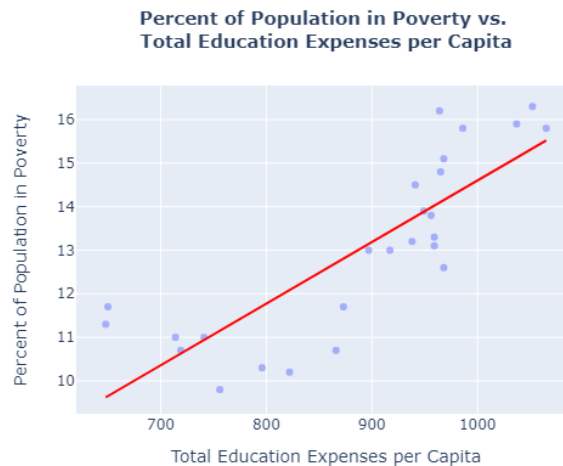


**Fig 7.** Poverty Percentage vs. Education Expenses in Ohio.

Another variable we focused on was welfare expenses. Overall, public welfare expenses have a weak positive relationship with poverty rates. There is a strong positive correlation between these two variables in Delaware, with a correlation coefficient of 0.84 (Fig 8a). There is a strong negative correlation between these two variables in North Dakota, with a correlation coefficient of -0.79 (Fig 8b).

**Fig 8a (left).** Poverty Percentage vs. Public Welfare Expenses in Delaware. **Fig 8b (right).** Poverty Percentage vs. Public Welfare Expenses in North Dakota.

## Section III: Machine Learning Model

We explored how machine learning could be fitted with the data to create an efficient model for predicting poverty percentages. We utilized the following factors related to poverty:

- Year
- State
- Drug overdose rate
- Violent crime rate
- Property crime rate
- Unemployment rate
- Party control of state government
- State general revenue
- State alcohol revenue
- State unemployment revenue
- State debt
- State property tax
- State alcohol tax
- State High School Diploma Rate
- State tobacco tax
- State individual income tax
- State insurance premium tax
- State motor fuel tax
- State public utility tax
- State unemployment payroll tax
- State spending on education
- State spending on social services
- State spending on police/fire services
- State spending on correctional facilities

- State spending on healthcare
- State Bachelor's Degree Rate

To represent the data correctly accounting for all the factors, we considered a LASSO model. The benefit of using this type of model is that it optimizes the data by filtering out the factors that do not have a strong correlation to our target: the state poverty percentages. The way that LASSO (Least Absolute Shrinkage and Selection Operator) works is by minimizing the mean squared error, which is the difference between the estimated values and actual value. This is done by introducing a hyperparameter called alpha, which is a coefficient that penalizes the weights (in this case, the factors) by reducing those unnecessary factors to a value of 0. This method enhances the model's accuracy and interpretability of the results.

To begin, we wanted to fit as much of the data that represented these poverty factors as possible. Our initial results showed that it was necessary to condense the range to 1999-2020 on account of the drug overdose rates: 1995 through 1998 had no values because the criteria for overdoses was updated in 1999. We did not want to impute four years worth of data with the chance that it might skew the results. We also had to convert all of the states and government parties into numerical values (dummy variables), so the model would run correctly. Next, we standardized the data to ensure that the factors were reflected accurately in the results. We then separated our independent variables (the factors listed above) as our X and then the dependent variable (state poverty percentage) as our Y. We tuned the model via a grid search to find the best alpha value, which was 0.005. Finally, we fitted the model using 70% of the data, using the remaining 30% to test the model.
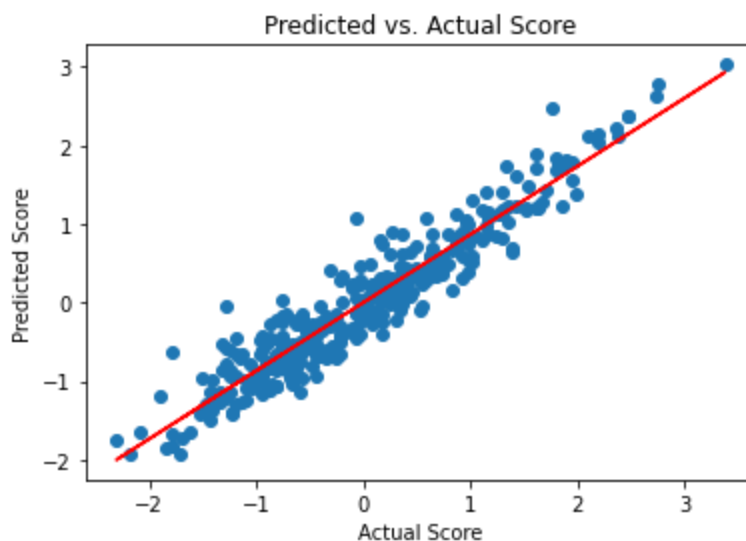
**Fig 9.** Results of Lasso ML Model: Predicted vs Actual Score

The scatter plot above is the result of comparing our predicted values against our actual values. We see that the points in the scatter plot follow in a linear fashion. This is from a visual standpoint, but it indicates that our model was fairly accurate with predicting poverty percentages. However, we also need to consider the $R^2$ value, which was 0.9011, or 90.11%. This can be interpreted to mean that 90% of the variation in our dependent variable, the state poverty percentages, can be explained from our independent variables (factors related to poverty).
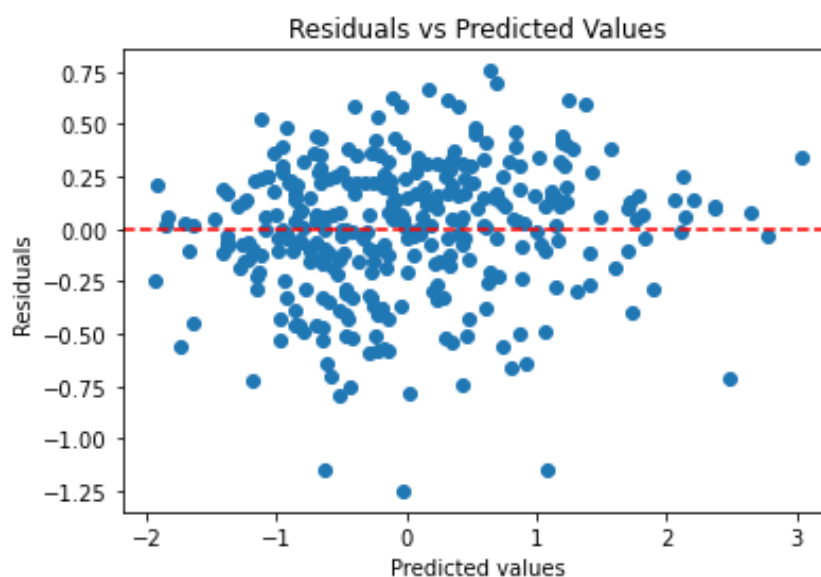
**Fig 10.** Results of LASSO ML model: Residuals vs Predicted Values

Figure 10 compares the predicted values to the residuals, the actual values minus the predicted values.  The red line is the horizontal axis, and you can see points clustering around the y=0 line.  These results are an indicator that our model is working well, and that picking the LASSO model, which is a linear regression model, was a great choice to use for our collected data.

With these results, we also examined which factors the model used for its predictions.  We provided the model 79 factors, which included the dummy variables we made for all 50 states and the government parties.  The results showed that the model used 66 of these variables, including most of the states and government parties.  This was to be expected as all the data we collected stemmed from the states themselves. The model used most of the factors that we listed at the beginning of this section, excluding state debt, drug overdoses, violent crime rates, alcohol revenue, healthcare expenses, and public welfare expenses. The factors that affected the model the most were year, Bachelor's degree attainment, and unemployment percentages.

**Section IV: Conclusion & Recommendations**

Through our data analysis, visualizations, and machine learning model predictions, we were able to address our initial exploratory questions, though more research should be done to investigate them further. On a federal level, we saw that the U.S. poverty level has varied between 1995 to 2020, falling to 11% in 2000, rising to 16% in 2012, and averaging around 14% overall. This rate does not trend in the same manner as inflation and federal spending on social services in the same time frame, but it seems to negatively correlate with Republican control of the Senate and Presidency (ignoring the effects of policy implemented from other administrations). At the state level, state poverty levels have a moderate negative correlation with High School Diploma and Bachelor's Degree attainment rates and a positive correlation with Unemployment and Drug Overdose death rates. As educational attainment rates rise and unemployment and drug overdose rates fall, we should expect to see a resultant fall in state poverty rates.

To our surprise, state tobacco revenue had only a slight negative correlation while state alcohol revenue seemed to have little correlation with state poverty rates. The only significant correlation between state poverty levels and state education spending was for education assistance spending. Additionally, we also found little correlation between state poverty levels and state spending on social services of public welfare and health expenses. Finally, we were successfully able to create and fit a LASSO machine learning model that accounted for 90% of the variation in our data set to predict state poverty levels. Though there may be some overfit, the model relied most heavily on state education attainment levels, year, and state unemployment rates to predict state poverty levels. Our residuals were evenly spread around the y=0 horizontal axis. Though some of our findings contradicted our initial hypotheses, we were able to answer our exploratory questions.

Using our findings, we can advise policymakers on where to focus their efforts to reduce national and state poverty rates. Most importantly, governments should enact policy to promote educational attainment, both at the high school and undergraduate levels; these were the most negatively correlated with state poverty levels. Intuitively, it seems logical education is vital to poverty levels as it opens the door to higher paying wages and stable living situations. Governments should provide greater funding for schools and create a better curriculum that keeps students engaged and motivated to complete their studies. Governments should also specifically target low income students to further their education, with scholarships for college and grants for school supplies.

Secondly, policymakers should create or promote more job opportunities for their populace. Unemployment rates positively correlate with poverty levels, so more jobs means reduced poverty. Governments can work with local businesses to allow for greater business growth and thus more jobs. Governments can also fund job training technical programs for highly skilled positions to allow for low-income individuals to gain better employment opportunities. Lastly, policymakers should focus on drug abuse, providing funds for more effective drug addiction facilities, programs, and education. Government programs can provide communities with information on how to treat drug overdoses, how to seek help, and how best to help loved ones through addiction. They can employ local community leaders, schools, and businesses to spread awareness.

By focusing on educational attainment, unemployment, and drug use, policymakers can best help their constituents and lower poverty rates. Our data analysis, visualizations, and machine learning model reveal that these issues are most correlated with poverty rates and should be addressed to reduce U.S. federal and state poverty rates.

**Works Cited**

- History, Art & Archives: United States House of Representatives(2021). "Party Government Since 1857." https://history.house.gov/Institution/Presidents-Coinciding/Party-Government/ Accessed on 09/21/2022.

- Iowa Community Indicators Program, Iowa State University (2021). *Annual Unemployment Rates by State* [Data set]. https://www.icip.iastate.edu/tables/employment/unemployment-states. Accessed 09/21/2022.

- Kaiser Family Foundation (2022). *Drug Overdose Death Rate (per 100,000 population)* [Data set]. https://www.kff.org/other/state-indicator/drug-overdose-death-rate-per-100000-population/?currentTimeframe=20&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D Accessed 09/21/2022.

- National Conference of State Legislatures(2022). *Legislative Partisan Composition Table* [Data set]. https://www.ncsl.org/research/about-state-legislatures/partisan-composition.aspx Accessed 09/21/2022.

- Unified Crime Reporting Statistics (2021). *State Crime: From the CORGIS Dataset Project* [Data set]. https://corgis-edu.github.io/corgis/csv/state_crime/ Accessed 09/20/2022.

- Urban Institute (2020). *State and Local Finance Data: Exploring the Census of Governments* [Data set]. https://state-local-finance-data.taxpolicycenter.org/pages.cfm Accessed 09/19/2022.

- U.S. Bureau of Labor Statistics(2022). "American Consumer Price Index: 1913 to 2022." https://www.rateinflation.com/consumer-price-index/usa-historical-cpi/ Accessed 09/20/2022.

- U.S. Bureau of Labor Statistics(2022). "Unemployment Rates for States, 2019 Annual Averages." https://www.bls.gov/lau/lastrk19.htm Accessed 09/27/2022.

- U.S. Bureau of Labor Statistics(2022). "Unemployment Rates for States, 2020 Annual Averages." https://www.bls.gov/lau/lastrk20.htm Accessed 09/27/2022.

- U.S. Census Bureau(2022). *Educational Attainment Tables* [Data set]. https://www.census.gov/topics/education/educational-attainment/data/tables.html Accessed 09/30/2022

- U.S. Census Bureau (2022). *Poverty Statistics: CPS & SAIPE (Time Series: various years)* [Data set]. https://www.census.gov/data/developers/data-sets/Poverty-Statistics.html Accessed 09/20/2022.

- White House Office of Management and Budget(2021). *Table 14.4–Total Government Expenditures by Major Category of Expenditure: 1948-2021*[Data set]. https://www.whitehouse.gov/omb/budget/historical-tables/ Accessed 09/21/2022.