Introduction

* Student clearly established the aim of the project.
* Student offered a clear roadmap of the report (i.e what is covered in the report).

Problem Statement and Background

The Low Income Home Energy Assistance Program (LIHEAP) has been administered by the Department of Health and Human Services (HHS) since 1981 to provide low-income households with financial assistance to cover their energy bills and weatherization of their homes.

LIHEAP is one of only 21 federally funded block grants as of 2020.[[1]](#footnote-1) The funding structure of a black grant dictates total sum of federal benefits is distributed by using a funding formula that allocates a percentage of the total funds appropriated by Congress to each grantee (U.S. state, territory, or indigenous tribe). States then distribute the funds to eligible households that apply for benefits.

The LIHEAP Formula was inherited from its predecessor program, the Low Income Energy Assistance Program (LIEAP), which only operated for one year in 1980.[[2]](#footnote-2) This “old” formula favored cold-weather states in terms of funding because it only used data from specific years before 1980, so funding percentages were static.

Congress' 1984 reauthorization of LIHEAP dictated that the formula should be updated yearly with recent population and energy data. The exact LIHEAP formula calculation is not published, but there is awareness of the general metrics and data sources used as formula inputs. This project aims to call upon the same data sources used by LIHEAP and predict the percentage allocation to each state, with one caveat.

The 1984 LIHEAP statute outlined two “hold-harmless” measures to attempt a more equitable funding distribution.[[3]](#footnote-3) Firstly, if the total LIHEAP appropriation for a given fiscal year exceeds $1.975 billion, then no state may receive less funding than they were allocated in 1984. Secondly, if the total appropriation exceeds $2.25 billion, then any state that would receive less than 1% under the formula calculation must receive whatever percentage they would have received of a $2.14 billion allocation. These two provisions, which are applied year to year, reduce the percentage for some states and increase it for others.

In my analysis, I have elected to exclude hold-harmless provisions and use the available data to predict the percentage that a state would have received without them. This allows me to more accurately assess the permutation importance of different variables in the prediction of the actual percentages. Hold-harmless provisions, which are triggered primarily by the size of the total allocation, reduce or increase allocations to an arbitrarily level that tells us very little about how the other data affected a state’s allocation in a given year.

Data

* Student outlined where their data came from.
* Student clearly specified:
  + the unit of observation;
  + the outcome of interest and how it is measured;
  + predictor variables of interest (and why they were selected);
  + potential issues in the data (e.g. missingness, coverage, etc.)
* Student articulate the steps they took to wrangle the data.

The Congressional Research Service report names several datasets that I pulled various predictors from. The unit of analysis is State-Year, and the final dataset includes all 50 states and Washington, DC for the years 2006 to 2019. There is no missingness in the data. The data wrangling process was very involved and required different steps for each dataset. I used pandas to manipulate and merge all dataframes from different sources, and the final dataframe used in my analysis had 714 observations and 36 predictors.

Dependent Variable

The dependent variable is the percent of LIHEAP funds allocated to each state in a given year, before hold-harmless measures were applied. Because the actual LIHEAP formula is not available, I received this data directly from the LIHEAP Program, whom I work with as a scholar with the Massive Data Institute. In order to combat right-skewness of the dependent variable, I applied a log transformation. When tuning and comparing my models, I tested models that used both Percent\_Allocation and log(Percent\_Allocation), and the log transformation led to better model performance.

Independent Variables

Temperature Variables

I used state-level data from the National Oceanic and Atmospheric Administration (NOAA) for variables relating to temperature, including Heating Degree Days (HDD), Cooling Degree Days (CDD), and a 30-year average of HDD weighted by state population. I used BeautifulSoup to scrape urls to download the “December” HDD and CDD files for each year, used urllib to download each file, and used pandas to read one as a fixed width file to add the data to a single data frame. I also created “Lag\_HDD” and “Lag\_CDD” variables, which is the state’s HDD or CDD from two years prior. The 30-year weighted HDD

Analysis

* Student described the methods/tools they explored in their project.
* Justified the tools/methods that they used.
* Adequately described what the tools/methods are doing.
* Note: Assume the reader is smart but doesn’t know programming/machine learning well. That is, be crystal clear about what you’re doing and why.

Results

* Student gave a detailed summary of their results.
* Student presented their results clearly and concisely.
* Student used visualizations (and tables) whenever possible/appropriate.
* Student highlighted some clear takeaways (“things learned”) and theoretical implications (“potential hypotheses”) from their analysis.

Discussion

* Student spoke on the “success” of their project (as defined in their proposal).
  + “Did you achieve what you set out to do? If not why?”
* Student articulate how they would expand the analysis if given more time

1. Block Grants: Perspectives and Controversies [↑](#footnote-ref-1)
2. Perl, Libby. The LIHEAP Formula [↑](#footnote-ref-2)
3. Cite LIHEAP Statute [↑](#footnote-ref-3)