**Applied Data Science Term Project Week 4-8**

Paul Galvez

Department of Data Science, Bellevue University

Summer DSC 680 T301: Applied Data Science

Dr. Catherine Williams

10/13/23

**Applied Data Science Term Project Week 1-4**

**Business Problem:**

The federal government created a program called WIC, which aids eligible families in the United States and federally recognized Native American tribes with family planning, healthcare, and food and nutrition assistance. For this project, the business problem being addressed is the WIC program funding in terms of sending resources to the states and tribes that are identified as having higher enrollment of kids in WIC. The project aims to predict the number of kids in the future based on state or tribe, which will funnel the money accordingly and reduce wasteful spending throughout the food and nutrition assistance for WIC families.

**Background/History:**

The federal government designed the WIC program to focus on food and nutritional assistance for families that apply and are identified as high risk for lacking the necessary resources to sustain a healthy lifestyle. Although WIC has numerous branch programs, the food and nutrition program granting money to families in need has one of the highest funding requirements because so many children under the age of five severely lack access to healthy and easily accessible foods. However, because American taxpayers fund the program, there needs to be some level of fiscal oversight to ensure the proper money is spent and allocated to the states and tribes with the most enrolled kids in the program. Therefore, it is in the public’s best interest to have a project where predicting the number of kids based on the previous year’s enrollments will give insight into improving the overall efficiency of WIC and reduce wasteful spending.

**Data Explanation (Data Prep):**

The dataset was pulled from Kaggle:

<https://www.kaggle.com/datasets/jpmiller/publicassistance>

The project focused on the dataset from the children participating CSV file and was uploaded into Juypter for cleaning and data preparation. The data columns read from left to right as the state or tribe and then the dates from the federal government fiscal year of 2013. The fiscal year for the federal government starts in October and ends in September of the following year—the dated fiscal year columns run from October 2012 through September 2013. The rows of the dataset read as the name of the state or tribe. The number of kids enrolled in WIC by state or tribe is the total number, not the average. A whole number allows for much greater accuracy in the prediction model because the number of kids is much closer to addressing the business problem the project aims to solve. However, a whole column at the end of the dataset focuses on the average number of kids. If there is a need to use or reference that number, it is available for possible use on aspects of the project. The columns with the months during the fiscal year need to be renamed for readability because there were sets of zeros following the months that served no purpose. Once the columns were renamed, the data was ready for plotting and visualization. Below is a picture of the dataset ready for the plotting and visualization stage of the project.

A screenshot of a computer

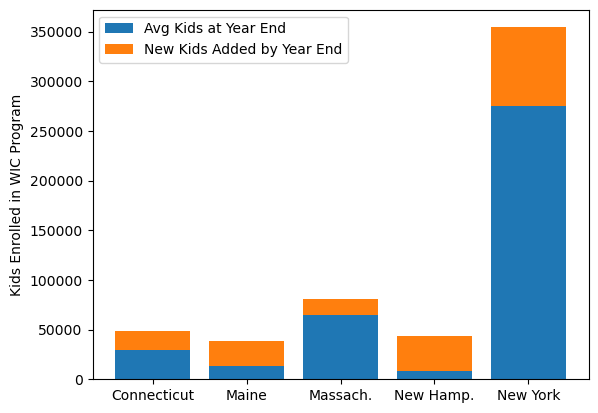
Description automatically generated

**Methods:**

The primary method for predicting the number of kids enrolled in WIC was to use the variables of state or tribe and the total number of kids enrolled in the program throughout the previous year or fiscal year 2013. Regression was the best predictive method because regression shows the independent variable, the state or tribe, with the dependent variable, the number of kids enrolled in FY 2013. The average number of children was used in the visualization of the dataset to show the highest numbers to address the business problem of managing the overall funding for the program. Regression was also an excellent choice for the methodology because it could be applied to other parts of WIC. For example, WIC also covers women who get health coverage during their pregnancies, which can be expensive considering the program covers most costs of the full term for pregnant women in WIC. So, ensuring proper funding is allocated to those most in need is a common theme we can apply to all parts of WIC.

**Analysis:**

The project yielded some exciting results that can be shown through the data visualizations. The average number of kids per state from some of the highest enrolled kids in those locations was visualized for comparison and planning. When we look at the highest average number of kids in the program in states and tribes with the highest number of enrollees versus the total number of new kids added to the program in FY 2013, New Hampshire and Maine saw the highest growth in terms of total number of new kids added to the program versus their average number of kids supported by WIC in those locations. The stacked bar chart below shows the project's findings for the averages and total number of new kids added for states with the highest number of added and averages.



For the linear regression model used to address the business problem, the project yielded some valuable information for funding and planning allocations for WIC. The model predicted that the number of kids added to the program versus the average number for the highest number of kids from the visual above was ninety-nine percent accurate. Moreover, the project’s findings can be applied to those states. The model can be applied to other states and other tribes within the dataset and to other sets of data throughout the program. For example, suppose there was a need to address predicting the cost of food purchased through WIC by enrollees. In that case, we can apply the model to that data to predict what those costs would look like by state, which is critical because the cost of food can change depending on location.

A screenshot of a computer code

Description automatically generated

**Conclusion:**

The project's findings suggest that the program is best allocated and planned for by the state by looking at that information from a future planning perspective from year to year. We know that the program covers kids from newborn to five years old, and the number of those kids can fluctuate every year depending on births in that state or tribe. However, the highest averages of kids throughout the years of WIC in those locations can be a reliable source of information regarding funding and planning. The overall recommendation from the project is to take the information and use that as a planning mechanism from year to year and not use the model as a planning tool for anything beyond the next fiscal year because of the rapidly changing number of kids enrolled in the program. Furthermore, the federal government plans for public assistance programs from year to year and often moves money around with little oversight. The project gives lawmakers and planners insight into forecasting and allows them to be more transparent about where WIC money goes and to whom those funds are supposed to support. In this case, the number of kids enrolled in the program for the neediest states and tribes.

**Assumptions:**

One of the fundamental assumptions about the dataset is that the number of kids enrolled in the WIC program is accurate by the state or tribe. The assumption of accuracy can call into question the need for such a data project because audiences will doubt the data, especially data from the federal government, which some people will never trust because of their lack of trust in the government. Also, the dataset is assumed to be completed yearly by the same collection method. If there is a change in the method or collection process, that adjustment should be considered should the model from this project be used for the business problem addressed, which is where the funding should be deployed based on the state’s needs.

**Limitations:**

The limitations placed on the findings are from year to year and cannot extend into other fiscal years. If the federal government planned WIC funding for multiple years, the model’s findings and project results could be applied to better allocate those monies to the appropriate states. Limitations placed on planning for the program also impact the ability of the project to predict other aspects of WIC, such as women’s pregnancy benefits and subsidizing the cost of food for kids and families enrolled in the program. Other project limitations include how the other program parts impact the business problem addressed in the findings and the model. The federal government could also adjust the amount of funds at the beginning of the fiscal year, which could limit the ability of the model and project’s findings to predict future kids in the program.

**Challenges:**

One key challenge was to link the findings to needs that translate into actions into a complex issue and program that the federal government runs. Historically, the government writes laws and regulations into everything within their preview, especially when that public assistance program is taxpayer-funded. Where tax dollars are being deployed, federal laws and regulations follow, creating red tape, difficulties, and complexities for even the most superficial changes in the running of the program. Persuading an audience of lawmakers and politicians to adapt the model could be a key challenge because of the amount of communication required for its implementation.

**Future Uses/Additional Applications:**

In the future, the model and predictions can be applied to other parts of the WIC program to understand better and manage the needs of the kids and families enrolled in the program and how growth or reduction amongst those most in need of the program’s many benefits. Again, the project was focused on the number of kids enrolled from newborn to five years old who need food and nutrition assistance that the federal government subsidizes as a way for those families to get the proper food into their households and eventually move those families into a more independent state where they can provide for themselves preventing the program from becoming burdensome on American taxpayers. The model can be applied to the various other branches of the program and can help with applications such as predicting the cost of food purchased for kids in the program and the number of women who used WIC during their pregnancy to supplement their lack of health insurance or those women who are underinsured.

**Recommendations:**

A couple of recommendations from the project should be implemented moving forward. First, the federal government should implement the model when planning for future fiscal year allocations regarding the WIC program. Because taxpayers are funding the program, the money being spent on public assistance will be examined and scrutinized at a much higher level than typical government programs. What most people are looking for from lawmakers is transparency and accountability. So, WIC needs a reliable model or plan that shows where the money is going and to whom the money is going in terms of families and kids. Second, the WIC program would benefit from predictive analytics because forecasting plays a pivotal role in any fiscal planning, and being as accurate as possible with taxpayer money helps create efficiencies in spending and reduces unnecessary or wasteful expenditures.

**Implementation Plan:**

The model should be implemented during the planning phases for the fiscal years, immediately followed by the data being released to the public for transparency. At the beginning of the fiscal year, the model can show how each state or tribe can and should benefit from funding allocations. With that in mind, the federal government should implement the model in other aspects of the program, not just the number of kids enrolled from state to state, which is the business problem addressed in the project. During the planning phase of the WIC program for the next fiscal year, the model can be implemented and scaled for whatever predictions those planners and allocators are accounting for or need to justify funding requirements to their audience.

**Ethical Assessment:**

Ethically, the predictions made by the project may only partially justify moving funding and money from one state to another based solely on the findings. For example, there are always going to be families and kids in need of government assistance, and the project does not consider the human aspect of the data and the more significant question of who needs more based on location alone may not necessarily be the best way to manage such an extensive program with so many branches and benefits. Also, the ethical considerations of not over or under-funding based solely on the project’s predictions should be a factor in determining the number of financial help states get from WIC. However, other variables such as income levels in the state and overall cost of living from state to state can be considered, as well as the number of kids predicted to be in the program.

**Ten Questions for the Project:**

1. How difficult is implementing the model into the WIC program?
2. How are the other aspects of the program impacted by the model and project?
3. Are there impacts to the funding based solely on state needs and the number of enrolled kids? Or are there other factors that impact the model?
4. Do the funding allocations for those states most need to be adjusted from year to year?
5. How does the federal government address the overall needs of WIC based on predictions alone?
6. Why does WIC only have funding from year to year and not planned for more in advance?
7. Are there other possible ways for the program to better managed without the moving of money or the removal of families from the program?
8. What can we do with the data to better improve the chances of these needy families moving from public assistance to independent living?
9. Would moving fund from one aspect of the program to another impact the kids who most in need?
10. How does the data from previous fiscal years impact the outcomes of the project and program?

**References**

*US Public Food Assistance 1 - WIC*. (n.d.). Retrieved October 14, 2023, from https://www.kaggle.com/datasets/jpmiller/publicassistance

Appendix A

Subject of the Appendix

The appendix shows the visuals used during the project for the purpose of laying out the predictive analytics for the dataset. The images below show the dataset as a completed data project that can be used for an audience to justify implementation and future use of the findings.

A colorful circle with text

Description automatically generated

A graph with blue rectangles

Description automatically generatedA screenshot of a computer

Description automatically generated