

**School Proficiency and the Impacts of Outside Factors**

Kelsey Puckett

A Thesis Presented in Partial Fulfillment of

The Requirements for the Degree

Bachelor of Data Analytics and Programming

CIS480

Dr. Kazuaki Kashihara

Submitted November 2025

## Thesis Proposal

Receiving an education is a key factor to a successful life. This all starts at a young age when a child goes to kindergarten. Here they learn the foundations of basic skills of which they will continue to build upon as they grow older. Luckily, there are legal measures put into place to help provide every child with the opportunity to learn. On the other hand, not every child's education is created equally. Unfortunately, factors outside of a student's control can impact their education. This project will look to answer the following question:

What external factors impact the proficiency grade of schools in the greater Phoenix area?

This analysis will be looking at the proficiency scores of students based on grades collected from the schools. The goal of this project is to uncover different patterns that align with school proficiency scores which can then be evaluated for correlation.

Growing up in Arizona, I have not only experienced the schooling available here but also seen the differences in learning from some of my peers at different schools. Playing sports growing up allowed me to connect with girls my age all over the valley. We spent time together both on the court and off, often doing homework or other school related tasks together. Although we would be learning the same topics, different peers would comprehend them differently. This led me to wonder what factors can impact success that a student does not have control of?

This project relates to the coursework for the data analytics and programming degree in two ways. First, this project relates to this degree through the ways data will be used and analyzed. In order to understand the different factors impacting a school's proficiency grade, various datasets will be cleaned, analyzed, made into visuals, and modeled. Different modeling

tools that have been taught throughout the duration of this degree may also be used to help outline patterns or trends. In addition to methods for analysis, different thought processes taught during this degree will also be applied. Second, this question relates to this degree in the scope of education in Arizona. Although this analysis will not be including college level education as part of the data, the schools and areas focused on in this project feed into the Maricopa Community Colleges.

In the real world, the significance of a strong and stable learning environment for students is crucial. Schools are shaping the future members of our society. Every student will one day be a part of the adult world and how they can contribute to it matters. The real-world application of this analysis can help outline where improvements to our society can be made and how to make them.

The primary objective of this project is to understand what outside factors impact a school's proficiency. Some potential areas that will be looked into are the median housing price, the amount of access to computers and laptops outside of school, income and poverty levels, school funding, and food access. Additional factors may be added during the analysis too. These areas were chosen to be looked into because of the impacts they can have on a student while at the same time being completely out of the student's control. In an article published by the Journal of Educational Psychology, it is stated that "Demographic characteristics had small but significant effects on initial score and growth rates [in students]" (Scammacca et al., 2020). The factors selected for this evaluation could have the impacts that the demographic characteristics from the article do too. As mentioned above, the primary research question is: What external factors impact the proficiency grade of schools in the greater Phoenix area? In addition to this, the following supplemental questions will also be asked:

Is there any correlation between the housing value and school proficiency? Does access to technology show any impacts? How does food availability in the area correlate to proficiency?

These supplemental questions are important because they help fine tune the goal of the project. The last question is important because it shifts the viewpoint of the project to a broader, real-world perspective.

Modeling will be an important part of this analysis. For the most part, regression models will be used because most of the variables are continuous. For laptop access and food access, those will be binary, so a classification model will be used to evaluate correlation. The project will include a model for each variable tested. Currently, there will be five models, but more may be added if necessary.

PolicyMap will be the primary data source for this project. PolicyMap provides a large variety of data which can be visualized using a map and different points or gradients. PolicyMap offers data on median housing prices, laptop access, school funding, and food access. All of these factors can be analyzed by location.

The primary dataset that will be used for evaluating student proficiency scores is from the Arizona Department of Education. This is a credible source with datasets that can be downloaded through their website.

For this project, the geographic level being analyzed will be the city of Phoenix with boundaries set by zip code. By setting these boundaries, clusters will be analyzed in a small yet meaningful groupings.

The next steps of this project will focus on narrowing down data sources, cleaning the data, and analyzing it. PolicyMap provides a large amount of data ranging in locations and dates

acquired. Narrowing down data sources will consist of researching how when the data was collected, looking into how accurate and complete the dataset is, and deciding whether the data within it is applicable to this analysis. Cleaning the data will involve removing any empty data values and evaluating outliers. For analyzing, different methods such as training models and creating visuals from them will be used to evaluate the correlation.

The timeline for this analysis will be 2024, based on the most recently reported school year. This timeframe was chosen because it will most accurately reflect the current impacts students are facing. It will be important for this project to use data that is relevant to the school proficiency scores. If the timelines for them do not align, they cannot be compared.

The scope of this analysis is external factors only. A future analysis could be done on internal factors such as school spending, student to teacher ratio, extracurricular activates offered by the school, and student involvement in school activities. Although it is interesting to see how those factors correlate to school proficiency, they are not within the scope for this analysis. It is important to stay within the scope of the project because adding additional layers like internal factors can blur the lines for correlation.

In addition to the external factors mentioned above, additional ones may be added to the report depending on the findings of the current ones. For example, looking at home ownership versus rentals and how that correlates to school proficiency could add depth to the analysis. Any additional fields added will be related to the ones already defined and the purpose of adding them will be to deepen the relationship of those categories to school proficiency.

At the end of this project, the goal is to have successfully completed an analysis that can point out which factors correlate to school proficiency levels, and which do not. To support the points made, visuals showing the comparisons will be used.

## Hypothesis and Literature Review

While in school, children can be affected by many factors outside of the classroom. Understanding these external factors that make an impact on a child's education is extremely important because it allows for changes to be made that can help a child succeed. This research paper will be focusing on the following topic:

What external factors impact the proficiency grade of schools in the greater Phoenix area?

Within this project, various areas will be explored such as housing, income levels, school funding, and the level of accessibility to food. Each of these factors has been chosen because of how impactful they can be to a child's life while at the same time the child has no control over them. As mentioned in the topic, this project will focus on the Phoenix area.

The first and primary data source for this paper will be PolicyMap. Policy map is a data source that allows the user to overlay various pieces of data within a specified geographical location. Data for the factors chosen for this project can be found within this resource. PolicyMap also provides many different levels and aspects to their data. For example, there are many different characteristics of housing that data is available for from this site. PolicyMap also supports grouping the data by a specified region or level. It can be grouped by zip code, city, county, state, and other groupings. For this project, data from the Arizona department of Education on school proficiency levels will be compared to the outside factors. The other data within PolicyMap was also sourced from reliable sources.

Within the last ten years, many studies have been done to help gain a better understanding of what impacts proficiency levels within schools. In 2021, a peer reviewed

academic journal published by the National Library of Medicine looked into different factors like ethnicity and socioeconomic status compared to school scores for students. As stated in the abstract for this journal, “achievement gaps remain a problem in the America’s education system, especially those between students from higher and lower income families,” (Scammaca et al., 2020). Within the study done, it was found that there were connections to lower incomes and socioeconomic statuses.

Another study done by Oklahoma State University looked into how having at home technology and internet impacted students ranging from 3<sup>rd</sup> to 11<sup>th</sup> grade. This study revealed that as children progress in their education, the impacts of not having access to technology become more severe due to the increase of complexity of their schoolwork (Whitacre et al., 2020). This trend is outlined in the fact sheet through graphs that show the difference in a child’s proficiency and their level of access to at home technology. In younger levels of schooling, the difference for having access to technology was slight, but as the grade levels increased, so did the proficiency levels.

Focusing on food security, a peer-reviewed research report on household food insecurity and academic outcomes looked into the impacts made by that. In this paper, new research was conducted in addition to previous studies to look at the links between the two. Through the study, it was found that “Children and adolescents who experience food insecurity are at risk for poorer academic outcomes,” and that food security has decreased since the study in result of the Covid-19 pandemic (Orihuela et al., 2023). The children within the study found that food security levels at home were negatively impacted in an academic setting.

The research done by the Journal of School Health, the Journal of Educational Psychology, and Oklahoma State University reveal the trend of how outside factors do impact a

child's proficiency levels within school. This trend can be seen when comparing proficiency to socioeconomic status, technology access, and food security. These studies have also shown that the impact to the student often follows them throughout their entire education. On the other hand, similar gaps in the research were shown throughout the different studies. The main gap is the lack of time the studies were conducted for. The research done by Scamacca had the longest time frame, which was only 24 months (Scamacca et al., 2020). Although trends can be seen through short term research and studies by having multiple checkpoints throughout the timeframe, having a longer study for the different areas could help provide a deeper understanding of these impacts.

These findings are significant to this project because they directly relate to the question being asked. The first journal mentioned, *Examining factors affecting reading and math growth and achievement gaps in grades 1-5: A cohort-sequential longitudinal approach* is connected to both the housing and income factors this project will be evaluating. Oklahoma State University's fact sheet on data gathered on technology and internet access will help support the part of this project looking into how those areas impact student proficiency. Lastly, the research report on food insecurity provides a solid foundation for researching the connection between food levels and school proficiency.

In addition to the primary research question stated above, supplemental research questions are:

Does median housing value correlate to school proficiency scores? Does technology access correlate to school proficiency? Does food security correlate to school proficiency?

These research questions have been chosen to deepen the connection of the outside factors to school proficiency scores.

The null hypothesis and alternative hypothesis for the primary research question are:

**Null Hypothesis (H<sub>0</sub>)**: Areas with poorer living conditions do not experience impacts to proficiency scores in the schools.

**Alternative Hypothesis (H<sub>1</sub>)**: Areas with poorer living conditions experience decreased proficiency scores in the schools.

These hypotheses are relevant because they look at the outside factors as a whole and compares them to the school proficiency score. This test will be one tailed because it is only focusing on lower values for living conditions compared to the rest

For the first supplemental question, focusing on median housing prices, the hypotheses are:

**Null Hypothesis (H<sub>0</sub>)**: There is no relationship between median housing value and school proficiency scores.

**Alternative Hypothesis (H<sub>1</sub>)**: Schools located in areas with above average median housing prices experience increased proficiency scores in schools.

The relevance in this hypothesis is that housing prices are related to socioeconomic statuses which can impact proficiency scores due to resource availability. This test will be two tailed because it is going to be looking at both below and above average prices. The next supplemental question is looking into how technology and internet availability impact the proficiency scores.

The hypotheses for that question are:

**Null Hypothesis (H<sub>0</sub>):** There is no relationship between technology access and school proficiency scores.

**Alternative Hypothesis (H<sub>1</sub>):** Areas experiencing lower technology access will have lower school proficiency scores.

Technology access outside of school is relevant because having access to those tools can impact a student's understanding greatly. This test will be one tailed because it will be looking at how limited to no access compares to having regular access. The third and final supplemental question is looking into the impacts of food insecurity, and the hypotheses are:

**Null Hypothesis (H<sub>0</sub>):** There is no relationship between food security and school proficiency scores.

**Alternative Hypothesis (H<sub>1</sub>):** Food insecurity decreases school proficiency scores.

This research question is relevant because of the intense mental impact food insecurity can have which can impact other areas of one's life. This test will be one tailed because it will be looking only at the lower values for food security levels and comparing that to the rest of the group.

The resources and literature chosen for this project will help deepen the connections between the independent and dependent variables. Each of the independent variables are important because of how impactful they can be to one's life. The different hypotheses formed will also aid in guiding this project by outlining what to focus of it is.

## Data Collection and Modeling

When performing a data analysis, there are many important aspects of it. There are the different hypotheses, modeling tactics, and creating a meaningful conclusion that can relay what was found are all essential, but they all rely on one thing: strong data sources. Within a data analysis, having data sources that have accurate and current data helps ensure the results are a good representation of what is being tested.

This research project is looking to answer the following question:

What external factors impact the proficiency grade of schools in the greater Phoenix area?

The supplemental questions for this project are:

Does median housing value correlate to school proficiency scores? Does technology access correlate to school proficiency? Does food security correlate to school proficiency?

In order to fully address all of the questions in the project, the data sources chosen must share data that represents the different factors within a specified location and time frame.

The primary data source for this project will be PolicyMap. PolicyMap is a tool that allows the user to visualize data points for various categories and set specific boundaries and filters to help outline what trends are happening (PolicyMap, n.d.). This tool will be used in many features of this project.

First, there is a dataset provided by the Arizona Department of Education that shows school performance, or proficiency scores. This will be used to view the scores in the Phoenix

area. These scores are collected for various different categories. For this research project, scores for mathematics and English language arts (ELA) will be used because they are standardized.

The next data source being used will be data from the United States Census through PolicyMap. This will be used to evaluate the variable of median housing value and will be compared to school proficiency scores, which is the first supplemental question. The timeframe for which this data was collected was the years 2019 to 2024. This source is appropriate because of its validity and assistance in answer what the research project is looking into. This data also does not need to be cleaned because if there is insufficient data, it is not used. The geographic scope of this source that will be used for the project is also the greater Phoenix area. Since the Census is a nationwide data source, it has more than enough data points for this project.

Another variable that will be using the United States Census though policy map is technology access. Data on this information is also gathered during the census within the same time frame and geographical location as the median housing value data. It also will not need to be cleaned outside of PolicyMap because insufficient values will not be shown.

The last data source that will be used for this analysis is United States Department of Agriculture, through PolicyMap as well. The data collected from this source will be important to the research project because it can provide insights to what areas within the Phoenix area was experiencing lower access to food. This is important because of the mental impacts not having access to food can have on people, including children. This dataset can be filtered to show low access and not low access on a color-coded map. This data was collected in 2019. Again, like the other source that can be used in PolicyMap, if there is insufficient data then that data will not be used. This is helpful because additional data cleaning will not be necessary.

Another important aspect of any data analysis project is data modeling. Data modeling is important because it can help discover trends and patterns in data and apply them to future scenarios. An accurate model can be extremely beneficial to businesses and companies because it can help them make better, more informed decisions based on previous outcomes.

When using data modeling, there are different models that serve different purposes. Some data models are predictive, and some are analytical. Predictive models are used to help see what future data points may be based on different values. Analytical models are used for discovering trends and patterns in data points that have already occurred. For this project, analytical models will be used. Based on the data being used, multiple linear regression modeling will be a good fit for this project. This model has been selected to potentially be the model because it will fit with the variables well. The independent variables are the median housing value, access to technology, and food access. These values are continuous. The median housing value is in dollars, and access to technology is measured as a percentage, and food insecurity is also a percentage.

As mentioned above, the independent variables will be the median housing value, technology access, and food security. Each of these factors plays a key role in the purpose of this project. First, the median housing value plays a key role because it can help outline the quality of life in the area. Much research has been done on the impacts of housing and schooling, including a peer reviewed academic journal published by the National Library of Medicine titled *Examining factors affecting reading and math growth and achievement gaps in grades 1-5: A cohort-sequential longitudinal approach*. This article ties how socioeconomic differences can impact school proficiency (Scamacca et al., 2020). Because of this journal, and other similar studies, this topic is related to the focus of this project. Second, technology access will be included in my

model due to their heavy use in many schools. As technology has advanced, it has become very common for homework to require some variety of technology access. In addition, as students learn more complex and difficult topics in school, having access to technology outside of school can help one deepen their understanding. Having that ability can help a student seek knowledge to succeed outside of the school setting. At the same time, having access to things like computers and cell phones is almost always outside of a student's control. For those reasons, this variable relates to my outcome of interest. Lastly, food security can play a major role in one's ability to focus and perform tasks well. When one does not have access to food on a regular basis, it can impact their mental and physical health. Although many schools do provide breakfast and lunch programs within the school to help provide lunches for students, food insecurity outside of the school is still impactful. Based on how detrimental it can be to not have enough food this variable also relates largely to the project.

## Data Visuals and Analysis

The focus of this research project is the impacts of outside factors on the proficiency scores in schools. The official research question is the following:

- What external factors impact the proficiency grade of schools in the greater Phoenix area?

Along with that, there are supporting questions that will help deepen the research done. Those questions are:

- Does median housing value correlate to school proficiency scores?
- Does technology access correlate to school proficiency?
- Does food security correlate to school proficiency?

These supporting questions will provide more context to what areas outside of a child's control can be related to decreased proficiency scores.

For visualization within the project, one of the primary visuals used is a geographical map. Within the maps, there are boundaries showing the distinctions between zip codes, color mapping to outline the different factors evaluated, and labels to show the school proficiency levels. There is also a legend with these visuals to show the different color gradient meanings. All of the maps have the zip code set as a dimension, and the other features are set to continuous measure being aggregated by the average.

The second style of visual used for this research project is a scatter plot with a trend line. For these plots, the student proficiency level is plotted on the Y-axis, and the outside factor is plotted on the X-axis. The trend line used is a linear line showing the relationship between the

variables. This visual was chosen because it allows the user to easier comprehend if there is any correlation between the proficiency scores and the specified variable. Having the trend line added also helps make the chart more easily understood. This style of visual is also very compatible with the structure of the data, which is continuous.

Both the geographical maps and scatter plots use the following filters:

- Only use data points within the city of Phoenix
- Only use zip codes within the Phoenix area

The first visual evaluates all three outside factors compared to the school proficiency scores. On the Y-axis is school proficiency, and the X-axis is median housing value. The size of points is the average level of food insecurity, and the color of the points are the percentage of household access to technology. The points plotted represent the zip codes in the Phoenix area. This visual is associated with the hypothesis:

Areas with poorer living conditions experience decreased proficiency scores in the schools.

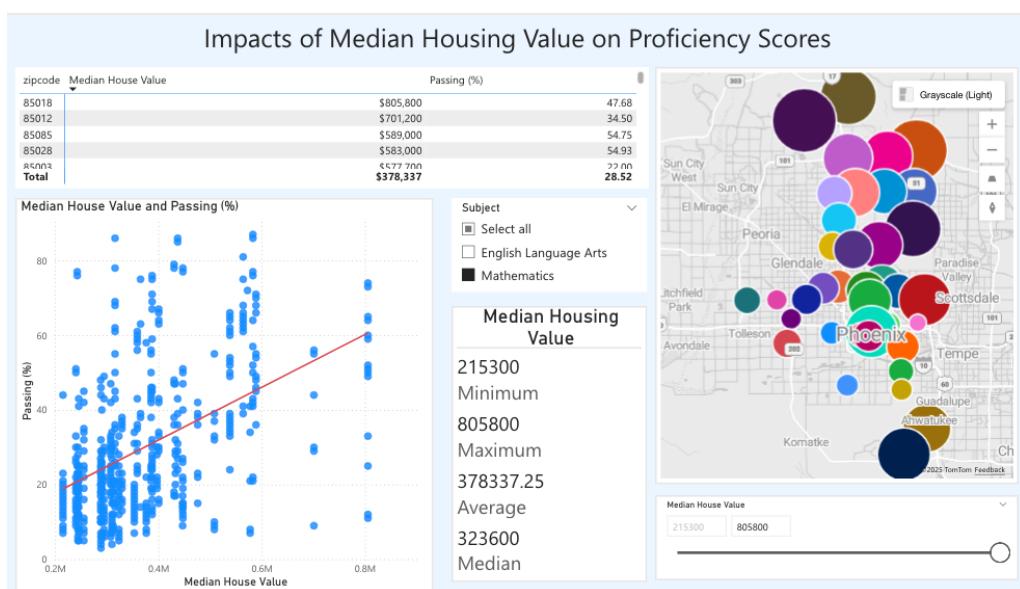
This is an alternative hypothesis that is one tailed. It is only focusing on if poorer living conditions have a negative impact. This visual is appropriate because it helps the user understand how the different factors all play a role with the proficiency levels. It related to my analytics goal because it shows the connection between the different factors when their values change. The structure of the data in this visual has all four values as continuous measures with the average set to the aggregation. A slicer will be incorporated with this visual. This interactivity was chosen because it can help the users analyze the different levels in the plot. The slicer will be set to the household access to technology values.

The next set of visuals both focus on the first supplemental question focusing on the median housing value. Both a map and a scatter plot are used for this. The hypothesis associated with these visuals is:

Schools located in areas with above average median housing prices experience increased proficiency scores in schools.

This is an alternative hypothesis, and it is two tailed. For this question, both lower and higher housing prices will be used to evaluate whether the proficiency scores are impacted.

For the geographical map, the detail is set to zip code, the color is set to the average median housing price, and a tooltip label is used to show the average proficiency value. This visual is appropriate because it helps show the connection between median housing prices and the proficiency levels with a map visualization. This can help the user better understand the different trends occurring in the data. This visual will not include any interactivity because it is only focusing on a small portion of the different factors being evaluated. Because of this, it is already easy to comprehend without additional interactions needed.



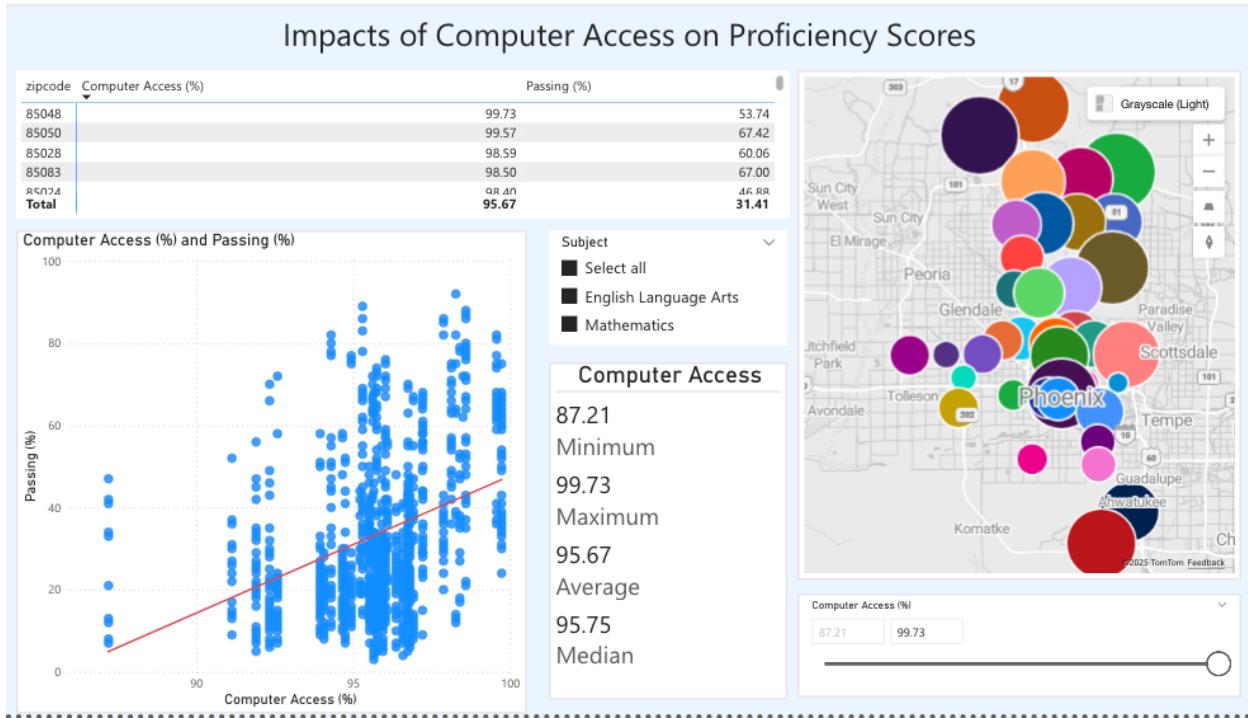
For the scatter plot on median housing value, the Y-axis is the proficiency level, and the X-axis is the average median housing value. There is also a trendline in this visual showing the linear relationship between the plots. This is appropriate because it allows the user to directly see the relationship between the two values. For this visual, both proficiency and median housing value are continuous and aggregated using average. An interactivity included in this visual is a filter that allows the user to filter the assessment type. This will help the user look more closely at the relationship between the value and proficiency.

The second supplemental questions focusing on technology access is visualized again with a geographical map and a scatter plot. These visuals help answer the hypothesis:

Areas experiencing lower technology access will have lower school proficiency scores.

This is an alternative hypothesis that is one tailed.

The geographical map used for this has zip code as the detail measured discretely. The percentage of households that have access to technology is shown as the color and continuously measured, and the proficiency is continuous and shown as a label. Both of these values are aggregated by average. There is a slicer for user interactivity to allow the user to narrow down the values being shown on the map. This style of visual was chosen because it can show the different connections between technology and proficiency is relationship to geographical location.



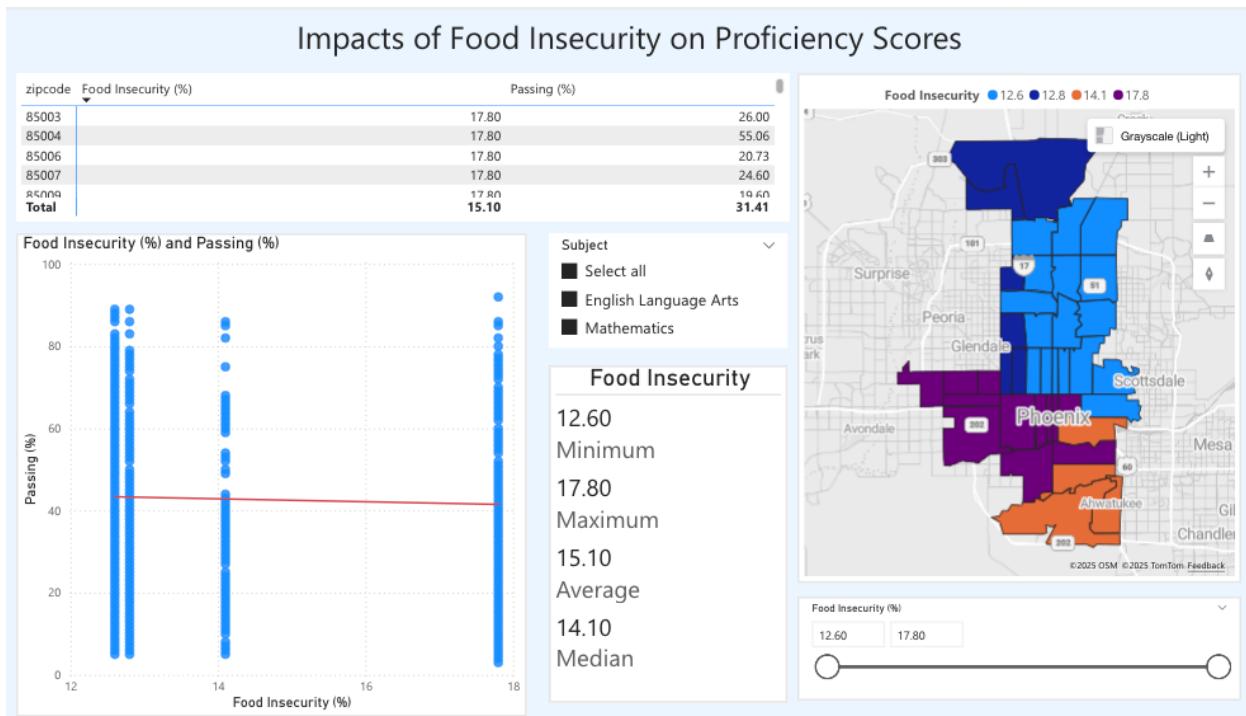
The scatter plot is also used for this portion of the research project because it can easily outline if there are connections or trends occurring between the two values. The Y-axis is set to proficiency, and the X-axis is set to the technology access percentage. Both values are again aggregated using the average and are continuous. Again, a filter will be added to this visual on the assessment type. This will be helpful for the user if they are wanting to get a closer look at a specific subject.

The last set of visuals are used to help deepen the connection between the last supplemental question focusing on food access. The hypothesis for this question is:

Food insecurity decreases school proficiency scores.

This is an alternative hypothesis that is one tailed. Like the other supplemental questions, both a geographical map and scatter plot will be used.

The geographical map is being used because it shows the geographical connection between the outside factor and proficiency. The detail for the map is set to zip code. The color is set to overall food insecurity rate. Like the other two maps, both of these values are continuous and aggregated by average. The user interactivity for this visual is a slicer with the food insecurity rate.



The scatter plot for this factor has the proficiency level on the Y-axis and the food insecurity on the X-axis. Both are continuous and aggregated by average. To help the user narrow down the data within the visual, a filter will be added with the assessment subject.

The different visuals chosen for this research project have been chosen because they can help the user gain a stronger understanding of how the different factors relate to proficiency scores. The different visuals also are compatible with the types of data being used (continuous).

In addition, when appropriate, some of the visuals will incorporate user interactivity to help the user better understand different aspects of the report.

## Hypothesis Summary and Results

This data analytics research project primarily focused on how different factors out of a student's control may impact their learning. The primary research question is:

- What external factors impact the proficiency grade of schools in the greater Phoenix area?

The supplemental research questions are:

- Does median housing value correlate to school proficiency scores?
- Does technology access correlate to school proficiency?
- Does food security correlate to school proficiency?

Each of these questions were given a hypothesis that predicted how student proficiency scores would be impacted by their values.

Starting with the primary question of this project, the hypothesis was:

- Areas with poorer living conditions experience decreased proficiency scores in the schools.

This hypothesis was not rejected. Each of the three factors, which will be looked into more closely when going over the secondary research question hypotheses, had negative impacts to the proficiency scores of students when the factors were on the lower scale. Based on the data models curated during this project, trends lines show that as desirable factors decrease, proficiency does too. Certain factors do have a stronger relationship to the scores, such as median housing value and computer access percentage. Like mentioned above, since each

supporting question directly relates to the main research question and hypothesis, the statistical reasoning to support the hypothesis will be explained with those hypotheses.

The three supporting hypotheses were:

- Schools located in areas with above average median housing prices experience increased proficiency scores in schools.
- Areas experiencing lower technology access will have lower school proficiency scores.
- Food insecurity decreases school proficiency scores.

The first hypothesis is related to the first supplemental question looking at median housing value.

This hypothesis was not rejected. The values from the linear regression model were:

- Coefficient: 6.7914e-5
- Intercept: 5.7817
- R-squared: 0.1790
- RMSE: 17.4451

Based on these values, as housing values increase, the proficiency levels of students also tend to rise. There is a positive relationship between median housing value and proficiency scores, which is shown with a positive coefficient. Based on this, the model supports the hypothesis, and it is not rejected.

The second hypothesis focuses on the second research question looking into technology access. The values from the linear regression were:

- Coefficient: 3.8520
- Intercept: -337.1496

- R-squared: 0.1414
- RMSE: 16.3551

Again, this hypothesis is not rejected. This is because the values from the regression show that there is a positive relationship between computer access and proficiency scores. This means that as one decreases or increases, the other will do the same. The R-squared score shows there is some correlation between computer access and proficiency scores. Because of that, this hypothesis is not rejected.

The last hypothesis looks into food insecurity and the relationship it shares with proficiency scores. The values from the linear regression were:

- Coefficient: -2.2363
- Intercept: 65.2137
- R-squared: 0.1071
- RMSE: 18.1929

This hypothesis is also not rejected. Since the coefficient is negative, it shows that as food insecurity increases, proficiency scores tend to decrease. The R-squared score shows that there is a connection between the values, even though it may be small. Because of this data, the hypothesis is not rejected.

All three supplemental hypothesis were not rejected. All three also support the main hypothesis. Because all were proven, the primary hypothesis is also not rejected.

## Project Summary and Reflection

This project focused on what factors outside of a student's control may impact school proficiency. The factors selected for this project were median housing value, food insecurity levels, and computer access at home. These were chosen based on how each could potentially put a child at a disadvantage while at the same time the child has no ability to control it.

For each of the factors selected, a correlating hypothesis was made. All three hypotheses predicted that as each living condition decreases, the proficiency level will too. After conducting the research project, the key findings show that to be true. As median housing value increases, proficiency tends to as well. When food insecurity levels increase, the inverse occurs for proficiency. When access to computers outside of school increases, so does school proficiency.

A major challenge encountered during this project is the inability to isolate each of the factors when comparing to school proficiency. This is called cofounding variables. While each of the hypotheses tested in this project were not rejected, it is also impossible to say that the factor being tested was the only factor causing the proficiency scores to lower. To help overcome this challenge, looking at the factors together instead of as separate values helped bridge some of the gaps. Doing this produced the best R-squared score, which was .2784.

A major lesson learned throughout this project is the importance how an analytics project is set up and the research done beforehand to ensure the most accurate and reliable tests are being done. For example, instead of just choosing the first data set that could potentially meet the needs of a data project, it can be beneficial to look at several potential sources and check which one is the most accurate for the project. Other important areas to consider are what the research question is truly asking and what methods will be used to answer it.

This research project has helped my development of data analytics by allowing me to go through the entire process of an analytics project. By seeing what work goes into selecting a topic, curating research questions and hypotheses, cleaning and modeling data, and conveying the findings to an audience, I know have more experience that can be applied to future projects in or out of school. From this analytics project, I could build upon it by looking more in depth on different factors. An example of some factors could be access to public transportation, type of homes (apartment, house, etc.), or pre-Covid19 scores compared to current.

## Real-World Implications

In America, schooling impacts everyone. At just age five, children start a thirteen-year journey where they learn valuable skills and lessons they can apply the rest of their lives. From learning basic math and reading skills in elementary school to learning about the history of different nations in high school, the information taught to students helps them live a successful life. Based on how important and impactful schooling is, it is just as important to understand what is limiting a child from reaching their full potential as a student.

Whether attending public school, private school, or even homeschool, all children in the United States of America go to school. Understanding what areas in our society impact a child's education is crucial. If a child is successful in the classroom, it can help them grow their confidence which can spread into other aspects of their life too. On the other hand, if a child feels unsure of themselves in school, it can be harmful. Creating confident and knowledgeable students starts with understanding why certain students perform better than others and how society can help bridge the gap between the two.

This research project focused on three major factors that are potentially impacting proficiency scores at schools. The first factor is median housing value. This value refers to the median monetary value of homes within each zip code in the city of Phoenix. Looking into this value provides practical impact because of how a child's home life can impact other areas. Additionally, if a family is able to afford a more expensive home, they may also have the ability to send their child to private school, pay for extra tutoring outside of school, or pay for different services that may help strengthen their child's understand of concepts covered in school. These same benefits may not be as accessible in zip codes with lower median housing values.

The next factor evaluated in this project is a food insecurity level. Food insecurity refers to when somebody does not have enough food. Not having a sufficient amount of food can impact many areas in one's life, including their mental health and ability to focus. If a child is struggling with food insecurity, negative impacts to their education may occur. This is a practical factor because it is a real issue faced by Americans. This factor is one that can be helped in the classroom through free meal programs, but that does not guarantee that the child is getting the proper nutrition outside of school. Another reason why this has a practical impact is because it is an issue that can be helped through law making and awareness. There are many organizations that help distribute food to those in need. By spreading awareness, more people can donate and in return the struggle with food insecurity can lessen.

The last factor this project looked into is access to computers outside of the classroom. Technology is integrated into almost every aspect of everyday life, including school. Outside of using a computer for schoolwork, having access to one can provide additional benefits for a child. An example of this is if a child was struggling with a concept in school, if they have a computer, they can easily access online resources to help strengthen their understanding. On the other hand, if a child does not have this same access, they might not have the opportunity to do this. This outside factor is practical because of how essential technology is. Access to a computer opens an entirely new world for children when it comes to their education.

As mentioned above, education is an aspect of life that every American is impacted by. Different laws require that all children receive some form of an education to help give them an equal opportunity at a successful future. This research project looks into the real-world problem of why some school proficiency scores are much higher than others. Understanding the root of this issue is the first step in making change. Although this research was only conducting on

median housing values, food insecurity levels, and computer access outside of school, it showed that those factors do make an impact. This analysis addresses how impactful different parts of society and life outside of the school setting can be to a child's education.

There are several everyday societal issues that this project relates to. First and primarily is the topic of education. This project looked at the education proficiency scores for schools within the city of Phoenix. Education is one of the backbones to society. Creating generations of educated people helps society become more successful in many ways. It allows for more opportunities to be taken and for more problems to be solved in the world. Finding the gaps in our education systems is important because it allows for issues to be addressed and resolved. If children are being negatively impacted by their surroundings to a degree of where it is impacting their ability to learn, it can cause many drawbacks for them in the future. If society is aware of the factors limiting a child's ability to learn, they are able to act.

A second everyday societal issue addressed by this research project is the negative impacts of income disparities. The main hypothesis for this project states that poorer living conditions negatively impact proficiency scores. Living conditions directly relate to income. The three outside factors are also all related to income levels. For example, if one has a higher income, they will most likely live in an area with a higher median housing value. Based on the findings in this report, that means that the school proficiency levels within that area will also be higher. If one has higher income, they will not be suffering from food insecurity. They would most likely have access to food and the ability to provide for their family. Lastly, when one has a higher income, they would be able to afford more technology. Again, a higher level of access to computers outside of school correlates to higher proficiency in school, as found in this project. All of these factors can be changed based on income. Of course, having a higher income does

not 100% guarantee that someone is living in an area with an above average median housing price or without food insecurity, but it does lessen the probabilities of poorer conditions.

The work done in this research project helps outline factors impacting schools within the city of Phoenix. Based on the conclusion of the analysis, the following is shown:

- Median housing value has a positive relationship with proficiency levels
- Computer access outside of the school has a positive relationship with proficiency levels
- Food insecurity rates have a negative relationship with proficiency levels

Understanding the relationship between these factors is the foundation of finding a solution. By knowing what areas in society correlate to either excelling proficiency scores or low proficiency scores, action can be taken. Having this information helps answer questions like what preventable circumstances are children experiencing that are negatively impacting them? Outside of the classroom, this research report can help society focus on supporting areas with financial disparities. Although this research project only looked at school proficiency scores within the city of Phoenix, the core of this project can be applied to other research project looking at a broader scale.

When research like what has been done in this project is done, it can make a real impact in our society. A potential next step after completing this research project is to bring awareness to the issues looked into. Students face many obstacles outside of the classroom that they have no control over. Studies like this can help shine light on what those problems are and how they can be minimized or completely resolved. Even though this specific research project only looked into factors within Phoenix, these issues are impacting students everywhere.

A secondary next step after this project is to look into more outside factors and how they impact students. In addition, different geographical locations could be studied to see how different disparities impact different locations. As a society, it is important to remember that current students are the future of our country. Because of that, their education matters.

In conclusion, this report looked at three outside factors focused on what outside factors are negatively impacting student proficiency scores. The three factors are median housing value, food insecurity levels, and computer access outside of school. For each factor, it was hypothesized that when one of the changes for the worse, the student proficiency score will follow. Datasets were selected, cleaned, merged into one, modeled, and turned into visuals within the span of this research project. After conducting the research, all hypotheses were not rejected. This was supported through both the results of linear regression models, and graphical models with trend lines. This research is important to society because of the impacts education make on our country. By highlighting what areas need improvement in our society to then improve a child's education, change can be made. Awareness to issues like food insecurity and limited access to technology can be helped through communities or even on a larger scale, lawmakers. Next steps for this project can help spread information on what students need to succeed. Sometimes it can be easy to look past the struggles faced by students, but it is important to remember that each child can make a difference.

Reference:

Arizona Department of Education (n.d.). *Accountability & Research data*.

<https://www.azed.gov/accountability-research/data>

*Food Access Research Atlas*. Food Access Research Atlas | Economic Research Service. (n.d.).

<https://www.ers.usda.gov/data-products/food-access-research-atlas>

Orihuela, C. A., Cox, C., Evans, R., & Mrug, S. (2023). Associations of household food insecurity with academic outcomes in early adolescents. *Journal of School Health*.

<https://doi.org/10.1111/josh.13358>

PolicyMap. (n.d.). <https://www.policymap-com.ezp.mesacc.edu/newmaps/e/mesacc>

Sciammacca, N., Fall, A.-M., Capin, P., Roberts, G., & Swanson, E. (2020, May). *Examining factors affecting reading and math growth and achievement gaps in grades 1-5: A cohort-sequential longitudinal approach*. Journal of educational psychology.

<https://pmc.ncbi.nlm.nih.gov/articles/PMC7560977/>

Whitacre, B., Sestak, J., & Hughbanks, R. (2020b, August 25). *Do home computers/internet access affect student performance? - Oklahoma State University*. Do Home Computers/Internet Access Affect Student Performance? | Oklahoma State University.

<https://extension.okstate.edu/fact-sheets/do-home-computers-internet-access-affect-student->

[performance.html#:~:text=Figures%201%20and%202%20show%20student%20proficienc](https://extension.okstate.edu/fact-sheets/do-home-computers-internet-access-affect-student-)

y%20scores, and%20is%20consistent%20across%20math%20and%20ELA%20performanc

e.