The Impact of the Covid-19 Pandemic on the Test Scores of Various Demographics Final Paper

As you wrap up your research and development, it is now time to put your work into words and formulate a final paper to document your progress and findings for the rest of the world to see. This template will serve as your general guide to organizing your thoughts and ideas and transforming your work into a comprehensive scientific paper. Your final paper will be composed of 8 main sections: **Abstract, Introduction, Background, Dataset, Methodology/Models, Results and Discussion, Conclusion, Acknowledgements,** and **References.** Below, each section and its respective composition is described (if you're looking for a more detailed description, **check out this guide**!).

1. Abstract

[< 250 words] The abstract should summarize the key aspects of your research in one paragraph, including motivation, methodology, and the most significant findings. Overall, your abstract is meant to be a concise summary of your research process, from the research question to the results. Remember, non-technical minded people might be reading the abstract, and they should be able to largely understand what's going on! *After filling out the questions below, combine all of your answers to compose your abstract!*

Answer the following questions in 1 or 2 sentences:

1. **What is your research problem or question? What are you trying to accomplish?**

How did Covid impact and affect the test scores of different demographics of students in various schools in New York?

2. **What is the background / context of the problem or question? Why is it important?**

By the conclusion of the school year, students had an average of five months' worth of math and four month's worth of reading behind them, according to an analysis of the pandemic's effects on K-12 student learning.[1] For example, student testing in 2021 was about ten points behind in math and nine points behind in reading, compared with matched students in previous years.[2] The pandemic expanded already-existing achievement and opportunity inequities, particularly harming historically underprivileged students. Children in low-income schools had seven months of unfinished arithmetic work at the end of the year, compared to six months for students in majority-Black schools.[3] High school students are now more likely to drop out, and seniors,

[1] Dorn, Emma, et al. "COVID-19 and education: The lingering effects of unfinished learning." *McKinsey*, 27 July 2021, https://www.mckinsey.com/industries/education/our-insights/covid-19-and-education-the-lingering-effects-of-unfinished-learning. Accessed 13 November 2022.

[2] ibid.

[3] ibid.

particularly those from low-income households, are less likely to continue their education after graduation. The crisis also had an effect on pupils' general health and well-being, with more than 35% of parents expressing severe or great concern for their kids' mental health.[4] The effects of the epidemic pose a threat to the hopes and possibilities for this generation far into adulthood. Their chances of enrolling in college and subsequently securing a rewarding profession that allows them to sustain a family may be adversely affected by the knock-on effects. Furthermore, analysis indicates that the pandemic's effects on schooling may result in today's students earning $49,000 to $61,000 less in their lifetime unless measures are implemented to remedy unfinished learning.[5]

3. **What was the overall approach used to answer the question?**

The overall approach that was used to answer this question involved using statistical analysis to examine the correlation between how certain factors such as being economically disadvantaged or being hispanic affected student proficiency scores on New York State regents exam tests in 2018 and 2021 in different schools.

4. **What were the significant results? Numbers should be understandable without reading the paper, so expressing differences as fractions or percentages may help the reader.**

Some of the significant results that came out of this research paper involved certain demographics such as those who were economically disadvantaged and those who were white. Specifically, in relation to how their demographic correlated with their school's percent proficiency difference in a subject exam between the years 2018 and 2021. For example, 100% of the students of the Academy for Language and Technology are economically disadvantaged and compared to 2021, the number of students who were proficient in English dropped by 49%. Furthermore, 72% of Akron High school's students are economically disadvantaged and compared to 2021, the number of students who were proficient in English dropped by 68%. However, on the other side of the spectrum, students of white descent did not fare as badly. For instance, Adirondack Middle School has an approximate demographic makeup of 100% white students. Their student's percent proficiency when it came to their Algebra I score did not change at all. Additionally, Yonkers Montessori Academy has a demographic makeup where 46% of its students are white. The students from this school showed a 5% increase in their proficiencies in their science scores.

5. **What are the major conclusions in relation to the problem or question?**

The research that was conducted as the basis for this paper was done in an effort to examine the COVID-19 pandemic's effects on student performance across various schools in New York. Specifically, it affected some demographics disproportionately when compared to others. Due to students being forced to learn from home, some without the necessary technology and adequate resources. Consequently, a lot of students took a significant hit to their academic progress. For instance, 100% of a student's demographic make up is economically disadvantaged and out of these students, 49% of them were no longer proficient in ELA after the COVID-19 pandemic. Whereas, students of white descent did not fare as badly. For instance, Adirondack Middle School

[4] ibid.
[5] ibid.

has an approximate demographic makeup of 100% white students. Their student's percent proficiency when it came to their Algebra I score did not change at all.

The research problem that this research paper attempts to address is, *How did Covid impact and affect the test scores of different demographics of students in various schools in New York?* This issue is significant for the following reasons. By the conclusion of the school year, students had an average of five months' worth of math and four month's worth of reading behind them, according to an analysis of the pandemic's effects on K-12 student learning.[6] For example, student testing in 2021 was about ten points behind in math and nine points behind in reading, compared with matched students in previous years.[7] The pandemic expanded already-existing achievement and opportunity inequities, particularly harming historically underprivileged students. Children in low-income schools had seven months of unfinished arithmetic work at the end of the year, compared to six months for students in majority-Black schools.[8] High school students are now more likely to drop out, and seniors, particularly those from low-income households, are less likely to continue their education after graduation. The crisis also had an effect on pupils' general health and well-being, with more than 35% of parents expressing severe or great concern for their kids' mental health. [9] The effects of the epidemic pose a threat to the hopes and possibilities for this generation far into adulthood. Their chances of enrolling in college and subsequently securing a rewarding profession that allows them to sustain a family may be adversely affected by the knock-on effects. Furthermore, analysis indicates that the pandemic's effects on schooling may result in today's students earning $49,000 to $61,000 less in their lifetime unless measures are implemented to remedy unfinished learning.[10] The overall approach that was used to answer this question involved using statistical analysis to examine the correlation between how certain factors such as being economically disadvantaged or being hispanic affected student proficiency scores on New York State regents exam tests in 2018 and 2021 in different schools. Some of the significant results that came out of this research paper involved certain demographics such as those who were economically disadvantaged and those who were white. Specifically, in relation to how their demographic correlated with their school's percent proficiency difference in a subject exam between the years 2018 and 2021. For example, 100% of the students of the Academy for Language and Technology are economically disadvantaged and compared to 2021, the number of students who were proficient in English dropped by 49%. Furthermore, 72% of Akron High school's students are economically disadvantaged and compared to 2021, the number of students who were proficient in English dropped by 68%. However, on the other side of the spectrum, students of white descent did not fare as badly. For instance, Adirondack Middle School has an approximate demographic makeup of 100% white students. Their student's percent proficiency when it came to their Algebra I score did not change at all. Additionally, Yonkers Montessori Academy has a demographic makeup where 46% of its students are white. The students from this school showed a 5% increase in their proficiencies in their science scores. The research that was conducted as the basis for this paper was done in an effort to examine the COVID-19 pandemic's effects on student

[6] Dorn, Emma, et al. "COVID-19 and education: The lingering effects of unfinished learning." *McKinsey*, 27 July 2021, https://www.mckinsey.com/industries/education/our-insights/covid-19-and-education-the-lingering-effects-of-unfinished-learning. Accessed 13 November 2022.

[7] ibid.

[8] ibid.

[9] ibid.

[10] ibid.

performance across various schools in New York. Specifically, it affected some demographics disproportionately when compared to others. Due to students being forced to learn from home, some without the necessary technology and adequate resources. Consequently, a lot of students took a significant hit to their academic progress. For instance, 100% of a student's demographic make up is economically disadvantaged and out of these students, 49% of them were no longer proficient in ELA after the COVID-19 pandemic. Whereas, students of white descent did not fare as badly. For instance, Adirondack Middle School has an approximate demographic makeup of 100% white students. Their student's percent proficiency when it came to their Algebra I score did not change at all.

2. Introduction

[~ 0.5 pages long] Clearly discuss your research question, its significance, and any relevant background information needed for the reader to fully understand the problem. Furthermore, highlight the type of problem you are working with (e.g. supervised/unsupervised & regression/classification), the nature of the data you are working with (e.g. numerical/categorical, language data, vision data), and the output of your project (e.g. labels, quantities, etc.).

This research paper addresses the research question of *How did Covid affect and impact the test scores of different demographics of students in various schools of New York?* We as humans quickly became aware of people who have been affected academically by this Covid pandemic. Schools were shut down and many people were forced to work and learn from home. As a result, people no longer had face to face interaction, with some losing vital motivation. As more and more people were sent to learn from their own homes, teachers were not able to check on their students as they might have done before the pandemic. In addition, some student's families do not have equal access to technology due to their circumstances. Consequently, some of these student's academic performances were affected significantly as this paper details. This paper addresses a problem where numerical data comes from the New York State education department. To analyze this data, statistical analysis was carried out and the result of this process was a percentage difference of scores for different demographics of students in the same school, across the years of 2018 and 2021.

3. Background

[~ 0.5 pages long] Describe 1-3 other articles/approaches that have been used to solve your research problem. Briefly highlight pros, cons, or unique contributions of other approaches and how they may relate to your research approach. *Consolidate the work you already completed in your Miniature Literature Review to write this section!*

Not a lot of work has been done in regards to examining the effects of Covid-19 on certain demographics of students with an Artificial Intelligence/Machine Learning approach, especially in New York. However, some analytical analysis has been done, with countries such as China. For instance, the paper, "*Compensating for academic loss: Online learning and student performance during the COVID-19 pandemic*", found that online learning, as opposed to no learning support, had a large positive effect on students' overall subsequent exam performance. But the content of the (recorded) lessons also mattered: those recorded by higher-quality teachers produced better exam results. Moreover, while the academic benefits of online learning were the same for rural and urban students, students who engaged in online learning with a computer profited more than those who used a smartphone. Furthermore, quantile DID (Difference in Differences Estimation) analysis showed that while any online learning benefited all students outside of the top performers, it was the academically-weaker students who benefited the most from online education and were more reactive to the quality of the teachers providing the online lessons. However, there are certain flaws with this data. For example, the data on student academic records and online educational practices comes from only three Middle Schools in one Chinese county. The external validity of these empirical results obviously requires investigation. In addition, the evaluation of nationwide online learning has not been possible due to data availability. Furthermore, area-specific COVID-19 conditions may have affected the educational practices that could be adopted by Chinese schools during the crisis. Had the treatment and control groups come from different areas, the effects of online education on exam results could have been confounded by economic and psychological pressure associated with area-specific COVID-19 outcomes. However, all of the students here come from urban schools in the same county, and were thus exposed to the same environment. The analysis does thus have greater internal validity, but at the cost of lower external validity.

4. Dataset

[~ 1 page long] Describe the dataset that you used for your project. State the type of data you are working with (numerical, language, vision), the number of samples you have, and how you split the data across training and testing datasets. Explain any data preprocessing procedure you carried out and show any visualizations of your dataset or features (if applicable). Be sure to fully describe the features of your dataset and explain the significance of the features with respect to your research problem. *Consolidate the work you already completed in your Dataset Selection to write this section!*

For our project, the dataset that we used to perform the analysis was New York State's yearly update of a "report card database". This Access database contains assessment results (elementary- and intermediate-level ELA, Math, and Science; Annual Regents; Total Cohort Regents; NYSESLAT; NYSAA), for the state, districts, public with charter schools, by county, and Need to Resource Capacity group. Specifically, I used the annual regents exams data which was documented in a numerical representation for each of these schools and their subgroups of students. Taking into account the complexity of the way in which the data is laid out, many data preprocessing steps were carried out. Firstly, all columns that had NaNs were replaced with zeros. In addition, all columns were converted to having an integer value stored in their respective places. Originally, these values were stored as string values and this conversion was done in order to prepare the data for analysis. Furthermore, all rows that contained a random and unexplained for "s" were removed from the dataset. Lastly, I performed one hot encoding on the "SUBGROUP" column, giving me a separate column for each and every one of the demographics provided by New York State documentation.

5. Methodology/Models

[~ 2 pages long] Describe the methods you used to solve your research problem. State the machine learning algorithms you used and briefly explain how each one works. Describe how you carried out your model learning procedure (e.g. split the dataset into train/test, removed some of the features, input the data into an sklearn model, etc.).

To solve the research problem, we mainly focused on a statistical approach. Initially, we attempted to come up with results using machine learning models. Unfortunately however, there were many data entries that were missing for some schools and their demographics of students. This made it difficult to carry out regression models on the data as hoped. We adapted my approach to a statistical outlook that involved the use of pandas and numpy methods. Firstly, both New York State regents exam data were loaded in from 2018 and 2021. Then for both years, we ended up separating the percent of students who were tested and were either economically disadvantaged, hispanic, or african american and stored these in variables. We also created a variable that was representative of the percentage of students who were proficient with their exam-taking. Next, I created a table with 5 columns which included the following: the school name, the subject of the exam taken, the percent proficiency of students of that school who took the exam in 2018 and 2021, the percent of a certain demographic of students who took that test for that student (variable), and the percentage difference for the regents exam scores between the years of

2018 and 2021. Lastly, I created a regplot of the results I came up with. Specifically, I plotted the percentage prevalence of a certain demographic against the percentage points difference in regents exam scores between 2018 and 2021. An example of this can be seen on the right. This ended up producing some very interesting results that will be discussed in the following section.

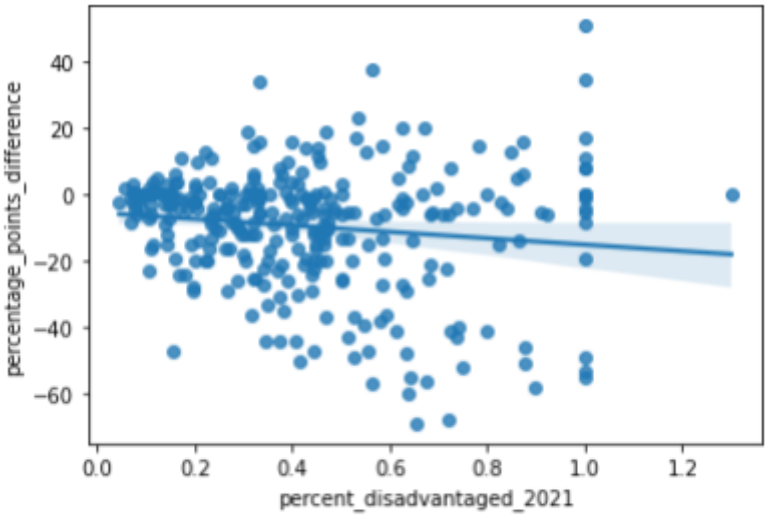

Figure 1
Regplot of Percent Disadvantaged Students (x) and the percentage points difference for that school (y)

6. Results and Discussion

[~ 2 pages long] Highlight your research findings. Describe the outcomes and metrics of your developed models. Include all relevant model metrics and make sure to highlight any hyperparameters you selected or modified for your models. Include visualizations, tables, and figures to depict your results. Clearly describe all figures and their significance. Furthermore, highlight any errors your method makes and discuss why you think your method may perform poorly in some cases. *Consolidate your Model Metrics and Hyperparameters Sheets to write the latter part of this section!*

When creating results to analyze in order to address my original research question of—"*How did Covid impact and affect the test scores of different demographics of students in various schools in New York?*"—I created charts and visualizations to do so. In total, 6 combinations of factors and test scores were plotted against each other. The first comparison that was conducted examined the potential correlation between the percentage of students who were disadvantaged in the year 2021 (x), against the percentage points difference in proficiency between 2018 and 2021 (y). This can be seen to the figure on the right. When one takes a look at this information, it becomes clear that students who are economically disadvantaged have been affected directly in their test scores after COVID. Additionally, a table output was created as can be seen below. This table

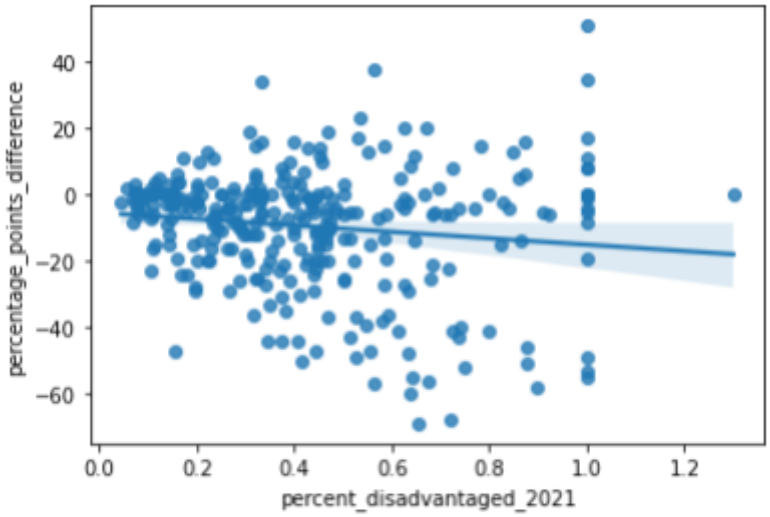

Figure 2.1
Regplot of Percent Economically Disadvantaged Students (x) and the percentage points difference for that school (y)

	ENTITY_NAME	SUBJECT	PER_PROF_2018	PER_PROF_2021	percent_disadvantaged_2021	percentage_points_difference
1	ACADEMY CHARTER SCHOOL	Regents Common Core Algebra I	68	88.0	0.625000	20.0
3	ACADEMY FOR LANGUAGE AND TECHNOLOGY	Regents Common Core English Language Art	75	26.0	1.000000	-49.0
5	ADIRONDACK MIDDLE SCHOOL	Regents Living Environment	100	100.0	0.312500	0.0
6	AKRON HIGH SCHOOL	Regents Common Core Algebra I	96	28.0	0.720000	-68.0
7	AKRON HIGH SCHOOL	Regents Living Environment	91	50.0	0.611111	-41.0
...	...	...	...	...	...	...
467	WINDSOR CENTRAL HIGH SCHOOL	Regents Living Environment	91	44.0	0.555556	-47.0
468	WOODLAND MIDDLE SCHOOL	Regents Common Core Algebra I	100	100.0	0.105263	0.0
469	WOODLAND MIDDLE SCHOOL	Regents Phy Set/Earth Sci	100	84.0	0.108108	-16.0
472	YONKERS MONTESSORI ACADEMY	Regents Living Environment	72	77.0	0.615385	5.0
473	YOUNG WOMENS LEADERSHIP SCHOOL	Regents Common Core Algebra I	76	50.0	0.500000	-26.0

Figure 2.2
(Table with Percentage Points Difference between 2018 and 2021 exam for a school and list of percent of students disadvantaged for that school)

provided an in-depth overview of the relationship between certain schools' demographics of students and their test scores for a particular subject exam for the years of 2018 and 2021. For example, 100% of the students of the Academy for Language and Technology are economically disadvantaged and compared to 2021, the number of students who were proficient in English dropped by 49%. Furthermore, 72% of Akron High school's students are economically disadvantaged and compared to 2021, the number of students who were proficient in

English dropped by 68%. A second comparison that was conducted involved the demographic of African American students. The plot generated above indicates the effects that COVID-19 likely had on these students, showing a relatively steep decline in proficiency scores on tests. Furthermore, this trend of certain demographics being affected due to the COVID-19 pandemic is also seen with students that have disabilities. With the plot generated to the left, it becomes noticeable that just as students who were of African American descent, students with disabilities also suffered a hit to their proficiencies when it comes to test taking after the COVID-19 pandemic.

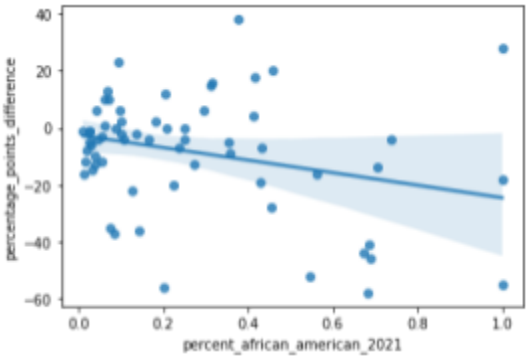

Figure 2.3
Regplot of Percent African American Students (x) and the percentage points difference for that school (y)

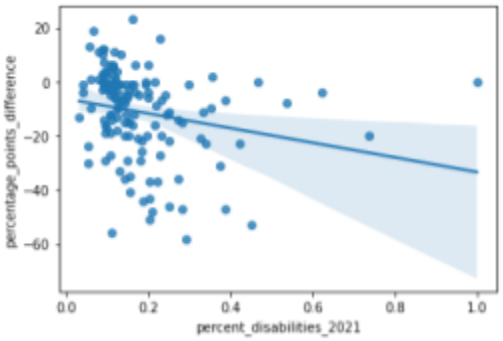

Figure 2.4
Regplot of Percent Students with Disabilities (x) and the percentage points difference for that school (y)

However, not every demographic of students was disproportionately affected by the pandemic as can be seen in the two following comparisons. For example, when the demographic of white students was taken into account with the plot shown on the right, it is clear to the observer that student scores did not differ significantly. Rather, it becomes apparent that the scores were relatively level before and after COVID, showing that the pandemic did not have as much of an impact when compared to other demographics. More specifically, Adirondack Middle School has an approximate demographic makeup of 100% white students. Their student's percent proficiency when it came to their Algebra I score did not change at all. Additionally, Yonkers Montessori Academy has a demographic makeup where 46% of its students are white. The students from this school showed a 5% increase in their proficiencies in their science scores. Furthermore, one final comparison was done that took a look at how economically sound students fared after the COVID-19 pandemic. With the plot shown to the left, the data plotted is clustered towards the left with one significant outlier. Generally, the trend seems to be that students who were not facing economic hardships tended to not suffer a drop in their scores, and in fact some saw an increase in their proficiencies. In summary, as we expected, students who were in economic hardship suffered noticeably as can be seen from large decreases in proficiency scores after the pandemic. This trend seems to have also affected other demographics such as African American students and students with disabilities. This draws the viewer's attention to the many factors that could have played a role in this effect. It sounds the

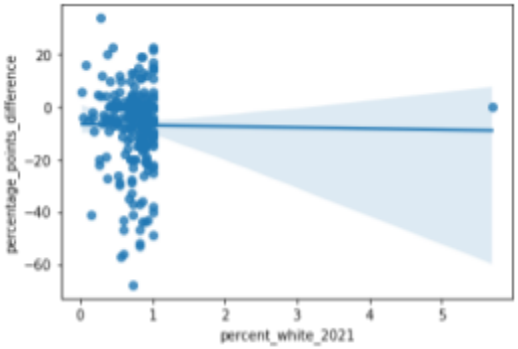

Figure 2.5
Regplot of Percent White Students (x) and the percentage points difference for that school (y)

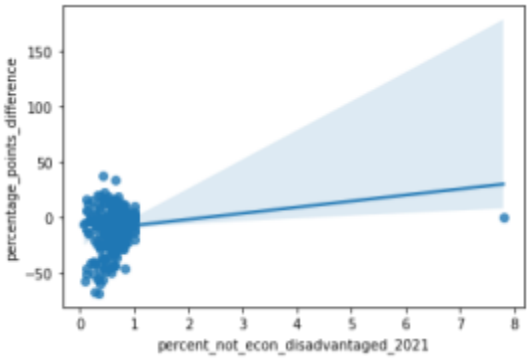

Figure 2.6
Regplot of Percent Not Economically Disadvantaged Students (x) and the percentage points difference for that school (y)

alarm that these students did not have as reliable access to resources, technology, and teaching. This should act as a warning to policymakers and other government officials that students who fall under these demographic categories are struggling and preventative action must be taken to stop any further decline in academic achievement.

7. Conclusion

[~ 0.5 pages long] Summarize your research paper clearly and concisely. Highlight key points regarding your goals, significance, methodology, and results. Describe why you think your model performed as well/poorly as it did. Finally, explain what next steps you would take with your research (e.g. try out different models, obtain more data, preprocess data differently, etc.).

The research that was conducted as the basis for this paper was done in an effort to examine the COVID-19 pandemic's effects on student performance across various schools in New York. Specifically, it affected some demographics disproportionately when compared to others. Due to students being forced to learn from home, some without the necessary technology and adequate resources. Consequently, a lot of students took a significant hit to their academic progress. For instance, 100% of a student's demographic make up is economically disadvantaged and out of these students, 49% of them were no longer proficient in ELA after the COVID-19 pandemic. Whereas, students of white descent did not fare as badly. For instance, Adirondack Middle School has an approximate demographic makeup of 100% white students. Their student's percent proficiency when it came to their Algebra I score did not change at all. To produce these results, a statistical approach was conducted to analyze the appropriate data. Firstly, both New York State regents exam data were loaded in from 2018 and 2021. Then for both years, we ended up separating the percent of students who were tested and the demographic they were categorized as and stored these in variables. We also created a variable that was representative of the percentage of students who were proficient with their exam-taking. Next, we created a table with 5 columns which included the following: the school name, the subject of the exam taken, the percent proficiency of students of that school who took the exam in 2018 and 2021, the percent of a certain demographic of students who took that test for that student (variable), and the percentage difference for the regents exam scores between the years of 2018 and 2021. Some next steps that I would like to take with my research would be to attempt to incorporate a machine learning model approach to the problem. Additionally, I would like to analyze across more than two years and even take a look at data from before 2018. However, with the results presented in this paper, it is our belief that it can be successfully used to draw the attention of lawmakers and other government officials to this pressing matter. With this data, it is apparent that certain demographics of students were affected more so compared to others and the necessary action must be taken to prevent any long standing damages to student academic achievement. Whether this be through increased technology access, accessible learning, or through the providal of adequate resources.

8. Acknowledgements

[2-3 sentences max] In this section, you should briefly thank or acknowledge any individuals or institutions that provided significant help and/or advice on your research, analysis, and report writing

I give my sincerest thanks to the New York State Education Department, for providing the data that fueled this project. I also give my thanks to Kush Khosla and the InspiritAI organization for giving me the opportunity to learn about and work with these data sets to form the research that I've done.

9. References

[No page limit] List all the articles and other sources that you used to compile this final paper. Citing your sources and keeping them carefully organized will greatly help in furthering your own research. There are many online citation resources that you can use. Additionally, there are many reference managers that simplify this process. Once you've compiled all your resources, add them here and add any relevant in-text citations.

[Your References Section Here]

- https://www.mckinsey.com/industries/education/our-insights/covid-19-and-education-the-lingering-effects-of-unfinished-learning