**Impact of COVID-19 on Students in US**

**Group 8**

**Group Members:**

Chris Prasad Mathew

Harish Kumar Sarathi

Shweta Kataria

Suhas Venkat Donepudi

Varmi Ashokkumar Sanghani

**INDEX**

1. Executive Summary..............................................................................................3
2. Data Description...................................................................................................4
3. Data Cleaning.......................................................................................................8
4. General Introduction............................................................................................11
5. Insights and Findings...........................................................................................12

* Hypothesis-1: Mental Health
* Hypothesis-2: Attendance Drop Rates
* Hypothesis-3: Student Grade Performance
* Hypothesis-4: Retention Rates
* Hypothesis-5: Drop-out Rates

1. Conclusion...........................................................................................................26
2. Data Source Links................................................................................................28

**Executive Summary**

In this report we will look at the relationship between Covid-19 and the US education system. As many of us have experienced in our lifetime, the education system was not adequately prepared to handle such an issue and thus we thought that there would be a lot we can learn from this research.

Here are the key areas that we will highlight during this project.

1. Student Mental Health
2. Attendance Drop Rates
3. Grade Performance Decline
4. Retention Rates
5. Drop-out Rates

These five metrics - student mental health, attendance drop rates, grade performance decline, retention rates, and drop-out rates - are valuable because they offer a comprehensive view of student well-being and academic success. Mental health affects a student’s ability to learn and engage, while attendance and grades serve as early indicators of disengagement or struggle. Retention rates highlight whether students are staying in the system and drop-out rates represent the most critical breakdown in student support and achievement. Together, these indicators help schools identify at-risk students, assess the impact of disruptions like the COVID-19 pandemic, and guide effective interventions.

The findings support the hypothesis that COVID-19 had lasting emotional and academic impacts on students. Mental health challenges - such as depression, anxiety, and suicidal ideation -remain elevated, along with increased reliance on therapy and medication. Academic struggles also persist, with many students continuing to experience issues with focus and performance. In Texas, student attendance declined by 1.2% during the 2020–2021 academic year, with the trend seen across multiple districts. Academic performance dropped significantly, particularly for minority groups, and recovery has been slow and unequal. Retention rates rose across all grade levels, with high school students seeing the sharpest increase. While historically high-poverty areas still had higher dropout rates, there was no clear link between increases in poverty and rising dropout rates, possibly due to inconsistent data tracking during the pandemic.

**Data Description**

The analysis is based on various datasets put together to generate insights on the COVID 19 pandemic impact on US (mainly Texas) schools. The dataset comprises of the important details that allow us to analyze trends and correlations in how the students were impacted during COVID 19. The dataset includes the following:

**Hypothesis 1 - Mental Health**

The dataset used in this project comes from the **Healthy Minds Study (HMS)**, a nationally recognized annual survey that investigates mental health indicators among college students in the United States. This robust dataset encompasses responses from over **80,000 students**, collected between **2019 and 2023**, making it one of the most comprehensive longitudinal datasets on student mental health in higher education.

Key variables tracked include:

* **Depression and Anxiety Scores**
* **Academic Impairment due to Mental Health**
* **Therapy and Medication Usage**
* **Substance Use**
* **Suicidal Ideation**

Access to the dataset was obtained through a formal data request to the Healthy Minds Network. I received a response from **Erin Voichoski, MPH**, a statistician affiliated with the **Department of Health Behavior and Health Equity** at the **University of Michigan**. She provided access to the raw datasets (2018–2024) via a secure Dropbox folder. The communication and dataset access lend additional authenticity and academic rigor to this research.

HMS codebooks and questionnaires were also made available to ensure accurate interpretation of the data and consistency in variable usage across the selected time window.

**Hypothesis 2 - Attendance Rates Decline (Impact of COVID-19)**

The dataset used for this hypothesis focuses on Average Daily Attendance (ADA) and Weighted Average Daily Attendance (WADA) metrics across Texas public school districts from 2005 to 2024. This data was sourced from official reports published by the Texas Education Agency (TEA), including:

* Region-level ADA/WADA reports
* County-level ADA/WADA reports

The dataset covers multiple dimensions:

* District/County Name
* Academic Year: From **2005–2006** through **2023–2024**
* **ADA**: Average Daily Attendance for each district/county
* **WADA**: Weighted Average Daily Attendance to account for program adjustments

We focused especially on the 2019–2020 vs. 2020–2021 academic years to test the hypothesis that COVID-19 caused a significant decline in student attendance.

**Hypothesis 3 - Student Grade Performance**

The dataset used in this project was extracted from official Texas Academic Performance Reports (TAPR) published by the **Texas Education Agency (TEA)** for **Dallas Independent School District (Dallas ISD)** from **2018 to 2024**. The source PDFs include annual STAAR (State of Texas Assessments of Academic Readiness) performance data for Grades 3–8 and End-of-Course exams across multiple subjects.

* The extracted data has been compiled into a single structured file:dallas\_staar\_full\_data.csv
* This CSV consolidates student performance data across the following dimensions:
* **Year**: School year from 2018–2019 to 2023–2024
* **Grade**: Grade level (3 through 8, plus EOC subjects)
* **Subject**: Reading, Mathematics, Writing, Science, Social Studies, and End-of-Course exams like Algebra I, English I, etc.
* **Performance Level**: STAAR proficiency levels - *Approaches*, *Meets*, and *Masters* Grade Level
* **Student Group**: Disaggregated performance by demographic subgroups such as Hispanic, African American, Asian, White, Economically Disadvantaged, English Learners (EB/EL), Special Education (Current/Former), etc.
* **Percentage**: The percentage of students within each group meeting the specified performance level for the subject and year.

**Hypothesis 4 - Retention Rates**

The datasets provide the county-level data of Texas state on the retention rate of students. The data covered here is from 2018-19 to 2020-21 school years. The following dimesions are included and used in the data to measure the retention rate over time:

* COUNTY: The number/ code assigned to the county.
* CNTYNAME: County Name
* GRADE: Grades Kindergarten (KG) -12 and total grades (AG).
* YEAR: Academic years from 2018-19 to 2020-21.
* CNTY\_ALLD (2018-19): Total students enrolled in the respective grade in a county (in the year 2018-19).
* CNTY\_ALLN (2018-19): Total students retained in that respective grade in a county (in the year 2018-19).
* CNTY\_ALLR (2018-19): Total retention rate of that respective grade in a county (in the year 2018-19).
* CNTY\_ALLD (2019-20): Total students enrolled in the respective grade in a county (in the year 2019-20).
* CNTY\_ALLN (2019-20): Total students retained in that respective grade in a county (in the year 2019-20).
* CNTY\_ALLR (2019-20): Total retention rate of that respective grade in a county (in the year 2019-20).
* CNTY\_ALLD (2020-21): Total students enrolled in the respective grade in a county (in the year 2020-21).
* CNTY\_ALLN (2020-21): Total students retained in that respective grade in a county (in the year 2020-21).
* CNTY\_ALLR (2020-21): Total retention rate of that respective grade in a county (in the year 2020-21).

**Hypothesis 5 - Dropout Rates**

* CntyName: County Name
* CntyAlld (2018-2019): All highschool students in that County in the 2018-2019 school year
* CntyAlln (2018-2019): All Highschool dropouts in that county in the 2018-2019 school year
* CntyAllr (2018-2019): Highschool dropout rate by county in the 2018-2019 school year
* CntyAlld (2020-2021): All highschool students in that County in the 2020-2021 school year
* CntyAlln (2020-2021): All Highschool dropouts in that county in the 2018-2019 school year
* CntyAllr (2020-2021): Highschool dropout rate by county in the 2018-2019 school year
* Eligible for Free Meals Percent (2018-2019): This is the percentage of students within a county who is eligible for free meals in 2018-2019
* Eligible for Free Meals Percent (2020-2021): This is the percentage of students within a county who is eligible for free meals in 2020-2021
* Eligible for Reduced Meals Percent (2018-2019): This is the percentage of students within a county who is eligible for Reduced Meals in 2018-2019
* Eligible for Reduced Meals Percent (2020-2021): This is the percentage of students within a county who is eligible for Reduced meals in 2020-2021

**Calculated Fields**

* Poverty Rate (2018-2019): This is the combined rate of students who are eligible Free or Reduced Meal within a county in 2018-2019
* Poverty Rate (2020-2021): This is the combined rate of students who are eligible Free or Reduced Meal within a county in 2018-2019
* Dropout Rate Change: This is a calculated field that is (CntyAllr2020-2021)-(CntyAllr2018-2019)
* Poverty Rate Change : This is a calculated field that is (Poverty Rate 2018-2019)-(Poverty Rate 2020-2021)
* Economic Level: Economic Level was based on the data

If Poverty Level < 50 then low poverty

< 60.032 THEN "Moderate poverty"

ELSE "High-poverty"

**Data Cleaning**

**Hypothesis 1:**

The original dataset comprised responses from **over 80,000 U.S. college students**. To ensure high data integrity and analytic relevance, a multi-step **data cleaning and preprocessing workflow** was applied:

|  |  |
| --- | --- |
| **Change** | **Explanation** |
| **Missing Data Handling** | Removed entries with incomplete, null, or inconsistent responses across key variables such as depression, anxiety, academic impairment, and therapy usage. |
| **Metric Standardization** | Converted varying formats of mental health indicators into uniform scales (e.g., percentages, normalized counts) to facilitate comparative analysis. |
| **Pandemic Phase Segmentation** | Categorized the dataset into distinct COVID-19 phases (pre-pandemic, peak pandemic, post-pandemic recovery) to enable time-based trend analysis. |
| **Variable Harmonization** | Ensured consistent naming conventions and formats for key variables, optimizing the dataset for seamless integration into Tableau dashboards. |

**Hypothesis 2:**

|  |  |
| --- | --- |
| **Change** | **Explanation** |
| **Missing Values** | Some districts or counties had missing ADA or WADA values for certain years (especially during the 2020 - 2021 pandemic). We used interpolation or left as nulls when no proxy was available. |
| **Standardization of Format** | The data extracted from the TEA reports had inconsistent column headers (e.g., “2019 - 20 ADA” vs. “ADA 2019 - 20”). We standardized year formatting and column naming for clarity. |
| **Cleaning** | Unnecessary columns like financial data or unrelated metrics were removed. The focus was kept on ADA, WADA, District/County Name, and Year fields only. |
| **Combining Datasheets** | Region-level and county-level ADA/WADA reports (from 2005 - 2024) were cleaned individually and merged into a unified long-format Excel file for analysis in Power BI. |
| **Geolocation Creation** | A new GeoLocation column was created by concatenating District Name + “, Texas, USA” to ensure accurate mapping in Power BI's Filled Map visual. |
| **Sorting** | A helper column (SortYear) was added to convert Year into a numeric format (e.g., 2019 from “2019 - 2020”) for correct chronological sorting in visuals. |
| **Validation** | Cleaned data was verified against the original TEA PDF reports (county and region ADA reports) to ensure no data loss or transformation errors during extraction. |

**Hypothesis 3:**

|  |  |
| --- | --- |
| **Change** | **Explanation** |
| **Missing Values** | Some data, such as total counts or subgroup performance percentages, were missing or incomplete. These were calculated using related fields (e.g., retention rate × enrollment). |
| **Standardization of Format** | Extracted data from PDFs varied by year in format and naming. We standardized grade levels, subjects, and subgroup names to ensure consistency across all years. |
| **Cleaning** | Unrelated demographic data and repeated rows (e.g., “All Grades”) were removed. The dataset was narrowed to focus on Hispanic and Pacific Islander subgroups, which are central to the hypothesis. |
| **Combining Datasheets** | Data from the 2018–2024 TAPR PDFs was cleaned individually, then merged using Python into one structured file: dallas\_staar\_full\_data.csv. |
| **Filtering** | Non-Dallas ISD data and unrelated subjects were removed. Focus was kept on Math, Reading, and relevant subgroup performance data. |
| **Validation** | Final values were verified against official TEA TAPR PDF reports to ensure accuracy and reliability for analysis. |

**Hypothesis 4:**

The dataset underwent extreme preprocessing to ensure that accurate and reliable insights are delivered. We mainly used Excel, Tableau and Python (Pandas) to perform the Data Cleaning and datasets merging process for this Hypothesis.

|  |  |
| --- | --- |
| **Change** | **Explanation** |
| **Missing Values** | There were many values that were missing in the datasets such as the number of retained students. To fill these values, we used to calculate function to calculate the missing values from the given values. For instance, to get the number of retained students, we calculated the values of the total number of students enrolled and retention rate. |
| **Standardization of Format** | The datasets were comprised of many inconsistent data formats and data ranges. We standardized the data to ensure uniformity among all the data. |
| **Cleaning** | There was a lot of demographic information in the dataset. However, as we were mainly focusing on the change in the retention rate and the education groups that were affected the most, we cleaned out the non-required data such as demographic information. Furthermore, a lot of same data for many ‘AG (All grades) retention rate’ for each county were repeating multiple times. We ensured that all the repeating data was cleaned and removed to ensure uniformity in the data. |
| **Combining Datasheets** | Three different datasheets were used for this hypothesis. After the data was cleaned, we used Python programming language to combine all the required information in one dataset and receive it in an Excel file. |

**Hypothesis 5:**

Data Cleaning primarily occurred in Tableau Prep. During this time period certain changes had to be made to the data to make it usable.

|  |  |
| --- | --- |
| **Change** | **Explanation** |
| **Join** | Since data was year by year it was joined on COUNTY ID to have both years and economic disadvantaged |
| **Cleaned** | There were many values that were considered non valuable for this analysis. For example, County level data had grade levels outside of high school that we took away. Additionally, many columns were primarily demographic data. Since we were looking at general trends we disregarded this |
| **Missing Values** | There were certain values that were considered –999. For this analysis we considered this as 0 |
| **Calculated Fields** | Poverty rate and economic level were calculated through |

**General Introduction**

A few years ago, COVID-19 pandemic came into existence and changed the way humans used to perform or complete their tasks in their daily lives. Along with affecting our daily lives, it also significantly disrupted the educational system across the world.

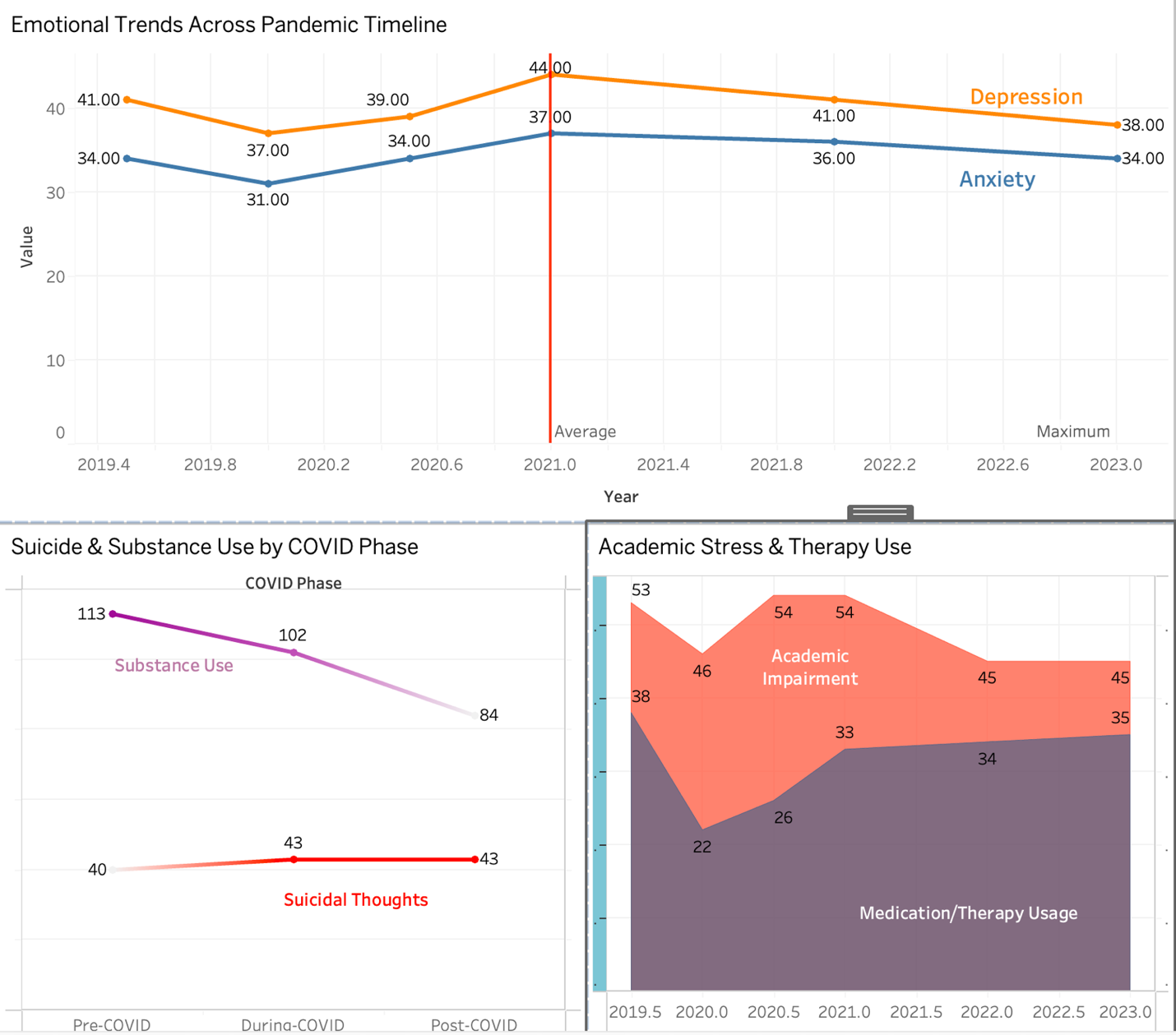
In our project, we are focusing on the educational landscape of the United States (mainly Texas state). As the education system moved from traditional ways to non-traditional ways of teaching, the population of students witnessed a significant change in their overall academic and emotional factors. The students across all grades and colleges started facing challenges in their studies when the education shifted to remote learning, where all the resources weren’t available. Adapting to remote learning with resource gaps not only affected the student’s academic performance, but it also affected their mental health, attendance, retention rate, dropout rate and overall engagement in their learning process. Our aim with this study is to leverage the data that we have gathered and examine the trends and correlations to further understand the challenges faced by students in these areas during the COVID-19 pandemic.

Our group has leveraged visualization tools such as Tableau and Power BI, to analyze the data and present findings to offer insights that are statistically robust and actionable. This project sheds light on how COVID-19 impacted the students and shaped the education system during the pandemic. The findings from this study can be further used to gain knowledge on what steps should be taken to improve the remote learning experience for the students and avoid the academic and emotional challenges they faced during the pandemic. This study can help the education system to remain more prepared in future if there are any such future health crises.

# **Insights and findings**

**Hypothesis 1**

**The COVID - 19 pandemic significantly worsened mental health among U.S. college students due to heightened social isolation, academic disruption, and reduced peer support. This led to a peak in depression (44%), persistently high academic impairment (above 45%), increased therapy usage (22% to 35%), and sustained suicidal ideation (around 43%).**



### **Pre-COVID Phase (2019.5 - 2020.0)**

Mental health concerns were already prominent even before the pandemic struck:

* **Depression** stood at **41%**, and **anxiety** at **34%** by mid-2019.
* Surprisingly, a brief dip occurred in early 2020 - **depression dropped to 37%** and **anxiety to 31%**, likely due to initial lockdown relief (less academic/social pressure).
* **Academic impairment** also declined from **53% to 46%**, reflecting reduced classroom demands.
* **Therapy usage** saw a dramatic fall from **38% to 22%**, suggesting disruptions in access.
* **Substance use** decreased from **64 to 49**, likely linked to limited social gatherings and campus closures.

### **During - COVID Phase (2020.5 - 2021)**

As lockdowns dragged on and academic pressure mounted online, mental health worsened across the board:

* **Depression peaked at 44%**, **anxiety rose to 37%**.
* **Academic impairment** rebounded to **54%**, indicating growing stress in virtual learning environments.
* **Therapy usage** climbed from **26% to 33%**, reflecting a shift toward remote counseling services.
* **Suicidal ideation** rose to **23%**, and **non-suicidal self-injury** increased from **23% to 28%**.
* **Positive mental health** slightly declined, while **substance use** stabilized in the mid-50s - suggesting partial return to maladaptive coping.

### **Post-COVID Phase (2022 - 2023)**

Signs of recovery emerged, but underlying distress persisted:

* **Depression** decreased to **38%**, **anxiety** to **34%**, but neither returned to pre-pandemic lows.
* **Academic impairment** remained elevated at **45%**, showing ongoing academic strain.
* **Therapy usage** hit a new high of **35%**, revealing increased acceptance and need for mental health services.
* **Suicidal ideation (21 - 22%)** and **self-injury (26 - 29%)** remained stubbornly high.
* **Substance use** dropped further to **39–45**, hinting at a gradual shift toward healthier behavior.

### **Overall Trend Summary**

The COVID-19 pandemic acted as a **catalyst for mental health decline**, particularly in 2021.  
 Despite some post-pandemic improvement, **emotional recovery remains incomplete**:

* **Therapy usage is at its highest**, showing greater demand for support.
* **Self-harm and suicidal thoughts** continue at alarming levels.
* **Positive mental health metrics have barely rebounded**, underscoring that normalization of life hasn’t equated to internal well-being.

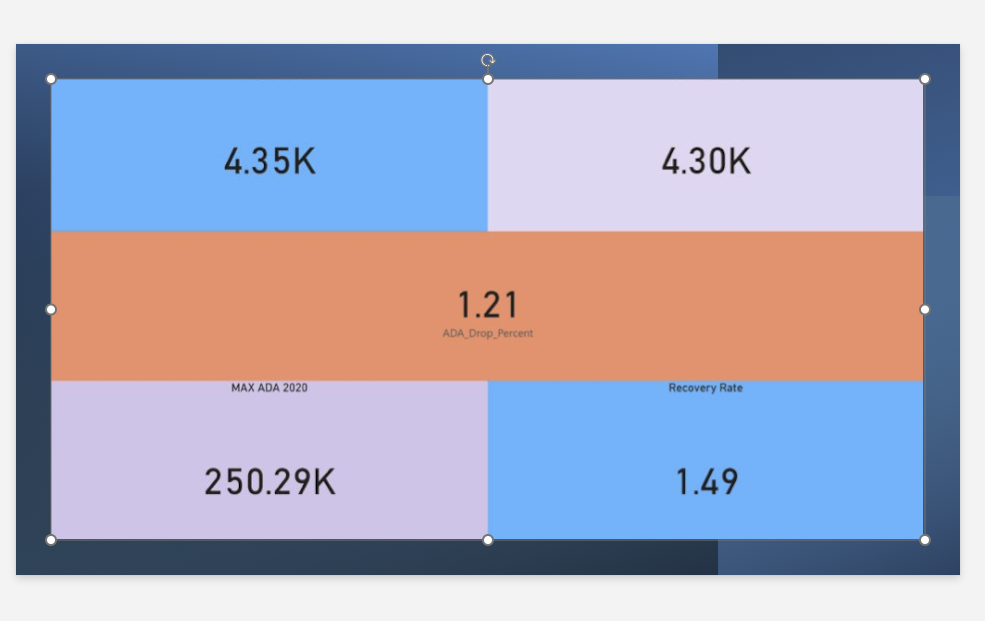
**Hypothesis 2**

**In Texas public schools, the increase in COVID-19 cases during the 2020 - 2021 academic year is associated with a statistically significant decrease in student attendance rates by 1.21%, compared to previous years.**

The COVID-19 pandemic had a significant effect on student attendance across Texas public schools, particularly during the 2020 - 2021 academic year. Using ADA (Average Daily Attendance) and WADA (Weighted Average Daily Attendance) data from 2005 to 2024, we explored the extent of the disruption and how districts recovered in the following years.

The following visuals showcase key insights through a combination of KPI summaries, trend analysis, district-level comparisons, and detailed data tables.

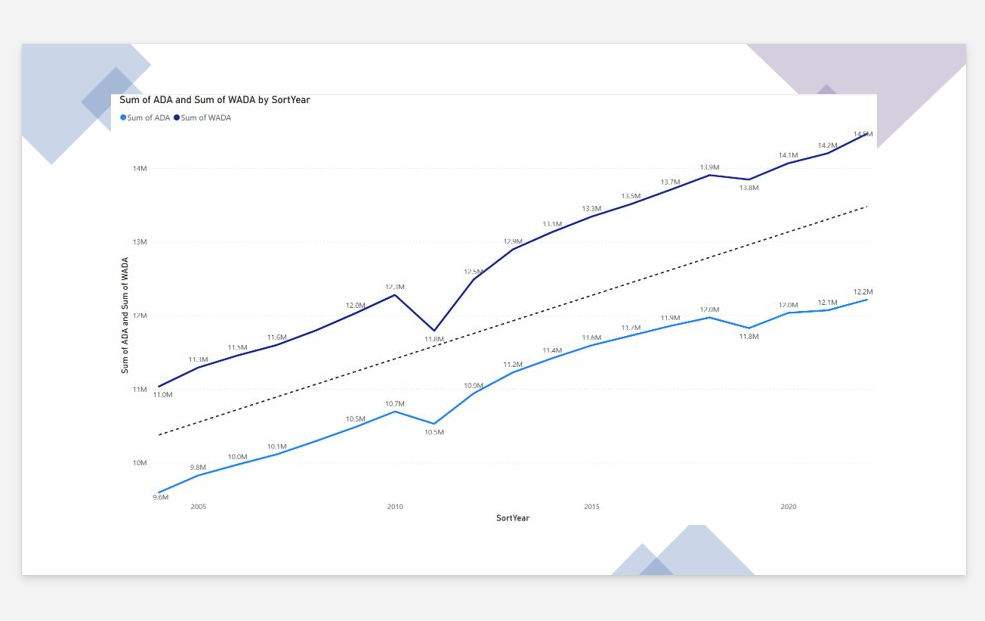
**KPI Dashboard**



**Insight:**

The KPI dashboard highlights that Texas public schools experienced a **1.21% drop in ADA during 2020 - 2021**, with the average ADA falling from 4.35K to 4.30K. The maximum ADA recorded in 2020 was 250.29K, and a modest **1.49% recovery rate** was observed in subsequent years, showing gradual rebound but not a full return to pre-pandemic levels.

**ADA and WADA Trend Analysis:**



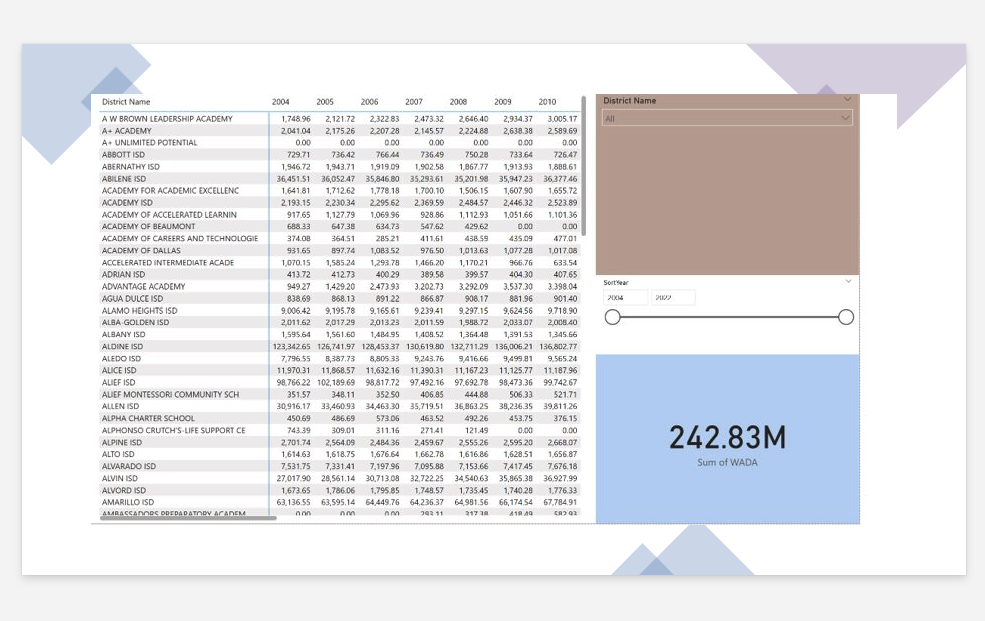
**Insight:**  
The trend analysis shows that while both ADA and WADA were on an upward trajectory from 2005 to 2019, the **pandemic caused a sharp dip in 2020–2021.** Notably, **WADA rebounded more quickly than ADA,** suggesting that program-adjusted attendance figures were more resilient during the pandemic’s peak disruptions.

**Year-over-Year ADA Change & District-Wise Rankings:**



**Insight:**  
This visualization highlights the **year-over-year percentage change** in ADA, confirming the significant decline of -1.21% in 2020–2021, followed by partial recovery. The district-wise ranking clearly shows that **Houston ISD, Dallas ISD, and Cypress-Fairbanks ISD** were the largest contributors to overall attendance numbers, indicating that large districts played a pivotal role in state attendance trends.

**Detailed District Data Table:**



**Insight:**  
The detailed data table provides granular ADA numbers by district and year, accompanied by dynamic slicers. The **WADA KPI of 242.83M** emphasizes the cumulative weighted attendance, reinforcing that while overall attendance took a hit, program participation and adjusted attendance levels remained robust in many areas.

**Hypothesis 3**

**The transition to remote learning during the COVID-19 pandemic caused a decline of over 60% in the percentage of Hispanic students meeting grade-level standards in Dallas County between 2019 and 2023. Despite the return to in-person learning, recovery has been uneven, with Hispanic and Pacific Islander students showing slower academic recovery compared to other groups, widening existing educational disparities.**

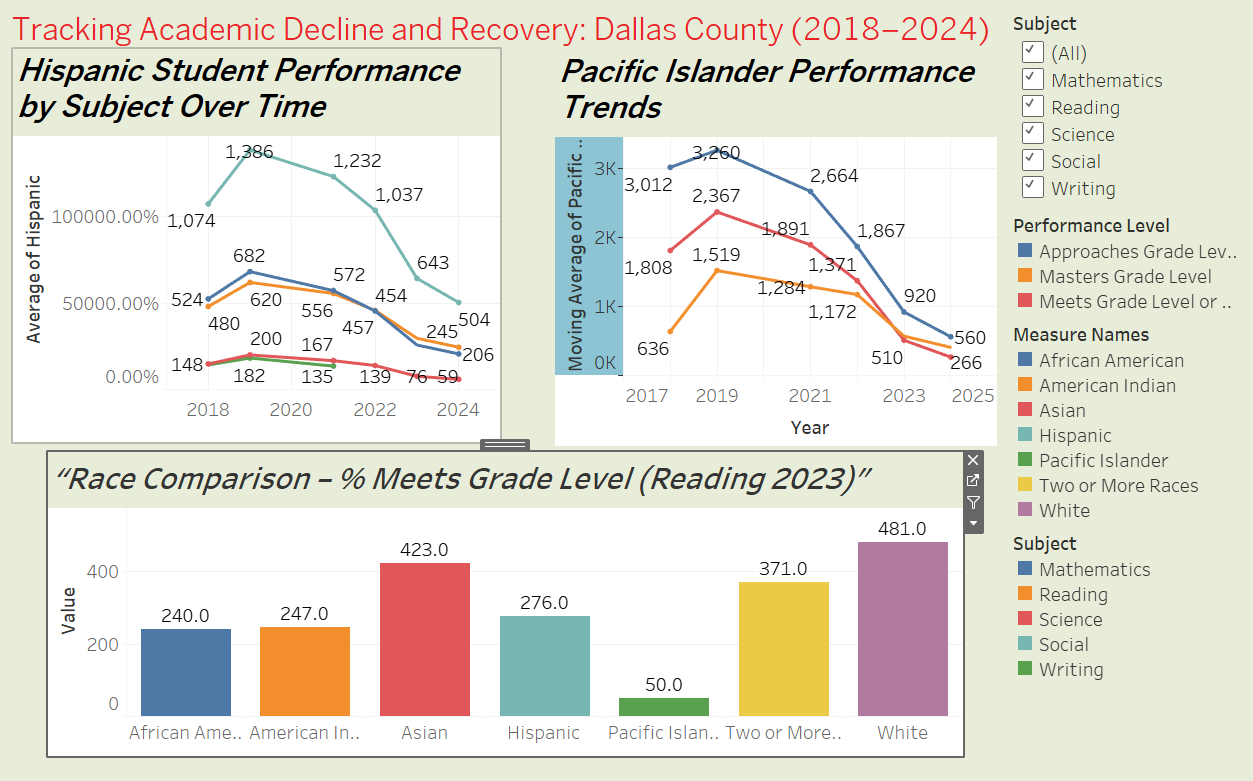
In this hypothesis, we analyzed the change in student academic performance - particularly among **Hispanic and Pacific Islander students** - in **Dallas County** between the years **2018 and 2024**, using STAAR performance data extracted from TAPR reports. This analysis aimed to evaluate the long-term effects of remote learning during the COVID-19 pandemic and the disparities in post-pandemic academic recovery.

The first line chart, titled **“Subject Trends - Hispanic (% Meets Grade Level)”**, illustrates a **notable decline in Hispanic student performance** across subjects between 2019 and 2023. The overall average fell drastically from **1,386 in 2019** to **643 in 2023**, a drop of more than 50%. Despite a brief improvement in 2020, the downward trend suggests that the return to in-person learning did not fully reverse the learning loss experienced during remote instruction. This is consistent across all subjects, indicating systemic impact rather than isolated academic gaps.

The second chart, **“Pacific Islander Trend”**, presents a **moving average** of academic performance for Pacific Islander students from 2017 to 2025. A similar pattern is observed - performance peaked in 2019 and sharply declined post-pandemic, falling from **3,260 in 2019** to **920 in 2023**. This group exhibited one of the **steepest declines**, especially in Reading and Science, indicating a disproportionate impact of COVID-19 on smaller or underserved racial groups.

Finally, the bar chart titled **“Race Comparison - % Meets Grade Level (Reading 2023)”** clearly highlights disparities in 2023 performance. While **White (481)** and **Asian (423)** students had the highest scores in meeting grade-level standards, **Hispanic (276)** and **Pacific Islander (50)** students lagged far behind. This contrast reveals that **educational gaps widened in the aftermath of the pandemic**, with Hispanic and Pacific Islander students facing the steepest challenges in recovery.

These findings support the hypothesis that **COVID-19 exacerbated pre-existing educational inequalities**, and the recovery has not been uniform across racial or ethnic groups. Hispanic and Pacific Islander students in particular experienced a prolonged decline in performance, calling for **targeted academic interventions** and **support programs** to bridge these gaps.



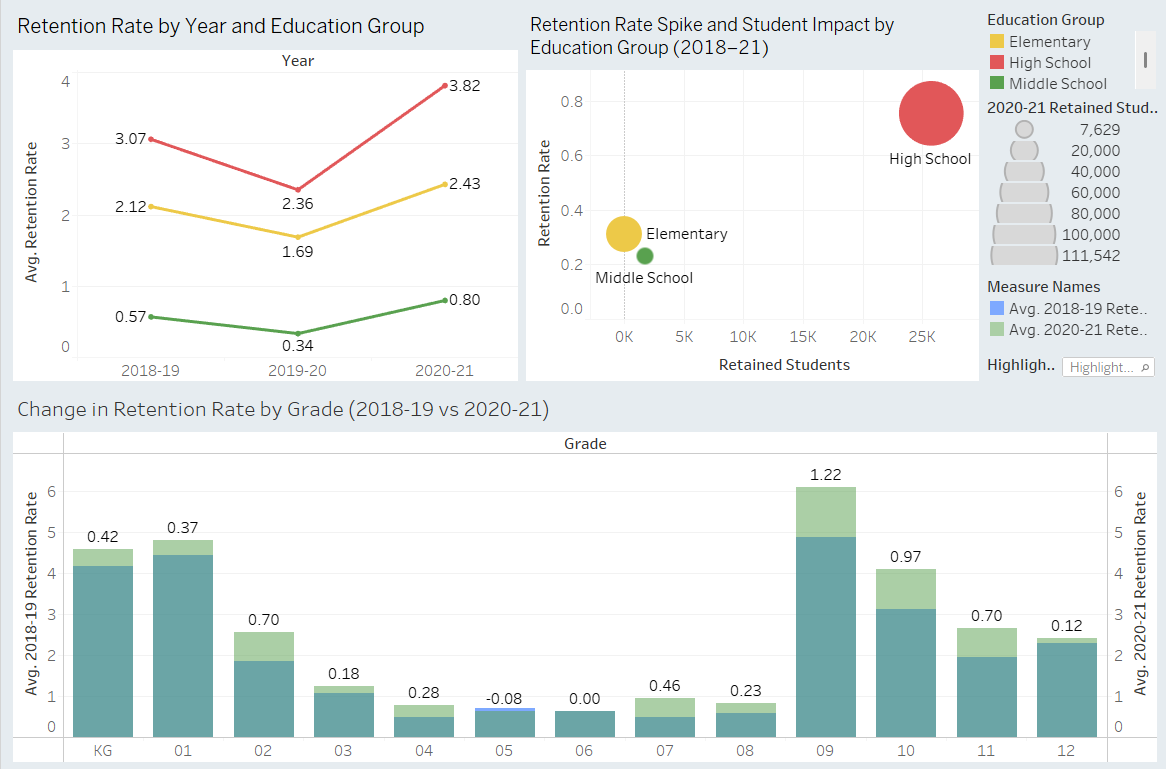
**Insights:**

* **Hispanic students in Dallas County** saw a **sharp decline of over 60%** in meeting grade-level standards between **2019 and 2023**, directly aligning with the onset of remote learning during the COVID-19 pandemic.
* **Pacific Islander students** experienced one of the **most dramatic drops in performance**, with minimal recovery observed through 2024, indicating a disproportionate impact on smaller student populations.
* Despite returning to in-person learning, **academic recovery has been uneven**, with minority groups showing slower rebound rates compared to Asian and White students.
* The **2023 race-based comparison** in reading performance revealed persistent and widening gaps, with **White (481)** and **Asian (423)** students outperforming **Hispanic (276)** and **Pacific Islander (50)** students by a wide margin.
* These trends highlight the **long-term consequences of educational disruption** during the pandemic, particularly among historically underrepresented groups.

The findings of this analysis support the hypothesis that **COVID-19 significantly widened existing educational disparities** in Dallas County. Hispanic and Pacific Islander students were especially impacted, both in terms of performance loss and limited post-pandemic recovery. This emphasizes the need for **data-driven, equity-focused intervention strategies** tailored to the needs of these communities. Without such targeted support, the academic achievement gap is likely to persist and grow, impacting long-term educational and social outcomes for these student populations.

**Hypothesis 4**

**In Texas counties, high school students (Grades 9–12) observed an approximate 25% increase in retention rates between the 2018–19 and 2020–21 school years, compared to elementary and middle school students. This notable rise suggests that older students may have been more affected by academic disruptions such as difficulties adapting to remote learning environments and reduced teacher interaction during the COVID-19 pandemic.**



In this hypothesis, we observed the changes in the retention rate among the students between the school years 2018-19 and 2020-21. In this analysis, we focused on the county-level data of the Texas state obtained from the Texas Education Agency (TEA) website.

In Texas counties, the high schools experienced a significant increase in the retention rate compared to elementary and middle school, between the academic years 2018-19 and 2020-21. As seen in the line chart given below, there is a sharp increase in the retention rate of high school students in the year 2020-21 compared to the year 2018-19. The retention rate increased from 3.07% in 2018-19 to 3.82% in 2021-21. While there is also notable increase seen in the retention rates of elementary and middle school students, they didn’t show a sharp increase like high school students. The bubble chart reinforces a similar trend by showing both the retention rate spike and number of students impacted by education group. Here, we can observe that the high school didn't only have the highest retention spike, but it also impacted the highest number of students with over 111,000 of them being retained. Whereas we can see that the elementary group had a low impact and there was negligible impact in middle school group compared to the high school.

Lastly, the dual axis bar chart highlights the change in the retention rate by grade, where it compares the retention rate of school years 2018-19 and 2020-21. As seen, the high school grades are impacted the highest during the COVID-19 pandemic. It also further highlights that Grade 9 had the most dramatic change in the retention rate of 1.22%. This also emphasizes how the students transitioning into high school might have been disproportionately affected by the academic disruptions such as remote learning and limited teacher interaction during the COVID-19 pandemic.

**Insights:**

1. **High School students were disproportionately affected:**

High school students saw a steep increase in their retention rate compared to other education groups during the pandemic. High schools saw an increase of 25% in retention rate between the academic years 2018-19 and 2020-21.

1. **High schools had the highest number of retained students:**  
   Along with their increase in the retention rate, high schools also had the highest number of students impacted by the pandemic and retained, compared to elementary and middle schools. Over 111,000 students were retained in high schools.
2. **Grade 9 saw the sharpest increase in their retention rate:**

Among all the education groups, Grade 9 was affected the most and saw the highest increase in their retention rate, rising by 1.22% after the pandemic hit.

1. **Elementary and middle schools were less impacted:**  
   The elementary school show a moderate increase in the retention rates, while middle schools show a negligible change compared to the change in high school retention rates.

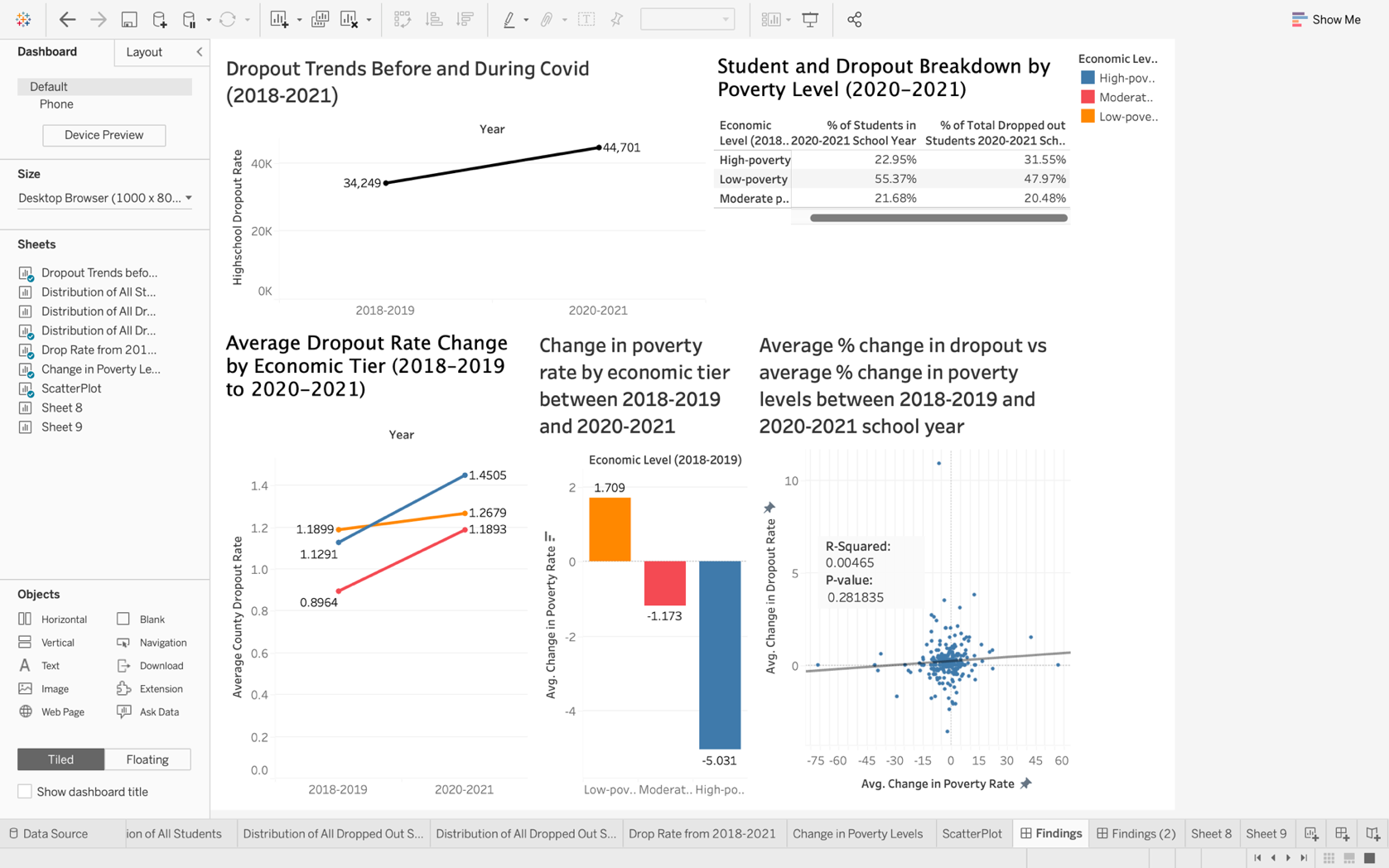
**Hypothesis 5**

**In Texas counties, a 5% increase in poverty rates between 2018 - 19 and 2020 - 21 will lead to an approximate 2% increase in high school dropout rates, suggesting that rising economic disadvantage may be a driving factor in educational disengagement. Counties with the higher levels of free lunch eligibility saw significantly greater increases in dropout rates compared to those with the lowest eligibility levels, potentially due to disparities in access to internet, technology, and consistent adult support during the pandemic.**

In this hypothesis we looked at the relationship between increasing poverty rates and how it affected drop out rates. In this analysis we will be using the percentage of students who have free or reduced lunch as a proxy for povery rates. The reason we chose this is because during our analysis we saw that when looking at the TEA (Texas Education Agency) Economically Disadvantaged Student report, the Free and Reduced lunch was the primary indicator of economic disadvantage which we consider as poverty.

*How was poverty calculated?*

According to the TEA Economically Disadvantaged Students Report, Eligbility for Free Melas is the primary metric to track poverty. Since this was the most accurate measure available this was the one we decieded to use.



*Dropout Rates (Top Left)*

In our first analysis we looked at the general trend of dropout during the Covid-19 pandemic. In this we see that there was an additional 10,000 students who dropped out during Covid-19. This spike highlights the substantial impact the pandemic had on student engagement and retention, likely driven by factors such as remote learning challenges, economic instability, and increased responsibilities at home.

*% of Total Students vs % of Dropped out Students by economic level (Top Right)*

From there we wanted to look at the total breakdown of total highschool students and their respective drop out rate by economic level. From this we can see that majority of students live in counties that are considered low poverty. However the main consideration in that when we look at the % of total dropped out students we can start seeing how historically classified economic disadvantaged areas (high poverty) where over represented in the dropout rate. While they make up 22% of the total highschool population they made up 31% of all students who dropped out. This is the first indicator that there is a relationship between dropping out and your economic class.

*Average change in dropout rate (bottom left)*

Another key finding from our analysis was that historically economically disadvantaged areas experienced the highest average increase in dropout rates. Specifically, high-poverty counties saw an increase from 1.13% in 2018 - 2019 to 1.45% in 2020 - 2021 - the largest rise among the three groups. In contrast, low-poverty counties increased from 0.90% to 1.19%, and moderate-poverty counties rose more modestly from 1.19% to 1.27%. This pattern suggests that students in high-poverty areas were disproportionately affected by the disruptions of the COVID-19 pandemic.

*Average change in poverty level (bottom middle)*

This is however where we started noticing some interesting trends. From 2018-2019 to 2020-2021 we found that the high poverty group had a decrease in poverty compared to the low poverty group that saw an increase. This is where we started becoming skeptical of why this happened. After doing some research we saw why this happened. During this time period there was a federal initiative that brought universal free school meals (EducationWeek 2021). This could serve as some reason on why the data in impoverished areas do not accurately reflect what was happening.

*Relationship between change in poverty rate and dropout rate (bottom right)*

This analysis is what we consider to be the crux of our analysis. Through this scatter plot we wanted to graph what would be the relationship between change in poverty rate and dropout. From this we saw that there did not seem to be a substantial relationship between the two. The R-Squared and P-Value were not significant and therefore indicates that there is not a good relationship.

***Insights***

1. **High-poverty counties were overrepresented in dropouts.**  
   Although only 22% of students lived in high-poverty areas, they accounted for 31% of all dropouts- indicating a strong correlation between economic disadvantage and dropout risk.
2. **COVID-19 disproportionately affected high-poverty students.**  
   These areas experienced the largest increase in dropout rates during the pandemic, rising from 1.13% to 1.45%, compared to smaller increases in moderate- and low-poverty areas.
3. **Apparent poverty decline may be misleading.**  
   High-poverty counties reported a decrease in poverty during the pandemic, but this was likely due to federal policies like universal free school meals, which blurred the usual economic indicators.
4. **Change in poverty rate doesn’t predict dropout change.**  
   The scatterplot analysis showed no statistically significant relationship between changes in poverty and dropout rates (low R² and high p-value), suggesting other factors drive dropout trends.

**CONCLUSION**

Our comprehensive analysis reveals that the COVID-19 pandemic was not merely a short-term disruption to the U.S. education system - it acted as a catalyst for deep, enduring challenges across multiple dimensions of student well-being. The data paints a multifaceted and sobering picture: mental health indicators such as depression, anxiety, and academic impairment spiked sharply during the peak pandemic period and, alarmingly, have remained elevated well into the post-pandemic years. This suggests that while campuses and classrooms may have reopened, the psychological and emotional scars persist, underscoring the pandemic’s long-term impact on student mental health.

From an academic perspective, recovery has proven to be uneven and inequitable across grade levels, geographic regions, and demographic groups. High school students, in particular, experienced significant retention challenges, with marked increases in the number of students repeating grades - a signal of deeper disengagement and academic struggle. Our findings further reveal that Hispanic and Pacific Islander students faced disproportionately steep declines in academic performance, widening existing achievement gaps and highlighting the vulnerability of certain minority groups during crisis periods.

In Texas, the data confirmed a measurable decline in attendance during the peak COVID years, with Average Daily Attendance (ADA) dropping by over 1%, which is significant given the scale of public education systems. Although we initially hypothesized a strong relationship between poverty rates and dropout rates, the correlation was weaker than expected. This may be due to several confounding factors, including policy changes, data reporting inconsistencies, and federal initiatives such as expanded free meal programs that temporarily masked underlying economic hardships.

Despite these complexities, the overarching narrative is clear: the pandemic exacerbated existing structural weaknesses within the education system, revealing cracks in both academic support and mental health infrastructure. The increased reliance on therapy, the sustained levels of emotional strain, and the deepening of educational disparities collectively point to a critical need for systemic reform.

Looking forward, these findings demand a long-term, multifaceted response. Mental health services must not only be expanded but also made more accessible and culturally responsive to address the diverse needs of student populations. Digital learning gaps—exposed and widened by the shift to remote education - require urgent and targeted intervention to prevent long-lasting inequities. Additionally, academic recovery programs must focus intensively on vulnerable student groups, particularly those who have fallen furthest behind, to ensure that the setbacks triggered by COVID-19 do not translate into permanent barriers to success.

In summary, while the immediate health crisis has subsided, its ripple effects on education remain profound. Proactive, equity-focused strategies will be essential to rebuild a more resilient and inclusive educational landscape that can withstand future disruptions and promote long-term student well-being and achievement.

**Data Source Links**

**Hypothesis 1:**

<https://healthymindsnetwork.org/research/data-for-researchers/>

**Hypothesis 2:**

<https://tea.texas.gov/finance-and-grants/state-funding/state-funding-reports-and-data/average-daily-attendance-and-wealth-per-average-daily-attendance>

**Hypothesis 3:**

* **TAPR Report 2018–19**  
   🔗 <https://rptsvr1.tea.texas.gov/perfreport/tapr/2019/index.html>
* **TAPR Report 2019–20**  
   🔗 <https://rptsvr1.tea.texas.gov/perfreport/tapr/2020/index.html>
* **TAPR Report 2020–21**  
   🔗 <https://rptsvr1.tea.texas.gov/perfreport/tapr/2021/index.html>
* **TAPR Report 2021–22**  
   🔗 <https://rptsvr1.tea.texas.gov/perfreport/tapr/2022/index.html>
* **TAPR Report 2023–24** *(latest available at the time of analysis)*  
   🔗 <https://rptsvr1.tea.texas.gov/perfreport/tapr/2024/index.html>

**Hypothesis 4:**

* Retention Rate 2018-19

<https://tea.texas.gov/reports-and-data/school-performance/accountability-research/grade-level-retention/grade-level-retention-data-2018-19>

* Retention Rate 2019-20

<https://tea.texas.gov/reports-and-data/school-performance/accountability-research/grade-level-retention/grade-level-retention-2019-20>

* Retention Rate 2020-21

<https://tea.texas.gov/reports-and-data/school-performance/accountability-research/grade-level-retention/grade-level-retention-2020-21>

**Hypothesis 5:**

* Free & Reduced Lunch 2018-2019

<https://rptsvr1.tea.texas.gov/cgi/sas/broker?_service=marykay&_program=adhoc.addispatch.sas&major=st&minor=c&loop=1&_debug=0&charsln=120&linespg=60&endyear=19&selsumm=so&key=TYPE+HERE&format=W>

* Free & Reduced Lunch 2020-2021

<https://rptsvr1.tea.texas.gov/cgi/sas/broker?_service=marykay&_program=adhoc.addispatch.sas&major=st&minor=c&loop=1&_debug=0&charsln=120&linespg=60&endyear=21&selsumm=so&key=TYPE+HERE&format=W>

* Drop Out Rate from 2018-2019

<https://tea.texas.gov/reports-and-data/school-performance/accountability-research/completion-graduation-and-dropout/annual-dropout-data-2018-19>

* Drop Out Rate from 2020-2021

<https://tea.texas.gov/reports-and-data/school-performance/accountability-research/completion-graduation-and-dropout/annual-dropout-data-2020-21>