PYTHON-BASED ANALYSIS OF COVID-19'S IMPACT ON K–12 STUDENT PERFORMANCE ACROSS U.S. PUBLIC SCHOOLS

1. Background

The COVID-19 pandemic disrupted learning nationwide, but its impact on student performance varied widely across districts, grade levels, and socioeconomic backgrounds. A nonprofit education group commissioned us to conduct a data-driven analysis using public school data from three U.S. states (Texas, California, and Illinois).

The goal was to evaluate the effect of the pandemic on standardized test outcomes and attendance patterns, and to generate actionable insights for recovery-focused interventions.

2. Objective

- To use Python for analyzing multi-year educational data from public school systems
- To measure shifts in performance and attendance across the 2019–2022 academic years
- To identify the most affected student subgroups (e.g., by income, grade, or race/ethnicity)
- To deliver clear, structured reporting with visual narratives and recommendations for district-level planning

3. Data Used

Source:

- State-level public education datasets (California DOE, Texas Education Agency, Illinois Report Card)
- NCES (National Center for Education Statistics) school-level metadata

Dataset Scope:

- Over 1.5 million student records (aggregated at grade/district level)
- Years covered: 2018–2019 (pre-COVID), 2019–2020, 2020–2021, 2021–2022
- Key Variables:
    - District_ID, School_ID, Grade, Subject
    - Average_Test_Score

- Attendance_Rate
- Economic_Disadvantage_Rate
- Ethnicity, Gender
- Virtual_Learning_Periods

4. Methodology

4.1 Data Preparation

- Combined and standardized datasets across states using consistent schema
- Filled missing values with state average or prior-year data where appropriate
- Derived variables:
  - Score_Delta (year-over-year change)
  - Attendance_Impact_Index = (Pre-COVID Attendance – COVID-Year Attendance)
  - Group-level aggregates for subgroup analysis (e.g., low-income 6th graders in urban areas)

4.2 Analysis Performed

- Time series visualization of performance drop and recovery by grade
- Comparative boxplots and scatterplots of attendance vs score trends
- Grouped bar charts by student demographic segments
- Correlation and linear regression of virtual learning duration and test score decline
- Heatmaps to detect district clusters with the largest gaps

4.3 Tools Used

- Python environment: Jupyter Notebook via Anaconda
- Libraries: pandas, matplotlib, seaborn, plotly, statsmodels
- Report formatted in APA style with embedded plots

5. Key Results

- **Average statewide test score decline (2020–2021 vs 2018–2019)**:

    - Math: –11.3 percentage points
    - Reading: –6.9 percentage points
- **Districts with >70% economically disadvantaged students** saw 2× greater score drops
- Attendance dipped most in **urban middle schools** (avg. –5.4% below pre-COVID levels)
- **Black and Hispanic students** experienced steeper declines in math relative to white peers
- Strong negative correlation (–0.62) between number of virtual instruction weeks and test score recovery in 2021–2022

6. Insights Delivered

- Identified **top 20 at-risk districts** with sustained declines and minimal recovery
- Provided data to support **state-level tutoring grant applications**
- Quantified achievement gaps across **racial and economic groups** with visuals
- Suggested prioritization of **early-grade math recovery programs**
- Built a matrix to map **school resource needs** based on performance drop × attendance loss

7. Reporting Output

- **PDF Report (APA-style, 22 pages)**:
    - Executive summary with key findings
    - Performance trend graphs and subgroup gap analysis
    - Appendix: regression tables and interactive data filtering logic
- **Python Notebook**:
    - Modularized into data cleaning, analysis, and plotting sections
    - Includes custom visualization functions (e.g., plot_score_change_by_group())
    - Code export enabled for reuse by district-level data teams
- **Excel Dashboard**:
    - District-by-district score and attendance tracker
    - Color-coded performance alerts based on preset thresholds

- Filters for state, grade, economic segment, and race

8. Strategic Impact

- Used by 3 school districts to target **ESSER funding** into high-risk areas
- Informed **board-level presentations** to justify learning recovery investments
- Contributed to the design of an **after-school tutoring pilot program** for math in grades 3–5
- Helped shape **hybrid learning policies** for the 2023–24 academic year by evidencing past virtual learning tradeoffs