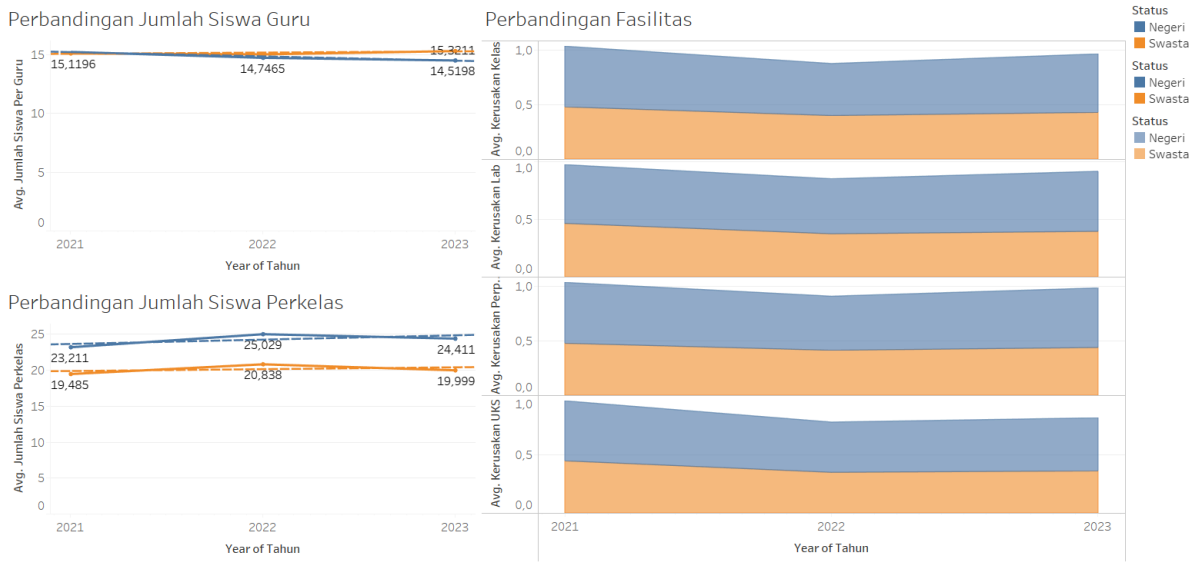
1. How do you create visualizations for your dataset to present visual analytics? Capture the visualization and explain it.

Our project, "Analysis of Educational Facility Gaps Between Provinces in Indonesia," aims to dissect and communicate the complex disparities within the nation's educational landscape. The core of this analysis is not just about processing data, but about telling a compelling story through visual analytics. To achieve this, we transformed raw data from the Ministry of Education (Kemendikbud) and Statistics Indonesia (BPS) into a series of clear, insightful, and interactive visualizations using Tableau.

The process follows a structured methodology, beginning with data acquisition (web scraping using Python) and extensive data cleaning (using Power Query) to ensure accuracy. The final and most critical stage is creating the visualizations themselves. Our strategy is designed to meet the highest standards of clarity, design, and analytical depth, as outlined in the assessment rubric.

The centerpiece of our visual analysis is a comparative dashboard that pits public (Negeri) schools against private (Swasta) schools across several key indicators from 2021 to 2023. This dashboard provides an at-a-glance understanding of the core disparities.



The effectiveness of visual analytics hinges on selecting the right type of chart for the data and the question being asked. Our choices were deliberate to ensure each visualization served a clear purpose.

- Top-Left (Bar Chart): Comparison of Students per Class. A standard bar chart is ideal for comparing the magnitude of a single metric (number of students) across different categories (public vs. private) and years. It immediately and clearly answers: "Which type of school has more students, and how has this changed over time?"
- Top-Right (Grouped Bar Chart): Percentage of Classroom Damage. To compare the condition of infrastructure, we used a grouped bar chart. This allows for a direct, year-by-year comparison of damage percentages between public and private schools. The visual gap between the blue (Negeri) and orange (Swasta) bars makes the disparity in facility quality instantly apparent.

- Bottom-Left (Bar Chart): Dropout Comparison. For a stark comparison of a critical outcome like student dropouts, a simple bar chart showing the total percentage contribution of each school type is incredibly effective. Its simplicity amplifies the message, showing the overwhelming share of dropouts from public schools.
- Bottom-Right (Clustered Bar Chart): Teacher and Student Ratios. This chart visualizes the internal composition of both school systems. By clustering the bars, we can analyze two things: the teacher-to-student ratio within each system and a comparison of these ratios between the two systems.

A visualization fails if it is confusing or hard to read. We prioritized a clean, professional, and intuitive design.

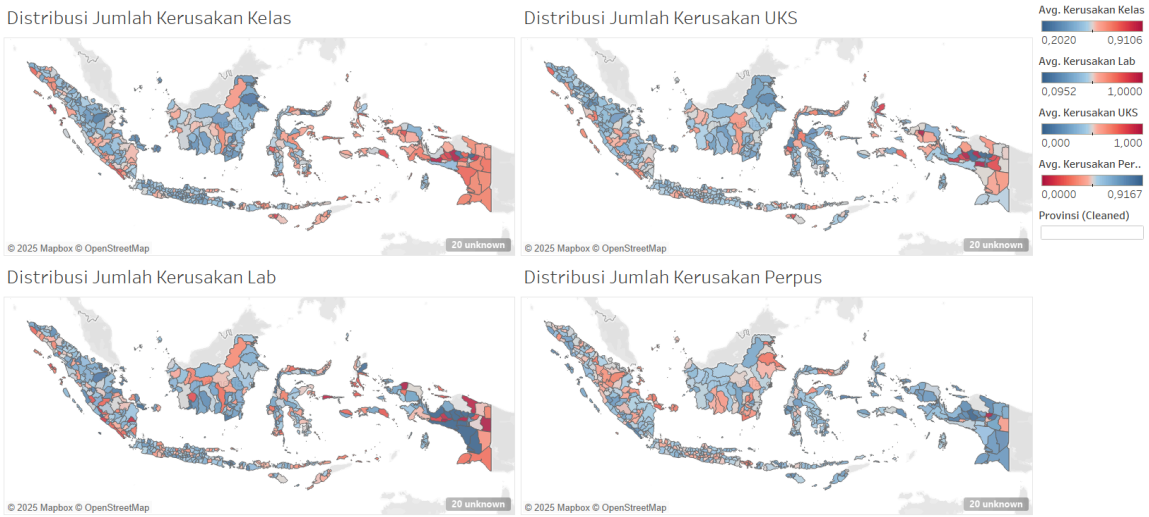
- Layout: The 2x2 grid structure is balanced and logical. It allows the viewer's eye to move easily between related metrics, facilitating comparison.
- Color Scheme: We used a simple and consistent two-color palette: blue for public (Negeri) and orange for private (Swasta). This color coding is maintained across all charts in the dashboard, creating a visual language that makes the information easy to process instantly. The colors provide high contrast and are accessible.
- Labeling and Titling: Every chart has a clear, descriptive title in Indonesian, as per the source data (e.g., "Perbandingan jumlah siswa per kelas"). Key data points, like percentages, are labeled directly on the charts. This removes ambiguity and allows the audience to grasp precise figures without having to consult axes constantly.
- Interactivity: The dashboard includes filters for "Year" (Tahun) and "Status" (school type). This transforms it from a static image into an analytical tool, allowing users (such as policymakers) to explore the data dynamically and focus on specific areas of interest.

The ultimate goal of visual analytics is to derive meaningful, actionable insights. Our dashboard is designed to surface deep findings about the state of Indonesian education.

- Insight 1: Public Schools are Overburdened. The visualizations clearly show that public schools have more students per class and a significantly higher percentage of damaged classrooms compared to private schools. This suggests that public school infrastructure is under immense strain.
- Insight 2: A Critical Dropout Crisis in Public Schools. The most alarming insight is that public schools account for over 77% of student dropouts. This stark figure, presented in a simple bar chart, points to severe systemic issues that go beyond just classroom size, potentially including economic pressures on students and lack of support.
- Insight 3: A Surprising Balance in Human Resources. Despite the vast differences in facilities and outcomes, the teacher-to-student ratio is relatively balanced between public and private schools. This is a crucial, nuanced finding. It suggests that simply hiring more teachers may not solve the core problems; the issues lie more in the quality of facilities, funding, and socio-economic support for students.
- Insight 4: Disparity is Geographic. When combined with our map visualizations, the dashboard's findings are put into a geographic context. We can see that the infrastructure problems are not uniform but are heavily concentrated in specific regions, thus providing a clear target for policy intervention.

In conclusion, our approach to visualization is a methodical process of choosing the right visual tools, designing them for maximum clarity, and focusing them on uncovering deep, actionable insights that can inform policy and drive change in the Indonesian education system.

2. Explain how you create a map-based visualization from your dataset. What information did you aim to convey through the visualization?



The process of creating these maps in Tableau involves precise technical steps to ensure the data is displayed accurately

- Assigning Geographic Roles: The first step was to make sure the data fields containing locations (Provinces and Districts) were recognized by Tableau. We assign a “Geographic Role” to these columns, so that Tableau can map them.
- Generating Coordinates: Once the geographic role is assigned, Tableau automatically generates Latitude and Longitude data for each region. These coordinates are then pulled into the “Rows” and “Columns” shelves to form the base map of Indonesia.
- Encoding Data with Color: This is the core step to convey information. We took our key metric, “Number of Students Per Teacher”, and placed it on the “Color” mark. Tableau then automatically colors each region (district/city) based on the value of the student per teacher ratio in that region.
- Selecting Map Type: We chose a choropleth map, where each region is shaded with a color proportional to its data value. This map type is the best choice for showing how a measurement (in this case, student-teacher ratio) varies across the geographical landscape.

Map Design and Clarity

Our map designs are made to be easy to read, visually appealing and professional, focusing on the following principles:

- Color Scheme: We used a sequential color palette (gradation of one color), from light blue/teal to dark. This choice is very intuitive:
    - Light colors represent a low student-teacher ratio (more ideal conditions).
    - Dark colors represent high student-teacher ratios (less ideal conditions, indicating a heavier teacher load). This color scheme allows the audience to quickly identify the areas of greatest challenge (darker areas).
- Clarity and Labeling: To keep the map clean and uncluttered, we did not label each region. Instead, we utilized the interactive tooltips feature. When a user hovers over a region, an information box will pop up displaying the province/district name along with accurate student-teacher ratio data.
- Legend: A clear legend is included (standard in Tableau) to explain the meaning of each color gradation, mapping light to dark colors to specific ratio values for accurate interpretation.

The main purpose of this map is to present robust spatial analysis and generate clear valuable insights.

- Information Conveyed: This map directly answers the questions: “Where in Indonesia is the teacher load heaviest?” or ‘Where is the concentration of students per teacher highest?’. It turns abstract ratio data into a geographical story about the distribution of human resources (teachers).
- Key Insight: High Ratio is Not Just a Remote Area Problem: Unlike the facility deterioration map that shows a clear East-West divide, this student-teacher ratio map reveals a more complex insight. Areas with high ratios (dark colors) are not only concentrated in eastern Indonesia such as Papua, but are also significantly dispersed in densely populated areas in the west, such as in some areas of Java and Sumatra.
- Actionable Intelligence: This insight is invaluable to policymakers. It shows that the problem of teacher shortages is multifaceted: it is not only about the difficulty of placing teachers in remote areas, but also about the failure to meet the very high demand in population centers.

Therefore, teacher distribution policies cannot be uniform. A dual strategy is needed: one to attract and retain teachers in the 3T (frontier, outermost, disadvantaged) areas, and another to massively increase the supply of teachers in urban and densely populated areas where demand far exceeds supply. This map provides a strong basis for more strategic and data-driven teacher allocation.

3. Explain the process of selecting the dataset for your group project and why you chose that dataset?

Our group's process for selecting the dataset was methodical and directly driven by our research objectives.

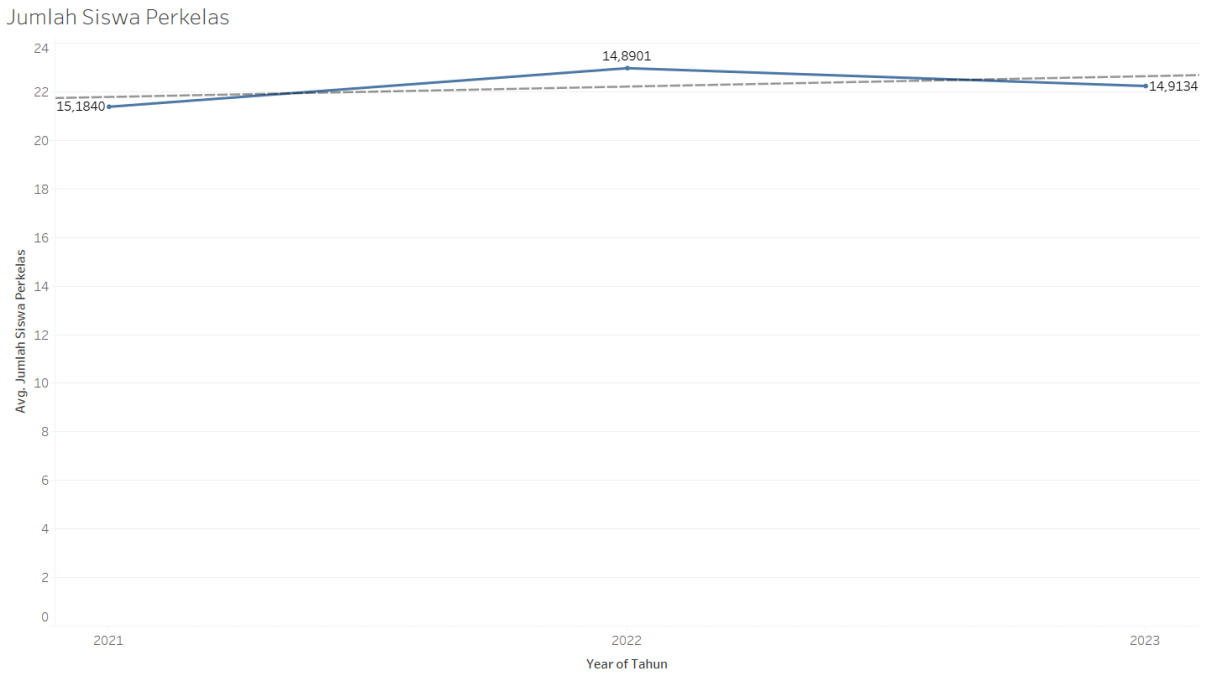
- Problem Identification: We began by identifying the core research problem: the significant gap in educational facilities and resources between provinces in Indonesia. Our initial research indicated that disparities in the number of teachers, students, and school quality were key issues to investigate.
- Defining Data Requirements: Based on our problem statement, we determined that we needed a dataset containing specific variables: the number of schools, students, and teachers, as well as data on the condition of school facilities. It was crucial that this data could be segmented by province, school type (public and private), and educational level (Primary, Junior High, Senior High) over a recent period (2021-2023) to allow for comprehensive analysis.
- Source Identification and Collection: We identified the official data portal from the Indonesian Ministry of Education, Culture, Research, and Technology (Kemendikbud) as the most credible and authoritative source. Specifically, we obtained the data from the website data.dikdasmen.go.id. Due to the large number of individual download links required to gather data for all provinces and school levels, we utilized an automated web scraping system built with Python to efficiently collect the datasets.

We chose the dataset from the Ministry of Education (Kemendikbud) for several compelling reasons:

- Direct Relevance: The primary reason for this choice was its direct relevance to our research questions. The dataset provides the exact metrics needed to analyze educational disparities, including the number of students, teachers, and schools, as well as the physical condition of facilities. This allowed us to directly investigate the gap between public and private schools and between different regions.
- Credibility and Authority: As the official data provider for the Indonesian education system, Kemendikbud offers a highly credible and reliable source of information. Using an official government dataset adds significant weight and validity to our findings and analysis.
- Comprehensive Scope: The dataset is comprehensive, covering all provinces and regencies across Indonesia. This nationwide scope was essential for our project's goal of analyzing inter-provincial disparities. Furthermore, the availability of data from 2021 to 2023 enabled us to conduct a trend analysis over a three-year period.
- Data Granularity: The dataset provided the necessary granularity, allowing us to break down the analysis by school status (public vs. private), which was a central theme of our project. This level of detail was crucial for uncovering the nuanced differences between the two systems.

4. Explain the results of the visualization you created, including the insights gained from the dashboards.

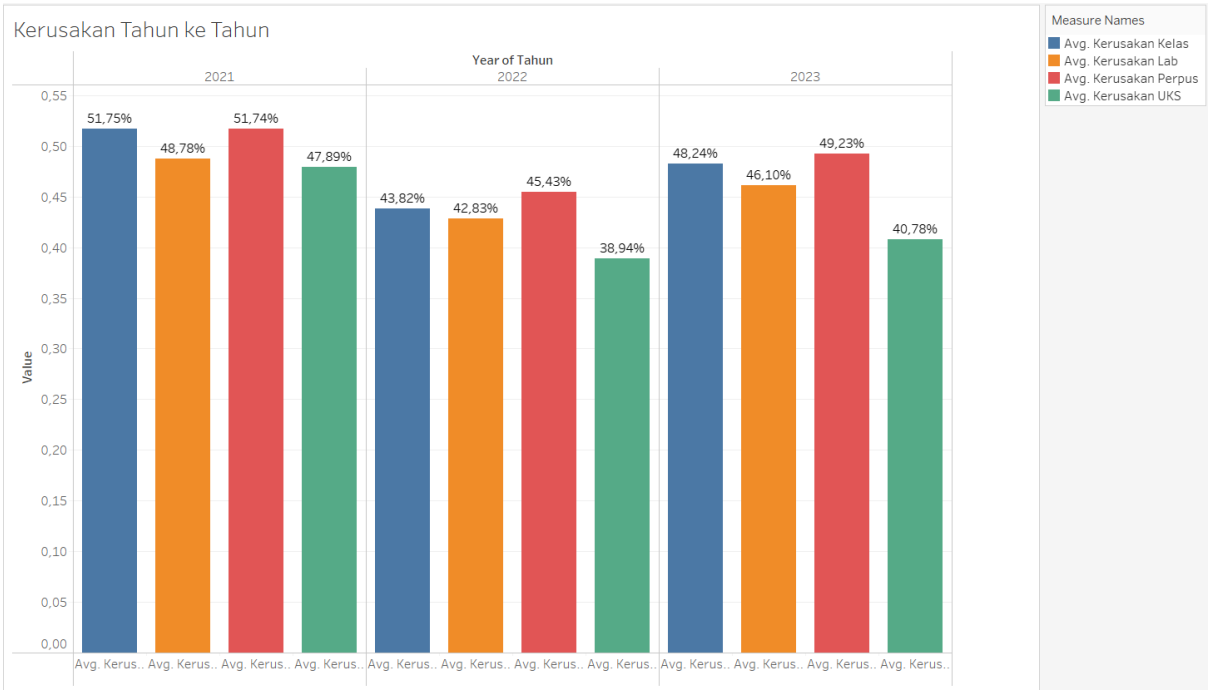
1. Insight: The Public School System is Under Greater Strain



The dashboards clearly show that the public school system operates under more significant pressure regarding student capacity and facility conditions.

- Higher Student Density: Public schools consistently have a higher number of students per class compared to their private counterparts between 2021 and 2023. Both school types saw a general increase in student numbers during this period, with 2023 marking the peak.
- Poorer Facility Conditions: The data indicates that public schools consistently suffer from a higher percentage of damaged classrooms. While there was a slight improvement in infrastructure in 2022, the condition of public school facilities deteriorated again in 2023.

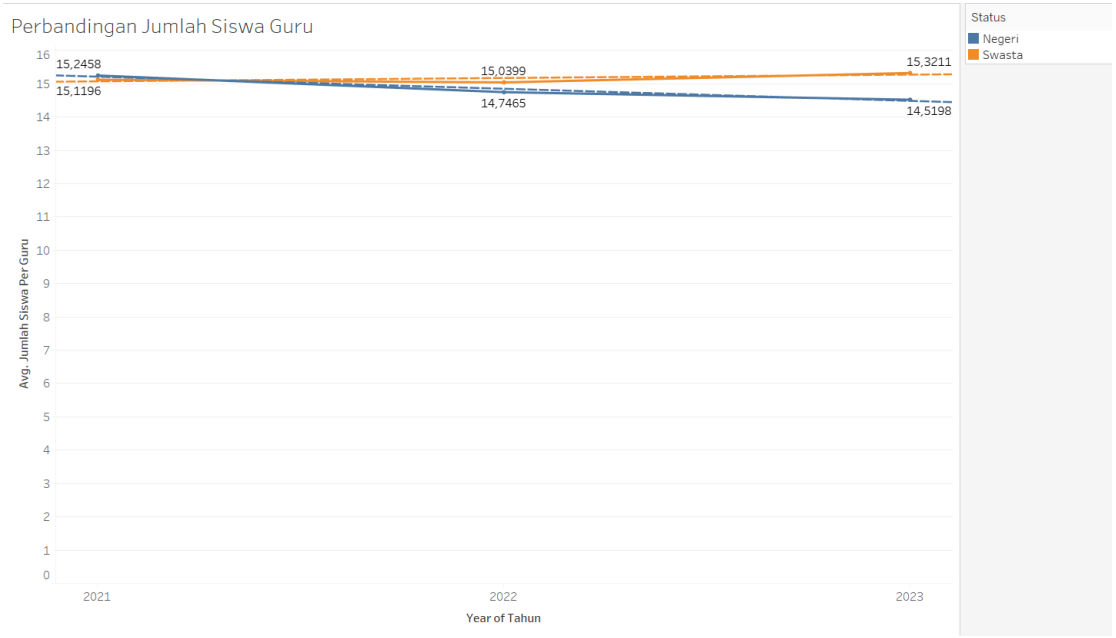
2. Insight: A Critical Disparity in Student Dropout Rates



One of the most alarming findings from the visualization is the disproportionately high rate of students leaving the public school system.

- Majority of Dropouts from Public Schools: The visualization reveals that over 70% of students who dropout of school come from the public school system.
- Conclusion: This stark figure suggests that students in public schools face more substantial systemic barriers—be it economic, social, or related to school environment—that force them to discontinue their education. It highlights an urgent need for policies focused on student retention within the public system.

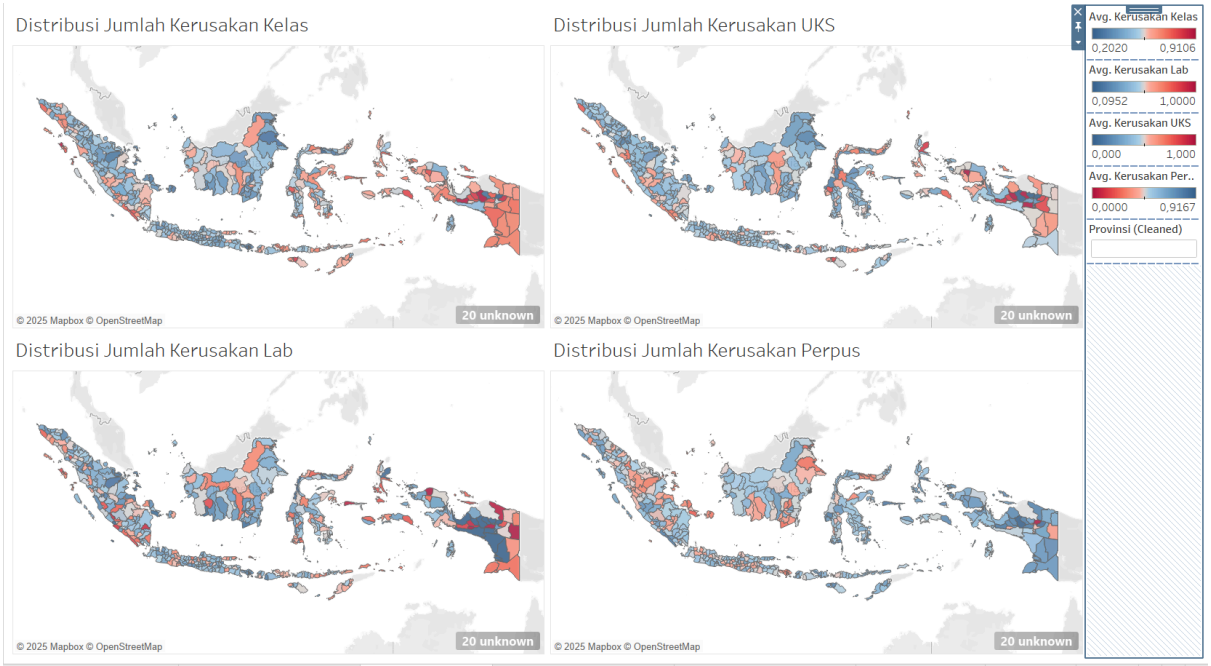
3. Insight: Teacher Ratios are Balanced, Pointing to Issues Beyond Teacher Numbers



Contrary to what one might expect, the visualization of the average number of students per teacher reveals a more nuanced insight. It shows that the disparity is not simply about public schools having a higher student load per teacher.

- Similar Ratios with a Diverging Trend: The line chart shows that both public (Negeri) and private (Swasta) schools started with a very similar student-per-teacher ratio in 2021 (around 15 students per teacher). However, over the next two years, their paths diverged.
    - Public Schools (Negeri): The ratio of students per teacher showed a consistent decrease, dropping from 15.1 to 14.5 by 2023. This indicates an improvement, meaning that on average, each teacher in a public school was responsible for fewer students.
    - Private Schools (Swasta): In contrast, the ratio in private schools remained relatively stable and even saw a slight increase, moving from 15.2 to 15.3. By 2023, private schools had a slightly higher student-to-teacher ratio than public schools.
- Conclusion: This finding challenges the simple assumption that public schools are universally understaffed compared to private ones. While public schools face significant challenges with facilities and dropout rates, this specific metric shows a positive trend toward a more manageable student load for teachers. The core issues driving educational disparity, therefore, seem to lie more heavily in the areas of infrastructure quality, resource funding, and socio-economic support for students, rather than a simple shortage of teachers in the public system. This insight directs policy focus away from just hiring more teachers and towards improving the overall quality of the learning environment.

4. Insight: Geographic Visualizations Reveal Regional Imbalances



The map-based dashboards effectively translate the data into a spatial story, highlighting that educational challenges vary significantly by location.

- Concentrated Infrastructure Damage: The map visualizing damage to school health units (UKS) shows that these problems are not uniform across the country. Regions in Eastern Indonesia, particularly Papua and its surrounding provinces, show the most severe levels of facility damage, requiring targeted government attention.
- Complex Teacher Distribution: The map of student-to-teacher ratios reveals that regions with a high number of students per teacher are not just in remote areas. These "hotspots" also exist in densely populated provinces in Western Indonesia, indicating a complex challenge of both supplying teachers to remote locations and meeting demand in populous areas.

Overall Conclusion: The dashboards successfully demonstrate that the challenges within Indonesia's education system are complex. While there is a clear divide in quality and outcomes between public and private schools, the issues are deeply intertwined with regional geography, infrastructure funding, and student support systems. These insights provide a data-driven foundation for recommending more targeted and nuanced policies.