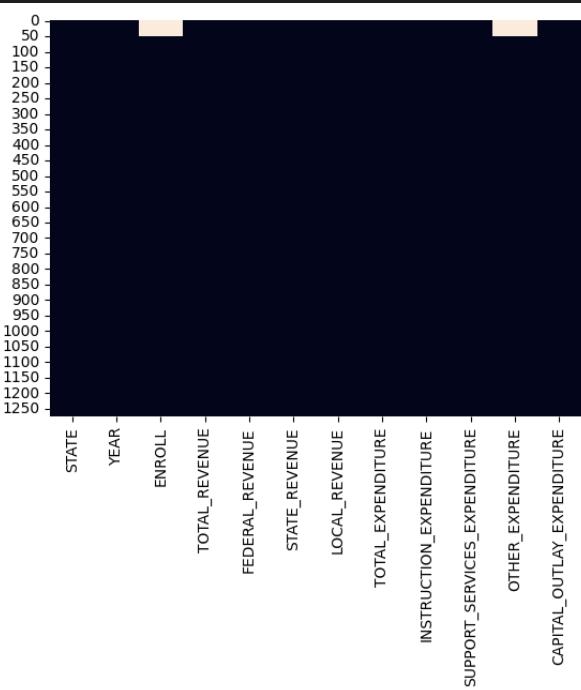
**US Student Analysis**

**Prepared by: Abdelrahman Harb**

**Prepared for: Rafi Ptashny, Chief Data Analyst**

**Date: 7 feb 2025**

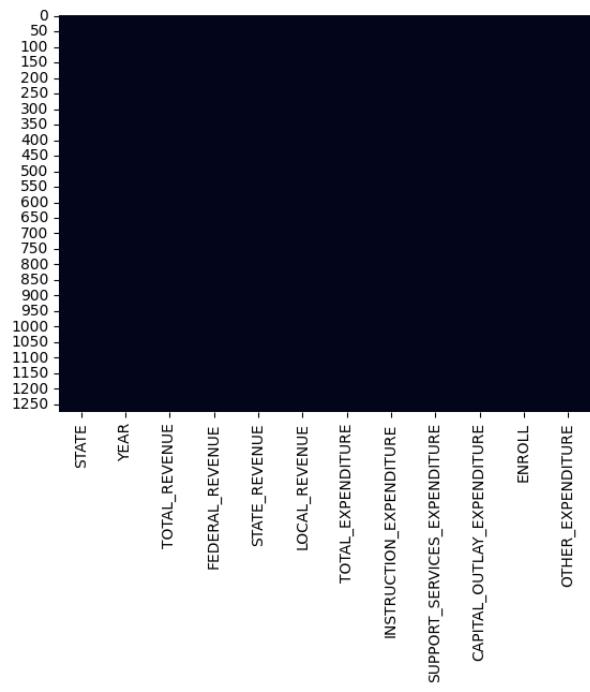
**Data Cleaning:**



I handled the missing data which was

Just numerical features with an imputer

That is known to be powerful one KNN



dropped duplicated rows and tested the effect of outliers.

**NOTE:**

I extracted new features PER\_STUDENT\_EXPENDITURE

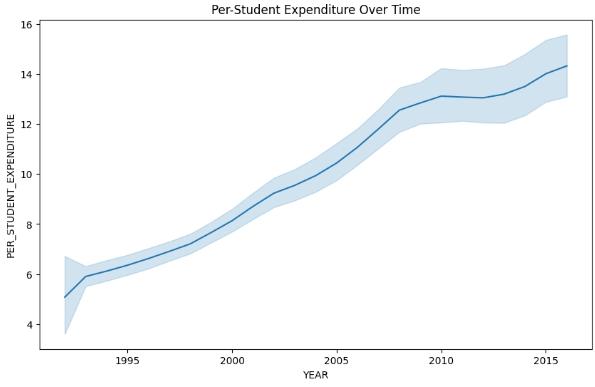
&

ENROLL\_CHANGE

For further analysis afterwards.

**Data Analysis:**

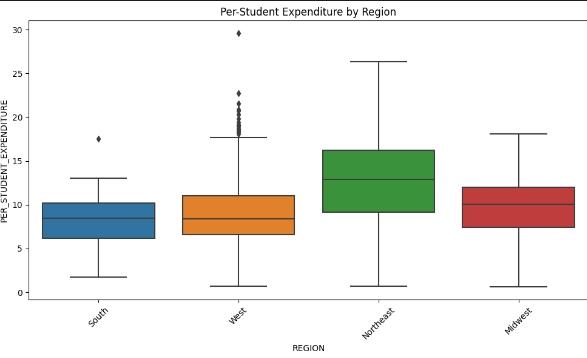
Investing and plotting trends over time i found



* **Upward Trend:** The most prominent feature is the clear upward trend in per-student expenditure over time. This indicates that, on average, the amount of money spent per student in the given context (likely US public education) has increased consistently from 1995 to 2015.
* **Steady Growth:** The increase appears relatively steady, with no dramatic spikes or drops. The line shows a gradual but consistent rise over the 20-year period.
* **Magnitude of Increase:** While the exact dollar amount isn't clear without units on the y-axis, we can observe a substantial relative increase. Expenditures seem to have more than doubled from the starting point in 1995 to the ending point in 2015.
* **Confidence Interval:** The shaded area suggests that there's some uncertainty about the exact values. The width of the shaded area gives a visual representation of this uncertainty. It's likely narrower in the earlier and later years, where there might be more data points, and possibly wider in the middle.

And Dividing State Features into geographic regions and conducting analysis examines per-student expenditure in public education across four regions of the United States: Midwest, Northeast, South, and West. The analysis is based on data that has been processed to calculate per-student expenditure and categorize states into these regions.

examines per-student expenditure in public education across four regions of the United States: Midwest, Northeast, South, and West. The analysis is based on data that has been processed to calculate per-student expenditure and categorize states into these regions i found



* **Northeast Leads in Spending:** The Northeast region exhibits the highest average per-student expenditure, at 13,022.25. This suggests a greater financial investment in education per student compared to other regions.
* **South Has Lowest Spending:** The South region demonstrates the lowest average per-student expenditure, at 8,249.50. This indicates potentially constrained resources allocated to education per student in this region.
* **Midwest and West in Mid-Range:** The Midwest and West regions fall within the mid-range for per-student spending. The Midwest has an average of 9,730.66, while the West has an average of 9,068.29.
* **Significant Regional Disparities:** The difference in per-student spending between the highest-spending (Northeast) and lowest-spending (South) regions is substantial ($4,772.75). This highlights significant financial disparities in educational investment across different regions.
* **Potential Implications:** These spending disparities could potentially lead to variations in educational resources, quality of instruction, and student outcomes across regions. Further investigation is warranted to explore the correlation between per-student expenditure and educational performance metrics.

**RECOMMENDATIONS:**

* **Resource Allocation Review:** Policymakers and educational stakeholders in lower-spending regions (especially the South) may need to review resource allocation strategies to ensure equitable educational opportunities for all students.
* **Performance Evaluation:** Evaluate educational performance metrics (e.g., standardized test scores, graduation rates) in conjunction with per-student expenditure to determine the impact of spending on student outcomes.

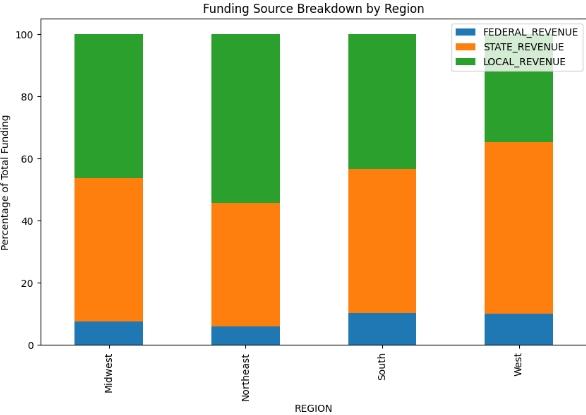
**OLS Model Results Investigation:**

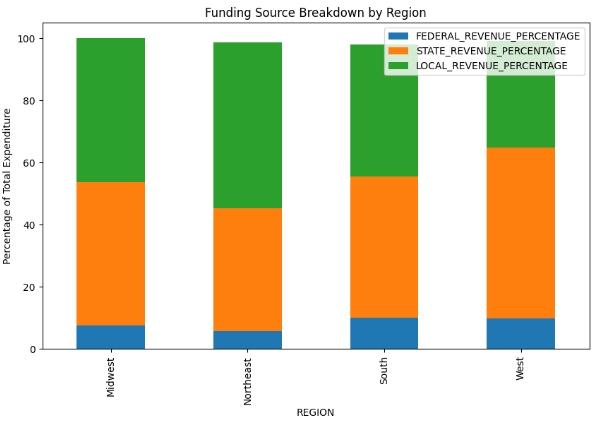
* **Highly Significant Difference:** The p-value (PR(>F)) for C(REGION) is extremely small (1.275420e-57, which is essentially 0 for practical purposes). This means there is a highly statistically significant difference in per-student expenditure between at least two of the regions. It's extremely unlikely that these differences are due to random chance.
* **Large F-statistic:** The F-statistic (99.274199) is very large, which further supports the conclusion of significant differences between region means.
* **Sum of Squares:** The sum of squares for C(REGION) (3866.406047) being larger than the residual sum of squares (16175.877745) indicates that a substantial portion of the total variability in per-student expenditure is explained by the region differences.

Interpretation :

that indicates a very strong and statistically significant difference in per-student expenditure between the regions.

**Founding Source Per Region:**



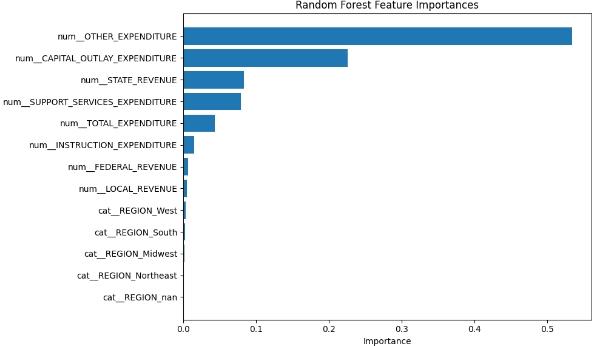


**Insights And Interpretation:**

* **Regional Variation in Funding Balance:**
  + **West:** The West region shows the highest reliance on Local funding and the lowest reliance on State funding among the four. This suggests a greater role of property taxes or other local revenue sources in funding education in the West.
  + **Northeast:** The Northeast region exhibits a relatively balanced distribution between State and Local funding, although Local funding still predominates.
  + **Midwest and South:** The Midwest and South regions demonstrate a more significant reliance on State funding compared to the West and Northeast.
* **Federal Contribution is Minor but Consistent:** Federal funding plays a relatively minor role in financing public education across all regions. The federal share is consistently the smallest segment in each bar, indicating its supplementary role.
* **Potential Implications for Equity:**
  + **Local Funding Disparities:** The heavy reliance on local funding, particularly in the West, could lead to disparities in per-student spending within states. Districts with higher property values can generate more local revenue, potentially resulting in wealthier districts having better-funded schools.
  + **State Role in Equalization:** State funding plays a crucial role in mitigating disparities caused by variations in local wealth. The greater reliance on state funding in the Midwest and South may reflect efforts to equalize funding across districts within those regions.
* **Federal Role in Targeted Support:** While the overall federal contribution is small, federal funding often targets specific needs or student populations (e.g., low-income students, special education). Its impact might be more significant in specific areas or programs.

**ML Modelling:**

I tried two kinds of models to investigate the relationship between the features and ENROLL “ the number of enrolled students to have better understanding of what features to track afterwards. And i found The Random Forest model provides better predictions of enrollment than Linear Regression. While both models show a good fit based on R-squared, the high MSE values suggest that there are likely other factors influencing enrollment that are not being captured by the current model. Further investigation, including feature importance analysis and addressing the high MSE, is necessary.



**SUMMARY**:

* **Dominant Influence of OTHER\_EXPENDITURE:** OTHER\_EXPENDITURE has by far the highest importance. This suggests that expenditures categorized as "other" (which could include various administrative, operational, or miscellaneous spending) are the strongest predictor of enrollment.
* **Significant Role of CAPITAL\_OUTLAY\_EXPENDITURE:** CAPITAL\_OUTLAY\_EXPENDITURE (spending on infrastructure, buildings, etc.) also plays a substantial role in predicting enrollment. This makes intuitive sense, as investments in facilities can influence student capacity and attractiveness of schools.
* **Moderate Influence of STATE\_REVENUE and SUPPORT\_SERVICES\_EXPENDITURE:** STATE\_REVENUE and SUPPORT\_SERVICES\_EXPENDITURE have a moderate impact. State funding and spending on support services (administration, counseling, etc.) are relevant, but less influential than the top two.
* **Lower Influence of Other Factors:** The remaining features, including instruction expenditure, federal and local revenue, and region, have relatively low importance in comparison. While they still contribute to the model's predictive power, their influence is less pronounced.

**Overall Conclusions and Recommendations:**

* Focus on Expenditure Analysis: The strong influence of expenditures, particularly OTHER\_EXPENDITURE and CAPITAL\_OUTLAY\_EXPENDITURE, suggests that a detailed breakdown and analysis of spending patterns are crucial to understanding enrollment trends. InvestInMind could focus on research that examines:
  + What constitutes "OTHER\_EXPENDITURE" and how it varies across districts or states.
  + The relationship between capital investments and enrollment growth or decline.
* Consider Non-Financial Factors: While funding sources and expenditures are important, the relatively low importance of "REGION" suggests that other non-financial factors might be at play. InvestInMind should consider exploring:
  + Demographic changes (population growth, migration patterns).
  + School quality and reputation (test scores, graduation rates, program offerings).
  + Parental choice and school choice programs.
  + Economic conditions and employment opportunities.
* Address Data Limitations: The high MSE values in the machine learning models, despite high R-squared, suggest that there might be data limitations. InvestInMind could advocate for:
  + Improved data collection and reporting on educational finances.
  + Standardized definitions and categorization of expenditures (especially "OTHER\_EXPENDITURE").
  + More granular data (e.g., at the district or school level).
* Targeted Support for Underserved Areas: Given the potential disparities in local funding, InvestInMind could focus its support on:
  + Districts with low LOCAL\_REVENUE or high poverty rates.
  + Areas where capital investments are needed to improve facilities and accommodate enrollment growth.
* Policy Advocacy: Based on the research and analysis, InvestInMind can advocate for:
  + More equitable funding formulas at the state and local levels.
  + Policies that promote efficient and effective use of educational resources.
  + Increased transparency and accountability in educational spending.
* Further Research: Conduct further research to:
  + Explore the causal relationships between expenditures and enrollment (using more advanced statistical techniques).
  + Investigate the impact of specific programs or interventions on enrollment and student outcomes.