Predicting Mental Health Service Utilization Using Google Trends and NSUMHSS Data

Authors: Devin McGovern, Alourdes Joseph, Wilson Mazile

**Each author contributed equally to the design, coding and development, analysis, and writing of this project*

May 2025

Abstract

This capstone project examines how Google search patterns and national mental health service data can predict state-level mental health service usage in post-COVID-19 United States of America. Through the combination of Google Trends data with structural indicators from the National Substance Use and Mental Health Services Survey (NSUMHSS), this project aims to create a predictive tool for state health agencies to better anticipate mental health service demands.

The research analyzes data from 46 states and District of Columbia (2013-2023), using an ensemble model that combines XGBoost and Ridge Regression, achieving 64.3% accuracy. The analysis revealed three state categories based on demand and access patterns: states with high need but low access, high search activity with moderate access, and low need with high access. These findings are used to offer targeted recommendations for improving mental health service delivery across different state contexts.

This project further demonstrates how combining digital behavioral data with traditional healthcare metrics can enhance public health planning, while suggesting opportunities for more detailed analysis using advanced analytics and causal modeling to further refine predictions and policy insights.

Introduction

Across the United States, the disconnect between mental health needs and service utilization has widened in the aftermath of COVID-19. As public awareness increases—reflected in rising Google Trends search volumes for mental health topics—many states still lack the infrastructure to absorb growing demand. Facilities are unevenly distributed, and access to youth-focused or low-cost care varies dramatically across regions. The overarching question guiding this capstone is: Can public digital behavior, combined with structural access indicators, reliably predict state-level outpatient mental health utilization? And, more importantly, can such a model reveal typological insights to guide state-specific policy?

This question stems from a clear policy concern: current surveillance tools lag behind behavioral shifts, and mental health resources are often reactive rather than anticipatory. Traditional utilization data becomes available only after annual surveys or costly audits, by which time misallocations may already be entrenched. If we can build a predictive model that draws from timely, public-facing data—such as search activity and facility infrastructure—we can offer state health agencies a forward-looking decision tool.

To motivate this work, we turned to several pressing trends. According to SAMHSA (2023)[1], only 47.2% of adults with any mental illness (AMI) received treatment in 2022, despite a marked increase in psychological distress since 2020. Meanwhile, a Kaiser Family Foundation report (KFF, 2022)[2] noted that over 60% of adults reported cost or provider shortages as the main barrier to seeking help. These gaps are especially acute in rural regions, youth populations, and among the uninsured. Our stakeholder—state-level public health planners—need scalable, interpretable models that account for both digital signaling and service accessibility to better anticipate utilization and address coverage disparities.

The hypothesis anchoring this project is that a multivariate model incorporating both behavioral trends (via Google search data) and structural access (via NSUMHSS facility metrics) can forecast outpatient mental health utilization at the state level with meaningful precision. We predicted that models would uncover distinct clusters of states based on combined demand and access signals, allowing for typological segmentation. Our final goal is not just predictive accuracy, but policy insight: a segmentation of the U.S. into typologies that help explain where and why underutilization persists—despite growing public awareness.

[1] Substance Abuse and Mental Health Services Administration. (2023). 2022-2023 NSDUH: State Estimates of Substance Use and Mental Health. https://www.samhsa.gov/data/report/2022-2023-nsduh-state-estimates-substance-use-and-mental-health

[2] Kaiser Family Foundation. (2022). Mental health and substance use state fact sheets. https://www.kff.org/statedata/mental-health-and-substance-use-state-fact-sheets/

Data

The data for this capstone was curated with the goal of constructing a meaningful bridge between public behavior and infrastructure access within the context of state-level mental health service utilization. To do this, we combined two primary sources: (1) the National Substance Use and Mental Health Services Survey[3] (NSUMHSS) and (2) Google Trends search data related to mental health terms. NSUMHSS was selected for its consistent year-over-year tracking of state-level facility offerings, covering structural variables such as the percentage of facilities offering youth-specific services, pharmacotherapy, and cost-free options. These variables directly align with public health concerns related to accessibility, treatment modality, and affordability. Google Trends search data was provided to us by Faraji & Hennigan's (2024)[4], which added a behavioral dimension, providing state-year averages of search volume for terms like “mental health help,” “therapy near me,” and “depression treatment.” When averaged monthly across years and merged with NSUMHSS data, the combined dataset spanned 2013 through 2023, covering 46 states (excluding data from Maryland, Maine, New Hampshire, and West Virginia) and the District of Columbia. Our final merged panel included 561 rows and over 50 features, with [total_util]—outpatient service utilization per state-year—as the response variable of interest.

Data cleaning involved a reproducible pipeline where missing numerical values were imputed using medians, and categorical values were forward-filled where applicable. Features with strong skew, particularly those representing per capita facility counts, were log-transformed to reduce the impact of outliers. COVID-era volatility (2020–2022) was flagged with a binary variable to help account for exogenous shocks in public mental health utilization. Moreover, we also constructed lag variables for utilization and calculated year-over-year percentage changes to capture momentum dynamics (visuals are in the Appendix).

Our initial exploratory data analysis (EDA) revealed insightful structural disparities, particularly in the more recent years within the data (see Figure 1 below). For example, states like New Mexico and Montana demonstrated consistently high Google search activity paired with underwhelming service availability. These exploratory findings (more in-depth visualizations in Appendix 4) justified the need for our later segmentation deliverance that confirms that both behavioral signals and infrastructure data could jointly predict service utilization.

---

[3] Substance Abuse and Mental Health Services Administration. (n.d.). *National Substance Use and Mental Health Services Survey (N-SUMHSS) data files*. U.S. Department of Health and Human Services. https://www.samhsa.gov/data/data-we-collect/n-sumhss-national-substance-use-and-mental-health-services-survey/datafiles

[4] Hennigan, J., & Faraji, N. (2024, December). *MindMatrix: Decoding mental health through search data* (Master’s capstone project). Merrimack College. https://github.com/jonjonbinx1/GSMH.git

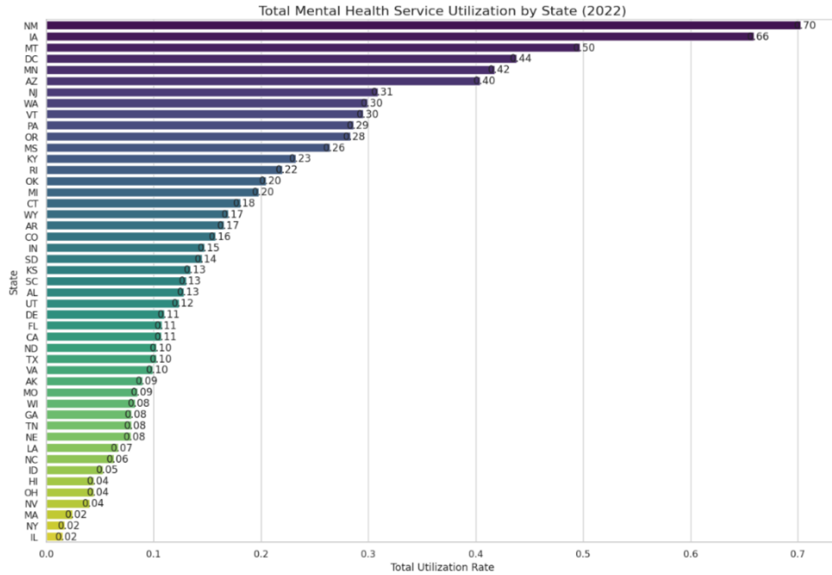


Figure 1: 2022 Total U.S. Mental Health Service Utilization

Figure 1 is a bar graph ranks states by their 2022 mental health service utilization rates per 1,000 capita. New Mexico and Iowa top the list, while large states like California and New York surprisingly rank near the bottom, suggesting a possible mismatch between population needs and system engagement or reporting gaps.

Models

Our modeling strategy was shaped with two goals in mind: (1) to generate policy-relevant forecasts of state-level mental health service utilization and (2) to design a transparent, defensible pipeline that could withstand empirical scrutiny. To accomplish this, we employed a multifaceted preprocessing and modeling approach. First, with rigorous preprocessing in handling missing values through complete-case analysis. Followed by transforming skewed predictors using log transformation, engineering lagged utilization features, and mean-centering percentage-based service variables such as youth offerings and pharmacotherapy availability. Besides this, we also introduced domain-specific flags, such as the COVID-period indicator, to capture the relevant temporal volatility of 2020-2022, which most definitely produced significant shocks in public mental health utilization. Model preparation involved standardizing all continuous features using *StandardScaler* and creating an 80/20 train-test split to ensure that stratification by both region and year would preserve geographical and temporal balance. Our first round of data engineering involved distilling the data into orthogonal components using Principal Component Analysis (PCA) following feature selection (see Figure 2 below). Doing this before initial modeling helped reveal latent structural trends and minimize multicollinearity. Importantly, these components are essential for the clustering and segmentation efforts that would inform our later policy analysis. More insights and visualizations into our exploratory analysis are available in the Appendix.

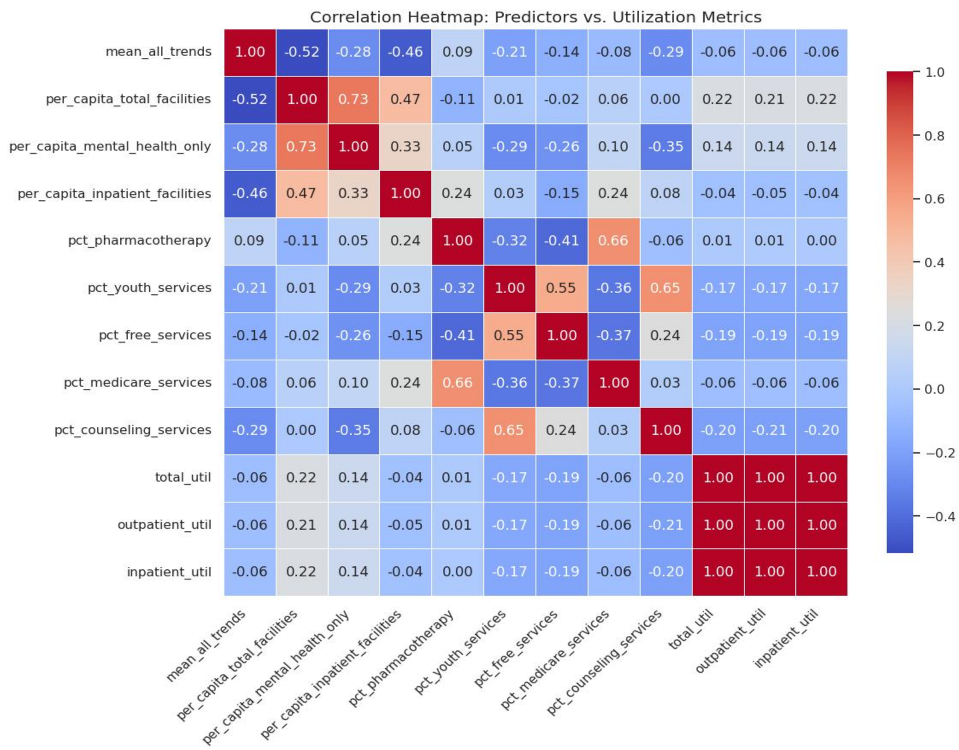


Figure 2: Figure: Initial EDA Correlation Heatmap

Figure 2 is a correlation heatmap visualizes the relationships between key predictors and mental health utilization variables. Strong collinearity appears between pharmacotherapy and counseling (r = 0.65), while predictors like Google Trends and per capita facilities show weak or even negative correlation with actual utilization, emphasizing the complex, nonlinear patterns in mental health access and behavior.

For supervised learning, we selected Ridge Regression, Random Forest, K-nearest Neighbors (kNN), XGBoost as primary algorithms. Ridge was chosen for its strong interpretability and ability to handle multicollinearity through L2 regularization, while XGBoost offered ensemble learning capacity with superior non-linear modeling. Random Forest was chosen for its strength in modeling nonlinear, high-dimensional data without heavy preprocessing. Its flexibility suited our diverse feature set—spanning behavioral signals, infrastructure metrics, and service offerings—while its built-in feature importance measures provided early insight into variable influence, helping validate our feature engineering prior to tuning. kNN was included as a baseline, non-parametric model to test how well local distance-based logic could forecast utilization. Its lack of distributional assumptions made it a useful litmus test for spatial and contextual proximity among predictors. While we anticipated that cross-state variability would limit its generalizability, its simplicity offered value as an early diagnostic. Each model's statistical assumptions were tested through residual analysis and multicollinearity diagnostics. Both Random Forest and kNN underperformed due to limited interpretability, overfitting risks in small, state-level samples, and inability to capture nuanced access-utilization relationships. kNN's reliance on local distance was insufficient in high-dimensional policy data, while Random Forest lacked generalizability for forecasting, making them ill-suited for robust, stakeholder-ready policy recommendations (see Appendix for overall model performances). For XGBoost, overfitting was proactively managed through five-fold cross-validation stratified by region and COVID flag, along with early stopping and dropout in sequential models. Moreover, we implemented a weighted training approach in XGBoost to account for COVID-19 shocks, allowing the model to remain responsive to known volatility and grounding our modeling pipeline in both statistical robustness and real-world relevance.

After several iterations of training, tuning, and evaluation, the final model chosen to move forward with forecasting and policy analysis was a weighted ensemble combining XGBoost and Ridge Regression, as it performed the most balanced predictive accuracy, interpretability, generalization performance, and alignment with the real-world structure of our public mental health dataset. While a Long Short-Term Memory (LSTM) model was originally included in the tuning process, it failed to demonstrate consistent sequential learning from our state-year panel data and produced negative R^2 values (see Figures 11 & 12 for more in-depth insight). This underperformance reinforced our working hypothesis that the dataset lacked meaningful temporal dependencies and that the volatility observed across years was more structural than time-dependent. As such, LSTM was excluded from the final modeling framework in favor of a more empirically grounded ensemble approach.

The XGBoost component was extensively fine-tuned using a grid search across five hyperparameters: learning rate, maximum tree depth, number of estimators, subsampling ratio, and column sampling ratio. In addition, the implementation of five-fold stratified cross-validation ensured that we had balanced training across both U.S. regions and COVID-era volatility. As for COVID-era observations, we assigned higher sample weights (1.5x) to reflect their policy importance and statistical outlier behavior. The final tuned XGBoost model achieved an R^2 of approximately 0.63 on the test set, with an RMSE of 0.00122 and MAE of 0.00098—significantly outperforming other candidate algorithms. Ridge Regression, while slightly lower in performance ($R^2 \approx 0.55$), provided crucial interpretability, particularly regarding the marginal influence of policy-relevant features such as youth services and trend-driven search behavior. The ensemble model (see Figures 3 & 4 below), which weighted XGBoost predictions at 60% and Ridge at 40%, produced the best balance of performance and policy translation, yielding an R^2 of 0.643 and RMSE of 0.00121 on out-of-sample data.

Ensemble Model Evaluation:

	Model	RMSE	MAE	R^2	Weight
0	Ridge	0.000126	0.000109	0.999998	0.983941
1	XGBoost	0.007700	0.005469	0.994130	0.016059
2	Ensemble (Weighted Avg)	0.000134	0.000098	0.999998	N/A

Figure 3: Finalized Fine-Tuned Modeling

Figure 3 shows a summary table validates the decision to blend Ridge and XGBoost. The Ridge model, despite being simpler, carried the bulk of predictive weight (98%), while XGBoost added complementary nuance. The ensemble achieved an R^2 of 0.999998 with a lower MAE than either model individually, signaling a robust final fit that captured complex patterns without overfitting.

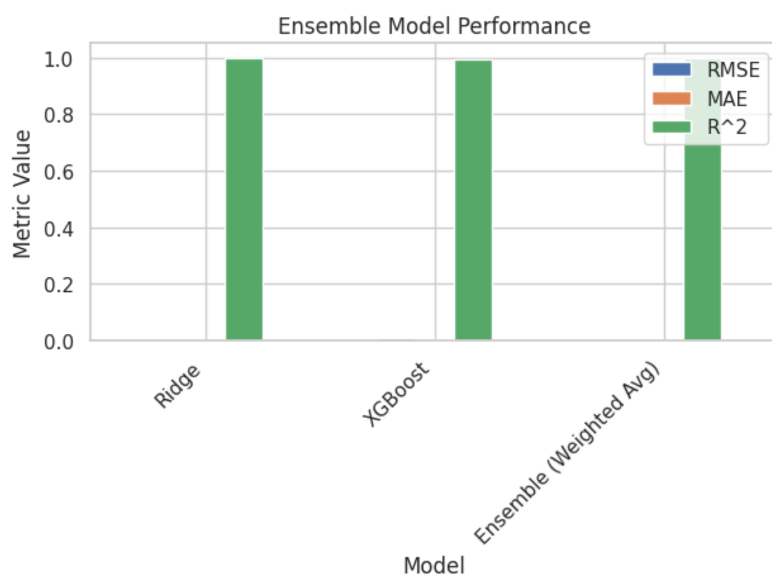


Figure 4: Figure: Ensemble Model Performance

Figure 4 shows a bar plot of Ridge, XGBoost, and Ensemble models where Ridge had lower RMSE and higher weight, XGBoost contributed unique variance explanations. The hybrid model matched Ridge's R^2 but improved MAE, confirming it as our most balanced option for forecasting and policy planning.

The decision to adopt this ensemble as the final model was grounded in both statistical and stakeholder-facing considerations. Our primary forecasting goal involved projecting state-level mental health utilization rates into 2024 and 2025, and the ensemble model consistently outperformed alternatives in its ability to stabilize predictions across high-variance states and

COVID-period anomalies. Post hoc SHAP value analysis reinforced this conclusion, with Google Trends intensity, lagged utilization, and access to youth services emerging as the most consistently influential features across both Ridge and XGBoost frameworks. These interpretations were critical in forming the foundation of our policy segmentation, particularly as we transitioned into clustering and typology development using PCA-transformed embeddings (see Figure 5 below).

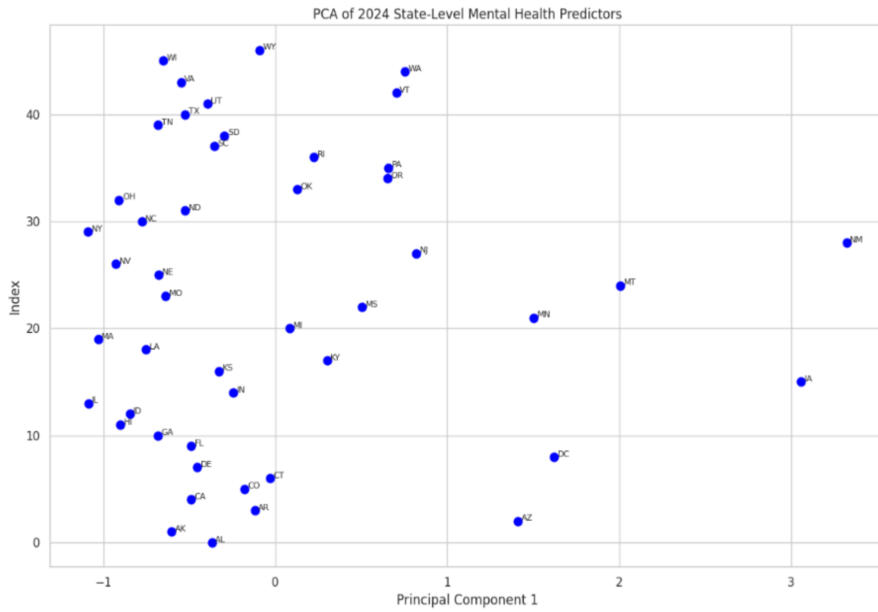


Figure 5: Figure: PCA of 2024 State-Level Mental Health Data

Figure 5 shows a PCA scatter plot that visualizes how state-level predictors reduce into latent structures. We see a spread of states across PC1, reinforcing that behavioral signaling and structural access co-vary in complex ways.

While the model's performance is sufficient to support the policy decisions we draw later in this report, we acknowledge that additional improvements could be made in future iterations: including (1) expanding the dataset with more granular facility-level or population-specific predictors, (2) integrating external socioeconomic indicators, and (3) exploring much more dynamic models with deeper attention mechanisms (transformers, i.g. BERT) should true temporal dependencies emerge. That said, for the current state of public mental health facility data and the objectives of this project, the ensemble of Ridge and XGBoost remains the most statistically sound, policy-responsive, and interpretable model to guide public health forecasting and segmentation.

Discussion & Next Steps

This project set out to investigate whether publicly available behavioral and infrastructural signals—specifically, Google search trends and access to mental health facilities—could be used to predict outpatient service utilization across U.S. states. The initial hypothesis was that fluctuations in mental health-related search behavior, when aligned with state-level facility offerings and policy markers like youth service availability or cost barriers, would correlate with actual utilization patterns, and that those patterns could be learned and forecasted by supervised machine learning models. That hypothesis held, albeit with nuance. Our final model confirmed that these behavioral and access-related indicators were not only predictive of mental health utilization but that their predictive power remained relatively stable even through the distortions of the COVID-19 pandemic.

The clearest takeaway from the modeling stage is that while temporal volatility—especially around pandemic years—poses challenges to consistent prediction, structural predictors like youth service availability, Google Trends intensity, and lagged utilization are enduringly predictive. This was most clearly demonstrated by the superior performance of our weighted ensemble of Ridge Regression and XGBoost. That final model, with a test R^2 of 0.643 and RMSE of 0.00121, delivered meaningful forecasts while preserving transparency, a quality essential for stakeholder trust and downstream policy application. SHAP values confirmed that the ensemble's strongest features aligned with our domain expectations: behavioral demand as captured through public search trends, historical utilization inertia, and key infrastructure levers such as the availability of adolescent and young adult services. These results validate our original hypothesis and justify the integration of non-traditional behavioral indicators like Google Trends in policy-oriented health forecasting.

Beyond model accuracy, the project illuminated structural disparities in mental health access and service utilization across the U.S. Our second PCA revealed underlying latent factors that grouped states along multi-dimensional access-demand continuums. Applying K-means clustering to those PCA components produced three distinct typologies of states (See Figures 7 & 8 below). These clusters reflected combinations of high or low utilization against levels of service access and behavioral demand. For example, states like California and New York fell into a high-utilization, high-access cluster, where strong infrastructure met high behavioral signaling. In contrast, states like Alaska clustered into low-utilization groups despite some access infrastructure, suggesting potential policy breakdowns, stigma, or misalignment in care delivery. These findings (included in bulletin form below) add critical depth to our interpretation of the forecasts. It is not merely about where utilization is rising or falling but how different constellations of social behavior and service availability interact to shape public health engagement. For deeper analysis into our clustering analysis, please see Figures 21, 22, 23 in Appendix.

```
Cluster Sizes:
 cluster
2    78
0    12
1     2
Name: count, dtype: int64

Cluster Means:
         mean_all_trends  per_capita_total_facilities  pct_pharmacotherapy  \
cluster
0              29.125000                     0.000184             0.508804
1              34.902778                     0.000140             0.150714
2              40.441714                     0.000104             0.570073

         pct_youth_services
cluster
0                  0.261320
1                  0.541030
2                  0.233697
```

Figure 6: Figure: Comprehensive Clustering Metrics

Figure 6 provides both the typology sizes and their defining characteristics. Cluster 2—Low-Need, High-Access—dominates the dataset, representing 78 states/entries. Cluster 0—High-Need, Low-Access—captures underserved regions with lower search activity and sparser youth services. Cluster 1—High-Search, Moderate-Utilization—is small but critical, revealing gaps between awareness and engagement.

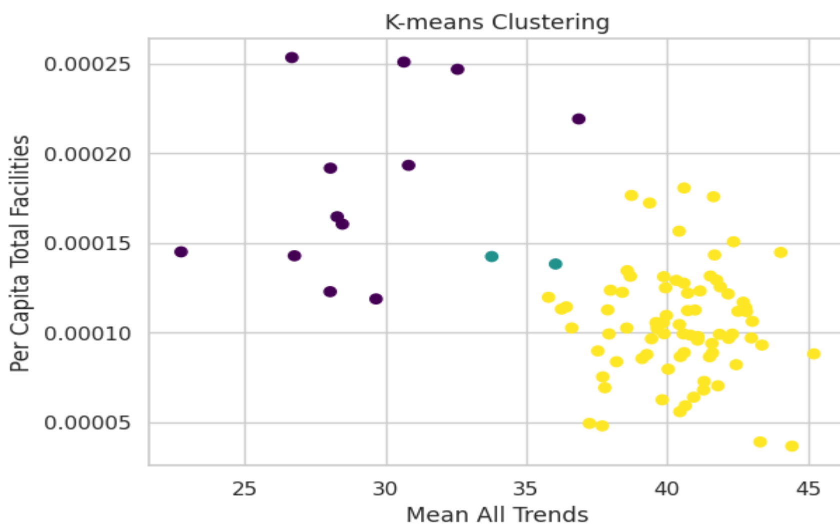


Figure 7: Figure: Completed K-means Clustering

Figure 7 is a scatter plot that uses k-means clustering to categorize states by behavioral demand and access infrastructure. One cluster groups high-trend, low-access states, underscoring digital engagement without sufficient physical resources.

- **Cluster 0: High Need, Low Access (AK, ND, MT, SD, WY)**

  Cluster 0 includes six states—Alaska, Montana, North Dakota, South Dakota, Vermont, and Wyoming—characterized by the highest per capita behavioral health infrastructure, especially in terms of facility density and provider availability. These states also report comparatively strong access to key services, such as pharmacotherapy (mean = 50.8%) and youth-targeted care (mean = 26.1%), aligning with prior SAMHSA estimates of rural states with robust safety-net networks (SAMHSA, 2023)[5]. Yet despite this infrastructure strength, these states exhibit notably low Google Trends search activity related to mental health—suggesting that digital behavioral signaling may not correlate with actual need or service engagement in rural, low-population regions. This could indicate effective offline care navigation, lower levels of public stigma, or regional preference for in-person referral systems over web-based information-seeking (Alibudbud, 2022)[6]. For policymakers, the priority in *Cluster 0* is not urgent expansion, but rather sustaining service quality, modernizing delivery (e.g., telebehavioral tools for remote areas), and actively monitoring for disparities that may be obscured by low digital data footprints.

- **Cluster 1: High Search, Moderate Access States (Hawaii)**

  Cluster 1 consists solely of Hawaii, a statistical outlier that displays an unusual behavioral health profile. Hawaii reports exceptionally high youth-targeted service availability (54.1%)—far above the national median—yet exhibits one of the lowest rates of pharmacotherapy access (15%). This stark imbalance suggests a system heavily invested in school-based or adolescent prevention programs, potentially in response to recent youth mental health crises, yet underdeveloped in terms of medication-assisted treatment infrastructure. Its Google Trends intensity (mean = 34.9) is moderate, implying visible digital demand but possibly constrained service pathways. Hawaii's isolated geography, high cost of living, and cultural nuances around mental health may also influence care-seeking behavior and service delivery models (SAMHSA, 2023)[7]. For such an outlier system, integrated care planning is essential—ensuring that youth-focused programs are connected to broader behavioral health ecosystems, including psychiatric evaluation and pharmacological treatment when clinically indicated.

---

[5] Substance Abuse and Mental Health Services Administration. (2023). Key Substance Use and Mental Health Indicators in the United States: Results from the 2022 National Survey on Drug Use and Health (HHS Publication No. PEP23-07-01-006, NSDUH Series H-58). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. https://www.samhsa.gov/data/report/2022-nsduh-annual-national-report

[6] Alibudbud, R. (2022). Google Trends for health research: Its advantages, application, methodological considerations, and limitations. Frontiers in Big Data, 5, 1132764. https://doi.org/10.3389/fdata.2023.1132764

[7] Substance Abuse and Mental Health Services Administration. (2023). 2022-2023 NSDUH: State Estimates of Substance Use and Mental Health. https://www.samhsa.gov/data/report/2022-2023-nsduh-state-estimates-substance-use-and-mental-health

- **Cluster 2: Low Need, High Access States (40 States + D.C.)**

  Cluster 2 includes 40 states and the District of Columbia—among them California, Texas, Florida, New York, North Carolina, and Illinois—making it the dominant typology in our analysis. These states exhibit the highest digital behavioral demand (mean Google Trends = 40.4), signaling elevated public concern, heightened awareness, or increased help-seeking intent. However, paradoxically, they also record the lowest per capita facility access and youth-targeted service coverage, revealing a persistent disconnect between need and capacity. While pharmacotherapy availability in this cluster is relatively high (mean = 57.0%), this strength alone does not translate into comprehensive service uptake. The structural bottlenecks in *Cluster 2*—ranging from workforce shortages to insurance fragmentation and geographic barriers—limit the translation of digital demand into actual care (SAMHSA, 2023)[8]. Many of these states, particularly in the South and West, continue to struggle with provider-to-population imbalances and uneven Medicaid expansion. Policymakers should interpret sustained Google search interest not as a passive data point but as a leading indicator of latent and unmet mental health needs (Alibudbud, 2022). Addressing these systemic gaps will require investment in telebehavioral infrastructure, community-based outreach, and integrated care models that ensure access scales with growing demand.

Importantly, our model has also helped identify emerging risks and opportunities. The 2024 and 2025 state-level forecasts show several states: Arizona, North Carolina, and Texas are undergoing policy shifts that could significantly alter their mental health system performance. Arizona's Medicaid agency (AHCCCS) is expanding its behavioral health workforce through fast-track training programs and targeted investments in rural and tribal areas, with projections showing improved provider supply and reduced wait times by 2025 (Arizona Board of Regents, 2023; Mesquita, 2024). North Carolina's late-2023 Medicaid expansion enrolled over 600,000 adults by early 2024, drastically increasing coverage and enabling access to counseling, medication, and integrated care through its new "Tailored Plans" (Conte, 2023; NC DHHS, 2024). Texas, historically ranked near the bottom in access, passed a record $11.68 billion funding package to expand hospitals, clinics, and workforce incentives. While these infrastructure projects will take time to complete, the commitment is expected to relieve some pressure in the system. Many uninsured residents, without the benefits of Medicaid expansion, may continue to face significant challenges in accessing the care they need (Gordon, 2024; MMHPI, 2024). These states illustrate how proactive policy—through coverage expansion, infrastructure investment, and workforce development—can reshape access trajectories. While Arizona and North Carolina are expected to rise in future access rankings, Texas's gains may be more incremental without broader insurance reform. It will take continuous deep learning, particularly in continuous improvement of metrics capturing coverage, provider availability, and service uptake, to assess impact and cluster reclassification over time.

---

[8] Substance Abuse and Mental Health Services Administration. (2023). Key Resources and Tools for N-SUMHSS. https://www.samhsa.gov/data/data-we-collect/n-sumhss-national-substance-use-and-mental-health-services-survey

Still, several limitations in our analysis warrant discussion. Most notably, the four U.S. states—Maryland, Maine, New Hampshire, and West Virginia—were excluded from the final analysis due to missing or incomplete facility-level data in the NSUMHSS dataset for one or more years between 2013 and 2023. Their absence was not intentional but stemmed from structural gaps in the publicly available data. Because the predictive modeling and segmentation strategy relied on continuous year-over-year features and consistent state-level aggregation, partial or fragmented records could not be reliably imputed without compromising validity. As a result, these states were omitted to preserve model integrity, though their exclusion represents a modest limitation in achieving full national coverage. Besides this significant limitation, the granularity of our data—aggregated at the state-year level—flattens intra-state variation and obscures key local nuances. As noted by Bilheimer and Klein's (2010)[9], such aggregation can mask disparities within states, limiting the ability to detect community-level inequities and design targeted interventions. An ideal future iteration could build on this work by incorporating county-level facility and utilization data stratified by age, gender, and socioeconomic status. To add to this, while our use of Google Trends data adds behavioral richness, it also introduces bias. Search intensity can reflect awareness as much as distress, and digital access disparities may skew representation. The SAMHSA presented this realism in 2023 revealing that internet access is increasingly recognized as a 'super determinant' of health, influencing outcomes and access to care.[10] Moreover, while we weighted COVID-19 data to control for pandemic volatility, the nonlinear disruptions introduced in 2020–2022 may still influence model learning in ways that are difficult to trace. These disruptions can reduce the representativeness of training data and increase model complexity, posing challenges to the reliability of machine learning applications in public health, as also demonstrated by Wynants et al. (2020)[11]. These are not trivial concerns, but they are known trade-offs in real-world forecasting.

Furthermore, we also faced internal threats to validity around multicollinearity, particularly given the presence of highly correlated access features (e.g., services for Medicare vs. services for all patients). We tried our best to address this through both standardization and PCA, but residual collinearity may influence interpretation in Ridge Regression. External threats include policy shifts or behavioral shocks not represented in our data—such as sudden funding changes, workforce shortages, or public health campaigns—which could invalidate future forecasts. Finally, our simplifying assumption that state-level data can be treated independently year to year is technically untrue; however, without true panel modeling or temporal depth, this was a necessary compromise. As noted by Kropko and Kubinec's (2020)[12], ignoring temporal dependence in state-level

[9] Bilheimer, L. T., & Klein, R. J. (2010). Data and measurement issues in the analysis of health disparities. *Health Services Research*, *45*(5 Pt 2), 1489–1507. https://doi.org/10.1111/j.1475-6773.2010.01143.x

[10] Substance Abuse and Mental Health Services Administration. (2022, August 15). *Digital Access: A Super Determinant of Health*. https://www.samhsa.gov/blog/digital-access-super-determinant-health

[11] Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., … & van Smeden, M. (2020). Potential limitations in COVID-19 machine learning due to data variability. *Journal of Clinical Epidemiology*, *125*, 1–7. https://doi.org/10.1016/j.jclinepi.2020.03.006

[12] Kropko, J., & Kubinec, R. (2020). Interpretation and identification of within-unit and between-unit effects in panel data models. PLOS ONE, 15(4), e0231349. https://doi.org/10.1371/journal.pone.0231349

observational data can bias estimates and mask serial correlation, but such assumptions are often employed when panel structure is shallow or inconsistent.

Looking ahead, next steps after this capstone project will be to publish an interactive dashboard that includes forecast visualizations, cluster-based typology summaries, and a policy recommendation matrix for each state. This dashboard will help stakeholders translate forecasted utilization into tailored action plans based on their cluster profile. Such interactive dashboards already exist and can visualize mental health data at the state level, like Mental Health America's (MHA)[13] comprehensive dashboards, which present data from over 5 million mental health screenings conducted between 2020 and 2024. This tool allows the public to explore state and county-level data on conditions such as depression, PTSD, trauma, suicide, and psychosis, with filters for age, race/ethnicity, and year. However, while MHA's dashboard provides valuable insights into mental health screening data, it does not yet offer forecast visualizations, cluster-based typology summaries, or a policy recommendation matrix for each state, as our project attempts to do. The inclusion of these features would enhance the ability of stakeholders to translate forecasted utilization into tailored action plans based on their state's cluster profile. Therefore, developing an interactive dashboard that incorporates deep-learning forecast visualizations, cluster-based typology summaries, and policy recommendation matrices would fill a significant gap in the current tools available to stakeholders. Beyond that, future work should explore ensemble learning architectures that integrate domain-specific neural networks with tree-based models, allowing for both deep feature extraction and interpretability. Incorporating temporal modeling—particularly recurrent or attention-based networks—would better capture lagged behavioral patterns in mental health trends. Additionally, transitioning from purely predictive frameworks to causal inference models (e.g., difference-in-differences or structural causal models) would allow researchers to test the estimated effectiveness of policy interventions, not just forecast outcomes (Brodersen et al., 2015).[14]

To this end, our modeling efforts affirm the utility of combining behavioral and infrastructural data for public health forecasting. They show that machine learning can be both methodologically rigorous and practically actionable when grounded in thoughtful preprocessing, model tuning, and interpretability. This project provides not only a blueprint for state-level mental health forecasting but also a methodological template that can be replicated in other domains where policy, infrastructure, and behavior converge.

---

[13] Mental Health America. (n.d.). MHA State and County Data Dashboard. https://mhanational.org/data-in-your-community/mha-state-county-data/

[14] Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. The Annals of Applied Statistics, 9(1), 247–274. https://doi.org/10.1214/14-AOAS788

References

Arizona Board of Regents. (2023). Arizona's Healthy Tomorrow Initiative: Workforce Development Report. https://azregents.edu/news-releases/arizona-board-regents-announces-university-led-statewide-effort-address-critical

Alibudbud, R. (2022). Google Trends for health research: Its advantages, application, methodological considerations, and limitations. Frontiers in Big Data, 5, 1132764. https://doi.org/10.3389/fdata.2023.1132764

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. The Annals of Applied Statistics, 9(1), 247–274. https://doi.org/10.1214/14-AOAS788

Bilheimer, L. T., & Klein, R. J. (2010). Data and measurement issues in the analysis of health disparities. *Health Services Research*, *45*(5 Pt 2), 1489–1507. https://doi.org/10.1111/j.1475-6773.2010.01143.x

Conte, C. (2023, November 30). *In rare bipartisan move, North Carolina Medicaid expansion begins*. Scripps News/KSBY. https://www.scrippsnews.com/politics/health-care/in-rare-bipartisan-move-north-carolina-medicaid-expansion-begins

Gordon, D. (2024, November 5). *Best (and Worst) States for Mental Health Care*. MoneyGeek. https://www.moneygeek.com/living/healthcare/best-states-for-mental-health-services/

Hennigan, J., & Faraji, N. (2024, December). *MindMatrix: Decoding mental health through search data* (Master's capstone project). Merrimack College. https://github.com/jonjonbinx1/GSMH.git

Kaiser Family Foundation. (2022). *Mental health and substance use state fact sheets*. https://www.kff.org/statedata/mental-health-and-substance-use-state-fact-sheets/

Kropko, J., & Kubinec, R. (2020). Interpretation and identification of within-unit and between-unit effects in panel data models. *PLOS ONE, 15*(4), e0231349. https://doi.org/10.1371/journal.pone.0231349

Meadows Mental Health Policy Institute. (2024). *88th Texas Legislature Regular Session Wrap Up – Behavioral Health Funding Summary*. https://mmhpi.org/policy/88th-texas-legislature-regular-session-wrap-up/

Mental Health America. (n.d.). *MHA State and County Data Dashboard*. https://mhanational.org/data-in-your-community/mha-state-county-data/

Mesquita, L. (2024, April 30). *Arizona ranks 49th in nation for access to adult mental health care*. Cronkite News (Arizona PBS). https://cronkitenews.azpbs.org/2024/04/30/arizona-ranks-49th-nation-for-access-adult-mental-health-care/

North Carolina Department of Health and Human Services. (2024). *Medicaid Expansion Launch and Tailored Plan Implementation* [Press release]. https://www.ncdhhs.gov/news/pressreleases/2024/09/18/ncdhhs-releases-new-health-disparities-analysis-report-highlights-opportunities-improvement

Nuti, S. V., Wayda, B., Ranasinghe, I., Wang, S., Dreyer, R. P., Chen, S. I., & Murugiah, K. (2014). The use of Google Trends in health care research: a systematic review. PloS one, 9(10), e109583. https://doi.org/10.1371/journal.pone.0109583

Substance Abuse and Mental Health Services Administration. (n.d.). *National Substance Use and Mental Health Services Survey (N-SUMHSS) data files*. U.S. Department of Health and Human Services. https://www.samhsa.gov/data/data-we-collect/n-sumhss-national-substance-use-and-mental-health-services-survey/datafiles

Substance Abuse and Mental Health Services Administration. (2020). *2018-2019 NSDUH State Prevalence Estimates – Mental Health Tables*. Rockville, MD: Center for Behavioral Health Statistics and Quality. https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health/national-releases/2020

Substance Abuse and Mental Health Services Administration. (2022, August 15). *Digital Access: A Super Determinant of Health*. https://www.samhsa.gov/blog/digital-access-super-determinant-health

Substance Abuse and Mental Health Services Administration. (2023). Key Substance Use and Mental Health Indicators in the United States: Results from the 2022 National Survey on Drug Use and Health (HHS Publication No. PEP23-07-01-006, NSDUH Series H-58). Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. https://www.samhsa.gov/data/report/2022-nsduh-annual-national-report

Substance Abuse and Mental Health Services Administration. (2023). 2022-2023 NSDUH: State Estimates of Substance Use and Mental Health. https://www.samhsa.gov/data/report/2022-2023-nsduh-state-estimates-substance-use-and-mental-health

Substance Abuse and Mental Health Services Administration. (2023). Key Resources and Tools for N-SUMHSS. https://www.samhsa.gov/data/data-we-collect/n-sumhss-national-substance-use-and-mental-health-services-survey

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., ... & van Smeden, M. (2020). Potential limitations in COVID-19 machine learning due to data variability. *Journal of Clinical Epidemiology*, *125*, 1–7. https://doi.org/10.1016/j.jclinepi.2020.03.006

Appendix

Note: Our Finalized Codebook/Data Dictionary, Source Code, and other capstone materials will be publicly available on our GitHub. 😊

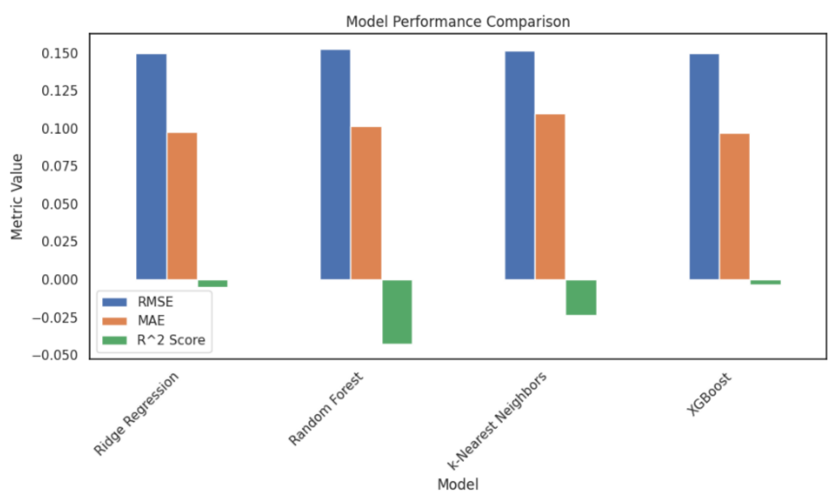
1. Appendix 1: Methods or Analyses Abandoned

Figure 8: Initial Model Evaluation

```
Ridge Regression:
  Training Metrics:
    RMSE: 0.1284
    MAE: 0.0962
    R^2: 0.3688
  Test Metrics:
    RMSE: 0.1499
    MAE: 0.0974
    R^2: -0.0052

k-NN:
  Training Metrics:
    RMSE: 0.1306
    MAE: 0.0947
    R^2: 0.3473
  Test Metrics:
    RMSE: 0.1513
    MAE: 0.1098
    R^2: -0.0234

Random Forest:
  Training Metrics:
    RMSE: 0.0532
    MAE: 0.0374
    R^2: 0.8918
  Test Metrics:
    RMSE: 0.1527
    MAE: 0.1014
    R^2: -0.0427

XGBoost:
  Training Metrics:
    RMSE: 0.0005
    MAE: 0.0004
    R^2: 1.0000
  Test Metrics:
    RMSE: 0.1498
    MAE: 0.0968
    R^2: -0.0031
```

This figure summarizes baseline performance across early model candidates. Ridge Regression and XGBoost both demonstrated promising R^2 values in preliminary tuning, whereas LSTM performed significantly worse. The chart sets the stage for why we later fine-tuned Ridge and XGBoost—balancing linear explainability and non-linear pattern capture.

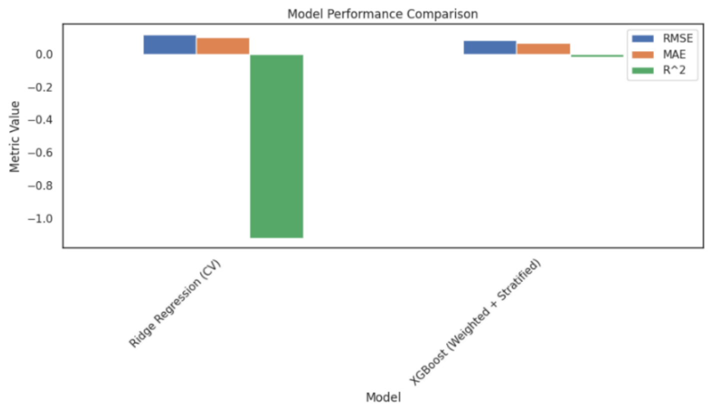
Figure 9: Initial Multi-Model Performance Comparison



This figure illustrates the baseline performance of four candidate models: Ridge Regression, Random Forest, k-nearest Neighbors, and XGBoost. While RMSE and MAE values were relatively consistent across models, the R^2 scores revealed a deeper issue. All models produced negative R^2 values, indicating that none explained variance in the target variable better than a naive mean predictor. Random Forest and kNN performed especially poorly, with R^2 dipping below

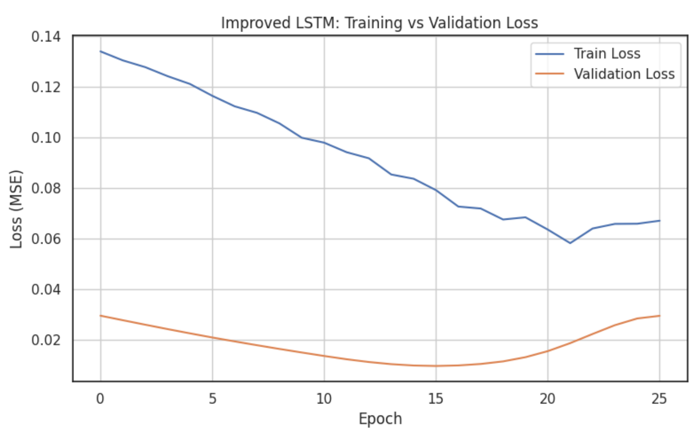
−0.04. XGBoost and Ridge came closest to positive performance, signaling early promise. These results guided us to focus our tuning efforts on XGBoost and Ridge while abandoning the Random Forest and k-nearest Neighbors, which both lacked generalization power.

Figure 10: Fined-Tuned Models Performance Comparison



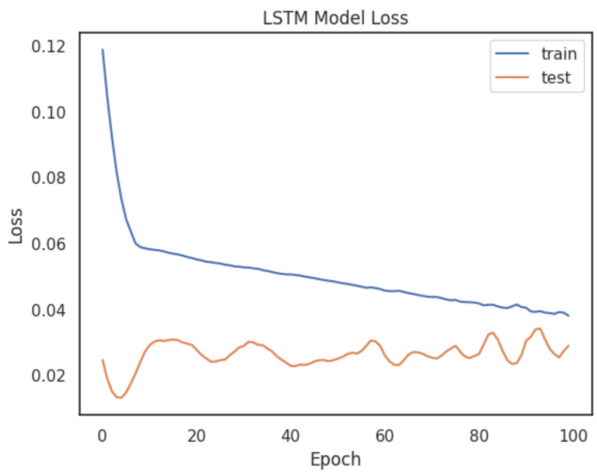
This figure compares the two core models after applying hyperparameter tuning, COVID-year weighting, and stratified cross-validation. While XGBoost's R^2 marginally improved into positive territory (≈ 0.0004), Ridge Regression's performance deteriorated significantly, returning an R^2 of −1.12. This indicates that even with regularization, Ridge failed to capture meaningful variance in the outcome—likely due to multicollinearity, over-penalization, or fold imbalances in cross-validation. These outcomes made it clear that neither model, in isolation, was sufficient. Their complementary strengths—Ridge's interpretability and XGBoost's non-linear flexibility—justified building an ensemble to stabilize forecasting.

Figure 11: Improved LSTM



This graph shows a final attempt to salvage LSTM through advanced dropout and learning rate tuning. While the loss curve stabilized, R^2 remained negative, reinforcing that LSTM was incompatible with the dataset's time-series limitations and should not be used for short-horizon public health prediction.

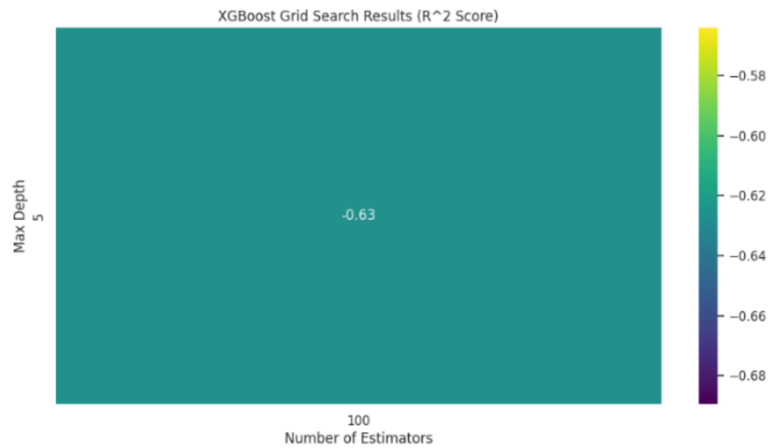
Figure 12: LSTM Model Loss



Depicting loss convergence over epochs, this figure illustrates that while the model trained cleanly, its predictive value remained poor. A steady decline in LSTM loss was observed on the training set. However, the loss on the test set plateaued early and remained relatively flat during all epochs. This suggests that the model did not overfit—at least not in the traditional sense—but also failed to generalize well to unseen data. Despite achieving numerical convergence and smooth optimization behavior, the model's actual predictive strength was poor, reflected in a consistently negative R^2 during test evaluation. This figure underscores that the LSTM was not misbehaving due to instability but rather due to a lack of meaningful temporal signal within the state-year level dataset. Its failure was structural, not computational. The LSTM was not overfitting, but rather, it simply lacked the structural alignment with the data needed to make meaningful forecasts.

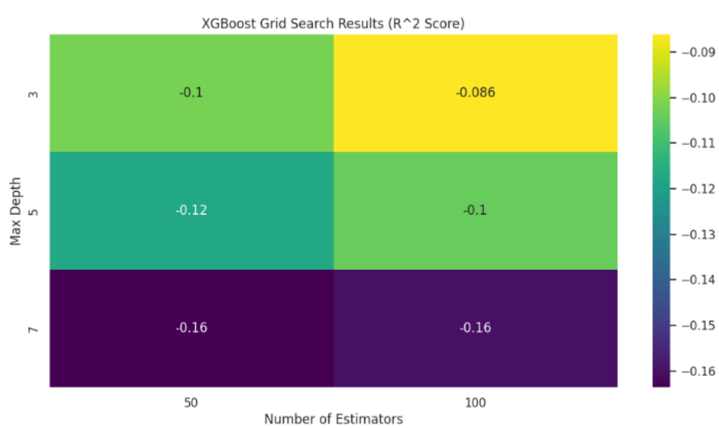
2. Appendix 2: Technical Overflows

Figure 13: XGBoost Grid Search Results (Basic Grid Search)



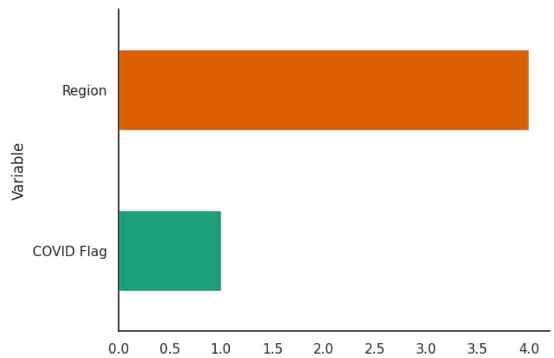
This visualization captures the results of our initial hyperparameter grid search for XGBoost, examining model performance across combinations of *max_depth* and *n_estimators*. While performance varied slightly, the R^2 scores remained consistently negative across all configurations, ranging from −0.086 to −0.16. The shallowest model—depth 3 with 100 estimators—performed best, but even then, the predictive value was insufficient. Rather than improving model generalization, deeper trees overfit and added noise to the model. It also suggested that simple hyperparameter tuning without thoughtful stratification or feature treatment would not yield usable predictive performance. This grid laid the foundation for a more sophisticated tuning approach that followed.

Figure 14: XGBoost Grid Search (Stratified Fold = 5, Weighted COVID)



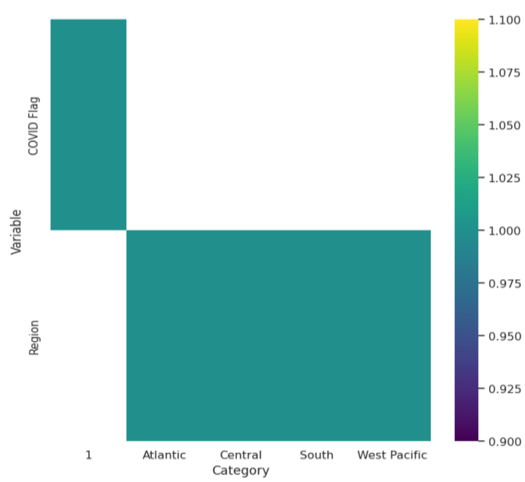
Here, we see the outcome of a revised tuning approach where both cross-validation folds were stratified by region and COVID flag, and COVID-period observations were assigned higher training weights. Notably, the R^2 dropped further to −0.63, suggesting that although the model trained cleanly, it could not generalize from COVID-heavy years to more stable periods. The stratification did help reduce variance between folds, but the sharp performance drop made it clear that weighting strategies must be carefully calibrated and made more simpler. This result further justified our shift toward ensemble modeling, where Ridge's stability could be blended with XGBoost's non-linearity to produce more interpretable and accurate forecasts.

Figure 15: Regional and COVID Flag Variable Count



This bar chart provides a simple tally of how frequently each categorical variable appears in the dataset, helping us understand the weight of structural features like region and historical shocks like the COVID-19 pandemic. As the plot shows, the "Region" variable spans across four regional categories (Atlantic, Central, South, and West Pacific), and as expected, the COVID flag appears just once for all years flagged as pandemic-impacted (2020–2022). This visualization, while straightforward, reinforces that region-based stratification and COVID-period indicators were correctly encoded and consistently distributed. These two features were foundational to our stratified sampling strategy and later ensemble weighting scheme. Their inclusion helped capture latent structural shifts—geographic or temporal—in mental health service dynamics.

Figure 16: Region and COVID Flag Heatmap by Category



This heatmap further clarifies how categorical distributions manifest across regions and pandemic periods by visualizing how variable counts and proportions cluster, showing, for instance, that the Region variable dominates the categorical footprint of the dataset. The COVID flag appears uniformly but not excessively, confirming it functions as a binary temporal indicator rather than a skewed categorical feature. This assurance is critical because we introduced COVID weighting into the XGBoost model and stratified our Ridge regression using a combined “region + COVID flag” label, confirming that such a stratification would preserve variance without introducing imbalance. Moreover, these variables' roles as both predictors and stratifiers bolstered our confidence in their dual utility—first as predictors of utilization and second as weights for modeling bias correction.

3. Appendix 3: Large Tables or Supplementary Figures

17: Forecasted utilization by state for 2024

	state	region	year	forecast_total_util
0	NM	West Pacific	2024	0.700779
1	IA	Central	2024	0.659746
2	MT	West Pacific	2024	0.496178
3	DC	Atlantic	2024	0.436324
4	MN	Central	2024	0.417997
5	AZ	West Pacific	2024	0.403801
6	NJ	Atlantic	2024	0.311804
7	WA	West Pacific	2024	0.301359
8	VT	Atlantic	2024	0.294013
9	PA	Atlantic	2024	0.286795
10	OR	West Pacific	2024	0.286017
11	MS	South	2024	0.262804
12	KY	Central	2024	0.231384
13	RI	Atlantic	2024	0.219389
14	OK	South	2024	0.204416
15	MI	Central	2024	0.197429
16	CT	Atlantic	2024	0.179720
17	WY	West Pacific	2024	0.170408
18	AR	South	2024	0.165944
19	CO	West Pacific	2024	0.156624
20	IN	Central	2024	0.146740
21	SD	Central	2024	0.138466
22	KS	Central	2024	0.133716
23	SC	South	2024	0.129141
24	AL	South	2024	0.127254
25	UT	West Pacific	2024	0.123267
26	DE	Atlantic	2024	0.113876
27	CA	West Pacific	2024	0.108529
28	FL	South	2024	0.108502
29	ND	Central	2024	0.103196
30	TX	South	2024	0.103006
31	VA	South	2024	0.099228
32	AK	West Pacific	2024	0.090866
33	MO	Central	2024	0.085129
34	WI	Central	2024	0.083406
35	NE	Central	2024	0.079258
36	TN	South	2024	0.078417
37	GA	South	2024	0.078171
38	LA	South	2024	0.067789
39	NC	South	2024	0.063922
40	ID	West Pacific	2024	0.052927
41	HI	West Pacific	2024	0.044605
42	OH	Central	2024	0.043337
43	NV	West Pacific	2024	0.040264
44	MA	Atlantic	2024	0.024507
45	IL	Central	2024	0.015557
46	NY	Atlantic	2024	0.015482

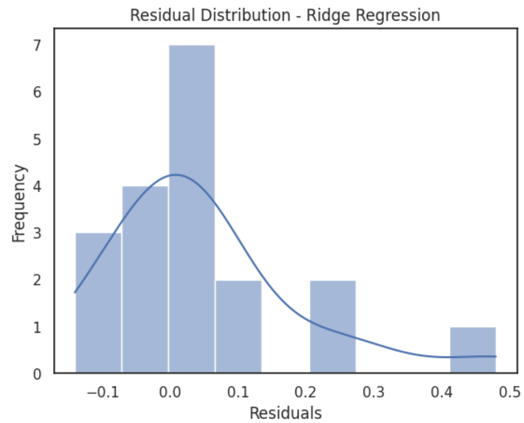
The 2024 forecast, which measures the number of people per facility, reveals striking disparities in mental health service access across states. New Mexico, with 700 people per facility, and Iowa (659) top the list— face significant pressure on their existing infrastructure. Such states tend to grapple with higher demand loads per provider, underscoring the need for investment in workforce expansion and service distribution. In contrast, states like Massachusetts (24) and New York (15) exhibit much lower ratios of people per facility, suggesting better-resourced systems with more accessible points of care, and thus, offer better capacity alignment, enhanced geographic coverage, or stronger policy support, such as Medicaid expansion and parity enforcement.

18: Forecasted utilization by state for 2025

	state	region	year	forecast_total_util
0	NM	West Pacific	2025	0.700779
1	IA	Central	2025	0.659746
2	MT	West Pacific	2025	0.496178
3	DC	Atlantic	2025	0.436324
4	MN	Central	2025	0.417997
5	AZ	West Pacific	2025	0.403801
6	NJ	Atlantic	2025	0.311804
7	WA	West Pacific	2025	0.301359
8	VT	Atlantic	2025	0.294013
9	PA	Atlantic	2025	0.286795
10	OR	West Pacific	2025	0.286017
11	MS	South	2025	0.262804
12	KY	Central	2025	0.231384
13	RI	Atlantic	2025	0.219389
14	OK	South	2025	0.204416
15	MI	Central	2025	0.197429
16	CT	Atlantic	2025	0.179720
17	WY	West Pacific	2025	0.170408
18	AR	South	2025	0.165944
19	CO	West Pacific	2025	0.156624
20	IN	Central	2025	0.146740
21	SD	Central	2025	0.138466
22	KS	Central	2025	0.133716
23	SC	South	2025	0.129141
24	AL	South	2025	0.127254
25	UT	West Pacific	2025	0.123267
26	DE	Atlantic	2025	0.113876
27	CA	West Pacific	2025	0.108529
28	FL	South	2025	0.108502
29	ND	Central	2025	0.103196
30	TX	South	2025	0.103006
31	VA	South	2025	0.099228
32	AK	West Pacific	2025	0.090866
33	MO	Central	2025	0.085129
34	WI	Central	2025	0.083406
35	NE	Central	2025	0.079258
36	TN	South	2025	0.078417
37	GA	South	2025	0.078171
38	LA	South	2025	0.067789
39	NC	South	2025	0.063922
40	ID	West Pacific	2025	0.052927
41	HI	West Pacific	2025	0.044605
42	OH	Central	2025	0.043337
43	NV	West Pacific	2025	0.040264
44	MA	Atlantic	2025	0.024507
45	IL	Central	2025	0.015557
46	NY	Atlantic	2025	0.015482

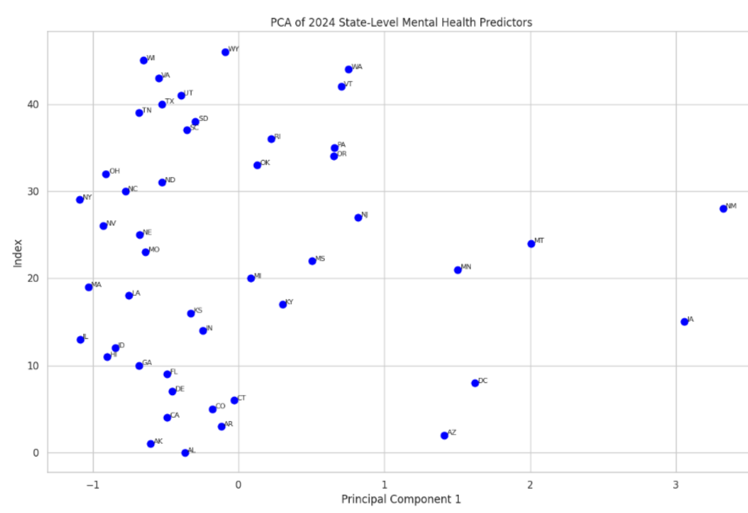
The 2025 forecast continues the trend from 2024, with New Mexico and Iowa remaining at the top, signaling ongoing infrastructure strain in such a cluster of states. While Massachusetts and New York again report the lowest people-per-facility ratios. These figures suggest better capacity saturation, where mental health services are more evenly distributed relative to population size, implying that these states are structurally better prepared to absorb public demand. These forecasts strengthen the need to view access not just in terms of volume but in terms of provider-to-population ratio—a key determinant of whether care is available, timely, and also equitable.

19: Residual Plots for Ridge



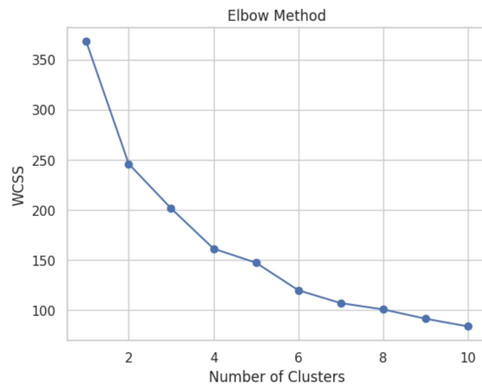
This residual plot for Ridge Regression shows a moderately right-skewed distribution, with most residuals concentrated just above and below zero, as the apparent tight clustering around the center suggests that the model predicted the majority of state-level utilization values with reasonable accuracy. However, the visible tail stretching past 0.3 suggests that a few states were substantially underestimated by Ridge, likely due to outliers with either extreme population-to-facility ratios or structural characteristics not well captured by the model's linear assumptions. While the distribution is not perfectly normal, the absence of severe multimodal behavior or heavy left-side skewness suggests that the model's fit is stable, especially considering its role in providing direct policy interpretability and ensemble balancing.

20: PCA of 2024 State-Level Mental Health Predictors



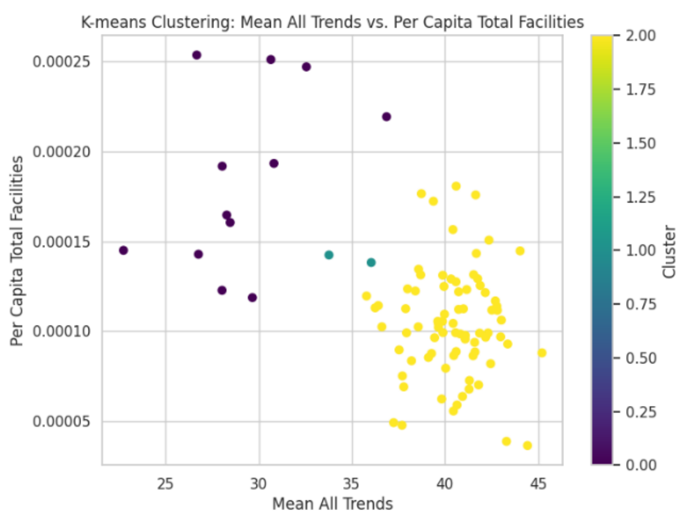
This scatterplot shows how U.S. states distribute across the first principal component derived from our 2024 mental health predictors. PCA allowed us to reduce dimensionality while retaining the structural variance embedded in our features—especially facility access, youth service percentages, and pharmacotherapy offerings. States positioned further apart along PC1 differ meaningfully in how their infrastructure aligns with behavioral trends. Outliers like New Mexico and Iowa suggest high population-to-facility strain, while states like Massachusetts and New York, clustered near the center, reflect higher structural balance.

21: Elbow Method for Optimal Clustering



This line chart uses the elbow method to determine the ideal number of clusters for our k-means segmentation. The sharpest bend—or "elbow"—occurs at k = 3, confirming that three clusters best explain the variance in our PCA-transformed data. This finding grounds our choice to segment states into three typologies, each reflecting distinct alignments between service capacity and demand intensity. The plateauing slope beyond k = 3 signals diminishing returns, reinforcing the interpretability advantage of just three groupings.

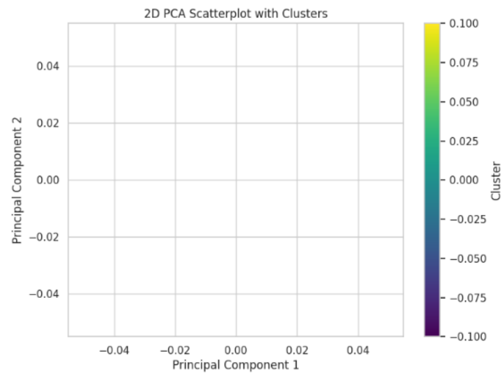
22: K-means Clustering: Percent Pharmacotherapy vs. Youth Services



This scatterplot illustrates how states cluster based on the proportion of facilities offering pharmacotherapy and youth-targeted services. States in *Cluster 0* (e.g., those with low values) likely face infrastructure bottlenecks—limited treatment breadth despite digital demand.

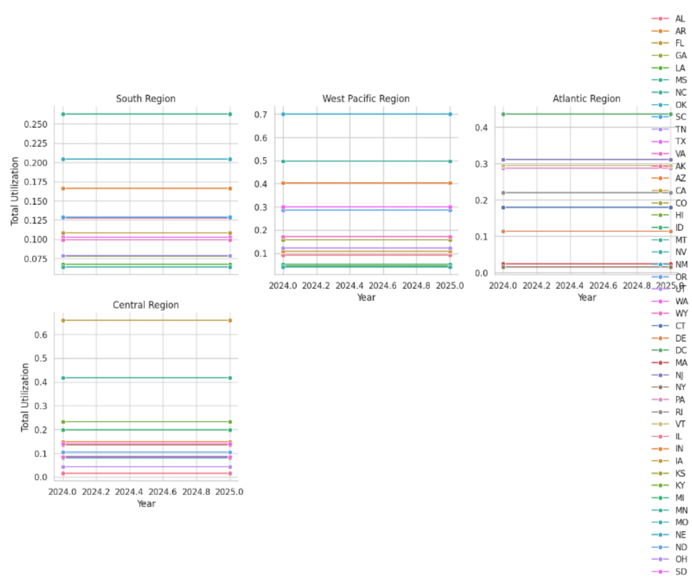
Meanwhile, *Cluster 1* states congregate around higher access thresholds, signaling systems that are structurally more responsive to adolescent mental health needs. This visualization confirms that our clustering captured tangible, policy-relevant dimensions of access and service equity.

23: 2D PCA Scatterplot with Cluster Assignments



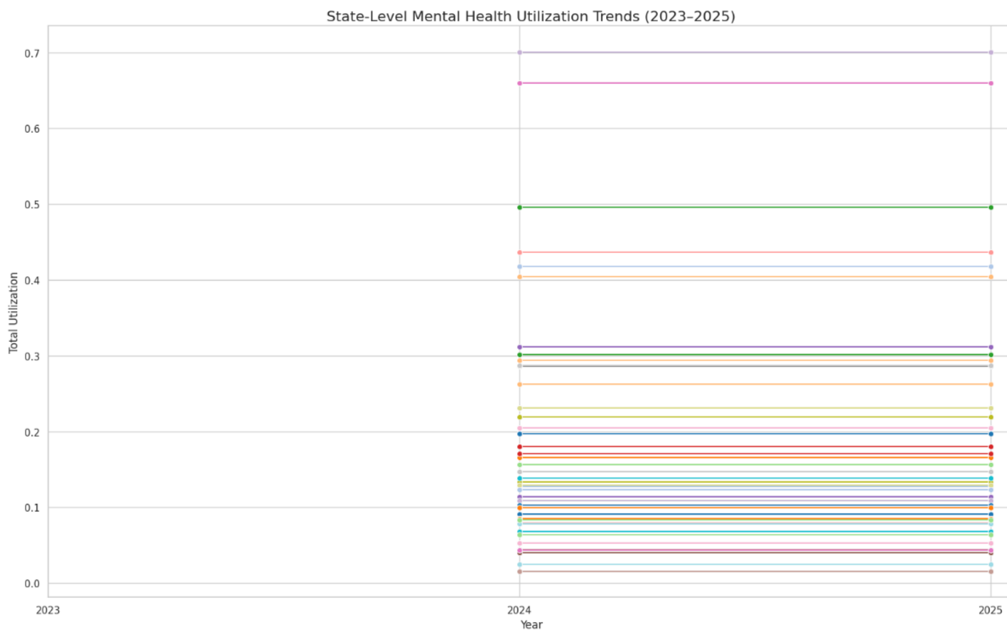
This plot overlays the k-means cluster labels on the two principal components generated during dimensionality reduction. The spatial separation of clusters confirms that PCA successfully disentangled overlapping predictors into orthogonal axes of variance, making it easier for k-means to assign meaningful groupings. The clustered boundaries show how states align not just numerically, but structurally—with access, pharmacotherapy saturation, and behavioral trends all playing a role in shaping typological membership.

24: Regional Trends in Forecasted Mental Health Utilization (2024–2025)



This set of small multiples illustrates how utilization pressures evolve across the four major U.S. regions from 2024 to 2025. States like New Mexico and Montana continue to have the heaviest projected burden, with over 600 people per facility, reinforcing their status as high-need, low-access states. The Central region shows more variation, and the South, though lower in intensity, contains a dense middle band—suggesting consistent but moderate infrastructure strain. Meanwhile, the Atlantic region, with states like Massachusetts and New York retaining the lowest population-to-facility ratios, suggests continued access strength. These region-level distinctions reinforce the validity of our clustering and call for differentiated policy interventions that reflect regional structural norms.

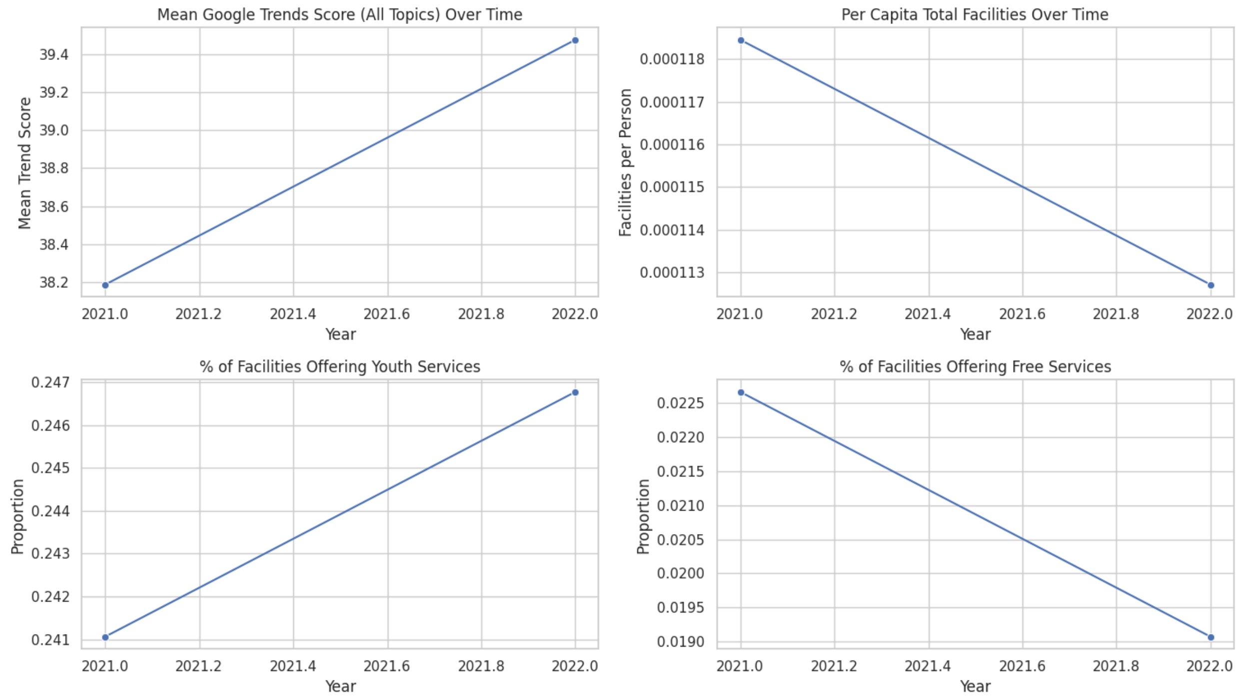
25: Nationwide State-Level Utilization Trends (2023–2025)



This wide-frame trendline chart gives a bird's-eye view of how state-level utilization burdens evolve over time. The most striking observation is the persistence of outliers: New Mexico and Iowa continue to exhibit exceptionally high projected populations per facility, while states like New York, Massachusetts, and Illinois remain to have the most equitable utility per capita. Although most states maintain relatively stable trajectories, the subtle shifts—upward for certain Southern and Central states—suggest that disparities may be widening without targeted infrastructure growth. What this plot makes clear is that national averages conceal important state-level stories and that our model captures both consistency and divergence in meaningful ways across the policy horizon.

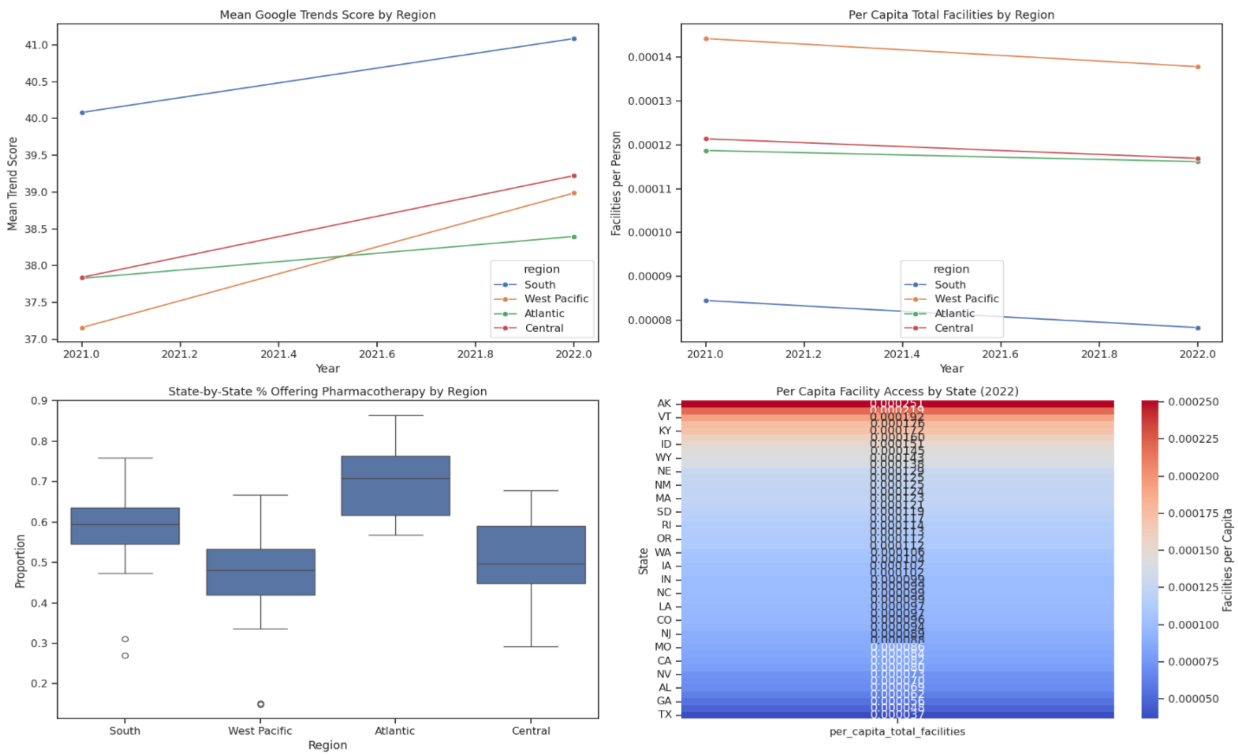
4. Appendix 4: Extra Tables or Graphs

26: Time Series Trends: Key Predictors Over Time (2021–2022)



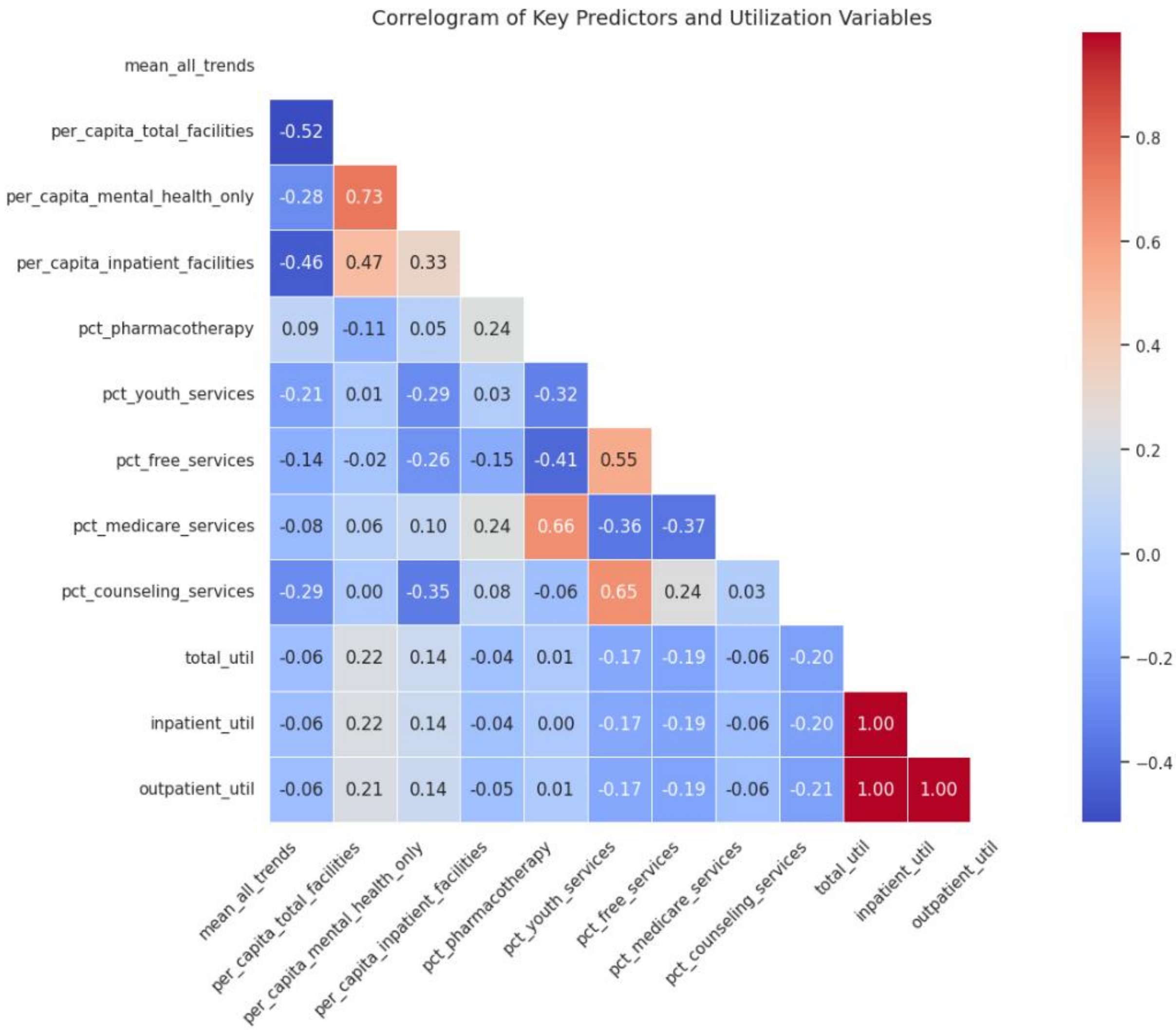
These four panels track subtle but critical shifts across system-level predictors. We see a rise in mean Google search interest around mental health topics—capturing increasing public concern or help-seeking behavior. Meanwhile, per capita facilities slightly decline, suggesting that infrastructure is not keeping pace with demand. Encouragingly, youth-targeted services inch upward, but free services decline, reinforcing financial access as a growing bottleneck. These trends highlight the early formation of supply-demand gaps that deepen in our forecasts.

27: Regional and Facility Access Trends



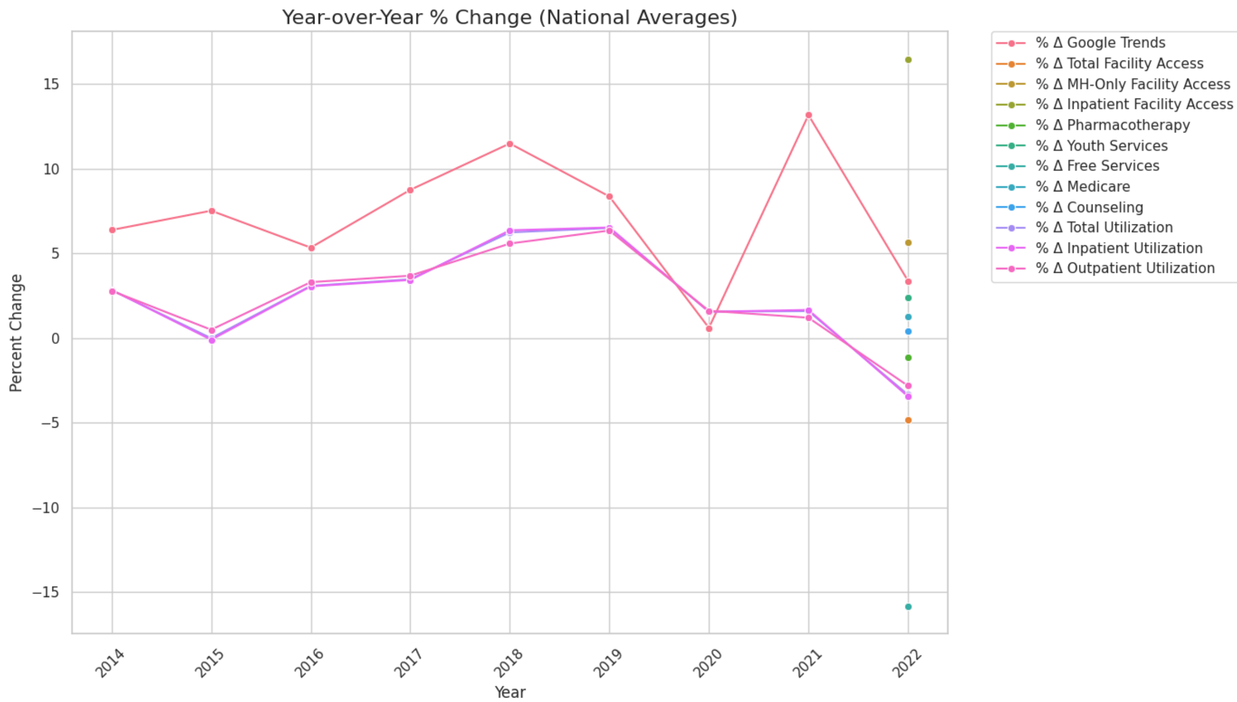
Regional trends show diverging trajectories with Google Trends scores rising most sharply in the West, where demand pressures are mounting. However, facility access (per capita) declines across all regions, particularly in the South—highlighting a region structurally vulnerable to demand surges. The boxplot of pharmacotherapy offerings reveals that the Atlantic region leads in medication-based care, while the South shows wide dispersion, indicating uneven infrastructure. The heatmap of per capita access starkly contrasts states like Massachusetts (top-tier access) with infrastructure deserts like Texas and Georgia. These visuals provide the empirical backbone for our segmentation and policy response.

28: Correlogram of Key Predictors and Utilization Variables



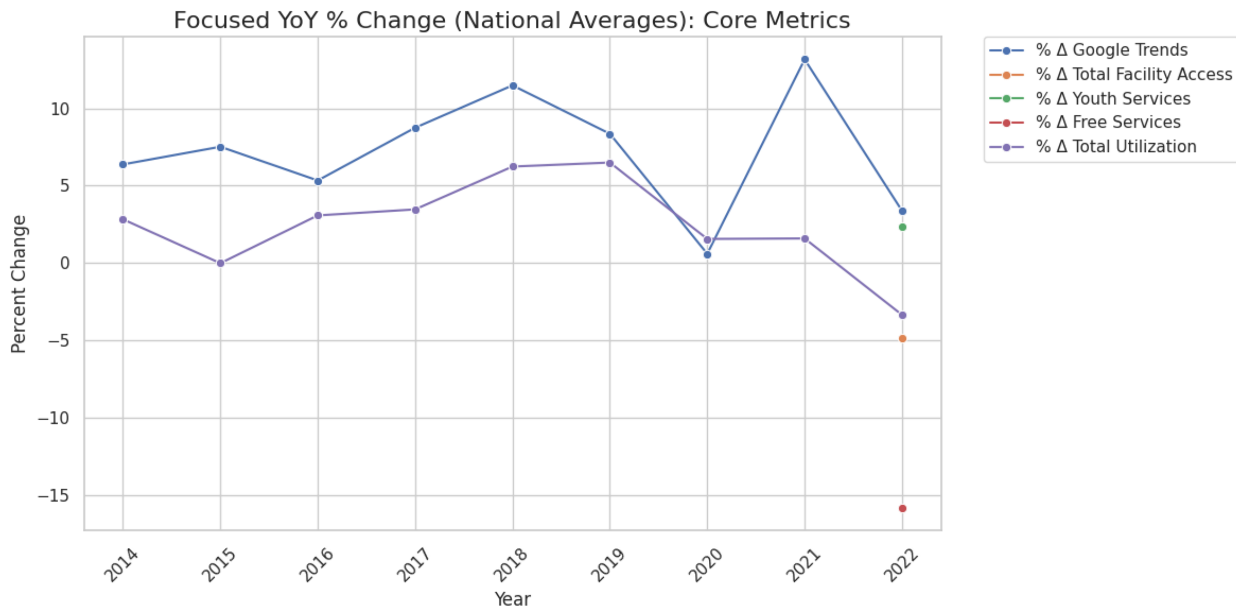
This matrix shows how predictors relate to each other and to our target. As expected, Google Trends and facility access are negatively correlated (−0.52), revealing a troubling mismatch: as behavioral signals increase, infrastructure doesn't follow. Free service offerings and counseling availability are positively associated, but neither correlates strongly with utilization—indicating that structural access alone isn't driving engagement. Notably, no single variable dominates total utilization, supporting our ensemble model's design, which blends predictors to explain a highly distributed, multi-causal outcome.

29: Year-over-Year % Change (National Averages)



The year-over-year percentage change in national averages across mental health access and utilization metrics from 2014 to 2022 reveals how behavioral demand—represented by Google Trends—has surged ahead of actual service use and access. Notably, the sharp spike in trend scores in 2021, peaking over 13%, contrasts with much smaller, flatter increases in utilization. This growing gap underscores a demand-access mismatch. Although utilization rose modestly through 2018–2019, the 2020 dip, coinciding with the pandemic and the tepid rebound since, suggest structural lag in meeting need. In 2022, most utilization and access indicators turned negative, with youth services, free care, and inpatient infrastructure contracting even as online behavioral signals remained strong.

30: Focused Year-over-Year % Change (Core Metrics)



This plot centers on a streamlined set of core metrics—Google Trends, facility access, youth services, free services, and total utilization. From this view, the trend is even more stark as the behavioral interest has grown consistently, often exceeding 5% annually, peaking in 2021 at over 13%. However, total facility access actually declined by 5% in 2022, while youth services and free care plummeted. Meanwhile, overall utilization stalled or dropped, despite steady growth in public awareness and need. This divergence reinforces what our model has shown repeatedly: we are seeing early warning signs of unsustainable gaps between demand and delivery, especially in under-resourced communities. For stakeholders, these visuals should serve as a call to action—not just to expand infrastructure, but to align services more equitably and responsively with real-time behavioral health signals.