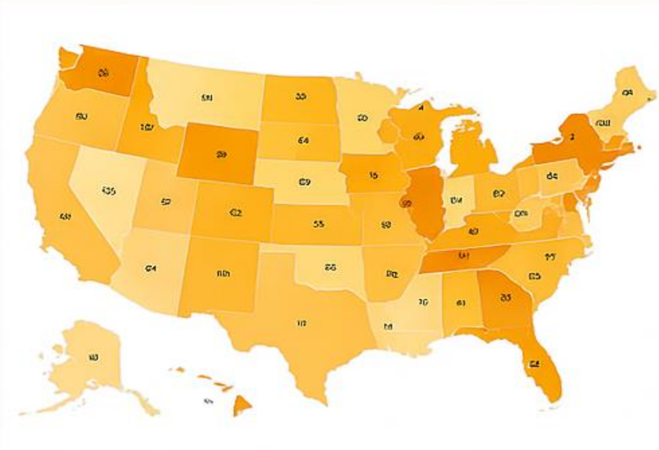
Kirti Kankaria

Market Saturation & Utilization:

State-County Health Ranking Analysis




Analyzing Medicare Market Saturation & Utilization: Strategic Implications for Health Systems and Policy Design

By: Kirti Kankaria

Executive Summary

This report presents a strategic analysis of how Medicare market saturation and utilization patterns affect county-level health behavior outcomes across the United States. Utilizing datasets from the Centers for Medicare & Medicaid Services (CMS) and the County Health Rankings initiative, this study explores the commercial and policy relevance of healthcare access and utilization patterns, focusing on Preventive Health Services (PHS) and Neurology. Through logistic regression and ensemble machine learning, the study identifies key socioeconomic and service availability variables influencing health rankings. The findings offer critical insights for government agencies, healthcare organizations, insurers, and policymakers aiming to optimize provider distribution, target interventions, and enhance health equity across diverse geographic markets.

Industry Context and Strategic Imperative

Healthcare delivery in the United States is transforming under the value-based care paradigm, prioritizing outcomes over volume. Geographic disparities, particularly in rural counties, undermine these objectives due to low provider availability and resource scarcity. These challenges are acute in Medicare-dependent populations, where limited provider presence and uneven service utilization can lead to deteriorating public health and higher costs. This study addresses the strategic understanding of how market saturation metrics, including fee-for-service penetration and dual-eligible concentration, serve as proxies for service accessibility and potential market inefficiencies.

From a business strategy standpoint, understanding healthcare access breakdowns allows healthcare systems and managed care organizations to target resource deployment effectively. Insights enable local governments and nonprofit entities to structure partnerships and funding proposals based on quantitative need rather than anecdotal demand. This project combines rigorous analysis with real-world applications extending beyond academic insights.

Research Objectives with Business and Policy Relevance

The core aim of this research is to provide actionable insights bridging data science and healthcare operations. Specifically, the project seeks to:

1. Identify primary operational and demographic variables predicting county health behavior outcomes.
2. Determine the influence of Medicare market saturation and service utilization patterns on county health behavior rankings, particularly in PHS and Neurology.

3. Evaluate which modeling approach—logistic regression or advanced ensemble models—offers a scalable, actionable framework for policymakers, public health leaders, and healthcare executives.

These goals inform statistical modeling and feed directly into business and policy decision-making processes, offering tools for risk stratification, performance measurement, and investment prioritization.

Stakeholder Value and Business Impact

The implications of this analysis are broad reaching. For federal and state agencies like CMS and the Department of Health and Human Services, the insights provide a roadmap for equitable resource allocation and provider reimbursement model refinement. Health systems expanding into underserved markets can leverage these results for workforce development, telehealth investments, and cross-sector collaboration with local agencies.

Private payers and Medicare Advantage plans can use this data to optimize network adequacy and member outreach strategies. Nonprofit health advocacy groups and community coalitions can derive empirical justification for grant applications and campaign targeting. Academically, the study supports efforts to model health determinants in a localized context, encouraging further integration of econometric methods in healthcare administration.

Analytical Methodology and Strategic Framework

Two robust public datasets were selected: the CMS Market Saturation and Utilization dataset and the County Health Rankings dataset. These were linked by FIPS codes to create a unified county-level view. The preprocessing strategy emphasized preserving business-relevant signal strength while ensuring model integrity. Median imputation was chosen for its balance between simplicity and robustness. Features were standardized for fair treatment in model estimation.

Steps involved:

1. **Exploratory Data Analysis (EDA)**: Initial data exploration to understand distributions, missing values, and potential correlations.

2. **Correlation and Causation Analysis**: Identifying relationships between variables and determining causative factors.

3. **Model Building**: Constructing logistic regression for interpretability and Random Forest/Gradient Boosting models for predictive accuracy.

4. **Dashboard Development**: Deploying a Power BI dashboard for interactive, on-demand access to insights for health administrators and field operators.

Key Insights from Preventive Health Services Domain

The PHS analysis revealed a direct relationship between population-level behavior and structural access indicators. Behavioral risk factors such as smoking, obesity, and alcohol misuse were significant predictors of worse health rankings, supporting investment in community health initiatives focusing on lifestyle modification.

Access-related indicators like higher primary care physician density and reduced preventable hospitalization rates emerged as protective factors, validating the business case for reinforcing primary care infrastructure as a cost-saving, outcome-improving measure. Logistic regression delivered an accuracy of 76%, while Random Forest and Gradient Boosting models improved predictive performance to 77% and 78%, respectively. The trade-off between interpretability and accuracy is crucial when choosing a modeling approach for business or policy deployment.

Strategic Findings from the Neurology Sector Analysis

The Neurology domain surfaced a slightly different set of priorities. Socioeconomic variables like uninsured adult population and dual-eligible user rates were highly correlated with worse health rankings. These factors reflect not just access gaps but deeper systemic vulnerabilities, including poverty and institutional inequity.

Conversely, the presence of more primary care physicians was again shown to be a mitigating factor, reinforcing its strategic value across specialties. delivered an accuracy of 70%, while Random Forest and Gradient Boosting models improved predictive performance to 76% and 73%, respectively. The trade-off between interpretability and accuracy is crucial when choosing a modeling approach for business or policy deployment. These results suggest that while Neurology services may require specialty-driven solutions, the infrastructure and economic levers available at the primary care level can significantly influence outcomes.

Interactive Business Intelligence via Power BI

To drive adoption among non-technical stakeholders, an interactive Power BI dashboard was developed. It allows users to slice data by state, health ranking class, and variable of interest. This tool supports scenario planning and performance auditing, enabling health system leaders to quickly identify at-risk areas and prioritize interventions. In commercial settings, such dashboards can be integrated into enterprise analytics platforms to inform everything from marketing strategy to provider contracting.

Policy Recommendations and Business Applications

Three tiers of recommendations emerge from the analysis:

1. For Neurology, targeted expansion of primary care access in counties with high uninsured and dual-eligible populations is essential. This can be pursued through telehealth, loan forgiveness for rural practitioners, or mobile clinic deployment.
2. For Preventive Health Services, investment in behavioral health outreach programs is critical. These should be culturally tailored and coupled with efforts to improve care coordination.

3. Strategically, organizations should adopt a dual-model approach: use interpretable models like logistic regression for governance and compliance purposes, and employ ensemble models in operational environments where predictive accuracy is paramount.

These recommendations are designed to bridge the gap between data science outputs and real-world actions, promoting scalable and sustainable health system improvements.

Limitations and Strategic Considerations

Despite its strengths, the analysis has limitations. Inconsistent or missing FIPS codes created data integration challenges that may bias results in certain geographies. Sparse data from some counties, particularly in frontier regions, limits generalizability. Ensemble models, although powerful, still lack industry-standard interpretability frameworks—raising barriers to adoption in regulatory settings.

Finally, the presence of counterintuitive statistical relationships reinforces the importance of integrating quantitative insights with local, qualitative knowledge. Business and policy leaders should view model outputs as directional guidance rather than deterministic rules.

Future Outlook and Scalability

Looking ahead, this model can be scaled both vertically and horizontally. Vertically, integrating time-series data will enable forecasting and impact evaluation over time. Horizontally, the framework can be extended to other specialties or social determinants such as housing and food access. The Power BI dashboard can evolve into a full-featured business intelligence platform with capabilities for predictive modeling, resource allocation simulations, and ROI forecasting for policy interventions.

From a strategic standpoint, embedding these analytics into the workflows of health systems, insurance payers, and community partners will be essential to achieving systemic impact. With the right investment in data governance and cross-sector collaboration, the approach outlined in this report can serve as a blueprint for transforming healthcare access and equity nationwide.

Acknowledgements

I would like to express my sincere gratitude to Professor VenuGopal Balijepally for his invaluable guidance and support throughout the preparation of this report. His insights and expertise were instrumental in shaping the direction and quality of my work.

References

- County Health Rankings & Roadmaps. (www.countyhealthrankings.org)
- Centers for Medicare & Medicaid Services (CMS) Public Use Files

---

Appendices

Appendix A: Correlations & Causation Variables
Appendix B:: Power BI Dashboard Screenshots
Appendix C: Data Dictionary
Appendix D: Model Performance Comparison Table
Appendix E: Model Performance Screenshot.

Appendix A

Correlations For PHS

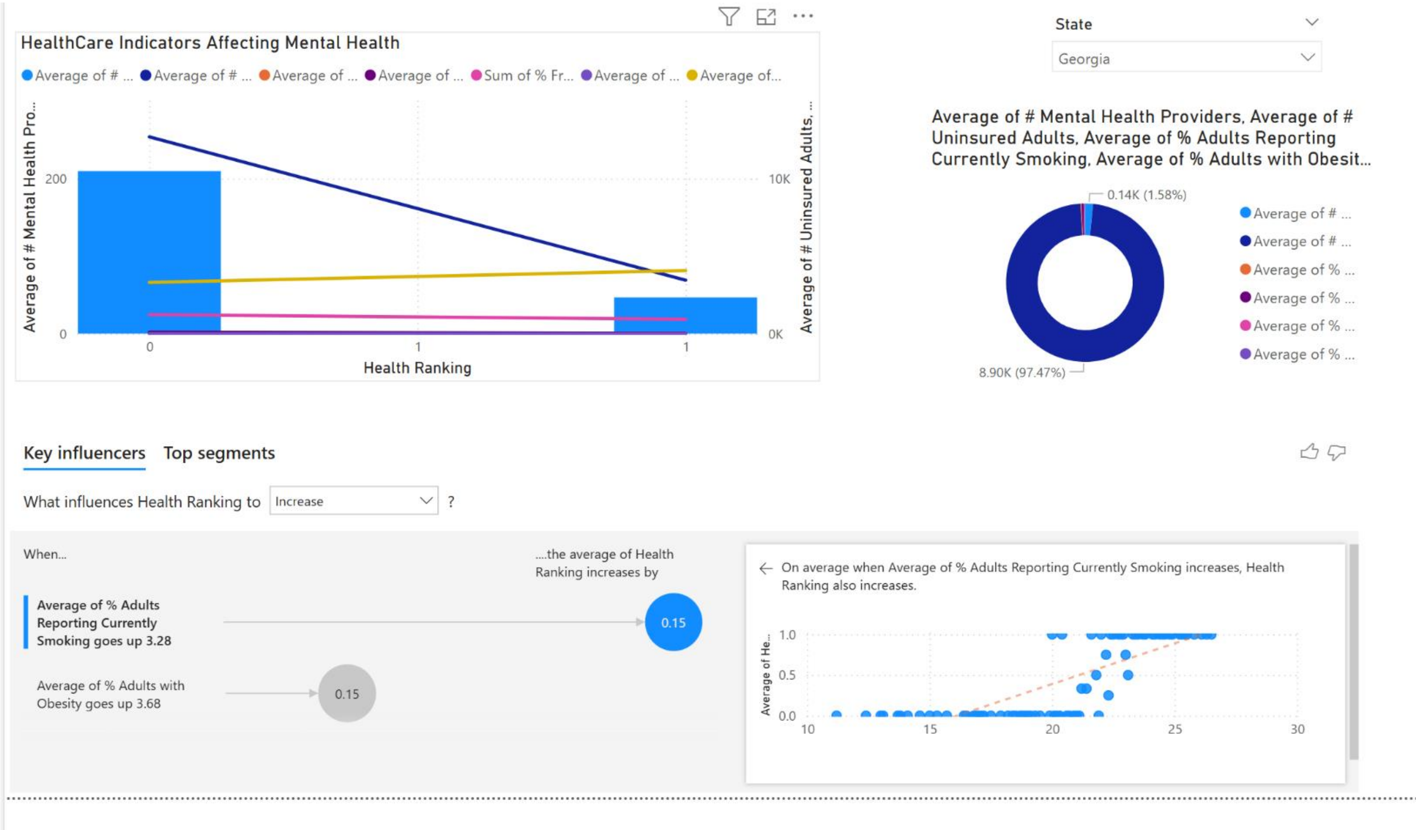
		Health Ranking	Preventable Hospitalization Rate	number_of_fee_for_service_beneficiaries	number_of_providers	average_number_of_users_per_provider	percentage_of_users_out_of_ffs_beneficiaries	number_of_users	average_number_of_providers_per_county	number_of_dual_eligible_users	percentage_of_dual_eligible_users_out_of_total_users	Deaths	% Fair or Poor Health	Average Number of Physically Unhealthy Days	Average Number of Mentally Unhealthy Days	% Adults Reporting Currently Smoking	% Adults with Obesity	% Excessive Drinking	% Uninsured	# Primary Care Physicians	Primary Care Physicians Rate	# Mental Health Providers	Mental Health Provider Rate	total_payment
Health Ranking	Pearson Correlation	1	.203**	-.175**	-.210**	-.048**	-.257**	-.185**	-.210**	-.073**	.356**	-.146**	.431**	.421**	.233**	.490**	.415**	-.164**	.119**	-.172**	-.224**	-.140**	-.067**	-.188**
	Sig. (2-tailed)		0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	3138	3071	3138	3138	3118	3138	3138	3138	2964	2964	2864	3137	3137	3137	3137	3137	3137	3138	2989	2991	2937	2937	3138
Preventable Hospitalization Rate	Pearson Correlation	.203**	1	-.056**	-.042*	-.171**	0.034	-.057**	-.042*	-0.022	.287**	-0.014	.493**	.459**	.377**	.470**	.453**	-.348**	.140**	-.060**	-.202**	-.065**	-.122**	-.059**
	Sig. (2-tailed)	0.000		0.002	0.020	0.000	0.062	0.002	0.020	0.240	0.000	0.441	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.001
	N	3071	3071	3071	3071	3063	3071	3071	3071	2941	2941	2856	3071	3071	3071	3071	3071	3071	3071	2933	2935	2887	2887	3071
number_of_fee_for_service_beneficiaries	Pearson Correlation	-.175**	-.056**	1	.966**	.238**	.179**	.993**	.966**	.825**	.075**	.953**	-.158**	-.196**	-.065**	-.334**	-.303**	-0.006	-.123**	.956**	.240**	.901**	.200**	.978**
	Sig. (2-tailed)	0.000	0.002		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.723	0.000	0.000	0.000	0.000	0.000	0.000
	N	3138	3071	3138	3138	3118	3138	3138	3138	2964	2964	2864	3137	3137	3137	3137	3137	3137	3138	2989	2991	2937	2937	3138
number_of_providers	Pearson Correlation	-.210**	-.042*	.966**	1	.161**	.239**	.968**	1.000**	.728**	0.028	.923**	-.193**	-.235**	-.071**	-.358**	-.320**	-0.008	-.150**	.930**	.264**	.849**	.200**	.961**
	Sig. (2-tailed)	0.000	0.020	0.000		0.000	0.000	0.000	0.000	0.000	0.127	0.000	0.000	0.000	0.000	0.000	0.000	0.637	0.000	0.000	0.000	0.000	0.000	0.000
	N	3138	3071	3138	3138	3118	3138	3138	3138	2964	2964	2864	3137	3137	3137	3137	3137	3137	3138	2989	2991	2937	2937	3138
average_number_of_users_per_provider	Pearson Correlation	-.048**	-.171**	.238**	.161**	1	0.032	.247**	.161**	.153**	.051**	.186**	-.119**	-.094**	-.050**	-.228**	-.266**	.051**	-.062**	.164**	.274**	.164**	.317**	.239**
	Sig. (2-tailed)	0.007	0.000	0.000	0.000		0.071	0.000	0.000	0.000	0.006	0.000	0.000	0.000	0.005	0.000	0.000	0.005	0.001	0.000	0.000	0.000	0.000	0.000
	N	3118	3063	3118	3118	3118	3118	3118	3118	2960	2960	2860	3118	3118	3118	3118	3118	3118	3118	2971	2973	2919	2919	3118
percentage_of_users_out_of_ffs_beneficiaries	Pearson Correlation	-.257**	0.034	.179**	.239**	0.032	1	.220**	.239**	.070**	-.260**	.111**	-.122**	-.142**	.046*	-.165**	-.070**	-.148**	-.124**	.111**	.055**	.075**	-.039*	.230**
	Sig. (2-tailed)	0.000	0.062	0.000	0.000	0.071		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.037	0.000
	N	3138	3071	3138	3138	3118	3138	3138	3138	2964	2964	2864	3137	3137	3137	3137	3137	3137	3138	2989	2991	2937	2937	3138
number_of_users	Pearson Correlation	-.185**	-.057**	.993**	.968**	.247**	.220**	1	.968**	.791**	.046*	.925**	-.168**	-.205**	-.071**	-.338**	-.310**	-0.006	-.125**	.932**	.242**	.868**	.194**	.993**
	Sig. (2-tailed)	0.000	0.002	0.000	0.000	0.000	0.000		0.000	0.000	0.012	0.000	0.000	0.000	0.000	0.000	0.000	0.724	0.000	0.000	0.000	0.000	0.000	0.000
	N	3138	3071	3138	3138	3118	3138	3138	3138	2964	2964	2864	3137	3137	3137	3137	3137	3137	3138	2989	2991	2937	2937	3138

Correlations For Neurology

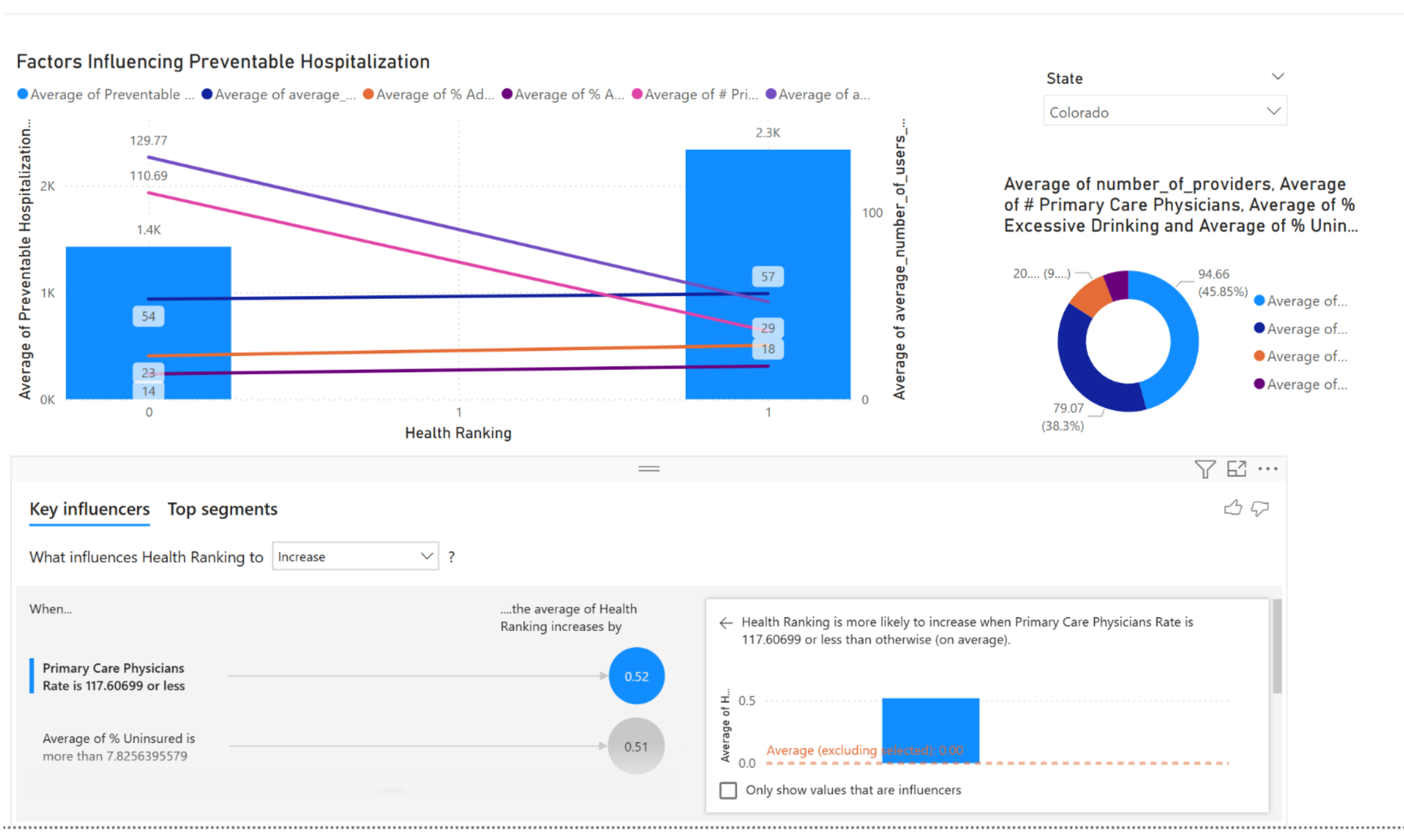
Correlations

		% Frequent Mental Distress	Average Number of Mentally Unhealthy Days	Average Number of Physically Unhealthy Days	Health Behaviors Quartile	% Fair or Poor Health	% Adults Reporting Currently Smoking	% Adults with Obesity	# Primary Care Physicians	Primary Care Physicians Ratio	# Mental Health Providers	Mental Health Provider Ratio	Preventable Hospitalization Rate	number_of_fee_for_service_beneficiaries	number_of_providers	average_number_of_users_per_provider	percentage_of_users_out_of_ffs_beneficiaries	number_of_dual_eligible_users	percentage_of_dual_eligible_users_out_of_total_users	total_payment	Population	Life Expectancy	% Frequent Physical Distress	# Uninsured Adults	Clinical Care Quartile
% Frequent Mental Distress	Pearson Correlation	1	.926**	.892**	.430**	.770**	.812**	.567**	-.218**	.294**	-.188**	.177**	.476**	-.216**	-.196**	0.002	0.017	-.115**	.236**	-.157**	-.189**	-.672**	.872**	-.119**	.363**
	Sig. (2-		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.955	0.386	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	2505	2505	2505	2505	2505	2505	2505	2459	2406	2463	2437	2501	2505	2505	1267	2505	944	944	2505	2391	2505	2505	2505	2505
Average Number of Mentally Unhealthy Days	Pearson Correlation	.926**	1	.789**	.285**	.632**	.638**	.399**	-.115**	.189**	-.089**	.088**	.402**	-.118**	-.108**	0.011	.041*	-.074*	.195**	-.093**	-.099**	-.542**	.739**	-.064**	.228**
	Sig. (2-	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.703	0.042	0.023	0.000	0.000	0.000	0.000	0.000	0.001	0.000
	N	2505	2505	2505	2505	2505	2505	2505	2459	2406	2463	2437	2501	2505	2505	1267	2505	944	944	2505	2391	2505	2505	2505	2505
Average Number of Physically Unhealthy Days	Pearson Correlation	.892**	.789**	1	.504**	.911**	.799**	.642**	-.222**	.331**	-.198**	.211**	.494**	-.230**	-.209**	-0.020	0.012	-.100**	.291**	-.162**	-.186**	-.709**	.982**	-.093**	.415**
	Sig. (2-	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.474	0.544	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	2505	2505	2505	2505	2505	2505	2505	2459	2406	2463	2437	2501	2505	2505	1267	2505	944	944	2505	2391	2505	2505	2505	2505
Health Behaviors Quartile	Pearson Correlation	.430**	.285**	.504**	1	.523**	.617**	.592**	-.280**	.319**	-.253**	.237**	.328**	-.302**	-.269**	-0.031	-0.018	-.125**	.279**	-.212**	-.254**	-.504**	.542**	-.164**	.655**
	Sig. (2-	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.273	0.363	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	2505	2505	2505	2505	2505	2505	2505	2459	2406	2463	2437	2501	2505	2505	1267	2505	944	944	2505	2391	2505	2505	2505	2505
% Fair or Poor Health	Pearson Correlation	.770**	.632**	.911**	.523**	1	.732**	.722**	-.166**	.348**	-.155**	.249**	.535**	-.180**	-.156**	-0.034	.046*	-0.030	.355**	-.111**	-.128**	-.742**	.947**	-0.031	.436**
	Sig. (2-	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.229	0.021	0.353	0.000	0.000	0.000	0.000	0.000	0.124	0.000
	N	2505	2505	2505	2505	2505	2505	2505	2459	2406	2463	2437	2501	2505	2505	1267	2505	944	944	2505	2391	2505	2505	2505	2505
% Adults Reporting Currently Smoking	Pearson Correlation	.812**	.638**	.799**	.617**	.732**	1	.715**	-.364**	.380**	-.337**	.260**	.492**	-.377**	-.334**	-0.048	-.052**	-.220**	.195**	-.278**	-.340**	-.738**	.820**	-.239**	.516**
	Sig. (2-	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.086	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	2505	2505	2505	2505	2505	2505	2505	2459	2406	2463	2437	2501	2505	2505	1267	2505	944	944	2505	2391	2505	2505	2505	2505
% Adults with Obesity	Pearson Correlation	.567**	.399**	.642**	.592**	.722**	.715**	1	-.324**	.348**	-.310**	.257**	.476**	-.331**	-.300**	-0.045	-0.001	-.193**	.174**	-.249**	-.274**	-.682**	.675**	-.158**	.470**
	Sig. (2-	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.112	0.954	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	2505	2505	2505	2505	2505	2505	2505	2459	2406	2463	2437	2501	2505	2505	1267	2505	944	944	2505	2391	2505	2505	2505	2505
# Primary Care Physicians	Pearson Correlation	-.218**	-.115**	-.222**	-.280**	-.166**	-.364**	-.324**	1	-.240**	.942**	-.178**	-.071**	.955**	.922**	-.061*	.168**	.724**	0.017	.778**	.962**	.238**	-.227**	.826**	-.231**
	Sig. (2-	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.031	0.000	0.000	0.592	0.000	0.000	0.000	0.000	0.000	0.000
	N	2459	2459	2459	2459	2459	2459	2459	2459	2404	2428	2402	2455	2459	2459	1262	2459	942	942	2459	2348	2459	2459	2459	2459
Primary Care Physicians Ratio	Pearson Correlation	.294**	.189**	.331**	.319**	.348**	.380**	.348**	-.240**	1	-.205**	.415**	.232**	-.232**	-.200**	-0.037	.067**	-.103**	.148**	-.140**	-.199**	-.253**	.344**	-.140**	.448**
	Sig. (2-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.195	0.001	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	2406	2406	2406	2406	2406	2406	2406	2404	2406	2379	2356	2402	2406	2406	1255	2406	941	941	2406	2297	2406	2406	2406	2406
# Mental Health Providers	Pearson Correlation	-.188**	-.089**	-.198**	-.253**	-.155**	-.337**	-.310**	.942**	-.205**	1	-.174**	-.076**	.900**	.843**	-0.048	.126**	.777**	.097**	.717**	.905**	.214**	-.204**	.732**	-.206**
	Sig. (2-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.087	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000
	N	2463	2463	2463	2463	2463	2463	2463	2428	2379	2463	2437	2459	2463	2463	1265	2463	944	944	2463	2350	2463	2463	2463	2463
Mental Health Provider Ratio	Pearson Correlation	.177**	.088**	.211**	.237**	.249**	.260**	.257**	-.178**	.415**	-.174**	1	.165**	-.194**	-.161**	-.080**	.053**	-.087**	0.037	-.115**	-.167**	-.150**	.217**	-.116**	.329**
	Sig. (2-	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.005	0.009	0.008	0.259	0.000	0.000	0.000	0.000	0.000	0.000
	N	2437	2437	2437	2437	2437	2437	2437	2402	2356	2437	2437	2433	2437	2437	1257	2437	943	943	2437	2324	2437	2437	2437	2437

Appendix B



HealthCare Indicators Affecting Mental Health
Average of # ... Average of # ... Average of ... Average of ... Sum of % Fr... Average of ... Average of...
Average of # Mental Health Pro...
Average of # Uninsured Adults, ...
200
0
10K
0K
0
1
1
Health Ranking
State
Georgia
Average of # Mental Health Providers, Average of # Uninsured Adults, Average of % Adults Reporting Currently Smoking, Average of % Adults with Obesit...
0.14K (1.58%)
8.90K (97.47%)
Average of # ...
Average of # ...
Average of % ...
Average of % ...
Average of % ...
Average of % ...
Key influencers
Top segments
What influences Health Ranking to
Increase
?
When...
....the average of Health Ranking increases by
Average of % Adults Reporting Currently Smoking goes up 3.28
0.15
Average of % Adults with Obesity goes up 3.68
0.15
On average when Average of % Adults Reporting Currently Smoking increases, Health Ranking also increases.
Average of He...
1.0
0.5
0.0
10
15
20
25
30



Factors Influencing Preventable Hospitalization
Average of Preventable ... Average of average_... Average of % Ad... Average of % A... Average of # Pri... Average of a...
Average of Preventable Hospitalization...
Average of average_number_of_users_...
129.77
110.69
1.4K
2.3K
2K
1K
0K
100
0
54
23
14
57
29
18
0
1
1
Health Ranking
State
Colorado
Average of number_of_providers, Average of # Primary Care Physicians, Average of % Excessive Drinking and Average of % Unin...
20.... (9....)
94.66 (45.85%)
79.07 (38.3%)
Average of...
Average of...
Average of...
Average of...
Key influencers
Top segments
What influences Health Ranking to
Increase
?
When...
....the average of Health Ranking increases by
Primary Care Physicians Rate is 117.60699 or less
0.52
Average of % Uninsured is more than 7.8256395579
0.51
Health Ranking is more likely to increase when Primary Care Physicians Rate is 117.60699 or less than otherwise (on average).
Average of H...
0.5
0.0
Average (excluding selected): 0.00
Only show values that are influencers

Appendix C

Key Variables and Why They Matter for Neurology:

1. **% Fair or Poor Health** – Neurological disorders significantly affect overall health status. Patients with poor health are more likely to need neurological services.
2. **Average Number of Physically Unhealthy Days** – Neurological disorders often lead to physical limitations and disabilities.
3. **Average Number of Mentally Unhealthy Days – (Highly Relevant!)** Neurological conditions, especially cognitive disorders like Alzheimer's or Parkinson's, can cause mental distress and cognitive decline.
4. **% Adults Reporting Currently Smoking** – Smoking is a major risk factor for stroke, aneurysms, and other neurological issues.
5. **% Adults with Obesity** – Obesity increases the risk of stroke and other conditions that lead to neurological disorders.
6. **% Adults with Diabetes** – Diabetes is a leading cause of neuropathy and increases the risk of stroke.
7. **# Primary Care Physicians & Ratio** – Primary care physicians often refer patients to neurologists for symptoms like chronic headaches, seizures, or cognitive decline.
8. **# Mental Health Providers & Ratio** – Mental health providers support patients with neurological disorders, as these conditions often have psychological impacts.
9. **Preventable Hospitalization Rate** – High rates may indicate inadequate management of neurological conditions that could otherwise be treated outpatient.
10. **number_of_fee_for_service_beneficiaries** – Many neurological services are used by older adults and Medicare beneficiaries.
11. **number_of_providers** – Adequate availability of neurologists ensures timely diagnosis and treatment.
12. **average_number_of_users_per_provider** – High provider workloads may lead to delayed diagnosis and poorer outcomes.
13. **percentage_of_users_out_of_ffs_beneficiaries** – This indicates the proportion of beneficiaries receiving neurological services.
14. **number_of_dual_eligible_users** – Dual-eligible individuals often have complex neurological conditions that require specialized care.
15. **percentage_of_dual_eligible_users_out_of_total_users** – Evaluates access disparities in neurological care.
16. **total_payment** – Higher payments may indicate greater use of specialized neurology services.
17. **Population** – Provides context for normalizing variables like provider ratios or service use.
18. **Life Expectancy** – Neurological conditions often reduce life expectancy, making this a relevant metric.
19. **% Frequent Physical Distress** – Chronic pain, muscle weakness, and mobility challenges are common in neurological disorders.
20. **% Frequent Mental Distress – (Highly Relevant!)** Depression, anxiety, and cognitive impairment are common among patients with neurological disorders.
22. **# Uninsured Adults** – Access to neurology care is crucial for managing chronic neurological conditions.

Key Variables and Why They Matter for PHS:

% Fair or Poor Health – Indicates overall population health, which preventive services aim to improve.
Average Number of Physically Unhealthy Days – Reflects the impact of preventive measures in reducing physical health issues.
% Adults Reporting Currently Smoking – Preventive care often targets smoking cessation, which directly affects long-term health.
% Adults with Obesity – Obesity prevention is a key goal of preventive health programs.
% With Access to Exercise Opportunities – Indicates whether the population has opportunities to engage in physical activity, a component of preventive care.
 # Primary Care Physicians & Ratio – Primary care physicians are often the first line of preventive care.
 # Mental Health Providers & Ratio – Mental health providers support preventive care in addressing mental health issues.
Preventable Hospitalization Rate – (Highly Relevant!) Reflects the success of preventive measures in avoiding hospitalizations.
number_of_fee_for_service_beneficiaries – Shows how many people are accessing fee-for-service models, which may include preventive services.
number_of_providers – More providers can lead to better access to preventive services.
average_number_of_users_per_provider – Reflects provider workload and potential gaps in preventive care.
percentage_of_users_out_of_ffs_beneficiaries – Indicates how well preventive services are being utilized within the fee-for-service population.
total_payment – Reflects the overall investment in preventive services.
Population – Helps normalize other variables for comparison.
Life Expectancy – (Potential Target Variable) Increased life expectancy can be a long-term indicator of the effectiveness of preventive services.
% Adults with Diabetes – Preventive services target diabetes management and prevention.
Health Behaviors Quartile & Clinical Care Quartile – Summarizes the quality of care and population health behavior, linked to preventive health.

Appendix D

Model For Neurology

ML/DL Prediction Models Summary

Model Parameters

(✔ = Feature selected in final model)

Feature	Logistic Regression	Random Forest	Gradient Boosting
% Adults Reporting Currently Smoking	✔	✔	✔
% Frequent Physical Distress	✔	✔	✔
% Adults with Obesity	✔	✔	✔
# Primary Care Physicians	✔	✔	✔
Preventable Hospitalization Rate	✔	✔	✔
Average Number of Physically Unhealthy Days	✔	✔	✔
% Fair or Poor Health	✔	✔	✔
Deaths	✔	✔	✔
# Uninsured Adults	✔	✔	✔
Life Expectancy	✔	✔	✔
number_of_dual_eligible_users	✔	✔	✔
average_number_of_users_per_provider	✔	✔	✔
number_of_fee_for_service_beneficiaries	✔	✔	✔
total_payment	✔	✔	✔

📊 Model Summary Metrics

Metric	Logistic Regression	Random Forest	Gradient Boosting
Accuracy	0.695	0.758	0.733

🧠 Group 1 (Class 0)

Metric	Logistic Regression	Random Forest	Gradient Boosting
Precision	0.725	0.756	0.729
Recall	0.684	0.805	0.789
F1-Score	0.704	0.780	0.758
Support	266	266	266

🧠 Group 2 (Class 1)

Metric	Logistic Regression	Random Forest	Gradient Boosting
Precision	0.664	0.761	0.737
Recall	0.706	0.706	0.668
F1-Score	0.685	0.733	0.701
Support	235	235	235

🧮 Macro Average

Metric	Logistic Regression	Random Forest	Gradient Boosting
Precision	0.695	0.759	0.733
Recall	0.695	0.755	0.729
F1-Score	0.694	0.756	0.730

🧮 Weighted Average

Metric	Logistic Regression	Random Forest	Gradient Boosting
Precision	0.696	0.759	0.733
Recall	0.695	0.758	0.733
F1-Score	0.695	0.758	0.731

Models For PHS

ML/DL Prediction Models Summary:

Model Parameters

✅ Selected Features Table (After Final Backward Elimination)

Feature	Logistic Regression	Random Forest	Gradient Boosting
% Adults Smoking	✔	✔	✔
% Adults with Obesity	✔	✔	✔
% Excessive Drinking	✔	✔	✔
% Uninsured	✔	✔	✔
% Fair or Poor Health	✔	✔	✔
Average Number of Physically Unhealthy Days	✔	✔	✔
Deaths	✔	✔	✔
Preventable Hospitalization Rate	✔	✔	✔
# Primary Care Physicians	✔	✔	✔
Avg Users/Provider (average_number_of_users_per_provider)	✔	✔	✔
% Users Not in FFS	✔	✔	✔
# Dual Eligible Users (absolute number)	✔	✔	✔
% Dual Eligible Users (share of total users)	✔	✔	✔

📊 Model Performance Summary (Final)

Overall Accuracy

Metric	Logistic Regression	Random Forest	Gradient Boosting
Accuracy	0.760	0.775	0.782

Group 1 (Class 0)

Metric	Logistic Regression	Random Forest	Gradient Boosting
Precision	0.761	0.764	0.776
Recall	0.773	0.814	0.807
F1-Score	0.767	0.788	0.791
Support	322	322	322

Group 2 (Class 1)

Metric	Logistic Regression	Random Forest	Gradient Boosting
Precision	0.757	0.789	0.788
Recall	0.745	0.735	0.755
F1-Score	0.751	0.761	0.771
Support	306	306	306

Macro Average

Metric	Logistic Regression	Random Forest	Gradient Boosting
Precision	0.759	0.777	0.782
Recall	0.759	0.774	0.781
F1-Score	0.759	0.775	0.781

Weighted Average

Metric	Logistic Regression	Random Forest	Gradient Boosting
Precision	0.760	0.776	0.782
Recall	0.760	0.775	0.782
F1-Score	0.759	0.775	0.782

Appendix E

Model for Neurology

```
=== Logistic Regression Summary (Backward Elimination) ===
                           Logit Regression Results
==============================================================================
Dep. Variable:         Health Ranking   No. Observations:                 2505
Model:                          Logit   Df Residuals:                     2490
Method:                           MLE   Df Model:                           14
Date:                Thu, 17 Apr 2025   Pseudo R-squ.:                  0.2718
Time:                        16:57:27   Log-Likelihood:                -1260.7
converged:                       True   LL-Null:                       -1731.3
Covariance Type:            nonrobust   LLR p-value:                6.206e-192
===============================================================================================================
                                                  coef    std err          z      P>|z|      [0.025      0.975]
---------------------------------------------------------------------------------------------------------------
const                                          -0.4096      0.067     -6.104      0.000      -0.541      -0.278
Deaths                                          1.4555      0.396      3.678      0.000       0.680       2.231
% Fair or Poor Health                          -1.1104      0.216     -5.152      0.000      -1.533      -0.688
Average Number of Physically Unhealthy Days    -3.1070      0.306    -10.163      0.000      -3.706      -2.508
% Adults Reporting Currently Smoking            0.3547      0.116      3.055      0.002       0.127       0.582
% Adults with Obesity                           0.4667      0.093      5.037      0.000       0.285       0.648
# Primary Care Physicians                      -2.2422      0.457     -4.906      0.000      -3.138      -1.346
Preventable Hospitalization Rate               -0.1891      0.061     -3.092      0.002      -0.309      -0.069
number_of_fee_for_service_beneficiaries         1.0334      0.356      2.903      0.004       0.336       1.731
average_number_of_users_per_provider           -0.1441      0.055     -2.629      0.009      -0.252      -0.037
number_of_dual_eligible_users                   1.3957      0.287      4.870      0.000       0.834       1.957
total_payment                                  -1.6885      0.439     -3.848      0.000      -2.548      -0.828
Life Expectancy                                -0.2623      0.094     -2.800      0.005      -0.446      -0.079
% Frequent Physical Distress                    4.7028      0.443     10.615      0.000       3.834       5.571
# Uninsured Adults                             -0.8614      0.289     -2.977      0.003      -1.429      -0.294
===============================================================================================================

=== Logistic Regression ===
Accuracy: 0.6946107784431138
Classification Report:
               precision    recall  f1-score   support

           0      0.725     0.684     0.704       266
           1      0.664     0.706     0.685       235

    accuracy                          0.695       501
   macro avg      0.695     0.695     0.694       501
weighted avg      0.696     0.695     0.695       501

Confusion Matrix:
 [[182  84]
 [ 69 166]]
```

Model for PHS

```
=== Logistic Regression Summary (Backward Elimination) ===
                           Logit Regression Results
==============================================================================
Dep. Variable:        Health Ranking   No. Observations:                 3138
Model:                         Logit   Df Residuals:                     3124
Method:                          MLE   Df Model:                           13
Date:               Thu, 17 Apr 2025   Pseudo R-squ.:                  0.3032
Time:                       17:05:35   Log-Likelihood:                -1515.0
converged:                      True   LL-Null:                       -2174.2
Covariance Type:           nonrobust   LLR p-value:                 6.031e-274
========================================================================================================================
                                                           coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------------------------------------------
const                                                   -0.2083      0.057     -3.675      0.000      -0.319      -0.097
Deaths                                                   1.0320      0.247      4.184      0.000       0.549       1.515
% Fair or Poor Health                                    1.1711      0.176      6.638      0.000       0.825       1.517
Average Number of Physically Unhealthy Days             -0.6257      0.152     -4.112      0.000      -0.924      -0.327
% Adults Reporting Currently Smoking                     1.1787      0.106     11.157      0.000       0.972       1.386
% Adults with Obesity                                    0.3005      0.084      3.559      0.000       0.135       0.466
% Excessive Drinking                                     0.2974      0.066      4.526      0.000       0.169       0.426
% Uninsured                                             -0.2979      0.057     -5.222      0.000      -0.410      -0.186
# Primary Care Physicians                               -2.3593      0.437     -5.404      0.000      -3.215      -1.504
Preventable Hospitalization Rate                        -0.1202      0.055     -2.177      0.030      -0.228      -0.012
average_number_of_users_per_provider                     0.2667      0.050      5.385      0.000       0.170       0.364
percentage_of_users_out_of_ffs_beneficiaries            -0.6213      0.051    -12.073      0.000      -0.722      -0.520
number_of_dual_eligible_users                            0.8206      0.157      5.239      0.000       0.514       1.128
percentage_of_dual_eligible_users_out_of_total_users     0.2740      0.046      5.927      0.000       0.183       0.365
========================================================================================================================

=== Logistic Regression ===
Accuracy: 0.7595541401273885
Classification Report:
               precision    recall  f1-score   support

           0      0.761     0.773     0.767       322
           1      0.757     0.745     0.751       306

    accuracy                          0.760       628
   macro avg      0.759     0.759     0.759       628
weighted avg      0.760     0.760     0.759       628
```