Summary of Performance Evaluation and Visualization

The model evaluation assessed the predictive performance of five distinct base models and a **Stacking Meta-Model** (Meta-Model) designed to combine their predictions. Key performance indicators used were the **Root Mean Squared Error (RMSE)**, **Mean Absolute Error (MAE)**, and the **coefficient of determination ($\mathbf{R^2}$)**.

---

1. Model Performance and Meta-Model Superiority

The evaluation demonstrated the Meta-Model's superior performance across all key metrics compared to the individual base models, confirming the benefit of the ensemble approach.

Model	RMSE	MAE	R2	Cross-Validated RMSE (CV_RMSE)
STACKING_META_MODEL	**3.42**	**2.39**	**0.825**	**3.82**
NCHS_year_race_sex_model	3.62	2.49	0.804	3.98
_133_year_race_sex_model	3.62	2.49	0.804	3.98
ssa_year_sex_model	4.56	3.17	0.690	4.61
state_sex_model	6.54	3.84	0.364	6.65
state_county_model	6.45×10^{11}	3.43×10^{11}	-6.20×10^{21}	1.92×10^{11}

Analysis:

- **Optimal Performance:** The Meta-Model achieved the **lowest RMSE (3.42)** and the **highest R^2 (0.825)**. This indicates that it explains 82.5% of the variance in the

target variable, marking a 5.5% improvement in R^2 over the best individual base models (NCHS and 133).
- **Base Model Performance:** Models utilizing detailed features (**NCHS** and **_133_**) were the strongest base learners.
- **Robustness to Instability:** The **state_county_model** exhibited catastrophic failure (resulting in nonsensical errors and a negative R^2), likely due to high-cardinality features leading to matrix singularity and instability. Crucially, the stacking methodology demonstrated resilience by successfully leveraging the strong models while mitigating the influence of the unstable base model.

---

2. Meta-Model Feature Contribution

To quantify the importance of each base model's prediction within the Meta-Model, **Permutation Importance** was calculated. This technique is robust to the multicollinearity issues that prevented traditional linear regression statistical analysis (T-statistics, F-statistics).

Analysis of Permutation Importance:

- The visualization (e.g., `meta_model_permutation_importance.png`) confirms that the Meta-Model's predictive power is overwhelmingly driven by the outputs of the best base models: **`nchs_year_race_sex_model_pred`** and **`_133_year_race_sex_model_pred`**.
- The importance scores for these predictions were significantly higher than those of the weaker base models, demonstrating that the Meta-Model effectively learned to **down-weight** the less reliable inputs, thus explaining its overall superior performance.

In conclusion, the evaluation confirms that the stacking ensemble is the most effective approach for this task, providing a measurable performance gain and exhibiting necessary stability when dealing with disparate and sometimes unstable base models.