

PROBLEM & OBJECTIVES

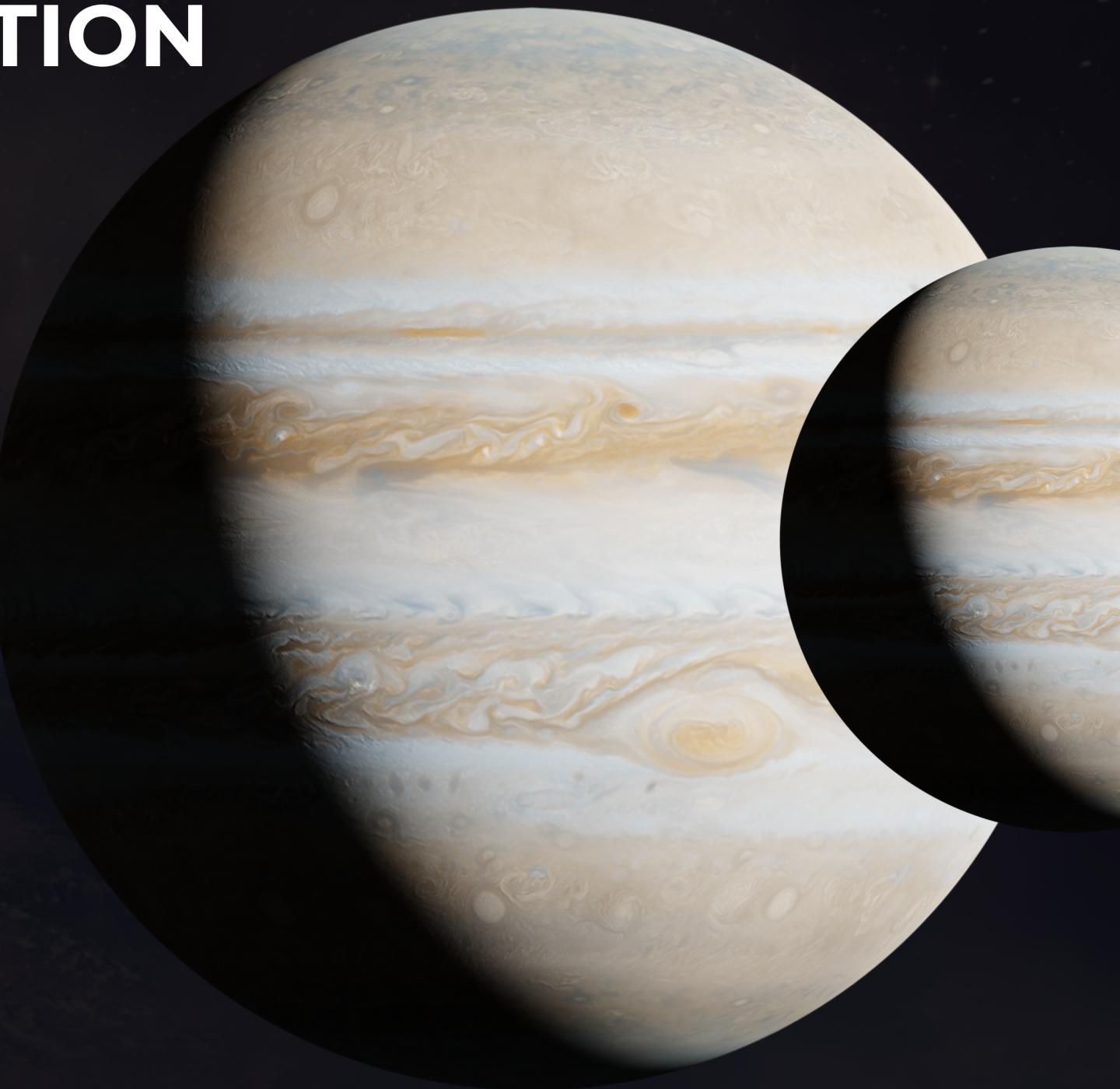
GALAXY WELL-BEING INDEX PREDICTION

Goal:

- Identify top socio-economic drivers of the Well-Being Index
- Predict well-being using historical galaxy data

Dataset:

- 3,097 galaxy-year observations
- 81 variables (health, education, income, etc.)



DATA CLEANING & FEATURE SELECTION



Feature Selection Strategy:

Dropped:

- All negatively or weakly correlated features
- Mid-correlation features with missing values

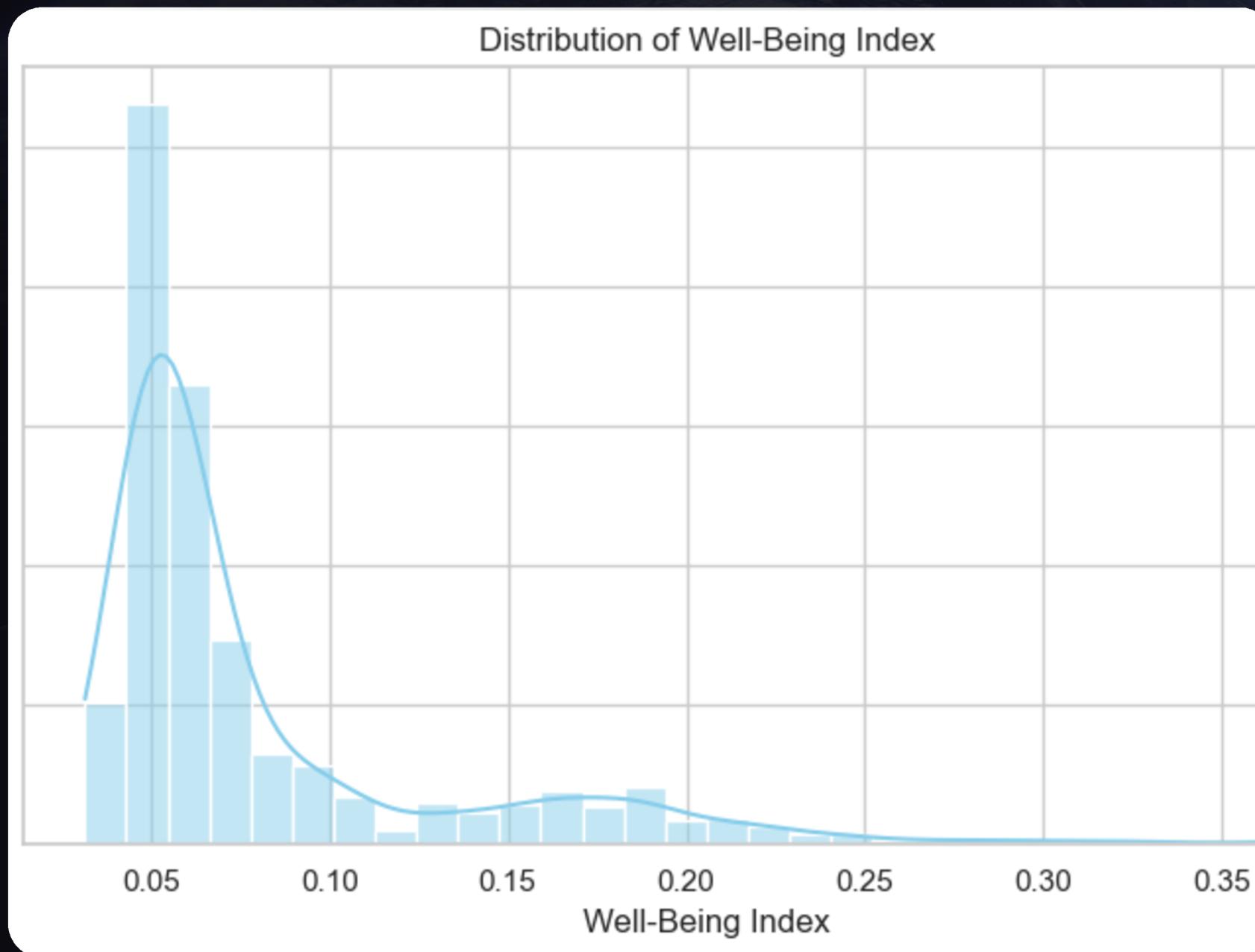
Retained:

- 30 high-correlation features (education, income, infrastructure)

Data Cleaning

- Imputed using KNN Imputer
- Created `_missing_flag` indicators
- Standardized and cleaned feature names

KEY FINDINGS & INSIGHTS



🌙 TOP POSITIVE PREDICTORS

- `Old age dependency ratio (+0.74)`
- `Digital Access & IDI indices`
- `Life Expectancy & Income Index`

🌙 TOP NEGATIVE PREDICTORS

- `Gender Inequality Index (-0.80)`
- `Youth dependency ratio`

🌙 IMPLICATION:

- Improved **education**, **digital access**, and **income equality** significantly raise well-being.

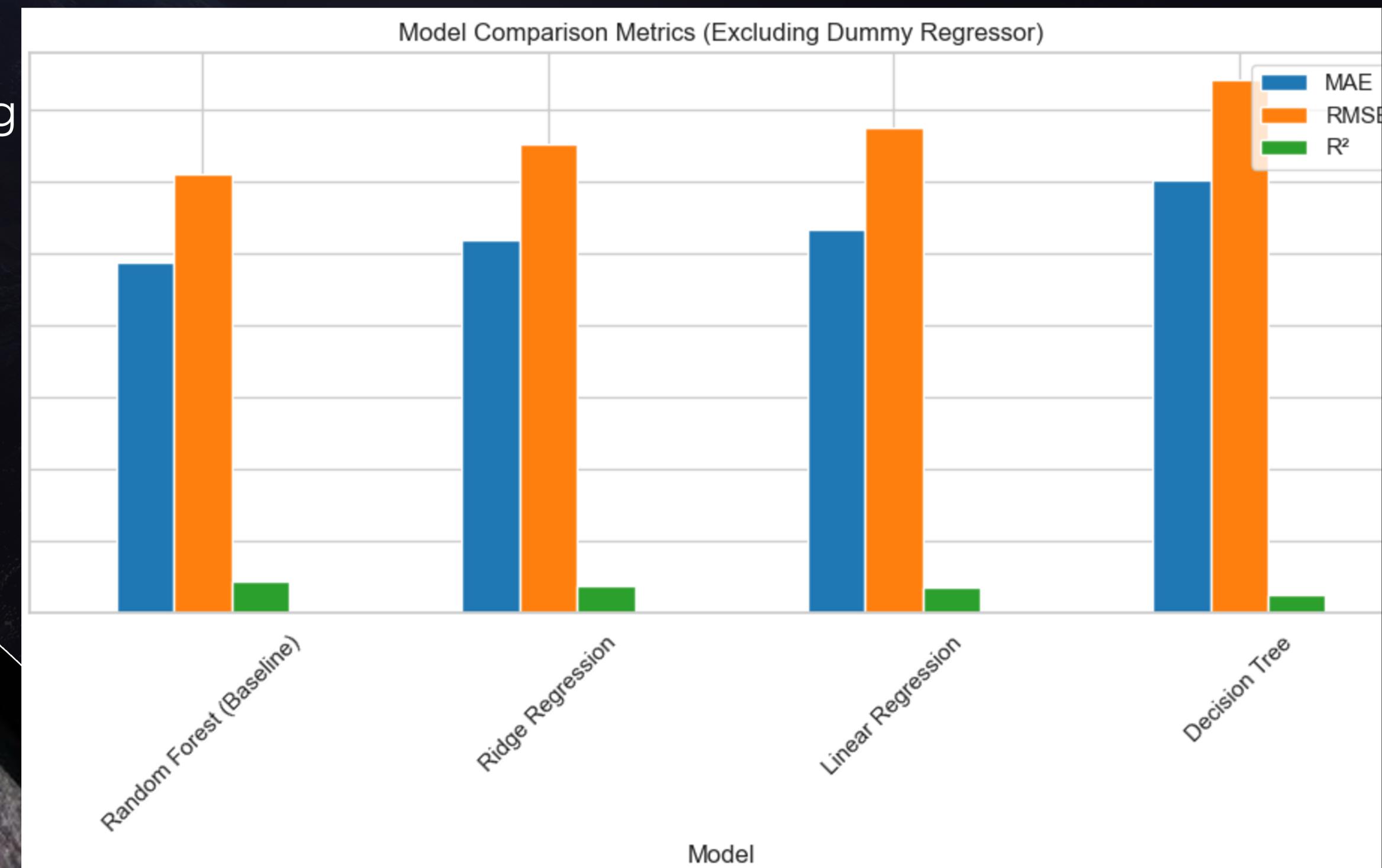
MODELING OVERVIEW

Models Compared

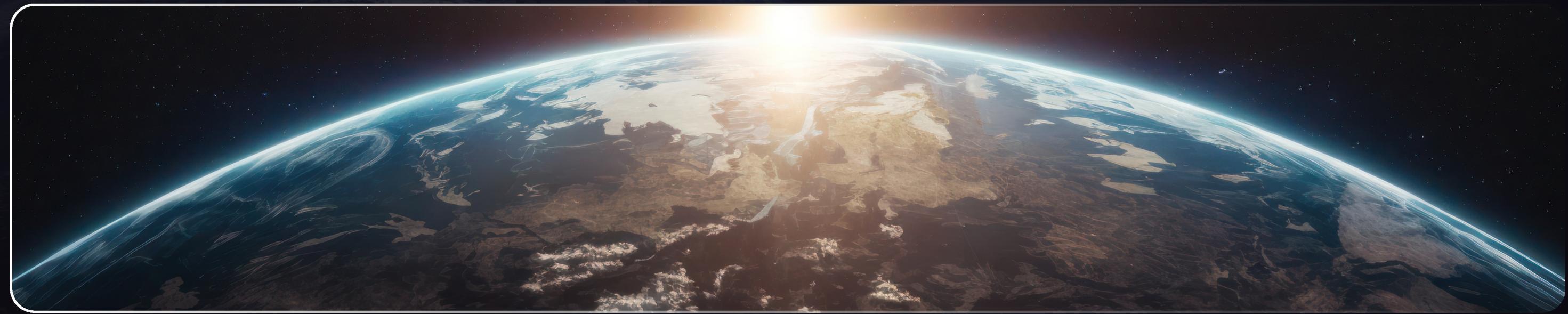
- Linear, Ridge, Decision Tree, Random Forest

Selected: Linear Regression

Chosen for interpretability and strong performance



RECOMMENDATIONS



1. Deploy Linear Regression

Simple, accurate, generalizable

2. Improve Key Drivers

Invest in: education, infrastructure, digital access

3. Enhance Data Collection

Reduce missingness in high-impact features

4. Retrain Yearly

Adapt to galactic policy changes and new trends