California House Price Prediction - Summary

Which features gave the best R²?

The project utilized the California Housing dataset, focusing on various features to predict median house values.

Among the features, Median Income emerged as the most significant predictor, exhibiting a strong positive correlation

with house prices. Other influential features included Housing Median Age, Total Rooms, and Population.

The linear regression model achieved an R^2 score of approximately 0.64, indicating that about 64% of the variance

in house prices could be explained by the model.

What could improve the model?

To enhance the model's performance, several strategies could be considered:

- Feature Engineering: Creating new features or transforming existing ones (e.g., rooms per household) could capture

more nuanced relationships.

- Handling Missing Values: Implementing more sophisticated imputation techniques might preserve data integrity

better than simple methods.

- Outlier Detection: Identifying and addressing outliers can prevent skewed model training.

- Advanced Algorithms: Exploring more complex models like Random Forests or Gradient Boosting Machines could

capture nonlinear relationships more effectively.

- Cross-Validation: Employing cross-validation techniques would provide a more robust assessment

of model performance.