Forecasting U.S. Rental Affordability

Time-Series Analysis of Top 20 High-Rent Markets

Data Science Bootcamp Final Project | Sejona Sujit Das

Project Overview & Dataset

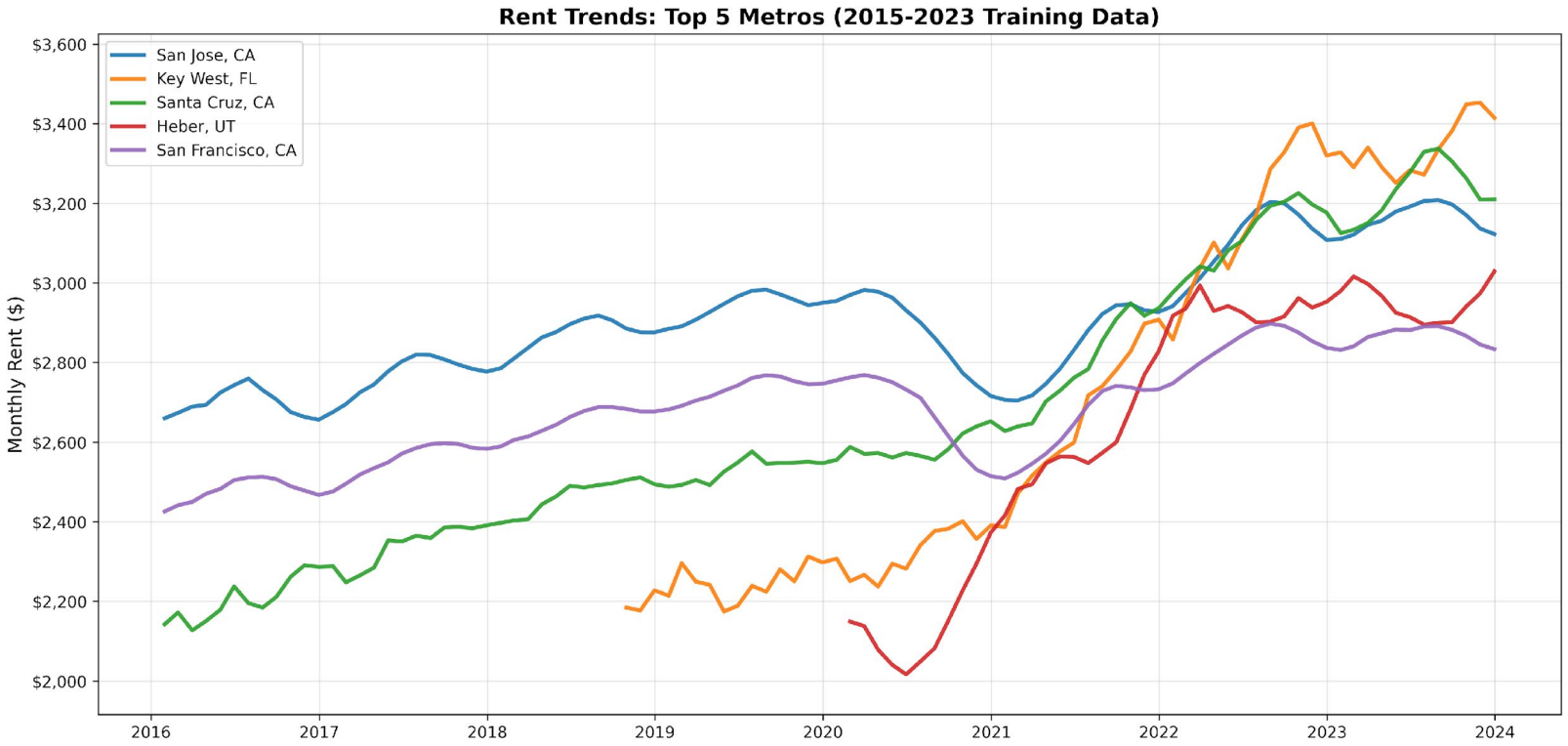
RESEARCH QUESTION

Can we build accurate forecasting models that substantially outperform naive baseline predictions?

- 7 models evaluated
- Proper temporal splits (train/val/test)

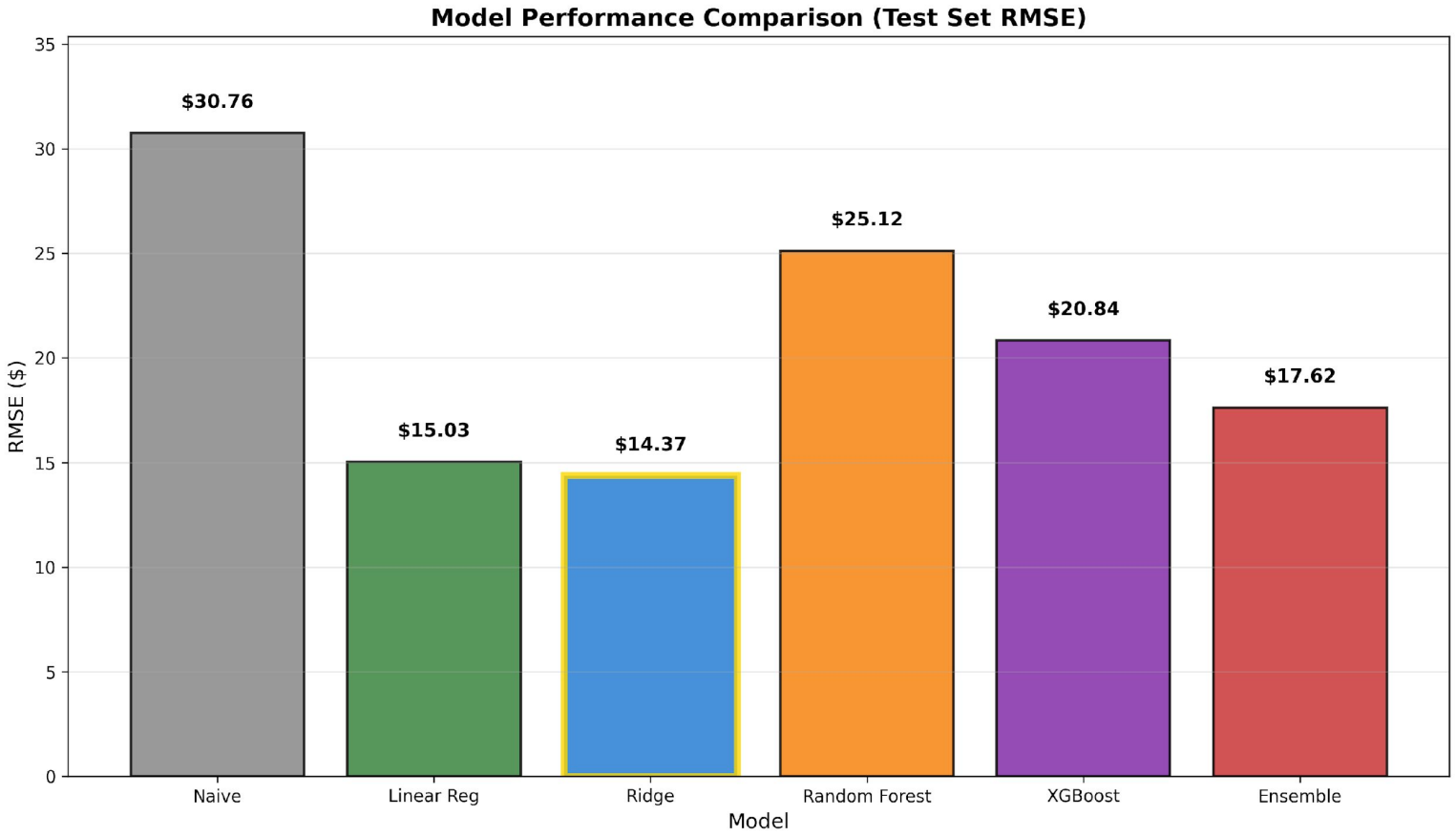
DATASET

- 2,451 observations (2015-2025)
- Top 20 highest-rent metros
- Zillow ZORI + FRED data
- 11 engineered features



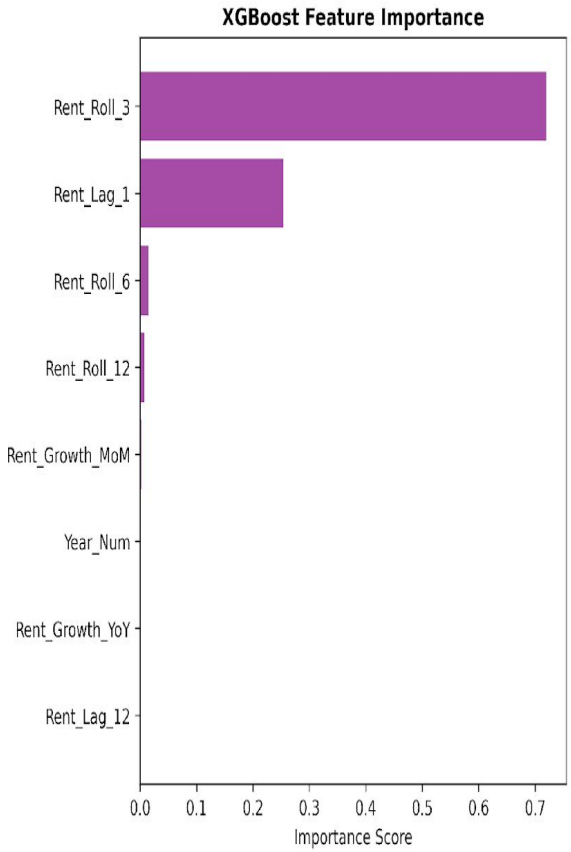
Model Performance

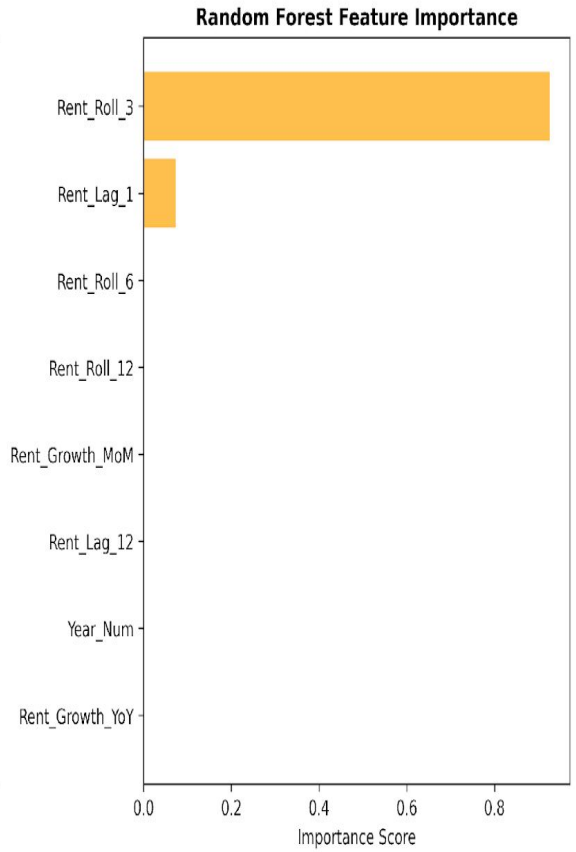
Best Model: Ridge Regression | Test RMSE: $14.37 | 53.3% improvement over baseline

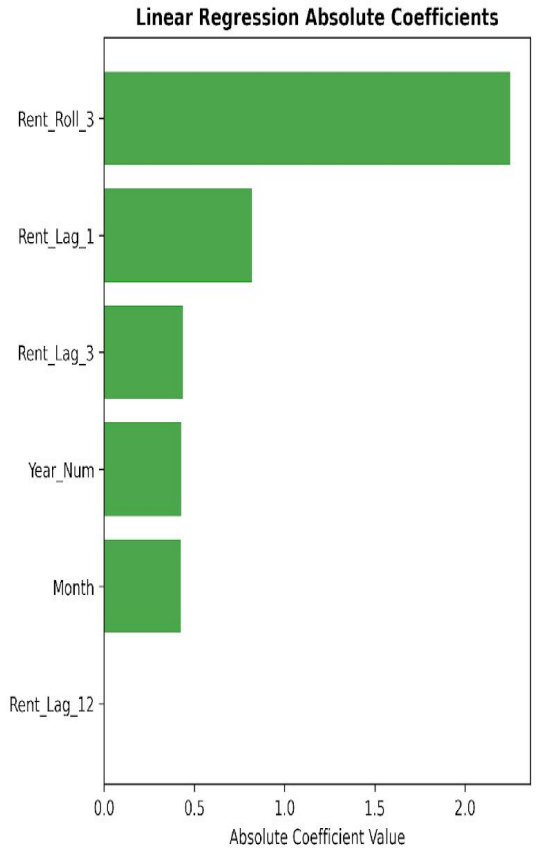


Key Findings: Feature Importance & Geographic Variation

Just 2 features explain 97% of predictions: 3-month rolling average (72%) + previous month rent (25%)



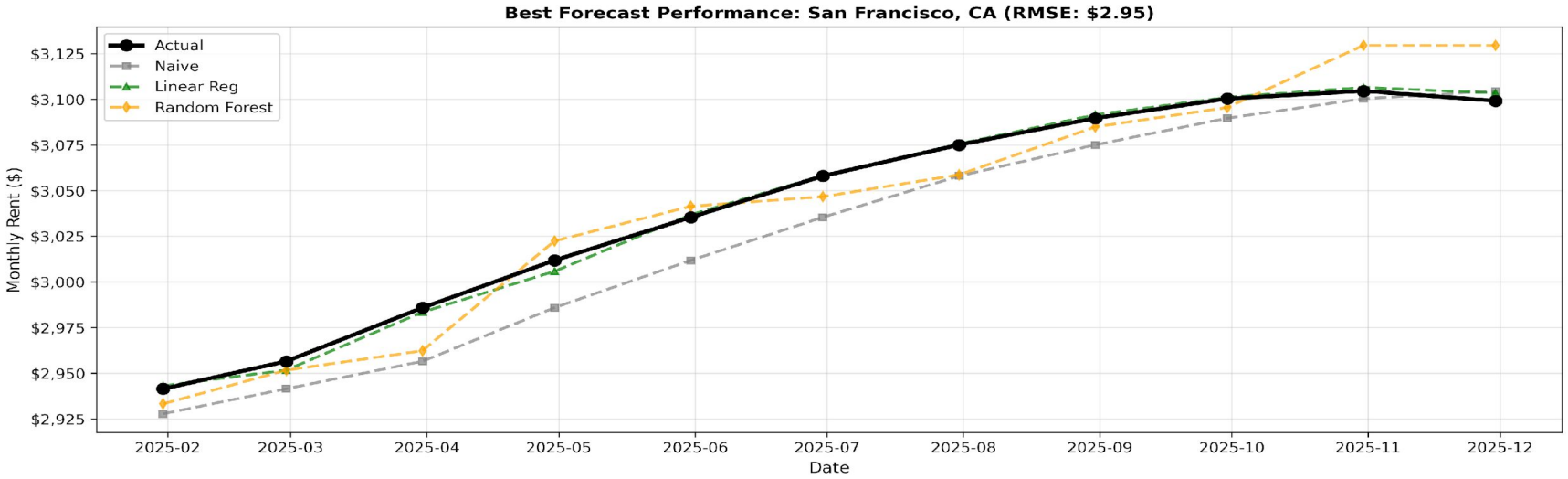


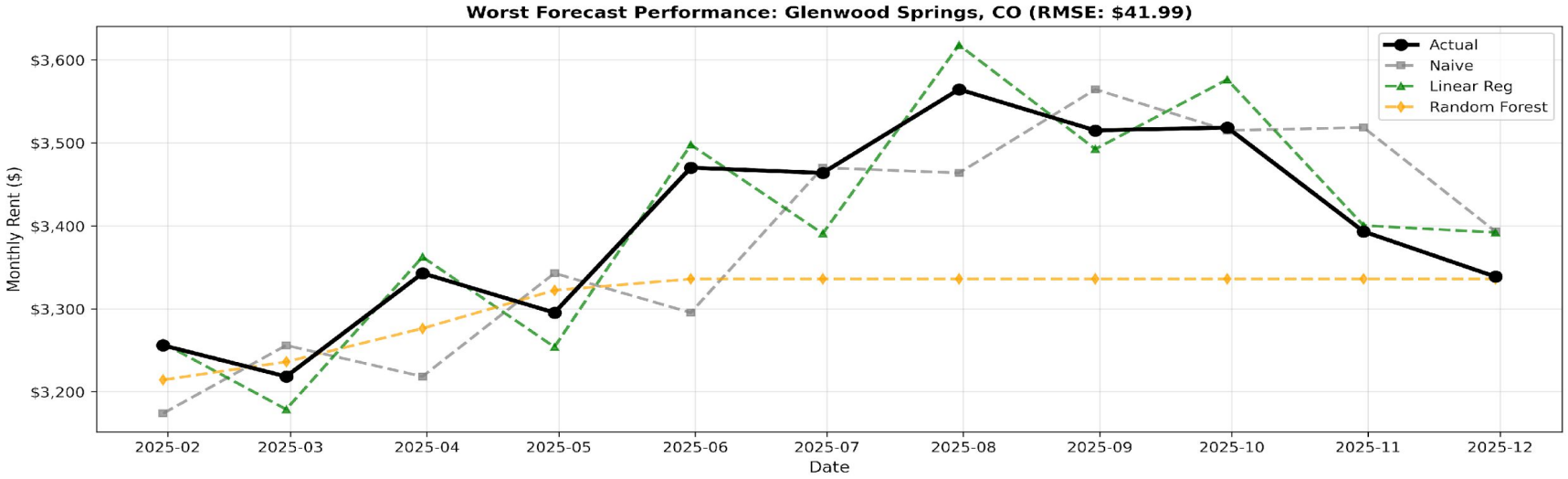


Key Findings: Feature Importance & Geographic Variation

Geographic Variation: San Francisco ($2.95 RMSE) vs Glenwood Springs ($41.99 RMSE)

Large, mature markets with diverse economies are highly predictable. Small resort markets with tourism dependence and seasonal volatility remain challenging.





Conclusions & Real-World Impact

KEY TAKEAWAYS

1. **Ridge/Linear beat Random Forest, etc.**
   Rental price dynamics are predominantly linear.
2. **Recent history dominates**
   3-month avg + lag explain 97% of variance
3. **High practical accuracy**
   ±$28 prediction intervals (95% CI)

REAL-WORLD IMPACT

1. **Property Managers**
   Optimize pricing & revenue forecasts
2. **Investors**
   Improve acquisition valuations
3. **Policymakers**
   Project affordability challenges

Conclusions & Real-World Impact

RECOMMENDATION

Deploy Ridge Regression (α=0.1) as production forecasting model for high-rent markets. Monitor residuals monthly for regime change detection. This simple, interpretable model provides optimal accuracy while remaining easy to maintain and explain to stakeholders.