ZIP-Level Home Price Forecasting

Current Forecasting Approach

Baseline Model

- Simple persistence model: usee laset price as the next month's forecast:

ARIMA Model

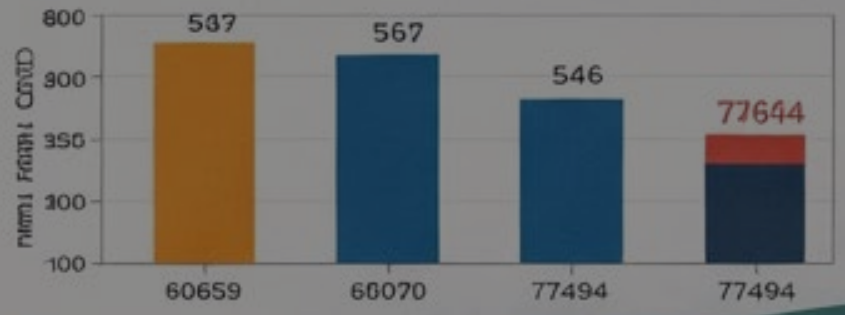
- Introduces trend analysis
- Considers both recent and

SARIMA Model

- Seasonal ARIMA: captures trend. and seasonality.
- Most accurate

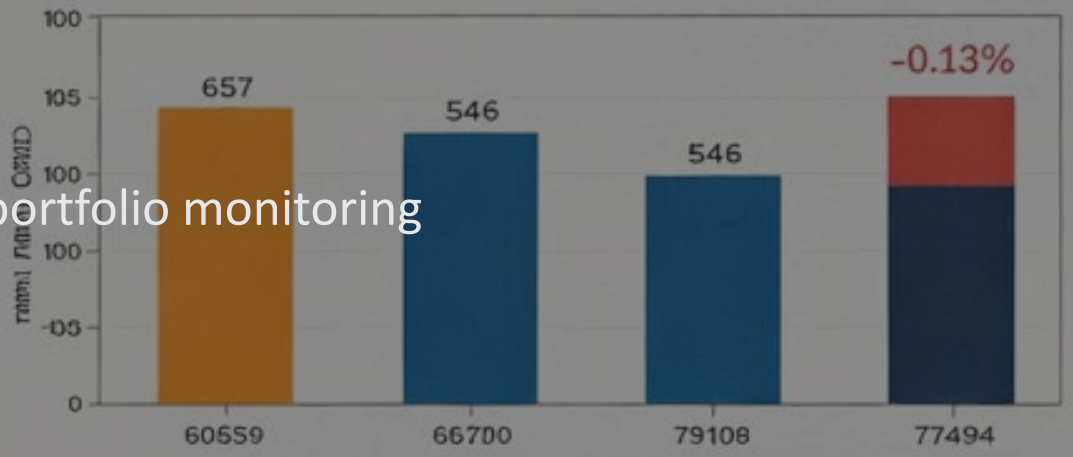
A practical risk signal for mortgage underwriting and portfolio monitoring

Test RMSE by ZIP: Baseline vs SARIMA



Example: How SARIMA Supports Loan Decisions

Test RMSE by ZIP (SARIMA) — Each bar = one ZIP



- For this set of ZIPs, SARIMA had lower forecast error, improving risk assessment.
- Red bar alet: ZIP 77494 is flagged as higher risk for a material declive of over 1% in the next month.
- Helps identify higher-risk areas to monitor more closely when making lending decisions

Rahaman Yusuf

ZIP-Level Home Price Forecasting

A practical risk signal for mortgage underwriting and portfolio monitoring

Rahaman Yusuf — Junior Data Scientist

Why this matters to a mortgage lender

Goal

Forecast next-month home values by ZIP and flag potential downside risk.

Stakeholder

Mortgage lenders and under writers

How you can use it

- Underwriting: adjust LTV / pricing where risk of decline is elevated
- Portfolio monitoring: scan ZIPs and focus review on the riskiest segments
- Early warning: flag a “material decline” if forecasted return ≤ -1% next month

Two outputs: price forecast + risk flag

1) Next-month price

Forecast the next monthly home value for each ZIP.

Example use: “Expected value next month ≈ $372K”

2) Material-decline flag

Convert the forecast into a simple risk signal for downside protection.

Flag = 1 if
forecasted return ≤ -1%

Example use: “Review this ZIP’s exposure before renewing / expanding.”

Progressive modeling: simple → smarter

We start simple and add structure step-by-step:

Baseline

"next month = last month"

ARMA

short-term patterns

ARIMA

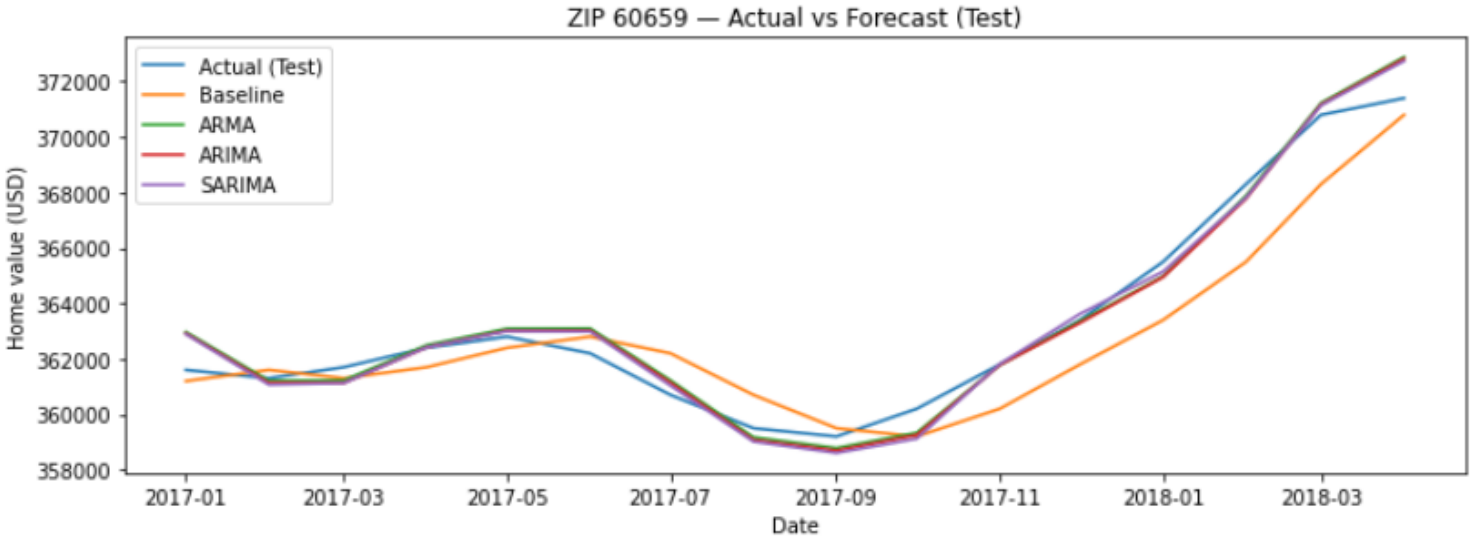
trend-aware

SARIMA

trend + seasonality

Models are selected by validation performance using RMSE and MAE (lower is better).

Example ZIP: actual vs forecast (test period)

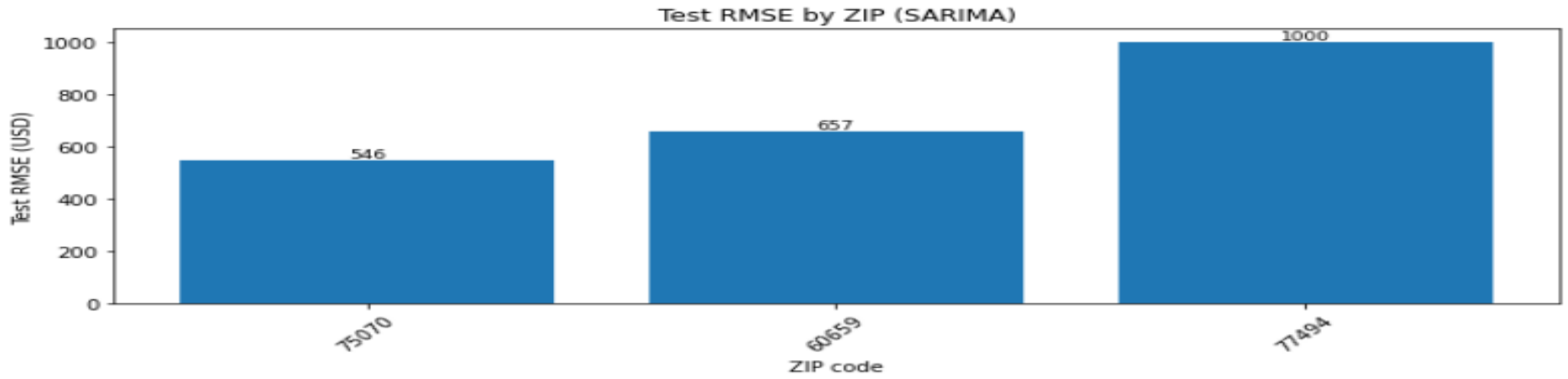


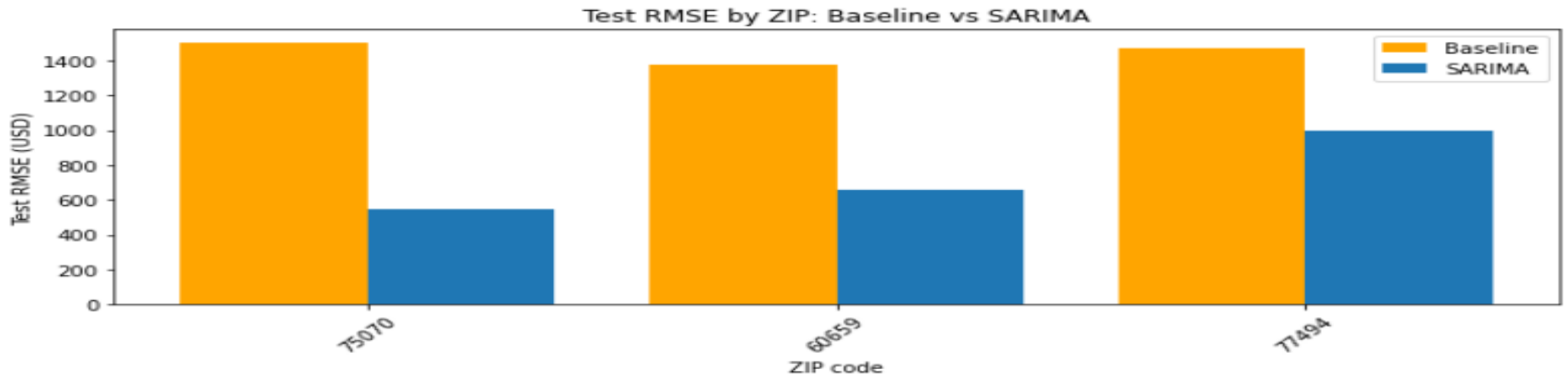
	model	MAE	RMSE	n
0	SARIMA(1, 1, 1)x(1, 0, 1, 12) on log	521.617679	656.538703	16
1	ARIMA(1, 1, 1) on log	525.750766	664.834330	16
2	ARMA(1, 1) on diff(log)	516.949359	668.943555	16
3	Baseline (persistence)	1125.000000	1373.408169	16

Key takeaway: the seasonal model (SARIMA) better tracks recurring market patterns than simpler baselines.

Portfolio view: compare ZIP performance & risk

	zip	rmse_baseline	rmse_sarima	rmse_improvement	next_month_pred_ret	risk_flag_decline_le_1pct
0	75070	1502.081889	546.174452	955.907436	0.000854	0
1	60659	1373.408169	656.538703	716.869467	-0.001336	0
2	77494	1470.544117	1000.258140	470.285977	0.005991	0





Interpretation: each bar corresponds to a ZIP you selected. Lower RMSE means more accurate monthly forecasts.

Conclusion & Next Steps

Recommended workflow

This project demonstrates a complete supervised time-series ML pipeline with:

- baseline + multiple ARIMA-family models,
- explicit validation-based model selection,
- holdout testing, diagnostics, and a risk-oriented decline alert.

Next steps:

Next steps

- Add exogenous macro variables (rates, unemployment), scale evaluation across many ZIPs, and improve the alert calibration for higher recall
- Add loan-level features (LTV, DTI), calibrate thresholds, and monitor performance over time.