

Project Proposal: Predictive Modeling and Visualization of U.S. Rental Trends Using Zillow Rent Index

1. Problem Identification

The U.S. rental housing market has experienced significant fluctuations over the past decade due to economic cycles, population shifts, and policy changes. Stakeholders lack accessible, data-driven tools to monitor, understand, and predict rental trends at a granular level across cities and neighborhoods. This limits their ability to make proactive decisions regarding housing policies, affordability initiatives, and investment strategies.

2. Problem Statement

How can we use historical rental data to forecast future rent prices across U.S. regions and visualize these trends in a way that supports effective decision-making by renters, policymakers, and investors?

3. Context

The rental market is a key component of the broader housing economy. Rising rents impact affordability, mobility, and housing security, especially in high-growth urban centers. With detailed rent data made available by Zillow's Observed Rent Index (ZORI), we now have the opportunity to apply machine learning and time series forecasting methods to anticipate future trends and present insights via interactive dashboards.

4. Criteria for Success

- A predictive model with strong performance on historical data (low RMSE/MAE)
- Forecasts of rent trends for the next 12–24 months, segmented by city or ZIP code
- A fully functional, interactive dashboard to explore trends and forecasts
- Clear, actionable insights tailored to different stakeholder needs
- Reproducible, well-documented code and visualizations

5. Scope of the Solution Space

Included:

- Historical trend analysis
- Time series and ML-based forecasting
- Regional comparisons and interactive visualizations
- ZIP code and city-level granularity

Excluded:

- Individual-level rent prediction
- Integration with real-time rent listings or transaction-level data

6. Constraints

- **Data limitations:** ZORI provides only median estimates; does not include individual rental contracts or socioeconomic covariates
- **Forecasting uncertainty:** Long-term projections are inherently uncertain, especially post-pandemic or in volatile markets
- **Geographic sparsity:** Some ZIP codes or regions may have sparse or inconsistent data
- **Computational resources:** Deep learning models may require GPU acceleration for efficiency

7. Stakeholders

- **Renters:** Seeking affordable housing options and long-term rent outlooks
- **Policymakers/City Planners:** Monitoring affordability, planning zoning and subsidies
- **Real Estate Investors/Developers:** Identifying profitable markets and trends
- **Researchers and Urban Economists:** Analyzing housing economics and demographics
- **Zillow & Similar Platforms:** Enhancing insights provided to users

8. Data Sources

Primary:

- **Zillow Observed Rent Index (ZORI)** from Kaggle
 - Monthly median rent estimates from 2010 to present
 - Coverage: ZIP code, city, county, and metro area
 - Granularity by number of bedrooms/property type

Data Acquisition:

- Kaggle dataset download
- Potential enrichment with U.S. Census data for context (optional)

9. Data Schema (Zillow Rent Index)

Below is a general structure of the dataset fields:

Column Name	Description
RegionID	Unique identifier for the geographic area
RegionName	Name of the city, ZIP code, or metro area
SizeRank	Ranking based on population or market size
RegionType	Type of region (Zip, City, County, Metro, State)
State	Two-letter state abbreviation

Metro	Associated metropolitan area
CountyName	Associated county name
2010-01, ..., 2025-05	Monthly median rent price values (wide format by date)

Note: The dataset is in wide format where each date is a separate column representing the median rent estimate for that month.

10. Approach and Methodology

a. Data Preprocessing and Exploration

- Parse and clean the ZORI dataset
- Reshape into time series format by ZIP code or city
- Visual EDA to identify patterns and anomalies

b. Predictive Modeling

- Baseline model: ARIMA or Prophet
- ML models: XGBoost, Random Forest
- Optional: Deep learning (LSTM) for complex sequences
- Evaluate using walk-forward validation (metrics: RMSE, MAE)

c. Visualization and Dashboarding

- Develop interactive dashboard using **Streamlit** or **Plotly Dash**
- Features:
 - Rent trends over time
 - Forecasting interface
 - Geographic drilldowns (e.g., choropleth maps)