Predictive Modeling and Visualization of U.S. Rental Trends Using Zillow Rent Index
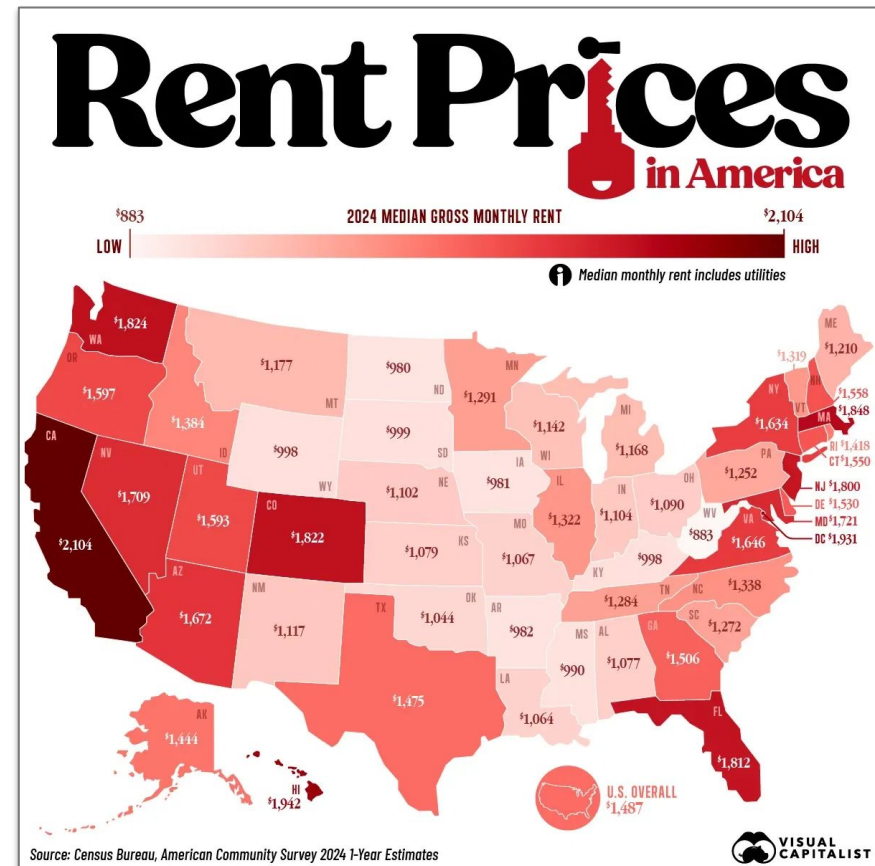
SUREKHA RAMANAIAH

The Problem

- The U.S. rental market has experienced significant volatility over the past decade.

- Zillow's ZORI dataset offers rich historical rental data, but it does not provide forward-looking insights.

- Stakeholders such as renters, policymakers, and investors need reliable forecasts to support decision-making.

Dataset Overview

- The project uses Zillow's ZORI dataset, which captures monthly median rents at the city level.

- The raw data is provided in a wide format, with each month represented as a separate column.

- Additional geographic features including city, state, metro area, and population rank enable more contextual modeling.

Data Wrangling

- Transformed the dataset from wide to long format to create a usable time-series structure.

- Generated a unified date index to ensure chronological consistency.

- Addressed missing data through a combination of filtering and forward-filling.

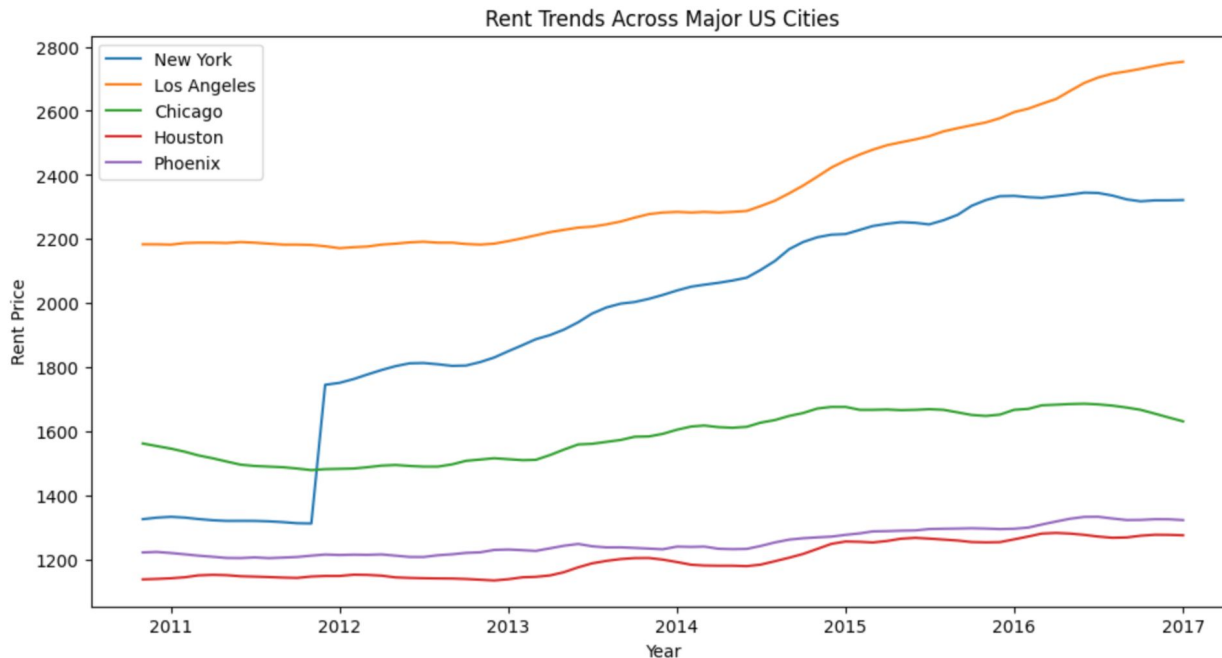- Capped extreme outliers to stabilize model training and reduce distortion.

EDA: Distribution Analysis

- Rental prices show a right-skewed distribution, driven by a small number of high-cost cities.

- This pattern highlights the importance of outlier treatment prior to modeling.
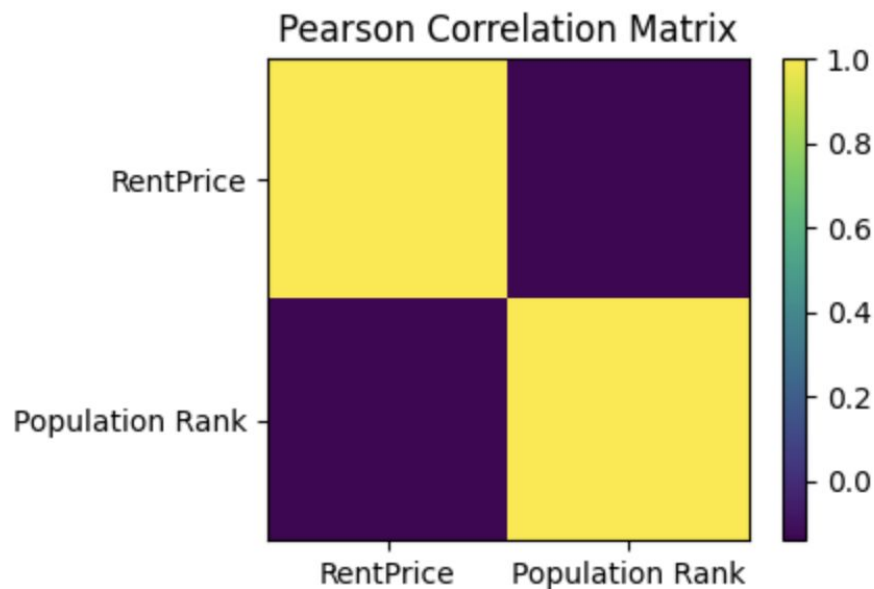

Distribution of Rent Prices

EDA: Temporal Trends

- Rent values across major U.S. cities exhibit persistent upward trends.

- The rate and magnitude of growth vary by city, reinforcing the need for localized forecasting.

Rent Trends Across Major US Cities

EDA: Geographic Relationships

- Pearson correlation coefficient between RentPrice and Population Rank was –0.142, indicating a very weak negative linear relationship.

- However, the relationship alone is insufficient for predictive accuracy, supporting a multivariate modeling approach.

Pearson Correlation Matrix

Preprocessing Approach

- Engineered temporal features such as year, month, and ordinal indices to capture seasonality and long-term trends.

- Encoded geographic attributes to support model generalization.

- Applied a chronological train test split to prevent data leakage and ensure forecasting validity.
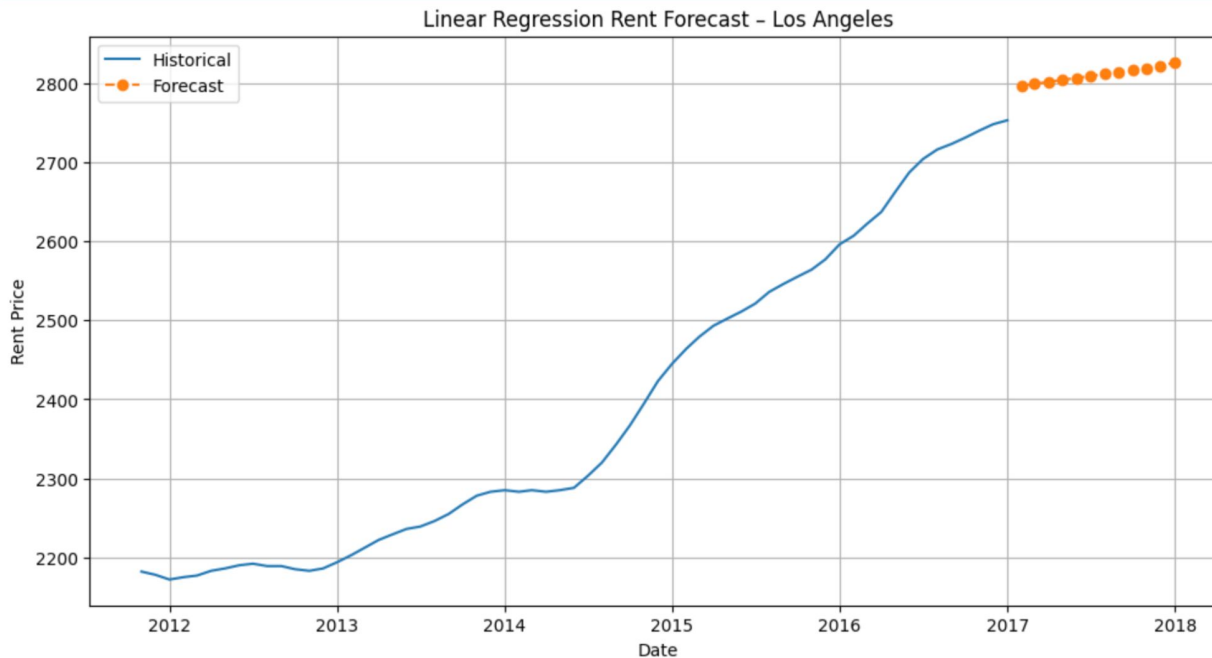
Modeling Strategy

- Evaluated Linear Regression, Random Forest, and Gradient Boosting models.

- Prioritized interpretability, stability, and computational efficiency.

- Linear Regression emerged as the strongest model given the engineered, largely linear feature structure.

Linear Regression Results

- MAE: 254.7

- RMSE: 495.1

- R^2: 0.43

- Linear Regression outperformed more complex models, delivering consistent and interpretable results.
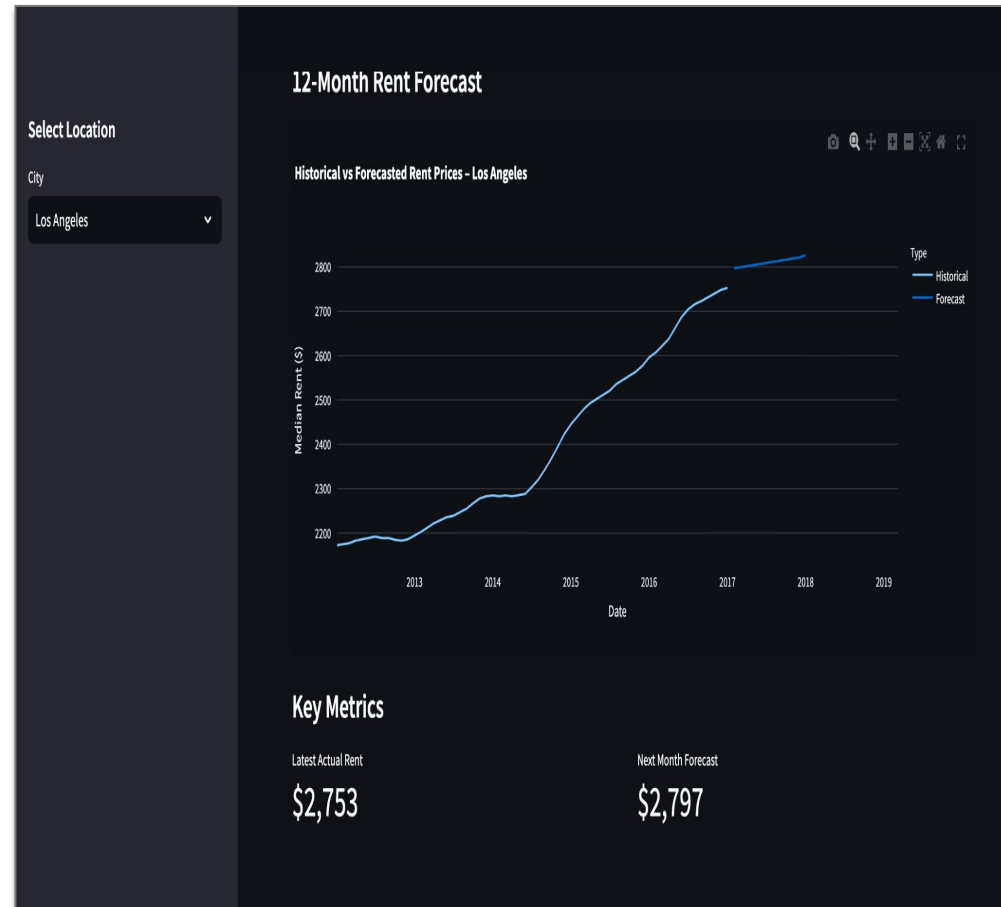
Forecast Visualization

- Produced 12-month rental forecasts for each city.

- Forecasts demonstrate smooth temporal patterns aligned with historical trends and seasonal variations.

Interactive Dashboard

- Developed an interactive Streamlit dashboard to make forecasts accessible.

- Users can select cities, view historical rent trends, and generate 12-month predictions.

- Includes key metrics such as the latest rent value and near-term forecast.

Conclusion

- Successfully developed a forecasting pipeline that transforms historical ZORI data into actionable insights.

- Linear Regression provided the best balance between performance and interpretability.

- The dashboard enhances usability, enabling stakeholders to explore trends with ease.

Future Work

- Integrate external predictors such as economic indicators or housing supply metrics.

- Experiment with advanced forecasting techniques (e.g., Prophet, XGBoost, LSTMs).

- Expand dashboard functionality and consider deployment for broader public use.