FORECASTING REAL ESTATE PRICES USING TIME SERIES ANALYSIS

Project Overview

The analysis will be Utilizing the Zillow housing dataset from April 1996 to April 2018. The project aims to identify top 5 high-potential zip codes for investment.

Business Understanding

Property investment is appealing due to its diverse revenue streams, but the key challenge is pinpointing ideal real estate locations. Our goal is to guide Kar-Dak Investments Group by identifying top-performing zip codes for high ROI in property purchases.

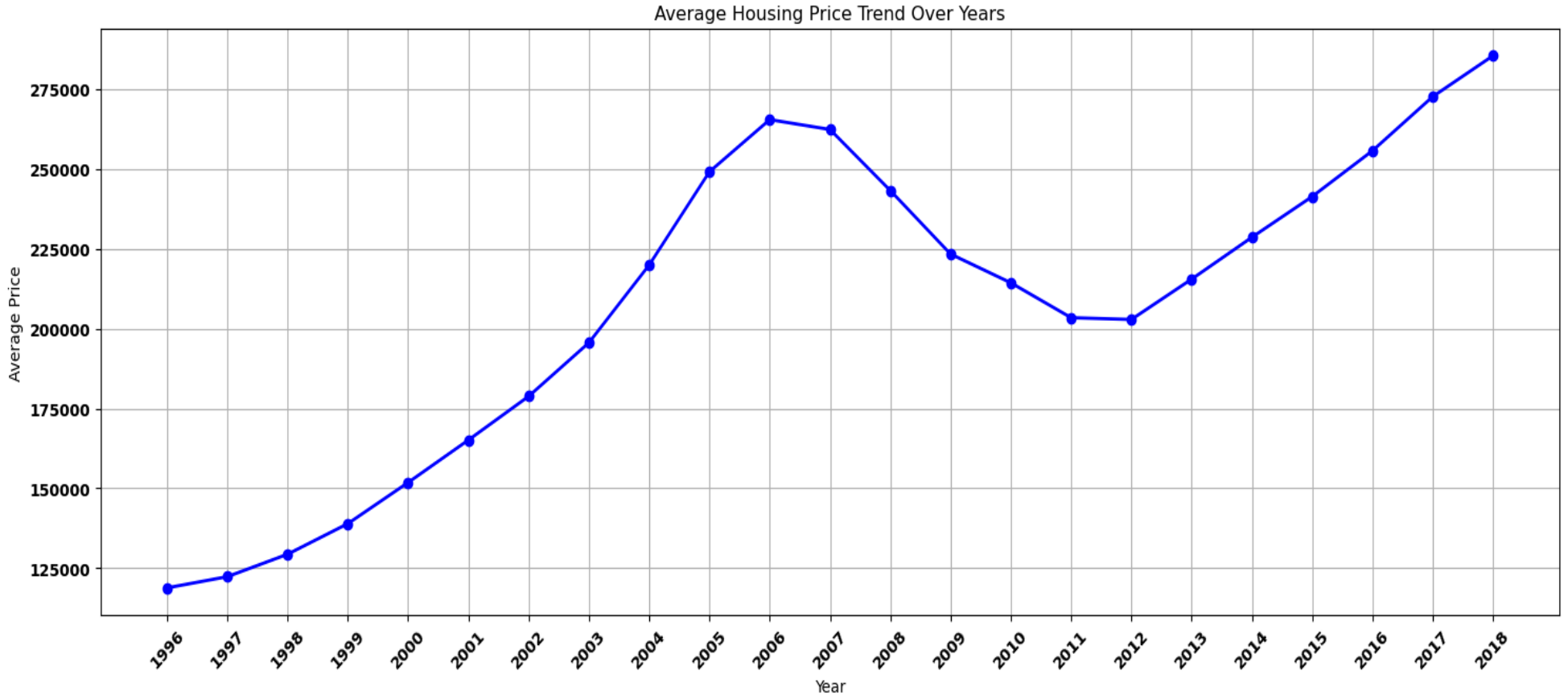
Objectives

- To identify the top 5 zip codes with the highest ROI.
- To develop time series models to forecast real estate prices for different zip codes over various time horizons.
- To establish cities that are optimal for both short-term and long-term investment

Data Understanding

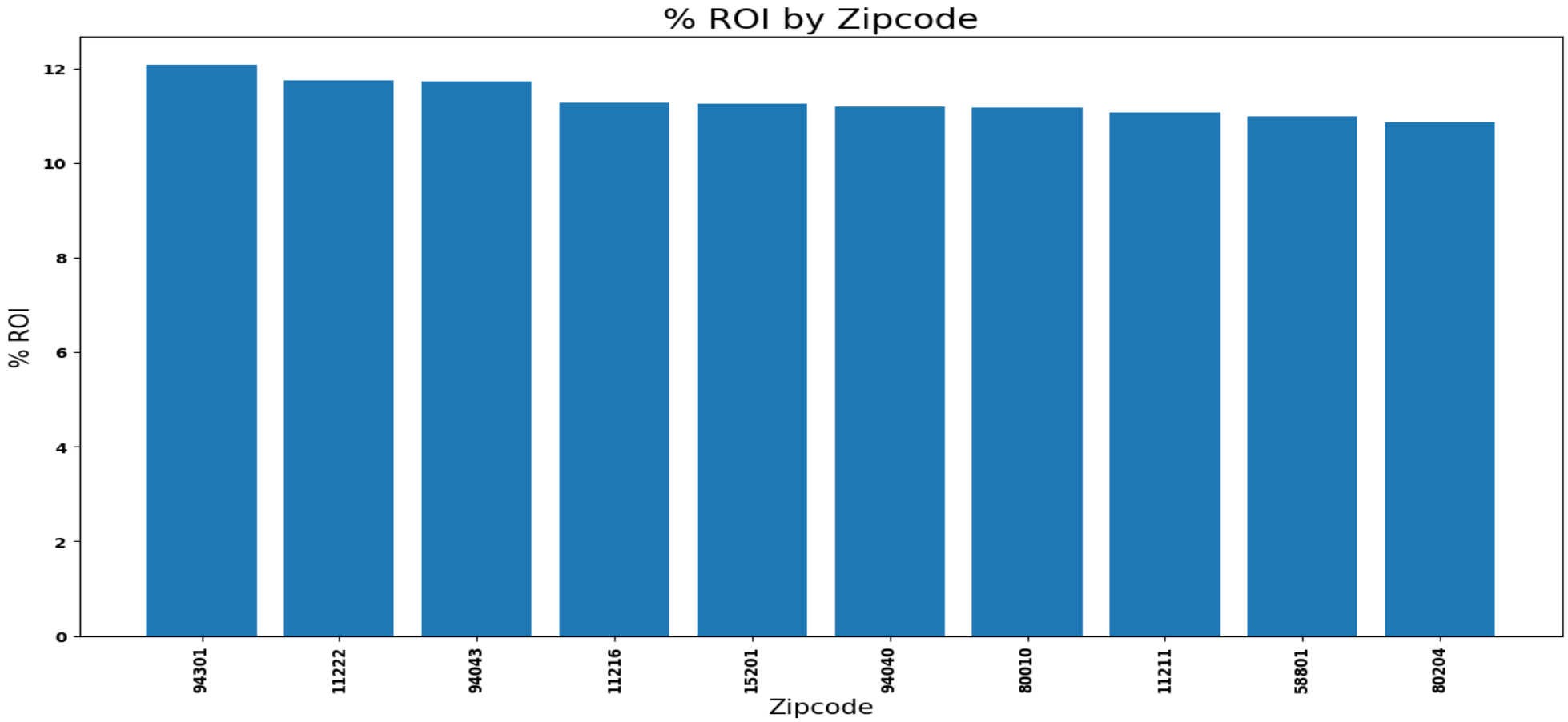
- Zillow Dataset contains information about House Prices between 1996-2018.
- Dataset Overview: Shape(14723 rows, 272 columns)
- Initially wide format with dates from column 8 onwards. Converted to long format for improved analysis.
- The dataset had some missing values in 220 columns .

Exploratory Data Analysis - EDA

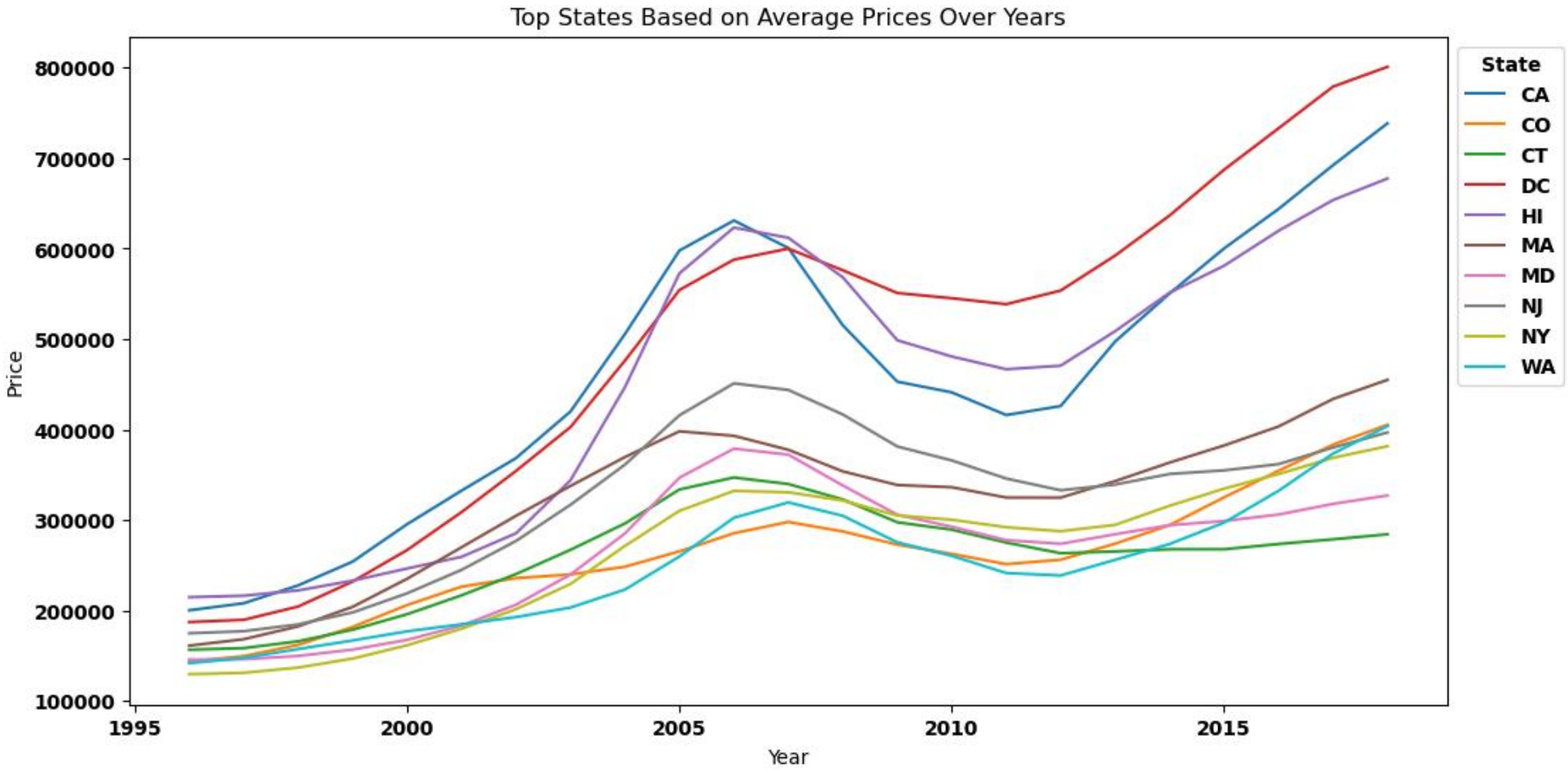


The average housing price has been increasing steadily
over the years, with a dip between 2007 and 2018. However, there was a notable deep from 2008 to 2012

Exploratory Data Analysis - EDA

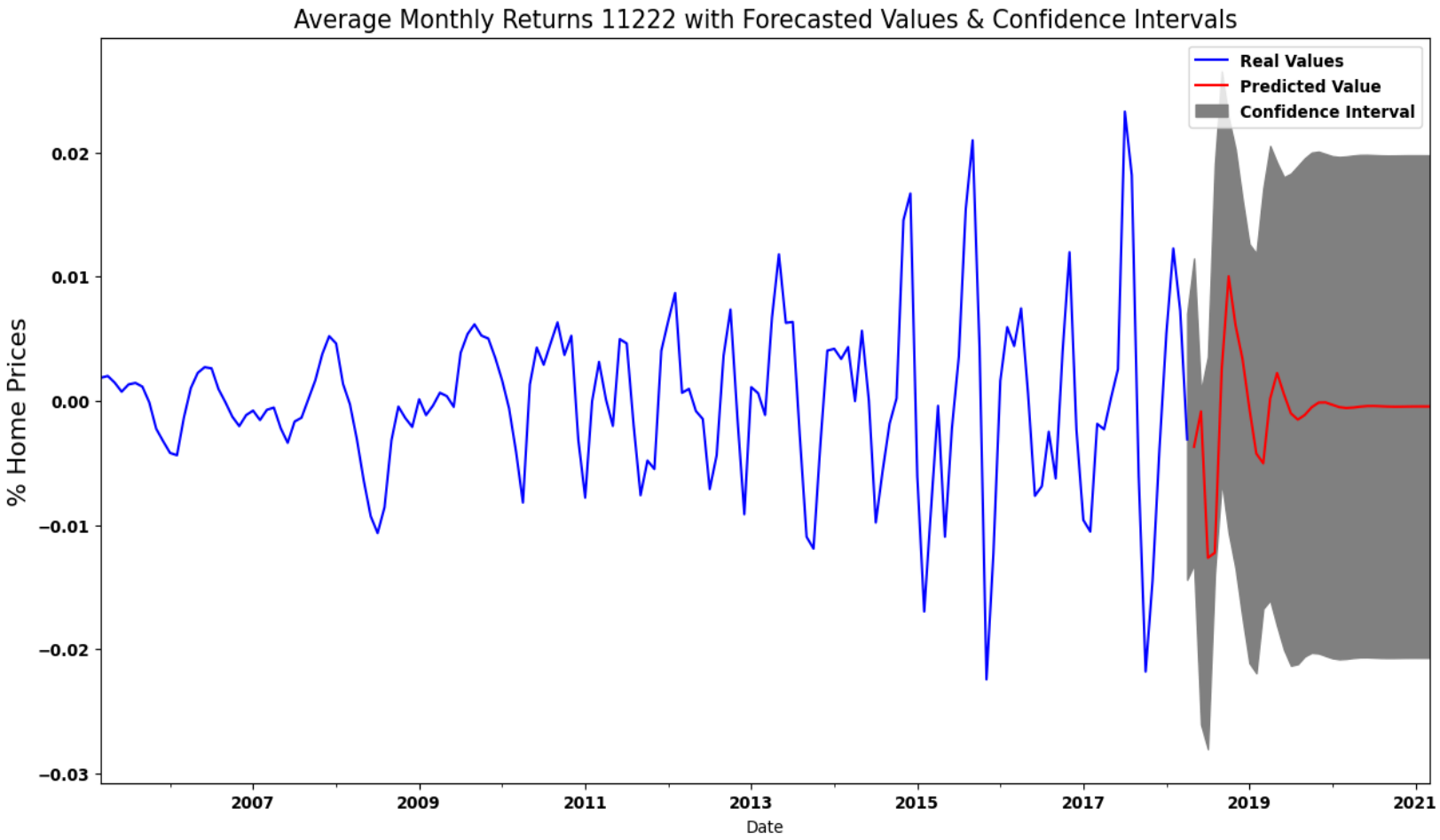


Exploratory Data Analysis - EDA

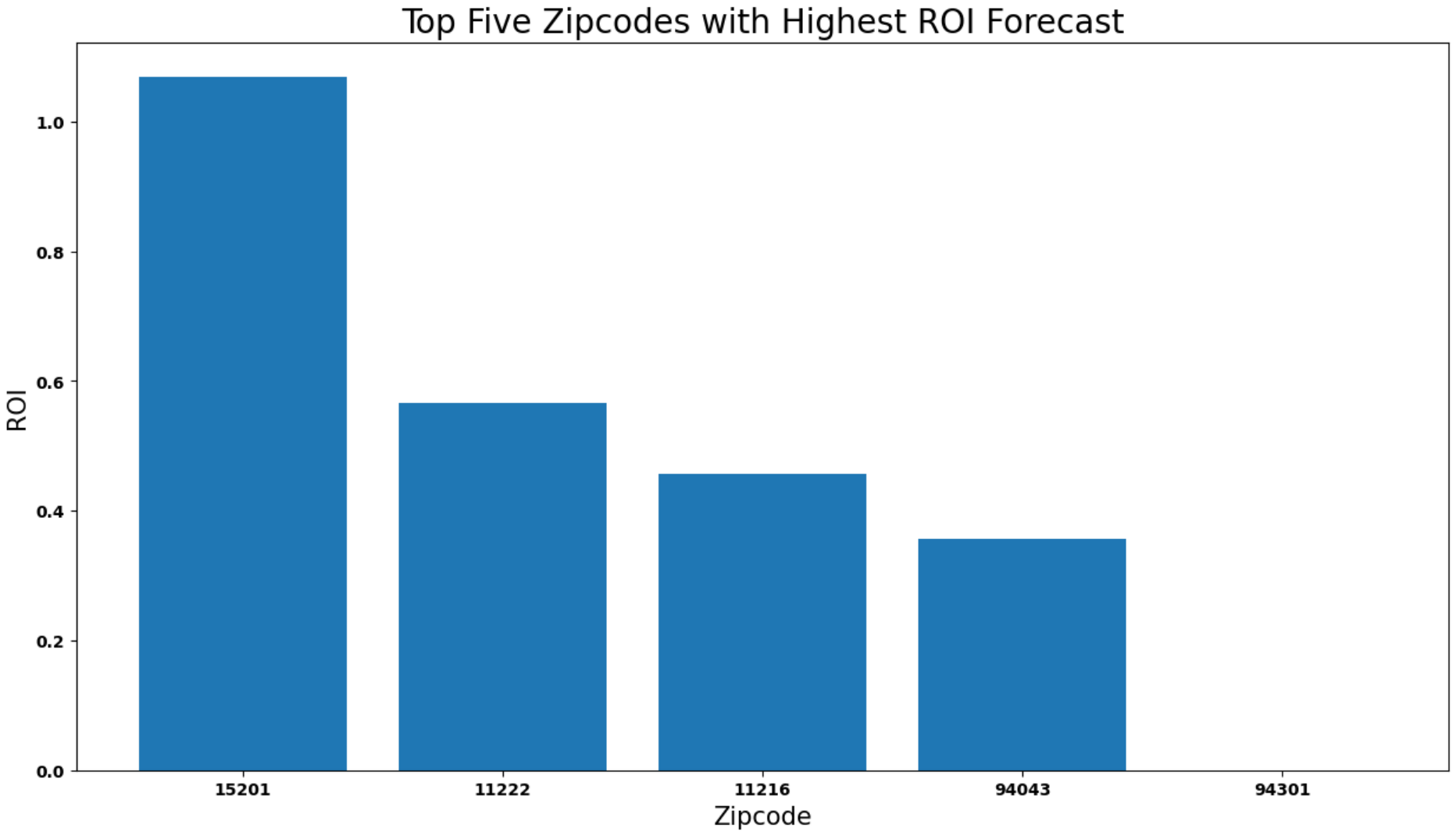


TREY research

Forecast Prediction



Forecast Prediction



Conclusions

Baseline ARIMA Model:

* Strong predictive performance, low RMSE for training and test data.
* Potential overfitting, excelling on training data.
* Clear alignment between actual and predicted values, capturing historical trends effectively.

SARIMA Model:

* Competitive predictive performance, low MSE across test data.
* Balanced parameter selection for simplicity and fitting.
* Well-aligned with historical trends, providing insightful forecasts.

Conclusions

- **Top zip codes:**
- 11216: NY: Kings
- 11222: NY: Kings
- 94043: CA: Santa Clara
- 94301: CA: Santa Clara
- 15201: PA: Allegheny

Recommendations And Next Steps

Diversify for Risk Mitigation

Explore New York

Prioritize High ROI

Evaluate Risk with CV

Consider City Attributes

Recommendations And Next Steps

- ✓ Enhance Dataset with Economic Indicators
- ✓ Optimize Model with Tuning Techniques

Thank You

Github Links

1. https://github.com/Sheila-machaha
2. https://github.com/Phelix-hub
3. https://github.com/Kiprotichemmanuel
4. https://github.com/4kipkorir
5. https://github.com/ihellenmwangi
6. https://github.com/WinnieOnduru
7. https://github.com/