From Past to Future:

A Journey into Predictive Analytics for Investment Success"

Presented By:

Risaneal Obondo

Ian Kedeyie

Prudence Coredo

Timothy Lenku

Marion Jelimo

Agenda

- Project Overview
- Business Understanding
- Data Understanding
- EDA & Modelling
- Conclusion & Recommendations

PROJECT OVERVIEW

Our project delves into the captivating realm of time series analysis, aiming to unravel the temporal dynamics that govern housing market trends.

Through this exploration, we seek not only to understand the past but to unveil insights that empower strategic decision-making for future investments.

BUSINESS UNDERSTANDING.

REAL ESTATE INVESTMENT FIRMS WEIGH ON MANY FACTORS WHEN DECIDING ON WHICH INVESTMENT TO MAKE.

THESE FIRMS' GOALS ARE TO ENHANCE THEIR DECISION-MAKING PROCESSES. WITH HISTORICAL REAL ESTATE MARKET DATA.

THE CHALLENGE IS TO LEVERAGE DATA SCIENCE TECHNIQUES TO IDENTIFY POTENTIAL INVESTMENT OPPORTUNITIES AND HELPING THE REAL ESTATE FIRMS IN MAKING INFORMED INVESTMENT DECISIONS

PROBLEM STATEMENT.

A Real Estate Investment Firm wants to know the top 5 best zip codes to invest in. As a Data Science consulting group, we have been tasked with finding the results.

The task at hand is to create a model that will inform us on the Real Estate investment market trends for the next 10 years.

Time series analysis.

OBJECTIVES

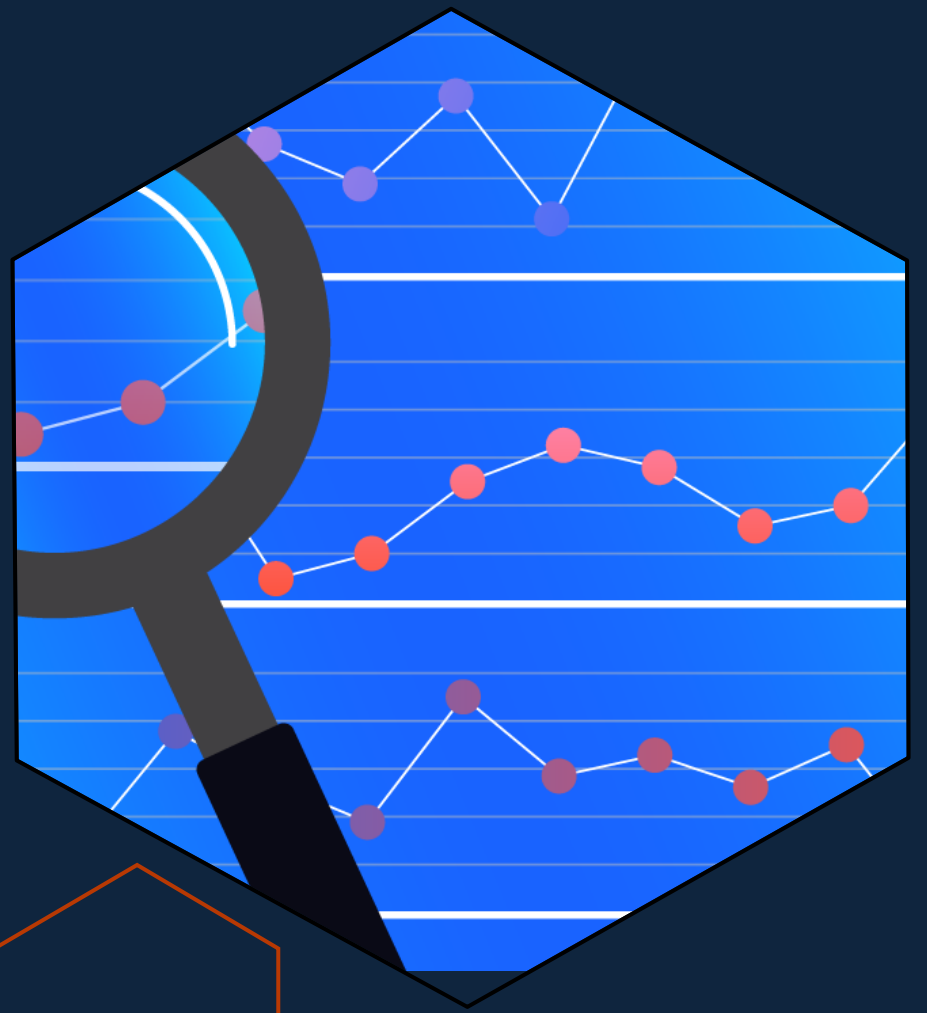
TO OPTIMIZE INVESTMENT PERFORMANCE THROUGH HOLISTIC DATA-DRIVEN STRATEGIES

TO SELECT THE 5 BEST ZIP CODES THAT OFFER BEST INVESTMENT OPPORTUNITIES.

TO PREDICT THE AVERAGE PRICE FOR THE TOP 5 ZIP CODES.

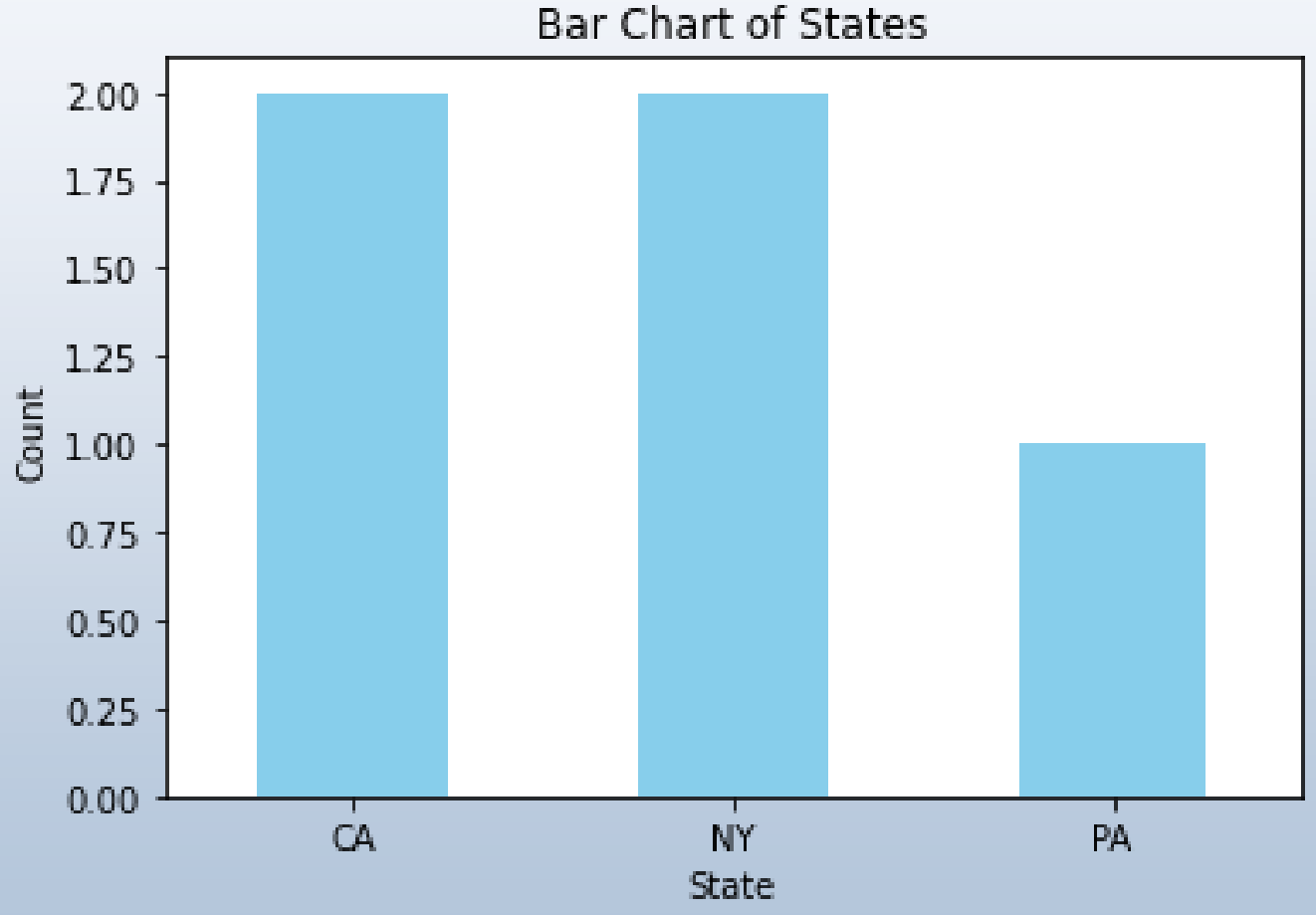
DATA UNDERSTANDING

- Data collected from the Zillow website. The dataset covers house prices from April 1996 to April 2018.
- It contains 14723 rows and 272 columns. Each row represents an individual property.
- Unnecessary columns such as RegionID, Metro were dropped during the data cleaning.
- Our working dataset had features that include; Zip Codes, City, State, Time and Price.
- The data set is seen in Wide format hence conversion to Long format is done for analysis.

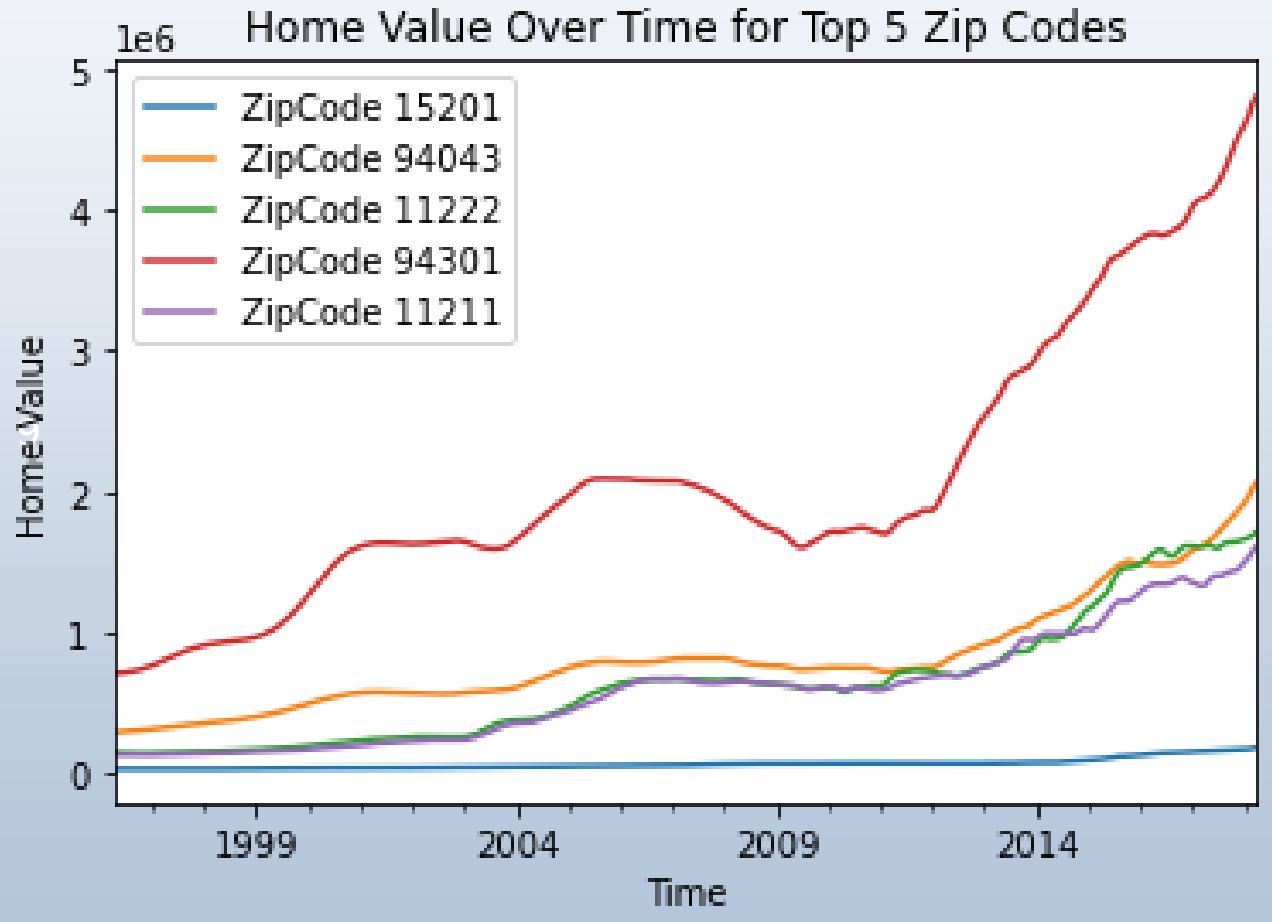


EXPLORATORY DATA ANALYSIS

- This chart shows the states where these best 5 zip codes are found as seen with the ROI (Return On Investment) and CAGR (Compound Annual Growth Rate) metrics.
- Two zip codes are found in California (CA) and New York (NY), while Pennsylvania (PA) has one zip code.

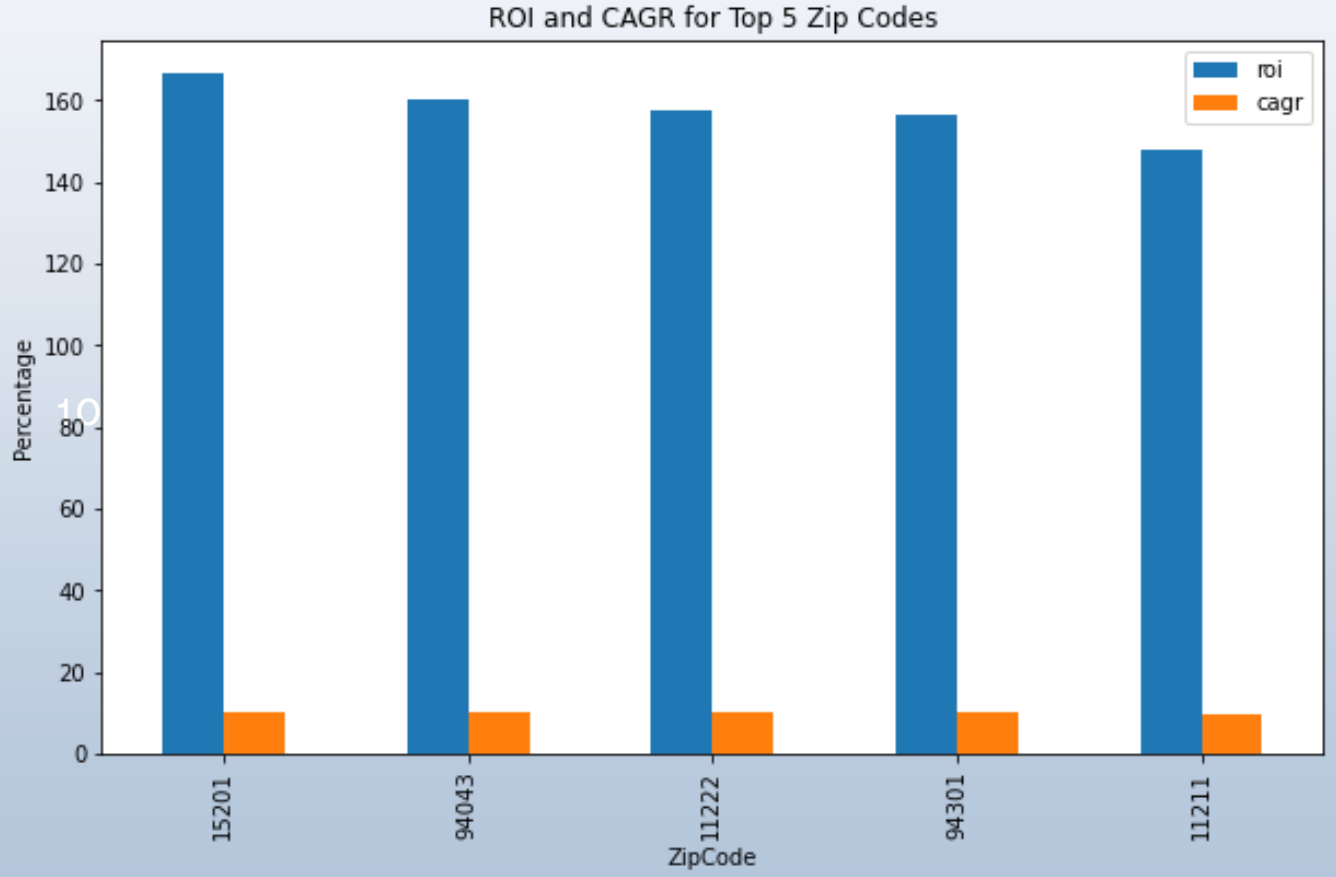


EXPLORATORY DATA ANALYSIS



- The time series line plot shows how home values have evolved over time for the selected zip codes. The zip code with the highest home value is 94301 located in the state of California.
- The top 5 zip codes were determined by the ROI values

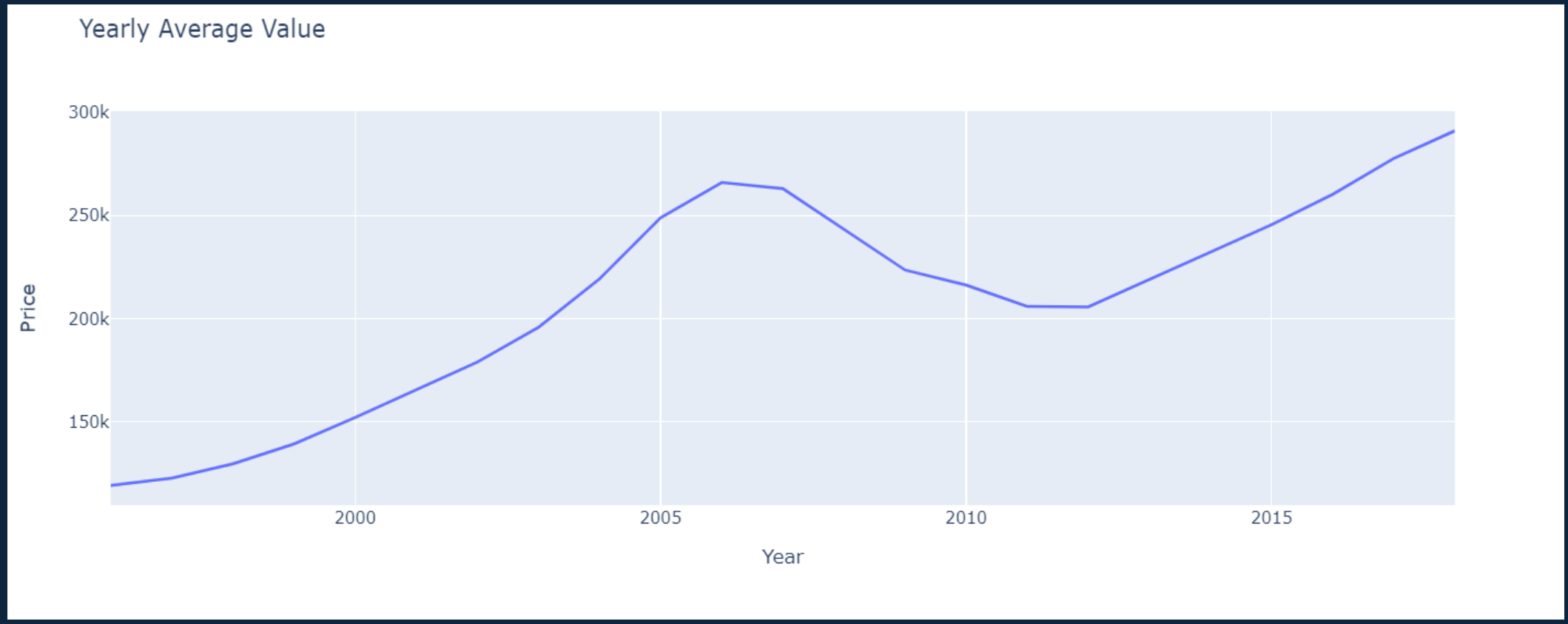
EXPLORATORY DATA ANALYSIS



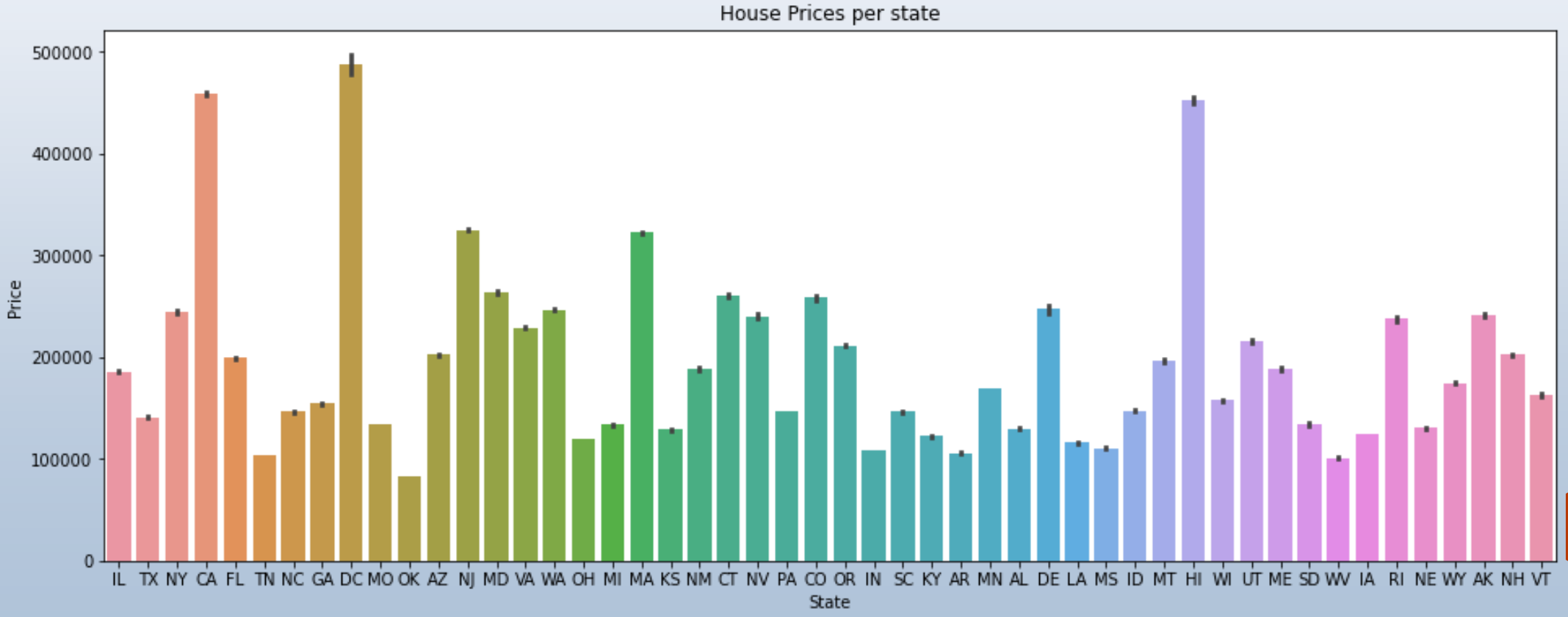
- The plot shows the ROI against CAGR for the best 5 zip codes.
- The property's ROI is significantly high compared to the CAGR.
- ROI is a measure of percentage increase of a property's value compared to the initial investment value; It increases greatly compared to the CAGR.
- However, CAGR may give a more stable representation of the property's overall performance.

EXPLORATORY DATA ANALYSIS –Time Series Line Plot.

The plot shows how the average home values change annually.



The chart shows average house price per state. Washington DC has the most expensive houses on average compared to other states. Following California.
Oklahoma has the least expensive houses.



Modelling

Four models were used but settled on the final one which seemed to be best in predicting.

The best model was Autoregressive which gave the lowest RMSE value.

Evaluation

Evaluated the performance of the best model and achieved a RMSE score of 102.09 and MAE of 73.32

CONCLUSIONS

	Investing	
In summary, the analysis highlights attractive investment prospects in identified zip codes and states, suggesting the potential for informed decisions. The outlined next steps, focusing on advanced models and collaborative strategies, pave the way for ongoing refinement and adaptation in response to the dynamic real estate landscape.		Promising zip codes for investment include 15201 in Pittsburgh, PA, and 11222 in Brooklyn, NY, driven by ongoing development and demand. California's tech-centric areas like 94043 (Mountain View) and 94301 (Palo Alto) present strong housing prospects, while diverse economies make California, New York, and Pennsylvania appealing states for real estate investment.

RECOMMENDATIONS

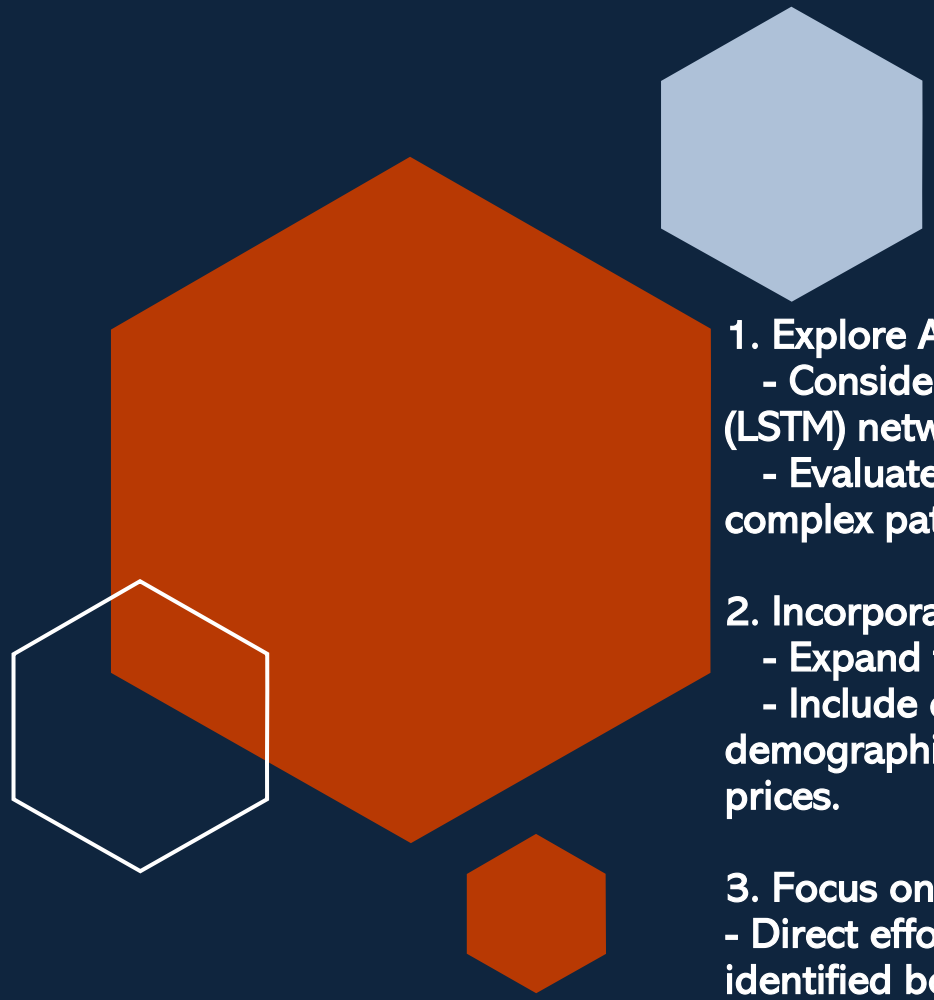
THE BEST STATES TO INVEST IN

- California
- New York
- Pennsylvania

BEST ZIPCODES TO INVEST IN

- 15201
- 94043
- 11222
- 94301
- 11211

NEXT STEPS



1. Explore Advanced Models:
   - Consider implementing more advanced models such as Long Short-Term Memory (LSTM) networks to potentially enhance predictive performance.
   - Evaluate the suitability of LSTM for time series forecasting, as it can capture complex patterns and dependencies over time.

2. Incorporate Additional Datasets:
   - Expand the dataset by incorporating more features beyond the current variables.
   - Include datasets that encompass a broader range of economic indicators, demographic factors, or other relevant variables that could influence real estate prices.

3. Focus on Best Zip Codes:
- Direct efforts towards modeling and analyzing real estate prices specifically for the identified best-performing zip codes (e.g., 15201, 94043, 11222, 94301, 11211). - Tailor models to the unique characteristics and dynamics of these high-potential areas for more precise and targeted predictions.

Thank you