INVESTING IN REAL ESTATE

Prepared by Group 10

GROUP 10 MEMBERS

- Crystal Wanjiru
- Miriam Nguru
- Paul Mbugua
- Kelvin Rotich
- Stephen Butiya
- Celiajoy Omiah

OUTLINE

INTRODUCTION
PROBLEM STATEMENT
OBJECTIVES
EXPLORATORY DATA ANALYSIS
MODELING
RECOMMENDATIONS
CONCLUSIONS

INTRODUCTION

In this real estate investment project, our goal is to empower investors with a sophisticated time series model using Zillow's dataset. Through advanced forecasting techniques, we aim to providing investors with accurate insights for strategic decision-making.

PROBLEM STATEMENT

Real estate investment stands as a profitable and ever-evolving industry, demanding meticulous analysis and strategic decision-making. A real estate investment firm is currently in search of insights to pinpoint the top zip codes offering promising investment opportunities. To tackle this inquiry, we leverage historical data sourced from Zillow Research

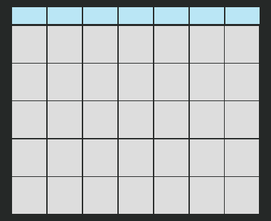
OBJECTIVES

* To identify the top 5 zip codes and states that offer the best investment potential in terms of real estate value. By analyzing historical trends and patterns, the project aims to provide actionable insights to the investment firm, enabling them to make informed decisions on where to allocate their resources.

* To analyze the historical data of the real estate value by looking into the monthly, quarterly, semi-annual and annual patterns over time.

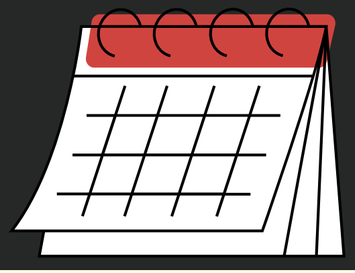
* To create an ARIMA model that will be able to predict future Real Estate Value.

Data Understanding



14723 rows of data

272 columns, with some as the house values for every region



The date values range from 1996 April to 2018 April

The rest of the columns are RegionID, RegionName which is zipcode, State, City, and SizeRank

DATA PREPARATION

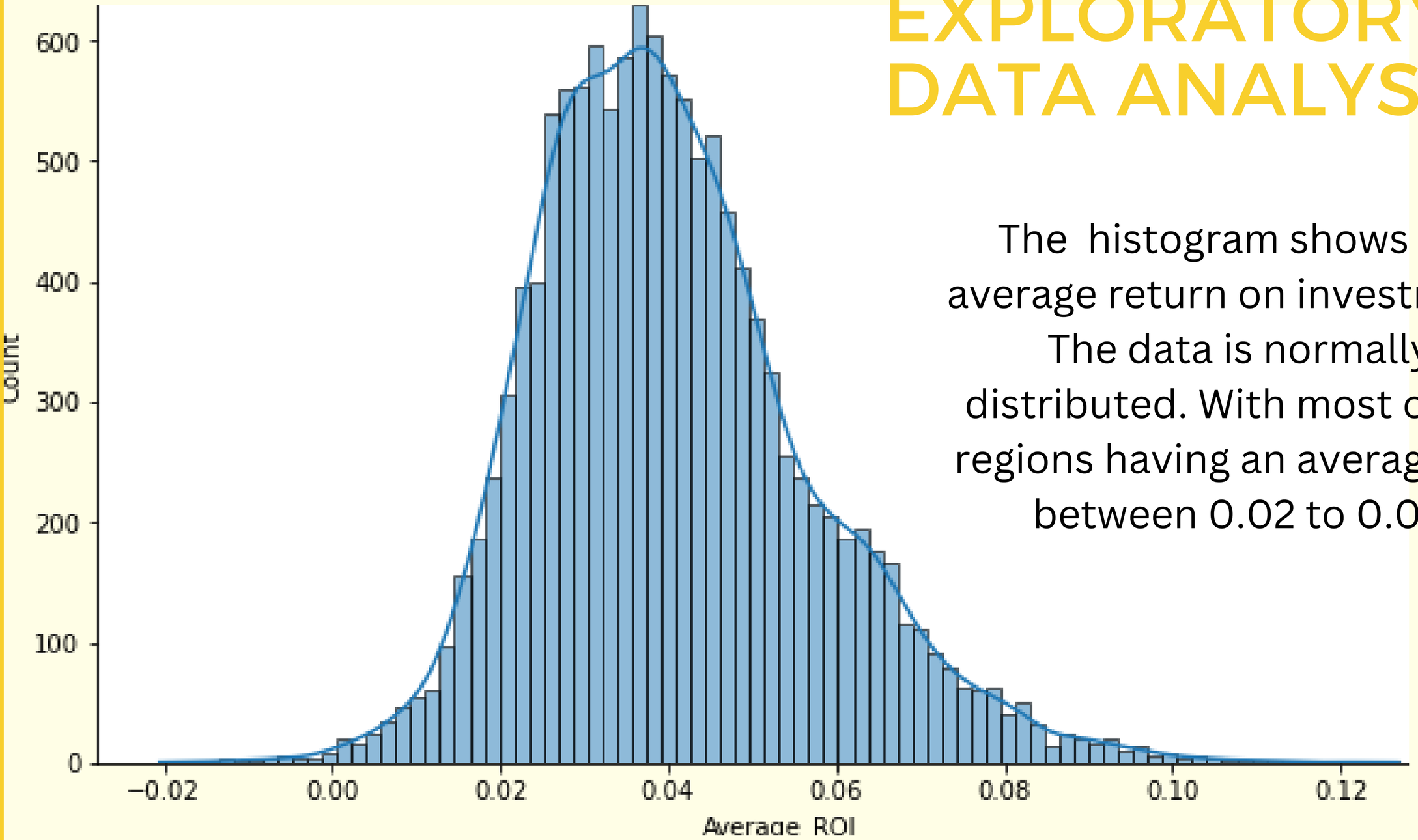
For the data preparation we dropped all the missing values leading to our data reducing from 14723 rows to 13680 rows

The data had no duplicate values but had outliers that we chose to keep because these were recorded events

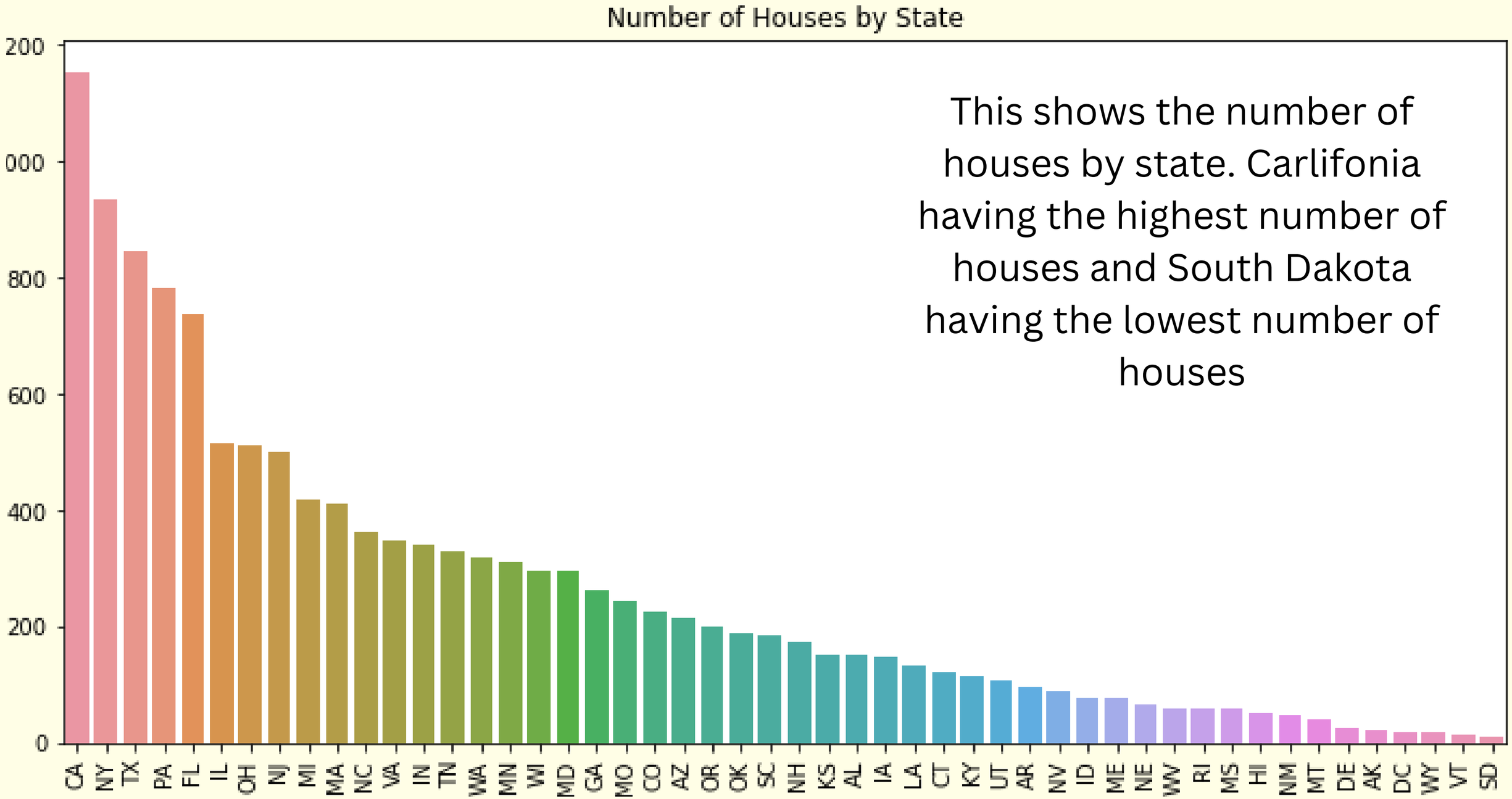
We created two Data Frames, one to do Exploratory Data analysis and the other a melted Data Frame for time series analysis and modeling

We also created a Returns on Investment column, from all the previous years

EXPLORATORY DATA ANALYSIS



The histogram shows the average return on investment . The data is normally distributed. With most of the regions having an average ROI between 0.02 to 0.06



This shows the number of houses by state. Carlifonia having the highest number of houses and South Dakota having the lowest number of houses

The zipcodes with the highest Return on investment are:

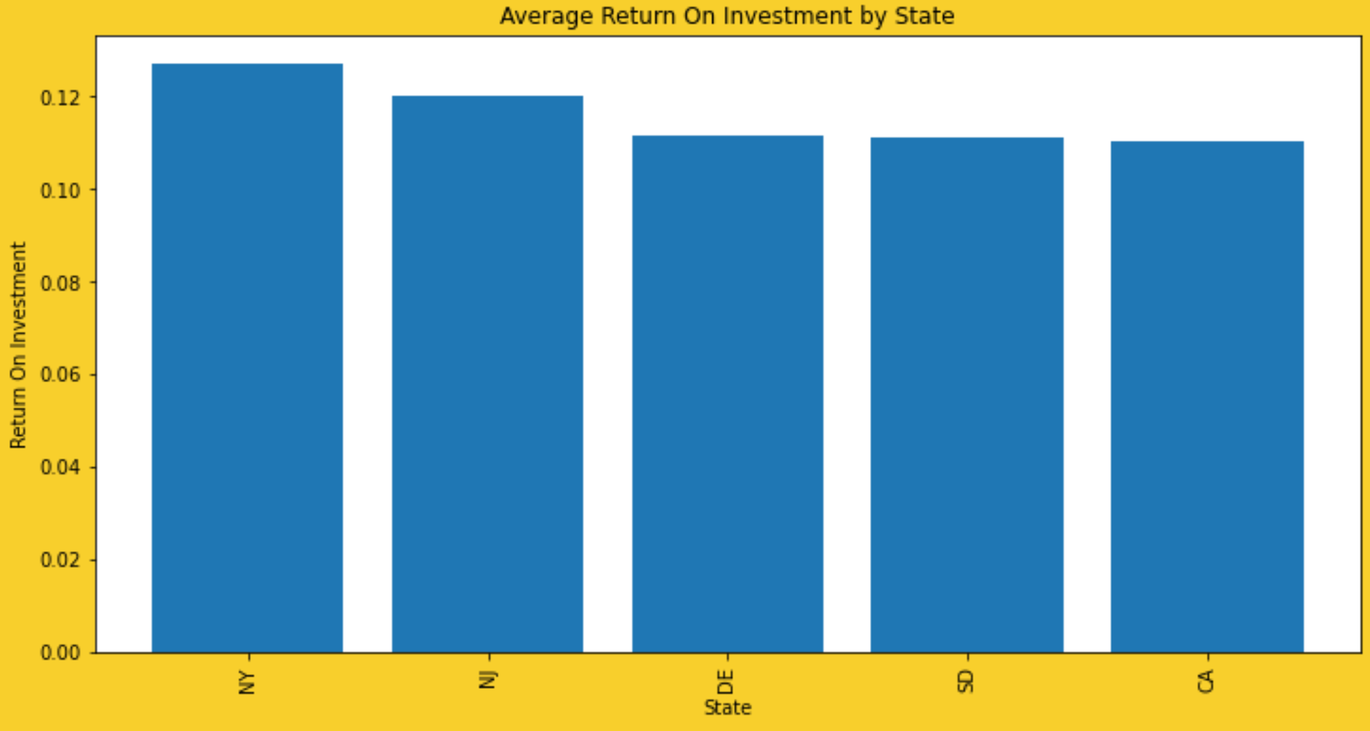
11211 - Brooklyn, New York

11222 - Brooklyn, New York

11216 - Brooklyn, New York

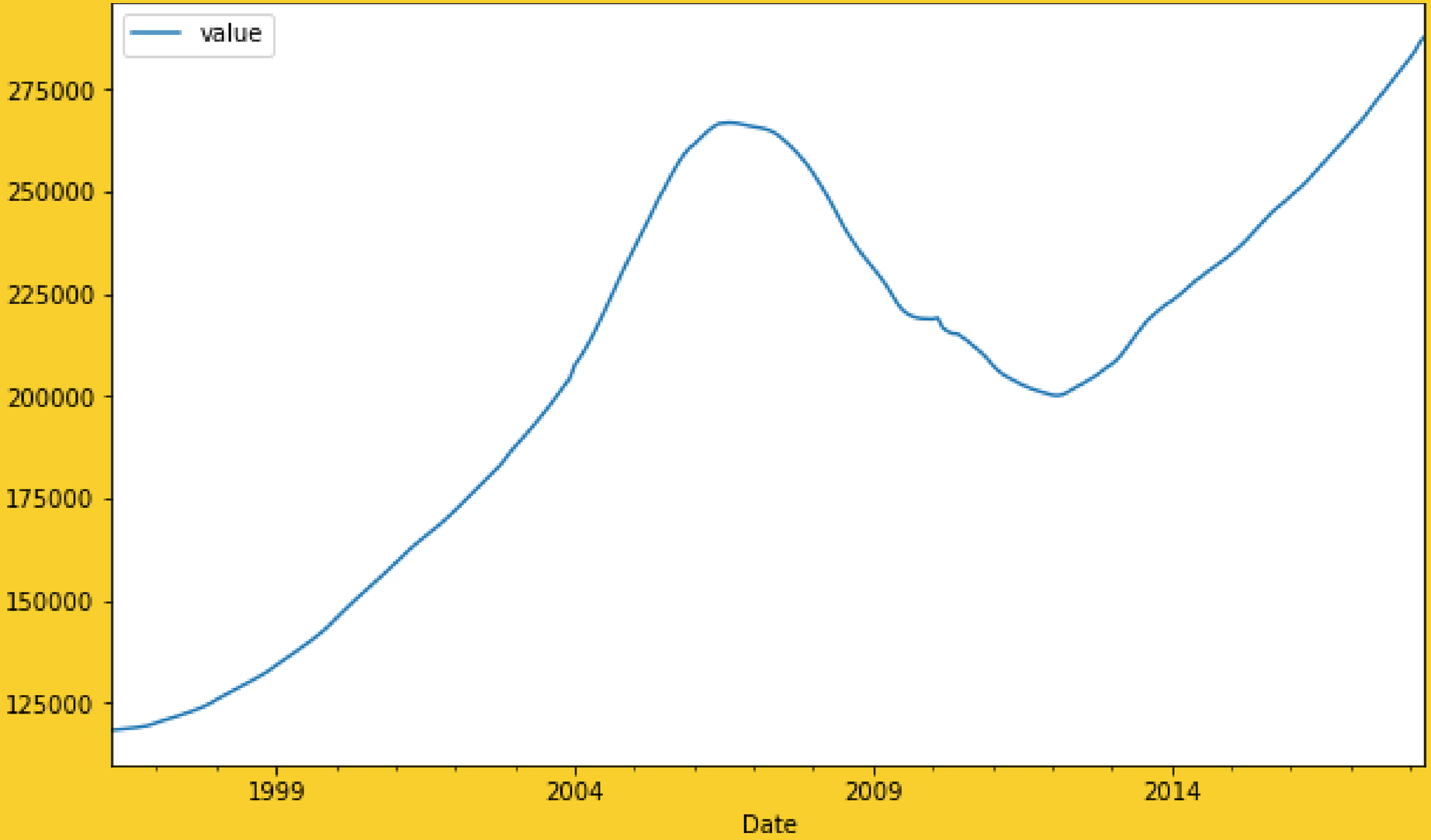
7302 - Jersey City, New Jersey

11930 - Amagansett, New York



The states with the highest Returns on investment are: New York, New Jersey, Delaware, South Dakota, and Califonia.

Time Series Analysis



The plot shows an upward trend of house values over time. Except for the fall between 2006 and 2013 that can be attributed to a recession and the market crash of 2008, The plot shows that the house values are appreciating over time

Modeling

ARIMA

	BASELINE MODEL	2ND ARIMA MODEL	3RD ARIMA MODEL	4TH ARIMA MODEL
RMSE	121.94	122.23	122.24	117.55
MAE	85.97	86.27	86.18	84.47

RMSE and MAE measure the amount of error that the model makes when predicting. From the above, the model with the least amount of error is the fourth model.

Prophet

	1ST PROPHET MODEL	2ND PROPHET MODEL	3RD PROPHET MODEL
RMSE	142.35	122.51	121.87
MAE	95.67	87.13	86.22

We also built another model, Prophet and the model with the lowest RMSE and MAE is the third model.

Between the ARIMA and Prophet models, The fourth ARIMA model performed best overall

RECOMMENDATIONS

REAL ESTATE

We recommend investing in Real Estate, the data showed an upward trend, indicating increasing house values over time

STATES AND REGIONS TO INVEST IN

we recommend investing in the following states, New York, New Jersey, Colorado, Carlifornia and Washington DC, from the analysis these states showed promising Returns on Investment. The best zipcodes were found within the states mentioned, these are, 11211 - Brooklyn, New York, 11222 - Brooklyn, New York, 11216 - Brooklyn, New York, 7302 - Jersey City, New Jersey and 11215 - Brooklyn, New York.

FORCASTING MODEL

As a way to mitigate risk we recommend using the model created to forecast future values of Real Estate.

CONCLUSION

From the project we were able to build a predictive model that will assist investors looking to invest in Real Estate in the United States.

Next steps

To collect more data on Real Estate Values - more data will better inform the model and lead to better predictive results.

Continuous model training to improve accuracy

THANK YOU!

LET‘S START YOUR INVESTMENT JOURNEY