

# REAL ESTATE Housing ● Market Forecasting.

---

A Moringa Group 3 Project



# TABLE OF CONTENTS



01

## Introduction

Includes the business problem and overview.

02

## Objectives

Business understanding and our goals for this project.

03

## Data Understanding

Exploring our dataset.

04

## Data Analysis

Diving deeper on the contents of our dataset.

05

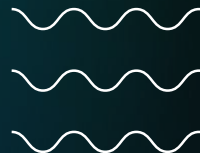
## Modeling and Forecasting

Includes graphs and models from our findings.

06

## Conclusion

Recommendation and next course of action.





01



Introduction.

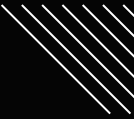
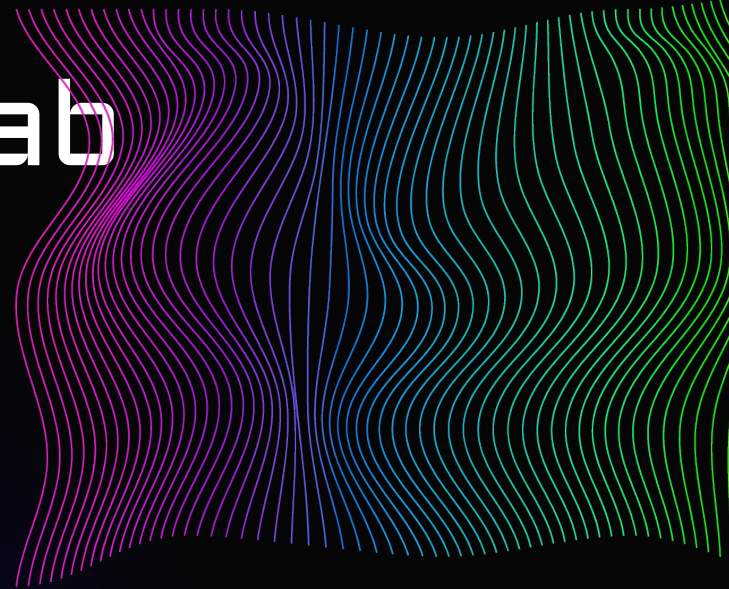
---





# Realestate datalab

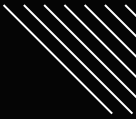
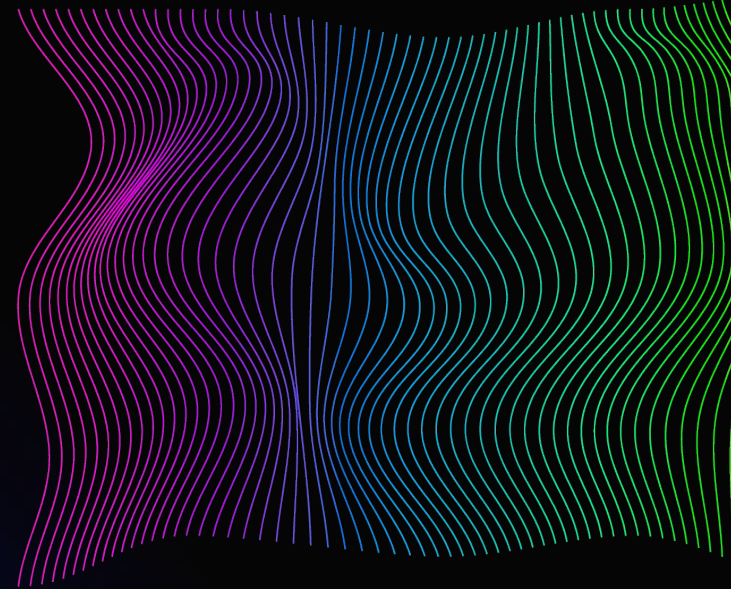
Realestate DataLab & KenyaData insights is a cutting-edge data firm based in Kenya, focused on interpreting real estate projections. Our expertise lies in dissecting these forecasts to understand their impact on housing prices. We leverage advanced forecasting models to capitalize on market opportunities, driving informed decisions in the real estate sector.





# Business Problem

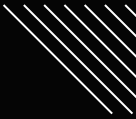
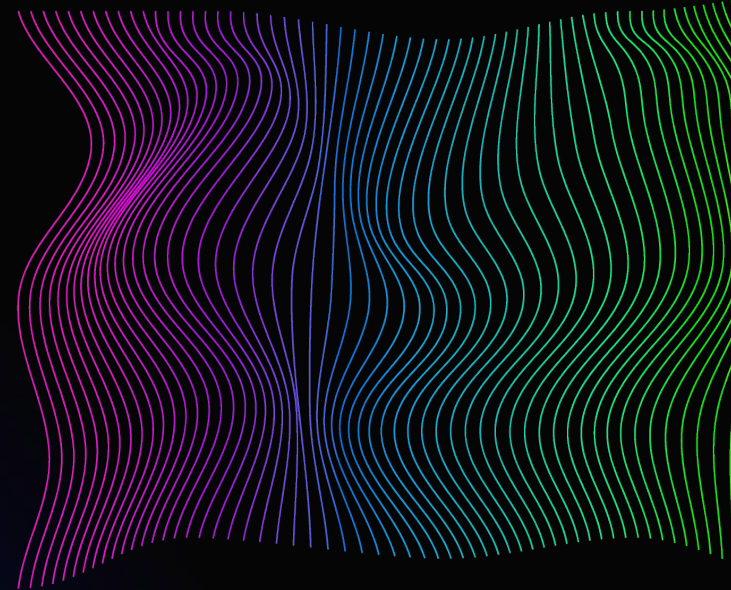
The core challenge lies in the limited insight and grasp of the US housing market among Kenyan investors. Furthermore, there's a necessity to pinpoint regions with potential investment prospects within the affordable housing sector. Amidst market upheavals like the 2008 housing crash, maintaining prediction accuracy and relevance poses another hurdle. Thus, the project's goal is to offer exhaustive data analysis and forecasting, facilitating informed investment decisions, and mitigating uncertainties.



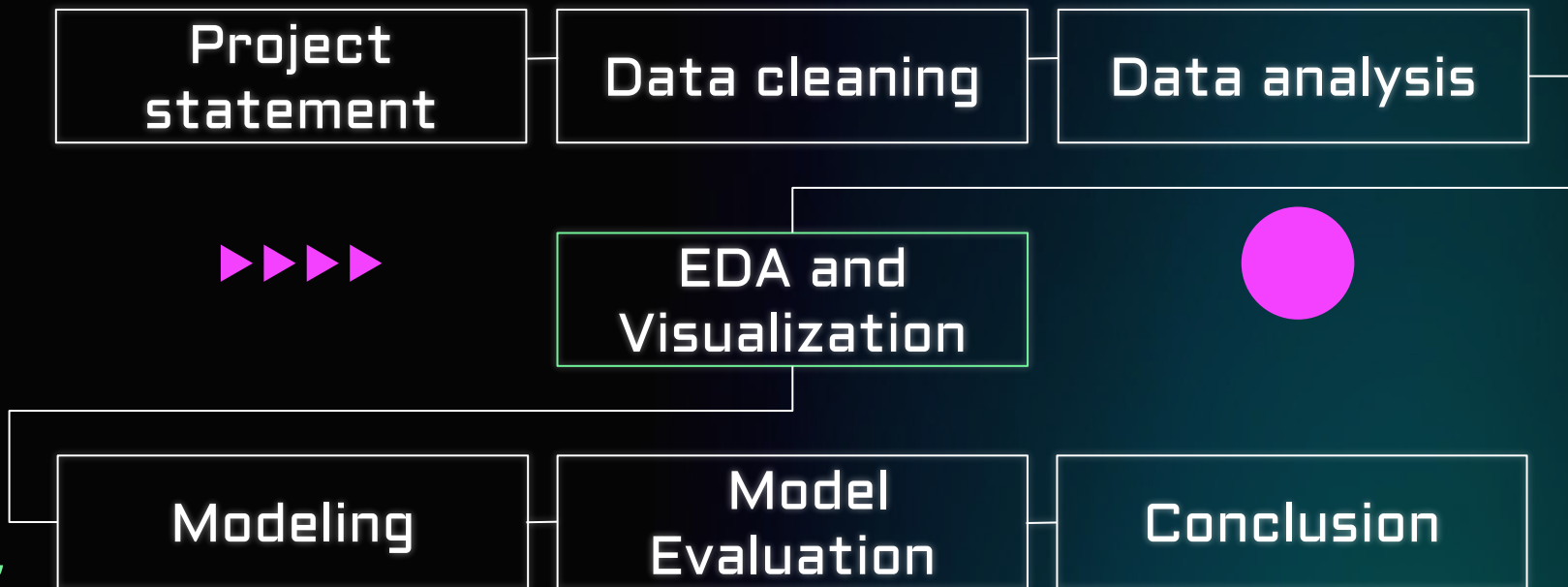


# Overview

Our client, a leading real estate investment firm in Kenya, facilitates local and diaspora investors' access to the US housing market. Specializing in affordable housing, they empower investors with strategic insights and data analysis, fostering financial inclusion and contributing to Kenya's economic growth.



# DATA PROJECT ARCHITECTURE





# Objectives

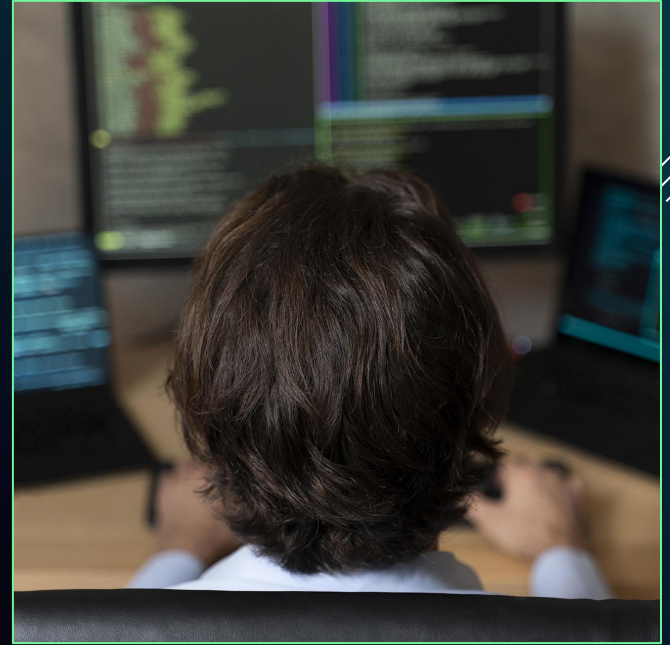
02





# BUSINESS OBJECTIVES

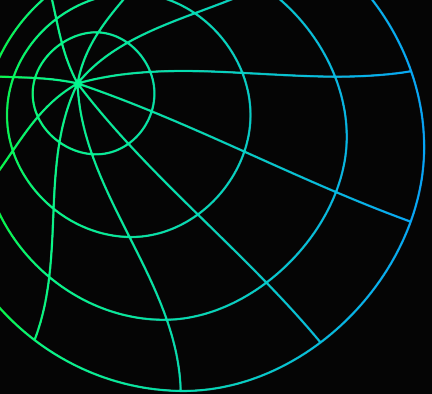
- Developing an accurate time series forecasting model using Zillow data.
- Providing market trend analysis to identify promising investment opportunities.
- Offering actionable investment strategies based on model insights.
- Enhancing the decision-making process for buyers, renters, and investors.
- Strengthening the investment portfolio of our clients through data-driven insights.





# STAKEHOLDERS

- Kenyans looking to invest in the US housing market
- Kenyan diaspora planning to buy or rent properties in the USA
- Real estate investment firms
- Financial analysts focusing on real estate investments
- Policy makers interested in foreign investments



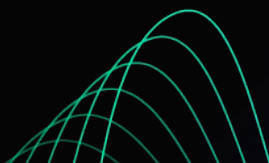
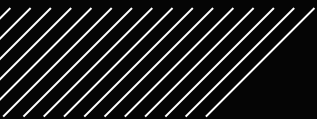
03



Data

Understanding.



---





# DATA SOURCE

Our analysis is based on `zillow_data.csv`, a dataset from Zillow containing historical housing prices across US regions from April 1996 to April 2018. It includes region details and monthly prices, helping us understand trends and forecast future market trends.





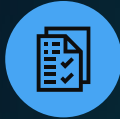
# FACTORS AFFECTING REAL ESTATE



## PRICE



CITY



STATE



TIME





# Data Analysis and visualization

---

04





# DATA PREPARATION

We took some preprocessing steps to ensure we had data that was easy to work with:

## Data Melting:

- The process of melting the data involves restructuring the dataset to make it more suitable for analysis. This typically entails converting columns into rows, which can help in performing various analyses.

## Handling Null Values:

- Null values, or missing data points, can hinder accurate analysis. Dealing with null values involves strategies such as imputation or removal to ensure data integrity and reliability.

## Adding ROI (Return on Investment):

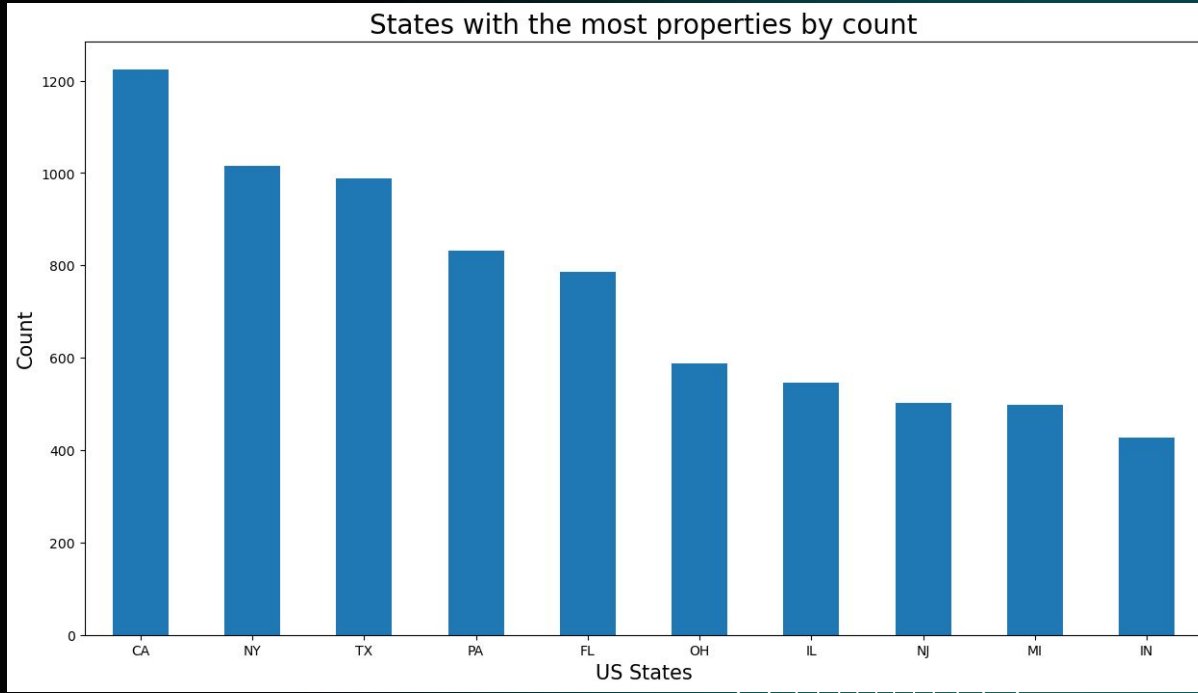
- Introducing ROI to the dataset allows us to assess the profitability of investments. ROI calculates the ratio of profit to the initial investment, providing valuable insights into investment performance.

## Incorporating Price Ranges:

- Price ranges offer a broader perspective on housing market dynamics, enabling us to categorize properties based on their price points. This segmentation helps in identifying market trends and target audiences effectively.

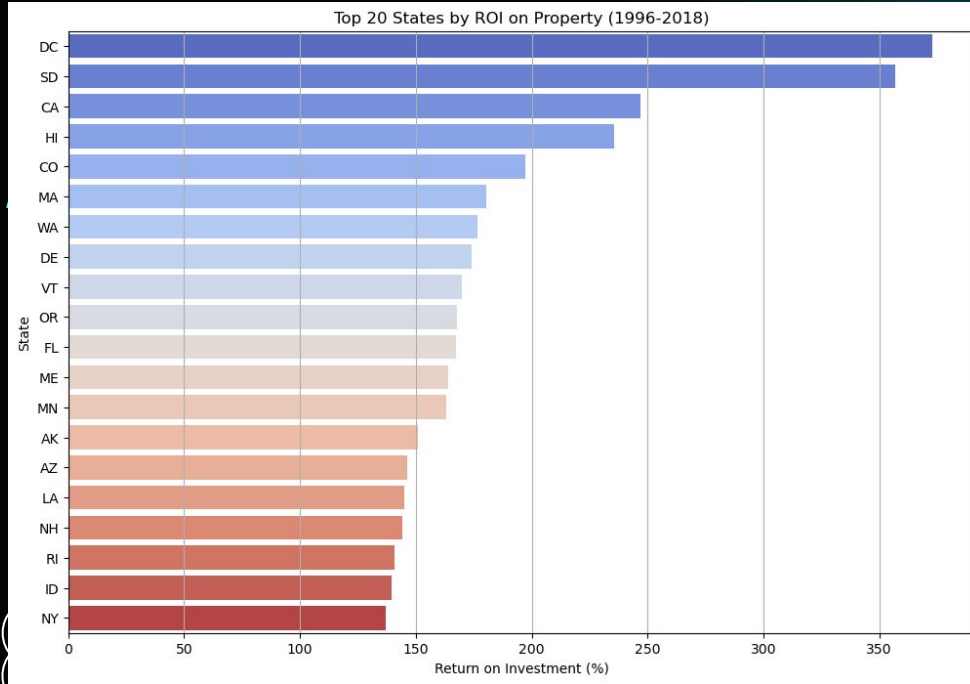


# Top States with the most properties





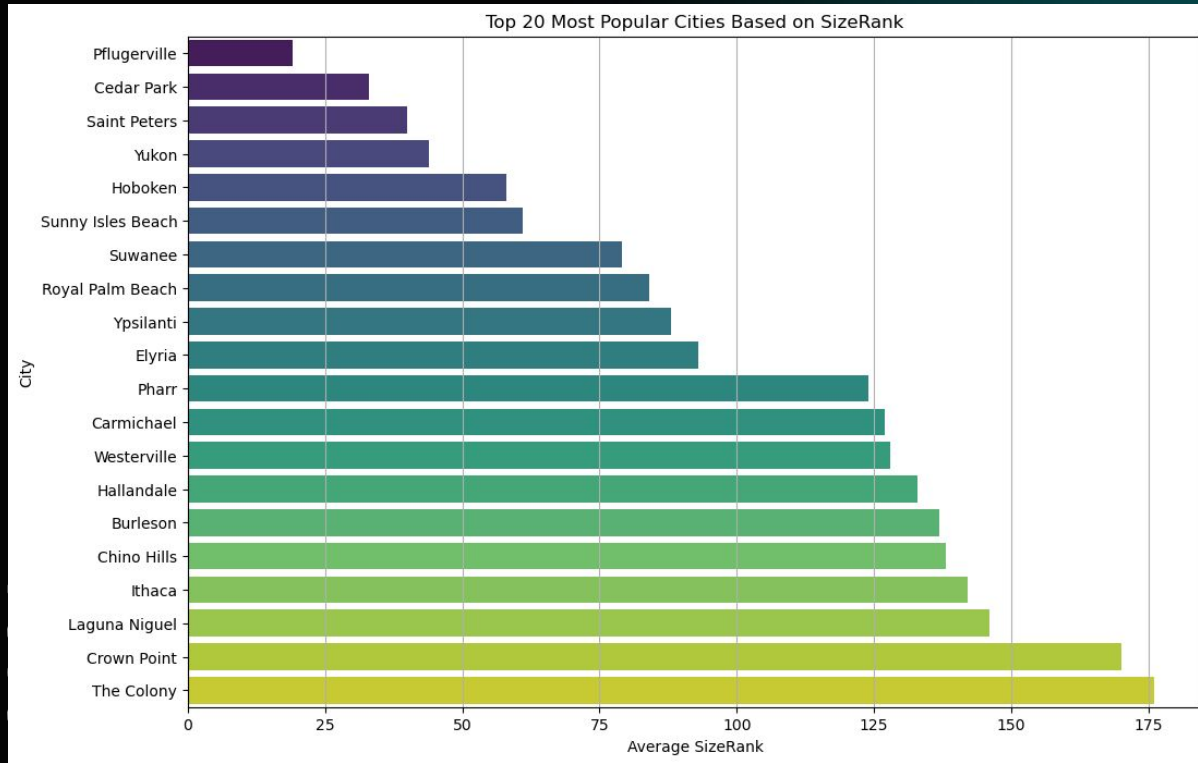
# Top 20 States by ROI on property



- ❖ A look at the top 20 states by ROI
- ❖ DC, SD and CA are the top 3 states in terms of ROI with well over 200%



# Top 20 Most popular cities



- ❖ The Colony, Crown Point and Laguna Niguel with well over 125% average size rank



05



# Modelling and Forecasting

---

# DATA Preparation and Modelling

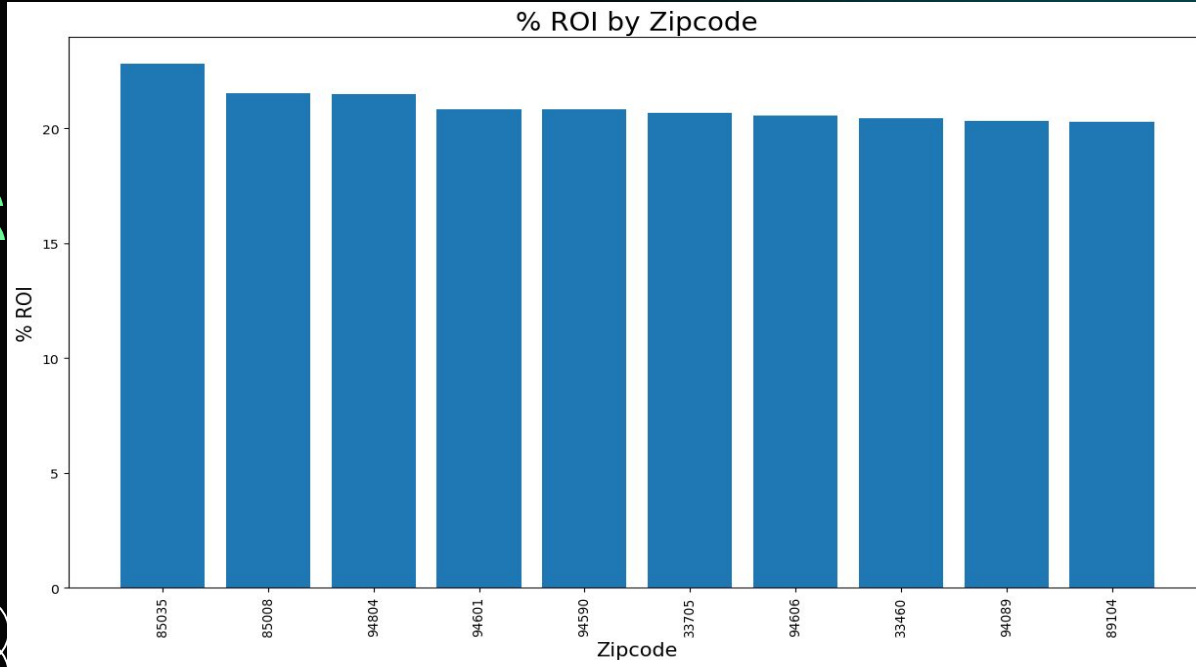
In this section we are going to get a better understanding about our data, what is going to be modelled and what we expect to model.

For this project we are going to be using Arima and Auto Arima models to helping forecasting data about the year and the price of properties in order to make predictions on the future prices.

In order to effectively and efficiently predict we are going to focus only on the top performing Zipcodes in terms of ROI



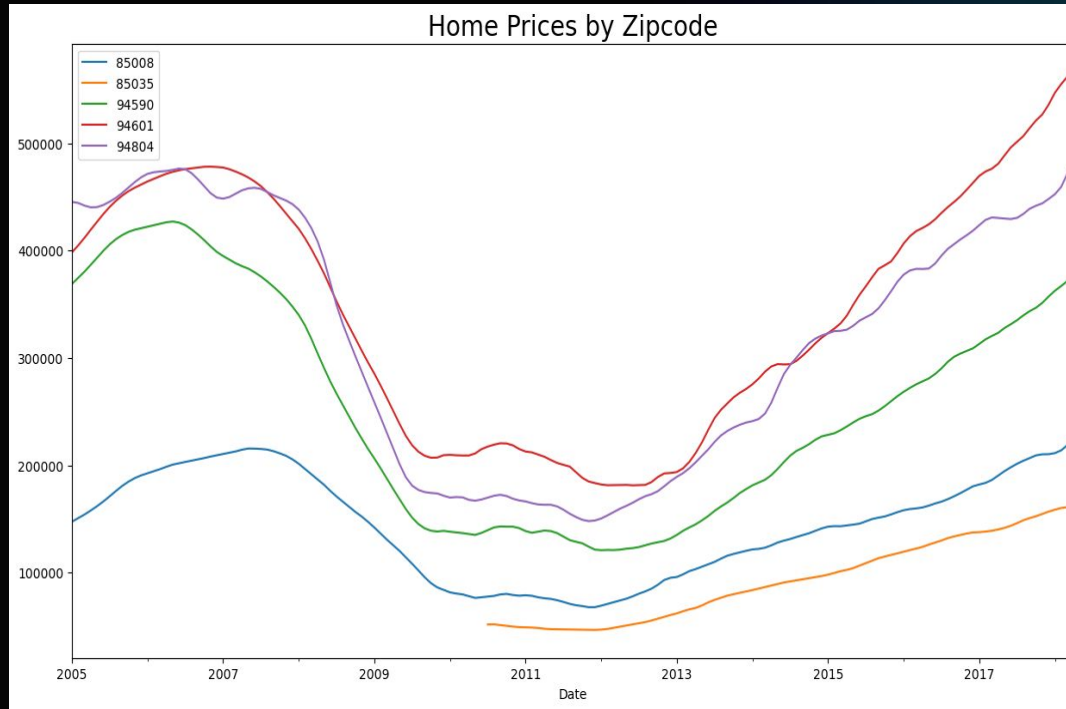
# Top performing Zipcodes in terms of ROI



The figure above shows the Zipcodes with the highest percentage Return On Investment(%ROI) in the country.

From this graph, zipcode 85035 seems to be the most profitable zipcode at 22.8% ROI from 2012 to 2018.

# Average house prices by zipcode



So in this case we are going to create individual time series of the zipcodes and plot them for a better understanding

It can be observed like before that there was a dip in the prices due to the 2008 market crash and then the prices continued to rise over time from 2012 all the way to 2018.

There is a drop after 2008 all the way to 2012 due to the financial crisis






# Model building

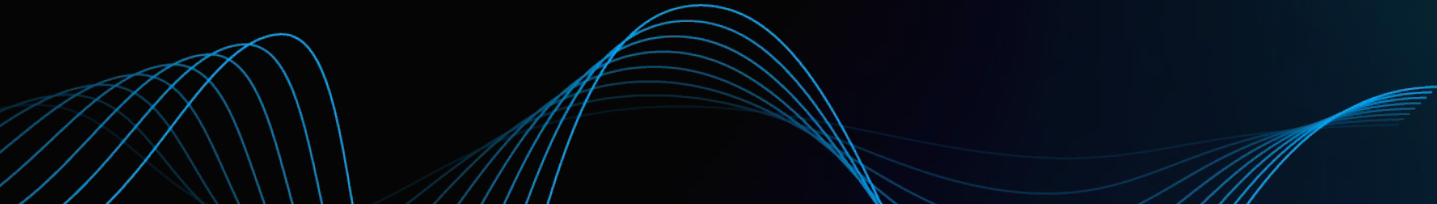

We first check for stationarity in all the zipcodes' time series by using rolling mean several times

Dickey Fuller test is also used in verifying stationarity

We then plot the acf and pacf figures to give us a better understanding of the data.



We then build our base models using auto arima on all time series data frames and we get the best order and parameters and we then use this to build our final model using `pd.arima`



# MODEL EVALUATION

Average Monthly Returns - 85035 - With Forecasted Values & Confidence Intervals



- ❖ The graph's forecasted values (red line) closely track the actual observed values (blue line), indicating the model effectively captures underlying data patterns. These accurate forecasts are vital for informed investment decisions and understanding market trends in the 85035 area.

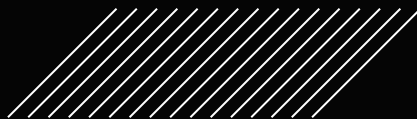




# Conclusion

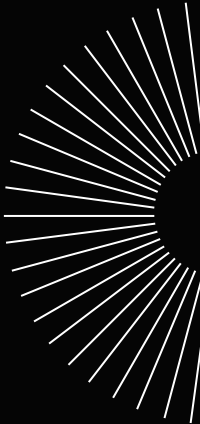


06



# Conclusions & Recommendations:

Our analysis identified several promising zip codes for investment, each with distinct advantages. Zip code 94804 stands out for its consistent rise in housing prices and strong ROI potential. Conversely, zip code 85035 should be avoided due to its negative ROI forecast. Through rigorous data analysis, we tailored recommendations to meet our client's investment objectives, emphasizing diversification and maximizing returns. To ensure sustained success, we recommend regular review of investment strategies and staying attuned to market dynamics. By leveraging data-driven insights, our client can navigate the dynamic US real estate market with confidence and precision, securing long-term profitability and growth.



# Conclusions & Recommendations:

**Zip code 94804 = Richmond California:** This area's housing prices have been steadily rising, and a high return on investment is expected. Its stable market dynamics and favourable price trends make it a good fit for our client's investment portfolio.

**Zip code 75217 = Dallas Texas** With its promising combination of affordability and appreciation potential, this area represents an excellent opportunity for our client. Our analysis shows a positive trajectory in housing prices, indicating a high ROI potential.

**Zip code 19143 = Kingessing Philadelphia County:** This zipcode demonstrates resilience in the face of market fluctuations, with consistent growth and promising investment opportunities. Its affordability and upward price trends make it an appealing option for our client looking for long-term returns.


**Zip code 60628 = Roseland Chicago Illinois:** This area has strong growth potential despite fluctuations in the overall market, especially in low-cost housing segments. Given the current favourable market conditions and anticipated growth, it is highly recommended for our client's investment plan.

**Zip code 48227 - Wayne County Detroit, Michigan:** This area offers our client an appealing investment opportunity because of its stability in the market and affordability. It has the potential to yield substantial returns over time due to its steady increase in housing prices and bright future prospects.

After that, the investor has the option to invest in any of the zip codes listed above, with the exception of 85035, which offers a negative return on investment.


# Future work & Deployment

Future work in this domain could focus on several key areas:



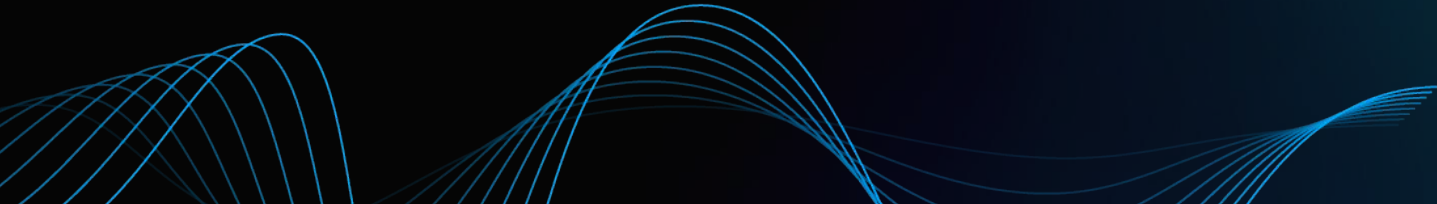

1. Expanding the dataset: Incorporating additional data sources in new Geographical areas, expand economic indicators, demographic trends, and policy changes, can provide a more comprehensive understanding of the factors influencing housing prices and enhance the accuracy of the forecasts.

2. Refining the modeling approach: Exploring advanced time series modeling techniques, such as SARIMA (Seasonal ARIMA) or machine learning algorithms like Long Short-Term Memory (LSTM) networks, could potentially improve the predictive power of the models and capture more complex patterns in the data.

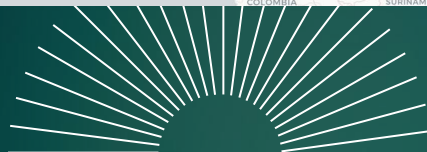
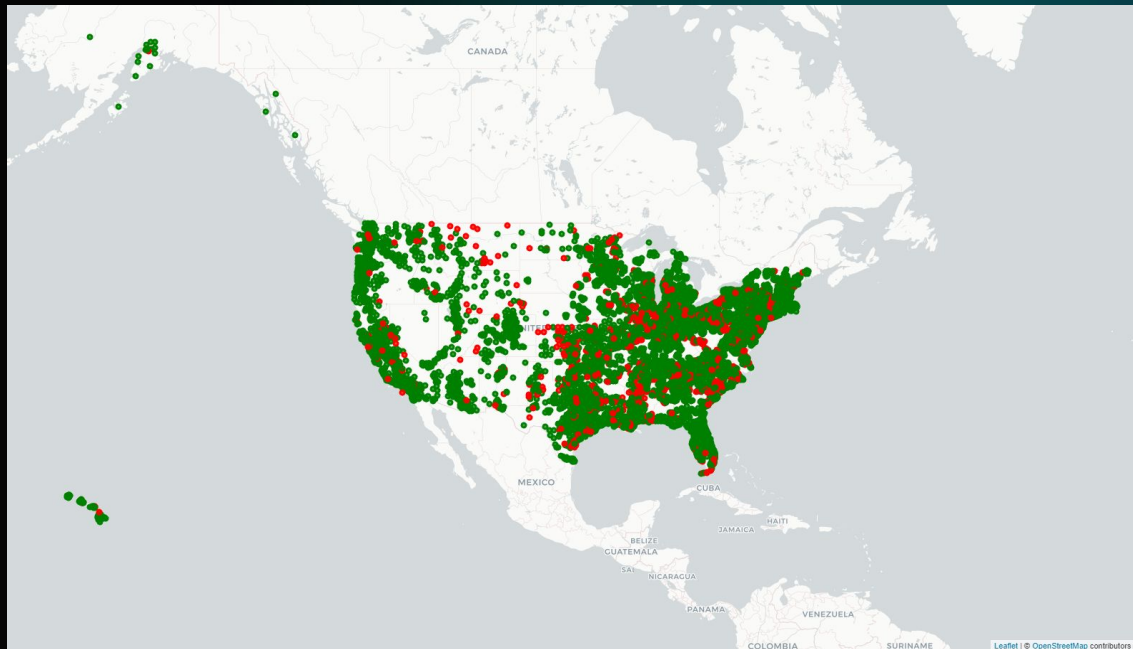


3. Developing a real-time monitoring system: Implementing a system that continuously updates the analysis with the latest housing market data and generates automated alerts for significant changes or investment opportunities could help the investment firm stay ahead of the curve and make timely decisions.

4. Conducting ongoing performance evaluation: Regularly assessing the performance of the recommended investments against the forecasted trends and ROI projections can help validate the effectiveness of the data-driven approach and identify areas for improvement



# US Zip Codes with Positive ROI





THANK  
YOU!!