

ANALYZING REAL ESTATE INVESTMENTS IN THE UNITED STATES

GROUP VII



Business Problem



The project seeks to identify the most advantageous zip code regions for American Properties Corporation's investment, underlining the significance of choosing areas with strong growth potential to optimize returns.

Project Objectives

- Define criteria for evaluating “**BEST**” investment zipcodes.
- Uncover key patterns, trends, and relationships within the data.
- Develop predictive time series model to forecast real estate prices per zipcode
- Provide investment recommendations.



Data Understanding



Dataframe Structure

- ❖ 14,722 rows
- ❖ 272 columns

Columns:

- ❖ RegionID,
- ❖ RegionName,
- ❖ City,
- ❖ State,
- ❖ Metro,
- ❖ CountyName,
- ❖ SizeRank
- ❖ The rest are dates

Data types:

- ❖ 219 : float columns.
- ❖ 49 : integer columns.
- ❖ 4 : object columns.

Missing values present
in 220 columns

Dataset consumes :
30.6 megabytes of
memory

No duplicates were
identified in any of the
columns

Data Cleaning



Extracting the datetimes in the columns.

Removing redundant columns. ie RegionID

Detecting and addressing missing values.

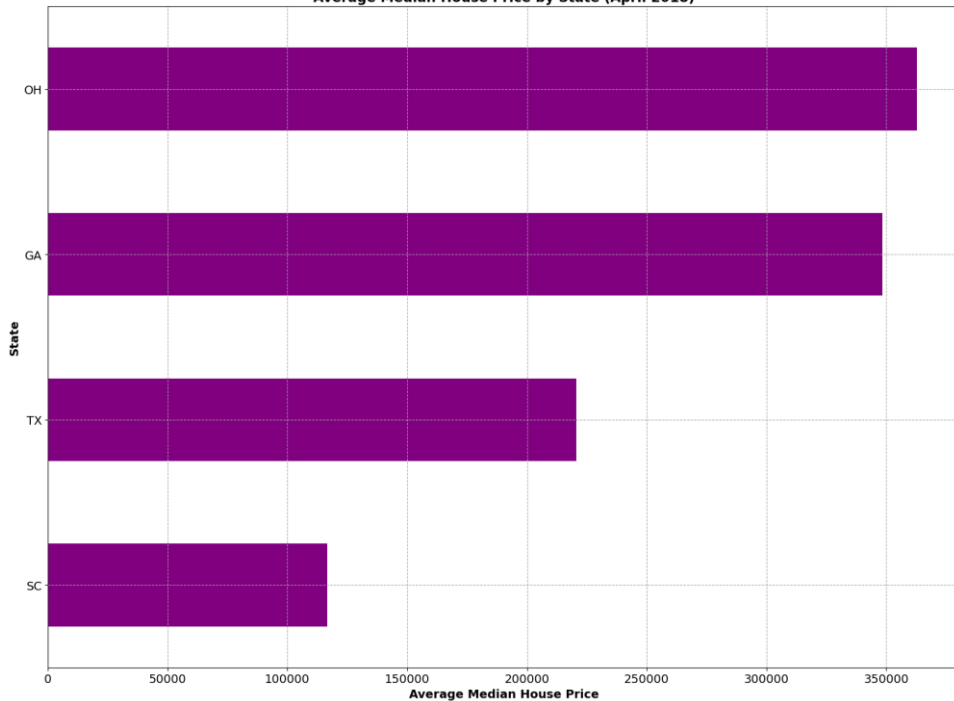
Renaming "RegionName" to "Zipcode"

Change categorical column datatypes to string.

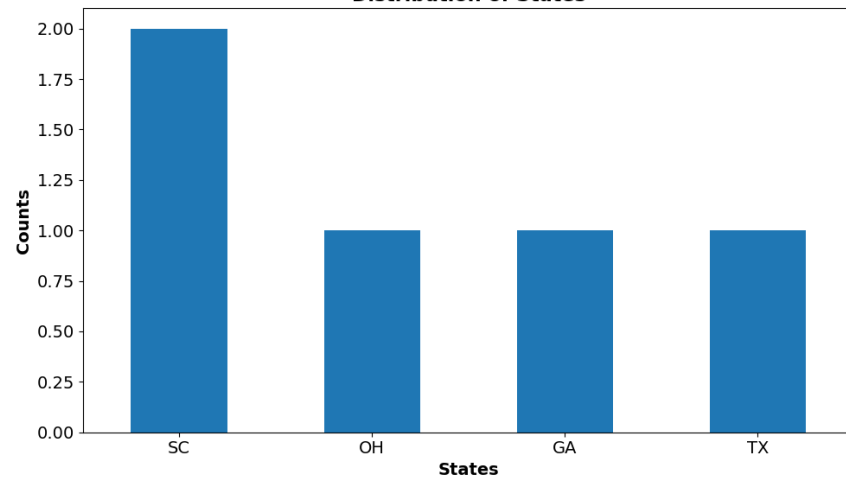
Exploratory Data Analysis



Average Median House Price by State (April 2018)



Distribution of States

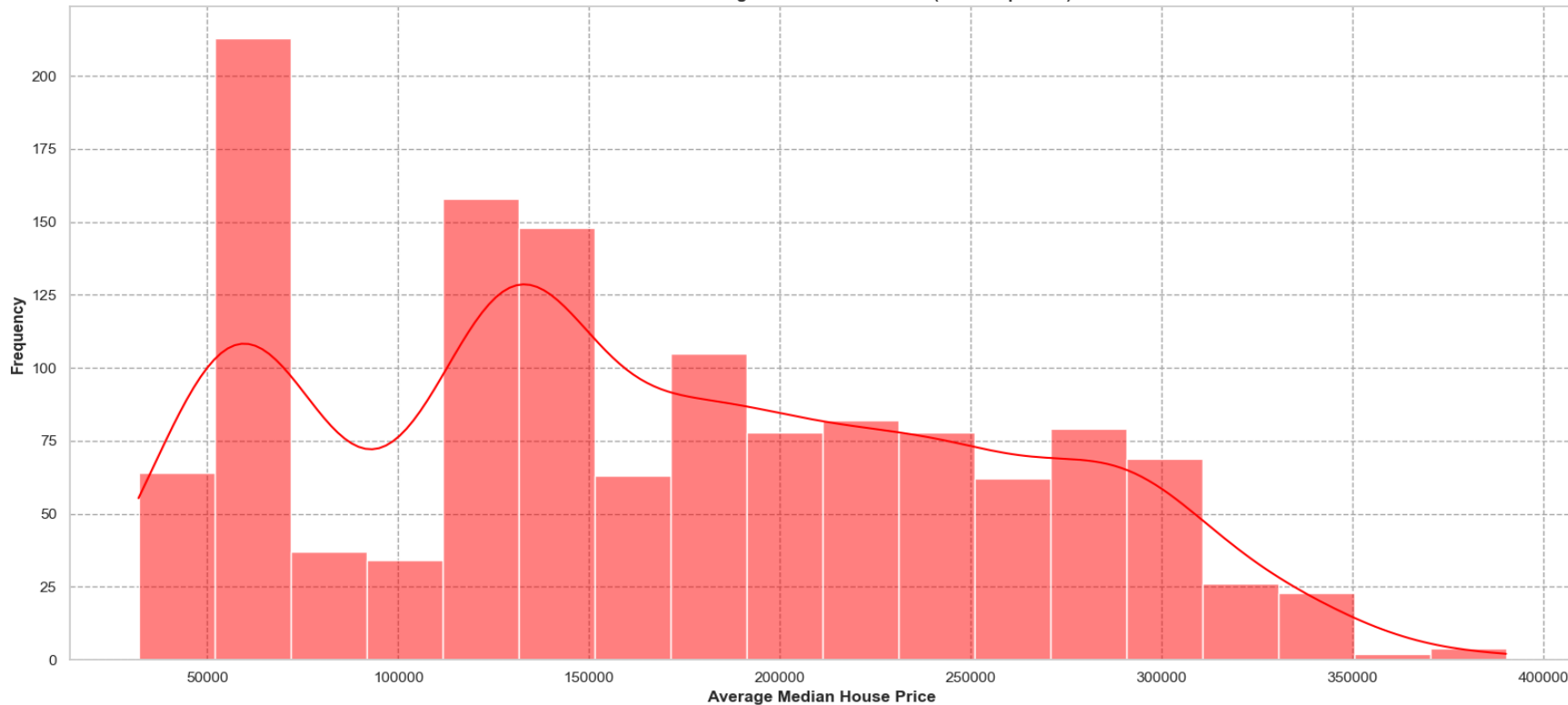


The distribution of states is based on our best 5 zipcodes

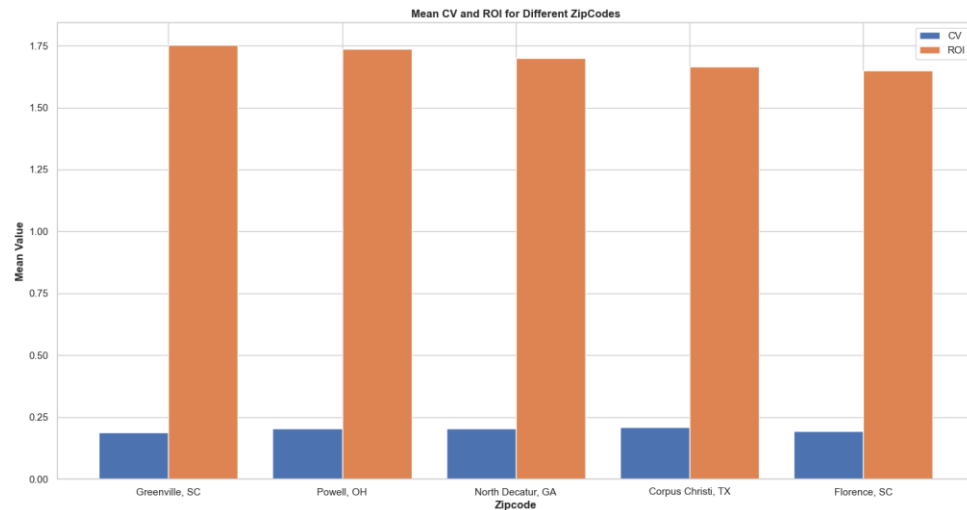
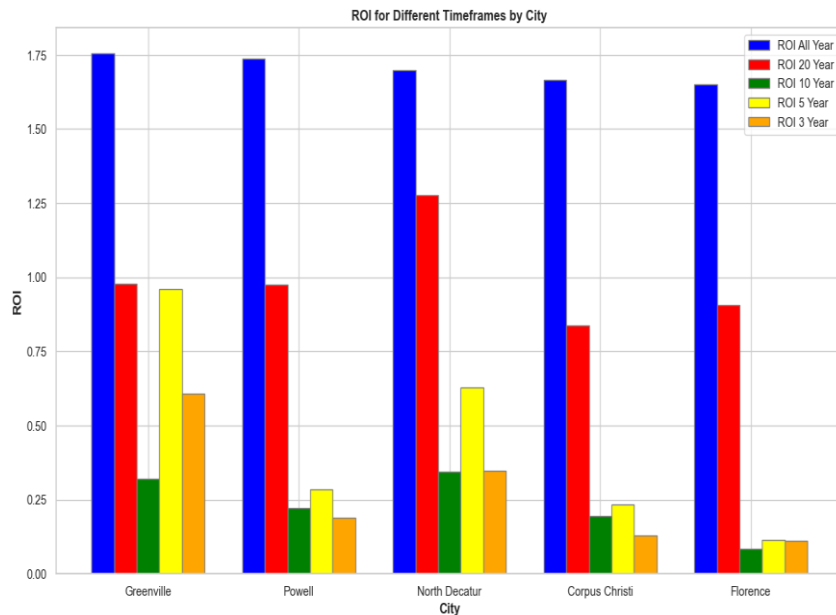
Exploratory Data Analysis



Distribution of Average Median House Prices (Best 5 Zipcodes)

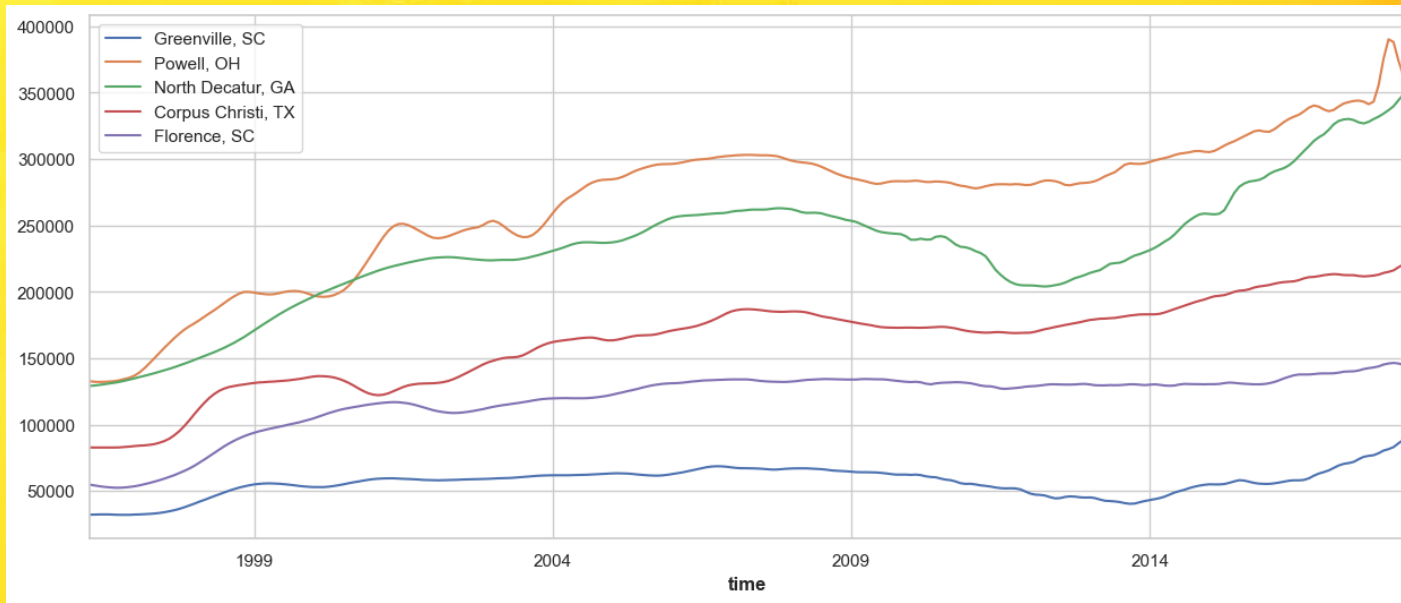


Exploratory Data Analysis



Time Series Modelling

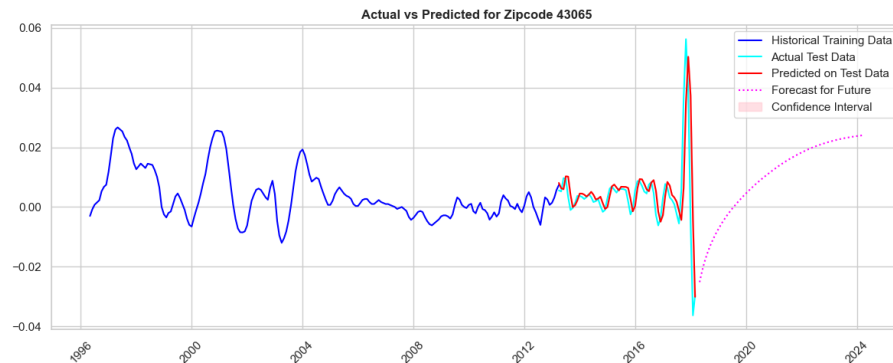
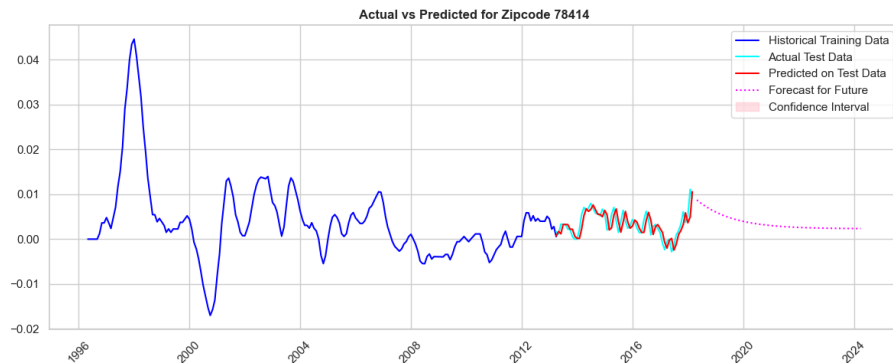
Median Property value per each Zip code over time



Future Forecasting



Stable predictions using best model (LSTM)



Top two promising Zip code:

- 78414:Corpus Christi, TX.
- 43065:Powell ,OH

Other Selected Zip codes:

- 78414:Corpus Christi, TX.
- 29611: Greenville, SC.
- 29501: Florence, SC

Conclusion

The LSTM model demonstrated strong predictive performance, outperforming the other models in terms of lower variation with the model's strengths evident in capturing non-linear trends and abrupt changes.



Recommendations

- Diversifying the investment portfolio across multiple states
- Investments can be focused on cities with higher ROI
 - Greenville, SC
 - North Decatur, GA
 - Powell, OH
- Good investment portfolio should balance ROI with risk
 - Greenville, SC has the lowest risk.
- Property value trends
 - Ohio has higher value property while S.Carolina has relatively lower prices which may be more attractive.
- Long term growth
 - Consider North Decature with highest ROI in 22yr period.



Next Steps

- Location matters in the investment portfolio consider:
 - City unique characteristics
 - Development prospects
 - Economic growth
- Consider other features that may influence ROI in the real estate:
 - Interest rates
 - Economic downtime
 - Policy changes
- Create a system that ensures the model responds to market dynamics in real time.

