TIME SERIES ANALYSIS AND FORECASTING FOR REAL ESTATE INVESTMENT

Group 5

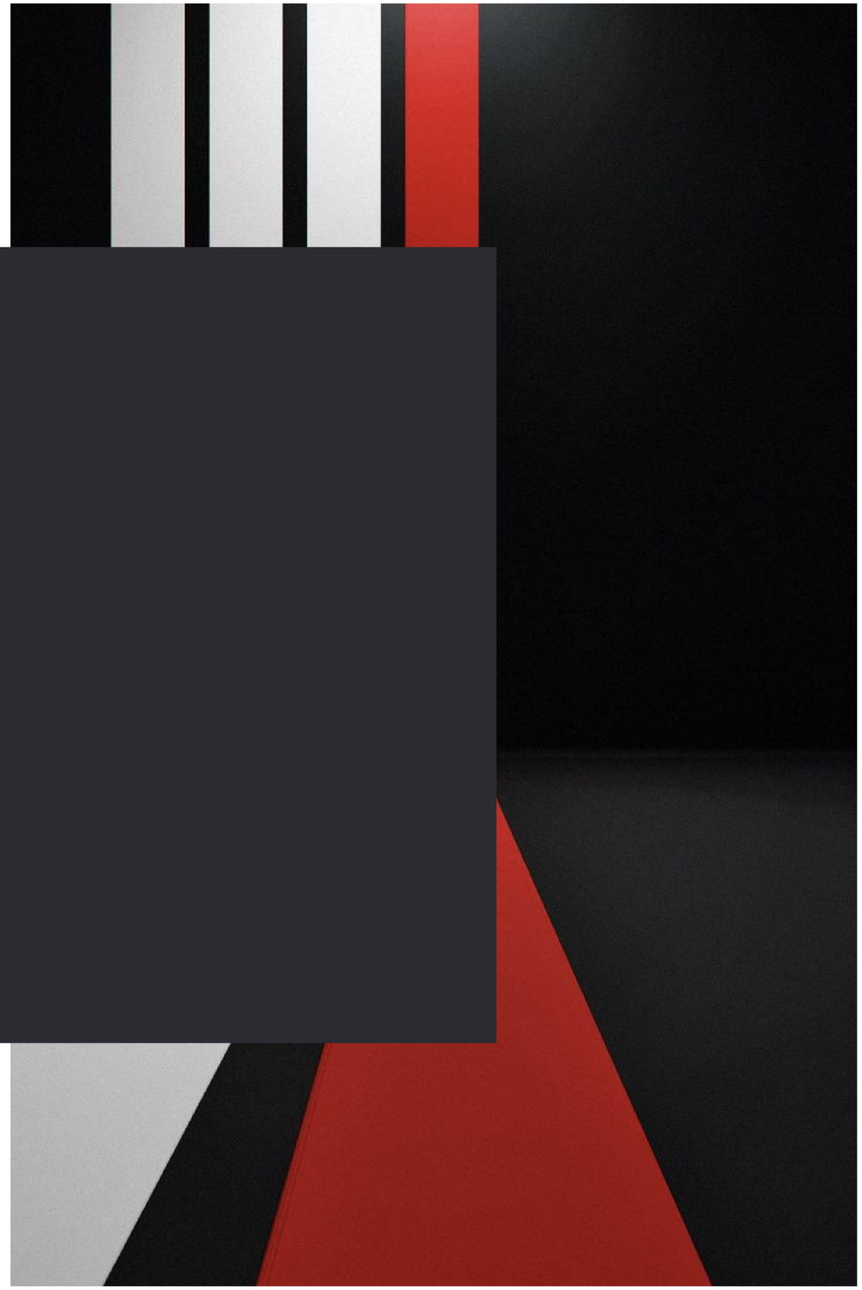
2023 - 2024

Group 5

Table Of Content

Project Overview

The project centers on improving forecasting for real estate prices, steering a make-believe investment firm toward the top 5 zip codes for prime investments. We're using advanced tools like ARIMA and Facebook Prophet to make sense of it all!



Business Understanding

- Predicting house prices aids in proactive risk management for homeowners and investors by leveraging historical patterns.

Particularly from the 2008 financial crisis This approach identifies early warning signs in current data, enabling informed decision-making and effective risk assessment

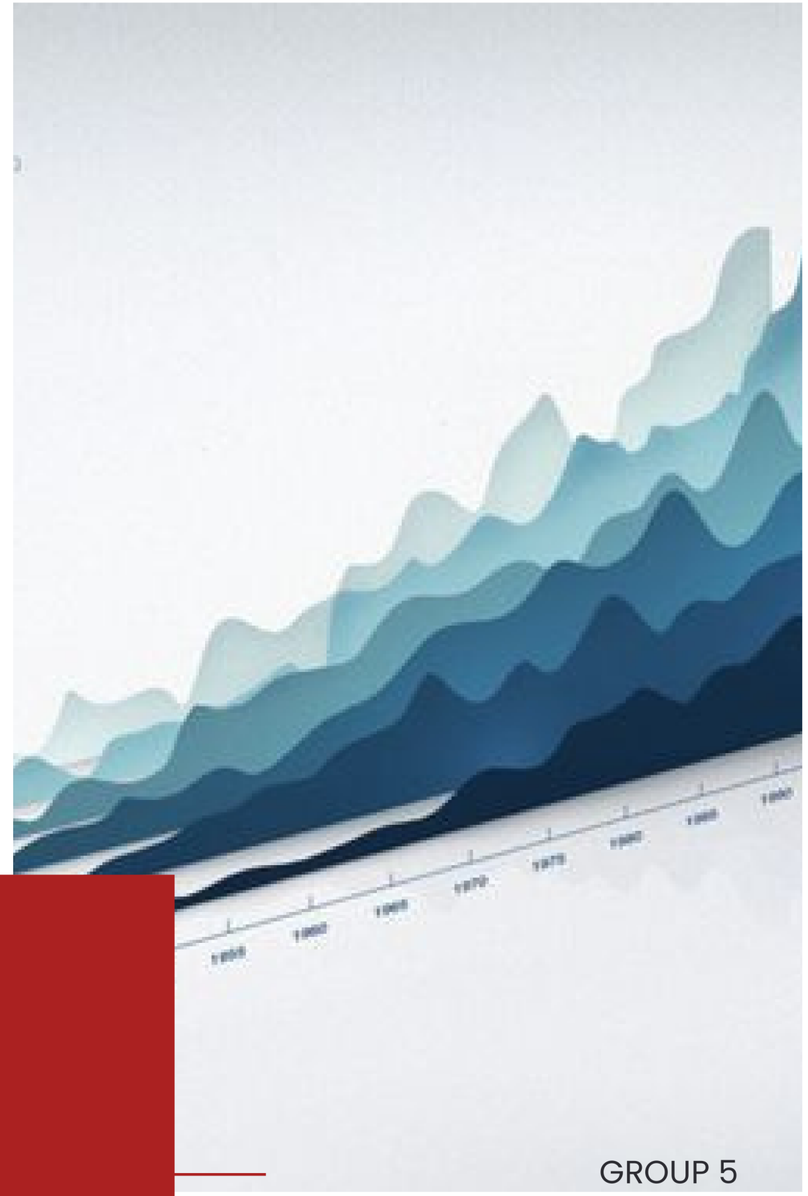
Problem Statement

The task at hand revolves around the essential requirement for accurate and timely predictions of housing prices. This is crucial to address and alleviate risks linked to housing price fluctuations, drawing lessons from historical events such as the 2008 financial crisis

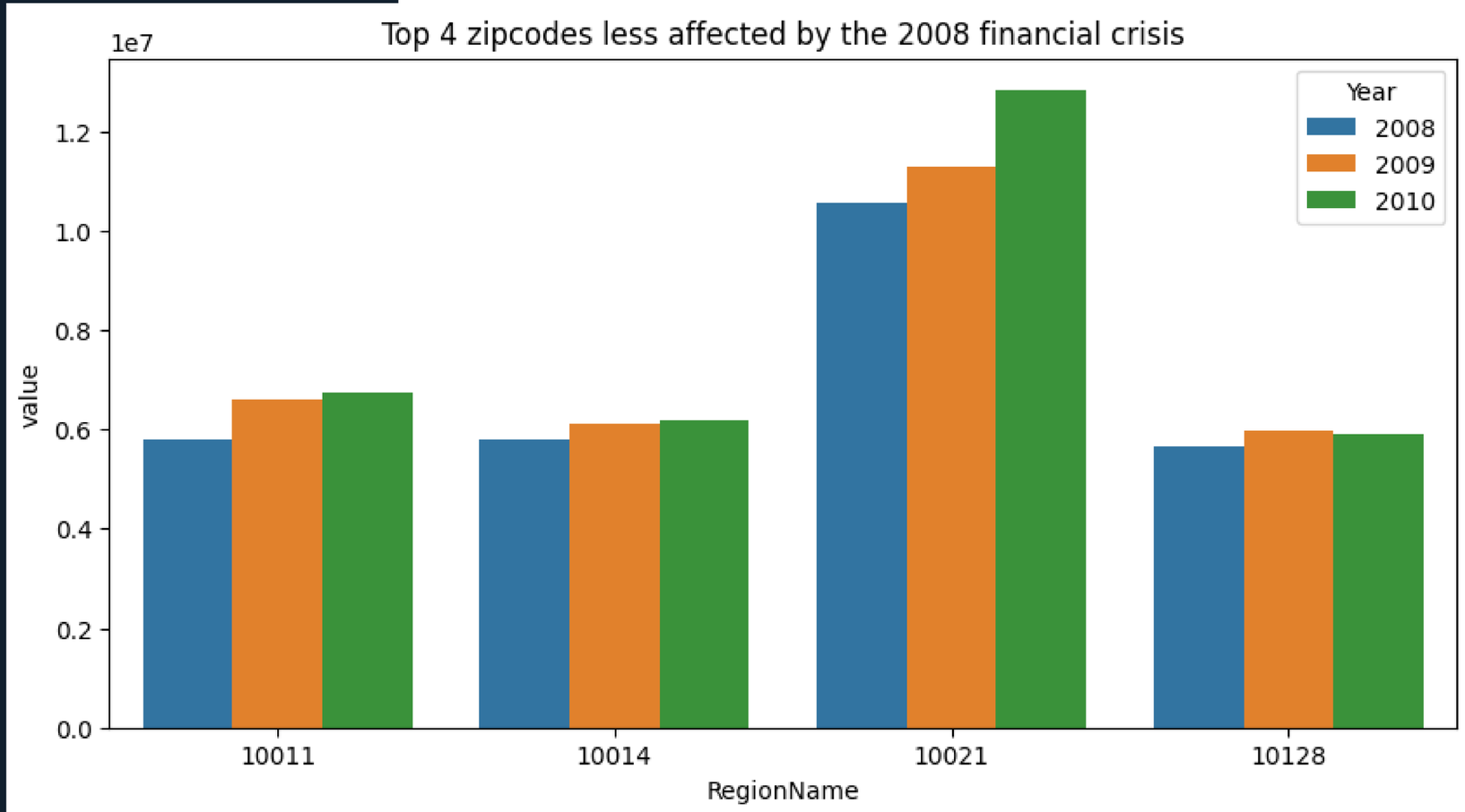
A more accurate prediction system is indispensable to mitigate risks effectively, enabling investors to make informed decisions that maximize returns in the dynamic landscape of real estate.

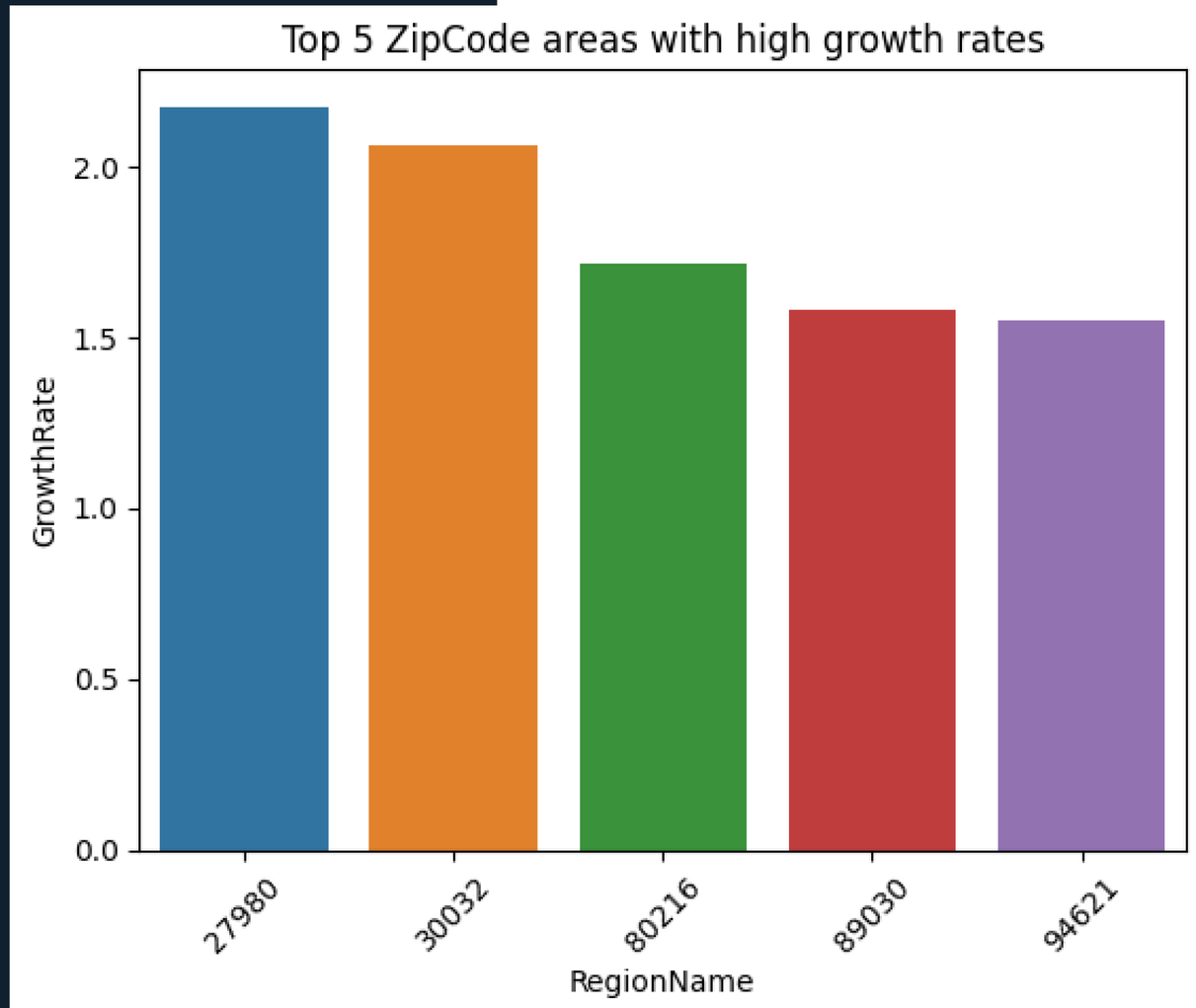
Data Understanding

- The dataset used for this project is sourced from Zillow Research and is provided in the file **zillow_data.csv**.
- The dataset contains a record of time, values, SizeRank, City, Metro, State, RegionID, RegionName, CountyName



Top 4 zipcodes less affected by 2008 financial crisis



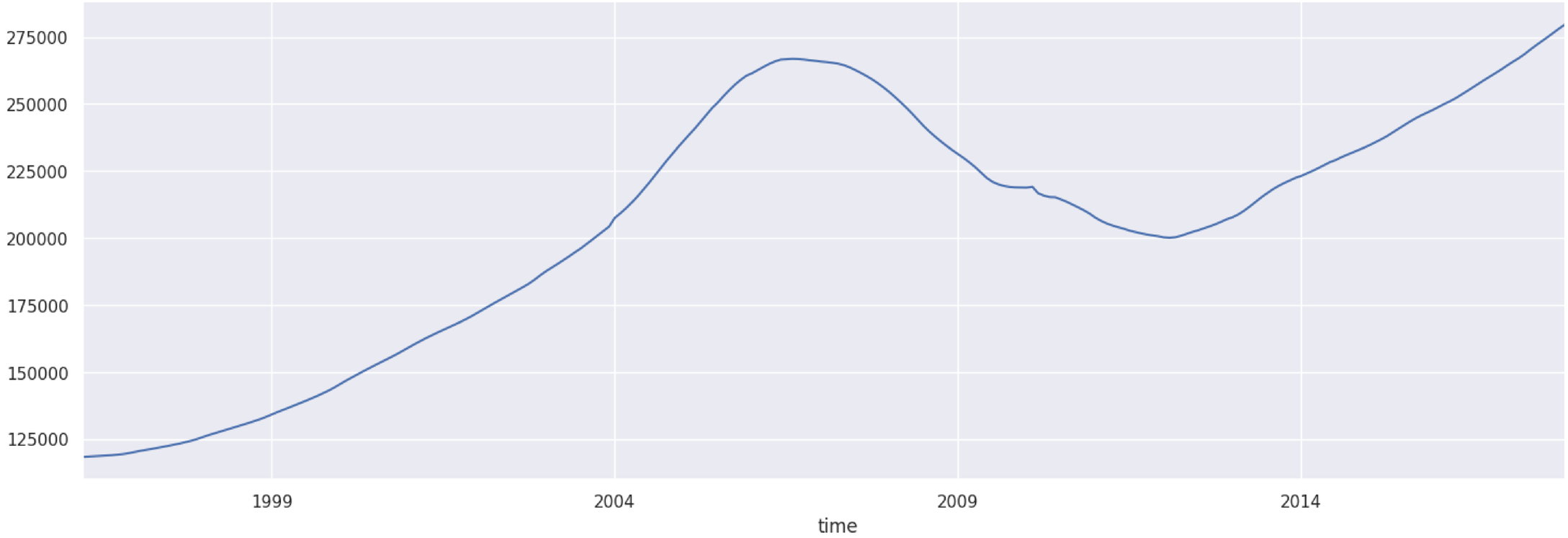


Top 5 zipcodes areas with hight growth rate

GROUP 5

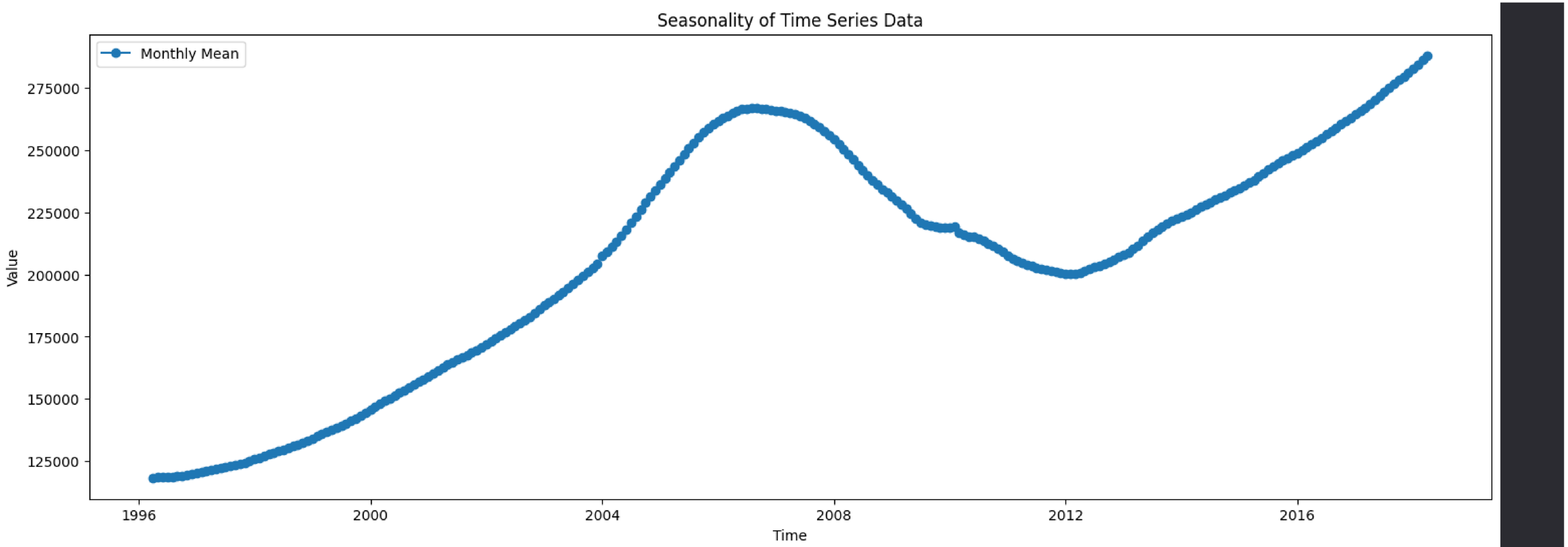
Visualizations :

Time Series Overview: Housing Sales Values Over Time



Seasonality

GROUP 5



Models Used

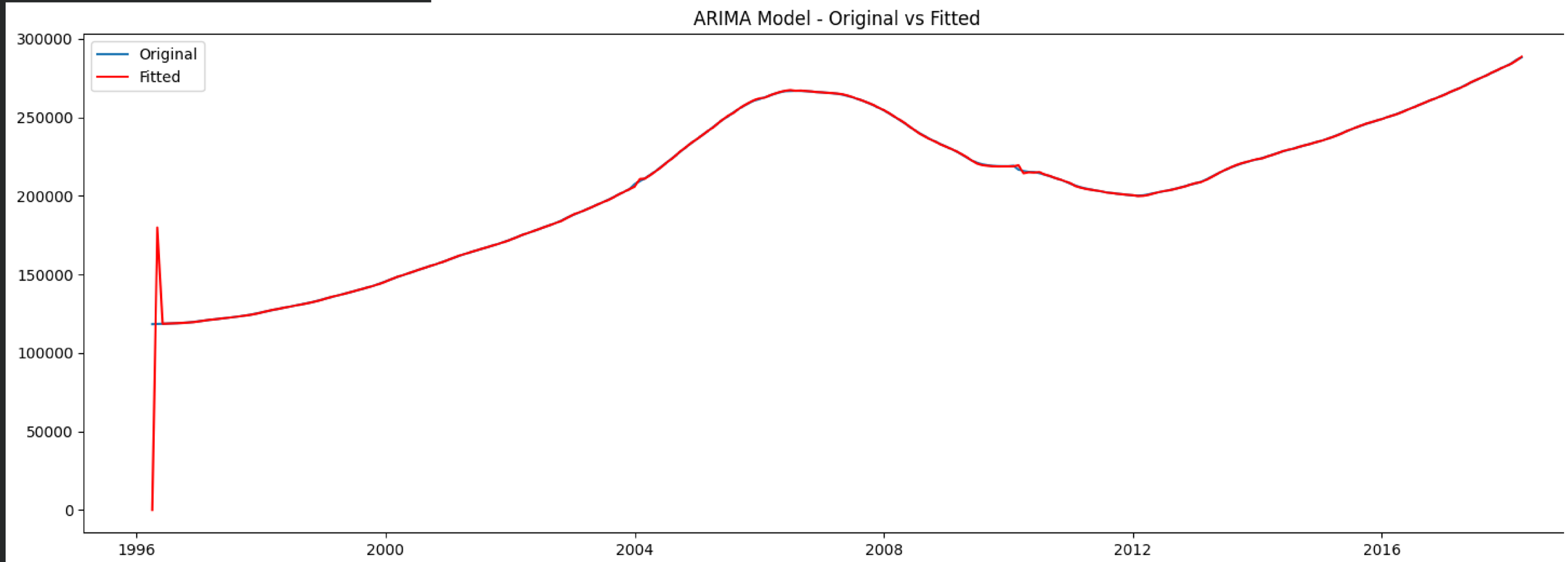
01 ARIMA

02 Facebook Prophet

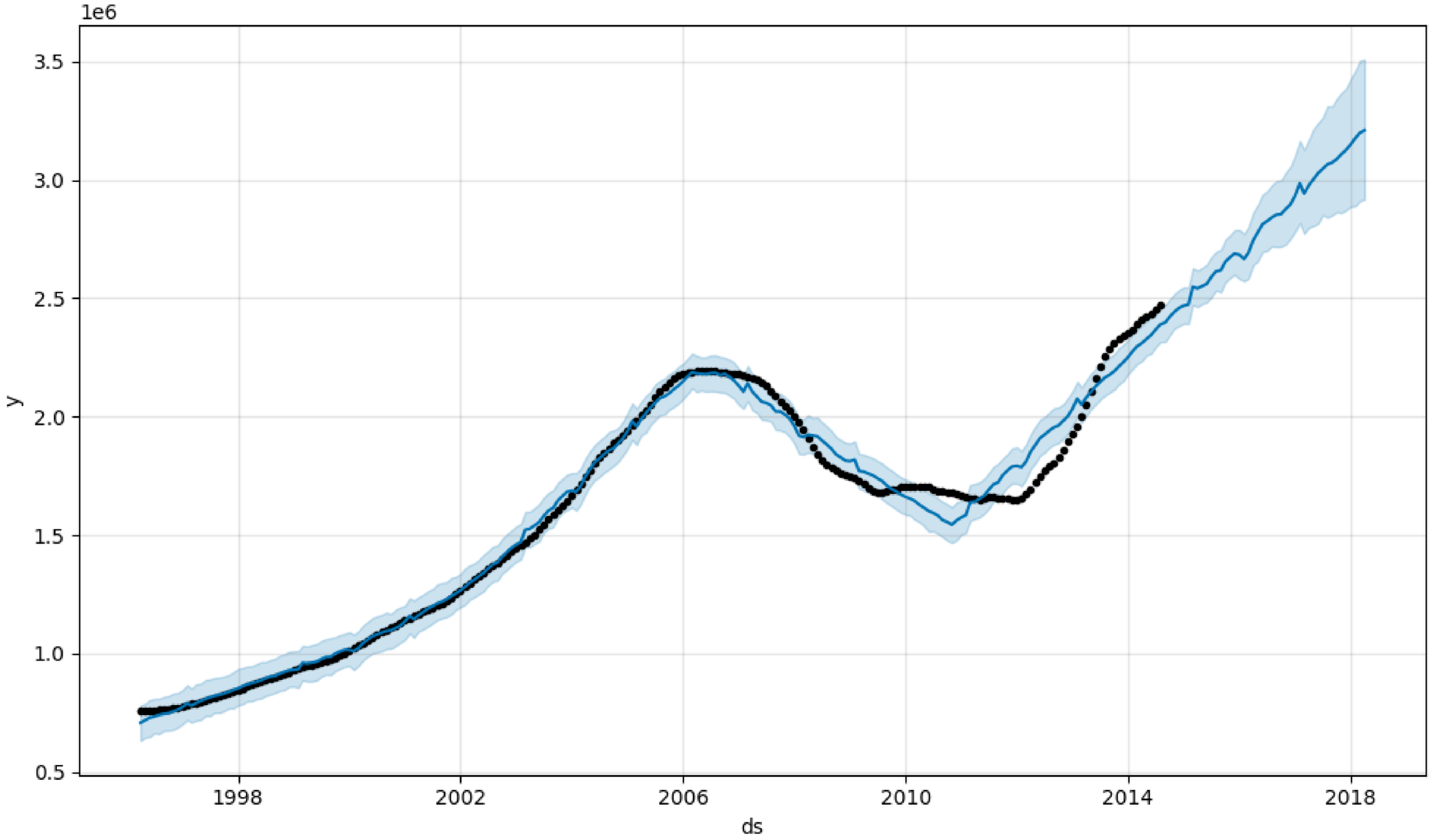
These models serve as our compass in navigating the dynamic landscape of real estate, providing valuable insights to guide our investment decisions. Let's explore how each model contributes to our quest for identifying the top 5 zip codes for optimal investment.

Arima Model

- The plotted graph shows the original time series values (in blue) and the fitted values predicted by the ARIMA model (in red). This visual representation helps evaluate how well the ARIMA model captures the patterns and fluctuations in the observed data.



Facebook Prophet



- The forecasted trend indicates a potential increase in housing prices in the upcoming seasons

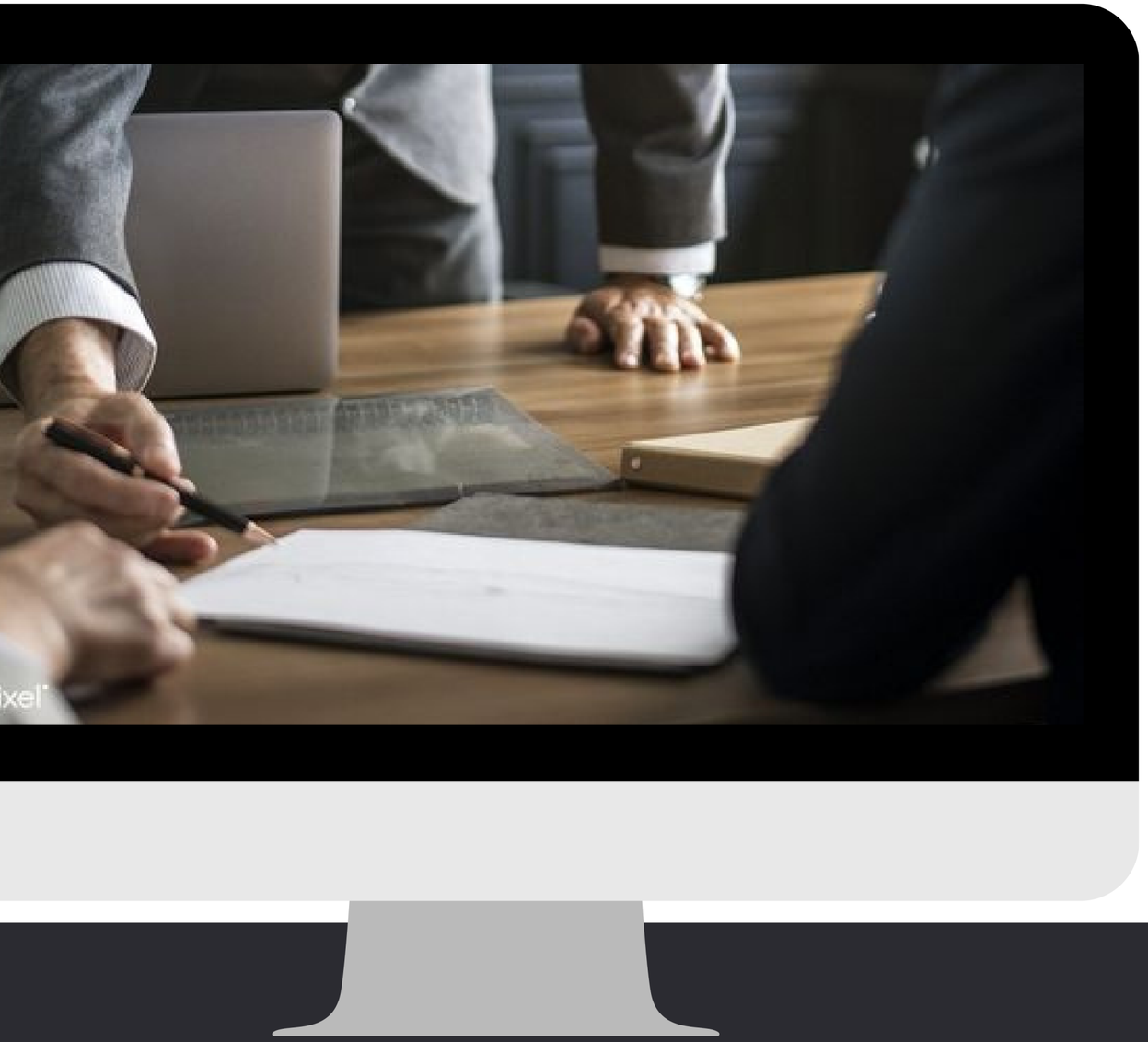
Evaluation

Facebook Prophet

The Facebook Prophet model achieved R2 scores of 74% and 73.66%. It exhibits flexibility with adjustable parameters, such as seasonality, Fourier order, and additional regressors for effective learning. The models for forecasting house values and zip codes have been saved for future use in production

Arima Model

The ARIMA model with the lowest AIC scores of **716** was the one of order **0,2,0**. This model gave us the best possible forecasted values to answer our business objective. To find the forecasted value for each zipcode, the forecasted value was averaged.



Challenges

- Using TimeSeriesSplit without data splitting leads to mismatches between real and forecasted values, hindering pipeline evaluation.

- Extended evaluation time; a solution is leveraging free cloud GPUs for model training, reducing computational load.

Conclusion & Recommendations

The project emphasizes the importance of structuring data in a long format with DateTime as the index for effective time series forecasting. ARIMA, SARIMAX, and Prophet models were employed, with the latter being preferred for their adaptability to seasonal trends. Challenges include tuning for overfitting and lengthy evaluation times.

Implement efficient ETL pipelines for seamless data processing and leverage GPU resources to expedite model evaluation. Monitor and deploy ARIMA and Prophet models for real-time forecasting, ensuring ongoing optimization and adaptation to market dynamics.

Next Steps

- Monitor the model's predictions and scores at production.
- Deploy both models for easier forecasting.

GROUP 5 PRESENTATION

END