Housing Price
Sentiment Advisor

*Richie Garafola * Joseph Garcia * Marc Agnew*

Our group is interested in the housing market as it relates to mortgage rates.

Project Description:

Our tool will analyze the 15 and 30 year fixed mortgage rates from the FRED economic database. We will look for correlations with Case Shiller housing prices. The fixed mortgage rates will act as indicators to help us understand the future outlook of the housing market. The economic sentiment will be gauged using NLP scraped from twitter. This tool will assess the correlations and judge if it is the right time to look for houses based on our predictions.

Questions:

Q1 Is there a correlation with mortgage rates and housing prices?

Q2 What states index is lower than the US National benchmark ?

Q3 What states index is higher than the US National benchmark ?

Q4 Do we foresee a slowdown or a correction in the future?

Q5 What is the overall sentiment of the housing market right now?

Sentiment Analysis

Our sentiment analysis tool scrapes twitter based on keywords, amount of tweets to be analyzed and amount of days back to scrape

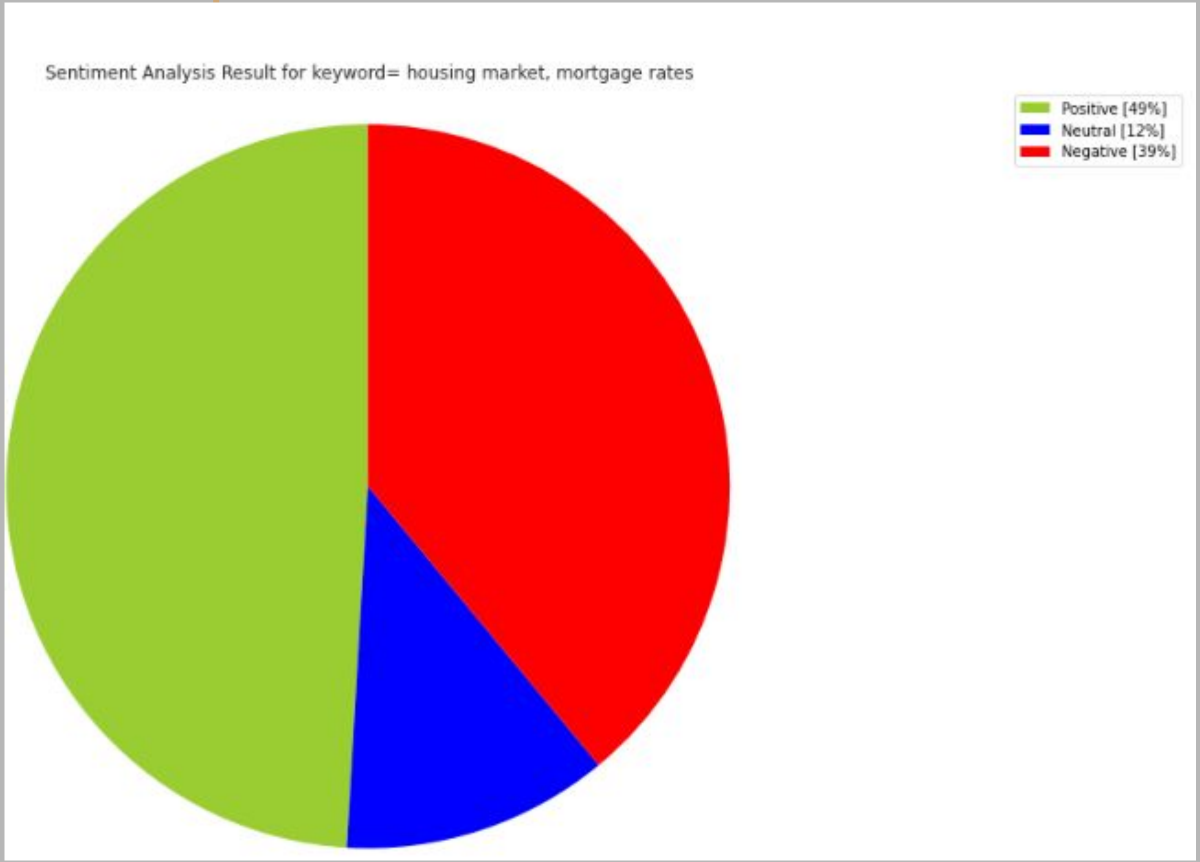
Based on these tweets, we use regex to clean the tweets and sentiment intensity analyzer to analyze the tweets and give us a polarity score for each tweet.

Using the sentiment from the intensity analyser we categorized our negative and positive tweets and ran several machine learning models.

Results vary according to each search. On Average, according to our test accuracy scores, our model predicts around 60%.

Housing Market / Mortgage rate Word Cloud

Sentiment based on Twitter



Logistics Regression model:
Test Set Accuracy : 63.52941176470588 %

Classification Report :

	precision	recall	f1-score	support
-1	0.64	0.98	0.78	55
1	0.00	0.00	0.00	30
accuracy			0.64	85
macro avg	0.32	0.49	0.39	85
weighted avg	0.42	0.64	0.50	85

Twitter Reports:

Training Naive Bayes model ComplementNB:
Test Set Accuracy : 63.52941176470588 %

Classification Report :

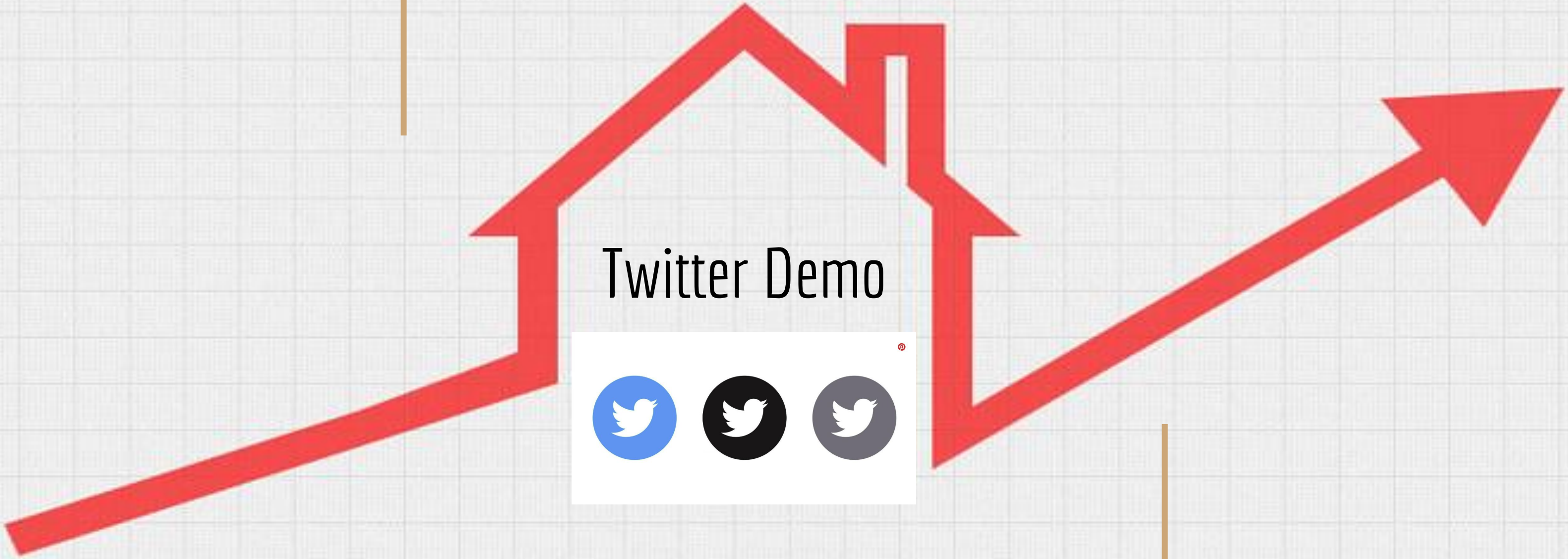
	precision	recall	f1-score	support
-1	0.69	0.78	0.74	55
1	0.48	0.37	0.42	30
accuracy			0.64	85
macro avg	0.59	0.57	0.58	85
weighted avg	0.62	0.64	0.62	85

Training Naive Bayes model MultinomialNB:
Test Set Accuracy : 62.35294117647059 %

Classification Report :

	precision	recall	f1-score	support
-1	0.64	0.96	0.77	55
1	0.00	0.00	0.00	30
accuracy			0.62	85
macro avg	0.32	0.48	0.38	85
weighted avg	0.41	0.62	0.50	85

Twitter Demo



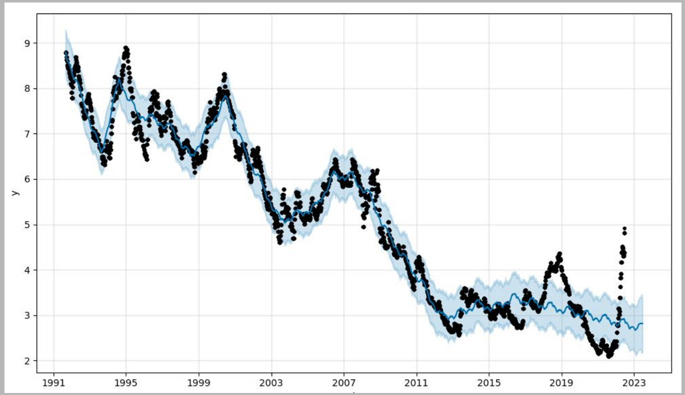
Project :

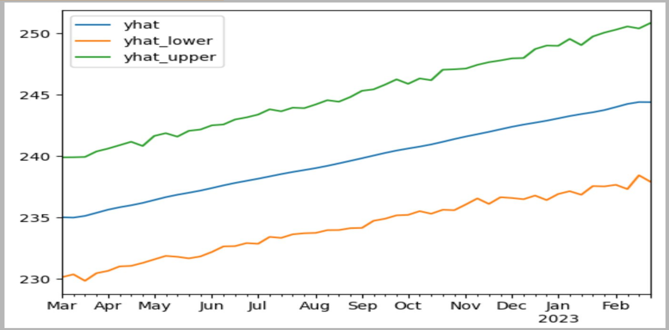
The Case-Shiller Home Price Index is delivered on a monthly basis. Although our data originates in 1987, 12 data points a year is not a lot. We decided to split our study into 2 as a work around.

1. We will modify our data by using a forward fill method which assigns the data that was delivered at the beginning of the month to fill all days in the month until the next delivery the following month. This will give us approximately 30x more data than the original dataset.
2. We want to be as unbiased as possible so we will also use the original dataset with less data points, but true to delivery. This will give us the real picture.

How will our tool forecast housing prices?

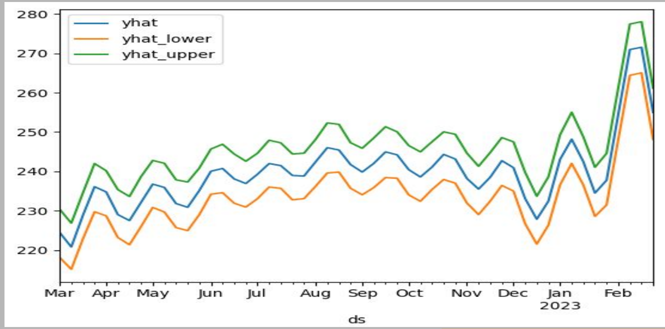
We will run a time series study using Prophet to analyze the 15 and 30 year fixed mortgage rate trend. We will also run a dynamic study on the state of the user's choice as well as the U.S. National Home Price Index.

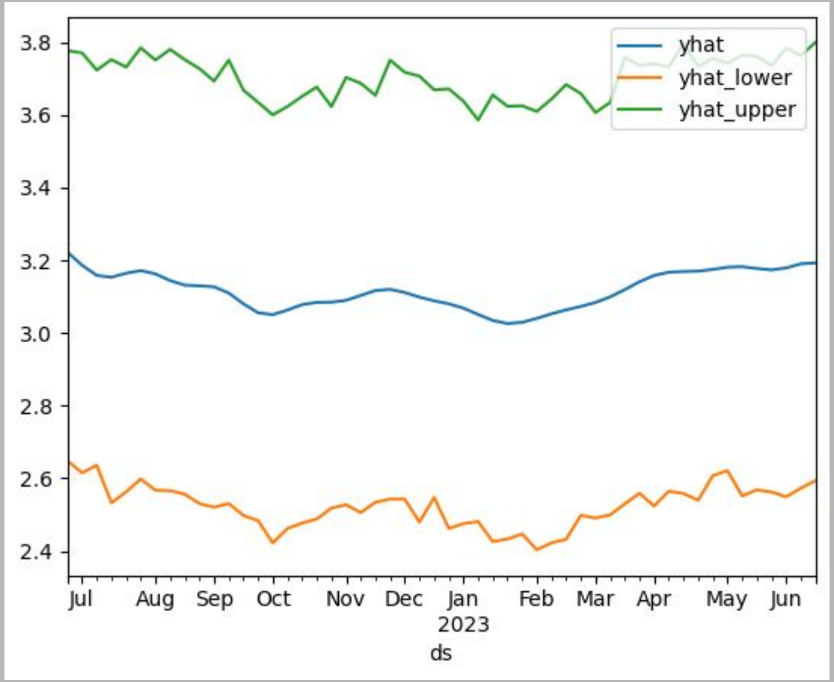




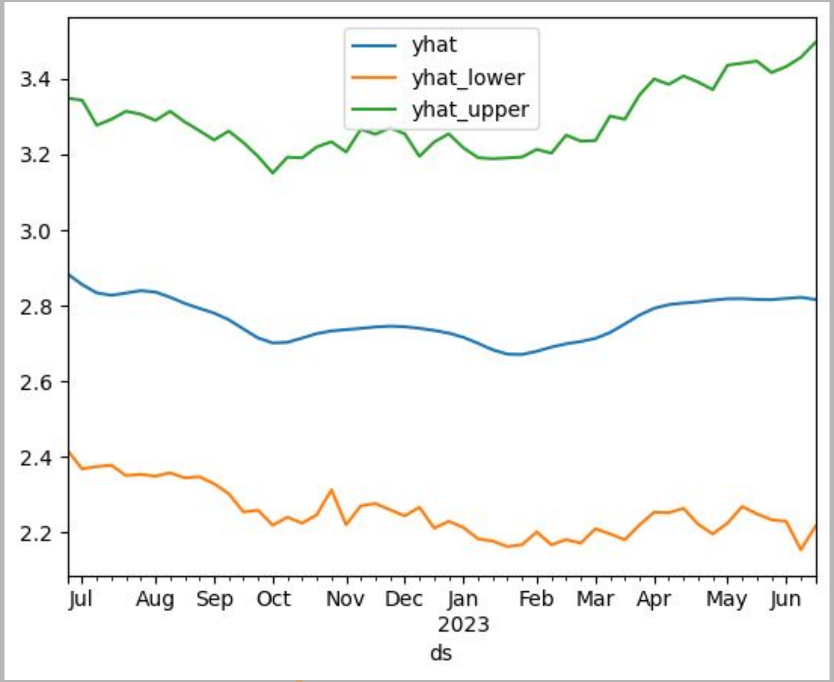
BiASED

UNBIASEd

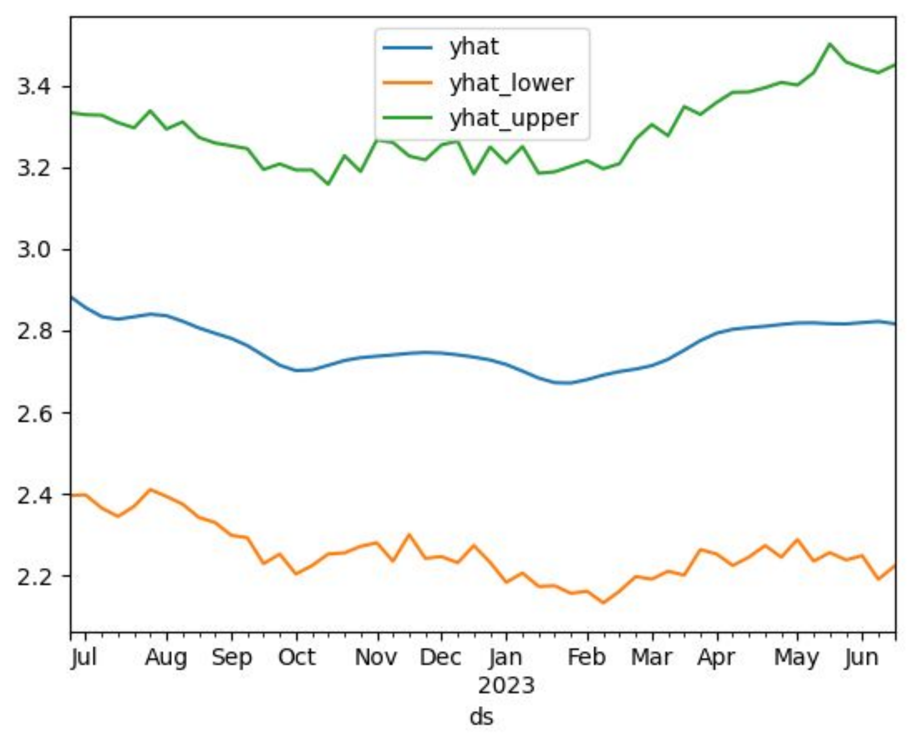


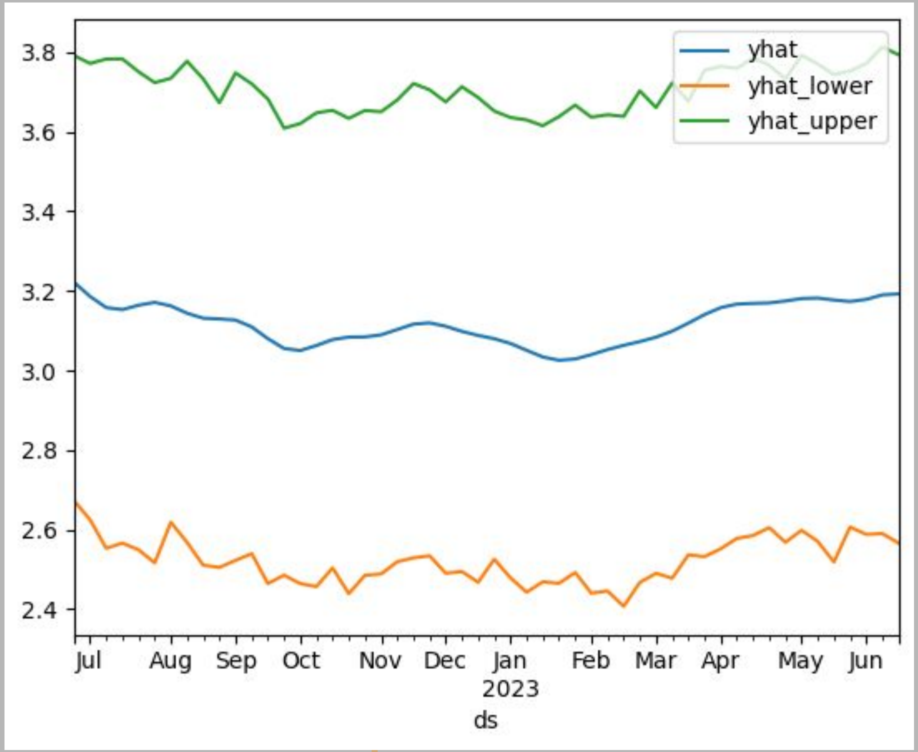


Biased 15 year Mortgage forecast



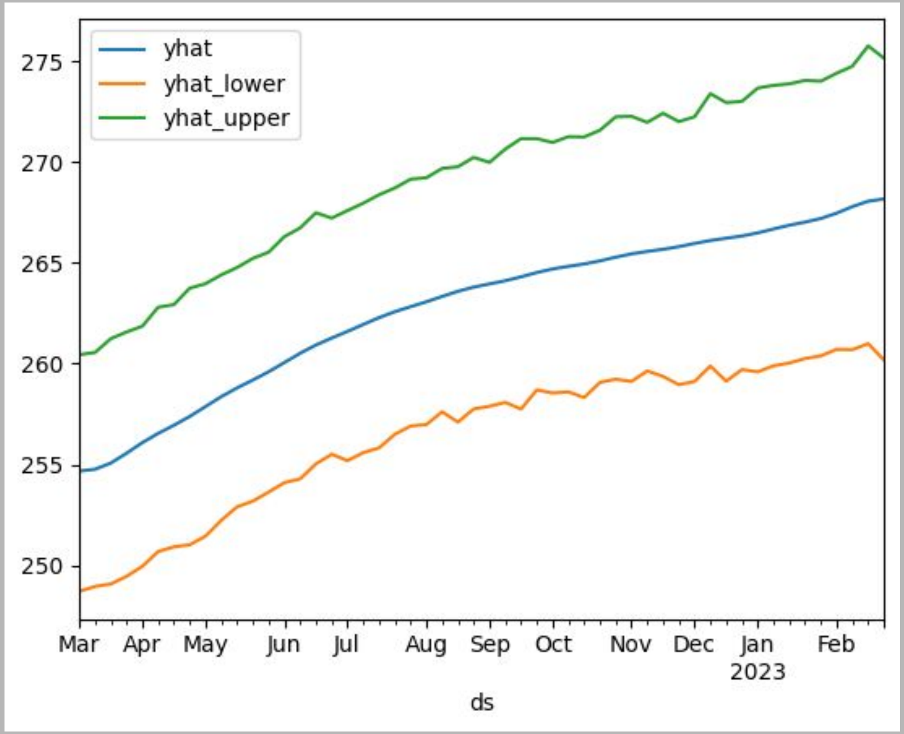
Non Biased 15 year Mortgage forecast



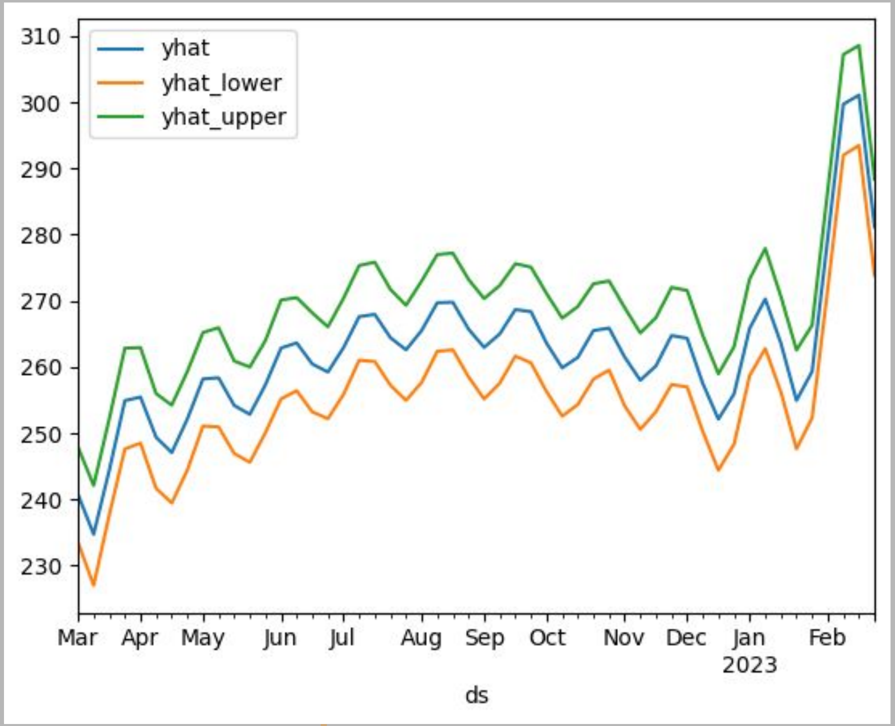


Biased 30 year Mortgage forecast

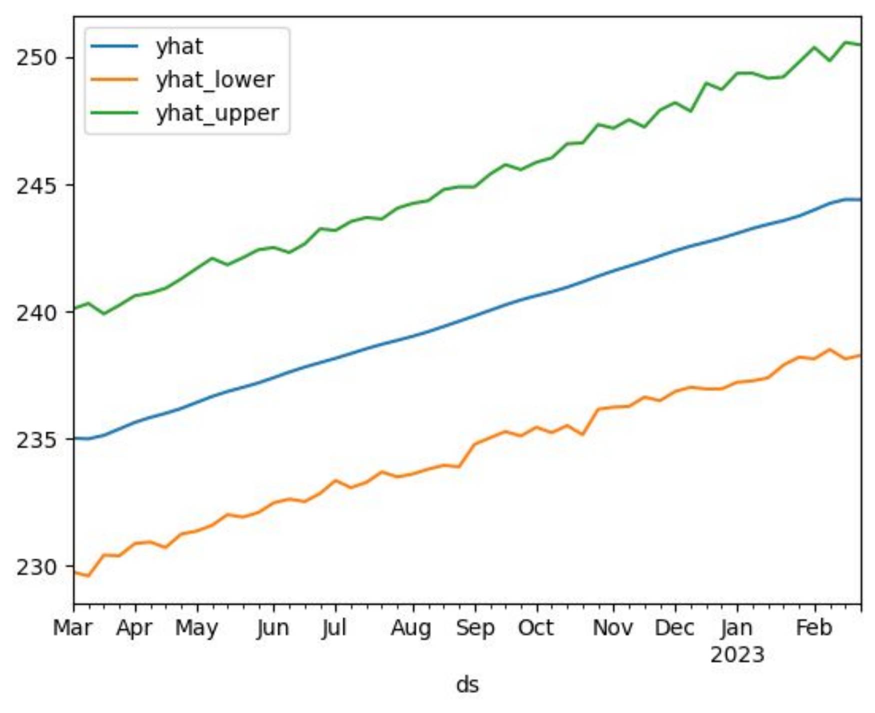
Non Biased 30 year Mortgage forecast



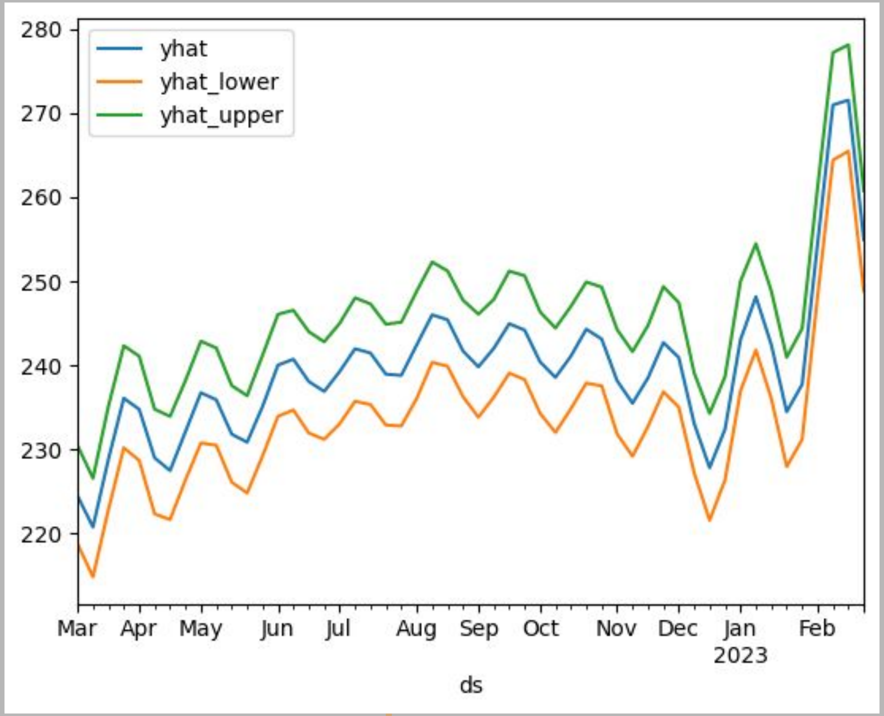
Biased US National Index forecast



Non Biased US National Index forecast



Biased NY State index forecast



Non Biased NY State Index forecast

Data Clean Up Process

- We found clean sources of data using an API

- Imported our data, concatenated, split, column manipulation

- Cleaned data, dropped NAs, counted and removed the nulls.

Data Analysis

NLP: WordCloud, Vader Sentiment Intensity Analyzer, PorterStemmer/ Lemmatizer

Machine Learning on NLP: **Naive Bayes** and **Logistic Regression**

Prophet:

Machine Learning on Index and Mortgage rates: SVM, ADABoostClassifier, Decision Tree Classifier, Passive Aggressive Classifier

Sklearn

Models used:

- ★ Linear - Logistic Regression, Passive Aggressive Classifier
- ★ Naive Bayes - Multinomial NB, Complement NB
- ★ SVM / SVC
- ★ Ensemble - AdaBoostClassifier
- ★ Decision Tree

Pre-processing/Metrics:

- ★ Train test split
- ★ Label Encoder
- ★ Standard Scaler
- ★ Classification Report
- ★ Accuracy Score
- ★ TfidfVectorizer

Machine Learning with Scikit-Learn

nltk

- ★ Work Tokenize
- ★ Sent Tokenize
- ★ Stopwords
- ★ WordNetLemmatizer
- ★ PorterStemmer
- ★ Vader - SentimentIntensityAnalyzer
- ★ Ngrams

NLTK

Downloads:

- ❖ nltk.download('vader_lexicon')
- ❖ nltk.download('stopwords')
- ❖ nltk.download('wordnet')
- ❖ nltk.download('omw-1.4')
- ❖ nltk.download('stopwords')
- ❖ nltk.download('vader_lexicon')
- ❖ nltk.download('punkt')

Tensorflow / Keras

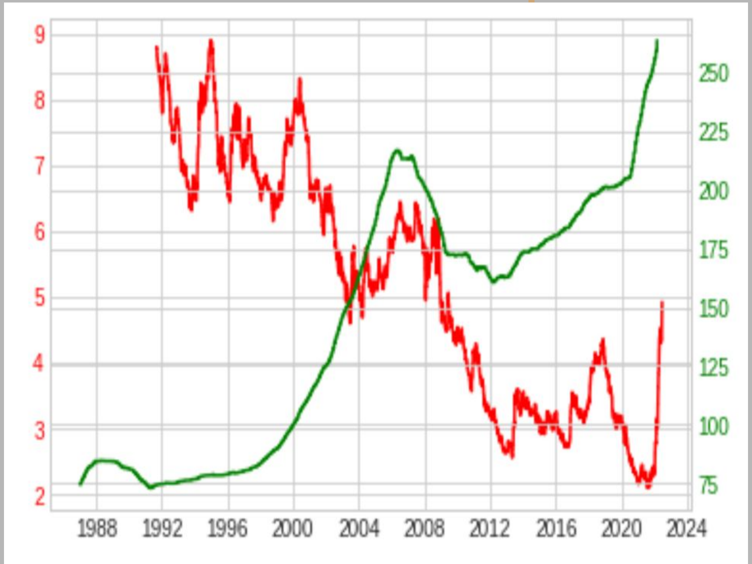
Model - Sequential

Layers - Dense

Preprocessing

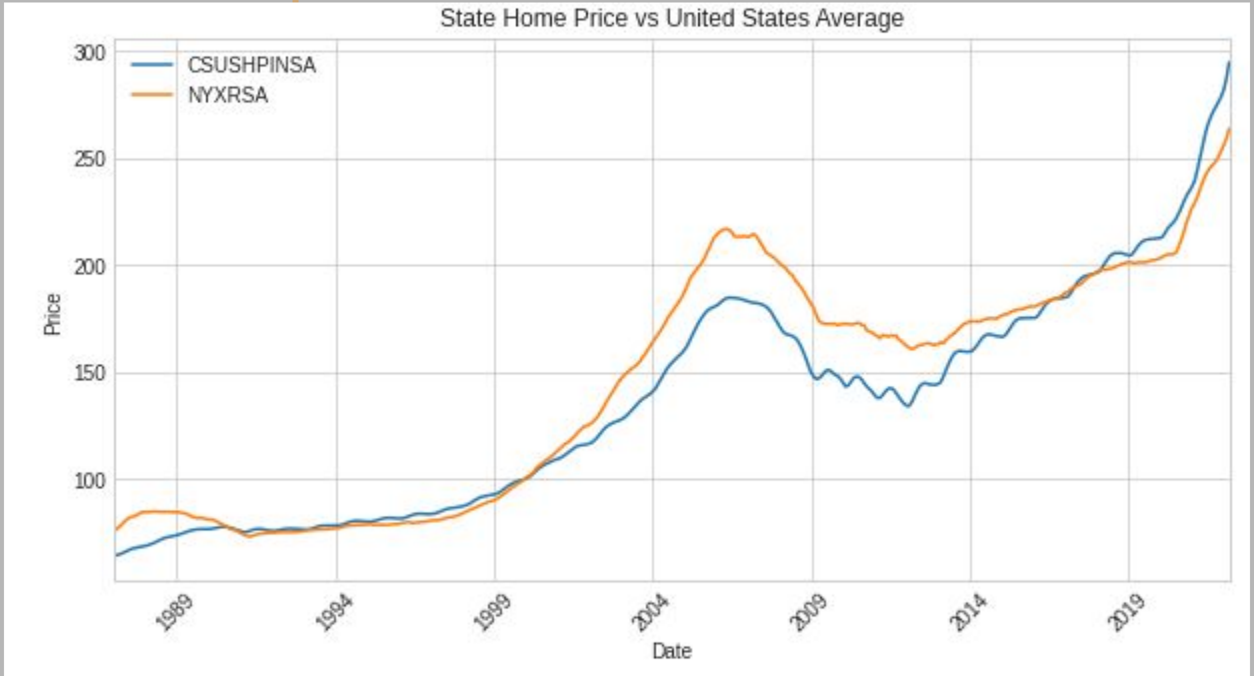
Tokenizer
Pad Sequences

15 Year Mortgage vs Washington

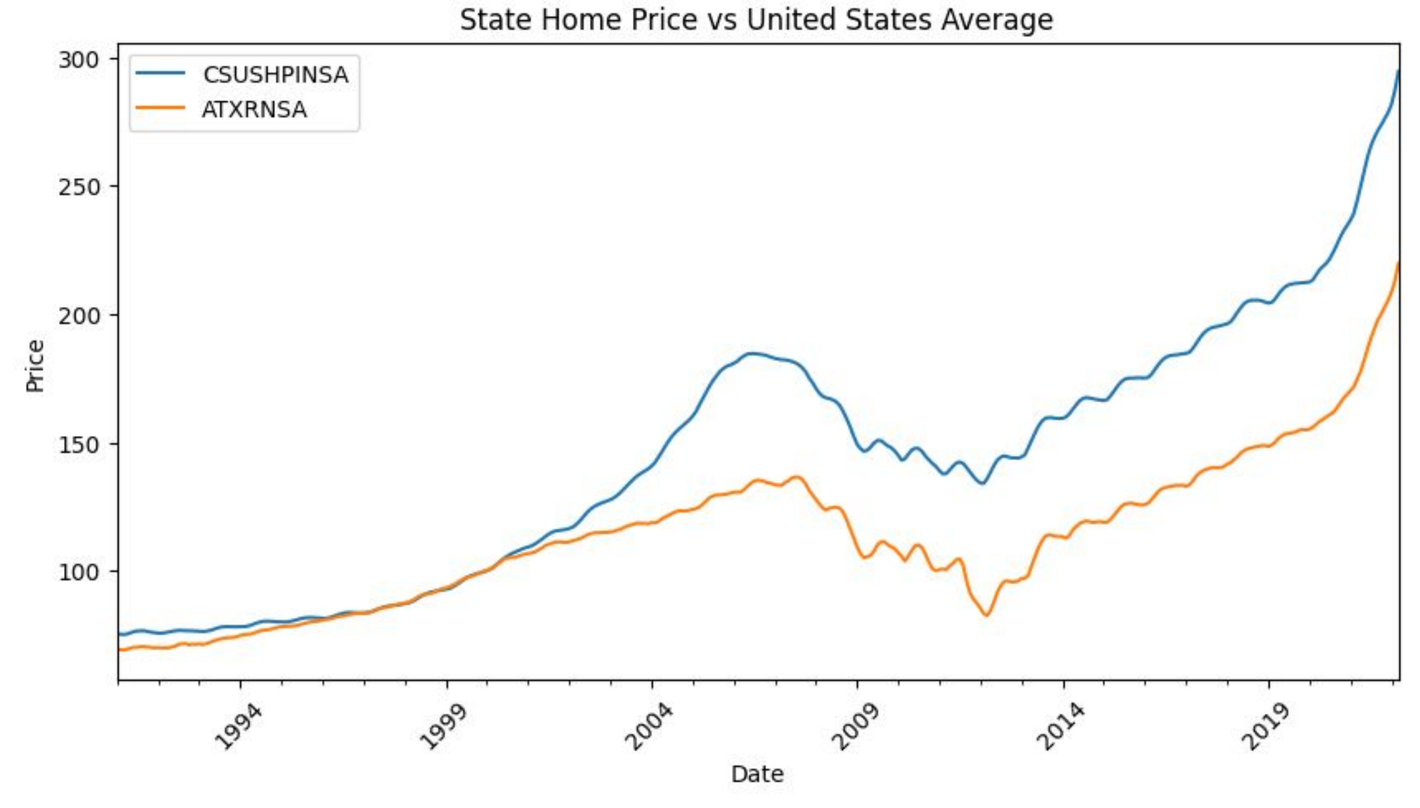


30 Year Mortgage Vs Washington

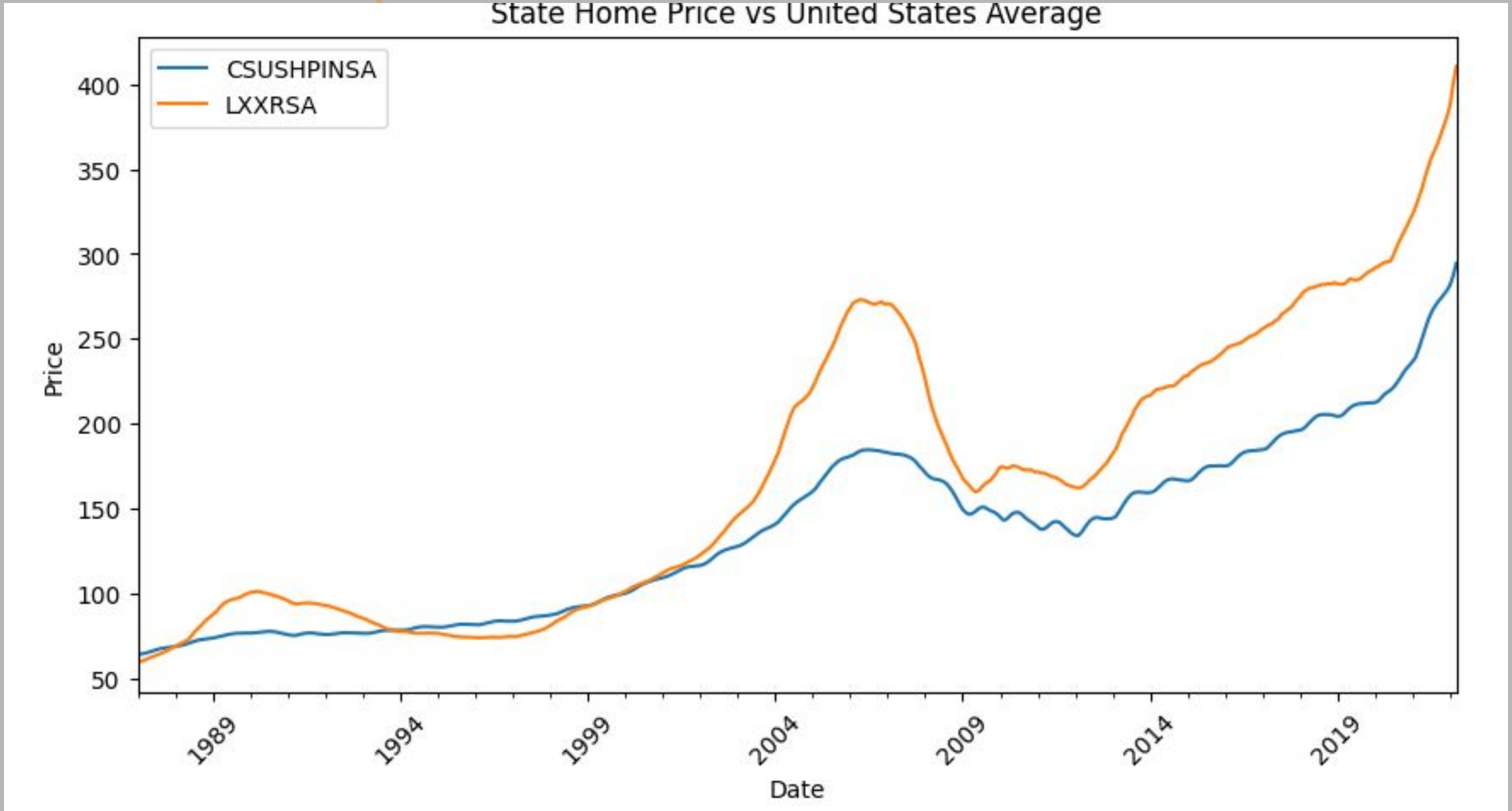
NY home price index vs United States Average



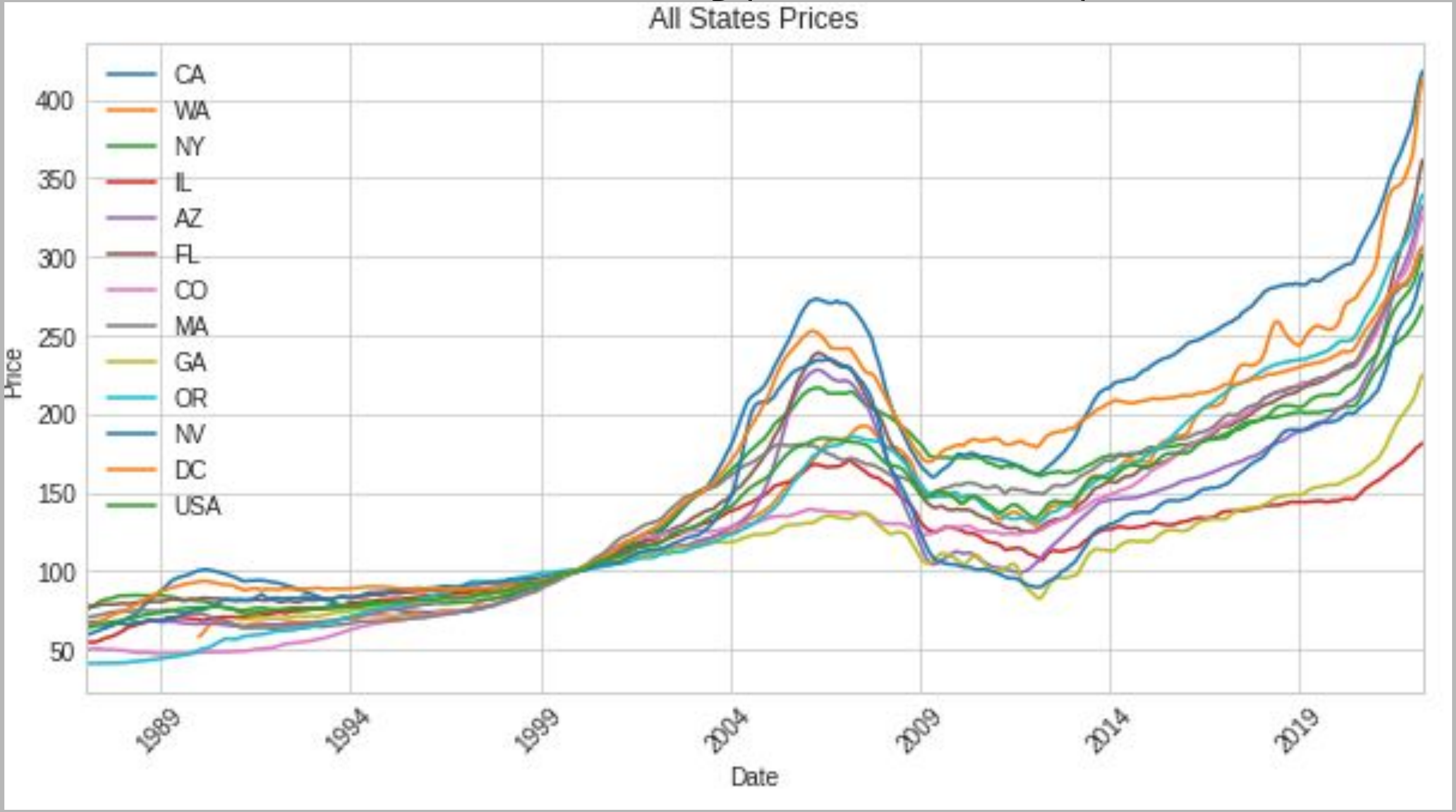
Georgia home price index vs United States Average



Los Angeles, Ca home price index vs United States Average



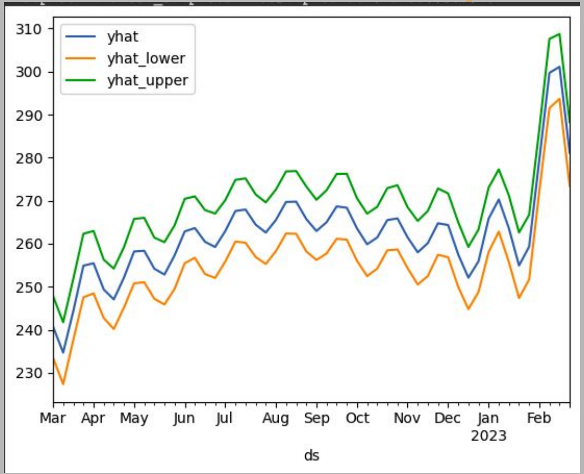
All available states housing price index - as per FRED



States Index in comparison to the us benchmark (Case Shiller)

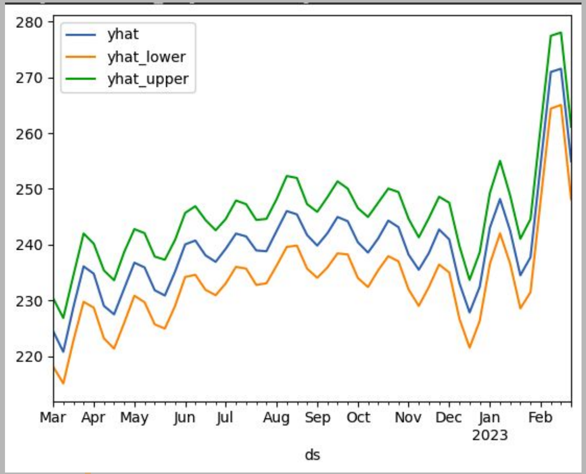
	More Expensive than the U.S. National Home Price Index
CA	True
WA	True
NY	True
IL	False
AZ	True
FL	True
CO	True
MA	True
GA	False
OR	True
NV	True
DC	True
USA	False

United States
Prophet Predictions



Prophet Predictions

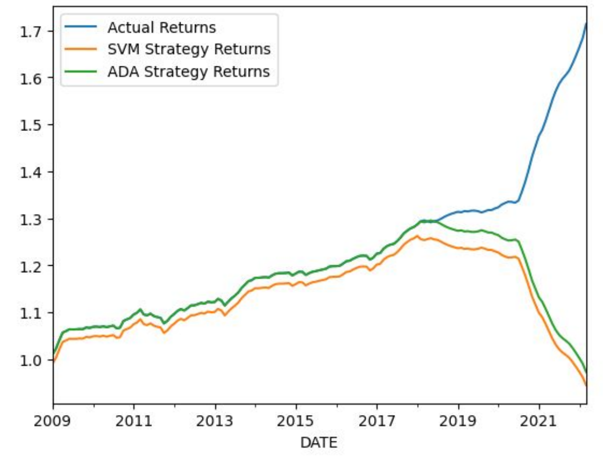
The state of Washington
Prophet Predictions



Machine Learning Outcomes

Answer 5 Questions

Machine Learning Outcomes



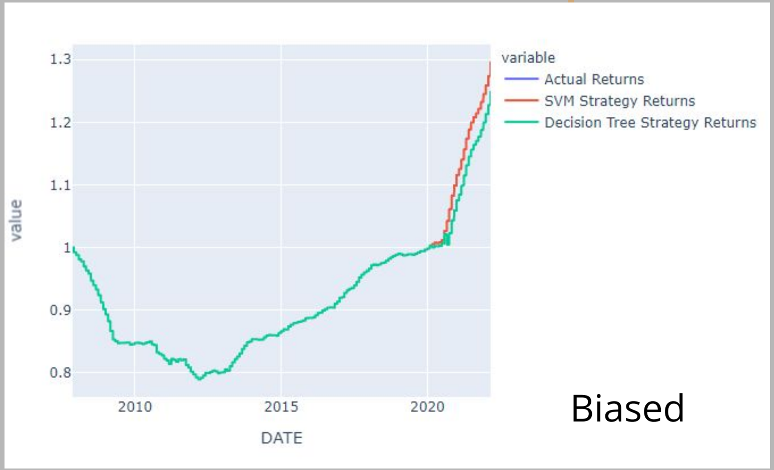
	precision	recall	f1-score	support
-1.0	0.00	0.00	0.00	56
1.0	0.99	1.00	0.99	5165
accuracy			0.99	5221
macro avg	0.49	0.50	0.50	5221
weighted avg	0.98	0.99	0.98	5221

Biased AdaBoost
Classification Report

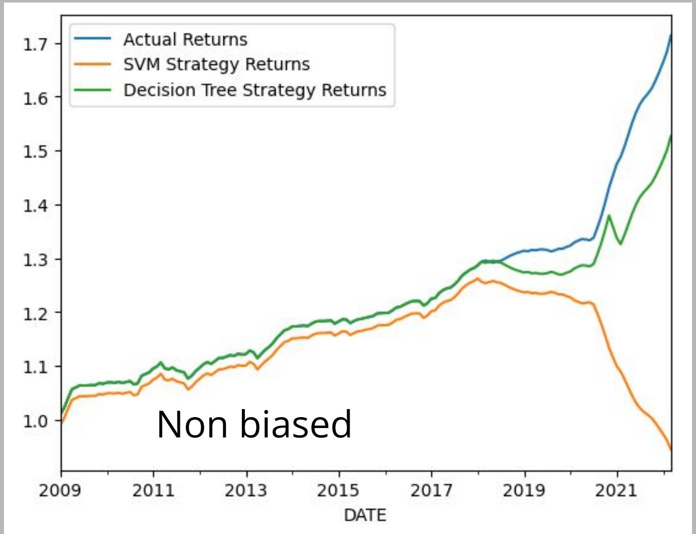
	precision	recall	f1-score	support
-1.0	0.10	0.12	0.11	43
1.0	0.65	0.62	0.64	116
accuracy			0.48	159
macro avg	0.38	0.37	0.37	159
weighted avg	0.51	0.48	0.49	159

Non Biased AdaBoost
Classification Report

Machine Learning Outcomes



	precision	recall	f1-score	support
-1.0	0.00	0.00	0.00	56
1.0	0.99	1.00	0.99	5165
accuracy			0.99	5221
macro avg	0.49	0.50	0.50	5221
weighted avg	0.98	0.99	0.98	5221



	precision	recall	f1-score	support
-1.0	0.17	0.09	0.12	43
1.0	0.71	0.83	0.76	116
accuracy			0.63	159
macro avg	0.44	0.46	0.44	159
weighted avg	0.56	0.63	0.59	159

Machine Learning Outcomes

Training Data Score: 0.9903356541619037
Testing Data Score: 0.9922069825436409

Training Data Score: 0.6938110749185668
Testing Data Score: 0.7766990291262136

Biased Passive Aggressive Classification Report

Test Set Accuracy : 99.09600997506234 %

Classification Report :

	precision	recall	f1-score	support
-1.0	0.00	0.00	0.00	29
1.0	0.99	1.00	1.00	3179
accuracy			0.99	3208
macro avg	0.50	0.50	0.50	3208
weighted avg	0.98	0.99	0.99	3208

Non Biased Passive Aggressive - Classification Report

Test Set Accuracy : 77.66990291262135 %

Classification Report :

	precision	recall	f1-score	support
-1.0	0.00	0.00	0.00	23
1.0	0.78	1.00	0.87	80
accuracy			0.78	103
macro avg	0.39	0.50	0.44	103
weighted avg	0.60	0.78	0.68	103

Postmortem

Although not for lack of effort, we were unsuccessful in building our dashboard. We had difficulties rendering plots into the panel dashboard. We understand we can not utilize certain libraries in panel but had a lot of trouble coming up with a solution as some charts required to be plotted with the unsupported libraries.

We also had trouble building our lambda functions to be executed in Lex.

Questions ?


Thank You