Top 5 Zip Codes to Invest Real Estate

Presented by Data Pros.

Agenda

- Project Overview
- Datasets Utilized
- Methodology
- How We Chose the Zip Codes
- Time Series Models and Analysis
- Recommendations
- Future Work
- Questions

Our Agenda for today:

- Project Overview
- Datasets Utilized
- Methodology
- How We Chose the Zip Codes
- Time Series Models and Analysis
- Recommendations
- Future Work
- Questions

Project Overview

- We are given a dataset of data from Zillow
- Our task is to find the top 5 zip codes
- Our method for narrowing down from state to city to zip code
- Forecasting trends with FaceBook Prophet
- Analysis
- Recommendations
- Conclusions/Additional Observations

Project Overview:

- We are given a dataset of data from Zillow
- Our task is to find the top 5 zip codes
- Our method for narrowing down from state to city to zip code
- Forecasting trends with FaceBook Prophet
- Analysis
- Recommendations
- Conclusions/Additional Observations

We will also talk about Future Work at the end and leave room for questions

The Data We Used

- Zillow Dataset - Flatiron School

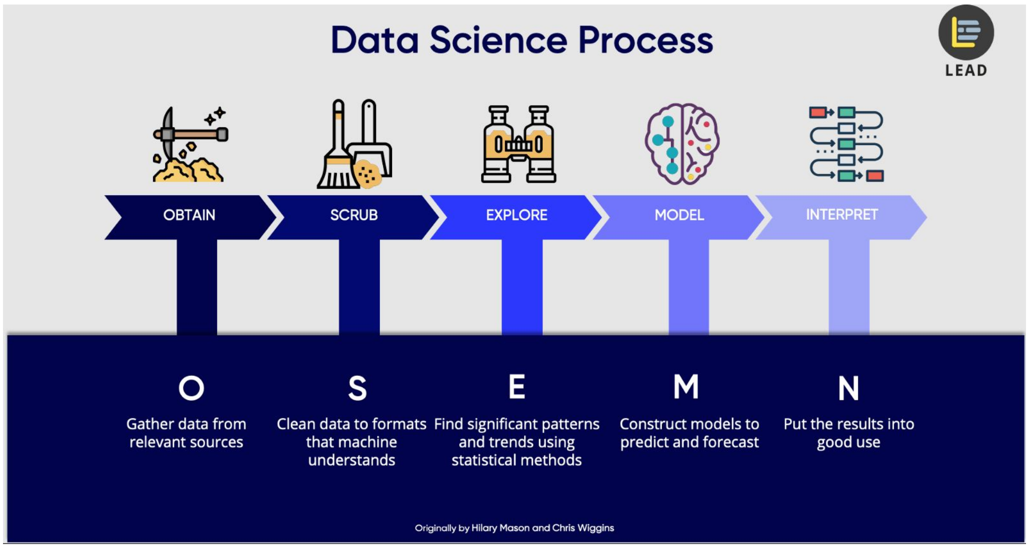
- U.S. Census Custom Dataset
  - Once we find our zip codes to search

We obtained the Zillow dataset from our partners over at Flatiron School.

The Custom Dataset we had to call the U.S. Census API with a custosmized list of zip codes to sort through

Our Methodology

Source : https://towardsdatascience.com/tagged/osemn



The process we used is OSEMN

Deciding on the State

Top 10 States in Numeric Growth: 2016 to 2017

Rank	Name	2010	2016	2017	Numeric growth
1	Texas	25,146,100	27,904,862	28,304,596	399,734
2	Florida	18,804,594	20,656,589	20,984,400	327,811
3	California	37,254,518	39,296,476	39,536,653	240,177
4	Washington	6,724,545	7,280,934	7,405,743	124,809
5	North Carolina	9,535,721	10,156,689	10,273,419	116,730
6	Georgia	9,688,690	10,313,620	10,429,379	115,759
7	Arizona	6,392,309	6,908,642	7,016,270	107,628
8	Colorado	5,029,325	5,530,105	5,607,154	77,049
9	Tennessee	6,346,295	6,649,404	6,715,984	66,580
10	South Carolina	4,625,381	4,959,822	5,024,369	64,547

Source: https://www.census.gov/newsroom/press-releases/2017/estimates-idaho.html#:~:text=DEC.,state%20population%20estimates%20released%20today

The article was a press release based on the data that the US Census collected. They said Idaho is the clear winner but upon my analysis of the states, Texas stands out.

Deciding on the State

Top 10 Most Populous States: 2017

Rank	Name	2010	2016	2017
1	California	37,254,518	39,296,476	39,536,653
2	Texas	25,146,100	27,904,862	28,304,596
3	Florida	18,804,594	20,656,589	20,984,400
4	New York	19,378,110	19,836,286	19,849,399
5	Pennsylvania	12,702,857	12,787,085	12,805,537
6	Illinois	12,831,565	12,835,726	12,802,023
7	Ohio	11,536,730	11,622,554	11,658,609
8	Georgia	9,688,690	10,313,620	10,429,379
9	North Carolina	9,535,721	10,156,689	10,273,419
10	Michigan	9,884,129	9,933,445	9,962,311

Source: https://www.census.gov/newsroom/press-releases/2017/estimates-idaho.html#:~:text=DEC.,state%20population%20estimates%20released%20today

Deciding on the State

Top 10 States in Percentage Growth: 2016 to 2017

Rank	Name	2010	2016	2017	Percent growth
1	Idaho	1,567,650	1,680,026	1,716,943	2.2
2	Nevada	2,700,691	2,939,254	2,998,039	2.0
3	Utah	2,763,889	3,044,321	3,101,833	1.9
4	Washington	6,724,545	7,280,934	7,405,743	1.7
5	Florida	18,804,594	20,656,589	20,984,400	1.6
6	Arizona	6,392,309	6,908,642	7,016,270	1.6
7	Texas	25,146,100	27,904,862	28,304,596	1.4
8	District of Columbia	601,766	684,336	693,972	1.4
9	Colorado	5,029,325	5,530,105	5,607,154	1.4
10	Oregon	3,831,072	4,085,989	4,142,776	1.4

Source: https://www.census.gov/newsroom/press-releases/2017/estimates-idaho.html#:~:text=DEC.,state%20population%20estimates%20released%20today

Deciding on the State

Most Populous	Numeric Growth	Percentage Growth
California	Texas	Idaho
Texas	Florida	Nevada
Florida	California	Utah
New York	Washington	Washington
Pennsylvania	North Carolina	Florida
Illinois	Georgia	Arizona
Ohio	Arizona	Texas
Georgia	Colorado	District of Columbia
North Carolina	Tennessee	Colorado
Michigan	South Carolina	Oregon

We see here there are some close runner ups which I would love to explore for future work.

Texas shows up as most populous and percentage growth which I think weighs more than just percentage growth.

Deciding on City and Zip Codes

```
n [23]: df_zillow_1['City'].value_counts()
```

```
ut[23]: Houston          86
        San Antonio      47
        Austin           38
        Dallas           33
        Fort Worth       26
                         ..
        Aldine            1
        Weatherford       1
        Mount Pleasant    1
        Canton            1
        Mart              1
        Name: City, Length: 540, dtype: int64
```

We did a count of the number of occurrences in the Zillow dataset of how many zip codes are per city in Texas and Houston looks like the clear winner.

Deciding on City and Zip Codes

```
df_texas_census.head()
```

Top Zip Codes

	NAME	S0101_C01_001E
[illegible]	ZCTA5 77084	104582
28	ZCTA5 77036	76605
80	ZCTA5 77095	72081
59	ZCTA5 77072	62162
63	ZCTA5 77077	57757

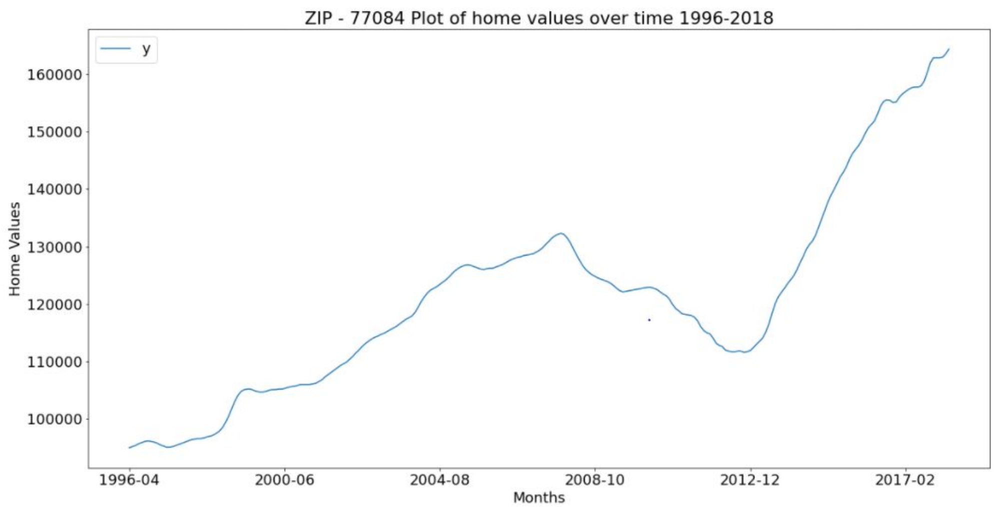
Population per zip code

We input Houston and its list of zip codes into the U.S. Census to get a list of populations for each zip code. Then we sort it and choose the top 5 most populated zip codes.

Our Zip Codes to Analyze

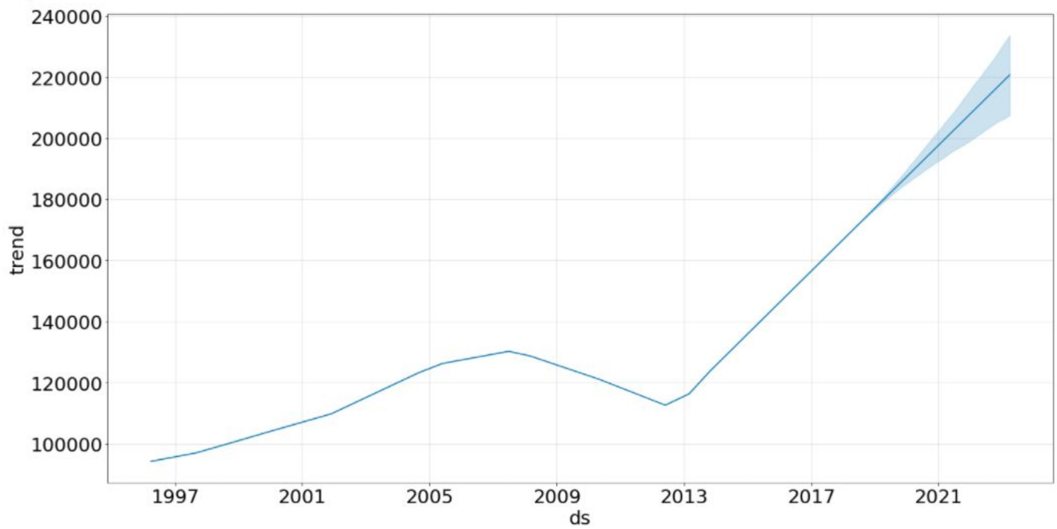
1) 77084
2) 77036
3) 77095
4) 77072
5) 77077

Models and Analysis - 77084



This is the initial plot of the data

Models and Analysis - 77084



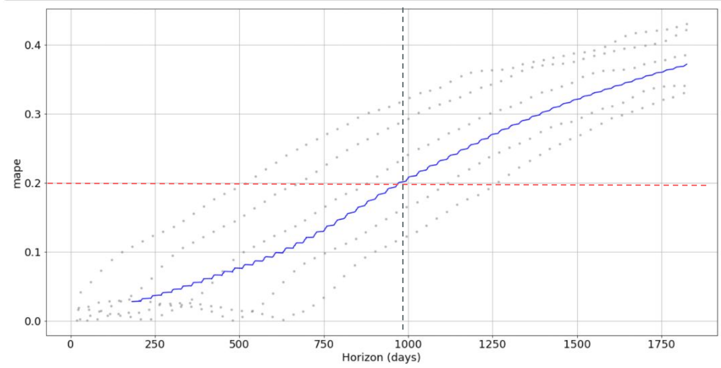
This is the time series forecast with FBProphet of 77084. For this and the future forecasts, Seasonality is not addressed because there is a huge upward trent in price after the 2008 market crash. This is not an indication of either seasonal nor cyclical nature. I’ve considered removing the data outliers and analyze the data from 1996 to 2010 but that is a huge chunk of data lost and I decided against that.

Models and Analysis - 77084

MAPE (Mean Average Percent Error) - Observation:

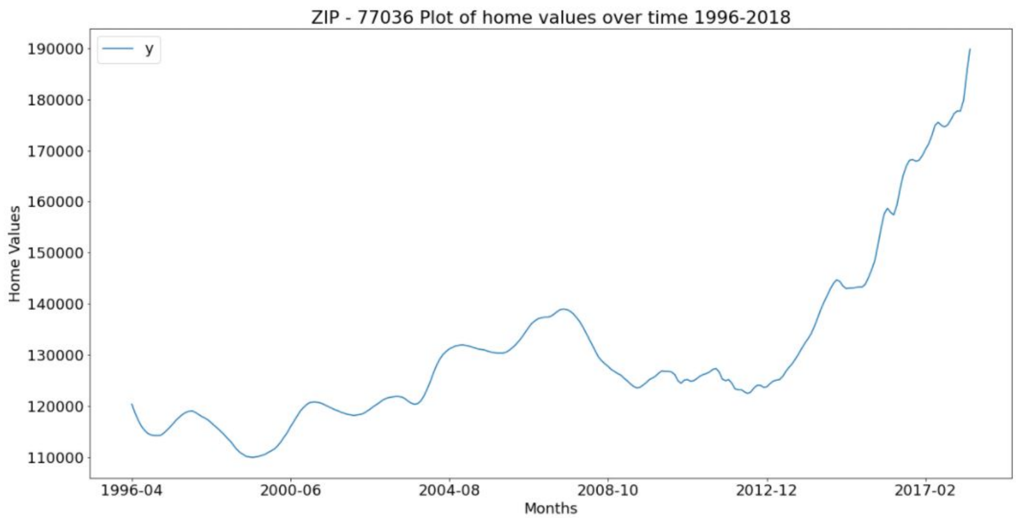
1. We see that MAPE increases over time
2. I am willing to tolerate MAPE of 0.1 to 0.2
   - This gets exceeded after about 1100 days
3. We will focus on MAPE as our main diagnostic metric.
   - Shows the model was about 80% accurate at 1000 days
   - Bullish prediction for the next 2-3 years
   - Supports the high upward trend we saw in the graph of all the data points for the zip code

[12]:
```
fig = plot_cross_validation_metric(cv_results, metric='mape', figsize=(20,10))
```

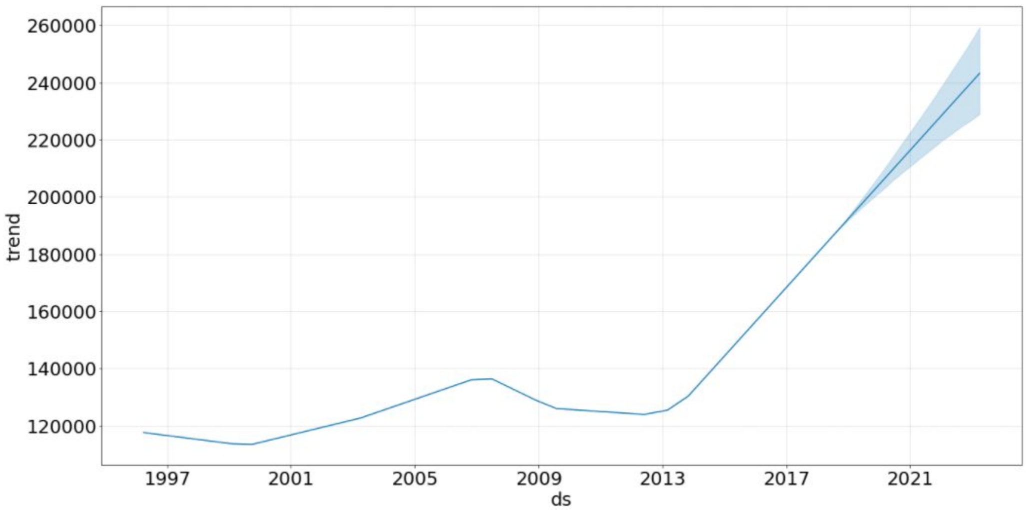


MAPE and forecast support the trend upward for this zip code

Models and Analysis - 77036



Models and Analysis - 77036

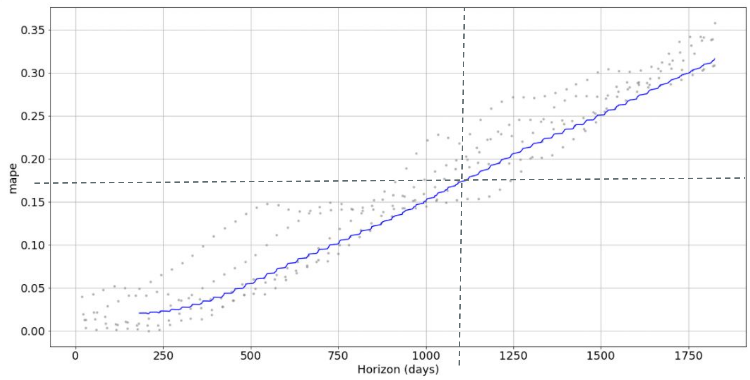


Models and Analysis - 77036

MAPE (Mean Average Percent Error) - Observation:

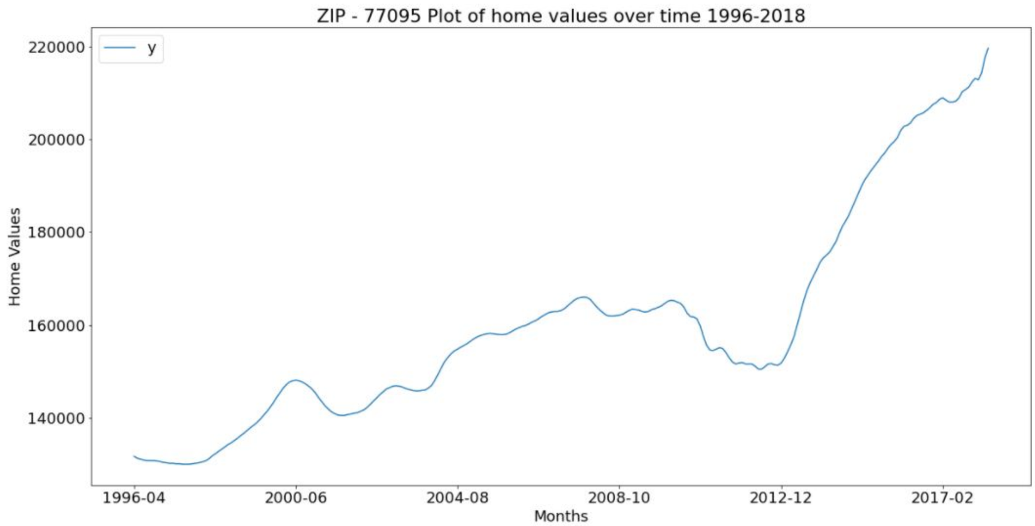
1. We see that MAPE increases over time
2. I am willing to tolerate MAPE of 0.1 to 0.2
    - This gets exceeded after about 1100 days
3. We will focus on MAPE as our main diagnostic metric.
    - Shows the model was about 85% accurate at 1000 days
    - Bullish prediction for the next 2-3 years
    - Supports the high upward trend we saw in the graph of all the data points for the zip code

n [12]:
```
fig = plot_cross_validation_metric(cv_results, metric='mape', figsize=(20,10))
```

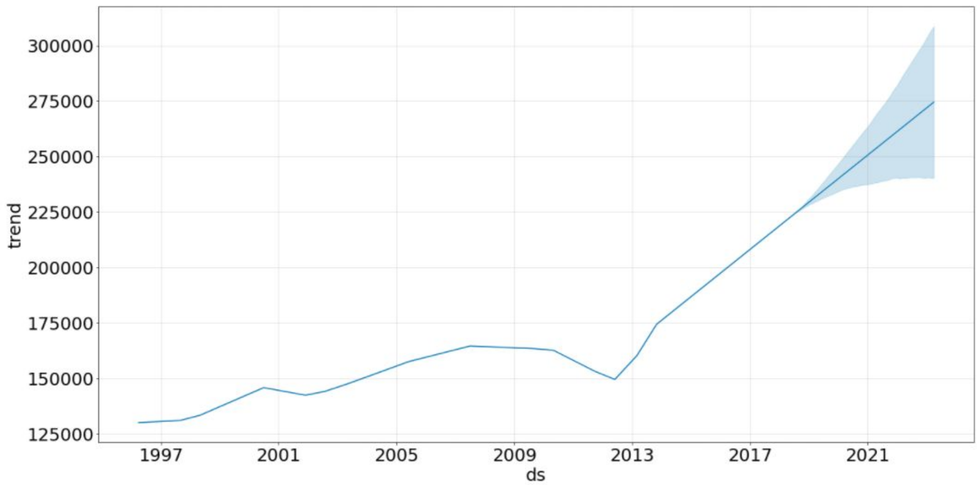


MAPE and forecast support the trend upward for this zip code

Models and Analysis - 77095



Models and Analysis - 77095



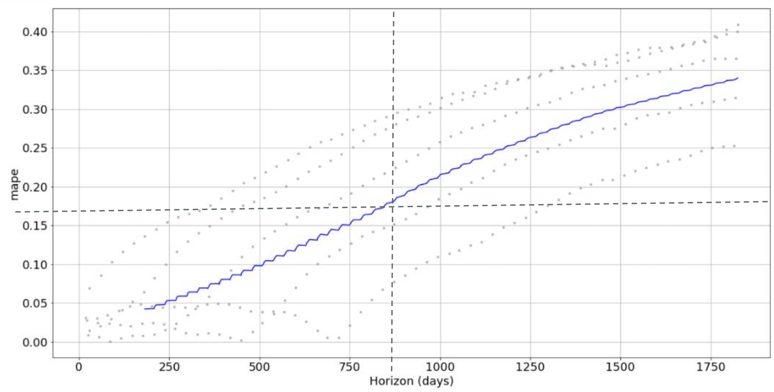
There is slightly more range in the forecast here for 77095. Cone not too narrow

Models and Analysis - 77095

MAPE (Mean Average Percent Error) - Observation:

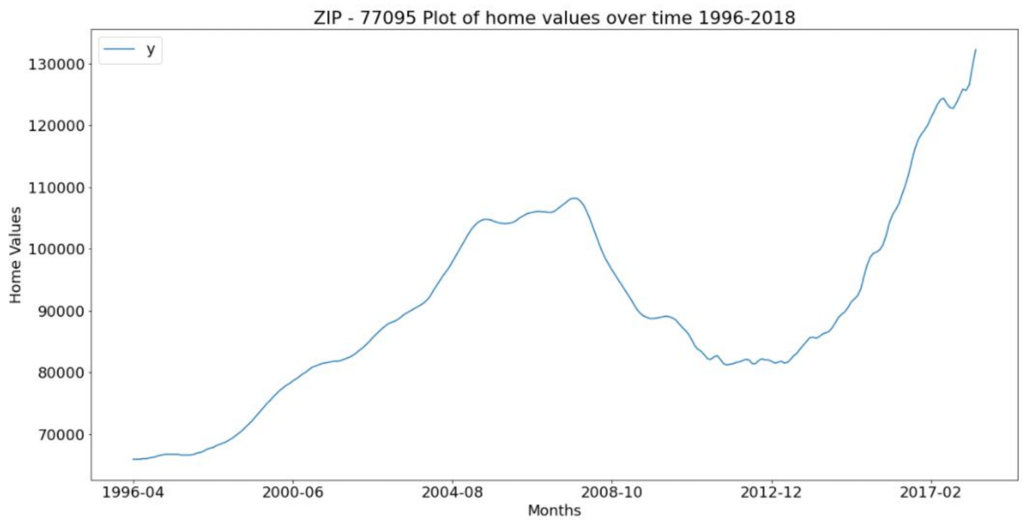
1. We see that MAPE increases over time
2. I am willing to tolerate MAPE of 0.1 to 0.2
   - This gets exceeded after about 800 days
3. We will focus on MAPE as our main diagnostic metric.
   - Shows the model was about 80% accurate at 800 days
   - Bullish prediction for the next 2-3 years
   - Supports the high upward trend we saw in the graph of all the data points for the zip code

In [12]:
```
fig = plot_cross_validation_metric(cv_results, metric='mape', figsize=(20,10))
```

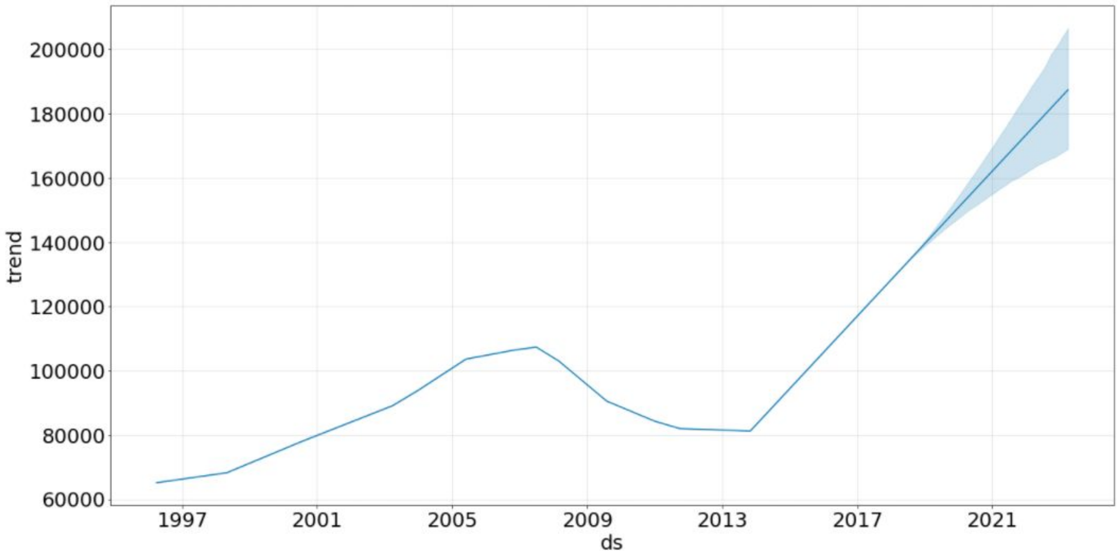


MAPE and forecast support the trend upward for this zip code

Models and Analysis - 77072



Models and Analysis - 77072

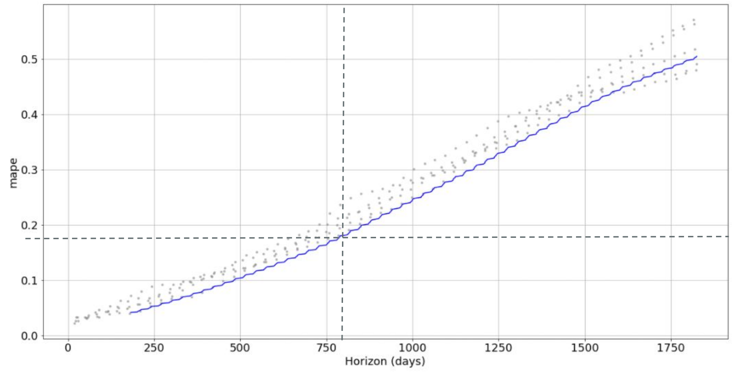


Models and Analysis - 77072

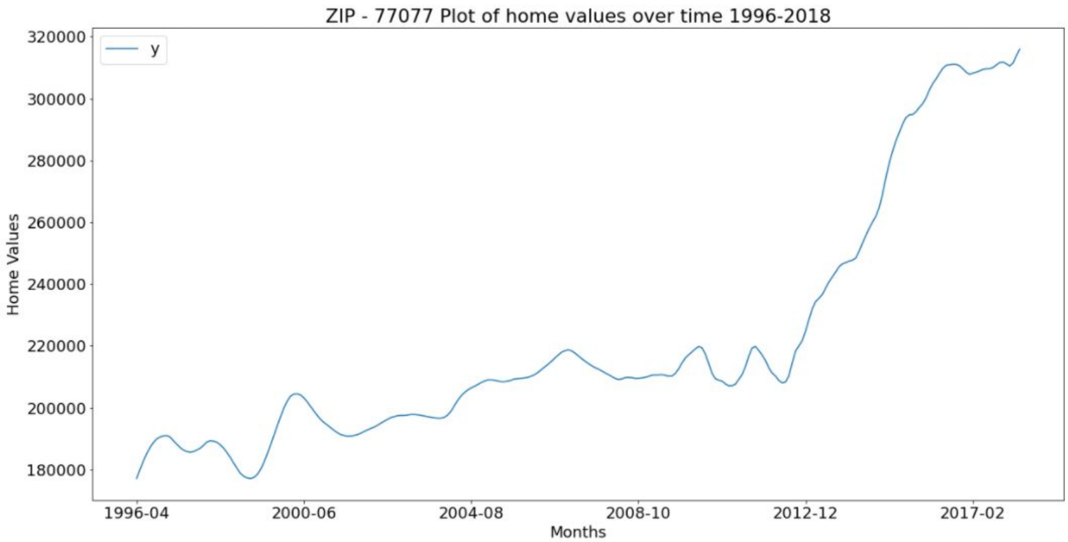
MAPE (Mean Average Percent Error) - Observation:

1. We see that MAPE increases over time
2. I am willing to tolerate MAPE of 0.1 to 0.2
    - This gets exceeded after about 800 days
3. We will focus on MAPE as our main diagnostic metric.
    - Shows the model was about 80% accurate at 800 days
    - Bullish prediction for the next 2-3 years
    - Supports the high upward trend we saw in the graph of all the data points for the zip code

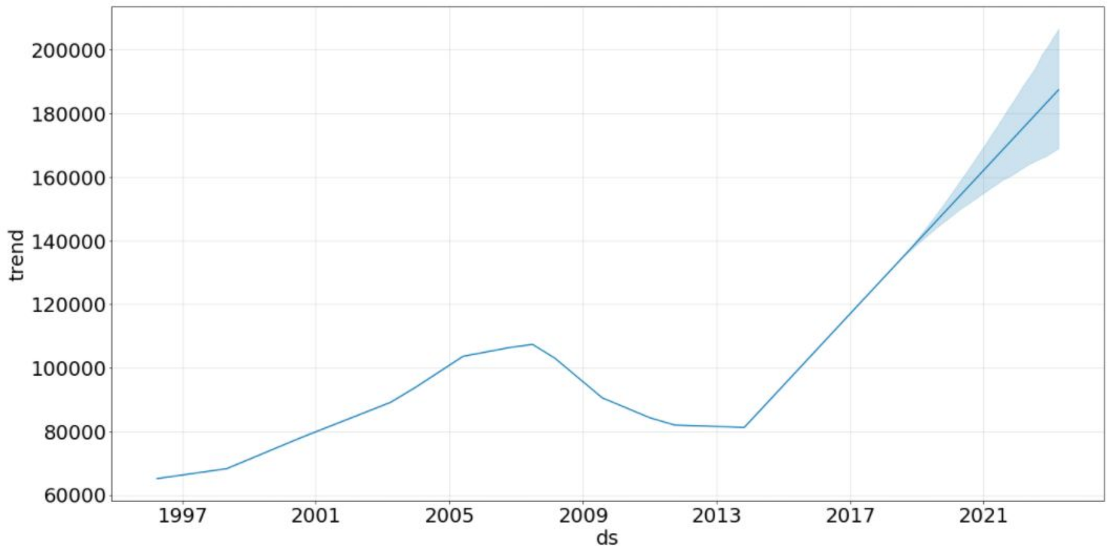
In [14]:
```
fig = plot_cross_validation_metric(cv_results, metric='mape', figsize=(20,10))
```



Models and Analysis - 77077



Models and Analysis - 77077

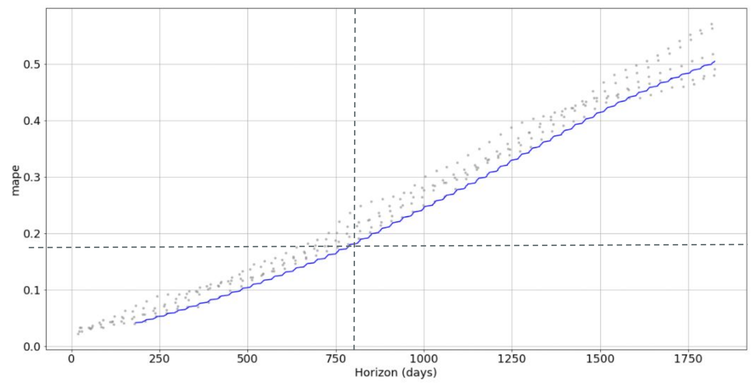


Models and Analysis - 77077

MAPE (Mean Average Percent Error) - Observation:

```
1. We see that MAPE increases over time
2. I am willing to tolerate MAPE of 0.1 to 0.2
    - This gets exceeded after about 800 days
3. We will focus on MAPE as our main diagnostic metric.
    - Shows the model was about 80% accurate at 800 days
    - Bullish prediction for the next 2-3 years
    - Supports the high upward trend we saw in the graph of all the data points for the zip code
```

```
In [14]: fig = plot_cross_validation_metric(cv_results, metric='mape', figsize=(20,10))
```



Recommendations

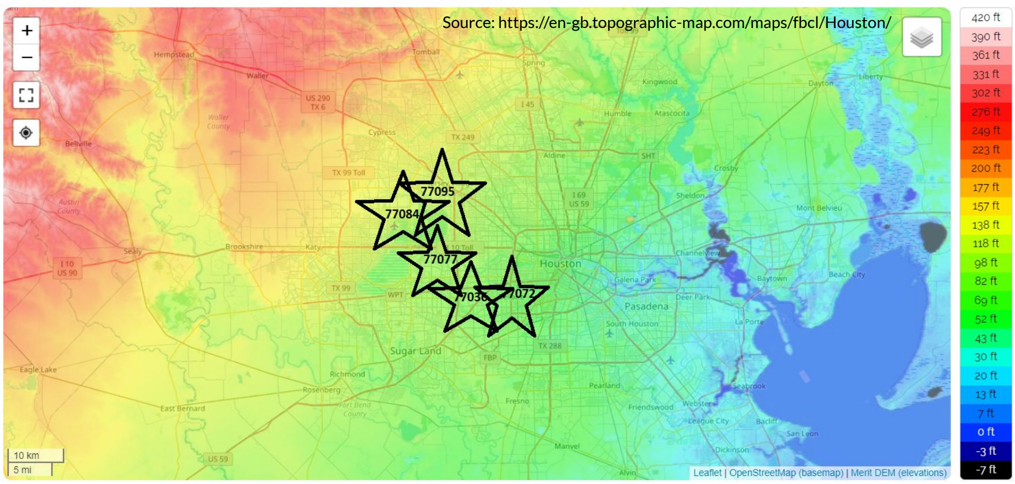
- All 5 zip codes show promise of growth. Good starting points.
  a. 77084
  b. 77036
  c. 77095
  d. 77072
  e. 77077

- High upward trends
  a. As seen in our models, shows promise of future growth

- Per the Census, Texas is one of the highest growing states
  a. Bonus - According to Apartments.com, Texas is a landlord friendly state.
     - Source: https://www.apartments.com/rental-manager/resources/article/top-10-u.s.-cities-for-buying-rental-property

Conclusion/Additional Observations

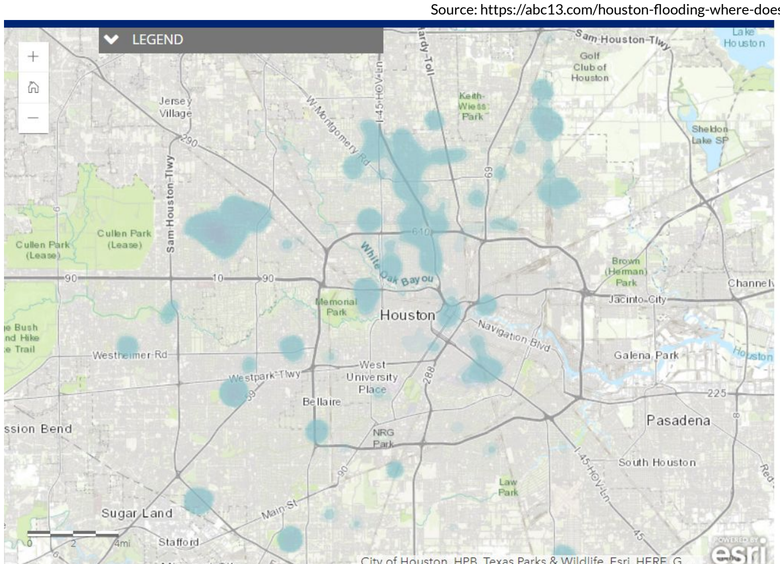
- Houston is near the Gulf of Mexico - flood warnings/hurricanes



Houston has flood zones. However, the choices that we have are in slightly higher elevation areas. I've provided the areas below on a topographic map.

Conclusion/Additional Observations

Source: https://abc13.com/houston-flooding-where-does-it-flood-in/5683641/



There is also a flood map of 311 calls from ABC news which talks about where specifically people were calling from saying their homes are flooded. These can help further narrow our choices down to the house/block level.

Future Work

1. Explore/Use Crime Data from the federal government
2. Zillow Word Cloud on the MLS database
    - Uses NLP (Natural Language Processing)
    - This can be useful for finding house patterns like how many bedrooms and bathrooms in each zip code
3. Utilize different time series tools other than Fbprophet
    - Limitations with tool for monthly data
4. Obtain Daily instead of monthly dataset of home values from same time range and re-apply analysis
5. Explore the other states that were runner-ups

If I had the time and resources dedicated, I would do the following:

1. Explore/Use Crime Data from the federal government
2. Zillow Word Cloud on the MLS database
    - Uses NLP (Natural Language Processing)
    - This can be useful for finding house patterns like how many bedrooms and bathrooms in each zip code
3. Utilize different time series tools other than Fbprophet
    - Limitations with tool for monthly data
4. Obtain Daily instead of monthly dataset of home values from same time range and re-apply analysis
5. Explore the other states that were runner-ups

Thank You!

Source: http://slworkshop.net/wp-content/uploads/2015/04/thank-you-wordle.jpg

Questions?

Source: https://www.quickanddirtytips.com/sites/default/files/images/7791/questions.jpg