



Phase 4: Time Series

DANIEL JOURNIC

Contents

1. Business Problem

2. The Data

3. Process

4. Conclusion

Business Problem:

A high end real estate developer has tasked us with finding properties In PA to invest in towards building commercial properties.

1. Identify 5 zip codes for investment
2. Forecast potential growth

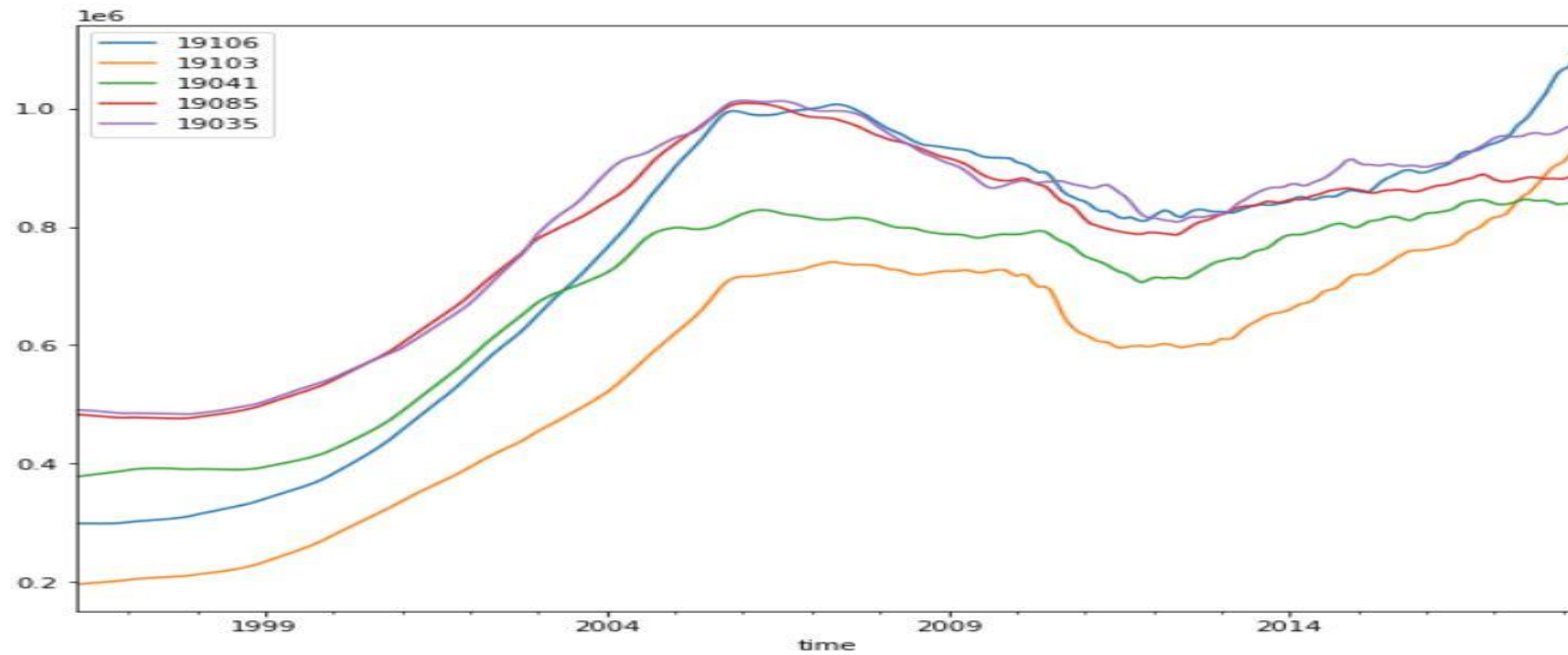
The Data

Zillow housing data:

	RegionID	RegionName	City	State	Metro	CountyName	SizeRank	1996-04	1996-05	1996-06	...	2017-07	2017-08	2017-09	2017-10	201
118	65133	17603	Lancaster	PA	Lancaster	Lancaster	119	87700.0	89200.0	90600.0	...	148400	149100	149800	150400	1513
125	64180	15601	Greensburg	PA	Pittsburgh	Westmoreland	126	36100.0	36000.0	36100.0	...	141200	142600	143900	144400	1447
194	65779	19111	Philadelphia	PA	Philadelphia	Philadelphia	195	84600.0	84500.0	84400.0	...	167100	168100	169300	170600	1719
282	65791	19124	Philadelphia	PA	Philadelphia	Philadelphia	283	43100.0	43000.0	42900.0	...	79800	80900	82200	83000	838
329	65933	19446	Lansdale	PA	Philadelphia	Montgomery	330	155300.0	154900.0	154600.0	...	305400	307700	309200	309300	3088

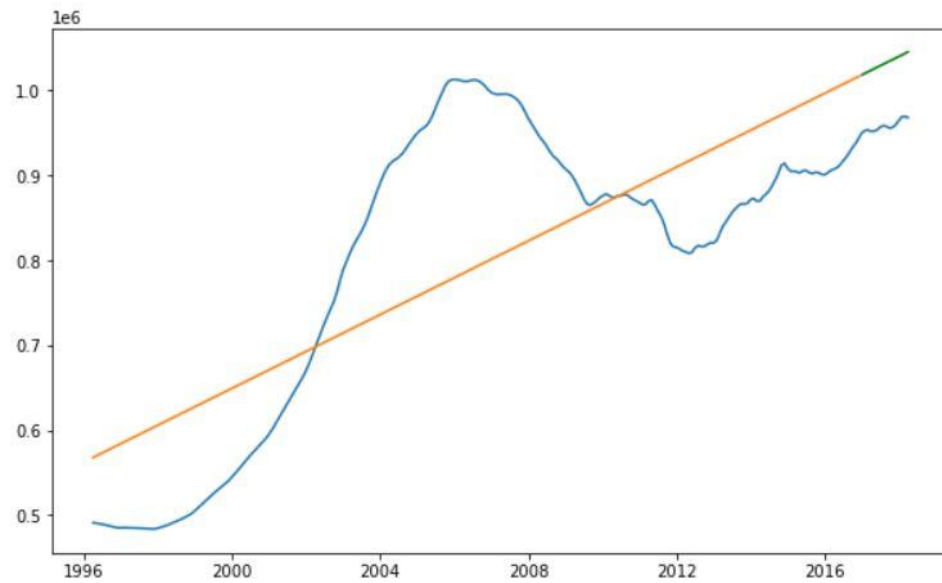
The Process

After melting the data, and plotting:

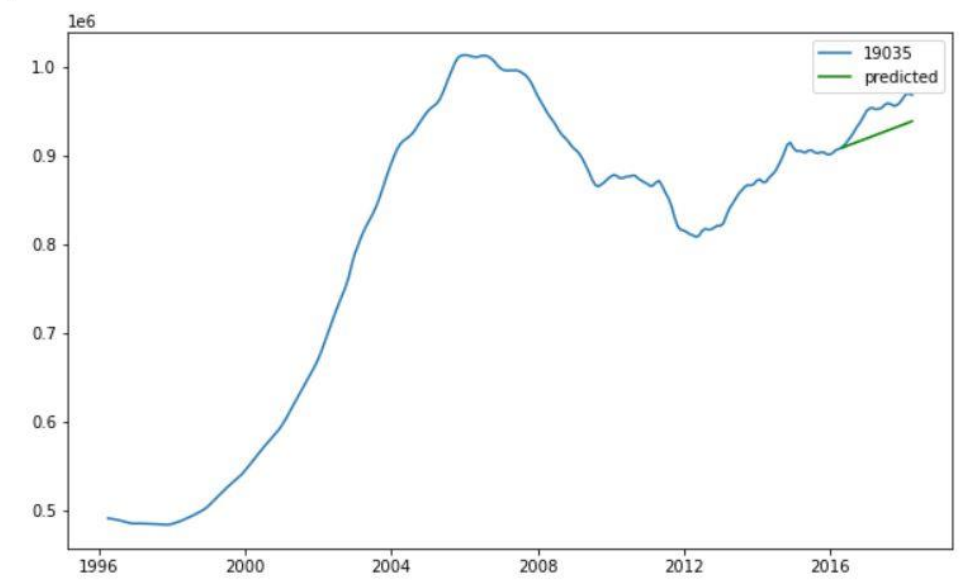


Process con't

BASELINE: LINEAR REGRESSION

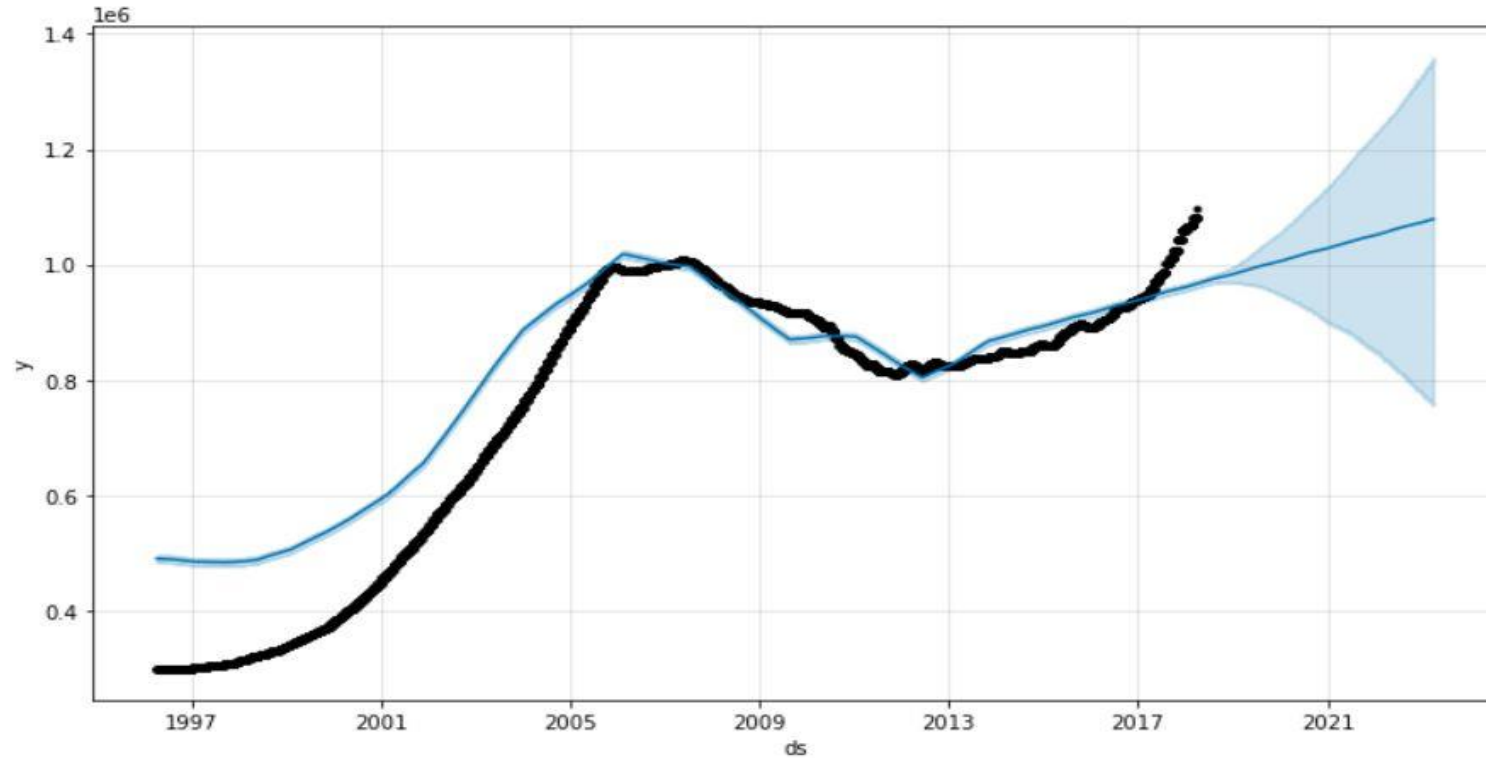


AUTO ARIMA



Process con't

Prophet:



Conclusions

Obviously, ARIMA did better than LR

Still more work to do with Prophet, but it has potential

Thank you

QUESTIONS?

DJOURNIC@GMAIL.COM