

Predicting Prices with Zillow Data

Time Series Forecasting and Analysis: Data Science

Orlando Vilar, Feb 2023

Introduction

Introduction

Overview & Business Understanding

- When the economic environment is challenging, investors might chose real assets as an investment to protect agains inflation;
- But how can one assess the best investment opportunities?
- I provide a framework for location ranking and house prices prediction for a real estate investment firm;
- **Relevance:** overall implication to investors decision-making process, realtors, contractors, and the general public.

Data & Analysis

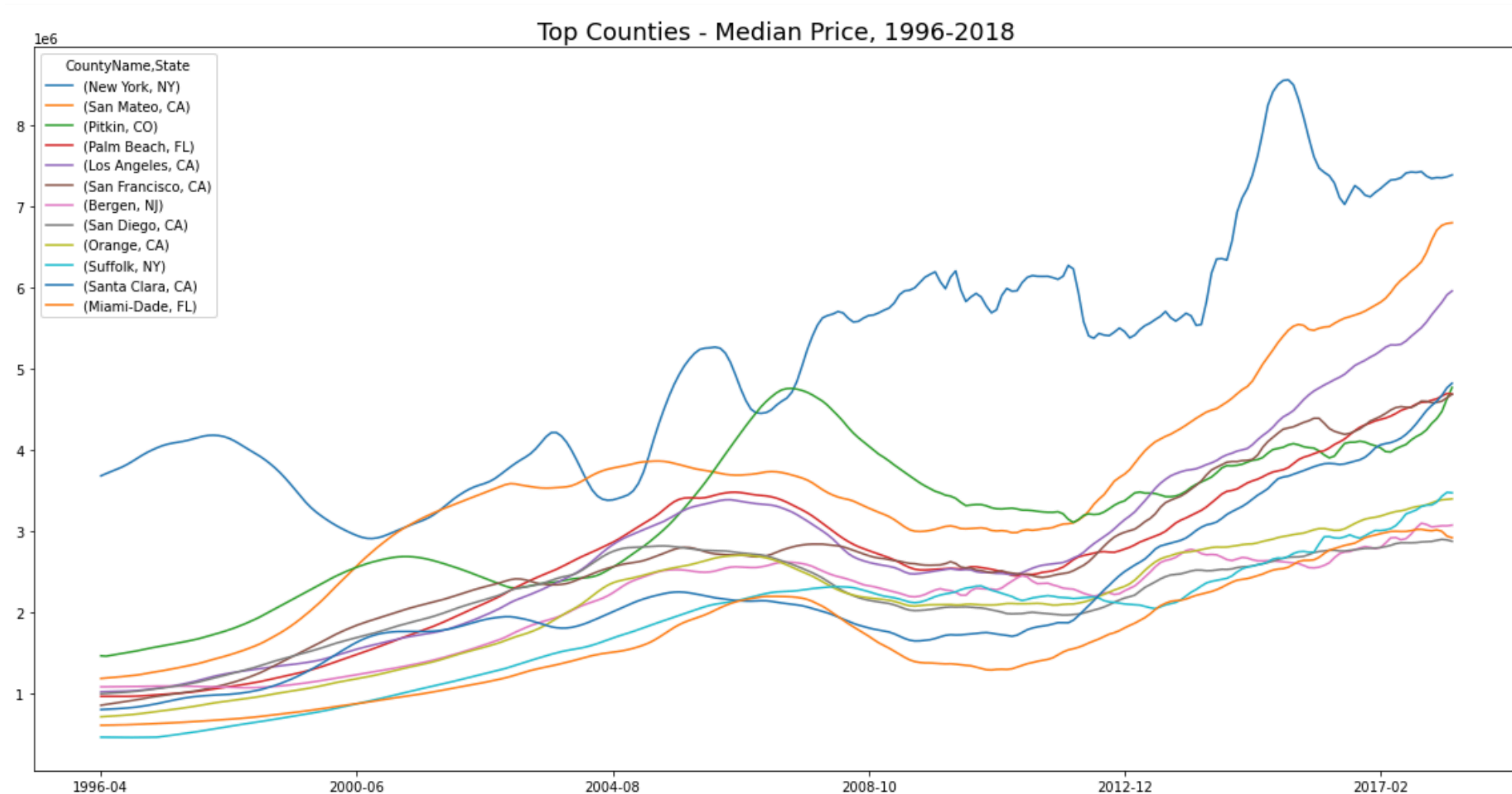
Data Understanding

Sources of Information

- Databases:
 - Zillow dataset;
 - More than 12.5 thousand locations from Apr. 1996 - Apr. 2018;
- Variables:
 - Prices, Region ID, Region Name, City, State, Metro area, County and Size Rank;
- Additional feature:
 - Matching of latitude and longitude to provide city/town location.

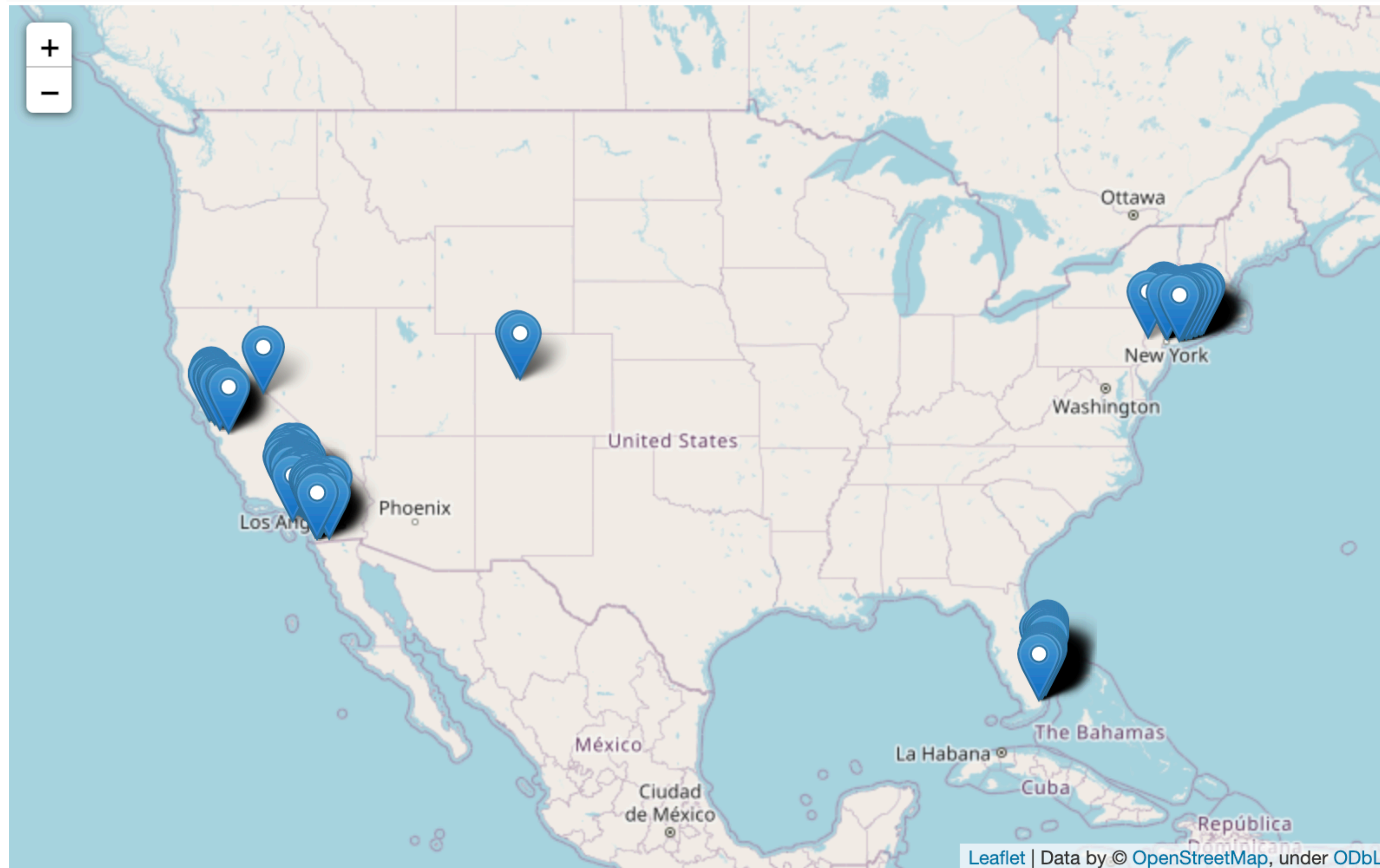
Counties by Median House Values

Top 10 most valuable locations



Counties by Median House Values

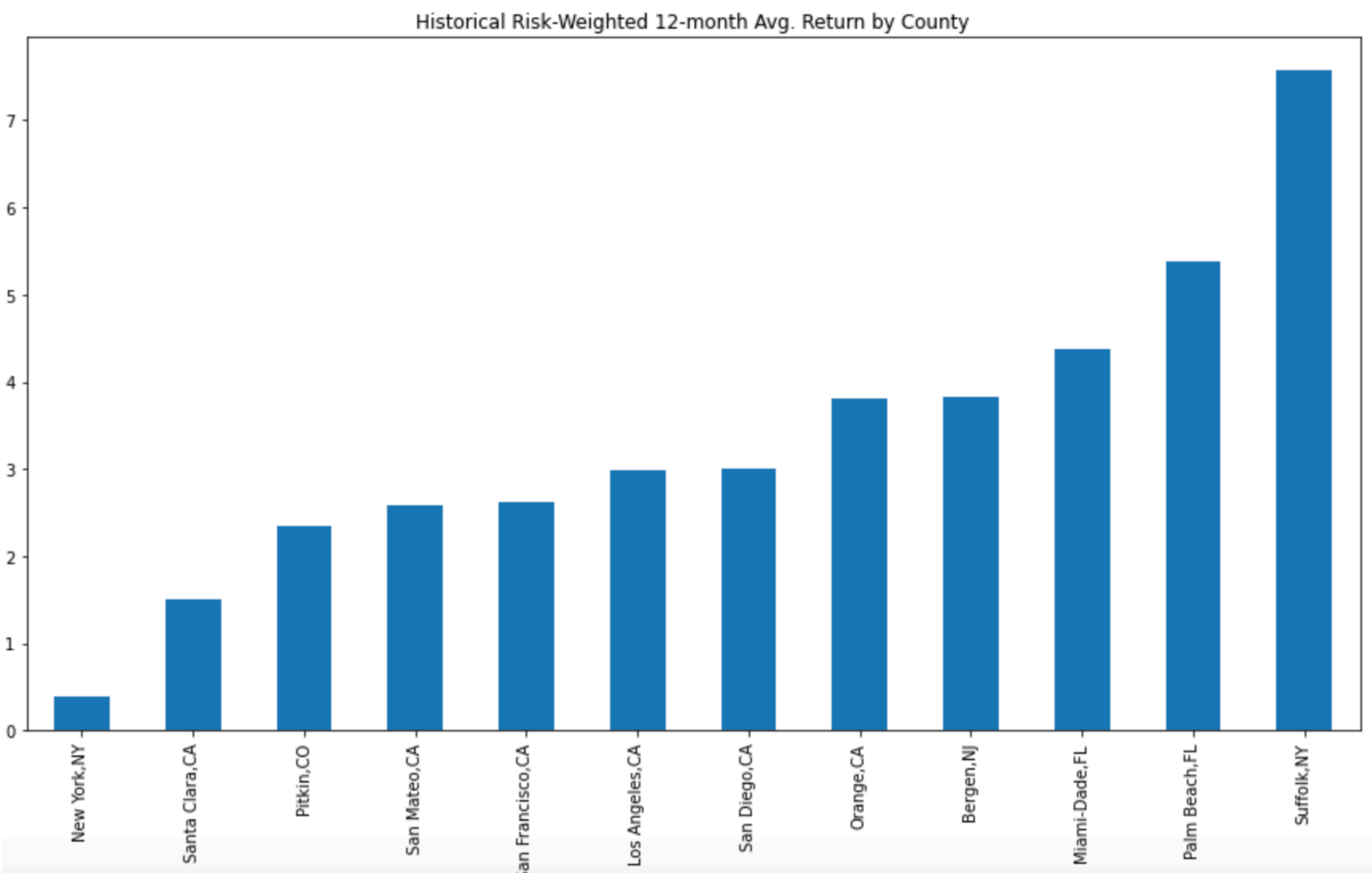
City/town location within the top 10 Counties



County Concentration
Over-representativeness of
locations in CA, FL and NY

Counties by Risk-Weighted Return

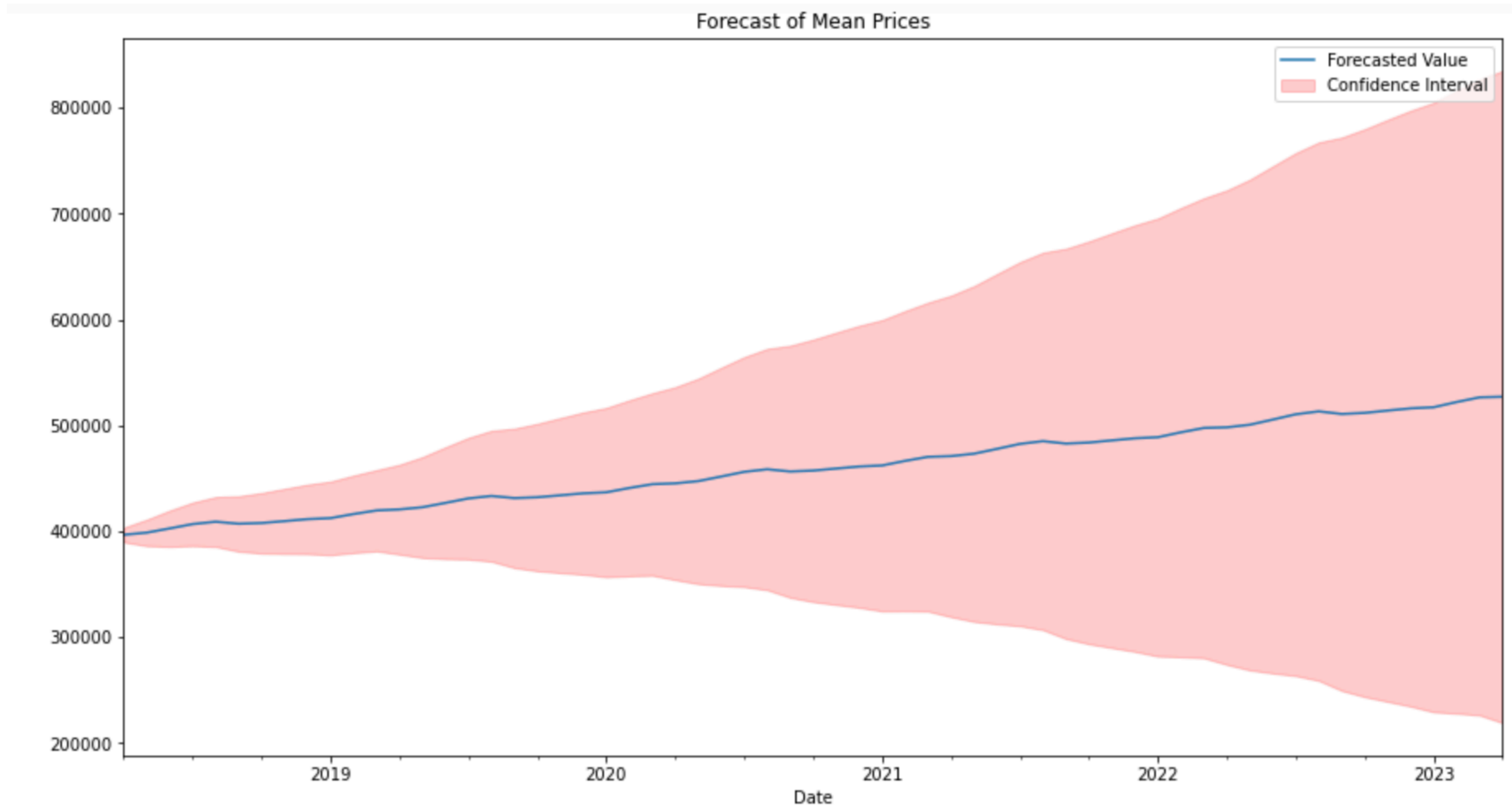
The ‘best bang for the buck’



Risk-Weighted Return
The best returns are located within Suffolk, NY; Palm Beach, FL and Miami-Dade, FL.

Forecasting Mean Prices

Suffolk, NY over a 60-month window



Closing Remarks

Strategies

Three key insights

- **Divide and Conquer:** investing at the top locations using a portfolio approach;
- **Rays of the Sun:** choosing one location and irradiating the investment strategy within an unique county;
- **Alternative:** investing in more established counties at the expense of more competition;
- Price prediction is more assertive over a shorter-term horizon (up to 24 months).

Limitations

Further analysis

- Adding demographic data about each specific area;
- Drilling down into one specific area for over/underpriced analysis;
- Embedding optimization under the Portfolio Theory framework.

Thank you!

<https://github.com/ovilar>