REAL ESTATE RADAR

By: Diahann Castellon: Data Engineer, Brandon Ingalz: Data Manager, Anushya Mani: Data Architect, Roland “Coy” Abellano: Senior Data Analyst, & Jimmy Nguyen: Data Scientist

Data Sourcing & ML Libraries Used

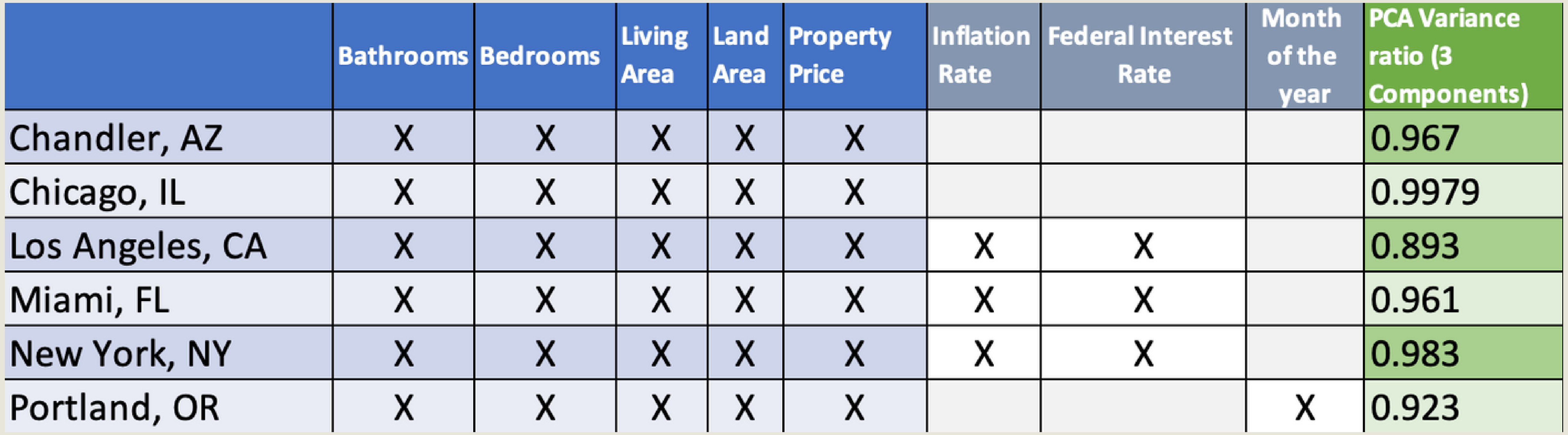
Data Sources Used:

- Zillow.com w/ Zillow Data Exporter (Google plug-in)
- PostgreSQL

Machine Learning Libraries:

- Python Pandas
- NumPy
- Matplotlib
- Seaborn
- Scikit-learn

What City fits your Housing Budget & Needs?



	Bathrooms	Bedrooms	Living Area	Land Area	Property Price	Inflation Rate	Federal Interest Rate	Month of the year	PCA Variance ratio (3 Components)
Chandler, AZ	X	X	X	X	X				0.967
Chicago, IL	X	X	X	X	X				0.9979
Los Angeles, CA	X	X	X	X	X	X	X		0.893
Miami, FL	X	X	X	X	X	X	X		0.961
New York, NY	X	X	X	X	X	X	X		0.983
Portland, OR	X	X	X	X	X			X	0.923

Factors not included:

•Real estate demand, per city/micro-city:

1. Seasonal impact?
2. Unemployment rates?
3. Education rating?
4. Safety?
5. Political impacts?

•HOA fees per property

•Last renovation date, per property

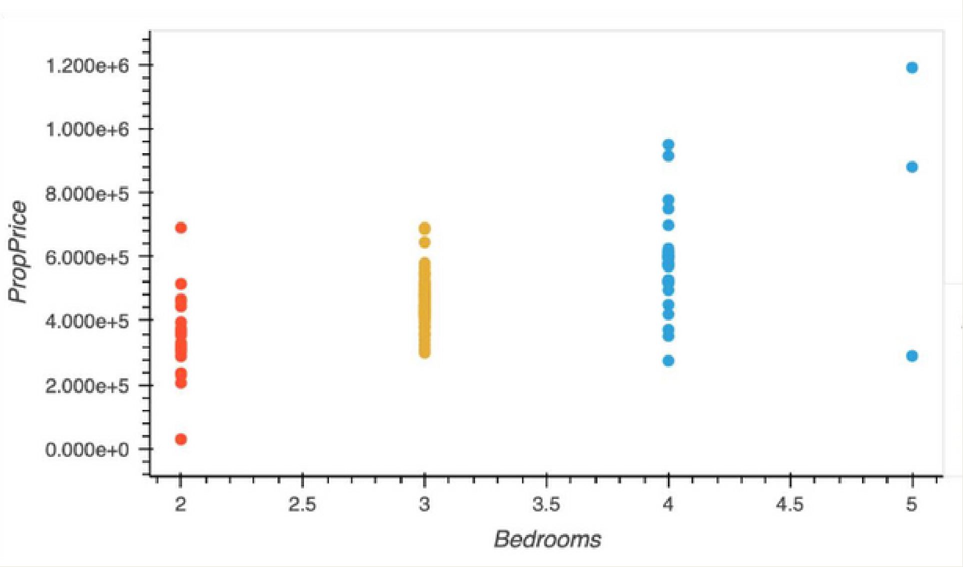
•Walkability, per property

•Other desirable property features such as: kitchen, basement, storage space, and parking

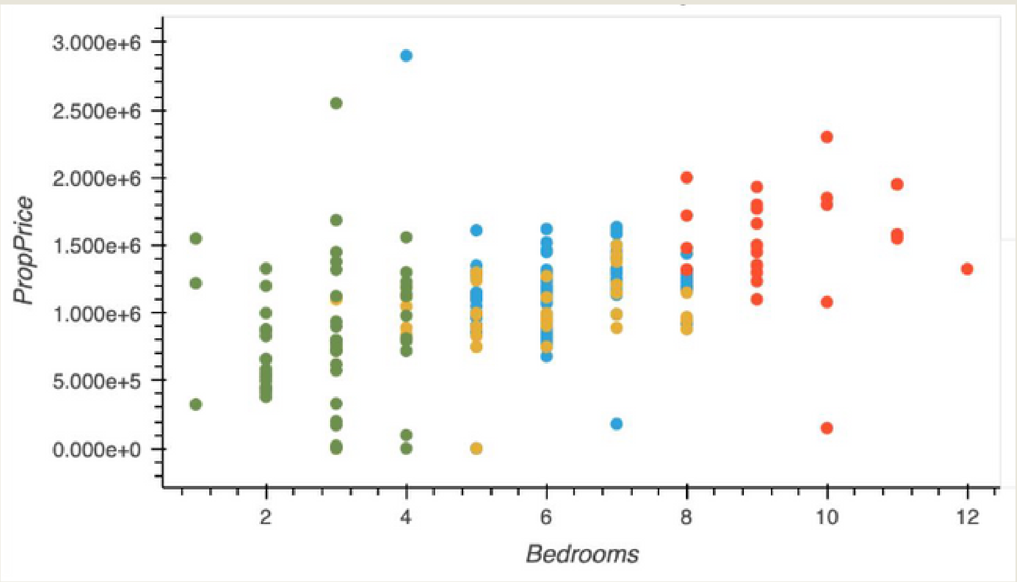
REAL ESTATE
RADAR

Property Clusters by City

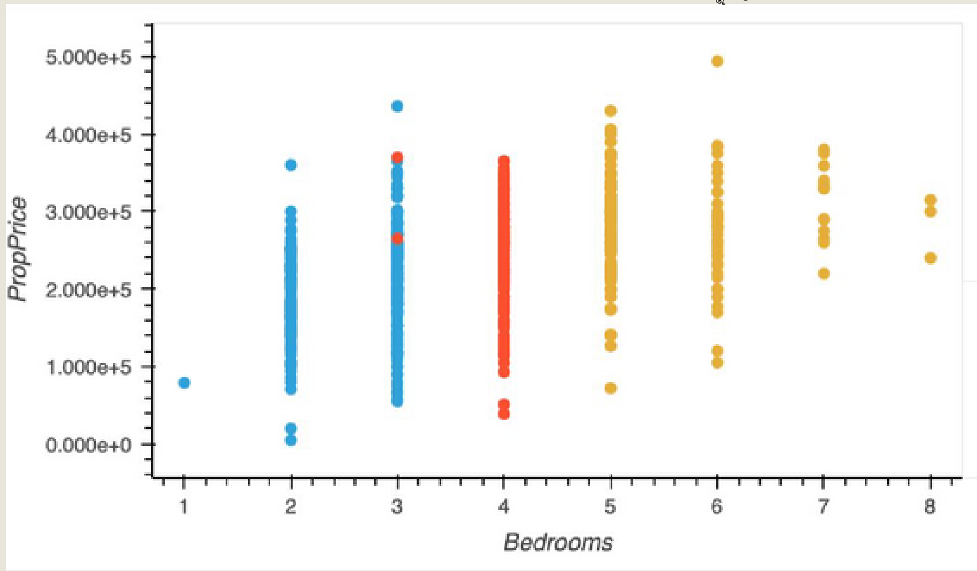
Chandler, AZ



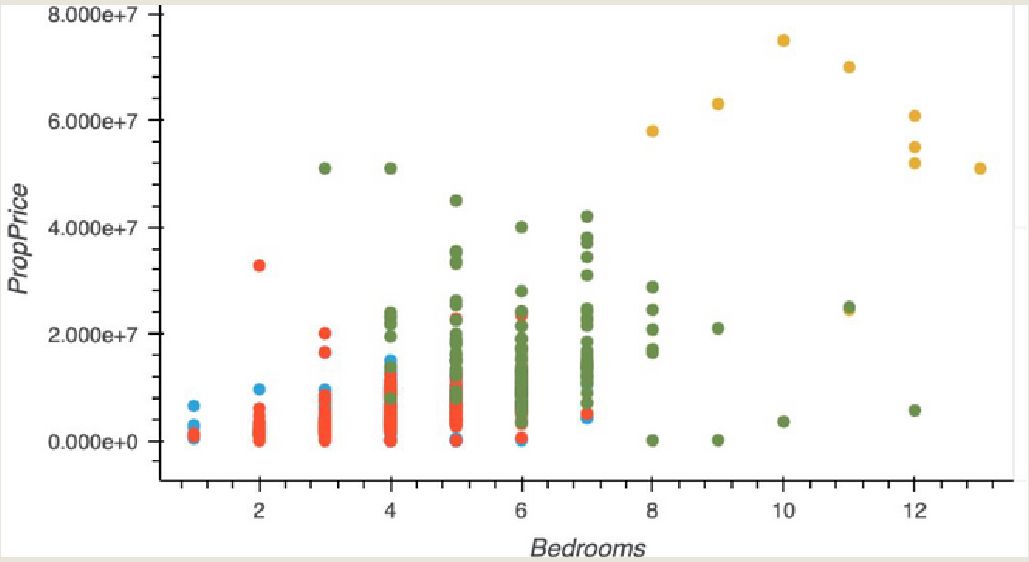
New York City, NY



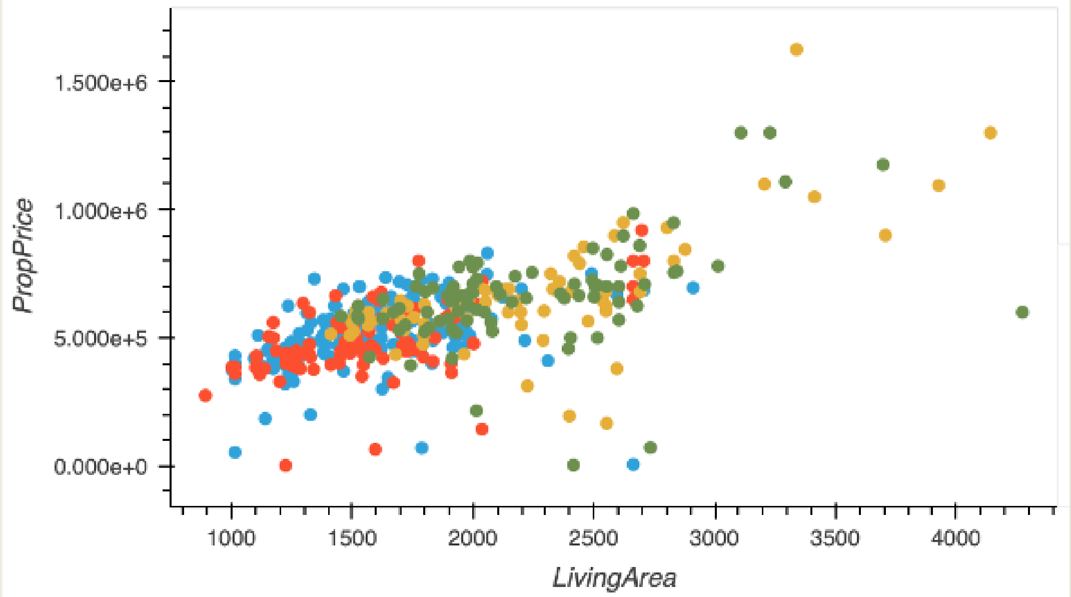
Chicago, IL



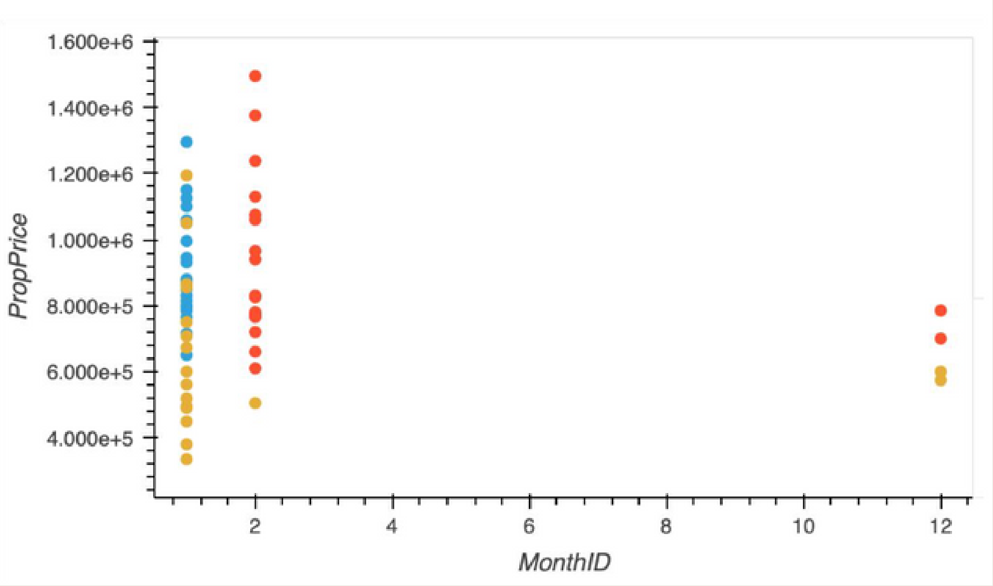
Los Angeles, CA



Miami, FL



Portland, OR



Potential Application

Create a calculator to pick the right city for you!

Where would you buy next?

Step 1: Pick the housing cluster(s) you can relate to the most, per city

Step 2: Normalize the (max$/feature), per feature across clusters. Use the inverse of these values

Note: Remove any feature that did not prove to be significant in the cluster analysis!

Step 3: Determine how important each feature is to you

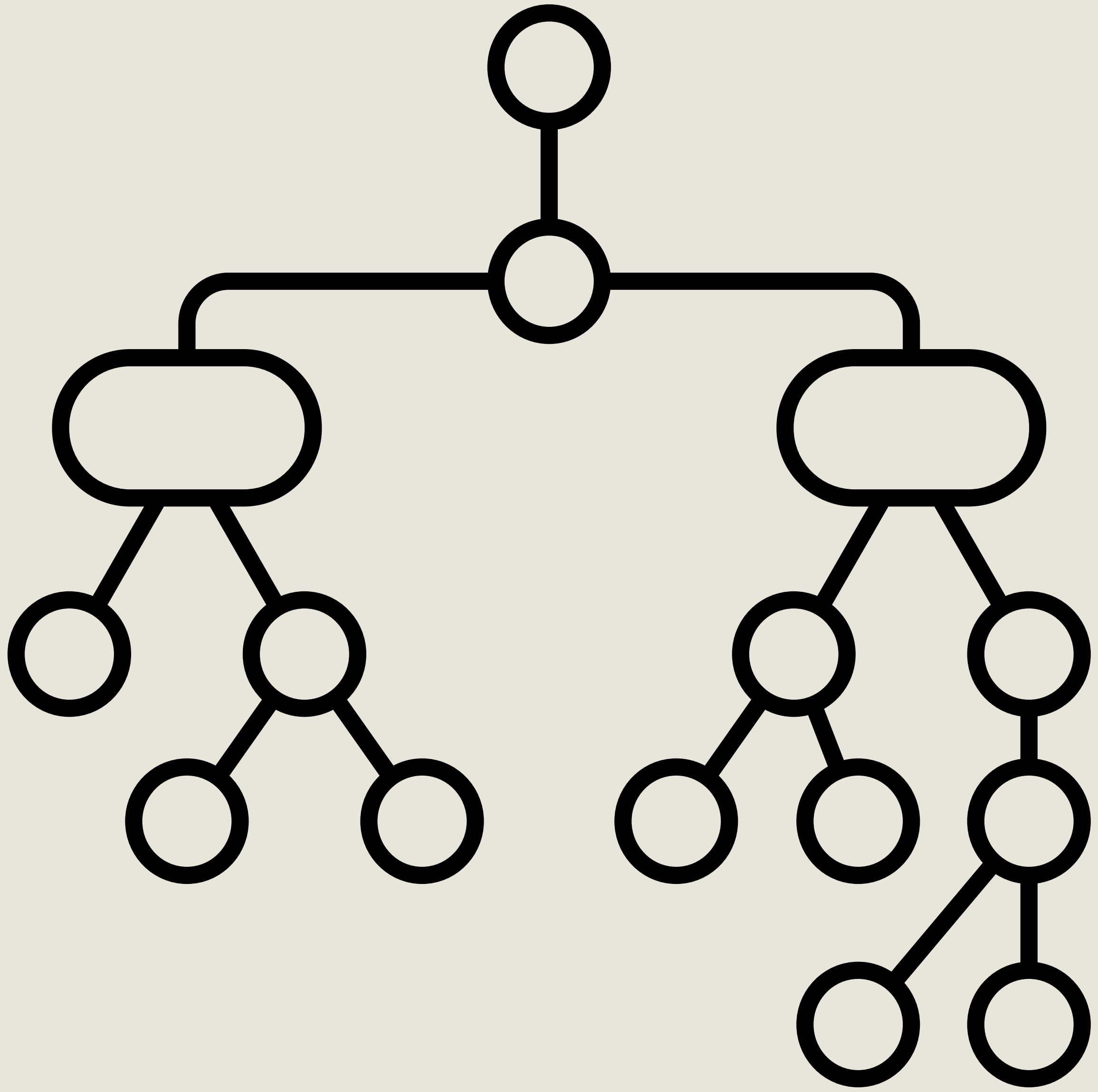
(Give a weightage for each property feature, sum should = 1):

Step 4: Calculate overall score for each cluster based on preferences, choose the cluster (and associated city) that scored the highest

(Example of scoring equation per city below):

ClusterX_City Y = (living_area_weight * ClusterX_CityY_norm[“living_area] + Land_area_weight * ClusterX_CityY_norm [“land_area“] + ...)

Predicting Whether Initial Home Price Offering Will Get Reduced



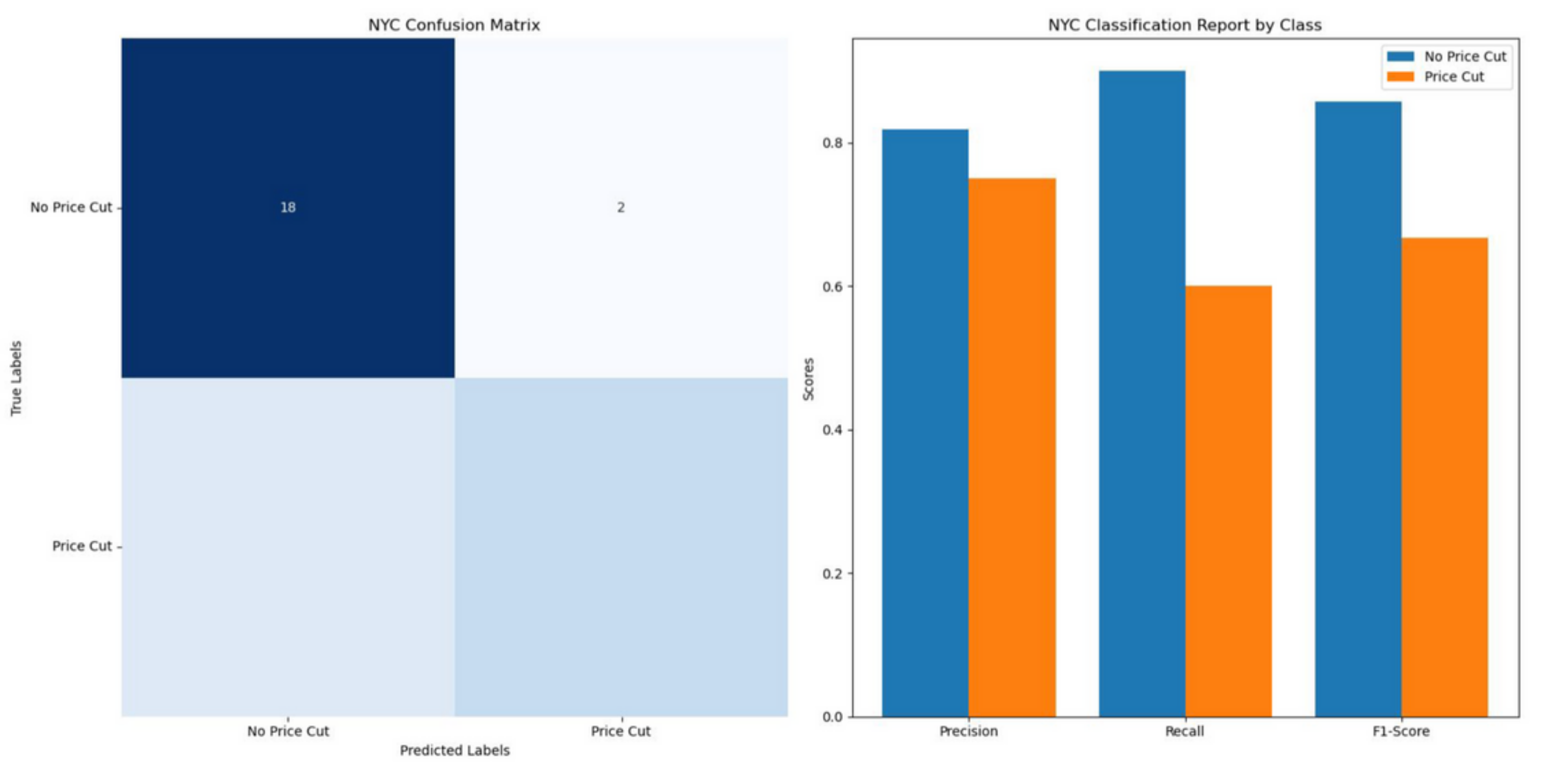
Supervised ML Models for Non-Linear Datasets:

- Decision Tree
- Random Forest

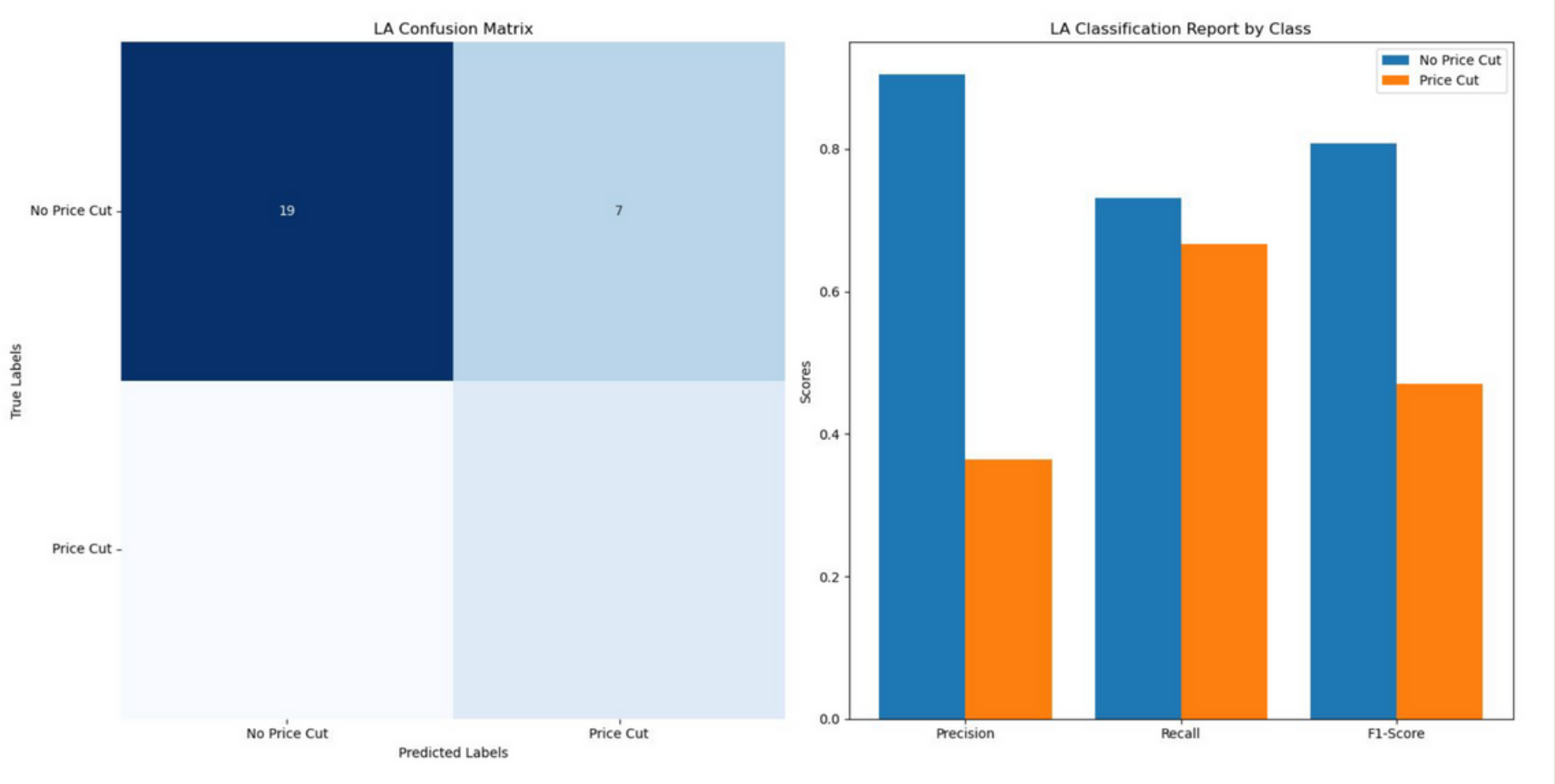
Creative with Feature Engineering is necessary

Supervised ML Model

New York City, NY

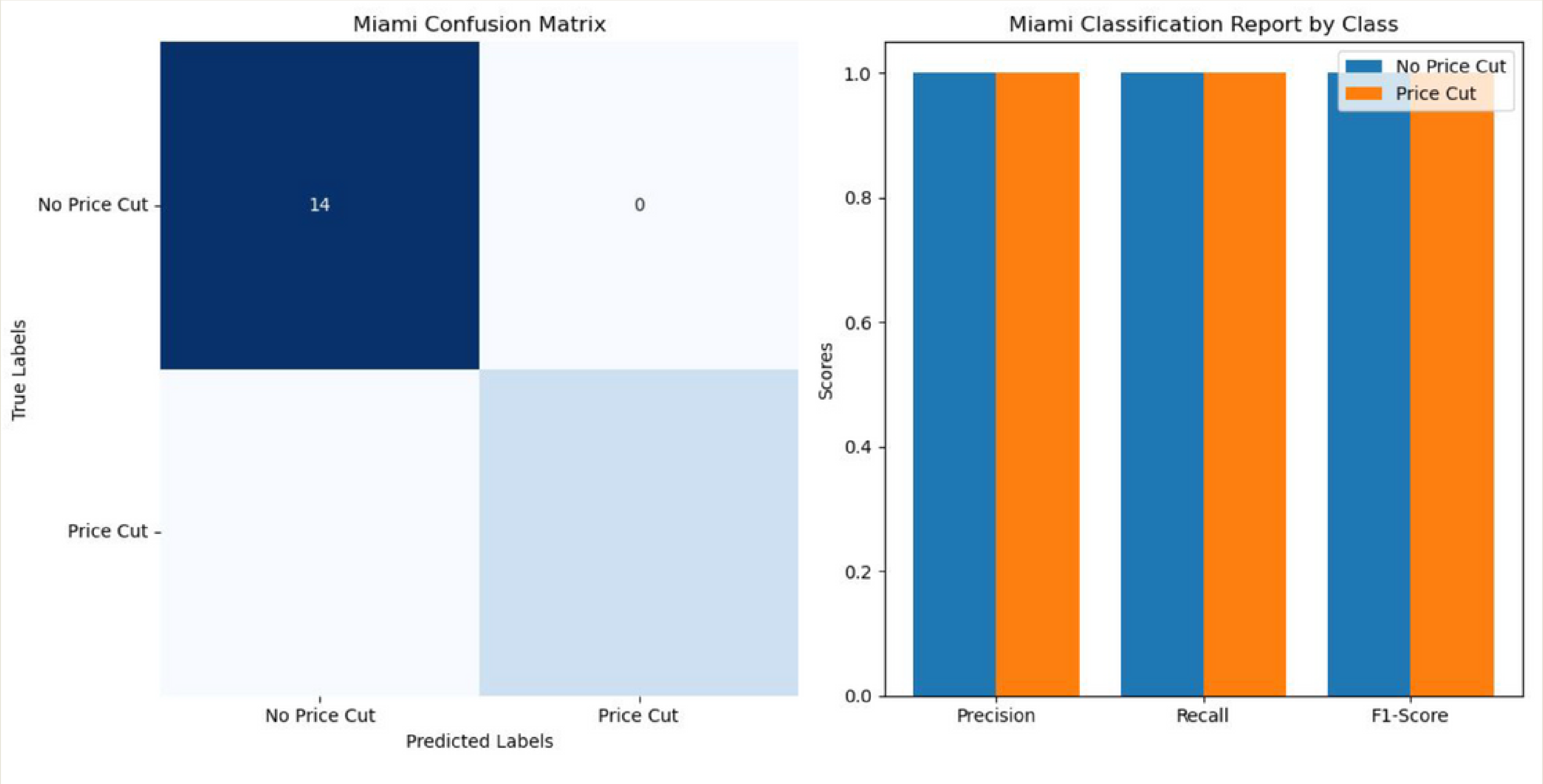


Los Angeles, CA

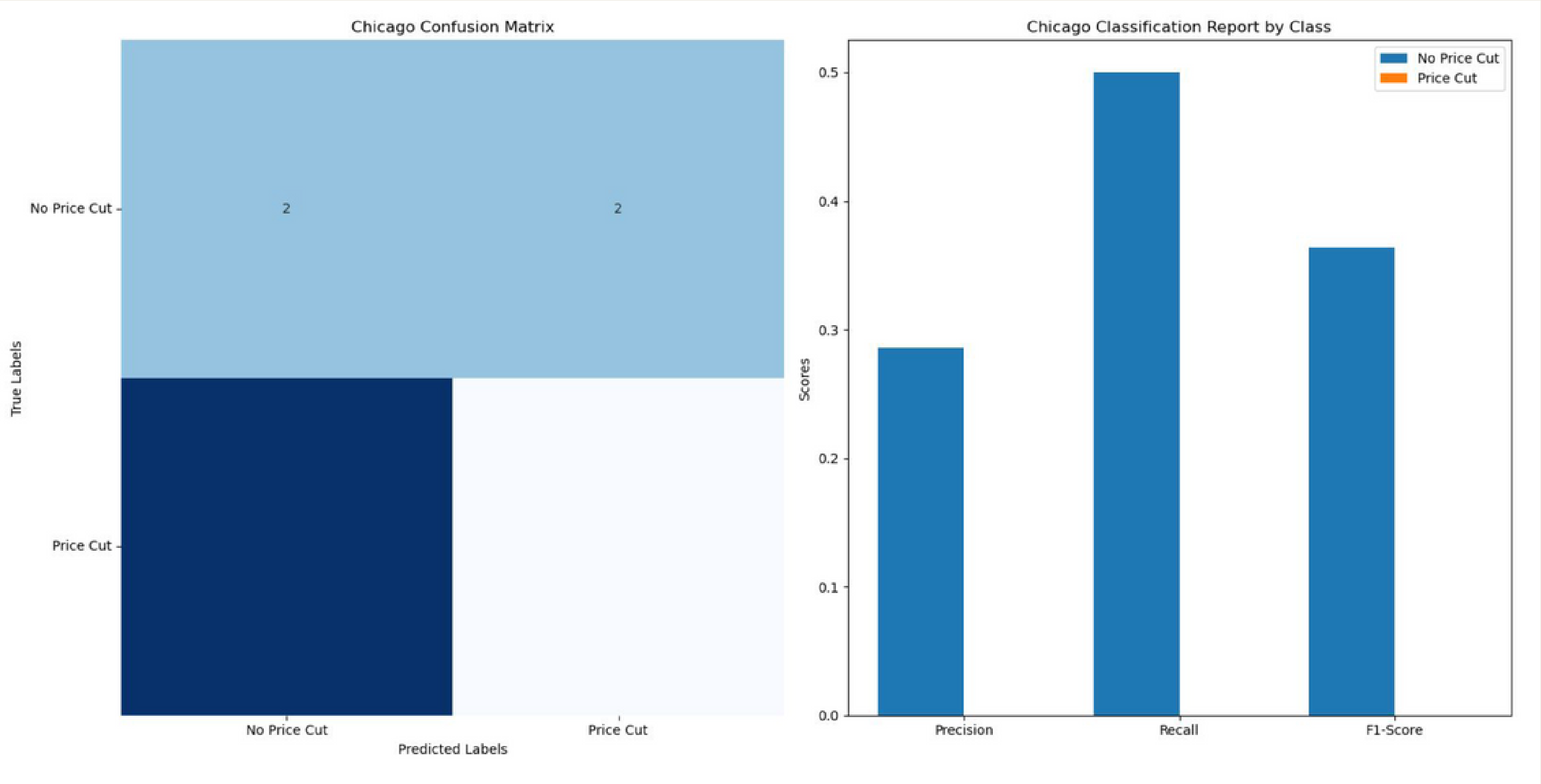


Supervised ML Model Continued

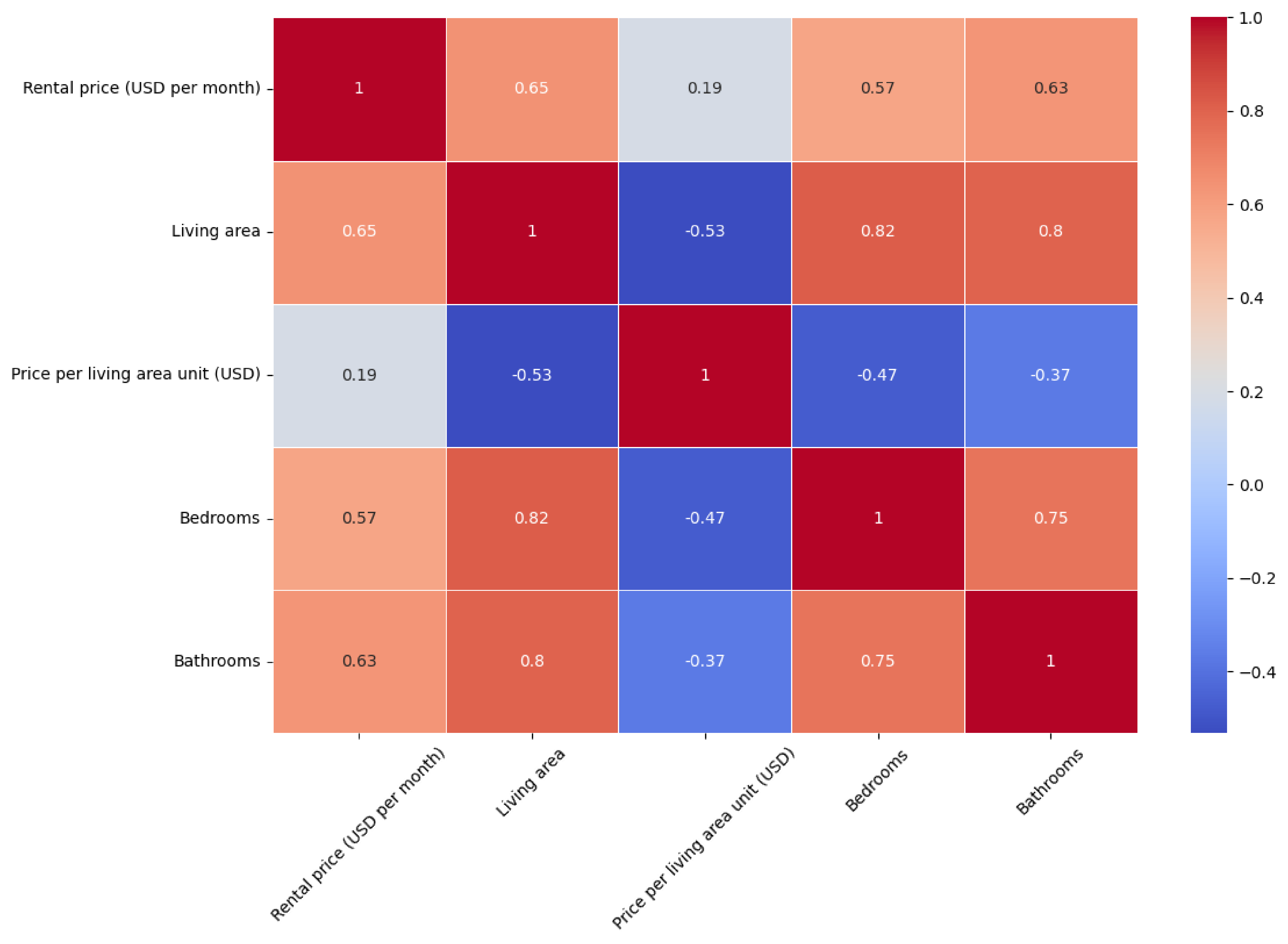
Miami, FL



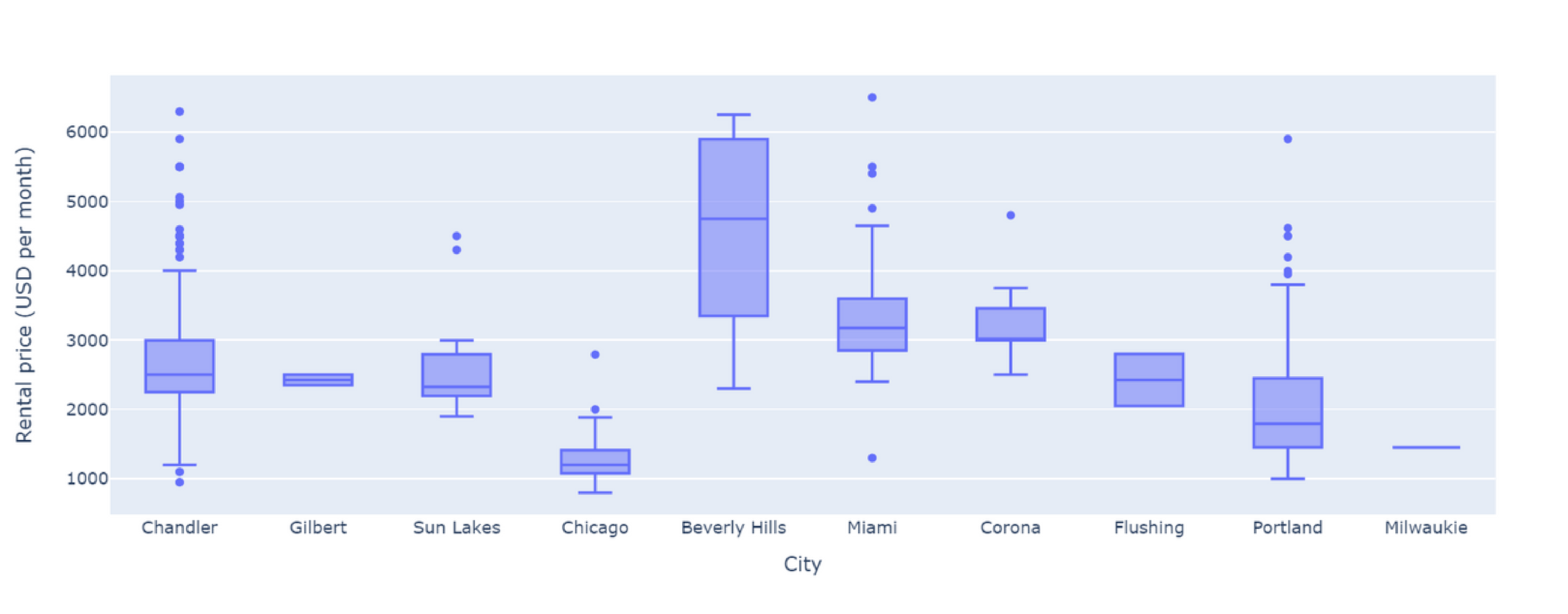
Chicago, IL

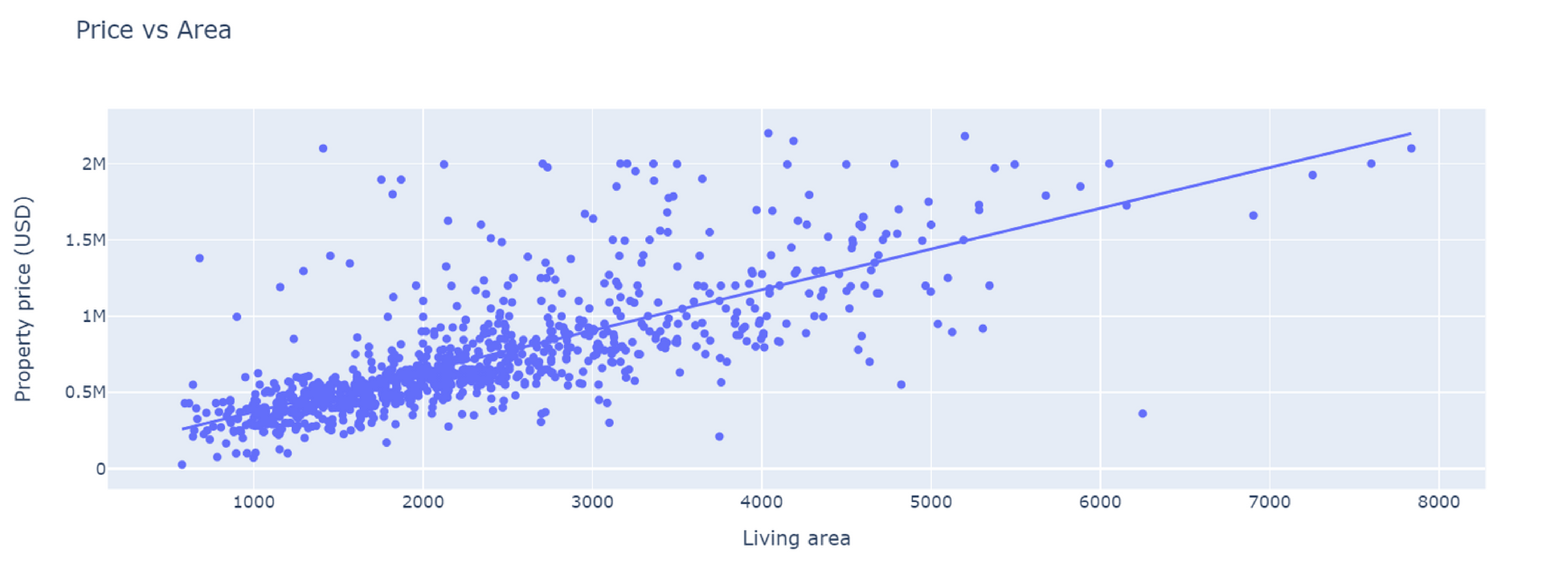


Correlation of Variables



	Rental price (USD per month)	Living area	Price per living area unit (USD)	Bedrooms	Bathrooms
Rental price (USD per month)	1	0.65	0.19	0.57	0.63
Living area	0.65	1	-0.53	0.82	0.8
Price per living area unit (USD)	0.19	-0.53	1	-0.47	-0.37
Bedrooms	0.57	0.82	-0.47	1	0.75
Bathrooms	0.63	0.8	-0.37	0.75	1


Rental price (USD per month)
6000
5000
4000
3000
2000
1000
Chandler
Gilbert
Sun Lakes
Chicago
Beverly Hills
Miami
Corona
Flushing
Portland
Milwaukie
City


Price vs Area
Property price (USD)
2M
1.5M
1M
0.5M
0
1000
2000
3000
4000
5000
6000
7000
8000
Living area

Limitations

- Parts of the United States do not disclose property sales to the public
- Need to integrate additional features such as school ratings or transportation scores
- Data set limited to **36 months**

Questions?