



FORECASTING COMMERCIAL REAL ESTATE VALUE

By Scott Fischer

The Challenge



FORECASTING A DIVERSE ASSET CLASS

Commercial Real Estate investment is a complex field with a variety of property types and approaches. These properties also adhere to regional market conditions, making proper valuation difficult.

There are also a variety of approaches that are taken in relation to investing in the property which in turn impacts the market price. This creates a very nuanced issue for buyers/sellers, who are often tasked with gathering all the relevant data prior to completing a transaction.

The barrier for entry is large for Commercial Real Estate. It's not uncommon for properties to be in contract for 6months - 2years+ with all the due diligence required.

Current Valuation Approach

While each transaction is unique, there are some relatively universal aspects considered when evaluating market value.

Recent Sales Comps

Emulating the market

What has sold recently like my property in the past 5 years?

Comps on Market

Competing with the Market

What's currently selling on the market similar to my property?

Financials

Investment Values

How much does the property net annually?
What are operating expenses?

Regional Analysis

Important local information

What are the area rental rates? Average rent?

Problems with Current Approach

Not "Data Centric"

Decisions on value are made re-actively on past sales comps (from 6 months ago) or from a speculative valuation.

Intangible Impacts

It's difficult to quantify exactly how certain area developments will impact property value.

Opportunity Cost

Aggregating the required information to determine sales price or decide on a purchase takes a lot of time.



DEVELOPING A SOLUTION

Creating one space for all these diverse consideration to be analyzed quickly in a data centric manner.

Indexing Value

Create a general valuation metric to judge if buying or selling in a relative high or low.

Estimating Value

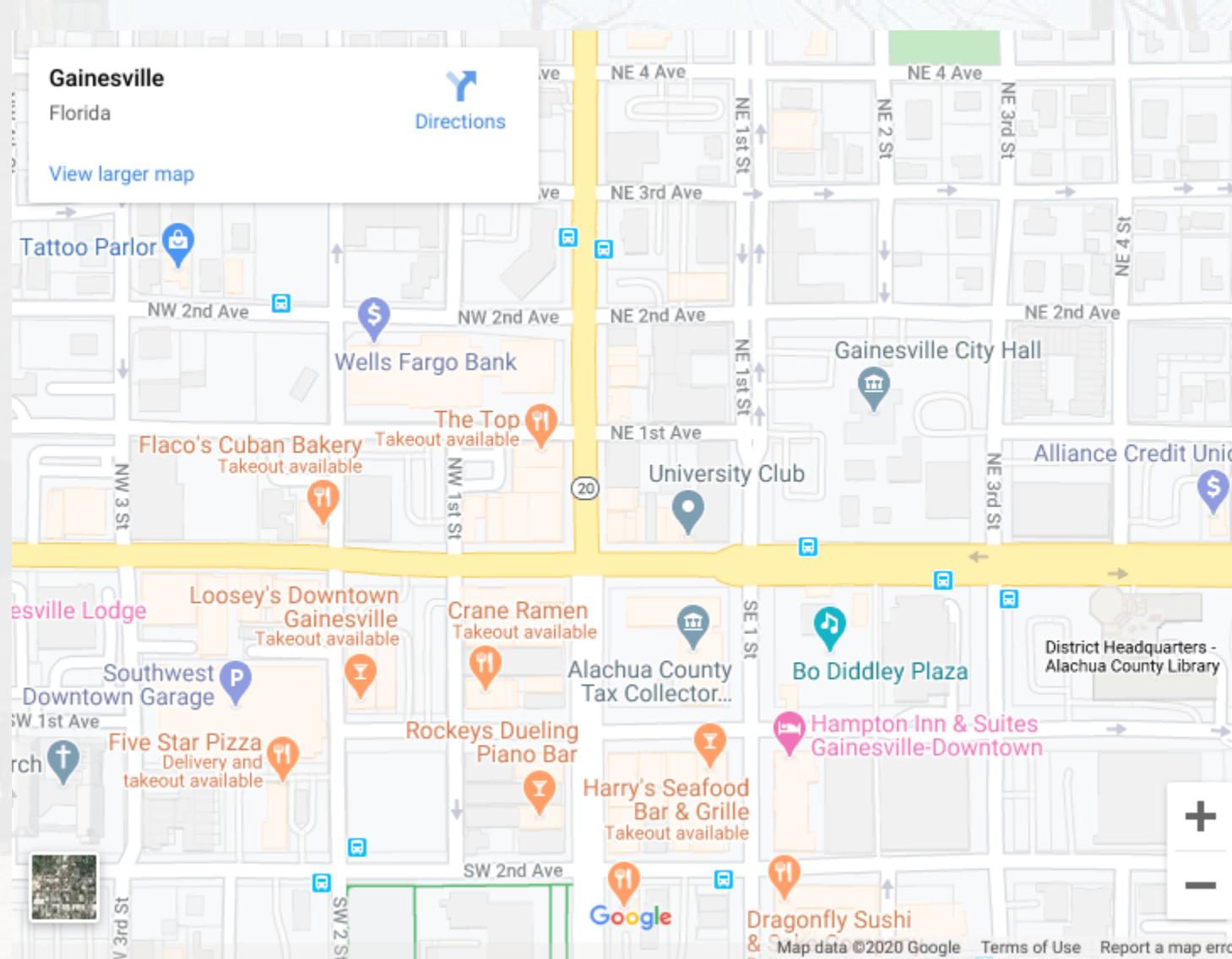
Supervised learning regression that uses index and general property information to estimate value.

Visualizing Value

interactive map based visualizations showcasing area trends. Unsupervised learning & aggregated property features.

The Data

Overview



Sources

Alachua County Property Appraiser

General Sales information that's publicly available. Represents all transactions recorded.

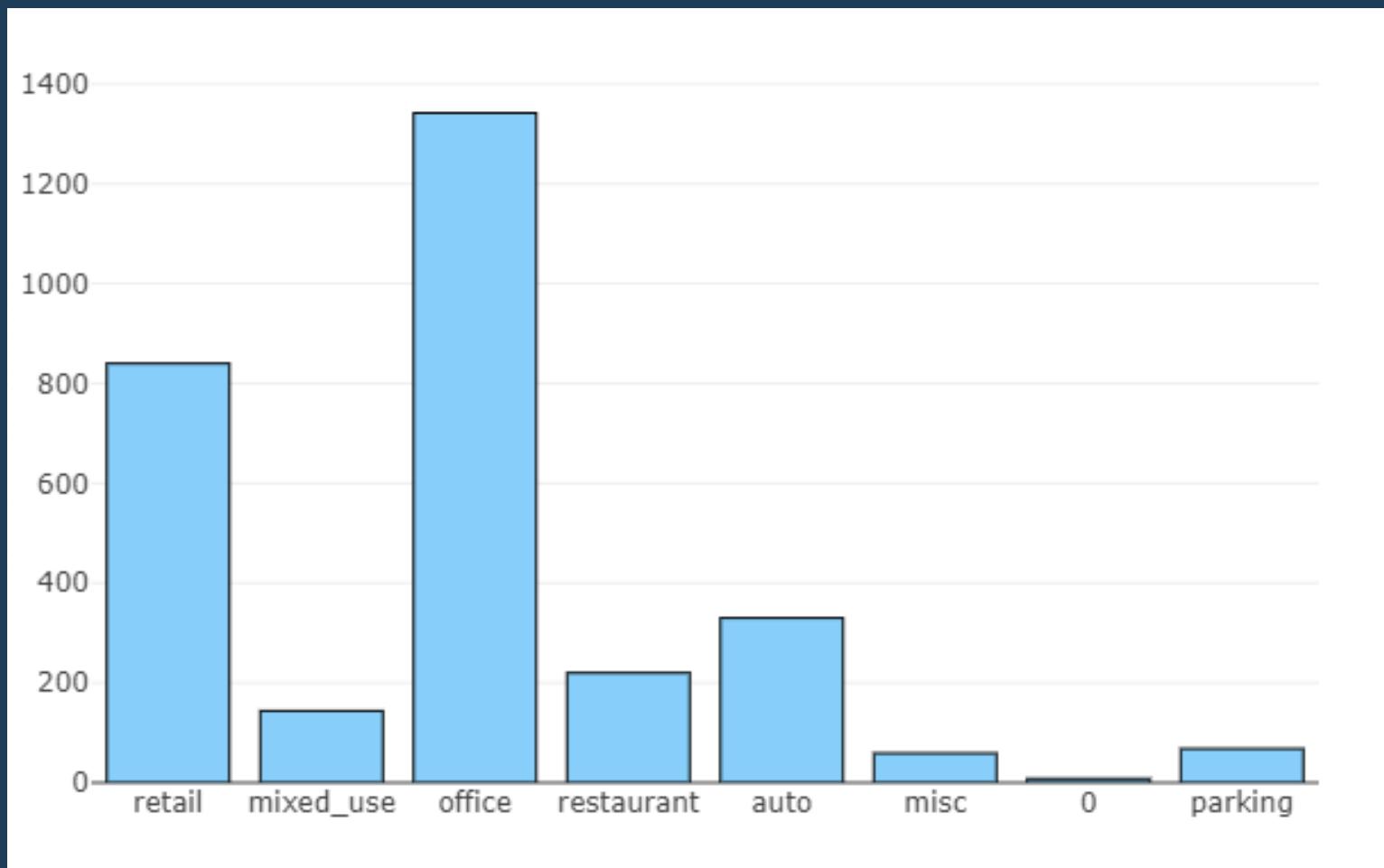
Department of Labor and Statistics

For a national rental index over the past 40 years.

Fed Interest Rate

Federal interest rates impact lending which in turn should affect pricing.

LIMITATIONS OF AVAILABLE DATA



SPARSITY IN PROPERTY TYPES

Office and Retail types dominate the market place

LIMITED MARKET TRANSACTIONS

After cleaning for errors, only about 3,000 transactions have occurred over a 40yr span

COVID 19 LIMITED ACCESS TO DATA

Certain government offices were closed, making some data sets impossible to acquire

PUBLIC DATA ISN'T DETAILED

Sale Date / Price / Vac / Current property Data

OUTLIER TRANSACTION TYPES

- Assignment Sales
- Assemblage Sales
- Portfolio Sales
- Transfer of ownership
- Tax Evasion
- Divorce / Dissolution

BUILDING THE INDEX

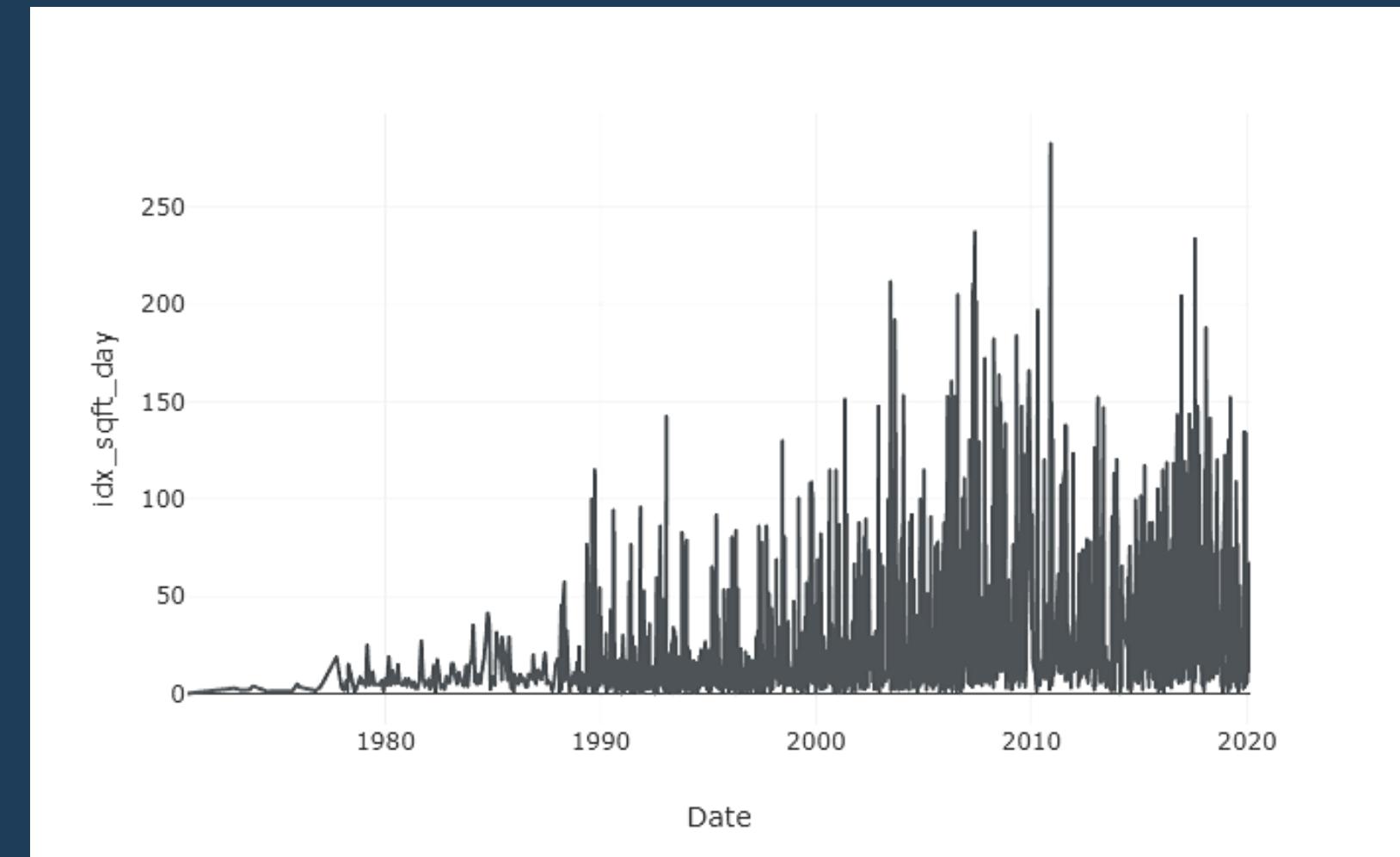


PRICE PER SQFT

Averaging by price per square foot per day provides a method to scale all historic property sales.

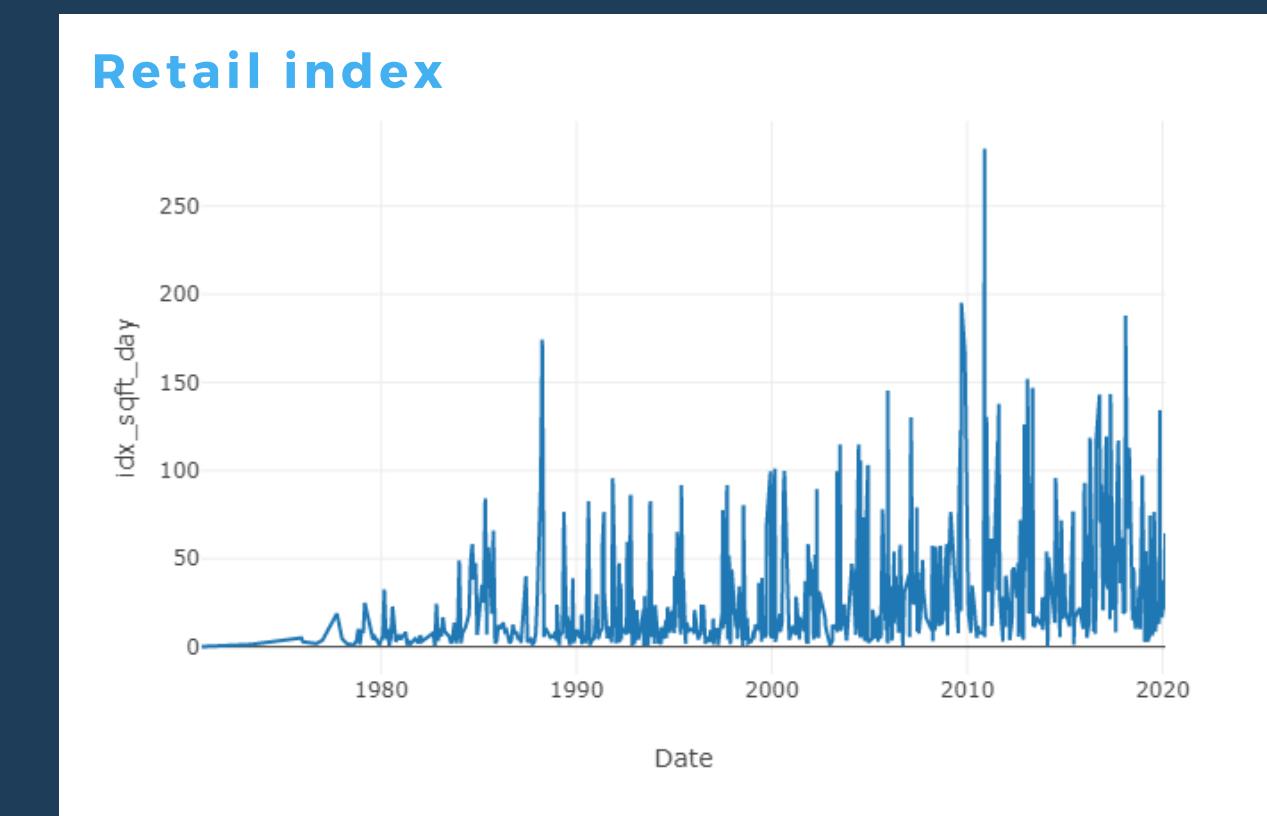
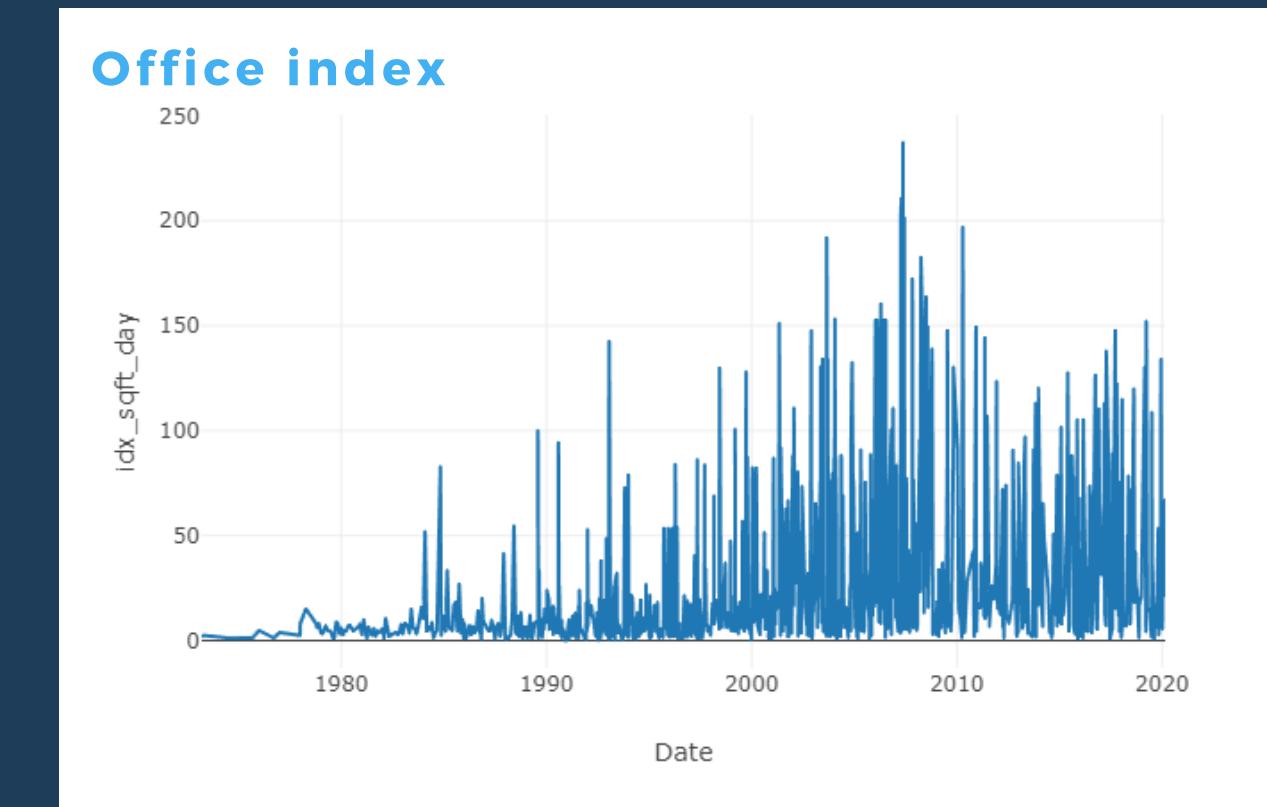
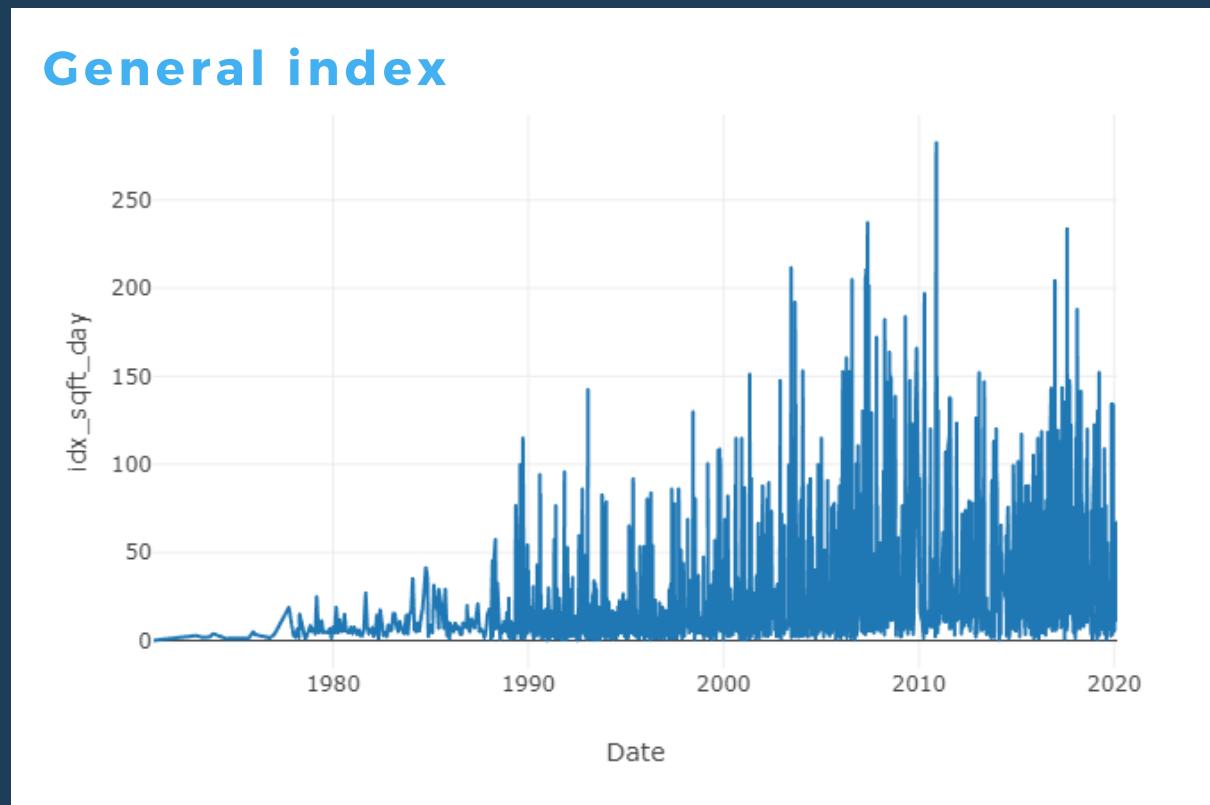
Limitations

- Data is sparse, sometimes only one sale will occur on a given day
- Asset classes are incredibly different in terms of price per square-foot; Sub-indexes are superior
- Annual aggregations are likely more valuable, but makes it difficult to make day by day decisions.



BUILDING THE INDEX

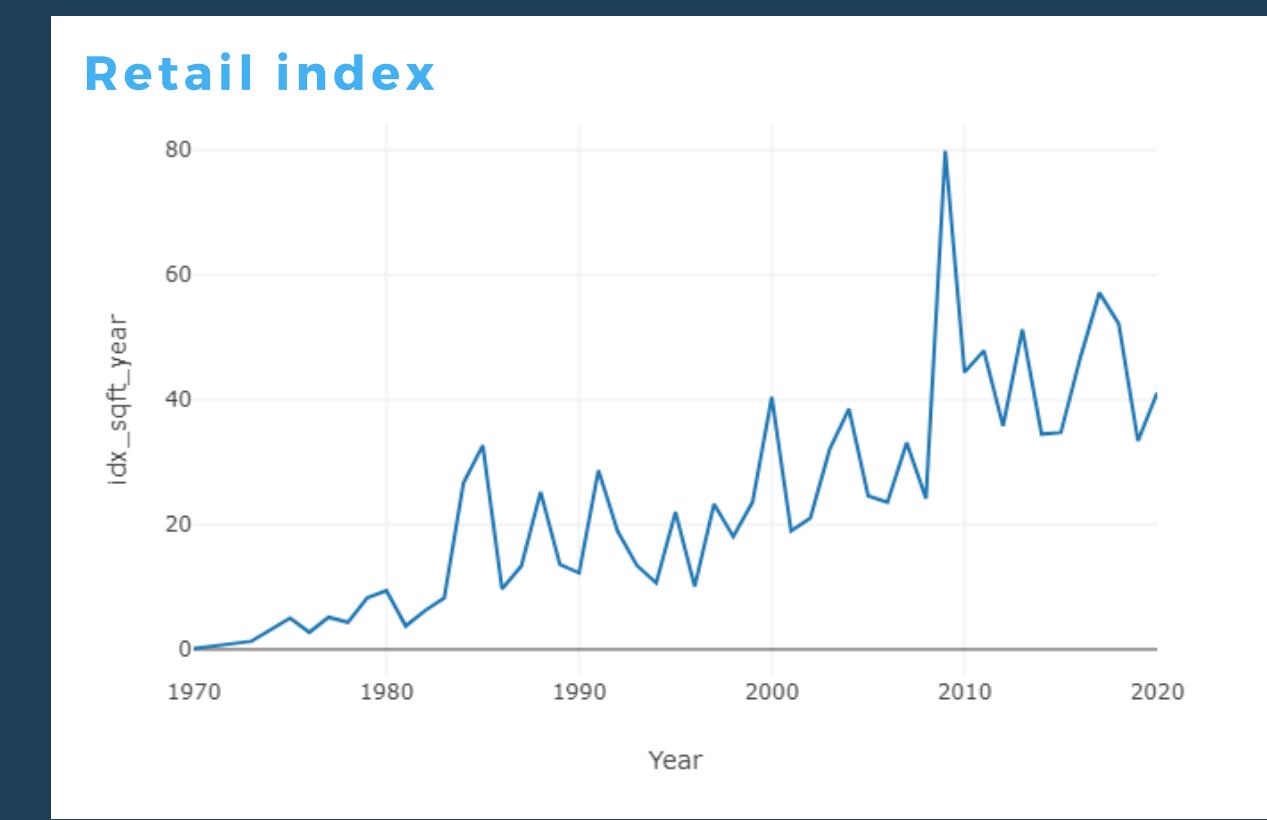
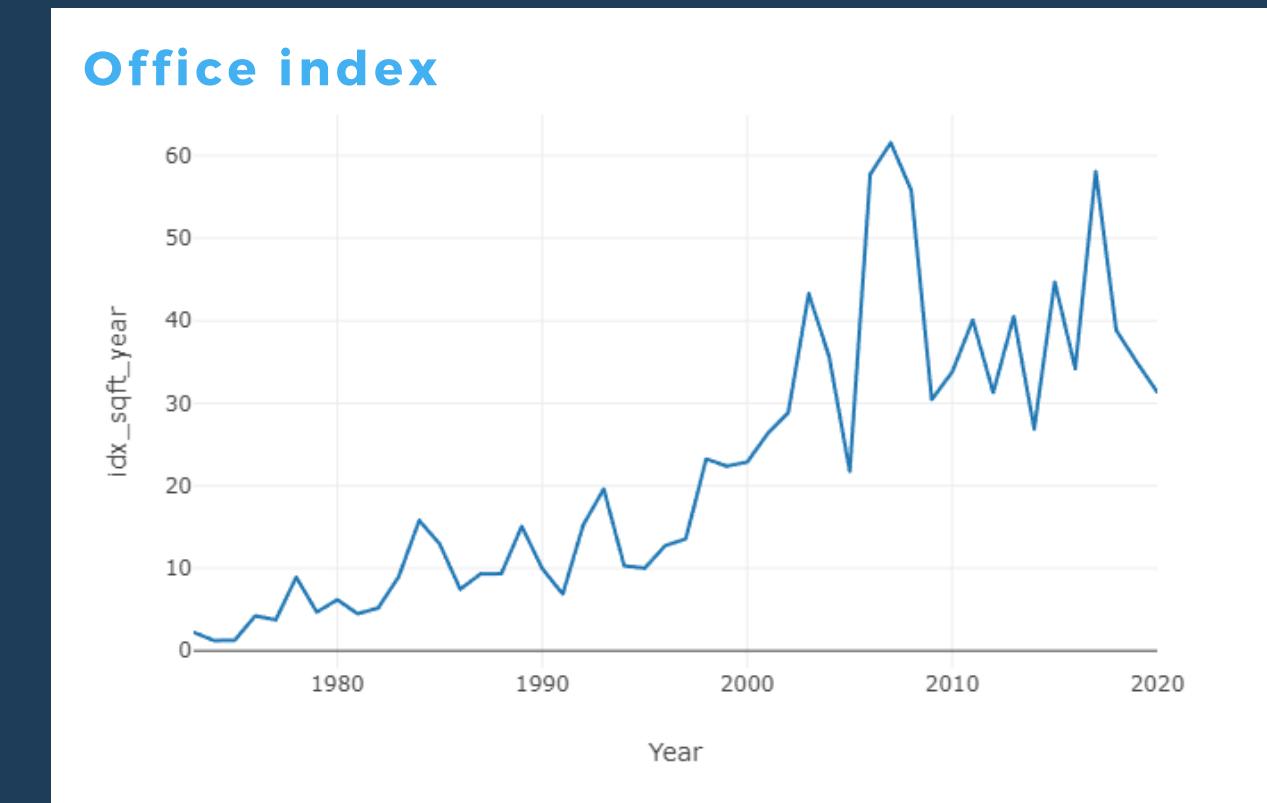
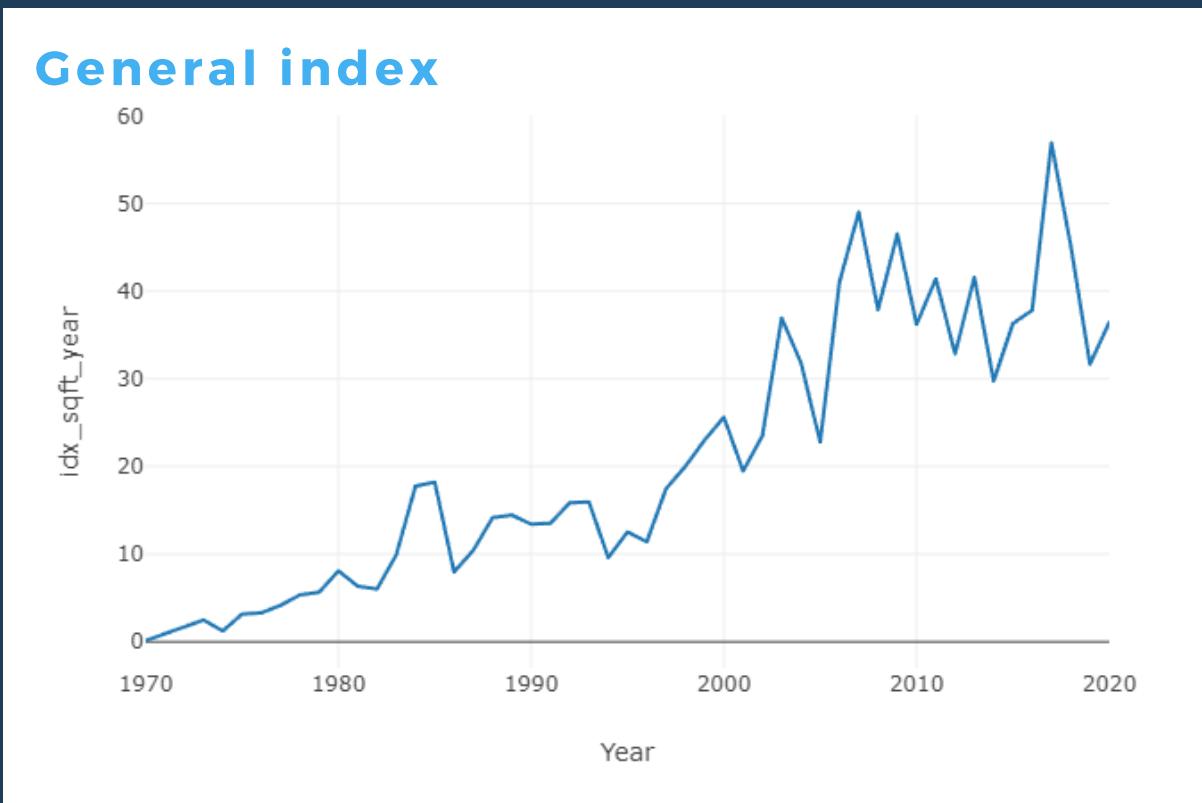
DAILY VALUES



BUILDING THE INDEX

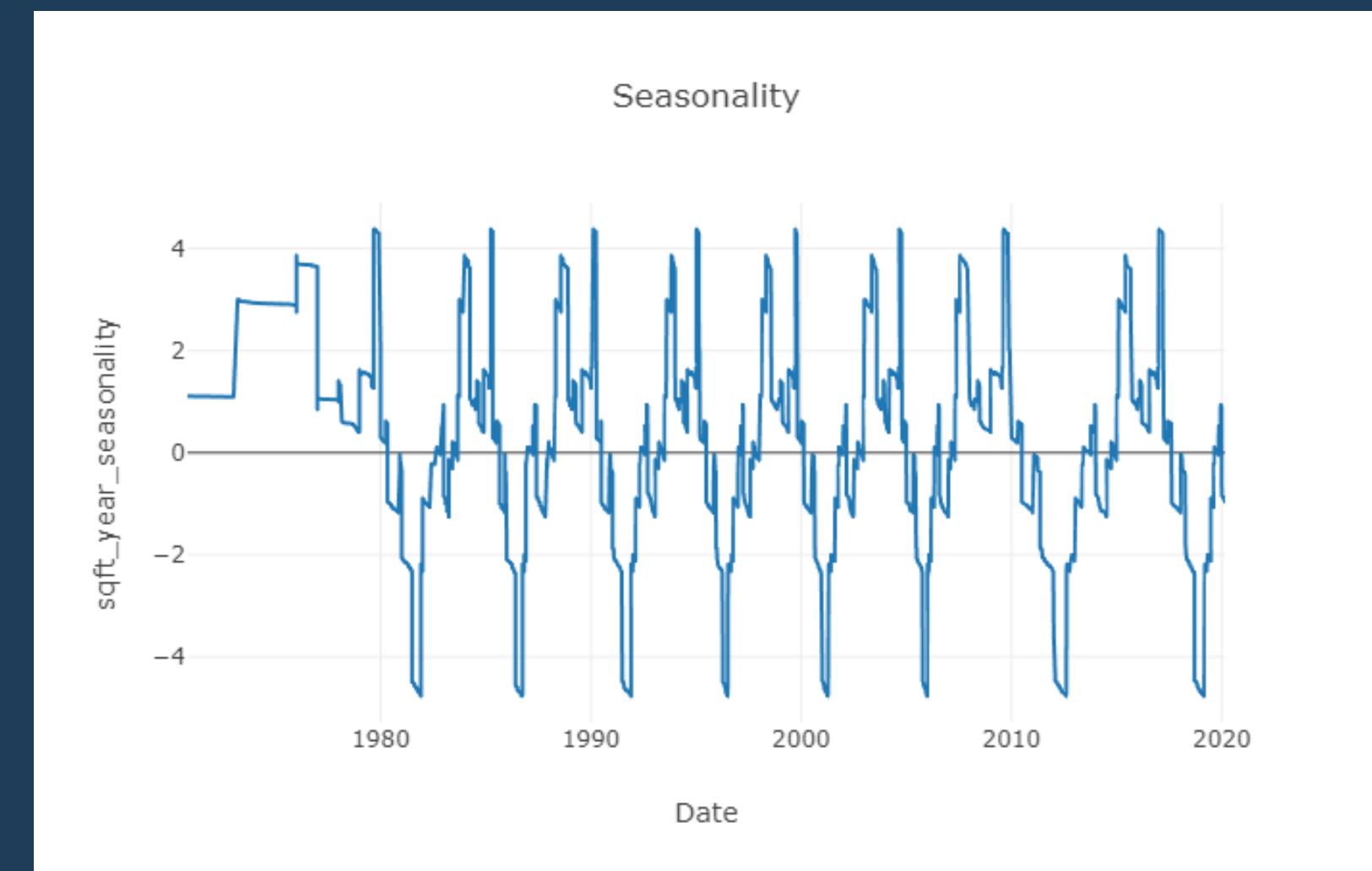
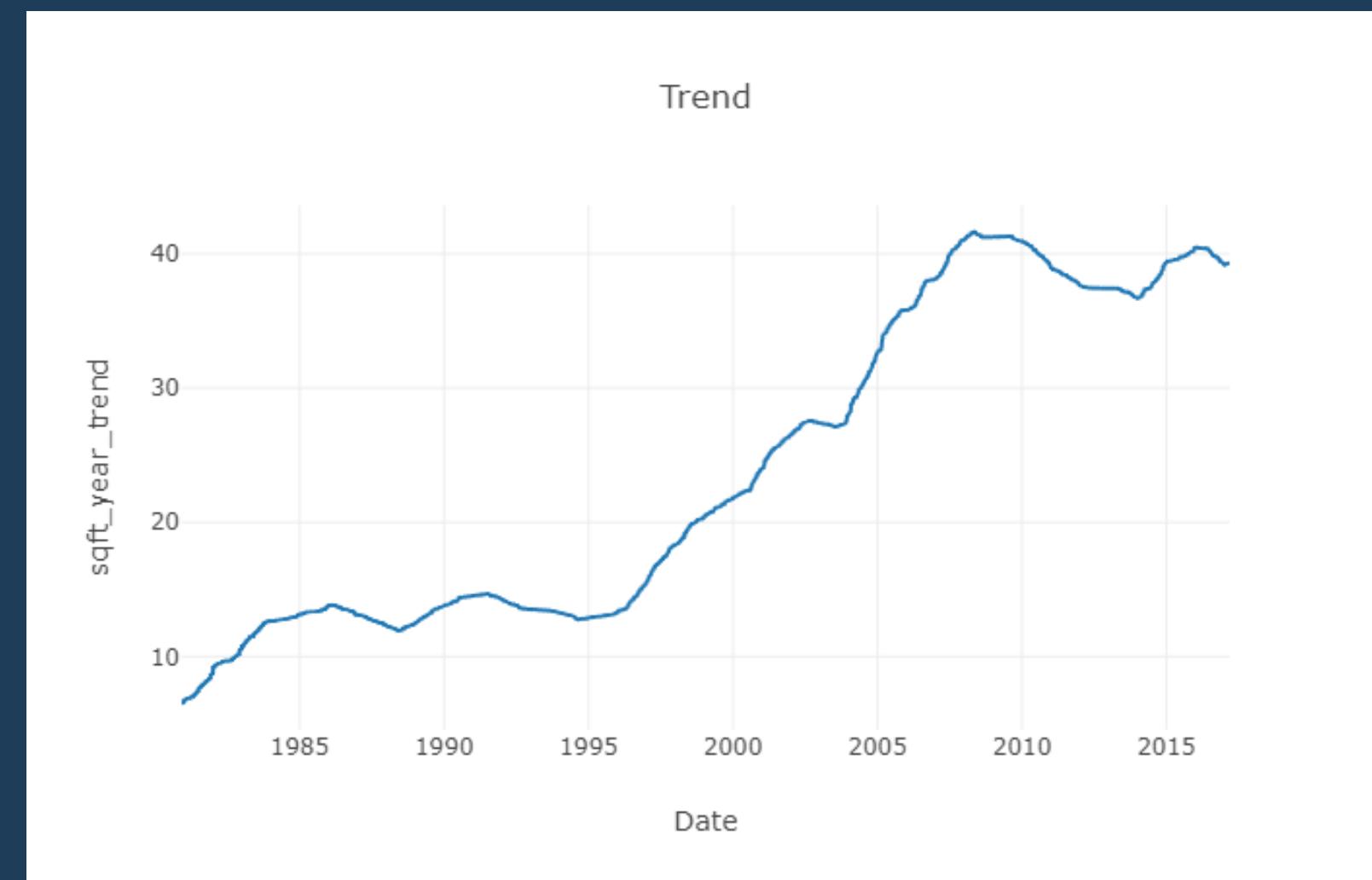


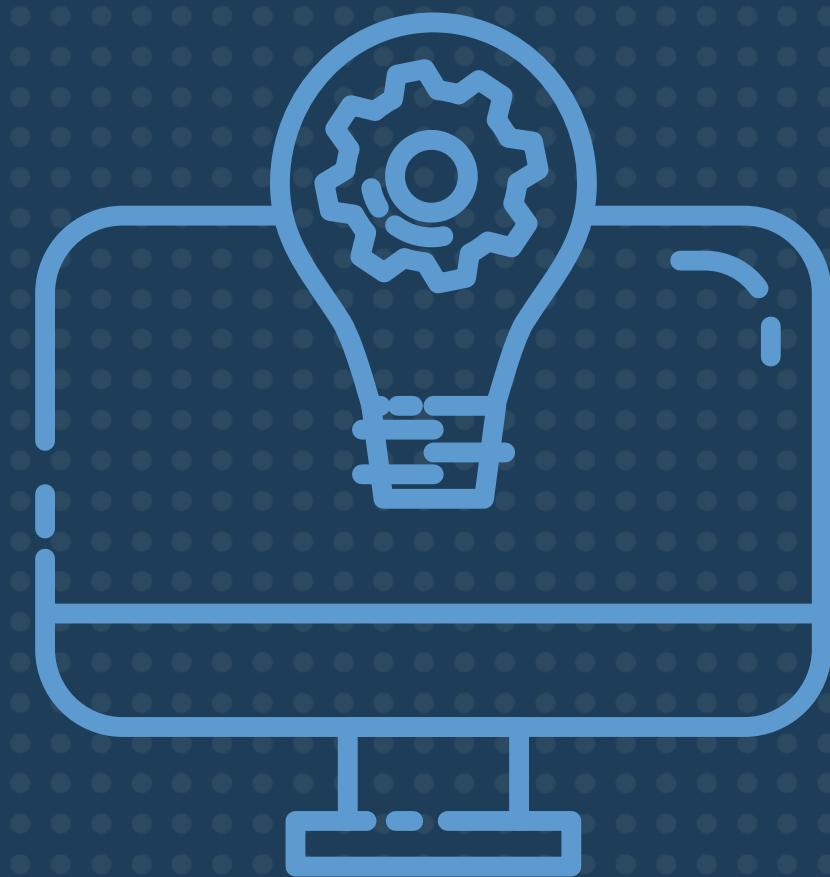
ANNUAL VALUES



BUILDING THE INDEX

GENERAL TREND / SEASONALITY





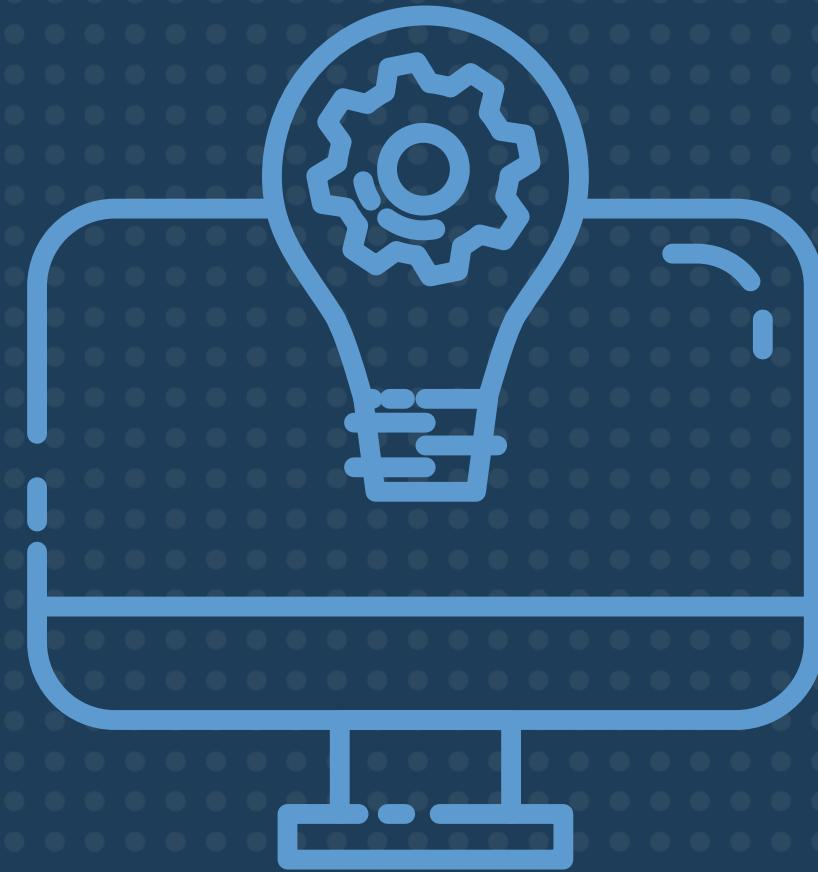
The Estimator

Supervised Learning Regression

Data Considerations:

- Huge outliers have been removed
- When general index is included in the calculation, data is filtered to avoid days with only a single transaction

The Estimator



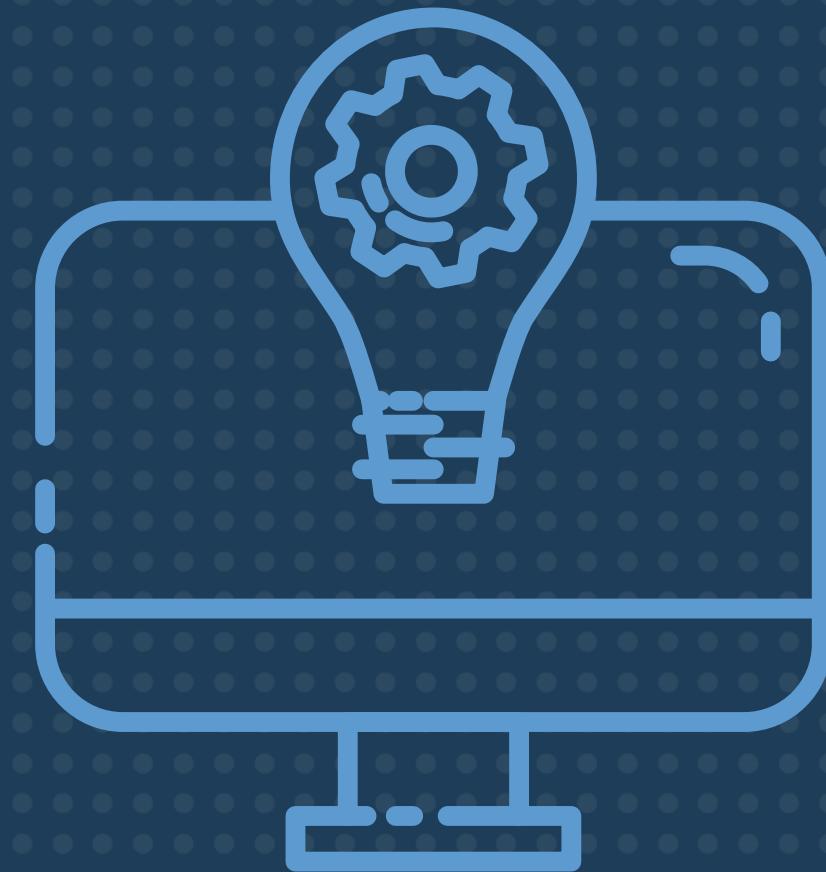
General Performance

The best performing regression was ensemble Random Forest.

In general, the results were not accurate, likely due to a few factors:

1. The difference in asset classes has too big of an impact on value
2. Data limitations had a heavy impact
3. Relative imbalance between property types

The Estimator



Tests & Scores

Regression 1

- Excluding daily index values

R2: .72

MAPE : 66%

Regression 2

- Includes daily index values
- Filtered to only include days where more than 2 transactions occurred.

R2: .53

MAPE : 45%

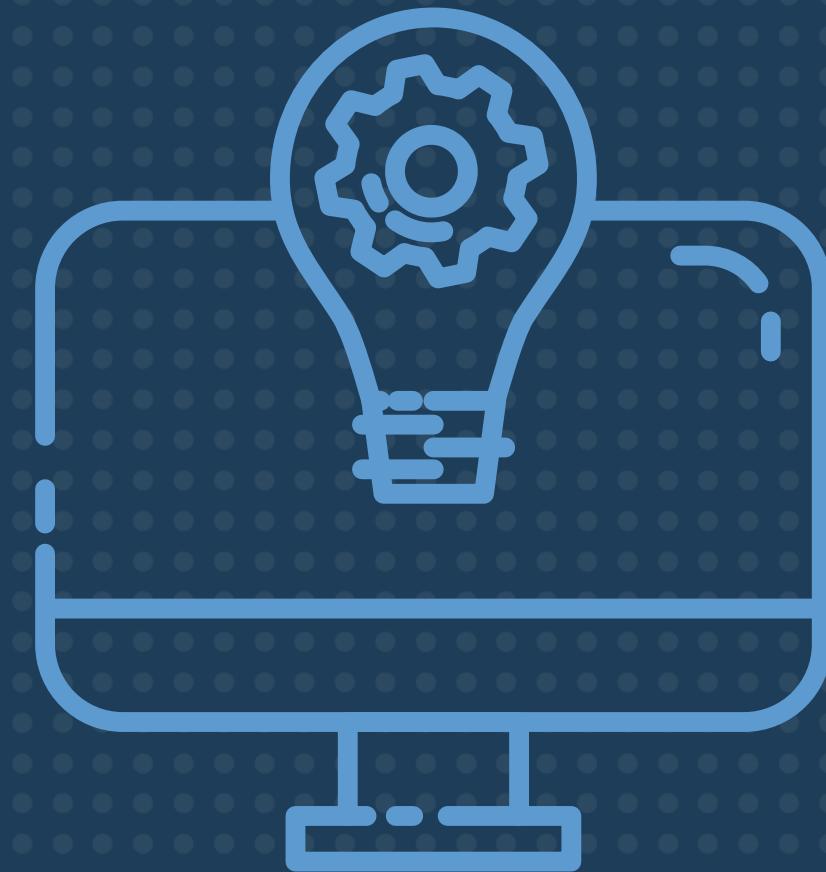
Regression 3

- Includes daily index values
- Filtered to only include days where more than 2 transactions occurred.
- Filtered to only include office property types.

R2: .73

MAPE : 34%

The Estimator



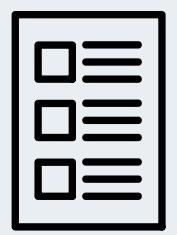
Final Thoughts

The index helped more than it hurt. It would have been very beneficial to have more transactions per day to see the proper impact.

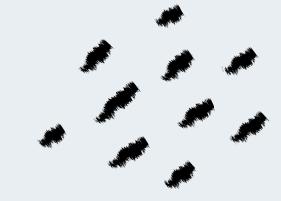
Estimating for specific property types (in this case office) yielded much better results. In the case of Regression 3, the general property index was surprisingly the most valuable feature.

Visualizing Value

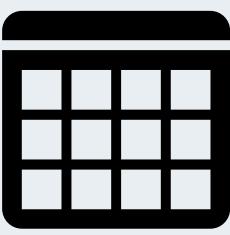
Map Based Visualizations



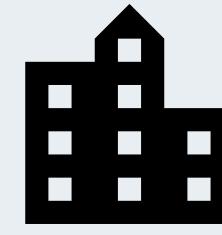
Frequency of sales



Unsupervised Clustering



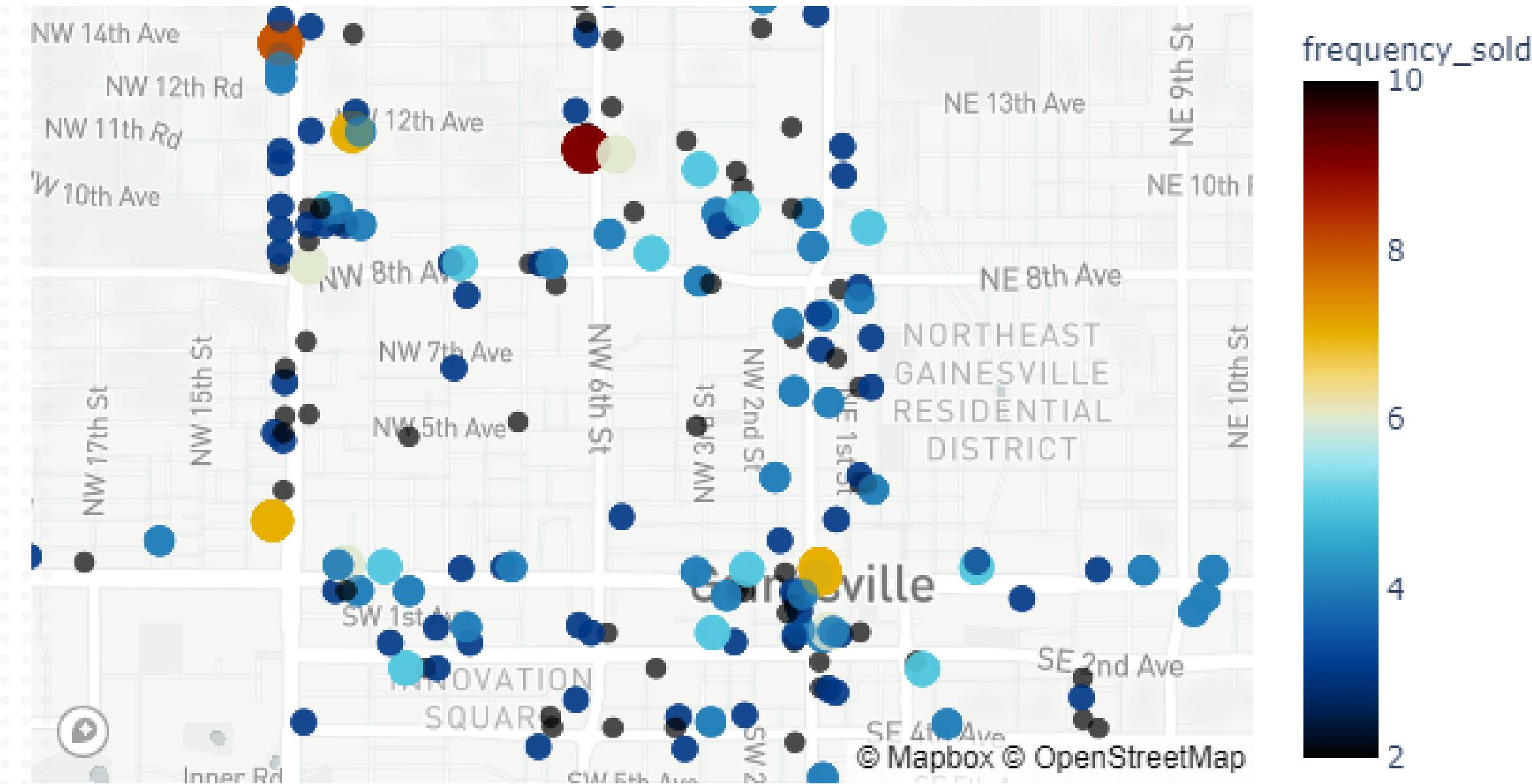
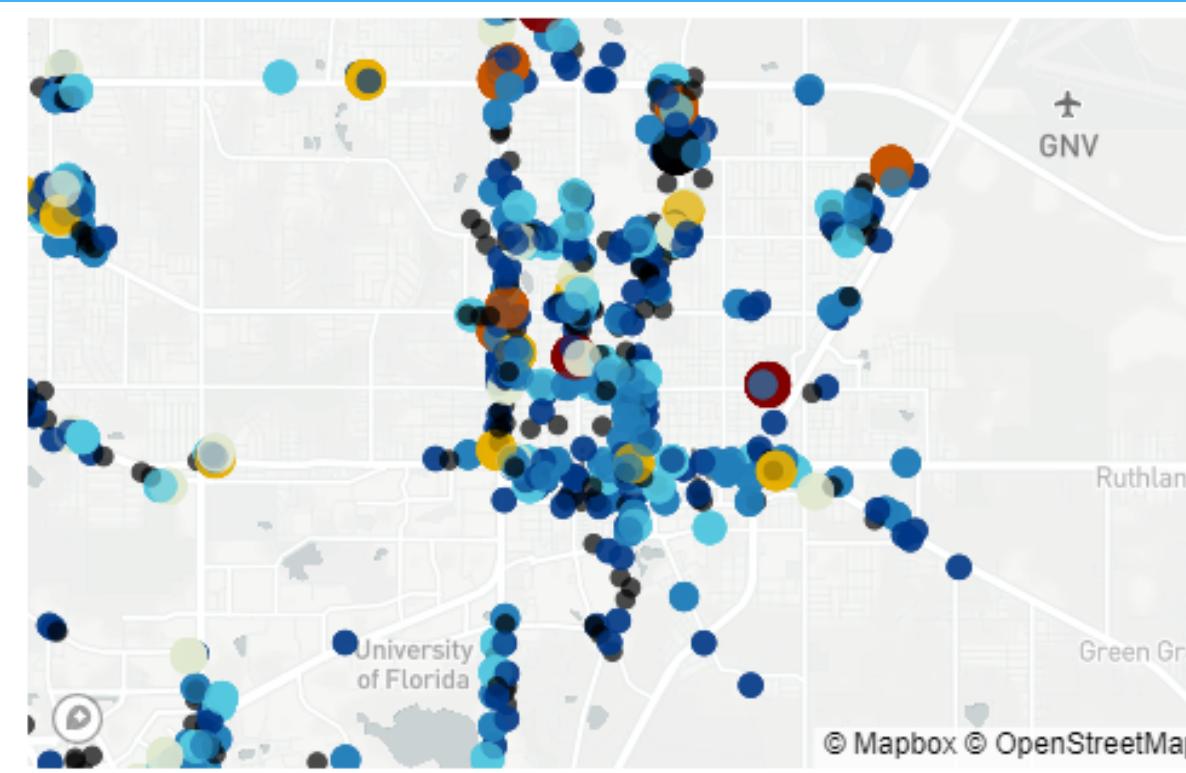
Days Between Sales



Property Types

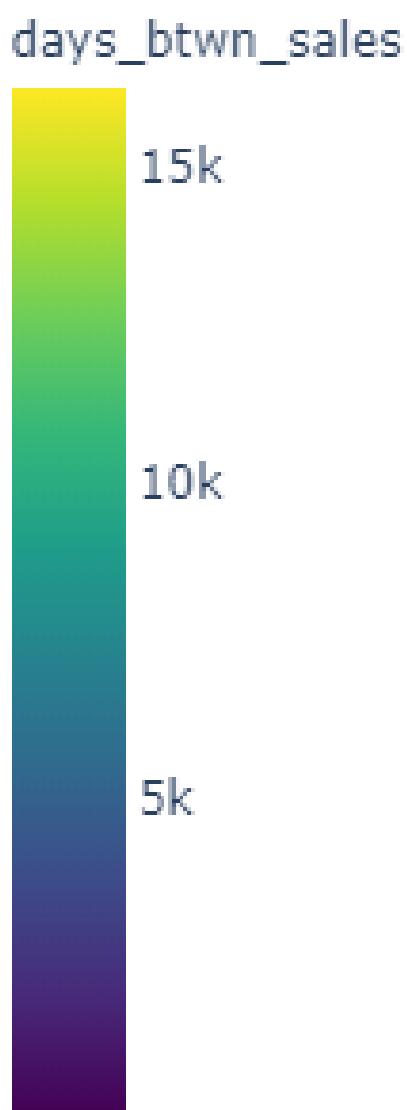
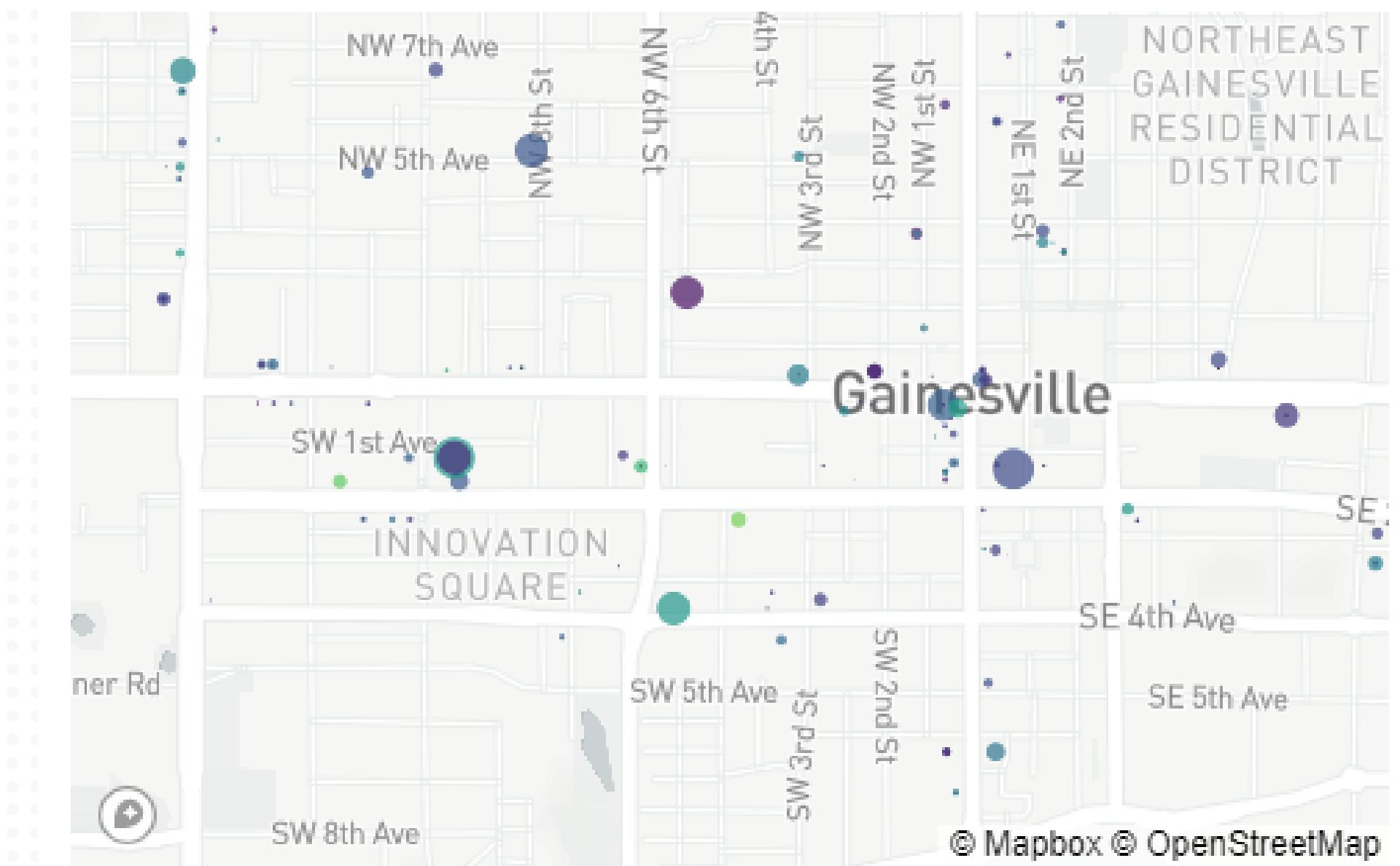
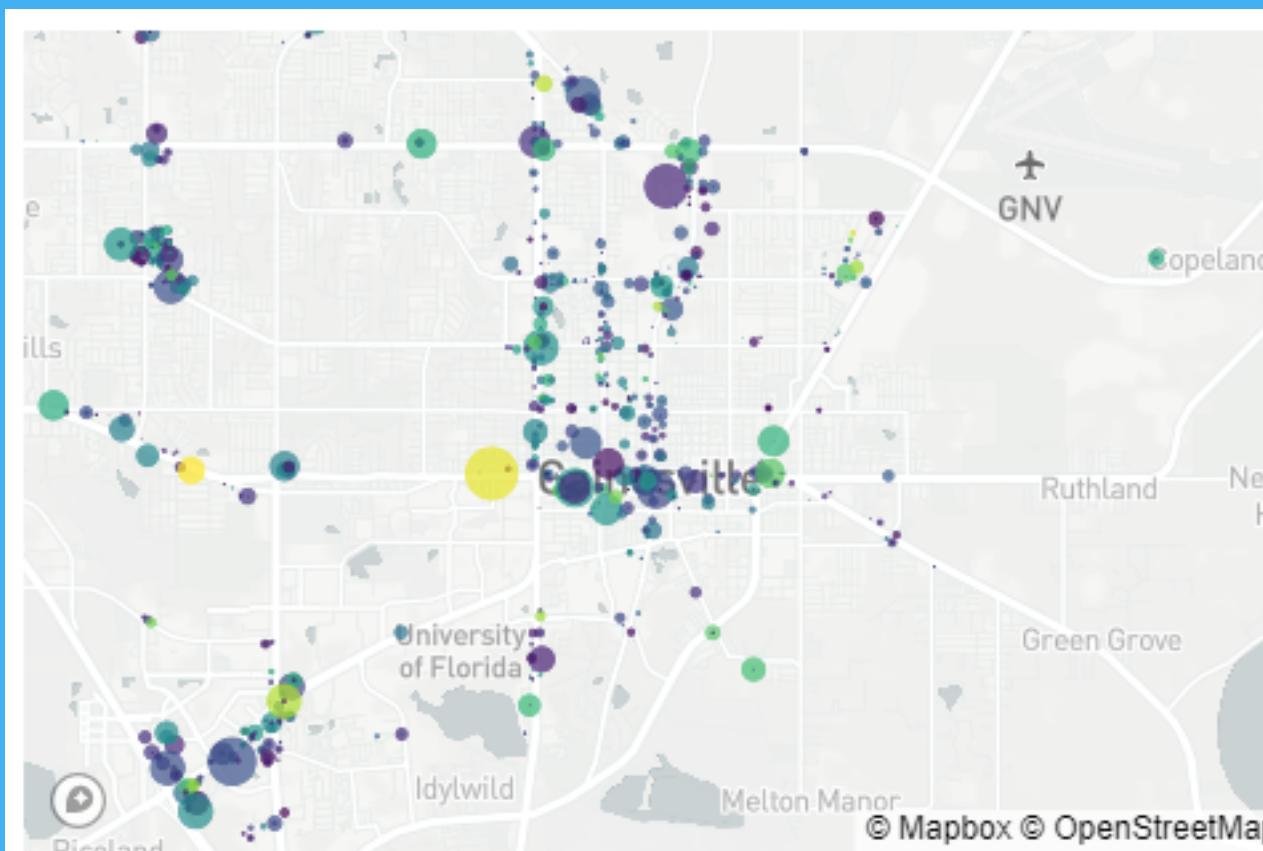
Frequency Sold

Color & Size based on frequency



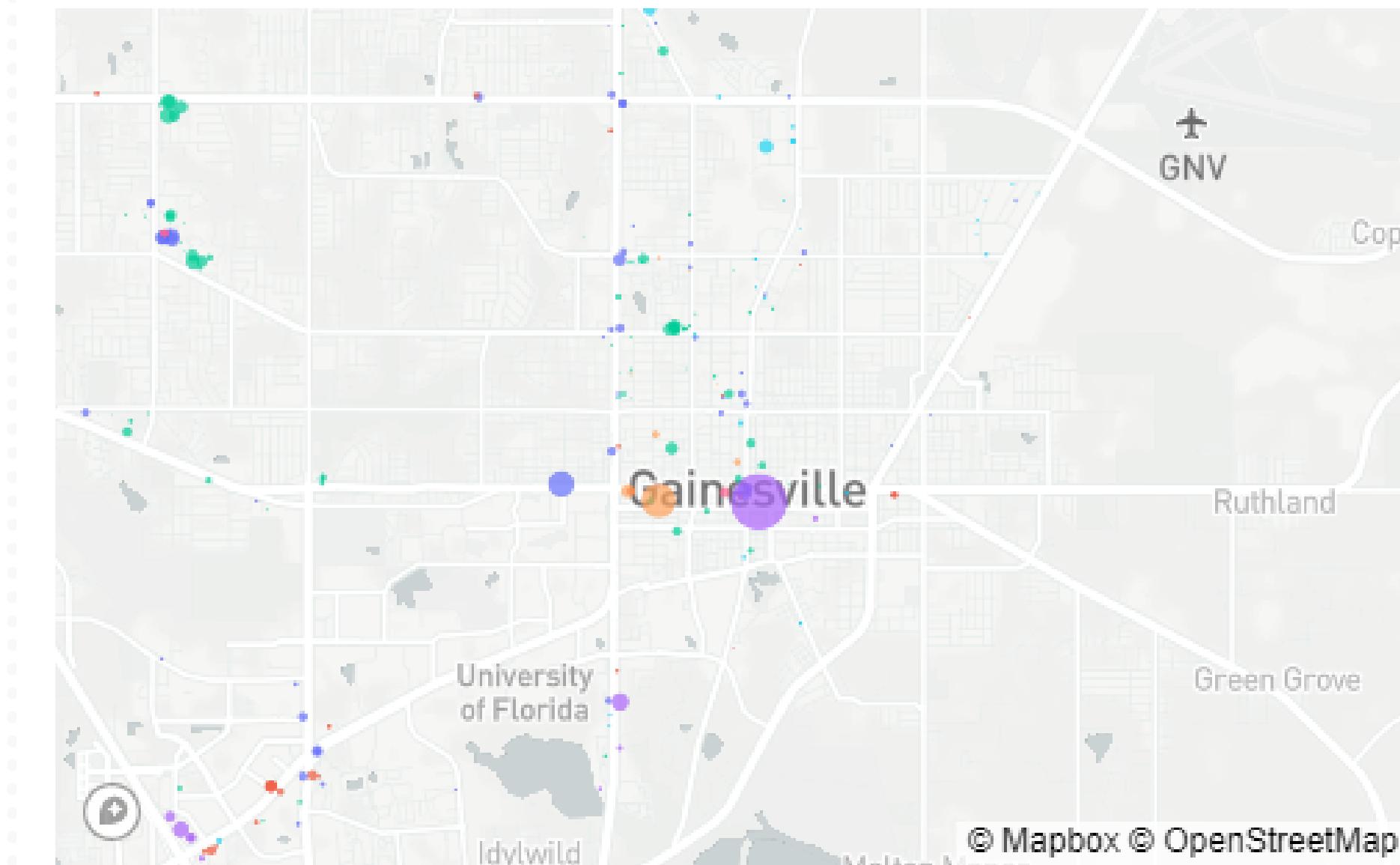
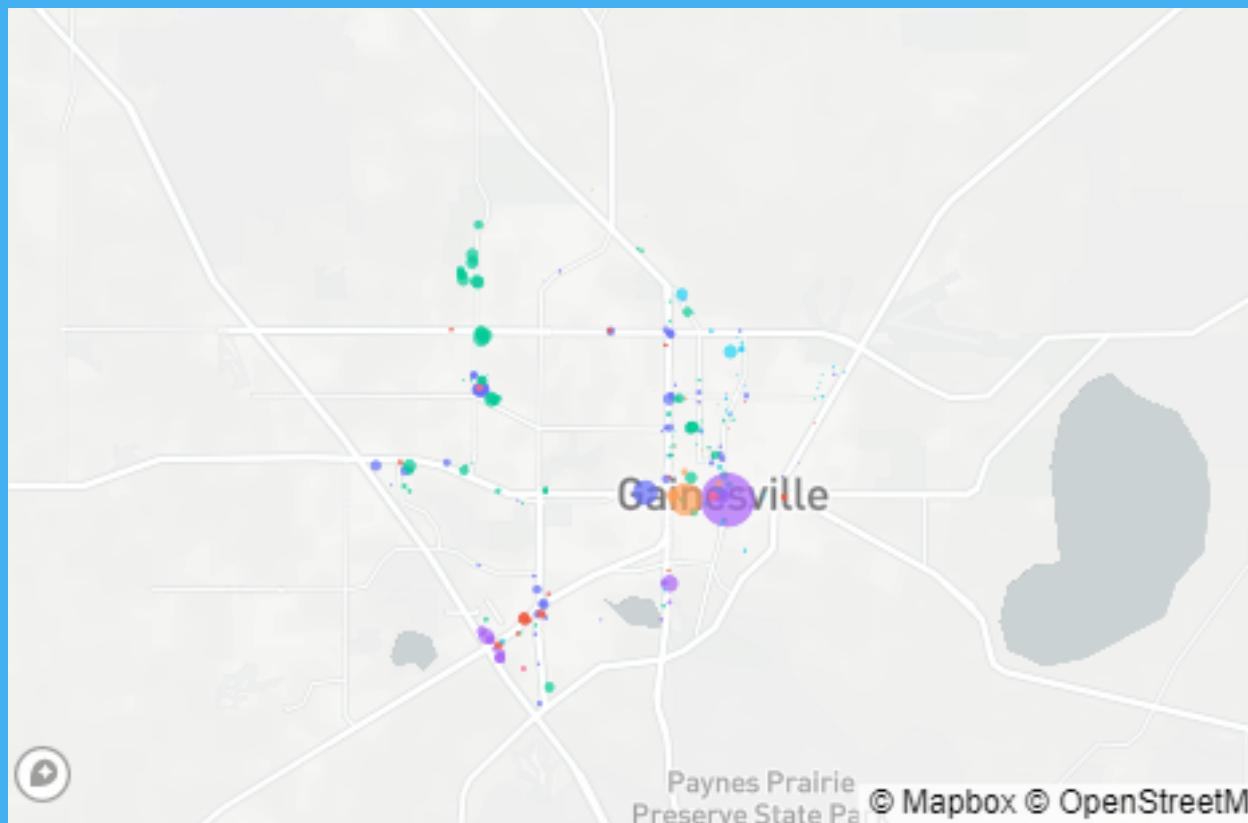
Days Between Sales

Color = Days
Size = % return



Past 5 Years

Color = Property Category
Size = Price per sqft



| Property_Cat |
|--------------|
| retail |
| restaurant |
| office |
| hotel |
| mixed_use |
| auto |
| misc |
| parking |

UNSUPERVISED CLUSTERING

HIERARCHICAL

Agglomerative clustering using gower distances yielded the most interesting clusters.

TYPE 0

1738 properties

Vast majority of properties on market fall into this bucket.

TYPE 1

291 properties

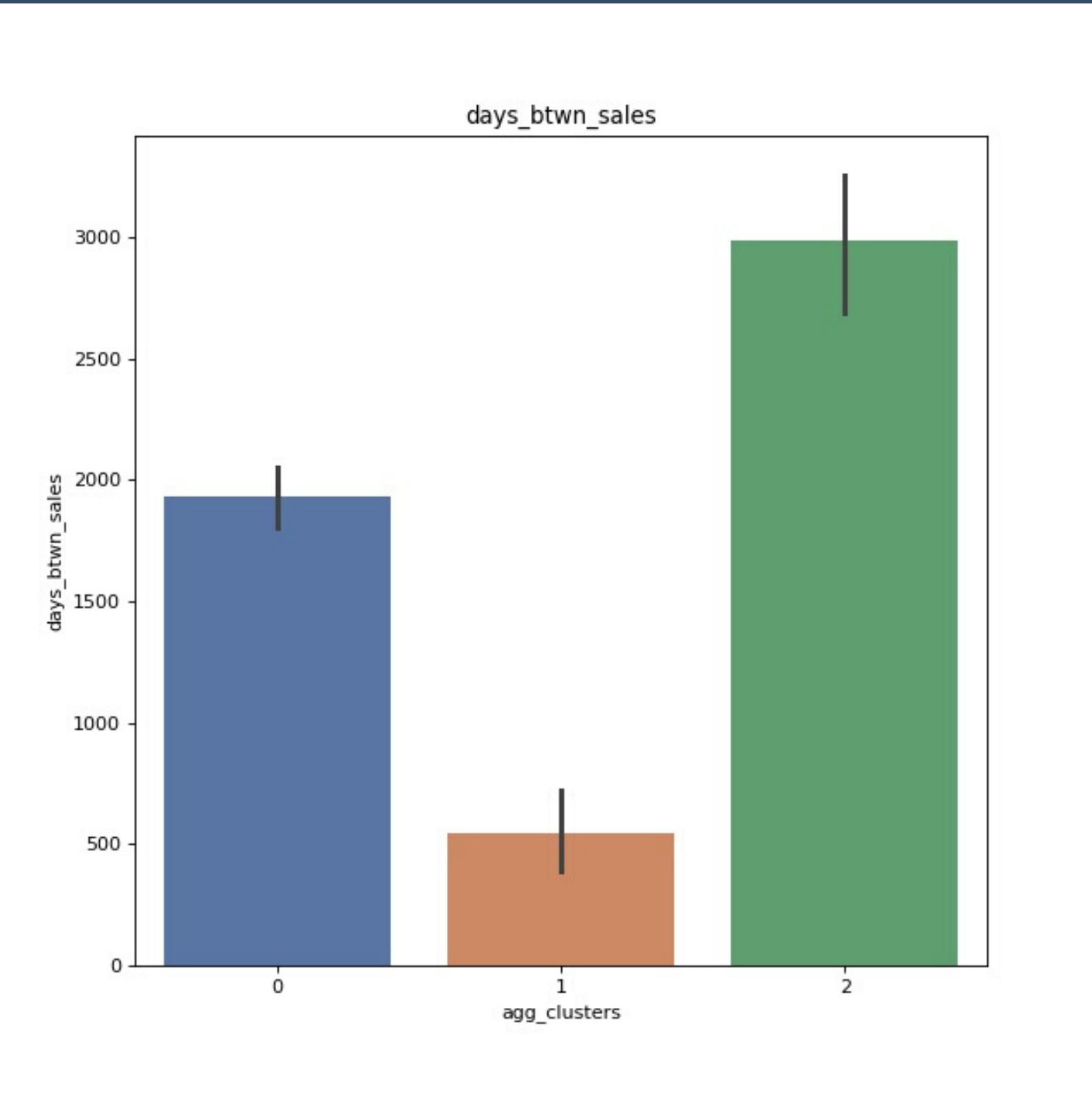
Vacant Properties and lots.

TYPE 2

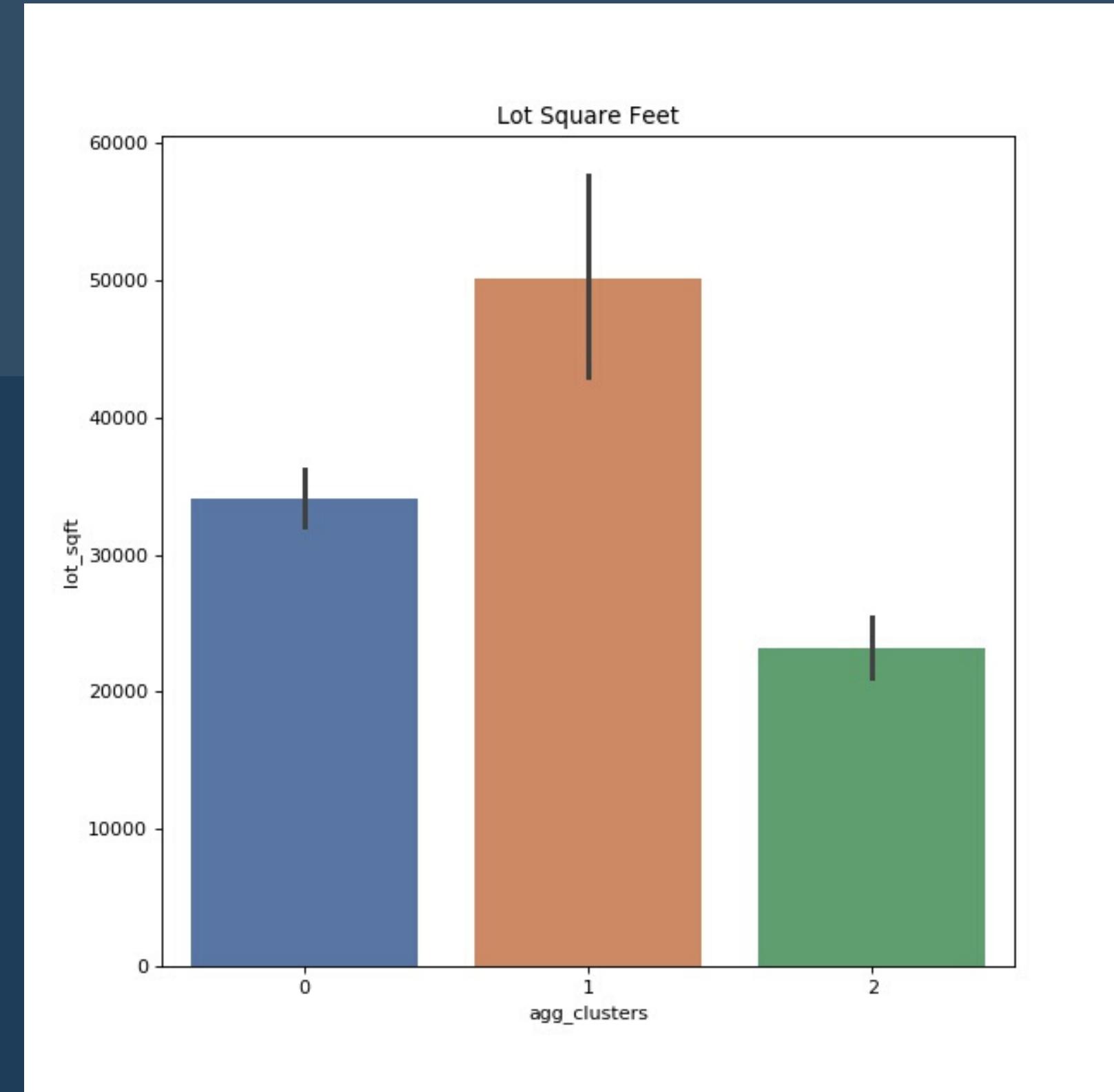
553 properties

Most expensive and longest between sales. Represent properties that swung index value.

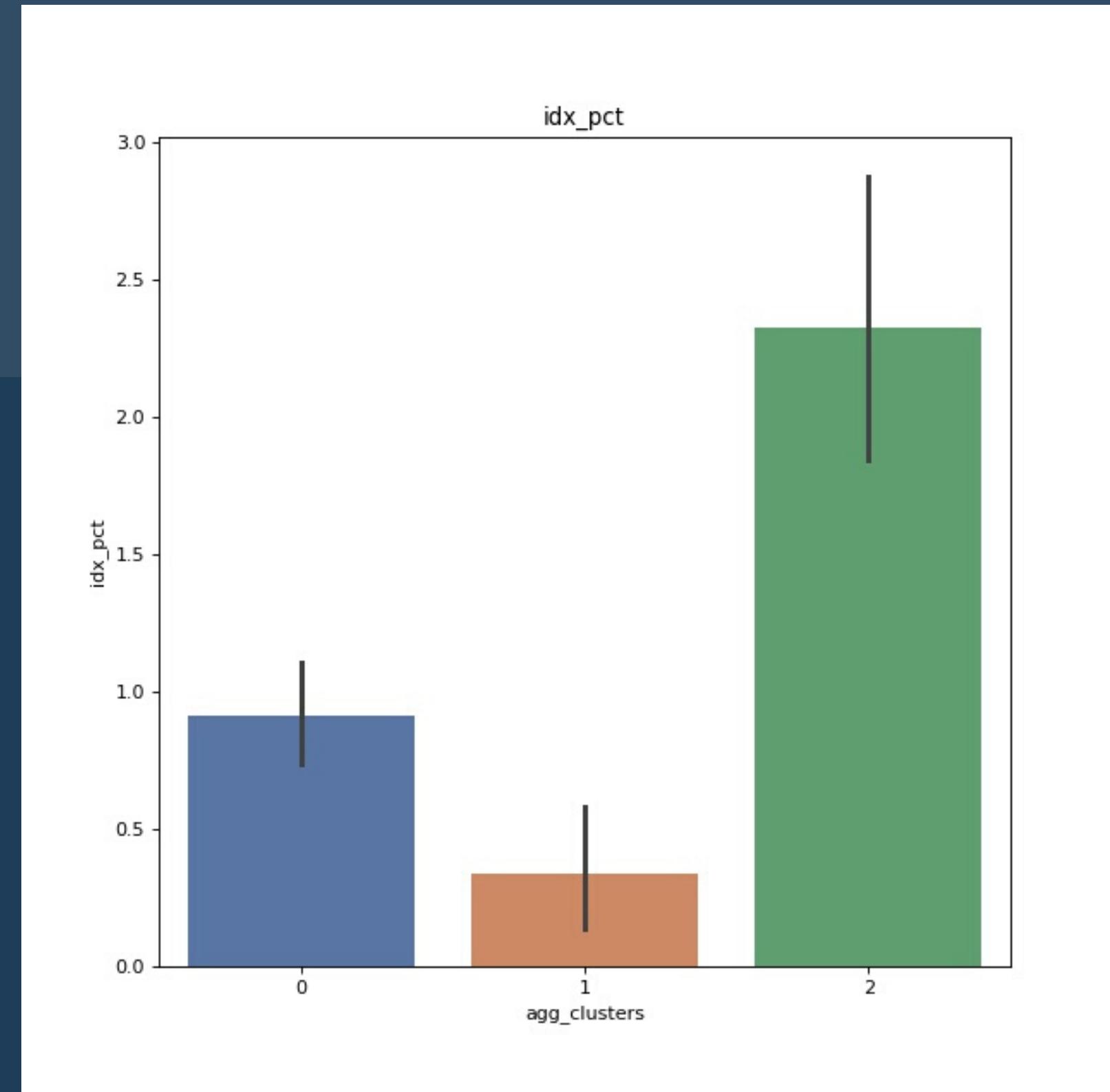
UNSUPERVISED CLUSTERING



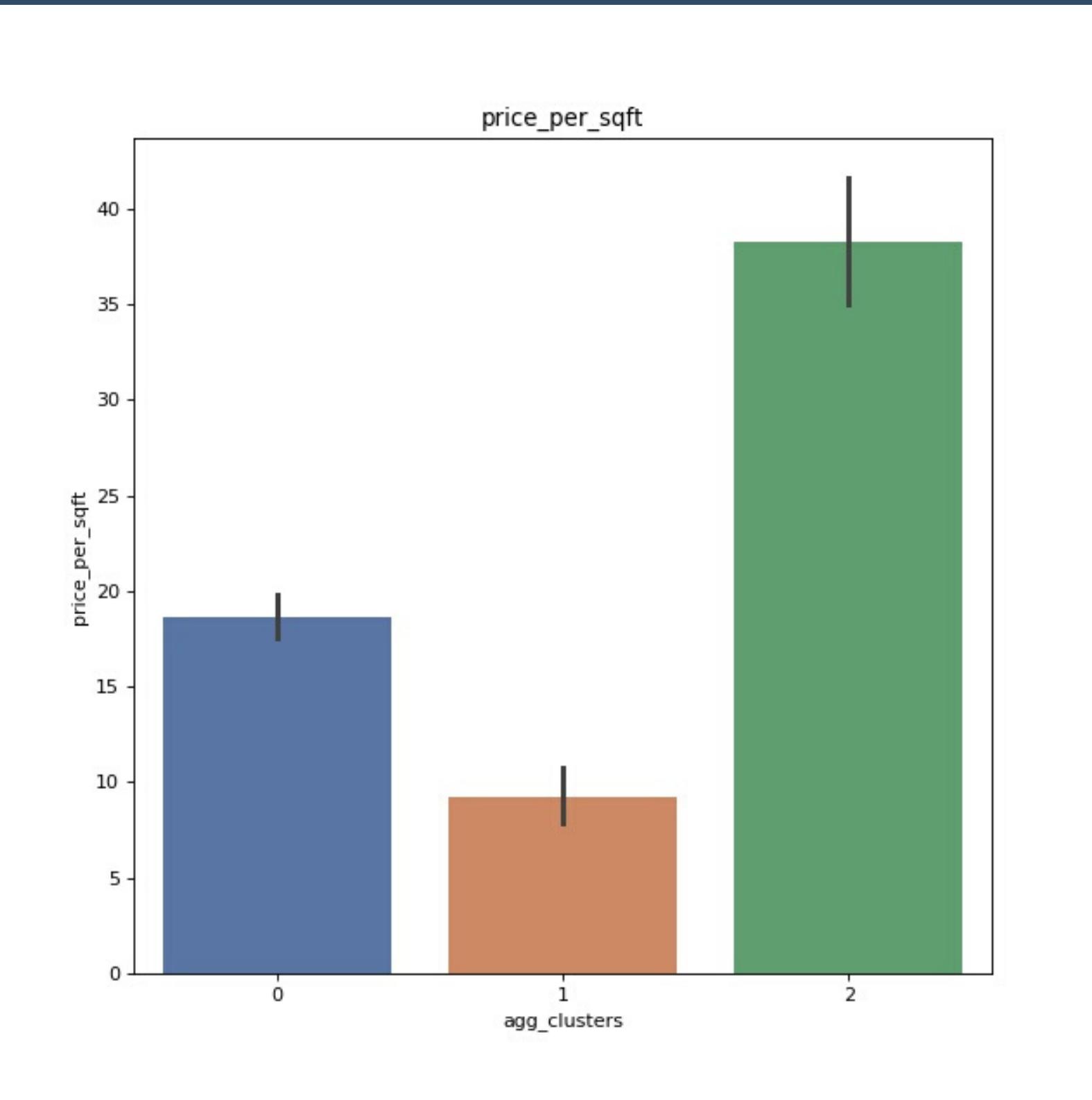
UNSUPERVISED CLUSTERING



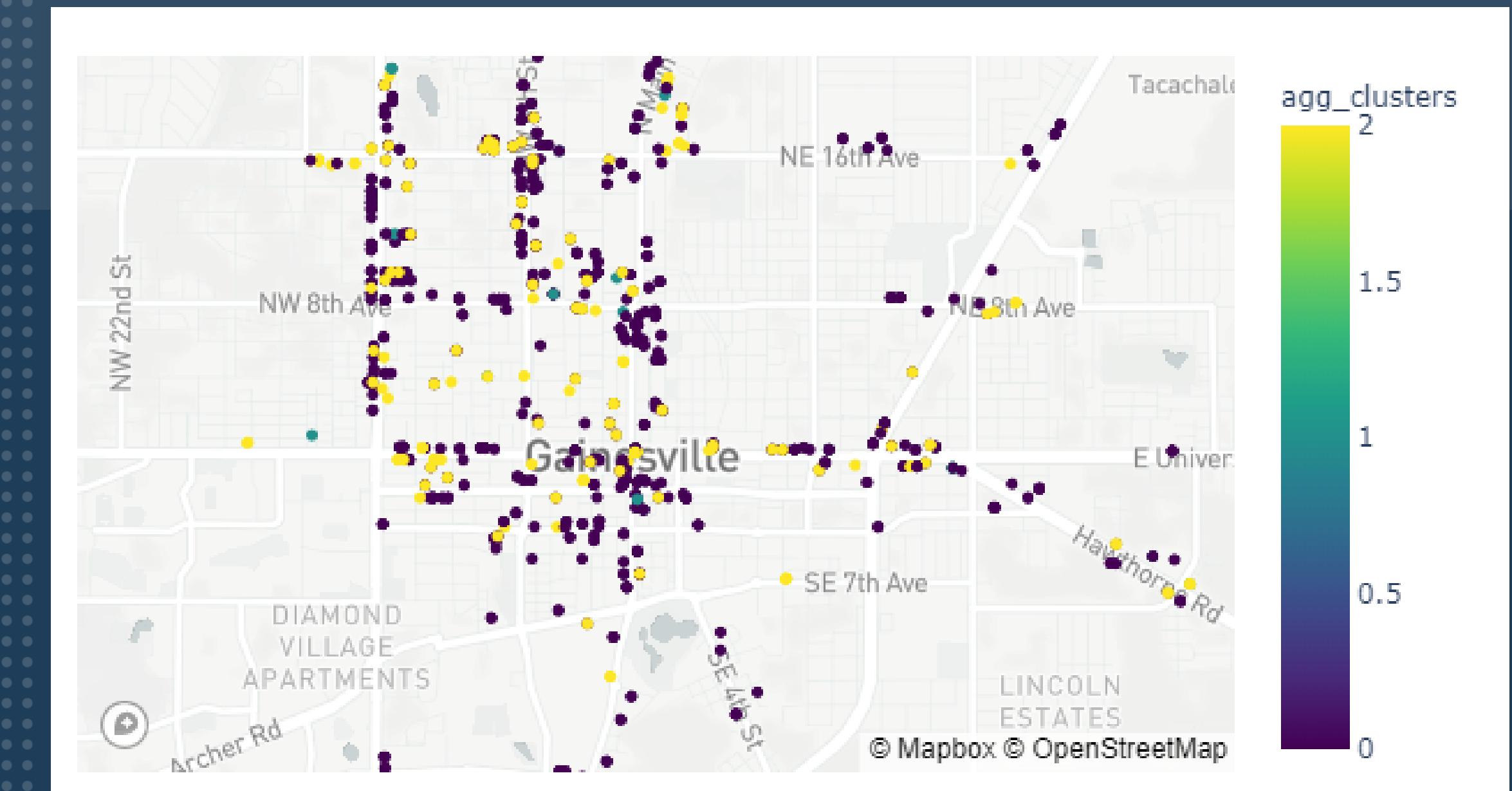
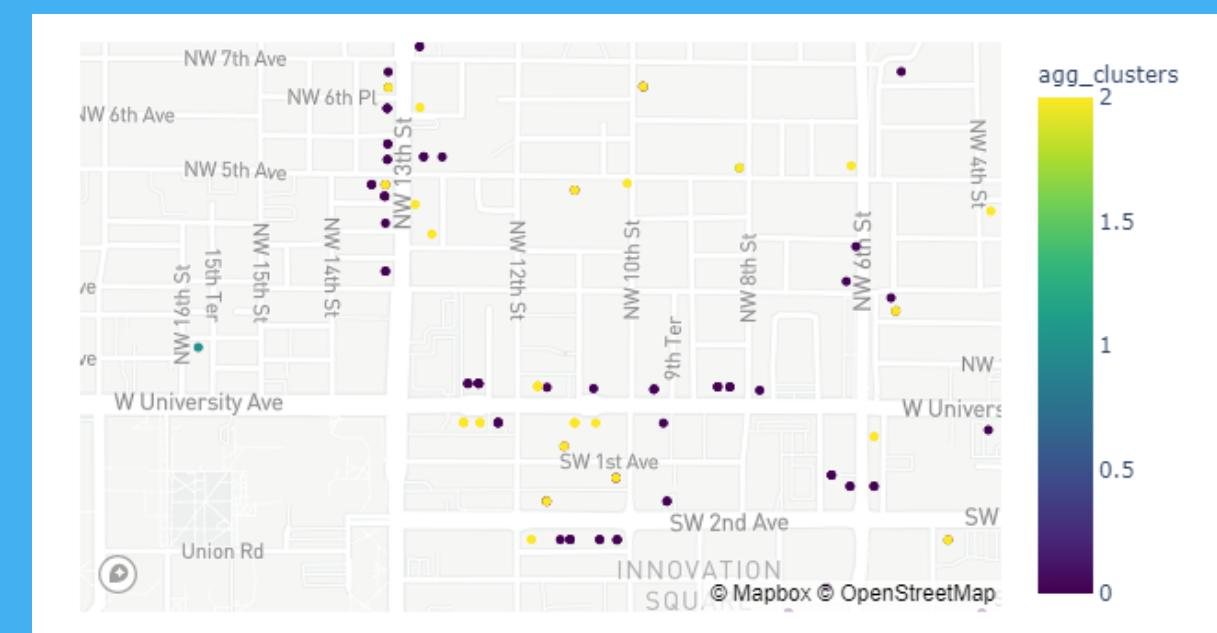
UNSUPERVISED CLUSTERING



UNSUPERVISED CLUSTERING



UNSUPERVISED CLUSTERING



Next Steps and Considerations

Dash Board

Generating all the data onto a single dashboard would provide the most value. This would allow buyers or sellers to check the relative value via the index, estimate the price for a specific property and acquire area insights in one location

Better Data

A more robust base data set would likely yield better results. A larger market (or grouping like kind markets) with multiple sales per day would help establish a more reliable index value.

Additional data on property and the area would be incredibly important as well. Building quality, financial information, traffic counts, and population density changes would all be important factors.

