

DSO 597

Undervalued Rental Property Identification

Final Project Presentation prepared by Group 2



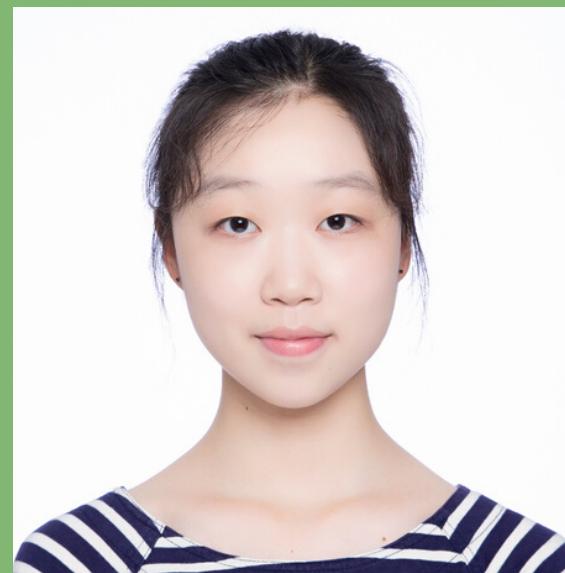
Meet Our Team



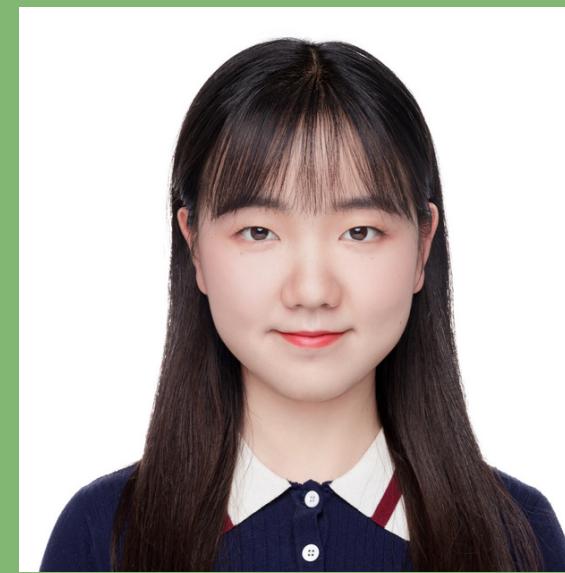
Samuel Zhu



Suixin Wang



Zheyi Tang



Bingxin Li



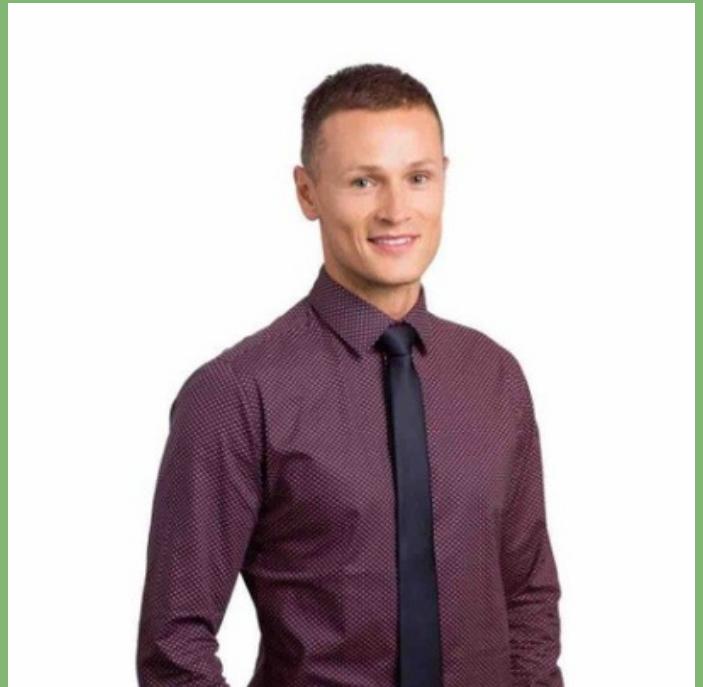
Feiran Zhang



Haowen Rui



Thank You!



Robertas Gabrys



Dylan Fonseca

Agenda

Introduction

01

Our Goal

General steps

Research & Analysis

02

Data and Sources

Data Cleaning

Exploratory Data Analysis

Final Dataset for Model

Model & Results

03

Model and Prediction

Markets suggested

Properties suggested



01 Introduction



Understand what factors drive the growth of rent
Create a predictive model
Identify the geographical areas that could yield high rent growth

Our Goal



DSO 597 | Group 2

General Steps



Research
background

Datasets
datasets of
interest

Data cleaning
normalize,
drop,
create new variables,
organize

Final data
merge

Model
4 seperate models
rank

Result
markets
properties

02 Research & Analysis



1. Yardi rental housing properties dataset
2. Neighborhood Characteristics by Census Tract dataset
3. Additional County-level data used in exploratory analysis
4. Other feature datasets

Data and sources



YARDI RENTAL HOUSING PROPERTIES DATASET

Yardi rental housing properties dataset, including details of the property, ratings from Yardi collected in December 2021. This is served as our main sources of features and target variables.

<https://www.yardimatrix.com/>

Data and sources(cont.)

NEIGHBORHOOD CHARACTERISTICS BY CENSUS TRACT DATASET

Neighborhood Characteristics by Census Tract dataset created by Harvard University's Opportunity Insights research center using census data like 2000 and 2010 Decennial Census, 2010-2016 American Community Survey, LEHD Origin-Destination Employment Statistics, etc.

We are able to extract insights from their free-and-open dataset, and build our model features based on their work.

<https://opportunityinsights.org/>

Data and sources(cont.)

ADDITIONAL COUNTY-LEVEL DATA USED IN EXPLORATORY ANALYSIS

HUD Fair Market Rent data

https://www.huduser.gov/portal/datasets/fmr/fmrs/FY2022_code/select_Geography.odn

Census Bureau Annual Resident Population Estimates in 2010-2015 time frame

<https://www2.census.gov/programs-surveys/popest/datasets/2010-2015/counties/totals/>

Data and sources(cont.)

OTHER FEATURE DATASETS

Starbucks

This Data Set is directly provided by professor via Slack.



Wholefoods

https://www.kaggle.com/adu47249/whole-foods?select=original_wf_data.csv

Hospitals

<https://www.kaggle.com/andrewmvd/us-hospital-locations>

Data and sources(cont.)



1. Yardi rental housing properties dataset
2. Neighborhood Characteristics by Census Tract dataset
3. Additional County-level data used in exploratory analysis
4. Other feature datasets

Data and sources



DSO 597 | Group 2

DATA CLEANING

Generate variable of interest:

"PROPERTY_CURRENT_RENT"

RENT per unit per square feet

Remove unimportant property types

Keep only "completed" properties

Build property AGE variable

Others...

Exploratory Data Analysis

Property Data

Census Tract Neighborhood Characteristics dataset

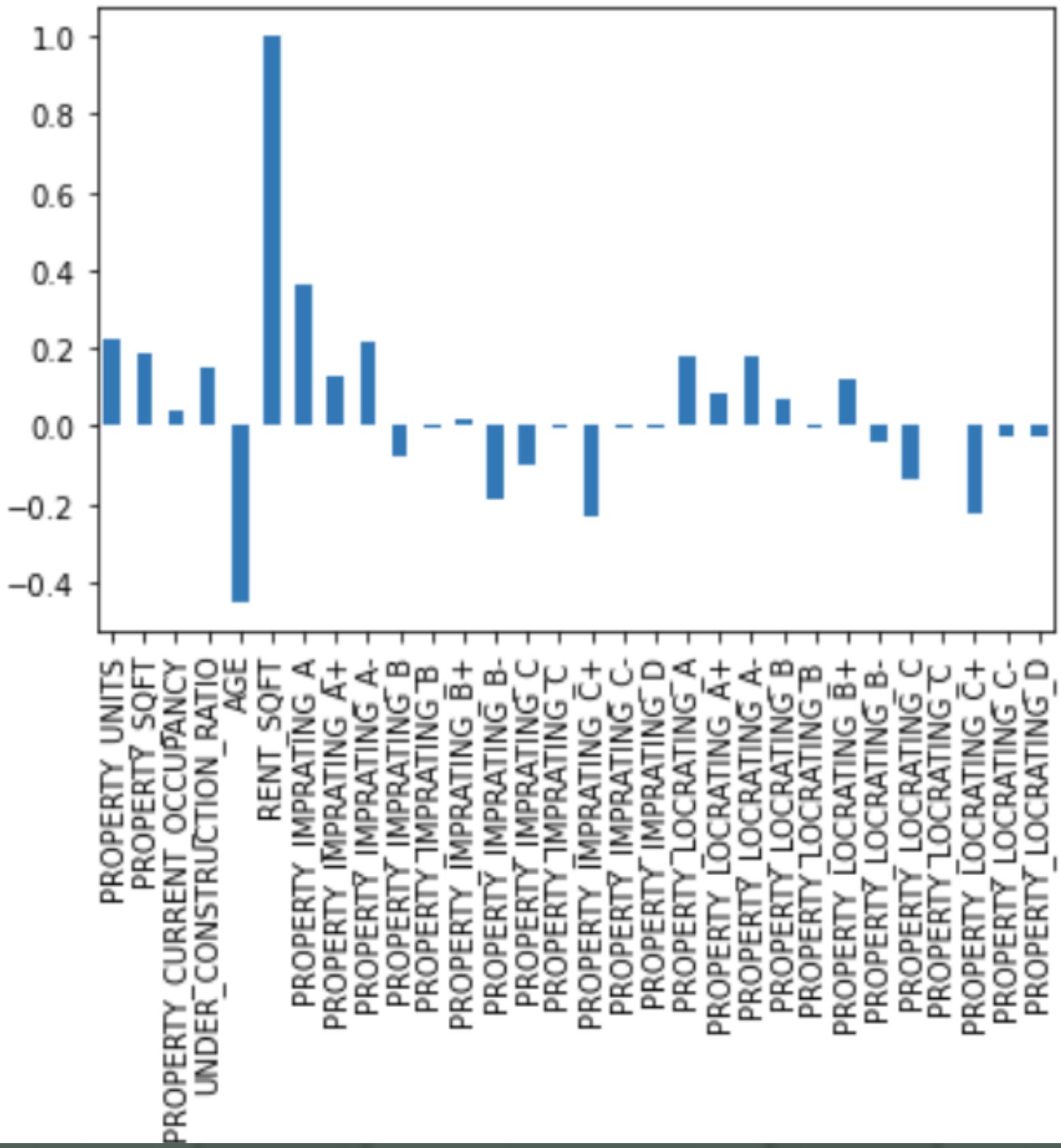
Other Feature Data

Property Data

Key variables of interest

| Column | Description |
|----------------------------|---|
| PROPERTY_ID | The Unique ID assigned to each property. |
| MARKETID | The "Market" of this Property's location defined by Yardi, roughly the size of 4 zipcode areas |
| COUNTY_NAME | The name of the county this Property locating in. |
| PROPERTY_CITY | The name of the city this Property locating in. |
| PROPERTY_STATE | The name of the state this Property locating in. |
| PROPERTY_ZIPCODE | The name of the zipcode this Property locating in. |
| PROPERTY_UNITS | Number of units of the property. |
| PROPERTY_SQFT | Number of square feet of the property. |
| PROPERTY_IMPRATING | The property's improvement and renovation score rated by Yardi, ranging from A+ to D. |
| PROPERTY_LOCRATING | The property's location score rated by Yardi, ranging from A+ to D. |
| PROPERTY_CURRENT_OCCUPANCY | The current occupancy percentage. |
| RENT_SQFT | The rent of all properties normalized by square feet and unit number. |
| UNDER_CONSTRUCTION_RATIO | The ratio of rental properties currently under construction within the specific MARKETID, indicating future supply. |
| AGE | The age of the property. |

```
# show only correlation with RENT_SQFT
subset.iloc[:, range(6, 36)].corr()['RENT_SQFT'].plot(kind='bar')
plt.show()
```



Census Data

Key variables of interest

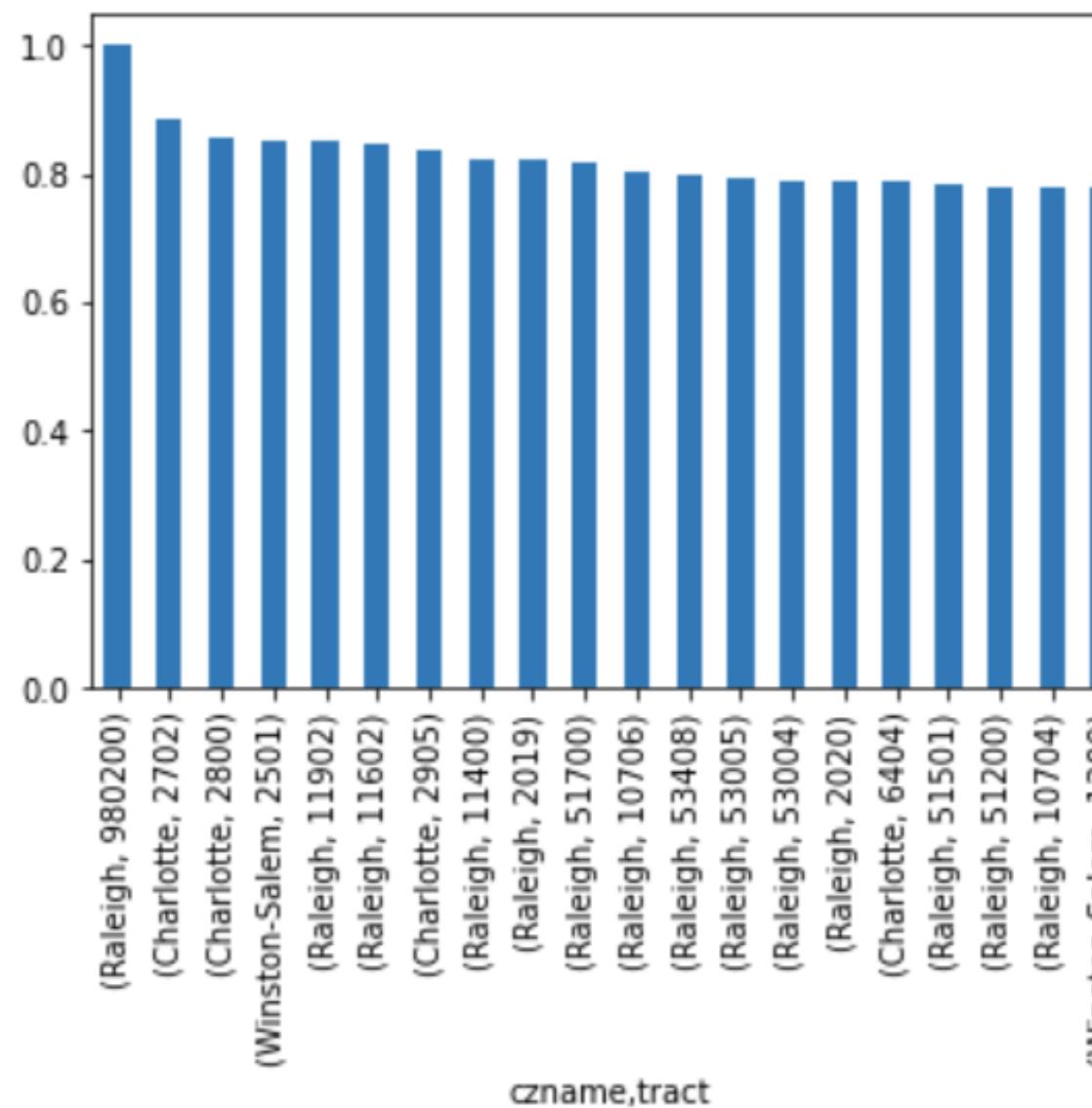
| Column | Description |
|------------------------------|---|
| ann_avg_job_growth_2004_2013 | Average annualized job growth rate |
| frac_coll_plus2010 | Number of people aged 25 or older who have a college degree or above |
| Job_density_2013 | Number of jobs per square mile in each tract. |
| Jobs_highpay_5mi_2015 | Number of jobs with earnings greater than \$3,333 per month in own and neighboring tracts within a radius of 5 miles. |
| Jobs_total_5mi_2015 | Total number of jobs in own and neighboring tracts within a radius of 5 miles. |
| Med_hhinc2016 | Median household income. |
| Poor_share2010 | Share of individuals in the tract below the federal poverty line. |
| Travelttime15_2010 | Share of workers whose commute is shorter than 15 minutes. |
| Popdensity2010 | Population density |
| mean_commutetime2000 | Mean commute time |
| rent_twobed2015 | Two-bedroom rent according to ACS (without normalization) |

Census Data (EDA)

Key variables of interest

high quality properties, job opportunity, college-grad population

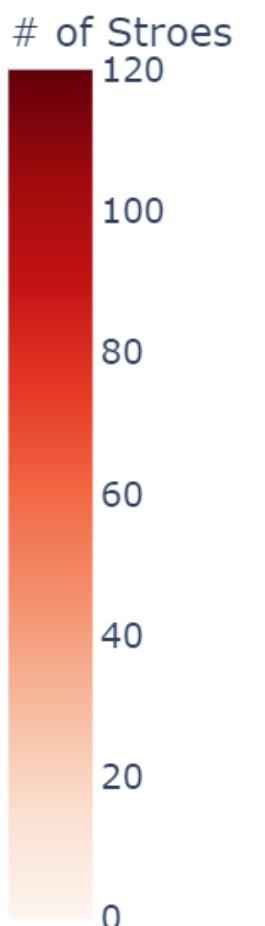
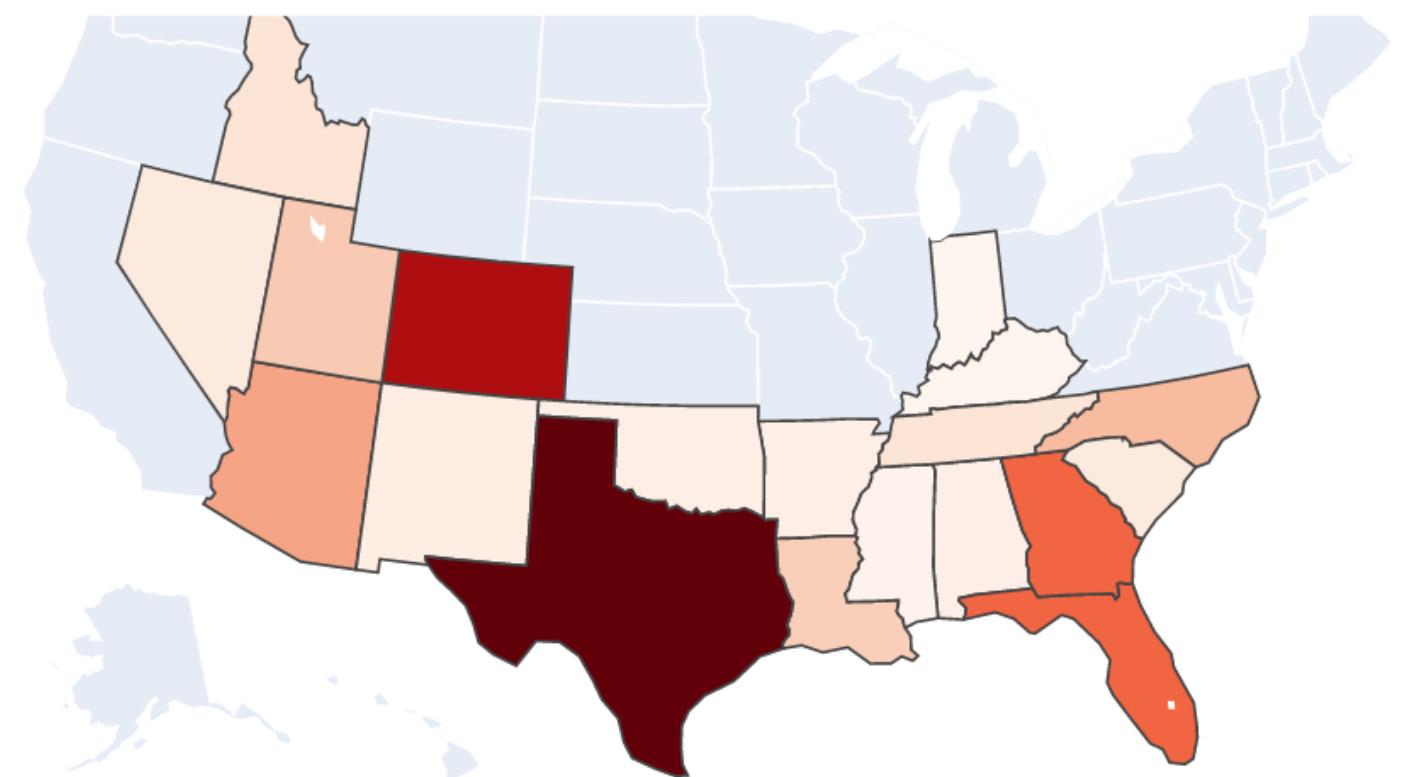
```
nc.groupby(['czname', 'tract'])['frac_coll_plus2010'].mean().sort_values(ascending=False).head(20).plot(kind='bar')  
plt.show()
```



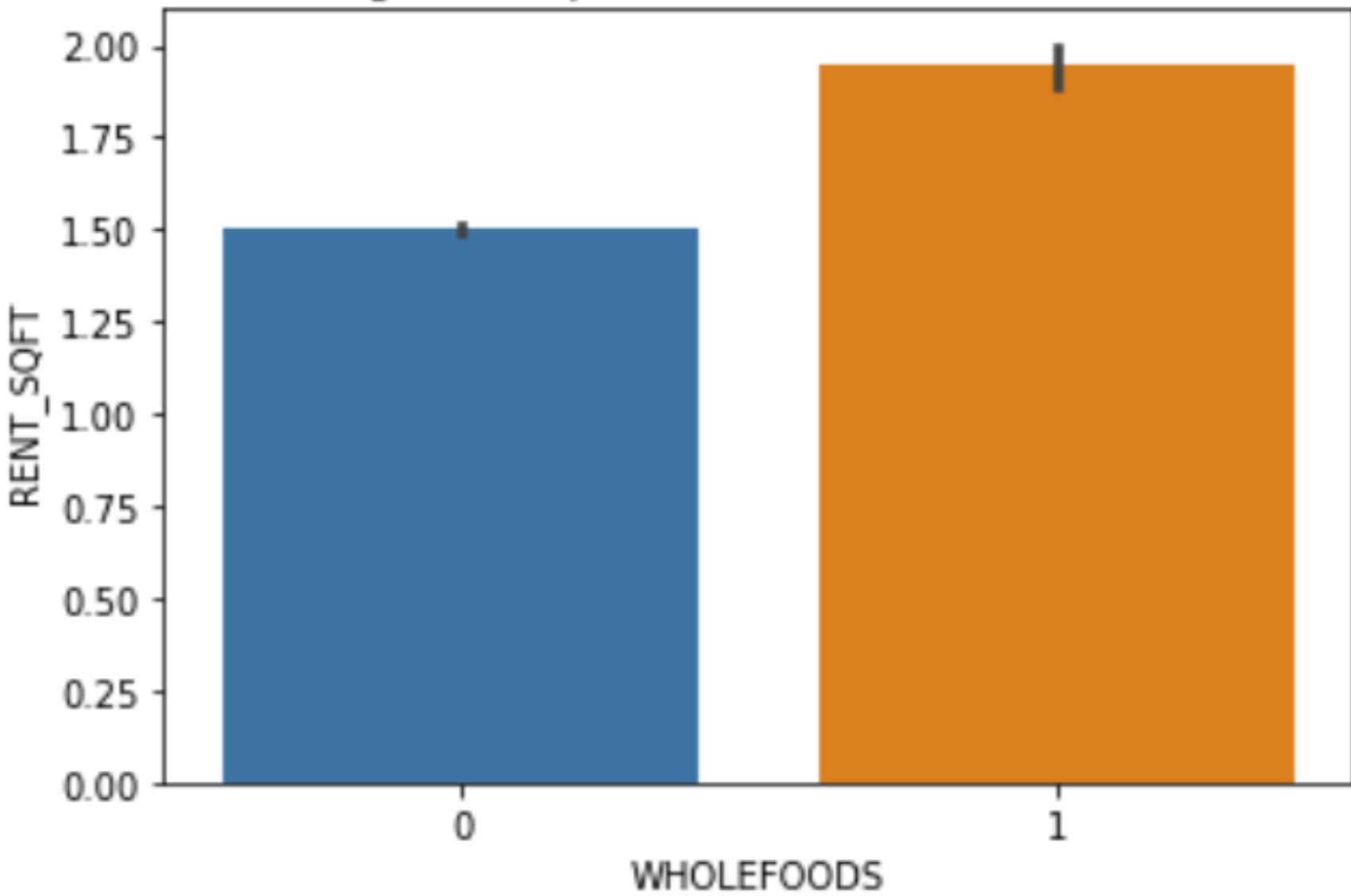
Other Feature Data (EDA)

Wholefoods

of Whole Foods Stores in Each States



Average Rents per SQFT vs. Whole Foods Stores



FINAL DATASET FOR MODEL

COMBINE ALL THOSE DATASETS

We used all three of the state#, county#, and tract# created by API package as keys to merge the *property* dataset and *tract_covariant* data set.



FINAL DATASET FOR MODEL

COMBINE ALL THOSE DATASETS

How did we calculate the number of Starbucks/Wholefoods/Hospitals around each property?

We have the longitude and latitude information for all the properties as well as amenities' nationwide locations. So, we used 'geodesic' package from 'geopy.distance' to calculate the distance and counted the number of each amenity with 1-5 miles range from each property.



03 Model & Results

Decision Tree
Random Forest
Extreme Gradient Boosting
Elastic Net

AutoSave (● Off) H ↻ 🔍 ⌂ ⌄ rank_with_data ▾ Search (Alt+Q) Zheying Tang ZT

File Home Insert Draw Page Layout Formulas Data Review View Help

Comments Share

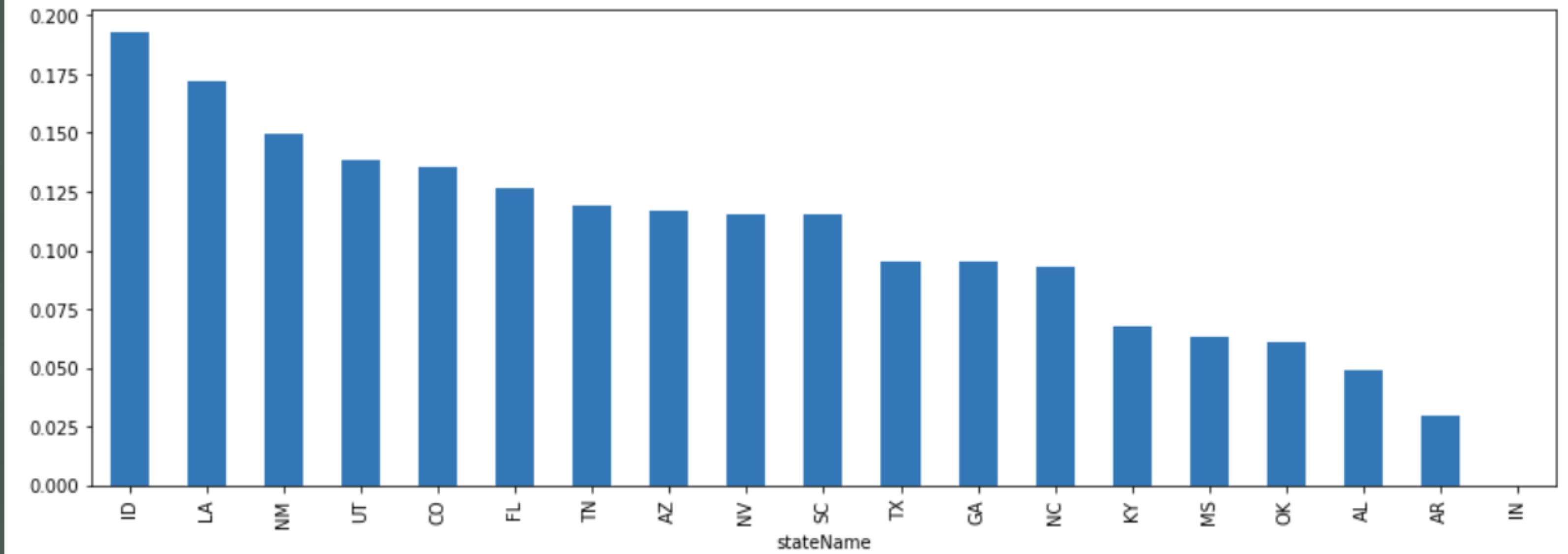
CW1 : X ✓ fx Final_Rank

| | A | CH | CI | CJ | CK | CL | CM | CN | CO | CP | CQ | CR | CS | CT | CU | CV | CW | CX |
|----|----------|-------------|-----------|-------------|--------------|---------|-------------|---------|---------|------------|--------------|----------|-------------|--------------|-----------|----------------|------------|----|
| 1 | PROPERTY | job_density | stateName | DT_result | DT_diff | DT_rank | RF_result | RF_diff | RF_rank | XGB_result | XGB_diff | XGB_rank | ENET_result | ENET_diff | ENET_rank | Final_Rank_AVG | Final_Rank | |
| 2 | 15_1028 | 2320.75 | FL | 1.815318035 | -0.168733336 | 2606 | 2.000215607 | -0.0841 | 3739 | 2.1594727 | -0.011138745 | 7216 | 1.646600994 | -0.245991893 | 370 | 3482.75 | 1246 | |
| 3 | 15_1160 | 3104.58 | FL | 1.757140242 | 0.070610962 | 15275 | 1.902330932 | 0.1591 | 20013 | 1.6584853 | 0.010501321 | 14351 | 1.704745968 | 0.038687566 | 15021 | 16165 | 18786 | |
| 4 | 15_12023 | 2085.6 | FL | 2.056480331 | -0.019016636 | 9742 | 1.908740746 | -0.0895 | 3475 | 2.056467 | -0.019022969 | 5057 | 1.833761853 | -0.125257925 | 1477 | 4937.75 | 2318 | |
| 5 | 15_1203 | 847.969 | FL | 1.774417582 | 0.117639878 | 17380 | 1.723940831 | 0.0858 | 17155 | 1.5438722 | -0.027572091 | 3220 | 1.582845356 | -0.003024368 | 9993 | 11937 | 12297 | |
| 6 | 15_1218 | 3925.65 | FL | 2.362903226 | 0.180648451 | 19184 | 1.854326643 | -0.0735 | 4346 | 1.776001 | -0.112603183 | 53 | 1.784857836 | -0.108177764 | 1814 | 6349.25 | 3875 | |
| 7 | 15_1309 | 675.841 | FL | 2.295670103 | 0.053198798 | 14293 | 1.865688291 | -0.1441 | 1470 | 1.8617439 | -0.145876202 | 25 | 1.718046784 | -0.211801031 | 555 | 4085.75 | 1643 | |
| 8 | 15_132 | 90232.7 | FL | 3.715455476 | 1.413442023 | 22438 | 1.882178833 | 0.2226 | 21212 | 1.5800157 | 0.026328054 | 18282 | 1.543354623 | 0.002514208 | 10809 | 18185.25 | 20752 | |
| 9 | 15_1385 | 1052.41 | FL | 1.913625947 | 0.024554839 | 12632 | 1.821644685 | -0.0247 | 8204 | 1.976057 | 0.057980437 | 21564 | 1.637060751 | -0.123518095 | 1507 | 10976.75 | 10733 | |
| 10 | 15_1386 | 1157.07 | FL | 1.377424167 | -0.441831268 | 83 | 2.093101191 | -0.1518 | 1305 | 2.4298353 | -0.015366413 | 5990 | 1.948475756 | -0.21042605 | 561 | 1984.75 | 458 | |
| 11 | 15_1403 | 3118.04 | FL | 1.892879353 | 0.209764762 | 19800 | 1.854173363 | 0.185 | 20600 | 1.5737357 | 0.005795752 | 12876 | 1.480939997 | -0.053511245 | 4202 | 14369.5 | 16269 | |
| 12 | 15_1408 | 6402.49 | FL | 2.080275283 | 0.045720684 | 13834 | 2.015257102 | 0.013 | 11741 | 2.0284686 | 0.019678308 | 16821 | 1.713221987 | -0.138791062 | 1292 | 10922 | 10649 | |
| 13 | 15_1410 | 771.948 | FL | 1.445850914 | 0.093375103 | 16368 | 1.554377163 | 0.1754 | 20396 | 1.2851456 | -0.028152742 | 3100 | 1.39181143 | 0.052509598 | 16204 | 14017 | 15708 | |
| 14 | 15_1412 | 382.409 | FL | 1.884413519 | 0.072432054 | 15370 | 1.761517332 | 0.0025 | 10772 | 1.785365 | 0.016062886 | 15943 | 1.714046813 | -0.024524752 | 7116 | 12300.25 | 12904 | |
| 15 | 15_1413 | 342.541 | FL | 3.086954961 | 0.306424625 | 21171 | 2.290452032 | -0.0307 | 7666 | 2.3080425 | -0.023217498 | 4054 | 2.037691508 | -0.137632263 | 1305 | 8549 | 6927 | |
| 16 | 15_1420 | 1005.48 | FL | 1.430992949 | 0.031482259 | 13040 | 1.625605657 | 0.1718 | 20321 | 1.4175652 | 0.021803346 | 17337 | 1.605067015 | 0.156957588 | 20845 | 17885.75 | 20515 | |
| 17 | 15_1427 | 7196.56 | FL | 1.068604651 | -0.277758227 | 730 | 1.65417403 | 0.118 | 18708 | 1.4895029 | 0.006715817 | 13197 | 1.420913183 | -0.039642172 | 5423 | 9514.5 | 8362 | |

diff = (pred-real)/real

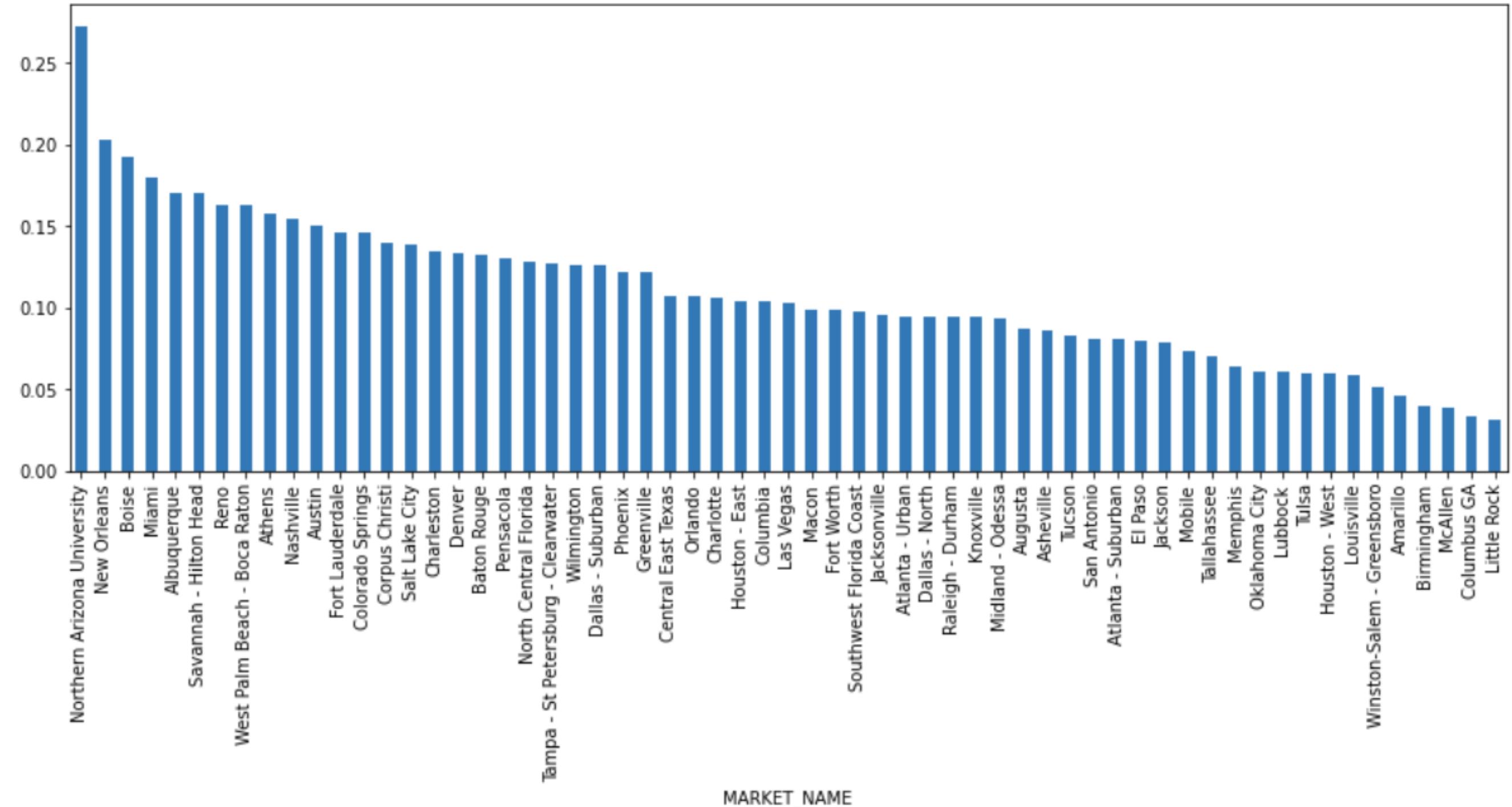
RESULT ANALYSIS

```
# abnormal rate by stateName  
plt.rcParams["figure.figsize"] = (15, 5)  
data_sort.groupby(['stateName'])['label'].mean().sort_values(ascending=False).plot(kind='bar')  
plt.show()
```



RESULT ANALYSIS

```
# abnormal rate by market
plt.rcParams["figure.figsize"] = (15, 5)
data_sort.groupby(['MARKET_NAME'])['label'].mean().sort_values(ascending=False).plot(kind='bar')
plt.show()
```



Property suggested

```
# Top 10 most abnormal properties  
sub_data.head(10)
```

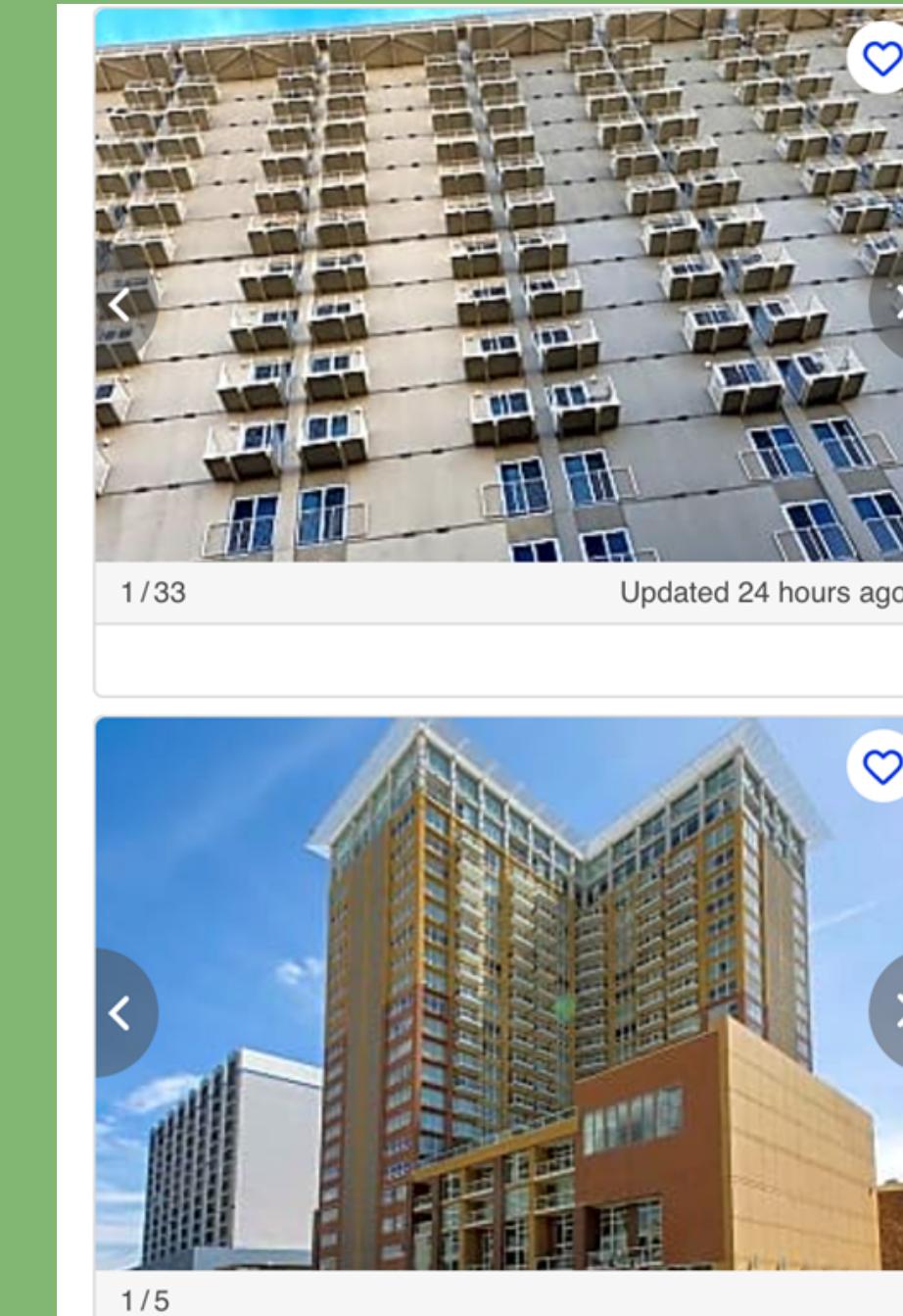
| PROPERTY_ID | MARKETID | MARKET_NAME | SUBMARKET_CODE | SUBMARKET_NAME | COUNTY_NAME | PROPERTY_NAME | PROPERTY_PRIORNAME\$ | PROP |
|-------------|----------|------------------|----------------|-----------------------|------------------|------------------|----------------------------------|----------|
| 76_531 | 76 | Reno | 1 | Reno - central | Washoe | Westlyn, The | Colonial, Ross, The | |
| 14_5266 | 14 | Las Vegas | 4 | Downtown Las Vegas | Clark | 211, The | Avenue 8 Studios, Campaige Place | 21 |
| 46_674469 | 46 | Raleigh - Durham | 4 | Hinton | Wake | Pullen Station | Pullen Park Lofts | |
| 73_41 | 73 | Birmingham | 1 | Birmingham - central | Jefferson | University Place | | NaN 951 |
| 15_453 | 15 | Miami | 15 | Coral Gables | Miami-Dade | Fountains, The | | NaN 23 |
| 116_250 | 116 | New Orleans | 11 | Metairie | Jefferson Parish | Premises | | NaN 3401 |
| 14_5292 | 14 | Las Vegas | 4 | Downtown Las Vegas | Clark | Desert Rose | Desert Rose/Summerplace | 29 |
| 38_1635 | 38 | Fort Worth | 49 | Great Southwest | Tarrant | Silver Bell | | NaN 3232 |
| 71_637 | 71 | Nashville | 14 | Southeast - west | Davidson | Mandolin | Warren House | 420 |
| 53_3 | 53 | McAllen | 21 | Brownsville - Central | Cameron | Reypres Square | | NaN 13 |

10 rows × 101 columns

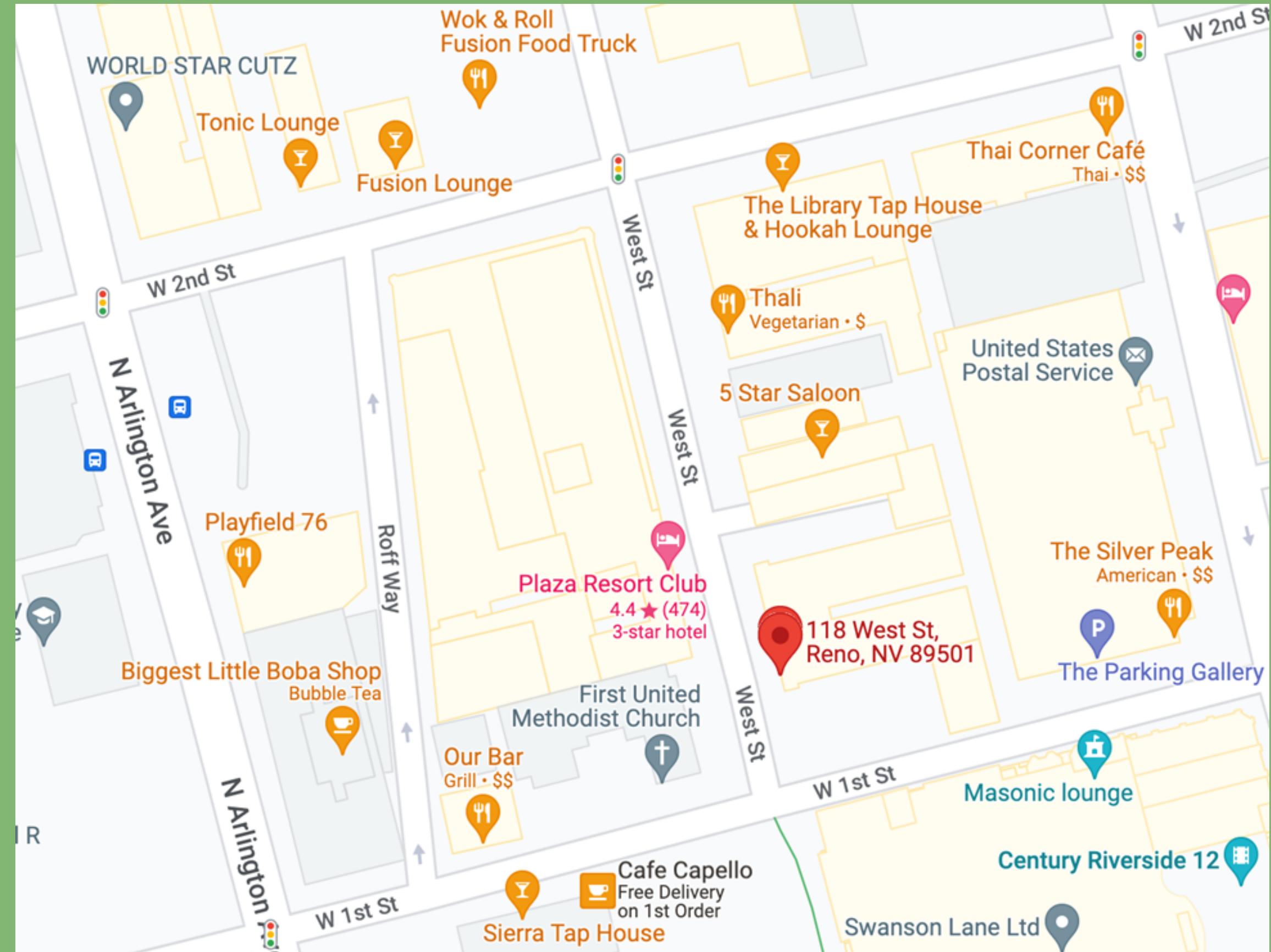
Examples of Undervalued Properties



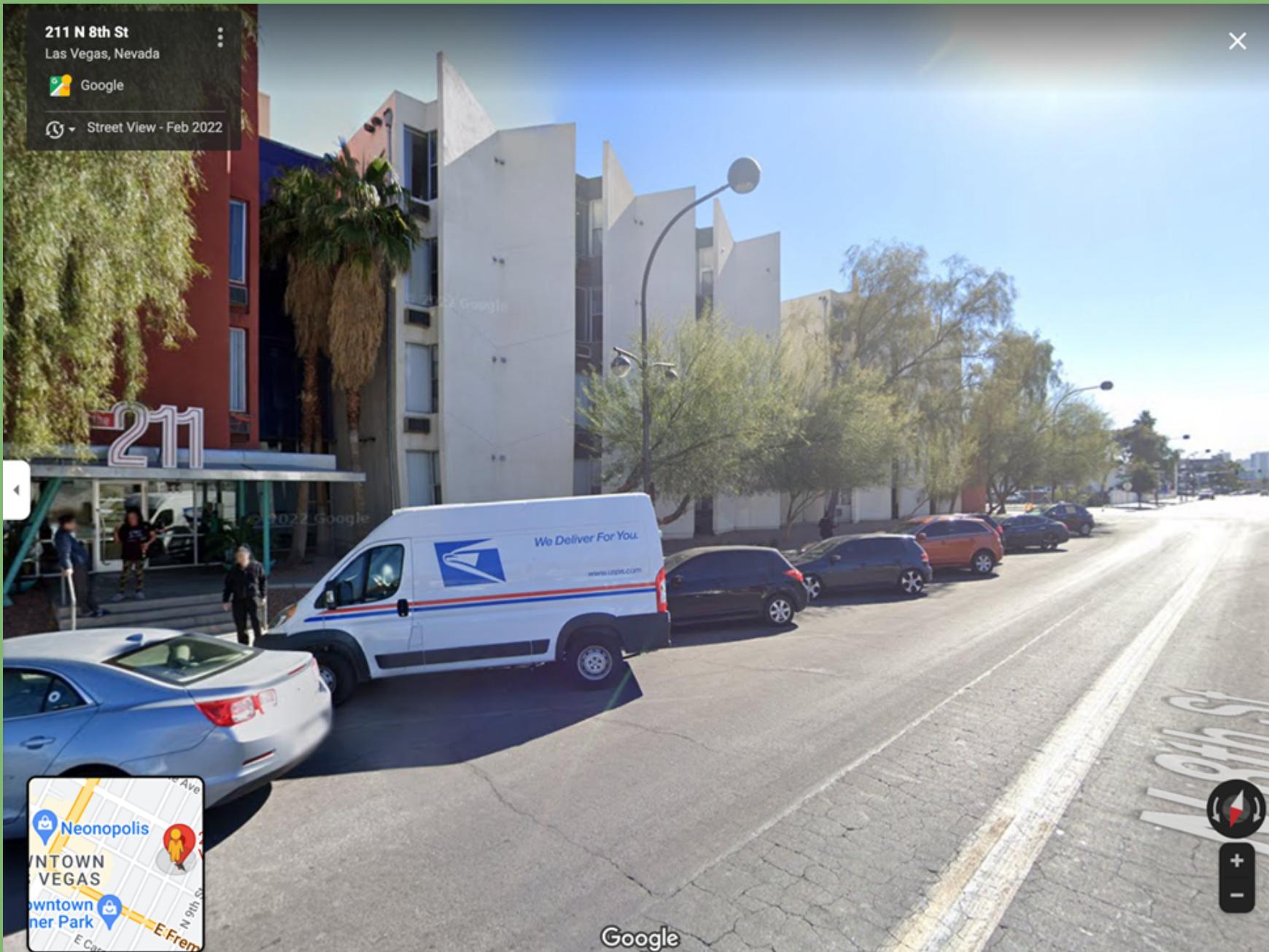
118 West Street, Reno, NV



Rent 1030



211 North 8th Street, Las Vegas, NV



Rent 807

30 Properties Sort by: Best Match ▾



Deals Cash Off

1/23 Updated 2 days ago

Top Amenities

Cats Allowed & Dogs Allowed

Southeast Las Vegas

(702) 859-5219

Contact Property

Available 4/5

The Vine on 8th
811 E Bridger Ave, Las Vegas, NV 89101

| | |
|----------------------|-------------------|
| Studio 1 Bath..... | \$1,245 |
| 325+ Sqft | 1 Floor Plan |
| 1 Bed 1 Bath..... | Contact for Price |
| 385+ Sqft | 1 Floor Plan |
| 2 Beds 1 Bath..... | Contact for Price |
| 516+ Sqft | 1 Floor Plan |

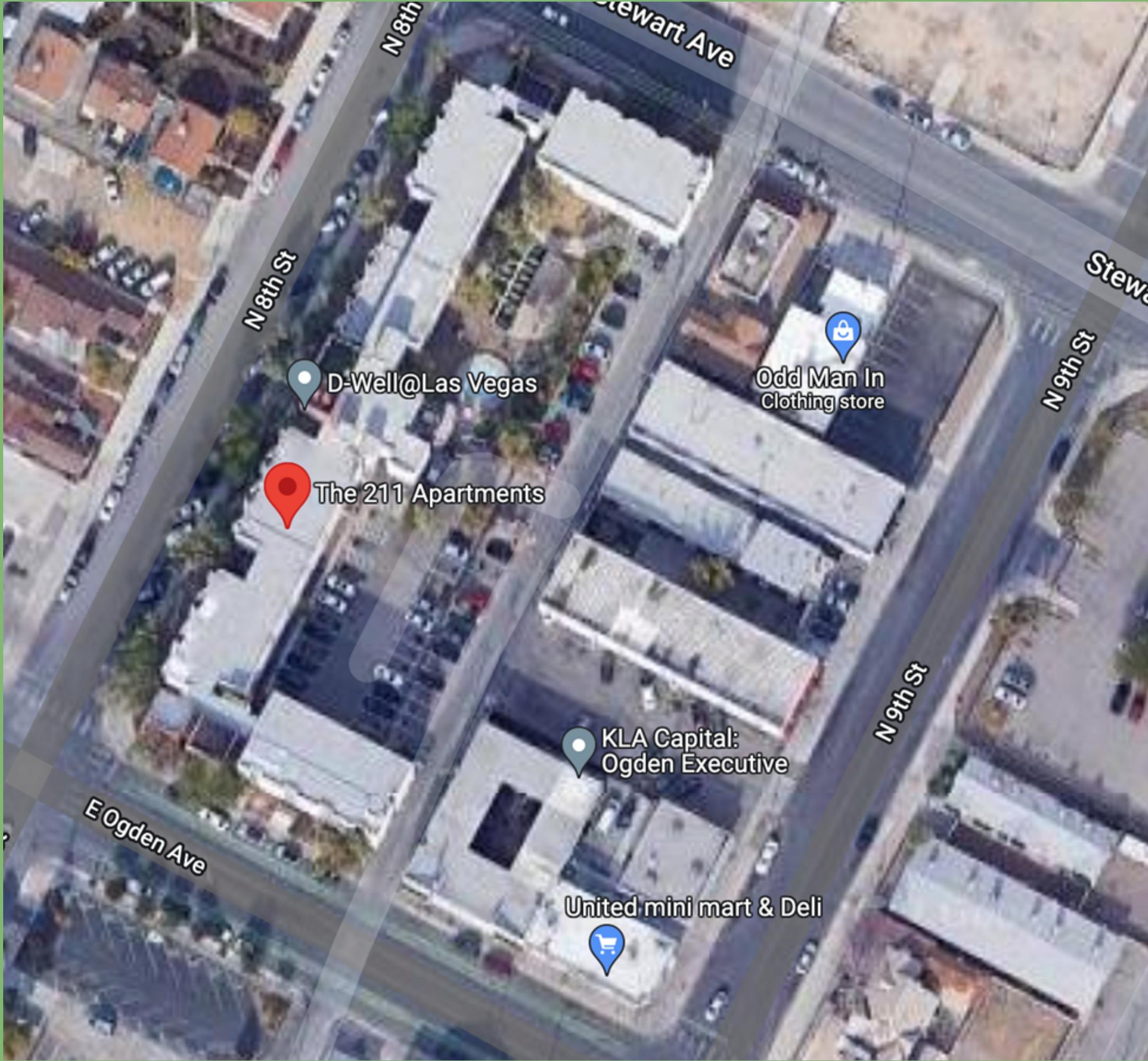
(702) 859-5219 **Contact Property**

Not Available

211 Apartments
211 N 8th St., Las Vegas, NV 89101

| | |
|----------------------|-----------------|
| Studio 1 Bath..... | \$695–\$900 |
| 160–335 Sqft | 12 Floor Plans |
| 1 Bed 1 Bath..... | \$1,500–\$1,600 |
| 655+ Sqft | 1 Floor Plan |

(702) 805-4748 **Contact Property**



Thank You

谢谢

Group members: Zhey ing Tang, Samuel Zhu, Suixin Wang, Feiran Zhang, Bingxin Li, Haowen Rui