Talia Berler
JMM 463
Professor Bugera
12/13/22

A Study of Change: Real Estate Analysis of Coconut Grove

Abstract – *As the neighborhood of Coconut Grove undergoes a period of rapid growth, this analysis reports on the area's real estate portfolio. After providing background on the history of the Grove, the borders and internal segmentation were decided based on zip code and personal familiarity with the area. The dataset from the property appraiser's site provided approximately 14,000 observations with 50 features, after feature generation. The data analysis focused on property types and number of sales across the board, then finding most frequent attributes of properties in each segment of the Grove. Predictive models were then designed; a linear regression was developed to predict the most recent sale amount of properties since 2020, and a logistic regression was developed to predict whether a property was a primary residence. Results showed that the Central Grove appears to be developing at a faster pace than either the North or South Grove and that property values have doubled in the last decade.*

1. Introduction

The neighborhood of Coconut Grove is alluring to most for its overgrown foliage obscuring eccentric homes off of Biscayne Bay and for its convenient location, south of the indulgent, party scene for which Miami is known. The area has been a coveted location for decades, but has experienced a boom in growth in recent years. The Miami Herald reports that, "2021 was a record year in the Grove for average residential sale price, average price per square foot, and number of homes sold" [1]. The Grove has changed extensively even in my lived experience as a resident, since 2010. Yet, the visible shift in the neighborhood's demographic makeup truly kicked off during the summer of 2020.

In the peak of the COVID-19 pandemic, wealthy buyers from California and New York flooded into South Florida, including Coconut Grove, for the weather and lack of state income-tax [1]. The new wave of migration has brought change to the physical appearance of the

Grove as well, with 50% of demolition permits issued in the city of Miami last year being for houses in the 5.6 square miles that make up the neighborhood, only 10% of Miami's total area [1]. Despite Miami's continuous growth and exposure from events such as the influx of Cuban refugees, the drug wars of the 80s, the Wynwood Walls and Art Basel scenes, and the golden age of the Miami Heat in the early 2010s, Coconut Grove had remained largely unchanged through the years. The area has a rich history, with cultural influences from across Latin America and the Caribbean. Coconut Grove is Miami's oldest community, established with the opening of the post office in 1873, long before the city was officially incorporated in 1896 [2]. Now, in the throes of the construction and development of the post-pandemic atmosphere, many residents of the area feel affronted by the rapidly transforming landscape. In recognition of this watershed period for my home, I embarked on this analysis of the current real estate market in Coconut Grove and its distinct geographic sections known to locals, but not officially designated. Using Python and relevant Python libraries, I will provide insight on the Grove's housing market today, define the internal sections of the neighborhood, and examine trends in each section in the hopes of understanding what the future may hold for the historical Miami neighborhood.

2. Data

2.1 Source and Segmentation

The data I used in this analysis originated from the Miami-Dade Property Appraiser's site, provided with the help of Professor Bugera [3]. Since Coconut Grove is a neighborhood, not a city, it has loosely defined borders that do not line up with legal indicators, such as zip codes. Both zip codes 33133 and 33129 are included in the limits of Coconut Grove, according to Google Maps (Fig 1). The northwest edge of the Google Maps-provided definition overlaps with

Brickell, so we limited the northmost edge to SW 32nd Road, just above the Vizcaya Estate.

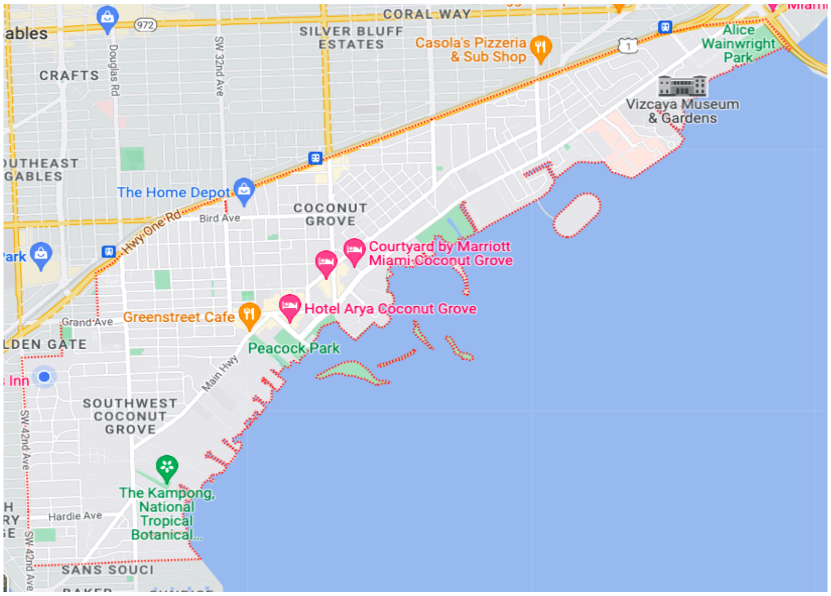

Figure 1. Google Maps designation of Coconut Grove area in the dotted red line border.

The moved border excludes properties in 33129, limiting our definition of Coconut Grove to the 33133 zip code.

Colloquially, the Grove is often divided into four parts: north, south, center, and west. I elected to view the middle area of the Grove as one section, combining the traditionally Caribbean West Grove, recently renamed Little Bahamas [2], and the surrounding Center Grove. These areas have become quite fluid due to gentrification, and difficult to properly separate [3]. Thus, we ended up with three approximately equally-sized segments (Fig 2). The South Grove begins just above Cocoplum Circle, extending northward until Franklin Ave–a relic of the Jim Crow-era, separating the Little Bahamas above from the traditionally white neighborhood to the south–and Armbrister Park, and bordered by LeJeune Road and Biscayne Bay on either side. The Central Grove then consists of the Little Bahamas area and the Center Grove, bordered by US 1 and the bay, and ending north of Kennedy Park along 22nd Ave. Finally, the North Grove is bordered to the north by US 1, includes Grove Isle, and ends above Vizcaya.

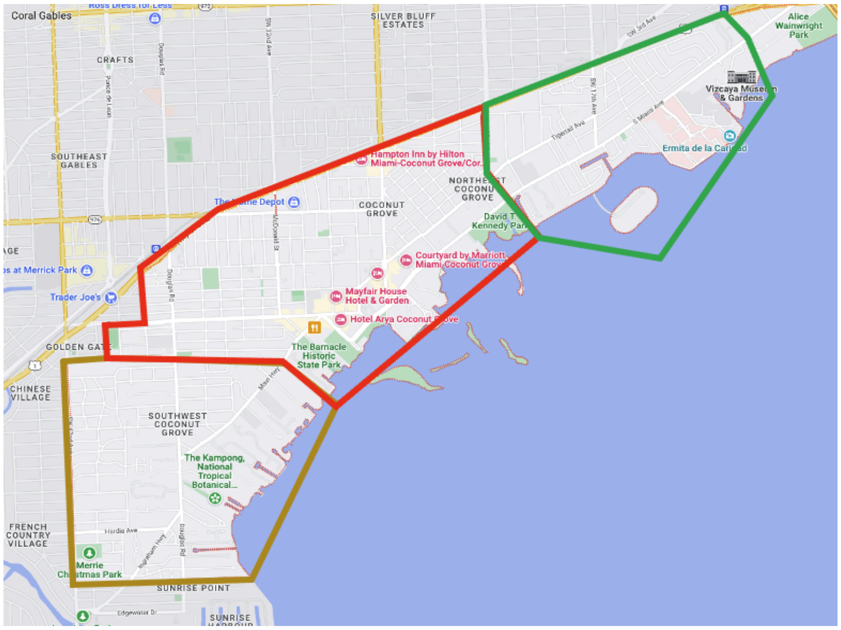

Figure 2. Professor Bugera and my designation of Coconut Grove and the 3 segments: North (green), Central (red), and South (brown).

2.2 Initial Load

After defining the borders of the Grove and its subsections, Professor Bugera assigned each property in the two zip codes to North, Central, South, or Outside, meaning it fell outside the borders we drew. The dataset with all the features of the properties uses a folio number to reference each individual property, so in a second dataset Professor Bugera assigned the area of the properties to their respective folio numbers. We then had the two datasets needed: one with all properties in 33133 and 33129 and their features given by the Property Appraiser's site, and another with the Coconut Grove properties inside the 33133 zip code assigned to one of the four subsection areas (Fig 3,4). To gather only the properties in Coconut Grove, we renamed the columns in the Areas dataset and inner merged the datasets on the folio number, excluding folio numbers that were not shared across both sets (Fig 5).

```
#Define Input Data
files = ["Talia_09281445_COCONUT_GROVE_COUNTYDATA.csv","CGROVE_AREAS.csv"]
#/content/drive/MyDrive/PROJECTS/INPUT/Talia_09281445_COCONUT_GROVE_COUNTYDATA.csv
```

Figure 3. File names of csv datasets to be loaded

```
(21850, 66) > Parcels: (21850, 66)
(14945, 3) > Areas: (14945, 3)
```

Figure 4. The shapes of the datasets before and after loading into Pandas DataFrames.

```
AREAS = AREAS.rename(columns = {"PARCEL20":"Folio", "XF_TYPE":"AREA"})
INPUTDATA1 = INPUTDATA.merge(AREAS, on="Folio", how="inner")
```

Figure 5. Merging the DataFrames with an inner merge on folio numbers.

After merging the datasets, I reviewed the variables and kept only the following features needed for the subsequent analysis:

1. FOLIO – reference number for each individual property
2. PROPERTYADDRESS – street address of the property
3. LAND – taxable value of the land
4. BLDG – taxable value of the building
5. TOTAL – combined taxable value of land and building
6. ASSESSED – taxable value of land and building with tax exemptions applied
7. HEX – Homestead Exemption, signifying a property as a primary residence with the value 25,000 (the amount of tax exemption granted for primary residences) or a 0
8. LANDUSE – current use of the land (ie. residential, commercial,etc.)
9. ZONING – the original intention of what the land should be used for
10. OWNER1 – owner of the property
11. OWNER2 – secondary owner of the property, if any
12. MAILINGADDRESS – mailing address of the property owner
13. MAILINGSTATE – mailing state of the property owner
14. MAILINGCOUNTRY – mailing country of the property owner
15. ADJUSTEDSQFT – taxable square footage of the property adjusted for features
16. LOTSIZE – size of the lot
17. BED – number of beds
18. BATH – number of bathrooms
19. STORIES – number of stories
20. UNITS – number of living units in the property (multi-residential, some condominiums)
21. YEARBUILT – year the property was built
22. EFFECTIVEYEARBUILT – year of most recent extensive renovation of the property

23. SALETYPE1 – specific type of most recent sale, more in depth information
24. SALEQUAL1 – whether most recent sale was qualified or unqualified, implying that the sale reflects the market value of the property (Q) or does not reflect market value (U)
25. SALEDATE1 – date of most recent sale
26. SALEAMT1 – amount the property was sold for most recently
27. SALETYPE2 – next most recent sale, see SALETYPE1
28. SALEQUAL2 – next most recent sale, see SALEQUAL1
29. SALEDATE2 – next most recent sale, see SALEDATE1
30. SALEAMT2 – next most recent sale, see SALEAMT1
31. SALETYPE3 – third most recent sale, see SALETYPE1 or 2
32. SALEQUAL3 – third most recent sale, see SALEQUAL1 or 2
33. SALEDATE3 – third most recent sale, see SALEDATE1 or 2
34. SALEAMT3 – third most recent sale, see SALEAMT1 or 2
35. XF1 – extra features of the property (pool, patio, etc.), if any
36. XF2 – other extra features, if any
37. XF3 – other extra features, if any
38. LIVINGSQFT – square footage of property with air conditioning
39. ACTUALSQFT – square footage of property including non-air conditioned areas
40. AREA – subsection of Coconut Grove assigned manually

2.3 Cleaning and Preprocessing

To handle missing values in the merged dataset, we looped through the columns, filling empty spaces with an empty string in the object columns and a zero in the numeric columns. All strings were also transformed into uppercase to standardize the dataset. Once missing values were handled, we were able to move on to feature generation (Fig 6,7).

```
STEP2DATA["MARKETVAL"] = STEP2DATA["TOTAL"].apply(lambda x: x*1.25)
# Define Property Type
STEP2DATA["PROPTYPE"] = np.where(STEP2DATA["LANDUSE"].str[:2]=="01","SINGLE",
                        np.where(STEP2DATA["LANDUSE"].str[:2].isin(["04","05"]),"CONDO",
                        np.where(STEP2DATA["LANDUSE"].str[:3].isin(["039"]),"HOTEL",
                        np.where(STEP2DATA["LANDUSE"].str[:2].isin(["03","08"]),
                                  np.where(STEP2DATA["UNITS"]<=4,"MULT-RESI","MULTI-COMM"),"OTHER"))))
# Define if Property is Primary Residence
STEP2DATA["PRIMARY"] = np.where(STEP2DATA["HEX"]>0,1,0)

# Define if Property has a pool
STEP2DATA["POOL"] = np.where(STEP2DATA["XF1"].str.upper().str.find("POOL")>=0,1,
                    np.where(STEP2DATA["XF2"].str.upper().str.find("POOL")>=0,1,
                    np.where(STEP2DATA["XF3"].str.upper().str.find("POOL")>=0,1,0)))

# Define if Company (True or False)
ENDINGS = ["LIMITED LIABILITY COMPANY","LLLC","LLC"," LC","LCC",
           "LIMITED PARTNERSHIP","LIMITED","LTD","LLLP","LLP","PLLC","PLC"," PL"," LP","PC",
           "PROFFESIONAL ASSOCIATION"," PA",
           "INCORPORATION","INCORPORATED","CORPORATION","COMPANY"," INC","CORP"," CO"]
STEP2DATA["COMPANY"] = False
for end in ENDINGS:
  endlen = len(end)
  STEP2DATA["COMPANY"] =  np.logical_or(STEP2DATA["COMPANY"],
                                        STEP2DATA["OWNER1"].str[-endlen:]==end)
  STEP2DATA["COMPANY"] =  np.logical_or(STEP2DATA["COMPANY"],
                                        STEP2DATA["OWNER2"].str[-endlen:]==end)
```

Figure 6. Feature generation; market value, property type, pool, and company.

```
# Define last sale amount and last sale date
STEP2DATA["LASTQSALEDT"] = 0
STEP2DATA["LASTQSALEAMT"] = 0

for I in ("1","2","3"):

  STEP2DATA["SALEDT"+I] = np.where(STEP2DATA["SALEDATE"+I]=="","0",
                                   STEP2DATA["SALEDATE"+I].str[6:10] +
                                   STEP2DATA["SALEDATE"+I].str[0:2] +
                                   STEP2DATA["SALEDATE"+I].str[3:5])
  STEP2DATA["SALEDT"+I] = STEP2DATA["SALEDT"+I].astype(int)

  STEP2DATA["LASTQSALEAMT"] = np.where((STEP2DATA["LASTQSALEDT"]==0) & (STEP2DATA["SALEQUAL"+I]=="Q"),
            STEP2DATA["SALEAMT"+I],STEP2DATA["LASTQSALEAMT"])

  STEP2DATA["LASTQSALEDT"] = np.where((STEP2DATA["LASTQSALEDT"]==0) & (STEP2DATA["SALEQUAL"+I]=="Q"),
            STEP2DATA["SALEDT"+I],STEP2DATA["LASTQSALEDT"])
```

Figure 7. Continuation of feature generation: last qualified sale date and amount.

We created features to provide the market value of each property using the total value, the type of property using the land use code, whether or not the property has a pool using the extra features, whether the owner is a private individual or a company, whether or not the property is a primary residence and when the last qualified sale was and what the property was sold for.

3. Data Analysis

3.1 Overview of Coconut Grove

The primary objective of my data analysis was to provide insight into the real estate market of Coconut Grove overall and then dive deeper into each subsection of the neighborhood. I first looked at the property counts and the types of properties (Fig 8). The majority of properties were in the Central Grove, with the second most outside of the Grove border, followed by a tie between the North and South Grove. This distribution could be due to the higher number of high-rise residential buildings in the Central Grove, or due to more commercial properties in the Central Grove and outside the borders of the Grove. Next, I wanted to find the average condo, multi-family, and single family home in the Grove. In order to find the average, I had to employ different measures of central valuation. The continuous numeric features often had outliers, so I found the median of value for each of those columns. Then, for the categorical columns I found the most frequently appearing value, or mode.

The results showed that single family homes were generally an individual's primary residence: a single story home, with 3 bedrooms and 2 bathrooms in the South Grove. Single family homes generally sat on a 7500 sq. ft. lot, had a pool, and were about 2400 sq. ft. The average last sale took place in 2009 for an average sale price of about $559,000, but the average market value sits at over $1 million dollars. Considering the lowered housing prices following the 2008 recession, and the increase in real estate values since then, this sharp rise is supported by the context.

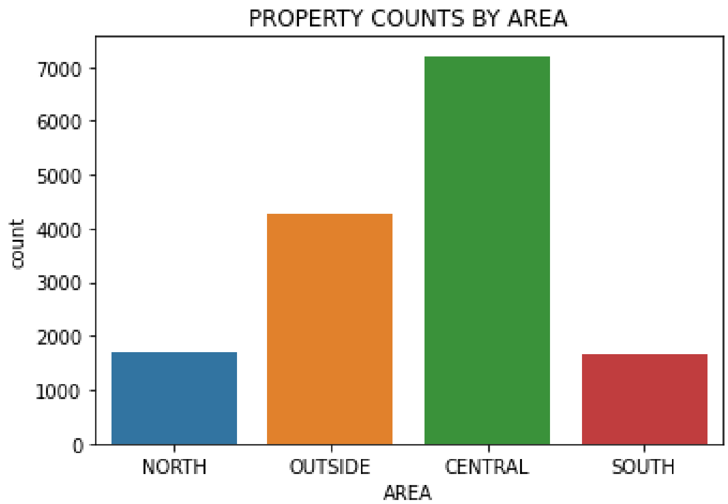

Figure 8. Property counts by area

Studying average condominium and multi-family properties proved to be slightly more challenging, as some of the fields were missing many values. The average condo was not a primary residence and was last sold in 2014 for an average of 430,000. These properties generally were in the Central Grove and had 2 bedrooms, 2 bathrooms, no pool, and 1425 sq. ft. of living space. The average condo was also built more recently than single family homes, with an average effective year built of 1996, as opposed to 1979. Finally, the average multi-family home was also in the Central Grove and owned by a company, rather than an individual. It follows then, considering the most common owners, that most of these properties were not primary residences. The rest of the fields cannot be properly evaluated, as multi-family homes contain features of the entire property and are not separated into the individual residential units.

3.2 Spotlight on Different Areas of the Grove

After studying an overview of the Grove, I focused my analysis on each area inside the Grove. I focused my efforts mainly on single family homes and condos, as these were the most common types of residential properties across all areas. Beginning with the North Grove, the average property was a single family home, last sold in 2010 for $680,000, though the current average market value is $1.3 million. These properties were usually a primary residence, had 3 bedrooms and 2 bathrooms, no pool, and were about 2500 sq. ft. sitting on a lot of around 6000 sq. ft. So, most homes in this area sit on larger lots, with homes taking up about a third of that area. There have been about 300 property sales since 2020 in the North Grove, about the same as the South Grove (Fig 9).

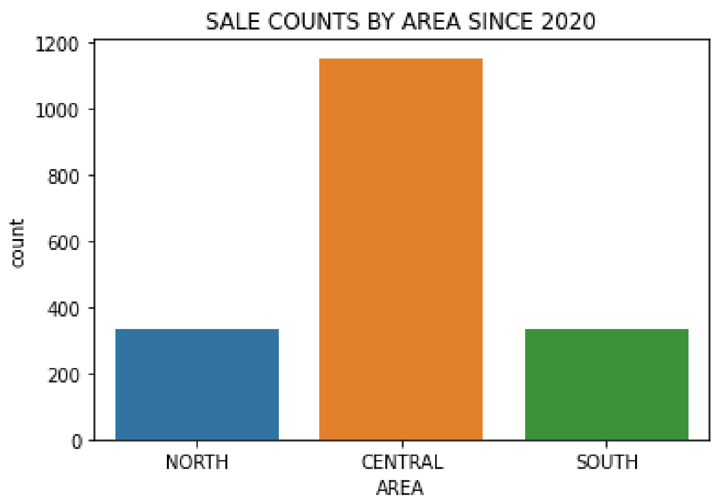

Figure 9. Number of qualified sales in each area of the Grove since 2020.

Moving on to the Central Grove, the average home was a condo, last sold in 2011 for $340,000, with an average current market value of $607,000. A property was likely to have 2 bedrooms and bathrooms, no pool, and an adjusted square footage of 1200 sq. ft., though the average lot size is unknown due to missing values. These properties were also usually not a primary residence. Properties in the Central Grove tended to be of lesser value and have less than half of the square footage of the average home in either the North or South Grove. Perhaps due to the more accessible pricing, this area had the most sales since 2020 with nearly 1200 residential properties purchased. The number of sales since COVID-19 in this area was approximately double the sales in the North and South Grove combined.

Finally, an average property in the South Grove was found to be quite similar to the North Grove. The average residential property in this area was a single family home last sold in 2011 for $765,000, but now has a market value of $1.6 million, more than double the average sale price from only 10 years ago. A South Grove property was likely to be a primary residence, with extra features of both a pool and a patio. These homes had 3 bedrooms, 2 bathrooms, an adjusted square footage of about 2600 on a lot of 9200 sq. ft. Similar to North Grove properties, the buildings took up a third of the lot in general, though South Grove lots tended to be larger. Finally, as previously stated, the number of properties sold in the South Grove since 2020 was the same as for the North Grove at about 300 sales.

3.3 Predictive Models

3.3.a Linear Regression

Once preliminary data analysis and exploration was complete, Professor Bugera and I designed predictive models based on the dataset. The first predictive model was a linear regression model to predict the last qualified sale amount. First, we separated the relevant numerical columns from the relevant categorical columns. Since condos and multi-family homes had many missing values, we focused our model on single family homes. We generated features defining whether a property was in the North or South areas, where a 1 indicated the property was in that area, and 0's in both features implied a property was in the Central area (Fig 10). Additionally, we created a feature containing the year and month of the most recent sale. Finally, we selected properties which were sold from 2020 to the present day for the model dataset.

```
MODELDATA1 = STEP2DATA[STEP2DATA["PROPTYPE"]=="SINGLE"]
MODELDATA1 = MODELDATA1[MODELDATA1["AREA"].isin(["SOUTH","NORTH","CENTRAL"])]
MODELDATA1 = MODELDATA1[CATVARS.union(NUMVARS)]
MODELDATA1["AREA_S"] = np.where(MODELDATA1["AREA"]=="SOUTH",1,0)
MODELDATA1["AREA_N"] = np.where(MODELDATA1["AREA"]=="NORTH",1,0)
MODELDATA1["YYYYMM"] = ((MODELDATA1["LASTQSALEDT"]/10000).astype(int) +
                        ((MODELDATA1["LASTQSALEDT"]/100).astype(int).mod(100)-1)/12)
#=(INT(B1/100)+MOD(B1,100)/12)*100

NUMVARS = set(NUMVARS.union({"AREA_S","AREA_N","YYYYMM"}))
```

Figure 10. Feature generation for linear regression.

With the model dataset prepared, the next step was to find correlations between variables and attempt to build the linear regression model. We created a correlation heatmap to find variables with strong positive correlations, particularly in relation to the last sale amount (Fig 12). Tax assessed features, bed and bath counts, lot size and adjusted square footage indicated strong relationships with the most recent sale amount. After reviewing the variable correlations, we split the model dataset into the target variable of the last sale amount and the remaining features selected for regression. With 635 observations, the train test split was set to create a test size of 20%. The variables that yielded the highest regression score using scikit-learn metrics were lot size, number of bathrooms, adjusted square feet and year of the last sale, returning a training score of 0.613 and a test score of 0.73 (Fig 11). Year had the strongest positive relationship with sale amount, followed by adjusted square feet, and lot size.

```
Run Simple Linear Regression:
Scoring Training Data: 0.6134823690282698
Scoring Test Data: 0.7297237590487657
```

```
Variable  LOTSIZE  has coefficient  30.633764561263163
Variable  BATH  has coefficient  -417813.69203994964
Variable  ADJUSTEDSQFT  has coefficient  1396.8152834255015
Variable  YYYYMM  has coefficient  470043.4895817354
```

Figure 11. Regression scores and feature coefficients

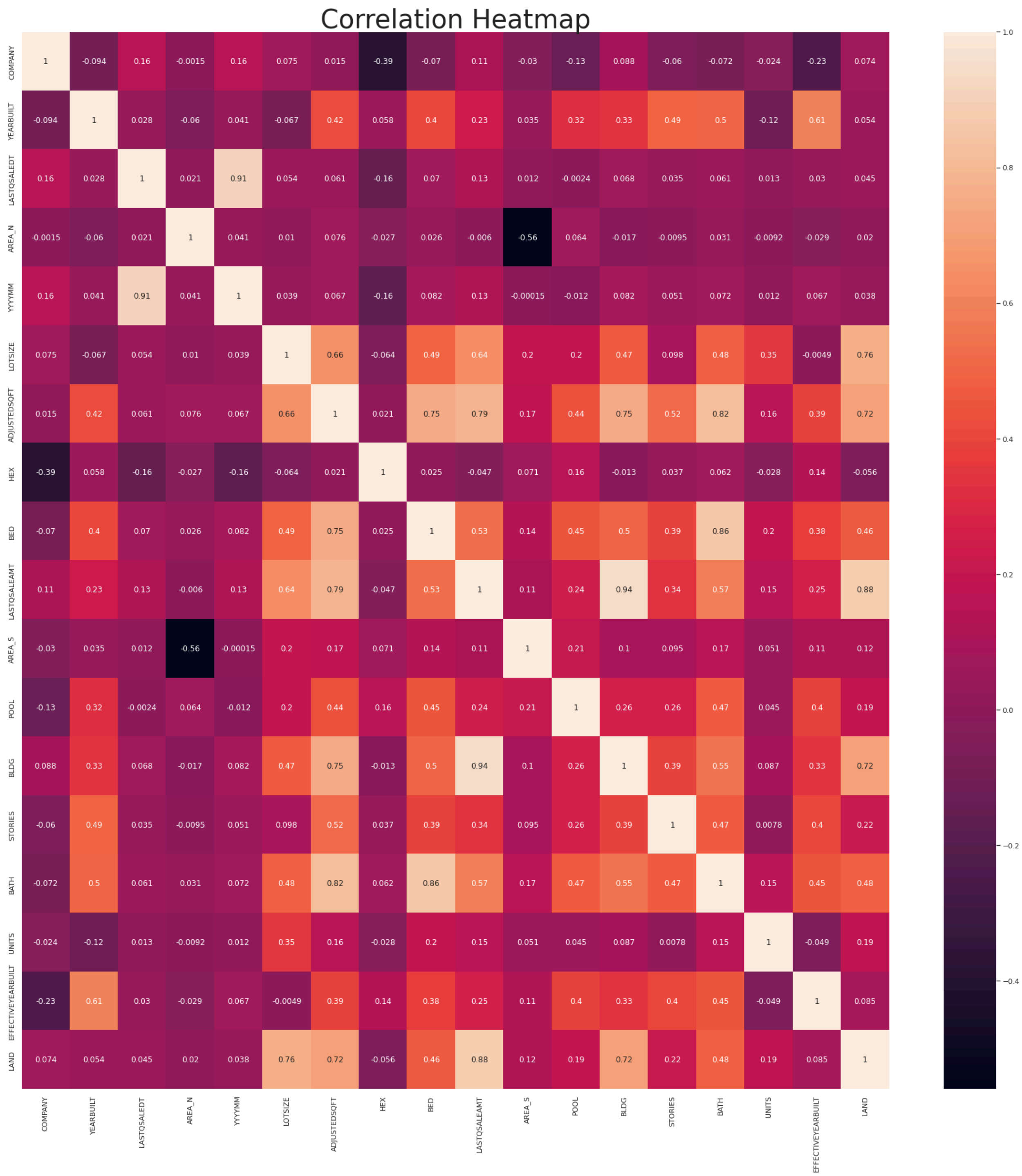

Figure 12. Correlation heatmap of variables in MODEL2DATA dataset.

3.3.b Logistic Regression

The second predictive model we designed was a logistic regression model to predict whether a property was a primary residence or not, as indicated by the HEX code. We used observations from all years and generated a new feature to specify whether each observation had the homestead exemption or not using a 0 or 1. After studying variable relationships using a seaborn template pairplot, the model dataset was separated into the target variable ‘HEXIND’ and the remaining variables. Since there were many observations to input, a test size of 50% was chosen. Using ROC curve metrics, the logistic regression model scored the highest, 0.727, using variables shown in Figure 13 and had relatively more true positives than false positives in the test set (Fig 14). The results were largely inconclusive as to what features most strongly indicated primary residency.

```
MODELVARS = ['LAND','BLDG','ADJUSTEDSQFT', 'BED','LOTSIZE','BATH','STORIES',
             'YEARBUILT', 'EFFECTIVEYEARBUILT', 'POOL', 'COMPANY',
            'LASTQSALEDT','LASTQSALEAMT']
```

Figure 13. Variables that produced the best predictive model for HEX codes.

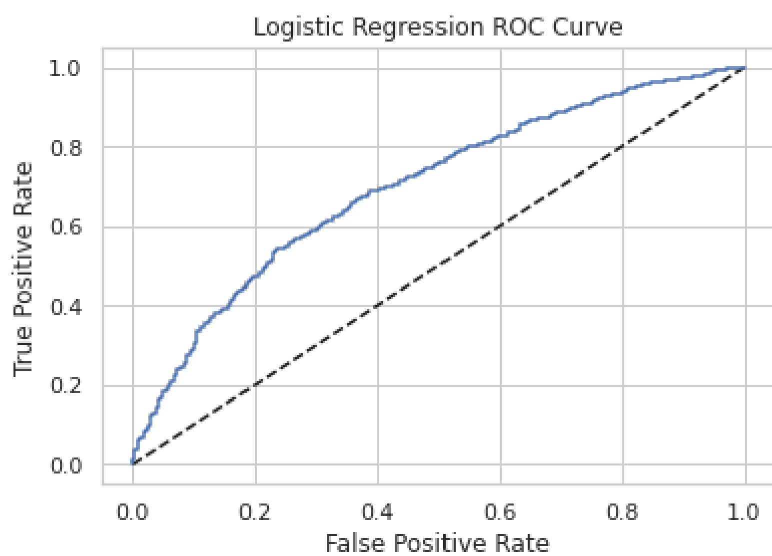

Figure 14. ROC curve of test data true positive to false positive measurements.

4. Conclusion

The results of my study of the real estate portfolio in Coconut Grove yielded responses that align with the changes occurring in the neighborhood today. Across the board, property values have doubled in the last decade. The Central Grove has the most residential properties at the lowest prices and the least number of primary residences, implying that this area will continue to develop and attract new buyers. This observation is also supported by the number of sales in this area since 2020, quadrupling the number of sales attributed to either the North and South Grove. The data analysis was limited by missing values and skewed data due to a vast range of property values and differences in data reporting between property types. The linear regression predictive model provided the insight that year of sale was the biggest indicator of the amount a Coconut Grove home sold for, followed by square footage of the home and size of the lot. The logistic regression to determine which features were the best indicators of primary residence was largely inconclusive. In future studies, the inclusion of historical data as far as sales and development would provide fascinating insights on the trajectory of Coconut Grove as it leaves its era of sleepy Miami enclave and steps into the spotlight of the metropolitan area's continued evolution.

5. References

1. Robertson, L. (2022, July 14). History bulldozed as charming Coconut Grove homes replaced by giant concrete cubes . *The Miami Herald*. Retrieved 2022, from https://www.miamiherald.com/news/local/community/miami-dade/coconut-grove/article263038363.html#storylink=cpy.
2. Smalls, C. I. (2022). ‘We need to know our history.' West Grove is now officially called Little Bahamas . *The Miami Herald*. Retrieved 2022, from https://eedition.miamiherald.com/popovers/dynamic_article_popover.aspx?artguid=2e77c810-e18f-41b1-b765-fb47467d995c.
3. https://www.miamidade.gov/Apps/PA/propertysearch/#/
4. https://www.google.com/maps/place/Coconut+Grove,+Miami,+FL/@25.7296835,-80.2318382,14z/data=!3m1!4b1!4m5!3m4!1s0x88d9b7c4a72958db:0x277e3b9f623100fb!8m2!3d25.7355248!4d-80.2377186