TO: Professor Abbass

FROM: Wanyi (Evelyn) Dong, Shihhung (Angela) Ma, Jenny Shang, Kaiping (Rory) Wang, Yiran (Jenny) Wang

DATE: December 1, 2019

SUBJECT: Data-Driven Real Estate Investing Report

Introduction & Executive Summary

This report serves as a summary of our analysis of data provided by Centro, a real estate asset management company. Our objective is to analyze data on a series of real estate property variables and identify how these variables affect the price per square foot. After a series of analysis, we propose the following recommendations which we will go into depth in the later sections of our report:

a. **Recommendation 1:** Allocate budget to different locations according to changes in 10-year US Treasury Rate and demographic information.
b. **Recommendation 2:** Select properties with amenities that positively contribute to price per sqft.
c. **Recommendation 3:** Invest more into buildings with either mainly four bedrooms or a mix of different units.

We firmly believe that these recommendations will help Centro effectively increase their return from real estate investments.

Problem Statement & Purpose

We were approached by the company Centro, an asset management company that currently focuses on quantitative investing in the real estate sector. Different from traditional real estate investment methodologies, Centro seeks to use data-driven analysis to assist in their asset selection process. The company's goal is to use quantitative data analysis to identify properties that have a high rent growth potential and sufficient market demand in order to realize a desirable return on investment (ROI). These properties should generate a stable and increasing stream of net operating income.

Our group's responsibilities are to derive insights and recommendations by analyzing the real estate data provided by Centro to help them achieve their goal. We conducted exploratory data analysis and regression analysis on property and sales data for apartments in the states of Florida, Georgia, North Carolina, and Texas. Our analysis focused on revealing relationships between price per square foot and some key variables that impact real estate valuation. The key variables that we looked at include:

- Building age
- Features and amenities
- Star rating
- Unit mix
- Location
- Vacancy
- Demographics

In the next sections of our report, we will dive into details about our data cleaning method, our analysis and visualization, and our recommendations.

Methodology

Our initial data was from two raw datasets. The first dataset contains information on 20,363 properties (rows) across 4 States (Florida, Georgia, North Carolina, and Texas) and 212 attributes (columns). The second dataset contains information on the sales of these properties. This dataset has 24,527 records (rows) and 200 attributes (columns). The properties have their own unique identifiers that are common across the two datasets, namely "SCP," "SCSitus_NumNam," and "SCAPN." Since the datasets were in an Excel workbook, we saved them into csv files with the names "Property_Data.csv" and "Sales_Data.csv" to be compatible for analysis in Python.

Our data cleaning started with the property information file. Since we only wanted to focus on certain key variables, we did not need to include all 212 columns. We first created another data frame and updated it by extracting only the columns that are relevant to our analysis. Next, we split the new dataset into four subsets based on which State the property is located. For backup purposes, these subsets were each converted into a separate csv file. Similarly, we created a new sales dataset by extracting 16 columns from the raw sales dataset that we think would be useful for this project, and saved it into a separate csv file for backup.

After extracting the key columns, we believe the dataset can be better analyzed if we combine property and sales data together. To do this, we merged the property files for each state with the sales data by using "SCAPN" as a common identifier. While merging the data, we used left join instead of inner join. We used the property dataset as a base to preserve as much property data points as possible to explore possible relationships between the variables. In addition, we wanted to explore the effect of external factors outside the data provided. Therefore, we obtained the daily U.S. 10 Year Treasury Rate and incorporated this value by looking up the sale date of each property. We chose the U.S. 10 Year Treasury Rate because real estate is a long-term investment and thus, we believe a long-term rate would be more appropriate. The combined data for each state is exported into four separate csv files.

Key Business Insights & Recommendations

After exploring the data provided by Centro and conducting data analysis, we observed some business insights and propose the corresponding recommendations to help Centro better understand how the business performs and how to react to the current trends.

Business Insight 1: Florida's building price is sensitive to changes in the 10-year Treasury Rate and has a positive correlation with the 10-year Treasury Rate while the building prices of the other three states are more robust to changes in the 10-year Treasury Rate.
Recommendation 1: We would recommend Centro to allocate more budgets in Georgia, North Carolina, and Texas when the 10-year Treasury Rate is expected to decrease or stay the same, and invest heavily in Florida when the 10-year Treasury Rate is expected to increase. In addition, Centro should use demographic information as as an initial selection for high-value regions, then perform further investigations.

Business Insight 2: Fitness center, business center, laundry facilities, and property manager on site are the most common amenities across the four states. Among these amenities, fitness center and business center have a significantly positive relationship with price per sqft across the states.
Recommendation 2: We would recommend Centro to focus on buildings with fitness centers and business centers as amenities. Centro should also conduct surveys to investigate the reasons behind the decrease in price per sqft caused by other amenities which are commonly supposed to add value to buildings.

Business Insight 3: There is a higher average price per sqft for buildings that are either mainly four bedrooms or a mix of different units, depending on the state, but the standard deviations for buildings with mixed units is much greater.

Recommendation 3: We would recommend Centro to invest in buildings with mainly four-bedroom units in Texas, Georgia and North Carolina, and buildings with a mixture of units in Florida.

Analysis & Visualization

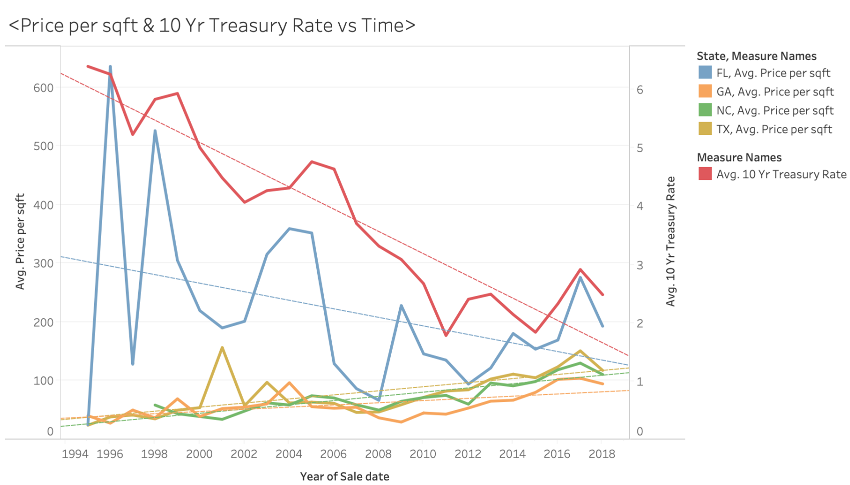
Geographic and demographic variables

Geographic and demographic information are important for real estate companies to determine building price per sqft. To understand how these features could influence building prices, we conducted our analysis in two steps: first, we created a dual-axis line chart to analyze how price per sqft changed across time for buildings in different states, and the relationship between the price per sqft and the 10-year Treasury Rate; second, we created eight heatmaps to analyze how demographic information, especially median household income and per capita income at zip code level, could influence the price per sqft in each region.

To compare the changes in housing prices of each state, we set the starting date as 1/1/1996 when all four states have available data. In addition, to eliminate the influence of some extreme and impossible values in price per sqft, we only included non-null values which are less than or equal to 50,000.

From the trendline of building prices of each state and the 10-year Treasury Rate across time, we found two patterns (see Figure 1 below). First, Florida's building price shows a downward trend since 1996, while building prices of the other three states (Georgia, North Carolina, Texas) show an upward trend. Second, Florida's building price is sensitive to changes in the 10-year Treasury Rate and has a positive correlation with the 10-year Treasury Rate; however, the building prices of the other three states are more robust to changes in the 10-year Treasury Rate. As a result, we recommend Centro to increase the budget allocation in Georgia, North Carolina, and Texas when the 10-year Treasury Rate is expected to decrease or stay the same, and invest heavily in Florida when the 10-year Treasury Rate is expected to increase dramatically.

Figure 1



The trends of Avg. Price per sqft and Avg. 10 Yr Treasury Rate for Sale date Year. Color shows details about Avg. Price per sqft and Avg. 10 Yr Treasury Rate. For pane Average of Price per sqft: Color shows details about State, Avg. Price per sqft and Avg. 10 Yr Treasury Rate. The data is filtered on Price per sqft and Sale date. The Price per sqft filter includes values less than or equal to 50000. The Sale date filter includes dates on or after 1/1/1996.

To understand how demographic features influence a building's price, we focused on household income and per capita income at zip code level, created heatmaps for the four states, and identified some areas with high growth potential (see Figure 2 in Appendix). The color of the map represents the amount of income. The size of the circle represents the average price per sqft in each zip code region. The color of the circle represents the number of buildings Centro owns.

By analyzing heatmaps, we found that the building price per sqft in each region is positively correlated with the median household income and the per capita income. Using this information, we could identify some areas that Centro doesn't invest heavily in and have the potential for a higher price. For example, we found that the city Ector in Texas has a building located in a high-income region and yields an average price per sqft of $4,048. However, Centro only owns one building in this region. As a result, we would recommend Centro to use the demographic information as an initial selection filter, find counties that have similar demographic groups as high-value counties, and conduct further investigations on building in these regions.

Amenities

Amenity tends to be an important contribution factor in determining buildings' values, and it is important for Centro to know what amenities people value more in order to increase profitability. Therefore, we decided to create word clouds for each of the four states to better understand what specific amenities are included in most buildings and how these main amenities affect their price per sqft.

Before performing the word cloud analysis, we further cleaned the data set to fit our needs. We only selected two columns, amenities and price per sqft for this analysis. For each data set by state, we first removed rows without a price per sqft value. Additionally, we also removed rows that have a null value in Amenities since those records cannot contribute any relevant information to the word clouds. The last step of data cleaning for this analysis is to remove outliers. We used interquartile range (IQR) to remove outliers and only included data points between Q1-1.5*IQR and Q3+1.5*IQR. After data cleaning was done, all prices per sqft in each state shows a smooth curve (see Figure 3 in Appendix).

Word clouds are visuals of words that give greater prominence to more frequently appeared words. That being said, bigger words simply mean that these words appear more often than the other. We used the word cloud package in Python to generate word clouds. First, we did not set any stop words, and the result showed all words that appeared in the Amenities column (see Figure 4 in Appendix). We then created a second word cloud for each state to show only the top ten words that appeared most frequently for better visualization (see Figure 5 in Appendix).

According to Figure 5, the most common amenities in all states are fitness center, property manager on site, laundry facilities and business center. Other common amenities include package service, picnic area and maintenance. Since the amenities across states all include fitness center, property manager on site, laundry facilities and business center, we selected these four to further analyze their impacts on their prices per sqft.

To further understand the relationship between amenities and price per sqft, we decided to show both average prices per sqft for buildings that do and do not have the four amenities at the same time for Centro to compare and visualize the results. Bar graphs in Figure 6 (see Figure 6 in Appendix) show the comparison between average price per sqft for buildings that contain and do not contain certain amenities. We can see that the average price per sqft across the four states for buildings that have business centers is much higher than buildings that do not have business centers and the average price per sqft across the four states for buildings that have fitness centers is also much higher than buildings that do not have fitness centers. Therefore, it is safe to assume that there is a positive relationship between having a fitness center and a business center, and price per sqft. However, this is not the case for laundry facilities and property manager. The average price per sqft for buildings that have these two amenities is lower or ambiguous compared to that of buildings that have the amenities. These two amenities do not appear to have positive relationships, thus no impacts for profitability.

For more detailed comparison, we calculated the dollar and percentage difference between with and without amenities in all states (see Figure 7). Fitness center increases price per sqft by $38 on average, and business center increases price per sqft by $26 on average. We can also see that Texas, North Carolina, and Georgia have very high percentages of increases with fitness center, especially North Carolina, where the average price per sqft nearly doubled. These insights can provide useful information for Centro to make an informed business decision when it comes to where to invest for buildings with a certain amenity.

As a conclusion, we would like to recommend Centro to focus more on buildings that have business center and fitness center as amenities as they are shown to increase values in the buildings.

Unit Mix

One aspect of the different buildings that may affect their price per sqft is the different unit mixes. In this analysis, the unit mix indicates the different type of units that are in one given building. The different types that are analyzed are the units that are mainly one bedroom, mainly two bedroom, mainly three bedroom, mainly four bedroom, mainly studio, and mixed. Any building that has more than 45% of its units be of one type of layout will then be classified as mainly x bedrooms. If no category is over 45% then it is classified as a mixed building.

The percentage of 45% was utilized because even though 50% would be a better indication if a building was more than one type, it would mean that too many buildings fall into the mixed category and it would be difficult to distinguish the differences between buildings in the mixed category. By setting the level at 45%, it still allows the building type to be mainly one type, without placing too many buildings into the mixed category.

To prevent outliers from distorting the data, the dataset was limited so that only rows with price per square foot that were greater than zero and less than 2,000 were incorporated. Additionally, to better see the different boxplots, each graph was limited to a price per sqft of $500. Boxplots for all four states: Florida, Texas, North Carolina, and Georgia were created to emphasize the differences between different unit mixes between states (see Figure 8 in Appendix).

Georgia, Texas, and North Carolina all have mainly four-bedroom buildings as the highest price per sqft, whereas in Florida the mixed unit buildings dominate. While the standard deviation for mixed buildings is a lot high in Florida than any other state (see Figure 9 in Appendix), it has a higher average price per square foot by almost double the other unit mixes, indicating that focusing on the high-priced mixed buildings would be efficient in Florida. Therefore, all states should focus on buildings with mainly four bedroom units, other than Florida where a mixture of units is more profitable.

Joint Relationships

We did individual analysis on how 10-year Treasury Rate, amenities, and the unit mixes affect a building's price per sqft, and now it is important for Centro to know the combination effect of all the factors to determine buildings' values. Therefore, we decided to create scatterplots for each of the four states to better understand the joint relationships and price distributions.

Before performing the joint relationships analysis, we further cleaned the data set. We combined the columns of 10-year Treasury Rate, amenities, unit mixes and price per square foot for this analysis. For each data set by state, we first removed rows that have a null value in price per square foot, since the price per square foot is what we want to determine. Secondly, we counted the number of keywords in amenities column of each listing. Additionally, to prevent outliers from distorting the data, the dataset was limited so that only rows with 10-year Treasury Rate that were greater than zero and less than 8 were incorporated. Scatterplots for all four states: Florida, Texas, North Carolina, and Georgia were created to explain the combined effect from all the factors between states.

Pair plots are visuals of joint relationships between each variable, while the color of the plots represent the unit mix of the building. We used the seaborn package in Python to generate colorful pair plots.
By analyzing pair plots in Figure 10 (see Figure 10 in Appendix), we found the highest price per sqft in each region happened when 10-year Treasury Rate is around 2-3. As the rate goes higher than that, the price per sqft is negatively correlated with the 10-year Treasury Rate. In addition, buildings with equal to or more than one amenity have relatively higher price per sqft in each region. Last but not least, we can see buildings in Florida with mixed units have the highest price, while in North Carolina, buildings with mainly four units have the highest price. Using this information, we could identify some indicators which Centro could take a look at when choosing the location and time of future investments.

Regression Model

In order to understand how the factors other than the ones mentioned above affect a building's price per sqft, and how they are different among states, we decided to create regression models for each state to better understand the contribution of each factor regarding the sale price of the buildings.

Before creating the regression models, we further cleaned the data set to fit our needs. We first removed the columns with highly correlated variables to avoid collinearity (see Figure 11 in Appendix). Secondly, we chose only numeric variables since they are more meaningful in this case. After the data set is cleaned, we used the linear model package in R to generate linear regression models for all four states.

By analyzing the results of regression models in Figure 12 (see Figure 12 in Appendix), we first found that the number of parking spaces per unit has a negative correlation in a high significance level with the price per sqft in Florida, Georgia and North Carolina. Secondly, the year the building was built has a high positive correlation in a high significance level with the price in Georgia, North Carolina and Texas. It looks like the later the building was built, the higher the price per sqft is. Looking more deeply into each individual model, we can see that in Florida, the land area of a building has a positive correlation in a high significance level with the price. In addition, the price per sqft in North Carolina is highly related with the building's floor area ratio. In Georgia, the price per sqft is highly related with both the building's star rating and the land area, while in Texas the price is more related to the number of floors.

Conclusion

Through conducting exploratory data analysis and regression analysis, three main recommendations were created to help Centro increase profits. By doing an in-depth dive on the demographics of each state, creating word clouds on most critical amenities, comparing different unit mixes of the states, and through analyzing joint relationships, we were also able to create a regression model. This regression model highlights other aspects that Centro can take into consideration to potentially further increase their profit.

To conclude, the three recommendations are:

a. Increase the budget allocation in Georgia, North Carolina, and Texas when the 10-year Treasury Rate is expected to decrease or stay the same, and invest heavily in Florida when the 10-year Treasury Rate is expected to increase dramatically.
b. Select properties that contain amenities that positively contribute to price per sqft.
c. Invest more into buildings that are mainly four bedrooms in Georgia, North Carolina, and Texas. Invest in buildings that have a mixture of units in Florida.

Future Work

We would also recommend Centro to continue recording what amenities are included in each building for more data and credible analysis in the future since there are a lot of null values in this analysis. Centro should also look further into the amenities that do not increase the price per sqft such as laundry facilities and property manager on site and investigate the reasons behind them to maximize potential profitability.

In regard to unit mixes, Centro should continue to record the different percentages of each unit type for each building. Additionally, further analysis can be conducted to see if changing the unit mix of a specific building would significantly change the profit from that building.

Centro can also consider incorporating additional external variables to select for optimal locations in various states. One example of an external variable would be the crime rate, which we believe could have a high correlation with property prices and overall profits for Centro. Other variables such as the schools nearby, the number of grocery stores or coffee chains in the area, or distance to the nearest bus stop could also analyzed.

Finally, we recommend Centro to use machine learning to select locations in the future. Automating the building of selection process through algorithms would make Centro's goal of finding better locations more efficient. By going another level higher than just using data analysis to select locations, Centro will be able to better distinguish itself from other asset management companies.

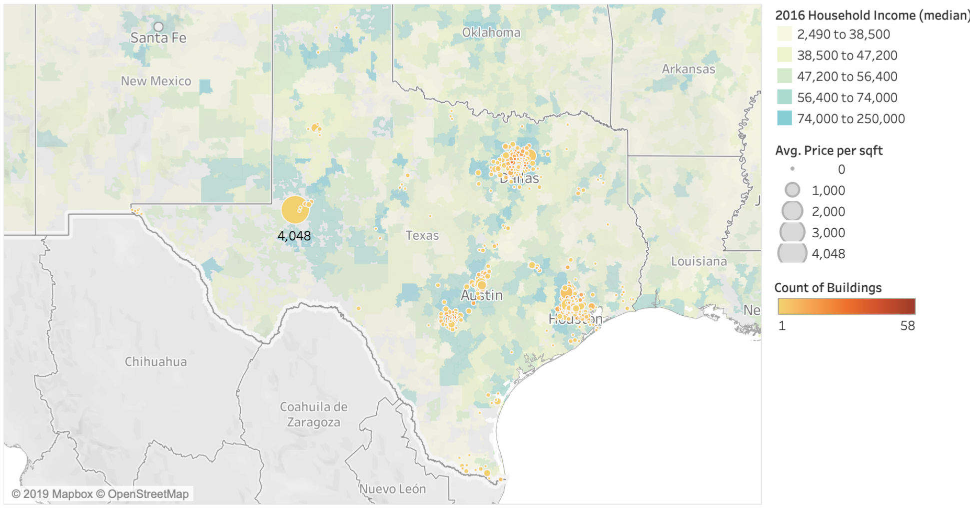
Team Contribution

Name	Contribution
Evelyn Dong	Recommendation 1 - geographic and demographic (heatmaps)
Angela Ma	Recommendation 2 – amenities (word clouds, bar graphs)
Jenny Shang	Data cleaning, methodology, problem statement, correlation matrix
Rory Wang	Joint relationships analysis (pair plot graphs), regression model
Jenny Wang	Recommendation 3 – unit mix, (bar graphs, boxplots)

Appendix

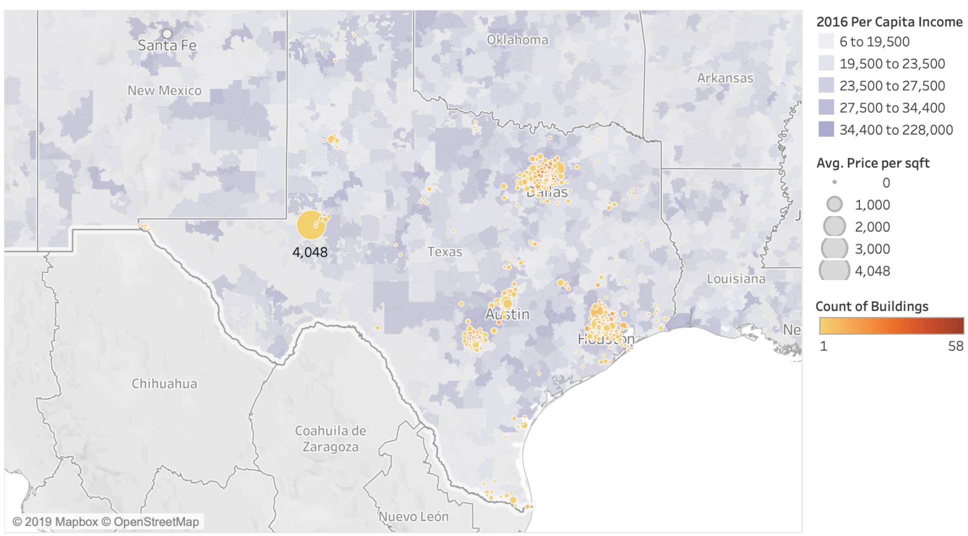
Figure 2: Heatmaps with Geographic Information for Four States

<Price per sqft vs Household Income (median) - TX>



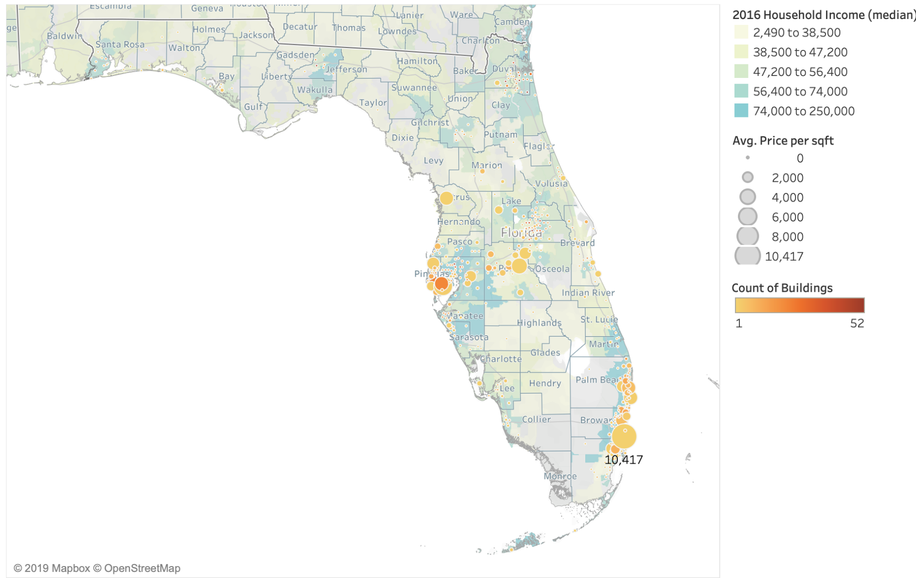
Map based on Longitude (generated) and Latitude (generated). Color shows count of Price per sqft. Size shows average of Price per sqft. The marks are labeled by average of Price per sqft. Details are shown for Zip code. Map coloring shows 2016 Household Income (median) by Zip Code. The data is filtered on Price per sqft and Sale date. The Price per sqft filter includes values less than or equal to 50000. The Sale date filter includes dates on or after 1/1/1996.

<Price per sqft vs Per Capita Income - TX>



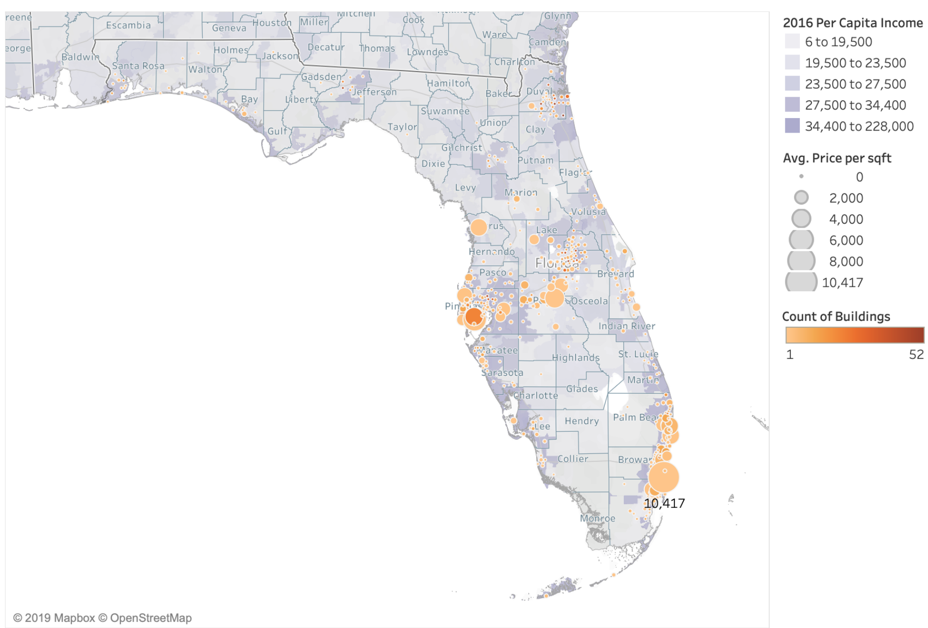
Map based on Longitude (generated) and Latitude (generated). Color shows count of Price per sqft. Size shows average of Price per sqft. The marks are labeled by average of Price per sqft. Details are shown for Zip code. Map coloring shows 2016 Per Capita Income by Zip Code. The data is filtered on Price per sqft and Sale date. The Price per sqft filter includes values less than or equal to 50000. The Sale date filter includes dates on or after 1/1/1996.

<Price per sqft vs Household Income (median) - FL>



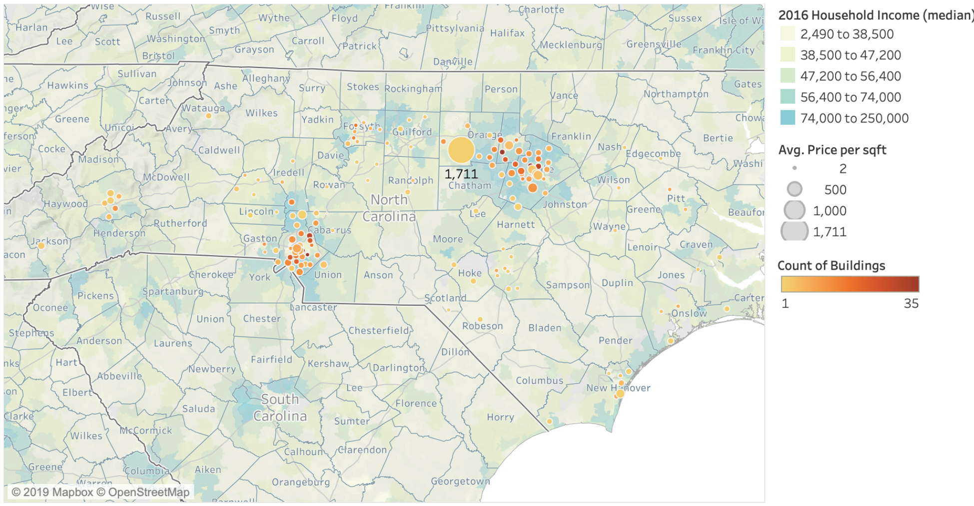
Map based on Longitude (generated) and Latitude (generated). Color shows count of Price per sqft. Size shows average of Price per sqft. The marks are labeled by average of Price per sqft. Details are shown for Zip code. Map coloring shows 2016 Household Income (median) by Zip Code. The data is filtered on Sale date and Price per sqft. The Sale date filter keeps non-Null values only. The Price per sqft filter includes values less than or equal to 50000.

<Price per sqft vs Per Capita Income - FL>



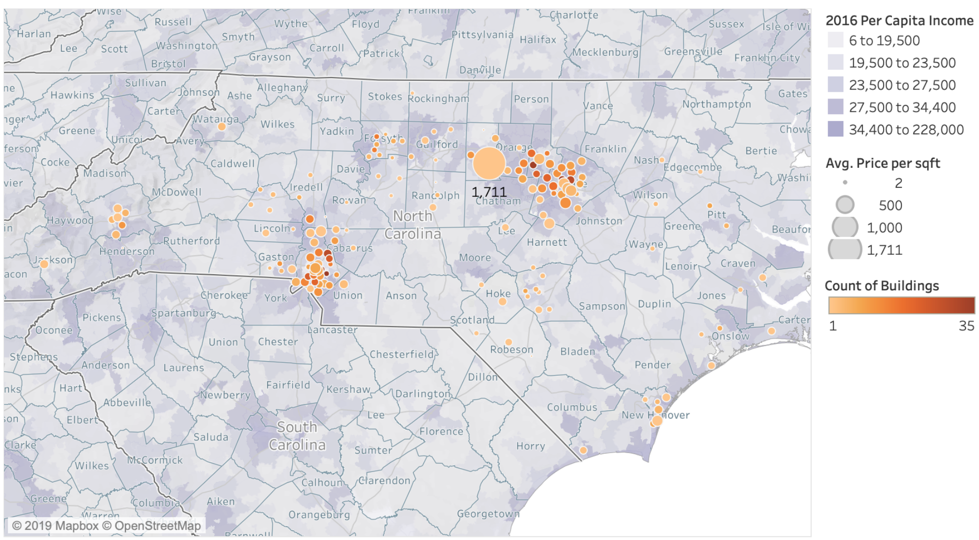
Map based on Longitude (generated) and Latitude (generated). Color shows count of Price per sqft. Size shows average of Price per sqft. The marks are labeled by average of Price per sqft. Details are shown for Zip code. Map coloring shows 2016 Per Capita Income by Zip Code. The data is filtered on Sale date and Price per sqft. The Sale date filter keeps non-Null values only. The Price per sqft filter includes values less than or equal to 50000.

<Price per sqft vs Household Income (median) - NC>



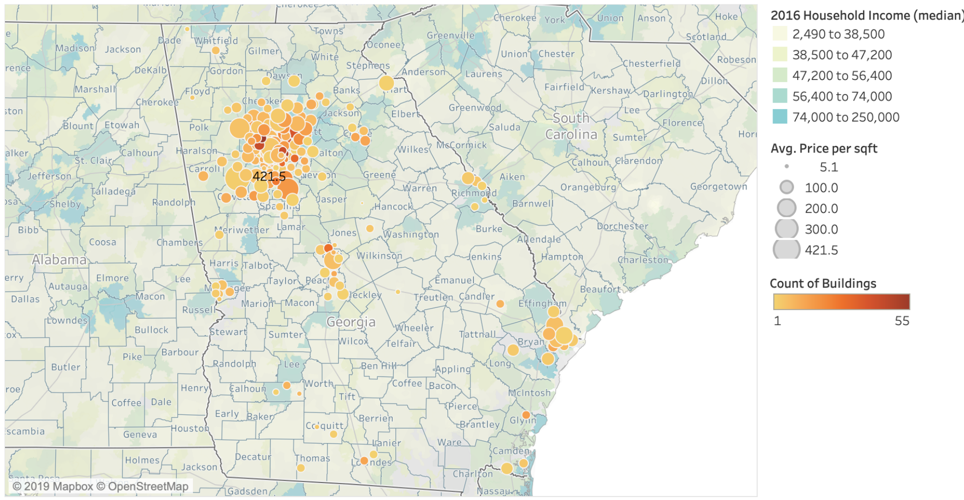
Map based on Longitude (generated) and Latitude (generated). Color shows count of Price per sqft. Size shows average of Price per sqft. The marks are labeled by average of Price per sqft. Details are shown for Zip code1. Map coloring shows 2016 Household Income (median) by Zip Code. The data is filtered on Sale date and Price per sqft. The Sale date filter keeps non-Null values only. The Price per sqft filter includes values less than or equal to 50000.

<Price per sqft vs Per Capita Income - NC>



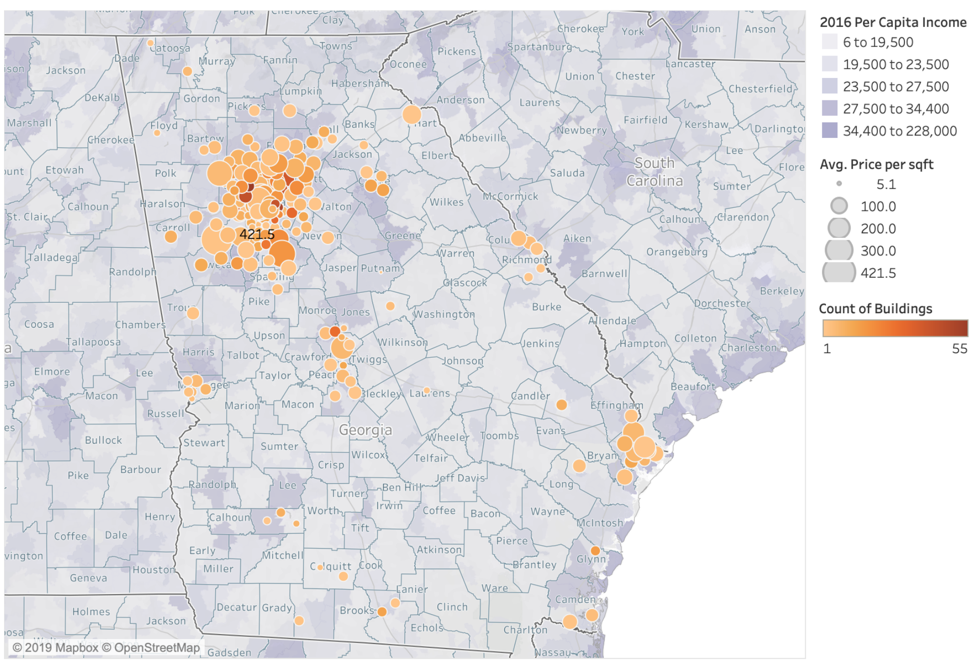
Map based on Longitude (generated) and Latitude (generated). Color shows count of Price per sqft. Size shows average of Price per sqft. The marks are labeled by average of Price per sqft. Details are shown for Zip code1. Map coloring shows 2016 Per Capita Income by Zip Code. The data is filtered on Sale date and Price per sqft. The Sale date filter keeps non-Null values only. The Price per sqft filter includes values less than or equal to 50000.

<Price per sqft vs Household Income (median) - GA>



Map based on Longitude (generated) and Latitude (generated). Color shows count of Price per sqft. Size shows average of Price per sqft. The marks are labeled by average of Price per sqft. Details are shown for Zip code1. Map coloring shows 2016 Household Income (median) by Zip Code. The data is filtered on Sale date and Price per sqft. The Sale date filter keeps non-Null values only. The Price per sqft filter includes values less than or equal to 50000.

<Price per sqft vs Per Capita Income - GA>



Map based on Longitude (generated) and Latitude (generated). Color shows count of Price per sqft. Size shows average of Price per sqft. The marks are labeled by average of Price per sqft. Details are shown for Zip code1. Map coloring shows 2016 Per Capita Income by Zip Code. The data is filtered on Sale date and Price per sqft. The Sale date filter keeps non-Null values only. The Price per sqft filter includes values less than or equal to 50000.

Figure 3: Data Cleaning for Four States

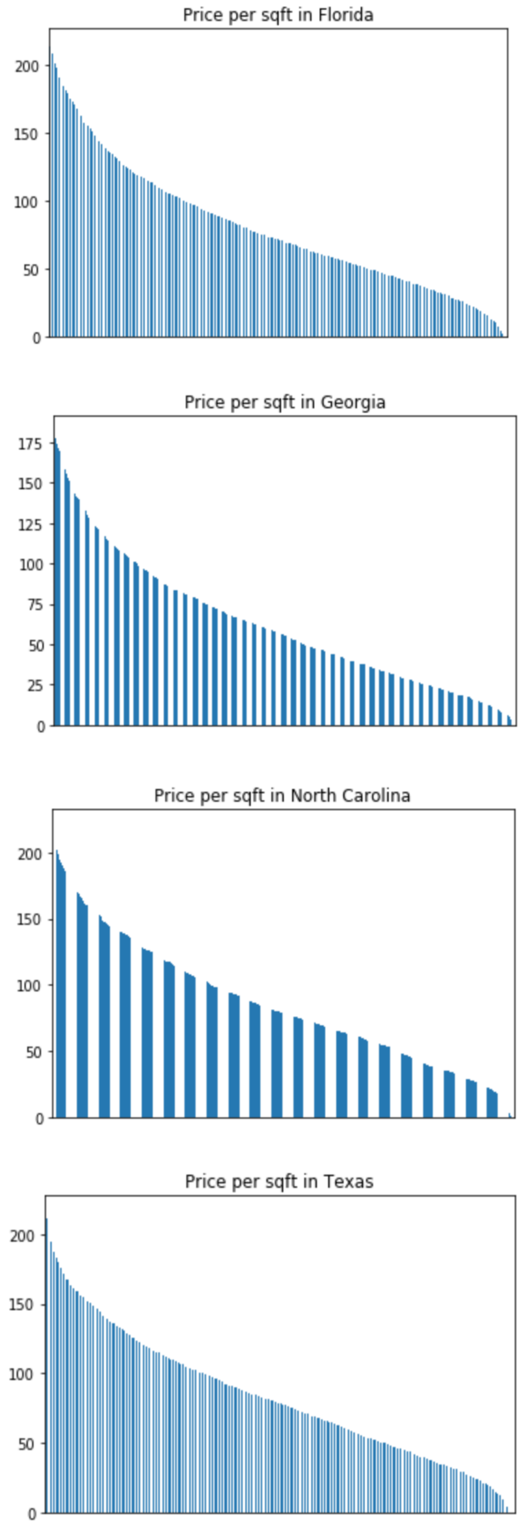


Figure 4: First Word Cloud for Four States

First Word Cloud for Florida

First Word Cloud for Texas

First Word Cloud for North Carolina

First Word Cloud for Georgia

Figure 5: Second Word Cloud for Four States

Second Word Cloud for Florida

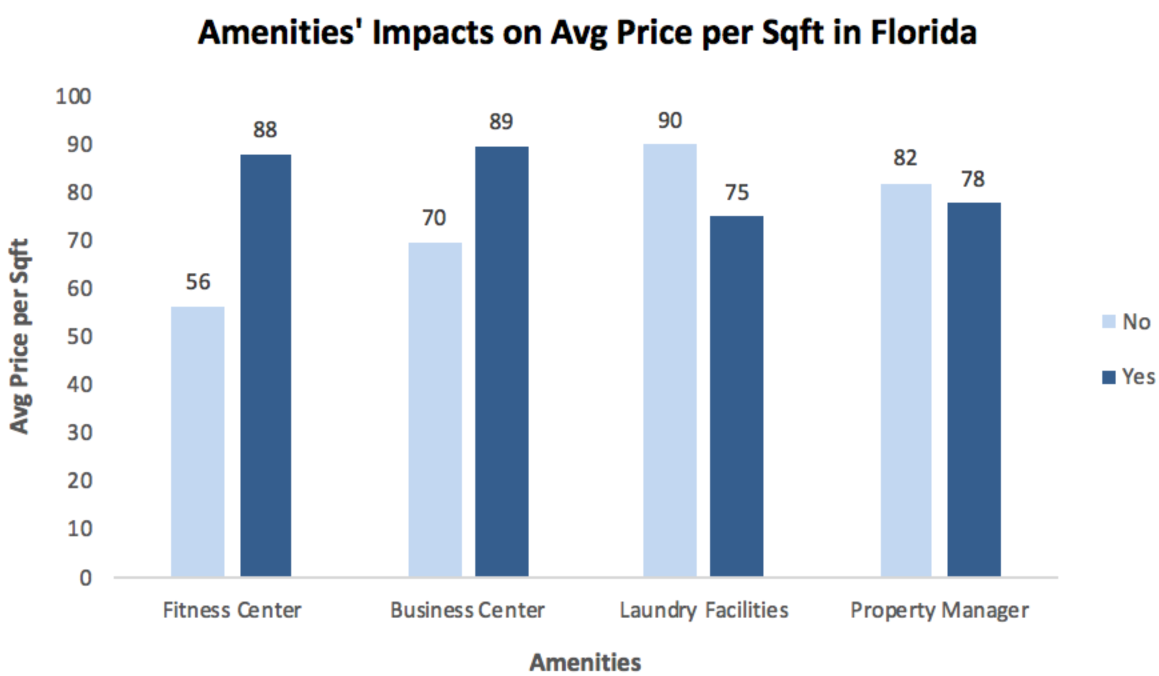
Second Word Cloud for Texas

Second Word Cloud for North Carolina

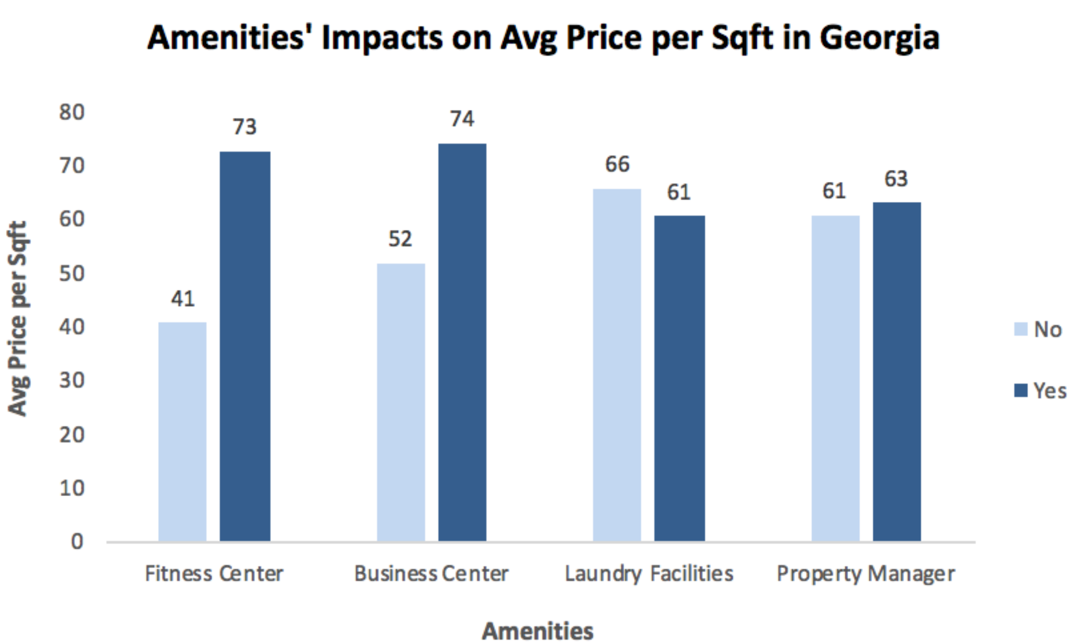
Second Word Cloud for Georgia

Figure 6: Average Price per sqft for Buildings that Contain and Not Contain Certain Amenities

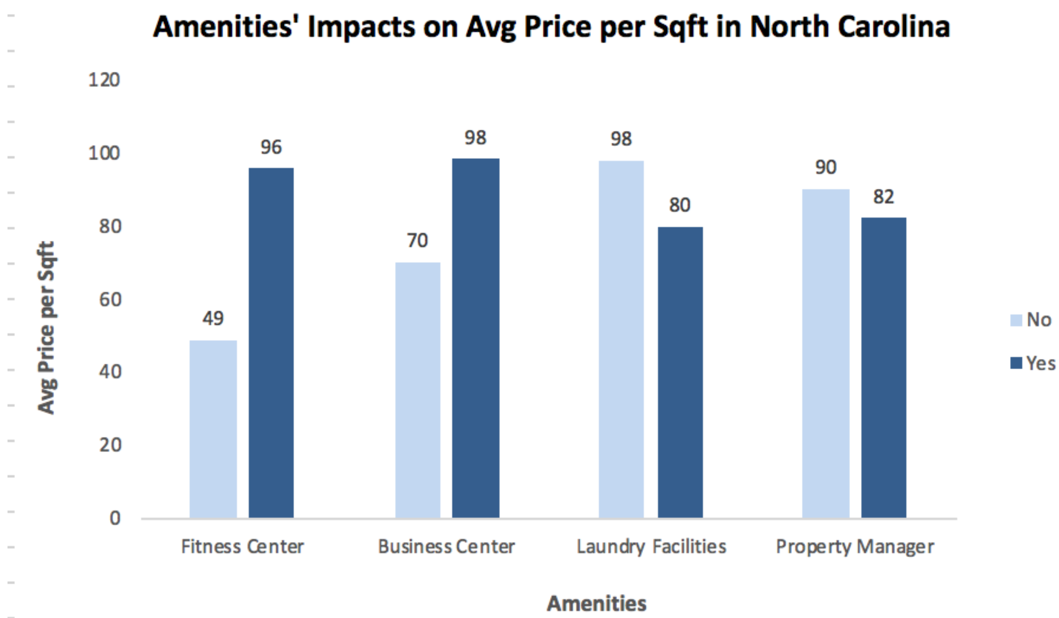
Florida



Georgia



North Carolina



Texas

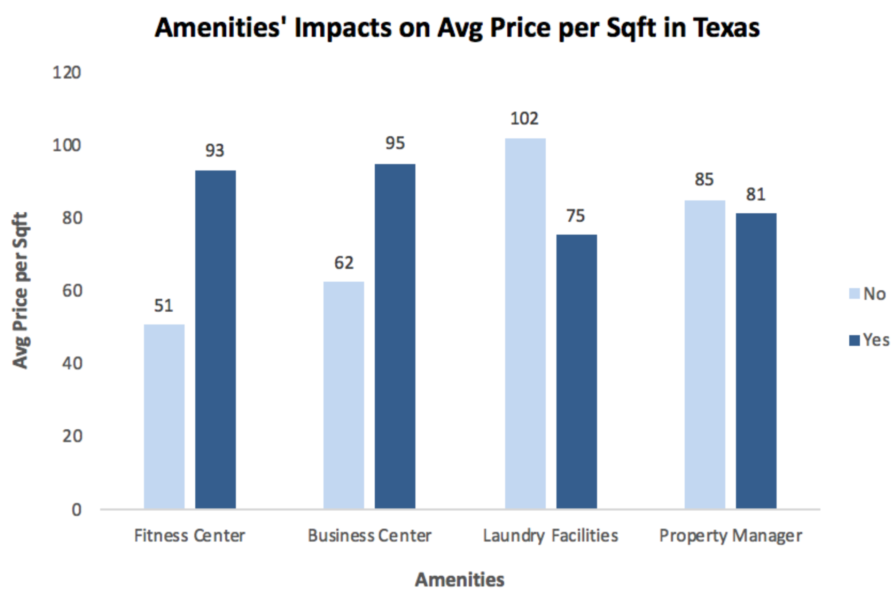
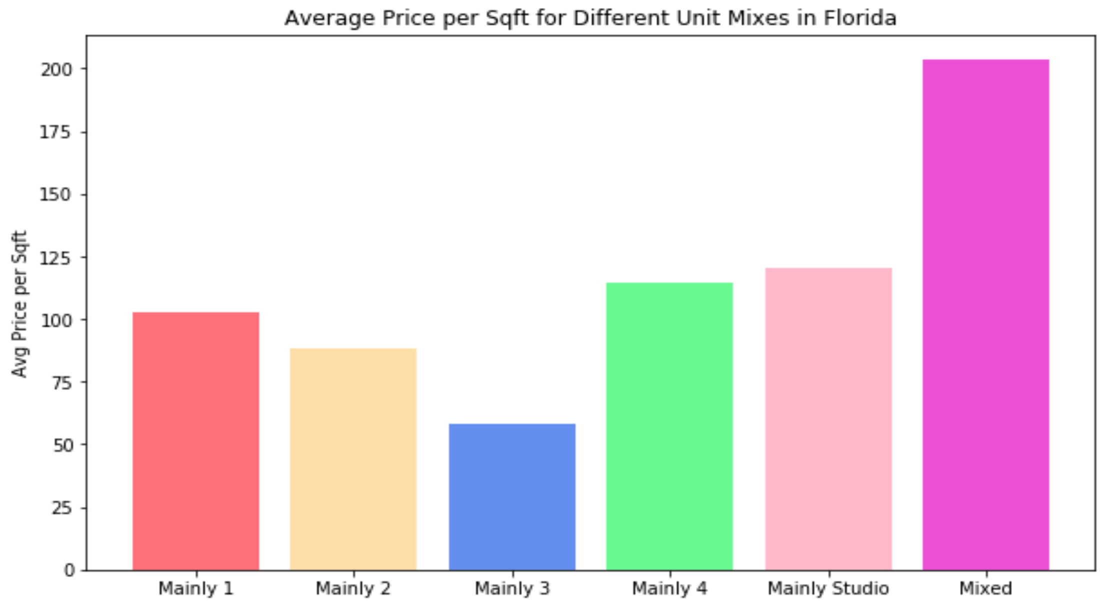
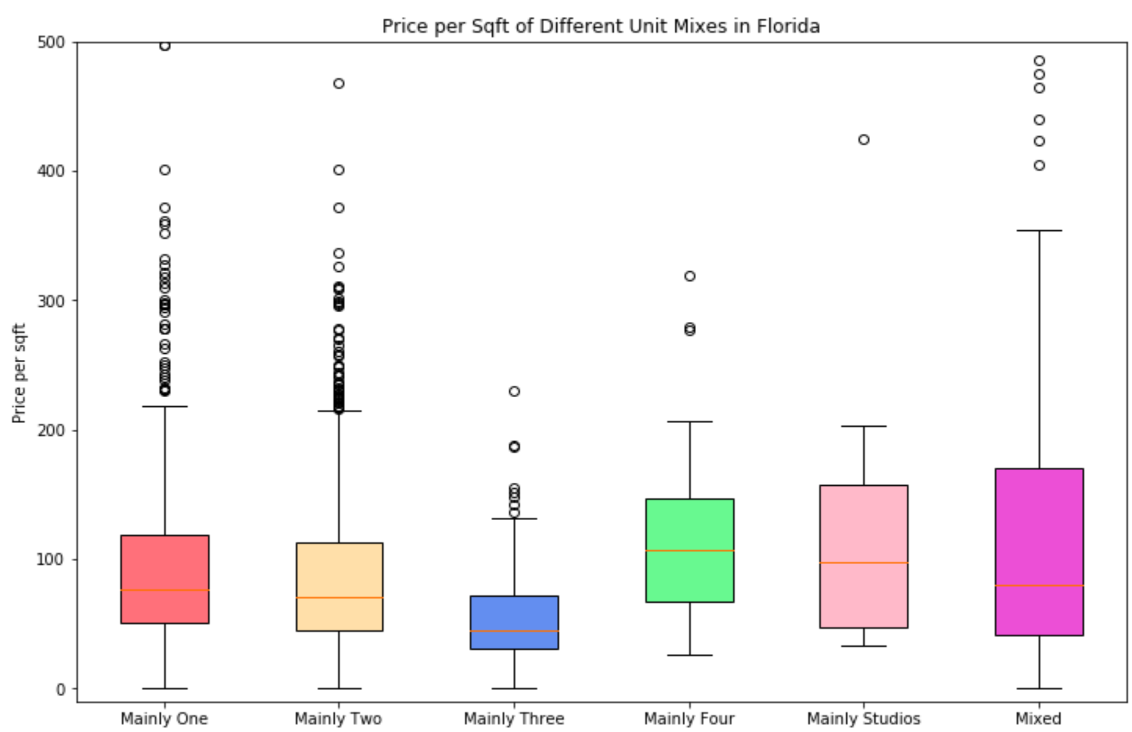


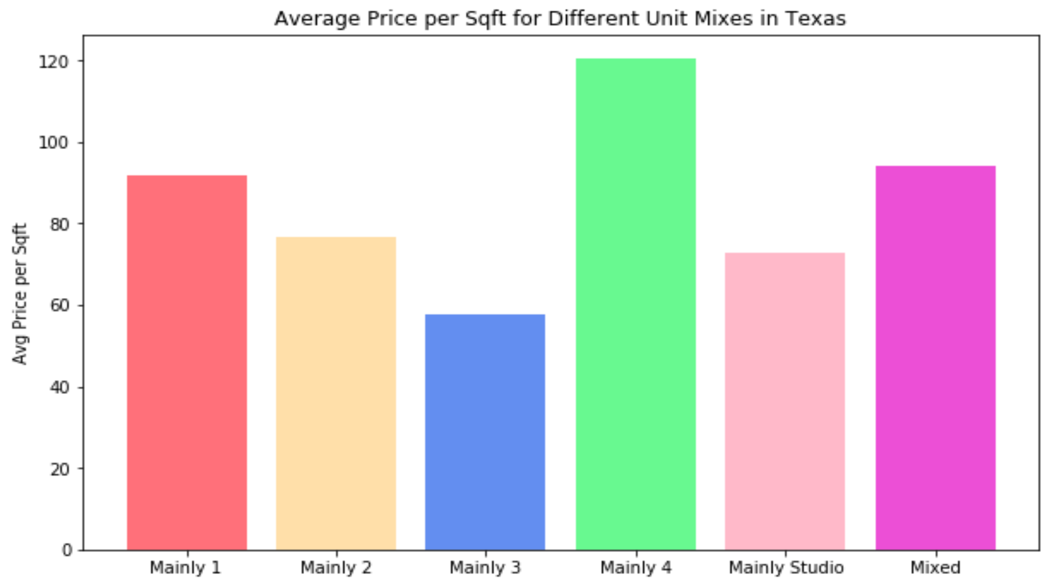
Figure 7: Difference between with and without Amenities in All States

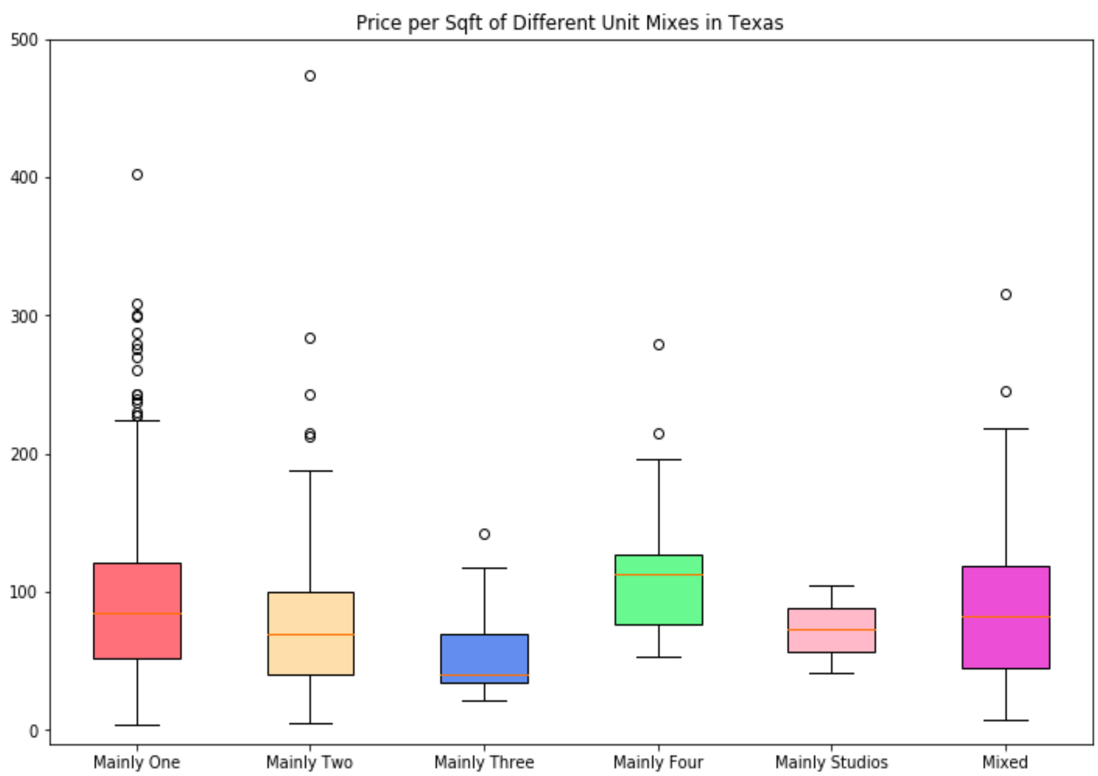
	$Difference with/without Amenities				% Difference with/without Amenities			
	Fitness Center	Business Center	Laundry Facilities	Property Manager	Fitness Center	Business Center	Laundry Facilities	Property Manager
Florida	31	20	-15	-4	56%	29%	-17%	-5%
Texas	43	33	-27	-4	84%	53%	-26%	-5%
North Carolina	47	28	-18	-8	97%	40%	-19%	-9%
Georgia	32	22	-5	2	78%	43%	-8%	4%
Average	38	26	-16	-4				

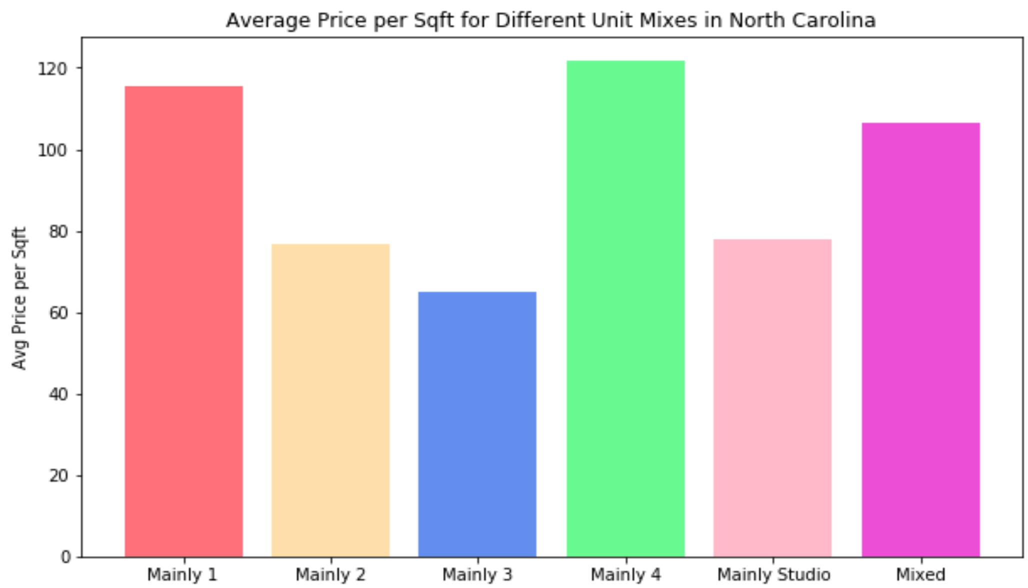
Figure 8: Average Price per sqft of Different Unit Mixes

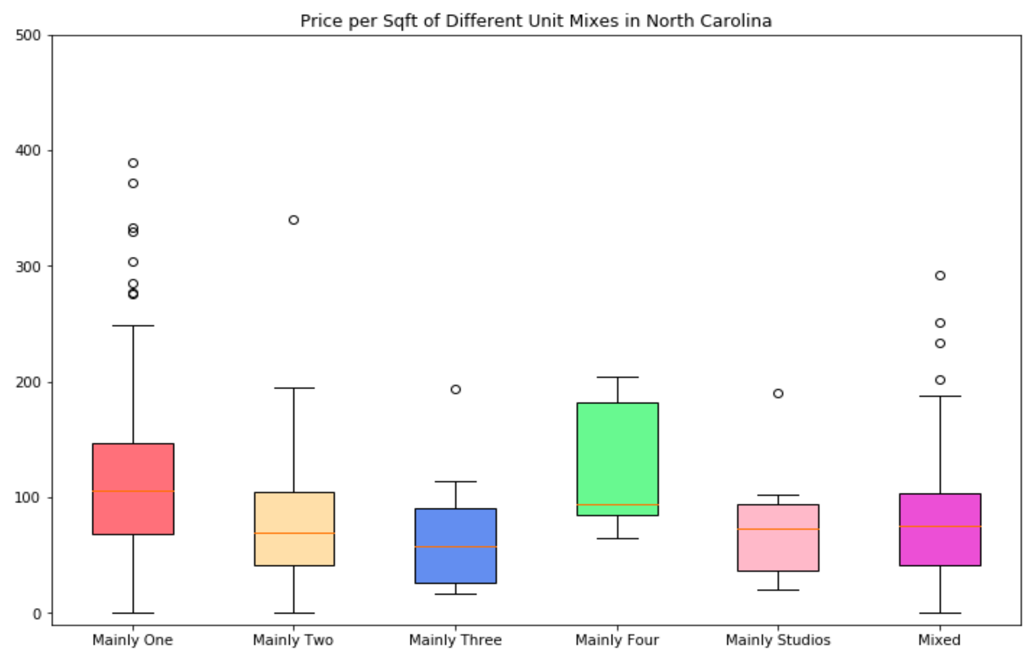


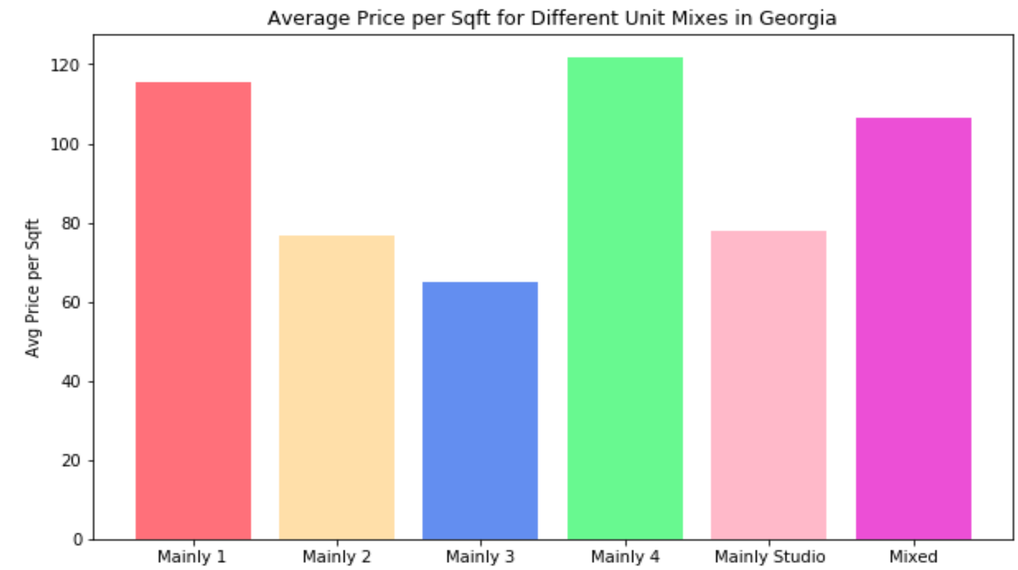



Average Price per Sqft for Different Unit Mixes in Texas
Avg Price per Sqft
0
20
40
60
80
100
120
Mainly 1
Mainly 2
Mainly 3
Mainly 4
Mainly Studio
Mixed


Price per Sqft of Different Unit Mixes in Texas
0
100
200
300
400
500
Mainly One
Mainly Two
Mainly Three
Mainly Four
Mainly Studios
Mixed


Average Price per Sqft for Different Unit Mixes in North Carolina
Avg Price per Sqft
0
20
40
60
80
100
120
Mainly 1
Mainly 2
Mainly 3
Mainly 4
Mainly Studio
Mixed


Price per Sqft of Different Unit Mixes in North Carolina
0
100
200
300
400
500
Mainly One
Mainly Two
Mainly Three
Mainly Four
Mainly Studios
Mixed


Average Price per Sqft for Different Unit Mixes in Georgia
Avg Price per Sqft
0
20
40
60
80
100
120
Mainly 1
Mainly 2
Mainly 3
Mainly 4
Mainly Studio
Mixed

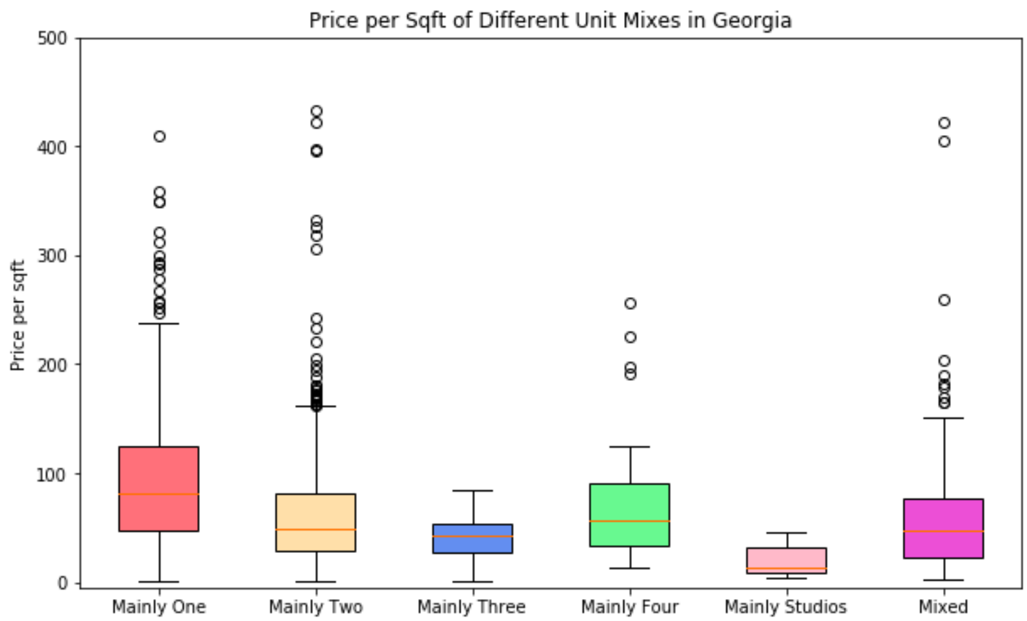
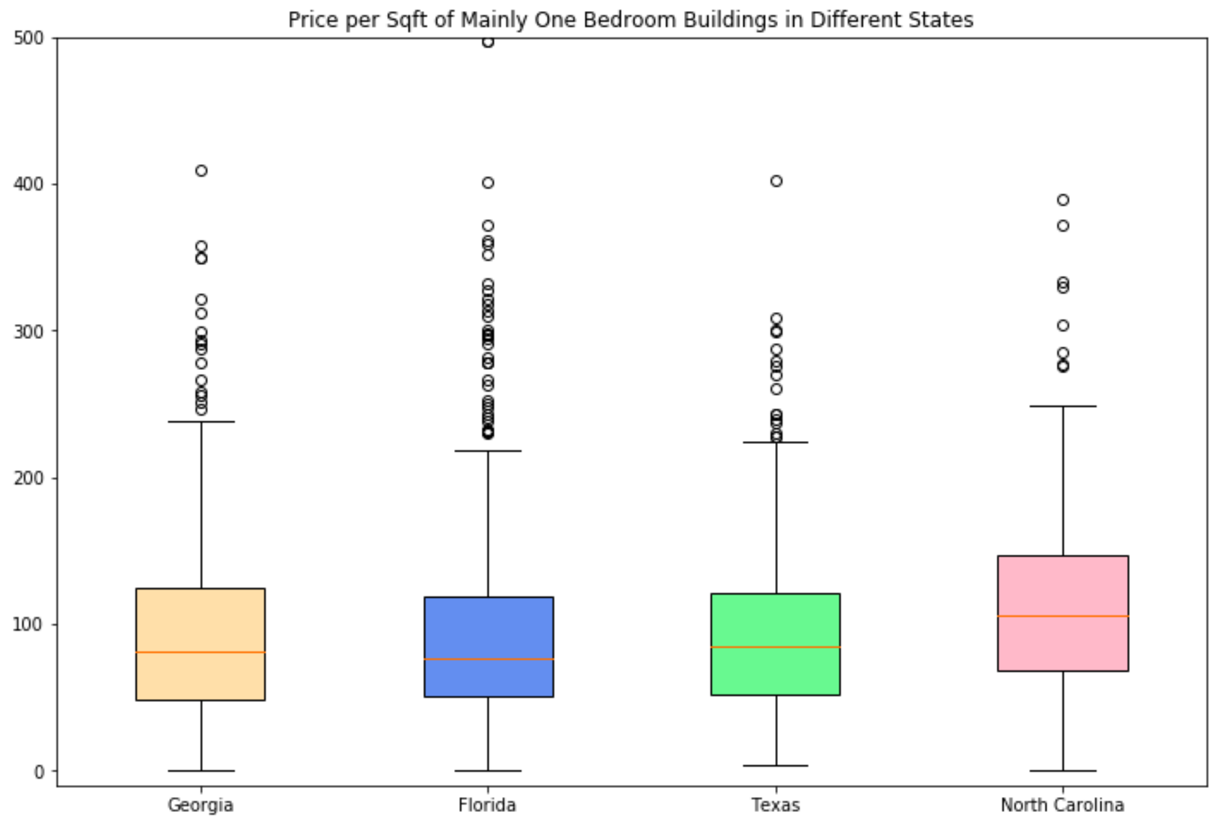
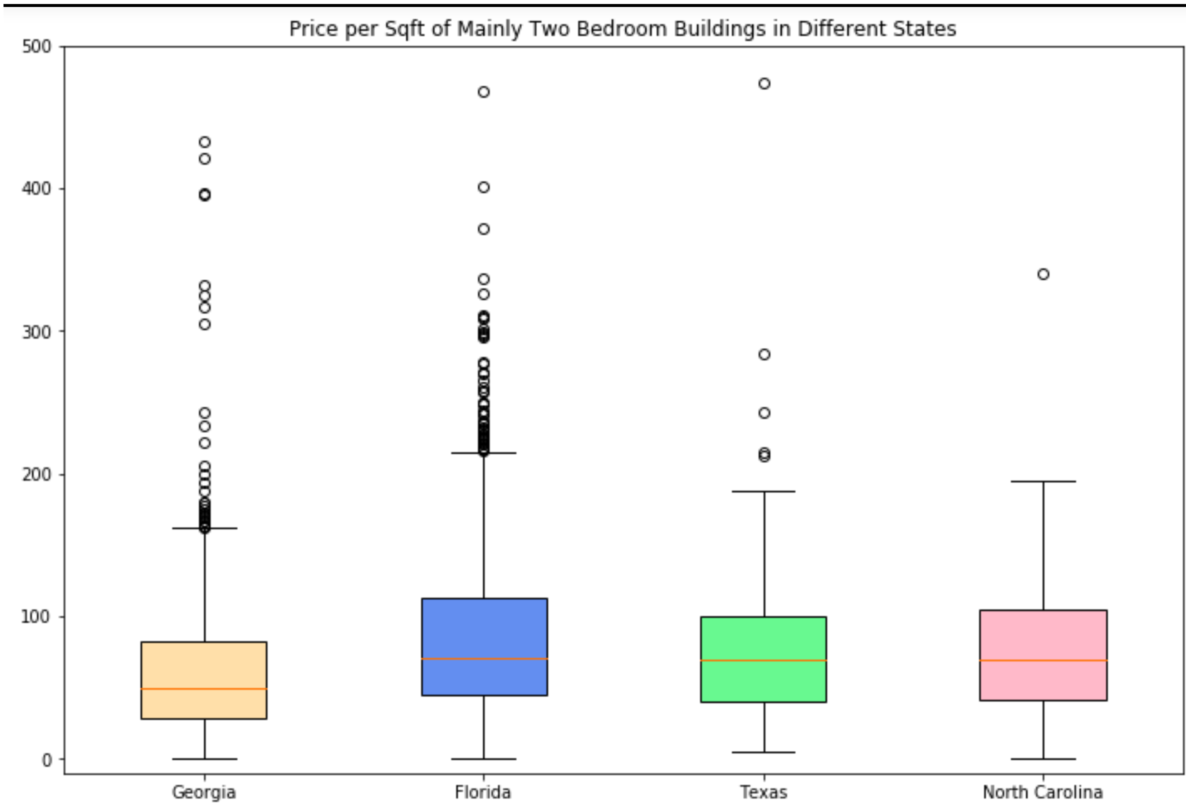
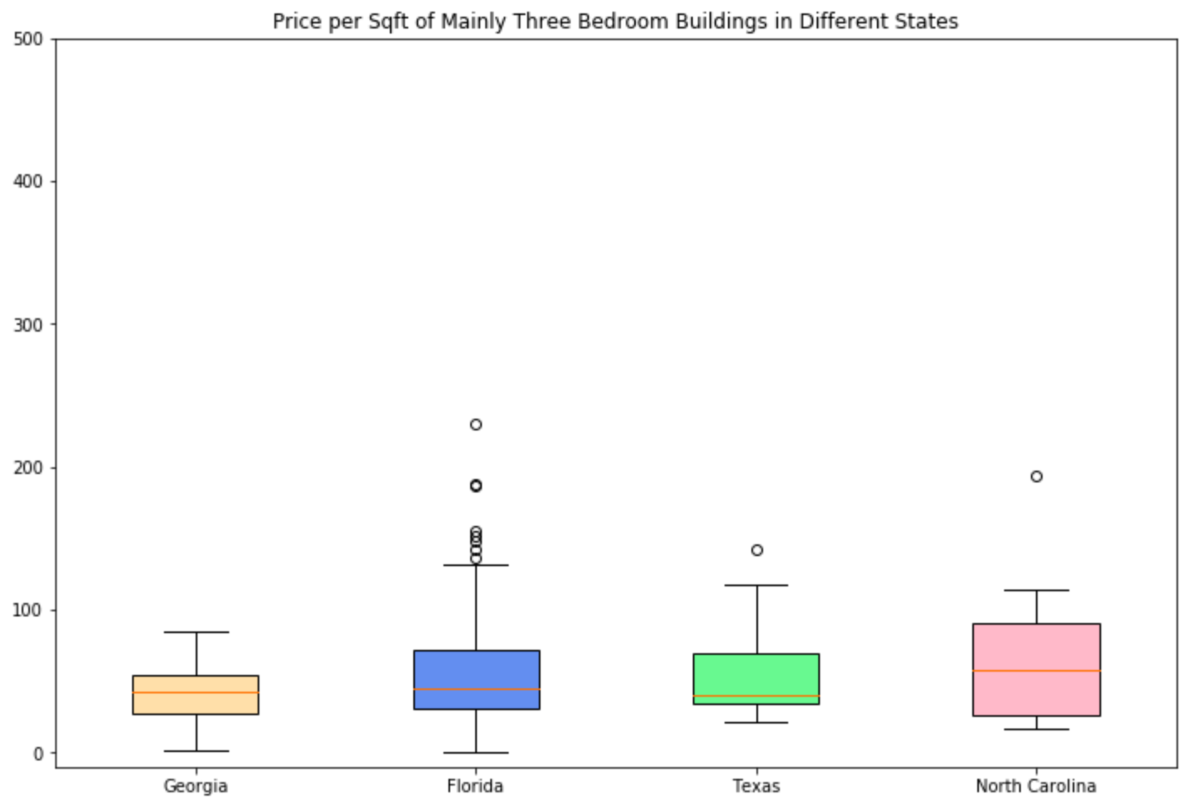
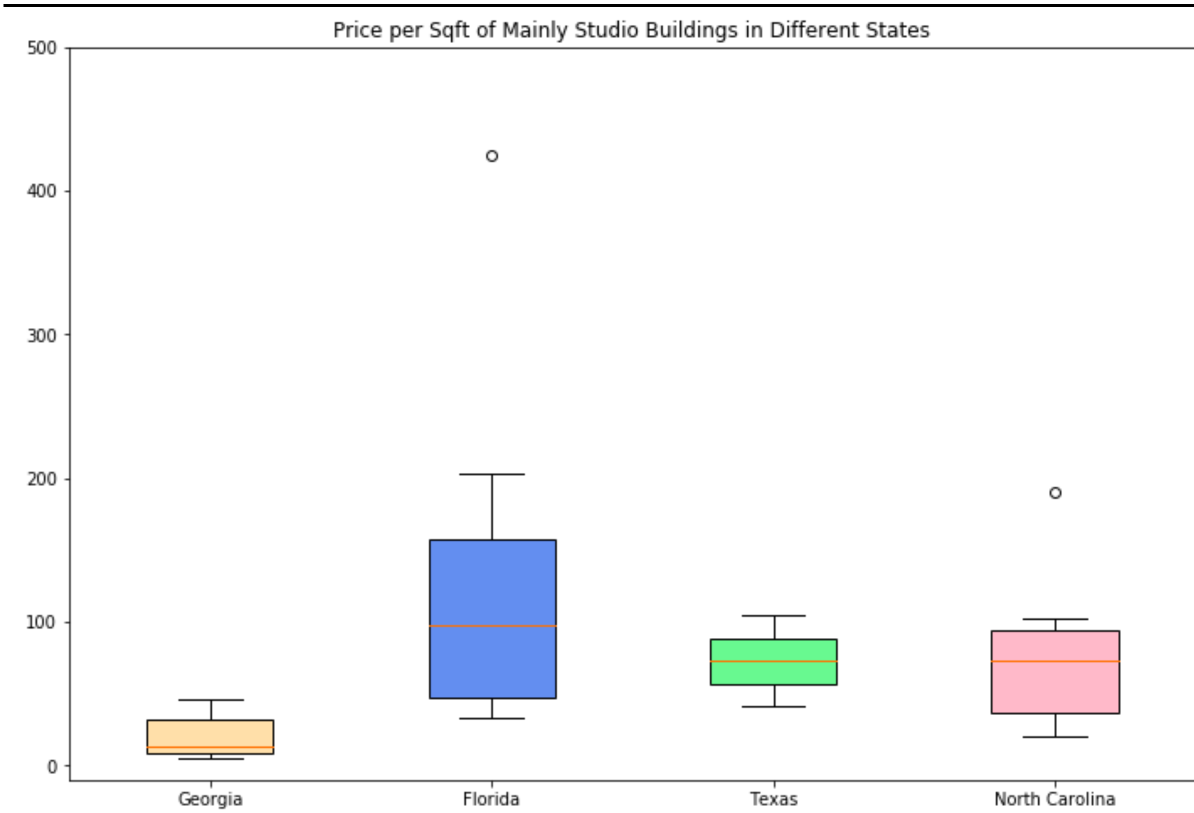

Price per Sqft of Different Unit Mixes in Georgia
Price per sqft
0
100
200
300
400
500
Mainly One
Mainly Two
Mainly Three
Mainly Four
Mainly Studios
Mixed

Figure 9: Boxplots for Different Types of Buildings






Price per Sqft of Mainly Three Bedroom Buildings in Different States
500
400
300
200
100
0
Georgia
Florida
Texas
North Carolina


Price per Sqft of Mainly Studio Buildings in Different States
500
400
300
200
100
0
Georgia
Florida
Texas
North Carolina

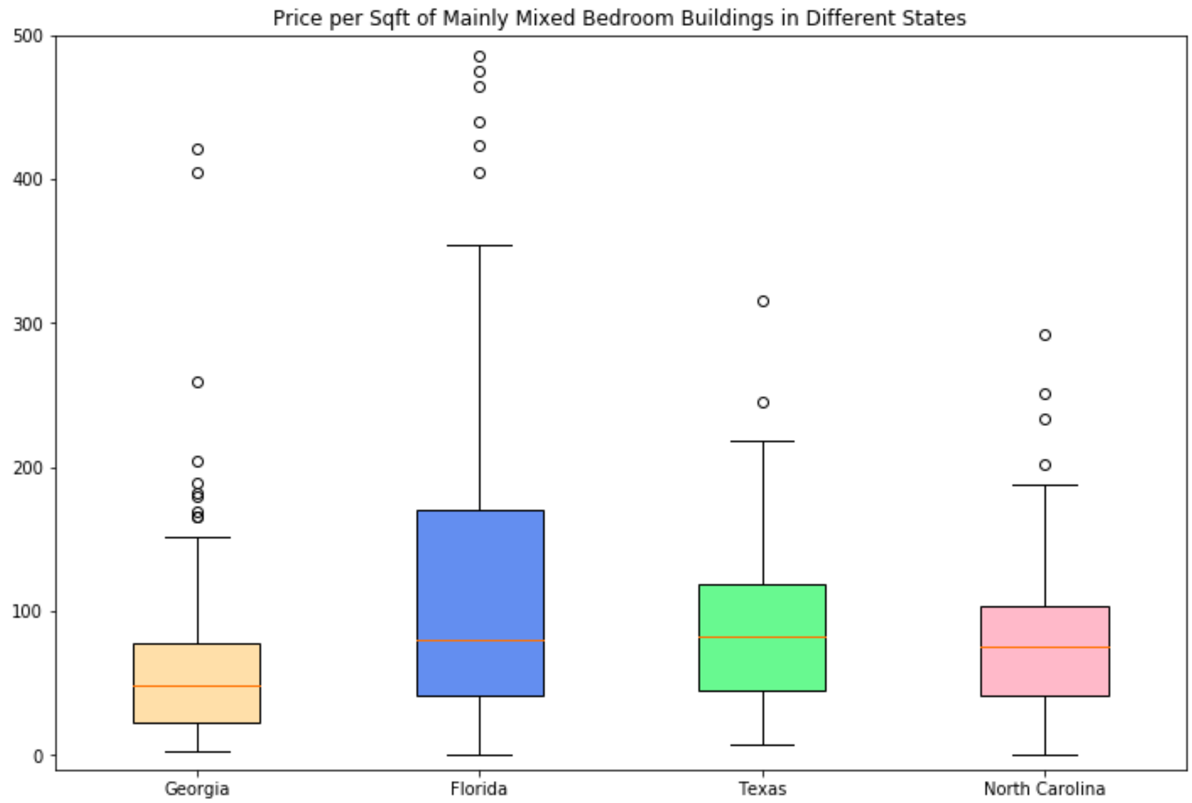
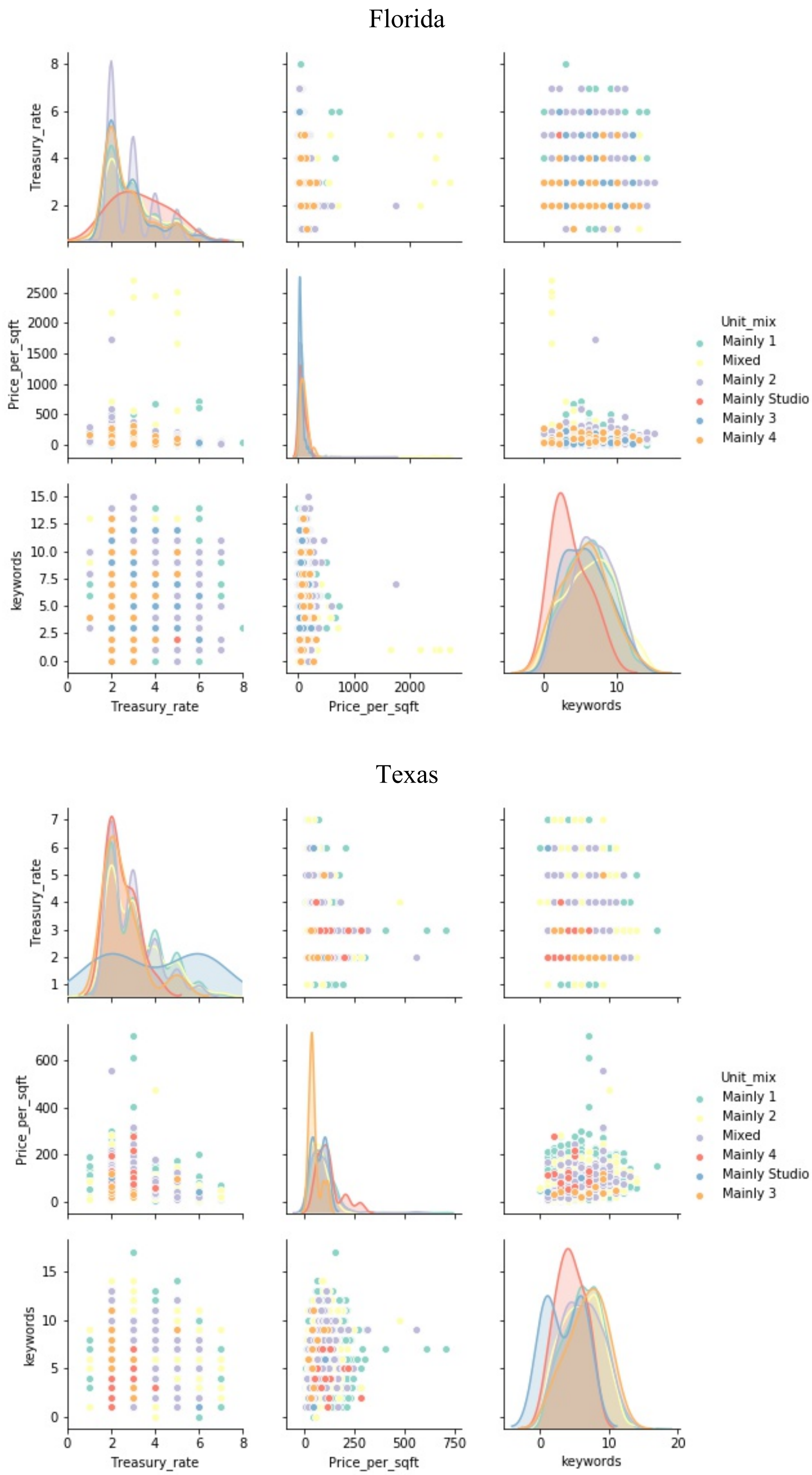
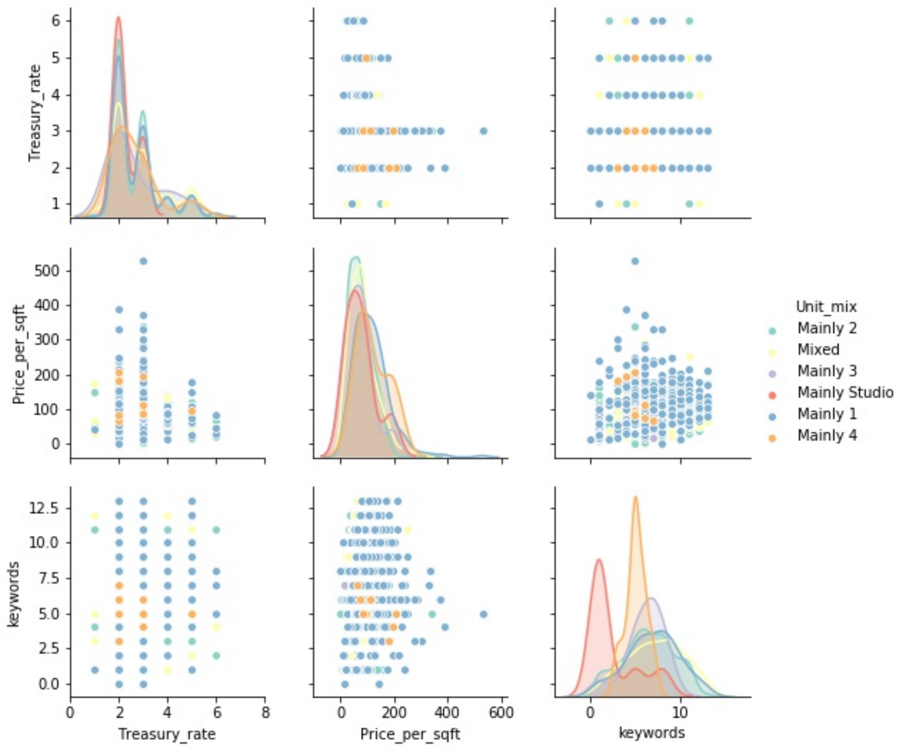

Price per Sqft of Mainly Mixed Bedroom Buildings in Different States
500
400
300
200
100
0
Georgia
Florida
Texas
North Carolina

Figure 10: Pair Plots for Joint Relationships



North Carolina



Georgia

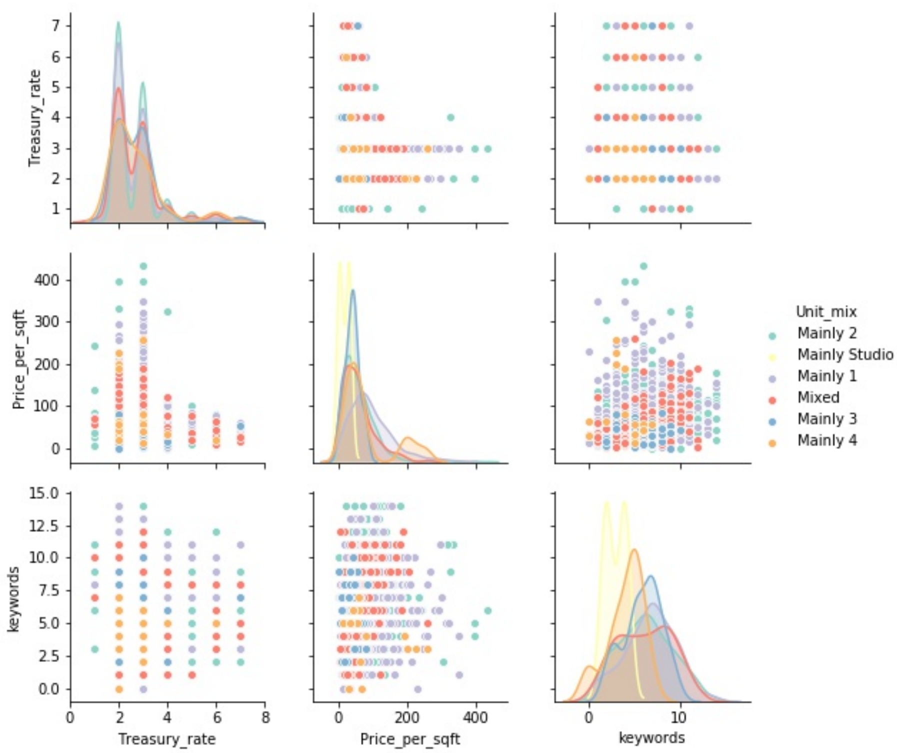
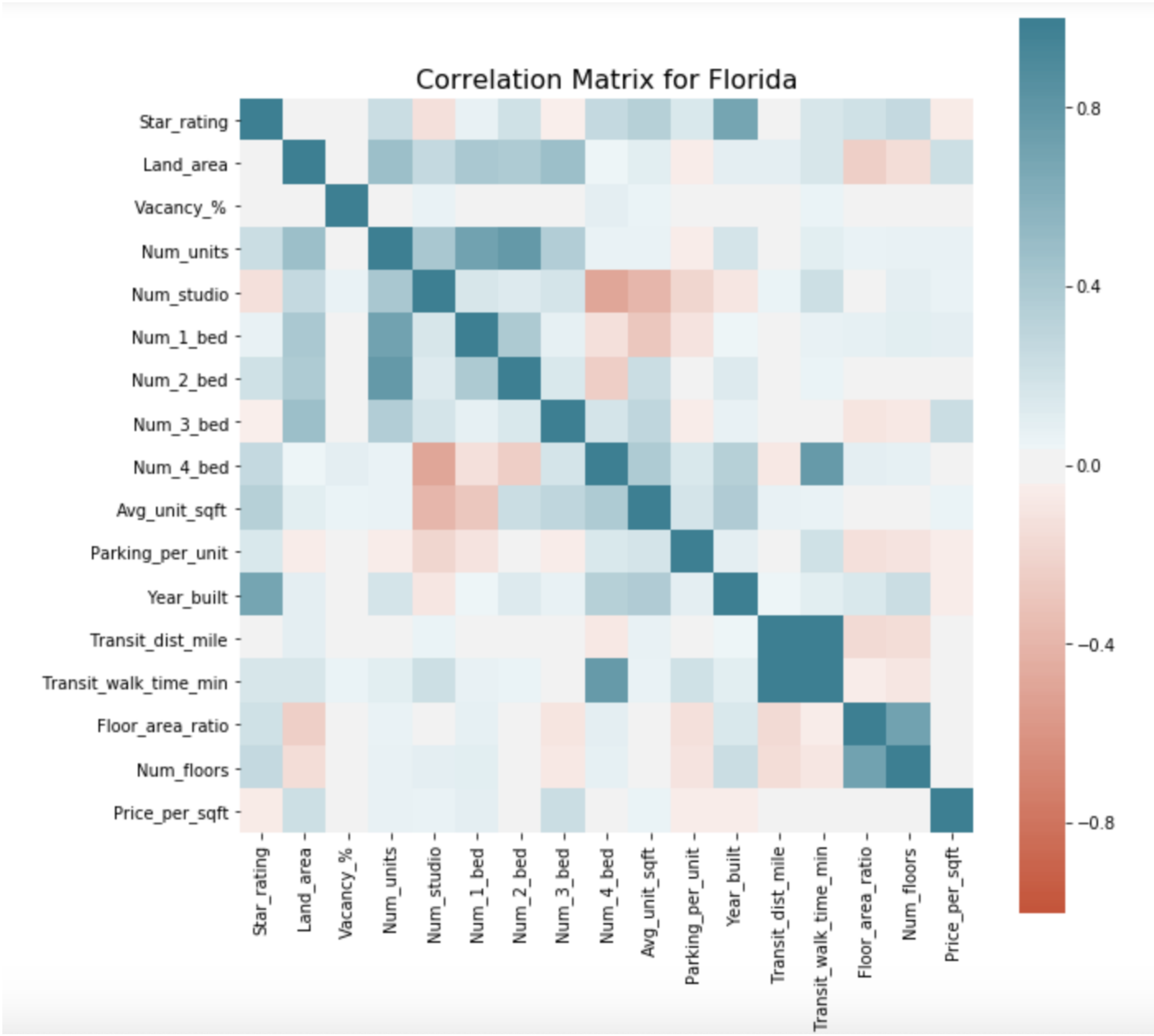
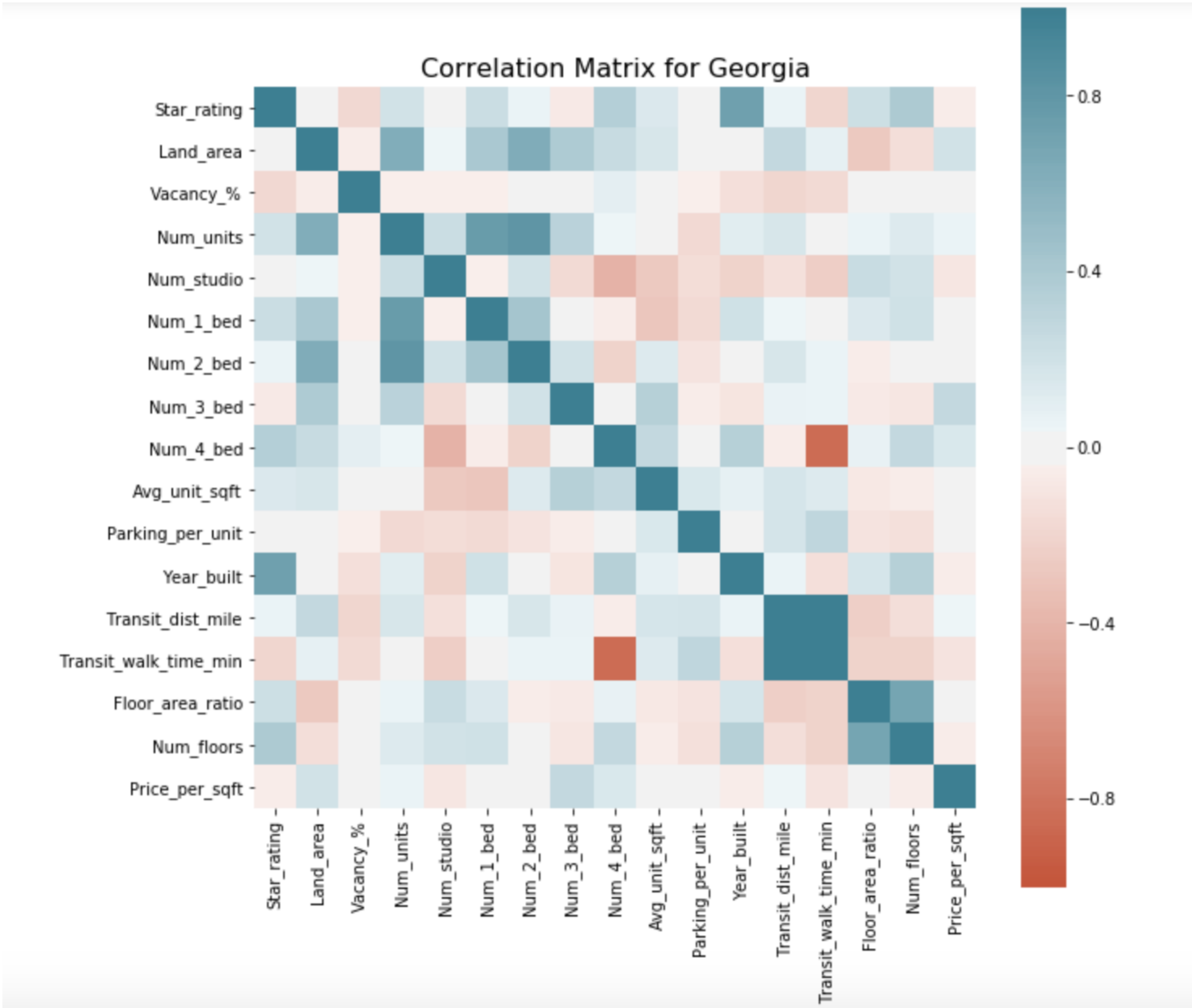
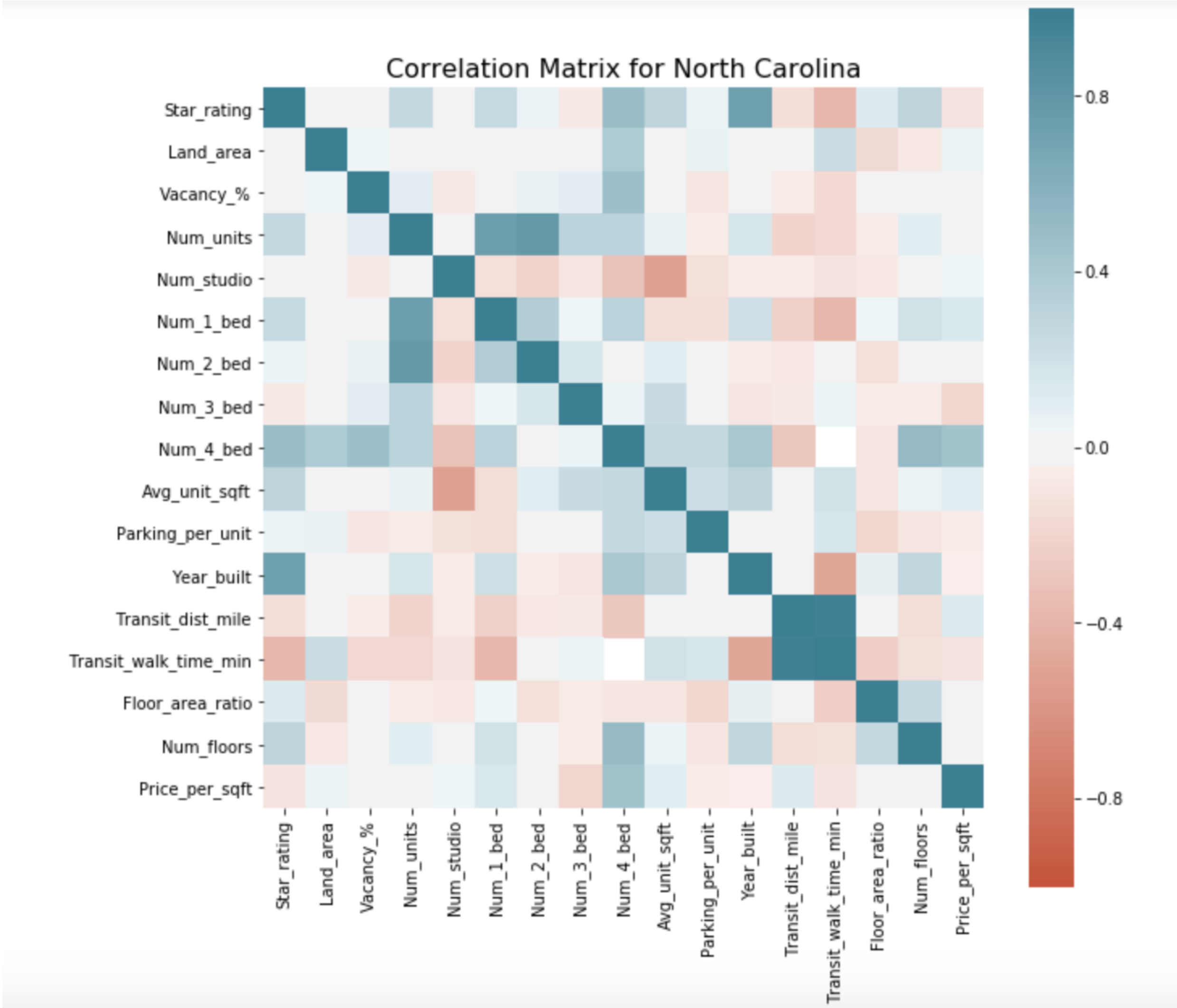


Figure 11: Correlation Matrix





Correlation Matrix for Georgia
Star_rating
Land_area
Vacancy_%
Num_units
Num_studio
Num_1_bed
Num_2_bed
Num_3_bed
Num_4_bed
Avg_unit_sqft
Parking_per_unit
Year_built
Transit_dist_mile
Transit_walk_time_min
Floor_area_ratio
Num_floors
Price_per_sqft
0.8
0.4
0.0
−0.4
−0.8


Correlation Matrix for North Carolina
Star_rating
Land_area
Vacancy_%
Num_units
Num_studio
Num_1_bed
Num_2_bed
Num_3_bed
Num_4_bed
Avg_unit_sqft
Parking_per_unit
Year_built
Transit_dist_mile
Transit_walk_time_min
Floor_area_ratio
Num_floors
Price_per_sqft
0.8
0.4
0.0
−0.4
−0.8

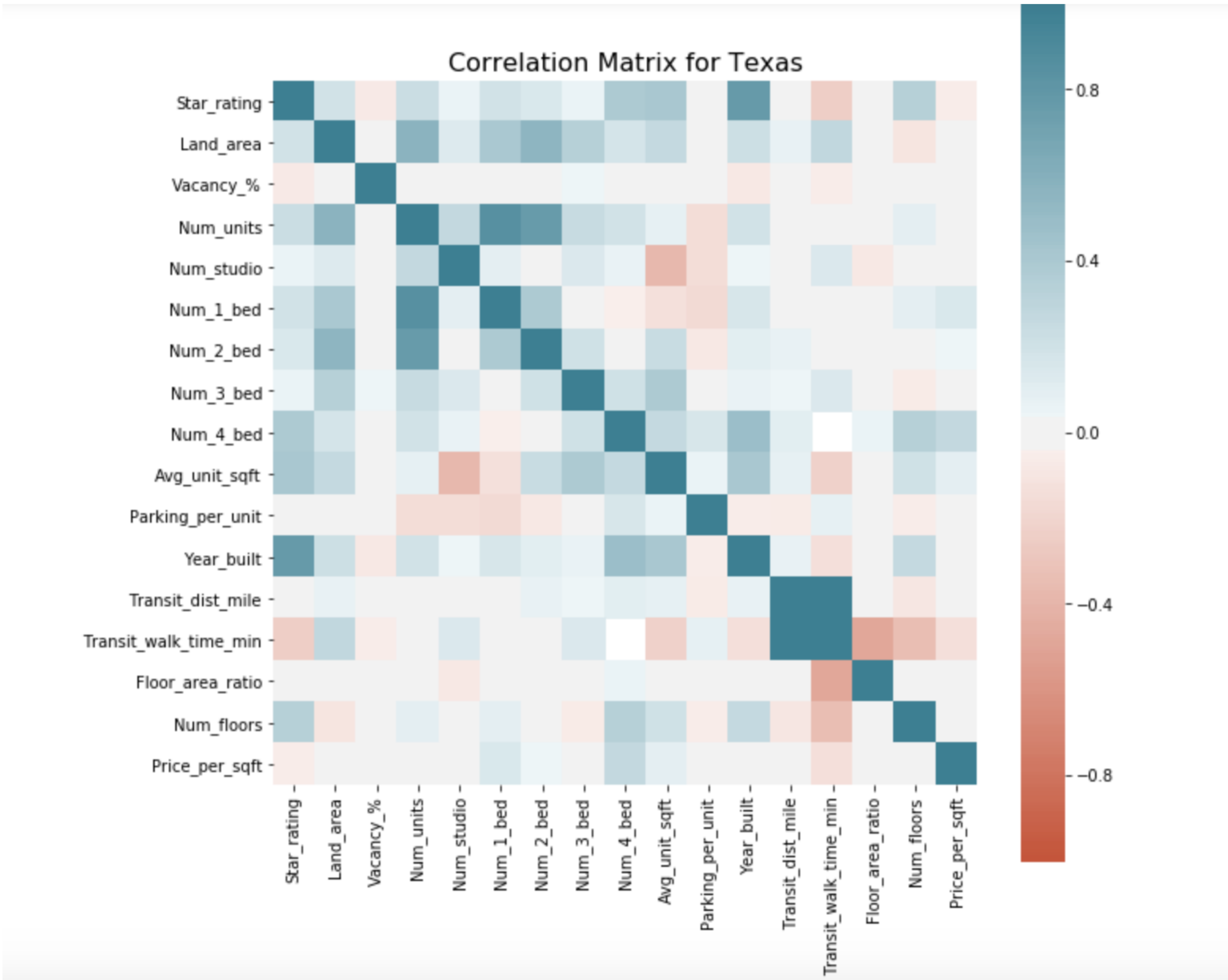

Correlation Matrix for Texas
Star_rating
Land_area
Vacancy_%
Num_units
Num_studio
Num_1_bed
Num_2_bed
Num_3_bed
Num_4_bed
Avg_unit_sqft
Parking_per_unit
Year_built
Transit_dist_mile
Transit_walk_time_min
Floor_area_ratio
Num_floors
Price_per_sqft
0.8
0.4
0.0
−0.4
−0.8

Figure 12: Regression Model for Four States

Florida

```
Call:
lm(formula = Price_per_sqft ~ Star_rating + Land_area + Num_units + 
    Avg_unit_sqft + Parking_per_unit + Year_built + Floor_area_ratio + 
    Num_floors, data = florida)

Residuals:
    Min      1Q  Median      3Q     Max 
-1064.5   -51.0    -6.6    37.2 18517.4 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -1.803e+03  1.490e+03  -1.210  0.22654    
Star_rating      -3.718e+01  1.654e+01  -2.248  0.02469 *  
Land_area         3.395e+00  6.091e-01   5.573 2.77e-08 ***
Num_units         2.271e-01  7.263e-02   3.126  0.00179 ** 
Avg_unit_sqft     9.084e-02  4.360e-02   2.084  0.03730 *  
Parking_per_unit -3.070e+01  1.190e+01  -2.580  0.00995 ** 
Year_built        9.287e-01  7.712e-01   1.204  0.22863    
Floor_area_ratio  1.236e+01  1.563e+01   0.791  0.42918    
Num_floors        3.042e+00  5.129e+00   0.593  0.55322    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 388.7 on 2467 degrees of freedom
  (4103 observations deleted due to missingness)
Multiple R-squared:  0.03885,	Adjusted R-squared:  0.03574 
F-statistic: 12.47 on 8 and 2467 DF,  p-value: < 2.2e-16
```

North Carolina

```
Call:
lm(formula = Price_per_sqft ~ Star_rating + Land_area + Num_units + 
    Avg_unit_sqft + Parking_per_unit + Year_built + Floor_area_ratio + 
    Num_floors, data = nc)

Residuals:
     Min       1Q   Median       3Q      Max 
-134.730  -24.196   -5.476   21.764  234.487 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)      -2.958e+03  2.421e+02 -12.218  < 2e-16 ***
Star_rating       1.465e+01  2.975e+00   4.924 1.00e-06 ***
Land_area         1.488e-01  1.752e-01   0.849 0.396042    
Num_units         1.602e-02  1.809e-02   0.886 0.376025    
Avg_unit_sqft    -3.204e-02  9.078e-03  -3.529 0.000437 ***
Parking_per_unit -9.218e+00  2.132e+00  -4.324 1.70e-05 ***
Year_built        1.519e+00  1.253e-01  12.130  < 2e-16 ***
Floor_area_ratio  1.388e+01  2.690e+00   5.161 3.01e-07 ***
Num_floors        2.105e+00  7.952e-01   2.648 0.008245 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 38.47 on 926 degrees of freedom
  (1930 observations deleted due to missingness)
Multiple R-squared:  0.483,	Adjusted R-squared:  0.4785 
F-statistic: 108.1 on 8 and 926 DF,  p-value: < 2.2e-16
```

Georgia

```
Call:
lm(formula = Price_per_sqft ~ Star_rating + Land_area + Num_units +
    Avg_unit_sqft + Parking_per_unit + Year_built + Floor_area_ratio +
    Num_floors, data = georgia)

Residuals:
    Min      1Q  Median      3Q     Max
-159.37  -27.94   -7.22   16.43 2248.58

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -2.336e+03  3.661e+02  -6.381 2.30e-10 ***
Star_rating       1.589e+01  3.982e+00   3.991 6.87e-05 ***
Land_area         8.460e-01  2.268e-01   3.730 0.000198 ***
Num_units        -3.242e-02  2.445e-02  -1.326 0.184953
Avg_unit_sqft    -2.915e-02  1.153e-02  -2.527 0.011609 *
Parking_per_unit -9.720e+00  2.718e+00  -3.576 0.000360 ***
Year_built        1.201e+00  1.882e-01   6.382 2.28e-10 ***
Floor_area_ratio  9.600e+00  4.130e+00   2.324 0.020233 *
Num_floors        2.245e+00  2.170e+00   1.035 0.301051
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 76.28 on 1585 degrees of freedom
  (2001 observations deleted due to missingness)
Multiple R-squared:  0.1512,    Adjusted R-squared:  0.1469
F-statistic:  35.3 on 8 and 1585 DF,  p-value: < 2.2e-16
```

Texas

```
Call:
lm(formula = Price_per_sqft ~ Star_rating + Land_area + Num_units +
    Avg_unit_sqft + Parking_per_unit + Year_built + Floor_area_ratio +
    Num_floors, data = texas)

Residuals:
    Min      1Q  Median      3Q     Max
-113.32  -23.85   -5.76   19.84  567.93

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      -3.823e+03  2.227e+02 -17.167  < 2e-16 ***
Star_rating       3.763e+00  2.333e+00   1.613  0.10691
Land_area        -3.113e-01  1.622e-01  -1.920  0.05507 .
Num_units         2.731e-02  9.628e-03   2.836  0.00461 **
Avg_unit_sqft    -1.768e-02  7.775e-03  -2.274  0.02309 *
Parking_per_unit  8.537e-01  1.390e+00   0.614  0.53917
Year_built        1.956e+00  1.152e-01  16.977  < 2e-16 ***
Floor_area_ratio -2.331e-02  2.576e-02  -0.905  0.36561
Num_floors        2.897e+00  6.302e-01   4.597 4.57e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.54 on 1901 degrees of freedom
  (9718 observations deleted due to missingness)
Multiple R-squared:  0.3584,    Adjusted R-squared:  0.3557
F-statistic: 132.8 on 8 and 1901 DF,  p-value: < 2.2e-16
```

Work Cited

10-Year Treasury Constant Maturity Rate. (2019, December 6). Retrieved from https://fred.stlouisfed.org/series/DGS10.