

X

AUSTIN, TX SHORT TERM RENTAL MARKET

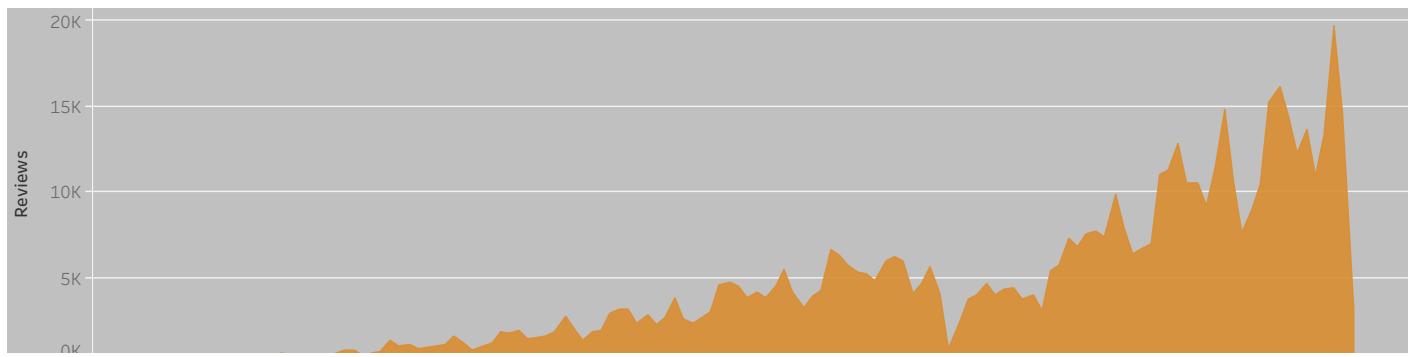
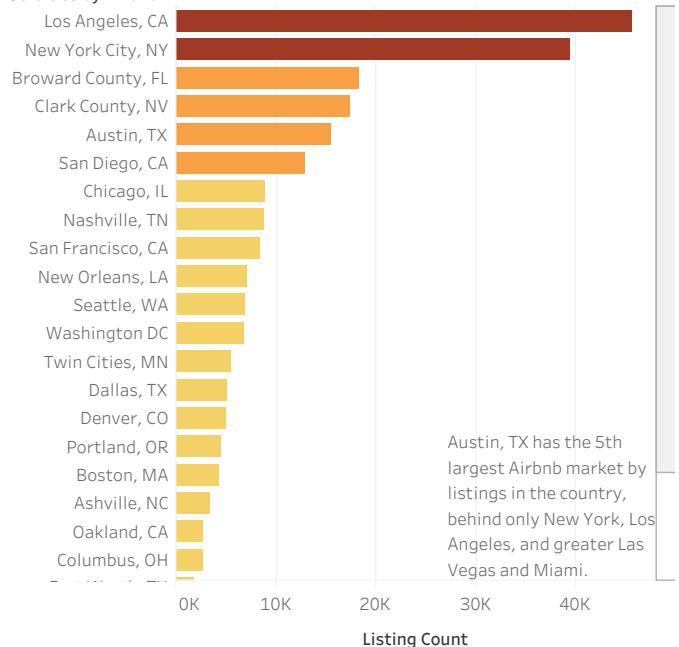
As Austin's reputation as a tech and tourism hotspot continues to soar, another trend has gained momentum: **the rise of Airbnb and short-term rentals**. With visitors flocking to experience the city's unique culture and events, the demand for alternative accommodations has surged. Airbnb listings offer travelers a diverse array of options, from cozy downtown lofts to charming bungalows nestled in Austin's eclectic neighborhoods.

Mirroring this increased demand, STR supply has risen in turn. As of December 2023, there are over 15,000 listings in the Austin area, more than Denver (4,971), New Orleans (7,075), or Nashville (8,850). Austin's STR market, like its residential housing market, is very competitive.

With that in mind, there may still be opportunities for the savvy investor. Analyzing the data, we can determine the viability of investing in the ATX STR market.

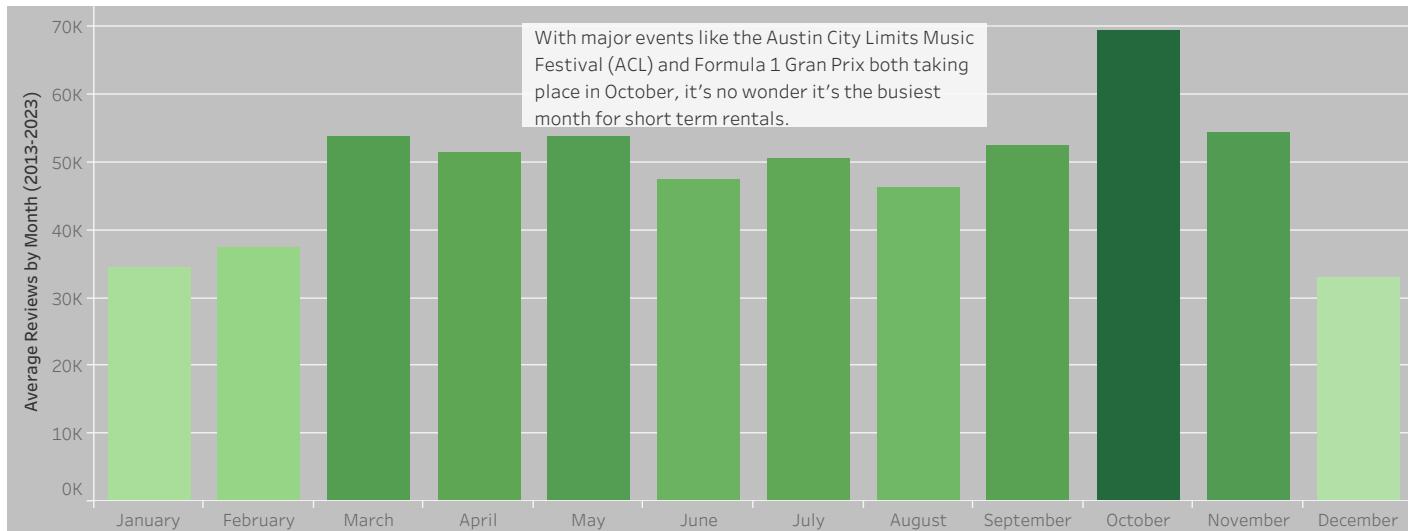
US Cities by Airbnb Listings

US Cities by Airbnb..



STR Seasonality Considerations

ATX STR Seasonality



Property Type Counts with Median Price per Night (Dec '23)

While Airbnb commands a substantial share of the online travel agency (OTA) market, it operates within a competitive landscape alongside platforms such as Vrbo, Booking.com, and Expedia. Despite this competition, Airbnb stands out as the leading player in the hotel-alternative sector.

Entire home \$209.5 5,839	Private room in home \$60.0 1,748	Entire condo \$140.0 1,260
Entire guesthouse \$109.0 695		
Entire rental unit \$97.0		



1) Introduction:
Austin and the Airbnb..

2) Seasonality of Short
Term Rental Market

3) Exploratory Data
Analysis

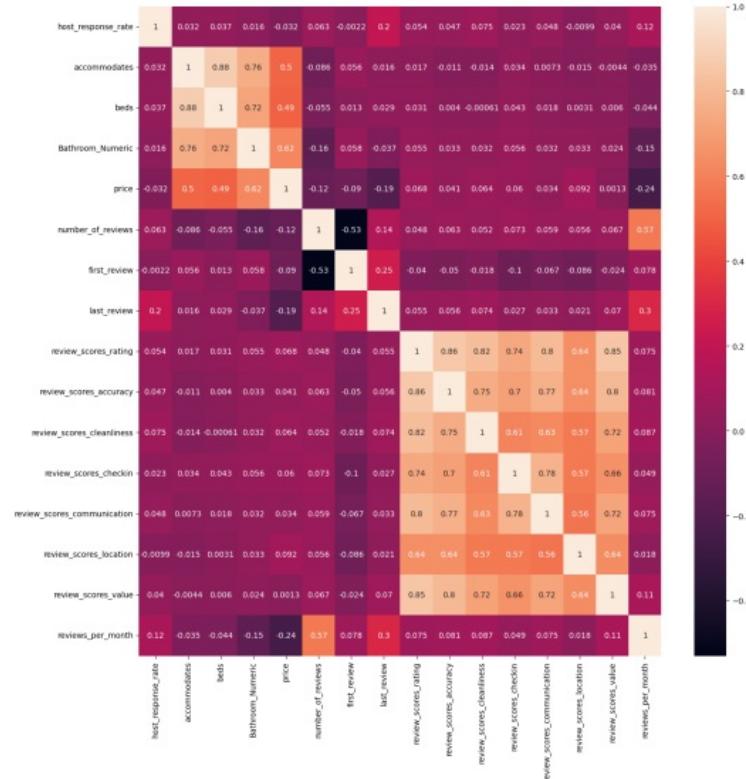
4) ATX by Zip Code

5) Introducing Clusters

6) Austin by Segments

7)
Conclusion/..

Exploratory Data Analysis



Correlation heatmap did not yield very strong or unexpected correlations.

Positive Correlation:

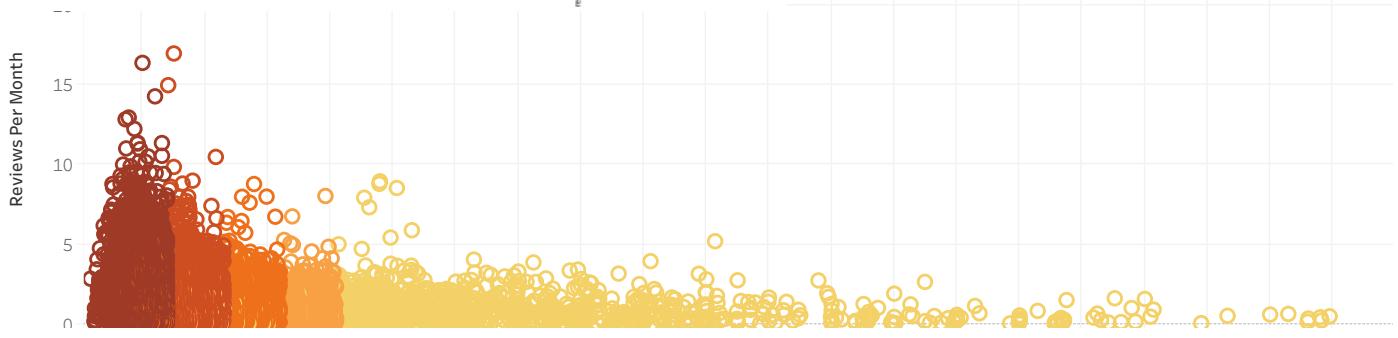
- 'beds' , 'accommodates' , 'Bathroom_Numeric';
- The 7x7 square of 'review_scores' and its subcategories;
- Number of reviews and reviews per month.

Negative Correlations:

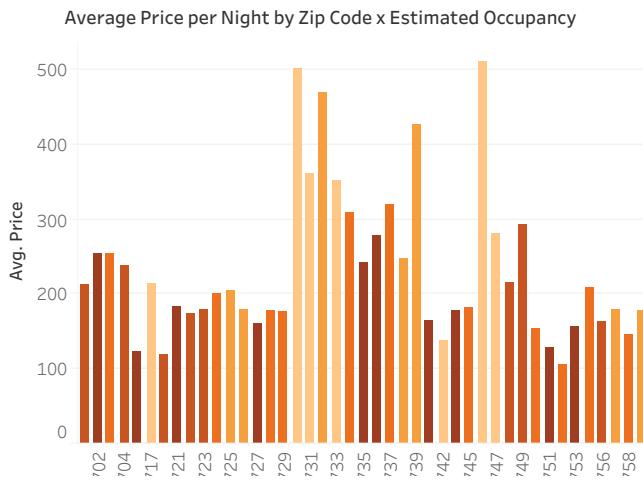
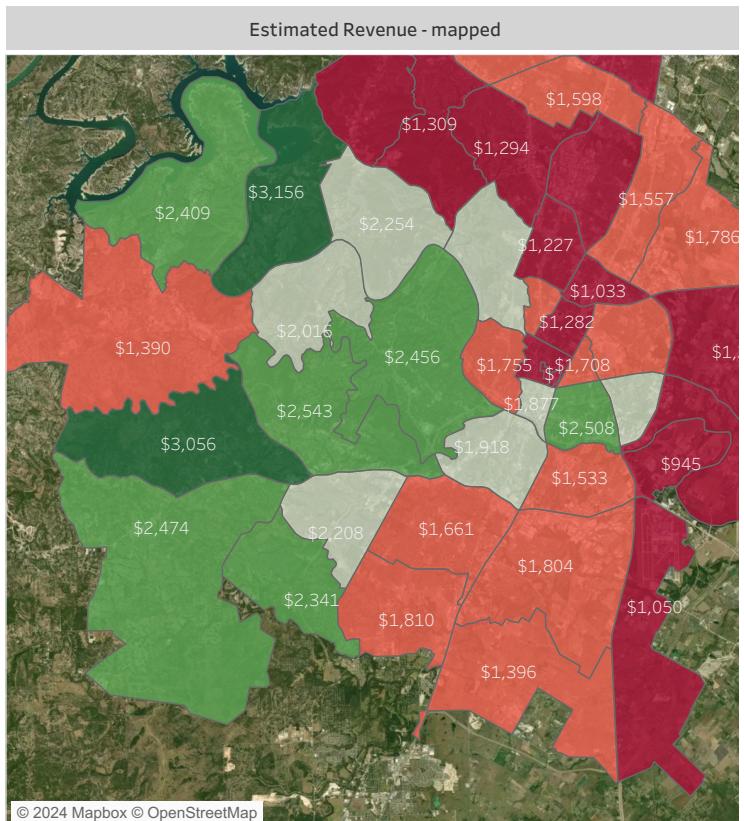
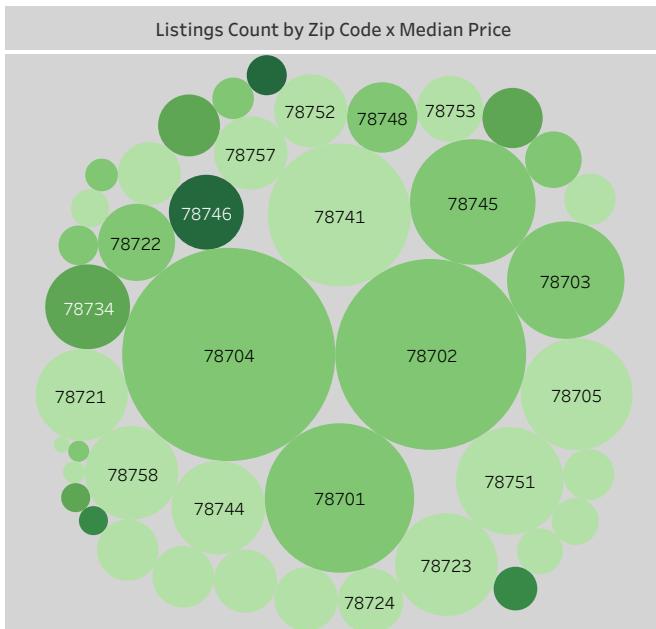
- Price and host response rate;
- First review and number of reviews;

Insights:

- Places that can accommodate more people by definition have more bedrooms and bathrooms.
- A high score in one would generally be correlated with other high scores. Generally indicative of a positive/negative experience.
- More frequently reviewed listings will also have a higher review count.
- The higher the price, the lower the response rate. More expensive properties, probably more vacancy/ less stays, balanced out by more money per stay.
- More established properties in the market will have m..



Austin Texas by Zip Code



Insights:

- 78704 and 78702 are the most populated/ listing dense zip codes.
- Highest priced zip codes have lower occupancy. 78732 is the highest occupancy of them -> highest est. revenue.
- The East side of Austin has relatively low estimated revenue, vs the West and North sides.

Segmenting Listings via Machine Learning

Using machine-learning, I clustered every listing into one of five segments that share a statistical profile.

First segment = **Mediocre Listings**

- below average review scores (~4.3)
- Lower age (by months) on the market

Second segment = **Established/Overachieving Listings**

- lowest average prices (low margin)
- highest average reviews (high volume)
- second highest estimated revenue (behind segment 4)

Third segment = **Average Listings**

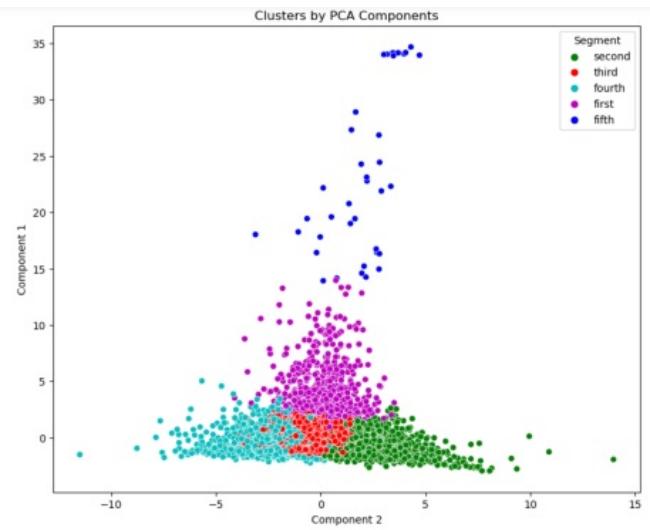
- Middle in Price, Number of Reviews, Estimated Revenue.
- Highest in Review Scores and Estimated Listing Age
- Lowest in accomodation numbers

Fourth segment = **Upscale/ Luxury Listings**

- Highest in accomodation, price, and estimated revenue
- High review scores and months on market
- High reviews on average, but distant second to Overachieving segment

Fifth segment = **Poor listings**

- lowest review count, review scores, revenue
- shortest time on market



```
1 df_pca_kmeans['Segment'].value_counts()
```

Segment	count
third	4357
second	1989
fourth	1287
first	733
fifth	38

Name: count, dtype: int64

Segment	accommodates		price		number_of_reviews		estimated_revenue		review_scores		est_listing_age	
	mean	median	mean	median	mean	median	mean	median	mean	median	mean	median
fifth	5.368421	4.0	198.052632	125.0	1.789474	1.0	374.272105	238.71	1.915526	2.00	8.631579	5.5
first	5.111869	4.0	148.272851	104.0	20.185539	10.0	917.160246	658.86	4.345566	4.43	18.803547	13.0
fourth	11.101787	10.0	509.508936	407.0	40.244755	23.0	3273.376970	2891.70	4.895532	4.95	32.311577	21.0

Geospatial Analysis by Segment



-Mediocre but also newer listings (less months since first review), appear most concentrated in the city core.

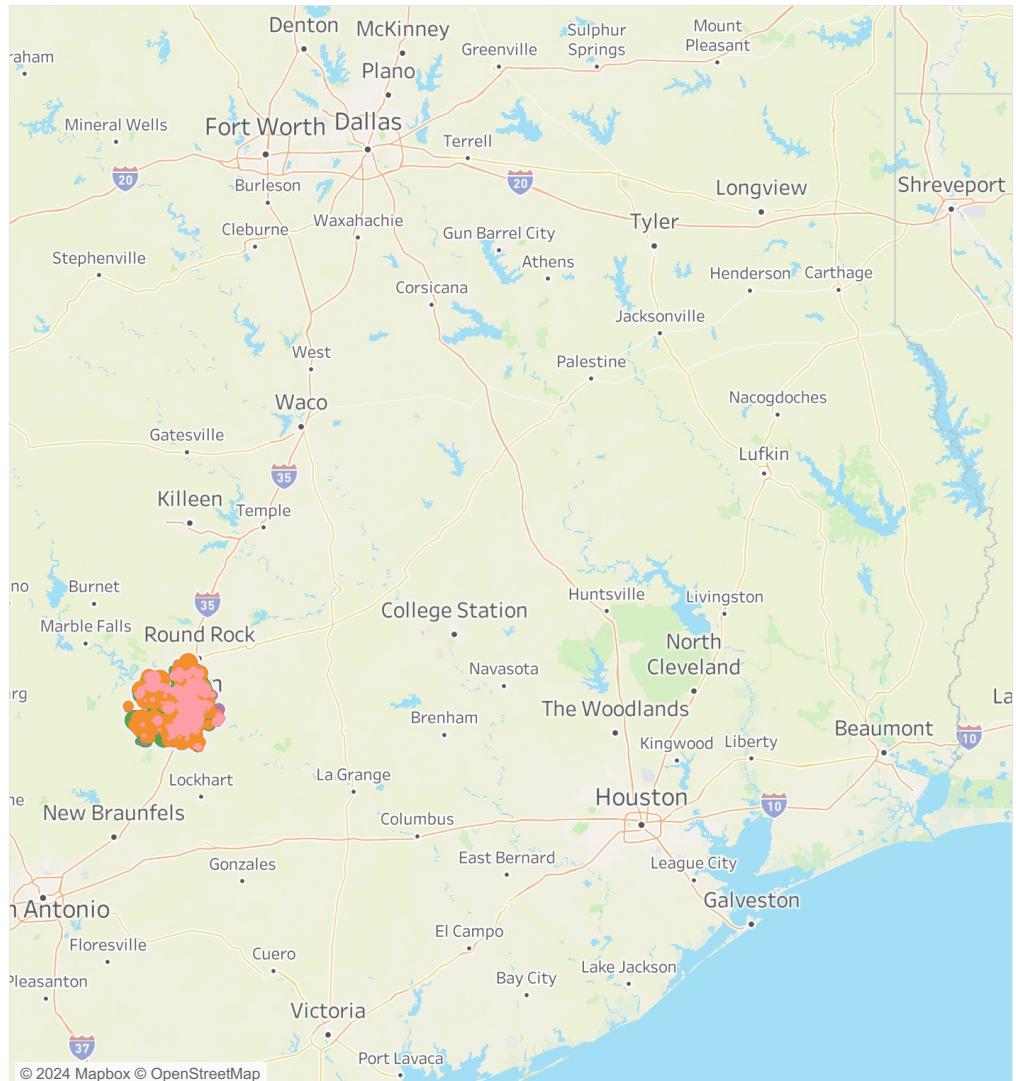
-Volume/Overachieving listings dominate the central city. They also appear ubiquitously across the city, a role that I expected only the average listings to hold.

-Average listings, as one might expect are across the city, but they seem overcrowded by volume overachievers across the board.

- Luxury listings appear in the main city as to the west of the city, especially West Lake Hills. They too seem positioned more diversely than expected.

-The 5th segment, poorly reviewed underperformers, are non-factors.

The Austin Airbnb market appears to be well



Conclusion and Recommendations

Across Austin, there is a robust inventory of high performing listings, average listings, and only a handful of not so high performing listings.

On a zip code/ neighborhood level, there are trends:

- 78704 and 78702 stand out as the most dense by listing count, and could reasonably be deemed the hub of Airbnb activity in the city.
- Neighborhoods to the west of this hub, especially 78746, 78732, and 78734, showed some of the highest prices per night and revenue figures on average.

On a city-wide level:

- we can see a mass of listings in the central city and a slightly lesser but still significant STR presence to the North, South, West and Southwest of Austin proper.
- relative lack of development to the East of Austin is notable.

From our machine-learning cluster analysis, we determined two distinct routes toward listing success:

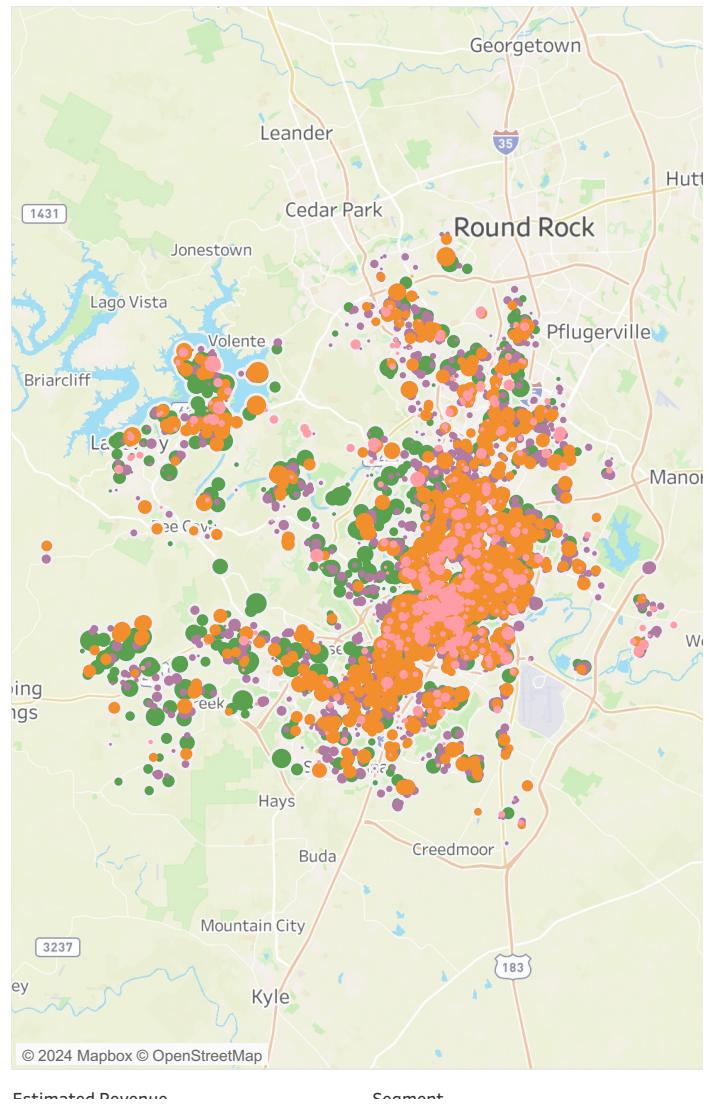
- Volume bookings at a competitive price or large and luxury listings.
- Surprisingly, we saw listings of these segment types coexisting in almost every neighborhood.

Recommendation:

- for any would-be STR investors, I would caution taking on an investment in the ATX area as it appears to be saturated.
- plenty of demand over the years has created a vast supply. Listings must be pareto-efficient to stand out.
- could still be a serious appreciation play, but with high prices and mortgage rates, cash flow will be harder to come by.

Limitations of this market analysis:

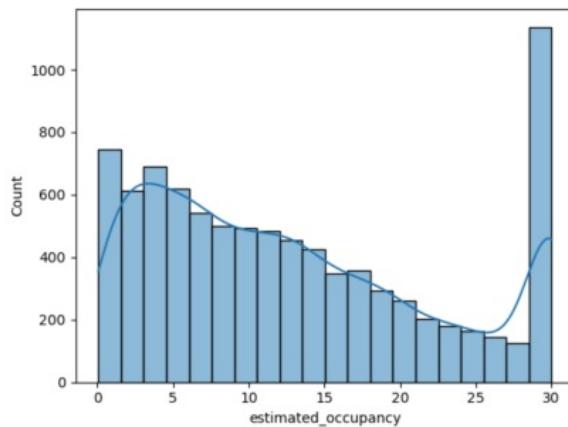
- lack of actual revenue figures in the data set necessitated an estimation based on review rates, which is not exactly representative of actual bookings.
- some places have better review rates than others, which can be a function of host involvement/prompting, listing category, or just chance.
- the 'bedrooms' variable did not come across in this scrape, so 'bathrooms' and 'accommodates' figures had to be used as proxies for listing size.
- the figures cited are based on an 'offseason' scrape, likely all prices and therefore revenue would be higher in a scrape during peak months.



About the Data

Downloaded from Inside Airbnb, (not endorsed by Airbnb or its competitors) an data/advocacy site with the mission to measure the impact that Airbnb has on affordability and housing in major tourist destinations across the globe.

Inside Airbnb conducts quarterly webscrapes of publicly available data from Airbnb, and therefore requires interpretation for key figures such as occupancy and revenue, as this information would be limited to hosts and internal Airbnb metrics.



Above: Seaborn histogram of entire_home listings by estimated occupancy, derived from each listing's reviews per month and minimum stay length.

Austin, Texas, United States

15 December, 2023 (Explore)

Country/City	File Name	Description
Austin	listings.csv.gz	Detailed Listings data
Austin	calendar.csv.gz	Detailed Calendar Data
Austin	reviews.csv.gz	Detailed Review Data
Austin	listings.csv	Summary information and metrics for listings in Austin (good for visualisations).

2. Data Wrangling

- Since we don't have an outright revenue column, we can derive it borrowing from the methodology outlined in Inside Airbnb's assumptions page: <http://insideairbnb.com/data-assumptions/>.
- We'll derive an ['estimated_occupancy'] metric by dividing ['reviews_per_month'] by 50% (in other words x2), times the minimum length of stay if greater than 3 nights, and capped at 3 if minimum stay lower than 3 nights.
 - not every guest will review. Airbnb CEO Brian Chesky uses 72% metric, which is likely optimistic; New York attorney general uses review rate of 30.5%. Inside Airbnb opted to split the difference with a 50% reviews to estimated bookings rate.
 - listings.csv dataset offers several 'minimum nights' metrics, including:
 - 'minimum_nights', minimum night stay for the listing
 - 'minimum_minimum_nights', the smallest minimum_night value from the calendar (looking 365 nights in the future)
 - 'maximum_minimum_nights' the largest minimum_night value from the calendar (looking 365 nights in the future),
 - 'minimum_nights_avg_ntm', the average minimum_night value from the calendar (looking 365 nights in the future)
 - opting to use 'minimum_nights_avg_ntm' for occupancy calculations. It may be slightly more accurate than 'minimum_nights'
 - 'estimated_occupancy' capped at 30 days for full occupancy.
 - Inside Airbnb caps occupancy at 70% to remain conservative; their goal is to demonstrate the effect which Airbnb's might impact the local housing supply.
- Next we'll derive ['estimated_revenue'] by multiplying ['price'] (*which can be changed by the host on a given date but for our purposes will remain static*) times ['estimated_occupancy'].

From Jupyter Notebook 6.3, an explainer on the methodology to derive estimated occupancy and estimated revenue.

By no means a perfect estimation, this derivation can still give us an idea on what listings are overperforming in the market.

See github link here: <https://github.com/jobachone/Austin-Airbnb>

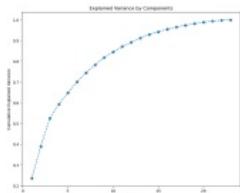
Worked primarily with 'listings.csv.gz' dataset, but also utilized the 'neighbourhoods.geojson' for geospatial analysis using folium python library, as well as the reviews.csv for time series data.

Pre data cleaning, the Doc 22 detailed listings dataset

Clustering: Unsupervised Machine Learning Technique

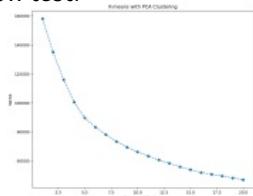
X

Using **Scikit-learn**, a machine learning library for Python, I analyzed the listings dataset and determined that 80% of its variance came down to 9 components.



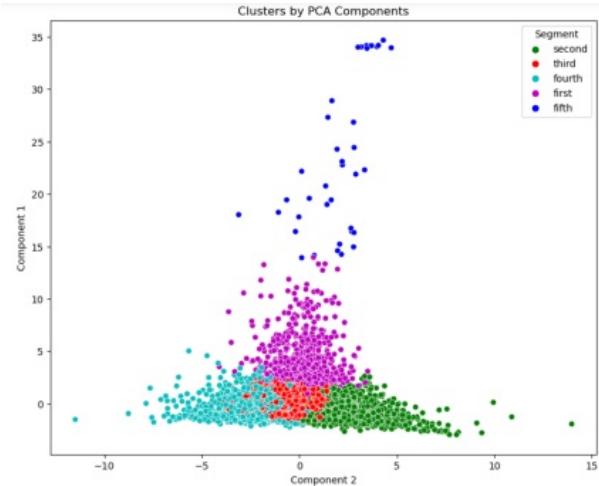
From there, I standardized the data using `StandardScaler()`, as the different components (pricing, review scores, number of reviews, age of listing in months) all had different orders of magnitude.

Then I applied PCA (Principal Component Analysis) followed by a k-means clustering algorithm in order to reduce the datasets dimensions and determined the number of clusters, which was 5 segments, via the elbow test.



After plotting a number of variable pairs with their clusters, and aggregating descriptive statistics for key variables to distinguish the segments, I was able to derive some key takeaways that distinguish the segments from one another. ->

accommodates	price	number_of_reviews	estimated_revenue	review_scores	est_listing_age
mean	median	mean	median	mean	median



```
1 df_pca_kmeans['Segment'].value_counts()
Segment
third    4357
second   1989
fourth   1287
first     733
fifth     38
Name: count, dtype: int64
```

Takeaways from Descriptive Statistics:

first segment - new and mediocre listings

- 2nd lowest listing age statistics (in months), only ahead of fifth segment.
- 2nd lowest median estimated revenue, 2nd lowest review scores.

second segment - established, low margin, high volume.

- highest mean number of reviews by a fair margin, and yet only 3rd in review numbers by median (3rd and fifth categories have wide margin between median and mean)
- lowest average prices by mean, and 2nd lowest (to first segment) in median.
- High review scores, but still 3rd behind fourth and third segment.
- Second highest estimated revenue, behind fourth segment, in both mean and median.

third segment - middle of everything, checks the box.

- middle in price by mean
- middle by review count
- middle estimated revenue
- curiously, highest review scores
- Also tied for highest median listing age, close first for mean listing age.

fourth segment - luxury/upscale/ large listings

- highest accommodation numbers by a wide margin: 10+ vs 4-5 for the other segments
- Highest price by a wide margin: median 407 vs other segments in the 100s.