Project Technical Report

Group 3 - Bionic Blobs

Group Members: Shirley Cho, Miko Le, Alex Reilly
Report Process Date: 5/11/2023-5/26/2023

Contents

Introduction

New York City (NYC) passed a short-term rental law on January 8, 2022 called Local Law 18 to address issues with many illegal short-term rentals in NYC which includes Airbnb rentals. This law prohibits people from renting out an entire house to guests for less than 30 days. This is NYC's way to address the possible skyrocketing housing prices in NYC, an already expensive city to live in.

Our interest in Airbnb data in NYC is to showcase the possible trends between the housing market and Airbnb over the years. Our second aim is to determine internal and external factors that made Airbnb grow so massively today in NYC. These internal factors include the types of amenities, the availability of Airbnbs, the amount of hosts, the variety of housing structures, cost, etc. that contribute to higher bookings. We also aim to postulate some external factors affecting Airbnb bookings such as seasonality, geography, crime rates, airplane prices, tourist attractions, other competing vendors of Airbnb like hotels, etc.

The end goal of our data is to provide an in-depth analysis on the effects of Airbnb on NYC to help organize plans that may mitigate the many illegal short-term rentals in NYC that bolster the housing prices of NYC.

Questions and their answers:

1. What are the most expensive and least expensive areas for Airbnbs in New York City?
   - Manhattan, Brooklyn, Staten Island, Queens, Bronx are the most expensive to least expensive boroughs in terms of average Airbnb price.
2. When is the most popular time of the year to book Airbnb in New York? (seasonal, bring in holidays)
   - Fall, Summer, Spring, Winter from most popular to least popular.
3. What are the allowed average stay lengths?
   - Most notably 1-7 days and 30 days for minimum stays
   - 25% listings suggest less than 365 days for a maximum stay whereas 75% of listings state they are okay with a maximum stay of over 365 days.
4. Count of Airbnbs bookings over time
   - A steady increase over time except for the year 2020, likely due to lockdown.
5. Count of properties per host
   - More than 50% of hosts own two or more properties.

Machine Learning Questions:
1. What amenities affect price most?
2. What factors will affect the review ratings of Airbnb? Can we use them to predict the ratings of Airbnb?

Business Questions:
1. Trends of average price of rental homes in new york from zillow data vs trend of average price of Airbnb in new york to see any trends by month, year.

- Positive correlation, but no real way to see if they affect one another.
2. Are hotels more expensive on average compared to Airbnbs?
   - Yes for all boroughs
3. Does the amount of crime in each borough have any impact on the amount of Airbnb bookings? How about tourist attractions?
   - There are some weak correlations, but probably no real effect/impact on Airbnb bookings.
4. Does Airplane fare prices affect AirBnB bookings and when is that?
   - Negative correlation. However unable to determine effect because there is no data suggesting that the people that booked these flights also booked an Airbnb in NYC

Data Sources and ETL

We used 8 datasets:

| Dataset	Topics of Interest	Format
NYC Airbnb Data	Bookings, Reviews, Price, Hosts, Ratings, Location, Amenities, etc.	CSV File
NYC Crime Data	Date occurred, Offense, Type of Crime, Location	CSV File
Zillow Housing Data	Average price on rental homes, count of Zillow listings, neighborhood, year, etc.	CSV File
NYC Tourist Attractions	Name, Location, Address	Web Scrape, Kafka
NYC Hotels	Hotel Name, Location, High-end Price, Low-end Price, Location, Address, Rating, etc.	CSV File
Airplane Prices	Fare price, destination airport, total distance traveled, total duration traveled, date of departure, etc.	CSV File
NYC Population	Population Count, Boroughs	Website

Data Sources:

1. *Get the Data*. Inside Airbnb. (n.d.). http://insideAirbnb.com/get-the-data
   - Airbnb Data for different Airbnbs across the world. Only used New York City
2. Dgomonov. (2019, August 12). *New York City Airbnb open data*. Kaggle.
   https://www.kaggle.com/datasets/dgomonov/new-york-city-Airbnb-open-data
   - Airbnb Data for New York City Airbnbs such as bookings, reviews, price
3. Diaz-Bérrio, G. (2017, November 16). *New York Hotels*. Kaggle.
   https://www.kaggle.com/datasets/gdberrio/new-york-hotels
   - Dataset that showcases locations of hotels in new york such as average price and postal code
4. (DOF), D. of F. (2021, June 18). *Hotels Properties Citywide: NYC open data*. Hotels Properties Citywide | NYC Open Data. https://data.cityofnewyork.us/City-Government/Hotels-Properties-Citywide/tjus-cn27
   - List of all hotel properties in New York City
5. Samoshyn, A. (2020, July 12). *New York City police crime data historic*. Kaggle.
   https://www.kaggle.com/datasets/mrmorj/new-york-city-police-crime-data-historic
   - Recorded Crime data happening over the years with location and date of complaint
6. *Housing Data*. Zillow. (2023, April 25). https://www.zillow.com/research/data/
   - Housing data offered by Zillow that includes many sources. Only used rental housing average price and number of listings for Zillow
7. Wong, D. (2022, October 18). *Flight prices*. Kaggle.
   https://www.kaggle.com/datasets/dilwong/flightprices?resource=download
   - Flight prices for flights. Only used data for flights going to New York City airports
8. Jain, N. (n.d.). *Top 50 NYC attractions you can't miss*. Attractions of America.
   https://www.attractionsofamerica.com/attractions/new-york-city-top-10-attractions.php
   - A website detailing the address, name, and description of 50 NYC tourist attractions

Other Cited Sources:

1. Wikimedia Foundation. (2023, April 3). *Timeline of Airbnb*. Wikipedia.
   https://en.wikipedia.org/wiki/Timeline_of_Airbnb
   - Wikipedia timeline of important events in Airbnb history
2. Herries, J. (n.d.). *New York population density*. ArcGIS Hub.
   https://hub.arcgis.com/maps/80f9b95a4ce0491091f1477710f6a91d
   - Population Density Heat map for NYC
3. *2023 Major Daily Holidays by Month*. Holiday Insights. (2023, January 1).
   https://www.holidayinsights.com/everyday.htm
   - List of holidays
4. http://www.citypopulation.de/en/usa/newyorkcity/
   - New York City population density by boroughs
5. https://rapidapi.com/trueway/api/trueway-geocoding
   - Api for geocoding to get latitude/longitude
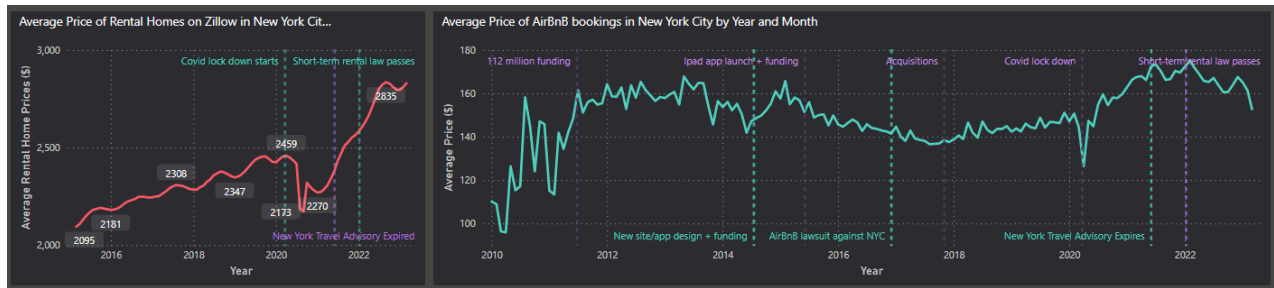
Visualizations
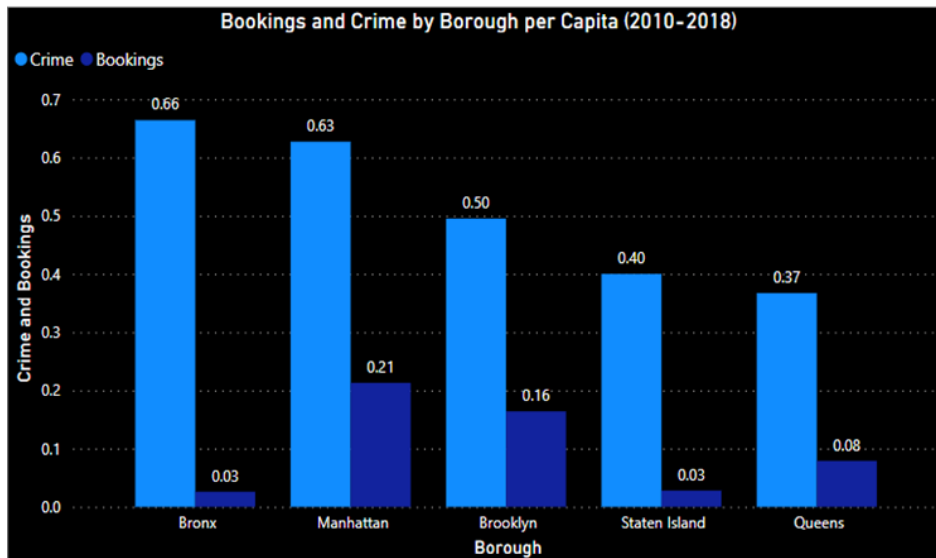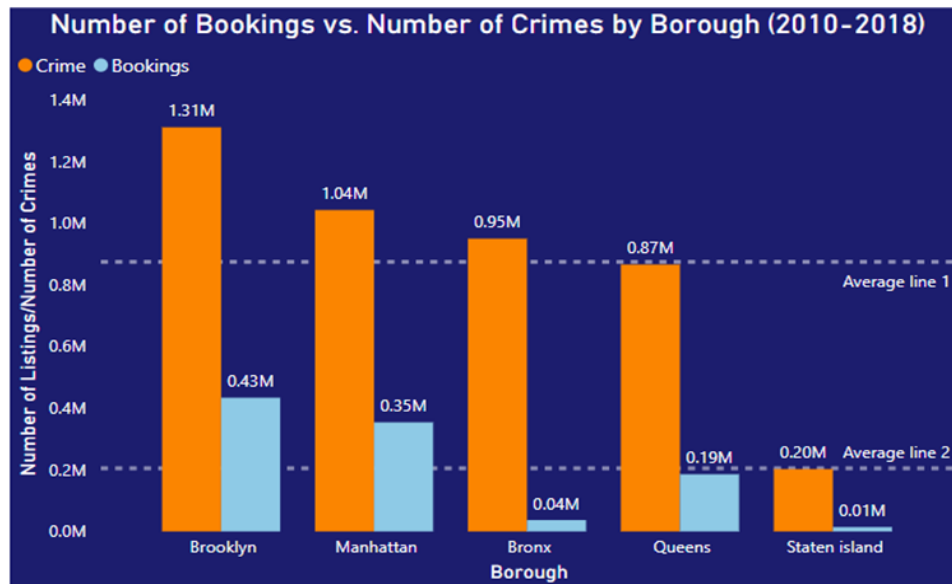
The first visual displays the comparison between the number of bookings and the number of reported crimes in New York City by boroughs. It is sorted by the number of crimes in descending order with the attempt to understand the relationship between the two data. For some boroughs, it seems that a higher number of crimes is leading to higher bookings. However, this trend was not consistent throughout all boroughs so a solid correlation could not be established. Our theory for the reason they data turned out like that is due to population density. It makes sense if there are more people in certain burrows, the number of both crimes and bookings can go up without having any correlation with one another.

The second visual showcases the ratio of both crime and bookings. This was done by dividing all the numbers for each borough by the population of those respective boroughs. In doing so, we were able to get the number of crimes and bookings per capita or per person. Even with this ratio, it is still difficult to determine any correlation between the crime and bookings. If we compare the ratios from Bronx and Manhattan, we can see that Bronx has the highest crime rate and the lowest bookings rate. Manhattan on the other hand has a significantly higher bookings rate (the highest) even though the crime rate is only .03 lower than the Bronx. After analyzing the charts by boroughs, it seems like there is little to no correlation between Airbnb bookings and crime rate in New York City.
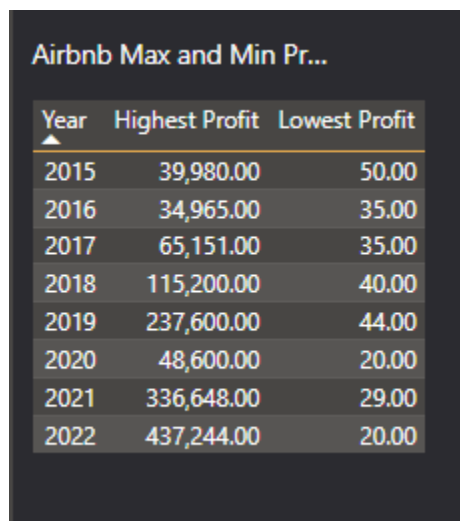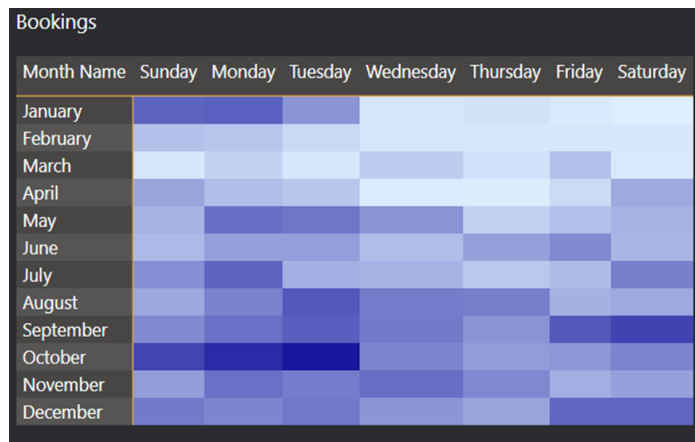

These are visuals that show the number of Airbnb bookings and number of crimes over time. By comparing these visuals, we can see which day or month has more crime or bookings. We found in our seasonality analysis that Fall season has the most bookings and that can be reflected here. In the bookings visuals, the month October seems to be the busiest month for Air bookings. Following that is January, this is logical as it falls under the holiday season with Christmas and New Year. With that being said, when it comes to deciding whether a correlation exists; it is still inconclusive. Let's look at both the crimes and bookings during the month of October. They both recorded a high number during that month. However if we look during March, only crime is high in number and bookings are at one of the lowest in the year. With this inconsistency with the overtime data, there is not a clear relationship between bookings and crime rate in NYC.

This visualization shows an overview of NYC Airbnb fluctuations over the years based on its average price of its bookings from 2009-2023 as well as the fluctuations of average price for rental homes on Zillow for NYC. These diagrams' main takeaway is the decrease in average price in 2020 that is most likely due to the Covid-19 pandemic for both charts and the rising housing price from 2021+ that rises along with Airbnb's average price from 2021+. The short-term rental law gets passed in 2022 and enforced in 2023 in which we see a decrease in Airbnb's average price as the short-term rental law is a law that basically forbids entire homes being listed on Airbnb for NYC thus removing many listings from Airbnb in 2023. We also see Zillow's housing price stop increasing shortly after. We don't have much data on 2023 so we cannot conclude any strong relationships between the two, but noted the recent positive uptrending together.
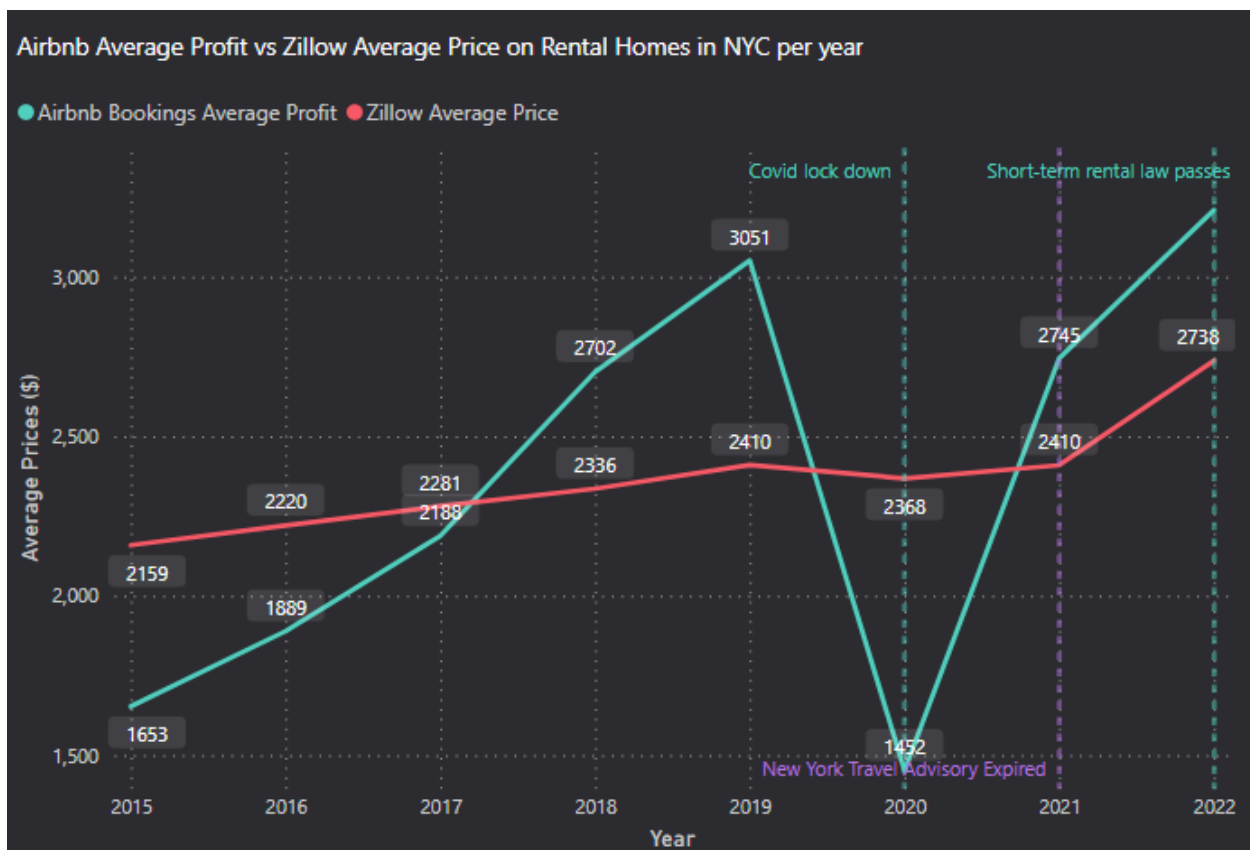
This visualization tries to compare whether Airbnb on average is more profitable than rental homes. The Airbnb trend line was made by taking all the bookings from 2009-2023 and filtering on only the room types being 'entire homes', 'entire apartment', 'entire condo', 'entire townhouse', etc. anything that sounded like an entire home and not a private room since those are cheaper and not comparable to Zillow rental homes in terms of price. Then I aggregated each by listings and years to get the sum of the profits that each listing did for that year. Then I averaged those values to get the average of profit per year for Airbnb bookings. We see that past 2017, the Airbnb bookings on average have more profit compared to the average Zillow rental housing prices. If we ignore the covid lockdown in 2020, we also see both Airbnb average profit and rental homes average prices trend upwards together for all years. In 2021, we see the sharp increase of Airbnb profit also showing a sharp increase in the average price of Zillow rental homes listings. Then in 2022, we will see short-term rental law passed. Unfortunately, we do not have enough 2023 data to further peak into the effects of this law on both Airbnb and rental homes.
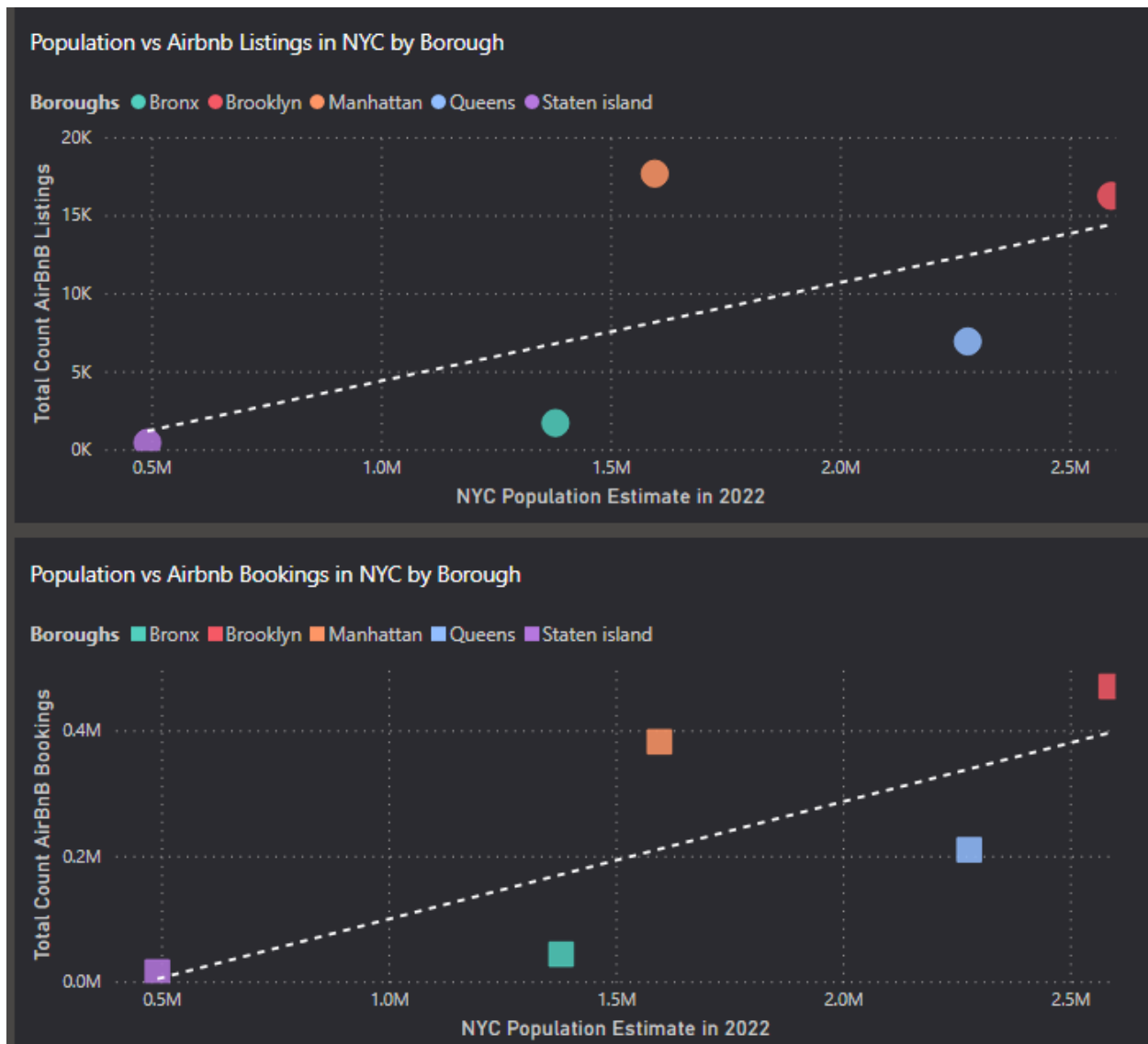
Airbnb Max and Min Pr...

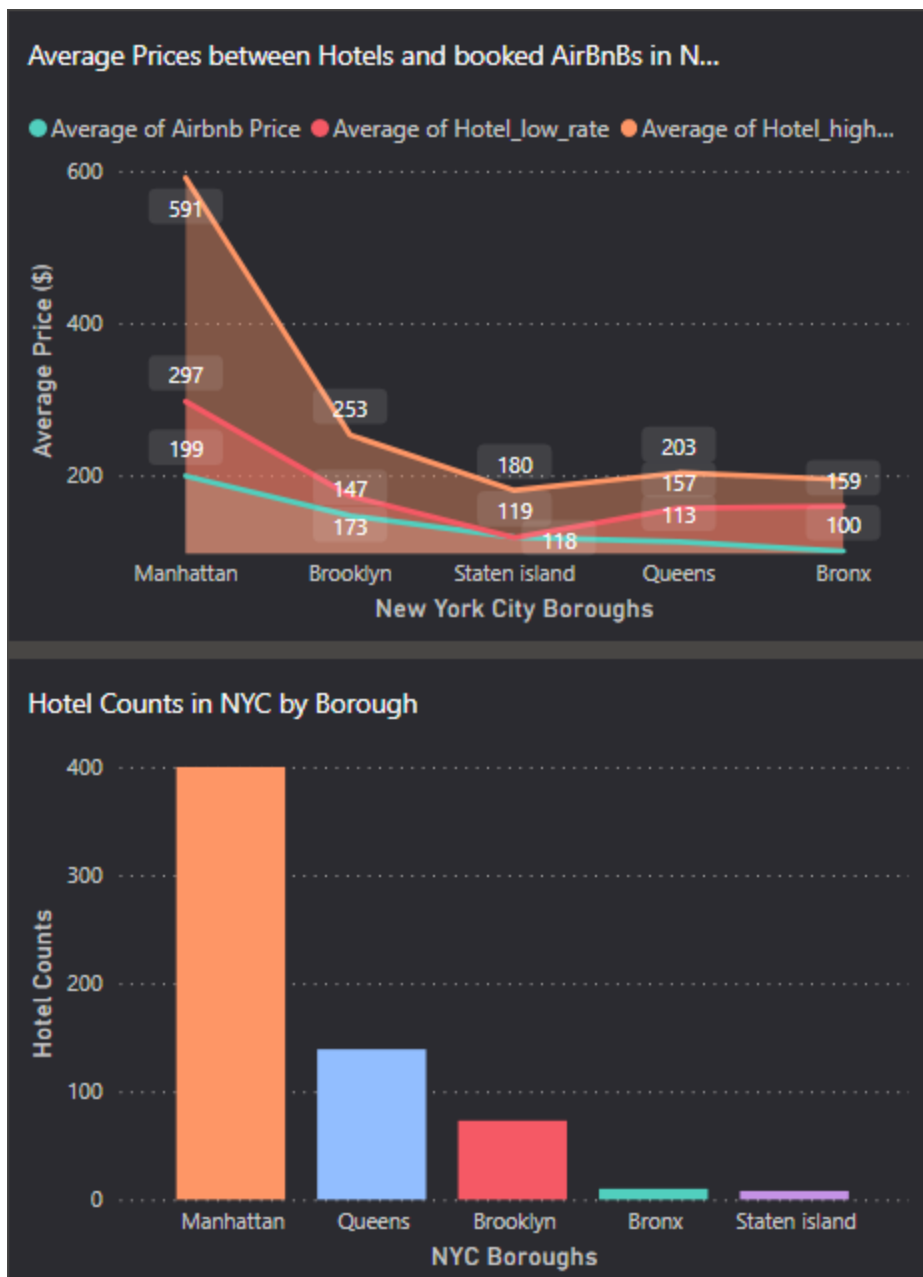| Year	Highest Profit	Lowest Profit
2015	39,980.00	50.00
2016	34,965.00	35.00
2017	65,151.00	35.00
2018	115,200.00	40.00
2019	237,600.00	44.00
2020	48,600.00	20.00
2021	336,648.00	29.00
2022	437,244.00	20.00

This visualization is a supplemental illustration at the yearly profit of the highest earner listing and the lowest earner listing for all the years. Showing that despite the average in the prior illustration being 2000-4000 price range, that there is such a wide range of earnings by Airbnb.

Count of AirBnB bookings vs Zillow Home Listings in NYC by years
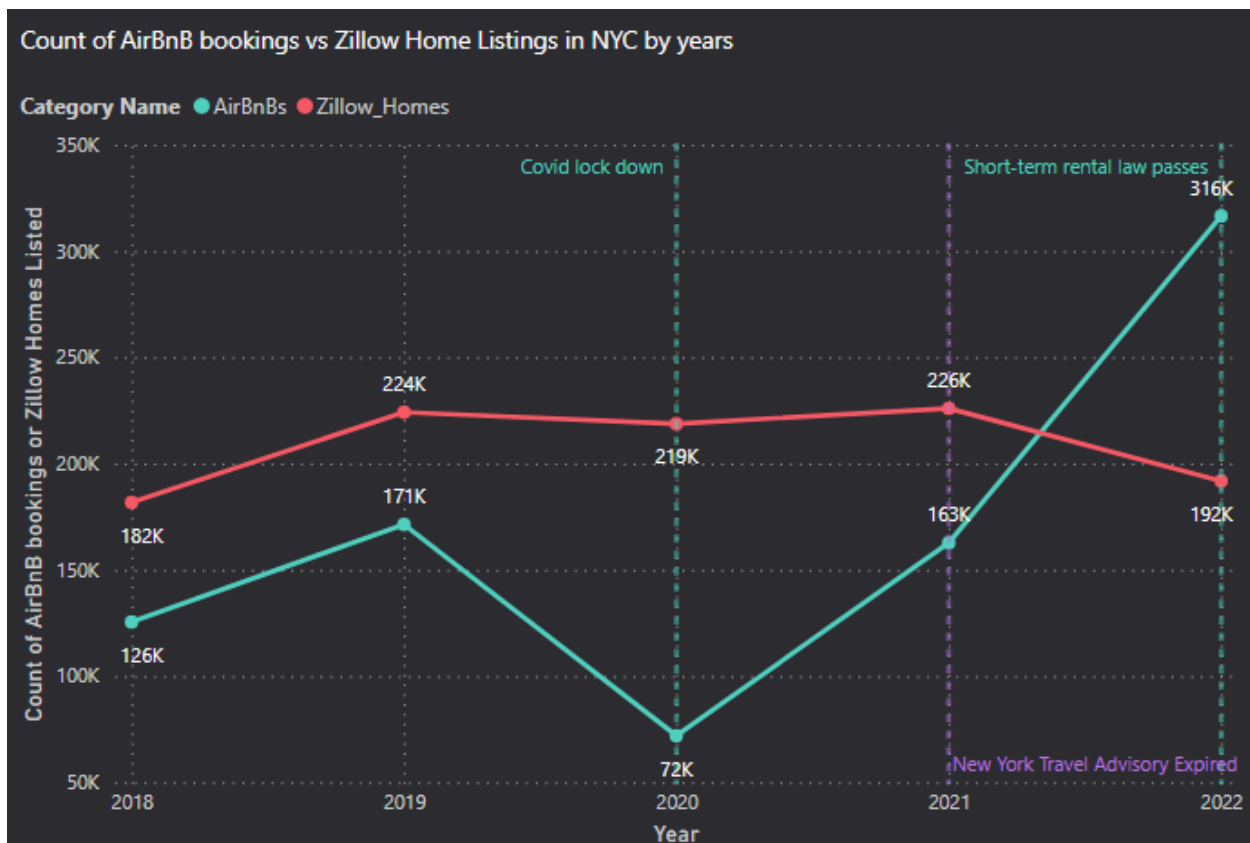
This visualization is the count of Airbnb bookings vs the count of Zillow home listings by each year. I charted these two together to compare if there is also a trend where more Airbnb bookings and the rise of popularity and usage of Airbnb would also result in a decrease in Zillow Home Listings. We see in 2018, the counts for both rise together and fall together during the covid pandemic. As lockdown ends in 2020, we start seeing an increase in both the Zillow homes and Airbnb counts, however as Airbnb bookings dramatically increase, we then see Zillow home listings count go down from 226k-192k. It is interesting to see the positive correlation from 2018-2021, but then a negative correlation from 2021-2022. There isn't enough information to determine if the increase in Airbnb bookings from 2021 is what caused the decrease in Zillow home listings.

Population vs Airbnb Listings in NYC by Borough

Boroughs ● Bronx ● Brooklyn ● Manhattan ● Queens ● Staten island

Population vs Airbnb Bookings in NYC by Borough

**Boroughs** ■ Bronx ■ Brooklyn ■ Manhattan ■ Queens ■ Staten island

These two listings are meant to show the effect of population density of borough locations on the number of listings and number of bookings for Airbnb. The question they try to answer is if population density is an important factor for the locations people choose to book an Airbnb at. Bronx and Queens are below the trendline, which suggests that for their population density, they have a below average amount of Airbnb bookings and listings. However, Manhattan is above the trendline which suggests that for its population density size, it has an above average amount of Airbnb bookings and listings. An interesting thing to note here is that while Manhattan has the most listings, they do not have the most bookings; that is actually held by Brooklyn.

Average Prices between Hotels and booked AirBnBs in N...

● Average of Airbnb Price ● Average of Hotel_low_rate ● Average of Hotel_high...

Average Price ($) vs New York City Boroughs

Manhattan: 591, 297, 199
Brooklyn: 253, 147, 173
Staten island: 180, 119, 118
Queens: 203, 157, 113
Bronx: 159, 100

Hotel Counts in NYC by Borough

Hotel Counts by NYC Boroughs: Manhattan ~400, Queens ~140, Brooklyn ~75, Bronx ~small, Staten island ~small

I took a look at other competitors of Airbnbs in New York City, namely hotels. Earlier we saw that Manhattan held the second highest number of bookings, despite being the highest in number of listings. There can be many factors causing this trend, but looking at hotel count, we see that Manhattan has the highest number of hotels by far and Brooklyn that took the highest amount of Airbnb bookings has a much lower hotel count. This could indicate that Manhattan had a lower booking count because people may be choosing to book into hotels over Airbnbs whereas Brooklyn doesn't have as many options of hotel choice as Manhattan and people might have to resort to booking at Airbnbs for Brooklyn. It is interesting to postulate that New York City could look into refurbishing their hotels as a possible choice to impact Airbnb bookings over something like the short-term rental law they passed. However, we also see that hotels on

average tend to be more expensive than Airbnbs across all boroughs. So further research on what features or amenities of hotels that make them attractive could be made to see if hotels truly had an impact on lowering Airbnb bookings because there isn't a strong correlation between the two.
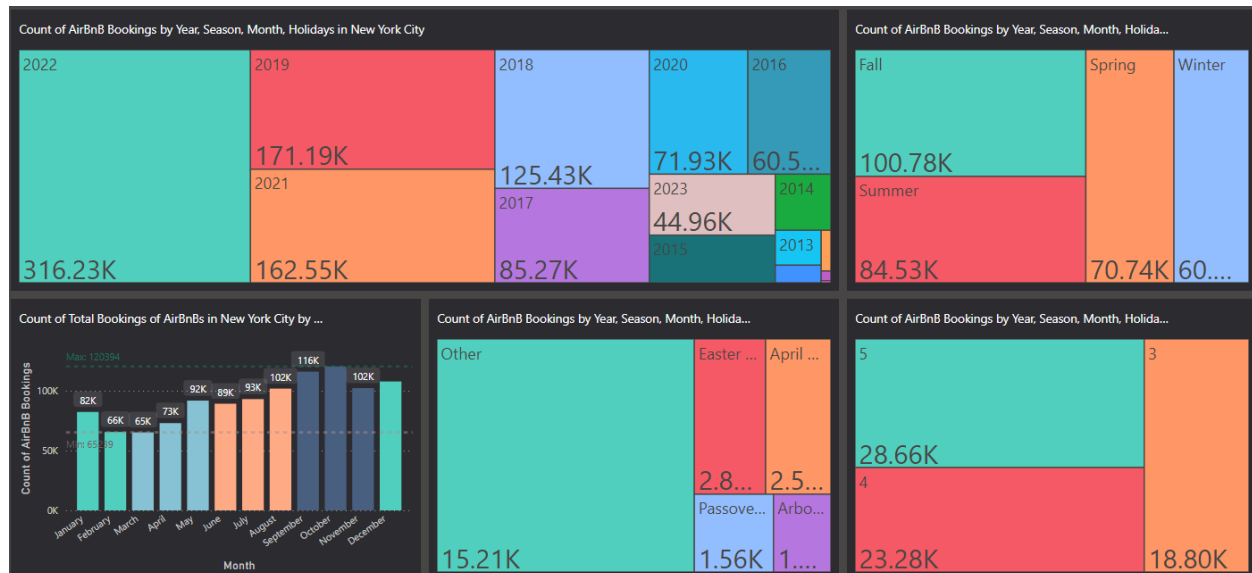

A tree map that breaks down the Airbnb bookings by year, season, month, and holidays.

The notable trends are that all the years have an increase in bookings except for 2020 and 2023 because 2020 most likely had less bookings due to covid lockdown and 2023 has less bookings because our data cuts off at march 2023. The most popular seasons for Airbnb bookings are Fall and Summer. There are no real conclusive trends about holidays and Airbnb bookings.

Possible Inverse Relationship between NYC AirBnB Booking Count and NYC Airplane Fare in 2022 per Months

● Proportion Air fare per distance travelled in 2022 ● Proportion of Count of AirBnB bookings in 2022 / max count

This visualization shows the negative correlation between Airbnb booking counts in NYC and airplane fare prices to NYC through the seasons. We again note the summer and fall seasons have the higher counts of bookings and we also see that negative correlation of summer and fall also having the cheapest airplane fare prices. This helps bolster that the seasonal trends for Airbnb bookings have some correlation with airplane fare prices. It is difficult to determine if the price of airplane fare causes more bookings or less bookings unless we have more data on if the person purchasing these tickets also chose to book an Airbnb or not.

Hosts by Number of Listings
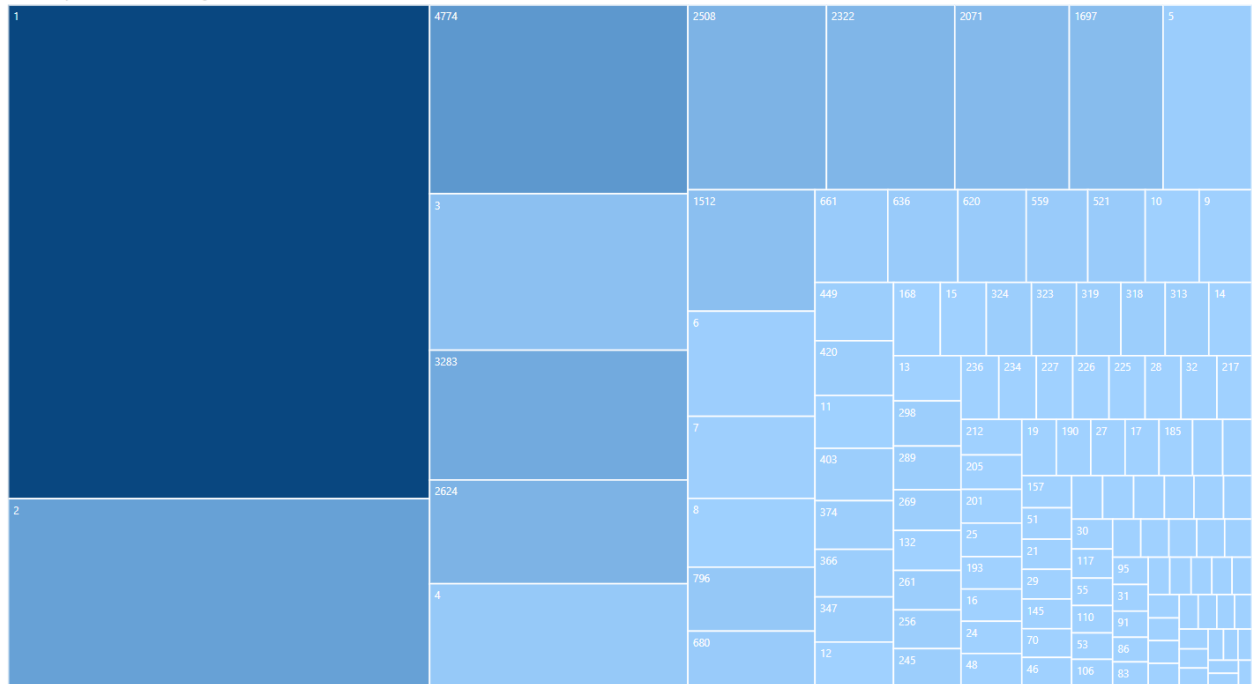
Since one of Airbnb's main marketing strategies is showing interpersonal relationships between hosts and tenants, we decided to examine the breakdown of how many listings each host has. Although 1 is the most popular number of listings per host, the majority of hosts have more than one listing, and the maximum is 4,774. The 10 hosts with the most listings have a combined total of approximately 15,000 listings.

Count of available listings and Average of price by Date
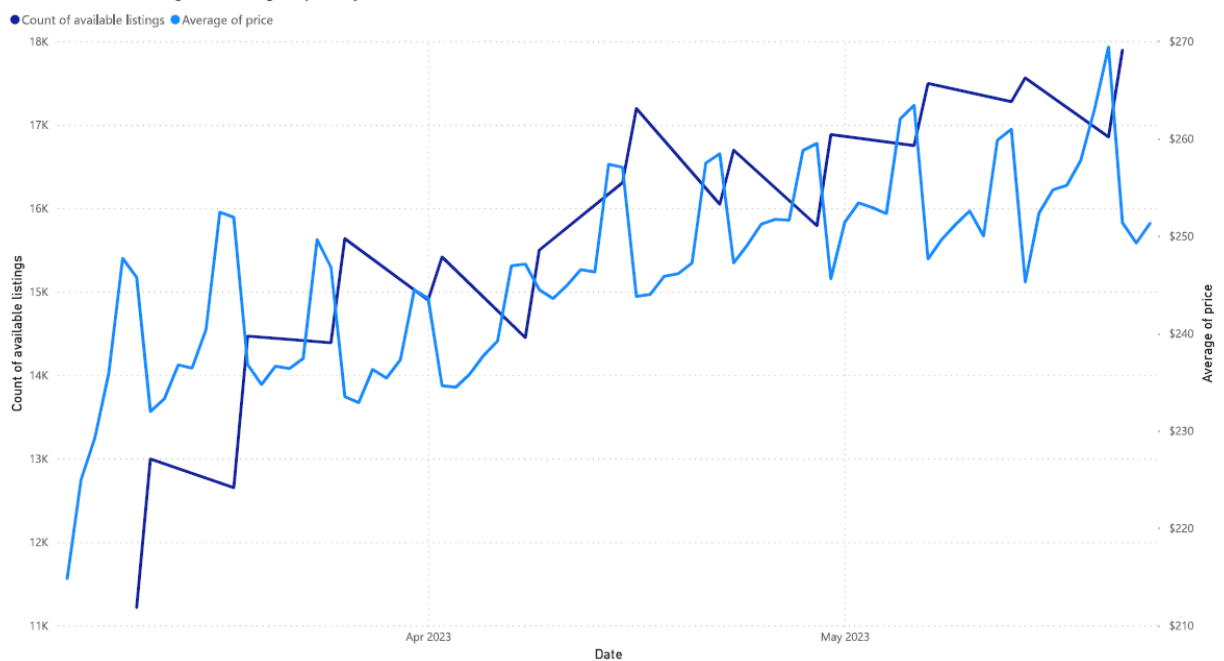
We examined the data for early 2023 and found periodic patterns with sharp increases and decreases in the count of available listings and the average price day-by-day. The two items fluctuated inverse to each other, so we homed in on the data to see what might be causing the pattern.

Average of price by Date
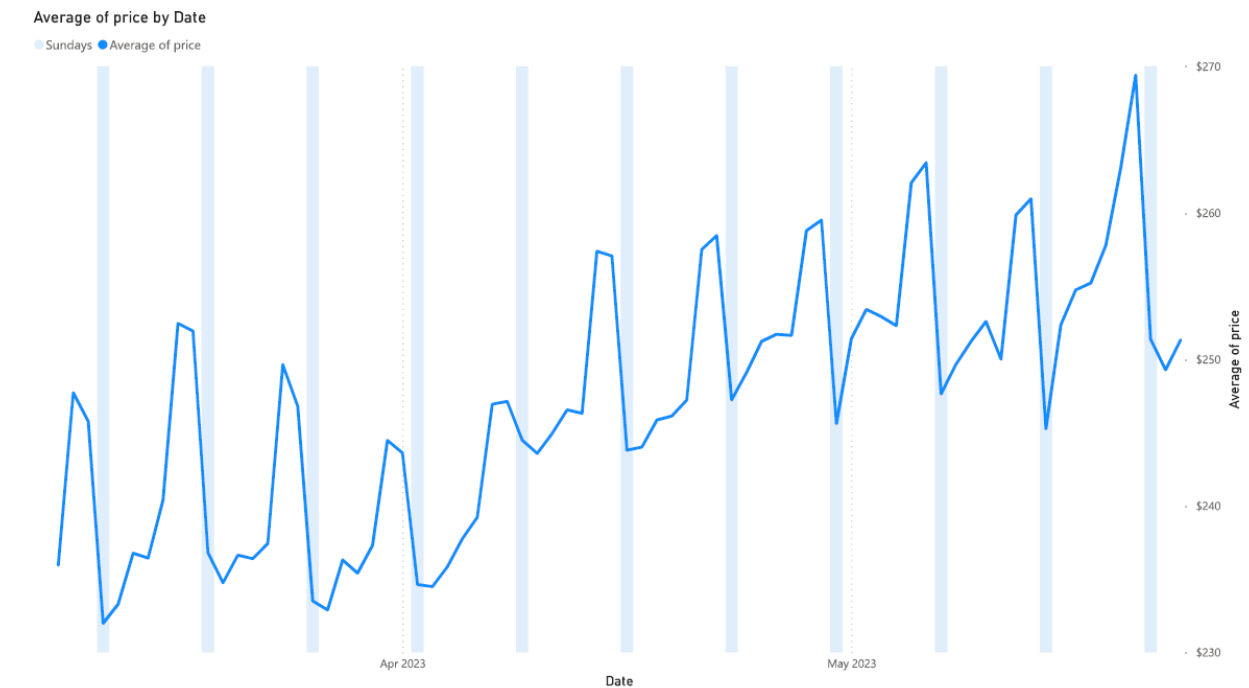
Upon examining the relationship between the day of the week and the previously discussed metrics, it became apparent that the sharp price drops occurred every Sunday, as availability increases.

Machine Learning

## Reviews Ratings

Actual Review Rating vs. Predicted
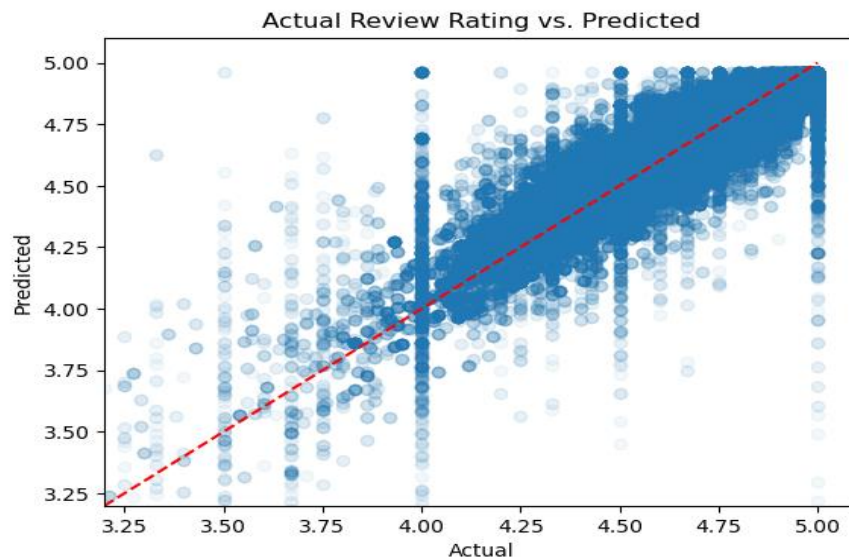
For this model, we used the LinearRegression algorithm on the Airbnb review rating score in New York City; with the range from 2009 to 2023. The Airbnb review ratings are scored 1-5, however there are not many listings with 1 or 2 ratings. With that, we chose to create a machine learning model that focuses primarily on predicting the higher review scores of 3.2-5.

The model returned a test score of 0.846 indicating that the model is able to predict the ratings correctly about 84.6% of the time. It also has a mean absolute error of .0536 so on average, the model's predictions are off by .0536 ratings point. Furthermore, this model returned the mean squared error of .006, this score demonstrates the squared difference between the actual ratings and predictive ratings. There are a few wrong predictions with this model, which can be seen most dominantly in 4 stars rating. We are unsure of the reason for this but with decently high test scores and low MSE and MAE scores; this is a good model to predict the performances of ratings.

Amenities

We also performed a supervised machine learning analysis on amenities and prices of the listings using a decision tree classification model. In order to build a robust model, we removed amenities that had fewer than 600 associated listings. We used Price as our target variable and Amenities as our predictor variables.

Decision tree trained on amenities
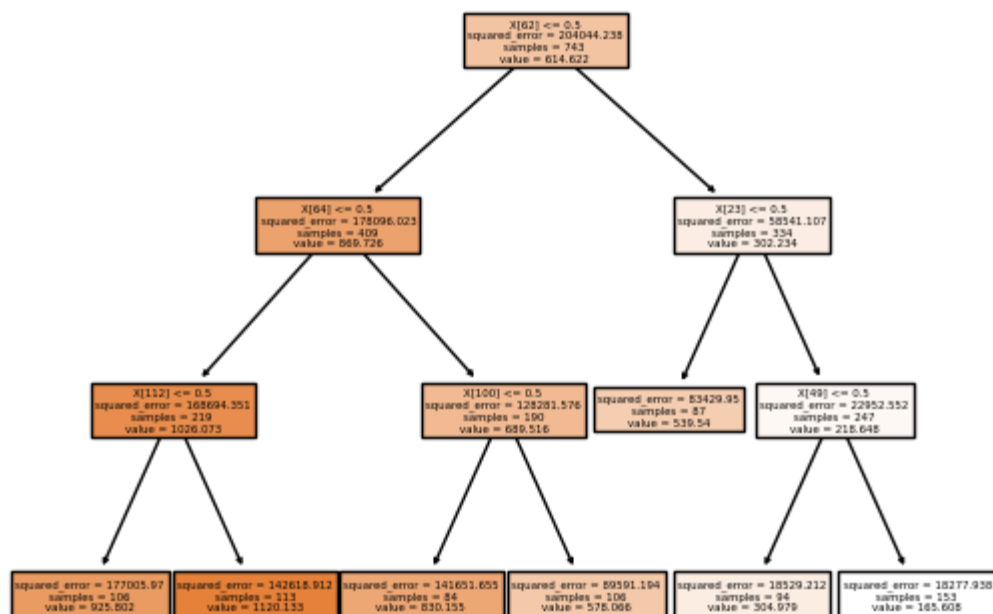
In order to reduce variation, we ran the test data 1,000 times with random states and took the average of the score.

```
#multiple runs aggregated

NumRuns = 1000

RunList = []



for i in range(NumRuns):

    X_train, X_test, y_train, y_test = train_test_split(X, y, random_state
= random.randint(1,100000000))

    dtree.fit(X_train,y_train)

    n_score = dtree.score(X_test,y_test)

    RunList.append(n_score)
```

```
range_RunList = min(RunList), max(RunList)

average_score = sum(RunList)/NumRuns



range_RunList, average_score
```

An anomaly that we found intriguing and would like to learn more about was that the model was slightly underfitted; the testing data consistently performed slightly better than the training data. We extracted a test score of .55, which is not a strong correlation, however, within the low-stakes context, we considered it acceptable.

We then attempted to boost the score with an AdaRegressor, which classifies data in small increments multiple times, assigning higher weights to misclassified data. Essentially, it "learns" from its classification mistakes. The new score was .27, and this method was deemed ineffective.

Conclusion

As a group, we came to the following broad conclusions regarding our data:

- Fall is the most popular season for Airbnb bookings.
- Airfare prices to NYC have an inverse relationship to Airbnb booking.
- Housing listings decrease as Airbnb bookings have a sharp increase.
- Crime does not have an impact on Airbnb bookings.
- The majority of Airbnb listings are long-term stays.