TEAM PROJECT

MSIN0093

Retrieved from: https://www.bbc.co.uk/news/technology-52184497

Wordcount: 3924

Table of Contents

**The restart option was exercised on the 3rd of November.*

1. Introduction

Founded in 2008, Airbnb is an online vacation rental service. As a platform, their website focuses on connecting travellers who are looking for accommodation besides conventional hotels, with local hosts who are looking to rent out space in their own homes. Airbnb is now available in 220 countries and has grown as an organisation, competing with major hotel chains globally.

However, the last year has seen a deterioration in the hospitality sector due to COVID-19. For this project, we are acting as consultants for Airbnb's executives, using pre-existing listings data to understand how COVID-19 affected listings in New York and to create an internal price prediction model for new hosts.

Currently, Airbnb does not offer a price predictor and instead asks their hosts to compare their offerings with other similar homes to determine a price as seen below from their website FAQs.

How should I choose my listing's price?

The price you charge for your listing is completely up to you. To inform your decision about what price to set, you can search for comparable listings in your city or neighbourhood to get an idea of market prices.

Retrieved from https://www.airbnb.co.uk/help/article/52/how-should-i-choose-my-listings-price

First, we define the dataset and the process of collecting and cleaning the data. We then perform an exploratory analysis to develop a better understanding of the differences between the 2019 and 2020 Airbnb listings. To expand our knowledge of the impact of COVID-19 on the major attributes of properties, we conduct additional descriptive statistics. Lastly, we create a price prediction model. Airbnb can use the model as a new feature on their website to offer a recommended price range for hosts putting up their listing for the first time. Initially, we conducted an OLS regression, to better understand our data and to analyse the coefficients. However, to increase the performance of our model, we conducted an additional Gradient Boosting Model which will ultimately be used for the price prediction feature. Using these methods and different types of analysis, we aim to understand the effects of COVID-19 on NY's listings and create a preliminary price prediction model.

2. Dataset

2.1 Description

The dataset contains data from Airbnb listings in New York. The dataset consists of two components. The first contains Airbnb listings from February to September 2019, and the second contains Airbnb listings from February to September 2020. February was chosen as the starting month for our 2020 dataset as COVID-19 had not yet had a substantial impact on the global economy. September was chosen as the end month as it was the latest data available. The same months were chosen in 2019 in order to analyse the impact COVID-19 had on Airbnb properties in New York.

The dataset contains a lot of information surrounding the Airbnb listings. The most revealing variables which are referenced throughout this report are the following:

Description of Main Dataset Variables

neighbourhood_group_cleansed	Classifies listings into the main areas of New York, including Staten Island, Brooklyn, Bronx, Manhattan and Queens.
property_type	Categorises listings into the main property types. These include condominiums, apartments, hotels, houses and other. Other consist of miscellaneous options such as boats or tents.
room_type	Describes the room type. The room types are private room, shared room, hotel room and entire apt/house.
accommodates	Reveals the number of people the listing accommodates.
bathrooms	Shows the number of bathrooms in the listing.
bedrooms	Shows the number of bedrooms the listing has.

price	Gives the price per night for the entire listing in USD.
security_deposit	Gives the amount of security deposit needed to book the listing in USD.
cleaning_fee	Gives the amount of the cleaning fee needed to book the listing in USD.
availability_365	The number of days the listing was available for guests to book in the past 365 days.
number_of_reviews	The total amount of reviews the specific listing has.
review_scores_rating	The current review score the listing has out of 100 total points.
reviews_per_month	The total amount of the reviews the listing has, divided by the total amount of months the listing has been on Airbnb.

2.2 Data collection process

The data was collected from the website http://insideairbnb.com/. It collects data from the Airbnb website every month. In order to attain the dataset, we downloaded the respective months in 2019 and 2020 separately as csv files. After including a column which stated the month from which the data stems, we concatenated the files together to create a 2019 and a 2020 dataset.

2.3 Data processing

In order to prepare the data for analysis and modelling, significant data cleaning was required. First, we deleted several columns. From the original 80 columns, 47 were deleted as they were deemed not valuable for our analysis.

Further cleaning steps included organising the variables into their correct data format and deleting unnecessary strings in the data. In addition, missing values were filled with the median values of their respective columns.

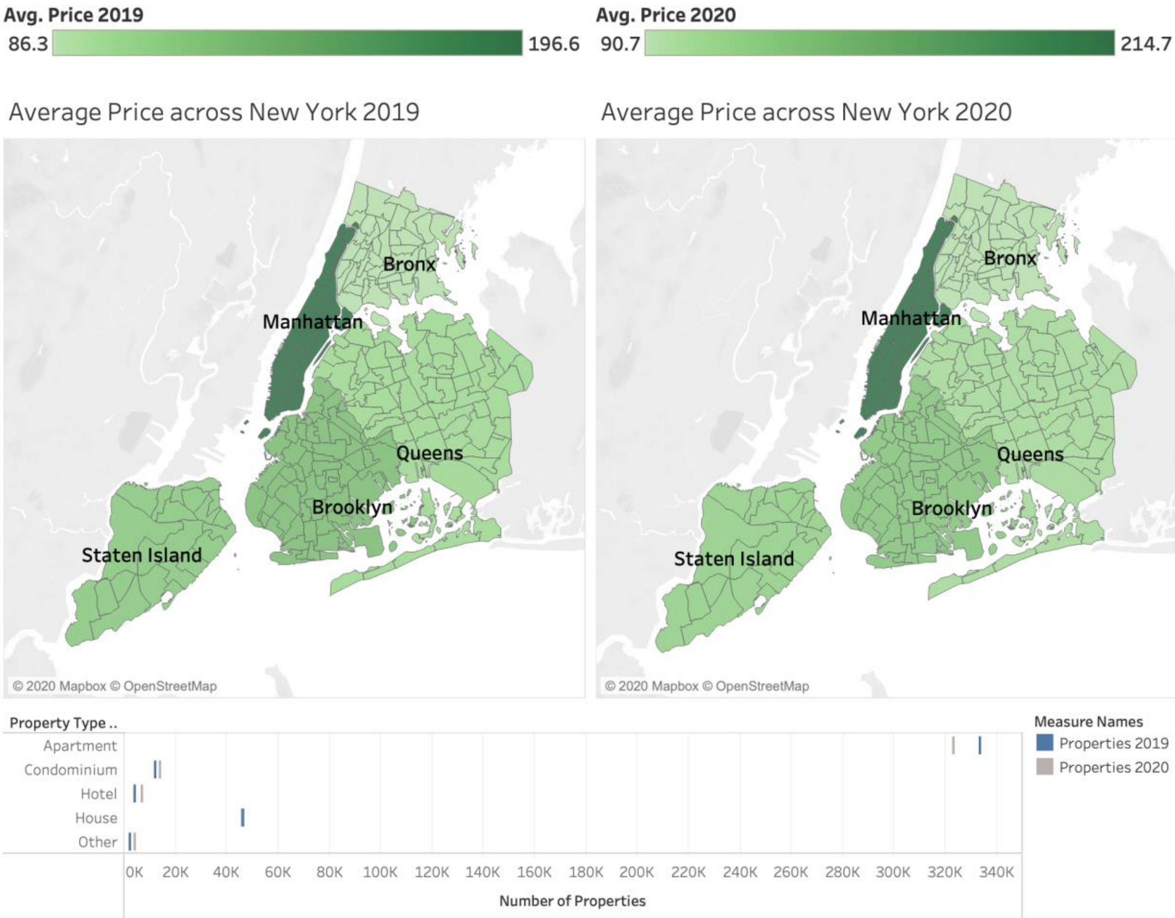
In order to achieve the grouped property types and room types, the many different accommodation types were mapped to four main categories for property and room type.

3. Analysis

3.1 Exploratory Analysis

In order to explore the data, we created visualisations to understand the data and generate insights about the variables and distribution within the dataset.

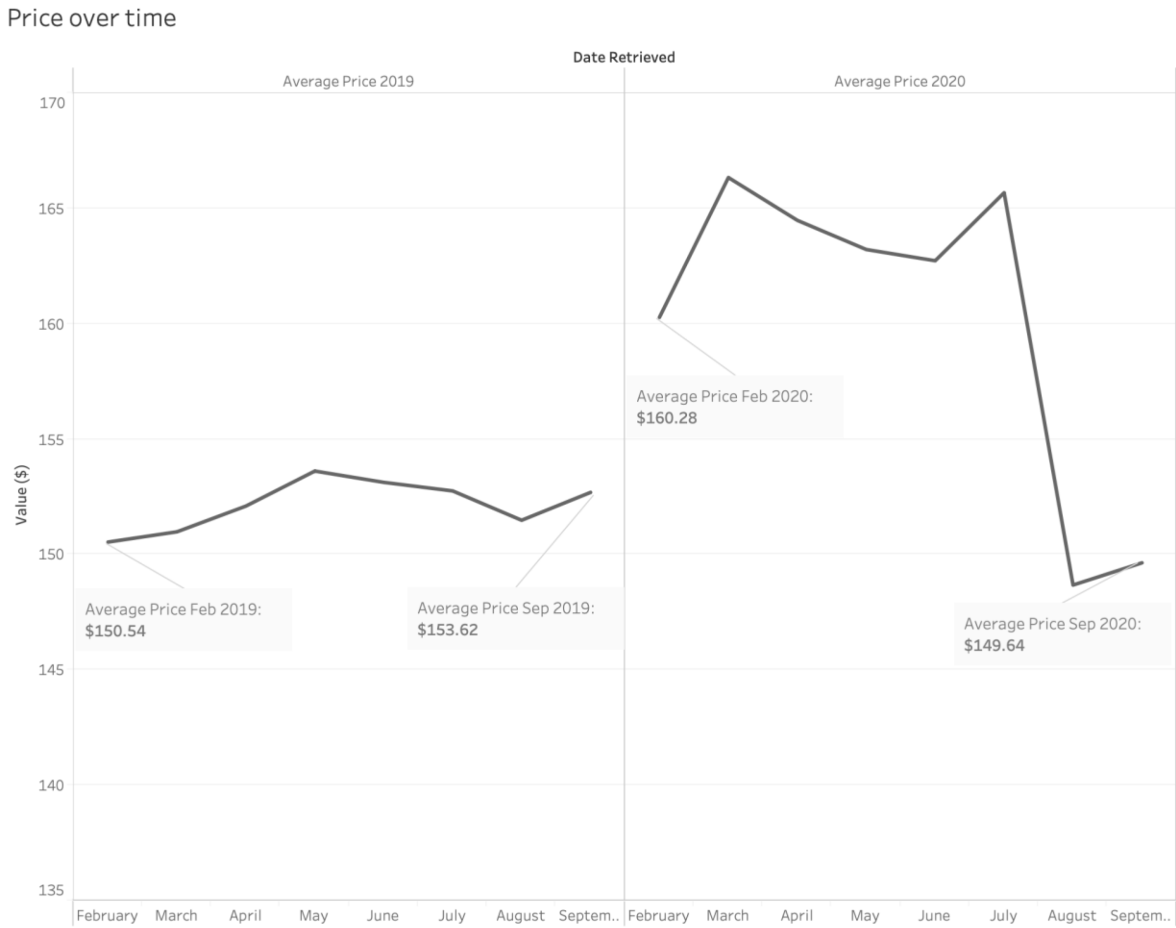
Figure 1: Price by Neighbourhood



The dashboard above displays the average price, split by the different neighbourhoods in New York. It showcases how Manhattan is the most expensive, while Bronx and Queens are cheaper. Also, it shows that although the relationship between the different neighbourhoods does not seem to change, there is a slight increase in the overall average prices from 2019 to 2020. Lastly, the number of properties across the different property types is depicted at the bottom to get an understanding of the different properties variations. For apartments, condominiums, hotels and other types, there has been a slight decrease from 2019 to 2020, while houses have not changed and apartments have seen an increase.

Building on the previous observations, we wanted to further investigate how the average price has been impacted from 2019 to 2020. In addition, we wanted to explore whether the pricing of properties was affected by seasonality or not.

Figure 2: Price variation 2019 & 2020

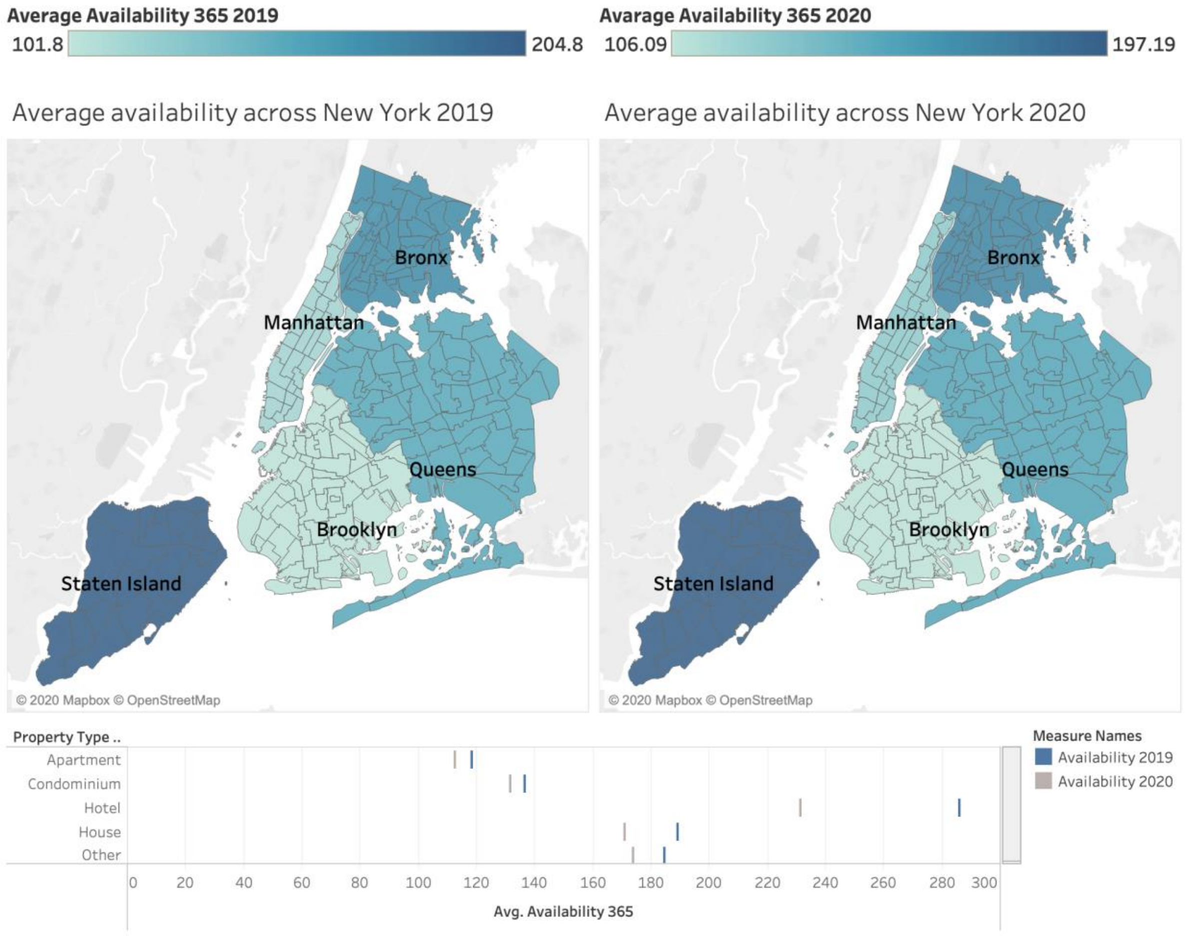


The graph above shows the movement of price from February to September in 2019 and 2020. For 2019, we can see that price increased slightly from $150.54 to $153.62 across the entire time-period, with a peak starting in May and generally higher prices across the summer months. It also shows that in our final month for 2019 it had an upward trajectory. Switching over to

2020, we can see more movement and at a different magnitude, possibly reflecting the uncertainties in the market due to COVID-19. It started with an upward trajectory from February, however, as cities started to go into lockdown in March, prices continued to decline slowly before seeing a peak in July months. From July to August, the prices dropped substantially, until an upward trajectory starting from August, as in 2019.

After getting a broad understanding of the price variable and its relationship with the seasons and neighbourhoods, we wanted to explore another important aspect, the availability of the properties.

Figure 3: Availabilities by Neighbourhood

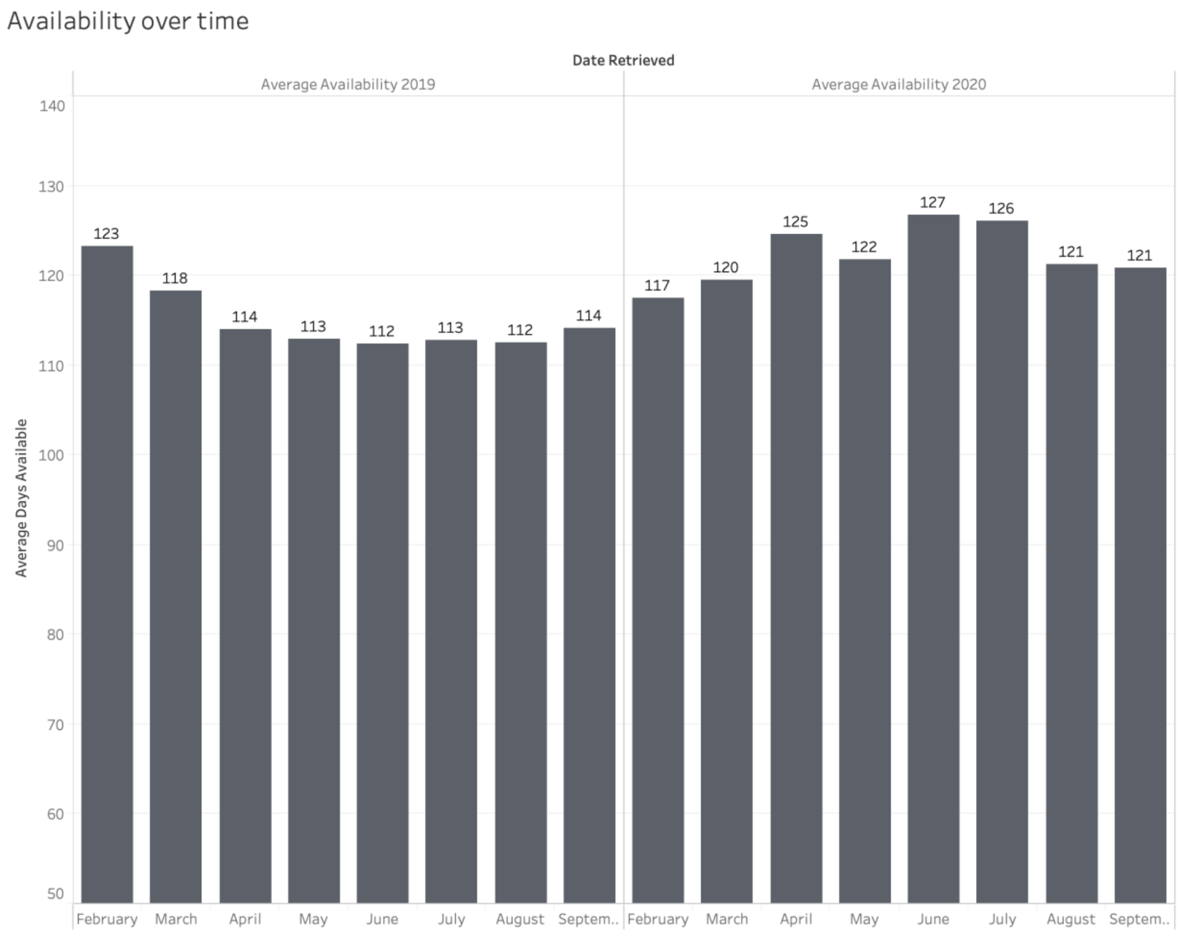


The dashboard above displays the average number of days properties are available on Airbnb, per neighbourhood. Staten Island has the properties with the highest availability, while Brooklyn properties have the lowest availability. The ratio between the neighbourhoods stayed the same across both 2019 and 2020, similar to prices. Looking at the availability split across property types on the bottom, we can see a decrease of average availability across all types from 2019 to 2020, with the largest decrease from hotels.

Overall, we have seen that there has been a decrease in average availability across the different property types. However, as we saw previously, there has also been an increase in the number of properties.

Continuing from the previous observations, we decided to inspect the average availability throughout the dataset from February to September, for both 2019 and 2020, to understand how the average availability has varied.

Figure 4: Availability variations 2019 & 2020



The graph above depicts how the average availability had a decreasing trend for 2019, with the lowest month being June. However, there was a slight increase in availability for 2020. Again, we see a somewhat chaotic distribution for 2020 instead of the smoother 2019 distribution. This is most likely due to the uncertainties of COVID-19 and the hosts not being sure how or whether to list their properties. In addition, we can conclude that although each type of property saw a decrease in their average availability, because of the increase in number of properties, the total availability has increased.

To further understand the relationships between the variables and price, we have conducted further analysis.

3.2 Descriptive Analysis

In order to analyse the impact of COVID-19 on the major attributes of properties, we performed summary statistics of the New York data in 2019 and 2020, outlined in the tables below.

Table 1: General Summary Statistics

	Review Scores Rating		**Availability 365**		**Reviews per month**		**Price**	
Year	**2019**	**2020**	**2019**	**2020**	**2019**	**2020**	**2019**	**2020**
Mean	94	94	115	122	1.5	1	152	160
St. deviation	8	8	133	140	1.5	1.2	237	403
Minimum	20	20	0	0	0	0	10	9
Maximum	100	100	365	365	68	66	10000	10000

From the summary statistics table above, we can obtain the following observations:

- The mean of availability 365 and price have increased from 2019 to 2020, while review scores rating has stayed the same.
- The mean, standard deviation, and the maximum of reviews per month decreased from 2019 to 2020.
- The standard deviation of availability 365 and price have also all increased from 2019 to 2020.

This indicates that availability of rooms were higher and prices were increased whereas number of monthly reviews were decreased. Furthermore, we can observe a sharp increase in standard deviation for price, which could highlight the uncertainties around pricing due to COVID-19.

Table 2: Summary Statistics by Room Type for Price

Room Type	Entire Home/ Apartment		Hotel Room		Private Room		Shared Room	
Year	**2019**	**2020**	**2019**	**2020**	**2019**	**2020**	**2019**	**2020**
Mean	211	207	236	312	89	108	70	90
Standard deviation	275	310	216	907	167	469	103	418
Minimum	10	9	27	38	10	10	10	10
Maximum	10000	10000	100	10000	10000	10000	1800	10000

Table 3: Summary Statistics by Room Type for Availability 365

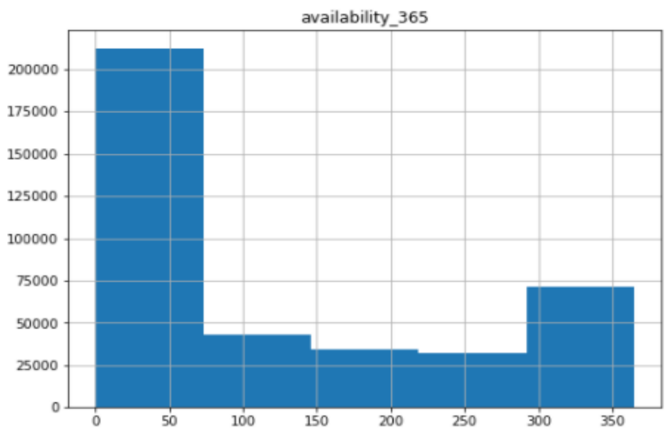
Room Type	Entire Home/ Apartment		Hotel Room		Private Room		Shared Room	
Year	**2019**	**2020**	**2019**	**2020**	**2019**	**2020**	**2019**	**2020**
Mean	114	123	275	257	112	116	162	168
Standard deviation	131	139	117	123	132	138	149	153
Minimum	0	0	0	0	0	0	0	0
Maximum	365	365	365	365	365	365	365	365

The availability variable shows how many days of the year hosts actually put up their properties for rent on Airbnb. If we investigate the summary statistics per room type and take the price and availability dimensions into account, we get some interesting findings. We see that NY's average price per apartment decreased where the average price of private, shared, and hotel rooms increased from 2019 to 2020. However, testing the mean difference in price between 2019 and 2020 revealed that the difference was not significant.

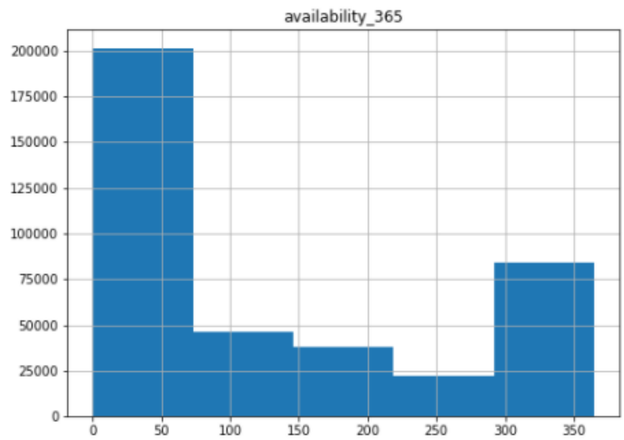
Additionally, the average availability of property types in NY increased from 2019 to 2020 except for hotels which decreased. This occurred as hotel listings were only introduced on Airbnb in October 2019. Therefore, our dataset includes only one month of hotel data in 2019 but eight months in 2020.

Figure 5: Histogram of Availability

2019



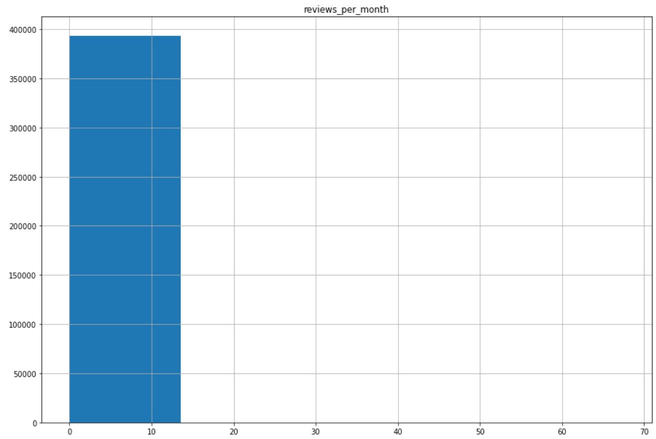
2020



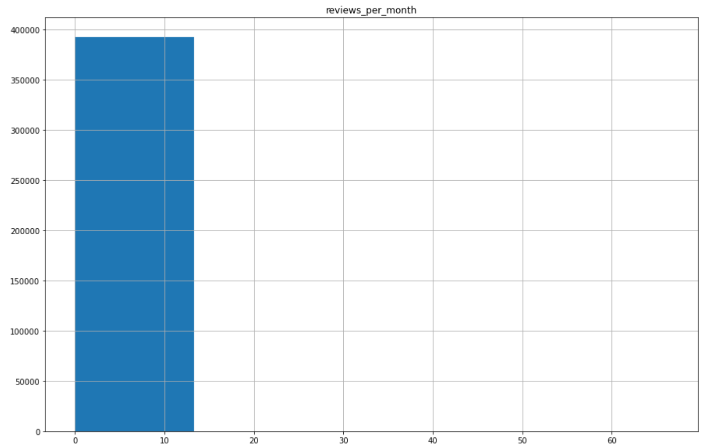
The histograms exhibit how there are two main peaks, those who don't offer their properties or only for very few days and those who list it all year round. It also depicts how the distribution is skewed as the majority of listings have minimal availability.

Figure 6: Histogram of Reviews per month

2019



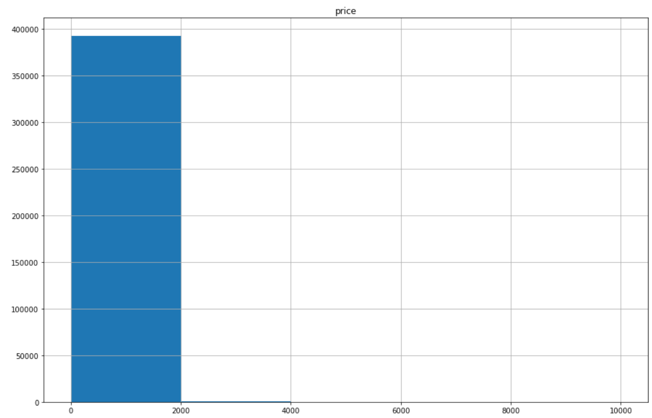
2020



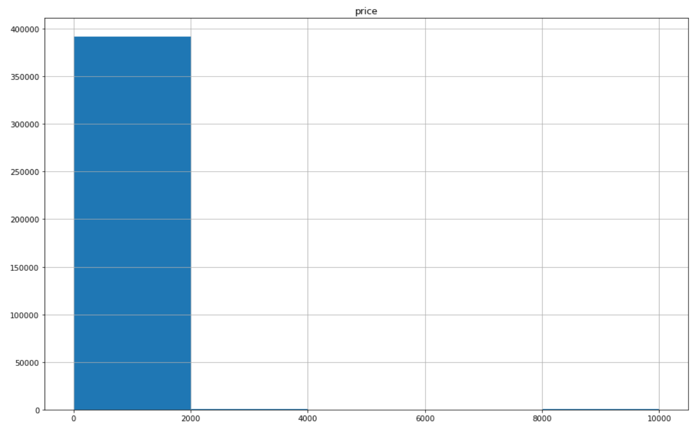
The histograms of reviews per month show that data is skewed whereby all properties have a number of monthly reviews between 0 and 12 for 2019 and 2020.

Figure 7: Histogram of Price

2019

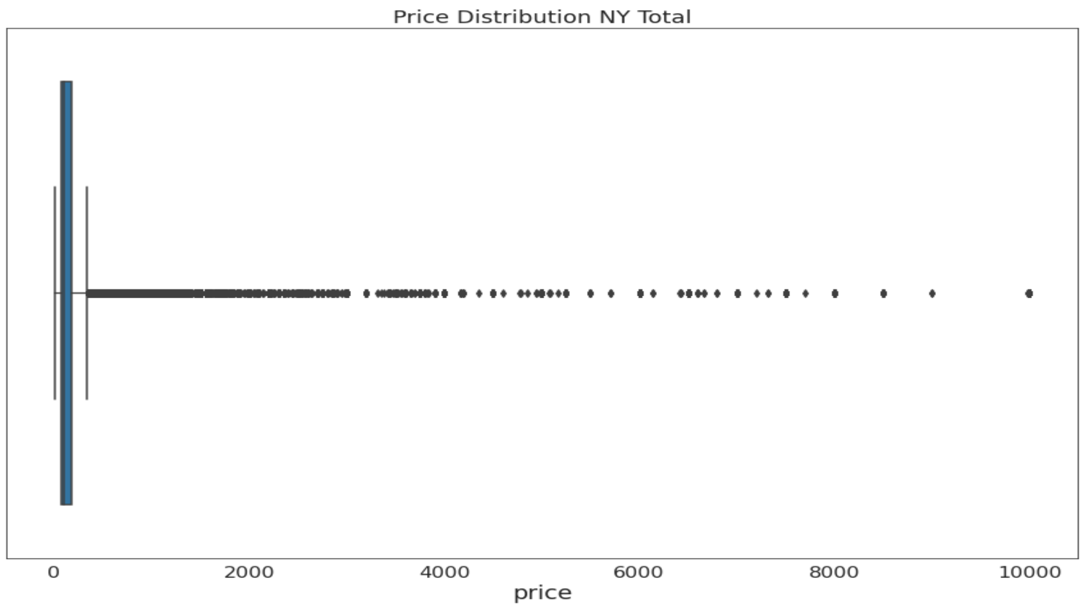


2020



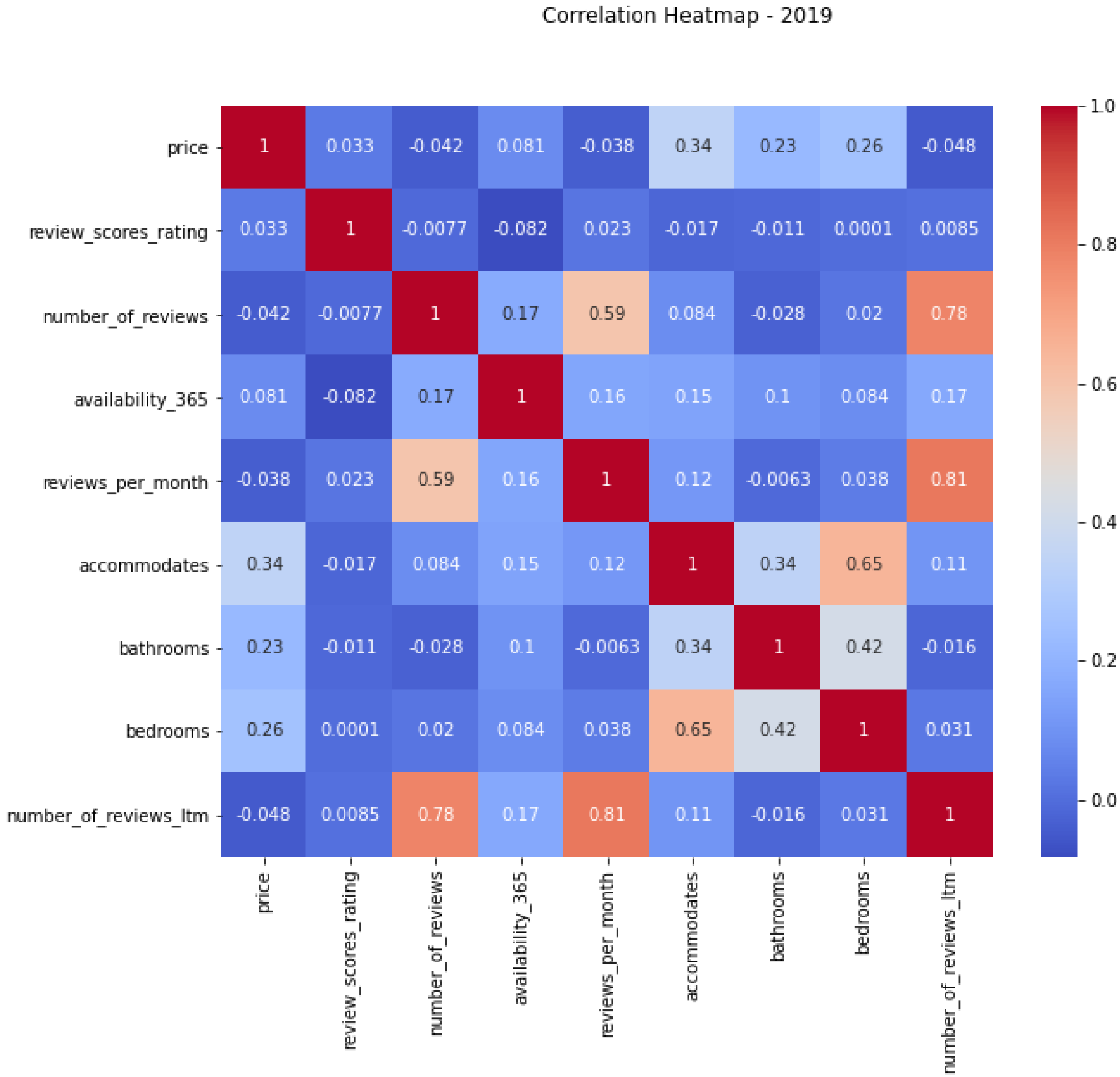
The price histograms give us an idea of how skewed the prices of Airbnb's in New York are. The vast majority of prices are in the lowest bin while a few very expensive listings skew the data. The boxplot below gives a more complete impression of the distribution of the price listings. Over 75% of the listings have a price per night below $170.

Figure 8: Price Distribution NY Total



To further investigate the dataset and changes, correlation heatmaps were generated in order to visualise any hidden correlations that would be interesting to explore further. The maps also offer a strong indication on which variables to include for our predictive models.

Figure 9: Correlation Heatmap (2019)

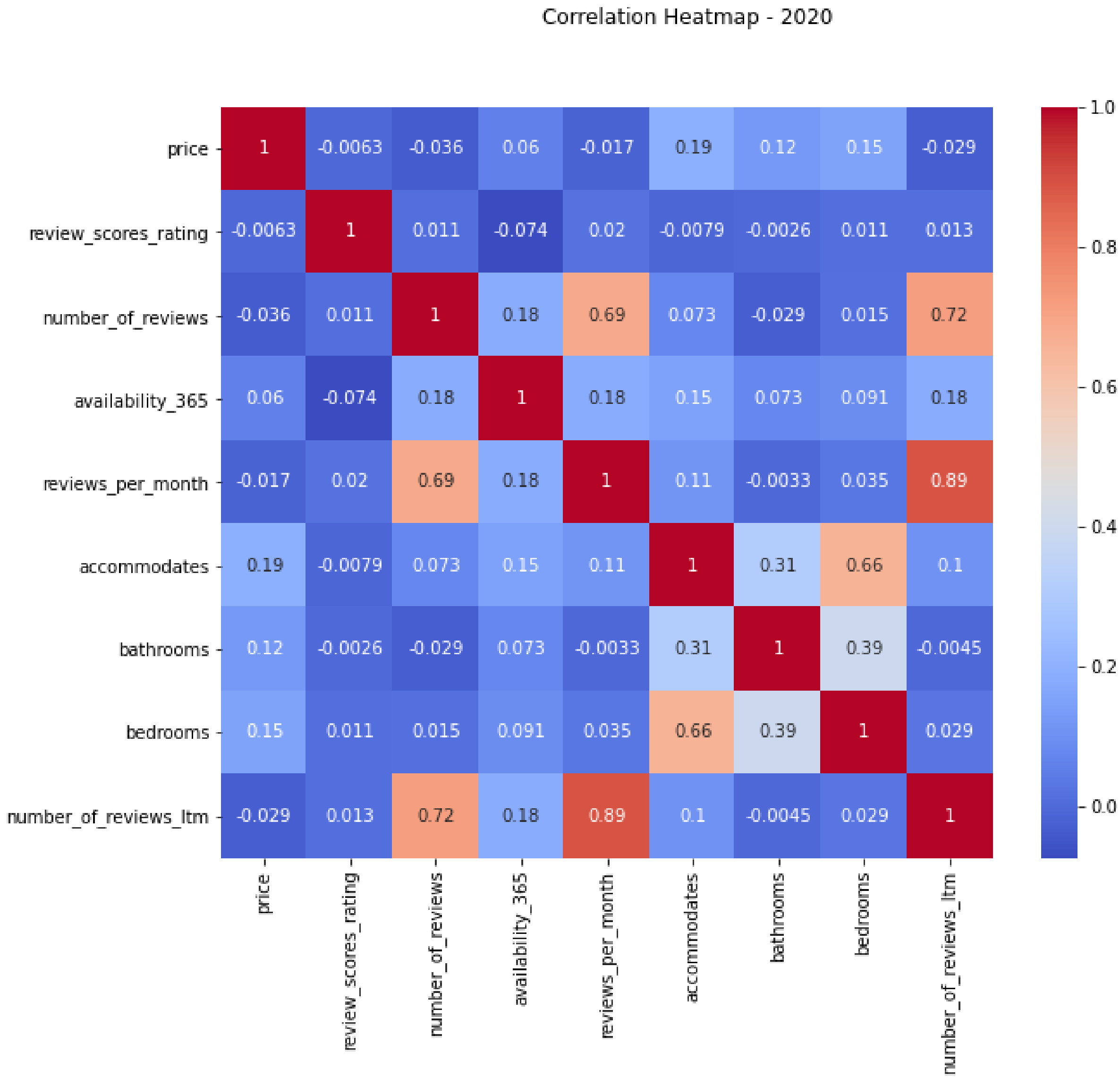


From the 2019 heatmap, we see that price is positively correlated with accommodates, bathrooms, and bedrooms. This makes sense as the bigger the property, the more expensive it would be.

Interestingly, price is negatively correlated with review scores rating, number of reviews, and reviews per month. This reflects how cheaper listings are associated with more visits than more expensive listings.

The 2019 correlation heat map shows a similar relationship to the 2020 map below. However, if we focus on the price correlations, we can see that for almost all variables, particularly accommodates, bathrooms and bedrooms, the magnitude has been nearly halved. This could mean that there is less rationale behind the hosts' price-setting in 2020 and that they are truly struggling to set the pricing themselves. This could be explained by all the uncertainties that COVID-19 brings, and that a pricing model would help them solve this issue.

Figure 10: Correlation Heat Map (2020)



The final part of the exploratory analysis looked at the exact effect of COVID-19 on the price of listings. We created a dummy variable with a value of 1 for 2020 and 0 for 2019, representing whether the year was impacted by COVID-19 or not.

In order to isolate the effects of COVID-19, we performed a fixed effects regression on the concatenated 2019 and 2020 dataset. Holding individual time-invariant characteristics constant, we regressed using the following equation:

$$log(Price)_{it} = \beta_0 + \beta_1 COVID_{it} + \beta_2 X_{it} + \gamma_t + \alpha i + \varepsilon it$$

The output highlights how holding variables constant, COVID-19 decreased price by 1.54%. Even though it seems like the price has only decreased by 1.54% because of COVID-19, looking at Figure 2 and increase in variance in the summary statistics shows that it brought greater uncertainties with it.

Having explored our dataset, established the relationships between the variables and gained some valuable insights we can start creating our price prediction models.

3.4 Prediction Models

3.4.1 Linear Model

In order to offer a price prediction for new Airbnb hosts, we initially conducted an ordinary least squares (OLS) regression. The OLS regression estimates the relationship between price and multiple variables from the dataset to create a model which predicts prices of future new listings.

For this, the 2019 and 2020 datasets were concatenated. The variables included in the regression reflect the information a new host in New York would have prior to posting their listing.

The variables include property type, room type, location, the estimated yearly availability, how many people the listing can accommodate and the price of the deposit and cleaning fee. Furthermore, the month in which the listing will be posted is included to account for seasonal price variations as explored in Figure 2.

Other variables such as ratings and reviews per month were not included in the model as these would not be available to new hosts.

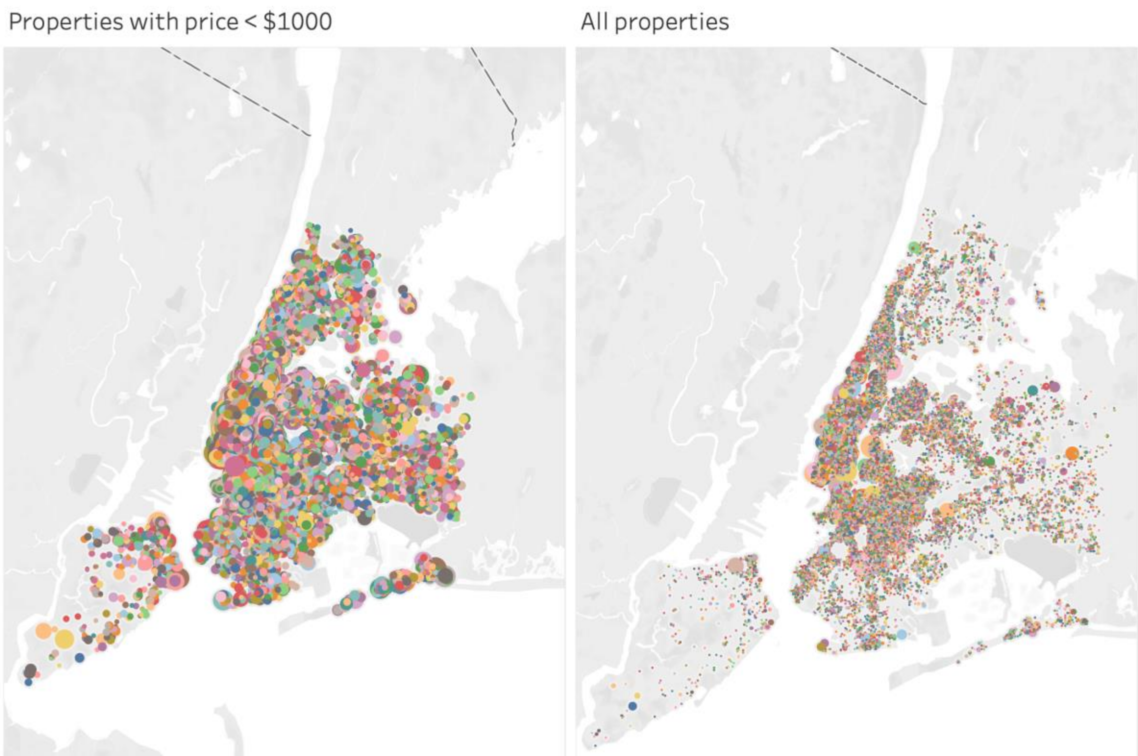
The following table depicts the variables included in the model and their correlation to the dependent variable price.

Table 4: OLS Variables and absolute Correlation to Price

price	1.000000
cleaning_fee	0.625596
accommodates	0.623641
bedrooms	0.606650
room_type_Private room	0.485747
neighbourhood_Manhattan	0.334118
bathrooms	0.299505
security_deposit	0.250808
neighbourhood_Queens	0.175561
neighbourhood_Brooklyn	0.161595
room_type_Shared room	0.128138
availability_365	0.126277
property_type_House	0.112192
property_type_Condominium	0.084052
neighbourhood_Staten Island	0.046651
property_type_Hotel	0.028337
Month	0.008450
property_type_Other	0.003778

Furthermore, some additional cleaning steps were necessary in order to achieve a model with a reasonable performance. The price distribution of all the listings in the dataset are very skewed as seen in Figure 8. To tackle this, the model was trained using data only including properties with a price per night below or equal to $1000. This still kept 99.4% of the listings in the dataset but reduced the effect of outliers on the model

Figure 11: Properties with Price < $1000 vs All Properties



The image above shows the location distribution of properties with a price below $1000 against all the properties in the dataset. The properties under the price cap seem to follow a similar distribution as all the properties. However, the variance between the different average prices has decreased tremendously. On the left image, the size of the dots, which represent the average price, are more equally sized than on the right image. This demonstrates that creating a pricing model with the properties where the price is less than $1000, reduces variance in price but still keeps the distribution of listings intact.

Furthermore, the dataset is skewed in the sense that there are many properties on the Airbnb website that are very rarely or never used, as seen in Figure 6.

The price recommendation should reflect the prices of listings which are in high demand. This is logical from a business standpoint as Airbnb would want to offer a price range for successful listings.

In order to have the model's prediction mirror the price range of successful listings, the top 50% of listings with the highest reviews per month were extrapolated to train the model. This means that the model takes the 25 000 top reviewed unique properties over all the 2019 and 2020 months into account.

Furthermore, before performing the linear model, we checked for multicollinearity in order to ensure the accuracy of our outputs. The rule of thumb for too high multicollinearity is a Variance Inflation Factor (VIF) of 10. We had to take out multiple variables such as bedrooms, bathrooms as these created multicollinearity issues. Furthermore, room type Hotel was deleted due to its relationship with property type Hotel. However, as seen in the following table, the figures kept in the model are below the threshold and therefore suited for the analysis.

Table 5: OLS Model's Features and their VIF

Feature	VIF
property_type_House	1.444022
property_type_Hotel	1.061532
property_type_Other	1.023534
property_type_Condominium	1.048986
accommodates	5.761650
neighbourhood_Manhattan	4.738402
neighbourhood_Staten Island	1.109336
neighbourhood_Queens	2.579202
neighbourhood_Brooklyn	4.308526
availability_365	2.648770
cleaning_fee	3.976470
security_deposit	1.346056
room_type_Private room	2.716452
room_type_Shared room	1.144391
Month	3.644783

The regression produced following coefficients which highlight the relationship they have with the dependent variable price.

Table 6: OLS Feature Output

	coef	std err	t	P>|t|	[0.025	0.975]
property_type_House	-0.7291	1.633	-0.446	0.655	-3.930	2.472
property_type_Hotel	53.0027	4.706	11.262	0.000	43.778	62.228
property_type_Other	19.1296	8.392	2.280	0.023	2.681	35.578
property_type_Condominium	22.6639	2.902	7.810	0.000	16.976	28.352
accommodates	19.1549	0.322	59.570	0.000	18.525	19.785
neighbourhood_Manhattan	82.8952	1.847	44.889	0.000	79.276	86.515
neighbourhood_Staten Island	-4.3169	5.463	-0.790	0.429	-15.025	6.391
neighbourhood_Queens	18.6055	2.164	8.596	0.000	14.363	22.848
neighbourhood_Brooklyn	28.3686	1.830	15.502	0.000	24.782	31.956
availability_365	0.0498	0.004	11.089	0.000	0.041	0.059
cleaning_fee	0.5970	0.015	39.992	0.000	0.568	0.626
security_deposit	0.0152	0.002	7.718	0.000	0.011	0.019
room_type_Private room	-34.5349	1.313	-26.294	0.000	-37.109	-31.961
room_type_Shared room	-55.9679	3.299	-16.967	0.000	-62.433	-49.502
Month	0.7027	0.197	3.575	0.000	0.317	1.088

From the property types, we can see that hotels have the largest impact on price, followed by condos, other types, houses and finally apartments (the reference dummy in the regression).

The regression also tells us which neighbourhoods have the biggest impact on price. Manhattan has by far the biggest impact, followed by Brooklyn, Queens, then the Bronx (the reference dummy in the regression) and finally Staten Island. Interestingly, Staten Island has a lower impact than the Bronx and Queens which contradicts the map in Figure 1. The reason for this is that many of the very highly-priced listings which were taken out for the model were situated in Staten Island. These outlier priced listings affected the map in Figure 1 but not the regression.

Furthermore, the regression shows how the entire apartmentt/home (reference dummy in the regression) has the most impact on price, followed by private rooms and shared rooms.

In addition, the accommodates variable tells us that for every additional person the listing can accommodate, the price increases by around $19.

Moreover, the results reveal how every additional day the listing is available, the estimated price increases by $0.05 per night.

Finally, every increase in the month is associated with a price increase of $0.7.

In order to analyse the performance of the model, we looked at the r-squared value and mean absolute error of the trained model against the test data set.

Table 7: OLS Accuracy Results

	MAE	R^2
Training	$47.1	57.6%
Test	$50.0	55.5%

The r-squared value of the model for the test set is 56%. This means the model explains 56% of the variation in the price of NY Airbnb listings. The similarity in r-squared between the training and test set demonstrates that our model is not overfitting.

A 56% accuracy is tolerable for a preliminary model. In order to increase the performance, additional variables such as proximity to popular venues, age of the building and perhaps the value of the interiors would increase the accuracy.

The mean absolute error of the model for the test set is $50. This means that on average, each prediction is $50 dollars different than the actual listed price. This gives the prediction a plus/minus window of around $25 per listing.

One of the limitations of our linear model is that there is assumed linearity between the independent and the dependent variables included in the regression. However, this is often not the case with complex relationships.

Therefore in order to overcome this limitation and to increase the accuracy of the model, we decided to perform a gradient boosting model.

3.4.2 Gradient Boosting Model

XGBoost (eXtreme Gradient Boosting) is an application of boosted decision trees which has been designed to reduce overfitting, thus, increasing performance. It has recently been a popular algorithm in applied machine learning (Nishida, 2020).

Extreme Gradiant Boosting offers superior performance and provides the benefit of generating automatically the features of importance in the model (Brownlee, 2020). Furthermore, decision tree models do not suffer from multicollinearity, which is a big advantage in our model since most variables are highly correlated, such as bathrooms and bedrooms.

The model was trained and tested in the same way as the linear regression model. However, we have kept multicollinear variables as they do not affect the output of decision trees.

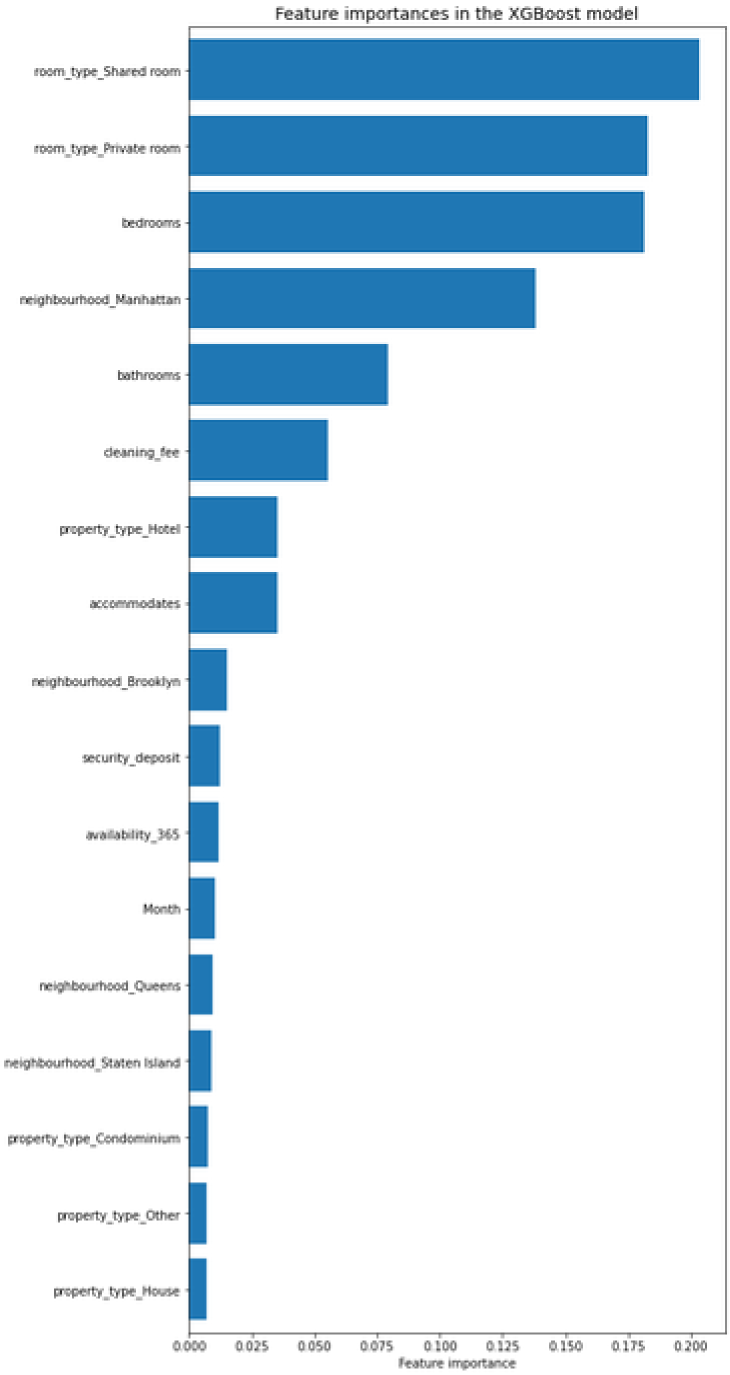
Table 8: XGBoosting Accuracy Results

	MAE	R^2
Training	$31.5	83.5%
Test	$39.5	67.5%

The results from the gradient boosting model seems to be slightly better than the linear regression model. The Mean Absolute error for the test set shows that on average, each prediction is $39.5 dollars different than the actual listed price compared to 50$ for the linear regression model. We have therefore chosen to use the gradient boosting model for the price prediction tool. Nonetheless, we can observe a high difference in R^2 between the test and training set which is an indication of overfitting.

From the graph below, we can observe that the mean features of importance in predicting the price are the room type, the different attributes of the rental and the neighbourhood.

Figure 12: Feature of Importance XGBoost model



3.5 Model Limitations

The best performing model was able to predict 64.9% of the variation in price. This means we still have a remaining 35.1% unexplained. The main reason for this are features not included in the data set which will be discussed in the conclusion.

4. Conclusion

4.1 Synthesis

The purpose of this project was to assess the impact of COVID-19 on New York's Airbnb properties and create a price prediction model for new listings.

1. The **descriptive analysis** demonstrated:
    - Differences in prices by geographical region in New York whereby Manhattan is the most expensive, while Bronx and Queens are the least expensive areas in both years.
    - An increase in movement and magnitude of price between 2019 and 2020, highlighting greater uncertainties in pricing.
    - A seasonal increase in price between June and August.
2. The **exploratory analysis** showed that:
    - Prices and availability increased from 2019 to 2020 but the number of monthly reviews decreased.
    - The **correlation heatmaps** show that price is positively correlated with accommodates, bathrooms, and bedrooms, but negatively correlated with review rating, number of reviews, and reviews per month.
    - The **fixed effects model** demonstrated that COVID-19 caused a 1.54% decrease in price.
3. The **prediction models** displayed that:
    - Hotels have had the largest price impact, and Manhattan have the biggest impact on price as seen by the **linear model**.
    - Room type and neighbourhood are the most important factors in determining price as seen by the **gradient boosting model**.

4.2 Recommendation

Based on the outcomes of the analysis, we recommend the following measure for Airbnb as a tool to monitor and control price fluctuations and maintain price differentiation as a key feature of the Airbnb platform.

Adopting a resilient price prediction model:

Airbnb should implement the prediction model on their website in order to help new homeowners set the price of their properties. This will help to support new users to list their properties in shorter periods of time and mimic the impact of potential external conditions. Furthermore, Airbnb will increase its customer experience by providing a better understanding of market price for hosts, which will help them maximise their occupancy and revenues.

Furthermore, from our analysis we discovered that COVID-19 has led to greater uncertainties in pricing, with price standard deviation almost doubling from 2019 to 2020. Providing a price prediction model will allow hosts to price their properties closer to the fair value and mitigate the uncertainties around price in times of crisis.

Therefore, by providing a better customer experience for hosts, Airbnb will be able to create value by acquiring new customers, since Airbnb captures value through fees set to hosts and guests.

4.3 Limitations

The main limitations which might potentially affect the outcomes of our analysis are:

1. Price and availability are also determined by other factors in the external environment such as economic, social, and political conditions. Also, intricacies in the internal environment such as competition would potentially impact the demand of users and price set by homeowners.

2. Our proposed predictive model lacks some information which are related to unobservable factors of properties such as age and quality.

3. Moreover, we believe that one important missing variable is the access to public transports, museums, and other points of interest.

4. The occupancy rate and demand are undetermined in our analysis due to the unavailability of data in this matter, which would have given us more information regarding the fair price.

5. Finally, the training set performs much better than on the test set. Therefore, the model is likely overfitted. The overfitting problem can be mitigated by taking a larger sample for the training set, or by removing the less important features in our analysis.

4.4 Next Steps

The further analysis which should be conducted is to:

1. **Analyse the movement, significance, and characteristics of Airbnb's demand, segment customers by demographics and cluster properties by value**. This will help to mitigate the endogeneity problem in our dataset. Furthermore, researchers found that units with verified photos (taken by Airbnb's photographers) generate additional revenue per year on average (Zhang, Lee, Singh and Srinivasan, 2017). Therefore, we **could also measure the causal impact of image quality,** using a difference-in-difference model.

2. **Evaluate the impact of Competition on Airbnb's prices:** Airbnb's data should be benchmarked against competitors' data to determine the areas where Airbnb should focus on to attract, solicit, and retain customers. In addition, this would help Airbnb understand whether the effects seen are indeed due to COVID-19 or internal factors.

3. **Perform cost-benefit analysis** to determine the value of the predictive model.

4. **Investigate the reasons for price fluctuations by region** to assess the possibility that some areas are more affected than others. This includes taking external health measures into consideration.

References

Airbnb. 2020. *How Should I Choose My Listing's Price?* [online] Available at: <https://www.airbnb.co.uk/help/article/52/how-should-i-choose-my-listings-price> [Accessed 10 December 2020].

BBC News. 2020. *Airbnb Hosts Defy Lockdown Laws With 'Covid-19 Retreats'*. [online] Available at: <https://www.bbc.co.uk/news/technology-52184497> [Accessed 10 December 2020].

Brownlee, J., 2020. Feature Importance And Feature Selection With Xgboost In Python. [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/> [Accessed 5 December 2020].

Nishida, K., 2020. Introduction To Extreme Gradient Boosting In Exploratory. [online] Medium. Available at: <https://blog.exploratory.io/introduction-to-extreme-gradient-boosting-in-exploratory-7bbec554ac7> [Accessed 4 December 2020].

Zhang, S., Lee, D., Singh, P. and Srinivasan, K., 2017. How Much Is an Image Worth? Airbnb Property Demand Estimation Leveraging Large Scale Image Analytics. *SSRN Electronic Journal*.