Regression Project: Pricing an Airbnb (11 Hyde Park)

Introduction

Our team has compiled data on the Airbnb rental prices in 11 Hyde Park through a deep market analysis. The 4 key variables that affect price we're discussing are: Room Type, Accommodations, Property Type, and Bedroom. The average nightly price of Airbnbs in 11 Hyde Park is $93.68 with a median of $75. This suggests our data is skewed to the right. A possible explanation for this is room type. In the dashboard bubble chart, there seems to be a positive relationship between number of accommodations and price. The trendline shows that for each additional accommodation, the price increases on average by $22.35. Since entire homes/ apartments typically accommodate more than the mean accommodation of 2.74 people, this increases the mean price. From this, we conclude that entire homes/ apartments have higher average prices and number of accommodations compared to private rooms. Property type also affects price, with apartments and houses having the most impact on the overall mean price ($93.68), since they have the highest counts as well as the highest average prices, $99.22 and $93.64 respectively. To continue, the dashboard map shows a positive relationship between average number of bedrooms and average price in 11 Hyde Park. A point of Airbnbs with 4 bedrooms has an average price of $269 while a point of 3 bed Airbnbs has an average price of $170. This is a common trend throughout the graph with the largest points, representing the most average bedrooms, having the highest average prices. Moreover, in the stacked bar chart, most Airbnbs have 1 bedroom. Since Airbnbs with 1 bed have the highest count, their prices on the map are the most typical in 11 Hyde Park.

Table 1:

	Mean	Median	Standard Deviation	Minimum	Maximum	Count
Accommodates	2.74	2	1.87282534	1	9	100
Bathrooms	1.305	1	0.55910733	0.5	3.5	100
Bedrooms	1.28	1	0.76646901	0	5	100
Beds	1.64	1	1.000202	1	5	100
Guests_included	1.34	1	0.90140183	0	6	100
Availability_30	11.33	7.5	11.2708214	0	30	100
Availability_365	224.36	309	143.552934	0	365	100
Review_Scores_rating	93.2133333	96	7.86764385	60	100	75
Distance from Boston Common	5.88192408	5.83050287	0.85408439	4.37829895	7.79938535	100
Price	93.68	75	61.9796122	31	349	100

Table 2:

Property Type	Count
House	66
Apartment	27
Townhouse	2
Loft	1
Condominium	1
Bed & Breakfast	3
Total	100

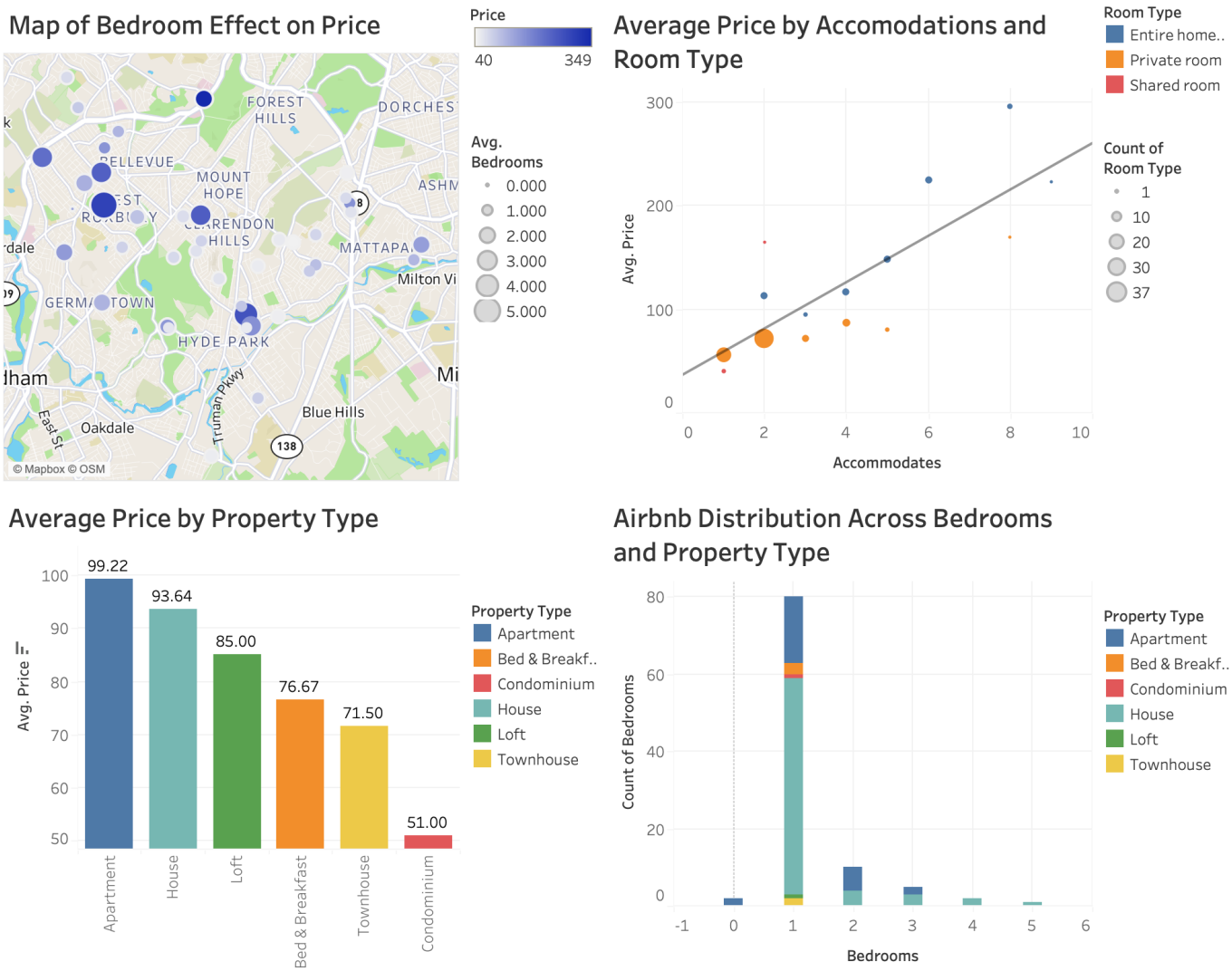
Table 3:

Room Type	Count
Private room	73
Entire home/apt	24
Shared room	3
Total	100

Table 4:

Property Room Type	Count
Private room in house	57
Entire house	10
Entire apartment	14
Private room in apartment	16
Shared room	3
Total	100

Dashboard:



Data Analysis

A. For each additional bedroom, the nightly Airbnb price increases on average by $63.69. The t stat for bedrooms is 12.652 which is greater than 2, meaning it's statistically significant, thus bedrooms have a significant effect on average price. The R^2 is 0.620 which means that 62% of the price can be explained by the variable number of bedrooms.

B. Holding beds constant, for each additional bedroom, the price increases on average by $48.30. The t-stat for bedrooms is 7.19 which is greater than 2 so the coefficient is statistically significant. While it still has an effect on average price, having the covariate of beds decreases the bedroom coefficient meaning it has less of an effect. The adjusted R^2 is 0.651 which means that 65.1% of the variation of price can be explained by the number of bedrooms. This is higher than the R^2 in regression A therefore it has a higher correlation.

C. ^price = 25.67 + 40.43* beds + 0.46* bed^2
The absolute value of the t-stat for bed squared is 0.124 which is less than 1, thus we cannot say there is a nonlinear relationship between number of beds and price. The t-stat for bed is 2.11 so it's statistically significant. The price of a rental with 2 beds is $108.37 and the price of a rental with 1 bed is $66.56 making the price change $41.81. The price change of rentals between 5 beds and 4 is $239.32 - $194.75 = $44.57. As the number of beds increases, the average price increases exponentially at a higher rate. The adjusted R^2 for this exponential regression is 0.465 which means that 46.5% of the prices are explained by beds in this regression. Overall this means that in comparison to the straight line regression of beds in part A, this exponential regression is worse at predicting price since the adjusted R^2 is lower.

D. Holding property room type, bedrooms, and beds constant, for every additional review, the price decreases on average by its coefficient of $0.29. This would make sense if most reviews are bad which is a common reason why people write reviews. Out of property room type, the variable which had the largest impact on price was private room in apartment with its |t-stat| = 3.39, while shared room had the lowest impact on price with |t-stat| = 1.29. Looking at the coefficients of each property type while holding constant other factors, property type has significant differences on price and their values make sense when comparing to entire apartment. The property room type with the only positive coefficient of 27.97 is entire house. This makes sense since entire house is bigger than entire apartment. The property room type which decreases price the most is private room in apartment with its coefficient of -43.95. This makes sense when comparing one private room in an apartment to an entire apartment.

E. Approximate Distance from Hyde Park T: we believe that Airbnbs closer to Hyde Park T stop would be priced higher due to transportation accessibility. We found the latitude and longitude of the station and used the same distance formula of Distance from Boston Commons.
The adjusted R^2 is 0.73 and the SEE is 32.03 while for part D the adjusted R^2 is 0.71 and the SEE is 33.51. So regression E is a slightly better predictor for price in comparison to regression D as the adjusted R^2 is 2% higher and the SEE is lower by 1.48. 73% of the average price can be

explained by the variables in our regression, thus it does a good job in explaining the variation in price.

F. The adjusted R^2 is 0.75 and the SEE is 31.06 for the final regression. Compared to the adjusted R^2 of 0.73 in regression E, the final regression is better at predicting price by 2%. The standard error decreased as well making our data more precise, from 32.03 in regression E to 31.06.

Regression Analysis

A. ^price = 67.54 -5.91(Approx Dist from Bos Common) + 8.81(TV) -21.33(Private room in house) + 23.18(Entire house) + 12.48(accommodate) + 26.44(bedrooms) -23.73(private room in apartment)

^price = 67.54 -5.91(4.38) + 8.81(1) -21.33(0) + 23.18(1) + 12.48(4) + 26.44(2) -23.73(0)= 176.44

95% CI: ^y ± 2*SEE = 176.44 ± (2)31.06 = (114.32, 238.56)

* for variable "accommodates": we believe 4 is the best value which yields a higher price. Since there are 2 beds, and a bed typically fits 2, a 2 bed unit can accommodate 4 people.
* for variable "Approximate distance from Boston Commons": the coefficient is negative. We found the smallest value (distance closest to Boston Commons) in our dataset and used it in this estimation. We can't use 0 as our Airbnb is in Hyde Park and not in Boston Commons.

B. https://www.redfin.com/MA/Boston/623-Cummins-Hwy-02136/unit-E/home/12431329
Estimated monthly mortgage: $1,686 /mo (estimated monthly mortgage on Redfin is constantly updated so our mortgage per month may not be the most up to date)

C. 1686 * 0.63 = $1062.18/ mo
   a. Expected revenue = 176.44* 20 = $3528.80
   $3528.8 - $1062.18 - $800 = $1666.62 expected profit for at least 20 days

   b. 176.44x -1062.18 - 800 = 0 → 176.44x = 1862.18 → x = 10.55
   The lowest number of nights to break even is around 11 nights

D. We believe it's profitable to purchase this condo and rent it as an Airbnb. Using our regression analysis, where 75% of the price can be explained by our variables, a minimum of 11 nights is needed to break even. This comes out to around 36.67% occupancy rate per month. We are confident that we can reach this occupancy rate to become profitable. The risk of this investment would be that Hyde Park is a more residential area of Boston so it would have less demand. People typically rent near the metropolitan area. Secondly, the US is currently under recession making housing investments less valuable. In recessions people are more reluctant to spend which lowers demand even more. The limitations of our analysis are variables that we didn't take into account. For instance, seasons may affect demand. Peak season like the holidays would cause

more demand in Airbnb rentals while off season like right after the holidays would cause stagnation in rentals.

E. Airbnb rentals have the ethical implication of gentrification, stripping a community's culture. According to data between 100 different cities and countries, short term rentals are growing in popularity leading to less permanent residents and less affordable housing. This had a negative impact as poorer residents are pushed out to different communities while transient richer communities take over. This can take place in Hyde Park - coined “A Small Town in the City”- stripping away the suburban culture. However, short term rentals have the potential of increasing tourism and businesses. Airbnb helps reduce overcrowding in highly visited areas by promoting less visited areas. Our Airbnb rental can level out highly populated areas like Boston Commons with lower traffic areas like Hyde Park.

Regression	A	B	C	D	E	F
beds		16.86	40.43	17.66	-0.75	
		(3.28)	(2.11)	(3.55)	(-0.12)	
bedrooms	63.69	48.30		25.99	26.90	26.44
	(12.65)	(7.19)		(3.21)	(3.21)	(3.66)
beds squared			0.46			
			(0.12)			
Bathrooms					-5.35	
					(-0.77)	
accommodates					14.29	12.48
					(3.75)	(4.82)
private room in apartment				-43.95	-20.90	-23.73
				(-3.39)	(-1.46)	(-2.04)
entire house				27.97	21.19	23.18
				(1.59)	(1.21)	(1.42)
private room in house				-34.62	-17.60	-21.33
				(-3.19)	(-1.41)	(-2.27)
shared room				-28.06	3.52	

				(-1.29)	(0.15)	
number of reviews				-0.29		
				(-1.96)		
approximate distance from Boston Commons					-3.67	-5.91
					(-0.70)	(-1.57)
approximate distance from Hyde Park T					1.26	
					(0.24)	
parking					-2.43	
					(-0.29)	
TV					7.09	8.81
					(0.81)	(1.16)
24-Hour Check-in					-3.08	
					(-0.30)	
superhost					-5.12	
					(-0.44)	
intercept	12.16	4.21	25.67	60.63	56.72	67.54
	(1.62)	(0.56)	(1.35)	(4.09)	(1.43)	(2.59)
R squared/ Adjusted R squared	0.62	0.65	0.47	0.71	0.73	0.75
SEE	38.39	36.61	45.33	33.51	32.03	31.06
number of observations	100	100	100	100	100	100