

# Predicting Airbnb Rates for a 2-Bedroom in Crown Heights

Airbnb, along with other online home-sharing businesses, has been under fire in New York (Abbruzzese & Plautz, 2014). A New York City law that would require home-sharing sites to provide detailed information on listings was slated to go in effect earlier this year was blocked. City officials believe that the majority of Airbnb listings are illegal and they need this information to prove it, but a federal judge felt that it violated the guarantee from illegal searches and seizures in the Fourth Amendment (Weiser & Goodman, 2019). For the moment, it seems as if the issue is a stalemate.

Although touting empowerment for the average person and addressing an undisputable need, Airbnb is disrupting more than just the hotel industry (“Inside Airbnb. Adding Data to the Debate,” n.d.). Landlords can generate more income by renting their properties through Airbnb rather than leasing to a tenant and are seeking to do so. This fact is putting more stress on the already difficult situation for low to middle-class families that rent (Wachsmuth, Chaney, Kerrigan, Shillolo, & Basalaev-Binder, 2018). A study by the Urban Politics and Governance Research Group at McGill University estimates that 7-13.5k long-term rentals have lost to the market resulting in a 1.9% rise in the cost of rent for long-term renters. The city believes that many of these are rent-regulated units that landlords would rather offer short-term but these laws will undoubtedly affect many people that use the income to offset the high cost of living in the city (Greenberg, 2018). All these issues aside, one has to admit that the flexibility and income gained from renting part of your home short-time is very appealing.

We have a two-family home in Crown Heights, that we are currently using as a single-family mostly due to the fact parts are uninhabitable and scary. We are slowly fixing it up and hope to have a two-bedroom apartment on the ground floor that we can rent. Of course, Airbnb is appealing because it would be great to rent it out part of the time and still be able to offer it to family members other times. For instance, my parents could stay during the summers. So I am interested in comparing what I would get in rent from a tenant versus Airbnb. (Whether or not we act on it will be a larger decision where the aforementioned issues would be considered.)

We have already been quoted by a real estate agent that our potential property could rent for \$2600-2800 each month. For comparison, Airbnb provides a calculator although many people feel that

it purposely lowballs rates (“Figure Out How Much Your Home Would Earn on Airbnb With This New Tool,” 2017). Airbnb gave me a rate of \$2,431 per month but only allows you to choose borough, room or apartment, and number of occupants. I would think that the neighborhood would be as important as the borough. It assumes that you will rent 15 days out of the month. That already sounds awesome.

There are several other calculators available. AirDNA is a popular tool for gaining insights on short-term rentals for Airbnb and a competitor, Vrbo (“AirDNA | Short-Term Rental Data & Analytics | Airbnb & Vrbo,” n.d.). They provide more robust information such as predictions on seasonal changes in rates and occupancy. My AirDNA results predicted that we could earn \$46,354 for the year and expect a 71% occupancy rate. That does better than Airbnb’s yearly estimate of \$29,172 at roughly 50% occupancy rate. At this point, I am wondering if I can predict the rate of a 2 bedroom apartment in Crown Heights myself.

Inside Airbnb (insideairbnb.com) is an independent site that scrapes publically available data from Airbnb and serves for research and analysis (“Inside Airbnb. Adding Data to the Debate.,” n.d.). They also provide some filters and key metric tools. The datasets for New York City used for this paper were scraped on September 19, 2019 (“Inside Airbnb,” n.d.). The site provides several options for download that contain varying levels and types of data. For this study, a simple clean summary file of listings and a very detailed file of listings were utilized.

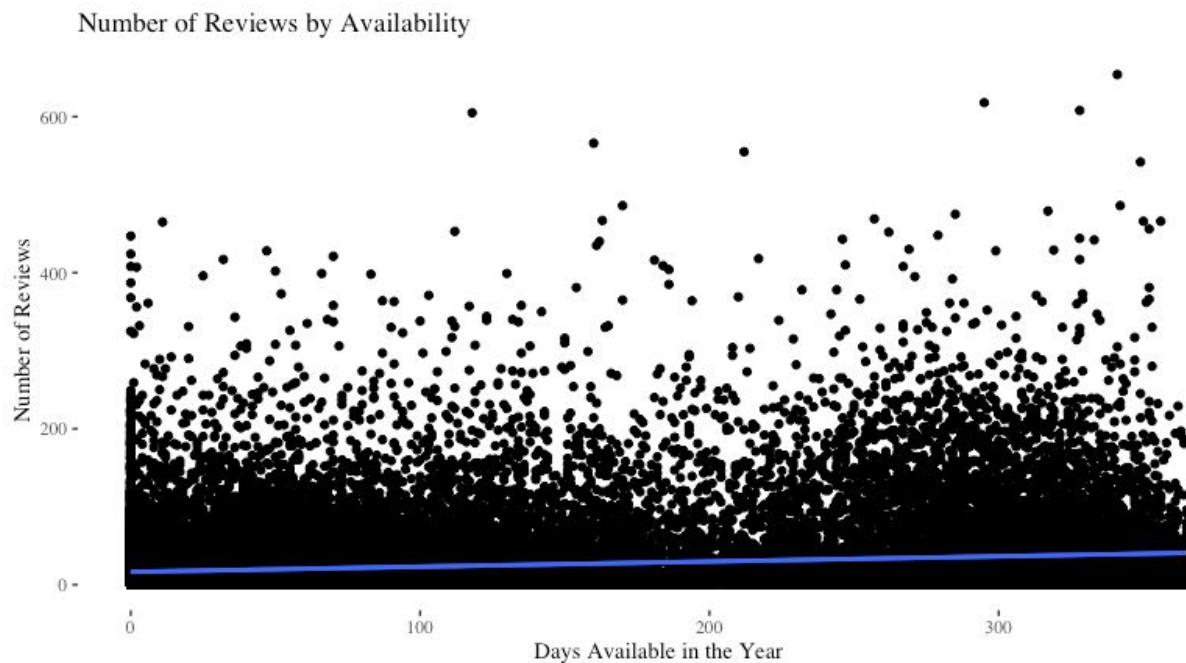
Beginning with the summary listings file, I was able to glean the following insights. The city has a total of 48,377 listings on Airbnb. The overwhelming majority of these are in Manhattan and Brooklyn. Interestingly, or not, 19 people decided to call their listings "Home away from home" which is the top title for a listing. And “Michael” is the most popular hostname. The average number of listings per host is 8.57 although most hosts have only one listing, which is why the median is 1. This inflated average is due to the number of hosts with very high numbers of listings. Sonder (NYC) is one of these hosts with 387 listings. Sonder is a property manager and many of their listings are categorized as a “hotel,” a newer category on Airbnb. They strive to provide a hotel experience that feels like a home (“Sonder, Taking Stay Further.,” n.d.).

The average nightly cost in all of NYC is 105. The highest price for one night was 10k a night. In fact, there are eight listings that are 10k a night. More than half are put up by Nycasa 46 in the Theater District, another property manager, but the first two are listed by individuals. The first one is a furnished room in an Astoria apartment with a minimum night stay of 100 that does has reviews albeit

from 2016 but just seems strange. The second is a luxury 1 bedroom apartment with stunning Manhattan views in Greenpoint with a min stay of five nights. The intended use could be as a film set location (Brydson, 2014).

When we break down the average cost of listing by borough Manhattan is the priciest at \$200 a night, then Brooklyn at \$123, Staten Island at \$119, Queens at \$99 and the Bronx \$86. Looking at neighborhoods in Brooklyn alone, the top five earners are Sea Gate (a gated community at the tip of Coney Island), Cobble Hill, Brooklyn Heights, Vinegar Hill, DUMBO and Clinton Hill. The majority of rentals in Brooklyn go between \$100-200. The average cost per night of Crown Heights is \$133 which is slightly above average for Brooklyn.

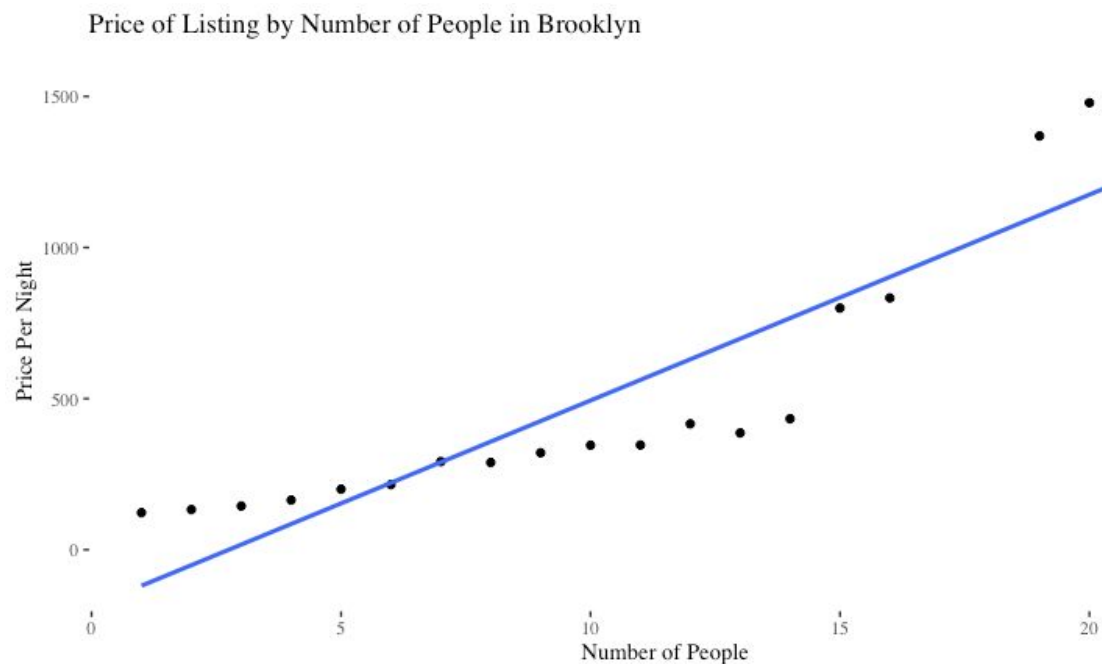
I decided to explore whether there was a relationship between availability and reviews. First I plotted the number of reviews over the number of days available. Looking at the chart, there did seem to be a slight rise. After calculating that for every day of availability, you will get 0.06645 more reviews, assuming a base 16.5 reviews, I determined that this was a weak correlation. The number of reviews may be more tied to the length of the listing rather than availability.



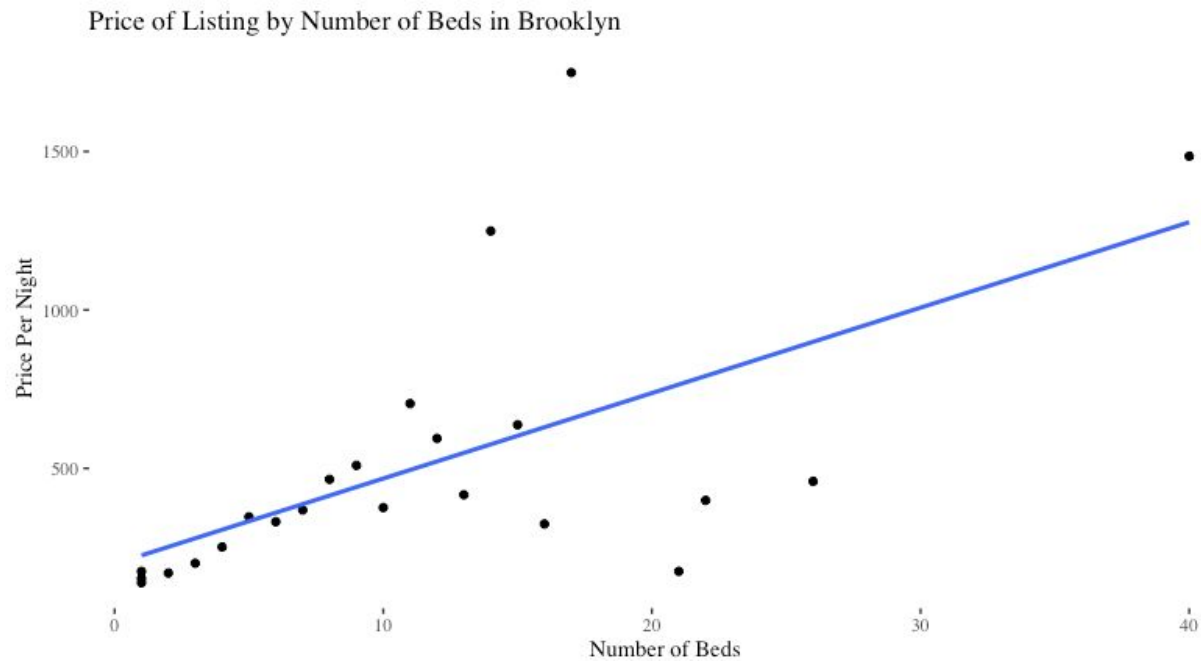
At this point, I realized I wasn't going to get the insights I was looking for from this dataset and needed to pull in the larger detailed set. This new set included 106 variables, some cleaner than others, including the number of bedrooms, bathrooms, beds, and people it can accommodate. I then created

my own dataset from this file with any values I felt would be applicable. The price values in the new set were a mess so I simply copied them over from the first dataset I was using.

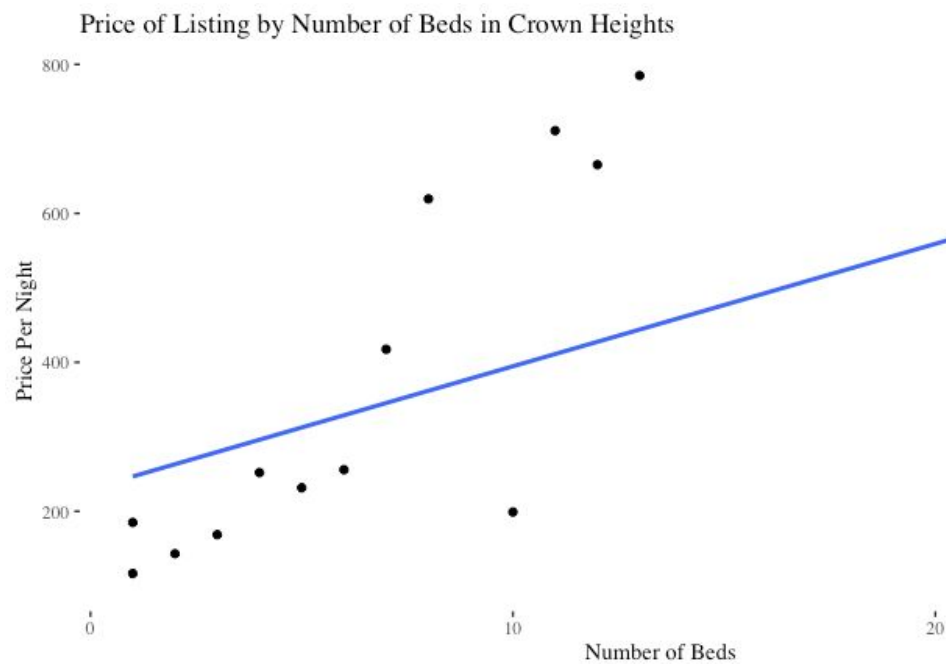
Focusing on “Entire home/apt” listings in Brooklyn, I wanted to see if there was a relationship between the number of people and price. I came up with an intercept of -187.26 and a coefficient of 68.12. This is a significant and strong relationship but you need at 3 or more to reach a positive baseline. For every person that a listing can accomodate there is an increase of \$68.12 in the price.



I then looked at the relationship between the number of beds and the price. Any value for beds that was 0, empty or NA, I changed the value to 1. My thinking that was that if a listing is claiming to accommodate people overnight there has to be at least one bed. This also created a strong correlation. There is an increase of \$26.96 in the price for every bed available in a listing with a base price of \$198.40 per night.



Finally, I focused in on Crown Heights specifically. Once again I had to covert any “bed” values that were 0, empty or NA to 1. I had a strong correlation again and I was able to make a prediction for the nightly rate or our property. There is an increase of \$16.44 for every bed added with a base price of \$230.30 for a full apartment in the neighborhood of Crown Heights.



I was able to predict that the nightly price for a two-bedroom in Crown Heights is \$263.17. If I assume a 71% occupancy rate that AirDNA provided, that would mean \$68,200.50 per year. If I assume 50% occupancy from Airbnb, that is \$48,028.53. Both of those are considerably higher than what we would get from a long-term tenant. Of course, these are dependent on my predictions.

Further considerations could include looking for connections between Landmarked homes and Airbnb prices. Staying in a Brooklyn brownstone with original details could attract tourists more than a small boxy hotel room in midtown. If a property is landmarked, does it go for more? If a property allows pets, can the get a higher rate?

In the end, I am still on the fence. I am not the landlord with multiple listings that is not on-premise, nor the person who sublets their place while they are away, so legality is not an issue (for now). I have a family and I can't say I am that comfortable with a revolving door into my home. My house is not like an apartment building, we will be in close quarters with whoever is below. Check out AirbnbHell.com and you'll find a plethora of horror stories about Airbnb renters trashing the place, using it for suspicious activities or engaging in threatening behavior. Tenants aren't without issue, but a good renter can be an asset by being dependable, caring for your home and being another set of eyes to make your home more secure. And it seems we are leaning in that direction.

# Bibliography

Abbruzzese, J., & Plautz, J. (2014, April 26). New York Goes to War Against Airbnb for Disrupting Hotel Business. Retrieved October 28, 2019, from Mashable website:

<https://mashable.com/2014/04/26/new-york-vs-airbnb/>

AirDNA | Short-Term Rental Data & Analytics | Airbnb & Vrbo. (n.d.). Retrieved October 29, 2019, from AirDNA - Short-Term Vacation Rental Data and Analytics website: <https://www.airdna.co/>

Brydson, N. (2014, July 7). Getting Creative with Locations Using Airbnb. Retrieved October 29, 2019, from Sunnyside Films website:

<https://sunnysidefilms.com/getting-creative-with-locations-using-airbnb/>

Figure Out How Much Your Home Would Earn on Airbnb With This New Tool. (2017, August 9).

Retrieved October 28, 2019, from

<https://lifehacker.com/figure-out-how-much-your-home-would-earn-on-airbnb-with-1797682122>

Greenberg, Z. (2018, July 18). New York City Looks to Crack Down on Airbnb Amid Housing Crisis. *The New York Times*. Retrieved from

<https://www.nytimes.com/2018/07/18/nyregion/new-york-city-airbnb-crackdown.html>

Inside Airbnb. Adding Data to the Debate. (n.d.). Retrieved October 28, 2019, from

<http://insideairbnb.com/index.html>

Inside Airbnb: New York City. Adding data to the debate. (n.d.). Retrieved October 29, 2019, from Inside Airbnb website: <http://insideairbnb.com>

Sonder, Taking Stay Further. (n.d.). Retrieved October 29, 2019, from <https://www.sonder.com>

Wachsmuth, D., Chaney, D., Kerrigan, D., Shillolo, A., & Basalaev-Binder, R. (2018). *The High Cost of Short-Term Rentals in New York City*. 49.

Weiser, B., & Goodman, J. D. (2019, January 3). Judge Blocks New York City Law Aimed at Curbing

Airbnb Rentals—The New York Times. Retrieved October 28, 2019, from New York Times Online website:

<https://www.nytimes.com/2019/01/03/nyregion/nyc-airbnb-rentals.html?auth=login-email&login=email>



# R Code

```
# Predictive Data Analysis
```

```
# InsideAirbnb New York City Listings [http://insideairbnb.com/get-the-data.html]
```

```
# Data scraped 09-12-2109
```

```
library(tidyverse)
```

```
library(dplyr)
```

```
library(lubridate)
```

```
library(ggthemes)
```

```
library(broom) #to help clean things up and remerge our dataframes
```

```
library(GGally) #to help run multiple pair-wise correlations
```

```
# Specified that the first row is the header, and check the formatting of names
```

```
nycListings <- read.csv("../DataSets/insideairbnb/listings.csv", header=TRUE, check.names = TRUE)
```

```
# This is a clean top level dataset
```

```
# Shows that it is being read as a dataframe
```

```
class(nycListings)
```

```
# The dataframe has 48377 objects with 16 variables
```

```
dim(nycListings)
```

# List all the variable names available

```
names(nycListings)
```

# Overview of dataset shows that there are 224 levels for Neighbourhood in the five boroughs

```
str(nycListings)
```

# A better overview...

```
glimpse(nycListings)
```

```
sum(is.na(nycListings))
```

# Almost 10k missing values

# Statistical summary

```
summary(nycListings)
```

# The median nightly cost of NYC is 105

# The lowest price is zero, so that should be dropped. The highest is 10k a night.

# Apparently 19 people decided to call their listing "Home away from home"

# Michael is the most popular host name

# Sonder (NYC) and Blueground appear in the top ten host names

# Brooklyn (19856) and Manhattan (21183) overwhelmingly have the most listings. Queens (5853), Bronx (1126) and Staten Island trails with (359)

# Brooklyn Neighbourhoods make up 3 of the top 5. Williamsburg, BedSty, Harlmen, Bushwick, Hell's Kitchen

# There are four types of rentals: Entire home/apt (24898), Hotel Room (435), Private Room (21852), Shared (1192)

# The median for number of listings per host id is 1 but the mean 8.57 due to a few high number hosts.

# Sonder (NYC) host has 387 listings. This company provides short housing in condo buildings where units haven't been sold yet and room type is a Hotel room. This could be due to the fact that some people have third-parties manage their airbnb's.

# There are eight listings that are 10k a night. More than half are put up by Nycasa 46 in the Theater District.

# The first one is a furnished room in Astoria apartment with a minimum nights at 100 and has reviews from 2016...seems strange.

# The second is a luxury 1 bedroom apartment with stunning Manhattan views in Greenpoint with a min stay of five nights. This could be used as a set location.

```
nycListings %>%
```

```
  filter(price == 10000)
```

# Median Cost by Borough

```
nycListings %>%
```

```
  group_by(neighbourhood_group) %>%
```

```
  summarize(meanPrice = mean(price))
```

# Bronx 85.8, Brooklyn 123, Manhattan 200, Queens 99.00, Staten Island 119

# Median Cost by Brooklyn Neighborhoods

```
# There are 48 neighborhoods in Brooklyn
```

```
bkMeanListings <- nycListings %>%
```

```
  filter(neighbourhood_group == "Brooklyn") %>%
```

```
  group_by(neighbourhood) %>%
```

```
  summarize(meanPrice = mean(price)) %>%
```

```
  arrange(desc(meanPrice))
```

```
summary(bkMeanListings)
```

```
head(bkMeanListings)
```

```
# Sea Gate      492.
```

```
# Cobble Hill   210.
```

```
# Brooklyn Heights  205.
```

```
# Vinegar Hill   199.
```

```
# DUMBO          193.
```

```
# Clinton Hill   182.
```

```
tail(bkMeanListings)
```

```
# Median Cost of Entire Apartment by Neighborhood
```

```
bkEntireMeanListings <- nycListings %>%
```

```
  filter(neighbourhood_group == "Brooklyn") %>%
```

```
  filter(room_type == "Entire home/apt") %>%
```

```
  group_by(neighbourhood) %>%
```

```
summarize(meanPrice = mean(price))%>%  
  
arrange(desc(meanPrice))  
  
summary(bkEntireMeanListings)  
  
hist(bkEntireMeanListings$meanPrice)  
  
# Sea Gate, a private gated community located at the tip of Coney Island, is the highest with 608 a  
night  
  
  
# Is there a relationship between availability and number of reviews  
  
ggplot(nycListings, aes(x=nycListings$availability_365, y=nycListings$number_of_reviews)) +  
geom_point()+  
  
labs(title="Number of Reviews by Availability")  
  
  
# Add line of best fit  
  
ggplot(nycListings, aes(x=nycListings$availability_365, y=nycListings$number_of_reviews)) +  
  
geom_point()+ geom_smooth(method = "lm")+  
  
stat_smooth(method="lm", se=FALSE)+  
  
labs(title="Number of Reviews by Availability") +  
  
xlab("Days Available in the Year") +  
  
ylab("Number of Reviews") +  
  
theme_tufte()  
  
  
lm_nycListings <- lm(number_of_reviews ~ availability_365, data=nycListings)
```

```
lm_nycListings
```

```
summary(lm_nycListings)
```

```
coef(lm_nycListings)
```

```
# (Intercept) availability_365
```

```
# 16.52898386    0.06648194
```

```
# For every day of availability, you will get 0.06645 more reviews, assuming a base 16.5 reviews. Not significant. Under .2 is a week correlation.
```

```
# Probably linked to length of listing rather than availability
```

```
# Number of people acommodated increases price in Brooklyn Entire Home/Apt
```

```
# Bring in detailed version of listings to get number of acommodated
```

```
nycListingsDetailed <- read.csv("../DataSets/insideairbnb/listings_detailed.csv", header=TRUE, check.names = TRUE)
```

```
# Create data set with data needed
```

```
nycListingsAcommodates <- select(nycListingsDetailed, id, name, host_id, host_since, host_response_rate, host_is_superhost, neighbourhood_cleansed, neighbourhood_group_cleansed, property_type, room_type, accommodates, bathrooms, bedrooms, beds, price)
```

```
head(nycListingsAcommodates)
```

```
# price needs to be cleaned before using, copy values from clean set
```

```
nycListingsAcommodates$price <- nycListings$price
```

```
# save it
```

```
write.csv(nycListingsAcommodates, "../DataSets/insideairbnb/listingsDetailedAcommodates.csv")
```

```
# Brooklyn rentals of entire home with mean price by number of people accommodated
```

```
bknEntireListingsAcommodates <- nycListingsAcommodates %>%
```

```
  filter(neighbourhood_group_cleansed == "Brooklyn") %>%
```

```
  filter(room_type == "Entire home/apt") %>%
```

```
  group_by(accommodates) %>%
```

```
  summarize(meanPrice = mean(price)) %>%
```

```
  arrange(desc(meanPrice))
```

```
summary(bknEntireListingsAcommodates)
```

```
head(bknEntireListingsAcommodates)
```

```
tail(bknEntireListingsAcommodates)
```

```
glimpse(bknEntireListingsAcommodates)
```

```
# Is there a relationship between number of people and median price
```

```
ggplot(bknEntireListingsAcommodates, aes(x=bknEntireListingsAcommodates$accommodates,  
y=bknEntireListingsAcommodates$meanPrice)) + geom_point()+
```

```
  stat_smooth(method="lm", se=FALSE)+
```

```
  labs(title="Price of Listing by Number of People in Brooklyn") +
```

```
xlab("Number of People") +
```

```
ylab("Price Per Night") +
```

```
theme_tufte()
```

```
lm_bknEntireListingsAcommodates <- lm(meanPrice ~ accommodates,  
data=bknEntireListingsAcommodates)
```

```
lm_bknEntireListingsAcommodates
```

```
#Coefficients:
```

```
# (Intercept) accommodates
```

```
# -187.26    68.12
```

```
# This is a significant and strong relationship but you need at 3 or more to reach a postive baseline. For  
every person that a listing can accomodate there is an increase of 68.12 in the price.
```

```
summary(lm_bknEntireListingsAcommodates)
```

```
coef(lm_bknEntireListingsAcommodates)
```

```
# Brooklyn rentals of entire home with mean price by number of beds
```

```
bknEntireListingsBeds <- nycListingsAcommodates %>%
```

```
filter(neighbourhood_group_cleansed == "Brooklyn") %>%
```

```
filter(room_type == "Entire home/apt") %>%
```



```
group_by(beds) %>%
```

```
summarize(meanPrice = mean(price))%>%
```

```
arrange(desc(meanPrice))
```

```
#Beds that are 0, empty "" and NA to be marked as 1
```

```
bknEntireListingsBeds$beds[bknEntireListingsBeds$beds==""] <- 1
```

```
bknEntireListingsBeds$beds[bknEntireListingsBeds$beds=="0"] <- 1
```

```
bknEntireListingsBeds$beds[bknEntireListingsBeds$beds==NA] <- 1
```

```
bknEntireListingsBeds$beds[20] <- 1
```

```
bknEntireListingsBeds$beds[20]
```

```
summary(bknEntireListingsBeds)
```

```
head(bknEntireListingsBeds)
```

```
tail(bknEntireListingsBeds)
```

```
glimpse(bknEntireListingsBeds)
```

```
# What is the relationship between number of beds and median price plotted out
```

```
ggplot(bknEntireListingsBeds, aes(x=bknEntireListingsBeds$beds,  
y=bknEntireListingsBeds$meanPrice)) + geom_point()+
```

```
stat_smooth(method="lm", se=FALSE)+
```

```
labs(title="Price of Listing by Number of Beds in Brooklyn") +
```

```
xlab("Number of Beds") +
```

```
ylab("Price Per Night") +
```

```
theme_tufte()
```

```
lm_bknEntireListingsBeds <- lm(meanPrice ~ beds, data=bknEntireListingsBeds)
```

```
lm_bknEntireListingsBeds
```

```
#Coefficients:
```

```
# (Intercept)    beds
```

```
# 198.40    26.96
```

```
summary(lm_bknEntireListingsBeds)
```

```
coef(lm_bknEntireListingsBeds)
```

```
# There is an increase of 26.96 in the price for every bed available in a listing with a base price of 198.40.
```

```
# If I want to narrow it to Crown Heights?
```

```
crownEntireListingsBeds <- nycListingsAcommodates %>%
```

```
filter(neighbourhood_group_cleansed == "Brooklyn") %>%
```

```
filter(neighbourhood_cleansed == "Crown Heights") %>%
```

```
filter(room_type == "Entire home/apt") %>%
```

```
group_by(beds) %>%
```

```
summarize(meanPrice = mean(price)) %>%
```

```
arrange(desc(meanPrice))
```

```
summary(crownEntireListingsBeds)
```

```
#Beds that are 0, empty "" and NA to be marked as 1
```

```
crownEntireListingsBeds$beds[crownEntireListingsBeds$beds==""] <- 1
```

```
crownEntireListingsBeds$beds[crownEntireListingsBeds$beds=="0"] <- 1
```

```
crownEntireListingsBeds$beds[crownEntireListingsBeds$beds==NA] <- 1
```

```
head(crownEntireListingsBeds)
```

```
hist(crownEntireListingsBeds$beds)
```

```
# What is the coefficient of number of beds and median price
```

```
ggplot(crownEntireListingsBeds, aes(x=crownEntireListingsBeds$beds,  
y=crownEntireListingsBeds$meanPrice)) + geom_point()+
```

```
stat_smooth(method="lm", se=FALSE)+
```

```
labs(title="Price of Listing by Number of Beds in Crown Heights") +
```

```
xlab("Number of Beds") +
```

```
ylab("Price Per Night") +
```

```
theme_tufte()
```

```
lm_crownEntireListingsBeds <- lm(meanPrice ~ beds, data=crownEntireListingsBeds)
```

```
lm_crownEntireListingsBeds
```

```
# (Intercept)    beds
```

```
# 230.30    16.44
```

```
summary(lm_crownEntireListingsBeds)
```

```
coef(lm_crownEntireListingsBeds)
```

```
# There is an increase of 16.44 for every bed added with a base price of 230.30.
```

```
# Can I make a prediction for a 2 bedroom apartment?
```

```
head(lm_crownEntireListingsBeds)
```

```
new_place <- data.frame("beds" = 2)
```

```
predict(lm_crownEntireListingsBeds, newdata=new_place)
```

```
# The price for a two bedroom in Crown Heights is $263.17.
```