Exploratory Data Analysis of Apartment Postings

Introduction

Craigslist is a website for posting local classified advertisements on various subjects including housing rentals. Postings are grouped by location for easy access. This report analyzes a dataset of Craigslist apartment postings in California. The aim of this report is to run visualizations to see how features are correlated to one another as well as provide summary statistics and answer key questions. By taking an empirical approach to evaluate the effect of features on price, I let the data craft its own story. Ultimately, this information may prove useful to potential renters searching for the ideal apartment. The data was downloaded from the Craigslist website. The analysis is performed in Rstudio.

1. Overview of the Dataset

1.1) Observations

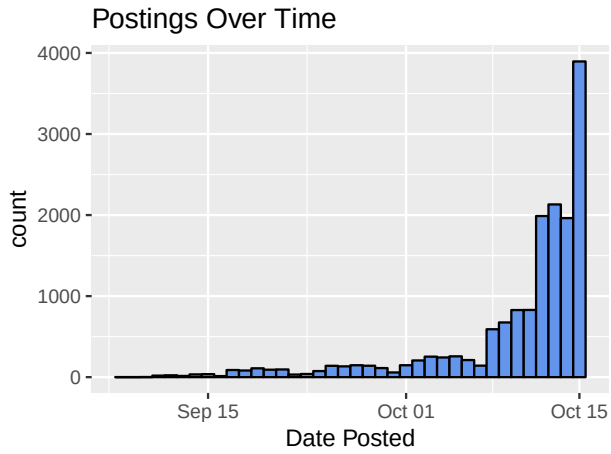
There are 21,948 initial observations and 20 features. Each unit of observation is a row corresponding to an apartment posting on Craigslist. Each column corresponds with features listed for title, text, date, price, etc of the apartment. The features are various data types such as numeric, integer, logical, character, and factor. There is information for the price of the apartment, size in square feet, location, and other factors important to consider when searching for an apartment.

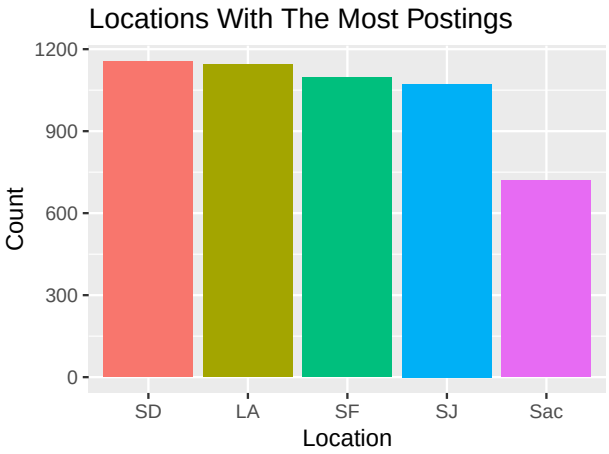
1.2) Data Quality

Quality of data is important for running accurate analyses. After taking a preliminary glance at the data using head() and str(), I found various errors in the form of outliers too extreme to be realistic. For instance, it's impossible an apartment will have a price of $0 or a size of only 1.0 sq ft. These errors were cleaned from the data set prior to analysis and are addressed in detail in part 5 "Limitations". After the data was cleaned, there are 21,718 observations left for analysis.

1.3) Date and Location Span

The postings span from 2018-09-08 to 2018-10-15, about one month. The time series bar plot shows most ads were posted near the end of the time period in mid-October. Locations of postings are mostly in California, from larger cities to smaller towns. The plot shows the five locations with the most postings are San Francicso, Los Angeles, San Diego, San Jose, and Sacramento.





1.4) Overview of Price

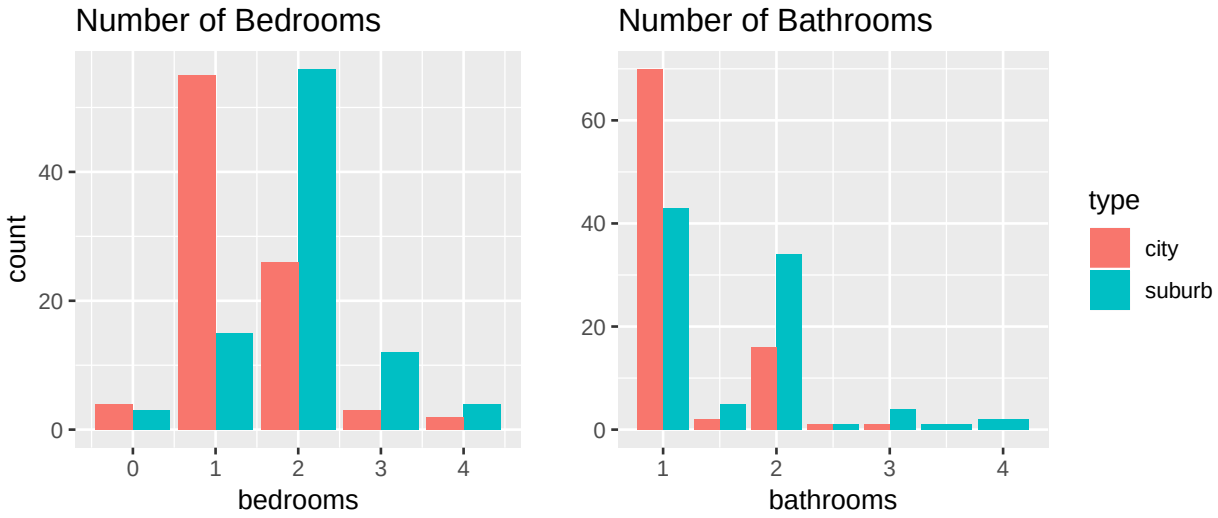
Price is one of the most important aspects of choosing an apartment. Important features such as size in sq ft, number of bedrooms, number of bathrooms, and location may affect the price of rent. From the summary statistics, I find the prices range from $0 to $17,700 and learn the mean price is $2,459 and the median price is $2,285. The histogram below is the distribution of price plotted for apartments, showing a skewed right-tailed distribution.

2. Initial Questions

2.1) Are apartments in suburbs more likely to be family-friendly (many bedrooms, pets allowed, etc) than apartments in major cities?

Subsets of the data were used to answer this question. For city apartments, a random sample of 90 postings were taken from the city of San Francisco. For suburban apartments, a sample of 90 postings were taken from places specifically defined as a "town" in the Bay Area. These towns include places like Atheron, Colma, and Corte Madera.

The mean number of city bedrooms is 1.3778 compared to the mean number of suburban bedrooms which is 1.9889. The mean number of city bathrooms is 1.2278 compared to the mean number of suburban bathrooms which is 1.6056. Suburbs on average have a higher number of bedrooms and bathrooms.



As shown in the bar charts, the number of bedrooms and bathrooms in suburban apartments outnumber the ones in the city. The distribution of city is skewed with most having one bedroom, whereas the distribution of suburban apartments is centered around two bedrooms. Most city apartments have one bathroom, while there is a more equal amount of suburban apartments with either one or two bathrooms. Overall, the suburbs have more 2+ bedroom and 2+ bathroom apartments (important for families needing more living space). Pet policy is examined using the proportion of city apartments that allow pets.

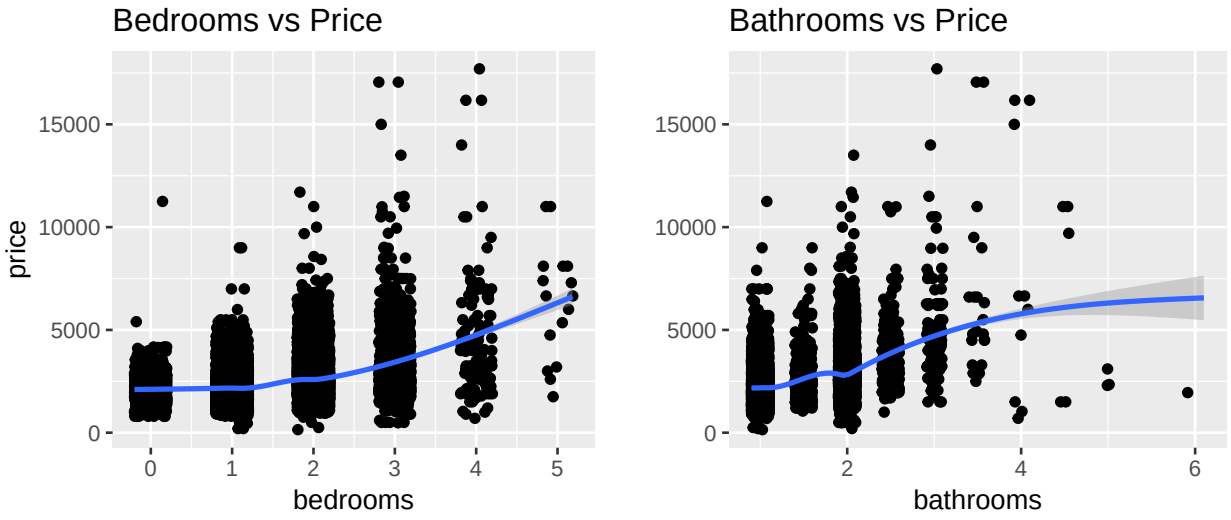
```
##
##       both       cats       dogs negotiable       none
## 0.48863636 0.03409091 0.03409091 0.00000000 0.44318182
```

This table is the proportion of suburban apartments that allow pets.

```
##
##       both       cats       dogs negotiable       none
## 0.50000000 0.03333333 0.03333333 0.00000000 0.43333333
```

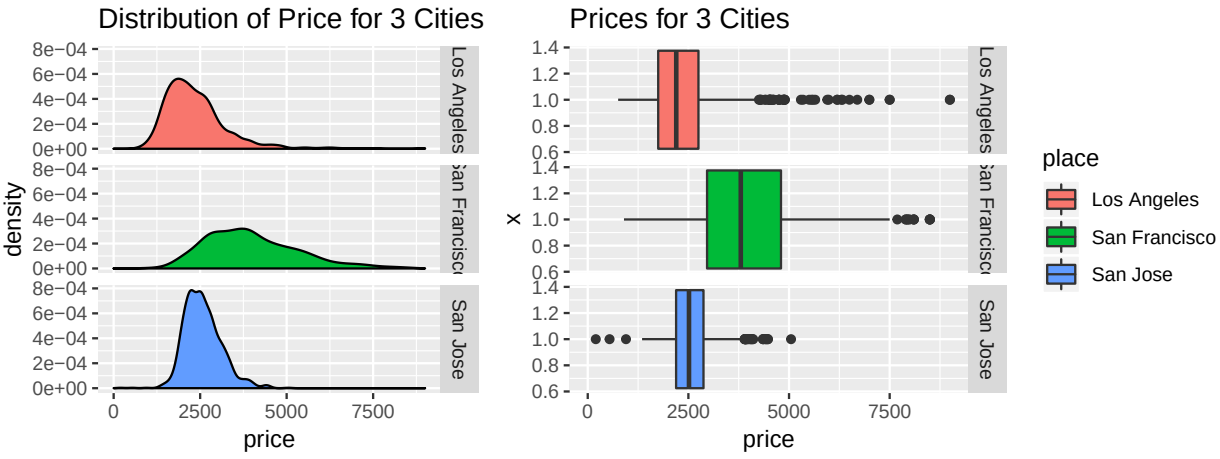
As shown by the tables, the city and suburban apartments come fairly close in terms of pet policy with surburban have a slightly higher proportion of 'both pets' allowed and city have a slightly higher proportion of 'none' allowed. By comparing the number of bedrooms, bathrooms, and pet policy proportions, suburban apartments appear to have a significantly higher advantage than city apartments. If families especially prefer larger apartments, then suburban apartments are the way to go.

2.2) Which adds more to rent: extra bedrooms, or extra bathrooms?

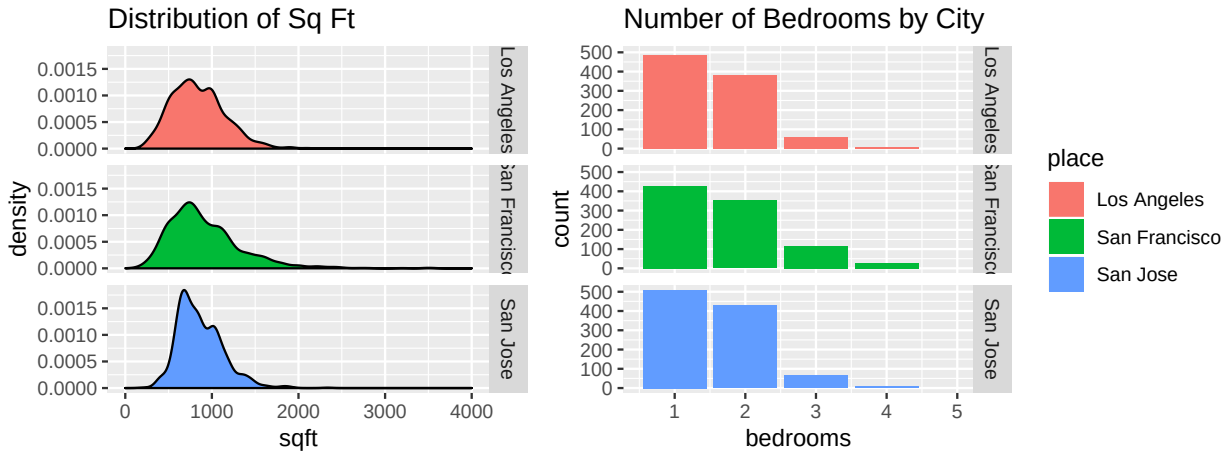


Extra bedrooms have a greater increase on rent. The scatterplots with model lines generated by the `method=gam` argument within geom_smooth() show a slightly steeper and upward curving line for bedrooms vs. price. This indicates for every increase in bedrooms, price increases slightly more than for every increase in bathrooms. This makes intuitive sense, since adding more bedrooms should increase apartment size more than adding more bathrooms, and higher prices will reflect this change in size. As the number of bedrooms goes up, price will increase even more for extra bedrooms judging by the upward curve of the line. However, price doesn't increase by as much (ie: is weakened) if adding extra bathrooms judging by the flatter slope at higher ends of the x axis. Important note: there is more variation in price and less observations available at higher numbers of bedrooms and bathrooms, so the accuracy of this effect measured in the data comes into question.

2.3) Do apartments in similar geographical areas tend to be similar?



For the purposes of this analysis, 'similarity' is defined by the similarity of price and size. Three cities are examined: San Francisco, Los Angeles, and San Jose. San Francisco has the greatest variability in price, while San Jose has the least. In other words, apartments in San Francisco are less similar to each other in price based on the flattened distribution. Apartments in San Jose are the most similar to each other in price based on the narrow shape of the distribution. Los Angeles is in the middle with a medium amount of variation.



Based on the shape of the distribution by sq ft, Los Angeles and San Francisco have more variation in size. More San Jose apartments tend to be around the same size, a little less than 1000 sq ft. Using the visual of the bar plot, I see that all three cities have mostly one or two bedroom apartments, with only very few have three or more. In conclusion, it depends on the region. Some cities, such as San Francisco, have greater variation in price and size among apartments and thus a wider range of different apartments available. Like San Francisco, Los Angeles has a wider range of sizes and prices available and thus its apartments are less similar. Other places such as San Jose have more apartments similar to each other in price and size.

3. What kinds of questions can be answered with this data set?

For my questions, I am interested in the best way to find apartments that would fit potential renters. Different renters have different requirements. For example, someone who is a student or new graduate may only need one bedroom and have a lower budget, whereas families may need more room and have a larger budget.

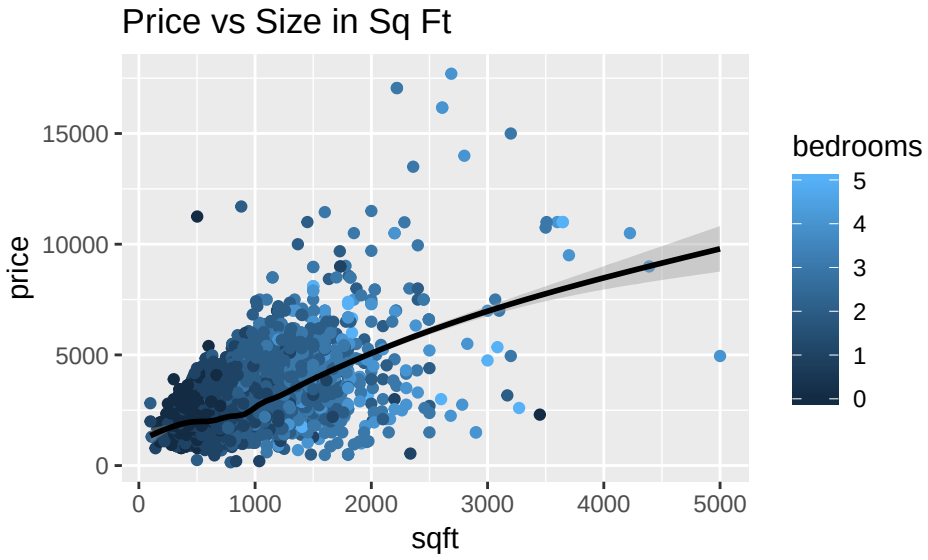
1. What are the common criteria for lowest cost apartments closest to major cities?
   - In other words, this question seeks the best apartment with low cost and good location near the workplace suitable for new grads, since young working professionals are more likely to move to major cities. These types of apartments may have certain factors in common.
2. How strong is the correlation between size and price?
   - I investigate the true relationship between price and size. Potential renters making a decision on apartment size would be interested to know if bigger apartments are actually more expensive or if other factors are more influential on price.
3. Does the rent per sq ft change for different areas?
   - Potential renters considering size, price, and location are interested in the relationship between these three variables. If they are considering different locations to move to, they may prefer a less expensive area with a lower price-to-size ratio.
4. Are there certain days of the week when cheaper apartments are posted?
   - People browsing Craigslist would be more likely to find a lower price posting if they know when they are posted. If there is a correlation, then there would be optimal times of the week to search Craigslist.
5. Is there a correlation between pet policy and price? Between parking and price?
   - Apartments that allow pets or have better parking may be comparatively more expensive than apartment that don't. Renters preferring a lower price may choose to forgo pets or have off-street parking.
6. What cities are most suitable for new graduates to live on a budget?
   - This question seeks to find the cities with the greatest number of suitable apartments that meet the criteria of low cost and ideal location. To answer this question, I could combine the Craigslist analysis with information from outside sources about cities offering the most new jobs. Graduates deciding on an ideal location could focus their search on these cities.

7. How much did the price of rent change throughout the month?
    - Though the data provided spans a relatively short period of time, we could still analyze price changes over the month. For practical purposes, perhaps there is an ideal time to search during the month when the lowest rent apartments are posted.
8. Is the data from Craigslist an accurate reflection of actual apartment rent in California?
    - This question is useful for evaluating the reliability of our data. Apartments on Craigslist might be more or less expensive than other sites. If Craigslist is more expensive, then seekers might want to check other sources of postings for the best fit. Outside data on California rent would need to be obtained to compare with Craigslist.
9. What's the most pet-friendly city?
    - Pet lovers who are deciding on a location would be interested to know where the greatest number of pet-frendly apartments are located. The criteria of "pet-friendly" could include not just pet policy, but also the presence of features such as yards or nearby parks.
10. What's the most expensive place to live?
    - With the rise of rent prices, certain places such as the Bay Area and Silicon Valley are becoming more expensive to live in. It would be interesting to see how the data can corroborate this, as well as useful for people considering moving to these areas. The rent for individual cities and towns can be compared, or a more general cross-county analysis can give us a bigger picture of price distribution across California. Additionally, we can see what features or amenities might be causing the price of apartments in this area to be so high.

4. Answering 5 questions about the data

4.1) How strong is the correlation between size and price?

Hypothesis: There is a strong correlation between size and price, and size is likely the strongest determinant of price overall.

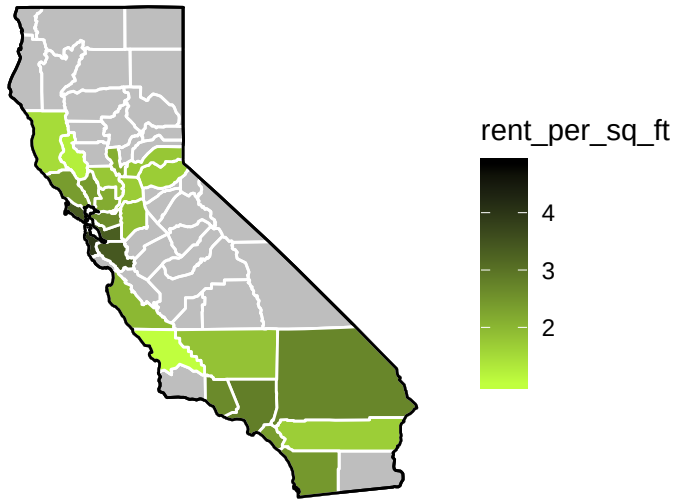


The Pearson correlation coefficient between sqft and price was found to be 0.5186, indicating a moderate uphill relationship. The scatterplot shows a clear correlation between sqft and price. Additionally, the color shows the positive relationship between bedrooms and price. While difficult to compare the impact of location vs the impact of size, plots drawn between latitude and price or longitude and price (not shown) do not depict a strong relationship, if any.

4.2) Does the rent per square feet change for different areas?

Hypothesis: Yes, the ratio will change depending on location. For instance, San Francisco may have a higher price-to-size ratio compared to Davis.

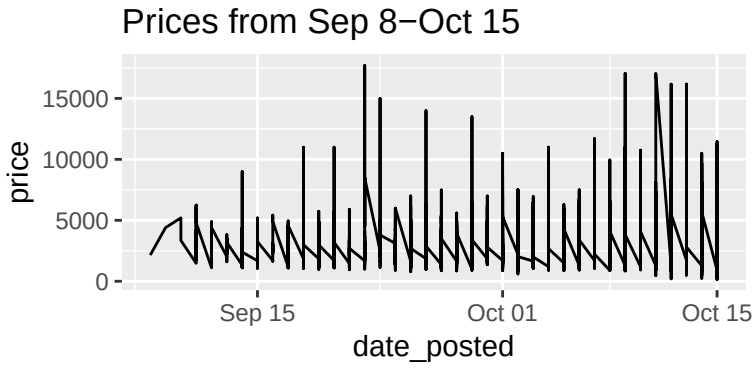
The rent per square feet can give us an idea of how rent changes from location to location. Calculating the price of one square foot will remove the variance in price due to differences in size, so that we can focus on the effect of location.



The change in colors on the map clearly show there is variation by county in the rent per sq ft. Places such as the San Francisco Bay Area and Los Angeles have the highest rent per sq ft, as shown by the darkest colors. THe most expensive counties contain California's major cities, which are also the areas with the most postings. The rent per sq ft decreases for more rural counties, as shown by the lighter color. (The gray areas have insufficient data)

4.3) How much did the price of rent change throughout the month?

Hypothesis: The price of rent doesn't change very much throughout a single month and stays relatively stable.



From the time series plot, we can see the price fluctuates throughout the month. There are increases in price near the second half of September and increases again approaching mid-October. This may be random fluctuations and not necessarily a sign of a pattern, and we will need more data to confirm if there are monthly patterns over the course of the year. Since the data was collected in the span of one month, trends over longer periods of time are not reflected. Perhaps various factors such as season, building regulations, and wages can influence the supply and demand of rentals over time.

4.4) What's the most pet-friendly city?

Hypothesis: A smaller city with more suburban-style residences will be the most pet-friendly city. Big metropolises such as San Francisco will be less pet-friendly, because apartments tend to be smaller.

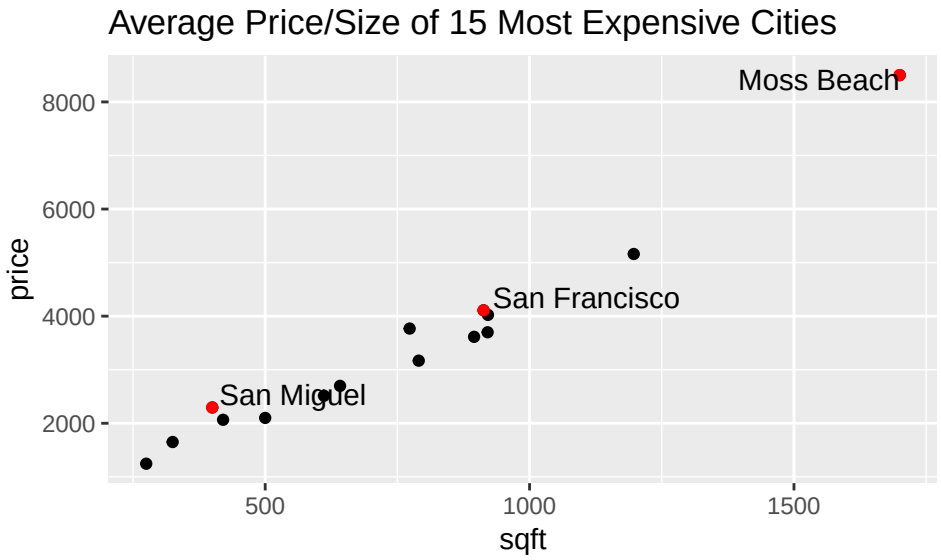
```
##          place both cats dogs negotiable none NA pets_allowed pets_sum
## 201    Rocklin  114    1    1          0    0  0    1.0000000      116
## 187  Pittsburg  136    0    0          0    2  0    0.9855072      138
## 209  Roseville  183    8    0          1    3  1    0.9846154      196
## 263      Vista  111    1    2          0    6  1    0.9500000      121
## 137 Long Beach   76   43    1          0    9  0    0.9302326      129
```

I calculated the proportion of postings that allowed pets (either both, cat, dog, or negotiable) by city. Only places with at least 100 postings that provided information about pets were examined, since places with less postings will have less accurate proportions. According to the data, Rocklin CA is the most pet-friendly city with the greatest proportion of postings that allow pets. Rocklin has a smaller population and does indeed have more of a suburban feel. Conversely, the least pet-friendly city is Palo Alto. Interestingly, most pet-unfriendly cities are located in Silicon Valley, likely because of more expensive real estate prices and smaller apartments.

4.5) What's the most expensive place to live?

Hypothesis: Cities in the Bay Area or Silicon Valley will be the most expensive in California.

This question can be interpreted in a number of ways. Both individual cities and towns were compared as well as across different counties. Without a doubt, San Francisco county is the most expensive county with a rent per sq ft of $6.5601. When ranking by cities, San Miguel has the highest rent per sq ft ratio. However, since there is only one posting for San Miguel, it is possible we need more observations to verify this. If subsetting by cities with at least fifty postings or more, San Francisco has the highest rent per sq ft ratio.



The scatterplot shows the 15 most expensive cities (determined by the rent to sq ft ratio) and their average price and sq ft. The red dots show the location of San Miguel and San Francisco. While San Miguel doesn't have the highest average price and size overall, its relative rent to sq ft ratio ranks the highest. Moss Beach takes the spot for the highest average price and size.

5. Limitations: Missing Values, Errors, and Outliers

From the original dataset, there are 26,316 missing values in total. Most are in the date_updated column with 13,139 missing. Since many posters don't update after their initial posting, it makes sense there are many postings that lack a date updated. Some missing values follow patterns in the sense that they are related to other missing values. For example, there are exactly 84 missing latitude and longitude, and 1048 missing for both bedrooms and bathrooms. Since these pairs of features go together, the postings must have originally lacked both their information.

There were two very extreme outliers for price: $34,083,742 and $9,951,095. Further inspection of these postings' texts shows the prices are in error because they list a price range and not just a single price. The $34,083,742 posting actually describes the price as "$3,408 - $3,742", not $34,083,742 as mistakenly inputted in price. Similarly, the $9,951,095 post actually lists $995 for a 1bdrm and $1,095 for a 2brm. These errors are corrected in the data by taking the mean of the prices.

Some postings were not actual apartment ads, but instead all manner of related postings such as postings by people looking for a place to live or companies advertising renters' insurance. Any observations with prices lower than $100 were removed from the dataset as irrelevant to the analysis of apartments. Similarly for size, any observation with sqft less than 50 or larger than 5000 were removed. By realistic standards, these postings are most likely not apartments or the number is taken in error. After 230 postings were removed from the dataset, the dataset was still sufficiently large enough for an accurate analysis.

The data set was generated by extracting web data from Craigslist. Bias exists in the sense that we only know limited information about apartments in a certain time window and state. There also aren't enough observations to make sound conclusions about specific locations. For example, San Miguel was found to be the most expensive city in the data set, but there was only one observation to support this. More diverse observations would be ideal to compare the cost of different regions, but there are just less apartments available in certain places compared to others.

We can't always infer conclusions about prices and location from the data. Since the data only came from Craigslist, information for rentals from other sources are not reflected. More data over a longer period of time could help identify long term trends and help refine precision of the overall analysis.

R Code

```
# ## ----setup, include=FALSE----------------------------------------------
knitr::opts_chunk$set(echo = FALSE, warning = FALSE, message = FALSE, fig.width = 5,
fig.height = 3)
require(ggplot2)
require(tidyverse)
require(ggpubr)
require(maps)
require(ggmap)
require(data.table)
require(kableExtra)
require(tm)
require(SnowballC)
require(wordcloud)
require(RColorBrewer)

## ------------------------------------------------------------------------
cl <- readRDS("cl_apartments.rds")
cl$date_posted <- as.Date(cl$date_posted)
cl$date_updated <- as.Date(cl$date_updated)

## ---- include=F--------------------------------------------------------
# Data Cleaning
cl <- cl[!(cl$price < 100 | cl$bathrooms == 0 | cl$sqft <50 | cl$sqft > 5000 ),]

which(cl == 34083742, arr.ind = TRUE)
which(cl == 9951095, arr.ind = TRUE)
cl[15961,]
cl[4531,]
cl[15961, "price"] = (3408 + 3742)/2
cl[4531,"price"] = (995 + 1095)/2

dim(cl)

## ---- echo=FALSE, fig.width = 8, fig.height=3--------------------------
time_plot <- ggplot(cl, aes(x = date_posted)) +
  geom_histogram(aes(y = ..count..), fill = "cornflowerblue", color = "black", binwidth=1) +
  labs(title = "Postings Over Time", x = "Date Posted")

top_five <- as.data.frame(head(sort(table(cl$place), decreasing=T), n=5))
names(top_five) <- c("Place", "Freq")
place_plot <- ggplot(top_five, aes(x = Place, y = Freq)) +
```

```
  geom_bar(stat="identity", aes(fill= Place)) +
  guides(fill=FALSE) +
  labs(title = "Locations With The Most Postings", x = "Location", y = "Count") +
  scale_x_discrete(breaks=c("San Francisco", "Los Angeles", "San Diego", "San Jose",
"Sacramento"), labels = c("SF", "LA", "SD", "SJ", "Sac"))

ggpubr::ggarrange(time_plot, place_plot, widths = c(1,1), heights = c(1,1))


## ---- include=FALSE-----------------------------------------------------
summary(cl$price)

## ---- echo=FALSE, fig.width = 5, fig.height=3---------------------------
regular <- filter(cl, price >= 10 & price <= 10000)
ggplot(regular, aes(x = price)) +
  geom_histogram(fill = "#FF9999", color = "black") +
  labs(title = "Distribution of Price")

## -----------------------------------------------------------------------
city <- filter(cl, place == "San Francisco")

city <- city[1:90,]

suburb <- filter(cl, place == "Atherton" | place == "Colma" | place == "Corte Madera" | place ==
"Danville" | place == "Fairfax" | place == "Hillsborough" | place == "Los Altos Hills" | place ==
"Los Gatos" | place == "Moraga" | place == "Portola Valley" | place == "Ross" | place == "San
Anselmo" | place == "Tiburon" | place == "Windsor" | place == "Woodside" | place ==
"Yountville")

suburb <- suburb[1:90,]

city$type <- c(rep("city"))
suburb$type <- c(rep("suburb"))



## ---- include=F--------------------------------------------------------
# Mean number of bedrooms
city_bdrm <- mean(city$bedrooms, na.rm=T)
suburb_bdrm <- mean(suburb$bedrooms, na.rm=T)

# Mean number of bathrooms
city_bth <- mean(city$bathrooms, na.rm=T)
```

```
suburb_bth <- mean(suburb$bathrooms, na.rm=T)


## ------------------------------------------------------------------------
city_suburb <- data.frame(rbind(city, suburb))

## ---- warning=FALSE, fig.width = 7---------------------------------------
bed_plot <- ggplot(city_suburb, aes(x = bedrooms, fill = type))+
  geom_bar(position = "dodge") +
  labs(title = "Number of Bedrooms") +
  guides(fill = FALSE)

bath_plot <- ggplot(city_suburb, aes(x = bathrooms, fill = type))+
  geom_bar(position = "dodge") +
  labs(title = "Number of Bathrooms", y="")

ggpubr::ggarrange(bed_plot, bath_plot, widths = c(1,1.3), heights = c(1,1))


## ------------------------------------------------------------------------
# Proportion of city apartments that allow pets
table(city$pets)/sum(table(city$pets))


## ------------------------------------------------------------------------
table(suburb$pets)/sum(table(suburb$pets))

## ---- warning=FALSE, fig.width = 7, fig.height = 3-----------------------
plot3 <- ggplot(cl, aes(x = jitter(bedrooms), y = price)) +
  geom_point() +
  geom_smooth(method = 'auto', se=T) +
  labs(title = "Bedrooms vs Price", x = "bedrooms")

plot4 <- ggplot(cl, aes(x = jitter(bathrooms), y = price)) +
  geom_point() +
  geom_smooth(method = 'auto') +
  labs(title = "Bathrooms vs Price", x = "bathrooms") +
  labs(y="")

ggpubr::ggarrange(plot3, plot4, widths = c(1,1.025), heights = c(1,1))


## ---- warning = FALSE, fig.width = 8, fig.height = 3--------------------
```

```
three_places <- filter(cl, place == "San Francisco" | place == "Los Angeles" | place == "San Jose")

plot1 <- ggplot(three_places, aes(x = price)) +
  geom_density(aes(fill = place)) +
  guides(fill=FALSE) +
  xlim(0,9000) +
  labs(title = "Distribution of Price for 3 Cities") +
  facet_grid(place~.)

plot2 <- ggplot(three_places, aes(x = 1, y = price)) +
  geom_boxplot(aes(fill = place)) +
  ylim(0, 9000) +
  coord_flip() +
  facet_grid(place~.) +
  labs(title = "Prices for 3 Cities")

ggpubr::ggarrange(plot1, plot2, widths = c(1,1.5), heights = c(1,1))


## ---- fig.width=8, fig.height=3----------------------------------------
plot5 <- ggplot(three_places, aes(x = sqft)) +
  geom_density(aes(fill = place)) +
  guides(fill=FALSE) +
  xlim(0, 4000)+
  facet_grid(place~.) +
  labs(title = "Distribution of Sq Ft")

plot6 <- ggplot(three_places, aes(x = bedrooms)) +
  geom_bar(aes(fill = place), position = "dodge") +
  facet_grid(place~.) +
  xlim(0.5,5) +
  labs(title = "Number of Bedrooms by City")

ggpubr::ggarrange(plot5, plot6, widths = c(1,1.3), heights = c(1,1))


## ----------------------------------------------------------------------
ggplot(cl, aes(x = sqft, y = price)) +
  geom_point(aes(color = bedrooms)) +
  labs(title = "Price vs Size in Sq Ft") +
  geom_smooth(method="auto", color = "black")
```

```
## ---- include=FALSE-----------------------------------------------------

cor(cl$price, cl$sqft, use = "complete.obs")


## ------------------------------------------------------------------------
# Calculate the rent per sq ft
cl$rent_per_sq_ft <- (cl$price)/(cl$sqft)
# New data frame of rent/sqft by county
counties <- aggregate(rent_per_sq_ft ~ county, cl, mean)

# Map making using 'maps' library
states <- map_data("state")
ca_df <- subset(states, region == "california")
ca_county <- subset(map_data("county"), region == "california")

ca_base <- ggplot(data = ca_df, mapping = aes(x = long, y = lat, group = group)) +
  coord_fixed(1.3) +
  geom_polygon(color = "black", fill = "gray") +
  theme_nothing() +
  geom_polygon(data = ca_county, fill = NA, color = "white") +
  geom_polygon(color = "black", fill = NA)

counties$subregion <- sapply(counties$county, tolower)

ca_rent <- inner_join(ca_county, counties, by = "subregion")

ditch_the_axes <- theme(
  axis.text = element_blank(),
  axis.line = element_blank(),
  axis.ticks = element_blank(),
  panel.border = element_blank(),
  panel.grid = element_blank(),
  axis.title = element_blank()
)

final_map <- ca_base +
  geom_polygon(data = ca_rent, aes(fill = rent_per_sq_ft), color = "white") +
  geom_polygon(color = "black", fill = NA) +
  theme_bw() +
  ditch_the_axes +
  scale_fill_gradient(low = "olivedrab1", high = "black")
```

```
final_map


## ---- fig.width = 4, fig.height=2----------------------------------------
ggplot(cl, aes(x = date_posted, y = price)) +
  geom_line() +
  labs(title = "Prices from Sep 8-Oct 15")

## ------------------------------------------------------------------------

# Calculate the proportion of pets allowed by place, then rank the places by decreasing order

pets1 <- cl %>%
  dcast(place ~ pets)

pets1$pets_allowed <- (pets1$both + pets1$cats + pets1$dogs + pets1$negotiable)/(pets1$both
+ pets1$cats + pets1$dogs + pets1$negotiable + pets1$none)

pets1$pets_sum <- rowSums(pets1[, 2:7])

pets2 <- subset(pets1, pets_sum > 100)

pets2 <- pets2[order(pets2$pets_allowed, decreasing = T),]

## ------------------------------------------------------------------------
head(pets2, n = 5)

## ---- include = F----------------------------------------------------
head(pets2[order(pets2$pets_allowed, decreasing = F),])

## ---- include = FALSE-----------------------------------------------
arrange(counties, desc(rent_per_sq_ft))

places <- aggregate(rent_per_sq_ft ~ place, cl, mean)

places <- places[order(places$rent_per_sq_ft, decreasing = T),]

places2 <- places[order(places$rent_per_sq_ft, decreasing = F),]

places <- places[1:5,]
```

```
places2 <- places2[1:5,]

places #top five most expensive by rent per sq ft ratio

places2 #five least expensive

## ------------------------------------------------------------------------
sum_prices <- aggregate(price ~ place, cl, mean)
sum_size <- aggregate(sqft ~ place, cl, mean)
prices_size <- merge(sum_prices, sum_size, by = "place")
prices_size$rent_per_sq_ft <- prices_size$price/prices_size$sqft
prices_size <- subset(prices_size, rent_per_sq_ft > 4)
ggplot(prices_size, aes(x = sqft, y = price)) +
  geom_point() +
  geom_point(data = subset(prices_size, place == "San Francisco" | place == "San Miguel" |
place == "Moss Beach"), color = "red") +
  geom_text(data = subset(prices_size, place == "San Francisco" | place == "San Miguel"),
aes(label = place), hjust = -.05, vjust = -.1) +
   geom_text(data = subset(prices_size, place == "Moss Beach"), aes(label = place), hjust = 1,
vjust = .7) +
  labs(title = "Average Price/Size of 15 Most Expensive Cities")`
```

Sources

https://en.wikipedia.org/wiki/List_of_cities_and_towns_in_the_San_Francisco_Bay_Area