

# Evaluation of Monthly Apartment Rental Prices in Honolulu, HI

TIFFANY EULALIO AND JAKE KOKI

Mathematics 373 (Elementary Statistics)

Mentor: Dr. Bjoern Kjos-Hanssen

---

*Renting an apartment is difficult, and choosing between larger living space, featured amenities, and affordability can be a daunting task. This is especially true for residents in Hawai'i, where housing costs are the highest in the nation. Thus, understanding what we are paying for when renting is critical. In this study, we analyze factors that influence the rental price of apartments in Honolulu, Hawai'i (HI). We focus on the correlation between an apartment's size (sq ft) and its rental price (\$). We also analyze the effect that amenities such as laundry units, parking spaces, pet allowance, floor number, and view type have on the rental price. We find a relatively high correlation between square footage and rental price ( $R^2 = 42.6\%$ ). We also observe that the view type of an apartment affects rental price, with ocean view incurring the highest cost, followed by mountain view, and lastly, city view. However, no correlation between the other amenities are apparent from our study.*

---

## Introduction

In 2016, Hawai'i was ranked as the most expensive place to live in the United States with an overall cost of living index of 167.4, and a particularly staggering housing index of 230.3, with respective average indices of 100 [1]. Due to the high cost of living, finding a house should be optimized by careful quantitative and qualitative comparisons of available properties. This is especial-

ly important for many first-time home owners, whose first home will most likely be an apartment. Thus, this report aims to examine apartment data, and from this, extract factors that affect price. From this, a reasonable extrapolation can be made of what factors make an apartment expensive, and hence what factors to look out for when house-hunting. Apartment data was extracted from [www.zillow.com](http://www.zillow.com) [2], a publicly accessible database of homes for sale. The apartment data was then analyzed with Microsoft Excel.



I am currently a mathematics major with a minor in chemistry looking to pursue further education in either applied mathematics or machine learning. (Jake)

I am currently a Computer Science and Biology double-major with a minor in Mathematics with the goal of attaining a PhD. I want to apply two of humankind's most useful tools—math and computers—to discovering ways of alleviating some of the problems that plague our world. (Tiffany)

Real-world data sets are often messy, unlike textbook statistics problems, and thus the most difficult part of writing this was working with apartment listings that were missing information.

The results here can be useful for anyone and will likely apply outside of the topic of apartments as well. The unexpected results pertaining to amenities really demonstrated the power of statistics to reveal patterns that aren't initially obvious.

## Methods

### DATA SOURCE

Apartment rental data was collected from the property sales website [www.zillow.com](http://www.zillow.com) during February of 2017. Using the website's built-in search filters, we narrowed our search to "Apartments", in "Honolulu, HI", listed as "For Rent", by a "Verified Source", which means that the apartment was listed by "a known and established property manager or property manager's representative" [7]. The results of this filtered search returned an average of 25 pages, with 26 apartment listings per page.

We collected the following data from a selection of these apartment listings: Rental Price (\$), Apartment Size (sq ft), Number of Beds, and Number of Baths. We also recorded floor number of the apartment unit, whether pets were allowed, the type of view from the apartment, and the availability of a laundry machine or parking space.

### DATA SELECTION

The data was collected in three separate batches, initially with a sample of 25 apartments. We used a random number generator (RNG) website [6] to obtain 25 numbers from 1–100 with no repeats allowed. We then considered the apartment listings as elements numbered in order, starting from the top of page 1, increasing towards the bottom of the page. This ordering continues through the pages until the last apartment listing on the last page is reached. Then, for the 25 randomly generated numbers, we find the corresponding apartment listing that was assigned that number. Data was collected from these 25 listings for the first batch. Next, we decided to add more data points to improve the accuracy of our experiment. We collected a second batch of numbers, which consisted of 10 more data points. The selection process of these apartments relied on two RNG's. The first RNG generated a number,  $P$ , which was between 1–10, and the second RNG generated a number,  $L$ , which was between 1–25.  $P$  determined which page number the apartment was on, and  $L$  determined which apartment listing on that page was selected. The last batch of 10 apartments was collected in the same manner as the second batch with an additional search filter to ensure that the apartments were at least 1,000 square feet. This last batch was necessary because our data was sparse for apartment sizes in this range. To better represent these apartments, we refined the Apartment Size search parameter.

## HANDLING MISSING VARIABLES

There were several cases where the randomly generated numbers lead us to select apartments which contained missing data fields for our variables of interest. In most cases, if a variable was missing, we kept the apartment listing and left the variable field empty. However, there were two variables that were imperative to our study, Apartment Size and Rental Price. If either of these two variables were missing, then we did not accept the apartment listing into our sample. In the first batch, we simply removed these apartments from the sample. In the second and third batches, if we noticed that these variables were missing, we selected the next apartment listing in the ordering. In some cases, the next listing was also missing vital variables, so we moved on to the subsequent listings until the Apartment Size and Rental Price variables were present. Our final sample size consisted of 42 apartments. The missing 3 apartments resulted from removing entries from our first batch of sampling.

### DATA ANALYSIS

After selecting the apartments for our study, we input our data into a Microsoft Excel document. We used the built-in features of the program to create our graphs and add any additional statistical analyses. One exception was the standard deviation of the residuals that was reported for the Apartment Size vs. Rental Price analysis. We calculated this value manually to ensure that the proper equation was used. The equation was obtained from a Statistics textbook used in college Statistics courses [5]. Regression analysis was performed and regression lines were added to all scatter plots. Due to the inability to alter colors in the box-and-whisker plots obtained from Microsoft Excel, we manually edited the colors in Adobe Photoshop. No other attributes of the graphs were edited other than the fill color of these plots.

## Results

### APARTMENT SIZE VS. RENTAL PRICE

The focus of our study was to determine whether there was a correlation between Apartment Size in square feet (AS) and Rental Price in US dollars (RP) for apartment rentals in Honolulu, HI. We performed a regression analysis (Figure 1), and our data revealed some correlation

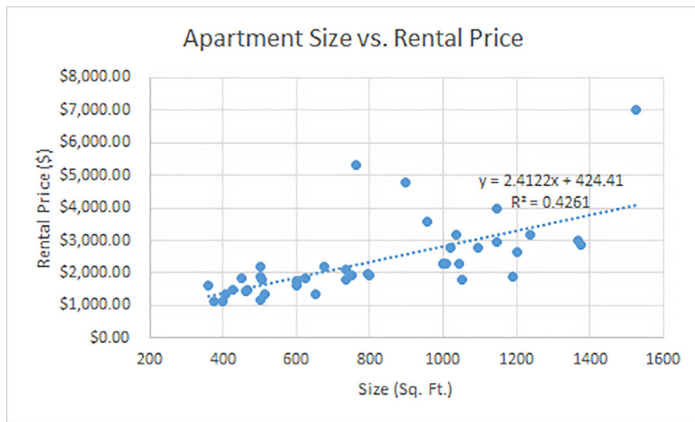


Figure 1 Apartment Size (sq ft) vs. Rental Price (\$) of apartments in Honolulu, HI.

between Apartment Size and Rental Price ( $R^2 = 42.6\%$ ). The regression line is represented by the equation  $RP = 2.4122 \cdot AS + 424.41$ . Rental price of apartments ranged from \$1,150 to \$7,000. The median price was \$1,956 ( $n = 42$ ,  $SD = \$1,178$ ). Apartment size ranged from 360 to 1,527 sq ft. The median size was 755.5 sq ft ( $n = 42$ ,  $SD = 318.96$  sq ft). The scatter plot of the residuals (Figure 2) shows no pattern, but reveals the presence of a few outliers. We also see that the variation in data points increases with the Apartment Size. In the histogram of the residuals (Figure 3), we see a unimodal normal distribution (Median = 0.15,  $Se = 0.884$ ). The three outliers can be seen more obviously in the histogram than in the scatter plot.

CONFOUNDING VARIABLES

View Type

Additional analyses were performed to determine whether apartment amenities behaved as confounding variables. We created subgroups of the data based on the view types of each apartment. The view types fell into three main categories: ocean ( $n = 5$ ), mountain ( $n = 7$ ), and city view ( $n = 17$ ). A comparison between the view types and their associated Apartment Size and Rental Price correlations can be seen overlaid in Figure 4. Apartments with city views had the strongest correlation between Apartment Size and Rental Price ( $R^2 = 61.4\%$ ), and a regression line described by  $RP = 1.6206 \cdot AS + 782.08$ . Apartments with ocean views had a weaker correlation ( $R^2 = 44.6\%$ ) and a regression line described by  $RP = 3.9736 \cdot AS + 742.94$ . The weakest correlation occurred with apartments with mountain views ( $R^2 = 3.4\%$ ),

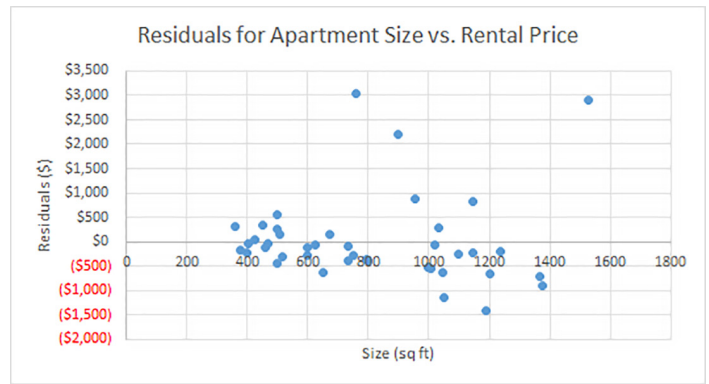


Figure 2 Residuals scatter plot for Apartment Size (sq ft) vs. Rental Price (\$) for apartments in Honolulu, HI.

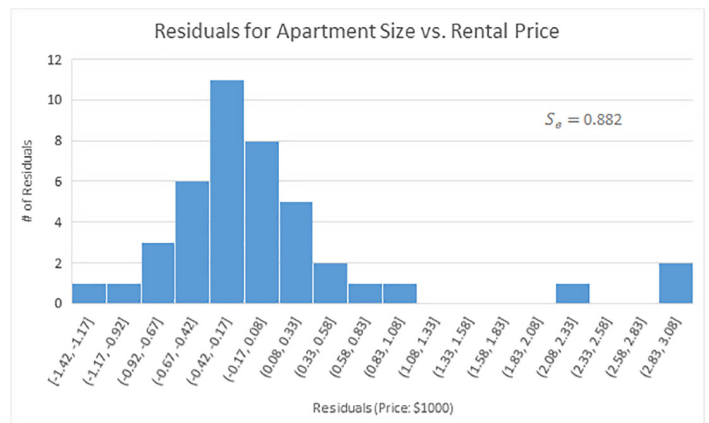


Figure 3 Residuals histogram for Apartment Size (sq ft) vs. Rental Price (\$) for apartments in Honolulu, HI.

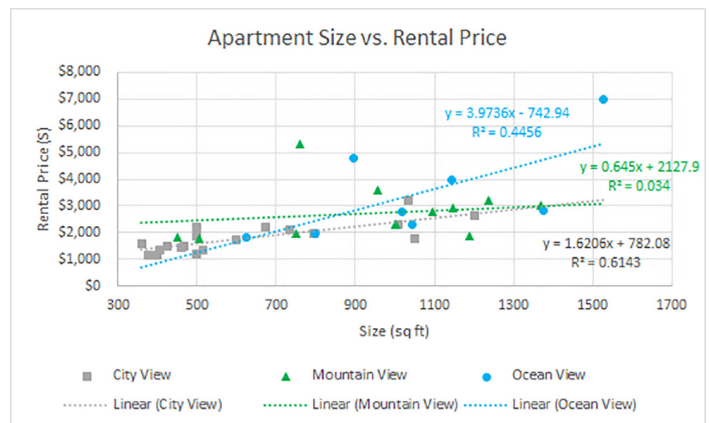
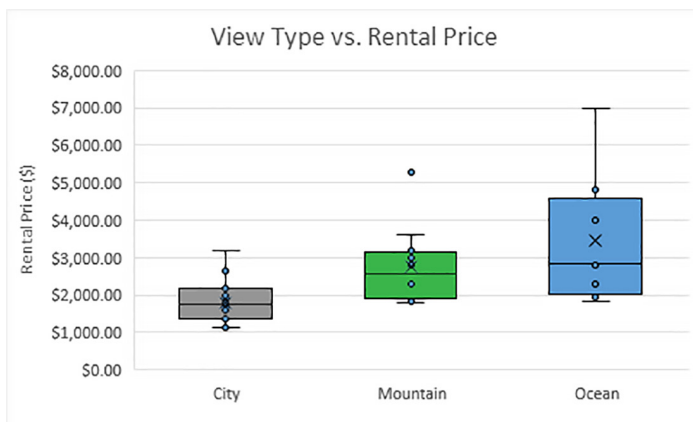
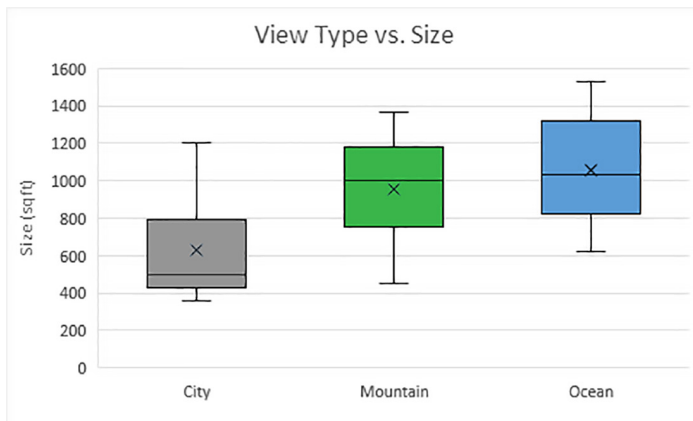


Figure 4 Apartment Size (sq ft) vs. Rental Price (\$) of apartments in Honolulu, HI, subgrouped by View Type (City, Mountain, Ocean).

which had a regression line described by  $RP = 0.645 \cdot AS + 2127.9$ . We made a direct comparison between the Rental Price of each view type (Figure 5). We see that city view apartments are the cheapest (Median = \$1,750, SD = \$545), followed by mountain view apartments (Median = \$2,550, SD = \$1,003), and lastly, ocean view apartments are the most expensive (Median = \$2,825, SD = \$1,758). A similar comparison is made for Apartment Size of the different view types (Figure 6). We see a similar pattern, with the city view apartments tending to be the smallest (Median = 500 sq ft, SD = 239 sq ft), mountain view apartments being larger (Median = 761 sq ft, SD = 279 sq ft), and ocean view apartments being the most spacious (Median = 896 sq ft, SD = 342 sq ft). Our error bars indicate that there is no significant difference between any of the view types. We can only make general statements about these relationships.



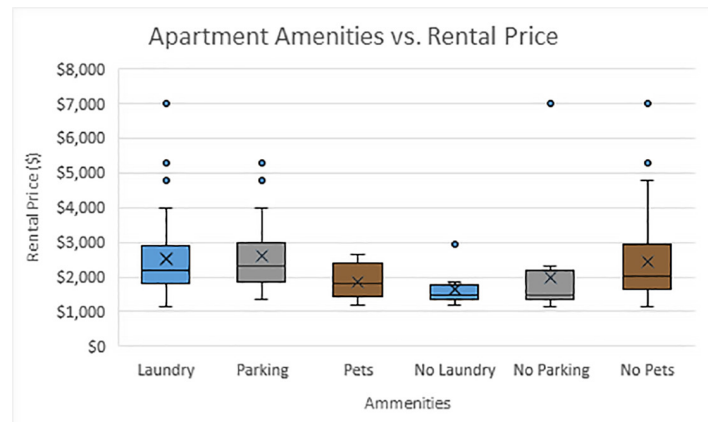
**Figure 5** Rental Price (\$) of apartments in Honolulu, HI based on the view type available.



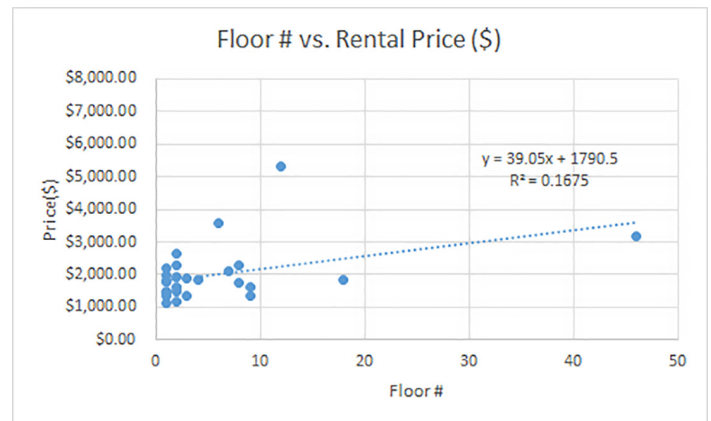
**Figure 6** Apartment Size (sq ft) of apartments in Honolulu, HI, based on view type available.

### Amenities

We analyzed the effect that amenities have on the Rental Price as well. We recorded data indicating whether an apartment unit contained a laundry machine, whether a parking space was provided and included in the rental price, and whether pets were allowed. The apartment rental price based on the availability (and unavailability) of these amenities were evaluated (Figure 7). We found no significant effects based on the availability of these amenities. We evaluated whether the floor number of an apartment unit influenced Rental Price (Figure 8). There was a very weak correlation between floor number and Rental Price ( $R^2 = 16.8\%$ ).



**Figure 7** Rental Price (\$) of apartments in Honolulu, HI, based on the availability of select amenities: laundry availability, parking availability, pet allowance.



**Figure 8** Floor number vs. Rental Price (\$) for apartments in Honolulu, HI.



## Discussion

### SIZE (SQ FT) VS. PRICE (\$)

In observational studies, an  $R^2$  between 30% and 50% is considered a “useful” correlation [5]. Since the regression line of Apartment Size vs. Rental Price showed an  $R^2$  value of 42.6%, it is reasonable to consider a relationship between how big an apartment is and how much it costs. Furthermore, we have good reason to trust a regression line if there is no discernible pattern in the residuals plot, which is to say a reliable regression has a normally distributed histogram of residuals. As we have observed, the residuals histogram showed a reasonably normal distribution with some skew to the right due to outliers, and thus we have a reliable regression line to draw conclusions from.

One thing to notice upon observing the regression line itself is the noticeable increase in price variation as square footage increases. A potential cause of this is due to the ambiguity of how apartment size is measured. For example, patio space, communal rooms, etc. can be counted in an apartment’s size and reasonably have less effect on the rental price. Another thing to look at is the equation for the regression line. We can interpret the slope as a roughly \$2.41 increase per square foot of an apartment. Interpreting the intercept is, although humorous, not very useful as it implies that a 0-square foot apartment will cost \$424.41.

Although we have found reasonable indication of a relationship between square feet and price, there are undoubtedly more factors that affect price other than size [3], [4]. Perhaps the more convincing indicator of price from our data was the type of view the apartment had.

### VIEW TYPE

Upon separation of the data points by “view,” we have three new Apartment Size vs. Rental Price regression lines for apartments with city, mountain, and ocean view. Apartments with city views had the highest Apartment Size vs. Rental Price correlation with an  $R^2$  at 61.4%. One interpretation for this is that cities tend to be more compacted, and thus land in the city is more valuable. Conversely, regions with mountain views tend to be less compacted, and thus the low  $R^2$  value at 3.4%. A similar explanation can account for the mountain view apartments having a lower slope than city apartments (0.645

and 1.62 respectively). It could be the case that the regression lines would be different if more data points were collected, and the low slope for mountain view could just be from the high leverage point at (761 sq ft, \$5,300). The regression line for apartments with ocean views, having the greatest slope of the three at 3.97 and a reasonable  $R^2$  of 44.6%, may perhaps indicate an almost standalone value of having an ocean view, as having it may give reason to charge more per square foot. Overall, as one might expect, an ocean view tends to cost more than a mountain view, which costs more than a city view. The ocean view and mountain view apartments tend to cluster around the higher square-foot range and seem to account for most of the “spread” of price, whereas the city apartments tend to cluster around the lower square footages. View type therefore seems to have more effect on price than square footage, if not at least complicate the narrative.

### AMENITIES

Perhaps the most surprising of our results was the effect that amenities had on apartment price. One can see that the lower bounds for amenity vs. no amenity were roughly similar, indicating that one could supposedly find an apartment with many amenities cheaper than another without the same amenities. The upper bounds, however, were found to be on average, higher for presence of amenity than without. The most peculiar result was that apartments that did not allow pets were higher costing than those that allowed pets.

### EXPERIMENTAL ERRORS

With experimental error primarily in the data collection phase, the first potential oversight is the use of only one source for the apartment data. There may be websites other than [www.zillow.com](http://www.zillow.com) that have more apartments, gather listings from more reliable sources, have more accurate data on variables like square footage, etc. Another issue was that apartment data would express certain categorical variables in a non-binary fashion, and defining what got a “0” for no and what got a “1” for yes became ambiguous. For example, some apartment entries may have said “no parking” while others may have said “street parking,” which in both cases we assigned a “0”. Another potential pitfall is that apartment data was provided by the landowner, and thus the accuracy of that data can vary based on this.

## Conclusion

Due to the inflated cost of living and housing in Hawai'i, it is important to understand and carefully analyze what we are buying. In this study, we determined what type of correlation between Apartment Size and Rental Price exists, and found that although apartment size plays a role in price, there are other confounding factors like view and amenities that perhaps play a bigger role in rental pricing. A somewhat expected result was that apartments with ocean views cost the most, and city views cost the least. The most surprising result, however, was that certain amenities like parking, allowance of pets, and laundry unit only nominally affected apartment price, which implies from a practical standpoint that one can be more "frivolous" with amenities in looking for an apartment. From here on out, further data analysis can be done with housing data with more judicious checking of source reliability. Given enough data, from not just Honolulu but all parts of Hawai'i, the ideas presented here can be applied to creating a predictive model for housing price—a tool undoubtedly necessary during economic inflation.

## References

- [1] "10 Timeless Factors That Affect Property Price | Propertylogy." [Online]. Available: <http://www.propertylogy.com/knowledge/10-timeless-factors-that-affect-property-price/>. [Accessed: 02-Mar-2017].
- [2] "A Peek Inside our Newest Zestimate: the Rent Zestimate," *Zillow Research*, 10-Mar-2011. [Online]. Available: <https://www.zillow.com/research/a-peek-inside-our-newest-zestimate-the-rent-zestimate-1076/>. [Accessed: 02-Mar-2017].
- [3] J. L. Pagliari Jr. and J. R. Webb, "On Setting Apartment Rental Rates: A Regression-Based Approach," *Journal of Real Estate Research*, vol. 12, no. 1, p. 37, Jul. 1996.
- [4] S. A. Gabriel and F. E. Nothaft, "Rental Housing Markets and the Natural Vacancy Rate," *Real Estate Economics*, vol. 16, no. 4, pp. 419–429, Dec. 1988.
- [5] R. D. D. Veaux, P. F. Velleman, and D. E. Bock, *Stats: Data and Models*, 4 edition. Boston: Pearson, 2015.
- [6] "RANDOM.ORG—True Random Number Service." [Online]. Available: <https://www.random.org/>. [Accessed: 02-Mar-2017].
- [7] "What is a Verified Source listing?," *Zillow Help Center*. [Online]. Available: <http://zillow.zendesk.com/hc/en-us/articles/216368218-What-is-a-Verified-Source-listing->. [Accessed: 02-Mar-2017].