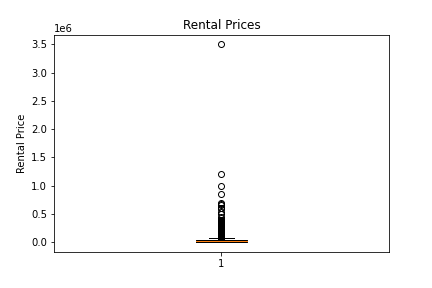
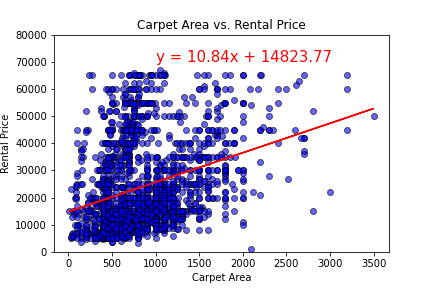
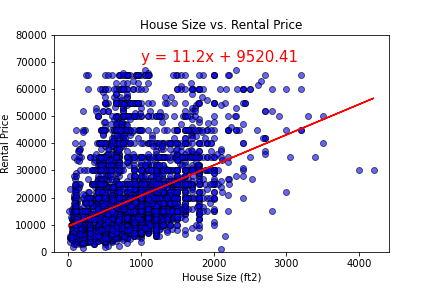
**Major Findings**

Our initial plots of the data all appeared very odd, with one group towering over the rest in all of our bar charts. So, to check for outliers, we first generated a boxplot, which is pictured below. As we had suspected, there were quite a few outliers, with one being far, far greater than every other value. So, we then made a new dataframe, containing only rental prices below the upper bound.

**Figure 1.** Rental Price boxplot.

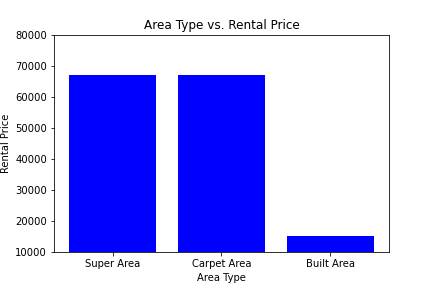
******Does the house size affect the rental price?**

**Figure 3.** Carpet Area vs. Rental Price.

**Figure 2.** House Size vs. Rental Price.

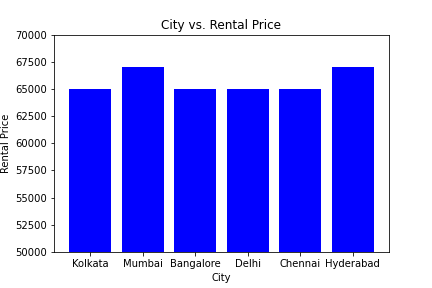
Yes, it appears that house size does indeed have some effect on rental price, though the correlation isn’t as strong as we had predicted. Figure 2 has an r-value of 0.394 and an r-squared value of 0.155. This r-value indicates a moderate positive correlation, though not a strong one. The r-squared value indicates that house size is responsible for approximately 15.5% of the variation in rental price, this value is lower than expected, but it is not totally unsurprising, as we would expect house size to be only one of several factors that influences rental prices.

Figure 3 is a similar graph, however house sizes measured via super area and built area (Area Type definitions explained in the Glossary) have been removed. We suspected that super area in particular, may have been skewing the data, as super area also includes common areas (lobbies, elevators, etc.) in its size measurements, meaning that these properties may appear very large based on house size measurements, but potentially not actually be very expensive. However, despite removing both super area and built area, the scatter plot and the correlation appears mostly unchanged, with an r-value of 0.325, and an r-squared value of 0.106.

**Does the area type affect the rental price?**

**Figure 4.** Area Type vs. Rental Price.

Yes, it appears that area type also affects rental price. A one-way ANOVA returned a P-value of 1.376e-93, a very low value, well below the threshold of statistical significance (P<0.05). To confirm that this wasn’t just due to the very low built area values (seen on the right in Figure 4), we then performed a T-test comparing super area to carpet area. This T-test also returned a P-value below 0.05, confirming statistical significance and allowing us to confidently reject the null hypothesis here.

**Does city/location affect the price?**

**Figure 5.** City vs. Rental Price.

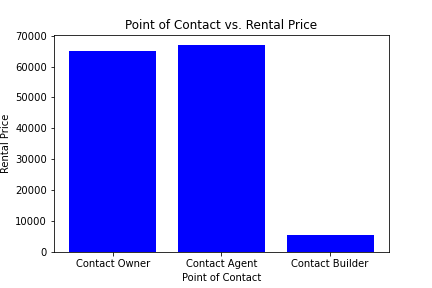
It also appears that city has an impact on rental prices. Similarly to above, we performed a one-way ANOVA which returned a P-value of 2.878e-309, an extremely low value. Based on this we can reject the null hypothesis here too. It appears here that Mumbai and Hyderabad are the most expensive cities for rental properties, but further testing is required to confirm this. Use of post-hoc analyses such as Tukey’s test or Bonferroni’s test would be useful in this regard.

**Icon

Description automatically generatedDoes the furnishing status affect the price?**

**Figure 6.** Furnishing Status vs. Rental Price.

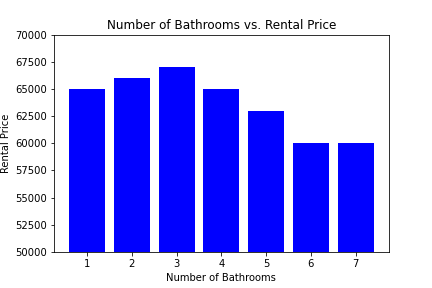
Again, the P-value obtained from a one-way ANOVA (P=2.049e-39) support the hypothesis that furnishing status affects rental price. Allowing us to reject the null hypothesis. Here, semi-furnished properties surprisingly appear to be the most expensive. However, as previously stated, further testing is needed to confirm or refute this.

**Does the point of contact affect the price?**

**Figure 7.** Point of Contact vs. Rental Price.

Point of contact seems to affect rental prices, as P=0**\***. As P<0.0001, we can confidently reject the null hypothesis here. It appears that rentals listed with their point of contact as “contact agent” tend to be more expensive, though further testing is needed.

**\*** This obviously can’t be an accurate P-value, but after much googling and discussing this odd P-value with the tutors, the most likely conclusion we found is that Python has a cut off for P-values, where P-values below a certain threshold will just appear as 0. To check that this wasn’t just being caused by the small Contact Builder column, we also did a T-test between Contact Owner and Contact Agent, which also returned P=0.

**Does the number of bathrooms affect the price?**

**Figure 8.** Number of Bathrooms vs. Rental Price.

As P=4.209e-277, there seems to be a statistically significant difference here too, confirming our hypothesis that number of bathrooms does impact rental price. However, the pattern observed is not as predicted, where we would expect that more bathrooms may be correlated with higher prices, it appears that three bathroom properties are, on average, the most expensive.

**Does the population affect prices?**

Population numbers for each city were obtained from the population API found on api-ninjas.com. A graph wasn’t made here as it would look identical to the City vs Rental Prices graph (Figure 5). A one-way ANOVA here allows us to reject the null hypothesis again, with a P-value of 2.878e-309.

In conclusion, it appears that all the variables examined impact on rental prices. House size was not as important as we had expected prior to this analysis, and it seems that rentals in Mumbai and Hyderabad tend to be the most expensive. Along with, surprisingly, semi-furnished and three-bathroom rentals. Though, further testing is required to confirm these findings.