Project 1
Stats 333
Shane McIntyre

Hello Vice President,

After conducting my analysis there are several factors that can influence the rent of the property size of your demand. The factors I found most influential are the following: Occupancy rate, time since last renovation, length of lease, distance to the airport, location, status of wiring, number of parking spots, the age of the building, distance to city center, number of floors, the size of the floor, distance to nearest hospital, the floor you're on, and if the lease is renewable or not. This is a total of 14 factors that can be influential to the rent. Each factor can either serve as a premium or a discount. The following factors will act as a discount: Time since last renovation if longer, length of lease if longer, location if further from city, age of the building, distance to city center if longer, the number of floors if more, the size of the floor, the distance to the nearest hospital if longer, and the floor you are on. The following factors will act as premiums: The occupancy rate, distance to the airport, the status of wiring, the number of parking spots, and if the lease is renewable or not. For example, if you want to rent out a space of 50,000 sqft and the lease you consider signing is 3 years long, within 4 years of renovation, a 80% occupancy rate, modern wiring, 10 parking spots, a new suburb, 10 miles from the airport, 20 year old building, distance to city center is 10 miles, 10 stories, 5 miles from a hospital, located on the second floor, and is renewable the rent per square foot is about $ 25.88, and thus the total rent is about $ 1,300,000. Depending on the discounts or premiums the range of the total rent could be around $1,000,000 - $ 1,500,000. What this tells us is if you emphasize the discounted factor and tried to find the leases that offer the best discounted terms, meaning an older building that's further from the city, etc. and limit the premiums such as lower occupancy rates, less parking, etc. you could find a lease with total cost closer to the $1,000,000 range, but if you get a bad deal it could cost you closer to the $1,500,000 range. If you are looking to be cheap on the initial lease deal, I would focus on finding an older building that is not recently renovated that has many stories and has larger floors. If you consider a longer lease this could also save you money. Location plays a very important role, the further from the city you are and further from a hospital can decrease the rent value. However, if the occupancy rate is higher and you are further from the airport this will increase your rent. Also, if it's a renewable lease with modern wiring and more parking spots the rent will increase. For example, from the expected price I already gave you from the earlier prediction of a total ret price of about $1,300,000, if we increase the length of the lease to 10 years it decreases the expected price to $1,150,000. With this, you can determine what factors are truly important to you, and what you wish to pursue in finding the best price for your needs.

Best,
Shane McIntyre

Technical Summary:

To figure out what model I should use to assess the rent I went with a two model method to compare the original full model to my reduced model. The full model I used was the response being Rent/SqftLease. This then allows you to determine variable and fixed factors, and for the fixed factors you would also divide that factor by SqftLease. Doing so standardizes everything to the same scale. Once all the factors were standardized to the response variable, I assessed the T value of each factor one at a time. Doing this, I did simple linear regression (to purposely asses the factor just to see how it affected the response all by itself) with each factor one by one to the response variable Rent/SqftLease. My null hypothesis was the factor didn't affect the response and my alternate hypothesis is that the given factor does affect the response. From this, if the factor assessed all by itself had a P-value of around 0.05, my null Hypothesis is rejected, and I marked that factor as important. The reason I assessed p-values around or less than 0.05 and not strictly less than 0.05 is because Rent is assessed by so many variables, that having some leniency is okay. After doing this for each of the 19 factors, any factor that I marked important I created a new multiple linear regression model. My new model accounted for 14 factors, compared to the full model of 19 factors. From this I conducted an F test to see if my reduced model assessed the response well. After doing the F test, I got a p-value of 2.939989e-05, which again rejects a null hypothesis of all the factors not affecting the response, meaning the factors are important. This shows my reduced model predicts the Rent/SqftLease well.

Variable Costs

- Rent
- SqftLease
- Wiring
- Occupancy
- SqftFloor
- Reno
- FloorsBldg
- FloorLease
- Firm Type
- Location
- Age
- Leaselength
- Renewable
- Restaurant
- Exercise

Fixed Costs

- DistCity
- DistAirp
- DriveAirp
- DistHosp
- Elevator

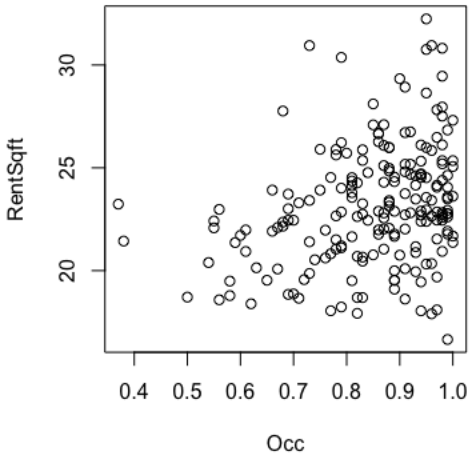
Parking

My overall model comes out to

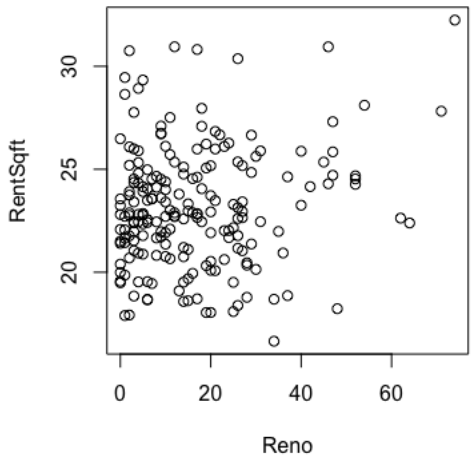
Y = ß0 + ß1X1 + ß2X2 + ß3X3+ ß4X4 + ß51X51 + ß52X52+ ß6X6 + ß7X7 + ß8X8 + ß9X9 + ß10X10 + ß11 X11+ ß12X12 + ß13X13 + ß14X14
Y = Rent/SqftLease
The ß's are the slope coefficients for its given factor, plus the ß of the intercept of the regression line. The ß tells us how much the response variable will change if that factor is changed by a unit of 1, assuming all other factors are held constant.
*Any values that were assessed qualitatively were translated numerically, if binary they are 0 and 1
*Plots show ranges of each factor
When assessing each factor, among all my factors there was a Response of over 70, I removed this Response (Response #148)

ß0 = Y intercept = 2.752e+01
Standard Error of Prediction = 2.188e+00

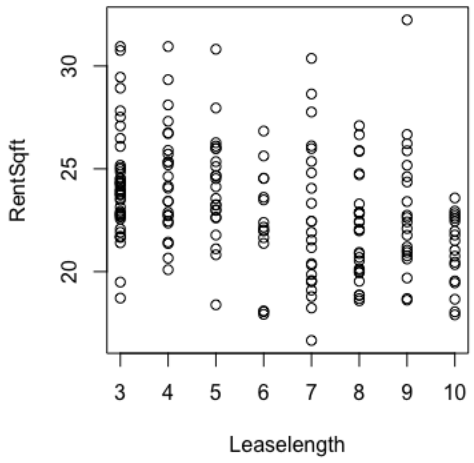
X1 = Value of Occupation
ß1 = Slope coefficient of Occupation = 1.676e+00
Standard Error of Prediction = 2.188e+00



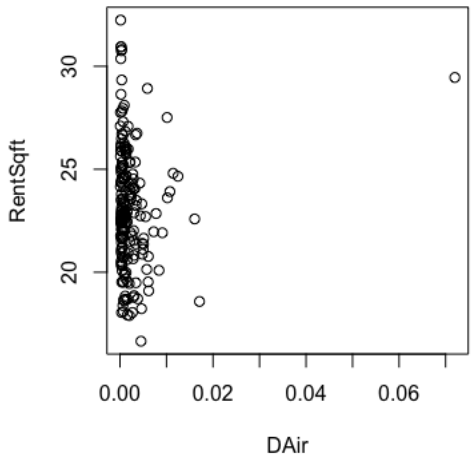
X2 = Value of log(Renovation + 1)
ß2 = Slope coefficient of Renovation = -2.679e-02
Standard Error of Prediction = 1.652e-01



X3 = Value of log(Leaselength + 2)
ß3 = Slope coefficient of Leaselength = -3.201e+00
Standard Error of Prediction = 4.653e-01

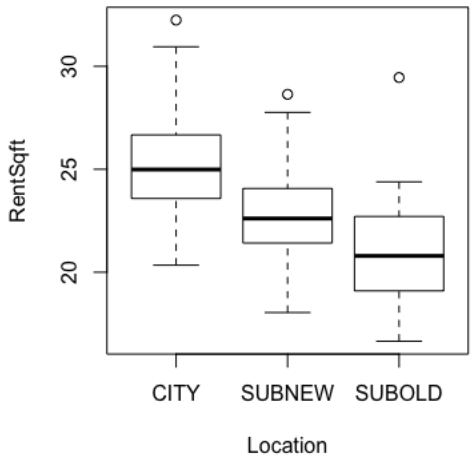


X4 = Value of log(Distance to Airport)
ß4 = Slope coefficient of Distance to Airport = -6.653e-02
Standard Error of Prediction = 4.008e-01



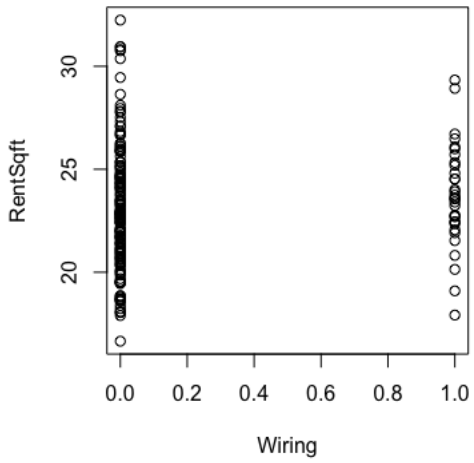
X51 = Value of Location (SubNew)
ß51 = Slope coefficient of Location (SubNew) = -2.428e+00
Standard Error of Prediction = 8.647e-01

X52 = Value of Location(SubOld)
ß52 = Slope coefficient of Location = -4.075e+00
Standard Error of Prediction = 7.397e-01

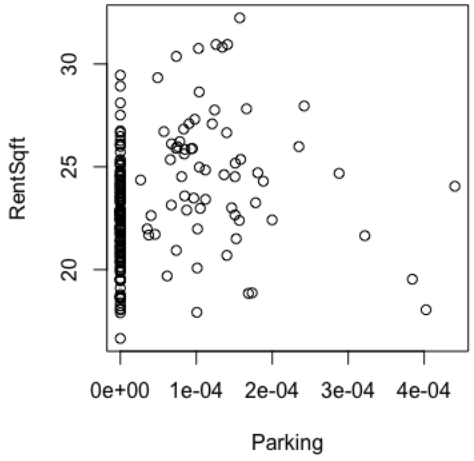


*I only assessed SUBNEW and SUBOLD since I already assessed drive time to city center

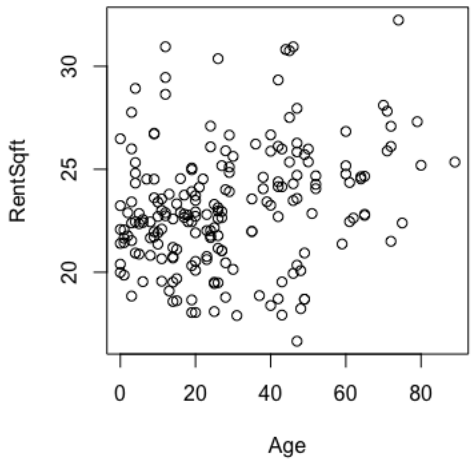
X6 = Value of Wiring
ß6 = Slope coefficient of Wiring = 4.914e-01
Standard Error of Prediction = 4.120e-01



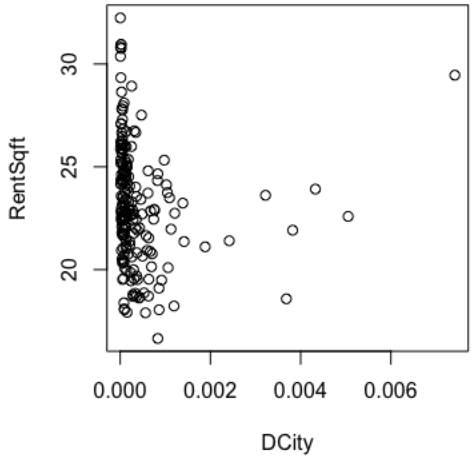
X7 = Value of Parking
ß7 = Slope coefficient of Parking = 5.641e+03
Standard Error of Prediction = 1.918e+03



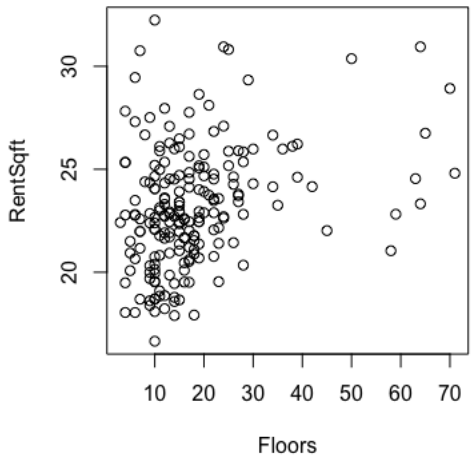
X8 = Value of Age
ß8 = Slope coefficient of Age = -3.219e-02
Standard Error of Prediction = 1.334e-02



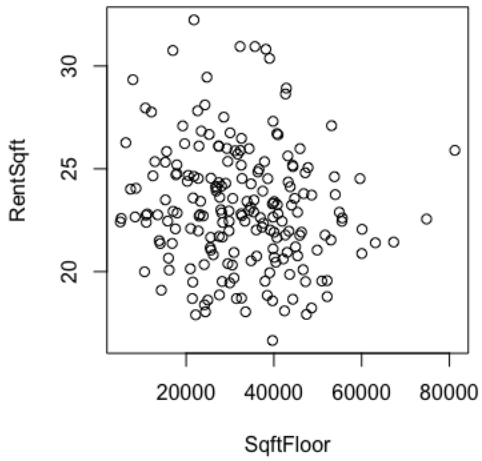
X9 = Value of log(Distance to City center)
ß9 = Slope coefficient of Distance to City center = -8.624e-01
Standard Error of Prediction = 3.097e-01



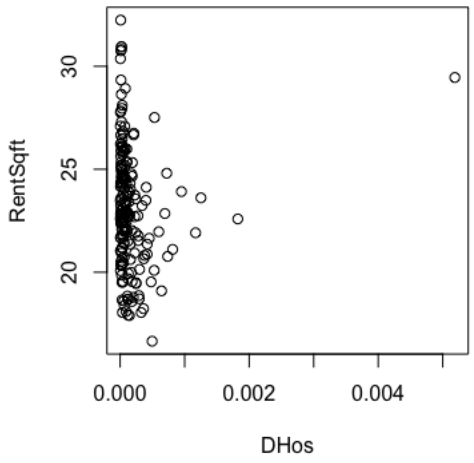
X10 = Value of Floors
ß10 = Slope coefficient of Floors = 1.464e-02
Standard Error of Prediction = 2.167e-02



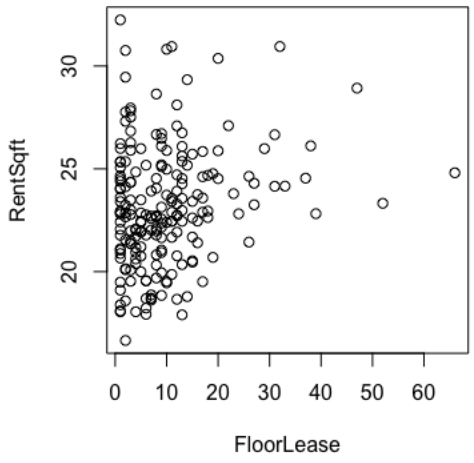
X11 = Value of SqftFloor
ß11 = Slope coefficient of SqftFloor = -1.052e-05
Standard Error of Prediction = 1.251e-05



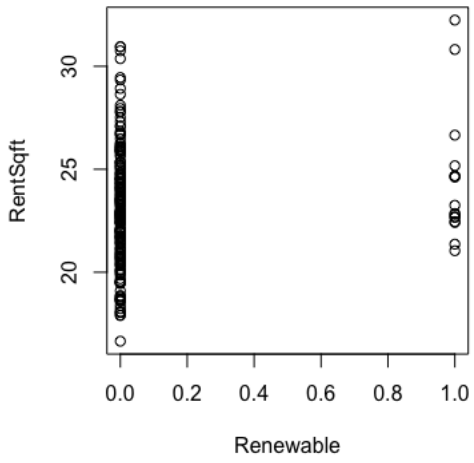
X12 = Value of log(Distance to Hospital)
ß12 = Slope coefficient of Distance to Hospital = 4.343e-01
Standard Error of Prediction = 2.346e-01



X13 = Value of FloorLease
ß13 = Slope coefficient of FloorLease = -3.119e-02
Standard Error of Prediction = 2.197e-02



X14 = Value of Renewable
ß14 = Slope coefficient of Renewable = 1.960e-01
Standard Error of Prediction = 5.511e-01



The factors I used were:

*Factors were either required or p-value close to 0.05 when assessed in my original t tests

- Occupation
    - Required
- Renovation
    - Required
    - Log was used to show for diminishing return, +1 to account for values of 0
- Leaselength
    - Required
    - Log was used to show for diminishing return, +2 to account for values of 0
- Distance to Airport
    - Required
    - Log was used to transform to decrease the p-value when assessed as the only factor in my original t tests
- Location
    - Required
- Wiring
    - Required
- Parking
    - Required
- Age
- Distance to City Center
    - Log was used to transform to decrease the p-value when assessed as the only factor in my original t tests
- Number of Floors
- Sqft of Floors
- Distance to hospital
    - Log was used to transform to decrease the p-value when assessed as the only factor in my original t tests

- Floor lease was on
- Renewable

The factors I left out in my model were:

- Drive Time to airport
    - I didn't use drive time since I was already accounting for the distance to the airport. So, an overall assessment of how far the airport is from the building is already assessed for
- Elevator
    - This proved to not have a significant impact on rent
- Restaurant
    - This proved to not have a significant impact on rent
- Exercise
    - This proved to not have a significant impact on rent
- Firmtype
    - This proved to not have a significant impact on rent

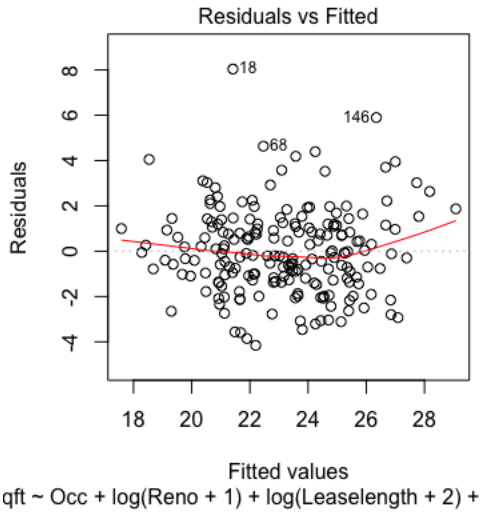
Overall:

My final regression equation comes out to:

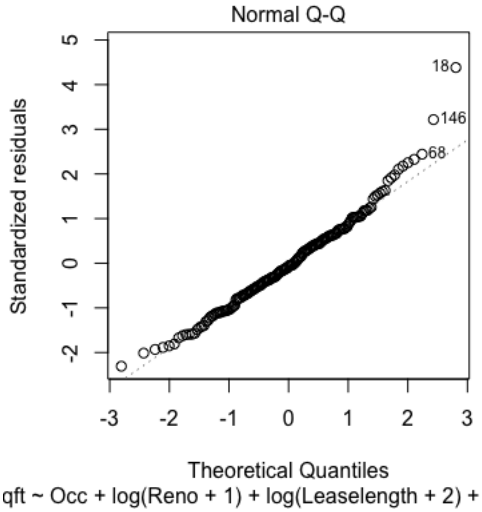
Y = 2.752e+01 +1.676e+00*X1 - 2.679e-02*X2 - 3.201e+00*X3 - 6.653e-02*X4 - 2.428e+00*X51 - 4.075e+00*X52+4.914e-01*X6 + 5.641e+03*X7 - 3.219e-02*X8 - 8.624e-01*X9 + 1.464e-02*X10 - 1.052e-05*X11+ 4.343e-01*X12 - 3.119e-02*X13 + 1.960e-01*X14

With F value of 4.721269 with a p-value of 2.939989e-05. This shows the factors are important in assessing the response (Null hypothesis is all factors don't affect response, this is rejected).

To test for linearity, normality, and homoskedasticity we look at the residual plots and qq plots:



From this, I would say linearity and homoscedasticity is not broken. The red line does bend up at the end, but its still less than the residuals of value 2, and given that rent is determined by many factors, some leniency is okay. The slight bend at the end suggests the landlord can try to rip off the tenant and give them a higher price than they should.



qft ~ Occ + log(Reno + 1) + log(Leaselength + 2) + log(

From the QQ plot, I'd say normality is not broken, almost all the dots are on the line, with just 2 out of 200 points deviating far from the line, which 2 out of 200 isn't enough to create a tail.