Airbnb_w_factors

Idaly Ferrales

2023-07-15

```
#Adding libraries and loading file
getwd()
```

```
## [1] "/Users/idalyferrales/OMSA folder"
```

```
setwd("/Users/idalyferrales/OMSA folder")
getwd()
```

```
## [1] "/Users/idalyferrales/OMSA folder"
```

```
#Import the Airbnb dataset
list.files(path=".", pattern=".csv", all.files=TRUE,
           full.names=TRUE)
```

```
##  [1] "./Airbnb_clean_dataset.csv"
##  [2] "./Airbnb_Data_USA.csv"
##  [3] "./Airbnb_dataset_eda.csv"
##  [4] "./Airbnb_with_factors.csv"
##  [5] "./airbnb_with_fips.csv"
##  [6] "./Berkshire.csv"
##  [7] "./binary_1.csv"
##  [8] "./binary_midterm.csv.numbers"
##  [9] "./binary.csv"
## [10] "./contrafund_.csv"
## [11] "./contrafund_Exam.csv"
## [12] "./contrafund_final.csv"
## [13] "./contrafund.csv"
## [14] "./Factor_HiTech_Midterm.csv"
## [15] "./factors.csv"
## [16] "./final_data.csv"
## [17] "./github.gatech.edu_raw_MGT-6203-Summer-2023-Canvas_Team-71_main_Brainstorm_walkability_impact_walkabili
ty_index.csv_token=GHSAT0AAAAAAAACWL4HSO7TEOBS7D4WNS5IZFJ3MCQ.txt"
## [18] "./Grades_Data.csv"
## [19] "./sample_airbnb_data_filtered_usa.csv"
## [20] "./UPS_KO_Exam.csv"
```

```
data <- read.csv("./Airbnb_with_factors.csv")
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

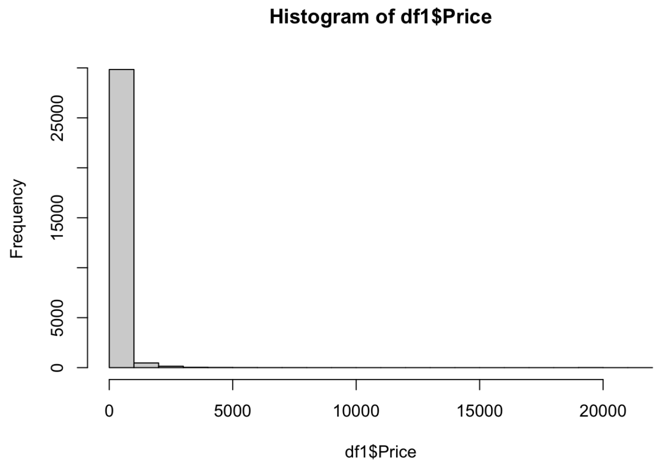
```
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

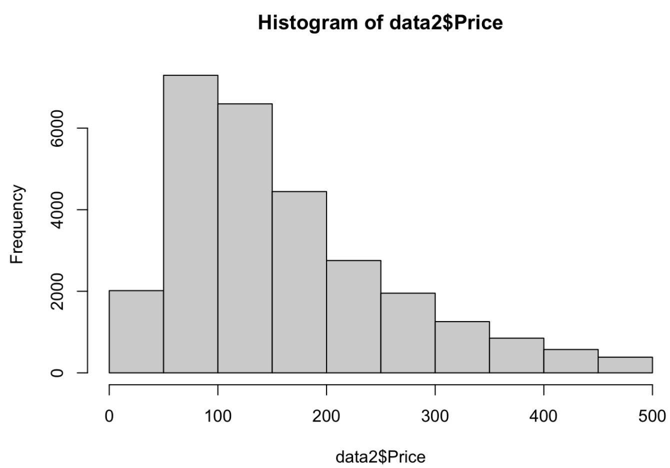
```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
# Converting our data file to a dataframe
df1 <- data.frame(data)
```

```
#Histogram of price with no data manipulation
hist(df1$Price)
```




```
#Storing our dataset under name 'data2' for price less than zero
data2 = data[data$Price < 500,]
hist(data2$Price)
```




```
#Manipulation of data on variables number of reviews with zero values
data2$number_of_reviews_ltm[data2$number_of_reviews_ltm == 0] <- 1
head(data2)
```

```
##        city Price availability_365 minimum_nights number_of_reviews_ltm
## 1 Asheville    79               29             30                     1
## 3 Asheville    79              362              2                     4
## 4 Asheville   196              158              1                    19
## 5 Asheville    72              168              1                    43
## 6 Asheville   150              204             28                     6
## 7 Asheville   149              362              1                    19
##         Room_type calculated_host_listings_count walk_index Avg.rent
## 1 Entire home/apt                              1  14.333333      947
## 3    Private room                              1  14.333330      947
## 4 Entire home/apt                              1  14.333333      947
## 5 Entire home/apt                             13   8.666667      947
## 6 Entire home/apt                              5  17.666667      947
## 7 Entire home/apt                             13   8.666667      947
##   Avg.hotel.price Avg.Home.Value Avg.Salary
## 1             122         306450      49930
## 3             122         306450      49930
## 4             122         306450      49930
## 5             122         306450      49930
## 6             122         306450      49930
## 7             122         306450      49930
```

```
set.seed(379) #setting seed
ind <- sample(2, nrow(data2), replace = TRUE, prob = c(0.7, 0.3)) #splitting data 70/30 for our linear regression
model
train2 <- data2[ind==1,] #This is our train set
test2 <- data2[ind==2,] #This is our test set
```

```
#Started with log linear model with availability_365, walkability index, number of review last 12 months, room ty
pe, city, minimum nights and calculated listings count
log_lin_all_except_avg = lm(log(Price) ~ availability_365+ walk_index + number_of_reviews_ltm + Room_type + city
+ minimum_nights + calculated_host_listings_count, data=train2)
summary(log_lin_all_except_avg)
```

```
##
## Call:
## lm(formula = log(Price) ~ availability_365 + walk_index + number_of_reviews_ltm +
##     Room_type + city + minimum_nights + calculated_host_listings_count,
##     data = train2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4850 -0.3549 -0.0366  0.3421  3.8000
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    4.919e+00  3.203e-02 153.581  < 2e-16 ***
## availability_365               2.789e-04  3.131e-05   8.907  < 2e-16 ***
## walk_index                     3.105e-03  1.253e-03   2.478  0.01324 *
## number_of_reviews_ltm         -1.718e-03  1.497e-04 -11.478  < 2e-16 ***
## Room_typeHotel room            9.837e-02  6.248e-02   1.575  0.11536
## Room_typePrivate room         -7.343e-01  8.521e-03 -86.173  < 2e-16 ***
## Room_typeShared room          -1.430e+00  3.812e-02 -37.517  < 2e-16 ***
## cityAustin                     2.328e-01  3.334e-02   6.982 2.99e-12 ***
## cityBoston                     1.889e-01  3.941e-02   4.793 1.66e-06 ***
## cityCambridge                  3.988e-01  6.204e-02   6.429 1.32e-10 ***
## cityChicago                   -6.186e-02  3.210e-02  -1.927  0.05399 .
## cityClark County               2.182e-01  3.022e-02   7.221 5.36e-13 ***
## cityColumbus                  -1.980e-01  4.215e-02  -4.697 2.65e-06 ***
## cityDenver                    -6.727e-02  3.920e-02  -1.716  0.08610 .
## cityFort Lauderdale            2.793e-01  3.055e-02   9.142  < 2e-16 ***
## cityJersey City               -4.687e-03  5.501e-02  -0.085  0.93211
## cityLos Angeles                1.836e-01  2.876e-02   6.383 1.77e-10 ***
## cityMonterrey                  6.197e-01  1.301e-01   4.763 1.92e-06 ***
## cityNashville                  3.007e-01  3.292e-02   9.133  < 2e-16 ***
## cityNew Orleans                2.284e-01  3.257e-02   7.014 2.38e-12 ***
## cityNew York City              1.811e-01  2.870e-02   6.309 2.87e-10 ***
## cityOakland                   -3.996e-02  4.052e-02  -0.986  0.32403
## cityPortland                  -1.973e-01  3.669e-02  -5.377 7.65e-08 ***
## cityRhode Island               3.945e-01  3.471e-02  11.367  < 2e-16 ***
## citySalem                     -7.673e-02  8.794e-02  -0.873  0.38290
## citySan Diego                  2.562e-01  3.192e-02   8.027 1.05e-15 ***
## citySan Francisco              2.762e-01  3.530e-02   7.824 5.37e-15 ***
## citySan Mateo County           2.551e-01  4.771e-02   5.348 8.99e-08 ***
## citySanta Clara County         6.505e-02  3.431e-02   1.896  0.05796 .
## citySanta Cruz County          3.892e-01  6.546e-02   5.946 2.79e-09 ***
## citySeattle                   -1.414e-03  3.617e-02  -0.039  0.96881
## cityTwin Cities MSA           -7.767e-02  3.630e-02  -2.140  0.03238 *
## cityWashington D.C.            1.024e-01  3.539e-02   2.893  0.00382 **
## minimum_nights                -4.010e-03  1.660e-04 -24.154  < 2e-16 ***
## calculated_host_listings_count  1.973e-04  3.469e-05   5.689 1.29e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5081 on 19797 degrees of freedom
## Multiple R-squared:  0.3541, Adjusted R-squared:  0.353
## F-statistic: 319.3 on 34 and 19797 DF,  p-value: < 2.2e-16
```

```
#log-log model of certain variables
regression_log_log_subset = lm(log(Price) ~ log(availability_365)+ log(walk_index) + log(number_of_reviews_ltm) ,
data=train2)
summary(regression_log_log_subset)
```

```
##
## Call:
## lm(formula = log(Price) ~ log(availability_365) + log(walk_index) +
##     log(number_of_reviews_ltm), data = train2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.62723 -0.41634  0.01218  0.45832  1.41962
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 4.841873   0.050297  96.266  < 2e-16 ***
## log(availability_365)       0.026088   0.004546   5.738 9.69e-09 ***
## log(walk_index)            -0.036688   0.016755  -2.190   0.0286 *
## log(number_of_reviews_ltm)  0.017871   0.002992   5.973 2.37e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6307 on 19828 degrees of freedom
## Multiple R-squared:  0.00338,    Adjusted R-squared:  0.00323
## F-statistic: 22.42 on 3 and 19828 DF,  p-value: 1.746e-14
```

```
#Linear linear model of availability, walkability index, number of reviews, room type and city
lin_lin = lm(Price ~ availability_365+ walk_index + number_of_reviews_ltm + Room_type + city, data=train2)
summary(lin_lin)
```

```
##
## Call:
## lm(formula = Price ~ availability_365 + walk_index + number_of_reviews_ltm +
##     Room_type + city, data = train2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -222.39  -58.98  -21.10   37.77  406.66
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.608e+02  5.606e+00  28.686  < 2e-16 ***
## availability_365        3.332e-02  5.413e-03   6.157 7.58e-10 ***
## walk_index              5.600e-02  2.194e-01   0.255 0.798571
## number_of_reviews_ltm  -2.537e-01  2.558e-02  -9.921  < 2e-16 ***
## Room_typeHotel room     2.813e+01  1.094e+01   2.571 0.010144 *
## Room_typePrivate room  -9.109e+01  1.488e+00 -61.223  < 2e-16 ***
## Room_typeShared room   -1.405e+02  6.677e+00 -21.043  < 2e-16 ***
## cityAustin              3.988e+01  5.840e+00   6.829 8.80e-12 ***
## cityBoston              2.015e+01  6.885e+00   2.927 0.003427 **
## cityCambridge           4.479e+01  1.085e+01   4.126 3.70e-05 ***
## cityChicago            -5.943e+00  5.618e+00  -1.058 0.290135
## cityClark County        3.787e+01  5.293e+00   7.154 8.73e-13 ***
## cityColumbus           -2.495e+01  7.383e+00  -3.379 0.000729 ***
## cityDenver             -1.232e+01  6.863e+00  -1.795 0.072710 .
## cityFort Lauderdale     5.001e+01  5.350e+00   9.349  < 2e-16 ***
## cityJersey City        -7.595e+00  9.623e+00  -0.789 0.429995
## cityLos Angeles         2.384e+01  5.026e+00   4.744 2.12e-06 ***
## cityMonterrey           1.028e+02  2.279e+01   4.511 6.50e-06 ***
## cityNashville           5.152e+01  5.766e+00   8.935  < 2e-16 ***
## cityNew Orleans         3.493e+01  5.704e+00   6.124 9.28e-10 ***
## cityNew York City       2.148e+01  5.024e+00   4.277 1.90e-05 ***
## cityOakland            -1.001e+01  7.097e+00  -1.411 0.158231
## cityPortland           -2.985e+01  6.427e+00  -4.645 3.43e-06 ***
## cityRhode Island        6.786e+01  6.079e+00  11.162  < 2e-16 ***
## citySalem              -2.071e+01  1.540e+01  -1.345 0.178740
## citySan Diego           4.211e+01  5.591e+00   7.532 5.19e-14 ***
## citySan Francisco       3.015e+01  6.171e+00   4.886 1.04e-06 ***
## citySan Mateo County    3.590e+01  8.356e+00   4.297 1.74e-05 ***
## citySanta Clara County  9.486e+00  6.007e+00   1.579 0.114321
## citySanta Cruz County   7.431e+01  1.147e+01   6.480 9.36e-11 ***
## citySeattle            -2.706e+00  6.336e+00  -0.427 0.669347
## cityTwin Cities MSA    -5.716e+00  6.358e+00  -0.899 0.368660
## cityWashington D.C.     9.837e+00  6.194e+00   1.588 0.112258
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89 on 19799 degrees of freedom
## Multiple R-squared:  0.2198, Adjusted R-squared:  0.2185
## F-statistic: 174.3 on 32 and 19799 DF,  p-value: < 2.2e-16
```

```r
#Interacting with variables using a log_log model, however not every variable was transformed, only numeric
regression1 = lm(log(Price) ~ log(availability_365)+ log(walk_index) + number_of_reviews_ltm + Room_type + city,
data=train2)
summary(regression1)
```

```
##
## Call:
## lm(formula = log(Price) ~ log(availability_365) + log(walk_index) +
##     number_of_reviews_ltm + Room_type + city, data = train2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.55925 -0.36235 -0.03091  0.35148  2.30080
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             4.8333048  0.0485975  99.456  < 2e-16 ***
## log(availability_365)   0.0245396  0.0037391   6.563 5.40e-11 ***
## log(walk_index)         0.0060483  0.0145867   0.415 0.678404
## number_of_reviews_ltm  -0.0010547  0.0001479  -7.133 1.02e-12 ***
## Room_typeHotel room     0.1275705  0.0634125   2.012 0.044259 *
## Room_typePrivate room  -0.7267699  0.0086183 -84.329  < 2e-16 ***
## Room_typeShared room   -1.4236301  0.0386947 -36.791  < 2e-16 ***
## cityAustin              0.2368619  0.0339112   6.985 2.94e-12 ***
## cityBoston              0.1335477  0.0399757   3.341 0.000837 ***
## cityCambridge           0.3361171  0.0629579   5.339 9.46e-08 ***
## cityChicago            -0.0593989  0.0326422  -1.820 0.068820 .
## cityClark County        0.2207637  0.0307460   7.180 7.21e-13 ***
## cityColumbus           -0.1858439  0.0428487  -4.337 1.45e-05 ***
## cityDenver             -0.0858762  0.0398645  -2.154 0.031236 *
## cityFort Lauderdale     0.3005270  0.0310542   9.677  < 2e-16 ***
## cityJersey City        -0.0708988  0.0558336  -1.270 0.204163
## cityLos Angeles         0.1641115  0.0292092   5.618 1.95e-08 ***
## cityMonterrey           0.5877617  0.1320791   4.450 8.63e-06 ***
## cityNashville           0.3107172  0.0334445   9.291  < 2e-16 ***
## cityNew Orleans         0.2279453  0.0331526   6.876 6.35e-12 ***
## cityNew York City       0.1621248  0.0292148   5.549 2.90e-08 ***
## cityOakland            -0.0346680  0.0411959  -0.842 0.400055
## cityPortland           -0.1902655  0.0372864  -5.103 3.38e-07 ***
## cityRhode Island        0.4036004  0.0352505  11.449  < 2e-16 ***
## citySalem              -0.0583245  0.0893035  -0.653 0.513697
## citySan Diego           0.2622007  0.0324772   8.073 7.23e-16 ***
## citySan Francisco       0.2314608  0.0358352   6.459 1.08e-10 ***
## citySan Mateo County    0.2413859  0.0484118   4.986 6.21e-07 ***
## citySanta Clara County  0.0654409  0.0348778   1.876 0.060630 .
## citySanta Cruz County   0.3843608  0.0665023   5.780 7.60e-09 ***
## citySeattle            -0.0047167  0.0367878  -0.128 0.897981
## cityTwin Cities MSA    -0.0706548  0.0368757  -1.916 0.055376 .
## cityWashington D.C.     0.0736865  0.0359921   2.047 0.040642 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5159 on 19799 degrees of freedom
## Multiple R-squared:  0.3341, Adjusted R-squared:  0.3331
## F-statistic: 310.5 on 32 and 19799 DF,  p-value: < 2.2e-16
```

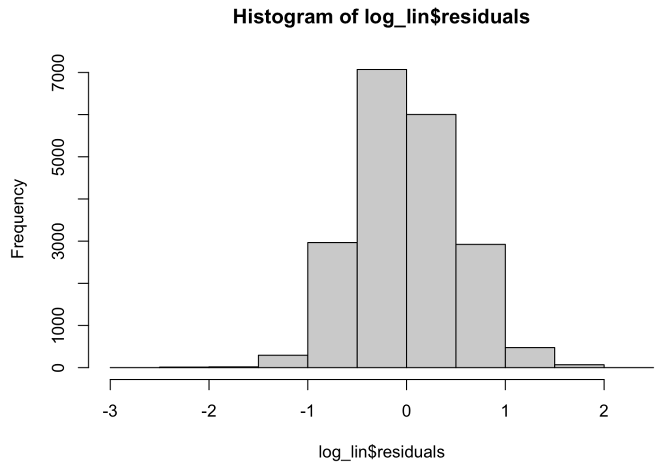
```
#Log linear model with availability, walkability index, number of reviews, room type and city
log_lin = lm(log(Price) ~ availability_365+ walk_index + number_of_reviews_ltm + Room_type + city, data=train2)
summary(log_lin)
```

```
##
## Call:
## lm(formula = log(Price) ~ availability_365 + walk_index + number_of_reviews_ltm +
##     Room_type + city, data = train2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.57045 -0.36082 -0.03155  0.35031  2.32072
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             4.891e+00  3.248e-02 150.598  < 2e-16 ***
## availability_365        2.362e-04  3.136e-05   7.530 5.27e-14 ***
## walk_index              2.601e-03  1.271e-03   2.046  0.04076 *
## number_of_reviews_ltm  -1.014e-03  1.482e-04  -6.840 8.13e-12 ***
## Room_typeHotel room     1.206e-01  6.340e-02   1.902  0.05719 .
## Room_typePrivate room  -7.277e-01  8.621e-03 -84.416  < 2e-16 ***
## Room_typeShared room   -1.427e+00  3.868e-02 -36.901  < 2e-16 ***
## cityAustin              2.323e-01  3.384e-02   6.865 6.82e-12 ***
## cityBoston              1.258e-01  3.989e-02   3.155  0.00161 **
## cityCambridge           3.305e-01  6.288e-02   5.255 1.49e-07 ***
## cityChicago            -6.810e-02  3.255e-02  -2.092  0.03642 *
## cityClark County        2.190e-01  3.067e-02   7.139 9.69e-13 ***
## cityColumbus           -1.916e-01  4.278e-02  -4.479 7.54e-06 ***
## cityDenver             -9.036e-02  3.977e-02  -2.272  0.02308 *
## cityFort Lauderdale     2.957e-01  3.100e-02   9.540  < 2e-16 ***
## cityJersey City        -7.485e-02  5.575e-02  -1.343  0.17942
## cityLos Angeles         1.568e-01  2.912e-02   5.383 7.43e-08 ***
## cityMonterrey           5.825e-01  1.320e-01   4.413 1.03e-05 ***
## cityNashville           3.074e-01  3.341e-02   9.202  < 2e-16 ***
## cityNew Orleans         2.227e-01  3.305e-02   6.737 1.66e-11 ***
## cityNew York City       1.566e-01  2.911e-02   5.380 7.52e-08 ***
## cityOakland            -4.011e-02  4.112e-02  -0.976  0.32931
## cityPortland           -1.985e-01  3.724e-02  -5.331 9.87e-08 ***
## cityRhode Island        4.032e-01  3.522e-02  11.448  < 2e-16 ***
## citySalem              -6.399e-02  8.924e-02  -0.717  0.47332
## citySan Diego           2.572e-01  3.239e-02   7.940 2.13e-15 ***
## citySan Francisco       2.235e-01  3.575e-02   6.250 4.19e-10 ***
## citySan Mateo County    2.466e-01  4.841e-02   5.094 3.54e-07 ***
## citySanta Clara County  6.111e-02  3.480e-02   1.756  0.07913 .
## citySanta Cruz County   3.890e-01  6.644e-02   5.855 4.85e-09 ***
## citySeattle            -1.160e-02  3.671e-02  -0.316  0.75204
## cityTwin Cities MSA    -7.193e-02  3.684e-02  -1.953  0.05086 .
## cityWashington D.C.     6.742e-02  3.589e-02   1.879  0.06031 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5157 on 19799 degrees of freedom
## Multiple R-squared:  0.3347, Adjusted R-squared:  0.3337
## F-statistic: 311.3 on 32 and 19799 DF,  p-value: < 2.2e-16
```

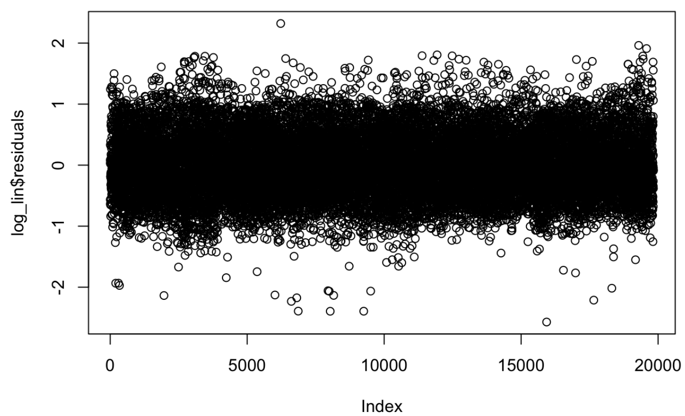
```r
#log linear with all variables
log_lin_all = lm(log(Price) ~ ., data=train2)
summary(log_lin_all)
```

```
##
## Call:
## lm(formula = log(Price) ~ ., data = train2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4850 -0.3549 -0.0366  0.3421  3.8000
##
## Coefficients: (4 not defined because of singularities)
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    4.919e+00  3.203e-02 153.581  < 2e-16 ***
## cityAustin                     2.328e-01  3.334e-02   6.982 2.99e-12 ***
## cityBoston                     1.889e-01  3.941e-02   4.793 1.66e-06 ***
## cityCambridge                  3.988e-01  6.204e-02   6.429 1.32e-10 ***
## cityChicago                   -6.186e-02  3.210e-02  -1.927  0.05399 .
## cityClark County               2.182e-01  3.022e-02   7.221 5.36e-13 ***
## cityColumbus                  -1.980e-01  4.215e-02  -4.697 2.65e-06 ***
## cityDenver                    -6.727e-02  3.920e-02  -1.716  0.08610 .
## cityFort Lauderdale            2.793e-01  3.055e-02   9.142  < 2e-16 ***
## cityJersey City               -4.687e-03  5.501e-02  -0.085  0.93211
## cityLos Angeles                1.836e-01  2.876e-02   6.383 1.77e-10 ***
## cityMonterrey                  6.197e-01  1.301e-01   4.763 1.92e-06 ***
## cityNashville                  3.007e-01  3.292e-02   9.133  < 2e-16 ***
## cityNew Orleans                2.284e-01  3.257e-02   7.014 2.38e-12 ***
## cityNew York City              1.811e-01  2.870e-02   6.309 2.87e-10 ***
## cityOakland                   -3.996e-02  4.052e-02  -0.986  0.32403
## cityPortland                  -1.973e-01  3.669e-02  -5.377 7.65e-08 ***
## cityRhode Island               3.945e-01  3.471e-02  11.367  < 2e-16 ***
## citySalem                     -7.673e-02  8.794e-02  -0.873  0.38290
## citySan Diego                  2.562e-01  3.192e-02   8.027 1.05e-15 ***
## citySan Francisco              2.762e-01  3.530e-02   7.824 5.37e-15 ***
## citySan Mateo County           2.551e-01  4.771e-02   5.348 8.99e-08 ***
## citySanta Clara County         6.505e-02  3.431e-02   1.896  0.05796 .
## citySanta Cruz County          3.892e-01  6.546e-02   5.946 2.79e-09 ***
## citySeattle                   -1.414e-03  3.617e-02  -0.039  0.96881
## cityTwin Cities MSA           -7.767e-02  3.630e-02  -2.140  0.03238 *
## cityWashington D.C.            1.024e-01  3.539e-02   2.893  0.00382 **
## availability_365               2.789e-04  3.131e-05   8.907  < 2e-16 ***
## minimum_nights                -4.010e-03  1.660e-04 -24.154  < 2e-16 ***
## number_of_reviews_ltm         -1.718e-03  1.497e-04 -11.478  < 2e-16 ***
## Room_typeHotel room            9.837e-02  6.248e-02   1.575  0.11536
## Room_typePrivate room         -7.343e-01  8.521e-03 -86.173  < 2e-16 ***
## Room_typeShared room          -1.430e+00  3.812e-02 -37.517  < 2e-16 ***
## calculated_host_listings_count  1.973e-04  3.469e-05   5.689 1.29e-08 ***
## walk_index                     3.105e-03  1.253e-03   2.478  0.01324 *
## Avg.rent                              NA         NA      NA       NA
## Avg.hotel.price                       NA         NA      NA       NA
## Avg.Home.Value                        NA         NA      NA       NA
## Avg.Salary                            NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5081 on 19797 degrees of freedom
## Multiple R-squared:  0.3541, Adjusted R-squared:  0.353
## F-statistic: 319.3 on 34 and 19797 DF,  p-value: < 2.2e-16
```

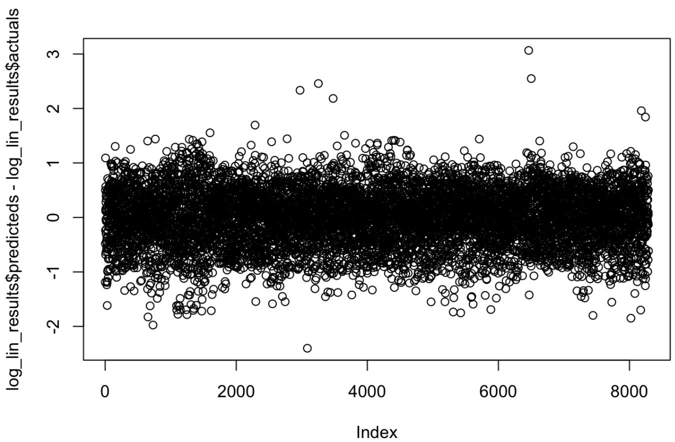
```
##log-linear model, plot training errors
hist(log_lin$residuals, breaks = 10)
```




```
plot(log_lin$residuals)
```




```
#get test errors and plot
log_lin_pred <- predict(log_lin, test2)
log_lin_results <- data.frame(cbind(actuals=log(test2$Price), predicteds=log_lin_pred))
plot(log_lin_results$predicteds-log_lin_results$actuals)
```




```
rmse <- sqrt(sum((exp(log_lin_results) - test2$Price)^2)/length(test2$Price))
```

```
#The RMSE of our model
rmse
```

```
## [1] 92.28303
```