Airbnb Price Regression Modeling

Barbara Flores

```r
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
```

```
Rows: 3,239
Columns: 75
$ id                        <dbl> 108061, 4394761, 4448604,~
$ listing_url               <chr> "https://www.airbnb.com/r~
$ scrape_id                 <dbl> 2.023062e+13, 2.023062e+1~
$ last_scraped              <date> 2023-06-18, 2023-06-18, ~
$ source                    <chr> "city scrape", "city scra~
$ name                      <chr> "Rental unit in Asheville~
$ description               <chr> "Walk to town in ten minu~
$ neighborhood_overview     <chr> "I love my neighborhood! ~
$ picture_url               <chr> "https://a0.muscache.com/~
$ host_id                   <dbl> 320564, 20447538, 5656896~
$ host_url                  <chr> "https://www.airbnb.com/u~
$ host_name                 <chr> "Lisa", "Brian", "Michael~
$ host_since                <date> 2010-12-16, 2014-08-23, ~
$ host_location             <chr> "Asheville, NC", "Ashevil~
$ host_about                <chr> "I am a long time residen~
$ host_response_time        <chr> "within an hour", "within~
$ host_response_rate        <chr> "100%", "100%", "100%", "~
$ host_acceptance_rate      <chr> "82%", "100%", "99%", "10~
$ host_is_superhost         <lgl> TRUE, TRUE, TRUE, FALSE, ~
$ host_thumbnail_url        <chr> "https://a0.muscache.com/~
$ host_picture_url          <chr> "https://a0.muscache.com/~
$ host_neighbourhood        <chr> NA, NA, NA, NA, NA, NA, N~
$ host_listings_count       <dbl> 2, 2, 3, 2, 1, 2, 4, 2, 2~
$ host_total_listings_count <dbl> 3, 2, 3, 4, 3, 5, 7, 2, 2~
$ host_verifications        <chr> "['email', 'phone']", "['~
$ host_has_profile_pic      <lgl> TRUE, TRUE, TRUE, TRUE, T~
$ host_identity_verified    <lgl> TRUE, TRUE, TRUE, TRUE, T~
```

```
$ neighbourhood                 <chr> "Asheville, North Carolin~
$ neighbourhood_cleansed        <dbl> 28801, 28801, 28804, 2880~
$ neighbourhood_group_cleansed  <lgl> NA, NA, NA, NA, NA, NA, N~
$ latitude                      <dbl> 35.60670, 35.61244, 35.62~
$ longitude                     <dbl> -82.55563, -82.55724, -82~
$ property_type                 <chr> "Entire rental unit", "En~
$ room_type                     <chr> "Entire home/apt", "Entir~
$ accommodates                  <dbl> 2, 3, 4, 2, 2, 2, 2, 8, 4~
$ bathrooms                     <lgl> NA, NA, NA, NA, NA, NA, N~
$ bathrooms_text                <chr> "1 bath", "1 bath", "1 ba~
$ bedrooms                      <dbl> 1, 2, 1, 2, 1, 1, NA, 3, ~
$ beds                          <dbl> 1, 1, 2, 2, 1, 1, 1, 6, 2~
$ amenities                     <chr> "[\"Induction stove\", \"~
$ price                         <chr> "$100.00", "$114.00", "$4~
$ minimum_nights                <dbl> 30, 1, 1, 30, 90, 30, 30,~
$ maximum_nights                <dbl> 365, 1125, 60, 1125, 1125~
$ minimum_minimum_nights        <dbl> 30, 1, 1, 30, 90, 2, 30, ~
$ maximum_minimum_nights        <dbl> 30, 1, 1, 30, 90, 30, 30,~
$ minimum_maximum_nights        <dbl> 1125, 1125, 1125, 1125, 1~
$ maximum_maximum_nights        <dbl> 1125, 1125, 1125, 1125, 1~
$ minimum_nights_avg_ntm        <dbl> 30.0, 1.0, 1.0, 30.0, 90.~
$ maximum_nights_avg_ntm        <dbl> 1125.0, 1125.0, 1125.0, 1~
$ calendar_updated              <lgl> NA, NA, NA, NA, NA, NA, N~
$ has_availability              <lgl> TRUE, TRUE, TRUE, TRUE, T~
$ availability_30               <dbl> 1, 13, 7, 1, 30, 5, 8, 10~
$ availability_60               <dbl> 14, 28, 12, 31, 60, 5, 38~
$ availability_90               <dbl> 44, 58, 23, 33, 90, 5, 68~
$ availability_365              <dbl> 319, 311, 261, 115, 243, ~
$ calendar_last_scraped         <date> 2023-06-18, 2023-06-18, ~
$ number_of_reviews             <dbl> 92, 312, 530, 76, 160, 50~
$ number_of_reviews_ltm         <dbl> 2, 29, 90, 3, 2, 13, 0, 1~
$ number_of_reviews_l30d        <dbl> 0, 2, 7, 0, 0, 2, 0, 0, 0~
$ first_review                  <date> 2011-09-21, 2015-04-01, ~
$ last_review                   <date> 2022-12-11, 2023-05-31, ~
$ review_scores_rating          <dbl> 4.51, 4.87, 4.77, 4.79, 4~
$ review_scores_accuracy        <dbl> 4.56, 4.94, 4.87, 4.81, 4~
$ review_scores_cleanliness     <dbl> 4.70, 4.91, 4.82, 4.96, 4~
$ review_scores_checkin         <dbl> 4.85, 4.98, 4.94, 4.97, 4~
$ review_scores_communication   <dbl> 4.80, 4.97, 4.94, 4.99, 5~
$ review_scores_location        <dbl> 4.84, 4.88, 4.89, 4.62, 4~
$ review_scores_value           <dbl> 4.49, 4.85, 4.89, 4.69, 4~
$ license                       <lgl> NA, NA, NA, NA, NA, NA, N~
$ instant_bookable              <lgl> FALSE, TRUE, FALSE, FALSE~
```

```
$ calculated_host_listings_count              <dbl> 2, 2, 2, 2, 1, 2, 4, 2, 2~
$ calculated_host_listings_count_entire_homes <dbl> 2, 2, 2, 2, 1, 2, 1, 2, 2~
$ calculated_host_listings_count_private_rooms <dbl> 0, 0, 0, 0, 0, 0, 3, 0, 0~
$ calculated_host_listings_count_shared_rooms <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0~
$ reviews_per_month                           <dbl> 0.64, 3.12, 5.54, 0.73, 1~


Call:
lm(formula = bedrooms ~ accommodates + bedrooms, data = Airbnb)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0516 -0.3853 -0.0264  0.2557  4.9736

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.30847    0.02085   14.80   <2e-16 ***
accommodates 0.35895    0.00366   98.07   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5571 on 2766 degrees of freedom
  (471 observations deleted due to missingness)
Multiple R-squared:  0.7766,    Adjusted R-squared:  0.7765
F-statistic:  9617 on 1 and 2766 DF,  p-value: < 2.2e-16


[1] "/Users/barbaraflores/Desktop/MIDS/IDS702_ModelingAndRepresentationOfData/20231001_Airbn|
```