Final Group Project

AirBnB Analytics

Group A15
Dhruv Ghurburrun, Michael Hiski, Johnna Liu, Vincent Lyu, Helena Martin, Dhruv Pathak

INTRODUCTION

- The goal of this analysis is to understand the main drivers of AirBnB prices in Brussels.

- **Main question: What price should 2 people be expected to pay if they plan to stay in an AirBnB in Brussels for 4 nights?**

EDA

- Dataset includes **74** variables on **5442** listings
- Variables include listings' price per night, amenities, #bedrooms, host response rate, neighborhood, etc.
- Top 4 property types: 'Entire rental unit' (52.7%), 'Private room in rental unit' (13.2%), 'Entire condominium (5.48%)', and 'Private room in residential home' (5.2%)

variable <chr>	min <dbl>	Q1 <dbl>	median <dbl>	Q3 <dbl>	max <dbl>	mean <dbl>	sd <dbl>	n <int>	missing <int>
accommodates	0	2.00	2.00	4.0000	16.0	3.0086365	1.770648e+00	5442	0
availability_30	0	0.00	3.00	19.0000	30.0	9.0852628	1.076625e+01	5442	0
availability_365	0	35.00	148.00	306.0000	365.0	166.5216832	1.340410e+02	5442	0
availability_60	0	0.00	21.00	45.0000	60.0	23.0020213	2.248474e+01	5442	0
availability_90	0	0.00	41.00	74.0000	90.0	39.2603822	3.454298e+01	5442	0
bathrooms	0	1.00	1.00	1.0000	19.5	1.1927555	5.640093e-01	5411	31
bedrooms	1	1.00	1.00	2.0000	40.0	1.4014963	1.046250e+00	4812	630
beds	0	1.00	1.00	2.0000	16.0	1.7116999	1.263640e+00	5359	83
host_listings_count	0	1.00	1.00	4.0000	2044.0	9.6409926	3.981678e+01	5440	2
host_total_listings_count	0	1.00	1.00	4.0000	2044.0	9.6409926	3.981678e+01	5440	2
maximum_minimum_nights	1	1.00	2.00	5.0000	1125.0	10.5024812	3.606691e+01	5441	1
maximum_nights	1	90.00	1125.00	1125.0000	8888888.0	2339.1253216	1.204863e+05	5442	0
maximum_nights_avg_ntm	1	365.00	1125.00	1125.0000	8888888.0	2472.0954604	1.204954e+05	5441	1
minimum_maximum_nights	1	365.00	1125.00	1125.0000	8888888.0	2458.0257306	1.204956e+05	5441	1
minimum_minimum_nights	1	1.00	2.00	4.0000	1125.0	9.9101268	3.585474e+01	5441	1
minimum_nights	1	1.00	2.00	4.0000	1125.0	10.2908857	3.619252e+01	5442	0
minimum_nights_avg_ntm	1	1.00	2.00	4.1000	1125.0	10.2651167	3.597720e+01	5441	1
number_of_reviews	0	2.00	8.00	35.0000	782.0	35.3724734	6.970097e+01	5442	0
number_of_reviews_l30d	0	0.00	0.00	1.0000	20.0	0.7614847	1.644381e+00	5442	0
number_of_reviews_ltm	0	0.00	1.00	5.0000	167.0	5.1352444	1.156730e+01	5442	0
price	0	46.00	65.00	92.0000	5000.0	87.1297317	1.323741e+02	5442	0
review_scores_accuracy	0	4.67	4.85	5.0000	5.0	4.7239023	4.523865e-01	4482	960
review_scores_rating	0	4.50	4.75	4.9225	5.0	4.5911197	6.504968e-01	4528	914

REGRESSION MODELS

Instead of (price_4_nights), we used log(price_4_nights) as its distribution is normal.

Main conclusions :

- Model 1: Statistically insignificant negative relationship between rating and price
- Model 2: Room type is a statistically significant predictor of price, except hotel rooms
- Model 3: Number of bedrooms, bathrooms, size of house are statistically significant positive predictors, whereas number of beds is insignificant.
- Model 4: Statistically insignificant positive correlation between being a superhost and price.
- Model 5: Statistically significant positive correlation between being instantly bookable and of price.
- Model 6: Location is a statistically significant predictor of price, except East Brussels.
- Model 7: Number of reviews is insignificant, but review score/rating is positively correlated and significant. Reviews/month and 30-day availability are significant and positively correlated.

MODEL RESULTS

	Model1	Model2	Model3	Model4	Model5	Model6	Model 7
#observations	2286	2655	2346	2024	2363	2363	2025
R squared	0.156	0.239	0.416	0.447	0.416	0.428	0.559
Adj. R Squared	0.154	0.237	0.413	0.444	0.413	0.424	0.555
Residual SE	0.530	0.508	0.451	0.439	0.453	0.449	0.392

RMSE_model1	RMSE_model2	RMSE_model3	RMSE_model4	RMSE_model5	RMSE_model6	RMSE_model7
0.518	0.504	0.441	0.418	0.441	0.441	0.37

Our Best Model 7

```
                                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                                    4.9696786  0.0839576  59.193  < 2e-16 ***
prop_type_simplifiedEntire rental unit        -0.0084988  0.0436885  -0.195 0.845778
prop_type_simplifiedOther                      0.1635574  0.0491512   3.328 0.000892 ***
prop_type_simplifiedPrivate room in rental unit -0.0685048  0.0616751  -1.111 0.266815
prop_type_simplifiedPrivate room in residential home 0.0507285  0.0685730   0.740 0.459524
review_scores_rating                           0.0383368  0.0139127   2.756 0.005913 **
room_typeHotel room                            0.0863199  0.0977376   0.883 0.377245
room_typePrivate room                         -0.4252994  0.0398456 -10.674  < 2e-16 ***
room_typeShared room                          -0.8333446  0.1175003  -7.092 1.82e-12 ***
bathrooms                                      0.0366305  0.0163751   2.237 0.025399 *
bedrooms                                       0.0340481  0.0110024   3.095 0.001998 **
accommodates                                   0.1202994  0.0064751  18.579  < 2e-16 ***
instant_bookableTRUE                           0.0677469  0.0187699   3.609 0.000314 ***
neighbourhood_simplifiedEast/Centre            0.0003808  0.0353283   0.011 0.991402
neighbourhood_simplifiedNorth East             0.0969861  0.0243075   3.990 6.85e-05 ***
neighbourhood_simplifiedNorth West            -0.1481909  0.0405403  -3.655 0.000263 ***
neighbourhood_simplifiedSouth/Centre           0.0494046  0.0272550   1.813 0.070031 .
reviews_per_month                             -0.0533501  0.0050452 -10.574  < 2e-16 ***
availability_30                                0.0166477  0.0008780  18.960  < 2e-16 ***

Residual standard error: 0.3923 on 2006 degrees of freedom
  (630 observations deleted due to missingness)
Multiple R-squared:  0.5586,    Adjusted R-squared:  0.5546
F-statistic:   141 on 18 and 2006 DF,  p-value: < 2.2e-16
```

We choose Model 7 because it has the lowest RMSE & the highest R-squared.

FINAL PREDICTION

- Question: What is the AirBnB price for 2 people for 4 nights in Brussels?

- We use Model 7 to answer.
  - Private room in rental unit
  - 2 people, 4 nights
  - Located in North-West Brussels
  - Number of reviews >= 10
  - Average rating score >= 4.5

- Expected price: 123.4 euros, 95% upper price is 131.6 Euros, and 95% lower price 115.7 Euros.