*****

AirBnB Project

- Milan

Data Summary

01

Character

24 character data
E.g. name, listing_url,
description and host_name

02

Date

5 date data
E.g. host_since, last_scraped,
frist_review, last_reveiw

03

Logical

8 logical data
Take a certain value based on a logical test
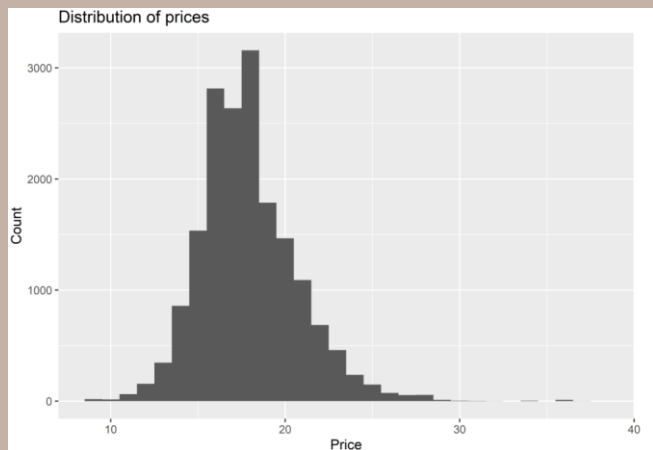E.g. bathrooms, host_is_superhost,
instant_bookable

04

Numeric

37 numeric data
E.g. host_listings_count, latitude,
accommodates, beds, minimum_nights
and availability_60

Distribution of prices

Price Distribution

Due to the highly right-skewed distribution of the original price data, we introduced a new variable to better capture price range and formed a log distribution
-- a large group of outliers was detected

Property Types

- 52 unique property types
- Top 4 most common property types make up more than 85% of total properties types

property_type	n	property_sum	property_perc
Entire rental unit	10178	17703	0.575
Private room in rental unit	2661	17703	0.15
Entire condominium (condo)	1830	17703	0.103
Entire loft	833	17703	0.0471
Private room in condominium (condo)	631	17703	0.0356
Entire residential home	274	17703	0.0155
Entire serviced apartment	189	17703	0.0107
Shared room in rental unit	182	17703	0.0103
Private room in residential home	168	17703	0.00949

	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	6.457 ***	6.697 ***	6.609 ***	6.660 ***	6.866 ***	6.674 ***
	(0.245)	(0.047)	(0.279)	(0.051)	(0.053)	(0.055)
prop_type_simplifiedEntire loft	-0.052					
	(0.230)					
prop_type_simplifiedEntire rental unit	-0.015					
	(0.144)					
prop_type_simplifiedOther	-0.074					
	(0.203)					
prop_type_simplifiedPrivate room in rental unit	-0.323					
	(0.167)					
number_of_reviews	-0.003 **	-0.004 ***	-0.004 **	-0.004 ***	-0.004 ***	-0.003 **
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
review_scores_rating	0.027					
	(0.043)					
room_typePrivate room		-0.341 ***	-0.414 ***	-0.364 ***	-0.281 **	-0.288 ***
		(0.093)	(0.100)	(0.093)	(0.088)	(0.080)
bathrooms_clean			0.049			
			(0.179)			
bedrooms				0.013		
				(0.204)		
beds				0.063		
				(0.091)		
host_is_superhostTRUE				-0.077		
				(0.149)		
instant_bookableTRUE					0.174	
					(0.092)	
neighbourhood_simplifiedEAST					-0.498 ***	-0.441 ***
					(0.132)	(0.121)
neighbourhood_simplifiedNORTH					-0.452 ***	-0.501 ***
					(0.095)	(0.087)
neighbourhood_simplifiedSOUTH					-0.586 ***	-0.563 ***
					(0.130)	(0.118)
neighbourhood_simplifiedWEST					0.001	-0.093
					(0.140)	(0.128)
availability_30						0.021 ***
						(0.003)
N	190	274	238	274	274	274
R2	0.072	0.088	0.107	0.100	0.218	0.354
logLik	-168.613	-265.892	-230.403	-264.073	-244.811	-218.579
AIC	353.226	539.783	476.805	538.146	505.621	455.158

*** p < 0.001; ** p < 0.01; * p < 0.05.

Comparison of 6 Models, including model names, significant predictors, adjusted R^2 and the Residual Standard Error

Best Model – model6.2

- fitted line fluctuates around the real line, which is within the range of 95% confidence interval
- highest R-squared (0.354)
- best fits our observations

Predict and Real Total Costs for 2 Person 4 Night