TRABAJO FINAL INFERENCIA ESTADÍSTICA II

Integrantes del grupo: Daniel Aramburu, Jorge Salas y Marc Pastor

Fuente de datos

En este trabajo vamos a analizar una base de datos de apartamentos de Airbnb de la ciudad de Barcelona. Los datos los recopilamos de los enlaces que hay más abajo, aunque hemos hecho ya la limpieza y el data wrangling porque estaban muy desordenados. En este enlace está el código de la limpieza de datos y un poco de análisis exploratorio (el código está en inglés porque este *dataset* lo usé el verano pasado para practicar en un proyecto). El archivo original contenía muchas más observaciones, pero contenía muchos NA's y celdas vacías, que rellené con la mediana (en el caso de variables numéricas) o la moda del barrio en el que se encuentra el apartamento.

http://insideairbnb.com/get-the-data.html

http://data.insideairbnb.com/spain/catalonia/barcelona/2021-02-09/data/listings.csv.g

Conjunto de datos:

El dataset tiene muchas variables pero no las usaremos todas. Principalmente nos basaremos en comparar precios y otras variables numéricas, agrupando por distintos atributos categóricos.

'data.frame':   7364 obs. of 156 variables:
$ host_since_year   : num 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
$ host_response_rate   : num 1 1 1 1 1 1 1 1 1 ...
$ host_acceptance_rate   : num 0.91 0.91 0.91 1 1 1 1 0.77 1 0.55 ...
$ host_is_superhost   : chr "FALSE" "FALSE" "FALSE" "TRUE" ...
$ host_neighbourhood   : chr "El GÃ²tic" "El GÃ²tic" "El GÃ²tic" "L'Antiga Esquerra de l'Eixample" ...
$ host_total_listings_count   : int 3 3 3 4 4 4 4 1 7 7 ...
$ host_has_profile_pic   : chr "TRUE" "TRUE" "TRUE" "TRUE" ...
$ host_identity_verified   : chr "TRUE" "TRUE" "TRUE" "TRUE" ...
$ neighbourhood   : Factor w/ 69 levels "","Camp d'en Grassot i Gràcia Nova",..:
$ latitude   : num 41.4 41.4 41.4 41.4 41.4 ...
$ longitude   : num 2.18 2.18 2.18 2.15 2.15 ...
$ is_location_exact   : chr "TRUE" "TRUE" "TRUE" "TRUE" ...
$ property_type   : chr "Apartment" "Apartment" "Apartment" "Apartment" ...
$ room_type   : chr "Private room" "Private room" "Entire home/apt" "Private room" ...
$ accommodates   : int 2 2 9 2 1 1 2 1 4 6 ...
$ bathrooms   : num 1 1 3 2 2 2 2 1 1 2 ...
$ bedrooms   : int 1 1 4 1 1 1 1 1 1 2 ...
$ beds   : int 1 1 6 1 1 1 1 1 2 4 ...

```
$ bed_type                              : chr  "Real Bed" "Real Bed" "Real Bed" "Real Bed" ...
$ square_feet                           : num  807 807 807 807 807 807 807 807 807 807 ...
$ price                                 : num  80 100 227 40 30 30 45 33 130 110 ...
$ security_deposit                      : num  100 150 200 0 0 0 0 0 150 500 ...
$ cleaning_fee                          : num  20 40 67 6 6 6 6 15 59 30 ...
$ guests_included                       : int  2 1 4 1 1 1 1 1 2 3 ...
$ extra_people                          : num  0 0 25 10 0 0 10 0 10 25 ...
$ minimum_nights                        : int  3 5 4 7 2 2 2 2 2 3 ...
$ maximum_maximum_nights                : int  90 120 1125 1125 730 1125 1125 65 364 365 ...
$ availability_30                       : int  16 30 30 26 14 21 17 29 30 30 ...
$ availability_60                       : int  16 60 60 52 31 44 43 59 60 60 ...
$ availability_90                       : int  16 90 90 75 46 70 69 83 80 90 ...
$ availability_365                      : int  88 180 180 348 318 345 344 358 336 365 ...
$ number_of_reviews                     : int  2 8 149 303 238 258 222 73 339 39 ...
$ review_scores_rating                  : num  100 68 91 94 95 96 95 94 94 88 ...
$ review_scores_accuracy                : num  10 8 10 10 10 10 10 10 10 9 ...
$ review_scores_cleanliness             : num  10 8 9 9 10 10 9 10 10 9 ...
$ review_scores_checkin                 : num  10 7 10 10 10 10 10 10 9 9 ...
$ review_scores_communication           : num  10 9 10 10 10 10 10 10 10 9 ...
$ review_scores_value                   : num  10 7 9 10 10 9 9 10 10 9 ...
$ instant_bookable                      : chr  "FALSE" "FALSE" "TRUE" "TRUE" ...
$ cancellation_policy                   : chr  "moderate" "moderate" "moderate" "moderate" ...
$ require_guest_profile_picture         : chr  "FALSE" "FALSE" "FALSE" "FALSE" ...
$ require_guest_phone_verification      : chr  "FALSE" "FALSE" "FALSE" "FALSE" ...
$ calculated_host_listings_count        : int  3 3 3 4 4 4 4 1 3 4 ...
$ calculated_host_listings_count_entire_homes : int  1 1 1 0 0 0 0 0 3 4 ...
$ calculated_host_listings_count_private_rooms: int  2 2 2 4 4 4 4 1 0 0 ...
$ calculated_host_listings_count_shared_rooms : int  0 0 0 0 0 0 0 0 0 0 ...
$ reviews_per_month                     : num  0.05 0.07 1.26 3.01 2.16 2.62 3.17 0.69 3.09 0.4 ...
$ email                                 : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
$ phone                                 : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
$ facebook                              : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 1 1 1 ...
$ reviews                               : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
$ manual_offline                        : Factor w/ 2 levels "Yes","No": 1 1 1 2 2 2 2 1 2 2 ...
$ jumio                                 : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
$ offline_government_id                 : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 2 2 2 ...
$ sent_id                               : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ selfie                                : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ government_id                         : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
$ identity_manual                       : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ work_email                            : Factor w/ 2 levels "Yes","No": 1 1 1 2 2 2 2 2 1 2 ...
$ TV                                    : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Cable_TV                              : Factor w/ 2 levels "Yes","No": 1 1 1 2 2 2 2 2 2 2 ...
$ Internet                              : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
$ Wifi                                  : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...
$ Air_conditioning                      : Factor w/ 2 levels "Yes","No": 1 1 1 2 2 2 2 2 1 1 1 ...
```

$ Wheelchair_accessible : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 1 ...	
$ Kitchen : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...	
$ Paid_parking_off_premises : Factor w/ 2 levels "Yes","No": 2 2 1 2 2 2 2 2 1 2 ...	
$ Elevator : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 2 1 ...	
$ Buzzerwireless_intercom : Factor w/ 2 levels "Yes","No": 2 2 1 1 2 2 1 1 1 1 ...	
$ Heating : Factor w/ 2 levels "Yes","No": 1 1 1 1 1 1 1 1 1 1 ...	
$ Familykid_friendly : Factor w/ 2 levels "Yes","No": 1 1 1 2 2 2 2 2 1 1 ...	
$ Washer : Factor w/ 2 levels "Yes","No": 1 2 1 1 1 1 1 2 1 1 ...	
$ Smoke_alarm : Factor w/ 2 levels "Yes","No": 1 2 1 2 2 2 2 2 2 2 ...	
$ Carbon_monoxide_alarm : Factor w/ 2 levels "Yes","No": 1 2 1 2 2 2 2 2 2 2 ...	
$ First_aid_kit : Factor w/ 2 levels "Yes","No": 2 2 1 2 2 2 2 2 2 2 ...	
$ Safety_card : Factor w/ 2 levels "Yes","No": 2 2 1 2 2 2 2 2 2 2 ...	
$ Fire_extinguisher : Factor w/ 2 levels "Yes","No": 1 2 1 2 2 2 2 2 2 2 ...	
$ Essentials : Factor w/ 2 levels "Yes","No": 1 2 1 1 1 1 1 1 1 1 ...	
$ Shampoo : Factor w/ 2 levels "Yes","No": 2 2 1 2 2 2 2 1 2 2 ...	
$ Hangers : Factor w/ 2 levels "Yes","No": 2 2 1 1 1 1 1 1 1 1 ...	
$ Hair_dryer : Factor w/ 2 levels "Yes","No": 2 2 1 1 1 1 1 1 1 1 ...	
$ Iron : Factor w/ 2 levels "Yes","No": 2 2 1 1 1 1 1 1 1 1 ...	
$ Laptop-friendly_workspace : Factor w/ 2 levels "Yes","No": 2 2 1 2 2 2 2 1 2 2 ...	
$ Private_entrance : Factor w/ 2 levels "Yes","No": 2 2 1 2 2 2 2 2 2 2 ...	
$ Outlet_covers : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ High_chair : Factor w/ 2 levels "Yes","No": 2 2 1 2 2 2 2 2 1 2 ...	
$ Stair_gates : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ Children's_books_and_toys : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ Window_guards : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ Table_corner_guards : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ Babysitter_recommendations : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ Crib : Factor w/ 2 levels "Yes","No": 2 2 1 2 2 2 2 2 1 2 ...	
$ Pack_'n_Playtravel_crib : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 1 2 ...	
$ Room-darkening_shades : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 1 2 2 ...	
$ Children's_dinnerware : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ Hot_water : Factor w/ 2 levels "Yes","No": 2 2 1 1 1 1 1 2 1 2 ...	
$ Body_soap : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ Bath_towel : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ Toilet_paper : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ Bed_linens : Factor w/ 2 levels "Yes","No": 2 2 1 2 1 2 1 2 1 2 ...	
$ Extra_pillows_and_blankets : Factor w/ 2 levels "Yes","No": 2 2 1 2 2 2 2 2 2 2 ...	
$ Ethernet_connection : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...	
$ Pocket_wifi : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 1 2 ...	

```
$ Microwave                  : Factor w/ 2 levels "Yes","No": 2 1 1 1 1 1 1 2 1 2 ...
$ Coffee_maker               : Factor w/ 2 levels "Yes","No": 2 1 1 1 1 1 1 2 1 2 ...
$ Refrigerator               : Factor w/ 2 levels "Yes","No": 2 1 1 1 2 1 1 2 1 2 ...
$ Dishwasher                 : Factor w/ 2 levels "Yes","No": 2 1 1 2 2 2 2 2 2 2 ...
$ Dishes_and_silverware      : Factor w/ 2 levels "Yes","No": 2 1 1 1 1 1 1 2 1 2 ...
$ Cooking_basics             : Factor w/ 2 levels "Yes","No": 2 1 1 1 1 1 1 2 1 2 ...
$ Oven                       : Factor w/ 2 levels "Yes","No": 2 1 1 2 2 2 2 2 2 2 ...
$ Stove                      : Factor w/ 2 levels "Yes","No": 2 1 1 1 1 1 1 2 1 2 ...
$ Single_level_home          : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Patio_or_balcony           : Factor w/ 2 levels "Yes","No": 2 2 2 1 2 2 1 2 2 2 ...
$ Luggage_dropoff_allowed    : Factor w/ 2 levels "Yes","No": 2 2 1 2 2 2 2 2 1 2 ...
$ Wide_hallways              : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ No_stairs_or_steps_to_enter: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Wide_entrance_for_guests   : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Flat_path_to_guest_entrance: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Well-lit_path_to_entrance  : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Extra_space_around_bed     : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Accessible-height_bed      : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Accessible-height_toilet   : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Wide_clearance_to_shower   : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ _toilet                    : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Wide_entryway              : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Host_greets_you            : Factor w/ 2 levels "Yes","No": 2 2 1 1 1 1 1 2 1 2 ...
$ Waterfront                 : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Beachfront                 : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 1 2 2 ...
$ Handheld_shower_head       : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Hot_water_kettle           : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Ceiling_fan                : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Beach_view                 : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Rain_shower                : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Bidet                      : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Heated_towel_rack          : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Balcony                    : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Printer                    : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Espresso_machine           : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Formal_dining_area         : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Day_bed                    : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Convection_oven            : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Standing_valet             : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Pillow-top_mattress        : Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
```

$ Memory_foam_mattress	: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 ...
$ En_suite_bathroom	: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Outdoor_seating	: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Mudroom	: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Full_kitchen	: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Paid_parking_on_premises	: Factor w/ 2 levels "Yes","No": 2 2 1 1 1 1 1 2 2 2 ...
$ Bedroom_comforts	: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Bathroom_essentials	: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Fixed_grab_bars_for_shower	: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ Shower_chair	: Factor w/ 2 levels "Yes","No": 2 2 2 2 2 2 2 2 2 2 ...
$ days_since_host	: num 4070 4070 4070 3918 3918 ...
$ days_since_first_review	: num 1401 3903 3818 3291 3579 ...
$ days_since_last_review	: num 1227 2802 371 364 277 ...