

Original SQL Schema (as shown in Phase 1 of Report)

Table Name: ListingsOG

Column Names:

- *id* - The unique identifier of that particular AirBnb listing.
- *listing_url* - link to listing on Airbnb.com
- *scrape_id* - ID used when scraping data from Airbnb.com. Not useful
- *last_scraped* - Date when data was last scraped. Not useful
- *name* - The name of the listing
- *description* - This category provides additional information about the listing and describes the listing in more detail.
- *neighborhood_overview* - This category describes the neighborhood in terms of where it is located, the type of ambience, and other descriptive information.
- *picture_url* - check to make sure url is active*
- *host_id* - contains the id for the host of the listing. For context
- *host_url* - link to host's profile on Airbnb.com
- *host_name* - name of the host
- *host_since* - how long the host has been active
- *host_location* - where the host is primarily located. This can be and is most often different from the location of the listing.
- *host_about* - An about paragraph regarding the host. This contains additional information about the host that would not be
- *host_response_time* - how long the host usually takes to reply to a request for booking.
- *host_response_rate* - how often the host responds to booking requests.
- *host_acceptance_rate* - how often the host accepts booking requests.
- *host_is_superhost* - Boolean for whether host is a superhost (T/F).
- *host_thumbnail_url* - Small photo of host used for listing thumbnail
- *host_picture_url* - Medium photo of host used in listing
- *host_neighbourhood* - same as *neighbourhood* column; redundant.
- *host_listings_count* - how many listings the host currently has on AirBnb.
- *host_total_listings_count* - total number of listings host has listed.
- *host_verifications* - Information from host that has been verified (e.g. email, phone, social media, etc.)
- *host_has_profile_pic* - Boolean for whether host has a profile picture (T/F).
- *host_identity_verified* - boolean value of whether the host's identity has been verified.
- *neighbourhood* - Original neighbourhood data as extracted from airbnb listing. Redundant to *Neighbourhood_cleansed* and *Neighbourhood_group_cleansed*.
- *neighbourhood_cleansed* - Neighbourhood data that has already been cleansed in the original dataset, D. We assume the original author used latitude/longitude data to generate this.
- *neighbourhood_group_cleansed* - Neighbourhood group data that has already been cleansed in the original dataset, D. We assume the original author used latitude/longitude data to generate this.
- *latitude* - the latitude coordinates of the listing's location.
- *longitude* - the longitude coordinates of the listing's location.
- *property_type* - The type of property (e.g. boat, entire apartment, private room, etc). More descriptive than *Room_type*.
- *room_type* - This defines whether the listing is a private room or an entire home/apartment.
- *accommodates* - The number of people the listing accommodates.
- *bathrooms* - The number of bathrooms at the listing.
- *bathrooms_text* - This category defines the number of bathrooms as well as if the bathroom is a shared or private bathroom.

- *bedrooms* - The number of bedrooms at the listing.
- *beds* - The number of beds at the listing.
- *amenities* - This category lists all the amenities the listing has in a list. Examples of amenities are kitchen, cable tv, heating, bathtub, etc,
- *price* - The price of the listing.
- *minimum_nights* - minimum nights to stay
- *maximum_nights* - maximum nights to stay
- *minimum_minimum_nights* - minimum minimum nights the customer must stay
- *maximum_minimum_nights* - maximum minimum nights the customer must stay
- *minimum_maximum_nights* - minimum maximum nights the customer must stay
- *maximum_maximum_nights* - maximum maximum nights the customer must stay
- *minimum_nights_avg_ntm* - average minimum nights the host has requested that customer stay in the next twelve months
- *maximum_nights_avg_ntm* - average maximum nights the host has requested that customer stay in the next twelve months
- *calendar_updated* - Blank column
- *has_availability* - A boolean value that indicates if the listing is available or not.
- *availability_30* - Availability in the month
- *availability_60* - Availability in two months
- *availability_90* - Availability in a quarter
- *availability_365* - Availability in a year
- *calendar_last_scraped* - same as last_scraped
- *number_of_reviews* - The number of reviews corresponding to that listing.
- *number_of_reviews_ltm* - Number of reviews in the last 12 months.
- *number_of_reviews_130d* - Number of reviews in the last 30 days.
- *first_review* - The date of the first review
- *last_review* - The data of the last review
- *review_scores_rating* - Overall score of the listing with scale 20 thru 100 (divide by 20 to get 1 - 5 stars).
- *review_scores_accuracy* - Score of the listing's accuracy with scale 2 thru 10 (divide by 2 to get 1 - 5 stars).
- *review_scores_cleanliness* - Score of the listing's cleanliness with scale 2 thru 10 (divide by 2 to get 1 - 5 stars).
- *review_scores_checkin* - Score of the listing's check-in with scale 2 thru 10 (divide by 2 to get 1 - 5 stars).
- *review_scores_communication* - Score of the listing's communication with scale 2 thru 10 (divide by 2 to get 1 - 5 stars).
- *review_scores_location* - Score of the listing's location with scale 2 thru 10 (divide by 2 to get 1 - 5 stars).
- *review_scores_value* - Score of the listing's value with scale 2 thru 10 (divide by 2 to get 1 - 5 stars).
- *license* - blank column; unnecessary
- *instant_bookable* - Boolean on whether the listing is instantly bookable (T/F).
- *calculated_host_listings_count* - Total number of listings from host (based on host_id)
- *calculated_host_listings_count_entire_homes* - Total number of listings that are entire homes from host (based on host_id)
- *calculated_host_listings_count_private_rooms* - Total number of listings that are private rooms from host (based on host_id)
- *calculated_host_listings_count_shared_rooms* - Total number of listings that are shared rooms from host (based on host_id)
- *reviews_per_month* - number of reviews per month

SQL Schema After Cleaning:

Table Name: Listings

Column Names:

- *name*
- *description*
- *picture_url*
- *host_id*
- *host_name*
- *host_since*
- *host_location*
- *host_response_time*
- *host_identity_verified*
- *room_type*
- *accommodates*
- *bathrooms*
- *bedrooms*
- *beds*
- *amenities*
- *price*
- *minimum_nights*
- *maximum_nights*
- *number_of_reviews*
- *review_scores_rating*
- *neighborhood*

Descriptions of these fields can be found in the section above.