This document contains the data dictionary for the Airbnb Rating Classification project. The project had the following tables:

- Calendar – CSV
- Listings – CSV
- Listings_detailed – CSV
- Neighbourhoods – CSV
- Neighbourhoods – GeoJSON
- Reviews – CSV
- Reviews_detailed - CSV

Listings Detailed Data Dictionary:

- id – Unique id number associated with a listing.
- listing_url – The listing's URL.
- scrape_id – The particular web scrape id where a single scrape_id is associated with each city/area.
- last_scraped – the date of when the listing was last web scraped.
- source – The source of web scraping. The values can either be 'city scrape' or 'previous scrape'.
- name - Name of the listing.
- description – Description of the listing.
- neighborhood_overview – Description of the listing's neighborhood provided by the host.
- picture_url – A URL of the picture of the property listing.
- host_id – Unique identifier associated with the host.
- host_url – URL of the listing's host/URL of the host's profile.
- host_name – Name of the listing's host.
- host_since – How long has the listing's host been on Airbnb as a host. This information is from the host's profile.
- host_location – The host's location.
- host_about – Contains the 'host about' information found on the host's profile page.
- host_response_time – The host's typical response time, which is categorical data.
- host_response_rate - The percentage or rate at which a host responds to guests' inquiries or booking requests. This metric is on a continuous scale.
- host_acceptance_rate - The percentage of booking requests that a host accepts out of the total number of requests they receive. Metric reflects a host's willingness/likelihood to accept guests into this listing. This metric is on a continuous scale.
- host_is_superhost - This field indicates whether a host has been awarded Superhost status by Airbnb. Superhost status is a recognition given to hosts who provide exceptional service, including high response and acceptance rates, a record of positive reviews, and a commitment to maintaining an active and welcoming listing on the platform. This status is reviewed and updated by Airbnb on a quarterly basis.

- host_thumbnail_url – The thumbnail photo URL of the host on the host's profile.
- host_picture_url – The URL of the host's picture found on the host's profile.
- host_neighbourhood – The host's neighborhood.
- host_listings_count – This field represents the number of active listings that a host currently has on Airbnb. From the time of the web scraping, this field indicates how many listings the host has that are actively being offered to guests and are searchable or bookable on the platform.
- host_total_listings_count – The field represents the total number of listings associated with the host, which includes both active/currently available for booking and inactive bookings. For any listing that is inactive, it can be inactive due to being temporarily unavailable or the host choosing to not currently offer the listing.
- host_verifications – The field contains an array indicating all the methods/channels on how the host verified their identity and reliability with Airbnb.
- host_has_profile_pic – A Boolean to indicate whether the host has a profile picture.
- host_identity_verified - A Boolean to indicate whether the host has gone through Airbnb's verification process.
- Neighbourhood – The listing's neighborhood.
- neighbourhood_cleansed - This field represents a standardized version of the neighborhood names where the listings are located compared to the 'Neighbourhood' column.
- neighbourhood_group_cleansed – When applicable, it groups neighborhoods into larger districts, areas, or boroughs to provide a more macro-level view of where listings are in a standardized format.
- latitude – The listing's latitude coordinates.
- longitude – The listing's longitude coordinates.
- property_type – A categorical variable that indicates the listing's offered accommodation type.
- room_type - A categorical variable that indicates the listing's offered room type. This field provides a simpler categorization compared to 'property_type' column.
- accommodates – The number of people the listing can accommodate.
- bathrooms – The number of bathrooms in the listing.
- bathrooms_text – Field provides more detailed information on the number of bathrooms and the type of bathroom.
- bedrooms – The number of bedrooms in the listing.
- beds – The number of beds in the listing.
- amenities - The field contains an array indicating all the amenities included in the listing.
- price – The field contains the cost to rent the listing per night.
- minimum_nights – The minimum number of nights a guest must book for a stay at the listed property.
- maximum_nights - The maximum number of nights a guest is allowed to book for a stay at the listed property.

- minimum_minimum_nights - Field represents the lowest number of minimum nights required for a booking that the host has dictated over the course of the listing appearing on the platform.
- maximum_minimum_nights - Field represents the highest number of minimum nights required for a booking that the host has dictated over the course of the listing appearing on the platform.
- minimum_maximum_nights - Field represents the lowest number of maximum nights allowed for a booking that the host has dictated over the course of the listing appearing on the platform.
- maximum_maximum_nights - Field represents the highest number of maximum nights allowed for a booking that the host has dictated over the course of the listing appearing on the platform.
- minimum_nights_avg_ntm – Field represents the average minimum nights required over the next twelve months (ntm). This field is important when considering the variation in minimum stay requirements throughout the year.
- maximum_nights_avg_ntm - Field represents the average maximum nights allowed over the next twelve months (ntm). This field is important when considering the variation in maximum stay allowed throughout the year.
- calendar_updated – The field represents when was the last time the host updated the listing's calendar to reflect changes in pricing or setting new minimum or maximum night requirements.
- has_availability - A Boolean field indicating whether the listing is currently available for booking.
- availability_30 – Field represents the number of days the listing is available for booking in the next 30 days from the current date.
- availability_60 – Field represents the number of days the listing is available for booking in the next 60 days from the current date.
- availability_90 – Field represents the number of days the listing is available for booking in the next 90 days from the current date.
- availability_365 – Field represents the number of days the listing is available for booking in the next 365 days from the current date.
- calendar_last_scraped – The field represents the last time the data was web-scraped, so the field effectively represents the current date.
- number_of_reviews – The field represents the listing's total number of reviews since the listing became available on the platform.
- number_of_reviews_ltm – Field represents the listing's total number of reviews received within the last twelve months (ltm).
- number_of_reviews_l30d - Field represents the listing's total number of reviews received within the last 30 days (l30d).
- first_review – A date field to represent the first review the listing received.
- last_review – A date field to represent the most recent review the listing received.
- review_scores_rating – The average overall review score rating.

- review_scores_accuracy – The average review score rating for how accurate the listing description and photos matched the guests' experiences.
- review_scores_cleanliness – The average review score rating for how satisfied the guests are with the cleanliness of the listing.
- review_scores_checkin – The average review score rating for how easy it was for the guests to check in.
- review_scores_communication – The average review score rating for how easy it was for the guests to communicate with the host.
- review_scores_location - The average review score rating for how satisfied the guests are with the listing's location, which takes into account factors such as convenience, safety, and proximity to points of interest.
- review_scores_value - The average review score rating for whether the guest considered the listing as a value given the amount paid to book the listing.
- license - This field indicates whether the host has provided a licensing number for their listing, which may be required by local regulations for short-term rentals.
- instant_bookable – A Boolean to indicate whether guests can book the listing instantly without waiting for the host's approval.
- calculated_host_listings_count – The field represents the total number of active listings the host has on Airbnb.
- calculated_host_listings_count_entire_homes – Field represents the total number of active listings the host has on Airbnb that are an entire home or apartment.
- calculated_host_listings_count_private_rooms – The field represents the total number of active listings the host has on Airbnb that are private rooms.
- calculated_host_listings_count_shared_rooms – The field represents the total number of active listings the host has on Airbnb that are shared rooms.
- reviews_per_month – An average number of reviews the listing receives per month.

Listings Data Dictionary:

- id - Unique id number associated with a listing.
- name - Name of the listing.
- host_id – Unique identifier associated with the host.
- host_name – Name of the listing's host.
- neighbourhood_group - When applicable, it groups neighborhoods into larger districts, areas, or boroughs to provide a more macro-level view of where listings are in a standardized format.
- neighbourhood - This field represents a standardized version of the neighborhood names for where the listing is located.
- latitude – The listing's latitude coordinates.
- longitude – The listing's longitude coordinates.
- room_type - A categorical variable that indicates the listing's offered room type.
- price – The field contains the cost to rent the listing per night.

- minimum_nights – The minimum number of nights a guest must book for a stay at the listed property.
- number_of_reviews – The field represents the listing's total number of reviews since the listing became available on the platform.
- last_review – A date field to represent the most recent review the listing received.
- reviews_per_month – An average number of reviews the listing receives per month.
- calculated_host_listings_count – Field represents the total number of active listings the host has on Airbnb.
- availability_365 – Field represents the number of days the listing is available for booking in the next 365 days from the current date.
- number_of_reviews_ltm – Field represents the listing's total number of reviews received within the last twelve months (ltm).
- license - This field indicates whether the host has provided a licensing number for their listing, which may be required by local regulations for short-term rentals.

Calendar Data Dictionary:
- listing_id – The unique identifier for each Airbnb listing.
- date – The specific date for which the listing's availability and price information are provided.
- available – Indicates whether the listing is available ('f' for false, meaning not available) on the given date.
- price – The original listed price for the property on the specified date, before any adjustments.
- adjusted_price – The price for the property on the specified date after adjustments, if any.
- minimum_nights – The minimum number of nights a guest must stay if booking for the given date.
- maximum_nights – The maximum number of nights a guest can stay when booking for the given date.

Neighbourhood Data Dictionary:
- neighbourhood_group – This column is intended to categorize listings into broader administrative or geographical regions but appears to be missing or not applicable in some datasets. It could refer to a higher-level grouping of areas, such as boroughs or districts in a large city.
- neighbourhood – Identifies the specific neighbourhood or district where the listing is located, offering a more detailed geographic classification than the neighbourhood_group. It helps users to understand the precise area of each listing.

Neighbourhoods.geojson Data Dictionary:
- type – This indicates the type of GeoJSON object. The root type is "FeatureCollection", which means this object is a collection of features.

- features – This is an array of Feature objects. Each feature represents a geographic feature and contains its own set of properties and geometry.
- For each object in the features array:
  - type – Specifies the type of GeoJSON object, in this case, "Feature".
  - geometry – Contains the geographic data of the feature.
    - type – The type of geometry, which can be one of several types such as "Point", "LineString", "Polygon", or "MultiPolygon".
    - coordinates – An array of points that define the geometry according to the geometry type. For "MultiPolygon", this is an array of polygons, where each polygon is defined by an array of linear rings (an array of points). The first linear ring represents the outer boundary of the polygon, and any subsequent rings represent holes in the polygon.
  - properties – An object containing properties of the feature. This can include any custom data associated with the feature.
    - neighbourhood: The name of the neighborhood.
    - neighbourhood_group: This is null in all provided features, which could mean that the neighborhood does not belong to a larger group or this information is not available.

Reviews Detailed Data Dictionary:
- listing_id – A unique identifier for the listing on the platform.
- id – A unique identifier for the review itself. This distinguishes each review from others, even for the same listing.
- date – The date when the review was posted. This appears to be in the format MM/DD/YY.
- reviewer_id – A unique identifier for the user who wrote the review. This helps in identifying and possibly tracking the review history of the user across different listings.
- reviewer_name – The name of the user who posted the review. It provides a human-readable identification for the reviewer, but names might not be unique.
- comments – The textual content of the review written by the user. This field contains the reviewer's thoughts, experiences, and opinions about their stay at the listing. It may include descriptions of the property, interactions with the host, and any other relevant information that the reviewer wishes to share.

Reviews Data Dictionary:

- listing_id – A unique identifier for each listing on the platform.
- date – The date when the review was posted. The format appears to be MM/DD/YY