# Airbnb Dataset

The Superhost certication program is used by Airbnb to provide guests with a clear and visible platform-initiated indicator of host quality.

Airbnb assesses whether a host has met the Superhost criteria once every three months. If a host meets the criteria, the host attains Superhost status and keeps that status until the next evaluation period. From the inception of this certification program (2009) until October 2018, Airbnb hosts must have met the following criteria over the previous 365 days to be awarded the Superhost status: (i) hosted at least 10 trips; (ii) maintained a 90% response rate when responding to guests requests; (iii) completed all confirmed reservations without cancellation; and (iv) received a 5-star review at least 80% of the time.

Airbnb uses specific “evaluation periods” to assess whether hosts have met the Superhost certification criteria. The three-month evaluation periods begin on January 1st, April 1st, July 1st, and October 1st. Each host’s Superhost status is reevaluated at the beginning of the subsequent evaluation period, usually within 10 days of the start of the evaluation period (i.e., for a January 1st evaluation period, each host’s status is adjusted, if necessary, by January 10th).

The dataset includes eight Superhost evaluation periods: April (2017, 2018), July(2016, 2017), October(2016, 2017), and January(2017, 2018). Each observation represents one active listing (or property) in one of the eight evaluation periods.

For variables which didn’t appear below but exist in the dataset, you don’t need to worry about them.

Airbnb Host ID: A unique identifier for the host.

Airbnb Property ID: A unique identifier for the property.

City\_x: City in which the property is located.

superhost\_period\_all: The evaluation period of this sample.

Scraped Date: Date when the information was scraped. The date is in the evaluation period. For example, when superhost\_period\_all=7, you may observe Scraped Date= 2017-02-05 which means the evaluation period is from Jan 2017 to March 2017.

host\_is\_superhost\_in\_period: Indicator (binary) showing if the host was a superhost during this period.

prev\_host\_is\_superhost\_in\_period: Indicator showing if the host was a superhost in the previous/last period.

Superhost: It is the same as host\_is\_superhost\_in\_period.

prev\_host\_is\_superhost: It is the same as prev\_host\_is\_superhost\_in\_period.

superhost\_change: Indicates a change in superhost status.

superhost\_change\_lose\_superhost: Indicates if the host lost the superhost status.

superhost\_change\_gain\_superhost: Indicates if the host gained the superhost status.

rating\_ave\_pastYear: Average rating of the listing over the past year.

numReviews\_pastYear: Number of reviews for the host over the past year.

numCancel\_pastYear: Number of cancellations for the host over the past year.

num\_5\_star\_Rev\_pastYear: Number of 5-star reviews received in the past year.

prop\_5\_StarReviews\_pastYear: Proportion of reviews that were 5 stars in the past year.

prev\_rating\_ave\_pastYear: Average rating of the host over the past year for the previous period. For example, if current period is from Jan to Mar 2018, then the last period would be for Oct to Dec 2017 and prev\_rating\_ave\_pastYear is the average rating of the host from Oct 1 2016 to Sept 30 2017.

prev\_numReviews\_pastYear: Number of reviews over the past year for the previous period.

prev\_numCancel\_pastYear: Number of cancellations over the past year for the previous period.

prev\_num\_5\_star\_Rev\_pastYear: Number of 5-star reviews over the past year for the previous period.

prev\_prop\_5\_StarReviews\_pastYear: Proportion of 5-star reviews over the past year for the previous period.

numReservedDays\_pastYear: Number of days the listing was reserved over the past year.

numReserv\_pastYear: Number of reservations made for the listing over the past year. A custom may make one reservation for multiple days.

prev\_numReservedDays\_pastYear: Number of days the listing was reserved over the past year for the previous period.

prev\_numReserv\_pastYear: Number of reservations over the past year over the past year for the previous period.

available\_days: Number of days the listing is available for booking from the 2nd month of this evaluation period to the next 6 months. For example, April 2017 is a Superhost evaluation month, which means that Airbnb hosts will be evaluated on the aforementioned Superhost criteria and will either obtain, retain, or lose their Superhost certication status. Then the 6-month period is from May 1, 2017 and October 31, 2017.

available\_days\_aveListedPrice: Average listed price for available days.

booked\_days: Number of days the listing is booked in the 6-month period. For instance, for the evaluation period from April 2017 to June 2017, 'booked\_days' represents the number of travel dates between May 1, 2017, and October 31, 2017 (6 months), whose booking dates occurred in this evaluation period for this listing. More specifically, if a consumer made a reservation for Sept 2nd's listing on May 14th, then this data point will be added to the 'booked\_days' for this evaluation period. The 'booked\_days' represents the listing's performance in each evaluation period.

booked\_days\_avePrice: Average booking price for booked days.

prev\_available\_days: Previous period's available days.

prev\_available\_days\_aveListedPrice: Previous period's average listed price for available days.

prev\_available\_days: Number of days the listing was available for booking for the previous evaluation period.

prev\_available\_days\_aveListedPrice: Average listed price for available days for the previous evaluation period.

prev\_booked\_days: Number of days the listing was booked for the previous evaluation period.

prev\_booked\_days\_avePrice: Average booking price for booked days for the previous evaluation period.

Property Type: The type of property or listing, such as apartment, house, etc.

Listing Type: The type of listing, such as entire home, private room, shared room, etc.

Created Date: The date when the property listing was created.

Zipcode: The postal code or ZIP code of the property's location.

Bedrooms: The number of bedrooms in the property.

Bathrooms: The number of bathrooms in the property.

Neighborhood: The neighborhood or locality where the property is situated.

Max Guests: The maximum number of guests the property can accommodate.

Cleaning Fee (USD): The fee charged for cleaning services for the property, in US dollars.

Minimum Stay: The minimum number of nights a guest must book to stay in the property.

Number of Photos: The total number of photos associated with the property listing.

Latitude: The geographical latitude coordinate of the property's location.

Longitude: The geographical longitude coordinate of the property's location.

Pets Allowed: An indicator (binary) showing whether pets are allowed in the property.

Instantbook Enabled: An indicator (binary) showing whether instant booking is enabled for the property.

prev\_Instantbook Enabled: An indicator (binary) showing whether instant booking was enabled for the property for the previous evaluation period.

Nightly Rate: The nightly rate for booking the property.

prev\_Nightly Rate: The nightly rate for booking the property in the previous evaluation period.

Number of Reviews: The total number of reviews for the property.

prev\_Number of Reviews: The total number of reviews for the property in the previous evaluation period.

Rating Overall: The overall rating of the property.

prev\_Rating Overall: The overall rating of the property in the previous evaluation period.

revenue: The revenue generated from the property.

occupancy\_rate: The occupancy rate of the property.

prev\_revenue: The revenue generated from the property in the previous evaluation period.

prev\_occupancy\_rate: The occupancy rate of the property in the previous evaluation period.

census\_tract: The census tract associated with the property's location.

tract\_total\_pop: The total population in the census tract.

tract\_white\_perc: The percentage of White population in the census tract.

tract\_black\_perc: The percentage of Black population in the census tract.

tract\_asian\_perc: The percentage of Asian population in the census tract.

tract\_housing\_units: The total number of housing units in the census tract.

zip\_total\_population: The total population in the ZIP code area.

zip\_hispanic\_or\_latino\_anyrace: The percentage of Hispanic or Latino population in the ZIP code area.

zip\_hispanic\_or\_latino\_anyrace\_percent: The percentage of Hispanic or Latino population in the ZIP code area.

zip\_white\_nothispanic: The percentage of non-Hispanic White population in the ZIP code area.

zip\_white\_nothispanic\_percent: The percentage of non-Hispanic White population in the ZIP code area.

zip\_black\_nothispanic: The percentage of non-Hispanic Black population in the ZIP code area.

zip\_black\_nothispanic\_percent: The percentage of non-Hispanic Black population in the ZIP code area.

zip\_asian\_nothispanic: The percentage of non-Hispanic Asian population in the ZIP code area.

zip\_asian\_nothispanic\_percent: The percentage of non-Hispanic Asian population in the ZIP code area.

tract\_count\_obs: The count of observations or data points in the census tract.

tract\_unique\_prices: The number of unique prices observed in the census tract.

Nightly Rate\_tractQuartile: The quartile of nightly rates within the census tract.

prev\_Nightly Rate\_tractQuartile: The quartile of nightly rates within the census tract in the previous evaluation period.

available\_days\_aveListedPrice\_tractQuartile: The quartile of average listed prices for available days within the census tract.

prev\_available\_days\_aveListedPrice\_tractQuartile: The quartile of average listed prices for available days within the census tract in the previous evaluation period.

tract\_superhosts: The count of superhosts in the census tract.

tract\_superhosts\_ratio: The ratio of superhosts to total hosts in the census tract.

tract\_prev\_superhosts: The count of superhosts in the census tract in the previous evaluation period.

tract\_prev\_superhosts\_ratio: The ratio of superhosts to total hosts in the census tract in the previous evaluation period.

tract\_price\_variance: The variance in prices within the census tract.

tractQuartilePrice\_variance: The variance in prices within the census tract quartiles.

prev\_host\_is\_superhost1: An additional field indicating the previous superhost status of the host.

prev\_year\_superhosts: The superhost status of the host in the previous year.

booked\_days\_period\_city: Number of days the listing was booked in the city during the evaluation period.

revenue\_period\_city: Revenue generated from the listing in the city during the evaluation period.

booked\_days\_period\_tract: Number of days the listing was booked in the census tract during the evaluation period.

revenue\_period\_tract: Revenue generated from the listing in the census tract during the evaluation period.

tract\_booking\_share: The share of bookings in the census tract.

tract\_revenue\_share: The share of revenue generated in the census tract.