

Airbnb Data

Source: <http://tomslee.net/airbnb-data-collection-get-the-data>

Variables:

- **room_id:** A unique number identifying an Airbnb listing. The listing has a URL on the Airbnb web site of http://airbnb.com/rooms/room_id
- **host_id:** A unique number identifying an Airbnb host. The host's page has a URL on the Airbnb web site of http://airbnb.com/users/show/host_id
- **room_type:** One of "Entire home/apt", "Private room", or "Shared room"
- **borough:** A subregion of the city or search area for which the survey is carried out. The borough is taken from a shapefile of the city that is obtained independently of the Airbnb web site. For some cities, there is no borough information; for others the borough may be a number. If you have better shapefiles for a city of interest, please send them to me.
- **neighborhood:** As with borough: a subregion of the city or search area for which the survey is carried out. For cities that have both, a neighbourhood is smaller than a borough. For some cities there is no neighbourhood information.
- **reviews:** The number of reviews that a listing has received. Airbnb has said that 70% of visits end up with a review, so the number of reviews can be used to estimate the number of visits. Note that such an estimate will not be reliable for an individual listing (especially as reviews occasionally vanish from the site), but over a city as a whole it should be a useful metric of traffic.
- **overall_satisfaction:** The average rating (out of five) that the listing has received from those visitors who left a review.
- **accommodates:** The number of guests a listing can accommodate.
- **bedrooms:** The number of bedrooms a listing offers.
- **price:** The price (in \$US) for a night stay. In early surveys, there may be some values that were recorded by month.
- **minstay:** The minimum stay for a visit, as posted by the host.
- **latitude and longitude:** The latitude and longitude of the listing as posted on the Airbnb site: this may be off by a few hundred metres. I do not have a way to track individual listing locations with

- `last_modified`: the date and time that the values were read from the Airbnb web site.

The data files for Sydney have been uploaded to the website corresponding the following survey dates:

Survey dates: 2014-09-26 (5567 listings), 2015-01-20 (7424 listings), 2015-03-20 (7871 listings), 2016-01-29 (13134 listings), 2016-06-11 (17343 listings), 2017-01-25 (23859 listings), 2017-03-12 (23297 listings), 2017-04-14 (24006 listings), 2017-05-17 (24419 listings), 2017-06-28 (25266 listings), 2017-07-23 (25506 listings).

You may like to use these files or another city, or data from several cities. Each csv file represents a single “survey” or “scrape” of the Airbnb web site for that city.