**Descriptive Analysis:**

The dataset chosen for analysis is for **Airbnb city data**. This dataset contains automated collection of data from the Airbnb web site for a specified city on or around a specific date .

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

**About Data:**

* The dataset gives a count of Airbnb listings in a city that is usually within 10% of the correct number. This level of accuracy provides a solid foundation for most policy and social impact discussions.
* The dataset includes latitude and longitude values for each listing. Street addresses of individual listings cannot be reliably inferred from these values.
* Price data is collected in $US. The value is usually per night and aggregated pricing information is therefore reliable
* Estimates of the proportions of listings and of revenue can be made, based using the number of reviews as a proxy for the relative number of visits, and the nightly price as a proxy for the relative income.

Below are the list of **attributes** that we have selected to carry out our analysis.

1.Host\_id

2.Room\_type

3.Country

4.City

5.Neighbourhood

6.Address

7.Reviews

8.Overall\_satisfaction

9.Accomodates

10.Bedrooms

11.Bathrooms

12.Price

13.Deleted

14.Minstay

15.last\_modified

16.Longitude

17.Latitude

18.Survey\_id

**Problems with the dataset:**

The data which is unavailable in our dataset is:

* occupancy rates: no occupancy data is collected.
* host income: without occupancy rates, host income is not available.
* specific addresses: exploration with available reverse geo coding databases suggests that the latitude and longitude values do not map reliably to specific addresses.
* guest information: no guest information is collected.

**Hypothesis to be explored with the data:**

This information can be used to estimate relative incomes and number of visits within a city. Absolute numbers of visits and incomes are not realistic, but relative numbers of visits would be reliable so long as the ratio of reviews to visits is the same within each group analyzed.







