

ML Group project:

Members:

- Mark Ditchburn
- Andrea Butera
- Kuno de Leeuw-Kent

Dataset:

The dataset selected is data on all listing activity and metrics in NYC, NY for Airbnb captured for the year 2019.

The dataset can be found [here](#), with the provider of the data [here](#).

Data Summary:

```
'data.frame':  48895 obs. of  16 variables:
 $ id                : int  2539 2595 3647 3831 5022 5099 5121 5178 5203 5238 ...
 $ name              : chr   "Clean & quiet apt home by the park" "Skylit Midtown Ca
 $ host_id           : int  2787 2845 4632 4869 7192 7322 7356 8967 7490 7549 ...
 $ host_name         : chr   "John" "Jennifer" "Elisabeth" "LisaRoxanne" ...
 $ neighbourhood_group : chr  "Brooklyn" "Manhattan" "Manhattan" "Brooklyn" ...
 $ neighbourhood     : chr  "Kensington" "Midtown" "Harlem" "Clinton Hill" ...
 $ latitude          : num  40.6 40.8 40.8 40.7 40.8 ...
 $ longitude         : num  -74 -74 -73.9 -74 -73.9 ...
 $ room_type         : chr   "Private room" "Entire home/apt" "Private room" "Entire
 $ price             : int  149 225 150 89 80 200 60 79 79 150 ...
 $ minimum_nights    : int   1 1 3 1 10 3 45 2 2 1 ...
 $ number_of_reviews  : int   9 45 0 270 9 74 49 430 118 160 ...
 $ last_review       : chr   "2018-10-19" "2019-05-21" "" "2019-07-05" ...
 $ reviews_per_month : num   0.21 0.38 NA 4.64 0.1 0.59 0.4 3.47 0.99 1.33 ...
 $ calculated_host_listings_count: int   6 2 1 1 1 1 1 1 1 4 ...
 $ availability_365   : int  365 355 365 194 0 129 0 220 0 188 ...
```

Data Visualisation:

