```
dtypes: float64(3), int64(7), object(6)
memory usage: 6.0+ MB
```

2 Start Tidying the Data

```
[8]: data.isnull().sum()
```

```
[8]: id                                   0
     name                                16
     host_id                              0
     host_name                           21
     neighbourhood_group                  0
     neighbourhood                        0
     latitude                             0
     longitude                            0
     room_type                            0
     price                                0
     minimum_nights                       0
     number_of_reviews                    0
     last_review                      10052
     reviews_per_month                10052
     calculated_host_listings_count       0
     availability_365                     0
     dtype: int64
```

It would make sense to drop any columns that involves reviews. The reviews do not tell us anything on the quality of the persons stay and there is far to many missing data points within the last_review and reviews_per_month columns.

```
[9]: to_drop = ['number_of_reviews','last_review','reviews_per_month']
```

```
[10]: data.drop(to_drop, inplace=True, axis=1)
```

```
[11]: data.head()
```

```
[11]:      id                                              name  host_id  \
      0  2539                  Clean & quiet apt home by the park     2787
      1  2595                            Skylit Midtown Castle     2845
      2  3647                THE VILLAGE OF HARLEM…NEW YORK !     4632
      3  3831                  Cozy Entire Floor of Brownstone     4869
      4  5022  Entire Apt: Spacious Studio/Loft by central park     7192

        host_name neighbourhood_group neighbourhood  latitude  longitude  \
      0      John              Brooklyn    Kensington  40.64749  -73.97237
      1  Jennifer             Manhattan       Midtown  40.75362  -73.98377
      2  Elisabeth            Manhattan        Harlem  40.80902  -73.94190
```