**Data description**

The original dataset is from Inside Airbnb of New York City (<http://insideairbnb.com>), which includes detailed listings data for New York City that compiled 07 September, 2020.

Link: (<http://data.insideairbnb.com/united-states/ny/new-york-city/2020-09-07/data/listings.csv.gz>)

The original dataset has 74 columns and 45756 observations. We manually screened out 32 columns that might be useful and saved as ‘listings\_new.csv’. Then, after data cleaning and NLP feature extraction, we got a data frame ‘df\_clean’ with 14,064 observations, which contains 30 features variables and 1 class variable. Price are divided into 5 categories and below is the data dictionary for features variables:

|  |  |
| --- | --- |
| **Column** | **Description** |
| host\_is\_superhost | ‘1’ for yes and ‘0’ for no |
| host\_has\_profile\_pic | ‘1’ for yes and ‘0’ for no |
| host\_identity\_verified | ‘1’ for verified and ‘0’ for not |
| neighbourhood\_group\_cleansed | 5 categories (0-4)  0 for Bronx  1 for Brooklyn  2 for Manhattan  3 for Queens  4 for Staten Island |
| latitude | numeric variable |
| longitude | numeric variable |
| room\_type | 4 categories (0-3)  0 for Entire home/apt  1 for Hotel room  2 for Private room  3 for Shared room |
| accommodates | 1-16 |
| bathroom\_share\_or\_not | ‘1’ for yes and ‘0’ for no |
| bedrooms | 1-6, 10, 21 |
| beds | 0-8, 10 |
| minimum\_nights | numeric variable |
| maximum\_nights | numeric variable |
| number\_of\_reviews | numeric variable |
| review\_scores\_rating | numeric variable |
| review\_scores\_accuracy | 9 categories (2-10) |
| review\_scores\_cleanliness | 9 categories (2-10) |
| review\_scores\_checkin | 9 categories (2-10) |
| review\_scores\_communication | 9 categories (2-10) |
| review\_scores\_location | 9 categories (2-10) |
| review\_scores\_value | 9 categories (2-10) |
| instant\_bookable | ‘1’ for yes and ‘0’ for no |
| reviews\_per\_month | numeric variable |
| bathrooms | 11 categories (0-10) |
| Pool | ‘1’ for yes and ‘0’ for no |
| BabyBath | ‘1’ for yes and ‘0’ for no |
| BabyMonitor | ‘1’ for yes and ‘0’ for no |
| LakeAccess | ‘1’ for yes and ‘0’ for no |
| BeachFront | ‘1’ for yes and ‘0’ for no |
| Piano | ‘1’ for yes and ‘0’ for no |