

Airbnb

December 29, 2020

1 AIR BnB

My goal for this project is to clean the data. The dataset is pulled from <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data> with the author being “Dgomonov”.

```
[1]: import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import pandas as pd
from scipy import stats
```

```
[2]: data = pd.read_csv('/Users/thomasoakley/Data_Science_Projects/Airbnb.csv')
```

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

```
[3]:
```

	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	
3	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	
4	Laura	Manhattan	East Harlem	40.79851	-73.94399	

	room_type	price	minimum_nights	number_of_reviews	last_review	\
0	Private room	149	1	9	2018-10-19	
1	Entire home/apt	225	1	45	2019-05-21	
2	Private room	150	3	0	NaN	
3	Entire home/apt	89	1	270	2019-07-05	
4	Entire home/apt	80	10	9	2018-11-19	