

```
pip install pymongo
```

```
Requirement already satisfied: pymongo in /usr/local/lib/python3.10/dist-packages (4.5.0)
Requirement already satisfied: dnspython<3.0.0,>=1.16.0 in /usr/local/lib/python3.10/dist-packages (from pymongo) (2.4.2)
```

```
import pymongo
import numpy as np
```

```
client = pymongo.MongoClient('mongodbsrv://iam_hema:hema1atha@hemz.fo3int4.mongodb.net/?retryWrites=true&w=majority')
```

```
db = client['sample_airbnb'] #use abisha # DB is an object created based on client that refers database called abisha
```

```
col = db['listingsAndReviews']
```

```
result = []
for i in col.find({},{'_id','name','description','neighborhood_overview','price','security_deposit','cleaning_fee','extra_people','guests_included','property_type','room_type','bed_type','cancellation_policy','number_of_reviews','review_scores','accommodates','bedrooms','beds','availability','amenities','location','address','host'}):
    result.append(i)
```

```
import pandas as pd
df = pd.DataFrame(result)
```

```
df[['availability_30','availability_60','availability_90','availability_365']] = df['availability'].apply(lambda i:pd.Series([i['availability_30'],i['availability_60'],i['availability_90'],i['availability_365']])))
df[['host_id','host_name','host_response_time']] = df['host'].apply(lambda i:pd.Series([i['host_id'],i['host_name'],i.get('host_response_time',0)])))
df[['street','country','country_code','location']] = df['address'].apply(lambda i:pd.Series([i['street'],i['country'],i['country_code'],i['location']])))
df[['location_type','coordinates','is_location_exact']] = df['location'].apply(lambda i:pd.Series([i['type'],i['coordinates'],i['is_location_exact']])))
df[['latitude','longitude']] = df['coordinates'].apply(lambda i:pd.Series([i[0],i[1]])))
df['average_rating'] = df['review_scores'].apply(lambda i: np.mean([i.get('review_scores_accuracy', 0), i.get('review_scores_cleanliness', 0)]))
```

```
df.drop('availability',axis = 1,inplace = True)
df.drop('review_scores',axis = 1,inplace = True)
df.drop('host',axis = 1,inplace = True)
df.drop('address',axis = 1,inplace = True)
df.drop('location',axis = 1,inplace = True)
df.drop('coordinates',axis = 1, inplace = True)
```

Double-click (or enter) to edit

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5555 entries, 0 to 5554
Data columns (total 34 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   _id                                    5555 non-null   object
1   name                                  5555 non-null   object
2   description                           5555 non-null   object
3   neighborhood_overview                 5555 non-null   object
4   property_type                         5555 non-null   object
5   room_type                            5555 non-null   object
6   bed_type                             5555 non-null   object
7   cancellation_policy                   5555 non-null   object
8   accommodates                          5555 non-null   int64
9   bedrooms                             5550 non-null   float64
10  beds                                 5542 non-null   float64
11  number_of_reviews                     5555 non-null   int64
12  amenities                             5555 non-null   object
13  price                                 5555 non-null   object
14  security_deposit                       3471 non-null   object
15  cleaning_fee                           4024 non-null   object
16  extra_people                           5555 non-null   object
17  guests_included                       5555 non-null   object
18  availability_30                        5555 non-null   int64
19  availability_60                        5555 non-null   int64
20  availability_90                        5555 non-null   int64
21  availability_365                       5555 non-null   int64
```

```

22 host_id          5555 non-null object
23 host_name        5555 non-null object
24 host_response_time 4167 non-null object
25 street           5555 non-null object
26 country          5555 non-null object
27 country_code     5555 non-null object
28 location_type    5555 non-null object
29 coordinates      5555 non-null object
30 is_location_exact 5555 non-null bool
31 latitude         5555 non-null float64
32 longitude        5555 non-null float64
33 average_rating   5555 non-null float64
dtypes: bool(1), float64(5), int64(6), object(22)
memory usage: 1.4+ MB

```

```
df.isnull().sum()
```

```

_id              0
name             0
description      0
neighborhood_overview 0
property_type    0
room_type        0
bed_type         0
cancellation_policy 0
accommodates     0
bedrooms         5
beds            13
number_of_reviews 0
amenities        0
price           0
security_deposit 2084
cleaning_fee     1531
extra_people     0
guests_included  0
availability_30  0
availability_60  0
availability_90  0
availability_365 0
host_id          0
host_name        0
host_response_time 1388
street           0
country          0
country_code     0
location_type    0
coordinates      0
is_location_exact 0
latitude         0
longitude        0
average_rating   0
dtype: int64

```

```
df.shape
```

```
(5555, 34)
```

```
df.size
```

```
188870
```

```

df['bedrooms'].fillna(0, inplace=True)
df['beds'].fillna(0, inplace=True)
df['cleaning_fee'].fillna('Not Specified', inplace=True)
df['security_deposit'].fillna('Not Specified',inplace = True)
df['host_response_time'].fillna('Not Specified',inplace = True)
df['is_location_exact'] = df['is_location_exact'].map({False:'No',True:'Yes'})
df['average_rating']=df['average_rating'].replace(0,'Unrated')

```

```
df.head()
```

	_id	name	description	neighborhood_overview	property_type	room_type	bed_type	cancellation_policy	accommodates	b
0	10006546	Ribeira Charming Duplex	Fantastic duplex apartment with three bedrooms...	In the neighborhood of the river, you can find...	House	Entire home/apt	Real Bed	moderate		8
1	1003530	New York City - Upper West Side Apt	Murphy bed, optional second bedroom available....	Great neighborhood - many terrific restaurants...	Apartment	Private room	Real Bed	strict_14_with_grace_period		2
2	10215858	Cozy Queen Guest Room&My	Mong Kok MTR E1 walk 100 meters, shopping, eat...	Electric Street, Mong Kok near there (Sai Yeun...	Apartment	Private room	Real Bed	strict_14_with_grace_period		2
3	10112159	Downtown Oporto Inn (room cleaning)	Tradicional building, with high ceilings next ...	Exciting, urban and dinamic, stay with us, nea...	Hostel	Private room	Real Bed	moderate		2
The										
df.to_csv('airbnb.csv', index=False)										
		Posto 6	room, toilet, bathro	zone is the district t...		home/apt				

Could not connect to the reCAPTCHA service. Please check your internet connection and reload to get a reCAPTCHA challenge.