# Airbnb price prediction

#### Out[164]:

	id	log_price	property_type	room_type	amenities	accommodates	bathrooms	bed_type
0	6901257	5.010635	Apartment	Entire home/apt	{"Wireless Internet","Air conditioning",Kitche	3	1.0	Real Be
1	6304928	5.129899	Apartment	Entire home/apt	{"Wireless Internet","Air conditioning",Kitche	7	1.0	Real Be
2	7919400	4.976734	Apartment	Entire home/apt	{TV,"Cable TV","Wireless Internet","Air condit	5	1.0	Real Be
3	13418779	6.620073	House	Entire home/apt	{TV,"Cable TV",Internet,"Wireless Internet",Ki	4	1.0	Real Be
4	3808709	4.744932	Apartment	Entire home/apt	{TV,Internet,"Wireless Internet","Air conditio	2	1.0	Real Be

5 rows × 29 columns

In [165]: ► data.shape

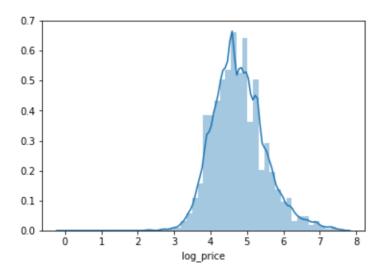
Out[165]: (74111, 29)

In [166]: | data.describe(include=['0'])

#### Out[166]:

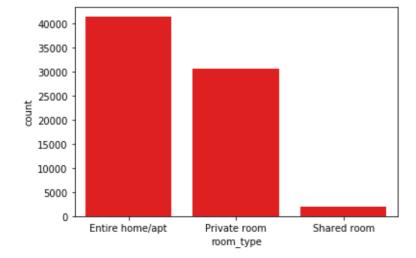
	property_type	room_type	amenities	bed_type	cancellation_policy	city	description	first_review
count	74111	74111	74111	74111	74111	74111	74111	58247
unique	35	3	67122	5	5	6	73479	2554
top	Apartment	Entire home/apt	8	Real Bed	strict	NYC	Hello, I've been running guest house for Korea	2017-01-01
freq	49003	41310	586	72028	32374	32349	8	293

C:\Users\user\Anaconda3\lib\site-packages\scipy\stats\stats.py:1713: FutureWarning: Usi
ng a non-tuple sequence for multidimensional indexing is deprecated; use `arr[tuple(se
q)]` instead of `arr[seq]`. In the future this will be interpreted as an array index, `
arr[np.array(seq)]`, which will result either in an error or a different result.
return np.add.reduce(sorted[indexer] \* weights, axis=axis) / sumval



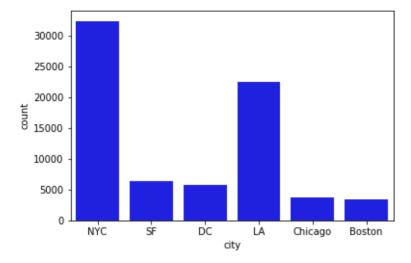
```
In [168]: N sns.countplot(x='room_type',color='red', data=data)
```

Out[168]: <matplotlib.axes.\_subplots.AxesSubplot at 0x202299b2f98>



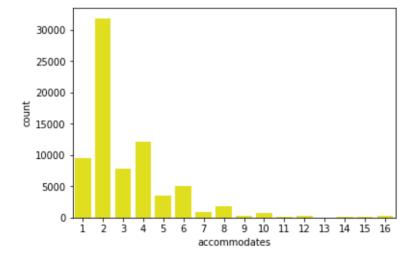
In [169]: N sns.countplot(x='city',color='blue', data=data)

Out[169]: <matplotlib.axes.\_subplots.AxesSubplot at 0x202299faeb8>



In [170]: ▶ sns.countplot(x='accommodates',color='yellow', data=data)

Out[170]: <matplotlib.axes.\_subplots.AxesSubplot at 0x2022a21d0f0>



Out[171]: <matplotlib.axes.\_subplots.AxesSubplot at 0x2022a2292e8>

id -	. 1	-0.0072	-0.0058	-0.004	-0.0031	-0.003	-0.0014	0.0029	0.00031	-0.0075	-0.0087
log_price -	-0.0072	1	0.57	0.36	0.11	-0.0022	-0.048	-0.032	0.091	0.47	0.44
accommodates -	-0.0058	0.57	1	0.51	0.18	-0.078	-0.085	0.042	-0.017	0.71	0.81
bathrooms -	-0.004	0.36	0.51	1	0.054	-0.14	-0.14	-0.041	0.011	0.59	0.53
deaning_fee <sup>-</sup>	-0.0031	0.11	0.18	0.054	1	-0.061	-0.069	0.11	0.035	0.11	0.13
latitude -	-0.003	-0.0022	-0.078	-0.14	-0.061	1	0.9	-0.018	-0.033	-0.059	-0.08
longitude -	-0.0014	-0.048	-0.085	-0.14	-0.069	0.9	1	-0.05	-0.049	-0.076	-0.082
number_of_reviews <sup>-</sup>	0.0029	-0.032	0.042	-0.041	0.11	-0.018	-0.05	1	0.012	-0.035	0.031
review_scores_rating -	0.00031	0.091	-0.017	0.011	0.035	-0.033	-0.049	0.012	1	0.011	-0.027
bedrooms -	-0.0075	0.47	0.71	0.59	0.11	-0.059	-0.076	-0.035	0.011	1	0.71
beds -	-0.0087	0.44	0.81	0.53	0.13	-0.08	-0.082	0.031	-0.027	0.71	1
	- þi	log_price	accommodates -	bathrooms -	deaning_fee -	latitude -	longitude –	number of reviews -	review_scores_rating -	bedrooms -	- Spaq

```
In [173]:
In [174]:
               for col in filedata:
                    if (filedata[col].isnull().any()):
                        print(col)
               first_review
               host_has_profile_pic
               host_identity_verified
               host_response_rate
               host_since
               last review
               neighbourhood
               thumbnail_url
               zipcode
In [175]:
               for col in filedata:
                    if (filedata[col].isnull().any()):
                        filedata=filedata.drop([col],axis=1)
               filedata
    Out[175]:
                       log_price
                                property_type
                                                                         amenities accommodates bed_type ca
                                            room_type
                                                 Entire
                                                               {"Wireless Internet","Air
                      5.010635
                                                                                                 Real Bed
                                   Apartment
                                                                conditioning", Kitche...
                                               home/apt
                                                 Entire
                                                               {"Wireless Internet","Air
                                                                                                 Real Bed
                       5.129899
                                   Apartment
                                                                conditioning", Kitche...
                                               home/apt
                                                 Entire
                                                             {TV,"Cable TV","Wireless
                       4.976734
                                                                                                 Real Bed
                                    Apartment
                                               home/apt
                                                                 Internet","Air condit...
                                                                        {TV,"Cable
                                                 Entire
                      6.620073
                                      House
                                                                TV",Internet,"Wireless
                                                                                                 Real Bed
                                               home/apt
                                                                       Internet",Ki...
                                                 Entire
                                                                {TV,Internet,"Wireless
                       4.744932
                                   Apartment
                                                                                                Real Bed
                                                               Internet","Air conditio...
                                               home/apt
In [176]:

▶ | filedata['bed_type'].value_counts()
   Out[176]: Real Bed
                                  72028
               Futon
                                    753
               Pull-out Sofa
                                    585
               Airbed
                                    477
               Couch
                                    268
               Name: bed_type, dtype: int64
               filedata=filedata.drop(['property_type','amenities','description'], axis=1)
In [177]:
            ▶ def replace_with_ohe(df, cat_columns):
In [178]:
                    for col in cat_columns:
                        df[col] = pd.Categorical(df[col])
                        dum = pd.get_dummies(df[col], prefix=col)
                        df=pd.concat([df, dum], axis=1)
                    df = df.drop(cat_columns, axis=1)
                    return df
```

### Out[179]:

r	room_type_Entire home/apt	name	instant_bookable	cleaning_fee	accommodates	log_price	
	1	Beautiful brownstone 1- bedroom	f	True	3	5.010635	0
	1	Superb 3BR Apt Located Near Times Square	t	True	7	5.129899	1
	1	The Garden Oasis	t	True	5	4.976734	2
	1	Beautiful Flat in the Heart of SF!	f	True	4	6.620073	3
	1	Great studio in midtown DC	t	True	2	4.744932	4
	0	Comfort Suite San Francisco	t	True	2	4.442651	5

Out[180]: <matplotlib.axes.\_subplots.AxesSubplot at 0x2022a2c94a8>

log_price -	1	0.57	0.11	0.6	-0.53	-0.22	-0.05	-0.04	-0.065	-0.039	0.099	-0.096	-0.05	0.13	0.039	0.037	0.031	-0.052	0.082	-0.057	-0.077	0.17
accommodates -	0.57	1	0.18	0.48	-0.45	-0.089	-0.035	-0.031	-0.055	-0.041	0.084	-0.17	-0.042	0.19	0.0054	0.032	0.015	0.07	0.052	0.077	-0.14	0.0021
deaning_fee -	0.11	0.18	1	0.21	-0.18	-0.083	-0.024	-0.017	-0.015	-0.0026	0.028	-0.35	0.09	0.25	0.0061	0.0051	0.015	0.0089	-0.048	0.069	-0.053	0.0075
room type Entire home/apt -	0.6	0.48	0.21	1	-0.94	-0.19	-0.056	-0.044	-0.068	-0.035	0.1	-0.16	-0.0035	0.15	0.035	0.013	0.027	0.018	0.073	0.029	-0.098	0.022
room type Private room -	-0.53	-0.45	-0.18	-0.94	1	-0.15	0.027	-0.022	0.048	0.0047	-0.036	0.15	0.014	-0.15	-0.033	-0.013	-0.02	-0.017	-0.07	-0.048	0.11	-0.014
room type Shared room -	-0.22	-0.089	-0.083	-0.19	-0.15	1	0.088	0.19	0.06	0.089	-0.2	0.052	-0.03	-0.021	-0.0067	-0.0026	-0.021	-0.0028	-0.0087	0.056	-0.022	-0.026
bed type Airbed	-0.05	-0.035	-0.024	-0.056	0.027	0.088	1	-0.0048	-0.0082	-0.0072	-0.47	0.023	0.0017	-0.022	-0.0031	-0.0012	0.0069	0.015	0.00088	-0.0013	-0.0079	
bed type Couch -	-0.04	-0.031	-0.017	-0.044	-0.022	0.19	-0.0048		-0.0061		-0.35	0.025	-0.0077					-0.0015		0.016		0.00021
bed type Futon -	-0.065	-0.055	-0.015	-0.068	0.048	0.06	-0.0082		1	-0.009	-0.6	0.0073	0.016					0.00013				-0.0054
	-0.039	-0.041	-0.0026		0.0047	0.089		-0.0054	-0.009	1	-0.52	0.00068	0.018					-0.0086		-0.01		0.00012
bed_type_Pull-out Sofa -		0.084											-0.017	0.038								
bed_type_Real Bed -	0.099		0.028	0.1	-0.036	-0.2	-0.47	-0.35	-0.6	-0.52	1	-0.025				0.0026			0.00065			
cancellation_policy_flexible -	-0.096	-0.17	-0.35	-0.16	0.15	0.052	0.023	0.025		0.00068		1	-0.39	-0.58	-0.026	-0.01	-0.029	-0.024		-0.0024		-0.0092
cancellation_policy_moderate -	-0.05	-0.042	0.09	-0.0035	0.014	-0.03	0.0017	-0.0077	0.016	0.018	-0.017	-0.39	1	-0.52	-0.023	-0.0089	-0.011	0.027	0.012	-0.0012	-0.029	0.029
cancellation_policy_strict -	0.13	0.19	0.25	0.15	-0.15	-0.021	-0.022	-0.016	-0.021	-0.016	0.038	-0.58	-0.52	1	-0.034	-0.013	0.033	-0.0019	-0.05	0.0049	0.019	-0.017
cancellation_policy_super_strict_30 -	0.039	0.0054	0.0061	0.035	-0.033	-0.0067	-0.0031	-0.0023	-0.0039	-0.0035	0.0066	-0.026	-0.023	-0.034	1	0.00059	0.046	0.00099	0.059	-0.025	-0.028	0.00034
cancellation_policy_super_strict_60 -	0.037	0.032	0.0051	0.013	-0.013	-0.0026	-0.0012	0.00091	-0.0015	-0.0014	0.0026	-0.01	-0.0089	-0.013	0.00059	1	0.0051	0.0006	-0.0044	0.011	-0.012	0.0017
city_Boston -	0.031	0.015	0.015	0.027	-0.02	-0.021	0.0069	-0.008	-0.0046	-0.0075	0.0064	-0.029	-0.011	0.033	0.046	0.0051	1	-0.051	-0.064	-0.15	-0.2	-0.068
city_Chicago -	-0.052	0.07	0.0089	0.018	-0.017	-0.0028	0.015	-0.0015	0.00013	-0.0086	-0.002	-0.024	0.027	-0.0019	0.00099	0.0006	-0.051	1	-0.066	-0.15	-0.2	-0.071
city_DC -	0.082	0.052	-0.048	0.073	-0.07	-0.0087	0.00088	0.0021	-0.0085	0.0087	0.00065	0.037	0.012	-0.05	0.059	-0.0044	-0.064	-0.066	1	-0.19	-0.25	-0.089
city_LA -	-0.057	0.077	0.069	0.029	-0.048	0.056	-0.0013	0.016	0.00062	-0.01	0.00055	-0.0024	-0.0012	0.0049	-0.025	0.011	-0.15	-0.15	-0.19	1	-0.58	-0.2
city_NYC -	-0.077	-0.14	-0.053	-0.098	0.11	-0.022	-0.0079	-0.012	0.01	0.012	-0.0042	0.01	-0.029	0.019	-0.028	-0.012	-0.2	-0.2	-0.25	-0.58	1	-0.27
city_SF -	0.17	0.0021	0.0075	0.022	-0.014	-0.026	-0.0014	0.00021	-0.0054	0.00012	0.004	-0.0092	0.029	-0.017	0.00034	0.0017	-0.068	-0.071	-0.089	-0.2	-0.27	1
	log_price -	accommodates -	deaning_fee -	oom_type_Entire home/apt -	room_type_Private room -	room_type_Shared room -	bed_type_Airbed -	bed_type_Couch -	bed_type_Futon -	bed_type_Pull-out Sofa -	bed_type_Real Bed -	ancellation_policy_flexible -	rellation_policy_moderate -	cancellation_policy_strict -	ion policy super strict 30 -	ion policy super strict 60 -	aty_Boston -	city_Chicago -	- aty_DC -	aty_LA -	aty_NYC -	dity_SF -

```
filedata['instant_bookable']
    Out[181]: 0
                         0
                         1
               2
                         1
               3
                         0
               4
                         1
               5
                         1
               6
                         1
               7
                         0
               8
                         0
               9
                         1
               10
                         0
                         0
               11
               12
                         0
               13
                         1
               14
                         0
               15
                         0
                         0
               16
                         0
               17
               18
                         0
In [182]:
               X = filedata.iloc[:,1:].drop(['name'], axis=1)
               Y = filedata['log_price']
In [183]:
            X
In [184]:
    Out[184]:
                                                                 room_type_Entire room_type_Private
                                                                                                  room_type_
                       accommodates cleaning_fee instant_bookable
                                                                        home/apt
```

filedata['instant\_bookable'] = (filedata['instant\_bookable'] == 't').astype(int)

In [181]:

True

```
Out[185]:
                          5.010635
                          5.129899
                1
                2
                          4.976734
                3
                          6.620073
                4
                          4.744932
                5
                          4.442651
                6
                          4.418841
                7
                          4.787492
                8
                          4.787492
                9
                          3.583519
                10
                          4.605170
                11
                          5.010635
                12
                          4.248495
                13
                          5.298317
                14
                          4.955827
                15
                          4.094345
                16
                          4.317488
                17
                          4.595120
                18
                          4.882802
In [186]:
                from sklearn.model_selection import train_test_split
                X_train, X_test, Y_train, Y_test = train_test_split(X, Y)
In [187]:
In [188]:
                X_train
    Out[188]:
                                                                    room_type_Entire
                                                                                     room_type_Private
                                                                                                       room_type_
                        accommodates cleaning_fee instant_bookable
                                                                           home/apt
                                                                                                 room
                 57252
                                    2
                                              True
                                                                 0
                                                                                  0
                                                                                                    1
                 18035
                                    4
                                              True
                                                                 0
                                                                                  1
                                                                                                    0
                 24338
                                    2
                                              True
                                                                                  0
                                                                                                    1
                 44248
                                    2
                                             False
                                                                 1
                                                                                  1
                                                                                                    0
                 51230
                                    2
                                                                 0
                                                                                  0
                                              True
                                                                                                    1
                 60244
                                    2
                                              True
                                                                                  0
                                                                                                    1
                 36415
                                    2
                                              True
                                                                 1
                                                                                  0
                                                                                                    1
                                    3
                                                                                                    0
                 15248
                                              True
                                                                 1
                                                                                  1
```

# **Linear Regression**

31263

52723

In [185]:

0

0

1

1

0

0

## **Random forest regression**

3

2

True

True

C:\Users\user\Anaconda3\lib\site-packages\sklearn\ensemble\forest.py:246: FutureWarnin
g: The default value of n\_estimators will change from 10 in version 0.20 to 100 in 0.2
2.

"10 in version 0.20 to 100 in 0.22.", FutureWarning)

accuracy= 0.532491983017879