

Aulya Yarzuki K.

Exploration Data Analysis AirBnb

Deskripsi Dataset

Sejak 2008, para tamu dan tuan rumah telah menggunakan Airbnb untuk bepergian dan menghadirkan tempat tinggal sementara. Kumpulan data ini mendeskripsikan aktivitas dan metriks pada AirBnb di New York City tahun 2019.

Gambaran Besar dari Dataset

Dataset ini terdiri dari 48895 baris dan 16 kolom. 16 kolom tersebut terdiri dari:

- `id`, `'name'`
- `'Host_id'`, `'Host_name'`
- `'Neighbourhood_group'`, `'neighbourhood'`,
- `'latitude'`, `'longitude'`,
- `'room_type'`, `'price'`,
- `'Minimum_nights'`, `'number_of_reviews'`,
- `'last_review'`, `'reviews_per_month'`,
- `'calculated_host_listings_count'`, `'availability_365'`

```
data.head()
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	r
0	2539	Clean & quiet apt home by the park	2787	John	Brooklyn	Kensington	40.64749	-73.97237	Private room	149		1
1	2595	Skylit Midtown Castle	2845	Jennifer	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225		1
2	3647	THE VILLAGE OF HARLEM....NEW YORK !	4632	Elisabeth	Manhattan	Harlem	40.80902	-73.94190	Private room	150		3
3	3831	Cozy Entire Floor of Brownstone	4869	LisaRoxanne	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89		1
4	5022	Entire Apt: Spacious Studio/Loft by central park	7192	Laura	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80		10

number_of_reviews	last_review	reviews_per_month	calculated_host_listings_count	availability_365
9	2018-10-19	0.21	6	365
45	2019-05-21	0.38	2	355
0	NaN	NaN	1	365
270	2019-07-05	4.64	1	194
9	2018-11-19	0.10	1	0

Tipe Data pada Setiap Kolom

```
In [100]: data.dtypes
```

```
Out[100]: id                int64  
          name              object  
          host_id           int64  
          host_name         object  
          neighbourhood_group object  
          neighbourhood      object  
          latitude          float64  
          longitude         float64  
          room_type         object  
          price             int64  
          minimum_nights    int64  
          number_of_reviews int64  
          last_review        object  
          reviews_per_month float64  
          calculated_host_listings_count int64  
          availability_365   int64  
          dtype: object
```

Missing Value pada Data

```
In [104]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 48895 entries, 0 to 48894
Data columns (total 16 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   id                                     48895 non-null  int64
1   name                                  48879 non-null  object
2   host_id                               48895 non-null  int64
3   host_name                             48874 non-null  object
4   neighbourhood_group                   48895 non-null  object
5   neighbourhood                         48895 non-null  object
6   latitude                             48895 non-null  float64
7   longitude                             48895 non-null  float64
8   room_type                             48895 non-null  object
9   price                                 48895 non-null  int64
10  minimum_nights                        48895 non-null  int64
11  number_of_reviews                     48895 non-null  int64
12  last_review                           38843 non-null  object
13  reviews_per_month                     38843 non-null  float64
14  calculated_host_listings_count         48895 non-null  int64
15  availability_365                       48895 non-null  int64
dtypes: float64(3), int64(7), object(6)
memory usage: 6.0+ MB
```

```
In [166]: data.isna().sum()
```

```
Out[166]: 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
```

Menampilkan Statistik Dasar dari Data (Mean, Median, Std, dll)

```
data.describe(include='all')
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price
count	4.889500e+04	48879	4.889500e+04	48874	48895	48895	48895.000000	48895.000000	48895	48895.000000
unique	NaN	47905	NaN	11452	5	221	NaN	NaN	3	NaN
top	NaN	Hillside Hotel	NaN	Michael	Manhattan	Williamsburg	NaN	NaN	Entire home/apt	NaN
freq	NaN	18	NaN	417	21661	3920	NaN	NaN	25409	NaN
mean	1.901714e+07	NaN	6.762001e+07	NaN	NaN	NaN	40.728949	-73.952170	NaN	152.720687
std	1.098311e+07	NaN	7.861097e+07	NaN	NaN	NaN	0.054530	0.046157	NaN	240.154170
min	2.539000e+03	NaN	2.438000e+03	NaN	NaN	NaN	40.499790	-74.244420	NaN	0.000000
25%	9.471945e+06	NaN	7.822033e+06	NaN	NaN	NaN	40.690100	-73.983070	NaN	69.000000
50%	1.967728e+07	NaN	3.079382e+07	NaN	NaN	NaN	40.723070	-73.955680	NaN	106.000000
75%	2.915218e+07	NaN	1.074344e+08	NaN	NaN	NaN	40.763115	-73.936275	NaN	175.000000
max	3.648724e+07	NaN	2.743213e+08	NaN	NaN	NaN	40.913060	-73.712990	NaN	10000.000000

minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_listings_count	availability_365
48895.000000	48895.000000	38843	38843.000000	48895.000000	48895.000000
NaN	NaN	1764	NaN	NaN	NaN
NaN	NaN	2019-06-23	NaN	NaN	NaN
NaN	NaN	1413	NaN	NaN	NaN
7.029962	23.274466	NaN	1.373221	7.143982	112.781327
20.510550	44.550582	NaN	1.680442	32.952519	131.622289
1.000000	0.000000	NaN	0.010000	1.000000	0.000000
1.000000	1.000000	NaN	0.190000	1.000000	0.000000
3.000000	5.000000	NaN	0.720000	1.000000	45.000000
5.000000	24.000000	NaN	2.020000	2.000000	227.000000
1250.000000	629.000000	NaN	58.500000	327.000000	365.000000

Data Harga Terendah

```
In [105]: min_price = data.price.min()  
min_price
```

```
Out[105]: 0
```

```
In [106]: min_neighbourhood = data.loc[data["price"] < 1]  
min_neighbourhood
```

```
Out[106]:
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	n
23161	18750597	Huge Brooklyn Brownstone Living, Close to it all.	8993084	Kimberly	Brooklyn	Bedford-Stuyvesant	40.69023	-73.95428	Private room	0		4
25433	20333471	★Hostel Style Room Ideal Traveling Buddies★	131697576	Anisha	Bronx	East Morrisania	40.83296	-73.88668	Private room	0		2
25634	20523843	MARTIAL LOFT 3: REDEMPTION (upstairs, 2nd room)	15787004	Martial Loft	Brooklyn	Bushwick	40.69467	-73.92433	Private room	0		2
25753	20608117	Sunny, Quiet Room in Greenpoint	1641537	Lauren	Brooklyn	Greenpoint	40.72462	-73.94072	Private room	0		2

Data Harga Tertinggi

```
In [107]: max_price = data.price.max()
max_price
```

```
Out[107]: 10000
```

```
In [108]: max_neighbourhood = data.loc[data["price"] > 9999]
max_neighbourhood
```

```
Out[108]:
```

	id	name	host_id	host_name	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	n
9151	7003697	Furnished room in Astoria apartment	20582832	Kathrine	Queens	Astoria	40.76810	-73.91651	Private room	10000		100
17692	13894339	Luxury 1 bedroom apt. - stunning Manhattan views	5143901	Erin	Brooklyn	Greenpoint	40.73260	-73.95739	Entire home/apt	10000		5
29238	22436899	1-BR Lincoln Center	72390391	Jelena	Manhattan	Upper West Side	40.77213	-73.98665	Entire home/apt	10000		30

Menghilangkan Kolom “name” dan “host_name”

```
airbnb = data.loc[:, ['id', 'host_id', 'neighbourhood_group',  
                    'neighbourhood', 'latitude', 'longitude', 'room_type', 'price',  
                    'minimum_nights', 'number_of_reviews', 'last_review',  
                    'reviews_per_month', 'calculated_host_listings_count',  
                    'availability_365']]  
airbnb
```

	id	host_id	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review
0	2539	2787	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	2018-10-19
1	2595	2845	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21
2	3647	4632	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	NaN
3	3831	4869	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05
4	5022	7192	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	2018-11-19
...
48890	36484665	8232441	Brooklyn	Bedford-Stuyvesant	40.67853	-73.94995	Private room	70	2	0	NaN
48891	36485057	6570630	Brooklyn	Bushwick	40.70184	-73.93317	Private room	40	4	0	NaN
48892	36485431	23492952	Manhattan	Harlem	40.81475	-73.94867	Entire home/apt	115	10	0	NaN

Isi *Missing Values* pada kolom “last_review” dengan Tanggal yang Paling Sering Muncul

```
In [113]: # Konversi data ke format penanggalan
airbnb["last_review"] = pd.to_datetime(airbnb["last_review"], format='%Y-%m-%d', errors = 'coerc
airbnb["last_review"]
```

```
Out[113]: 0      2018-10-19
1      2019-05-21
2           NaT
3      2019-07-05
4      2018-11-19
...
48890          NaT
48891          NaT
48892          NaT
48893          NaT
48894          NaT
Name: last_review, Length: 48895, dtype: datetime64[ns]
```

```
In [114]: airbnb.dtypes
```

```
Out[114]: id                int64
          host_id           int64
          neighbourhood_group object
          neighbourhood      object
          latitude           float64
          longitude          float64
          room_type          object
          price              int64
          minimum_nights     int64
          number_of_reviews  int64
          last_review        datetime64[ns]
          reviews_per_month  float64
          calculated_host_listings_count int64
          availability_365    int64
          dtype: object
```

```
In [115]: airbnb['last_review'] = airbnb['last_review'].fillna(airbnb['last_review'].mode()[0])
airbnb
```

Out[115]:

	id	host_id	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	re
0	2539	2787	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	2018-10-19	
1	2595	2845	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21	
2	3647	4632	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	2019-06-23	
3	3831	4869	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05	
4	5022	7192	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	2018-11-19	
...
48890	36484665	8232441	Brooklyn	Bedford-Stuyvesant	40.67853	-73.94995	Private room	70	2	0	2019-06-23	
48891	36485057	6570630	Brooklyn	Bushwick	40.70184	-73.93317	Private room	40	4	0	2019-06-23	
48892	36485431	23492952	Manhattan	Harlem	40.81475	-73.94867	Entire home/apt	115	10	0	2019-06-23	
48893	36485609	30985759	Manhattan	Hell's Kitchen	40.75751	-73.99112	Shared room	55	1	0	2019-06-23	

Isi *Missing Values* pada kolom “review_per_month” dengan Rata-Rata Data

```
In [117]: airbnb['reviews_per_month'].fillna(airbnb['reviews_per_month'].mean())  
airbnb
```

Out[117]:

	id	host_id	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review
0	2539	2787	Brooklyn	Kensington	40.64749	-73.97237	Private room	149	1	9	2018-10-19
1	2595	2845	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21
2	3647	4632	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	2019-06-23
3	3831	4869	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05
4	5022	7192	Manhattan	East Harlem	40.79851	-73.94399	Entire home/apt	80	10	9	2018-11-19
...
48890	36484665	8232441	Brooklyn	Bedford-Stuyvesant	40.67853	-73.94995	Private room	70	2	0	2019-06-23
48891	36485057	6570630	Brooklyn	Bushwick	40.70184	-73.93317	Private room	40	4	0	2019-06-23
48892	36485431	23492952	Manhattan	Harlem	40.81475	-73.94867	Entire home/apt	115	10	0	2019-06-23
48893	36485609	30985759	Manhattan	Hell's Kitchen	40.75751	-73.99112	Shared room	55	1	0	2019-06-23
48894	36487245	68119814	Manhattan	Hell's Kitchen	40.76404	-73.98933	Private room	90	7	0	2019-06-23

Membuat Kategori Harga

longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month	calculated_host_listings_count	availability_365	range_price
-73.97237	Private room	149	1	9	2018-10-19	0.21	6	365	Cheap
-73.98377	Entire home/apt	225	1	45	2019-05-21	0.38	2	355	Cheap
-73.94190	Private room	150	3	0	2019-06-23	NaN	1	365	Cheap
-73.95976	Entire home/apt	89	1	270	2019-07-05	4.64	1	194	Cheap
-73.94399	Entire home/apt	80	10	9	2018-11-19	0.10	1	0	Cheap
...
-73.94995	Private room	70	2	0	2019-06-23	NaN	2	9	Cheap

```
airbnb["range_price"].value_counts()
```

```
: Cheap      48656
   Medium     223
   Expensive   16
   Name: range_price, dtype: int64
```

Filter semua data dengan last_review > 2019

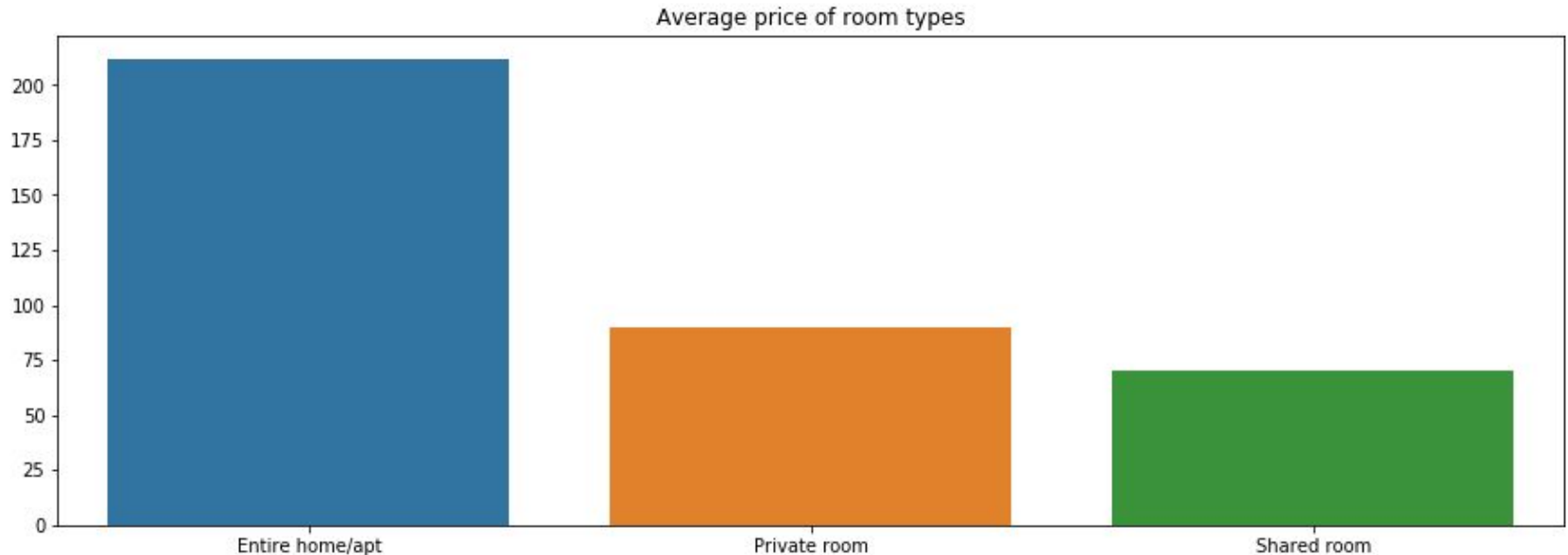
Filter semua data dengan last_review > 2019

```
In [131]: latest_review = airbnb.loc[(airbnb['last_review'] > '2018-12-31')]  
latest_review
```

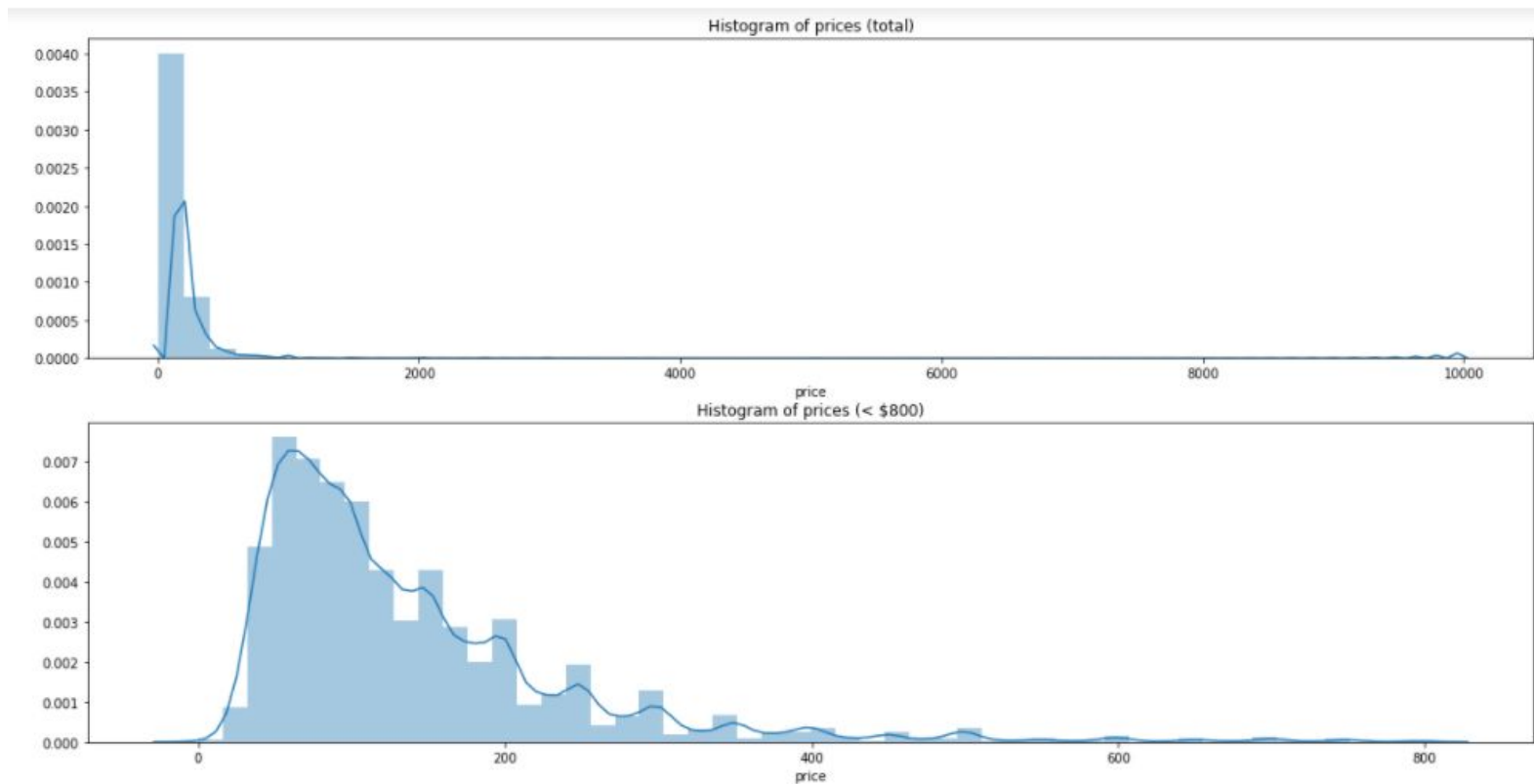
host_id	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	last_review	reviews_per_month
2845	Manhattan	Midtown	40.75362	-73.98377	Entire home/apt	225	1	45	2019-05-21	0.38
4632	Manhattan	Harlem	40.80902	-73.94190	Private room	150	3	0	2019-06-23	NaN
4869	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire home/apt	89	1	270	2019-07-05	4.64
7322	Manhattan	Murray Hill	40.74767	-73.97500	Entire home/apt	200	3	74	2019-06-22	0.59
8967	Manhattan	Hell's Kitchen	40.76489	-73.98493	Private room	79	2	430	2019-06-24	3.47
...
8232441	Brooklyn	Bedford-Stuyvesant	40.67853	-73.94995	Private room	70	2	0	2019-06-23	NaN

Visualisasikan kolom "price" untuk membandingkan harga kamar yang tersedia

```
In [165]: d = airbnb.groupby('room_type')['price'].mean().sort_values(ascending=False)
fig, ax= plt.subplots(figsize=(15,5))
sns.barplot(d.index.tolist(), d.values, ax=ax)
ax.set_title('Average price of room types')
fig.show()
```



Distribusi kolom 'price' pada Dataset AirBnb



Visualisasi harga dibandingkan dengan ketersediaan ruangan dalam hari

```
In [38]: x = airbnb['price']  
y = airbnb['availability_365']  
plt.xlabel('Harga')  
plt.ylabel('Tersedianya Ruangan dalam Hari')  
plt.scatter(x,y)
```

Out[38]: <matplotlib.collections.PathCollection at 0x187d0051e08>

