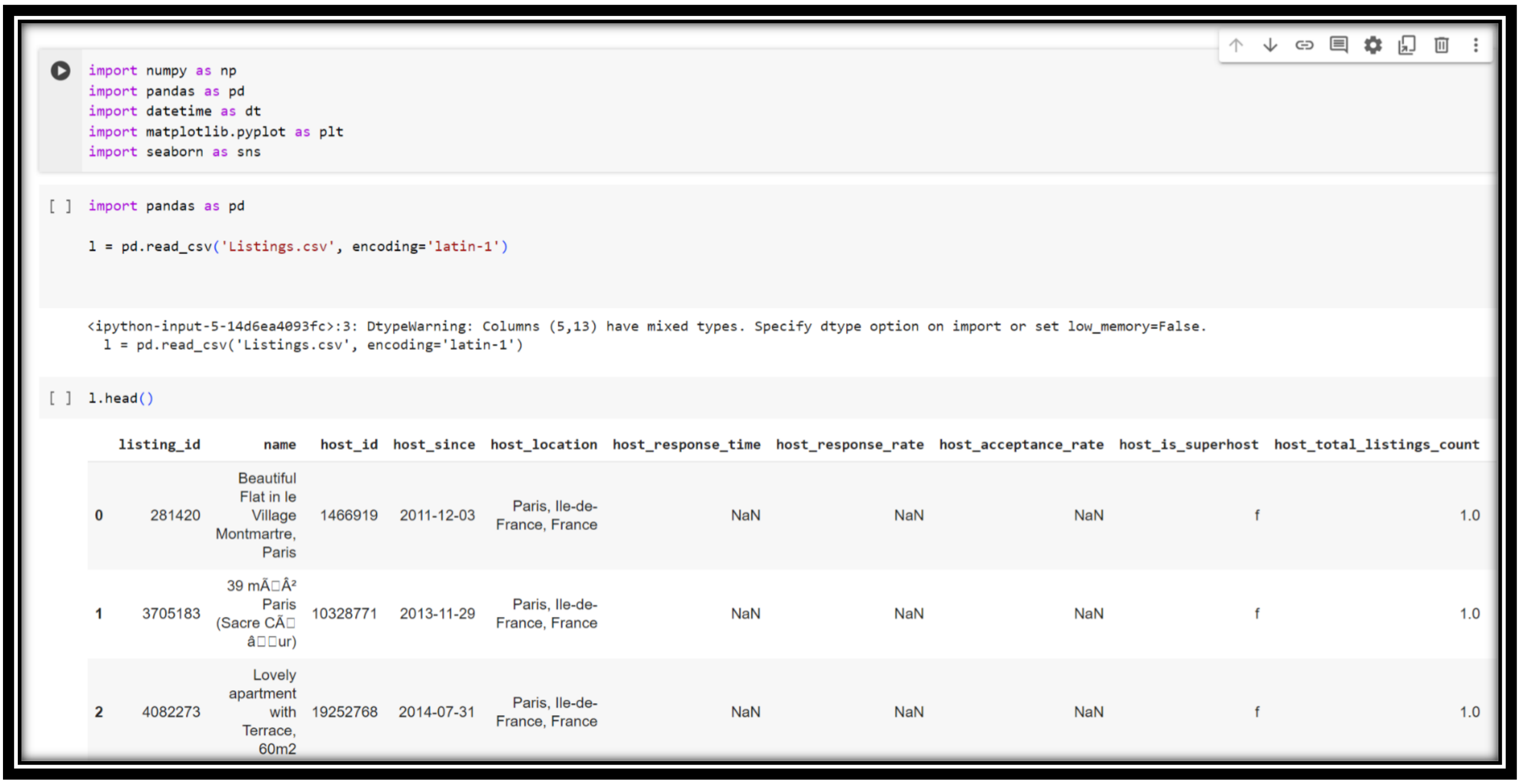
# Air BnB

Amjad Alomani

Aisha Aloumi

Instructor: Ali El-kassas

# 1-Data Loading and Initial Exploration:



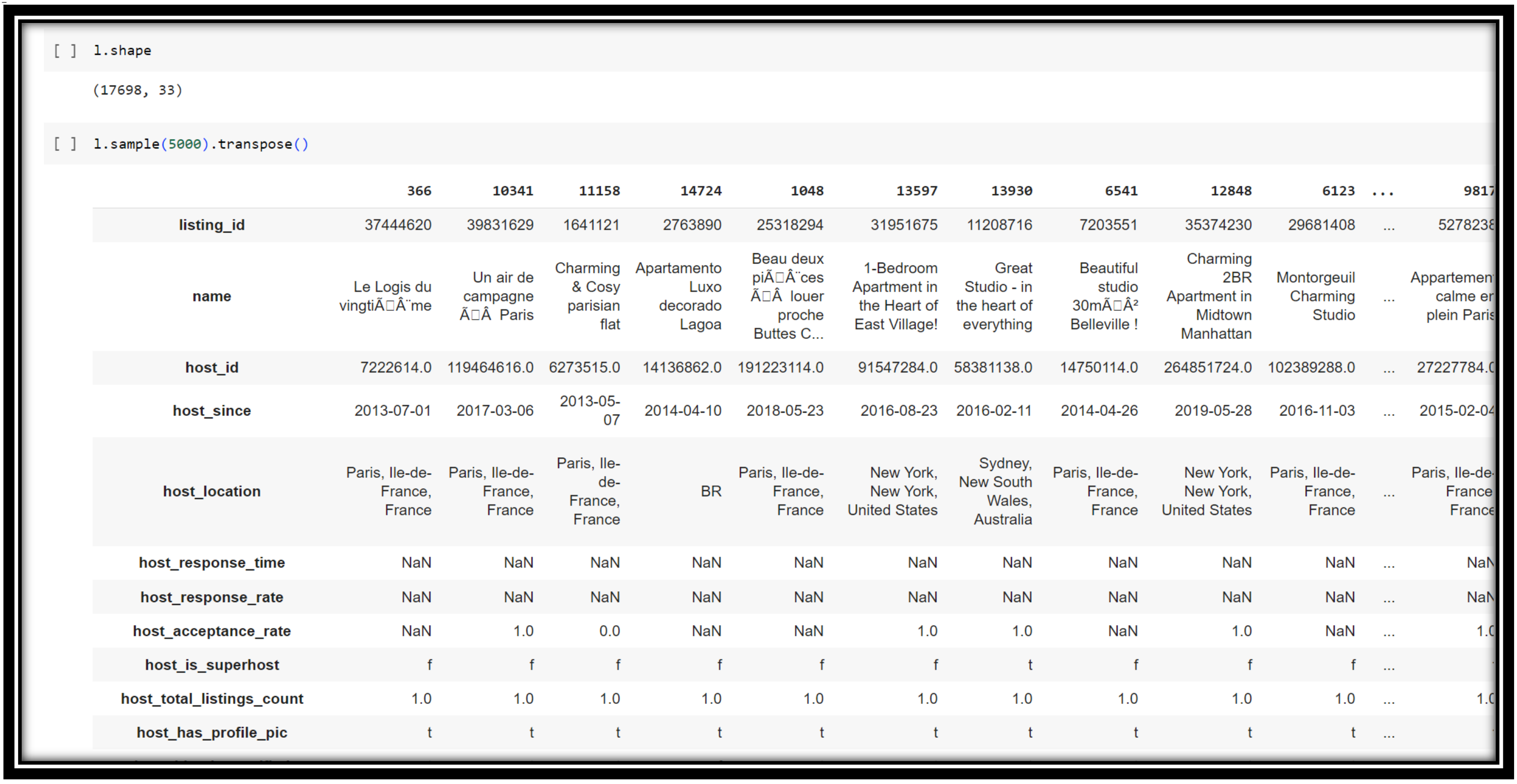
```
import numpy as np
import pandas as pd
import datetime as dt
import matplotlib.pyplot as plt
import seaborn as sns
```

```
[ ] import pandas as pd

    l = pd.read_csv('Listings.csv', encoding='latin-1')
```

```
<ipython-input-5-14d6ea4093fc>:3: DtypeWarning: Columns (5,13) have mixed types. Specify dtype option on import or set low_memory=False.
  l = pd.read_csv('Listings.csv', encoding='latin-1')
```

```
[ ] l.head()
```

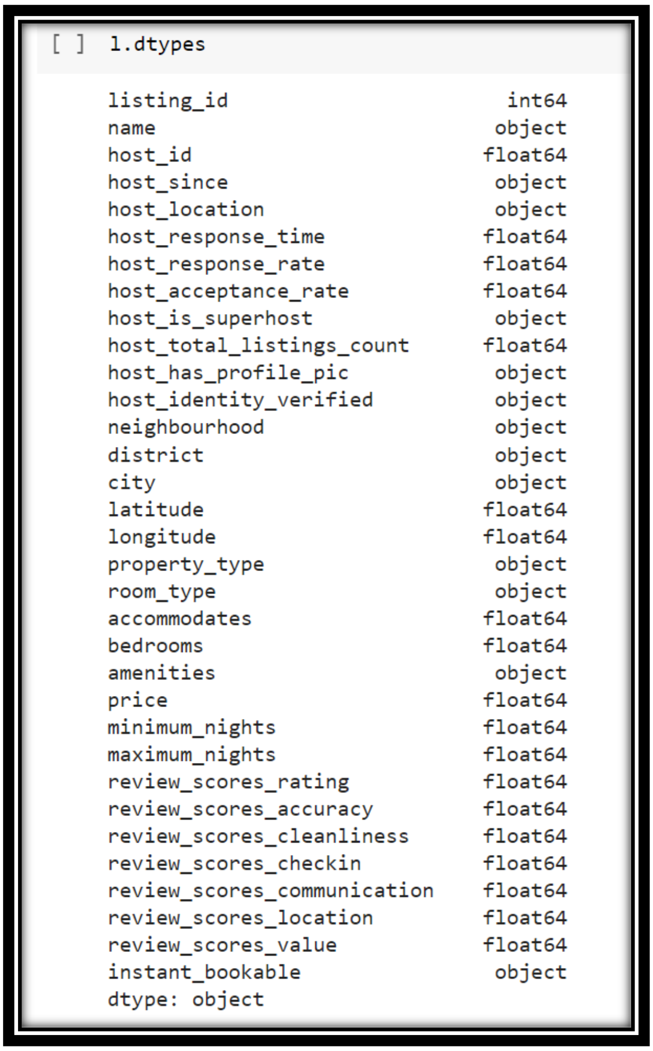
| | listing_id | name | host_id | host_since | host_location | host_response_time | host_response_rate | host_acceptance_rate | host_is_superhost | host_total_listings_count |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 281420 | Beautiful Flat in le Village Montmartre, Paris | 1466919 | 2011-12-03 | Paris, Ile-de-France, France | NaN | NaN | NaN | f | 1.0 |
| 1 | 3705183 | 39 mÃ□Â² Paris (Sacre CÃ□ â□□ur) | 10328771 | 2013-11-29 | Paris, Ile-de-France, France | NaN | NaN | NaN | f | 1.0 |
| 2 | 4082273 | Lovely apartment with Terrace, 60m2 | 19252768 | 2014-07-31 | Paris, Ile-de-France, France | NaN | NaN | NaN | f | 1.0 |



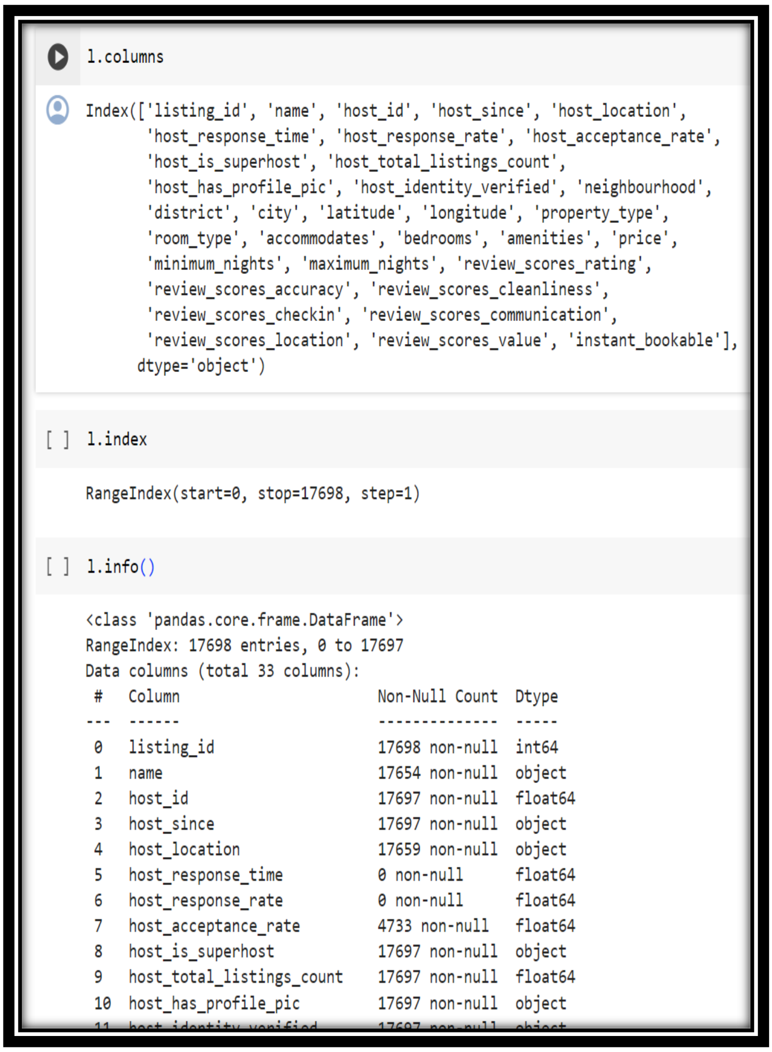
```
[ ] l.shape

(17698, 33)
```

```
[ ] l.sample(5000).transpose()
```

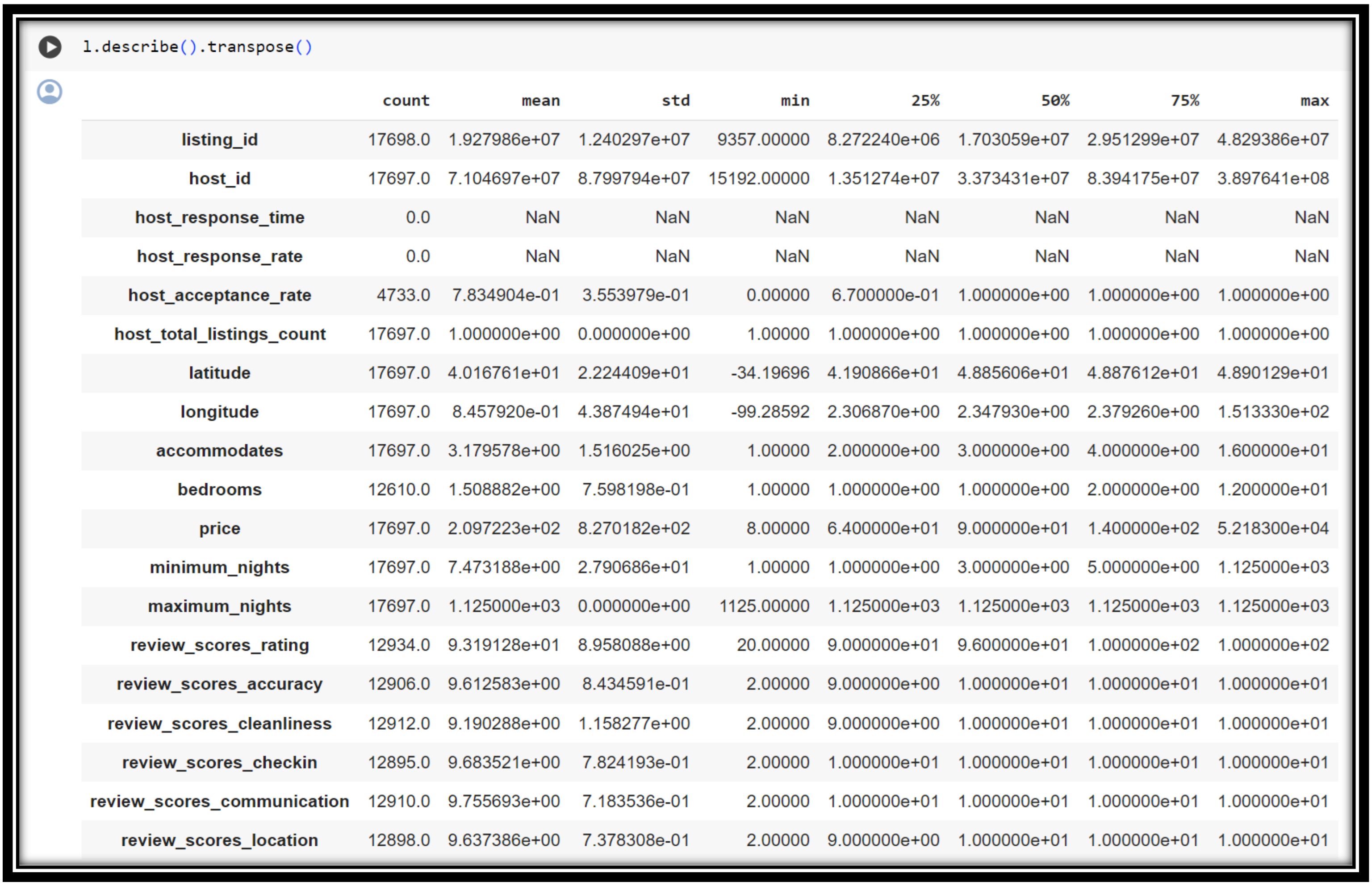
| | 366 | 10341 | 11158 | 14724 | 1048 | 13597 | 13930 | 6541 | 12848 | 6123 | ... | 9817 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| listing_id | 37444620 | 39831629 | 1641121 | 2763890 | 25318294 | 31951675 | 11208716 | 7203551 | 35374230 | 29681408 | ... | 5278238 |
| name | Le Logis du vingtiÃ□Â¨me | Un air de campagne Ã□Â  Paris | Charming & Cosy parisian flat | Apartamento Luxo decorado Lagoa | Beau deux piÃ□Â¨ces Ã□Â  louer proche Buttes C... | 1-Bedroom Apartment in the Heart of East Village! | Great Studio - in the heart of everything | Beautiful studio 30mÃ□Â² Belleville ! | Charming 2BR Apartment in Midtown Manhattan | Montorgeuil Charming Studio | ... | Appartement calme en plein Paris |
| host_id | 7222614.0 | 119464616.0 | 6273515.0 | 14136862.0 | 191223114.0 | 91547284.0 | 58381138.0 | 14750114.0 | 264851724.0 | 102389288.0 | ... | 27227784. |
| host_since | 2013-07-01 | 2017-03-06 | 2013-05-07 | 2014-04-10 | 2018-05-23 | 2016-08-23 | 2016-02-11 | 2014-04-26 | 2019-05-28 | 2016-11-03 | ... | 2015-02-04 |
| host_location | Paris, Ile-de-France, France | Paris, Ile-de-France, France | Paris, Ile-de-France, France | BR | Paris, Ile-de-France, France | New York, New York, United States | Sydney, New South Wales, Australia | Paris, Ile-de-France, France | New York, New York, United States | Paris, Ile-de-France, France | ... | Paris, Ile-de-France France |
| host_response_time | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN |
| host_response_rate | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | ... | NaN |
| host_acceptance_rate | NaN | 1.0 | 0.0 | NaN | NaN | 1.0 | 1.0 | NaN | 1.0 | NaN | ... | 1. |
| host_is_superhost | f | f | f | f | f | f | t | f | f | f | ... | |
| host_total_listings_count | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | 1.0 | ... | 1. |
| host_has_profile_pic | t | t | t | t | t | t | t | t | t | t | ... | |



```
[ ]  l.dtypes

listing_id                       int64
name                            object
host_id                        float64
host_since                      object
host_location                   object
host_response_time             float64
host_response_rate             float64
host_acceptance_rate           float64
host_is_superhost               object
host_total_listings_count      float64
host_has_profile_pic            object
host_identity_verified          object
neighbourhood                   object
district                        object
city                            object
latitude                       float64
longitude                      float64
property_type                   object
room_type                       object
accommodates                   float64
bedrooms                       float64
amenities                       object
price                          float64
minimum_nights                 float64
maximum_nights                 float64
review_scores_rating           float64
review_scores_accuracy         float64
review_scores_cleanliness      float64
review_scores_checkin          float64
review_scores_communication    float64
review_scores_location         float64
review_scores_value            float64
instant_bookable                object
dtype: object
```



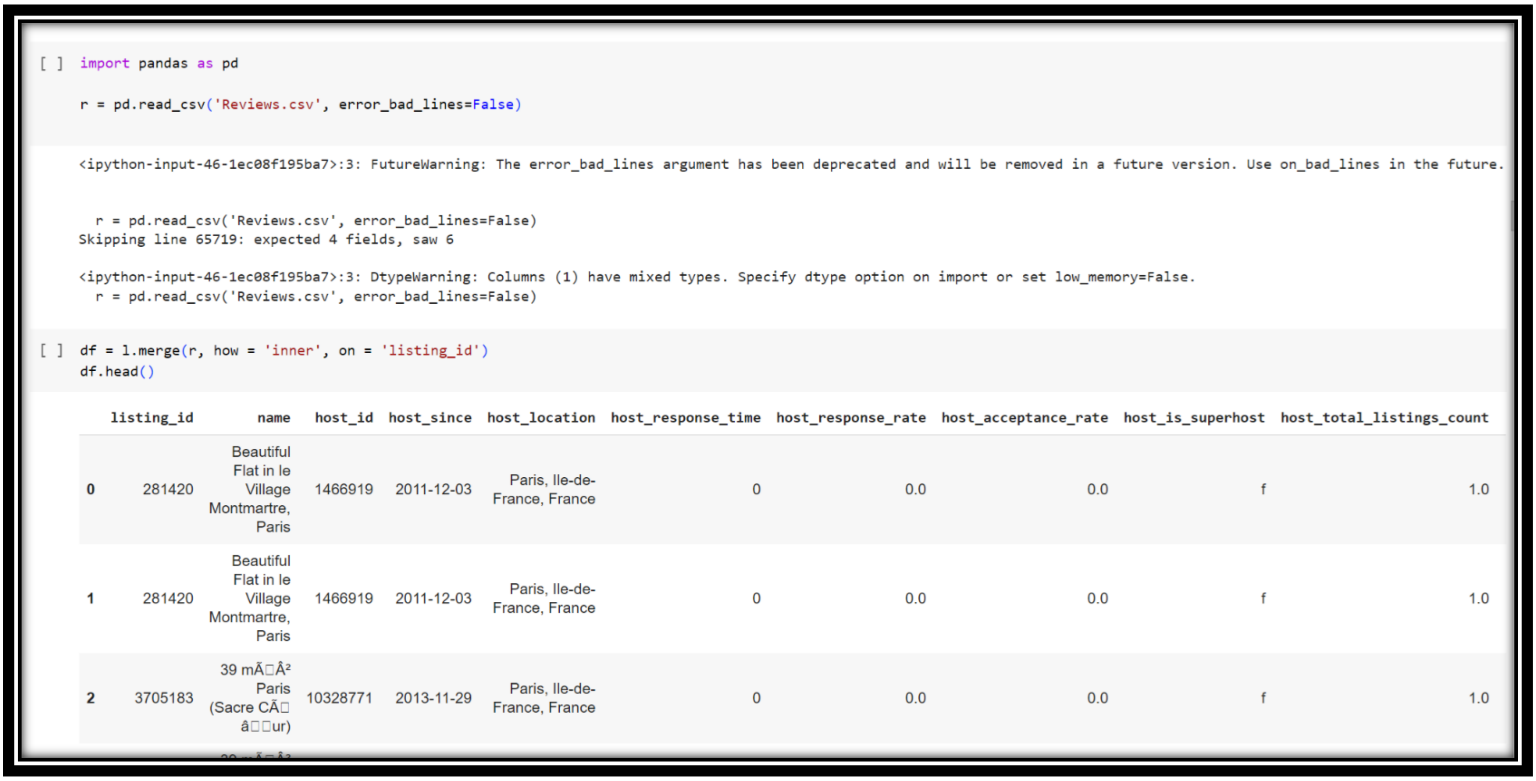
```
l.columns

Index(['listing_id', 'name', 'host_id', 'host_since', 'host_location',
       'host_response_time', 'host_response_rate', 'host_acceptance_rate',
       'host_is_superhost', 'host_total_listings_count',
       'host_has_profile_pic', 'host_identity_verified', 'neighbourhood',
       'district', 'city', 'latitude', 'longitude', 'property_type',
       'room_type', 'accommodates', 'bedrooms', 'amenities', 'price',
       'minimum_nights', 'maximum_nights', 'review_scores_rating',
       'review_scores_accuracy', 'review_scores_cleanliness',
       'review_scores_checkin', 'review_scores_communication',
       'review_scores_location', 'review_scores_value', 'instant_bookable'],
      dtype='object')

[ ]  l.index

RangeIndex(start=0, stop=17698, step=1)

[ ]  l.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 17698 entries, 0 to 17697
Data columns (total 33 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   listing_id                   17698 non-null  int64
 1   name                         17654 non-null  object
 2   host_id                      17697 non-null  float64
 3   host_since                   17697 non-null  object
 4   host_location                17659 non-null  object
 5   host_response_time           0 non-null      float64
 6   host_response_rate           0 non-null      float64
 7   host_acceptance_rate         4733 non-null   float64
 8   host_is_superhost            17697 non-null  object
 9   host_total_listings_count    17697 non-null  float64
 10  host_has_profile_pic         17697 non-null  object
 11  host_identity_verified       17697 non-null  object
```



```
l.describe().transpose()
```

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **listing_id** | 17698.0 | 1.927986e+07 | 1.240297e+07 | 9357.00000 | 8.272240e+06 | 1.703059e+07 | 2.951299e+07 | 4.829386e+07 |
| **host_id** | 17697.0 | 7.104697e+07 | 8.799794e+07 | 15192.00000 | 1.351274e+07 | 3.373431e+07 | 8.394175e+07 | 3.897641e+08 |
| **host_response_time** | 0.0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **host_response_rate** | 0.0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| **host_acceptance_rate** | 4733.0 | 7.834904e-01 | 3.553979e-01 | 0.00000 | 6.700000e-01 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 |
| **host_total_listings_count** | 17697.0 | 1.000000e+00 | 0.000000e+00 | 1.00000 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 | 1.000000e+00 |
| **latitude** | 17697.0 | 4.016761e+01 | 2.224409e+01 | -34.19696 | 4.190866e+01 | 4.885606e+01 | 4.887612e+01 | 4.890129e+01 |
| **longitude** | 17697.0 | 8.457920e-01 | 4.387494e+01 | -99.28592 | 2.306870e+00 | 2.347930e+00 | 2.379260e+00 | 1.513330e+02 |
| **accommodates** | 17697.0 | 3.179578e+00 | 1.516025e+00 | 1.00000 | 2.000000e+00 | 3.000000e+00 | 4.000000e+00 | 1.600000e+01 |
| **bedrooms** | 12610.0 | 1.508882e+00 | 7.598198e-01 | 1.00000 | 1.000000e+00 | 1.000000e+00 | 2.000000e+00 | 1.200000e+01 |
| **price** | 17697.0 | 2.097223e+02 | 8.270182e+02 | 8.00000 | 6.400000e+01 | 9.000000e+01 | 1.400000e+02 | 5.218300e+04 |
| **minimum_nights** | 17697.0 | 7.473188e+00 | 2.790686e+01 | 1.00000 | 1.000000e+00 | 3.000000e+00 | 5.000000e+00 | 1.125000e+03 |
| **maximum_nights** | 17697.0 | 1.125000e+03 | 0.000000e+00 | 1125.00000 | 1.125000e+03 | 1.125000e+03 | 1.125000e+03 | 1.125000e+03 |
| **review_scores_rating** | 12934.0 | 9.319128e+01 | 8.958088e+00 | 20.00000 | 9.000000e+01 | 9.600000e+01 | 1.000000e+02 | 1.000000e+02 |
| **review_scores_accuracy** | 12906.0 | 9.612583e+00 | 8.434591e-01 | 2.00000 | 9.000000e+00 | 1.000000e+01 | 1.000000e+01 | 1.000000e+01 |
| **review_scores_cleanliness** | 12912.0 | 9.190288e+00 | 1.158277e+00 | 2.00000 | 9.000000e+00 | 1.000000e+01 | 1.000000e+01 | 1.000000e+01 |
| **review_scores_checkin** | 12895.0 | 9.683521e+00 | 7.824193e-01 | 2.00000 | 1.000000e+01 | 1.000000e+01 | 1.000000e+01 | 1.000000e+01 |
| **review_scores_communication** | 12910.0 | 9.755693e+00 | 7.183536e-01 | 2.00000 | 1.000000e+01 | 1.000000e+01 | 1.000000e+01 | 1.000000e+01 |
| **review_scores_location** | 12898.0 | 9.637386e+00 | 7.378308e-01 | 2.00000 | 9.000000e+00 | 1.000000e+01 | 1.000000e+01 | 1.000000e+01 |

## 2- Merge listings with reviews



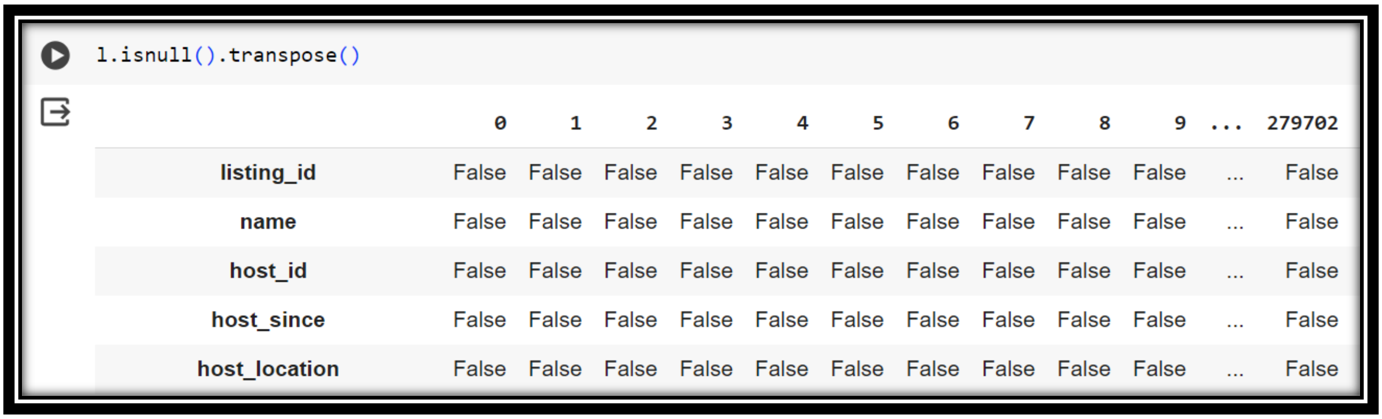
```
[ ] import pandas as pd

    r = pd.read_csv('Reviews.csv', error_bad_lines=False)
```

```
<ipython-input-46-1ec08f195ba7>:3: FutureWarning: The error_bad_lines argument has been deprecated and will be removed in a future version. Use on_bad_lines in the future.


  r = pd.read_csv('Reviews.csv', error_bad_lines=False)
Skipping line 65719: expected 4 fields, saw 6

<ipython-input-46-1ec08f195ba7>:3: DtypeWarning: Columns (1) have mixed types. Specify dtype option on import or set low_memory=False.
  r = pd.read_csv('Reviews.csv', error_bad_lines=False)
```

```
[ ] df = l.merge(r, how = 'inner', on = 'listing_id')
    df.head()
```

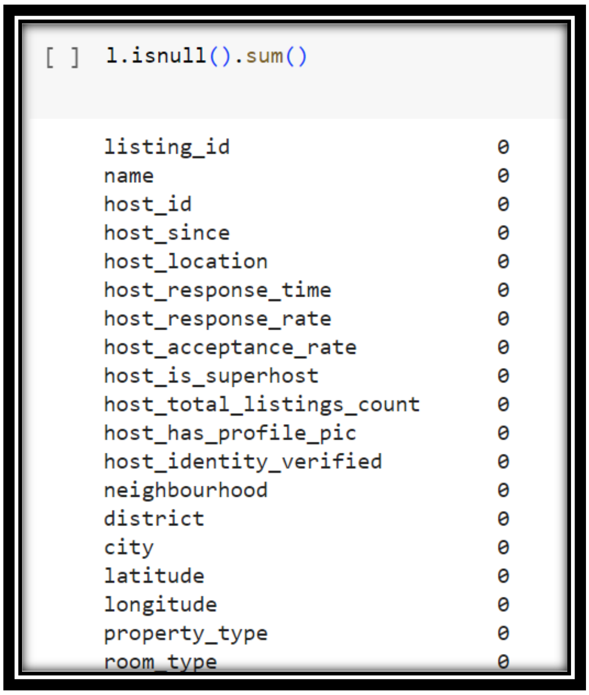
| | listing_id | name | host_id | host_since | host_location | host_response_time | host_response_rate | host_acceptance_rate | host_is_superhost | host_total_listings_count |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 281420 | Beautiful Flat in le Village Montmartre, Paris | 1466919 | 2011-12-03 | Paris, Ile-de-France, France | 0 | 0.0 | 0.0 | f | 1.0 |
| **1** | 281420 | Beautiful Flat in le Village Montmartre, Paris | 1466919 | 2011-12-03 | Paris, Ile-de-France, France | 0 | 0.0 | 0.0 | f | 1.0 |
| **2** | 3705183 | 39 mÃ□Â² Paris (Sacre CÃ□ â□□ur) | 10328771 | 2013-11-29 | Paris, Ile-de-France, France | 0 | 0.0 | 0.0 | f | 1.0 |

## 3- Data cleaning:

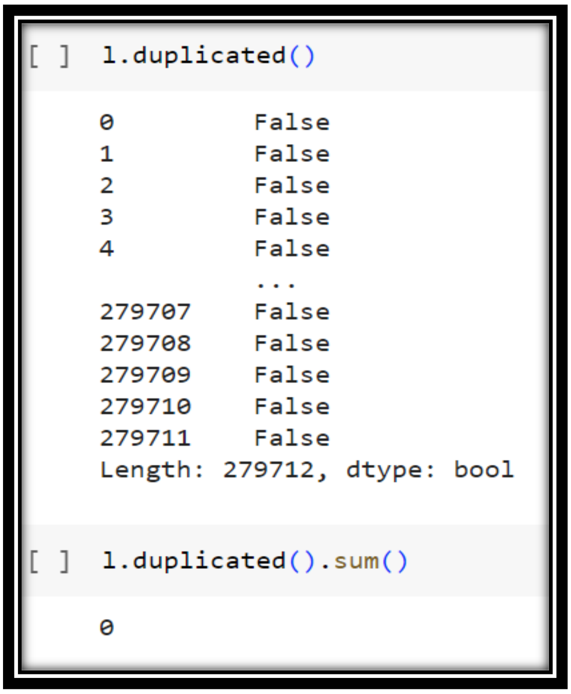


```
l.isnull().transpose()
```

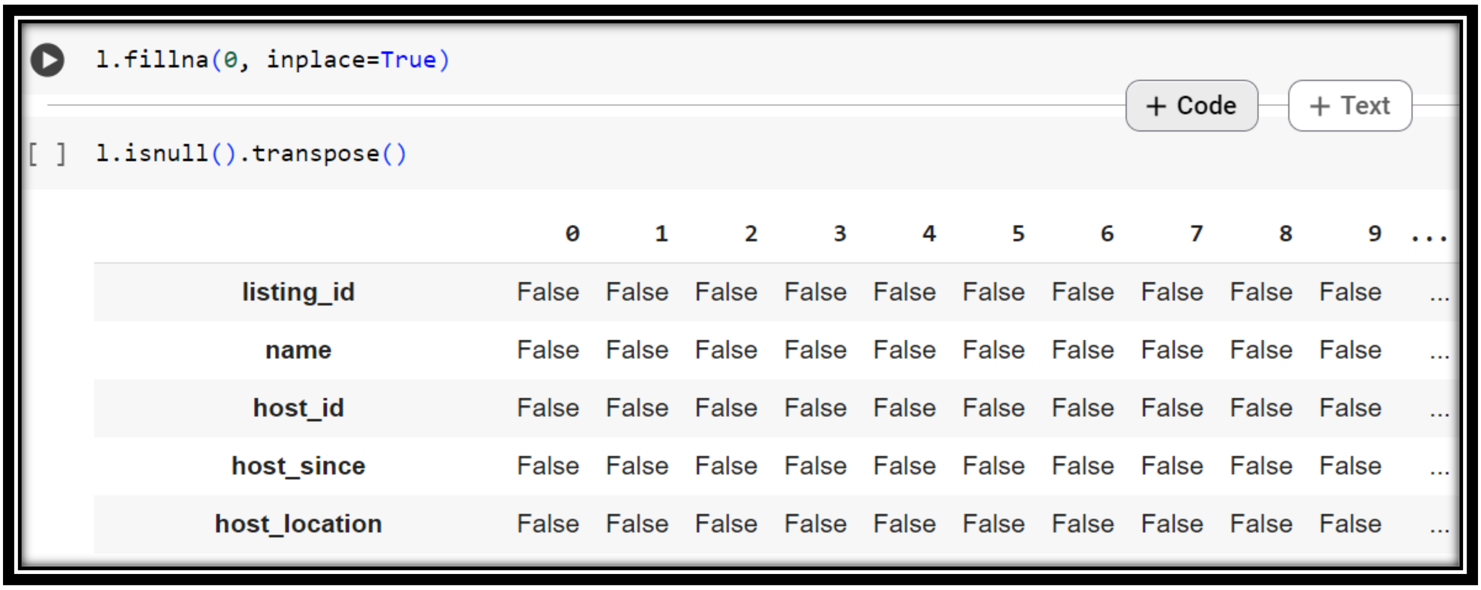
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | ... | 279702 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **listing_id** | False | False | False | False | False | False | False | False | False | False | ... | False |
| **name** | False | False | False | False | False | False | False | False | False | False | ... | False |
| **host_id** | False | False | False | False | False | False | False | False | False | False | ... | False |
| **host_since** | False | False | False | False | False | False | False | False | False | False | ... | False |
| **host_location** | False | False | False | False | False | False | False | False | False | False | ... | False |



```
[ ] l.isnull().sum()

    listing_id                    0
    name                          0
    host_id                       0
    host_since                    0
    host_location                 0
    host_response_time            0
    host_response_rate            0
    host_acceptance_rate          0
    host_is_superhost             0
    host_total_listings_count     0
    host_has_profile_pic          0
    host_identity_verified        0
    neighbourhood                 0
    district                      0
    city                          0
    latitude                      0
    longitude                     0
    property_type                 0
    room_type                     0
```



```
[ ] l.duplicated()

    0         False
    1         False
    2         False
    3         False
    4         False
              ...
    279707    False
    279708    False
    279709    False
    279710    False
    279711    False
    Length: 279712, dtype: bool

[ ] l.duplicated().sum()

    0
```

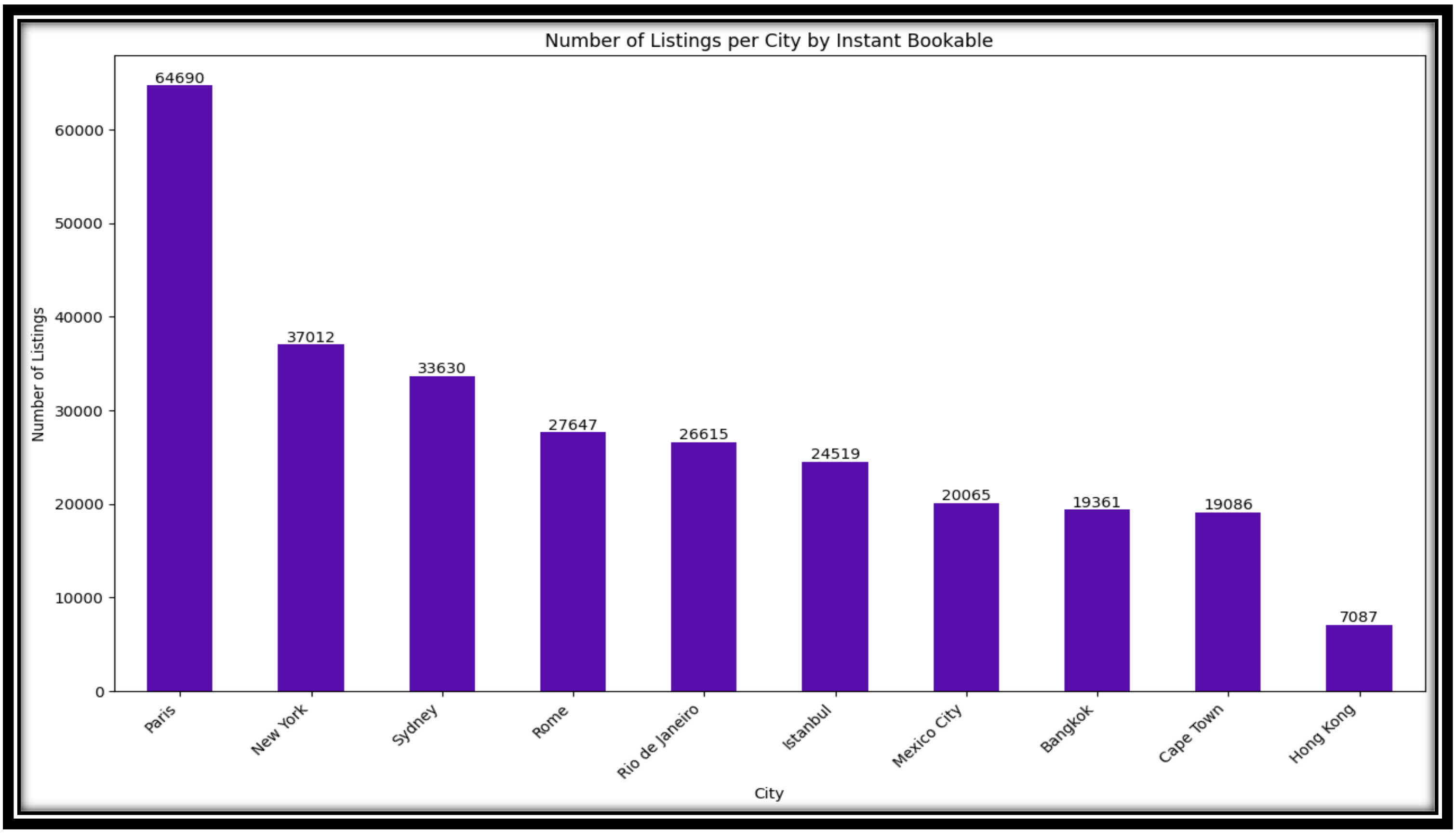


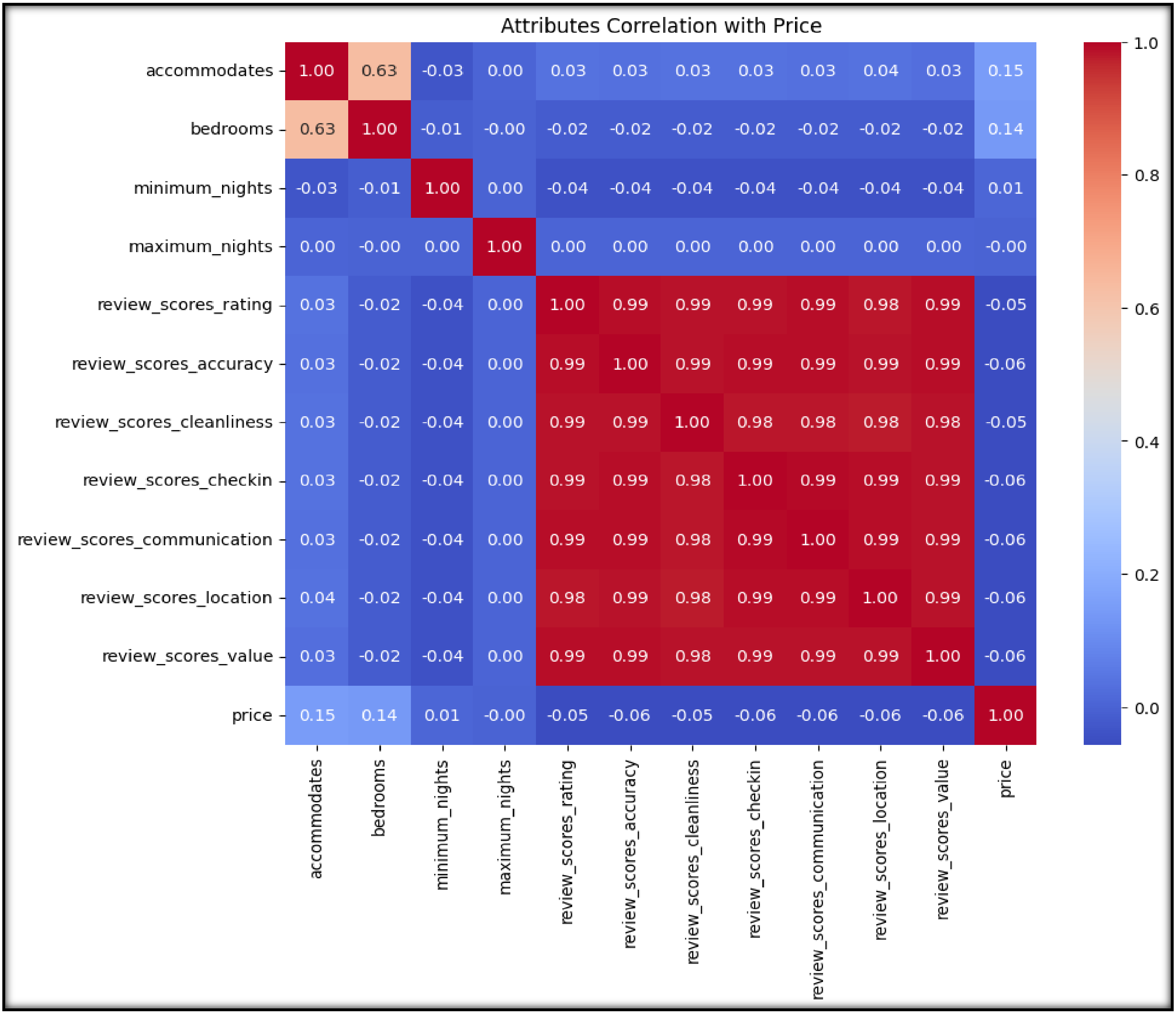
```
l.fillna(0, inplace=True)
```

```
[ ] l.isnull().transpose()
```

|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | ... |
|---|---|---|---|---|---|---|---|---|---|---|---|
| listing_id | False | False | False | False | False | False | False | False | False | False | ... |
| name | False | False | False | False | False | False | False | False | False | False | ... |
| host_id | False | False | False | False | False | False | False | False | False | False | ... |
| host_since | False | False | False | False | False | False | False | False | False | False | ... |
| host_location | False | False | False | False | False | False | False | False | False | False | ... |

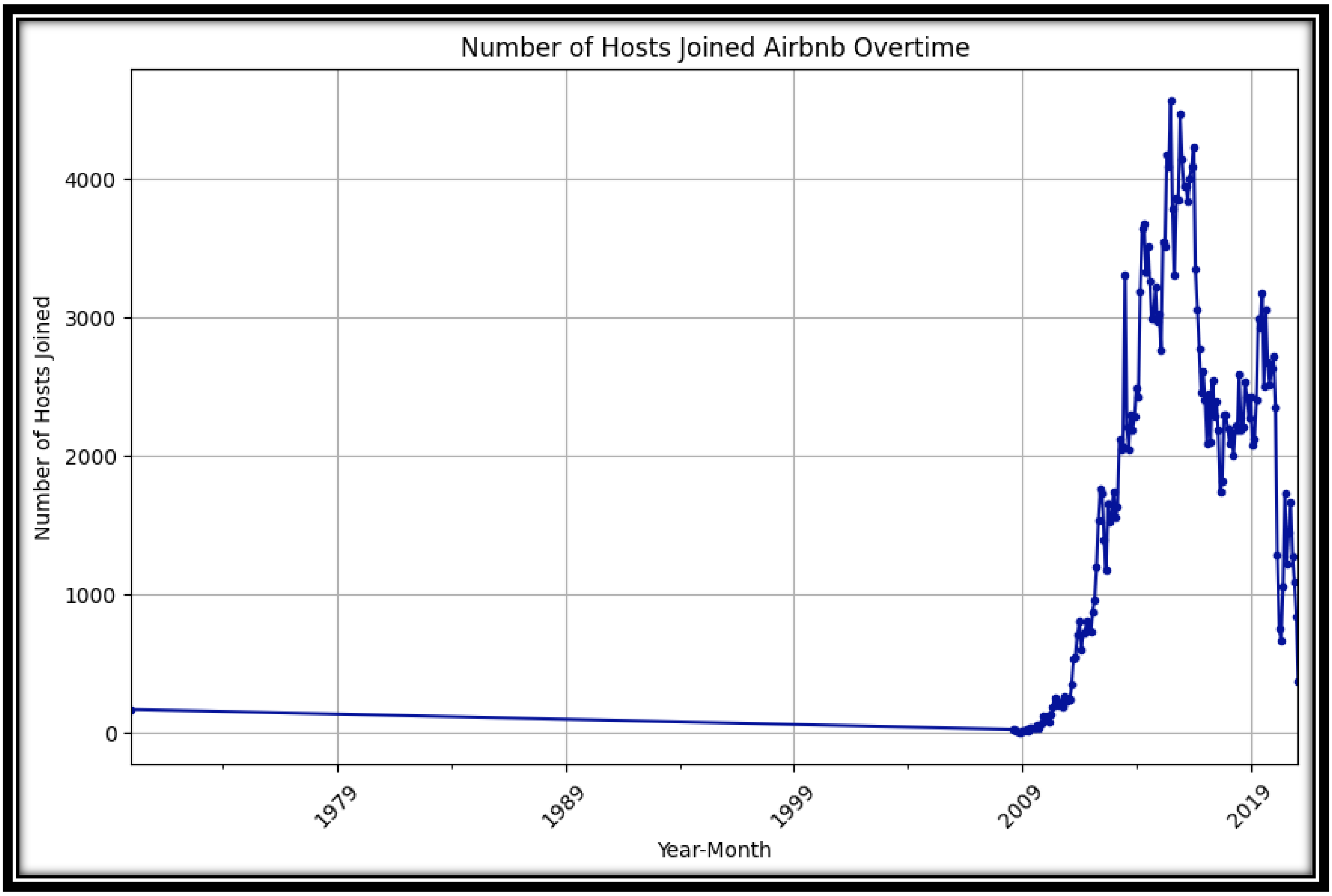
## 4-Exploratory Data Analysis :

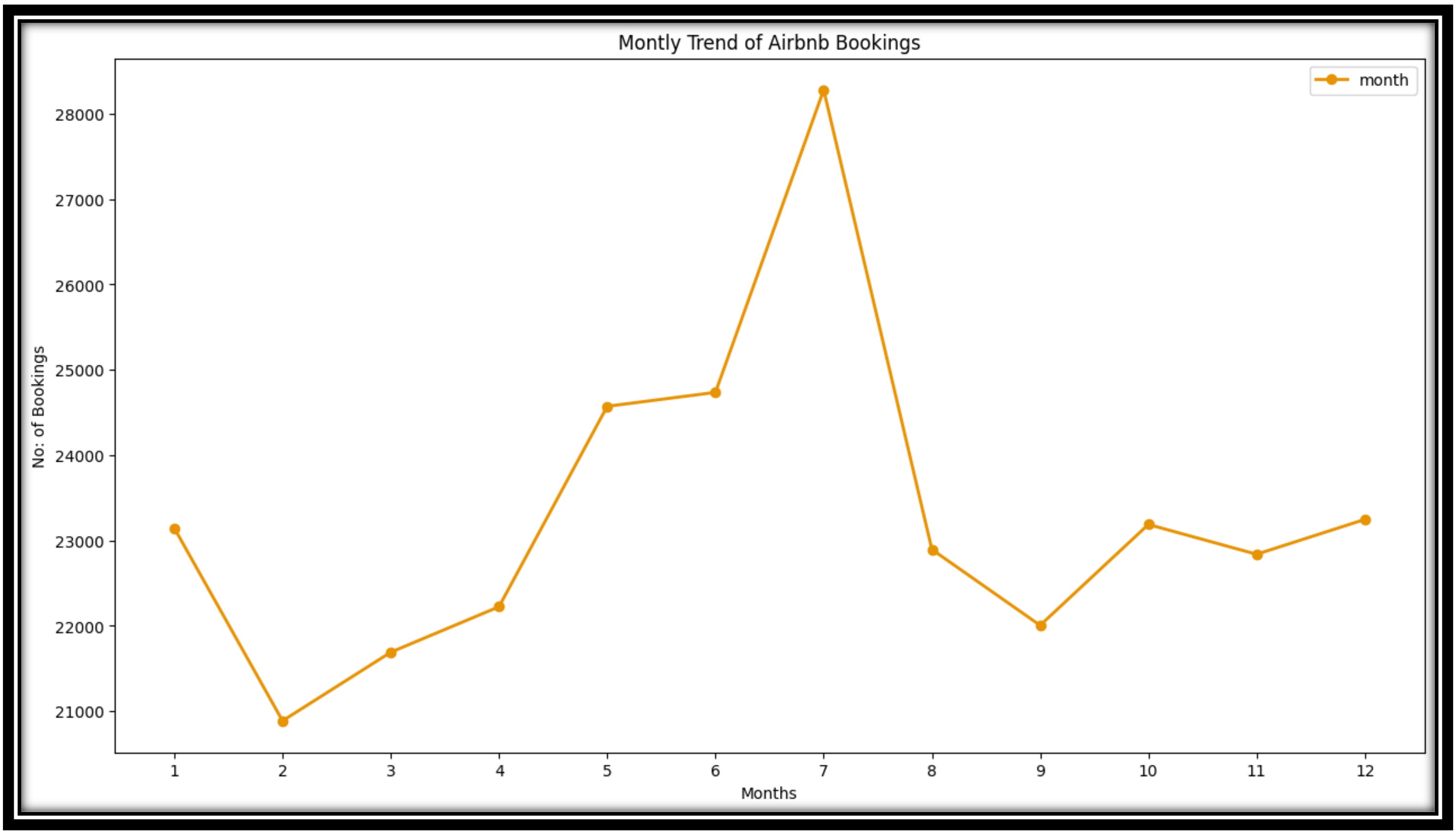


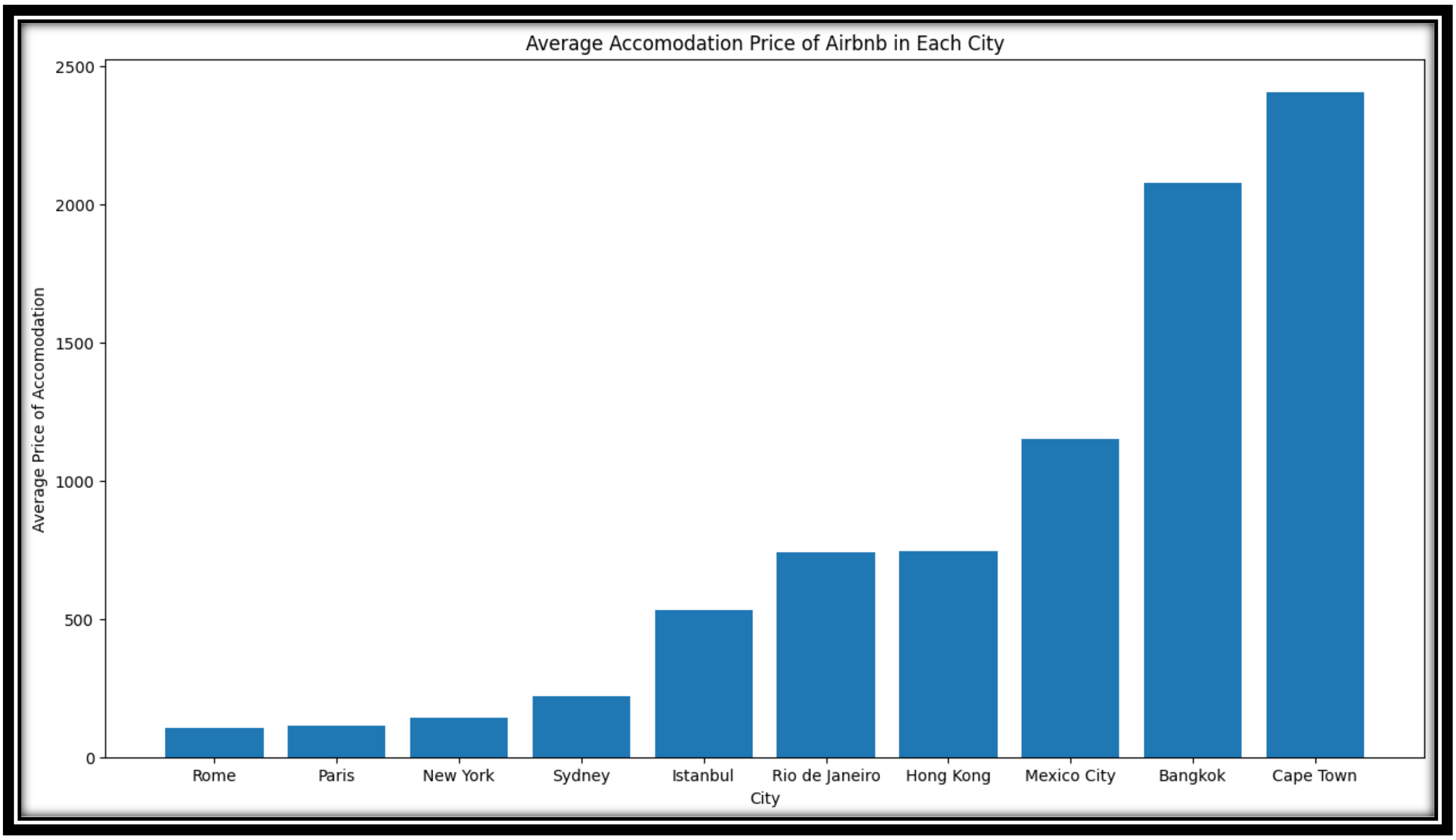


Attributes Correlation with Price

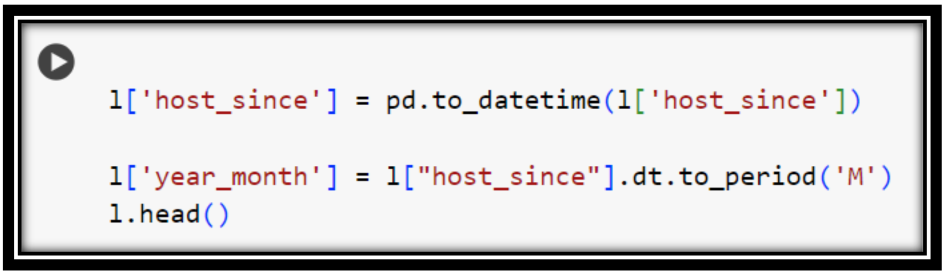
| | accommodates | bedrooms | minimum_nights | maximum_nights | review_scores_rating | review_scores_accuracy | review_scores_cleanliness | review_scores_checkin | review_scores_communication | review_scores_location | review_scores_value | price |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| accommodates | 1.00 | 0.63 | -0.03 | 0.00 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.04 | 0.03 | 0.15 |
| bedrooms | 0.63 | 1.00 | -0.01 | -0.00 | -0.02 | -0.02 | -0.02 | -0.02 | -0.02 | -0.02 | -0.02 | 0.14 |
| minimum_nights | -0.03 | -0.01 | 1.00 | 0.00 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | -0.04 | 0.01 |
| maximum_nights | 0.00 | -0.00 | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.00 |
| review_scores_rating | 0.03 | -0.02 | -0.04 | 0.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.98 | 0.99 | -0.05 |
| review_scores_accuracy | 0.03 | -0.02 | -0.04 | 0.00 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | -0.06 |
| review_scores_cleanliness | 0.03 | -0.02 | -0.04 | 0.00 | 0.99 | 0.99 | 1.00 | 0.98 | 0.98 | 0.98 | 0.98 | -0.05 |
| review_scores_checkin | 0.03 | -0.02 | -0.04 | 0.00 | 0.99 | 0.99 | 0.98 | 1.00 | 0.99 | 0.99 | 0.99 | -0.06 |
| review_scores_communication | 0.03 | -0.02 | -0.04 | 0.00 | 0.99 | 0.99 | 0.98 | 0.99 | 1.00 | 0.99 | 0.99 | -0.06 |
| review_scores_location | 0.04 | -0.02 | -0.04 | 0.00 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 | 1.00 | 0.99 | -0.06 |
| review_scores_value | 0.03 | -0.02 | -0.04 | 0.00 | 0.99 | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 1.00 | -0.06 |
| price | 0.15 | 0.14 | 0.01 | -0.00 | -0.05 | -0.06 | -0.05 | -0.06 | -0.06 | -0.06 | -0.06 | 1.00 |


Number of Hosts Joined Airbnb Overtime
Number of Hosts Joined
0
1000
2000
3000
4000
1979
1989
1999
2009
2019
Year-Month

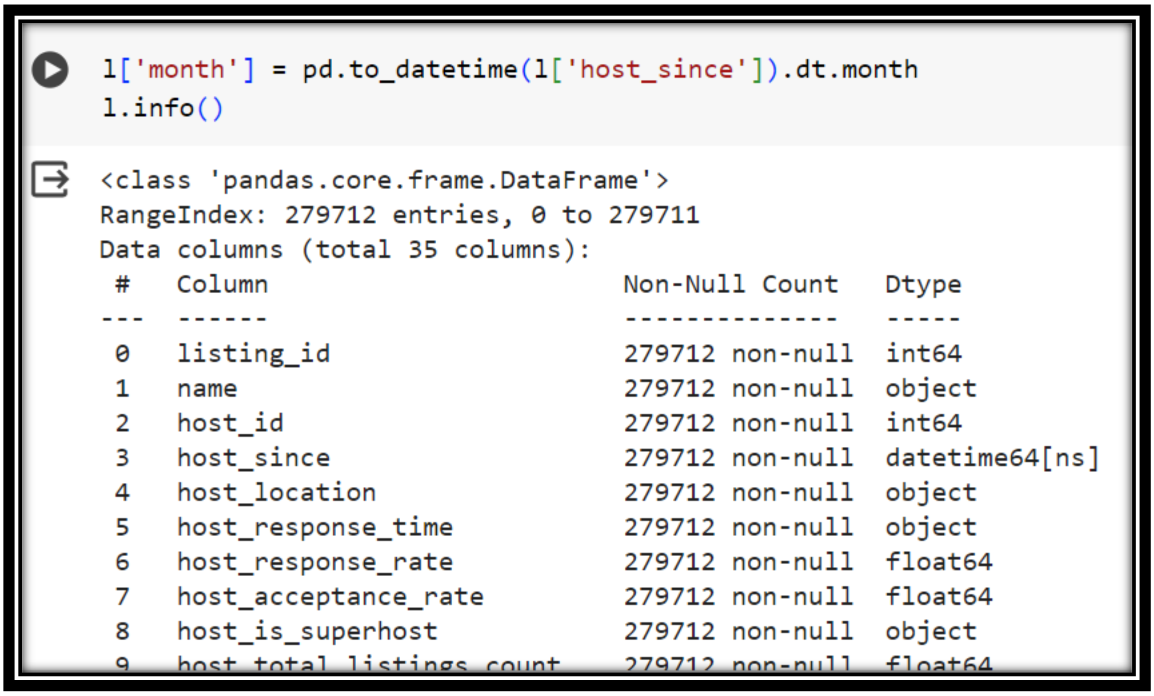

Montly Trend of Airbnb Bookings
month
No: of Bookings
28000
27000
26000
25000
24000
23000
22000
21000
1
2
3
4
5
6
7
8
9
10
11
12
Months


Average Accomodation Price of Airbnb in Each City
Average Price of Accomodation
2500
2000
1500
1000
500
0
Rome
Paris
New York
Sydney
Istanbul
Rio de Janeiro
Hong Kong
Mexico City
Bangkok
Cape Town
City

## 5-Future Engineering:



```
l['host_since'] = pd.to_datetime(l['host_since'])

l['year_month'] = l["host_since"].dt.to_period('M')
l.head()
```



```
l['month'] = pd.to_datetime(l['host_since']).dt.month
l.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 279712 entries, 0 to 279711
Data columns (total 35 columns):
 #   Column                       Non-Null Count   Dtype
---  ------                       --------------   -----
 0   listing_id                   279712 non-null  int64
 1   name                         279712 non-null  object
 2   host_id                      279712 non-null  int64
 3   host_since                   279712 non-null  datetime64[ns]
 4   host_location                279712 non-null  object
 5   host_response_time           279712 non-null  object
 6   host_response_rate           279712 non-null  float64
 7   host_acceptance_rate         279712 non-null  float64
 8   host_is_superhost            279712 non-null  object
 9   host_total_listings_count    279712 non-null  float64
```