

```

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

df=pd.read_csv('/content/AB_NYC_2019.csv')

df

{"summary":{"\n  \"name\": \"df\", \n  \"rows\": 48895, \n  \"fields\": [\n    {\n      \"column\": \"id\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 10983108, \n        \"min\": 2539, \n        \"max\": 36487245, \n        \"num_unique_values\": 48895, \n        \"samples\": [\n          317905, \n          34205267, \n          12342297\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\" \n      }, \n      {\n        \"column\": \"name\", \n        \"properties\": {\n          \"dtype\": \"string\", \n          \"num_unique_values\": 47905, \n          \"samples\": [\n            \"Luxurious Penthouse 3bed/2bath Apt w/Gym, Doorman\", \n            \"MANHATTAN LIVING!\", \n            \"Wonderful private room near Two Bridges II\" \n          ], \n          \"semantic_type\": \"\", \n          \"description\": \"\" \n        }, \n        {\n          \"column\": \"host_id\", \n          \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 78610967, \n            \"min\": 2438, \n            \"max\": 274321313, \n            \"num_unique_values\": 37457, \n            \"samples\": [\n              1504257, \n              5592151, \n              208938947\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\" \n          }, \n          {\n            \"column\": \"host_name\", \n            \"properties\": {\n              \"dtype\": \"category\", \n              \"num_unique_values\": 11452, \n              \"samples\": [\n                \"Eki\", \n                \"Laine\", \n                \"Elen\" \n              ], \n              \"semantic_type\": \"\", \n              \"description\": \"\" \n            }, \n            {\n              \"column\": \"neighbourhood_group\", \n              \"properties\": {\n                \"dtype\": \"category\", \n                \"num_unique_values\": 5, \n                \"samples\": [\n                  \"Manhattan\", \n                  \"Bronx\", \n                  \"Queens\" \n                ], \n                \"semantic_type\": \"\", \n                \"description\": \"\" \n              }, \n              {\n                \"column\": \"neighbourhood\", \n                \"properties\": {\n                  \"dtype\": \"category\", \n                  \"num_unique_values\": 221, \n                  \"samples\": [\n                    \"Stuyvesant Town\", \n                    \"Eltingville\", \n                    \"Stapleton\" \n                  ], \n                  \"semantic_type\": \"\", \n                  \"description\": \"\" \n                }, \n                {\n                  \"column\": \"latitude\", \n                  \"properties\": {\n                    \"dtype\": \"number\", \n                    \"std\": 0.054530078057371895, \n                    \"min\": 40.49979, \n                    \"max\": 40.91306, \n                    \"num_unique_values\": 19048, \n                    \"samples\": [\n                      40.75913, \n                      40.68314, \n                      40.72126\n                    ], \n                    \"semantic_type\": \"\", \n                    \"description\": \"\" \n                  } \n                } \n              } \n            } \n          } \n        } \n      } \n    } \n  ] \n}

```

```

}\n    },\n    {\n        \"column\": \"longitude\", \n        \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": \n0.04615673610637182, \n            \"min\": -74.24442, \n            \"max\": -\n73.71299, \n            \"num_unique_values\": 14718, \n            \"samples\": [\n                -73.88892, \n                -73.87851, \n                -73.97175\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\"\n        }\n    },\n    {\n        \"column\": \"room_type\", \n        \"properties\": {\n            \"dtype\": \"category\", \n            \"num_unique_values\": 3, \n            \"samples\": [\n                \"Private room\", \n                \"Entire home/apt\", \n                \"Shared room\"\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\"\n        }\n    },\n    {\n        \"column\": \"price\", \n        \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 240, \n            \"min\": 0, \n            \"max\": 10000, \n            \"num_unique_values\": 674, \n            \"samples\": [\n                519, \n                675, \n                488\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\"\n        }\n    },\n    {\n        \"column\": \"minimum_nights\", \n        \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 20, \n            \"min\": 1, \n            \"max\": 1250, \n            \"num_unique_values\": 109, \n            \"samples\": [\n                160, \n                60, \n                2\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\"\n        }\n    },\n    {\n        \"column\": \"number_of_reviews\", \n        \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 44, \n            \"min\": 0, \n            \"max\": 629, \n            \"num_unique_values\": 394, \n            \"samples\": [\n                12, \n                144, \n                314\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\"\n        }\n    },\n    {\n        \"column\": \"last_review\", \n        \"properties\": {\n            \"dtype\": \"object\", \n            \"num_unique_values\": 1764, \n            \"samples\": [\n                \"2016-07-26\", \n                \"2018-05-21\", \n                \"2019-02-27\"\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\"\n        }\n    },\n    {\n        \"column\": \"reviews_per_month\", \n        \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 1.6804419952744627, \n            \"min\": 0.01, \n            \"max\": 58.5, \n            \"num_unique_values\": 937, \n            \"samples\": [\n                1.7, \n                0.28, \n                2.14\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\"\n        }\n    },\n    {\n        \"column\": \"calculated_host_listings_count\", \n        \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 32, \n            \"min\": 1, \n            \"max\": 327, \n            \"num_unique_values\": 47, \n            \"samples\": [\n                37, \n                17, \n                121\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\"\n        }\n    },\n    {\n        \"column\": \"availability_365\", \n        \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 131, \n            \"min\": 0, \n            \"max\": 365, \n            \"num_unique_values\": 366, \n

```

```
\ "samples\": [\n                335,\n                309,\n                249\n            ],\n            \"semantic_type\": \"\",\n            \"description\": \"\"\n        }\n    ],\n    \"type\": \"dataframe\", \"variable_name\": \"df\"}
```

```
df.head(10)
```

```
{
  "summary": "{\n    \"name\": \"df\",\n    \"rows\": 48895,\n    \"fields\": {\n      \"column\": \"id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 10983108,\n        \"min\": 2539,\n        \"max\": 36487245,\n        \"num_unique_values\": 48895,\n        \"samples\": [\n          317905,\n          34205267,\n          12342297\n        ],\n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      },\n      {\n        \"column\": \"name\",\n        \"properties\": {\n          \"dtype\": \"string\",\n          \"num_unique_values\": 47905,\n          \"samples\": [\n            \"Luxurious Penthouse 3bed/2bath Apt w/Gym, Doorman\",\n            \"MANHATTAN LIVING!\",\n            \"Wonderful private room near Two Bridges II\"\n          ],\n          \"semantic_type\": \"\", \n          \"description\": \"\"\n        },\n        {\n          \"column\": \"host_id\",\n          \"properties\": {\n            \"dtype\": \"number\",\n            \"std\": 78610967,\n            \"min\": 2438,\n            \"max\": 274321313,\n            \"num_unique_values\": 37457,\n            \"samples\": [\n              1504257,\n              5592151,\n              208938947\n            ],\n            \"semantic_type\": \"\", \n            \"description\": \"\"\n          },\n          {\n            \"column\": \"host_name\",\n            \"properties\": {\n              \"dtype\": \"category\",\n              \"num_unique_values\": 11452,\n              \"samples\": [\n                \"Eki\",\n                \"Laine\",\n                \"Elen\"\n              ],\n              \"semantic_type\": \"\", \n              \"description\": \"\"\n            },\n            {\n              \"column\": \"neighbourhood_group\",\n              \"properties\": {\n                \"dtype\": \"category\",\n                \"num_unique_values\": 5,\n                \"samples\": [\n                  \"Manhattan\",\n                  \"Bronx\",\n                  \"Queens\"\n                ],\n                \"semantic_type\": \"\", \n                \"description\": \"\"\n              },\n              {\n                \"column\": \"neighbourhood\",\n                \"properties\": {\n                  \"dtype\": \"category\",\n                  \"num_unique_values\": 221,\n                  \"samples\": [\n                    \"Stuyvesant Town\",\n                    \"Eltingville\",\n                    \"Stapleton\"\n                  ],\n                  \"semantic_type\": \"\", \n                  \"description\": \"\"\n                },\n                {\n                  \"column\": \"latitude\",\n                  \"properties\": {\n                    \"dtype\": \"number\",\n                    \"std\": 0.054530078057371895,\n                    \"min\": 40.49979,\n                    \"max\": 40.91306,\n                    \"num_unique_values\": 19048,\n                    \"samples\": [\n                      40.75913,\n                      40.68314,\n                      40.72126\n                    ],\n                    \"semantic_type\": \"\", \n                    \"description\": \"\"\n                  },\n                  {\n                    \"column\": \"longitude\",\n                    \"properties\": {\n                      \"dtype\": \"number\",\n                      \"std\": 0.04615673610637182,\n                      \"min\": -74.24442,\n                      \"max\": -\n
```

```

73.71299,\n          \"num_unique_values\": 14718,\n          \"samples\": [\n\n            -73.88892,\n            -73.87851,\n            -73.97175\n          ],\n          \"semantic_type\": \"\", \n          \"description\": \"\"\n        },\n        {\n          \"column\": \"room_type\", \n          \"properties\": {\n            \"dtype\": \"category\", \n            \"num_unique_values\": 3,\n            \"samples\": [\n              \"Private room\", \n              \"Entire home/apt\", \n              \"Shared room\"\n            ],\n            \"semantic_type\": \"\", \n            \"description\": \"\"\n          },\n          {\n            \"column\": \"price\", \n            \"properties\": {\n              \"dtype\": \"number\", \n              \"std\": 240,\n              \"min\": 0,\n              \"max\": 10000,\n              \"num_unique_values\": 674,\n              \"samples\": [\n                519,\n                675,\n                488\n              ],\n              \"semantic_type\": \"\", \n              \"description\": \"\"\n            },\n            {\n              \"column\": \"minimum_nights\", \n              \"properties\": {\n                \"dtype\": \"number\", \n                \"std\": 20,\n                \"min\": 1,\n                \"max\": 1250,\n                \"num_unique_values\": 109,\n                \"samples\": [\n                  160,\n                  60,\n                  2\n                ],\n                \"semantic_type\": \"\", \n                \"description\": \"\"\n              },\n              {\n                \"column\": \"number_of_reviews\", \n                \"properties\": {\n                  \"dtype\": \"number\", \n                  \"std\": 44,\n                  \"min\": 0,\n                  \"max\": 629,\n                  \"num_unique_values\": 394,\n                  \"samples\": [\n                    12,\n                    144,\n                    314\n                  ],\n                  \"semantic_type\": \"\", \n                  \"description\": \"\"\n                },\n                {\n                  \"column\": \"last_review\", \n                  \"properties\": {\n                    \"dtype\": \"object\", \n                    \"num_unique_values\": 1764,\n                    \"samples\": [\n                      \"2016-07-26\", \n                      \"2018-05-21\", \n                      \"2019-02-27\"\n                    ],\n                    \"semantic_type\": \"\", \n                    \"description\": \"\"\n                  },\n                  {\n                    \"column\": \"reviews_per_month\", \n                    \"properties\": {\n                      \"dtype\": \"number\", \n                      \"std\": 1.6804419952744627,\n                      \"min\": 0.01,\n                      \"max\": 58.5,\n                      \"num_unique_values\": 937,\n                      \"samples\": [\n                        1.7,\n                        0.28,\n                        2.14\n                      ],\n                      \"semantic_type\": \"\", \n                      \"description\": \"\"\n                    },\n                    {\n                      \"column\": \"calculated_host_listings_count\", \n                      \"properties\": {\n                        \"dtype\": \"number\", \n                        \"std\": 32,\n                        \"min\": 1,\n                        \"max\": 327,\n                        \"num_unique_values\": 47,\n                        \"samples\": [\n                          37,\n                          17,\n                          121\n                        ],\n                        \"semantic_type\": \"\", \n                        \"description\": \"\"\n                      },\n                      {\n                        \"column\": \"availability_365\", \n                        \"properties\": {\n                          \"dtype\": \"number\", \n                          \"std\": 131,\n                          \"min\": 0,\n                          \"max\": 365,\n                          \"num_unique_values\": 366,\n                          \"samples\": [\n                            335,\n                            309,\n                            249\n                          ],\n                          \"semantic_type\": \"\", \n                          \"description\": \"\"\n                        }\n                      }\n                    }\n                  }\n                }\n              ],\n              \"type\": \"dataframe\", \"variable_name\": \"df\"}

```

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

```
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
dtype: int64
```

```
df.dtypes
```

```
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
```

```
df['last_review']=pd.to_datetime(df['last_review'])
```

```
df['last_review'].dtypes
```

```
dtype('<M8[ns]')
```

```
df['name']
```

```
0          Clean & quiet apt home by the park
1                Skylit Midtown Castle
2      THE VILLAGE OF HARLEM....NEW YORK !
```

```

3           Cozy Entire Floor of Brownstone
4   Entire Apt: Spacious Studio/Loft by central park
      ...
48890   Charming one bedroom - newly renovated rowhouse
48891   Affordable room in Bushwick/East Williamsburg
48892   Sunny Studio at Historical Neighborhood
48893   43rd St. Time Square-cozy single bed
48894   Trendy duplex in the very heart of Hell's Kitchen
Name: name, Length: 48895, dtype: object

```

here we cant eliminate null value with any other name of apt so will just drop the row with null values

```

df=df.dropna(subset=['name'])
df['name']

0           Clean & quiet apt home by the park
1           Skylit Midtown Castle
2   THE VILLAGE OF HARLEM....NEW YORK !
3           Cozy Entire Floor of Brownstone
4   Entire Apt: Spacious Studio/Loft by central park
      ...
48890   Charming one bedroom - newly renovated rowhouse
48891   Affordable room in Bushwick/East Williamsburg
48892   Sunny Studio at Historical Neighborhood
48893   43rd St. Time Square-cozy single bed
48894   Trendy duplex in the very heart of Hell's Kitchen
Name: name, Length: 48879, dtype: object

df['name'].isnull().sum()

0

df.isnull().sum()

id           0
name         0
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  10042
reviews_per_month  10042

```

```

calculated_host_listings_count    0
availability_365                  0
dtype: int64

df['last_review'].isnull().sum()

10042

df['last_review']

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: 48879, dtype: datetime64[ns]

```

since the last review stand for latest reviewed so its in the form of dates so we will fill the null value with no review statemet. so lets assume that there was no review from host_id

```

df['last_review']=df['last_review'].fillna('No Review')

<ipython-input-127-e16aee4c8108>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
    df['last_review']=df['last_review'].fillna('No Review')

df['last_review']

0      2018-10-19 00:00:00
1      2019-05-21 00:00:00
2              No Review
3      2019-07-05 00:00:00
4      2018-11-19 00:00:00
...
48890         No Review
48891         No Review
48892         No Review
48893         No Review
48894         No Review
Name: last_review, Length: 48879, dtype: object

```

```

df['last_review'].isnull().sum()
0
df.isnull().sum()
id                0
name              0
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       0
reviews_per_month 10042
calculated_host_listings_count 0
availability_365  0
dtype: int64

df['reviews_per_month'].isnull().sum()
10042

df['reviews_per_month']
0      0.21
1      0.38
2      NaN
3      4.64
4      0.10
...
48890   NaN
48891   NaN
48892   NaN
48893   NaN
48894   NaN
Name: reviews_per_month, Length: 48879, dtype: float64

```

since the review are floating number and these column contain the rate for review for each month so we can actually replace it with mean

```

review_mean=df['reviews_per_month'].mean()

review_mean
1.3734098926281637

```



```
df['reviews_per_month']=df['reviews_per_month'].replace(np.nan,review_mean)
```

```
<ipython-input-135-678fea126995>:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
df['reviews_per_month']=df['reviews_per_month'].replace(np.nan,review_mean)
```

```
df['reviews_per_month']
```

```
0      0.21000  
1      0.38000  
2      1.37341  
3      4.64000  
4      0.10000
```

```
...  
48890    1.37341  
48891    1.37341  
48892    1.37341  
48893    1.37341  
48894    1.37341
```

```
Name: reviews_per_month, Length: 48879, dtype: float64
```

```
df['reviews_per_month'].isnull().sum()
```

```
0
```

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

```
id      0  
name    0  
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    0  
reviews_per_month    0  
calculated_host_listings_count    0
```

```

availability_365      0
dtype: int64

df['host_name']

0      John
1    Jennifer
2    Elisabeth
3    LisaRoxanne
4      Laura
...
48890    Sabrina
48891    Marisol
48892    Ilgar & Aysel
48893      Taz
48894    Christophe
Name: host_name, Length: 48879, dtype: object

```

since host name is specified to a specific identity so we will eliminate missing value

```

df=df.dropna(subset=['host_name'])
df['host_name'].isnull().sum()

0

df.isnull().sum()

id      0
name    0
host_id  0
host_name  0
neighbourhood_group  0
neighbourhood  0
latitude  0
longitude  0
room_type  0
price    0
minimum_nights  0
number_of_reviews  0
last_review  0
reviews_per_month  0
calculated_host_listings_count  0
availability_365  0
dtype: int64

```

since data is cleaned we move to next task to check duplicate value throughout the dataset

```

duplicate=df.duplicated()

```

```

duplicate
0      False
1      False
2      False
3      False
4      False
...
48890   False
48891   False
48892   False
48893   False
48894   False
Length: 48858, dtype: bool

dupli_row = df[duplicate]

dupli_row

{"repr_error": "Out of range float values are not JSON compliant:
nan", "type": "dataframe", "variable_name": "dupli_row"}

num_dupli = duplicate.sum()

num_dupli

0

```

Hence there are no duplicate value in the dataset

```

df.head()

{"summary": "{\n  \"name\": \"df\",\n  \"rows\": 48858,\n  \"fields\": [\n    {\n      \"column\": \"id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 10982893,\n        \"min\": 2539,\n        \"max\": 36487245,\n        \"num_unique_values\": 48858,\n        \"samples\": [\n          28539464,\n          13182297\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"name\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 47884,\n        \"samples\": [\n          \"Huge Western Room on the Top Floor of a Townhouse\",\n          \"Spacious and Comfy private room in Brooklyn Ctown\",\n          \"Modern Light Filled 1 Bedroom in East Williamsburg\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"host_id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 78623888,\n        \"min\": 2438,\n        \"max\": 274321313,\n        \"num_unique_values\": 37425,\n        \"samples\": [\n          105315839,\n          11774630,\n          22042865\n        ],\n        \"semantic_type\": \"\"

```

```

{"description": "", "host_name": "Amitis", "category": "Thikshan", "samples": ["Elen"], "semantic_type": "", "description": "", "neighbourhood_group": "Manhattan", "category": "Bronx", "samples": ["Queens"], "semantic_type": "", "description": "", "neighbourhood": "Stuyvesant Town", "category": "Eltingville", "samples": ["Briarwood"], "semantic_type": "", "description": "", "latitude": 40.054527703874443896, "min": 40.49979, "max": 40.91306, "num_unique_values": 19039, "samples": [40.6322, 40.77474, 40.81793], "semantic_type": "", "description": "", "longitude": 0.04615938872998777, "min": -74.24442, "max": -73.71299, "num_unique_values": 14716, "samples": [-73.93177, -73.86977, -73.97486], "semantic_type": "", "description": "", "room_type": "Private room", "num_unique_values": 3, "samples": ["Entire home/apt", "Shared room"], "semantic_type": "", "description": "", "price": 240, "min": 0, "max": 10000, "num_unique_values": 674, "samples": [519, 675, 488], "semantic_type": "", "description": "", "minimum_nights": 20, "min": 1, "max": 1250, "num_unique_values": 108, "samples": [160, 60, 2], "semantic_type": "", "description": "", "number_of_reviews": 44, "min": 0, "max": 629, "num_unique_values": 394, "samples": [12, 144, 314]

```

```

n      ],\n      \"semantic_type\": \"\", \n
\"description\": \"\" \n      },\n      {\n      \"column\":
\"last_review\", \n      \"properties\": {\n      \"dtype\":
\"category\", \n      \"num_unique_values\": 1765, \n
\"samples\": [\n      \"2018-03-11 00:00:00\", \n      \"2018-
05-21 00:00:00\", \n      \"2017-01-13 00:00:00\" \n      ], \n
\"semantic_type\": \"\", \n      \"description\": \"\" \n      } \n
      },\n      {\n      \"column\": \"reviews_per_month\", \n
\"properties\": {\n      \"dtype\": \"number\", \n      \"std\":
1.4978156780049712, \n      \"min\": 0.01, \n      \"max\": 58.5, \n
\"num_unique_values\": 938, \n      \"samples\": [\n      1.54, \n
4.33, \n      6.83 \n      ], \n      \"semantic_type\": \"\", \n
\"description\": \"\" \n      },\n      {\n      \"column\":
\"calculated_host_listings_count\", \n      \"properties\": {\n
\"dtype\": \"number\", \n      \"std\": 32, \n      \"min\": 1, \n
\"max\": 327, \n      \"num_unique_values\": 47, \n
\"samples\": [\n      37, \n      17, \n      121 \n
      ], \n      \"semantic_type\": \"\", \n
\"description\": \"\" \n      },\n      {\n      \"column\":
\"availability_365\", \n      \"properties\": {\n      \"dtype\":
\"number\", \n      \"std\": 131, \n      \"min\": 0, \n
\"max\": 365, \n      \"num_unique_values\": 366, \n
\"samples\": [\n      335, \n      309, \n      249 \n
      ], \n      \"semantic_type\": \"\", \n      \"description\": \"\" \n
      } \n      } \n      ] \n      }\", \"type\": \"dataframe\", \"variable_name\": \"df\"}

```

now lets change column name since they contain _ in between lets make them clean

```

header=['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']

df.columns=header

df.head()

{"summary": "{\n  \"name\": \"df\", \n  \"rows\": 48858, \n  \"fields\":
[\n    {\n      \"column\": \"Id\", \n      \"properties\": {\n
\"dtype\": \"number\", \n      \"std\": 10982893, \n      \"min\":
2539, \n      \"max\": 36487245, \n      \"num_unique_values\":
48858, \n      \"samples\": [\n      4285876, \n
28539464, \n      13182297 \n      ], \n      \"semantic_type\":
\"\", \n      \"description\": \"\" \n      }, \n      {\n
\"column\": \"Name\", \n      \"properties\": {\n      \"dtype\":
\"string\", \n      \"num_unique_values\": 47884, \n
\"samples\": [\n      \"Huge Western Room on the Top Floor of a
Townhouse\", \n      \"Spacious and Comfy private room in Brooklyn
Ctown\", \n      \"Modern Light Filled 1 Bedroom in East

```

```

Williamsburg\","],\n      \"semantic_type\": \"\",\n      \"description\": \"\",\n      \"column\": \"Host Id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 78623888,\n        \"min\": 2438,\n        \"max\": 274321313,\n        \"num_unique_values\": 37425,\n        \"samples\": [\n          105315839,\n          11774630,\n          22042865\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\",\n        \"column\": \"Host Name\",\n        \"properties\": {\n          \"dtype\": \"category\",\n          \"num_unique_values\": 11450,\n          \"samples\": [\n            \"Amitis\",\n            \"Thikshan\",\n            \"Elen\"\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\",\n          \"column\": \"Neighbourhood Group\",\n          \"properties\": {\n            \"dtype\": \"category\",\n            \"num_unique_values\": 5,\n            \"samples\": [\n              \"Manhattan\",\n              \"Bronx\",\n              \"Queens\"\n            ],\n            \"semantic_type\": \"\",\n            \"description\": \"\",\n            \"column\": \"Neighbourhood\",\n            \"properties\": {\n              \"dtype\": \"category\",\n              \"num_unique_values\": 221,\n              \"samples\": [\n                \"Stuyvesant Town\",\n                \"Eltingville\",\n                \"Briarwood\"\n              ],\n              \"semantic_type\": \"\",\n              \"description\": \"\",\n              \"column\": \"Latitude\",\n              \"properties\": {\n                \"dtype\": \"number\",\n                \"std\": 0.054527703874443896,\n                \"min\": 40.49979,\n                \"max\": 40.91306,\n                \"num_unique_values\": 19039,\n                \"samples\": [\n                  40.6322,\n                  40.77474,\n                  40.81793\n                ],\n                \"semantic_type\": \"\",\n                \"description\": \"\",\n                \"column\": \"Longitude\",\n                \"properties\": {\n                  \"dtype\": \"number\",\n                  \"std\": 0.04615938872998777,\n                  \"min\": -74.24442,\n                  \"max\": -73.71299,\n                  \"num_unique_values\": 14716,\n                  \"samples\": [\n                    -73.93177,\n                    -73.86977,\n                    -73.97486\n                  ],\n                  \"semantic_type\": \"\",\n                  \"description\": \"\",\n                  \"column\": \"Room Type\",\n                  \"properties\": {\n                    \"dtype\": \"category\",\n                    \"num_unique_values\": 3,\n                    \"samples\": [\n                      \"Private room\",\n                      \"Entire home/apt\",\n                      \"Shared room\"\n                    ],\n                    \"semantic_type\": \"\",\n                    \"description\": \"\",\n                    \"column\": \"Price\",\n                    \"properties\": {\n                      \"dtype\": \"number\",\n                      \"std\": 240,\n                      \"min\": 0,\n                      \"max\": 10000,\n                      \"num_unique_values\": 674,\n                      \"samples\": [\n                        519,\n                        675,\n                        488\n                      ],\n                      \"semantic_type\": \"\",\n                      \"description\": \"\",\n                      \"column\": \"Minimum Nights\",\n                      \"properties\": {\n                        \"dtype\": \"number\",\n                        \"std\": 20,\n                        \"min\": 1,\n                        \"max\": 1250,\n                        \"num unique values\": 108,\n
```

```

\"samples\": [\n          160,\n          60,\n          2\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\",\n        \"column\":\n        \"Number Of Reviews\",\n        \"properties\": {\n          \"dtype\":\n          \"number\",\n          \"std\": 44,\n          \"min\": 0,\n          \"max\": 629,\n          \"num_unique_values\": 394,\n        },\n        \"samples\": [\n          12,\n          144,\n          314\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\",\n        \"column\":\n        \"Last Review\",\n        \"properties\": {\n          \"dtype\":\n          \"category\",\n          \"num_unique_values\": 1765,\n        },\n        \"samples\": [\n          \"2018-03-11 00:00:00\",\n          \"2018-05-21 00:00:00\",\n          \"2017-01-13 00:00:00\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\",\n        \"column\":\n        \"Reviews Per Month\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\":\n          1.4978156780049712,\n          \"min\": 0.01,\n          \"max\": 58.5,\n          \"num_unique_values\": 938,\n        },\n        \"samples\": [\n          1.54,\n          4.33,\n          6.83\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\",\n        \"column\":\n        \"Calculated Host Listings Count\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 32,\n          \"min\": 1,\n          \"max\": 327,\n          \"num_unique_values\": 47,\n        },\n        \"samples\": [\n          37,\n          17,\n          121\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\",\n        \"column\":\n        \"Availability 365\",\n        \"properties\": {\n          \"dtype\":\n          \"number\",\n          \"std\": 131,\n          \"min\": 0,\n          \"max\": 365,\n          \"num_unique_values\": 366,\n        },\n        \"samples\": [\n          335,\n          309,\n          249\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n    ],\n  },\n  \"type\": \"dataframe\", \"variable_name\": \"df\"}

```

```
from sklearn.preprocessing import MinMaxScaler
```

```
mm=MinMaxScaler()
```

```
df['price_mm']=mm.fit_transform(df[['Price']])
```

```
df['price_mm']
```

```

0      0.0149
1      0.0225
2      0.0150
3      0.0089
4      0.0080
...
48890  0.0070
48891  0.0040
48892  0.0115

```

```
48893    0.0055
48894    0.0090
Name: price_mm, Length: 48858, dtype: float64
```

```
df.head(10)
```

```
{"summary":{"\n  \"name\": \"df\",\n  \"rows\": 48858,\n  \"fields\": [\n    {\n      \"column\": \"Id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 10982893,\n        \"min\": 2539,\n        \"max\": 36487245,\n        \"num_unique_values\": 48858,\n        \"samples\": [\n          28539464,\n          13182297\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Name\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 47884,\n        \"samples\": [\n          \"Huge Western Room on the Top Floor of a Townhouse\",\n          \"Spacious and Comfy private room in Brooklyn Ctown\",\n          \"Modern Light Filled 1 Bedroom in East Williamsburg\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Host Id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 78623888,\n        \"min\": 2438,\n        \"max\": 274321313,\n        \"num_unique_values\": 37425,\n        \"samples\": [\n          105315839,\n          11774630,\n          22042865\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Host Name\",\n      \"properties\": {\n        \"dtype\": \"category\",\n        \"num_unique_values\": 11450,\n        \"samples\": [\n          \"Amitis\",\n          \"Thikshan\",\n          \"Elen\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Neighbourhood Group\",\n      \"properties\": {\n        \"dtype\": \"category\",\n        \"num_unique_values\": 5,\n        \"samples\": [\n          \"Manhattan\",\n          \"Bronx\",\n          \"Queens\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Neighbourhood\",\n      \"properties\": {\n        \"dtype\": \"category\",\n        \"num_unique_values\": 221,\n        \"samples\": [\n          \"Stuyvesant Town\",\n          \"Eltingville\",\n          \"Briarwood\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Latitude\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 0.054527703874443896,\n        \"min\": 40.49979,\n        \"max\": 40.91306,\n        \"num_unique_values\": 19039,\n        \"samples\": [\n          40.6322,\n          40.77474,\n          40.81793\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Longitude\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 0.04615938872998777,\n        \"min\": -74.24442,\n        \"max\": -
```



```

73.71299,\n          \"num_unique_values\": 14716,\n          \"samples\": [\n\n            -73.93177,\n            -73.86977,\n            -73.97486\n          ],\n          \"semantic_type\": \"\", \n          \"description\": \"\"\n        },\n        {\n          \"column\": \"Room Type\", \n          \"properties\": {\n            \"dtype\": \"category\", \n            \"num_unique_values\": 3,\n            \"samples\": [\n              \"Private room\", \n              \"Entire home/apt\", \n              \"Shared room\"\n            ],\n            \"semantic_type\": \"\", \n            \"description\": \"\"\n          },\n          {\n            \"column\": \"Price\", \n            \"properties\": {\n              \"dtype\": \"number\", \n              \"std\": 240,\n              \"min\": 0,\n              \"max\": 10000,\n              \"num_unique_values\": 674,\n              \"samples\": [\n                519,\n                675,\n                488\n              ],\n              \"semantic_type\": \"\", \n              \"description\": \"\"\n            },\n            {\n              \"column\": \"Minimum Nights\", \n              \"properties\": {\n                \"dtype\": \"number\", \n                \"std\": 20,\n                \"min\": 1,\n                \"max\": 1250,\n                \"num_unique_values\": 108,\n                \"samples\": [\n                  160,\n                  60,\n                  2\n                ],\n                \"semantic_type\": \"\", \n                \"description\": \"\"\n              },\n              {\n                \"column\": \"Number Of Reviews\", \n                \"properties\": {\n                  \"dtype\": \"number\", \n                  \"std\": 44,\n                  \"min\": 0,\n                  \"max\": 629,\n                  \"num_unique_values\": 394,\n                  \"samples\": [\n                    12,\n                    144,\n                    314\n                  ],\n                  \"semantic_type\": \"\", \n                  \"description\": \"\"\n                },\n                {\n                  \"column\": \"Last Review\", \n                  \"properties\": {\n                    \"dtype\": \"category\", \n                    \"num_unique_values\": 1765,\n                    \"samples\": [\n                      \"2018-03-11 00:00:00\", \n                      \"2018-05-21 00:00:00\", \n                      \"2017-01-13 00:00:00\"\n                    ],\n                    \"semantic_type\": \"\", \n                    \"description\": \"\"\n                  },\n                  {\n                    \"column\": \"Reviews Per Month\", \n                    \"properties\": {\n                      \"dtype\": \"number\", \n                      \"std\": 1.4978156780049712,\n                      \"min\": 0.01,\n                      \"max\": 58.5,\n                      \"num_unique_values\": 938,\n                      \"samples\": [\n                        1.54,\n                        4.33,\n                        6.83\n                      ],\n                      \"semantic_type\": \"\", \n                      \"description\": \"\"\n                    },\n                    {\n                      \"column\": \"Calculated Host Listings Count\", \n                      \"properties\": {\n                        \"dtype\": \"number\", \n                        \"std\": 32,\n                        \"min\": 1,\n                        \"max\": 327,\n                        \"num_unique_values\": 47,\n                        \"samples\": [\n                          37,\n                          17,\n                          121\n                        ],\n                        \"semantic_type\": \"\", \n                        \"description\": \"\"\n                      },\n                      {\n                        \"column\": \"Availability 365\", \n                        \"properties\": {\n                          \"dtype\": \"number\", \n                          \"std\": 131,\n                          \"min\": 0,\n                          \"max\": 365,\n                          \"num_unique_values\": 366,\n                          \"samples\": [\n                            335,\n                            309,\n                            249\n                          ],\n                          \"semantic_type\": \"\", \n                          \"description\": \"\"\n                        },\n                        {\n                          \"column\": \"price_mm\", \n                          \"properties\": {\n                            \"dtype\": \"number\", \n                            \"std\": 11.5,\n                            \"min\": 0,\n                            \"max\": 100,\n                            \"num_unique_values\": 100,\n                            \"samples\": [\n                              11.5,\n                              10,\n                              10\n                            ],\n                            \"semantic_type\": \"\", \n                            \"description\": \"\"\n                          }\n                        }\n                      }\n                    }\n                  }\n                }\n              }\n            }\n          }\n        }\n      }\n    }\n  }\n}

```

```

{"properties\": {\n      \"dtype\": \"number\", \n      \"std\": 0.02402323855251984, \n      \"min\": 0.0, \n      \"max\": 1.0, \n      \"num_unique_values\": 674, \n      \"samples\": [\n        0.0519, \n        0.0675, \n        0.0488\n      ], \n      \"semantic_type\": \"\", \n      \"description\": \"\"\n    }\n  }\n], \"type\": \"dataframe\", \"variable_name\": \"df\"}

```

```
df.describe()
```

```

{"summary": "{\n  \"name\": \"df\", \n  \"rows\": 8, \n  \"fields\": [\n    {\n      \"column\": \"Id\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 13052000.972540405, \n        \"min\": 2539.0, \n        \"max\": 36487245.0, \n        \"num_unique_values\": 8, \n        \"samples\": [\n          19023349.934565477, \n          19691143.5, \n          48858.0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Host Id\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 91354574.9132282, \n        \"min\": 2438.0, \n        \"max\": 274321313.0, \n        \"num_unique_values\": 8, \n        \"samples\": [\n          67631688.28595112, \n          30791331.0, \n          48858.0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Latitude\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 17261.574743766392, \n        \"min\": 0.054527703874443896, \n        \"max\": 48858.0, \n        \"num_unique_values\": 8, \n        \"samples\": [\n          40.72894117462851, \n          40.72307, \n          48858.0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Longitude\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 17296.343123442795, \n        \"min\": -74.24442, \n        \"max\": 48858.0, \n        \"num_unique_values\": 8, \n        \"samples\": [\n          -73.95217028736339, \n          -73.95568, \n          48858.0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Price\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 17084.65198416182, \n        \"min\": 0.0, \n        \"max\": 48858.0, \n        \"num_unique_values\": 8, \n        \"samples\": [\n          152.74030864955586, \n          106.0, \n          48858.0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Minimum Nights\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 17214.410981803765, \n        \"min\": 1.0, \n        \"max\": 48858.0, \n        \"num_unique_values\": 7, \n        \"samples\": [\n          48858.0, \n          7.012444226124688, \n          5.0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Number Of Reviews\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 17238.541119040983, \n        \"min\":

```

```

0.0,\n          \"max\": 48858.0,\n          \"num_unique_values\": 8,\n\"samples\": [\n          23.273097547996233,\n          5.0,\n48858.0\n          ],\n          \"semantic_type\": \"\",\n\"description\": \"\"\n          }\n          },\n          {\n          \"column\":\n\"Reviews Per Month\",\n          \"properties\": {\n          \"dtype\":\n\"number\",\n          \"std\": 17270.66751399549,\n          \"min\":\n0.01,\n          \"max\": 48858.0,\n          \"num_unique_values\": 8,\n\"samples\": [\n          1.3732663042349025,\n          1.22,\n48858.0\n          ],\n          \"semantic_type\": \"\",\n\"description\": \"\"\n          }\n          },\n          {\n          \"column\":\n\"Calculated Host Listings Count\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 17255.482503072184,\n          \"min\": 1.0,\n          \"max\": 48858.0,\n          \"num_unique_values\": 6,\n          \"samples\": [\n          48858.0,\n          7.148368742068852,\n          327.0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          },\n          {\n          \"column\": \"Availability 365\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\":\n17229.832291421804,\n          \"min\": 0.0,\n          \"max\": 48858.0,\n          \"num_unique_values\": 7,\n          \"samples\": [\n          48858.0,\n          112.80142453641164,\n          227.0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          },\n          {\n          \"column\": \"price_mm\",\n          \"properties\":\n{\n          \"dtype\": \"number\",\n          \"std\":\n17273.85730085874,\n          \"min\": 0.0,\n          \"max\": 48858.0,\n          \"num_unique_values\": 8,\n          \"samples\": [\n          0.015274030864955587,\n          0.0106,\n          48858.0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          }\n          ]\n          }\", \"type\": \"dataframe\"}

```

```

selected=['Latitude','Longitude','price_mm','Reviews Per
Month','Number Of Reviews']

```

```

plt.figure(figsize=(12, 8)) # Set the figure size
num_rows = 2
num_cols = 3

```

```

for i, column in enumerate(selected, 1):
    plt.subplot(num_rows, num_cols, i)
    sns.boxplot(y=df[column],palette='rocket')
    plt.title(f'Box Plot of {column}')

```

```

plt.tight_layout()
plt.show()

```

<ipython-input-160-dfd262cca7f9>:7: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(y=df[column],palette='rocket')  
<ipython-input-160-dfd262cca7f9>:7: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(y=df[column],palette='rocket')  
<ipython-input-160-dfd262cca7f9>:7: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

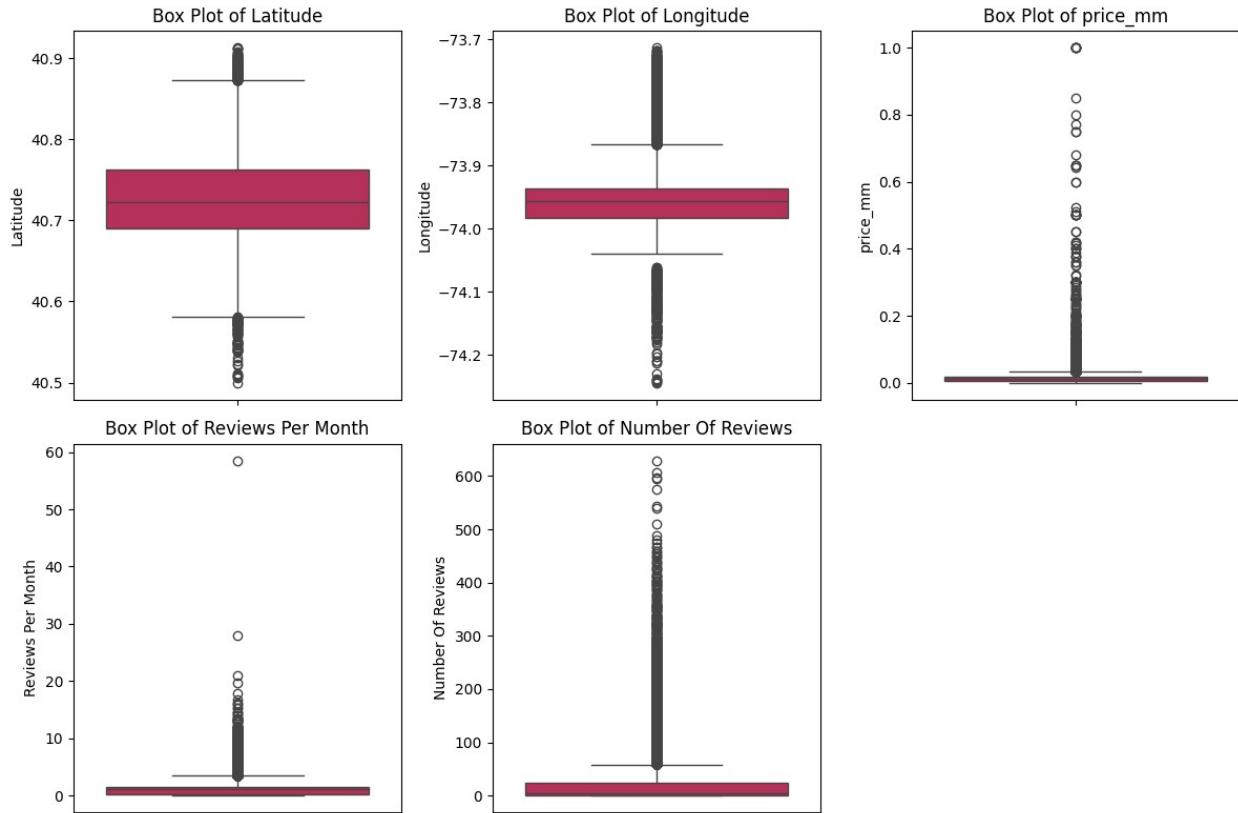
```
sns.boxplot(y=df[column],palette='rocket')  
<ipython-input-160-dfd262cca7f9>:7: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(y=df[column],palette='rocket')  
<ipython-input-160-dfd262cca7f9>:7: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.boxplot(y=df[column],palette='rocket')
```



```
q1 = df['price_mm'].quantile(0.25)
q3 = df['price_mm'].quantile(0.75)
iqr = q3 - q1
lower_bound = q1 - 1.5 * iqr
upper_bound = q3 + 1.5 * iqr

q1,q3,iqr

(0.0069000000000000001, 0.0175, 0.0106000000000000002)

lower_bound,upper_bound

(-0.0090000000000000005, 0.033400000000000006)

outliers = df[(df['price_mm'] > upper_bound) | (df['price_mm'] <
lower_bound)]

outliers

{"summary":{"\n  "name": "outliers",\n  "rows": 2971,\n  "fields": [\n    {\n      "column": "Id",\n      "properties":\n      {\n        "dtype": "number",\n        "std": 11774597,\n        "min": 15396,\n        "max": 36457700,\n        "num_unique_values": 2971,\n        "samples": [\n          33660966,\n          18907166,\n          33689606\n        ],\n        "semantic_type": "\",\n        "description": "\"\"\""}\n  ]}
```

```

\n      },\n      {\n        \"column\": \"Name\", \n        \"properties\": {\n          \"dtype\": \"string\", \n          \"num_unique_values\": 2950, \n          \"samples\": [\n            \"Brooklyn Factory Loft\", \n            \"Chelsea Studio for Super Bowl!\", \n            \"Gorgeous Midtown West LOFT 3BEDS/1.5BATH XL\"\n          ], \n          \"semantic_type\": \"\", \n          \"description\": \"\" \n        }, \n        {\n          \"column\": \"Host Id\", \n          \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 91101066, \n            \"min\": 8730, \n            \"max\": 274079964, \n            \"num_unique_values\": 2446, \n            \"samples\": [\n              204537864, \n              153735730, \n              22129776\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\" \n          }, \n          {\n            \"column\": \"Host Name\", \n            \"properties\": {\n              \"dtype\": \"category\", \n              \"num_unique_values\": 1444, \n              \"samples\": [\n                \"B\", \n                \"Jason\", \n                \"Charlie\"\n              ], \n              \"semantic_type\": \"\", \n              \"description\": \"\" \n            }, \n            {\n              \"column\": \"Neighbourhood Group\", \n              \"properties\": {\n                \"dtype\": \"category\", \n                \"num_unique_values\": 5, \n                \"samples\": [\n                  \"Brooklyn\", \n                  \"Bronx\", \n                  \"Queens\"\n                ], \n                \"semantic_type\": \"\", \n                \"description\": \"\" \n              }, \n              {\n                \"column\": \"Neighbourhood\", \n                \"properties\": {\n                  \"dtype\": \"category\", \n                  \"num_unique_values\": 114, \n                  \"samples\": [\n                    \"Williamsbridge\", \n                    \"South Slope\", \n                    \"Cobble Hill\"\n                  ], \n                  \"semantic_type\": \"\", \n                  \"description\": \"\" \n                }, \n                {\n                  \"column\": \"Latitude\", \n                  \"properties\": {\n                    \"dtype\": \"number\", \n                    \"std\": 0.03950426241223145, \n                    \"min\": 40.53076, \n                    \"max\": 40.88671, \n                    \"num_unique_values\": 2607, \n                    \"samples\": [\n                      40.75615, \n                      40.85029, \n                      40.684\n                    ], \n                    \"semantic_type\": \"\", \n                    \"description\": \"\" \n                  }, \n                  {\n                    \"column\": \"Longitude\", \n                    \"properties\": {\n                      \"dtype\": \"number\", \n                      \"std\": 0.03401027045076513, \n                      \"min\": -74.20295, \n                      \"max\": -73.73529, \n                      \"num_unique_values\": 2478, \n                      \"samples\": [\n                        -73.97601, \n                        -74.01065, \n                        -73.98486\n                      ], \n                      \"semantic_type\": \"\", \n                      \"description\": \"\" \n                    }, \n                    {\n                      \"column\": \"Room Type\", \n                      \"properties\": {\n                        \"dtype\": \"category\", \n                        \"num_unique_values\": 3, \n                        \"samples\": [\n                          \"Entire home/apt\", \n                          \"Private room\", \n                          \"Shared room\"\n                        ], \n                        \"semantic_type\": \"\", \n                        \"description\": \"\" \n                      }, \n                      {\n                        \"column\": \"Price\", \n                        \"properties\": {\n                          \"dtype\": \"number\", \n                          \"std\": 777, \n                          \"min\": 335, \n                          \"max\": 10000, \n                          \"num_unique_values\": 352, \n                          \"samples\": [\n                            8500, \n                            680, \n                            1100\n                          ], \n                          \"semantic_type\": \"\", \n                        }\n                      }\n                    }\n                  }\n                }\n              }\n            }\n          }\n        }\n      }\n    }\n  }\n}

```



```

{"summary":{"\n  \"name\": \"n_df\", \n  \"rows\": 45887, \n  \"fields\": [\n    {\n      \"column\": \"Id\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 10919192, \n        \"min\": 2539, \n        \"max\": 36487245, \n        \"num_unique_values\": 45887, \n        \"samples\": [\n          532749, \n          27527902, \n          19236716\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\" \n      }, \n      {\n        \"column\": \"Name\", \n        \"properties\": {\n          \"dtype\": \"string\", \n          \"num_unique_values\": 44971, \n          \"samples\": [\n            \"Dashing 1BR in Midtown East by Sonder\", \n            \"One bedroom apartment in mid-town\", \n            \"Williamsburg Duplex\" \n          ], \n          \"semantic_type\": \"\", \n          \"description\": \"\" \n        }, \n        {\n          \"column\": \"Host Id\", \n          \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 77570721, \n            \"min\": 2438, \n            \"max\": 274321313, \n            \"num_unique_values\": 35358, \n            \"samples\": [\n              8748472, \n              6645096, \n              67567916\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\" \n          }, \n          {\n            \"column\": \"Host Name\", \n            \"properties\": {\n              \"dtype\": \"category\", \n              \"num_unique_values\": 11006, \n              \"samples\": [\n                \"Fernando And Lenin\", \n                \"Jenya\", \n                \"Abby & Kurt\" \n              ], \n              \"semantic_type\": \"\", \n              \"description\": \"\" \n            }, \n            {\n              \"column\": \"Neighbourhood Group\", \n              \"properties\": {\n                \"dtype\": \"category\", \n                \"num_unique_values\": 5, \n                \"samples\": [\n                  \"Manhattan\", \n                  \"Bronx\", \n                  \"Queens\" \n                ], \n                \"semantic_type\": \"\", \n                \"description\": \"\" \n              }, \n              {\n                \"column\": \"Neighbourhood\", \n                \"properties\": {\n                  \"dtype\": \"category\", \n                  \"num_unique_values\": 219, \n                  \"samples\": [\n                    \"West Brighton\", \n                    \"Briarwood\", \n                    \"Breezy Point\" \n                  ], \n                  \"semantic_type\": \"\", \n                  \"description\": \"\" \n                }, \n                {\n                  \"column\": \"Latitude\", \n                  \"properties\": {\n                    \"dtype\": \"number\", \n                    \"std\": 0.055328702281769976, \n                    \"min\": 40.49979, \n                    \"max\": 40.91306, \n                    \"num_unique_values\": 18749, \n                    \"samples\": [\n                      40.80146, \n                      40.86868, \n                      40.71852\n                    ], \n                    \"semantic_type\": \"\", \n                    \"description\": \"\" \n                  }, \n                  {\n                    \"column\": \"Longitude\", \n                    \"properties\": {\n                      \"dtype\": \"number\", \n                      \"std\": 0.046474279041928655, \n                      \"min\": -74.24442, \n                      \"max\": -73.71299, \n                      \"num_unique_values\": 14532, \n                      \"samples\": [\n                        -73.94884, \n                        -73.98914, \n                        -73.77109\n                      ], \n                      \"semantic_type\": \"\", \n                      \"description\": \"\" \n                    }, \n                    {\n                      \"column\": \"Room Type\", \n                      \"properties\": {\n                        \"dtype\": \"category\", \n                        \"num_unique_values\": 3, \n                        \"samples\": [\n                          \"Private

```



```
room",\n      {\n        "column": "Entire home/apt",\n        "properties": {\n          "dtype": "number",\n          "std": 68,\n          "min": 0,\n          "max": 334,\n          "num_unique_values": 322,\n          "samples": [\n            197,\n            119,\n            173\n          ],\n          "semantic_type": "\n",\n          "description": "\n"}\n    },\n    {\n      "column": "Price",\n      "properties": {\n        "dtype": "number",\n        "std": 19,\n        "min": 1,\n        "max": 1250,\n        "num_unique_values": 105,\n        "samples": [\n          20,\n          23,\n          62\n        ],\n        "semantic_type": "\n",\n        "description": "\n"}\n    },\n    {\n      "column": "Minimum Nights",\n      "properties": {\n        "dtype": "number",\n        "std": 45,\n        "min": 0,\n        "max": 629,\n        "num_unique_values": 393,\n        "samples": [\n          38,\n          439,\n          153\n        ],\n        "semantic_type": "\n",\n        "description": "\n"}\n    },\n    {\n      "column": "Number Of Reviews",\n      "properties": {\n        "dtype": "number",\n        "std": 1753,\n        "min": 0,\n        "max": 2018-10-22 00:00:00,\n        "num_unique_values": 937,\n        "samples": [\n          1.7,\n          0.28,\n          2.33\n        ],\n        "semantic_type": "\n",\n        "description": "\n"}\n    },\n    {\n      "column": "Last Review",\n      "properties": {\n        "dtype": "number",\n        "std": 31,\n        "min": 1,\n        "max": 327,\n        "num_unique_values": 47,\n        "samples": [\n          37,\n          17,\n          121\n        ],\n        "semantic_type": "\n",\n        "description": "\n"}\n    },\n    {\n      "column": "Category",\n      "properties": {\n        "dtype": "number",\n        "std": 130,\n        "min": 0,\n        "max": 365,\n        "num_unique_values": 366,\n        "samples": [\n          315,\n          309,\n          249\n        ],\n        "semantic_type": "\n",\n        "description": "\n"}\n    },\n    {\n      "column": "Availability 365",\n      "properties": {\n        "dtype": "number",\n        "std": 0.006814543836984588,\n        "min": 0.0,\n        "max": 0.0334,\n        "num_unique_values": 322,\n        "samples": [\n          0.019700000000000002,\n          0.0119,\n          0.0173\n        ],\n        "semantic_type": "\n",\n        "description": "\n"}\n    },\n    {\n      "column": "Calculated Host Listings Count",\n      "properties": {\n        "dtype": "number",\n        "std": 0.006814543836984588,\n        "min": 0.0,\n        "max": 0.0334,\n        "num_unique_values": 322,\n        "samples": [\n          0.019700000000000002,\n          0.0119,\n          0.0173\n        ],\n        "semantic_type": "\n",\n        "description": "\n"}\n    },\n    {\n      "column": "Reviews Per Month",\n      "properties": {\n        "dtype": "number",\n        "std": 1.5169661417575975,\n        "min": 0.01,\n        "max": 58.5,\n        "num_unique_values": 937,\n        "samples": [\n          1.7,\n          0.28,\n          2.33\n        ],\n        "semantic_type": "\n",\n        "description": "\n"}\n    },\n    {\n      "column": "price_mm",\n      "properties": {\n        "dtype": "number",\n        "std": 0.006814543836984588,\n        "min": 0.0,\n        "max": 0.0334,\n        "num_unique_values": 322,\n        "samples": [\n          0.019700000000000002,\n          0.0119,\n          0.0173\n        ],\n        "semantic_type": "\n",\n        "description": "\n"}\n    }\n  ],\n  "shared_home/apt": {\n    "column": "Shared home/apt",\n    "properties": {\n      "dtype": "number",\n      "std": 68,\n      "min": 0,\n      "max": 334,\n      "num_unique_values": 322,\n      "samples": [\n        197,\n        119,\n        173\n      ],\n      "semantic_type": "\n",\n      "description": "\n"}\n  }\n}
```

```
\["description\": \["\n      }\n    ]\n  ], "type": "dataframe", "variable_name": "n_df"}
```

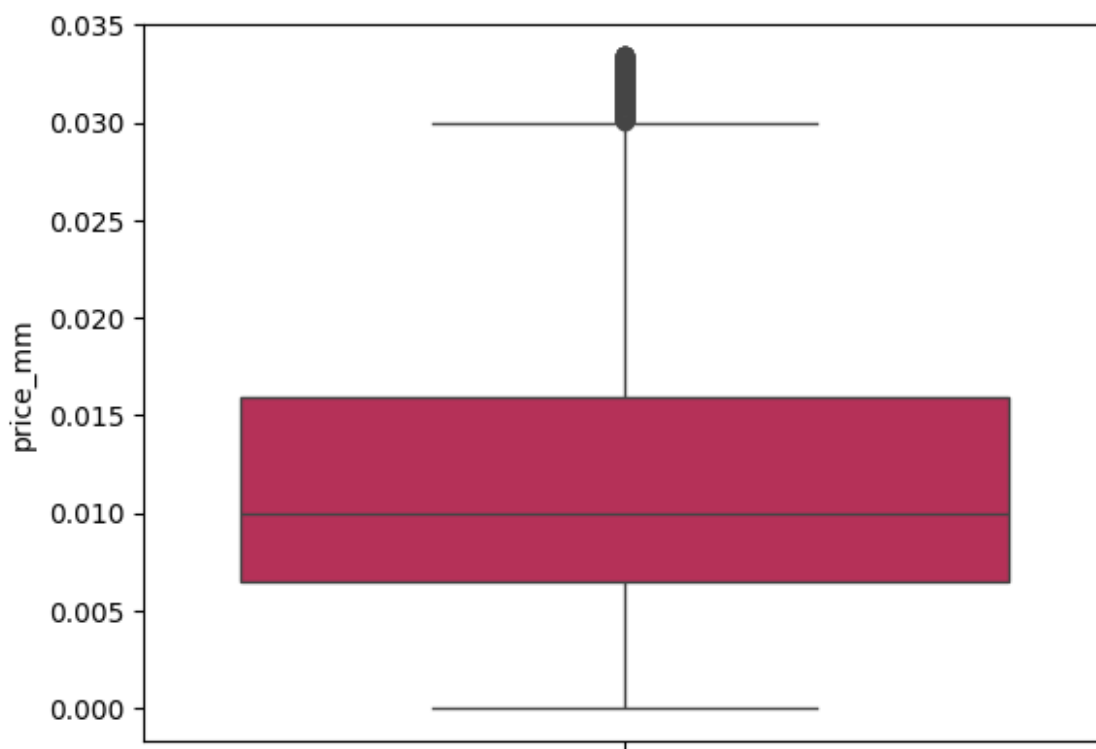
```
sns.boxplot(y=n_df['price_mm'],palette='rocket')
```

```
<ipython-input-168-196b2ed5d206>:1: FutureWarning:
```

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

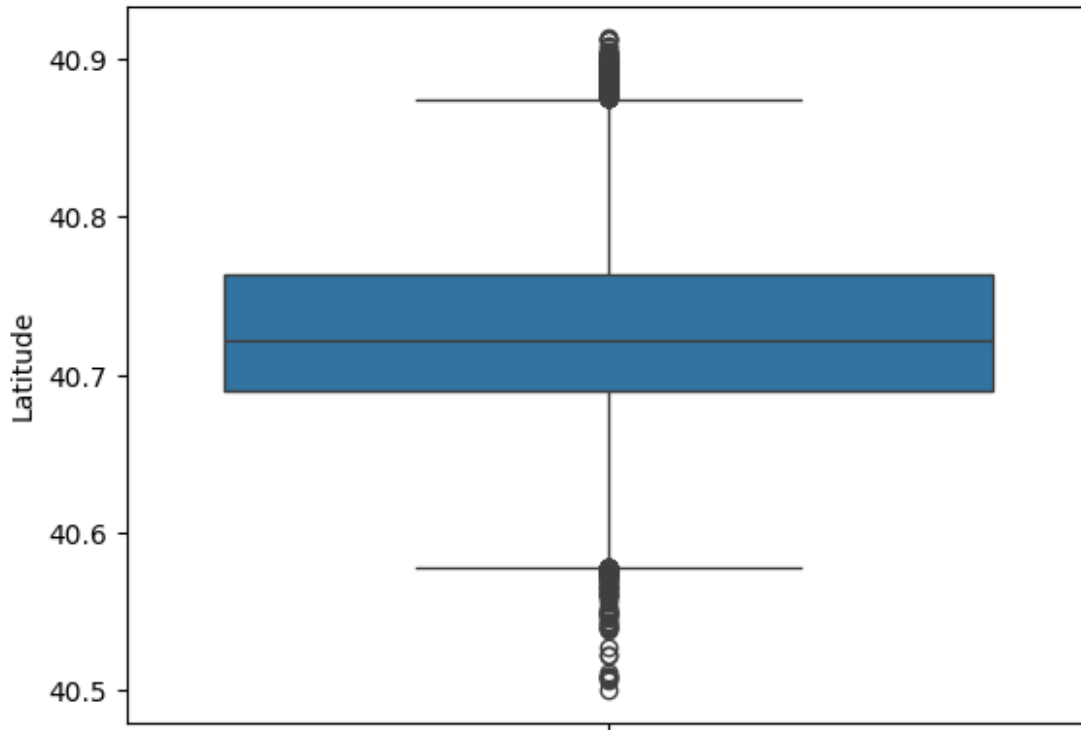
```
sns.boxplot(y=n_df['price_mm'],palette='rocket')
```

```
<Axes: ylabel='price_mm'>
```



```
sns.boxplot(n_df['Latitude'])
```

```
<Axes: ylabel='Latitude'>
```



```

q1 = n_df['Latitude'].quantile(0.25)
q3 = n_df['Latitude'].quantile(0.75)
iqr = q3 - q1
lower_bound = q1 - 1.5 * iqr
upper_bound = q3 + 1.5 * iqr

q1,q3,iqr
(40.68923, 40.763385, 0.07415499999999753)

upper_bound,lower_bound
(40.8746175, 40.577997500000001)

outliers_lat = n_df[(n_df['Latitude'] > upper_bound) |
(n_df['Latitude'] < lower_bound)]

outliers_lat

{"summary":{"\n  \"name\": \"outliers_lat\",\n  \"rows\": 346,\n  \"fields\": [\n    {\n      \"column\": \"Id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 10144251,\n        \"min\": 63610,\n        \"max\": 36438336,\n        \"num_unique_values\": 346,\n        \"samples\": [\n          16892955,\n          22087093,\n          32744748\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Name\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 345,\n
```

```

\"samples\": [\n          \"Gorgeous Private room near NYC\", \n
\"*Spacious NYC gem* Direct train to central sights!\", \n
\"Bright 2- Room apartment in a house next to train.\", \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n
n      }, \n      {\n          \"column\": \"Host Id\", \n          \"properties\": {\n
n              \"dtype\": \"number\", \n              \"std\": 80140593, \n
\"min\": 35944, \n              \"max\": 273672846, \n
\"num_unique_values\": 271, \n              \"samples\": [\n
21444167, \n              28190733, \n              28189446 \n              ], \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n
n      }, \n      {\n          \"column\": \"Host Name\", \n
\"properties\": {\n              \"dtype\": \"string\", \n
\"num_unique_values\": 245, \n              \"samples\": [\n
\"Ayesha\", \n              \"Claudia\", \n              \"Camille\" \n
n          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      }, \n      {\n          \"column\":
\"Neighbourhood Group\", \n          \"properties\": {\n              \"dtype\":
\"category\", \n              \"num_unique_values\": 5, \n              \"samples\":
[\n              \"Staten Island\", \n              \"Manhattan\", \n
\"Brooklyn\" \n              ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      }, \n      {\n          \"column\":
\"Neighbourhood\", \n          \"properties\": {\n              \"dtype\":
\"category\", \n              \"num_unique_values\": 35, \n              \"samples\": [\n
\"Olinville\", \n              \"Lighthouse Hill\", \n
\"Oakwood\" \n              ], \n          \"semantic_type\":
\"\", \n          \"description\": \"\" \n      }, \n      {\n
\"column\": \"Latitude\", \n          \"properties\": {\n              \"dtype\":
\"number\", \n              \"std\": 0.15356401466283245, \n              \"min\":
40.49979, \n              \"max\": 40.91306, \n              \"num_unique_values\":
332, \n              \"samples\": [\n              40.87829, \n              40.5771, \n
40.89984 \n              ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      }, \n      {\n          \"column\":
\"Longitude\", \n          \"properties\": {\n              \"dtype\":
\"number\", \n              \"std\": 0.10483924449624134, \n              \"min\": -
74.24442, \n              \"max\": -73.83186, \n              \"num_unique_values\":
339, \n              \"samples\": [\n              -74.1366, \n              -
73.91164, \n              -73.83434 \n              ], \n          \"semantic_type\":
\"\", \n          \"description\": \"\" \n      }, \n      {\n
\"column\": \"Room Type\", \n          \"properties\": {\n              \"dtype\":
\"category\", \n              \"num_unique_values\": 3, \n              \"samples\": [\n
\"Entire home/apt\", \n              \"Private room\", \n
\"Shared room\" \n              ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      }, \n      {\n          \"column\": \"Price\", \n
\"properties\": {\n              \"dtype\": \"number\", \n              \"std\": 58, \n
\"min\": 25, \n              \"max\": 309, \n              \"num_unique_values\":
101, \n              \"samples\": [\n              101, \n              84, \n
42 \n              ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      }, \n      {\n          \"column\":

```

```

\"Minimum Nights\", \n      \"properties\": { \n          \"dtype\":
\"number\", \n          \"std\": 23, \n          \"min\": 1, \n
\"max\": 365, \n          \"num_unique_values\": 18, \n
\"samples\": [ \n              2, \n              1, \n              28 \n          ], \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n      } \n
      }, \n      { \n          \"column\": \"Number Of Reviews\", \n
\"properties\": { \n          \"dtype\": \"number\", \n          \"std\":
35, \n          \"min\": 0, \n          \"max\": 205, \n
\"num_unique_values\": 80, \n          \"samples\": [ \n              8, \n
38, \n              1 \n          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      } \n      }, \n      { \n          \"column\":
\"Last Review\", \n          \"properties\": { \n          \"dtype\":
\"category\", \n          \"num_unique_values\": 133, \n
\"samples\": [ \n              \"2016-08-10 00:00:00\", \n              \"2019-
06-22 00:00:00\", \n              \"2019-04-20 00:00:00\" \n          ], \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n      } \n
      }, \n      { \n          \"column\": \"Reviews Per Month\", \n
\"properties\": { \n          \"dtype\": \"number\", \n          \"std\":
1.4936227975965681, \n          \"min\": 0.02, \n          \"max\": 9.33, \n
\"num_unique_values\": 167, \n          \"samples\": [ \n              4.93, \n
0.58, \n              2.6 \n          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      } \n      }, \n      { \n          \"column\":
\"Calculated Host Listings Count\", \n          \"properties\": { \n
\"dtype\": \"number\", \n          \"std\": 2, \n          \"min\": 1, \n
\"max\": 37, \n          \"num_unique_values\": 8, \n          \"samples\":
[ \n              1, \n              6, \n              13 \n          ], \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n      } \n
      }, \n      { \n          \"column\": \"Availability 365\", \n
\"properties\": { \n          \"dtype\": \"number\", \n          \"std\":
138, \n          \"min\": 0, \n          \"max\": 365, \n
\"num_unique_values\": 166, \n          \"samples\": [ \n              178, \n
350, \n              62 \n          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      } \n      }, \n      { \n          \"column\":
\"price_mm\", \n          \"properties\": { \n          \"dtype\":
\"number\", \n          \"std\": 0.005833278794369201, \n          \"min\":
0.0025, \n          \"max\": 0.0309, \n          \"num_unique_values\":
101, \n          \"samples\": [ \n              0.010100000000000001, \n
0.008400000000000001, \n              0.004200000000000001 \n          ], \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n      } \n
      } \n  ] \n } \", \"type\": \"dataframe\", \"variable_name\": \"outliers_lat\"}

```

```

ne_df = n_df[(n_df['Latitude'] < upper_bound) & (n_df['Latitude'] >
lower_bound)]

```

```

ne_df

```

```

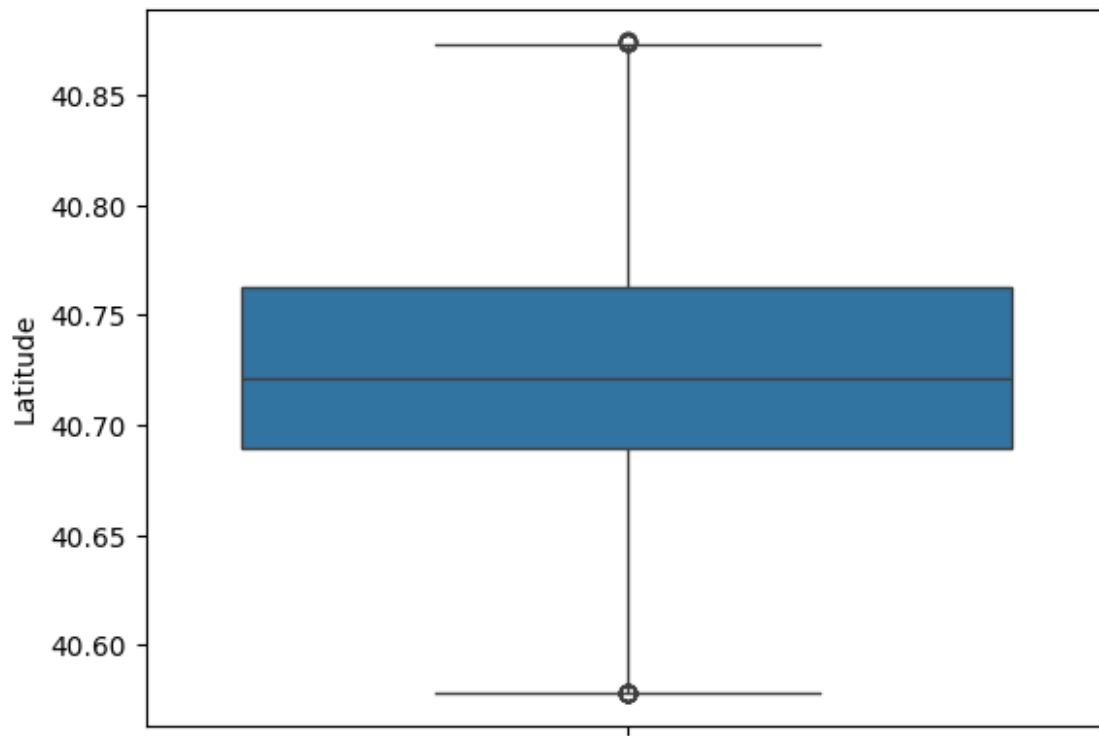
{ \"summary\": { \n      \"name\": \"ne_df\", \n      \"rows\": 45541, \n
\"fields\": [ \n          { \n              \"column\": \"Id\", \n              \"properties\":
{ \n                  \"dtype\": \"number\", \n                  \"std\": 10921092, \n
\"min\": 2539, \n                  \"max\": 36487245, \n

```

```
\ "num_unique_values\ ": 45541,\n          \ "samples\ ": [\n
4530160,\n          14698446,\n          16814818\n          ],\n
\ "semantic_type\ ": \ "\",\n          \ "description\ ": \ "\",\n
n      },\n      {\n          \ "column\ ": \ "Name\ ",\n          \ "properties\ ": {\n
\ "dtype\ ": \ "string\ ",\n          \ "num_unique_values\ ": 44633,\n
\ "samples\ ": [\n          \ "Modern and Spacious\ ",\n          \ "Cozy 2
Bedroom in Murray Hill!\ ",\n          \ "A great room in
Manhattan !!!\ "\n          ],\n          \ "semantic_type\ ": \ "\",\n
\ "description\ ": \ "\",\n          },\n          {\n          \ "column\ ":
\ "Host Id\ ",\n          \ "properties\ ": {\n          \ "dtype\ ": \ "number\ ",\n
n          \ "std\ ": 77484547,\n          \ "min\ ": 2438,\n          \ "max\ ":
274321313,\n          \ "num_unique_values\ ": 35100,\n
\ "samples\ ": [\n          229739739,\n          70091,\n
39528519\n          ],\n          \ "semantic_type\ ": \ "\",\n
\ "description\ ": \ "\",\n          },\n          {\n          \ "column\ ":
\ "Host Name\ ",\n          \ "properties\ ": {\n          \ "dtype\ ":
\ "category\ ",\n          \ "num_unique_values\ ": 10935,\n
\ "samples\ ": [\n          \ "Paul & Ewa\ ",\n          \ "Florence And
Peter\ ",\n          \ "Melinda\ "\n          ],\n          \ "semantic_type\ ": \ "\",\n
\ "description\ ": \ "\",\n          },\n          {\n          \ "column\ ": \ "Neighbourhood Group\ ",\n
\ "properties\ ": {\n          \ "dtype\ ": \ "category\ ",\n
\ "num_unique_values\ ": 5,\n          \ "samples\ ": [\n
\ "Manhattan\ ",\n          \ "Bronx\ ",\n          \ "Queens\ "\n
n          ],\n          \ "semantic_type\ ": \ "\",\n
\ "description\ ": \ "\",\n          },\n          {\n          \ "column\ ":
\ "Neighbourhood\ ",\n          \ "properties\ ": {\n          \ "dtype\ ":
\ "category\ ",\n          \ "num_unique_values\ ": 193,\n
\ "samples\ ": [\n          \ "Ridgewood\ ",\n          \ "Norwood\ ",\n
\ "Concourse\ "\n          ],\n          \ "semantic_type\ ": \ "\",\n
\ "description\ ": \ "\",\n          },\n          {\n          \ "column\ ":
\ "Latitude\ ",\n          \ "properties\ ": {\n          \ "dtype\ ":
\ "number\ ",\n          \ "std\ ": 0.05373802900495381,\n          \ "min\ ":
40.57801,\n          \ "max\ ": 40.87448,\n          \ "num_unique_values\ ":
18417,\n          \ "samples\ ": [\n          40.64993,\n
40.68806,\n          40.75953\n          ],\n          \ "semantic_type\ ":
\ "\",\n          \ "description\ ": \ "\",\n          },\n          {\n
\ "column\ ": \ "Longitude\ ",\n          \ "properties\ ": {\n
\ "dtype\ ": \ "number\ ",\n          \ "std\ ": 0.04571626844150896,\n
\ "min\ ": -74.17459,\n          \ "max\ ": -73.71299,\n
\ "num_unique_values\ ": 14331,\n          \ "samples\ ": [\n          -
73.8474,\n          -73.9199,\n          -73.76828\n          ],\n          \ "semantic_type\ ": \ "\",\n
\ "description\ ": \ "\",\n          },\n          {\n          \ "column\ ": \ "Room Type\ ",\n
\ "properties\ ": {\n          \ "dtype\ ": \ "category\ ",\n
\ "num_unique_values\ ": 3,\n          \ "samples\ ": [\n          \ "Private
room\ ",\n          \ "Entire home/apt\ ",\n          \ "Shared room\ "\n
          ],\n          \ "semantic_type\ ": \ "\",\n          \ "description\ ": \ "\",\n
          },\n          {\n          \ "column\ ": \ "Price\ ",\n          \ "properties\ ":
```

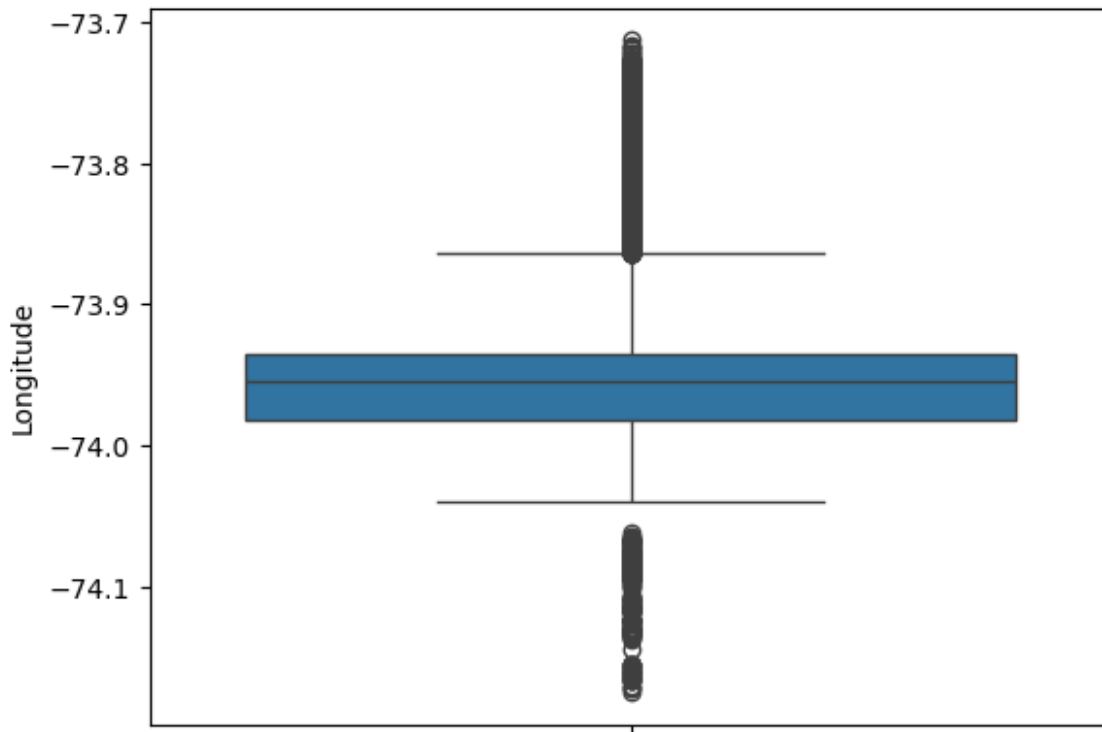


```
<Axes: ylabel='Latitude'>
```



```
sns.boxplot(ne_df['Longitude'])
```

```
<Axes: ylabel='Longitude'>
```

```

q1 = ne_df['Longitude'].quantile(0.25)
q3 = ne_df['Longitude'].quantile(0.75)
iqr=q3-q1
upper_bound = q3 + 1.5*iqr
lower_bound = q1 - 1.5*iqr

outlier_long = ne_df[(ne_df['Longitude'] > upper_bound) |
(ne_df['Longitude'] < lower_bound)]

outlier_long

{"summary":{"\n  \"name\": \"outlier_long\",\n  \"rows\": 2517,\n  \"fields\": [\n    {\n      \"column\": \"Id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 9670053,\n        \"min\": 32363,\n        \"max\": 36484363,\n        \"num_unique_values\": 2517,\n        \"samples\": [\n          34229869,\n          13370393,\n          26076877\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Name\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 2469,\n        \"samples\": [\n          \"A very conducive private area near to\ntrain statn\",\n          \"Fort Pl B&B\",\n          \"Chill pad\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Host Id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 84153518,\n        \"min\": 3415,\n        \"max\": 274225617,\n        \"num_unique_values\": 1646,\n        \"samples\": [\n
```

```

150019486,\n      33555400,\n      155977656\n],\n{\n  \"semantic_type\": \"\",\n  \"description\": \"\",\n  \"column\": \"Host Name\", \n},\n{\n  \"dtype\": \"category\", \n  \"num_unique_values\": 1215,\n  \"samples\": [\n    \"Mo\", \n    \"Nicola\", \n    \"H. Ann\" ],\n  \"semantic_type\": \"\", \n  \"description\": \"\", \n  \"column\": \"Neighbourhood Group\", \n},\n{\n  \"dtype\": \"category\", \n  \"num_unique_values\": 4,\n  \"samples\": [\n    \"Staten Island\", \n    \"Brooklyn\", \n    \"Queens\" ],\n  \"semantic_type\": \"\", \n  \"description\": \"\", \n  \"column\": \"Neighbourhood\", \n},\n{\n  \"dtype\": \"category\", \n  \"num_unique_values\": 87,\n  \"samples\": [\n    \"Westerleigh\", \n    \"Flushing\", \n    \"City Island\" ],\n  \"semantic_type\": \"\", \n  \"description\": \"\", \n  \"column\": \"Latitude\", \n},\n{\n  \"dtype\": \"number\", \n  \"std\": 0.06698578850410555,\n  \"min\": 40.57801,\n  \"max\": 40.87372,\n  \"num_unique_values\": 2337,\n  \"samples\": [\n    40.86975,\n    40.86674,\n    40.63513 ] ,\n  \"semantic_type\": \"\", \n  \"description\": \"\", \n  \"column\": \"Longitude\", \n},\n{\n  \"dtype\": \"number\", \n  \"std\": 0.101552152655777,\n  \"min\": -74.17459,\n  \"max\": -73.71299,\n  \"num_unique_values\": 2324,\n  \"samples\": [\n    -73.74517,\n    -73.85836,\n    -73.78187 ] ,\n  \"semantic_type\": \"\", \n  \"description\": \"\", \n  \"column\": \"Room Type\", \n},\n{\n  \"dtype\": \"category\", \n  \"num_unique_values\": 3,\n  \"samples\": [\n    \"Private room\", \n    \"Entire home/apt\", \n    \"Shared room\" ],\n  \"semantic_type\": \"\", \n  \"description\": \"\", \n  \"column\": \"Price\", \n},\n{\n  \"dtype\": \"number\", \n  \"std\": 55,\n  \"min\": 10,\n  \"max\": 325,\n  \"num_unique_values\": 184,\n  \"samples\": [\n    69,\n    120,\n    74 ] ,\n  \"semantic_type\": \"\", \n  \"description\": \"\", \n  \"column\": \"Minimum Nights\", \n},\n{\n  \"dtype\": \"number\", \n  \"std\": 9,\n  \"min\": 1,\n  \"max\": 365,\n  \"num_unique_values\": 25,\n  \"samples\": [\n    30,\n    90,\n    2 ] },\n  \"semantic_type\": \"\", \n  \"description\": \"\", \n  \"column\": \"Number Of Reviews\", \n},\n{\n  \"dtype\": \"number\", \n  \"std\": 54,\n  \"min\": 0,\n  \"max\": 629,\n  \"num unique values\": 220,\n  \"samples\": [\n    131,\n
```

```

98,\n          66\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          },\n          {\n            \"column\":\n            \"Last Review\",\n            \"properties\": {\n              \"dtype\":\n              \"category\",\n              \"num_unique_values\": 407,\n              \"samples\": [\n                \"2014-05-24 00:00:00\",\n                \"2017-06-21 00:00:00\",\n                \"2018-04-06 00:00:00\",\n                ],\n              \"semantic_type\": \"\",\n              \"description\": \"\"\n            },\n            {\n              \"column\": \"Reviews Per Month\",\n              \"properties\": {\n                \"dtype\": \"number\",\n                \"std\": 2.217425115739985,\n                \"min\": 0.01,\n                \"max\": 20.94,\n                \"num_unique_values\": 624,\n                \"samples\": [\n                  5.56,\n                  6.11,\n                  5.13,\n                  ],\n                \"semantic_type\": \"\",\n                \"description\": \"\"\n              },\n              {\n                \"column\":\n                \"Calculated Host Listings Count\",\n                \"properties\": {\n                  \"dtype\": \"number\",\n                  \"std\": 3,\n                  \"min\": 1,\n                  \"max\": 43,\n                  \"num_unique_values\": 15,\n                  \"samples\": [\n                    8,\n                    15,\n                    1,\n                    ],\n                  \"semantic_type\": \"\",\n                  \"description\": \"\"\n                },\n                {\n                  \"column\": \"Availability 365\",\n                  \"properties\": {\n                    \"dtype\": \"number\",\n                    \"std\": 132,\n                    \"min\": 0,\n                    \"max\": 365,\n                    \"num_unique_values\": 344,\n                    \"samples\": [\n                      337,\n                      62,\n                      191,\n                      ],\n                    \"semantic_type\": \"\",\n                    \"description\": \"\"\n                  },\n                  {\n                    \"column\":\n                    \"price_mm\",\n                    \"properties\": {\n                      \"dtype\":\n                      \"number\",\n                      \"std\": 0.0055424678255653675,\n                      \"min\": 0.001,\n                      \"max\": 0.0325,\n                      \"num_unique_values\": 184,\n                      \"samples\": [\n                        0.0069000000000000001,\n                        0.012,\n                        0.0074,\n                        ],\n                      \"semantic_type\":\n                      \"\",\n                      \"description\": \"\"\n                    },\n                    {\n                      \"column\": \"outlier_long\"\n                    }\n                  ],\n                \"type\": \"dataframe\", \"variable_name\": \"outlier_long\"

```

```

new_df= ne_df[(ne_df['Longitude'] < upper_bound) & (ne_df['Longitude']
> lower_bound)]

```

```

ne_df

```

```

{"summary": "{\n  \"name\": \"ne_df\",\n  \"rows\": 45541,\n  \"fields\": [\n    {\n      \"column\": \"Id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 10921092,\n        \"min\": 2539,\n        \"max\": 36487245,\n        \"num_unique_values\": 45541,\n        \"samples\": [\n          4530160,\n          14698446,\n          16814818,\n          ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      {\n        \"column\": \"Name\",\n        \"properties\": {\n          \"dtype\": \"string\",\n          \"num_unique_values\": 44633,\n          \"samples\": [\n            \"Modern and Spacious\",\n            \"Cozy 2 Bedroom in Murray Hill!\",\n            \"A great room in Manhattan !!!\",\n            ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\":

```

```

\"Host Id\",\\n      \"properties\": {\\n      \"dtype\": \"number\",\\n
n      \"std\": 77484547,\\n      \"min\": 2438,\\n      \"max\":
274321313,\\n      \"num_unique_values\": 35100,\\n
\"samples\": [\\n      229739739,\\n      70091,\\n
39528519\\n      ],\\n      \"semantic_type\": \"\",\\n
\"description\": \"\"\\n      }\\n      },\\n      {\\n      \"column\":
\"Host Name\",\\n      \"properties\": {\\n      \"dtype\":
\"category\",\\n      \"num_unique_values\": 10935,\\n
\"samples\": [\\n      \"Paul & Ewa\",\\n      \"Florence And
Peter\",\\n      \"Melinda\"\\n      ],\\n
\"semantic_type\": \"\",\\n      \"description\": \"\"\\n      }\\n
n      },\\n      {\\n      \"column\": \"Neighbourhood Group\",\\n
\"properties\": {\\n      \"dtype\": \"category\",\\n
\"num_unique_values\": 5,\\n      \"samples\": [\\n
\"Manhattan\",\\n      \"Bronx\",\\n      \"Queens\"\\n
n      ],\\n      \"semantic_type\": \"\",\\n
\"description\": \"\"\\n      }\\n      },\\n      {\\n      \"column\":
\"Neighbourhood\",\\n      \"properties\": {\\n      \"dtype\":
\"category\",\\n      \"num_unique_values\": 193,\\n
\"samples\": [\\n      \"Ridgewood\",\\n      \"Norwood\",\\n
\"Concourse\"\\n      ],\\n      \"semantic_type\": \"\",\\n
\"description\": \"\"\\n      }\\n      },\\n      {\\n      \"column\":
\"Latitude\",\\n      \"properties\": {\\n      \"dtype\":
\"number\",\\n      \"std\": 0.05373802900495381,\\n      \"min\":
40.57801,\\n      \"max\": 40.87448,\\n      \"num_unique_values\":
18417,\\n      \"samples\": [\\n      40.64993,\\n
40.68806,\\n      40.75953\\n      ],\\n      \"semantic_type\":
\"\",\\n      \"description\": \"\"\\n      }\\n      },\\n      {\\n
\"column\": \"Longitude\",\\n      \"properties\": {\\n
\"dtype\": \"number\",\\n      \"std\": 0.04571626844150896,\\n
\"min\": -74.17459,\\n      \"max\": -73.71299,\\n
\"num_unique_values\": 14331,\\n      \"samples\": [\\n      -
73.8474,\\n      -73.9199,\\n      -73.76828\\n      ],\\n
\"semantic_type\": \"\",\\n      \"description\": \"\"\\n      }\\n
n      },\\n      {\\n      \"column\": \"Room Type\",\\n
\"properties\": {\\n      \"dtype\": \"category\",\\n
\"num_unique_values\": 3,\\n      \"samples\": [\\n      \"Private
room\",\\n      \"Entire home/apt\",\\n      \"Shared room\"\\n
],\\n      \"semantic_type\": \"\",\\n      \"description\": \"\"\\n
}\\n      },\\n      {\\n      \"column\": \"Price\",\\n      \"properties\":
{\\n      \"dtype\": \"number\",\\n      \"std\": 68,\\n
\"min\": 0,\\n      \"max\": 334,\\n      \"num_unique_values\":
322,\\n      \"samples\": [\\n      197,\\n      119,\\n
173\\n      ],\\n      \"semantic_type\": \"\",\\n
\"description\": \"\"\\n      }\\n      },\\n      {\\n      \"column\":
\"Minimum Nights\",\\n      \"properties\": {\\n      \"dtype\":
\"number\",\\n      \"std\": 19,\\n      \"min\": 1,\\n
\"max\": 1250,\\n      \"num_unique_values\": 104,\\n
\"samples\": [\\n      20,\\n      23,\\n      62\\n

```

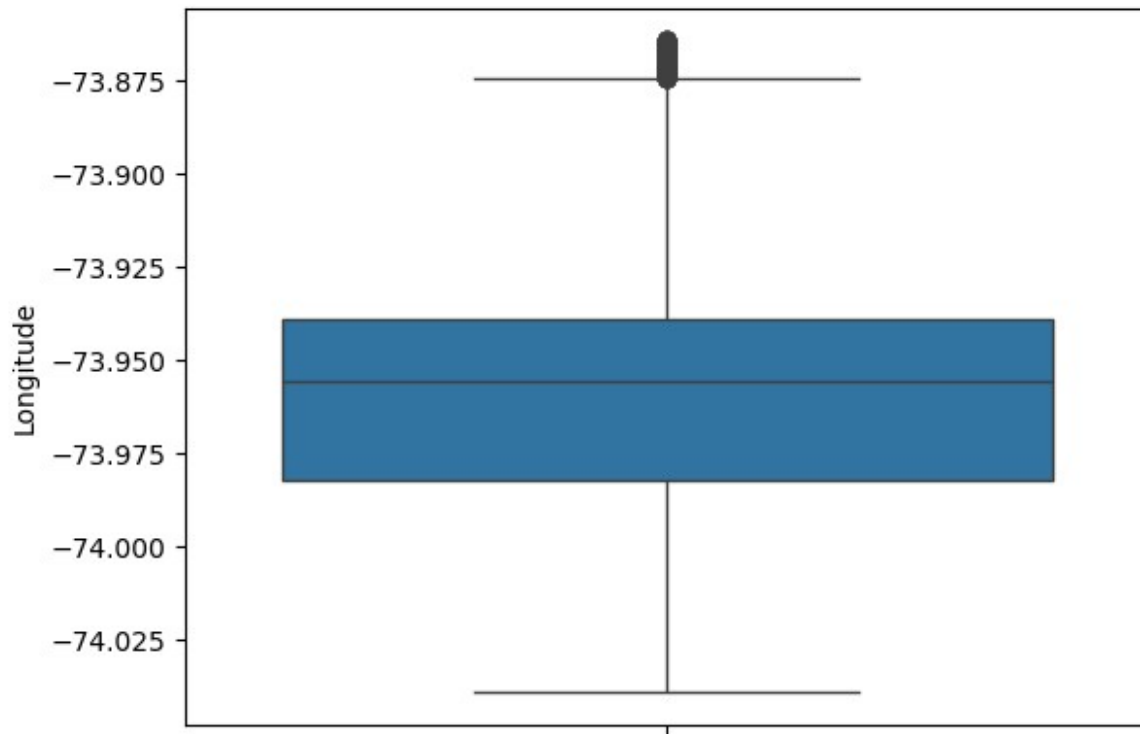
```

n      ],\n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      },\n      {\n          \"column\":
\"Number Of Reviews\", \n          \"properties\": {\n              \"dtype\":
\"number\", \n              \"std\": 45, \n              \"min\": 0, \n
\"max\": 629, \n              \"num_unique_values\": 393, \n
\"samples\": [\n                  38, \n                  439, \n                  153 \n
n      ],\n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      },\n      {\n          \"column\":
\"Last Review\", \n          \"properties\": {\n              \"dtype\":
\"category\", \n              \"num_unique_values\": 1753, \n
\"samples\": [\n                  \"2015-05-31 00:00:00\", \n                  \"2018-
10-22 00:00:00\", \n                  \"2017-05-30 00:00:00\" \n
n      ],\n          \"semantic_type\": \"\", \n          \"description\": \"\" \n
n      },\n      {\n          \"column\": \"Reviews Per Month\", \n
\"properties\": {\n              \"dtype\": \"number\", \n              \"std\":
1.5171042015533407, \n              \"min\": 0.01, \n              \"max\": 58.5, \n
\"num_unique_values\": 935, \n              \"samples\": [\n                  1.7, \n
0.28, \n                  2.33 \n
n      ],\n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      },\n      {\n          \"column\":
\"Calculated Host Listings Count\", \n          \"properties\": {\n
\"dtype\": \"number\", \n              \"std\": 31, \n              \"min\": 1, \n
\"max\": 327, \n              \"num_unique_values\": 47, \n
\"samples\": [\n                  37, \n                  17, \n                  121 \n
n      ],\n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      },\n      {\n          \"column\":
\"Availability 365\", \n          \"properties\": {\n              \"dtype\":
\"number\", \n              \"std\": 130, \n              \"min\": 0, \n
\"max\": 365, \n              \"num_unique_values\": 366, \n
\"samples\": [\n                  315, \n                  309, \n                  249 \n
n      ],\n          \"semantic_type\": \"\", \n          \"description\": \"\" \n
n      },\n      {\n          \"column\": \"price_mm\", \n
\"properties\": {\n              \"dtype\": \"number\", \n              \"std\":
0.006816688961496291, \n              \"min\": 0.0, \n              \"max\":
0.0334, \n              \"num_unique_values\": 322, \n              \"samples\": [\n
0.019700000000000002, \n                  0.0119, \n                  0.0173 \n
n      ],\n          \"semantic_type\": \"\", \n
\"description\": \"\" \n      } \n
n      ],\n      \"type\": \"dataframe\", \"variable_name\": \"ne_df\"}

```

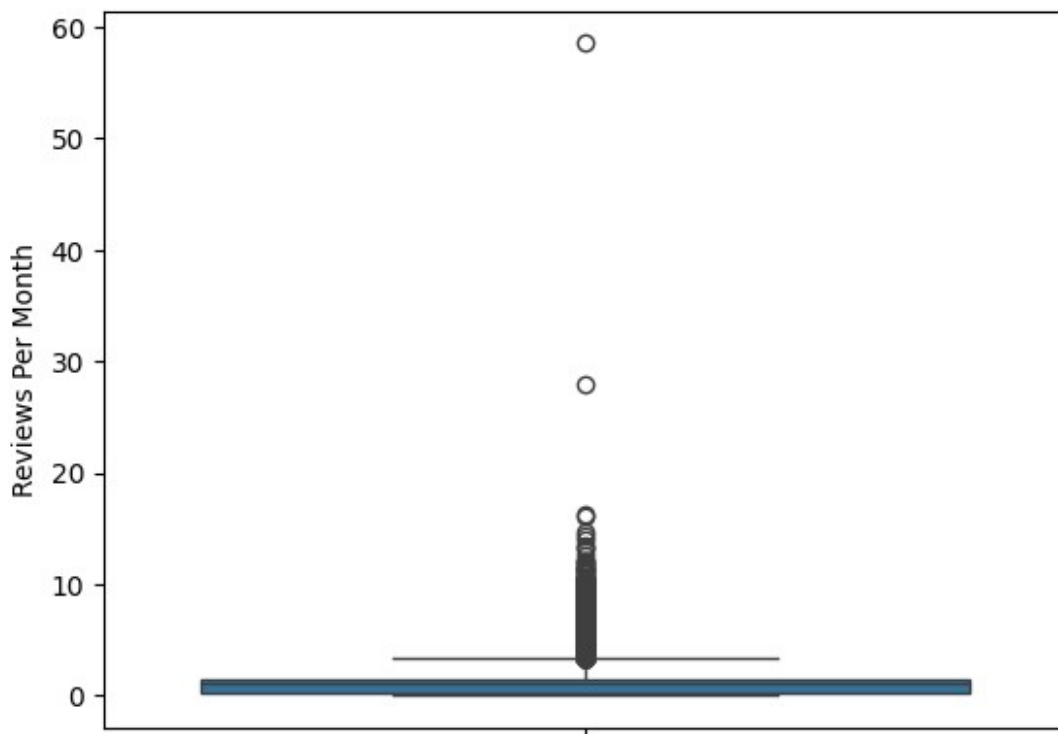
```
sns.boxplot(new_df['Longitude'])
```

```
<Axes: ylabel='Longitude'>
```



```
sns.boxplot(new_df['Reviews Per Month'])
```

```
<Axes: ylabel='Reviews Per Month'>
```



```

q1 = new_df['Reviews Per Month'].quantile(0.25)
q3 = new_df['Reviews Per Month'].quantile(0.75)
iqr= q3 -q1
upper_bound = q3 +1.5*iqr
lower_bound = q1 - 1.5*iqr

q1,q3,iqr
(0.25, 1.52, 1.27)

upper_bound,lower_bound
(3.425, -1.655)

outliers_review = new_df[(new_df['Reviews Per Month'] > upper_bound) |
(new_df['Reviews Per Month'] < lower_bound)]

outliers_review
{"repr_error":"0","type":"dataframe","variable_name":"outliers_review"}

new1_df = new_df[(new_df['Reviews Per Month'] < upper_bound) &
(new_df['Reviews Per Month'] > lower_bound)]

new1_df
{"summary":{"\n  \"name\": \"new1_df\", \n  \"rows\": 39325, \n
  \"fields\": [\n    {\n      \"column\": \"Id\", \n      \"properties\": {\n
        \"dtype\": \"number\", \n      \"std\": 10944178, \n
        \"min\": 2539, \n      \"max\": 36487245, \n
        \"num_unique_values\": 39325, \n      \"samples\": [\n
        35005936, \n      21560584, \n      33689069\n      ], \n
        \"semantic_type\": \"\", \n      \"description\": \"\" \n
      }, \n      {\n      \"column\": \"Name\", \n      \"properties\": {\n
        \"dtype\": \"string\", \n      \"num_unique_values\": 38565, \n
        \"samples\": [\n      \"Furnished Apartment in Morningside\", \n
        \"Sunny Private Room w/Exposed Brick\", \n      \"Cozy bedroom in
downtown Manhattan\" \n      ], \n      \"semantic_type\": \"\", \n
        \"description\": \"\" \n      }, \n      {\n      \"column\":
        \"Host Id\", \n      \"properties\": {\n      \"dtype\": \"number\", \n
        \"std\": 73870494, \n      \"min\": 2438, \n      \"max\":
        274321313, \n      \"num_unique_values\": 31121, \n
        \"samples\": [\n      16789954, \n      92634775, \n
        129179250\n      ], \n      \"semantic_type\": \"\", \n
        \"description\": \"\" \n      }, \n      {\n      \"column\":
        \"Host Name\", \n      \"properties\": {\n      \"dtype\":
        \"category\", \n      \"num_unique_values\": 9735, \n
        \"samples\": [\n      \"Mitra\", \n      \"Narmeen\", \n
        \"Debbie & Hilel\" \n      ], \n      \"semantic_type\": \"\", \n
        \"description\": \"\" \n      }, \n      {\n      \"column\":

```

```

\"Neighbourhood Group\", \n      \"properties\": { \n      \"dtype\":
\"category\", \n      \"num_unique_values\": 4, \n      \"samples\":
[ \n      \"Manhattan\", \n      \"Bronx\", \n
\"Brooklyn\", \n      ], \n      \"semantic_type\": \"\", \n
\"description\": \"\", \n      }, \n      { \n      \"column\":
\"Neighbourhood\", \n      \"properties\": { \n      \"dtype\":
\"category\", \n      \"num_unique_values\": 120, \n
\"samples\": [ \n      \"Flatiron District\", \n      \"Little
Italy\", \n      \"Murray Hill\", \n      ], \n
\"semantic_type\": \"\", \n      \"description\": \"\", \n
      }, \n      { \n      \"column\": \"Latitude\", \n      \"properties\":
{ \n      \"dtype\": \"number\", \n      \"std\":
0.05246522020967552, \n      \"min\": 40.57804, \n      \"max\":
40.87446, \n      \"num_unique_values\": 17124, \n      \"samples\":
[ \n      40.67821, \n      40.8188, \n      40.61735 \n
      ], \n      \"semantic_type\": \"\", \n      \"description\": \"\", \n
      }, \n      { \n      \"column\": \"Longitude\", \n      \"properties\": { \n      \"dtype\": \"number\", \n      \"std\":
0.029954369575799, \n      \"min\": -74.03942, \n      \"max\": -
73.86418, \n      \"num_unique_values\": 11651, \n      \"samples\":
[ \n      -74.00681, \n      -73.95366, \n      -73.98808 \n
      ], \n      \"semantic_type\": \"\", \n      \"description\": \"\", \n
      }, \n      { \n      \"column\": \"Room Type\", \n      \"properties\": { \n      \"dtype\": \"category\", \n
\"num_unique_values\": 3, \n      \"samples\": [ \n      \"Private
room\", \n      \"Entire home/apt\", \n      \"Shared room\", \n
      ], \n      \"semantic_type\": \"\", \n      \"description\": \"\", \n
      }, \n      { \n      \"column\": \"Price\", \n      \"properties\":
{ \n      \"dtype\": \"number\", \n      \"std\": 68, \n
\"min\": 0, \n      \"max\": 334, \n      \"num_unique_values\":
320, \n      \"samples\": [ \n      217, \n      18, \n
      68 \n      ], \n      \"semantic_type\": \"\", \n
\"description\": \"\", \n      }, \n      { \n      \"column\":
\"Minimum Nights\", \n      \"properties\": { \n      \"dtype\":
\"number\", \n      \"std\": 21, \n      \"min\": 1, \n
\"max\": 1250, \n      \"num_unique_values\": 104, \n
\"samples\": [ \n      40, \n      23, \n      62 \n
      ], \n      \"semantic_type\": \"\", \n
\"description\": \"\", \n      }, \n      { \n      \"column\":
\"Number Of Reviews\", \n      \"properties\": { \n      \"dtype\":
\"number\", \n      \"std\": 33, \n      \"min\": 0, \n
\"max\": 354, \n      \"num_unique_values\": 277, \n
\"samples\": [ \n      15, \n      129, \n      176 \n
      ], \n      \"semantic_type\": \"\", \n
\"description\": \"\", \n      }, \n      { \n      \"column\":
\"Last Review\", \n      \"properties\": { \n      \"dtype\":
\"category\", \n      \"num_unique_values\": 1748, \n
\"samples\": [ \n      \"2019-02-27 00:00:00\", \n      \"2019-
04-16 00:00:00\", \n      \"2016-11-11 00:00:00\", \n
      ], \n

```



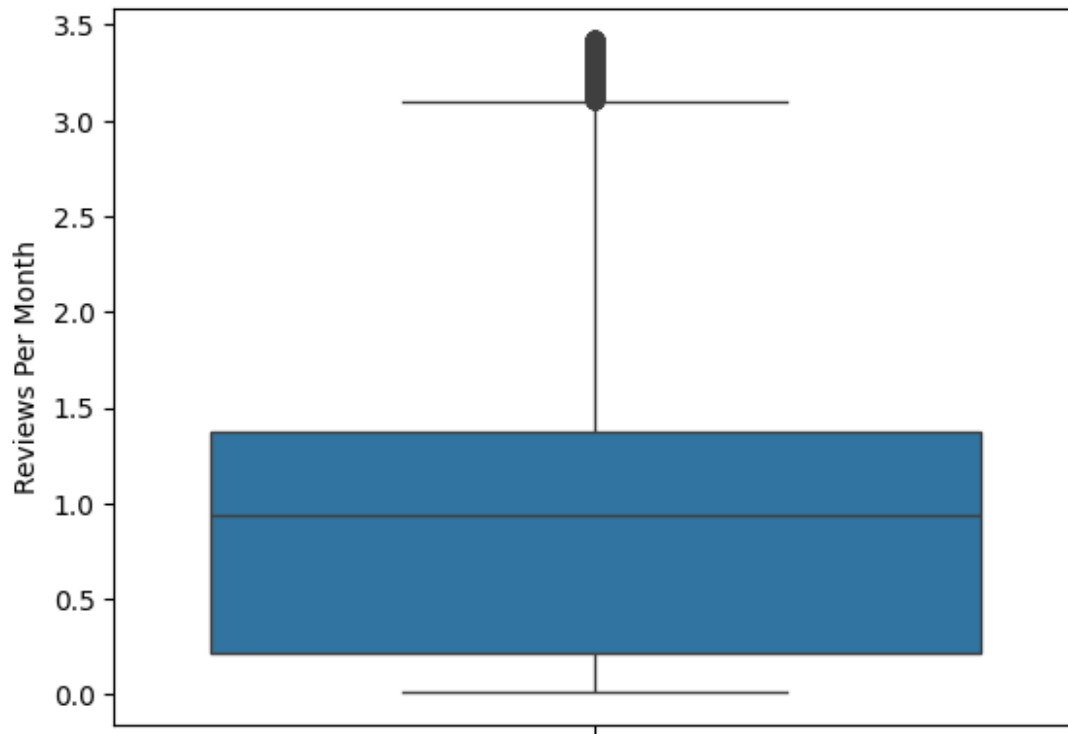
```

{"semantic_type": "\",\n      \"description\": \"\"\n    },\n    {\n      \"column\": \"Reviews Per Month\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 0.8359903183556833,\n        \"min\": 0.01,\n        \"max\": 3.42,\n        \"num_unique_values\": 343,\n        \"samples\": [\n          2.29,\n          0.36,\n          2.0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      {\n        \"column\": \"Calculated Host Listings Count\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 33,\n          \"min\": 1,\n          \"max\": 327,\n          \"num_unique_values\": 47,\n          \"samples\": [\n            49,\n            17,\n            37\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\": \"Availability 365\",\n          \"properties\": {\n            \"dtype\": \"number\",\n            \"std\": 129,\n            \"min\": 0,\n            \"max\": 365,\n            \"num_unique_values\": 366,\n            \"samples\": [\n              221,\n              265,\n              304\n            ],\n            \"semantic_type\": \"\",\n            \"description\": \"\"\n          },\n          {\n            \"column\": \"price_mm\",\n            \"properties\": {\n              \"dtype\": \"number\",\n              \"std\": 0.0068592574772327755,\n              \"min\": 0.0,\n              \"max\": 0.0334,\n              \"num_unique_values\": 320,\n              \"samples\": [\n                0.0217,\n                0.0018000000000000002,\n                0.0068000000000000005\n              ],\n              \"semantic_type\": \"\",\n              \"description\": \"\"\n            }\n          }\n        }\n      ],\n      \"type\": \"dataframe\", \"variable_name\": \"new1_df\"}

```

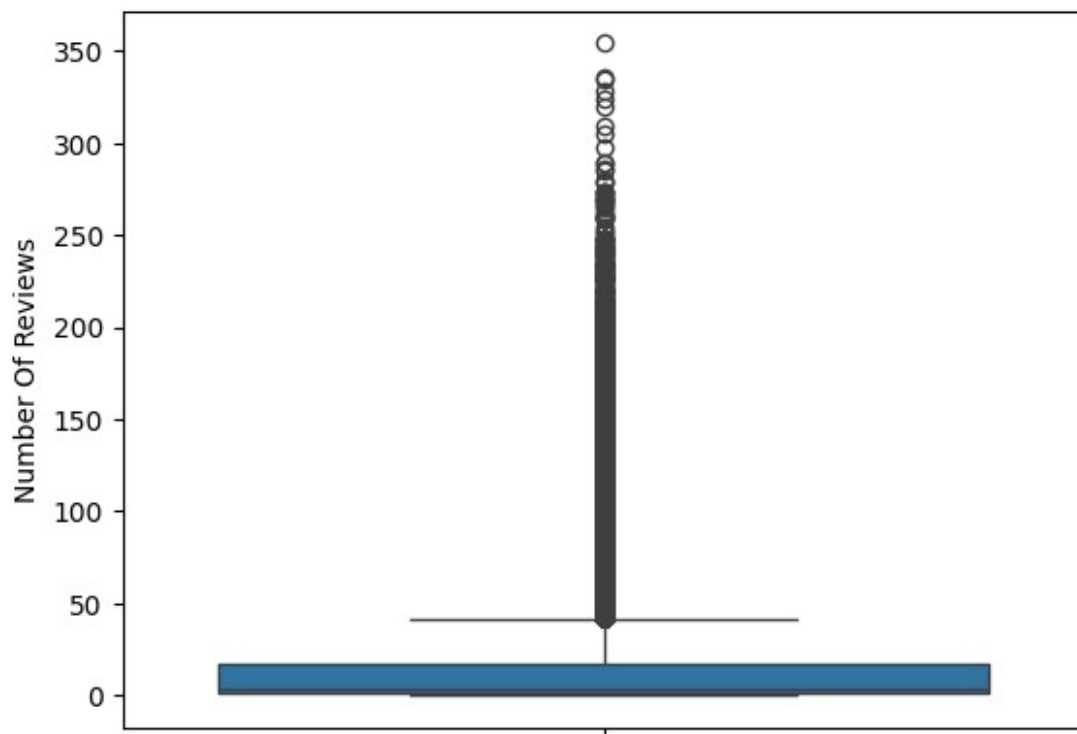
```
sns.boxplot(new1_df['Reviews Per Month'])
```

```
<Axes: ylabel='Reviews Per Month'>
```



```
sns.boxplot(new1_df['Number Of Reviews'])
```

```
<Axes: ylabel='Number Of Reviews'>
```



```

q1 = new1_df['Number Of Reviews'].quantile(0.25)
q3 = new1_df['Number Of Reviews'].quantile(0.75)
iqr= q3 -q1
upper_bound = q3 +1.5*iqr
lower_bound = q1 - 1.5*iqr

q1,q3,iqr

(1.0, 17.0, 16.0)

upper_bound,lower_bound

(41.0, -23.0)

outlier_num = new1_df[(new1_df['Number Of Reviews'] > upper_bound) |
(new1_df['Number Of Reviews'] < lower_bound)]

outlier_num

{"repr_error":"0","type":"dataframe","variable_name":"outlier_num"}

data = new1_df[(new1_df['Number Of Reviews'] < upper_bound) &
(new1_df['Number Of Reviews'] > lower_bound)]

data

{"summary":{"\n  \"name\": \"data\",\n  \"rows\": 34185,\n  \"fields\": [\n    {\n      \"column\": \"Id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 10735894,\n        \"min\": 2539,\n        \"max\": 36487245,\n        \"num_unique_values\": 34185,\n        \"samples\": [\n          6327529,\n          16346945,\n          13392941\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Name\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 33514,\n        \"samples\": [\n          \"Gorgeous Williamsburg Apt + balcony\",\n          \"New Luxury Brooklyn Apartment Near Manhattan\",\n          \"Gorgeous West Village Hidden Gem\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Host Id\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 76701959,\n        \"min\": 2438,\n        \"max\": 274321313,\n        \"num_unique_values\": 27474,\n        \"samples\": [\n          119464278,\n          7556128,\n          3710931\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Host Name\",\n      \"properties\": {\n        \"dtype\": \"category\",\n        \"num_unique_values\": 8844,\n        \"samples\": [\n          \"Mitra\",\n          \"Allon\",\n          \"Naveen\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    },\n    {\n      \"column\": \"Neighbourhood Group\",\n
```

```

\ "properties\ ": { \ n      \ "dtype\ ": \ "category\ ", \ n
\ "num_unique_values\ ": 4, \ n      \ "samples\ ": [ \ n
\ "Manhattan\ ", \ n      \ "Bronx\ ", \ n      \ "Brooklyn\ " \
n      ], \ n      \ "semantic_type\ ": \ "\", \ n
\ "description\ ": \ "\ \ n      } \ n      }, \ n      { \ n      \ "column\ ":
\ "Neighbourhood\ ", \ n      \ "properties\ ": { \ n      \ "dtype\ ":
\ "category\ ", \ n      \ "num_unique_values\ ": 120, \ n
\ "samples\ ": [ \ n      \ "Gravesend\ ", \ n      \ "Fort
Hamilton\ ", \ n      \ "Williamsburg\ " \ n      ], \ n
\ "semantic_type\ ": \ "\", \ n      \ "description\ ": \ "\ \ n      } \
n      }, \ n      { \ n      \ "column\ ": \ "Latitude\ ", \ n      \ "properties\ ":
{ \ n      \ "dtype\ ": \ "number\ ", \ n      \ "std\ ":
0.05215537554769396, \ n      \ "min\ ": 40.57804, \ n      \ "max\ ":
40.87446, \ n      \ "num_unique_values\ ": 16108, \ n      \ "samples\ ":
[ \ n      40.79967, \ n      40.76769, \ n      40.67002 \ n
], \ n      \ "semantic_type\ ": \ "\", \ n      \ "description\ ": \ "\ \ n
} \ n      }, \ n      { \ n      \ "column\ ": \ "Longitude\ ", \ n
\ "properties\ ": { \ n      \ "dtype\ ": \ "number\ ", \ n      \ "std\ ":
0.029936279725976858, \ n      \ "min\ ": -74.03942, \ n      \ "max\ ": -
73.86418, \ n      \ "num_unique_values\ ": 11171, \ n      \ "samples\ ":
[ \ n      -73.98832, \ n      -73.94748, \ n      -73.91423 \ n
], \ n      \ "semantic_type\ ": \ "\", \ n      \ "description\ ": \ "\ \ n
} \ n      }, \ n      { \ n      \ "column\ ": \ "Room Type\ ", \ n
\ "properties\ ": { \ n      \ "dtype\ ": \ "category\ ", \ n
\ "num_unique_values\ ": 3, \ n      \ "samples\ ": [ \ n      \ "Private
room\ ", \ n      \ "Entire home/apt\ ", \ n      \ "Shared room\ " \ n
], \ n      \ "semantic_type\ ": \ "\", \ n      \ "description\ ": \ "\ \ n
} \ n      }, \ n      { \ n      \ "column\ ": \ "Price\ ", \ n      \ "properties\ ":
{ \ n      \ "dtype\ ": \ "number\ ", \ n      \ "std\ ": 69, \ n
\ "min\ ": 0, \ n      \ "max\ ": 334, \ n      \ "num_unique_values\ ":
319, \ n      \ "samples\ ": [ \ n      157, \ n      22, \ n
235 \ n
], \ n      \ "semantic_type\ ": \ "\", \ n
\ "description\ ": \ "\ \ n      } \ n      }, \ n      { \ n      \ "column\ ":
\ "Minimum Nights\ ", \ n      \ "properties\ ": { \ n      \ "dtype\ ":
\ "number\ ", \ n      \ "std\ ": 22, \ n      \ "min\ ": 1, \ n
\ "max\ ": 1250, \ n      \ "num_unique_values\ ": 103, \ n
\ "samples\ ": [ \ n      122, \ n      200, \ n      364 \ n
], \ n      \ "semantic_type\ ": \ "\", \ n      \ "description\ ": \ "\ \ n
} \ n      }, \ n      { \ n      \ "column\ ": \ "Number Of Reviews\ ", \ n
\ "properties\ ": { \ n      \ "dtype\ ": \ "number\ ", \ n      \ "std\ ":
9, \ n      \ "min\ ": 0, \ n      \ "max\ ": 40, \ n
\ "num_unique_values\ ": 41, \ n      \ "samples\ ": [ \ n      10, \ n
2, \ n      11 \ n
], \ n      \ "semantic_type\ ": \ "\", \ n
\ "description\ ": \ "\ \ n      } \ n      }, \ n      { \ n      \ "column\ ":
\ "Last Review\ ", \ n      \ "properties\ ": { \ n      \ "dtype\ ":
\ "category\ ", \ n      \ "num_unique_values\ ": 1741, \ n
\ "samples\ ": [ \ n      \ "2016-04-23 00:00:00\ ", \ n      \ "2017-
02-22 00:00:00\ ", \ n      \ "2014-08-15 00:00:00\ " \ n      ], \ n
\ "semantic_type\ ": \ "\", \ n      \ "description\ ": \ "\ \ n      } \

```

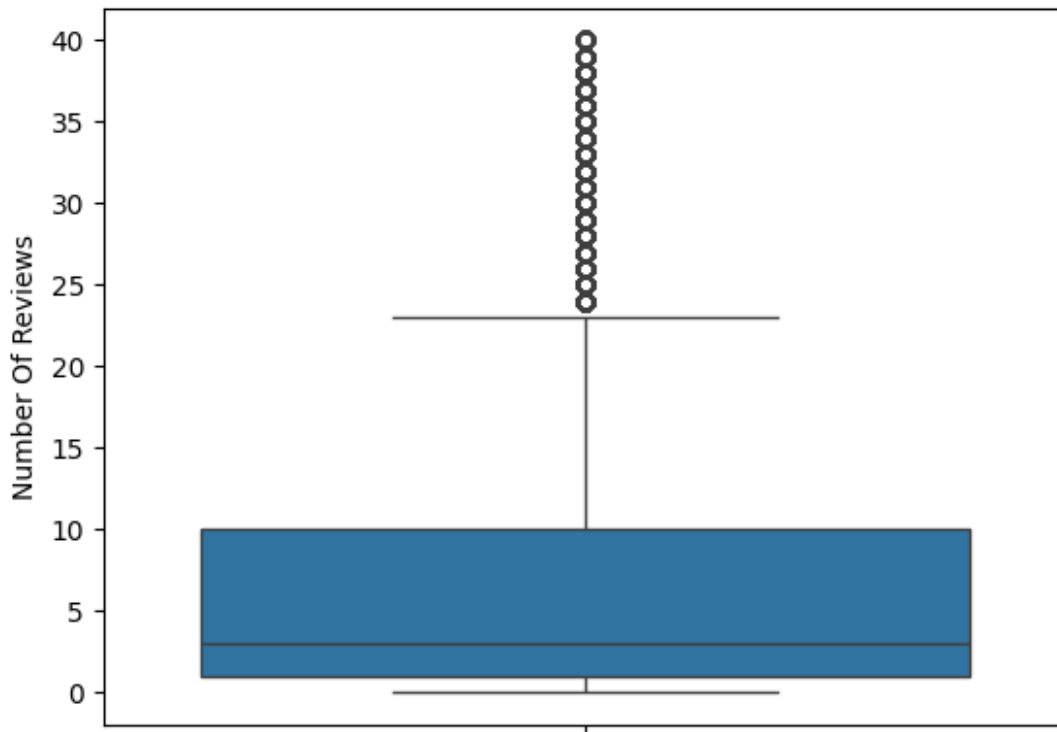
```

n    },\n    {\n        \"column\": \"Reviews Per Month\", \n        \"properties\": {\n            \"dtype\": \"number\", \n            \"std\": 0.7359449523493259, \n            \"min\": 0.01, \n            \"max\": 3.42, \n            \"num_unique_values\": 343, \n            \"samples\": [\n                1.92, \n                2.2, \n                1.48\n            ], \n            \"semantic_type\": \"\", \n            \"description\": \"\"\n        }, \n        {\n            \"column\": \"Calculated Host Listings Count\", \n            \"properties\": {\n                \"dtype\": \"number\", \n                \"std\": 35, \n                \"min\": 1, \n                \"max\": 327, \n                \"num_unique_values\": 47, \n                \"samples\": [\n                    49, \n                    17, \n                    37\n                ], \n                \"semantic_type\": \"\", \n                \"description\": \"\"\n            }, \n            {\n                \"column\": \"Availability 365\", \n                \"properties\": {\n                    \"dtype\": \"number\", \n                    \"std\": 128, \n                    \"min\": 0, \n                    \"max\": 365, \n                    \"num_unique_values\": 366, \n                    \"samples\": [\n                        352, \n                        34, \n                        324\n                    ], \n                    \"semantic_type\": \"\", \n                    \"description\": \"\"\n                }, \n                {\n                    \"column\": \"price_mm\", \n                    \"properties\": {\n                        \"dtype\": \"number\", \n                        \"std\": 0.00691723493673568, \n                        \"min\": 0.0, \n                        \"max\": 0.0334, \n                        \"num_unique_values\": 319, \n                        \"samples\": [\n                            0.015700000000000002, \n                            0.0022, \n                            0.0235\n                        ], \n                        \"semantic_type\": \"\", \n                        \"description\": \"\"\n                    }, \n                    }\n            }\n        ], \n        \"type\": \"dataframe\", \n        \"variable_name\": \"data\"
    }

sns.boxplot(data['Number Of Reviews'])

<Axes: ylabel='Number Of Reviews'>

```



```
data.shape
```

```
(34185, 17)
```

```
data.head()
```

```
{
  "summary": {
    "name": "data",
    "rows": 34185,
    "fields": [
      {
        "column": "Id",
        "properties": {
          "dtype": "number",
          "std": 10735894,
          "min": 2539,
          "max": 36487245,
          "num_unique_values": 34185,
          "samples": [
            6327529,
            16346945,
            13392941
          ],
          "semantic_type": "",
          "description": ""
        }
      },
      {
        "column": "Name",
        "properties": {
          "dtype": "string",
          "num_unique_values": 33514,
          "samples": [
            "Gorgeous Williamsburg Apt + balcony",
            "New Luxury Brooklyn Apartment Near Manhattan",
            "Gorgeous West Village Hidden Gem"
          ],
          "semantic_type": "",
          "description": ""
        }
      },
      {
        "column": "Host Id",
        "properties": {
          "dtype": "number",
          "std": 76701959,
          "min": 2438,
          "max": 274321313,
          "num_unique_values": 27474,
          "samples": [
            119464278,
            7556128,
            3710931
          ],
          "semantic_type": "",
          "description": ""
        }
      },
      {
        "column": "Host Name",
        "properties": {
          "dtype": "category",
          "std": 76701959,
          "min": 2438,
          "max": 274321313,
          "num_unique_values": 27474,
          "samples": [
            119464278,
            7556128,
            3710931
          ],
          "semantic_type": "",
          "description": ""
        }
      }
    ]
  }
}
```

```

\"num_unique_values\": 8844,\n          \"samples\": [\n
\"Mitra\", \n          \"Allon\", \n          \"Naveen\" \n          ], \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n          } \n
    }, \n    { \n          \"column\": \"Neighbourhood Group\", \n
\"properties\": { \n          \"dtype\": \"category\", \n
\"num_unique_values\": 4, \n          \"samples\": [\n
\"Manhattan\", \n          \"Bronx\", \n          \"Brooklyn\" \n
          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n          } \n    }, \n    { \n          \"column\":
\"Neighbourhood\", \n          \"properties\": { \n          \"dtype\":
\"category\", \n          \"num_unique_values\": 120, \n
\"samples\": [\n          \"Gravesend\", \n          \"Fort
Hamilton\", \n          \"Williamsburg\" \n          ], \n
\"semantic_type\": \"\", \n          \"description\": \"\" \n          } \n
    }, \n    { \n          \"column\": \"Latitude\", \n          \"properties\":
{ \n          \"dtype\": \"number\", \n          \"std\":
0.05215537554769396, \n          \"min\": 40.57804, \n          \"max\":
40.87446, \n          \"num_unique_values\": 16108, \n          \"samples\":
[\n          40.79967, \n          40.76769, \n          40.67002 \n
          ], \n          \"semantic_type\": \"\", \n          \"description\":
\"\" \n          } \n    }, \n    { \n          \"column\": \"Longitude\", \n
\"properties\": { \n          \"dtype\": \"number\", \n          \"std\":
0.029936279725976858, \n          \"min\": -74.03942, \n          \"max\": -
73.86418, \n          \"num_unique_values\": 11171, \n          \"samples\":
[\n          -73.98832, \n          -73.94748, \n          -73.91423 \n
          ], \n          \"semantic_type\": \"\", \n          \"description\":
\"\" \n          } \n    }, \n    { \n          \"column\": \"Room Type\", \n
\"properties\": { \n          \"dtype\": \"category\", \n
\"num_unique_values\": 3, \n          \"samples\": [\n          \"Private
room\", \n          \"Entire home/apt\", \n          \"Shared room\" \n
          ], \n          \"semantic_type\": \"\", \n          \"description\":
\"\" \n          } \n    }, \n    { \n          \"column\": \"Price\", \n
\"properties\": { \n          \"dtype\": \"number\", \n          \"std\": 69, \n
\"min\": 0, \n          \"max\": 334, \n          \"num_unique_values\":
319, \n          \"samples\": [\n          157, \n          22, \n
          235 \n          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n          } \n    }, \n    { \n          \"column\":
\"Minimum Nights\", \n          \"properties\": { \n          \"dtype\":
\"number\", \n          \"std\": 22, \n          \"min\": 1, \n
\"max\": 1250, \n          \"num_unique_values\": 103, \n
\"samples\": [\n          122, \n          200, \n          364 \n
          ], \n          \"semantic_type\": \"\", \n          \"description\":
\"\" \n          } \n    }, \n    { \n          \"column\": \"Number Of Reviews\", \n
\"properties\": { \n          \"dtype\": \"number\", \n          \"std\":
9, \n          \"min\": 0, \n          \"max\": 40, \n
\"num_unique_values\": 41, \n          \"samples\": [\n          10, \n
          2, \n          11 \n          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\" \n          } \n    }, \n    { \n          \"column\":
\"Last Review\", \n          \"properties\": { \n          \"dtype\":

```

```

\"category\", \n          \"num_unique_values\": 1741, \n
\"samples\": [ \n          \"2016-04-23 00:00:00\", \n          \"2017-02-22 00:00:00\", \n          \"2014-08-15 00:00:00\" ], \n
\"semantic_type\": \"\", \n          \"description\": \"\", \n
n    }, \n    { \n          \"column\": \"Reviews Per Month\", \n
\"properties\": { \n          \"dtype\": \"number\", \n          \"std\": 0.7359449523493259, \n          \"min\": 0.01, \n          \"max\": 3.42, \n
\"num_unique_values\": 343, \n          \"samples\": [ \n          1.92, \n          2.2, \n          1.48 \n          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\", \n          } \n    }, \n    { \n          \"column\":
\"Calculated Host Listings Count\", \n          \"properties\": { \n
\"dtype\": \"number\", \n          \"std\": 35, \n          \"min\": 1, \n
\"max\": 327, \n          \"num_unique_values\": 47, \n
\"samples\": [ \n          49, \n          17, \n          37 \n
n          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\", \n          } \n    }, \n    { \n          \"column\":
\"Availability 365\", \n          \"properties\": { \n          \"dtype\":
\"number\", \n          \"std\": 128, \n          \"min\": 0, \n
\"max\": 365, \n          \"num_unique_values\": 366, \n
\"samples\": [ \n          352, \n          34, \n          324 \n
n          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\", \n          } \n    }, \n    { \n          \"column\":
\"price_mm\", \n          \"properties\": { \n          \"dtype\":
\"number\", \n          \"std\": 0.00691723493673568, \n          \"min\":
0.0, \n          \"max\": 0.0334, \n          \"num_unique_values\": 319, \n
\"samples\": [ \n          0.015700000000000002, \n          0.0022, \n
0.0235 \n          ], \n          \"semantic_type\": \"\", \n
\"description\": \"\", \n          } \n    } \n  ] \n
n}, \"type\": \"dataframe\", \"variable_name\": \"data\"}

```