

Exploratory Analysis

Problems:

Load the NYC AirBnB Truncated Dataset. This dataset is a mirror of the full NYC AirBnB dataset found at Kaggle, but only contains the first 10,000 entries.

<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>

For the numerical features,

- 1) Display histograms for the numerical features.
- 2) Construct the scatter plots of price with each of the numerical features.
- 3) Display the correlation histogram.
- 4) Which numerical feature appears to be most predictive of the price?
- 4) Using numerical features to predict the renting price with linear model. Write down the predict function.
- 6) Calculate the RSS cost.

```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
```

```
data =
pd.read_csv("https://raw.githubusercontent.com/tipthederiver/Math-7243-2020/master/Datasets/NYCAirBnB/train.csv")
data
```

```
{"rows":
[{"last_review":"10/19/2018","minimum_nights":"1","availability_365":365,"longitude":-73.97237,"number_of_reviews":"9","_deepnote_index_column":"0","host_name":"John","latitude":"40.64749","room_type":"Private room","calculated_host_listings_count":"6","host_id":"2787","name":"Clean & quiet apt home by the park","reviews_per_month":"0.21","neighbourhood_group":"Brooklyn","neighbourhood":"Kensington","price":"149"},
{"last_review":"5/21/2019","minimum_nights":"1","availability_365":355,"longitude":-73.98377,"number_of_reviews":"45","_deepnote_index_column":"1","host_name":"Jennifer","latitude":"40.75362","room_type":"Entire home/apt","calculated_host_listings_count":"2","host_id":"2845","name":"Skylit Midtown Castle","reviews_per_month":"0.38","neighbourhood_group":"Manhattan","neighbourhood":"Midtown","price":"225"},
```

```
{"last_review":"nan","minimum_nights":"3","availability_365":"365","longitude":"-73.9419","number_of_reviews":"0","_deepnote_index_column":"2","host_name":"Elisabeth","latitude":"40.80902","room_type":"Private room","calculated_host_listings_count":"1","host_id":"4632","name":"THE VILLAGE OF HARLEM...NEW YORK !","reviews_per_month":"nan","neighbourhood_group":"Manhattan","neighbourhood":"Harlem","price":"150"}, {"last_review":"7/5/2019","minimum_nights":"1","availability_365":"194","longitude":"-73.95976","number_of_reviews":"270","_deepnote_index_column":"3","host_name":"LisaRoxanne","latitude":"40.68514","room_type":"Entire home/apt","calculated_host_listings_count":"1","host_id":"4869","name":"Cozy Entire Floor of Brownstone","reviews_per_month":"4.64","neighbourhood_group":"Brooklyn","neighbourhood":"Clinton Hill","price":"89"}, {"last_review":"11/19/2018","minimum_nights":"10","availability_365":"0","longitude":"-73.94399","number_of_reviews":"9","_deepnote_index_column":"4","host_name":"Laura","latitude":"40.79851","room_type":"Entire home/apt","calculated_host_listings_count":"1","host_id":"7192","name":"Entire Apt: Spacious Studio/Loft by central park","reviews_per_month":"0.1","neighbourhood_group":"Manhattan","neighbourhood":"East Harlem","price":"80"}, {"last_review":"6/22/2019","minimum_nights":"3","availability_365":"129","longitude":"-73.975","number_of_reviews":"74","_deepnote_index_column":"5","host_name":"Chris","latitude":"40.74767","room_type":"Entire home/apt","calculated_host_listings_count":"1","host_id":"7322","name":"Large Cozy 1 BR Apartment In Midtown East","reviews_per_month":"0.59","neighbourhood_group":"Manhattan","neighbourhood":"Murray Hill","price":"200"}, {"last_review":"10/5/2017","minimum_nights":"45","availability_365":"0","longitude":"-73.95596","number_of_reviews":"49","_deepnote_index_column":"6","host_name":"Garon","latitude":"40.68688","room_type":"Private room","calculated_host_listings_count":"1","host_id":"7356","name":"BlissArtsSpace!","reviews_per_month":"0.4","neighbourhood_group":"Brooklyn","neighbourhood":"Bedford-Stuyvesant","price":"60"}, {"last_review":"6/24/2019","minimum_nights":"2","availability_365":"220","longitude":"-73.98493","number_of_reviews":"430","_deepnote_index_column":"7","host_name":"Shunichi","latitude":"40.76489","room_type":"Private room","calculated_host_listings_count":"1","host_id":"8967","name":"Large Furnished Room Near B'way","reviews_per_month":"3.47","neighbourhood_group":"Manhattan","neighbourhood":"Hell's Kitchen","price":"79"}, {"last_review":"7/21/2017","minimum_nights":"2","availability_365":"0","longitude":"-73.96723","number_of_reviews":"118","_deepnote_index_column":"8","host
```

```
_name": "MaryEllen", "latitude": "40.80178", "room_type": "Private  
room", "calculated_host_listings_count": "1", "host_id": "7490", "name": "Co  
zy Clean Guest Room - Family  
Apt", "reviews_per_month": "0.99", "neighbourhood_group": "Manhattan", "nei  
ghbourhood": "Upper West Side", "price": "79"},  
{ "last_review": "6/9/2019", "minimum_nights": "1", "availability_365": "188  
", "longitude": "-  
73.99037", "number_of_reviews": "160", "_deepnote_index_column": "9", "host  
_name": "Ben", "latitude": "40.71344", "room_type": "Entire  
home/apt", "calculated_host_listings_count": "4", "host_id": "7549", "name":  
": "Cute & Cozy Lower East Side 1  
bdrm", "reviews_per_month": "1.33", "neighbourhood_group": "Manhattan", "ne  
ighbourhood": "Chinatown", "price": "150"},  
{ "last_review": "6/22/2019", "minimum_nights": "5", "availability_365": "6"  
", "longitude": "-  
73.96545", "number_of_reviews": "53", "_deepnote_index_column": "10", "host  
_name": "Lena", "latitude": "40.80316", "room_type": "Entire  
home/apt", "calculated_host_listings_count": "1", "host_id": "7702", "name":  
": "Beautiful 1br on Upper West  
Side", "reviews_per_month": "0.43", "neighbourhood_group": "Manhattan", "ne  
ighbourhood": "Upper West Side", "price": "135"},  
{ "last_review": "6/23/2019", "minimum_nights": "2", "availability_365": "39  
", "longitude": "-  
73.98867", "number_of_reviews": "188", "_deepnote_index_column": "11", "hos  
t_name": "Kate", "latitude": "40.76076", "room_type": "Private  
room", "calculated_host_listings_count": "1", "host_id": "7989", "name": "Ce  
ntral Manhattan/near  
Broadway", "reviews_per_month": "1.5", "neighbourhood_group": "Manhattan",  
"neighbourhood": "Hell's Kitchen", "price": "85"},  
{ "last_review": "6/24/2019", "minimum_nights": "4", "availability_365": "31  
4", "longitude": "-  
73.98779", "number_of_reviews": "167", "_deepnote_index_column": "12", "hos  
t_name": "Laurie", "latitude": "40.66829", "room_type": "Private  
room", "calculated_host_listings_count": "3", "host_id": "9744", "name": "Lo  
vely Room 1, Garden, Best Area, Legal  
rental", "reviews_per_month": "1.34", "neighbourhood_group": "Brooklyn", "n  
eighbourhood": "South Slope", "price": "89"},  
{ "last_review": "7/5/2019", "minimum_nights": "2", "availability_365": "333  
", "longitude": "-  
73.96113", "number_of_reviews": "113", "_deepnote_index_column": "13", "hos  
t_name": "Claudio", "latitude": "40.79826", "room_type": "Private  
room", "calculated_host_listings_count": "1", "host_id": "11528", "name": "W  
onderful Guest Bedroom in Manhattan for  
SINGLES", "reviews_per_month": "0.91", "neighbourhood_group": "Manhattan",  
"neighbourhood": "Upper West Side", "price": "85"},  
{ "last_review": "10/31/2018", "minimum_nights": "90", "availability_365": "0"  
", "longitude": "-  
74.00525", "number_of_reviews": "27", "_deepnote_index_column": "14", "host  
_name": "Alina", "latitude": "40.7353", "room_type": "Entire  
home/apt", "calculated_host_listings_count": "1", "host_id": "11975", "name":
```

```
": "West Village Nest -  
Superhost", "reviews_per_month": "0.22", "neighbourhood_group": "Manhattan",  
"neighbourhood": "West Village", "price": "120"},  
{ "last_review": "6/29/2019", "minimum_nights": "2", "availability_365": "46",  
"longitude": "-73.95352", "number_of_reviews": "148", "_deepnote_index_column": "15", "host_name": "Allen & Irina",  
"latitude": "40.70837", "room_type": "Entire home/apt", "calculated_host_listings_count": "1", "host_id": "15991", "name": "Only 2 stops to Manhattan  
studio", "reviews_per_month": "1.2", "neighbourhood_group": "Brooklyn", "neighbourhood": "Williamsburg",  
"price": "140"},  
{ "last_review": "6/28/2019", "minimum_nights": "2", "availability_365": "321", "longitude": "-73.97185",  
"number_of_reviews": "198", "_deepnote_index_column": "16", "host_name": "Jane", "latitude": "40.69169",  
"room_type": "Entire home/apt", "calculated_host_listings_count": "1", "host_id": "17571", "name": "Perfect for Your Parents +  
Garden", "reviews_per_month": "1.72", "neighbourhood_group": "Brooklyn", "neighbourhood": "Fort Greene",  
"price": "215"},  
{ "last_review": "7/1/2019", "minimum_nights": "1", "availability_365": "12", "longitude": "-73.99501",  
"number_of_reviews": "260", "_deepnote_index_column": "17", "host_name": "Doti", "latitude": "40.74192",  
"room_type": "Private room", "calculated_host_listings_count": "1", "host_id": "18946", "name": "Chelsea  
Perfect", "reviews_per_month": "2.12", "neighbourhood_group": "Manhattan", "neighbourhood": "Chelsea",  
"price": "140"},  
{ "last_review": "6/22/2019", "minimum_nights": "3", "availability_365": "21", "longitude": "-73.94694",  
"number_of_reviews": "53", "_deepnote_index_column": "18", "host_name": "Adam And Charity",  
"latitude": "40.67592", "room_type": "Entire home/apt", "calculated_host_listings_count": "1",  
"host_id": "20950", "name": "Hip Historic Brownstone Apartment with Backyard",  
"reviews_per_month": "4.44", "neighbourhood_group": "Brooklyn", "neighbourhood": "Crown Heights",  
"price": "99"},  
{ "last_review": "nan", "minimum_nights": "7", "availability_365": "249", "longitude": "-73.94872",  
"number_of_reviews": "0", "_deepnote_index_column": "19", "host_name": "Sing", "latitude": "40.79685",  
"room_type": "Entire home/apt", "calculated_host_listings_count": "2", "host_id": "17985",  
"name": "Huge 2 BR Upper East Cental Park", "reviews_per_month": "nan",  
"neighbourhood_group": "Manhattan", "neighbourhood": "East Harlem", "price": "190"},  
{ "last_review": "12/28/2011", "minimum_nights": "3", "availability_365": "0", "longitude": "-73.95718",  
"number_of_reviews": "9", "_deepnote_index_column": "20", "host_name": "Chaya", "latitude": "40.71842",  
"room_type": "Entire home/apt", "calculated_host_listings_count": "1", "host_id": "21207",  
"name": "Sweet and Spacious Brooklyn Loft", "reviews_per_month": "0.07", "neighbourhood_group": "Brooklyn", "nei
```

```

ghbourhood":"Williamsburg","price":"299"},
{"last_review":"7/1/2019","minimum_nights":"2","availability_365":"347",
"longitude":-73.97706,"number_of_reviews":"130","_deepnote_index_column":"21","host_name":"Lisel","latitude":"40.68069","room_type":"Private room","calculated_host_listings_count":"6","host_id":"22486","name":"C BG CtyBGd HelpsHaiti rm#1:1-4","reviews_per_month":"1.09","neighbourhood_group":"Brooklyn","neighbourhood":"Park Slope","price":"130"},
{"last_review":"1/1/2019","minimum_nights":"1","availability_365":"364",
"longitude":-73.97798,"number_of_reviews":"39","_deepnote_index_column":"22","host_name":"Lisel","latitude":"40.67989","room_type":"Private room","calculated_host_listings_count":"6","host_id":"22486","name":"C BG Helps Haiti Room#2.5","reviews_per_month":"0.37","neighbourhood_group":"Brooklyn","neighbourhood":"Park Slope","price":"80"},
{"last_review":"7/2/2019","minimum_nights":"2","availability_365":"304",
"longitude":-73.97865,"number_of_reviews":"71","_deepnote_index_column":"23","host_name":"Lisel","latitude":"40.68001","room_type":"Private room","calculated_host_listings_count":"6","host_id":"22486","name":"C BG Helps Haiti Rm #2","reviews_per_month":"0.61","neighbourhood_group":"Brooklyn","neighbourhood":"Park Slope","price":"110"},
{"last_review":"6/19/2019","minimum_nights":"2","availability_365":"233",
"longitude":-73.94028,"number_of_reviews":"88","_deepnote_index_column":"24","host_name":"Nathalie","latitude":"40.68371","room_type":"Entire home/apt","calculated_host_listings_count":"2","host_id":"25183","name":"MAISON DES SIRENES1,bohemian apartment","reviews_per_month":"0.73","neighbourhood_group":"Brooklyn","neighbourhood":"Bedford-Stuyvesant","price":"120"}],
"row_count":9999,"column_count":15,"columns":
[{"name":"name","dtype":"object","stats":
{"unique_count":9928,"categories":
[{"name":"Loft Suite @ The Box House Hotel","count":11},
{"name":"9927 others","count":9980},
{"name":"Missing","count":8}],
"nan_count":8}},
{"name":"host_id","dtype":"int64","stats":
{"unique_count":8628,"histogram":
[{"count":3810,"bin_end":4142121.7,"bin_start":2571},
{"count":1689,"bin_end":8281672.4,"bin_start":4142121.7},
{"count":1133,"bin_end":1.24212231000000001e7,"bin_start":8281672.4},
{"count":709,"bin_end":1.65607738e7,"bin_start":1.24212231000000001e7},
{"count":608,"bin_end":2.07003245e7,"bin_start":1.65607738e7},
{"count":727,"bin_end":2.48398752000000003e7,"bin_start":2.07003245e7},
{"count":407,"bin_end":2.89794259000000002e7,"bin_start":2.48398752000000003e7},
{"count":320,"bin_end":3.31189766e7,"bin_start":2.89794259000000002e7},
{"count":381,"bin_end":3.72585273000000004e7,"bin_start":3.31189766e7},

```



```

{"count":215,"bin_end":41398078,"bin_start":3.72585273000000004e7}], "min": "2571", "max": "41398078", "nan_count": 0}},
{"name": "host_name", "dtype": "object", "stats":
{"unique_count": 3714, "categories": [{"name": "Michael", "count": 98},
{"name": "3713 others", "count": 9891},
{"name": "Missing", "count": 10}], "nan_count": 10}},
{"name": "neighbourhood_group", "dtype": "object", "stats":
{"unique_count": 5, "categories": [{"name": "Manhattan", "count": 4583},
{"name": "Brooklyn", "count": 4488}, {"name": "3
others", "count": 928}], "nan_count": 0}},
{"name": "neighbourhood", "dtype": "object", "stats":
{"unique_count": 171, "categories": [{"name": "Williamsburg", "count": 944},
{"name": "Bedford-Stuyvesant", "count": 742}, {"name": "169
others", "count": 8313}], "nan_count": 0}},
{"name": "latitude", "dtype": "float64", "stats":
{"unique_count": 7579, "histogram":
[{"count": 3, "bin_end": 40.548615999999996, "bin_start": 40.50868},
{"count": 30, "bin_end": 40.588552, "bin_start": 40.548615999999996},
{"count": 106, "bin_end": 40.628488, "bin_start": 40.588552},
{"count": 748, "bin_end": 40.668424, "bin_start": 40.628488},
{"count": 2743, "bin_end": 40.70836, "bin_start": 40.668424},
{"count": 3220, "bin_end": 40.748295999999996, "bin_start": 40.70836},
{"count": 1628, "bin_end": 40.788232, "bin_start": 40.748295999999996},
{"count": 1136, "bin_end": 40.828168, "bin_start": 40.788232},
{"count": 338, "bin_end": 40.868104, "bin_start": 40.828168},
{"count": 47, "bin_end": 40.90804, "bin_start": 40.868104}], "min": "40.50868", "max": "40.90804", "nan_count": 0}},
{"name": "longitude", "dtype": "float64", "stats":
{"unique_count": 6351, "histogram": [{"count": 2, "bin_end": -74.188774999999999, "bin_start": -74.23986}, {"count": 8, "bin_end": -74.137689999999999, "bin_start": -74.188774999999999}, {"count": 12, "bin_end": -74.086604999999999, "bin_start": -74.137689999999999}, {"count": 32, "bin_end": -74.035519999999999, "bin_start": -74.086604999999999}, {"count": 2324, "bin_end": -73.984434999999999, "bin_start": -74.035519999999999}, {"count": 6031, "bin_end": -73.93335, "bin_start": -73.984434999999999}, {"count": 1310, "bin_end": -73.882265, "bin_start": -73.93335}, {"count": 159, "bin_end": -73.83118, "bin_start": -73.882265}, {"count": 82, "bin_end": -73.780095, "bin_start": -73.83118}, {"count": 39, "bin_end": -73.72901, "bin_start": -73.780095}], "min": "-74.23986", "max": "-73.72901", "nan_count": 0}},
{"name": "room_type", "dtype": "object", "stats":
{"unique_count": 3, "categories": [{"name": "Entire home/apt", "count": 5811}, {"name": "Private room", "count": 4054}, {"name": "Shared room", "count": 134}], "nan_count": 0}},
{"name": "price", "dtype": "int64", "stats":
{"unique_count": 389, "histogram":
[{"count": 9939, "bin_end": 1009, "bin_start": 10},
{"count": 41, "bin_end": 2008, "bin_start": 1009},
{"count": 7, "bin_end": 3007, "bin_start": 2008},

```

```

{"count":3,"bin_end":4006,"bin_start":3007},
{"count":3,"bin_end":5005,"bin_start":4006},
{"count":2,"bin_end":6004,"bin_start":5005},
{"count":1,"bin_end":7003,"bin_start":6004},
{"count":1,"bin_end":8002,"bin_start":7003},
{"count":0,"bin_end":9001,"bin_start":8002},
{"count":2,"bin_end":10000,"bin_start":9001}], "min": "10", "max": "10000",
"nan_count": 0}}, {"name": "minimum_nights", "dtype": "int64", "stats":
{"unique_count": 75, "histogram":
[{"count": 9953, "bin_end": 125.9, "bin_start": 1},
{"count": 22, "bin_end": 250.8, "bin_start": 125.9},
{"count": 20, "bin_end": 375.70000000000005, "bin_start": 250.8},
{"count": 2, "bin_end": 500.6, "bin_start": 375.70000000000005},
{"count": 0, "bin_end": 625.5, "bin_start": 500.6},
{"count": 0, "bin_end": 750.40000000000001, "bin_start": 625.5},
{"count": 0, "bin_end": 875.30000000000001, "bin_start": 750.40000000000001},
{"count": 1, "bin_end": 1000.2, "bin_start": 875.30000000000001},
{"count": 0, "bin_end": 1125.10000000000001, "bin_start": 1000.2},
{"count": 1, "bin_end": 1250, "bin_start": 1125.10000000000001}], "min": "1", "
max": "1250", "nan_count": 0}},
{"name": "number_of_reviews", "dtype": "int64"},
{"name": "last_review", "dtype": "object"},
{"name": "reviews_per_month", "dtype": "float64"},
{"name": "calculated_host_listings_count", "dtype": "int64"},
{"name": "availability_365", "dtype": "int64"},
{"name": "_deepnote_index_column", "dtype": "int64"}]]

```

```

data = data.dropna()
data = data.reset_index(drop=True)
data

```

```

{"rows":
[{"last_review": "10/19/2018", "minimum_nights": "1", "availability_365": "365", "longitude": "-73.97237", "number_of_reviews": "9", "_deepnote_index_column": "0", "host_name": "John", "latitude": "40.64749", "room_type": "Private room", "calculated_host_listings_count": "6", "host_id": "2787", "name": "Clean & quiet apt home by the park", "reviews_per_month": "0.21", "neighbourhood_group": "Brooklyn", "neighbourhood": "Kensington", "price": "149"},
{"last_review": "5/21/2019", "minimum_nights": "1", "availability_365": "355", "longitude": "-73.98377", "number_of_reviews": "45", "_deepnote_index_column": "1", "host_name": "Jennifer", "latitude": "40.75362", "room_type": "Entire home/apt", "calculated_host_listings_count": "2", "host_id": "2845", "name": "Skylit Midtown Castle", "reviews_per_month": "0.38", "neighbourhood_group": "Manhattan", "neighbourhood": "Midtown", "price": "225"},
{"last_review": "7/5/2019", "minimum_nights": "1", "availability_365": "194", "longitude": "-73.95976", "number_of_reviews": "270", "_deepnote_index_column": "2", "host

```

```
_name":"LisaRoxanne","latitude":"40.68514","room_type":"Entire
home/apt","calculated_host_listings_count":"1","host_id":"4869","name"
:"Cozy Entire Floor of
Brownstone","reviews_per_month":"4.64","neighbourhood_group":"Brooklyn
","neighbourhood":"Clinton Hill","price":"89"},
{"last_review":"11/19/2018","minimum_nights":"10","availability_365":"
0","longitude":"-
73.94399","number_of_reviews":"9","_deepnote_index_column":"3","host_n
ame":"Laura","latitude":"40.79851","room_type":"Entire
home/apt","calculated_host_listings_count":"1","host_id":"7192","name"
:"Entire Apt: Spacious Studio/Loft by central
park","reviews_per_month":"0.1","neighbourhood_group":"Manhattan","nei
ghbourhood":"East Harlem","price":"80"},
{"last_review":"6/22/2019","minimum_nights":"3","availability_365":"12
9","longitude":"-
73.975","number_of_reviews":"74","_deepnote_index_column":"4","host_na
me":"Chris","latitude":"40.74767","room_type":"Entire
home/apt","calculated_host_listings_count":"1","host_id":"7322","name"
:"Large Cozy 1 BR Apartment In Midtown
East","reviews_per_month":"0.59","neighbourhood_group":"Manhattan","ne
ighbourhood":"Murray Hill","price":"200"},
{"last_review":"10/5/2017","minimum_nights":"45","availability_365":"0
","longitude":"-
73.95596","number_of_reviews":"49","_deepnote_index_column":"5","host_
name":"Garon","latitude":"40.68688","room_type":"Private
room","calculated_host_listings_count":"1","host_id":"7356","name":"Bl
issArtsSpace!","reviews_per_month":"0.4","neighbourhood_group":"Brookl
yn","neighbourhood":"Bedford-Stuyvesant","price":"60"},
{"last_review":"6/24/2019","minimum_nights":"2","availability_365":"22
0","longitude":"-
73.98493","number_of_reviews":"430","_deepnote_index_column":"6","host
_name":"Shunichi","latitude":"40.76489","room_type":"Private
room","calculated_host_listings_count":"1","host_id":"8967","name":"La
rge Furnished Room Near B'way
","reviews_per_month":"3.47","neighbourhood_group":"Manhattan","neighb
ourhood":"Hell's Kitchen","price":"79"},
{"last_review":"7/21/2017","minimum_nights":"2","availability_365":"0"
,"longitude":"-
73.96723","number_of_reviews":"118","_deepnote_index_column":"7","host
_name":"MaryEllen","latitude":"40.80178","room_type":"Private
room","calculated_host_listings_count":"1","host_id":"7490","name":"Co
zy Clean Guest Room - Family
Apt","reviews_per_month":"0.99","neighbourhood_group":"Manhattan","nei
ghbourhood":"Upper West Side","price":"79"},
{"last_review":"6/9/2019","minimum_nights":"1","availability_365":"188
","longitude":"-
73.99037","number_of_reviews":"160","_deepnote_index_column":"8","host
_name":"Ben","latitude":"40.71344","room_type":"Entire
home/apt","calculated_host_listings_count":"4","host_id":"7549","name"
:"Cute & Cozy Lower East Side 1
```



```

bdrm", "reviews_per_month": "1.33", "neighbourhood_group": "Manhattan", "neighbourhood": "Chinatown", "price": "150"},
{"last_review": "6/22/2019", "minimum_nights": "5", "availability_365": "6", "longitude": "-73.96545", "number_of_reviews": "53", "_deepnote_index_column": "9", "host_name": "Lena", "latitude": "40.80316", "room_type": "Entire home/apt", "calculated_host_listings_count": "1", "host_id": "7702", "name": "Beautiful 1br on Upper West Side", "reviews_per_month": "0.43", "neighbourhood_group": "Manhattan", "neighbourhood": "Upper West Side", "price": "135"}], "row_count": 8666, "column_count": 15, "columns": [{"name": "name", "dtype": "object", "stats": {"unique_count": 8613, "categories": [{"name": "Loft Suite @ The Box House Hotel", "count": 11}, {"name": "Loft Suite @ Box House Hotel", "count": 4}, {"name": "8611 others", "count": 8651}], "nan_count": 0}}, {"name": "host_id", "dtype": "int64", "stats": {"unique_count": 7372, "histogram": [{"count": 3424, "bin_end": 4039429.6, "bin_start": 2571}, {"count": 1514, "bin_end": 8076288.2, "bin_start": 4039429.6}, {"count": 904, "bin_end": 1.21131468e7, "bin_start": 8076288.2}, {"count": 560, "bin_end": 1.61500054e7, "bin_start": 1.21131468e7}, {"count": 549, "bin_end": 20186864, "bin_start": 1.61500054e7}, {"count": 566, "bin_end": 2.42237226e7, "bin_start": 20186864}, {"count": 399, "bin_end": 2.82605812e7, "bin_start": 2.42237226e7}, {"count": 236, "bin_end": 3.22974398e7, "bin_start": 2.82605812e7}, {"count": 280, "bin_end": 3.63342984e7, "bin_start": 3.22974398e7}, {"count": 234, "bin_end": 40371157, "bin_start": 3.63342984e7}], "min": 2571, "max": 40371157, "nan_count": 0}}, {"name": "host_name", "dtype": "object", "stats": {"unique_count": 3363, "categories": [{"name": "Michael", "count": 87}, {"name": "Mike", "count": 66}, {"name": "3361 others", "count": 8513}], "nan_count": 0}}, {"name": "neighbourhood_group", "dtype": "object", "stats": {"unique_count": 5, "categories": [{"name": "Manhattan", "count": 3932}, {"name": "Brooklyn", "count": 3907}, {"name": "3 others", "count": 827}], "nan_count": 0}}, {"name": "neighbourhood", "dtype": "object", "stats": {"unique_count": 168, "categories": [{"name": "Williamsburg", "count": 796}, {"name": "Bedford-Stuyvesant", "count": 660}, {"name": "166 others", "count": 7210}], "nan_count": 0}}, {"name": "latitude", "dtype": "float64", "stats": {"unique_count": 6780, "histogram": [{"count": 2, "bin_end": 40.548615999999996, "bin_start": 40.50868}, {"count": 28, "bin_end": 40.588552, "bin_start": 40.548615999999996}, {"count": 92, "bin_end": 40.628488, "bin_start": 40.588552}, {"count": 646, "bin_end": 40.668424, "bin_start": 40.628488}, {"count": 2403, "bin_end": 40.70836, "bin_start": 40.668424}, {"count": 2780, "bin_end": 40.748295999999996, "bin_start": 40.70836}, {"count": 1381, "bin_end": 40.788232, "bin_start": 40.748295999999996}, {"count": 1000, "bin_end": 40.828168, "bin_start": 40.788232},

```

```

{"count":290,"bin_end":40.868104,"bin_start":40.828168},
{"count":44,"bin_end":40.90804,"bin_start":40.868104}], "min": "40.50868", "max": "40.90804", "nan_count":0}},
{"name": "longitude", "dtype": "float64", "stats":
{"unique_count":5799, "histogram":[{"count":1, "bin_end":-74.18877499999999, "bin_start":-74.23986}, {"count":8, "bin_end":-74.13768999999999, "bin_start":-74.18877499999999}, {"count":10, "bin_end":-74.08660499999999, "bin_start":-74.13768999999999}, {"count":27, "bin_end":-74.03551999999999, "bin_start":-74.08660499999999}, {"count":2007, "bin_end":-73.98443499999999, "bin_start":-74.03551999999999}, {"count":5210, "bin_end":-73.93335, "bin_start":-73.98443499999999}, {"count":1142, "bin_end":-73.882265, "bin_start":-73.93335}, {"count":148, "bin_end":-73.83118, "bin_start":-73.882265}, {"count":78, "bin_end":-73.780095, "bin_start":-73.83118}, {"count":35, "bin_end":-73.72901, "bin_start":-73.780095}], "min": "-74.23986", "max": "-73.72901", "nan_count":0}},
{"name": "room_type", "dtype": "object", "stats":
{"unique_count":3, "categories":[{"name": "Entire home/apt", "count":5142}, {"name": "Private room", "count":3417}, {"name": "Shared room", "count":107}], "nan_count":0}},
{"name": "price", "dtype": "int64", "stats":
{"unique_count":367, "histogram":
[{"count":8638, "bin_end":1014.4, "bin_start":16}, {"count":18, "bin_end":2012.8, "bin_start":1014.4}, {"count":3, "bin_end":3011.2, "bin_start":2012.8}, {"count":0, "bin_end":4009.6, "bin_start":3011.2}, {"count":3, "bin_end":5008, "bin_start":4009.6}, {"count":1, "bin_end":6006.4, "bin_start":5008}, {"count":0, "bin_end":7004.8, "bin_start":6006.4}, {"count":1, "bin_end":8003.2, "bin_start":7004.8}, {"count":0, "bin_end":9001.6, "bin_start":8003.2}, {"count":2, "bin_end":10000, "bin_start":9001.6}], "min": "16", "max": "10000", "nan_count":0}}, {"name": "minimum_nights", "dtype": "int64", "stats":
{"unique_count":70, "histogram":
[{"count":8633, "bin_end":125.9, "bin_start":1}, {"count":16, "bin_end":250.8, "bin_start":125.9}, {"count":15, "bin_end":375.70000000000005, "bin_start":250.8}, {"count":1, "bin_end":500.6, "bin_start":375.70000000000005}, {"count":0, "bin_end":625.5, "bin_start":500.6}, {"count":0, "bin_end":750.40000000000001, "bin_start":625.5}, {"count":0, "bin_end":875.30000000000001, "bin_start":750.40000000000001}, {"count":0, "bin_end":1000.2, "bin_start":875.30000000000001}, {"count":0, "bin_end":1125.10000000000001, "bin_start":1000.2}, {"count":1, "bin_end":1250, "bin_start":1125.10000000000001}], "min": "1", "max": "1250", "nan_count":0}},
{"name": "number_of_reviews", "dtype": "int64", "stats":
{"unique_count":375, "histogram":
[{"count":6256, "bin_end":61.6, "bin_start":1}, {"count":1145, "bin_end":122.2, "bin_start":61.6},

```

```
{
  "count": 662, "bin_end": 182.8, "bin_start": 122.2},
  {"count": 376, "bin_end": 243.4, "bin_start": 182.8},
  {"count": 125, "bin_end": 304, "bin_start": 243.4},
  {"count": 59, "bin_end": 364.6, "bin_start": 304},
  {"count": 23, "bin_end": 425.2, "bin_start": 364.6},
  {"count": 15, "bin_end": 485.8, "bin_start": 425.2},
  {"count": 2, "bin_end": 546.4, "bin_start": 485.8},
  {"count": 3, "bin_end": 607, "bin_start": 546.4}], "min": "1", "max": "607", "na_n_count": 0}}, {"name": "last_review", "dtype": "object"}, {"name": "reviews_per_month", "dtype": "float64"}, {"name": "calculated_host_listings_count", "dtype": "int64"}, {"name": "availability_365", "dtype": "int64"}, {"name": "_deepnote_index_column", "dtype": "int64"}]]
```

1. Display histograms for the numerical features.

```
display(type(data))
display(data.head(5))
display(data.tail(5))
print("Our datasets has dimensions", data.shape)
#print("Our datasets has dimensions", data.shape[0], "by",
data.shape[1])
```

pandas.core.frame.DataFrame

	name	host_id
host_name \		
0	Clean & quiet apt home by the park	2787
John		
1	Skylit Midtown Castle	2845
Jennifer		
2	Cozy Entire Floor of Brownstone	4869
LisaRoxanne		
3	Entire Apt: Spacious Studio/Loft by central park	7192
Laura		
4	Large Cozy 1 BR Apartment In Midtown East	7322
Chris		

	neighbourhood_group	neighbourhood	latitude	longitude	
room_type \					
0	Brooklyn	Kensington	40.64749	-73.97237	Private
room					
1	Manhattan	Midtown	40.75362	-73.98377	Entire
home/apt					
2	Brooklyn	Clinton Hill	40.68514	-73.95976	Entire
home/apt					
3	Manhattan	East Harlem	40.79851	-73.94399	Entire
home/apt					
4	Manhattan	Murray Hill	40.74767	-73.97500	Entire
home/apt					

	price	minimum_nights	number_of_reviews	last_review
reviews_per_month \				
0	149	1	9	10/19/2018
0.21				
1	225	1	45	5/21/2019
0.38				
2	89	1	270	7/5/2019
4.64				
3	80	10	9	11/19/2018
0.10				
4	200	3	74	6/22/2019
0.59				

	calculated_host_listings_count	availability_365
0	6	365
1	2	355
2	1	194
3	1	0
4	1	129

	name	host_id	host_name \
8661	The Wielingen	40371157	Jimmy
8662	Perfect Location - Meticulously Kept Flat	12620454	Will
8663	Garden Apt in Historic Brownstone!	2060383	Lisa
8664	East Village Private Room & Terrace	39956905	Can
8665	Cosy apartment in Carroll Gardens	33064750	Suzan

	neighbourhood_group	neighbourhood	latitude	longitude \
8661	Bronx	Longwood	40.82417	-73.90156
8662	Brooklyn	Bushwick	40.70442	-73.92484
8663	Brooklyn	Cobble Hill	40.68732	-73.99245
8664	Manhattan	East Village	40.72811	-73.98453
8665	Brooklyn	Carroll Gardens	40.68282	-73.99774

	room_type	price	minimum_nights	number_of_reviews
last_review \				
8661	Private room	95	1	182
1/1/2019				
8662	Entire home/apt	220	5	27
1/1/2017				
8663	Entire home/apt	147	3	23
6/16/2019				
8664	Private room	95	2	1
8/29/2015				
8665	Entire home/apt	160	5	2
8/8/2017				

	reviews_per_month	calculated_host_listings_count
availability_365		
8661	3.82	2

```

0
8662          0.57          1
0
8663          0.51          1
2
8664          0.02          2
0
8665          0.06          1
0

```

Our datasets has dimensions (8666, 15)

```

names = list(data)
print(names)

```

```

['name', 'host_id', 'host_name', 'neighbourhood_group',
'neighbourhood', 'latitude', 'longitude', 'room_type', 'price',
'minimum_nights', 'number_of_reviews', 'last_review',
'reviews_per_month', 'calculated_host_listings_count',
'availability_365']

```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 8666 entries, 0 to 8665
```

```
Data columns (total 15 columns):
```

#	Column	Non-Null Count	Dtype
0	name	8666 non-null	object
1	host_id	8666 non-null	int64
2	host_name	8666 non-null	object
3	neighbourhood_group	8666 non-null	object
4	neighbourhood	8666 non-null	object
5	latitude	8666 non-null	float64
6	longitude	8666 non-null	float64
7	room_type	8666 non-null	object
8	price	8666 non-null	int64
9	minimum_nights	8666 non-null	int64
10	number_of_reviews	8666 non-null	int64
11	last_review	8666 non-null	object
12	reviews_per_month	8666 non-null	float64
13	calculated_host_listings_count	8666 non-null	int64
14	availability_365	8666 non-null	int64

```
dtypes: float64(3), int64(6), object(6)
```

```
memory usage: 1015.7+ KB
```

Only interested in numerical values/columns so int64 and float64 will be taken into consideration.

```

from matplotlib import pyplot as plt
fig, ax = plt.subplots()

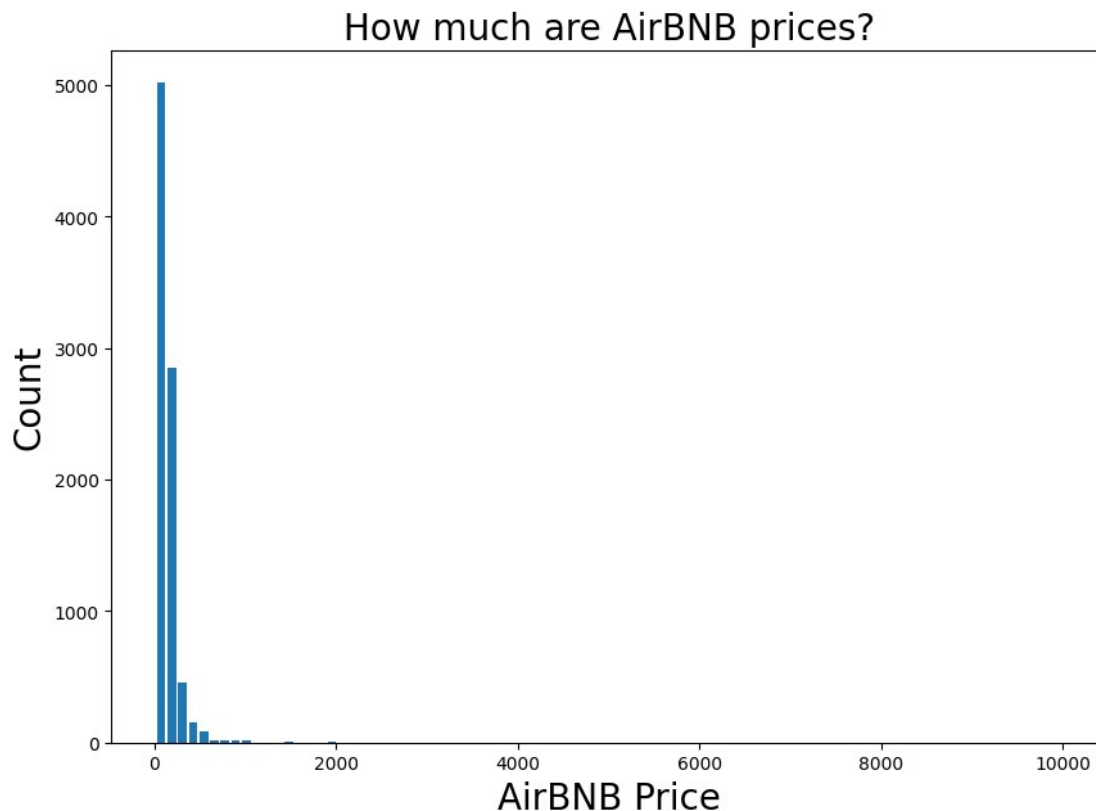
```



```

ax.hist(data.price,bins=85, rwidth=.8)
ax.set_xlabel("AirBNB Price", fontsize=20)
ax.set_ylabel("Count", fontsize=20)
ax.set_title("How much are AirBNB prices?", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()

```



Obtain minimum, maximum and average AirBNB Prices:

```

print("Minimum AirBNB Price:", min(data.price))
print("Maximum AirBNB Price:", max(data.price))
print("Average AirBNB Price:", np.mean(data.price))

```

```

Minimum AirBNB Price: 16
Maximum AirBNB Price: 10000
Average AirBNB Price: 153.09889222247864

```

There are some prices that are outliers for AirBNB prices. So all prices above 1500 are cut and plotted down below. We define a logical index $z = \text{data.price} < 1500$, which will only return that values for which the index z is true. This was taken from Professors master tutorial.

```

from matplotlib import pyplot as plt
fig, ax = plt.subplots()

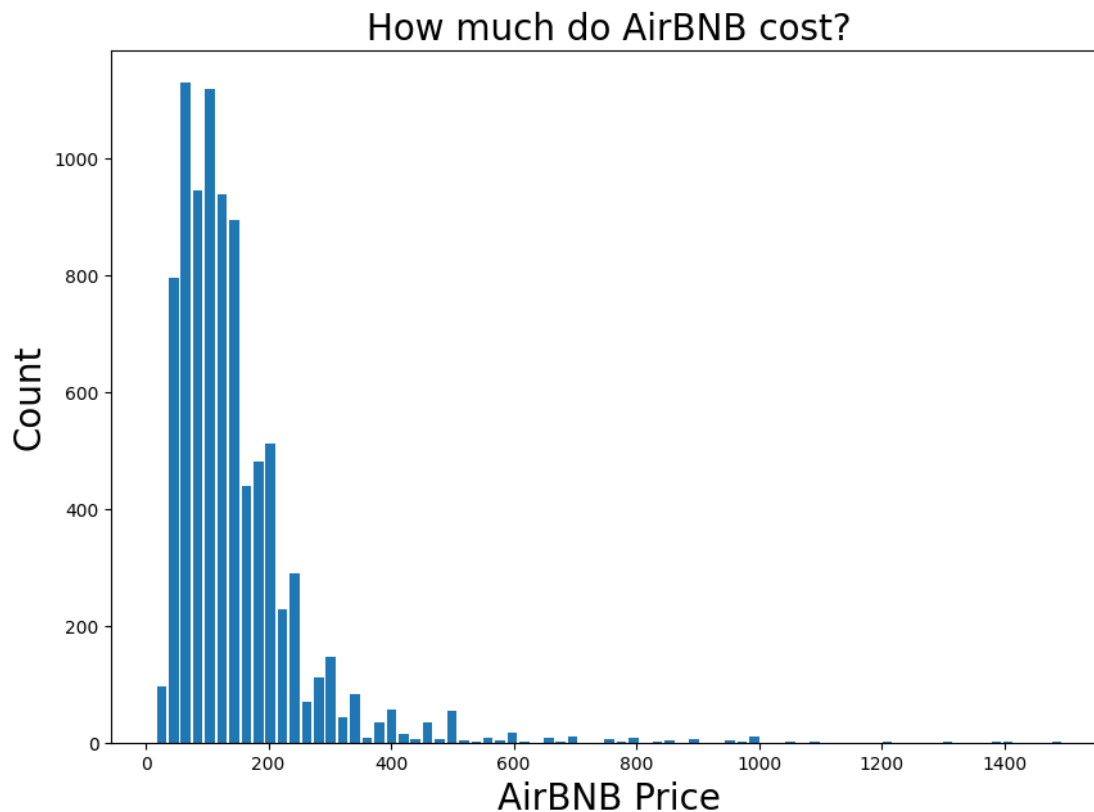
```

```

z = data.price < 1500

ax.hist(data.price[z],bins=75, rwidth=.8)
ax.set_xlabel("AirBNB Price", fontsize=20)
ax.set_ylabel("Count", fontsize=20)
ax.set_title("How much do AirBNB cost?", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()

```



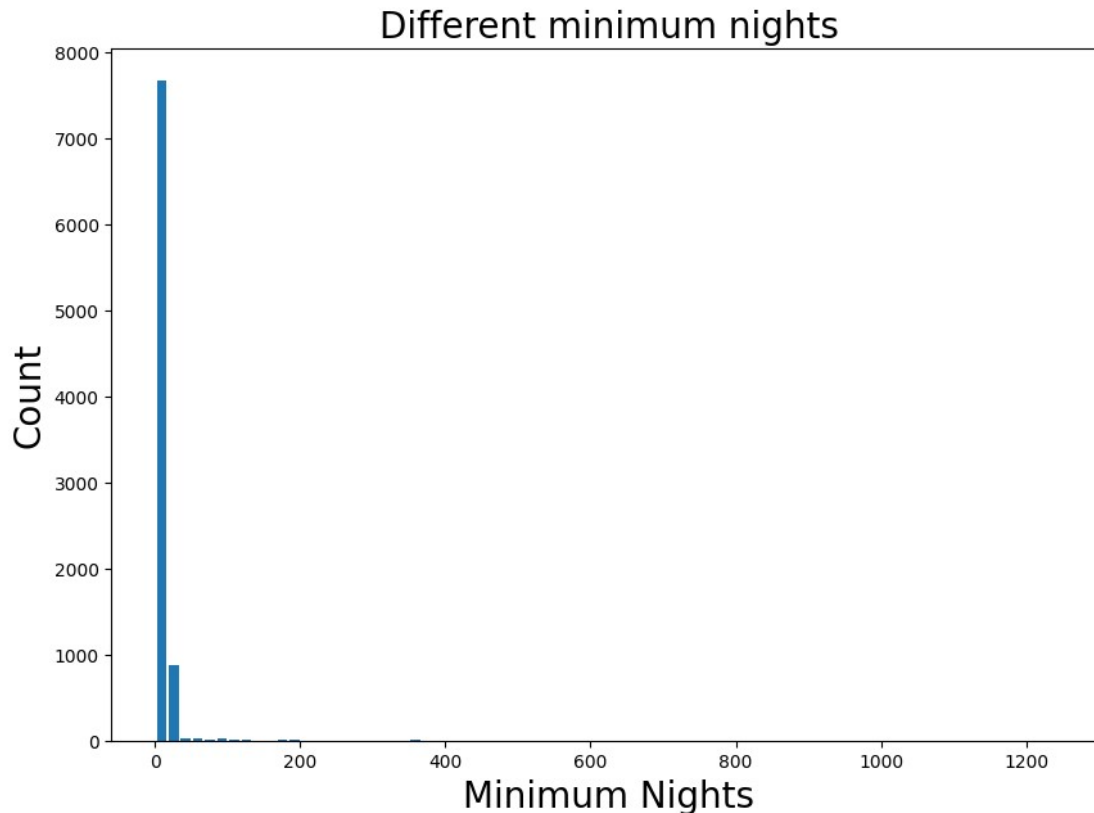
Next up is minimum nights

```

from matplotlib import pyplot as plt
fig, ax = plt.subplots()

ax.hist(data.minimum_nights,bins=75, rwidth=.8)
ax.set_xlabel("Minimum Nights", fontsize=20)
ax.set_ylabel("Count", fontsize=20)
ax.set_title("Different minimum nights", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()

```



Obtain minimum, maximum and average for minimum nights:

```
print("Minimum(lowest) minimum nights stay:",
min(data.minimum_nights))
print("Maximum(highest) minimum nights stay:",
max(data.minimum_nights))
print("Average minimum nights stay:", np.mean(data.minimum_nights))
```

```
Minimum(lowest) minimum nights stay: 1
Maximum(highest) minimum nights stay: 1250
Average minimum nights stay: 7.764943457189014
```

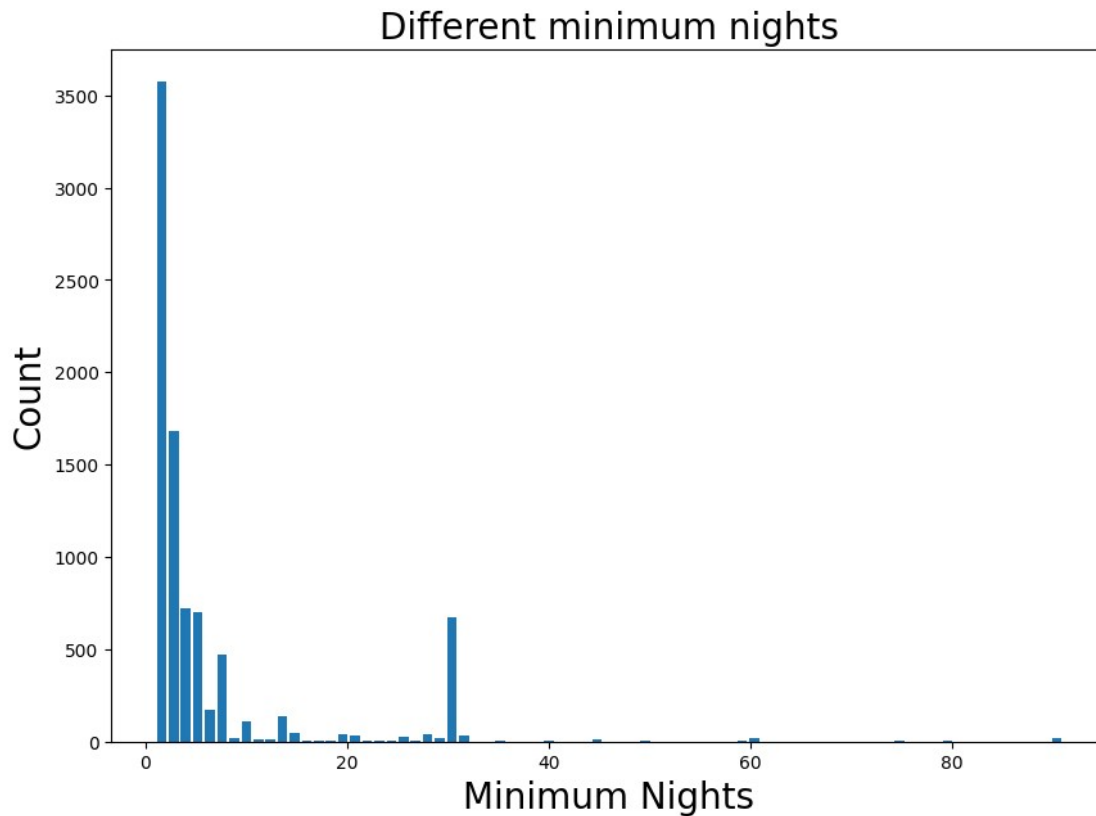
As was done for prices, there are some minimum nights that are outliers. So all minimum nights above 100 are cut and plotted down below. We define a logical index $z = \text{data.minimum_nights} < 100$, which will only return that values for which the index z is true. This was taken from Professors master tutorial.

```
from matplotlib import pyplot as plt
fig, ax = plt.subplots()
```

```
z = data.minimum_nights < 100
```

```
ax.hist(data.minimum_nights[z],bins=75, rwidth=.8)
ax.set_xlabel("Minimum Nights", fontsize=20)
ax.set_ylabel("Count", fontsize=20)
```

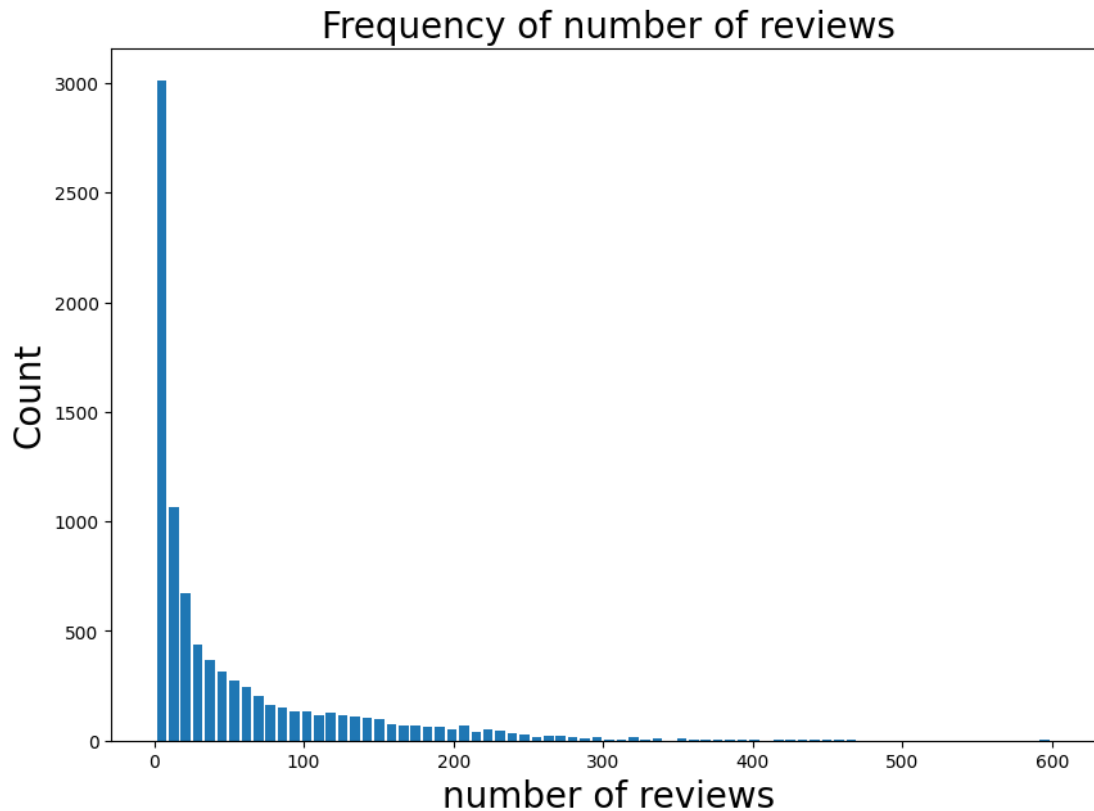
```
ax.set_title("Different minimum nights", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()
```



Next is number_of_reviews

```
from matplotlib import pyplot as plt
fig, ax = plt.subplots()

ax.hist(data.number_of_reviews, bins=75, rwidth=.8)
ax.set_xlabel("number of reviews", fontsize=20)
ax.set_ylabel("Count", fontsize=20)
ax.set_title("Frequency of number of reviews", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()
```



Obtain minimum, maximum and average for number of reviews:

```
print("Minimum number of reviews:", min(data.number_of_reviews))
print("Maximum number of reviews:", max(data.number_of_reviews))
print("Average number of reviews:", np.mean(data.number_of_reviews))
```

```
Minimum number of reviews: 1
Maximum number of reviews: 607
Average number of reviews: 52.20609277636741
```

As was done for previous numerical values, there are some number of reviews that are not that useful as their frequency is low and they are outliers. So all number of reviews above 200 are cut and plotted down below. We define a logical index $z = \text{data.number_of_reviews} < 200$, which will only return that values for which the index z is true. This was taken from Professors master tutorial.

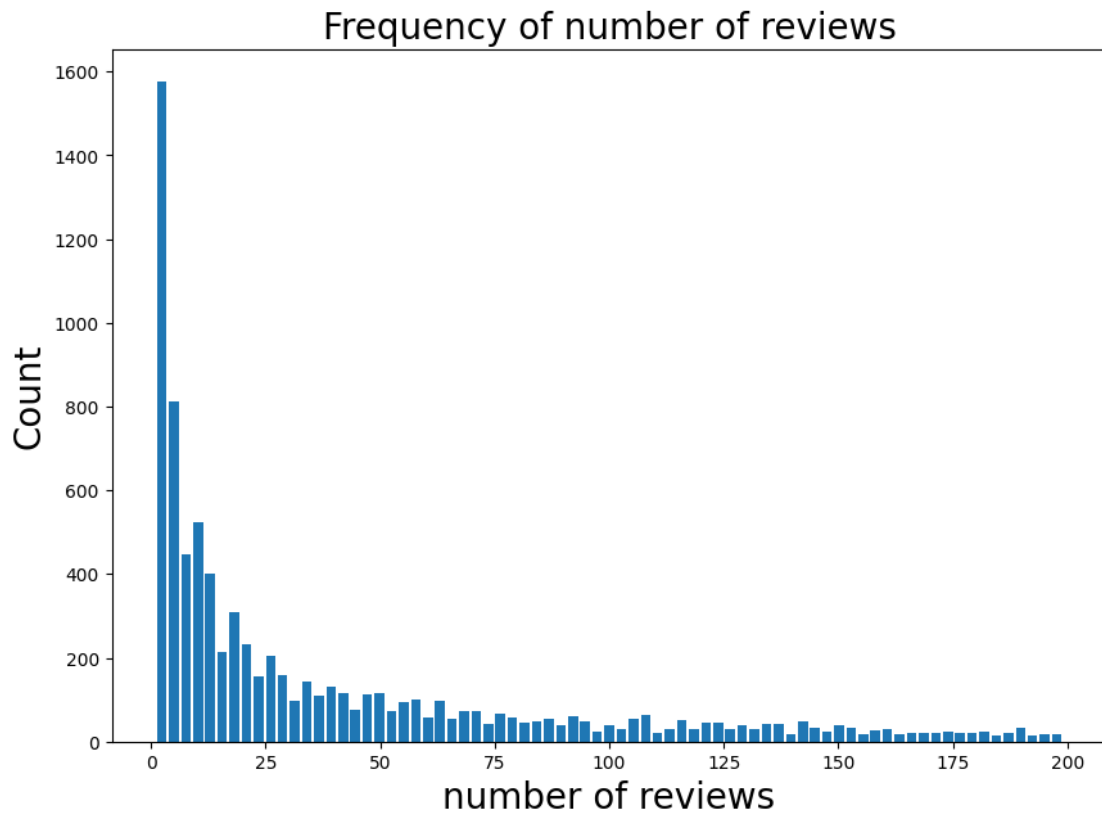
```
from matplotlib import pyplot as plt
fig, ax = plt.subplots()
```

```
z = data.number_of_reviews < 200
```

```
ax.hist(data.number_of_reviews[z], bins=75, rwidth=.8)
ax.set_xlabel("number of reviews", fontsize=20)
ax.set_ylabel("Count", fontsize=20)
ax.set_title("Frequency of number of reviews", fontsize=20)
```



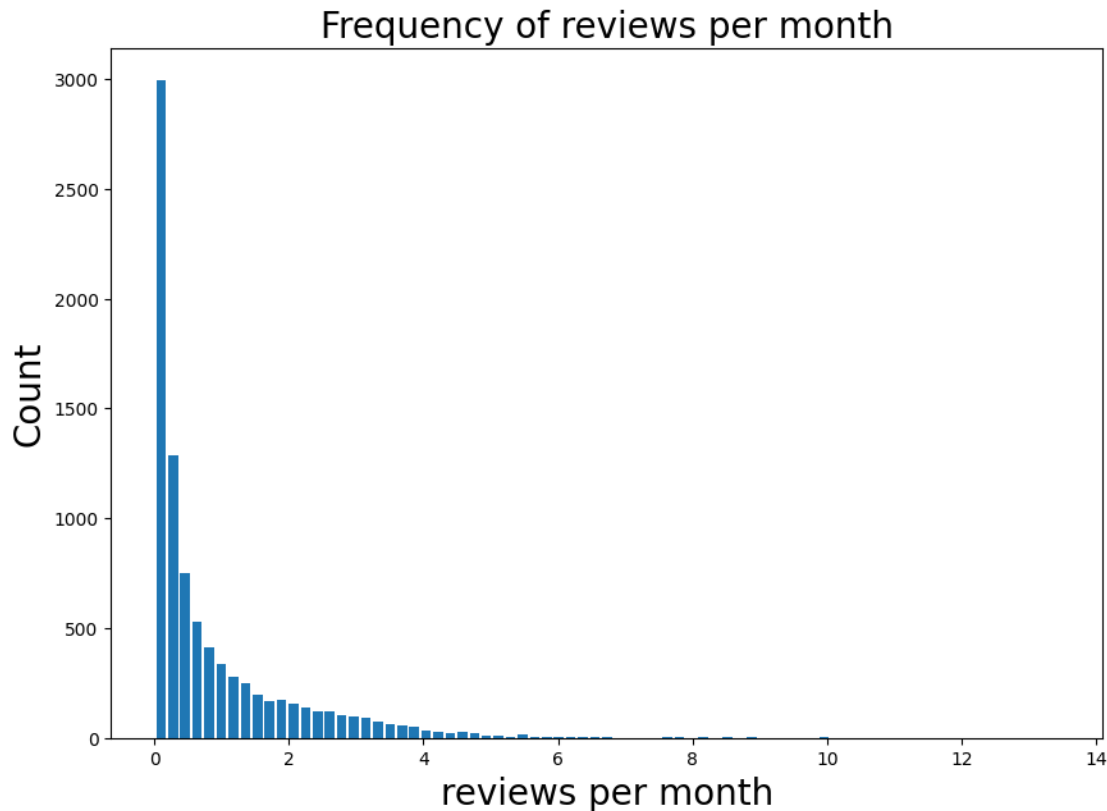
```
fig.set_size_inches(10, 7)
plt.show()
```



Next is reviews_per_month

```
from matplotlib import pyplot as plt
fig, ax = plt.subplots()

ax.hist(data.reviews_per_month, bins=75, rwidth=.8)
ax.set_xlabel("reviews per month", fontsize=20)
ax.set_ylabel("Count", fontsize=20)
ax.set_title("Frequency of reviews per month", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()
```



Obtain minimum, maximum and average for reviews per month:

```
print("Minimum reviews per month:", min(data.reviews_per_month))
print("Maximum reviews per month:", max(data.reviews_per_month))
print("Average reviews per month:", np.mean(data.reviews_per_month))
```

```
Minimum reviews per month: 0.01
Maximum reviews per month: 13.45
Average reviews per month: 0.8762924071082391
```

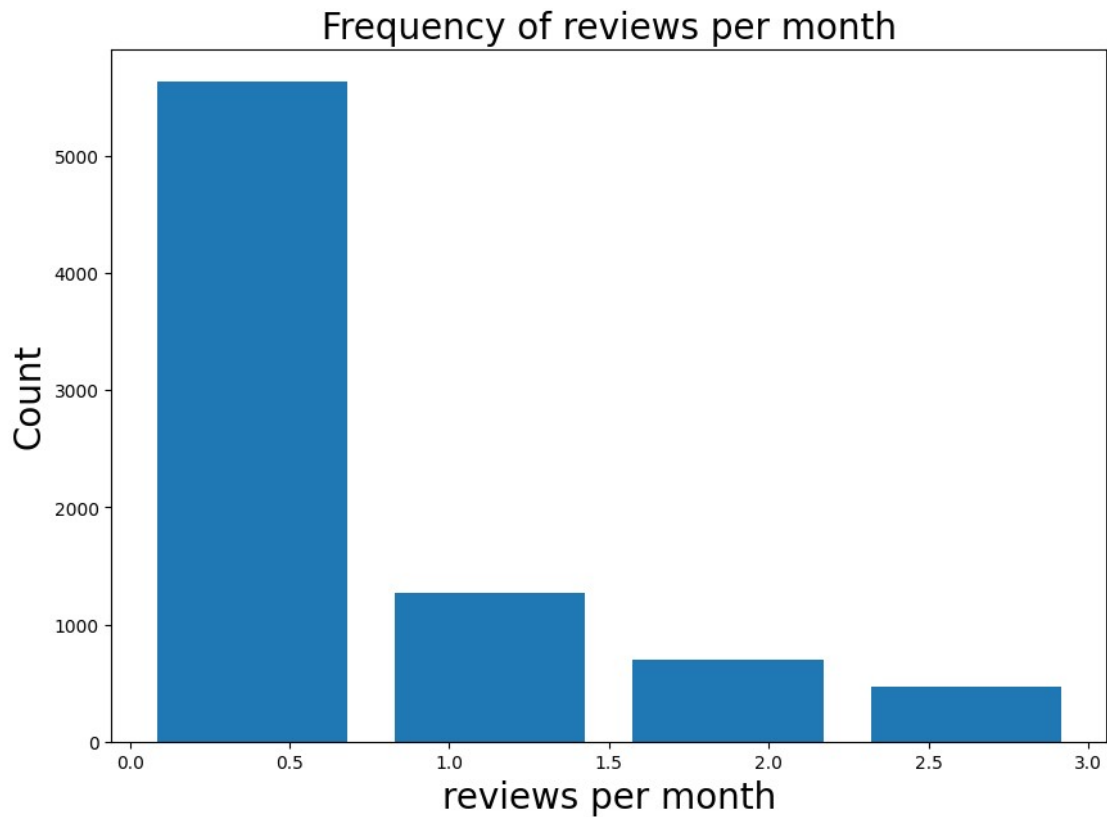
As was done for previous numerical values, there are some reviews per month that are not that useful as their frequency is low and they are outliers. So all number of reviews above 3 are cut and plotted down below. We define a logical index $z = \text{data.reviews_per_month} < 3$, which will only return that values for which the index z is true. This was taken from Professors master tutorial.

```
from matplotlib import pyplot as plt
fig, ax = plt.subplots()
```

```
z = data.reviews_per_month < 3
```

```
ax.hist(data.reviews_per_month[z], bins=4, rwidth=.8)
ax.set_xlabel("reviews per month", fontsize=20)
ax.set_ylabel("Count", fontsize=20)
ax.set_title("Frequency of reviews per month", fontsize=20)
```

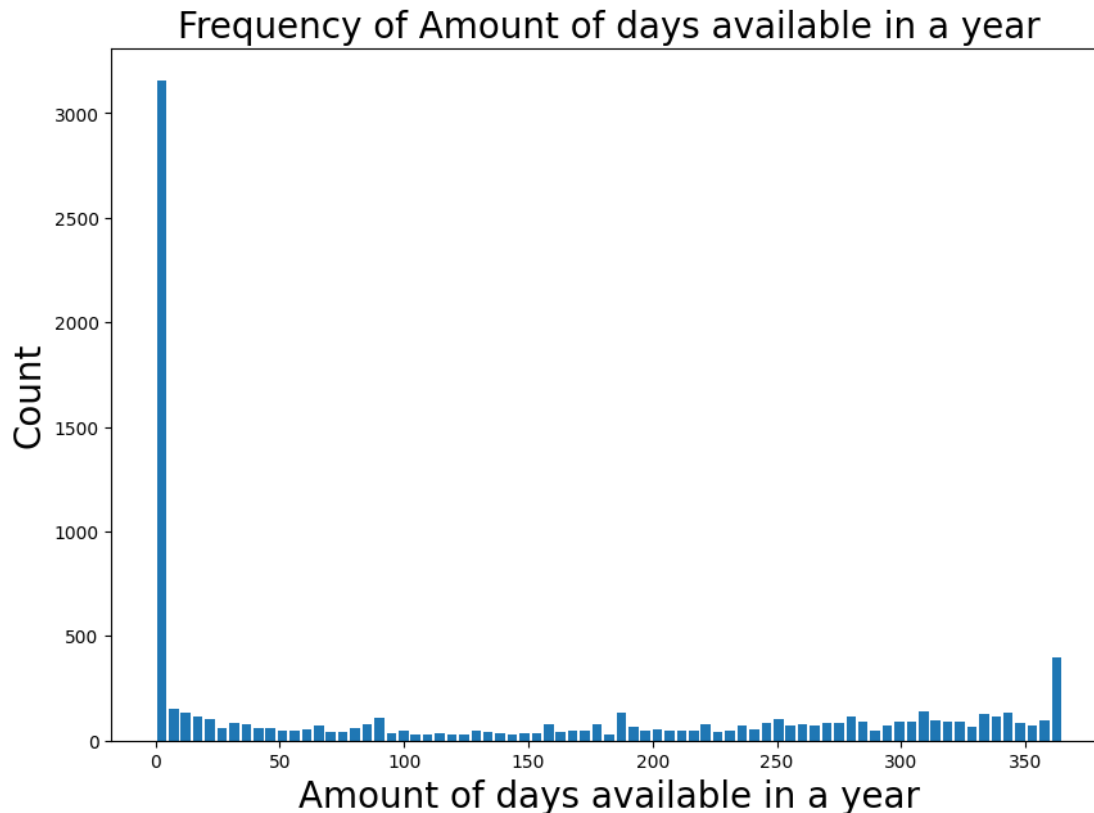
```
fig.set_size_inches(10, 7)
plt.show()
```



Next up is availability_365

```
from matplotlib import pyplot as plt
fig, ax = plt.subplots()

ax.hist(data.availability_365, bins=75, rwidth=.8)
ax.set_xlabel("Amount of days available in a year", fontsize=20)
ax.set_ylabel("Count", fontsize=20)
ax.set_title("Frequency of Amount of days available in a year",
             fontsize=20)
fig.set_size_inches(10, 7)
plt.show()
```



Obtain minimum, maximum and average for availability_365

```
print("Minimum availability in a year:", min(data.availability_365))
print("Maximum availability in a year:", max(data.availability_365))
print("Average availability in a year:",
      np.mean(data.availability_365))
```

Minimum availability in a year: 0

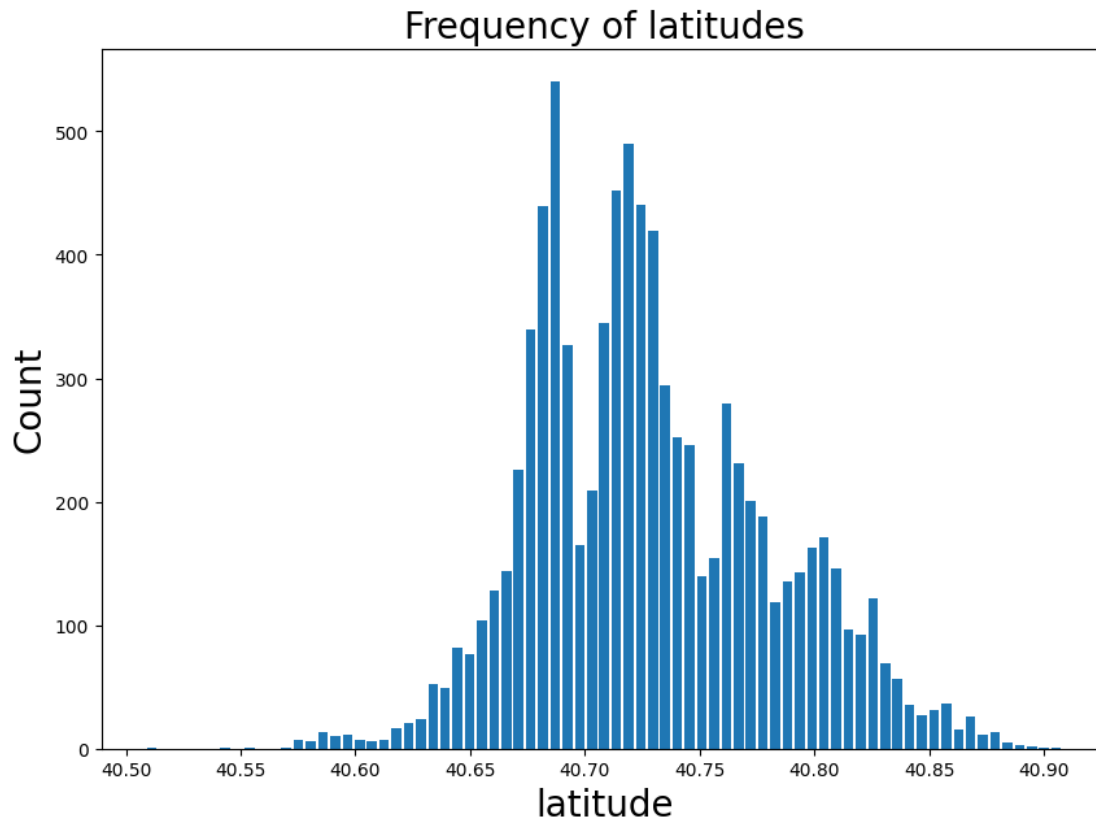
Maximum availability in a year: 365

Average availability in a year: 131.81513962612507

Next up is latitude

```
from matplotlib import pyplot as plt
fig, ax = plt.subplots()
```

```
ax.hist(data.latitude, bins=75, rwidth=.8)
ax.set_xlabel("latitude", fontsize=20)
ax.set_ylabel("Count", fontsize=20)
ax.set_title("Frequency of latitudes", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()
```



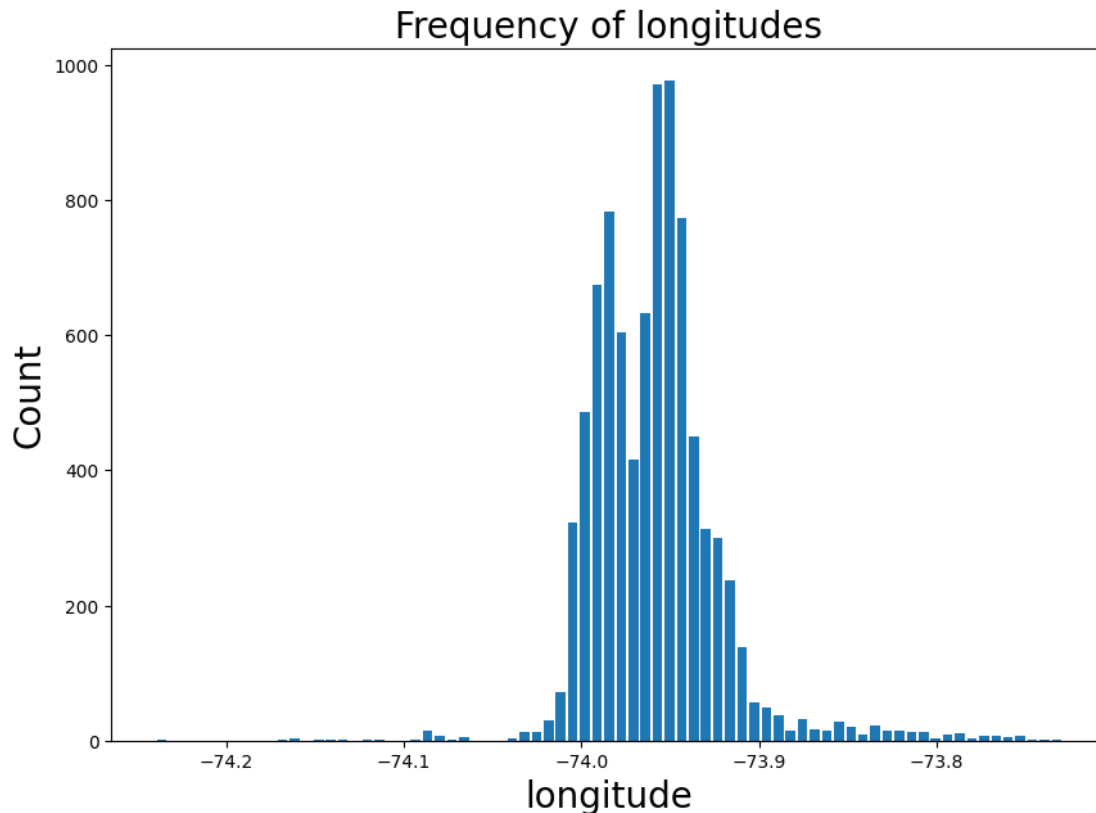
Obtain minimum, maximum and average for latitude

```
print("Minimum latitude:", min(data.latitude))  
print("Maximum latitude:", max(data.latitude))  
print("Average latitude:", np.mean(data.latitude))
```

```
Minimum latitude: 40.50868  
Maximum latitude: 40.90804  
Average latitude: 40.72876700553889
```

Next up is longitude

```
from matplotlib import pyplot as plt  
fig, ax = plt.subplots()  
  
ax.hist(data.longitude, bins=75, rwidth=.8)  
ax.set_xlabel("longitude", fontsize=20)  
ax.set_ylabel("Count", fontsize=20)  
ax.set_title("Frequency of longitudes", fontsize=20)  
fig.set_size_inches(10, 7)  
plt.show()
```

2) Construct the scatter plots of price with each of the numerical features.

```
nums = data.select_dtypes(include=['int64', 'float64'])
display(nums.head(5))
```

```
{
  "rows": [
    {
      "minimum_nights": "1",
      "availability_365": "365",
      "longitude": "-73.97237",
      "number_of_reviews": "9",
      "_deepnote_index_column": "0",
      "latitude": "40.64749",
      "calculated_host_listings_count": "6",
      "host_id": "2787",
      "reviews_per_month": "0.21",
      "price": "149"
    },
    {
      "minimum_nights": "1",
      "availability_365": "355",
      "longitude": "-73.98377",
      "number_of_reviews": "45",
      "_deepnote_index_column": "1",
      "latitude": "40.75362",
      "calculated_host_listings_count": "2",
      "host_id": "2845",
      "reviews_per_month": "0.38",
      "price": "225"
    },
    {
      "minimum_nights": "1",
      "availability_365": "194",
      "longitude": "-73.95976",
      "number_of_reviews": "270",
      "_deepnote_index_column": "2",
      "latitude": "40.68514",
      "calculated_host_listings_count": "1",
      "host_id": "4869",
      "reviews_per_month": "4.64",
      "price": "89"
    },
    {
      "minimum_nights": "10",
      "availability_365": "0",
      "longitude": "-73.94399",
      "number_of_reviews": "9",
      "_deepnote_index_column": "3",
      "latitude": "40.79851",
      "calculated_host_listings_count": "1",
      "host_id": "7192",
      "reviews_per_month": "0.1",
      "price": "80"
    },
    {
      "minimum_nights": "3",
      "availability_365": "129",
      "longitude": "-73.975",
      "number_of_reviews": "74",
      "_deepnote_index_column": "4",
      "latitude": "40.74767",
      "calculated_host_listings_count": "1",
      "host_id": "7322",
      "reviews_per_month": "0.59",
      "price": "200"
    }
  ],
  "row_count": 5,
  "column_count": 9,
  "columns": [
    {
      "name": "host_id",
      "dtype": "int64",
      "stats":

```

```

{"unique_count":5,"histogram":
[{"count":2,"bin_end":3240.5,"bin_start":2787},
{"count":0,"bin_end":3694,"bin_start":3240.5},
{"count":0,"bin_end":4147.5,"bin_start":3694},
{"count":0,"bin_end":4601,"bin_start":4147.5},
{"count":1,"bin_end":5054.5,"bin_start":4601},
{"count":0,"bin_end":5508,"bin_start":5054.5},
{"count":0,"bin_end":5961.5,"bin_start":5508},
{"count":0,"bin_end":6415,"bin_start":5961.5},
{"count":0,"bin_end":6868.5,"bin_start":6415},
{"count":2,"bin_end":7322,"bin_start":6868.5}], "min":"2787","max":"732
2","nan_count":0}}, {"name":"latitude","dtype":"float64","stats":
{"unique_count":5,"histogram":
[{"count":1,"bin_end":40.662592,"bin_start":40.64749},
{"count":0,"bin_end":40.677693999999995,"bin_start":40.662592},
{"count":1,"bin_end":40.692796,"bin_start":40.677693999999995},
{"count":0,"bin_end":40.707898,"bin_start":40.692796},
{"count":0,"bin_end":40.723,"bin_start":40.707898},
{"count":0,"bin_end":40.738102,"bin_start":40.723},
{"count":1,"bin_end":40.753204,"bin_start":40.738102},
{"count":1,"bin_end":40.768306,"bin_start":40.753204},
{"count":0,"bin_end":40.783408,"bin_start":40.768306},
{"count":1,"bin_end":40.79851,"bin_start":40.783408}], "min":"40.64749"
,"max":"40.79851","nan_count":0}},
{"name":"longitude","dtype":"float64","stats":
{"unique_count":5,"histogram":[{"count":1,"bin_end":-
73.979792,"bin_start":-73.98377},{count":0,"bin_end":-
73.975814,"bin_start":-73.979792},{count":2,"bin_end":-
73.971836000000001,"bin_start":-73.975814},{count":0,"bin_end":-
73.967858,"bin_start":-73.971836000000001},{count":0,"bin_end":-
73.96388,"bin_start":-73.967858},{count":0,"bin_end":-
73.959902,"bin_start":-73.96388},{count":1,"bin_end":-
73.955924,"bin_start":-73.959902},{count":0,"bin_end":-
73.951946,"bin_start":-73.955924},{count":0,"bin_end":-
73.947968,"bin_start":-73.951946},{count":1,"bin_end":-
73.94399,"bin_start":-73.947968}], "min":"-73.98377","max":"-
73.94399","nan_count":0}}, {"name":"price","dtype":"int64","stats":
{"unique_count":5,"histogram":
[{"count":2,"bin_end":94.5,"bin_start":80},
{"count":0,"bin_end":109,"bin_start":94.5},
{"count":0,"bin_end":123.5,"bin_start":109},
{"count":0,"bin_end":138,"bin_start":123.5},
{"count":1,"bin_end":152.5,"bin_start":138},
{"count":0,"bin_end":167,"bin_start":152.5},
{"count":0,"bin_end":181.5,"bin_start":167},
{"count":0,"bin_end":196,"bin_start":181.5},
{"count":1,"bin_end":210.5,"bin_start":196},
{"count":1,"bin_end":225,"bin_start":210.5}], "min":"80","max":"225","n
an_count":0}}, {"name":"minimum_nights","dtype":"int64","stats":
{"unique_count":3,"histogram":

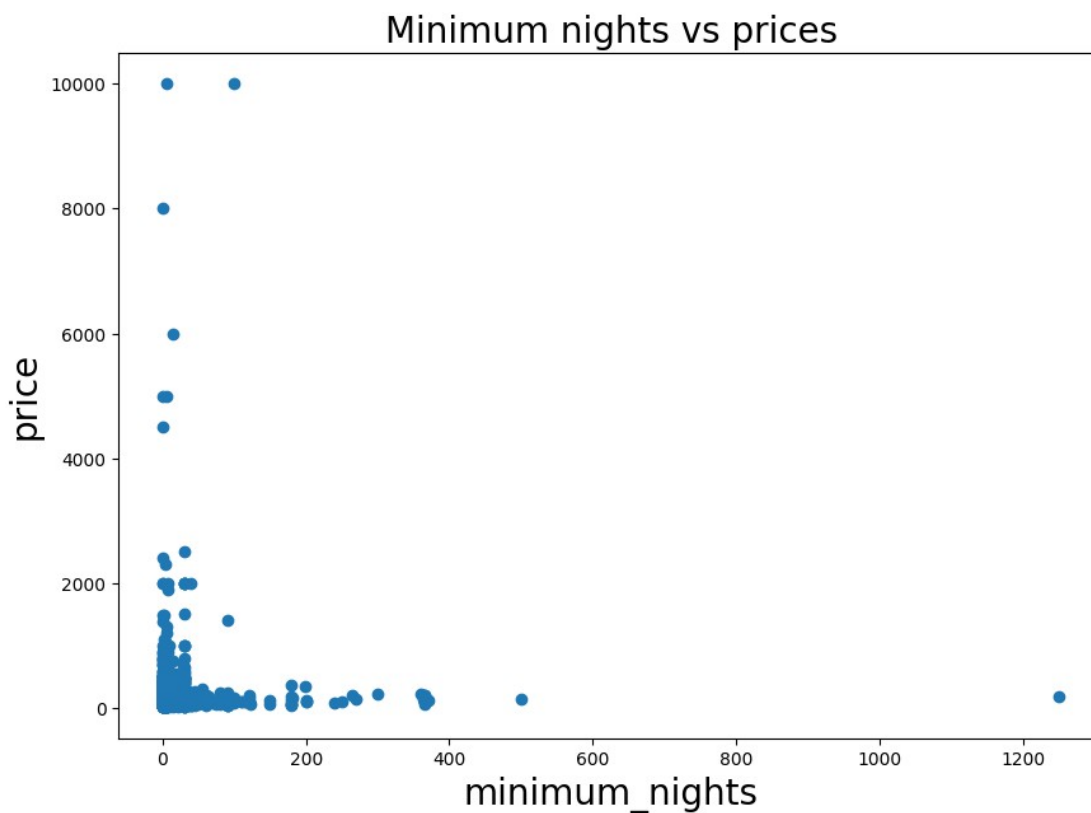
```

```
[{"count":3,"bin_end":1.9,"bin_start":1},
{"count":0,"bin_end":2.8,"bin_start":1.9},
{"count":1,"bin_end":3.7,"bin_start":2.8},
{"count":0,"bin_end":4.6,"bin_start":3.7},
{"count":0,"bin_end":5.5,"bin_start":4.6},
{"count":0,"bin_end":6.4,"bin_start":5.5},
{"count":0,"bin_end":7.3,"bin_start":6.4},
{"count":0,"bin_end":8.2,"bin_start":7.3},
{"count":0,"bin_end":9.1,"bin_start":8.2},
{"count":1,"bin_end":10,"bin_start":9.1}],{"min":"1","max":"10","nan_count":0}},{"name":"number_of_reviews","dtype":"int64","stats":
{"unique_count":4,"histogram":
[{"count":2,"bin_end":35.1,"bin_start":9},
{"count":1,"bin_end":61.2,"bin_start":35.1},
{"count":1,"bin_end":87.30000000000001,"bin_start":61.2},
{"count":0,"bin_end":113.4,"bin_start":87.30000000000001},
{"count":0,"bin_end":139.5,"bin_start":113.4},
{"count":0,"bin_end":165.60000000000002,"bin_start":139.5},
{"count":0,"bin_end":191.70000000000002,"bin_start":165.60000000000002},
{"count":0,"bin_end":217.8,"bin_start":191.70000000000002},
{"count":0,"bin_end":243.9,"bin_start":217.8},
{"count":1,"bin_end":270,"bin_start":243.9}],{"min":"9","max":"270","nan_count":0}},{"name":"reviews_per_month","dtype":"float64","stats":
{"unique_count":5,"histogram":
[{"count":3,"bin_end":0.554,"bin_start":0.1},
{"count":1,"bin_end":1.008,"bin_start":0.554},
{"count":0,"bin_end":1.4620000000000002,"bin_start":1.008},
{"count":0,"bin_end":1.9160000000000001,"bin_start":1.4620000000000002},
{"count":0,"bin_end":2.37,"bin_start":1.9160000000000001},
{"count":0,"bin_end":2.8240000000000003,"bin_start":2.37},
{"count":0,"bin_end":3.278,"bin_start":2.8240000000000003},
{"count":0,"bin_end":3.732,"bin_start":3.278},
{"count":0,"bin_end":4.186,"bin_start":3.732},
{"count":1,"bin_end":4.64,"bin_start":4.186}],{"min":"0.1","max":"4.64",
"nan_count":0}},
{"name":"calculated_host_listings_count","dtype":"int64","stats":
{"unique_count":3,"histogram":
[{"count":3,"bin_end":1.5,"bin_start":1},
{"count":0,"bin_end":2,"bin_start":1.5},
{"count":1,"bin_end":2.5,"bin_start":2},
{"count":0,"bin_end":3,"bin_start":2.5},
{"count":0,"bin_end":3.5,"bin_start":3},
{"count":0,"bin_end":4,"bin_start":3.5},
{"count":0,"bin_end":4.5,"bin_start":4},
{"count":0,"bin_end":5,"bin_start":4.5},
{"count":0,"bin_end":5.5,"bin_start":5},
{"count":1,"bin_end":6,"bin_start":5.5}],{"min":"1","max":"6","nan_count":0}},{"name":"availability_365","dtype":"int64","stats":
{"unique_count":5,"histogram":
[{"count":1,"bin_end":36.5,"bin_start":0},
```

```
{
    "count":0,"bin_end":73,"bin_start":36.5},
    {"count":0,"bin_end":109.5,"bin_start":73},
    {"count":1,"bin_end":146,"bin_start":109.5},
    {"count":0,"bin_end":182.5,"bin_start":146},
    {"count":1,"bin_end":219,"bin_start":182.5},
    {"count":0,"bin_end":255.5,"bin_start":219},
    {"count":0,"bin_end":292,"bin_start":255.5},
    {"count":0,"bin_end":328.5,"bin_start":292},
    {"count":2,"bin_end":365,"bin_start":328.5}], "min": "0", "max": "365", "na_n_count": 0}}, {"name": "_deepnote_index_column", "dtype": "int64"}]]
```

```
fig, ax = plt.subplots()
```

```
ax.plot(nums.minimum_nights,nums.price,'o')
ax.set_xlabel("minimum_nights", fontsize=20)
ax.set_ylabel("price", fontsize=20)
ax.set_title("Minimum nights vs prices", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()
```



Cut outliers using indexes from before. From Professors master tutorial

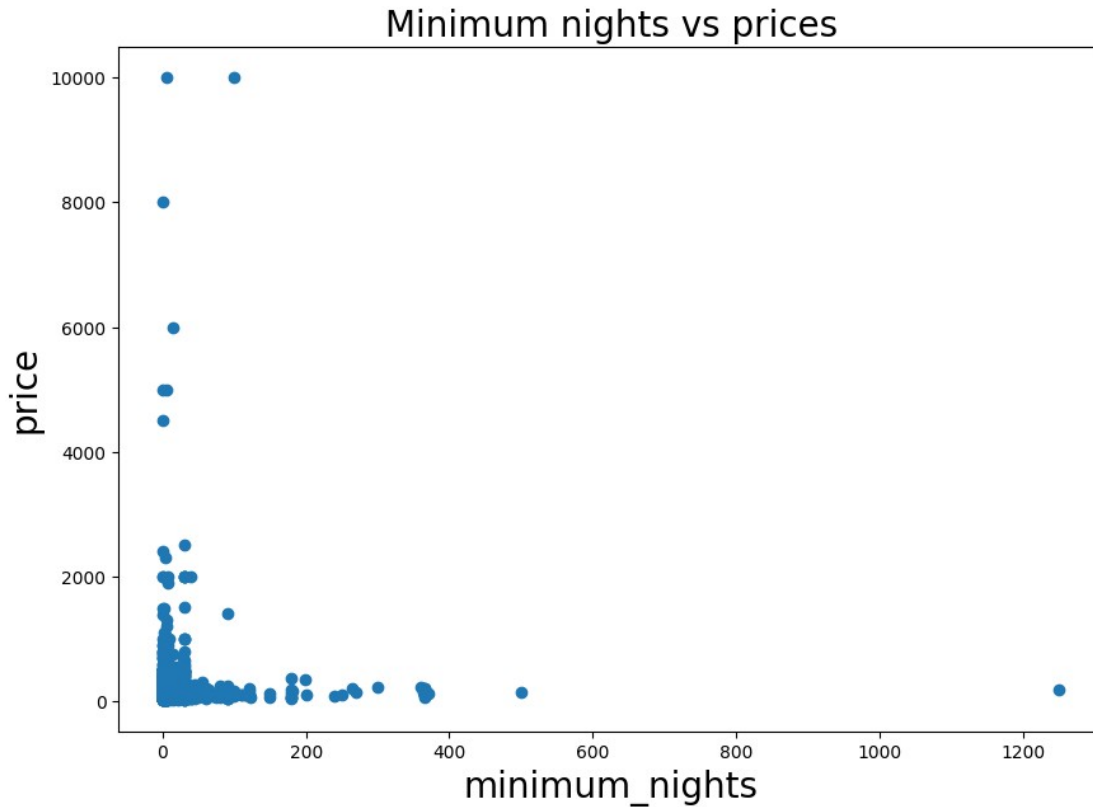
```
fig, ax = plt.subplots()
```

```
ax.plot(nums.minimum_nights[z],nums[z].price,'o')
ax.set_xlabel("minimum_nights", fontsize=20)
```

```

ax.set_ylabel("price", fontsize=20)
ax.set_title("Minimum nights vs prices", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()

```



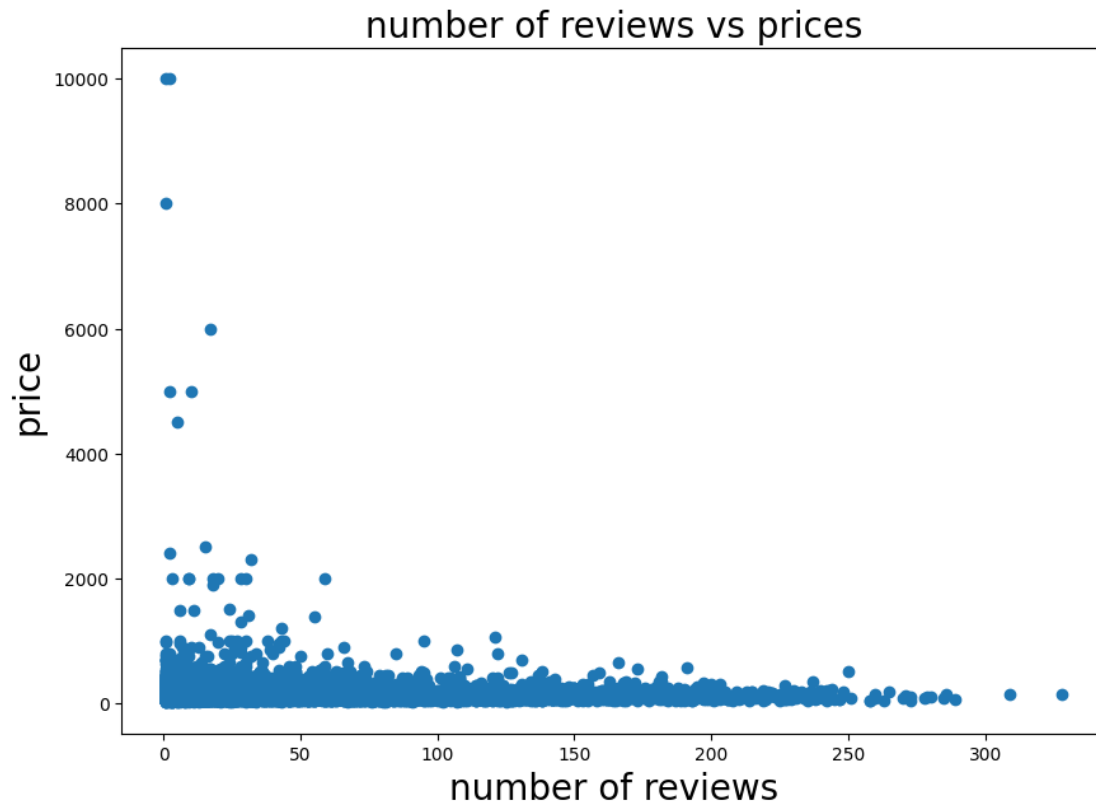
```

numsc = nums[z] ## Here, the c stands for "cleaned"

fig, ax = plt.subplots()

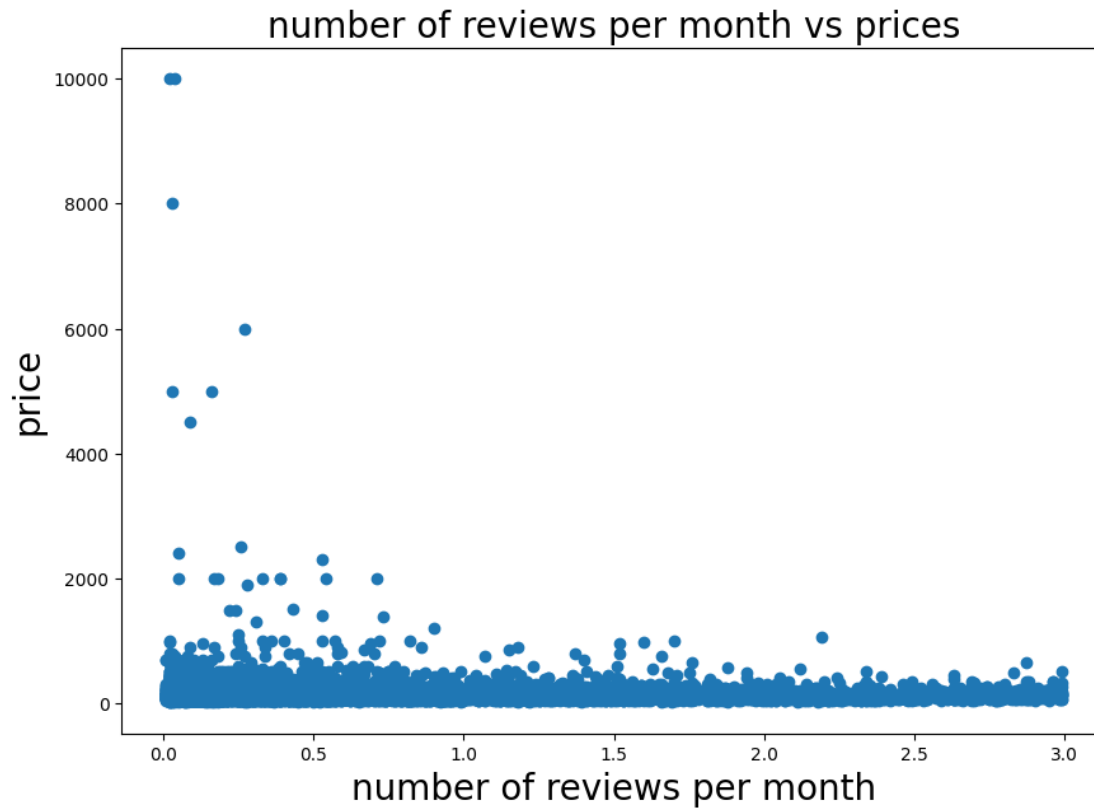
ax.plot(numsc.number_of_reviews, numsc.price, 'o')
ax.set_xlabel("number of reviews", fontsize=20)
ax.set_ylabel("price", fontsize=20)
ax.set_title("number of reviews vs prices", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()

```

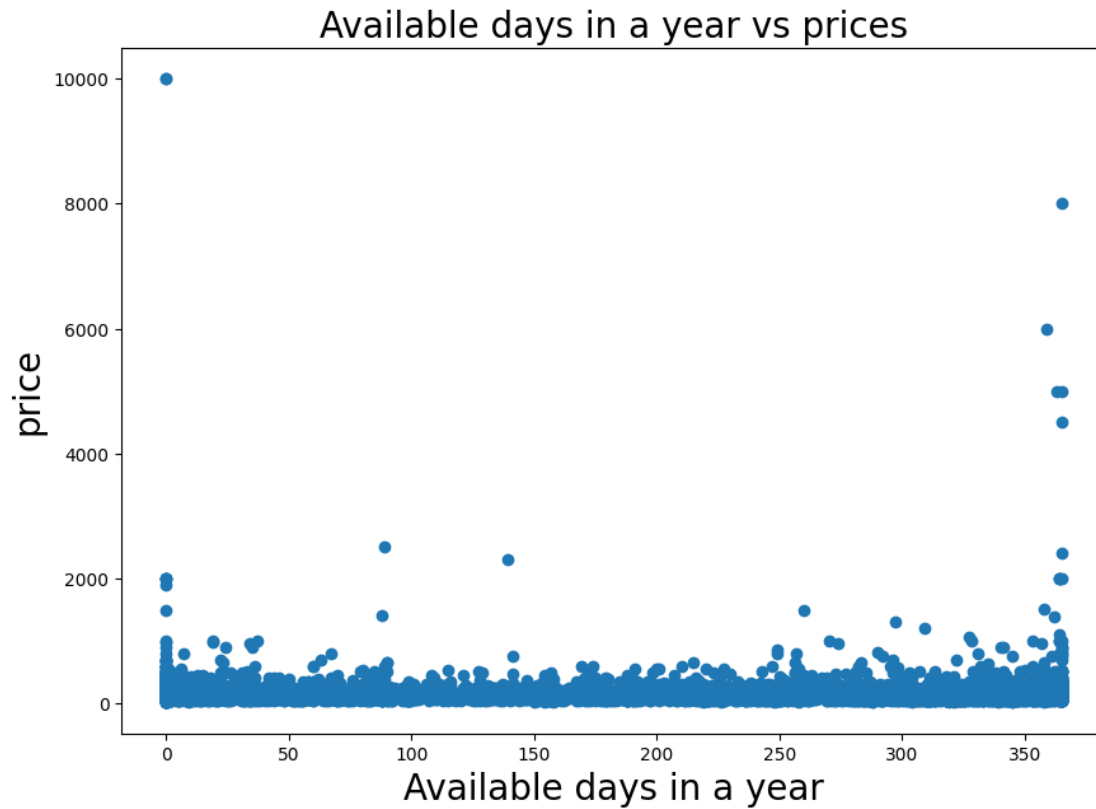
```
fig, ax = plt.subplots()
```

```
ax.plot(numsc.reviews_per_month,numsc.price,'o')
ax.set_xlabel("number of reviews per month", fontsize=20)
ax.set_ylabel("price", fontsize=20)
ax.set_title("number of reviews per month vs prices", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()
```



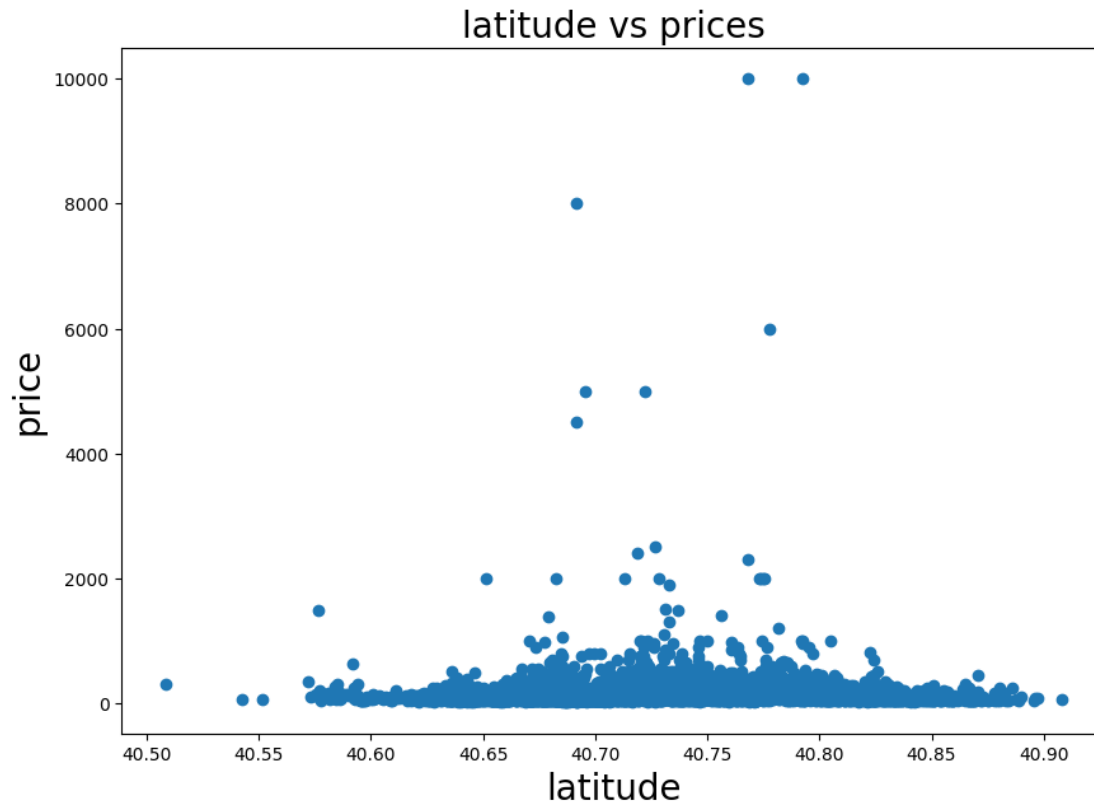
```
fig, ax = plt.subplots()
```

```
ax.plot(numsc.availability_365,numsc.price,'o')
ax.set_xlabel("Available days in a year", fontsize=20)
ax.set_ylabel("price", fontsize=20)
ax.set_title("Available days in a year vs prices", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()
```



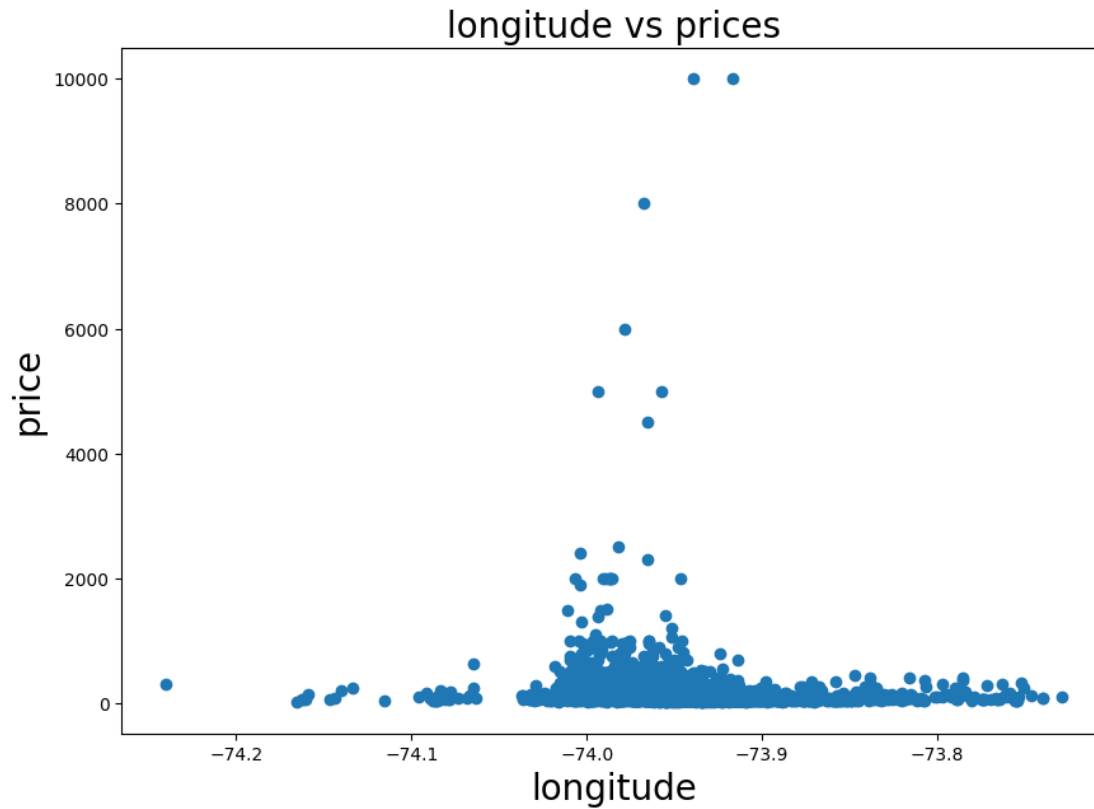
```
fig, ax = plt.subplots()
```

```
ax.plot(numsc.latitude,numsc.price,'o')
ax.set_xlabel("latitude", fontsize=20)
ax.set_ylabel("price", fontsize=20)
ax.set_title("latitude vs prices", fontsize=20)
fig.set_size_inches(10, 7)
plt.show()
```



```
fig, ax = plt.subplots()
```

```
ax.plot(numsc.longitude,numsc.price,'o')  
ax.set_xlabel("longitude", fontsize=20)  
ax.set_ylabel("price", fontsize=20)  
ax.set_title("longitude vs prices", fontsize=20)  
fig.set_size_inches(10, 7)  
plt.show()
```

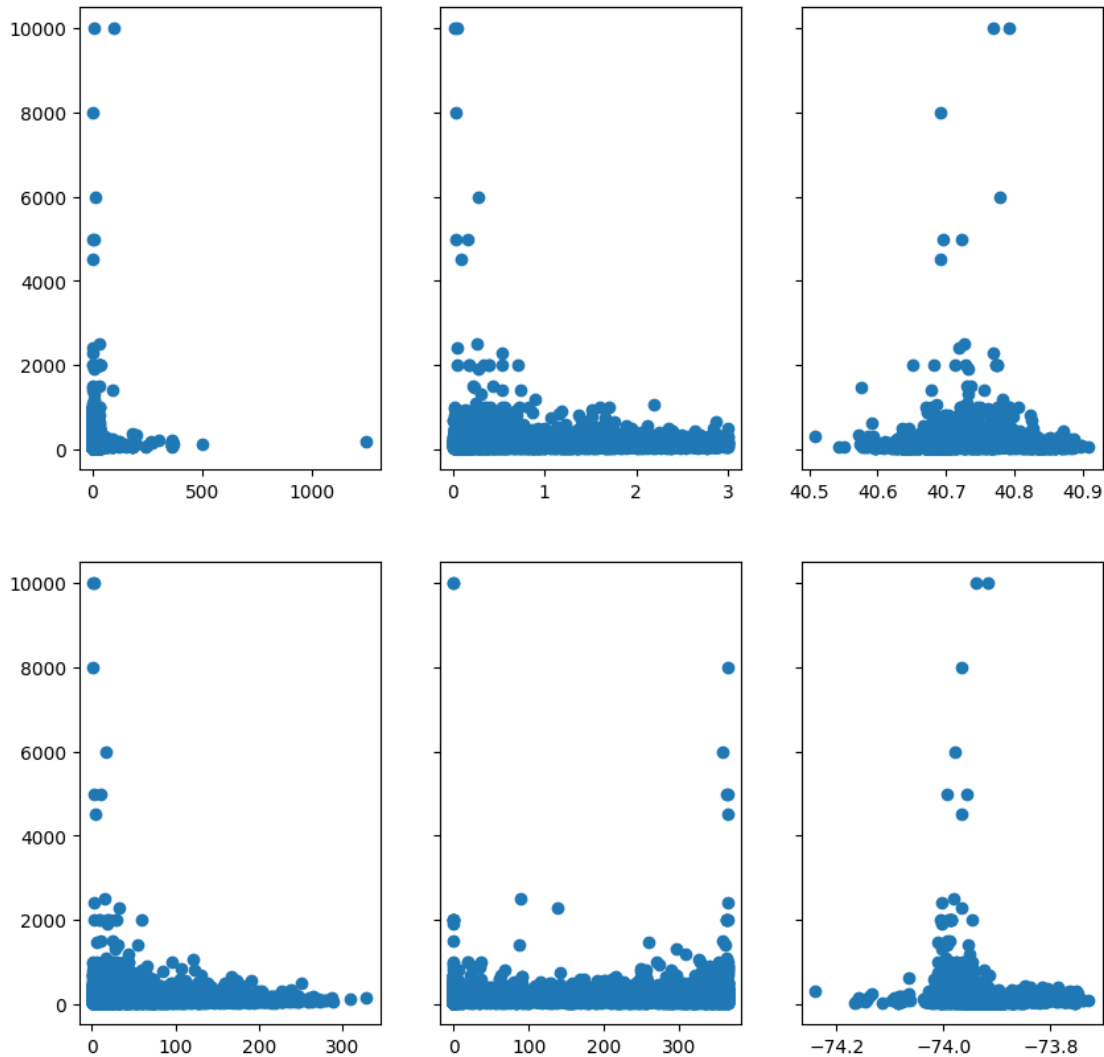


Subplots

```
fig, axes = plt.subplots(2,3, sharey=True)
fig.set_size_inches(10, 10)
```

```
axes[0,0].plot(numsc.minimum_nights,numsc.price,'o')
axes[1,0].plot(numsc.number_of_reviews,numsc.price,'o')
axes[0,1].plot(numsc.reviews_per_month,numsc.price,'o')
axes[1,1].plot(numsc.availability_365,numsc.price,'o')
axes[0,2].plot(numsc.latitude,numsc.price,'o')
axes[1,2].plot(numsc.longitude,numsc.price,'o')
```

```
plt.show()
```



Correlation

```
import pandas as pd
```

```
#url = 'https://raw.githubusercontent.com/tipthederiver/Math-7243-2020/master/Datasets/NYCAirBnB/train.csv'
```

```
data =  
pd.read_csv("https://raw.githubusercontent.com/tipthederiver/Math-7243-2020/master/Datasets/NYCAirBnB/train.csv") # Load NYC Data
```

```
data
```

```
{ "rows":  
[{"last_review": "10/19/2018", "minimum_nights": "1", "availability_365": "365", "longitude": "-73.97237", "number_of_reviews": "9", "_deepnote_index_column": "0", "host_name": "John", "latitude": "40.64749", "room_type": "Private"}]
```

```
room","calculated_host_listings_count":"6","host_id":"2787","name":"Clean & quiet apt home by the park","reviews_per_month":"0.21","neighbourhood_group":"Brooklyn","neighbourhood":"Kensington","price":"149"},
{"last_review":"5/21/2019","minimum_nights":"1","availability_365":"355","longitude":"-73.98377","number_of_reviews":"45","_deepnote_index_column":"1","host_name":"Jennifer","latitude":"40.75362","room_type":"Entire home/apt","calculated_host_listings_count":"2","host_id":"2845","name":"Skylit Midtown Castle","reviews_per_month":"0.38","neighbourhood_group":"Manhattan","neighbourhood":"Midtown","price":"225"},
{"last_review":"nan","minimum_nights":"3","availability_365":"365","longitude":"-73.9419","number_of_reviews":"0","_deepnote_index_column":"2","host_name":"Elisabeth","latitude":"40.80902","room_type":"Private room","calculated_host_listings_count":"1","host_id":"4632","name":"THE VILLAGE OF HARLEM...NEW YORK !","reviews_per_month":"nan","neighbourhood_group":"Manhattan","neighbourhood":"Harlem","price":"150"},
{"last_review":"7/5/2019","minimum_nights":"1","availability_365":"194","longitude":"-73.95976","number_of_reviews":"270","_deepnote_index_column":"3","host_name":"LisaRoxanne","latitude":"40.68514","room_type":"Entire home/apt","calculated_host_listings_count":"1","host_id":"4869","name":"Cozy Entire Floor of Brownstone","reviews_per_month":"4.64","neighbourhood_group":"Brooklyn","neighbourhood":"Clinton Hill","price":"89"},
{"last_review":"11/19/2018","minimum_nights":"10","availability_365":"0","longitude":"-73.94399","number_of_reviews":"9","_deepnote_index_column":"4","host_name":"Laura","latitude":"40.79851","room_type":"Entire home/apt","calculated_host_listings_count":"1","host_id":"7192","name":"Entire Apt: Spacious Studio/Loft by central park","reviews_per_month":"0.1","neighbourhood_group":"Manhattan","neighbourhood":"East Harlem","price":"80"},
{"last_review":"6/22/2019","minimum_nights":"3","availability_365":"129","longitude":"-73.975","number_of_reviews":"74","_deepnote_index_column":"5","host_name":"Chris","latitude":"40.74767","room_type":"Entire home/apt","calculated_host_listings_count":"1","host_id":"7322","name":"Large Cozy 1 BR Apartment In Midtown East","reviews_per_month":"0.59","neighbourhood_group":"Manhattan","neighbourhood":"Murray Hill","price":"200"},
{"last_review":"10/5/2017","minimum_nights":"45","availability_365":"0","longitude":"-73.95596","number_of_reviews":"49","_deepnote_index_column":"6","host_name":"Garon","latitude":"40.68688","room_type":"Private room","calculated_host_listings_count":"1","host_id":"7356","name":"BlissArtsSpace!","reviews_per_month":"0.4","neighbourhood_group":"Brooklyn"}
```

```

yn", "neighbourhood": "Bedford-Stuyvesant", "price": "60"},
{"last_review": "6/24/2019", "minimum_nights": "2", "availability_365": "220", "longitude": "-73.98493", "number_of_reviews": "430", "_deepnote_index_column": "7", "host_name": "Shunichi", "latitude": "40.76489", "room_type": "Private room", "calculated_host_listings_count": "1", "host_id": "8967", "name": "Large Furnished Room Near B'way", "reviews_per_month": "3.47", "neighbourhood_group": "Manhattan", "neighbourhood": "Hell's Kitchen", "price": "79"},
{"last_review": "7/21/2017", "minimum_nights": "2", "availability_365": "0", "longitude": "-73.96723", "number_of_reviews": "118", "_deepnote_index_column": "8", "host_name": "MaryEllen", "latitude": "40.80178", "room_type": "Private room", "calculated_host_listings_count": "1", "host_id": "7490", "name": "Cozy Clean Guest Room - Family Apt", "reviews_per_month": "0.99", "neighbourhood_group": "Manhattan", "neighbourhood": "Upper West Side", "price": "79"},
{"last_review": "6/9/2019", "minimum_nights": "1", "availability_365": "188", "longitude": "-73.99037", "number_of_reviews": "160", "_deepnote_index_column": "9", "host_name": "Ben", "latitude": "40.71344", "room_type": "Entire home/apt", "calculated_host_listings_count": "4", "host_id": "7549", "name": "Cute & Cozy Lower East Side 1 bdrm", "reviews_per_month": "1.33", "neighbourhood_group": "Manhattan", "neighbourhood": "Chinatown", "price": "150"}], "row_count": 9999, "column_count": 15, "columns": [{"name": "name", "dtype": "object", "stats": {"unique_count": 9928, "categories": [{"name": "Loft Suite @ The Box House Hotel", "count": 11}, {"name": "9927 others", "count": 9980}, {"name": "Missing", "count": 8}], "nan_count": 8}}, {"name": "host_id", "dtype": "int64", "stats": {"unique_count": 8628, "histogram": [{"count": 3810, "bin_end": 4142121.7, "bin_start": 2571}, {"count": 1689, "bin_end": 8281672.4, "bin_start": 4142121.7}, {"count": 1133, "bin_end": 1.24212231000000001e7, "bin_start": 8281672.4}, {"count": 709, "bin_end": 1.65607738e7, "bin_start": 1.24212231000000001e7}, {"count": 608, "bin_end": 2.07003245e7, "bin_start": 1.65607738e7}, {"count": 727, "bin_end": 2.48398752000000003e7, "bin_start": 2.07003245e7}, {"count": 407, "bin_end": 2.89794259000000002e7, "bin_start": 2.48398752000000003e7}, {"count": 320, "bin_end": 3.31189766e7, "bin_start": 2.89794259000000002e7}, {"count": 381, "bin_end": 3.72585273000000004e7, "bin_start": 3.31189766e7}, {"count": 215, "bin_end": 41398078, "bin_start": 3.72585273000000004e7}], "min": "2571", "max": "41398078", "nan_count": 0}}, {"name": "host_name", "dtype": "object", "stats": {"unique_count": 3714, "categories": [{"name": "Michael", "count": 98}, {"name": "3713 others", "count": 9891}, {"name": "Missing", "count": 10}], "nan_count": 10}}, {"name": "neighbourhood_group", "dtype": "object", "stats": {"unique_count": 5, "categories": [{"name": "Manhattan", "count": 4583}, {"name": "Brooklyn", "count": 4488}, {"name": "3

```



```

others", "count":928}], "nan_count":0}},
{"name": "neighbourhood", "dtype": "object", "stats":
{"unique_count":171, "categories": [{"name": "Williamsburg", "count":944},
{"name": "Bedford-Stuyvesant", "count":742}, {"name": "169
others", "count":8313}], "nan_count":0}},
{"name": "latitude", "dtype": "float64", "stats":
{"unique_count":7579, "histogram":
[{"count":3, "bin_end":40.548615999999996, "bin_start":40.50868},
{"count":30, "bin_end":40.588552, "bin_start":40.548615999999996},
{"count":106, "bin_end":40.628488, "bin_start":40.588552},
{"count":748, "bin_end":40.668424, "bin_start":40.628488},
{"count":2743, "bin_end":40.70836, "bin_start":40.668424},
{"count":3220, "bin_end":40.748295999999996, "bin_start":40.70836},
{"count":1628, "bin_end":40.788232, "bin_start":40.748295999999996},
{"count":1136, "bin_end":40.828168, "bin_start":40.788232},
{"count":338, "bin_end":40.868104, "bin_start":40.828168},
{"count":47, "bin_end":40.90804, "bin_start":40.868104}], "min": "40.50868
", "max": "40.90804", "nan_count":0}},
{"name": "longitude", "dtype": "float64", "stats":
{"unique_count":6351, "histogram": [{"count":2, "bin_end":-
74.18877499999999, "bin_start":-74.23986}, {"count":8, "bin_end":-
74.13768999999999, "bin_start":-74.18877499999999},
{"count":12, "bin_end":-74.08660499999999, "bin_start":-
74.13768999999999}, {"count":32, "bin_end":-
74.03551999999999, "bin_start":-74.08660499999999},
{"count":2324, "bin_end":-73.98443499999999, "bin_start":-
74.03551999999999}, {"count":6031, "bin_end":-73.93335, "bin_start":-
73.98443499999999}, {"count":1310, "bin_end":-73.882265, "bin_start":-
73.93335}, {"count":159, "bin_end":-73.83118, "bin_start":-73.882265},
{"count":82, "bin_end":-73.780095, "bin_start":-73.83118},
{"count":39, "bin_end":-73.72901, "bin_start":-73.780095}], "min": "-
74.23986", "max": "-73.72901", "nan_count":0}},
{"name": "room_type", "dtype": "object", "stats":
{"unique_count":3, "categories": [{"name": "Entire
home/apt", "count":5811}, {"name": "Private room", "count":4054},
{"name": "Shared room", "count":134}], "nan_count":0}},
{"name": "price", "dtype": "int64", "stats":
{"unique_count":389, "histogram":
[{"count":9939, "bin_end":1009, "bin_start":10},
{"count":41, "bin_end":2008, "bin_start":1009},
{"count":7, "bin_end":3007, "bin_start":2008},
{"count":3, "bin_end":4006, "bin_start":3007},
{"count":3, "bin_end":5005, "bin_start":4006},
{"count":2, "bin_end":6004, "bin_start":5005},
{"count":1, "bin_end":7003, "bin_start":6004},
{"count":1, "bin_end":8002, "bin_start":7003},
{"count":0, "bin_end":9001, "bin_start":8002},
{"count":2, "bin_end":10000, "bin_start":9001}], "min": "10", "max": "10000"
, "nan_count":0}}, {"name": "minimum_nights", "dtype": "int64", "stats":
{"unique_count":75, "histogram":

```

```

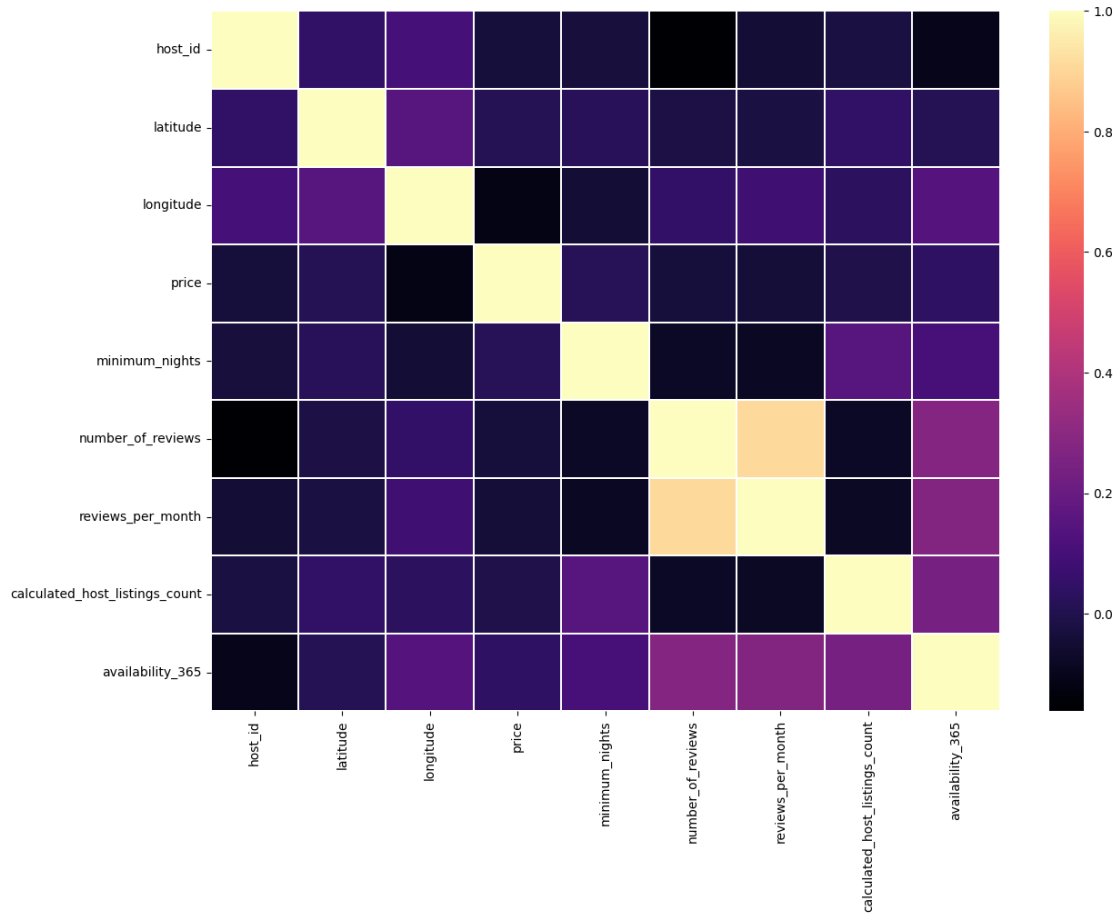
[{"count":9953,"bin_end":125.9,"bin_start":1},
{"count":22,"bin_end":250.8,"bin_start":125.9},
{"count":20,"bin_end":375.70000000000005,"bin_start":250.8},
{"count":2,"bin_end":500.6,"bin_start":375.70000000000005},
{"count":0,"bin_end":625.5,"bin_start":500.6},
{"count":0,"bin_end":750.4000000000001,"bin_start":625.5},
{"count":0,"bin_end":875.3000000000001,"bin_start":750.4000000000001},
{"count":1,"bin_end":1000.2,"bin_start":875.3000000000001},
{"count":0,"bin_end":1125.1000000000001,"bin_start":1000.2},
{"count":1,"bin_end":1250,"bin_start":1125.1000000000001}], "min": "1", "
max": "1250", "nan_count": 0}},
{"name": "number_of_reviews", "dtype": "int64"},
{"name": "last_review", "dtype": "object"},
{"name": "reviews_per_month", "dtype": "float64"},
{"name": "calculated_host_listings_count", "dtype": "int64"},
{"name": "availability_365", "dtype": "int64"},
{"name": "_deepnote_index_column", "dtype": "int64"}]]

nums = data.select_dtypes(include=['int64','float64']) # Get only
numerical features

import seaborn as sns
fig,ax = plt.subplots(figsize=(14, 10))

#sns.set(font_scale=1.4)
sns.heatmap(numsc.corr(), ax=ax, linewidths=0.05,cmap="magma")
plt.show()

```



```
numsc.corr()
```

```
{ "rows":
[{"minimum_nights": "0.01566496805026751", "availability_365": "0.03904043742614429", "longitude": "-0.11691794344005815", "number_of_reviews": "-0.03663568369819261", "_deepnote_index_column": "price", "latitude": "0.012653145941728405", "calculated_host_listings_count": "-0.006132355022074966", "host_id": "-0.0370096731639517", "reviews_per_month": "-0.04135236176228002", "price": "1.0"},
{"minimum_nights": "0.10285495011238892", "availability_365": "1.0", "longitude": "0.1388523846534466", "number_of_reviews": "0.2818310412566302", "_deepnote_index_column": "availability_365", "latitude": "0.013881710792440662", "calculated_host_listings_count": "0.2386015717846589", "host_id": "-0.10647226902382971", "reviews_per_month": "0.27186704430125064", "price": "0.03904043742614429"},
{"minimum_nights": "1.0", "availability_365": "0.10285495011238892", "longitude": "-0.04462986667486845", "number_of_reviews": "-0.07630091378766615", "_deepnote_index_column": "minimum_nights", "latitude": "0.02209412164496461", "calculated_host_listings_count": "0.15025381792394402", "host_id": "-0.034618322419276565", "reviews_per_month": "-0.08873451245860663", "price": "0.01566496805026751"}],
```

```

{"minimum_nights":"0.02209412164496461","availability_365":"0.01388171
0792440662","longitude":"0.14807722580577928","number_of_reviews":"-
0.020837020129735786","_deepnote_index_column":"latitude","latitude":"
1.0","calculated_host_listings_count":"0.04429440029532216","host_id":
"0.045236854576407845","reviews_per_month":"-
0.024934175986113496","price":"0.012653145941728405"},
{"minimum_nights":"0.15025381792394402","availability_365":"0.23860157
17846589","longitude":"0.03173167265215009","number_of_reviews":"-
0.07820295209442936","_deepnote_index_column":"calculated_host_listing
s_count","latitude":"0.04429440029532216","calculated_host_listings_co
unt":"1.0","host_id":"-0.023314758667583663","reviews_per_month":"-
0.08419977685366752","price":"-0.006132355022074966"},
{"minimum_nights":"-
0.07630091378766615","availability_365":"0.2818310412566302","longitud
e":"0.04786103200360704","number_of_reviews":"1.0","_deepnote_index_co
lumn":"number_of_reviews","latitude":"-
0.020837020129735786","calculated_host_listings_count":"-
0.07820295209442936","host_id":"-
0.16197418164647162","reviews_per_month":"0.9128424039650588","price":
"-0.03663568369819261"},{"minimum_nights":"-
0.034618322419276565","availability_365":"-
0.10647226902382971","longitude":"0.10012834418413424","number_of_revi
ews":"-
0.16197418164647162","_deepnote_index_column":"host_id","latitude":"0.
045236854576407845","calculated_host_listings_count":"-
0.023314758667583663","host_id":"1.0","reviews_per_month":"-
0.044715947673164086","price":"-0.0370096731639517"},
{"minimum_nights":"-
0.08873451245860663","availability_365":"0.27186704430125064","longitu
de":"0.07867492824876011","number_of_reviews":"0.9128424039650588","_d
eepnote_index_column":"reviews_per_month","latitude":"-
0.024934175986113496","calculated_host_listings_count":"-
0.08419977685366752","host_id":"-
0.044715947673164086","reviews_per_month":"1.0","price":"-
0.04135236176228002"},{"minimum_nights":"-
0.04462986667486845","availability_365":"0.1388523846534466","longitud
e":"1.0","number_of_reviews":"0.04786103200360704","_deepnote_index_co
lumn":"longitude","latitude":"0.14807722580577928","calculated_host_li
stings_count":"0.03173167265215009","host_id":"0.10012834418413424","r
eviews_per_month":"0.07867492824876011","price":"-
0.11691794344005815"}], "row_count":9, "column_count":9, "columns":
[{"name":"host_id","dtype":"float64","stats":
{"unique_count":9,"histogram":[{"count":2,"bin_end":-
4.577676348182445e-2,"bin_start":-0.16197418164647162},
{"count":5,"bin_end":7.042065468282271e-2,"bin_start":-
4.577676348182445e-2},
{"count":1,"bin_end":0.18661807284746987,"bin_start":7.042065468282271
e-2},
{"count":0,"bin_end":0.302815491012117,"bin_start":0.18661807284746987
}],

```

```
{"count":0,"bin_end":0.41901290917676426,"bin_start":0.302815491012117},
{"count":0,"bin_end":0.5352103273414114,"bin_start":0.41901290917676426},
{"count":0,"bin_end":0.6514077455060586,"bin_start":0.5352103273414114},
{"count":0,"bin_end":0.7676051636707057,"bin_start":0.6514077455060586},
{"count":0,"bin_end":0.8838025818353529,"bin_start":0.7676051636707057},
{"count":1,"bin_end":1,"bin_start":0.8838025818353529}], "min": "-0.16197418164647162", "max": "1.0", "nan_count":0}},
{"name":"latitude","dtype":"float64","stats":
{"unique_count":9,"histogram":
[{"count":7,"bin_end":7.755924161249785e-2,"bin_start":-2.4934175986113496e-2},
{"count":1,"bin_end":0.1800526592111092,"bin_start":7.755924161249785e-2},
{"count":0,"bin_end":0.28254607680972055,"bin_start":0.1800526592111092},
{"count":0,"bin_end":0.3850394944083319,"bin_start":0.28254607680972055},
{"count":0,"bin_end":0.48753291200694326,"bin_start":0.3850394944083319},
{"count":0,"bin_end":0.5900263296055546,"bin_start":0.48753291200694326},
{"count":0,"bin_end":0.692519747204166,"bin_start":0.5900263296055546},
{"count":0,"bin_end":0.7950131648027773,"bin_start":0.692519747204166},
{"count":0,"bin_end":0.8975065824013886,"bin_start":0.7950131648027773},
{"count":1,"bin_end":1,"bin_start":0.8975065824013886}], "min": "-0.024934175986113496", "max": "1.0", "nan_count":0}},
{"name":"longitude","dtype":"float64","stats":
{"unique_count":9,"histogram":[{"count":2,"bin_end":-5.226149096052332e-3,"bin_start":-0.11691794344005815},
{"count":4,"bin_end":0.10646564524795349,"bin_start":-5.226149096052332e-3},
{"count":2,"bin_end":0.2181574395919593,"bin_start":0.10646564524795349},
{"count":0,"bin_end":0.3298492339359651,"bin_start":0.2181574395919593},
{"count":0,"bin_end":0.441541028279971,"bin_start":0.3298492339359651},
{"count":0,"bin_end":0.5532328226239768,"bin_start":0.441541028279971},
{"count":0,"bin_end":0.6649246169679827,"bin_start":0.5532328226239768},
{"count":0,"bin_end":0.7766164113119884,"bin_start":0.6649246169679827},
{"count":0,"bin_end":0.8883082056559942,"bin_start":0.7766164113119884}
```

```
}, {"count":1,"bin_end":1,"bin_start":0.8883082056559942}], "min": "-  
0.11691794344005815", "max": "1.0", "nan_count":0}},  
{"name": "price", "dtype": "float64", "stats":  
{"unique_count":9, "histogram": [{"count":5, "bin_end": -  
5.226149096052332e-3, "bin_start": -0.11691794344005815},  
{"count":3, "bin_end": 0.10646564524795349, "bin_start": -  
5.226149096052332e-3},  
{"count":0, "bin_end": 0.2181574395919593, "bin_start": 0.1064656452479534  
9},  
{"count":0, "bin_end": 0.3298492339359651, "bin_start": 0.2181574395919593  
},  
{"count":0, "bin_end": 0.441541028279971, "bin_start": 0.3298492339359651}  
},  
{"count":0, "bin_end": 0.5532328226239768, "bin_start": 0.441541028279971}  
},  
{"count":0, "bin_end": 0.6649246169679827, "bin_start": 0.5532328226239768  
},  
{"count":0, "bin_end": 0.7766164113119884, "bin_start": 0.6649246169679827  
},  
{"count":0, "bin_end": 0.8883082056559942, "bin_start": 0.7766164113119884  
}, {"count":1,"bin_end":1,"bin_start":0.8883082056559942}], "min": "-  
0.11691794344005815", "max": "1.0", "nan_count":0}},  
{"name": "minimum_nights", "dtype": "float64", "stats":  
{"unique_count":9, "histogram":  
[{"count":5, "bin_end": 2.013893878725402e-2, "bin_start": -  
8.873451245860663e-2},  
{"count":2, "bin_end": 0.12901239003311465, "bin_start": 2.013893878725402  
e-2},  
{"count":1, "bin_end": 0.23788584127897527, "bin_start": 0.129012390033114  
65},  
{"count":0, "bin_end": 0.34675929252483595, "bin_start": 0.237885841278975  
27},  
{"count":0, "bin_end": 0.4556327437706966, "bin_start": 0.3467592925248359  
5},  
{"count":0, "bin_end": 0.5645061950165572, "bin_start": 0.4556327437706966  
},  
{"count":0, "bin_end": 0.6733796462624179, "bin_start": 0.5645061950165572  
},  
{"count":0, "bin_end": 0.7822530975082785, "bin_start": 0.6733796462624179  
},  
{"count":0, "bin_end": 0.8911265487541392, "bin_start": 0.7822530975082785  
}, {"count":1,"bin_end":1,"bin_start":0.8911265487541392}], "min": "-  
0.08873451245860663", "max": "1.0", "nan_count":0}},  
{"name": "number_of_reviews", "dtype": "float64", "stats":  
{"unique_count":9, "histogram": [{"count":3, "bin_end": -  
4.577676348182445e-2, "bin_start": -0.16197418164647162},  
{"count":3, "bin_end": 7.042065468282271e-2, "bin_start": -  
4.577676348182445e-2},  
{"count":0, "bin_end": 0.18661807284746987, "bin_start": 7.042065468282271  
e-2},
```



```
{"count":1,"bin_end":0.302815491012117,"bin_start":0.18661807284746987
},
{"count":0,"bin_end":0.41901290917676426,"bin_start":0.302815491012117
},
{"count":0,"bin_end":0.5352103273414114,"bin_start":0.4190129091767642
6},
{"count":0,"bin_end":0.6514077455060586,"bin_start":0.5352103273414114
},
{"count":0,"bin_end":0.7676051636707057,"bin_start":0.6514077455060586
},
{"count":0,"bin_end":0.8838025818353529,"bin_start":0.7676051636707057
},{ "count":2,"bin_end":1,"bin_start":0.8838025818353529}], "min": "-
0.16197418164647162", "max": "1.0", "nan_count":0}},
{"name":"reviews_per_month","dtype":"float64","stats":
{"unique_count":9,"histogram":
[{"count":5,"bin_end":2.013893878725402e-2,"bin_start":-
8.873451245860663e-2},
{"count":1,"bin_end":0.12901239003311465,"bin_start":2.013893878725402
e-2},
{"count":0,"bin_end":0.23788584127897527,"bin_start":0.129012390033114
65},
{"count":1,"bin_end":0.34675929252483595,"bin_start":0.237885841278975
27},
{"count":0,"bin_end":0.4556327437706966,"bin_start":0.3467592925248359
5},
{"count":0,"bin_end":0.5645061950165572,"bin_start":0.4556327437706966
},
{"count":0,"bin_end":0.6733796462624179,"bin_start":0.5645061950165572
},
{"count":0,"bin_end":0.7822530975082785,"bin_start":0.6733796462624179
},
{"count":0,"bin_end":0.8911265487541392,"bin_start":0.7822530975082785
},{ "count":2,"bin_end":1,"bin_start":0.8911265487541392}], "min": "-
0.08873451245860663", "max": "1.0", "nan_count":0}},
{"name":"calculated_host_listings_count","dtype":"float64","stats":
{"unique_count":9,"histogram":
[{"count":4,"bin_end":2.4220200831699235e-2,"bin_start":-
8.419977685366752e-2},
{"count":2,"bin_end":0.132640178517066,"bin_start":2.4220200831699235e
-2},
{"count":2,"bin_end":0.2410601562024327,"bin_start":0.132640178517066}
,
{"count":0,"bin_end":0.34948013388779947,"bin_start":0.241060156202432
7},
{"count":0,"bin_end":0.45790011157316624,"bin_start":0.349480133887799
47},
{"count":0,"bin_end":0.5663200892585329,"bin_start":0.4579001115731662
4},
{"count":0,"bin_end":0.6747400669438998,"bin_start":0.5663200892585329
},
,
```

```
{
  "count": 0, "bin_end": 0.7831600446292665, "bin_start": 0.6747400669438998
},
{
  "count": 0, "bin_end": 0.8915800223146332, "bin_start": 0.7831600446292665
},
{
  "count": 1, "bin_end": 1, "bin_start": 0.8915800223146332
}], "min": "-0.08419977685366752", "max": "1.0", "nan_count": 0}},
{
  "name": "availability_365", "dtype": "float64", "stats": {
    "unique_count": 9, "histogram": [
      {
        "count": 1, "bin_end": 4.174957878553265e-3, "bin_start": -0.10647226902382971
      },
      {
        "count": 3, "bin_end": 0.11482218478093624, "bin_start": 4.174957878553265e-3
      },
      {
        "count": 1, "bin_end": 0.2254694116833192, "bin_start": 0.11482218478093624
      },
      {
        "count": 3, "bin_end": 0.3361166385857022, "bin_start": 0.2254694116833192
      },
      {
        "count": 0, "bin_end": 0.44676386548808517, "bin_start": 0.3361166385857022
      },
      {
        "count": 0, "bin_end": 0.5574110923904682, "bin_start": 0.44676386548808517
      },
      {
        "count": 0, "bin_end": 0.6680583192928512, "bin_start": 0.5574110923904682
      },
      {
        "count": 0, "bin_end": 0.7787055461952341, "bin_start": 0.6680583192928512
      },
      {
        "count": 0, "bin_end": 0.8893527730976171, "bin_start": 0.7787055461952341
      },
      {
        "count": 1, "bin_end": 1, "bin_start": 0.8893527730976171
      }
    ]
  }, "min": "-0.10647226902382971", "max": "1.0", "nan_count": 0}},
{
  "name": "_deepnote_index_column", "dtype": "object"
}]}
```

Reviews per month and longitude appears to have a negative correlation with price, While availability per year seems to have a positive correlation with price

```
new = numsc.drop(["host_id", "calculated_host_listings_count"],
axis=1)
new
```

```
{
  "rows": [
    {
      "minimum_nights": "1", "availability_365": "365", "longitude": "-73.97237", "number_of_reviews": "9", "_deepnote_index_column": "0", "latitude": "40.64749", "reviews_per_month": "0.21", "price": "149"
    },
    {
      "minimum_nights": "1", "availability_365": "355", "longitude": "-73.98377", "number_of_reviews": "45", "_deepnote_index_column": "1", "latitude": "40.75362", "reviews_per_month": "0.38", "price": "225"
    },
    {
      "minimum_nights": "10", "availability_365": "0", "longitude": "-73.94399", "number_of_reviews": "9", "_deepnote_index_column": "3", "latitude": "40.79851", "reviews_per_month": "0.1", "price": "80"
    },
    {
      "minimum_nights": "3", "availability_365": "129", "longitude": "-73.975", "number_of_reviews": "74", "_deepnote_index_column": "4", "latitude": "40.74767", "reviews_per_month": "0.59", "price": "200"
    },
    {
      "minimum_nights": "45", "availability_365": "0", "longitude": "-73.95596", "number_of_reviews": "49", "_deepnote_index_column": "5", "latitude": "40.68688", "reviews_per_month": "0.4", "price": "60"
    },
    {
      "minimum_nights": "2", "availability_365": "0", "longitude": "-73.97237", "number_of_reviews": "9", "_deepnote_index_column": "0", "latitude": "40.64749", "reviews_per_month": "0.21", "price": "149"
    }
  ]
}
```



```

73.96723","number_of_reviews":"118","_deepnote_index_column":"7","lati
tude":"40.80178","reviews_per_month":"0.99","price":"79"},
{"minimum_nights":"1","availability_365":"188","longitude":"-
73.99037","number_of_reviews":"160","_deepnote_index_column":"8","lati
tude":"40.71344","reviews_per_month":"1.33","price":"150"},
{"minimum_nights":"5","availability_365":"6","longitude":"-
73.96545","number_of_reviews":"53","_deepnote_index_column":"9","latit
ude":"40.80316","reviews_per_month":"0.43","price":"135"},
{"minimum_nights":"2","availability_365":"39","longitude":"-
73.98867","number_of_reviews":"188","_deepnote_index_column":"10","lat
itude":"40.76076","reviews_per_month":"1.5","price":"85"},
{"minimum_nights":"4","availability_365":"314","longitude":"-
73.98779","number_of_reviews":"167","_deepnote_index_column":"11","lat
itude":"40.66829","reviews_per_month":"1.34","price":"89"}], "row_count
":8075,"column_count":7,"columns":
[{"name":"latitude","dtype":"float64","stats":
{"unique_count":6431,"histogram":
[{"count":2,"bin_end":40.548615999999996,"bin_start":40.50868},
{"count":27,"bin_end":40.588552,"bin_start":40.548615999999996},
{"count":86,"bin_end":40.628488,"bin_start":40.588552},
{"count":611,"bin_end":40.668424,"bin_start":40.628488},
{"count":2213,"bin_end":40.70836,"bin_start":40.668424},
{"count":2631,"bin_end":40.748295999999996,"bin_start":40.70836},
{"count":1270,"bin_end":40.788232,"bin_start":40.748295999999996},
{"count":922,"bin_end":40.828168,"bin_start":40.788232},
{"count":270,"bin_end":40.868104,"bin_start":40.828168},
{"count":43,"bin_end":40.90804,"bin_start":40.868104}], "min":"40.50868
","max":"40.90804","nan_count":0}},
{"name":"longitude","dtype":"float64","stats":
{"unique_count":5535,"histogram":[{"count":1,"bin_end":-
74.188774999999999,"bin_start":-74.23986},{ "count":8,"bin_end":-
74.137689999999999,"bin_start":-74.188774999999999},
{"count":8,"bin_end":-74.086604999999999,"bin_start":-
74.137689999999999},{ "count":24,"bin_end":-
74.035519999999999,"bin_start":-74.086604999999999},
{"count":1874,"bin_end":-73.984434999999999,"bin_start":-
74.035519999999999},{ "count":4885,"bin_end":-73.93335,"bin_start":-
73.984434999999999},{ "count":1050,"bin_end":-73.882265,"bin_start":-
73.93335},{ "count":134,"bin_end":-73.83118,"bin_start":-73.882265},
{"count":60,"bin_end":-73.780095,"bin_start":-73.83118},
{"count":31,"bin_end":-73.72901,"bin_start":-73.780095}], "min":"-
74.23986","max":"-73.72901","nan_count":0}},
{"name":"price","dtype":"int64","stats":
{"unique_count":361,"histogram":
[{"count":8047,"bin_end":1014.4,"bin_start":16},
{"count":18,"bin_end":2012.8,"bin_start":1014.4},
{"count":3,"bin_end":3011.2,"bin_start":2012.8},
{"count":0,"bin_end":4009.6,"bin_start":3011.2},
{"count":3,"bin_end":5008,"bin_start":4009.6},
{"count":1,"bin_end":6006.4,"bin_start":5008},

```

```
{"count":0,"bin_end":7004.8,"bin_start":6006.4},
{"count":1,"bin_end":8003.2,"bin_start":7004.8},
{"count":0,"bin_end":9001.6,"bin_start":8003.2},
{"count":2,"bin_end":10000,"bin_start":9001.6}],"min":"16","max":"1000
0","nan_count":0}},{"name":"minimum_nights","dtype":"int64","stats":
{"unique_count":70,"histogram":
[{"count":8044,"bin_end":125.9,"bin_start":1},
{"count":14,"bin_end":250.8,"bin_start":125.9},
{"count":15,"bin_end":375.70000000000005,"bin_start":250.8},
{"count":1,"bin_end":500.6,"bin_start":375.70000000000005},
{"count":0,"bin_end":625.5,"bin_start":500.6},
{"count":0,"bin_end":750.40000000000001,"bin_start":625.5},
{"count":0,"bin_end":875.30000000000001,"bin_start":750.40000000000001},
{"count":0,"bin_end":1000.2,"bin_start":875.30000000000001},
{"count":0,"bin_end":1125.10000000000001,"bin_start":1000.2},
{"count":1,"bin_end":1250,"bin_start":1125.10000000000001}],"min":"1","
max":"1250","nan_count":0}},
{"name":"number_of_reviews","dtype":"int64","stats":
{"unique_count":261,"histogram":
[{"count":5167,"bin_end":33.7,"bin_start":1},
{"count":1217,"bin_end":66.4,"bin_start":33.7},
{"count":634,"bin_end":99.10000000000001,"bin_start":66.4},
{"count":459,"bin_end":131.8,"bin_start":99.10000000000001},
{"count":316,"bin_end":164.5,"bin_start":131.8},
{"count":142,"bin_end":197.20000000000002,"bin_start":164.5},
{"count":88,"bin_end":229.90000000000003,"bin_start":197.2000000000000
2},{ "count":38,"bin_end":262.6,"bin_start":229.90000000000003},
{"count":12,"bin_end":295.3,"bin_start":262.6},
{"count":2,"bin_end":328,"bin_start":295.3}],"min":"1","max":"328","na
n_count":0}},{"name":"reviews_per_month","dtype":"float64","stats":
{"unique_count":299,"histogram":
[{"count":3958,"bin_end":0.30800000000000005,"bin_start":1.0e-2},
{"count":1267,"bin_end":0.6060000000000001,"bin_start":0.308000000000
0005},
{"count":744,"bin_end":0.9040000000000001,"bin_start":0.60600000000000
01},
{"count":529,"bin_end":1.2020000000000002,"bin_start":0.90400000000000
01},
{"count":404,"bin_end":1.5000000000000002,"bin_start":1.20200000000000
02},
{"count":294,"bin_end":1.7980000000000003,"bin_start":1.50000000000000
02},{ "count":284,"bin_end":2.096,"bin_start":1.7980000000000003},
{"count":239,"bin_end":2.394,"bin_start":2.096},
{"count":193,"bin_end":2.692,"bin_start":2.394},
{"count":163,"bin_end":2.99,"bin_start":2.692}],"min":"0.01","max":"2.
99","nan_count":0}},
{"name":"availability_365","dtype":"int64","stats":
{"unique_count":366,"histogram":
[{"count":3759,"bin_end":36.5,"bin_start":0},
{"count":352,"bin_end":73,"bin_start":36.5},
```

```
{"count":381,"bin_end":109.5,"bin_start":73},
{"count":217,"bin_end":146,"bin_start":109.5},
{"count":328,"bin_end":182.5,"bin_start":146},
{"count":408,"bin_end":219,"bin_start":182.5},
{"count":421,"bin_end":255.5,"bin_start":219},
{"count":524,"bin_end":292,"bin_start":255.5},
{"count":655,"bin_end":328.5,"bin_start":292},
{"count":1030,"bin_end":365,"bin_start":328.5}], "min": "0", "max": "365",
"nan_count": 0}}, {"name": "_deepnote_index_column", "dtype": "int64"}]}
```

```
!pip install statsmodels==0.13.2
```

```
Collecting statsmodels==0.13.2
```

```
  Downloading statsmodels-0.13.2-cp39-cp39-
manylinux_2_17_x86_64.manylinux2014_x86_64.whl (9.9 MB)
----- 9.9/9.9 MB 70.0 MB/s eta
```

```
0:00:00
```

```
Requirement already satisfied: scipy>=1.3 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
statsmodels==0.13.2) (1.9.1)
```

```
Collecting patsy>=0.5.2
```

```
  Downloading patsy-0.5.2-py2.py3-none-any.whl (233 kB)
----- 233.7/233.7 KB 46.8 MB/s eta
```

```
0:00:00
```

```
Requirement already satisfied: packaging>=21.3 in /shared-libs/python3.9/py-
core/lib/python3.9/site-packages (from statsmodels==0.13.2) (21.3)
```

```
Requirement already satisfied: numpy>=1.17 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
statsmodels==0.13.2) (1.23.3)
```

```
Requirement already satisfied: pandas>=0.25 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
statsmodels==0.13.2) (1.2.5)
```

```
Requirement already satisfied: pyparsing!=3.0.5,>=2.0.2 in /shared-
libs/python3.9/py-core/lib/python3.9/site-packages (from
packaging>=21.3->statsmodels==0.13.2) (3.0.9)
```

```
Requirement already satisfied: pytz>=2017.3 in
/shared-libs/python3.9/py/lib/python3.9/site-packages (from
pandas>=0.25->statsmodels==0.13.2) (2022.2.1)
```

```
Requirement already satisfied: python-dateutil>=2.7.3 in /shared-
libs/python3.9/py-core/lib/python3.9/site-packages (from pandas>=0.25-
>statsmodels==0.13.2) (2.8.2)
```

```
Requirement already satisfied: six in
/shared-libs/python3.9/py-core/lib/python3.9/site-packages (from
patsy>=0.5.2->statsmodels==0.13.2) (1.16.0)
```

```
Installing collected packages: patsy, statsmodels
```

```
Successfully installed patsy-0.5.2 statsmodels-0.13.2
```

```
WARNING: You are using pip version 22.0.4; however, version 22.2.2 is
available.
```

```
You should consider upgrading via the '/root/venv/bin/python -m pip
install --upgrade pip' command.
```

```

from sklearn import linear_model
import statsmodels.api as sm

x =
new[['latitude','longitude','minimum_nights','number_of_reviews','reviews_per_month','availability_365']]
y = new['price']

```

```

# with sklearn
regr = linear_model.LinearRegression()
regr.fit(x, y)

```

```

print('Intercept: \n', regr.intercept_)
print('Coefficients: \n', regr.coef_)

```

```

# with statsmodels
x = sm.add_constant(x) # adding a constant

```

```

model = sm.OLS(y, x).fit()
predictions = model.predict(x)

```

```

print_model = model.summary()
print(print_model)

```

```

Intercept:
-71725.85115343943
Coefficients:
[ 1.40349515e+02 -8.94539421e+02 -2.34759992e-02 -1.31289724e-01
 -8.84071482e+00  1.28803040e-01]

```

OLS Regression Results

```

=====
=====
Dep. Variable:                price    R-squared:
0.020
Model:                        OLS      Adj. R-squared:
0.019
Method:                        Least Squares    F-statistic:
27.61
Date:                        Tue, 27 Sep 2022    Prob (F-statistic):
8.44e-33
Time:                        01:47:25    Log-Likelihood:
-55953.
No. Observations:            8075    AIC:
1.119e+05
Df Residuals:                8068    BIC:
1.120e+05
Df Model:                    6
Covariance Type:            nonrobust

```

```

=====
=====
                                coef      std err          t      P>|t|
[0.025      0.975]
-----
-----
const                -7.173e+04    6536.222    -10.974    0.000    -
8.45e+04    -5.89e+04
latitude             140.3495     53.137      2.641    0.008
36.187      244.512
longitude            -894.5394     79.088    -11.311    0.000    -
1049.572    -739.507
minimum_nights       -0.0235      0.112     -0.209    0.835    -
0.244      0.197
number_of_reviews    -0.1313      0.135     -0.970    0.332    -
0.397      0.134
reviews_per_month    -8.8407      9.212     -0.960    0.337    -
26.899      9.218
availability_365      0.1288      0.021      6.080    0.000
0.087      0.170
=====
=====
Omnibus:                19407.313    Durbin-Watson:
1.852
Prob(Omnibus):          0.000    Jarque-Bera (JB):
238170735.483
Skew:                   24.704    Prob(JB):
0.00
Kurtosis:               842.902    Cond. No.
4.80e+05
=====
=====

```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

[2] The condition number is large, 4.8e+05. This might indicate that there are strong multicollinearity or other numerical problems.

$$y_i = -7.173e+04 + 140.3495x_1 - 894.5394x_2 - 0.0235x_3 - 0.1313x_4 - 8.8407x_5 + 0.1288x_6$$

```
import pandas as pd
```

```
import numpy as np
```

```
from sklearn.linear_model import LinearRegression
```

```
# independent variable
```

```
X =
```

```
new[['latitude', 'longitude', 'minimum_nights', 'number_of_reviews', 'revi
```

```
ews_per_month','availability_365']]

# output variable
y = new['price']

# using the linear regression model
model = LinearRegression()

# fitting the data
model.fit(X, y)

# predicting values
y_pred = model.predict(X)
df = pd.DataFrame({'Actual': y, 'Predicted': y_pred})

print(' RSS : '+ str(np.sum(np.square(df['Predicted'] -
df['Actual']))))
```

RSS : 493462251.608315

Created in Deepnote