

```
In [21]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Exercise Pandas

For these exercises we are using a [dataset \(https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data/kernels\)](https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data/kernels) provided by Airbnb for a Kaggle competition. It describes its offer for New York City in 2019, including types of appartments, price, location etc.

1. Create a dataframe

Create a dataframe of a few lines with objects and their poperties (e.g fruits, their weight and colour). Calculate the mean of your Dataframe.

2. Import

- Import the table called `AB_NYC_2019.csv` as a dataframe. It is located in the Datasets folder. Have a look at the beginning of the table (head).
- Create a histogram of prices

3. Operations

Create a new column in the dataframe by multiplying the "price" and "availability_365" columns to get an estimate of the maximum yearly income.

3b. Subselection and plotting

Create a new Dataframe by first subselecting yearly incomes between 1 and 100'000. Then make a scatter plot of yearly income versus number of reviews

4. Combine

We provide below and additional table that contains the number of inhabitants of each of New York's boroughs ("neighbourhood_group" in the table). Use `merge` to add this population information to each element in the original dataframe.

5. Groups

- Using `groupby` calculate the average price for each type of room (`room_type`) in each `neighbourhood_group`. What is the average price for an entire home in Brooklyn ?
- Unstack the multi-level Dataframe into a regular Dataframe with `unstack()` and create a bar plot with the resulting table

6. Advanced plotting

Using Seaborn, create a scatter plot where x and y positions are longitude and latitude, the color reflects price and the shape of the marker the borough (`neighbourhood_group`). Can you recognize parts of new york ? Does the map make sense ?