

```

In [2]: from bs4 import BeautifulSoup
import urllib3
import re
import time
import pandas as pd
from pandas import Series, DataFrame

## ステップ1
http = urllib3.PoolManager()
url = "https://suumo.jp/jj/chintai/ichiran/FR301FC001/?ar=030&bs=040&ta=13&sc=13102&cb=0.0&ct=9999999&et"
response = http.request('GET', url)
soup = BeautifulSoup(response.data, "html.parser")

## ステップ2
# ページ数を取得
s = soup.find("div", {"class": "pagination pagination_set-nav"})
num_pages = int(s.find_all("a")[-2].string)

# 全てのページのURLを作成
urls = []
urls.append(url)
for i in range(num_pages-1):
    pg = str(i+2)
    url_page = url + '&pn=' + pg
    urls.append(url_page)

names = [] #マンション名
addresses = [] #住所
locations0 = [] #立地1つ目 (最寄駅/徒歩~分)
locations1 = [] #立地2つ目 (最寄駅/徒歩~分)
locations2 = [] #立地3つ目 (最寄駅/徒歩~分)
ages = [] #築年数
heights = [] #建物高さ

```

```

heights = [] #建物高さ
floors = [] #階
rent = [] #賃料
admin = [] #管理費
others = [] #敷/礼/保証/敷引,償却
floor_plans = [] #間取り
areas = [] #専有面積
detail_urls = [] # 詳細URL

## ステップ4
for url in urls:
    time.sleep(3)
    response = http.request('GET', url)
    soup = BeautifulSoup(response.data, "html.parser")

    apartments = soup.find_all('div', class_='cassetteitem')

## ステップ5
    for apartment in apartments:

        room_number = len(apartment.find_all('tbody'))

        name = apartment.find('div', class_='cassetteitem_content-title').text
        address = apartment.find('li', class_='cassetteitem_detail-col1').text

        for i in range(room_number):
            names.append(name)
            addresses.append(address)

## ステップ6
        sublocation = apartment.find('li', class_='cassetteitem_detail-col2')
        cols = sublocation.find_all('div')
        for i in range(len(cols)):
            text = cols[i].find(text=True)
            for j in range(room_number):
                if i == 0:
                    locations0.append(text)
                elif i == 1:
                    locations1.append(text)
                elif i == 2:

```

```
locations2.append(text)
```

```
## ステップ7
```

```
age_and_height = apartment.find('li', class_='cassetteitem_detail-col3')
age = age_and_height('div')[0].text
height = age_and_height('div')[1].text

for i in range(room_number):
    ages.append(age)
    heights.append(height)
```

```
## ステップ8
```

```
table = apartment.find('table')
rows = []
rows.append(table.find_all('tr'))

data = []
for row in rows:
    for tr in row:
        cols = tr.find_all('td')
        if len(cols) != 0:
            _floor = cols[2].text
            _floor = re.sub('[\r\n\t]', '', _floor)

            _rent_cell = cols[3].find('ul').find_all('li')
            _rent = _rent_cell[0].find('span').text
            _admin = _rent_cell[1].find('span').text

            _deposit_cell = cols[4].find('ul').find_all('li')
            _deposit = _deposit_cell[0].find('span').text
            _reikin = _deposit_cell[1].find('span').text
            _others = _deposit + '/' + _reikin

            _floor_cell = cols[5].find('ul').find_all('li')
            _floor_plan = _floor_cell[0].find('span').text
            _area = _floor_cell[1].find('span').text

            _detail_url = cols[8].find('a')['href']
            _detail_url = 'https://suumo.jp' + _detail_url
```

```

        text = [_floor, _rent, _admin, _others, _floor_plan, _area, _detail_url]
        data.append(text)

    for row in data:
        floors.append(row[0])
        rent.append(row[1])
        admin.append(row[2])
        others.append(row[3])
        floor_plans.append(row[4])
        areas.append(row[5])
        detail_urls.append(row[6])
    print(data)

## ステップ9
#各リストをシリーズ化
names = Series(names)
addresses = Series(addresses)
locations0 = Series(locations0)
locations1 = Series(locations1)
locations2 = Series(locations2)
ages = Series(ages)
heights = Series(heights)
floors = Series(floors)
rent = Series(rent)
admin = Series(admin)
others = Series(others)
floor_plans = Series(floor_plans)
areas = Series(areas)
detail_urls = Series(detail_urls)

suumo_df = pd.concat([names, addresses, locations0, locations1, locations2, ages, heights, floors, rent,
suumo_df.columns=['マンション名', '住所', '立地1', '立地2', '立地3', '築年数', '建物の高さ', '階層', '賃料料', '管理費', '敷
suumo_df.to_csv('suumo_cyuoku.csv', sep = '\t', encoding='utf-16')

```

/Users/kai/anaconda3/lib/python3.6/site-packages/urllib3/connectionpool.py:847: InsecureRequestWarning: Unverified HTTPS request is being made. Adding certificate verification is strongly advised. See: <https://urllib3.readthedocs.io/en/latest/advanced-usage.html#ssl-warnings>

```
.../urllib3.readthedocs.io/en/latest/advanced-usage.html#ssl-warnings  
(https://urllib3.readthedocs.io/en/latest/advanced-usage.html#ssl-warnings)
```

```
InsecureRequestWarning)
```

```
/Users/kai/anaconda3/lib/python3.6/site-packages/urllib3/connectionpool.py:847: InsecureRequestWarning:  
Unverified HTTPS request is being made. Adding certificate verification is strongly advised. See: https://urllib3.readthedocs.io/en/latest/advanced-usage.html#ssl-warnings
```

```
(https://urllib3.readthedocs.io/en/latest/advanced-usage.html#ssl-warnings)
```

```
InsecureRequestWarning)
```

```
[['4階', '10.4万円', '12000円', '10.4万円/-', 'ワンルーム', '32.92m2', 'https://suumo.jp/chintai/jnc_000047505747/?bc=100148802254'], ['4階', '11万円', '12000円', '11万円/-', 'ワンルーム', '32.77m2', 'https://suumo.jp/chintai/jnc_000040125469/?bc=100146064214'], ['4階', '11万円', '12000円', '11万円/-', 'ワンルーム', '32.77m2', 'https://suumo.jp/chintai/jnc_000047505745/?bc=100146716725'], ['7階', '11.9万円', '12000円', '11.9万円/-', '1DK', '32.92m2', 'https://suumo.jp/chintai/jnc_000047505748/?bc=100145784414'], ['11階', '12.2万円', '12000円', '12.2万円/-', '1DK', '32.77m2', 'https://suumo.jp/chintai/jnc_000047505740/?bc=100144314658'], ['4階', '11万円', '24000円', '11万円/-', 'ワンルーム', '32.77m2', 'https://suumo.jp/chintai/jnc_000047505744/?bc=100143996163'], ['13階', '12.3万円', '12000円', '12.3万円/-', '1DK', '32.77m2', 'https://suumo.jp/chintai/jnc_000047505742/?bc=100148113467'], ['11階', '12.2万円', '24000円', '12.2万円/-']]
```

In [ ]: