

数据获取作业 1:

源代码:

```
import requests
import bs4
import csv
import threading

# config here
REGION_NAMES = {"东城": "dongcheng", "西城": "xicheng",
                 "海淀": "haidian", "朝阳": "chaoyang"}
PAGE_NUM = 5
TERM_IN_PAGE_NUM = 30

BASE_URL = "https://bj.lianjia.com/ershoufang/"
CSV_HEADERS = ["区域", "楼盘名称", "总价", "面积", "单价"]

result_list = []

def one_page_pipeline(region_name: str, url: str):
    URL_TEXT = requests.get(url).text
    print("target url = {}".format(url))
    soup = bs4.BeautifulSoup(URL_TEXT, features="lxml")
    INFO_LIST = soup.select("#content > div.leftContent > ul")[0]
    for CHILD in range(1, TERM_IN_PAGE_NUM + 1):
        INFO = INFO_LIST.select(
            "li:nth-child({}) > div.info.clear".format(CHILD))[0]
        NAME = INFO.select("div.title > a")[0].text
        TOTAL_PRICE = INFO.select("div.priceInfo > div.totalPrice.totalPrice2 > span")[0].text + "万"
        AREA = INFO.select("div.address > div")[0].text.split('|')[1][1:-1]
        UNIT_PRICE = INFO.select(
            "div.priceInfo > div.unitPrice > span")[0].text
        RESULT = [region_name, NAME, TOTAL_PRICE, AREA, UNIT_PRICE]
        result_list.append(RESULT)

threads = []

for REGION_NAME, REGION_PINYIN in REGION_NAMES.items():
    for PAGE in range(1, PAGE_NUM + 1):
        URL = BASE_URL + REGION_PINYIN + "/pg{}/".format(PAGE)
        thread = threading.Thread(
            target=one_page_pipeline, args=[REGION_NAME, URL])
```

```

        threads.append(thread)
        thread.start()

for thread in threads:
    thread.join()

result_list.sort()

with open("res1.csv", "w", newline="", encoding="utf8") as f:
    f_csv = csv.writer(f)
    f_csv.writerow(CSV_HEADERS)
    f_csv.writerows(result_list)

```

csv 文件内容:

区域,楼盘名称,总价,面积,单价

东城,08 年商品房,小户型,2 梯 6 户,集中供暖,有门禁,460 万,27.38 平米,"168,006 元/平"  
 东城,70 年产权商品房 精装修 有燃气 满五年唯一 朝南,619 万,63.35 平米,"97,712 元/平"  
 东城,东二环 东直门外大街 春秀路小区中国港湾家属楼,685 万,60.63 平米,"112,981 元/平"  
 东城,东二环 西水井胡同 东西通透 正规两居室 电梯房,738 万,74.95 平米,"98,466 元/平"  
 东城,东二环 阳光都市 紧挨工体三里屯,2200 万,240.55 平米,"91,458 元/平"  
 东城,东二环内中间楼层,东南两居满五年,电梯楼.,628 万,59.87 平米,"104,894 元/平"  
 东城,东交民巷 满五唯一 中高层 独立小区 可免费停车,620 万,49.38 平米,"125,557 元/平"  
 东城,东单 金宝街 东堂子 中间楼层 户型好 诚意出售,525 万,44.32 平米,"118,457 元/平"  
 东城,东南三居室 全天采光 封闭式管理 人车分流 距地铁 50 米,1130 万,97.45 平米,"115,957 元/平"  
 东城,东四十条 二环里大两居 独门独院 业主诚售,750 万,73.09 平米,"102,614 元/平"  
 东城,东四地铁 板楼 精装小户型 低层双南二居 东四西大街,668 万,52.17 平米,"128,043 元/平"  
 东城,东城 西革新里 110 号院 满五年唯一,362 万,44.44 平米,"81,459 元/平"  
 东城,东城二环里新改造小区靠公园地铁,430 万,45.59 平米,"94,319 元/平"  
 东城,东城二环里长安街畔,70 年产权,贡院九号,精装,1480 万,156.83 平米,"94,370 元/平"  
 东城,东城区 南北通透两居室 中间层,535 万,72.8 平米,"73,490 元/平"  
 东城,东城区 小两居室 近地铁 精装修 满五年唯一,529 万,64.56 平米,"81,940 元/平"  
 东城,东城区 09 年商品房社区 东向两居室,666 万,80.77 平米,"82,457 元/平"  
 东城,东城区 2001 年商品房 南向两居 双阳台 满五唯一,913 万,114.17 平米,"79,969 元/平"  
 东城,东城区 2015 朝南两居 精装修,588 万,71.35 平米,"82,411 元/平"  
 东城,东城区 向新西里 全南 2 居 满五唯一诚心出售,528 万,54.47 平米,"96,935 元/平"  
 东城,东城区 定安里 南北通透两居室,450 万,57.3 平米,"78,535 元/平"  
 东城,东城区 悠胜美苑 03 年新楼 商品房 视野好,545 万,70.17 平米,"77,669 元/平"  
 东城,东城区 次新小区 一梯四户 使用面积大 东南向高层,698 万,67.53 平米,"103,362 元/平"  
 东城,东城区东直门外 十字坡东里 田字格三居室 中间楼层,789 万,72.37 平米,"109,024 元/平"  
 东城,东城区南北 2 居室,无遮挡采光好,485 万,68.62 平米,"70,680 元/平"  
 东城,东城区和平里中街 3 号院 户型好,1130 万,95.27 平米,"118,611 元/平"  
 东城,东城区崇文门 草厂八条 平房,270 万,16.3 平米,"165,645 元/平"  
 东城,东城区崇文门国瑞城东区南北两居室,935 万,79.57 平米,"117,507 元/平"  
 东城,东城区崇文门国瑞城中区 两居室朝东视野直观好,830 万,69.69 平米,"119,099 元/平"

东城,东城区广渠家园 2009 年楼龄 地铁 7 号线 板楼一梯三户,695 万,62.34 平米,"111,486 元/平"

东城,东城区的一居室, 南向无遮挡, 有电梯, 楼龄新,390 万,47.25 平米,"82,540 元/平"

东城,东城区聚龙花园南北三居室满五唯一,1290 万,180.58 平米,"71,437 元/平"

东城,东城区, 90 年钢混结构, 户型方正, 视野好, 阳光充足,615 万,79.39 平米,"77,466 元/平"

东城,东城区, 三居, 东南向, 中间楼层, 有车位! ,725 万,66.45 平米,"109,105 元/平"

东城,东城区, 西南两居室高楼层无遮挡 满五年,880 万,91.57 平米,"96,102 元/平"

东城,东城和平里两居带厅 封闭小区环境好 楼龄新,768 万,65.82 平米,"116,682 元/平"

东城,东城和平里五区一居室 值得选择,555 万,40.57 平米,"136,801 元/平"

东城,东城天天家园 板楼南北通透 2 居 一梯两户带电梯 采光好,960 万,126.73 平米,"75,752 元/平"

东城,东城广渠门交管局家属楼电梯 2 居不临街视野开阔,756 万,70.72 平米,"106,901 元/平"

东城,东城龙体区域南北通透 2 居室近龙潭湖公园,550 万,48.22 平米,"114,061 元/平"

东城,东城, 赵堂子胡同小区, 精装自住, 中粮和商务部家属楼,565 万,41.66 平米,"135,622 元/平"

东城,东堂子胡同, 精装修两居室, 随时可看,568 万,65.47 平米,"86,758 元/平"

东城,东直门 东城区 东方银座 西南向两居室,870 万,113.17 平米,"76,876 元/平"

东城,东花市北里中区 3 室 2 厅 南 北,1038 万,100.27 平米,"103,521 元/平"

东城,东西通透 明厨明卫 中间楼层 临近公园 业主诚心出售,575 万,52.95 平米,"108,594 元/平"

东城,东黄城根南街 南北通透 央产房满五年封闭式管理可停车,640 万,58.3 平米,"109,778 元/平"

东城,中海紫御公馆 开发商精装两居 对小区花园,1150 万,91.86 平米,"125,191 元/平"

东城,中海紫御公馆明厨明卫双南向大卧室两居,1028 万,90.08 平米,"114,121 元/平"

东城,二环内新上东南两居室 诚心出售,660 万,51.17 平米,"128,982 元/平"

## 数据获取作业 2:

源代码:

```
import requests
import bs4
import json

HEADERS = {
    "accept": "application/json, text/plain, */*",
    "accept-encoding": "gzip, deflate, br",
    "accept-language": "zh",
    "app-name": "xtzx",
    "cache-control": "no-cache",
    "django-language": "zh",
    "pragma": "no-cache",
    "referer": "https://www.xuetangx.com/university/all",
    "sec-ch-ua": "\"Google Chrome\";v=\"95\", \"Chromium\";v=\"95\", \";Not A Brand\";v=\"99\"",
    "sec-ch-ua-mobile": "?0",
    "sec-ch-ua-platform": "\"Linux\"",
    "sec-fetch-dest": "empty",
    "sec-fetch-mode": "cors",
    "sec-fetch-site": "same-origin",
```

```

        "terminal-type": "web",
        "user-agent": "Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/95.0.4638.54 Safari/537.36",
        "x-client": "web",
        "xtbz": "xt"
    }

result_list = []
page_num = 1
while (True):
    webpage = requests.get(
        "https://www.xuetangx.com/api/v1/lms/get_org_list/?page={}&page_size=100&apid=10000&query=&order=3".format(page_num), headers=HEADERS)
    INFO_JSON = json.loads(webpage.text)
    if len(INFO_JSON["data"]["org_list"]) == 0:
        break
    for INFO in INFO_JSON["data"]["org_list"]:
        result = {}
        result["name"] = INFO["name"]
        result["num"] = INFO["product_num"]
        result_list.append(result)
    page_num += 1
with open("res2.json", "w", encoding="utf8") as f:
    json.dump(result_list, f, ensure_ascii=False, indent=4)

```

Json 文件内容:

```

[
    {
        "name": "清华大学",
        "num": 423
    },
    {
        "name": "北京大学",
        "num": 24
    },
    {
        "name": "北京师范大学",
        "num": 61
    },
    {
        "name": "国防科技大学",
        "num": 15
    },
]

```

```
{
  "name": "西安交通大学",
  "num": 102
},
{
  "name": "哈尔滨工业大学",
  "num": 38
},
{
  "name": "华南理工大学",
  "num": 48
},
{
  "name": "南开大学",
  "num": 41
},
{
  "name": "复旦大学",
  "num": 9
},
{
  "name": "南京大学",
  "num": 9
},
{
  "name": "中国科学技术大学",
  "num": 7
},
{
  "name": "圣彼得堡国立大学",
  "num": 15
},
{
```