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&appid=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

    },

    {