

# PYTHON

## 数据预处理-1

编译环境:

PyCharm 2019.2.5 (Community Edition)

Build #PC-192.7142.56, built on November 19, 2019

Runtime version: 11.0.4+10-b304.77 amd64

VM: OpenJDK 64-Bit Server VM by JetBrains s.r.o

Windows 10 10.0

GC: ParNew, ConcurrentMarkSweep

Memory: 1963M

Cores: 8

Registry:

Non-Bundled Plugins:

python 版本: 3.7 (Anaconda3)

scrapy 版本: 1.8

### 作业 1:

1、爬取新房数据的爬虫:

Spider.py

```
import scrapy
import re
from q19_scrapy.items import Q19ScrapyItem
```

```
class mySpider(scrapy.spiders.Spider):
    name = "loufang"
```

```

allowed_domains = ["www.bj.fang.lianjia.com/"]
start_urls = []
for page in range(1, 5):
    url = "https://bj.fang.lianjia.com/loupan/nhs1pg{}/".format(page)
    start_urls.append(url)

def parse(self, response):
    item = Q19ScrapyItem()
    for each in response.xpath("/html/body/div[4]/ul[2]/*"):

        item['name'] = each.xpath("div/div[1]/a/text()").extract()
        item['location1'] = each.xpath("div/div[2]/span[1]/text()").extract()
        item['location2'] = each.xpath("div/div[2]/span[2]/text()").extract()
        item['location3'] = each.xpath("div/div[2]/a/text()").extract()
        item['house_type'] = each.xpath("div/a/span[1]/text()").extract()
        size = each.xpath("div/div[3]/span/text()").extract()
        price = each.xpath("div/div[6]/div[1]/span[1]/text()").extract()
        unit = each.xpath("div/div[6]/div[1]/span[2]/text()").extract()
        if unit:
            unit[0] = unit[0].replace("\xa0", " ")
        if size:
            item['size'] = [re.findall(r'\d+', size[0])[0]]

        if unit[0] == '元/平(均价)':
            item['avg_price'] = price
            item['sum_price'] = []
        else:
            item['sum_price'] = price
            item['avg_price'] = []

        if item['name'] and item['location1'] and item['location2'] and
item['location3'] and \
            item['house_type'] and item['size'] and (item['sum_price'] or
item['avg_price']):
            yield (item)

```

## Items.py

```

# -*- coding: utf-8 -*-

# Define here the models for your scraped items
#
# See documentation in:
# https://docs.scrapy.org/en/latest/topics/items.html

import scrapy

```

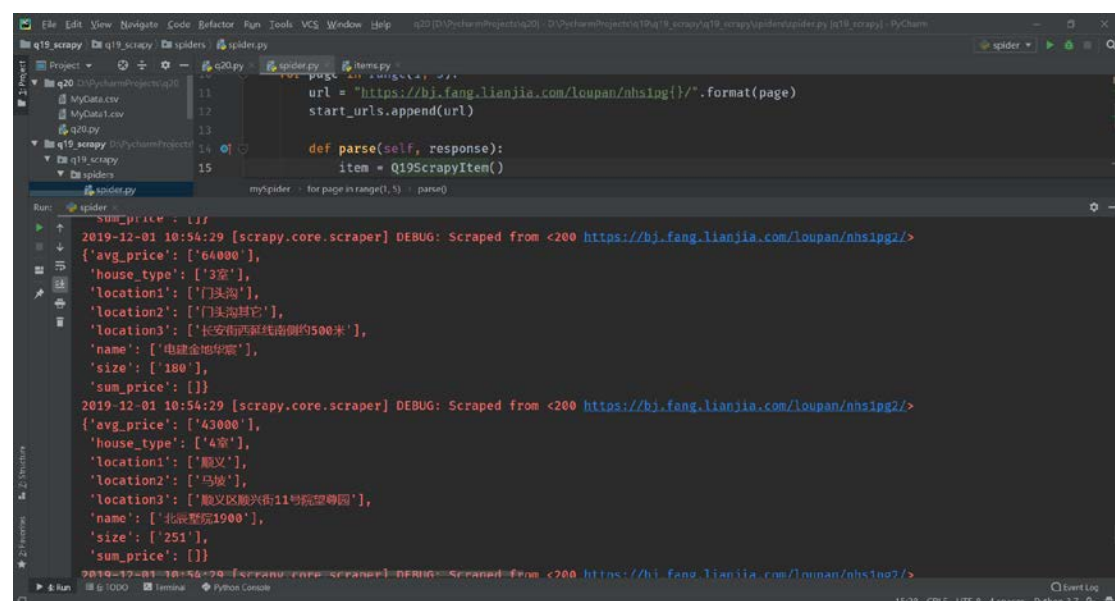
```

class Q19ScrapyItem(scrapy.Item):
    name = scrapy.Field()
    location1 = scrapy.Field()
    location2 = scrapy.Field()
    location3 = scrapy.Field()
    house_type = scrapy.Field()
    size = scrapy.Field()
    sum_price = scrapy.Field()
    avg_price = scrapy.Field()
    pass

```

name 为楼盘名称，location 为位置，分为三部分，house\_type 为房型，size 为建面大小，sum\_price 为总价价格，avg\_price 为均价。

运行：



The screenshot shows a PyCharm IDE with a file named 'spider.py' open. The code in the file is as follows:

```

11 url = "https://bj.fang.lianjia.com/loupan/nhs1pg{}/".format(page)
12 start_urls.append(url)
13
14 def parse(self, response):
15     item = Q19ScrapyItem()

```

The Run console shows the output of the spider, including the following data:

```

2019-12-01 10:54:29 [scrapy.core.scraper] DEBUG: Scraped from <200 https://bj.fang.lianjia.com/loupan/nhs1pg2/>
{'avg_price': ['64000'],
 'house_type': ['3室'],
 'location1': ['门头沟'],
 'location2': ['门头沟其它'],
 'location3': ['长安街西线南侧约500米'],
 'name': ['电建金地华寓'],
 'size': ['180'],
 'sum_price': []}

2019-12-01 10:54:29 [scrapy.core.scraper] DEBUG: Scraped from <200 https://bj.fang.lianjia.com/loupan/nhs1pg2/>
{'avg_price': ['43000'],
 'house_type': ['4室'],
 'location1': ['顺义'],
 'location2': ['马坡'],
 'location3': ['顺义区顺兴街11号院望都园'],
 'name': ['北宸墅院1900'],
 'size': ['251'],
 'sum_price': []}

```

结果：

```

{"name": ["天恒水岸壹号"], "location1": ["房山"], "location2": ["良乡"], "location3": ["良乡大学城西站地铁南侧 400 米，刺猬河旁"], "house_type": ["3 室"], "size": ["184"], "sum_price": ["1000"], "avg_price": []}

```

```

{"name": ["观唐云鼎"], "location1": ["密云"], "location2": ["密云其它"], "location3": ["密溪路 39 号院"], "house_type": ["4 室"], "size": ["170"], "sum_price": ["460"],

```

"avg\_price": []}

{"name": ["运河铭著"], "location1": ["通州"], "location2": ["北关"], "location3": ["商通  
大道与榆东一街交叉口,温榆河森林公园东 500 米"], "house\_type": ["2 室"], "size": ["100"],  
"sum\_price": [], "avg\_price": ["46000"]}

{"name": ["万年广阳郡九号"], "location1": ["房山"], "location2": ["长阳"], "location3": ["  
长阳清苑南街与汇商东路交汇处西北角"], "house\_type": ["3 室"], "size": ["139"],  
"sum\_price": [], "avg\_price": ["48500"]}

{"name": ["首开璞瑅公馆"], "location1": ["丰台"], "location2": ["方庄"], "location3": ["  
紫芳园五区"], "house\_type": ["4 室"], "size": ["236"], "sum\_price": [], "avg\_price":  
["106000"]}

{"name": ["华远裘马四季"], "location1": ["门头沟"], "location2": ["大峪"], "location3": ["  
增产路 16 号院"], "house\_type": ["3 室"], "size": ["156"], "sum\_price": [], "avg\_price":  
["60000"]}

{"name": ["懋源钧云台"], "location1": ["丰台"], "location2": ["玉泉营"], "location3": ["  
西三环康庄东路于丰台东路交汇西南角, 康庄东路 9 号院"], "house\_type": ["4 室"], "size":  
["251"], "sum\_price": ["2900"], "avg\_price": []}

{"name": ["V7 九间堂"], "location1": ["通州"], "location2": ["潞苑"], "location3": ["通燕  
高速耿庄桥北出口中化石油对面"], "house\_type": ["4 室"], "size": ["220"], "sum\_price":  
["1700"], "avg\_price": []}

{"name": ["北京城建国誉府"], "location1": ["房山"], "location2": ["长阳"], "location3": ["  
地铁良乡大学城北站东北约 1300 米, 文昌东路与阜盛东街交汇处"], "house\_type": ["4 室  
"], "size": ["143"], "sum\_price": [], "avg\_price": ["48000"]}

{"name": ["御汤山熙园"], "location1": ["昌平"], "location2": ["昌平其它"], "location3": ["小汤山疗养院西侧，紧邻安泗路，距离北六环 61 号出口约 2000 米"], "house\_type": ["4 室"], "size": ["300"], "sum\_price": ["1800"], "avg\_price": []}

{"name": ["兴创荣墅"], "location1": ["大兴"], "location2": ["大兴其它"], "location3": ["采育镇育胜街与福源路交叉口西侧 350 米路南"], "house\_type": ["3 室"], "size": ["240"], "sum\_price": ["1100"], "avg\_price": []}

{"name": ["北科建泰禾丽春湖院子"], "location1": ["昌平"], "location2": ["沙河"], "location3": ["中关村北延新核心，沙河水库边（地铁昌平线沙河站向南 800 米）"], "house\_type": ["5 室"], "size": ["468"], "sum\_price": ["1020"], "avg\_price": []}

{"name": ["景粼原著"], "location1": ["朝阳"], "location2": ["中央别墅区"], "location3": ["京密路与顺黄路交叉路口往西 200 米"], "house\_type": ["4 室"], "size": ["479"], "sum\_price": ["4000"], "avg\_price": []}

{"name": ["首创禧瑞山"], "location1": ["顺义"], "location2": ["马坡"], "location3": ["牛栏山镇昌金路与右堤路交界处"], "house\_type": ["3 室"], "size": ["141"], "sum\_price": ["830"], "avg\_price": []}

{"name": ["观山源墅"], "location1": ["房山"], "location2": ["良乡"], "location3": ["阳光北大街与多宝路交汇处西南(理工大学北校区西侧)"], "house\_type": ["3 室"], "size": ["92"], "sum\_price": [], "avg\_price": ["47500"]}

{"name": ["北辰墅院 1900"], "location1": ["顺义"], "location2": ["马坡"], "location3": ["顺义区顺兴街 11 号院望尊园"], "house\_type": ["3 室"], "size": ["120"], "sum\_price": [], "avg\_price": ["43000"]}

{"name": ["利锦府府上"], "location1": ["朝阳"], "location2": ["东坝"], "location3": ["东

五环七棵树出口沿东坝中街向东 1500 米, 红松园北里 18 号院"], "house\_type": ["4 室"],  
"size": ["280"], "sum\_price": [], "avg\_price": ["60000"]}

{"name": ["中骏西山天璟"], "location1": ["门头沟"], "location2": ["城子"], "location3": ["  
西山永定楼北 300 米"], "house\_type": ["4 室"], "size": ["117"], "sum\_price": [],  
"avg\_price": ["65000"]}

{"name": ["泰禾昌平拾景园"], "location1": ["昌平"], "location2": ["南邵"], "location3": ["  
地铁昌平线南邵站西 100 米"], "house\_type": ["3 室"], "size": ["210"], "sum\_price":  
["1100"], "avg\_price": []}

{"name": ["国瑞熙墅"], "location1": ["昌平"], "location2": ["北七家"], "location3": ["北  
七家镇定泗路与七北路交叉口东行 200 米"], "house\_type": ["3 室"], "size": ["314"],  
"sum\_price": ["1400"], "avg\_price": []}

{"name": ["领秀翡翠墅"], "location1": ["丰台"], "location2": ["丰台其它"], "location3": ["  
王佐镇长青路南侧, 长青路 88 号院"], "house\_type": ["3 室"], "size": ["137"], "sum\_price":  
[], "avg\_price": ["51000"]}

{"name": ["绿地海珀云翡"], "location1": ["大兴"], "location2": ["大兴其它"], "location3":  
["兴亦路京开高速东侧 (黄村镇第一中心小学对面) "], "house\_type": ["2 室"], "size":  
["102"], "sum\_price": [], "avg\_price": ["70000"]}

{"name": ["都丽华府"], "location1": ["平谷"], "location2": ["平谷其它"], "location3": ["  
新平南路与林荫南街交汇处向西 100 米"], "house\_type": ["2 室"], "size": ["90"],  
"sum\_price": ["300"], "avg\_price": []}

{"name": ["华远和墅"], "location1": ["大兴"], "location2": ["大兴其它"], "location3": ["  
南六环磁各庄桥沿南中轴向南 2 公里"], "house\_type": ["5 室"], "size": ["295"],

"sum\_price": ["1580"], "avg\_price": []}

{"name": ["中粮京西祥云"], "location1": ["房山"], "location2": ["长阳"], "location3": ["地铁稻田站北 800 米, 西邻京深路"], "house\_type": ["4 室"], "size": ["115"], "sum\_price": [], "avg\_price": ["58000"]}

{"name": ["燕西华府"], "location1": ["丰台"], "location2": ["丰台其它"], "location3": ["王佐镇青龙湖公园东 1500 米, 泉湖西路 1 号院 (七区) , 泉湖西路 1 号院 (六区) "], "house\_type": ["4 室"], "size": ["60"], "sum\_price": [], "avg\_price": ["49800"]}

{"name": ["天恒水岸壹号"], "location1": ["房山"], "location2": ["良乡"], "location3": ["良乡大学城西站地铁南侧 400 米, 刺猬河旁"], "house\_type": ["2 室"], "size": ["108"], "sum\_price": ["600"], "avg\_price": []}

{"name": ["紫辰院"], "location1": ["丰台"], "location2": ["岳各庄"], "location3": ["岳各庄北桥东北角 200 米处"], "house\_type": ["4 室"], "size": ["266"], "sum\_price": [], "avg\_price": ["120000"]}

{"name": ["天恒摩墅"], "location1": ["房山"], "location2": ["房山其它"], "location3": ["周口店镇政府东 200 米"], "house\_type": ["3 室"], "size": ["134"], "sum\_price": ["370"], "avg\_price": []}

{"name": ["天资华府"], "location1": ["房山"], "location2": ["长阳"], "location3": ["长阳 CSD 办公楼南侧 (京港澳高速长阳出口西 1000 米)"], "house\_type": ["5 室"], "size": ["93"], "sum\_price": [], "avg\_price": ["42000"]}

{"name": ["檀香府"], "location1": ["门头沟"], "location2": ["门头沟其它"], "location3": ["京潭大街与潭柘十街交叉口"], "house\_type": ["3 室"], "size": ["208"], "sum\_price": [], "avg\_price": ["55000"]}

```
{"name": ["观山源墅"], "location1": ["房山"], "location2": ["良乡"], "location3": ["阳光  
北大街与多宝路交汇处西南（理工大学北校区西侧）"], "house_type": ["3 室"], "size":  
["290"], "sum_price": [], "avg_price": ["47500"]}
```

```
{"name": ["中国铁建花语金郡"], "location1": ["大兴"], "location2": ["瀛海"], "location3":  
["南海子公园西侧(南五环旧忠桥向南第二个红绿灯西 300 米)"], "house_type": ["3 室"],  
"size": ["150"], "sum_price": [], "avg_price": ["66000"]}
```

```
{"name": ["中国铁建万科翡翠长安"], "location1": ["门头沟"], "location2": ["门头沟其它  
"], "location3": ["西长安街沿线与六环交汇处西南 1.2 公里处"], "house_type": ["3 室"],  
"size": ["166"], "sum_price": [], "avg_price": ["57000"]}
```

```
{"name": ["电建金地华宸"], "location1": ["门头沟"], "location2": ["门头沟其它"],  
"location3": ["长安街西延线南侧约 500 米"], "house_type": ["3 室"], "size": ["180"],  
"sum_price": [], "avg_price": ["64000"]}
```

```
{"name": ["北辰墅院 1900"], "location1": ["顺义"], "location2": ["马坡"], "location3": ["  
顺义区顺兴街 11 号院望尊园"], "house_type": ["4 室"], "size": ["251"], "sum_price": [],  
"avg_price": ["43000"]}
```

```
{"name": ["西山燕庐"], "location1": ["门头沟"], "location2": ["门头沟其它"], "location3":  
["长安街延线南约 500 米"], "house_type": ["3 室"], "size": ["123"], "sum_price": [],  
"avg_price": ["65000"]}
```

## 2、数据处理部分

将 json 文件转化为 csv 文件 MyData.csv:

```
"name","location1","location2","location3","house_type","size","sum_price","avg_pri  
ce"
```



"中国铁建花语金郡","大兴","瀛海","南海子公园西侧(南五环旧忠桥向南第二个红绿灯西 300 米)","3 室","150","","66000"

"中国铁建万科翡翠长安","门头沟","门头沟其它","西长安街沿线与六环交汇处西南 1.2 公里处","3 室","166","","57000"

"电建金地华宸","门头沟","门头沟其它","长安街西延线南侧约 500 米","3 室","180","","64000"

"北辰墅院 1900","顺义","马坡","顺义区顺兴街 11 号院望尊园","4 室","251","","43000"

"西山燕庐","门头沟","门头沟其它","长安街延线南约 500 米","3 室","123","","65000"

"北科建泰禾丽春湖院子","昌平","沙河","中关村北延新核心, 沙河水库边(地铁昌平线沙河站向南 800 米)","5 室","468","1020",""

"绿地海珀云翡","大兴","大兴其它","兴亦路京开高速东侧(黄村镇第一中心小学对面)","2 室","102","","70000"

"都丽华府","平谷","平谷其它","新平南路与林荫南街交汇处向西 100 米","2 室","90","300",""

"中粮京西祥云","房山","长阳","地铁稻田站北 800 米, 西邻京深路","4 室","115","","58000"

"燕西华府","丰台","丰台其它","王佐镇青龙湖公园东 1500 米, 泉湖西路 1 号院(七区), 泉湖西路 1 号院(六区)","4 室","60","","49800"

"天恒水岸壹号","房山","良乡","良乡大学城西站地铁南侧 400 米, 刺猬河旁","2 室","108","600",""

"紫辰院","丰台","岳各庄","岳各庄北桥东北角 200 米处","4 室","266","","120000"

"天恒摩墅","房山","房山其它","周口店镇镇政府东 200 米","3 室","134","370",""

"兴创荣墅","大兴","大兴其它","采育镇育胜街与福源路交叉口西侧 350 米路南","3 室",  
",240","1100",""

"景粼原著","朝阳","中央别墅区","京密路与顺黄路交叉路口往西 200 米","4 室",  
",479","4000",""

"首创禧瑞山","顺义","马坡","牛栏山镇昌金路与右堤路交界处","3 室","141","830",""

"观山源墅","房山","良乡","阳光北大街与多宝路交汇处西南(理工大学北校区西侧)","3 室",  
",92","",47500"

"首创远洋禧瑞天著","亦庄开发区","通州其它","科创十一街与经海九路交汇处西南角(亦庄  
线次渠南站 200 米)","2 室","89","",52695"

"首开璞瑅公馆","丰台","方庄","紫芳园五区","4 室","236","",106000"

"绿城西府海棠","石景山","石景山其它","隆恩寺路北人大附石景山分校西侧","3 室",  
",90","",52024"

"颐璟万和","大兴","黄村火车站","新源大街 4 号线义和庄地铁向北约 500 米","3 室",  
",80","",55016"

"泰禾金府大院","丰台","西红门","南四环地铁新宫站南 800 米","4 室","175","1400",""

"北京东湾","通州","万达","通惠北路 88 号","1 室","58","",68500"

"华发中央公园","房山","良乡","北京市房山区良乡大学城站南 900 米","3 室",  
",85","",41000"

"天恒水岸壹号","房山","良乡","良乡大学城西站地铁南侧 400 米,刺猬河旁","3 室",  
",184","1000",""

"观唐云鼎","密云","密云其它","密溪路 39 号院","4 室","170","460",""

"运河铭著","通州","北关","商通大道与榆东一街交叉口,温榆河森林公园东 500 米","2 室"

","100","","46000"

"万年广阳郡九号","房山","长阳","长阳清苑南街与汇商东路交汇处西北角","3 室  
","139","","48500"

"华远裘马四季","门头沟","大峪","增产路 16 号院","3 室","156","","60000"

"懋源钓云台","丰台","玉泉营","西三环康庄东路于丰台东路交汇西南角, 康庄东路 9 号院  
","4 室","251","2900",""

"V7 九间堂","通州","潞苑","通燕高速耿庄桥北出口中化石油对面","4 室","220","1700",""

"北京城建国誉府","房山","长阳","地铁良乡大学城北站东北约 1300 米, 文昌东路与阜盛东  
街交汇处","4 室","143","","48000"

"御汤山熙园","昌平","昌平其它","小汤山疗养院西侧, 紧邻安泗路, 距离北六环 61 号出口  
约 2000 米","4 室","300","1800",""

"华远和墅","大兴","大兴其它","南六环磁各庄桥沿南中轴向南 2 公里","5 室  
","295","1580",""

"天资华府","房山","长阳","长阳 CSD 办公楼南侧 (京港澳高速长阳出口西 1000 米) ","5 室  
","93","","42000"

"檀香府","门头沟","门头沟其它","京潭大街与潭柘十街交叉口","3 室","208","","55000"

"观山源墅","房山","良乡","阳光北大街与多宝路交汇处西南 (理工大学北校区西侧) ","3 室  
","290","","47500"

## Py 源文件:

```
import numpy as np
import pandas as pd
```

```
fileNameStr = 'D:\\PycharmProjects\\q20\\MyData.csv'
fileNameStr1 = 'D:\\PycharmProjects\\q20\\MyData1.csv'
df = pd.read_csv(fileNameStr, encoding='utf-8', dtype=str)
```

```

df['name'] = df['name'].str.strip()
df['location1'] = df['location1'].str.strip()
df['location2'] = df['location2'].str.strip()
df['location3'] = df['location3'].str.strip()
df['house_type'] = df['house_type'].str.strip()
df['size'] = df['size'].str.strip()
df['sum_price'] = df['sum_price'].str.strip()
df['avg_price'] = df['avg_price'].str.strip()
df['size'] = df['size'].astype(np.int)

df['sum_price'].fillna(0, inplace=True)
df['avg_price'].fillna(0, inplace=True)
df['sum_price'] = df['sum_price'].astype(np.int)
df['avg_price'] = df['avg_price'].astype(np.float)
df['avg_price'] = df['avg_price']/10000

df.loc[(df.sum_price == 0), 'sum_price'] = \
    (df[df['sum_price'] == 0]['avg_price'] * df[df['sum_price'] == 0]['size']).map(lambda x:
"%d" % x)

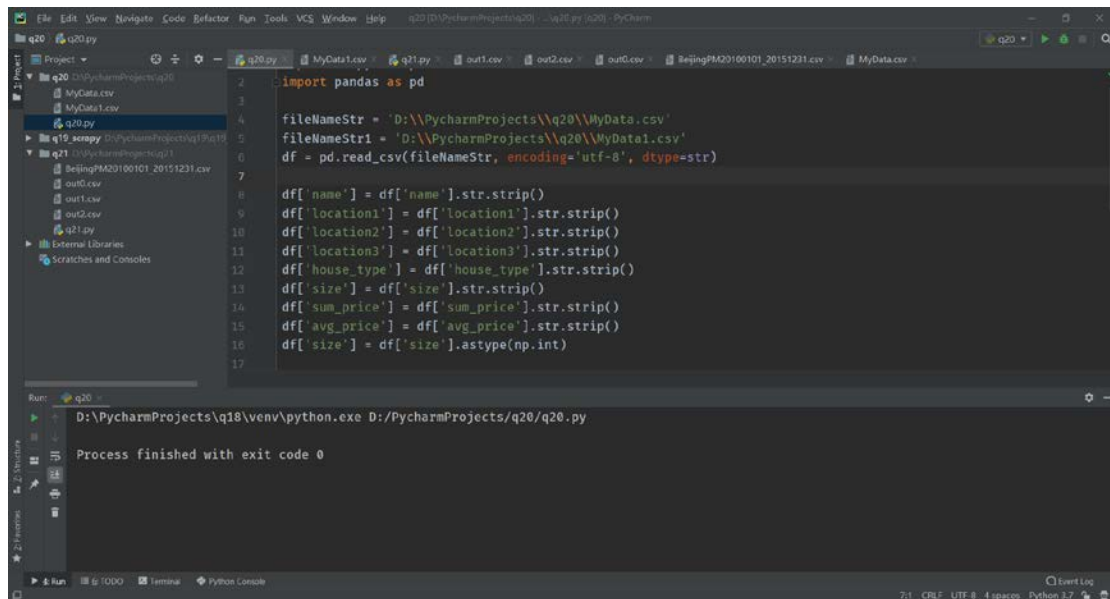
df.loc[(df.avg_price != 0), 'avg_price'] = df[df['avg_price'] != 0]['avg_price'].map(lambda x:
"%0.4f" % x)
df.loc[(df.avg_price == 0), 'avg_price'] = \
    (df[df['avg_price'] == 0]['sum_price'] / df[df['avg_price'] == 0]['size']).map(lambda x:
"%0.4f" % x)

df.to_csv(fileNameStr1)

```

**结果：**

**运行：**



```
1 import pandas as pd
2
3
4 fileNameStr = 'D:\\PycharmProjects\\q20\\MyData.csv'
5 fileNameStr1 = 'D:\\PycharmProjects\\q20\\MyData1.csv'
6 df = pd.read_csv(fileNameStr, encoding='utf-8', dtype=str)
7
8 df['name'] = df['name'].str.strip()
9 df['location1'] = df['location1'].str.strip()
10 df['location2'] = df['location2'].str.strip()
11 df['location3'] = df['location3'].str.strip()
12 df['house_type'] = df['house_type'].str.strip()
13 df['size'] = df['size'].str.strip()
14 df['sum_price'] = df['sum_price'].str.strip()
15 df['avg_price'] = df['avg_price'].str.strip()
16 df['size'] = df['size'].astype(np.int)
17
```

Run: q20  
D:\PycharmProjects\q18\venv\python.exe D:/PycharmProjects/q20/q20.py  
Process finished with exit code 0

## MyData1.csv:

,name,location1,location2,location3,house\_type,size,sum\_price,avg\_price

0,中国铁建花语金郡,大兴,瀛海,南海子公园西侧(南五环旧忠桥向南第二个红绿灯西 300 米),3 室,150,990,6.6000

1,中国铁建万科翡翠长安,门头沟,门头沟其它,西长安街沿线与六环交汇处西南 1.2 公里处,3 室,166,946,5.7000

2,电建金地华宸,门头沟,门头沟其它,长安街西延线南侧约 500 米,3 室,180,1152,6.4000

3,北辰墅院 1900,顺义,马坡,顺义区顺兴街 11 号院望尊园,4 室,251,1079,4.3000

4,西山燕庐,门头沟,门头沟其它,长安街延线南约 500 米,3 室,123,799,6.5000

5,北科建泰禾丽春湖院子,昌平,沙河,中关村北延新核心, 沙河水库边 (地铁昌平线沙河站向南 800 米) ,5 室,468,1020,2.1795

6,绿地海珀云翡,大兴,大兴其它,兴亦路京开高速东侧 (黄村镇第一中心小学对面) ,2 室,102,714,7.0000

7,都丽华府,平谷,平谷其它,新平南路与林荫南街交汇处向西 100 米,2 室,90,300,3.3333

8,中粮京西祥云,房山,长阳,地铁稻田站北 800 米, 西邻京深路,4 室,115,667,5.8000

- 9,燕西华府,丰台,丰台其它,"王佐镇青龙湖公园东 1500 米, 泉湖西路 1 号院 (七区) , 泉湖西路 1 号院 (六区) ",4 室,60,298,4.9800
- 10,天恒水岸壹号,房山,良乡,良乡大学城西站地铁南侧 400 米, 刺猬河旁,2 室,108,600,5.5556
- 11,紫宸院,丰台,岳各庄,岳各庄北桥东北角 200 米处,4 室,266,3192,12.0000
- 12,天恒摩墅,房山,房山其它,周口店镇政府东 200 米,3 室,134,370,2.7612
- 13,兴创荣墅,大兴,大兴其它,采育镇育胜街与福源路交叉口西侧 350 米路南,3 室,240,1100,4.5833
- 14,景粼原著,朝阳,中央别墅区,京密路与顺黄路交叉路口往西 200 米,4 室,479,4000,8.3507
- 15,首创禧瑞山,顺义,马坡,牛栏山镇昌金路与右堤路交界处,3 室,141,830,5.8865
- 16,观山源墅,房山,良乡,阳光北大街与多宝路交汇处西南 (理工大学北校区西侧) ,3 室,92,437,4.7500
- 17,首创远洋禧瑞天著,亦庄开发区,通州其它,科创十一街与经海九路交汇处西南角 (亦庄线次渠南站 200 米) ,2 室,89,468,5.2695
- 18,首开璞瑅公馆,丰台,方庄,紫芳园五区,4 室,236,2501,10.6000
- 19,绿城西府海棠,石景山,石景山其它,隆恩寺路北人大附石景山分校西侧,3 室,90,468,5.2024
- 20,颐璟万和,大兴,黄村火车站,新源大街 4 号线义和庄地铁向北约 500 米,3 室,80,440,5.5016
- 21,泰禾金府大院,丰台,西红门,南四环地铁新宫站南 800 米,4 室,175,1400,8.0000
- 22,北京东湾,通州,万达,通惠北路 88 号,1 室,58,397,6.8500
- 23,华发中央公园,房山,良乡,北京市房山区良乡大学城站南 900 米,3 室,85,348,4.1000

24,天恒水岸壹号,房山,良乡,良乡大学城西站地铁南侧 400 米, 刺猬河旁,3  
室,184,1000,5.4348

25,观唐云鼎,密云,密云其它,密溪路 39 号院,4 室,170,460,2.7059

26,运河铭著,通州,北关,商通大道与榆东一街交叉口, 温榆河森林公园东 500 米,2  
室,100,459,4.6000

27,万年广阳郡九号,房山,长阳,长阳清苑南街与汇商东路交汇处西北角,3 室,139,674,4.8500

28,华远裘马四季,门头沟,大峪,增产路 16 号院,3 室,156,936,6.0000

29,懋源钓云台,丰台,玉泉营,"西三环康庄东路于丰台东路交汇西南角, 康庄东路 9 号院",4  
室,251,2900,11.5538

30,V7 九间堂,通州,潞苑,通燕高速耿庄桥北出口中化石油对面,4 室,220,1700,7.7273

31,北京城建国誉府,房山,长阳,地铁良乡大学城北站东北约 1300 米, 文昌东路与阜盛东街  
交汇处,4 室,143,686,4.8000

32,御汤山熙园,昌平,昌平其它,小汤山疗养院西侧, 紧邻安泗路, 距离北六环 61 号出口约  
2000 米,4 室,300,1800,6.0000

33,华远和墅,大兴,大兴其它,南六环磁各庄桥沿南中轴向南 2 公里,5 室,295,1580,5.3559

34,天资华府,房山,长阳,长阳 CSD 办公楼南侧 (京港澳高速长阳出口西 1000 米) ,5  
室,93,390,4.2000

35,檀香府,门头沟,门头沟其它,京潭大街与潭柘十街交叉口,3 室,208,1144,5.5000

36,观山源墅,房山,良乡,阳光北大街与多宝路交汇处西南 (理工大学北校区西侧) ,3  
室,290,1377,4.7500

## 作业 2:

### 1、py 源文件

```

import matplotlib.pyplot as plt
import pandas as pd

fileNameStr = 'D:\\PycharmProjects\\q21\\BeijingPM20100101_20151231.csv'
fileNameStr1 = 'D:\\PycharmProjects\\q21\\out0.csv'
fileNameStr2 = 'D:\\PycharmProjects\\q21\\out1.csv'
fileNameStr3 = 'D:\\PycharmProjects\\q21\\out2.csv'
df = pd.read_csv(fileNameStr, encoding='utf-8', usecols=[0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

df.dropna(axis='index', how='all', subset=['PM_Dongsi', 'PM_Dongsihuan',
'PM_Nongzhanguan', 'PM_US Post'], inplace=True)

df['sum'] = df[['PM_Dongsi', 'PM_Dongsihuan', 'PM_Nongzhanguan', 'PM_US
Post']].sum(axis=1)
df['count'] = df[['PM_Dongsi', 'PM_Dongsihuan', 'PM_Nongzhanguan', 'PM_US
Post']].count(axis=1)
df['avg'] = round(df['sum'] / df['count'], 2)

df_mean1 = df.groupby(['year'])[['avg']].mean()
df_mean2 = df.groupby(['year', "month"])[['avg']].mean()
df_mean1 = pd.DataFrame(df_mean1)
df_mean2 = pd.DataFrame(df_mean2)

print(df_mean1)
print(df_mean2)

df.groupby(['year'])[['avg']].mean().plot()
df.groupby(['year', "month"])[['avg']].mean().plot()
plt.show()

df.to_csv(fileNameStr1)
df_mean1.to_csv(fileNameStr2)
df_mean2.to_csv(fileNameStr3)

```

## 2、 结果

运行：



```

df.dropna(axis='index', how='all', subset=['PM_Dongsi', 'PM_Dongsihuan', 'PM_Nongzhanguan', 'PM_US Post'], inplace=True)
df['sum'] = df[['PM_Dongsi', 'PM_Dongsihuan', 'PM_Nongzhanguan', 'PM_US Post']].sum(axis=1)
df['count'] = df[['PM_Dongsi', 'PM_Dongsihuan', 'PM_Nongzhanguan', 'PM_US Post']].count(axis=1)
df['avg'] = round(df['sum'] / df['count'], 2)

df_mean1 = df.groupby(['year'])['avg'].mean()

```

year	avg
2010	104.045730
2011	99.093240
2012	90.530768
2013	98.402683
2014	93.917709
2015	85.858937

year	month	avg
2010	1	90.403670
	2	97.239940
	3	94.046544
	4	80.072423
	5	87.071913

## out0.csv 的一部分:

,No,year,month,day,hour,season,PM\_Dongsi,PM\_Dongsihuan,PM\_Nongzhanguan,PM\_US Post,sum,count,avg

23,24,2010,1,1,23,4,,,129.0,129.0,1,129.0  
 24,25,2010,1,2,0,4,,,148.0,148.0,1,148.0  
 25,26,2010,1,2,1,4,,,159.0,159.0,1,159.0  
 26,27,2010,1,2,2,4,,,181.0,181.0,1,181.0  
 27,28,2010,1,2,3,4,,,138.0,138.0,1,138.0  
 28,29,2010,1,2,4,4,,,109.0,109.0,1,109.0  
 29,30,2010,1,2,5,4,,,105.0,105.0,1,105.0  
 30,31,2010,1,2,6,4,,,124.0,124.0,1,124.0  
 31,32,2010,1,2,7,4,,,120.0,120.0,1,120.0  
 32,33,2010,1,2,8,4,,,132.0,132.0,1,132.0  
 33,34,2010,1,2,9,4,,,140.0,140.0,1,140.0  
 34,35,2010,1,2,10,4,,,152.0,152.0,1,152.0  
 35,36,2010,1,2,11,4,,,148.0,148.0,1,148.0  
 36,37,2010,1,2,12,4,,,164.0,164.0,1,164.0  
 37,38,2010,1,2,13,4,,,158.0,158.0,1,158.0  
 38,39,2010,1,2,14,4,,,154.0,154.0,1,154.0  
 39,40,2010,1,2,15,4,,,159.0,159.0,1,159.0  
 40,41,2010,1,2,16,4,,,164.0,164.0,1,164.0  
 41,42,2010,1,2,17,4,,,170.0,170.0,1,170.0  
 42,43,2010,1,2,18,4,,,149.0,149.0,1,149.0  
 43,44,2010,1,2,19,4,,,154.0,154.0,1,154.0  
 44,45,2010,1,2,20,4,,,164.0,164.0,1,164.0  
 45,46,2010,1,2,21,4,,,156.0,156.0,1,156.0  
 46,47,2010,1,2,22,4,,,126.0,126.0,1,126.0  
 47,48,2010,1,2,23,4,,,90.0,90.0,1,90.0

48,49,2010,1,3,0,4,,,63.0,63.0,1,63.0  
49,50,2010,1,3,1,4,,,65.0,65.0,1,65.0  
50,51,2010,1,3,2,4,,,55.0,55.0,1,55.0  
51,52,2010,1,3,3,4,,,65.0,65.0,1,65.0  
52,53,2010,1,3,4,4,,,83.0,83.0,1,83.0

### **out1.csv:**

year,avg

2010,104.04572982326042  
2011,99.0932403834184  
2012,90.53876763535511  
2013,98.40268334475678  
2014,93.91770864253918  
2015,85.85893721461186

### **out2.csv:**

year,month,avg

2010,1,90.40366972477064  
2010,2,97.23994038748137  
2010,3,94.04654442877292  
2010,4,80.0724233983287  
2010,5,87.0719131614654  
2010,6,109.03893805309734  
2010,7,123.4260752688172  
2010,8,97.68343195266272  
2010,9,122.79273504273505  
2010,10,118.78436657681941  
2010,11,138.38403614457832  
2010,12,97.1157469717362  
2011,1,44.87369985141159  
2011,2,150.29017857142858  
2011,3,57.99198717948718  
2011,4,91.72067039106145  
2011,5,65.10814606741573  
2011,6,108.79465541490858  
2011,7,107.38648648648649  
2011,8,103.7338003502627  
2011,9,94.96940194714882  
2011,10,145.5568181818182  
2011,11,109.43496503496503  
2011,12,108.72139973082099  
2012,1,118.92238805970149  
2012,2,84.44202898550725  
2012,3,96.47432432432433

2012,4,87.83588317107093  
2012,5,90.96671490593343  
2012,6,96.63418079096046  
2012,7,80.64970930232558  
2012,8,81.1653290529695  
2012,9,59.95224719101124  
2012,10,94.95135135135135  
2012,11,87.43696275071633  
2012,12,109.18729641693811  
2013,1,183.19533783783783  
2013,2,113.56647321428571  
2013,3,114.57275537634408  
2013,4,63.047805555555556  
2013,5,89.14853494623655  
2013,6,111.35484722222225  
2013,7,74.93284366576819  
2013,8,67.92362903225806  
2013,9,85.71787499999998  
2013,10,102.20880376344086  
2013,11,85.14629166666667  
2013,12,90.3177688172043  
2014,1,107.91174731182795  
2014,2,160.5138988095238  
2014,3,103.18325268817203  
2014,4,92.16069444444442  
2014,5,64.95850806451614  
2014,6,59.15463888888888  
2014,7,91.79994623655915  
2014,8,65.66822341857336  
2014,9,68.23266666666665  
2014,10,135.26973118279568  
2014,11,106.33752777777777  
2014,12,76.62251344086023  
2015,1,110.02276881720428  
2015,2,103.4455505952381  
2015,3,94.48344086021504  
2015,4,79.39705555555553  
2015,5,61.16751344086024  
2015,6,60.33238888888889  
2015,7,60.22952956989249  
2015,8,45.89607526881719  
2015,9,50.92475000000001  
2015,10,77.2577419354839  
2015,11,125.80318055555553

2015,12,162.17891129032256

图表:

