

作业1：通过爬虫爬取链家的新房数据，并进行预处理。

<https://bj.fang.lianjia.com/loupan/>

- 最终的csv文件，应包括以下字段：名称，地理位置（3个字段分别存储），房型（只保留最小房型），面积（按照最小值），总价（万元，整数），均价（万元，保留小数点后4位）；
- 对于所有字符串字段，要求去掉所有的前后空格；
- 如果有缺失数据，不用填充。
- 找出总价最贵和最便宜的房子，以及总价的中位数
- 找出单价最贵和最便宜的房子，以及单价的中位数





公园十七区

住宅

在售

关注

顺义 / 后沙峪 / 中央别墅区火沙路和裕庆路交汇口北500米

3室 / 4室

建面 90-144m²

55711 元/平(均价)

总价475万/套起

免费专车

明星户型

限竞房

多轨交汇

三甲医院



V7九间堂

别墅

在售

通州 / 潞苑 / 通燕高速耿庄桥北出口中化石油对面

4室 / 5室 / 6室 / 8室

建面 220-420m²

1700 万/套(总价)

总价1700万/套起

私属庭院

入户花园

环线房

近主干道

名称，地理位置（3个字段分别存储），房型（只保留最小房型），面积（按照最小值，整数），均价（元，整数），总价（万元，保留小数点后4位），结果示例如下：

公园十七区，顺义，后沙峪，中央别墅区火沙路和裕庆路交汇口北500米，3室，90，55711，501.3990。注：最后一项501.3990由面积乘以单价计算得出。

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，220，45455，1700.0000。注：45455由1700万除以面积220得出，只保留整数。



作业2：计算北京空气质量数据

1. 汇总计算PM指数年平均值的变化情况

2. 汇总计算10-15年PM指数和温度月平均数据的变化情况

- No: 行号
- year: 年份
- month: 月份
- day: 日期
- hour: 小时
- season: 季节
- PM: PM2.5浓度 ($\mu\text{g}/\text{m}^3$)
- DEWP: 露点 (摄氏温度) 指在固定气压之下, 空气中所含的气态水达到饱和而凝结成液态水所需要降至的温度。
- TEMP: Temperature (摄氏温度)
- HUMI: 湿度 (%)
- PRES: 气压 (hPa)
- cbwd: 组合风向
- lws: 累计风速 (m/s)
- precipitation: 降水量/时 (mm)
- lprec: 累计降水量 (mm)



作业2：雾霾指数数据分析

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
No	year	month	day	hour	season	PM_Taiyua	PM_US Po	PM_Xiaoh	DEWP	HUMI	PRES	TEMP	cbwd	lws	precipitatio	lprec
1	2010	1	1	0	4	NA	NA	NA	-26	69.79	1024	-22	NE	1.0289	NA	NA
2	2010	1	1	1	4	NA	NA	NA	-26	76.26	1024	-23	NE	2.5722	NA	NA
3	2010	1	1	2	4	NA	NA	NA	-27	69.56	1023	-23	NE	5.1444	NA	NA
4	2010	1	1	3	4	NA	NA	NA	-27	69.56	1023	-23	NE	7.7166	NA	NA
5	2010	1	1	4	4	NA	NA	NA	-27	69.56	1022	-23	NE	9.7744	NA	NA
6	2010	1	1	5	4	NA	NA	NA	-26	76.26	1022	-23	NE	11.8322	NA	NA
7	2010	1	1	6	4	NA	NA	NA	-25	76.46	1021	-22	NE	14.4044	NA	NA
8	2010	1	1	7	4	NA	NA	NA	-24	70.26	1021	-20	NE	16.9766	NA	NA
9	2010	1	1	8	4	NA	NA	NA	-23	70.49	1021	-19	NE	19.0344	NA	NA
10	2010	1	1	9	4	NA	NA	NA	-22	70.71	1021	-18	NE	21.6066	NA	NA
11	2010	1	1	10	4	NA	NA	NA	-20	77.39	1022	-17	NE	24.1788	NA	NA
12	2010	1	1	11	4	NA	NA	NA	-18	77.75	1021	-15	NE	27.2655	NA	NA
13	2010	1	1	12	4	NA	NA	NA	-17	77.92	1020	-14	NE	29.8377	NA	NA
14	2010	1	1	13	4	NA	NA	NA	-16	78.1	1019	-13	NE	32.9244	NA	NA
15	2010	1	1	14	4	NA	NA	NA	-15	84.87	1018	-12	NE	35.4066	NA	NA

- No: 行号
- year: 年份
- month: 月份
- day: 日期
- hour: 小时
- season: 季节
- PM: PM2.5浓度 (ug/m^3)
- DEWP: 露点 (摄氏温度) 指在固定气压之下，空气中所含的气态水达到饱和而凝结成液态水所需要降至的温度。
- TEMP: Temperature (摄氏温度)
- HUMI: 湿度 (%)
- PRES: 气压 (hPa)
- cbwd: 组合风向
- lws: 累计风速 (m/s)
- precipitation: 降水量/时 (mm)
- lprec: 累计降水量 (mm)



以上作业以报告形式提交，需要将核心代码贴在报告中，并在报告中给出最终的csv内容（截取前50条数据即可），以及分析结论。

文件名为学号，文件格式为pdf，提交平台为爱课堂。

作业提交截止时间：2021.12.6 24点

