

数据分析修炼手册

目录

数据分析修炼手册	1
前言	1
数据分析师如何分类?	2
数据分析师的具体工作职责和工作内容有哪些?	3
如何在业余时间成为数据分析师?	5
数据分析实战与运用	8
如何用 Excel 做数据分析?	9
如何用 Tableau 做数据可视化?	15

前言

知乎：路人甲

微博：玩数据的路人甲

微信公众号：一个程序员的日常

在知乎分享已经有一年多了，之前一直有朋友说我的回答能整理成书籍了，一直偷懒没做，最近有空仔细整理了知乎上的回答和文章另外也添加了一些新的内容，完成了几本小小的电子书，这一本是有关于[数据分析](#)方面的。

在此之前已经在公众号更新过 [Python](#) 和 [Python 爬虫总结](#) 电子书，这本书一共被浏览了二十多万次，下载了几万次，也收到反馈很多人通过这本电子书开启了 Python 或者 Python 爬虫之路，也有不少小伙伴通过这本电子书找到了一些心仪的工作。

还有另外几本包括我的一些数据分析方面的读书笔记、增长黑客的读书笔记、机器学习十大算法等等内容。将会在我的微信公众号：[一个程序员的日常](#)进行更新，同时也可以关注我的知乎账号：[路人甲](#) 及时关注我的最新分享用数据讲故事。

在公众号后台回复关键词：

「1」：可以得到 Python 和 Python 爬虫电子书

「2」：可以得到数据分析电子书

数据分析师如何分类？

入行之后，我才发现数据分析其实可以分为两种：一种类似产品经理、一种偏向数据挖掘，类似产品经理向更加注重业务，对业务能力要求比较高；数据挖掘向更加注重技术，对算法代码能力要求比较高。

首先我说说这两种方向共同需要的技术面，当然以下只是按照数据分析入门的标准来写：

1. SQL（数据库），我们都知道数据分析师每天都会处理海量的数据，这些数据来源于数据库，那么怎么从数据库取数据？如何建立两表、三表之间的关系？怎么取到自己想要的特定的数据？等等这些数据选择问题就是你首要考虑的问题，而这些问题都是通过 SQL 解决的，所以 SQL 是数据分析的最基础的技能，零基础学习 SQL 可以阅读这里：[SQL 教程 w3cschool](#)
2. 统计学基础，数据分析的前提要对数据有感知，数据如何收集？数据整体分布是怎样的？如果有时间维度的话随着时间的变化是怎样的？数据的平均值是什么？数据的最大值最小值指什么？数据相关与回归、时间序列分析和预测等等，这些在网易公开课上倒是有不错的教程：[哈里斯堡社区大学公开课：统计学入门 全 24 集 网易公开课](#)
3. Python 或者 R 的基础，这一点是必备项也是加分项，在数据挖掘方向是必备项，语言相比较工具更加灵活也更加实用。至于学习资料：R 语言我不太清楚，Python 方向可以在廖雪峰廖老师的博客里看 Python 教程，面向零基础。

再说说两者有区别的技能树：

1. 数据挖掘向

我先打个前哨，想要在一两个月内快速成为数据挖掘向的数据分析师基本不可能，做数据挖掘必须要底子深基础牢，编程语言基础、算法、数据结构、统计学知识样样不能少，而这些不是你自习一两个月就能完全掌握的。

所以想做数据挖掘方向的，一定要花时间把软件工程专业学习的计算机基础课程看完，这些课程包括：数据结构、算法，可以在这里一探究竟：[如何学习数据结构？](#)

在此之后你可以动手用 Python 去尝试实现数据挖掘的十八大算法：[数据挖掘 18 大算法实现以及其他相关经典 DM 算法](#)

2. 产品经理向

产品经理向需要你对业务感知能力强，对数据十分敏感，掌握常用的一些业务分析模型套路，企业经常招聘的岗位是：商业分析、数据运营、用户研究、策略分析等等。这方面的学习书籍就很多，看得越多掌握的方法越多，我说几本我看过的或者很多人推荐的书籍：《增长黑客》、《网站分析实战》、《精益数据分析》、《深入浅出数据分析》、《啤酒与尿布》、《数据之魅》、《Storytelling with Data》

总之一百个人眼中有一百个哈姆雷特，一百个数据分析师对快速入门有一百种方法，但是万变不离其中，以上是我的愚见，希望能对你有所启发。

因为我本人是属于商业数据分析师的角色，所以在接下来的讨论中主要围绕产品经理向的数据分析师。

数据分析师的具体工作职责和工作内容有哪些？

有关于这一块，我现在给大家贴几个链接。

- 1、<https://www.zhihu.com/question/20129061>「数据分析师的具体工作职责和工作内容有哪些？」
- 2、<http://www.pmcaff.com/discuss/index/315805294081088>「数据分析师的日常工作有哪些？如何入门数据分析？」

以下是我作为一个商业数据分析师自己的理解，我把数据分析师的工作日常总结为下面几个方面。

「1」产生数据

我这里所说的产生数据，并不是说去做开发或是做数据采集。我这里所说的数据是每当业务上有新的功能点需要开发上线的时候，数据分析师需要去围绕着这些功能会产生哪些业务变化、这个功能上线的目的是什么、上线之后该如何衡量效果等一系列问题，在功能上线前做好数据的埋点、以及可以衡量最终效果的指标。这样当功能上线的时候，你可以快速的衡量业务效果。

「2」提供数据

第一步的数据产生有方便了第二步的数据提供。提供数据可能是作为一个数据分析师每天都要做的事情，甚至有时候大半天都在做这件事情。数据需求的来源是多方的，各种业务方以及产品经理。商业数据分析师是公司业务方面的数据的出口，为了能准确的给需求方提供数据，你需要跟业务方有充分的沟通，对公司的数据维度有详细的了解。

「3」解释数据

作为数据分析师，是不能止步于提供数据的。提供数据之余，你也要会解释数据，这些数据是怎么来的？好的数据坏的数据都需要去知道，这样才能取长补短。

- 春节期间用户购买 vip 的数量翻了一倍，为什么翻倍了？
- 上周的用户留存降低了几个百分点，为什么降低？
- 知乎的最近一周的活跃用户答题量下滑几个百分点？

等等这些问题，都需要你一步一步的深入挖掘，而这背后的真正原因能够快速的做出预警或者给之后的发展一些很好的 idea.

「4」探索数据

光有解释数据是不够的，因为数据分析并不是解决能看出来的问题，还要能提出发现解决一些探索性的问题。

- 给新用户怎样的激励，才能让他们的次日留存达到最佳？比方说 **Linkedin** 探索出来的让新用户拥有六个及以上好友可以很好提高他们的留存。

探索数据通常是一个长期的比较大的项目，探索数据并不存在一个标准的答案，也通常可能是几个月出不了一个好的结论。

「5」影响数据

A/B Test 是为了探索更好的方向，更受用户喜欢的功能。通过数据论证可以说服业务方听询数据分析师的意见。推送的消息文本应该怎么发？发给几个实验组看看效果；这个按钮应该设计

成什么颜色，配上什么文字？多设计几套上线几个不同分组看效果。通过不断的 A/B Test，数据分析师会更好的辅助产品的迭代，影响数据的产生。

如何在业余时间成为数据分析师？

以下方法仅适用开发出身或有语言基础的同学。

我的情况跟题主是一样的，软件工程出身，在之前一直在做 Java/Python 开发，无意之中接触到了爬虫，接触了一堆一堆的数据，之后慢慢开始对数据感兴趣，开始学习一些数据分析的技能：凭着爬虫以及数据分析的技能我找到了理想中的数据分析岗位。所以我对自己的评价一直是：**科班出身的软件工程师、野路子的数据分析师。**

从 Java/Python 开发到 Python 爬虫到数据分析，有很多人是看着我一步一步走过来的，我想在这里根据我自己这段野路子的学习经历说一说非科班如何利用业余时间学数据分析，当然只是数据分析，不是数据挖掘。

数据分析师的能力分为：**分析能力和业务能力**，**分析能力决定一个数据分析师的下限，业务能力决定一个数据分析师的上限。**

因为是开发出身，我的方法与其他人的不同，

我把数据分析的学习分为两个大的阶段，：

学习的第一阶段：这一阶段是为了培养自己做分析的能力

第一阶段的基础是要有数据

分析能力相对来说是死板的，通常你需要学会掌握下面这些技能：

- 数据分析常用哪些算法？
- 需要用哪些库？
- 如何进行计算？
- 进行可视化，每一种图表的用途是什么？

为了解决上面抛出的问题，你可以仔细学习阅读下面的书籍：

1. [利用 Python 进行数据分析 \(豆瓣\)](#)
2. [SciPy and NumPy \(豆瓣\)](#)

当然，如果你不想用 Python 用 R 去搞这么复杂的计算，那你也可以直接把清理好的数据扔在一些做计算和可视化的工具里，我比较熟悉的就是：Tableau，以及据说不错的 Power BI.当然最后你发现还是学习 Python 或者 R 的数据分析库来的实在，因为他们更加灵活。

在第一阶段用工具结合 Python 我留下了一些文章，可以作为参考，这些都是在爬虫抓了很多数据之后，先有数据后结合教程书籍分析学习产生的内容。

1. [链家网的租房数据做些有意思的事情？](#)
2. [豆瓣读书分析报告](#)
3. [爬取 6.6w+豆瓣电影之后的分析故事](#)

学习的第二阶段：这一阶段是为了培养自己做业务的能力

在第一阶段的学习你已经学会了：

- 对数据的整体感知
- 知道什么时候该用什么图表
- 掌握了一些基本的算法和分析库

这一阶段，技术不再是你的障碍，**重点是要有 IDEA**

这一阶段数据分析的步骤通常为：

- 我想要做什么？分析什么？
- 是否有数据能提供支持？把你想做的抽象成数据。
- 数据获取/数据清洗
- 开始进行分析（语言、工具）

我们能看到这四个步骤后两个步骤是在第一阶段学习的内容，所以我们只要把精力聚焦在前两个步骤就可以了，那么怎样培养自己数据分析的 IDEA，我的方法是：多学多看多做。

多看：看一些实战 IDEA 较多的书籍

1. [数据之美 \(豆瓣\)](#)
2. [啤酒与尿布 \(豆瓣\)](#)
3. [深入浅出数据分析 \(豆瓣\)](#)

多做：

网络上有很多数据分析的案例，很简单易懂，在深层的技术点也不难，你完全可以按照他的思路重新做一遍，就变成了你的思路了。

数据冰山的数据分析文章：

1. [生活中的数据犀利哥之一：开篇及选车](#)
2. [生活中的数据犀利哥之二：排队](#)
3. [生活中的数据犀利哥之三：找座](#)
4. [生活中的数据犀利哥之四：看趋势](#)
5. [生活中的数据犀利哥之五：找错因果关系](#)
6. [生活中的数据犀利哥之六：快递员送货](#)
7. [北上广深哪里过得更潇洒](#)
8. [沧海横流，看行业起伏（2015 年）](#)
9. [和颐事件之后：怎么定酒店更有安全感？](#)
10. [老鹿玩数据——不光是求婚神器（一）](#)
11. [老鹿玩数据——不光是求婚神器（二）](#)
12. [《春节自救指南》之数据分析](#)

还有我曾经写的几篇文章：

1. [如何判断一场知乎 live 的质量？](#)
2. [豆瓣 5.6 分的《西游伏妖篇》有水军吗？](#)

团支书的一些回答和文章：

1. [怎样才能以最少的钱租个靠近地铁的房子？](#)
2. [「上海富了周围，北京坑了周围」？ - 知乎](#)
3. [如何不吹牛地形容北京有多大？ - 知乎](#)

等等几位的文章都可以用来学习参考。

通过以上数据分析算是入门了，至于如何进阶，我也在摸索中，希望有机会我可以更新到这篇回答里，也欢迎在学习数据分析的同学一起交流。

数据分析实战与运用

自从我的知乎读者到达十万之后，我经常会受到各种私信问我为什么能有这么多读者，有什么小技巧可以分享，我经常给到的回复是：**数据分析**。我的知乎能在一年多时间里从零到十五万是我一步一步分析到实践的结果，那么简单来说一说我从哪些方面做的。

1.经常分析研究精华回答

精华回答就是知乎社区或者话题下赞同数量最多的回答，答案赞同越多答主曝光率越高，那么被关注的机会就越大，那么如何写一个高赞精华回答就很重要，如何写高赞的回答？把擅长的话题或者知乎的根话题下所有精华回答拉出来，拿一个笔记本对每一个回答做归类做分析，学习高赞回答的文风、思路，对我来说收获非常大。

当然，并不是所有的高赞回答都会增粉的，那么这些就是需要结合具体的回答以及回答者的涨粉情况去慢慢分析琢磨了。

2.经常分析研究涨粉快的用户

有很多小伙伴都知道，从今年年初开始，我每个月会固定抓一次知乎粉丝过万的所有用户，并且做用户的粉丝增长数量和速度排名，这些有什么用？可以拎出每个月粉丝增长数量前十或者二十的用户进行分析，他们为什么涨粉？有什么我可以借鉴的地方，当然有很多东西是不可复制的，也是不能复制的。

3.什么时间发什么内容？

在第一点的时候我说到高赞回答是增粉最快的手段了，就像微信公众号一样不同的时间发文章阅读量会不一样，因为这和平台用户的活跃时间有关。

通过对知乎的数据的监控分析，我知道了大部分知乎用户在平常的时候每天的早晨九点到十点半会活跃并且容易点赞，在晚上的九点之后容易活跃并且点赞，在周末的时候用户容易在下午的三到四点点赞，当然远远不止这些，对此我以前做过一点解释：[路人甲：为什么知乎的众多答主答题后获得的赞是在「一觉睡醒」后才发现的呢？](#)

通过不断地测试分析我发现在计算机领域，早晨八点钟发回答或者文章一天下来点赞数量比其他时间段发表好，在晚上的时候发表娱乐相关的内容效果比较好。同样我发现我的读者特别喜欢在我早晨发表回答的时候点赞，下午的时候我的重度读者几乎不在刷知乎。

数据分析对我的帮助远远不止这些，更多的时候她能帮助我看清很多东西，包括但不限于：

- 1.通过数据我能发现电影是否有水军：[豆瓣 5.6 分的《西游伏妖篇》评论有水军吗？](#)
- 2.通过数据我能判断知乎 Live 质量：[不交智商税，如何判断一场知乎 live 的质量？](#)
- 3.通过数据帮助业务人员发现一些他们发现不了的东西：[大数据对物流管理有什么影响？](#)
- 4.克服我的选择困难症，用大众点评数据帮我选合适的餐厅：[约会聚餐如何正确选择餐馆？](#)
- 5.用新浪微博数据揭露华为海军的真面目：[路人甲：如何评论华为官方对闪存事件的回复](#)

如何用 EXCEL 做数据分析？

其实在正式开始工作之前我是很排斥用 Excel 做分析的，那时候我用 Tableau 和 Python，有时候甚至做个平均值也扔到 Tableau 算一下。我一直认为 Python 和 Tableau 运用比较自由，而 Excel 在我的心目中一直是比较刻板的，我觉得这个印象可能跟初高中的时候的计算机课程有关。

不过直到工作之后，我才发现 Excel 其实是大家用着最多的，业务人员会用 Excel、财务人员会用 Excel，总之人人都会使用 Excel，而作为数据分析师会用 Excel 做分析是与业务人沟通的一座桥梁。

当然，当我开始使用 Excel 做分析之后我发现 Excel 还有如下的优点：

1. 轻便，不需要多余的工具或语言环境
2. 可视化非常方便
3. 简单易上手，对于非技术是个很好的选择
4. 学习成本低、学习速度快

我这里用 Excel 做的一个很简单的统计和分析的例子：我这边有一份数据是来自安居客网站的二手房的一些数据。我们接到来自上级的任务：我们需要在上海的高档小区中做一些产品的推广，请给我找到这些高档社区。

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
房屋编码	发布时间	小区	年代	类型	面积	朝向	楼层	装修程度	房屋单价	参考首付	参考总价	经度	纬度	小区物业费总户数	容积率	绿化率	
5.82E+08	2016年8月24日	鸿丰香榭里		2006 公寓	399.1	南	低层(共8层)	简单装修	132805	1590	5300	121.4409	31.2142	12	35	2 50%	
5.82E+08	2016年8月24日	鸿丰香榭里		2006 公寓	399.1	南	低层(共8层)	简单装修	132805	1590	5300	121.4409	31.2142	12	35	2 50%	
5.73E+08	2016年8月13日	嘉里华庭		2005 公寓	175.9	南	中层(共38层)	精装修	112563	594	1980	121.4432	31.21765	5.5	534	5 53%	
5.82E+08	2016年8月24日	皓源苑		2010 公寓	205	南北	低层(共6层)	豪华装修	107317	660	2200	121.4337	31.20958	未知	62	暂无	0%
5.82E+08	2016年8月24日	嘉里华庭		2004 公寓	175	南	低层(共38层)	精装修	100571	528	1760	121.4432	31.21765	未知	534	5	53%
5.8E+08	2016年8月22日	淮海名邸		2016 普通住宅	234	南北	中层(共9层)	豪华装修	98290	690	2300	121.4311	31.20713	未知		暂无	0%
5.81E+08	2016年8月23日	嘉里华庭		2002 公寓	146	南	中层(共38层)	简单装修	97945	429	1430	121.4432	31.21765	5.5	534	5	53%
5.73E+08	2016年8月13日	嘉里华庭		2005 公寓	153.9	南	高层(共38层)	精装修	96567	464	1480	121.4432	31.21765	5.5	534	5	53%
5.81E+08	2016年8月24日	延安西路4		1990 公寓	55.6	南北	低层(共6层)	精装修	93115	155.4	518	121.4438	31.22224	未知	112	2.5	22%
5.81E+08	2016年8月24日	延安西路4		1990 公寓	55.6	南北	低层(共6层)	精装修	93115	155.4	518	121.4438	31.22224	未知	112	2.5	22%
5.77E+08	2016年8月19日	金色贝拉维		2007 公寓	150	南北	中层(共20层)	精装修	90000	420	1350	121.4096	31.19929	未知	700	2.7	50%
5.78E+08	2016年8月19日	东方剑桥		2001 公寓	147	西南	高层(共36层)	精装修	88435	390	1300	121.4445	31.22204	3.2	483	4.6	40%
5.78E+08	2016年8月19日	东方剑桥		2001 公寓	147	西南	高层(共36层)	精装修	88435	390	1300	121.4445	31.22204	3.2	483	4.6	40%
5.77E+08	2016年8月19日	金色贝拉维		2008 公寓	150	南北	高层(共19层)	豪华装修	86656	390	1300	121.4096	31.19929	未知	700	2.7	50%
5.77E+08	2016年8月19日	金色贝拉维		2008 公寓	150	南北	高层(共19层)	豪华装修	86656	390	1300	121.4096	31.19929	未知	700	2.7	50%
5.77E+08	2016年8月19日	联泰和城		2008 公寓	68.9	南	高层(共8层)	精装修	85631	177	590	121.4161	31.2072	1.9	56	1.9	40%
5.81E+08	2016年8月23日	古北嘉年华		2004 公寓	171	南北	中层(共35层)	简单装修	84795	435	1450	121.4149	31.20039	2.2	700	1.9	52%
5.75E+08	2016年8月16日	天山华庭		2010 公寓	100	南北	低层(共15层)	精装修	82000	246	820	121.3933	31.22127	1.98元/平	540	2.54	40%
5.81E+08	2016年8月23日	红德苑		2000 公寓	137	南北	中层(共11层)	精装修	78832	324	1080	121.4193	31.22578	未知	280	1.7	33%
5.81E+08	2016年8月23日	东方雅苑		2000 公寓	149	南	中层(共31层)	豪华装修	77181	345	1150	0	0	未知	248	暂无	35%
5.81E+08	2016年8月23日	东方雅苑		2000 公寓	149	南	中层(共31层)	豪华装修	77181	345	1150	0	0	未知	248	暂无	35%
5.81E+08	2016年8月24日	新海小区		1977 老公房	43	南	共5层	精装修	76744	99	330	121.4413	31.22068	未知	128	2.8	19%
5.81E+08	2016年8月24日	新海小区		1977 老公房	43	南	共5层	精装修	76744	99	330	121.4413	31.22068	未知	128	2.8	19%
5.81E+08	2016年8月24日	新海小区		1977 老公房	43	南	共5层	精装修	76744	99	330	121.4413	31.22068	未知	128	2.8	19%
5.82E+08	2016年8月25日	四季康园		2003 公寓	141	南北	低层(共18层)	简单装修	75595	324	1080	121.4013	31.19855	3	288	2	40%
5.82E+08	2016年8月25日	四季康园		2003 公寓	141	南北	低层(共18层)	简单装修	75595	324	1080	121.4013	31.19855	3	288	2	40%
5.78E+08	2016年8月20日	凤城北苑		2000 公寓	128.1	南	低层(共18层)	精装修	76478	294	980	121.407	31.20371	未知	370	1.6	35%
5.8E+08	2016年8月22日	华山花园		1996 公寓	131	南	中层(共27层)	豪华装修	76335	300	1000	121.4409	31.21824	8	112	2.1	30%
5.82E+08	2016年8月25日	新天地河滨		2011 公寓	145	南北	中层(共33层)	精装修	75862	330	1100	121.4017	31.22276	未知	953	2.5	45%
5.77E+08	2016年8月18日	风云汇恒里		2004 公寓	101	南	中层(共24层)	精装修	75247	228	760	121.4292	31.21461	未知	125	2	34%
5.82E+08	2016年8月25日	新华名门		1999 公寓	109.5	南北	高层(共6层)	简单装修	74899	246	820	121.4346	31.20860	1.5	185	1.5	30%
5.81E+08	2016年8月23日	武夷花园		2002 普通住宅	103	南北	高层(共20层)	精装修	74757	231	770	121.4222	31.21767	2	390	3.1	46%
5.76E+08	2016年8月18日	虹桥新城		2004 公寓	118	南北	中层(共11层)	精装修	72711	257.4	858	121.4128	31.22256	1.5	1185	2.1	41%
5.76E+08	2016年8月18日	虹桥新城		2004 公寓	118	南北	中层(共11层)	精装修	72711	257.4	858	121.4128	31.22256	1.5	1185	2.1	41%
5.76E+08	2016年8月18日	虹桥新城		2004 公寓	118	南北	中层(共11层)	精装修	72711	257.4	858	121.4128	31.22256	1.5	1185	2.1	41%
5.76E+08	2016年8月18日	虹桥新城		2004 公寓	118	南北	中层(共11层)	精装修	72711	257.4	858	121.4128	31.22256	1.5	1185	2.1	41%
5.8E+08	2016年8月21日	江苏路186		1998 公寓	35.1	南	中层(共7层)	精装修	72649	76.5	255	121.4364	31.22815	未知	120	1.9	13%

接到任务之后，我们对照着数据做一个拆解，高档社区应该有什么特点呢？房屋单价高是一个很重要的指标，小区的绿化率也可以做一定的参考，装修的一定要不错。

1、所以我们先筛选出精装修和豪华装修的二手房。

AZ

排序

漏斗

筛选

清除

高级

分列

删除

重复项

数据验证

合并

计算

模拟分析

组合

取消组合

小计

F	G	H	I	J	K	L	M	N
面积	朝向	楼层	装修程度	房屋单价	参考首付	参考总价	经度	纬度
399.1	南	低层(共8层)	简单装修					
399.1	南	低层(共8层)	简单装修					
175.9	南	中层(共38层)	精装修					
205	南北	低层(共6层)	豪华装修					
175	南	低层(共38层)	精装修					
234	南北	中层(共9层)	豪华装修					
146	南	中层(共38层)	简单装修					
153.3	南	高层(共38层)	精装修					
55.6	南北	低层(共6层)	精装修					
55.6	南北	低层(共6层)	精装修					
150	南北	中层(共20层)	精装修					
147	西南	高层(共36层)	精装修					
147	西南	高层(共36层)	精装修					
150	南北	高层(共19层)	豪华装修					
150	南北	高层(共19层)	豪华装修					
68.9	南	高层(共8层)	精装修					
171	南北	中层(共35层)	简单装修					
100	南北	低层(共15层)	精装修					
137	南北	中层(共7层)	精装修					
149	南	中层(共31层)	豪华装修					
149	南	中层(共31层)	豪华装修					
43	南	共5层	精装修					
43	南	共5层	精装修					
43	南	共5层	精装修					
141	南北	低层(共18层)	简单装修					
141	南北	低层(共18层)	简单装修					
128.1	南	低层(共18层)	精装修					
131	南	中层(共27层)	豪华装修					

排序

A↑ 升序

Z↓ 降序

按颜色: 无

筛选器

按颜色: 无

选择一个

搜索

☒ (全选)

☒ 豪华装修

☒ 简单装修

☒ 精装修

☒ 毛坯

☒ 中等装修

清除筛选

76335	300	1000	121.4409	31.2182
-------	-----	------	----------	---------

2、我们想在以上的已经筛选出的数据里面找到平均房屋单价最高的小区。于是很多眼疾手快的小伙伴直接选择房屋单价这一列进行降序排序。如果你这么做确实可以找出二手房价格最高的房子所属小区。但是注意审题，我们是要找出的单价最高的小区。

排序警告

当前所选内容以外的数据不会进行排序。

您要做什么操作？

☒ 扩展选定区域

☐ 以当前选定区域排序

取消 排序

房屋单价	参考首付
132805	159
132805	159
112563	59
107317	66
100571	52
98290	69
97945	42
96567	44
93115	155
93115	155
90000	42
88435	39
88435	39
86666	39
86666	39
85631	17
84795	43

3、所以这个时候我们应该按照小区给一个小区的房屋单价，这时候就要用到我们的强大的数据透视表了。选择我们要透视的数据，然后点击入中的数据透视表。

创建数据透视表

请选择要分析的数据。

☒ 选择一个表或区域

表/区域: 第一节课-排序!\$A\$1:\$R\$538

☐ 使用外部数据源

选择要放置数据透视表的位置。

☒ 新工作表

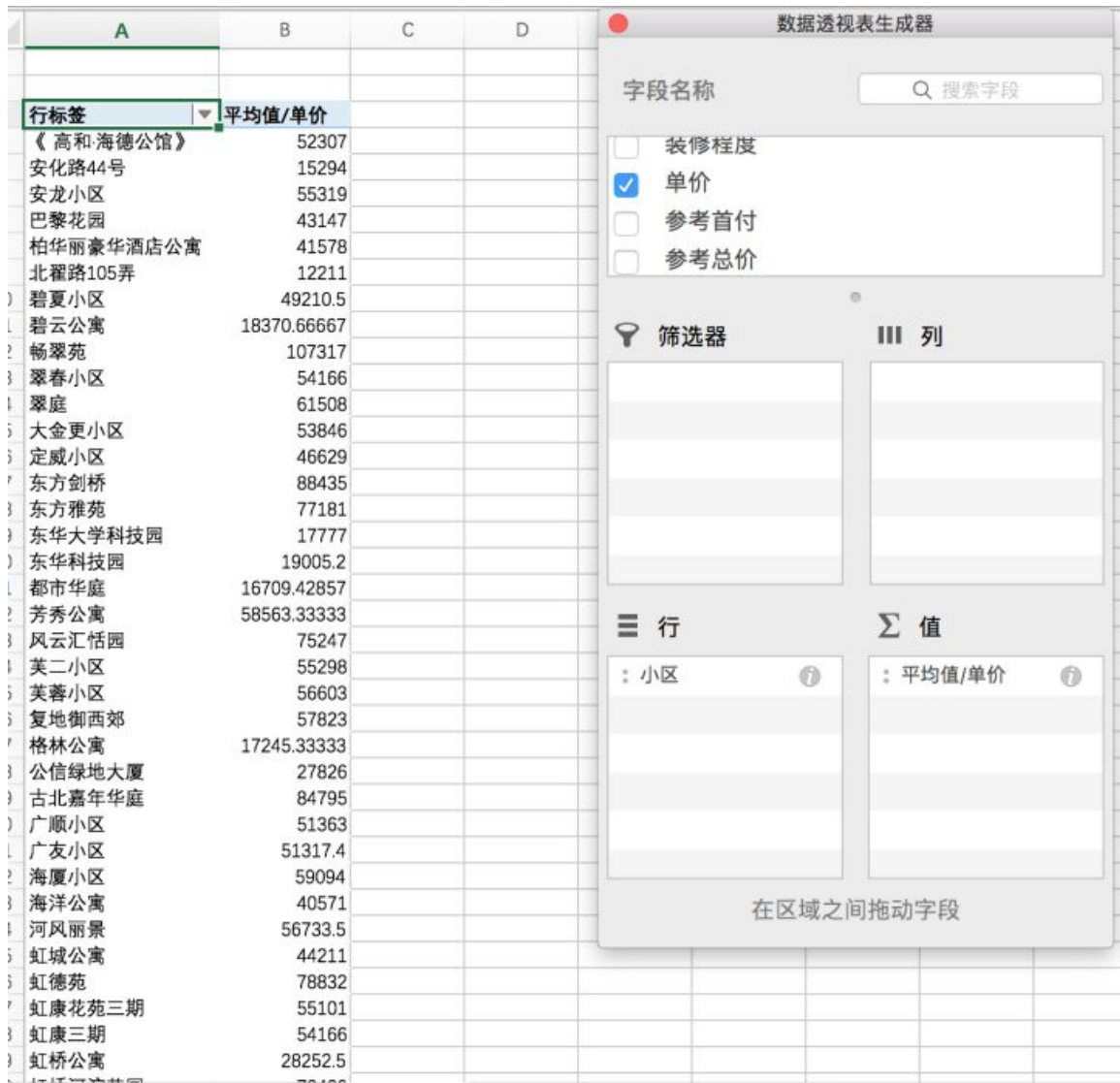
☐ 现有工作表

表/区域: 第一节课-排序!\$A\$1:\$R\$538

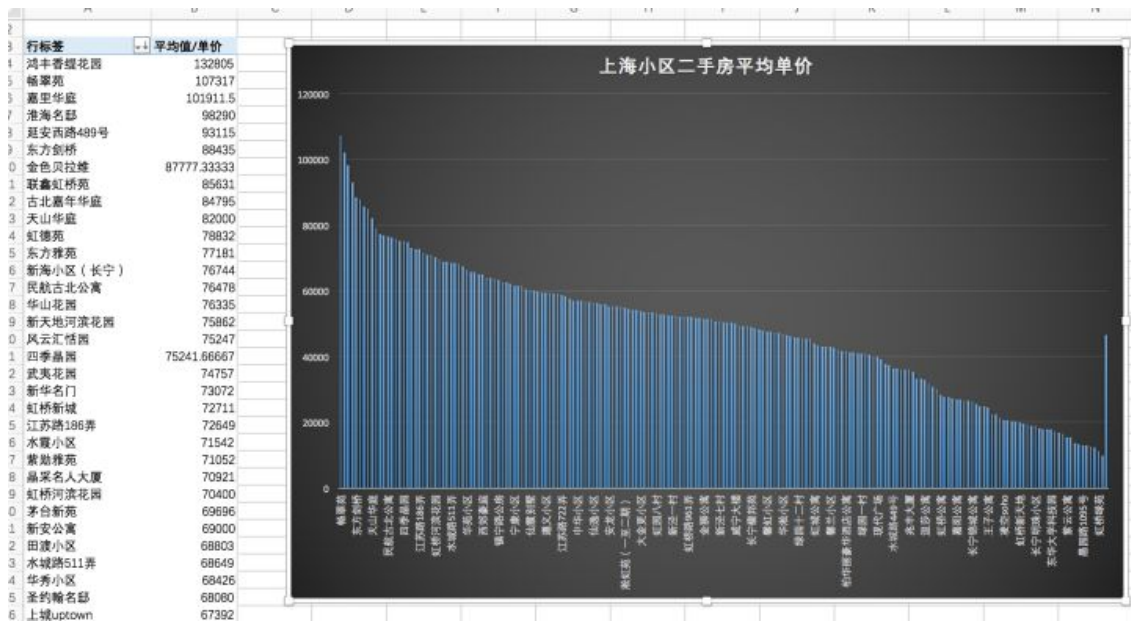
取消 确定

房屋单价	参考首付
132805	159
132805	159
112563	59
107317	66
100571	52
98290	69
97945	42
96567	44
93115	155
93115	155
90000	42
88435	39
88435	39
86666	39
86666	39
85631	17
84795	43

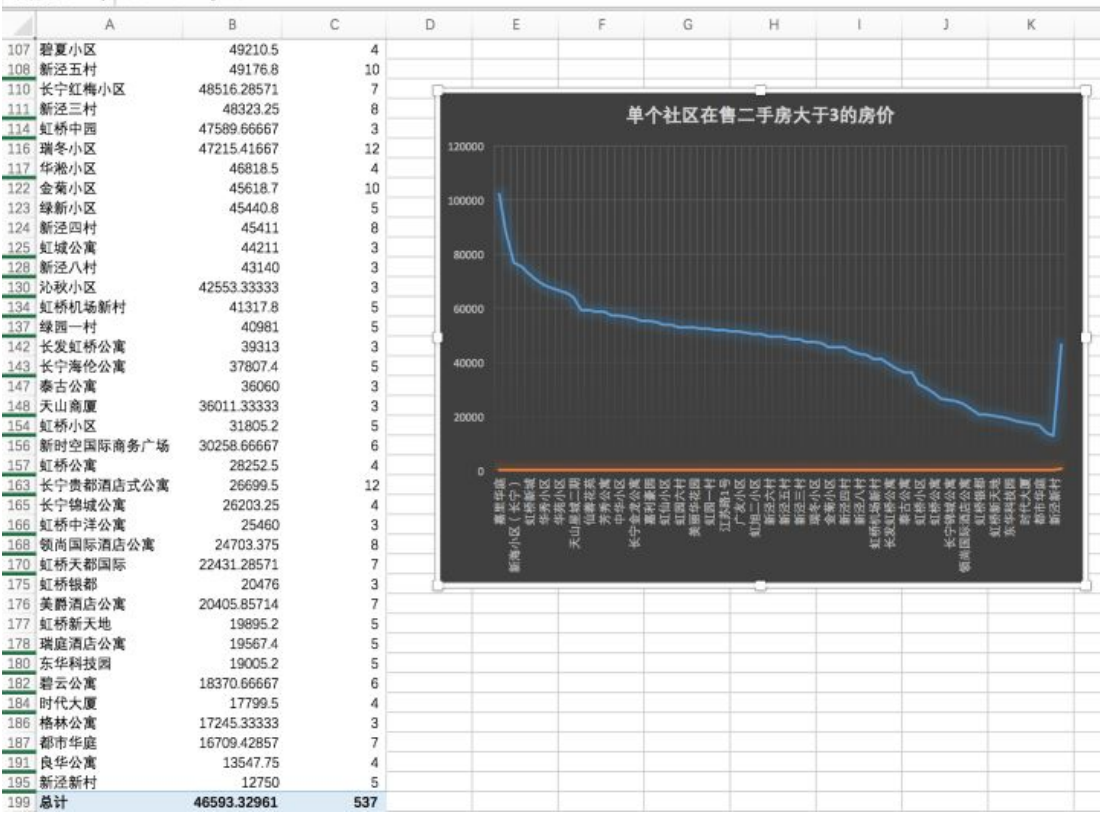
4、选择行和列，并且选择我们要聚合的方式，注意我们这里是按照小区的二手房平均单价进行聚合，所以如下操作。



最后先按照二手房平均单价进行排序，点击插入中的图表，选择柱状图。



然后大家就开开心心的把这个图表交给了上级，但是我们忽略了一个问题，因为我们采集的数据并不是全量数据，有的小区可能只有一个二手房在出售如果这个单个房子房价很高这样会影响我们的分析结果，所以我们再次对每个小区的房子进行筛选，筛选有三套及以上二手房在售的小区。



以上只是一个简单的例子。总之，我认为 Excel 是一款小而强大的分析工具。但也仅仅是工具而已，和众多的数据分析工具一样，它能帮助我们实现我们的分析思路 and 目的，但是它并不能告诉我们怎么去分析，做数据分析还需要我们掌握基础的统计学基础、掌握常用的数据分析的模型框架。

排名第一的回答其实已经讲了很多细致的方法论和基础知识了，但是还是缺少一些生动形象的案例是实用的商业场景：[知乎用户：怎样用 Excel 做数据分析？](#)

下面我来推荐一些课程和学习资料，帮助大家一步步学习 Excel 数据分析。课程因为会有练习和项目模拟，可以从实际的应用场景出发掌握好相关技能。看博客和书籍可以更灵活地巩固知识，建议大家结合着来。

1、[小蚊子数据分析_新浪博客](#)

我觉得转行做数据分析或者在产品经理和运营岗位的想懂一点数据分析的话都会去看这个博客，我学习 Excel 的时候第一个想到的就是这个博客，以及这个博客作者的那本畅销书《谁说菜鸟不会数据分析》，当然如果喜欢纸质书也推荐你买一本。

2、《Excel 2010 数据处理与分析实战技巧精粹》

除了上面说的那个博客之外，还有这本书籍。本书在对 Excel 技术论坛上上百万个提问的分析与提炼的基础上，汇集了用户在使用 Excel 进行数据处理与分析过程中最常见的需求，全书通过生动形象的案例来讲解 Excel 的操作，如何通过 Excel 进行分析。个人认为这本书类似于深入浅出系列，但是很遗憾深入浅出系列没有 Excel 部分。

最后，说来说去其实 Excel 也只是一个工具，数据分析师不应该拘泥于工具，而应该更加注重于实现，哪个工具或哪种方法能帮助我们在最快的时间内得出我们想要看的数据，那就毫不犹豫的选择它。

如何用 TABLEAU 做数据可视化？

排在前面的答主回答的很精彩，但是罗列了一堆教学资源往往让初学者无法下手，只是点击一个收藏按钮就没了。所以这里我推荐两种易上手的方法，如果你直接开始学习多用点时间一周内就能比较系统的学会 Tableau 基础。

我从去年开始学习 Tableau 现在在用 Tableau 辅助工作，有两种方法我觉得都可行去系统的学习 Tableau 基础，第一是《人人都是数据分析师：Tableau 应用实战》之前我们部门给公司做数据分析特训，也是送的这本书。

第二是看官网的视频，官网的视频还是比较系统的，并且比较友好的是很多视频已经被翻译成了中文，所以这些免费的官方资源就能带你入门了，甚至还为你准备了数据源。为什么要舍近求远呢？

Tableau Desktop

您是否希望通过 Tableau Desktop 处理数据？请观看以下视频获取相关指导！了解如何连接、浏览、分析、演示和共享数据。

3 部视频	入门	34 MIN
10 部视频	连接到数据	54 MIN
20 部视频	可视化分析	111 MIN
7 部视频	仪表板和故事	38 MIN
9 部视频	绘制地图	36 MIN
15 部视频	计算	60 MIN
6 部视频	为什么 Tableau 这样做？	32 MIN
9 部视频	入门指导	36 MIN

当然上面说的一些基础，如果后续还想学习一些进阶的内容，做出一些炫酷的可视化，建议可以看看官方网站：[Gallery](#)，这个网站上面官方会选出由世界各地的 Tableau 爱好者每天上传的最优秀的 tableau 作品。下载你感兴趣的作品到自己的 Tableau 上面研究优秀的作品是如何做出来的对自己的促进很大。

至于你说如何学习典型的商务分析的能力，这主要跟 Tableau 就无关了，珠吞是跟业务相结合，建议你去看看《精益数据分析》，这本书详细些了一些电影的商业模式。

PS: Tableau 是收费软件，而且收费比较高，我们公司采购 Lisense 都是一笔不小的消费。所以如果你学生的话可以在 [Tableau 学生版](#) 他们的官网下载到学生版本，上传学生证认证即可免费使用。如果你不是学生的话你也可以在 Tableau 试用期下载一个时光驻留器。