
9 月 30 日大鹏老师数据分析师（Python） 岗位职责解读+答疑

数据分析师(Python)前置课请戳：https://study.163.com/topics/python01_m/

数据分析师 (Python) 报名请戳：<https://url.163.com/ipb>

数据分析师（Python）岗位职责解读 （本科/1-3 年/10k-20k）

职位描述

- 1.负责企业业务数据的挖掘和分析，撰写分析报告；
- 2.为业务部门提供基础性的数据支撑服务，包括数据规整、清洗、自动报表提供等；
- 3.和产品团队协作，对产品进行数据分析和跟踪，促使产品得到不断的优化。

岗位职责具体描述

- 1.本科以上学历，数学类，工商管理类，计算机类专业优先，专业作为参考，具体不做限定；
- 2.1-2 年数据分析相关工作，有互联网、咨询公司等经验优先；
- 3.掌握 EXCEL VBA/SQL/spss 等基本的分析工具；
- 4.熟练掌握 Python 软件 精通使用 Python 语言实现各种统计分析和预测算法，拥有 Python 语言集成产品开发经验者优先；
- 5.良好的数据敏感度，能注意到数据的异常和错误以及挖掘到数据中隐蔽的有价值信息，具备良好的商业分析能力和输出结论的能力；
- 6.良好的团队协作能力。

大家好！我也是第一次和大家进行文字直播，中间环节限定的是大家禁言，所以有任何问题大家整理一下，最后互动环节一起聊聊。今晚的内容我会从一个岗位描述开始，延伸讲讲数据分析师“具体做什么”，“技术体系”以及“成长建议”给大家讲讲。今晚主题是“就业”，到这个层面就不一定完全是技术能力决定的，所以欢迎在最后和大家交流入职的一些问题，我们就直接开始了~

在分析之前，肯定有同学想问，数据分析师（Python）这个行业到底怎么样，从事数据分析师（Python）这个方向的职位需要掌握哪些技能？我在这里简单说明一下。

数据分析师非常要注意对数据、数据指标的解读，通过对数据的分析，进行业务监控，建立分析体系来解决商业问题，最高级别是行业未来发展的趋势分析，从而及时对战略进行不断优化。

从职位薪水来说，数据分析行业的高薪主要分布在长三角、珠三角和京津地区。北京、上海和深圳的薪水位列第一方阵，均薪在 10k+；杭州、宁波和广州位列第二方阵，均薪在 9k+；其他沿海及内陆区域中心城市，如南京、重庆、苏州、无锡等位于第三方阵，均薪在 8k 左右。2018 年呈上升趋势，发展“钱”途良好。数据分析师入门难度相比很多行业并不算高，所以就长远来看，若想在行业内保持竞争力，就要在业务方向积累经验建立起自己的壁垒，或者在技术方面有所建树，这样下来就不用太担心就业问题了。

说了这么多，我不知道大家对数据分析师（Python）这个岗位了解多少，肯定有不少同学会存在一些疑问，数据分析师（Python）到底要做哪些工作呢？它的岗位要求是什么？自己距离这些要求还差哪些等都是需要我们去学习的地方。我相信大家也浏览过很多的招聘网站，包括我们之前的体验课和直播课中也讲到了这方面的内容。我这里针对刚才的招聘信息给大家解读一下。

首先我们来看看“职位描述”，这也是大家要了解的第一个问题——“具体做什么？”

职位描述

1.负责企业业务数据的挖掘和分析，撰写分析报告；

2.为业务部门提供基础性的数据支撑服务，包括数据规整、清洗、自动报表提供等；

3.和产品团队协作，对产品进行数据分析和跟踪，促使产品得到不断的优化。

这里大家要清楚一个概念“数据 workflow”，这一点在我的第五次直播课程中花了 80 分钟给大家讲了一下（《大数据 —— 从概念到落地》）

针对现有的数据分析师的工作，我这里简单讲讲这两个主要的方向

→ ①面向研究报告，②面向产品开发

《大数据 —— 从概念到落地》直播课请戳：<https://url.163.com/v8R>

数据分析师（Python）报名请戳：<https://url.163.com/ipb>

目的针对研究报告的时候，数据工作者核心是要辅助上层决策，在这个过程中数据分析师需要对接数据部门，从数据清洗工作到数据分析，到最后总结出一个结果报告，甚至最后还要做汇报，这样的工作中流程中，分析师面临的不仅仅是数据研究的工作，还有繁杂的数据整理工作，以及最后的结果呈现 —— 数据报告的撰写。

当目的是面向产品开发时，数据工作者核心就是配合产品前端后端了，属于是中间算法层面，再做数据研究的同时也要做数据清洗，这里数据分析师省去了作报告的工作，核心就在于数据处理及分析的效率上，当然还包括了结果的准确性，同时也要了解一下产品层面的东西。且往后甚至会深入到开发层面，优化代码逻辑、处理报错等等。

所以这里我放上一张图，这也是大家未来入职后需要长期思考的，自己处理 workflow 中的什么位置，自己的上下游的工作分别是什么，自己的工作的意义是什么？



除此之外，大家不管是做研究报告还是配合产品开发，都需要去了解自己所在行业的相关知识，这个是需要长期积累的，是一个非常核心的“职场软实力”，这个也是在未来岗位中需要边工作边学习的。

接下来，我们再来看一看数据分析师的技术体系层面的东西，这里我们把“岗位具体职责”拿出来看看：

岗位职责具体描述

- 1.本科以上学历，数学类，工商管理类，计算机类专业优先，专业作为参考，具体不做限定；
- 2.1-2 年数据分析相关工作，有互联网、咨询公司等经验优先；
- 3.掌握 EXCEL VBA/SQL/spss 等基本的分析工具；
- 4.熟练掌握 Python 软件 精通使用 Python 语言实现各种统计分析和预测算法，拥有 Python 语言集成产品开发经验者优先；
- 5.良好的数据敏感度，能注意到数据的异常和错误以及挖掘到数据中隐蔽的有价值信息，具备良好的商业分析能力和输出结论的能力；
- 6.良好的团队协作能力。

这里我从第三点开始讲：“掌握 EXCEL VBA/SQL/spss 等基本的分析工具”

这里说的基本分析工具中 excel (vba 语言在这里应该是指的辅助 excel 的宏相关的技巧，而非开发) 其实算是现有业务人员的主流数据分析工具了，目前各行各业都会有数据分析，且在数据分析前期阶段，大量的时间会用于数据清洗、整合，同时在涉及到和不同人对接的时候，主要的数据对象就会是 excel；sql 是一个数据库的工具，在业务 workflow 中，数据分析师也会存在自己从数据库中提取、筛选、初步处理数据，然后 spss 则是起到了数据分析的功能，特别是一些较专业的数据统计分析（很多数据分析工作也会在 excel 中做）。

总体来说，现在大部分业务人员主要的技术平台还是 excel，既能做数据整理，也可以做一定分析，支持大量图表，最后也可以出成果报表甚至部分 BI 商业仪表盘等。

然而在座的大家如果本身就在做一些数据分析相关工作，可以回想一下自己的

工作时间主要放在哪里。其实大多数业务数据分析师，80%的工作可能放在前期数据整合、清洗阶段，而整理了很久后，最后花掉的 20%的时间做数据分析，其实这里很多时间是由于技术平台问题给浪费掉了。

而且确实以 excel 为主的分析体系会有一些缺陷，比如：

- ① 当数据量较大时，excel 处理数据效率低问题
- ② 多个数据文件批量整合、处理效率低问题
- ③ 算法局限
- ④ 图表类型限制，图表逻辑不明确
- ⑤ 由于是第三方软件，和产品开发对接问题等

所以我们再来看第四点：“熟练掌握 Python 软件，精通使用 Python 语言实现各种统计分析和预测算法，拥有 Python 语言集成产品开发经验者优先”

如果通过写代码(Python、R)来实现数据研究 ,以上问题都会迎刃而解 ,Python 是一个简单易学，但已经非常成熟的语言，不仅仅可以快速做数据清洗，也支持多种计算方法。python 就好比你的另一只“手”，你用它去帮你完成你想做的事儿：



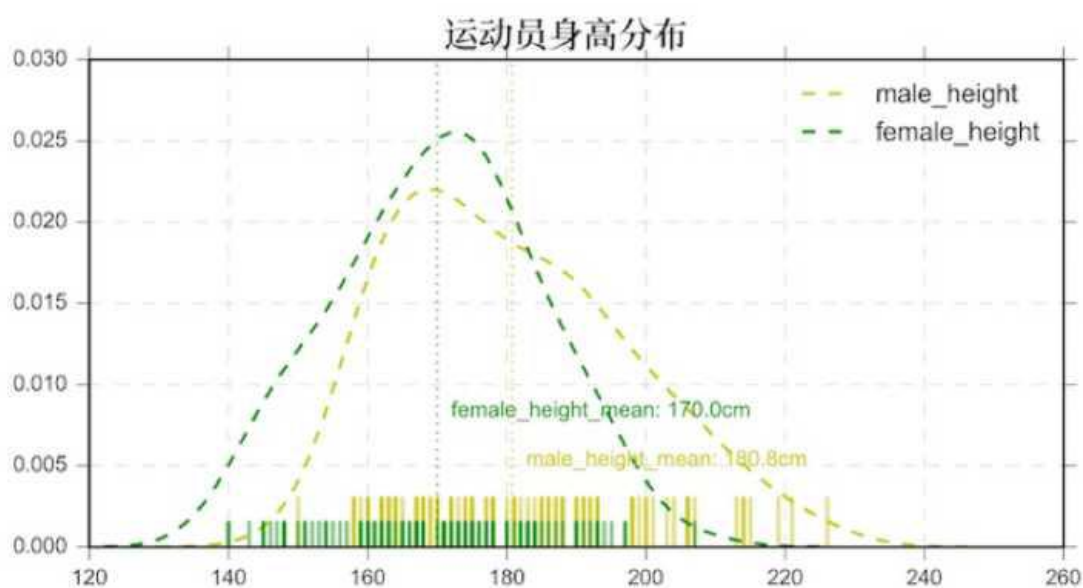
这是一个新路子 and 老路子的一个图示

用 python 做数据分析过程中，大家首先需要掌握 python 的基础知识及语法，然后就需要学习核心的分析工具了，这里几个核心的包就比如 numpy、pandas、matplotlib，掌握了他们后基本就可以实现 excel 的所有功能了，且大家可以通过写

代码，实现模块化运作，高效率完成重复性工作等等。

同时在分析层面，除了基本统计运算之外，对于较复杂的算法（比如机器学习中要涉及到的回归，分类，聚类）都有强大的工具包使用，但大家需要了解算法本身的逻辑，这里数学基础要求较高，需要长期理解学习。

另外在数据可视化层面，图表本身应该是辅助数据研究，而 python 的 matplotlib、seaborn 支持非常多种数据图表（面向过程分析），而课程中讲的 bokeh 也可以实现动态交互图表（面向结果展示），这些工具包也都是开源的，通过 python 调用就可使用。



这里我拿前置课的一个数据研究案例做个例子

这个图的源数据是两百多条包含不同项目的运动员数据，且包含的信息有三点：

①男女运动员身高分布密度曲线，②男女运动员身高平均值，③男女运动员样本数具体分布

如果想初步查看男女运动员的身高数据分布，分析逻辑就是 数据按照性别分组整理 → 制作直方图/面积图 → 分别求出身高平均值做参考直线 → 图表叠加

在 python 中可以通过 seaborn 的 distplot 方法直接生成图表（会自动按照性别统计做出 2 套图表），但如果要通过 python 实现，除了知道一行代码之外，还需要了解图表本身的生成逻辑。

数据分析师（Python）前置课是免费的，戳链接，即可领取课程：

https://study.163.com/topics/python01_m/

人生苦短，你用 Python！抓紧学起来吧！

对于 python 做数据分析师的优势可以在之前的直播课程中查看:

数据分析师（Python）体验课请戳：<https://url.163.com/8ub>

数据分析师（Python）之前的免费直播课请戳：<https://url.163.com/v8R>

数据分析师（Python）报名请戳：<https://url.163.com/ipb>

我们再来看第五点：良好的数据敏感度，能注意到数据的异常和错误以及挖掘到数据中隐蔽的有价值信息，具备良好的商业分析能力和输出结论的能力。

我想通过这一点延伸一点告诉大家，数据工作者的软实力——数据感知，这个能力是需要长期数据工作慢慢总结的，可以说是一种“数据经验”，这个能力也是抛开技术之外的核心竞争力，它需要你在对行业熟悉的基础上，对数据有很强的认知，面对问题可以快速找到研究方向，从数据中挖掘答案。

好了现在我针对不同的层次的小伙伴儿，带来一些个人成长的建议，首先对于数据分析工作，我把这里面的几个技术环节放一张图给大家看看：



数据分析工作本身的方法套路基本就是数据前期处理 → 数据分析 → 数据可

视化，小伙伴们可以分别看看这三个阶段中，自己的痛点问题在哪里：是工作效率不高？还是无法深入挖掘问题？具体的技术层面我在最后的互动环节给大家解答，我现在先给不同阶段的小伙伴们一些成长建议。

“痛点问题在哪里”，这个大家也根据自己的情况思考一下：

1、面临毕业的大学生

(1) 首先思考自己选择什么行业，每个行业都需要数据分析

(2) 技术体系上，可以给自己 2-3 个月革新一下，比如以 python 为技术平台成为一个写代码的数据分析师，熟悉数据工作的整套流程工作，把技术先抓起来，实际项目时间在工作中再去学习

(3) 如果有时间，可以尝试一些算法实战项目，比如 kaggle 上的一些竞赛题，提高自己的算法能力

目标岗位薪资：按照 8k-12k 的目标争取，在岗位上学习 1-2 年后，往 15-20k 的目标再争取



2、刚入职 2 年内的业务人员，薪资在 10k 以下

(1) 花时间思考目前工作的痛点问题，这个阶段的数据分析师很多时间会花在前期的数据工作上，能否找到办法优化工作方法，提高效率，可以尝试通过写代码完成，比如 python、excel 的 vba 语言，sql，给自己安排一个学习计划

(2) 开始花时间学习算法层面的内容，补一下统计学，如果开始写代码做数据

分析，可以研究一些机器学习的内容，基本的回归、分类、聚类方法

(3) 沉淀行业知识，给自己在 3 年后顶一个目标，如果想转管理岗应该怎么扩充知识，积累自己的数据敏感性

目标岗位薪资：按照 12-15k 的水平做下一步跳槽（慎重思考是否转行），1-2 年后按照 20k 的水平再争取

3、企业中层，未来准备组建数据团队等

(1) 花时间了解目前数据挖掘主流方法（python，R），以及数据层面的行业实践

(2) 重点培养数据感知能力，提高对数据的理解和把控能力，了解数据市场动态等

这一部分就不做薪资建议了，但有个点，目前行业整体都在转型，17，18 年特别明显，所以行业本身也在做各种尝试，那么也代表了市场不成熟，所以定价体系等等也还在完善，所以这个创新也是要尝试的。

最后再罗列一下如果准备学习 python 语言的一些学习方向：

①python 基础语言语法

②numpy 工具包：用于科学计算、矩阵算法（其实 excel 的列与列的相互计算就是矩阵计算）

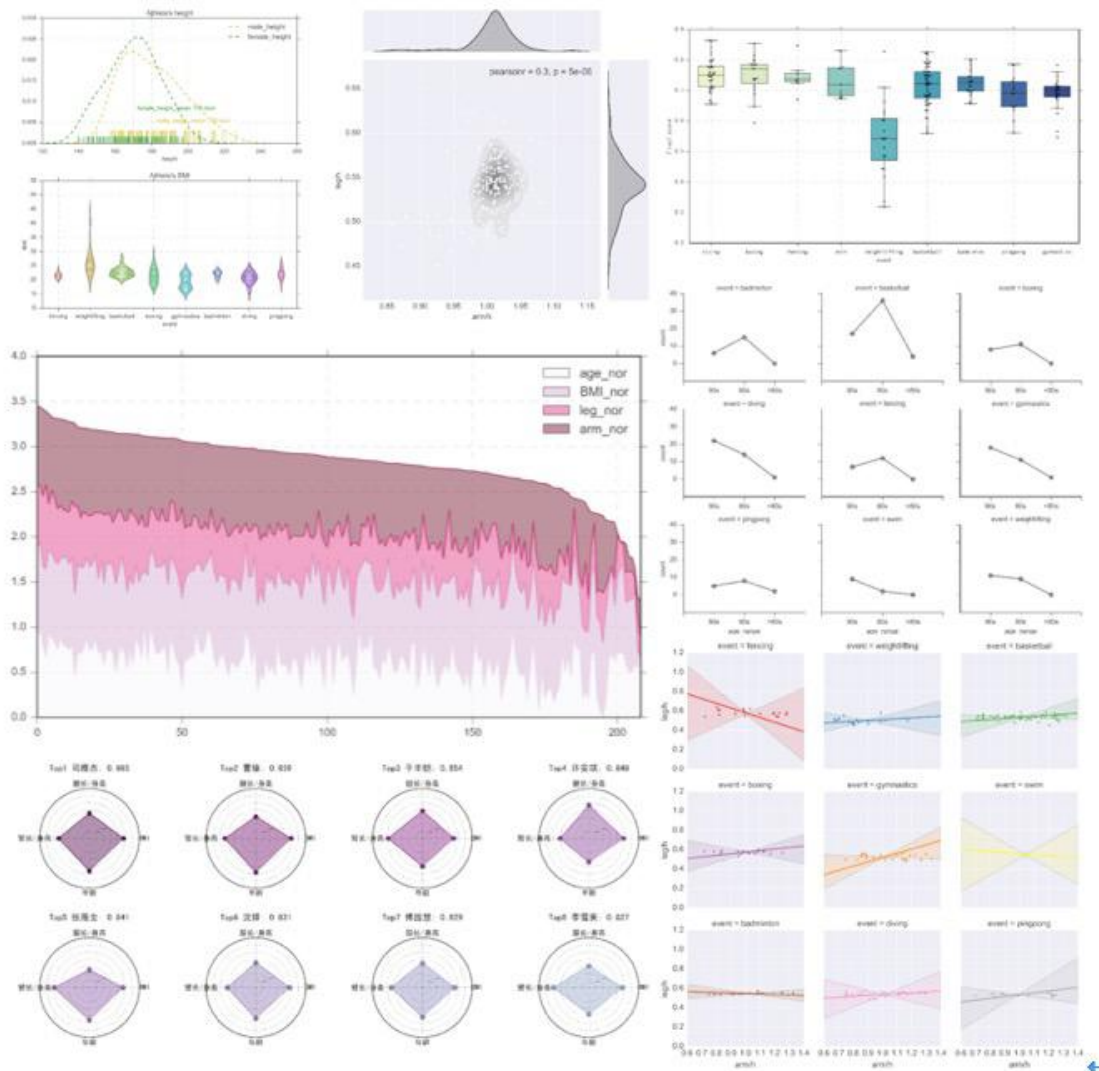
③pandas 工具包：用于数据分析，学习它之后就可以“像用 excel”一样使用 python 做数据分析

④matplotlib 工具包：基本绘制图表的工具包

⑤seaborn 工具包：结合 matplotlib，制作较复杂可视化图表

⑥熟悉基本回归、分类、聚类方法 按照需求去使用相关工具包 比如 scikit-learn 等

⑦bokeh 工具包：面向结果呈现，制作可交互的可视化图表



这是我们前置课将会学到的内容（免费），大家可以看看。

数据分析师（Python）前置课请戳：

https://study.163.com/topics/python01_m/

数据分析师（Python）免费体验课请戳： <https://url.163.com/8ub>

数据分析师（Python）报名请戳： <https://url.163.com/ipb>