

十周入门数据分析学习大纲

“我是文科生出身，可以学习数据分析吗？”

“我没有编程基础，可以成为数据分析师吗？”

“学习数据分析必须学习 R 和 Python 吗？”

... ..

其实，数据分析没有想象中那么难，入门也没有那么多条条框框。

我认识的 HR 转数据分析，市场营销转数据运营的，大有人在。

数据分析重要的是问题的理解、分析的思路、分析的流程及结果的解读，工具和编程都是实现这些分析思路的手段之一。

不过优秀的数据分析师也并非几日能速成，但入门也有入门的捷径。

盘算了一下数据分析的入门知识，大体分为以下这些，只要拿出你大学时啃高数的状态，每周夯实一个基础，基本能学成。

学习大纲：

数据分析的思维和方法

Excel 进阶

数据库理解和 SQL 入门

数理统计学

数据分析软件应用

数据可视化

常见的业务分析模型

Python/R 语言掌握

业务理解和指标设计

增长黑客：数据驱动增长

第一周：培养数据分析思维

为什么数据分析思维很重要？

如果我们在分析一个问题前，思维缺失就像下面图中所表达的一样，往往不知道问题从哪里下手，即使拿到数据也是一脸懵逼。

所以我们要通过训练数据分析思维，帮助在遇到问题时，大家脑中能快速梳理出分析的切入点，甚

至是分析的思路，这一点很重要。

常用的一些思维方式：

1、金字塔/结构化思维

把待分析问题按不同方向去分类，然后不断拆分细化，能全方位的思考问题，一般是先把所有能想到的一些论点先写出来，然后在进行整理归纳成金字塔模型。主要通过思维导图来写我们的分析思维。

2、公式化思维

在结构化的基础上，这些论点往往会存在一些数量关系，使其能进行+、-、×、÷的计算，将这些论点进行量化分析，从而验证论点。所谓指标体系，就是这么梳理得来的。

3、业务化思维

业务化即是深入了解业务情况，结合该项目的具体业务进行分析，并且能让分析结果进行落地执行。用结构化思考+公式化拆解得出的最终分析论点再很多时候表示的是一种现象，不能体现产生结果的原因。所以需要继续去用业务思维去思考，站在业务人员或分析对象的角度思考问题，深究出现这种现象的原因或者通过数据推动业务。

增加业务思维方法：贴近业务，换位思考，积累经验。

同时，这样的思维模式在一些特定业务场景下，还衍生出一些基础的分析方法，比如象限法、多维法、假设法、指数法、二八法、对比法、漏斗法，这个对未来构建分析模型都有帮助。

思维模型的好处是他能提供一种视角或思维框架，从而帮助你建立起观察事物和分析问题的视角。通过对思维模型的学习和训练，能提高你成功的可能性。

第二周：Excel 技能进阶

学习 Excel 是一个循序渐进的过程

基础的：简单的表格数据处理、筛选、排序

函数和公式：常用函数、高级数据计算、数组公式、多维引用、function

可视化图表：图形图示展示、高级图表、图表插件

数据透视表、VBA 程序开发

按照我习惯的方法，先过一遍基础，知道什么是，然后找几个 case 练习。多逛逛 excelhome 论坛，平常多思考如何用 excel 来解决问题，善用插件。

函数和数据透视表是两个重点，结合业务场景来学习，可参考《谁说菜鸟不会数据分析》。

制作数据模板必须掌握的 excel 函数

日期函数：day, month, year, date, today, weekday, weeknum。日期函数是做分析模板的必备，可以用日期函数来控制数据的展示，查询指定时间段的数据。

数学函数：product, rand, randbetween, round, sum, sumif, sumifs, sumproduct

统计函数：large, small, max, min, median, mode, rank, count, countif, countifs, average, averageif, averageifs。统计函数在数据分析中具有举足轻重的作用，求平均值，最大值，中位数，众位数都用得到。

查找和引用函数：choose, match, index, indirect, column, row, vlookup, hlookup, lookup, offset, getpivotdata。这几个函数的作用不用多说，特别是 vlookup，不会这个函数基本上复杂一点的报表寸步难行。

文本函数：find, search, text, value, concatenate, left, right, mid, len。这几个函数多半用在数据整理阶段使用。

逻辑函数：and, or, false, true, if, iferror。

数据透视表

数据透视表的作用是把大量数据生成可交互的报表，数据透视表具有这样一些重要功能：分类汇总、取平均、最大最小值、自动排序、自动筛选、自动分组；可分析占比、同比、环比、定比、自定义公式。

第三周：学习数据库原理和 SQL

做数据分析，数据从哪里来？数据库！怎么取数据？写 SQL！

做数据分析，取数、清洗数据，基本都要依赖 SQL。

初入门阶段，对于数据库不必精通，只需了解常用的数据库类型，能够在现有的表格里查询出数据，能够更新数据对数据进行重编码，知道怎么增加添加数据，把数据变得规整就行。理解主键，索引等含义和用处。导入导出数据可以使用工具，分析数据可以使用 ODBC 或者其他的接口对数据库进行连接。取数的排序，做数据的交集并集，数据转换，数据表合并等这些，最好也能掌握。

这里我总结了几个核心技能：

技能一：学会用 select 语句添加字段和找出需要的数据

直接给一个随时可以套用的万金油模板吧：

```
select cola,colb,colc into newtable from oldtable where cola='x' and colb is not null;
```

基本上，学会这个就可以完全的查出大部分的数据了。

select 后面是一个个的字段，要哪个选哪个。有 into 意味着放到一张新表里面，没有就是查询出来。where 之后的就是我们的条件，等于某个值，或者是不是空值，是最常用的几种查询方式吧。

还有一种 select 也用的非常多：select cola from oldtable group by cola;

这个语句是看看 cola 有多少种值的可能性。

select 进阶学习，可能要讲讲 join, union，以及多个查询组合成的嵌套查询，或者是子查询的模式，以及模糊查询。这个后面我会再花篇幅写出来给大家参考。

技能二：学会 alter 学会增加，减少字段

`alter` 可以做的事情很多，增加字段，减少字段，增加主键减少主键等等，非常常用。

1. 增加字段：`alter table tablename addcolname varchar;`

这样就可以增加一个空字段，`varchar` 是一种数据类型。

2. 减少字段：`alter table tablename dropcolumn colname;`

这样就去掉了一个原有的字段。

技能三：学会 `update` 学会更新数据更新数据

大概常用的有两种，一种是更新成一个固定值：

`update table setcol=1;`

另一种是从另一张表里面更新，这种方法，在处理一些小型数据的时候经常会导出，然后导入到数据库，就可以用：

`update table setcol=tableb.col from tableb where table.id=tableb.id;`

里面 `table` 和 `tableb` 是两张表，然后通过两张表的 `id` 关联起来，学会这个书写结构就行。

第四周：数理统计学

统计学是数据分析师必备的基础知识之一，是一组用于汇总数据和量化给定观测样本域属性的工具。

单独的原始观察数据只是数据，还不能变为我们想要的信息或知识。有了原始数据，那么接下来的问题是：

- 什么是最常见或可预期的观测？
- 观测的限制条件是什么？
- 数据是什么样子的？

回答这些问题，我们需要借助一些统计工具来得出一些结论。借助统计学，你的分析深度、专业度和科学性都会有很大提升。

所以这一周，我们需要掌握统计学的以下几大概念：

1. 集中趋势(中数、众数、平均数)
2. 变异（四分位数、四分位距、异常值、方差）
3. 归一化(标准分数)
4. 正态分布
5. 抽样分布（中心极限、抽样分布）
6. 估计(置信度、置信区间)
7. 假设检验
8. T 检验

推荐书籍：吴喜之-《统计学·从数据到结论》

第五周：数据分析软件应用

有了数据分析思维基础，懂一些统计学知识之后，我们就可以着手开始相对专业的分析，用可视化的方法探索数据的规律。

这一周，除了 Excel，你需要掌握一个傍身的数据分析工具。

考虑到快速入门，这里暂时放一放 SPSS、R、Python 一类工具，先掌握 BI 工具的运用，帮助快速熟悉起数据分析的流程。知名的 BI 产品有 Tableau，PowerBI，还有国内的 FineBI，网上都有体验版和免费版下载。处理好的数据拿来放 BI 分析，分分钟就能出很漂亮的可视化，比 Excel 的图表高级多了，而且绝大多数人都能轻松上手。

BI 需要掌握数据的连接，连不上数据怎么分析。还有仪表盘 Dashboard 的概念，知道绝大多数图表适用的场景和怎么绘制，维度和指标的区分。一些数据的清洗，如果 BI 掌握得透彻也可以放 BI 处理，但不熟悉还是用 SQL 处理吧。

第六周：数据可视化

可视化看似是简单的步骤，但也是有造诣的。可视化说白了是一种表达，数据分析结果表达的是否到位，领导是否认同，工资涨不涨，全靠这一纸 dashboard（当然还有你“讲故事”的功力）。

如何选择最佳的图表类型？趋势性、相关性、分布性、周期性、地理位置分布性.....

颜色和字体等细节样式方面，如何进行更加美观的调配。

布局设计原则,故事性布置可视化仪表板，报告的标题和结论注释，以及整体展现的逻辑性。

还有很多可视化的陷阱，都是值得花一周探究的。

第七周：常见的业务分析模型

基于一些数据分析方法，如象限法、多维法、假设法、指数法、二八法、对比法、漏斗法，在特定业务场景下，还衍生了通用的业务分析模型，常用的有购物篮分析模型，RFM 模型，漏斗分析模型，客户生命周期，以及预测、聚类分析等挖掘模型。

这一周我们要掌握常用的分析模型，最好能有深刻的认识，直接套用到实际的业务场景中，活学活用。

第八周：Python/R 语言掌握

到这一周，数据分析的入门之路基本上完成一大半。

本着提升自己，以及加大自己求职和面试的筹码，掌握 Python 或 R 绝对是加分项。

有关数据分析的编程语言有 Python 和 R 语言。R 语言倾向于统计分析、绘图等。统计学家或者学统计学的喜欢用 R 语言，而我更青睐学习 Python，因为 Python 是面向未来的语言，无论从流行度、可用性还是学习难度来讲，Python 都是最好的入门语言。

当然，如果可以的话，再掌握一下 R 语言是最好不过的，学习嘛，永无止境。

Python 有很多分支，但我们学习的主题是数据分析。

入门推荐《深入浅出 Python 编程》

从 code academy 开始学起，完成上面的所有练习。Code academy 涵盖了 Python 基本概念。当完成了 code academy 练习之后，看看这个 lpython notebook:

其次，掌握三个库 Numpy、Pandas、Matplotlib

Numpy 是利用 Python 科学计算的基础包，对 Numpy 好的掌握将会帮助你有效地使用其他工具例如 Pandas。包括 N 维数组，索引，数组切片，整数索引，数组转换，通用函数，使用数组处理数据，常用的统计方法等等。

Numpy BasicsTutorial, Index Numpy 遇到 Numpy 陌生函数，查询用法，推荐！

Pandas 包含了高级的数据结构和操作工具，能使得 Python 数据分析更加快速和容易。包含 series, data frams，从一个 axis 删除数据，缺失数据处理等等。

Pandas BasicsTutorial, Index Pandas 遇到陌生函数，查询用法，推荐！

Matplotlib 是一个强大的 Python 可视化库。几行代码就能绘制出散点图、折线图、直方图、柱状图、箱线图等。

第九周：业务理解和指标设计

到了第九周，大家可以发现，这个学习计划更多是偏业务的数据分析，可见业务理解的重要性。但业务理解需要多年的积累，没有掌握的捷径。刚入门也不会拷问太多业务上的问题，更多时候是考验你逻辑思维能力和数据分析的方法。所以简单花一周时间了解各行业的业态，各业务的通用指标。

其次，指标体系。几乎一个数据岗的岗位要求都会涉及这样一句话：“负责建立和优化部门的数据指标体系”。事实上目前太多的数据岗主要工作都是不断完善与优化数据指标体系，而对层面的工作是比较少的，即使岗位叫做数据分析师。一个优秀的的数据指标体系，不仅能让你快速解决数据需求，洞察出可能会被忽略的价值数据，还能反映出你目前最需解决的业务问题。所以，这一周还要掌握梳理业务指标的思路。

第十周：数据驱动&增长黑客

这个话题比较严肃，需要把数据分析师这个岗位放到整个公司甚至整个行业来看。任何一个岗位的设定都是要为企业带来价值或利润的，数据分析师也是如此。要搞清未来价值所在，升职加薪之路才会明朗。

目前有一个非常火的词来形容一类数据分析师，叫“增长黑客”，所做的事也很时髦，叫“数据驱动”。字面意思很好理解，就是利用数据分析技术给业务带来增长，驱动企业转型。

本周，你要了解业界不管是传统行业还是互联网行业数据驱动业务增长的成功案例。

未来成为增长黑客要做好哪些准备？

日常工作中，如何让自己避免成为“取数机”？

如何推动数据分析工作并得到老板重视？

工作中有哪些工作细节，去潜移默化的改变身边人尤其是老板对数据价值的认知？