



The Complete Guide to Landing a Career in Data

数据行业终极职业指南

章节列表

1. 简介
2. 职业选择：数据分析师、数据科学家和数据工程师
3. 稀缺技能：21世纪最性感的工作
4. 工作：数据分析师薪资是多少？工作机会散布在哪里？
5. 角色和技能：你需要掌握哪些技能才能找到工作
6. 软技能：成功必备品质
7. 面试：成功攻略
8. 作品集：找到工作的秘密武器
9. 首个数据工作：什么样的工作，如何找到这样的工作
10. 数据的未来：预测分析简介
11. 总结



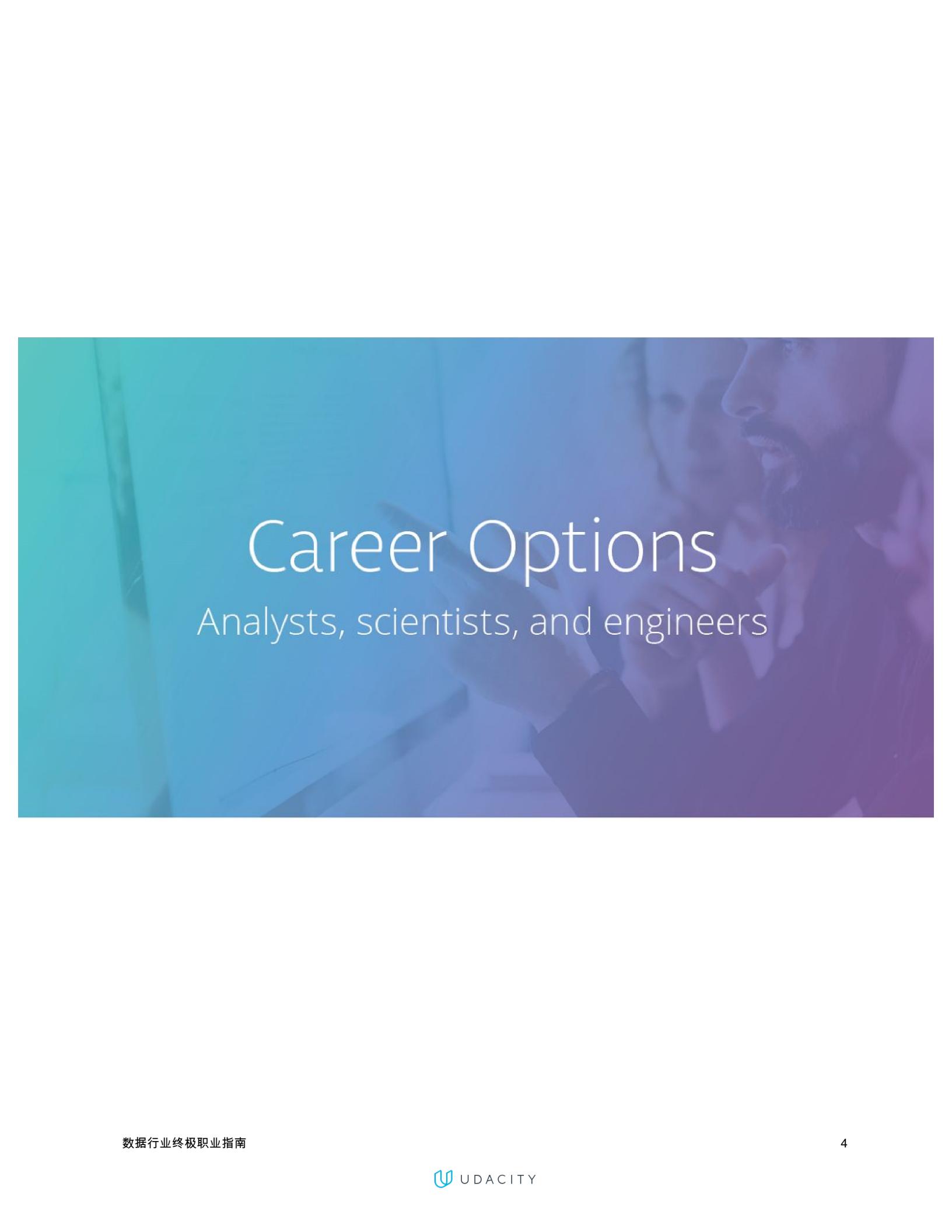
在选择职业道路时，你可以追求平稳，也可以选择冒险。数据职业同时提供了这两种选择。一方面，它是一个安全的选择——对数据人才的需求在继续增加，并没有缩减的迹象。当你掌握数据技能后，几乎在任何行业都会成为炙手可热的人才。另一方面，外面是个全新的世界，我们产生了庞大的数据，数据让很多精彩的新事物成为可能。但是，为了利用所有这些数据，我们需要借助目前所掌握的方法和策略来发明新的方法，意味着未来会实时发生变革。站在数据行业前沿的人也是站在技术发展前沿的人。

好消息是，无论你选择哪条道路——安全的道路、危险的道路，还是二者兼之，都会有无数的工作机会来满足任何对数据感兴趣的人士的需求。

但是，如何实际行动起来呢？

这份指南就可以派上用场。在 Udacity，我们非常有幸能与世界上最具前瞻性的公司合作，并能够接触到数据领域最具创新思维的人士。我们的招聘合作伙伴，如 Facebook, Google, Tableau, mongoDB 等，提供这一领域最吸引人的职位机会，并且是了解企业对数据人才有何期望的重要信息来源。

我们汇总了所有资料，并整理出了这份指南，我们的专业课程导师提供了独特的见解和经验。如果你想进入数据领域，那么这份指南最合适不过了。



Career Options

Analysts, scientists, and engineers

每当你发短信、发推文、在 Facebook 上晒照片、点击链接或在线购买商品时，都会生成数据。想一想，世界上有超过 [35 亿的互联网用户](#) 和 [20 亿的手机用户](#)，这该会产生多少数据啊。

幸运的是，数据成倍增加后，人们收集、组织和分析数据的能力也提高了。数据存储比以往成本更低，数据处理能力比以往更强大，各种工具更容易接触，使我们能够从大量数据中挖掘信息并为企业提供市场情报。在最近几年，数据分析的应用越来越广泛，从股价预测到预防家庭火灾，不一而足。

所有这些数据分析都需要大量数据人才。引申含义是：现在是进入数据领域的最佳时机。在 2011 年，[麦肯锡全球研究所](#) 指出所有这些数据分析都需要大量数据人才。引申含义：现在是进入数据领域的最佳时机。在 2011 年，麦肯锡全球研究所预测到 2018 年，美国可能会缺少 150 万了解如何利用数据分析做出有效决策的人才。实际上现在马上将要超过这一数字。[CrowdFlower](#) 最近公布的一项数据科学报告指出，“在接受问卷调查的数据科学家人群中，有 83% 的人士认为现今的数据科学家处于短缺状态”，这一结果很让人震惊。

赶快把握机会吧。

已经踏出成为专业数据人士的第一步？考虑下以下三大职业选择吧：数据分析师、数据科学家和数据工程师。

数据分析师

数据分析师本质上是初级数据科学家。如果你刚开始接触数据这一行业，并急于尝试进入这一领域，那么数据分析师是最佳起点。数据分析师不需要具有数学或研究背景以便开发新的算法，但是他们能够熟练地使用现有的工具来解决问题。

技能和工具

数据分析师需要对以下五大核心技能有基本掌握：编程、统计学、机器学习、数据再加工和数据可视化。要想成为一名成功的数据分析师，除了技术技能之外，还需要注重细节，能够有效地呈现结果。

工作和职责

数据分析师会受到企业内经验更丰富的数据专业人士的指导。在他们的指导下，数据分析师需要获取、处理和总结数据。数据分析师要保证数据收集的质量，按照相关人士的要求定期查询五大数据库，处理数据问题以便及时得出解决方案。他们还会以文字或图表的形式呈现易于看懂的结论。

对于所有数据专业人士来说，无论职位级别高低如何，对数据始终保持好奇心、密切检查不断演变的最佳做法和工具都至关重要。

数据科学家

有些公司将“数据科学家”和“数据分析师”看做同等职位，但是二者之间在技能和经验方面存在一定的差别。虽然数据科学家和数据分析师在公司具有相同的使命——从庞大的数据池中得出结论，但是数据科学家的工作需要更复杂的技能，以便处理量更大、变换速度更快的数据。

因此，数据科学家能够在没有指导下开展调查研究，并处理开放性问题。数据科学家通常在量化领域具有很高的学历，例如计算机科学、物理学、统计学或应用数学，并且能够开发新的算法来解决数据问题。

数据科学家在公司内的作用非常大，因为他们能够发现新的业务机会或通过发现隐藏的数据模式（例如发现令人惊喜的客户行为或发现潜在的存储集群故障），为企业节省开支。

技能和工具

数据分析师可能研究的是仅从单一来源获得的数据，而数据科学家研究的是不同来源的数据。数据科学家会使用 Hadoop（最常用的分布式文件系统处理框架）等工具，使用 Python 和 R 等编程语言，并将高级数学和统计学知识运用到工作中。

不同的单位和项目对技能的要求有所不同，下面的 Data Science London 示例展示了数据科学家工具包的复杂程度：

[Sort of a] Data Scientist Toolkit

- Java, R, Python... (bonus: Clojure, Haskell, Scala)
- Hadoop, HDFS & MapReduce... (bonus: Spark, Storm)
- HBase, Pig & Hive...(bonus: Shark, Impala, Cascalog)
- ETL, Webscrapers, Flume, Sqoop... (bonus: Hume)
- SQL, RDBMS, DW, OLAP...
- Knime, Weka, RapidMiner... (bonus: SciPy, NumPy, scikit-learn, pandas)
- D3.js, Gephi, ggplot2, Tableau, Flare, Shiny...
- SpSS, Matlab, SAS... (the enterprise man)
- NoSQL, Mondo DB, Couchbase, Cassandra...
- And Yes! ... MS-Excel: *the most used, most underrated DS tool*

Data Science London

来源：[Data Science London](#)

数据科学家具备的最有价值的非技术技能是强烈的好奇心。数据科学家应该积极提出问题并寻求解决方案，从而挖掘出可能会让企业发生变革的宝贵信息。

工作和职责

数据科学家本质上就是利用数据来解决企业问题。他们分析数据、做出推断并获得解决方案，从而提出可以行动的建议。数据分析师是对过去的数据进行总结，而数据科学家是对未来做出决策。

数据科学家可以明确地判断如何优化网站，以便更好地留住客户，如何宣传产品，实现更强大的客户生命周期价值，或者如何细调交付流程，以便加快速度和尽量减少浪费。

数据工程师

数据工程师负责构建稳健、容错的数据渠道，并将杂乱无序的数据清理、转换和汇总成数据库或数据源。数据工程师通常都是软件工程师。数据工程师的职责不再是数据分析，而是负责编译和安装数据库系统、编写复杂的查询、扩展到多个机器上，并部署灾难恢复系统。

数据工程师本质上是为数据分析师或数据科学家打好基础，使他们能够轻松地检索所需的数据，并进行评估和实验。

技能和工具

数据科学家会从数据中发现规律，而数据工程师负责确保数据能从源头流畅地抵达目的地，以便处理数据。

因此，数据工程师需要熟练掌握以下知识和技能：

- 基于 Hadoop 的技术，例如 MapReduce、Hive 和 Pig
- 基于 SQL 的技术，例如 PostgreSQL 和 MySQL
- NoSQL 技术，例如 Cassandra 和 MongoDB
- 数据仓库解决方案

工作和职责

数据工程师负责幕后工作，使数据分析师和数据科学家能够更高效地完成他们的工作。

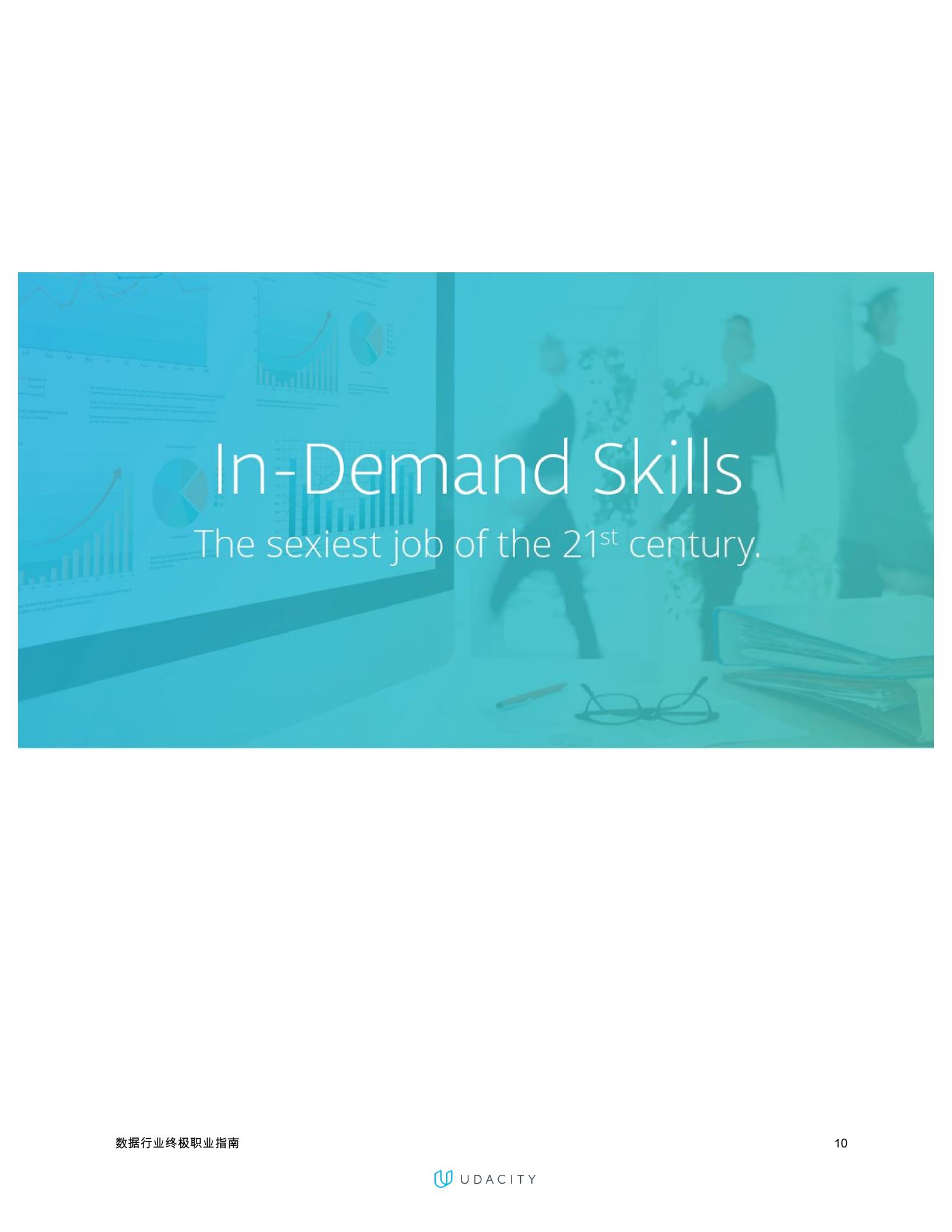
底线

对于数据行业来说，你有很多选择。如果你想要探索这一领域，则可以考虑以下三大主要选择：数据分析师、数据科学家和数据工程师。

作为大数据分析和可视化初创企业 Datavore Labs 的联合创始人，Sanjay Venkateswarulu 对这些分支领域的发展历史做出了以下比喻：

“数据分析师逐渐演变成了这三大学科或更专业的学科。我认为随着现代医学的诞生，医生也经历了相同的专业化发展历程。一开始是村长或长者扮演着主要角色，但是随着工具越来越细化，现在出现了普通医师、外科医生和神经外科医生。”

如果你刚开始接触数据科学这一领域，建议你先瞄准分析师岗位，用 Venkateswarulu 的比喻就是“普通医师”。当你积攒了各种技能和经验后，你将能够成长为数据科学家或数据工程师。



In-Demand Skills

The sexiest job of the 21st century.

很少有哪个职位能像数据科学家这样获得最夸张的比喻了。最著名的当属[哈佛商业评论](#)将数据科学家称为“21世纪最性感的工作”。虽然会有这样的溢美之词，但是负面评论也不可避免，肯定有一些否定的声音，甚至夹杂着一些末日言论，指出在10年之内，这一岗位就会被自动化完全替代。

这绝不会发生。

数据科学职业的重要性

无论你对数据科学很性感这一说法有怎样的看法，但是绝对无法忽略的是数据以及我们分析、组织和语境化思考数据问题的能力继续保持重要性。Glassdoor根据大量就业数据和员工反馈，发布了[美国25大最佳职位](#)榜单，猜猜排名第一的是哪个？数据科学家！

虽然这一职位保持不变，但是毫无疑问，数据科学家的具体工作职责会发生变化。随着机器学习等技术越来越普遍，深度学习等新兴领域越来越受到研究人员和工程师以及招聘公司的关注，数据科学家继续引领着创新浪潮和技术发展之路。

企业被数据淹没，渴望独特见解

更重要的是存在需求。福布斯杂志在2015年发表的文章《[2016年十大最难填补的职位](#)》指出“合格的应聘者处于短缺状态。”

随着大数据的不断涌现和追踪数据的需求在不断增长，雇主一直都在寻求数据科学家。但是合格的候选人士却处于短缺状态。这一领域太新了，以至于美国劳工统计局都未将它作为一个职业加以统计。但是数千家公司都在急寻可以将大批量的数据整理成实用信息的数据人才，从分析信用卡数据以便在目标市场展开营销和广告宣传的初创企业到福特汽车和普华永道等企业巨头们无不例外。正如技术市场研究公司Forrester在2016年4月的一份报告中提到，“企业淹没在数据中，渴望获得独特见解”。

工作与生活的平衡

事实是，如果你想进入新的职业领域或打算改行，那么数据科学家是个非常有吸引力的职业选择。不仅仅是市场对这一职位的需求庞大，而且给出的薪资也很诱人！但是这一岗位却可以在工作与生活之间实现完美平衡。根据Glassdoor的另一项研究（[25大最具工作生活平衡度的职位](#)调查研究），数据科学家实际上是排名第一的职位。

但是，虽然现在你已经对这一热门领域很激动了（在读了这么多内容后，谁不激动啊？），还是要注意，虽然需求非常大，薪资不断增长，工作生活平衡度也很好，但是要想在这一领域取得成功，还是会面临一些实际挑战。或许最关键的就是技术挑战。

变革的步伐

上面所引用的福布斯文章的信息大部分来源于人力资源管理协会 (SHRM) 与一家叫做 CareerCast 的加利福利亚公司共同展开的一项调查研究。[在 SHRM 网站](#)上发表这一研究结果的作者 Tony Lee 指出：

技术飞速发展使得很多技能瞬间变得过时，而这一现象也导致岗位培训紧俏起来，并且应聘者本来可以完美符合要求的技能在上岗一年之后就过时了。

现今，人们都在广泛讨论现代教育模式逐渐转变成了终身学习模式。Udacity 总裁兼联合创始人 Sebastian Thrun 在 NPR 上提到了这一点，并且指出传统意义的现代大学无法跟上科技的发展步伐，因此无法充分支持学生实现职业发展目标或满足企业渴望引入新型人才的需求。

“纳米学位项目是绝对最前沿的理念。大学本质上已经落后了。因此，现今的大学很少会教授数据科学等新事物。我认为世界已经从短暂教育转变成了终身教育。越来越多的人需要从一个岗位跳到另一个新的岗位，这就导致我们需要从一次拿到学位之后再也不会去继续接受教育的模式转变成融入日常生活、工作时间的教育。我认为我们的学生不仅需要免费教育，还需要跟上时代的教育。”

“工作”的新定义

如今的就业环境已经发生翻天覆地的变化，每天都会出现新型职位，工作任期大大缩短了。“工作”的含义被重新定义，需要体现出技术、流动性、灵活性和全球关联性。鉴于这些变革，我们能够清晰地理解取得成功需要掌握的技能至关重要。

在最近的一篇[《华尔街日报》文章](#)中，作者 Elizabeth Dwoskin 引用了 RJ Analytics 的一项报告中的内容：

“数据科学家要掌握的五大技能是数据分析、数据挖掘、机器学习和编程语言 R 和 Python。”

数据科学家的工作职责是什么？

现在我们再回到《哈佛商业评论》将数据科学家称作 21 世纪最性感的工作这篇即权威又值得怀疑的文章。在这篇文章中，作者用比较华丽的辞藻描述了数据科学家的工作职责：

“数据科学家的职责是在大批量的数据中寻找规律。他们喜欢探索周围的世界。他们在数字领域悠闲地游荡着，能够将大量的无序数据整理出来，并使分析过程变得可行。他们会

识别丰富的数据来源，并与其他可能不完整的数据来源整合到一起，然后整理最终的结果数据集。现今社会竞争异常激烈，挑战总是不断变化，数据不断涌现进来，而数据科学家可以帮助决策者从专门的分析流程转变成与数据对话模式。”

谁会完成这种工作？

“可以将数据科学家看做数据黑客、分析师、沟通者和受信任的顾问的综合型人才。”

如果你觉得这听起来很浪漫、令人兴奋，没错，还性感，那么数据科学家正合你的口味。但是如果你觉得这段描述对你来说太抽象了，那么不妨考虑以下这段总结：

学习如何：

- 从各种数据库、数据格式和数据来源中整理、提取、转换和加载数据
- 使用探索性数据分析技巧从复杂的数据集中发现有意义的联系、模式或趋势
- 对无标签的数据进行分类，或利用应用统计学和机器学习预测未来
- 通过有效的数据可视化图表传达数据分析结果

这段内容来自 Udacity [数据分析师纳米学位项目](#) 的摘要，清晰地总结了数据科学家的日常工作，他们使用什么工具和技巧并如何应用所学的知识。

总之，我认为我们可以保证的是数据一直都存在着，那些研究数据的人士也不会被科技淘汰。如果你想找一份充满机遇、供不应求、报酬丰厚并且具有无与伦比的工作生活平衡度的工作，那么赶紧选择数据科学家吧！

Salaries

What do data analysts make,
and where do they make it?

数据是 21 世纪新的发展前沿，几乎所有领域的企业都被潮水一样涌进的新信息淹没了。大量的数据为我们提供了详细的信息，包括客户何时、何地、多久使用某个产品，以及某个流程在接近无限数量的接触点的效果如何。

但是如果没有人来整理、评估这些数据，从中发现可以采取行动的见解并以图表或口头形式展现出来，那么这些数据对企业来说就是毫无价值的。

这时候就需要你来发挥作用了。

对高技术人才的需求居高不下

数据分析师总是供不应求，原因很简单，这项工作本身要求就很高。数据科学汇集了统计学、商业情报、社会学、计算机科学和通信科学。你必须是数字天才，同时善于沟通；你需是一名分析高手，具有抽象思维。

正如高级网络技术经理 Erik Berger 最近提出的观念，数据科学家不仅仅要掌握正确的技能：

“当然，你需要具备技术技能，才能提取数据并进行统计学分析，但是更需要掌握无形的技能，能够发现规律或不符合规律的现象并体现到报告中。要擅长数据科学，你需要完全理解你正在分析的业务的性质——仅仅查看数字只是故事的一半。”

这种对技术技能和敏锐能力的要求正是数据科学家一直处于高需求短缺状态的原因，因此非常受到重视，并且薪资报酬很高。

数据分析师的薪资范围

根据我们的合作伙伴 PAYSA 给出的数据，数据分析师的薪资范围可以达到 8.5 万到 12.5 万美元！

随着对数据专业人士的需求不断增长，数据岗位之间的交集也越来越多。数据分析师经常与数据科学家（注意：某些企业将这两种职位合二为一了）、数据库管理员、数据工程师密切合作，并且可能还会随着企业的不断发展，与其他职位的人士合作。

下面是一些数据分析师的薪资范围数据供你参考，数据来自 [DataJobs](#)。

关于以下数据职位的全国薪资范围：

- 数据分析师（入门级）: \$50,000-\$75,000
- 数据分析师（具有经验）: \$65,000-\$110,000
- 数据科学家： \$85,000-\$170,000

- 数据库管理员（入门级）：\$50,000-\$70,000
- 数据库管理员（具有经验）：\$70,000-\$120,000
- 数据工程师（初级/通才）：\$70,000-\$115,000
- 数据工程师（领域专家）：\$100,000-\$165,000

上述每种职位对于数据获取、分析和交付来说贡献都很大。进入数据分析师这一职业领域之后，随着你不断磨练技能，将会有许多的选择机会。

请务必注意，由于数据行业的人才危机和动态发展状况，薪酬市场比较混乱。目前，薪资主要取决于企业愿意花多少钱来填补他们的迫切需求。

这些工作散布在哪里

毫无疑问，数据工作会明显集中在全国的主要科技枢纽地区：旧金山和纽约。

虽然旧金山仅代表了 [Dice](#) 上发布的职位机会中的 7%，但是有 24% 的大数据职位机会都来自这里。纽约和附近的华盛顿特区/巴尔的摩地区拥有第二多和第三多的大数据职位机会。

但是，初级数据分析师在全国很多其他大都市地区也有职位机会。波士顿和西雅图分别在 Dice 上发布了 3% 的职位机会，但是分别提供了 7% 和 6% 的大数据职位机会，意味着这些地区有大量你正在寻找的工作机会。费城和洛杉矶也如此。

要记住的一点是，对于工作地点来说，你有很多的选择。此外，还有一个好消息：如果你想待着的家乡没有数据分析师岗位，还有比以往任何时候都要多的合同工、自由工作者和远程工作机会。你可以自由选择对你的目标、生活习惯和经验来说最合适的工作。

底线

无论你是打算改行还是刚要进入工作岗位，选择数据分析师都保证会给你带来丰厚的收入和丰富的经验。

Roles and Skills

What you need to know to get hired.

很多机构都强调成为数据科学家需要综合掌握多个领域的知识，包括软件开发、数据挖掘、数据库、统计学、机器学习和数据可视化。

实际上并非始终这样。你肯定没必要了解终身受用的数据相关信息才能找到数据工作。为了判断哪些机会最适合你以及可以展示什么样的技能，一定要学会如何仔细阅读数据科学职位说明。这样，你将能够申请你已经具备必要技能的职位，或针对想要进入的岗位掌握具体的数据技能。

四个数据科学职位场景

“数据分析师”通常是用来概括各种相差极大的职位的统称。为了帮助你浏览各种职位机会，我们通过研究四大常见场景看看四个不同类型的数据科学职位：

1. 你是数据科学家和分析师：在很多公司，数据科学家就等同于数据分析师。你的工作任务可能包括从 MySQL 数据库中提取数据、成为 Excel 表格达人并生成基本的数据可视化内容（例如线形图和条形图）。偶尔需要分析 A/B 测试的结果，或成为公司 Google Analytics（分析）帐号的负责人。像这样的情形很适合想要学习知识、有抱负的数据科学家。一旦你能够顺利处理日常工作后，这样的企业就可以给你提供很好的环境，帮助你尝试新事物并拓展你的技能。
2. 你是招聘的首批数据人才：似乎有很多公司在出现大量用户流量（以及越来越多的数据量）时，才会想着找一个人来设置庞大的数据基础架构，帮助公司往前发展。他们还希望对方能提供分析。你会在“数据科学家”和“数据工程师”标题下看到此类招聘信息。对于这种情形，你可能就是他们的首位数据人才，所以无论是统计学专家还是机器学习专家都不重要。具有软件工程师背景的数据科学家可能会擅长此类职位，更重要的是，数据科学家可以为正式版代码带来有意义的数据类贡献，并提供基本的见解和分析。可能不会有太多的初级数据科学家指导机会，所以你会有更多的机会自己去尝试实验并成长起来。但是，因为指导很少，所以遇到挫折或止步不前的风险也更高。
3. 数据就是业务：有很多企业的产品就是数据（或数据分析平台）。在这种情况下，数据分析或机器学习将会要求很高。如果你拥有正式数学、统计学或物理学背景，并且希望继续朝着学术道路发展，那么这种情形最适合你。在这种情形下，数据科学家很可能更侧重于生成以数据为依据的产品，而不是回答公司的运营问题。这类公司通常都直接面向消费者并拥有大量数据，或者提供基于数据的服务。
4. 公司以数据为依据：有很多公司都属于此类情形。对于这种岗位，你可能会加入到成熟的数据科学家团队。通常，这类公司要么寻找的是通才，要么就是填补团队缺少人才的特殊岗位，例如数据可视化或机器学习。在参加这类公司的面试时，一些更重要的技能是熟悉“大数据”工具（例如 Hive 或 Pig），并拥有处理混乱、“实际生活”数据集的经验。

希望上述内容使你能够大概了解“数据科学家”这一职位的含义。上述四大情形的每个情形都寻求不同的技能、技术和经验级别。但是，所有这些职位信息都会说是“数据科学家”，所以一定要仔细阅读职位说明，了解你将加入什么样的团队、面临什么样的挑战、拥有哪些发展机会，以及需要具备哪些技能。

8 帮助你拿下 Offer 的数据技能

当你开始思考哪个数据职业道路适合自己后，则务必掌握下面这些基本技能，无论你选择哪个道路，这些技能都很重要。下面是无论选择哪种岗位都应具备的八大核心数据科学技能：

基本工具：无论你要面试哪种类型的企业，都很可能需要了解如何使用行业工具。这意味着 R 或 Python 等统计学编程语言，以及 SQL 等数据库查询语言。

基本统计学：至少对统计学有基本的了解对数据科学家来说很关键。你应该熟悉统计学测试、分布、最大似然估计值等。回忆下你的基础统计学课程吧！机器学习也一样，但是一个更重要的统计学知识则是了解不同技巧何时是有效（或无效）的方法。对于所有类型的企业来说，统计学都很重要，尤其是以数据为依据的企业，它们的产品并不侧重于数据，产品利益相关者将依靠你来作出决策并设计/评估实验。

机器学习：如果你就职于一家拥有大量数据的企业，或者公司的产品本身就以数据为依据，那么你可能需要熟悉机器学习方法，包括 k 近邻算法、随机预测、组合方法——所有的机器学习流行语。这些技巧当中的很多技巧的确可以使用 R 或 Python 库来实现，因此如果你不是世界领先的算法专家也没关系。更重要的是，能够了解大致概念并完全明白何时该使用什么样的技巧。

多元微积分和线性代数：在面试中，你可能需要讲述你在其他地方得出的机器学习或统计学结果。即使不会遇到这样的面试问题，面试人员也可能会问一些基础多元微积分或线性代数问题，因为它们是很多技巧的基础部分。你可能会疑问，市面上已经有很多开箱即用型 sklearn 或 R 实现，为何数据科学家还需要了解这些知识。答案是，在某些时候，数据科学团队可能有必要自己构建实现方法。对于产品受到数据的界定，预测效果的小小改善或算法优化就会为公司带来巨大效益的企业，了解这些概念最为重要。

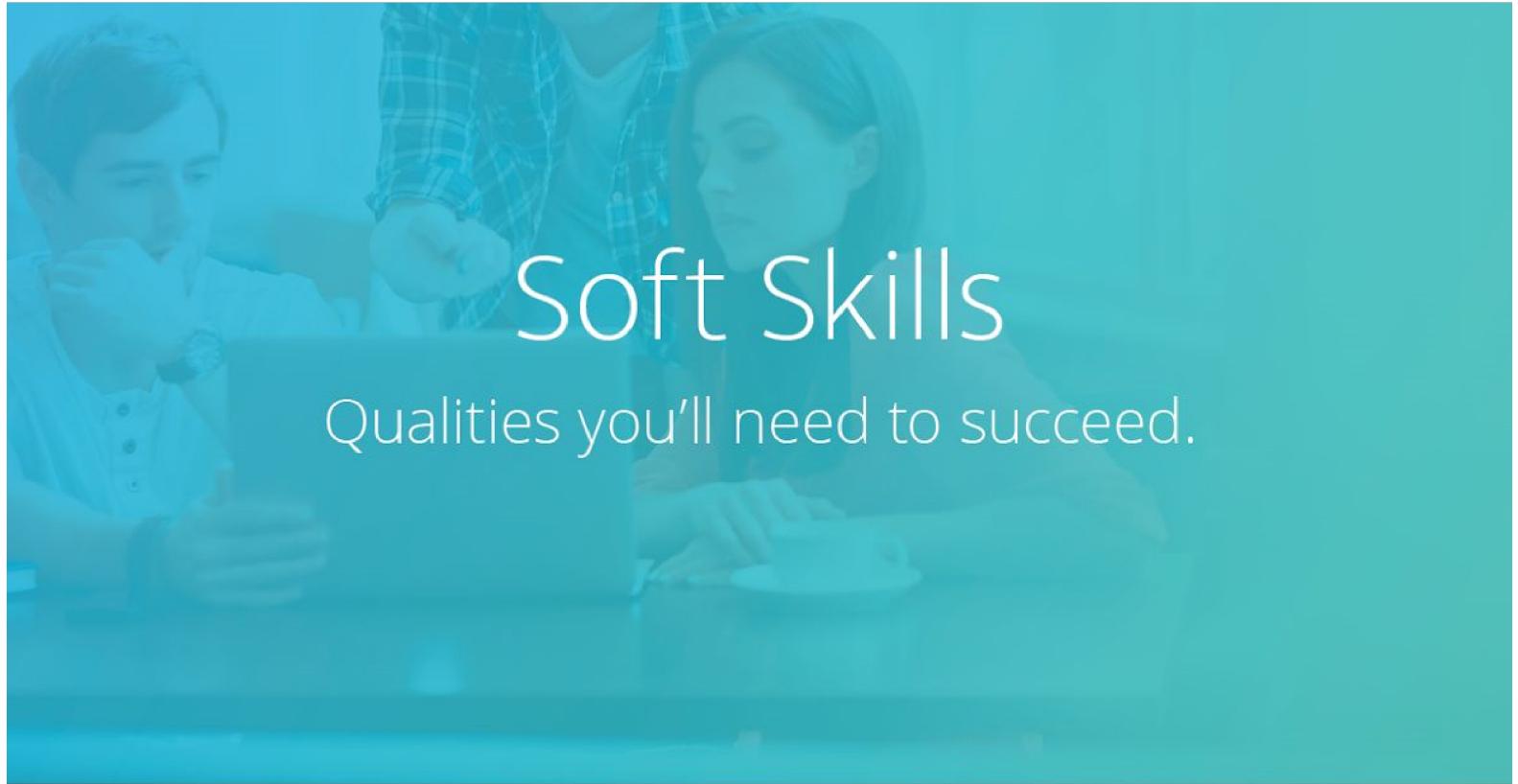
数据再加工：通常，你要分析的数据都很混乱、难以处理。因此，一定要知道如何处理数据中的缺陷。一些数据缺陷包括缺失值、字符串格式不一致（例如“New York”与“new york”与“ny”）、日期格式不一致（“2014-01-01”与“01/01/2014”，UNIX 时间与时间戳等等）。对于一些小型公司，你是首批数据人才，或者在以数据为依据的企业，产品与数据不相关（尤其是因为企业通常发展太快，并没有太注重数据整洁性），这种技能最为重要，但是所有人都有必要掌握该技能。

数据可视化和传达信息：可视化和传达数据异常重要，尤其对于一些新兴公司，它们首次做出以数据为依据的决策，或者数据科学家在公司内被视作可以帮助其他人做出以数据为依据的决策，更是如此。传达数据是指向受众描述发现结果或技巧（包括技术技巧和非技术技巧）的使用方法。对于可视化来说，熟悉 `ggplot` 和 `d3.js` 等数据可视化工具帮助非常大。不仅仅要熟悉可视化数据的工具，还要了解可视化编码数据和传达信息背后的规则。

软件工程师：如果你面试的是更小型的公司，并且是首批数据科学招聘岗位，那么有必要具有扎实的软件工程背景。你将负责处理各种数据日志记录，并且可能需要开发以数据为依据的产品。

像数据科学家一样思考问题：这类公司希望你是一名以数据为依据的问题解决者。意味着在面试过程中，他们可能会问你一些高级问题，例如关于公司可能想要运行的一项测试或想要开发的一款以数据为依据的产品。务必能够思考哪些事物很重要，哪些不重要。作为数据科学家，你应该如何与工程师和产品经理交流？你会使用什么方法？何时应该使用近似法？

数据科学正在被定义成一个领域。寻找工作很大程度上就是寻找企业需求和你的技能相匹配的公司，因为他们正在掌握这些技能。



Soft Skills

Qualities you'll need to succeed.

人们喜欢对一年发表看法。2017 年将是...之年？什么？数据？肯定不可能，因为我们讨论数据（无论是大数据还是其他类型的数据）已经有一段时间了。但是，Glassdoor 却将数据科学家评为 2016 年美国[最佳职位](#)。并非只有他们做出了这一评价。[CareerCast.com](#) 也将数据科学家列为最佳职位。这两大研究被广泛引用，使数据科学家成为现今最热门的工作之一，未来也会这样。那么如何为数据科学这一职业道路做好准备呢？

稀缺技能

[Crowdflower](#) 的团队在对数千个 LinkedIn 职位信息进行分析后，得出了你需要掌握的五大稀缺技能：SQL、Hadoop、Python、Java 和 R。关于这一研究得出的结论，我们可以阅读 Seamus Breslin 在 [KD Nuggets](#) 上发表的一篇文章，他在其中再次提到了 Crowdflower 的发现结果，并给出了一些精彩的个人观点。

例如，除了 Hadoop、Python 和 R 之外，他还添加了数据可视化、统计学，还有创造力！这很棒，稍后你会发现还很重要。

[Smart Data Collective](#) 团队也对 LinkedIn 进行了大量的研究，他们的发现结果很大程度上都和 Crowdflower 的结果相似，但是他们还对经验程度的不同进行了有趣的研究。例如：

“我们发现高级和首席数据科学家不太可能需要强调技术技能。相反，他们的技能可能包括商业智力、领导力、策略和管理能力。如果你想进入负责将以数据为依据的见解变成策略行动的岗位，那么则需要掌握所有这些技能。”

要查看他们的完整研究结果，可以在 [rjmetrics.com](#) 上阅读[《数据科学的状态》](#)这篇文章，你会找到很到位的概括性总结，其中将以下技能列为热门数据科学技能：数据分析、R、Python、数据挖掘和机器学习。

通用技能和独特技能

要深入了解数据科学技能，不妨阅读 Bob Hayes 在 [Business 2 Community](#) 网站上发表的《数据科学领域的十大技能》。Hayes 的研究既严谨又令人印象深刻，更重要的是，他按照职位角色来划分技能。例如，在自称具有业务经理角色的数据专业人群当中，86% 拥有项目管理技能。而自称承担开发工程师角色的人群当中，这一数据下降到了 46%。他在总结部分指出某些技能是通用技能，无论承担何种职位，其中包括：管理结构化数据、数学、数据挖掘和可视化工具，以及产品设计和开发。他还指出了特定于具体角色的技能，例如成功的开发工程师拥有编程技能，而研究人员具有机器学习技能。几乎每个角色都应掌握的首要技能是什么？沟通能力。数据必须有效传达给他人，这样才能有效地采取行动。

软技能和数据科学精神

事实上，沟通等软技能很明显对在数据科学领域取得成功来说很关键。Linda Burtch 是一名招聘高管并专长于量化业务科学，并且是 Burtch Works Executive Recruiting 的负责人。她在一篇广传阅的文章[《成为数据科学家必须掌握的技能》](#)中列出了一些相关技能，并专门强调了三大“非技术”技能：求知欲、商业敏锐性和沟通技能。她还在[她的网站上引用了 Frank Lo 的一篇博文](#)，Frank Lo 是 Wayfair 的数据科学总监，他将求知欲列为数据科学家应掌握的“首要无形”技能。他为何这么认为呢？

“数据科学的精髓是发现。”

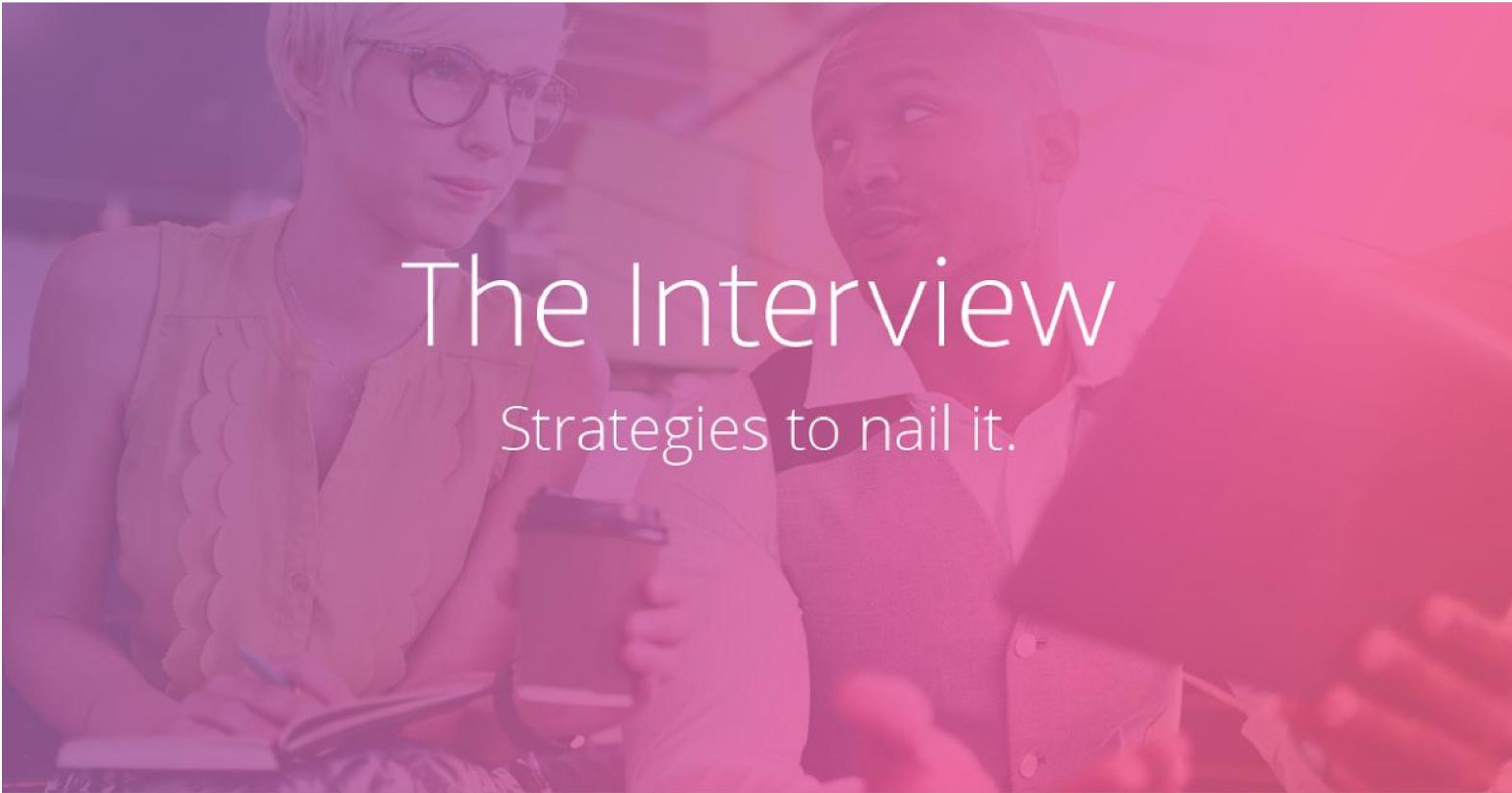
—Wayfair 数据科学总监 Frank Lo

Lo 还提出了以下劝诫性建议：

“不要根据学位来筛选应聘者。

有些人认为最优秀的数据科学家都拥有博士学位。根据我的面试和审核应聘者个人经验，我认为学术学位应该是最后考虑的因素。”

根据我们的上述所有分析，以及 Frank Lo 的观点，我们可以毫无疑问地支持以下做法：通过掌握关键技术技能和培养非技术技能并利用企业对合格人才日益增长的需求，自己摸索一条在数据科学领域走向成功的职业道路。



The Interview

Strategies to nail it.

在过去一年里，你不断磨炼自己的技能。晚上都会花时间完成作业和项目，不去陪朋友和家人，脏衣服堆的越来越高，有时甚至忙到深夜。你出于兴趣爱好为本地市政当局编写项目，并将自己的发现结果洋洋洒洒地写成各种精彩的帖子（或者至少你自己认为很精彩）。当朋友提到他们部门营销业绩不理想时，你的思绪却飘到别的地方，思考如何发现有意义的数据并处理结果。

现在你想找一份工作。虽然你在数据分析师这一领域投入了很多精力，但是最终结果很大程度上取决于面试过程。

毫无疑问的是，精心准备和了解流程预期不仅仅让你顺利通过数据分析师岗位面试，而且会让你脱颖而出，成为最合格的应聘者。

即使你现在没有在积极找工作，依然在学习技能并处理各种项目，现在也可以练习面试问题了，这样的话，在六到十二月之后，你将已经做好必要的准备工作，成功吸引潜在雇主的注意并赢得梦想的工作机会。

了解你的领域

Katie Malone 是一名物理学博士和优达学城的前讲师，参加过 Groupon 和湾区及芝加哥其他公司的面试。她经常会遇到以下几个基本问题：

- 你认为数据科学家是干什么的？
- 你认为数据科学家应该具备哪些最重要的技能？

精心准备的答案不仅体现出你对数据职业的兴趣和投入，而且展现了你的沟通技能。记住，面试你的人可能是没有技术背景的团队主管或 HR 总监，在这种情况下，你需要用最简单的术语解释某些概念。

Nick Gustafson 是优达学城的前数据科学家，指出了你要准备好深入讨论数据分析技能和工具。他建议准备好讨论类似于以下主题的内容：

- 对于给定的特定问题，使用哪个机器学习模型（例如分类与递归）
- 不同类型的分类模型和不同类型的回归模型之间的利弊
- 如何训练、测试和验证结果。控制模型复杂度的不同方式。
- 如何对无法观察的数量生成模型（例如使用贝叶斯方法，在使用该方法时如何选择先验分布）。
- 各种数值优化技巧（最大似然率、最大后验概率）
- 对于特定的业务需求，哪种类型的数据很重要，你会如何收集此类数据。
- 处理数据集中的相关特征，如何降低数据维度。

如果你发现自己无法回答这些问题，别担心。可以请求提供更多信息或相关示例。但是也要准备讨论理论知识。你需要熟悉擅长这一领域。

提前复习

对数据分析的各种工具和方法都能流畅自信地谈论表明你提前做足了功课。复习课程内容和笔记并阅读最新的科技博客及行业简报或许很有用。前优达学城数据工程师 Krasnoshtan Dmytro (现在就职于 Google) 通过熟练掌握以下概念来为面试做准备：

- 线性回归/多项式回归
- 决策树
- 降维
- 聚类

并经常阅读 [《Data Science Weekly》](#) 和 [《Machine Learning Mastery》](#)，以及通过参加 [Hacker Rank 竞赛](#) 和 [Kaggle 竞赛](#) 磨炼自己的技能。

复习以前的内容并继续锤炼和使用这些技能只会让你更深入地理解相关资料。

介绍自己

毫无疑问，面试人员会请你稍加介绍你做过的项目。正如 Katie Malone 提到的，潜在雇主始终会问这些问题。

你可以乘此机会讲述你对数据问题的解决方法，以及如何制作报告并分享你的结果。选一个你非常喜欢的项目，兴趣可以为演示加分。确保你能解释：

- 对于你要尝试解决的问题，为何选择了相关模型。
- 你的数据具有什么特征。
- 你是如何测试和验证结果的。
- 你从该项目中学到了什么。

并能够进行推断、简单地介绍你的技能，回答类似于下面的问题：

- 当你获得新的数据集时，你会如何处理该数据集以查看是否满足你对给定项目的需求？
- 你会如何处理大数据集？如果某个项目具有数十 G 或更多的数据，你会如何处理该项目？

了解公司

除了知识和技能之外，雇主还希望求职者能融入公司和企业文化。不言而喻，你需要努力调查你要面试的公司，不仅要查看他们的产品，而且要尽力了解他们的企业文化。想想你想要在他们公司工作的几个原因（稳定的薪水除外）。

能够回答以下问题：

- 你希望加入公司的哪个项目？
- 你会开始处理什么样的数据？
- 你觉得你可以给团队带来什么样的独特技能？

如果你能提供相关示例，那更好。

超越基本问题

自问自答一些数据科学面试中经常出现的问题并没有让朋友或同学向你提问这些问题效果强，模拟面试不仅让你能够组织和表达自己的想法，而且可以在一定的压力下锻炼自己（但是要为出现让人焦虑不安的面试这一可能性做好准备）。

联系你在这一领域认识的人，问问他们的面试过程如何，如果他们要寻找与你的技能对口的合适数据分析师，他们会提出什么样的问题。

Lewis Kaneshiro 是优达学城的前讲师和 [Streamlio](#) 的联合创始人。他不仅经历过让人精疲力尽的面试，而且面试过 KPCB Fellows 暑期实习求职者。在寻找出色的应聘者时，Lewis 会问：线性回归需要哪些假设？

“令人惊奇的是，多年来，这个问题出现在了多个面试中，让面试者能够区分出将线性模型看做“R/Python 中的函数”（甚至更糟糕“Excel 中的函数”）和能将这些模型应用到实际数据中的人士。”

在 Lewis 的大部分面试经历中，他都能够自信地讲述和演示（通过白板）这些假设。他还暗示包含将会与每个假设产生冲突的图表展示很重要。“很简单，但是学生如何忽略这些条件，则容易盲目地应用模型，而不会理解根本的使用案例，并且不会想到需要标准化、不齐量调整、异常值检测或其他现实问题。他们在实际工作中还会需要太多的监督。学生有时候认为招他们来就是要向数据中应用各种很酷的模型，但实际上有 90% 以上的工作都是通过线性模型和数据标准化/验证来完成的。”

数据科学面试

在与数据相关的职位面试期间你应该准备好讨论的问题和话题。

- 对于数据科学家来说，什么是最重要的技能？
- 对业务需求来说，哪种数据很重要？
- 你会专注于[企业内]的哪些数据并开始处理这些数据？
- 讨论各种数值优化技巧。
- 展示出对训练、测试和结果验证的理解。
- 解释不同类型的回归模型和不同类型的分类模型之间的利弊。
- 线性回归需要假设什么？
- 当你获得新的数据集时，你会如何处理该数据集以查看是否满足你对给定项目的需求？
- 你会如何处理大数据集？
- 如何检测异常值？
- 如何控制模型复杂度？
- 如何对无法观察的数量生成模型？

Lewis 强调在面试中展现出对数据有浓厚的兴趣、保持好奇心、具有不懈动力和坚韧不拔的任性很关键。

“我发现兴趣和激情加上好奇心和知识毅力每次都能超越简单的“才华”或“智力”。我们淘汰了那些拥有博士学位的纯粹学术人士，并选择拥有激情或愿意深入研究数据并处理混乱状况的本科毕业生。例如，我们正在考虑一名看起来并不是最优秀的夏季实习生，但是他的 Kaggle 方法和个人项目（通过收集公交抵达时间来预测实际抵达时间）表明他愿意处理混乱状况并坚持过来。这比想要发明新算法的优秀应聘者更实用。后者可能更适合博士项目，这样的话，可能就不是很好的员工。”

最后，Lewis 想知道被面试者如何检测异常值？

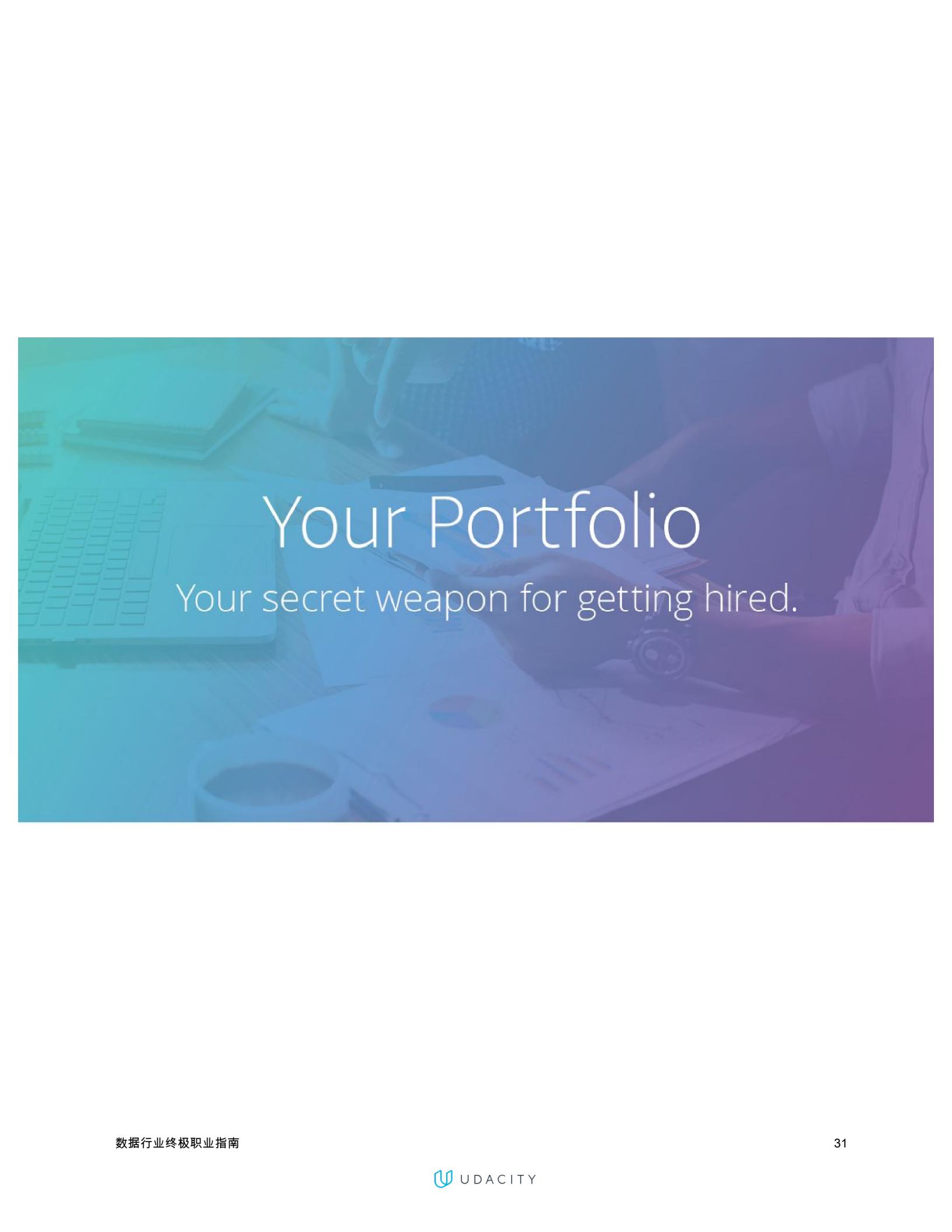
“我经历过的一些最佳面试都以这个问题开始并继续下去。它可以将只会应用算法来清理数据的学术人士与具有实际经验或对现实世界好奇的人士区分开来。花费大部分时间来理解原始数据并使用正确的方法和模型始终比疯狂地应用每个已知的 R 模型并检查准确性要强。我认为这是最有经验的面试者会提出的问题。”

要回答这个问题，Lewis 建议查看 Sebastian Thrun 和 Katie Malone 的 [《机器学习简介》](#) 中介绍的实用方法：“使用 LR 将 [预测模型拟合到数据集中]，排名错误，并抛出前 10% 的错误，然后再迭代地自拟合以检测稳定性。”

底线

如果你问一百位不同的数据专业人士他们在面试中遇到了什么样的问题，可能会获得一百个不同的答案。幸运的是，你自己并不需要收集这一信息。有无数的[清单](#)和[资源](#)中包含大量的示例问题，你可以拿来做练习，并为面试当天[做好准备](#)。

熟练掌握你的领域、复习项目并重新激起对数据的兴趣将对成功面试非常有帮助。赶紧练习吧，测验自己、挑战自己，并讲述处理过的问题，希望你会在面试过程中表现优异，并拿下一直期待的 offer。



Your Portfolio

Your secret weapon for getting hired.

在优达学城的就职 [Intersect 峰会](#) 上，一名学员询问如何掌握数据分析经验才能找到一份工作。优达学城的高级课程开发人员 Mat Leonard 给出的建议是创建小型数据分析项目，然后放到网上，[他自己的博客](#)就是采取这种方式。他指出，小型项目可以帮助你深入了解分析方法或学习新的技巧。将项目放到网上可以构建作品集，向潜在雇主展示你可以成功地回答数据方面的问题。

下面是 Mat 提出的一些关于如何成功地构建数据分析作品集并帮助你拿下 Offer 的策略。

借助几项技术就可以轻松地将项目上传到网络上。

首先，Jupyter notebook 是组合文字、代码和图片的精彩工具。

记事本可以转换为标记文件以用在网络框架中，例如 Pelican。最后，你可以将博客免费托管到 GitHub 上。为此，你需要对 git 和 GitHub 有基本的了解：如何 stage、commit 和 push 更改。你可以在优达学城的[版本控制课程](#)中了解 git 和 GitHub。你还需要熟练掌握命令行工具，你可以在[此处](#)进行学习。

构建博客

要构建博客，你可以使用 [Pelican](#)——一款用 Python 编写的静态网站生成器。Pelican 使用标记文件作为博客帖子，并自动创建归档、类别和标签。Pelican 提供了多种主题背景，你也可以自己创建一个主题背景并展现个人风采。

The screenshot shows a blog post on the 'Matatat.org' website. The title of the post is 'A/B Testing with Sampyl'. Below the title, it says 'Jan 31, 2016 | tags: Statistics Python Sampyl'. The main content of the post discusses the author's work on building a Bayesian model to infer the firing rate of neurons over time, mentioning the use of PyMC and Theano packages.

此外，因为 Pelican 创建的是静态网站，你可以将其免费托管到 GitHub 上。

要详细了解如何设置自己的博客，请点击[此处](#)，同时注意，你需要拥有[Git 和 GitHub](#) 经验，以及 Python 和 shell 命令行经验。

你的数据项目：从何处开始

要为你的新网站填充内容，你需要完成一些小型数据项目。Kaggle 会举行[数据科学竞赛](#)，对你来说是一个不错的开始。这些数据通常已经清理过，所以你可以专注于构建模型。Kaggle 还托管了各种与竞赛无关的[数据集](#)供你探索，例如希拉里的电子邮件。其他资源包括[datasets subreddit](#)、[Data.gov](#)，以及可以在此[Quora 话题](#)下找到的很多其他资源。很多城市都有公共数据集，例如[旧金山](#)和[纽约市](#)。你还可以通过[抓取网页数据](#)来收集数据，就像我在[Yelp](#) 上收集数据一样。

最重要的是寻找你感兴趣的数据，并提出你想要回答的问题。利用这段时间学习新技巧，并深入研究你已经知道的方法。处理小型数据项目并上传到网络上非常有助于你实现职业发展目标，尤其可以向雇主展示你喜欢处理数据，并且希望能继续提高自己的能力。希望在我的上述指导下，你能够创建自己的在线作品集并向全世界展示你最精彩的作品！

Your First Data Job

What it will be, and how to get it.

无论你处在数据这一职业道路的哪个阶段，思考为了面试需要准备的所有技能依然让人生畏。通常，数据工作者来自以下三大不同的背景领域：

1. 从零开始
2. 扎实的编程背景
3. 扎实的数学背景

根据你的起点，你认为哪条道路最适合成为你的首份数据科学工作？你可以使用哪些技能以最高效的方式奠定基础？

研究这三种情形（零经验、有编程经验但是没有数学经验、有数学经验但是没有编程经验）很有用，可以了解掌握终极数据技能需要具备的基础能力。

从零开始

你想成为一名数据分析师？恭喜你！你选择了一个报酬丰厚、工作地点灵活并且非常稳定的职业，在未来的很多年都会继续强劲发展下去。

当然，前期需要刻苦学习并磨炼必要的技能，然后才能从中受益。下面是一些建议，可以帮助你掌握相关工具并成为炙手可热的数据分析师。

编程

编程是进行数据分析所必须具备的技能。它是让数据分析师超越业务分析师之上的核心技能。你需要熟练掌握一到多门编程语言，从 Python 或 R 开始，并充分掌握最常用的数据科学库和程序包（例如 ggplot2、reshape2、numpy、pandas 和 scipy）的发展状况。

统计学

如果拥有出色的编程技能，但是不会解读数据，那又有什么用呢？对数据分析来说，了解统计学知识很关键，包括统计学测试、分布和最大似然估计值。

描述统计学和推论统计学都要掌握。前者是指描述样本特性的量化方式；后者是预测分析，通过解读样本推断出更大的总体的特性。你需要知道一些基础知识，很多在高中或大学都学过（均值、中值、众数；标准偏差和方差；假设检验），并在此基础上掌握更复杂的统计学技能（不同类型的数据分布：标准常态、指数/泊松、二项式、卡方，以及显著性检验：Z 检验、t 检验、惠特尼 U 检定、卡方分布、方差分析）。

除了描述统计学和推论统计学之外，数据分析师还要擅长统计学实验设计。也就是系统地选择参数以便得出有效且具显著意义的结果。例如，你需要决定要收集多少个样本、每个因素之间应如何相关联、如何选择好的对照组和检验组，等等。要使用 A/B 检验等工具和幂次定律等概念执行强大的实验设计，最佳做法是使用“SMART (Specific (明确)、Measurable (可衡量)、Actionable (可行动)、Realistic (现实) 和 Timely (及时)) 实验”这一理念。

数学

数据分析师的语言就是数字，因此拥有扎实的数据基础是成为数据分析师的基石。

你至少应该熟悉大学代数知识。你需要将以前所认为的“应用问题”（现实中等同于业务问题）转变成数学表达式；你要能够操纵代数表达式并求解方程式；并且要能够绘制不同类型的函数，深入理解函数图表与等式之间的关系。

此外，熟练掌握多元微积分和线性代数对数据分析师很有帮助。思考：矩阵运算、点积、特征值和特征向量，以及多元导数。

机器学习

多元微积分和线性代数以及统计学构成了[机器学习](#)的基础，使数据专业人士能够根据大批量的数据提出预测性或计算性建议。作为数据分析师，你不需要发明新的机器学习算法（这一高级技能可以使你成为合格的数据科学家），但是你应该了解这些领域的常见概念。一些示例包括主成分分析、神经网络、支持向量机和 K 均值聚类。注意，你可能不需要了解这些算法背后的理论和实现细节，但是应该了解各自的利弊以及何时该（以及何时不该）将它们应用到数据集中。

数据分析师需要知道三大类型的[机器学习](#)：监督式学习、非监督式学习和强化学习。

在监督式学习中，“学习者”（计算机程序）被提供了两个数据集：训练集和检验集。计算机从训练集中的加标签示例中“学习”规律，从而能够准确地在检验集中识别无标签示例。目标是让学习者能够发现规则，并帮助识别检验集中的元素。监督式学习让手机能够识别你的声音，电子邮箱能够过滤垃圾邮件。你将用到的具体工具包括：

- 决策树
- 朴素贝叶斯分类
- 普通最小二乘回归
- 逻辑回归

- 神经网络
- 支持向量机
- 以及组合方法

当你需要从给定的“无标签”数据集中发现隐含的关系并得出隐藏结构时，就要用到非监督式学习。非监督式学习使 Netflix 能够向你推荐你喜欢的电影，使 Amazon 能够预测你会喜欢的商品。你将用到的具体工具包括：

- 聚类算法
- 主成分分析 (PCA)
- 奇异值分解 (SVD)
- 和独立成分分析 (ICA)

最后，强化学习适用于位于监督式和非监督式学习极端示例之间的情形，即对于每个预测性步骤或操作，都有某种形式的反馈，但是没有精确的标签或误差测定。如果你想要判断如何最大化奖励（例如在机器人控制、国际象棋、双陆棋、跳棋和其他软件代理可以学习的竞技活动中），则可以应用强化学习。你将用到的具体工具包括：

- Q 学习
- TD 学习
- 和遗传算法

数据整理

成为数据分析师要掌握的最后三大重要技能是操纵、展示和解读数据。要将原始资料转变成有条理的实用数据集，数据整理（亦称为“数据挖掘”）就派上用场了。这一流程是指收集和清理数据，以便轻松地探索和分析数据。

你需要掌握信息存储的中心枢纽——数据库系统（包括基于 SQL 的数据库和不急于 SQL 的数据库）。如果熟悉关系数据库（例如 PostgreSQL、mySQL、Netezza 和 Oracle 以及 Hadoop、Spark 和 [MongoDB](#)），则很有用。

对数据整理来说很关键的其他概念和工具包括正则表达式、数学变换和 Python 字符串库（用于字符串处理）。你还需要了解如何解析常见文件格式（例如 csv 和 xml 文件），以及如何通过 log-10 变换将非正态分布转换为正态分布。

现在听起来可能会觉得压力很大，尤其是当你不仅不了解所涉及的技能，而且对某些术语本身都不熟悉。注意，所有这些技能都是可以叠加的：每掌握一项技能，都能帮助你学习下一个技能，然后继续学习下一个技能，直到你成为完全合格的数据分析师并功成名就。

数据可视化

清理、整理、组织、整合并解释数据之后，你要能够直观地展现你的发现结果，以便利益相关者（包括对数据一窍不通的人士）能够完全理解你的发现结果。如果你无法清晰有效地传达你的想法，那么你的数据分析繁琐工作不会受到任何认可。

熟悉诸如 ggplot、matplotlib、sea born 和 D3.js 等[数据可视化](#)工具将会有帮助。当然，不仅仅要熟悉实际地展示数据的工具，还要了解可视化编码数据背后的原理。要取得成功，你需要完全了解业务背景，以便知道如何调整你的数据可视化结果，使其具有最高的关联性。

数据直觉

在了解了上述技能的技术知识后，你需要知道如何思考、如何提出正确的问题。你可以在之后的整个人生中分析一个数据集，并以各种格式可视化你的解释，并展现出大量的结果。现实情况是，你只有有限的时间和空间来处理同事的问题，并分析手头的数据。因此，一定要形成一种观念：哪些事情重要，哪些不重要。

努力深入了解你所在的领域，无论是股票市场还是消费性包装产品。花时间尽可能了解更多的数据集，例如参加 Kaggle 竞赛，了解如何避免死胡同。学习发现所分配任务“问题背后的问题”，深入挖掘，换句话说，发现推动分析数据这一需求的精确业务问题。

掌握编程背景知识

要应聘数据分析师职位，接着你需要掌握以下知识。

基础知识

- 统计学：你需要通过对不同类型的分布运用正确的方法、技巧或统计学检验，严格地解析、推导和比较不同类型的数据。请参阅上述关于具体工具和技能的分类。
- 概率学：为了得出准确的结论，数据分析师要能够推理事件曾经或未来发生的概率。请参阅上述关于具体工具和技能的分类。

高级知识

- 多元微积分/线性代数：这些高级数学技能比统计学和概率学知识的重要性要低，但是如果你想知道机器学习的实际工作原理，那么了解这些知识肯定很有用。此外，如果你的预想是某一天从数据分析师发展成为数据科学家，那么多元微积分和线性代数将有助于你构建自己的算法。

掌握数学背景知识

或许你是一名数学天才，但是没有编程背景。下面的指南可以逐步指导你掌握对数据分析师来说很关键的编程知识。

基础知识

- 变量、控制流、循环、函数：这些编程的基本概念。请了解并爱上这些概念
- 调试：你的代码第一次可能无法正常运行，或者出现意外情形时代码就会崩溃。发生这种情况时，你需要弄清楚问题出在哪，为何会发生。这时候调试技能就派上用场了。

- 面向对象的编程：学习如何将代码组织成面向对象的设计模式，这样才能轻松地重复利用、测试和与他人分享代码。

高级知识

- 数据结构：如果能熟悉堆栈、队列、列表、数组、哈希图、优先队列、try 语句和图表，那就更好了。在某些情形下，一种数据结构会比其他结构（在内存使用情况和运行时效率方面）更好，如果你能理解这些关系，则能够优化你的程序，使运行速度更快、效率更高。这样就会让你的团队刮目相看，并在众多数据专业人士当中脱颖而出。
- 算法：知道在何种情形下运用何种算法可以将程序的运行时间从几天缩短到几个小时，或者将内存占用量从几个 G 降到几百兆。努力学习分而治之 (D&C) 算法、贪婪算法、动态规划、线性规划和图形算法（深度与广度与遍历、最小生成树和两个节点之间的最短路径）。
- 软件设计模式：想要让你的代码稳健、可重复利用并可测试？很多先驱型软件工程师和计算机科学家开发了软件设计模式，可以帮助你实现这些要求。请熟练掌握这些工具，从而擅长于数据分析。

底线

数据分析是一个快速发展的领域，有很多声音和你一样渴望知道该学习什么，按照什么样的顺序学习。各种不同的信息会令人困惑、感到畏惧并让人气馁。

从我们推出的这份可靠指南中得出你需要真正学习哪些技能以便获得第一份数据分析师工作，以及从何处开始，取决于你的具体背景。

无论你是刚刚起步还是想拓展现有技能，数据分析师要投入的精力都很多，但是我们保证，回报却更高。

The Future of Data

A look at predictive analytics.

以前，数据分析经常都是回顾过去。发生了什么，意义是什么？这种模式已经完全过时了。对于现代企业掌握的大批量数据，数据分析的一切都是向前看。有什么样的预期、如何朝前发展？我们进入了预测分析时代。

预测分析

人们对预测分析给出了很多精彩的定义，某些偏于技术，某些语言表述相对平淡些。以下分别给出了精彩的示例：

“预测分析描述的是任何具有以下四大特性的数据挖掘方法：

1. 强调预测性（而不是描述、分类或集群）
2. 几小时或几天内就能完成的快速分析（而不是常规的数个月传统数据挖掘）
3. 强调发现结果与企业的关联性（而不是象牙塔一样的学术分析）
4. 越来越强调易用性，因此让企业用户能够使用这些工具

—Gartner IT Glossary

“预测分析是一门使用数据做出信息更充分的决策的科学方法。预测分析有助于发现数据中的隐藏规律和联系，从而帮助我们更自信地预测未来可能会发生的情况，并为企业提供更有价值、可以采取行动的见解。”

—Predictive Analytics for Dummies

越来越多的企业将预测分析应用到数据策略中。[MarketsandMarkets](#) 目前的一项研究“预测全球预测分析市场将从 2015 年的 27.4 亿美元增长到 2020 年的 92 亿美元，复合年增长率 (CAGR) 达到 27.4%。”[TDWI Research](#) 的一项新研究指出了企业为何使用预测分析的五大原因：

- 预测趋势
- 理解客户
- 改善业务绩效
- 推动策略性决策制定
- 预测行为

预测分析很明显具有强大的吸引力，能够形成这种可以采取行动、具有前瞻性的分析快速成为品牌战胜竞争对手的关键因素。但是，这些新方法产生了一个关键现实：现今的有效数据分析需要两大要素：工具和会使用工具的人。报名参加 Udacity 的[业务分析师纳米学位项目](#)吧。

优达学城业务分析师纳米学位课程

市面上有越来越多的优秀工具可以帮助企业运用预测分析。该纳米学位课程的合作伙伴之一——Alteryx 就是领先的提供商。我们喜欢他们的解决方案有多方面的原因，尤其是使用他们的工具不需要编程经验：

要在竞争中拔得头筹并提高结算盈亏的底线，企业需要了解未来的发展状况，以便准备更充分，盈利能力更强。预测分析使企业能够通过分析当前和历史数据来提高竞争优势，更好地了解和预测未来事件，通过更准确的决策改善业务成果。

遗憾的是，大多数预测分析工具都比较复杂、耗费时间，并需要编程或数据科学及统计学技能。Alteryx 对编程技能没有要求，降低了预测分析的复杂性，使数据分析师和数据科学家都能掌控整个预测分析周期。

我们构建业务分析纳米学位课程是为了帮助你实现两大目标：

1. 你将学会在任何涉及做出有效的以数据为依据的业务决策的岗位上取得成功所需的必要技能。
2. 你将熟练掌握两个软件工具包：Alteryx，一款使分析师能快速准备、混合和分析数据的工具；以及 Tableau，一款强大的数据可视化工具。

为什么要学习 Udacity 业务分析师纳米学位项目？

要回答这一问题，我们想请你回忆下我们在前面提到的两点。首先，预测分析很热门。这一市场在不断发展，越来越多的企业加入进来。其次，优秀的工具需要熟练的用户。因此，随着对这些工具的需求增多，对会高效成功地使用这些工具的人才需求也增多。

作为业务分析师纳米学位项目的毕业生，这时候你就可以把握机会了。

借助你将掌握的技能和将展现的软件熟练度，你将很快就可以对选择加入的企业带来价值，从实施软件解决方案（例如 Alteryx 或 Tableau）到生成可以采取行动的预测性数据（这些数据将显著改善未来企业发展状况）。



Conclusion

正如我们在这份指南的简介中提到的，数据这一职业既是安全可靠的选择，又是一个前沿职业道路。在读了我们的所有见解之后，希望你能清楚地明白数据行业可以提供各式各样的入门起点。虽然无论是何种职位，你都需要掌握特定的技能和工具，但是该领域涉及的范围如此广泛，可以向各种经验水平和侧重领域的人士敞开大门。

我们的[数据分析师纳米学位项目](#)专门为你进入这一异常热门的职业发展领域做好准备。我们与 Facebook 和 mongoDB 等领先公司的专家合作，开发了内容广泛的课程表，几乎涵盖了我们在此指南中提到的所有主题内容，从使用 Python 清理数据和使用 R 探索和总结数据，到创建可视化图表及运行 A/B 测试，无所不包。我们还会讲解特定的领域，例如机器学习和预测分析。当然，如果你想深入学习这些领域，我们也推出了专门的课程！

我们的课程以项目为基础，将帮助你不仅学会成功所需的稀缺技能，并且可以构建帮助你找到工作的作品集。

"我因为缺少相关经验，找工作碰了很多壁。我在这门课程中构建的项目让我的简历很出彩，并帮助我找到了工作。" —Soumya Ranjan Mohanty，数据分析师纳米学位项目毕业生，现在是 Gramener 的一名数据专员

希望你能将从本指南中读到的知识运用到你所选的职业道路上。有一点可以确定的是，掌握数据技能在很长一段时间内都会是抢手的人才。如果你想选择数据这一职业道路，那么可以阅读这份指南，并且我们的数据分析师纳米学位项目会帮助你走向成功。

准备好加入抢手数据专家的下一代人才大军了吗？马上加入我们与 Facebook、mongoDB 和 Tableau 官方联合打造的[数据分析师纳米学位项目](#)吧！