

第1章

数据分析概述

1.1 何谓数据分析

数据是指以时间为轴，记录人物、地点、事件和方法等生活各个维度的数字字符。数据会随着时间不断累积，也会随着科技、生活观念等变化而呈现出不同的特性。消费者去商场用现金支付的方式购买了一件心仪的衣服，商店的日销售报告中记录了此次交易的金额、数量、款式和型号。当消费者采用的是刷卡的支付方式时，银行的日流水单以及商店的 Pose 机刷卡记录就产生了一笔实时交易数据。如果该消费者还是这家商店的会员，那么该商店就拥有了该消费者部分基本信息以及多次购买产品的交易记录。随着互联网、自动化科技的发展，消费者更多地参与了线上交易，那么线上交易平台会产生消费者常用地址、联系方式、偏好产品、产品型号、消费额度和消费频率等全面而及时的消费数据。

在传统的商业和社会环境下，人们对于数据的利用是非常有限的。企业和商家们利用自身的营销数据汇编成财务报告、信息披露报告，用来为管理层或者潜在投资者提供企业经营状况的参考资料。政府各个机构拥有的人口、宏观指标、地区发展、部门业务发展等各方面的数据为定期政府报告等特定事项提供数据服务。数据成为我们衡量过去发展状况和业绩水平的一种度量衡。传统意义上对于数据利用存在的缺失是不可忽视的。

首先，传统意义上对于数据的利用形成了无数个数据孤岛。宏观数据、调研数据、社会化数据和企业数据之间存在数据孤岛，而政府内部和企业内部同样存在数据孤岛。各个政府部门、甚至每个政府部门内部都有自身因专门的业务内容而产生专项数据，但是这些专项数据仅仅在服从专项需求时才被局部调用。企业内部也是这样，除了必要的信息披露之外，企业各部门之间的数据也是缺乏协同和共享机制的。数据孤岛的存在比

我们想象得还要多而广泛，也极大地降低了对数据的利用。

其次，传统数据存在缺失和错误的比率较大。在商务系统和互联网尚未发展的情况下，企业和社会运营的数据很多是通过人工的方式进行记录的，因此很容易存在数据缺失和数据失实的情况。更重要的是，很多数据的产生需要大量的人力、物力去完成，在不可估量商业价值的情况下，数据的累积往往具有很强的延时性。

再次，数据的价值被低估，缺乏专业的数据分析人员对数据的商业和社会价值进行分析。我们会发现，传统意义上的数据管理是基于某些特定的目的和需求，例如，定期的信息披露，盈余管理和预测等。但是这些目的和需求都不是为了能够创造价值而设立的，更多的是一种业务和管理层面的辅助。在缺乏商业利益动机的前提下，也就没有专业数据分析师存在的必要。

然而，大数据时代拥有的数据量是足够大的。在互联网的世界里，每分钟 Facebook 平均有 600 次的访问量，并有新增用户 28 万；Amazon 每分钟销售高达 8.3 万美元；全球 IP 网一分钟能够传输 639TB 的数据；你需要花费 5 年的时间才能看完互联网上一秒钟传输的视频。同时，大数据时代的数据开始逐步走向多元化的趋势。数据来源包括移动数据、店面交易、网络行为、定位信息、电商、用户调查、社会网络以及企业 CRM 等。

当今社会是一个大数据的社会，信息高度发达，数据信息更是呈爆炸式增长，每天全世界都在产生着巨大的数据，大到一个跨国公司，小到一个社区的小卖部，都不可避免地要与各种数据打交道。面对众多的数据，无论是管理者、经营者还是政策的制定者，都面临着管理好数据、发现数据的规律以及从数据中获得价值的问题。在这一章中，我们将首先探讨什么是数据分析。

1.1.1 数据分析定义概述

数据分析是数据的摄取，进行数据分析离不开数据的支持。数据实际上是一种观测值，是信息的外在表现形式，也是实验、策略、观察和调查的结果，以数量形式来表现。原始的数据往往具有数量巨大且杂乱无章的特点，很多时候给人的感觉就是让人眼花缭乱，不知所云。这样的数据是没有任何意义的，需要对其进行分析。

那么数据分析就是将大量且杂乱无章的数据进行整理、归纳和提炼，从中寻找出数据的内在规律，从而获得需要的信息。数据分析的过程，实际上就是对数据进行汇总和理解吸收的过程，也是为了提取有用的信息和形成结论而对数据加以研究和概括总结的过程。通过对数据进行分析，以求最大化地开发数据，发挥数据的作用。

数据分析是一种有组织有目的处理数据并使数据成为信息的过程，其根本目的是集中、萃取和提炼。在实际工作中，其最终是为了帮助经营者和决策者做出判断，以便采取正确有效的行动。例如，企业的高层希望通过市场分析和研究，把握当前产品的市场动向，从而制订合理的产品研发和销售计划。因此，在经济生活中，经济决策实际上就是一种“数据决策”，“用数据说话”是众多企业经营者和决策者的共识。

大数据时代，数据的形态和体量都发生了很大的变化，缺乏数据分析的数据本身是

不具备商业价值的。数据分析的确能够为大数据时代带来质的飞跃。常规报表、查询、多维分析、警报——这数据分析的前四个等级都只能展示已经发生的历史状况，但是数据分析不仅仅如此。统计分析能够帮助我们找到触发事件发生的相关因素、确认最为有效的潜在交易方案。预报可以告诉我们未来股市预期变动或者是企业未来盈利水平预期。预测建模可以帮助金融机构预测新的金融产品的潜在客户。运筹优化能够帮助企业在限定的条件下把握最优的业务机会。

数据分析的核心思路就是要与实际业务、商业目的和运营目标相结合，进而为社会、经济和个体创造价值。数据分析与业务流程相结合可以体现为五个基本步骤，包括认知、运营、交互、销售和维持。商业运营要与数据分析的关键指标紧密联系，用数据提高产品市场营销效率和推广效率。大数据的维护和累积能够为商业运营描绘完整的企业画像、客户画像。大数据画像包括了解企业或者客户的基本信息、需求倾向、用户行为等。通过追踪核心的数据指标，进一步完善企业或者客户画像，进而将其转化成为产品元素和营销战略。通过数据分析，我们可以知道通过什么渠道以最小的成本将竞争对手的客户转化为自身的客户，进而创造营业收入。通过大数据与运营维护的结合可以很大程度上提高客户满意度，降低客户的流失率。

目前数据分析实践的运用主要体现在物联网、定位服务、客户制成以及反欺诈领域。首先是物联网领域。以 UPS（美国快递公司）为例，UPS 每天通过 5 万台快递车派送约 1630 万个包裹。UPS 在每台快递车上都安装了传感器，并且通过传感器传输数据分析，制定每天每台车少跑一英里的运营战略，该战略为 UPS 每年实现了约 3000 万美元的盈利。其次是定位服务。以美洲银行为例，美洲银行为其客户提供汉堡王的优惠券。该优惠券以美洲银行客户刷卡记录数据为基础，判断汉堡王潜在竞争对手的客户，并对这些客户进行了定向、定位的优惠券推送。该项营销战略既维护了美洲银行客户，也为汉堡王实现了创收。再次是客户支撑。通过文本挖掘、自然语言处理、情感分析等手段，对客户评论、客户投诉、海外舆情、媒体报道数据进行分类处理，进而充分掌握客户潜在的需求，达到及时有效维护客户的商业目的。最后是反欺诈领域。最典型的例子就是保险公司骗保。我们可以通过神经网络分析等多元的数据分析方法及时识别和判断已有的欺诈模式和潜在的欺诈人群，进而有效地进行客户管理，确保企业运营和效益。

在传统的数据分析模式下，通常是先提出假设检验，然后带着问题去进行数据分析。在大数据时代下，更重要的是关注小数据完善和收集的同时，构建完善的数据交互平台。在先有数据的基础上，在数据中找寻新的思路和创新机遇，进而实现价值的飞跃。在数据爆炸和新媒体时代的背景下，文字、图片、视频、网络数据等新兴的数据模式使得我们需要掌握和运用全新的数据处理方式。同时，还需要对数据进行生命周期的管理，对非结构数据进行筛选和标签化。数据分析看重的是数据的多元性和数据的质量，需要构建起大数据谱系，同时结合数据的特性采用不同的数据分析方法、分析工具和分析模型。因此，数据分析需要较为综合的思维和能力。

1.1.2 数据分析方法

对数据进行分析的方法有很多,归纳起来主要包括统计分析方法、运筹学分析方法、财务分析方法和图表分析方法。下面分别简单介绍以上所述分析方法。

(1) 统计分析方法:是指对收集到的数据进行整理归类并解释的分析过程,主要包括描述性统计或推断性统计。其中,描述性统计以描述和归纳数据的特征以及变量之间的关系为目的,主要涉及数据的集中趋势、离散程度和相关程度,其代表性指标是平均数、标准差和相关系数等。推断性统计是用样本数据来推出总体特征的一种分析方法,包括总体参数估计和假设检验,代表性方法是 Z 检验、T 检验和卡方检验等。

(2) 运筹学分析方法:是在管理领域中运用的数学方法,该方法能够对需要管理的对象(如人、财务和物等)进行组织从而发挥最大效益。运筹学分析常使用数学规划分析,如线性规划、非线性规划、整数规划和动态规划等,也可以运用运筹学中的理论(如图论、决策论和库存论等)来进行分析预测。运筹学分析方法常用在企业的管理中,如服务、库存、资源分配、生产和产品可靠性分析等诸多领域。

(3) 财务分析方法:是以财务数据及相关数据为依据和起点来系统分析和评估企业过去和现在的经营成果、财务状况以及变动情况,从而了解过去、分析现在和预测未来,达到辅助企业的经营和决策的目的。财务分析方法包括比较分析法、趋势分析法和比率分析法等。

(4) 图表分析方法:是一种直观形象的分析方法,其将数据以图表的形式展示出来,使数据形象、直观和清晰,让决策者更容易发现数据中的问题,提高数据处理和效率。图表分析主要针对不同的数据分析类型,采用不同的图表类型将数据单独或组合展示出来,常见的图表如柱形图、条形图、折线图和饼图等。

1.2 数据分析步骤

数据分析通常可以分为明确目的、收集数据、处理数据、分析数据和撰写报告这几个步骤。

1.2.1 明确目的

在进行数据分析时,首先需要明确分析的目的。在接收到数据分析的任务时,首先需要搞清楚为什么要进行这次分析、这次数据分析需要解决的是什么问题、应该从哪个方面切入进行分析以及什么样的分析方法最有效等问题。在确定总体目的后,可以对目标进行细化,将分析的目标细化为若干分析要点,厘清具体的分析思路并搭建分析框架,搞清楚数据分析需要从哪几个角度来进行,采用怎样的分析方法最有效。

只有这样才能为接下来的工作提供有效的指引，保证分析完整性、合理性和准确性，使数据分析能够高效进行，分析结果保证有效和准确。

总之，在开展数据分析之前，要考虑清楚，为什么要开展数据分析？通过这次数据分析要解决什么问题？只有明确数据分析的目标，数据分析才不会偏离方向，否则得出的数据分析结果不仅没有指导意义，甚至可能将决策者引入歧途，导致严重的后果。

1.2.2 收集数据

收集数据是在明确数据分析的目的后，获取需要数据的过程，其为数据的分析提供直接的素材和依据。收集数据的来源包括两种方式，第一种方式就是直接来源，也称为第一手数据，数据来源于直接的调查或现实的结果。第二种方式称为间接数据，也可称为第二手数据，数据来源于他人的调查或实验，是结果加工整理后的数据。通常，数据来源主要有以下几种方式：

(1) 公司或机构自己的业务数据库，存放着大量相关业务数据。这个业务数据库就是一个庞大的数据资源，需要有效地利用起来。

(2) 公开出版物，比如《中国统计年鉴》、《中国社会统计年鉴》、《世界经济年鉴》、《世界发展报告》等统计年鉴或报告。

(3) 互联网，网络上发布的数据越来越多，特别是搜索引擎可以帮助我们快速找到所需要的数据。例如，国家及地方统计局网站、行业协会网站、政府机构网站、传播媒体网站和大型综合门户网站等，上面都可能有我们需要的数据。

(4) 进行市场调查。在数据分析时，如果要了解用户的想法和需求，通过以上三种方式获得数据会比较困难，这时就可以采用市场调查的方法收集用户的想法和需求数据。市场调查是指运用科学的方法，有目的、有系统地收集、记录、整理有关市场营销的信息和资料，分析市场情况，了解市场现状及其发展趋势，为市场预测和营销决策提供客观、正确的数据资料。市场调查可以弥补其他数据收集方式的不足，但进行市场调查所需的费用较高，而且会存在一定的误差，故仅作参考之用。

所以，在实际工作中，获取数据的方式有很多，根据不同的需要有不同的获取途径，如对本公司的经营状况的分析，可以从公司自由的业务数据库获取。对于一些专业数据，可以从公开的出版物获取，如年鉴或分析报告等。随着互联网的发展，获取数据的途径更为广阔，通过搜索引擎，可以快速找到需要的数据，如到国家或地方统计局的网站、行业组织的官方网站或行业信息网站等。

1.2.3 处理数据

在获得数据后，需要对数据进行处理。数据处理是指对收集到的数据进行加工整理，形成适合数据分析的样式，它是数据分析前必不可少的一步工作。数据处理的基本目的是从大量的、杂乱的且难以理解的数据中抽取并推导出对解决问题有价值

意义的数据。

数据处理常常需要对数据进行清理、转换、提取、汇总和计算。通常情况下，收集到的数据都需要进行一定的处理才能用于后面的数据分析工作，即使是再“干净”的原始数据也需要先进行一定的处理才能使用。

1.2.4 分析数据

数据分析需要从数据中发现有关信息，一般需要通过软件来完成。在进行数据分析时，数据分析人员根据分析的目的和内容确定有效的数据分析方法，并将这种方法付诸实施。当前数据分析一般都是通过软件来完成的，简单实用的有大家熟悉的 Excel，专业高端的软件有 SPSS 和 SAS 等。

1.2.5 撰写报告

在完成数据分析后，需要将分析结果展示出来并形成分析报告。数据分析报告一般包括封面、目录、分析内容和总结这几个部分。数据分析报告是对数据分析过程的总结和归纳，分析报告需要描述出数据分析的过程和分析的结果，并且要给出分析的结论。数据分析报告应该结构清晰且主次分明。分析报告应该具有一定的逻辑性，一般可以按照发现问题、总结问题原因和解决问题这一流程来描述。在分析报告中，每一个问题都必须要有明确的结论，一个分析对应一个结论，切忌贪多，结论应该基于严谨的数据分析，不能主观臆测。同时，分析报告应该通俗易懂，使用图表和简洁的语言来描述，不要使用过多的专业名词，要让看报告的人能够看懂。

最后，好的分析报告一定要有建议和解决方案。作为决策者，需要的不仅仅是找出问题，更重要的是建议或解决方案，以便决策者在做决策时参考。所以，数据分析师不仅需要掌握数据分析方法，还要了解和熟悉业务，这样才能根据发现的业务问题，提出具有可行性的建议或解决方案。

1.3 数据分析前景

1.3.1 数据分析的作用

数据分析在管理上有着十分重要的作用，它产生的价值来源于详尽而真实的数据，是一个企业的管理走向正规化、决策走向合理化的重要环节。数据分析在实际工作中能够及时纠正经营和生产中的错误，使企业的管理者能够了解企业现阶段的经营状况，知道企业业务的发展和变动的情况，及时对企业的运营有一个深入了解。通过数据分析，

可以对企业的计划进度进行分析，实时了解经营情况。同时，在了解企业当前状况的同时，提供了科学管理的依据。数据分析也可以有效帮助决策者对未来的发展趋势进行预测，为制定经营方向、运营目标以及决策提供有效的参考和依据，最大限度地规避风险。

1.3.2 数据分析就业前景

就算你不是数据分析师，数据分析技能也是未来必不可少的工作技能之一。在数据分析行业发展成熟的国家，90%的市场决策和经营决策都是通过数据分析研究确定的。

数据分析师，是指在互联网、金融、电信、医疗、旅游、零售等多个行业专门从事数据的采集、清洗、处理、分析，能够利用统计数据、定量分析和信息建模等技术制作业务报告、进行行业研究、评估和预测，从而为企业或所在部门提供商业决策的新型数据分析人才。

2015年2月，美国白宫正式命名DJ Patil担任首席数据科学家和制定数据策略的副首席技术官。DJ Patil曾在LinkedIn、eBay、PayPal、Skype和风险投资公司Greylock Partners等诸多硅谷知名公司工作过，积累了丰富的经验，在上任之后将会扮演负责政府大数据应用开发专家的角色，尤其是针对奥巴马的医疗改革方案。美国政府正在用实际行动告诉全世界，其已经意识到要充分利用他们的数据。

IDC(互联网数据中心)预测，目前每年数据的生产量超过8ZB，2020年将达到40ZB，如图1-1所示，属于大数据的时代已经到来。

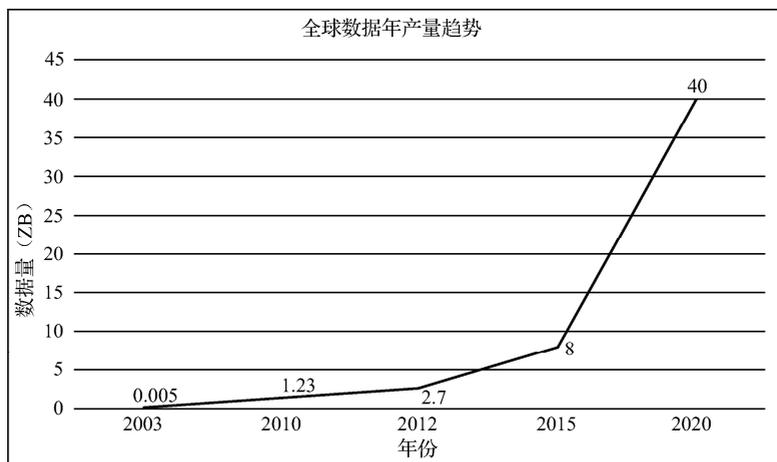


图 1-1 全球数据年产量趋势

数据生产量“拐点”已至，将开始爆发式增长。我们正处在一个数据量爆发增长的时代，当今的信息产业呈现出前所未有的繁荣，新的互联网技术不断涌现，从传统互联网的PC终端，到移动互联网的智能手机，再到物联网传感器，技术革新使数据生产能力呈指数级提升。

同时, 大数据时代可视化趋势明显, 开始重视展示数据的在线动态模式以及分布形态。数据可视化是一种新的数据分析手段、一种叙事手段, 并且包含了思考和批判的思维。通过数据可视化的方式, 能够探查数据之间的关联。随着技术的发展, 可视化将推动数据实时动态, 以及自动化更新和发布的发展。数据可视化提供了一条清晰有效地传达与沟通信息的渠道, 具体体现在交互性、可视性和多维性。在这里, 交互性是指用户能够方便地通过交互界面实现数据的管理、计算与预测。可视性是指数据可以用图像、二维图形、三维图形和动画等方式来展现, 并可对其模式和相互关系进行可视化分析。多维性是指可以从数据的多个属性或变量对数据进行切片、钻取、旋转等, 以此剖析数据, 从而多角度、多方面分析数据。

数据分析需要大数据的价值在于挖掘。在大数据时代, 可视化图表工具不可能“单独作战”。通常数据可视化都是和数据分析功能组合, 数据分析又需要数据接入整合、数据处理、ETL 等数据功能, 发展成为一站式的大数据分析平台。

1. 国外数据分析就业前景

在欧美日等发达国家, 数据分析行业不仅仅在企业的运营管理中起到举足轻重的作用, 也在政府的社会治理等方面发挥着重要作用。2012 年的美国总统大选中, 奥巴马就利用数据分析武器, 来了解不同选民的需求, 设计并策划合理有效的“自我营销”事件, 最终在选举中击败劲敌罗姆尼赢得连任, 此案已经传为“数据分析”制胜的佳话。

数据分析行业在发达国家, 不仅仅在企业中有大量的从业人员, 而且发展出很多具有规模的专业性服务机构。这些专业的服务机构有的来源于信息技术公司, 如 IBM、惠普、微软; 有的则来自数据库软件公司, 如甲骨文; 更有的来自传统行业如亚马逊、沃尔玛; 当然也有一些新兴的专业咨询企业, 在投资公司和私募基金的资金支持下, 获得飞速的发展, 如 Mu Sigma, TeraData 等。其中美国有近万家专门从事数据分析的服务公司, 年营业额达到几千亿美元, 英国有三千多家, 日本有一千多家, 瑞典也有五百多家有影响的数据分析服务公司。

从数据分析师职业来看, 数据分析行业在国外从业人群众多。在美国, 几乎所有大中型企业里都有专业的数据分析人才从事相关的数据分析工作, 数量有数百万之多, 日本有十五万多, 瑞典也有十多万名数据分析专业技术人员。数据分析高端人才的需求这几年仍在迅速扩张和增加, 数据分析人才的供应量远远赶不上需求量, 缺口很大。

从 20 世纪 90 年代起, 欧美国家开始大量培养数据分析师, 直到现在, 对数据分析师的需求仍然长盛不衰, 而且还有扩展之势。根据美国劳工部预测, 到 2018 年, 数据分析师的需求量将增长 20%。

IDC (互联网数据中心) 发布预测报告称, 2017 年大数据技术和市场将增至 324 亿美元, 实现 27% 的年复合增长率。此外, 还预测基于大数据的决策解决方案将开始取代或影响知识工作者角色, 这势必引发人才转型。

数据分析行业在国外历史已久, 伴随着互联网技术、信息技术、通信技术的发展, 目前已经非常成熟, 并远远领先国内的发展水平, 据估计, 这一差距至少要有 5~10 年。

2. 国内数据分析就业前景

自改革开放以来，随着国内经济的快速发展以及在各大行业与国际接轨的步伐不断扩大，国内的数据分析行业从2003年开始觉醒和渐热，如今已经过14年的发展。这期间数据科学相关职业从少到多、认证协会从无到有、数据分析挖掘工作从模糊到清晰。如今，中国的数据分析行业经过十多年的磨砺，正迎来辉煌灿烂的井喷式发展期。

2004年至2006年是数据分析行业的起步阶段；从2006年到2010年，数据分析行业已经全面成型，相关的培养方案和课程体系进一步完善，全国性行业协会的申请工作正式开展。我国数据分析师人数从零起步，猛增至近万人。数据人才的分布领域也从最初的分析评估业和金融业，迅速扩展到会计师、投融资机构、政府审批和企业管理等众多领域，涉及的行业从银行保险等金融行业到分析服务业、制药业、石油和天然气行业以及IT行业，数据分析师迅速成为国内炙手可热的职业之一。

近两年国内市场对数据分析师职位的需求逐步涌现。根据猎聘网数据显示，全国中高端职位中数据分析师职位由2014年初的200多个职位逐步增长到接近3000个职位，数据分析师职位无论从绝对数到相对数量而言都出现了快速增长的态势。二线城市目前对于数据分析师的需求相对滞后。分析师职位主要集中在互联网、金融、消费品、制药和医疗等行业，其中互联网和金融行业的分析师职位数超过了80%。目前数据分析师的薪酬水平高于行业平均水平，体现出数据分析师以及数据的价值正在逐渐被市场所认可。

数据分析师职位的大量涌现和对数据分析师市场价值的认可主要是基于数据分析3.0时代的到来。1954—2005年，计算机设备广泛应用，数据库初步形成；2005—2013年，互联网蓬勃发展，互联网公司为了解决自身数据量较大、数据复杂的问题引入了解决数据问题的分析工具；2013年至今，传统行业开始引入互联网行业中运用的数据分析方法，数据分析3.0时代开启。2013年至今，数据相关企业迅速发展，包括为数据提供分析、服务、软件和硬件相关的商业化和开源公司。鉴于互联网行业对于大数据分析成功的经验，市场开始重视数据和数据分析对创造商业价值的重大潜力。

2011年，“云计算”的概念风靡世界，并开始在全国推广，国内一些大型互联网公司如阿里巴巴等建成了一大批以“云计算技术”和“云存储技术”为概念的“云计算中心”，并投资开发多个开发区。这为数据采集后的存储、处理、传输和分析提供了基础。数据分析师职业有了更加具体的应用方向。

自2012年开始，“大数据”一词横空出世，国外的一些行业领导者开始提出“大数据时代”的概念。“大数据”一开始就不止步于理论，它对大量和复杂数据的处理，在技术上提出了新的拓展思路和方向。随着互联网技术的提速、第四代移动互联网的广泛应用、社交媒体的移动化，各行各业在数据的内容、结构、复杂程度和数量方面都呈现出几何倍增的特征。很多企业的分析师对如何更好地利用海量数据为政府管理、企业运营等决策提供科学的依据展开探讨。这也为数据分析师这一职业的快速发展开拓了巨大的空间。CSDN的一项调查报告指出，国内的大数据应用目前多集中在互联网领域，

并且有超过 56% 的企业在筹备和发展大数据研究。未来 5 年, 94% 的公司都需要数据分析专业人才。

埃森哲一项分析报告曾指出, 数据分析人才价值倍增的原因在于业务分析法已经从企业的辅助角色跃升至核心地位, 并能够帮助企业制定许多重要的决策和流程。对处于这一发展趋势最前沿的互联网行业而言, 分析法已经成为一项企业战略性能力。即便是在分析法仍处于起步阶段的电子和高科技等行业, 分析人才也是企业未来高速发展的关键所在。在报告中, 在所调查的包括分析服务业、银行业、石油天然气行业、通信技术行业等七大传统行业内, 新增的数据分析就业机会在中国的发展速度仅次于美国, 在 2015 年增加 30 500 人, 74% 的新增数据分析专家工作将会出现在中国、印度和巴西; 尽管美国提供了最多的数据分析就业机会, 但是, 中国、印度和巴西的数据分析职业发展速度更快, 并且只需要短短十年, 中国和印度就将在这些行业中雇佣近一半的数据分析人才。

如今, 我们已经进入了企业发展日新月异的“互联网+”时代——一个用数据说话的时代, 也是一个依靠数据竞争的时代。目前在世界 500 强企业中, 有 90% 以上都建立了数据分析部门。IBM、微软、Google 等知名巨头公司都在积极投资数据业务、建立数据部门、培养数据分析团队。各国政府和越来越多的企业意识到数据和信息已经成为企业的智力资产和资源, 数据的分析和处理能力正在成为企业日益倚重的技术手段。我国在互联网行业热钱涌动的又一波浪潮下, 对数据分析方面人才的需求更加迫切, 培养力度更是空前。随着大数据在国内的发展, 大数据相关人才却出现了供不应求的状况, 大数据分析师更是被媒体称为“未来最具发展潜力的职业之一”。

从目前来看, 在未来五年, 互联网、金融及医疗行业将会创造大多数的数据科学相关职位。互联网行业将积累大量的数据, 传统金融行业转型面临巨大的数据科学相关职位的缺口; 对于医疗行业来说, “3521 工程”, 即建设国家级、省级和地市级三级卫生信息平台, 加强公共卫生、医疗服务、新农合、基本药物制度、综合管理 5 项业务应用, 建设健康档案和电子病历 2 个基础数据库和 1 个专用网络建设, 当前全国有数十个省份在搭建省级的信息化平台、100 多个城市在不同程度上搭建市级平台, 以及区域医疗建设和医联体等, 都会积累大量的数据, 而且未来利用大数据解决医疗问题是面临的急需解决的问题。

根据对阿里巴巴、星图数据、钱方银通、和堂金融等公司的访谈及调研, 并根据相关数据做出的预测显示, 到 2018 年, 数据分析师的职位空缺将达到近 40 000 人, 而且各行各业均会对数据科学相关岗位产生很大的需求。

总之, 数据分析是一门技术也是一门艺术, 数据分析起源于生活, 也为生活创造着新的价值。从事数据分析需要累积多元化的知识和素质, 包括统计学、机器学习、工程、可视化、深刻行业知识、强化数据库能力、精练信息的能力、运筹学等。数据分析师还需要具备怀疑态度以及创造能力, 才能将数据的技术和艺术相结合, 使得数据分析能够与业务相结合, 更加贴近我们的生活。多元化的学识背景以及对于生活的感知能够造就一名优秀的数据分析师。大数据时代已经来临, 数据分析行业的急速扩展必然给数据分

析师们带来广阔的发展空间。数据分析师是一门需要掌握多元数据分析技术的职位，是拥有生活感知、经济分析能力的高端人才就业岗位。

习 题

1. 什么是数据分析？
2. 数据分析的方法有哪些？
3. 简述数据分析的步骤。