

第 1 篇

商业数据分析理论篇

对于企业来讲，数据分析可以辅助企业优化流程，降低成本，提高营业额，往往我们把这类数据分析定义为商业数据分析。商业数据分析的目标是利用大数据为职场人做出迅捷、高质、高效的决策，提供可规模化的解决方案。商业数据分析的本质在于创造商业价值，驱动企业业务增长。

本篇我们将介绍商业数据分析的思维技巧、常用的分析模型和必备的技能等，这也是我们后续深入学习 Tableau 的基础。

第 1 章

商业数据分析思维与技巧

在商业数据分析中，数据分析思维是框架式的指引，实际分析问题时还需要有很多“技巧工具”。这就好比在中学里解一元二次方程式，可以用公式法、配方法、直接开平方法、因式分解法。数据分析也是有技巧的，在一些通用分析场景下可以快速使用，同时对未来构建合适的商业数据分析模型也是有帮助的。本章我们主要介绍商业数据分析的思维、技巧与常用工具，以及商业数据分析师的必备技能等内容。

1.1 商业数据分析的思维

互联网企业的日常运营往往以某个业务平台为核心，其中数据和数据分析是不可或缺的环节。通过平台为目标用户群提供产品或服务，用户在使用产品或服务过程中产生交易数据，根据这些数据洞察反推用户的需求，创造更多符合用户需求的增值产品和服务，再重新投入到运营过程中，从而形成一个完整的业务闭环，实现企业真正意义上数据驱动业务的增长，如图 1-1 所示。

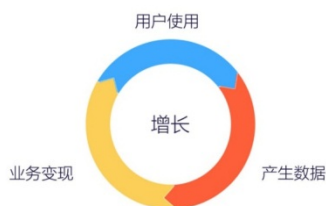


图 1-1 数据驱动业务

为什么说数据分析思维是非常重要的呢？这是由于我们在分析一个问题前如果思维缺失就会像图 1-2 中所表达的这样，往往不知道从哪里下手，在这个时候就需要学习与提高我们的数据分析思维了。

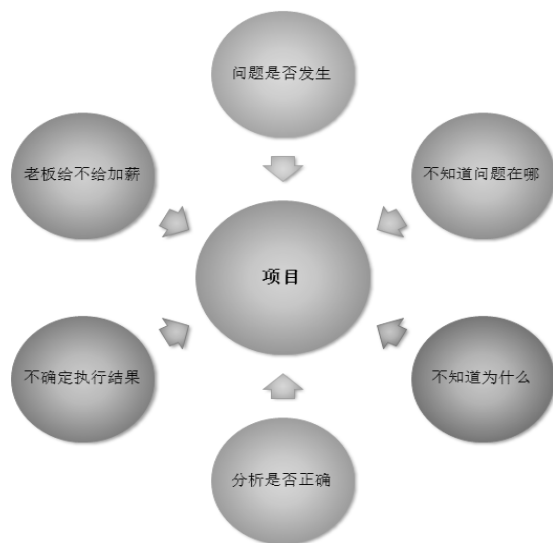


图 1-2 分析遇到的思维困境

1.1.1 结构化：按不同的方向分类

结构化思维可以看作是金字塔思维，把需要分析的问题按不同方向分类，然后不断拆分细化才能全方位地思考问题，一般先把所有能想到的想法写出来，然后整理归纳成金字塔模型。可以通过思维导图来阐述我们的分析过程。

例如，现在有一个线下销售的产品，发现 12 月份的销售额出现下降，与去年同期相比下降了 10%。首先可以观察时间趋势下的波动，是突然暴跌还是逐渐下降。然后按照不同地区分析一下差异，有没有地区性的影响因素。此外，还可以从外部的角度分析一下现在的市场环境怎么样。具体的结构化分析过程如图 1-3 所示。

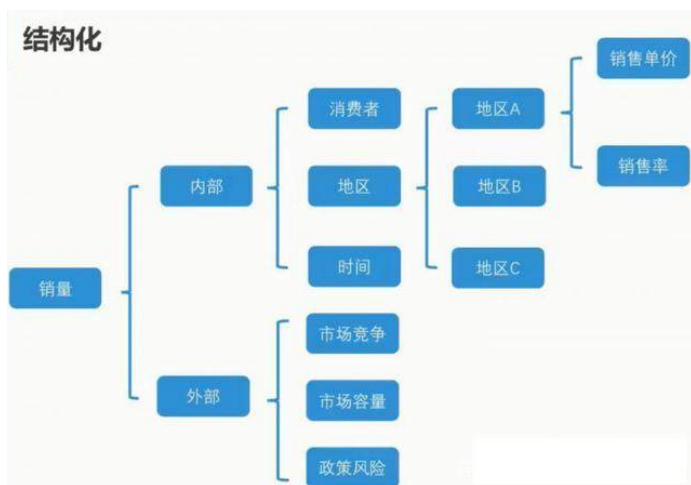


图 1-3 结构化思维

1.1.2 公式化：对数据进行量化分析

在结构化的基础上，分析的变量往往会存在一些数量关系，使其能进行计算，将分析过程进行量化，从而验证我们的观点是否正确。它是对现象的数量特征、数量关系与数量变化进行分析的方法。企业的销售数据公式化思维如图 1-4 所示。



图 1-4 公式化思维

1.1.3 业务化：结合具体业务进行分析

业务化就是深入了解业务情况，结合该项目的具体业务进行分析，并且能让分析结果落地。用结构化思考和公式化拆解得出的最终分析结果在很多时候表示的是一种现象，不能体现原因。所以需要继续用业务的思维去思考，站在业务人员的角度思考问题，深究出现这种现象的原因，从而实现通过数据推动业务。

提升业务思维的主要途径如下：

- 贴近业务：工作中多和一线的业务人员进行沟通。
- 换位思考：站在业务人员和用户的角度进行思考。
- 积累经验：从成功和失败的经历中总结业务特点。

1.2 商业数据分析的 7 种技巧

在数据分析中，结构化、公式化和业务化 3 种核心数据分析思维是框架性的指引，实际应用中还是需要很多技巧的。下面列举几种常用的数据分析技巧，分别是象限法、多维法、假设法、指数法、二八法、对比法和漏斗法。

1.2.1 象限法：运用坐标的方式分析

通过对两种维度的划分，运用坐标的方式表达出想要的价值，由价值直接转变为策略，从而进行一些落地的推动。象限法是一种策略驱动的思维，广泛应用于战略分析、产品分析、市场分析、客户管理、用户管理和商品管理等。

RFM 模型把客户按最近一次消费（Recency）、消费频率（Frequency）、消费金额（Monetary）3 个维度分成 8 个象限，如图 1-5 所示。

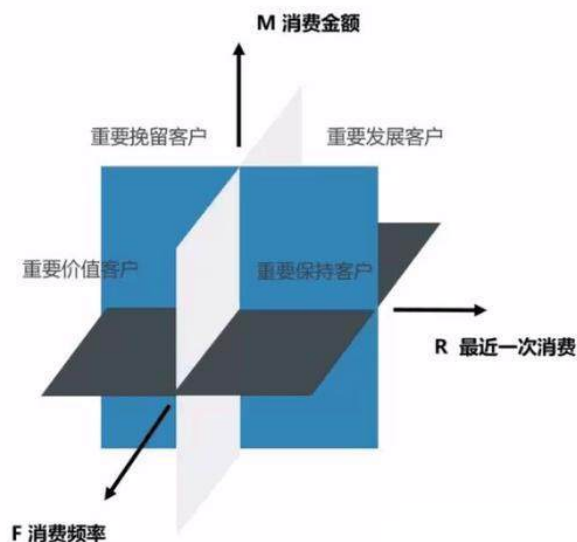


图 1-5 象限法的 RFM 模型

1.2.2 多维法：运用多维的方法呈现

多维法是指对分析对象从多个维度进行分析，这里一般是 3 个维度，每个维度有不同的数据分类，代表总数据的立方体就被分割成一个个小方块，落在同一个小方块的数据拥有同样的属性，这样可以通过对比小方块内的数据进行分析。

图 1-6 是某企业产品销售额的数据立方体，虽然只有日期、地区和产品 3 个维度，但是根据这个立方体，已经能够解决很多管理者急需了解的问题，可以通过切片实现每类产品的销售额、每个月每类产品的销售额等操作。

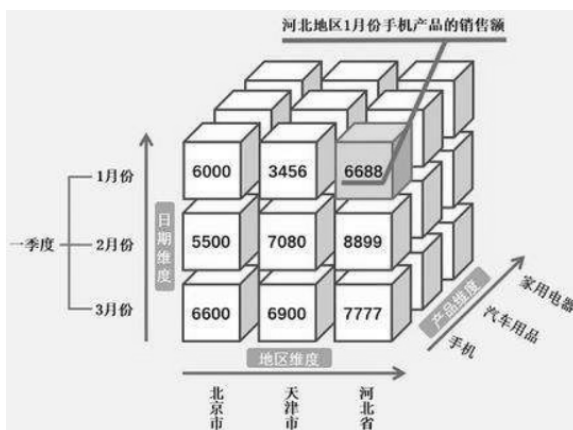


图 1-6 多维法的数据立方体

1.2.3 假设法：运用假设推理的方式

在一些情况下，如进入新市场的销量、商品提价后销量的变化情况，可能没有明细数据进行分析，就需要用到假设法。假设法就是假设一个变量或者比率成立，然后根据部分数据进行反推，这是一种启发思维的技巧，一般过程是先假设后验证再判断出分析结果。

如果你是一名电商数据分析师，想将商品提价，分析一下销售额会有怎样的变化。解答思路：如果确定销量会下降，那么会下降多少？这里就要假设商品流量情况、提价后转化率的变化情况，然后根据历史数据汇总出销量下降多少，从而得出销售额的变化情况。

假设商品提价后销量一定会下跌，那么销量会下跌多少？

- 首先假设流量不会有变化，流量和渠道营销正相关，商品价格影响转化率，那么现在确定转化率的波动幅度。
- 参照平时的转化率（例如 15%），预估提价后的转化率变化，假设各类型用户对价格敏感度不同，将用户划分为忠诚型、普通型、羊毛型等类型。
- 不同类型用户的数量不同，反应也不同。凭借经验可以做出一般忠诚用户转化率较高，羊毛用户几乎不会发生转化的假设。

1.2.4 指数法：整合与计算多个指标

指数法是把某个数据多个指标按一定的整合与计算转化为同度量的一个值，这个度量值称为指数。例如，在一场游戏竞技比赛中要确定该场的最有价值团队成员（MVP），则需要根据击杀数、死亡数、助攻数等指标综合计算出一个得分，得分高的为 MVP。

指数法常用的有线性加权、反比例和对数 3 种：线性加权是把每个指标乘以一个系数后相加；反比例是用数学上的反比例函数 $y=k/x$ 变化后再计算；对数是数学中所说的对数，一般以 2 或者 10 为底数。如何使用指数法，没有统一标准，一般是根据经验实现，将无法利用的数据加工成可利用的。

1.2.5 二八法：运用帕累托法则分析

二八法又名帕累托法则(Pareto's principle),也叫巴莱特定律、关键少数法则(Vital FeRule)、不平衡原则等,被广泛应用于社会学及企业管理学等,是19世纪末20世纪初意大利经济学家帕累托发现的。他认为,在任何一组东西中,最重要的只占其中一小部分,约20%,其余80%尽管是多数,却是次要的。

例如,在个人财富上可以说世界上20%的人掌握80%的财富。而在数据分析中,则可以理解为20%的数据产生了80%的效果,需要围绕这20%的数据进行挖掘。二八法是抓重点分析,适用于任何行业,找到重点,发现其特征,然后可以思考如何让其余的80%向这20%转化,提高效果。

1.2.6 对比法：运用对比法对比分析

对比法就是用两组或两组以上的数据进行比较,常见的是用于在时间维度上的同比和环比、定基比,与竞争对手的对比、类别之间的对比、特征和属性对比等。对比法可以发现数据变化规律,使用非常频繁,常常与前面的技巧结合使用。

对比分析基于相同数据标准由其他影响因素所导致的数据差异,目的在于找出差异后进一步挖掘差异背后的原因,从而找到优化的方法。

对比分析法的维度可以分为同比、环比、定基比。

- 同比:例如,去年6月与今年6月的对比,同比一般被看作是基于相同数据维度的时间同期对比,也可以看作基于时间维度的影响因素对比。例如相同的营销活动在不同的渠道投放所带来的转化数据,也可以看作是同比。
- 环比:例如6月份与5月份数据的对比,这是从时间维度的对比,也可以以周期性维度对比,例如第一阶段推广投放了20个渠道,第二阶段推广投放了25个渠道,那么第二阶段与第一阶段的推广环比上涨还是下降,进而找出变化的原因。
- 定基比:针对一个基准数据的对比,例如在各个推广渠道中,渠道B与渠道A相比,渠道C与渠道A相比,而两者的比值是渠道B与渠道C的定基比。

1.2.7 漏斗法：挖掘有价值的新用户

漏斗法即使用漏斗图对用户进行转化率分析,有点像倒金字塔,是一个流程化思考方式,常用于新用户的开发、购物转化率这些有变化和一定流程的分析中。不过,单一的漏斗分析是没有用的,不能得出什么结果,要与其他相结合,如与历史数据的对比等。

漏斗分析是一套流程式数据分析,能够科学反映用户行为状态以及从起点到终点各阶段用户转化率情况的重要分析模型。漏斗分析模型已经广泛应用于网站用户行为分析和App用户行为分析的流量监控、产品目标转化等日常数据运营与数据分析的工作中。

漏斗分析对用户行为分析来说是不可或缺的,最常用的指标是转化率和流失率。例如,有100

人访问某电商网站，有 30 人单击注册，有 10 人注册成功。这个过程共有 3 步：第一步到第二步的转化率为 30%，流失率为 70%；第二步到第三步转化率为 33%，流失率 67%；整个过程的转化率为 10%，流失率为 90%。

1.3 商业数据分析的典型方法

数据对于商业中的决策者来说非常重要，不同领域的企业都在寻求利用数据的最佳方案。零售领域发展非常迅速，数据的应用也相当全面，例如零售商会分析数据并发展特殊的客户画像，以了解零售商的痛点，对大数据的全面分析甚至可以影响与操纵客户的决策。

下面我们简单介绍一些零售业分析中采用的典型方法。

1.3.1 推荐引擎

推荐引擎对零售商来说非常重要，可以作为用户行为预测的工具，零售商倾向于使用推荐引擎作为影响客户购买决策的重要方式。此外，商品推荐一方面使零售商能够提升销售额，另一方面还可以主导消费趋势。推荐引擎就可以根据客户的偏好计算出商品相似性指数，从而相应地提供商品或服务的推荐。

推荐引擎利用大量的数据筛选来获取有价值的商业洞察。通常，推荐引擎使用协同过滤或内容过滤的模型来实现，主要考虑客户过去的行为或一系列的产品特征，例如人口统计数据、消费偏好数据、先前购物体验等，都会被纳入模型中。

1.3.2 购物篮分析

购物篮分析被视为零售业中数据分析的经典方法。零售商多年来一直从该方法中获利，这个过程主要取决于梳理客户交易历史数据。消费者未来的决策和选择可以通过这个工具进行预测。挖掘篮子中的产品以及客户的所有喜欢、不喜欢和预览记录，从而有助于零售商在货架管理、价格制定和内容放置方面的决策。

购物篮分析通常通过规则挖掘算法进行，需要事先将数据进行处理，转换成以订单为基础的格式，在此基础上，建立产品之间的关联联系，并且进一步应用关联规则进行分析，这些分析有助于改善零售商的发展战略和营销技巧，从而极大地提升营销效果。

1.3.3 保修分析

零售领域中的保修分析包括监控保修索赔、检测欺诈活动、降低成本和提高质量等。该过程涉及数据挖掘和文本挖掘，从而实现识别索赔模式，并通过细分分析将数据转换成可以执行的实时计划和建议。

通常保修分析可以使企业专注于检测保修索赔中的异常情况，尤其是对于互联网数据平台，可以加快保修索赔的过程。

1.3.4 价格优化分析

为客户和零售商提供合适的价格是优化分析的优势。价格形成的过程不仅取决于生产的成本，还取决于客户的消费力和竞争对手的报价。数据分析工具将这个问题提升到一个新的水平。该方法先假定客户对价格变化是敏感的，使用实时优化模型，零售商有机会吸引客户，保持关注度并实现个人定价方案。

价格优化工具从多个渠道来源获得数据，能够定义价格的灵活性，并且考量位置、客户的个人购买意愿、季节性和竞争对手的定价等因素，计算极值和频率表等，以进行预测变量和利润响应的变量评估和最优分布等。

1.3.5 库存管理

库存管理是指在紧急时能够使用的库存，以供将来使用。零售商的目标是在适当的时间、适当的条件、适当的地方提供合适的产品。为了实现这样的目标，零售商必须对库存和供应链进行深入的分析。

机器学习算法和数据分析平台可以挖掘各要素和供应链之间的模式与相关性，通过不断调整参数，算法可以算出最优的库存和库存策略，分析师可以发现合适的模式，并制定合理的销售策略以及优化物流管理等。

1.3.6 选址分析

商业选址是企业经营模式对场地的具体要求，同时也反映了商业企业经营策略、开店能力、扩张能力等。招商工作不仅仅要了解对象的具体门店要求，还要了解对象的发展策略和扩张意图，只有这样才能完成招商任务，并达到物业价值最大化、经营持久化、业态租金的均衡化，其中比较有名的是星巴克（Starbucks）选址策略。

数据分析对解决选址类型的问题非常有效。通常为了解决这种类型的问题，需要进行大量的分析。数据分析师在探索客户数据时，会非常关注人口的相关因素，其中邮政编码和地理位置为分析市场潜力提供了基础。此外，分析师还可以对零售商进行网络分析，期望能够通过连接所有关联点而找到最终解决方案。

1.3.7 情感分析

情感分析不是零售行业的新工具，随着数据科学开始流行，它的金钱和时间成本已经降低很多。

数据分析师可以通过从社交网络和在线服务反馈收到的信息来进行品牌与客户间的情感分析，社交媒体资源也相对容易获得，这就是为什么在社交平台上进行分析要容易得多。

情感分析使用语言处理来跟踪客户的积极或消极词汇，这些反馈成为服务改进的背景。分析师在自然语言处理、文本分析的基础上进行情感分析，以提取正面、中性或负面的情绪。算法会分析所有有意义的词语，被发现的情绪都会属于某些类别，最后的输出是类别的情绪评级和文本的整体情绪表现。

1.3.8 品类分析

品类分析已成为零售业务的重要环节，包含增加产品销售和推广的绝大多数活动和战略。它有助于通过各种渠道影响客户的决策过程，商品轮换有助于保持产品的新鲜和更新，精美的包装和品牌保留了客户的注意力，增强了视觉吸引力。

此外，品类结构优化是对品类定义阶段定义好的品类分类、品类角色，经过一段时间的实际运营后，对相应的指标达成情况进行综合评估，对品类定义直到单品构成的一个改善优化过程。品类结构优化经常被简单地理解成商品的淘汰过程，这种认识具有片面性，因为这样做的结果不是品类结构的慢慢优化，极有可能导致品类不断地缩小。

1.3.9 生命价值预测

在零售业中，客户生命价值是客户在整个客户与公司业务关系中对公司的利润总值，可以获得客户生命周期中的两个重要指标：历史销量和预测销量。所有预测都是基于客户最近交易的历史数据进行，品牌的客户生命周期也是这样进行，其中比较典型的是基于客户生命周期的 LRFM 价值分析。

通常，客户生命价值模型收集、分类和清理有关客户偏好、消费、最近购买的细节和行为的数据，处理完这些数据后，我们可以得到现有客户和潜在客户价值的线性方程。该算法还可以发现客户特征与其选择之间的相互依赖关系，此外统计方法的应用还有助于确定客户的购买模式。

1.3.10 反欺诈分析

反欺诈的检测是零售商的挑战之一，欺诈检测的主要目的是为了防止经济损失。客户可能会遭受退货和交付类型的欺诈、权利滥用、信用风险等，这些案件只会破坏零售商的声誉。一旦成为这种情况的受害者，可能会永远破坏客户的宝贵信任，因此保护公司声誉的唯一有效方法是比欺诈者领先一步。

为欺诈检测而开发的算法不仅应该识别欺诈并杜绝其发生的机会，还应预测未来的欺诈活动，这就是为什么深度神经网络被证明是有效的原因。应用常见的降维法来识别隐藏的规律，并用来标记活动和欺诈的交易，在某种程度上提高了零售商保护客户和公司的能力。

1.3.11 留存分析

留存分析是一种分析用户参与活跃程度的模型，考察进行初始行为的用户中有多少人会进行后续行为，是用来衡量产品对用户价值高低的重要方法，可以帮助回答以下问题：

- 新客户在未来的一段时间内是否完成了你期望完成的行为？
- 改进新注册用户的引导流程，改善用户注册后的参与程度，如何验证？
- 改善产品功能是否有效，如新增邀请好友功能，是否有人因此而多使用产品？

标准的留存分析模型具有灵活条件配置——根据具体需求筛选初始行为或后续行为的细分维度，针对用户属性筛选合适的分析对象。留存分析具有如下价值：

第一，留存率是判断产品价值最重要的标准，揭示了产品保留用户的能力。

留存率实际上是一种转化率，即由初期不稳定的用户转化为活跃用户、稳定用户、忠诚用户。随着统计数字的变化，运营人员可看到不同时期用户的变化情况，从而判断产品对用户的吸引力。

第二，宏观上把握用户生命周期长度以及定位产品可以改善之处。

通过留存分析，可以查看新功能上线之后对不同群体的留存是否带来不同效果，可以判断产品新功能是否提高了用户的留存率，结合版本更新、市场推广等诸多因素砍掉使用频率低的功能，实现快速迭代验证，并制定相应的策略。

1.4 商业数据可视化分析工具

商业数据分析的工具分为非编程类和编程类。大部分商业数据分析师，对编程都比较陌生，因此我们这里仅仅介绍一些非编程类的数据可视化工具。由于熟悉 Excel 和 SQL 语言是对每一位数据分析师的最基础要求，因此这里不做介绍。

1.4.1 Tableau Desktop

Tableau 是桌面系统中最简单的商业智能软件之一，没有强迫用户编写自定义代码，新控制台也可以完全自定义配置，不仅能够监测信息，还提供了完整的分析能力。Tableau 简单、易用、快速，一方面归功于斯坦福大学的突破性技术，集计算机图形学、人机交互和数据库系统于一身的跨领域技术，其中最耀眼的就是 VizQL 可视化查询语言和混合数据架构；另一方面在于 Tableau 专注于处理简单的结构化数据，即已整理好的数据——Excel、数据库等。

针对 Tableau、Qlik、TIBCO Software、SAS、Microsoft、SAP、IBM 和 Oracle 这 8 家数据可视化产品和服务提供商的调查，分别从知名度、流行度和领导者 3 个角度进行分析。从知名度来看，8 家厂商几乎不分先后，只有微小的差距；从流行度来看，SAP、IBM 和 SAS 占据前 3 位，分别占比 19%、18%和 17%；从领导者来看，Tableau 以 40%的优势遥遥领先。

Tableau Desktop 可以制作出绚丽的仪表板，例如 Ryan Sleeper 制作的有史以来收入最高的 10 名演员的仪表板，可以让我们了解有史以来票房收入最高的 10 名演员、他们的电影、总收入以及关键的接受点等，如图 1-7 所示。

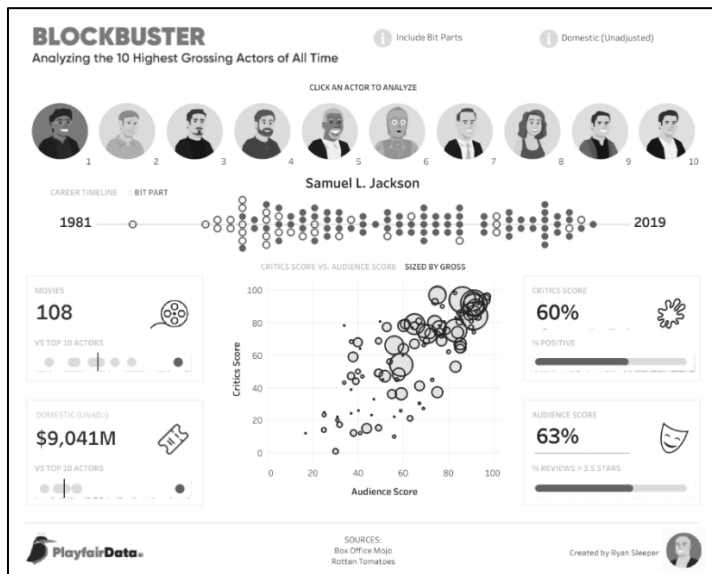


图 1-7 收入最高的前 10 名演员

1.4.2 Microsoft Power BI

Microsoft Power BI 是一套商业分析工具，可以连接数百个数据源、简化数据准备并提供即席查询。即席查询（Ad Hoc）是用户根据自己的需求灵活地选择查询条件，系统能够根据用户的选择生成相应的统计报表。即席查询与普通应用查询最大的不同是普通的应用查询是定制开发的，而即席查询是由用户自定义查询条件的。

它是微软发布的最新可视化工具，整合了 Power Query、Power Pivot、Power View 和 Power Map 等一系列工具的经验成果，所以使用过 Excel 做报表和 BI 分析的从业人员可以快速使用，甚至可以直接使用以前的模型。此外，Excel 2016 提供了 Power BI 插件。

Microsoft Power BI 是一种业务分析服务，可提供见解，以实现快速、明智的决策。

- 将数据转换为令人赞叹的视觉对象，并在任何设备上与大家共享。
- 在一个视图中直观浏览和分析本地数据和云端数据。
- 协作并共享自定义仪表板和交互式报表。
- 通过内置管控和安全性在整个组织中进行缩放。

Microsoft Power BI 也可以制作出美观的仪表板，其可视化效果如图 1-8 所示。



图 1-8 Power BI Desktop 仪表板

1.4.3 Smartbi Insight

思迈特商业智能与大数据分析软件（Smartbi Insight）是企业级商业智能和大数据分析平台，经过多年的持续发展，整合了各行业的数据分析和决策支持的功能需求，在传统 BI 到自助 BI 再到智能 BI 的历史进程中不断创新与探索，具有分布式云计算、直观的流式建模、拖曳式操作、实用的统计分析、可视化数据探索等特点。

Smartbi Insight 产品定位于一站式大数据服务平台，对接各种业务数据库、数据仓库和大数据平台，进行加工处理、分析挖掘与可视化展现；满足各种数据分析应用需求，如大数据分析、自助探索分析、地图可视化、移动管理驾驶舱、指挥大屏幕、企业报表平台等。Smartbi 功能设计全面完整，覆盖数据管理、数据提取、数据分析、数据分享四大环节，帮助客户从数据角度描述业务现状、分析业务原因、预测业务趋势、驱动业务变革。

Smartbi 可视化使业务的重心回归到问题的原因分析和解决方案设计上，其可视化效果如图 1-9 所示。

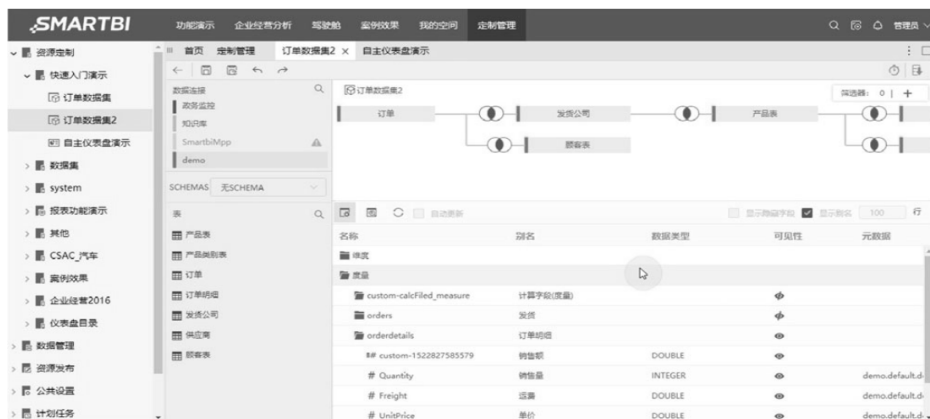


图 1-9 Smartbi 数据可视化

1.4.4 QlikView

QlikView 是一个完整的商业分析软件，开发者和分析者可以使用 QlikView 构建和部署应用，各种终端用户可以高度可视化、功能强大和创造性的方式互动分析业务信息，是一个具有完全集成 ETL 工具的应用开发环境，拥有强大的 AQL 分析引擎和一个高度直觉化、使用简单的用户界面，让开发者能从多种数据库里提取和清洗数据，建立强大、高效的应用。

QlikView 是一个可升级的解决方案，完全利用了基础硬件平台，用大量数据进行业务分析，由开发工具 (QlikView Local Client)、服务器组件 (QlikView Server)、发布组件 (QlikView Publisher) 以及其他应用接口组成，支持多种发布方式，还可以与其他系统进行集成，使业务的重心回归到问题的原因分析和解决方案设计上，其可视化效果如图 1-10 所示。

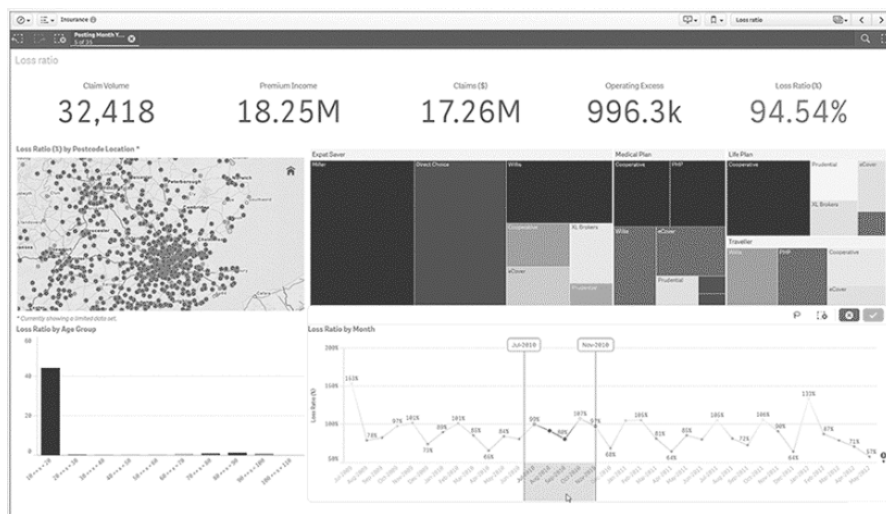


图 1-10 QlikView 仪表板

1.4.5 FineBI

FineBI 是帆软软件有限公司推出的一款商业智能产品，可以通过最终业务用户自主分析企业已有的信息化数据，帮助企业发现并解决存在的问题，协助企业及时调整策略做出更好的决策，增强企业的可持续竞争性。

- 完善的数据管理策略。FineBI 支持丰富的数据源连接，以可视化的形式帮助企业进行多样数据管理，极大地提升了数据整合的便利性和效率。
- 可连接多种数据源。FineBI 支持 30 多种大数据平台和 SQL 数据源，支持 Excel、TXT 等文件数据集，支持多维数据库、程序数据集等各种数据源。
- 可视化管理数据，用户可以方便地以可视化形式来对数据进行管理，简单易操作。
- 多种数据处理方式。用户可以通过 FineBI 以可视化的形式来对数据进行多种操作，如过滤、分组汇总、新增列、字段设置、排序等功能，完完全全掌控数据。

- 智能权限继承。管理员只需配置基础的数据关联和权限，分析数据的用户一定要在其权限范围内操作，而且数据集的关联也可以自动继承，提升双方效率。
- 良好的使用体验。用户在 FineBI 中的每一个操作都可以增加、删除和修改，并且提供预览。

FineBI 数据可视化效果如图 1-11 所示。

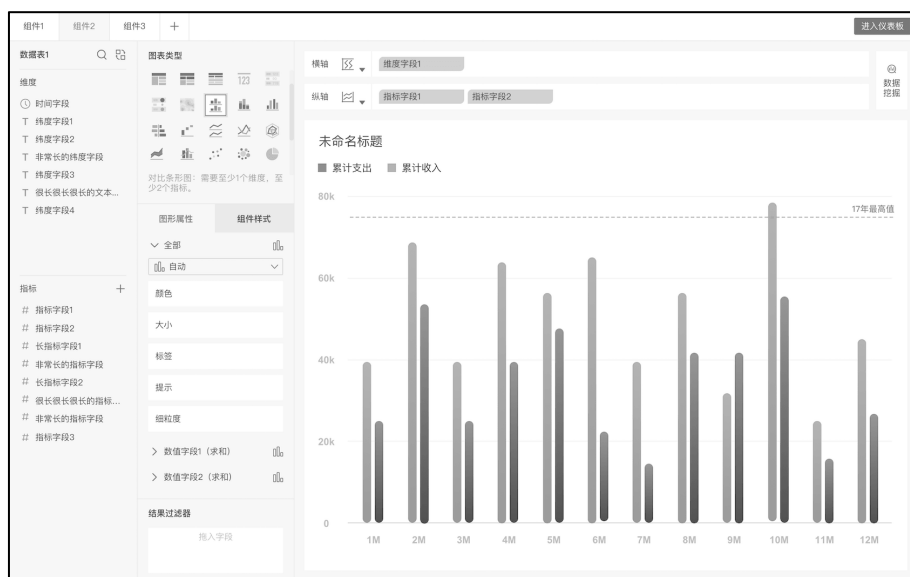


图 1-11 FineBI 数据可视化

1.5 商业数据分析的步骤与阶段

面对海量的数据，很多人都不知道如何准备、如何开展，如何得出结论，下面为大家介绍商业数据分析的基本步骤。

商业数据分析都应该以业务场景为起始思考点，以业务决策作为终点。数据分析应该先做什么、后做什么？基于此，我们提出了商业数据分析流程的 5 个基本步骤，如图 1-12 所示。



图 1-12 商业数据分析流程

第一步，挖掘业务含义，理解数据分析的背景、前提以及想要关联的业务场景是什么。

第二步，制定分析计划，如何对场景拆分，如何推断。

第三步，从分析计划中拆分出需要的数据，真正落地分析本身。

第四步，从数据结果中判断提炼出商务洞察。

第五步，根据数据结果洞察，最终产出商业决策。

我们一般以商业回报来定位数据分析的不同阶段，因此我们将商业数据分析分为 4 个阶段，如图 1-13 所示。

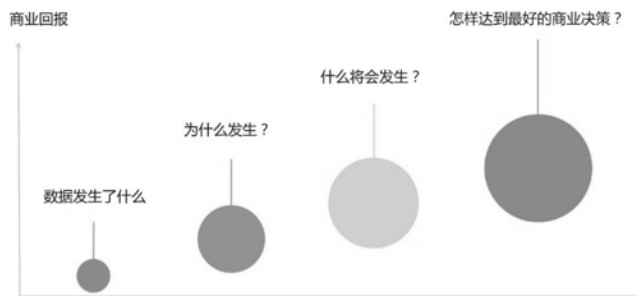


图 1-13 商业数据分析的阶段

阶段 1：观察数据当前发生了什么。

首先，基本的数据展示可以告诉我们发生了什么。例如，公司上周投放了新的搜索引擎 A 的广告，想要比上一周下来，新渠道 A 比现有渠道 B 的情况如何，A、B 各自带来了多少流量，转化效果如何？又比如，新上线的产品有多少用户喜欢，新注册流中注册的人数有多少。这些都需要通过数据来展示结果，都是基于数据本身提供的“发生了什么”。

阶段 2：理解为什么发生。

如果看到了渠道 A 比渠道 B 带来更多的流量，这时我们就要结合商业来进一步判断这种现象的原因。这时候我们可以进一步通过数据信息进行深度拆分，也许是某个关键字带来的流量，也许是该渠道更多地获取了移动端的用户。这种数据深度分析判断成为商业分析的第二个进阶，也能够提供更多商业价值上的体现。

阶段 3：预测未来会发生什么。

当我们理解了渠道 A、B 带来流量的高低，就根据以往的知识预测未来会发生什么。在投放渠道 C 和渠道 D 的时候，猜测渠道 C 比渠道 D 好，哪一个节点比较容易出问题；我们也可以通过数据挖掘的手段，自动预测判断 C 和 D 渠道之间的差异，这就是数据分析的第三个进阶，预测未来会发生的结果。

阶段 4：商业决策。

所有工作中最有意义的还是商业决策，通过数据来判断应该做什么。商业数据分析的目的就是商业结果。当数据分析的产出可以直接转化为决策或直接利用数据做出决策时，才能直接体现出数据分析的价值。

此外，在数据分析的过程中会有很多因素影响到我们的决策，那么如何找到这些因素呢？内外因素分解法是把问题拆成 4 部分，包括内部因素、外部因素、可控和不可控，然后再一步步解决每一个问题，如图 1-14 所示。

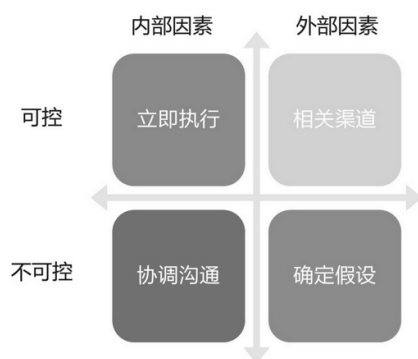


图 1-14 影响数据分析的因素

例如，某社交招聘类网站，其盈利模式一般是向企业端收费，其中一个收费方式是购买职位的广告位。业务人员发现，“发布职位”的数量在过去的6个月中有缓慢下降的趋势。根据内外因素分解法，我们可以从以下4个角度依次去分析可能的原因。

- 内部可控因素：产品近期上线更新、市场投放渠道变化、产品黏性、新老用户留存问题、核心目标的转化。
- 外部可控因素：市场竞争对手近期行为、用户使用习惯的变化、招聘需求随时间的变化。
- 内部不可控因素：产品策略（移动端/PC端）、公司整体战略、公司客户群定位。
- 外部不可控因素：互联网招聘行业趋势、整体经济形势、季节性变化。

有了内外因素分解法，我们就可以较为全面地分析数据指标，避免可能遗失的影响因素，从而对症下药。

1.6 商业数据分析师的必备技能

商业数据分析师（BA）的技能可以分为两大类：商业技能和分析技能。在商业领域，BA应当具备战略管理、商业过程、产品/运营或市场以及相关行业的背景知识，还要有与各个部门沟通与协调的能力。

1. 商业问题解析

这一阶段作为项目的起始阶段，主要致力于从商业角度对项目目标与需求进行解析，并将相关需求转化为可供数据挖掘以及分析的具体问题。这一阶段囊括了前5项内容：商业分析规划、需求信息检索、需求动态更新、战略分析以及需求分析与问题定义。具体而言包括定义商业目标、寻找商业需求、相关利益分析、数据问题转化等。

2. 相关数据理解

在项目进行到数据阶段之后，BA的首要任务在于通过进一步明确数据质量需求来开展数据收集与选择，并且做到形成初步的数据全局意识，以及能够从一些数据集分支当中发现隐藏信息。例如，我们需要什么样的数据，这些数据是否能够获得，这些数据有哪些重要特征。

3. 数据收集准备

这一阶段的主要任务在于从多个渠道收集数据，并且对数据进行清洗、集成、变换、归约等一系列的预处理，从而为下一步做好基础。大数据专家 Borat 曾经调侃道：“在数据科学中，80%的时间用在了准备数据上，20%的时间用在了抱怨准备数据上。”干净有条理的数据是整个项目成功最重要的一步，然而在实际操作过程中，会发现绝大多数的数据都是不完整、不一致甚至杂乱无章的，这样的数据几乎无法达到进入分析阶段的要求，因而细致且完整的预处理显得尤为重要。

在建模方面，描述与预测方面的统计学相关知识都是必要的，具体应用包括 A/B 测试、时间序列预测以及分类建模等，所运用到的软件包括 R、SAS、SPSS、Google Analytics 等；数据收集与处理方面，数据库的管理与存储、数据预处理以及可视化处理都是 BA 的必备技能，相关软件包括 SQL、Excel、R、Python、Tableau 等。

4. 分析与建模

将最终的数据需要根据具体的商业需求应用于模型当中，具体步骤包括确定建模方法、确定变量、选择分析方法、运行并测试模型。

5. 检查与确认

在模型应用之前进行最终的评估与审查，确保模型结果与商业问题所探求的内容一致。为此，BA 应当做到对模型结果进行反复细致的检查，并且能够回答商业项目的问题。

6. 报告与可视化

作为衔接商业与技术的桥梁，BA 工作中非常重要的任务就是与各方沟通。在这一阶段，能够将模型结果与商业问题相结合，通过讲故事的方式将生硬枯燥的数字变成可行性强的方案。在一个项目周期结束之后，BA 应当根据相应的 KPI 指标评估项目结果，并寻找存在问题以及提出改进建议。

通常，商业数据分析师一般使用 Microsoft Excel、Microsoft PowerPoint、Microsoft Access、SQL、Google Analytics 和 Tableau 等软件，具体需要使用的软件和工具取决于职位和企业的要求。这些工具帮助分析师收集和分类数据、创建图表、编写文档和可视化视图等，不一定需要具备非常出色的编程或数据库技能，但是如果你具备这些技能，就将是锦上添花。

1.7 练习题

1. 简述什么是商业数据分析的思维以及主要方法。
2. 简述什么是商业数据分析的技巧以及主要内容。
3. 结合工作实际，简述几种商业数据分析的典型方法。
4. 尝试下载和安装几种主要的商业数据可视化软件。
5. 收集整理有关 Tableau Desktop 的文档等学习资料。
6. 简述商业数据分析的详细步骤和阶段。
7. 列举商业数据分析需要的必备技能。

第 2 章

商业数据分析的模型

作为一名商业数据分析师，假如没有数据思维或者没有思考模型怎么办？思维模型会给你提供一种视角或思维框架，从而帮助你建立起观察事物和看待世界的视角。通过对思维模型的学习，能提高你成功的可能性，并帮你避免失败。

本章主要介绍一些常用的数据模型，通过这些模型，能够帮助读者快速上手数据分析，减少自我探索过程中的弯路。

2.1 KANO 分析模型

KANO 模型是东京理工大学教授狩野纪昭（Noriaki Kano）发明的对用户需求分类和优先排序的有用工具，该模型是受行为科学家赫兹伯格的双因素理论启发而提出的，体现了产品性能和用户满意之间的非线性关系，主要是通过标准化问卷进行调研，根据调研结果对各因素属性归类，解决产品属性的定位问题，以提高客户满意度，如图 2-1 所示。

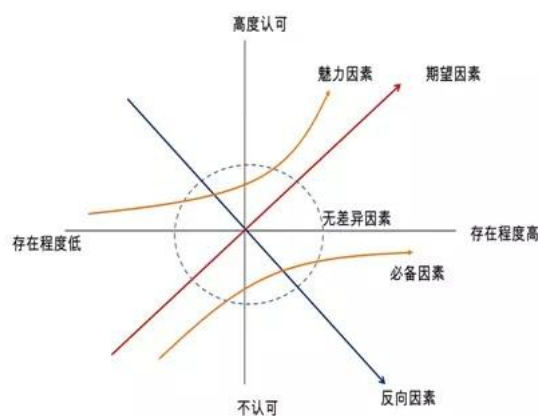


图 2-1 KANO 分析模型

(1) 魅力因素：又称兴奋型需求，指不会被顾客过分期望的需求。对于魅力型需求，随着满

足顾客期望程度的增加，顾客满意度会急剧上升，一旦得到满足，即使表现并不完善，顾客表现出的满意状况也是非常高的。

(2) 期望因素：也称为意愿型需求，是指顾客的满意状况与需求的满足程度成比例关系的需求。此类需求得到满足或表现良好的话，客户满意度会显著提高，企业提供的产品和服务水平超出顾客期望越多，顾客的满意状况越好。

(3) 必备因素：也称为必备型需求、理所当然需求，是顾客对企业提供的产品或服务因素的基本要求，是顾客认为产品“必须有”的属性或功能。当其特性不充足（不满足顾客需求）时，顾客很不满意；当其特性充足（满足顾客需求）时，顾客也可能不会因而表现出满意。

(4) 无差异因素：无论提供或不提供此需求，用户满意度都不会改变，用户根本不在意。无论提供与否，对用户体验均无影响，不会导致顾客满意或不满意。

(5) 反向因素：用户根本没有此需求，提供后用户满意度反而会下降，又称逆向型需求，是指引起强烈不满的质量特性和导致低水平满意的质量特性，因为并非所有的消费者都有相似的喜好。

总之，在实际操作中，企业首先要全力以赴地满足顾客的基本型需求，保证顾客提出的问题得到解决，重视顾客认为企业有义务做到的事情。然后，企业应尽力去满足顾客的期望型需求，这是质量的竞争性因素，提供顾客喜爱的额外服务或产品功能，使其优于竞争对手并有所不同，使顾客满意。最后争取实现顾客的兴奋型需求，为企业建立忠实的客户群。

2.2 5W2H 分析模型

5W2H 分析法又叫“七问分析法”，是二战中美国陆军兵器修理部首创的，简单、方便，易于理解、使用，富有启发意义，广泛用于企业管理和技术活动，对于决策和执行性的活动措施也非常有帮助，还有助于弥补考虑问题的疏漏。用 5 个以 W 开头的英语单词和两个以 H 开头的英语单词进行设问，发现解决问题的线索，寻找发明思路，进行设计构思，从而搞出新的发明项目，这就叫作 5W2H 法。

提出疑问与发现问题和解决问题是极其重要的，创造力高的人都具有善于提问题的能力。众所周知，提出一个好的问题就意味着问题解决了一半。发明者在设计新产品时，常常提出：为什么（Why）；做什么（What）；何人做（Who）；何时（When）；何地（Where）；如何（How）；多少（How much）。这就构成了 5W2H 法的总框架，如图 2-2 所示。

- Why: 你的购买动机/目的是什么？比如企业运用（拍电影、宣传片）——新用户注册获取。
- What: 用户主要购买的产品是哪些——产品销量获取。
- Who: 用户特征，性别构成、年龄分布、地域分布、学历分布、收入分布、注册时间。
- When: 购买时间分布、购买间隔分布。
- Where: 购买渠道，例如官网、App、天猫、京东、苏宁、线下线上代理商。
- How: 用户的支付方式销量分布，用户更喜欢用什么支付。
- How much: 价格段销量分布（用户更容易接受哪个价位）。

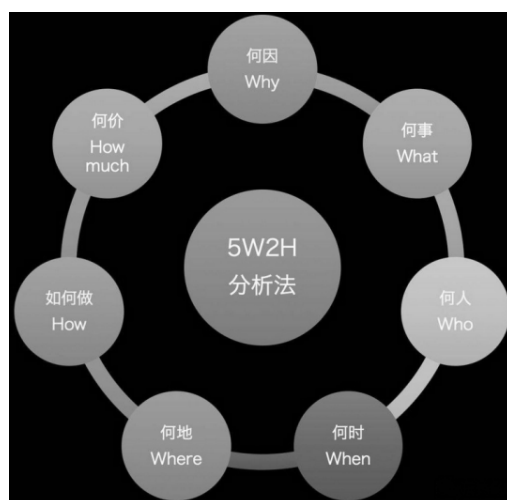


图 2-2 5W2H 分析模型

如果现行的做法或产品经过 7 个问题的审核已无懈可击，便可认为这一做法或产品可取。如果 7 个问题中有一个答复不能令人满意，则表示这方面有改进余地。如果哪方面的答复有独创的优点，就可以扩大产品这方面的效用。5W2H 模型的优势如下：

- 可以准确界定、清晰表述问题，提高工作效率。
- 有效分析事件的本质，完全抓住了事件的主骨架，把事件打回原形思考。
- 简单、方便，易于理解、使用，富有启发意义。
- 有助于思路的条理化，杜绝盲目性。有助于全面思考问题，从而避免在流程设计中遗漏项目。

2.3 战略钟分析模型

战略钟模型（SCM）是由克利夫·鲍曼（Cliff Bowman）提出的。“战略钟”是分析企业竞争战略选择的一种工具，为企业的管理人员和数据分析师提供了一套思考竞争战略和取得竞争优势的方法，如图 2-3 所示。战略钟模型假设不同企业的产品或服务的适用性基本类似，顾客购买时选择其中一家而不是其他企业，可能有以下原因：

- (1) 这家企业的产品和服务的价格比其他公司低。
- (2) 顾客认为这家企业的产品和服务具有更高的附加值。

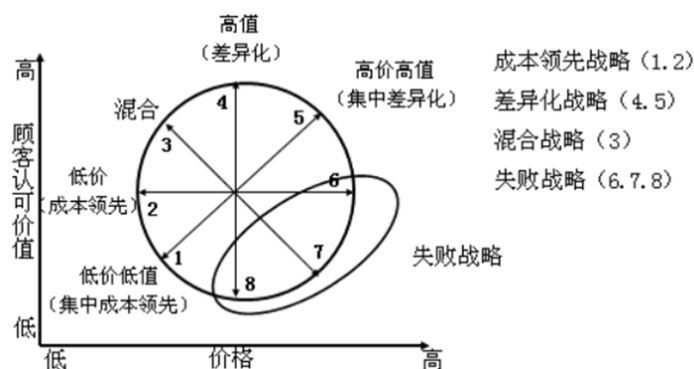


图 2-3 战略钟分析模型

- 低价低附加值战略：采用途径 1 的企业关注的是对价格非常敏感的细分市场的情况。企业采用这种战略是在降低产品或服务附加值的同时降低产品或服务的价格。
- 低价格战略：采用途径 2 的企业是建立企业竞争优势的典型途径，即在降低产品或服务价格的同时，提升产品或服务的质量。这种竞争策略容易被竞争对手模仿。要想通过这一途径获得成功，企业必须取得成本领先地位。
- 混合战略：采用途径 3 的企业在为顾客提供可感知的附加值同时保持低价格。这种高品质低价格的策略能否成功既取决于企业理解和满足客户需求的能力，又取决于是否有保持低价格策略的成本基础，并且难以被模仿。
- 差异化战略：采用途径 4 的企业以相同和略高于竞争对手的价格向顾客提供可感受的附加值，目的是通过提供更好的产品和服务来获得更多的市场份额，或者通过稍高的价格提高收入。企业可以通过采取有形差异化战略或无形差异化战略来获得竞争优势。
- 集中的差异化战略：采用途径 5 的企业可以采用高品质高价格策略在行业中的竞争，即以特别高的价格为用户提供更高的产品和服务的附加值，但是采用这样的竞争策略意味着企业只能在特定的细分市场中参与经营和竞争。
- 高价撇脂战略：采用途径 6、7、8 的企业一般都处在垄断经营地位，完全不考虑产品的成本和附加值。企业采用这种经营战略的前提是市场中没有竞争对手提供类似的产品和服务，否则竞争对手很容易夺得市场份额，并很快削弱该企业的市场地位。

2.4 RFM 客户价值模型

RFM 模型是衡量客户价值和客户创利能力的重要工具和手段。在众多的客户关系管理分析模式中，RFM 模型是被广泛提到的。根据美国数据库营销研究所 Arthur Hughes 的研究，客户数据库中有 3 个神奇的要素，即最近一次消费 (Recency)、消费频率 (Frequency)、消费金额 (Monetary)，这 3 个要素构成了数据分析最好的指标，即如图 2-4 所示。

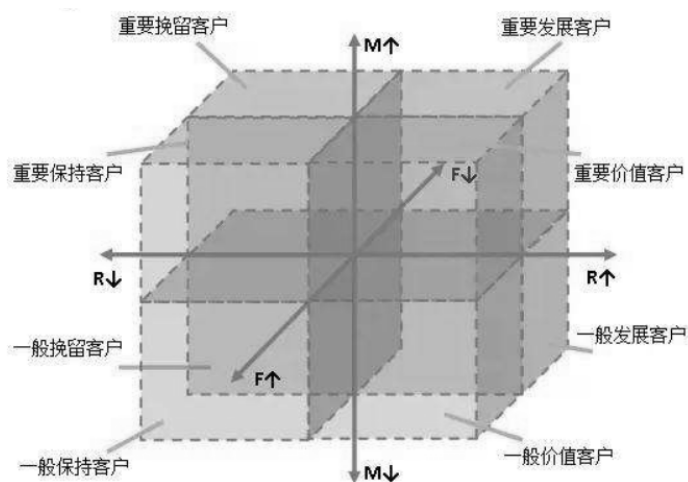


图 2-4 RFM 客户价值模型

在 RFM 模式中，R（Recency）表示客户购买的时间有多远，F（Frequency）表示客户在时间内购买的次数，M（Monetary）表示客户在时间内购买的金额。一般的分析型 CRM 着重在对于客户贡献度的分析，RFM 则强调以客户的行为来区分客户。

（1）最近一次消费

“最近一次消费”是指上一次购买的时候——顾客上一次是几时来店里、上一次根据哪本邮购目录买东西、什么时候买的车，或最近一次在超市买早餐是什么时候。

理论上，最近一次消费时间越近的顾客应该比较好的顾客，对提供即时的商品或是服务也最有可能会有反应。买过你的商品、服务或是曾经光顾你商店的消费者是最有可能再向你买东西的顾客。另外，要吸引一个几个月前才上门的顾客购买，比吸引一个一年多以前来过的顾客要容易得多。

（2）消费频率

消费频率是顾客在限定的期间内所购买的次数。我们可以说最常购买的顾客也是满意度最高的顾客。如果相信品牌及商店忠诚度，那么最常购买的消费者，忠诚度也最高。增加顾客购买的次数意味着从竞争对手处夺取市场占有率，由别人的手中赚取营业额。

（3）消费金额

消费金额是客户在时间内购买的金额，是所有数据库报告的支柱，也可以验证“帕雷托法则”（Pareto's Law），即公司 80% 的收入来自 20% 的顾客。

RFM 模型较为动态地显示了一个客户的全部轮廓，对个性化的沟通和服务提供了依据。同时，如果与该客户打交道的的时间足够长，就能够较为精确地判断该客户的长期价值（甚至是终身价值），通过改善 3 项指标的状况，从而为更多的营销决策提供支持。

2.5 用户行为分析模型

用户行为指用户为获取、使用产品或服务采取的各种行动，首先要认知熟悉，然后试用，再决定是否继续消费使用，最后成为产品或服务的忠实用户。

习惯将用户的行为指标分为三大类，即黏性、活跃和产出，每个分类可以包含多个行为指标来共同衡量用户在这3类中的行为表现，进而区分用户的行为特征，对用户进行分类或者综合评定，如图2-5所示。



图 2-5 用户行为分析模型

- 黏性（Stickiness）主要关注用户在一段时间内持续访问和使用网站的情况，更强调一种持续的状态，这里将“访问频率”和“访问间隔时间”两个指标归到了黏性的分类。
- 活跃（Activity）则更多地针对用户每次的访问过程，考察用户访问中的参与度，所以对统计期中用户的每次访问取了平均值，选择“平均停留时长”和“平均访问页面数”来衡量活跃。黏性和活跃从用户的访问情况衡量用户可能创造的价值。
- 产出（Outcomes）直接根据网站的业务衡量用户创造的直接价值输出，如电子商务网站可以选择“订单数”和“客单价”，一个衡量产出的频率；另一个衡量平均产出价值的大小。

在统计用户行为指标进行分析时，需要注意选择合适的时间段，时间段的长度不能过短，不然无法体现用户长期和持续性的行为特征，黏性指标的分析会不准确；同时短期的用户行为也会误导对用户整体特征和价值的判断，有可能用户在该段时间内极度活跃或者极度低调，也可能用户在短时间内创造了高产出，但从长期看用户创造的价值并没有那么高。

不同的用户分类群体可能会有不同的行为表现，我们可以通过分析各种用户分类的用户行为指标来区分各类用户的特征及对网站的期望要求，进而针对各类用户群体进行调整和定向的营销推广。

可以从3个角度对用户进行细分，即流失用户与留存用户、新用户与老用户、单次购买用户和二次购买用户。基于这3类细分，对每个分类的用户购买商品进行比较分析，计算每类用户的数量，以及占总用户数的比例，每类用户的同比和环比增长率等。此外，还需要注意必须与总体水平的比

较才能得出合理的结论，“与总体比较”不是直接的相减，而是一个差异的幅度体现，只有这样才能明确哪些商品更加符合用户的预期。

2.6 关联分析模型

关联分析又称关联挖掘，就是在交易数据、关系数据或其他信息载体中查找存在于项目集合或对象集合之间的频繁模式、关联、相关性或因果结构。关联分析是一种简单、实用的分析技术，就是发现存在于大量数据集中的关联性或相关性，从而描述一个事物中某些属性同时出现的规律和模式。

关联分析的一个典型例子是购物篮分析，通过合理的啤酒和尿布的货架摆放或捆绑销售可提高超市的服务质量和效益，如图 2-6 所示。该过程通过发现顾客放入其购物篮中的不同商品之间的联系，分析顾客的购买习惯，帮助零售商制定营销策略等。



图 2-6 啤酒与尿布

- 事务：每一条交易称为一个事务。
- 项：交易的每一个物品称为一个项。
- 项集：包含零个或多个项的集合。
- k-项集：包含 k 个项的项集。
- 支持度计数：一个项集出现在几个事务当中，它的支持度计数就是几。
- 支持度：支持度计数除以总的事务数。
- 频繁项集：支持度大于或等于某个阈值的项集。
- 前件和后件：对于规则 $A \rightarrow B$ ，A 叫作前件，B 叫作后件。
- 置信度：对于规则 $A \rightarrow B$ ， $\{A, B\}$ 的支持度计数除以 A 的为这个规则的置信度。
- 强关联规则：大于或等于最小支持度阈值和最小置信度阈值的规则。

Apriori 算法是挖掘产生布尔关联规则所需频繁项集的基本算法，也是最著名的关联规则挖掘算法之一。Apriori 算法是根据有关频繁项集特性的先验知识而命名的。它使用一种称作逐层搜索的迭代方法，k-项集用于探索 (k+1)-项集。首先，找出频繁 1-项集的集合，记作 L_1 ， L_1 用于找出频繁 2-项集的集合 L_2 ，再用于找出 L_3 ，如此下去，直到不能找到频繁 k-项集。找出每个 L_k 都

需要扫描一次数据库。

由于 Apriori 方法的效率仍然不能令人满意。2000 年，Han Jiawei 等人提出了基于频繁模式树（FP-tree）的发现频繁模式的算法 FP-growth。它通过两次扫描事务数据库，把每个事务所包含的频繁项目按其支持度降序压缩存储到 FP-tree 中。在以后发现频繁模式的过程中，不需要再扫描事务数据库，而仅在 FP-Tree 中进行查找即可，并通过递归调用 FP-growth 的方法来直接产生频繁模式，因此在整个发现过程中也不需要产生候选模式。

2.7 SWOT 分析模型

来自麦肯锡咨询公司的 SWOT 分析方法是一种常见的战略分析模型，包括分析企业的优势（Strengths）、劣势（Weaknesses）、机会（Opportunities）和威胁（Threats），如图 2-7 所示。因此，SWOT 分析实际上是对企业内外部条件各方面内容进行综合和概括，进而分析组织的优劣势、面临的机会和威胁的一种方法。通过 SWOT 分析，可以帮助企业把资源和行动聚集在自己的强项和有最多机会的地方，并让企业的战略变得明朗。



图 2-7 SWOT 分析模型

SWOT 分析法即态势分析法，就是将与研究对象密切相关的各种主要内部优势、劣势和外部的机会和威胁等通过调查列举出来，并依照矩阵形式排列，然后用系统分析的思想把各种因素相互匹配起来加以分析，从中得出的一系列相应的结论，而结论通常带有一定的决策性。运用这种方法，可以对研究对象所处的情景进行全面、系统、准确的研究，从而根据研究结果制定相应的发展战略、计划以及对策等。

（1）优势与劣势分析（SW）

由于企业是一个整体，并且竞争优势来源非常广泛，因此在做优劣势分析时必须从整个价值链的每个环节上将企业与竞争对手做详细的对比，例如产品是否新颖、制造工艺是否复杂、销售渠道是否畅通以及价格是否具有竞争性等。衡量一个企业及其产品是否具有竞争优势只能站在现有潜在用户角度上，而不是站在企业的角度上。

(2) 机会与威胁分析 (OT)

企业必须分析替代品给公司的产品或服务带来的是“灭顶之灾”还是提供了更高的利润或价值；购买者转而购买替代品的转移成本；公司可以采取什么措施来降低成本或增加附加值来降低消费者购买盗版替代品的风险。

从整体上看，SWOT 可以分为两部分：第一部分为 SW，主要用来分析内部条件；第二部分为 OT，主要用来分析外部条件。利用这种方法可以从中找出对自己有利的、值得发扬的因素，以及对对自己不利的、要避开的东西，发现存在的问题，找出解决办法，并明确以后的发展方向。

2.8 鱼骨图分析模型

鱼骨图指的是一种发现问题“根本原因”的分析方法。鱼骨图由日本管理大师石川馨先生所发明，故又名石川图。鱼骨图是一种发现问题“根本原因”的方法，也可以称为“Ishikawa”或者“因果图”。其特点是简捷实用，深入直观。它看上去有些像鱼骨，问题或缺陷（后果）标在“鱼头”处，如图 2-8 所示。

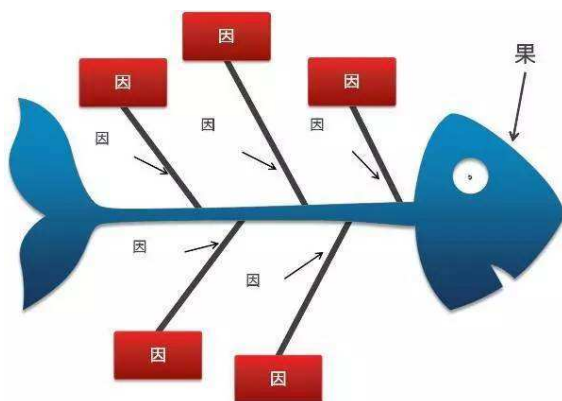


图 2-8 鱼骨图分析模型

问题的特性总是受到一些因素的影响，我们通过头脑风暴法找出这些因素，并将它们与特性值一起，按相互关联性整理而成的层次分明、条理清楚，并标出重要因素的图形就叫特性要因图、特性原因图。因其形状如鱼骨，所以叫鱼骨图。它是一种透过现象看本质的分析方法。鱼骨图也用在生产中，用来形象地表示生产车间的流程。

制作鱼骨图的两个步骤：分析问题原因/结构、绘制鱼骨图。

(1) 分析问题原因/结构

- A. 针对问题点，选择层别方法（如人、机、料、法、环等）。
- B. 按头脑风暴分别对各层别类别找出所有可能原因（因素）。
- C. 将找出的各要素进行归类、整理，明确其从属关系。
- D. 分析选取重要因素。
- E. 检查各要素的描述方法，确保语法简明、意思明确。

(2) 鱼骨图的绘图过程

- A. 填写鱼头（按为什么不好的方式描述），画出主骨。
- B. 画出大骨，填写大要因。
- C. 画出中骨、小骨，填写中小要因。
- D. 用特殊符号标识重要因素。



注意 绘图时，应保证大骨与主骨成 60 度夹角，中骨与主骨平行。

比如，某炼油厂情况，采用鱼骨图分析法对其市场营销问题进行解析，“鱼头”部分表示需要解决的问题，即该炼油厂产品在市场中所占份额少。可以把产生该炼油厂市场营销问题的原因概括为 5 类，即人员、渠道、广告、竞争和其他。在每一类中包括若干造成这些原因的可能因素，如营销人员数量少、销售点少、缺少宣传策略、进口油广告攻势等。将 5 类原因及其相关因素分别以鱼骨分布态势展开，形成鱼骨分析图。

2.9 波士顿矩阵分析

波士顿矩阵（BCG Matrix）又称市场增长率-相对市场份额矩阵、四象限分析法等。波士顿矩阵由美国著名的管理学家、波士顿咨询公司创始人布鲁斯·亨德森于 1970 年首创。

波士顿矩阵认为一般决定产品结构的基本因素有两个，即市场引力与企业实力。市场引力包括整个市场的销售量（额）增长率、竞争对手强弱及利润高低等，其中最主要的是反映市场引力的综合指标——销售量增长率。企业实力包括市场占有率、技术、设备、资金利用能力等。其中，市场占有率是决定企业产品结构的内在要素，直接显示出企业竞争实力。

销售量增长率与市场占有率既相互影响又互为条件，两个因素相互作用，会出现 4 种不同性质的产品类型：①销售量增长率和市场占有率“双高”的产品群（明星类产品）；②销售量增长率和市场占有率“双低”的产品群（瘦狗类产品）；③销售量增长率高、市场占有率低的产品群（问题类产品）；④销售量增长率低、市场占有率高的产品群（金牛类产品），如图 2-9 所示。

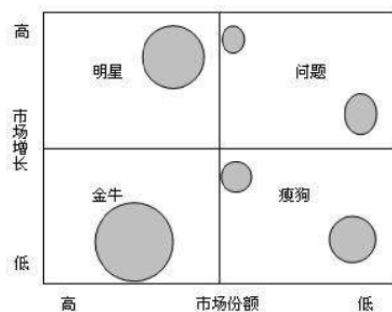


图 2-9 波士顿矩阵分析

- 金牛：在低增长市场上具有相对高的市场份额的业务将产生健康的现金流，它们能用于

向其他方面提供资金，发展业务。

- **瘦狗**：在低增长市场上具有相对低的市场份额的业务经常是中等现金流的使用者。由于其虚弱的竞争地位，它们将成为现金的陷阱。
- **明星**：在高增长市场上具有相对高的市场份额通常需要大量的现金以维持增长，但具有较强的市场地位并将产生较高的报告利润，它们有可能处在现金平衡状态。
- **问题**：在迅速增长的市场上具有相对较低市场份额的业务需要大量的现金流入，以便为增长筹措资金。

按照波士顿矩阵的原理，产品市场占有率越高，创造利润的能力越大；另外，销售量增长率越高，为了维持其增长及扩大市场占有率所需的资金亦越多。这样可以使企业的产品结构实现产品互相支持、资金良性循环的局面。按照产品在象限内的位置及移动趋势的划分，形成了波士顿矩阵的基本应用法则。

如果某一产品从问题产品（包括从瘦狗产品）变成金牛产品的移动速度太快，说明其在高投资与高利润率的明星区域的时间很短，因此对企业提供利润的可能性及持续时间都不会太长，总的贡献也不会大；相反，如果产品发展速度太慢，在某一象限内停留时间过长，则该产品也会很快被淘汰。

2.10 波特五力分析模型

波特五力模型是迈克尔·波特（Michael Porter）于20世纪70年代初提出的，被广泛应用于很多行业的战略制定。波特认为在任何行业中，无论是国内还是国际，无论是提供产品还是提供服务，竞争的规则都包括在5种竞争力量内。这5种竞争力就是企业间的竞争、潜在新竞争者的进入、潜在替代品的开发、供应商的议价能力、购买者的议价能力，如图2-10所示。这5种竞争力量决定了企业的盈利能力和水平。

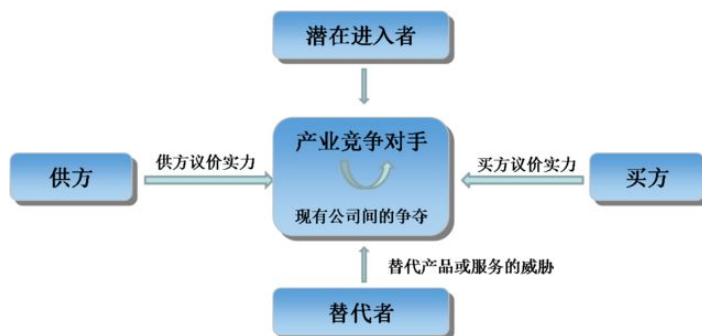


图 2-10 波特五力分析模型

1. 供应商的议价能力

供方主要通过其提高投入要素价格与降低单位价值质量的能力来影响行业中现有企业的盈利

能力与产品竞争力。供方力量的强弱主要取决于他们所提供给买主的是什么投入要素，当供方所提供的投入要素价值构成了买主产品总成本的较大比例、对买主产品生产过程非常重要或者严重影响买主产品的质量时，供方对于买主的潜在讨价还价力量就会增强。

2. 购买者的议价能力

购买者主要通过其压价与要求提供较高的产品或服务质量的能力来影响行业中现有企业的盈利能力。

3. 新进入者的威胁

新进入者在给行业带来新生产能力、新资源的同时，将希望在已被现有企业瓜分完毕的市场中赢得一席之地，这就有可能会与现有企业发生原材料与市场份额的竞争，最终导致行业中现有企业盈利水平降低，严重的话还有可能危及这些企业的生存。

4. 替代品的威胁

两个处于同行业或不同行业中的企业，可能会由于所生产的产品是互为替代品，从而在它们之间产生相互竞争行为，这种源自于替代品的竞争会以各种形式影响行业中现有企业的竞争战略。

5. 同业竞争者的竞争程度

大部分行业中的企业，相互之间的利益都是紧密联系在一起的，作为企业整体战略一部分的各企业竞争战略，其目标都在于使得自己的企业获得相对于竞争对手的优势，所以在实施中就必然会产生冲突与对抗现象，这些冲突与对抗就构成了现有企业之间的竞争。此外，现有企业之间的竞争常常表现在价格、广告、产品介绍、售后服务等方面。

2.11 PDCA 执行模型

PDCA 循环是美国质量管理专家休哈特博士首先提出的，由戴明采纳、宣传，获得普及，所以又称戴明环。PDCA 循环的含义是将质量管理分为 4 个阶段，即计划（Plan）、执行（Do）、检查（Check）、纠正（Action），如图 2-11 所示。在质量管理活动中，要求把各项工作按照做出计划、计划实施、检查实施效果，然后将成功的纳入标准、不成功的留待下一个循环去解决。



图 2-11 PDCA 执行模型

(1) P (Plan, 计划) 包括方针和目标的确定，以及活动规划的制定。

(2) D (Do, 执行) 根据已知的信息, 设计具体的方法、方案和计划布局; 再根据设计和布局进行具体运作, 实现计划中的内容。

(3) C (Check, 检查) 总结执行计划的结果, 分清哪些对了或错了, 明确效果, 找出问题。

(4) A (Action, 纠正) 对总结检查的结果进行处理, 对成功的经验加以肯定, 并予以标准化; 对于失败的教训, 引起重视; 对于没有解决的问题, 提交给下一个 PDCA 循环去解决。

PDCA 循环可以使我们的思想方法和工作步骤更加条理化、系统化、图像化和科学化, 具有如下特点:

(1) 大环套小环、小环保大环、推动大循环。

PDCA 循环作为质量管理的基本方法, 各级部门根据企业的方针目标, 都有自己的 PDCA 循环, 层层循环, 形成大环套小环, 小环里面套更小的环。大环是小环的母体和依据, 小环是大环的分解和保证。

(2) 不断前进、不断提高。

PDCA 循环就像爬楼梯, 一个循环运转结束, 生产的质量就会提高一步, 然后制定下一个循环, 再运转、再提高, 不断前进, 不断提高。

(3) 门路式上升。

PDCA 循环不是在同一水平上循环, 而是每循环一次就解决一部分问题, 取得一部分成果, 工作就前进一步, 水平就提高一步。每通过一次 PDCA 循环, 都要进行总结, 提出新目标, 再进行第二次 PDCA 循环, 使品质治理的车轮滚滚向前。

2.12 AARRR 模型

AARRR 也叫用户增长模型, 是 Acquisition、Activation、Retention、Revenue、Refer 这 5 个单词的缩写, 分别对应用户生命周期中的 5 个重要环节, 如图 2-12 所示。模型的提出者认为, 所有创新型、成长型的企业都应该按照这个模型来做增长。



图 2-12 AARRR 模型

(1) 用户获取 (Acquisition)

在这个阶段，最关心的数据是下载量。通常用下载量来衡量一个应用的用户规模和是否成功。不过，下载了应用不等于一定会安装，安装了应用也不等于一定会使用该应用。所以激活量成为这个层次中大家最关心的数据。此外，CAC（用户获取成本）也是需要去关注的的数据。

(2) 提高活跃度 (Activation)

首先需要关注的指标是 DAU（日活跃用户）、MAU（月活跃用户）。这两个指标基本上说明了应用当前的用户群规模。用户启动不等于活跃，每次启动时间很短，活跃度其实并不高。所以其实还要看另外两个指标：每次启动平均使用时长和每个用户每日平均启动次数。

(3) 提高留存率 (Retention)

下载、安装、使用、卸载或者遗忘，这是用户在每个应用中的生命周期。成功的应用就是那些能尽量延长用户的使用寿命、最大化价值的应用。对于大部分应用，应该关心的是 1-Day Retention 和 7-Day Retention。通常用户新安装使用后的前几天是流失比例最大的时期。

(4) 获取收入 (Revenue)

关于收入，对应的指标是 ARPU（平均每个用户收入）、ARPPU（平均每付费用户收入）。两者之间的关系与付费用户占全部用户的比例有关。ARPU 是和时间段相关的指标，还要多看一个指标：LTV（生命周期价值）。LTV 是用户在生命周期内创造的收入总计，可以看成是一个长期累计的 ARPU 值。用户平均的 LTV = 每月 ARPU * 用户按月计的平均生命周期。LTV - CAC 的差值可以视为该应用从每个用户身上获取的利润。

(5) 自传播 (Refer)

自传播（或者说病毒式营销）是最近 10 年才被广泛研究的营销方法。K 因子这个术语并非起源于市场学或软件业，而是来源于传染病学——对，就是研究真正的病毒传播的科学。K 因子量化了感染的概率，即一个已经感染了病毒的宿主所能接触到的所有宿主中会有多少宿主被其传染上病毒。

2.13 ABC 分类模型

ABC 分类法 (Activity Based Classification) 的全称应为 ABC 分类库存控制法，又称帕累托法、柏拉图分析、主次因分析法，平常我们也称之为“80 对 20”规则，EMBA、MBA 等主流商管教育均有介绍 ABC 分类法，对管理者组织决策等都有重要影响。

ABC 分类法是项目管理中常用的一种方法。它是根据事物在技术或经济方面的主要特征进行分类排队，分清重点和一般，从而有区别地确定管理方式的一种分析方法。因为它把被分析的对象分成 A、B、C 三类，所以又称为 ABC 分析法，如图 2-13 所示。

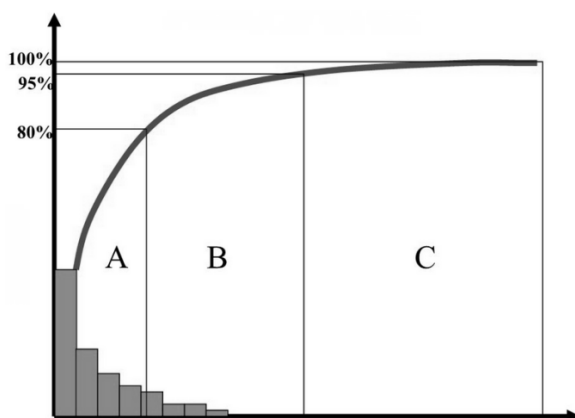


图 2-13 ABC 分类模型

ABC 分类法是由意大利经济学家维尔弗雷多·帕累托首创的，是储存管理中常用的分析方法，也是经济工作中一种基本工作和认识方法。ABC 分析的应用在储存管理中比较容易取得以下成效：第一，压缩总库存量；第二，解放被占压的资金；第三，使库存结构合理化；第四，节约管理力量。

1879 年，帕累托在研究个人收入的分布状态时发现少数人的收入占全部人收入的大部分，而多数人的收入却只占一小部分，并将这一关系用图表示出来，就是著名的帕累托图。该分析方法的核心思想是在决定一个事物的众多因素中分清主次，识别出少数的但对事物起决定作用的关键因素和多数但对事物影响较少的次要因素。1951 年，管理学家戴克（H.F.Dickie）将其应用于库存管理，命名为 ABC 法。1963 年，彼得·德鲁克（P.F.Drucker）将这一方法推广到全部社会现象，使 ABC 法成为企业提高效益、普遍应用的管理方法。

在 ABC 分析法的分析图中，有两个纵坐标、一个横坐标、几个长方形、一条曲线。左边纵坐标表示频数，右边纵坐标表示频率，以百分数表示。横坐标表示影响质量的各项因素，按影响大小从左向右排列。曲线表示各种影响因素大小的累计百分数。一般地，是将曲线的累计频率分为 3 级，与之相对应的因素分为 3 类：

- A 类因素，发生累计频率为 0%~80%，是主要影响因素。
- B 类因素，发生累计频率为 80%~95%，是次要影响因素。
- C 类因素，发生累计频率为 95%~100%，是一般影响因素。

2.14 杜邦分析模型

企业在对经营情况进行分析时，大多主要利用财务报表数字通过财务指标计算反映其盈利能力、营运能力、财务风险、股东回报水平等，其中最有代表性的应属杜邦分析模型，这种分析方法最早由美国杜邦公司使用，故名杜邦分析法，如图 2-14 所示。

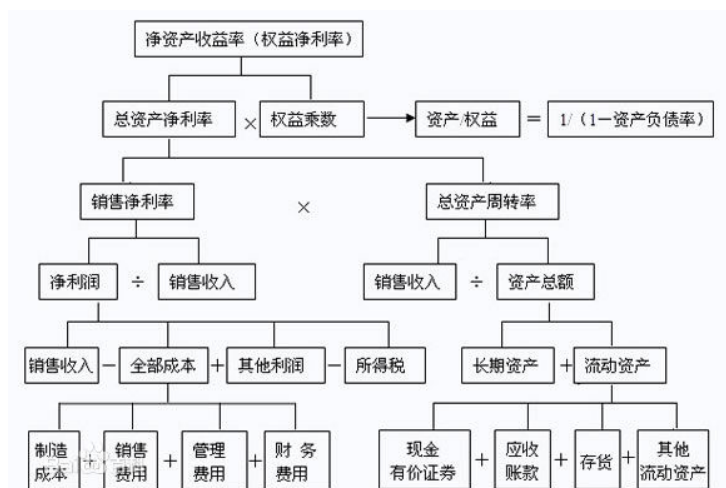


图 2-14 杜邦分析模型

杜邦分析模型通过对净资产收益率的分解，从销售净利率、总资产周转率、权益乘数三方面考察对股东回报的影响，而这三方面又涵盖了采购管理、成本费用控制、资产营运水平、财务杠杆的税盾效应等更深入的因素，从而由表及里、由外而内、由果至因地体现历时财务状况。

杜邦分析法的基本思路如下：

(1) 权益净利率也称权益报酬率，是综合性最强的财务分析指标，是杜邦分析的核心。

(2) 总资产净利率是影响权益净利率最重要的指标，具有很强的综合性。资产净利率取决于销售净利率和总资产周转率的高低。总资产周转率反映总资产的周转速度。

(3) 权益乘数表示企业的负债程度，反映公司利用财务杠杆进行经营活动的程度。资产负债率高，权益乘数就大，说明公司负债程度高，公司会有较多的杠杆利益，但风险也高；反之，资产负债率低，权益乘数就小，说明公司负债程度低，公司会有较少的杠杆利益，但相应所承担的风险也低。

从企业绩效评价的角度来看，杜邦分析法只包括财务方面的信息，不能全面反映企业的实力，有很大的局限性。在实际运用中需要加以注意，必须结合企业的其他信息加以分析。

(1) 对短期财务结果过分重视，有可能助长公司管理层的短期行为而忽略长期的价值创造。

(2) 财务指标反映的是企业过去的经营业绩，衡量工业时代的企业能够满足要求；但在目前的信息时代，顾客、供应商、雇员、技术创新等因素对企业经营业绩的影响越来越大，杜邦分析法在这些方面是无能为力的。

(3) 在目前的市场环境中，企业的无形资产对提高企业长期竞争力至关重要，杜邦分析法却不能解决无形资产的估值问题。

2.15 平衡计分卡模型

杜邦分析模型等指标设计看似非常精巧,由于财务分析主要依据过去的数据或过去经济活动在分析时点结果的数据,因此它就成了企业过往生产经营成果的集中体现。利用这些数据,也只能分析过去干得如何,那么如何对企业经营未来的情况做预测呢?答案是依仗另一个模型——平衡计分卡模型。

平衡计分卡是从财务、顾客、内部流程、学习成长4个角度,将组织的战略落实为可操作的衡量指标和目标值的一种新型绩效管理体系,如图2-15所示。设计平衡计分卡的目的是要建立“实现战略制导”的绩效管理体系,从而保证企业战略得到有效的执行。因此,人们通常称平衡计分卡是加强企业战略执行力最有效的战略管理工具。

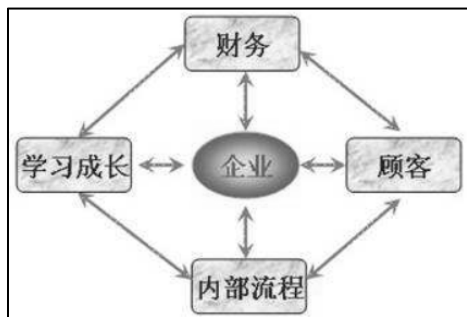


图 2-15 平衡计分卡模型

平衡记分卡方法的引入改变了企业以往只关注财务指标的考核体系缺陷。仅仅关注财务指标会使企业过分关注一些短期行为而牺牲一些长期利益,比如员工的培养和开发、客户关系的开拓和维护等。平衡记分卡的优点在于从企业的4个方面来建立衡量体系,即财务、顾客、业务管理、人员的培养和开发。这4个方面是相互联系、相互影响的,其他3类指标的实现最终保证了财务指标的实现。

实际上,平衡计分卡方法打破了传统的只注重财务指标的业绩管理方法。平衡计分卡方法认为,组织应从4个角度审视自身业绩,即学习成长、内部流程、顾客、财务,包含5项平衡:

(1) 财务指标和非财务指标的平衡。企业考核的一般是财务指标,而对非财务指标(顾客、内部流程、学习成长)的考核很少,即使有对非财务指标的考核,也只是定性的说明,缺乏量化的考核,缺乏系统性和全面性。

(2) 企业的长期目标和短期目标的平衡。平衡计分卡是一套战略执行的管理系统,如果以系统的观点来看平衡计分卡的实施过程,那么战略是输入、财务是输出。

(3) 结果性指标与动因性指标之间的平衡。平衡计分卡以有效完成战略为动因,以可衡量的指标为目标管理的结果,寻求结果性指标与动因性指标之间的平衡。

(4) 企业组织内部群体与外部群体的平衡。在平衡计分卡中,股东与客户为外部群体,员工和内部业务流程是内部群体,可以平衡这些群体间利益的重要性。

(5) 领先指标与滞后指标之间的平衡。财务指标是一个滞后指标，只能反映公司上一年度发生的情况，不能告诉企业如何改善业绩和可持续发展。顾客、内部流程、学习成长这3项指标可使企业达到领先指标和滞后指标之间的平衡。

2.16 营销漏斗模型

营销漏斗模型（见图 2-16）常用于营销过程，是将非潜在客户逐步变为客户的转化量化模型。营销漏斗模型的价值在于量化营销过程各个环节的效率，帮助找到薄弱环节。

营销的环节指的是从获取用户到最终转化成购买整个流程中的一个子环节。相邻环节的转化率是指用数据指标来量化每一个步骤的表现。整个漏斗模型就是先将整个购买流程拆分成一个个步骤，然后用转化率来衡量每一个步骤的表现，最后通过异常的数据指标找出有问题的环节，从而解决问题、优化步骤，最终达到提升整体购买转化率的目的。整体漏斗模型的核心思想可以归为分解和量化。



图 2-16 营销漏斗模型

比如分析电商的转化，我们要做的就是监控每个层级上的用户转化，寻找每个层级的可优化点。对于没有按照流程操作的用户，专门绘制他们的转化模型，缩短路径提升用户体验。

用户生命周期衍生于“客户生命周期”的概念。

用户从不同角度有不同的分类法：

- 从用户活跃度来看，可以分为僵尸用户、低频用户、活跃用户和深度用户
- 从用户对平台的价值来看，可以分为种子用户、普通用户、核心用户。
- 从用户价值来看，分为无效用户、潜在用户、跟进用户、成单用户。

不同的行业有不同的分类，不同的分类也有不同的运营策略。从用户生命周期的角度来解析和管理用户是常见的数据分析方式。

从用户活跃度来看，可能就要构建一个用户活跃度模型：

- 流失用户：有一段时间没有再打开产品（曾经打开过产品），那么我们就视为流失用户，根据产品的属性可以按30天、60天、90天等划分。
- 不活跃用户：有一段时间没有使用产品，为了和流失区分开来，需要选择无交集的时间范围。例如，流失用户是60天以上没打开产品，那么不活跃则是0~60天没打开。
- 回流用户：有一段时间没用产品，之后突然回来再次使用，就称为回流用户。回流用户是活跃用户，且是由流失用户或不活跃用户召唤回来的。
- 活跃用户：一段时间内打开过产品。
- 忠诚用户：也可以叫超级活跃用户，长期持续使用产品，比如连续四周或者一个月内活跃15天等。

用户活跃可以简化为一个简单的公式：新增用户的数量要大于流失用户的增加量。不同业务情况需要各自衡量怎么去构建这个模型。

2.17 SCP 分析模型

SCP 分析模型(见图 2-17)由美国哈佛大学产业经济学权威乔·贝恩(Joe S.Bain)、谢勒(Scherer)等人于 20 世纪 30 年代建立。该模型提供了一个既能深入具体环节又有系统逻辑体系的市场结构(Structure)一市场行为(Conduct)一市场绩效(Performance)的产业分析框架。SCP 框架的基本含义是,市场结构决定企业在市场中的行为,而企业行为又决定市场运行在各个方面的经济绩效。



图 2-17 SCP 分析模型

SCP 分析模型主要用于分析企业经营者受到外部冲击时企业的战略调整及行为变化，从特定行业结构、企业行为和经营绩效 3 个角度来分析外部冲击的影响。

- 外部冲击：主要是指企业外部经济环境、政治、技术、文化、消费习惯等因素的变化。
- 行业结构：主要是指外部各种环境的变化对企业所在行业可能的影响，包括行业竞争的变化、产品需求的变化、细分市场的变化、营销模型的变化等。
- 企业行为：主要是指企业针对外部的冲击和行业结构的变化，有可能采取的应对措施，包括企业方面对相关业务单元的整合、业务的扩张与收缩、营运方式的转变、管理的变革等一系列变动。
- 经营绩效：主要是指在外部环境方面发生变化的情况下，企业在经营利润、产品成本、市场份额等方面的变化趋势。

应用 SCP 分析模型的场景如下：

- (1) 分析外部竞争环境对企业造成的冲击。

通过对替代产品可获得性、产品差异性、增长率、变更性/周期性来分析外部竞争环境对企业造成的冲击。

(2) 从行业结构分析外部冲击对企业造成的影响。

行业结构是指特定的市场中的企业在数量、份额、规模上的关系。

(3) 从企业行为分析外部冲击对企业造成的影响。

企业行为是市场结构、经济绩效的联系纽带，通过各种策略对潜在进入者施加压力，从而影响市场结构。分析的角度包含营销（定价、产量、广告、新产品研发）、产能变化、企业纵向整合、内部运营效率（成本控制、物流、过程发展、组织效能）等。

(4) 从经营绩效分析外部冲击对企业造成的影响。

经营绩效是指特定市场结构下通过特定企业行为使某一产业在价格、产量、成本、利润、产品质量、品种及技术进入等方面达到的状态，主要是通过对财务方面的营利性以及价值创造、科技的发展、雇佣对象的变化来分析企业当前的经营绩效。

2.18 逻辑树分析模型

麦肯锡逻辑树又称问题树、演绎树或分解树等，很多咨询公司分析问题最常使用的工具就是“逻辑树”。逻辑树是将问题的所有子问题分层罗列，从最高层开始，并逐步向下扩展，如图 2-18 所示。

把一个已知问题当成树干，然后开始考虑这个问题和哪些相关问题或者子任务有关。每想到一点，就给这个问题（也就是树干）加一个“树枝”，并标明这个“树枝”代表什么问题。一个大的“树枝”上还可以有小的“树枝”，以此类推，找出问题的所有相关联项目。逻辑树主要是帮助我们理清自己的思路，不进行重复和无关的思考。

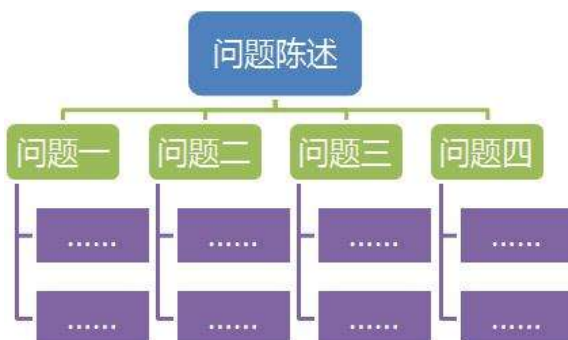


图 2-18 逻辑树分析模型

那逻辑树怎么运用于数据分析呢？

一个门店近半年销售额下滑很严重，可是不知道到底是什么问题。这个时候就可以按照逻辑树的分析框架进行分析。把所有影响展现量的因素都罗列出来，然后一个一个地进行分析，看是不是

这个因素的原因，这样就可以把店铺的问题都找出来了。

逻辑树分析法最大的好处就是可以帮你理清分析的思路。如果没有这种分析，可能会很乱，一下分析这个，一下分析那个，进行了很多重复的分析、很多没有意义的思考。先做好逻辑树框架之后，基本上就能避免这些问题。

逻辑树分析也有缺点，就是它依赖于一个人的经验和对业务的专业程度。例如，如果你不知道影响展现量的因素有哪些，那么就算分析到了出现问题在于展现量，你也没有办法继续下面的分析。如果你经验不足、专业能力不强，就很容易漏掉很重要的地方。例如，你对竞争环境分析能力不强，就可能会漏掉这一块，而真正影响展现量的可能就是竞争环境。因此，出现这种情况的话就会走偏了。

2.19 波士顿三四矩阵

波士顿三四矩阵是由波士顿咨询集团（BCG）提出的，主要用于分析一个成熟市场中企业的竞争地位，如图 2-19 所示。

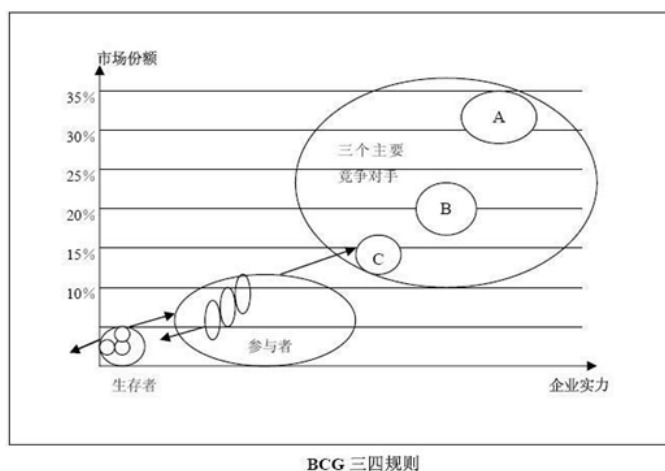


图 2-19 波士顿三四矩阵

在一个稳定的竞争市场中，参与市场竞争的参与者一般分为 3 类：领先者、参与者、生存者。领先者一般是指市场占有率在 15% 以上，可以对市场变化产生重大影响的企业，如在价格、产量等方面；参与者一般是指市场占有率在 5%~15% 之间的企业，这些企业虽然不能对市场产生重大的影响，但是它们是市场竞争的有效参与者；生存者一般是局部细分市场填补者，这些企业的市场份额都非常低，通常小于 5%。

在有影响力的领先者之中，企业的数量绝对不会超过 3 个，而在这 3 个企业之中最有实力的竞争者的市场份额又不会超过最小者的 4 倍。这个模型是由下面两个条件决定的：

(1) 在任何两个竞争者之间，2 比 1 的市场份额似乎是一个均衡点。在这个均衡点上，无论

哪个竞争者要增加或减少市场份额都显得不切实际，而且得不偿失。这是一个通过观察得出的经验性结论。

(2) 市场份额小于最大竞争者的 $1/2$ 就不可能有效参与竞争。这也是经验性结论，但是不难从经验曲线的关系中推断出来。

通常，上述两个条件最终导致这样的市场份额序列：每个竞争者的市场份额都是紧随其后的竞争者的 1.5 倍，而最小的竞争者的市场份额不会小于最大者的 $1/4$ 。

“三四规则”只是从经验中得出的一种假设，并没有经过严格的证明，但是这个规则的意义非常重要：在经验曲线的效应下，成本是市场份额的函数。倘若两个竞争者拥有几乎相同的市场份额，那么谁能提高相对市场份额，谁就能同时取得在产量和成本两个方面的增长；与所付出的代价相比，得到的可能会更多。但是对市场竞争的领先者而言，可能得到的好处反而少了。在任何主要竞争者的激烈争夺情况下，最有可能受到伤害的则是市场中最弱小的生存者。

这个理论正好可以解释中国彩电行业的几次降价行为、各个企业竞争态势的变化。在长虹第一次降价后，企业的成本和产量都得到收益，使它迅速成为市场占有率最大的企业。康佳和 TCL 则紧随其后。经过几次的降价后，企业已经不能再靠价格因素来扩大市场份额了，企业必须创造新的竞争优势。

2.20 练习题

1. 结合实际工作，列举商业数据分析的主要模型。
2. 详细阐述 RFM 客户价值模型的基本概念及应用过程。
3. 详细阐述用户行为分析模型的基本概念及应用过程。
4. 详细阐述营销漏斗模型的基本概念及应用过程。
5. 结合日常工作介绍一种本书中没有介绍的模型。

第 2 篇

Tableau 之新手入门篇

本篇我们将介绍 Tableau 的基础知识，包括 Tableau Desktop 的简介，连接到文件和数据库等数据源，Tableau 的维度和度量、连续和离散等基本操作，Tableau 的表计算、创建字段、创建参数、聚合函数等高级操作，Tableau 制作条形图、饼形图、散点图、树形图、环形图、漏斗图等 14 类视图。这一篇是读者学好 Tableau 软件必须要掌握的技能，同时也是学习后续内容的基础。

第 3 章

Tableau Desktop 基础入门

Tableau Desktop 是基于斯坦福大学突破性技术的软件应用程序，可以分析实际存在的所有结构化数据，可以在几分钟内生成美观的图表、坐标图、仪表盘与报告。

利用 Tableau 简便的拖放式可以自定义视图、布局、形状、颜色等，帮助展现自己的数据视角。本章将详细介绍 Tableau Desktop 2019.1 的新功能、数据类型、运算符等内容。

3.1 Tableau Desktop 简介

我们可以在 Tableau 的官方网站 (<http://www.Tableau.com/zh-cn/products/trial>) 下载最新的免费试用版本。在安装 Tableau Desktop 之前，我们首先需要确保计算机满足条件：操作系统为 Windows Server 2008\2012、Windows 7\8\8.1\10。此外，Tableau Public 的安装过程与 Tableau Desktop 基本相同。

Tableau Desktop 的安装比较简单，这里不做介绍。安装 Tableau Desktop 后，可以通过双击桌面上的 Tableau Desktop 图标打开软件。此外，Tableau 文件通常存储在“我的 Tableau 存储库”文件夹中，该文件夹一般位于“文档”文件夹中，如图 3-1 所示。

名称	修改日期	类型	大小
地图源	2018/9/18 5:10	文件夹	
服务	2018/12/15 7:34	文件夹	
工作簿	2018/9/18 5:10	文件夹	
扩展	2018/9/18 5:10	文件夹	
日志	2019/2/19 21:32	文件夹	
书签	2018/9/18 5:10	文件夹	
数据源	2018/9/18 5:10	文件夹	
形状	2018/9/18 5:10	文件夹	
Preferences.tps	2018/9/18 5:10	Tableau 首选项文...	1 KB

图 3-1 我的 Tableau 存储库

3.1.1 新增主要功能

截至 2019 年 3 月，Tableau Desktop 的最新版本是 2019.1。它是一款功能强大的可视化数据分析软件，与之前的版本相比，该版本的功能有较大的提升，新功能包括 Ask Data、Tableau Prep Conductor、Tableau Mobile 应用、导出到 PowerPoint、Google Adwords 连接器等，大大提高了 Tableau 的数据处理能力和分析能力。

1. Ask Data

通过 Ask Data（数据问答），用户只需简单地输入一个问题，例如“本月我的销售额是多少？”Tableau 将会给予用户一个交互式的可视化结果，在此基础上，用户可以继续探索，提出新的问题。在这一过程中，用户无须深入了解数据结构，也无须其他设置或掌握编程技能。

Ask Data 背后的复杂算法是通过对个人意图的理解而驱动，而不是通过关键词驱动，有助于得出更为相关的结果。Ask Data 可以使用于任何已发布的数据源，无须改动数据或训练算法。此外，Ask Data 的解析逻辑可自动断句、解读模糊的语言，使人们可以轻松地以自然口语的方式提问。

2. Tableau Prep Conductor

Tableau 2019.1 还引入了 Tableau Prep Conductor，它是名为 Tableau Data Management 包的一部分。Tableau Prep Conductor 是适用于 Tableau Online 和 Tableau Server 的新插件产品，让 IT 部门、数据管家和分析师能够在安全可靠的可扩展服务器环境中以一致的方式提供准备好的数据。通过提供可视化、细节的清理历史记录、数据连接背后的详细信息以及流程不合规运行时的警报，Tableau Prep Conductor 使人们对其数据更加信赖。

3. 全新的 iOS 和 Android 版 Tableau Mobile

重新设计的 Tableau Mobile 提供 iOS 和 Android 的通用访问，可以带来更加直观的体验，在 Tableau 2019.1 发布之后推出。Tableau Mobile 使 Tableau Server 和 Tableau Online 仪表盘触手可及，可以随时了解自己的数据。全新的应用软件提供了同步收藏夹和离线交互功能，其中包括突出显示和工具提示，因此即便在旅途中也可以访问数据，利用强大的集成搜索来浏览项目并查找具体内容。

4. 混合内容类型

新增在 Tableau Server 和 Tableau Online 上浏览内容的新方法，可以在同一个页面查看任一项目中的全部数据源、工作簿和视图，还可以在项目范围内按内容类型进行筛选。

5. 导出到 PowerPoint

通过一键式操作将 Tableau 仪表盘直接整合到演示文稿中。现在，能够以高分辨率图片的形式将可视化从 Tableau Server 或 Tableau Online 轻松导出到 PowerPoint，同时提供指向原始工作簿的链接。

6. 自动化仪表盘手机布局

可以更加轻松地仪表盘制作移动布局。现在，Tableau 会自动为新仪表盘添加手机布局。你

可以通过编辑功能来自定义移动布局，也可以启用“自动布局模式”，让 Tableau 将你对默认布局所做的更改不断同步到手机布局中。

7. Google Adwords 连接器

可以从自己的 Google Adwords 数据获取更深入的见解。使用我们新推出的连接器在 Tableau 中连接和分析 Google 广告指标。

8. URL 动作方面的改进

现在，URL 动作可以将仪表板中的特定 Web 区域作为目标，从而让作者能够通过更强的控制功能，更加顺畅地链接到数据源之外的信息。改进集成效果并为仪表板添加更有意义的内容，设计更加精密的应用式仪表板。

9. 矢量地图

通过矢量背景地图，我们在 Tableau Online 中实现了更加顺畅的地图绘制体验。当通过平移、放大、缩小等操作来探索地理空间数据时，地图会比以往任何时候都更加清晰和顺畅。标签和图标会根据视图自动调整大小和形状，淡入和淡出非常顺畅。

10. 嵌套排序方面的改进

可以针对多个维度进行嵌套排序，同时还更新了排序对话框和工具栏排序按钮。

3.1.2 “开始” 页面

Tableau Desktop 工作簿文件与 Excel 工作簿十分类似，包含一个或多个工作表，可以是普通工作表、仪表板或故事。通过这些工作簿文件，可以对结果进行组织、保存和共享。打开 Tableau 时自动创建一个空白工作簿，也可以创建新工作簿，方法是选择“文件”→“新建”。

可以通过执行以下操作之一打开现有的工作簿：

- 单击开始页面上的工作簿缩略图图像。
- 选择“文件”→“打开”，使用“打开”对话框导航到该工作簿的位置。
- 双击 Windows 资源管理器中的任意工作簿文件。
- 将任意工作簿文件拖到 Tableau Desktop 图标或运行中的应用程序上。

Tableau Desktop 的开始页面由 3 个窗格组成，即“连接”“打开”和“探索”，可以从中连接数据、访问最近使用的工作簿以及浏览 Tableau 社区的内容，如图 3-2 所示。

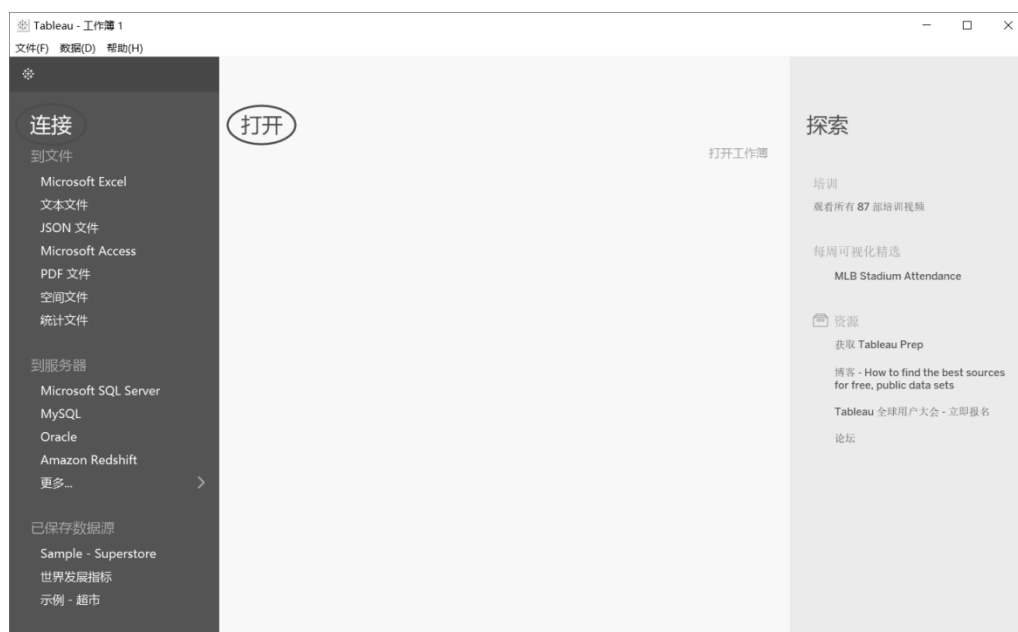


图 3-2 Tableau Desktop 开始页面

1. 连接

- 连接“到文件”：可以连接存储在 Microsoft Excel 文件、文本文件、Access 文件、Tableau 数据提取文件和统计文件等数据源。
- 连接“到服务器”：可以连接存储在数据库中的数据，如 Tableau Server、Microsoft SQL Server 或 Oracle 和 MySQL 等。
- 已保存数据源：快速打开之前保存到“我的 Tableau 存储库”目录的数据源，默认情况下显示一些已保存数据源的示例。

2. 打开

在“打开”窗格可以执行以下操作：

- 访问最近打开的工作簿：首次打开 Tableau Desktop 时，此窗格为空，随着创建和保存新工作簿，此处将显示最近打开的工作簿。
- 锁定工作簿：可通过单击工作簿缩略图左上角的锁定图标将工作簿锁定到开始页面。

3.1.3 “数据源”页面

在建立与数据的初始连接后，Tableau 将引导我们进入“数据源”页面，也可以通过在工作簿任意位置单击“显示开始页面”按钮返回开始页面，重新连接数据源，如图 3-3 所示。

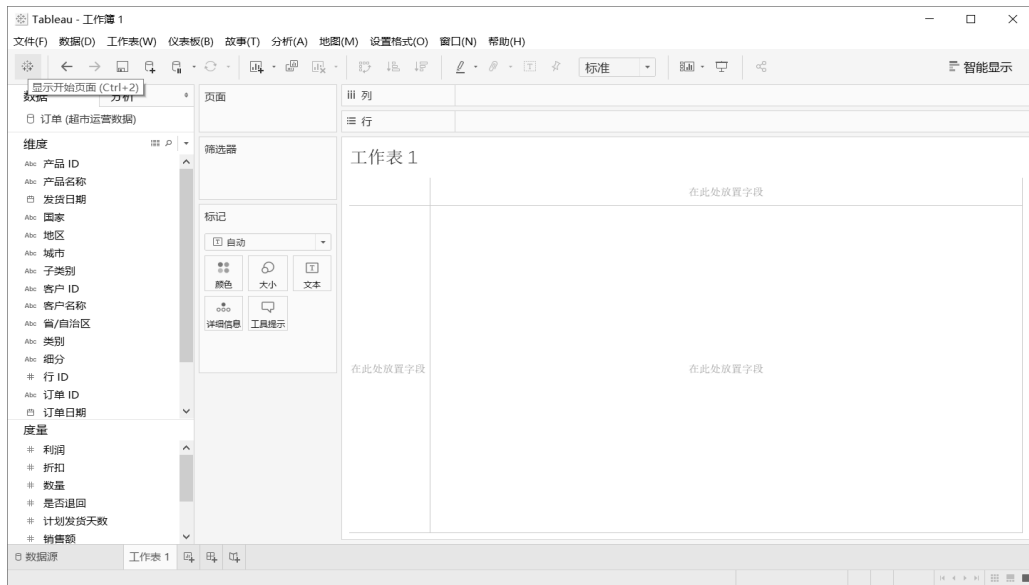


图 3-3 单击“显示起始页面”按钮

页面外观和可用选项会根据连接的数据类型而异。“数据源”页面通常由 3 个主要区域组成：左侧窗格、画布和网格，如图 3-4 所示。



图 3-4 “数据源”页面

1. 左侧窗格

“数据源”页面的左侧窗格显示有关 Tableau Desktop 连接数据的详细信息。对于基于文件的数据，左侧窗格可能显示文件名和文件中的工作表；对于关系数据，左侧窗格可能显示服务器、数据库或架构、数据库中的表。

2. 画布

连接大多数关系数据和基于文件的数据后,我们可以将一个或多个表拖到画布区域的顶部以设置 Tableau 数据源。当连接多维数据集数据后,“数据源”页面的顶部会显示可用的目录或要从中进行选择的查询和多维数据集。

3. 网格

通过使用网格我们可以查看数据源中的字段和前 1000 行数据,还可以使用网格对 Tableau 数据源进行一般的修改,如排序/隐藏字段、重命名字段/重置字段名称、创建计算、更改列/行排序或添加别名。

此外,根据连接的数据类型,单击“管理元数据”按钮以导航到元数据网格。元数据网格会将数据源中的字段显示为行,以便能够快速检查 Tableau 数据源的结构并执行日常管理任务,如重命名字段或一次性隐藏多个字段,如图 3-5 所示。



图 3-5 “元数据”网格

3.2 数据类型

Tableau 支持字符串、日期/日期时间、数字和布尔数据类型。这些数据类型会以正确的方式自动进行处理。如果创建自己的计算字段,就需要注意如何在公式中使用和组合不同的数据类型,如不能将字符串与数字相加。此外,许多在定义计算时可供使用的函数仅适用于特定数据类型,如 DATEPART()函数只能接受日期/日期时间数据类型作为参数。

3.2.1 主要的数据类型

数据源中的所有字段都具有一种数据类型。数据类型反映了该字段中存储信息的种类，如整数、日期和字符串。字段的数据类型在“数据”窗格中由图标标识。Tableau Desktop 主要数据类型的图标如图 3-6 所示。

下面介绍 Tableau 支持的数据类型。

1. 字符串 (STRING)

字符串是由零个或多个字符组成的序列。例如，“Wisconsin”“ID-44400”和“Tom Sawyer”都是字符串，字符串通过单引号或双引号进行识别。引号字符本身可以重复包含在字符串中，如“O”Hanrahan”。

2. 日期/日期时间 (DATE/DATETIME)

日期或日期时间，如“January 23,1972”或“January 23,1972 12:32:00 AM”。如果要将以长型格式编写的日期解释为日期/日期时间，就要在两端放置#符号。例如，“January 23,1972”被视为字符串数据类型，而#January 23,1972#被视为日期/日期时间数据类型。

3. 数值型

Tableau 中的数值可以为整数或浮点数。对于浮点数，聚合的结果可能并非总是完全符合预期。例如，可能发现 SUM 函数返回值为-1.42e-14，求和结果正好为 0，出现这种情况的原因是数字以二进制格式存储，有时会以极高的精度级别舍入。

4. 布尔型 (BOOLEAN)

包含 TRUE 或 FALSE 值的字段，当结果未知时会出现未知值。例如，表达式 $7 > \text{Null}$ 会生成未知值，自动转换为 Null。

此外，还有地理型，可以根据需要将省市字段转换为具有经纬度坐标的字段。

图标	说明
Abc	文本值
📅	日期值
🕒	日期和时间值
#	数值值
T/F	布尔值 (仅限关系数据源)
🌐	地理值 (用于地图)

图 3-6 Tableau Desktop 主要数据类型的图标

3.2.2 更改数据类型

在日常工作中，Tableau 可能会将字段标识为错误的数据类型。例如，可能会将包含日期的字段标识为整数而不是日期，可以在“数据源”页面上更改曾经作为原始数据源一部分的字段数据类型。

在“数据源”页面单击字段的字段类型图标，从下拉列表中选择一种新数据类型，如图 3-7 所示。



图 3-7 在“数据源”页面更改数据类型

如果使用数据提取，就要确保在创建数据提取之前已经进行所有必要的数据类型更改，否则数据可能不准确。例如，Tableau 把原始数据源中的浮点字段解释为整数，生成的浮点字段部分精度会被截断。

如果要在“数据”窗格中更改字段的数据类型，就要单击字段名称左侧的字段类型图标，然后从下拉列表中选择一种新数据类型，如图 3-8 所示。

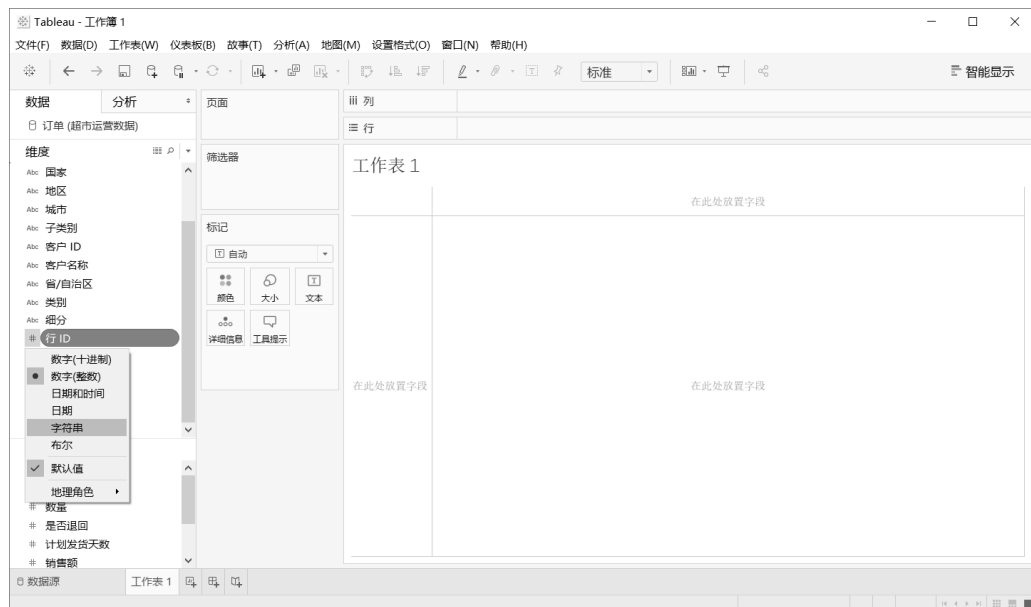


图 3-8 在“数据”窗格更改数据类型

若要在视图中更改字段的数据类型，则要在“数据”窗格中右击某个字段，选择“更改数据类型”，然后选择适当的数据类型，如图 3-9 所示。

此外，由于数据库中数据的精度比 Tableau 可以建模的精度高，因此将这些值添加到视图中时状态栏右侧将显示一个精度警告对话框。



图 3-9 在“数据”视图更改数据类型

3.3 文件类型

可以使用多种不同的 Tableau 专用文件类型保存工作，有工作簿、书签、打包数据文件、数据提取和数据连接文件。

3.3.1 工作簿文件

- 工作簿 (.twb)：Tableau 工作簿文件具有 .twb 文件扩展名，工作簿中含有一个或多个工作表，有零个或多个仪表板和故事。

3.3.2 打包工作簿

- 打包工作簿 (.twbx)：Tableau 打包工作簿具有 .twbx 文件扩展名。打包工作簿是一个 zip 文件，包含一个工作簿以及任何提供支持的本地文件数据源和背景图像，适合与不能访问该数据的其他人共享。

3.3.3 其他文件类型

- 数据提取 (.tde)：Tableau 数据提取文件具有 .tde 文件扩展名，提取文件是部分或整个数据源的一个本地副本，可用于共享数据、脱机工作和提高数据库性能。
- 数据源 (.tds)：Tableau 数据源文件具有 .tds 文件扩展名，是连接经常使用的数据源的快捷方式，不包含实际数据，只包含连接到数据源所必需的信息和在“数据”窗格中所做的修改。
- 打包数据源 (.tdsx)：Tableau 打包数据源文件具有 .tdsx 文件扩展名，是一个 zip 文件，包含数据源文件 (.tds) 和本地文件数据源，可使用此格式创建一个文件，以便与不能访问该数据的其他人共享。
- 书签 (.tbn)：Tableau 书签文件具有 .tbn 文件扩展名，书签包含单个工作表，是快速分享所做工作的简便方式。

3.4 运算符及优先级

运算符用于执行程序代码运算，会针对一个以上操作数项目进行运算的。例如， $2+3$ 的操作数是 2 和 3，运算符是“+”。Tableau 支持的运算符有算术运算符、逻辑运算符、比较运算符。

3.4.1 算术运算符

- + (加法)：此运算符应用于数字时表示相加；应用于字符串时表示串联；应用于日期时，将天数与日期相加。例如，`'abc'+def='abcdef'`；`#April 15,2004#+15=#April 30,2004#`。
- - (减法)：此运算符应用于数字时表示减法；应用于表达式时表示求反；应用于日期时，可用于从日期中减去天数，还可用于计算两个日期之间的天数差异。例如，`7-3=4`；`-(7+3)=-10`；`#April 15,2004#-#April 8,2004#=7`。
- * (乘法)：此运算符表示数字乘法。例如，`5*4=20`。
- / (除法)：此运算符表示数字除法。例如，`20/4=5`。
- % (求余)：此运算符表示数字余数。例如，`5%4=1`。
- ^ (乘方)：此符号等效于 POWER 函数，用于计算数字的指定次幂。例如，`6^3=216`。

3.4.2 逻辑运算符

- AND：逻辑运算且，两侧必须使用表达式或布尔值。

例如，`IIF(Profit=100 AND Sales=1000,"High","Low")`，如果两个表达式都为 TRUE，结果就为 TRUE；如果任意一个表达式为 UNKNOWN，结果就为 UNKNOWN；其他情况结果都为 FALSE。

- OR: 逻辑运算或，两侧必须使用表达式或布尔值。

例如，`IIF(Profit=100 OR Sales=1000,"High","Low")`，如果任意一个表达式为 TRUE，结果就为 TRUE；如果两个表达式都为 FALSE，结果就为 FALSE；如果两个表达式都为 UNKNOWN，结果就为 UNKNOWN。

- NOT: 逻辑运算符否，可用于对另一个布尔值或表达式求反。

例如，`IIF(NOT(Sales=Profit),"Not Equal","Equal")`。

3.4.3 比较运算符

Tableau 有丰富的比较运算符，有 == 或 =（等于）、>（大于）、<（小于）、>=（大于或等于）、<=（小于或等于）、!= 和 <>（不等于），用于比较两个数字、日期或字符串，并返回布尔值（TRUE 或 FALSE）。

3.4.4 运算符优先级

所有运算符都按特定顺序计算，如 $2*1+2$ 等于 4 而不等于 6，因为 * 运算符始终在 + 运算符之前计算。表 3-1 显示了计算运算符的顺序，第一行具有最高优先级，同一行中的运算符具有相同优先级，如果两个运算符具有相同优先级，在公式中就从左向右进行计算。

表 3-1 运算符优先级

优先级	运算符	优先级	运算符
1	-（求反）	5	==、>、<、>=、<=、!=
2	^（乘方）	6	NOT
3	*/、%	7	AND
4	+、-	8	OR

可以根据需要使用括号，括号中的运算符在计算时优先于括号外的运算符，从内部的括号开始向外计算，如 $(1+(2*2+1)*(3*6/3))=31$ 。

3.5 Tableau 软件概况

Tableau 公司成立于 2003 年，是由斯坦福大学的三位校友 Christian Chabot（首席执行官）、Chris Stole（开发总监）以及 Pat Hanrahan（首席科学家）在远离硅谷的西雅图注册成立的。其中，Chris Stole 是计算机博士，Pat Hanrahan 是皮克斯动画工作室的创始成员之一，曾负责视觉特效渲染软件的开发，两度获得奥斯卡最佳科学技术奖，至今仍在斯坦福担任教授职位，教授计算机图形课程。

Tableau 公司主要面向企业数据提供可视化服务，是一家商业智能软件提供商。企业运用 Tableau 授权的数据可视化软件对数据进行处理和展示，不过 Tableau 的产品并不局限于企业，其他机构甚至个人都能很好地运用 Tableau 软件进行数据分析工作。数据可视化是数据分析的完美结果，能够让枯燥的数据以简单友好的图表形式展现出来。可以说，Tableau 在抢占细分市场，也就是大数据处理末端的可视化市场，目前市场上并没有太多这样的产品。同时，Tableau 还为客户提供解决方案服务。

3.5.1 Tableau Desktop

“所有人都能学会的业务分析工具”，这是 Tableau 官方网站上对 Tableau Desktop 的描述。确实，Tableau Desktop 的简单、易用程度令人发指，这也是软件的最大特点。使用者不需要精通复杂的编程和统计原理，只需要把数据直接拖放到工作簿中，通过一些简单的设置就可以得到想要的可视化图形。

Tableau Desktop 的学习成本很低，使用者可以快速上手，这无疑对日渐追求高效率和本成本控制的企业来说具有巨大吸引力，特别适合日常工作中需要绘制大量报表、经常进行数据分析或需要制作图表的人使用。简单、易用并没有妨碍 Tableau Desktop 拥有强大的性能，它不仅能完成基本的统计预测和趋势预测，还能实现数据源的动态更新。

Tableau Desktop 不同于 SPSS，SPSS 作为统计分析软件，比较偏重于统计分析，使用者需要有一定数理统计基础，虽然功能强大且操作简单、友好，但输出的图表与办公软件的兼容性及交互方面有所欠缺。Tableau Desktop 是一款完全的数据可视化软件，专注于结构化数据的快速可视化，使用者可以快速进行数据可视化并构建交互界面，用来辅助人们进行视觉化思考，并没有 SPSS 强大的统计分析功能。

总之，快速、易用、可视化是 Tableau Desktop 最大的特点，能够满足大多数企业、政府机构数据分析和展示的需要，以及部分大学、研究机构可视化项目的要求，而且特别适合企业使用，毕竟 Tableau 自己的定位是业务分析和商业智能。在简单、易用的同时，Tableau Desktop 极其高效，数据引擎的速度极快，处理上亿行数据只需几秒就可以得到结果，用其绘制报表的速度也比程序员制作传统报表快 10 倍以上。

Tableau Desktop 还具有完美的数据整合能力，可以将两个数据源整合在同一层，甚至可以将一个数据源筛选为另一个数据源，并在数据源中突出显示，这种强大的数据整合能力具有很大的实用性。Tableau Desktop 还有一项独具特色的数据可视化技术——嵌入地图，使用者可以用经过自动地理编码的地图呈现数据，这对于企业进行产品市场定位、制定营销策略等有非常大的帮助。

3.5.2 Tableau Prep

2018 年 4 月份，Tableau 推出全新的数据准备产品——Tableau Prep，官方下载地址为 <https://www.Tableau.com/products/prep>。定位于如何帮助人们以快速可靠的方式对数据进行合并、

组织和清理，进一步缩短从数据获取见解所需的时间。简而言之，Prep 是一款简单易用的数据处理工具（部分 ETL 工作）。

之所以使用 Tableau Prep，是因为我们在使用 BI 工具进行数据可视化展示时数据常常不具有适合分析的形制（数据模型），很难应对复杂的数据准备工作。因此，我们需要一种更方便的工具来搭建我们需要的数据模型。

Tableau Desktop 的 key 无法直接激活 Prep，根据官方介绍，需要进入 Tableau client portal，使用 administrator 账户登录，即可看到已经购买的 Desktop Key，以及 1 个 Prep 的 Key，此 Key 可以激活与 Desktop Key 关联电脑上的 Prep。

Prep 保持了与 Tableau Desktop 一致的蓝色基调 UI，默认英语，未支持多语言选择。界面分为 3 部分，左边第一部分进行数据链接，中间是最近使用过的操作流程及预设的展示操作流程（还是大家熟悉的超市数据），右侧是一些教学资源。

3.5.3 Tableau Online

Tableau Online 是 Tableau Server 的软件即服务托管版本，让商业分析比以往更加快速轻松。可以利用 Tableau Desktop 发布仪表板，然后与同事、合作伙伴或客户共享，利用云智能随时随地、快速找到答案。

利用 Tableau Online 可以省去硬件与安装时间。利用 Web 浏览器或移动设备中的实时交互式仪表板可以让公司上下每一个人都成为分析高手，在仪表板上批注、分享发现。可以订阅和获得定期更新，这一切都在敏捷安全的软件即服务 Web 平台上完成。可以从几个用户着手，随后在需要时按需添加。

利用云智能可以在世界任意地点发现数据背后的真相。无论在办公室、家里还是在途中，均可查看仪表板，进行数据筛选、下钻查询或将全新数据添加到分析工作中；可以在现有报表未能预计的方面获得对这些问题的新见解；还可以在 Web 上编辑现有视图，利用 Tableau 飞速数据引擎完成这一切，让问题随问随答。

Tableau Online 可连接云端数据和办公室内的数据。Tableau Online 还与 Amazon Redshift、Google BigQuery 保持实时连接，同时可连接其他托管在云端的数据源（如 Salesforce 和 Google Analytics）并按计划安排刷新，或从公司内部向 Tableau Online 推送数据，让团队轻松访问，按设定的计划刷新数据，在数据连接发生故障时获得警报。

3.5.4 Tableau Server

Tableau Server 是一种新型的商业智能，传统的商业智能系统往往很笨重、复杂，需要运用专业人员和资源进行操作和维护，一般由企业专门设立的 IT 部门进行维护，不过 IT 技术人员通常缺乏企业其他人员的商业背景，这种鸿沟导致对系统利用的低效率和时间滞后。

Tableau Server 非常简单、易用，一般人都能学会，是一种真正自助式的商业智能，速度比传

统商业智能快 100 倍。更重要的是，Tableau Server 是一种基于 Web 浏览器的分析工具，是可移动式的商业智能，用 iPad、Android 平板也可以进行浏览和操作，而且 Tableau 的 iPad 和 Android 应用程序都已经过触摸优化处理，操作起来非常容易。

Tableau Server 的工作原理是，由企业服务器安装 Tableau Server，并由管理员进行管理，将需要访问 Tableau Server 的人作为用户添加（无论是进行发布、浏览还是管理）。Tableau Server 还必须为用户分配许可级别，不同许可级别具有不同的权限，为自定义视图并与其进行交互的用户提供 Interactor 许可证，为只能查看与监视视图的用户提供 Viewer 许可证。

被许可的用户可以将自己在 Tableau Desktop（只支持专业版）中完成的数据可视化内容、报告与工作簿发布到 Tableau Server 中与同事共享。同事可以查看你共享的数据并进行交互，通过共享的数据源以极快的速度进行工作。这种方式可以更好地管理数据的安全性，如用户通过 Tableau Server 可以安全地共享报告，不再需要通过电子邮件发送带有敏感数据的电子表格。

值得一提的是，在全球最大的商业智能用户调查中，Tableau 在客户忠诚度、实施速度、最低实施成本和拥有成本方面都排名第一，击败了包括 IBM、甲骨文、微软、SAS 在内的众多 BI 供应商。

3.5.5 Tableau Public

Tableau Public 是 Tableau 的免费版本，适合所有想要在 Web 上讲述交互式数据故事的人。作为服务交付，Tableau Public 可以立即启动并运行。Tableau Public 可以连接到数据、创建交互式数据可视化内容，并将其直接发布到自己的网站，通过所发现的数据内在含义引导读者，让他们与数据互动，发掘新的见解，这一切不用编写代码即可实现。

3.5.6 Tableau Mobile

Tableau Mobile 可以帮助用户随时掌握数据，需要搭配 Tableau Online 或 Tableau Server 账户才能使用，可以通过 [Tableau.com/zh-cn/products/trial](https://tableau.com/zh-cn/products/trial) 下载免费试用版。

Tableau Mobile 可以快速流畅地查看数据，提供快捷、轻松的数据处理途径。从提出问题到取得见解只需要几次轻触。

Tableau Mobile 的主要功能如下：

- 随处编写和查看，编写一次仪表板可以在任何设备上查看。
- 脱机快照，即使在脱机状态下也能够以高分辨率图像形式使用，仅限 iPad。
- 订阅，在需要将重要信息发送至收件箱，立刻向 Tableau Mobile 订阅工作簿。
- 灵活，Tableau Mobile 提供适用于 iPad、Android 和移动浏览器的应用。
- 内容安全性，内容加密保存在设备上，并且安全连接 Tableau Online 和 Server。
- 共享，与团队轻松协作，轻触屏幕即可通过电子邮件发送分析结果或数据。

3.5.7 Tableau Reader

Tableau Reader 是一款免费桌面应用程序，用来与 Tableau Desktop 中生成的可视化数据进行交互。利用 Tableau Reader 可以筛选、向下钻取和查看数据明细，一直详细到用户允许的程度。

3.6 练习题

1. 到官方网站下载和安装 Tableau Desktop 最新版本的软件。
2. 打开软件的开始页面，尝试导入一个 Excel 文件并核查。
3. 将 Tableau 软件的运算符和优先级与其他软件进行比较。