

# 数据分析方法论

确定分析思路需要以营销、管理等理论为指导。这些跟数据分析相关的营销、管理等理论统称为数据分析方法论。可以把方法论理解为指南针，在分析方法论的指导下我们才去开展数据分析，这样分析的结果才具有指导意义，而不会出现南辕北辙的情况。

## 1. 数据分析方法论与数据分析区别

数据分析方法论主要用来指导数据分析师进行一次完整的数据分析，它更多的是指数据分析思路，比如主要从哪几方面开展数据分析，各方面包含什么内容和指标？数据分析方法论主要从宏观角度指导如何进行数据分析，它就像是一个数据分析的前期规划，指导着后期数据分析工作的开展。

而数据分析法则是指具体的分析方法，例如我们常见的对比分析、交叉分析、相关分析、回归分析、聚类分析等数据分析法。数据分析法主要从微观角度指导如何进行数据分析。

比如下表我们以服装制作为例子，对比理解数据分析方法论：

	数据分析	服装制作
方法论	5W2H	服装设计图
工具	Excel、SPSS 等	剪刀、缝纫机、电熨斗等
技术	交叉分析、回归分析、聚类分析、相关分析等	平面、立体裁剪等，合缝、包缝、骑缝等

图表：数据分析与服装制作对比图

## 2. 数据分析方法论的重要性

很多人在做数据分析时，经常遇到这几个问题：不知从哪方面入手开展分析；分析的内容和指标常常被质疑是否合理、完整，而自己也说不出个所以然来。对这些问题常常感到困扰。

数据分析方法论主要有以下几个作用：



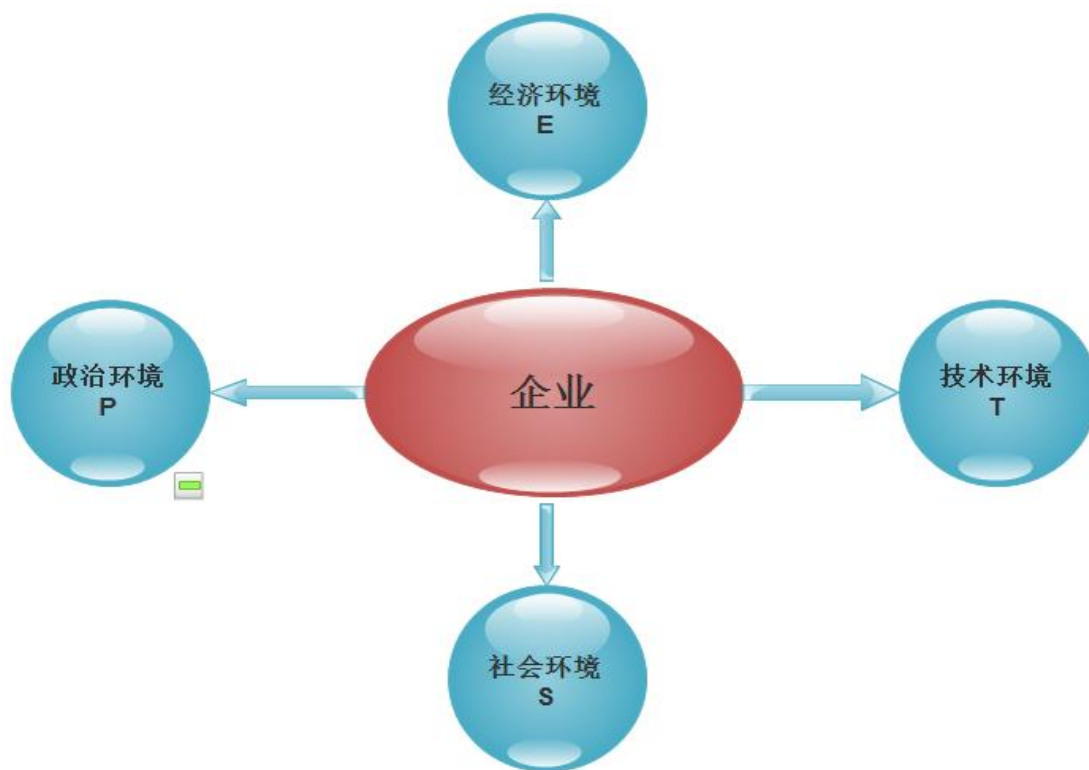
- 理顺分析思路，确保数据分析结构体系化。
- 把问题分解成相关联的部分，并显示它们之间的关系。
- 为后续数据分析的开展指引方向。
- 确保分析结果的有效性生及正确性。

如果没有数据分析方法论的指导，整个数据分析报告虽然各方面都涵盖到，但会给人感觉还缺点什么。其实就是报告主线不明，各部分的分析逻辑不清。

## 3. 常用的数据分析方法论

### 3.1. PEST 分析法

PEST 分析法用于对宏观环境的分析。**宏观环境又称一般环境，是指影响一切行业和各种企业的各种宏观力量。**对宏观环境因素作分析时。由于不同行业和企业有其自身特点和经营需要，分析的具体内容会有差异，但一般都应对**政治(Political)**、**经济(Economic)**、**技术(Technological)**和**社会(Social)**这四大类影响企业的主要外部环境因素进行分析，这种方法简称为 **PEST 分析法**。



- **政治环境：**构成政治环境的关键指标有:政治体制、经济体制、财政政策、税收政策、产业政策、投资政策、专利数量、国防开支水平、政府补贴水平、民众对政治的参与度等。
- **社会环境：**构成社会文化环境的关键指标有:人口规模、性别比例、年龄结构、出生率、死亡率、种族结构、妇女生育率、生活力式、购买习惯、教育状况、城市特点、宗教信仰状况等因素。
- **技术环境：**构成技术环境的关键指标有:新技术的发明和进展、折旧和报废速度、技术更新速度、技术传播速度、技术商品化速度、国家重点支持项目、

国家投入的研发费用、专利个数、专利保护情况等因素。

- **经济环境:** 构成经济环境的关键指标有:GDP 及增长率、进出口总额及增长率、利率、汇率、通货膨胀率、消费价格指数、居民可支配收入、失业率、劳动生产率等。

### 3.2. 5W2H 分析法

5w2H 分析法是以五个 w 开头的英语单词和两个 H 开头的英语单词进行提问，从回答中发现解决问题的线索，即**何因 (Why)**、**何事(What)**、**何人(Who)**、**何时(When)**、**何地(Where)**、**如何做(How)**、**何价(How much)**，这就构成了 5W2H 分析法的总框架。

该方法简单、方便，易于理解和使用，富有启发意义，广泛用于企业营销、管理活动，对于决策和执行性的活动措施非常有帮助，也有助于弥补考虑问题的疏漏。其实对任何事情都可以从这七大方面去思考，对于不善分析句题的人，只要多练习即可上手，所以同样它也适用于指导建立数据分析框架。

现在以用户购买行为分析为例，来学习 5W2H 分析法。



### 3.3. 逻辑树分析法

逻辑树又称问题树、演绎树或分解树等。它是将问题的所有问题分层罗列，从最高层开始，并逐步向下扩展。

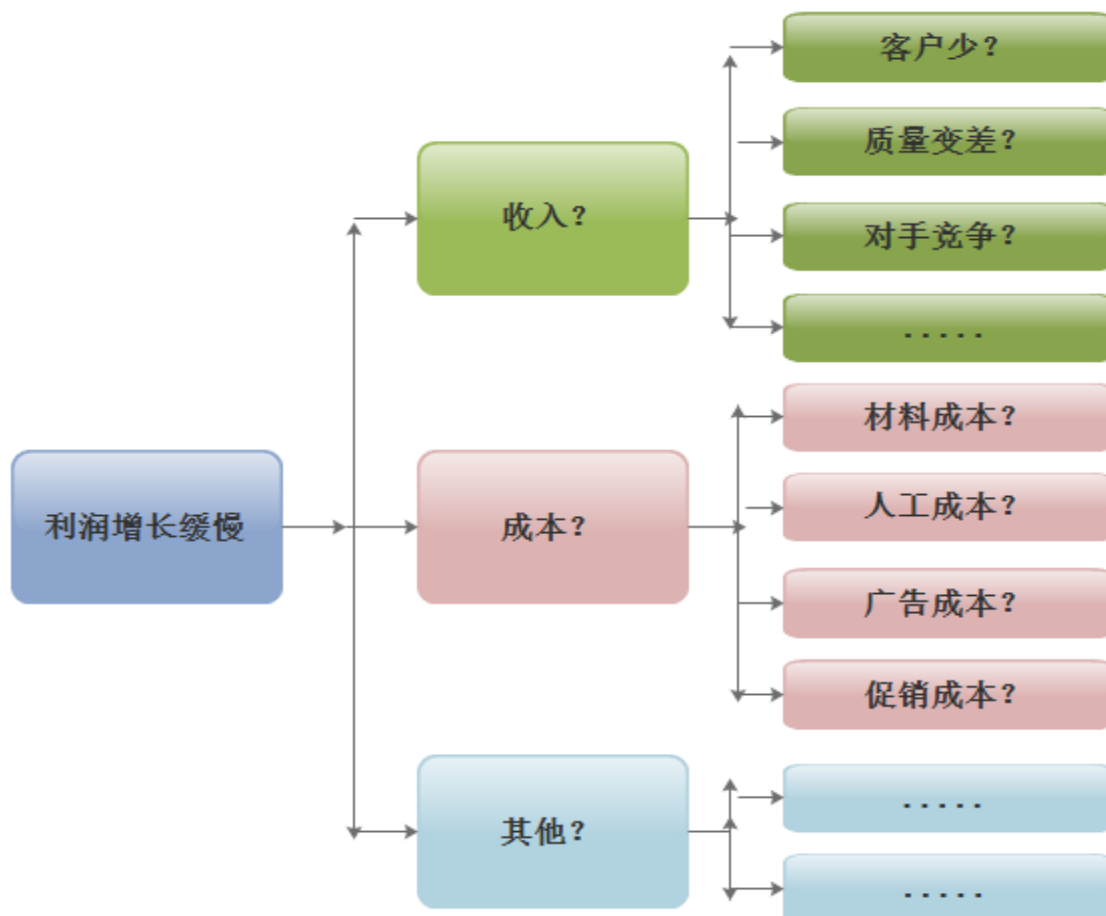
把一个已知问题当成树干，然后开始考虑这个问题和哪些相关问题有关。每想到一点，就给这个问题所在的树干加一个“树枝”，并标明这个“树枝”代表什么问题。

逻辑树能保证解决问题的过程的完整性，它能把工作细分为便于操作的任务，确定各部分的优先顺序，明确地把责任落实到个人。

逻辑树的使用必须遵循以下三个原则。

- **要素化**：把相同的问题总结归纳成要素。
- **框架化**：将各个要素组织成框架。遵守不重不漏的原则。
- **关联化**：框架内的各要素保持必要的相互关系，简单而不独立。

在使用逻辑树的时候，尽量把涉及的问题或要素考虑周全。



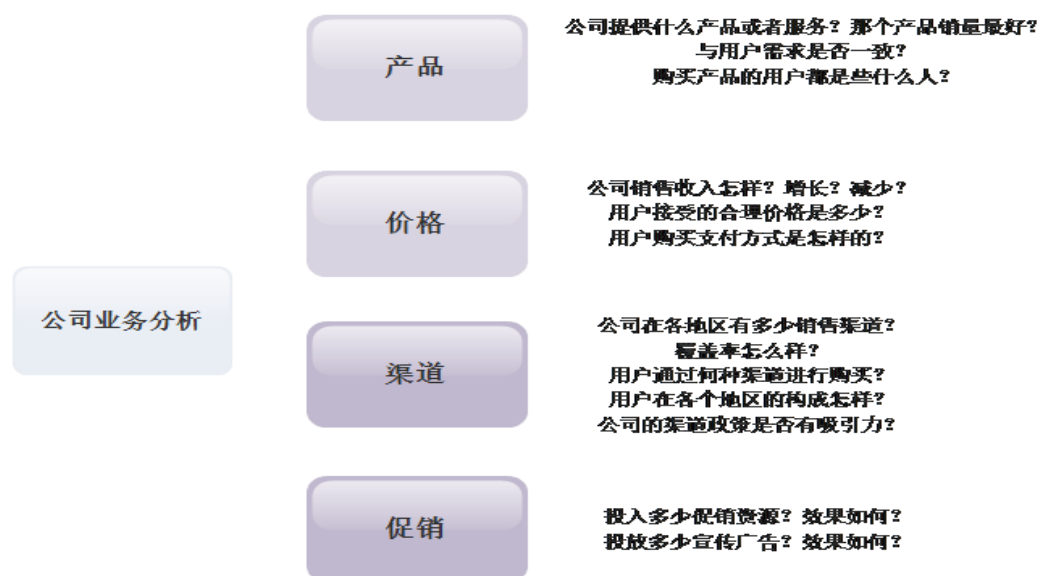
图：逻辑树分析法在利润分析中的应用

### 3.4. 4P 营销理论

**4P 营销理论**产生于 20 世纪 60 年代的美国，它是随着营销组合理论的提出而出现的。营销组合实际上有几十个要素，这些要素可以概括为 4 类：**产品(Product)、价格(Price)、渠道(Place)、促销(Promotion)**。

- **产品**：从市场营销的角度来看，产品是指能够提供给市场，被人们使用和消费并满足人们某种需要的任何东西，包括有形产品、服务、人员、组织、观念或它们的组合。
- **价格**：是指顾客购买产品时的价格，包括基本价格、折扣价格、支付期限等。影响定价的主要因素有三个：需求、成本与竞争。
- **渠道**：是指产品从生产企业流转 to 用户手上全过程中所经历的各个环节。
- **促销**：是指企业通过销售行为的改变来刺激用户消费，以短期的行为（比如让利、买一送一，营销现场气氛等等）促成消费的增长，吸引其他品牌的用户或导致提前消费来促进销售的增长。广告、宣传推广、人员推销、销售促进是一个机构促销组合的四大要素。

接下来，我们就用 4P 营销理论对公司业务运营的整体情况做分析。在大家实际工作中，需要根据实际业务情况进行调整，灵活运用，切忌生搬硬套。只有深刻理解公司业务的同时才能较好地进行业务方面的数据分析，否则将脱离业务实际，得出无指导意义的结论，犹如纸上谈兵，甚至贻笑大方。



### 3.5. 用户行为理论

网站分析的发展已经较为成熟，有一套成熟的分析指标。比如 IP、PV、页面停留时间、跳出率、回访者、新访问者、回访次数、回访相隔天数、流失率、关键字搜索、转化率、登录率，等等。遇到这么多指标，所有的指标都要采用吗？什么指标该采用？什么指标又不该采用，各指标之间有何联系，哪个指标先分析？哪个指标后分析？

这么多指标，对于我们来说下手还是懵的，所以需要梳理它们之间的逻辑关系，比如利用用户使用行为理论进行梳理。

用户使用行为是指用户为获取、使用物品或服务所采取的各种行动，用户对产品首先需要有一个认知、熟悉的过程，然后试用，再决定是否继续消费使用，最后成为忠诚用户。

现在我们可利用用户使用行为理论，梳理网站分析的各关键指标之间的逻辑关系，构建符合公司实际业务的网站分析指标体系：



图：用户使用行为理论在网站分析中的应用