

面对数据异常，我们经常会出现“好像是 A 原因引起的？”“貌似和 B 原因也相关？”“有可能是 C 操作不当”的主观臆测。

或者，拿到一个分析议题，分析“11 月销售数据下降的原因”，是先从产品层面，还是渠道层面着手的茫然无措。

显然，这样的思维是乱的。

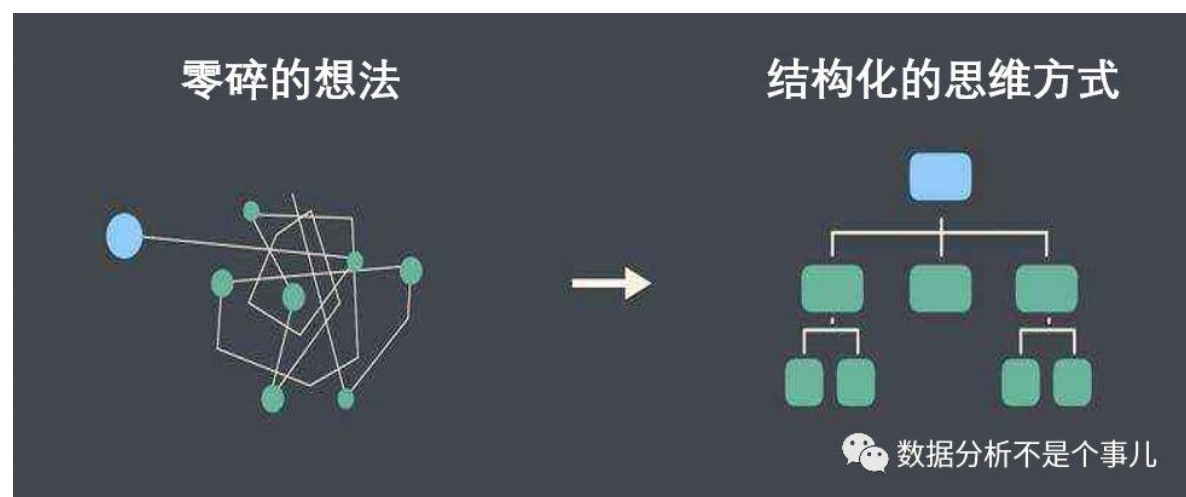
做数据分析，首先你得具备看待一个事物的逻辑化思维，其次用数据去证明他。

我们会经常听说两种推理模式，一种是归纳，一种是演绎。这也是麦肯锡思维当中很经典的两个方法，工作中所有的问题，都可以用归纳或者演绎的形式进行拆分，我喜欢把这个过程称为“解构”。

这两种思维模式能够帮助数据分析师完成原始的业务逻辑积累，在此基础上快速定位业务问题，提升分析效率。

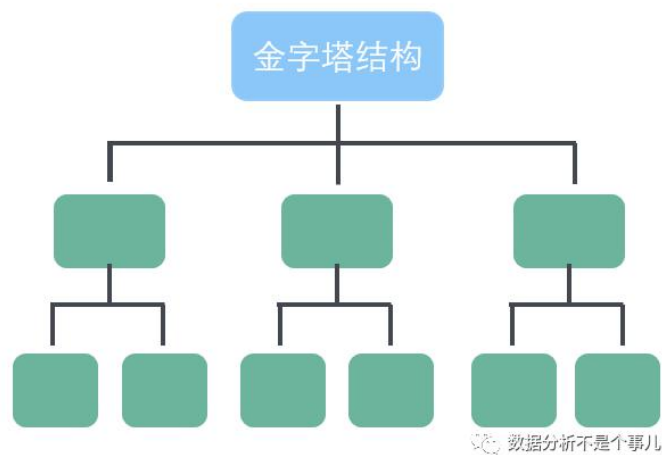
一、结构化思维

归纳其实就是把复杂问题分解成多种单一因素的过程，并且将这些因素加以归纳和整理，使之条理化、纲领化。这个过程犹如抽丝剥茧，将一团乱麻理地条条顺顺。



如何练习结构化思维，这其中会运用一个很重要工具，那就是金字塔模型。

根据《金字塔原理》，“任何事情都可以归纳出中心论点，由中心论点出发，可由三至七个论据支撑，每个一级论点可以衍生出其他的分论点。”如此发散开来，就可以形成以下的金字塔结构思考方式。



但是在你还没有掌握这种结构化思维方式时，直接用这种思考方式是有一定难度的。这时候就可以采用金字塔原理中的 MECE 法则去思考结构。具体的操作方式是：

- A. 尽可能列出所有思考的要点
- B. 找出关系，进行分类。

他的原则是论点之间相互独立，不重叠；论据穷尽划分，不遗漏。

举个例子：

现在有一个线下销售的产品。我们发现 8 月的销售额度下降，和去年同比下降了 20%。我想先观察时间趋势下的波动，看是突然暴跌还是逐渐下降。再按照不同地区的数据看一下差异，有没有地区性的因素影响。我也准备问几个销售员，看一下现在的市场环境怎么样，听说有几家竞争对手也缩水了，是不是这个原因。

用结构化思维梳理，就是：

用这种方式思考，能确保思考的点成体系，逻辑严谨，要素相互之间不凌乱不打架，思考的点都穷尽。长期练习这种方法，不仅更容易找到逻辑结构，也更容易培养你的结构化思维。

具体，可以阅读书籍：《金字塔思维》

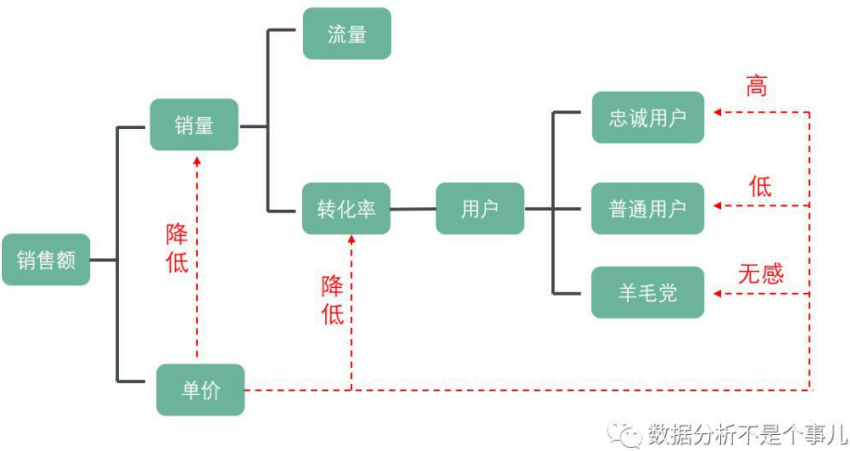
二、假说演绎思维

以情况为起点的推理方法是归纳推理，以规则为起点的推理方法可以称之为演绎推理。

比如：

某自营电商网站，现在想将商品提价，让你分析下销售额会有怎样的变化？

首先可以确定销量会下降，那么下降多少？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的情况，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，问题是销量下跌多少？

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。

找出平时的转化率（譬如为 20%），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，那么将用户划分忠诚 XX、普通 XX、羊毛 XX.....

不同用户层次数量不同，反应不同。忠诚用户转化率变化极低，羊毛几乎不会转化....这些数据可以凭借经验做出假设，最后汇总。

具体的变化情况都可以根据过往的数据来拟合，统计学上也有一些科学的预测模型，后面讲数理统计知识时会有涉及。

假设先行就是以假设作为思考的起点，先提出问题，然后用 MECE 原则梳理关联因素间的结构关系。

小结

归纳和演绎的思维是数据分析初期必备的，面试考察逻辑思维也无非是这两点。实际情况中可针对不同的项目要求进行组合应用。在经过一定阶段的训练后，可以帮助提升业务熟悉程度，完成业务的初始积累后，后续的分析过程中就可以逐步减少拓展推理的层级及组合，逐步提升问题原因定位的效率。

三、指标化思维

上述的分析思维，帮助我们去定性问题，接下来我们要介入数据的方式，去定量分析。首要掌握指标化的思维。

假设有一家电商公司，我们想要了解网站运营的情况如何？运营人员向我们描述，我们的网站的流量很高啊，比淘宝差一点，比京东好一点，每天都有大量的新用户，老用户下单也很活跃啊。

我就疑惑了，流量高是多少？大量的新用户怎么衡量？一个手机注册了算新用户还是新下单的用户？下单活跃又是怎么个活跃法？

这样的问题相信只能凭运营人员的经验来判断，而经验带来的“后果”往往是拍脑袋式的决策。

如果用指标化的思维，应该用 PV 和 UV 去衡量流量，新用户下单数和占比去评价网站的拉新，新老买家占比等指标去衡量用户活跃。

很明显，指标就是用来定义、评价和衡量业务的一个标准。比如网站相关用户访问量、停留时长、跳出率等。销售相关销售量、销售额、客单价等。应该很好理解。

指标的设定有两个经验：

A.“有总比没有强”。对于要监控的事物，能有指标的尽量要有指标。

B.“一个好的指标应该是用来衡量具体且可量化的事物”。比如，用户访问量、停留时长、跳出率等。

指标体系

有指标是否就够了呢？指标按照结构化思维可以形成一个体系，如销售分析指标体系，生产指标体系，电商行业指标体系。

一家企业建立的数据分析体系通常细分到了具体可执行的部分，可以根据设定的某个指标异常变化，相应立即执行相应的方案，来保证运营的正常进行。

附上一张电商行业的指标体系，各运营模块的指标体系网上一搜一大把，可以参考着建立，后面我也会在我的文章里阐述。

建立指标体系的思路：

向上

可以按业务职能结构划分，映射出更多维度，比如渠道，运营，产品等相关模块，将相关指标映射到主要模块，通过简单快速的沟通，快速定位问题原因。

向下

可以按因果结构划分，也就是指标分解，利用公式的方法。比如营收=日活*付费率*arpu 等指标因果关系进行划分，通过定位指标波动，定位最细指标，辅助维度下转，能够清楚的问题原因。

就像枝丫一样，从主干不断延伸，将业务用指标评价量化，逐渐形成一个健全的数据分析体系。

四、维度分析思维

最后，站在分析的角度讲一下维度思维。

当你有了指标，可以着手进行分析，数据分析大体可以分三类，第一类是利用维度分析数据，第二类是使用统计学知识如数据分布假设检验，最后一类是使用机器学习。这里我们主要了解维度分析法。

维度是观察数据的角度，例如“时间”、“地区”、“产品”。在具体分析中，我们可以把它认为是分析事物的角度。时间是一种角度、地区是一种角度，产品也是一种角度，所以它们都能算维度。

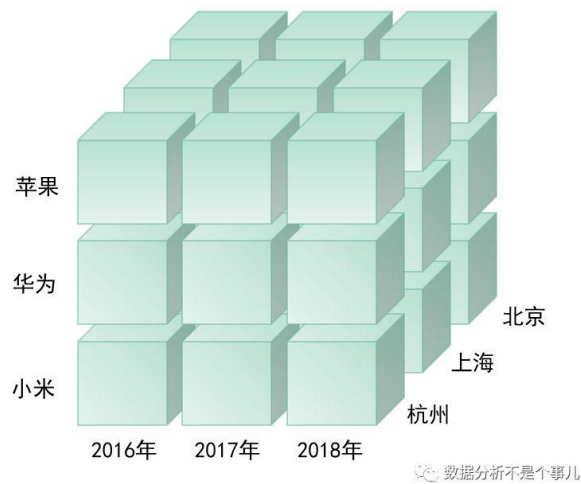
当我们有了维度后，就能够通过不同的维度组合，形成数据模型。数据模型不是一个高深的概念，它就是一个多维立方体。

这个概念最早来源于商业智能 OLAP 技术。数据按照事实表 (Fact Table) 和维表 (Dimension Table) 的形式存在。事实表用来记录具体事件，比如销量、销售额、售价、折扣等具体的数值信息。维度表是对事实表中事件的要素的描述信息，比如时间、城市、品牌、机型等。

这是一个最简单的星形模型的实例。

事实表里面主要包含两方面的信息：维和度量。维的具体描述信息记录在维表，事实表中的维属性只是一个关联到维表的键，并不记录具体信息；度量一般都会记录事件的相应数值，比如这里的产品的销售数量、销售额等。维表中的信息一般是可以分层的，比如时间维的年月日、地域维的省市县等，这类分层的信息就是为了满足事实表中的度量可以在不同的粒度上完成聚合，比如 2016 年商品的销售额，来自上海市的销售额等。

下图举例一个简化的分析模型，分别由产品、城市、时间这三个维度组成，实际数据分析中，维度远不止三个。



在数据库中，可能是这样两张表：

品牌	机型	容量	产品型号	上市日期	售价
华为	mate20	8G+128G	HWM23	2018年10月	4499
华为	mate20pro	8G+128G	HWMP22	2018年10月	6399
苹果	X	128G	APPX02	2017年9月	6399
苹果	XS	256G	APPXS3	2018年9月	10099
苹果	XS MAX	256G	APPXSP3	2018年9月	10999
小米	8	6G+64G	XM8063	2018年5月	2299
.....

订单号	产品型号	用户名	手机号	下单时间	实付价	折扣	国家	省/直辖市	城市/区	地址
10023791	HWMP22	wuliashe	13250507893	2018年11月12	6199	0.885	中国	北京市	西城区	*****
10023792	HWMP22	kikie	17898454546	2018年11月12	6099	0.885	中国	江苏省	南京市	*****
10023793	APPXS3	mommy	15750508868	2018年11月12	9999	0.9	中国	上海市	虹口区	*****
.....

我们可以将品牌作为维度，分析手机的销量情况，也可以将时间作为维度，分析每一年手机市场的份额情况。

多维分析操作包括：钻取（Drill-down）、上卷（Roll-up）、切片（Slice）、切块（Dice）以及旋转（Pivot）。

钻取（Drill-down）：在维的不同层次间的变化，从上层降到下一层，或者说是将汇总数据拆分到更细节的数据，比如通过对 2018 年华为的总销售数据进行钻取来查看各个手机型号的销售数据。

上卷（Roll-up）：钻取的逆操作，即从细粒度数据向高层的聚合。如将江苏省、上海市和浙江省的销售数据进行汇总来查看江浙沪地区的销售数据。

切片（Slice）：选择维中特定的值进行分析，比如只选择苹果手机的销售数据，或 2017 年的手机销售数据。

切块（Dice）：选择维中特定区间的数据进行分析，比如选择 2016 年 2017 年的销售数据。

旋转 (Pivot)：即维的位置的互换，就像是二维表的行列转换，如图中通过旋转实现产品维和地域维的互换。

为什么这边花那么多笔墨去讲维度和度量呢，一者是在梳理分析思路时，常常会按照几个大的维度类去划分层级，多面分析，如时间维、地域维、产品维，帮助我们成为“多面分析手”。另一方面，BI 商业智能在操作也基于维度一说，熟悉维度和数据模型的原理，能更好的理解这个工具。（没错，这玩意儿后面要学）

好了，花了一天的时间整理了数据分析的思维，大家慢慢消化。

为了更好的巩固，推荐大家去读《金字塔原理》，可以用思维导图的方式做好知识点整理，工具推荐 Xmind 和幕布。

不算作业不做监督，但如果你后台发我的话，我看到都会反馈。