### 数据分析一方面自己要会将数据制作成图形便于他人解读，另外一方面，我们自己也必须学会解读别人的图形，或者解读自己过去曾经制作的图形。每个数据图形背后都是数据，而不同水平和能力的人在同样图形上看到的东西可能会不同。

解读数据会有很多方法，我们这里给大家提示性介绍十个方法。

上期内容：①辅助线 ②找差异、找变化 ③找奇异点、特殊群体 ④找转折点、拐点 ⑤找特征

本期内容：⑥找问题 ⑦找源头**Root Cause** ⑧找关系 ⑨找驱动 ⑩找规律

**6、找问题**

对数据分析人员来讲找问题就是找数据，只要问对了问题，就能够找到对的合适的数据，通过数据可以分析出具体的问题，并能够找到具体的方案。在一定意义上来讲，问对问题比找到答案更加重要。只要问题问对了，答案基本上非常容易获得。有人甚至夸张地说，问对问题，事情就解决了一半。

我作为资深的数据分析师，我常用以下五个问题来确保我们的分析能够深入到最深层次，能够挖掘到事物的本质，能够找到问题的根源。这五个问题永远都是数据分析师应该掌握的内容，否则就仅仅是数据处理人员，称不上数据分析师。

What – 存在什么差异或者变化？这个差异和变化是什么？是向好还是向坏的方向发展的？是好现象还是坏现象？是需要警惕还是需要挖掘创新的？What is the change? ——这个变化是什么？这个变化的本质是什么？

Who – 这个变化的主体是谁？客体是谁？是谁在变化或者是谁引导了变化？是谁造成的变化？这个变化的责任方应该是谁？如果是好的话，谁该领功，是坏的话，谁该负责？他为什么要引领这个变化或者他有什么样的动机要让这个变化发生？背后的利益或者情感的驱动是什么？

When – 这个变化是什么时间发生的？是最新发生的吗？是过去就一直在发生着，只是现在变化大了才发现的？这种变化持续多久了？多长时间才发生了这么大的变化？是快速还是慢速？按照这个速度发生变化下去，多长时间可以发生质的变化？这种变化是否需要阻止？最迟需要什么时候阻止？这个变化是否需要加速？如果需要，到什么时候需要达成什么结果？

Where – 这个变化在哪里发生的？发生的环境要素是什么？牵扯到哪些部门、组织或者利益相关方？每个利益相关方的诉求是什么？会对该事物的发生和发展有什么影响？这个变化发生的环境因素是必要条件还是充分条件？是什么条件达到了才产生了这个变化？这个条件是否可以人为创造？如果不能人为创造，自然界将有什么可能形成这样的条件？

How – 如何调整？如何强化？如何弱化？如何创造事物发生的条件让事物按照我们期望的方向进行变化？如何才能改变现在的状况？这个变化是怎样发生的？这个发生能否避免或者重复？

4W1H的五问方法可以让数据分析人员能够把握问题的关键，能够找到问题的根源。在执行过程中需要钻研精神，不放过任何一个细微的变化与差别，对所有的问题有追根寻底的精神。刚开始学习这五问方法的时候，容易陷入一个窘境，就是有太多的问题要问了，非常少量的数据需要花费大量的时间去追根究底。其实，当你对公司业务熟悉之后，在追问问题根源的时候，也会更有重点，当然这也可能是遗漏问题的根源。越想走捷径，越容易出遗漏的问题。

**7、找源头 Root Cause**

上节中4W1H的方法可以让我们通过问问题找到问题的关键和问题的根源，找到发生变化的原因，从而能够找到解决问题的方法，是一个非常全面的方式。而在进行数据分析的时候，找到问题的根源，问题的根，却是一个比较见功底的活儿。

一个公司中的管理问题往往根源都在管理上，也就是都在老总对公司的管理上。常言道，问题都在前三排，根源就在主席台。很多公司的管理问题都可以追究到企业老总的管理方法和管理理念、管理方式、管理思路的问题。为何公司令行不止？因为老总言行不一致；为何很多人收贿受贿？因为老总就是这样的人；为什么很多人光说不做？因为老总只喜欢那些说得好听的；为什么大家都不愿意实干？因为出错之后，老总就会惩罚，不干无错。。。太多的问题都容易指向老总的问题。

有些问题的根源来自于公司的潜规则或者长期以来养成的习惯，有些公司的员工不作为，是因为长期以来大家养成了懒散的习惯，没有尽职尽责的精神。制度流程的问题往往是管理中常见的问题根源。门在哪儿开，人们就从哪儿走。门开错了位置，大家就会走错位置。公司的制度和流程缺少严谨性和严肃性，往往导致很多制度和流程最终流为形式；如果制度和流程在不同的人身上发生不同的变化，那么制度和流程的严肃性就会被质疑而失去其权威性；如果制度和流程失去权威性，很多人就会钻空子。

随着公司规模的变化、人数的增多、业务的复杂程度增加，原有的制度和流程可能会发生变化，出现不适用的现象，需要根据公司发展需要进行调整。数据能够显现出一些问题，当这些问题出现后，数据分析人员需要根据公司的情况，提出相关的建议和意见，协同相关的部门对制度和流程进行修订，并追踪制度流程修订后的结果，用数据来追踪效果，确保这种修订是好的、可行的、有效的。就如前面分析的月度销售额曲线图案例那样，公司的目标制定流程不科学，需要重新设定公司目标制定和分解的流程；另外公司员工激励的制度不合理，不能有效地激励员工超额完成任务，需要协同人力资源部修订员工激励方案，确保员工能够把能力发挥到极致，让公司切实把握所有的发展机会来发展，不能因为员工的问题影响公司的持续。

在挖掘数据现象背后的Root Cause的时候，需要我们对事物的本质有清醒的认知。同样是营销和销售，每个公司对这两个部门的定义会有很大的区别。在快速消费品公司，营销主导销售，公司的销售模式是通过广告吸引消费者购买，而销售需要做好产品的分销，确保消费者去购买的店面中有可销售的产品，不断货；而在药品、医疗器械、耐用消费品中，销售起到非常重要的作用，在终端消费者决策中起到非常重要的影响，而营销则是为销售提供各种销售工具的。而这种差异来自于什么呢？购买者的决策机制。当购买者的决策机制是由感性的品牌喜好决定的时候，营销主导销售；当消费者的购买决策复杂，需要专业知识、理性决策等的时候，需要专业的销售人员或者专家在销售点能够临门一脚地把我成交过程。这里的Root Cause就是购买者的购买决策过程。由此你可以推断下去，B2B的销售往往都是销售主导营销，或者营销为销售提供支持服务。

**8、找关系**

通过数据找到事物之间的关系是我们比较容易做到的。通过散点图，我们看到相关的分布状况，我们就能够直观地判断两者之间的关系是否是相关关系，相关性有多高，随机误差的影响有多大。



有时候我们需要对数据进行一定的调整才能看到其相关性，比如可以通过对数关系、幂指关系等处理后再看其相关关系。相关性高可以有正相关和负相关，相关性低，说明两个关系不紧密。

相关性研究需要比较谨慎，一方面要有足够的丰富的专业知识，同时还不要违背常识。比如，如果说监狱中85%罪犯都喜欢吃米饭，我们是否可以得出结论说米饭是导致犯罪的原因？调查研究显示癌症患者中只有18%的人是重烟民，但90%的人都吃米饭，有65%的人每天至少吃一顿米饭，那么是否可以得出结论说米饭是导致癌症的原因？纽约股票交易所的股票指数与华尔街上女性的裙子长短呈正相关，股票下跌，裙子变短；股票上涨，裙子变长，那么华尔街女性的裙子长度是导致纽约股市涨跌的原因吗？

这些是无法立住脚的分析判断，但数据分析软件或者算法所给予我们的只有这些相关性，只有真实的数据和指标，我们解读的时候必须加入常识的成分，不能单纯地从数据角度去判定事物间的关系。

**9、找驱动**

数据无法直接告诉我们事物变化背后的驱动要素是什么，只能告诉我们相关的关系或者两种指标之间的变化关系，我们需要自己找寻事物变化的驱动关系。

汽车加入汽油，通过发动机的启动，我们就能够行驶几百公里，这是典型的汽油驱动发动机、发动起驱动车轮前行的逻辑关系，但现实世界，特别是商业环境下，这种单纯的关系是不存在，甚至我们认为存在的关系，也会变得更加复杂。

广告可以扩展我们品牌的知名度、提高消费者的认知度，从而提高我们的销售量，但广告和销售量之间的关系并不像发动机那样明显。我们把广告投放到电视媒体、广播媒体和互联网媒体，让我们的品牌有更多的曝光率，但是曝光率和购买率之间还有很大的距离，我们如何来评测广告的效果，站在一个公司的角度，是投入和产出之间的关系，而站在消费者的角度，则是看到与想买与可买与最后购买之间的转化率关系。

数据分析要从多个维度进行，一方面要从公司的维度，广告是公司的投入，销售收入是公司的产出，这是一个投入产出之间的关系，另外一方面，站在用户的角度，如果广告消费冗余，看到广告会更加腻烦，本来还希望购买你的产品，但看到你们比较恶心的广告或者不断骚扰我们日常生活的广告之后就不想再买了。因为消费者知道，你花费了广告费打广告，我如果购买你的商品，我购买的价格中包含了你的广告费，且你的广告这么恶心，对我造成了骚扰，为了不让你继续骚扰我们，我不想购买你们的产品，这样我从口袋中掏出来的钱不会花费在无聊甚至恶心的广告上，不会再对我们造成骚扰，最好你们能够从地球上消失，这样我在看网页、看视频、看我喜欢的内容的时候不会出现你们无聊的恶心广告。消费者会越来越成熟，你是否能够让广告更加符合他们自身的需求，这会变得越来越重要。

我们在进行数据分析的时候经常会分析转化率的问题。销售的分析带有转化率分析，我们获得了多少客户的信息、有多少比例愿意跟我们接触，有多少人愿意了解我们的产品，有多少人对我们的产品感兴趣，又有多少人对我们的产品产生了购买意向，最后有多少人最终购买了我们的产品，这整个过程的数据体现出各个阶段的转化率，不同的转化率代表着我们营销和销售的效率问题。



我们数据只能分析从一个阶段到另外一个阶段的转化率问题，却无法发现从一个阶段到下一个阶段转化的驱动要素，我们可以通过各种方式采集数据，但现阶段的技术无法实现驱动效果分析，我们能做的是通过各种不同的实验方法，然后分析不同方法的效率和效果问题。这种方法我们称为A/B测试。

所谓的A/B测试，就是我们针对同类的客户，采取A方案的方法和B方案的方法，然后来对比A方案和B方案的方法之间的效率差异、效果差异，然后确定那种方法更加有效率。数据分析最终的结果可以评测A和B的效率与效能问题，但无法揭示A为什么有效、B为什么低效，或者背后的逻辑是什么，这就需要人为的参与解读和理解，而这些理解往往是建立在假设或者臆想之上的，没有足够的证据证明其驱动过程和效果。

A/B测试是一种方法，是数据分析人员与公司的管理者达成一致进行实验的一种方法，如果管理者本身不想进行测试，这个过程就很难实现。因为管理者有其管理目标，A/B测试总要锁定一定的条件，导致管理者在过程中的自由裁决力度减弱，在这个方面，管理者为了保证既定考核目标的实现，往往会采取其喜欢的方式来管理，而让整个测试结果不可用。为了推行A/B测试，需要公司的高层给予支持，对比A/B两个方案，让管理者在背负相关责任上要有明确的合理的适合科学实验精神的标准，否则这个测试就会带来不一样的结果，甚至得不到希望得到的结论。

我曾经主导过一个A/B测试，其费用投入差不多在2亿元人民币左右的广告费用，但在过程中，因为投入成本高，管理者的业绩压力巨大，在出现问题之后，需要做些调整，但这些调整会让整个测试失去可对比性。无可否认，测试只能向公司管理目标倾斜，最后只好放弃，感觉非常遗憾。一方面，这个测试的标的过大，导致很多管理者的压力太大；另外一方面缺少最高层领导的支持和理解，他们也不想放弃短时期内貌似可行的方案，会让整个对比试验夭折。短期利益的驱使让整个A/B测试会中途夭折，这是数据分析人员需要在开始之前就要做好预案的。当然，这需要经验积累和阅历，同时更加需要专业的权威性支持。

**10、找规律**

寻找规律是数据分析与挖掘的最基本的目标，也是唯一的目标，数据分析的根本目的还是指导未来的经营实践。通过数据分析找到事物发展的规律，从而指导我们的企业经营实践，这样我们就能够对未来做出更加准确的预测，能够把控未来的发展方向。

比如，我们研究互联网行业整个行业的发展状况，纳斯达克股票交易市场的综合指数基本能够反映长周期的市场情况，也能够反映整个市场的价值。如下图：



这代表着一个技术从发展到应用普及所走的道路。Gartner机构研究了上百种技术的路线，得出了类似的结论。感兴趣的朋友可以用谷歌搜索Gartner Hype Cycle，如果对本行业的技术应用阶段性的分析感兴趣，可以在谷歌搜索中加入行业的关键词，比如说数字医疗： Gartner Hype Cycle Digital Healthcare。

基于这个技术的发展曲线，我们对互联网的20多年的发展历程做了分析，我们得到这样一个图形。



互联网发展经历了三个发展阶段，第一个阶段是技术发现的第一个5年，第一个五年中，大家对技术的认知还处在一个不断完善的过程，这个技术到底是什么，大家都在盲人摸象的阶段。最初大家对互联网的认知就认为互联网是信息传播的一种方式，可以替代以前的媒体，包括纸质媒体（报纸、杂志、期刊等）、电视媒体、广播媒体，能够给人们带来更多更方便的信息，所以，当时衡量互联网的关键指标就是点击率，如果一个网站被点击次数更多就更加值钱。在这个阶段，投资者认可，技术也在不断创新，在资本的追逐下，更多的公司被创立，并开始了很多的技术创新，一种新的技术可以吸引大量的资金投资，就如《硅谷》美剧中所讲的故事，如果一个技术被发现很有商业价值，各种投资公司追逐下，这个公司的估值会大幅度放大，一个压缩技术能够让一家公司估值超过10亿元美金。

第二个阶段，随着技术应用价值被重新估值，技术泡沫逐渐被穿破，媒体和社会都口诛笔伐，导致大家对技术的信心在下降，进入幻灭期，这个时候资本更加理性，商业应用也更加关注商业本身，而不是技术本身，大家专注思考实际应用价值，一些新的应用型公司成立。在过热期和幻灭期两个阶段是第一个黄金投资期，目前互联网大型公司基本都是这个时期成立的。

随着技术泡沫期消退，技术创新应用被不断地挖掘，新的应用出现，只要技术够给力，巨量级的技术应用公司发展起来。互联网在这个时期出现了微信、微博、whatsapp、Uber、滴滴打车、百度外卖等巨量级的应用，技术的价值被逐渐深度挖掘。

通过对互联网技术发展历史的研究，我们可以来研究大数据技术的未来发展的可能性，经过Gartner的研究，各种技术都遵循雷同的发展曲线，所以大数据也基本上会遵循这样的发展路线。在目前的阶段，大数据技术到底在什么阶段我们需要有充分的认知，这样我们就能够在数据分析的时候，理解大数据本身所展现出来的价值。



互联网的本质是信息传播的方式，经历了1.0、2.0、3.0的发展，现在在向4.0发展。Web1.0的时期，我们认为互联网是媒体，是传播；Web2.0时代，消费者可以发出自己的声音，消费者产生内容，UGC产生，消费者的力量逐步显现，同时，作为消费者交流的平台，电商应用开始发展起来，亚马逊、eBay， 淘宝、京东这样的公司开始发展起来；当到了Web3.0时代，消费者因为交流而集结成群，形成了各种社群，消费者彼此之间的交流互动成为互联网的主题，而信息传播、交易则成了辅助，用户之间的关系发生的更大的变化，原先不怎么联系的社会关系变得更加紧密，而之前经常联系的亲戚关系反而不再是新交流沟通模式下的交流沟通的主体。所以，互联网技术随着发展逐渐改变了媒体（把纸媒都干掉了）、改变了人们交流沟通的方式，改变了人们的社会关系，改变了我们交易的方式，也改变了企业做生意的方式，逐渐由原来的线下交易线下交付，逐渐转变成为线上交易、线下交付（O2O）。

而大数据的本质跟互联网还是有着本质的区别。大数据本质上是一种数据和信息的加工方式，这种加工方式将代替人们的思考方式。过去我们通过掌握的信息、看到的和听到的，经过大脑的加工形成意识判断，从而做出相关的决策，所以我们过度依赖人们大脑的聪明程度。而随着大数据技术的发展，人工智能将成为信息加工的主要方式，计算机将逐步代替人类大脑的大部分功能，通过计算技术、分析技术、大数据处理技术，逐步取代人们的大脑，产生智慧，影响着我们的各种决策，包括商业上的决策。如果说互联网改变了我们的沟通方式的话，大数据将改变我们的思考方式和做事方式，必然也会改变我们企业的生产方式，最终会改变我们生活中的各种物质要素，最终改变我们的生存方式。

我们知道，任何一个行业的发展规律都遵循一个扁平的S曲线，上图中的技术发展路线也是一个S曲线，但在现代社会中，我们对技术抱有更大的期望，所以会影响整个行业的发展曲线发生变化。掌握一些基本的规律，能够让我们在做数据分析的时候能够清楚地知道现在和未来发展的趋势，通过数据分析过程中的假设和验证更加有效。掌握的知识越多、经验越丰富我们就能够提出更多的假设，就能够进行更多的验证，然后证实更多的数据发现。



全文摘自《企业经营数据分析-思路、方法、应用与工具》赵兴峰著

**上期内容：《看懂数据分析图标的10个基本方法》（上）**

1. 辅助线 ②找差异、找变化 ③找奇异点、特殊群体 ④找转折点、拐点 ⑤找特征