### 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。

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

**1、辅助线**——强化差异的辅助线让我们更加清楚认识事物的规律性、特征性

我们举一个例子。



上图是一家快速消费品公司的月度销售额曲线图，总共是2年半左右的数据。从图形上我们可以看到这家公司的月度销售额在不断地上升，呈现稳定增长的态势，多数月份都是小幅波动，只有部分月份波动幅度比较大一些。

如果仅仅停留在这个层级的解读上，我们可能还没有看到这个公司销售管理的问题。我们添加一下“辅助线”，将上面的图形用下面的方式来“辅助解读”一下。



添加辅助线之后，你会发现在每年的11-12月份销售额不增长或者呈下降趋势，而每年的1-2月份实现井喷式增长，而到了3月份有大幅度下滑。这个带有春节的特征，但作为快速消费品来讲，消费者年前大幅度购买存年货是一种可能。而在这种可能下，业务人员为了达成年度销售目标，在每年的12月份做好促销，从而让本年度的销售目标超额完成。为了追究原因，数据分析人员需要去查看市场，了解具体情况。

实际考察发现，这个公司每年的12月份，货架上的产品几乎空了，是销售团队“技术性断货”。是因为销售团队完成了当年的任务目标，担心今年完成的太多，明年的销售目标定太高，给自己带来完成目标的压力，所以技术性断货。而产品是非常受消费者欢迎的，当消费者发现自己喜欢的产品断货了，等到产品重新上架的时候，消费者会一下子购买很多，以免下次断货不好买到，所以造成1-2月份的销售额大幅度提升了。这种现象我们称作“压货”，压着货物惜售现象。

压货会造成公司营销费用的浪费和销售机会损失，会阻碍公司的发展速度，影响公司市场占有率的提升。同时，因为供货不及时会造成渠道和终端抢货，无形中形成过高的库存，甚至让公司对终端销量的预测失效，影响决策的准确性。

压货现象的存在是公司在销售目标制定上的不科学，武断地将去年的销售额作为基础，生硬地添加一个比例来作为新一年的销售额目标，对市场需求、公司发展、竞争对手等估算不足，不能按照市场情况制定合理的销售额目标；另外，公司的超额完成任务的奖励上不足，导致销售团队超额达成当年销售目标的积极性不足，从而宁愿牺牲很小的超额完成奖励，也不愿意明年背负更多指标。这是人力资源设定的绩效激励机制的不足。

同样的图，我们再做第二条辅助线，如下图。



通过添加这一条“辅助线”，你会看到整个公司的年度增长竟然呈现一个“大阶梯状”，而不是我们常见的“斜坡式”。一个公司销售额的增长肯定是一个斜坡式，因为随着公司品牌的推广、渠道的推广，知名度在慢慢地提高，而不是一下子就提高了，只会是一传十十传百地递进式增长，而不会是一年每个月变化不大，第二年一下子上到一个台阶之后，继续在这个台阶上持续销售相同的数量而不增长。辅助线这种特征的销售额在一定意义上反应了公司的“限量销售”，公司一个月就只有这么多货物，只有这么大的产量，产多少卖多少，多了生产不出来。第二年上了一个台阶，说明扩大了产能，供应量增加了，第三年又是通过产能扩大来实现的销售增长。

仔细审查下公司实际的产量，如果不是产能限制导致的这个图形，那么就是销售团队的管理问题。该公司产能利用远远未饱和，销售人员按量销售，一个核心的问题在于销售人员的月度指标是按照全年指标除以12来设计的，销售人员完成了当月的指标之后，就没有了积极性去继续销售更多的产品。当第二年销售目标重新设定的更高了，仍然能够快速地完成，说明这家公司的产品还是有非常大的市场潜力没有被挖掘，销售人员只按照其目标和奖励机制在工作，并不是为了公司的发展和市场占有率提升，这导致整个公司丧失了大量的销售机会。背后的根本原因两个：1）销售目标分解不合理，年度/12，没有考虑从1月份到12月份的增速；2）销售超额完成激励不足，销售人员没有动力去卖超过月度目标的产品。

辅助线是一个强大的工具，就如我们中学里学习几何要利用辅助线解题一样，我们也需要利用辅助线来解读我们的数据。必要的时候，在图上添加几条辅助线，能够让我们更好地解读数据背后的故事。

**2、找差异、找变化**

找差异和找变化是我们天生的一种技能，我们人类认知大自然的时候就开始认知差异和变化，并从不断地大自然的变化中寻找规律，日出日落，我们有了利用影子来计时的发明——日晷(gui)，或者叫做日规。

如果事物没有什么变化，我们往往得不出一个明确的结论，当有变化的时候，我们寻找事物变化的驱动因素，从而形成事物变化的驱动要素，如果我们希望事物做出某种变化，我们就按照我们从事物变化中总结出来的驱动要素，配置这样的驱动要素，让事物做出变化，这是我们积极主动推动事物发生变化。如果我们能够有成功的驱动要素，我们按照成功的驱动要素积极去构筑这个成功的条件，我们就能够取得成功，企业也能够取得成功。这也是我们经常喜欢读传记、读别人成功的故事。

一个事物的变化往往有必要条件和充分条件。充分条件的意思是只要这个条件达到了，事物必然会按照预定的方向呈现出变化；而必要条件是指要想事物发生变化，这些条件必须具备，但仅仅具备这些条件，事物还是不能按照预期做出相应的变化。但当我们用数据或者现象事实来做分析的时候，如果找出的是充分条件，我们就成功完成了分析工作，但如果只是找出了必要条件，我们距离成功还有很长的路。要想验证我们找出的条件是必要条件还是充分条件，我们就需要通过重复实验来完成，如果我们不能通过论证来达成的话。

我们听了太多成功的故事，当外部环境发生变化的时候，我们沿用故事中主人翁成功的方法的时候，发现原来成功的充分条件可能仅仅是必要条件而已。我们听了马云太多的成功故事和心灵鸡汤，现在80后的人基本上都是喝着马云的鸡汤而长大的，马云创业成功的故事激励着很多人去创业，去尝试，但马云成功的外部环境对比现在已经发生了翻天覆地的变化，他的成功路径已经无法复制，他的鸡汤虽然营养丰富，但这些基本都是必要条件，而非充分条件。

做数据分析的专业人士需要对事物的变化有足够的敏感性，不要因为变化微小而忽略，更不能因为变化不常见而不见。敏锐地觉察到事物细微的变化是我们比别人更早发现问题、发现规律、发现事物发展方向的捷径。如果不是这样，等大家都看得到变化的时候，你所看到的，跟别人看到的就没有不同了。

“春江水暖鸭先知”与“温水煮青蛙”的故事有相反，我们在阅读大量数据的时候，我们能够感受到事物的微小变化，能够做到“春江水暖鸭先知”，如果我们感受不到事物的微小变化，我们就会是“温水中的青蛙”，因为感知不到微小变化而逐步被同化。

有次回农村老家，是2014年，我发现78岁的伯母都能够拿着智能手机跟远在外地的堂兄聊微信聊半天。虽然不识字，但按住就说话，放开就发送，这么简单的操作。之前我不觉得这个简单的设计多么强大，自从我发现78岁的文盲老太太都会用微信语音聊天的时候，我真的被感动了，这个全民普及智能通信的时代，所有人都在网上成为网络的一个分子，这必然会产生裂变式的效应。未来智能手机将还会更加简单，不需要按键，比Siri还聪明，能够感知到你想要的和想说的，一个78岁的农村文盲老太太也能够直接跟手机对话，完成老太太希望完成的事情，比如说跟儿子通话、跟儿子视频、看孙子玩耍等等。

找变化是人的本能技能，但寻找变化背后的驱动因素需要专业能力，需要对事物的理解，需要一些基本的常识，如果在企业中，那就需要对企业经营、行业规则等有较为深刻的认识，能够辨别是充分条件还是必要条件。

**3、找奇异点、特殊群体**

与传统统计学不同的是，大数据时代下，我们的数据集都是全部数据，不是抽样数据，如果存在奇异点，不会是因为抽样造成的、也不是随机造成的，是切切实实存在的，而奇异点的存在必然就有其道理——存在即为合理。而奇异点的深度研究能够让我们感知到一些新的变化或者细微的变化，非常小众的东西有可能会成为受大众欢迎的东西。

我始终不认为大数据更加关注事物或者数据的相关关系、关联关系，而不太重视数据间的逻辑关系、因果关系。为了更快地商业化数据分析结果，我们可以将发现的关联关系、相关关系、共生关系应用到实践中去，但追求数据现象逻辑规律的努力不能停止。我们需要了解事物为什么发生，为什么存在？我们需要知道事物的逻辑规律，我们才能更好地预知世界，才能产生更多的知识和智慧。

而特殊点、奇异点、特殊群体的存在为我们揭示了世界的差异所在，而这个奇异点未来会不会成为主流，未来会不会成为一种趋势，这都是需要我们探究的。探究奇异点的存在，让我们发现更多新生的东西，是我们做数据分析的乐趣所在。

我在一家婴幼儿奶粉公司做商业智能的时候，发现有个用户购买了大量的三段奶粉，而这个用户本身并没有婴幼儿，我原以为这个用户是买来送礼的，因为并没有转卖的机会，我们的分销渠道覆盖率非常高的。但是连续两个月、三个月都购买了大量的奶粉，应该是4-5个婴幼儿的用量。后来我们追踪下去，发现这个用户是买来给自己家里术后住院的家人的，家人手术后无法进食，而我们的奶粉是全配方奶粉，是非常全面的营养配方，只要吃奶粉，所有的营养多全面了。我们内部开会的时候我提出一个想法，我们公司的奶粉除了婴幼儿市场外还有一个值得关注的市场，那就是老年人和手术后营养补充市场。这个市场应该也不小，值得关注开发对应的产品。

这就是通过关注奇异点的数据，并深度挖掘之后，找到新产品开发路径的方法。如果不关注奇异点，或者发现奇异点之后就放弃了，这是非常可惜的。一些奇异点能够给我们带来创新的启发。

就如海尔洗衣机在农村市场发现有人用洗衣机来洗土豆，然后开发了用来清洗土豆的洗衣机一样，一个奇异点的数据可以给我们带来新的探索。如果我们把握好了奇异点的数据，然后研究下去，就会发现完全不同于我们常规理解的内容，我们就能够有了新的启发点、新的引爆点。

**4、找转折点、拐点**

数据的转折点和拐点是非常重要的，甚至是非常关键的分析要点。事物发展都有惯性的，而转折点和拐点突破了事物常规发展的惯性，完全去到另外一个方向，这需要强大的力量来扭转，这个拐点和转折点的分析能过让我们认识这个强大力量的真实实力。

一个公司销售额持续增长，忽然间下滑了，那么这个下滑的驱动因素是什么？如果忽略了这个驱动因素，有可能公司就像高山滚雪球，越往下速度越快，而雪球越滚越大，最后一发不可收拾，届时就彻底晚了。

中国经济的发展过程，这个图形我们前面分析过。中国经济发展的35年中我们有好几次的拐点，每次拐点都是我们释放了巨大的机会导致了新的突破式增长。



其实仔细研究中国经济的发展历史，你会发现很多个拐点，有的拐点很大，比如说1981年-1982年的快速增长，1984-1985年的持续下滑，1990-1991年的再次快速增长，1992-1993年后的进一步下滑，1999-2000年的止跌稳步上升，2007年的巅峰时期到2008年的大幅度下滑。研究这些拐点，能够让我们认识到一个国家经济发展增速的驱动力，以及驱动力发挥规程中的阻碍力量，以及每个驱动力的极限。

研究行业数据也要认识到数据的拐点，一个行业快速发展到逐步稳定，到后期的持续下滑，然后逐步进入不增长的阶段，每个行业都会有一个发展历程和周期，都会有不同的驱动要素。

研究公司的发展历史也可以看出一些常见的规律。一个企业发展到20-30人就需要开始建立分工的组织体系；当发展到50-100人的时候，就需要建立相对完善和严谨的制度和流程；当发展到250人-300人的时候，企业需要开始考虑每个人的职业规划，公司的职级、薪级体系需要完善；当公司规模达到500人以上的时候，需要不同的业务领域或者分组管理，开始考虑业务的中长期规划，有种子业务、发展业务和核心业务，需要在人员管理上做出不同的调整，在业务安排上有不同的激励措施；当达到1000人以上时，企业又开始了面对新的挑战。以上所说的人数规模是大概的，不同的行业、不同的业务复杂程度、人员文化水平肯定面临瓶颈期的时间是不同的。

行业环境也会大大影响企业瓶颈期的到来。快速发展的行业，企业的瓶颈期到来的就晚一些，行业发展停滞时，企业的瓶颈期到来的就早些。原因很简单，当企业乘坐行业的大船快速发展的时候，只要发展，大问题也不是问题，当行业发展停滞，企业的发展也受到影响，企业不发展，小问题都会成为大问题，甚至是致命的问题。

**5、找特征**

事物的发展总会留下历史的证据，而这个证据是需要不断挖掘的，这就是数据分析的经验所在。看过美剧CSI (Crime Scene Investigation, 犯罪现场调查)的朋友都会知道，所有犯罪证据的收集都是时候，根据现场留下的各种证据然后尽兴推测演绎还原等方式来推演过去发生了什么。

其实只要事物发生发展的过程中记录了数据，就会留下事物发生和发展的影子，就能够推演过去发生了什么，为什么会发生，怎么发生的，然后找寻其中的驱动要素或者发生发展的规律，这就是从数中找寻证据。

比如说前面我们研究的一个公司月度销售额的曲线图，因为我们采取了压货的策略，必然在销售数据上呈现了下行，而断货恢复后，消费者购买里反弹，导致销售额上升，这个变化是“压货”这个行为造成的结果在数据上的反应。如果我们的数据真实、准确、可靠，会留下事物发生过程的影子，是无法造假的，只有你发现不了，没有不留痕迹的数据。

任何的事物发展都有其自身的特征，如果能够对事物的数据特征有更好的认知，就很容易从数据的形态表现上看到事物是怎么发生和发展的。这就是我们说的，要看数据表现出来的特征。

识别特征是需要经验的。如果不知道什么叫做压货，什么叫做囤货，为什么要压货，为什么要囤货，那么就无法识别出上例数据中的压货现象。这也是为什么作为数据分析师跟其他的技术人员、编码人员职业不同的原因，数据分析师类似老中医，经验越丰富、经历越多，其价值越高。越有经验的数据分析师越能够从一些简单的图形中看出问题所在，时间积累的经验为数据分析提供了更多的知识。虽然在技能上、在软件操作上、在代码编写上落后于年轻一代，但在数据解读和数据识别上，资历深厚的数据分析师能够给出更加丰富的解读。

换一句话来表达这个意思，我们可以说，越懂得人性，越能挖掘数据背后的故事。如果我们拥有更多社会学、心理学、经济学、管理学等多学科的知识，我们能够了解企业中会发生什么，我们在解读数据过程中就能够有更加丰富的想象和假设，然后根据想象和假设寻找想对应的验证，然后佐证我们发现的问题。如果没有这个过程，数据分析只会停留在对数据本身特征的描述上，不能解释数据所反映的事物的发展变化规律。

做股票交易的朋友会有各种看图的本领，比如说一个5日均线、10日均线、40日均线与主交易线穿插代表什么，什么是大阳线、什么是大阴线、什么时候预示着上涨行情，什么时候预示着下跌行情，什么是金叉，什么是死叉，什么是交叉，这些都是经验的积累，也有数据分析的结果。通过图形本身的特征预示着事物背后的相关规律，这是从数据的角度去解读事物发生和发展规律的基本方法。

为了寻找数据的特征，可以采用各种方法，一种方法就是呈现在图形上，肉眼来观察具体数据集的特征，另外一种方法就是我们常用的描述统计的方法，通过观察描述统计后得到的指标找到数据分布的特征，从而能够更准确地理解数据集，包括数据集的平均值、极距、最大值、最小值、中值、峰度、偏度等信息。

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

**下期预告：《看懂数据分析图标的10个基本方法》（下）**

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