



第3章 用数据分析解决问题

前面我们讲了每一种分析方法的使用和案例。但是在实际工作中，业务问题比较复杂，是无法用单独的某一个分析方法来解决的。

这一章通过学习如何将单独的分析方法组合起来解决问题。学会以后，再遇到其他问题就可以灵活应用。随着处理的问题越来越多，工作经验越来越丰富，相信你会积累起解决实际问题的经验。

3.1 数据分析解决问题的过程

最近我一直咳嗽去了医院，我和医生发生了下列对话：

医生：“咳嗽多久了？”

我：“最近一周都在咳嗽。”

医生：“看下哪里出问题了。我发现你嗓子没发炎，但是在流鼻涕。”

我：“为什么会流鼻涕呢？”

医生：“因为最近天气变冷了，而你穿得少，导致感冒了。”

我：“那如何应对呢？”

医生：“我给你开些感冒药，同时要注意多穿衣服。”

看病的过程其实就像数据分析解决问题的过程（图3-1）：第1步明确问题；第2步分析原因；第3步提出建议。

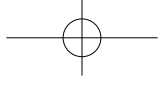


图3-1 看病的过程

在现实生活中，其实也是用这个过程来解决问题的。数据分析解决问题的过程如图3-2所示。



图3-2 数据分析解决问题的过程



120 数据分析思维：分析方法和业务知识

第1步：明确问题。

通过观察现象把问题定义清楚，这是数据分析的第1步。只有明确了问题，才能围绕这个问题展开后面的分析。如果一开始问题就定义错了，那么再怎么分析，也只能是白白浪费时间。

第2步：分析原因。

这一步是分析问题发生的原因，可以通过下面两个问题把原因搞清楚：

- (1) 哪里出了问题？
- (2) 为什么会出现这个问题？

对应看病这个例子，医生一开始在想“哪里出了问题”，是嗓子导致咳嗽，还是流鼻涕导致咳嗽？最后发现，问题出在流鼻涕。然后，医生在想“为什么会出现这个问题”，围绕这个问题去查找原因。

第3步：提出建议。

找到原因就完事了吗？还不行，要找到对应的解决办法才是分析的终点。所以，找到原因以后，还要针对原因给出建议，或者提出可以实施的解决方案。

当完成这3步，才能把分析的过程清楚呈现给决策者。这就是使用数据进行有条理的分析，从而准确做出决策的能力。

接下来通过一个案例来学习每一步如何具体去做。

案例背景：某店铺上半年（1—6月）完成的利润，与年初制定的月平均盈利500万元的目标还有很大差距。如果按目前销售进度，到年底没有办法完成全年6000万元的总利润目标（全年总利润目标=月平均利润目标×12=500×12=6000万元）。销售总监要求你尽快找到利润没有达标的原因，以及拿出完成年度目标的方案。

3.2 如何明确问题？

我们来学习“用数据分析解决问题”的第1步：明确问题。

数据分析的目标是为了解决生活或工作中遇到的问题。例如你想找数据分析相关的工作，想知道哪些城市找到工作的机会更大，怎么办？再例如，你是公众号的运营，发现最近一周的新增用户数下降了，怎么办？

面对这些问题，数据分析都可以帮助你。数据分析解决问题的第1步是定义清楚问题，这决定了后面工作的成败（图3-3）。

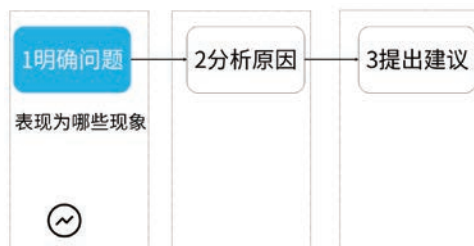


图3-3 明确问题

3.2.1 明确问题的常见错误

有一个段子是说一个小伙子看到一个老头和他脚边的一只狗。然后小伙子问：“你的狗咬人么？”老头说：“不咬人。”于是小伙子弯腰摸了摸这条狗，结果被咬了一口。小伙子气地说：“你不是说你的狗不咬人么？”老头说：“这不是我的狗。”

像这种定义问题的错误在工作中随处可见。例如领导告诉你：“可能是客单价高，最近利润下降了。”听到领导这番话以后，你将问题定义为“客单价高导致利润下滑，怎么办？”这样定义问题有什么风险呢？

答案是，在定义问题中就已经包括了“原因”（在这个案例中是“客单价高”）。这样后面所有的工作都变成了收集与价格相关的数据。那么这就导致分析者忽略了其他可能的原因，而只限制在“客单价”这个视野范围内。明确问题决定了你的分析范围，而错误定义问题会缩小分析的范围（图3-4）。

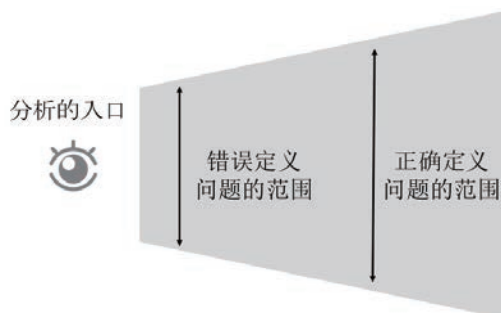


图3-4 明确问题决定了你的分析范围

上面的案例是“明确问题”时常见的错误：定义问题时，分析者根据自己的经验“主观”地把思考限定在“我觉得的范围内”。这种错误也叫做“确认偏误”，就是人们总倾向于寻找证据来支持自己已经相信的事情，从而抵制不同的看法。我们要时刻清晰地提醒自己：数据分析不是主观的臆断，是一种客观的分析。

3.2.2 如何明确问题？

明确问题可以图3-5所示的方法来进行。

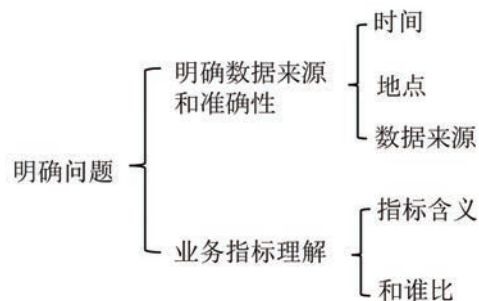


图3-5 明确问题的方法

1) 明确数据来源和准确性

为什么要明确数据来源和准确性呢？

我们日常看一篇新闻，想知道这篇新闻是“假新闻”还是“真新闻”，最靠谱的办法是看来源。如果是官方发布的，那么可以保证新闻的可靠性。如果是从哪个八卦平台看到的，那么新闻的可靠性就不能保证了。

同样地，在明确问题时，也要保证数据本身是准确的，才能进行后面的步骤。所以，要明确数据来源和准确性。

如何做呢？可以从时间、地点、数据来源确认问题表现为哪些现象，还可以通过向相关人员提问的方式来沟通清楚。例如可以这样提问：

针对时间：这是观察哪个时间范围的数据发现问题的？

针对地点：这是哪个地区的数据？

针对数据来源：数据来自哪里？是否核对过数据没有问题？

在本章案例中，可以通过与相关人员沟通，了解清楚“利润没有达到目标”表现为哪些现象。例如，可以观察销售部门（数据来源）提供的上半年1—6月（时间）全国（地点）每个月的实际利润，并和每个月的计划利润对比，如表3-1所示。

表3-1 上半年的利润

| | 1月份 | 2月份 | 3月份 | 4月份 | 5月份 | 6月份 | 总计 | 平均 |
|--------------|--------|--------|--------|--------|--------|--------|---------|--------|
| 实际利润 (万元) | 46.60 | 89.25 | 286.87 | 378.69 | 520.08 | 669.88 | 1991.37 | 331.90 |
| 计划利润 (万元) | 200.00 | 250.00 | 350.00 | 500.00 | 800.00 | 900.00 | 3000.00 | 500.00 |

从表中可以看出，公司上半年的利润下达计划为3000万元（表中计划利润总计），实际完成1991.37万元（表中实际利润总计）。上半年的利润缺口=计划利润总计（3000万元）-实际利润总计（1991.37万元）=1008.63万元，也就是1—6月完成的利润是落后于计划的。

目前要解决的问题是：上半年（1—6月）利润落后于计划的原因是什么？公司如何在下半年（7—12月）完成4008.63万元（上半年利润缺口1008.63万元+下半年利润目标3000万元）的利润目标？

2) 业务指标理解

对于业务指标，可以分析指标含义、和谁比。

在这个案例里，是用“利润”这个指标，那么可以这样问：利润这个指标是怎么定义的？

通过与业务部门沟通，明确了指标定义：利润=销售收入-销售成本-营业外支出。其中，销售收入=客单价×用户数，这里的用户数是指进入店铺的人数。销售成本主要包括为实现销售而支出的相关费用和商品采购成本，也就是销售成本=销售费用+商品采购成本。营业外支出是指各项非营业性支出，例如罚款支出、捐赠支出等。

此外，当要解决的问题中有提到“高、低、大、小”等字眼，要问清楚和谁比。这里就用

到了前面章节的对比分析方法。为什么要找到比较对象呢？问题本质上是指现状与理想状态之间的差距。因此不知道理想的状态是什么，就无法比较，从而就无法确定问题。所以，在定义问题时，要去弄清楚“比较对象”是谁（图3-6）。

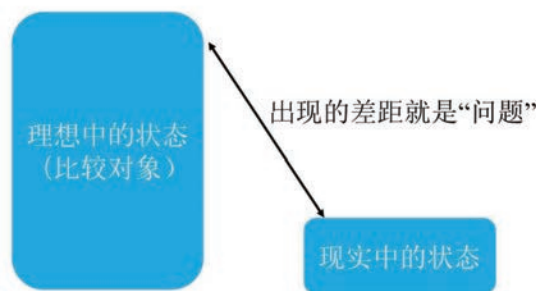


图3-6 弄清楚“比较对象”是谁

在上述的案例中，“与上半年的计划利润相比，实际完成利润落后于计划”，这里的比较对象是“上半年的计划利润”。

3.2.3 总结

这一节学习了“用数据分析解决问题”的第1步：明确问题，主要包含以下内容：

- (1) 明确问题的常见错误。在定义问题时，注意不要加入分析者的“主观猜测”，导致无法分析其他可能的原因。
- (2) 如何明确问题。可以用图3-5所示的方法。

3.3 如何分析原因？

这一节我们来学习“数据分析解决问题”的第2步：分析原因。

影响问题的原因可能有很多种，如果把所有原因都分析一遍，那么这个工作量是非常大的。根据“计算的平方律”定律，当一个问题要考虑的因素的复杂度变为原来的2倍，那么解决问题需要的时间就会变成原来的4倍。所以，在分析原因的过程中，要优先分析那些关键的因素。

例如，你早上着急上班出门，突然发现钥匙不见了。现在的问题是：钥匙在哪里？你想了下钥匙可能在这些地方（问题发生的原因）：衣服、卧室、厨房、卫生间、客厅等。你不会把所有地方都翻一遍，因为找这么多地方，工作量很大，这么大工作量还不如换一把锁划算。那怎么办呢？你会先想：昨天钥匙可能放到哪里了？这种在大量可能性中，通过快速判断最可能的做法，就是优先分析关键因素。

如何知道哪些是“关键的因素”呢？可以在分析的过程中使用这3步来分析（图3-7）：

- (1) 使用“多维度拆解分析方法”对问题进行拆解，将一个复杂问题细化成各个子问题；

(2) 对拆解的每个部分，使用“假设检验分析方法”找到哪里出了问题。分析的过程可以用“对比分析方法”等多个分析方法来辅助完成；

(3) 在找到哪里出了问题以后，可以多问自己“为什么出现了这个问题”，然后使用“相关分析方法”进行深入分析。

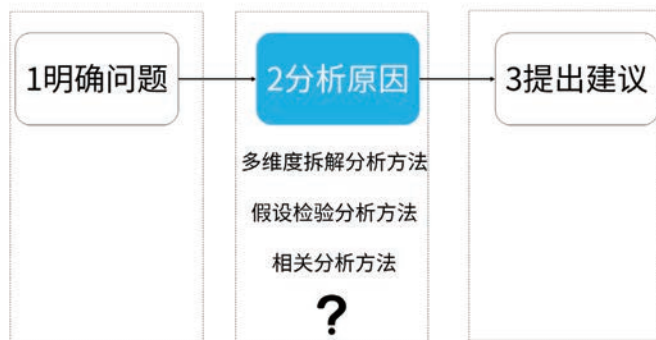


图3-7 分析原因

现在，通过本章的案例来看下如何分析原因。

在数据分析的第1步明确了要解决的问题是：上半年利润落后于计划的原因是什么？公司如何在下半年完成4008.63万元的利润目标？需要先找到上半年利润落后于计划的原因，才能针对具体的原因制定出详细的改进建议，从而保证在下半年完成目标。

为了找到“哪里出了问题”，可以使用多维度拆解分析方法对“利润”这个指标，按指标定义进行拆解，如图3-8所示。

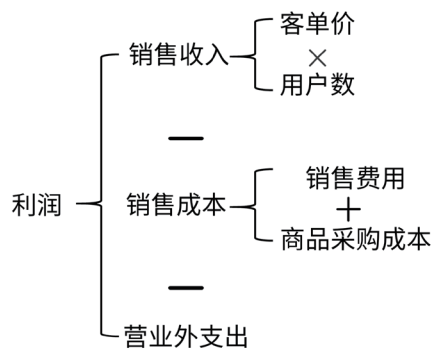


图3-8 对“利润”进行拆解

至于拆解到什么程度，没有统一的标准，要根据对业务的理解和实际问题灵活把握。接下来使用假设检验分析方法对“多维度拆解分析方法”里面的每个部分进行验证。

1) 营业外支出

提出假设：营业外支出增加导致上半年利润落后于计划，如图3-9所示。

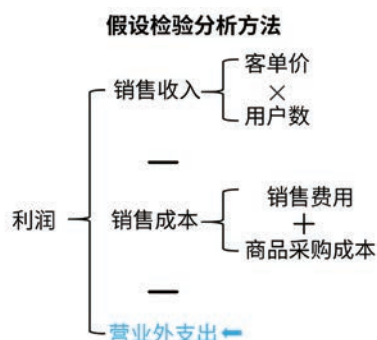


图3-9 考查营业外支出

收集数据：营业外支出汇总如表3-2所示。

表3-2 上半年的营业外支出

| | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 |
|---------------|----|----|----|----|----|----|
| 营业外支出 (万元) | 0 | 0 | 0 | 0 | 0 | 0 |

得出结论：从表中可以发现，上半年未发生营业外支出，因此可以不用考虑该因素对利润的影响，所以假设“营业外支出增加导致上半年利润落后于计划”不成立。

2) 销售成本

提出假设：销售成本提高导致上半年利润落后于计划，如图3-10所示。

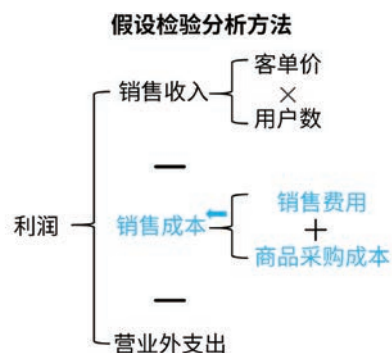


图3-10 考查销售成本

收集数据：将今年上半年和去年上半年销售成本进行对比分析，如表3-3所示。

表3-3 上半年的销售成本对比

| | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 总销售 | 总销售同比 |
|----------------|--------|---------|---------|---------|---------|---------|----------|-------|
| 今年销售成本 (万元) | 423.92 | 853.48 | 1910.94 | 2451.22 | 3323.67 | 4160.64 | 13123.87 | 3.08% |
| 去年销售成本 (万元) | 464.09 | 1052.34 | 2022.41 | 2409.86 | 2832.19 | 3938.14 | 12719.03 | |

表3-3中的总销售同比是（今年上半年的总销售成本13123.87万元-去年上半年的总销售成本12719.03万元）/今年上半年的总销售成本13123.87万元=3.08%，说明今年上半年的总销售成本同比去年是增加了3.08%。

为方便观察数据随时间发生了哪些变化，将表3-3中的数据绘制成图3-11。

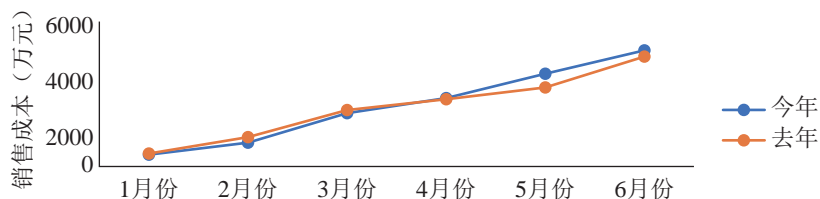


图3-11 上半年销售成本对比

从图3-11中可以看出，对比去年上半年，今年上半年的销售成本出现了明显的增长。是什么原因造成销售成本增长呢？

接下来对“销售成本”的组成部分进行分析。因为销售成本=销售费用+商品采购成本，所以分别来分析销售费用和商品采购成本的情况。

（1）分析销售费用。

收集今年上半年的销售费用数据，如表3-4所示。

表3-4 上半年的销售费用

| | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 |
|-----------|------|--------|--------|--------|-------|--------|
| 销售费用 (万元) | 46.8 | 111.19 | 206.52 | 259.66 | 330.2 | 399.22 |

为了方便观察数据随时间发生了哪些变化，将表3-4中的数据绘制成图3-12。

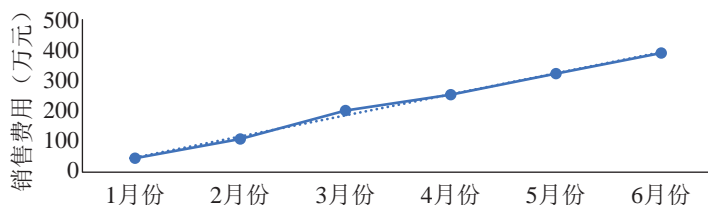


图3-12 上半年的销售费用

从图中可以看出，上半年的销售费用是逐步上升的，这会不会就是造成利润减少的原因呢？

答案是不一定。只单独分析出销售费用是增长的，还不能判断它就是造成利润落后于计划的根本原因。因为销售费用是用来提升销售增长的，如果增加销售费用能带来销售收入的增长，那也是合理的。所以，需要将销售费用和销售收入结合起来加以分析，这就需要用到一个指标：费率比，费率比=投入的费用/产生的销售收入。费率比越低，说明通过投入的费用带动销售增长的效果越好。

将上半年的销售费用和产生的销售收入按月份排列，并计算出费率比，如表3-5所示。

表3-5 上半年的费率比

| | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 |
|--------|--------|--------|---------|---------|---------|---------|
| 费用(万元) | 46.8 | 111.19 | 206.52 | 259.66 | 330.2 | 399.22 |
| 销售(万元) | 470.52 | 942.73 | 2197.81 | 2829.91 | 3843.75 | 4830.52 |
| 费率比 | 9.95% | 11.79% | 9.40% | 9.18% | 8.59% | 8.26% |

对数据进行可视化设计来观察变化，如图3-13所示。

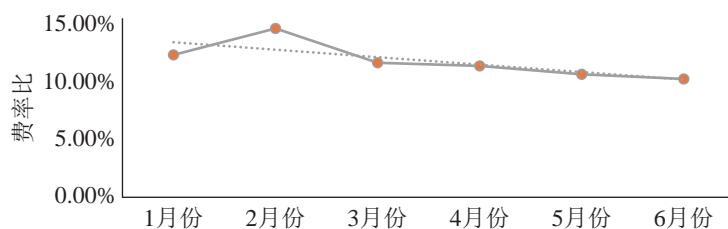


图3-13 上半年的费率比

从图中可以发现，上半年的费率比是逐步下降的，也就是说，销售费用的增长也提升了销售收入的增长。所以销售费用的上升并不是利润不达标的主要原因。

(2) 分析商品采购成本。

将今年上半年和去年上半年的商品采购成本进行对比分析，如表3-6所示。

表3-6 上半年的商品采购成本对比

| | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 总计 | 平均 |
|--------------|--------|---------|---------|---------|---------|---------|----------|---------|
| 今年商品采购成本(万元) | 377.12 | 742.29 | 1704.42 | 2191.56 | 2993.47 | 3761.42 | 11770.28 | 1961.71 |
| 去年商品采购成本(万元) | 452.72 | 1031.96 | 1988.89 | 2370.05 | 2779.17 | 3872.08 | 12494.87 | 2082.48 |

从表中可以发现，今年上半年的平均采购成本（1961.71万元）低于去年上半年的平均采购成本（2082.48万元）。由此可以判断，商品采购成本不是影响利润不达标的主要原因。

通过多维度拆解分析方法，对销售费用和采购成本分析后，可以判断假设“销售成本增加导致利润水平下降”不成立。

3) 销售收入

提出假设：销售收入减少导致上半年利润落后于计划，如图3-14所示。

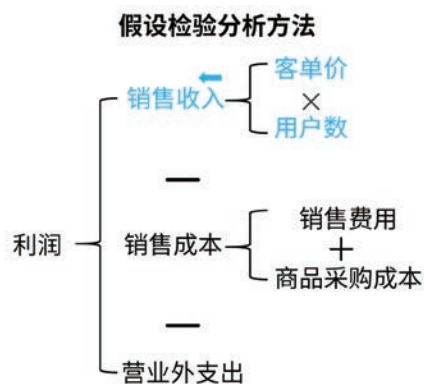


图3-14 考查销售收入

收集数据：将今年上半年和去年上半年的数据进行比较分析，如表3-7所示。

表3-7 上半年的销售收入对比

| | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 总销售 | 总销售同比 |
|--------------------|--------|---------|---------|---------|---------|---------|----------|--------|
| 今年 销售收入 (万元) | 470.52 | 942.73 | 2197.81 | 2829.91 | 3843.75 | 4830.52 | 15115.24 | -8.79% |
| 去年 销售收入 (万元) | 587.56 | 1336.34 | 2593.69 | 3093.48 | 3731.26 | 5100.84 | 16443.17 | |

得出结论：表3-7中的总销售同比是（今年上半年的总销售收入15115.24万元-去年上半年的总销售收入16443.17万元）/今年上半年的总销售收入15115.24万元=-8.79%，今年上半年的总销售收入同比去年下降了8.79%，说明销售收入减少是导致上半年利润落后于计划的主要原因之一。

是什么原因造成销售收入下降的呢？接下来对“销售收入”的组成部分进行分析。因为销售收入=客单价×用户数，分别来分析客单价和用户数的情况。

（1）分析客单价。

将今年上半年和去年上半年的客单价进行对比分析，如表3-8所示。

表3-8 上半年的客单价

| | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 平均客单价 |
|------------------|------|------|------|------|------|------|-------|
| 今年 客单价 (元) | 2394 | 2335 | 3980 | 2856 | 3533 | 3180 | 3026 |
| 去年 客单价 (元) | 2889 | 2644 | 3211 | 2634 | 2568 | 3117 | 2844 |

从表3-8中可以发现，今年上半年的平均客单价是3026元（平均客单价=1—6月每个月客单价之和/6），去年上半年的平均客单价为2844元，说明今年上半年的平均客单价是高于去年的，因此客单价并不是影响销售收入下降的主要因素。

(2) 分析用户数。

将今年上半年和去年上半年的用户数进行对比分析，如表3-9所示。

表3-9 上半年的用户数

| | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 平均用户数 | 平均用户数同比 |
|-------|------|------|------|-------|-------|-------|-------|---------|
| 今年用户数 | 1965 | 4037 | 5522 | 9909 | 10880 | 15190 | 7917 | -1717 |
| 去年用户数 | 2034 | 5054 | 8078 | 11744 | 14530 | 16365 | 9634 | |

从表3-9中可以发现，今年上半年的平均用户数是7917人（平均用户数=1—6月每个月用户数之和/6），去年上半年的平均用户数是9634人，今年上半年的平均用户数是低于去年上半年的，因此用户数减少是销售收入下降的主要原因。

但是如果只分析到这里，这样的分析结果无法产生实际的意义。因为只看用户数下降，还不能决定“接下来要采取哪些具体的措施才能解决问题”。所以，要接着分析导致“用户数”下降的原因是什么（图3-15）。

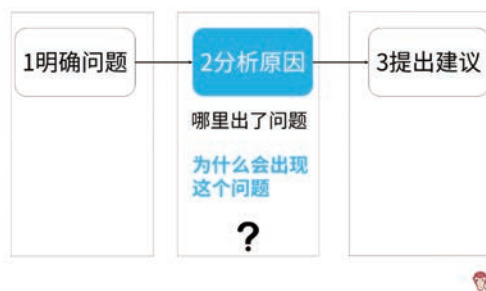


图3-15 分析原因

现在把“用户数”按用户、产品、竞品这3个维度进行拆解。根据具体问题可以决定优先展开分析哪个维度来找到问题发生的原因。对于本章的案例，用户数下降是因为用户跑到其他竞争对手购买了，所以可以从“竞品维度”来比较用户数。通过和竞争对手比较发现，公司的用户数比同行竞争对手低很多（图3-16）。

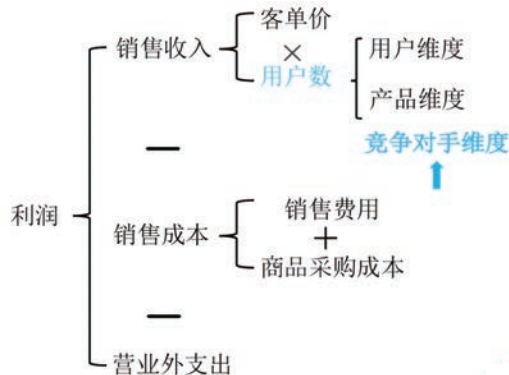


图3-16 考查竞争对手维度

为了进一步找到用户数低的原因，可以梳理本店铺和竞争对手店铺的业务流程，方便从业务流程提出假设（图3-17）：

第1步：用户在店铺中选择自己想要购买的商品；

第2步：用户比较不同店铺中商品的价格；

第3步：用户付款购买后，售后部门安排送货上门。

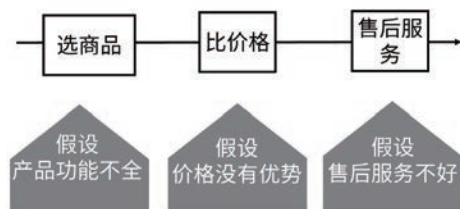


图3-17 从业务流程提出假设

对于第1步，可以提出假设：供用户选择的商品不全，用户找不到想购买的商品。

对于第2步，可以提出假设：价格相对于竞争对手没有优势。

对于第3步，可以提出假设：售后服务不能让用户满意。

可以和竞争对手的数据进行对比，例如通过问卷调查或者电话访谈，来分析本店铺和竞争对手有哪些差异。整个分析思路如图3-18所示。

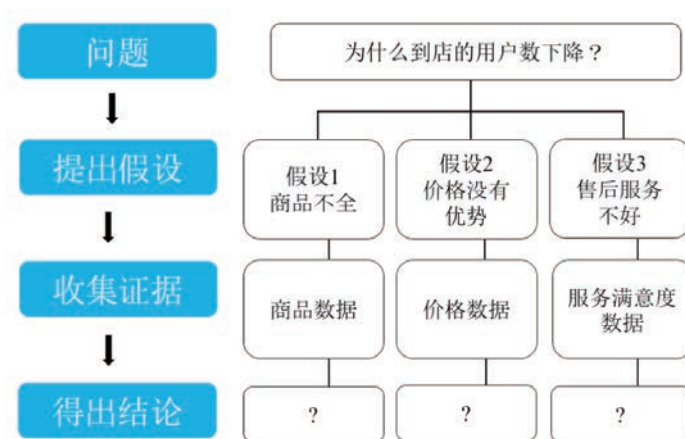


图3-18 分析思路

先来看假设1：供用户选择的商品不全，用户找不到想购买的商品。

通过对比发现，本店铺的商品种类并不比竞争对手少。只是店铺为了处理功能上相对欠缺的老产品，没有及时展示出新产品型号，导致用户不能及时直观地了解新产品。

所以假设1不成立，用户数下降不是因为店铺商品不全，而是因为具有新功能的产品没有被及时展示在店铺中，导致追求新功能的那一部分用户转去购买竞争对手的产品（图3-19）。

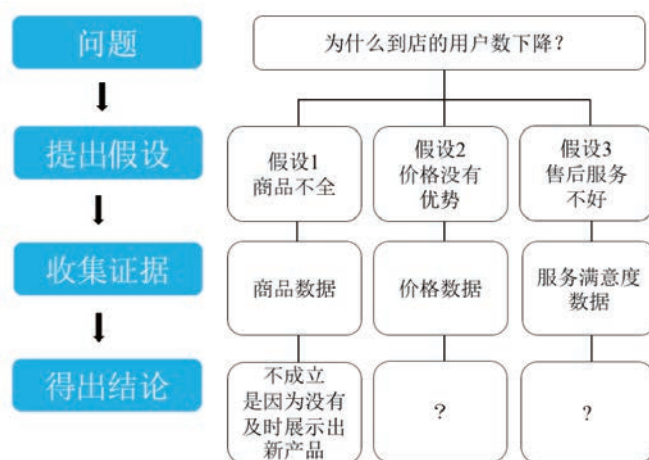


图3-19 假设1不成立

再来看第2个假设：价格相对于竞争对手没有优势。

通过市场调查发现，本店铺在价格定位上和竞争对手相同。但是，上半年本店铺共计开展过6场促销活动，而竞争对手共计开展过10场促销活动。部分用户通过市场比价，被竞争对手促销的优惠价格吸引，导致本店铺的用户流失。

所以假设2成立，在相同的商品定价基础上，本店铺的促销活动频次少于竞争对手，导致用户数下降（图3-20）。

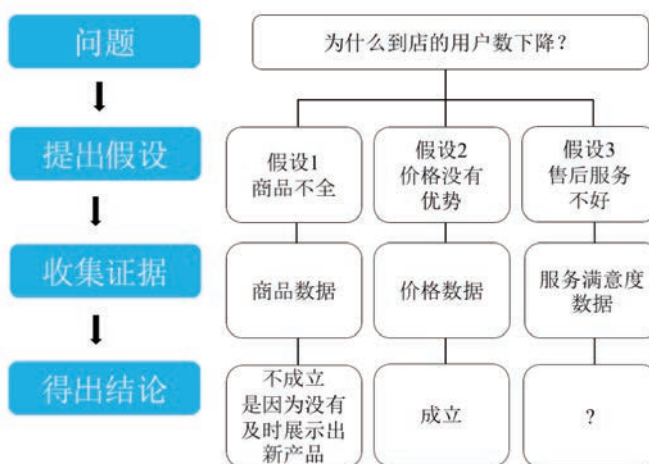


图3-20 假设2成立

再来看下假设3：售后服务过程不能让用户满意。

本店铺的商品在用户购买后，需要为用户提供送货上门等售后服务。如果在售后服务的过程中给用户带来不好的体验，会降低品牌良好口碑的传播，降低用户的复购率。

通过用户回访，请他们对售后服务过程的质量进行打分，满分为100分。分值越高，说明用户对服务的过程越满意。上半年的用户服务满意度如表3-10所示（用户满意度分值为每个月的平均分）。

表3-10 上半年的用户服务满意度

| | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 |
|-------|----|----|----|----|----|----|
| 满意度打分 | 89 | 81 | 73 | 65 | 59 | 47 |

为方便观察数据随时间发生了哪些变化，将表格中的数据绘制成图3-21。

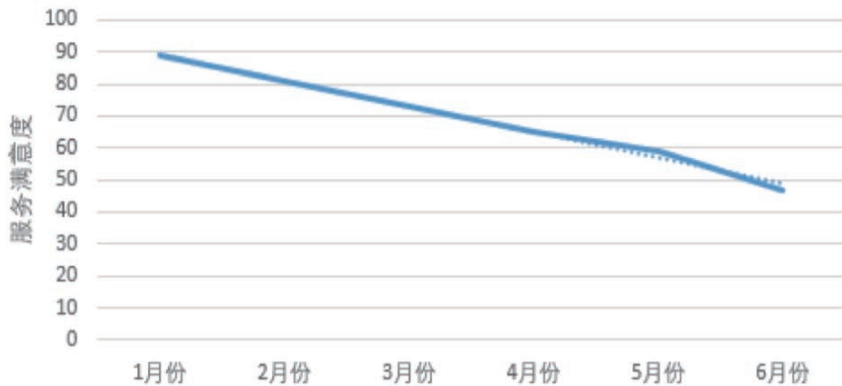


图3-21 用户服务满意度

从图3-21中可以发现，上半年的满意度数值有下降趋势，可以验证本店铺售后服务的水平也在下降。服务质量的下降会造成用户数下降，所以假设3也是成立的。

通过以上分析，上半年的用户数下降，是因为：

- (1) 具有新功能的产品没有被店铺及时展出，导致追求新功能的用户转向去购买竞争对手的产品；
- (2) 本店铺的促销活动少于竞争对手，导致用户数下降；
- (3) 服务质量水平的下降造成了用户数下降。

原因1可以快速解决，店铺及时展出新品就可以。原因2和原因3不是立马能解决的。公司资源有限，一个阶段只能集中解决一个问题，这两个因素中，哪个对用户数下降影响更大呢？

这时候可以通过相关分析，来确定影响更大的因素。前文介绍了一种应用相关分析的模式（图3-22），就是使用相关分析来找出哪些因素与分析目标有相关关系。

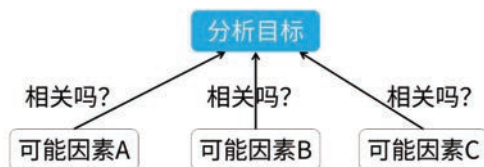


图3-22 相关分析模式

我们可以分别计算商品价格、用户满意度和用户数的相关系数，观察这些相关系数的大小，可以得知哪个因素对用户数影响更大。

用户数和商品价格比、用户满意度的数据如表3-11所示。其中，商品价格比是用本店铺的平均价格和市场竞争对手的平均价格相比得到，例如1月本店铺的平均价格为100元，竞争对手的平均价

格为434.78元，那么1月商品价格比为：本店铺平均价格（100元）/竞争对手平均价格（434.78元） ≈ 0.23 。该指标主要是反映商品定价相对于竞争对手是偏高还是偏低，分值越大说明定价越高。

表3-11 用户数据

| 月份 | 用户数 | 商品价格比 | 用户满意度（分数） |
|-----------|-------|-------|-----------|
| 1月 | 1965 | 0.23 | 89 |
| 2月 | 4037 | 0.83 | 81 |
| 3月 | 5522 | 1.12 | 73 |
| 4月 | 9909 | 0.97 | 65 |
| 5月 | 10880 | 1.15 | 59 |
| 6月 | 15190 | 1.23 | 47 |
| 和用户数的相关系数 | | 0.773 | 0.992 |

对数据进行相关分析，可以得出用户数和商品价格比的相关系数是0.773，用户数和用户满意度的相关系数是0.992。

通过相关分析，发现跟用户数相关度最高的因素是用户满意度，所以优先提升服务满意度，即提升店铺的售后服务水平，可以显著提高用户数。

现在复盘下这个案例是如何分析的。可以用图3-23把前面的分析组织起来，这样和其他人沟通的时候，对方就知道你是如何对问题展开分析的。

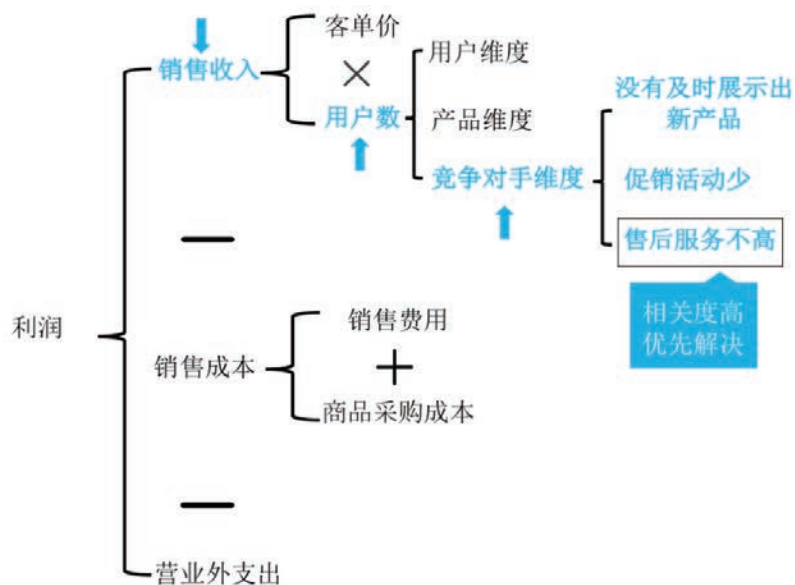


图3-23 案例分析思路

3.4 如何提出建议？

这一节我们来学习“如何用数据分析解决问题”的第3步：提出建议（图3-24）。也就是根

据第2步找到的原因提出建议。在提出建议这一步，常用的分析方法是回归分析或者AARRR模型分析方法。

如何决策？

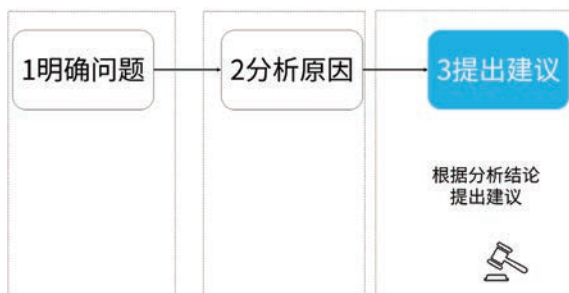


图3-24 提出建议

下面我们先学习什么是回归分析，然后通过本章案例来看如何应用。

3.4.1 回归分析

一般问题发生的原因有很多种，所以第2步找到的原因不会只有一种。这些原因和目标都相关，如果在第3步提出建议阶段，只是说a和b高度相关，那么，你的领导是无法根据“相关”采取具体行动的。

在本章的案例中，虽然找到了利润没达到目标的原因，却不知道要改善到什么程度才能实现下半年的业务目标。这时候就需要“回归分析”来计算出某个原因能够对目标造成“多大程度”的影响。可以计算出，利润和销售收入的相关系数是0.999，说明这两种数据之间是高度相关关系（表3-12）。

表3-12 销售数据

| 月份 | 利润（万元） | 销售收入（万元） |
|------|--------|----------|
| 1月 | 46.6 | 470.52 |
| 2月 | 89.25 | 942.73 |
| 3月 | 286.87 | 2197.81 |
| 4月 | 378.69 | 2829.91 |
| 5月 | 520.08 | 3843.75 |
| 6月 | 669.88 | 4830.52 |
| 相关系数 | 0.999 | |

根据数据，绘制出利润和销售收入的相关关系散点图，如图3-25所示。

现在想要通过销售收入来预测利润，也就是，知道了销售收入要达到多少以后，就可以实现下半年的利润4008.63万元。那么如何预测呢？能想到的办法是，在散点图上画一条穿过这些点的直线，使这条直线尽量接近各个数据点，这样的直线叫作最佳拟合线，如图3-26所示。

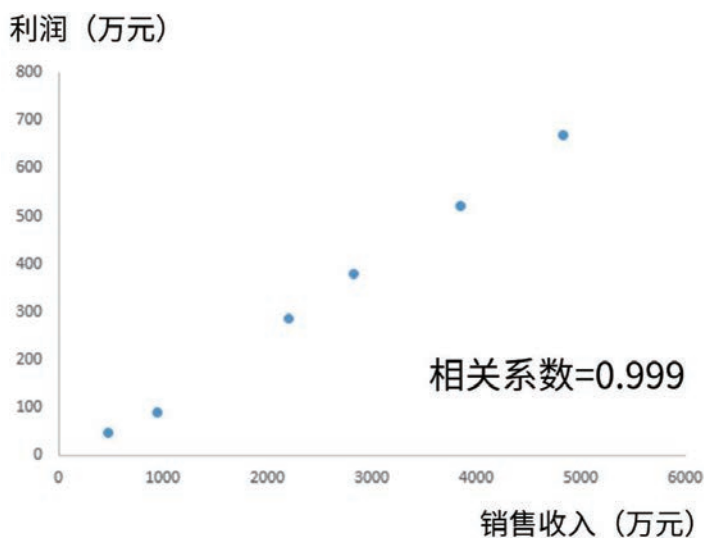


图3-25 散点图

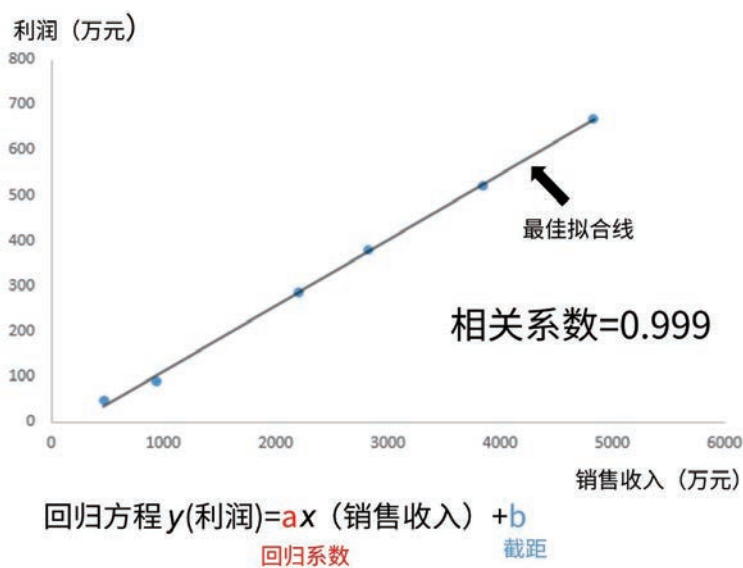


图3-26 最佳拟合线

你无法使得这条直线穿过每一个点，但是，如果两种数据存在相关关系，那么应该可以保证每一个点合理地接近你所绘制的直线。这样一来，就可以根据销售收入预测出利润的值。

如果你还记得高中几何课程的话，一定能回想起一个直线方程，也就是 $y=ax+b$ ，这个方程叫作回归方程，对应的这个直线叫作回归线。 a 叫作回归系数， b 叫作截距。别被这些高大上的名字吓到，它就是高中见过的直线方程。在这个例子中， x 表示销售收入， y 表示利润。接下来的问题就是，如何求出 a 和 b 的值。可以使用Excel进行回归分析，得到回归方程里面 a 和 b 的值，具体操作如下（图3-27）：

第1步，单击“数据”选项卡下的“数据分析”功能；

第2步，选择“回归”后单击“确定”按钮，就会跳出回归分析的对话框。

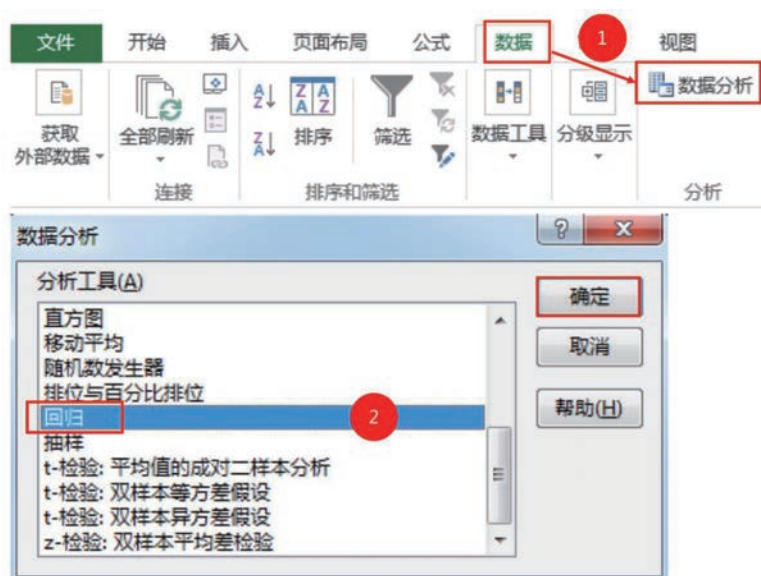


图3-27 使用Excel进行回归分析

跳出回归分析的对话框之后的操作步骤如下（图3-28）：

第1步，在“X值输入区域”和“Y值输入区域”选择对应的数据。这里需要注意，一般用横轴的X值表示“输入”（能够控制的数据，也叫作自变量），用纵轴的Y值表示“输出”（预测结果，也叫作因变量）。因为是要通过销售收入来预测出利润，所以，“X值输入区域”选择“销售收入”这一列，“Y值输入区域”选择“利润”这一列。注意，在选择数据的时候不能包括列名，因为列名不是数；

第2步，勾选“线性拟合图”；

第3步，单击“确定”按钮。

回归方程 $y(\text{利润})=ax(\text{销售收入})+b$



图3-28 回归分析的对话框

这样就得到了图3-29的散点图。选择散点图上任意一个点，右击，在弹出的快捷菜单中选择“添加趋势线”选项。

选择任意一个点
鼠标右键

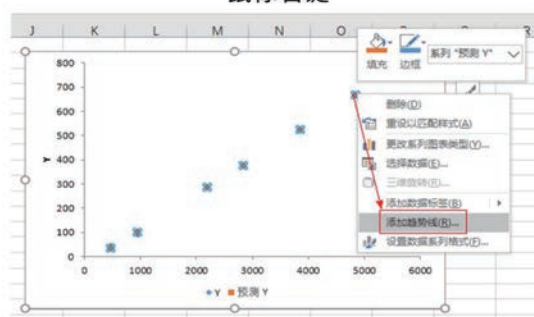


图3-29 散点图

这样就得到了最佳拟合线，然后在出现的页面中勾选最下面的“显示公式”，就会在散点图上显示回归方程（图3-30）。可以把横轴和纵轴修改成自己想要的名称。如何理解这个回归方程呢？

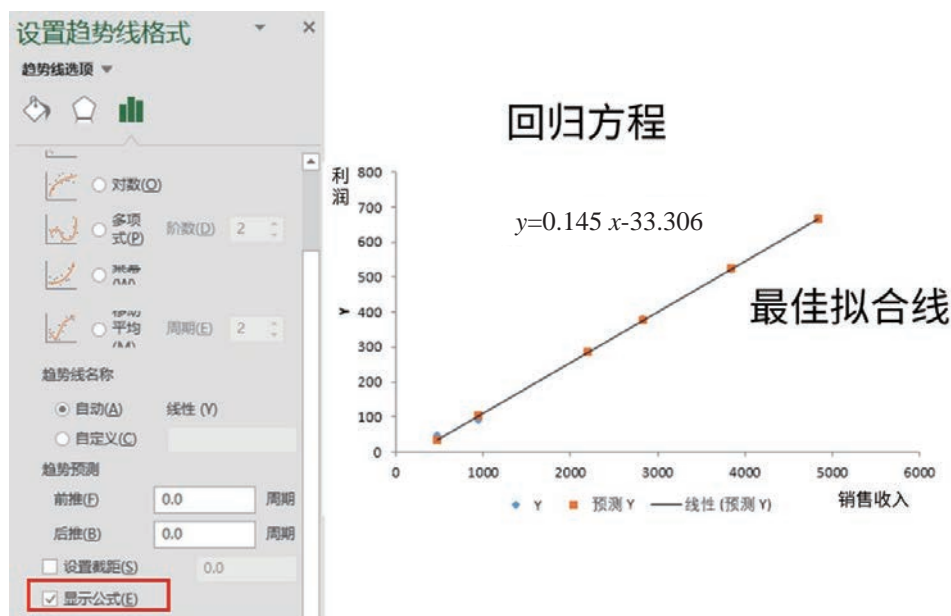


图3-30 显示公式

这个回归方程可以理解为： y （利润）= $0.145x$ （销售收入）-33.306。前面说到，下半年需要完成利润是4008.63万元，也就是 y （利润）=4008.63万元。代入回归方程中就可以算出 x （销售收入）=27875.42万元。也就是说，根据公司目标，要实现 y （利润）=4008.63万元，需要将销售收入提升到27875.42万元（图3-31）。

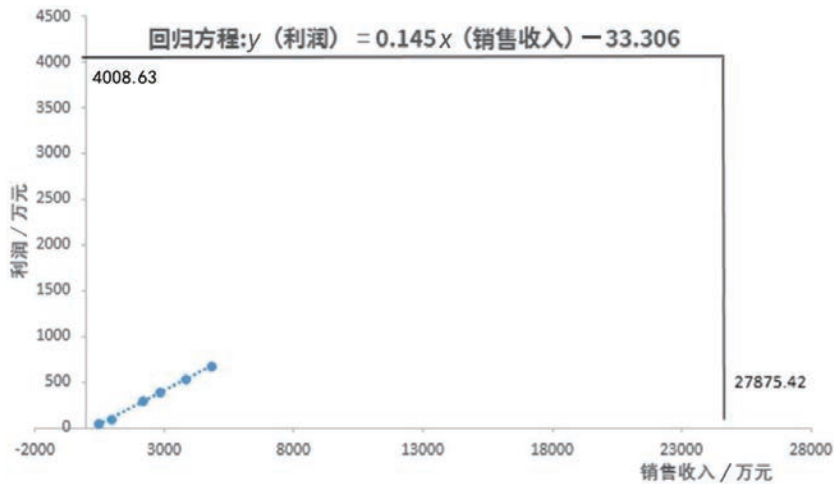


图3-31 预测销售收入

使用回归分析需要注意：

(1) 回归分析有很多类型，前面案例里的回归方程只有一种自变量，这种回归分析叫作一元线性回归分析。相比于其他类型的回归分析，一元线性回归分析不仅简单，而且分析结果也容易被非专业的人理解。

(2) 一元线性回归分析的前提是两种数据之间要有相关关系。所以，要先判断两种数据有相关关系，才能使用一元线性回归分析。

(3) 合理分配资源。前面的案例是知道了 y 的值，想知道 x 的值是多少。还有一种情况是，知道了 x 的值，想知道 y 的值是多少。例如 x 是投入广告的费用， y 是产生的收益，这样在推广前就可以知道，投入的成本（ x ，广告费用）能预期产生多少收益（ y ，产生的收益）。当决策者有多种推广方案要选择的时候，就可以根据回归分析，知道把有限的资源投入到哪里才能发挥出最好的效果。

3.4.2 回归分析应用

前面通过回归分析，知道了根据公司目标，下半年要实现利润4008.63万元，需要将销售收入提升到27875.42万元。接下来需要考虑，采取哪些措施才能将销售收入提升到27875.42万元。

在第2部分的分析原因，知道了销售收入下降是因为用户数减少。接下来用回归分析来看看销售收入和用户数的关系（图3-32）。



图3-32 分析用户数

通过对销售收入和用户数进行相关分析，发现两种数据的相关系数是0.980，具有高度相关关系，可以使用一元线性回归分析（表3-13）。

表3-13 销售收入与用户数

| 月份 | 销售收入 (万元) | 用户数 |
|------|-----------|-------|
| 1月 | 470.52 | 1965 |
| 2月 | 942.73 | 4037 |
| 3月 | 2197.81 | 5522 |
| 4月 | 2829.91 | 9909 |
| 5月 | 3843.75 | 10880 |
| 6月 | 4830.52 | 15190 |
| 相关系数 | 0.980 | |

通过回归分析, 得到回归方程是: $y = 0.3318x - 107.85$ (图3-33)。该公式可以理解为: y (销售收入) $= 0.3318 \times x$ (用户数) $- 107.85$ 。

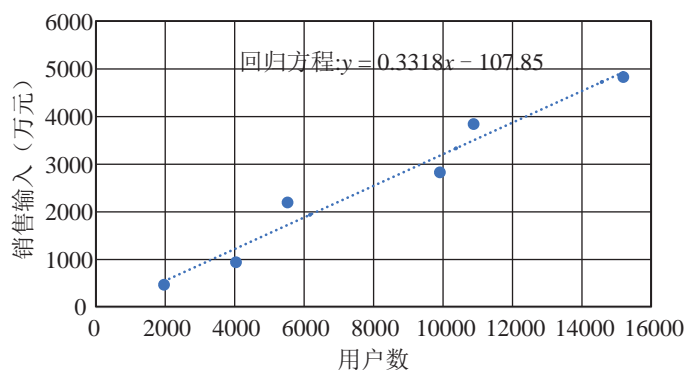


图3-33 散点图与回归方程

前面知道了销售收入要提升到27875.42万元, 也就是 y (销售收入) $= 27875.42$ 万元, 把它代入回归方程可以计算出 x (用户数) $= 84338$ 人。也就是说, 要实现销售收入为27875.42万元的目标, 需要将到店的用户数提升到84338人 (图3-34)。

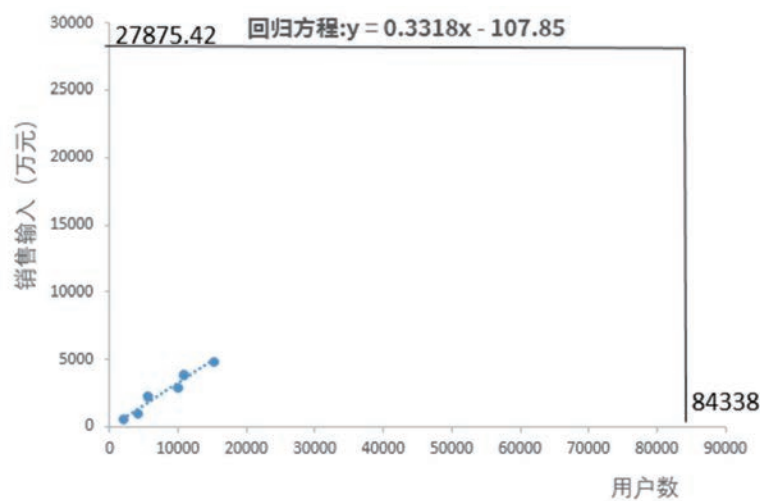


图3-34 预测用户数

通过分析，为公司下半年的经营提出以下建议：

如果想要完成年度的6000万元利润目标，建议在保持目前商品采购成本不变的前提下，通过增加销售收入的方式来保证利润目标。下半年需要将销售收入目标设定为27875.42万元，平均每个月销售目标设定为4546.90万元（27875.42万元/6）。下半年需要吸引84338人到店消费，平均每个月到店的用户数是14057人（84338人/6）。根据相关分析方法，要想提高到店用户数，需要优先提升店铺的售后服务水平。

建议可将此作为关键绩效指标，细化分解到具体的部门去执行，并与日常考核联系起来，这样才能确保全年利润目标顺利完成（表3-14）。

表3-14 措施与目标

| 措施 | 目标 |
|----------------|--------------------------------------------|
| 将月度销售目标分解到个人 | 将4546.90万元的月度目标分解到400个员工，人均11.61万元/月 |
| 将月度销售目标分解到各个分店 | 将4546.90万元的月度目标分解到20家分店，每家店销售目标是227.35万元/月 |
| 制定个人销售激励办法 | 对于完成月度个人目标，售后服务水平达标的，给予100元的奖励 |
| 制定集体销售激励办法 | 对于完成月度目标的店长，售后服务水平达标的，给予1%的奖励 |

3.5 总结

数据分析解决问题的过程如图3-35所示。

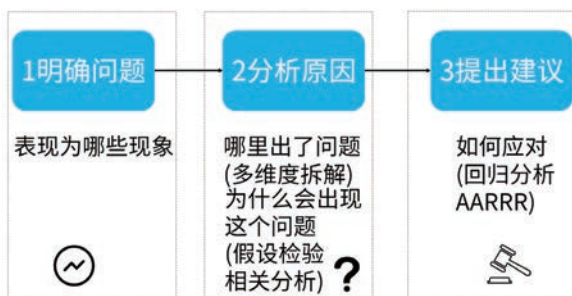


图3-35 数据分析解决问题的过程

第1步：明确问题。

通过观察现象，把问题定义清楚，这是数据分析的第1步。只有明确了问题，才能围绕这个问题展开后面的分析。如果一开始问题就定义错了，那再怎么分析，也是白费时间。

第2步：分析原因。

这一步是分析问题发生的原因，可以通过下面两个问题把原因搞清楚：①哪里出了问题？②为什么会出现这个问题？具体分析步骤如下：

（1）使用“多维度拆解分析方法”对问题进行拆解，将一个复杂问题细化成各个子问题；

（2）对拆解的每个部分，使用“假设检验分析方法”找到哪里出了问题。分析的过程可以用“对比分析方法”等多个分析方法来辅助完成；

(3) 在找到哪里出了问题以后，可以多问自己“为什么出现了这个问题”，然后使用“相关分析方法”进行深入分析。

第3步：提出建议。

找到原因就完事了吗？还不行，要找到对应的办法才是分析的终点。所以，找到原因以后，还要针对原因给出建议，或者提出可以实施的解决方案。在决策这一步，常用的分析方法是回归分析或者AARRR分析。需要注意的是：

(1) 做决策的选项不能太多。太多的选项不仅会增加决策的成本，还会让人迷失，无从下手。相对简单的问题，需要4个选项左右；相对复杂的问题，需要4~7个选项；

(2) 决策要是可以落地的具体措施，这样决策者才能根据措施，合理安排资源，把措施变成行动。

本章作者介绍

猴子，中国科学院大学硕士，“猴子·数据分析学院”创始人，公众号“猴子数据分析”创始人，前IBM工程师。其“分析方法”课程入围知乎年度口碑榜TOP 10，首创的“闯关游戏学习数据分析模式”深受用户喜欢。