

打造数据驱动的组织：第二年
2015 年 4 月 6 号

原文链接：http://www.p-value.info/2015/04/creating-data-driven-organization-two_6.html

这是我在 Warby Parker 打造数据驱动的组织过程中写的第三篇文章。[第一篇](#)讲的是我在 2013 年作为首席数据科学家刚加入公司的一些初步想法，[第二篇](#)讲的是第一年的进展，这一篇则会详细说明第二年的情况，这一年里面我们取得了相当大的进展。

数据字典（Data Dictionary）的力量

我们从去年秋天开始放弃了 Tableau，转而使用商业智能工具 Looker。我们的分析师对此赞不绝口，这使我下定决心专注于 Looker，仅使用 Tableau 作为给高层管理者提供报表的工具。

我们在前六个月并没有全面铺开 Looker，不过这并不是因为将 Looker 整合进系统太复杂，而是我们的数据源实在太繁杂。如果我们想要可信的数据，我们就必须和经销商们达成一致并确定好数据字典的口径，包括在整个商业链条中各个术语的规定，包括对客户各个属性的精确定义，包括销售渠道的梳理等等。

要打造数据驱动的文化和管理，我们就需要像这样对每个名词做好解释，并在整个公司内达成共识，下面是一些实施的细节。

设立并验证数据字典

计划的第一步是每次专注于一项数据来源，并打通数据上下游的流动。在上游，有“掌握”着商业逻辑的部门，他们定义了公司运营中的各类术语，在下游，则是每天产生大量数据的业务部门，这些数据可以用于交叉验证。这两个部门有着从我们其他业务系统导出的包含原始数据的报表，更重要的是，他们知道如何度量不同的数据，知道数据的组成。比如说，在一次促销行为中，就包含了礼品卡，抵价优惠，赠品等诸多组成部分，这些知识由业务部门告诉我们，然后我们将其导入 Looker 产生数据集，并通过 Looker 的计算对原始报表的每一行进行验证。这个过程产生了非常有趣的结果并对组织产生了巨大的影响。

我们从中学到了四点经验，且让我一一道来。首先，业务部门提供的计算逻辑得出的结果与报表上的数字有时并不一致。原因在于，这些报表在经年累月的时间里是由不同的人打理的，并且报表中总有各类特殊情况（而未标注）。正因为要从报表中找出计算逻辑如此之难，我们转而建议业务部门按照 If...Then 的陈述方法梳理其逻辑，在这个过程中我们双方都收获颇丰。

第二点，有时候在不同的部门之间对同一数据有不一样的度量。虽然报表上看似名称一样，但是数据实际上不是同步的。同样地，找出这些不一致以后让双方讨论究竟应该选择哪种度量方式，甚至哪种情况下应该使用不同的名称，这也会是一次非常有意义的对话。再举个例子，我们的财务部门和产品战略部门对“预售数量”有着不同的理

解，经过对话之后，我们有了“预售数量”和“产品预售数量”两个定义清晰的名称，现在公司内部也不会因为这个词而产生误解。

第三点，因为我们使用 SQL 进行验证，我们很容易找出那些不符合条件的行的根本原因。我们标记出了那些业务部门且尚未考虑到的特殊情况，比如如何处理分拆订单。当我们向管理层解释那些特殊情况时，他们的反应是“这种情况不可能发生！”，但是当看到这些事实以后，我们得以对此进行内部的改进并在以后做好准备。

最后一个经验是，我们在那些 Excel 表格中发现一部分计算逻辑并不是最优化的，而其原因在于企业的 ERP（Enterprise Resource Planning）软件能力有限，而公司只能采取变通的办法来适应 ERP。因此当我们发现这点以后，我们就可以去问管理层：“在理想的状况下，你会想要怎么定义这一逻辑？”在此之后，我们就可以据此对流程进行改进，并使得每个部门都有一个更简单，更清楚，并且更符合逻辑的数据表。

但你可以想象到的另一面是，和不同部门的人一起整理数据来源，将其导入 Looker，验证它（这个过程最花时间），然后再和他们逐一确认，这个过程是相当缓慢而痛苦的。尽管如此，最先开展这一工作的几个部门都认识到了这一过程的重大收获，他们对报表中的名称有了更深的了解，并拥有了一个 Looker 这样可以信赖，将很多事情自动化并处理到位的系统。根据这些部门的积极反馈，我们的 CEO 决定将整个公司的数据都导入 Looker 并进行验证，且开放重要数据以供分析师制作报表并提供建议。

对于这一流程，我再怎么强调其对数据驱动的的管理的重要性也不为过。就算我们明天就停止使用 Looker，我们仍然有我们的数据字典，并且在不同部门之间都保持着一致性。这改变了我们整个公司在数据方面的沟通方式。

之后我们把所有的知识都整理成册并挂在了我们内部的网页上。它有着像书一样的阅读方式（使用 gitbook.io），列举了我们的数据来源，处理数据的方式，以及数据字典。每个人都能根据这本书找到所需的知识并理解其含义。

数据民主

我们现在在 Looker 有了完整的数据集，并且它们能够被轻松的分割，具有高可信度。现在在公司内，许多报表都会自动产生并发送到相关部门去，对于另一部分的报表，我们将其聚集到一定数量并发送以供分析。比如说，Warby Parker 给客户发送客户调查的邮件以得到反馈，我们通过 Looker 就可以将不同客户的反馈以一种一目了然的方式发送到不同的经销商那里。

接下来要做的几件事

接下来我们会很多的培训。我们之前刚刚请了 [dashingd3js](#) 的 Sebastian Gutierrez 给我们十几个分析师做了数据可视化的培训。接下来还会有统计方面的培训，不过这将是给 Warby Parker 的中层管理人员提供的。你可以从我上一篇文章中回忆起来我们对分析师进行统计相关的培训并没有很好的效果，因此这次我希望这个培训能使得管理层对统计有更大的需求，能够看懂报表中更多的数据，从而倒逼分析师们能够在统计方面更下工夫。

此外，我还在完善我的《分析师能力矩阵》，这篇文章会对不同级别的分析师在数据处理，数据分析和数据可视化三个方面提出不同层次的要求，而分析师可以对自己的职业路径有个更好的规划，并根据不同的层次一步步自我提升。我希望这一切能让分析师们对他们的工作更满意，并且工作起来更有效率。

更宽泛的来说，我想要推广更有预见性的分析，包括更多的预测模型甚至对供应链的随机模拟模型。

对数据的需求越来越大

从分析部门层面回到公司层面，我发现我们对数据的胃口越来越大，而这是一个相当有价值的问题。比如说在第一年里，我们的分析师在进行原始的地理分析之后提出需要更多的数据，于是我们提供给他们包括 ZIP codes（美国邮政编码），CBSA（大城市地区分块），DMA（收视率地理分布）等一系列的数据集。结合客户资料和销售数据，我们的分析师们给出了前所未有的详尽而有深度的分析报告。上个礼拜，又有人来向我们索要当地社区级别的数据来分析新开分店对当地已有分店的影响。

几天前，我参加 Warby Parker 的管理层季度会议，其中的主题之一就是更多的数据，更直观的展示以及更数据驱动的管理。我能感觉到整个公司的氛围的变化，并且这一切都是积极而有利的。

和此前一样，欢迎在明年继续来观看我们的进展，再会。