

第 4 章

数据获取与处理

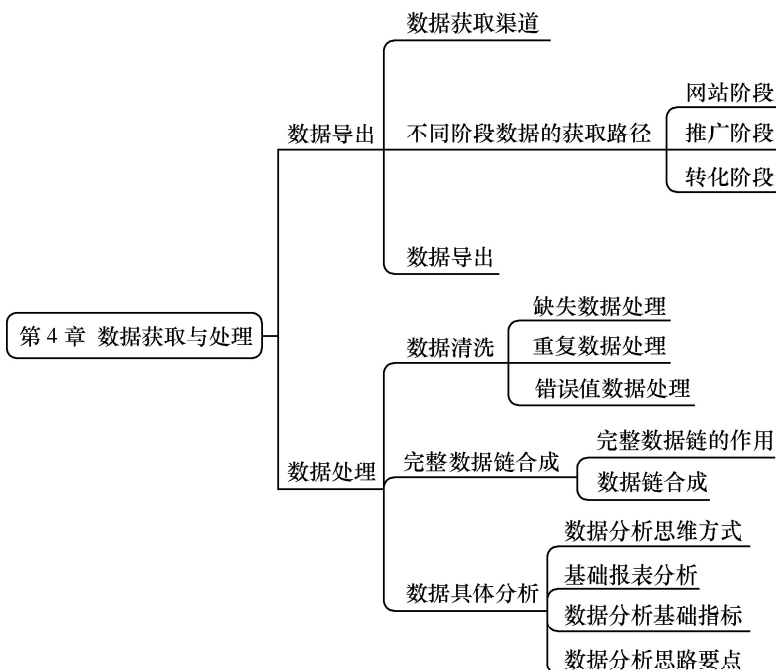
技能目标

- ❖ 掌握 SEM 各阶段数据的导出方法
- ❖ 能够独立完成数据的清洗工作
- ❖ 掌握数据链合成的方法，并对合成后的数据链进行分析

本章导读

随着推广的深入，会产生大量的数据，有咨询工具部分的数据、推广账户后台的数据、网站统计工具的数据，各种数据分布在不同的平台。我们需要把不同阶段的数据导出，并把数据处理成有价值的、可供分析的格式，而不再是零零散散的数据；再根据数据链计算相关量指标和质指标，并用来指导和修正投放方向和投放策略。





4.1 数据导出

数据导出，顾名思义就是把各个平台上提供的数据导出到本地。因为数据来自于不同的平台，所以数据格式难免会有出入，需要对格式不一致的数据进行处理，以符合我们的数据分析要求。

4.1.1 数据获取渠道

数据获取的渠道一共分为三大块，首先是推广阶段的数据，也就是账户后台提供的相关数据；其次是网站统计阶段的数据，也就是百度统计相关报告提供的的数据；最后是在线咨询工具数据。

1. 账户数据报告

数据报告是获取自身账户推广数据最直接的入口，在百度推广后台“数据报告”中提供了多个维度的数据报告可供下载。登录百度推广账户，进入搜索推广版块，点击“推广报告”选项卡或者底部的“数据报告”选项卡均可以进行数据报告的下载。数据报告

目前提供基础报告、维度分析报告、产品效果报告、定制报告、其他报告等，其中基础报告和维度分析报告是最常用的数据报告。

(1) 基础报告

基础报告里包含账户、推广计划、推广单元、关键词四个层级的数据报告，另外还包含基础创意报告、高级创意报告、组件创意报告、搜索词报告，如图4.1所示。各报告统计的数据指标包括：

- 消费：指定范围内的点击消费总金额。
- 点击量：指定范围内获得的点击次数，即潜在客户点击推广结果的次数。
- 平均点击价格：每次点击平均价格，即消费 / 点击量。
- 展现量：指定范围内获得的展现次数，即推广结果覆盖了多少网民。



图 4.1 百度数据报告

(2) 维度分析报告

维度分析报告里包含分时分析报告、左侧 / 上方排名分析报告、地域报告、转化报告、无效点击报告、推广排行榜报告等。



注意

统计报告中的数据并非实时数据，系统可能无法提供最近半小时内的展现量、点击量、消费和点击率数据等。另外，由于反作弊系统对于无效展现的排查会引发数据修正，最近1小时内的展现数据可能会发生波动。无效展现是与正常展现相对而言的，是经系统确认非正常检索行为带来的展现。包括机器产生的检索和人为不以搜索为目的产生的检索行为，如SEO刷词、爬虫流量、双击检索等。

2. 百度统计相关报告

百度统计提供各种维度的访客行为分析报告，如图 4.2 所示，这些报告可以帮助企业全方位改善网站质量与推广方案。



图 4.2 百度统计报表

(1) 流量分析报告

流量分析——多少人来过企业网站？流量分析模块包含流量趋势分析、实时访客、跨屏分析等报告，能告诉 SEM 专员谁来过企业网站，他们分别来自什么地域，以及这些访客是否足够关注你的网站等。

(2) 来源分析报告

流量来源——访客从什么途径来？流量来源模块包含全部来源、搜索引擎、搜索词等报告，能告诉 SEM 专员投放哪些关键词可以让更多的潜在客户找到企业网站，哪些关键词还需要进一步优化，哪种媒介推广方式更有效果，哪种推广方式需要改善等。

(3) 百度推广报告

推广效果——推广效果如何？推广效果模块包含百度推广报告，如搜索推广、网盟推广等，能告诉 SEM 专员投放的关键词中哪些访问及转化效果更好，哪些需要进一步优化等。

(4) 访问分析报告

网站访问——网站够好么？该模块包含受访页面、入口页面、页面热力图、页面上下游等报告，能告诉 SEM 专员网站上哪些页面最受网民欢迎，页面的热点区域，访客习惯的访问轨迹等。

(5) 访客分析报告

访客特征——来访的访客什么样？访客特征模块包含地域分布、访客属性和系统环境等报告，能告诉 SEM 专员访客从什么地区来，使用何种系统环境，男女年龄比例等信息。

(6) 转化分析报告

转化情况——转化效果怎么样？该模块包括转化概况、转化路径、订单分析等报告，能帮助企业监控转化效果，有针对性地发现问题、提升转化。

(7) 优化分析报告

网站诊断——网站流畅吗？该模块包括 SEO 分析、搜索词排名、升降榜等报告，能帮助 SEM 专员检查网站对百度搜索引擎的友好程度，并提出优化的官方建议，评估网站推广效果或查找变化原因。

3. 咨询工具数据

咨询工具都具有数据导出的功能，例如网站商务通，在主界面上显示有“历史记录”字样，如图 4.3 所示，点击“历史记录”即可进入访客记录下载界面，选择对应时段的报告下载即可。



图 4.3 历史记录入口

“导出报表”功能是历史记录的主要功能之一，它可以方便地将历史记录的数据进行分类或排序后生成一套系统的报表数据，如图 4.4 所示。



图 4.4 历史记录导出

4.1.2 不同阶段数据的获取路径

数据越全面，推广的分析和推演越接近真实情况，所以要尽可能把所有的数据都获取到，并进行汇总和整理。主要有推广阶段数据、网站阶段数据、转化阶段数据。

1. 推广阶段

推广阶段的数据主要来源于各推广平台后台的数据报告，例如百度凤巢后台数据报告、360 点睛营销平台数据报告、搜狗推广后台数据报告。目前各搜索引擎的后台都已日趋成熟，直接登录账户进入后台的数据中心，下载对应数据报告即可。

2. 网站阶段

网站阶段数据的获取，大多中小企业都是直接外挂的第三方统计软件，如 GA、百度统计、CNZZ、51LA 等，虽然大多数统计工具在功能上相差无几，但是统计数据还是有微小的差别。例如对百度统计和 CNZZ 两个平台难以取舍，可以两个平台都上，当双重保险。平时的分析中，CNZZ 使用得更普遍，如果你的网站使用了百度竞价，那么百度统计是更好的选择。或者两种统计代码全部部署到网站上去。现在统计代码基本都是异步加载，不用太过担心网站打开速度的问题。

3. 转化阶段

转化阶段的数据主要来源于咨询工具（例如网站商务通、乐语、live800、53kf、百度商桥等）和公司 CRM 系统，主流的在线客服系统在数据留存方面已经相当成熟，在客服软件后台直接选定时间段下载即可，CRM 系统可直接导出销售记录。

4.2 数据处理

数据处理是数据分析过程中非常重要的一个环节，没有数据处理，就没办法进行分析。如何将数据处理成便于分析的样式。这就需要在数据处理上好好下一番功夫。

4.2.1 数据清洗

从各个渠道获取到的数据在格式上难免存在一些问题，数据清洗的过程就是对数据中的缺失值、重复值、错误值、文本型数值等纠错的过程。

1. 缺失数据处理

数据缺失主要有记录的缺失和记录中某个字段信息的缺失，两者都会造成统计结果的不准确。销售阶段的数据在整理时可能会人为造成数据缺失，使数据的完整性受到破坏。如果是主关键词缺失，该条记录已经失去了分析的意义，直接删除该条记录；如果是某个属性值缺失，以零补缺，如图 4.5 所示。

2. 重复数据处理

数据重复一般是指一行记录的信息是完全一样的，多数为人为重复录入。一条记录多次录入，会影响分析结果的准确性，在错误的分析上分析出错误的结果，所以重复数

据也是不允许存在的，要做好去重处理，如图 4.6 所示。

关键词	展现量	点击量	消费	点击率	平均点击价格	平均排名
网络营销课程	143	8	134.95	5.59%	16.87	2.63
	79	4	41.44	5.06%	10.36	4
哪里有网络营销培训	17	1	5.47	5.88%	5.47	

图 4.5 缺失数据处理

关键词	展现量	点击量	消费	点击率	平均点击价格	平均排名
网络营销课程	143	8	134.95	5.59%	16.87	2.63
网络营销学习	79	4	41.44	5.06%	10.36	4
网络营销课程	143	8	134.95	5.59%	16.87	2.63

图 4.6 重复数据处理

3. 错误值数据处理

数据格式错误，如时间、日期、数值、全半角等显示格式不一致。例如销售表格，可能由很多名销售的表格合并而成，所以会存在格式不一致、全半角字符等问题，这类的问题均需要进行格式统一化。格式不一致的数据如图 4.7 所示。

日期	账户	推广计划	推广单元	关键词	展现量	点击量	消费	点击率	平均点击价格
2017-2-20	A账户	网络营销	营销-培训	四川网络营销培训	3	0	0	0.00%	0
2017-2-20	A账户	网络营销	营销-培训	网络营销培训	12	3	51.77	25.00%	17.26
2017-2-20	A账户	网络营销	营销-培训	成都网络营销培训	1	0	0	0.00%	0
02-20-17	A账户	网络营销	网络营销	四川网络营销	9	0	0	0.00%	0
2017-2-20	A账户	网络营销	网络营销	网络营销	6 2	2	16.42	3.23%	8.21
2017-2-20	A账户	网络营销	网络营销	成都网络营销	5	1	8.67	0.2	8.67
2017-2-20	A账户	网络营销	营销-哪里好	网络营销哪里好	31	1	9.41	3.23%	9.41

图 4.7 数据格式不一致

4.2.2 完整数据链合成

在数据分析中，得到一些阶段性的数据，例如推广阶段数据、网站阶段数据、咨询转化阶段数据。对阶段性数据进行分析得到的是阶段性质指标，阶段性质指标能说明的问题往往比较片面，而数据分析需要纵观全局，所以需要把阶段性的数据进行对接，得到一条完整的数据链，才能在最大程度上反映出推广的过程，以及整个过程中出现的问题，进而根据问题来进行优化调整。

1. 完整数据链的作用

完整数据链的作用主要包括以下六个方面。

(1) 完整的数据链更能反映推广真实情况

形成完整数据链后，真实的推广情况都会反映到数据链变化上，这样能对问题的排查和解决提供更多的便捷，驱动营销决策。

(2) 指导关键词调价

完整的数据链可以快速制作出关键词 ROI 值表，如图 4.8 所示，即关键词—关键词消费—关键词对话量—关键词留电量—关键词对话成本—关键词留电成本。根据这一关键词维度的数据链，可以计算出关于关键词的一系列指标，如点击成本、对话成本、留电成本、转化成本、点击率、进入率、转化率等。通过对量指标和质指标的分析，可以全方位对该关键词的“能力”进行评估，从而对关键词出价进行考量，是提高出价，还是降低出价。

关键词名称	消费	展现	点击	平均排名	点击率	转化	转化率	转化成本	出价	平均点击价格
上海有哪些医院有儿科	5461.27	758	72	2.51	9.50%	14	19.44%	390.09	98.06	75.85
上海哪些医院儿科好	5397.16	1571	73	2.41	4.65%	13	17.81%	415.17	93.72	73.93
长宁区妇幼保健院	3987.32	1419	365	1.22	25.72%	77	21.10%	51.78	11.36	10.92
上海医院排名	3033.46	711	79	2.75	11.11%	3	3.80%	1011.15	26.81	38.4
上海儿科	2703.19	282	11	1.68	3.90%	3	27.27%	901.06	365.88	245.74
脑瘫早期症状	1830.21	306	63	1.08	20.59%	7	11.11%	261.46	35.23	29.05
脑瘫好治疗吗	1824.18	242	30	1.91	12.40%	5	16.67%	364.84	56.26	60.81
儿童医院	1795.94	184	7	3.26	3.80%	2	28.57%	897.97	255.66	256.56
上海儿童医院	1774.75	341	40	1.1	11.73%	0	0.00%	#DIV/0!	65.55	44.37
上海哪家医院儿科好	1699.46	862	16	2.72	1.86%	1	6.25%	1699.46	58.99	106.22
脑瘫治疗最佳时间	1577.28	338	26	4.1	7.69%	2	7.69%	788.64	50.24	60.66
上海浦东儿科	1396.9	197	10	3.14	5.08%	2	20.00%	698.45	148.7	139.69
上海正规儿童医院	1364.22	1324	18	3.89	1.36%	4	22.22%	341.06	92.65	75.79
上海 脑瘫治疗	1351.43	113	9	3.06	7.96%	1	11.11%	1351.43	195.48	150.16
上海医保儿童医院	1313.86	2713	16	3.6	0.59%	3	18.75%	437.95	96.68	82.12
上海公立医院	1261.07	250	54	1.1	21.60%	9	16.67%	140.12	23.88	23.35
上海妇幼保健院	1233.76	145	20	3.23	13.79%	3	15.00%	411.25	158.26	61.69
上海最好的儿童医院	1196.87	79	32	0.72	40.51%	0	0.00%	#DIV/0!	75.88	37.4
闵行区妇幼保健院	1158.33	772	108	1.49	13.99%	8	7.41%	144.79	11.36	10.73
儿童哮喘	1094.17	374	42	1.35	11.23%	0	0.00%	#DIV/0!	55	26.05
上海哪家医院治疗儿童哮喘	1088.58	724	14	2.19	1.93%	3	21.43%	362.86	96.42	77.76
上海儿科	1038.63	257	8	4.99	3.11%	3	37.50%	346.21	106.95	129.83
上海儿科医院	947.27	235	4	3.23	1.70%	1	25.00%	947.27	365.88	236.82
上海最好的儿童医院	917.18	87	5	3.19	5.75%	1	20.00%	917.18	284.73	183.44
儿童癫痫	865.06	470	40	1.87	8.51%	3	7.50%	288.35	28.96	21.63
普陀区妇幼保健院	856.04	271	80	1.02	29.52%	11	13.75%	77.82	11.36	10.7

图 4.8 指导关键词调价

(3) 指导推广地域选择

完整的数据链可以快速制作出地域 ROI 值表，如图 4.9 所示，即地域—地域展现量—地域点击量—地域消费—地域咨询量—地域留电量，从而计算出地域的投放成本高低。对于投放成本低的地域，说明该地域获客成本较低，可以作为重点投放地域，反之亦然。

(4) 指导推广时段选择

在同等投放力度下，可以计算出时段—时段消费—时段点击量—时段访客数—时段对话数—时段转化数，同理计算出时段对话成本、转化成本等指标，然后选择转化成本低的时段作为重点时段进行投放（说明我们的意向客户在该时段比较密集）。

(5) 更合理的分配预算

每个账户每天的预算总额是固定的，如何实现效益最大化是至关重要的。根据完整数据链可以得出计划—计划展现量—计划点击量—计划访客数—计划对话数—计划信息

数—计划转化数，从而可以计算出每个计划的点击成本、信息成本、转化成本。到这里我们应该都明白，预算需要向转化成本低的计划倾斜，这样就可以在预算不变的情况下，获取到更多的转化。

地域	展现量	点击量	消费	留电数	留电成本
四川省成都市	1951040	50867	248706.37	1166	213.29877
四川省南充市	200187	6973	27193.53	88	309.01739
四川省达州市	195663	7050	25017.12	116	215.66483
四川省绵阳市	282089	5983	24736.02	157	157.55427
四川省乐山市	112728	3962	21047.19	62	339.47081
四川省宜宾市	131033	4062	16605.11	76	218.48829
四川省德阳市	130923	3653	15928.51	44	362.01159
四川省巴中市	101478	4043	15528.81	94	165.20011
四川省凉山州	97920	3711	13906.56	110	126.42327
四川省泸州市	108153	3434	13609.49	61	223.10639
四川省眉山市	89886	3100	13150.91	60	219.18183
四川省广元市	95989	3254	12395.89	76	163.10382
四川省广安市	91056	3024	12078.76	62	194.81871
四川省内江市	89286	2752	11233.94	47	239.02
四川省遂宁市	128324	2790	10983.42	55	199.69855
四川省未知地域市	65588	3975	10685.87	#N/A	#N/A
四川省资阳市	74797	2401	9820.99	39	251.82026
四川省自贡市	82975	2318	9554.52	52	183.74077
四川省雅安市	55793	1794	7685.66	35	219.59029
四川省攀枝花市	54392	1658	7611.64	24	317.15167
四川省阿坝州	27790	937	3416.91	13	262.83923
四川省甘孜州	22558	737	2611.39	14	186.52786

图 4.9 指导推广地域选择

(6) 选择更合适的着陆页

网站的着陆页各不相同，究竟什么样的着陆页更能迎合潜在目标客户，获得更多的转化呢？可以从完整数据链 URL 维度来进行，URL—URL 进入数—URL 对话数—URL 信息数—URL 转化数。URL 转化数 / URL 进入数 = URL 效率值（仅供参考）。URL 效率值可在一定程度上说明该 URL 是否符合潜在目标客户的需求，从而在现有的 URL 中选择出更优的 URL 来作为着陆页，同时还可以提取转化比较高的着陆页的元素，添加到新设计的页面中去做测试，争取使着陆页越来越符合潜在目标客户的要求。

2. 数据链合成

数据链合成主要包括以下几个步骤。

(1) 数据准备

各个阶段的数据均需要围绕某一关键词进行数据梳理。

- 地域：地域消费、地域访问量、地域咨询数、地域转化数。
- 计划：计划消费、计划访问量、计划咨询数、计划转化数。
- 关键词：关键词消费、关键词访问量、关键词咨询数、关键词转化数。
- 时段：时段消费、时段访问量、时段咨询数、时段转化数。
- URL：URL 进入数、对话数、留电数、跳出量等。

(2) VLOOKUP 函数

函数语法：VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])。函数的参数分别对应的是：VLOOKUP(查找值，查找范围，查找列数，精确匹配或者近似匹配)。我们在 SEM 数据分析中，通常要求精确到每个计划、关键词等，所以几乎都

使用精确匹配，该项的参数一定要选择为 FALSE，否则返回值可能是错误的。

lookup_value 是一个很重要的参数，它可以是数值、文字字符串或引用地址。我们常常用的是引用地址。使用这个参数时，有两点要特别提醒：

table_array 是搜寻的范围，col_index_num 是范围内的列数。

参数 range_lookup 是个逻辑值，我们常常输入 0 或者 FALSE，其实也可以输入 1 或者 TRUE。

(3) 完整数据链合成（以地域维度为例）

首先准备好地域维度消费报告（如图 4.10 所示）、地域维度对话量（如图 4.11 所示）、地域维度电话量（如图 4.12 所示）、地域维度转化量（如图 4.13 所示），数据准备好之后，就可以开始数据的匹配调用了。首先地域维度消费报告中已经有地域的展现量、点击量和消费这三项，通过查找“城市”这个值，在“地域维度对话量”表找到对应城市的对话量。应用 VLOOKUP 函数可以把两个表按照“城市”一一匹配起来，如图 4.14 所示。

账户	地域	城市	展现量	点击量	消费
A账户	四川	成都	2151080	43058	353989.5
A账户	四川	绵阳	241992	5815	34823.53
A账户	四川	达州	223042	6901	32090.33
A账户	四川	南充	187002	6161	30985.84
A账户	四川	宜宾	135607	3951	20665.7
A账户	四川	德阳	134486	3513	20085.02
A账户	四川	巴中	96243	3722	17507.23
A账户	四川	乐山	106842	2873	16419.83
A账户	四川	泸州	113230	3247	16117.64
A账户	四川	广安	89268	3015	15396.5
A账户	四川	眉山	88597	2750	15137.21
A账户	四川	凉山	102929	3375	14959.73
A账户	四川	内江	95655	2813	14664.78
A账户	四川	广元	93288	3099	14081.87
A账户	四川	资阳	75995	2470	13641.19
A账户	四川	自贡	90180	2272	12373.72

图 4.10 地域维度消费报告

地域	对话数
四川省成都市	1193
四川省	195
四川省绵阳市	189
四川省达州市	137
四川省巴中市	86
四川省南充市	84
四川省广安市	82
四川省广元市	79
四川省凉山州	77
四川省泸州市	73
四川省宜宾市	66
四川省眉山市	62
四川省遂宁市	58
四川省内江市	57
四川省德阳市	55
四川省自贡市	44

图 4.11 地域维度对话量

地域	留电数
四川省成都市	946
四川省	155
四川省绵阳市	138
四川省达州市	102
四川省巴中市	61
四川省南充市	61
四川省广安市	64
四川省广元市	50
四川省凉山州	61
四川省泸州市	48
四川省宜宾市	49
四川省眉山市	51
四川省遂宁市	42
四川省内江市	45
四川省德阳市	43
四川省自贡市	35

图 4.12 地域维度电话量

地域	转化数
四川成都市	33
四川	10
四川绵阳市	6
四川达州市	9
四川广安市	6
四川南充市	5
四川巴中市	3
四川凉山州	4
四川眉山市	2
四川广元市	8
四川宜宾市	8
四川泸州市	5
四川内江市	8
四川德阳市	4
四川遂宁市	5
四川自贡市	2
四川资阳市	7

图 4.13 地域维度转化量

账户	地域	展现量	点击量	消费	对话数
A账户	四川成都市	4151000	18058	358389.5	??
A账户	四川绵阳市	841992	5815	34823.53	
A账户	四川达州市	223042	6901	32090.33	
A账户	四川南充市	187002	6161	30985.84	
A账户	四川宜宾市	135607	3251	20665.7	
A账户	四川德阳市	134486	3513	20085.02	
A账户	四川巴中市	96243	3722	17507.23	
A账户	四川乐山市	106842	2873	16419.83	
A账户	四川泸州市	113230	3247	16117.64	
A账户	四川广安市	89268	3015	15396.5	
A账户	四川眉山市	88597	2750	15137.21	
A账户	四川凉山州	109820	2275	14959.72	
A账户	四川内江市				
A账户	四川广元市				
A账户	四川资阳市	75995	2470	13641.19	
A账户	四川	90180	2272	12373.72	

地域	对话数
四川省成都市	1193
四川省	195
四川省绵阳市	189
四川省达州市	137
四川省巴中市	86
四川省南充市	84
四川省广安市	82
四川省广元市	79
四川省凉山州	77
四川省泸州市	73
四川省宜宾市	66
四川省眉山市	62
四川省遂宁市	58
四川省内江市	57
四川省德阳市	55
四川省自贡市	44
陕西省西安市	43
四川省资阳市	40

图 4.14 数据链实现过程

4.2.3 数据具体分析

数据分析是从“数据获取”到“信息转化”的过程，换言之就是获取到数据后，数据本身并没有什么价值，而有价值的是我们从数据中提取出来的转化信息。

1. 数据分析思维方式

数据分析的目的是解决现实中的某个问题或者满足现实中的某个需求。在这个从数据到信息的过程中，无论遇到多少复杂多变的数据布局，肯定是有一些固定的思路或者称之为思维方式，来作为统一的分析模式。

(1) 对照

俗称对比，单独看一个数据是绝对化的数字，不能代表什么，只有跟另一个数据做对比，才能比较出数据价值。图 4.15 为今天销量的柱状图，对数据分析思维来说，今天的销量就是个数字，无数据价值和意义；图 4.16 为今天与昨天的销量，经过跟昨天的销量做对比，就会发现，今天的销量实际比昨天的销量差了一大截。这样数据就有了自己的数据价值。

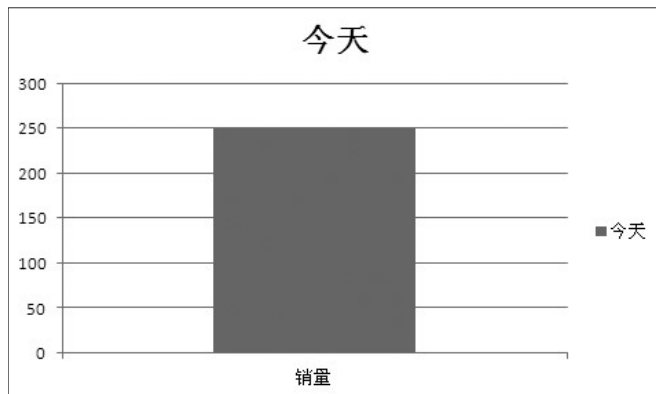


图 4.15 今天销量

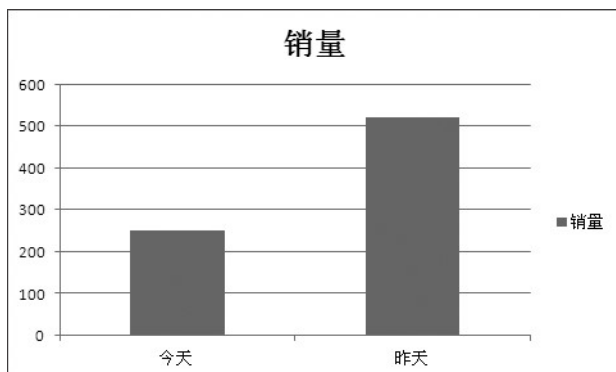


图 4.16 今天销量 VS 昨天销量

这是最基本的思路，也是最重要的思路。在现实数据分析中的应用非常广，大到业绩汇报，小到选款测款、监控店铺数据等工作，这些过程都是在做“对照”。数据分析人员获取到数据后，若数据是单独的，无法进行对比参照的话，就不能用来作为判断依据，也就没有实际意义，也就不能从数据中提取出有价值的信息了。

(2) 拆分

分析这个词从字面上来理解，就是拆分和解析。因此可见拆分在数据分析中的重要性。我们回到第一个思维方式“对照”上面来，当某个维度可以对比的时候，我们选择对比。在对比后发现发现问题需要找出原因的时候却没法可对比，怎么办？这个时候，“拆分”就闪亮登场了。

例如：运营人员小美，经过对比店铺的数据，发现今天的销售额只有昨天的 50%，这个时候，我们再怎么对比销售额这个维度，已经没有意义了。这时就需要对销售额这个维度（或指标）做进一步分解。

根据电商基本公式：销售额 = 成交用户数 × 客单价，成交用户数 = 访客数 (UV) × 转化率。

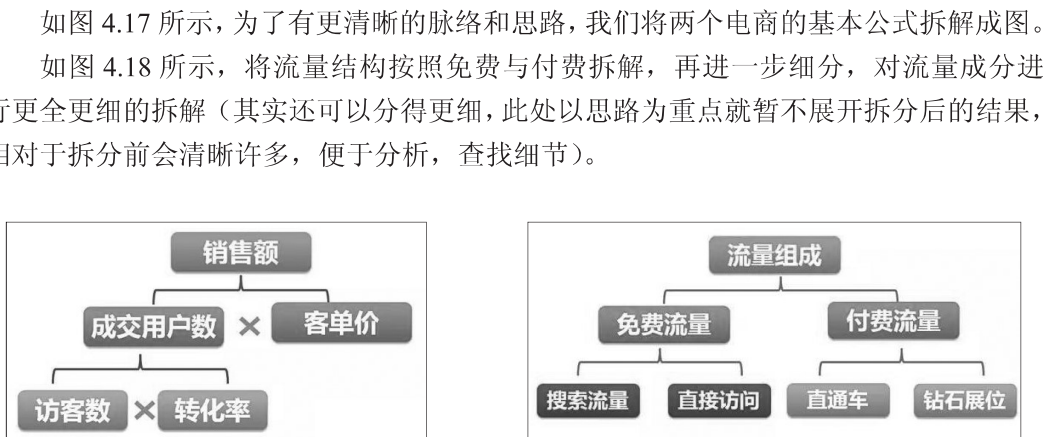


图 4.17 销售额计算

图 4.18 流量组成

例如“流量多还是少”的论断是不利于采取决策的，将流量细分到免费流量和付费

流量上来判断，是“免费流量过少，付费流量占比太大”还是“免费流量很稳定，付费流量不够”，判断越细致，越利于解决问题进而采取有效精准的措施。

(3) 降维

面对一大堆不同维度的数据束手无策，看着屏幕发呆……你是否有过这样的经历？当数据的维度太多的时候，我们不可能分析每一个维度。对一些相关联的数据指标，只筛选出数据堆中能够合并的维度即可，如图 4.19 所示。

日期	浏览量	访客数	访问深度	销售额	销售量	订单数	成交用户数	客单价	转化率
2017/7/1	2584	957	2.7	9045	96	80	67	135	7%
2017/7/2	3625	1450	2.5	9570	125	104	87	110	6%
2017/7/3	2572	1286	2	12780	130	108	90	142	7%
2017/7/4	4125	1650	2.5	15345	143	119	99	155	6%
2017/7/5	3699	1233	3	8362	107	89	74	113	6%
2017/7/6	4155	1286	3.2	14040	130	108	90	156	7%

图 4.19 降维

这么多的数据维度，其实不必每个都分析。我们知道了“成交用户数 / 访客数 = 转化率”，所以当存在这种维度时，是可以通过其他两个维度计算出来的，这时就可以“降维”，即三选二。例如，当数据堆中含有成交用户数、访客数和转化率三个数据维度时，三选二可只留访客数与转化率。而“成交用户数 × 客单价 = 销售额”这三个数据维度也可以三选二，即选择销售额与成交用户数。

综上，通常我们只关心对自己有用的数据，当某些维度的数据跟我们的分析无关时，就可以直接过滤掉，也能达到“降维”目的。

(4) 增维

增维和降维是相对应的，有降必有增。当目前的数据维度不能很清晰全面地解释问题时，我们就需要对数据做一个运算，多增加一个指标。

例如，我们发现一个搜索指数和一个宝贝数，这两个指标一个代表需求，一个代表竞争，有很多人用“搜索指数 / 宝贝数 = 倍数”来代表一个词的竞争度（仅供参考）。这种做法，就是在给数据增维。增加维度有一种叫法称之为“辅助列”。

需要强调的是，做好“增维”和“降维”之前，必须要充分、熟练掌握数据本身的意义与相互之间的关系，有目的地对数据进行转换处理和运算，才能最终达到分析数据的目的。

(5) 假说

当我们根据现有数据分析不出头绪时，可以采用“假说”。假说是统计学的专业名词，俗称假设。当我们不知道结果，或者可能面临几种选择的时候，那么我们就召唤“假说”吧，即先假设有了结果，然后运用逆向思维来进行分析。

从结果到原因，要有什么样的原因，才能产生这种结果。这有点寻根的味道。我们可以知道，现在满足了多少因，还需要多少因。如果是多选的情况下，我们就可以通过这种方法来找到最佳路径（决策）。我们在做未来一年的年度计划或者月度计划的时候，都是对未来拿不准的，这个时候就要假设一个目标，倒推分析达成这个目标需要做什么；或者某项营销活动的结果可能有好几种，那么设想针对不同的结果，应该有什么样的数据现象，然后需要什么样的准备。

当然，“假说”的威力不仅仅如此，除了结果可以假设，过程也是可以假设的。我们回到数据分析的目的，就会知道只有明确了问题和需求，才能选择分析的方法。

以上内容，就是 5 个核心的思维方式：对照、拆分、降维、增维以及假说，大家在简单数据分析时可尝试着套用。

2. 基础报表分析

账户数据统计分析作为 SEM 工作中最为重要最为核心的一个工作阶段，很多 SEM 专员都曾经在这个阶段中迷失过、困惑过。到底该怎么去统计数据？怎么去分析数据？怎么去根据数据来调整账户？这无疑是很很多 SEM 专员心中最为头疼、最为棘手的事情了。

通常账户后台中的数据给人的感觉是又多又杂。“多”指的是数据量过于庞大，“杂”指的是数据过于复杂。

一个账户在日常管理中少不了要做以下几份数据：

(1) 各个推广平台的汇总数据（如图 4.20 所示）。

日期	账户日常数据报告						
	展现	点击	消费	点击率	均价	转化	转化成本
12	180133	3803	11867	2.11%	3.12	39	304.29
13	214463	4082	15122	1.90%	3.70	40	378.06
14	244460	4061	16327	1.66%	4.02	54	302.36
15	242217	4257	16784	1.76%	3.94	57	294.46
16	255780	4659	17975	1.82%	3.86	43	418.02
17	247969	4833	18252	1.95%	3.78	57	320.20
18	298929	5392	1850	1.80%	0.34	58	31.90
19	350631	5692	18114	1.62%	3.18	59	307.02
20	343388	5442	17611	1.58%	3.24	68	258.99
21	304786	5535	17527	1.82%	3.17	52	337.06
22	278421	5464	16120	1.96%	2.95	50	322.39
23	295148	5680	17394	1.92%	3.06	64	271.78
24	304647	6106	17437	2.00%	2.86	48	363.27
25	277678	5155	16443	1.86%	3.19	47	349.85
26	285683	4974	16645	1.74%	3.35	46	361.84
27	229850	4859	14368	2.11%	2.96	54	266.07
28	238716	5295	17415	2.22%	3.29	53	328.59
29	164935	5497	18396	3.33%	3.35	67	274.56
30	169954	5377	18806	3.16%	3.50	43	437.35
总计	7020824	140686	434221	63.50%	3.09	1535	282.88

图 4.20 各平台汇总数据

- “展现量”指的是在推广平台上的曝光量
- “点击量”是给网站带来的流量
- “消费”指投入的推广费用
- “点击率” = 点击 / 展现
- “点击均价”指点击一次所支付的平均费用
- “转化成本” = 总消费 / 转化

这份数据不只是一是要给自己看，更是要给领导看，这是评估账户推广效果的一份汇总数据，记录着账户中每日的总消费以及投资回报率，是每日必做的一份数据。

(2) 每个推广账户的各病种分类转化数据（以医疗行业为例）（如图 4.21 所示）。

如果你对账户中的计划和单元进行了一个病种标注，那这份数据是很容易汇总出来的。某个病种每天消费多少、点击多少、转化多少，不是当领导这样问你的时候，你才

想着去统计这份数据。这份数据不是做给领导看的，而是自己在分析账户时，真正需要去认真对待的一份数据。

病种	展现	点击	消费	点击率	均价	转化	转化成本
脱发	19480	473	2880.58	2.43%	0.16	15	192.04
水泡	7543	145	748.7	1.92%	0.19	16	46.79
跖疣	5085	261	765.23	5.13%	0.34	7	109.32
寻常疣	237	24	92.5	10.13%	0.26	1	92.50
扁平疣	4834	339	1499.67	7.01%	0.23	13	115.36
头皮癣	2404	74	419.63	3.08%	0.18	2	209.82
花斑癣	1963	117	483.68	5.96%	0.24	6	80.61
紫癜	1261	77	267.26	6.11%	0.29	3	89.09
行业词	5769	380	782.61	6.59%	0.49	12	65.22
酒糟鼻	648	26	783.92	4.01%	0.03	2	391.96
疥疮	396	20	73.07	5.05%	0.27	1	73.07
灰指	125	6	23.64	4.80%	0.25	1	23.64
荨麻疹	1681	86	290.98	5.12%	0.30	10	29.10
牛皮癣	873	37	142.68	4.24%	0.26	1	142.68
湿疹	618	49	346.44	7.93%	0.14	4	86.61
青春痘	1984	74	477.75	3.73%	0.15	4	119.44
掌跖脓疱病	184	10	58.37	5.43%	0.17	1	58.37
总和	55085	2198	10136.71	3.99%	0.22	99	102.39

图 4.21 病种分类数据

(3) 每个推广账户的病种地区转化数据（如图 4.22 所示）。

病种	脱发			水泡			跖疣		
	消费	转化	转化成本	消费	转化	转化成本	消费	转化	转化成本
湖北	3376.82	25	135.07	4272.49	21	203.45	3319.93	23	144.34
河南	3141.3	37	84.90	4095.09	35	117.00	3899.58	33	118.17
湖南	2126.57	19	111.92	2219.4	16	138.71	2113.62	18	117.42
河南	211.42	1	211.42	261.08	3	87.03	229.41	4	57.35
广东	213.57	2	106.79	453.82	3	151.27	332.11	1	332.11
山西	205.41	4	51.35	174.65	1	174.65	143.23	3	47.74
山东	45.58	1	45.58	22.21	0	#DIV/0!	37.44	2	18.72
甘肃	39.9	0	#DIV/0!	4.49	0	#DIV/0!	3.88	0	#DIV/0!
辽宁	234.68	4	58.67	397.72	5	79.54	398.38	6	66.40
吉林	249.82	5	49.96	232.01	1	232.01	230.96	7	32.99
黑龙江	0.93	0	#DIV/0!	0	0	#DIV/0!	0	0	#DIV/0!
云南	25.03	0	#DIV/0!	112.68	1	112.68	16.41	0	#DIV/0!
贵州	62.91	3	20.97	17.94	0	#DIV/0!	7.66	0	#DIV/0!
重庆	43.15	0	#DIV/0!	58.26	3	19.42	19.9	1	19.90
陕西	14.01	0	#DIV/0!	6.99	0	#DIV/0!	2.93	1	2.93
其它区域	105.64	3	35.21	62.79	2	31.40	36.89	4	9.22
合计:	10096.74	104	97.08	12391.62	91	136.17	10792.3	103	104.78

图 4.22 病种地域数据分析

如果你的账户是全国推广，是很有必要每周将各个推广地区的投放情况进行一次数据统计的。作为一名精细化 SEM 专员，每月上级给出一个总的推广费用，你是否会将这些推广费用事先划分到各个推广地区去，a 地区划分多少费用投放，b 地区划分多少费用投放，以及每个地区所投放的病种分别划分多少费用投放，后期你都要制作一份地区数据来跟踪好这些数据，做好反馈和调整。

数据统计虽然复杂，但如果你对竞价具有这种钻研琢磨的精神，这些数据统计都不在话下，当然你还会挖掘统计到更有价值的数，这也是我们做竞价的精髓之一。

汇总的数据也就是各个推广平台的汇总数据，如图 4.20 所示；一个推广平台总的

消费，我们知道了；但是这个总的消费是怎么得来的，于是我们将汇总数据进行了分解统计，也就是统计出了每个推广账户的各病种分类转化数据，如图 4.21 所示。

3. 数据分析基础指标

数据分析对于搜索引擎竞价推广账户来说至关重要。因为每次点击都会产生消费，如果数据分析工作没做好，就会造成广告费用的浪费。那么，百度竞价推广账户日常需要分析哪些基础数据呢？

(1) 网站展现量

在网民搜索查询时，如果账户内符合网民搜索需求的关键词被触发，该关键词所对应的创意将出现在搜索结果页，称之为关键词和创意的一次展现。一段时间内创意获得的展现次数就称之为“展现量”。

想让用户搜索到你，就需要有足够的展现量。如果展现量过低，说明选择的搜索引擎和关键词出现了问题。首先应分析网站受众人群和搜索引擎使用人群是否匹配，关键词数量是否适中。总的来说，可以通过增加关键词数量、调整关键词价格、设置关键词匹配模式等方面来解决展现量过低问题。

(2) 网站访问量

网站访问量是指网站流量（traffic），是用来描述访问一个网站的用户数量以及用户所浏览的网页数量等指标。常用的统计指标包括网站的独立用户数量、总用户数量（含重复访问者）、网页浏览数量、每个用户的页面浏览数量、用户在网站的平均停留时间等。

- 访问数（IP）：指独立 IP 数。00:00 ~ 24:00 内相同 IP 地址只被计算一次。
- 独立访客（UV）：访问网站的一台计算机客户端为一个访客。00:00-24:00 内相同的客户端只被计算一次。
- 综合浏览量（PV）：即页面浏览量或页面被看了多少次，如果一个页面被一个访客看了多次就多次累积，每次刷新也被计算一次。

(3) 网站点击率

企业推广结果展现时，如果网民对企业的推广结果感兴趣，希望进一步了解企业的产品/服务，可能会点击访问企业的网站。一段时间内企业获得的点击次数称为“点击量”。

点击率的计算公式为：点击量 / 展现量。当点击率过低时首先检查是否是账户出了问题，如账户质量过低，有可能是关键词创意无法吸引用户点击或是排名太靠后，此时可以尝试重新修改创意或出价。

(4) 网站抵达率

抵达率 = 访问量 / 点击量。如果抵达率偏低，可以考虑以下三方面原因：

- 创意与推广页面主题相关性不大，可通过修改相应创意或着陆页 URL 解决。
- URL 打不开或者打开速度太慢，查看推广 URL 速度报告以进行排查。
- 推广页面用户体验欠佳。抵达率低说明你的网站体验不好，网站链接可能出现了问题，造成无法访问或者访问速度慢的问题。

(5) 转化率

网站转化率（conversion rate）是指用户进行了相应目标行动的访问次数与总访问

次数的比率。相应的行动可以是用户登录、用户注册、用户订阅、用户下载、用户购买等一些用户行为，因此网站转化率是一个广义的概念。简而言之，就是当访客访问网站的时候，把访客转化成网站常驻用户，也可以理解为访客到用户的转换。

转化率的计算公式为：转化量 / 点击量。当转化率过低时要排查是哪些因素造成的。首先查看关键词是否适合受众人群搜索习惯和是否够精准，或是否推广页面不够吸引人。分析时可以结合页面访问时长和页面跳出率等其他推广数据，然后根据分析结果做出相应的优化调整。

(6) 网站跳出率

跳出率：指在某个时间段内，只浏览了一页即离开网站的访问次数占总访问次数的比例。对于某页面的跳出率算法是从这个页面进入网站没有再点击其他页面即离开的次数 / 所有进入这个页面的次数。对于整个网站跳出率的算法是只浏览一个页面即离开的访问次数 / 进入网站的总次数。

网站跳出率是评价一个网站性能的重要指标。跳出率高，说明网站用户体验做得不好，用户进去就跳出去了；反之如果跳出率较低，说明网站用户体验做得不错，用户能够找到自己需要的内容。而且以后他可能还会再来光顾你的网站，提高了用户黏度，慢慢地可以积累大量的网站用户。

(7) 平均访问时长

平均访问时长的长短是网站分析的一个重要指标，通常用于评估网站的用户体验，可以用于指导网站以及页面的改善。平均访问时长越长，说明网站或页面对用户的吸引力越强，能带给用户的有用信息越多，用户越喜爱；反之，对用户的吸引力越差，可用的有用信息越少，也说明网站需要优化了。

(8) 转化次数

转化次数是指访客到达转化目标页面的次数，说明该页面用户体验好，访客有黏度，对网页有兴趣且进行了多次访问，成交意向较高。

(9) 推广关键词

首先查看投放的关键词流量报告，查看哪些词带来了流量，哪些词带来了对话用户？哪些词带来了咨询用户并留下联系方式或上门报名或直接购买产品等，即形成转化的词，我们通过数据还可以观察推测哪些词是被恶意点击的。

筛选关键词，对于竞价咨询率比较高，有意向能够达成转化的这类词，可以加大投入，增加此类相关词，并做着陆页策划，写好创意；相反，对转化率低、消费高的或者没有转化率的词，可以考虑暂停或直接删掉。

(10) 平均点击价格

平均点击价格的调整可以更好地控制推广成本。首先找出单次点击成本过高的词和日点击次数过多的词，对重点词调整出价。通过不断调整、优化关键词质量度（质量度是根据关键词的点击率、关键词与创意的相关性、账户整体表现等多个因素计算出来的。搜索者越感兴趣，越多人点击，访问者体验越好，质量度越高），用最低的价格达到最佳的排名位置，节省推广费用。

关键词排在所有推广结果的最后一名，或是唯一一个可以展现的推广结果，则点击价格为该关键词的最低展现价格。其他情况下，理论上每次点击价格 = (下一名出价 * 下一名质量度) / 关键词质量度 + 0.01 元。

对百度搜索推广来说，综合排名指数（CRI），即出价与质量度的乘积，才是排名真正的衡量标准。也就是说，如果关键词有很好的质量度就有可能在出价较低的情况下排名靠前。

4. 数据分析思路要点

这里以医疗行业为例，在进行数据分析时，主要有以下几个方面的思路要点。

(1) 高消费高转化病种分析方向

- 高消费高转化的病种为什么消费那么多？分别是哪些词在消费？成本是否在可控范围？
- 高消费高转化的病种分别是哪些词产生转化？我们能否让这些产生转化的词继续保留转化下去？
- 高消费的病种是不是我们主推的病种？
- 高消费高转化的病种里面是否包括一部分有消费无转化的关键词？我们是否找出高消费无转化的词做好调整，让高消费高转化的病种成本更低。

(2) 高消费低转化病种分析方向

- 高消费低转化的病种为什么消费那么多？是哪些词在消费？为什么这些消费的词没有产生转化？要怎么去调整好这些高消费无转化的词。
- 高消费低转化的病种分别是哪些词产生转化？我们能否让这些产生转化的词继续保留转化下去？让高消费低转化的病种成本能够控制在正常的范围之内？
- 高消费低转化的病种是不是我们主推的病种？我们是否考虑将高消费低转化的病种转化成本先压低下去，让一个总体的数据成本能更低些？

(3) 低消费高转化病种分析方向

- 低消费高转化的病种分别是哪些词产生转化？为什么这类病种消费这么少就能产生这么多转化？
- 我们是否将这类低消费高转化的病种进一步加大扩展投放，让这些低消费高转化的病种能带来更多转化？

(4) 低消费低转化病种分析方向

- 低消费低转化的病种转化成本是否在正常范围之内？又分别是哪些词在消费？产生转化的词又是哪些？
- 对低消费低转化的病种接下来我们该怎么去做好优化和调整工作，能让这类病种发挥出最大的存在价值？这些我们都要去进行深度思考和规划总结。

上面我们分别通过四个方向进行了深度分析和思考。总的分析方向告诉我们：有消费的病种为什么消费那么多，有转化的病种怎么让这些病种继续保留转化，以及该怎么对这些病种做好优化及管理工作。

本章总结

- 数据获取渠道：推广数据报告、百度统计相关报告、咨询工具数据历史记录。
- 数据阶段：主要是推广阶段数据、网站统计阶段数据、咨询工具 & CRM 阶段数据。
- 数据处理：数据缺失处理、数据重复处理、数据格式不一致处理、CSV 格式分列处理。
- 完整数据链的作用、完整数据链合成、VLOOKUP 函数使用。
- 数据分析思维方式：对照、拆分、降维、增维、假设。
- 基础指标分析：“消费”“点击率”“点击均价”“转化”“转化成本”。

本章作业

1. 说出数据分析需要的数据由哪几大部分组成。
2. 说出 VLOOKUP 函数的每一个参数的含义。
3. 独立完成数据链合成的实战（自行准备数据）。
 - 合成地域维度完整数据链（地域—地域展现量—地域点击量—地域访问量—地域信息量—地域转化量）。
 - 合成关键词维度完整数据链（关键词—关键词展现量—关键词点击量—关键词访问量—关键词信息量—关键词转化量）。
4. 登录课工场，按要求完成预习作业。



随手笔记

A series of horizontal dashed lines for writing notes.

