

天猫婴幼儿商品交易数据分析报告

邓乔予
2023.01

1 分析背景与目的

日常业务产生的数据蕴含巨大的商业价值, 本文利用天池上公开的天猫婴幼儿商品交易数据进行分析, 以期挖掘数据背后的价值信息, 寻找提升商户销量的方法, 发现平台当前存在的问题, 给运营与营销部门提供数据参考与建议。

2 数据预处理

2.1 数据认识

天猫婴幼儿数据集由两张表组成, 表一记录了用户的购买记录, 详细描述了用户的订单信息, 共 29971 条记录。表二为婴儿信息, 每一条记录了婴幼儿的特征, 共 953 条记录。表 1 展示了表中的字段名及其含义。

表 1 数据集字段信息

表名	字段	中文名称	含义
表一 购买记录	buy_mount	购买数量	用户对商品的成交量
	user_id	用户 ID	用户在电商平台的唯一身份标识
	auction_id	购买行为编号	购买行为唯一编号
	cat1	商品一级分类	反映用户购买的商品属于哪个一级分类
	cat_id	商品二级分类	反映用户购买的商品属于哪个二级分类
	property	商品属性	有长度、大小、品牌、颜色等一系列属性
	day	购买时间	用户购买婴儿商品的时间戳
表二 婴幼儿信息	user_id	用户 ID	用户在电商平台的唯一身份标识
	birthday	出生日期	婴儿出生日期
	gender	性别	婴儿性别: 0 为女性, 1 为男性

2.2 数据清洗

数据清洗主要包括三个部分: 删除重复值、缺失值处理以及异常值处理。由于本数据集中不包含重复值和缺失值, 因此本次数据清洗只包含异常值处理。

… (1) 购买数量异常值

为了使异常值的剔除更加符合业务逻辑, 异常值依据各年龄段婴幼儿奶粉推荐食用量确定, 表 2 展示了各年龄段奶粉最大食用量。

表 2 各年龄段奶粉最大食用量

奶粉类型			一段								
月龄	奶粉（平均）	一平均奶粉量	每日推荐喂哺次数	每日最大喂哺次数	天数	奶粉克数小计	最小罐	370g罐数	最大罐	850g罐数	
0~2周		2	4.4	9~10	10	14	1232	370	3	850	1
2~4周		3		7~8	8	14	1478.4		4		2
4~8周		4		6~7	7	28	3449.6		9		4
2~6个月		5		5~6	6	120	15840		43		19
1段总用量						22000		59		26	
奶粉类型			二段								
月龄	奶粉（平均）	一平均奶粉量	每日推荐喂哺次数	每日最大喂哺次数	天数	奶粉克数小计	最小罐	370g罐数	最大罐	850g罐数	
6~12个月		6	5	3~4	4	180	21600	370	58	850	25
奶粉类型			三段								
月龄	奶粉（平均）	一平均奶粉量	每日推荐喂哺次数	每日最大喂哺次数	天数	奶粉克数小计	最小罐	370g罐数	最大罐	850g罐数	
12~36个月		4	10	2~3	3	720	86400	370	234	850	102
奶粉类型			四段								
月龄	奶粉（平均）	一平均奶粉量	每日推荐喂哺次数	每日最大喂哺次数	天数	奶粉克数小计	最小罐	370g罐数	最大罐	850g罐数	
36~72个月		4	10	2~3	3	1080	129600	370	350	850	152

表 2 中的各项数据参考美赞臣铂睿婴幼儿奶粉喂哺标准[1-5]，计算出每一个时期所需要的奶粉最大罐数。对于需求方来说，天猫为 B2C 平台，不存在企业批量购货的现象，因此购买奶粉的需求方为个人用户。根据表 2 的计算结果，婴幼儿奶粉需求量最大的年龄段在 12~72 个月之间，一年最多消耗 51 罐 850g 奶粉。根据行业的调查报告显示[6]，双十一和 618 两个购物节是囤货高峰。假设用户在两个购物节购买相同量的产品，则 26 罐为用户每次能够购买的最大罐数，也是本次销量数据的异常阈值。

(2) 婴儿特征异常值

数据集中的婴幼儿年龄存在异常值 28，-1 和-2。-1 认为是提前囤货不做处理，-2 和 28 应该是用户误填信息，在分析婴幼儿年龄与产品选择之间的关系时，过滤掉这部分的信息。婴幼儿性别中也存在异常值 2，在分析性别与产品选择之间的关系时，同样将其过滤掉。

3 分析框架

本次数据集中没有过程指标，只有一个销量指标作为结果指标。因此，本文分析的目的有三个部分：分析当前销量的状况、发生此状况的原因以及未来提升的空间。

(1) 销量状况

主要依据销量指标的时间序列来观察销量的走势，计算销量在每个月份的环比和同比，找出发生异常的时间节点。

(2) 异常原因

总结发生销量异常的具体原因，给出发生异常的产品线。

(3) 销量提升空间

销量差的话，给出销量的缺口大小、挽救时间节点、具体的运营对象以及用户画像。

4 分析内容

4.1 销量基本情况

图 1 展示了婴幼儿奶粉销售情况，依据月度销售情况，可以看出 14 年的销售走势与 13 年类似，5 月以及 11 月出现销量高峰，1 月以及 2 月销量低迷。同时，由于 2015 年 2 月也出现销量断崖式的下降，因此本文主要聚焦于分析各年度 1-2 月的数据，找出销量低迷的原因。

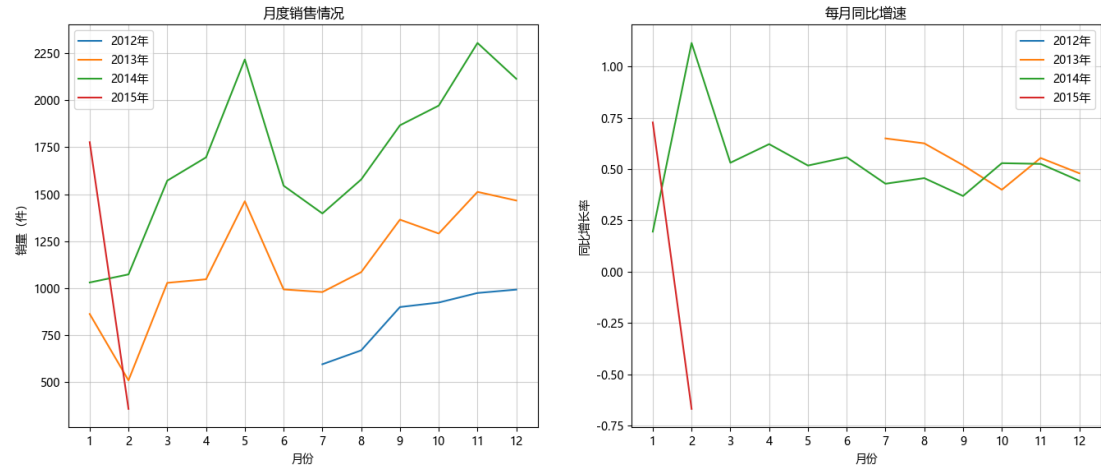


图 1 婴幼儿奶粉销售情况

在预处理的过程中，发现 2015 年 2 月只有 1 日至 5 日的数据，因此数据下降一部分是由数据缺失导致的。但 2013 年 2 月的销量数据也呈现断崖式的下降，同时，2014 年 1 月同

比增速（19.51%）与 2015 年 1 月（72.69%）的同比增速相比较，也表现较差。由于我国春节假期一般都在每年 1-2 月份，假期期间部分快递停运，销量下滑的原因可能与春节假期有关。

4.2 销量异常时间定位

对基本的销售情况有所了解之后，本文假设 1-2 月的销量下滑与春节假期有关。表 3 展示了 2013-2015 年春节的具体时段，本次分析讨论春节前 30 日的销售情况是否异常。

表 3 春节具体时段

年份	时段
2013	2 月 9 日-2 月 15 日
2014	1 月 30 日-2 月 6 日
2015	2 月 19 日-2 月 25 日

从图 2 可以看出闰月初一到十七，同比增速为 57.55%，而 15 年增速为 40.61%，低于目标 57.55%，销量状况不佳。计算 2014 年春节前 30 日的销量为 783 罐，同比增速 57.55%，得到 2015 年春节前 30 日目标销量为 783*1.57 为 1720 罐，而目前总销量为 1101 罐，还有 619 罐的缺口，平均每天 48 罐。

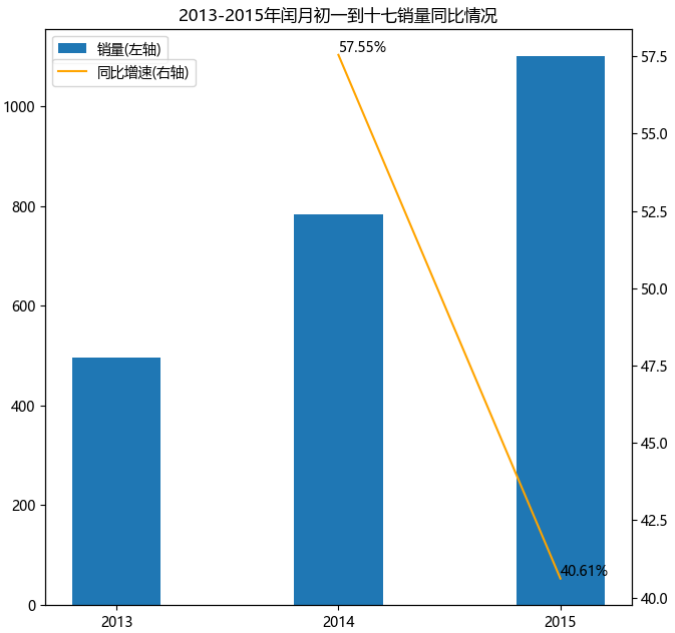


图 2 闰月初一到初十七销量同比情况

4.3 销量异常原因

图 3 展示了各产品大类的销售情况，28 大类，38 大类以及 50022520 大类销售情况良好，同比增长率超过了 65%，但销售量位于第二和第三的 50008168 与 50014815 销售情况不佳，同比增长率仅有 18%。122650008 大类的销售量同比增长率尽管低于 57.55%，但由于销量较低，是销量下降的次要问题点。

由于数据集中只包含了有关商品类别与销售量的数据集，只能考察了二级产品数量与商品销售量之间的关系，发现 50008168 和 50014815 大类的旧产品同比增长为负值（-7%和-6%），旧产品的销售情况不乐观，新产品带来的增长量也仅有 20%左右(图 4)。但旧产品的销售情况的原因不能够进一步深挖，因为缺少定价、渠道、推广还是产品的生命周期数据。

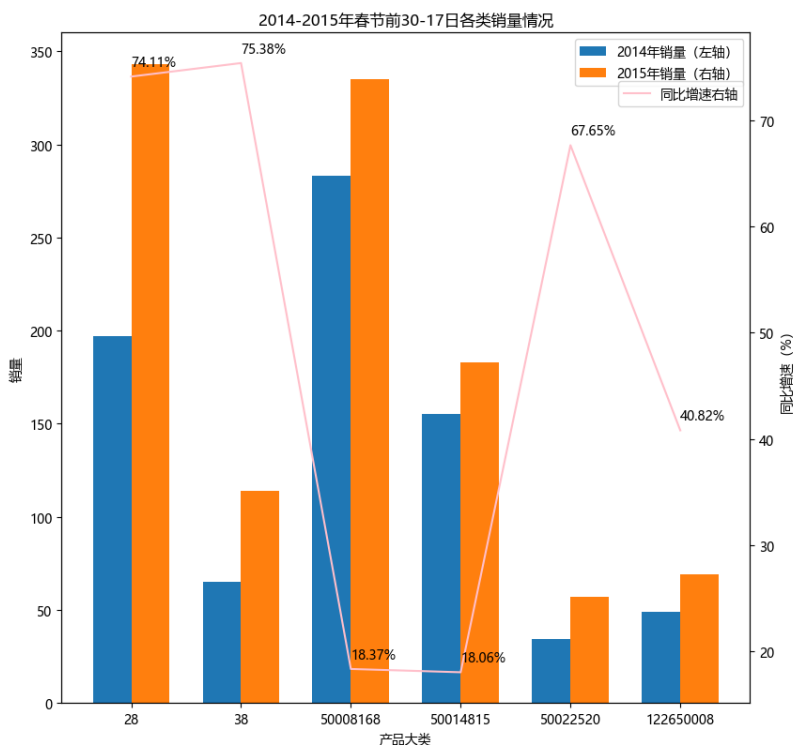


图 3 春节前 30-17 日各类销量情况

除了销量异常的下降，图 3 也包含销量异常的上涨。28 大类和 38 大类增长情况良好，同比增长率分别为 74.11%和 75.38%。本数据集中包含了产品二级分类，对二级分类中的产品进行深挖，发现 28 大类的销量上涨可能有二级产品数量增多导致（与 2014 年相比，增加了 27 种二级产品），计算 2015 年订单中只包含 2014 年二级产品的销量总和，发现 28 大类的销量仅 218 罐（同比增长 10.66%），说明 2015 年 28 大类销量上涨原因是由新二级产品带来的。

表 4 各年份二级产品数量

产品代码	二级产品数量		
	2013	2014	2015
28	58	61	86
38	10	9	13
50008168	26	30	38
50014815	32	34	47
50022520	13	16	22
122650008	6	8	7

38 大类和 50022520 的二级产品数量尽管分别增加了 4 种和 6 种，但新产品的开发仍然是销量上涨的原因，2015 年旧产品的销量仅为 79 罐（同比上涨 21%）和 37 罐（同比上涨 8%）。

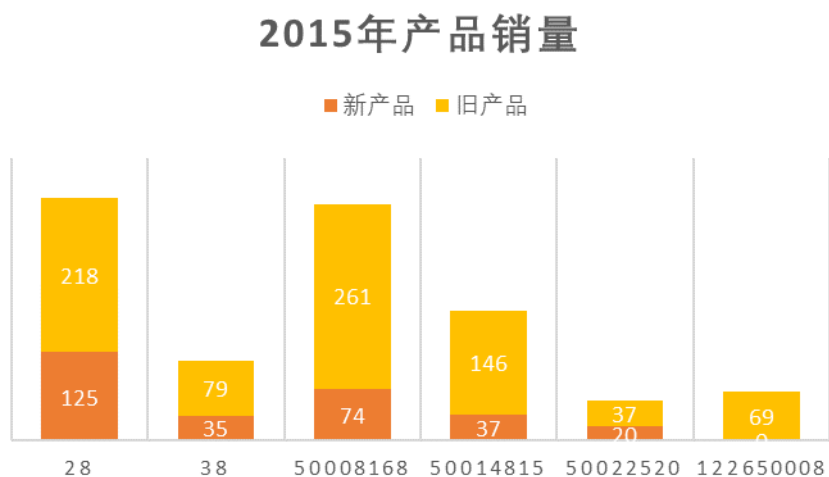
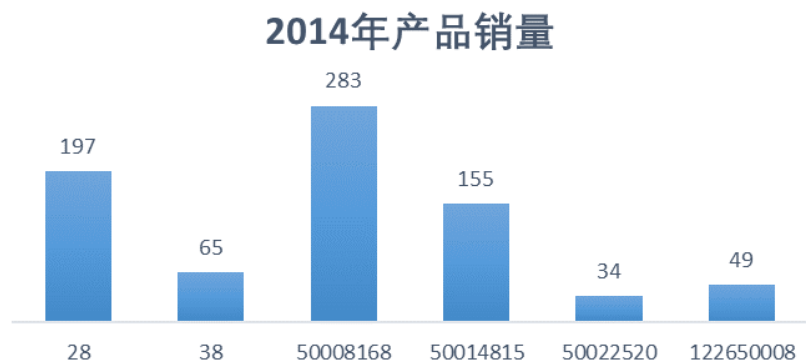


图 4 产品销量

4.4 销量推广计划

(1) 推广时限

从往期日销售情况来看，春节前 30 日的销售量将逐步下降，且闰月廿六将会下降到 10 罐以下。假设廿六到三十的日销量为 5 罐，那么闰月十八到廿五的日均销量需要达到 74 罐。需要在接下来的八天作 1~2 轮推广计划，并将闰月廿六作为保底冲刺节点。

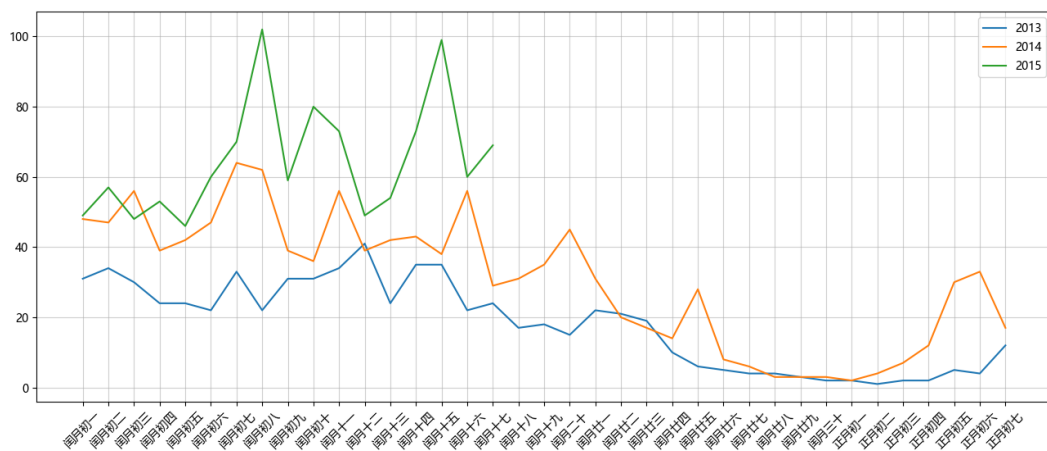


图 5 春节前 30 日日销售情况

(2) 推广产品

28 大类：观察日销售数据，发现每隔 3-7 天销售量会陡增。需要确定是否是自然增长还是推广带来的增长。若是自然增长，接下来还将会再有 2~3 次陡增，当日销售量在 30 罐之上。若是推广计划带来的增长，参考闰月初七以及闰月廿四的推广计划，在未来一周做 1~2 次推广。

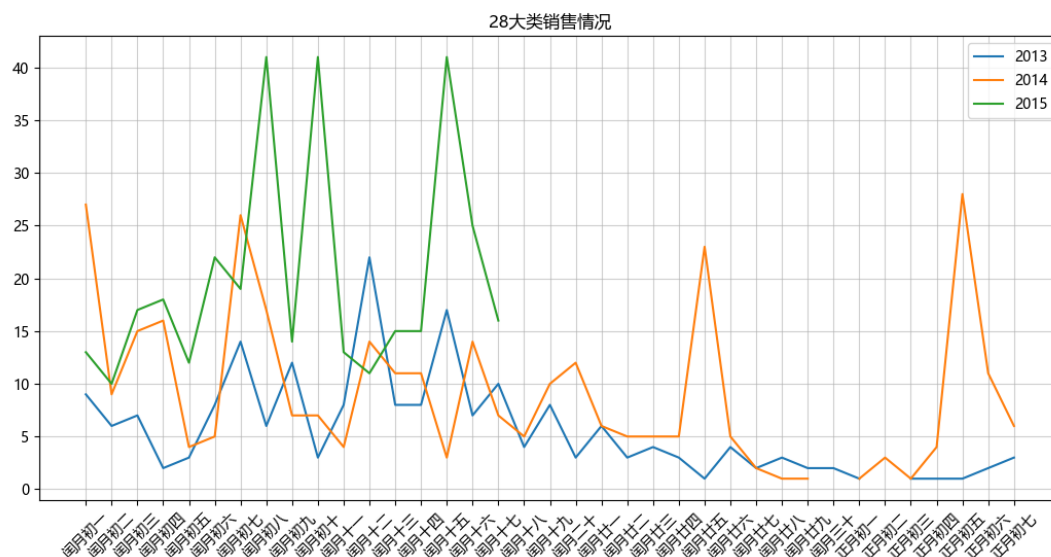


图 6 28 大类销售情况

38 大类：2014-2015 年的销量数据不是很稳定，但 2015 年个别日期销量猛增达到 20 罐，参考当前是否有推广计划，分析是否由推广计划带来的销量增加。若推广计划有效，考虑延长推广日期。

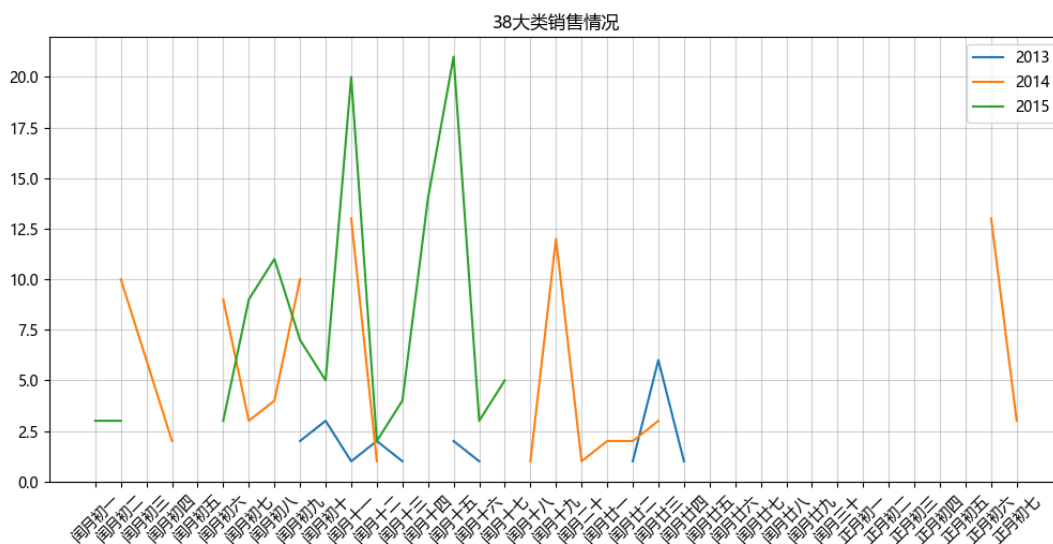


图 7 38 大类销售情况

50008168 大类：2013~2014 年的销售数据显示，该大类的销量从闰月十五开始将会持续下降，参考过往是否有推广计划，若推广之后销量仍然没有上涨，不考虑推广这类产品。

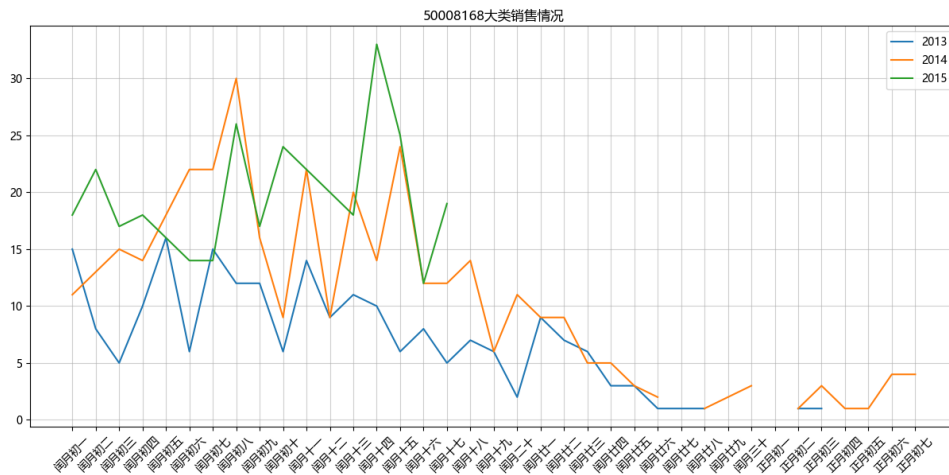


图 8 5008168 大类销售情况

50014815 大类：过往数据显示闰月二十到闰月廿二，该大类的销售量将会有小幅度增长。参考是否此前有推广计划。

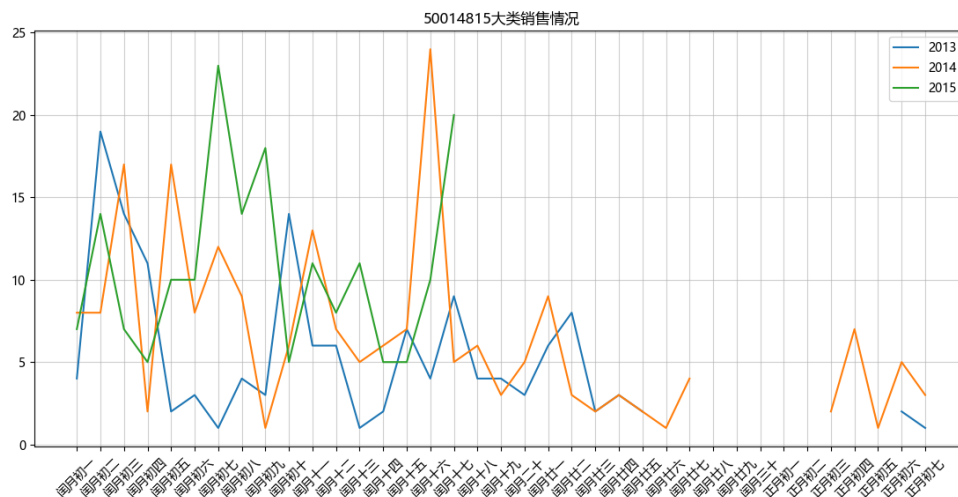


图 9 50014815 大类销售情况

50022520 大类和 122650008 大类：两大类销量不稳定且销量大都处于 10 罐以内，在有限的资源条件下，不考虑作为推广计划的对象。

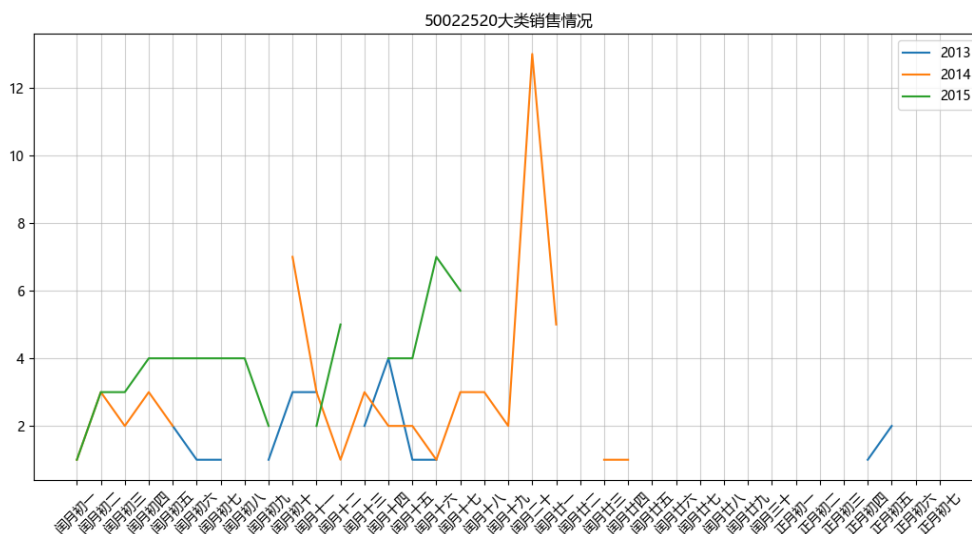


图 10 50022520 大类销售情况

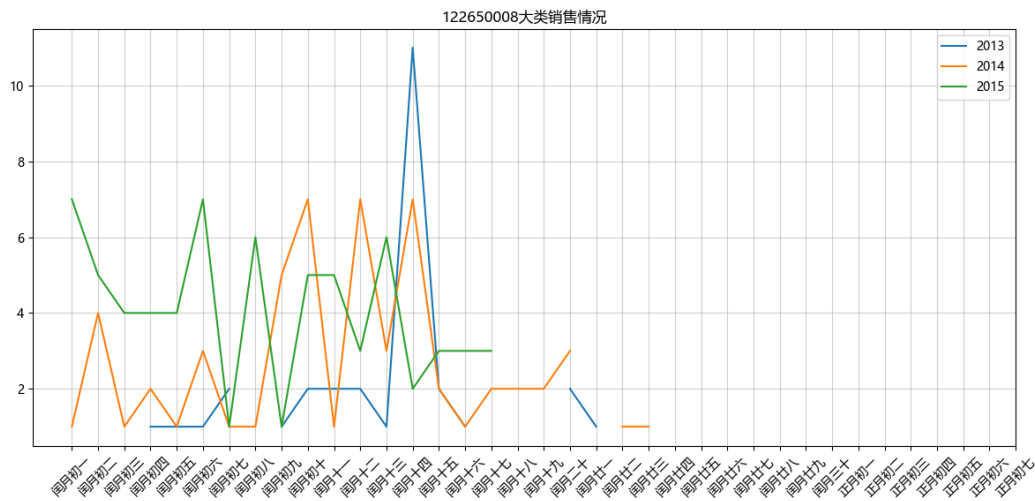


图 11 122650008 大类销售情况

(3) 推广人群

对比 2013-2015 年春节前 30-17 日的用户数增长情况，发现在 2015 年 1-2 月 28 大类、38 大类的用户数增长情况良好，剩余大类用户数增长不足。用户数由新老用户两个部分组成，计算 2014 与 2015 年春节前 30-17 日的新老用户数，发现购买产品的均为新用户，说明销量增长主要通过拉新来达成（2014 年与 2015 年的用户购买罐数均值分别为 1.35 和 1.34，因此不认为是客单价影响了销量增长）。

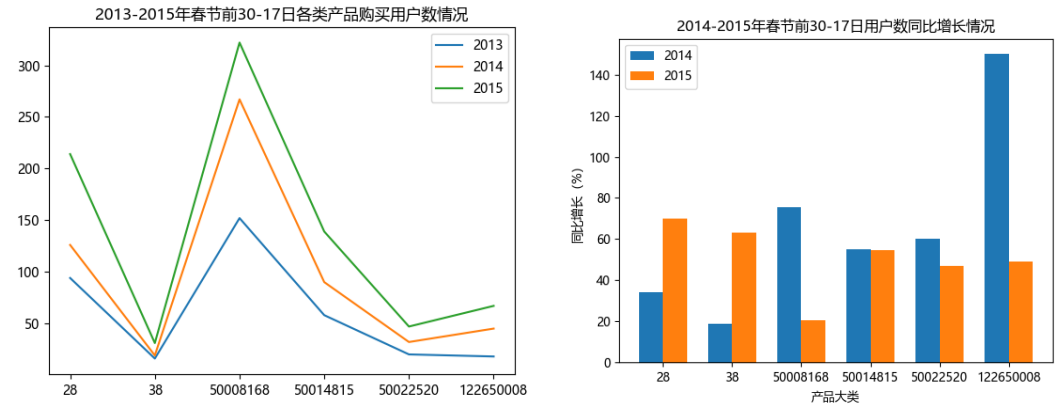


图 12 用户数量变化

结合婴儿特征和用户的购买记录，得到 0-6 周岁婴幼儿产品销售情况如下，与 2015 年的销售情况有较大差异。包含婴儿特征的数据集中有 103 条 2012 年的购买记录，297 条 2013 年的购买记录，424 条 2014 年的购买记录，51 条 2015 年的购买记录。2015 年记录较少，需要得到更多的抽样数据，才能对当前的用户行为进行分析。

本文也假设婴幼儿性别会影响产品类别，使用 t 检验对检验性别的差异是否会带来购买数量的差异，检验结果不显著，婴幼儿性别与产品购买罐数没有关系。

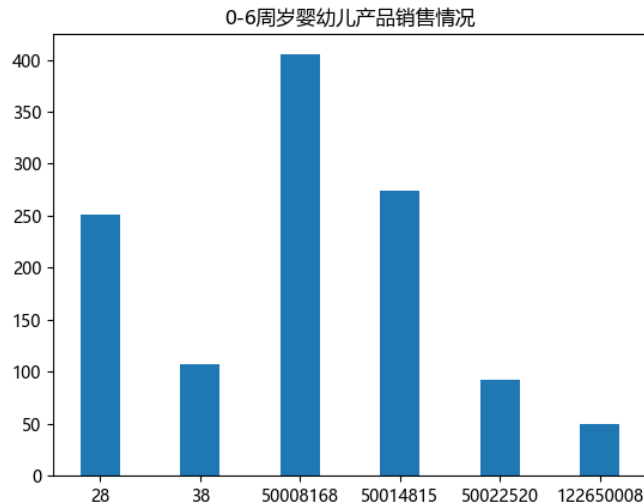


图 13 0-6 周岁婴幼儿产品销售情况

5 结论与建议

销量基本情况：春节前 30 日-17 日总销量为 1101 罐，同比增速为 40.61%，低于目标值 57.55%，销售状况有待提高。距离春节还有 13 天，有 619 罐的缺口，未来 8 天的日均销售量需要达到 74 罐，廿六到三十的日销量达到 5 罐才能达到目标

销量异常原因：50008168 和 50014815 两大类的销售增速低于 20%，是销量下降的主要原因。旧产品的销量为负增长，新产品的销量同比增长也仅有 20%，需要进一步收集相关渠道、推广、产品信息探究更深层的原因。

销量推广计划：未来 8 天作 1~2 轮推广计划，将廿六作为最后的冲刺节点。

28 大类的销售量最高，可以作为主推产品，其次是 38 大类。

在资源有剩余的情况下，再考虑推广 5000168 大类、50014815 大类。

参考资料

1. 1 段奶粉喂哺标准：[顺丰包邮 荷兰进口 美赞臣铂睿 A2 婴儿配方奶粉 1 段 0-6 个月 850g*1 罐-tmall.com 天猫](#)
2. 2 段奶粉喂哺标准：[新上市 美赞臣铂睿 A2 蛋白婴儿配方奶粉 2 段（6-12 月）850g*1 罐-天猫超市-天猫 Tmall.com-上天猫，就购了-理想生活上天猫](#)
3. 3 段奶粉喂哺标准：[美国美赞臣铂睿 A2 蛋白系列幼儿奶粉\(12-36 月龄.3 段\)850g*4 罐荷兰-tmall.com 天猫](#)
4. 4 段奶粉喂哺标准：[【官方】美赞臣铂睿 4 段（3 岁或以上）850gx1 罐儿童宝宝配方奶粉-天猫超市-天猫 Tmall.com-上天猫，就购了-理想生活上天猫](#)
5. 2012 年全年中国婴幼儿奶粉市场报告：[中国婴幼儿奶粉市场报告：四大品牌占近半市场-财经--人民网 \(people.com.cn\)](#)
6. 本土婴幼儿奶粉电商消费报告：[国双发布本土婴幼儿奶粉电商消费报告|奶源 网易新闻 \(163.com\)](#)