

A/B Test

1. 平台体系建设
2. 如何做A/B测试
3. 误区&总结

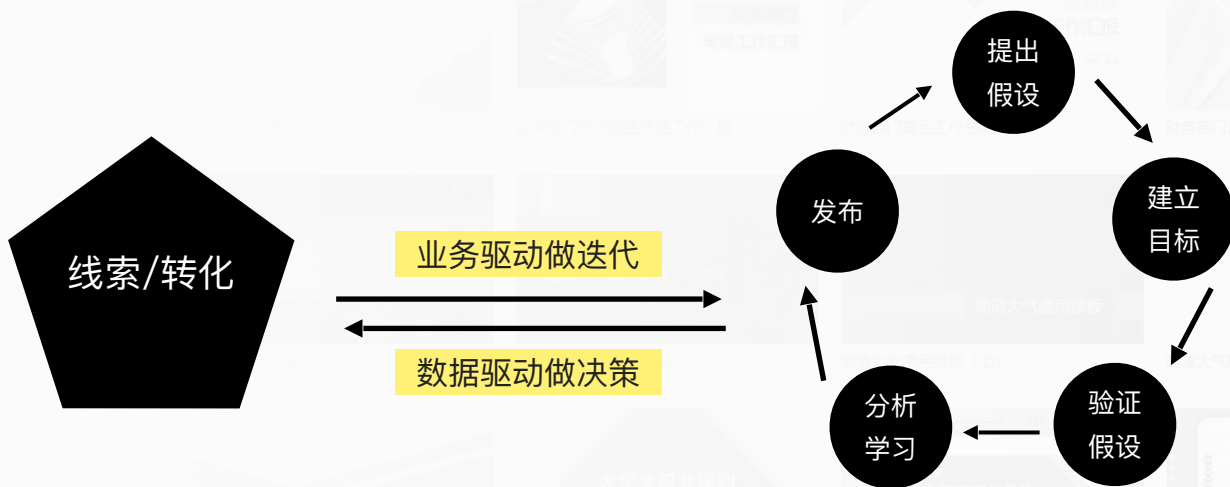
-- 刘爱霞

01

平台体系建设

A/B实验定义

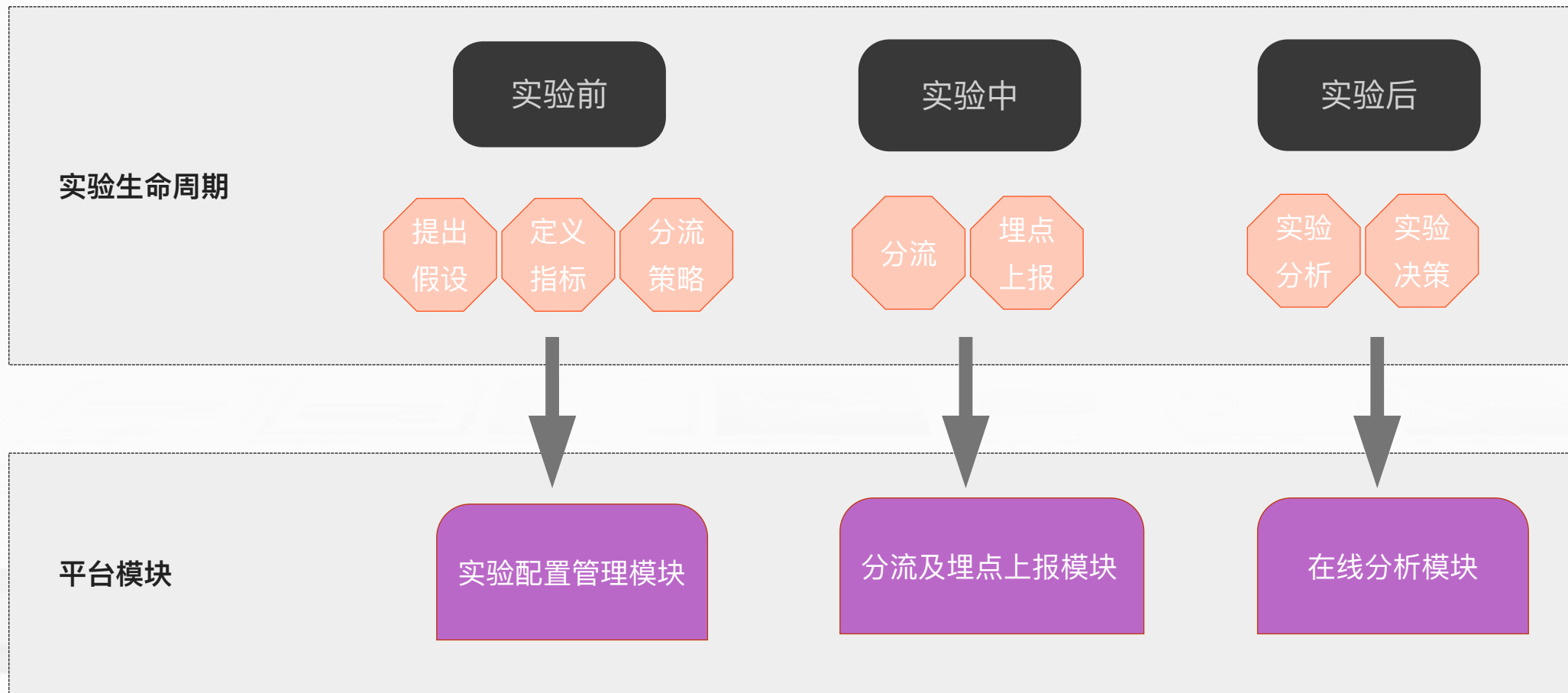
A/B实验，简单来说，就是为同一个目标制定两个版本或多个版本的方案，在同一时间维度，分别让组成成分相同（相似）的A/B群组分别采用这些版本，收集各群组的体验数据和业务数据，最后分析、评估出最好版本，正式采用。



1.1

实验生命周期

- 1. 实验前，提出该实验假设，定义实验成功的指标，确定分流策略；
- 2. 实验中，即验证假设的阶段，根据配置阶段的分流策略进行分流和埋点上报；
- 3. 实验后，进行实验分析与学习，并基于实验报告决定是否发布；



1.2

平台模块系分

实验配置管理模块

版本管理

流量选择

AB实验管理

流量正交检查

流量分层/分桶

分流及埋点上报模块

sdk接入

获取实验配置

本地分流计算

埋点上报

在线分析模块

数据指标池

显著性分析

趋势分析

对比分析

发布

02

怎样做A/B测试



2.1

现状分析

从数据中发现问题，也需要回到数据中去解决问题

每个业务都有自己的**北极星指标**，但是分析业务数据，需要确定当前最关键的改进点。这个关键改进点往往是随着时间在变化的。

eg:

1. 例如应用刚上线的时候，流量可能是关键改进点；
2. 流量上来以后，**注册率**可能成为关键改进点；
3. 再后来，客户**留存率**成为关键改进点；
4. 业务平稳后，就要整个流程分析，找到**木桶最短板 – 业务环节里最薄弱**的地方，这个地方很可能是关键改进点。

数据分析运用的比较多的便是**漏斗模型**，用户路径中每一个关键步骤都会出现流失，转化率下降的问题，只是说这个流失是否在合理范围内，和业内的平均数据相比，这个数据是否还存在可优化的空间。

eg: 营销活动

1. 进入页面 - 2. **查看利益点** - 3. **填写目标车型** - 4. 填写城市 - 5. 填写手机号码 - 6. 提交留资线索

比如正常的2->3步骤的转化率是5%，但是我们这边只有2%，这就是问题。

2.2

目标设置

核心指标 + 辅助指标 + 反向指标

核心指标：（查看利益点 -> 填写目标车型）的转化率提升，即（2% -> 5%）

辅助指标：最终的留咨线索转化率上升

反向指标：页面跳失率不能受到太大影响

2.3

流量分配

确定测试的样本量 + 确定实验周期

1. 样本量太小，会导致参数估计值的大小和符号违反经济理论和实际经验，使结果不可信。

每个测试版本需要日流量至少1000uv。

2. 没有上限的样本量？NO！

样本容量太大，势必会造成人力、物力和财力的很大浪费（成本问题）

路人甲有款产品，月活用户几百万。他要进行产品迭代，搞了个A/B测试，A版是原版，B版是新版1，C是新版2。为了追求更精确的结果，他给每个版本都分配了25万用户。经过了一周的测试，陆仁乙开心地发现C版拥有超过原版数个百分点的转化能力，同时喜闻乐见地发现B版的25万用户因为B版本体验非常不好，有数万用户删除了应用.....

3. [样本量计算器](#)

4. 实验周期

需要保证实验周期，能够刚好覆盖产品的一般使用周期。

eg：如果一个产品的使用周期是30天，比如美柚，那它的实验周期就需要大于等于30天，才能更完全的覆盖用户，从而保

2.4

结果分析

在保证核心指标明显提升的基础上，辅助指标也有一定的提升

1. 统计学分析

统计显著性达到95%或以上并且维持一段时间，实验可以结束；如果在95%以下，则可能需要延长测试时间；如果很长时间统计显著性不能达到95%甚至90%，则需要决定是否中止试验。

统计显著性指标的数值 = $1 - \text{统计学假设检验中的P值 (p-value)}$ ，P值表示样本间的差异由随机因素所导致的概率。

2. 业务分析

在统计学上的可信度确定以后，我们还要对AB测试结果做**业务分析**。业务分析是为了找到优化版本比原始版本好或差的**真正原因**，是AB测试结果分析的重要部分。我们要基于自己所从事行业的业务领域知识，结合客户心理、使用习惯、直觉本能、喜好倾向等角度来分析，把数字后面的道理挖掘出来，用来更好指导我们未来的行动。

3.其他指标说明

pv、uv、转化数、转化率、跳出率...

4. 将已经做好的实验结果进行整理分类，沉淀下来，方便后来人调取实验库

03

总结误区

出两种效果的方向，但产品和效果设计师有时想不清楚，哪一种效果用户会更喜欢呢？于是把一堆盲目效果同时上线AB。

eg：做过最多的一次AB，是有56个实验组，把各种效果参数拆开排列组合，完全以数据结论为导向来设置实验。最后上线之后，数据有好有坏，用户反馈也一大堆，然后再根据数据和用户反馈去拆解，为什么实验组A好，为什么实验组B不好；为什么年龄大一点大用户会吐槽效果1，年龄小一点大用户会吐槽效果2。

这就是想不清楚实验目的的AB最终会带来的困境。所以说，还是要基于产品核心能力和市场，有最起码的判断和假设。再去用AB去验证这个假设。

2. 指标找错，一切全费

设定**北极星指标**

3. AB的结果是好的，一全量并没有什么卵用？（本质为增长乏力）

1. 每次AB带来的变动很小，最终反应在结果上数据的波动非常小
2. 做撬动市场的新功能，但最终没能打成战略目标的提升

这时候，市场在提醒你，也许应该换一种思路来做产品。

4. 一定要做的改变，需要做AB吗？

3.1

总结

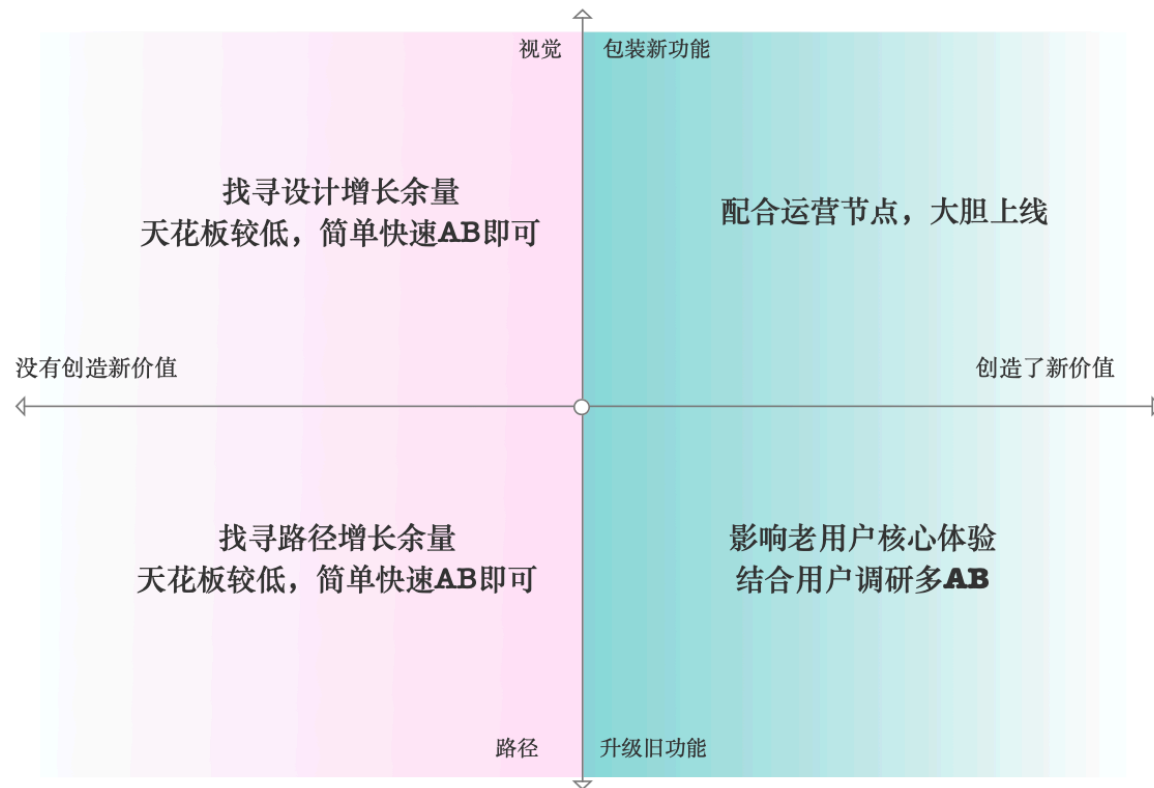
1. 忘记AB

有人在提AB，都需要询问

- (1) AB的目的是什么？
- (2) 你觉得哪个方案会更好？
- (3) 你是否做过市场分析和用户调研？

这三个问题能够有效帮产品梳理需求的意义。如果仅仅以“我拿不准哪个方案会更好”来进行AB，那么最终团队会陷入无限AB无限浪费资源的死循环。

2. 寻找业务的AB Test规律 (增长类和核心类)



THANKS

多多指教