SENSORS Analytics

初创公司构建数据分析平台

桑文锋@SensorsData

SENSORS
Analytics

Why, What & How

- 数据分析平台的概念
- 现有解决方案
- 推荐方案

SENSORS Analytics

数据分析的目的

Why

1

运营监控

- 上个月销售表现如何?
- 近期活跃用户数变化趋势?

2

产品改进

- 用户粘性如何?
- 新功能的使用情况怎样?

3

商业决策支持

- 是否要开展天津地区业务?

SENSORS Analytics

数据分析平台的概念

What

SENSORS Analytics

如果现状是...

SENSORS Analytics

工程师老王负责处理所有跑数据的需求

SENSORS Analytics

如果现状是...

每个需求都是一个新的脚本?

SENSORS Analytics

工程师老王负责处理所有跑数据的需求，直到有一天…

SENSORS Analytics

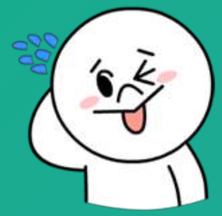
如果现状是...

SENSORS Analytics

公司终于有了数据仪表盘...

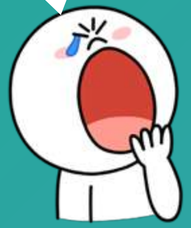
真是高大上！

明明昨天
一个机房挂了，但是流量还在涨……



用户量下跌了，但是根本看不出来问题来自哪里……

这些泛泛的指标很难指导决策，不看也罢……

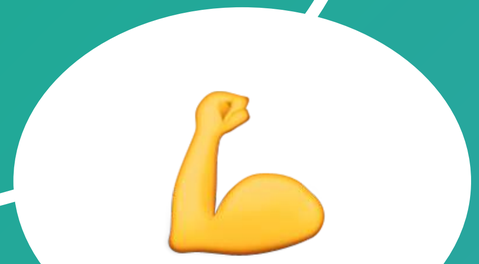


SENSORS Analytics

人人都是数据分析师

Self-service Data Analytics

让参与业务的人真正掌握数据！



SENSORS Analytics

数据分析平台——

- 适应公司的快速发展
- 将繁杂数据抽象为简洁的模型
- 让每个业务参与者能够用数据驱动决策
- 数据可反馈于线上

SENSORS Analytics

数据分析解决方案

现有常用方案

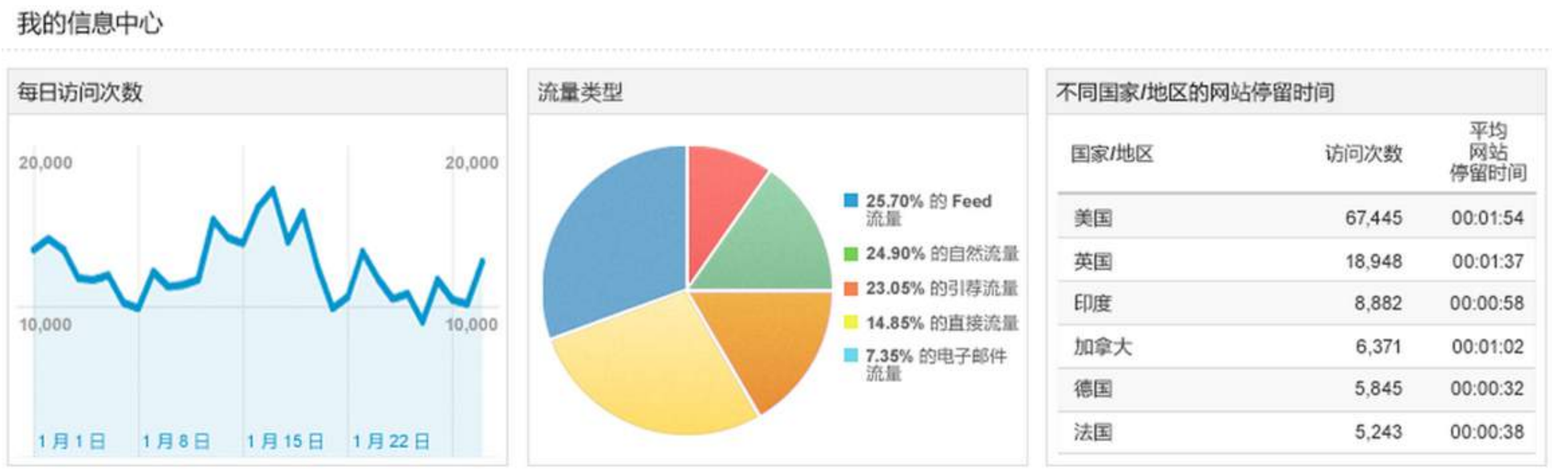
1. 第三方统计服务
2. 业务数据库写SQL
3. 基于日志写统计脚本

SENSORS Analytics

现有常用方案

1 第三方统计服务

Google Analytics　友盟 UMENG　Baidu 统计



SENSORS Analytics

现有常用方案

1 第三方统计服务

Google Analytics　友盟 UMENG　Baidu 统计

SENSORS Analytics

现有常用方案

1 第三方统计服务

Google Analytics　友盟 UMENG　Baidu 统计

SENSORS Analytics

现有常用方案

1 第三方统计服务

好处

- 使用简单
- 免费

SENSORS Analytics

现有常用方案

1 第三方统计服务

👎 不足

- 无法与业务数据交叉分析
- 分析能力较弱，无法覆盖深度分析
- 指标无法自定义
- 数据无法取回
- 数据安全存在顾虑

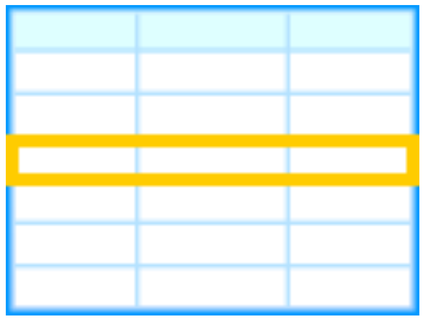
SENSORS Analytics

现有常用方案

2 业务数据库写SQL

业务数据库

SQL



导出数据

分析处理

SENSORS Analytics

现有常用方案

2 业务数据库写SQL

好处

- 可根据需求灵活定制
- 数据准确、实时
- 可分析业务数据

SENSORS Analytics

现有常用方案

2 业务数据库写SQL

👎 不足

- 历史状态被覆盖

业务数据库

数据仓库

时间

SENSORS Analytics

现有常用方案

2 业务数据库写SQL

👎 不足

- 计算能力有限，无法水平扩展
- 开发维护代价大
  - 需额外开发工作量
  - 查询逻辑随着业务的演进复杂化，不好维护（SQL、脚本、结果数据）
  - 和业务数据无法解耦
  - 随分析需求增加字段、数据表

SENSORS Analytics

现有常用方案

3 基于日志写统计脚本

python

SENSORS Analytics

现有常用方案

3 基于日志写统计脚本

好处

- 与业务数据库解耦

SENSORS Analytics

现有常用方案

3 基于日志写统计脚本

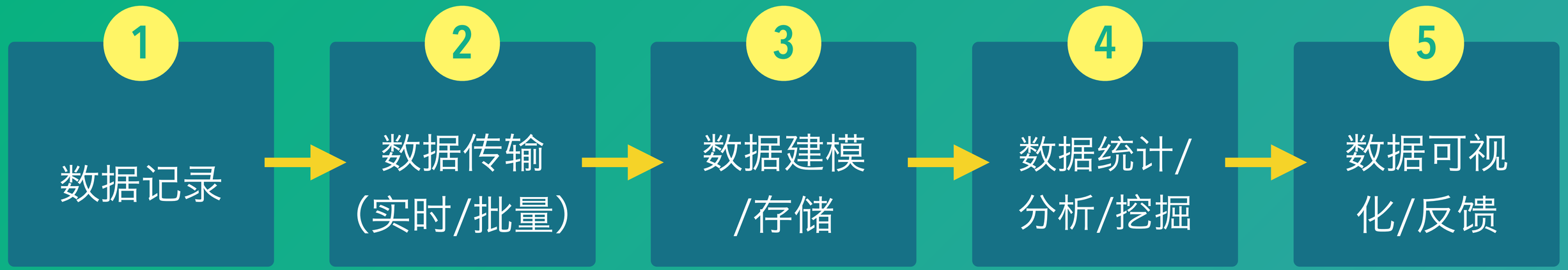
👎 不足

- 开发效率低（2天/个，重复开发）
- 准确性无法保证
- 计算能力有限
- 有技术门槛
  - 打好日志是一件很难的事情
  - 数据流难以管理

SENSORS Analytics

数据分析平台的推荐方案

How



SENSORS Analytics

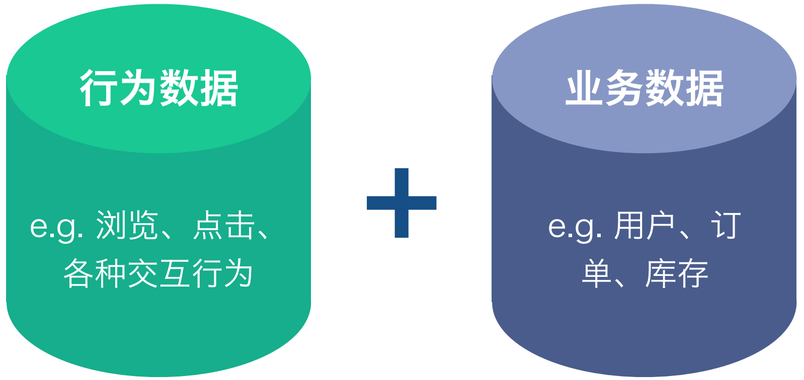
1 数据记录

- 生成高质量的源数据
  - 全
  - 准

SENSORS Analytics

1 数据记录

- 数据类型
  - 行为数据
  - 业务数据：用户、订单、库存
- 数据规范
  - 行为数据：操作系统、应用版本、是否WIFI、屏幕尺寸、设备型号、商品ID、商品价格等。
  - 用户属性数据：性别、年龄、婚姻状况、注册时间、收入级别、是否有小孩等。



SENSORS Analytics

1 数据记录

- 数据格式
  - 非格式化文本 Vs. Json、Thrift、Protocol Buffer、Avro
- 数据采集点
  - 尽量在后端打
  - 前端打（压缩、加密、批量）
- 数据落地
  - 写网络
  - 写本地文件

SENSORS Analytics

2 数据传输

- 需关注的问题
  - 时效性（实时？批量？）
  - 可靠性（丢？重？）
  - 扩展性
- 方案
  - FTP
  - Kafka
  - Scribe、Flume

SENSORS Analytics

3 数据建模/存储

- 数据模型抽象
  - Event:
    - Event Type + Properties + UserID
  - User Profile:
    - UserID + Properties（年龄、所在地、Tag等）
- Event：记录所有的历史状态变更

SENSORS Analytics

3 数据建模/存储

- ETL (Extract, Transform and Load)
  - ID-Mapping
  - Merge
  - 批量 or 实时

SENSORS Analytics

3 数据建模/存储

- 存储
  - 单机文件
  - 关系型数据库（Mysql、Vertica、Teradata）
  - Nosql (HBase、MongoDB)
  - HDFS

SENSORS Analytics

4 数据统计/分析/挖掘

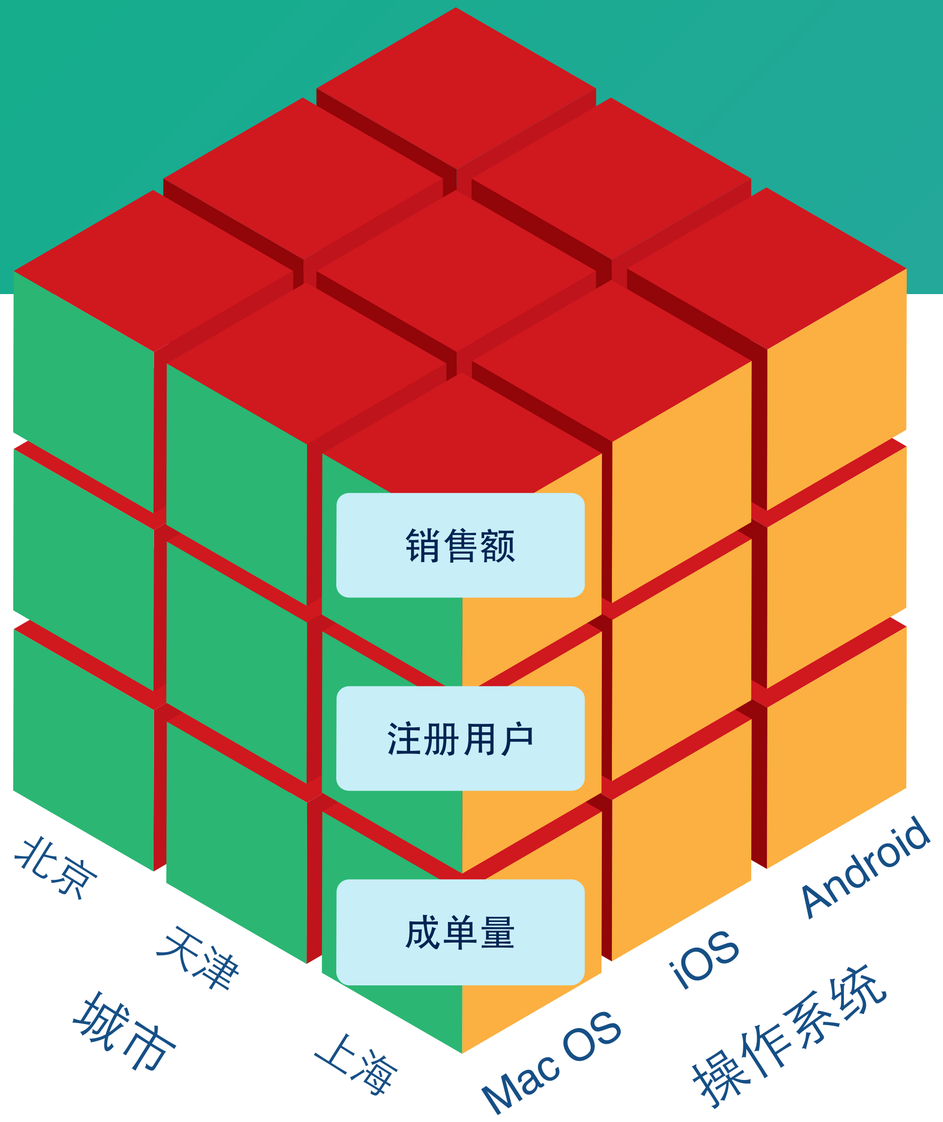
- 批处理

- 交互式

SENSORS Analytics

4 数据统计/分析/挖掘

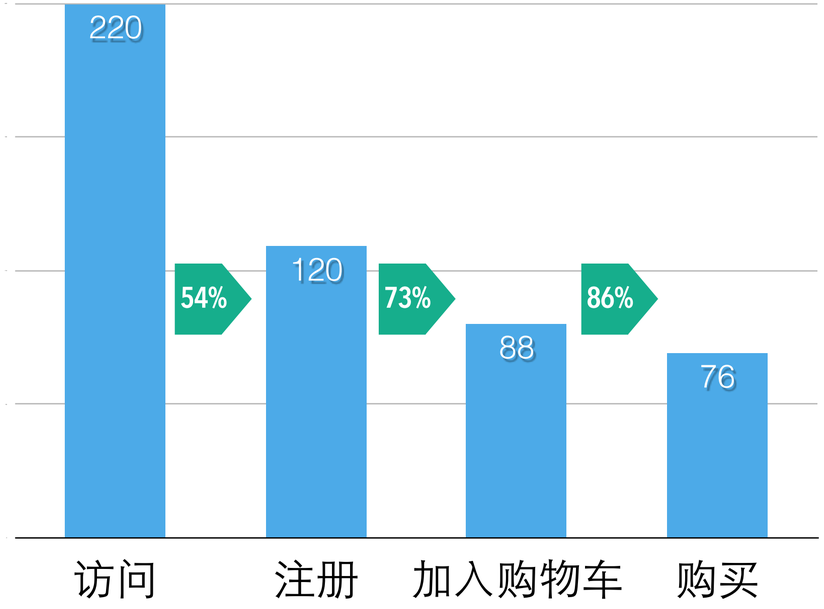
- OLAP (Online Analytical Processing)
  - 维度
  - 指标
  - 数据魔方



SENSORS Analytics

4 数据统计/分析/挖掘

- 分析模型
  - 漏斗
  - 留存

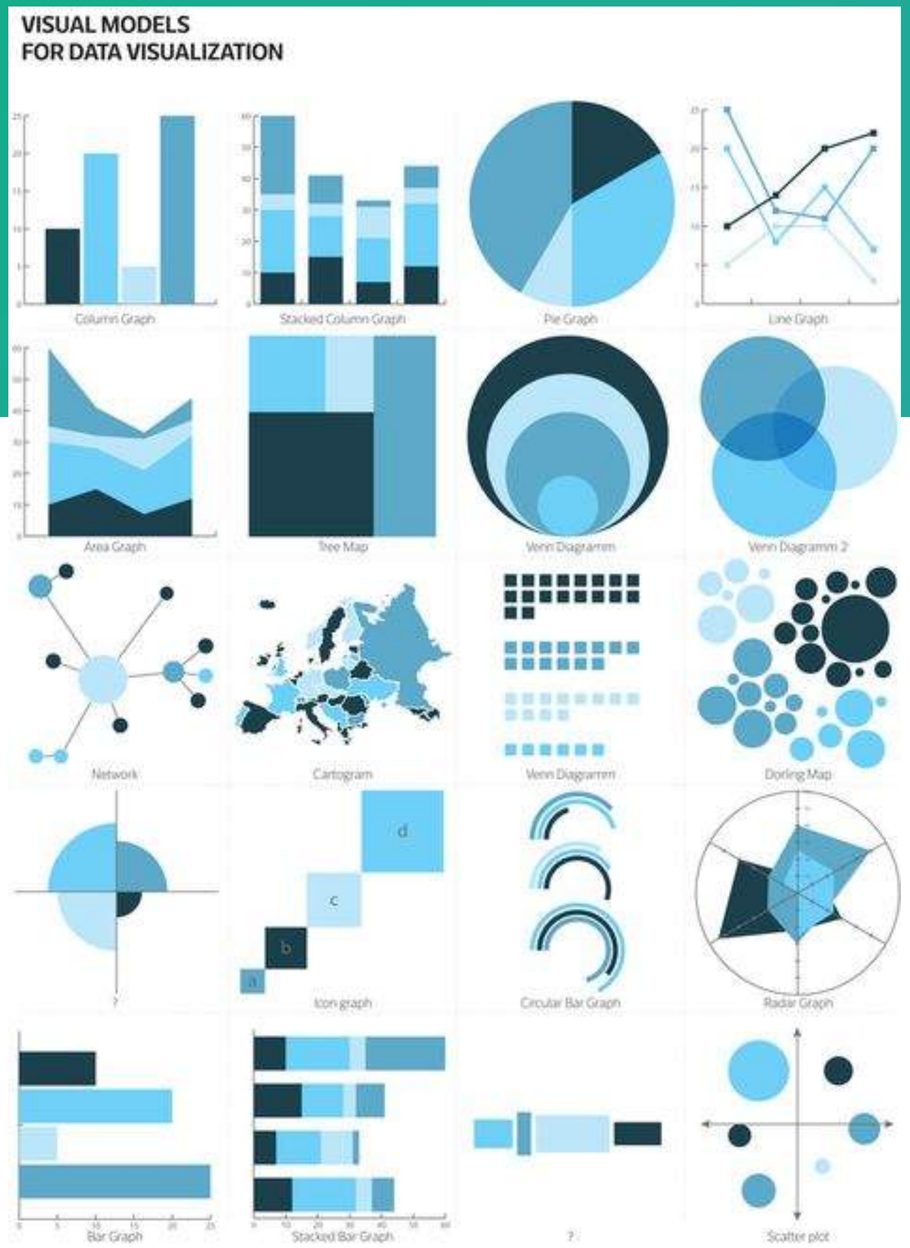


Date	Installs	1	2	3	4	5	6	7	8	9	10	11	12
May 23	10,021	43%	42%	31%	28%	26%	22%	18%	16%	9%	7%	6%	3%
May 24	10,091	42%	41%	36%	34%	27%	24%	17%	14%	8%	7%	6%	
May 25	11,232	46%	41%	34%	32%	29%	23%	19%	13%	10%	8%		
May 26	13,164	43%	40%	36%	30%	22%	18%	16%	16%	9%			
May 27	14,081	41%	38%	33%	31%	29%	22%	17%	14%				
May 28	15,755	46%	42%	36%	33%	24%	24%	19%					
May 29	16,072	46%	43%	37%	34%	26%	23%						
May 30	18,243	42%	40%	33%	32%	27%							
May 31	19,001	46%	41%	33%	31%								
Jun 1	19,687	51%	40%	36%									
Jun 2	20,201	53%	41%										
Jun 3	21,444	60%											

SENSORS Analytics

5 数据可视化/反馈

- 展现方式
  - 曲线
  - 柱状图
  - 饼状图
  - 热力图
  - 地域分布



图片来源：https://www.behance.net/gallery/Visual-Models-for-Data-Visualization/4078769

SENSORS Analytics

5 数据可视化/反馈

- 可视化工具
  - OpenCharts
  - HighCharts
  - ECharts
  - Tableau Software
  - Oracle BIEE

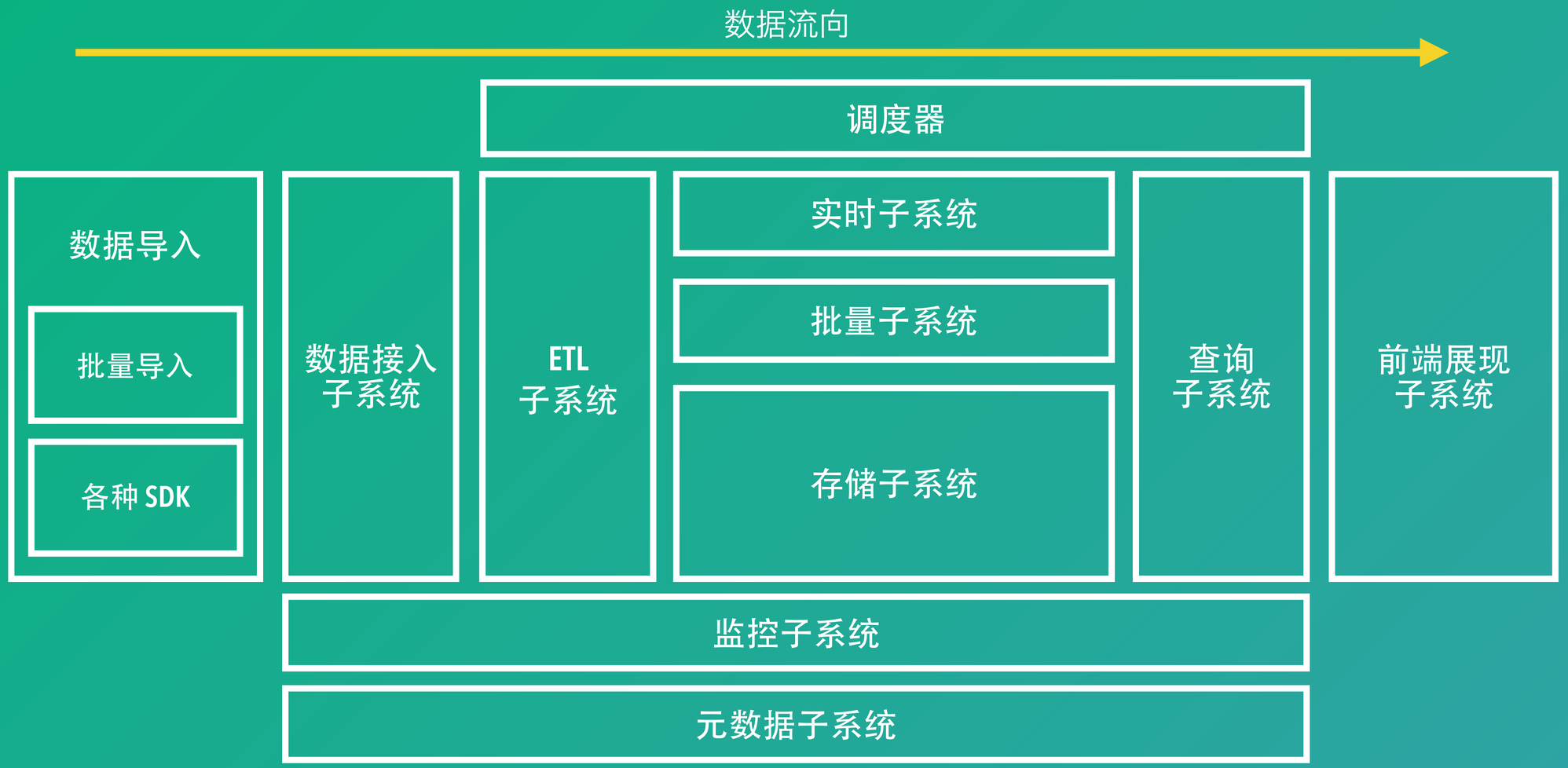
SENSORS Analytics

5 数据可视化/反馈

- 数据分析的结果直接反馈到产品系统中，提升产品体验（BI只是数据分析的很小一部分）
- 反馈方式
  - 推送
  - 个性化推荐
  - 风控
  - CRM集成

SENSORS Analytics

总体架构



SENSORS Analytics

开发代价

- 合适的人
- 3-5名数据工程师
- 开发6个月

SENSORS Analytics

要点

完备的数据

让团队轻松
获取数据

定义关键指标

SENSORS
Analytics

SENSORS Analytics

产品定位与特点

- 我们的定位：
  - 私有化部署的大数据分析产品
- 我们的特点：
  - 私有化部署
  - 强有力的多维分析
  - 属于你的数据仓库

SENSORS Analytics

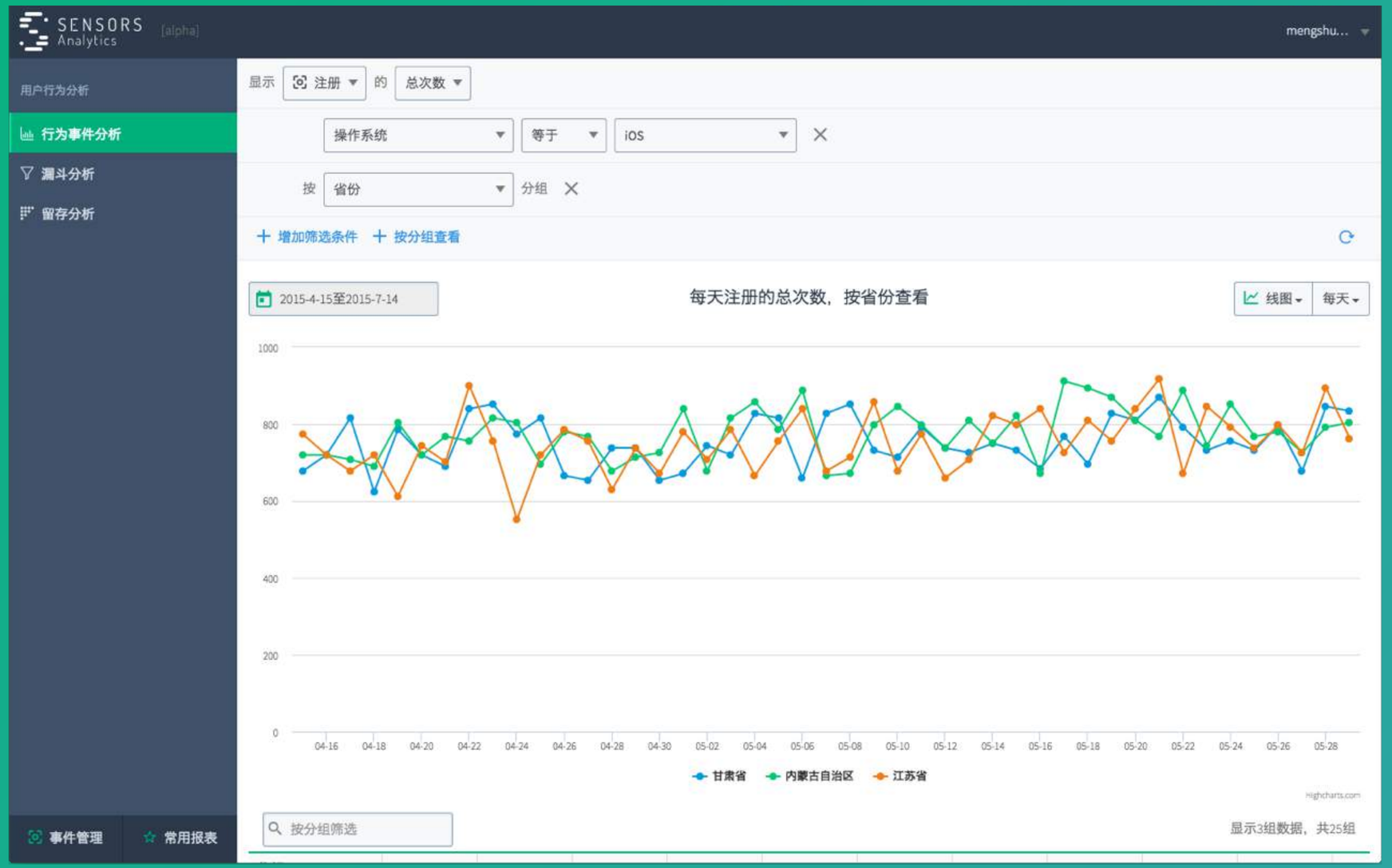
演示网址

demo.sensorsdata.cn

（可联系微信sangwf申请试用）

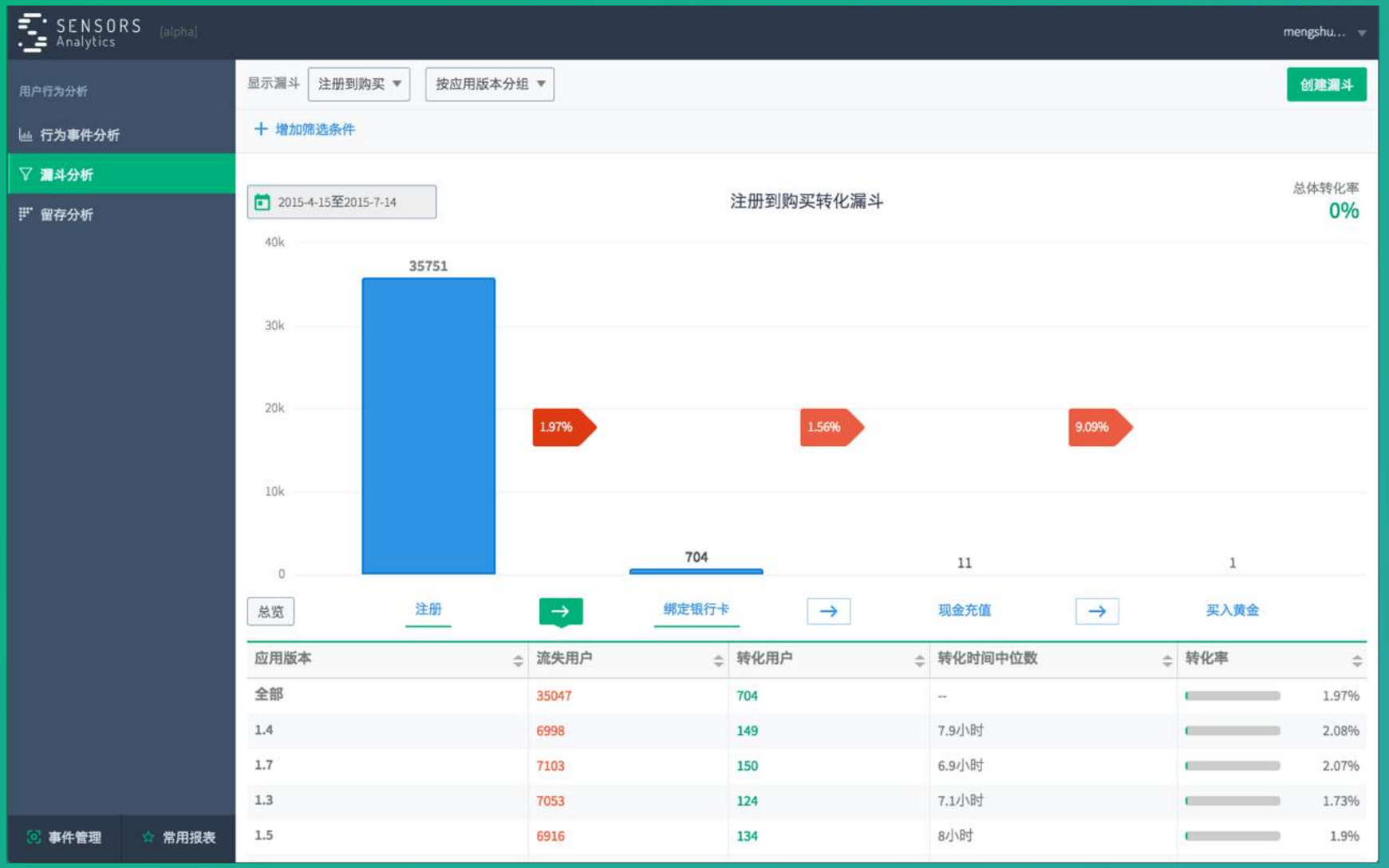
SENSORS Analytics

功能1：用户事件分析



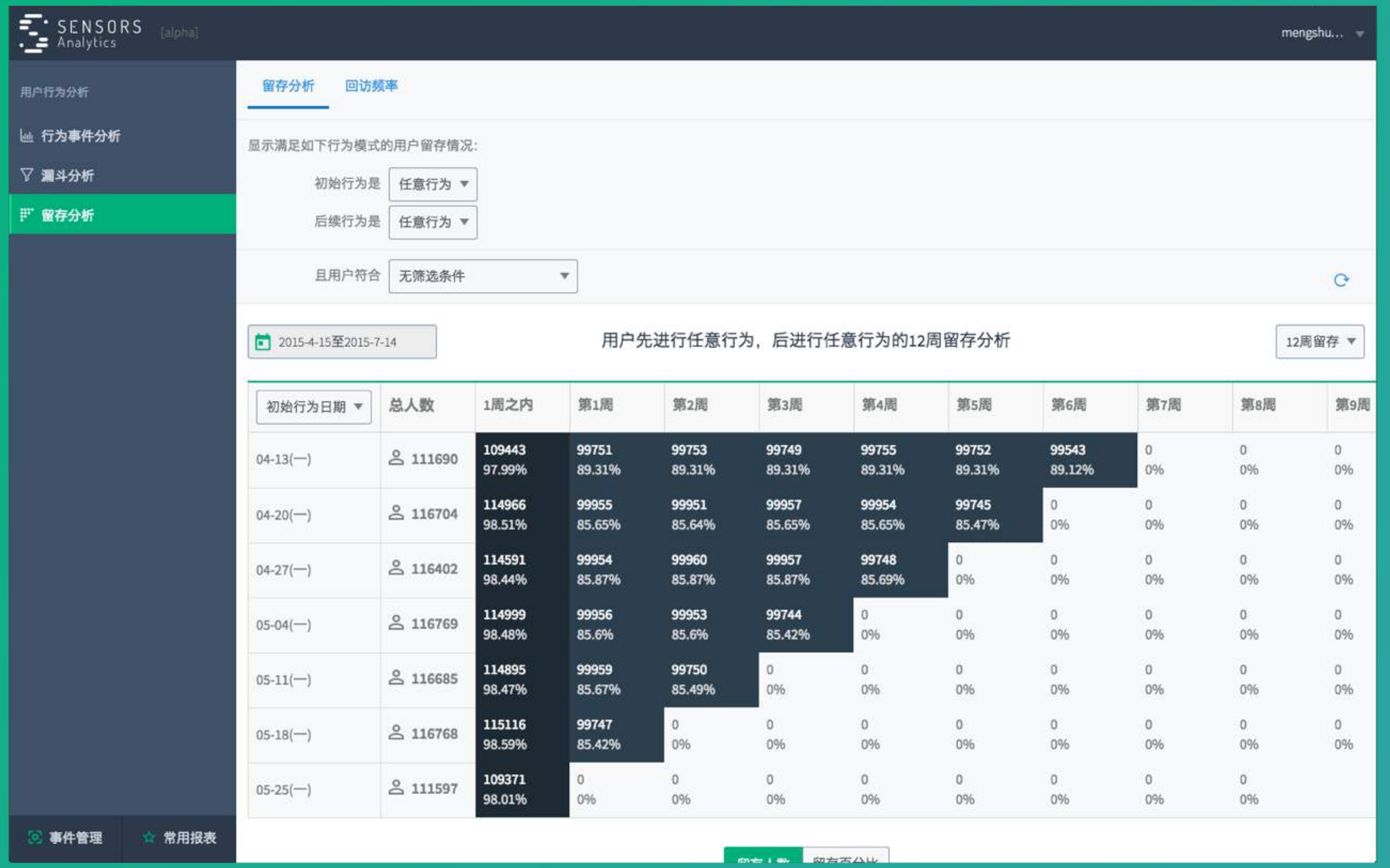
SENSORS Analytics

功能2：漏斗分析



SENSORS Analytics

功能3：留存分析



初始行为日期	总人数	1周之内	第1周	第2周	第3周	第4周	第5周	第6周	第7周	第8周	第9周
04-13(一)	111690	109443 97.99%	99751 89.31%	99753 89.31%	99749 89.31%	99755 89.31%	99752 89.31%	99543 89.12%	0 0%	0 0%	0 0%
04-20(一)	116704	114966 98.51%	99955 85.65%	99951 85.64%	99957 85.65%	99954 85.65%	99745 85.47%	0 0%	0 0%	0 0%	0 0%
04-27(一)	116402	114591 98.44%	99954 85.87%	99960 85.87%	99957 85.87%	99748 85.69%	0 0%	0 0%	0 0%	0 0%	0 0%
05-04(一)	116769	114999 98.48%	99956 85.6%	99953 85.6%	99744 85.42%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%
05-11(一)	116685	114895 98.47%	99959 85.67%	99750 85.49%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%
05-18(一)	116768	115116 98.59%	99747 85.42%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%
05-25(一)	111597	109371 98.01%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	0 0%	

SENSORS Analytics

没有银弹

- 数据与业务系统紧密相关，无孔不入
- SensorsAnalytics 为中等规模（>= A轮）的互联网公司解决核心数据分析问题，且具有足够的开放性

SENSORS Analytics

Q&A

Q 小公司是否要一步到位?

A 数据源的打印是最关键的一步

SENSORS Analytics

Q&A

Q 老板、产品、运营和技术如何配合做好数据平台建设?

A 老板支持，肯投入；产品、运营抽象好需求，主动学习；技术授人以渔。

SENSORS Analytics

Q&A

Q 电商运营如何利用数据?

A 大众点评case。

SENSORS Analytics

Q&A

Q 如何建立数据的技术及分析部门?

A 人员配备（平台、策略、策略工程化、BI工程师）

SENSORS
Analytics

微信：sangwf

SENSORS Analytics

试用流程

	业务介绍	需求梳理	需求确认	试用准备	开始试用
神策	介绍SA功能	1）需求梳理建议 2）需求梳理模板	1）确认需求满足情况 2）给出event梳理模板	1）给出试用SA的准备工作清单（机器、SDK、日志格式样例）	1）产品使用指导
用户	介绍业务、及现有数据统计情况	1）收集运营、产品、市场、COO等岗位的现有报表信息 2）相关数据需求方的潜在需求以及日常工作中的数据使用场景 3）梳理形成指标/维度表格	1）根据报表需求和event梳理模板，梳理产品event list 2）确认日志升级计划	1）SA部署环境准备 2）日志升级准备 3）历史数据导入准备	1）开始数据分析 2）问题、效果反馈
耗时	2小时	1周	2天	1周	2-4周