

SEM Bigdata  
森马大数据

×



大数据项目组  
1996-2019 SEMIR  
BIG DATA TEAM

×

FUN Elysian  
Fields

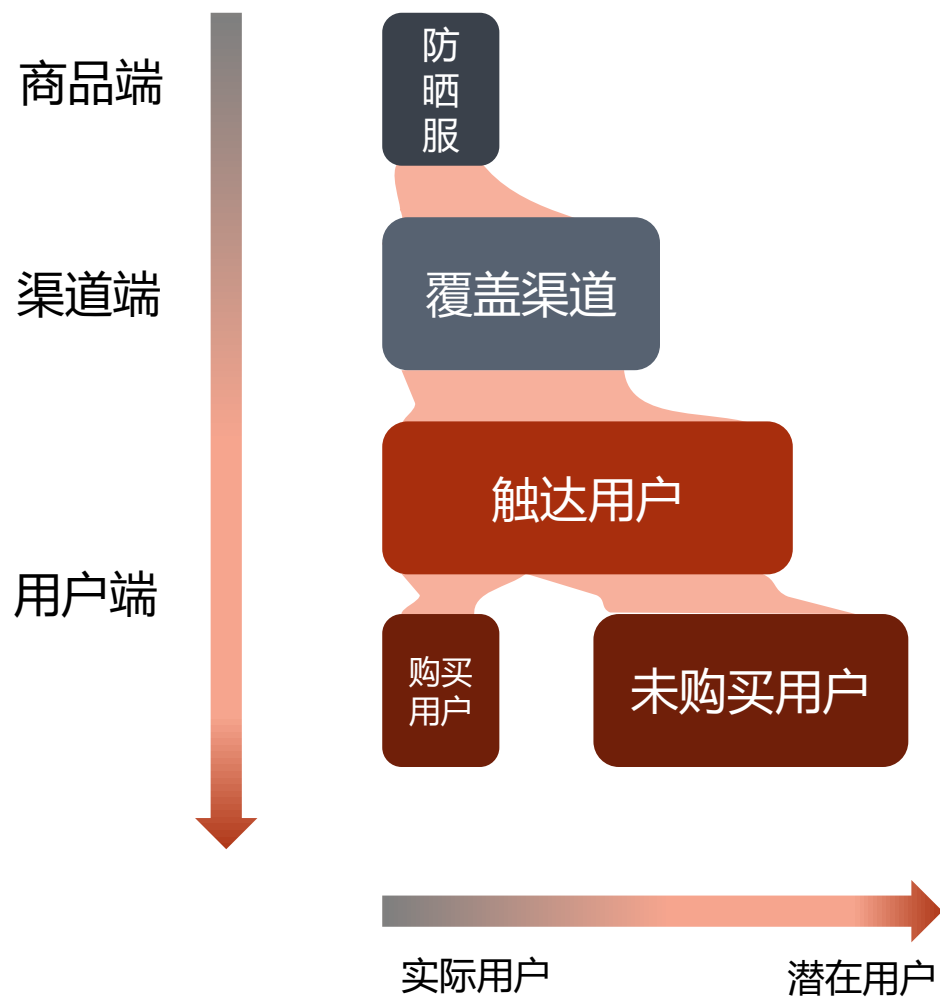
# SUNSCREEN

Clothes  
防晒服专题分析

2019.05

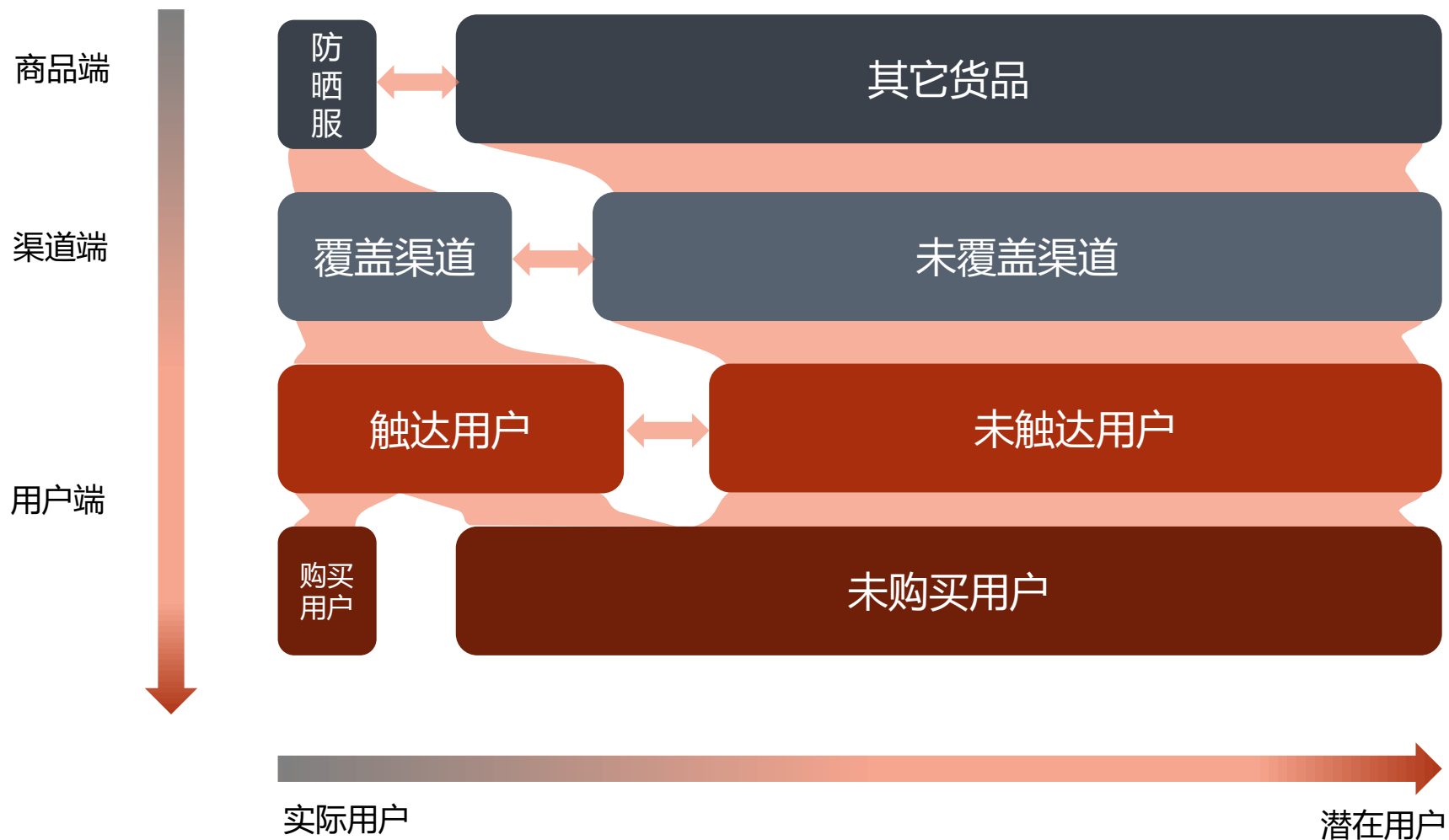
大数据项目组-郑乐君  
[www.semirbigdata.cn](http://www.semirbigdata.cn)

# 分析概述



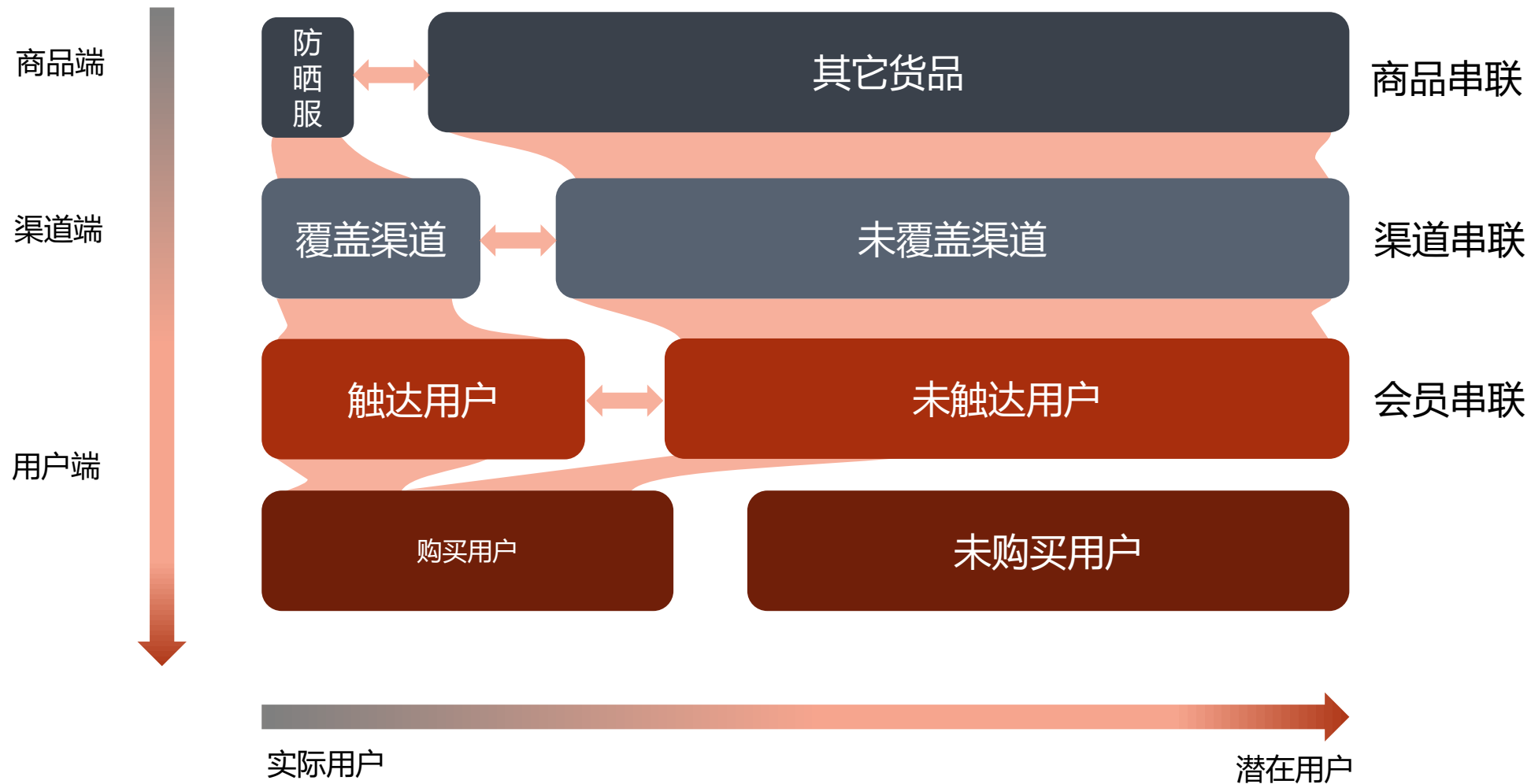
流程：从商品传递到用户  
目的：增加购买用户

# 分析概述



流程：从商品传递到用户  
目的：增加购买用户  
手段：  
货品串联  
渠道串联  
用户串联

# 分析概述



# CONTENTS

**01**

**商品串联**

**02**

**渠道串联**

**03**

**会员串联**

# 计划新增项目：商品开发面辅助分析

- 1, 渠道串联：品类潜在门店的分布，辅助开发端口
- 2, 会员串联：品类潜在消费者的标签，辅助开发端口
- 3, 商品串联：利用关联搭配，找搭配，辅助模块开发

商品面猜想：

防晒服重功能还是重搭配

地域对品类的关联

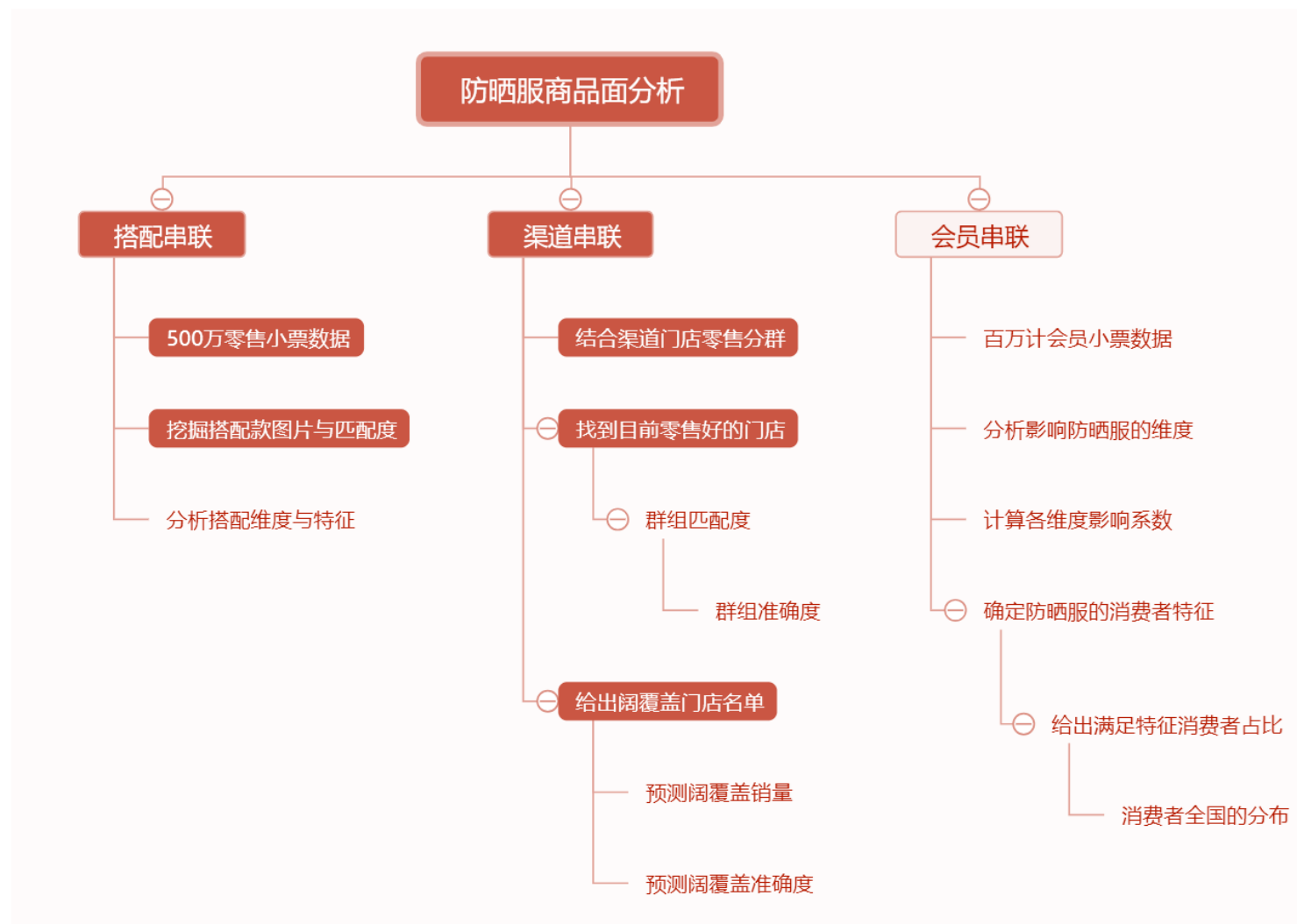
## 数据维度

数据时间：2019.5.21前所有小票数据

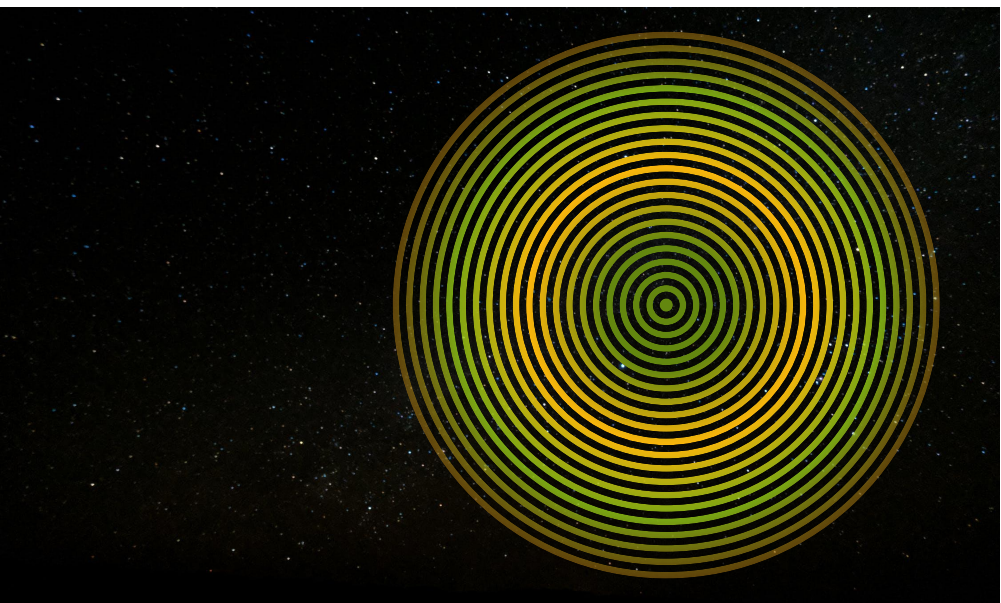
其它限制：产品季节-19#S1&S2&S3&S4，小品牌-森马，渠道-除海外

数据体量：6323975

# 分析思路







/01

商品串联：

利用关联搭配，找搭配，辅助模块开发



# 商品串联： 利用关联搭配，找搭配，辅助模块开发

 款号： 13039080002		提升度 72.22%	组合数 16
		69.57%	36
		59.58%	25
		45.03%	27
		41.79%	21

数据时间： 2019.5.21前所有小票数据  
其它限制： 产品季节-19#S3&S4， 小品牌-森马， 渠道-除海外  
数据体量： 2745268

# 商品串联： 利用关联搭配，找搭配，辅助模块开发



款号：13029080066



组合数	提升度
45	56.79%
41	47.96%
23	32.84%
28	32.99%
18	30.55%

数据时间：2019.5.21前所有小票数据  
其它限制：产品季节-19#S1&S2，小品牌-森马，渠道-除海外  
数据体量：3578707

# 商品串联： 利用关联搭配，找搭配，辅助模块开发

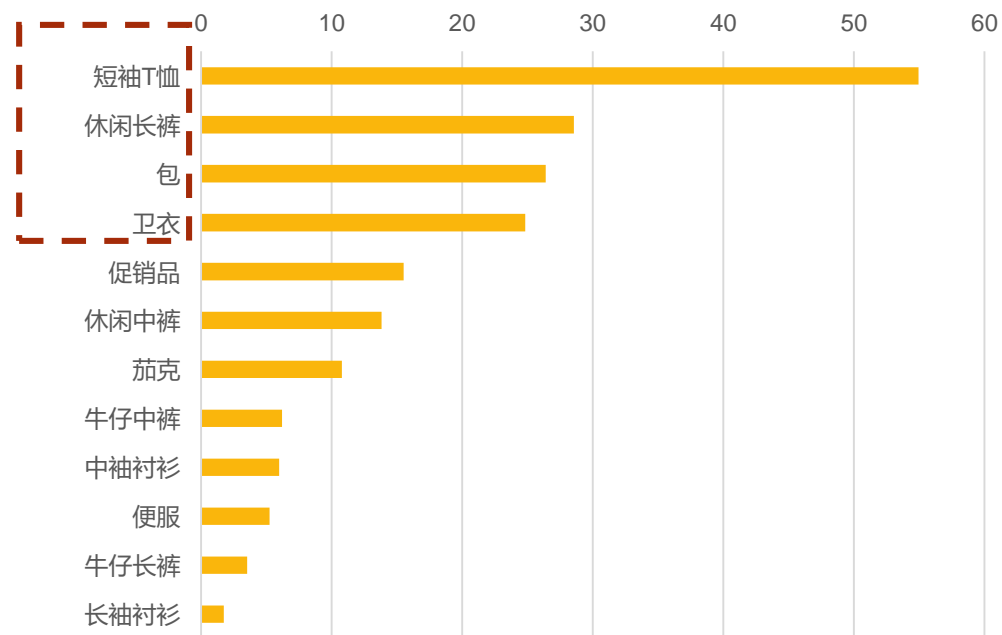


组合数	提升度
121	89.77%
43	66.89%
31	62.81%
27	55.96%
32	54.06%

数据时间：2019.5.21前所有小票数据  
其它限制：产品季节-19#S1&S2，小品牌-森马，渠道-除海外  
数据体量：3578707

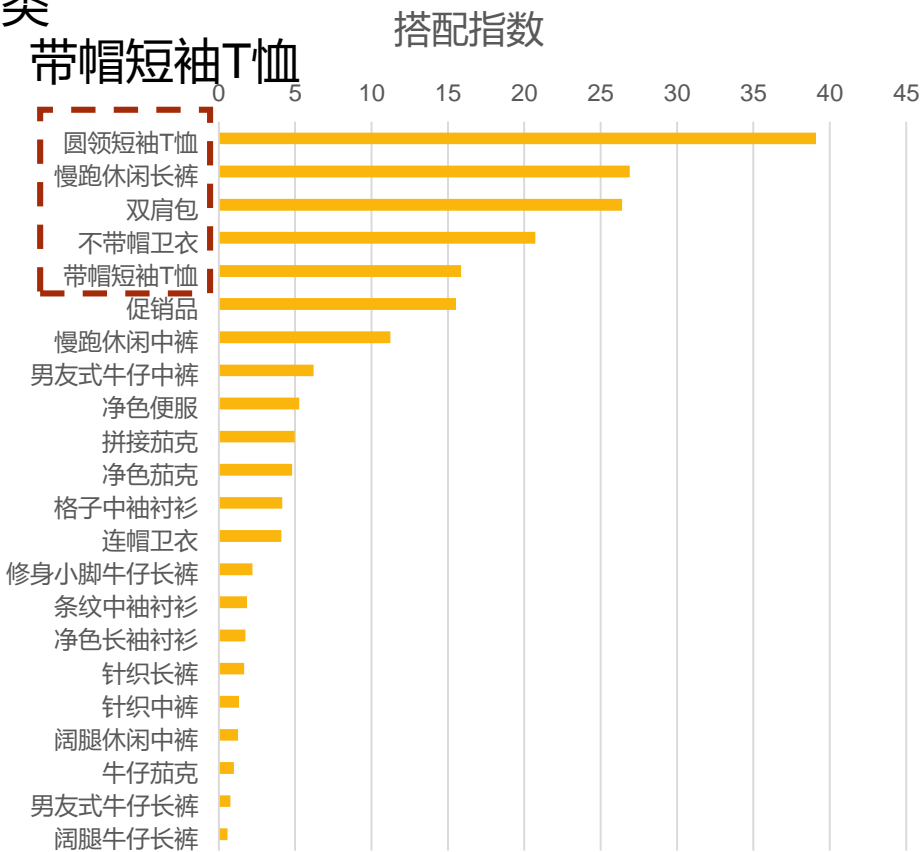
# 搭配中类

从商品串联的角度看，对防晒服搭配提升的较大的品类  
短袖T恤，休闲长裤，包，卫衣



# 搭配小类

从商品串联的角度看，对防晒服搭配提升的较大的品类  
圆领短袖T恤，慢跑休闲长裤，双肩包，不带毛卫衣，带帽短袖T恤



数据时间：2019.5.21前所有小票数据  
其它限制：产品季节-19#S1&S2&S3&S4，小品牌-森马，渠道-除海外  
数据体量：6323975



# 最佳商品组合

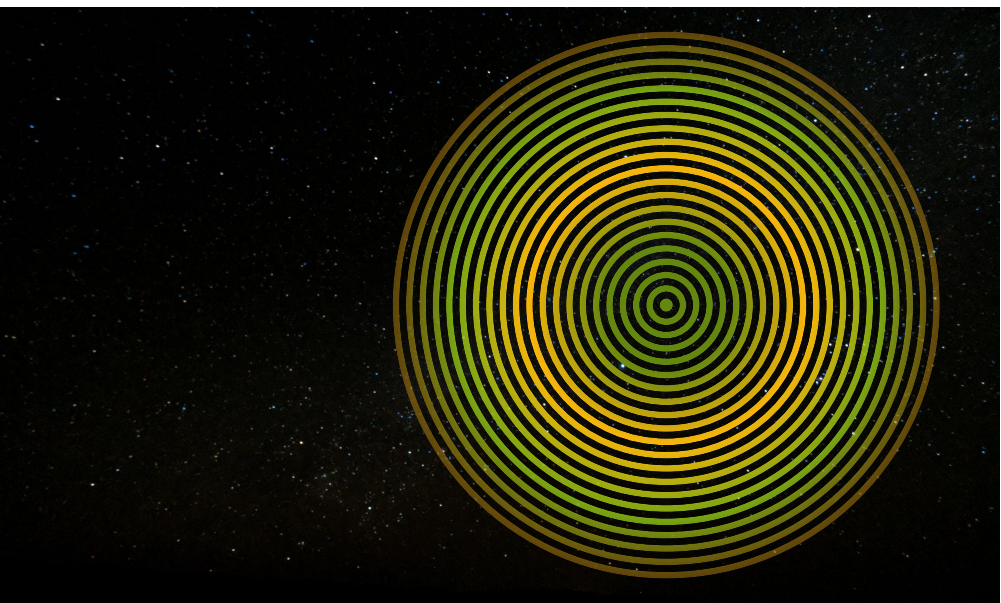


防晒服



圆领短袖T恤  
慢跑休闲长裤  
双肩包  
不带帽卫衣  
带帽短袖T恤  
慢跑休闲中裤

提升效果预计为**5.589**  
较其它搭配售卖概率提升度



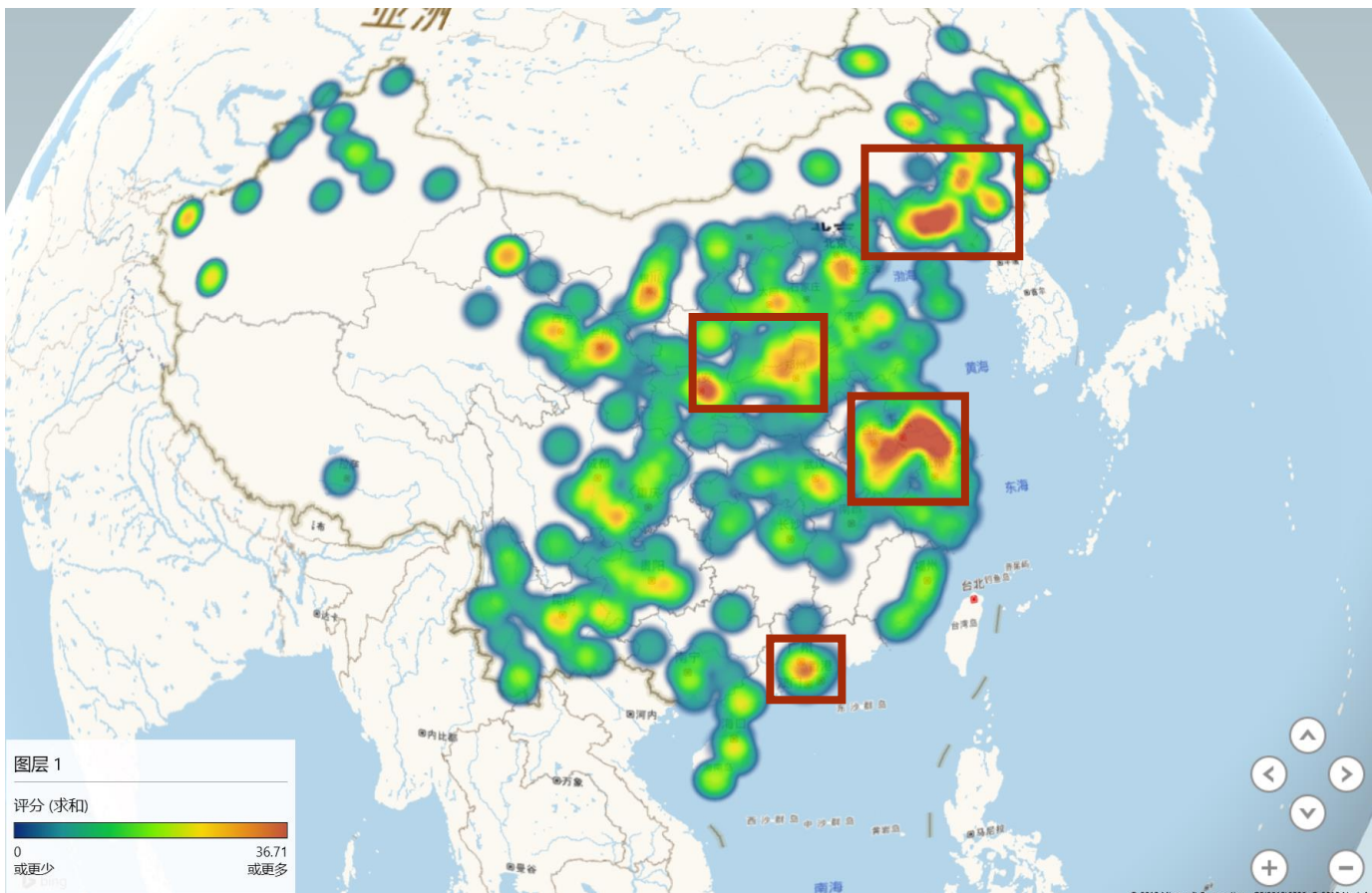
/02

## 渠道串联：

品类潜在门店的分布，和零售预测，辅助开发端口



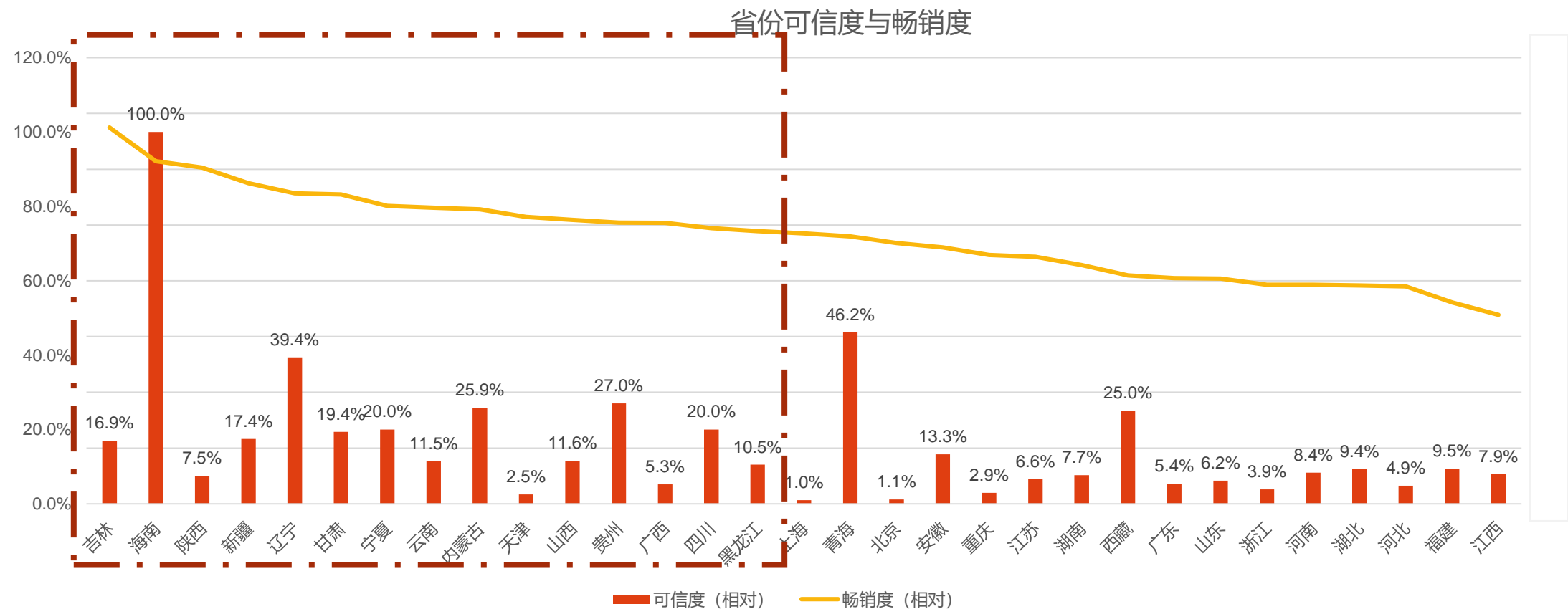
# 目前零售好的分布



左图显示的是防晒服适合零售的分布热力图

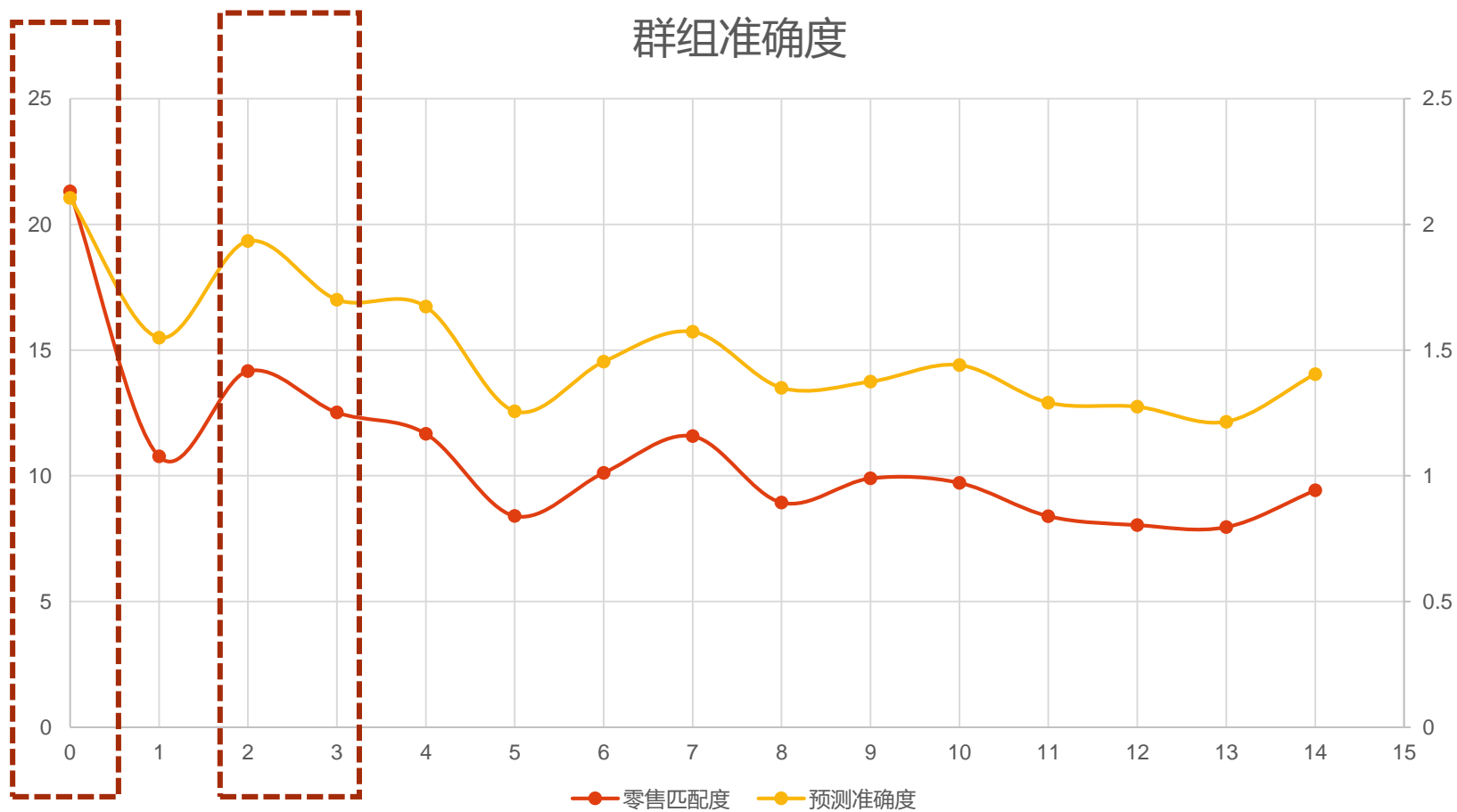
东北部，中西部，东部，南部  
部分区域相对零售更好

# 省份可信度与畅销度



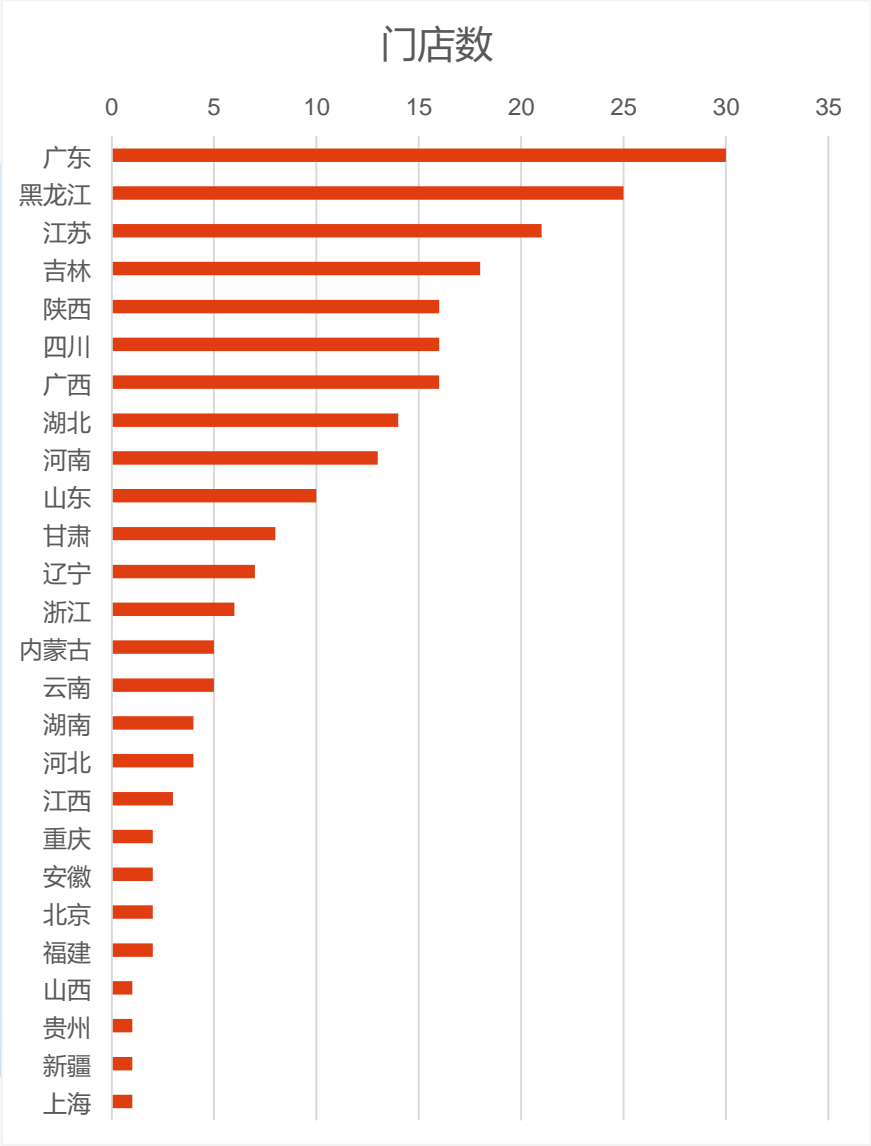
省份按畅销度排名，但是不同省份的覆盖率不同所以准确度也会有所差异  
推荐关注省份：如吉林，陕西，新疆，辽宁，甘肃，云南，内蒙，山西，贵州

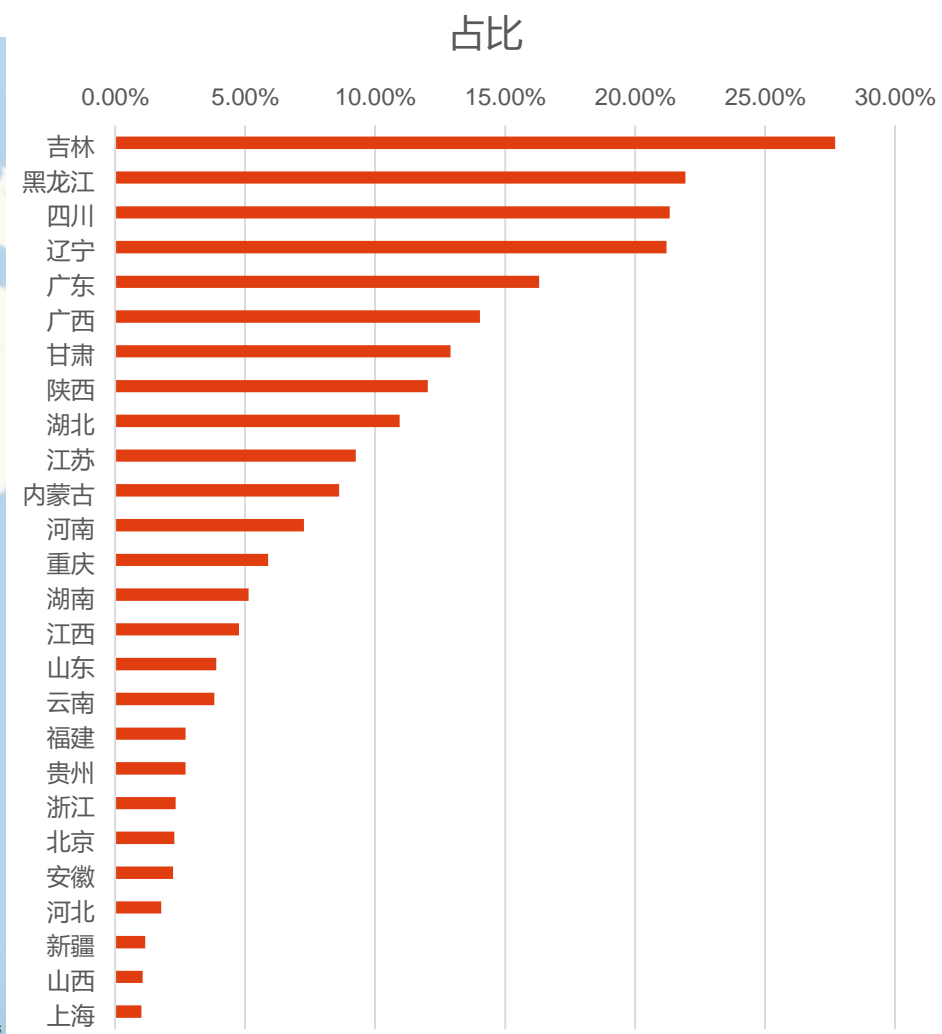
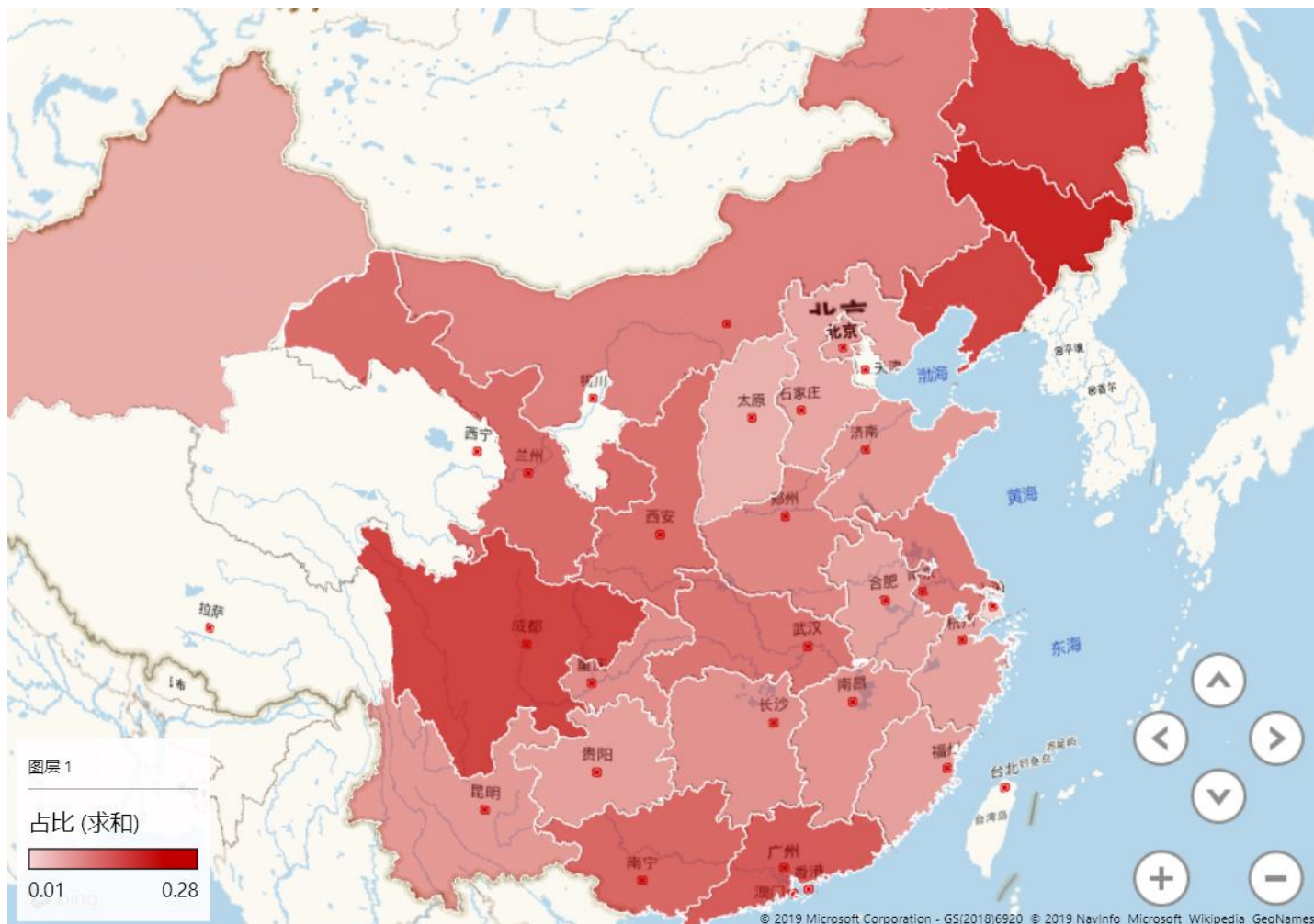
# 群组分布



推荐群组0, 2, 3, 号群组

# 群组门店数量分布



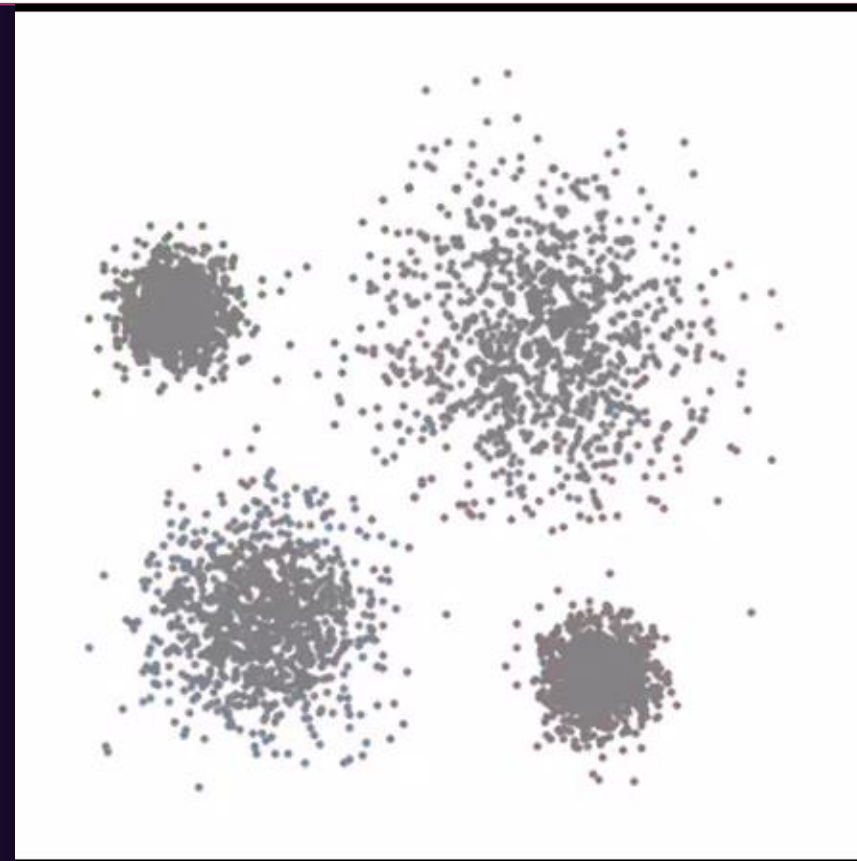




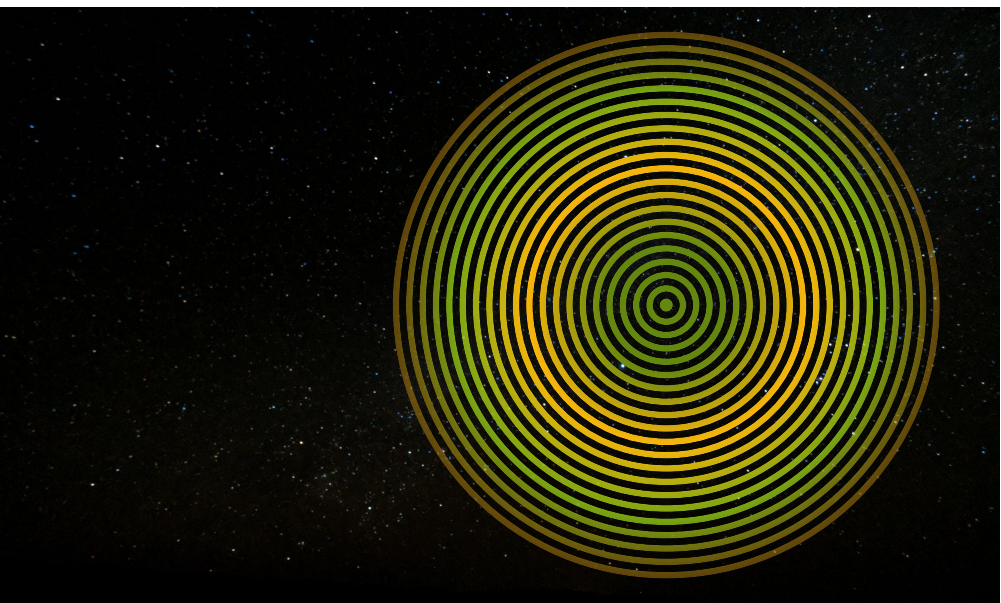
## 群组计算：

以上如何预测相似门店的

采用优化后的Mean-Shift算法，智能化  
计算规律宽度，既保证数据的稳定性又保  
证可用性。



Mean-Shift 是一种基于滑动窗口的聚类算法。也可以说它是一种基于质心的算法，这意思是它是通过计算滑动窗口中的均值来更新中心点的候选框，以此达到找到每个簇中心点的目的。然后在剩下的处理阶段中，对这些候选窗口进行滤波以消除近似或重复的窗口，找到最终的中心点及其对应的簇。看看下面的图解



/03

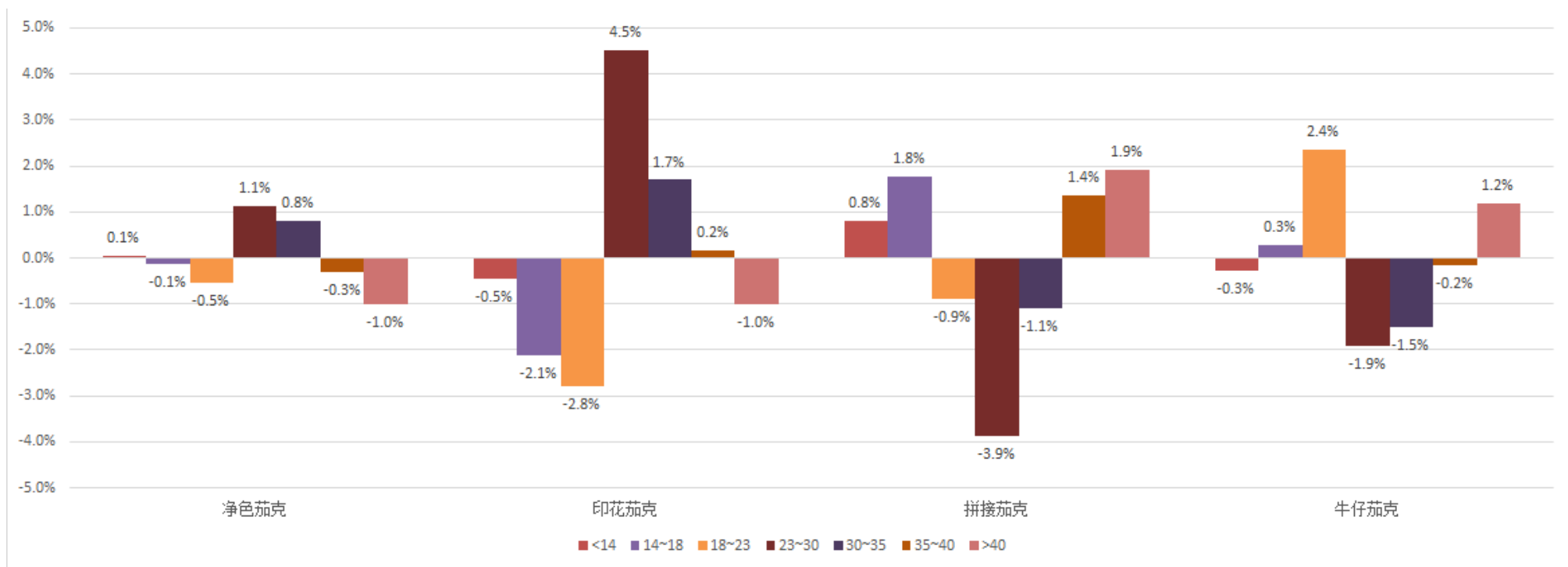
## 会员串联：

品类潜在消费者的标签，和零售预测，  
辅助开发端口



# 消费者年龄分布

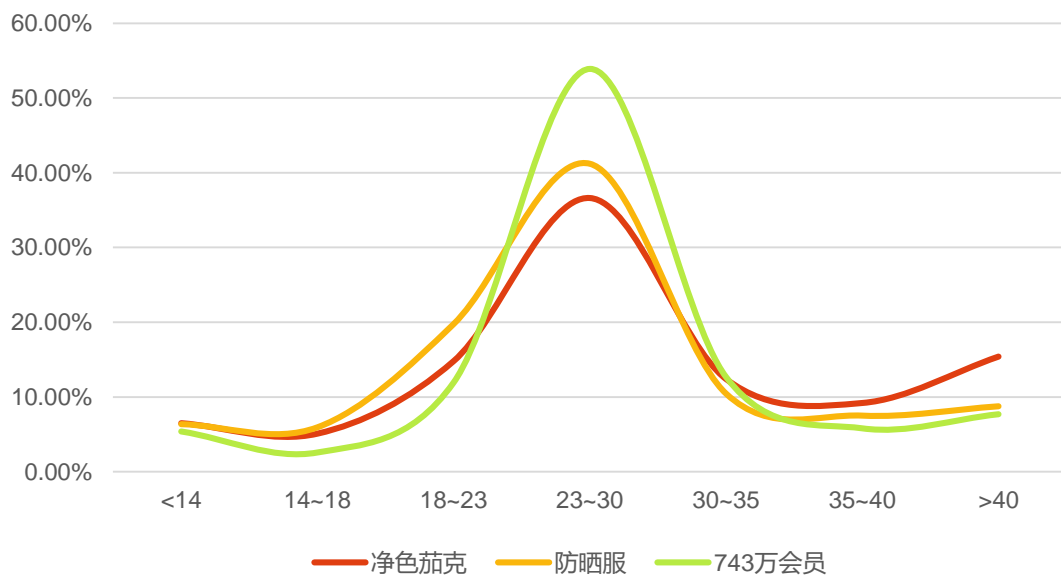
净色夹克相对整体和夹克的消费者年龄分布相似，印花夹克偏重于23~40岁，拼接夹克偏重于两边



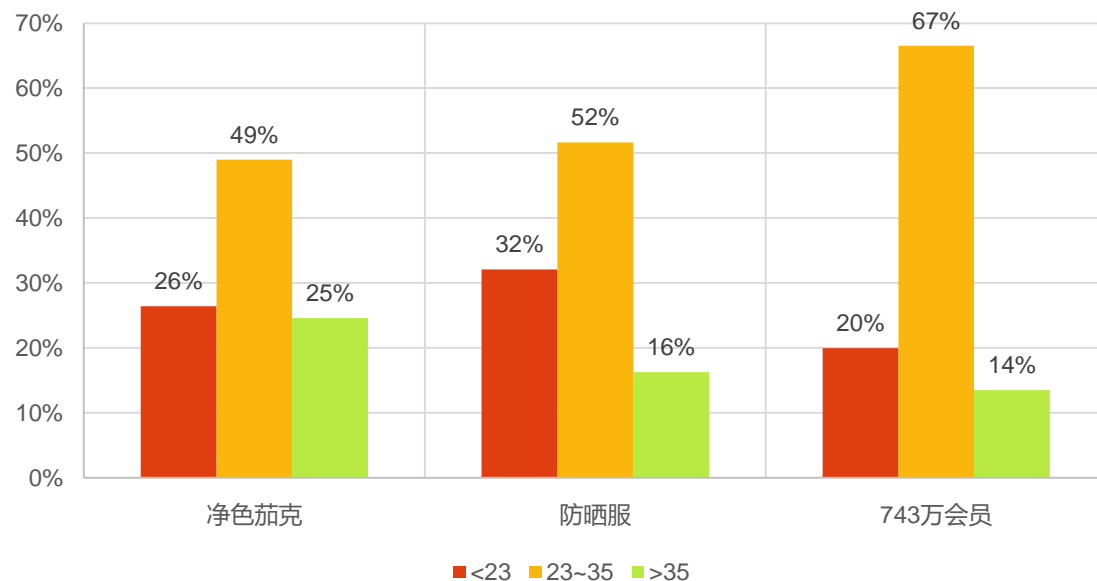
# 净色夹克与防晒服的会员年龄分布

净色夹克相对整体会员的年龄分布少了大量的23~30年龄段用户，有较大潜力  
防晒服相对整体会员和净色夹克都更加年轻，偏向于23岁以下用户

消费者年龄分布



年龄段分布





# Thanks

[www.semirbigdata.cn](http://www.semirbigdata.cn)

**SEM** Bigdata  
森马大数据

×



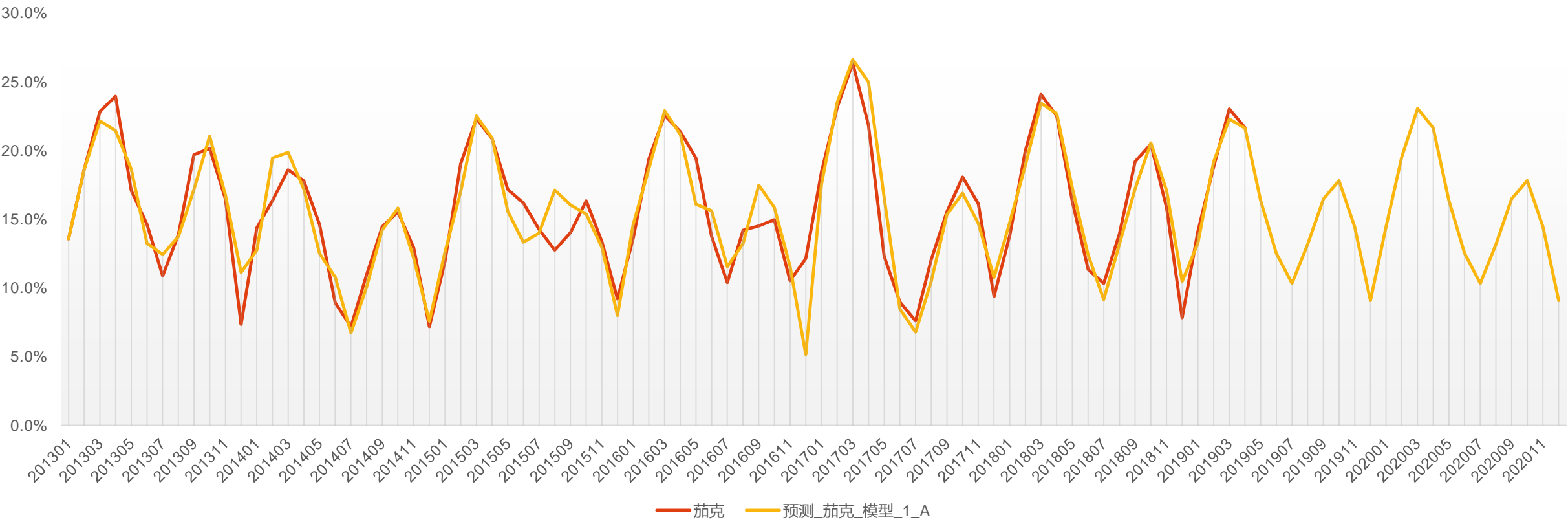
大数据项目组  
1996-2019 SEMIR  
BIG DATA TEAM

×

**FUN** Elysian  
Fields

# 零售预测曲线

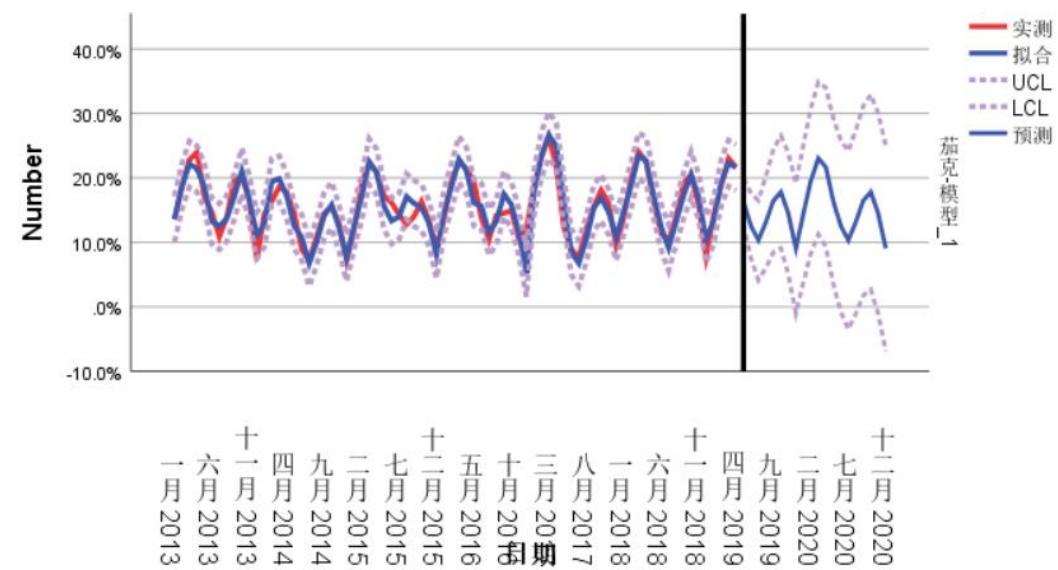
夹克预测模型



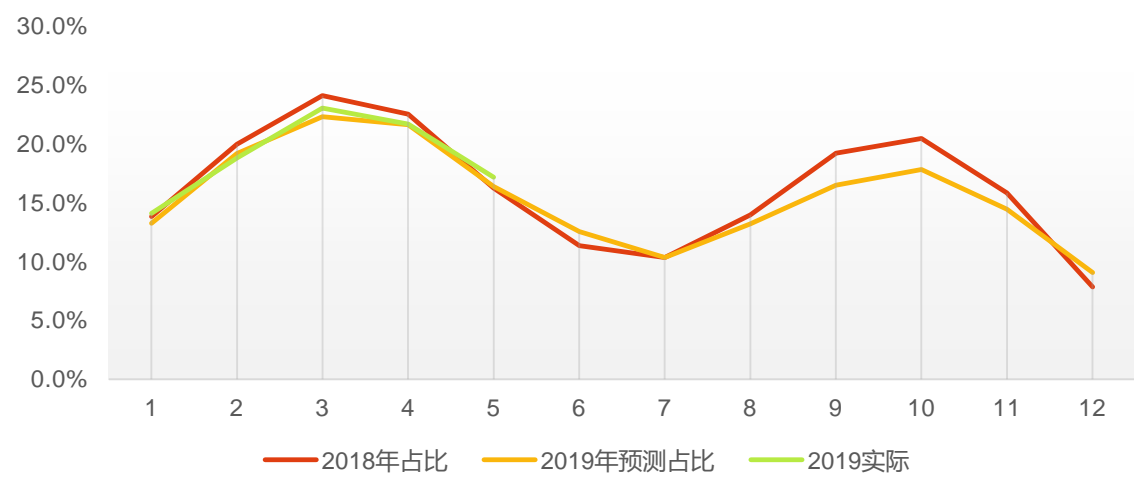
历史数据测算误差小于1.4%。

利用2017年数据预测2018年  
与直接利用2017年数据  
对比实际2018数据，误差下降了51%

# 置信区间

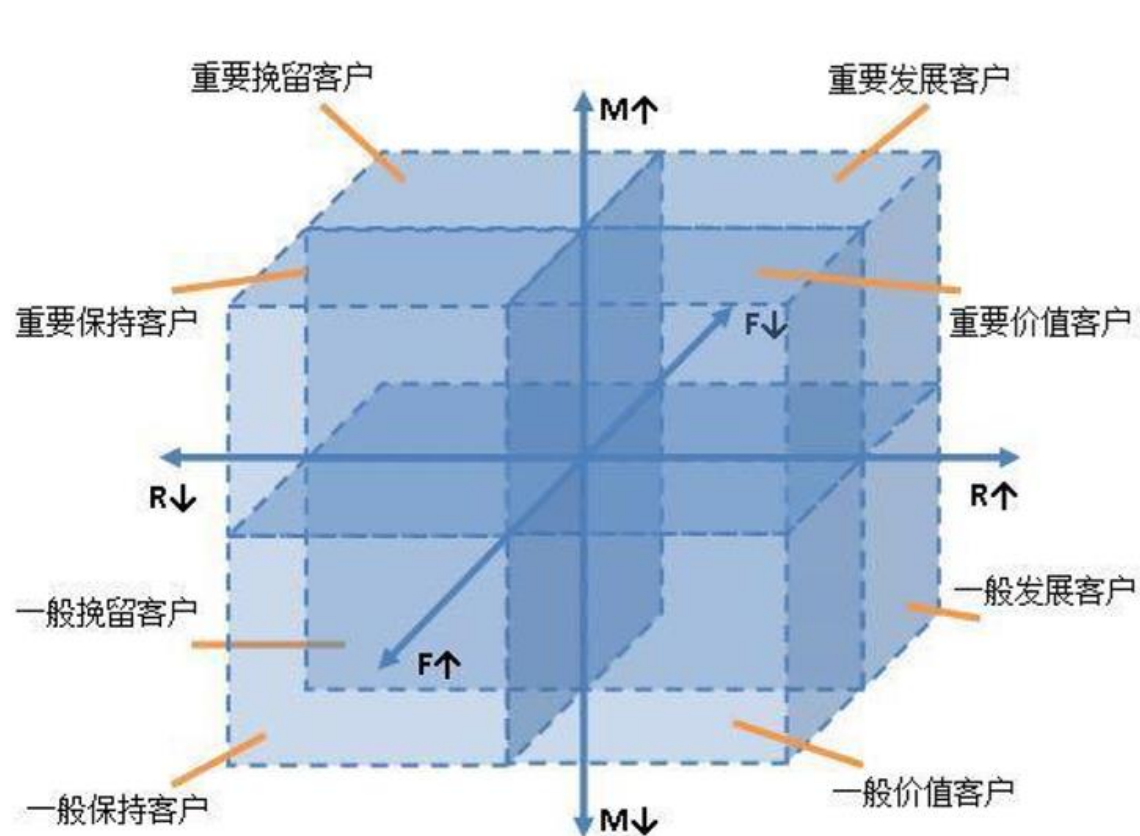


夹克占比两年差异





# 数字化体系 – 丰富的数据模型识别更多潜在消费者



影响因素

应用场景

**R**  
最近一次消费  
(Recency)

1. 店铺记忆强度
2. 接触机会多少
3. 回购周期

决定接触策略  
决定接触频次  
决定刺激力度

**F**  
消费频率  
(Frequency)

1. 品牌忠诚度
2. 店铺熟悉度
3. 客户会员等级
4. 购买习惯养成

决定资源投入  
决定营销优先级  
决定活动方案

**M**  
消费金额  
(Monetary)

1. 消费能力
2. 产品认可度

决定推荐商品  
决定折扣门槛  
决定活动方案

建模目标	找出核心客群	找出驱动核心顾客消费因子	从非核心顾客中找出潜在顾客	找出潜力顾客转化为核心客群的驱动因子
数据分析和建模方法	RFM 二维聚类 & 8/2 分布切分	配适最佳分类算法 如：XGboost、SVM、逻辑回归、贝叶斯、神经网络	KNN近邻算法 或 基于逻辑回归的Lookalike	配适最佳分类算法 如：XGboost、SVM、逻辑回归、贝叶斯、神经网络
模型评估方法	聚类结果的 差分性和 覆盖性	利用 N-Fold，从数据包中抽取 15~20%构建验证集合	利用已知历史数据中的核心客 群构建验证集合	利用 N-Fold，从数据包中抽取15~20%构 建验证集合

# 预测，品类模型，会员数据