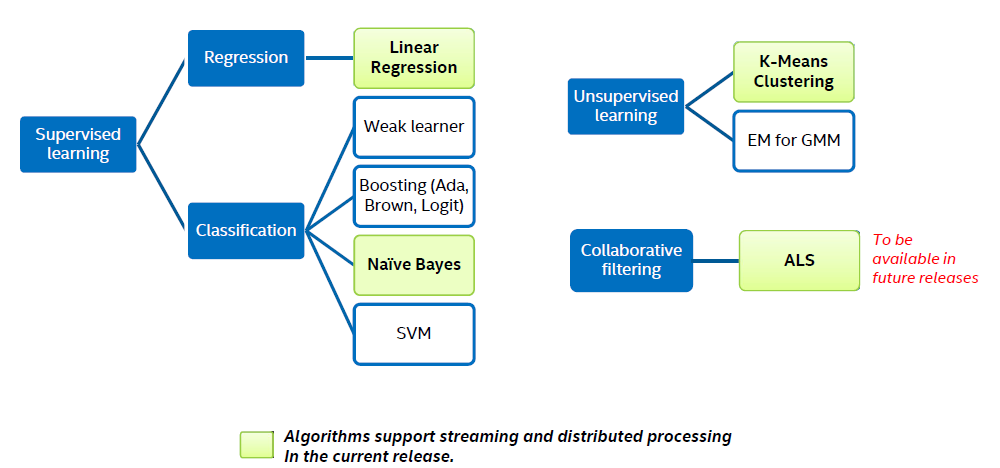
# 背景知识:总体人工智能



# 客户最短成交路径

场景是：为了实现客户快速成交！需要知道客户在短时间内的成交路径是怎么走的，找出最优的选择。比如：自注册开始，1个小时之内，有成交的客户。是什么吸引他们快速成交的？线上和线下有没有什么区别，什么样的优惠更有诱惑力。

# 客户流失预测



本图来自于Cachecache售前资料

# 营销自动化

会员运营平台在对潜客/粉丝、客户及会员（统称客户）进行运营时会收集数据，包含基本信息、交易数据，以及在客户交互场景下的行为数据（例如：优惠券事件、权益事件、客户交互等等）；这些数据会被数据中台采集并处理，数据中台会进行计算，输出各类客户标签（事实标签、统计标签、模型标签及预测标签）；客户标签直接与营销平台整合，进行触发式营销或者基于标签的客群主动营销，实现真正意义的营销自动化；

# 客服

数据中台输出客户画像，在一对一服务的场景下，客服/导购能更直接的了解当前服务客户的360度视图，为客户提供更加个性化/精准的服务及营销动作。

# 效果评估

# 动态线性规划

资源的最优配比/库存管理、车联网、最优运输工具选择、活动方案优化

资源优化调度

工业合作适配评估

配送优化

干线配送是指货车在多个城市组成的物流网络上进行配送，它是物流成本中占比最高的环节。因为各地区需求复杂且不均衡，很多物流公司或运输部门都会面临空车行驶率太高的问题，需要进行整体优化，提高车辆利用效率，减少配送成本。

多个城市组成的物流网络中，客户有一批货物需要配送，每个货物指定了装货地点，卸货地点，装卸货的时间窗。假设你是承运方，需要使用提供的一些车辆（不一定都使用，不使用则不计入成本）把所有货物送往目的地，使得配送的总成本尽可能少。

为了简化起见，假定各辆车的行驶速度是固定值，而且不允许中途卸货（只能到达目的地再卸货）。

例：有A，B，C，D四个地点（城市），各地点之间距离如图1(a)（距离单位是公里）。有3辆车可用，每辆车的规格都一样，最大容积=10立方米，最大载重量=1000公斤，车辆行驶速度20公里/小时。前2辆车初始位置是A，运送货物后需要返回A，第3辆车初始位置是B，全部运送完成后需要返回C。

有下列货物需要配送：

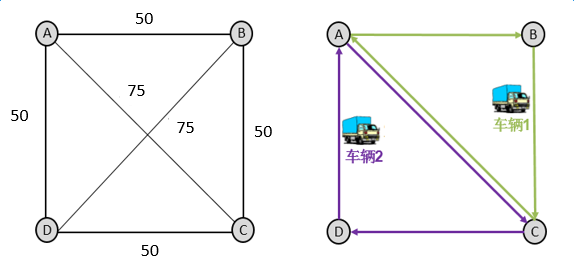
货物1：从A送到B，体积8立方米，重量500公斤，指定装货时间窗[9:00,10:00]（即最早9:00开始装货，最晚10:00开始装货，后面类似），卸货时间窗[10:00,12:00]；

货物2：从C送到A，体积10立方米，重量500公斤，指定装货时间窗[11:00,12:00]，卸货时间窗[12:00-15:00]；

货物3：从C送到D，体积6立方米，重量100公斤，指定装货时间窗[10:00,12:00]，卸货时间窗[11:00,15:00]；

货物4：从A送到C，体积7立方米，重量100公斤，指定装货时间窗[8:00,9:00]，卸货时间窗[10:00,12:00]之间.

每个货物的装(卸)时间为20分钟。



(a)地点和距离 (b)配送方案

*图1：同城配送示例*

图1(b) 提供了一种配送方案：

车辆1在A点装货物1，送到B点卸货，再到C点装货物2，送回A点卸货；

车辆2从A点装货物4，送到C点卸货，再装货物3，送到D点卸货，再返回A点；

当然，还有其他的各种配送方案，需要选择最优方案输出。

4**．输入数据格式：**

地点数据：location表

|  |  |
| --- | --- |
| 字段名 | 含义 |
| id | 地点id |
| name | 地点名称 |

道路数据：road表

|  |  |
| --- | --- |
| 字段名 | 含义 |
| start | 起点（对应location表的id） |
| end | 终点（对应location表的id） |
| distance | 距离（公里） |

货物数据：goods表

|  |  |
| --- | --- |
| 字段名 | 含义 |
| id | 货物id |
| start | 装货点（对应location表的id） |
| end | 卸货点（对应location表的id） |
| volume | 体积（立方米） |
| weight | 重量（公斤） |
| t1 | 最早装货时间（用字符表示，例如0930表示9:30，下同） |
| t2 | 最晚装货时间 |
| t3 | 最早卸货时间 |
| t4 | 最晚卸货时间 |

车辆数据：vehicle表

|  |  |
| --- | --- |
| 字段名 | 含义 |
| id | 车辆id |
| start | 初始点 |
| end | 返回点 |
| volume | 容积（立方米） |
| weight | 载重量（公斤） |
| speed | 速度（公里/小时） |
| cost\_day | 每天固定费用（元/天） |
| cost\_dist | 每公里费用（元/公里） |

装卸货过程所用时间：

在同一地点无论是装货还是卸货，或者是卸货之后再装货，该过程所用时间都设为*u* = 20分钟。

**5．输出数据格式：**

result表

|  |  |
| --- | --- |
| 字段名 | 含义 |
| goods\_id | 货物id（对应goods表的id） |
| vehicle\_id | 车辆id（对应vehicle表的id） |
| time\_load | 开始装货时间 |
| time\_unload | 开始卸货时间 |

另外，根据result表可以得到每辆车的行驶顺序，从而计算出总的成本（优化目标）,将此结果写在输出表最后一行。

总成本 = ∑每辆车的成本 = ∑ （车辆每天固定成本\*天数 + 车辆每公里成本\*公里数）。

天数按自然日（0:00-24:00），只要该天有配送任务，就算一整天。

**6. 约束条件：**

* 完全配送：所有货物都要完成配送（输入数据的车辆足够完成配送）。
* 车辆起止点：车辆从初始点出发，完成配送后必需回到返回点。
* 容积约束：每辆车在任何时刻装载的货物体积之和不得大于车辆容积。
* 重量约束：每辆车在任何时刻装载的货物重量之和不得大于车辆载重量。

时间窗约束：每个货物的装（卸）货开始时间必须在装（卸）货时间窗内。

店铺发货 🡪 立体仓库的位置 🡪 补货、配货

配货对每个区：配货的宽度、深度

## 道路交通优化

交通问题：设某种物资有m个产地和n个销地。产地Ai的产量为ai(i=1,2,…,m)；销地Bj的销量bj(j=1,2,…,n)。从第i个产地向第j个销地运输每单位的屋子的运价为Cij。这就是由多个产地供应多个销地的单品种物资运输问题。问如何调运这些物资才能使得总运费达到最小？

调度问题： 对一系列装货点（上车点）和（或）卸货点（下车点），组织适当的行车路线，使车辆有序地通过它在满足一定的约束条件如（货物需求量、发送量、交货时间、车辆容量限制、行驶里程限制、时间限制）下，达到一定目标（如路程最短、费用最少、时间尽量少、使用车辆数尽量少等）

优化问题：它是指一定数量的客户，各自有不同数量的货物需求，配送中心向客户提供货物，由一个车队负责分送货物，组织适当的行车路线，目标是使得客户的需求得到满足，并能在一定的约束下，达到诸如路程最短、成本最小、耗费时间最少等目的。

## 立体仓库优化

## 供应链优化

# 模拟

促销模拟 （可以归因到销售预测模式中）

## 预算模拟

重大活动的模拟；年度预算模拟；会有很多的业务方面的不同路径、不同事项的选择。。

# 最值问题

# 数据可视化

智慧楼宇

自动驾驶

高铁运行实时诊断

CDN实时故障诊断与自修复

拥有一块既能展现业务全貌又支持实时监控的数据看板，成为了数据化运营的起点。阿里云DataV数据可视化是专精于地理信息展现与业务数据融合的大数据可视化工具，旨在帮助用户通过图形化界面轻松搭建可视化应用，满足会议展览、业务监控、风险预警、地理信息分析等多种业务展示需求。

商业活动实时可视化

# 精细化管理

# 边界值

对于销售预测的满量、少量、精准量

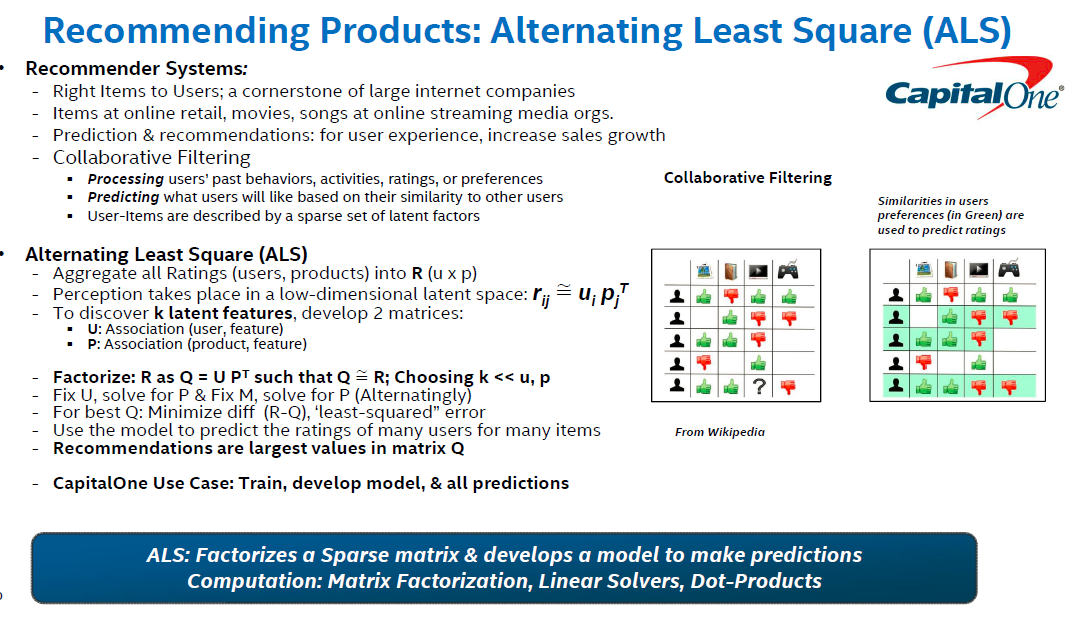
惩罚因子、奖赏因子

# 事后评估

转化率评测

ROI分析

# 推荐



ALS 是用户-商品-评分的混合推荐算法

因为某些用户对某些商品并没有进行评分。是缺失的

ALS就是为了去填充这些缺失的评分项，从而为该用户推荐 这些估计的评分里面相对较高的商品

保险公司的神经网络推荐：



基于消费者的消费等级及上一笔消费产品，导购可根据系统推荐推送不同满减金额的优惠券

## 协同过滤

例如：

员工可在操作台看到以下信息，

* 根据顾客的购买记录，查看到”猜你喜欢”的商品及商品搭配建议
* 根据消费者不同的消费行为习惯，不同的导购话术建议

从而完成：

产品偏好及搭配建议

销售一对一的个性化服务

# 聚类

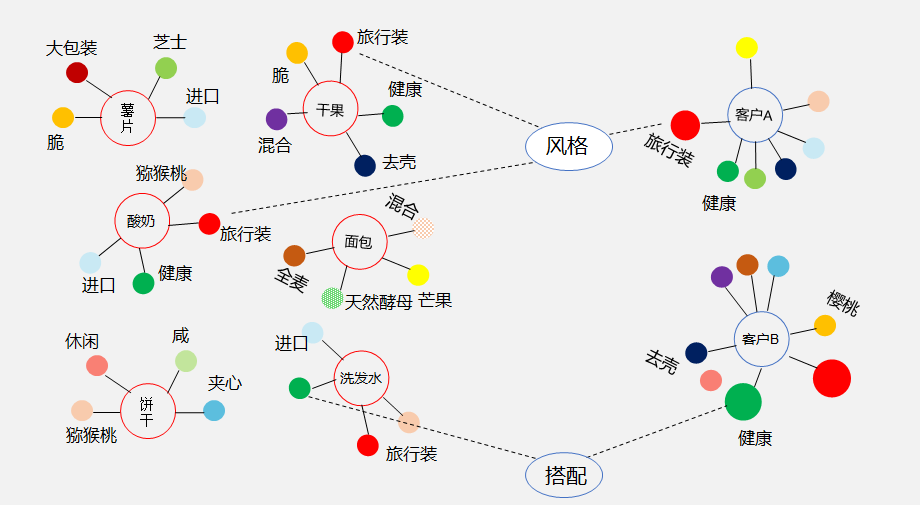
客群分析：静态客群、动态客群、客群快照；预测客群、自定义客群

营销洞察

对商店/便利店进行聚类

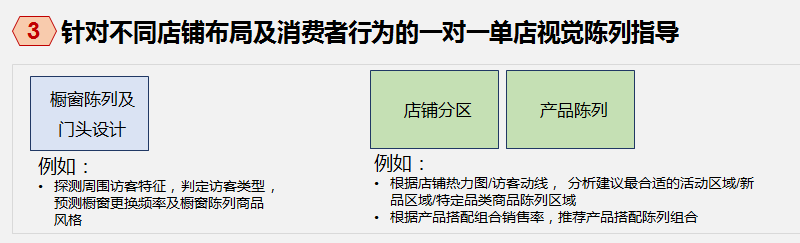
# 协同过滤

人货匹配引擎、商品推荐(猜你喜欢、购物车推荐、感兴趣客群、感兴趣人群特征)、用户推荐(感兴趣商品种类、购物的价格区间、)



# 智能选址

# 店面布局热力图



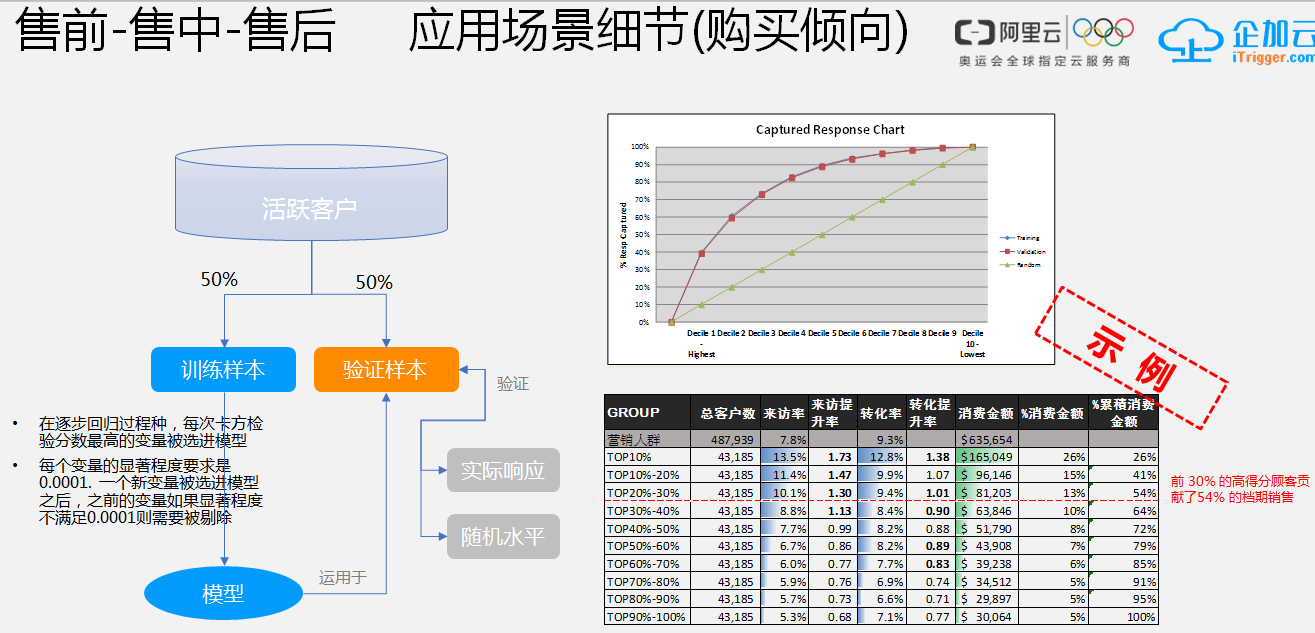
# 监控告警 与 阈值建设

价格促销

# 趋势分析 形式分析

万事达 用户 商户 倾向性 模型

服装行业的服装商品趋势



本图来自于Cachecache

# 对比分析

竞争对手分析

# 相对性分析

# 客户信息收集

人员到店面热力图—高德地图案例

# 区间一族

价格区间

# 历史记录

# 点击流

购买历史、收藏历史、浏览历史、加入购物车、打分评价

# 预测

在预测中的归因分析，比如天气温度对商品的销售预测

背景

XX医院近期门诊量激增，医院院长希望了解门诊量增长的动因，对医院运营产生的影响，以及下月门诊量可能的走势，以做好预案

客户需求

* 分析医院近期门诊情况，提交门诊分析月报
* 进行下月门诊量预测

问题描述

1. 根据提供的数据，提交一份门诊业务分析报告给院长
2. 根据医院提供的历史三年的门诊量数据，预测下月门诊量，并提出进一步完善门诊量预测方法的方案建议

提供尽可能完善的全景分析，帮助医院管理者了解门诊业务的状态及导致异常的原因，定位门诊业务中存在的关键问题

库存预测

销售预测(品类、库存)

**根据店铺不同的特点为线下每家店铺预测SKU层级的销售**

* 产品铺货
* 店铺促销活动
* 供应链的优化

例如：

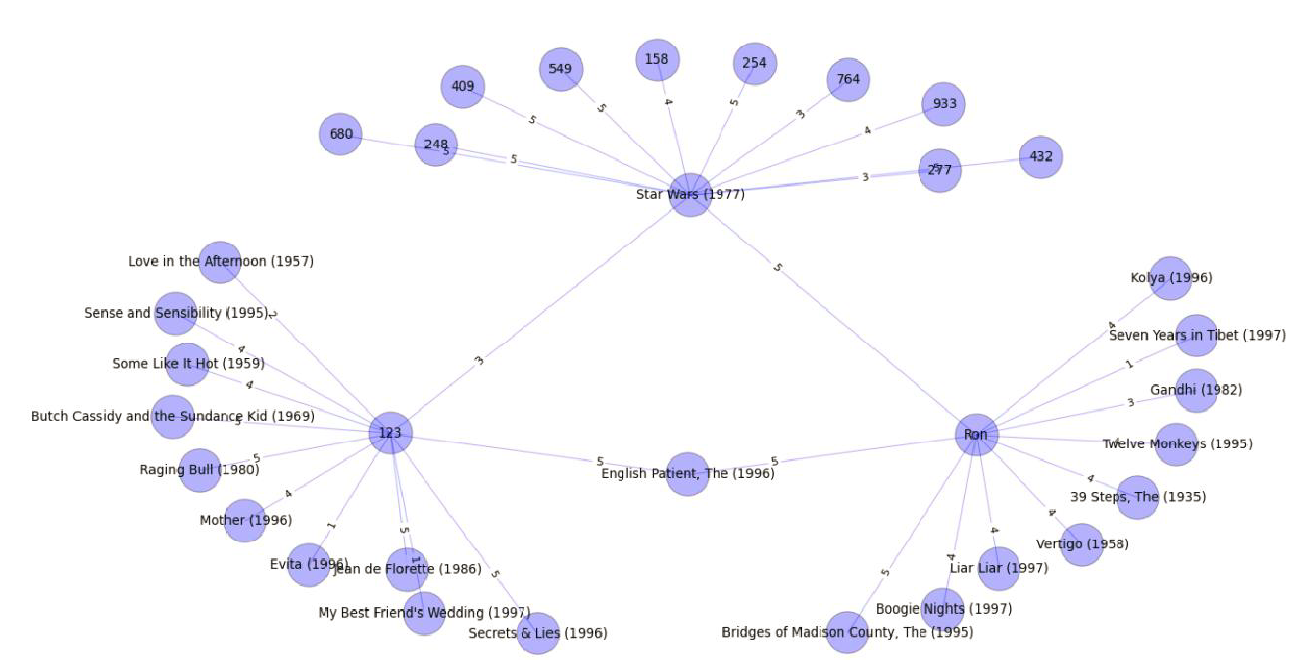
* 根据单店历史产品销售记录/周边环境/消费者特征，预测产品潜在销售及所需库存并进行铺货
* 根据单店的销售情况，建议最适合的促销活动来促进销售或清理库存

弹性供应链

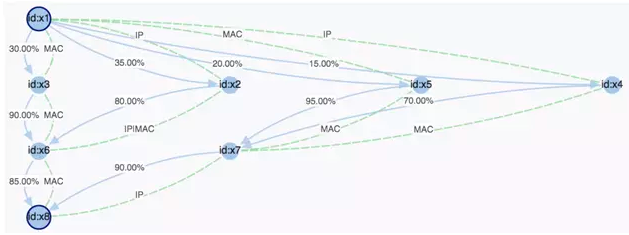
# 熵值评定

# 关联分析

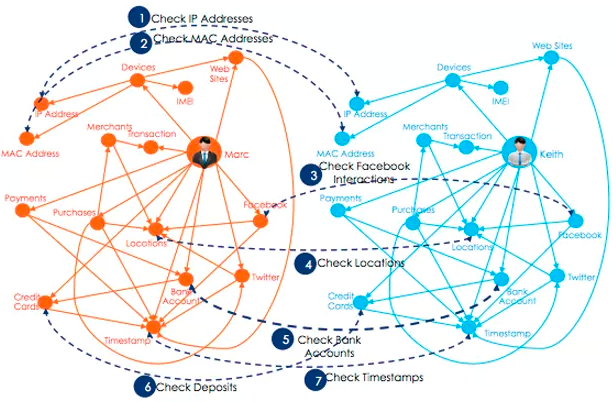
电影院需要知道 电影被观看的次数 和 该电影的平均评级 之间的关系



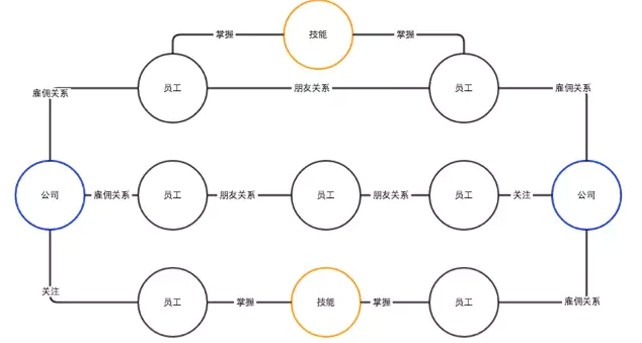
## 反洗钱



## 金融案例反欺诈



## 社交网络图谱



## 企业关系图谱



# RFM及其它相关算法

# 不同数据源的打通 OneID/Uni\_ID

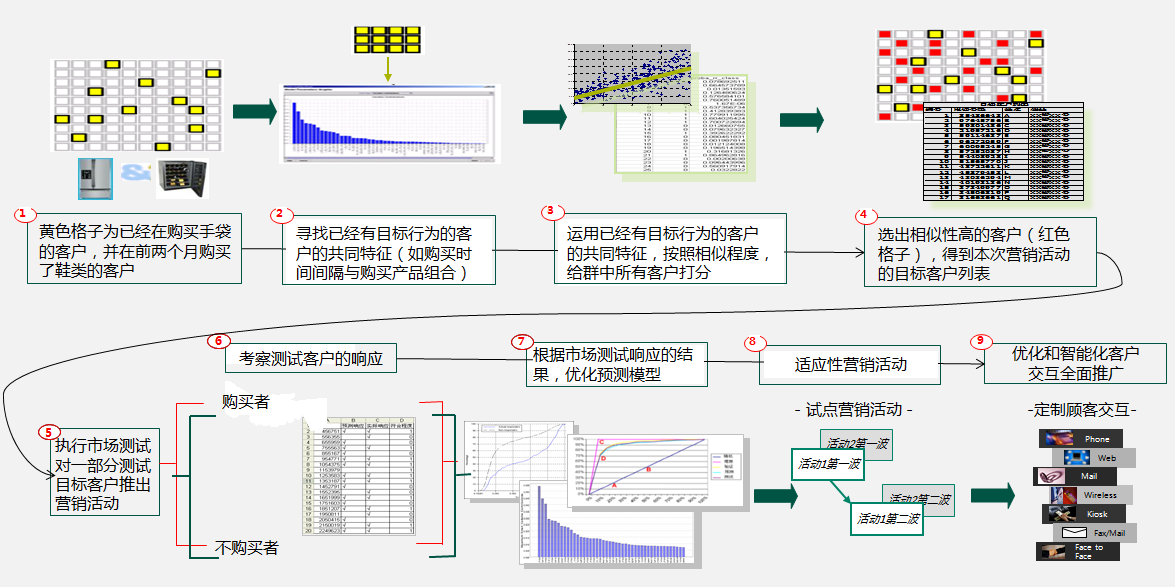
# NLP

由于电子病历数据量比较大，很多医院电子病历都是将电子病历已txt文件的方式存储，为了实现结构化需要将原电子病历文本进行数据解析。

医院原始电子病历文本，具有数据量大，格式不统一，需要开发一套便于电子病历解析引擎。

针对医院原始电子病历文本，按照业务需求既定的解析规则，解析成相应的二维数据格式。

# 千人千面



# 其他

## 数据倾斜

### 场景描述

用户访问页面日志表 table\_a, 表字段如表1所示，日数据量1000亿，共有1000万个 item\_id， 其中有 1% 的 item\_id 单个日记录数超过 500万。

表1：用户访问日志表 table\_a

|  |  |  |
| --- | --- | --- |
| 字段 | 类型 | 字段说明 |
| ds | string | 日期 |
| user\_id | string | 用户id |
| shop\_id | string | 店铺id |
| page\_id | string | 页面id |
| item\_id | string | 商品id |
| visit\_time | string | 访问时间yyyy-MM-dd HH:mm:ss |

商品类目信息表 table\_b，表字段如下表2所示，日数据量 10 亿。

表2：商品类目信息表 table\_b

|  |  |  |
| --- | --- | --- |
| 字段 | 类型 | 字段说明 |
| ds | string | 日期 |
| item\_id | string | 商品id |
| cate\_id | string | 商品类目id |
| cate\_name | string | 类目名称 |
| item\_deail | string | 商品详情 |

在类目分析时，分析师希望知道每个类目下的商品每天的访问PV和UV，结果表table\_c如表3所示。

表3：结果表table\_c

|  |  |  |
| --- | --- | --- |
| 字段 | 类型 | 字段说明 |
| ds | string | 日期 |
| cate\_name | string | 商品类目名称 |
| pv | bigint | 当天访问次数 |
| uv | bigint | 当天访问用户数 |

### 问题描述

在实际关联的时候，会产生长尾现象[也称为数据倾斜]，即有一些节点运行时间非常长，请分析产生的原因，并给出解决方法。

## 拥堵时段计算

### 场景描述

信号灯路口每2分钟计算出一个拥堵指数值，当拥堵指数大于2的时候，认为路口发生异常，表结构如表4所示，其中数据样例如表5所示,其中 2016-12-18 00:00:00 表示时间段 [2016-12-18 00:00:00 ~ 2016-12-18 00:02:00)

表4 路口拥堵指数表 table\_d

|  |  |  |
| --- | --- | --- |
| 字段 | 类型 | 字段说明 |
| inter\_id | string | 路口id |
| time | string | 2分钟时间段yyyy-MM-dd HH:mm:ss |
| cong\_index | double | 拥堵指数 |
| ds | string | 发生日期yyyyMMdd |

表5 路口拥堵指数示例数据

|  |  |  |  |
| --- | --- | --- | --- |
| inter\_id | time | cong\_index | ds |
| 1 | 2016-12-18 00:00:00 | 0.4 | 20161218 |
| 1 | 2016-12-18 00:02:00 | 2.6 | 20161218 |
| 1 | 2016-12-18 00:04:00 | 2.1 | 20161218 |
| 1 | 2016-12-18 00:06:00 | 1.4 | 20161218 |
| 1 | 2016-12-18 00:08:00 | 0.7 | 20161218 |
| 1 | 2016-12-18 00:10:00 | 0.8 | 20161218 |
| 1 | 2016-12-18 00:12:00 | 2.9 | 20161218 |
| 1 | 2016-12-18 00:14:00 | 1.3 | 20161218 |
| 1 | 2016-12-18 00:16:00 | 2.2 | 20161218 |
| 1 | 2016-12-18 00:18:00 | 1.6 | 20161218 |
| 1 | 2016-12-18 00:20:00 | 0.3 | 20161218 |
| 1 | 2016-12-18 00:22:00 | 0.7 | 20161218 |
| 1 | 2016-12-18 00:24:00 | 0.9 | 20161218 |
| 1 | 2016-12-18 00:26:00 | 1.1 | 20161218 |
| 1 | 2016-12-18 00:28:00 | 2.3 | 20161218 |
| 1 | 2016-12-18 00:30:00 | 1.7 | 20161218 |

一天之内，上一个异常时段的结束时间到下一个异常时段的开始时间间隔小于10分钟则将两个异常时段合并。

### 问题描述

结果表输出格式如下表6所示。

表6 异常时段合并结果表 table\_e

|  |  |  |
| --- | --- | --- |
| 字段 | 类型 | 字段说明 |
| inter\_id | string | 路口id |
| time\_period | string | 合并后的时间段[yyyy-MM-dd HH:mm:ss,yyyy-MM-dd HH:mm:ss] |
| ds | string | 发生日期yyyyMMdd |

其中上表表5中的样例数据产生的结果如表6所示。

|  |  |  |
| --- | --- | --- |
| inter\_id | time\_period | ds |
| 1 | [2016-12-18 00:02:00,2016-12-18 00:18:00) | 20161218 |
| 1 | [2016-12-18 00:28:00,2016-12-18 00:28:00) | 20161218 |

## 3 数据接入

### 场景描述

某出租车运营商会实时将出租车的 gps 数据追加写入到服务器指定目录的文件中，一天一个目录，同一个服务器下面一天按照小时只写一份文件，数据字段和顺序如表7所示，各个字段之间使用’\u001’ 分隔。

表 7 出租车gps 日志文件字段说明

|  |  |  |
| --- | --- | --- |
| 序号 | 字段 | 说明 |
| 1 | log\_time | 日志记录时间yyy-MM-dd HH:mm:ss |
| 2 | vhc\_id | 车辆id |
| 3 | vhc\_color | 颜色 |
| 4 | vhc\_no | 车牌号 |
| 5 | lng | 经度 |
| 6 | lat | 纬度 |
| 7 | dir | 形式方向 |
| 8 | ds | 发生日期yyyyMMdd |

为了提供更好的计算，通常需要将数据上云，以阿里云的 MaxCompute 和 DataHub 为例，以一方面需要将数据实时写入Datahub，供下游节点实时计算，另一方面需要将数据按天归档到 MaxCompute, 用作离线计算。其中DataHub 和 MaxCompute 的Schema结构至少包含如表8和表9所示的字段。

表8 出租车gps 数据 datahub topic Schema

|  |  |  |
| --- | --- | --- |
| 字段 | 类型 | 字段描述 |
| log\_time | string | 日志记录时间yyy-MM-dd HH:mm:ss |
| vhc\_id | string | 车辆id |
| vhc\_color | string | 颜色 |
| vhc\_no | string | 车牌号 |
| lng | double | 经度 |
| lat | double | 纬度 |
| dir | string | 行驶方向 |
| ds | string | 日期yyyMMdd |

表9 出租车GPS数据MaxCompute 表 Schema

|  |  |  |
| --- | --- | --- |
| 字段 | 类型 | 字段描述 |
| log\_time | string | 日志记录时间yyy-MM-dd HH:mm:ss |
| vhc\_id | string | 车辆id |
| vhc\_color | string | 颜色 |
| vhc\_no | string | 车牌号 |
| lng | double | 经度 |
| lat | double | 纬度 |
| dir | string | 行驶方 |
| ds | string | 日期yyyMMdd |

### 要求

写入Dtahub和MaxCompute的数据不能丢失。

写入 Datahub的数据需要均匀分布在Datahub的shard中，且写入MaxCompute的时候不能有太多小文件。

写入datahub 的数据需要产出接入数据日志报表[datahub的数据统计可以使用阿里云流计算]，包括每5分钟往datahub写数据记录数，datahub接收到的数据数和平均延迟

写入 MaxCompute的同样需要接入数据日志报表，包括小时往MaxCompute写写入数据量和写入MaxCompute中的数据量。

需要考虑多台服务器存在的情况。

请给出相关解决方案和设计实现代码。

## 4 数据缺失值处理

### 场景描述

如异常值合并中的情况，由于一些客观条件，会造成某个时间片拥堵指数缺失的情况，在展现的时候需要对数据进行填充，保证数据曲线的完整性，通常需要对这些异常值进行插值填充；另一方面由于数据采集的异常值存在，会导致计算出来的拥堵指数也存在异常值[特别大或者特别小],这些数据也需要剔除，当做缺失值进行插值平滑。

### 要求

使用箱形图进行异常值过滤，使用拉格朗日插值或者牛顿插值对缺失值进行插值处理

结合城市路口超过10000个的情况，给出相关实现。

## 生产效率评估

## 人货场规则管理器

杨伦彪发的从店铺、商品、店员、会员四个角度规划的规则管理器，如下：

**店铺**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **主题** | **主题含义** | **涉及指标** | **计算规则** | **计算周期** | **备注** |
| 销售好但库存风险高的店铺 | 靠堆库存来完成销售的店铺 | 店铺类型、城市、商圈、近3周新款销售额排名、库销比变化值 | 1. 近3周销售排名在前30% 2.库销比变化值的斜率变化 | 周 | 该主题旨在找到库存潜在风险过大的店铺，通过及时促销或者调拨调整 |
| 销售好但店员不均衡的店铺 | 店员不均衡的店铺 | 店铺类型、城市、商圈、近3周销售额排名、店员销售额方差 | 1.近3周销售排名在前30% 2.近3周店员销售额方差变化 | 周 | 该主题定位到店员能力偏差过大的店铺，存在店员冲突风险和人员流失风险 |
| 竞争性缺乏的店铺 | 店员水平差距小且都较低 | 店铺类型、城市、商圈、近3周店员销售排名、店员销售额方差 | 1.近3周店员销售排名在后30% 2.近3周店员销售方差变化 | 周 | 该主题旨在找到店员能力普遍缺乏的店铺，存在整体经营风险 |
| 整体销售好但会员销售差的店铺 | 会员能力低的店铺 | 店铺类型、城市、商圈、销售额、会员销售排名 | 1.销售额排名在30% 2.会员销售额排名与整体销售的差异排名 | 周 | 该主题定义会员销售能力较差的优秀店铺，从而判断店铺风险的大小 |
| 会员销售好但有效会员数差的店铺 | 会员销售占比高，但有效会员数下降的店铺 | 店铺类型、城市、商圈、近3周会员销售额排名、有效会员数变化值 | 1.近3周会员销售排名在30% 2.有效会员数变化值得斜率变化 | 周 | 该主题找到潜在有效会员出现流失的店铺 |
| 业绩好但离职倾向高的店铺 | 业绩好但店铺离职倾向高的店铺 | 店铺类型、城市、商圈、销售额、店员指标达成率方差变化值、平台活跃度、平台活跃度变化 | 1.近3周销售额排名在前30% 2.店员平台活跃度在后30% 3.店员指标达成率方差变化的斜率 4.店员平台活跃度方差变化的斜率 | 周 | 该主题旨在找出有店员离职风险的店铺 |
| 销售好但库存不足的店铺 | 库存不足的店铺 | 店铺类型、城市、商圈、近3周新款销售额排名、库销比变化值 | 1. 近3周销售排名在前30% 2.库销比变化值的斜率变化 | 周 | 该主题旨在找到库存不足的店铺，通过及时补货或上新调整 |
| 销售好但折扣低的店铺 | 低折扣支撑销售的店铺 | 店铺类型、城市、商圈、销售额、销售折扣 | 1.销售额排名在前30% 2.销售折扣排名 | 周 | 该主题旨在找到销售折扣支撑搞销售额的店铺，适合清库 |
| 销售好但客单数低的店铺 | 客流有潜在风险的高销售店铺 | 店铺类型、城市、商圈、销售额、客单数 | 1.销售额排名在前30% 2.客单数排名 | 周 | 该主题旨在找到成单数下滑的店铺，店内流量出现问题 |
| 销售好但连带率低的店铺 | 店员连带能力有问题的店铺 | 店铺类型、城市、商圈、销售额、连带率 | 1.销售额排名在前30% 2.连带率排名 | 周 | 该主题旨在找到连带率下滑的店铺，多件成交需要提高 |
| 销售好但结构不合理的店铺 | 结构不合理的店铺 | 店铺类型、城市、商圈、新款销售额、系列占比偏差值、类别占比偏差值、款式占比偏差值 | 1.新款销售排名在前30% 2.系列占比偏差值 3.类别占比偏差值 4.款式占比偏差值 | 月 | 该主题旨在找到销售结构不合理的店铺，定位需要调整 |
| 会员流失风险的店铺 | 存在会员流失风险的店铺 | 店铺类型、城市、商圈、有效会员数量变化值 | 1.有效会员数量变化值 | 周 | 该主题旨在找到会员流失风险的店铺 |
| 持续下滑风险店铺 | 业绩下滑的店铺 | 销售额环比 | 1.销售环比排名 | 周 | 该主题旨在找到业绩持续下滑的店铺 |
| 进步空间最大的店铺 | 有巨大进步空间的店铺 | 销售额环比（分拆更多指标） | 1.销售环比排名 | 周 | 该主题旨在找到业绩进步巨大的店铺 |
| 销售好但成长率不足的店铺 | 销售出现瓶颈的店铺 | 销售额、销售额环比 | 1.销售额排名前30% 2.销售额环比排名 | 周 | 该主题定位现有模式成长空间有限的店铺 |

**商品**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **主题** | **主题含义** | **涉及指标** | **计算规则** | **计算周期** | **备注** |
| 销售好但库存风险高的商品 | 库存风险高的畅销商品 | 近3周新款销售额排名、库销比变化值 | 1. 近3周销售排名在前30% 2.库销比变化值的斜率变化 | 周 | 该主题旨在找到库存潜在风险过大的商品，及时促销 |
| 整体销售好但销售不合理的商品 | 结构不合理的商品 | 新款销售额、尺码占比偏差值、颜色占比偏差值 | 1.新款销售排名在前30% 2.系列占比偏差值 3.类别占比偏差值 4.款式占比偏差值 | 周 | 该主题旨在找到销售结构不合理的商品 |
| 连带性最强的商品 | 连带量最多的商品 | 搭配率 | 1.搭配率排名 | 周 | 该主题旨在找到搭配性最强的商品 |
| 补货预警的商品 | 需要补货的商品 | 近3周新款销售额排名、库销比变化值 | 1. 近3周销售排名在前30% 2.库销比变化值的斜率变化 | 周 | 该主题旨在找到库存不足的商品，通过及时补货或上新调整 |
| 销售好但折扣低的商品 | 低折扣支撑销售的商品 | 销售额、销售折扣 | 1.销售额排名在前30% 2.销售折扣排名 | 周 | 该主题旨在找到销售折扣支撑搞销售额的商品，适合清库 |
| 销售好但平台投诉率高的商品 | 低折扣支撑销售的商品 | 销售额、平台投诉排名 | 1.销售额排名在前30% 2.平台投诉排名 | 周 | 该主题旨在找到平台投诉最多的商品，商品质量可能有问题 |
| 调拨预警的商品 | 需要调拨的商品 | 近3周新款各店销量的方差、库销比的方差 | 1. 近3周新款各店销量的方差在30% 2.库销比方差的排名 | 周 | 该主题旨在找到亟需调拨的商品 |

**店员**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **主题** | **主题含义** | **涉及指标** | **计算规则** | **计算周期** | **备注** |
| 业绩好但有离职倾向的店员 | 业绩好但存在离职倾向的店员 | 店铺类型、城市、商圈、销售额、店员指标达成率方差变化值、平台活跃度、平台活跃度变化 | 1.近3周销售额排名在前30% 2.店员平台活跃度在后30% 3.店员指标达成率方差变化的斜率 4.店员平台活跃度方差变化的斜率 | 周 | 该主题旨在找出有离职风险的优秀店员 |
| 业绩好但折扣低的店员 | 业绩好但依靠低折扣的店员 | 店铺类型、城市、商圈、销售额、销售折扣 | 1.销售额排名在前30% 2.销售折扣排名 | 周 | 该主题旨在找到销售折扣支撑搞销售额的店员，适合清库 |
| 业绩好但客单数低的店员 | 客流有潜在风险的高销售店员 | 店铺类型、城市、商圈、销售额、客单数 | 1.销售额排名在前30% 2.客单数排名 | 周 | 该主题旨在找到成单数下滑的店员，店内流量出现问题 |
| 业绩好但连带率低的店员 | 连带能力有问题的店员 | 店铺类型、城市、商圈、销售额、连带率 | 1.销售额排名在前30% 2.连带率排名 | 周 | 该主题旨在找到连带率下滑的店员，多件成交需要提高 |
| 业绩好但会员销售差的店员 | 会员能力低的店员 | 店铺类型、城市、商圈、销售额、会员销售排名 | 1.销售额排名在30% 2.会员销售额排名与整体销售的差异排名 | 周 | 该主题定义会员销售能力较差的优秀店员，从而判断店员风险的大小 |
| 会员数量多但有效会员低的店员 | 会员数量高，但有效会员数下降的店员 | 店铺类型、城市、商圈、近3周会员数量排名、有效会员数变化值 | 1.近3周会员数量排名在30% 2.有效会员数变化值得斜率变化 | 周 | 该主题找到潜在有效会员出现流失的店员 |
| 能力不均衡的店员 | 销售好但各项指标有显著出入的店员 | 店铺类型、城市、商圈、销售额，折扣、客单数、连带率、价格排名方差 | 1.销售额排名在30% 2.折扣、客单数、连带率、价格排名方差排名 | 周 | 该主题找到能力不均衡的优秀销售人员 |
| 退步最大的店员 | 业绩下滑的店员 | 店铺类型、城市、商圈、销售额环比 | 1.销售环比排名 | 周 | 该主题旨在找到业绩持续下滑的店员 |
| 进步空间最大的店员 | 有巨大进步空间的店员 | 销售额环比（分拆更多指标） | 1.销售环比排名 | 周 | 该主题旨在找到业绩进步巨大的店员 |
| 业绩好但成长率低的店员 | 销售出现瓶颈的店员 | 销售额、销售额环比 | 1.销售额排名前30% 2.销售额环比排名 | 周 | 该主题定位现有模式成长空间有限的店员 |

**会员**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **主题** | **主题含义** | **涉及指标** | **计算规则** | **计算周期** | **备注** |
| 消费高但频次低的会员 | 消费能力强但消费频次不高的会员 | 近3月消费金额排名、近3月消费频次 | 1.近3月消费金额排名在前30% 2.近3月消费频次排名 | 周 | 本主题旨在找到需要加强沟通和联系的顾客 |
| 消费低但频次高的会员 | 消费能力低但频次很高的会员 | 近3月消费金额排名、近3月消费频次 | 1.近3月消费金额排名在后30% 2.近3月消费频次排名 | 周 | 本主题旨在找到粉丝顾客，做推广和促销 |
| 准流失会员 | 有效分数（RFM模型）在60分以上，但上次消费时间在60天以前 | 有效分数、消费时间 | 1.有效分数在60分以上 2.上次消费时间在60天以前 | 周 | 本主题定位准流失的优质会员，注重激活 |
| 消费高但连带率低的会员 | 消费金额高但连带低的会员 | 近3月消费金额、连带率 | 1.近3月消费金额排名在前30% 2.连带率排名 | 月 | 本主题定位高消费群体中连带低的会员，注重连带销售 |
| 消费高但服务导购不稳定的会员 | 消费金额高，服务导购不固定的会员 | 近3月消费金额、不同服务导购人数 | 1.近3月消费金额排名在前30% 2.服务导购人数排名 | 月 | 本主题定位高消费群体中专属导购不稳定的会员，加强维护 |
| 消费高但消费门店不稳定的会员 | 消费金额高，消费门店不固定的会员 | 近3月消费金额、消费门店数 | 1.近3月消费金额排名在前30% 2.消费门店数排名 | 月 | 本主题定位高消费群体中专消费门店不稳定的会员，加强维护 |
| 新品喜好的会员 | 消费金额高且上新及消费的人群 | 近3月消费金额、消费时间与上新时间的差值的平均值、折扣 | 1.近3月消费金额排名在前30% 2.折扣在90%以上 2.消费时间与上新时间差值的平均值排名 | 月 | 本主题定位高消费群体中新品偏好的会员 |
| 折扣喜好的会员 | 消费金额高但折扣低的会员 | 近3月消费金额、折扣 | 1.近3月消费金额排名在前30% 2.折扣率排名 | 月 | 本主题定位高消费群体中折扣低的会员，折扣和活动喜好 |
| 积分使用率高的会员 | 积分使用占总金额比例的会员 | 积分使用额/（消费总金额+积分使用额） | 积分使用额/（消费总金额+积分使用额）排名 | 月 | 本注定定位偏好积分使用的会员，推动积分使用 |

**1、扩展客户互动渠道。**很多会员，可能只接触了我们1个渠道，多的可能2-3个。那会员对我们品牌的感知会比较片面和单一。那从收益上客户的消费潜力也无法最大化挖掘。因此我们可以通过数据中台，筛选出互动渠道少的顾客群，然后针对性推荐新的渠道，分别策划其感兴趣的点来吸引最终在新的渠道也产生互动。比如有的会员只在巧虎之家有过互动，他并不清楚还有商城，或者视频又有什么奖励，平时哪能发优惠券或者举办什么活动。那我们可以分析他本身在巧虎之家互动的产品、价格范围、喜欢的优惠、每次互动时长、平均活跃周期，推荐视频乐园给客户，告知可在视频乐园观看视频学习更多早教知识，还能积攒积分，兑换优惠券和礼品。那么进一步还可以引导关注服务号，在积分商城活跃起来。客户越活跃，产生的消费越高。那给客户去触发的方式和推荐的产品和优惠券范围等，就可以通过数据中台分析，在业务中台来给客户做触发。最终在业务中台的效果分析来长期追踪效果。**2、挽回预流失淡漠客户。**即使品牌已经有很高的复购率，但是总有一批人介于还没流失但有流失危险的区域。这部分客户我们可以去根据互动周期的降低、积分活动参与数降低、短信活动反馈降低、登录次数减少、登录时长缩短等来判定。一旦判定出客户群，就可以根据分析出的他们的软肋，来策划挽留方案。如客户是某一次对产品体验不好，进而变得淡漠，还是近几次服务没有得到及时响应而淡漠，如果各方分析都不是，那需要确定是否竞品接入，通过问卷或电话来采集。确定原因后，可以策划一定的给优惠券/给积分/老带新福利/解决服务难点/解释产品体验不好的原因重新指导等方案，通过业务中台来触发后，设定数据跟踪周期为1天、3天、1周、2周、1月、2月等来持续跟踪客户挽回效果。**3、挖掘KOL群体，推动其带动一般会员。**如在数据中台筛选互动渠道大于等于4，且活跃周期小于等于15天，客单价在300元以上，总消费次数7次以上，总互动次数20次以上（这里可以根据点赞、观看视频、商城活动、微信回复、友介较多、口碑宣传比较多）的客户，打标签为KOL群体。然后给这批群体下发一些任务，比如参与一些活动发朋友圈，或者在购买页面上传好的评价，或者给他们的妈妈帮群体，分享内购优惠券，指引他们去特定渠道来参与，线下或线上都行。那么推动和效果追踪就可以在业务中台去跟踪。当然他们人员的执行是相当重要的。

**◆识别职业特性，寻找更贴切互动话题。**从交易地址关键词，根据算法逻辑组合，判断出客户最大可能的职业特征。同时根据消费能力组合出客户职业之上的最可能的级别，结合其他行为数据来策划不同职业特征客群的最合适互动话题。

场景实例：**中等购买力、中等互动程度的“白领”群体**

中等互动频次

中等购买力

1500≥客单价≥500

浏览至下单时长≤8min

或：客件数≥3且1500≥客单价≥350

地址关键词

最近下单时间0-180天

最近点击时间0-60天

最近观看视频0-50天

最近登录会员中心0-50天

60天上述互动次数2-5次

大厦，中心，广场，公司，有限，科技园，软件园，孵化器，信息，科技

关系：并且+或者

最近活跃：≤180天

关系：或者

次数：≥1

关系：前5或者，6并且

再来分析要推动的互动话题，依据具体的互动内容、购买商品偏好，来分析对应客群可能的心理生活特征，去策划：

**分析推测：**

该父母群体职业稳定，收入中等，重视孩子兴趣培养，对孩子的创意精神和舞台表现力有培养期待。但工作关系互动并非全职带孩那般频繁，基于客群特征互动却也不少。

近期看过舞台剧1次

TV观看“创意”类节目

观看次数大于等于3

观看时长≥50min

有过点赞和评论

互动内容

适应年龄：3-4岁

商品内容：好奇心套餐

购买商品

关系：并且

商品时间0-90天

关系：或者+并且

**互动内容策划方向**

根据上述结论，我们可考虑的互动内容去贴切客群方案如：

互动时间：周四、周五晚（父母群体将要周末，提前策划孩子兴趣内容）

互动内容：《家庭小品表演》、《儿童生活小知识》、《趣味天体》

互动时长：准备时间：40min，互动时间：50min

运营支持：业务中台策划全部流程，运营团队编辑内容，短信/微信推送客户群体。

效果跟踪：活动发起后，实时跟踪互动内容、互动频次有无提升，是否有下单购买新的相关产品，互动产品类型是否符合预定轨迹等。