

销量预测系统

Sales Forecasting System

WHY

为什么要做销量预测？

HOW

预测怎么做出来的，要考虑哪些因素？

为什么要做销量预测

需求预测驱动一致的运营计划；支持全品类千万级SKU的需求预测

支持采销业务

滚动的需求预测 (Rolling demand forecasting)

91天短期
12个月中长期

支持运营运作

促销预测

智能补货

库内补货

调拨计划

滞销预测

GMV预测

促销模拟

营销计划

单量预测

订单生产
计划

库存网络规
划

运营规划

ECLP预测

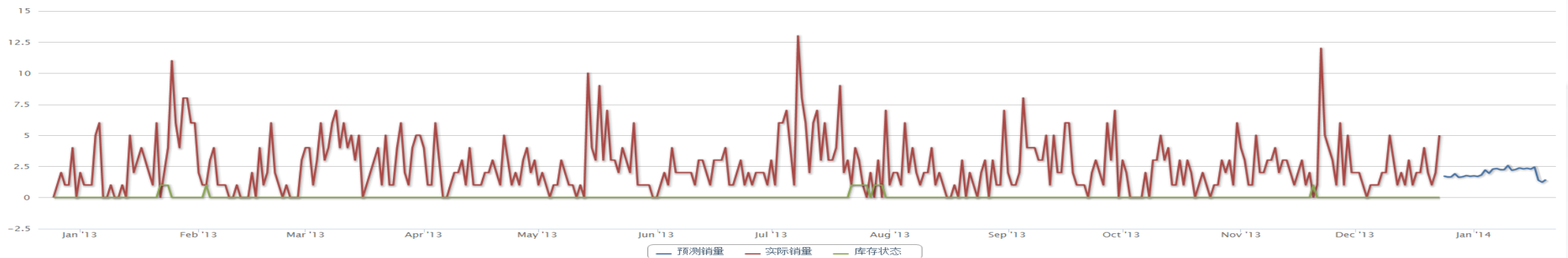
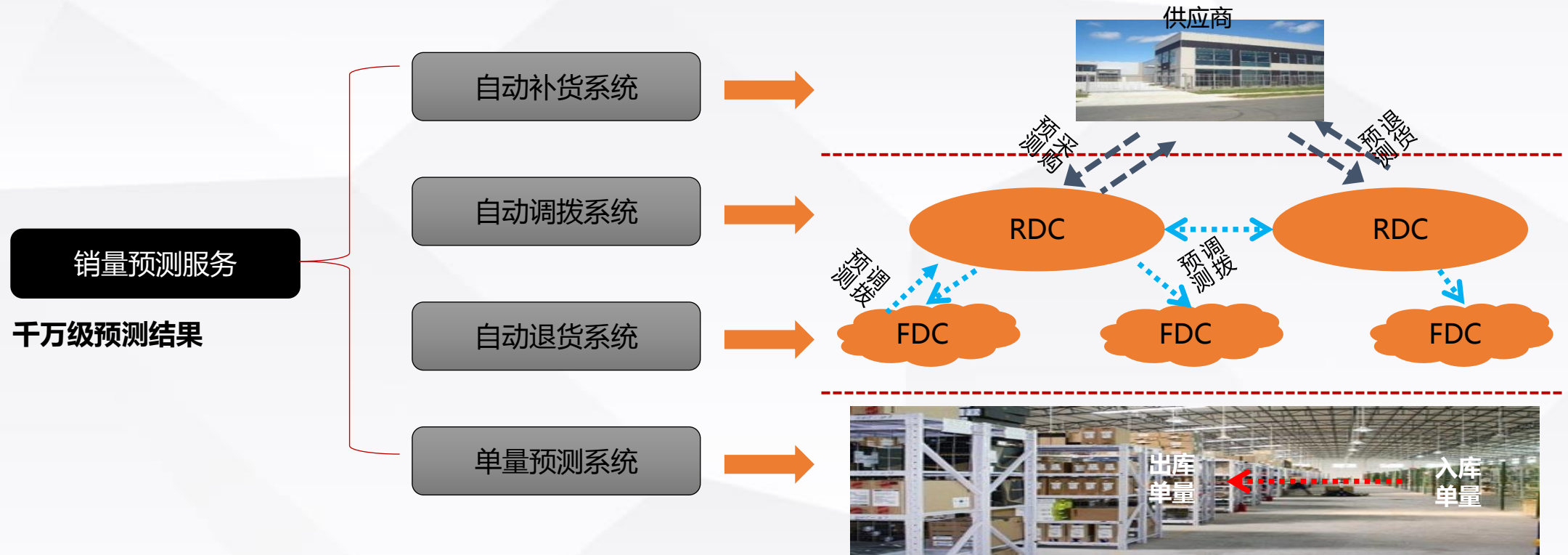
库内补货

劳动力计划

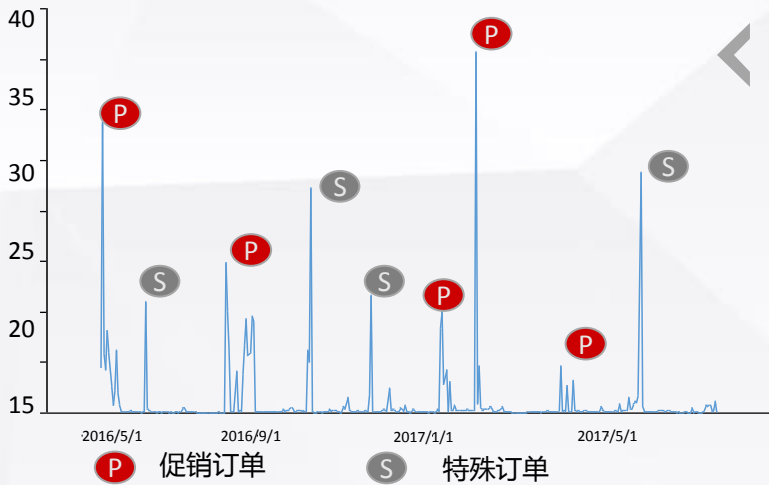
存货布局

销量预测覆盖范围

JD.COM 京东



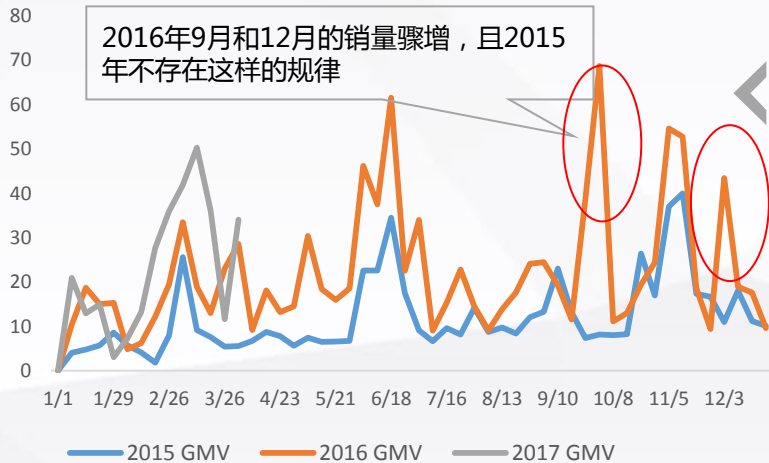
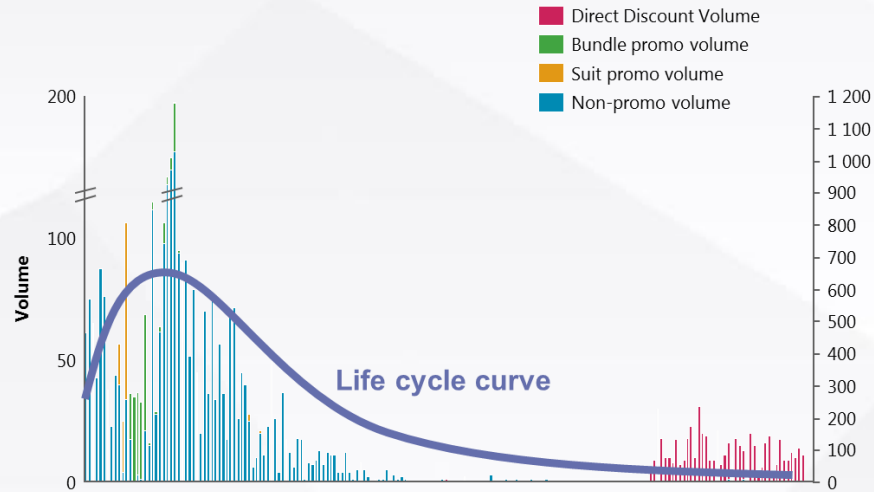
销量预测的难点



促销订单：互联网电商的销售高度依赖促销，京东促销订单占比约70%，促销带来的销量波动不稳定性大且很难寻找规律

特殊订单：由企业客户，B端商家等其它渠道带来的不规律订单

生命周期：基于生命周期的商品在持续没有销量的情况下，因为清仓而又带来再次销售的小波峰



不同的运营策略导致，品类的销售数据并不规则

9/12月的销量剧增，导致在对2016年9月，12月做预测时，无法根据历史趋势来准确预测销量

销量预测不稳定性：

促销复杂多样，同一商品多个促销叠加繁琐

广告位流量信息难解析

不同种商品交错着光环，蚕食效应无法全面考虑

运营策略导致商品销售数据

生命周期趋势受到清仓而又表现不稳定性

1. 数据处理

2. 模型预测

3. 常规预测

4. 促销预测

5. 预测发布

历史

中小件

销量

--收货地址城市对应的配送中心口径

统计

图书

--收货地址城市对应的图书配送中心口径

保税仓

--收货地址城市对应的保税仓覆盖口径

大家电

--出库地对应的配送中心口径

1. 数据处理

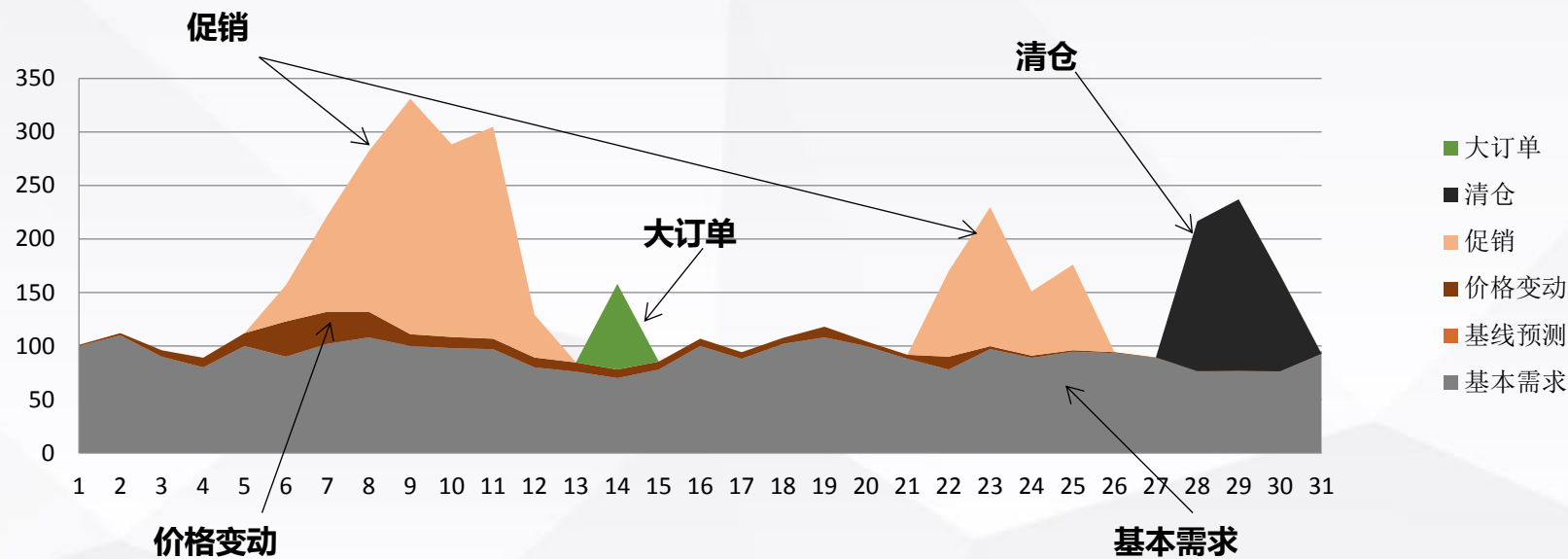
2. 模型预测

3. 促销预测

4. 常规预测

5. 预测发布

- 基础数据处理
- 价格清洗
- 节假日调制
- 促销处理
- 动态周期选择
- 季节调整
- 极值平滑
- 库存异常处理
- 预测单元选择
- 天气信息
- 大单过滤等等。。。。。。



1. 数据处理

2. 模型预测

3. 常规预测

4. 促销预测

5. 预测发布

- 线下模型训练
- 多模型KPI PK胜出
- 标准化输入和输出

预测建模分类	适用问题/场景	建模方法
简单预测建模	各种问题 (长尾品, 新品, 不动销, 销售样本点少...)	1. 周平均 2. 历史同期 3. 加权平均 4. 加限拦截 5. 新品模型 6. 不动销模型 7. 保守模型 8. 新品预测
影响因子预测建模	促销模型、价格模型、节假日、流量模型 (618, 双十一, 促销, 十一, 新年, 首图广告, 品类日...)	1. 一元线性回归模型 2. 多元线性回归模型 3. 非线性回归模型
历史销量预测建模	趋势建模、周期性建模、季节性建模 (四季, 开学季, 四六级考试前...)	1. 指数平滑模型/状态空间统计模型 2. Arima模型 3. Holt winters 4. 动态趋势模型
机器学习/综合预测建模	综合预测因子和历史销量建模 (长期有销量, 稳定性产品...)	1. 动态回归模型 2. 神经网络模型 3. GBDT 4. Deep Learning

1. 数据处理

2. 模型预测

3. 常规预测

4. 促销预测

5. 预测发布

通过对历史促销数据的学习

促销的
时段

开始时间及截至时间

促销的
力度

让利的幅度

促销的
形式

满减、每满减、买赠、免邮、组套优惠

促销的
范围

参与的品类、客户级别

促销商
品池

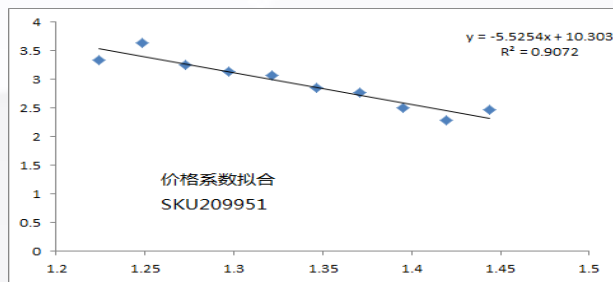
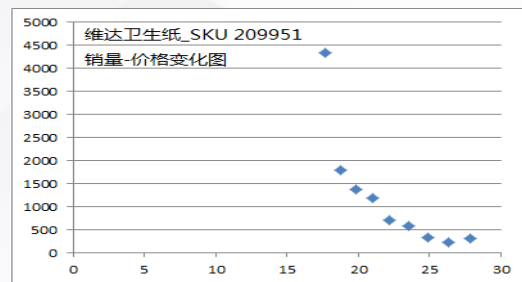
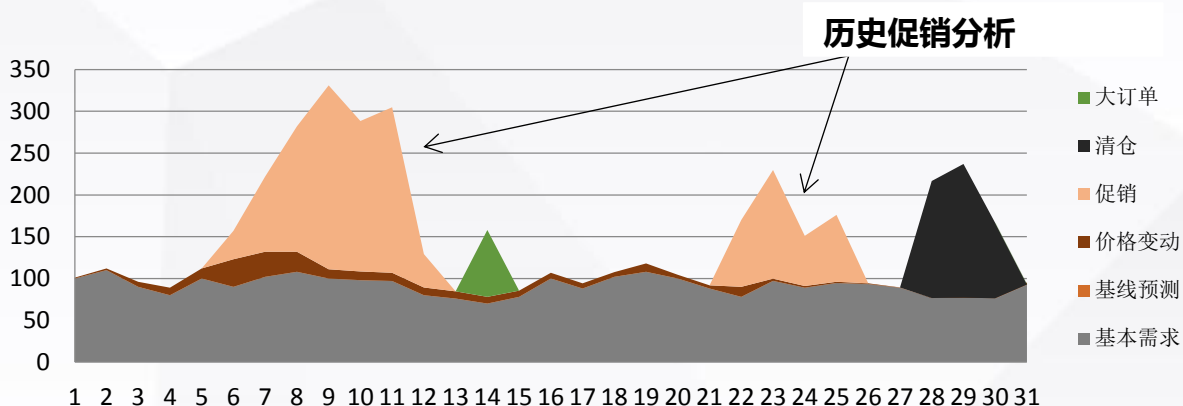
商品联合促销范围品类

营销资
源

页面位置及站外推广渠道

效果拆
分

多种促销叠加后效果拆分



预测的方法-促销预测（机器学习模型）

JD.COM 京东

1. 数据处理

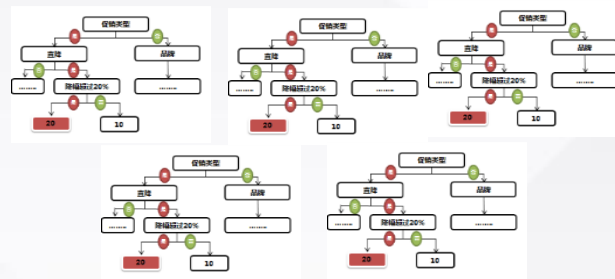
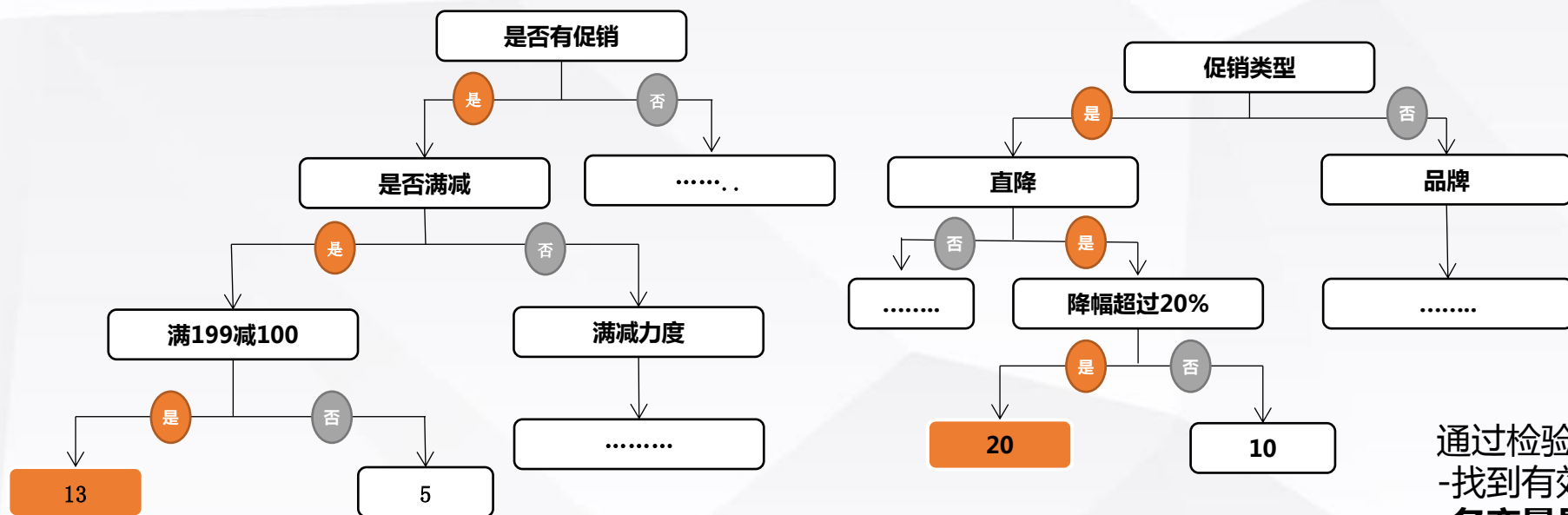
2. 模型预测

3. 常规预测

4. 促销预测

5. 预测发布

机器学习模型能够“学习”各促销特性之间复杂的关系，以及它们对于最终销售额的影响



通过检验数百种决策树，机器学习算法

-找到有效/无效的预测变量

-各变量是否重要

-根据大量促销特性做出预测

- 但根据若干组解释变量随机创造出数千种决策树
- 每一种决策树都是一个弱预测模型
- 然而，数千种模型的集合创造出功能超级强大的最终模型!

经“训练”，机器学习算法可以模拟多种促销活动，而且速度非常快

1. 数据处理

2. 模型预测

3. 常规预测

4. 促销预测

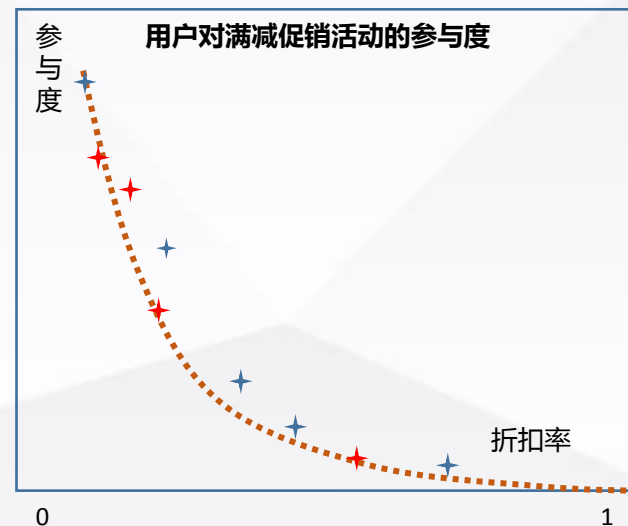
5. 预测发布

GBDT可以抽象为如下形式：

- GBDT的目标是串行学习出K棵树： $M = (T_1, T_2, \dots, T_K)$
- 假定第*i*棵树 T_i 对应的函数为 f_i ，则对于任意一个特征向量 X ，其预测值等于 $f(X) = \sum_{k=1}^K f_k(X)$
- Recall，我们的目标是最小化下式：

$$\begin{aligned} & \sum_{i=1}^n L(f(X_i), y_i) + \Omega(M) \\ &= \sum_{i=1}^n L\left(\sum_{k=1}^K f_k(X_i), y_i\right) + \sum_{k=1}^K \Omega(T_k) \\ &= \sum_{i=1}^n L(f_K(X_i) + \widehat{y_i^{K-1}}, y_i) + \sum_{k=1}^K \Omega(T_k) \end{aligned}$$

其中 $\widehat{y_i^{K-1}} = \sum_{k=1}^{K-1} f_k(X_i)$ 为前K-1棵树对 X_i 的预测值



1. 数据处理

2. 模型预测

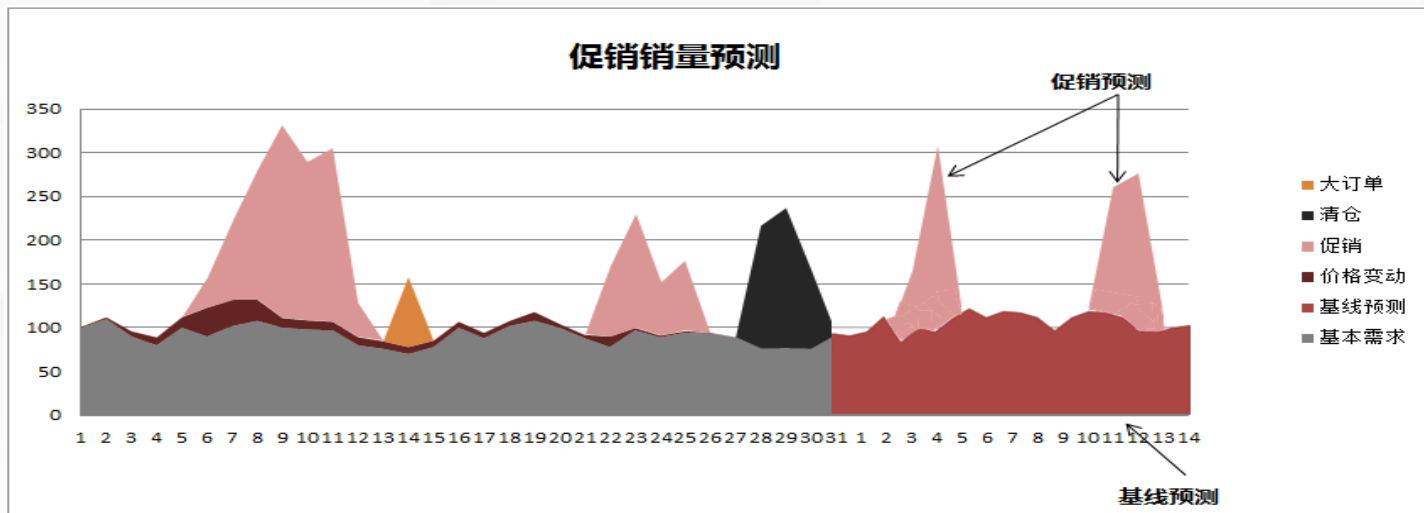
3. 常规预测

4. 促销预测

5. 预测发布

在分析历史促销的基础上，通过获取**未来**促销计划，力度，资源位等通过多种模型预测一个商品在该促销中(GMV,Sales)的提升力度，输出**叠加在基线预测**的促销预测

- **历史模型**-通过已知的历史促销活动，对相同条件下的相同促销的销售额进行合理的预测
- **回归模型**-用于对一些变量进行推断，例如促销额对于销量提升额，然后根据回归（弹性）模型进行的给出一个预测
- **机器学习模型**-机器学习模型用于对产品和促销特征之间的复杂关系进行理解和推断，“学习”各促销特性之间复杂的关系，以及它们对于最终销售额的影响



1. 数据处理

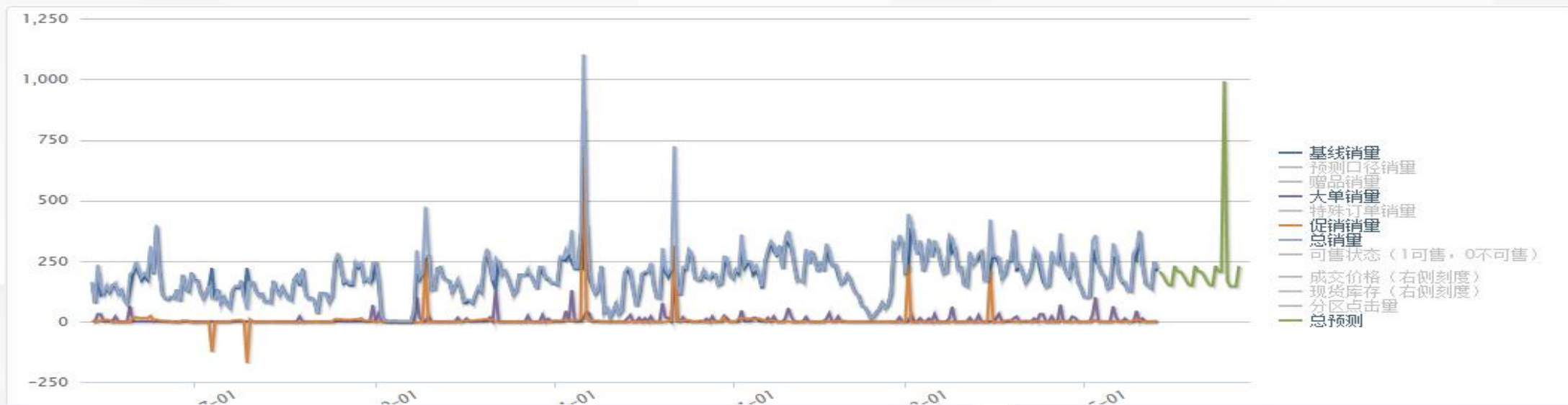
2. 模型预测

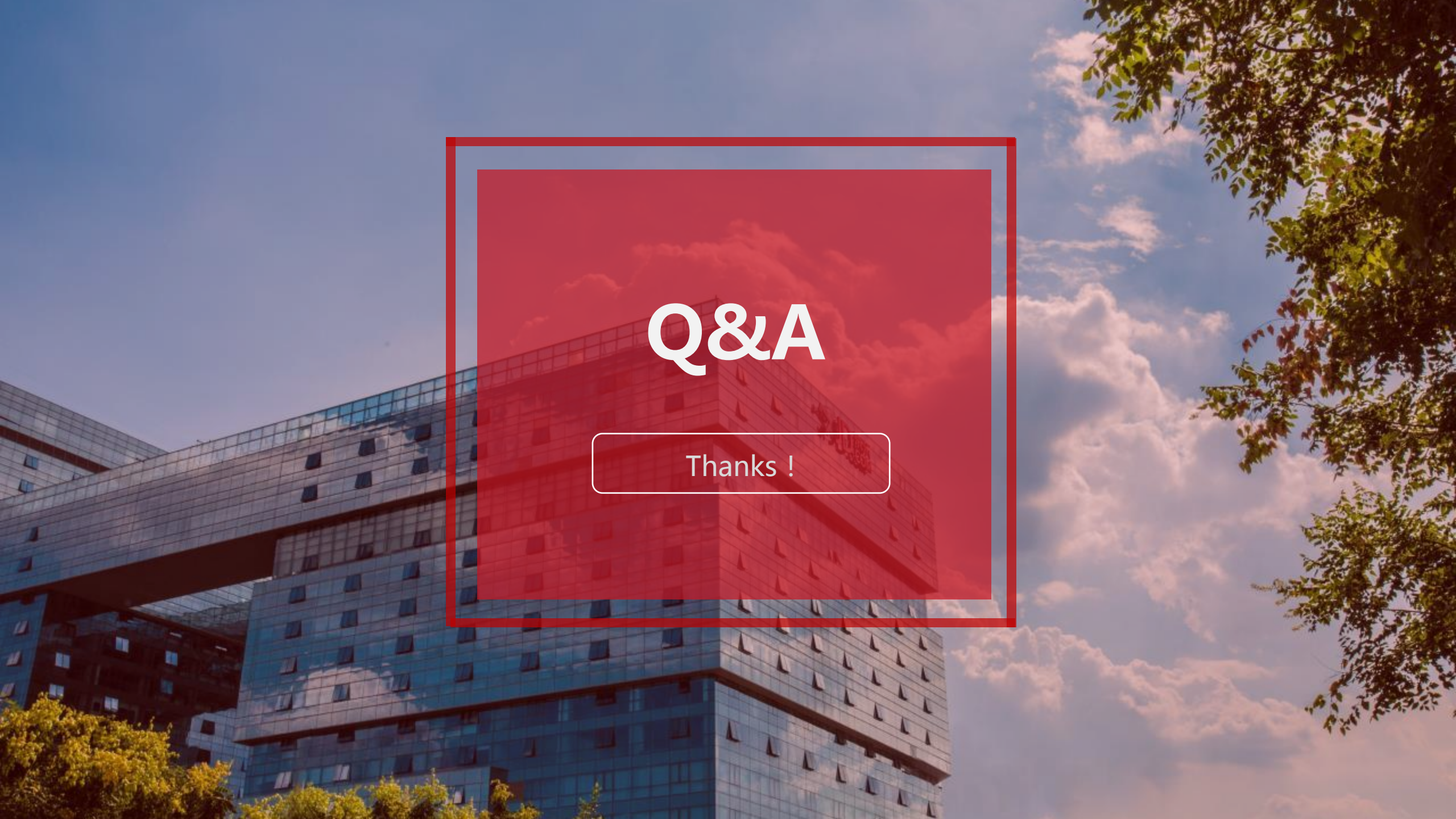
3. 预测调整

4. 促销预测

5. 预测发布

- 通过WEB共享
- 内外部协同产生一致需求预测
- 异常检测
- 预测均值和标准差，高斯分布





Q&A

Thanks !