

LLM在微盟BI领域的探索

微盟 / 王成

Outline

- 基于生成式的ABI
- 微盟BI之路
- 基于LLM的ABI实践与探索
- 总结与展望

基于生成式的ABI能给我们带来什么

- 更强的AI任务能力
- 更拟人化的交互结果
- 更通用的理解能力
- 更低的门槛

微盟BI之路



中国领先的智慧商业服务提供商

微盟致力于为商家提供数字化新商业操作系统
成为企业数字化转型的最佳伙伴

6900+

公司员工数

16万+

帮助企业数字化转型

10+

全国分子公司数

1600+

全国服务商1600家
覆盖全国30多个省市

50强

2020在线经济（上海）50强

微盟产品和服务布局



微盟数据建设



基于LLM的ABI实践与探索

ABI实践方向



ABI实践难点

规范化的数据宽表 + AI能力 = 优秀的ABI

- 数据表零散，数据字段较多
- 字段命名、格式不规范
- 不同场景AI能力有区别
- AI需要跟下游业务场景关联
- AI缺乏行业知识

ABI实践-NL2SQL

复杂场景的准确率

- 同比、环比等
- 多表查询

表信息

- 数据表多、字段多
- 字段格式不统一

专业名词

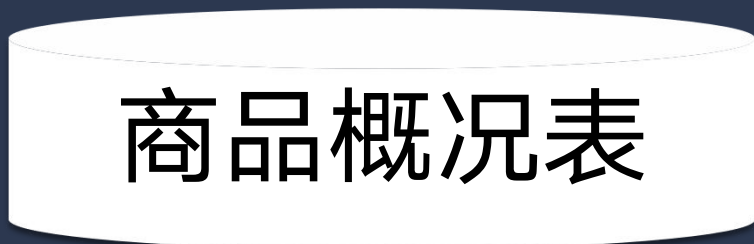
- ROI
- GMV

ABI实践-NL2SQL

- 多表问题（非对称问题）
 - LLM
 - Embedding
 - Ensemble+Rerank

ABI实践-NL2SQL

- 多表的问题->Embedding Finetune（冷启动）



基于LLM生成



问题列表
请提供最近一周内商详浏览量最高的商品名称和浏览量是多少？
能否给出销售模式为线上、渠道为电商的商品在过去一个月内的支付金额和下单件数？
可以提供节点类型为 'A'，且售后率最高的商品名称和售后率是多少吗？
请提供在组织架构树id为 '123' 的节点下，商详页退出率最低的商品名称和退出率是多少？
可以提供最近一周内商详访问支付转化率最高的商品名称和转化率是多少吗？

ABI实践-NL2SQL

- SQL生成
 - Prompt Engineering
 - Pre-Train、SFT
 - MAS

ABI实践-NL2SQL

- Prompt Engineer

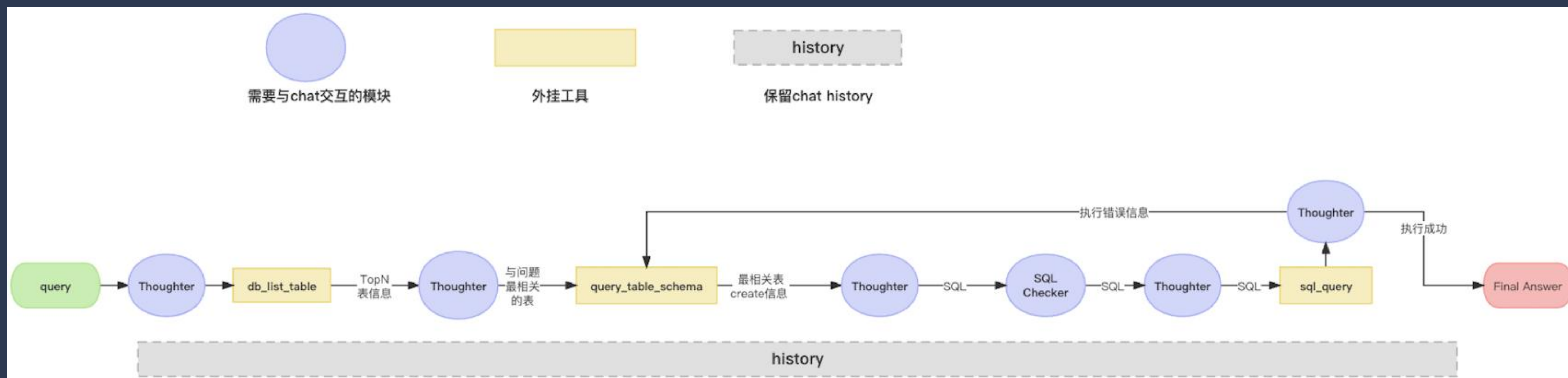
Prompt需要加上
数据字段说明以
及数据示例

```
Schema:
CREATE TABLE 内置_交易域_全部订单事实表 (
  订单创建时间 STRING ,
  订单更新时间 STRING ,
  下单时间 STRING ,
  下单日期 STRING ,
  下单周 STRING ,
  下单月 STRING ,
  订单渠道类型ID STRING ,
  下单时段 STRING COMMENT '下单时段, 如00,01',
  支付时间 STRING ,
  支付时段 STRING COMMENT '支付时段, 如00,01',
  支付周 STRING ,
  支付月 STRING ,
  支付日期 STRING ,
  支付方式名称 STRING COMMENT '支付方式名称(可有多)',
  订单完成时间 STRING ,
  ...
  组织名称 STRING ,
  门店编号 STRING ,
  满减邮折扣金额 BIGDECIMAL ,
  订单附加费 BIGDECIMAL COMMENT '订单附加费(商品附加费)',
  订单附加费抵扣金额 BIGDECIMAL COMMENT '订单附加费(商品附加费)抵扣金额',
  秒杀优惠金额 BIGDECIMAL ,
  售后应退积分金额 BIGDECIMAL ,
  售后应退余额 BIGDECIMAL
);

示例数据:[{'会员优惠金额': 0.0, '订单自定义信息': None, '门店编号': '23409992', '客户姓名': None, '礼品卡抵扣金额': 0.0, '开票信息电话号码': None}]
Question: 本月订单同比增长多少
```

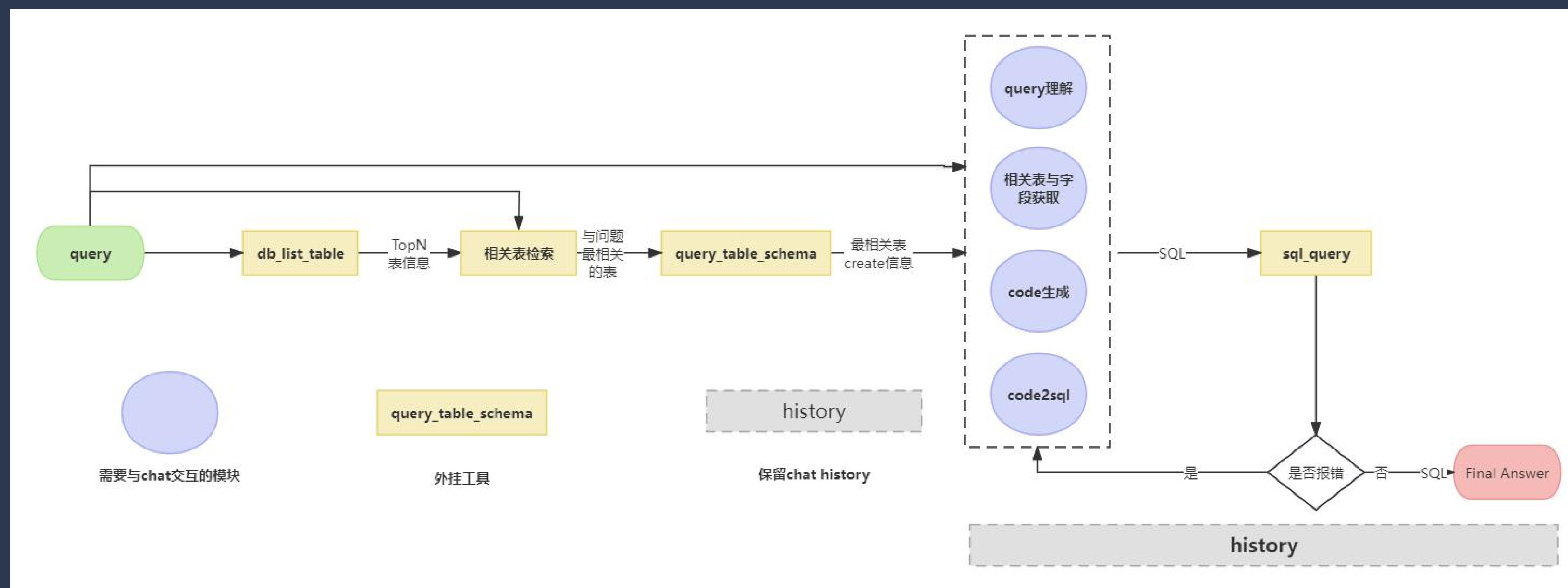
ABI实践-NL2SQL

- MAS



ABI实践-NL2SQL

- MAS



ABI实践-NL2SQL

最近一个月订单
相比上月环比变
化情况

213

保存

选择数据源

内置数据源

数据表

输入搜索关键字

内置_商品域_在售商...

内置_商品域_商品入...

内置_商品域_商品出...

内置_商品域_仓库维表

内置_商品域_库存规...

内置_商品域_库存商...

内置_配置域_自定义...

内置_角色域_客户收...

内置_角色域_微客所...

内置_角色域_微客基...

内置_角色域_微客等...

内置_财务域_分销业...

内置_配置域_安全维...

数据集

输入搜索关键字

1_1

5

4

3

2

1

SQL

运行 参数配置

1 SELECT

2 COUNT(*) AS current_month_total_orders,

3 (

4 SELECT COUNT(*)

5 FROM 内置_交易域_全部订单事实表

6 WHERE 下单月 = DATE_FORMAT(DATE_SUB(NOW(), INTERVAL 1 MONTH), '%Y年%m月')

7) AS last_month_total_orders,

8 (

9 COUNT(*) - (

10 SELECT COUNT(*)

11 FROM 内置_交易域_全部订单事实表

运行结果 历史记录

NO	NO	NO
current_month_total_orders	last_month_total_orders	change_percentage
976	1585	-38.42271293375394

智能生成SQL 写作 画图

智能生成SQL

((SELECT COUNT(*) FROM 内置_交易域_全部订单商品事实表 WHERE 订单创建时间 >= '2023-11-25' AND 订单创建时间 <= '2023-12-25') - COUNT(*) / COUNT(*) * 100 AS change FROM 内置_交易域_全部订单商品事实表;

使用 复制

您好, 请告诉我创作要求, 我可以帮您创作哦!

最近一个月订单相比上月环比变化情况如何

SELECT COUNT(*) AS current_month_total_orders, (SELECT COUNT(*) FROM 内置_交易域_全部订单事实表 WHERE 下单月 = DATE_FORMAT(DATE_SUB(NOW(), INTERVAL 1 MONTH), '%Y年%m月')) AS last_month_total_orders, (COUNT(*) - (SELECT COUNT(*) FROM 内置_交易域_全部订单事实表 WHERE 下单月 = DATE_FORMAT(DATE_SUB(NOW(), INTERVAL 1 MONTH), '%Y年%m月'))) / (SELECT COUNT(*) FROM 内置_交易域_全部订单事实表 WHERE 下单月 = DATE_FORMAT(DATE_SUB(NOW(), INTERVAL 1 MONTH), '%Y年%m月')) * 100 AS change_percentage FROM 内置_交易域_全部订单事实表 WHERE 下单月 = DATE_FORMAT(NOW(), '%Y年%m月');

使用 重做 复制

ABI实践-NL2SQL

最近一个月
订单同比去
年变化情况

32131231

选择数据源

内置数据源

数据表

输入搜索关键字

内置_商品域_在售商...

内置_商品域_商品入...

内置_商品域_商品出...

内置_商品域_仓库维表

内置_商品域_库存规...

数据集

输入搜索关键字

1_1

5

4

运行 参数配置

1 SELECT

2 COUNT(*) AS current_month_order_count,

3 (SELECT COUNT(*) FROM 内置_交易域_全部订单事实表 WHERE 下单月 = DATE_SUB(CURDATE(), INTERVAL 1 YEAR)) AS last_year_month_order_count,

4 (COUNT(*) - (SELECT COUNT(*) FROM 内置_交易域_全部订单事实表 WHERE 下单月 = DATE_SUB(CURDATE(), INTERVAL 1 YEAR))) / (SELECT COUNT(*) FROM 内置_交易域_全部订单事实表 WHERE 下单月 = DATE_SUB(CURDATE(), INTERVAL 1 YEAR)) * 100 AS change_percentage

5 FROM 内置_交易域_全部订单事实表

6 WHERE 下单月 = DATE_FORMAT(CURDATE(), '%Y年%m月');

运行结果 历史记录

current_month_order_count	last_year_month_order_count	change_percentage
976	0	-

智能生成SQL 写作 画图

最近一个月订单量同比去年变化情况

SELECT

COUNT(*) AS current_month_order_count,

(SELECT COUNT(*) FROM 内置_交易域_全部订单事实表 WHERE 下单月 = DATE_SUB(CURDATE(), INTERVAL 1 YEAR)) AS last_year_month_order_count,

(COUNT(*) - (SELECT COUNT(*) FROM 内置_交易域_全部订单事实表 WHERE 下单月 = DATE_SUB(CURDATE(), INTERVAL 1 YEAR))) / (SELECT COUNT(*) FROM 内置_交易域_全部订单事实表 WHERE 下单月 = DATE_SUB(CURDATE(), INTERVAL 1 YEAR)) * 100 AS change_percentage

FROM 内置_交易域_全部订单事实表

WHERE 下单月 = DATE_FORMAT(CURDATE(), '%Y年%m月');

文字创作 使用 重做 复制

ABI实践-NL2SQL

基于订单交易表
统计门店最近半
年整体GMV情
况如何

213

选择数据源

内置数据源

数据表

输入搜索关键字

内置_商品域_在售商...

内置_商品域_商品入...

内置_商品域_商品出...

内置_商品域_仓库维表

内置_商品域_库存规...

内置_商品域_库存商...

内置_配置域_自定义...

内置_角色域_客户收...

内置_角色域_微客所...

内置_角色域_微客基...

内置_角色域_微客等...

内置_财务域_分销业...

内置_配置域_安全通...

数据集

输入搜索关键字

1_1

5

4

3

2

1

SQL

出库分析

运行

参数配置

1 SELECT

2 SUM(订单总金额) AS 总体GMV,

3 SUM(订单实收金额) AS 实际GMV,

4 SUM(订单应收金额) AS 预期GMV

5 FROM

6 内置_交易域_内部订单支付每日统计事实表

7 WHERE

8 统计日期 >= '2023-06-28' AND 统计日期 <= '2023-12-25';

运行结果

历史记录

No	No	No
总体GMV	实际GMV	预期GMV
757667467.03	705584750.53	746024406.44

智能生成SQL

写作

画图

使用

复制

您好, 请告诉我创作要求, 我可以帮您创作哦!

最近一个月订单相比上月环比变化情况如何

SELECT
COUNT(*) AS previous_month_total_order
s,
(SELECT COUNT(*) FROM 内置_交易域_全
部订单事实表 WHERE 下单日期 >= '2023-12-01'
AND 下单日期 <= '2023-12-31') AS current_mo
nth_total_orders,
((SELECT COUNT(*) FROM 内置_交易域_全
部订单事实表 WHERE 下单日期 >= '2023-12-01'
AND 下单日期 <= '2023-12-31') - COUNT(*)) /
COUNT(*) * 100 AS change_percentage
FROM 内置_交易域_全部订单事实表
WHERE 下单日期 >= '2023-11-01' AND 下单
日期 <= '2023-11-30';

使用

复制

您好, 请告诉我创作要求, 我可以帮您创作哦!

基于订单交易表统计门店最近半年整体GMV
情况如何

SELECT
SUM(订单总金额) AS 总体GMV,
SUM(订单实收金额) AS 实际GMV,
SUM(订单应收金额) AS 预期GMV
FROM
内置_交易域_内部订单支付每日统计事实表
WHERE
统计日期 >= '2023-06-28' AND 统计日期 <
= '2023-12-25';

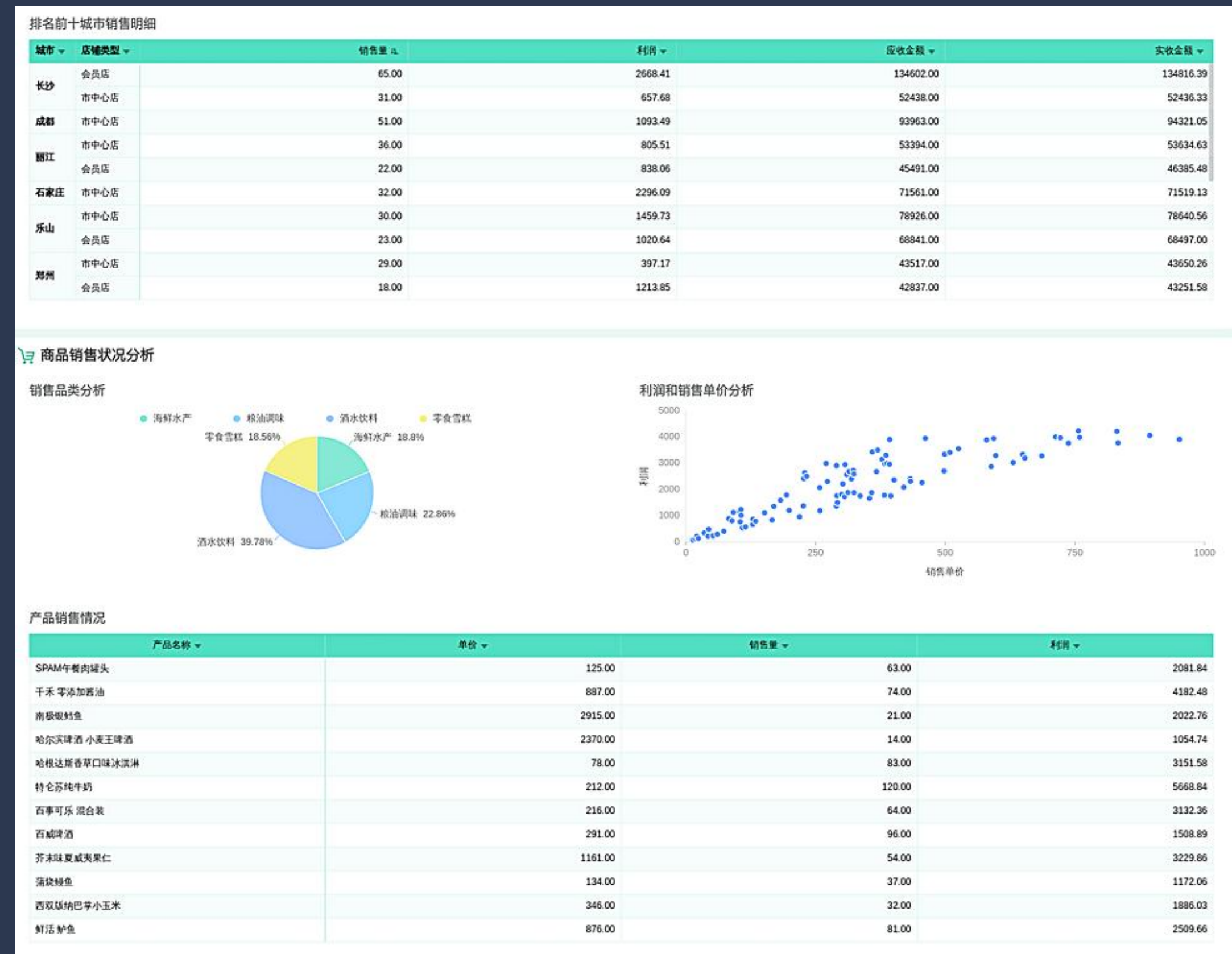
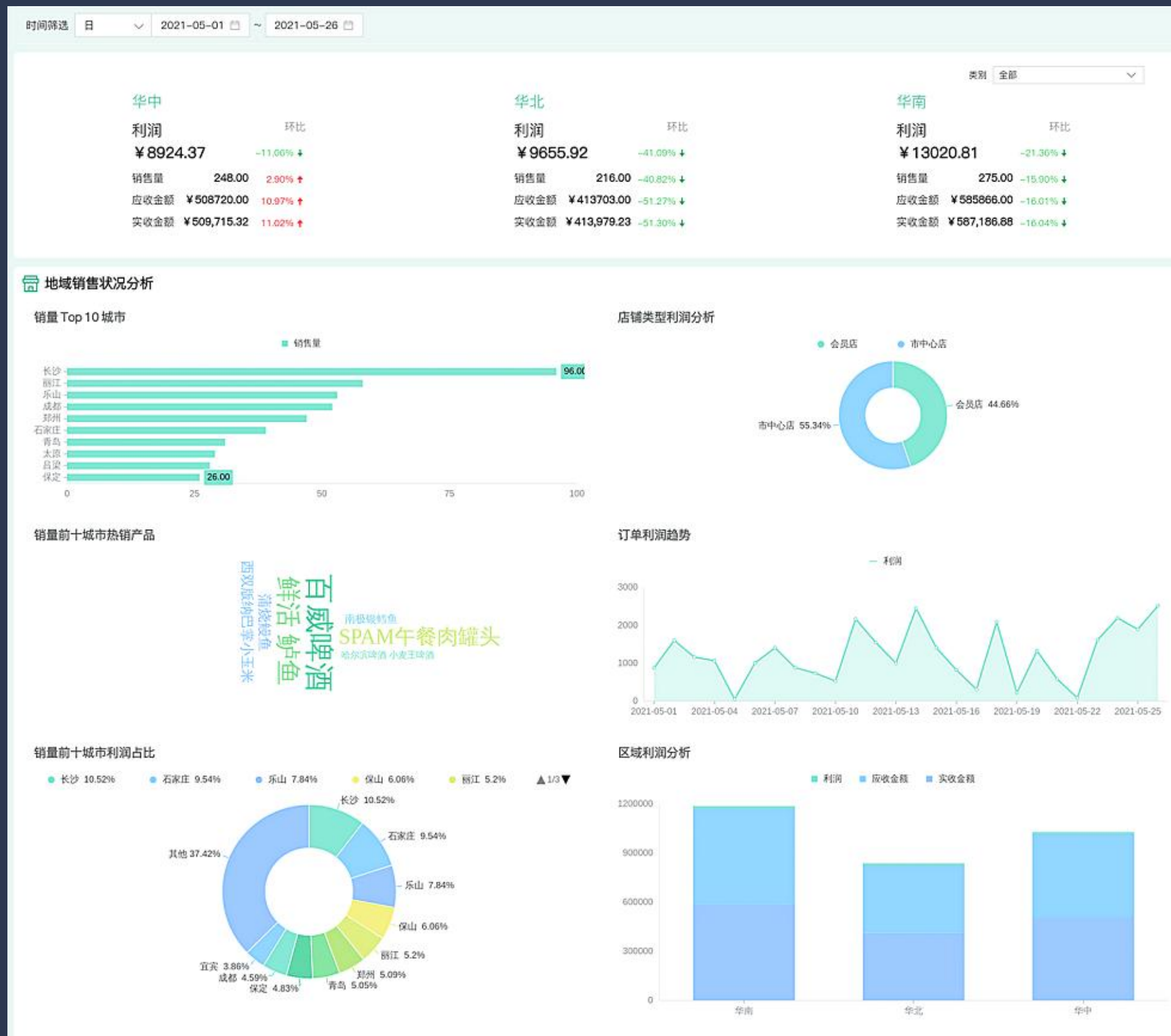
使用

重做

复制

您好, 请告诉我创作要求, 我可以帮您创作哦!

ABI实践-图表生成



ABI实践-图表生成

基于具体问题

- 最近30天的订单情况
- 商品的库存变化情况
- 最近活动的转化率

基于抽象问题

- 针对老客户召回，帮我做相关分析
- 根据订单表，做一些有意思的分析

ABI实践-图表生成

- 维度与度量的选择
- 字段粒度的选择
- 字段的聚合操作（sum、avg等等）
- 对抽象问题的理解
- 字段之间的组合关系

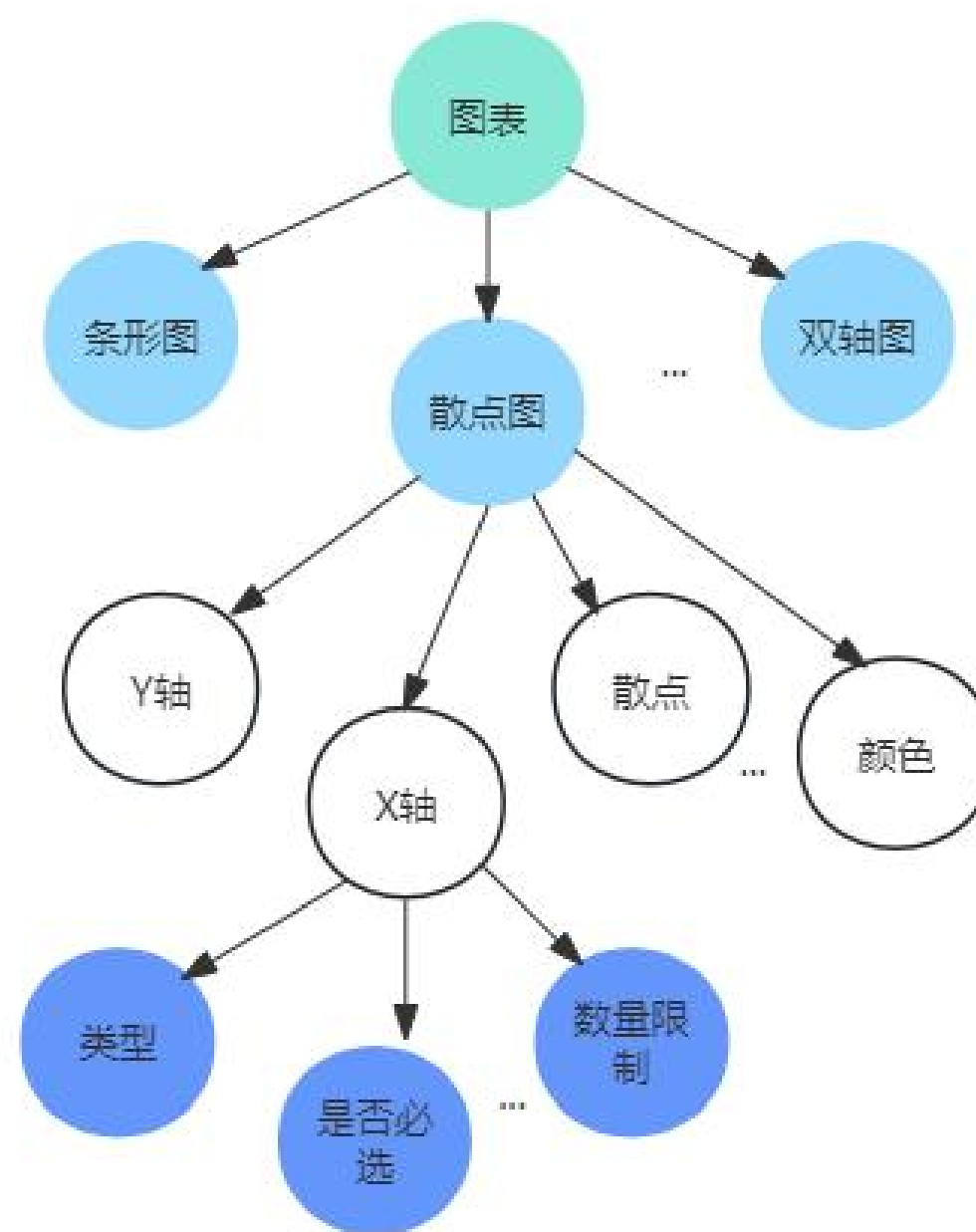
有没有老客户召回相关的数据分析？



```
{
  "solved": "true",
  "reason": "根据给定的配置，点击曝光表中存在度量召回沉默客户数和维度月份，\n              可以解决用户的问题并已完成配置生成",
  "config": {
    "type": "stackBar",
    "query": {
      "datasetId": 4075154027848702,
      "queryFields": {
        "dimensions": [{
          "fieldClassify": "dimension",
          "fieldType": "datetime",
          "datasetId": 4075154027848702,
          "fieldId": "月份",
          "title": "月份",
          "dateType": "day",
          "name": "月份"
        }],
        "measures": [{
          "fieldClassify": "measure",
          "fieldType": "number",
          "datasetId": 4075154027848702,
          "fieldId": "召回沉默客户数",
          "title": "召回沉默客户数",
          "aggrType": "sum",
          "name": "召回沉默客户数"
        }]
      },
      "aggrMode": 1,
      "submit": 1,
      "limit": 100
    },
    "aggrMode": 1,
    "submit": 1,
    "limit": 100
  }
}
```


ABI实践-图表生成

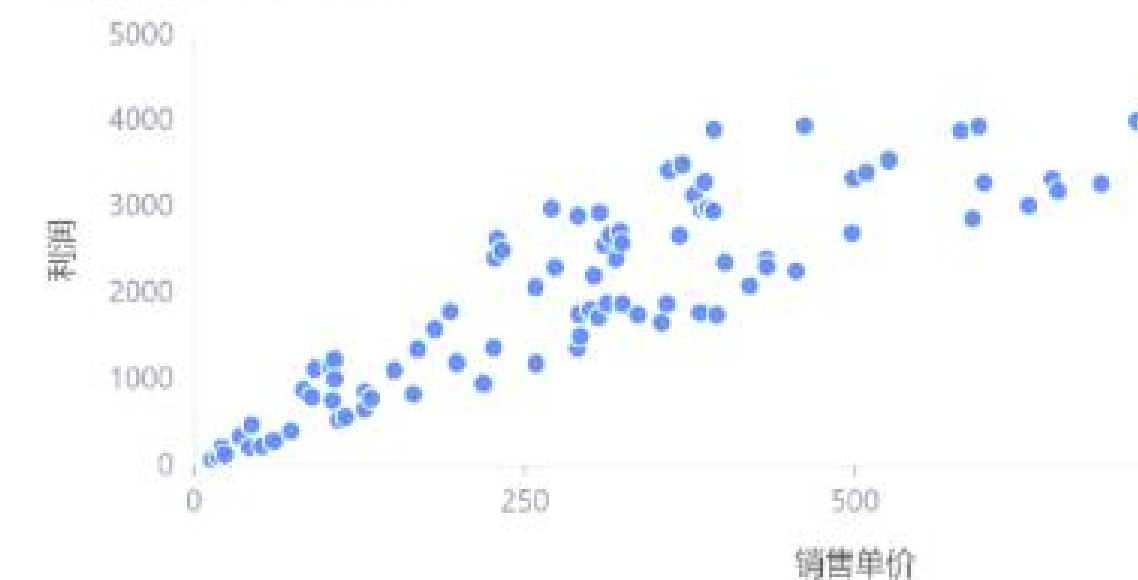
店铺类型利润分析



订单利润趋势



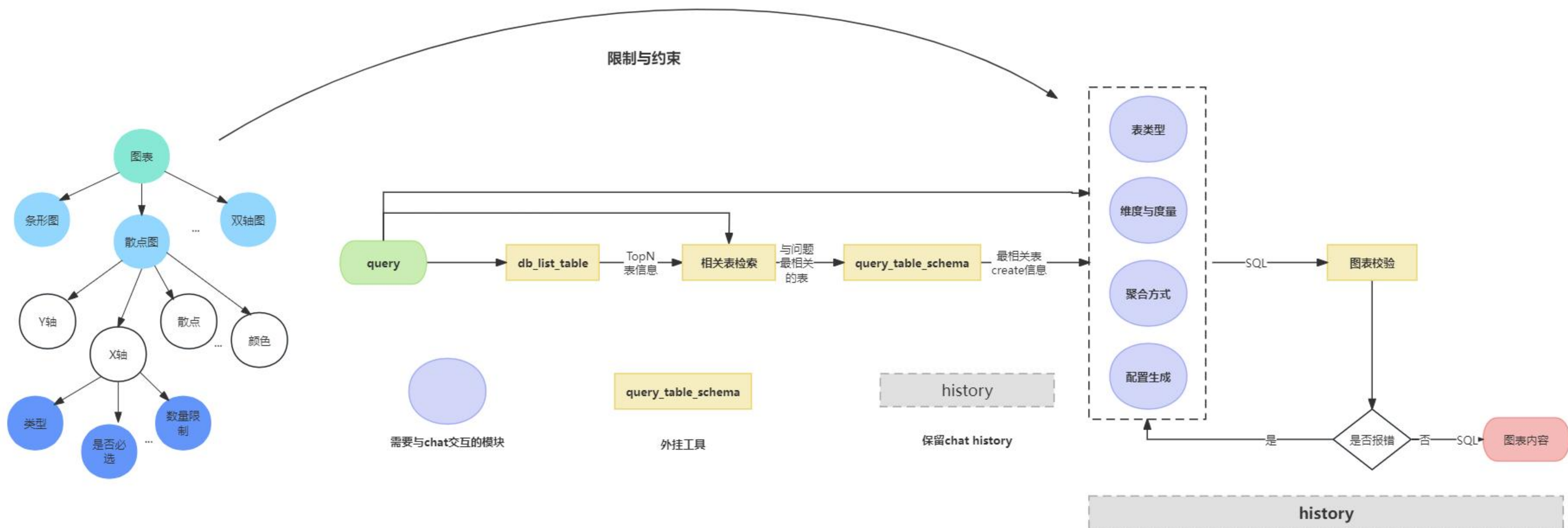
利润和销售单价分析



ABI实践-图表生成

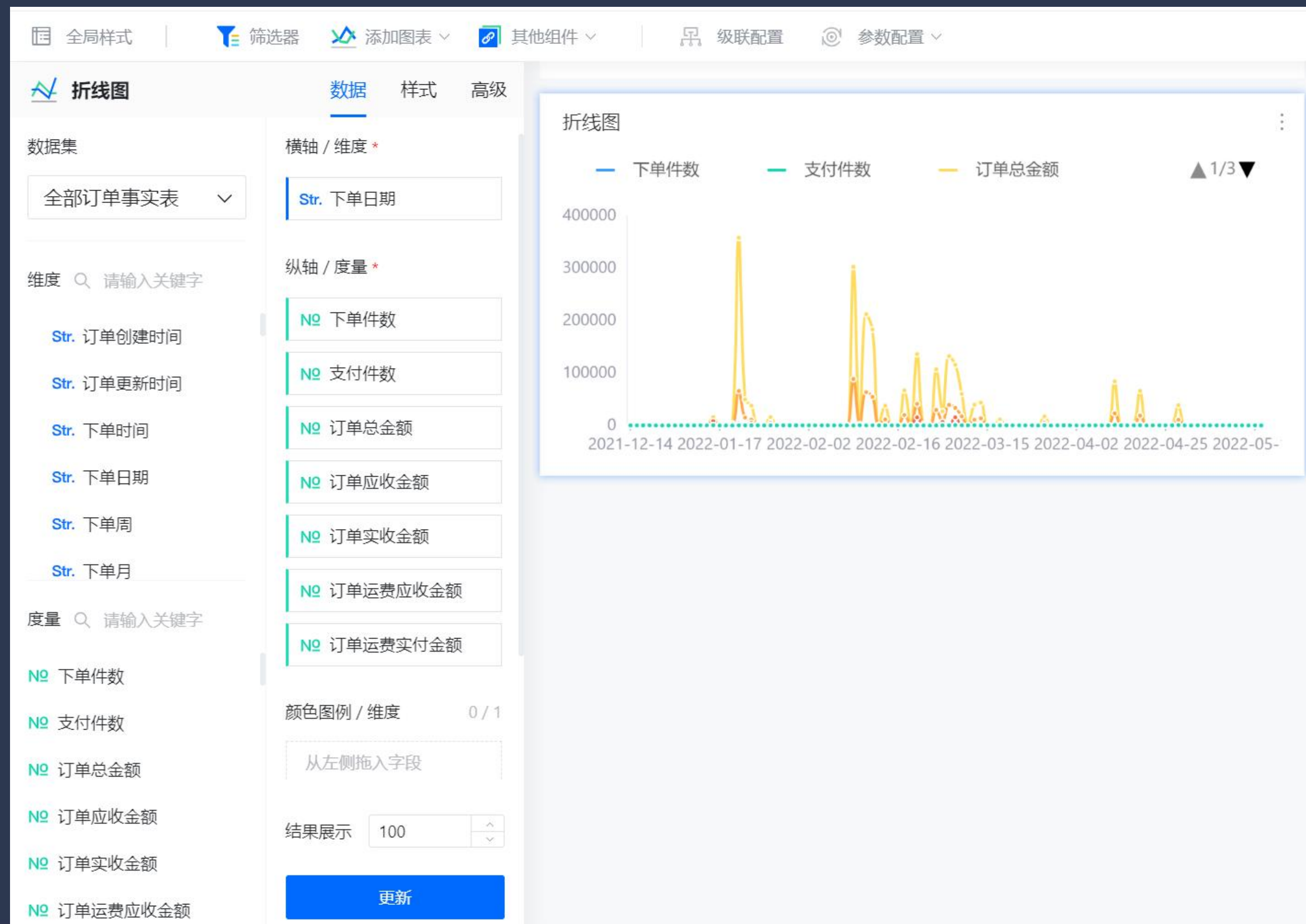
- 图表生成
 - Copilot
 - SFT
 - MAS

ABI实践-图表生成



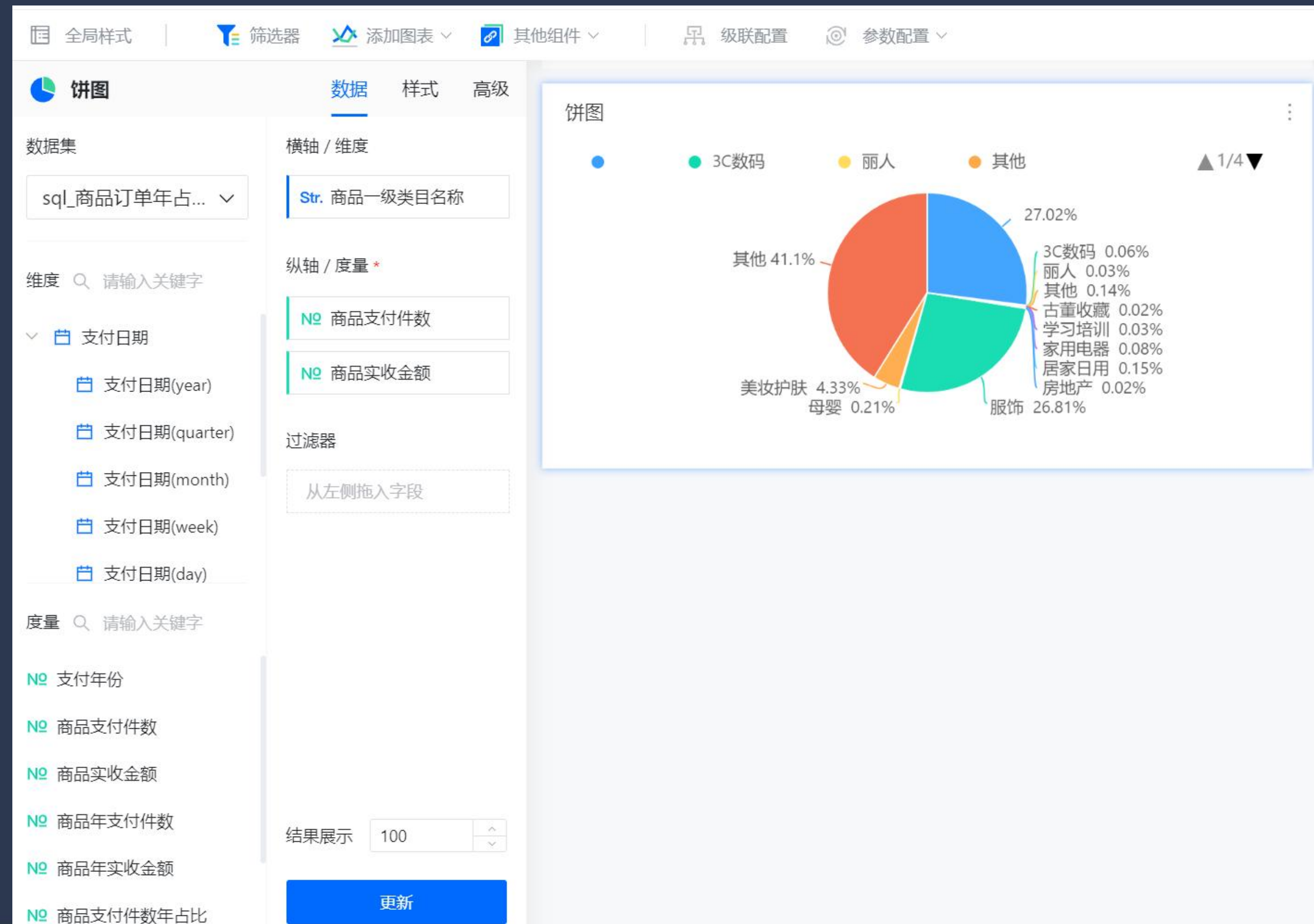
ABI实践-图表生成

用折线图展示一下最近半年的订单情况



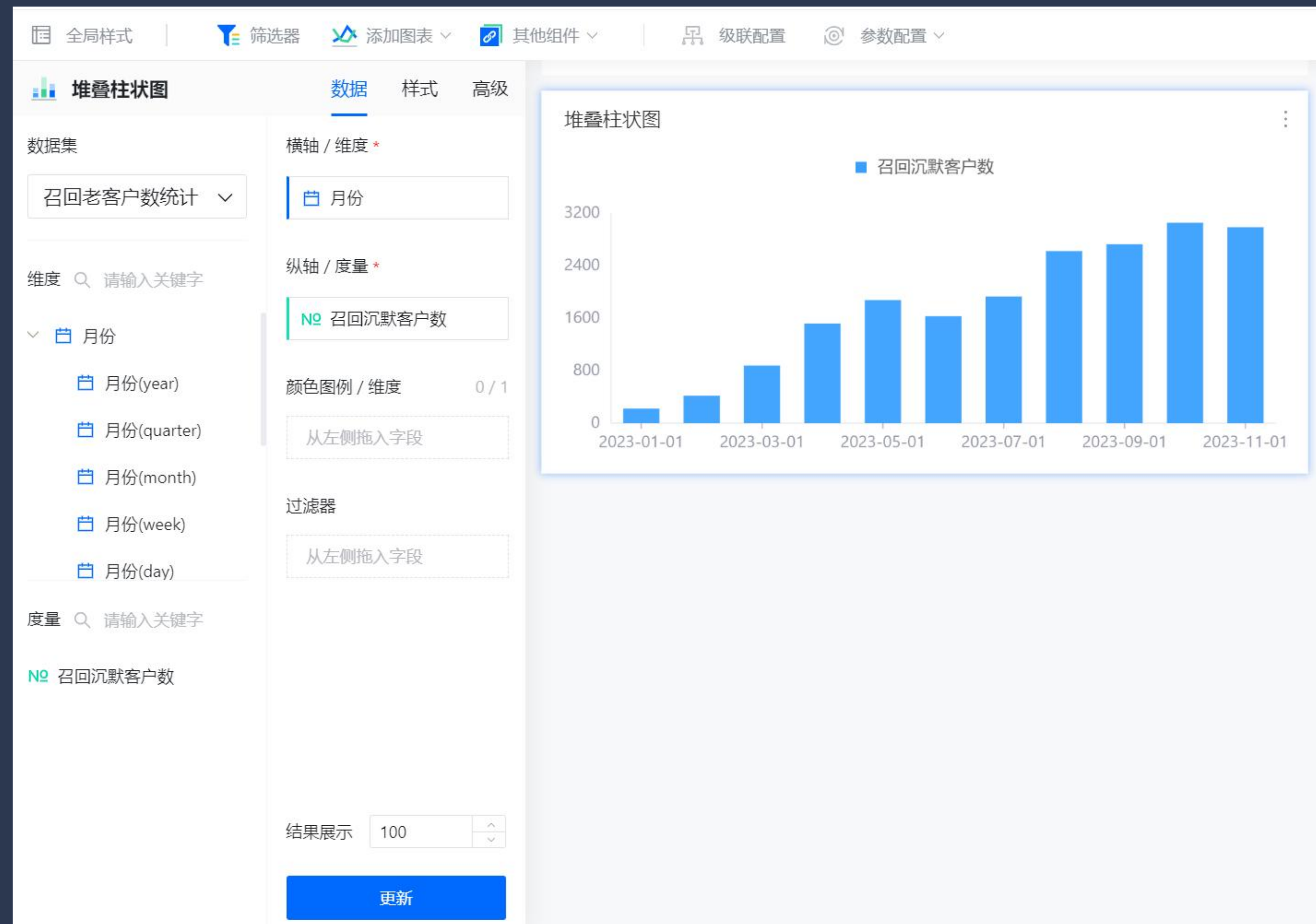
ABI实践-图表生成

饼图展示订单结合商品类目的分析



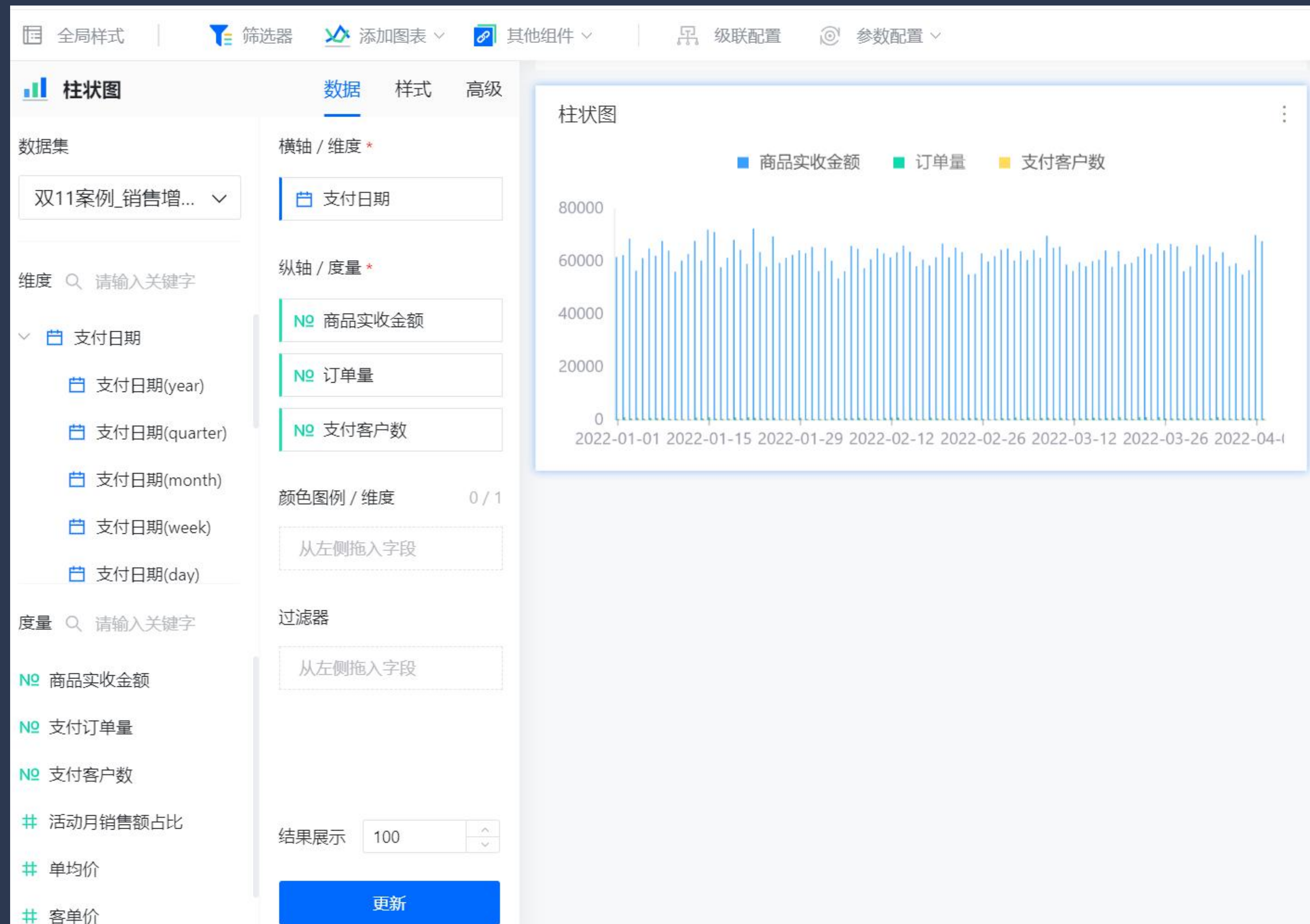
ABI实践-图表生成

有没有老客户召回相关的数据分析



ABI实践-图表生成

基于订单表，帮我生成有价值的分析



ABI实践-数据解读

多视角分析

统计视角解
读



业务视角解
读

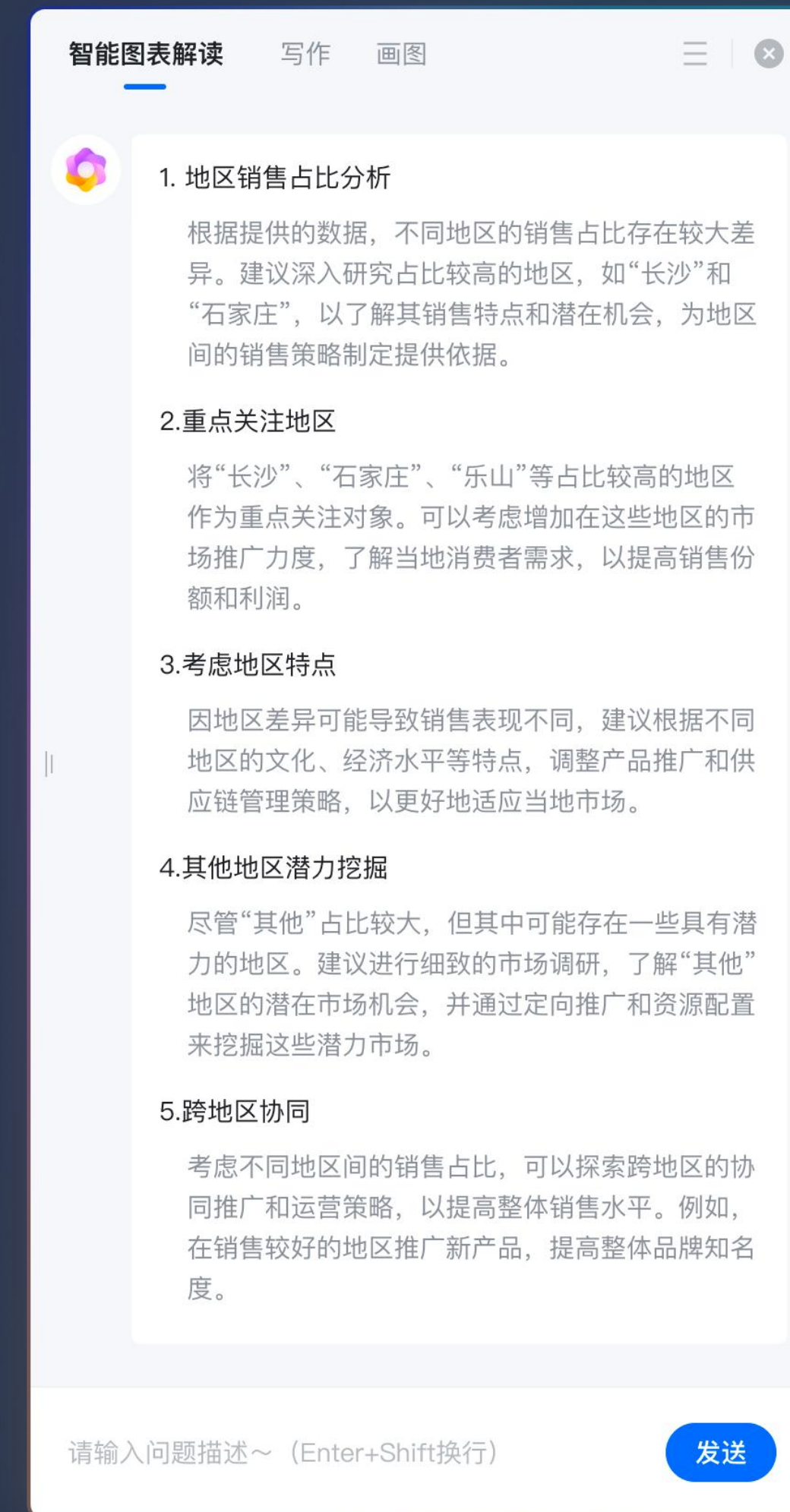
多表联动

单表解读



多表解读

ABI实践-数据解读



总结与展望

ABI实践方向



展望

- 宽表、指标建设
- 不同场景、任务下的效果提升
- 支持模糊化的需求产出
- 基于行业know-how的数据分析、洞察

THANKS

软件正在重新定义世界

Software Is Redefining The World