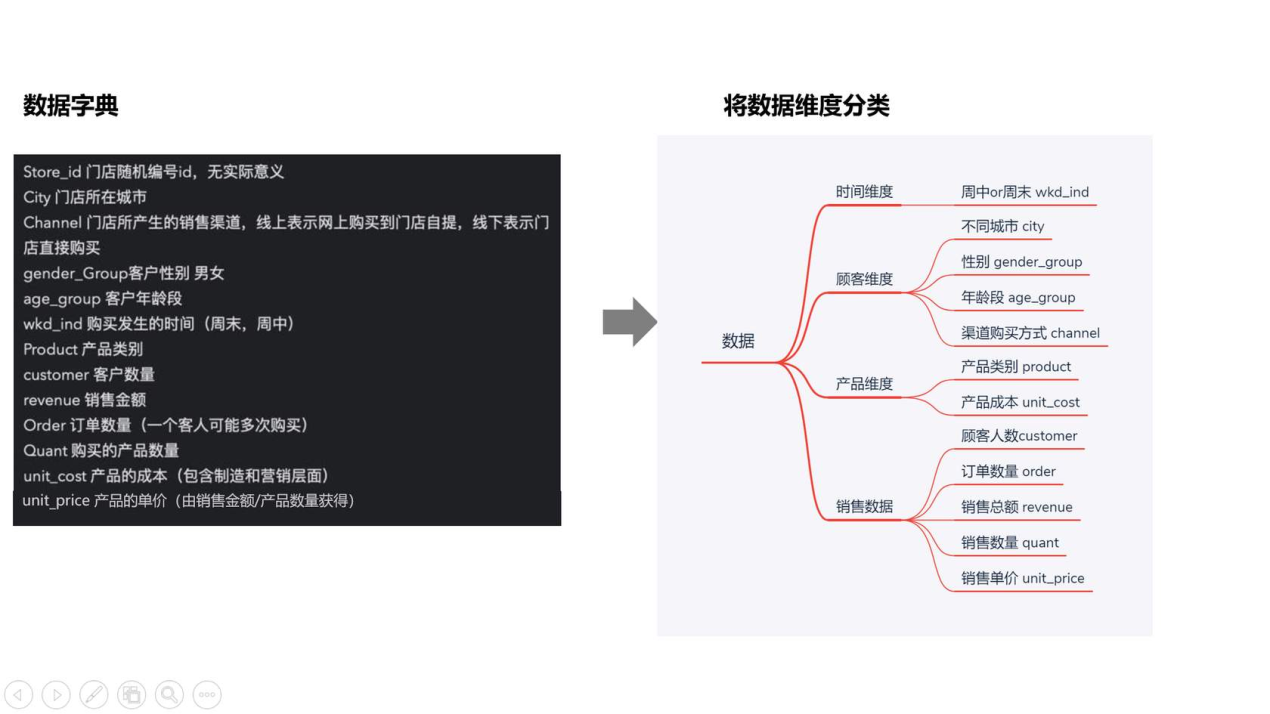
Week5：优衣库销售产品分析 plt.rcparams[‘font.sans-serif’]=[‘simhei’]



一、问题界定

整体销售情况随着时间的变化是怎样的？

不同产品的销售情况是怎样的？

顾客偏爱哪一种购买方式？

销售额和产品成本之间的关系怎么样？

二、数据查看、简单清洗与整理

三、数据探索与可视化

1、整体销售情况随着时间的变化是怎样的？

（1）思路：

①变量选择：

销售情况可以用销售数量quant和销售金额revenue体现，时间则用周末/周中wkd\_ind表示

②数据关系：

对比销售数据在周中和周末以及线上和线下分组下的分布情况

③图表：柱状图

（2）具体步骤：

①在数据中添加用于计数的列count，每行赋值1

②查看数据分布情况

（3）代码及代码结果：

（4）业务结论：

2、不同产品的销售情况是怎样的？

（1）思路：

①变量选择：

不同产品用产品类别product来区分，销售情况用销售数量quant和销售总额revenue来表示

②数据关系：

对比不同类别产品在销售数量和销售总额上的区别

③图表：柱状图

（2）具体步骤：

（3）代码及代码结果：

（4）业务结论：

3、顾客偏爱哪一种购买方式？

思路：

①变量选择：细分为什么样的顾客偏好什么样的购买方式：

顾客数量用customer的合计来表示，顾客可以用所在城市city、性别gender\_Group、年龄段age\_group来分类，购买方式用购买渠道channel来表示

②数据关系：

对比不同特征的顾客在不同渠道上的分布

③图表：柱状图

（2）具体步骤：

（3）代码及代码结果：

（4）业务结论：

对于渠道来说：绝大部分的城市偏好线下购买，只有广州是唯一线上超过线下的地方；全年龄段以及无论男女都更偏好在线下商店购买优衣库的产品

对于城市来说：深圳、杭州和武汉是订单数最多的城市，均在3500以上，同时深圳和杭州几乎全是线下销售

性别和年龄段：同时男女客户均集中分布在20~44岁之间，其中女性远多于男性，超过部分主要来自线下购物

4、销售额和产品成本之间的关系怎么样？

思路：

①变量选择：

销售额可以用销售金额revenue和销售数量quant来度量，产品成本使用unit\_cost表示

②数据关系：

分析销售金融和销售数量与产品成本的相关关系

③图表：

热力图

散点图

相关系数表

（2）具体步骤：

（4）业务结论：

通过是优衣库销售数据的分析，我们得到以下结论：

·数据主要集中在深圳、杭州和武汉等城市，北京和南京数据较少

·无论性别如何，和年龄如何，选择线下购买的客户都远高于线上购买的

·优衣库的客户年龄主要分布在20~44岁之间，且女性多于男性

·周末的客户数量和销售总额显著高于工作日的，但是销售额均值超出值并不多

·T恤、当今新品是最畅销的产品，T恤和配件是利润总额最高的商品，但是配件的单件利润要显著高