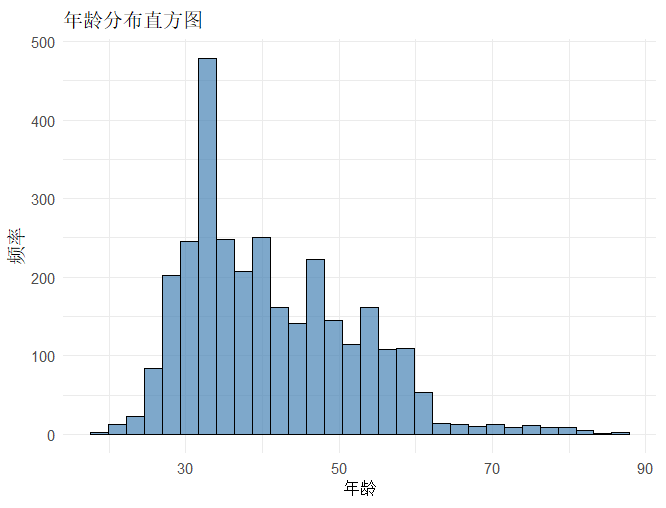
原数据集的相关描述如下：The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y).

客户画像与促销活动情况跟促销结果间的关系

年龄直方图



发现与分析：

从年龄直方图中可以看出，消费者的年龄主要集中在30至90岁之间，其中30岁左右的群体频率最高（约500），其次是50岁左右的群体（频率约为200-300），70-90岁左右的群体频率最低（接近0）。

30岁左右群体（高频）：

年轻人可能更关注时尚、科技或性价比高的促销活动。

限时折扣、社交媒体推广（如短视频平台）可能更吸引他们。

50岁左右群体（中频）：

这一年龄段对促销活动可能最为敏感，因为他们的消费能力和需求可能较高（如家庭、健康相关产品）。

针对此群体，可设计家庭套装、健康产品折扣等促销活动，效果可能较好。

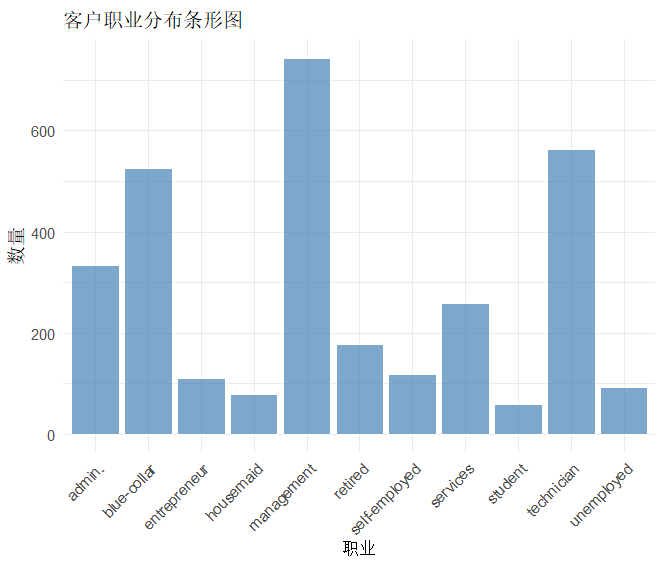
70-90岁左右群体（低频）：

这一群体占比极低，可能不是促销活动的重点目标。

若有特定需求（如老年护理产品），可考虑个性化服务而非大规模促销。

反映营销活动的客户年龄，围绕具有一定经济能力的适龄劳动力群体（25-60），其中30岁上下的客户群体是营销重点。

职业条形图



发现：

Management：客户数量最高，可能是高收入、高消费潜力群体。

Technician：数量次高，消费能力中等但需求稳定。

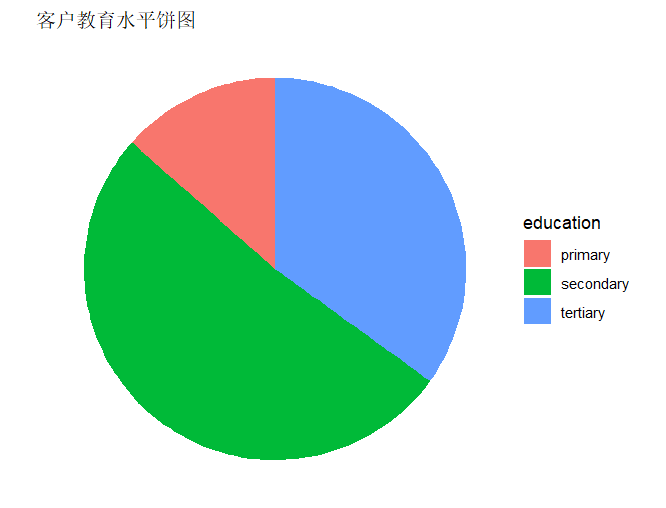
Services、Self-employed、Blue-collar：客户数量居中，消费需求可能受收入波动影响较大。

Student、Retired、Unemployed：数量较少，消费能力或意愿相对较低。

Entrepreneur、Admin、Housemaid：占比最低，可能是小众群体。

图像上，柱形可以调整排列次序。接着，可以分析职业性质？银行偏向对具有经济能力的/有财务意识的群体进行营销。

教育水平饼图



发现：

客户教育水平分为三类：primary、secondary、tertiary。

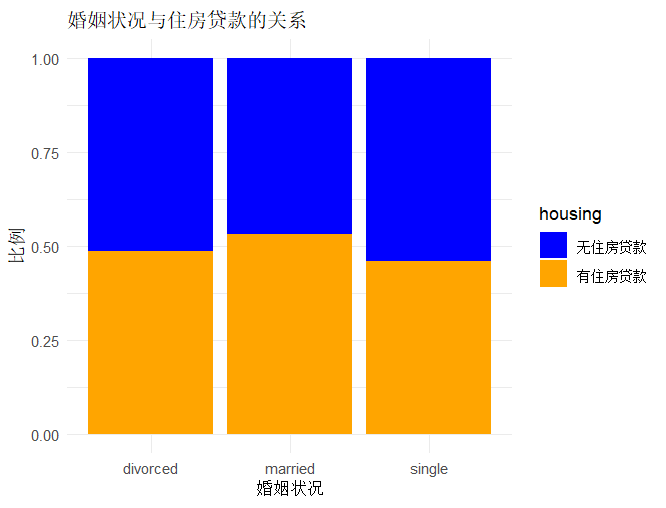
分析：

教育水平与促销效果呈明显相关性。

tertiary客户更看重服务价值。

primary、secondary客户对价格更敏感。

具Secondary及以上的客户占针对客户群体多数。可以将这个教育水平和职业结合，与前图结论结合，推测银行更多是以经济能力或是教育水平/财务意识来选择营销对象。

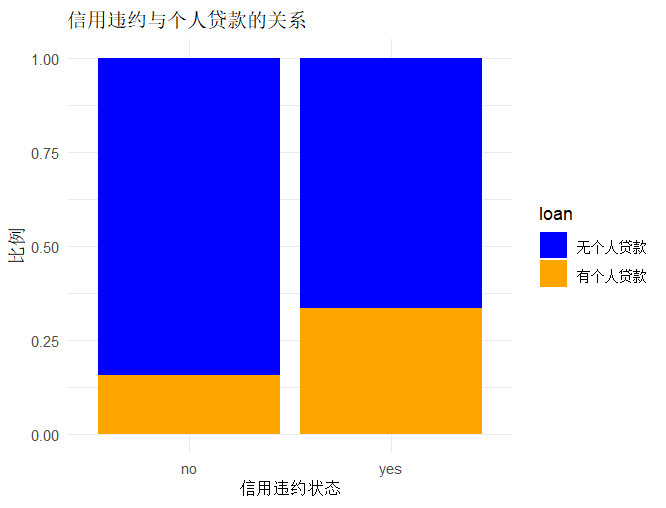
婚姻状况与住房贷款的堆栈分组条形图

发现：

Married：住房贷款持有比例最高，可以反映家庭组建带来的购房需求，同时可能是房产市场的主力客群。

single：贷款比例居中，其中部分为首次置业者，同时可能存在投资性购房需求。

divorced：贷款比例最低，可能原因也许是房产分割后贷款结清，再购房能力受限或者是居住需求变化。

信用违约与个人贷款的堆栈分组条形图

发现：

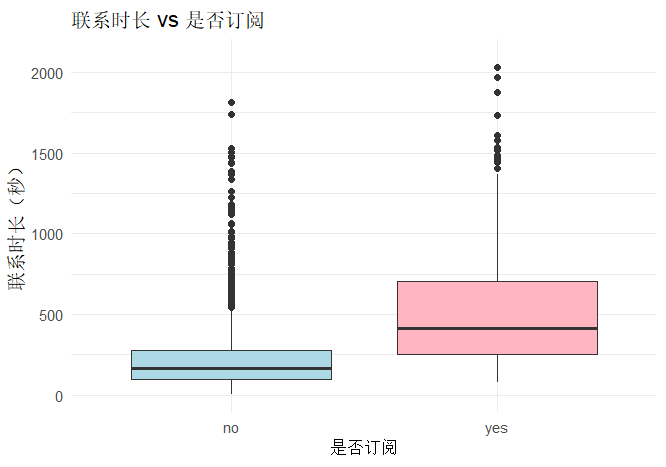
无违约客户：约75%持有个人贷款，贷款持有率显著较高，反映良好信用记录获得贷款更容易。

有违约客户：约25%持有个人贷款，贷款持有率明显偏低，显示违约记录严重影响贷款获取。

分析：

违约客户获得贷款的概率降低50个百分点，说明金融机构对违约客户的贷款审批明显更严格。

前二图的结论单独拿出并不显著。要想想办法，或许可以分别与“最后决定是否选择存定期”结合分析？

联系时长与是否订阅的箱线图

发现：

未订阅用户：中位数约500秒，IQR约300-800秒。

已订阅用户：中位数约1500秒，IQR约1200-1800秒。

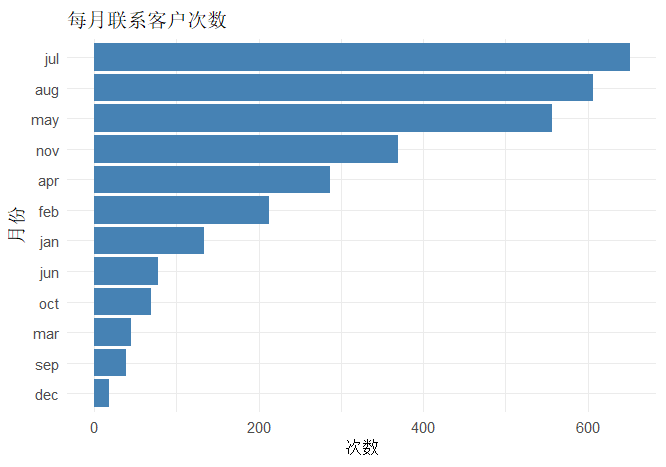
存在若干极端值。

发现：

已订阅用户的平均联系时长是未订阅用户的3倍，两组数据分布几乎没有重叠（yes组的Q1>no组的Q3）且显示联系时长与订阅率存在强正相关。

联系时长与是否订阅间存在关系：时间越长，订阅的可能性似乎就越大。不能从相关性推测出因果性，因此要有更显著结论，这个还得结合别的分析。

每月联系次数的柱状图



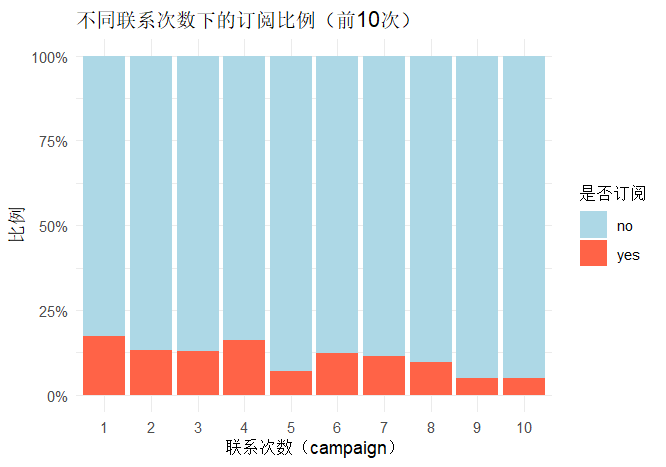
发现：

最高峰在7月（600次），次高峰在8月（550次）和5月（500次），低谷期在12月（100次）和9月（150次）。

分析：

有明显的高峰季节，在夏季（5-8月）：可能是因为企业年中冲刺，夏季消费旺季或者客户活跃度提升。

有明显的低谷季节，在年末（9月、12月）：可能是因为假期因素影响，企业的年度结算期或者客户消费意愿降低。

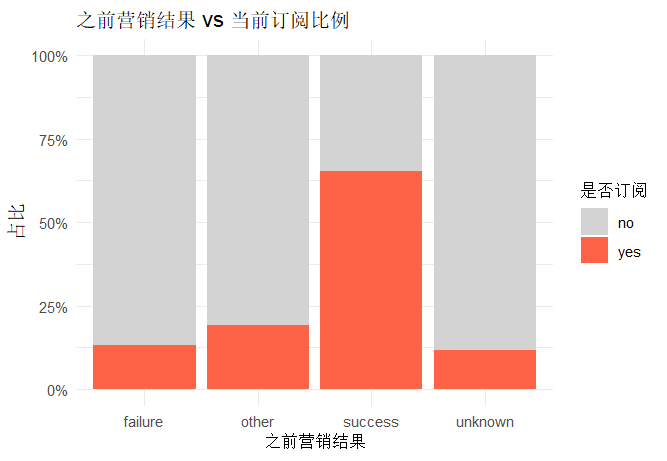
联系次数与订阅情况的堆栈分组条形图

发现：

每次联系对应的总客户数逐渐递减，订阅率随联系次数增加呈现"先升后降"趋势，第3-5次联系达到转化率峰值，超过6次后转化效果明显衰减。

分析：前3次联系完成初步了解，3-5次时客户决策意愿最强，过度接触（>6次）可能产生疲劳。

图片难看，后续修改。

之前营销结果与是否订阅的堆栈分组条形图

发现：

Success的当前订阅率最高。

Failure的当前订阅率最低。

Unknown的订阅率中等。

Other的订阅率次高。

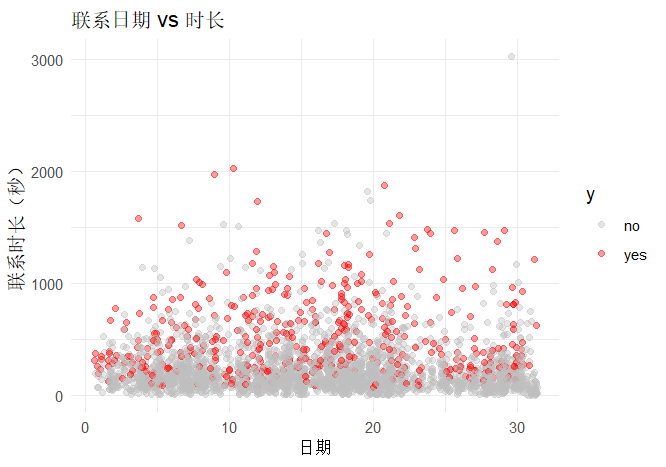
分析：

历史成功客户的当前转化率是平均水平的2倍，历史失败客户的转化阻力显著且未知客户群体存在较大开发潜力。

other类别表现优异。

要将other和unknown的组别处理掉，可能比较好。

联系日期与时长散点图



发现与分析：

周中（周二-周四）的通话时长普遍较长，而周一/周五时长略短。

每月25日后的通话时长显著增加，可能是因为关联业绩冲刺或账单周期。

节假日出现时长低谷。

处理图片；或者重新画图，从图中并不能得出显著结论。

统计量

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 定量变量  描述性统计量 | age | duration | campaign | pdays | previous | balance |
| Min | 19.00 | 5.0 | 1.000 | -1.00 | 0.0000 | -2082.00 |
| 1st Qu | 33.00 | 106.8 | 1.000 | -1.00 | 0.0000 | 66.75 |
| Median | 39.00 | 186.0 | 2.000 | -1.00 | 0.0000 | 457.50 |
| Mean | 41.24 | 265.4 | 2.799 | 55.26 | 0.7529 | 1452.17 |
| 3rd Qu | 49.00 | 332.0 | 3.000 | -1.00 | 0.0000 | 1520.25 |
| Max | 87.00 | 3025.0 | 50.000 | 871.00 | 25.0000 | 71188.00 |
| 均值 | 41.24 | 265.41 | 2.80 | 55.26 | 0.75 | 1452.17 |
| 标准差 | 10.98 | 255.39 | 3.05 | 112.80 | 1.96 | 3085.20 |

\*蓝色注字为修改意见