RFM：

RFM分析用于客户细分，依据三个指标：最近一次购买时间（Recency）、购买频率（Frequency）、以及购买金额（Monetary value）。

数据预处理：先将有数据缺失值的行删除，因为没有规律和参考，无法将其缺失值补充。然后将账户间转账记录和收入记录都分别单独存储为csv文件用做其他用途，保留商户交易记录。最后根据商户类型将商户分成了十大类（分类依据根据商户名称/商户信息决定，查了信息，可能需要文献？）分别是：Bank，Indoor Entertainment，Fashion & Clothing，Kids，Arts & Craft，Grocery，Daily，Large Retailers，Medical，Food & Drink。

针对每个类型计算RFM指标：

特定时间都为该数据集中表示的总时间（一年）

Recency（最近一次购买时间）：计算每个客户自最后一次购买以来的天数。

Frequency（购买频率）：计算每个客户在特定时间内的购买次数。

Monetary（购买金额）：计算每个客户在特定时间内的总购买金额。

基于RFM指标，每个指标被分为四个等级（四分位数），每个客户根据他们的行为在每个指标上获得得分（1到4）。得分越低代表客户表现越好（例如，1表示在该指标上表现最好）。

RFM客户细分格式如下：

超级客户 (Best Customers): 这些客户通常被标记为“111”，表示他们最近购买（Recency得分为1），购买频率高（Frequency得分为1），并且消费金额高（Monetary得分为1）。这类客户的忠诚度高，是保持关系的重点。

高价值客户：

忠诚价值客户 (Loyal High-Value Customers): 如果客户的Frequency和Monetary得分都是1，但Recency得分不是1，这意味着他们购买频率高并且消费金额也高，但有一段时间没有购买了。这类客户对企业来说非常重要，因为他们表现出高度的忠诚和购买力，但可能需要通过促销或特别关注来鼓励他们重新进行购买。

新兴价值客户 (New High-Value Customers): 如果客户的Recency和Monetary得分都是1，但Frequency得分不是1，这表明他们最近进行了购买，且购买金额较大，但购买频率不高。这类客户有成为忠诚客户的潜力，企业应该采取措施促使他们更频繁地购买。

新兴忠诚客户 (New Loyal Customers): 如果客户的Recency和Frequency得分都是1，但Monetary得分不是1，这说明他们最近有购买，且购买频率高，但单次购买金额不大。这些客户显示出对品牌的忠诚，可能是通过小额但频繁的购买建立起来的。对这些客户的策略应该是增加他们的平均购买金额。

忠诚客户 (Loyal Customers): 这类客户的Frequency得分通常为1，显示他们购买频率很高。对这类客户，企业可能会考虑推出忠诚度计划或促销活动以进一步增加其粘性。

大额消费客户 (Big Spenders): Monetary得分为1的客户。他们每次购买可能金额较大，但频率和最近一次购买时间可能不理想。对这类客户，企业可能会推送高价值产品的促销信息。

近期客户 (New Customers): Recency得分为1的客户，表示他们是最近才进行购买。这些客户有潜力转化为忠诚客户，因此需要通过优质服务和产品来维持他们的兴趣。

需要激活的客户 (At Risk Customers): 这些客户在所有三个维度的得分可能不是最佳，比如“322”或“233”。他们以前可能购买频繁，但已经有一段时间没有购买了。企业需要采取措施重新激活这些客户的购买欲望。

流失客户 (Lost Customers): 得分如“444”表示这些客户很久没有购买，购买频率低且消费金额低。这类客户可能已经转向竞争对手或不再对产品感兴趣。

之后根据分类情况进行对应的反馈/服务设计

可视化我之后考虑要不要做

CLV：

CLV也是先计算RFM指标，但这回根据RFM分析的结果来看，只分析了最值得分析的一个商户类型，Large Retailers。然后计算了客户年龄（T），较长的T值可能意味着客户已经有了较长的购买历史。

基于RFM的指标，使用BG/NBD（Beta Geometric/Negative Binomial Distribution）模型来预测客户未来的购买行为。但准确率不高，没有太多参考价值。

然后使用Gamma-Gamma模型来估计客户在未来交易中的平均利润。

使用BG/NBD和Gamma-Gamma模型的输出来计算每个客户的预期终身价值。

最终CLV数值越高，客户的价值也就越大。可以针对结果来关注某些高价值客户（这个标准是依托于商户本身流水的，可以看看有没有文献提供相关信息，以便于找到分辨高价值的界限，或者还是用四分位数来区分价值）。